

RAINFALL RUNOFF MODEL USING PROBABILITY DISTRIBUTED MODEL
(PDM)

YAZRE BIN YUSOF

UNIVERSITI TEKNOLOGI MALAYSIA

RAINFALL RUNOFF MODEL USING PROBABILITY DISTRIBUTED MODEL (PDM)

YAZRE BIN YUSOF

A thesis is submitted in fulfillment
of the requirements for the award of the
Master of Science (Engineering Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

JANUARY 2014

*To my beloved mother, Noryati binti Bajuri,
My father, Yusof bin Mohamad akip
My sister, Yazni Yusof and Yazwani Yusof
Thank you for all of your greatest support and everlasting love*

*Millions of thankful to my supervisor and co-supervisor
Assoc. Prof. Dr. Fadhilah Yusof
Assoc. Prof. Dr. Zalina Mohd Daud
Dr. Shariffah Suhaila Syed Jamaludin*

*To my friends, Mr. Rudi Yanto and Norain binti Ahmad Nordin who always support
and guide me
And also my best friend
Nabila 'Izzaty Mohd Yusof, Wan Anis Farina Wan Mohd Yuseri
Adam bin Samsudin, Muhammad Azrin bin Ahmad
and Abdul Rahman Mohamad Kasim
Thank you for your concern, intelligence and guidance in life
Thank you for everything*

ACKNOWLEDGEMENTS

Praise be to Allah, Lord of the Worlds, and prayers and peace be upon the Messenger of Allah. To you belongs all praise for your guidance and your care.

First and foremost, I would like to express my profound thanks to my supervisor, Associate Prof Dr. Fadhilah binti Yusof, who has suggested the dissertation topic and directed the research. I thank her for her incisive guidance, supervision and encouragement to improve my work.

I would like also to thank Dr. Shariffah Suhaila Syed Jamaludin and Assoc. Prof. Dr. Zalina binti Mohd Daud as my co-supervisor for their support and advises to improve my project dissertation. In addition, I would like to thank Mr. Rudi Yanto for sharing his knowledge in giving me some input about the software used in this research.

Finally, I would like to thank all those who have contributed to this work.

ABSTRACT

Probability Distributed Model (PDM) is widely used in analyzing the hydrological behavior. One of the applications is involved in rainfall runoff modeling to forecast the occurring of flood in Johor Bahru, Malaysia. In this study, Pareto distribution is used to represent the storage capacity of PDM. By providing the best fit between observed and simulated discharges, a Genetic Algorithm (GA) method has been applied to optimize the optimal parameter of PDM. The performance of PDM is accessed through the calibration and validation of different data with the same parameter. The model was applied to Sungai Pengeli (Station B) and the performance was assessed using the values R-squared value. A strong relationship between observed and calculated discharge was detected. In order to forecast water level days ahead, the error prediction employs the Autoregressive Moving Average (ARMA) model which is one of the model of time series. From the analysis, the trend of change in water level of station Sungai Pengeli (Station B) is quite accurately captured by using PDM with to small differences of the water level between actual and forecast values.

ABSTRAK

Model Kebarangkalian Pengedaran (PDM) digunakan secara meluas dalam menganalisis tindak balas hidrologi. Salah satu aplikasi ini terlibat dalam pemodelan hujan lari untuk meramal berlakunya banjir di Johor, Malaysia. Dalam kajian ini, pengedaran Pareto digunakan untuk mewakili kapasiti penyimpanan PDM. Dengan menyediakan pelepasan simulasi dan pemerhatian yang baik, kaedah Algoritma Genetik (GA) telah digunakan untuk mengoptimumkan parameter optimum PDM. Pencapaian PDM dinilai melalui penentukuran dan pengesahan data yang berbeza dengan parameter yang sama. Ia boleh dilihat melalui stesen yang dipilih iaitu Sungai Pengeli (Stesen B) iaitu pencapaian terbaik ialah nilai R-kuasa dua yang tinggi. Oleh itu, hubungan yang kuat antara pemerhatian dan pengiraan pelepasan telah dikenalpasti. Untuk meramal paras air pada masa depan, ramalan kesilapan melalui model Purata Bergerak Autoregresi (ARMA) digunakan berdasarkan siri masa. Daripada analisis ini, perubahan dalam paras air stesen Sungai Pengeli (Stesen B) diperolehi dengan agak tepat dengan menggunakan PDM kerana perbezaan kecil paras air antara nilai sebenar dan ramalan.

TABLE OF CONTENTS

CHAPTER	ITEM	PAGE
	TITLE	i
	ABSTRACT	ii
	ABSTRAK	iii
	TABLE OF CONTENTS	iv
	LIST OF TABLES	x
	LIST OF FIGURES	xi
	LIST OF ABBREVIATIONS	xv
	LIST OF SYMBOLS	xvi
	LIST OF APPENDICES	xvii
1	INTRODUCTION	
	1.1 Research Background	1
	1.2 Problem Statement	2
	1.3 Objectives of the Study	3
	1.4 Scope of the Study	3
	1.5 Significance of the Study	4
	1.6 Organization of the Report	4

2	LITERATURE REVIEW	
	2.1 Introduction	6
	2.2 Basic Components of Cycle Hydrology	7
	2.3 Hydrologic Modeling	8
	2.4 Climate, Weather and Seasons of Malaysia	11
	2.5 Rainfall Characteristics	11
	2.6 Runoff Characteristics	12
	2.6.1 Rainfall abstraction	14
	2.6.2 Water Level-Discharge Data	15
	2.7 Calibration and Validation of Data	16
	2.8 Forecasting Technique	16
	2.9 Conclusion	18
	3 RESEARCH METHODOLOGY	
	3.0 Introduction	18
	3.1 Probability Distributed Model (PDM) Description	19
	3.2 Soil Moisture Storage	20
	3.3 Surface and Subsurface Storage	28
	3.4 Model Parameters	31
	3.5 Data Collection and Preparation	32
	3.5.1 The Rating Curve Parameters	32
	3.5.2 Lost Data: Mean Imputations	33

3.5.3	Parameter Optimization using Genetic Algorithm (GA) method	34
3.5.4	Calibration and Validation	37
3.6	Forecasting Process	40
3.6.1	Error Prediction	40
3.6.2	Box-Jenkins Method	42
3.6.2.1	Autoregressive Model, AR (p)	42
3.6.2.2	Moving Average Model, MA (q)	43
3.6.2.3	Autoregressive Moving Average Model (ARMA)	44
3.6.2.4	Analyzing of Time Series	45
3.6.2.5	Components of Time Series	45
3.6.2.6	Model Identification	45
3.6.2.6.1	Non Stationary Series	46
3.6.2.6.2	Seasonal Data	47
3.6.2.6.3	Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)	48
3.6.2.7	Model Estimation	48
3.6.2.8	Forecasting Future Values	50

RESULTS AND DISCUSSION

4	PARAMETER OPTIMISATION	
4.1	Introduction	51
4.2	Site description	52
	4.2.1 Hydrological Stations	54
	4.2.2 Flood History	56
4.3	Descriptive of Data	57
	4.3.1 Rainfall Data	57
	4.3.2 Water Level Data	60
4.4	Lost Data: Mean Imputations	61
4.5	The Rating Curve Parameter	62
4.6	The Correlation Between The Rainfall and Discharge Data	66
	4.6.1 The correlation between the Rainfall and Discharge	68
4.7	PDM calibration and validation	68
	4.7.1 The PDM calibration	78
	4.7.2 PDM validation	81
	4.7.3 Performance Evaluations	81
	4.7.3.1 Coefficient of correlation and Nash-Sutcliffe	83

	Coefficient	
	4.7.3.2 Root Mean Square Error	84
4.8	Conclusion	84
5	FORECASTING PROCESS	
5.1	Introduction	85
5.2	The sources of data	86
5.3	Data Analysis using Box-Method method	86
	5.3.1 Station A	86
	5.3.1.1 Model Identification	87
	5.3.1.2 Parameter Estimation Process	93
	5.3.1.3 Forecasting process	94
	5.3.2 Station B	96
	5.3.2.1 Model Identification	97
	5.3.2.2 Parameter Estimation Process	103
	5.3.2.3 Forecasting process	104
5.4	Performances of Water Level	105
	5.4.1 Station A	109
	5.4.2 Station B	109
5.5	Conclusion	111

6	CONCLUSION AND RECOMENDATION	
6.1	Introduction	111
6.2	Summary and Conclusion	112
6.3	Recommendation	113
	REFERENCES	113-115
	APPENDICES	
	Appendix A	116
	Appendix B	118
	Appendix C	119

LIST OF TABLES

TABLE NO	TITLE	PAGE
3.1	Parameters of Probability Distributed Model	35
4.1	The Data Used in Every Station	38
4.2	Descriptive Statistics of Rainfall Data	57
4.3	The Maximum Rainfall in Station Rancangan Ulu Sebul	61
4.4(a)	Evapotranspiration Data with some Missing Values	61
4.4(b)	Evapotranspiration data after simple mean imputation applied	62
4.5(a)	First Step	63
4.5(b)	Second step	64
4.6(a)	The optimized parameters in Station A	64
4.6(b)	The optimized parameters in Station B	81
4.7	Formulas for Coefficient of Correlation and Nash-Sutcliffe Coefficient	82
4.8	The accuracy of model performance	89
4.9	The accuracy of model performance between two stations	89
5.1	Final Estimate of Parameters	93
5.2	The Forecast Values of Water Levels Using Box-Jenkins Method	94
5.3	The Water Level Preferences	98
5.4	Final Estimate of Parameters	103

LIST OF FIGURES

FIGURE NO	TITLE	PAGE
3.1	The probability distributed model (PDM) model structure	19
3.2	Representation of runoff production by using basin behaviour	20
3.3	Representation of basin storage with different depts. With probability density function	21
3.4	Direct runoff generated from the stores	23
3.5	The PDM model structure	28
3.5	Genetic Algorithms	35
3.6	The general structure of Genetic Algorithm (Gerr and Cheng, 1997)	50
3.7	The Steps of Box-Jenkins Methodology	50
4.1	The Districts under state of Johor	52
4.2	Type of Major Soils in Johor (Department of Agriculture)	53
4.3	The physical map shows the selected hydrological stations in Johor	54
4.4	The flow path of Sungai Sayong with generated Geographic Information System (GIS)	56
4.5	Histogram Chart on Rainfall Data with Elapse Time in Station Rancangan Ulu Sebol	59
4.6(a)	Histogram Chart on Water Level of Station A	60
4.6(b)	Histogram Chart on Water Level of Station B	60

4.7(a)	The Relationship between Flow and Water Level in Station A	61
4.7(b)	The Relationship between Flow and Water Level in Station B	61
4.8(a)	The Correlation between rainfall and discharge in Station A	62
4.8(b)	The Correlation between rainfall and discharge in Station B	63
4.8(c)	The Correlation Graph between Rainfall and flow data of Station A and Station B	64
4.9(a)	The Relationship between Storage Capacity and Distribution Function for Station A	65
4.9(b)	The Relationship between Storage Capacity and Distribution Function for Station A	65
4.10(a)	Model Calibration before optimization procedure applied in Station A	66
4.10(b)	Model Calibration before optimization procedure applied in Station B	69
4.11(a)	Graph of cumulative discharge after the first initialization with the performance of model in Station A	70
4.11(b)	Probability distributed interacted with storage capacity for Station B	72
4.11(b)	Graph of cumulative discharge after the second initialization with the performance of model in Station A	72
4.11(b)	Graph of cumulative discharge after the third initialization with the performance of model in Station A	73
4.12(a)	Graph of cumulative discharge after the first initialization with the performance of model in	74

	Station B	
4.12(b)	Graph of cumulative discharge after the second initialization with the performance of model in Station B	74
4.12(c)	Graph of cumulative discharge after the third initialization with the performance of model in Station B	75
4.13(a)	Model Calibration with Parameter Optimization in Station A	76
4.13(b)	Model Calibration with Parameter Optimization in Station B	77
4.14(a)	Data validation in Station A	78
4.14(b)	Data validation in Station B	79
5.1	Time Series Plot for Residual of Discharge	87
5.2	The Trend Analysis for Residual of Discharge	88
5.3	Autocorrelation for η_t	89
5.4	Partial Autocorrelation for η_t	90
5.5	Trend Analysis Plot for $\nabla\eta_t$	91
5.6	Autocorrelation for $\nabla\eta_t$	92
5.7	Partial Autocorrelation for $\nabla\eta_t$	95
5.8	The Graph for Forecast Values of Water Levels in Station A	96
5.10	Time Series Plot for Residual of Discharge in Station B	97
5.11	The Trend Analysis for Residuals of Discharge	100
5.12	Autocorrelation for η_t	101
5.13	Partial Autocorrelation for η_t	102
5.14	Trend Analysis Plot for $\nabla\eta_t$	105
5.15	Autocorrelation for η_t	106
5.16	Partial Autocorrelation for $\nabla\eta_t$	106
5.17(a)	The Graph for Forecast Values of Water Levels	107

	on 2 November 2013	
5.17(b)	The Graph for Forecast Values of Water Levels on 4 November 2013	107
5.17(c)	The Graph for Forecast Values of Water Levels on 5 November 2013	107
5.17(d)	The Graph for Forecast Values of water levels on 18 November 2013	108
5.17(e)	The Graph for Forecast Values of Water Levels on 30 November 2013	108
5.17(f)	The Graph for Forecast Values of Water Levels on 1 December 2013	108
5.17(g)	The Graph for Forecast Values of Water Levels on 13 December 2013	109
5.17(h)	The Graph for Forecast Values of Water Levels on 15 December 2013	110
5.18(a)	The Actual and Forecast Water Level in Station A	110
5.18(b)	The Actual and Forecast Water Levels in Station B	110

LIST OF ABBREVIATIONS

PDM	- Probability Distributed Model
ACF	- Autocorrelation Function
AR	- Autoregressive
ARMA	- Autoregressive Moving Average

LIST OF SYMBOLS

f_c	- Rainfall Factor,
b_e	- Exponent in actual evaporation function,
c_{\min}	- Exponent in actual evaporation function,
c_{\max}	- Maximum store capacity,
b	- Exponent parameter,
k_g	- Groundwater recharges time constant,
b_g	- Exponent of recharge function,
s_t	- Soil tension storage capacity,
k_1	- Time constants of cascade of first linear reservoir,
k_2	- Time constants of cascade of second linear reservoir,
k_b	- Base flow time constant,
q_c	- Constant flow representing returns or abstraction,

LIST OF APPENDICES

APPENDICES	TITLE	PAGE
A	Output of PDM parameters in Station A	116
	Output of PDM parameters in Station B	
B	Output of Final Result of Calibration data for Station A	118
	Output of Final Result of Calibration data for Station B	
C	Comparison between Actual and Forecast Water Level of Station A	119
	Comparison between Actual and Forecast Water Level of Station B	120

CHAPTER 1

INTRODUCTION

1.1 Research Background

Malaysia is one of the countries that is located in humid tropical region and considered as a rapidly developing country. The development of the agricultural and industrialization contributes to natural disturbance such as flood and hydrological process. According to Chan (1996), flood risks have increased since 1985 due to development of urban river corridors. When the precipitation has reached into the earth and when rainfall does not fully infiltrate the soil, thus it makes way to other surface systems such as surface storage and subsurface storage. Water begins to fill the depression storage on the soil surface. Once the rate of precipitation exceeds the infiltration, water starts to flow over to the land and it moves laterally towards the stream, which is called rainfall runoff.

There are many rainfall runoff models that are widely used such as TOPMODEL, TOPKAPI, Horton Overland Flow Model and Hewlett Runoff Model (Keith J. Beven, 1997). However, this study focused on Probability Distributed Model (PDM) that has a main role in analyzing the hydrological behavior. PDM has been widely used in many countries such as England, Hong Kong, India and others (Young and Reynard, 2004; Blyth and Bell, 2004; Moore, 2007). This model has been implemented in numerous processes, such as runoff production, subsurface storage, groundwater recharge and others. PDM was discovered in the 1960s by researchers. According to Moore and Clarke (1981), PDM has no actual estimation between the distribution of storage and characteristics of catchment. Then, this research was continued by Hosing and Clark (1990), Moore (1985) and Clarke (1983). PDM model is not only targeted on the surface runoff but also applied on snowmelt model, macro scale hydrological model and others.

This research focused only on Johor River, Malaysia. According to Department of Irrigation and Drainage (DID), Johor, major floods occurred over the past decades and also most recently in December 2006 and January 2007. In addition, it caused massive flooding and sometimes destroyed properties and killing people at each year.

1.2 Problem Statement

The factors that are considered in flood occurrence in Johor are topography, topographic slope, soil, land use, climate change, lithology, drainage and human factors. These factors were determined by previous researchers (World Meteorological Organization, 2008). The management of poor dams and the storage capacity has to be considered so that we can reduce the frequency of having floods in selected area. Although there are many models discovered in flood prediction, the problem still remains. Therefore, Probability Distributed Model (PDM) would be used for Johor

Bharu River so that the factors can be minimized or overcome to reduce the flood occurrence problems.

1.3 Objectives of the Study

The objectives of this study are:

- i. To apply PDM model with Pareto distribution to represent the storage capacity to Johor river basin.
- ii. To assess the performance of the model.
- iii. To validate the model by predicting the future water level

1.4 Scope of the Study

By using Lazarus software, the Probability Distributed Model (PDM) software is utilized in analyzing the hydrological data. In this study, the data sets consist of inputs and outputs needed in analyzing the hydrological data. Therefore, a part of watershed located in Johor, Malaysia has been selected in this research. Hydrological data and digital topographic data were acquired from Department of Irrigation and Drainage (DID), Hydrological Unit and Department of Survey and Mapping Malaysia (JUPEM).

1.5 Significance of the Study

The purpose of the study is to show on how Probability Distributed Model (PDM) is used in solving the problems of flood by applying some PDM techniques to the problems. In this study, we used Probability Distributed Model (PDM) to get a better and precise model in forecasting the frequency of flooding in Johor Bharu areas and also give important contributions to the hydrologist. PDM model is different than other models because the storage capacity is assumed to be distributed using the probability distribution. PDM model is also well-structured compared to the concepts of other models (Willems et al., 2001). In forecasting the process, PDM model would be updated by using state correction and error prediction techniques either in offline forecasting or real time forecasting. Besides that, PDM model also represents a broad range of runoff behavior by using a minimum number of model parameters.

1.6 Organization of the Report

This report consists of six chapters. Chapter 1 provides an introduction to the research study as an overview of the current research. It consists of the background of the study, problem statement, objectives, scope and significance of the study. Chapter 2 discusses the literature of the study. This chapter then continues by presenting historical researches of the rainfall characteristics, climate, weather and seasons of Malaysia, runoff characteristics, hydrologic process, hydrologic modeling, surface runoff, and forecasting technique. The research methodology is presented in Chapter 3. In this chapter, we will give some illustrations about how the Probability Distributed model (PDM) model behaviour is used. In this chapter also describes the research methodology about forecasting process. The data analyses are presented in Chapter 4 and Chapter 5, where the results are discussed in detail. Finally, Chapter 6 presents the summary, conclusions and recommendations for future study.

REFERENCES

- Beven, Keith J. (1997). *Rainfall-Runoff Modeling The primer*, United Kingdom, Wiley Publishers.
- D.Ward, Andy and W.Trimble, Stanley (2004). *Environmental Hydrology Second Edition*, United States of America, Lewis Publishers.
- Moore, R.J (2007). *The Probability Distributed Moisture rainfall-runoff model*, Centre For Ecology and Hydrology, Wallingford, Oxon, United Kingdom.
- Moore, R.J (1985). *The Probability-Distributed Principle and Runoff Production at Point and Basin Scales*, Hydrological Sciences Journal, 273-297.
- Leong Weng Chin (2007). *Study of Malaysian Urban Rainfall-Runoff characteristics: Case study of Sungai Kayu Ara, Damansara, Selangor*. Doctor Philosophy, Universiti Teknologi Malaysia, Skudai.
- Muhammad Muneer Anoum and Amir Hassan Mohd Kassim (2000). *Verification of Runoff Coefficient of Local Malaysian Urban Catchment*, Universiti Teknologi Malaysia, Skudai.
- Beven, Keith J. (2001). *Rainfall-Runoff Modeling – The Primer*, John Wiley and Sons Ltd, New York.

- Moore, R.J and Weiss, G, (1980). *Recursive parameter estimation of a non-linear flow forecasting model using the extended Kalman Filter Real-time hydrological forecasting and control*, Institute of Hydrology, Wallingford, United Kingdom.
- Bowerman B.L., O'Connell R.T and Koehler A.B. (2005). *Forecasting, Time Series and Regression*. (4th Edition), United of States.
- J.H. Holland (1975). *Adaptive in Natural and Artificial Systems*. Ann Arbor University of Michigan Press.
- Moore, R.J (1992). *PDM: A generalized rainfall-runoff model for real-time*, National Rivers Authority River Flow Forecasting system, United Kingdom.
- Moore, R.J., (1983). *Flood Forecasting Technique*, Bangkok, Thailand
- Moore, R.J and Clarke, R.T., (1981). *A Distribution Function Approach to Rainfall-Runoff Modeling*, 1367-1382.
- Zhao, R.J., Zhuang, Y, Fang L.R., Lin, X.R. and Zhang Q.S., (1980). *The Xinanjiang Model in Hydrological Forecasting*, IAHS Publ.no. 129, 351-356.
- Pieter Cabus (2008). *River Flow Prediction through Rainfall–Runoff Modeling with a Probability-Distributed Model (PDM) in Flanders, Belgium*, Belgium
- Masoud Bakhtyari Kia, Saied Pirasteh, Biswajeet Pradhan, Ahmad Rodzi Mahmud, Wan Nor Azmin Sulaiman, Abbas Moradi (2012). *An artificial neural network model for flood simulation using GIs Johor River Basin*, Malaysia
- Kuok King Kuok1, Sobri Harun, Pou Chan Chiu (2011). *Auto-Calibration of Daily and Hourly Tank Model's Parameters using Genetic Algorithm*, Universiti Teknologi Malaysia.

Jia Liu and Dawei Han (2013). *Impact of Data Time Interval on Real-Time Hydrological Forecasting*, Water and Environmental Management Research Centre, University of Bristol, UK

Kazuaki Hiramatsu and Masayoshi Harada (2012). *Optimizing Parameters for Two Conceptual Hydrological Models Using a Genetic Algorithm: A Case Study in the Dau Tieng River Watershed, Vietnam*

J. Liu and D. Han (2008). *On selection of the optimal data time interval for real-time Hydrological forecasting*, University of Bristol, China

Danielle R. Tripp, Jeffrey D. Niemann (2008). *Evaluating the Parameter Identifiability And Structural Validity of a Probability-Distributed Model for Soil Moisture*, Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, United of States