GENERALIZED LINEAR MODEL APPROACH FOR MODELING RAINFALL OCCURRENCE

HII SHIUN LEH

UNIVERSITI TEKNOLOGI MALAYSIA

GENERALIZED LINEAR MODEL APPROACH FOR MODELING RAINFALL OCCURRENCE

HII SHIUN LEH

A dissertation submitted in partial fulfillment of the

requirements for the award of the degree of

Master of Science (Mathematics)

Faculty of Science

Universiti Teknologi Malaysia

JUNE, 2014

To my dearest family members.

ACKNOWLEDGEMENT

First and foremost I would like to thank the God for always guiding me in the right direction and fulfilling my desire to complete this educational journey smoothly.

I would also like to express my gratitude towards my dearest supervisor, Dr. Shariffah Suhaila Syed Jamaludin for supporting, advising and guiding me patiently throughout this study. I would not have completed this research without her supervision. I shall always be grateful to Dr. Shariffah for the guidance.

Besides that, I would like to thank my mother and my sister for the continuous supports given to me during my hard times conducting this study. The enormous mental support motivates me to proceed with my work. Thank you.

ABSTRACT

Rainfall modeling has been used to identify the characteristics of the rainfall occurrence and rainfall amount. The daily rainfall data was obtained from the Malaysian Drainage and Irrigation Department for six selected rain gauge stations over period of 32 years ranging from 1980 to 2011 in Peninsular Malaysia. The purpose of this study is to model the rainfall occurrence that varies as a function of time of year using Generalized Linear Model (GLM) and compare rainfall patterns between stations or regions based on the smoothing curves. The rainfall occurrence was fitted with a two-state first order Markov chain model in this study. The transition probabilities for first order Markov chain model were calculated using the maximum likelihood method. Since the transition probabilities for all stations are not stationary, smoothing curves to model the transition probabilities using Fourier series throughout the year were suggested. Analysis of deviance table was obtained by using statistical program *R* project. The result shows that all the stations display a bimodal pattern of rainfall with two distinct peaks except station Hospital Pontian for transition probabilities of dry day to rainy day and station Ipoh for transition probabilities of rainy day to rainy day. Based on the observations, rainfall modeling can be applied in agriculture sector.

ABSTRAK

Pemodelan hujan telah digunakan untuk mengenal pasti ciri-ciri kejadian hujan dan jumlah hujan. Data hujan harian telah diperolehi daripada Jabatan Pengairan dan Saliran Malaysia untuk enam tolok hujan yang terpilih dalam tempoh 32 tahun dari tahun 1980 hingga 2011 di Semenanjung Malaysia. Tujuan kajian ini adalah untuk memodelkan kejadian hujan menggunakan *Generalized Linear Model* (GLM) dan membandingkan corak hujan antara stesen atau wilayah berdasarkan lengkung licin. Model raintaian Markov peringkat pertama dengan dua keadaan telah digunakan untuk memodelkan kejadian hujan dalam kajian ini. Kebarangkalian peralihan untuk model raintaian Markov peringkat pertama telah dikira menggunakan kaedah maksimum kebolehjadian. Kebarangkalian peralihan untuk semua stesen tidak pegun, jadi lengkung licin untuk memodelkan kebarangkalian peralihan menggunakan siri *Fourier* sepanjang tahun telah dicadangkan. Jadual untuk analisis deviance telah diperolehi dengan menggunakan program statistik *R* projek. Hasilnya menunjukkan bahawa semua stesen memaparkan corak bimodal hujan dengan dua puncak yang berbeza kecuali stesen Hospital Pontian bagi kebarangkalian peralihan dari hari kering kepada hari hujan dan stesen Ipoh bagi kebarangkalian peralihan dari hari hujan kepada hari hujan. Berdasarkan pemerhatian, pemodelan hujan boleh digunakan dalam sektor pertanian.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CTMC	-	Continuous Time Markov Chain
CV	-	Coefficient of Variation
DTMC	-	Discrete Time Markov Chain
GCM	-	General Circulation Model
GLM	-	Generalized Linear Model
LSD	-	Log Series Distribution
MGTPD	-	Mixed Geometric with Truncated Poisson Distribution
MLE	-	Maximum Likelihood Estimation
MLPD	-	Mixed Log Series with Poisson Distribution
MLSD	-	Mixed Log Series Distribution
MLTPD	-	Mixed Log Series with Truncated Poisson Distribution
MMM	-	Modified Markov Model
NSRP	-	Neyman-Scott Rectangular Pulses

LIST OF SYMBOLS

μ	-	Mean
σ	-	Standard deviation
$kurt$	-	Kurtosis
$skew$	-	Skewness
x_i	-	Vector of explanatory variables
β	-	Vector of parameters
h	-	Link function
p_{00}	-	Transition probability from a dry day to a dry day
p_{01}	-	Transition probability from a dry day to a wet day
p_{10}	-	Transition probability from a wet day to a dry day
p_{11}	-	Transition probability from a wet day to a wet day
m	-	Number of harmonics
D	-	Deviance
H_0	-	Null hypothesis

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Water is one of our most precious natural resources. All living things, animals and plants depend on water. Rainfall is the main component of water and energy cycles. It is an important process for energy exchanges between the atmosphere, ocean and land. It determines the Earth's climate. Rainfall serves useful purposes for agriculture, industry and domestic consumption. In contrast, rainfall can cause problems. When there is no rainfall because of the dry weather, this will bring many effects to water supply, natural ecosystems and agriculture. In agriculture, drought can kill the crops and also increase erosion. Meanwhile, excessive rainfall can cause floods. This will cause a big loss to human, economic and environmental systems. Excessively wet weather can cause the growth of fungus and bacteria. Therefore, it is important to manage the rainfall appropriately so that the demand of rainfall can be maintained. Hence, it is vital to

know the characteristic of the rainfall occurrence and the rainfall pattern for the purpose of managing water resource systems.

According to National Research Council (2001), different rainfall trends indicate more volatile rainfall patterns and increase in average temperature. Due to climate change, Malaysia is undergoing higher rainfall throughout this century. Malaysia had experienced an increase in average temperatures and heavier rainfall compared to the previous years. In addition, from the Fifth Assessment Report by the Inter-Governmental Panel on Climate Change (IPCC), the rainfall intensity is increasing and Malaysia's average temperature is increasing by 3 to 5℃. Therefore, it is important to study the rainfall process by modeling the rainfall data. Various studies of modeling rainfall process had been conducted by researchers to give a better understanding of rainfall patterns and its characteristics.

According to Malaysian Meteorological Department, uniform temperature, high humidity and abundant rainfall are the characteristic features of the climate of Malaysia. Peninsular Malaysia receives average rainfall of 2500 millimeters. However, major climate change in Malaysia has caused the major flood and flash flood. Due to the high intensity and few days' duration rainfall, monsoon flood happens during the North-East Monsoon between November and March. On the other hand, high rainfall intensity will cause flash flood due to rapid development in the urban area. River overflows more rapidly than normal. Since flood can caused immense damages to the urban area and death if the flood is occurred seriously, therefore it is important to manage water resources systems.

1.2 Research Background

Rainfall modeling is a topic of study by many researchers from different specialized fields such as hydrology, meteorology, climatology and atmospheric physics. Rainfall modeling is important for planning, designing and managing water resource systems. It can also applied in agriculture and engineering. Rainfall modeling has been used to identify the characteristics of the rainfall occurrence and rainfall amount. Rainfall modeling generally describes the nature of rainfall and its time sequence mathematically.

The rainfall data often display a characteristic that is a substantial proportion of the values are zeroes. The zero values represent the dry days while the non-zero values represent the rainy days. Hence, two models are recommended for modeling rainfall data. One of the models is a model for dependent binary data to model rainfall occurrence on a particular day which is conditional on previous rainfall occurrence while another one is a model for rainfall amount to model rainfall intensity on rainy days for example gamma, exponential or mixed exponential distribution.

The rainfall occurrence is significant on modeling agricultural and hydrological impacts. The occurrence of rainfall at each station is considered as an independent random event. A model of rainfall occurrence is a model that provides a sequence of dry and wet days. Markov chain models can be used to fit the rainfall occurrence. Therefore, a study of Markov chain models is conducted to describe the occurrence of rainfall.

In general, Markov chains had been used to model the daily rainfall occurrence. A study on the sequence of daily rainfall occurrence had been done by Gabriel and Neumann (1962). The daily rainfall occurrence of rain at Tel Aviv, Israel was successfully analyzed by fitting with a probability model in the form of first-order

Markov chain. Markov chain models were used to describe the distribution of sequences of wet and dry days during the rainy season. After that, a seven state chain where every state corresponding to a different amounts of rain was analyzed by Haan *et al*. (1976). Moreover, Markov chains model of higher than first order was applied. Precipitation as a chain-dependent process which can reproduce the probability distribution of maximum daily amounts of precipitation was proposed by Katz (1977). Zero-, first- and second-order Markov chains were fitted to the daily rainfall data. Meanwhile, hourly rainfall observation was studied by Lund and Grantham (1977). A 12^{th} order Markov chains were fitted to the hourly rainfall data. Nguyen and Rousselle (1981) proposed that first- and second-order Markov chains were fitted to the consecutive hourly rainfall. According to Stern and Coe (1984), non-stationary second-order Markov chains are used to model daily rainfall occurrence at a single site.

There are many distribution models that can be used in the statistical studies of the distribution of rainfall, i.e., binomial, exponential, Gamma, Weibull and Kappa distributions. For example, daily rainfall occurrence during July and August from 1961 to 2007 in eastern China was analyzed using binomial generalized linear models. (Wang 2010). Woolhiser and Roldan (1982) used three-parameter mixed exponential distribution to describe the rainfall amounts on wet days. However, Todorovic (1975) proposed that exponential distribution does not fitted to the daily rainfall very well. The gamma distributions with parameters which may different with the time of year are usually fitted to the probability distribution of rainfall amounts on wet days. The two-parameter gamma distribution was used by Katz (1977), Buishand (1978) and Stern and Coe (1984) to describe the occurrence of rainfall amount on wet days.

In this study, Generalized Linear Models (GLMs) approach is chosen in the rainfall modeling to analyze the occurrence of the rainfall. GLMs approach for rainfall modeling has advantages for hydrological design and simulation studies. According to Stern and Coe (1982), the methods used for the analysis of GLMs are almost the same as the methods used for normally distributed data. The close terminology related in fitting

curves to the data then changes the prospect of analyzing non-stationary Markov chains. Moreover, a standard computer package can be used to fit the models. For example, the package GLIM (Baker and Nelder, 1978) and ASLGREG (Skyes, 1992) were used for testing the models for daily rainfall data. In addition to this, R package is provided to fit and simulate GLMs to daily rainfall data.

1.3 Problem Statement

Recently, climate change becomes a major issue in Malaysia. Rainfall are projected to experience greater fluctuations. Climate change may impacts agriculture, land resources, water resources and biodiversity in Malaysia. Climate change may affect water resources by changing the water levels and temperatures. Thus, food supply, health, industry, transportation and ecosystem integrity may be affected. One of the impacts to the agriculture is decreasing the productivity of agricultural. Therefore, it is important to find out the characteristics of rainfall occurrence and demonstrate the viability of approach for modeling the rainfall pattern.

Drought is defined as a significant dry period with no rainfall. In recent years, an extended period of dry weather had happened across Malaysia. One of the most severe droughts was 1997/98 El Nino events. These events brought severe drought resulting in water crises in many parts of Malaysia. Droughts have led to agricultural problems, water shortages and an increase in wildfires. Besides, climate change may also increase the rainfall intensity due to more water vapor could evaporate into the atmosphere which result in the intense storms. Floods are the most significant natural disasters in Malaysia due to the excessive rainfall. The floods occur due to the northeast monsoon and uncontrolled built-up areas with inadequate drainage in urban. Floods can cause crop damages, building structure damages and even cause death. Therefore, rainfall modeling

can be useful in water resources planning and management in order to solve these two disasters.

Markov chain models are often used to describe the daily rainfall occurrence. Also, generalized linear models (GLMs) are usually fitted using maximum likelihood method. GLMs were established in the statistical literature and first applied to the modeling of daily rainfall data by Stern and Coe (1984) since it is equivalent to a two-state Markov chain model for daily rainfall occurrence. Therefore, GLMs can handle variables with discrete variables such as rainfall occurrence as well as non-normal distributions such as rainfall amount on a wet day. GLMs were used in analyzing rainfall data because they can account for climate change over a long time scale. This indicates GLMs can be used to interpret historical rainfall data and changes in rainfall patterns. Besides that, a wide range of models can be formulated as generalized linear models and easily fitted using statistical program *R* project.

1.4 Research Objectives

- To profile the characteristics of rainfall occurrence in terms of their descriptive statistics.

- To model the rainfall occurrence that varies as a function of time of year using Generalized Linear Model (GLM).

- To compare rainfall patterns between stations or regions based on the smoothing curves.

1.5 Scope of the Study

In this study, daily rainfall occurrence was analyzed using binomial generalized linear models. Rainfall occurrence was analyzed by fitting a two-state, first-order Markov chain. Transition probabilities of the first-order Markov chain and the parameters of the binomial distribution were estimated by Maximum Likelihood Estimation (MLE). The parameters were estimated by using statistical program *R* project. In this study, the rainfall modeling was analyzed based on daily rainfall series. A complete daily rainfall data were obtained from different stations to achieve the objectives.

1.6 Significance of the Study

Rainfall is very important because all living things cannot survive without water. Rainfall is important for resources of water in Earth. Therefore, rainfall plays an important role for hydrological and agricultural studies. In agricultural, it is useful for planning irrigation systems using groundwater, rainfed agriculture and rice cultivation. In the study of hydrological process, analyzing rainfall data is also beneficial in rainfall-runoff and infiltration process. Sometimes, the result of simulation is not reliable due to the inadequate of analysis of the past rainfall data. Thus, the rainfall data of this study was analyzed to model rainfall occurrence accurately and to compare rainfall patterns between stations.

REFERENCES

Ababaei, B., Sohrabi, T. and Mirzaei, F. (2012). Simulation of Rainfall Occurrence in Qazvin Synoptic Station using Probability Models in Farsi. *Physical Geography Researchs*, 44(79), 91-110.

Aksoy, H. (2000). Use of Gamma Distribution in Hydrological Analysis. *Turkish Journal of Engineering Environmental Science*, 24, 419-428.

Audu, E. B. (2012). A Descriptive Analysis of Rainfall for Agricultural Planning in Lokoja Local Government Area of Kogi State, Nigeria. *International Journal of Science and Technologi*, 2(12).

Baker, R. J. and Nelder, J. A. (1978). The GLIM system, Release 3. *Numerical Algorithms Group, Oxford.*

Bardaie, M. Z. and Salam, A. C. (1981). A Stochastic Model of Daily Rainfall for Universiti Pertanian Malaysia, Serdang. *Pertanika*, 4(1), 1-9.

Buishand, T. A. (1978). Some Remarks on the Use of Daily Rainfall Models. *Journal of Hydrology*, 36, 295-308.

Cannarozzo, M, Noto, L. V. and Viola, F. (2006). Spatial Distribution of Rainfall Trends in Sicily (1921-2000). *Phys. Chem Earth*, 31, 1201-1211.

Cazacioc, L. and Cipu, E. C. (2005). Evaluation of the Transition Probabilities for Daily Precipitation Time Series Using a Markov Chain Model. *3rd International Colloqium – Mathematics in Engineering and Numerical Physics*, 12, 82-92.

Chapman, T. G. (1997). Stochastic Models for Daily Rainfall in the Western Pacific. *Mathematics and Computers in Simulation (MATCOM)*, 43(3), 351-358.

Chin, E. H. (1977). Modeling Daily Precipitation Occurrence Process with Markov Chain. *Water Resources Research*, 13, 949-956.

Coe, R. and Stern, R. D. (1982). Fitting Models to Daily Rainfall Data. *Journal of Applied Meteorology*, 21, 1024-1031.

Deni, S. M., Jemain, A. A. and Ibrahim, K. (2009). Fitting optimum order Markov Chain Models for Daily Rainfall Occurrences in Peninsular Malaysia. *Theoretical and Applied Climatology*, 97, 109-121.

Deni, S. M. and Jemain, A. A. (2009). Mixed Probability Models for Dry and Wet Spells. *Statistical Methodology*, 6(3), 290-303.

Dumont, A. G., Boyce, D. S. (1974). Probabilistic simulation of weather variables. *Journal of Agricultural Engineering Research*, 19(4), 131-145.

Dunn, P. K. (2004). Occurrence and Quantity of Precipitation can be Modelled Simultaneously. *International Journal of Climatology*, 24: 1231-1239.

Folwer, H. J., Kilsby, C. G., O'Connell, P. E. And Burton, A. (2000). A Stochastic Rainfall Model for the Assessment of Regional Water Resource Systems under Changed Climatic Conditions. *Hydrology and Earth System Sciences.* 4, 261-280.

Fu, G., Charles, S. P., Chiew, F. H. S. and Teng, J. (2011). Statistical Downscaling of Daily Rainfall for Hydrological Impact Assessment. *19th International Congress on Modeling and Simulation, Perth, Australia*, 12-16th December 2011, Perth, Western Australia.

Gabriel, K. R. and Neumann, J. (1962). A Markov Chain Model for Daily Rainfall Occurrences at Tel Aviv. *Quarterly Journal of Royal Meteorological Society*, 88, 90-95.

Green, J. R. (1964). A Model for Rainfall Occurrence. *Journal of the Royal Statistical Society*, 26(2), 345-353.

Haan, C. T., Allen, D. M. and Street, J. O. (1976). A Markov Chain Model of Daily Rainfall. *Water Resources Research*, Vol. 12(3). 443-449.

Hughes, J. P., Guttorp, P. and Charles, S. P. (1999). A Non-Homogeneous Hidden Markov Model for Precipitation. *Applied Statistics*, 48, 15-20.

Jayawardene, H. K. W. I., Sonnadara, D. U. J. and Jayewardene, D. R. (2005). Trends of Rainfall in Sri Lanka over the Last Century. *Sri Lankan Journal of Physics*, Vol. 6. 7-17.

Karl, T. R. and Knight, R. W. (1998). Secular Trends of Precipitation Amount, Frequency, and Intensity in the United States. *Bulletin of the American Meteorological Society*, 79(2), 231-241.

Katz, R. W. (1977). Precipitation as a Chain-dependent Process. *Journal of Applied Meteorology*, 16, 671-676.

Lana, X. and Burgueno, A. (1998). Daily Dry-Wet Behaviour in Catalonia (NE Spain) from the Viewpoint of Markov Chains. *International Journal of Climatology*, 18(7), 793-815.

Longobardi, A. and Villani, P. (2009). Trend Analysis of Annual and Seasonal Rainfall Time Series in the Mediterranean Area. *International Journal of Climatology*, Vol. 30. 1538-1546.

Lund, I. A. and Grantham, D. D. (1977). Persistence, Runs and Recurrence of Precipitation. *Journal of Applied Meteorology*, 16, 346-358.

Madsen, H. and Thyregod, P. (2011). Introduction to General and Generalized Linear Models. *Boca Raton, FL, USA: Chapman and Hall.*

Mangaraj, A. K., Sahoo, L. N. and Sukla, M. K. (2013). A Markov Chain Analysis of Daily Rainfall Occurrence at Western Orissa of India. *Journal of Reliability and Statistical Studies*, Vol. 6(1), 77-86.

Mehrotra, R. and Sharma, A. (2007). A Multisite Stochastic Downscaling Model for Daily Rainfall Occurrence with Long Term Persistence. *International Congress on Modeling and Simulation*, 10-13rd December 2007, University of Canterbury, Christchurch, New Zealand.

Mimikou, M. (1983). Forecasting Daily Precipitation Occurrence with Markov Chain of Seasonal Order. *International Symposium on Hydrometeology,* June 13-17, 1982, Denver, Colorado, J Am Water Resources Association, 219-224.

Modarres, R. and Sarhadi, A. (2009). Rainfall Trends Analysis of Iran in the Last Half of the Twentieth Century. *Journal of Geophysical Research: Atmospheres (1984-2012)*, 114(D3).

Nguyen, V-T-V. and Rousselle, J. (1981). A Stochastic Model for the Time Distribution of Hourly Rainfall Depth. *Water Resources Research,* 17(2), 399-409.

Nnaji, A. O. (2001). Forecasting Seasonal Rainfall for Agricultural Decision-making in Northern Nigeria. *Agricultural and Forest Meteorology* 107, 193-205.

Pant, B. C. and Shivhare, R. P. (1998). Markov Chain Model for Study of Wet or Dry Spells at AF Station, Sarsawa during Southwest Monsoon Season. *Vatavaran*, 22(1), 37-50.

Piyadasa, P. M. and Sonnadara, D. U. J. (2010). Analysis of Wet and Dry Behavior of Weather through Markov Models. *Proceedings of the Technical Sessions*, 26, 25-32.

Quintana, J. and Aceituno, P. (2006). Trends and Interdecadal Variability of Rainfall in Chile. *8 ICSHMO, Foz de Iguazu*, 371-372.

Rahman, M. S. (1999). A Rainfall Simulation Model for Agricultural Development in Bangladesh. *Discrete Dynamics in Nature and Society*, 5, 1-7.

Richardson, C. W. and Wright, D. A. (1984). WGEN: A Model for Generating Daily Weather Variables. *U. S. Department of Agriculture, Agricultural Research Serivce*, ARS-8, 88.

Robertson, A. W., Ines, A. V. M. and Hansen, J. W. (2007). Downscaling of Seasonal Precipitation for Crop Simulation. *Jounral of Applied Meteorology and Climatology.* 46, 677-693.

Rupakumar, K., Pant, G. B., Parthasarathy, B. and Sontakke, N. A. (1992). Spatial and Subseasonal Patterns of the Long-Term Trends of Indian Summer Monsoon Rainfall. *International Journal of Climatology*, 12(3), 257-268.

Shin, D. W., Baigorria, G. A., Lim, Y. K., Cocke, S., LaRow, T., O'Brien, J. J. and Jones, J. W. (2009). Assessing Crop Yield Simulations with Various Seasonal Climate Data. *18th Conference on Applied Climatology, 90th AMS meeting*, 17-21st January, 3.4, Atlanta, GA.

Skyes, A. M. (1992). Regression Volume 1 Version 1.1. *British APL Association*, 1991.

Srikanthan, R. and McMahon, T. A. (2001). Stochastic Generation of Annual, Monthly and Daily Climate Data: A Revies. *Hydrology and Earth System Sciences*, 5(4), 653-670.

Stern, R. D. and Coe, R. (1982). Fitting Models to Daily Rainfall Data. *Journal of Applied Meteorology*, 21, 1024-1031.

Stern, R. D. and Coe, R. (1984). A Model Fitting Analysis of Daily Rainfall Data. *Journal of Royal Statistical. Society Series A.* 147(1), 1-34.

Su, B. D., Jiang, T. and Jin, W. B. (2006). Recent Trends in Observed Temperature and Precipitation Extremes in the Yangtze River Basin, China. *Theoretical and Applied Climatology*, 83, 139-151.

Suhaila, J., Deni, S. M., Zawiah Zin, W. A. N. and Jemain, A. A. (2010). Trends in Peninsular Malaysia Rainfall Data During the Southwest Monsoon and Northeast Monsoon Seasons: 1975-2004. *Sains Malaysiana*, 39(4): 533-542.

Suhaila, J., Kong, C. Y., Fadhilah, Y. and Foo, H. M. (2011). Introducing the Mixed Distribution in Fitting Rainfall Data. *Open Journal of Modern Hydrology*, 1, 11-22.

Todorovic, P. (1975). A Stochastic Model of n-Day Precipitation. *Journal of Applied Meteorology.* 14, 17-24.

Wang, Y., Yan, Z. W. and Chandler, R. E. (2010). An Analysis of Mid-Summer Rainfall Occurrence in Eastern China and Its Relationship with Large-Scale Warming Using Generalized Linear Models. *International Journal of Climatology,* 30: 1826-1834.

Williams, C. B. (1952). Sequences of Wet and Dry Days Considered in Relation to Logarithmic Series. *Quarterly Journal of the Royal Meteorological Society,* 78, 91.

Woolhiser, D. A. and Roldan, J. (1982). Stochastic Daily Precipitation Models, 1. A Comparison of Occurrence Processes. *Water Resources Research*, 18(5), 1451-1459.

Yusuf, A. U., Adamu, L. and Abdullahi, M. (2014). Markov Chain Model and Its Application to Anuual Rainfall Distribution for Crop Production. *American Journal of Theoretical and Applied Statistics*, 3(2): 39-43.