Model Identification and Validation for a Heating System using MATLAB System Identification Toolbox

Muhammad Junaid Rabbani¹, Kashan Hussain², Asim-ur-Rehman khan³, Abdullah Ali⁴
Electrical Engineering, National University of Computer & Emerging Sciences-FAST
Karachi, Pakistan
Email: junaid.rabbani@nu.edu.pk, kashan.husain@nu.edu.pk, asim.rehman@nu.edu.pk, abdullah.ali@nu.edu.pk

Abstract. This paper proposed a systematic approach to select a mathematical model for an industrial heating system by adopting system identification techniques with the aim of fulfilling the design requirement for the controller. The model identification process will begin by collecting real measurement data samples with the aid of MATLAB system identification toolbox. The criteria for selecting the model that could validate model output with actual data will based upon: parametric identification technique, picking the best model structure with low order among ARX, ARMAX and BJ, and then applying model estimation and validation tests. Simulated results have shown that the BJ model has been best in providing good estimation and validation based upon performance criteria such as: final prediction error, loss function, best percentage of model fit, and co-relation analysis of residual for output.

Index Terms: System Identification, ARX Model, ARMAX Model, BJ Model, Validation.

1. Introduction
The idea of this research work initiates from an essential requirement of designing a control system for an industrial heating system used earlier by G. Dullerud and R. Smith [1] and also presented by [2]. The heating system as shown in Fig. 1, is a simple Single Input Single Output System (SISO), the input voltage (U) drives a 300 Watt halogen lamp, suspended 2 inches above the metal plate. The output (Y) is a thermocouple measurement, attached underneath the plate to measure heating temperature. While, first step to design a controller required: mathematical model that describe physical properties of a heating system that need to be controlled [3]-[4]-[5]. This paper proposes system identification techniques to determine an efficient mathematical model for a heating system that could validate the real measurement data.

System identification is the class of control system engineering that determines physical characteristics of a plant and represent them in the form of mathematical expression by using real time measurement or experimental data [2]-[6]. After collecting an experimental data, most essential stages of model identification process can be summarized as follows, (1) select between parametric or non parametric modelling technique, (2) choosing the model structure, and finally (3) model...
estimation and validation criteria [7]. All these stages have a significant impact on model complexity, final prediction error, loss function, and percentage of model fitness.

System identification can be categorized based on their technique such as parametric model identification and non-parametric model identification. Non-parametric identification technique in comparison with parametric identification technique is relatively simple but less accurate. Whereas, parametric model identification provides complete model description that truly describes the physical dynamics of plant. The model structures for parametric identification are Auto-Regressive Exogenous (ARX), Auto-Regressive Moving Average Exogenous Model (ARMAX), Box-Jenkins (BJ) and State Space Model that are discussed in [8]-[9].

Furthermore, in system identification, model estimation of plant is carried out using measurement data. The objective of model estimation is to minimize the error function between measured and predicted output. Various model estimation approaches can be adopted in parametric estimation such as; prediction error method, least square, instrumental variable and maximum likelihood [9]. Whereas, model validation is vital to find the model credibility whether, model is capable to produce measured data or not. It can be achieved by examine best percentage of model fitting and correlation analysis of residual function between input and output [7].

Fig. 1. Heating system block diagram

2. METHODOLOGY
MATLAB System Identification tool box also presented by [10]-[11]-[12] has been used to determine mathematical model for a heating system, which provide ease and computer based simulation with multiple iteration facility to pick the best model that prove maximum model estimation and validation in comparison with real system. This paper used real measurement data, taken directly from the heating system for model identification discussed in [13]. The integral input/output samples data taken at desired sample interval of a heating system are as follow:

\[ U = \text{input voltage (Volt)} \]
\[ Y = \text{Temperature (deg. C)} \]
Sampling interval: 2 seconds
Total number of samples: 801

First load heating system measurement data such as input voltage, sampling time and output temperature in MATLAB and then initialize input signal as \((u)\) vector and output signal as \((y)\) vectors in MATLAB work space. Now open the system identification tool box (‘ident’) and import input and output data array by setting its starting point as zero and sampling interval to 2.0. Fig. 2 shows the time domain representation of observed data; both output temperature and input voltage is plotted with respect to sampling time period. It can be observed that the relation between input voltage and output temperature is linear. On the other hand, Fig. 2 represents time domain representation of both
functions, after subtracting the mean value from the measurement data to remove offset. The reason of subtracting the mean value from each signal is: to build a linear model that would describe deviated response from the physical equilibrium.

Finally the entire measurement data was divided in two parts, keeping the first one reserved for estimation and other for validation. Two-third data from (1 - 533 samples) is selected for estimation and one-third data from (534 - 801) has been reserved for validation. Fig. 3 shows the estimation and validation data set as red and quasi line colors respectively. The estimation data are used as an input, and these data sets are input to the model estimator to predict the model as discussed in [2], [3].

This paper compared ARX, ARMAX and BJ model structures to find the best model for a heating system by adopting systematic approach based upon these stages: (i) choosing parametric model identification technique, (ii) picking the best model structure among ARX, ARMAX and BJ model with low order, (iii) apply model estimation and validation techniques by evaluating model performance based upon these performance criteria: (1) minimizing FPE and LF, (2) choose model structure which provide highest percentage of model fit, (3) auto-correlation and cross-correlation analysis of residual for output should be inside the confident region.

Fig. 2. Heating system real measurement data before and after removing mean
3. MODELING OVERVIEW

A. Auto Regressive Model (ARX)

The ARX model is the simplest model describing input to output path as a linear difference equation. The model output is the sum of input and white noise, which enter directly into the system considered as a measurement noise. The ARX model provides maximum model estimation and validation result while model order id high [3]. The ARX model is defined as:

$$A(q)y(t) = B(q)u(t) + e(t)$$  \hspace{1cm} (1)

Where A and B polynomials are output and input respectively, and e(t) is the white noise.

$$A(q) = 1 + a_1q^{-1} + \ldots + a_nq^{-n_a}$$  \hspace{1cm} (2)

$$B(q) = b_1q^{-1} + \ldots + b_nq^{-n_b}$$  \hspace{1cm} (3)

$$y(t) + a_1y(t-1) + \ldots + a_ny(t-n_a) = b_1u(t-1) + \ldots + b_nu(t-n_b) + e(t)$$  \hspace{1cm} (4)

B. Auto Regressive and Moving Average Model (ARMAX)

The ARMAX model has an advantage over ARX model, because ARX model was not good in describing the property of dynamics disturbance. While ARMAX model includes flexibility to model disturbance separately by adding numerator polynomial $C(z^{-1})$. ARMAX model can be useful to handle disturbance, enter at the input [3]. This gives the model:

$$A(q)y(t) = B(q)u(t) + C(q)e(t)$$  \hspace{1cm} (5)

Where, A and B polynomials will remain same as in ARX model, C polynomials can be written as

$$C(q) = 1 + c_1q^{-1} + \ldots + c_nq^{-n_c}$$  \hspace{1cm} (6)
C. Box-Jenkins (BJ)

Box-Jenkins (BJ) model structure provides further development of output error, by filtering noise through $C(q)/D(q)$ transfer function. The BJ model is useful when noise enter late in the process [3], such as measurement noise.

$$y(t)= \frac{B(q)}{F(q)} u(t) + \frac{C(q)}{D(q)} e(t)$$

(7)

4. RESULTS

A. ARX Model Estimation and Validation Results

Firstly, to compute ARX model with suitable order and delay, Fig. 4 shows the best fits for ARX model as a function of: number of parameters used for various combination of $A(q)$ and $B(q)$ polynomials. Fig. 5 shows the percentage of best fit between measured and estimated model output for different $A(q)$ and $B(q)$ polynomials orders. The best estimation for the model is achieved; when the order of $A(q)$ and $B(q)$ polynomials is selected as eight and ten respectively that’s give 69.64% of the best fit. Furthermore, model validation test is done by investigating residual analysis. Figure 6, shows model credibility under the performance of residual analysis. Residual analysis has been done to quantify the error between the predicted output and measured output based on estimated model from the validation data set. Residual analysis consists of both whiteness and independence test to check whether the residuals are uncorrelated and independent from past inputs that prove good model validation. The quality of the estimated model can be judged if the autocorrelation function resides inside the confidence interval comprises of two straight lines. Figure 6, shows the auto-correlation and cross-correlation analysis of residual for ARX model at various orders; it can be observed that residual is inside the confident region (.18 to -.18) for all ARX model order combination. The estimated model equation of the ARX along with the $A(q)$ and $B(q)$ polynomials are as follow,

$$A(q). y(t) = B(q). u(t) + e(t)$$

(8)

$$A(q) = 1- 1.22 q^{-1} + 0.1457 q^{-2} + 0.0451 q^{-3} + 0.18819 q^{-4} + 0.1366 q^{-5} + 0.08447 q^{-6} + 0.0149 q^{-7} + 0.05336 q^{-8}$$

(9)

$$B(q) = 0.2454 q^{-2} + 0.2631 q^{-3} + 0.1645 q^{-4} - 0.06359 q^{-5} - 0.1046 q^{-6} - 0.1062 q^{-7} - 0.0464 q^{-8} - 0.1641 q^{-9} - 0.09716 q^{-10}$$

(10)

Fig. 4. ARX models best fit for its validation data
B. ARMAX Model Estimation and Validation Results

Similarly, ARMAX model has been designed to identify the best possible model. Unlike ARX model here, it is also important to find C(q) polynomial along with A(q) and B(q) polynomials to structure noise model as well. Fig. 7, shows best fit between measured and estimated output at different A(q), B(q) and C(q) polynomials orders. Best result is achieved, when order is selected 2, 2, 2 for A(q), B(q) and C(q) polynomials respectively, it provides best possible model fit up to 75% in comparison with other ARMAX models at low order. Figure 8, illustrates the auto-correlation and cross-correlation analysis of residual for ARMAX model at various orders; it can be observed that residual is within the confident range (0.2 to -0.2) for those ARMAX models that fit up to 75%, ensuring that residuals are not correlated and independent from past input for the desired polynomial order. While residual lie outside the confident range for ARMAX model estimating up to 65% model best fit. The estimated mathematical model of heating system using ARMAX model can be written as;
\[ A(q)y(t) = B(q)u(t) + C(q)e(t) \]  

(11)

Where, \( A, B \) and \( C \) polynomials are as follow,

\[ A(q) = 1 - 1.913q^{-1} + 0.9141q^{-2} \]  

(12)

\[ B(q) = 0.8145q^{-3} + 0.7895q^{-4} \]  

(13)

\[ C(q) = 1 - 0.7041q^{-1} + 0.009244q^{-2} \]  

(14)

\[ y(t) - 1.913y(t-1) + 0.9y(t-2) = 0.81u(t-3) + 0.79u(t-4) + e(t) - 0.7e(t-1) + 0.0092e(t-2) \]  

(15)

Fig. 7. Best fit % between measured and estimated output at different ARMAX model order
C. BJ Model Estimation and Validation Results

Finally, BJ model has been designed for heating system, Fig. 9, illustrates best fit % between measured and estimated model output. It is noticeable here; BJ model provides maximum model best fit result among all model structures. The efficiency of model is 91.03 %, achieved after choosing B(q), C(q), F(q) and D(q) polynomials order up to 7, 5, 8, 9 respectively. Figure 9, illustrate the auto-correlation and cross-correlation analysis of residual for BJ model structure; it can be observed that again residual is inside the confident region (0.2 to -0.2), with tolerance range ± 1%. It further passes both witness and independence test by indicating; residuals are not correlated and independent from past input that proves the model’s capability of validating actual data.

Fig. 8. Co-relation analysis of ARMAX model for model validation

Fig. 9. Best fit % between measured and estimated output of BJ model
Figure 11 and Table 1; summarize model estimation and validation results among ARX, ARMA and BJ model structures. It is required to choose best mathematical model for a heating system by analyzing and judging all performance parameters. It can be observed that ARX and ARMAX models provide best fit up to 70% and 75% respectively with low FPE and LF. Whereas, BJ model has a tendency to validate model fit up to 91% as well as provides low FPE and LF. It has also been noticed that all model structures passed validation test under residual analysis.

Figure 11. Best fit % between measured and estimated output of ARX, ARMAX and BJ models
5. CONCLUSION
This paper has focused the system identification techniques to select best mathematical model that would be suitable to translate physical behavior of a heating system. Among ARX ARMAX and BJ modeling structure, it seems reasonable to pick BJ model as a final choice; because it provides maximum model estimation and validation up to 91 %, in comparison with ARX and ARMAX models that give 65% and 75% respectively. It also has been observed that all model structures have less FPE and LF as well, passing model validation test under residual analysis. On the other hand, BJ model structure increases model complexity due to increase in system order in contrast with ARMAX model. So, there is a tradeoff between system order and best percentage of model fit.

6. REFERENCES
[1] G. Dullerud and R. Smith, "The experimental validation of robust control models for a heat experiment: A linear matrix inequality approach," in Decision and Control, 1995., Proceedings of the 34th IEEE Conference on, 1995, pp. 3486-3491.
[2] T. Mulyana, M. Than, M. Nor and N. A. Mustapha, "Identification of heat exchanger qad model BDT 921 based on hammerstein-wiener model," in International Seminar on the Application of Science & Mathematics 2011.
[3] L. Ljung, System identification - Theory for the User, Prentice-Hall, USA, 1999.
[4] T. Söderström and P. Stoica, System Identification. Prentice-Hall, Inc., 1988.
[5] R. Pintelon and J. Schoukens, System Identification: A Frequency Domain Approach. Wiley-IEEE Press, 2004.
[6] M. Gontikaki and A. van Schijndel, "Application of system identification methods to implement Comsol models into external simulation environments," 2009.
[7] F. Tjärnström, Variance Expressions and Model Reduction in System Identification. Department of Electrical Engineering, Linköpings universitet, 2002.
[8] W. Lai, M. Rahmat and N. A. Wahab, "MODELING AND CONTROLLER DESIGN OF PNEUMATIC ACTUATOR SYSTEM WITH CONTROL VALVE," in international journal on smart sensing and intelligent system, vol. 5, no. 3, pp. 624-644, 2012.
[9] P. Czop and G. Wszołek, "Advanced model structures applied to system identification of a servo-hydraulic test rig," Issues, vol. 1, pp. 2010.
[10] L. Ljung, "MATLAB System Identification Toolbox User's Guide. The Mathworks," Inc., Sherborn, Massachusetts, 1986.
[11] B. Ninness and A. Wills, "An identification toolbox for profiling novel techniques," in 14th IFAC Symposium on System Identification, 2006, pp. 301-307.
[12] A. Wills, A. Mills and B. Ninness, "A matlab software environment for system identification," in Proceedings of the 15th IFAC Symposium on System Identification, St. Malo, France, 2009.
[13] Lathauwer L. Database for the identification of systems: FECG data EAST/SISTA KU leuven, belgium. 2008-02-22).[2013-02-8].Available: http://ww.esat.kuleuven.ac.be/sista/daisy

Table I Model estimation and validation result

| Model           | LF     | FPE     | Best fit |
|-----------------|--------|---------|----------|
| ARX (8,9,2)     | 0.0538576 | 0.056889 | 69.64   |
| ARMAX(2,2,2,3)  | 0.0579158 | 0.059544 | 75.24   |
| BJ (75892)      | 0.054687 | 0.5243  | 91.03   |