Modeling Automotive Palm Oil Biodiesel Engine using Multi-Objective Optimization Differential Evolution Algorithm

ROBIAH AHMAD[1]*, MOHD ZAKIMI ZAKARIA[2], AZUWIR MOHD NOR[2], HISHAMUDDIN JAMALUDDIN[3]

[1]*UTM Razak School
Universiti Teknologi Malaysia
54100 Jalan Semarak, Kuala Lumpur
MALAYSIA
[2]School of Manufacturing Engineering
Universiti Malaysia Perlis
Pauh Putra Main Campus, 02600 Arau, Perlis
MALAYSIA
[3]Department of Applied Mechanics
Faculty of Mechanical Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru
MALAYSIA
robiah@ic.utm.my*, zakimizakaria@unimap.edu.my, azuwir@unimap.edu.my,
hishamj@fkm.utm.my

Abstract: - Modeling automotive palm oil biodiesel engine using the propose algorithm called multi-objective optimization differential evolution (MOODE) is carried out in this paper. The performance of the system is analyzed to study the dynamic behaviors of using biodiesel in running the engine. An adequate modeling of the system is needed to design an exclusive engine controller. The biodiesel engine is treated as a black box where the acquired input-output data is used in the modeling processes. Two objective functions are considered for optimization; minimizing the number of term of a model structure and minimizing the mean square error between actual and predicted outputs. Nonlinear auto-regressive with exogenous input (NARX) model is used to represent the mathematical model of the investigated system. To obtain an optimal model for representing the dynamic behavior of automotive palm oil biodiesel engine, the model validity tests have been applied. Therefore, MOODE can be applied for other processes to find an optimal model structure as long as those processes have reliable input output data.

Key-Words: - System identification, multi-objective optimization, differential algorithm, mathematical model, palm oil biodiesel, NARX model

1 Introduction

System identification is the science and art of developing mathematical models of dynamic systems from measured input-output data. This method is a very large topic, with different techniques that depends on the characteristics of the estimated models such as linear, nonlinear, hybrid, and nonparametric [1]. In this chapter, the literature reviews are presented. The study is concentrated on selecting a compact and adequate model structure that is one of the important parts in system identification. In short, numerous researched in system identification are focused on selecting appropriate model to be used in controlling the dynamic systems. Beside, parameter estimation algorithms have also attracted many researchers to propose new algorithms that are suitable for the model parameters to be estimated.

Generally, constructing models from observed input-output data is a fundamental element in engineering and science. In system identification, it involves some theories and concepts in order to estimate a good and adequate model. These concepts are [1]:

i. *Model*. A model presents the relationship between observed input-output data. It allows for prediction of characteristics of the system. Commonly, the relationship is expressed in mathematical equation.
ii. *True Description*. It is difficult to achieve a true description of the system to be modelled. Thus, this concept is convenient to assume such a description as an abstraction. It has the same properties as a model, but typically much more complex.
iii. *Model class*. This concept presents the size of model that needs to be considered for estimating the relationship between observed data. It could be a set that can be parameterized by a finite-dimensional parameter such as ARX (Auto-Regressive with eXogenous input) models with input-output lags.
iv. *Complexity*. The complexity of a model is much dependent on the model class. The higher model class is the higher number of complexity. This could be the dimension of a vector that parameterizes the model in a smooth way.
v. *Information*. It is more likely to model class which concerns both information provided by the observed data and prior information about the system to be modelled.
vi. *Estimation*. This process to find a suitable model class and the particular model parameters to represent the investigated system. Commonly, the data used for this concept is called estimation data. It is fashionable to call this process learning to produce a good model.
vii. *Validation*. This process to ensure the estimated model can be used to other data beside the estimation data. The other data is commonly called validation data. Another term for this process is generalization.
viii. *Model fit*. Finally this process is measured on how well the selected model is able to fit the observed data.

One of the main problems in system identification is to represent a dynamic system with an adequate and parsimonious model. A dynamic system can be considered as a black box and can be modeled based on measured input-output data [2]. The task of system identification is to develop an adequate and parsimonious mathematical model representing a linear or nonlinear system. This task has been studied by many researchers [3-8]. An ultimate objective in modeling a dynamic system is to produce a compact model structure with only significant terms included in the final model.

Differential evolution (DE) is one of EA was first published as a technical report by Storn and Price [9]. They developed DE with simple algorithm and easy to use but reliable and versatile function optimizer. Since then, DE is the most chosen algorithm to global optimization in various applications [10-14]. DE is a population-based optimizer likes all EAs. It starts randomly chosen initial points from the initial population by sampling the objective function involved. Further, DE was extended to multi-objective optimization (MOO) problem by Babu et al. [15] called Multi-Objective Differential Evolution (MODE). Babu et al. [15] applied MODE in solving optimization problems in chemical engineering. Industrial adiabatic styrene reactor was considered as a problem to be optimized by considering productivity, selectivity and yield as the main objectives. The results were compared with NSGA and shown its superiority.

In this study, the proposed algorithm is named Multi-Objective Optimization using Differential Evolution (MOODE). This proposed algorithm is applied in system identification for model structure selection, whereas two objective functions are considered: model predictive error and model complexity.

An automotive diesel engine fuelled with palm oil biodiesel is considered in this case study. This system is referred as POB system. The palm waste was converted to palm oil biodiesel that is used as an alternative energy to feed the POB system. The performance of the system is analyzed to study the dynamic behaviors of using biodiesel in running the engine. An adequate modeling of the system is needed to design an exclusive engine controller. In this case study, MOODE is applied to obtain some models to be chosen from for POB system. These identified models are validated using model validity tests. These tests are applied to select a model that is adequate to represent the dynamic behavior of the POB system.

Since the process systems have reliable input output data, the proposed algorithm integrating with system identification technique shall be used as a modeler to predict the behavior of those systems. Therefore, MOODE can be applied in another process as shown in the case study from this paper.

2 Model Structure Representation

Representing a dynamic nonlinear system from acquired input output data needs to define the type of model representation. Most nonlinear systems are modeled and identified by using mathematical and signal models, block diagram models, and simulation models. In this study, the mathematical and signal model is considered. A very common polynomial linear discrete-time system model representations is ARX (AutoRegressive with eXegenous input) model where the system output can be predicted by the past inputs and outputs of the system [2]. This model is defined as

$$y(t) = C + a_1 y(t-1) + \cdots + a_{n_y} y(t-n_y) + b_1 u(t-1) + \cdots + b_{n_u} u(t-n_u) + e(t) \quad (1)$$

where output, input, and noise signal are represented by $y(t)$, $u(t)$, and $e(t)$ respectively, while C, n_y, and n_u representing a constant, the maximum output and input lags in the model respectively. The coefficients of the model are represented by $a_1 \dots a_{n_y}$ and $b_1 \dots b_{n_u}$. The nonlinear version for ARX model is called NARX model. Chen and Billings [16] presented a Non-linear AutoRegressive Moving Average with eXogeneous inputs (NARMAX) model which provides a wide class of a nonlinear model representation with the special case as a NARX model. The NARX model can be defined as

$$y(t) = F^l\left(C, y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u)\right) + e(t) \quad (2)$$

where $F^l(.)$ is a polynomial non-linear function with l degree of nonlinearity. The NARX model can be transformed into a linear regression model represented by

$$y(t) = \sum_{i=1}^{M} \theta_i \emptyset_i(t) + e(t), \quad n_y \le t \le N \quad (3)$$

where θ_i and $\emptyset_i(t)$ are unknown coefficients or parameters and regressors respectively, M is the maximum number of terms of the regressors and N is the size of data. The maximum number of possible terms, *Lt* in the NARX model in Equation (2) can be calculated as [16]

$Lt = M + 1$ where $M = \sum_{i=1}^{l} n_i$ and $n_i = \frac{n_{i-1}(n_y+n_u+i-1)}{i}, \quad n_0 = 1$ (4)

For example, a NARX model with $n_y = n_u = 3$ and $l = 2$, a second order degree of nonlinearity, would contain 28 terms respectively. Thus, the possible models need to be considered can be calculate as $2^{Lt} - 1$, which is 268 435 455. Thus, increasing the orders of input and output lags and degree of nonlinearity, will increase the maximum number of terms of NARX model and the possible models that need to be searched. Thus, the user defined parameters such as the orders of input and output lags, and degree of nonlinearity will affect the difficulty of model structure selection. The search space become large and impractical when large user defined parameters are used. Therefore, selecting significant terms to be included in the final model become challenging and needs an automated and suitable tool for this task.

3 Modeling procedures

3.1 Model Validity Tests

The final procedure in system identification is model validation. After a set of models are identified, the model validity tests are used to ease decision maker to choose a good model. The main objective of model validation is to check whether the model fits the data adequately without any biased. In this study, three model validity tests are considered such as one step ahead prediction (OSA), model predicted output (MPO), and correlation tests.

First model validity test which is OSA can be described as [17]:

$$\hat{y}_{OSA}(t) = \hat{F}\left(y(t-1), \ldots, y(t-n_y), u(t-1, \ldots, ut-nu\right. \quad (5)$$

where the OSA output or predicted output is based on the previous actual output and input data, while $\hat{F}$ is an estimate of the non-linear function $f(.)$.

Second model assessment used in this study is MPO test [18]. This test has slightly different compared with OSA test, whereas it based on previous predicted output and input data. MPO is given by:

$$\hat{y}_{MPO}(t) = \hat{F}\left(\hat{y}(t-1), \ldots, \hat{y}(t-n_y), u(t-1), \ldots, u(t-n_u)\right) \quad (6)$$

A further study for model validation is correlation tests to validate the models obtained by MOODE. Correlation tests can be applied to both linear and nonlinear systems. The following equations show the correlation functions [19] used in this study:

$$\left.\begin{aligned}
&\emptyset_{\varepsilon\varepsilon}(\tau) = E[\varepsilon(t)\varepsilon(t-\tau)] = \delta(t), && \forall\tau \\
&\emptyset_{u\varepsilon}(\tau) = E[u(t)\varepsilon(t+\tau)] = 0, && \forall\tau \\
&\emptyset_{\bar{u}^2\varepsilon}(\tau) = E[\bar{u}^2(t)\varepsilon(t+\tau)] = 0, && \forall\tau \\
&\emptyset_{\bar{u}^2\varepsilon^2}(\tau) = E[\bar{u}^2(t)\varepsilon^2(t+\tau)] = 0, && \forall\tau \\
&\emptyset_{(u\varepsilon)\varepsilon}(\tau) = E[\varepsilon(t)u(t)\varepsilon(t+\tau+1)] = 0, && \tau \geq 0,
\end{aligned}\right\} \quad (7)$$

where $\emptyset$, $\varepsilon(t)$, $u(t)$, and $\delta(t)$ represent the standard correlation function, the residual sequence, the input of system, and the delta function or impulse function, respectively, and E[.] is the expectation operator. An identified model is adequate if all five correlation functions as shown in Equation (7) are satisfied. These functions can be analyzed visually by plotting on a graph paper with the confidence bands. The confidence bands plotted on the graphs reveal if the identified model is adequate or not. The confidence bands are estimated as 95% confidence limits that are approximately $\pm 1.96/\sqrt{N}$ where *N* is the data length.

All the model validity tests are used to check the selected models obtained by MOODE algorithm. To select a good model among the selected models, the results from the all model validations are observed. If the selected model is good in OSA and MPO tests, it is not guarantee to satisfy the correlation tests. Meanwhile, if the selected model is satisfied the correlation tests, then it is biased for OSA and MPO tests. Thus, all three model validity tests must fulfill in order to choose a good and adequate model for the real process data.

3.2 Description Data of POB System

The input-output data was collected from an experimental setup. The developed system is DOHC Mitsubishi Diesel Engine with capacity engine of 2000 cc using direct injection. The procedures and experimental setup to obtain the input-output data of the system were designed by Azuwir et al. [20]. The acquired input-output data were used for the identification purposes. The block diagram of experimental setup to acquire the input-output data of system is shown in Fig. 1.

A PRBS signal was used to excite the actuator that gives voltage to the engine system. The input of the system is voltage with values between 1 V and 1.2 V. The POB system was operated with speed range of around 2100 RPM. The speed sensor is used to count the speed of engine in operation. Thus, the output of the system is the speed of engine in RPM (revolutions per minute). The collection of data consists of 720 data points. There are two separation sets for this data. The first data set consists of 300 data and it is used for the estimation set. Meanwhile, the second data set consists of 420 data is used for validation set. The input-output data of POB system is illustrated in Fig. 2.

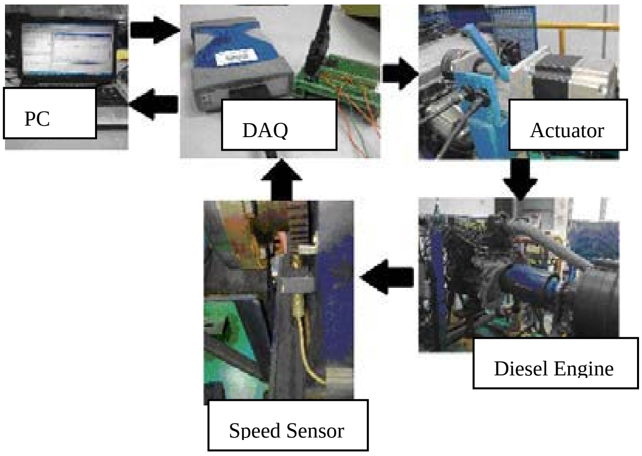

Fig.1. The experimental setup for collecting data from POB system

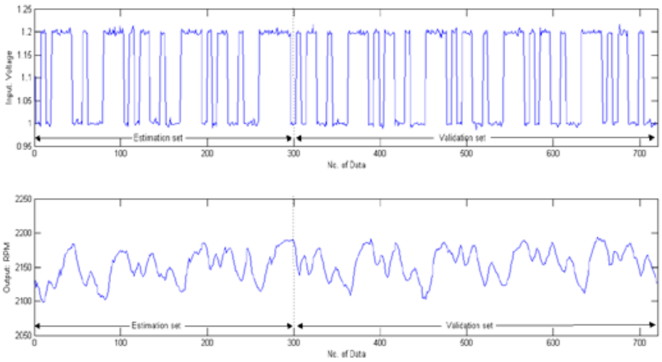

Fig. 2. Input and output data of POB system

3.3 MOODE for Model Structure Selection

The MOODE algorithm is the integration of DE into NSGA-II procedure but without the DE selection process. This technique was proposed by Peng et al.[21] and in this study it is adapted to system identification problem. The main difference compared with single-objective optimization of DE [22] is to produce a set of possible solutions instead of a single solution. The set of solutions are represented as points along the Pareto-optimal front. The Pareto-optimal front is a dense set of the selected solutions between two contradicting objective functions i.e. model predictive accuracy and model complexity that are considered in selection of model structure. Therefore, the algorithm is designed to generate population of possible solutions that scatters as points along Pareto-optimal front that optimized both objective functions. In MOO, the potential solution is selected from the Pareto-optimal front obtained. The flowchart of model structure determination that is formulated by using MOODE is shown in Fig. 3. This flowchart shows the implementation of MOODE that produces a set of optimal model structures.

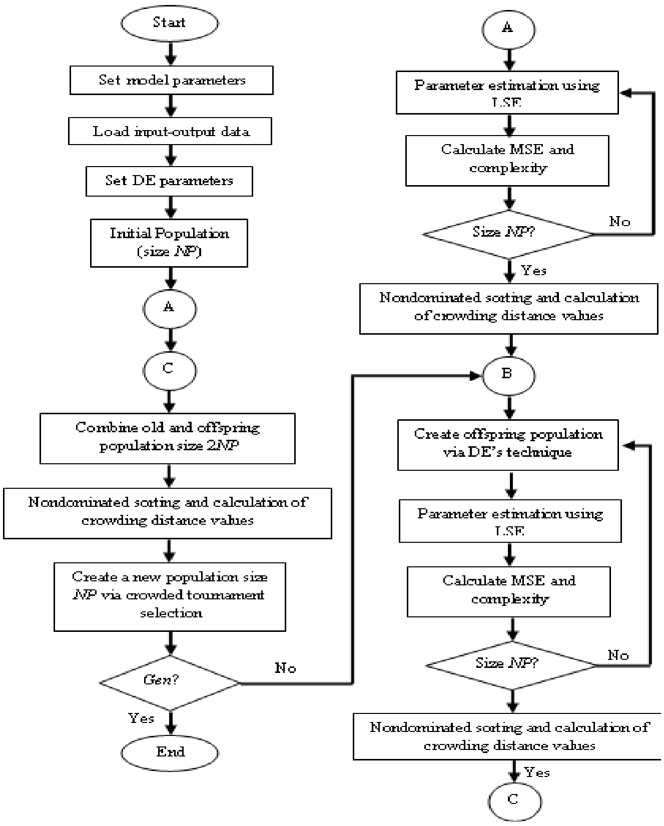

Fig. 3. MOODE's flowchart

4 Results and Discussion

4.1 Simulated System

Systems with known mathematical model structures were excited with the same input signals are called simulated systems. In this study, two simulated systems used in Billings and Wei [23] are considered to show the effectiveness of the proposed algorithm. In this study, those simulated systems are referred as S1 and S2. Table 1 lists the description of each simulated systems.

Table 1. Description of the simulated systems

Systems	Descriptions
S1	y(t)= 0.5y(t-1)-0.45y(t-3)+1.2u(t-3)+0.1u(t-4)
S2	y(t)= 0.797y(t-1)-0.237y(t-3)+0.441u(t-1)+ 0.105y(t-4)u(t-4)+0.333u(t-3)u(t-5)

4.1.1 Test the Effectiveness of MOODE

All simulated systems were excited with random white signals in the range [-0.5, 0.5]. The number of input-output data generated was 400. All 400 data points were used for model estimation and model test. At the initializing stage, the proposed algorithm used the same input-output lags and degree of nonlinearity, where five input-output lags (nu = ny = 5) and two degree of nonlinearity (l = 2) were considered for S1 and S2 systems. The maximum number of model terms is 66 and the the possible models to be chosen from is $2^{66} – 1$ equal to $7.38×10^{19}$. The possible model developed for each simulated system could be nonlinear. The genetic parameters of the algorithm used as listed in Table 2.

In MOODE the solutions produced as a set of solution where those solutions are along a single front known as Pareto-optimal front. The main objective of multi-objective optimization is to arrange all the possible solutions in the Pareto-optimal front. Since all solutions are along the Pareto-optimal fronts, they can be said as potential solutions to be selected from. In this case, the exact model structure of the simulated systems is expected to be positioned along the Pareto-optimal front among the other possible solutions. Fig. 4 and Fig. 5 illustrate the Pareto-optimal front of each simulated system.

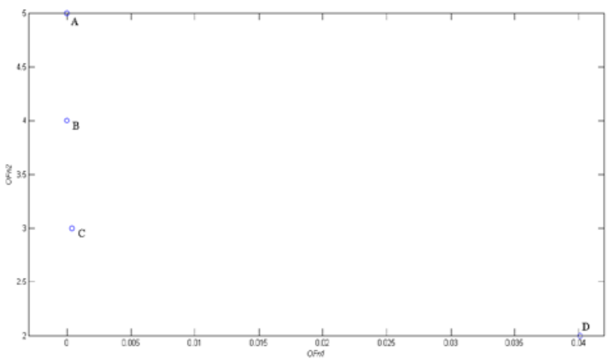

Fig. 4. Pareto-optimal front for S1 system

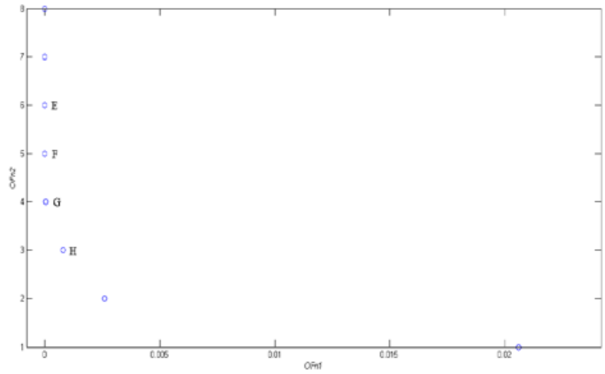

Fig. 5. Pareto-optimal front for S2 system

Table 2. MOODE parameters used in this study

MOODE Parameters	Values
CR	0.7
MR	0.3
NP	50
Gen	100
L	0.1
H	0.4

As can be seen in Fig. 4 and Fig. 5, there are four solutions from each Pareto-optimal front and they are marked. These solutions indicate the possible models of each simulated system. The number of possible solutions for each simulated system is lower than 50, which is the maximum number of solution that equals to the population size used in this study. A reason behind this situation is some solutions are repeated that were counted as single solution. These solutions are referred as repeated solution. Details of marked models are shown in Table 3 and Table 4 for each simulated system.

The exact model structure of S1 system comprises of four terms as shown in Table 4. Four possible models obtained by MOODE were marked as A, B, C, and D as shown in Fig. 4. These models have different model structures where the model complexity of those models is 5, 4, 3, and 2 for A, B, C, and D models respectively. Referring to Table 3, model B has the exact model structure and coefficient as the simulated model. Even though the model developed is nonlinear system with larger number of terms to test the linear S1 system, MOODE is able to capture the exact model structure of S1 system and model B has the most repeated solution with 16 repetitions, which is 32% of the population size. For S2 system, four out of eight possible models were marked as E, F, G, and H. The exact model structure for S2 system comprises of five terms. Model F has the same terms and coefficients as the simulated model as shown in Table 4. Further, Model F has higher number of repeated solution compared with the other models which is 9 repetitions. Thus, a possible model with higher number of repeated solution could be given priority in selecting the final model. This observation is shown by the models which are B and F models with the higher number of repeated solutions.

Table 3. Details of models for S1 system

Terms	Actual	A	B	C	D
Constant	--	--	--	--	--
$y(t\text{-}1)$	0.5000	0.5000	0.5000	0.5474	0.5127
$y(t\text{-}2)$	--	--	--	--	--
$y(t\text{-}3)$	0.4500	-0.4500	-0.4500	-0.4490	--
$u(t\text{-}2)$	--	-4.02×10^{-16}	--	--	--

u(*t*-3)	1.2000	1.2000	1.2000	1.1940	1.1470
u(*t*-4)	0.1000	0.1000	0.1000	--	--
Number of repeated solution		13	16	10	11
*OFn*1		1.66×10^{-31}	8.66×10^{-32}	3.88×10^{-04}	4.01×10^{-02}
*OFn*2		5	4	3	2

Table 4. Details of models for S2 system

Terms	Actual	E	F	G	H
Constant	--	--	--	--	--
y(*t*-1)	0.7970	0.7970	0.7970	0.7999	0.8047
y(*t*-2)	--	--	--	--	--
y(*t*-3)	-0.2370	-0.2370	-0.2370	-0.2380	-0.2429
y(*t*-4)	--	--	--	--	--
y(*t*-5)	--	--	--	--	--
u(*t*-1)	0.4410	0.4410	0.4410	0.4429	0.4422
u(*t*-3)*u*(*t*-5)	0.3330	0.3330	0.3330	0.3289	--
Number of repeated solution		1	9	5	6
*OFn*1		2.82×10^{-32}	1.06×10^{-31}	3.61×10^{-05}	8.01×10^{-04}
*OFn*2		6	5	4	3

4.2 Modeling POB system

A NARX model is used as model representation for the system. The model parameters used are five input-output lags (nu = ny = 5) and two degree of nonlinearity (l = 2), with the genetic parameters of MOODE used is given in Table 2. The tradeoff between two objective functions i.e. model predictive error (OFn1) and model complexity (OFn2) from the final generation is illustrated. This illustration called Pareto-optimal front to show the number of models that represent the POB system.

Fig. 6 shows the non-dominated models obtained from the final generation using the MOODE algorithm. The non-dominated models consist of 21 models of representing the dynamic behavior of POB system. Four out of those models are randomly selected and marked as D1, D2, D3, and D4. All selected models are chosen based on their complexity whereas the model predictive error of each model is not much different. The highest model complexity is 31 terms while the lowest model complexity is 2 terms only as shown in Fig. 6 which gives big difference between these two. However, the model complexity is given priority in choosing the compact model but adequately capture dynamic of the system [8]. The detail of the selected models is listed in Table 5.

Each selected model shows the nonlinear model which consists of single term and cross term as can be seen in Table 5. The model predictive errors of optimal models are all within the range of 1.10 – 1.20×10-03. However, the model complexity of marked models are diverse where the lowest model complexity is 2 terms for model D4 while the highest model complexity is 10 terms for model D1. Observing to the number of repeated solution, model D3 indicates the higher value of repeated solutions with 5 repetitions. Meanwhile, the second higher value of repeated solution comes from models D1 and D4. From the previous study on the simulated systems, the model with the highest number of repeated solution should be given priority to be chosen. Therefore, model D3 is a possible model to be chosen to represent the dynamic behavior of POB system. However, all models have to be validated using the model validity tests, before selecting a good and adequate model.

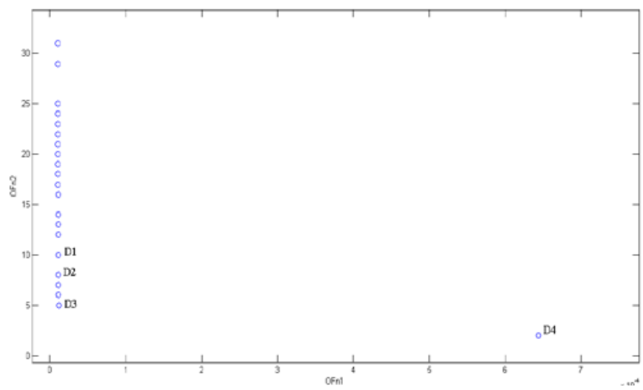

Fig. 6. The non-dominated models obtained for POB system

Table 5. Details of selected models for POB system

4.3 Model Validation

Three model validity tests are considered in this study i.e. OSA test, MPO test, and correlation tests. First set of data which consist of 300 data points are used for estimation set, while the second set of data is used for the validation tests which consist of 420 data points. Figs. 7 to 10 illustrate the OSA test for each selected model. MPO test for the selected models are shown in Figures 11 to 14. These figures show the output of POB system which is engine speed superimposed on predicted model output and its residual plot.

As can be seen in Figs. 10 and 14, the results for model D4 show that the OSA and MPO outputs do not follow the system outputs. Average of the residuals of model D4 is more than 1% indicating that model D4 is unable to capture the dynamic of the system. The results for models D2 and D3 show that the OSA and MPO outputs follow the system outputs very well. These are shown in Figs. 6 and 7 for OSA test and Figs. 12 and 13 for MPO test. However, both models are biased in process model where most of their residuals are only positive values for model D2 and negative values for model D3.

For model D1, both validation tests show good results with the OSA and MPO outputs follow the system output very well without biased. This can be shown in Figs. 7 and 11. The MSE values in estimation and validation sets for OSA and MPO tests are listed in Table 6 and Table 7, respectively. Model D1 shows the lowest MSE values in both estimation and validation sets. Therefore, model D1 is given priority to be chosen in representing the dynamic behavior of the POB system.

Terms	D1	D2	D3	D4
Constant	-0.9930	--	--	--
$y(t\text{-}1)$	2.0000	1.0800	1.0600	--
$y(t\text{-}1)y(t\text{-}3)$	-0.2090	-0.0646	--	--
$y(t\text{-}1)u(t\text{-}1)$	-0.5480	-0.0010	-0.0015	0.0285
$y^2(t\text{-}2)$	--	--	-0.0098	0.4500
$y(t\text{-}2)y(t\text{-}5)$	-0.0111	--	--	--
$y(t\text{-}2)u(t\text{-}4)$	0.0028	--	--	--
$y^2(t\text{-}3)$	--	0.0731	--	--
$y(t\text{-}3)y(t\text{-}4)$	--	--	--	--
$y(t\text{-}3)y(t\text{-}5)$	-0.0354	-0.0476	-0.0224	--
$y(t\text{-}3)u(t\text{-}1)$	0.5470	--	--	--
$u^2(t\text{-}3)$	--	0.0126	--	--
$u(t\text{-}3)u(t\text{-}5)$	0.0098	-0.0128	0.0154	--
$u^2(t\text{-}5)$	0.0049	0.0176	--	--
Number of repeated solution	4	2	5	4
$OFn1\times10^{-05}$	1.10	1.12	1.16	64.4
$OFn2$	10	8	5	2

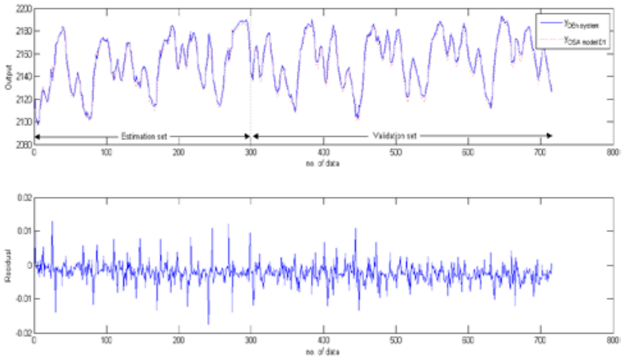

Fig. 7 The engine speed of the POB system superimposed on OSA output of model D1 and its residual plot

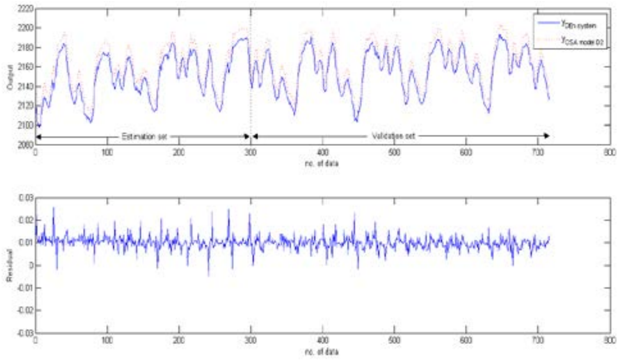

Fig. 8 The engine speed of the POB system superimposed on OSA output of model D2 and its residual plot

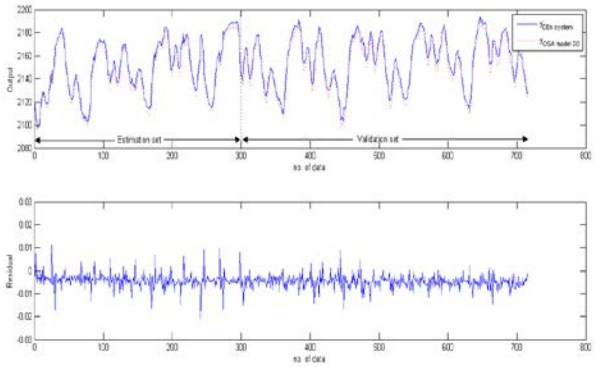

Fig. 9 The engine speed of the POB system superimposed on OSA output of model D3 and its residual plot

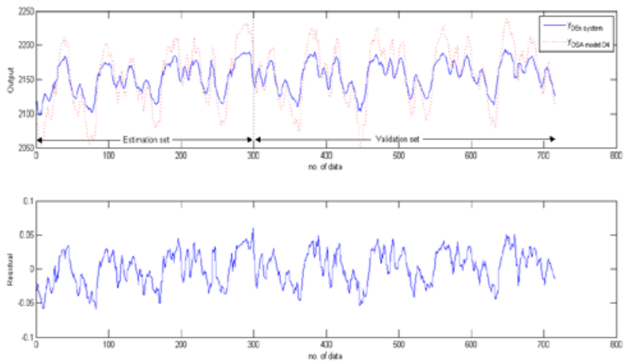

Fig. 10 The engine speed of the POB system superimposed on OSA output of model D4 and its residual plot

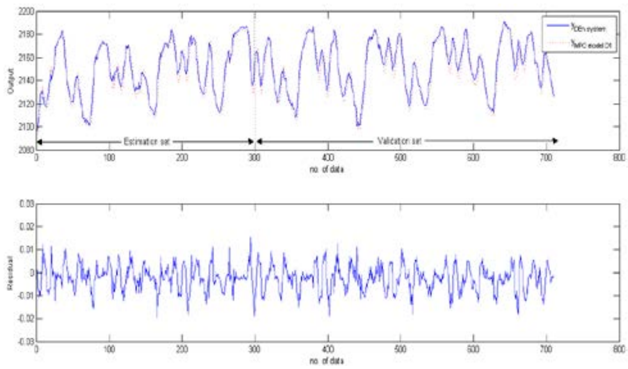

Fig. 11. The engine speed of the POB system superimposed on MPO output of model D1 and its residual plot.

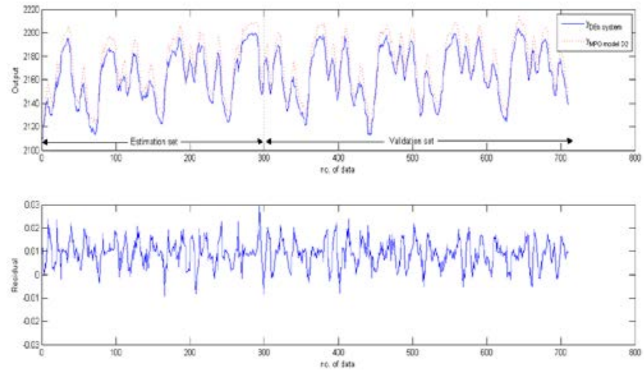

Fig. 12. The engine speed of the POB system superimposed on MPO output of model D2 and its residual plot.

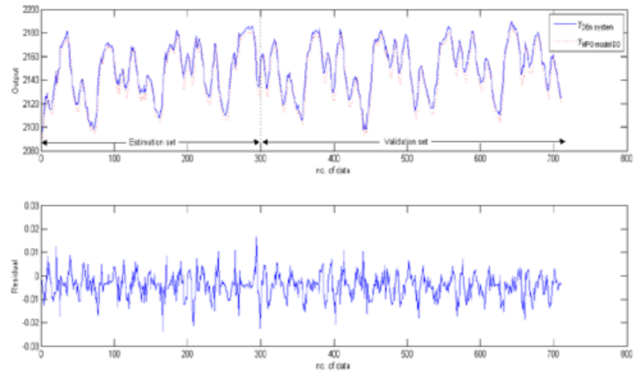

Fig. 13. The engine speed of the POB system superimposed on MPO output of model D3 and its residual plot.

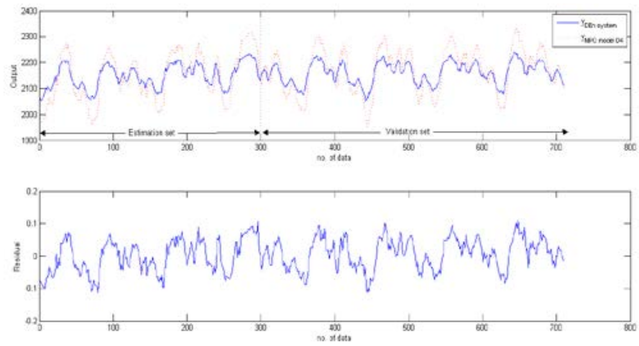

Fig. 14. The engine speed of the POB system superimposed on MPO output of model D4 and its residual plot.

Table 6. MSE estimation and validation values of OSA tests for POB system

Models	OSA test	
	MSE	
	Estimation set	Validation Set
D1	1.39×10^{-5}	1.44×10^{-5}
D2	1.18×10^{-4}	9.94×10^{-5}
D3	2.79×10^{-5}	3.04×10^{-5}
D4	6.38×10^{-4}	5.36×10^{-4}

Table 7. MSE estimation and validation values of MPO tests for POB system

Models	MPO test	
	MSE	
	Estimation set	Validation Set
D1	3.20×10^{-5}	3.76×10^{-5}
D2	1.18×10^{-4}	1.09×10^{-4}
D3	4.11×10^{-5}	4.48×10^{-5}
D4	2.60×10^{-3}	2.20×10^{-3}

The illustrations of correlation tests for each model are shown in Figs. 15 to 18. The correlation tests for model D4 are laid outside of the 95% confidence band for each correlation function of Equation (7). These correlation functions as shown in Fig. 18 are clearly not satisfied. Therefore, this model is inadequate to represent the dynamic behavior of POB system. The correlation tests for other models (D1, D2, and D3) also unsatisfactory. Two out of five correlation functions are unsatisfied: $\emptyset_{\varepsilon\varepsilon}$ and $\emptyset_{(u\varepsilon)\varepsilon}$ as shown in Figs. 15 to 17. However, these models are better than model D4 in term of correlation tests.

Considering all the model validity tests, model D1 is the best model among the selected models to represent the dynamic of POB system. Even though model D1 does not satisfied all the correlation tests, it still can be selected as the best model if based on MPO and OSA tests. This study is focused on selecting an adequate and parsimonious final model based on three model validity tests from the final models in the Pareto-optimal front produced by the developed algorithm. Therefore, details for models that do not satisfy the correlation tests are not discussed here because it is beyond the scopes of work defined. Model D1 can be expressed as:

$$\begin{aligned} y(t) = {} & -0.993 + 2.00y(t-1) \\ & - 0.209y(t-1)y(t-3) \\ & - 0.548y(t-1)u(t-1) \\ & - 0.0111y(t-2)y(t-5) \\ & + 0.0028y(t-2)u(t-4) \\ & - 0.0354y(t-3)y(t-5) \\ & + 0.547y(t-3)u(t-1) \\ & + 0.0098u(t-3)u(t-5) \\ & + 0.0049u^2(t-5) + e(t) \end{aligned}$$

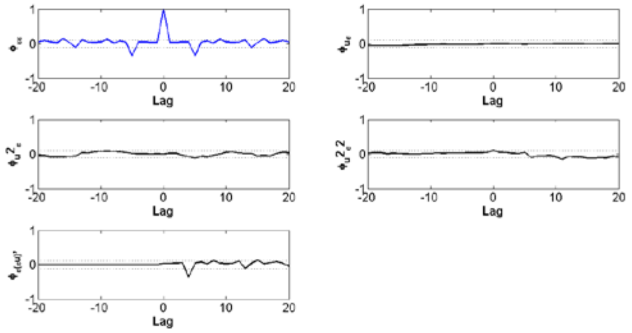

Fig. 15. Correlation tests for model D1

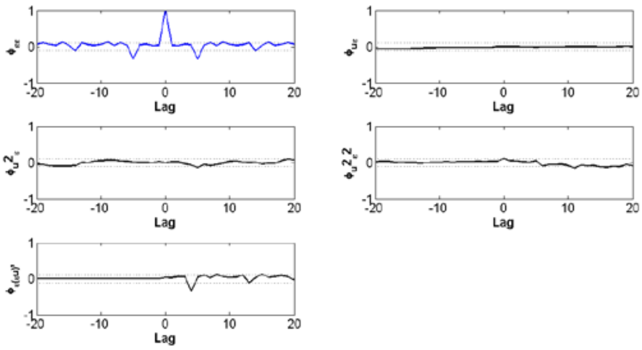

Fig. 16. Correlation tests for model D2

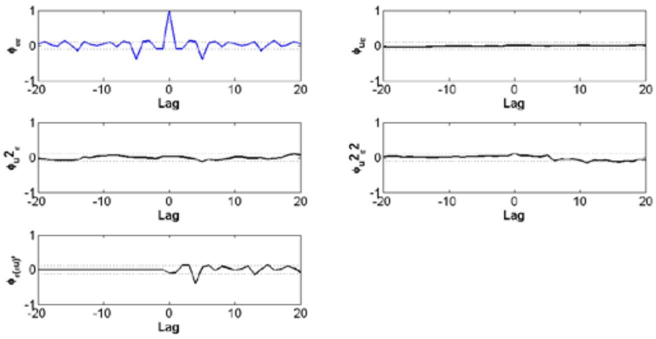

Fig. 17. Correlation tests for model D3

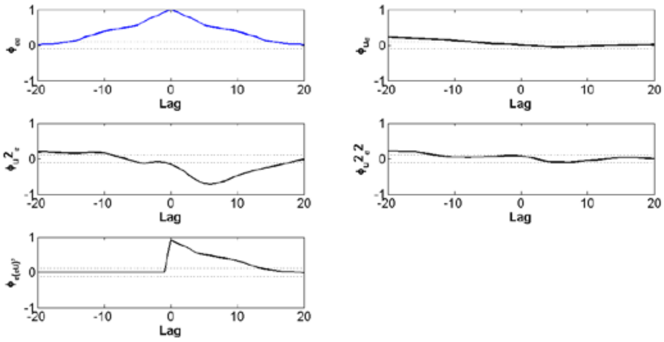

Fig. 18. Correlation tests for model D4

5 Conclusion

In this paper, the input-output data collected from the automotive diesel engine was considered. Modeling POB system using MOODE produces a set of possible models to be chosen from. Four models from the non-dominated front of final generation of MOODE algorithm are selected. The OSA and MPO tests indicate that each model is good in estimating the predicted output of POB system. However, not all models fit to represent the dynamic behavior of this system. The results of correlation tests show that model D1 is the most satisfied compared with the other models. The needs of model validity tests i.e. OSA, MPO and correlation tests to select a good and adequate model in representing the systems to be used concurrently. Therefore, the model validation stage is the important procedure in selecting a good model from the obtained models by MOODE algorithm. The results presented in this study show that the proposed algorithm can be applied as an alternative algorithm to model the dynamic behavior of any process systems.

Acknowledgment. The authors would like to thank Universiti Malaysia Perlis (UniMAP) and Universiti Teknologi Malaysia (UTM) under the

Grant University Project Vote 04H55 for the financial support provided throughout the course of this research.

References:

[1] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, pp. 1-12, 2010.

[2] L. Ljung, *System Identification Theory for the User*. New Jersey: Prentice-Hall Inc, 1999.

[3] R. Ahmad, H. Jamaluddin, and M. A. Hussain, "Model Structure Selection for a Discrete-time Non-linear System using Genetic Algorithm," *Journal of Systems and Control Engineering, Part I. Proc. Instn. Mech. Engrs.*, vol. 218, pp. 85-98, 2004.

[4] W.-D. Chang, "Nonlinear System Identification and Control Using a Real-Coded Genetic Algorithm," *Applied Mathematical Modelling*, vol. 31, pp. 541-550, 2007.

[5] L. Ferariu and A. Patelli, "Multiobjective Genetic Programming for Nonlinear System Identification," in *Adaptive and Natural Computing Algorithms*. vol. 5495, M. Kolehmainen, Toivanen, P., and Beliczynski, B., Ed., ed: Springer Berlin / Heidelberg, 2009, pp. 233-242.

[6] N. Kondo, T. Hatanaka, and K. Uosaki, "Nonlinear Dynamic System Identification Based on Multiobjectively Selected RBF Networks," in *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making (MCDM 2007)*, 2007, pp. 122-127.

[7] B. Subudhi and D. Jena, "A Differential Evolution based Neural Network Approach to Nonlinear System Identification," *Applied Soft Computing*, vol. 11, pp. 861-871, 2011.

[8] M. Z. Zakaria, H. Jamaluddin, R. Ahmad, and S. M. R. Loghmanian, "Comparison between Multi-Objective and Single-Objective Optimization for the Modeling of Dynamic Systems," *Journal of Systems and Control Engineering, Part I. Proc. Instn. Mech. Engrs.*, vol. 226, pp. 994-1005, 2012.

[9] R. M. Storn and K. V. Price, "Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces," vol. Technical Report TR-95-012, p. ICSI, 1995.

[10] L. d. S. Coelho, V. C. Mariani, and J. V. Leite, "Solution of Jiles–Atherton Vector Hysteresis Parameters Estimation by Modified Differential Evolution Approaches," *Expert Systems with Applications*, vol. 39, pp. 2021-2025, 2012.

[11] T. W. Liao, "Two Hybrid Differential Evolution Algorithms for Engineering Design Optimization," *Applied Soft Computing*, vol. 10, pp. 1188-1199, 2010.

[12] X. Wang and G. Xu, "Hybrid Differential Evolution Algorithm for Traveling Salesman Problem," *Procedia Engineering*, vol. 15, pp. 2716-2720, 2011.

[13] Y. Wang, B. Li, and T. Weise, "Estimation of Distribution and Differential Evolution Cooperation for Large Scale Economic Load Dispatch Optimization of Power Systems," *Information Sciences*, vol. 180, pp. 2405-2420, 2010.

[14] U. Yüzgeç, "Performance Comparison of Differential Evolution Techniques on Optimization of Feeding Profile for an Industrial Scale Baker's Yeast Fermentation Process," *ISA Transactions*, vol. 49, pp. 167-176, 2010.

[15] B. V. Babu, P. G. Chakole, and J. H. Syed Mubeen, "Multiobjective Differential Evolution (MODE) for Optimization of Adiabatic Styrene Reactor," *Chemical Engineering Science*, vol. 60, pp. 4822-4837, 2005.

[16] S. Chen and S. A. Billings, "Representations of Nonlinear Systems: The NARX Model," *International Journal of Control*, vol. 49, pp. 1013-1032, 1989.

[17] R. Ahmad, "Identification of Discrete-Time Dynamic Systems Using Modified Genetic Algorithm," PhD PhD Thesis, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, Johor Bahru, 2004.

[18] J. L. Wright, "New evapotranspiration crop coefficients," *J. of Irrig. and Drain. Div.*, vol. 108, pp. 57-74, 1982.

[19] S. A. Billings and W. S. F. Voon, "Correlation Based Model Validity Tests for Non-Linear Models," *International Journal of Control*, vol. 44, pp. 235-244, 1986.

[20] M. N. Azuwir, M. Z. Abdulmuin, and A. H. Adom, "Modelling and Validation of Automotive Engine Fuelled with Palm Oil Biodiesel," *International Journal of Engineering and Technology,* vol. 3, pp. 582-586, 2011.

[21] C. Peng, H. Sun, and J. Guo, "Multi-Objective Optimal PMU Placement using a Non-dominated Sorting Differential Evolution Algorithm," *International Journal of Electrical Power & Energy Systems,* vol. 32, pp. 886-892, 2010.

[22] R. M. Storn and K. V. Price, "Differential Evolution - A Simple and Efficient Heuristics for Global Optimization over Continuous Spaces," *Journal of Global Optimization,* vol. 11, pp. 341-359, 1997.

[23] S. A. Billings and H. L. Wei, "An Adaptive Orthogonal Search Algorithm for Model Subset Selection and Non-linear System Identification," *International Journal of Control,* vol. 81, pp. 714-724, 2008.