INVESTIGATION ON EFEFCT OF INPUT ORDER SELECTION IN SYSTEM IDENTIFICATION

NURFA BINTI FAHAR

**This report is submitted
in fulfillment of the requirement for the degree of
Bachelor of Mechanical Engineering (Structure and Materials)**

Faculty of Mechanical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2016

DECLARATION

I declare that this project report entitled "Investigation On Effect Of Input Order Selection In System Identification" is the result of my own work except as cited in the references

Signature	:	....
Name	:	NURFA BINTI FAHAR
Date	:	30/6/2016

APPROVAL

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 (Structure & Materials).

Signature : ....

Name of Supervisor : MD FAHMI ABD SAMAD

Date : 29/6/16

DEDICATION

To my beloved mother and father

ABSTRACT

System identification techniques are discussed and applied in investigation on effect of input order selection. The performance of each identified input order was investigated to determine its effect and to identify the best model structure. The design of the experiments and the data processing will be discussed using MATLAB's system identification graphical user interface (GUI) named 'IDENT' found in System Identification Toolbox. Basically, system identification is the technique of developing models of dynamic systems from experimental input and output data. To perform system identification, one or more test input data sequences and the measured output data sequences are required for the system being modelled. Input order selection is an essential pre-processing stage in problems such as machine learning. Model to be used is autoregressive with exogenous input (ARX) model. In this study, 3 different equations were used throughout the investigation. Every equation was evaluated based on several indicators. The procedure of discovering a decent model can be isolated into four assignments which are analysis configuration, model type selection, parameter estimation and model validation. In conclusion, the effectiveness of all indicators help user to do input order selection in system identification. The indicators that were used throughout this investigation allows the user to identify the most accurate model and select a complexity which leads to a good reproduction of input properties of true system.

ABSTRAK

Teknik pengenalan sistem dibincangkan dan digunakan dalam siasatan mengenai kesan pemilihan perintah input. Prestasi setiap perintah input dikenal pasti itu disiasat untuk menentukan kesannya dan untuk mengenal pasti struktur model terbaik. Reka bentuk eksperimen dan pemprosesan data yang akan dibincangkan dengan menggunakan antara muka MATLAB ini pengenalan sistem pengguna grafik (GUI) yang dinamakan 'IDENT' dijumpai di dalam Sistem Pengenalan Toolbox. Pada dasarnya, pengenalan sistem ialah teknik membangunkan model sistem dinamik daripada data input dan output eksperimen. Untuk melaksanakan pengenalan sistem, satu atau lebih urutan data input ujian dan diukur urutan data output diperlukan untuk sistem yang dimodelkan. pemilihan perintah Input merupakan peringkat pra-pemprosesan penting dalam masalah seperti pembelajaran mesin. Model yang akan digunakan adalah autoregresif dengan model input luaran (ARX). Dalam kajian ini, 3 persamaan yang berbeza telah digunakan sepanjang siasatan. Setiap persamaan dinilai berdasarkan beberapa penunjuk. Prosedur menemui model yang baik boleh diasingkan kepada empat tugasan yang konfigurasi analisis, pemilihan jenis model, anggaran parameter dan pengesahan model. Kesimpulannya, keberkesanan semua penunjuk membantu pengguna untuk melakukan pemilihan untuk input dalam pengenalan sistem. Petunjuk yang digunakan sepanjang siasatan ini membolehkan pengguna untuk mengenal pasti model yang paling tepat dan memilih kerumitan yang membawa kepada pembiakan yang baik ciri-ciri input sistem benar

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Md Fahmi B. Abd Samad @ Mahmood for the continuous support of my Degree study, my research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my research.

Besides my supervisor, I would like to thank the rest of my seminar examiner, Dr Mohd Basri Bin Ali which also my personal advisor for my Degree study and Dr. Siti Hajar Binti Sheikh Md Fadzullah for their encouragement, insightful comments, and hard questions.

I thank my fellow friends, Nur Afifah Faizah Binti Mohd Rafie and Siti Nur Ariffah Alia for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we had in the last four years. Also I would like to thank my course mates for giving me their support, patience and encouragement.

Last but not the least, I would like to thank my family, my parents Mr. Fahar Bin Ruslan and Mrs. Sidah Binti Yusof for giving birth to me at the first place and supporting me spiritually and financially throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\hat{\theta}$	=	Estimated parameter vector
$\emptyset$	=	Matrix regressor
ϵ_M	=	Model set
σ^2	=	Constant variance
Ψ	=	Constraints
θ	=	Parameter vector
D_M	=	Open subset
Z	=	Vector output

CHAPTER 1

INTRODUCTION

1.1 Background Of Study

The zone of system identification is a standout amongst the most vital ranges in building on the grounds that a large portion of the dynamical system conduct can be acquired misusing system identification methods. In addition, system identification is a fundamental necessity in regions, for example, control, correspondence, power system and instrumentation for acquiring a model of a system of interest or another system to be produced. The accessibility and unwavering quality of the configuration methods of system identification did extend the application fields past the extent of mechanical applications. The identification problem requires a set of model structures, a validation criterion and an aim (Ljung, 1994). Basically, system identification is the technique of developing models of dynamic systems from experimental input and output data. To perform system identification, one or more test input data sequences and the measured output data sequences are required for the system being modeled.

This project aims at investigating the effect of input order selection in system identification. System identification is utilized to locate a reduced and precise mathematical model of a dynamic system. The design of the experiments and the data processing will be discussed using MATLAB's system identification graphical user interface (GUI) named 'IDENT' found in System Identification Toolbox. The study will analyze the effect of input order selection in system identification. In particular, a graphical user interface will be produced to connect the user with a few MATLAB capacities involves the system identification operations. The GUI may be used for building exact, disentangled models of complex systems from uproarious time-arrangement information. It provides tools for creating linear mathematical models of dynamic systems based on observed input/output data. The identification techniques provided with this toolbox are useful for applications ranging from control system design and signal processing to time-series analysis and vibration analysis.

Models describe the relationship between one or more measured input signals, $u(t)$, and one or more measured output signals, $y(t)$. The System Identification Toolbox software supports modelling input and output signals measured in either the time or frequency domain. For linear models, the general symbolic model description is given by Eq. (1.1) below:

$$y = Gu + He \tag{1.1}$$

where G is an operator that describes the system dynamics from the input to the output. G is often called a transfer function between u and y. H is an operator that describes the properties of the additive output disturbance and is called a disturbance model, or noise model. The actual disturbance contribution to the output, $H(e)$, has real significance and contains all the known and unknown influences on the measured y not included in the input u. Therefore, if the experiment was repeated with the same input, $H(e)$ explains why the output signal is different.

The procedure of discovering a decent model can be isolated into four assignments which are analysis configuration, model type selection, parameter estimation and model validation. In experiment design, the input signal(s) and the sampling interval that is used in the identification experiment should be decided (Godfrey, 1993, Ljung, 1999). Signal range and frequency content ought to be considered and in what working focuses the model will be utilized. Great analysis configuration is important to get educational estimation information that can be utilized for estimation of helpful models. Next, in model type selection, the function that is suitable to portray the connection between the regressors and the output should be resolved. There are several versatile model types available for both linear and nonlinear relations (Sjoberg et al., 1995). Nonlinear model sorts have a tendency to have an expansive number of parameters, notwithstanding for few regressors, because of the scourge of dimensionality. Countless make it important to have a lot of estimation information. The parameters connected with the picked model sort must be assessed. This is finished by minimizing some paradigm in view of the distinction in the middle of estimations and forecasts from the model. This is frequently the simplest undertaking to handle, however tedious. The evaluated model must be approved to verify that it is sufficient for its proposed use. Forecast and reenactment execution, model mistakes and soundness are critical to check. Prediction and simulation performance, model errors and stability are important to check. The input/output data used for estimation should not be reused for validation, but instead a new data set should be used (Ljung, 1999). The importance of the model validation cannot be overrated.

1.2 Problem Statement

Picking a model structure is typically the initial move towards its estimation. The request of the model should be determined before the relating parameters are estimated. The viability of the model-based control plan depends in light of the nature of the model that is utilized for control. It is imperative that an exact model that can portray the dynamic behavior of the system.

Consider the linear statistical model as Eq. (1.2):

$$z(n) = \sum_{j=1}^{p} \theta_j f(\varphi_j(n)) + e(n) \tag{1.2}$$

where z(n) is the observed system output, θ_j is an unknown system parameter, $\varphi_j(n)$ is a regressor, $e(n)$ is an independent Gaussian variable with zero-mean and constant variance σ^2, f is a nonlinear mapping of the regresors, and n is a sample index point. Let the regressors be described as a linear expansion of the observed system output, input and noise as Eq. (1.3):

$$\phi(n) = [1, z(n-1), \cdots, z(n-n_z), u(n), \cdots, u(n-n_u), e(n-1), \cdot, e(n-n_e)]^T \tag{1.3}$$

where u is the input. For the special case where f is a nonlinear mapping of polynomial form it may include a variety of nonlinear terms, such as terms raised to an integer power, products of current and past inputs, past outputs, or cross-terms. In addition, the nonlinear mapping, f can be described by a wide variety of nonlinear functions such as a sigmoid.

System identification is a field of study that uses observe input-output data from a system to build mathematical models for dynamical systems. These models are really approximations of genuine procedures. A portion of the models utilized as a part of system identification likewise consolidate aggravations that the system may incorporate. This makes it perfect to utilize system identification to demonstrate the dynamical conduct of genuine procedures. It is suggested that system identification methods used to get a numerical model for this project. This model could help in the outline of the control plan. The system identification procedures utilized will likewise comprehend the impacts that the input order has on the model. The cross relationship between inputs are utilized to gauge the impact that the input order has.

1.3 Objectives

There are two objectives of this project. Firstly, to perform system identification using linear difference equation model. Secondly, to investigate effect of input order on performance of identification.

1.4 Scope Of Project

There are three scope of this project. Firstly, the identification will perform using GUI Ident in MATLAB. Next, model to be used is autoregressive with exogenous input (ARX) model. Lastly, to undergo performance comparison which will be made from various perspectives such as time plot and data spectra.

CHAPTER 2

LITERATURE REVIEW

2.1 System Identification

System of identification is selection of a model for a process such as studied system or device under test using a limited number of measurements of the input and outputs which may be disturbed by noise and a priori system knowledge (Jan Swevers, 2006). There are four main steps involved in system identification and these are data acquisition, model structure selection, parameter estimation and model validation (Ljung, 1990). The data record or data acquisition is the determination of inputs and outputs. Inputs sign can be browse typical operation or particularly planned identification tests (Steen Toffner-Clausen, 2012). Moreover, input flag additionally can be look over persistency of excitation or ideal excitation which is maximally educational and negligible instability.

Model structure determination is performed by deciding a limited arrangement of models, ordinarily inside of a sure most extreme detail, and enumeratively testing the models for prescient exactness and miserliness. The choice of determination depends on certain data foundation where some settled rules are Akaike's information criterion, B-information criterion and Φ-information criterion (Veres, 1991). In the control system field, the issue of model request choice is available in association with system identification hones, which depend on the estimation of a scientific model for a dynamical system, beginning from exploratory input-output information (Kongens Lyngby, 2013). This issue is applicable for parametric system identification strategies, which utilize a limited dimensional parameter vector in the quest for the best portrayal of the system. Such methods require the decision of the model sort, of the model request and of the model parameterization. These decisions should be possible as indicated by:

i. a priori considerations, which are independent from the particular set of data used;

ii. a preliminary data analysis, which can help in the determination of the model order and also in the choice between the use of a linear or a non-linear model;

iii. a comparison among various model structures, which relies on the estimation of different types of models and on the comparison based on a pre-defined fit function;

iv. the validation of the estimated model, which uses the original estimation data to evaluate how well the model is able to describe their features, i.e. how much the data obtained from the estimated system agree with the estimation data.

The parameter estimation step decides inside of the arrangement of models, the model that is the best guess or gives the best clarification of the observed data. The estimation of the model parameters relates to the minimization of the chosen criterion. The decision of basis relies on upon the accessible data about and the motivation behind the model. Model validation is conceivably the most vital stride in the model building arrangement. It is likewise a standout amongst the most disregarded. Regularly the approval of a model appears to comprise of simply citing the R^2 measurement from the fit (which measures the part of the aggregate variability in the reaction that is represented by the model). Unfortunately, a high R^2 quality does not ensure that the model fits the information well. Utilization of a model that does not fit the data well can't give clever responses to the basic building or experimental inquiries under investigation (NIST, 2013). Figure 2.1 shows the system identification loop.

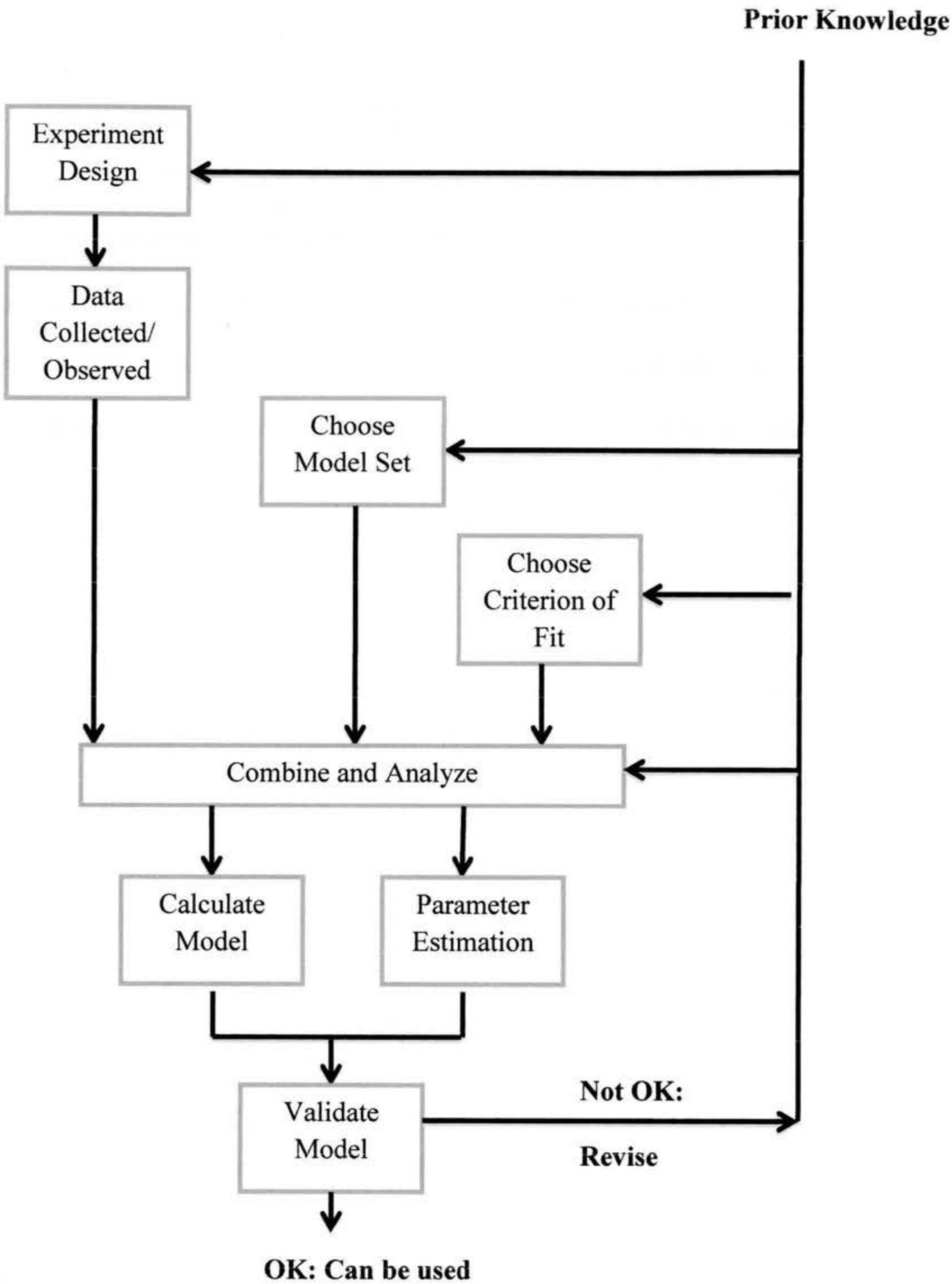

Figure 2.1: Flow Chart of Methodology

2.2 Mathematical Model

A model is a representation or a reflection of a system or a process. We manufacture models in light of the fact that they assist us with defining our issues, compose our contemplations, comprehend our information, impart and test that comprehension, and make forecasts. A model is consequently a scholarly device. Mathematical model is an arrangement of mathematical statements which can be utilized to process the time-space development of a physical system (Bellomo and Preziosi, 1995). Mathematical models are rearranged representations of some true element which can be in equations or computer code. Mathematical models are planned to copy vital components while forgetting inessentials. The characterization of mathematical models depends on the suspicions about variables (the things which change), parameters (the things we don't change), and functional forms (the relationship between the two). Model structure choice in system identification essentially includes the quest for an ideal model structure among numerous option models.

2.2.1 Classification of models

i. Linear vs. nonlinear

On the off chance that every one of the administrators in a mathematical model display linearity, the subsequent mathematical model is characterized as linear. A model is thought to be nonlinear something else. The meaning of linearity and nonlinearity is subject to setting, and linear models may have nonlinear expressions in them. For instance, in a measurable linear model, it is accepted that a relationship is linear in the parameters, yet it might be nonlinear in the indicator variables (Adam Kilicman, 2011). Additionally, a differential mathematical statement is said to be linear in the event that it can be composed with linear differential administrators, however it can in any case have nonlinear expressions in it. In a mathematical programming model, if the target capacities and requirements are spoken to totally by linear comparisons, then the model is viewed as a linear model. In the event that one or a greater amount of the target capacities or imperatives are spoken to with a nonlinear comparison, then the model is known as a nonlinear model.

ii. Static vs. dynamic

A dynamic model records for time-ward changes in the condition of the system, while a static (or enduring state) model ascertains the system in equilibrium, and in this way is time-invariant. Dynamic models regularly are spoken to by differential equations.

iii. Explicit vs. implicit

On the off chance that the greater part of the input parameters of the general model are known, and the output parameters can be figured by a limited arrangement of calculations, the model is said to be explicit. In any case, some of the time it is the output parameters which are known, and the relating inputs must be tackled for by an iterative technique, for example, Newton's strategy (if the model is linear) or Broyden's system (if non-linear). For instance, a plane motor's physical properties, for example, turbine and spout throat territories can be expressly ascertained given an outlined thermodynamic cycle (air and fuel stream rates, weights, and temperatures) at a particular flight condition and power setting, yet the motor's working cycles at other flight conditions and power settings can't be explicitly calculated from the consistent physical properties.

iv. Discrete vs. continuous

A discrete model regards objects as discrete, for example, the particles in an atomic model or the states in a factual model; while a constant model speaks to the articles in a ceaseless way, for example, the speed field of liquid in funnel streams, temperatures and anxieties in a strong, and electric field that applies consistently over the whole model because of a point charge.

v. Deterministic vs. probabilistic (stochastic)

A deterministic model is one in which each arrangement of variable states is remarkably dictated by parameters in the model and by sets of past conditions of these variables. Hence, a deterministic model dependably performs the same route for a given arrangement of starting conditions. On the other hand, in a stochastic model, as a rule called a "measurable model" where irregularity is available, and variable states are not depicted by novel qualities, yet rather by probability distribution.

2.2.2 Testing Techniques

2.2.2.1 White box testing technique

White box testing is a test case design method that uses the control structure of the procedural design to derive test cases. White box testing can uncover implementation errors such as poor key management by analyzing internal workings and structure of a piece of software. White box testing is applicable at integration, unit and system levels of the software testing process. In white box testing the tester needs to have a look inside the source code and find out which unit of code is behaving inappropriately (ME Khan, 2011). Some important types of white box testing techniques are briefly described below: