COMPARISON OF MODELLING BETWEEN ARX AND ARMAX MODEL IN SYSTEM IDENTIFICATION

SITI NUR NADHIRAH BINTI NONCHIK

A report submitted
in fulfillment of the requirements for the degree of
Bachelor of Mechanical Engineering

Faculty of Mechanical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2018

SUPERVISOR'S DECLARATION

I hereby declare that I have read this project report and in my opinion this report is sufficient in terms of scope and quality for the award of the degree of Bachelor of Mechanical Engineering.

Signature : ……………………….

Supervisor's Name : ……………………….

Date : ……………………….

DECLARATION

I declare that this project report entitled "Comparison of Modelling between ARX and ARMAX Model in System Identification" is the result of my own work except as cite in the references.

Signature	: ……………………….
Name	: ……………………….
Date	: ……………………….

DEDICATION

To my beloved father, Nonchik Bin Adam
and mother, Fedzilah Bt Yusof

ABSTRACT

This research presents the comparison of modelling between ARX and ARMAX in System Identification. System Identification is a methodology to explain the dynamic behavior by building mathematical models using measurement of the system's input and output signals. The field of system identification is now widely used in most of the industrial projects in which the identification software have a wide circulation in industrial world. There are several type of general models in system identification that consist of AR model, ARX model, ARMAX model, Box-Jenkins model and Output-Error model but the main focus of this project are using ARX and ARMAX model. The aim of this research is able to simulate modelling using ARX model and ARMAX model and to compare the modelling performance of ARX and ARMAX model based on selected performance indicators. Specifically, the performance indicators that were used includes best fit value, final prediction error value and mean square error value. In a completion of the analysis, all simulation is conducted using the 'ident' graphical user interface in MATLAB R2015b and the least square method is utilized to estimate the parameters of the ARX and ARMAX models structure in this research. The results generally showed that ARX model structure is slightly better than ARMAX model structure in terms of model best fit, final prediction error and mean square error due to an additional input variable in the model. Thus, ARMAX could not provide better fit value caused by the random disturbance provided. However, by implementation the real data, the results showed that ARMAX model structure is better compared to ARX model. In conclusion, the better performance of both ARX and ARMAX model is still depending on the data distribution.

ABSTRAK

Kajian ini membentangkan perbandingan antara model ARX dan ARMAX menggunakan pengenalpastian sistem. Pengenalpastian sistem adalah kaedah untuk menerangkan tingkah laku dinamik dengan membina model matematik menggunakan system input dan output signal. Bidang pengenalpastian sistem kini digunakan secara meluas dalam kebanyakan projek industri di mana perisian pengenalan mempunyai peredaran yang meluas di dunia perindustrian. Terdapat beberapa jenis model dalam sistem pengenalan yang terdiri daripada model AR, model ARX, model ARMAX, Model Box-Jenkins dan model Output-Error tetapi tumpuan utama projek ini adalah menggunakan model ARX dan ARMAX. Tujuan penyelidikan ini adalah untuk mensimulasikan pemodelan menggunakan model ARX dan model ARMAX dan membandingkan prestasi pemodelan model ARX dan ARMAX berdasarkan penanda prestasi terpilih. Khususnya, penanda prestasi yang digunakan merangkumi nilai fit, FPE dan MSE. Untuk menyelesaikan analisis, semua simulasi dijalankan dengan menggunakan 'ident' GUI di MATLAB R2015b dan kaedah kuasa dua terkecil digunakan untuk menganggarkan penanda prestasi struktur model ARX dan ARMAX dalam kajian ini. Hasil keputusan secara amnya menunjukkan bahawa struktur model ARX adalah lebih baik daripada struktur model ARMAX dari segi nilai fit, FPE dan MSE disebabkan oleh input tambahan dalam model. Oleh itu, ARMAX tidak dapat memberikan nilai yang lebih baik disebabkan oleh gangguan rawak yang disediakan. Bagaimanapun, dengan pelaksanaan data sebenar, hasilnya menunjukkan bahawa struktur model ARMAX lebih baik berbanding dengan model ARX. Kesimpulannya, prestasi ARX dan ARMAX yang lebih baik masih bergantung kepada pengagihan data.

ACKNOWLEDGEMENTS

All praises to Allah S.W.T for helping and blessing through the entire problem that I faced during this period of final year project and to complete this final report.

Firstly, I would like to thanks my parents who are always give me their full support and encourage me to complete this project. I also want to take this opportunity to express my deepest appreciation to my supervisor, Dr. Fahmi Bin Abd Samad for his supervision, encouragement, suggestion, advice and trust throughout the duration in completion this project.

Finally, I would like to extend my sincere gratitude to each of my friends and everyone that always motivate me for giving a good cooperation in comment, opinion and support me directly or indirectly to complete this project.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARMAX	-	Autoregressive-moving average with exogenous terms
ARX	-	Autoregressive with exogenous terms
DC	-	Direct Current
FPE	-	Final Prediction Error
GUI	-	Graphical User Interface
MATLAB	-	Matrix Laboratory
MSE	-	Mean Square Error
SISO	-	Single Input Single Output
PSM	-	Projek Sarjana Muda

LIST OF SYMBOLS

y	-	Output
u	-	Input
w	-	Measured Disturbance
$y(t)$	-	Current Output
$y(t-k)$	-	Finite number of past outputs
$u(t-k)$	-	Input at lag k
$e(t)$	-	Noise
n_a	-	Number of poles
$n_b - 1$	-	Number of Zeros
n_k	-	Dead-time
q^{-1}	-	Delay operator
b	-	The slope of the regression line.
a	-	The intercept point of the regression line and y axis
$\bar{X}$	-	x average
$\bar{Y}$	-	y average
$u2$	-	Input (heating power)
$y2$	-	Output (temperature of the outflow air)
R^2	-	R^2 Coefficient
θ, Z^N	-	Loss Function
d	-	Total number of estimated data
N	-	The length of data record
n_a	-	Number of past output terms used to predict the current output.
n_b	-	Number of past input terms used to predict the current output.
n_k	-	Delay from input to the output in terms of the number of samples.
n_c	-	Number of past values of the disturbance signal.
$\hat{Y}$	-	Vector of n predictions
Y	-	Vector of the observed values

CHAPTER 1

INTRODUCTION

1.1 Background

Mathematical model can take very different forms depending on the system under study, which may range from social, economic, environmental, mechanical to electrical system. Generally, the inner mechanism of economic, social or environmental systems are not widely known or recognize and often only small data sets are available, while previous understanding of mechanical and electrical systems is at high level, and experiments can easily carried out. Hence, system identification is one of the method that commonly used to develop a suitable mathematical model of a particular dynamic system (Mediliyegedara et al., 2004).

System identification is a methodology to explain the dynamic behavior by building mathematical models using measurement of the system's input and output signals (Saifizi, Ab Muin Sazali & Mohamad, 2013). In order to carry out prediction and simulation that require wide applications including biology, meteorology, mechanical engineering, economics, physiology and model-based control design, system identification tools can be used in resulting dynamic mathematical model (Singh & Ajith B, (2014). System identification is applied to many objects from huge systems involving gas turbine, reactors, airplanes and even geology of the earth to a small system which are servo DC motor and electromagnetic valves. It is mostly used in three areas that consist of modelling and simulation, control design and prediction.

1.2 Problem Statement

System identification is an area of control system where the range between theory and practical is not very well pronounced. The field of system identification is now widely used in most of the industrial projects in which the identification software have a wide circulation in industrial world. Furthermore, to build a model in industry, a carefully designed identification experiment is carried out. There are several type of general models in system identification that consist of AR model, ARX model, ARMAX model, Box-Jenkins model and Output-Error model. Hence, the focus of this project is to investigate the comparison and to clarify the difference between ARX and ARMAX model in order to find a suitable model for identification.

1.3 Objective

The objectives of this research are:

1. To simulate modelling using ARX and ARMAX model.
2. To compare the modelling performance of ARX and ARMAX model based on several selected performance indicators.

1.4 Scope

In order to achieve the objective, the scopes are prepared as shown below:

1. All simulation is conducted using the ‘ident’ graphical user interface in MATLAB.
2. MATLAB is also used to make data acquisition based on simulated data in the form of Single-Input-Single-Output (SISO) system.
3. The performance of modelling will be decided based on several indicators provided in Graphical User Interface (GUI).
4. The least square method is utilized to estimate the parameters of the ARX and ARMAX models in this research.

CHAPTER 2

LITERATURE REVIEW

2.1 System Identification

2.1.1 Introductions

System identification is a technique to develop a mathematical model of a specific dynamic system using the measurement of systems input and output signals. Both of the systems input and output can be seen as an interface between the actual application and the world of mathematical control theory and model abstraction by using a combination of observed data; 1) Basic mechanics and dynamics, 2) Prior knowledge of relationship between signals (Rivera, 2004). The models can be divided into three types that are a white box, a black box and a gray box but the main focus of this topic will be black box. The black box is a completely empirical description of the dynamics of a system for which essentially no information is known a priori.

2.1.2 Dynamic System

There are a few terms of a system that is accessible which ranged from loose description to severe mathematical formulations. One type of a system called *open systems* produces observable signals. It is commonly called *outputs* and are contemplated to be an object in which different variables connect at different types of time and space scales and it is influenced by external stimuli. *Input* can be operated by the observer and as for the *disturbances,* it can be categorized into those that is directly measured and that are only can be detected through its influence on the output (Ljung, 2012). The graphical model of a general open system that is acceptable for system identification and the dissimilarity among measured disturbances and inputs is frequently insignificant for the modeling process as can be seen in Figure 2.1.

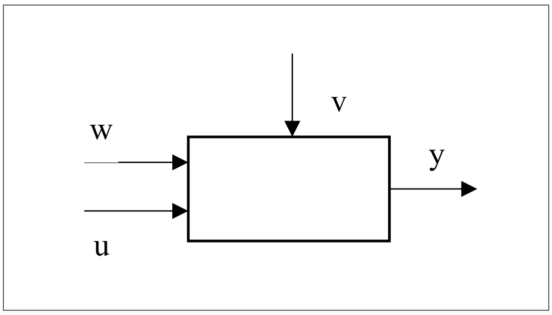


Figure 2.1 A system with output *y,* input *u*, measured disturbance *w*, and unmeasured disturbance *v* (Ljung, 2012)

As shown in Figure 2.2, the solar-heated house is considered as an example of a system. The system runs in a way that sun heats the air on the solar panel. The air then flows into heat storage which is a box filled with pebbles. The stored energy can later be transferred to the house. This system is represented in Figure 2.3 and the record of data obtained over fifty hour period and the variables sampled every ten minutes are shown in Figure 2.4 (Ljung, 2012).

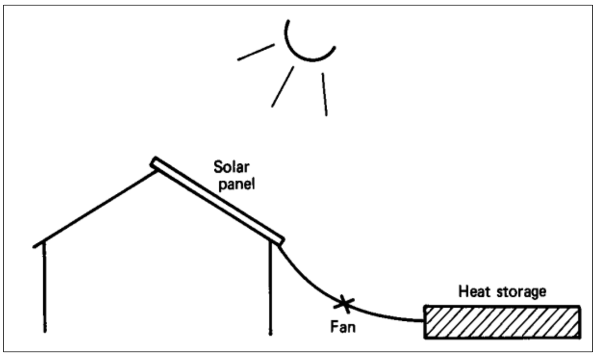


Figure 2.2 A solar-heated house.

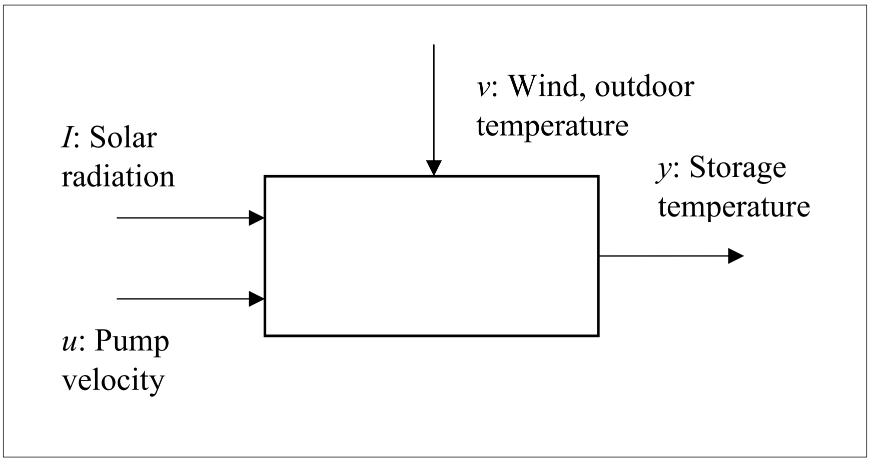
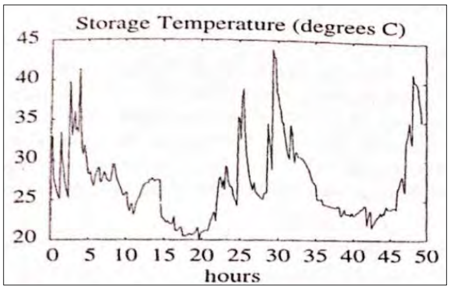
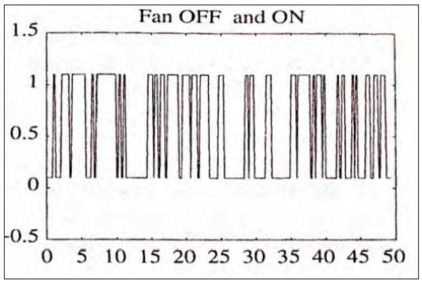


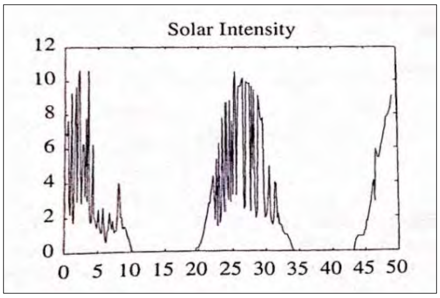
Figure 2.3 The solar-heated house system: *u*: input; *w*: measured disturbance; *y*: output; *v*: unmeasured disturbances (Ljung,2012).



(a) Storage temperature



(b) Pump Velocity



(c) Solar intensity

Figure 2.4 Storage temperature *y*, pump velocity *u*, and solar intensity *I* over a 50 hour period. Sampling Interval: 10 minutes.

2.1.3 Mathematical Models

Additive sensor noise term v(.) in Figure 2.5 denote errors produce from the measurement process. W(.) represents input disturbances while a white noise signal, it is usually presumed to represent v(.).

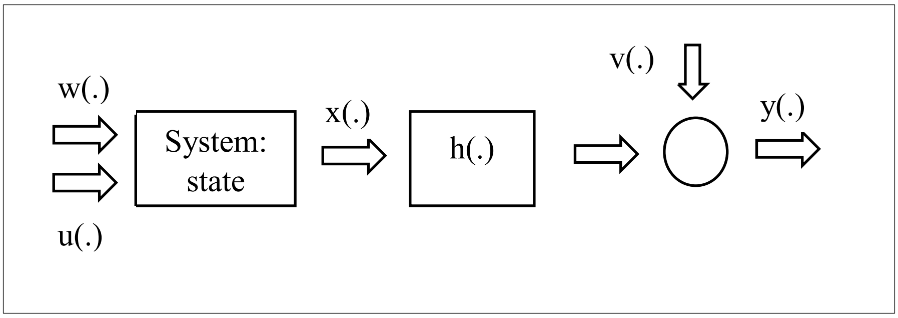


Figure 2.5 Basic structure of mathematical model (Keesman, 2011)

By considering the basic structure of system shown in Figure 2.5, set of standard differential equations with additive sensor noise shown in equation (2.1) and (2.2) as it represent standard description of a finite-dimensional system (Keesman, 2011).

Discrete-time:

$$x(t+1) = f(t, x(t), u(t), w(t); \vartheta), \qquad x(0) = x_0$$
$$y(t) = h(t, x(t), u(t); \vartheta) + v(t), \qquad t \in \mathbb{Z}^+ \qquad (2.1)$$

Continuous-time:

$$\frac{dx(t)}{dt} = f(t, x(t), u(t), w(t); \vartheta), \qquad x(0) = x_0$$
$$y(t) = h(t, x(t), u(t); \vartheta) + v(t), \qquad t \in \mathbb{R} \qquad (2.2)$$

Due to the availability of trial data and the ideal modification of a mathematical model into simulation code, the discrete-time form may be apply in system identification while the continuous-time will only be used for demonstration. According to Keesman (2011), these classification also tell the difference between linear and nonlinear, time-invariant and time-varying, static and dynamic systems.