

SPACE TIME RAINFALL MODELING USING HIDDEN MARKOV MODEL

TAN WEI LUN

A thesis submitted in fulfilment of the  
requirements for the award of the degree of  
Doctor of Philosophy (Mathematics)

Faculty of Science  
Universiti Teknologi Malaysia

MARCH 2017

To My Beloved Family

## ACKNOWLEDGEMENT

I would like to express my deepest appreciation and gratitude to my main supervisor, Assoc. Prof. Dr. Fadhilah Yusof, and to my co-supervisor, Prof Dr. Zulkifli Yusop, for their continuous support and encouragement throughout my studies. During my research process, they have given me extensive supervision, criticism, advice and guidance. Without their encouragement and enthusiasm, this thesis would not been completed.

Apart from both of my supervisors, I would also like to thank the Department of Irrigation and Drainage, Malaysia, for providing the valuable rainfall data that have been applied in my research work. My sincere appreciation is extended to the fund and sponsorship from the MyBrain15 Scholarship; their financial support is gratefully acknowledged.

In addition, I would like to express my deepest gratitude to Assoc. Prof. Dr. Liew Ju Neng of the School of Environmental and Natural Resource Sciences, UKM, for the valuable information. My gratitude also goes to all of my Professors, Senior Lecturers and in general all the staff of the Department of Mathematical Sciences, UTM, and Centre for Environmental Sustainability and Water Security (IPASA), UTM, for their important guidance and valuable remarks.

Moreover, my sincere appreciation also extends to my lovely family members for their unconditional support and encouragement, from the beginning of my research, until its successful completion. Last but not least, I would like to express my appreciation to all my friends and others who have provided me with their beneficial assistance during numerous occasions. For all their contribution, support, guidance and patience, only God can repay them.

## ABSTRACT

Statistical modeling of rainfall in space-time scales is essential in providing information on the behavior of the rainfall process at a particular region. One of the conventional ways of rainfall modeling is done through studies of the probabilistic structure of the rainfall. In the last decade, many statistical rainfall models were done regardless of the atmospheric information. A model which succeeded in incorporating atmospheric information will be useful in studies of climate variability or climate change. Therefore, this study applied the hidden Markov model (HMM) and non-homogeneous hidden Markov model (NHMM) to model daily rainfall of 40 stations in Peninsular Malaysia over a period of 34 years during the Northeast monsoon and Southwest monsoon. Four different models for rainfall amounts namely single exponential distribution, mixture of two exponential distributions, single gamma distribution and mixture of two gamma distributions were examined for the non-zero rainfall amount. The relationship between the local rainfall process and the large scale atmospheric process were investigated through the behavior of the composite wind anomalies at 850hPa with omega vertical velocity at 500hPa on each hidden state from HMM. The HMM was then extended to NHMM by including the time-varying covariates (atmospheric variables) into the model. So far, the most popular algorithm used for the parameter estimation of HMM was Baum-Welch algorithm, but it was only guaranteed to find a local maximum with a high dependency on initial parameters chosen. Hence, this study also proposed a parameter estimation, segmental *K*-means algorithm that sacrifices some of Baum-Welch's generality for computational efficiency. The findings here showed that the segmental *K*-means algorithm is able to improve the conventional model with a reduced computational time. The performances of the HMM and NHMM are assessed through the comparison between the observed rainfall data with the simulated rainfall data. For the rainfall occurrences, the HMM is considered as a very well fit for the tropical regions because it can capture fairly well the rainfall data in Peninsular Malaysia. It is found that the rainfall process in Peninsular Malaysia is associated to the atmospheric composites: low rainfall probabilities are characterized by a high pressure system and high rainfall probabilities are accompanied by a low pressure system. The HMM is able to reproduce the wet/dry spells for most of the stations but overestimated on the short duration of the wet/dry spell (one or two days wet/dry spell). For the rainfall amount, the NHMM has exploited all the mechanisms related to the atmospheric information and rainfall data, and able to reproduce and predict the interannual variability during the Northeast monsoon.

## ABSTRAK

Permodelan statistik hujan dalam skala reruang dan masa adalah penting dalam menyediakan maklumat tentang tingkah laku proses hujan di kawasan tertentu. Salah satu cara konvensional pemodelan hujan adalah dilakukan melalui kajian struktur kebarangkalian hujan. Dalam dekad yang lalu, kebanyakan model hujan statistik telah dibina tanpa mengambil kira maklumat atmosfera. Model yang berjaya menggabungkan maklumat atmosfera adalah berguna dalam kajian kebolehubahan iklim atau perubahan iklim. Oleh itu, kajian ini menggunakan Model Markov Terpendam (HMM) dan Model Markov Terpendam Tak Homogen (NHMM) untuk memodelkan hujan harian bagi 40 stesen di Semenanjung Malaysia dalam tempoh 34 tahun semasa Monsun Timur Laut dan Monsun Barat Daya. Empat model yang berlainan untuk amaun hujan iaitu taburan eksponen tunggal, taburan campuran dua eksponen, taburan gamma tunggal dan taburan campuran dua gamma telah dikaji untuk amaun hujan tak sifar. Hubungan antara proses penurunan hujan tempatan dan proses atmosfera berskala luas disiasat melalui kelakuan anomali angin komposit 850hPa dengan halaju menegak omega 500hPa di setiap keadaan terpendam dari HMM. HMM kemudiannya diperluas kepada NHMM dengan memasukkan kovariat masa berubah-ubah (pembolehubah atmosfera) dalam model tersebut. Setakat ini, algoritma yang paling popular digunakan untuk anggaran parameter HMM adalah algoritma Baum-Welch, tetapi ia dijamin hanya untuk mencari maksimum tempatan dengan pergantungan yang tinggi kepada parameter awal yang dipilih. Oleh itu, kajian ini juga mencadangkan suatu anggaran parameter, algoritma bersegmen *K-means* yang menghapuskan beberapa tatacara biasa Baum-Welch untuk kecekapan pengiraan. Dapatan kajian menunjukkan bahawa algoritma bersegmen *K-means* dapat meningkatkan kecekapan model konvensional dengan pengurangan masa pengiraan. Prestasi daripada HMM dan NHMM dinilai melalui perbandingan di antara data hujan cerapan dengan data hujan simulasi. Untuk kejadian hujan, model HMM dianggap sesuai untuk rantau tropika kerana ia dapat merakam data curah hujan dengan baik di Semenanjung Malaysia. Didapati proses penurunan hujan di Semenanjung Malaysia boleh dikaitkan dengan komposit atmosfera: kebarangkalian hujan yang rendah dicirikan oleh sistem tekanan tinggi dan kebarangkalian hujan yang tinggi disertai dengan sistem tekanan yang rendah. HMM dapat menghasilkan semula rentetan basah/kering untuk kebanyakan stesen tetapi terlebih anggaran pada rentetan basah/kering (satu atau dua hari rentetan basah/kering) yang singkat tempohnya. Bagi amaun hujan, NHMM telah menerokai semua mekanisme yang berkaitan dengan maklumat atmosfera dan data hujan, serta dapat menghasilkan semula dan meramal kebolehubahan antara tahunan semasa monsun timur laut.

## TABLE OF CONTENTS

<b>CHAPTER</b>	<b>TITLE</b>	<b>PAGE</b>
	<b>DECLARATION</b>	ii
	<b>DEDICATION</b>	iii
	<b>ACKNOWLEDGEMENT</b>	iv
	<b>ABSTRACT</b>	v
	<b>ABSTRAK</b>	vi
	<b>TABLE OF CONTENTS</b>	vii
	<b>LIST OF TABLES</b>	xi
	<b>LIST OF FIGURES</b>	xiv
	<b>LIST OF ABBREVIATIONS</b>	xxi
	<b>LIST OF SYMBOLS</b>	xxiii
	<b>LIST OF APPENDICES</b>	xxv
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 Introduction	1
	1.2 Background of the Study	2
	1.3 Statement of the Problem	3
	1.4 Objectives of the Study	4
	1.5 Scope of the Study	5
	1.6 Significance of the Study	6
	1.7 Structure of the Thesis	7
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>8</b>
	2.1 Introduction	8
	2.2 The Rainfall Occurrence Process	8

2.2.1	The Markov Chain Approach	9
2.2.2	Alternating Renewal Process	12
2.3	The Rainfall Amount Process	14
2.3.1	Multi-State Markov Model	15
2.3.2	The Weather State Model Approach	17
2.4	Hidden Markov Model	19
2.5	Non-Homogeneous Hidden Markov Model	21
2.6	El Niño-Southern Oscillation (ENSO)	25
2.6.1	ENSO indices	27
<b>3</b>	<b>METHODOLOGY</b>	<b>30</b>
3.1	Introduction	30
3.2	Flow Chart	30
3.3	Study Site Description	33
3.3.1	Rainfall Data	33
3.3.2	Atmospheric Data	35
3.4	Stochastic Process	35
3.5	Markov Chain	37
3.5.1	Discrete Time Markov Chain	37
3.5.2	First Order Markov Chain	38
3.6	Basic of the Hidden Markov Model	40
3.7	Hidden Markov Model	41
3.7.1	Parameterization for the Rainfall Process	42
3.7.1.1	Bernoulli Distribution	42
3.7.1.2	Gamma Distribution	42
3.7.1.3	Exponential Distribution	43
3.7.2	Parameterization for the Hidden Process	44
3.8	Non-Homogeneous Hidden Markov Model	44
3.8.1	Parameterization for the Hidden Process	45
3.9	The Atmospheric Variables	46
3.9.1	Singular Value Decomposition	46
3.9.2	Climatological anomalies	47
3.9.3	El Niño-Southern Oscillation	49

3.9.4	Student's $t$ Test	49
3.10	Parameter Estimation	50
3.10.1	Likelihood	50
3.10.2	The Expectation-Maximization (EM) Algorithm	53
3.10.3	The $K$ -Means Algorithm	56
3.11	Model Selection	57
3.12	Viterbi Algorithm	58
3.13	Scaling	59
3.14	Simulation and Prediction	61
3.15	Model Assessment	62
3.15.1	Pearson Correlation Coefficient	62
3.15.2	Survival Function	63
3.15.3	Quantile-Quantile Plots	67
3.15.4	Root Mean Squared Error (RMSE)	67
<b>4</b>	<b>HIDDEN MARKOV MODEL FOR RAINFALL OCCURRENCE</b>	<b>69</b>
4.1	Introduction	69
4.2	Descriptive Statistics of the Rainfall Data	69
4.3	Model Fitting	70
4.4	Estimated Parameters and Hidden States	73
4.5	Physical Interpretation on the Hidden States	83
4.6	Assessing the Performance of EM Algorithm and $K$ -means Algorithm	92
4.7	Summary	115
<b>5</b>	<b>HIDDEN MARKOV MODEL FOR RAINFALL AMOUNT</b>	<b>117</b>
5.1	Introduction	117
5.2	Description of Atmospheric Data	102
5.3	Fitting the Hidden Markov Model to the Rainfall Amount	120



5.4	Estimated Parameters and Hidden States	123
5.5	Assessing the Performance of HMM on Rainfall Amount	134
5.6	Non-homogeneous Hidden Markov Model	138
5.7	Synoptic Conditions	143
5.8	Inter-annual and Inter-decadal Variability	150
5.9	Comparison between the Simulations from HMM and NHMM.	157
5.10	Short-Term Prediction from HMM and NHMM	164
5.11	Summary	169
<b>6</b>	<b>CONCLUSION AND RECOMMENDATION</b>	<b>172</b>
6.1	Introduction	172
6.2	Conclusion	172
6.3	Recommendation for Future Research	174
	<b>REFERENCES</b>	<b>175</b>
	Appendices A - H	188-243

## LIST OF TABLES

TABLE NO.	TITLE	PAGE
3.1	Geographical coordinates of the 40 selected rainfall stations in Peninsular Malaysia from 1975-2008. <i>C</i> Center; <i>E</i> East; <i>NW</i> Northwest; <i>SW</i> Southwest; <i>W</i> West	34
3.2	The examined atmospheric variables to input into the NHMM from the NCEP reanalysis dataset.	35
3.3	The frequency of State 1 during the Northeast monsoon.	49
3.4	The dry spell distribution in Station 1 (Janda baik) during the Northeast monsoon. Group 1 and Group 2 represent the observed and simulated data via the EM algorithm, respectively.	64
3.5	Log-rank test statistic for Station 1 (Janda Baik) during the Northeast monsoon	65
4.1	Summary of the descriptive statistics of rainfall data during the Northeast monsoon and Southwest monsoon season. Sn-Station; Total- Total rainfall amount ( <i>mm</i> ) for the monsoon season; Mean-Daily mean ( <i>mm</i> ); SD- Standard deviation ( <i>mm</i> ); n(w)-Number of wet days; P(w)-Probability of wet days; Re-Regional.	72
4.2	Bayesian information criterion (BIC) for two to six states HMMs during the Northeast monsoon and Southwest monsoon. Log ( <i>L</i> ): log-likelihood	73
4.3	Estimated parameters from the EM algorithm and <i>K</i> -means algorithm during the Northeast monsoon	74
4.4	Estimated parameter from the EM algorithm and <i>K</i> -mean algorithm during the Southwest monsoon	79

4.5	Rainfall statistics for the identified hidden states from the EM algorithm and <i>K</i> -means algorithm during Northeast monsoon	84
4.6	Rainfall statistics for the identified hidden states generated using the EM algorithm and <i>K</i> -means algorithm during the Southwest monsoon	88
4.7	The bias values on the observed inter-station correlation during the Northeast and Southwest monsoon.	94
4.8	RMSE for the observation wet/dry spell versus EM and <i>K</i> -means simulation wet/dry spell	107
4.9	The results of the log-rank test statistics on the survival function curves. <i>No</i> represents no significant difference; <i>Yes</i> represents a significant difference; <i>NEM</i> represents the Northeast monsoon; <i>SWM</i> represents the Southwest monsoon. <i>Total</i> represents the total number of “No”.	114
5.1	Skill scores of the summary variables. CC is the mean absolute Pearson correlation among all the station between summary variable and the observed rainfall amount and CV the percentage of correlation explained by the first summary variable	119
5.2	The Bayesian Information criterion (BIC) for two- to ten-states HMMs and the four different amounts distribution during the Northeast monsoon	121
5.3	The Bayesian Information criterion (BIC) for two- to ten-states HMMs and the four different amounts distribution during the Southwest monsoon	122
5.4	Estimated HMM parameters for rainfall amounts during the Northeast monsoon	123
5.5	Estimated HMM parameters for rainfall amounts during the Southwest monsoon	126
5.6	Rainfall statistics for the identified hidden state of HMM during the Northeast monsoon	132
5.7	Rainfall statistics for the identified hidden state of HMM during the Southwest monsoon	133
5.8	Comparison of the BIC for the models during Northeast monsoon. <b>S6</b> six hidden states model; <b>n1</b> Precipitable water; <b>n2</b> Relative humidity 700hPa; <b>n3</b> Specific humidity 700hPa; <b>n4</b> Specific humidity 850hPa	138

5.9	Comparison of the BIC for the models during Southwest monsoon. <b>S5</b> five hidden states model; <b>m1</b> Zonal wind 700hPa; <b>m2</b> Zonal wind 850hPa; <b>m3</b> Meridional wind 200hPa; <b>m4</b> Meridional wind 1000hPa.	139
5.10	Rainfall statistics for the identified hidden state of NHMM during the Northeast monsoon	140
5.11	Rainfall statistics for the identified hidden state of NHMM during the Southwest monsoon	142
5.12	Summary of the hidden state patterns during the Northeast monsoon	146
5.13	Summary of the hidden state patterns during the Southwest monsoon	150
5.14	The years used to create the SST anomaly composites during the Northeast monsoon	152
5.15	The years used to create the SST anomaly composites during the Southwest monsoon	154
5.16	RMSE for the predicted versus observed results.	169

## LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
2.1	Structure of the HMM	19
2.2	Structure of the NHMM	22
2.3	El Niño. The red color represents the warmer than normal tropical Pacific sea surface temperature and the blue color represents the cooler than normal tropical Pacific sea surface temperature. Available from: < <a href="http://www.esrl.noaa.gov/psd/">http://www.esrl.noaa.gov/psd/</a> >	26
2.4	La Niña. The red color represent the warmer than normal tropical Pacific sea surface temperature and the blue color represents the cooler than normal tropical Pacific sea surface temperature. Available from: < <a href="http://www.esrl.noaa.gov/psd/">http://www.esrl.noaa.gov/psd/</a> >	26
2.5	Location of the Darwin and Tahiti in which its sea level pressures contribute to the Southern Oscillation Index. Available from: < <a href="http://www.esrl.noaa.gov/psd/">http://www.esrl.noaa.gov/psd/</a> >	28
2.6	Location of the Niño regions for measuring sea surface temperature. Available from: < <a href="http://www.esrl.noaa.gov/psd/">http://www.esrl.noaa.gov/psd/</a> >	29
3.1	Flowchart summarizing the methodological framework for this study.	32
3.2	The coordinate of the rainfall stations and rainfall regions in Peninsular Malaysia. <i>NW</i> northwestern; <i>SW</i> southwestern; <i>E</i> eastern; <i>C</i> central; <i>W</i> western.	37
3.3	Graphical representation of a wet or dry spell occurrence	63
4.1	Error bar at various rainfall regions for each region in Peninsular Malaysia (S1 refers to Station 1). Note: ‘x’ indicates the mean and the line indicates standard error.	72

- 4.2 (i) and (ii) The estimated state sequence. (iii) and (iv) The accumulated days in the state over 34 years during the Northeast monsoon via the (a) EM algorithm and (b) *K*-means algorithm. 85
- 4.3 HMM rainfall probabilities from a reddish (dry) to bluish (wet) color and anomaly composites of 500hPa omega vertical velocity (contour, Pa/s) with winds at 850hPa for each hidden state during the Northeast monsoon from 1975 to 2008. Only statistically significant wind vectors at 95% level are plotted. 88
- 4.4 (i) and (ii) The estimated state sequence. (iii) and (iv) The accumulated days according to state over 34 years during the Southwest monsoon generated from the (a) EM algorithm and the (b) *K*-means algorithm 89
- 4.5 HMM rainfall probabilities from reddish (dry) to bluish (wet) color and anomaly composites of 500hPa omega vertical velocity (contour, Pa/s) with winds at 850hPa for each hidden state during the Southwest monsoon for the period of 1975–2008. Only wind vectors that are statistically significant at 95% level are plotted. 92
- 4.6 The daily rainfall probability between 40 stations for the observed data versus the simulated data. NE refers to the Northeast monsoon and SW refers to Southwest monsoon 93
- 4.7 The observed versus simulated Spearman’s inter-station correlations. NE refers to the Northeast monsoon and SW refers to the Southwest monsoon. 94
- 4.8 Comparison of the wet spell between the observed data (blue), simulated data from the EM algorithm (green), and simulated data from the *K*-means algorithm (red) during the Northeast monsoon. 96
- 4.9 Comparison the dry spell between the observed (blue), simulated from EM algorithm (green) and simulated from *K*-means algorithm (red) during the Northeast monsoon 97
- 4.10 The number of wet spells for a wet spell length of: (a) 1 day; (b) 2 days; (c) 3 days; (d) 4 days; (e) 5 days; (f) 6 days; (g) 7 days; and (h) 8 days during the Northeast monsoon. 99
- 4.11 The number of dry spells for a wet spell length of: (a) 1 day; (b) 2 days; (c) 3 days; (d) 4 days; (e) 5 days; (f) 6

	days; (g) 7 days; and (h) 8 days during the Northeast monsoon.	100
4.12	Comparison of the wet spells between the observed data (blue), simulated data from EM algorithm (green), and simulated data from <i>K</i> -means algorithm (red) during the Southwest monsoon.	102
4.13	Comparison of the dry spell between the observed data (blue), data simulated from EM algorithm (green), and data simulated from <i>K</i> -means algorithm (red) during the Southwest monsoon.	103
4.14	The number of wet spells for a wet spell length of: (a) 1 day; (b) 2 days; (c) 3 days; (d) 4 days; (e) 5 days; (f) 6 days; (g) 7 days; and (h) 8 days during the Southwest monsoon.	104
4.15	The number of dry spells for a wet spell length of: (a) 1 day; (b) 2 days; (c) 3 days; (d) 4 days; (e) 5 days; (f) 6 days; (g) 7 days; and (h) 8 days during the Southwest monsoon.	105
4.16	Observed versus mean simulated wet duration distribution of rainfall from the EM algorithm and <i>K</i> -means algorithm at each station during the Northeast monsoon.	109
4.17	Observed versus mean simulated dry duration distribution of rainfall from the EM algorithm and <i>K</i> -means algorithm at each station during the Northeast monsoon.	110
4.18	Observed versus mean simulated wet duration distribution of rainfall from the EM algorithm and <i>K</i> -means algorithm at each station during the Southwest monsoon.	112
4.19	Observed versus mean simulated dry duration distribution of rainfall from the EM algorithm and <i>K</i> -means algorithm at each station during the Southwest monsoon.	113
5.1	Coordinates of the thirty-five gridded reanalysis data (95°E-110°E, 0-10°N)	118
5.2	BIC for non-zero rainfall models by single exponential, mixture of two exponential, single gamma, and mixture of two gamma distribution.	121

5.3	The BIC for non-zero rainfall models via single exponential, a mixture of two exponentials, single-Gamma, and a mixture of two Gamma distributions during the Southwest monsoon.	122
5.4	HMM rainfall parameters (a)-(f) rainfall occurrence probability from reddish (dry) to bluish (wet) and (g)-(l) rainfall amount on days receiving larger than 0.3 mm rainfall during the Northeast monsoon. Note: The size of the circles refers to the value of occurrence probability and amount of rainfall.	130
5.5	HMM rainfall parameters (a)-(e) rainfall occurrence probability from reddish (dry) to bluish (wet) and (f)-(j) rainfall amount on days receiving larger than 0.3 mm rainfall during the Southwest monsoon. Note: The size of the circles refers to the value of occurrence probability and amount of rainfall.	131
5.6	(a) The estimated state sequence (b) The accumulated state over 34 years during the Northeast monsoon	132
5.7	(a) The estimated state sequence (b) The accumulated state over 34 years during the Southwest monsoon.	134
5.8	Quantile-quantile plots of observed versus simulated rainfall amounts for all the stations during the Northeast monsoon.	136
5.9	Quantile-quantile plots of observed versus simulated rainfall amounts for all the stations during the Southwest monsoon.	137
5.10	BIC for models with different combinations of atmospheric variables during the Northeast monsoon.	139
5.11	BIC for models with different combinations of atmospheric variables during the Southwest monsoon.	140
5.12	(a) The estimated state sequence (b) The accumulated days in particular states over the 34 year Northeast monsoon rainfall data	141
5.13	(a) The estimated state sequence (b) The accumulated days in particular states over 34 years of Southwest monsoon rainfall data	142
5.14	The specific humidity of 850hPa for the mean of Nov–Feb	144
5.15	NHMM rainfall parameters including the first summary	



	variable for a specific humidity of 850hPa: (a) rainfall occurrence probabilities from reddish (dry) to bluish; (b) rainfall amount on days receiving larger than 0.3 mm rainfall, from light bluish (low amounts) to dark bluish (high amounts); and (c) Mean of the specific humidity of 850hPa during the Northeast monsoon. Note: The size of the circles refers to the value of rainfall probability and amounts.	146
5.16	The 700hPa zonal wind for the mean of May–Aug	148
5.17	NHMM rainfall parameters including the first summary variable for the 700hPa zonal wind: (a) rainfall occurrence probabilities from reddish (dry) to bluish; (b) rainfall amount on days receiving larger than 0.3 mm rainfall from light bluish (low amounts) to dark bluish (high amounts); and (c) Mean of the 850hPa specific humidity during the Southwest monsoon. Note: The size of the circles refers to the value of probability and amounts.	149
5.18	(a) Inter-annual variability and (b) 9-year running mean of state frequency and NINO 3.4 index (multiple of 10) during the Northeast monsoon.	152
5.19	(a) Inter-annual variability and (b) 9-year running mean of state frequency and NINO 3.4 index (multiple of 10) during the Southwest monsoon	154
5.20	Composites of seasonal mean (November to February) sea surface temperature anomalies defined as years in the upper 15% of the inter-annual distribution of the state frequency during the Northeast monsoon. The number of seasons in each composite is given in brackets. The shading represents 90% statistical significant according to a two-tailed Student <i>t</i> -test. The solid lines depict positive contours and the negative contours are shown as dashed lines for which the contour interval is 0.25.	155
5.21	Composites of seasonal mean (May to August) sea surface temperature anomalies defined as years in the upper 15% of the inter-annual distribution of the state frequency during the Southwest monsoon. The number of seasons in each composite is given in brackets. The shading represents 90% statistical significant according to a two-tailed Student <i>t</i> -test. The positive contours are represented with solid lines and the negative contours are shown as dashed lines for which the contour interval is 0.25.	156

- 5.22 Inter-annual variability of HMM and NHMM-simulated during the Northeast monsoon, (a) seasonally-averaged rainfall amount, (b) occurrence frequency, and (c) rainfall intensity (averaged rainfall amount on wet days). Plotted is the median of 100 HMM and 100 NHMM simulations averaged over all 40 stations. 158
- 5.23 Inter-annual variability of HMM and NHMM-simulated during Southwest monsoon: (a) seasonally-averaged rainfall amount; (b) occurrence frequency; and (c) rainfall intensity (averaged rainfall amount on wet days). Plotted is the median of 100 HMM and 100 NHMM simulations averaged over all 40 stations. 159
- 5.24 The boxplot of the entire range of 100 simulations during the Northeast monsoon, (a) seasonally-averaged rainfall amount, (b) occurrence frequency, (c) rainfall intensity (averaged rainfall amount on wet days) for (i) The HMM, and (ii) The NHMM. The line depicts the observation data averaged over all 40 stations. 161
- 5.25 The boxplot of the entire range of 100 simulations during the Southwest monsoon, (a) seasonally-averaged rainfall amount, (b) occurrence frequency, and (c) rainfall intensity (averaged rainfall amount on wet days) for (i) The HMM and (ii) NHMM. The line depicts the observation data averaged over all 40 stations. 163
- 5.26 Inter-annual variability of HMM and NHMM prediction during the Northeast monsoon, (a) seasonally-averaged rainfall amount, (b) occurrence frequency, and (c) rainfall intensity (averaged rainfall amount on wet days). Plotted is the median of 100 HMM and 100 NHMM predictions averaged over all 40 stations. 166
- 5.27 Inter-annual variability of HMM and NHMM prediction during the Southwest monsoon, (a) seasonally-averaged rainfall amount, (b) occurrence frequency, (c) and rainfall intensity (averaged rainfall amount on wet days). Plotted is the median of 100 HMM and 100 NHMM prediction averaged over all 40 stations. 167
- 5.28 The boxplot of the entire range of 100 predictions during the Northeast monsoon; (a) The HMM and (b) The NHMM for; (i) seasonally-averaged rainfall

amount; (ii) occurrence frequency; and (iii) rainfall intensity (averaged rainfall amount on wet days). The line depicts the observation data averaged over all 40 stations.

168

5.29

The boxplot of the entire range of 100 predictions during the Southwest monsoon: (a) The HMM and (b) The NHMM for: (i) seasonally-averaged rainfall amount; (ii) occurrence frequency; and (iii) rainfall intensity (averaged rainfall amount on wet days). The line depicts the observation data averaged over all 40 stations.

169

**LIST OF ABBREVIATIONS**

AIC	-	Akaike's Information Criterion
ARMA	-	Autoregressive Moving Average
BIC	-	Bayesian Information Criterion
C	-	Central
CART	-	Classification and Regression Trees
CDF	-	Cumulative distribution function
E	-	Eastern
EM	-	Expectation Maximization
ENSO	-	El-Niño Southern Oscillation
FO	-	Frequency Occurrence
GCM	-	General Circulation Model
HMM	-	Hidden Markov Model
MCMC	-	Markov Chain Monte Carlo
MMD	-	Malaysian Meteorology Department
MJJA	-	May-June-July-August
NCAR	-	National Center for Atmospheric Research
NCEP	-	National Centers for Environmental Prediction
NDJF	-	November-December-January-February
NHMM	-	Non-Homogeneous Hidden Markov Model
NOAA	-	National Oceanic and Atmospheric Administration
NW	-	Northwestern
PC	-	Principal Component
QQ	-	Quantile-Quantile
SD	-	Standard Deviation
RMSE	-	Root Mean Square Error
SST	-	Sea Surface Temperature

SVD	-	Singular Value Decomposition
SW	-	Southwestern
TPM	-	Transition Probability Matrix
W	-	Western

## LIST OF SYMBOLS

$a_{ij}$	-	Transition probability from $i$ to $j$
$n_{ij}$	-	The number of transition from $i$ to $j$
$M$	-	The number of stations
$N$	-	The number of the hidden states
$S$	-	The hidden sequence
$\pi$	-	Initial Probability
$G$	-	The emission probability
$p_{(is)}$	-	The probability of occurrence at station $i$ for state $s$
$c$	-	Threshold of rainfall
$\Gamma(.)$	-	Gamma function
$\alpha$	-	Shape parameter for Gamma distribution
$\beta$	-	Scale parameter for Gamma distribution
$\Sigma$	-	The variance-covariance matrix
$k$	-	The number of mixture component
$g$	-	The number of grid nodes
$T$	-	The number of observation
$f_t$	-	The forward variable at time $t$
$b_t$	-	The backward variable at time $t$
$L$	-	Likelihood
$v$	-	The number of parameters of the model

$sc$	-	Scaling coefficient
$S_p^2$	-	Pooled sample variance
$Sv(.)$	-	Survival function
$SE(.)$	-	Standard error of a function
$z$	-	The significance of the log odds ratio
$cp$	-	Cumulative probability
$F^{-1}(.)$	-	The inverse of the cumulative distribution function
<b>R</b>	-	The multivariate rainfall process
<b>W</b>	-	The complete data
<b>I</b>	-	The identity matrix
$\mu$	-	Mean vector
<b>X</b>	-	Vector of atmospheric process
$\mathbf{X}_{(std)}$	-	Standardized atmospheric variable

**LIST OF APPENDICES**

<b>APPENDIX</b>	<b>TITLE</b>	<b>PAGE</b>
A	Hidden Markov model using EM algorithm by Matlab coding	188
B	Hidden Markov model using <i>K</i> -means algorithm by Matlab coding	195
C	Output from the Viterbi algorithm	200
D	The wet and dry spell from observation, EM <i>K</i> -means algorithm	212
E	The log-rank test statistics on the survival function.	237
F	The data for the 9-years running mean of state Frequency and Niño 3.4 index (multiple 10)	238
G	The year choose for SST anomaly composites	240
H	Publications / proceedings	242



## **CHAPTER 1**

### **INTRODUCTION**

#### **1.1 Introduction**

The weather is the state of the atmosphere such as the wind speed and direction, air temperature and pressure, relative and specific humidity, rainfall occurrence, and rainfall amounts in a particular region over a short-term period. In view of this, climate describes the weather pattern of a particular region over a long-term period and is also defined as the average weather pattern of a region. Malaysia is located in the equatorial zone, situated in the approximate northern latitude between 1°N and 6°45'N and the approximate eastern longitude of 99°36'E to 104°24'E. The atmospheric temperature in Malaysia is uniformly high throughout the year. The climate of Malaysia has a seasonal rhythm because Malaysia experiences the Northeast monsoon from November until February and the Southwest monsoon from May until August. The Northeast monsoon brings about heavier rainfall, especially in the eastern region, whereas the Southwest monsoon contributes towards the dry period in Peninsular Malaysia.

Rainfall has an important impact on human and physical environments because all living things cannot live without water. With the rapid growth of population and industrialization, the management of water resources has become an increasing concern in Malaysia. The analysis of rainfall behavior, particularly regarding rainfall amount and rainfall occurrence, is useful for managing water consumption. This thesis therefore investigates the characteristics of monsoon

rainfall associated with the atmospheric circulation in Peninsular Malaysia on the spatial and temporal scale.

## 1.2 Background of the Study

A stochastic rainfall model is always preferred compared to a deterministic model due to the complexity, randomness, and dynamic behavior of rainfall. Stochastic rainfall modeling on a space and time scale is essential in providing information on the probabilistic structure of the rainfall process in certain regions. It is crucial to predict the trend in rainfall for managing water resources and natural disaster planning preventions such as flood or drought. Furthermore, simulations from the stochastic rainfall model can be used as input to stream flow, crop growth, runoff, soil loss, and flooding models. Therefore, knowledge based on analysis of rainfall characteristics is essential in order to understand rainfall patterns for designing, planning, and managing various water resource systems.

Stochastic models on rainfall are mostly developed in isolation, i.e. without reference to atmospheric information. Atmospheric information may be included as part of the rainfall model to produce simulations, which are consistent with the given atmospheric information. In addition, a model, which incorporates atmospheric information, would be useful in studies investigating the impact of climate change. Therefore, the stochastic rainfall model, also known as the "weather state model" or "downscaling model", has been developed for this purpose (Hay *et al.*, 1991; Bardossy and Plate, 1991; Hughes *et al.*, 1993). The weather state model can be used to generate realistic rainfall simulations using historical rainfall data and atmospheric data. Furthermore, the weather state model may also be used to study the impact of climate change and variability of rainfall using atmospheric simulations from a General Circulation Model (GCM).

A weather state model, namely the Hidden Markov model (HMM), has been developed to condition the daily rainfall in regard to the available atmospheric

information at multiple sites. The Hidden Markov model is a doubly stochastic process in which the rainfall observation distribution depends on several unobserved discrete states (Rabiner and Juang, 1986). The Hidden Markov models have become popular tools for modeling dependent random variables in such diverse areas such as DNA recognition, speech processing, and rainfall modeling. For rainfall modeling, the hidden (unobserved) states of HMM can be used to interpret the various patterns of circulation anomalies (Robertson *et al.*, 2004; Robertson *et al.*, 2005; Greene *et al.*, 2008). The HMM can be extended to model non-stationary processes by incorporating time-varying atmospheric variables, which is known as the non-Homogeneous Hidden Markov model (NHMM). This model exhibits unobserved weather states and serves as a link between the local rainfall process and large-scale atmospheric information.

This thesis presents the modeling of daily rainfall during 2 monsoon seasons using the Hidden Markov model. The distinct hidden states of HMM are interpreted by relating the local rainfall process with atmospheric circulation. Therefore, the HMM can be extended to NHMM by including the atmospheric variable into the model for which the patterns of the atmospheric variable correspond to the hidden states. The variability of the Northeast and Southwest monsoon over Peninsular Malaysia is analyzed and this variability will then be related to atmospheric information using time scales from daily to multi-decadal.

### **1.3 Statement of the Problem**

Malaysia receives a large amount of rain every year. Therefore, floods are the most significant natural hazard in Malaysia in terms of population affected, frequency, flood duration, area extent, and social-economic damage (Youssef *et al.*, 2011). According to Feng and Lu (2010), about 40% of total economic loss caused by natural disasters is due to flooding. Spatial and temporal rainfall modeling have important impacts on physical environments. The trend in future rainfall may be easier to predict when the rainfall characteristics are known. However, a more recent

issue that has been of concern is the climate change and global warming impact on rainfall. Nevertheless, most studies so far have only concentrated on stochastic rainfall models, which fail to incorporate atmospheric information. The stochastic rainfall model that does not include atmospheric information will not be so useful in studies investigating the effect of climate change or global warming. Therefore, prediction of the variability of rainfall for future periods under different climate change scenarios is essential in order to provide necessary information for high-quality climate-related studies.

In order to assess the effect of climate change on the rainfall trend in Malaysia, it is necessary to use a stochastic rainfall model that can incorporate atmospheric information. This study will focus on rainfall modeling using a Hidden Markov model by associating synoptic atmospheric patterns to the local rainfall in Peninsular Malaysia. The incorporation of the time-varying atmospheric variable into the NHMM can be used to generate realistic rainfall simulations, for example: an extreme rainfall event.

#### **1.4 Objectives of the Study**

The objectives of the research are:

- 1 To model rainfall occurrences and rainfall amounts on multi-site rainfall stations using a Hidden Markov Model (HMM).
- 2 To compare the parameter estimation of HMM using *K*-means and EM algorithm for rainfall occurrences.
- 3 To assess the performance of HMM on the atmospheric composite with rainfall probabilities in Peninsular Malaysia.
- 4 To model rainfall amounts using a non-homogenous Hidden Markov model (NHMM).
- 5 To assess the performance of NHMM on the interannual and interdecadal variability in Peninsular Malaysia.

## 1.5 Scope of the Study

The scopes of this research consists of the following:

1. Forty rainfall stations from Peninsular Malaysia are selected. The rainfall data is obtained from the Malaysian Meteorology Department (MMD). The rainfall record over 34 years (1975-2008) is thus compiled;
2. Thirty-five atmospheric variables with different pressure levels are considered as potential candidates for the NHMM input. Thirty-five grid nodes that cover Peninsular Malaysia and the sea surrounding are reduced via singular value decomposition (SVD). The atmospheric data were obtained from the National Centers for Environment Prediction (NCEP) reanalysis data.
3. Two seasonal monsoons, which are the Northeast monsoon from November until February and the Southwest monsoon from May to August, are considered in this study.
4. The optimum number of hidden states for HMM and the number of atmospheric variables included in NHMM are determined via the Bayesian Information Criterion (BIC).
5. The binomial distribution is used to model the rainfall occurrence at each station in HMM. Four probability distribution functions for non-zero rainfall amount in each station, namely: single exponential distribution, single gamma distribution, the mixture of two exponential distribution, and the mixture of two gamma distribution, were examined and selected based on the BIC.

## 1.6 Significance of the Study

Climate change is undoubtedly one of the most important global environmental issues today. The impact of climate change affects the rainfall pattern, for example, increases the severity of floods and causes longer drought, and heavier thunderstorms. However, little work has been done on rainfall that is associated with the large atmospheric circulations in Malaysia. Therefore, a model called the Hidden Markov model was proposed in this study.

This thesis first fits the rainfall occurrence and rainfall amounts via HMM in Peninsular Malaysia and then identifies the physical definition of each state in the HMM with large-scale atmospheric behaviors and rainfall pattern. The state sequence of the HMM that categorizes each day into a state provides an exhaustive description of the rainfall process and is able to accurately estimate the future rainfall process. This model can provide a useful descriptive analysis of the rainfall in Malaysia.

In general, the EM algorithm is used to estimate the parameters of HMM. Another parameter estimation algorithm called segmental  $K$ -means for the HMM is also used in this study. The segmental  $K$ -means may provide more flexibility for the algorithm in the rainfall-modeling framework.

The HMM can be extended to NHMM by including atmospheric variables into the model. The goals of this work are to analyze the subseasonal to multidecadal variability of monsoon rainfall and produce simulations or predictions, which are consistent with the included atmospheric information. The state sequence of the NHMM can be used to classify the ENSO into a few states and the relationship between ENSO and monsoon rainfall can be investigated. The findings from NHMM are useful for the assessment of the impact of climate change and tend to result in a good model for descriptive and predictive modeling of the rainfall in Malaysia.

## 1.7 Structure of the Thesis

This thesis is organized into six chapters. The first chapter begins with the introduction to the research. This chapter also presents the background, the problem statement, the objectives, and the scope of research.

Chapter 2 provides an overview of the literature on rainfall modeling based on rainfall occurrence and rainfall amounts.

Chapter 3 outlines the description of rainfall data. Parameter estimation via EM algorithm and *K*-means algorithm; algorithm for find the optimum states paths; steps to generate the simulations and predictions; and methods for model assessments, are also discussed in detail.

Chapter 4 presents the HMM results in regard to rainfall occurrence. The results estimated from the *K*-means and EM algorithm are then compared. The relationship between the rainfall process and atmospheric circulations are interpreted in detail.

Chapter 5 shows the results of the HMM and NHMM in regard to rainfall amounts. The atmospheric variables are described and the atmospheric variable to be included into NHMM are selected. The comparison between the simulations and predictions from the HMM and NHMM is also discussed.

Finally, Chapter 6 concludes the study. This chapter summarizes the study and conclusion based on the analysis and results of this study. This chapter also suggests some recommendations for future study.

## REFERENCES

- Aldrian, E. and Dwi Susanto, R. (2003). Identification of Three Dominant Rainfall Regions within Indonesia and their Relationship to Sea Surface Temperature. *International Journal of Climatology*. 23(12), 1435-1452.
- Ailliot, P., Thompson, C. and Thomson, P. (2009). Space-time Modelling of Precipitation by Using a Hidden Markov Model and Censored Gaussian Distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 58(3), 405-426.
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*. 19(6), 716-723.
- Anagnostopoulou, C. H. R., Maheras, P., Karacostas, T. and Vafiadis, M. (2003). Spatial and Temporal Analysis of Dry Spells in Greece. *Theoretical and Applied Climatology*. 74(1-2), 77-91.
- Bardossy, A. and Plate, E. J. (1991). Modeling Daily Rainfall Using a Semi-Markov Representation of Circulation Pattern Occurrence. *Journal of Hydrology*, 122(1-4), 33-47.
- Bardossy, A. and Plate, E. J. (1992). Space-Time Models for Daily Rainfall Using Atmospheric Circulation Patterns. *Water Resources Research* 28(5), 1247-1259
- Barnston, A. G., Chelliah, M. and Goldenberg, S. B. (1997). Documentation of a Highly ENSO-Related SST Region in the Equatorial Pacific. *Atmosphere-Ocean*. 35(3), 367-383.
- Baum, L. (1972). An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes. *Inequalities* 3, 1-8
- Bellenger, H., Guilyardi, É., Leloup, J., Lengaigne, M. and Vialard, J. (2014). ENSO Representation in Climate Models: from CMIP3 to CMIP5. *Climate Dynamics*. 42(7-8), 1999-2018.



- Bellone, E., Hughes, J. P. and Guttorp, P. (2000). A Hidden Markov Model for Downscaling Synoptic Atmospheric Patterns to Precipitation Amounts. *Climate Research*. 15(1), 1-12.
- Bernardara, P., De Michele, C., and Rosso, R. (2007). A Simple Model of Rain in Time: An Alternating Renewal Process of Wet and Dry States with a Fractional (non-Gaussian) Rain Intensity. *Atmospheric Research*. 84(4), 291-301.
- Besson, L. (1924). On the Probability of Rain. *Monthly Weather Review*. 52(6), 308.
- Bjerknes, J. A. B. (1969). Atmospheric Teleconnections from the Equatorial Pacific. *Monthly Weather Review*. 97(3), 163-172.
- Boughton, W. C. and Hill, P., (1997). *A Design Flood Estimation Procedure Using Data Generation and a Daily Water Balance Model*. Technical Report. CRC for Catchment Hydrology, Monash University, Melbourne.
- Boughton, W.C., 1999. *A Daily Rainfall Generating Model for Water Yield and Flood Studies*. Technical Report. CRC for Catchment Hydrology, Monash University, Melbourne
- Bretherton, C. S., Smith, C. and Wallace, J. M. (1992). An Intercomparison of Methods for Finding Coupled Patterns in Climate Data. *Journal Climate*. 5(6), 541–560.
- Brijker, J. M., Jung, S. J., Ganssen, G. M., Bickert, T. and Kroon, D. (2007). ENSO Related Decadal Scale Climate Variability from the Indo-Pacific Warm Pool. *Earth and Planetary Science Letters*. 253(1), 67-82.
- Buishand, T. A. (1977). *Stochastic Modelling of Daily Rainfall Sequences*. Mededelingen Landbouwhogeschool, Wageningen (Netherlands). 1-211.
- Buishand, T. A. (1978). Some Remarks on the Use of Daily Rainfall Models. *Journal of Hydrology*. 36(3-4), 295-308.
- Chang, C. P., Wang, Z., Ju, J. and Li, T. (2004). On the Relationship between Western Maritime Continent Monsoon Rainfall and ENSO during Northern Winter. *Journal of Climate*. 17(3), 665-672.
- Chapman, T. (1997). Stochastic Modelling of Daily Rainfall in the Western Pacific. *Mathematics and Computers in Simulation*. 43(3-6), 351-358.
- Charles, S. P., Bates, B. C., Whetton, P. H. and Hughes, J. P. (1999a). Validation of Downscaling Models for Changed Climate Conditions: Case Study of Southwestern Australia. *Climate Research*. 12(1), 1-14.

- Charles, S. P., Bates, B. C. and Hughes, J. P. (1999b). A Spatiotemporal Model for Downscaling Precipitation Occurrence and Amounts. *Journal of Geophysical Research*. 104 (D24): 31657-31669.
- Charles, S. P., Bates, B. C., Smith, I. N. and Hughes, J. P. (2004). Statistical Downscaling of Daily Precipitation from Observed and Modelled Atmospheric Fields. *Hydrological Processes*. 18(8), 1373-1394.
- Cheang, B. K., (1993). Interannual Variability of Monsoons in Malaysia and its Relationship with ENSO. *Proceedings of the Indian Academy of Sciences-Earth and Planetary Sciences*. 102(1), 219-239.
- Chen, T. C., Tsay, J. D., Yen, M. C. and Matsumoto, J. (2013). The Winter Rainfall of Malaysia. *Journal of Climate*. 26(3), 936-958.
- Chin, E. H. (1977). Modeling Daily Precipitation Occurrence Process with Markov Chain. *Water Resources Research*. 13(6), 949-956.
- Cochran, W. G. (1938). An Extension of Gold's Method of Examining the Apparent Persistence of One Type of Weather. *Quarterly Journal of the Royal Meteorological Society*. 64(277), 631-34.
- Cowden, J. R., Mihelcic, J. R. and Watkins D. W. (2007). Stochastic Rainfall Occurrence Modeling in Benin. *Proceedings of World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat*. 15-19 May. Tampa, FL, USA, 1-8.
- Dai, A., Trenberth, K. E. and Qian, T. (2004). A Global Dataset of Palmer Drought Severity Index for 1870-2002: Relationship with Soil Moisture and Effects of Surface Warming. *Journal of Hydrometeorology*. 5(6), 1117-1130.
- Dambul, R. (2010). Monsoon Indicators for Borneo. *Geografia: Malaysian Journal of Society and Space*. 6(3), 1-12.
- De Groen, M. M. (2002). *Modeling Interception and Transpiration at Monthly Time Steps: IHE Dissertation 31*. PhD Thesis. IHE-Delft, Swets and Zeitlinger, Lisse, The Netherlands.
- Deni, S. M., Jemain, A. A. and Ibrahim, K. (2008a). Fitting Optimum Order of Markov Chain Models for Daily Rainfall Occurrences in Peninsular Malaysia. *Theoretical and Applied Climatology*. 97(1-2), 109-121.
- Deni, S. M., Jemain, A. A. and Ibrahim, K. (2008b). The Spatial Distribution of Wet and Dry Spells over Peninsular Malaysia. *Theoretical and Applied Climatology*, 94(3-4), 163-173.

- Feng, L. H., Lu, J. (2010). The Practical Research on Flood Forecasting Based on Artificial Neural Networks. *Expert Systems with Applications*. 37(4), 2974–2977
- Forney, G. D., Jr. (1973). The Viterbi Algorithm. *Proceedings of the IEEE*. 61(3), 268-278.
- Foufoula-Georgiou, E. and Lettenmeier, D. (1987). A Markov Renewal Model for Rainfall Occurrence. *Water Resources Research*. 23(5), 875-884.
- Fu, G., Charles, S. P. and Kirshner, S. (2012). Daily Rainfall Projections from General Circulation Models with a Downscaling Nonhomogeneous Hidden Markov Model (NHMM) for South-eastern Australia. *Hydrological Processes*. 27(25), 3663-3673.
- Gabriel, K. R., and Neumann, J. (1962). A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv. *Quarterly Journal of Royal Meteorology Society*. 88(375), 90-95.
- Gold, E. (1929). Note on the Frequency of Occurrence of Sequences in a Series of Events of Two Types. *Quarterly Journal of the Royal Meteorological Society*. 55(231), 307–309.
- Gallo, E., Le Breton, A and Martin, S. (1981). *A Model for Weather Cycles Based on Daily Rainfall Occurrence*. In Cosnard, M., Demongeot, J. and Breton, A. L. (Eds.) *Rhythmes in Biology and Other Fields of Application* (pp. 303-318). London: Springer-Verlag.
- Gelati, E., Madsen, H. and Rosbjerg, D. (2010). Markov-Switching Model for Nonstationary Runoff Conditioned on El Niño Information. *Water Resources Research*, 46(2), W02517, doi:10.1029/2009WR007736.
- Goel, M., Khanna, P. and Kishore, J. (2010). Understanding Survival Analysis: Kaplan-Meier Estimate. *International Journal of Ayurveda Research*. 1(4), 274.
- Goodspeed, M. J. and Pierrehumbert, C. L. (1975). *Synthetic Input Data Time Series for Catchment Model Testing*. In Chapman, T. G. and Dunin, F. X. (Eds.) *Prediction in Catchment Hydrology*. (pp. 359-370). Canberra: Australian Academy of Science.
- Gottschalck, J., Roundy, P. E., Schreck III, C. J., Vintzileos, A. and Zhang, C. (2013). Large-Scale Atmospheric and Oceanic Conditions during the 2011–12 DYNAMO Field Campaign. *Monthly Weather Review*. 141(12), 4173-4196.
- Green, J. R. (1964). A Model for Rainfall Occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)*. 26, 345-353.

- Greene, A. M., Robertson, A. W. and Kirshner, S. (2008). Analysis of Indian Monsoon Daily Rainfall on Subseasonal to Multidecadal Time-scales Using a Hidden Markov Model. *Quarterly Journal of the Royal Meteorological Society*. 134(633), 875-887.
- Greene, A. M., Robertson, A. W., Smyth, P. and Triglia, S. (2011). Downscaling Projections of Indian Monsoon Rainfall Using a Non-homogeneous Hidden Markov Model. *Quarterly Journal of the Royal Meteorological Society*. 137(655), 347-359.
- Haan, C. T., Allen, D. M. and Street, J. O. (1976). A Markov Chain Model of Daily Rainfall. *Water Resources Research*. 12(3), 443-449.
- Hamlet, A. F., Mote, P. W., Clark, M. P. and Lettenmaier, D. P. (2005). Effects of Temperature and Precipitation Variability on Snowpack Trends in the Western United States. *Journal of Climate*. 18(21), 4545-4561.
- Hay, L., McCabe, G. J., Wolock, D. M. and Ayers, M. A. (1991). Simulation of Precipitation by Weather Type Analysis. *Water Resources Research*. 27(4), 493-501.
- Haylock, M. and McBride, J. (2001). Spatial Coherence and Predictability of Indonesian Wet Season Rainfall. *Journal of Climate*. 14(18), 3882-3887.
- Hendon, H. H. (2003). Indonesian Rainfall Variability: Impacts of ENSO and local Air-Sea Interaction. *Journal of Climate*. 16(11), 1775-1790.
- Hughes, J. P., Lettenmaier, D. P. and Guttorp, P. (1993). A Stochastic Approach for Assessing the Effect of Changes in Synoptic Circulation Patterns on Gauge Precipitation. *Water Resources Research*. 29(10), 3303-3315.
- Hughes, J. P. and Guttorp, P. (1994a). A Class of Stochastic Models for Relating Synoptic Atmospheric Patterns to Regional Hydrologic Phenomena. *Water Resources Research*. 30(5), 1535-1546..
- Hughes, J. P. and Guttorp, P. (1994b). Incorporating Spatial Dependence and Atmospheric Data in a Model of Precipitation. *Journal of Applied Meteorology* 33(12), 1503-1515.
- Hughes, J. P., Guttorp, P. and Charles, S. P. (1999). A Non-homogeneous Hidden Markov Model for Precipitation Occurrence. *Journal of the Royal Statistical Society (Series C): Applied Statistics*. 48(1), 15-30
- Jimoh, O. D. and Webster, P. (1996). The Optimum Order of a Markov Chain Model for Daily Rainfall in Nigeria. *Journal of Hydrology*. 185(1-4), 45-69.

- Jones, P. G. and Thornton, P. K. (1997) Spatial and Temporal Variability of Rainfall Related to a Third-Order Markov Model. *Agricultural and Forest Meteorology*. 86(1-2), 127-38.
- Juang, B. and Rabiner, L. R. (1990). The Segmental *K*-Means Algorithm for Estimating Parameters of Hidden Markov Models. *IEEE Transactions on Acoustics, Speech and Signal Processing*. 38(9), 1639-1641.
- Juneng, L. and Tangang, F. T. (2005). Evolution of ENSO-Related Rainfall Anomalies in Southeast Asia Region and its Relationship with Atmosphere–Ocean Variations in Indo-Pacific Sector. *Climate Dynamics*. 25(4), 337-350.
- Juneng, L., Tangang, F. T. and Reason, C. J. (2007). Numerical Case Study of an Extreme Rainfall Event during 9–11 December 2004 over the East Coast of Peninsular Malaysia. *Meteorology and Atmospheric Physics*. 98(1-2), 81-98.
- Juneng, L. and Tangang, F. T. (2008). Level and Source of Predictability of Seasonal Rainfall Anomalies in Malaysia using Canonical Correlation Analysis. *International Journal of Climatology*. 28(9), 1255-1267.
- Kabiri, R., Bai, V. R. and Chan, A. (2015). Assessment of Hydrologic Impacts of Climate Change on the Runoff Trend in Klang Watershed, Malaysia. *Environmental Earth Sciences*. 73(1), 27-37
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*. 77(3), 437–471.
- Katz, R. W. (1977). An Application of Chain-Dependent Process to Meteorology. *Journal of Applied Probability*. 14(3), 598-603.
- Katz, R. W. (1981). On Some Criteria for Estimating the Order of Markov Chain. *Technometrics*. 23(3), 243-249.
- Katz, R. W. (1985). *Probabilistic models*. In Murphy, A. H. and Katz, R. W. (Eds.) *Probability, statistics, and decision making in the atmospheric sciences* (pp. 351-362). Boulder, CO: Westview Press.
- Kiem, A. S. and Franks, S. W. (2001). On the Identification of ENSO-Induced Rainfall and Runoff Variability: a Comparison of Methods and Indices. *Hydrological Sciences Journal*. 46(5), 715-727.

- Kioutsoukis, I., Melas, D. and Zanis, P. (2008). Statistical Downscaling of Daily Precipitation over Greece. *International Journal of Climatology*. 28(5), 679-691.
- Kottegoda, N. T., Natale, L. and Raiteri (2004). Some Considerations of Periodicity and Persistence in Daily Rainfalls. *Journal of Hydrology*. 296(1-4), 23-37.
- Lee, H. S. (2015). General Rainfall Patterns in Indonesia and the Potential Impacts of Local Seas on Rainfall Intensity. *Water*. 7(4), 1751-1768.
- Lennartsson, J., Baxevani, A. and Chen, D. (2008). Modelling Precipitation in Sweden Using Multiple Step Markov Chains and a Composite Model. *Journal of Hydrology*. 363(1-4), 42-59
- Levinson, S. E., Rabiner, L. R. and Sondhi, M. M. (1993). An Introduction to the Application of the Theory of Probabilistic functions of a Markov Process to Automatic Speech Recognition. *The Bell System Technical Journal*. 62(4), 1035-1074.
- Li, J., Xie, S. P., Cook, E. R., Morales, M. S., Christie, D. A., Johnson, N. C., Chen, F., D'Arrigo, R., Fowler, A. M., Gou, X. and Fang, K. (2013). El Niño Modulations over the Past Seven Centuries. *Nature Climate Change*. 3(9), 822-826.
- Liew, J. N. (2001). Kajian Struktur Stokastik Hujan di Malaysia Pendekatan Model Markov Terpendam Tak Homogen. MSc Thesis, Universiti Kebangsaan Malaysia, Bangi.
- Lim, J. T. (1976). Rainfall Minimum in Peninsular Malaysia during the Northeast Monsoon. *Monthly Weather Review*. 104(1), 96-99.
- Longley, R. W. (1953). The Length of Dry and Wet Periods. *Quarterly Journal of the Royal Meteorological Society*. 79(342), 520-527.
- Mares, I., Mares, C. and Stanciu, P. (2008). A Hidden Markov Model for the Orsova Discharge Level in Springtime. *BALWOIS 2008*. 27-31 May 2008. Ohrid, Republic of Macedonia, 1-7.
- Ngai, S. T. (2013). *Penurunan Skala Curahan Harian di Malaysia Menggunakan Model Markov Terpendam Tak Homogen*. MSc Thesis, Universiti Kebangsaan Malaysia, Bangi.
- Newnham, E. V. (1916). The Persistence of Wet and Dry Weather. *Quarterly Journal of the Royal Meteorological Society*. 42(179), 153-162.

- Pillai, P. A. and Mohankumar, K. (2009). Role of TBO and ENSO Scale Ocean–Atmosphere Interaction in the Indo-Pacific Region on Asian Summer Monsoon Variability. *Theoretical and Applied Climatology*. 97(1-2), 99-108.
- Priya, P., Mujumdar, M., Sabin, T. P., Terray, P. and Krishnan, R. (2015). Impacts of Indo-Pacific Sea Surface Temperature Anomalies on the Summer Monsoon Circulation and Heavy Precipitation over Northwest India–Pakistan Region during 2010. *Journal of Climate*. 28(9), 3714-3730.
- Qian, J. H., Robertson, A. W. and Moron, V. (2010). Interactions among ENSO, the Monsoon, and Diurnal Cycle in Rainfall Variability over Java, Indonesia. *Journal of the Atmospheric Sciences*. 67(11), 3509-3524.
- Rabiner, L. R., and Juang B. H. (1986). An Introduction to Hidden Markov Models. *IEEE ASSP Magazine*. 3(1), 4-16.
- Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*. 77(2), 257-286.
- Racsko, P., Szeidl, L. and Semenov, M. (1991). A Serial Approach to Local Stochastic Weather Models. *Ecological Modeling*. 57, 27–41.
- Rasmusson, E. M. and Carpenter, T. H. (1982). Variations in Tropical Sea Surface Temperature and Surface Wind Fields Associated with the Southern Oscillation/El Niño. *Monthly Weather Review*. 110(5), 354-384.
- Richardson, C. W. (1981). Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation. *Water Resources Research*. 17(1), 182–90.
- Robertson, A. W., Kirshner, S. and Smyth, P. (2004). Downscaling of Daily Rainfall Occurrence over Northeast Brazil Using a Hidden Markov Model. *Journal of Climate*. 17(22), 4407-4424.
- Robertson, A. W., Kirshner, S., Smyth, P., Charles, S. P. and Bates, B. C. (2005). Subseasonal-to-interdecadal Variability of the Australian Monsoon over North Queensland. *Quarterly Journal of the Royal Meteorological Society*. 132(615), 519-542.
- Robertson, A. W., Ines, A. V. M. and Hansen, J. W. (2007). Downscaling of Seasonal Precipitation for Crop Simulation. *Journal of Applied Meteorology and Climatology*. 46(6), 677-693.
- Robertson, A. W., Moron, V. and Swarinoto, Y. (2009). Seasonal Predictability of Daily Rainfall Statistics over Indramayu District, Indonesia. *International Journal of Climatology*. 29(10), 1449-1462.

- Roldan, J. and Woolhiser, D. A. (1982). Stochastic Daily Precipitation Models: 1. A Comparison of Occurrence Processes. *Water Resources Research*. 18(5), 1451–1459.
- Ropelewski, C. F. and Halpert, M. S. (1987). Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation. *Monthly Weather Review*. 115(8), 1606-1626.
- Salimun, E., Tangang, F., Juneng, L., Behera, S. K. and Yu, W. (2014). Differential Impacts of Conventional El Niño versus El Niño Modoki on Malaysian Rainfall Anomaly during Winter Monsoon. *International Journal of Climatology*. 34(8), 2763-2774.
- Sansom, J. (1997). A Hidden Markov Model for Rainfall Using Breakpoint Data. *Journal of Climate*. 11, 42-53.
- Schoof, J. T. and Pryor, S. C. (2008). On the proper Order of Markov Chain Model for Daily Precipitation Occurrence in the Contiguous United States. *Journal of Applied Meteorology and Climatology*. 47(9), 2477-2486.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*. 6(2), 461-464.
- Selvalingam, S. and Miura, M. (1978). Stochastic Modelling of Monthly and Daily Rainfall Sequences. *Water Resources Bulletin*. 14(5), 1105-1120.
- Shaw, E. M., Beven, K. J., Chappell, N. A. and Lamb, Rob. (2010). *Hydrology in Practice*. (4th ed). New York: CRC Press.
- Small, M. J. and Morgan, D. J. (1986). The Relationship Between a Continuous-Time Renewal Model and a Discrete Markov Chain Model of Precipitation Occurrence. *Water Resources Research*. 22(10), 1422– 1430.
- Smith, T. M., Reynolds, R. W., Livezey, R. E. and Stokes, D. C. (1996). Reconstruction of Historical Sea Surface Temperatures using Empirical Orthogonal Functions. *Journal Climate*. 9(6), 1403–1420.
- Srikanthan, R. and McMahon, T. A. (1983). Stochastic Simulation of Daily Rainfall for Australian Stations. *Transaction of the ASAE*. 26(3), 754-759.
- Stern, R. D. and Coe, R. (1984). A Model Fitting Analysis of Daily Rainfall Data. *Journal of the Royal Statistical Society Series A*. 147(1), 1–34.
- Taewichit, C., Soni, P., Salokhe, V. M. and Jayasuriya, H. P. W. (2013). Optimal Stochastic Multi-States First-Order Markov Chain Parameters for Synthesizing



- Daily Rainfall Data Using Multi-Objective Differential Evolution in Thailand. *Meteorological Application*. 20(1), 20-31.
- Tan, K. S., Chiew, F. H. S. and Srikanthan, R. (2005). Comparison of Two Stochastic Spatial Daily Rainfall Generation Approaches. *International Congress on Modelling and Simulation*. December 12-25. Melbourne, 1922-1928.
- Tan, W. L., Yusof, F. and Yusop, Z. (2013). Non-homogenous Markov Model for Daily Rainfall Amount in Peninsular Malaysia. *Jurnal Teknologi*. 63(2), 75-79.
- Tan, W. L., Yusof, F. and Yusop, Z. (2016). Subseasonal to Multidecadal Variability of Northeast Monsoon Daily Rainfall over Peninsular Malaysia using a Hidden Markov Model. *Theoretical and Applied Climatology*. 1-10. doi:10.1007/s00704-016-1795-9.
- Tangang, F. T. (2001). Low Frequency and Quasi-Biennial Oscillations in the Malaysian Precipitation Anomaly. *International Journal of Climatology*. 21(10), 1199-1210.
- Tangang, F. T. and Juneng, L. (2004). Mechanisms of Malaysian Rainfall Anomalies. *Journal of climate*. 17(18), 3616-3622.
- Thom, H. C. S. (1958). A Note on the Gamma Distribution. *Monthly Weather Review*. 86(4), 117-22.
- Thyer, M., Kuczera, G. (2000). Modelling Long-term Persistence in Hydro-climatic Time Series Using a Hidden State Markov Model. *Water Resource Research*. 36 (11), 3301-3310.
- Thyer, M. and Kuczera, G. (2003a). A Hidden Markov Model for Modelling Long-term Persistence in Multi-site Rainfall Time Series 1. Model Calibration Using a Bayesian Approach. *Journal of Hydrology*. 275(1-2), 12-26
- Thyer, M. and Kuczera, G. (2003b). A Hidden Markov Model for Modelling Long-term Persistence in Multi-site Rainfall Time Series. 2. Real Data Analysis. *Journal of Hydrology*. 275(1-2), 27-48.
- Todorovic, P. and Woolhiser, D. A. (1975). A Stochastic Model of N-Day Precipitation. *Journal of Applied Meteorology*. 14(1), 17-24.
- Tolika, K. and Maheras, P. (2005). Spatial and Temporal Characteristics of Wet Spells in Greece. *Theoretical and Applied Climatology*, 81(1-2), 71-85.
- Tong, H. (1975). Determination of the Order of a Markov Chain by Akaike's Information Criterion. *Journal of Applied Probability*. 12(3), 488-497.

- Tudhope, A. W., Chilcott, C. P., McCulloch, M. T., Cook, E. R., Chappell, J., Ellam, R. M., Lea, D. W., Lough, J. M. and Shimmield, G. B. (2001). Variability in the El Niño-Southern Oscillation through a Glacial-Interglacial Cycle. *Science*, 291(5508), 1511-1517.
- Von Storch, H., and Zwiers, F. (1999) *Statistical Analysis in Climate Research*. Cambridge: Cambridge University Press.
- Wallace, J. M., Rasmusson, E. M., Mitchell, T. P., Kousky, V. E., Sarachik, E. S. and Storch, H. V. (1998). On the Structure and Evolution of ENSO-Related Climate Variability in the Tropical Pacific: Lessons from TOGA. *Journal of Geophysical Research: Oceans*. 103(C7), 14241-14259.
- Wan, H., Zhang, X. and Barrow E. M. (2005) Stochastic Modeling of Daily Precipitation for Canada. *Atmosphere-Ocean*. 43(1), 23-32.
- Wang, B., Wu, R. and Fu, X. (2000). Pacific-East Asian Teleconnection: How Does ENSO Affect East Asian Climate?. *Journal of Climate*. 13(9), 1517-1536.
- Wang, B., Yang, J., Zhou, T. and Wang, B. (2008). Interdecadal Changes in the Major Modes of Asian-Australian Monsoon Variability: Strengthening Relationship with ENSO since the Late 1970s. *Journal of Climate*. 21(8), 1771-1789.
- Ward, P. J., Jongman, B., Kumm, M., Dettinger, M. D., Weiland, F. C. S. and Winsemius, H. C. (2014). Strong Influence of El Niño Southern Oscillation on Flood Risk around the World. *Proceedings of the National Academy of Sciences*. 111(44), 15659-15664.
- Wilby, R. L. (1994). Stochastic Weather Type Simulation for Regional Climate Change Impact Assessment. *Water Resources Research*. 30(12), 3395-403.
- Wilby, R. L. (1998). Statistical Downscaling of Daily Precipitation using Daily Airflow and Seasonal Teleconnection Indices. *Climate Research*. 10, 163-178.
- Wilks, D. S. (1989). Conditioning Stochastic Daily Precipitation Models on Total Monthly Precipitation. *Water Resources Research*. 25(6), 1429-39.
- Wilks, D. S. (1992). Adapting Stochastic Weather Generation Algorithms for Climate Change Studies. *Climatic Change*. 22(1), 67-84.
- Wilks, D. S. (1998). Multisite Generalizations of a Daily Stochastic Precipitation Generation Model. *Journal of Hydrology*. 210(1-4), 178-191

- Wilks, D. S. (1999). Interannual Variability and Extreme-Value Characteristics of Several Stochastic Daily Precipitation Models. *Agricultural and Forest Meteorology*. 93(3), 153-169.
- Wilks, D. S. and Wilby, R. L. (1999). The Weather Generation Game: A Review of Stochastic Weather Models. *Progress in Physical Geography*. 23(3), 329-357.
- Wilks D. W. (2011). *Statistical Methods in the Atmospheric Sciences*. (3<sup>rd</sup> ed). Boston: Academic Press.
- Williams, C. B. (1947). The Log Series and its Applications to Biological Problems. *Journal of Ecology*. 34, 253-272.
- Williams, C. B. (1952). Sequences of Wet and Dry days Considered in Relation to the Logarithmic Series. *Quarterly Journal of the Royal Meteorological Society*. 78(335), 91-96.
- Wilson, L. L., Lettenmaier, D. P. and Skillingstad, E. (1992). A Hierarchical Stochastics Model of Large-scale Atmospheric Circulation Patterns and Multiple-Station Daily Precipitation. *Journal of Geophysical Research*. 97(D3), 2791-2809.
- Wong, C. L., Venneker, R., Uhlenbrook, S., Jamil, A. B. M. and Zhou, Y. (2009). Variability of Rainfall in Peninsular Malaysia. *Hydrology and Earth System Sciences Discussions*. 6(4), 5471-5503.
- World Meteorological Organisation WMO (1975). *Droughts and Agriculture*. WMO Tech. Note No. 138.
- Woolhiser, D. A. and Pegram, G. G. S. (1979). Maximum Likelihood Estimation of Fourier Coefficients to Describe Seasonal Variations of Parameters in Stochastic Daily Precipitation Models. *Journal of Applied Meteorology*. 18(1), 34-42
- Woolhiser, D.A. and Roldan, J. (1982) Stochastic Daily Precipitation Models: 2. A Comparison of Distribution of Amounts. *Water Resourced Research*. 18(5), 1461-1468.
- Wyrtki, K. (1985). Water Displacements in the Pacific and the Genesis of El Niño Cycles. *Journal of Geophysical Research: Oceans*. 90(C4), 7129-7132.
- Yates, R. D. and Goodman, D. J. (2005). *Probability and Stochastic Processes*. (2<sup>nd</sup> ed). Danvers: John Wiley and Sons, Inc.
- Youssef, A.M., Pradhan, B, and Hassan, A.M. (2011). Flash Flood Risk Estimation along the St. Katherine Road, Southern Sinai, Egypt using GIS based Morphometry and Satellite Imagery. *Environmental Earth Sciences*. 62(3), 611-623.

Zucchini, W. and Guttorp, P. (1991). A Hidden Markov Model for Space-time Precipitation. *Water Resources Research*. 27(8), 1917-1923.