TIME SERIES MODELING USING MARKOV AND ARIMA MODELS

MOHD KHAIRUL IDLAN BIN MUHAMMAD

A report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Engineering (Civil – Hydraulic & Hydrology)

Faculty of Civil Engineering
Universiti Teknologi Malaysia

JANUARY 2012

DEDICATION

Special dedication to my beloved father and mother
Mr. Muhammad bin Ismail
and
Madam Siti Maznah binti Abdullah
and
My inspiration…

Jazakumullahu khairan for all love and inspiration
throughout the entire creation of this thesis.

ACKNOWLEDGEMENT

Assalammualaikum w.b.t.

Alhamdulillah, all praise to Allah S.W.T for the gift of life and what I have achieved today.

Appreciation goes to my family for their prayers, moral and financial support. May Allay reward you abundantly.

My sincere and deepest gratitude goes to my supervisor, Dr. Sobri Harun for his guidance, encouragement and support in completing this master project.

My gratitude to Dr. Muhammad Askari for his invaluable suggestions, guidance, and encouragement.

Last but not least, to all my lecturers, classmates and friends, their help and supports are really appreciated and will be remembers forever, InsyaALLAH. Thank you all

.

ABSTRACT

Streamflow forecasting plays important roles for flood mitigation and water resources allocation and management. Inaccurate forecasting will cause losses to water resources managers and users. The suitability of forecasting method depends on type and number of available data. Thus, the objective of this study are to propose the streamflow forecasting methods using Markov and ARIMA models and to inspect the accuracy of Markov and ARIMA models in forecasting ability. Streamflow data of Sungai Bernam, Selangor was used. Minitab and Microsoft Excel were used to model ARIMA and Markov respectively. Criteria performance evaluation procedure that being used in this study were Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Chi-square test of Normality to inspect the forecasting accuracy of the different models. The tentative model that best fits the criteria and meets the requirement for ARIMA model is ARIMA $(1,1,1)(0,1,1)^{12}$. From the criteria performance evaluation procedure, ARIMA model has better performance of model for forecasting than Markov model in this study. Therefore, ARIMA model has the ability to accurately predict the future monthly streamflow for Sungai Bernam.

ABSTRAK

Peramalan aliran sungai memainkan peranan yang penting untuk kawalan banjir dan pengurusan air. Peramalan yang tidak tepat akan menyebabkan kerugian kepada pihak pengurusan sumber air dan juga kepada pengguna. Kesesuaian kaedah peramalan bergantung kepada jenis dan jumlah data yang tersedia. Maka, objektif kajian ini adalah untuk mencadangkan kaedah peramalan aliran sungai dengan menggunakan model Markov dan ARIMA dan untuk memeriksa ketepatan model Markov dan ARIMA dalam membuat peramalan. Data aliran sungai Sungai Bernam telah digunakan. Minitab digunakan untuk memodelkan model ARIMA dan Microsoft Excel digunakan untuk memodelkan model Markov. Prosedur penilaian prestasi kriteria yang digunakan dalam kajian ini ialah Mean Absolute Percentage Error (MAPE), Root Mean Squared error (RMSE) dan ujian Chi-Squared untuk memeriksa ketepatan peramalan model-model yang berlainan. Tentatif model yang terbaik sesuai dengan kriteria dan memenuhi kehendak untuk model ARIMA ialah ARIMA $(1,1,1)(0,1,1)^{12}$. Dari prosedur penilaian prestasi kriteria, model ARIMA mempunyai prestasi yang lebih baik dalm membuat ramalan berbanding dengan model Markov. Justeru, model ARIMA mempunyai keupayaan untuk meramalkan dengan tepat aliran sungai di masa hadapan untuk Sungai Bernam.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ACF	-	Autocorrelation Function
AD	-	Anderson Darling
AR	-	Autoregressive
ARIMA	-	Autoregressive Integrated Moving Average
DF	-	Degree of Freedom
K-S	-	Kolmogorov-Smirnov
LSE	-	Least Squared Error
MA	-	Moving Average
MAPE	-	Mean Absolute Percentage Error
PACF	-	Partial Autocorrelation Function
RMSE	-	Root Mean Square Error
R^2	-	Coefficient of Determination
S	-	Standard Deviation
SE	-	Standard Error
Sg.	-	Sungai
χ^2	-	Chi-square

CHAPTER 1

INTRODUCTION

1.1 Background of Study

According to Bowerman and O'Connell (1993), predictions of future events and conditions are called forecasts, and the act of making such predictions is called forecasting. In many types of organizations, forecasting is very important as predictions of future events must be incorporated into the decision-making process. In forecasting events that will occur in the future, information concerning events that have occurred in the past must be relied.

In order to prepare forecasts, past data need to be analyzed to identify a pattern that can be used to describe it. Then, this pattern is extrapolated or extended into the future. This forecasting technique rests on the assumption that the pattern that has been identified will continue in the future to give good predictions. If the data pattern that has been identified does not persist in the future, this indicates that the forecasting technique used is likely to produce inaccurate predictions (Bowerman and O'Connell, 1993).

Most forecasting problems involve the use of time series data. In this study, time series is used to prepare forecasts. Time series is formed from measurements of a variable taken at regular intervals over time. It is a stochastic process which amounts to a sequence of random variables. The hydrologic data of streamflows fall under the category of time series (Gupta, 1989). Time series can be used in application of forecasting of future values of a time series from current and past values, and can be used to forecast streamflow (Box and Jenkins, 1976). Time series plots can reveal patterns such as random, trends, level shifts, periods or cycles, unusual observations, or a combination of patterns.

Streamflow forecasting plays important roles for flood mitigation and water resources allocation and management. In water management, the high quality streamflow forecast and efficient use of this forecast can give considerable economic and social benefits. Short-term forecasting like hourly and daily forecasting is crucial for flood warning and defense while long-term forecasting which is based on monthly, seasonal or annual time series is very useful for reservoir operation, irrigation management decision, drought mitigation and managing river treaties (Shalamu, 2009).

Recently, due to the increase in data availability from metering stations, real time data retrieval and increasing computational capability with the development of more robust methods and computer techniques, time series models have become quite popular in streamflow forecasting (Wang, 2006). A considerable number of forecasting models and methodologies have been developed and applied in streamflow forecasting due to importance of hydrologic forecasting. In this study, Markov and ARIMA model have been used in the modeling of monthly streamflow processes.

The Markov process considers that the value of streamflow at one time is correlated with the value of the streamflow at an earlier period (i.e. a serial or autocorrelation exists in the time series). In a first-order Markov process, this correlation exists in two successive values of the events (Gupta, 1989).

The first order Markov model states that the value of a variable x in one time period is dependent on the value of x in the preceding time period plus a random component. Thus, the synthetic streamflow represent a sequence of numbers, each of which consists of two parts, which are deterministic and random parts (Gupta, 1989).

Autoregressive Integrated Moving Average (ARIMA) which is often called method of Box-Jenkins time series has good accuracy for short-term forecasting, but less good accuracy for long-term forecasting. Usually, it will tend to become flat for a sufficiently long period. ARIMA model ignores the independent variable completely, and uses past and present values of dependent variable to produce accurate short-term forecasting (Hendranata, 2003).

ARIMA is suitable when the observation of time series is statistically related to the dependent. The purpose of this model is to determine good statistical relationships between the variables that being predicted and the historical value of these variables, so that forecasting can be performed with the model (Hendranata, 2003).

1.2 Problem Statement

There are many time series forecasting methods can be used to predict the streamflow. However, not all of these methods can produce accurate forecasts. Inaccurate forecasting will cause losses to water resources managers and users. The suitability of forecasting method depends on type and number of available data. ARIMA and Markov models must be inspected to determine the ability of this method to provide accurate and reasonable monthly streamflow forecasting. Through statistical methods, the accuracy of both models for forecasting monthly streamflow will be tested and evaluated. ARIMA modeling approach and Markov model was employed to the data set to further investigate the behavioral change in the streamflow. The result of the study can be used as a reference guideline to the flood control as Markov and ARIMA models best suited for short-term forecasting.

1.3 Justification of the Study

Monthly streamflow forecasting is an integral part of drought, irrigation and reservoir operation management. Stochastic data generation aims to provide alternative hydrologic data sequences that are likely to occur in future to assess the reliability of alternative systems designs and policies, and to understand the variability in future system performances. It is also very important to develop a stochastic hydrologic model to generate the monthly streamflows and thus to estimate the future streamflows. Through this model, it is wish that the problem on water shortage can be reduced. Forecasting also can be used to give warning of extreme events like drought (Joomizan, 2010).

1.4 Aim and Objectives

The aim of this paper is to forecast streamflow by using appropriate time series modeling approach. To achieve this aim, the following objectives have been identified:

1. To propose the streamflow forecasting methods using Markov and ARIMA models.

2. To inspect the accuracy of Markov and ARIMA models in forecasting ability.

1.5 Scope of Study

In this study, two models of time series are used which are Markov model and ARIMA model to predict the behavior of streamflow. Streamflow data of Sungai Bernam, Selangor for the period of 1960 to 2010 were used for the application of the model. The study area that located in southeast Perak and northeast Selangor is semi developed area and the size is 186km^2.

Streamflow data were obtained from station Sg. Bernam at Tanjung Malim (Station No. 3615412). The data which is monthly streamflow were collected from the Department of Irrigation and Drainage, Kuala Lumpur. Computer program that being used for ARIMA model is Minitab 15 and Microsoft Excel is used for Markov model.