

THE STUDY OF DYNAMIC RESPONSE USING ARX MODEL IN EXTRACTION PROCESS

Z. M. Yusoff^{1,*}, M. A. A. Aziz², N. F. Razali¹, S. A. Nordin¹, A. F. Z. Abidin³ and N. M. Zain¹

¹Faculty of Electrical Engineering, Universiti Teknologi MARA, 81750 Pasir Gudang, Johor, Malaysia

²Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

³Faculty of Engineering Technology, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia

Published online: 05 October 2017

ABSTRACT

This work presents a model using system identification approach namely as ARX to represent the dynamic response for essential oil extraction process. A fresh set of data under feed in disturbance was collected using MATLAB Simulink. The 3000 samples of data was collected by using PRBS as an input and temperature in °C as an output. The collected data was separated into two groups; training data and estimation data by using interlacing technique. The model estimation was done by using linear regression method. The robustness of the model was evaluated by using best fit (R^2), OSA, root mean square error (RMSE), correlation analysis and residual analysis (histogram). Based on validation results, the ARX model was successfully capturing the dynamic response of extraction process by provide the high best fit, low RMSE error and normally distributed by producing small mean and variance.

Keywords: auto-regressive with exogenous input (ARX); pseudo-random binary (PRBS); interlacing technique; prediction error method; one-step ahead prediction (OSA).

Author Correspondence, e-mail: zakiah9018@johor.uitm.edu.my

doi: <http://dx.doi.org/10.4314/jfas.v9i4s.8>



1. INTRODUCTION

Extraction process is the process of breaking the material pockets with the aims to extract the aromatic compounds from the plant [1-3]. It is very highly aromatic and more concentrated than the original plant and can be found mostly in flowers and leaves. Recently, the essential oil gained great popularity and high demand from industries since it can be used in many applications such as in aromatherapy, medical, as a food flavour, insect repellent, household merchandise, tumour activity, cosmetic and etc. Essential oil or sometimes called as a volatile oil contains a mixture of terpenes, oxygenated terpenes, sesquiterpenes and oxygenated sesquiterpenes. The quality of the essential oil is depend on the composition of the active constituents and can be validated by using Gas Chromatography/Mass Spectrometry (GC-MS) analysis. There are many existing methods to extract this valuable compound such as steam temperature, hydro distillation, soxhlet extraction, solvent extraction, superheated water extraction, microwave hydro diffusion and gravity, ultrasound-assisted distillation, cold pressing, supercritical carbon dioxide, supercritical fluid extraction and etc. [4-5]. All these methods have their own advantages and disadvantages. The steam distillation is one of the conventional method to extract the essential oil and most preferable in industries and research domain. The steam distillation technique becomes a benchmark to the other methods to compare the quality and quantity of essential oil. The advantages of this method are low operational cost, clean, in terms of productivity and low maintenance cost. Existing research found that quality of the essential oil is quite similar and more effective as compared with the supercritical fluid extraction method. The loss of valuable constituents, the efficiency and thermal degradation problems also can be prevent by steam distillation. It does also can produce more volume of oil as compared with the recent techniques such as superheated and soxhlet. During the extraction process, many significant parameters need to be consider such as temperature, size, extraction time, pressure and flow rate since its will give large effect in terms of quality and quantity [2, 6]. Existing research stressed that the steam temperature is one of the most significant parameters that should give more attention since it will influence the production yield and quality [7-8].

Recently, the application of system identification growth rapidly and becomes one of the

important tool for many sector in order to develop the sophisticated algorithms. Academically, system identifications mean that the process to build the mathematical model from the set of input output data to describe the underlying process [9-11]. Usually, modelling using system identification approach is necessary to study the dynamic behaviour about the extraction process. In a series of existing papers stated that modelling is necessary for controller design, predict and study the dynamic of a process, to prevent any dangerous situation, predict the energy consumption and optimize the operation condition. Modelling can be divided into two groups which are linear and nonlinear model. The Finite Impulse Response (FIR), Auto-regressive with Exogenous Input (ARX) model, Auto-regressive Moving Average with Exogenous Input (ARMAX), Output-Error (OE) and Box Jenkins model are the list of the linear model. Meanwhile, the example for the nonlinear model such as artificial neural Network (ANN), Nonlinear Auto-regressive with exogenous Input (NARX) [12], Neural Network Auto-regressive with Exogenous Input (NNARX) and NARMAX [10, 24].

In order to produce the excellent result in controller design, the dynamic model of extraction process can be model by using linear model. The linear system identification starts in the beginning of 1990s and already uses many applications such as extraction process, robot, boiler system, heat exchanger, furnace, manufacturing and etc. [13]. Linear model is much simpler and can represent the dynamic system in small operating range as compared with the nonlinear model. The common linear models are Auto-Regressive with Exogenous (ARX), Auto-Regressive Moving Average with Exogenous (ARMAX), Output Error and Box-Jenkins model [1, 14-17]. One of the most popular linear models is ARX model. Existing research in the track of linear system identification reported that the linear ARX model is adequate to represent and capture the dynamic behaviors of extraction process [18-19]. It does can provide the best performance at high frequency as compared with the ARMAX model and commonly used in industry since its simple in determining the parameters. Meanwhile, the nonlinear model is applied when the process operate in wide operating range. Consequently, the nonlinear models able to capture more accurately the process dynamic thus, making the adequate for controlling a real process in a wide region of operation [9].

The linear system identification consists of four steps. It start with the experiment design to

collect the input output data from the system. The collective data must be informative enough to represent the dynamic system. The data can be collected either open or closed loop system using a computer. The main important thing during the data collection process is the type of input signal. The input should contain these 2 main characteristics:

- Must be fluctuating and persistently exciting to represent all the interest's dynamic.
- Should have different level of amplitude

The example of commonly input signal are Pseudo Random Binary Sequence (PRBS), step function, white noise, sine wave random binary sequence (RBS), random multi-level and Gaussian signal. The selection of input signal is depended on the applications. However, previous publication reported that the PRBS input signal is the most popular because of signal varies between two values; its power spectrum and the signal can be concentrated in frequency ranges. Then, the collected data must go to the pre-processing stages before further used to identify the undesired noise that can affect the result. Filtering, magnitude scaling and removing outliers are the examples of pre-processing. Filtering means that the process of removing unwanted noise from the input output data. Then, removing is the process of removing outliers that will affect the estimated model and done before training the collected data. Meanwhile, the magnitude scaling is the process to scale the process variable into the same class. The other important thing during data collection is the sampling interval. Existing research suggest that the sampling interval should be in between 1% to 5% of the time constant. There should have some knowledge on the process dynamic so that the sampling interval is not too small or too large.

Second step is the model order selection either by heuristic method or numerical approach. The example for heuristic method is trial and error method and for numerical approach such as Akaike's Final Prediction Error (FPE), Lipchitz function, Rissanen's minimum description length and normalized sum of squared error (NSSE). The selection of the method is depending on the purpose of application.

Third is the model estimation. The purpose of the model estimation is to estimate the unknown parameters with the low residual. The model can be classified as a good model when it shows the small prediction error. There are two establish approaches which are correlation and prediction error method (PEM). The difference between these two approaches is the first method using the least square, meanwhile the second method is using the concept

of minimizes the loss function. However, PEM also known as a residual is most preferable for most researchers [20-21].

Finally is the model validation. This is the important step before the best model chosen deliver to the user. The chosen model should represent the same dynamic response as a real system. The estimated model will be validating either by graphical or statistical method. The graphical approach means that the result will be interpret by using the common sense and more simple as compared to the graphical approach. Meanwhile, the statistical approach produces the results using the statistical method. There are some common test applied in validation stage such as one-step ahead prediction (OSA prediction), residual histogram and correlation function [22]. The most popular validation method is the OSA prediction. It calculate the model fit by measure the difference between the predicted output and measured output, also call as a residual.

2. MATERIAL AND METHOD

2.1. ARX Structure

Generally, the block diagram for the dynamic system is shown in Fig. 1 where $u(t)$ is represent the input, $y(t)$ is represent the output and $e(t)$ is the error.

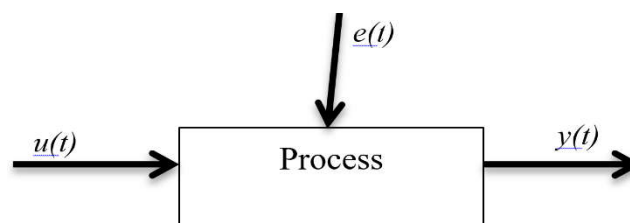


Fig.1. The dynamic system

The Auto-Regressive with Exogenous (ARX) model was design to represent the system dynamic of extraction process. The model structure can be describes as Equation (1) [23]:

$$A(q)y(t) = B(q)u(t - nt) + e(t) \quad (1)$$

where $u(t)$ = system input, $y(t)$ = system output, nt = time delay, $e(t)$ = noise disturbance and $A(q)$ and $B(q)$ = polynomials. The equations for $A(q)$ and $B(q)$ are defined as Equation (2) and (3) respectively:

$$A(q) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a} \quad (2)$$

$$B(q) = 1 + b_0 + b_1q^{-1} + \dots + b_{n_b}q^{-n_b} \quad (3)$$

where q^{-1} represents the backward shift generator and $[n_a, n_b, n_k]$ represent the number of poles, number of zeros and time delay in ARX model [23]. Fig. 2 shows the general structure for ARX model.

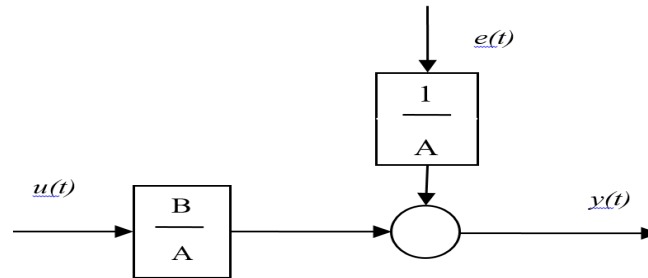


Fig.2. ARX's model structure

2.2. Model Estimation

The unknown parameters of ARX structure were estimate by using linear regression approach by using the concept of minimizing the sum of square residual. The definition of regress is calculate the value of output $y(t)$ at any point in time domain based on previous outputs and inputs. The general equation for linear regression as state in Equation (4):

$$y(t) + a_1y(t-1) + \dots + a_{n_a}y(t-n_a) = b_1u(t-1) + \dots + b_{n_b}u(t-n_b) + e(t) \quad (4)$$

where $y(t)$ = output at time t , $a_1 \dots a_n$ and $b_1 \dots b_n$ = estimated parameters, $y(t-1) \dots y(t-n_a)$ = past outputs, $u(t-1) \dots u(t-n_b)$ = inputs and $e(t)$ = disturbance. Next, rearrange the Equation (4) in terms of past input and past output.

$$y(t) = -a_1y(t-1) - \dots - a_{n_a}y(t-n_a) + b_1u(t-1) + \dots + b_{n_b}u(t-n_b) + e(t) \quad (5)$$

Then, introduction of regression vector as Equation (6):

$$\varphi(t) = [-y(t-1) - y(t-2) - \dots - y(t-n_a) \quad u(t-1) \dots u(t-n_b)]^T \quad (6)$$

The adjustable parameters are:

$$\theta = [a_1 \dots a_{n_a}, b_1 \dots b_{n_b}]^T \quad (7)$$

Lastly, Equation (5) can be representing in matrix form as in Equation (8):

$$y(t) = \varphi^T(t)\theta + e(t) \quad (8)$$

where $\varphi(t)$ denote as the regression vector.

2.3. Model Validation

The main objective of the model validation is to validate the dynamic model that represents the system response as a real process. The validation was done by using the one part of experimental data that was separated by using interlacing technique. There are many techniques apply in model validation such as one-step ahead prediction (OSA prediction), residual analysis, coefficient of determination root mean square error and correlation analysis.

2.4. System Identification

The system identification was applied in order to determine the dynamic response of the extraction process. To get the mathematical model for ARX, it involves five stages which are data collection, data separation, model order selection, model estimation and lastly is model validation as shown in Fig. 3.

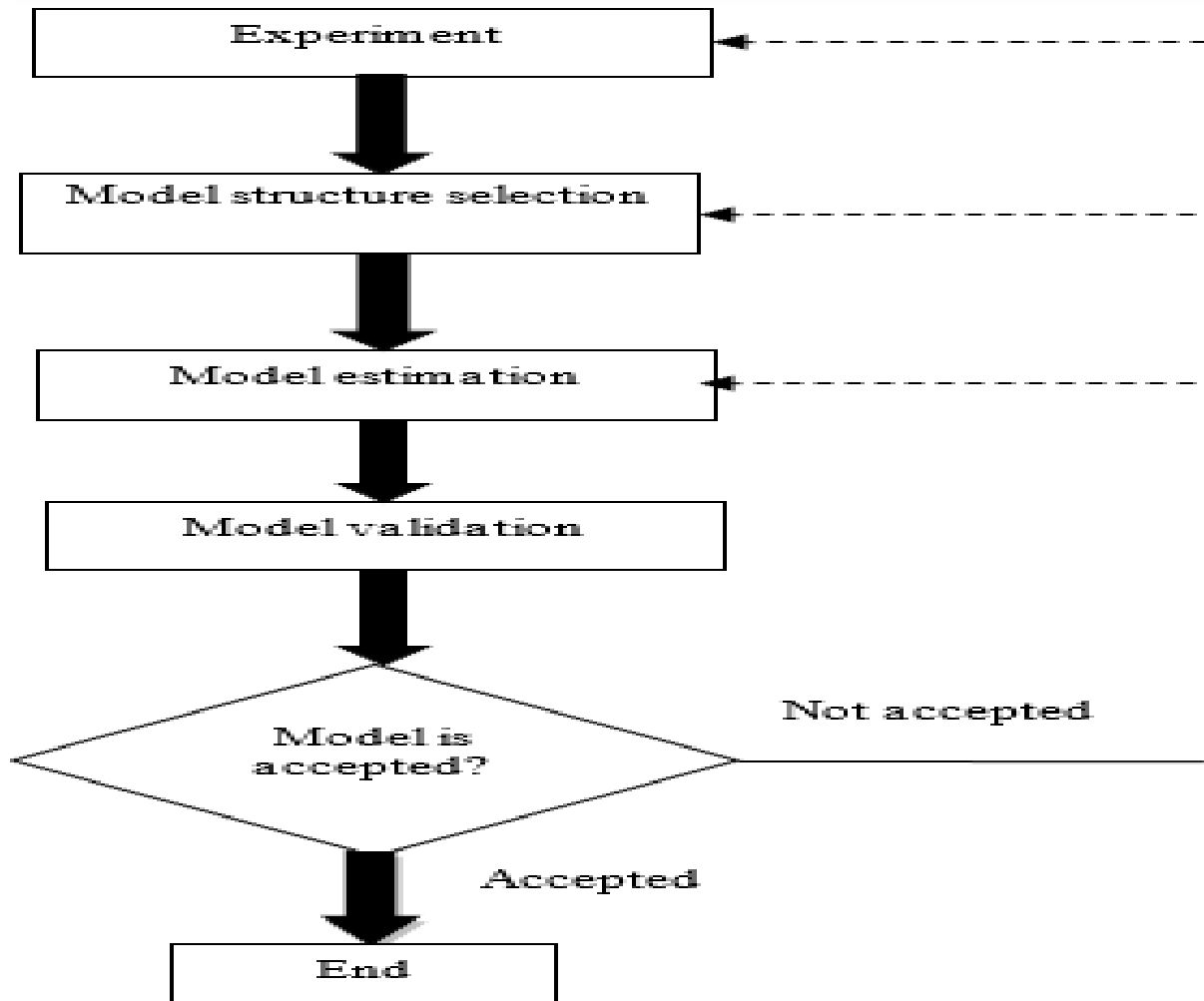


Fig.3. The flow of system identification

2.5. Data Collection

The data acquisition (DAQ) card was attached to the extraction plant to collect the input output data. The data collection was done by using MATLAB programming with 1 second sampling period. Fig. 4 shows the data collection from essential oil extraction process where (a) input PRBS with proportional band, PB (0.4) denoted as $u(t)$ and (b) output, steam temperature denoted as $y(t)$ in $^{\circ}\text{C}$. The data collection was done under feed-in disturbance by using the MATLAB software. The input output data were saving in mat file format. From the data, it shows that the output temperature is depending on the PRBS switching. The data consists of 3000 samples, starting from 76°C to 92°C .

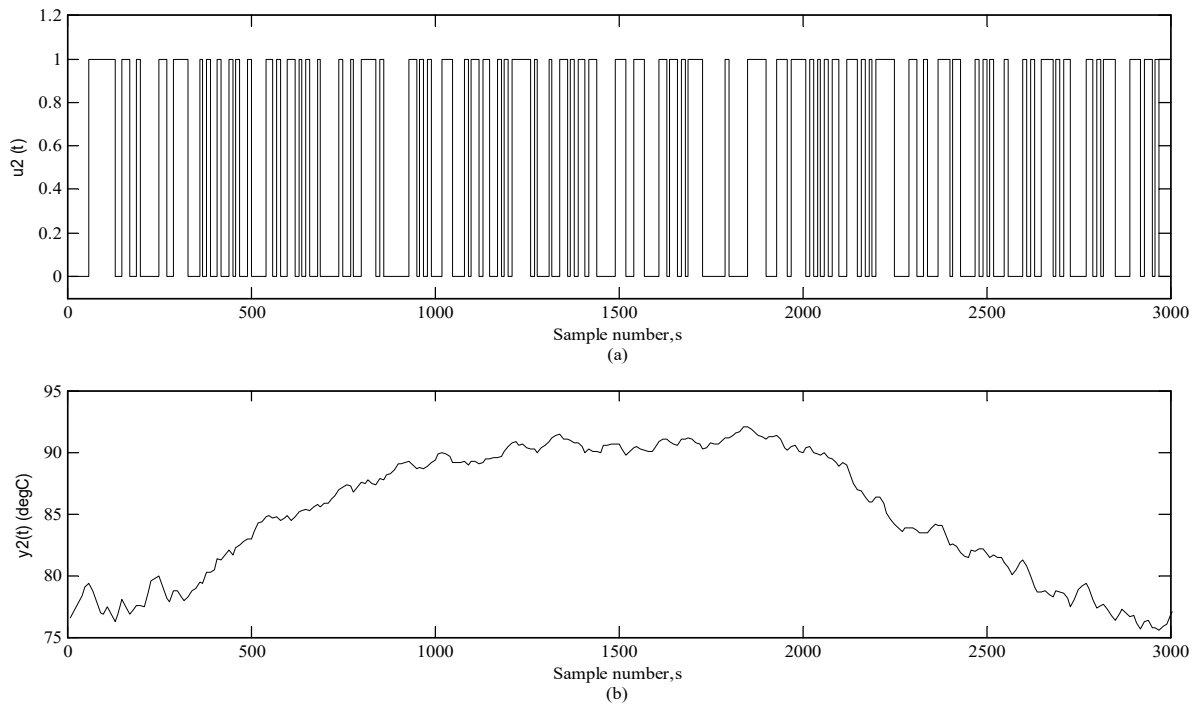


Fig.4. Data collection (a) input (b) output

From Fig. 4, it can be noted that the starting temperature is at 76 °C and keep increasing until 92 °C. Then, the steam temperature suddenly drops at the sample 1930th because of the system is suddenly disturbed by feed-in disturbance.

2.6. Data Separation

The data was separated into two groups as shown in Fig. 5 namely as estimation data and validation data by using interlacing technique. Fig. 5 (a) is estimation data and (b) is the estimation data. Each group will have 1500 samples with 2 seconds interval. The odd sample's (1, 3, 5, 7...) is estimation while the even sample is validation. Table I tabulated the separation data.

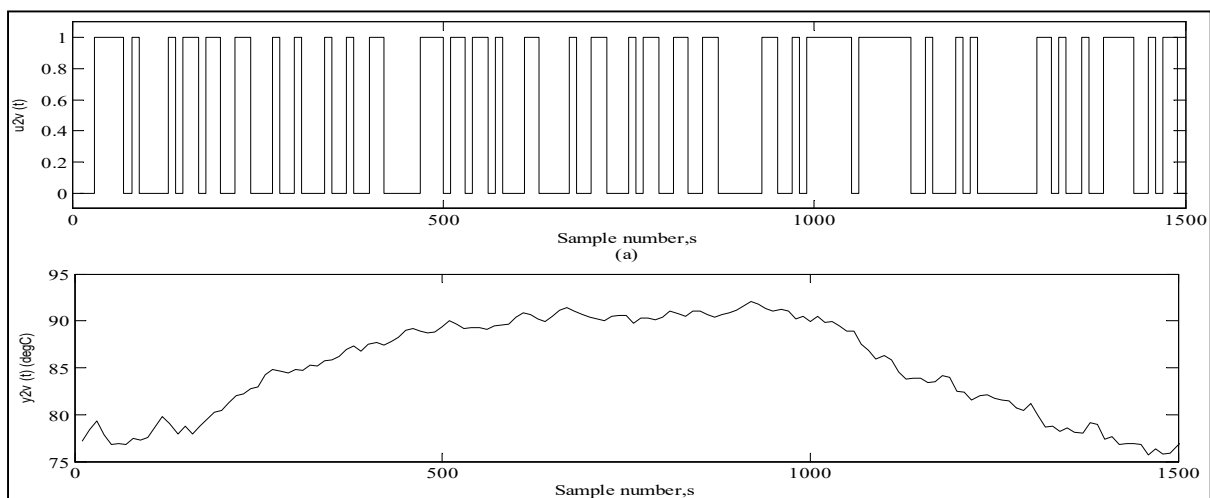


Fig.5. (a) Estimation data

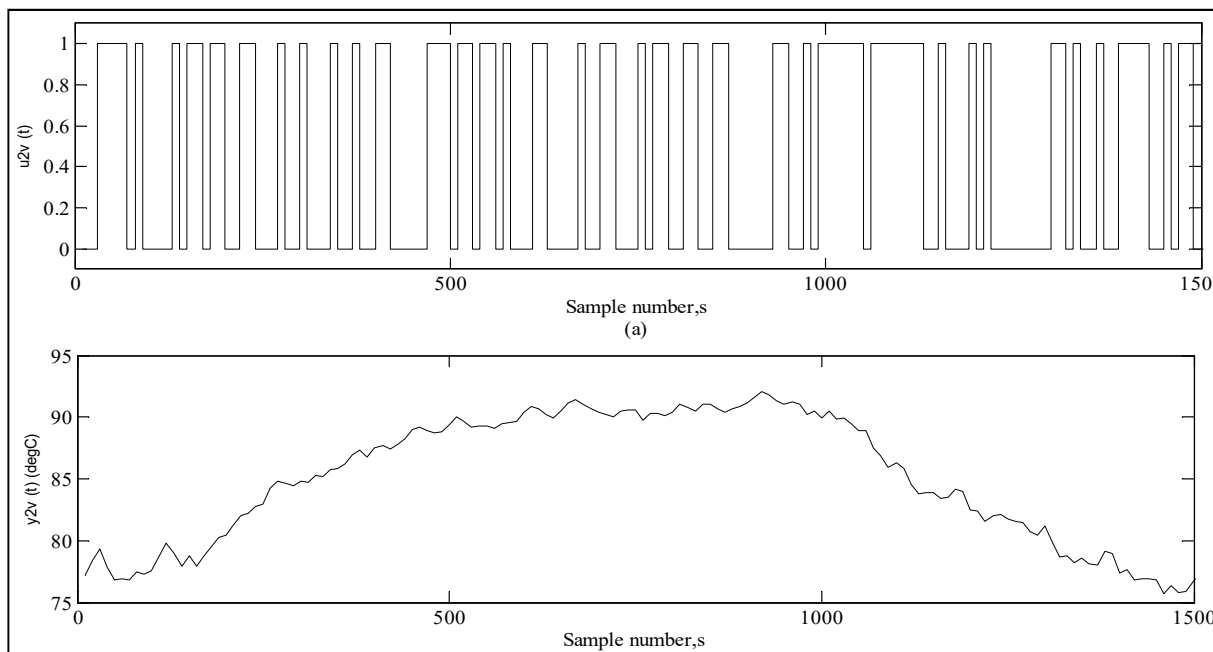


Fig.5. (b) Validation data

Table 1. Estimation and validation data

Data	Estimation Data	Validation Data
(u, y)	z_{ue} and z_{ye}	z_{uv} and z_{yv}

From Table 1, the estimation data is consisting of 2 sets of data which is z_{ue} (estimation input) and z_{ye} (estimation output). Meanwhile, the validation data consists of z_{uv} validation input) and z_{yv} (validation output).

2.7. Model Order Selection

The ARX model with first order was selected based on the previous work. Many existing study reported that the first order is adequate to represent the heating process. So, ARX 111 was chose to represent the essential oil extraction process [20].

2.8. Model Estimation

The model estimation to calculate the parameter for ARX was done by using the linear regression method. The error is representing in terms of loss function. The function arx was use to estimate the unknown parameter of $A(q)$ and $B(q)$. The equation for calculating the loss function based on Akaike’s Final Prediction Error (FPE) as below:

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta) \tag{9}$$

where $y(t)$ = actual output, $\hat{y}(t|\theta)$ = predicted output and θ = estimated parameter.

2.9. Model Validation

The main objective of the model validation is to make sure the selected model can represent the same response as the real process. The statistical analysis using best fit denoted as R^2 , one-step ahead prediction (OSA) residual and correlation analysis were applied by comparing estimated model with the validation data. The equation for best fit as state as below:

$$\text{Best fit} = 100 \left[1 - \frac{\text{norm}(\hat{y}-y)}{\text{norm}(y-\bar{y})} \right] \quad (10)$$

where y is the actual output and \hat{y} is the average output. The higher best fit, R^2 indicates that the model is more accurate.

To be more accurate, the residual in RMSE value between predicted and validation was measured using OSA method. This was done using MATLAB command line using `iddata` function. The OSA prediction is representing as Equation [10]:

$$\hat{y} = \hat{f}(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), \varepsilon(t-1, \hat{\theta}), \dots, \varepsilon(t-n_\varepsilon, \hat{\theta})) \quad (11)$$

$\hat{f}(\cdot)$ denoted as the estimate of $f(\cdot)$, meanwhile $\varepsilon(t, \hat{\theta})$ denoted as residual.

The residual analysis was done by using histogram. The MATLAB command line using the function `hist` was used to measure the mean and variance.

3. RESULTS AND DISCUSSION

The result for ARX111 in modelling the steam temperature for the extraction process is shown in Fig. 6 where (a) measured output, (b) OSA output and (c) residual. The result clearly shows that the predicted output is in good agreement with the measured output when the residual lies in between -0.3 and 0.3. The best fit, R^2 is 97.68% and RMSE value is 0.0145.

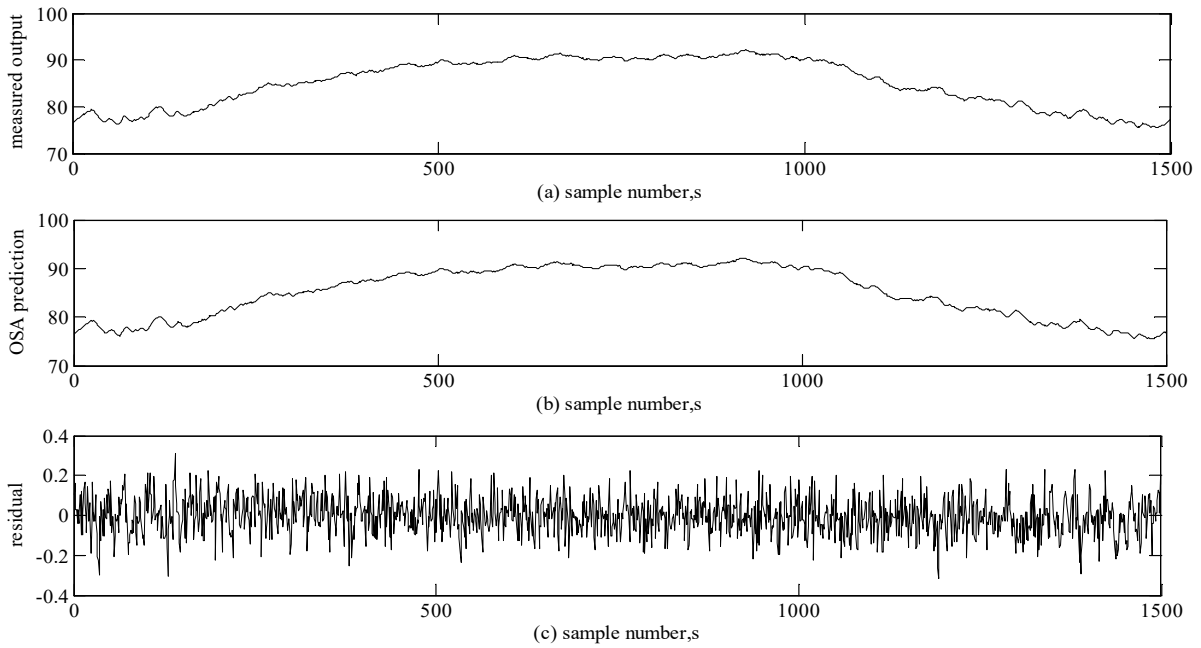


Fig.6. (a) measured output (b) OSA (c) residual

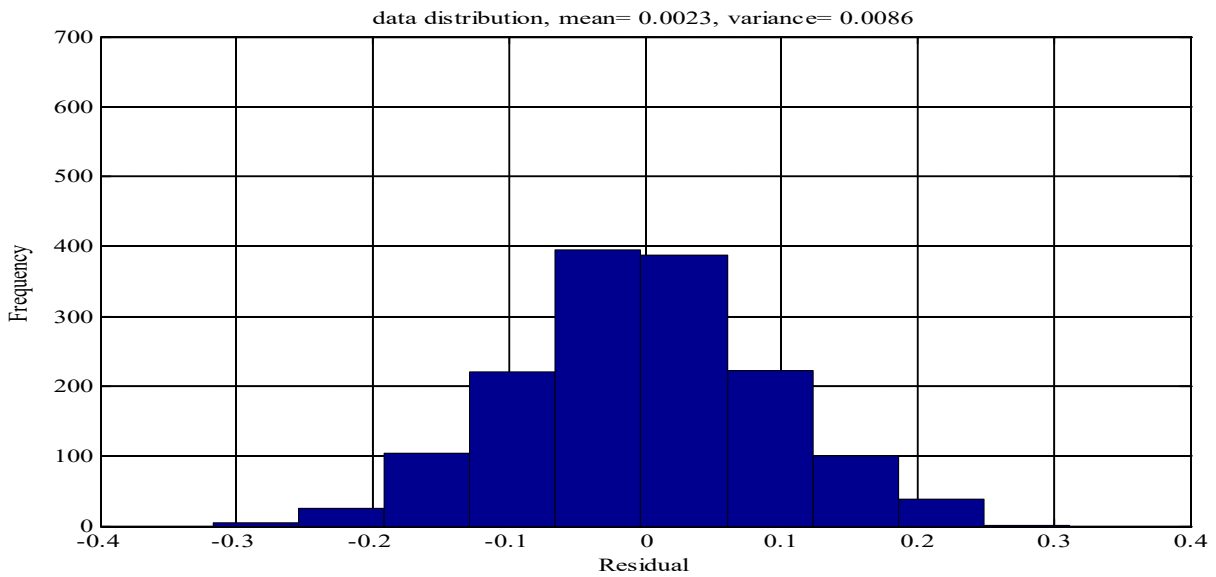


Fig.7. Residual histogram

Fig. 7 shows that the residual is normally distributed by producing small mean and variance. The mean is 0.0023 meanwhile the variance is 0.0086 respectively.

The result for discrete transfer function for ARX 111 is tabulate in Table 2. The transfer function is derived by numerator, $B(q)$ over denominator, $A(q)$.

Table 2. Transfer function for ARX111

Model's Name	ARX Model	Transfer Function
ARX111	$A(q) = 1 - q^{-1}$ $B(q) = 0.01227q^{-1}$	$arx2 = \frac{0.01227z^{-1}}{1 - z^{-1}}$

Next, the above transfer function was rearranged to represent in equation form for easier interpretation and observation as below:

Table 3. Estimated ARX model

Model	ARX Model
ARX111	$y(t) + y(t - 1) = 0.01227u(t - 1) + e(t)$

Fig. 8 depicted the result for correlation analysis using auto-correlation function (ACF) test and cross-correlation function (CCF) test. The ACF test of the residual and CCF test between input residual as shows in Fig. 8 (a) and Fig. 8 (b) respectively. The dashed line in both figure indicated that the 95% confidence limit. Referring to Fig. 8 (a) indicated that the residual is correlated and some of the correlation coefficients are detected outside confidence limit. The correlation was detected at lags 1, 2, 3, 4 and 5. Meanwhile, for CCF test, it showed that there is a correlation between input and residual since some correlation coefficients lie outside confidence limit at lags 2, 3, 4, 5, 6 and 7.

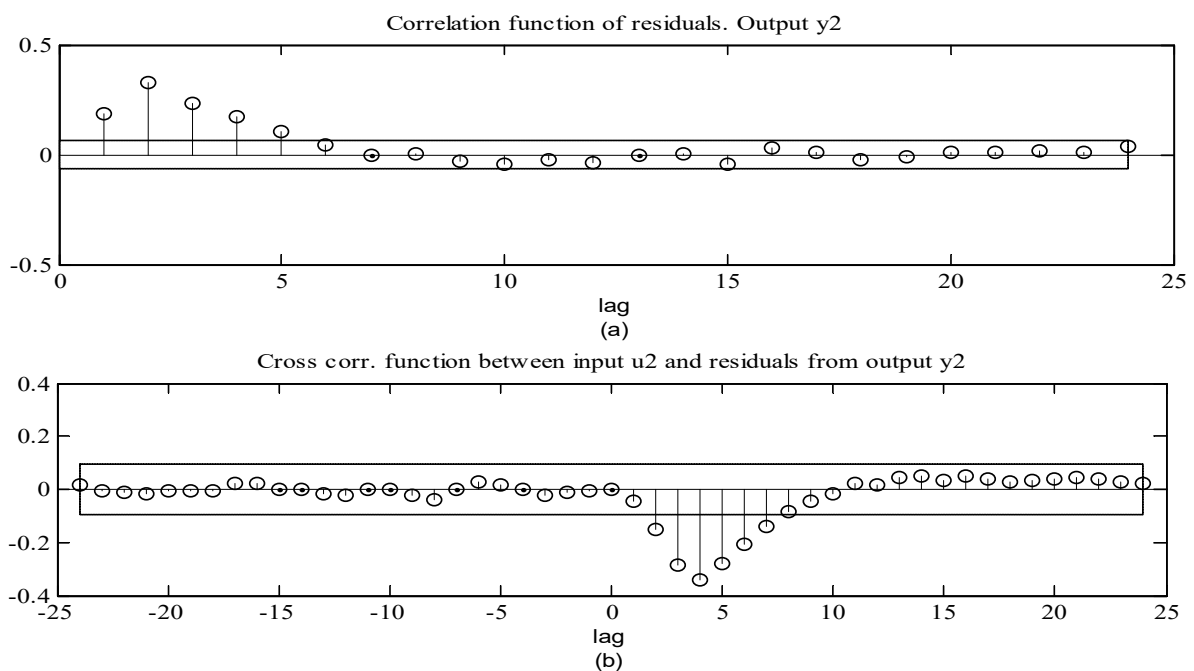


Fig.8. Correlation test

4. CONCLUSION

From the results, it can be conclude that the ARX model with first order is enough to represent the system dynamic of extraction process by produce the good best fit, RMSE, mean and variance. The model was able to capture the dynamic behavior of the extraction process.

5. ACKNOWLEDEGMENTS

The authors acknowledge the support from the UiTM Pasir Gudang, IRAGS under project code: 600-RMI/IRAGS 5/3 (57/2015) and faculty of Electrical Engineering, UiTM Selangor for facilities and financial provided during this research. We also would like to express our gratitude to all members for the idea and endless support.

6. REFERENCES

- [1] Yusoff Z M, Muhammad Z, Rahiman M H, Taib M N. ARX modeling for down-flowing steam distillation system. In IEEE 8th International Colloquium on Signal Processing and its Applications, 2012, pp. 485-490
- [2] Muhammad Z, Yusoff Z M, Rahiman M H, Taib M N. The self-tuning fuzzy PID to regulate the steam temperature in real-time process. In IEEE Conference on Systems, Process and Control, 2015, pp. 28-33
- [3] Hidro P. Effect of temperature on kaffir lime oil by using Hydro-diffusion steam distillation system. Malaysian Journal of Analytical Sciences, 2013, 17(2):326-339
- [4] Franco-Vega A, Ramírez-Corona N, Palou E, López-Malo A. Estimation of mass transfer coefficients of the extraction process of essential oil from orange peel using microwave assisted extraction. Journal of Food Engineering, 2016, 170:136-143
- [5] Huang C Y, Kuo J M, Wu S J, Tsai H T. Isolation and characterization of fish scale collagen from tilapia (*Oreochromis sp.*) by a novel extrusion–hydro-extraction process. Food Chemistry, 2016, 190:997-1006
- [6] Yusoff Z M, Muhammad Z, Ahmad N D, Rahiman M H, Taib M N. Hybrid fuzzy plus PID controller of hydro-diffusion steam distillation essential oil extraction system: Design and performance evaluation. AIP Conference Proceedings, 2016, 1774(1):1-7
- [7] Santos J S, Deolindo C T, Esmerino L A, Genovese M I, Fujita A, Marques M B, Rosso N D, Dagher H, Valese A C, Granato D. Effects of time and extraction temperature on

phenolic composition and functional properties of red rooibos (*Aspalathus linearis*). *Food Research International*, 2016, 89:476-487

[8] Kaushik P, Dowling K, Adhikari R, Barrow C J, Adhikari B. Effect of extraction temperature on composition, structure and functional properties of flaxseed gum. *Food Chemistry*, 2017, 215:333-340

[9] Shariff H M, Rahiman M H. Black-box modeling of steam distillation pilot-scale using M-level PRS perturbation. In *IEEE 5th Control and System Graduate Research Colloquium*, 2014, pp. 85-90

[10] Yassin I M, Zabidi A, Ali M S, Tahir N M, Abidin H Z, Rizman Z I. Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm. *International Journal on Advanced Science, Engineering and Information Technology*, 2016, 6(5):630-637

[11] Yusoff Z M, Muhammad Z, Rahiman M H, Taib M N. Analysing the dynamic behaviour of hydro-diffusion system using step test. In *IEEE Conference on Systems, Process and Control*, 2015, pp. 16-21

[12] Yassin I M, Jailani R, Ali M, Baharom R, Hassan A, Rizman Z I. Comparison between cascade forward and multi-layer perceptron neural networks for NARX functional electrical stimulation (FES)-based muscle model. *International Journal on Advanced Science, Engineering and Information Technology*, 2017, 7(1):215-221

[13] Favoreel W, De Moor B, Van Overschee P. Subspace state space system identification for industrial processes. *Journal of Process Control*, 2000, 10(2):149-155

[14] Bai F, Liu Y, Sun K, Bhatt N, Del Rosso A, Farantatos E, Wang X. Input signals selection for measurement-based power system ARX dynamic model response estimation. In *IEEE PES T&D Conference and Exposition*, 2014, pp. 1-7

[15] Sarwar R, Cho H, Cox S J, Mago P J, Luck R. Field validation study of a time and temperature indexed autoregressive with exogenous (ARX) model for building thermal load prediction. *Energy*, 2017, 119:483-496

[16] Barraza-Barraza D, Tercero-Gómez V G, Beruvides M G, Limón-Robles J. An adaptive ARX model to estimate the RUL of aluminum plates based on its crack growth. *Mechanical Systems and Signal Processing*, 2017, 82:519-536

[17] Yusoff Z M, Muhammad Z, Nordin M N, Rahiman M H, Taib M N. Real time PID control for hydro-diffusion steam distillation essential oil extraction system using Gradient

Descent tuning method. In IEEE Control and System Graduate Research Colloquium, 2012 IEEE, pp. 288-293

[18] Ismail N, Tajjudin N, Rahiman M H, Taib M N. Modeling of dynamic response of essential oil extraction process. In 5th IEEE International Colloquium on Signal Processing and Its Applications, 2009, pp. 298-301

[19] Hambali N, Rahiman M H, Taib M N, Yassin A I. A review on nonlinear modeling for steam distillation system and its significance. In IEEE 12th International Colloquium on Signal Processing and Its Applications, 2016, pp. 328-333

[20] Kim D, Cai J, Ariyur K B, Braun J E. System identification for building thermal systems under the presence of unmeasured disturbances in closed loop operation: Lumped disturbance modeling approach. *Building and Environment*, 2016, 107:169-180

[21] Brunot M, Janot A, Carrillo F, Gautier M. A separable prediction error method for robot identification. *IFAC-PapersOnLine*, 2016, 49(21):487-492

[22] Yusoff Z M, Muhammad Z, Rahiman M H, Taib M N. PRBS perturbation and modeling of hydro-diffusion steam distillation essential oil extraction system under feed-in disturbance. In IEEE 8th International Colloquium on Signal Processing and its Applications, 2012, pp. 480-484

[23] Xu Y, Miao Z, Fan L. Deriving ARX models for synchronous generators. In IEEE North American Power Symposium, 2016, pp. 1-6

[24] Zabidi A, Tahir N M, Yassin I M, Rizman Z I. The performance of binary artificial bee colony (BABC) in structure selection of polynomial NARX and NARMAX models. *International Journal on Advanced Science, Engineering and Information Technology*, 2017, 7(2):373-379

How to cite this article:

ZM Yusoff, MAA Aziz, NF Razali, SA Nordin, AFZ Abidin, NM Zain. The study of dynamic response using arx model in extraction process. *J. Fundam. Appl. Sci.*, 2017, 9(4S), 145-159.