

UNIVERSITI TEKNOLOGI MARA

**THE ESTIMATION ON MISSING
DAILY RAINFALL DATA AT
MACHANG AND AMPANG STATION
USING GENERALIZED LINEAR
MODEL**

NURUL AISHAH BINTI RAHMAN

Thesis submitted in fulfillment
of the requirements for the degree of
Master of Science
(Statistics)

Faculty of Computer and Mathematical Sciences

September 2019

ABSTRACT

A complete rainfall data plays an important role in climatological, hydrological and meteorological studies. However, the rainfall data is not always available due to several reasons such as malfunction of the instrument and power failure in recording the rainfall data. The presence of missing data could affect the continuity and consistency of the rainfall data. Thus, this study aims to estimate the missing daily rainfall data using the existing selected imputation methods, simple arithmetic average, normal ratio method, inverse distance weighting method, correlation coefficient weighting method and geographical coordinate. Despite of their popularity and simplicity, rainfall pattern is scarcely considered when estimating the missing data. Generally, the rainfall data in Malaysia have a unimodal or bimodal pattern which could be summarized by fitting a smooth curve to the estimated values obtained using generalized linear model (GLM). This model is appropriate to be used in evaluating the estimated values obtained by the imputation methods as mentioned above. In order to assess the performance of the improved estimated values obtained through GLM, the distance between target station and neighbouring stations also the rainy and non-rainy period with respect to different level of missingness are considered. In this study, the rainfall stations that located within 25 km and 35 km from target station are considered as neighbouring stations. Historical daily rainfall data from the periods of 1975 to 2014 at Kelantan and Selangor region are used for the analysis. The performance of the imputation methods is quantified using mean absolute error, root mean squared error and similarity index. The results indicated that there is improvement in estimating the missing daily rainfall data since GLM provide a useful information for describing the rainfall patterns. A further investigation also revealed that the neighbouring stations located within 25 km are enough to be considered in producing sufficient estimation. In addition, during rainy period and non-rainy period, the imputation method with consideration to GLM consistently well performed for the distance of 25 km at Ampang station. Thus, it can be concluded that GLM has proven to be selected as the proposed methods in providing better estimation results for missing daily rainfall data treatment.

ACKNOWLEDGEMENT

Firstly, I wish to thank Allah for giving me the opportunity to embark on my Master and for completing this long and challenging journey successfully. I would like to express my gratitude to my supervisor, Assoc. Prof. Dr. Sayang Mohd Deni and my co-supervisor, Dr. Norazan Mohamed Ramli for the continuous support, valuable comments and ideas when assisting me in all time of research and thesis writing.

Furthermore, I also would like to express my appreciation to the staff of Department of Drainage, Ampang for providing the data, knowledge and assistance. Also, thanks to the sponsorship from Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Ministry of Higher Education and Institute of Research Management and Innovation (IRMI), UiTM, 600-RMI/FRGS 5/3(58/2014). Special thanks to my friends especially, Siti Nur Zahrah Amin Burhanuddin for the simulating discussions and all the fun we have had in the last four years.

Finally, I dedicate this piece of victory to my family especially my parents, Rahman Md Zan and [REDACTED] for providing me with continuous support and encouragement throughout my life. I am grateful forever for your love. Alhamdulillah.

TABLE OF CONTENTS

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	x
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xiv
CHAPTER ONE: INTRODUCTION	1
1.1 Overview	1
1.2 Problem Statement	5
1.3 Research Questions	6
1.4 Research Objectives	7
1.5 Scope and Limitations	7
1.6 Significance of Study	8
1.7 Thesis Outline	8
CHAPTER TWO: LITERATURE REVIEW	9
2.1 Introduction	9
2.2 Missing Data	9
2.3 Possible Reasons of Missingness in Rainfall Data	10
2.3.1 Malfunction of Gauging Devices	11
2.3.2 Power Failure	11
2.3.3 Meteorological Extreme	11
2.3.4 Absence of Observer	12
2.3.5 Improper Data Registration	12
2.3.6 Relocation of the Rainfall Stations	12

2.3.7	Changes of Rainfall Stations Surrounding	13
2.3.8	Other Possible Reasons of Missingness	13
2.4	The Impact of Missingness in Rainfall Data	13
2.4.1	Lack of Rainfall Data in Performing Hydrological Modelling	14
2.4.2	Loss of Relevant Information on the Rainfall Data	14
2.4.3	Prevent Important Analysis on the Rainfall Data	14
2.4.4	Increase Level of Complexity of the Analysis	15
2.4.5	Lack of Information for Planning and Management of Water Resources	15
2.4.6	Other Impacts of Missingness in Rainfall Data	15
2.5	Estimation of Missing Rainfall Data	16
2.5.1	Imputation Methods	16
2.5.2	Artificial Intelligence and Machine Learning Techniques	20
2.5.3	Geostatistical Methods	22
2.5.4	Related Studies on the Estimation of Missing Rainfall Data in Malaysia	23
2.6	Stochastic Rainfall Modelling	24
2.6.1	Neyman-Scott Rectangular Pulse Model (NSRP)	25
2.6.2	Bartlett Lewis Rectangular Pulse Model (BLRP)	26
2.6.3	Generalized Linear Model (GLM)	27
2.6.4	Other Stochastic Rainfall Modelling	31
2.7	Concluding Remark	33
 CHAPTER THREE: RESEARCH METHODOLOGY		34
3.1	Introduction	34
3.2	Scope and Data Description	34
3.3	Rainfall Measurement	38
3.4	Imputation Methods for Estimating Missing Rainfall Data	43
3.4.1	Simple Arithmetic Average (SAA)	45
3.4.2	Normal Ratio Method (NRM)	45
3.4.3	Inverse Distance Weighting Method (IDW)	46
3.4.4	Correlation Coefficient Weighting Method (CCW)	46
3.4.5	Geographical Coordinate (GC)	47
3.5	Generalized Linear Model (GLM)	47