

UNIVERSITI TEKNOLOGI MARA

**THE ENHANCEMENT OF NORMAL
RATIO METHOD THROUGH
MULTIPLE IMPUTATION
APPROACH IN ESTIMATING
MISSING DATA WITH OUTLIERS
FOR PENINSULAR MALAYSIAN
RAINFALL DATASET**

**SITI NUR ZAHRAH BT AMIN
BURHANUDDIN**

PhD

October 2020

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.


Name of Student : Siti Nur Zahrah Bt Amin Burhanuddin

Student I.D. No. : 2012678162

Programme : Doctor of Philosophy in Information Technology and
Quantitative Sciences – CS990

Faculty : Computer and Mathematical Sciences

Thesis Title : Enhancement of Normal Ratio Method through
Multiple Imputation Approach in Estimating Missing
Data with Outliers for Peninsular Malaysian Rainfall
Dataset

Signature of Student : 

Date : October 2020

ABSTRACT

The complete rainfall dataset is very important in representing the climatological characteristics precisely, especially for hydrological and meteorological studies. It is also contributed to effective and efficient environmental management. However, the rainfall data is highly vulnerable to the missing problem due to the dynamic process of the climatic variable. Furthermore, the data is exposed to the seasonal activities that could contribute to the uncertainty and irregularity variations in the rainfall amount which will cause the presence of outliers in the dataset. These situations will affect the quality of the rainfall dataset and subsequently provide inaccurate information to the users. Concerning this situation, this study attempts to develop a practical and reliable approach to treat the missing values in the effort to provide a good quality dataset for the public domain. Spatial estimation method, i.e. normal ratio method was considered in this study to estimate the missing rainfall data. Various efforts were proposed to improve the performance of the method, however, there are lacking works on robustifying the method so that it can perform well for the dataset that contains outliers. Therefore, this study aims to propose the enhancement of normal ratio methods for imputing the missing values in the daily rainfall dataset with outliers. The robust statistics (i.e. trimmed mean, median, and geometric median) were adopted in the proposed methods to make them less affected by the outliers. The normal ratio method was commonly implemented through single imputation approach, but this approach encounters with the limitation of not considering uncertainty in missing values. Thus, this study has proposed a multiple imputation approach based on block bootstrap to overcome the limitation of single imputation approach as well as improving the performance of the existing multiple imputation approach incorporated in Amelia package. Block bootstrap was firstly introduced in the proposed multiple imputation approach (named as NRMI-Bboot) to enhance the performance when dealing with the rainfall time series. The performance of each estimation method was evaluated based on five performance criteria at six different levels of missing data (5%, 10%, 15%, 20%, 25%, and 30%) and three levels of outlying data (5%, 10%, and 15%) that have been created in the dataset. Complete 40 years daily rainfall data from 22 meteorology stations were considered for the analysis purpose. Four target stations were selected as the representative of the main regions in Peninsular Malaysia (northwest, east, west, and southwest). The capability of the estimation methods was further verified using distribution fitting. The adoption of the robust statistics in the proposed estimation methods associated with the NRMI-Bboot approach has provided an improvement to the estimation results, especially when dealing with the dataset that contains extreme outliers. The block bootstrap ensured that the original rainfall time series structure was preserved within each monsoon block and consequently produced more accurate estimation results. This indicates the advantages of the proposed estimation methods and multiple imputation approach in their role of providing accurate imputed values for missingness in Peninsular Malaysian daily rainfall dataset.

ACKNOWLEDGEMENT

Firstly, I wish to thank Allah for giving me the opportunity to embark on my study and for completing this journey successfully. My gratitude and very special thanks go to my dearest supervisor Assoc. Prof. Dr. Sayang Mohd Deni, and co-supervisor, Dr. Norshahida Shaadan and Dr. Norazan Mohamed Ramli. Thank you for the guidance and full support in assisting me with the study and for contributing the brilliant ideas to produce such valuable research.

Furthermore, I would like to express my appreciation to the Ministry of Higher Education for the scholarship during my study and to Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) for providing the facilities and assistance. Also, thanks to the staff of the Department of Irrigation and Drainage, Ampang for providing all the necessary data and knowledge.

Special thanks to my best friend, Noridayu Binti Mah Hashim. Thank you for the moral support and ups and downs experiences that we shared since we know each other. Also, thanks to my partner Nurul Aishah Binti Rahman and Rosadibah Mohd Towel who have accompanied and supported me in completing this journey.

Moreover, I also would like to express my thanks to all lecturers and colleagues, Dr S. Sarifah Radiah, Dr Yantizpura, Dr Zuraida, Dr Norizam, Norsalwani, Syafikah, Asmui', Zaid, Zamri, Amira, Hamzah, Muqhlisah, Najihah, Aisyah Sim, and Syeilla for helping me during the study. Also, thanks to my friends Khairun, Mazaty, Mazita, Al-Amin, Azira and all persons who have been directly and indirectly involved in completing my PhD journey.

The warmest thanks to my dearest sister Siti Nur Faizah and brothers, Md Farid and Mohamad Faiz for continuous financial and moral support. Finally, this thesis is specially dedicated to my dearest father, Amin Burhanuddin bin Hussin, and mother, Rohani Binti Mohamed for fully supporting and endlessly loving me during this long and challenging journey. This piece of victory is specially dedicated to both of you. Alhamdulillah.

TABLE OF CONTENT

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENT	vi
LIST OF TABLES	x
LIST OF FIGURES	xiv
LIST OF SYMBOLS	xvi
LIST OF ABBREVIATIONS	xviii
CHAPTER ONE: INTRODUCTION	1
1.1 Research Background	1
1.2 Research Gap	5
1.3 Problem Statement	6
1.4 Research Objectives	9
1.5 Significance of the Study	9
1.6 Scope and Limitation of the Study	10
1.7 Organization of the Thesis	12
CHAPTER TWO: LITERATURE REVIEW	14
2.1 Introduction	14
2.2 Rainfall Data Measurement	14
2.3 Data Quality	18
2.4 Importance of Quality Data	18
2.5 The Existence of Outliers in Rainfall Data and Its Implications on Data Analysis	19
2.6 Implications of Missing Values on Rainfall Data Analysis	21
2.7 Missing Data Treatment Approach	22
2.7.1 The Discard Approach	23