BALANCED STOCHASTIC REALIZATION ALGORITHM FOR DEVELOPMENT OF RAINFALL MODEL

by

FAHIMY BIN AZHARI

A dissertation submitted to Universiti Sains Malaysia for partial fulfillment of the requirement for the degree of Master of Science

MARCH 2014

ACKNOWLEDGEMENT

In the name of Allah, Most Gracious, Most Merciful

First, I thank the Almighty Allah for giving me inspiration, patience and good health which enabled me to complete this thesis. I owe the greatest gratitude to my supervisor, Dr. Rosmiwati binti Mohd Mokhtar for her patience, motivation, enthusiasm and immense knowledge. She provides me the opportunity to work on an exciting and interesting project. I learned so much from her during discussion we have had, not forgetting her constructive comments and warm encouragement that help me a lot during this research.

I also owe gratitude to my dearest family. Without my family's support, encouragement and understanding, I would not be the person I am today.

My appreciation goes to my lovely wife, Siti Khadijah binti Mohd Bakhori and my children, Muhammad Luqman bin Fahimy, for their continuous support, patience and love during this research.

Finally, I would like to thank all my friends and colleagues for their continuous support and guidance throughout this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANFIS	Adaptive Network Based Fuzzy Inference System
ANN	Artificial Neural Network
AWS	Automatic Weather Station
BF	Best Fit
BSR	Balanced Stochastic Realization
CCA	Canonical Correlation Analysis
FCA	Fuzzy Comprehensive Assessment
FSM	Fuzzy Similarity Methods
GP	Genetic Programming
LQ	Linear Quadratic
MMD	Malaysian Meteorological Department
MOESP	Multivariable Output Error State Space
MOSTI	Ministry of Science, Technology and Innovation
MSE	Mean Square Error
N4SID	Numerical Algorithm for Subspace State Space System Identification
NOAA AVHRR	National Oceanic & Atmospheric Administration Advanced Very High Resolution Radiometer
RAM	Random Access Memory
SFC	Simple Fuzzy Classification
SVD	Singular Value Decomposition
VAF	Variance Accounted For

ALGORITMA PENJELMAAN STOKASTIK TERIMBANG UNTUK PEMBANGUNAN MODEL HUJAN

ABSTRAK

Ramalan hujan merupakan komponen penting dalam amaran banjir kepada orang awam terutamanya di kawasan iklim tropika seperti Malaysia. Ramalan ini boleh digunakan untuk membuat keputusan sama ada amaran banjir patut dikeluarkan kepada orang awam terlebih dahulu. Di dalam penyelidikan ini, algoritma sub ruang Penjelmaan Stokastik Terimbang (BSR) digunakan untuk membangunkan model hujan bagi Malaysia. Analisa simulasi dilakukan menggunakan perisian MATLAB. Model dibangunkan menggunakan data hujan yang diperolehi daripada Jabatan Meteorologi Malaysia (MMD). Dalam kajian ini, data hujan tahun 2001 hingga 2005 dari stesen Kota Bharu telah digunakan untuk menerbitkan model. Model ini kemudiannya digunakan untuk meramal hujan bagi stesen yang sama untuk tahun 2006 hingga 2010. Keputusan menunjukkan prestasi dan ketepatan model yang baik. Untuk menilai prestasi model selanjutnya, model Kota Bharu digunakan untuk membuat ramalan pada lokasi yang berbeza dan corak taburan hujan yang berbeza di Malaysia. Terdapat lima klasifikasi lokasi utama iaitu kawasan pantai timur (Kuantan dan Kuala Terengganu), kawasan utara (Butterworth), kawasan barat (Ipoh), kawasan selatan (Senai) dan Malaysia timur (Miri dan Sandakan). Daripada analisa, model ini menunjukkan prestasi yang mantap apabila diuji di lokasi yang berbeza. Hasilan bermakna daripada kajian ini boleh digunakan bagi penyelidikan selanjutnya berkenaan ramalan banjir di Malaysia khususnya dan kawasan iklim tropika umumnya.

BALANCED STOCHASTIC REALIZATION ALGORITHM FOR DEVELOPMENT OF RAINFALL MODEL

ABSTRACT

Rainfall forecasting is an important component of flood warning to the public especially in tropical climate region like Malaysia. The forecasts can be used to make decisions about whether warnings of floods should be issued to the general public in advance. In this research, Balanced Stochastic Realization (BSR) subspace algorithm is used to develop a rainfall model for Malaysia. Simulation analysis is done using MATLAB software. The model is developed using a rainfall data obtained from Malaysia Meteorological Department (MMD). In this study, rainfall data in 2001 until 2005 from Kota Bharu station is used to simulate the model. The model is then used to predict rainfall of the same station in 2006 until 2010. The results reveal good model performance and accuracy. To further evaluate the model performance, Kota Bharu model is used to make forecasting of different places and different rain pattern in Malaysia. There are five main region classifications. Those are east coast area (Kuantan and Kuala Terengganu), northern area (Butterworth), western area (Ipoh), southern area (Senai) and east Malaysia (Miri and Sandakan). From analysis, the model also demonstrates high robustness when tested to different locations. The significant outcome from this study will able to assist further investigation on flood forecasting study in Malaysia, specifically, and tropical climatic region, generally.

CHAPTER 1

INTRODUCTION

1.1 Research Background

Flood is known the most frequent disaster that happened around the world (International Federation of Red Cross and Red Crescent Societies, 1998). Flood is a major disaster that could cause catastrophic damage to buildings and structures also including loss of life. Power transmission also can be damaged by flood that could affect the electricity power supply to the affected area. Moreover, water supply loss that caused by flood will result in loss of drinking water or severe water contamination. Lack of clean water combined with human sewage in the flood waters raises the risk of waterborne diseases, which can include cholera, typhoid, cryptosporidium and many other diseases. Damage to roads and transport infrastructure will result in difficulty to mobilize aid or to provide emergency health treatment to the affected. Flood waters typically overwhelmed farm land, making the land unworkable and preventing crops from being planted or harvested, which can lead to shortages of food both for humans and farm animals. Entire harvests for a country can be lost in extreme flood circumstances. Some tree species may not survive prolonged and severe flooding.

Predicting floods before happening allows for safety measure to be taken and people to be notified so that they can be prepared in advance for flooding conditions. For example, farmers can relocate animals from low areas and emergency services can plan ahead emergency action if needed. Emergency services can also make provisions to have enough resources available ahead of time to respond to emergencies as they occur. Flood risk assessment has widely been used in flood insurance, disaster warning, disaster evaluation, flood influence evaluation, and improving the public's flood risk awareness.

This research adopts subspace algorithm to develop a model of rainfall in Malaysia. There has been little study on subspace algorithm for weather forecasting purposes. In this study, a model is created and tested using rainfall data obtained by MMD. Simulation was done to model the prediction output and compared with existed data. Validation data were overlaid to validate and analyze the accuracy of subspace algorithm model. This research is focusing on subspace algorithm as a method of developing a time series forecasting model.

1.2 Objectives of Research

In this project, subspace algorithm is used to develop a rainfall forecasting model from rainfall data obtained from MMD. The objectives of this project are;

i) To develop a rainfall model by using a BSR algorithm.

ii) To test the performance accuracy of the algorithm and the simulation.

1.3 Scope of Research

The simulation work in this study uses MATLAB software Version 6.1 running on laptop Windows XP, 32-bit operating system having 1.73 GHz processor and 2 GB RAM (Random Access Memory). The availability of the rainfall data used in the study is limited to year 2001 until 2010. During evaluation analysis, only one or two station locations will be used to represent region classification made for Malaysia. Kota Bharu was chosen as subject in order to obtain the model. Subsequently the model is tested at other predetermined places as described in Section 3.2.4 in Chapter 3 to verify its performance and robustness.

1.4 Thesis Outline

The chapters of this thesis are organized as follows. Chapter 1 contains the introduction and motivation of conducting the research. This chapter also contains objectives and scope of project. Chapter 2 introduces what is basic idea of forecasting and flood forecasting. Existing methods for flood forecasting and brief introduction of subspace method will be discussed. Chapter 3 is devoted for explaining research methodology. This chapter deals with research framework and software selection. Subspace algorithm that will be used in developing rainfall forecasting model also will be detailed out in this chapter. Chapter 4 presents simulation test result. Rainfall data analysis using subspace algorithm will be discussed in this chapter. Finally, in Chapter 5, conclusion on the simulation and recommendation for future research are presented.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Flood is an overflow of an expanse of water that submerges land. Flood is one of common natural disaster happened in most area of the world. On top of that, it can be destructive and bring harm to human life. Millions of dollars and countless lives can be saved each year from accurate predictions of flood forecasting. Indeed, many people rely upon this information, from dam operators to public citizens, from farmers to sewage treatment plants. In this chapter a review on flood forecasting model will be discussed. It also followed by a review on subspace method which will be used as a method to perform the prediction.

2.2 What is Forecasting Model?

The process of making statements about events whose actual outcomes have not yet been observed is called forecasting. A ordinary example might be estimation of some variable of interest at some specified future date. It might refer to formal statistical methods employing time series, cross-sectional or longitudinal data, or alternatively to less formal judgmental methods.

Forecasting can be explained as predicting what the future will look like. Apparently, there is no single right forecasting method to use. Method selection should be considered on conditions and objectives. Forecasting has application in many situations. Economic forecasting such as oil price forecasting using genetic algorithm (Guo *et al*., 2012), environment forecasting such as earthquake magnitude prediction by using artificial neural network (Abdulrahman *et al*., 2012) and health forecasting such as predicting virus existence using adaptive network based fuzzy inference system (ANFIS) (Ucar and Karahoca, 2011).

Forecasting modeling is the process by which a model is chosen or created to try to best predict the probability of an outcome. In many cases the model is chosen on the basis to try to guess the probability of an outcome given a set amount of input or output data. One needs to understand how the data are compiled and how they relate to the forecasting issues. The data is examined to understand the structure and main features, and summarize the information available. Although a mathematical model is not a real system, it provides initial guess for making prediction. A model will help to verify overall performance and robustness of the final implementation.

2.2.1 Flood Forecasting

Floods are among the most devastating forces of nature especially in terms of human hardship and economical loss (see Figure 2.1). Flood appeared to be occurring more frequently due to climate change. The flood protection of properties is a highly important issue due to the damage, danger and other hazards associated to it to human life, properties, and environment.

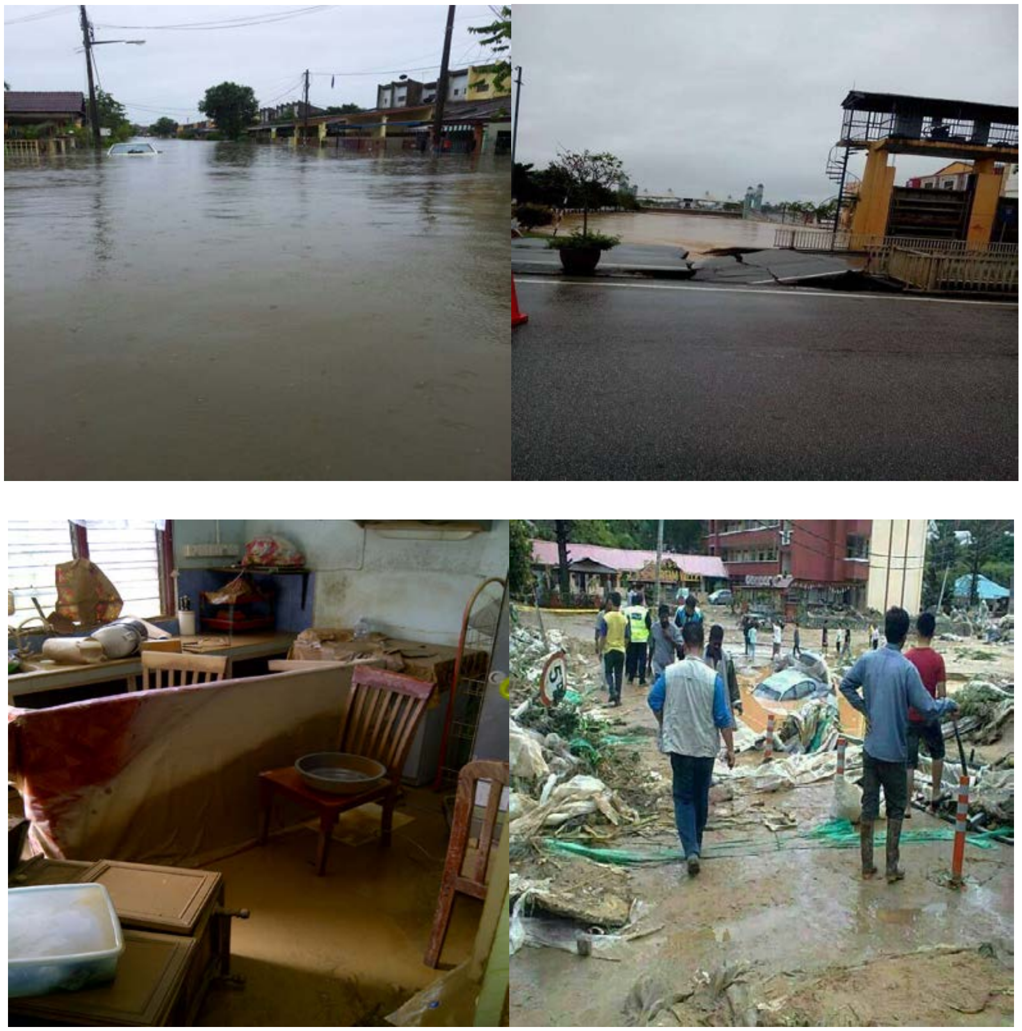

Figure 2.1: Floods often cause damage to homes and businesses

Flood forecasting is an important component of flood warning, where the distinction between the two is that the outcome of flood forecasting is a set of forecast time-profiles of channel flows or river levels at various locations, while "flood warning" is the task of making use of these forecasts to make decisions about whether warnings of floods should be issued to the general public or whether previous warnings should be retracted.

2.2.2 Existing Methods for Flood Forecasting

Ji *et al*. (2012) have demonstrated the flood forecast model based on Genetic Programming (GP) and Artificial Neural Network (ANN). Although it is difficult to simulate the forming of flood through rainfall, but practice has approved that applying data mining technique of flood forecasting of Double Reservoir is achievable.

On the other hand, fuzzy mathematics also can be used as assessment method in flood disaster validation (Jiang *et al*., 2009). The fuzzy comprehensive assessment (FCA), simple fuzzy classification (SFC) and the fuzzy similarity methods (FSM) are type of fuzzy mathematics that can be used to evaluate disaster risk. The percentage of validation accuracy in the studied area can be obtained depending on different zones; lowest risk, lower risk, medium, higher risk and highest risk zone.

Besides that, system identification method can be applied to estimate the parameters in the nonlinear time series flood model (Chen and Dyke, 2007). In this model, water flow dynamics is considered as a system. River flow and rainfall are considered as the system input and water surface level is considered as the system output. The models are purely data based and it is very flexible model structures and feasible approximation of the real model

2.2.3 Malaysian Weather Forecasting

In recent years, weather monitoring has become increasingly automated. An unmanned weather observation station is normally equipped with an automated weather system consisting suite of meteorological sensors housed in instrument shields. Then it is connected to a field processing unit or data logger and it is stored in data storage system. Automatic Weather System (AWS) as shown in Figure 2.2, used by MMD to measure the amount of rainfall precipitation and updating the data every minutes, 24 hours a day without human intervention. Besides that, AWS is capable of measuring atmospheric pressure, temperature, humidity, wind speed and direction and global solar radiation.

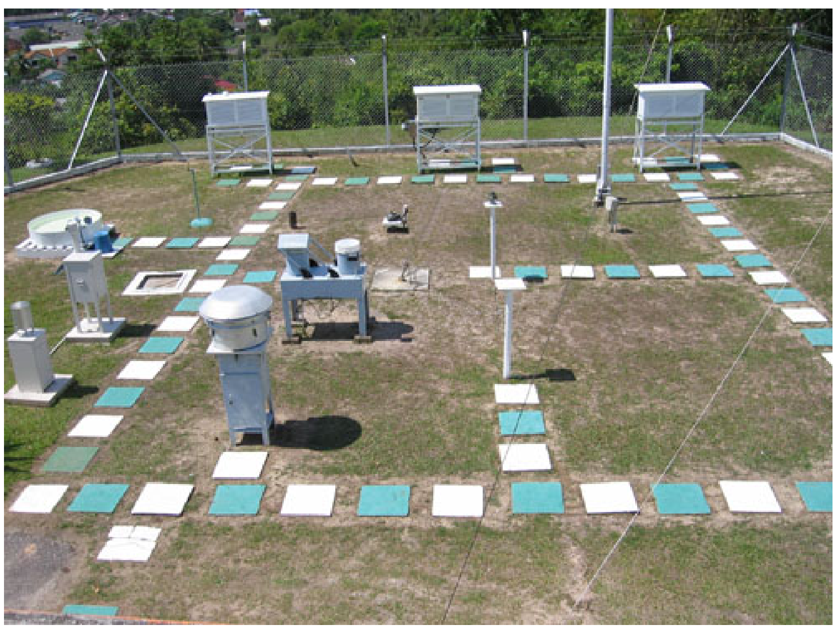

Figure 2.2: Automatic Weather System at Mersing Meteorological Station (Automatic Weather Stations, 2014)

MMD provides weather forecast in several different categories that consists of four main categories. Each category describes the weather in morning, afternoon and night for maximum seven days of forecasting. The first category is General Weather Forecast that comprises of three categories of weather forecast that is Peninsular Malaysia, Sabah and Sarawak. Consequently, the second category, State Weather forecast, enabled user to gain information of weather forecast of each state in Malaysia independently. On the other hand, third category that is District Weather Forecast give access of weather forecast of every district in each state in Malaysia. Finally, Major Towns / Tourist Destinations category provide information of weather forecast in major town and tourist destinations. Nevertheless, MMD weather forecast only mentioned the possibility of raining or not raining in their forecast and apparently not mentioning the amount of rainfall precipitation expected.

Besides that, there are studies conducted regarding weather forecasting in Malaysia. Tank model and weather surveillance radar data was used for flood forecasting at Sungai Gombak catchment (Tahir and Che-Hamid, 2013). The tank model uses hydrological and meteorological data such as rainfall, stream flow, and water level as an input to develop the model. This research shown that the development of the tank model for Sungai Gombak catchment was successful and was able to provide a reliable water level forecast at Jalan Tun Razak. With adaptive adjustment in this model, the average simulation errors for the average mean square error and root mean square error were reduced from 0.13 m and 0.384 m to 0.025 m and 0.065 m.

Meanwhile, Billa *et al*. (2006) have developed a rainfall intensity model to process National Oceanic & Atmospheric Administration Advanced Very High Resolution Radiometer (NOAA AVHRR) satellite data for rainfall intensity in an attempt to improve quantitative precipitation forecasting as input to operational hydro-meteorological flood early warning. The general extent of the study area covers the land and territorial water of Peninsular Malaysia. The thermal bands in the multispectral Advanced Very High Resolution Radiometer data were processed for brightness temperature. Data were further processed to determine cloud height and classification performed to delineate clouds in three broad classes of low, middle, and high. Rainfall intensity was assigned to the 1-D cloud model to determine the maximum rain rate as a function of maximum cloud height and minimum cloud model temperature at a threshold level of 235 K. This raises the opportunity of simulating rainfall runoff for the river catchment through the coupling of a suitable hydro-dynamic model to provide early warning prior to the actual rainfall event.

On the other hand, El-Shafie *et al*. (2011) have proposed ANFIS model to forecast the rainfall for Klang River in Malaysia on monthly basis. Statistical data from 1997 to 2008 was obtained from Klang gates dam data to train and test the ANFIS and ANN models. The optimum structure and optimum input pattern of model was determined through trial and error. Different combinations of rainfall were produced as inputs and the result indicate that the ANFIS model showed higher rainfall forecasting accuracy and lower root mean square error 0.052 compared to the ANN model that is 0.074.

2.3 Subspace Method

Subspace algorithms have been emerged in the last decades. It is another option to prediction error methods for the estimation of linear dynamical systems (Bauer, 2003). Numerical feasibility and simplicity and have been the main justification in favor of the method. The advantage of subspace method is it offers numerically reliable algorithms for computing state-space descriptions directly from data. Subspace algorithms have their origination in the algorithms of and Ho and Kalman (1966) and Zeiger and McEwen (1974). Subsequently, it has been developed further leading to the three most well known algorithms; Canonical Correlation Analysis (CCA), Numerical Algorithm for Subspace State Space System Identification (N4SID) and Multivariable Output Error State Space (MOESP).

.

CCA is a method of calculating the linear relationship between two multidimensional variables (Hotelling, 1936). It finds two bases, one for each variable, that are optimal with respect to correlations and, at the same time, it finds the corresponding correlations. In other words, it finds the two bases in which the correlation matrix between the variables is diagonal and the correlations on the diagonal are maximized. The dimensionality of these new bases is equal to or less than the smallest dimensionality of the two variables.

Recently, N4SID algorithm portrays a special attention in the system identification field (Hachicha *et al*., 2013). It is powerful alternative to the classical system identification method based on iterative methods. The main idea of this method is the oblique projection of subspaces generated by the block Hankel matrices formed by input/output data system. Others geometric and mathematics tools of linear algebra like Singular Value Decomposition (SVD) are used to extract the order of the system which contain the parameters of the estimated model.

The MOESP algorithm proposed by Verhaegen (1994), which require the knowledge of the Hankel matrices, is especially based on the orthogonal projection. From it, the matrices with specially structured row or columns are able to be revealed.

2.3.1 Applications using Subspace Method

There has been growing interest in mathematical models for all kinds of applications, such simulation, prediction, fault diagnosis, quality and safety monitoring, state estimation and signal processing. Subspace method has been utilized in various areas such fault detection, system monitoring, prediction and forecasting, optimization and quality control. Mathematical models such as subspace algorithms are highly useful in those situations in which, experimenting with the real system is too expensive or difficult.

Early detection of damage allows increased expectations of reliability, safety and reduction of the maintenance cost (Mevel, 1999). This research describe validation technique to monitor the health of a structure in operating conditions (e.g. rotating machinery, civil constructions subject to ambient excitations, etc.) and to detect slight deviations in a model derived from in-operation measured data. A statistical local approach based on covariance-driven stochastic subspace identification is proposed.

In many application of signal processing, especially in communication and biomedics, preprocessing is necessary to remove noise from data recorded by multiple sensors (Vorobyov and Cichocki, 2002). Subspace filtering is used as a noise reduction technique for electroencephalogram applications.

In this study, the subspace algorithm is adopted in order to develop a rainfall model, that is an important step to make rainfall forecast. Eventually it is possible to develop a flood forecasting after all this process are tested and validated. The algorithm is flexible and fast, as iterations are not required. This would be a great advantage to other methods in term of computational efficiency. Furthermore, current MMD rainfall forecasting only describe whether it is rain or not for maximum 7 days forecasting without able to forecast rainfall precipitation amount. Therefore, this would be a great novelty research to develop a rainfall model, thus would enable future work for rainfall precipitation forecasting to be done.

2.4 Summary

Literature review on flood forecasting and some application using subspace method have been discussed in this chapter. The importance of flood forecasting could help government and authorities to make plan in order to face flood disaster effectively.

In the next chapter, subspace algorithm to develop a rainfall forecasting model will be discussed.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this chapter, research methodology and work procedure will be described. Research framework is used to segregate jobs that will be done in this research. This chapter basically will describe regarding software selection, rainfall data, construction of rainfall data matrices, subspace algorithm implementation, simulation and performance analysis.

3.2 Research Framework

Figure 3.1 shows the flow of procedures that will be carried out in this study. It starts by choosing the MATLAB software for algorithm coding and program simulation. Next is rainfall data collection. Data available for this research come in excel format obtained from MMD. Then, data is arranged into matrices using Teoplitz matrix described in section 3.5. Subsequently, the data is simulated using BSR algorithm according to the requirement of the experimental. Eventually the results of simulation are validated and analyzed using validation methods that will be described in section 3.6 of this chapter.

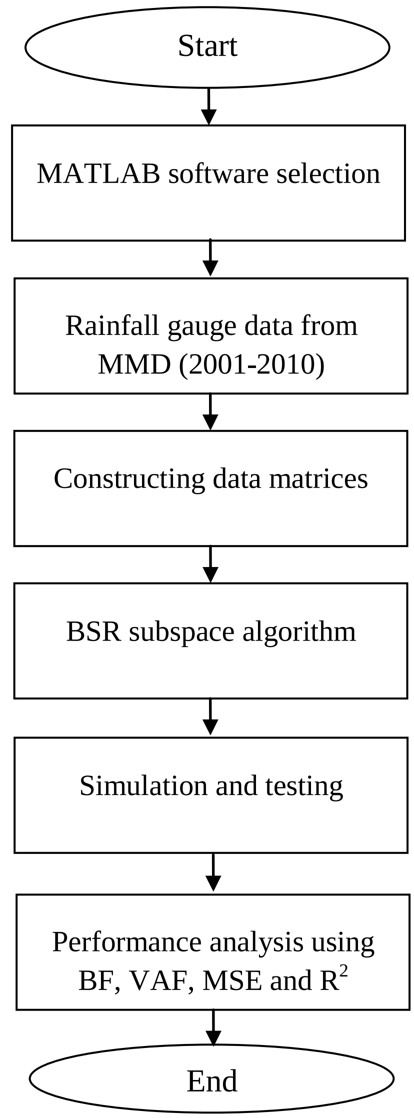

Figure 3.1: Project framework

3.3 Software Selection

MATLAB were chosen as the software to be used in this project. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces. MATLAB users come from various backgrounds of engineering, science, and economics. MATLAB is widely used in academic and research institutions as well as industrial enterprises.

3.4 Rainfall Data

Rainfall data can provide a quick overview of wet and dry periods of a certain place. Rain is measured using a rain gauge also known as a udometer or a pluviometer which gathers and measures the amount of liquid precipitation over a set period of time. Most rain gauges generally measure the precipitation in millimeters. Figure 3.2 shows example of rain gauge instrument.

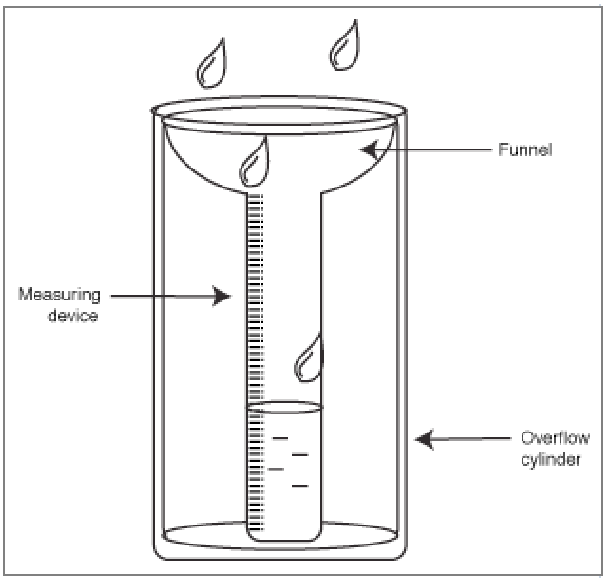

Figure 3.2: Rainfall gauge (Goldstein, 1999)

3.4.1 Malaysia Rainfall Data

For this research, Malaysia rainfall data is obtained from MMD that is under Ministry of Science, Technology and Innovation (MOSTI). Malaysia's climate is categorized as equatorial, being hot and humid throughout the year. The weather in Malaysia is characterized by two monsoon regimes, the Southwest Monsoon from late May to September, and the Northeast Monsoon from November to March. The Northeast Monsoon brings heavy rainfall, particularly to the east coast states of Peninsular Malaysia (Refer to Figure 3.3) and western Sarawak, whereas the Southwest Monsoon normally signifies relatively drier weather. The transition period in between the monsoons is known as the inter monsoon period (General Climate of Malaysia, 2013).

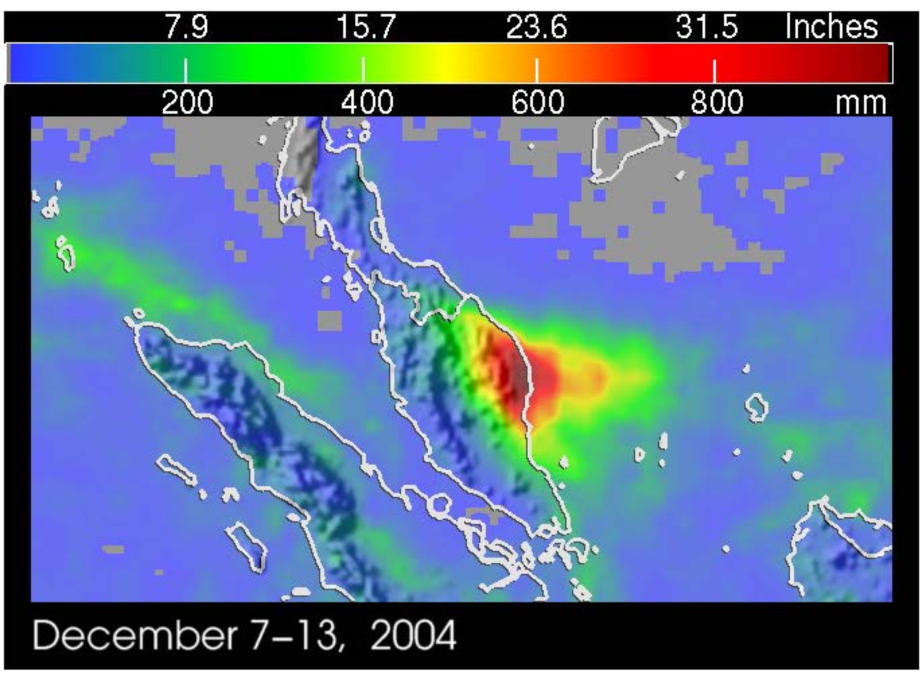

Figure 3.3: Peninsular Malaysia Precipitation Map on December 2004 showing heavy precipitation on the east coast, causing floods (Pierce, 2004)

3.4.2 Rainfall Data Analysis

Kota Bharu is chosen as subject of analysis of this algorithm. The key factor of choosing Kota Bharu in Kelantan is due to flood that are regularly happen nearly every year during the North East monsoon season. Severe flooding occurred in 1926 and 1967. In the 1967 floods, 84% of the Kelantan populations of 537,000 people were badly affected (Abu Hafizd, 2006). Data obtained are in monthly basis from January to December from 2001 until 2010. Rainfall data is recorded in millimeter for reference.

Basically the analysis is done in two categories. Firstly, is to obtain rainfall model based on data provided by Kota Bharu rain gauge station. Secondly, by using Kota Bharu as rain model, other location of different places rainfall will be forecasted. The reason of forecasting other places using Kota Bharu model is to identify the performance and robustness of the algorithm. Different places of location have different rain pattern and precipitation. It is divided into five prominent area of Malaysia as shown in Figure 3.4.

In East coast of Peninsular Malaysia, Kuala Terengganu and Kuantan as subject is chosen. The rain pattern is quite similar due to its neighboring location. Referring to rainfall data collected in 2005 as example, from Figure 3.5, it can be observed that rain pattern is similar in east cost, seeing high rain fall throughout October to December.

Figure 3.4: Location of test subject in Malaysia

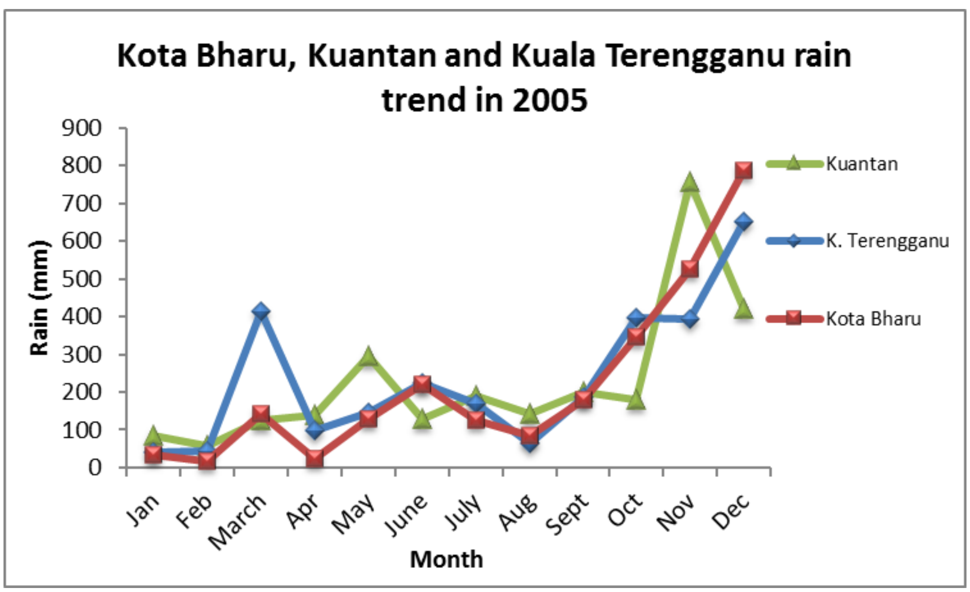

Figure 3.5: Kota Bharu, Kuantan and Kuala Terengganu rain trend in 2005

As shown in Figure 3.6, northern, western and southern area of Peninsular Malaysia seems to have average rainfall distribution throughout the year. Average of rainfall is not more than 300 mm per month. This location is different from Kota Bharu that has high rainfall during October until December exceeding other location average data. Western and southern part show same similarity of rain fall trend. However, the monthly rainfall pattern of northern area shows two periods of maximum rainfall separated by two periods of minimum rainfall. The primary maximum generally occurs in October until November while the secondary maximum generally occurs in April until May.

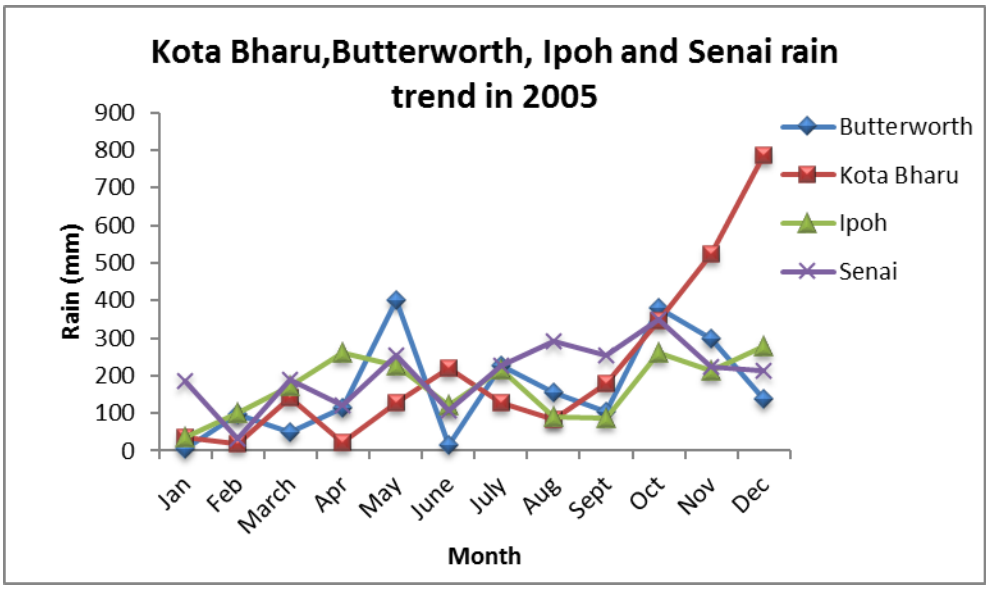

Figure 3.6: Kota Bharu, Butterworth, Ipoh and Senai rain trend in 2005

The coastal areas of Sarawak and northeast Sabah experience a maximum rainfall during the northeast monsoon months of December to March as shown in Figure 3.7.

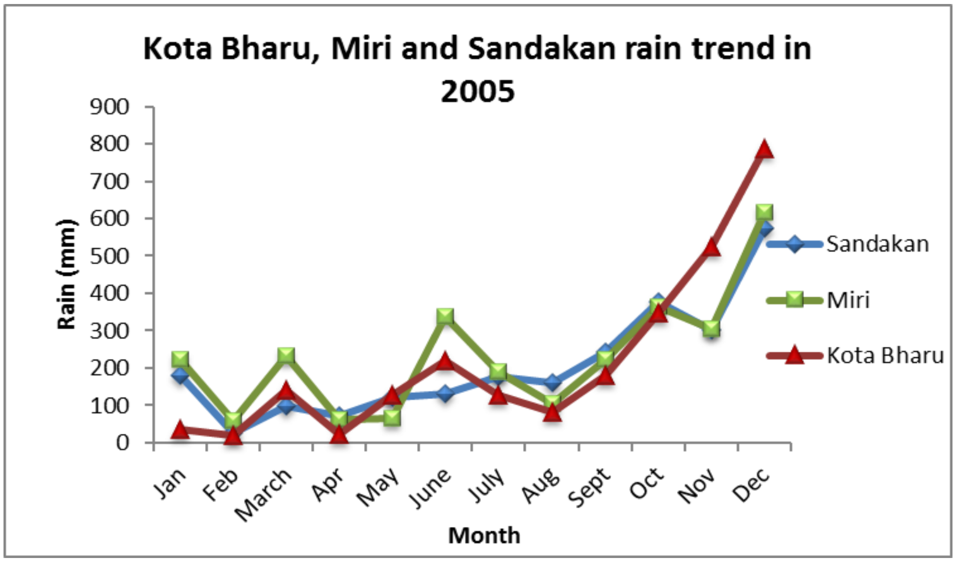

Figure 3.7: Kota Bharu, Miri and Sandakan rain trend in 2005.

3.4.3 Data Classification

The ten years data from Kota Bharu station is divided into two, that is, data from 2001 until 2005 is used for modeling (60 months) and data from 2006 until 2010 is used for testing (60 months). Data from other stations (representing northern, western, southern, east coast and east Malaysia) of year 2006 until 2010 (also 60 months for each) will be used for further evaluation of the algorithm.

3.5 Stochastic Subspace Algorithm

Stochastic subspace identification algorithm compute state space model from given output data only (Van-Overschee and De-Moor, 1996). A finite collection of data be given by $\{y(t),\ t = 0, 1, \ldots, N + 2k - 2\}$, where $k > 0$ and N is sufficiently large.

Assume that the given data is a finite sample from stationary process. The innovation state space model for y is given by

$$x(t+1) = Ax(t) + Ke(t) \tag{3.1a}$$

$$y(t) = Cx(t) + e(t) \tag{3.1b}$$

Where $e(t)$ is a white noise with mean zero, $x(t) \in R^n$ is the state vector, K is the Kalman filter gain and $A \in R^{n \times n}$ and $C \in R^{l \times n}$ are system matrices. n represents order of the system and l represents number of output. Figure 3.8 shows the block diagram of stochastic system. Kalman filter is a dynamic system that recursively produces the state estimates x(t+1) and x(t|t) by updating the old estimates based on the received output data y(t).

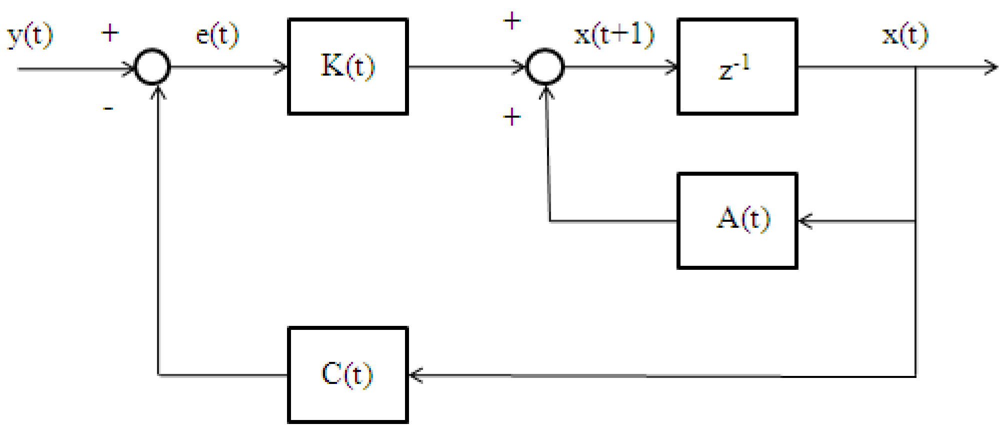

Figure 3.8: Block diagram of stochastic system

Rainfall data for modeling is organized in Toeplitz matrix or diagonal-constant matrix, named after Otto Toeplitz. It is a matrix in which each descending diagonal from left to right is constant. Define the Toeplitz matrix as (Katayama, 2005)

$$\breve{Y}_{0|k-1} = \begin{bmatrix} y(k-1) & y(k) & \cdots & y(N+k-2) \\ y(k-2) & y(k-1) & \cdots & y(N+k-3) \\ \vdots & \vdots & & \vdots \\ y(0) & y(1) & \cdots & y(N-1) \end{bmatrix} \in R^{kp \times N} \tag{3.2}$$

And the Hankel matrix as

$$Y_{k|2k-1} = \begin{bmatrix} y(k) & y(k+1) & \cdots & y(k+N-1) \\ y(k+1) & y(k+2) & \cdots & y(k+N) \\ \vdots & \vdots & & \vdots \\ y(2k-1) & y(2k) & \cdots & y(N+2k-2) \end{bmatrix} \in R^{kp \times N} \tag{3.3}$$

where $k > n$, and the number of column of block matrices is N. Let k be the present time and $Y_p = \breve{Y}_{0|k-1}$ and $Y_f = Y_{k|2k-1}$, respectively. The sample covariance matrices of the given data are defined by

$$\frac{1}{N} \begin{bmatrix} Y_p \\ Y_f \end{bmatrix} \begin{bmatrix} Y_p^T & Y_f^T \end{bmatrix} = \begin{bmatrix} \Sigma_{pp} & \Sigma_{pf} \\ \Sigma_{fp} & \Sigma_{ff} \end{bmatrix} \tag{3.4}$$

The linear quadratic (LQ) decomposition is constructed in the form of

$$\frac{1}{\sqrt{N}} \begin{bmatrix} Y_p \\ Y_f \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} \tag{3.5}$$

Then it follows that

$$\Sigma_{fp} = L_{21} L_{11}^T, \qquad \Sigma_{ff} = L_{21} L_{21}^T + L_{22} L_{22}^T, \qquad \Sigma_{pp} = L_{11} L_{11}^T$$