

Time-Varying Water Temperature Modelling of Steam Distillation Pilot Plant using NARX-based Binary Particle Swarm Optimisation Structure Selection

Najidah Hambali, Mohd Nasir Taib, Ahmad Ihsan Mohd Yassin, Mohd Hezri Fazalul Rahiman*

Abstract—Many studies in current years has concentrated on both linear and nonlinear modelling in the real nonlinear system applications. This study reports a nonlinear modelling for a time-varying process of water temperature by utilising a Binary Particle Swarm Optimisation (BPSO) algorithm based on Nonlinear Auto-Regressive with eXogenous input (NARX) structure. The model structure selection of polynomial NARX has been concentrated on BPSO algorithm for system identification of Steam Distillation Pilot Plant (SDPP). Several model's selection criteria such as Akaike Information Criterion (AIC), Model Descriptor Length (MDL), and Final Prediction Error (FPE) were investigated. The results demonstrated that all criterion models were considered valid and accurate representations of the system. The accuracy was evaluated by the high R-squared, small MSE value and passed all the correlation and histogram tests.

Index Terms—System Identification, NARX, Particle Swarm Optimisation, Distillation Column, Temperature.

I. INTRODUCTION

A prediction model can be achieved from system identification based on input and output with the implementation of mathematical model creation process [1]. The model creation needs the input and output data without the former knowledge of the system [2]. A linear and nonlinear modelling are included in the system identification [3], [4].

An essential oil extractions [5]–[8] and a waste water treatment [9] can be implemented by employing a distillation column system, which is one of the popular methods. The nonlinear dynamic behaviour has been displayed by a broad application of the distillation column in chemical operations [10]–[12]. Auto-Regressive with eXogenous input (ARX) [9], [13]–[15], Auto-Regressive Moving Average with eXogenous input (ARMAX) [16] and Nonlinear Auto-Regressive with eXogenous input (NARX) [17],[18] were

among the linear and nonlinear modelling that has been utilised for the distillation column identification. In addition, the time-varying dynamic behaviour is a growingly crucial area in nonlinear system identification due to the substantial challenges type that determined the precise quality forecasting [19]. Various efforts have been established for the nonlinear and time-varying NARX model development [20]–[22].

Unfortunately, some restriction due to the robustness of particular input array has been presented in a linear model for the nonlinear system [4]. Furthermore, the nonlinear model showed greater presentation, potentiality and decent compared to lesser correctness and inadequate fit resulted in linear model modelling for nonlinear system representation [9], [23]. The adequacy of nonlinear dynamic behaviours of the nonlinear system is demonstrated substantially and can be simply applied to control design. Furthermore, a time-varying and nonlinearity behaviour of the real system such as distillation column has been a serious matter in such cases like loss detection [24] which most existing formulated models have a limitation in identifying this physical progression [21]. The extended investigation of the nonlinear and time-varying model allows in advocated analysis of nonlinear system dynamics, thus demonstrate the complexity of the system. So far, however, there has been little discussion about NARX for the steam distillation column in [6]–[8], [16], [25]. In addition, no research has been found that applied Binary Particle Swarm Optimisation (BPSO) algorithm for the mentioned time-varying distillation column process above using polynomial NARX model.

Time-varying dynamic behaviour is a progressively vital area in nonlinear system identification. [20]–[22] have magnificently developed the nonlinear and time-varying NARX model for semisubmersible platform, dielectric elastomer actuators, and human EEG data respectively. Recently, time-varying modelling has proven the strength and accuracy finding of the nonlinearity process. These highly nonlinear responses were revealed through the accurate nonlinear model by selecting the significant terms in the model.

The structure selection aim is model parsimony which it should capable of clarifying the data dynamics by means of the last regressor terms number [26]. Model parsimony aims the best model with the least complexity between multiple model structures. Information criteria such as Akaike Information Criterion (AIC), Final Prediction Error (FPE) and Model Descriptor Length (MDL) are used to impose parsimony by integrating difficulty drawbacks in accession to

This paper is submitted on 31st December 2016. Accepted on 3rd March 2017. The first author would like to thank SLAB/SLAI Kementerian Pendidikan Tinggi Malaysia for their scholarship.

All the authors are currently working with Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA (e-mail*: hezrif@ieec.org).

residual fit [27]–[29]. The BPSO is an altered of the original PSO algorithm for binary optimisation problem solving [30]. The particle values [18] and discretised version [31] for the solution of a binary problem have been employed using a PSO algorithm. Data from previous research has identified the capability search, convergence behaviour and algorithm accuracy improvement [32] in the study by BPSO execution. The proposed BPSO method [33]–[35] was not only demonstrated higher precision but also skilful in the model fit enhancement and correlation violations (CRV) number decline [29]. Moreover, for structure selection optimisation, BPSO needed the least iterations with well fitness values for convergence [36]. NARX model with BPSO offered a reliable model fit for essential oil identification [18] and appropriately concealed the dynamics of the time invariant system. Additionally, the previous study also presented optimal swarm size for convergence using BPSO-based NARX for DC Motor [29], [37].

In addition, the full range excitation is required due to the adequate energy of input signal to promise nonlinear system's competency for the system dynamic nonlinearity presentation [38]. Therefore, the identification experiment is very important. The suitability of Pseudo Random Binary Signal (PRBS) signals as process inputs lead to the common application because of broad range excitation of amplitudes and frequencies [10], [14], [16],[39]–[45].

The purpose of this study is to presents a time-varying nonlinear modelling for water temperature by utilising a Binary Particle Swarm Optimisation (BPSO) algorithm. It is based on the polynomial Nonlinear Auto-Regressive with eXogenous input (NARX) for Steam Distillation Pilot Plant (SDPP) identification using PRBS input signal. Several models selection criteria such as Akaike Information Criterion (AIC), Model Descriptor Length (MDL), and Final Prediction Error (FPE) were investigated in order to evaluate its fitness.

This paper has been divided into six parts. The second part deals with the system identification background. In the third part, the experimental design is presented. The methodology that has been used in this paper is also reported in part four. All the results and discussion has been shown in part five. Lastly, the conclusion and future work are discussed in the final part.

II. SYSTEM IDENTIFICATION

The NARX is a model without the residual terms like NARMAX and an extension of ARX.

The NARX model is presented as

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

f^d is the estimated model with corresponding maximum lags and input signal time delay, n_y , n_u and n_k , while $y(t)$ and $u(t)$, are the output and input, respectively. The pick of model structures is required for parameter prediction of y that depends on lagged y , u , and ε terms.

A. Polynomial NARX

NARX polynomial model representation is given by

$$y(t) = \sum_{m=1}^{n_p} P_m \theta_m + \varepsilon(t) \quad (2)$$

The m -th regression term, P_m and the m -th regression parameter, θ_m are presented in the polynomial expansion for n_p the number of terms.

The Least Squares (LS) problem's formulation and solution have been involved in identification. The matrix form is

$$P\theta + \varepsilon = y \quad (3)$$

where y is the real reflections, θ is a coefficient vector and P is a regressor matrix.

B. Model Structure Selection : Binary Particle Swarm Optimisation (BPSO)

The PSO technique is established on evolutionary computation and swarm philosophy. The convergence quality improvement and the algorithm adaptation in problems solving contributed to numerous established variants such as Vanilla and Binary PSO algorithms.

The velocity and position update equations are among the Vanilla PSO algorithm,

$$V_{id} = V_{id} + C_1(P_{best} - X_{id}) \times rand_1 + C_2(G_{best} - X_{id}) \times rand_2 \quad (4)$$

The best particle's fitness, P_{best} and the best particle's solution, G_{best} attained by the swarm with compounding of C_1 and C_2 , the cognition and social learning rate, respectively. While the particle velocity, V_{id} and the particle position X_{id} , are used together with uniformly-distributed random numbers, $rand_1$ and $rand_2$ which is between 0 and 1.

The value of V_{id} has been employed for a particle positions alteration.

$$X_{id} = X_{id} + V_{id} \quad (5)$$

In binary optimisation problem solving, the probabilities of change have been demonstrated in the BPSO particle positions rather than the actual solution as in (4) and (5). The bit change process is between 0 and 1, as stated below;

$$\text{bin string} = \begin{cases} 1, & X_{id} \geq 0.5 \\ 0, & X_{id} < 0.5 \end{cases} \quad (6)$$

The bit will vary from its current condition to another (either 0 to 1 or 1 to 0) for probability value bigger than 0.5. Otherwise, the bit will sustain for the particle value is smaller than 0.5.

BPSO convergence is affected by various parameters such a velocity bounding parameters (V_{min} and V_{max}), position bounding parameters (X_{min} and X_{max}) and parameters (C_1 and C_2) for the influence control of the swarm and self-cognition.

In the swarm of BPSO for polynomial NARX structure selection, it involves a linear least squares solution. A $1 \times m$ solutions vector, X_{id} has been transmitted for each one particle. QR factorisation has been employed for the forecasting of the parameter value, θ_R for the reduced P matrix (θ_R).

$$P_R \theta_R + \varepsilon = y \quad (7)$$

$$P_R = Q_R R_R \quad (8)$$

$$g_R = Q_R^T y \quad (9)$$

$$R_R \theta_R = g_R \quad (10)$$

Next, the value of θ_R can be estimated by reorganising and solving (10).

$$\theta_R = R_R^T g_R \quad (11)$$

C. Model Estimation

Based on the model parameters, θ , the residuals Normalised Sum Squared Error (NSSE), $V_{NSSE}(\theta, Z^N)$ is;

$$V_{NSSE}(\theta, Z^N) = \frac{1}{2N} \sum_{t=1}^N \varepsilon^2(t, \theta) \quad (12)$$

The selection of model order can be done by using several established models selection criteria such as Akaike Information Criterion (AIC), Model Descriptor Length (MDL), and Final Prediction Error (FPE) as shown in (13), (14) and (15) respectively.

$$V_{AIC} = \left(1 + 2 \frac{d}{N}\right) V_{NSSE}(\theta, Z^N) \quad (13)$$

$$V_{MDL} = \left(1 + \log(N) \frac{d}{N}\right) V_{NSSE}(\theta, Z^N) \quad (14)$$

$$V_{FPE} = \left(\frac{1 + \frac{d}{N}}{1 - \frac{d}{N}}\right) V_{NSSE}(\theta, Z^N) \quad (15)$$

where the number of estimated parameters and the data points amount are presented as d and N , respectively. $V_{NSSE}(\theta, Z^N) \rightarrow 0$ will leads to minimum values of V_{AIC} , V_{MDL} and V_{FPE} when $d \rightarrow 1$. Models with the lowest V_{AIC} , V_{MDL} and V_{FPE} scores or fitness values conform the principle of parsimony as the smallest quantity of parameters were necessary to supply the most beneficial fit for the data.

The collected data using SDPP will be separated into two divisions; training set for model estimation and testing set for model validation. For model estimation, the sample data are utilized for estimation to fit the NARX model. Three pre-processing (PP) methods namely magnitude scaling, block division, and interlacing are presented.

Per a standard range, the magnitude scaling method will scale the data. The scaling technique is required when there is an unsuitable input-output data sets range. The scaling technique formulation is

$$y = \frac{(y_{max} - y_{min})(x_{max} - x_{min})}{x_{max} - x_{min}} + y_{min} \quad (16)$$

$(y_{max} - y_{min})$ and $(x_{max} - x_{min})$ are the data range after and prior to scaling, respectively. y and x are the data before and rescaled, respectively.

Typically, the equal 50 % division is for training and testing. Two methods of division exist, namely, block division and interlacing. Only one of these methods is used at one time.

In block division, the training set ($D_{training}$) consists of the first half dataset and the testing set ($D_{testing}$) contains the second half dataset.

The interlacing method divides the dataset based on the position of the data into training and testing sets. The training set contains odd positions data, while even positions data are allotted for the testing set.

For odd N case, $D_{1,3,5,\dots,N} = D_{training}$ and $D_{2,4,6,\dots,N-1} = D_{testing}$.

Otherwise, for even N case, $D_{2,4,6,\dots,N-1} = D_{training}$ and $D_{1,3,5,\dots,N} = D_{testing}$.

D. Model Validation

A measurement of the model's ability for future value prediction in One-Step Ahead (OSA) prediction is based on its previous data as given by;

$$\hat{y} = \hat{g}(z(t)) \quad (17)$$

The estimated nonlinear model, \hat{g} and the regressors, $z(t)$ are utilised in (17). For the NARX model, the $z(t)$ representation is given below.

$$z(t) = \begin{bmatrix} y(t-1), y(t-2), \dots, y(t-n_y), \\ u(t-n_k-1), u(t-n_k-2), \dots, u(t-n_k-n_u) \end{bmatrix} \quad (18)$$

The magnitude testing of residuals for regression and model fitting problems are solved by employing the standard methods, Sum Squared Error (SSE) and Mean Squared Error (MSE).

The SSE equation of length n for a residual vector ε is given by;

$$SSE = \sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (19)$$

where y_i is the discovered value, and \hat{y}_i is the projected value at point i .

Similar to the SSE equation, but the MSE equation is divided by n , the number of samples as stated below;

$$MSE = \frac{\sum_{i=1}^n (e_i)^2}{n} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (20)$$

A good model fit results from low values of SSE and MSE from the residuals magnitude.

The R-Squared technique is employed for the goodness of fit model measurement. The R-Squared is reported as;

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (21)$$

where the actual and estimated observations at interval i , are described as y_i and \hat{y}_i , respectively. n is the number of observations and \bar{y} is the mean value of y .

The model validity for the nonlinear model can be done using correlation tests for identification by deciding the residuals' whiteness. The following cross correlation tests are needed for the fitness exhaustive test of the nonlinear model;

$$\theta_{u\varepsilon^2}(\tau) = E[u(t-\tau)\varepsilon^2(t)] = 0, \forall \tau \quad (22)$$

$$\theta_{u'^2\varepsilon}(\tau) = E[(u^2(t-\tau) - \bar{u}^2(\tau))\varepsilon(t)] = 0, \forall \tau \quad (23)$$

$$\theta_{u'^2\varepsilon^2}(\tau) = E[(u^2(t-\tau) - \bar{u}^2(\tau))\varepsilon^2(t)] = 0, \forall \tau \quad (24)$$

where:

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

$E[\blacksquare]$: mathematical expectation of the correlation function.

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

τ : lag space.

$u(t)$: observed input at time t .

For any coefficients that are lying beyond the confidence band, it reveals a significantly large correlation. The 95 % confidence band is expected in correlation tests as of finite quantity of data length availability.

The distribution of the residuals in system identification is showed in a histogram analysis. The histogram requires a symmetric bell-shaped distribution of the white noise with the most frequency amounts clustered in the central and narrowing off equally at the end of the tails.

III. EXPERIMENTAL DESIGN

An SDPP system has been used coil-type water immersion heater to generate steam with two resistive temperature detectors (RTD) PT-100. The water and steam temperature are monitored using both of the RTDs. 1 V to 5 V signals are obtained from the RTDs' resistance output for specific several temperature series. The boiling water allows steam exceeds through the raw material during the extraction process. Then, the condenser condenses the steam first before it changes into oil and hydrosol in the collector. A plant with a 1.5 kW, 240 V and 50 Hz power has been used for immersion heating element with specific sampling time. For modelling purposes, a MATLAB software is utilised for plant integration and data collection. Fig. 1 illustrates a pilot plant of the essential oil extraction system.

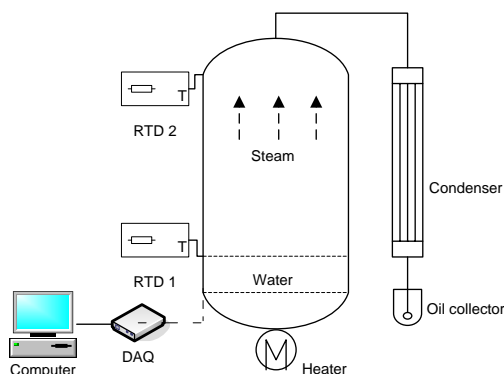


Fig. 1. Steam Distillation Pilot Plant (SDPP) System

IV. METHODOLOGY

The nonlinear water temperature dataset that consists of 18, 000 data points from the SDPP system with the implementation of the PRBS input is demonstrated in Fig. 2. In addition, a total of four PP combinations were employed in this analysis. The PP methods are depicted in Table I in terms of PP Code.

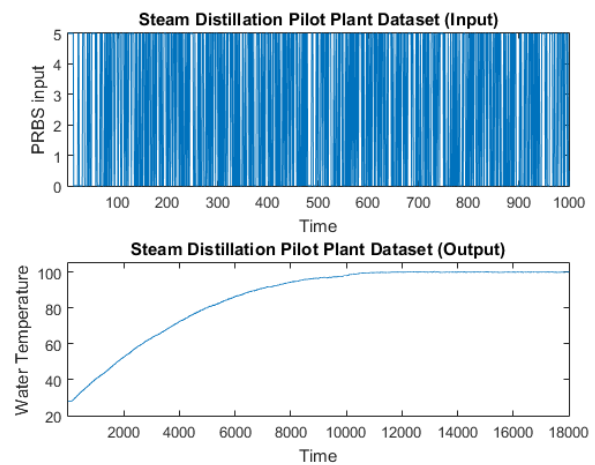


Fig. 2. SDPP Dataset

TABLE I
PRE-PROCESSING TECHNIQUE AND RESPECTIVE CODES

PP Code	Pre-processing Method
00	No magnitude scaling, block division
01	No magnitude scaling, interleaving
10	Magnitude scaling, block division
11	Magnitude scaling, interleaving

For optimisation purposes, various swarm sizes, maximum iterations, and random seeds parameters have been combined. Higher swarm sizes create larger potential in global minima searching by the number of agents based on optimisation time and computational cost. Termination of the PSO search will be driven by the achievement of the objective or the finding of maximum iterations for each experiment. In addition, the initial random seed for each V_{id} and X_{id} particles have influence on the final optimisation PSO result. The results consistency will be decided by different initial random seeds for repeated experiments. The selected parameter values as listed in Table II are examined which has been implemented by [29], [36] for the fitness function optimal convergence exploration.

TABLE II
BPSO PARAMETER SETTINGS FOR STRUCTURE SELECTION [29], [36]

Parameter	Value
Fitness Criterion	AIC, FPE, MDL
Swarm size	10, 20, 30, 40, 50
Maximum Iterations	500, 1000, 1500
Initial Random 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

V. RESULTS AND DISCUSSION

Table III indicates the magnitude scaling from PP 10 and PP 11 appeared to have a positive effect on fitness. On the other side, the magnitude scaling resulted in a negative effect on CRV from PP 10 and 11 based on the minimum fitness searching. Hence, these high number of CRV leads to highly biased predictions.

TABLE III
EFFECT OF PRE-PROCESSING METHOD ON FITNESS AND TOTAL CRV
BASED ON MINIMUM FITNESS SEARCH

PP	Criteri- on	Fitness Value			Total CRV
		Min	Average	Max	
00	AIC	5.4905×10^{-3}	5.9162×10^{-3}	6.3549×10^{-3}	65
	FPE	5.5696×10^{-3}	5.9629×10^{-3}	6.6241×10^{-3}	140
	MDL	5.3399×10^{-3}	5.6995×10^{-3}	6.1529×10^{-3}	65
01	AIC	5.7183×10^{-3}	6.0288×10^{-3}	6.4284×10^{-3}	52
	FPE	5.8063×10^{-3}	6.1875×10^{-3}	6.7334×10^{-3}	52
	MDL	5.5615×10^{-3}	5.8595×10^{-3}	6.2266×10^{-3}	52
10	AIC	4.3213×10^{-6}	4.6457×10^{-6}	5.2024×10^{-6}	95
	FPE	4.3975×10^{-6}	4.9249×10^{-6}	6.0915×10^{-6}	95
	MDL	4.1847×10^{-6}	4.5497×10^{-6}	4.8951×10^{-6}	95
11	AIC	4.5570×10^{-6}	4.8525×10^{-6}	5.6715×10^{-6}	181
	FPE	4.6021×10^{-6}	4.9651×10^{-6}	6.1327×10^{-6}	174
	MDL	4.3892×10^{-6}	4.6071×10^{-6}	5.2881×10^{-6}	129

Complementary to this, based on the minimum CRV searching as illustrated in Table IV, interleaving without magnitude scaling recorded the best combination of both CRV and fitness from PP 01. In this search, the BPSO-based NARX model managed to cut down the CRV numbers and maintained the low fitness values in PP 01 while other PP methods demonstrated high CRV numbers. Therefore, the best results were obtained using PP method 01. This is based on the lowest total CRV from both training and testing sets and low fitness value. In addition, lower CRV values contribute to the uncorrelated residuals which represent a good model fit.

TABLE IV
EFFECT OF PRE-PROCESSING METHOD ON FITNESS AND TOTAL CRV
BASED ON MINIMUM TOTAL CRV SEARCH

PP Code	Criterion	Min Total CRV	Fitness Value
00	AIC	61	5.9268×10^{-3}
	FPE	37	5.7732×10^{-3}
	MDL	62	5.5275×10^{-3}
01	AIC	9	6.1741×10^{-3}
	FPE	11	6.7334×10^{-3}
	MDL	10	5.7969×10^{-3}
10	AIC	47	5.0208×10^{-6}
	FPE	95	4.3975×10^{-6}
	MDL	54	4.8866×10^{-6}
11	AIC	52	4.7407×10^{-6}
	FPE	117	4.9478×10^{-6}
	MDL	51	4.5265×10^{-6}

As presented in Table V, the models were compared for PP 01 based on the model's selection criteria as in Eq. (13) – (15). High R-squared and low MSE values have been reported by all the criterion designated to the acceptable fitting results. Not only that, the residuals of all criterion exhibited low correlation, as indicated by the small number of CRV. Consequently, the AIC model was found to be slightly better compared to FPE and MDL models because of the best combination of lower fitness values and the least CRV on the

testing set.

TABLE V
MODEL VALIDATION SUMMARY

Fitness Criterion	Evaluation Criterion	Training Set	Testing Set
AIC	Times Found	3	
	AIC	6.1741×10^{-3}	6.4916×10^{-3}
	FPE	6.3207×10^{-3}	6.6457×10^{-3}
	MDL	5.9703×10^{-3}	6.2773×10^{-3}
	R-squared (%)	99.9975	99.9974
	CRV	3	6
FPE	Times Found	3	
	AIC	6.4393×10^{-3}	6.8001×10^{-3}
	FPE	6.7334×10^{-3}	7.1107×10^{-3}
	MDL	6.1622×10^{-3}	6.5075×10^{-3}
	R-squared (%)	99.9976	99.9975
	CRV	3	8
MDL	Times Found	2	
	AIC	5.9948×10^{-3}	6.3335×10^{-3}
	FPE	6.1371×10^{-3}	6.4839×10^{-3}
	MDL	5.7969×10^{-3}	6.1244×10^{-3}
	R-squared (%)	99.9976	99.9975
	MSE	9.7700×10^{-3}	1.0317×10^{-2}

The BPSO-based NARX justification results representing the AIC model from PP 01 are presented. Fig. 3 shows good model resulted by high R-squared. Additionally, small MSE value is demonstrated in Fig. 4. The model created using this PP method recorded little violations in the correlation plots (Fig. 5 until Fig. 7) and well distributed of the white noise in histogram tests (Fig. 8).

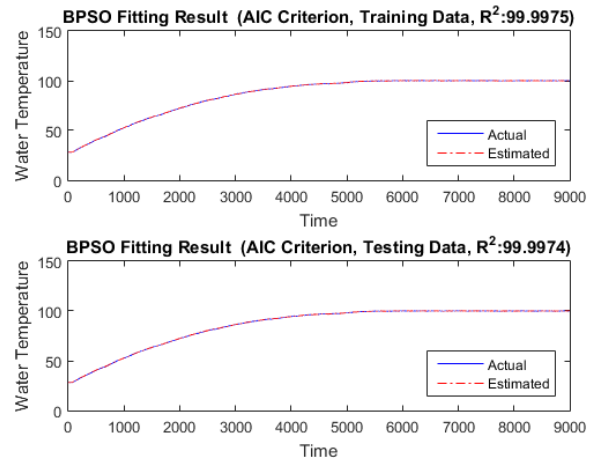


Fig. 3. BPSO Water Temperature Model Fit

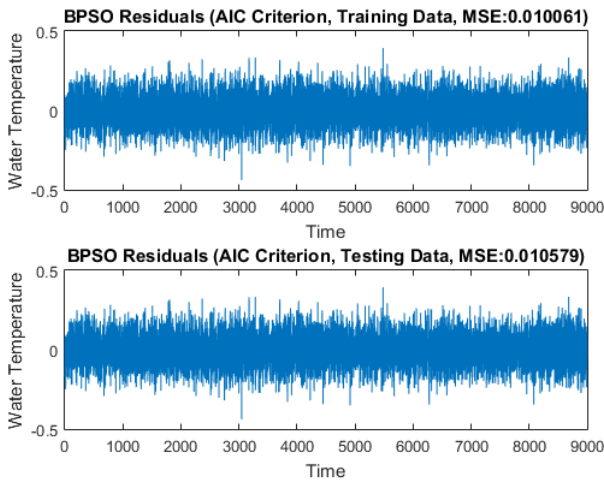


Fig. 4.BPSO Water Temperature Residual Plot

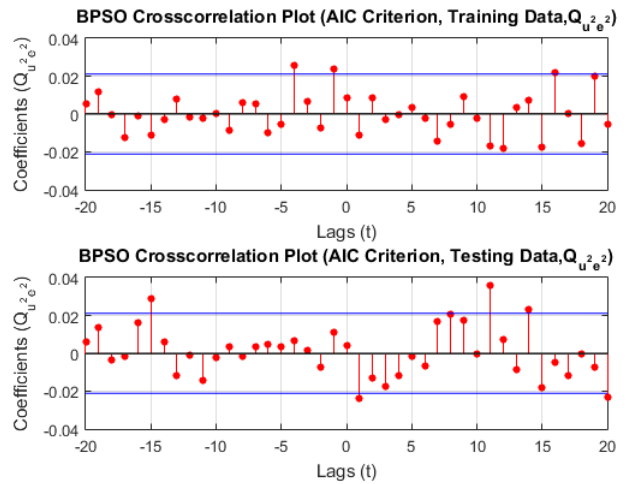


Fig. 7.BPSO Water Temperature Correlation Test (3/3)

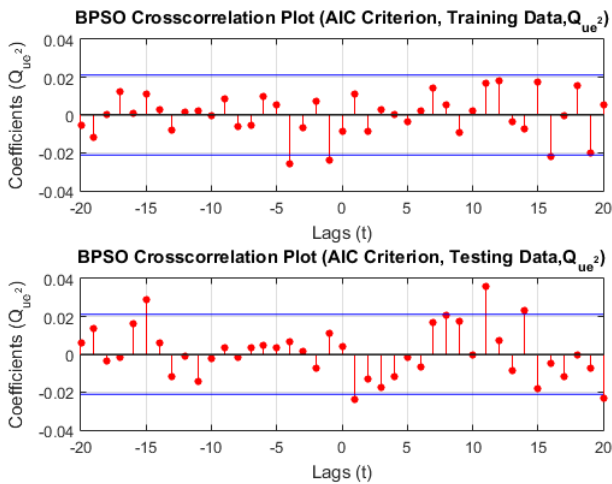


Fig. 5.BPSO Water Temperature Correlation Test (1/3)

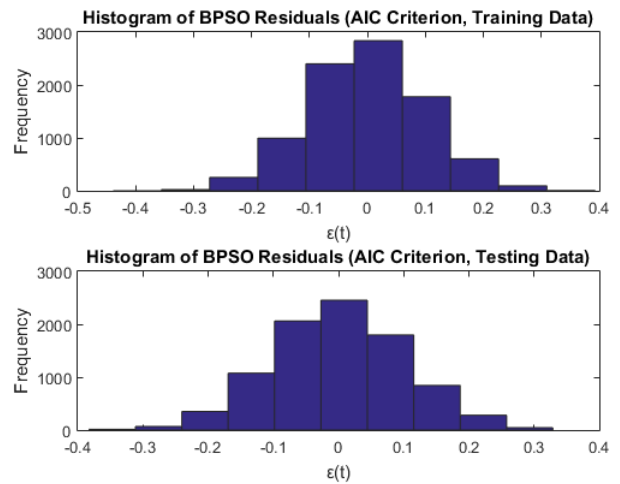


Fig. 8.BPSO Water Temperature Histogram of Residuals

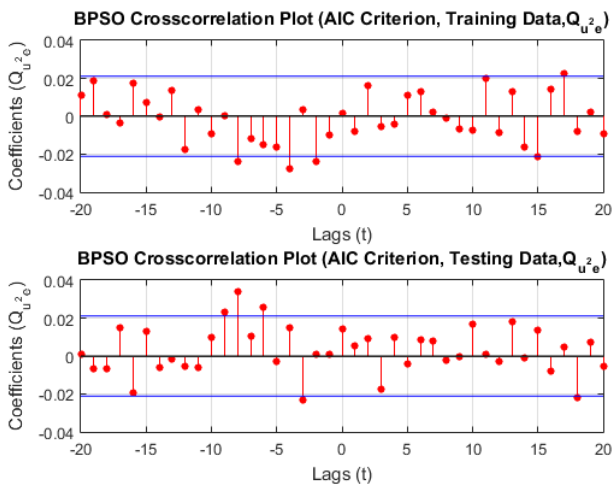


Fig. 6.BPSO Water Temperature Correlation Test (2/3)

Thus, the AIC residual displayed small correlation, as indicated within 95 % confidence limit in the correlation plots with low deviations number. Therefore, all criterion models generated using PP method 01 for BPSO-based NARX were measured efficient and exact illustrations of the nonlinear system.

The Water Temperature BPSO-NARX models for AIC, FPE and MDL models are in Table IV.

TABLE VI
OUTPUT MODEL

Criterion	Output Model
AIC	$y(t) = 2.0880 \times 10^{-4} u(t-4) + 1.0017 y(t-2) + 8.3542 \times 10^{-3} y(t-1) * y(t-3) - 2.8747 \times 10^{-3} y(t-1) * y(t-2) - 5.4969 \times 10^{-3} y(t-2) * y(t-3) + \epsilon(t)$
FPE	$y(t) = 1.7413 \times 10^{-4} u(t-4) + 1.0019 y(t-2) + 2.1629 \times 10^{-7} u(t-1) * y(t-4) + 1.9333 \times 10^{-3} y(t-1) * y(t-4) + 3.0495 \times 10^{-3} y(t-1) * y(t-3) - 5.8121 \times 10^{-4} y(t-3) * y(t-3) - 4.4207 \times 10^{-3} y(t-2) * y(t-2) + \epsilon(t)$
MDL	$y(t) = 1.0019 y(t-2) - 8.1381 \times 10^{-5} u(t-2) * u(t-3) + 1.9125 \times 10^{-3} y(t-3) * y(t-4) + 6.9128 \times 10^{-3} y(t-1) * y(t-1) - 8.8446 \times 10^{-3} y(t-1) * y(t-2) + \epsilon(t)$

Based on Table VI, the output models selected using different criteria for PP 01 were reported. Consequently, the AIC and MDL models were found to be more effective with the minimum number of parameters of the output model. The

parsimonious model structures are an indication of good model fit as the smallest quantity of parameters were necessary to supply the most beneficial fit for the data.

VI. CONCLUSION

The nonlinear modelling for water temperature using BPSO algorithm based NARX structure was presented. The model structure selection of polynomial NARX has been employed on BPSO algorithm for the SDPP identification. Fitness and CRV investigation were conducted based on several models selection criteria such as AIC, FPE, and MDL. The combination of interleaving without magnitude scaling technique (PP 01) recorded the best combination due to its lowest total CRV from both training and testing sets and low fitness value. Subsequently, the AIC model of PP 01 was observed to be a slightly greater model because it had low fitness values, besides the least CRV on the testing set compared to FPE and MDL. Moreover, AIC and MDL recorded lesser parameter number of the output model. Overall, all criterion models generated using PP 01 for BPSO-based NARX were considered valid and accurate representations of the system with AIC model presented as the best model. It can be seen through the high R-squared, small MSE value, uncorrelated residuals and a minimum number of parameters in the output model. The results are encouraging and should be explored with another type of dataset, for instance, the steam temperature dataset that can also be collected using SDPP.

REFERENCES

- [1] N. Togun, S. Baysec, and T. Kara, "Nonlinear modeling and identification of a spark ignition engine torque," *Mech. Syst. Signal Process.*, vol. 26, pp. 294–304, Jan. 2012.
- [2] T. G. Ling, M. F. Rahmat, A. R. Husain, and R. Ghazali, "System identification of electro-hydraulic actuator servo system," in *2011 4th International Conference on Mechatronics (ICOM)*, 2011, pp. 1–7.
- [3] S. A. Billings, S. Chen, and R. J. Backhouse, "The identification of linear and non-linear models of a turbocharged automotive diesel engine," *Mech. Syst. Signal Process.*, vol. 3, no. 2, pp. 123–142, Apr. 1989.
- [4] J. Paduart, L. Lauwers, J. Swevers, K. Smolders, J. Schoukens, and R. Pintelon, "Identification of nonlinear systems using Polynomial Nonlinear State Space models," *Automatica*, vol. 46, no. 4, pp. 647–656, Apr. 2010.
- [5] A. Daryasafar, A. Ahadi, and R. Kharrat, "Modeling of Steam Distillation Mechanism during Steam Injection Process Using Artificial Intelligence," *Sci. World J.*, vol. 2014, no. 246589, 2014.
- [6] N. Kasuan, Z. Yusuf, M. N. Taib, M. H. F. Rahiman, N. Tajuddin, and M. A. A. Aziz, "Robust steam temperature regulation for distillation of essential oil extraction process using hybrid fuzzy-PD plus PID Controller," *World Acad. Sci. Eng. Technol.*, vol. 71, pp. 932–937, 2010.
- [7] N. Kasuan, M. N. Taib, and M. H. Fazalul Rahiman, "Model Reference Adaptive Controller to regulate steam temperature in distillation process for essential oil extraction," in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, 2011, pp. 298–303.
- [8] N. Kasuan, M. Yunus, M. H. F. Rahiman, S. R. S. Aris, and M. N. Taib, "Essential oil composition of Kaffir lime: Comparative analysis between controlled steam distillation and hydrodistillation extraction process," in *2009 IEEE Student Conference on Research and Development (SCORED)*, 2009, pp. 479–482.
- [9] D. O. Araromi, J. A. Sonibare, and J. O. Emuoyibofarhe, "Fuzzy identification of reactive distillation for acetic acid recovery from waste water," *J. Environ. Chem. Eng.*, vol. 2, no. 3, pp. 1394–1403, Sep. 2014.
- [10] I. Yunan, I. M. Yassin, S. F. S. Adnan, and M. H. F. Rahiman, "Identification of essential oil extraction system using Radial Basis Function (RBF) Neural Network," in *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, 2012, pp. 495–499.
- [11] S. Dudret and K. Beauchard, "Stability and asymptotic observers of binary distillation processes described by nonlinear convection/diffusion models," *Am. Control Conf. (ACC)*, 2012. *IEEE*, 2012., 2012.
- [12] A. Bachnas, "Linear parameter-varying modelling of a high-purity distillation column," *Master Sci. Diss.*, 2012.
- [13] M. H. F. Rahiman, M. N. Taib, and Y. M. Salleh, "Performance of multi-step-ahead-prediction ARX for steam temperature in a self-refilling distillation essential oil extraction system," in *2007 International Conference on Control, Automation and Systems*, 2007, pp. 1696–1699.
- [14] Z. Muhammad, Z. M. Yusoff, M. H. F. Rahiman, and M. N. Taib, "Modeling of steam distillation pot with ARX model," in *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, 2012, pp. 194–198.
- [15] N. Tajjudin, N. Ismail, M. H. F. Rahiman, and M. N. Taib, "Model Predictive Control using ARX model for steam distillation essential oil extraction system," in *2010 International Conference on Intelligent and Advanced Systems*, 2010, pp. 1–5.
- [16] N. N. Mohammad, N. Kasuan, M. H. F. Rahiman, and M. N. Taib, "Steam temperature control using fuzzy logic for steam distillation essential oil extraction process," in *2011 IEEE Control and System Graduate Research Colloquium*, 2011, pp. 53–58.
- [17] M. N. N. Nordin, M. H. F. Rahiman, R. Adnan, Z. M. Yusoff, and I. M. Yassin, "Optimizations of NARX lag space selection for a Multi-Layer Perceptron (MLP)-based model of a down-flowing steam distillation system using Particle Swarm Optimization (PSO)," in *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, 2012, pp. 533–538.
- [18] F. Awadz, I. M. Yassin, M. H. F. Rahiman, M. N. Taib, A. Zabidi, and H. A. Hassan, "System identification of essential oil extraction system using Non-Linear Autoregressive Model with Exogenous Inputs (NARX)," in *2010 IEEE Control and System Graduate Research Colloquium (ICSGRC 2010)*, 2010, pp. 20–25.
- [19] H. Jin, X. Chen, L. Wang, K. Yang, and L. Wu, "Dual learning-based online ensemble regression approach for adaptive soft sensor modeling of nonlinear time-varying processes," *Chemom. Intell. Lab. Syst.*, vol. 151, pp. 228–244, 2016.
- [20] E. Yazid, M. S. Liew, S. Parman, and V. J. Kurian, "Identification of transfer functions from surge motion response of a semisubmersible platform using time-varying NARX model," *Appl. Ocean Res.*, vol. 54, pp. 53–66, Jan. 2016.
- [21] W. R. Jacobs, E. D. Wilson, T. Assaf, J. Rossiter, T. J. Dodd, J. Porritt, and S. R. Anderson, "Control-focused, nonlinear and time-varying modelling of dielectric elastomer actuators with frequency response analysis," *Smart Mater. Struct.*, vol. 24, no. 5, p. 55002, May 2015.
- [22] Y. Zhao, S. A. Billings, H. Wei, F. He, and P. G. Sarrigiannis, "A new NARX-based Granger linear and nonlinear causal influence detection method with applications to EEG data," *J. Neurosci. Methods*, vol. 212, no. 1, pp. 79–86, Jan. 2013.
- [23] H. M. Shariff, M. H. Marzaki, M. Tajjudin, and M. H. F. Rahiman, "System identification for steam distillation pilot plant: Comparison between linear and nonlinear models," in *2013 IEEE 3rd International Conference on System Engineering and Technology*, 2013, pp. 263–268.
- [24] I. A. Kougioumtzoglou and P. D. Spanos, "An identification approach for linear and nonlinear time-variant structural systems via harmonic wavelets," *Mech. Syst. Signal Process.*, vol. 37, no. 1–2, pp. 338–352, May 2013.
- [25] N. Kasuan, N. Ismail, M. N. Taib, and M. H. Fazalul Rahiman, "Recurrent adaptive neuro-fuzzy inference system for steam temperature estimation in distillation of essential oil extraction process," in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, 2011, pp. 1–6.
- [26] S. Billings and H. Wei, "An adaptive orthogonal search algorithm for model subset selection and non-linear system identification," *Int. J. Control*, 2008.
- [27] E. M. A. M. Mendes and S. A. Billings, "An alternative solution to the model structure selection problem," *IEEE Transactions on Systems Man and Cybernetics Part A: Systems and Humans*. 01-Nov-2001.
- [28] I. M. Yassin, "Nonlinear Auto-Regressive Model Structure SElection Using Binary Particle Swarm Optimization Algorithm," *PhD Thesis*, no. 1, pp. 1–5, 2014.

- [29] I. Yassin, M. Taib, R. Adnan, and S. Alam, "Extended analysis of bpso structure selection of nonlinear auto-regressive model with exogenous inputs (NARX) of direct current motor," *Songklanakarın J. Sci. Technol.*, vol. 36, no. 6, pp. 683–699, 2014.
- [30] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, 1997, vol. 5, pp. 4104–4108.
- [31] Y. Liu, Z. Zhang, and Z. Liu, "Customized configuration for hierarchical products: component clustering and optimization with PSO," *Int. J. Adv. Manuf. Technol.*, 2011.
- [32] K. K. Bharti and P. K. Singh, "Opposition chaotic fitness mutation based adaptive inertia weight BPSO for feature selection in text clustering," *Appl. Soft Comput.*, vol. 43, pp. 20–34, 2016.
- [33] R. Keshavarzi, M. Akhlaghi, and F. Emami, "Binary PSO algorithm assisted to investigate the optical sensor based plasmonic nano-bi-domes," *Opt. - Int. J. Light Electron Opt.*, vol. 127, no. 19, pp. 7670–7675, 2016.
- [34] G. Haixiang, L. Yijing, L. Yanan, L. Xiao, and L. Jinling, "BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification," *Eng. Appl. Artif. Intell.*, vol. 49, pp. 176–193, 2016.
- [35] Z. Beheshti, S. M. Shamsuddin, and S. Hasan, "Memetic binary particle swarm optimization for discrete optimization problems," *Inf. Sci. (Ny)*, vol. 299, pp. 58–84, 2015.
- [36] I. M. Yassin, M. N. Taib, R. Adnan, M. K. M. Salleh, and M. K. Hamzah, "Effect of swarm size parameter on Binary Particle Swarm optimization-based NARX structure selection," in *2012 IEEE Symposium on Industrial Electronics and Applications*, 2012, pp. 219–223.
- [37] I. M. Yassin, M. N. Taib, R. Adnan, M. K. M. Salleh, and M. K. Hamzah, "Effect of swarm size parameter on Binary Particle Swarm optimization-based NARX structure selection," in *2012 IEEE Symposium on Industrial Electronics and Applications*, 2012, pp. 219–223.
- [38] A. Soni, "Control-relevant system identification using nonlinear Volterra and Volterra-Laguerre models," *Dr. Diss. Univ. Pittsburgh.*, 2006.
- [39] D. Rivera, H. Lee, H. Mittelmann, and M. Braun, "High-Purity Distillation," *IEEE Control Syst. Mag.*, vol. 27, no. 5, pp. 72–89, Oct. 2007.
- [40] M. Deflorian and S. Zaglauer, "Design of experiments for nonlinear dynamic system identification," *IFAC 18th World Congr. Milano 2011*, 2011.
- [41] O. Nelles, *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*. Springer, 2001.
- [42] M. W. Braun, D. E. Rivera, A. Stenman, W. Foslien, and C. Hrenya, "Braun, M. W., et al. 'Multi-level pseudo-random signal design and 'model-on-demand' estimation applied to nonlinear identification of a RTP wafer reactor.' American Control Conference, 1999. Proceedings of the 1999. Vol. 3. IEEE, 1999.," in *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*, 1999, vol. 3, pp. 1573–1577.
- [43] M. Hafner, M. Schüler, O. Nelles, and R. Isermann, "Fast neural networks for diesel engine control design," *Control Eng. Pract.*, vol. 8, no. 11, pp. 1211–1221, Nov. 2000.
- [44] A. U. Levin and K. S. Narendra, "Control of nonlinear dynamical systems using neural networks. II. Observability, identification, and control.," *IEEE Trans. Neural Netw.*, vol. 7, no. 1, pp. 30–42, Jan. 1996.
- [45] R. Isermann and N. Muller, "Nonlinear identification and adaptive control of combustion engines," *IFAC-Workshop Adapt. Learn. Control signal Process. 29–31 August 2001, Como, Italy. 2001.*, 2001.