

RAIN GAUGE NETWORK OPTIMIZATION IN A
TROPICAL AREA TOWARDS EFFICIENT
HYDROLOGICAL DATA ACQUISITION

MOHD ZAHARIFUDIN BIN MUHAMAD ALI

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

**RAIN GAUGE NETWORK OPTIMIZATION IN A
TROPICAL AREA TOWARDS EFFICIENT
HYDROLOGICAL DATA ACQUISITION**

MOHD ZAHARIFUDIN BIN MUHAMAD ALI

**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY (CIVIL ENGINEERING - WATER
RESOURCES ENGINEERING)**

**FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2019

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **MOHD ZAHARIFUDIN BIN MUHAMAD ALI**

Matric No: **KHA 120004**

Name of Degree: **DOCTOR OF PHILOSOPHY**

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

**“RAIN GAUGE NETWORK OPTIMIZATION IN A TROPICAL AREA
TOWARDS EFFICIENT HYDROLOGICAL DATA ACQUISITION”**

Field of Study: **CIVIL ENGINEERING (WATER RESOURCES**

ENGINEERING)

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature

Date:

Subscribed and solemnly declared before,

Witness’s Signature

Date:

Name:

Designation:

RAIN GAUGE NETWORK OPTIMIZATION IN A TROPICAL AREA TOWARDS EFFICIENT HYDROLOGICAL DATA ACQUISITION

ABSTRACT

An adequate and reliable rain gauge network is essential for observing rainfall data in hydrology and water resource applications. For this purpose, rain gauge stations are installed in the catchment area of the river that forms a rain gauge network. Normally, a rain gauge network will be developed in accordance with the hydrological purpose and evaluated to extend to the network size in order to increase data accuracy. The increasing number of rain gauge network and augmentation of the existing rain gauge networks without proper planning and design has resulted in the high density of rain gauge stations compared to the recommendation made by the World Meteorological Organization (WMO). This resulted in increasing maintenance costs, but at the same time, creates the possibility of redundancy of stations within the catchment area. Due to this factor, it is the aim of this study to review the rain gauge network in a specific catchment to establish the optimal number of stations so that efficient rainfall data acquisition can be obtained. Two new optimization approaches have been developed in this study for the rain gauge network optimization and to prioritize the rain gauge stations, first by coupling the cross-validation technique with the geostatistical method (CV-Geo), and second, using the modified Particle Swarm Optimization (MPSO) technique. The spatial interpolation error of the spatial rainfall distribution, measured as the Root Mean Square Error (E_{rms}) optimization criterion, is applied to a rain gauge network in a tropical urban area. The total daily rainfall data from the 55 rain gauge stations were used to perform the optimization process for seven flood events. The optimization aimed to reduce the number of rain gauge stations in the existing network that could be hypothetically redundant. By using the two new methods, CV-Geo and

MPSO, the number of stations in the existing rain gauge network could be optimized based on the lowest E_{rms} value of spatial interpolation error. The optimized rain gauge network exhibited a better semivariogram structure, especially in terms of nugget value that has been drastically improved. However, MPSO had shown a slightly better nugget value since it has recorded the lowest value of nugget. The rain gauge stations were prioritized based on their importance in the network. Four stations, namely T02, N03, N06, and N21 were considered ineffective and could, therefore, be relocated within the study area or eliminated from the existing network. A preliminary evaluation of the optimized network without the four stations showed satisfactory results in flood hydrograph simulation using a lump hydrologic model. Three out of four flood hydrograph simulations have yielded the NSE, r , and R^2 values more than 0.75, which have indicated that the optimized network is efficient enough to produce rainfall data to simulate a flood hydrograph. The optimized rain gauge network exhibited a better semi variogram structure and lowered spatial interpolation error.

Keywords: rain gauge network, cross-validation, geostatistical analysis, optimization, Particle Swarm Optimization.

PENGOPTIMUMAN RANGKAIAN STESEN HUJAN DI KAWASAN TROPIKA KE ARAH PENCERAPAN DATA HIDROLOGI YANG CEKAP

ABSTRAK

Rangkaian stesen hujan yang mencukupi dan boleh dimanfaatkan adalah penting untuk mencerap data hujan untuk kegunaan aplikasi hidrologi dan sumber air. Untuk tujuan ini, beberapa stesen hujan dipasang di kawasan tadahan sungai yang membentuk rangkaian stesen hujan. Lazimnya, rangkaian stesen hujan akan dibangunkan mengikut keperluan hidrologi tertentu dan dinilai untuk meluaskan saiz rangkaian bagi meningkatkan ketepatan data. Peningkatan saiz rangkaian stesen hujan dan peluasan rangkaian stesen hujan sedia ada tanpa perancangan dan reka bentuk yang betul telah mengakibatkan kepadatan stesen hujan yang tinggi berbanding dengan yang disyorkan oleh Pertubuhan Meteorologi Sedunia (WMO). Ini menyebabkan peningkatan kos penyelenggaraan, dan pada masa yang sama, mewujudkan kemungkinan berlakunya pertindanan stesen hujan di kawasan tadahan. Disebabkan oleh faktor ini, kajian ini bertujuan untuk mengkaji semula rangkaian stesen hujan di kawasan tadahan tertentu untuk mendapatkan bilangan stesen yang optimum ke arah pengumpulan data hujan yang cekap. Dua pendekatan pengoptimuman baharu dibangunkan dalam kajian ini untuk mengoptimumkan rangkaian stesen hujan dan menyenaraikan stesen hujan mengikut keutamaan, pertama dengan menggabungkan teknik *cross-validation* dengan kaedah *geostatistical* (CV-Geo) dan kedua menggunakan teknik *Particle Swarm Optimization* (PSO) terubahsuai (MPSO). Ralat sisipan ruang bagi taburan hujan ruangan, *Root Mean Square Error* (E_{rms}) diadaptasi sebagai kriteria pengoptimuman terhadap rangkaian stesen hujan di kawasan bandar tropika. Proses pengoptimuman dijalankan untuk tujuh kejadian banjir dengan menggunakan data hujan harian pada 55 stesen hujan. Tujuan pengoptimuman adalah untuk mengurangkan bilangan stesen hujan

yang secara hipotesis adalah bertindan di dalam rangkaian sedia ada. Dengan menggunakan dua kaedah baharu, CV-Geo dan MPSO, bilangan stesen hujan dalam rangkaian sedia ada dapat dioptimumkan berdasarkan nilai ralat sisipan ruangan, E_{rms} yang terendah. Hasil rangkaian stesen hujan yang dioptimumkan mempamerkan struktur semivariogram yang lebih baik, terutamanya dari segi nilai *nugget* yang telah ditambahbaik secara drastik. Walau bagaimanapun, MPSO telah menunjukkan nilai *nugget* sedikit lebih baik kerana ia mencatatkan nilai *nugget* terendah. Stesen hujan telah disusun berdasarkan kepentingan mereka dalam rangkaian. Empat stesen hujan, iaitu T02, N03, N06, dan N21 dianggap tidak berkesan dan boleh dipindahkan dari lokasi asal ke lokasi baru dalam kawasan kajian atau dihapuskan dari rangkaian sedia ada. Penilaian awal rangkaian yang dioptimumkan tanpa empat stesen hujan tersebut menunjukkan keputusan yang memuaskan dalam simulasi hidrograf banjir menggunakan model hidrologi. Tiga dari empat simulasi hidrograf banjir telah menghasilkan nilai NSE, r , dan R^2 lebih daripada 0.75, yang menunjukkan bahawa rangkaian yang dioptimumkan cukup memadai untuk menghasilkan data hujan bagi mensimulasikan hidrograf banjir. Rangkaian stesen hujan dioptimumkan mempamerkan struktur *semivariogram* yang lebih baik dan ralat sisipan ruangan yang terendah.

Keywords: rangkaian stesen hujan, *cross-validation*, analisa *geostatistical*, pengoptimuman, *Particle Swarm Optimization*.

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the Most Merciful

First and foremost, I am grateful to ALLAH, who has given me the strength, knowledge, and required understanding to complete this thesis. I would like to express my utmost appreciation to my supervisor, Associate Professor Dr Faridah Othman for her support, guidance, motivation and constructive criticisms throughout this study period. I would like to thank the Public Service Department of Malaysia and the Department of Drainage and Irrigation, Malaysia for the financial support during the study leave granted. I would also like to thank the University of Malaya for providing the Research Grants to support this study. I am most grateful and would like to thank those who have assisted, guided, and supported me which have led to in my studies leading to the completion of this thesis. Finally, I would like to extend my deepest gratitude to my parents, my wife and daughters, who have always given me unremitting support during the study period.

TABLE OF CONTENTS

Abstract	iii
Abstrak	v
Acknowledgements	vii
Table of Contents	viii
List of Figures	xii
List of Tables.....	xv
List of Symbols and Abbreviations.....	xvii
List of Appendices	xix
CHAPTER 1: INTRODUCTION.....	1
1.1 Background.....	1
1.2 Problem statement	4
1.3 Aim and Objectives of the Study.....	7
1.4 Significance of the Study.....	8
1.5 Thesis Structure	9
CHAPTER 2: LITERATURE REVIEW.....	11
2.1 Introduction.....	11
2.2 Rain Gauge Network	12
2.2.1 Set up a rain gauge network	14
2.3 Spatial Rainfall Interpolation.....	16
2.3.1 Thiessen-polygon Method.....	16
2.3.2 Inverse Distance Weighting Method	17
2.3.3 Geostatistical method	18
2.3.3.1 Spatial Interpolation	21

2.4	Application in Rain Gauge Network Optimization	22
2.4.1	Statistical Method.....	22
2.4.2	Probabilistic Method	22
2.4.3	Geostatistical approach.....	24
2.4.4	Artificial Intelligence and a combination of methods	26
2.4.5	Rainfall Data for Optimization.....	27
2.5	Particle Swarm Optimization (PSO).....	29
2.5.1	PSO Application in Rain Gauge Network Optimization.....	34
2.6	Conclusion Remark	35
 CHAPTER 3: METHODOLOGY		37
3.1	Introduction.....	37
3.2	Description of Study Area	37
3.3	Overview of Methodology.....	40
3.4	Input Data	42
3.4.1	Rainfall dataset and flood events.....	42
3.4.2	Discharge Data Sets.....	47
3.5	Preliminary Data Analysis	48
3.5.1	Rainfall Data Analysis.....	48
3.5.2	Discharge Data Analysis	49
3.6	Methodology for the Selection of Variogram Model	49
3.6.1	Geostatistical Method.....	50
3.6.1.1	Ordinary Kriging and Semivariogram Model	51
3.6.1.2	Indicators of Spatial Rainfall Distribution Interpolation.....	53
3.6.2	Analytical Hierarchy Procedure (AHP).....	55
3.6.2.1	Priority Weight of AHP's Criteria	55
3.6.2.2	Semivariogram Model Evaluation	59

3.7	Methodology for Rain Gauge Network Optimization	60
3.7.1	Coupling Cross Validation and Geostatistical (CV-Geo)	61
3.7.1.1	Leave-One-Out and Add-One-In Cross Validation Technique.	61
3.7.1.2	Geostatistical and Spatial Rainfall Interpolation.....	62
3.7.2	Modified Particle Swarm Optimization (MPSO).....	63
3.7.2.1	Modification on PSO Equation	64
3.7.2.2	Evaluation process.....	66
3.7.3	Application of Proposed Optimization Methods.....	68
3.7.3.1	Application of CV-Geo Method.....	69
3.7.3.2	Application of MPSO Method	71
3.7.3.3	Objective Function of optimization.....	74
3.8	Verification of Optimized Rain Gauge Network by Hydrological Tank Model ...	74
3.8.1	Tank Model Development.....	77
3.8.1.1	Tank Model Set Up	77
3.8.1.2	Calibration and Validation of TM-UKRB Parameters.....	80
3.8.2	Application of TM-UKRB for optimized rain gauge network verification	

CHAPTER 4: RESULTS AND DISCUSSIONS	84	
4.1	Introduction.....	84
4.2	The result of Preliminary Data Analysis	85
4.2.1	Descriptive Statistic and Preliminary Number of Stations.....	85
4.2.2	Rainfall-Elevation Correlation	88
4.2.3	Rainfall-Discharge Correlation	92
4.3	Selection of the Best Semivariogram Model	93
4.3.1	AHP's Priority Weight	93
4.3.2	Selected Semivariogram Model	94

4.3.3	Semivariogram of Existing Rain Gauge Network.....	100
4.4	The evaluation result of MPSO-IP	101
4.4.1	Controlled Value of Random Number (r1 and r2).....	101
4.4.2	Multi-run with Uncontrolled Value of Random Number.....	108
4.5	Optimization of Rain Gauge Network.....	113
4.5.1	Optimum Rain Gauge Network.....	113
4.5.1.1	Variography structure of the optimized network.	123
4.5.1.2	Redundant rain gauge station	126
4.6	TM-UKRB development	130
4.6.1	Calibration and Validation Result of TM-UKRB	130
4.6.2	Optimal Network Performance in Simulating Flood Events.....	134
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS.....		138
5.1	Introduction.....	138
5.2	The selection of the appropriate semivariogram model	138
5.3	The performance of MPSO.....	139
5.4	The rain gauge network optimization	140
5.5	Recommendations for future research	141
References		143
List of Publications and Papers Presented		151
Appendix A.....		152
Appendix B		154
Appendix C		156

LIST OF FIGURES

Figure 1.1: Framework of hydrological network analysis and redesign in line with the World Meteorological Organization (2008).....	2
Figure 2.1: Typical example of Thiessen Polygon.....	17
Figure 2.2: Typical example of Inverse Distance Weight method.....	18
Figure 2.3: Example of semivariogram.....	19
Figure 2.4: A typical procedure of the PSO algorithm	32
Figure 3.1: Location of the research area.....	39
Figure 3.2: Long-term mean monthly rainfall in the study area.....	39
Figure 3.3: Summary of research methodology.....	41
Figure 3.4: Location of the rain gauge stations mapped on DEM of the study area.....	45
Figure 3.5: AHP model for semivariogram model selection.....	56
Figure 3.6: Summary of the CV-Geo method employed to optimize the number of rain gauges in the network.....	70
Figure 3.7: Summary of the MPSO method employed to optimize the number of rain gauges in the network.....	73
Figure 3.8: Flow chart of Tank Model development and application for verification of optimized rain gauge network.....	76
Figure 3.9: Tank Model set up for TM-UKRB.....	79
Figure 4.1: Optimum number of rain gauge stations, N	87
Figure 4.2: Pearson correlation, r between rainfall and elevation for 3 February 2009 .	89
Figure 4.3: Pearson correlation, r between rainfall and elevation for 3 March 2009	89
Figure 4.4: Pearson correlation, r between rainfall and elevation for 18 September 2011	90
Figure 4.5: Pearson correlation, r between rainfall and elevation for 13 December 2011	90
Figure 4.6: Pearson correlation, r between rainfall and elevation for 7 March 2012	91

Figure 4.7: Pearson correlation, r between rainfall and elevation for 18 April 2012	91
Figure 4.8: Pearson correlation, r between rainfall and elevation for 21 August 2012 ..	92
Figure 4.9: Pearson correlation, r between rainfall and discharge.	93
Figure 4.10: Semivariogram generated by Spherical and Gaussian models, a) rainfall data on 3 March 2009 and b) rainfall data on 18 April 2012.....	98
Figure 4.11: Spatial rainfall map generated by, a) Spherical model for rainfall data on 3 March 2009, b) Gaussian model for rainfall data on 3 March 2009, c) Spherical model for rainfall data on 18 April 2012 and d) Gaussian model for rainfall data on 18 April 2012.....	99
Figure 4.12: Convergence graphs of TF_1 ($N=25$) for a) 1000 iterations b) 2000 iterations.....	104
Figure 4.13: Convergence graphs of TF_2 ($N=25$) for a) 1000 iterations b) 2000 iterations.....	105
Figure 4.14: Convergence graphs of TF_3 ($N=25$) for a) 1000 iterations b) 2000 iterations.....	105
Figure 4.15: Convergence graphs of TF_4 ($N=25$) for a) 1000 iterations b) 2000 iterations.....	105
Figure 4.16: Convergence graphs of TF_5 ($N=25$) for a) 1000 iterations b) 2000 iterations.....	106
Figure 4.17: Convergence graphs of TF_1 ($N=50$) for a) 1000 iterations b) 2000 iterations.....	106
Figure 4.18: Convergence graphs of TF_2 ($N=50$) for a) 1000 iterations b) 2000 iterations.....	106
Figure 4.19: Convergence graphs of TF_3 ($N=50$) for a) 1000 iterations b) 2000 iterations.....	107
Figure 4.20: Convergence graphs of TF_4 ($N=50$) for a) 1000 iterations b) 2000 iterations.....	107
Figure 4.21: Convergence graphs of TF_5 ($N=50$) for a) 1000 iterations b) 2000 iterations.....	107
Figure 4.22: Plotting of E_{rms} against the total number of stations in the network for the optimization result of CV-Geo (LOO and AOI) and MPSO.	118

Figure 4.23: Frequency rate (F_r) map of stations remaining in the optimal network: a) optimized by LOO, b) optimized by AOI and c) optimized by MPSO. 129

Figure 4.24: Hydrograph of calibration result using rainfall data on 18 September 2011. 132

Figure 4.25: Sensitivity analysis of Upper Sg Klang Tank Model parameter. 133

Figure 4.26: Comparison of the observed and simulated flood hydrograph of the optimal rain gauge network. 137

University of Malaya

LIST OF TABLES

Table 2.1: Statistical figures of the hydrological stations of the Drainage and Irrigation Department (DID, 2009).	13
Table 2.2: The minimum station's density recommended (area in km ² per station), (World Meteorological Organization, 2008).	15
Table 3.1: Events of the daily rainfall data used in this study.	45
Table 3.2: List of rain gauge stations in the existing network.	46
Table 3.3: Discharge dataset and period of simulation used in the hydrologic model. ..	48
Table 3.4 Mathematical equations of the variogram models	53
Table 3.5: Comparative weighting score for pairwise matrix comparison.	57
Table 3.6: Random Index for a number of criteria (Saaty, 1980)	59
Table 3.7: S_r value for each rank position	60
Table 3.8: Parameters value of SPSO-IP and MPSO-IP	66
Table 3.9: Testing Function parameters	67
Table 3.10: ID of candidate rain gauge network of a domain solution for networks that contain 10-station out of 25 stations	71
Table 3.11: TM-UKRB model's parameters	80
Table 3.12: Calibration and validation data	80
Table 3.13: The lower and upper limit of model parameters	82
Table 4.1: The descriptive statistical value of rainfall datasets.	87
Table 4.2: Results of priority weight and its associated comparative numerical value and Consistency Ratio of pairwise matrix comparison of the criteria versus objective	94
Table 4.3: Results of the variogram model ranking based on the Final Score value	95
Table 4.4: Results of variography parameter value from geostatistical analysis	97
Table 4.5: Value of variogram parameters of the existing network and their corresponding E_{rms} value for selected flood events	100

Table 4.6: Results of single runs with a controlled random number for 1000 iterations.	103
Table 4.7: Results of single runs with a controlled random number for 2000 iterations.	104
Table 4.8: Results of the best value extracted from 50 runs with an uncontrolled random number for 1000 iterations.	109
Table 4.9: Results of the best value extracted from 50 runs with an uncontrolled random number for 2000 iterations.	109
Table 4.10: Statistical t-test between MPSO-IP and SPSO-IP algorithms for 1000 and 2000 iterations and 25 swarm particle sizes at $\alpha=0.05$	111
Table 4.11: Summary of E_{rms} and E_{ma} of optimized rain gauge network by CV-Geo (LOO and AOI) and MPSO with selected subset stations.	120
Table 4.12: The summary result of the total number of stations in the optimum network and their density by CV-Geo (LOO and AOI) and MPSO.	121
Table 4.13: The comparison of variogram parameters values between the existing network and optimized network by CV-Geo (LOO and AOI) and MPSO.	125
Table 4.14: Frequency rate, F_r of subset stations chosen in the optimized rain gauge network.	128
Table 4.15: The calibration result of TM-UKRB	130
Table 4.16: The NSE value for validation process	131
Table 4.17: Sensitivity Index, SI of TM-UKRB Model Parameter	134
Table 4.18: Validation of the optimal rain gauge network	135

LIST OF SYMBOLS AND ABBREVIATIONS

AE	:	Absolute Error
AHP	:	Analytical Hierarchy Procedure
AOI	:	Add-One-In
<i>AW</i>	:	alternatives' priority weight
<i>AWP</i>	:	alternatives' weighted priority
<i>CI</i>	:	Consistency Index
<i>CR</i>	:	Consistency Ratio
<i>cr</i>	:	convergence rate
CV-Geo	:	Coupling Cross Validation and Geostatistical
DEM	:	Digital Elevation Model
<i>df</i>	:	degree of freedom
DID	:	Department of Drainage and Irrigation
Eas	:	Average Standard Error
<i>E_{ma}</i>	:	Mean Absolute Error
<i>E_{ms}</i>	:	Mean Standardized Error
<i>E_{rms}</i>	:	Root-Mean-Square Error
<i>E_{rmss}</i>	:	Root-Mean-Square Standardized Error
<i>E_{ss}</i>	:	sum squared error
<i>F_r</i>	:	frequency rate
GIS	:	Geographical Information System
ID	:	position of solution from 1 to 3 268 760
IDW	:	Inverse Distance Weight
IP	:	identification of the positive natural number
LOO	:	Leave-One-Out
LS	:	Least Squares Method

MPSO	:	Modified Particle Swarm Optimization
NN	:	National Hydrological Network
NSE	:	Nash-Sutcliffe efficiency index
NWP	:	Numerical Weather Prediction
OAT	:	one-at-a-time
OK	:	Ordinary Kriging
PSO	:	Particle Swarm Optimization
<i>PW</i>	:	priority weights
QMS	:	Quality Management System
RF	:	rainfall
<i>RI</i>	:	Random Index
SA	:	sensitivity analysis
SI	:	Sensitivity Index
SMART	:	Storm Water Management and Road Tunnel
SPSO	:	Standard Particle Swarm Optimization
TM-UKRB	:	Tank Model of Upper Klang River Basin
TN	:	<i>Infobanjir</i> Telematics' Network
TSP	:	travel sale person
UKRB	:	Klang River basin
WMO	:	World Meteorological Organization

LIST OF APPENDICES

Appendix A: LOO algorithm in Matlab code.....	152
Appendix B: AOI algorithm in Matlab code.....	154
Appendix C: MPSO algorithm in Matlab code.....	156

University of Malaya

CHAPTER 1: INTRODUCTION

1.1 Background

A rain gauge network is a hydrological network meant to collect rainfall data and facilitate hydrology applications, such as hydrologic modeling (Hongliang et al., 2015), flood forecasting (Kar, Lohani, Goel, & Roy, 2015), flash flood prediction (Volkman, Lyon, Gupta, & Troch, 2010) as well as water resource analysis. Rainfall data from a rain gauge network is used to compute spatial rainfall information in terms of point-based, areal average and spatial variability. Accurate rainfall data is crucial for hydrology and water resource-related projects at the planning, design and operational levels (Adhikary, Yilmaz, & Muttil, 2015).

The development of a rain gauge network is an evolutionary process, beginning with the initial development of a basic network, followed by periodic reviews for upgrading to achieve an optimum network (Vivekanandan & Jagtap, 2013b). River basin managers around the world adopt this process in rain gauge network design and optimization. A rain gauge network is reviewed to optimize the appropriate number of point rainfall stations, as studied by Bastin, Lorent, Duque, and Gevers (1984) and Pardo-Igu'zquiza (1998). The review process can adopt the procedure suggested by the World Meteorological Organization (WMO) as illustrated in Figure 1.1.

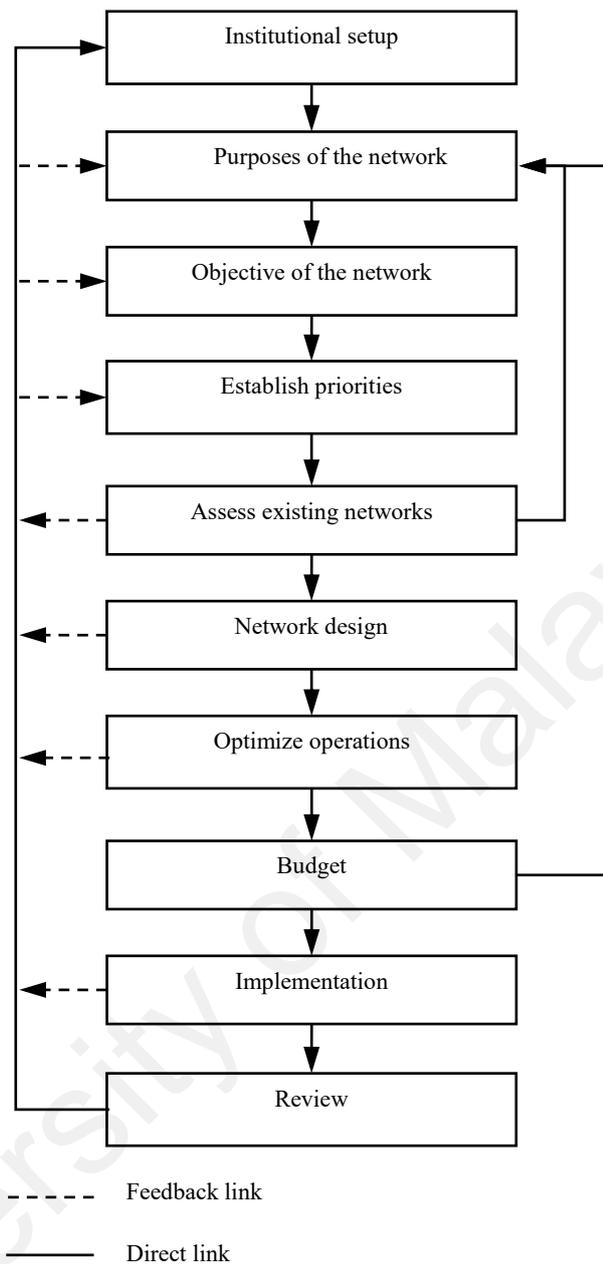


Figure 1.1: Framework of hydrological network analysis and redesign in line with the World Meteorological Organization (2008).

The framework illustrates the process to review an existing rain gauge network based on the purpose and objective of a rain gauge network. This will determine the priority of each rain gauge. Then, a network will be designed and optimized for operations, limited by the budget constraint. At this stage, the budget also influences the purpose of the analysis and redesign of a network. Furthermore, the budget significantly depends on the economic condition in term of global and micro-scale levels. The policy and vision

of an organization concerning the review process also play a primary role in influencing the finance. However, the redesign and optimization stages are the technical part in which there are widely explored by researchers to contribute in the review process (Adhikary et al., 2015; Al-Abadi & Al-Aboodi, 2014; Bakhtiari, Kermani, & Bordbar, 2013; Kar et al., 2015; Leach, Kornelsen, Samuel, & Coulibaly, 2015; Mishra & Coulibaly, 2014; Shaghaghian & Abedini, 2013)

In Malaysia, the development of the rain gauge network began in 1972, initiated by the Department of Drainage and Irrigation (DID). The initial network is called National Hydrological Network that comprised 86 principal rain gauge stations and 647 secondary rainfall stations (Jabatan Pengairan dan Saliran, 2009). The principal stations are defined as permanent or fixed stations and equipped with self-recording instrument. Meanwhile, the secondary stations are similar to principal station but they are installed for short-term or project basis which is subjected to review after certain period, normally for 5 to 10 years. The network was expanded each year, and by 1985, the densities of the rain gauge stations established by DID had met the requirements as recommended by the World Meteorological Organization (WMO). To date, with the increasing need of hydrological analysis for sustainable water resources planning, development, and management, the National Hydrological Network was expanding to collect more data for analysis. Currently, DID maintain more than 1000 hydrological stations throughout the country for hydrological data collection for water resources, flood and drought monitoring (Jabatan Pengairan dan Saliran, 2017).

A common but effective way to represent spatial rainfall data is by using a spatial rainfall map. The map generated using rainfall data from the rain gauge networks involves point-based rainfall that is used to compute the average areal rainfall for spatial rainfall mapping. The map is produced through an interpolation process using point

rainfall values. Accurate rainfall maps are crucial for any hydrology application. Thus, the level of spatial interpolation error determined the accuracy of the map. The ability to reduce the error will ensure the goodness of result in the hydrological analysis, and it is an excellent opportunity to explore.

A rain gauge network with a high number of stations could not guarantee the quality of the spatial rainfall map because it immeasurably depends on how the stations are distributed within the area. In many cases, the rain gauge stations were installed without proper design especially the stations that had been installed at an early stage. When the network is expanded to procure more rainfall data, the redundancy of the station could occur. In addition, the climate change factor that affects the entire world influences the pattern of rainfall distribution in terms of spatial and temporal scale. These factors motivate the need for a review process on the existing rain gauge network.

1.2 Problem statement

In operational hydrology, rainfall data presented as spatial rainfall map are generated via spatial interpolation process. Through the spatial interpolation method, point rainfall data are converted into spatial rainfall data for hydrological analysis such as to check water availability within a catchment and to give early flood warning information to the public via a flood forecasting model or system.

The spatial rainfall data are commonly measured as a catchment or areal rainfall (Bastin et al., 1984; Bras & Colon, 1978; Bras & Rodríguez-Iturbe, 1976; Pardo-Igu'zquiza, 1998). The catchment rainfall is calculated using rainfall data from several stations within a catchment area. Therefore, the accuracy of the catchment rainfall significantly depends on the adequacy and reliability of the rainfall data measured from every station in the station network. Furthermore, the adequacy and reliability of the

rainfall data depend on the appropriate location of the station to gauge the rainfall that reaches the ground (Bastin et al., 1984; Cheng, Lin, & Liou, 2008).

In conventional practice, the location of the rain gauge station depends on a few factors such as the cost of development, access to the site, and maintenance aspect. In addition, Shaghaghian and Abedini (2013) have listed the factors that influence the rain gauge network design which are the overall objective of the network design, the process considered, the attribute under consideration, the temporal scale or sampling interval in, the spatial scale, the topographic setting, types of precipitation, the nature of the objective function used for design and the algorithm used for minimization of maximization the objective function value. These factors are relatively subjective from one basin to another. On the other hand, it is recommended that factors like the elevation effect (Feki, Slimani, & Cudennec, 2012; Goovaerts, 2000), spatial distribution of rainfall (Xu, Xua, Chen, Zhang, & Li, 2013), and purpose of the station to be developed are to be taken into consideration. However, to incorporate these factors in the design, the location and rain gauge network are hard to implement. Instead, the instinct of personal experience takes over this task.

The conventional use of the geostatistical method for producing spatial rainfall distribution has been enhanced by the advanced development of the Geographical Information System (GIS). Today, the GIS applications have been facilitated by the built-in interpolation methods in geostatistical analysis. This has offered an advantage in analyzing spatial rainfall data. The Kriging technique variant is one of the example methods which require a variogram model to compute estimated data at the interpolation point.

The Kriging technique is quite similar to the Inverse Distance Weight (IDW) method that needs the weight of the observed data to estimate the value at a certain location.

The IDW derives the weight based on the distance of the interpolated point to the observed data point only. However, the Kriging incorporated the distance with the spatial correlation of the observed data arrangement. The spatial correlation is evaluated using the variogram model before the estimation at an unobserved point can be done. Furthermore, several variogram models are available in the ArcGIS tools for use in spatial interpolation using the Kriging technique. Gaussian, Exponential, Pentaspherical, Spherical, and Tetraspherical are the most popular models. However, for practicality of modeling the experimental datasets, only one variogram model should be selected (Ly, Charles, & Degré, 2011).

A variogram model selection can be justified by many performance indicators. However, fulfilling them concurrently remains questionable. This has made researchers consider in their analyses only the indicator that most achievable. In GIS application, five performance indicators are available which are root-mean standardized error, mean error, average standard error, root-mean-square standardized error and mean-square error (Johnston, Hoef, Krivoruchko, & Lucas, 2003). In order to evaluate the accuracy of spatial interpolation of a spatial rainfall map, all indicators are recommended to be emphasized. Thus, a multi-criteria decision-making tool which is able to incorporate the indicators in the decision-making process is an excellent opportunity to be adopted.

In a tropical urban area like Kuala Lumpur, the capital of Malaysia, the global climate change has affected the frequency of flash events in term of rainfall magnitude and spatial distribution. The main authority of hydrological work, the Drainage and Irrigation Department of Malaysia (DID) has installed many rain gauge stations within the Kuala Lumpur area (243km²) to monitor the flash floods events. To date, there are at least 3 networks developed with a total of more than 100 stations for hydrological purposes such as flood monitoring and water resources study. The density is vastly more

than the number recommended by the World Meteorological Organization (WMO) within the range of 10km² per station to 20km² per station (World Meteorological Organization, 2008). Indeed, these stations incurred a vast commitment in terms of maintenance costs.

In such a case, the rain gauge network needs to be optimized to ensure that the network is able to collect sufficient and less variation in spatial rainfall data by ensuring an optimum number of stations in the network. The network size and the existing location of the station need to be reviewed in order to optimize the network. In this case, the application of the conventional method and evolutionary computation are very beneficial to come up with a new approach. The Cross-Validation technique is an example of a conventional method that is widely used in various fields. On the other hand, Particle Swarm Optimization is an evolutionary computation widely applied especially in the electrical engineering field. However, a modification task must be conducted to apply suitable methods to overcome the rain gauge network problem.

1.3 Aim and Objectives of the Study

This study aims to optimize the number of rain gauges for effective rainfall data acquisition. To achieve the aim, four objectives are chosen for this research and listed as follow:

- a. To select an appropriate semivariogram model for spatial rainfall interpolation process,
- b. To adapt the Particle Swarm Optimization algorithm for rain gauge network optimization design,
- c. To optimize the number of rain gauges in the rain gauge network using the Cross-Validation technique and modified Particle Swarm Optimization method,

- d. To validate the optimized rain gauge network using an established hydrological model.

1.4 Significance of the Study

This study enhances the spatial rainfall mapping for a better presentation, instead of the numerical form. However, along with the process, several added values to the hydrological methodology are developed which has contributed to the significance of this study.

The results of this study can help hydrologist in the process of analyzing rainfall data using the geostatistical method in selecting the appropriate semivariogram model. The exact selection of the semivariogram models can be made, and the error in spatial interpolation at the ungauged location will be subsequently reduced.

The new methodology developed in this study offers a new perspective to hydrologist to optimize the rain gauge network, especially in geostatistical and artificial intelligence applications. This is parallel to the Fourth Industrial Revolution in which artificial intelligence is one of the emerging applications.

The ability to optimize rain gauge network will reduce the maintenance costs because the numbers of stations are reduced to the optimum size. Furthermore, the spatial rainfall map produced by the optimal network has minimal error. Thus, the spatial rainfall map has a better presentation of the rainfall data.

A better spatial rainfall map that is produced by the enhancement of spatial interpolation can be utilized for bias correction of the satellite rainfall estimation and Numerical Weather Prediction (NWP) data. It is the new direction in which such data is used in hydrological and water resource applications, but it must be corrected using

ground observed data prior to use. The output of this study offers a good option for the bias correction of the NWP data.

1.5 Thesis Structure

This thesis is divided into five chapters. The contents of each chapter are elaborated as follows:

This Chapter is an introduction which covers the background of the study, the problem statements, the aim and objectives of the study, the significance of the study and the structure of this thesis.

Chapter 2 covers a review of relevant literature material in the previous research works regarding rain gauge station network and the available optimization approaches in this field. A brief discussion regarding the introduction of rainfall measurement, the development of the rain gauge network, the method of spatial rainfall mapping and interpolation and the research conducted on the rain gauge network design or optimization are presented. A new proposed method that is the Particle Swarm Optimization method is also reviewed for modification and adaptation to achieve the objectives of this study.

The methodologies adopted in this thesis to answer the research objectives are explained in detail in Chapter 3. It covers the study area and the available rain gauge network, followed by the explanation on the input data used and the preliminary data analysis involved. This chapter also explains how the methodology was developed using the geostatistical method, cross-validation technique and the Particle Swarm Optimization method to solve the optimization problem of the rain gauge network. The verification method is included in this chapter in order to verify the optimized rain gauge network.

The results and discussions are presented in Chapter 4. The results from the applied methodology are presented along with the discussion about the research finding. Generally, it is divided into four major subchapters to discuss the result of the selected semivariogram model, the developed reference optimized rain gauge network, optimized rain gauge network using Cross Validation technique and modified Particle Swarm Optimization method and the validation of the optimized rain gauge network.

The last chapter of this thesis, Chapter 5, presents the conclusion of the research findings. The recommendations for potential future studies are included for the reader to enhance the methodology and/or improve the current results.

University of Malaya

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

A rain gauge network is a hydrological network which consists of a few numbers of the rain gauge station. The network has been designed to measure rainfall amount that reaches to the ground for hydrologic and water resource purposes. The rain gauge instrument installed could be an automatic (equipped with a telemetric system for real-time data collection or non-telemetry which collected manually) or manual. The measured rainfall data is normally archived for future purposes.

Rainfall amount that is measured by the rain gauge network is a point value which is not directly used in a hydrological application or rainfall data illustration, for instance, hydrologic modeling and spatial rainfall mapping. For this purpose, point rainfall data must be converted into areal rainfall. Several methods are available to calculate the areal rainfall and the most common techniques used in any hydrological application are Thiessen Polygon and Inverse Distance Weighting schemes due to their simplicity in theory and calculation. Besides, an advanced spatial interpolation method that is geostatistical analysis is also being applied for the hydrologic application. Somehow, whatever methods used in the hydrologic application, an appropriate quantity of rainfall data is essential to produce acceptably accurate output.

The main source of the rainfall data comes from a rain gauge network. Appropriate rainfall data is determined by the size of the network. A dense rain gauge network is certainly producing an appropriate quantity of rainfall data. However, to develop such a dense rain gauge network incurred high development and maintenance cost. For this reason, a rain gauge network needs to be optimized for an optimum size so that the rainfall data measured is able to produce an analysis output with acceptable accuracy.

This chapter presents a review of the studies that have been carried out for a rain gauge network optimization by the researchers and scientists. The achievement of method or procedure applied by them are presented and also with the advantages and disadvantages. In addition, new potential methods for rain gauge network optimization are explained as well.

2.2 Rain Gauge Network

The rain-gauge network is a network formed of several rainfall stations for a particular area. The individual station within the network will gauge the rainfall amount for certain rainfall event. The rainfall data collected by the rain gauge station were normally used to check on the water availability within the certain catchment and to give early warning to the public via a flood forecasting model. To ensure the network can execute, it needs a sufficient and reliable rainfall data. But in the real situation, sufficient and reliable rainfall data is very hard to get due to several problems such as missing data, instrumentation breakdown, lack of maintenance and an inadequate amount of rainfall collected by the station. These problems can be a threat to other hydrological application that used the rainfall data as the main input data. For instance, a flood forecasting model cannot be calibrated if the rainfall data is insufficient and this could lead to disseminating wrong flood warning information.

Division of Water Resources Management and Hydrology, Drainage and Irrigation Department of Malaysia (DID) has developed hydrological stations network as early as the 1970s. To date, there are more than 1,000 numbers of stations in the networks with various types of hydrological stations (telemetry or non-telemetry) that have been developed and maintained throughout Malaysia (Table 2.1). The department is implementing Quality Management System (QMS)-MS ISO 9001:2008 within the

hydrological data management scope to ensure the quantity and quality of collected hydrological data from the network.

Table 2.1: Statistical figures of the hydrological stations of the Drainage and Irrigation Department (DID, 2009).

Region	Rainfall	Evaporation	Streamflow	River Water Quality	River Suspended sediment
Peninsula	710	23	92	70	78
Sabah	83	4	37	32	31
Sarawak	297	25	39	5	-
Total	1090	47	168	107	109

Conventionally practiced in Malaysia, the Jabatan Pengairan dan Saliran (2009) has listed several factors that should be considered to develop rain gauge network:

- a. The minimum density of gauges as recommended by the World Meteorological Organization (2008)
- b. Gauge type - In Malaysia, a tipping bucket of 0.5mm is used because of the high intensity of rainfall received.
- c. The location of the rain gauge
- d. Observation and communication method for data collection.
- e. The cost of development
- f. The access to the site
- g. Maintenance aspect.

A comprehensive review by Shaghaghian and Abedini (2013) on factors that affected a typical rain gauge network design have stated that the factors that should be taken into consideration are, but not limited to the overall objective of the network design, the process considered, the attribute under consideration, the temporal scale or

sampling interval in, the spatial scale, the topographic setting, types of precipitation, the nature of the objective function used for design and the algorithm used for minimization of maximization the objective function value.

Rain gauge network must be designed in such a way that it can record rainfall data for certain purposes. Either for water resource study or for flood monitoring, the minimum numbers of rain-gauge should be in the network to record a sufficient enough of rainfall data that is a benefit to its purpose. The network with minimum numbers of rain gauge station is distinguished as an optimum rain gauge network for rainfall data acquisition.

The number of rain gauge stations in the base network may be different based on the temporal resolution of the rainfall data that required to be measured. Normally, the higher temporal resolution of rainfall data such as monthly and annually needs few numbers of the stations as compared with a smaller temporal resolution like hourly rainfall data. However, this rule should be assessed according to the climatological characteristic of the specific site.

2.2.1 Set up a rain gauge network

Normal practice in hydrological network development is to construct the initial network with a few numbers of rain gauge stations, a so-called base network for specific objective of rainfall data collection, for instance drainage and irrigation, flood mitigation and water reservoir project. Then, the network is to be augmented to improve the accuracy of the rainfall data from the available data sets.

In any development of the hydrological network, it must be in accord with the guideline that has been explained by the World Meteorological Organization (2008). The density of a rain gauge network depends on the geographical categories and the

type of rain gauge instrument. The general density of the rain gauge network as recommended by the WMO is tabulated in Table 2.2.

Table 2.2: The minimum station's density recommended (area in km² per station), (World Meteorological Organization, 2008).

Area	Rainfall station	
	Recording	Non-recording
Coastal	9,000	900
Mountains	2,500	250
Interior plains	5,750	575
Hilly/undulating	5,750	575
Small islands	250	25
Urban areas	10 - 20	-
Polar/arid	100,000	10,000

The climate of the river basin also influences the rain gauge network design. All factors should be considered before designing a rain gauge network but to incorporate them in design is almost impossible. Thus, the appropriate analysis must be carried out to find the most important factors to be considered for rain gauge network optimization or design.

Rain-gauge network must be designed in such a way that it can record rainfall data for certain purposes. The minimum number of rain gauges should be in the network to record sufficient enough rainfall data for the benefit of the purpose. According to Table 2.2, for the urban area, a rain gauge station should cover area between 10 km² to 20 km². Thus, if a 500 km² basin is assumed to be considered, the recommended number of rain gauge station should be in a range of 25 to 50 stations. However, the number of stations somehow do not guarantee the accuracy of the spatial rainfall interpolation within the basin.

2.3 Spatial Rainfall Interpolation

In hydrologic and water resources application, rainfall recorded by a rain gauge network is converted into areal rainfall according to the basin or subbasin. Areal rainfall is commonly presented as a single or spatial rainfall value for each basin and spatial mapping.

The areal rainfall of river basin is estimated by using the averaging or interpolating method for hydrological application. For this purpose, Thiessen-polygon method and Inverse Distance Weighted Method are two common methods can be used to estimate areal rainfall. Areal rainfall estimated through these methods is still a single point value at the center of the river basin or catchment. These methods have been used widely in the hydrological application and it is still relevant today.

2.3.1 Thiessen-polygon Method

Thiessen-polygon method has been commonly used to estimate the basin or areal rainfall in the hydrological application included in the hydrologic software such as HEC-HMS. Its simplicity in concept and calculation make it easy and practical to be applied. The concept of Thiessen weight method is that the individual station will cover a certain area out of the total basin proportionally with adjacent stations. The Thiessen Polygon must be constructed first to calculate the coverage area of every station within the basin area as illustrated in Figure 2.1. The equation below is used to calculate the estimated areal rainfall.

$$P_t = \frac{\sum P_i \cdot A_i}{\sum A_i} \quad \text{Equation 2.1}$$

where,

P_t = estimated areal rainfall

P_i = recorded rainfall at i station

A_i = coverage area of station i

i = number of stations (1, 2, 3, ..., n)

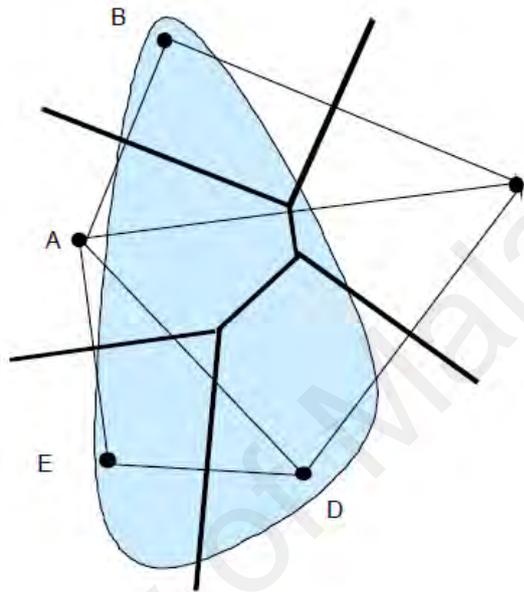


Figure 2.1: Typical example of Thiessen Polygon

2.3.2 Inverse Distance Weighting Method

Inverse Distance Weighting (IDW) scheme is a point-wise estimator. This scheme is manipulated to estimate areal rainfall which assuming that the value can be measured at the centroid of the basin. The centroid will be the reference to calculate the weight based on the distance to the nearest station in the quadrants as depicted by Figure 2.2. IDW relies on the theory that the interested interpolated point is more influenced by closer gauged location than by gauged location further away. In other words, if a set of values of gauged points are arrived at, then the values at ungauged points which located within the set can be calculated (Ly et al., 2011). Suppose that A, B, C and D are the

rain-gauge station and d_A , d_B , d_C and d_D are the distance, respectively. The estimated rainfall at point Y (point that under consideration) will be calculated as,

$$P_Y = \frac{\left(\frac{1}{d_A}\right)^2 \cdot P_A + \left(\frac{1}{d_B}\right)^2 \cdot P_B + \left(\frac{1}{d_C}\right)^2 \cdot P_C + \left(\frac{1}{d_D}\right)^2 \cdot P_D}{\left(\frac{1}{d_A}\right)^2 + \left(\frac{1}{d_B}\right)^2 + \left(\frac{1}{d_C}\right)^2 + \left(\frac{1}{d_D}\right)^2} \quad \text{Equation 2.2}$$

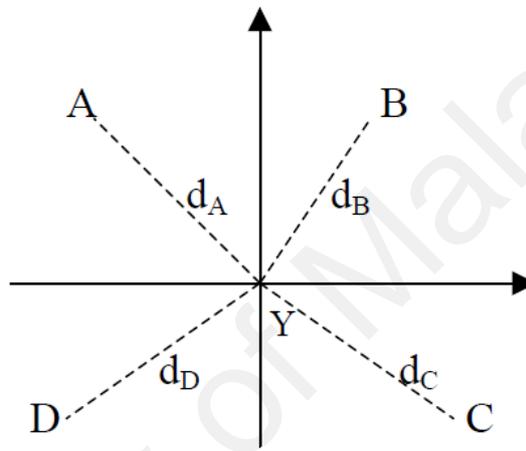


Figure 2.2: Typical example of Inverse Distance Weight method.

2.3.3 Geostatistical method

The geostatistical method was originally developed to study mining activity (Journel & Huijbregts, 1978). It is an advanced method of studying spatial datasets in vast research fields. With the geostatistical method, the datasets are modeled based on the spatial variation between each data point and presented via semivariograms (Xu et al. (2013), Shaghaghian and Abedini (2013), Putthividhya and Tanaka (2012), F. Othman, Akbari, and Samah (2011), Zhang and Yao (2008), Garcia, Peters-Lidard, and Goodrich (2008), Cheng et al. (2008)).

A semivariogram is a graphical diagram that explains the relationship between the variability of a dataset and the distance of individual data in a certain direction. A typical semivariogram example is illustrated in Figure 2.3. The x -axis in the diagram is a group distance between two dataset locations, also known as separation lag. The y -axis represents the variability measurement of the dataset group distance that is measured as semivariance. In addition, the geostatistical characteristics of the studied dataset are inferred from the semivariogram properties (sill, range and nugget) after fitting the studied dataset to an appropriate variogram model. In a typical semivariogram, the nugget is a value of initial variability in the smallest group distance, including the measurement error. The variability value rises from the initial value up to the sill, where the line is off or flattened. The sill value can be read from the semivariogram where the line is off and the partial sill is calculated by subtracting the nugget value from the sill. The range is the distance value extracted from the semivariogram at the sill's location on the diagram that is beyond this range and where the autocorrelation measure is zero.

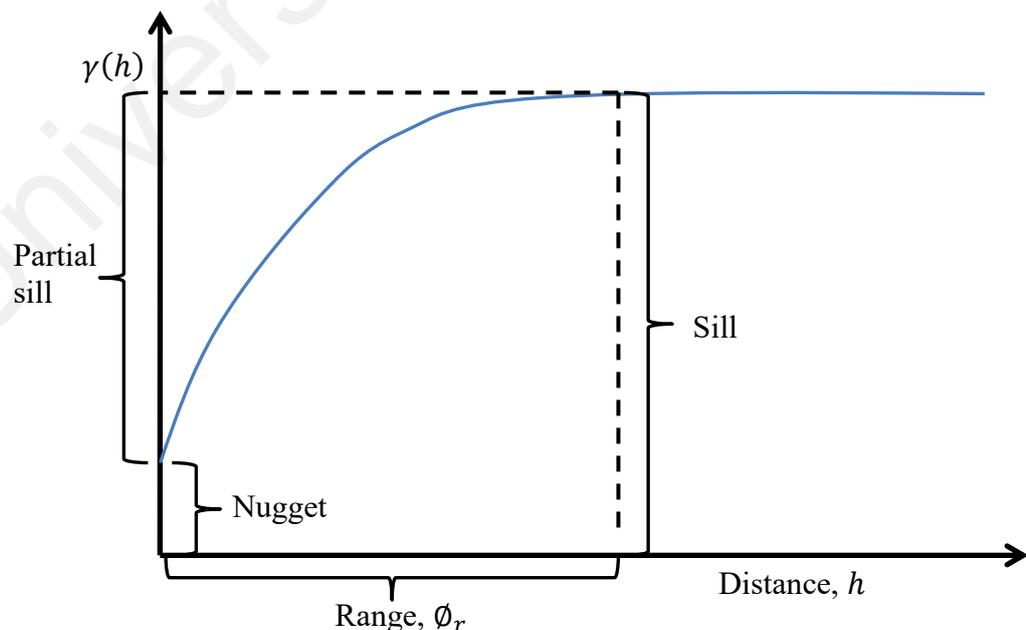


Figure 2.3: Example of semivariogram.

A dataset can be modeled using Equation 2.3,

$$\gamma(h) = \frac{1}{2n} \sum [g(x) - g(x+h)]^2 \quad \text{Equation 2.3}$$

Where is $\gamma(h)$ the semi-variance, $[g(x) - g(x+h)]$ is the difference value of paired dataset, and n is the size of the dataset. The semivariogram properties are calculated using a variogram model that fits the dataset. Fitting the experimental dataset semivariogram to an appropriate variogram model is an important stage in the geostatistical analysis. Several variogram model candidates are available to fit the experimental dataset, for instance, Spherical, Tetraspherical, Pentaspherical, Exponential and Gaussian. The use of the variogram model depends on the purpose of the analysis. For instance, a study conducted by Z. Xuesong and Srinivasan (2009) to compare the geostatistical approaches used variogram that best fits the experimental data in each analysis. The approach to use the best fits variogram model is efficient in the analysis but it is not practical in other applied study since the experimental data has its own characteristic. More practical, the best variogram model fits the experimental dataset is selected for geostatistical analysis (Ly et al., 2011; F. Othman et al., 2011).

In GIS application, 5 performance indicators of best-fit measure are available: root-mean standardized error, mean error, average standard error, root-mean-square standardized error and mean-square error (Johnston et al., 2003). However, to the best of knowledge, all studies of the Geostatistical analysis used only some indicator and none considered all indicators while fitting the variogram model. It is beneficial to recommend that to consider all indicators to evaluate the accuracy of spatial interpolation as well as the smoothness of the produced map.

To conduct such a task, a multi-criteria decision-making tool which will be able to incorporate the indicators in the decision-making process such as Analytical Hierarchy Procedure (AHP) is needed. The AHP is a simple structured approach using the criterion and the alternative decisions in hierarchy form to analyze the decision-making process. It can consider both numerical and non-numerical form of criterion. Based on this reason, the AHP has offered a great opportunity to fill in the gap.

2.3.3.1 Spatial Interpolation

The variogram model is used to represent the spatial correlation of the experimental dataset using variography parameters. The variography parameters are then used for spatial interpolation process to estimate the ungauged rainfall. The general equation for spatial rainfall interpolation is as follows:

$$z_{\text{est}}(s_o) = \sum_{i=1}^m \lambda_i z(s_i) \quad \text{Equation 2.4}$$

Where z_{est} is the estimated rainfall value at the location (s_o) and λ_i is the weight associated with the observed value $z(s_i)$. The weight λ_i is calculated based on the distance from the observed data to the predicted location and their spatial variation using the variogram model. The sum of all weights λ_i must be equal to 1 to ensure that the predicted value is unbiased.

Based on the literature, the Kriging interpolation has multi variants. Among Kriging variants associated with the geostatistical method are Ordinary, Simple, Universal, Indicator, Probability, Disjunctive and Co-Kriging, and so many others. All of these Kriging variants quantify the spatial structure of the data and the prediction error.

2.4 Application in Rain Gauge Network Optimization

A rain gauge network needed to be optimized for an optimum number of stations in the network and their spatial distribution to ensure that the network would be able to collect sufficient and less variation in spatial rainfall data. To achieve this, the researcher had to set an objective function to be evaluated as indicator or criteria of design or optimization work. In the literature, most of the cases of rain gauge network optimization are minimizing the objective function and this depends on the variable that is selected as the objective function. For instance, Pardo-Igúzquiza (1998) used the accuracy of mean areal rainfall and cost of data collection as an objective function that had to be minimized. On the other hand, there was a case of maximizing and minimizing the objective function simultaneously as studied by Volkmann et al. (2010).

2.4.1 Statistical Method

Optimization of the rain gauge network has been done by researchers using several methods. Earlier research on rain gauge network assessment has been conducted using classical methods, such as statistical and probabilistic approaches. Nemeč and Askew (1986) explained the philosophy of hydrological network design using statistical moments of mean and variance. Sorman and Balkan (1983) applied the same statistical moments to redesign the rain gauge network in the Kizilirmak River basin. However, limitations with the statistical ability to explain precise rainfall data have encouraged the application of the probability theory.

2.4.2 Probabilistic Method

A probabilistic approach called the entropy method has also been used to design hydrological station networks. This method is also known as the Shannon Entropy (Shannon & Weaver, 1949) and can be utilized to model system information through transmitting and receiving information as entropy values. The probability distribution

logarithm serves to measure the entropy value. According to the literature, this method has been used to study the influence of seasonal discharge information on the discharge networks of river basins (Mishra & Coulibaly, 2014). Mishra and Coulibaly (2010) also applied the entropy method to assess a discharge station network in a Canadian river basin.

A new entropy application approach called maximum information minimum redundancy (MIMR) was proposed by Chao, Singh, and Mishra (2012) to design a streamflow gauge and water level network. The entropy method has also been applied to evaluate rain gauge network performance for the appropriate selection of rain gauge stations in a number of studies by Krstanovic and Singh (1992a), Krstanovic and Singh (1992b), Yoo, Jung, and Lee (2008), Ridolfi, Montesarchio, Russo, and Napolitano (2011), Ridolfi et al. (2012) and Vivekanandan, Roy, and Chavan (2012). Moreover, this method has been coupled with the kriging technique to optimize the number of rain gauge stations in a network (Awadallah (2012), Chen, Wei, and Yeh (2008), Wei, Chiang, Wey, Yeh, and Cheng (2010) and Yeh, Chen, Wei, and Chen (2011)). In these studies, the locations of new stations were determined prior to be applying the methodology to evaluate the locations' effectiveness. This method was able to prioritize the number of candidate stations within the studied network.

The advantage of the entropy method in rain gauge network evaluation is that only rainfall data are needed. However, the entropy value is estimated and it is essentially dependent on the probability distribution used in the analysis. It is sensitive to the assumption in the probability distribution while making the estimation (Alfonso, Ridolfi, Gaytan-Aguilar, Napolitano, & Russo, 2014). Therefore, the entropy values depend on some assumptions that can influence the result.

2.4.3 Geostatistical approach

Another method that is quite extensively used in designing and optimizing rain gauge network is a geostatistical method. The geostatistical analysis is a recent method of designing and optimizing rain gauge networks applied by researchers. It is a robust method of studying environmental datasets from spatial or spatiotemporal perspectives. The geostatistical method can estimate the variable values under study through spatial interpolation as well as estimated variance.

Earlier publications of geostatistical applications for rain gauge network optimization are based on variance reduction, for instance, studies by Pardo-Igu'zquiza (1998), Barca, Passarella, and Uricchio (2008) and Cheng et al. (2008). Cheng et al. (2008) optimized a rain gauge network by introducing new stations and relocating existing stations based on the total areal percentage. Acceptable rainfall estimation accuracy was achieved at the stations and optimization was done based on trial and error. Pardo-Igu'zquiza (1998) minimized the variance of data collection estimation and the cost of designing an optimal rain gauge network. In the mentioned study, the variance of estimation represented the accuracy measure of the areal rainfall estimated from synthetic rainfall datasets. The optimization algorithm was developed by coupling the geostatistical and simulated annealing methods. These methods exhibited the ability to make good rainfall data estimations in the optimized rainfall network. Barca et al. (2008) applied this method to extend an existing rain gauge network for crop protection purposes.

Feki, Slimani, and Cudennec (2010) have carried out a research on rain-gauge network optimization using 3 types of kriging interpolation method (kriging with an external drift, regression-kriging and co-kriging) to map monthly rainfall data to evaluate the rain gauge network performance. The research has revealed that among

tested method the use of kriging with external drift and regression-kriging will generate isohyet map of monthly rainfall more likely with topographic pattern whereas co-kriging method will generate a smooth isohyet map contour zone. To date, the geostatistical method has emerged in the latest publications on rain gauge network design and optimization, e.g. Putthividhya and Tanaka (2012), Shaghaghian and Abedini (2013) and Feki, Slimani, and Cudennec (2016).

Geostatistical rainfall estimation is greatly dependent on the rain gauge network configuration. A good configuration tends to produce less variance, meaning that the spatial rainfall estimation is more accurate. As a general rule, every study on rain gauge network design requires to have an existing network of rain gauge in place before an optimal rain gauge network is configured. One could differentiate among various studies in reference to the support size of estimation. For instance, Pardo-Igu'zquiza (1998) and Barca et al. (2008) assessed optimal rain gauge networks based on the randomization of existing rain gauge networks. Pardo-Igu'zquiza (1998) also considered the support size for observation to be point-wise and that of estimation to be block-wise. Meanwhile, Cheng et al. (2008) adapted the support size for observation and estimation to be point-wise and employed trial and error on a predetermined rain gauge network to achieve an optimal rain gauge setup.

Nonetheless, the predetermine approaches have a tendency for bias at certain stations selected. Therefore, evaluating each station in the network is an alternative way to counter this issue, for instance by applying the cross-validation technique. Yeasmin and Pasha (2008) applied leave-one-out (LOO) cross-validation to examine the optimal number of rain gauges based on estimated runoff by removing stations one by one. This approach is simple and appropriate for evaluating individual stations. However, revalidation is recommended by adding the rain gauge stations one by one into the

existing, base rain gauge network, since both approaches produce different station combinations.

In such circumstance, the LOO cross-validation technique and geostatistical method have high potential to be used together to produce the best solutions, but the recommendation mentioned in the previous paragraph should be addressed. For instance, an opposite process to LOO has to be introduced as an enhancement of the LOO method to evaluate the station's combination in a network. In other words, the cross-validation techniques can be used as a generator of the candidate of optimal rain gauge network. This can be coupled with geostatistical analysis for evaluating different station combinations. In addition, the geostatistical method is an advanced and robust method for the analysis of spatial datasets like spatial rainfall distribution produced by the rain gauge network.

2.4.4 Artificial Intelligence and a combination of methods

Another way of optimization method is tailoring an optimization method with others and this is also known as a hybrid approach. Ruiz-Cárdenas, Ferreira, and Schmidt (2010) adapted this approach to hybrid the genetic algorithm with local search operator and this algorithm has outperform the ordinary Genetic Algorithm and Simulated Annealing to design the ozone monitoring network. In hydrology field, Pardo-Igúzquiza (1998) used the advantages of a simulated annealing process to minimize the number and location of rainfall station to get its optimal number to estimate the spatial rainfall. However, the study was conducted using artificial rainfall data set.

Geostatistical method combined with machine learning algorithm (artificial neural network) was tailored by Foresti, Pozdnoukhov, Tuia, and Kanevski (2010) to map precipitation data. This method was successfully applied to estimate the extreme precipitation in the study area. Aziz, Yusof, Daud, Yusop, and Kasno (2016) used the

variance reduction approach using the geostatistical and Particle Swarm Optimization (PSO) to determine number and location of optimal rain gauge network in Johor, Malaysia for Northeast monsoon season (November – February). The same approach was adapted by Attar, Abedini, and Akbari (2018) to design an optimum rain gauge network using geostatistical coupled with the Artificial Bee Colony (ABC). The method was able to priorities the rain gauge stations in south-western part of Iran (subtropical region), whereby the climate is long, hot, dry in summers and short, cool in winters.

The intelligent algorithm is a robust approach to be applied for network optimization. However, the appropriate objective function coupled with the robust algorithm of optimization using an appropriate temporal resolution of input data will offer a great opportunity of getting expected result in research. Thus, the strategy of coupling or tailoring the process should be critically designed for research. The climate type and region of the study area are also important to consider.

2.4.5 Rainfall Data for Optimization

The primary input data for rain-gauge network optimization is rainfall depth from each station in the network. The rainfall depth is available in many temporal scales, for instances hourly, daily, monthly or annually. The appropriate temporal scale for analysis will justify the purpose of analysis. Thus, it is important to select the right temporal scale.

The main criterion to determine the selection of rainfall dataset temporal scale is not only relying on the optimization objective but the climate characteristic of the study area as well. Feki et al. (2010) had selected monthly rainfall data in the arid area to compare the three interpolation methods of kriging series. The same temporal scale has been applied by Chen et al. (2008) in their research using kriging and entropy method to design rainfall network in Taiwan which is subtropical climate. Another study in

Taiwan, Cheng et al. (2008) adapted two temporal scales, first is the annual rainfall and second hourly rainfall to optimize the rain gauge network. The hourly rainfall was used to analyse different type of storms such as 'Mei-Yu', convective, typhoon and Frontal. Meanwhile, the annual rainfall was retrieved irrespective of storm type.

In the tropical monsoon area, Putthividhya and Tanaka (2012) had used the monthly and annual rainfall data to redesign the optimal network and predict the spatial rainfall. A study on spatial rainfall characteristic for the tropical region with the almost fully urbanized area by F. Othman et al. (2011) had used daily rainfall data based on storm event in the semi-variogram analysis. The daily rainfall as well was adapted by Ly et al. (2011) to set up an algorithm for interpolation of spatially daily rainfall at basin scale using the geostatistical algorithm.

Previous studies on evaluating or designing rainfall networks have used various rainfall data time intervals, from minute to annual scale. Most studies employed large time intervals, such as monthly and/or annual data (Feki et al. (2010), Chen et al. (2008), Yeh et al. (2011), Awadallah (2012), Putthividhya and Tanaka (2012), Vivekanandan et al. (2012), Shaghaghian and Abedini (2013); Yong, Hyunglok, Jongjin, and Minha (2014) and Feki et al. (2016)). To the best of the authors' knowledge, only a few studies have used the daily time interval, for instance, Krstanovic and Singh (1992a, 1992b), Barca et al. (2008), Yoo et al. (2008) and Mishra and Coulibaly (2010). In an extensive study, Ridolfi et al. (2011) used multi-time interval data, in which more details of time scale resolution were analyzed for a better result.

The temporal resolution of rainfall dataset must be chosen which is suitable to the climate of the study area. The type of storm in the study area should be distinguished prior to determining the form of temporal resolution of rainfall dataset. A tropical region

like in Malaysia, a convective rainfall is dominant especially in city areas like Kuala Lumpur. Based on the DID flood report, it may be noticed that the temporal pattern of rainfall events was inconsistent from one event to another. A short duration rainfall normally in 1 to 3 hours is likely to happen and the rainfall depth for this duration is equivalent to a daily rainfall record. In such a case, the daily rainfall format is appropriate to be selected to avoid analysis complexity. Thus, it is better to limit the scope to focus on performance.

2.5 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is one of the computational methods in computer science that solves the optimization problem via iteration process to improve the computed solution. It was introduced by Eberhart and Shi (1995) to simulate the social behaviour that imitates the movement a flock of birds of searching for their foods. From here, it was simplified to be an optimization operator of a problem.

The PSO solves the optimization problem of a population of solutions which is also called as swarm of particles that distributed randomly in the swarm. The particles will move within the swarm to search the potential solution and each particle will interact with each other to find the best solution. In the searching process, the particles move in multi-direction of search space with different velocity towards the position of potential solutions. The searching process continued to another position with different velocity in the best current potential solution. The particles will keep moving until they stop at the same best solution.

The PSO algorithm is easy to apply as compared with other mathematical computation methods. The simple analogy of the algorithm and less number of parameters to be considered makes it convenient to code using a computer. However, as a metaheuristic method, it does not offer a good convergence to the optimal solution.

Despite, the PSO has been applied to a vast research field as reviewed by Poli, Kennedy, and Blackwell (2007) and Poli (2008).

To implement the PSO, the objective function, $f(x)$ to solve the problem must be determined. Variable x is the solution to the problem of $f(x)$ that giving the objective function value either maximum or minimum and this depends on the objective to solve the problem. To search for a solution, the initial population (P) with the number of particles (N) has to develop within the upper and lower boundary limit of x (X^{UL} and X^{LL} , respectively) and these particles move with initial velocity set (V) as follows:

$$P = \{x_1^j, x_2^j, x_3^j, \dots, x_N^j\} \quad \text{Equation 2.5}$$

$$V = \{v_1^j, v_2^j, v_3^j, \dots, v_N^j\} \quad \text{Equation 2.6}$$

Where, j is the number of iterations and j equal to zero is referred to as the initial value. The objective function $f(x)$ is evaluated using P value and it depends on the optimization problem (assume that to find the global minimum). At the early stage, the P value at j equal to zero is assumed to be the personal best value (p_{pb}) for every particle. And the lowest p_{pb} value is assumed to be the global best value (p_{gb}). For the next iteration, the P value is updated using the new computed V value as in Equation 2.7 and Equation 2.8 and the objective function is computed using the current P value. The p_{pb} value for each particle is updated by taking the p_{pb} value with the corresponding lowest objective function value. Whereas, the global best value (p_{gb}) is updated from the current lowest of p_{pb} value. This process is continued until all particles reach at the global best value.

$$v_{p=1,2,\dots,N}^{j+1} = v_{p=1,2,\dots,N}^j + c_1 \times r_1 \times [p_{pb} - x_{p=1,2,\dots,N}^j] + c_2 \times r_2 \times [p_{gb} - x_{p=1,2,\dots,N}^j] \quad \text{Equation 2.7}$$

$$x_{p=1,2,\dots,N}^{j+1} = x_{p=1,2,\dots,N}^j + v_{p=1,2,\dots,N}^{j+1}$$

Equation 2.8

Where,

c_1 = a cognitive constant coefficient

c_2 = a social constant coefficient

r_1 and r_2 = random number independently selected from uniformly distributed in the interval of [0,1].

Again, the $f(x)$ values are computed using the current P value. The p_{pb} value for each particle is updated by taking the p_{pb} value with the corresponding lowest $f(x)$ value. Whereas, the p_{gb} is updated from the current p_{pb} value. This process is continued and stopped when all particles reach at the same global best value or the specified maximum number of iterations is reached. These conditions are distinguished as termination criteria and at this point, normally, the PSO is assumed to have converged to the solution of the problem. The general procedure of the PSO algorithm is illustrated in Figure 2.4.

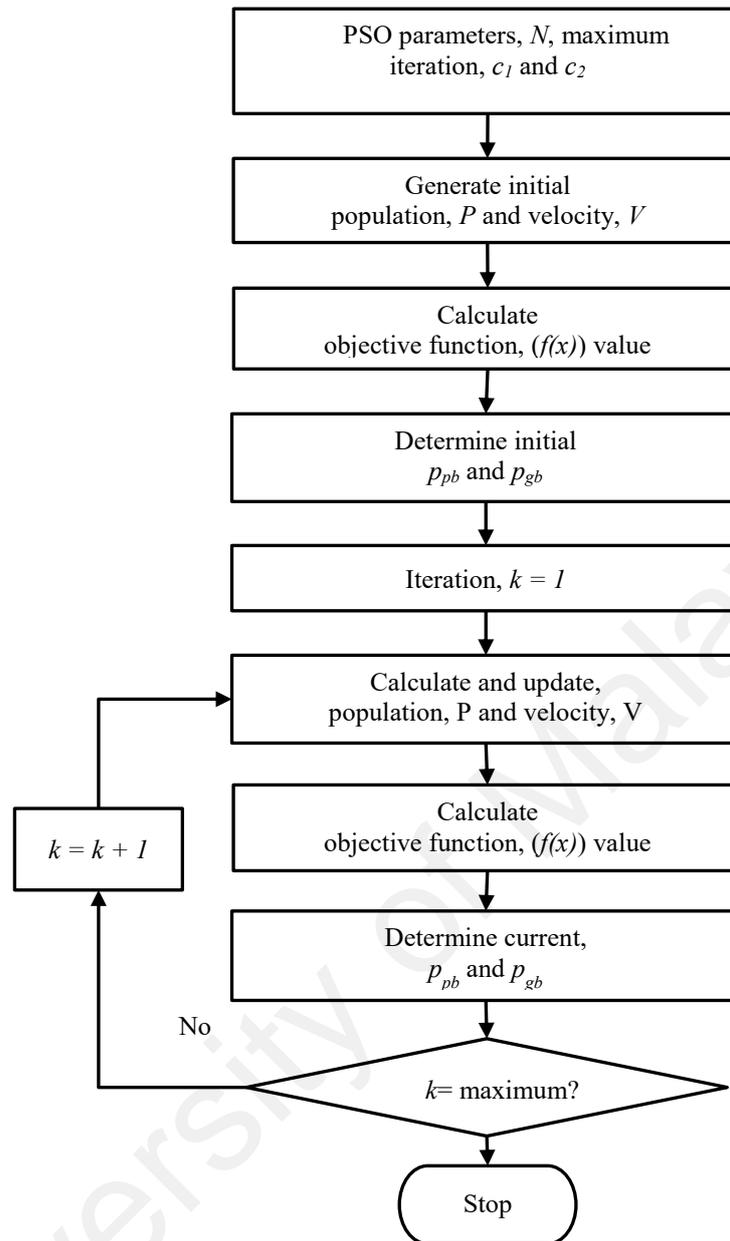


Figure 2.4: A typical procedure of the PSO algorithm

The drawback of PSO is the slow convergence towards the solution of the optimization problem. To overcome this problem Y. Shi and Eberhart (1998) have introduced the scaling factor called inertia weight (w) which control the previous value of velocity and the current velocity. The inertia weight is able to improve the convergence of PSO by setting the inertia weight, decreasing linearly with iteration

from the maximum value (w_{max}) of 0.9 to the minimum value (w_{min}) of 0.4 as follows:

$$w = w_{max} - \left(\frac{w_{max} - w_{min}}{j_{max}} \right) \cdot j \quad \text{Equation 2.9}$$

Where j and j_{max} are the iteration number and maximum number of iterations, respectively. By incorporating the inertia weight into Equation 2.7, the updating equation for velocity becomes as follows:

$$v_{p=1,2,\dots,N}^{j+1} = w \times v_{p=1,2,\dots,N}^j + c_1 \times r_1 \times [p_{pb} - x_{p=1,2,\dots,N}^j] + c_2 \times r_2 \times [p_{gb} - x_{p=1,2,\dots,N}^j] \quad \text{Equation 2.10}$$

The cognitive and social coefficients are considered the most dominant terms modified by various researcher to improve the learning rate of both components. However, in most standard PSO applications, for instance, in Mu, Cao, and Wang (2009) and Ravizi (2012), the c_1 and c_2 coefficients are set to 2.

Safaei, Tavakkoli-Moghaddam, and Kiassat (2012) used the time-varying value for w , c_1 and c_2 . The initial and last value of w was [0.9, 0.4] and for the c_1 , it was [2.5, 0.5]. Whereas, for the c_2 term, a varying intervals of [0.5, 2.5] had been used. In another study by Eberhart and Shi (2001), they used the value of w which randomly varied in the range of [0.5, 1.0] that was calculated by $[0.5 + (\text{rand}/2)]$ and the c_1 and c_2 values were fixed as 1.494. These parameter values were found successful to track the optimization task of the dynamic system in their study.

In the literature, the effort to improve the convergence characteristic of the PSO has succeeded to solve the particular studied problem. However, to apply it to a discrete

problem such as rain gauge network optimization, modification on the PSO algorithm needs to be done. This is a new contribution to the hydrology field.

2.5.1 PSO Application in Rain Gauge Network Optimization

To the best of knowledge, the PSO has been used only by Aziz et al. (2016) as a tool of optimization for rain gauge network design. In their study, the swarm's particles only randomly explore the possible optimum network based on the number of station selected by different particles. It is a great opportunity to explore the ability of the swarm's particles of PSO to solve a new optimization problem of rain gauge network evaluation whereby all possible combination of stations by each number of stations selected are evaluated.

Since then, early establishment of rain gauge network is based on ad-hoc assumption; the topology is more oriented toward being dense. Thus, the rain gauge network optimization problem consists of two problems, first, if the network is too dense the existing network is optimized for the optimum size and second if the network is sparse, it is expanded to the optimum size. Most of the studies regarding the rain gauge network evaluation or design are dealing with the first case, for instances Adhikary et al. (2015), Kar et al. (2015), Bakhtiari et al. (2013), Shaghaghian and Abedini (2013); Vivekanandan and Jagtap (2013a, 2013b). The main reason for this is that the rain gauge network involves quite a huge allocation to the organizations for operation and maintenance. In fact, the allocation is directly proportional to the number of stations in the network and it is influenced by the economic condition of the organization. In this perspective, a good review made by Mishra and Coulibaly (2009) stated that the reduction of hydrometric network density in Canada is due to the reduction in the allocation and also the changes of management focus of the government.

To optimize the network (irrespective of reduction and/or increase in the network size), it is all about to get the appropriate combination of stations to form an optimum network. It is similar to the combinatorial case where the intention is to choose a certain number of the stations without considering its sequences. This condition makes the rain gauge network optimization problem differ from the case of travel salesman problem (TSP) that is solved using the PSO method as applied by Abdel-Kader (2011); Fan (2010); Goldberg, Goldberg, and Souza (2008); X. H. Shi, Liang, Lee, Lu, and Wang (2007); Tasgetiren, Suganthan, and Pan (2007); Y. Xuesong et al. (2012).

In order to apply the PSO to solve the rain gauge network optimization problem, the PSO algorithm must have to go through the modification process. Since the PSO is used to solve a problem where the solution is in continuous number, thus it is essential to modify the algorithm to works for the discrete number as solution. This is an enhancement of PSO and it is fruitful to explore for a new contribution to the artificial intelligent and hydrologic application.

2.6 Conclusion Remark

Rain gauge network augmentation from the ad-hoc network has increased the density of the rain gauge. In an urban tropical catchment area like Klang river basin, at least 3 networks have been developed for hydrological purposes such as flood monitoring and water resources study. The rain gauge station density is about 13 km² per station, almost near to the maximum density recommended by the World Meteorological Organization (WMO). Indeed, these stations incurred a vast commitment in terms of maintenance costs. Thus, it is very essential to evaluate the existing rain gauge network for optimum size and prioritize the stations.

One of the latest methods applied by researchers to optimize rain gauge networks is geostatistical analysis. It is a robust method of studying environmental datasets from

spatial or spatiotemporal perspectives, like rainfall. To enable the method to evaluate the existing rain gauge network through the optimization process, cross-validation technique with two different techniques as mentioned in the previous section are adapted in this study as a generator of the candidate of optimal rain gauge network prior to analysis for an optimum network.

More advanced approach, the application of artificial intelligent of PSO has a great opportunity to apply for the rain gauge network optimization. The PSO is a well-established method in research whereby it has a simple analogy of the algorithm, easy to apply compared to other methods, and less parameter to be considered makes the algorithm convenient to code using a computer. In addition, the method uses the actual variable value in the analysis. Moreover, the algorithm has been applied successfully for the same purpose in a tropical basin. However, the suggested improvements to overcome the slow convergence rate problem and to increase the ability of exploration of the swarm's particles of PSO need to be considered. Therefore, modifications to the PSO algorithm are tailored to improve the method prior to applying it for a discrete problem such as rain gauge network optimization.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter elaborates in detail the methodology applied to fulfill the four objectives that have been determined to achieve the aim of this study. It begins with the description of the study area in section 3.2 and followed by section 3.3 which explains the general overview of the overall methodology.

Next, the input data collection and classification process is elaborated in section 3.4, where all the required data are gathered and sorted according to the uses of the data in the methodologies. Then, preliminary data analysis is carried out and explained in section 3.5.

In section 3.6, the methodology to achieve the first objective that is to select the semivariogram using the AHP method is presented. The methodologies for the second and third objectives are explained in section 3.7 which are to optimize the number of rain gauge in the rain gauge network using the Cross-Validation technique and modified Particle Swarm Optimization method.

The last methodology elaborated in section 3.8 is to validate the optimized rain gauge network using an established hydrological model in order to achieve the fourth objective of this study.

3.2 Description of Study Area

The study area comprises the upper part of the Klang River basin (UKRB), which is located in the federal territory of Kuala Lumpur, Malaysia, and some parts of the state of Selangor (Figure 3.1). The UKRB approximately located between $101^{\circ}36'E$ to $101^{\circ}51'E$ and $3^{\circ}4'E$ and $3^{\circ}24'E$. The basin covers about 584 km^2 of the catchment area. The northern part of the study area is about 1,366 m above sea level and is covered with

virgin forest. The southern part is a fully urbanized, almost flat city area, and located about 16 m above sea level. This area has a high density of residential population.

The climate in the study area is influenced by the monsoon system which is categorized into 4 seasons, two main monsoons and two transitional monsoons. The main monsoon seasons occur from December to March (also known as the northeast monsoon) and from June to September (also known as the southwest monsoon). The transitional monsoons occur from March to June and September to December. Usually, the study area receives heavy convective rainfall in March, April, October and November. These four months received mean monthly rainfall more than 250mm. This scenario is explained by a long-term monthly mean (from 1 January 1995 to 31 December 2015) as illustrated in Figure 3.2. In addition, the months of May and September received quite high rainfall amount which is more than 200mm. Based on the historical flood record, most of the flash flood events occurred in these months. In general, the annual and monthly mean rainfall received is relatively uniform with about 2,573 mm per year and 216 mm per month.



Figure 3.1: Location of the research area.

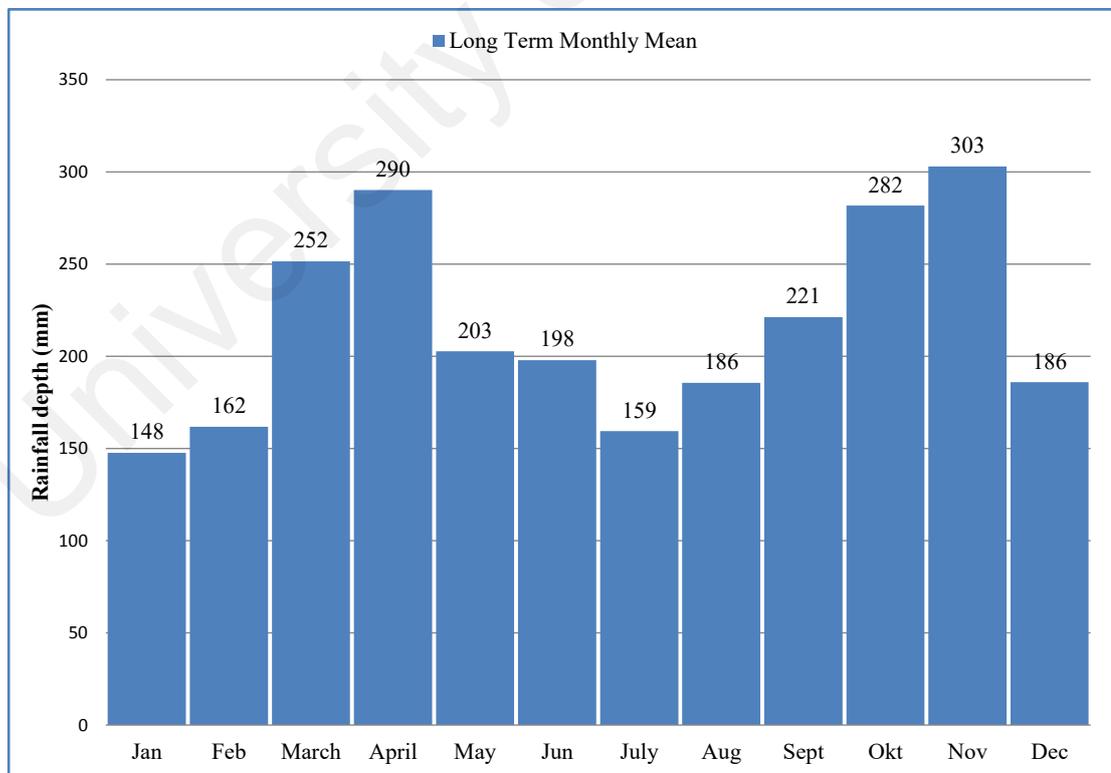


Figure 3.2: Long-term mean monthly rainfall in the study area.

3.3 Overview of Methodology

The methodologies have been designed in this study according to the study's objectives. They are inter-related one to another and started with the data collection and screening process. The data are used in the geostatistical analysis to analyze the performance of the semivariogram models based on the performance indicators. The outputs from this analysis are evaluated using the AHP application to select the best semivariogram model to be used in the next geostatistical analysis. To the best knowledge, the geostatistical and AHP application is a new methodology in order to select the variogram model that fits to model the experimental rainfall dataset.

Next, the selected semivariogram model is used in the geostatistical method in the optimization process to obtain the optimum rain gauge network size. Two new methods, cross-validation technique coupled with geostatistical and MPSO are proposed as optimization tools for a single objective of optimization problem based on the E_{rms} .

Finally, a hydrological lump-model of tank model for the study area is developed to validate the optimum rain gauge network produced by the optimization tasks. The summary of the methodology in this study is illustrated in Figure 3.3.

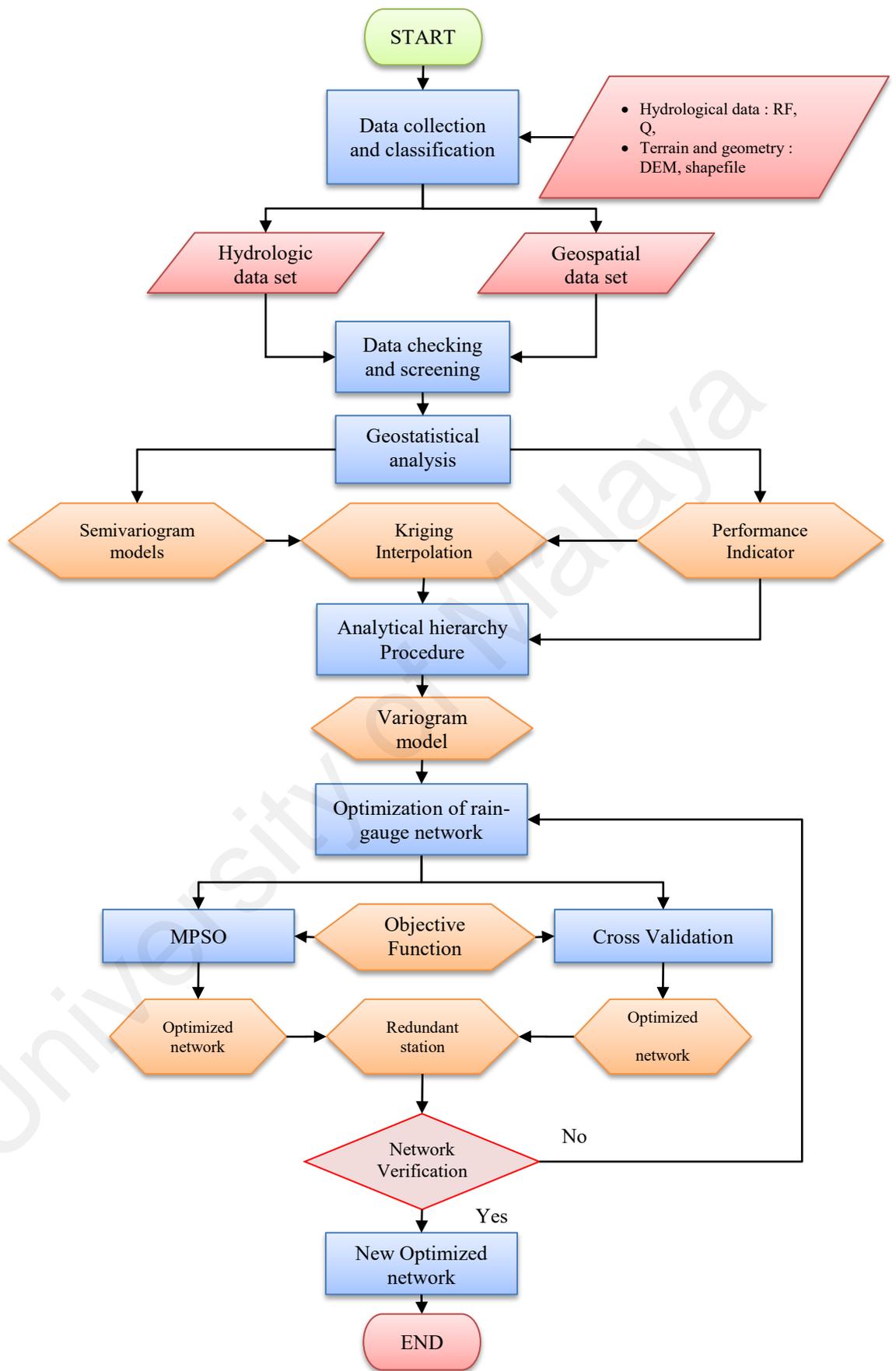


Figure 3.3: Summary of research methodology.

3.4 Input Data

The UKRB contains three rain gauge networks available for hydrologic and water resource purposes, namely the Storm Water Management and Road Tunnel hydrological station (SMART), National Hydrological Network (NN) and *Infobanjir* Telematics' Network (TN). All stations in these networks were installed with automatic rain gauges and equipped with telemetry devices except for NN, which is a non-telemetry station. The NN is in function since 1972 for general hydrological and water resource application purposes. Meanwhile, the TN was set up in 2000 to facilitate an online real-time flood monitoring system by the Drainage and Irrigation Department (DID) via the *Infobanjir* webpage. SMART is the latest network designed in 2007 for the Storm Water Management and Road Tunnel Project to solve the flood problem in the Klang River basin and to reduce traffic congestions in the Kuala Lumpur city centre. All networks are operated and monitored separately by different divisions of Drainage and Irrigation Department (DID).

3.4.1 Rainfall dataset and flood events

As part of the study methodology, flood events in the study area are determined. The floods record in the period of 2007 to 2012 was investigated for flood events to collect rainfall data for analysis because SMART began to operate in 2007. Floods record was obtained from DID. 12 flood events during the period are available to be considered. All rainfall stations in the study area and adjacent to the boundary were determined. Preliminarily, 56 rainfall stations were available for consideration.

To ensure robust analysis and results in this study, the flood events and rainfall datasets were examined. Good datasets were extracted based on the availability and completeness of rainfall data. For this purpose, the rainfall stations and flood events were filtered through the following process:

- a. Rainfall stations with more than 10% missing data based on 12 flood events were rejected.
- b. The remaining rainfall stations were used to assess the validity of the flood events to be used according to 3 criteria:
 - i. Percentage of rainfall stations with missing data. The flood events at stations with more than 10% missing data were rejected. Next, the remaining missing rainfall values for the station and for each flood event were estimated using the Inverse Distance Weight (IDW) method.
 - ii. The average rainfall value for each flood event was computed and flood events whose average rainfall was less than or equal to 10 mm were removed from the analysis. The threshold value was adopted from DID, based on whose records, an average rainfall of 10 mm produced negligible or minor flood events.
 - iii. The effective maximum rainfall value was determined for each flood event. Flood events with maximum rainfall of less than 60 mm were rejected as they may possibly generate insignificant floods according to the DID flood report.
- c. Flood events that met any one of the criteria in b were excluded from further analysis. The final flood events used in this study and brief rainfall information are tabulated in Table 3.1.

The filtering process yielded 55 rainfall stations and 7 flood events. For each station, the daily and time series of 15 minutes time interval rainfall data were obtained from the DID hydrological database (NIWA-Tideda software version 4). The daily data set is used for rain gauge network optimization. Meanwhile, the 15-minute time series data is used to develop the hydrological model for validation of optimized rain gauge network. These stations are arranged in Table 3.2 according to network type and location in the

study area and are denoted as an existing rainfall network. The existing rainfall network consists of the TN, NN and SMART rainfall stations, wherein TN has 9 stations, SMART has 21 stations and NN has 25 stations. Forty-four stations are located in the study area, which consists of 8 TN stations, 17 NN stations and 19 SMART stations. The other 11 stations are located outside the study area, which consists of 8 NN stations, 1 TN station and 2 SMART stations. The study area with 55 rain gauge stations and three networks was mapped on the Digital Elevation Model (DEM) of the study area, as illustrated in Figure 3.4.

The methodology applied involved an optimization task to select the appropriate number of rain gauge stations located within the catchment area. Among the 44 stations in the catchment area, 19 stations (SMART station) have been designed according to the WMO guideline for flood monitoring purpose. Meanwhile, 25 stations (8 TN stations and 17 NN stations), are categorized as hypothetical redundant because they were installed based on ad-hoc assumption. This set of stations will be evaluated and selected to remain in the optimum rain gauge network. Thus, an optimum rain gauge network will be consists of 19 stations (SMART station) plus with the rain gauge stations selected from the hypothetical redundant station. Therefore, the expected result of an optimum number of stations in the study area could be 19 to 44 stations. However, all 55 rain gauge stations are used in the analysis, for which the rainfall data from all stations were required in the optimization task. Those hypothetical redundant stations are the first 25 stations listed in Table 3.2.

Table 3.1: Events of the daily rainfall data used in this study.

Year	Date	Minimum Rainfall (mm)	Maximum Rainfall (mm)	Mean Rainfall (mm)	Standard Deviation (mm)
2009	3 February 2009	0	100	19.1	25.15
	3 March 2009	0	134	38	28.01
2011	13 December 2011	0	142	36.1	30.47
	18 September 2011	0	87	31.1	24.08
2012	7 March 2012	0	239.5	78.4	55.09
	18 April 2012	0	111	19.3	25.49
	21 August 2012	0	130	24.5	27.74

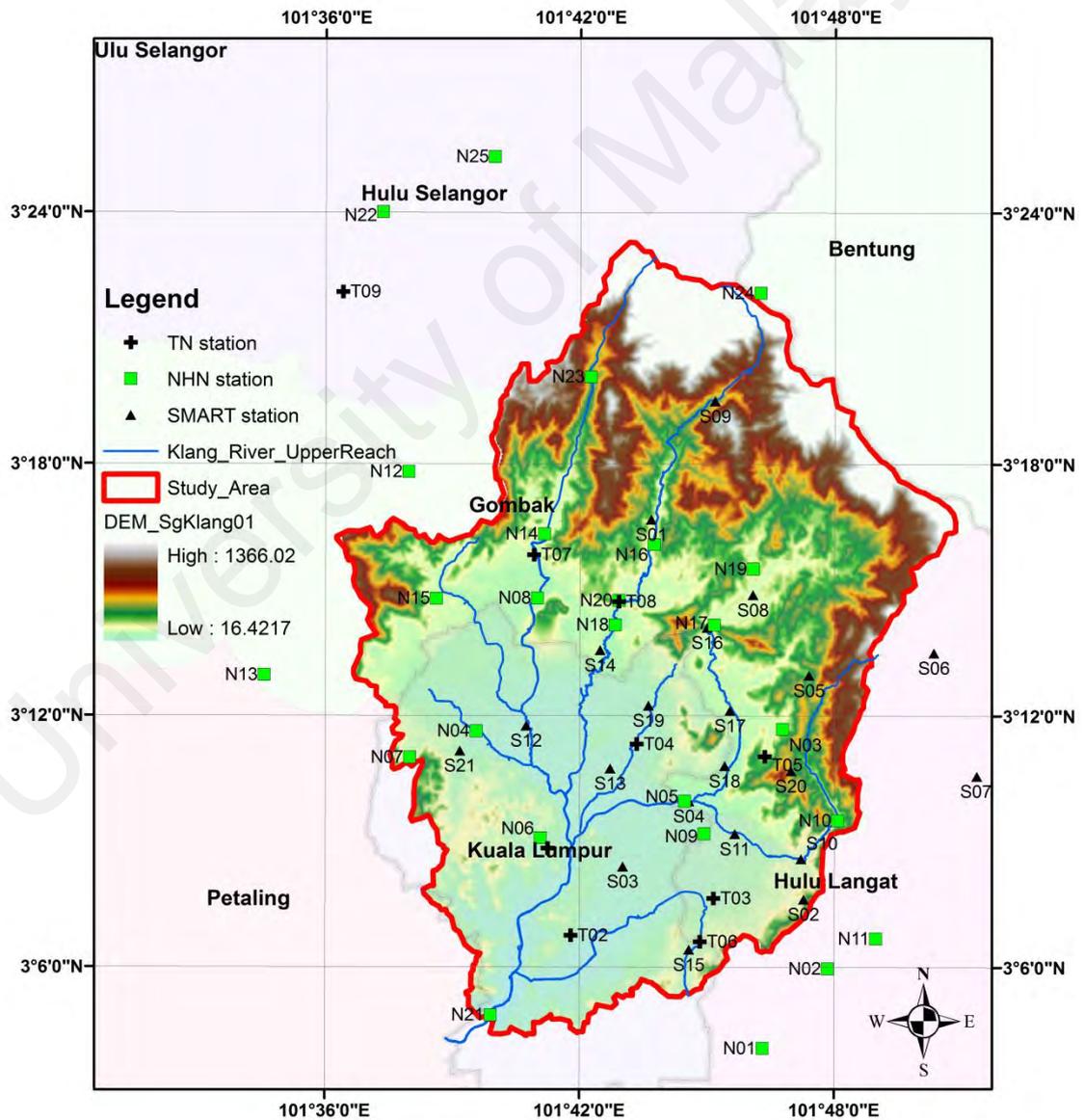


Figure 3.4: Location of the rain gauge stations mapped on DEM of the study area.

Table 3.2: List of rain gauge stations in the existing network.

Station ID	Station Name	Lat (⁰ N)	Long (⁰ E)	Network	Position ¹
T01	JPSWilayah_Tele	3.1472	101.6875	TN	In
T02	LebohPasar_Tele	3.1126	101.6966	TN	In
T03	PandanIndah_Tele	3.1274	101.7526	TN	In
T04	AirPanas_monthly_Tele	3.1887	101.7224	TN	In
T05	BktAntarabangsa_Tele	3.1837	101.7727	TN	In
T06	KgCherasBaru_Tele	3.1101	101.7474	TN	In
T07	EmpanganBatu_Tele	3.2639	101.6819	TN	In
T08	SimpangTiga_Tele	3.2453	101.7153	TN	In
N03	Km10 Ulu Kelang di UK Height	3.1947	101.7797	NN	In
N04	Jln Sg. Udang di segambut	3.1938	101.6594	NN	In
N05	Kg. Berembang di Keramat	3.1660	101.7413	NN	In
N06	JPS Wilayah Persekutuan	3.1514	101.6847	NN	In
N07	Ldg. Edinburgh Site 2	3.1833	101.6333	NN	In
N08	Klm T. Banjir Batu PO di Empat Tin	3.2467	101.6833	NN	In
N09	JPS Ampang	3.1531	101.7489	NN	In
N10	Pemasokan Ampang	3.1583	101.8014	NN	In
N14	Kg. Sg. Tua	3.2722	101.6861	NN	In
N15	Taman Ehsan di Kepong	3.2466	101.6436	NN	In
N16	Ibu Bekalan Km. 16, Gombak	3.2681	101.7292	NN	In
N17	Empangan Genting Kelang	3.2361	101.7528	NN	In
N18	Ibu Bekalan Km. 11, Gombak	3.2361	101.7139	NN	In
N19	Kg.Kuala Seleh	3.2583	101.7681	NN	In
N20	Kg. Kerdas(This station shifted from Gombak Damsite)(SMART)	3.2458	101.7153	NN	In
N21	Jam. Petaling di Jln Kelang Lama	3.0809	101.6652	NN	In
N23	Air Terjun Sg.Batu	3.3347	101.7042	NN	In
S01	SMART_Apartment UIA at Jln Gombak	3.2780	101.7280	SMART	In
S02	SMART_Bukit Ampang at Jln Ampang Hulu Langat	3.1270	101.7880	SMART	In
S03	SMART_Bulatan Kg Pandan at Jln Tun Razak	3.1400	101.7170	SMART	In
S04	SMART_DS Sg Kelang Sg Ampang Confluence	3.1660	101.7430	SMART	In
S05	SMART_IBMBS at Kg Kemensah	3.2160	101.7900	SMART	In
S08	SMART_Kuala Seleh at Empangan Kelang Gate	3.2480	101.7680	SMART	In
S09	SMART_Pusat Pengajian Luar UM at Jln Gombak	3.3250	101.7530 0	SMART	In
S10	SMART_Sg Ampang at Bukit Belacan	3.1430	101.7870	SMART	In
S11	SMART_Sg Ampang at Kg Melayu Ampang	3.1530	101.7610	SMART	In

Table 3.2, continued

Station ID	Station Name	Lat (⁰N)	Long (⁰E)	Network	Position¹
S12	SMART_Sg Batu at Jln Chendurah	3.1960	101.6790	SMART	In
S13	SMART_Sg Bunus at Jln Tun Razak	3.1790	101.7120	SMART	In
S14	SMART_Sg Gombak at Jln Changkat Tmn Greenwood	3.2260	101.7080	SMART	In
S15	SMART_Sg Kerayong at Jln Kuari	3.1070	101.7430	SMART	In
S16	SMART_Sg Kelang at Empangan Kelang Gate	3.2350	101.7500	SMART	In
S17	SMART_Sg Kelang at Jln AU 5C6	3.2020	101.7590	SMART	In
S18	SMART_Sg Kelang at Jln AU 31	3.1800	101.7570	SMART	In
S19	SMART_SM Pendidikan Khas at Genting Kelang	3.2040	101.7270	SMART	In
S20	SMART_The Peak at Tmn TAR	3.1780	101.7830	SMART	In
S21	SMART_Tmn Sri Sinar	3.1860	101.6530	SMART	In
N01	Sg. Raya Bt.9 Hulu Langat	3.0678	101.7719	NN	Out
N02	Sg. Serai Bt.12 Hulu Langat	3.0994	101.7975	NN	Out
N11	Batu 14, Hulu Langat(Balai Polis)	3.1114	101.8164	NN	Out
N12	Taman Templer	3.2969	101.6328	NN	Out
N13	Kg. Melayu Sg. Buloh	3.2161	101.5761	NN	Out
N22	Ldg. Sg. Gapi	3.4003	101.6225	NN	Out
N24	Genting Sempah	3.3681	101.7708	NN	Out
N25	Taman Desa Kelisa	3.4222	101.6664	NN	Out
T09	Serendah_Tele	3.3684	101.6068	TN	Out
S06	SMART_Kem Zone 6 at Sg Congkak Resorts	3.2250	101.8390	SMART	Out
S07	SMART_Kg Pangsun at Hulu Langat	3.1760	101.8560	SMART	Out
Note: ¹ In the catchment area.					

3.4.2 Discharge Data Sets

Based on the filtered 7 flood events, discharge data set for the 15 minutes time interval at the final outlet of the UKRB that is near station N21 was obtained from the DID hydrological database (NIWA-Tideda software version 4). The discharge data set is used together with the rainfall data set to develop a hydrological model for the UKRB to validate the optimized rain gauge network. However, for the 7 flood events, only 4

events that have a complete time series of discharge data set. The discharge data set are tabulated in Table 3.3.

Table 3.3: Discharge dataset and period of simulation used in the hydrologic model.

Date	Peak flood
3 March 2009	3 March 2009, 06:30pm
18 September 2011	18 September 2011, 04:00pm
18 April 2012	18 April 2012, 06:15pm
21 August 2012	21 August 2012, 06:00pm

3.5 Preliminary Data Analysis

The collected data sets were undergone preliminary analysis to explore and understand the characteristic of the data. This analysis is important to ensure that only good data sets are used and accurate approaches adopted in the further analysis of the optimization stage.

3.5.1 Rainfall Data Analysis

The rainfall datasets were analysed based on 2 indicators, descriptive statistics and correlation test between the rainfall amount and ground elevation. The descriptive statistics is used to compute the characteristic of the data such as mean, standard deviation, variance, the coefficient of variation, maximum and minimum value. These statistical parameters are essential to preliminarily estimate the optimum number of rain gauge station using the statistical method.

Prior to using the rainfall data in the further analysis, the correlation of rainfall data (denoted as y) with the ground elevation (denoted as x) is assessed to check the availability of relationship between these variables. This is important to confirm for better geostatistical analysis done in the next analysis. The Pearson correlation test is

applied and the equation given in Equation 3.1 is used to calculate the Pearson correlation, r :

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad \text{Equation 3.1}$$

Based on this formula, a strong positive correlation indicated if the value of r is close to +1, and if r is close to -1, a strong negative correlation is indicated. The elevation of the stations was extracted from the Digital Elevation Model (DEM) as illustrated in Figure 3.4.

3.5.2 Discharge Data Analysis

It is important to ensure a good response discharge data towards rainfall data. Thus, the correlation between rainfall and discharge data is checked using statistical correlation test. For this purpose, the Pearson correlation test as explained previously is used. The raw rainfall hyetograph would not be in same ordinate with the flood hydrograph with responsible for reproducing total hydrograph for an event. But, it is certain that the rising part of the hydrograph is generated by the rainfall. Thus, to enable the Equation 3.1 to compute the Pearson correlation, r the rainfall data within the rising part of hydrograph was accumulated. Then, the time-series of accumulated rainfall and discharge were used in Equation 3.1 to compute the Pearson correlation, r value.

3.6 Methodology for the Selection of Variogram Model

The methodology of selecting the most appropriate variogram model for spatial rainfall mapping using the AHP is presented. First of all, the geostatistical analysis is performed on 7 daily rainfall data to obtain the variography parameters (sill, range and nugget) for 5 variogram model candidates (Gaussian, Exponential, Pentaspherical, Spherical, and Tetraspherical). These parameter values are used to generate a spatial rainfall map with the prediction error indicators (root- mean standardized error, mean

error, average standard error, root-mean-square standardized error and mean-square error). The prediction error indicators are set as criteria in the AHP method and serve to evaluate the variogram model's performance to produce the best spatial rainfall map. To select the best variogram model for spatial rainfall mapping, the AHP method is used to rank the best variogram models for the final decision.

3.6.1 Geostatistical Method

The geostatistical analysis is a way to study the environmental datasets related to space and time or the distribution of the datasets. Fundamentally it consists of three main elements. First, the candidate data are characterized based on the correlation of spatial data distribution using the variogram model. Second, the variogram model is applied to estimate the space-based data through an optimal interpolation method. Then, the distribution is simulated to generate the datasets for the space domain using the variogram model. Further details of the geostatistical analysis are explained by Journel and Huijbregts (1978), Goovaerts (1997) and Chiles and Delfiner (1999).

In modeling the spatial correlation of an experimental dataset (in this study it is daily rainfall data) using geostatistical analysis, Equation 2.3 is used. Equation 2.3 yields the semi-variance since the product on the right side is divided by 2. The semi-variance is plotted against lag to produce the semivariogram. Then the experimental dataset is fitted to a curve that represents the spatial correlation of the variography parameters. The experimental semivariogram is fitted with the 5 variogram model candidates used to calculate the variography parameters for spatial rainfall mapping using the Kriging interpolation method.

The Kriging interpolation variants, namely Ordinary, Simple, Universal, Indicator, Probability, Disjunctive and Co-Kriging, are associated with the geostatistical method. All of these Kriging variants quantify the spatial structure of the data and the prediction

error (Johnston et al., 2003). The selection of the Kriging method is dependent on the dataset used in this study. Since there is only one variable, that is the rainfall data used, and the constant mean of the dataset is assumed to be reasonable (Johnston et al., 2003), then the Ordinary Kriging (OK) method is adapted for the spatial interpolation stage. The geostatistical analysis is done using the ArcGIS 9.3 software toolbox.

3.6.1.1 Ordinary Kriging and Semivariogram Model

The OK method is used to estimate the value of spatial interpolation based on assumption that the observed data has constant mean but unknown within the study area (Johnston et al., 2003). The OK models the observed data using Equation 3.2;

$$R(s) = \mu + \varepsilon(s) \quad \text{Equation 3.2}$$

Where the $R(s)$ is the value for a location of (s) in coordinate of (x, y) , μ is the constant mean and $\varepsilon(s)$ is a random error.

To estimate the value R_{est} at the prediction location (s_o) , OK uses the weighted average λ_i of the observed value $R(s_i)$ as formed in Equation 3.3. The weight λ_i is calculated based on the distance of observed data to the prediction location and their spatial variation using the variogram model. The sum of all weights λ_i , must be equal to one to ensure that the predicted value is unbiased.

$$R_{est}(s_o) = \sum_{i=1}^m \lambda_i R(s_i) \quad \text{Equation 3.3}$$

Basically, the variogram models have similarity on the variography properties such as sill, range and nugget as illustrated in Figure 2.3. Theoretically, the semivariogram started at zero value. The measurement will be raised up to the sill where the line will be off or almost flat. However, due to the measurement error, the semivariogram has offset

value at the origin called the nugget effect. This nugget is a value of initial variability in the smallest group distance (lag). The sill value can be read at where the line off and partial sill value can be calculated by the residual of the sill and the nugget value. The range is the lag value extracted from the diagram at the sill location on the diagram which if beyond this range, the autocorrelation measure is zero.

The dissimilarity between the variogram models is due to how their mathematical equation presents the experimental data. The mathematical equations of variogram models are tabulated in Table 3.4. The spherical model presents the semivariogram curve linearly increasing at the early distance or lag and gradually change before it reaches the range. In contrast, the Gaussian model presents the parabolic form within the range value up to the sill. The Exponential model's curve is quite similar to Spherical model at the early distance but exponentially increasing to the sill value as the distance increase to the range. The Tetraspherical and Pentaspherical are differentiated by the mathematical formula from other variogram model. The Gaussian can be unstable numerically in the kriging interpolation if the nugget effect is not considered.

Table 3.4 Mathematical equations of the variogram models

Variogram Models	Equations
Spherical	$\gamma(h, \phi) = \begin{cases} \phi_s \left[\frac{3 \ h\ }{2 \phi_r} - \frac{1}{2} \left(\frac{\ h\ }{\phi_r} \right)^3 \right] & \text{for } 0 \leq \ h\ \leq \phi_r \\ \phi_s & \text{for } h \leq \ h\ \end{cases}$
Tetraspherical	$\gamma(h, \phi) = \begin{cases} \frac{2\phi_s}{\pi} \left[\arcsin\left(\frac{\ h\ }{\phi_r}\right) + \frac{\ h\ }{\phi_r} \sqrt{1 - \left(\frac{\ h\ }{\phi_r}\right)^2} + \frac{2\ h\ }{3\phi_r} \left(1 - \left(\frac{\ h\ }{\phi_r}\right)^2\right)^{\frac{3}{2}} \right] & \text{for } 0 \leq \ h\ \leq \phi_r \\ \phi_s & \text{for } \phi_r < \ h\ \end{cases}$
Pentasperical	$\gamma(h, \phi) = \begin{cases} \phi_s \left[\frac{15 \ h\ }{8 \phi_r} - \frac{5}{4} \left(\frac{\ h\ }{\phi_r}\right)^3 + \frac{3}{8} \left(\frac{\ h\ }{\phi_r}\right)^5 \right] & \text{for } 0 \leq \ h\ \leq \phi_r \\ \phi_s & \text{for } h \leq \ h\ \end{cases}$
Exponential	$\gamma(h, \phi) = \phi_s \left[1 - \exp\left(-\frac{3\ h\ }{\phi_r}\right) \right]$
Gaussian	$\gamma(h, \phi) = \phi_s \left[1 - \exp\left(-3 \left(\frac{\ h\ }{\phi_r}\right)^2\right) \right]$

Where,

$\gamma(h, \phi)$ = semi-variance

ϕ_s = sill

ϕ_r = range

h = lag (distance different) between dataset pair

3.6.1.2 Indicators of Spatial Rainfall Distribution Interpolation

The accuracy of the spatial rainfall map produced by the geostatistical method is evaluated by spatial interpolation error using a cross-validation approach. For this purpose, hold-one-out cross-validation technique was applied where a station is removed one by one and the rainfall magnitude is estimated using the variogram parameters. Then, the spatial interpolation errors are computed by Root-Mean-Square-Error (E_{rms}), Average Standard Error (E_{as}), Mean Standardized Error (E_{ms}) and Root-

Mean-Square Standardized Error (E_{rms}). The respective equations for each indicator are listed below.

$$E_{rms} = \sqrt{\frac{\sum_{i=1}^n (R_{est} - R_{obs})^2}{n}} \quad \text{Equation 3.4}$$

$$E_{as} = \sqrt{\frac{\sum_{i=1}^n \sigma_i}{n}} \quad \text{Equation 3.5}$$

$$E_{ms} = \frac{\sum_{i=1}^n \left(\frac{(R_{est} - R_{obs})}{\sigma_i} \right)^2}{n} \quad \text{Equation 3.6}$$

$$E_{rms} = \sqrt{\frac{\sum_{i=1}^n \left(\frac{(R_{est} - R_{obs})}{\sigma_i} \right)^2}{n}} \quad \text{Equation 3.7}$$

Where;

R_{obs} : Observed rainfall value at the rain gauge;

R_{est} : Interpolation estimated rainfall value at the rain gauge;

σ : Standard error of estimated value at the rain gauge;

n : Number of rain gauge station.

The best spatial rainfall map produced by the geostatistical method should have an appropriate value of each indicator. The E_{ms} should be close to zero but the E_{rms} near 1. However, the E_{rms} should be small and the E_{as} nears the E_{rms} value. Most researchers considered one or 2 indicators in their studies because these criteria are rarely achieved. In this study, this problem is solved using the multi-criteria decision-making tool.

3.6.2 Analytical Hierarchy Procedure (AHP)

Analytical Hierarchy Procedure (AHP) is adapted to execute the multi-criteria decision making. The AHP is a simple structured approach of the criteria and the alternative decisions in hierarchy form to analyze the decision-making process. It is able to consider both numerical and non-numerical forms of criteria. Based on this reason, the AHP has been applied in vast application either technical or non-technical fields.

The AHP method was applied to analyse the decision-making process to justify the best variogram model that produced an accurate spatial rainfall map by the geostatistical method. By using the Kriging interpolation technique, 5 alternatives of variogram models namely Spherical, Tetraspherical, Pentaspherical, Exponential, and Gaussian are structured for evaluation based on 4 spatial interpolation indicators (E_{rms} , E_{as} , E_{ms} and E_{rmss}).

The Analytical Hierarchy Procedure (AHP) was proposed by Thomas L. Saaty in the 1970s. The AHP is a methodology of multi-criteria decision-making for qualitative or quantitative study through an evaluation of a set of variables in the hierarchical structure (Saaty, 1990). It consists of 4 evaluation stages in the hierarchical structures, i.e. modeling, assessment, ranking and conclusion. In the modeling stage of the hierarchical structure, the study objective is placed at the top of the structure. Meanwhile, the criteria, sub-criteria and alternatives are structured at the bottom. To rank the alternatives for a decision, the criteria and sub-criteria are evaluated to set the priority weights used to assess the alternatives' attributes.

3.6.2.1 Priority Weight of AHP's Criteria

The aim of this analysis is to select the best variogram model from 5 candidate models based on 4 prediction error criteria of spatial rainfall mapping. The AHP model used in this study is shown in Figure 3.5. Based on this figure, two steps of pairwise

matrix evaluation are involved, whereby the first step is to evaluate the pairwise matrix of 4 prediction error criterion of spatial rainfall mapping and the second step is for the 5 candidate models of semivariogram.

In the first step of pairwise matrix evaluation, the preference for each criterion is compared using the pairwise matrix comparison (Saaty, 1980). A comparative scale with numerical values of 1 to 9 represents the influence of one criterion on another (Saaty & Vargas, 1991). The criteria are labelled C1 for E_{rms} , C2 for E_{as} , C3 for E_{ms} and C4 for E_{rms} . These criteria are compared and tabulated in Table 3.5 and the assignment of numerical values is based on the best output of spatial rainfall mapping. To achieve the best spatial rainfall mapping (Johnston et al., 2003), criterion C2 must be the same as C1, which is the smallest value. Criterion C4 must be nearest to 1 but criterion C3 must approach 0.

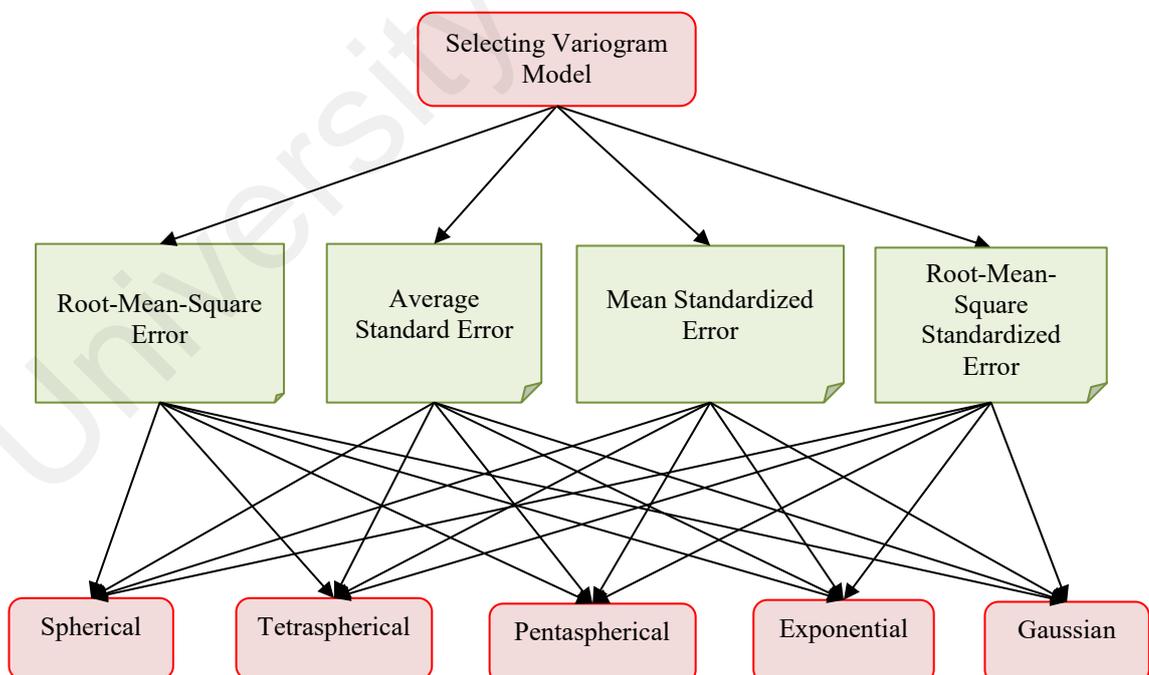


Figure 3.5: AHP model for semivariogram model selection.

Table 3.5: Comparative weighting score for pairwise matrix comparison.

Criteria	Comparative Weighting Score																Criteria	
	More importance than								Equal	Less importance than								
C1	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	C2
C1	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	C3
C1	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	C4
C2	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	C3
C2	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	C4
C3	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	C4

In a real case, spatial rainfall mapping that fulfils all these criteria is difficult to achieve. Furthermore, the Consistency Ratio (*CR*) of the pairwise matrix must be less than 0.1 to ensure the priority weights (*PW*) of the criteria are reliable for analysis and it is the measurement of the consistency of the decision as structured by the AHP. The *PW* is calculated by first assigning the comparative numerical value of the criterion. For instance, in Set 1, a comparative numerical value 4 is assigned for C1 over C3. This meant that C1 is more important than C3. In contrast, a comparative numerical value 0.11 (or 1/9) is assigned for C1 over C4. This meant that the C4 is greatly important than C1. After the pairwise matrix is created, each comparative value is normalized using the sum of each column according to Equation 3.8, where the X_{ij} is the normalized comparative numerical value, C_{ij} is the comparative numerical value, $\sum_{i=1}^n C_{ij}$ is the sum of the column for the comparative value and notation of i and j is the row and column, respectively.

$$X_{ij} = \frac{C_{ij}}{\sum_{i=1}^n C_{ij}} \quad \text{Equation 3.8}$$

Then, the PW is calculated by averaging the normalized value of the comparative value for each row using **Equation 3.9**.

$$PW_{ij} = \frac{\sum_{j=1}^n X_{ij}}{n} \quad \text{Equation 3.9}$$

Where the $\sum_{j=1}^n X_{ij}$ is the sum of the normalized comparative value of row and n is the number of criteria.

The CR value is computed using Equation 3.10, where the CI is the Consistency Index which calculated using Equation 3.11 and the RI is the Random Index. The RI value is representing the random consistency of developed pairwise matrix and the value is based on the number of criteria in AHP structure as tabulated in Table 3.6. Since the number of the criteria in this study is 4, then the RI value is 0.9. The λ_{max} in Equation 3.11 is the value of the sum of the weighted sum of the pairwise matrix with the PW of each criterion.

$$CR = \frac{CI}{RI} \quad \text{Equation 3.10}$$

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad \text{Equation 3.11}$$

The numerical value selection of a pairwise matrix is crucial because it determines the consistency and robustness of the decision made. To achieve this, the pairwise matrix comparison using comparative numerical is simulated until the CR value is less than 0.1. As a result, there are 3 sets of pairwise matrices with different comparative numerical values and the associated priority weights are used in the analysis. A detailed calculation of PW and the CR is given by Coulter, Coakley, and Sessions (2006), Ishizaka and Lusti (2006) and Saaty (2008).

Table 3.6: Random Index for a number of criteria (Saaty, 1980)

Number of criteria, n	1	2	3	4	5	6	7
RI	0	0	0.58	0.9	1.12	1.24	1.32

3.6.2.2 Semivariogram Model Evaluation

The second step of pairwise matrix evaluation is to rank the candidate models of semivariogram using the PW of each criterion. The alternatives (variogram models) are evaluated using the criterion value and the PW value of the criterion in the first step. The criterion value of each alternatives is re-scale to the comparative scale of Saaty (1980), thus transforming the values to be dimensionless. Then the transformed value of the criterion for each alternative is used in the pairwise matrix evaluation of the candidate models of semivariogram to get the alternative's priority weight (APW) using Equation 3.8 to Equation 3.11.

To rank the alternatives for decision making, Equation 3.12 is used to calculate the alternatives' weighted priority (AWP) value, where the AW_i is the alternative's priority weight and n is the number of criteria. The PW_i is the priority weight value for each criteria. The alternatives' rank position is produced by sorting the AWP values from highest to lowest.

$$AWP = \sum_{i=1}^n AW_i \cdot PW_i \quad \text{Equation 3.12}$$

The AHP method produced a rank for decision making for a single data set. However, in this study 7 rainfall data sets are used to select the best variogram model. The AHP results have different rank set of alternatives for each data set and it is unable to make a decision from multi-datasets. To overcome this problem, a new set of scores is introduced by multiplying the probability (f_r) of the alternatives being placed at a rank position with S_r value. The value f_r is computed by counting the frequency of the

alternatives placed at a rank position divided by the total number of the dataset (7 in this study). The S_r value is a factor that consists of a set of numbers in descending order from 5 (number of alternatives) to 1. The S_r value is assigned according to a rank position as tabulated in Table 3.7. The Final Score is then calculated by summing up the score as in Equation 3.13 where the r is the rank position. The Final Score value is sorted in descending order to produce the final rank for decision making. The best model to be used is the alternative that ranked in the top place.

$$Final\ score = \sum_{r=1,2,3\dots} f_r \cdot S_r \quad \text{Equation 3.13}$$

Table 3.7: S_r value for each rank position

Rank Position, r	1	2	3	4	5
S_r	5	4	3	2	1

3.7 Methodology for Rain Gauge Network Optimization

This section presents a discussion on the methodology for rain gauge network optimization used to obtain the optimal rain gauge network. It begins with an explanation of each method, followed by a detailed explanation of the proposed method application. Two new methods were introduced, namely, 1) cross-validation technique coupled with the geostatistical method and 2) MPSO method to optimize the number of rain gauge stations in the rain gauge network studied based on daily rainfall data. All developed algorithms are coded and analysed using MATLAB version R2013a, software by MathWork.

3.7.1 Coupling Cross Validation and Geostatistical (CV-Geo)

This section discusses the cross-validation technique and geostatistical method used to obtain the optimal rain gauge network. It begins with an explanation of each method, followed by a detailed explanation of the proposed method application. In the proposed method, the cross-validation technique was coupled with the geostatistical method (CV-Geo) to optimize the number of rain gauge stations in the rain gauge network studied based on daily rainfall data.

Generally, the geostatistical analysis of different network configurations will produce different values of variography parameters and its associated spatial interpolation error. This concept was used to optimize the rain gauge number in the studied network. Also, the rain gauge network was configured with different combinations of rain gauge number using two different techniques adapted from the cross-validation approach, Leave-One-Out (LOO) and Add-One-In (AOI) whereby adapted from Bastin et al. (1984) and Kassim and Kottegoda (1991). Each network was analyzed to compute the variography parameters (sill, range, and nugget) and the spatial interpolation error. The rain gauge network with the lowest error value is usually identified as an optimized rain gauge network. In other words, the best rain gauge network configured.

3.7.1.1 Leave-One-Out and Add-One-In Cross Validation Technique

LOO cross-validation is commonly used to evaluate the performance of variables in a dataset. It involves a simple process of leaving out a variable from the dataset temporarily and evaluating the remaining variables for their performance. This process is repeated until all variables are evaluated and a conclusion is drawn regarding their performance.

The LOO approach was employed in this study for optimization to generate a rain gauge network with 25 stations. Throughout the optimization process, the intention was

to remove one of the hypothetically ineffective rain gauge stations at a time before being combined into the existing network (as a candidate optimum rain gauge network) for evaluation based on the optimization criterion. It was assumed that the stations omitted in every repetition were unrelated to each other to produce a better optimized network. However, each station is in fact very important in a rain gauge network to produce an accurate spatial rainfall distribution. Thus, the LOO output required validation and in order to do so, the AOI cross-validation technique was introduced.

AOI is essentially opposite of LOO. If LOO is intended to remove one station from the dataset, AOI is executed to temporarily transfer one station at a time from the dataset to be combined into the existing network (as an optimum rain gauge network candidate) prior to optimization criterion evaluation. It was assumed that the added stations should remain in the optimized network. Despite both techniques having different assumptions, it is essential to evaluate the results produced by both techniques for an unbiased decision.

3.7.1.2 Geostatistical and Spatial Rainfall Interpolation

At this stage, the geostatistical application is as explained in section 3.6.1. Meanwhile, the semivariogram model used is adopted from the result of methodology in section 3.6. The model used to fit the experimental dataset to calculate the semivariogram properties. Fitting the experimental dataset semivariogram to the variogram model is an important stage in the geostatistical analysis.

To generate a smooth semivariogram curve, the Least Squares Method (LS) was adopted to fit the selected semivariogram model to the dataset. LS is a common method of fitting the spatial dataset to the candidate variogram model (Yoon & Lahiri, 2002). LS employs a simple estimation of the sum of squared error, E_{ss} , between the semivariance of the dataset and the semivariance estimated by the variogram model

using variography properties. The aim of LS is to find the best variography property values (sill, nugget and range) of the spherical model that minimize E_{ss} in Equation 3.14 to produce the best semivariogram curve. The estimated variography properties are used to calculate E_{rms} through the spatial interpolation method.

$$E_{ss} = \sum_{k=1}^l \left(\gamma(h)_k - \gamma(h, a)_{sph_k} \right)^2 \quad \text{Equation 3.14}$$

Spatial interpolation was applied to re-estimate the rainfall value at the measured point using LOO cross-validation prior to calculating E_{rms} . In this study, the Ordinary Kriging (OK) method was used to carry out spatial rainfall interpolation based on the assumption that the observed rainfall data had a constant mean but was unknown within the study area as explained in section 3.6.1.1.

3.7.2 Modified Particle Swarm Optimization (MPSO)

In this study, the MPSO was developed from the standard version of PSO to solve the rain gauge network optimization problem. The standard version of PSO algorithm relies on the velocity (v_i) and the position of the swarm particles (x_i) that was calculated using Equation 2.7 and Equation 2.8, respectively. Originally, both equations are applied to the real number problem. However, the rain gauge network optimization is a discrete problem, where an optimum network is selected from all possible combinations of a number of rain gauge stations. For instance, if 10 stations are selected from 25 stations, all possible combinations will be from 1 to 3,268,760, which are positive integer number. In this study, the set of positive integer number is set as determinant of the candidate optimum networks or domain of solution for PSO. Based on this reason, the original PSO is modified in this study. The modification process is

explained in next sub-section to enable that the PSO algorithm works for rain gauge network optimization case.

3.7.2.1 Modification on PSO Equation

The previous studies classified this problem as a discrete domain problem. Kennedy and Eberhart (1997) solved this problem using the binary-based approach. The algorithm was modified to ensure that the changes in the variables were updated, either zero or one in each iteration. It was a simple modification but had a significant improvement for the PSO performance that simply converted the velocity of the particles into integer form by the *round(.)* operator. This concept was applied by X. H. Shi et al. (2007) to solve the traveling salesman problem and the generalized traveling salesman problem. In their algorithm, the velocity was rounded to the nearest integer value and this solves the studied problems which involved the permutation of the salesman locations along his travel.

Thus, in this study, the first modification was to suit the PSO algorithm for rain gauge network optimization problem. A similar approach as in X. H. Shi et al. (2007) was applied by implying the round operator (*round(.)*) into Equation 2.7. However, the *round(.)* operator was introduced for each component in Equation 2.7. This was to ensure that the efficient interactions of cognitive and social components are preserved while exploring the swarm for the best position. This modification was represented in Equation 3.15. Meanwhile, the populations were still updated using Equation 2.8.

$$\begin{aligned}
 v_{p=1,2,\dots,N}^{j+1} &= \text{round}(w \times v_{p=1,2,\dots,N}^j) \\
 &+ \text{round}\{c_1 \times r_1 \times [p_{pb} - x_{p=1,2,\dots,N}^j]\} \\
 &+ \text{round}\{c_2 \times r_2 \times [p_{gb} - x_{p=1,2,\dots,N}^j]\}
 \end{aligned}
 \tag{Equation 3.15}$$

The first modification did not overcome the premature convergence issue. Therefore, the second modification was introduced to improve the modified PSO algorithm for better convergence behaviour. In the literature, extensive research was carried out to improve the convergence rate of PSO by modifying the value of the PSO's parameters. For instance, Safaei et al. (2012) used the time-varying value for inertia weight, cognitive and social coefficient. In their study, the initial and last value of inertia weight was [0.9, 0.4] and for the social coefficient was [2.5, 0.5]. In contrast, the cognitive coefficient used varying intervals of [0.5, 2.5]. In another study by Eberhart and Shi (2001), they used the value of w which randomly varied in the range of [0.5, 1.0] that was calculated by $[0.5+(\text{rand}/2)]$ and the c_1 and c_2 value were fixed as 1.494. These parameter values were found successful to track the optimization task of the dynamic system in their study.

Motivated by the randomly varying concept, in this study, a randomized value was adopted for all parameters in the modified PSO algorithm to improve the convergence rate and learning rate of the cognitive and social component for a better solution. The range of values of the parameters was set up based on the established values in the literature for each parameter. The maximum value of the parameters was adopted from Ravizi (2012), where the maximum value used for w was 0.9, while for c_1 and c_2 , was 2.5, respectively. The minimum value for the parameters of w was set as 0.1. On the other hand, the minimum value for the parameters value of c_1 and c_2 were set as 1.5 that is adopted and rounded up from Eberhart and Shi (2001). Based on these modification processes, the final version of the modified PSO is called MPSO.

The MPSO algorithm was evaluated prior to use for the rain gauge network optimization process. For this purpose, the MPSO algorithm was compared to the standard modified PSO (SPSO) where this algorithm only used Equation 3.15 and

Equation 2.8 with the standard value of the parameters. In the evaluation process, the SPSO and MPSO were renamed as SPSO-IP and MPSO-IP, respectively where ‘-IP’ used to denote that the algorithms work for the positive integer number. The parameters values for both algorithms are tabulated in Table 3.8.

Table 3.8: Parameters value of SPSO-IP and MPSO-IP

Parameter	Description	SPSO-IP	MPSO-IP
Domain	Variables form	Integer	Integer
N	Particle number	25 and 50	25 and 50
$Itermax$	Maximum number of iterations	1000 and 2000	
w_{min}	Minimum inertia weight	0.4	Randomly from 0.1 to 0.9
w_{max}	Maximum inertia weight	0.9	
c_1	Cognitive coefficient	2	Randomly from 1.5 to 2.5
c_2	Social coefficient	2	Randomly from 1.5 to 2.5

3.7.2.2 Evaluation process

In the evaluation process, both PSO algorithms were applied to minimize five common test functions, namely Rosenbrock (TF_1), Cube (TF_2), Ackley (TF_3), Griewank (TF_4) and Sphere (TF_5) for their optimum best value. All of these test functions yield 0 as the optimum value. The equations of each test function are as stated in Equation 3.16 to Equation 3.20.

$$TF_1 = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \quad \text{Equation 3.16}$$

$$TF_2 = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^3)^2 + (x_i - 1)^2] \quad \text{Equation 3.17}$$

$$TF_3 = -20 \exp \left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \right) - \exp \left(\frac{1}{n} \sum_{i=1}^n \cos 2\pi x_i \right) + 20 + e \quad \text{Equation 3.18}$$

$$TF_4 = \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \quad \text{Equation 3.19}$$

$$TF_5 = \sum_{i=1}^n x_i^2 \quad \text{Equation 3.20}$$

The testing was set up for 10 dimensions problem condition in a range of -30 to 30 for each dimension. The tests were conducted for a different number of iterations and the number of swarm particles. By using two scenarios of swarm particles size (25 and 50 swarm particle), the tests were executed for 1000 and 2000 iterations. Meanwhile, the maximum velocity is set at 10 per cent of the population size as applied by Xuedan, Qiang, Haiyan, and Lili (2009). The detailed specifications of the test function are tabulated in Table 3.9.

Table 3.9: Testing Function parameters

Test Function	Function Name	Dimension (n)	Range	f(x) Optimum value,	Solution at Optimum Value, (x)
TF ₁	Rosenbrock	10	[-30,30] ⁿ	0	1,1,1,1,1,1,1,1,1,1
TF ₂	Cube	10	[-30,30] ⁿ	0	1,1,1,1,1,1,1,1,1,1
TF ₃	Ackley	10	[-30,30] ⁿ	0	0,0,0,0,0, 0,0,0,0,0
TF ₄	Grienwank	10	[-30,30] ⁿ	0	0,0,0,0,0, 0,0,0,0,0
TF ₅	Sphere	10	[-30,30] ⁿ	0	0,0,0,0,0, 0,0,0,0,0

Two stages of performance evaluation were involved, at the first stage, both algorithms were executed for a single run to optimize the test function under the controlled values of random numbers of r_1 and r_2 . In other words, both algorithms at each iteration were applied using the same value of the random numbers r_1 and r_2 . This stage was important in order to evaluate the algorithm's ability to converge to an optimized test function value in equal condition.

At the second stage, the algorithms were executed for a multi-consecutive run without controlling the random value c . The best result from the multi-consecutive run was selected to represent the optimization result. In addition, statistical analysis and hypothesis t-test was carried out on the multi-run result to justify the performance of the MPSO-IP algorithm. For this purpose, 50 multi-consecutive runs for both algorithms were adopted for sufficient statistical data analysis, as done by Yang, Yuan, Yuan, and Mao (2007).

3.7.3 Application of Proposed Optimization Methods

Prior to optimization, the existing rain gauge network (55 stations) was divided into two datasets: 1) a dataset containing the rain gauge stations being evaluated (25 stations) and denoted by $[R=(r_1, r_2, r_3, \dots, r_m)]$ and 2) a dataset of the remaining rain gauge stations (30 stations) denoted by $[E=(e_1, e_2, e_3, \dots, e_n)]$. The E dataset contains 11 stations located outside the study area. These stations are important for ensuring the continuity of the spatial interpolation of the distributed rainfall within the study area, especially at the catchment boundary. The interpolation of spatial rainfall distribution at the catchment boundary would be lost if the stations were not considered, which could increase the uncertainty and error of the rainfall distribution spatial interpolation.

Every rain gauge station in each dataset contains spatial information of the longitude (x), latitude (y) and rainfall magnitude (z). A rain gauge station in each dataset is denoted as follows:

$$r_m = (x_m^r, y_m^r, z_m^r) \quad \text{Equation 3.21}$$

$$e_n = (x_n^e, y_n^e, z_n^e) \quad \text{Equation 3.22}$$

Where m and n are the numbers of stations in the datasets, while r and e are the individual stations in the datasets, respectively.

The main objective of the optimization is to prioritize the rain gauge stations in the R dataset to produce an optimum number of rain gauge stations in the network. This task involves two stages of evaluation: 1) generating a candidate optimum rain gauge network using the stations selected from the R dataset and combining them with the E dataset, and 2) evaluating the candidate network based on the evaluation criteria by using the spatial information (x, y, z) of each station.

3.7.3.1 Application of CV-Geo Method

In the first stage of the optimization process, the stations were selected through LOO and AOI cross-validation. LOO was aimed to determine the stations that are less important in the optimal network. In contrast, AOI was assigned to determine the stations that are more important in the optimal network. The optimal network candidates generated by LOO and AOI were evaluated next using the geostatistical method for spatial interpolation error (E_{rms}) based on the variography parameters.

The geostatistical analysis of different network sizes produced different variography parameter values and associated spatial interpolation errors (E_{rms}). By manipulating this relationship, optimization was carried out to determine the optimum rain gauge network for every number of rain gauges selected based on the lowest E_{rms} . In other words, the E_{rms} value was set as an optimization criterion in the optimization task. The methodology employed to evaluate the optimized rain gauge network in this study is summarized and presented in Figure 3.6.

3.7.3.2 Application of MPSO Method

Application of the MPSO method has a different approach from the previous method to create rain gauge network candidates and to optimize it. The MPSO maintains the original PSO algorithm that is creating the domain of solution before the optimization process is executed.

To generate a domain of solution, all possible combinations of rain gauge station based on the number of rainfall stations required relative to the total number of rain stations evaluated (set R) is created. Each combination of optimum network candidates is generated where non-repeated rain gauge station is allowed. These optimum network candidates are distinguished by their position in the domain of solution. For example, to get a domain solution for networks that contain 10-station out of 25 stations, 3 268 760 candidates of optimum network generated and for each network, the stations were combined non-repeatedly. The network candidates were distinguished by their position from 1 to 3 268 760 as an ID solution as illustrated below:

Table 3.10: ID of candidate rain gauge network of a domain solution for networks that contain 10-station out of 25 stations

ID solution	1	2	3	4	5	6	7	...	3268760
rain gauge station	r1	...	r16						
	r2	...	r17						
	r3	...	r18						
	r4	...	r19						
	r5	...	r20						
	r6	...	r21						
	r7	...	r22						
	r8	...	r23						
	r9	...	r24						
	r10	r11	r12	r13	r14	r15	r16	...	r25

The optimization process is executed to optimize the rain gauge network that has 1 to 25 stations from the R dataset, one after another. The domain of solution for each size of networks is then generated as explained in the previous paragraph. Next, the MPSO algorithm is applied to optimize (reducing the objective function value) that is analyzed using the geostatistical method as explained in Chapter 2 and subchapter 3.6.2 to obtain an optimum network for a respective number of stations selected. This process is repeated for a network size of 1 to 24 stations to obtain an optimum rain gauge network for each network size based on the lowest spatial interpolation error (E_{rms}). The summary of the MPSO application is summarized in Figure 3.7.

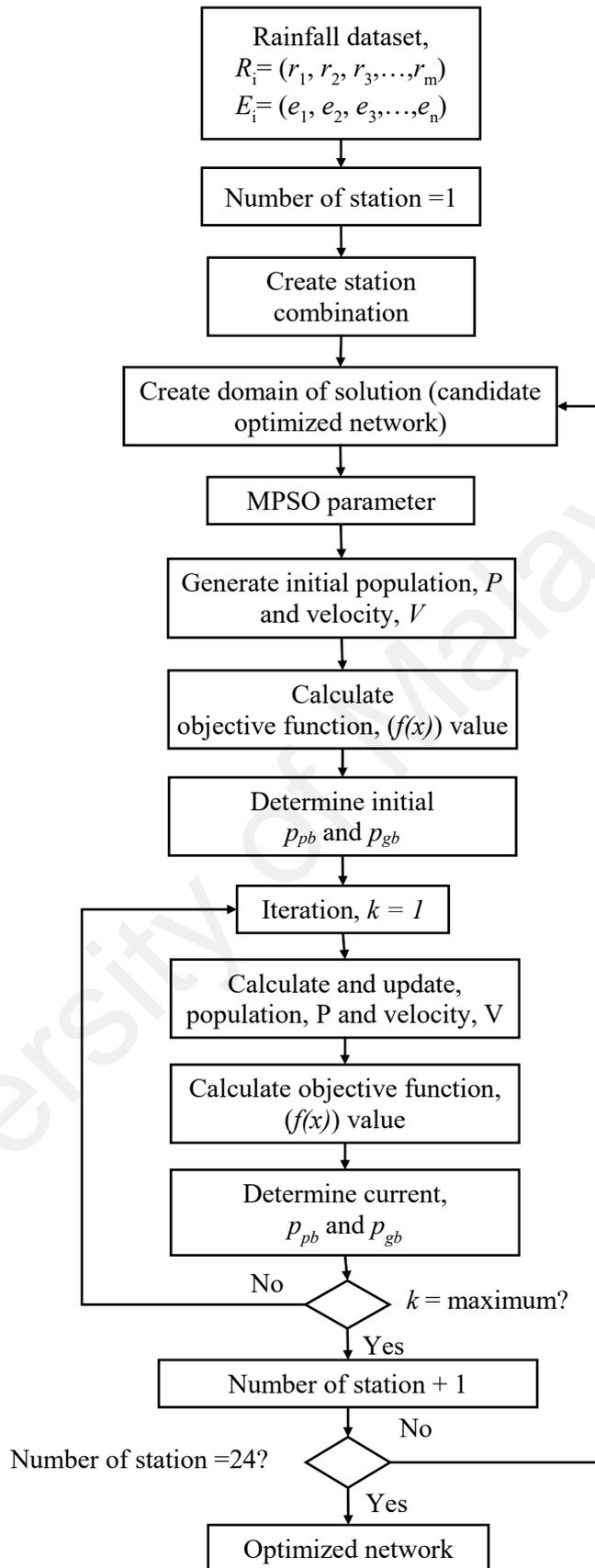


Figure 3.7: Summary of the MPSO method employed to optimize the number of rain gauges in the network.

3.7.3.3 Objective Function of optimization

The configured rain gauge networks are evaluated using spatial interpolation error. The evaluation objective is to produce the networks with less spatial interpolation error and this is measured as Root-Mean-Square-Error (E_{rms}). The E_{rms} is adapted to be the performance indicator in network evaluation and optimization. The E_{rms} value is minimized in both methods.

To calculate the E_{rms} value, considering a network to be analyzed, one by one station in the network is omitted and the rainfall value at the omitted point is estimated using the value of variography model parameters. Using these estimated values the E_{rms} is calculated using Equation 3.23 below. An effective network should produce E_{rms} value close to zero or the lowest as compared with other networks. The E_{rms} is expressed as:

$$E_{rms} = \sqrt{\sum_{i=1}^n (R_{obs} - R_{est})^2} \quad \text{Equation 3.23}$$

Where;

R_{obs} : Observed rainfall value at the rain gauge;

R_{est} : Interpolated estimate rainfall value at the rain gauge;

n : Number of rain gauge station.

3.8 Verification of Optimized Rain Gauge Network by Hydrological Tank Model

The optimized rain gauge network is verified using hydrological lump-model to confirm the reliability of rainfall data produced to simulate the observed flow. In the verification stage, the effect on the hydrologic model efficiency is examined by different input rainfall datasets.

For this purpose, the Modified Tank Model by DID is employed as the hydrological model. Figure 3.8 illustrates the verification process of the optimized rain gauge network. In the model development stage, the Tank Model for Upper Klang river basin (TM-UKRB) is developed through an automated calibration technique using PSO method to obtain the appropriate TM-UKRB's parameters value. Then, the model is validated based on the model efficiency to simulate the observed flow. The TM-UKRB with calibrated parameters is used to simulate the observed flow using the rainfall data sets produced by the existing network and optimized network to compare the results.

University of Malaya

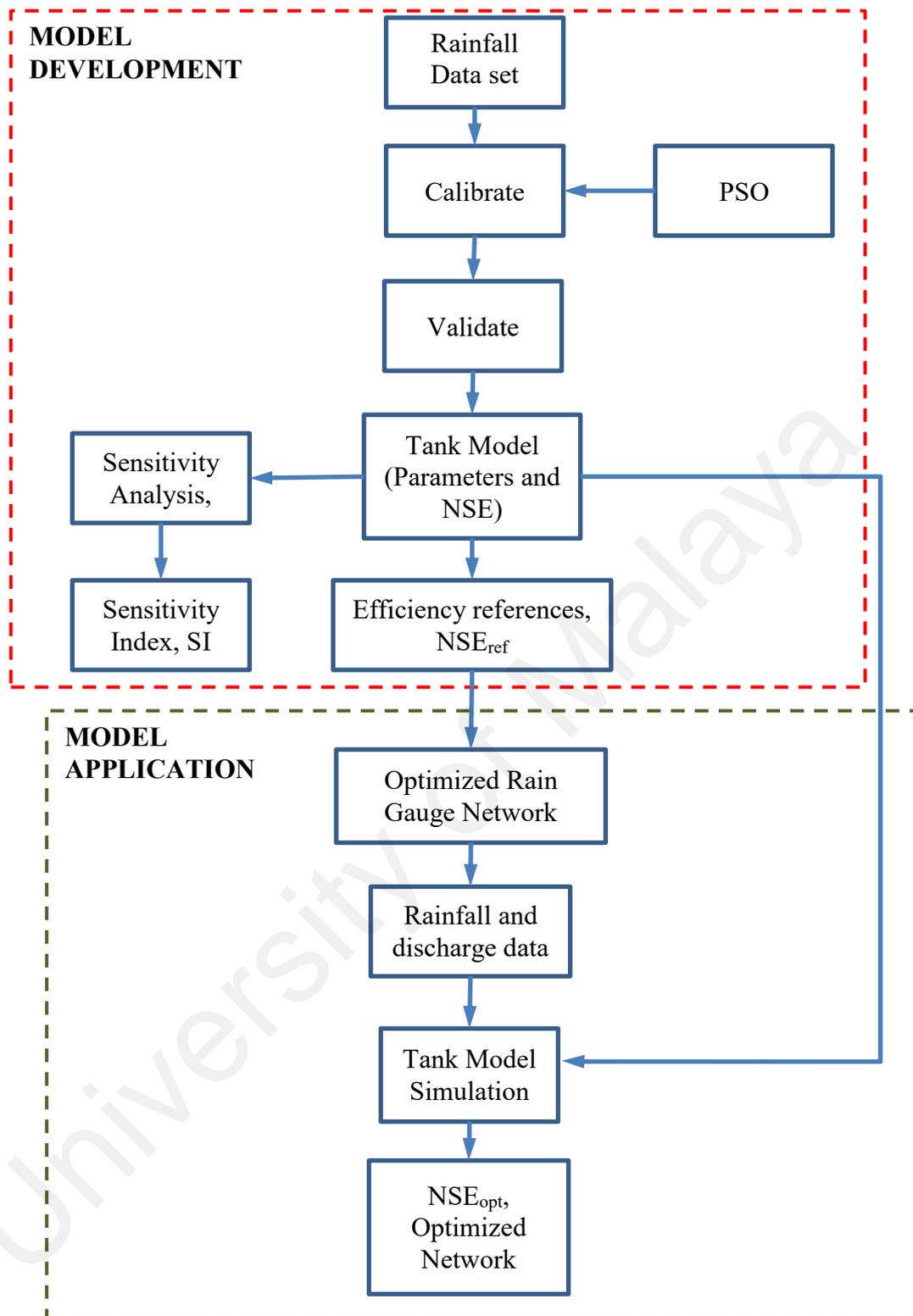


Figure 3.8: Flow chart of Tank Model development and application for verification of optimized rain gauge network

3.8.1 Tank Model Development

The development of TM-UKRB involves three processes such as model set up, calibration and validation of the model. The processes are explained in the following section.

3.8.1.1 Tank Model Set Up

The TM-UKRB was set up using three tanks organized in vertical order as depicted in Figure 3.9. The top tank represents the topsoil layer and rainfall as an input. It modeled the hydrological process that consists of rainfall, surface evaporation, infiltration into the next soil layer, storage volume, maximum storage depth and surface runoff. The subsequent tank represents the intermediate soil layer. The infiltration from the top tank is the main input to the tank's storage. This tank modeled the interaction between the input and intermediate runoff as well as infiltration that percolated into the last tank that counts the tank storage constricted by its maximum depth of storage. The last tank represents the base flow generated by the storage in the tank as the residue of the infiltration from the intermediate tank and the available storage without further infiltration. The combination of surface runoff, intermediate flow and base flow is the total flow of the river basin.

The TM-UKRB has 12 parameters that influence the hydrological process to produce the total flow of the river basin from rainfall, as tabulated in Table 3.11. The total flow is calculated using the parameters in the hydrologic process which is computed as follows:

Step 1: The rainfall (RF) over the river basin (Tank 1) will be affected by evaporation losses (Ev1) and increases the initial tank storage (TS1) up to the maximum storage depth (X1). If the storage in the Tank 1 exceeds the X1, then the surface flow (Q1) will be produced by the tank, otherwise, the flow is zero. The current

storage (TS1) in the tank is the residue of storage by the Q1. The infiltration (I1) into the second tank is calculated by multiplying the infiltration coefficient (K4) by the current tank storage. This process is explained in Equation 3.24, Equation 3.25 and Equation 3.26, respectively.

$$TS1_0 = RF - Ev1 + TS1 \quad \text{Equation 3.24}$$

$$Q1 = K1(TS1 - X1)^M \text{ if } TS1 > X1, \text{ otherwise } Q1 = 0 \quad \text{Equation 3.25}$$

$$I1 = K4(TS1 - Q1) \quad \text{Equation 3.26}$$

Step 2: The infiltration (I1) from the first tank will be percolated into the second tank storage (TS2) and increases the current storage. If the tank storage more than the maximum storage depth (X2), the intermediate flows (Q2) will be generated out of the second tank. The current storage in the tank is the residue of storage by the Q2. The infiltration (I2) into the second tank is calculated by multiplying the infiltration coefficient (K5) by the current tank storage. This process is explained in Equation 3.27, Equation 3.28 and Equation 3.29, respectively.

$$TS2_0 = I1 + TS2 \quad \text{Equation 3.27}$$

$$Q2 = K2(TS2 - X2)^M \text{ if } TS2 > X2, \text{ otherwise } Q2 = 0 \quad \text{Equation 3.28}$$

$$I2 = K4(TS2 - Q2) \quad \text{Equation 3.29}$$

Step 3: The infiltration (I2) from the second tank will be percolated into the last tank storage (TS3) and increases its current storage. The baseflow (Q3) is generated by multiplying the runoff coefficient (K3) by TS3 as in Equation 3.31.

$$TS3 = I2 + TS3_0 \quad \text{Equation 3.30}$$

$$Q_3 = K_3 \cdot TS_3$$

Equation 3.31

Step 4: The total flow (Q_T) out from the river basin is the summation of the Q_1 , Q_2 and Q_3 .

$$Q_T = Q_1 + Q_2 + Q_3$$

Equation 3.32

All elements in the hydrological process of tank model are measured in millimetre (mm). Thus, to convert the flow into the System International unit (m^3/s), Q_T is multiplied by the river basin area and divided by the time interval in second (s). This hydrological process is coded and simulated using MATLAB version R2013a, software of MathWork.

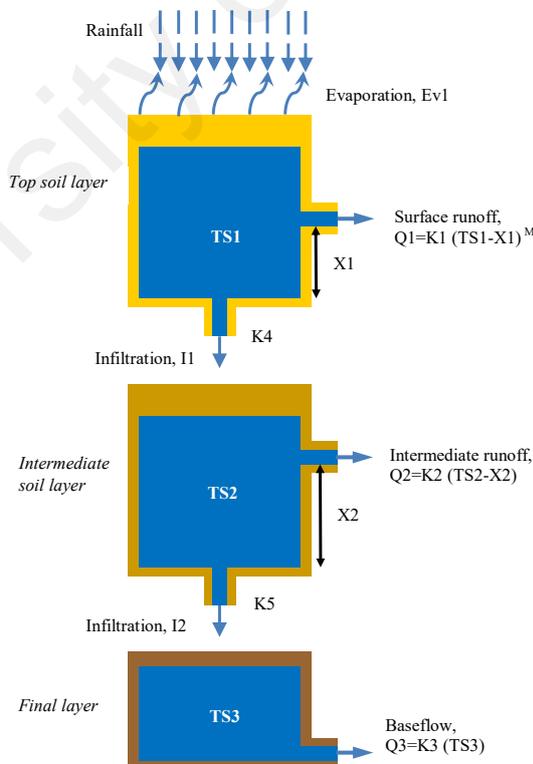


Figure 3.9: Tank Model set up for TM-UKRB

Table 3.11: TM-UKRB model's parameters

No.	Parameter	Description
1	TS1	Storage coefficient in Tank 1
2	TS2	Storage coefficient in Tank 2
3	TS3	Storage coefficient in Tank 3
4	Ev1	Evaporation of water body in Tank 1
5	M	Power coefficient of storage to calculate the direct runoff in Tank 1
6	X1	Level of storage in Tank 1 just before the Q1 generated
7	X2	Level of storage in Tank 2 just before the Q2 generated
8	K1	Runoff coefficient in Tank 1
9	K2	Runoff coefficient in Tank 2
10	K3	Runoff coefficient in Tank 3
11	K4	Infiltration coefficient for infiltration occurred from Tank 1
12	K5	Infiltration coefficient for infiltration occurred from Tank 2

3.8.1.2 Calibration and Validation of TM-UKRB Parameters

The areal rainfall data series in the 15-minute interval is used as input data to the TM-UKRB. It is calculated using the Thiessen Polygon method by rainfall data series of the rain gauge network. The rainfall data series is used to simulate flow at the calibration point which is located at the final outlet of the river basin (at station N21, Sg Klang at Jalan Klang Lama). Based on the seven flood events as given in Table 3.1, only four flood events have sufficient discharge data that can be utilized for calibration and validation of the model as tabulated in Table 3.12. The data dated on 18 September 2011 is used for calibration and the rest of data are used for the validation process.

Table 3.12: Calibration and validation data

Date	Period of simulation
3 March 2009	3 March 2009, 00:00am to 4 March 2009, 06:00am
18 September 2011	18 September 2011, 00:00am to 18 September 2011, 06:00am
18 April 2012	18 April 2012, 00:00am to 19 April 2012, 06:00am
21 August 2012	21 August 2012, 00:00am to 22 August 2012, 06:00am

The TM-UKRB is calibrated and validated based on the Nash-Sutcliffe efficiency index (NSE) to ensure the reliability value of the model parameters (Nash & Sutcliffe, 1970). The NSE is calculated as:

$$NSE = 1 - \frac{\sum(Q_i - F_i)^2}{\sum(Q_i - Q)^2} \quad \text{Equation 3.33}$$

Where,

Q_i =the simulated flow at time i

Q = the average observed flow, $\frac{\sum Q_i}{N}$

F_i =observed flow at time i

N = number of data

The automatic calibration approach is employed using the SPSO method. The SPSO is used to maximize the objective function that is the NSE of the model. The SPSO specification for the calibration process is set as follows:

- a. 12 dimensions to represent 12 model parameters
- b. maximum number of iterations of 1000
- c. 25 number of swarm particles
- d. the lower and the upper limit of each parameter are tabulated in Table 3.13 below.

Table 3.13: The lower and upper limit of model parameters

Parameter	Lower limit	Upper limit
TS1	0	35
TS2	0	25
TS3	0	20
Ev1	0	1
M	1.001	1.6
X1	0	25
X2	0	15
K1	0.0001	0.3
K2	0.0001	0.3
K3	0.0001	0.3
K4	0.0001	0.3
K5	0.0001	0.3

The calibrated parameters of TM-UKRB are evaluated via sensitivity analysis (SA). The SA is carried out to assess which parameters influence the model efficiency most. It is also important for a better evaluation of the parameters and better estimation of model output to reduce the uncertainty in the model (Lenhart, Eckhardt, Fohrer, & Frede, 2002). The SA is conducted by using one-at-a-time method (OAT) where the sensitivity of the parameters are calculated by changing a parameter value independently at a time while the rest of parameters are held at the calibrated value (Hamby, 1995). The range of changing for the value of the parameter that is from -50% to +50% of the calibrated value is used to compute the NSE value. The Sensitivity Index (SI) is used for sensitivity measurement (Hoffman & Gardner, 1983) and calculated based on Equation 3.34, where i is the number of parameters.

$$SI_i = \frac{NSE_i^{max} - NSE_i^{min}}{NSE_i^{max}} \quad \text{Equation 3.34}$$

In addition, the Pearson correlative, r and coefficient of determination, R^2 of the simulated hydrograph and the observed hydrograph is used to evaluate the model

efficiency. The R^2 is calculated using Equation 3.35, where p is the simulated value and q is the observed value.

$$R^2 = \frac{n \sum pq - \sum p \cdot \sum q}{\sqrt{[n \sum p^2 - (\sum p)^2][n \sum q^2 - (\sum q)^2]}} \quad \text{Equation 3.35}$$

3.8.2 Application of TM-UKRB for optimized rain gauge network verification

The main objective of the TM-UKRB development is as a tool to validate the optimized rain gauge network. The TM-UKRB is used to simulate the observed flow using different rainfall datasets which are produced by each optimized rain gauge network for 4 flood events which are used in the calibration and validation process (Table 3.12 and Table 3.13).

The areal rainfall data series in the 15-minute interval for each optimized rain gauge network is calculated using the Thiessen Polygon method and used as input data to the TM-UKRB to simulate the observed flow. The NSE of the model and the relative error of the hydrograph are calculated to analyse the reliability of the optimized rain gauge network.

CHAPTER 4: RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter presents and discusses the results of the methodology described in Chapter 3. Basically, this chapter presents 5 results i.e. data input for preliminary analysis, the best semivariogram model selected, PSO algorithm modification, optimization of the rain gauge network and the validation of the optimal station network.

The preliminary data analysis (rainfall and discharge) result has been presented in section 4.2 to give a general description of the data used in this study. The descriptive statistical values of the datasets have been used and in addition, the Pearson correlation is adopted for rainfall-elevation and rainfall-discharge relationships evaluation before the geostatistical method is employed in subsequent methodologies.

In section 4.3, the results of the new application of geostatistical and AHP methods to select the semivariogram model are discussed. The selected best semivariogram model is used in the rain gauge network optimization process.

One of the novelties of this thesis is the application of the modified PSO method in the rain gauge network optimization process. Thus, the experimental results of how this method is modified are presented in section 4.4. Next, in section 4.5, the results of the optimization of the rain gauge network using two new methods, CV-Geo and MPSO were presented and discussed.

The validation result of the optimum network produced using the hydrological tank model has been presented in section 4.6. This section also presents the performance of the optimum rain gauge network to simulate flood events for different rainfall data from the optimum rain gauge network.

4.2 The result of Preliminary Data Analysis

4.2.1 Descriptive Statistic and Preliminary Number of Stations

The utilization of statistical nature is a conventional method applied to design the appropriate rain gauge number from the base network. It gives a basic guide to determine the required number of rain gauge stations in a network using the statistic parameters value and spatial variation of rainfall data set. In this case, the coefficient of spatial variation of rainfall (C_v) from the base network is utilized for determining the optimum number of rain gauges. If there are already some rain gauges in the catchment, the optimal number of stations that should exist to have an assigned percentage of error in the estimation of mean rainfall is obtained by statistical analysis as:

$$N = \left(\frac{C_v}{\varepsilon} \right)^2 \quad \text{Equation 4.1}$$

where,

N = optimal number of stations,

ε = allowable degree of error in the estimate of mean rainfall and

C_v = coefficient of variation of rainfall values at then existing m stations.

If there are m stations in a catchment and P_1, P_2, \dots, P_m is the recorded rainfall at a known time at 1, 2, m station, then the coefficient of variation C_v is calculated as:

$$C_v = \frac{100 \times \sigma_{m-1}}{\bar{P}} \quad \text{Equation 4.2}$$

Where

$$\sigma_{m-1} = \frac{\sqrt{\sum_{i=1}^m P_i^2 - m \times \bar{P}^2}}{(m-1)} \quad \text{Equation 4.3}$$

P_i is daily average precipitation at i station and \bar{P} is the average rainfall of 'm' number of stations, given by:

$$\bar{P} = \frac{\sum_{i=1}^m P_i}{m} \quad \text{Equation 4.4}$$

For purpose of C_v calculation, σ_{m-1} is used when number of stations, m , in the network is less than 30 otherwise σ_m can also be used. In this study, N is generated using three ε values, 10%, 15% and 20% to evaluate the appropriate number of rain gauge stations.

The calculated N values were plotted against ε and the result is illustrated by Figure 4.1. The result showed that at $\varepsilon=10\%$ (the common value applied) the number of stations will be in a range of 41 to 164. This range is considered too vast for the study area as the catchment area is about 584km². According to the WMO's guide for the urban area, the recommended number of stations shall be in a range of 29 (20km² per station) to 58 (10km² per station) number of stations. To trade off the number of the station by increasing the $\varepsilon=15\%$ and $\varepsilon=20\%$, the range is decreased to a range of 18 to 73 stations and 10 to 41 stations, respectively. It is observed that the range was reduced when the allowable degree of error, ε increased. In such a case, it is difficult to determine the right number of stations as accuracy remained as the main factor. However, this could be as a preliminary benchmark. The result of all statistical parameter values including C_v and N for each ε is tabulated in Table 4.1.

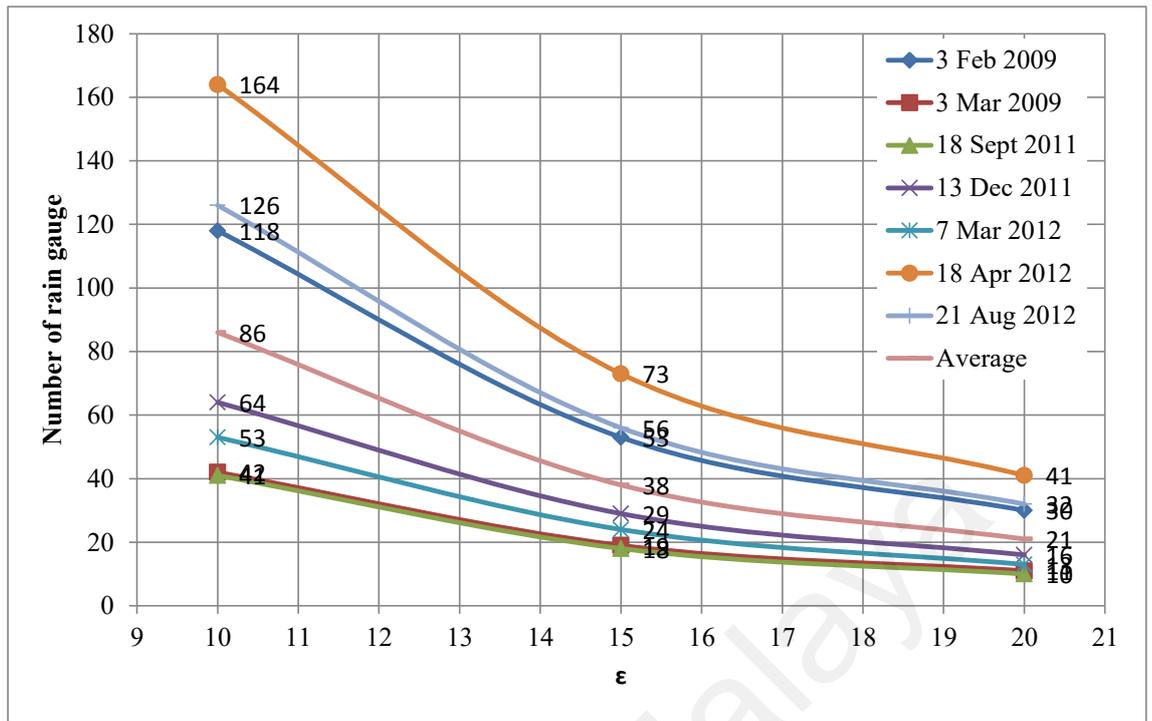


Figure 4.1: Optimum number of rain gauge stations, N

Table 4.1: The descriptive statistical value of rainfall datasets.

Statistical parameter	3 February 2009	3 March 2009	18 September 2011	13 December 2011	7 March 2012	18 April 2012	21 August 2012
Maximum	100.0	134.0	87.0	142.0	239.5	111.0	130.0
Minimum	0	0	0	0	0	0	0
Mean	23.98	45.03	37.42	40.71	79.93	22.35	27.71
Standard Error	3.93	4.42	3.61	4.92	8.79	4.32	4.69
Standard Deviation	26.07	29.33	23.95	32.65	58.29	28.65	31.11
Variance	679.57	860.44	573.78	1066.05	3397.74	821.06	967.64
Cv	108.722	65.135	64.0164	80.204	72.931	128.207	112.253
Number of station (ε=10%)	118	42	41	64	53	164	126
Number of station (ε=15%)	53	19	18	29	24	73	56
Number of station (ε=20%)	30	11	10	16	13	41	32

Note: Only stations in the catchment were considered.

4.2.2 Rainfall-Elevation Correlation

One of the advantages of using geostatistical methods is that some of the criteria that relate to one another can be used in the analysis. In this study, the rainfall data used may have a relationship with the station's elevation. If a significant relationship exists, the elevation factor should be taken into account in the geostatistical analysis to obtain an accurate result.

To determine the correlation between rain data with elevation, Pearson correlation (r) is used. The results are illustrated in Figure 4.2 until Figure 4.8. The r value results are in a range of -0.393 to 0.185. A strong negative correlation indicated if the value of r is close to -1. Thus, this result shows that there is no correlation between the rainfall amount and the station's elevation for all datasets. In addition, the computed coefficient of determination, R^2 in a range of -1.174 to -0.495, which indicated the same inference.

This analysis is quite simple but essential to the geostatistical analysis. To obtain an accurate result from the geostatistical analysis all elements that have good correlation should be incorporated into the analysis. However, based on the result, it is justified that for this study the elevation is not needed to be incorporated in the geostatistical analysis.

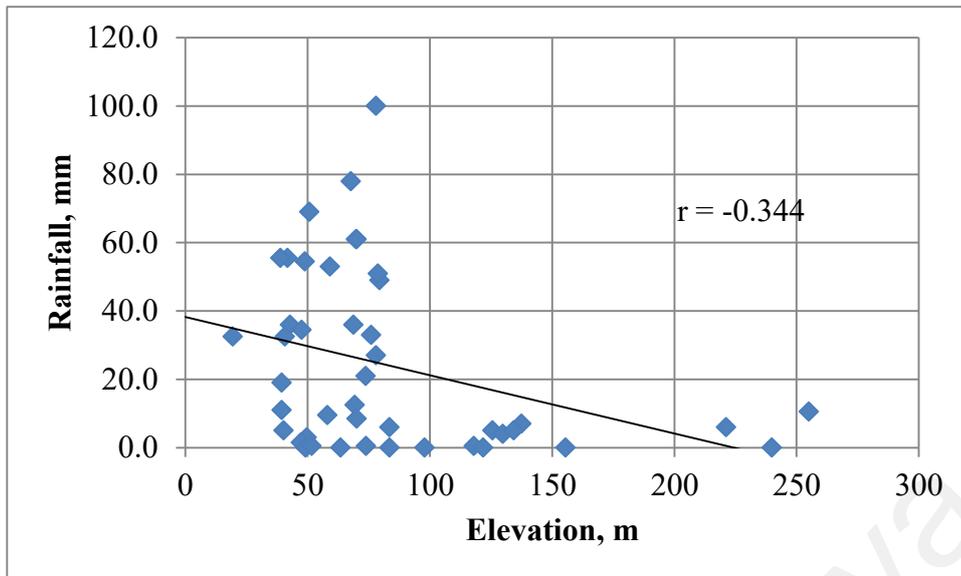


Figure 4.2: Pearson correlation, r between rainfall and elevation for 3 February 2009

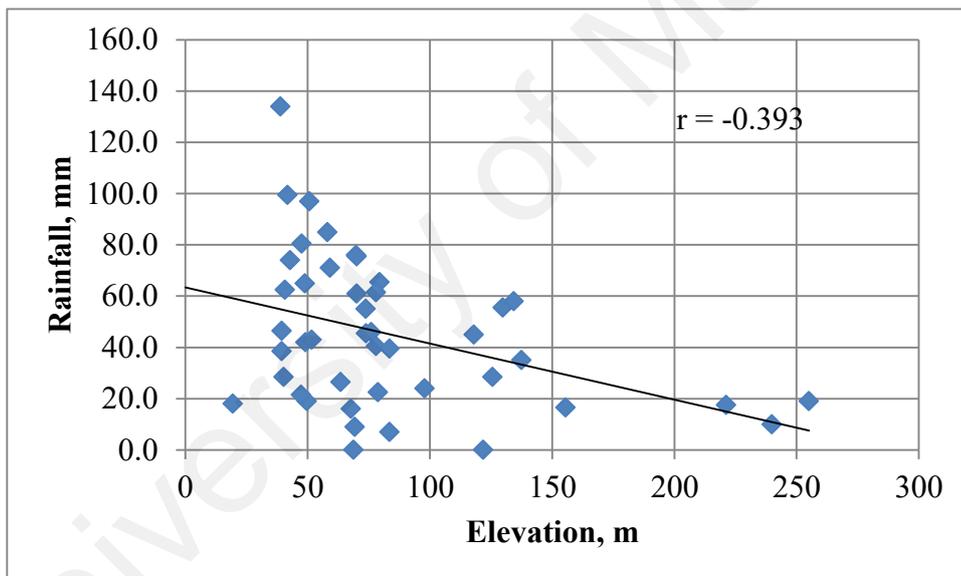


Figure 4.3: Pearson correlation, r between rainfall and elevation for 3 March 2009

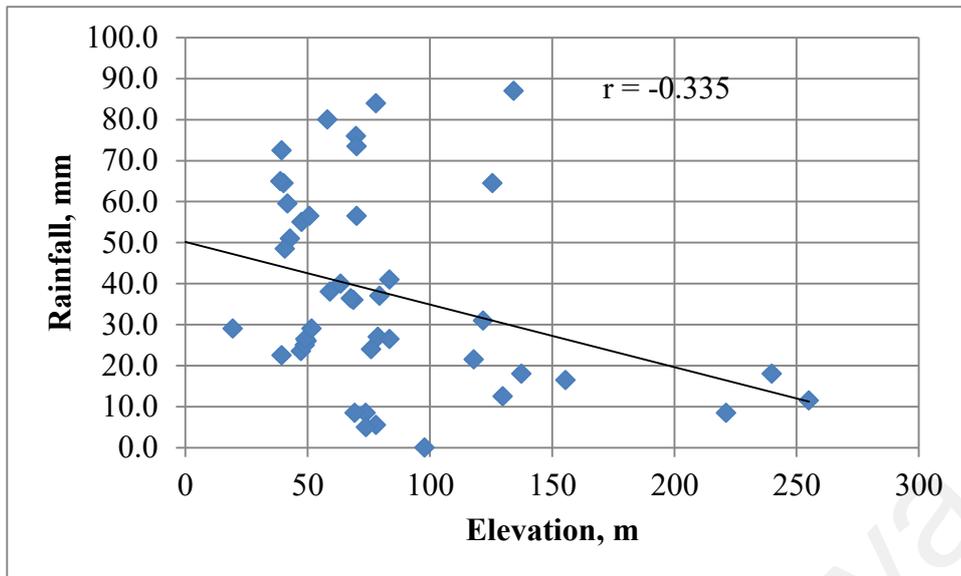


Figure 4.4: Pearson correlation, r between rainfall and elevation for 18 September 2011

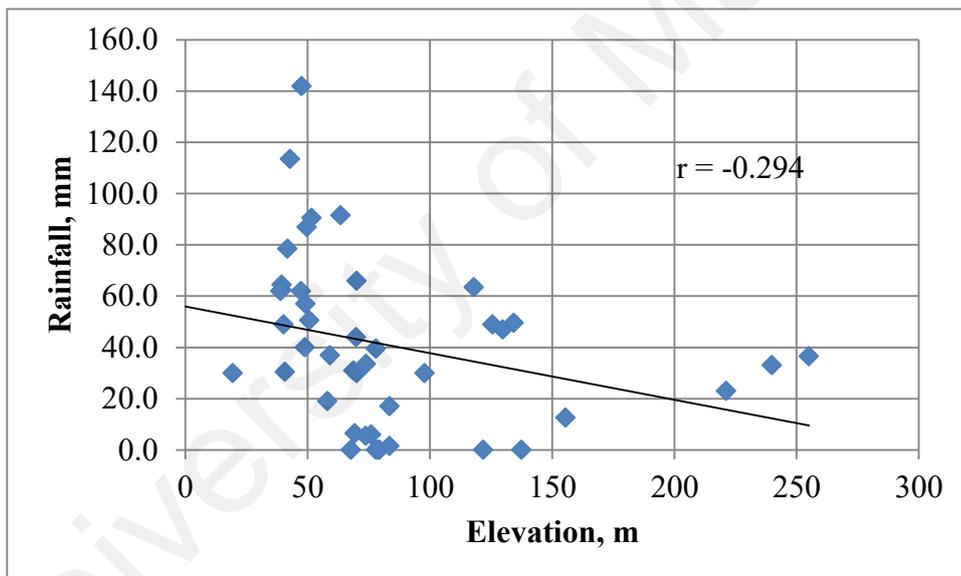


Figure 4.5: Pearson correlation, r between rainfall and elevation for 13 December 2011

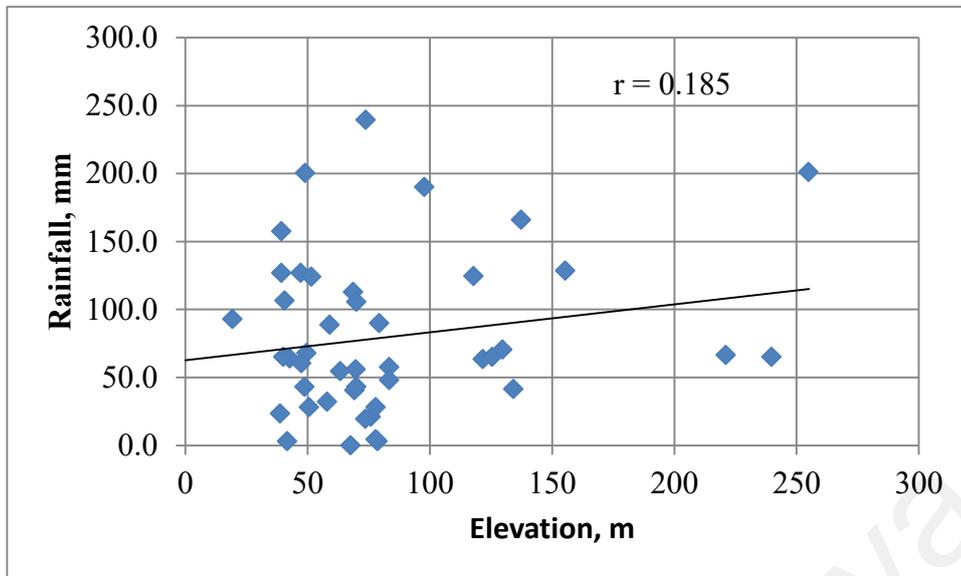


Figure 4.6: Pearson correlation, r between rainfall and elevation for 7 March 2012

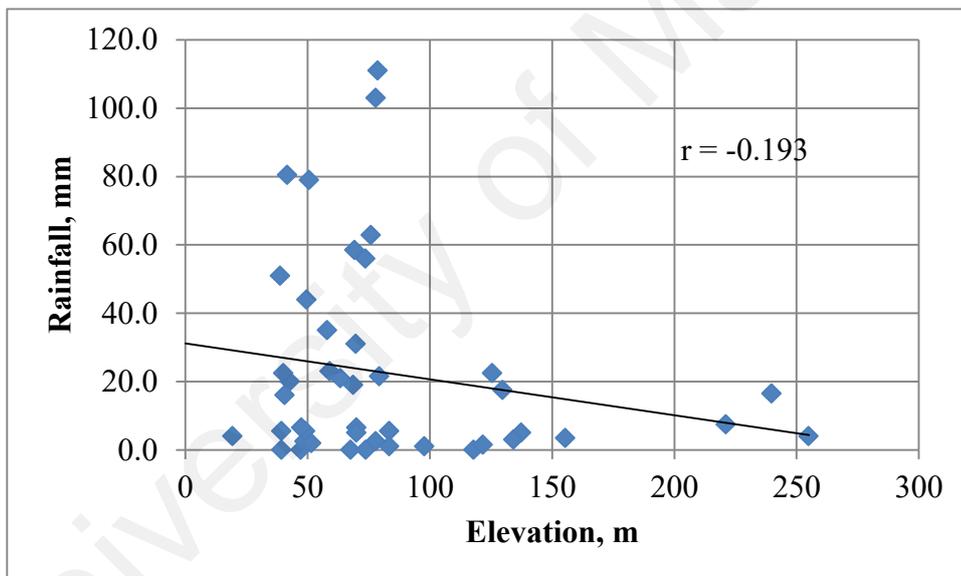


Figure 4.7: Pearson correlation, r between rainfall and elevation for 18 April 2012

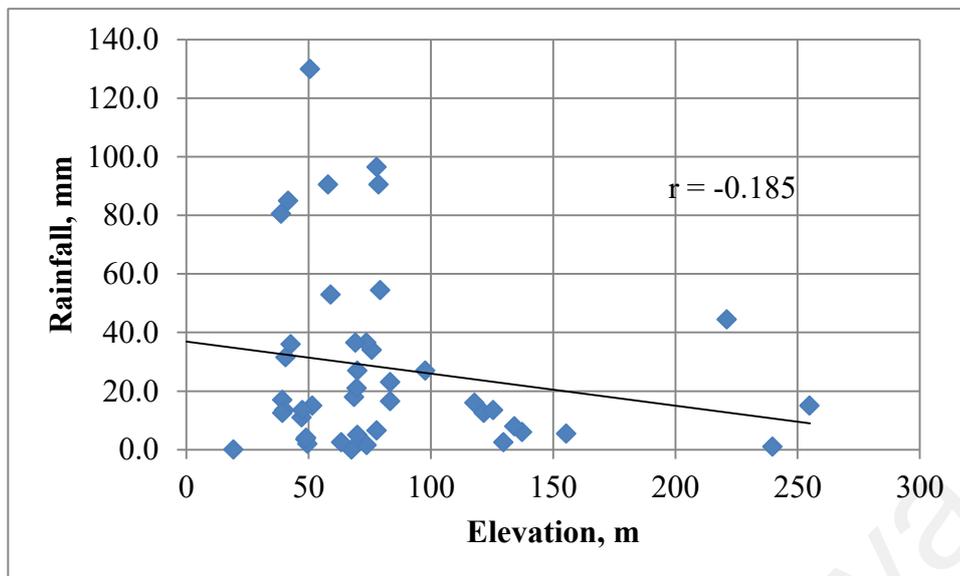


Figure 4.8: Pearson correlation, r between rainfall and elevation for 21 August 2012

4.2.3 Rainfall-Discharge Correlation

The correlation between rainfall value and discharge is very important for hydrological modeling. A good correlation will produce an accurate model and vice versa. Thus this relationship is assessed using r value.

In this study, only four discharge datasets are available. The r values were computed and the result is presented in Figure 4.9. Based on the result, the lowest and maximum r values calculated are 0.723 and 0.771, respectively. These values indicated that a good correlation exists between the observed rainfalls and discharge. Thus, this result infers that the datasets are good enough to produce an efficient hydrologic model.

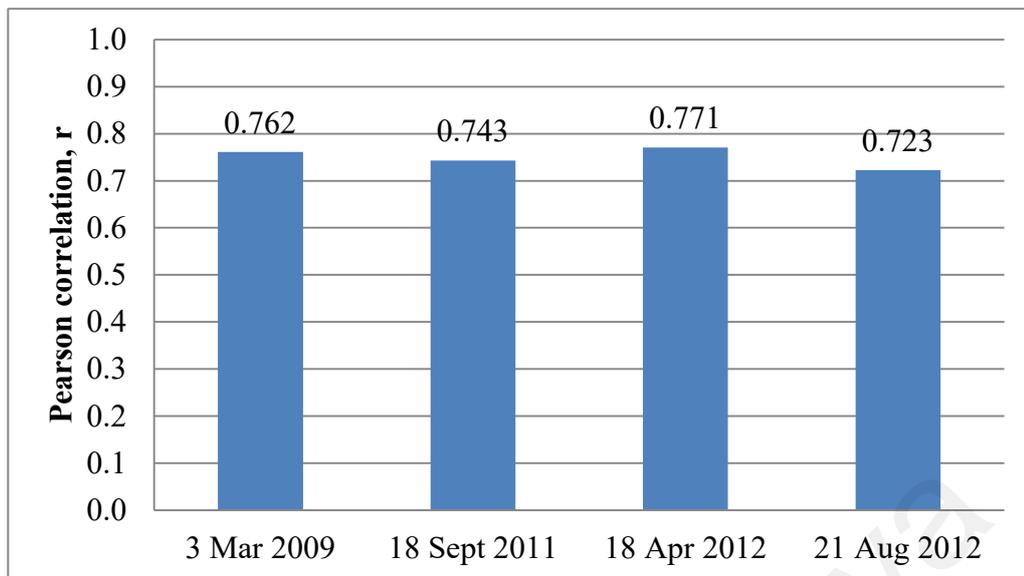


Figure 4.9: Pearson correlation, r between rainfall and discharge.

4.3 Selection of the Best Semivariogram Model

The successful selection of the best semivariogram model relied on the priority weight of criteria and sub-criteria designed in the AHP's model. For this purpose, in this section, the priority weight of AHP model is presented and discussed. Next, the evaluation result on alternatives of the semivariogram model is discussed in detail to select the best.

4.3.1 AHP's Priority Weight

The results of priority weights in pairwise matrix analysis for criteria versus goal in the study are presented in Table 4.2. As mentioned earlier, in this study 3 sets of priority weight were employed. All priority weight sets produced satisfactory Consistency Ratio values of less than 0.1 to ensure the consistency of calculated priority weight. These 3 weight sets were used to observe the results for different levels of influence among the criteria. The *CR* value in Set 1 is 0.095 or 9.5% inconsistency in the decision, which is the highest and in Set 2, the *CR* value is the lowest (0.012 or 1.2% inconsistency). Whereas in Set 3 has moderate *CR* value that is 0.039 (3.9% inconsistency).

In Set 1, criterion C4 had 71.3% influences on the decision, but criterion C3 had the least influence with only 4.6%. Criteria C1 and C2 had an equal influence of 12.1%. However, in Set 2 and Set 3 criterion C4 still had the greatest weight, but criterion C3 increased up to 26.7% in Set 2 and 20.6%, in Set 3. Meanwhile, criteria C1 and C2 had the influence of only 6% and 8.5% on the decision, respectively. In other words, in Set 1 the dominant criterion was C4 (E_{mss}) and there was no emphasis on criterion C3 (E_{ms}). On the other hand, in Set 2 and Set 3, criterion C4 remained important with the same contribution as criterion C3 on the evaluation process. These 3 scenarios are employed to ensure the robustness of the decision made by AHP

Table 4.2: Results of priority weight and its associated comparative numerical value and Consistency Ratio of pairwise matrix comparison of the criteria versus objective

Set No.	Priority Weight	Factor	C1	C2	C3	C4	Consistency Ratio
1	0.121	C1	1.00	1.00	4.00	0.11	0.095
	0.121	C2	1.00	1.00	4.00	0.11	
	0.046	C3	0.25	0.25	1.00	0.11	
	0.713	C4	9.00	9.00	9.00	1.00	
2	0.060	C1	1.00	1.00	0.20	0.11	0.012
	0.060	C2	1.00	1.00	0.20	0.11	
	0.267	C3	5.00	5.00	1.00	0.33	
	0.612	C4	9.00	9.00	3.00	1.00	
3	0.085	C1	1.00	1.00	0.33	0.17	0.039
	0.085	C2	1.00	1.00	0.33	0.17	
	0.206	C3	3.00	3.00	1.00	0.20	
	0.623	C4	6.00	6.00	5.00	1.00	

4.3.2 Selected Semivariogram Model

At this stage, the Final Score value is used to evaluate the alternatives. Based on the Final Score value the variogram models are ranked for decision making. The results of the variogram model rank based on the Final Score value for each set is tabulated in Table 4.3. Based on the table, each set unanimously resulted in the same model ranking

at the top, that is, the Spherical model. The minimum Spherical model's score value is 3.56 (in Set 2) compared with the maximum score in this study that is 5. This means that the Spherical model demonstrated good performance to fulfil all criteria with at least 71.2% and the maximum performance is 75.6% (score of 3.78 in Set 1).

The Gaussian and Tetraspherical models are ranked second and third in Set 2 and Set 3, respectively. In contrast, for Set 1 the Pentaspherical ranked second followed by the Gaussian model in third place. The Gaussian model's score recorded was equal in Set 2 and Set 3 with 3.56. This score is equal to the minimum score of the Spherical model. This result signifies that the Spherical and Gaussian models perform well in producing a spatial rainfall map that fulfils the criteria explained by Johnston et al. (2003). As for the rest, the performances of the variogram models were average and were ranked by the AHP method in bottom place.

Table 4.3: Results of the variogram model ranking based on the Final Score value

Pairwise Matrix Set		Set 1		Set 2		Set 3	
Variogram model rank position	1	Spherical	3.78	Spherical	3.56	Spherical	3.67
	2	Pentaspherical	3.11	Gaussian	3.56	Gaussian	3.56
	3	Gaussian	2.89	Tetraspherical	3.00	Tetraspherical	2.78
	4	Tetraspherical	2.89	Exponential	2.56	Pentaspherical	2.67
	5	Exponential	2.33	Pentaspherical	2.33	Exponential	2.33

This result was compared with the findings by F. Othman et al. (2011). They conducted an analysis of a storm event on spatial rainfall distribution in the same study area. By using a total of 28 rain gauge stations, they found that the Gaussian model had slightly better performance in producing spatial rainfall estimation compared to the

Spherical and Exponential models. There is a good agreement on the findings regarding the performance of the Gaussian and Exponential models. But contradiction arises for the Spherical model. Applying the AHP method has ranked the Spherical model at the top, indicating that the model has good performance.

This is possibly due to 2 reasons, one of which is the number of rain gauge stations selected. There were 55 stations selected in this study, which is twice more than by F. Othman et al. (2011). Thus, network density in this study is much greater. In the geostatistical analysis, the accuracy of spatial rainfall distribution is reliant on the density of the network used (Goovaerts, 1999). The higher the density of the rain gauge network used, the more accurate the spatial rainfall map produced is.

The second reason is the storm event selected for analysis. In this study, the rainfall data were extracted based on recent flood events on a cumulative, daily basis covering the years 2009 until 2012. This is to ensure that the data extracted for analysis are data with a good response from a hydrological perspective. However, F. Othman et al. (2011) selected storm events from the same year. According to the findings of Feki et al. (2010), the estimated variance and interpolation system in the geostatistical analysis are sensitive to season and region. As aforementioned, the climate feature in this study area is influenced by the monsoon seasons.

The difference in performance between the Spherical and Gaussian can be explained by how the model presents the correlation of the semivariogram within the range value. The Gaussian has a high response to the nugget effect but the Spherical is not. The nugget effect contributes to high error while the model used in the spatial interpolation process at unmeasured points. This factor caused the Gaussian to perform less than the Spherical model. Thus, it is important to assess the variography structure of these two models that are represented by variography parameters to justify the result.

The variography parameters values of the data sets for the Spherical and Gaussian model are tabulated in Table 4.4. The Nuggets recorded by Gaussian are higher than Spherical but the sill values do not much differ. The range values are higher for the Spherical model. This result shows that the Spherical model has better variography structure than Gaussian because the Spherical has a lower nugget and has higher range value. This meant that the Spherical has a more spatial correlation at longer distance and less measurement error. By using the variography structure, semivariogram of Spherical and Gaussian for two selected data sets are plotted in Figure 4.10. It shows that the Spherical model fits the data set slightly better as compared to the Gaussian model.

Table 4.4: Results of variography parameter value from geostatistical analysis

Date	Variogram Model	Range (km)	Nugget	Semi variance / Sill
3 February 2009	Spherical	14.714	205.05	652.8
	Gaussian	12.592	272.31	654.89
3 March 2009	Spherical	11.949	127.07	861.10
	Gaussian	10.245	252.02	862.73
18 September 2011	Spherical	18.956	385.66	658.18
	Gaussian	16.772	429.15	662.97
13 December 2011	Spherical	7.144	585.02	1069.82
	Gaussian	6.085	660.12	1070.51
7 March 2012	Spherical	13.535	272.92	3776.92
	Gaussian	11.162	719.44	3768.34
18 April 2012	Spherical	16.057	44.624	712.594
	Gaussian	14.160	150.79	722.88
21 August 2012	Spherical	14.557	38.027	917.687
	Gaussian	13.317	182.77	941.37

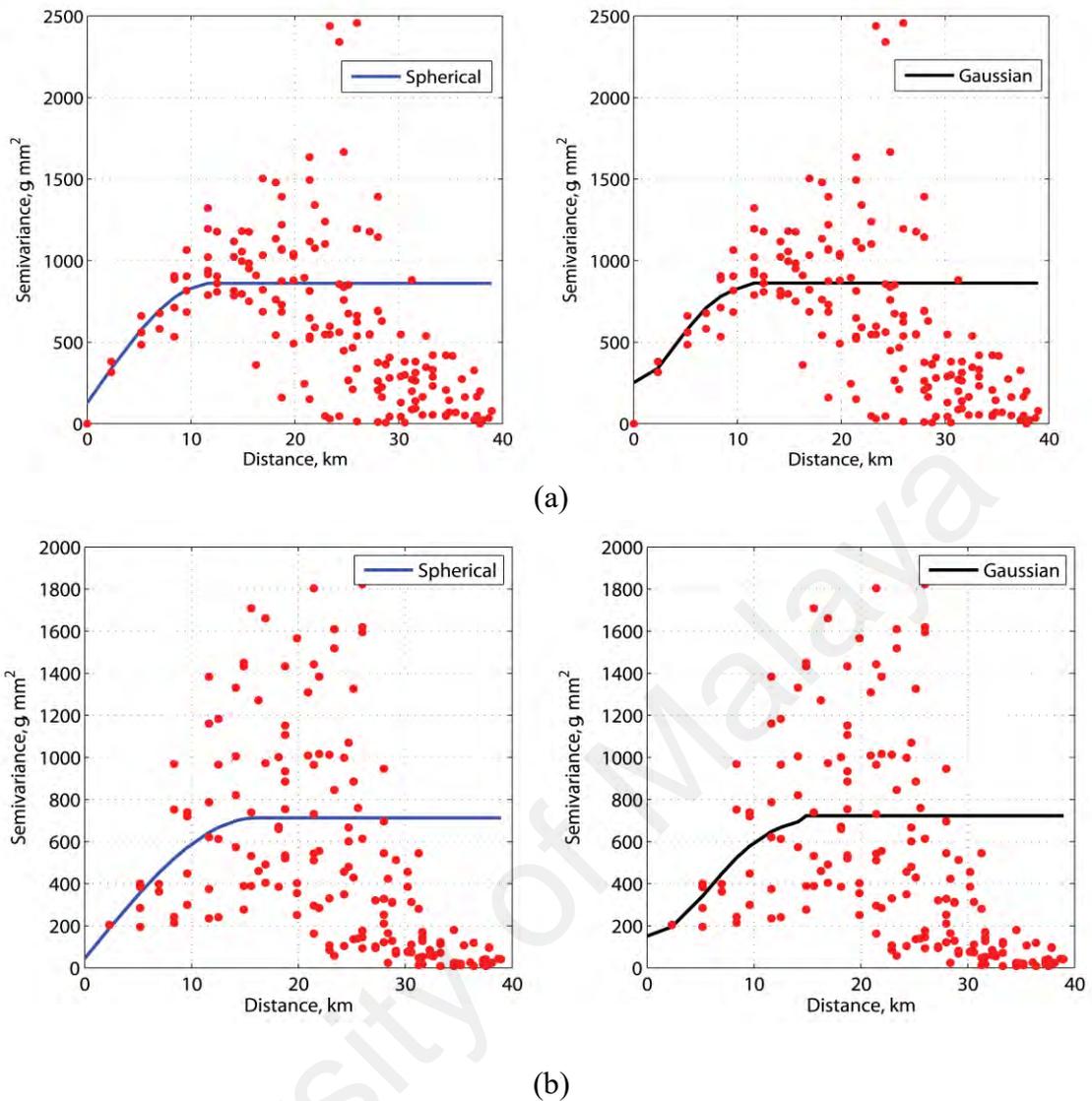


Figure 4.10: Semivariogram generated by Spherical and Gaussian models, a) rainfall data on 3 March 2009 and b) rainfall data on 18 April 2012.

A good spatial rainfall map can be produced by a good semivariogram structure. Apparently, based on the semivariogram structure, the Spherical model is able to produce good spatial rainfall map. Spatial rainfall maps for the same selected events are generated and shown in Figure 4.11. The figure illustrated that the spatial rainfall map produced by the Spherical model is smoother than that with the Gaussian model, especially in the area marked with a red boundary. The smooth line in Spherical is produced from accurate spatial interpolation. The smooth line justifies that the model has better spatial interpolation as explained in the previous paragraph.

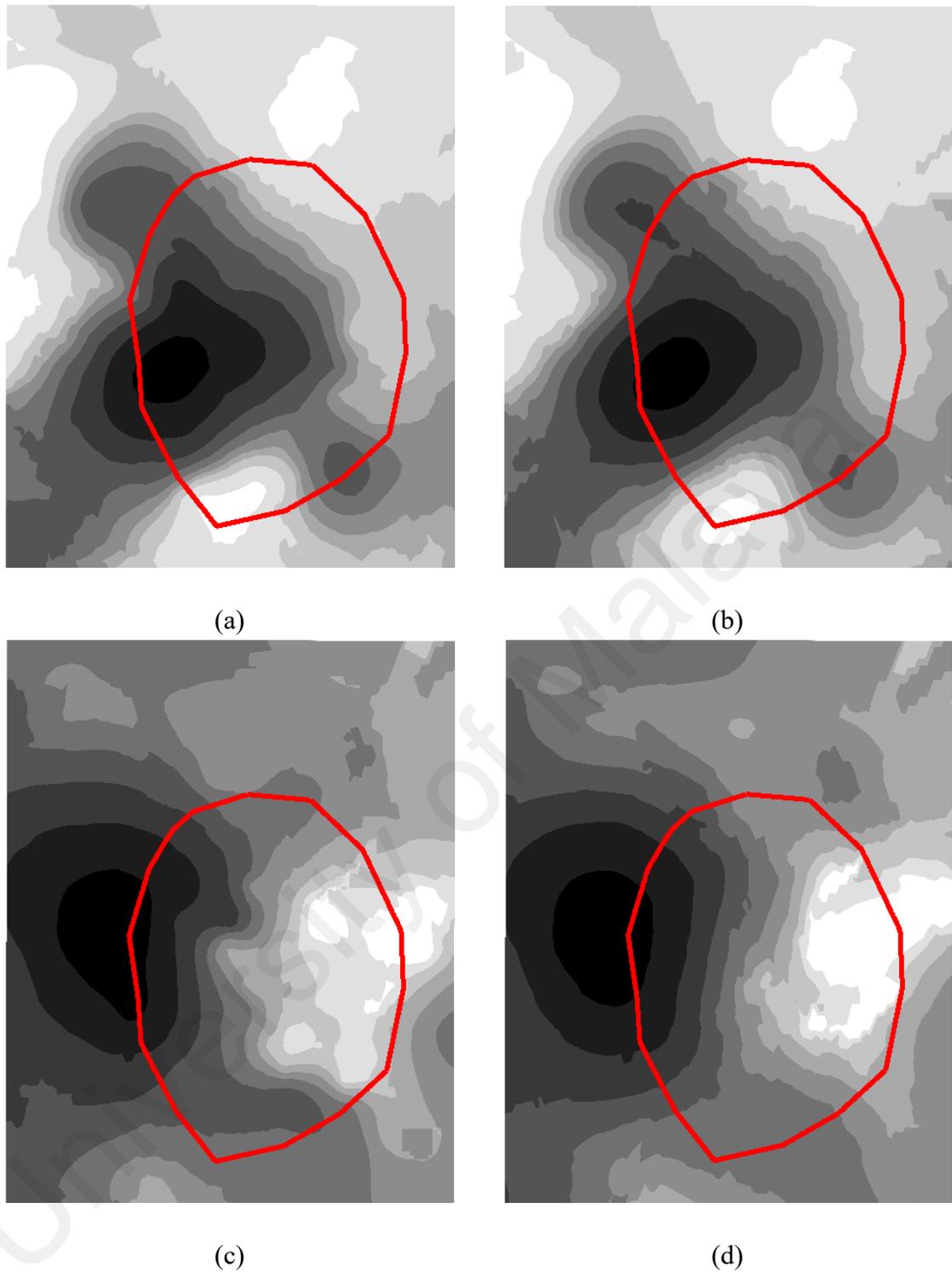


Figure 4.11: Spatial rainfall map generated by, a) Spherical model for rainfall data on 3 March 2009, b) Gaussian model for rainfall data on 3 March 2009, c) Spherical model for rainfall data on 18 April 2012 and d) Gaussian model for rainfall data on 18 April 2012

4.3.3 Semivariogram of Existing Rain Gauge Network

Prior to optimizing the rain gauge network, it is important to know the variogram properties of the existing rain gauge network. For this purpose, the geostatistical analysis was conducted on the existing network that contained 55 rain gauge stations with seven rainfall datasets. Detailed of variogram parameters values are tabulated in Table 4.5. Based on Table 4.5, the rainfall dataset for 7 March 2009 had the highest sill value of 3 703.179 mm². The remaining datasets had sill values ranging from 450.123 mm² (18 September 2011) to 1 017.863 mm² (21 August 2012). Although high differences in sill values were recorded, the semivariogram of the datasets was plotted with more or less similar range values except for the rainfall data on 13 December 2012 (3.754 km). The mean range value was about 13.289 km, whereas the maximum value was 18.182 km. Among the datasets, the rainfall data for 18 September 2011 had the highest nugget (246.107 mm²). Besides, three datasets (13 December 2011, 18 April 2012 and 21 August 2012) recorded nugget values very close to zero. The computed E_{rms} value of spatial interpolation error was in the 19.505 to 33.175 range with a mean E_{rms} value of 25.184.

Table 4.5: Value of variogram parameters of the existing network and their corresponding E_{rms} value for selected flood events.

Flood Event	Sill, C_1	Range, a (km)	Nugget, C_0	E_{rms}
3 February 2009	693.097	15.341	47.137	23.891
3 March 2009	930.505	12.118	23.511	24.281
18 September 2011	450.123	16.109	246.107	24.636
13 December 2011	983.093	3.754	2.376×10^{-8}	33.175
7 March 2012	3703.179	13.532	41.483	30.389
18 April 2012	992.719	18.182	4.359×10^{-8}	20.408
21 August 2012	1017.863	13.989	1.106×10^{-8}	19.505
Mean	1 252.940	13.289	51.177	25.184
Maximum Value	3 703.179	18.182	246.107	33.175
Minimum Value	450.123	3.754	1.106×10^{-8}	19.505

4.4 The evaluation result of MPSO-IP

The objective of this evaluation was to assess the convergence rate of the MPSO-IP against the SPSO-IP to minimize the test functions value for two scenarios. The first scenario is both algorithms were run at one time using similar value of r_1 and r_2 by controlling the value every number of iterations. Second scenario is to run the algorithms for 50 consecutive run without controlling the value of r_1 and r_2 . The convergence rate denoted as cr was defined as the ratio of the residue iteration where the swarm particles reached the best-optimized value over the maximum number of iterations. It was calculated using Equation 4.5 and the cr value should be in a range of [0, 1]. For a specific number of N and maximum number of iterations, the higher cr value illustrated the better performance of the algorithms to reach to the solution of the problem.

$$cr = 1 - \frac{i_B}{I_{max}} \quad \text{Equation 4.5}$$

Where, i_B = number of iteration for the best value, and

I_{max} = maximum number of iterations (1000 or 2000)

For this purpose, the SPSO-IP and MPSO-IP algorithms were coded and run using MATLAB R2013a. All evaluation stages explained in Chapter 3 were run using a laptop with Intel Core i7-2.4GHz processor and 8 GB memory. The obtained results were analyzed to evaluate the performance of the algorithms in terms of the convergence rate and the minimized best value for each test function.

4.4.1 Controlled Value of Random Number (r_1 and r_2)

The SPSO-IP and MPSO-IP algorithms were run with similar value of random numbers but different value of maximum number of iterations, I_{max} and N . Two values of N (25 and 50) and I_{max} (1000 and 2000) were applied in this study. The test results

for I_{max} of 1000 and 2000 iterations for different N values are tabulated in Table 4.6 and Table 4.7, respectively.

Based on the results, apparently the MPSO-IP has higher cr value for all test functions, irrespective of the value of both N sizes and I_{max} . The MPSO-IP has recorded the minimum average value of cr as lower as 0.852 (for $N=50$ and for I_{max} of 1000). On the other hands, the SPSO-IP has recorded the maximum average value of cr as higher as 0.419 (for $N=50$ and for I_{max} of 2000). In term of the average number of iteration wherein the algorithms stop at the best-optimized value for both I_{max} values, the MPSO-IP has stop at 58 numbers of iterations for N equal to 25 and 148 number of iterations for N equal to 50. According to Equation 4.5, the value of cr will be increased if the I_{max} increased and i_B was unchanged. Therefore, the cr value for MPSO-IP has increased and its value was higher than the cr value of SPSO-IP. These finding indicated that the MPSO-IP was independent of the maximum number of iterations. Meanwhile, SPSO-IP is still dependent on the maximum number of iterations.

Since, the MPSO-IP is likely independent of the maximum number of iterations, the comparisons of the best-optimized value of the test function were done based on the N value to assess the behavior of the algorithms for the same I_{max} . Based on results in Table 4.6 and Table 4.7, for I_{max} values of 1000, the MPSO-IP has reached to the actual best value of the test function for TF_1 when N value increased from 25 to 50 but not for SPSO-IP. In contracts, the SPSO-IP has slightly better result for TF_2 as compared to MPSO-IP. As for the TF_3 and TF_5 , the results show that both algorithms are equally performed for both values of N . However, both algorithms produced best-optimized value slightly higher for TF_4 when N value increased from 25 to 50.

For I_{max} values of 2000, both the MPSO-IP and SPSO-IP have reached the actual best value of the test function for TF_1 when N value increased from 25 to 50. In

contracts, the SPSO-IP has slightly better result for TF_2 as compared to MPSO-IP. The TF_3 and TF_5 results show that no changes for both algorithms. On the other hands, for TF_4 MPSO-IP has produced lower best-optimized value for both values of N as compared to SPSO-IP.

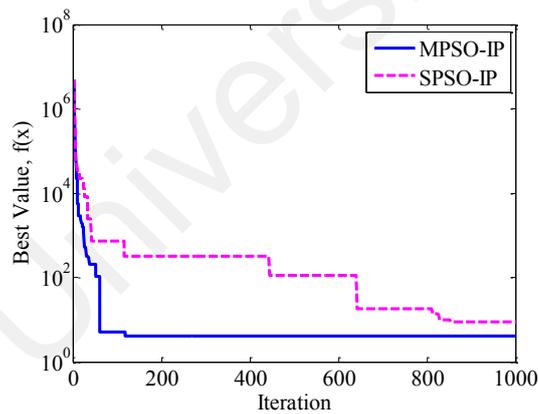
In this evaluation, the random numbers in both algorithms were controlled to ensure that the same set of random numbers was applied in the test. Thus, the results produced by both algorithms were comparable. Based on the observation, it is can be inferred that the MPSO-IP and SPSO-IP are equally performed to minimize the test functions value. The obvious different performance between both algorithms is that the MPSO-IP has advantage to converge the solution faster as compared to SPSO-IP. These inferences were justified by the convergences graph of each TF for each test condition as plotted in Figure 4.12 to Figure 4.21. The only disadvantage can be noticed is that the possibility of the algorithms trapped at the global solution.

Table 4.6: Results of single runs with a controlled random number for 1000 iterations.

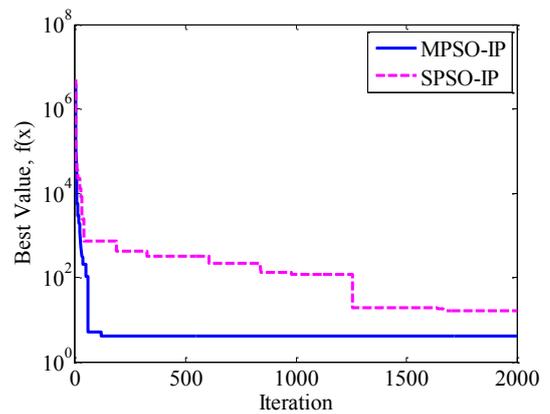
Test Function	Algorithm	N=25			N=50		
		Best Value	Iteration at Best Value	<i>cr</i>	Best Value	Iteration at Best Value	<i>cr</i>
TF_1	SPSO-IP	9.000	853	0.147	16.000	848	0.152
	MPSO-IP	4.000	117	0.883	0.000	194	0.806
TF_2	SPSO-IP	102.000	811	0.189	9.000	505	0.495
	MPSO-IP	205.000	56	0.944	205.000	111	0.889
TF_3	SPSO-IP	8.88E-16	812	0.188	8.88E-16	802	0.198
	MPSO-IP	8.88E-16	44	0.956	8.88E-16	55	0.945
TF_4	SPSO-IP	0.0710	804	0.196	0.0943	805	0.195
	MPSO-IP	0.0862	31	0.969	0.0929	326	0.674
TF_5	SPSO-IP	0	812	0.188	0	546	0.454
	MPSO-IP	0	44	0.956	0	55	0.945
Average	SPSO-IP		818	0.182		701	0.299
	MPSO-IP		58	0.942		148	0.852

Table 4.7: Results of single runs with a controlled random number for 2000 iterations.

Test Function	Algorithm	N=25			N=50		
		Best Value	Iteration at Best Value	cr	Best Value	Iteration at Best Value	cr
TF ₁	SPSO-IP	16.000	1674	0.163	0.000	1634	0.183
	MPSO-IP	4.000	117	0.942	0.000	194	0.903
TF ₂	SPSO-IP	0.000	1663	0.169	0	1534	0.233
	MPSO-IP	205.000	56	0.972	205	111	0.945
TF ₃	SPSO-IP	8.88E-16	1602	0.199	8.88E-16	996	0.502
	MPSO-IP	8.88E-16	44	0.978	8.88E-16	55	0.973
TF ₄	SPSO-IP	0.1587	623	0.689	0.1011	623	0.689
	MPSO-IP	0.0862	31	0.985	0.0929	326	0.837
TF ₅	SPSO-IP	0	1154	0.423	0	1029	0.486
	MPSO-IP	0	44	0.978	0	55	0.973
Average	SPSO-IP		1343	0.329		1163	0.419
	MPSO-IP		58	0.971		148	0.926

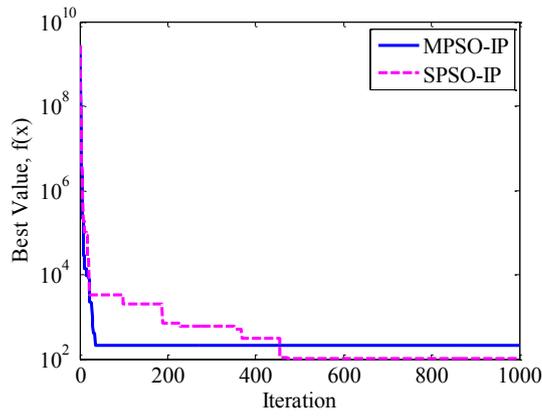


a)

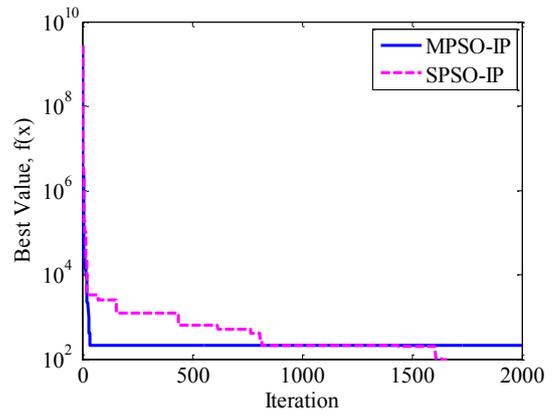


b)

Figure 4.12: Convergence graphs of TF₁ (N=25) for a) 1000 iterations b) 2000 iterations.

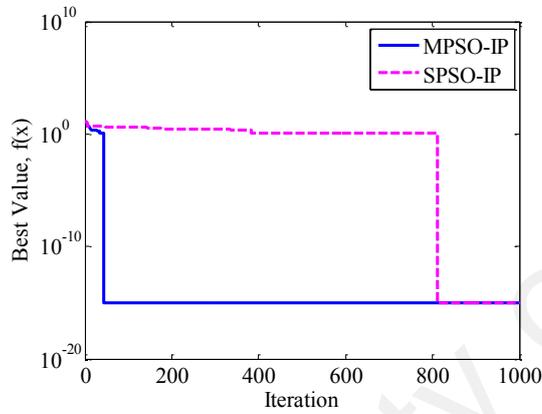


a)

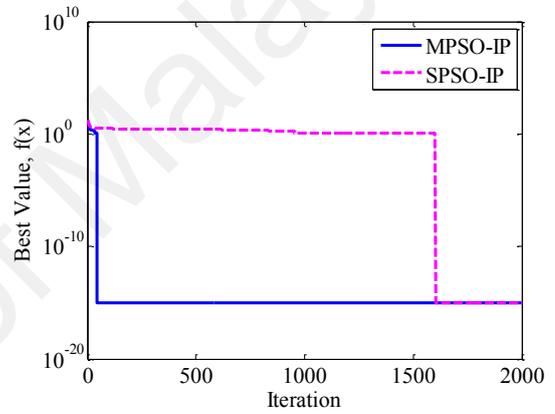


b)

Figure 4.13: Convergence graphs of TF_2 ($N=25$) for a) 1000 iterations b) 2000 iterations.

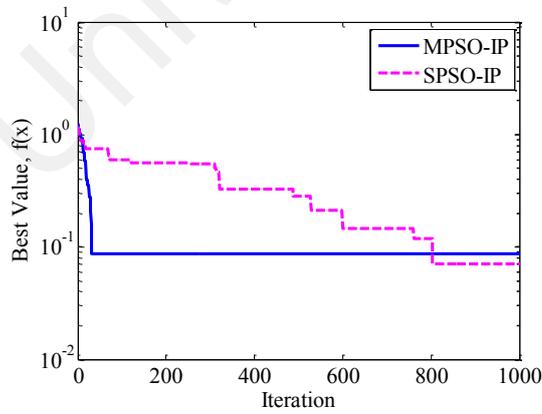


a)

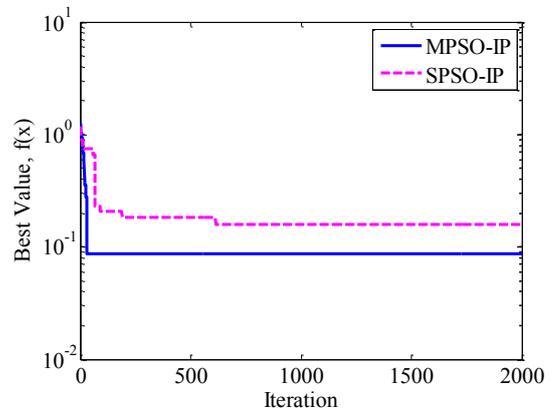


b)

Figure 4.14: Convergence graphs of TF_3 ($N=25$) for a) 1000 iterations b) 2000 iterations.

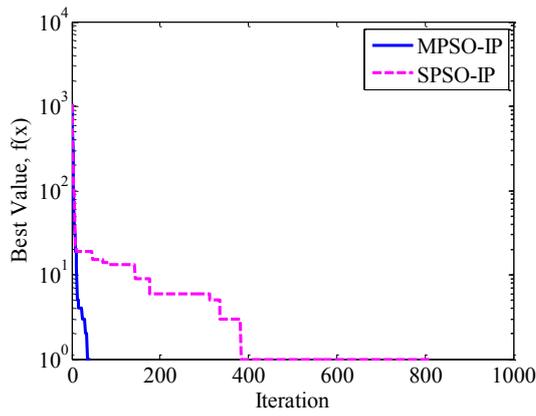


a)

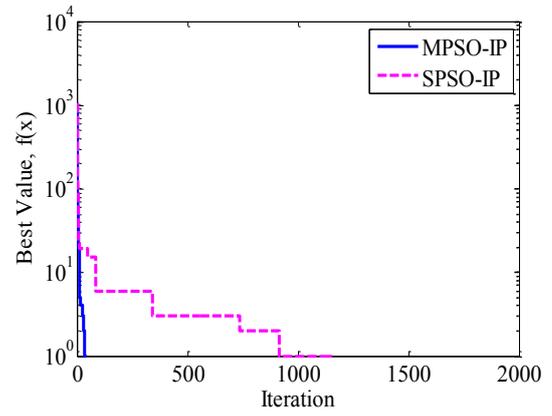


b)

Figure 4.15: Convergence graphs of TF_4 ($N=25$) for a) 1000 iterations b) 2000 iterations.

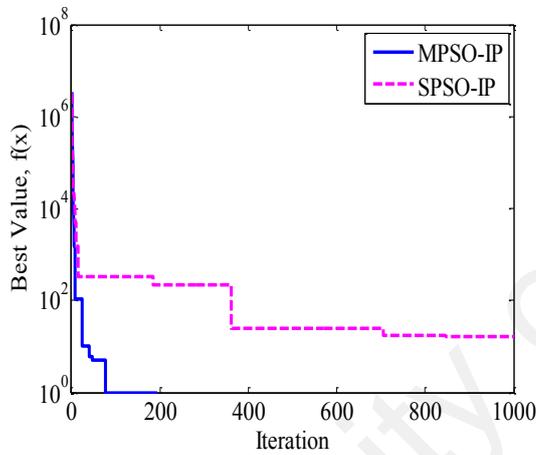


a)

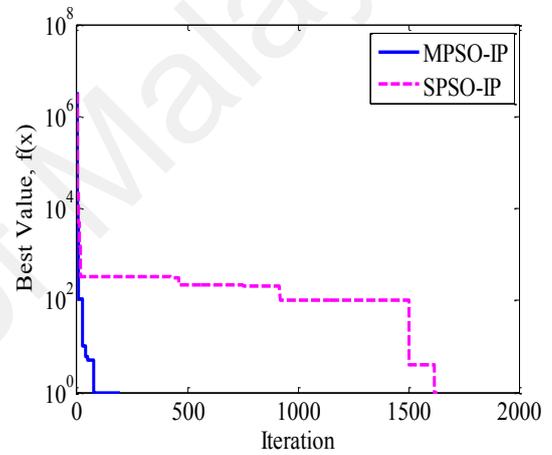


b)

Figure 4.16: Convergence graphs of TF_5 ($N=25$) for a) 1000 iterations b) 2000 iterations.

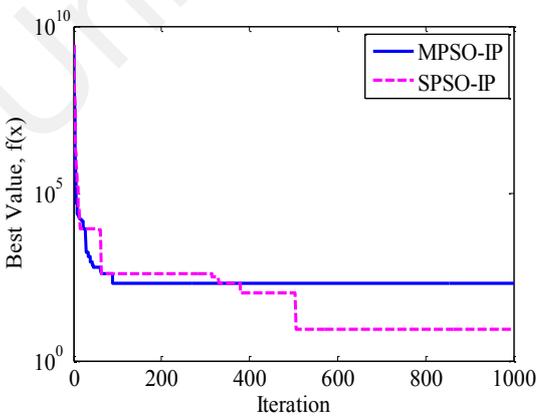


a)

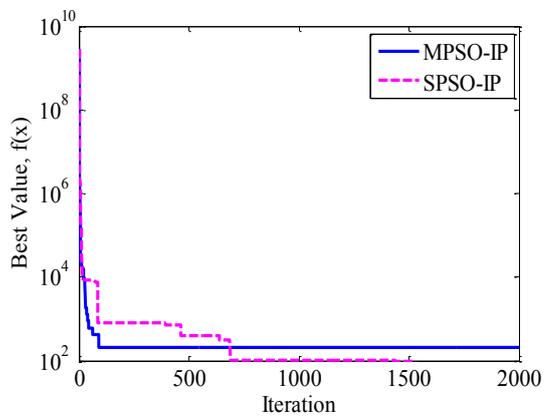


b)

Figure 4.17: Convergence graphs of TF_1 ($N=50$) for a) 1000 iterations b) 2000 iterations.

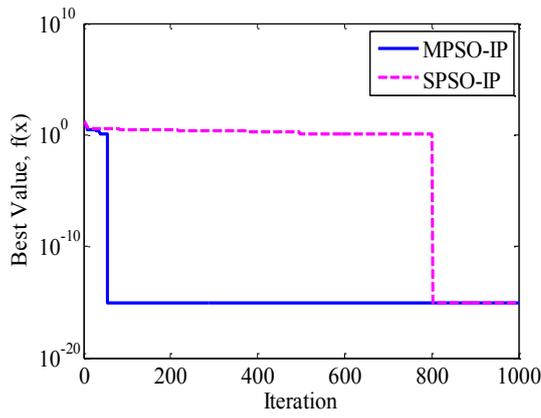


a)

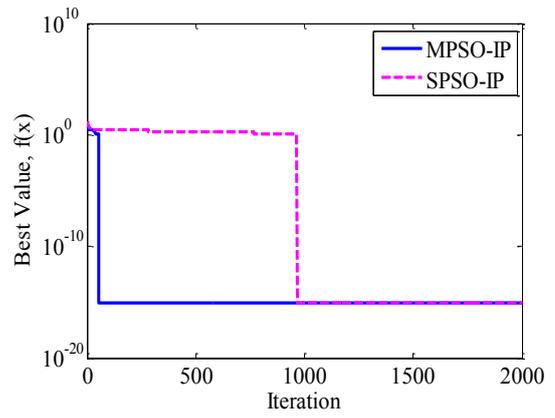


b)

Figure 4.18: Convergence graphs of TF_2 ($N=50$) for a) 1000 iterations b) 2000 iterations.

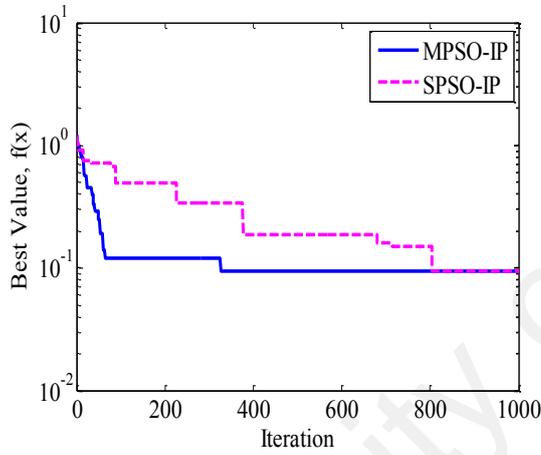


a)

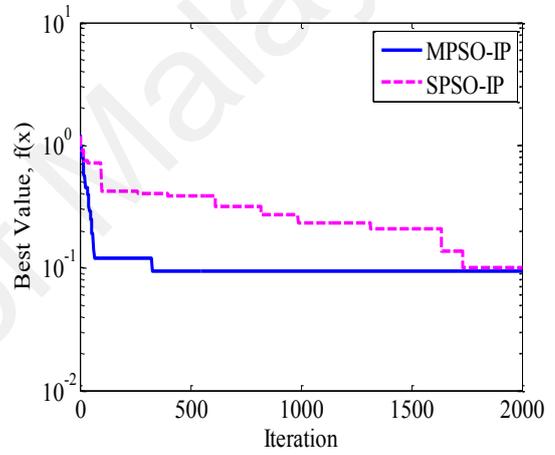


b)

Figure 4.19: Convergence graphs of TF_3 ($N=50$) for a) 1000 iterations b) 2000 iterations.

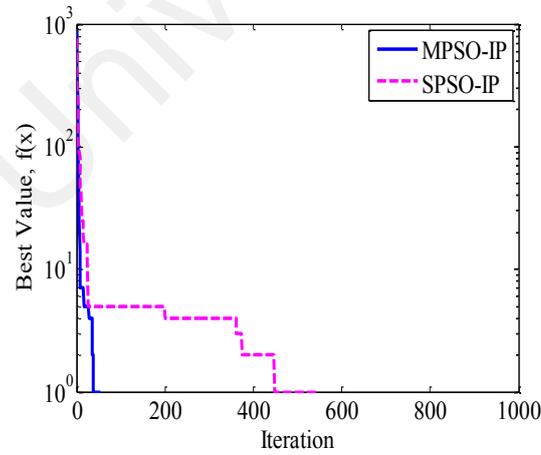


a)

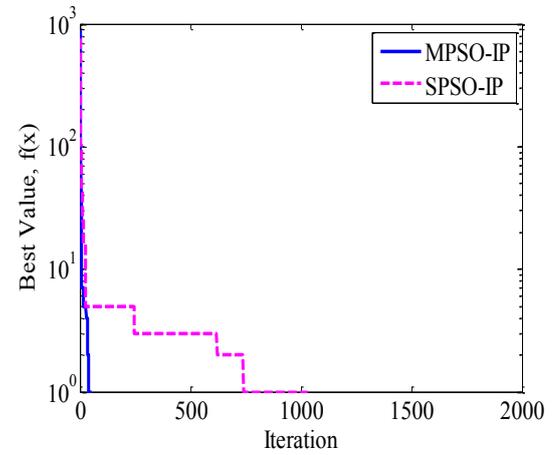


b)

Figure 4.20: Convergence graphs of TF_4 ($N=50$) for a) 1000 iterations b) 2000 iterations.



a)



b)

Figure 4.21: Convergence graphs of TF_5 ($N=50$) for a) 1000 iterations b) 2000 iterations.

4.4.2 Multi-run with Uncontrolled Value of Random Number

In literature, most of the PSO algorithms did not reach to the actual optimum value of the test function, for instances in Safaei et al. (2012), Jin (2011) and X. H. Shi et al. (2007). The success of the optimization task using the PSO algorithm can be assured by the repetition of the runs. Indeed, it is a common practice in the optimization task to solve the high dimension of the test function. In this study, all test functions were set to 10 dimensions problem. In the single run test with controlled random number values, it apparently showed that some of the test functions did not reach the actual optimum value, which was zero. Thus, for the next stage of the evaluation, 50 consecutive runs were applied and the lowest best result of the optimum value was extracted to represent the optimization result.

The extracted best value of the multi-run test result with uncontrolled random number value for 1000 and 2000 maximum number of iterations are tabulated in Table 4.8 and Table 4.9, respectively. Based on the result, for both maximum number of iterations and the N size, the TF_1 , TF_2 and TF_5 reached the target optimized the best value. However, the TF_3 did not reach the target best value, but both algorithms had the equal best value for both maximum number of iterations and N size. This was quite similar for TF_4 with $N=25$; however, for the test with $N=50$ and both iterations, only the SPSO-IP reached the target best value. Regardless, for the best value of both maximum number of iterations and N size, the MPSO-IP had a higher cr value.

Table 4.8: Results of the best value extracted from 50 runs with an uncontrolled random number for 1000 iterations.

Test Function	Algorithm	N=25			N=50		
		Best Value	Iteration at Best Value	<i>cr</i>	Best Value	Iteration at Best Value	<i>cr</i>
<i>TF</i> ₁	SPSO-IP	0.000	472	0.528	0.000	811	0.189
	MPSO-IP	0.000	47	0.953	0.000	162	0.838
<i>TF</i> ₂	SPSO-IP	0.000	820	0.180	0.000	554	0.446
	MPSO-IP	0.000	44	0.956	0.000	61	0.939
<i>TF</i> ₃	SPSO-IP	8.88E-16	377	0.623	8.88E-16	334	0.666
	MPSO-IP	8.88E-16	30	0.970	8.88E-16	19	0.981
<i>TF</i> ₄	SPSO-IP	0.0312	862	0.138	0	885	0.115
	MPSO-IP	0.0312	58	0.942	0.0312	31	0.969
<i>TF</i> ₅	SPSO-IP	0	377	0.623	0	334	0.666
	MPSO-IP	0	30	0.970	0	19	0.981
Average	SPSO-IP		581	0.418		583	0.416
	MPSO-IP		48	0.958		58	0.942

Table 4.9: Results of the best value extracted from 50 runs with an uncontrolled random number for 2000 iterations.

Test Function	Algorithm	N=25			N=50		
		Best Value	Iteration at Best Value	<i>cr</i>	Best Value	Iteration at Best Value	<i>cr</i>
<i>TF</i> ₁	SPSO-IP	0.000	1671	0.165	0.000	1199	0.401
	MPSO-IP	0.000	291	0.855	0.000	92	0.954
<i>TF</i> ₂	SPSO-IP	0.000	1075	0.463	0.000	905	0.548
	MPSO-IP	0.000	47	0.977	0.000	51	0.975
<i>TF</i> ₃	SPSO-IP	8.88E-16	709	0.646	8.88E-16	571	0.715
	MPSO-IP	8.88E-16	33	0.984	8.88E-16	18	0.991
<i>TF</i> ₄	SPSO-IP	0.0246	1604	0.198	0	1333	0.334
	MPSO-IP	0.0246	53	0.974	0.0246	219	0.891
<i>TF</i> ₅	SPSO-IP	0	709	0.646	0	571	0.715
	MPSO-IP	0	33	0.984	0	18	0.991
Average	SPSO-IP		1153	0.424		916	0.523
	MPSO-IP		91	0.955		79	0.960

Based on these results, it can be inferred that the MPSO-IP had the ability to optimize the test functions to the actual optimum value with high convergence rate, irrespective of the iteration and N size. In contrast, the SPSO-IP had very low convergence rate, though it had the equal ability to reach the actual optimum value. These inferences indicated that the MPSO-IP had a better performance in terms of the convergence rate as compared to SPSO-IP, but they were equal in terms of producing a reliable result. To justify the inferences, a t-test of paired two samples for means that was applied by Safaei et al. (2012) in their analysis was used to compare the optimization results of 50 consecutive runs.

The statistical t-test at a significant level of $\alpha=0.05$ with 49 degree of freedom (df) which was the critical $t_{0.05}$ that was equal to 2.01 was used to compare the mean optimized test function value, μ that was obtained from MPSO-IP and SPSO-IP. In this test, the output from MPSO-IP was assessed; it was statistically different from the SPSO-IP output. At this point, a hypothesis was developed as (Null hypothesis, H_0 : μ_1 (MPSO-IP) = μ_2 (SPSO-IP)), which meant that the mean optimized value of MPSO-IP was equal to the mean optimized value of SPSO-IP. Thus, it implied that MPSO-IP and SPSO-IP had equal performance. Alternative hypothesis, H_a : μ_1 (MPSO-IP) > μ_2 (SPSO-IP), inferred that the mean optimized value of MPSO-IP was higher than SPSO-IP. In other words, SPSO-IP was better than MPSO-IP.

The obtained results from the hypothesis test are tabulated in Table 4.10. The results showed that all tests were to accept the H_0 and only 4 tests came out to reject the H_0 which involved TF_2 and TF_4 . The rejection of TF_2 occurred at 2000 maximum number of iterations and both sizes of N . The rejection result in TF_4 occurred when the $N=50$ for both iterations. In these cases, the SPSO-IP was better than the MPSO-IP.

Table 4.10: Statistical t-test between MPSO-IP and SPSO-IP algorithms for 1000 and 2000 iterations and 25 swarm particle sizes at $\alpha=0.05$.

Iteration	Particle Swarm Size	TF	Algorithm	Best Value	Mean	Standard Deviation	t_{cal}	Result
1000	25	TF_1	SPSO-IP	0	17.3	17.1217	0.858	Accept Ho
			MPSO-IP	0	20.26	14.5642		
		TF_2	SPSO-IP	0	152.96	248.2271	0.007	Accept Ho
			MPSO-IP	0	153.28	148.0563		
		TF_3	SPSO-IP	8.8818×10^{-16}	8.8818×10^{-16}	0	1.769	Accept Ho
			MPSO-IP	8.8818×10^{-16}	0.0735	0.2941		
		TF_4	SPSO-IP	0.0312	0.1148	0.0534	0.499	Accept Ho
			MPSO-IP	0.0312	0.1209	0.0664		
		TF_5	SPSO-IP	0	0	0	1.769	Accept Ho
			MPSO-IP	0	0.06	0.2399		
	50	TF_1	SPSO-IP	0	10.66	11.1860	-	Accept Ho
			MPSO-IP	0	8	10.7457		
		TF_2	SPSO-IP	0	68.3	149.9579	0.729	Accept Ho
			MPSO-IP	0	87.84	121.5555		
TF_3		SPSO-IP	8.8818×10^{-16}	8.8818×10^{-16}	0	*NA	*NA	
		MPSO-IP	8.8818×10^{-16}	8.8818×10^{-16}	0			
TF_4		SPSO-IP	0	0.1030	0.0599	2.320	Reject Ho	
		MPSO-IP	0.0312	0.1322	0.0647			
TF_5		SPSO-IP	0	0	0	*NA	*NA	
		MPSO-IP	0	0	0			

*NA – Not applicable.

Table 4.10, continued

Iteration	Particle Swarm Size	TF	Algorithm	Best Value	Mean	Standard Deviation	t_{cal}	Result
2000	25	TF_1	SPSO-IP	0	13.48	13.6161	0.441	Accept Ho
			MPSO-IP	0	14.88	14.7699		
		TF_2	SPSO-IP	0	56.1	125.8214	3.915	Reject Ho
			MPSO-IP	0	194.02	228.2014		
		TF_3	SPSO-IP	8.8818×10^{-16}	8.8818×10^{-16}	0	1.000	Accept Ho
			MPSO-IP	8.8818×10^{-16}	0.0245	0.1733		
		TF_4	SPSO-IP	0.0245	0.1120	0.0491	1.093	Accept Ho
			MPSO-IP	0.0246	0.1228	0.0465		
		TF_5	SPSO-IP	0	0	0	1.000	Accept Ho
			MPSO-IP	0	0.02	0.1414		
	50	TF_1	SPSO-IP	0	58.32	358.3638	-0.883	Accept Ho
			MPSO-IP	0	13.52	16.3161		
		TF_2	SPSO-IP	0	42.04	98.9172	2.231	Reject Ho
			MPSO-IP	0	91.46	115.654		
		TF_3	SPSO-IP	8.8818×10^{-16}	8.8818×10^{-16}	0	*NA	*NA
			MPSO-IP	8.8818×10^{-16}	8.8818×10^{-16}	0		
		TF_4	SPSO-IP	0	0.0957	0.0422	4.252	Reject Ho
			MPSO-IP	0.0246	0.1358	0.0553		
TF_5		SPSO-IP	0	0	0	*NA	*NA	
		MPSO-IP	0	0	0			

*NA – Not applicable.

It was observed that 4 tests were unable to be executed, which involved the TF_3 and TF_5 for both maximum number of iterations and $N=50$. This was due to the equal results computed by both algorithms for the 50 consecutive runs. Thus, it made no difference at all between the paired mean and as a consequence, the t-test was unable to be applied. In other words, both algorithms were equal in performances.

4.5 Optimization of Rain Gauge Network

The rain gauge network has been optimized using geostatistical methods through two different optimization approaches, cross-validation and MPSO. For cross-validation method, LOO and AOI were applied to configure the optimum rain gauge network. LOO and AOI are opposite to each other in the way of configuring the best rain gauge network, where LOO reduces the number of rain gauge stations; on the others hand, AOI expands network size from the base network size to existing network size. Meanwhile, MPSO applies swarms intelligence to get the best network out of all combinations of the stations.

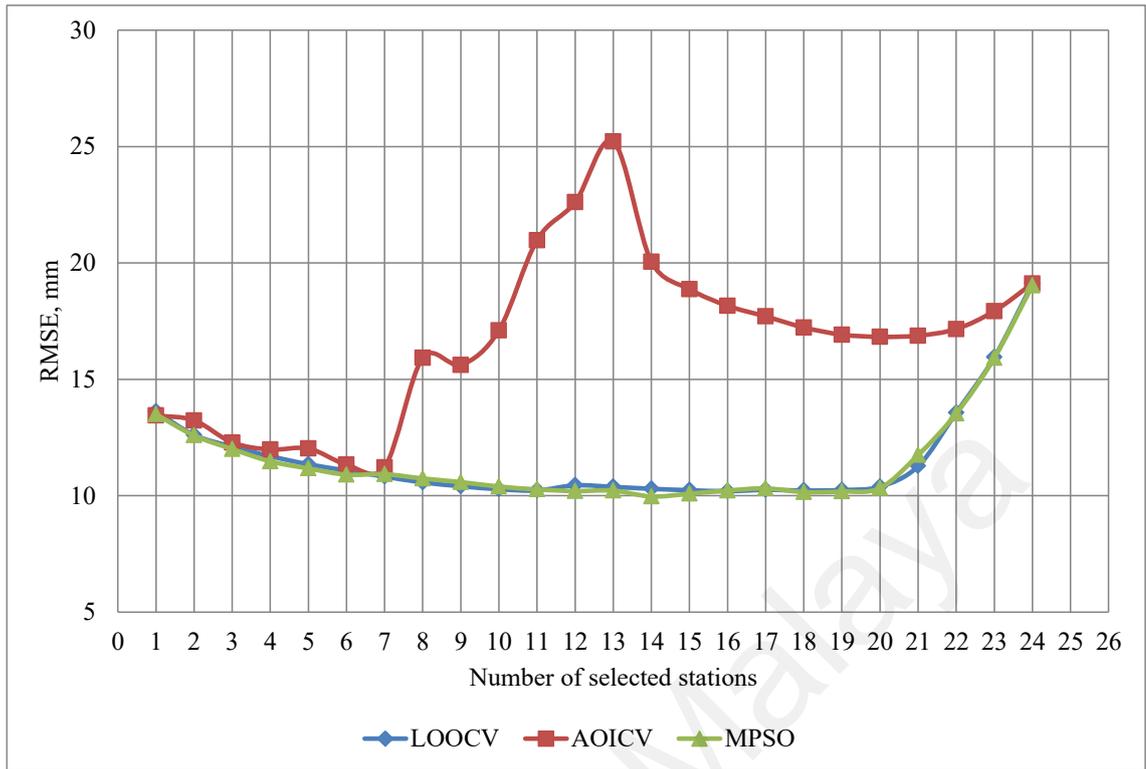
This subchapter presents the results and findings of the optimization process using the methodology that has been designed as described in Chapter 3.

4.5.1 Optimum Rain Gauge Network

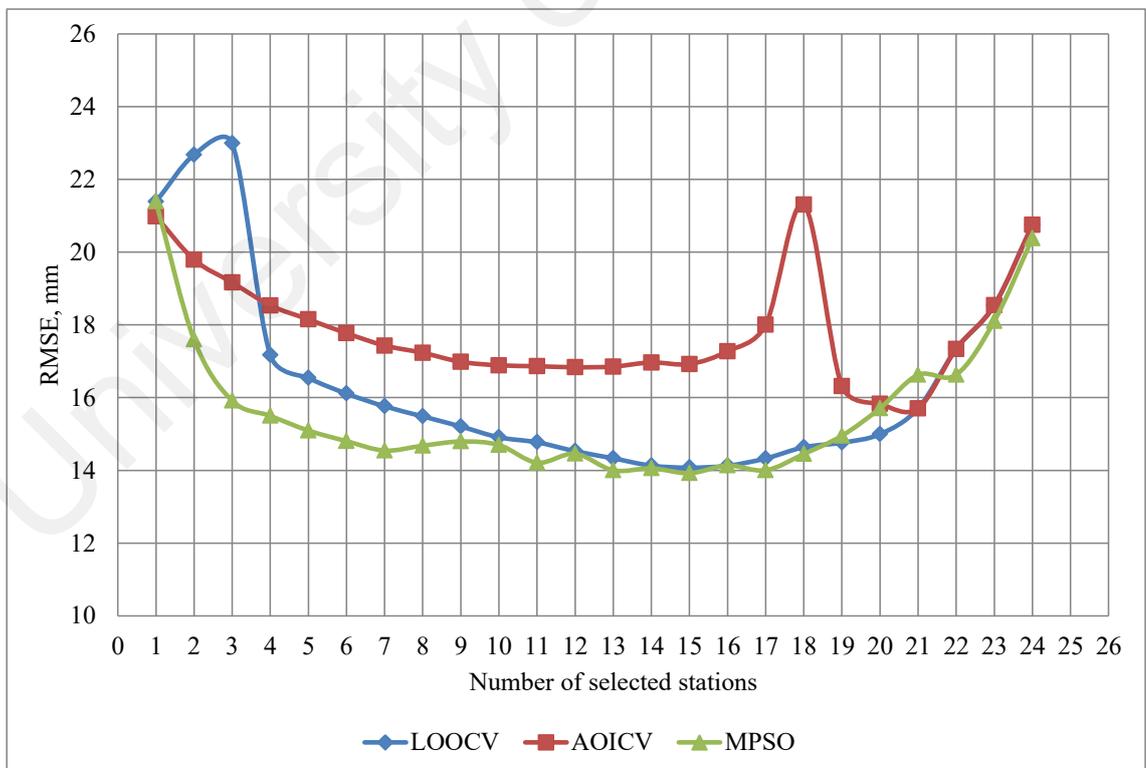
The main objective of this study is to obtain the optimal number of rain gauge stations in the network. Prior to optimizing the network, a candidate optimal network was configured by cross-validation technique, LOO and AOI. The AOI was introduced to re-validate the LOO and to explore the configured candidate networks for the better optimal network because the LOO would be biased towards certain rainfall datasets. Both methods will come up with E_{rms} values for every size of the optimal network. The

rain gauge network optimization results based on the E_{rms} value of spatial interpolation error for LOO and AOI were plotted against the total number of rain gauge stations in the optimized network. The results are illustrated in Figure 4.22, (a) to (g).

Generally, the results show that the E_{rms} value is not necessarily lowest for maximum or a minimum number of stations in the network, in fact, the lower E_{rms} occurs between the minimum and a maximum number of stations. However, there are results that had shown the almost equal value of E_{rms} for different size of networks as plotted in the subplot *a* for the LOO technique, subplot *d* for AOI technique and subplot *g* for both techniques. Another characteristic exhibited by the results is the availability of multi-point of minimum E_{rms} value as shown in subplot *c* for both technique and *f* for AOI. It is important to emphasize that this study was conducted based on the single objective of the optimization process. Thus, the network with the lowest E_{rms} was selected as the best-optimized network for both techniques and for each data set. It is worth it to incorporate the cost in the optimization as in Alfonso, Lobbrecht, and Price (2010) but the information from DID stated that the cost of operation and maintenance of each station are equal for the study area. In this case, the cost will incline with the increasing of the number of stations which mean that the selection of an optimal network will always go to the lower number of station irrespective the E_{rms} value even though the appearances of lowest value. In addition, in an optimization case, the maximum or minimum value of the objective function is the target to achieve.

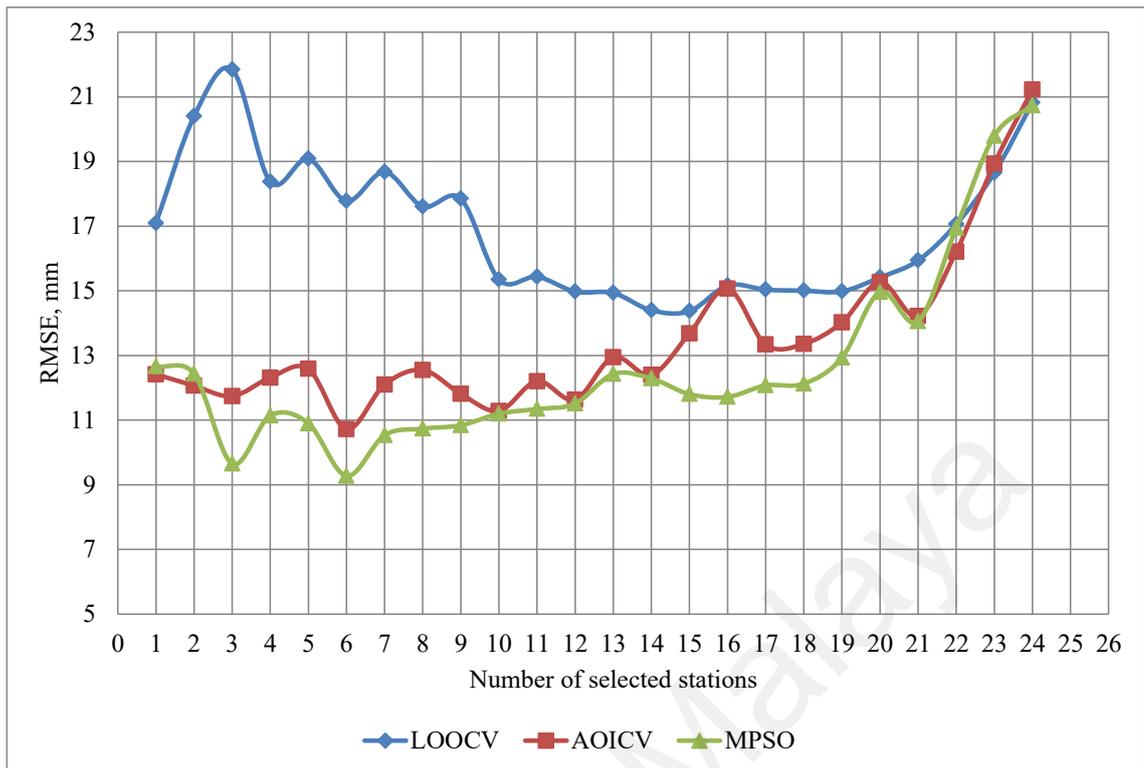


(a) 3 February 2009

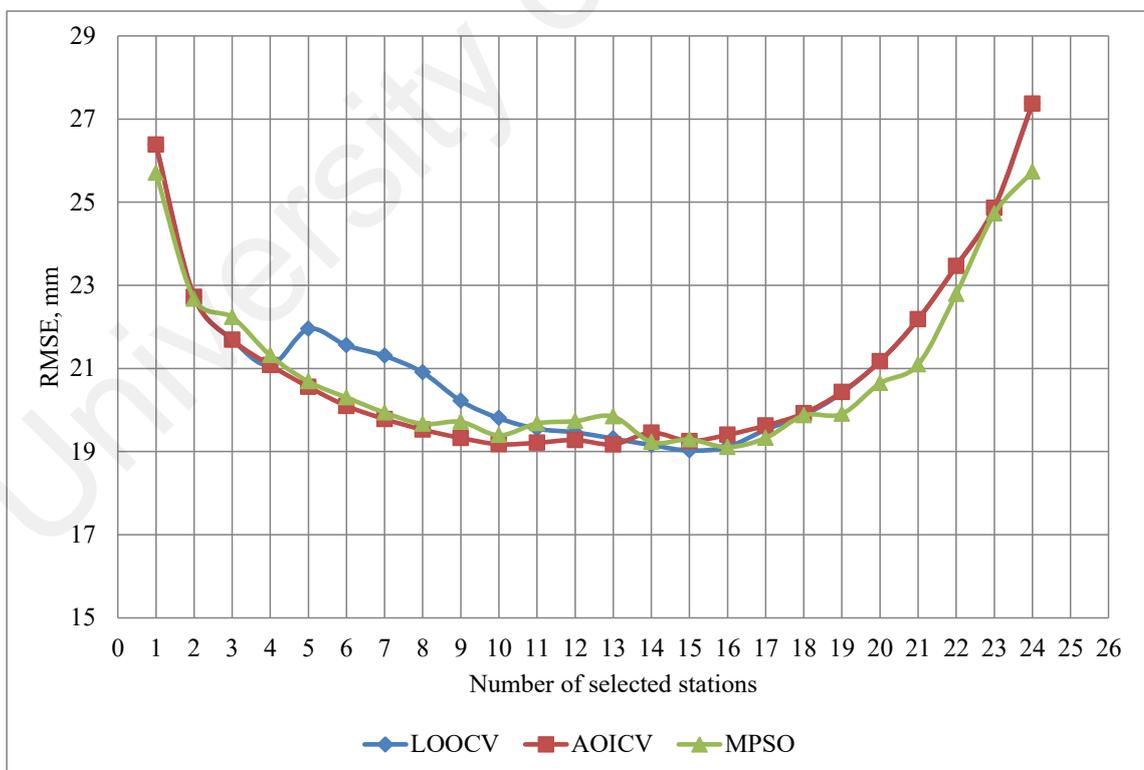


(b) 3 March 2009

Figure 4.22, continued

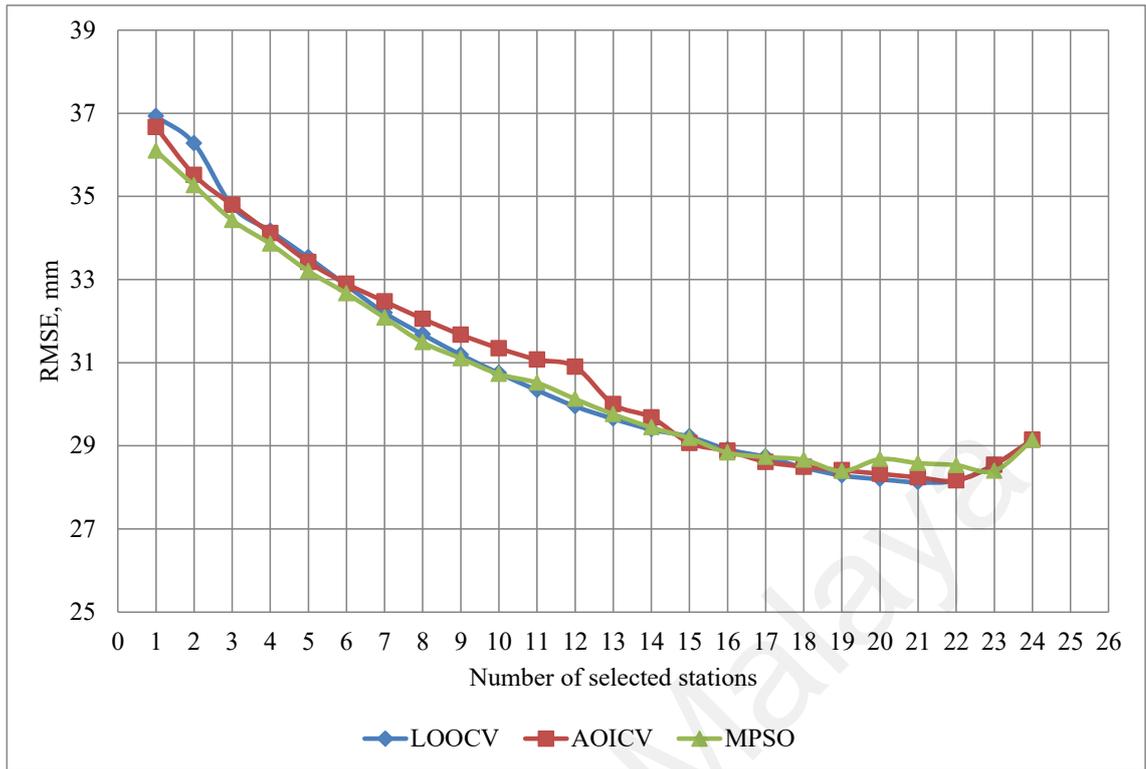


(c) 18 September 2011

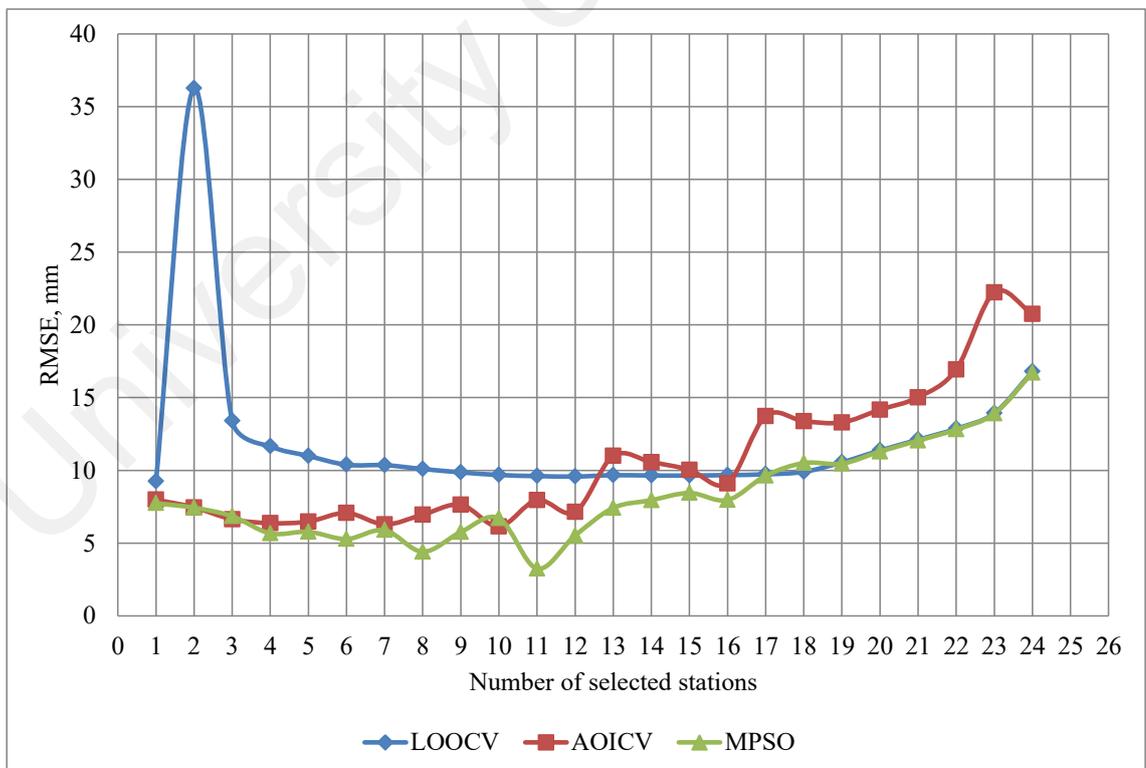


(d) 13 December 2011

Figure 4.22, continued

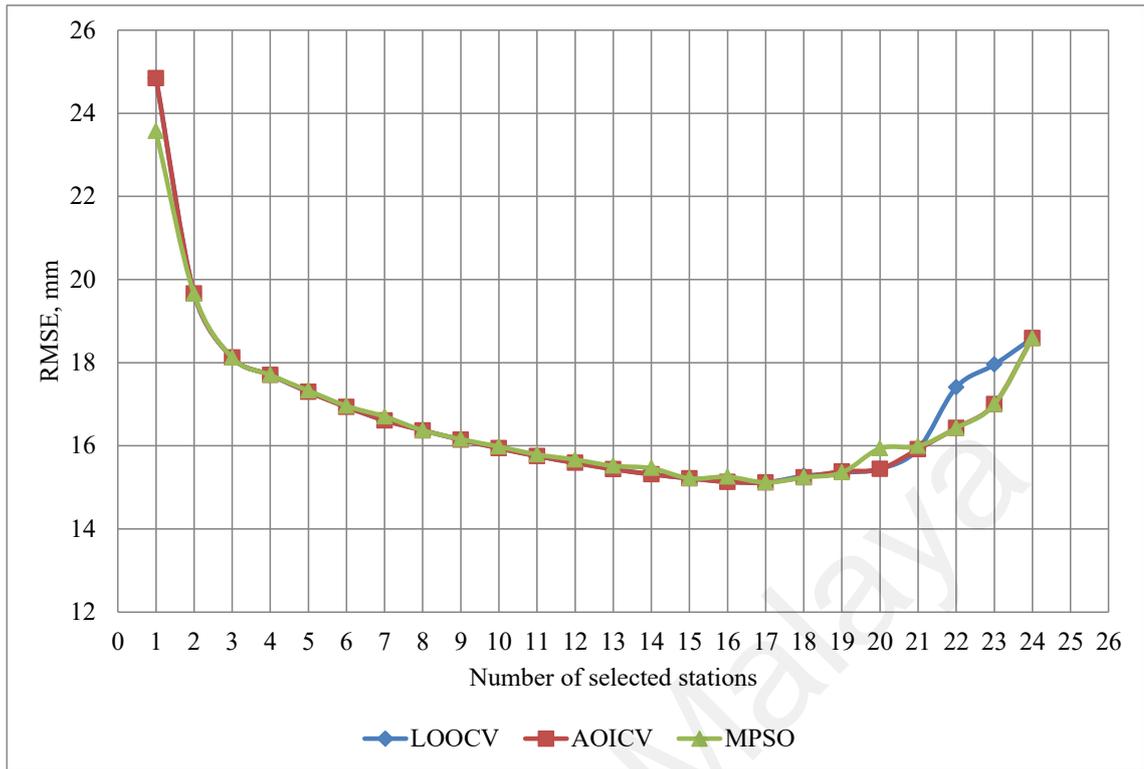


(e) 7 March 2012



(f) 18 April 2012 and

Figure 4.22, continued



(g) 21 August 2012.

Figure 4.22: Plotting of E_{rms} against the total number of stations in the network for the optimization result of CV-Geo (LOO and AOI) and MPSO.

Further assessment is carried out on the best-optimized network based on the lowest E_{rms} value and the Mean Absolute Error, E_{ma} was calculated as a measurement of the average E_{rms} of each number of subset stations. For each E_{rms} , their corresponding of the subset stations selected and the overlap stations between the LOO, AOI and MPSO were extracted from Figure 4.22 and tabulated in Table 4.11. In addition, the information on the total number of stations in the optimum network and their density were calculated and tabulated in Table 4.12.

The results given by all methods for each dataset were differed except for 21 August 2012, which had an E_{rms} value of 15.12. All methods presented similar results for the total number of stations in the network (36) with 17 selected subset stations and the

overlap of the selected subset stations (17) as well. Another quite similar result from all methods was the rainfall dataset for 7 March 2012. The numbers of stations in the network for LOO, AOI and MPSO were 40, 41 and 38, respectively, and the E_{rms} values were very close (28.122, 28.165 and 28.39, respectively). The AOI method produced 22 subset stations, which is one station more than LOO (21 stations). Meanwhile, the MPSO produced 19 subset stations and 18 subset stations were found overlapped.

Another almost similar E_{rms} value result was for the rainfall dataset on 13 December 2011. The E_{rms} values computed for LOO, AOI and MPSO were 19.026, 19.175 and 19.106, respectively. For this dataset, LOO selected two stations more for the total number of stations in the network and the selected subset stations than AOI (32 and 13, respectively). On the other hand, the MPSO produced 16 subset stations to form an optimal network contained a total of 35 numbers of stations. Out of the subset station selected by all method, 11 stations were found overlapped.

Table 4.11: Summary of E_{rms} and E_{ma} of optimized rain gauge network by CV-Geo (LOO and AOI) and MPSO with selected subset stations.

Rainfall Data	Error	Methods			
		Existing	LOO	AOI	MPSO
3 Feb 2009	E_{rms} Optimum	23.891	10.195	11.219	9.966
	E_{ma}	-	11.554	16.660	11.518
	Number of selected stations	25	16	7	14
	Overlap Station	5			
3 March 2009	E_{rms} Optimum	24.281	14.075	15.704	13.923
	E_{ma}	-	16.469	17.853	15.610
	Number of selected stations	25	15	21	15
	Overlap Station	13			
18 Sept 2011	E_{rms} Optimum	24.636	14.368	10.724	9.274
	E_{ma}	-	16.924	13.504	12.668
	Number of selected stations	25	15	6	6
	Overlap Station	1			
13 Dec 2011	E_{rms} Optimum	33.175	19.026	19.175	19.106
	E_{ma}	-	21.338	21.049	20.941
	Number of selected stations	25	15	13	16
	Overlap Station	11			
7 March 2012	E_{rms} Optimum	30.389	28.122	28.165	28.390
	E_{ma}	-	30.811	30.939	30.745
	Number of selected stations	25	21	22	19
	Overlap Station	18			
18 April 2012	E_{rms} Optimum	20.408	9.239	6.146	3.249
	E_{ma}	-	11.958	10.600	8.310
	Number of selected stations	25	12	10	11
	Overlap Station	4			
21 Aug 2012	E_{rms} Optimum	19.505	15.120	15.120	15.120
	E_{ma}	-	16.798	15.943	16.713
	Number of selected stations	25	17	17	17
	Overlap Station	17			

Table 4.12: The summary result of the total number of stations in the optimum network and their density by CV-Geo (LOO and AOI) and MPSO.

Flood Event	Total Number of Station*			Density, km ² /station*		
	LOO	AOI	MPSO	LOO	AOI	MPSO
3-Feb-09	35	26	33	16.7	22.5	17.7
3-Mar-09	34	40	34	17.2	14.6	17.2
18-Sep-11	34	25	25	17.2	23.4	23.4
13-Dec-11	34	32	35	17.2	18.3	16.7
7-Mar-12	40	41	38	14.6	14.2	15.4
18-Apr-12	31	29	30	18.8	20.1	19.5
21-Aug-12	36	36	36	16.2	16.2	16.2
Mean	34	32	33	16.8	18.5	18.0
Minimum	31	25	25	18.8	23.4	23.4
Maximum	40	41	38	14.6	14.2	15.4

Noted = *only stations located inside the catchment area considered.

The optimized networks produced by both methods were different in terms of the network size for each dataset. The minimum total numbers of stations in the optimized network by LOO is 31 stations. However, the AOI and the MPSO have produced an equal minimum number of stations (25 stations each). Meanwhile, the maximum numbers of stations were 40, 41, and 38, respectively. These results made the mean total numbers of stations were produced by LOO, AIO and MPSO were 34, 32 and 33, respectively and it can be concluded that the optimal size of rain gauge network is a range of 25 to 41 stations. These figures were compared with the output of the statistical result in subchapter 4.2.1, where at ϵ values of 20%, the range of stations is 10 to 41 stations. Between these two ranges, the statistical method has a wider range of value but both have a similar maximum range, 41 stations. A wider range of stations indicates the inconsistency of the network size. Moreover, the WMO's guide proposed the station s range between 29 to 58 stations for an urban area. Thus, the CV-Geo and MPSO results likely fulfil the WMO's guideline.

The density of the total number of stations in the optimized network over the study area was calculated and compared with the standard set by the World Meteorological Organization (WMO). For municipal areas, the WMO standard is the range of 10 km² to 20 km² per station (World Meteorological Organization, 2008). In this study, LOO fulfilled the WMO standard for all datasets with a minimum density of 18.8 km² per station and the maximum density of 14.6km² per station. The AOI, three out of seven results produced a density of more than 20 km² per station whereas, the average and minimum station densities are fulfilled by the WMO requirement. The MPSO come out with only one result that has density more than 20 km² but the average and minimum station densities have still met the WMO standard.

It is essential to note that the configuration of candidate networks was actually pre-determined for CV-Geo (LOO and AOI), with 325 candidate optimal networks considered in the optimization process for each flood event. In actual application, station selection in the network configuration was based on a combinatorial case (Pardo-Igu'zquiza, 1998). For instance, the number of possible combinations to select a subset of stations (r) from the number of stations (N) was determined with equation (8).

$${}^N C_r = \frac{N!}{r!(N-r)!} \quad \text{Equation 4.6}$$

The actual number of possible combinations based on the total of 25 stations ranged from 25 to 5,200,300. In the geostatistical analysis, different network configurations will produce different variography structures that are used for the spatial interpolation task, which will result in different error values. In this study, the combinatorial case has been incorporated in the MPSO to enhance the ability to explore more possible combinations. Based on the result in Table 4.11, MPSO comes out with 4 datasets that have the lowest E_{rms} value and 6 datasets with lowest E_{ma} value. These results proved that the MPSO has slightly better performance compared to the CV-Geo.

4.5.1.1 Variography structure of the optimized network.

The network optimized in this study should have had a better variography structure than the existing network after optimization. To investigate this matter, the variogram properties of each best network for each dataset (Table 4.11 and Table 4.12) were extracted and compared with the existing network. The comparison of variogram properties for the existing and optimized networks with both methods is tabulated in Table 4.13. The sill value comparison in the table indicates little difference after the number of stations was optimized. The mean sill value of the existing network was $1,252.94\text{mm}^2$, whereas with LOO, AOI and MPSO the values were $1,158.82\text{ mm}^2$, $1,293.82\text{ mm}^2$ and $1,245.30\text{ mm}^2$, respectively. However, there was an increase in the range value for both methods. AOI recorded the highest range of 15.59 km compared to LOO with 14.14 km, which is due to the low number of stations configured by both methods. Thus, a lower number of stations in the rain gauge network increased the range value. This also indicates that the rainfall data of the optimized network had a better correlation among stations.

The nugget values after the network was optimized were tremendously reduced compared with the existing network. MPSO recorded the lowest mean nugget value of 0.28 mm, whereas LOO produced a nugget value of 35.1 mm compared with the existing network value of 51.18 mm. Moreover, the existing network had a minimum nugget value of $1.106 \times 10^{-8}\text{ mm}$, which reduced to $4.7 \times 10^{-9}\text{ mm}$ for LOO and $4.32 \times 10^{-10}\text{ mm}$ for AOI following network optimization. These results showed that both methods improved the variogram structure for the nugget value, especially the MPSO method. Apparently, the optimized networks produced the lowest spatial interpolation error, which was the reason for nugget value improvement with both methods.

Although the optimized network by both techniques has a lowest spatial rainfall interpolation error, better variography structures and less variance of interpolation, it is important to prioritize the evaluated stations to distinguish which stations are classified as redundant in the network. The redundant stations could be removed in the first place to obtain an optimal network and perhaps rely on the financial constraint. Thus, we were conducted redundant station evaluation to prioritize them.

University of Malaya

Table 4.13: The comparison of variogram parameters values between the existing network and optimized network by CV-Geo (LOO and AOI) and MPSO.

	Sill, γ				Range, a (km)				Nugget			
	Existing	LOO	AOI	MPSO	Existing	LOO	AOI	MPSO	Existing	LOO	AOI	MPSO
3-Feb-09	693.097	494.519	828.514	558.166	15.341	12.815	19.690	11.411	47.140	1.400E-08	3.840E-09	1.962
3-Mar-09	930.505	1112.163	1042.418	1079.950	12.118	16.863	16.058	15.780	23.510	2.550E-08	8.870E-08	5.846E-08
18-Sep-11	450.123	491.922	1094.234	719.263	16.109	19.783	19.469	13.567	246.100	246.000	0.799	1.286E-08
13-Dec-11	983.093	1048.175	954.911	986.235	3.754	6.082	5.235	5.652	2.376E-08	4.700E-09	1.800E-07	2.872E-07
7-Mar-12	3703.179	3544.457	3737.750	3697.311	13.532	13.348	14.863	13.562	41.480	1.630E-07	19.600	1.390E-06
18-Apr-12	992.719	591.834	569.865	866.255	18.182	16.153	19.864	19.701	4.359E-08	1.030E-08	4.320E-10	6.483E-08
21-Aug-12	1017.863	828.638	828.638	809.925	13.989	13.950	13.950	13.623	1.106E-08	8.370E-08	8.370E-08	1.143E-07
Mean	1252.940	1158.815	1293.761	1245.301	13.289	14.142	15.590	13.328	51.176	35.143	2.914	0.280
Mini	450.123	491.922	569.865	558.166	3.754	6.082	5.235	5.652	1.106E-08	4.700E-09	4.320E-10	1.286E-08
Max	3703.179	3544.457	3737.750	3697.311	18.182	19.783	19.864	19.701	246.100	246.000	19.600	1.962

4.5.1.2 Redundant rain gauge station

In rain gauge network optimization, the result may depend on the type of rainfall event used. The optimized network will consist of different rain gauge stations from an event to another, but there will be certain stations that appear frequently in the optimized network. In this study, the optimization task carried out using the datasets showed there were stations that overlapped frequently for every event (Table 4.11), which demonstrated the great importance to the network. In contrast, a few stations were incorporated in the optimized network less frequently and these can be considered redundant stations with less influence on the spatial rainfall distribution.

Therefore, to prioritize the hypothetical redundant stations selected for the optimized network using the two methods with seven flood events, the frequency rate (F_r) was calculated by dividing the station frequency by the total number of flood events (seven). The F_r value signifies the station's importance to remain in the rain gauge network and the redundant station should have a lower F_r value. The F_r value was in the $[0, 1]$ range and it was sorted in descending order, as seen in Table 4.14. The F_r value also plotted for a better presentation of which station is often selected in the optimized network as shown in Figure 4.23. To evaluate the redundant station, first, the threshold F_r value of 0.5 was set to benchmark the redundant station. Those stations with F_r values below 0.5 were deemed less important. Then the stations in this category were checked for overlap between the two methods.

Based on the results, 6 stations were below the LOO threshold value. On the other hand, the AOI and MPSO resulted in 9 and 10 stations, respectively. Those stations were compared with the stations overlapping in threshold F_r value and there were only four stations (T06, N08, N15, and N20). These four stations can be inferred to be ineffective and were classified as redundant stations. An analysis of individual stations

indicated that stations T06 and N20 had the lowest average rainfall depths (16.7 mm and 26.8 mm respectively) for seven flood events. Smaller rainfall value is normally less effect on the spatial rainfall interpolation and thus these stations were likely ineffective. In addition, these stations were located quite close to another station, which was not listed as redundant. For instance, station N20 is adjacent to station T08 at a distance of only 5.57 m. Meanwhile, station T06 is separate at a distance of 599.17 m from station S15. Station N08 and N15 are located on the west side of the catchment area, which has fewer stations. This indicates the needs for additional stations to improve the interpolation error of this area. For this reason, it was good to relocate T06 and N20 to the west to enhance the sparse stations in this area. This is an improvement opportunity to be explored in future studies since it was not included in the scope of this study. The F_r values are plotted in Figure 4.23 for a better representation of which stations are often selected in an optimized network.

The four redundant stations were also evaluated using the variability of receiving rainfall along the available records and were measured as the probability of zero rainfall that is calculated by $P(z = 0) = 1 - p$, where p is the probability of wet days in which the station recorded non-zero rainfall values. This evaluation is adopted from Yoo et al. (2008) where it is used in their study to compare the mixed and continuous distribution function's application on entropy theory for rain gauge network evaluation. However, the probability of zero rainfall is adopted to validate the selection of redundant stations based on the higher probability of zero rainfall value. The higher probability value of zero rainfall shown that the location is less efficient to gauge rainfall thus it's justified the selection of the station as redundant.

As mentioned earlier, the N20 is very close to T08 and meanwhile, the T06 is near to the S15. Thus, the P values of N20 and T06 were compared to those of T08 and S15,

respectively. It was observed that N20 and T06 had higher P values (0.648 and 0.671, respectively) than T08 and S15 (0.55 and 0.507, respectively). This supports the reason why N20 and T06 can be classified as redundant stations but not T08. However, both N08 and N15 had moderate P values (0.436 and 0.451, respectively).

Table 4.14: Frequency rate, F_r of subset stations chosen in the optimized rain gauge network.

LOOCV		AOI		PSO	
ID	F_r	ID	F_r	ID	F_r
N14	1.00	N19	1.00	T03	0.86
N19	0.86	N10	0.86	T05	0.86
T05	0.86	T05	0.71	N14	0.86
T07	0.86	N09	0.71	N19	0.86
N04	0.86	T01	0.71	T07	0.71
N18	0.86	T08	0.71	N04	0.71
T03	0.71	N18	0.57	N06	0.71
N06	0.71	T03	0.57	N10	0.71
N09	0.71	N06	0.57	N23	0.71
T01	0.71	N16	0.57	N05	0.57
N16	0.71	N21	0.57	N07	0.57
N21	0.71	T02	0.57	N09	0.57
T02	0.71	N05	0.57	N16	0.57
N10	0.57	T04	0.57	N18	0.57
N05	0.57	N07	0.57	N21	0.57
N23	0.57	N03	0.57	T01	0.43
T04	0.57	N14	0.43	T02	0.43
T08	0.57	T07	0.43	T04	0.43
N17	0.57	N04	0.43	T06	0.43
N07	0.43	N23	0.43	T08	0.43
N03	0.43	N17	0.43	N03	0.43
T06	0.43	N08	0.43	N17	0.43
N08	0.43	T06	0.29	N08	0.29
N20	0.29	N15	0.29	N15	0.14
N15	0.14	N20	0.14	N20	0.14

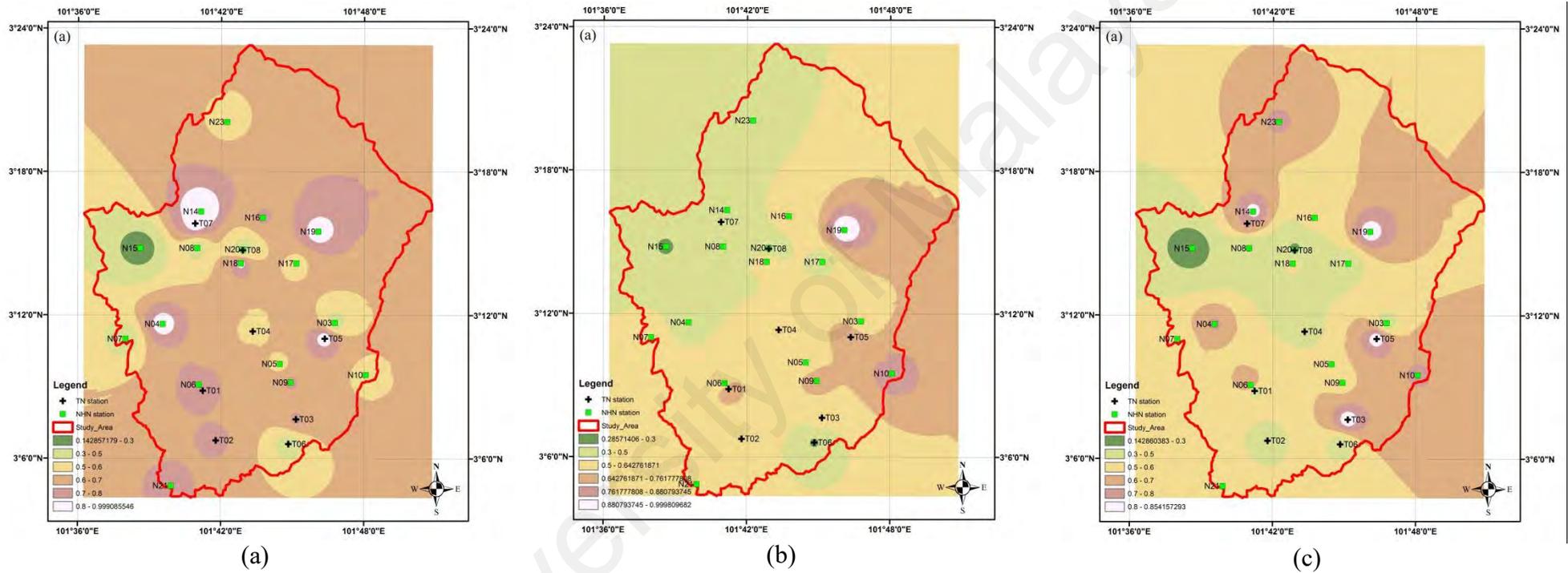


Figure 4.23: Frequency rate (F_r) map of stations remaining in the optimal network: a) optimized by LOO, b) optimized by AOI and c) optimized by MPSO.

4.6 TM-UKRB development

The optimal network that is without the redundant stations was evaluated to simulate the flood events used in the optimization process. The network should be able to simulate the flood events with a satisfactory level of model efficiency compared to the simulation used the existing network. For this purpose, the methodology that was explained in Chapter 3 is applied and the results are presented and discussed. The result is presented in two parts, first is the calibration result and second is the performance of the optimized network to simulate the flood events.

4.6.1 Calibration and Validation Result of TM-UKRB

The TM-UKRB model parameters are calibrated using SPSO for the best NSE value and the model performance was evaluated using three performance indicators, NSE, r and R^2 . The calibration result of TM-UKRB model parameters and model performance indicators are tabulated in Table 4.15.

Table 4.15: The calibration result of TM-UKRB

Calibration date on 18 September 2011	Parameter	Value
Performance Indicator	NSE	0.9937
	r	0.9968
	R^2	0.9937
Model Parameter	Ts1	7.6854
	Ts2	2.3234
	Ts3	0.2534
	Ev	0.5884
	M	1.1851
	X1	15.1911
	X2	9.0543
	K1	0.0381
	K2	0.1087
	K3	0.2106
K4	0.0969	
K5	0.0137	

Based on this result, the hydrologic model of TM-UKRB has demonstrated high model efficiency. The calibration process produced NSE values greater than 0.9. This result is supported by the r and R^2 values which are also more than 0.9. As all indicators value showed high model efficiency and a strong correlation between the simulated and observed flood, the TM-UKRB is expected to be a substantial model.

The TM-UKRB model's parameter values have undergone validation process and the results of validations are tabulated in Table 4.16. Based on this table, the high NSE values for validation results were produced. The minimum NSE value for validation recorded for data dated on 3 March 2009. Meanwhile, the maximum value is recorded by data on 21 August 2012, 0.9406. Based on this result, the model with the calibrated value parameters can be considered as a good model to be used for hydrologic simulation. The calibration hydrograph is illustrated in Figure 4.24. For further analysis, the calibrated parameters values of TM-UKRB were evaluated for sensitivity towards the model efficiency using Sensitivity Index (SI).

Table 4.16: The NSE value for validation process

Calibration	Validation		
18 September 2011	3 March 2009	18 April 2012	21 August 2012
0.99370	0.85970	0.93230	0.94060

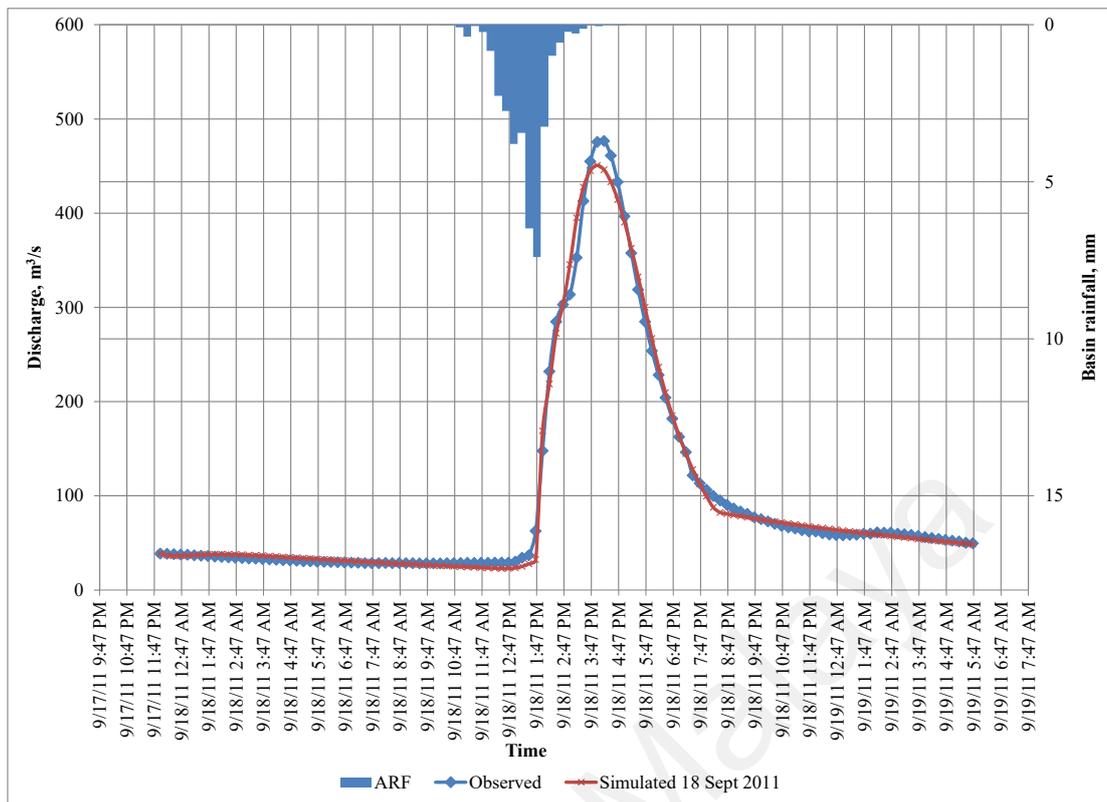


Figure 4.24: Hydrograph of calibration result using rainfall data on 18 September 2011.

A sensitivity analysis was conducted on the parameters using OAT approach and the result is presented in Figure 4.25. The figure shows that at the NSE more than 0.9, 5 parameters, K4, X2, EV, X1 and K2 effect the performance of the model significantly. Graphically, three parameters, namely K4, X2, and EV are very sensitive for both increase and decrease. However, X1 and K2 are only sensitive as the value decrease.

The SI value for all parameters was calculated using Equation 3.34 and tabulated in Table 4.17. The SI value less than 0.05 is considered as very less sensitive to negligible and on the other hand, SI more than 0.05 is considered as a medium to high sensitive (Lenhart et al., 2002). Based on the result, five parameters mentioned above have a higher SI value (more than 0.05). The K4 has the highest SI value, 0.426406. Meanwhile, X2, Ev and X1 have more or less similar value (0.26824, 0.240243 and

0.235316, irrespectively). On the other hand, K2 has a SI value of 0.114811. The rest of the parameters have SI value less than 0.01. Theoretically, the K4 is representing the infiltration rate of the topsoil layer and K2 is representing the runoff coefficient of the intermediate flow and the X1 and X2 are representing maximum storage that can be retained by the topsoil and intermediate soil layer. These parameters are very sensitive to calibrate in hydrological tank model. However, from Figure 4.25, the sensitive parameters are able to contribute a high NSE value (0.9) of the model in a range of $\pm 25\%$ in changes. Within this range, the parameters can be considered as a robust set of parameter for the TM-UKRB model.

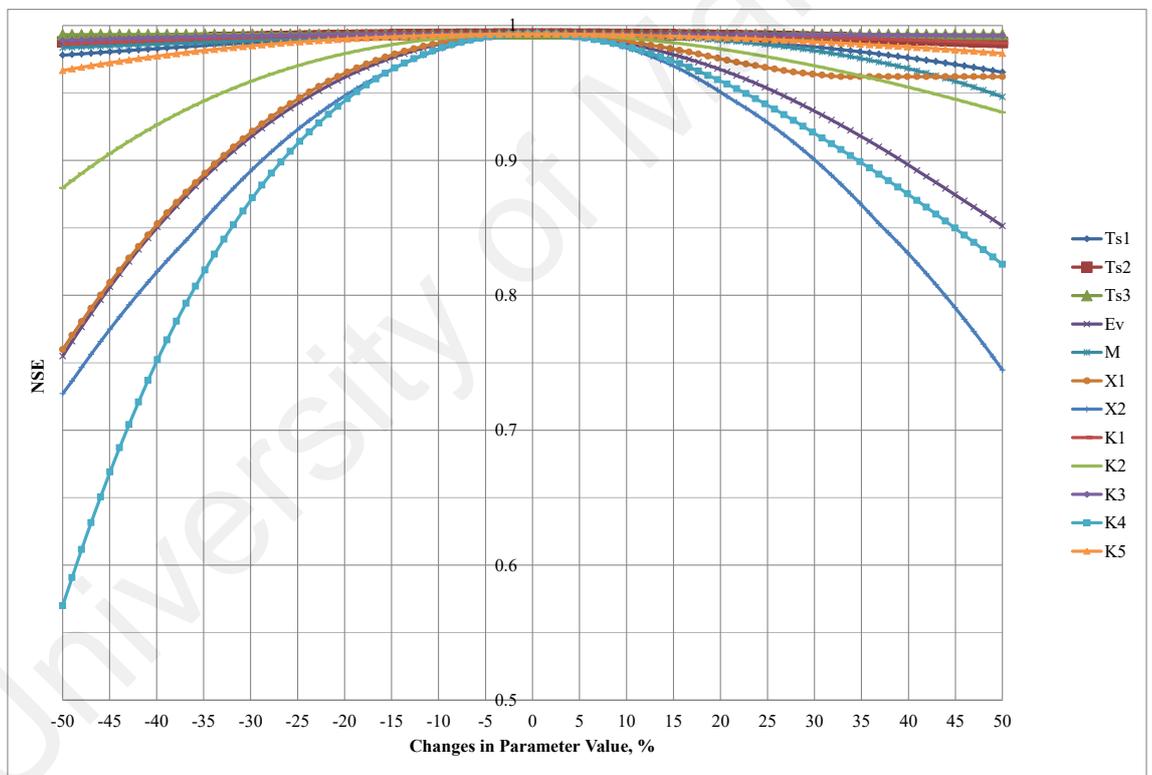


Figure 4.25: Sensitivity analysis of Upper Sg Klang Tank Model parameter.

Table 4.17: Sensitivity Index, SI of TM-UKRB Model Parameter

Parameters	SI
Ts1	0.028553
Ts2	0.006521
Ts3	0.000484
Ev	0.240243
M	0.046975
X1	0.235316
X2	0.26824
K1	0.007564
K2	0.114811
K3	0.005041
K4	0.426406
K5	0.027215

4.6.2 Optimal Network Performance in Simulating Flood Events

The optimal network without the four redundant stations was preliminarily evaluated to simulate the flood events used in the optimization process. The network should be able to simulate the flood events with a satisfactory level of model efficiency than the existing network. For this purpose, a hydrological Tank Model, TM-UKRB was used to simulate floods at stations located downstream of the catchment area near station N21.

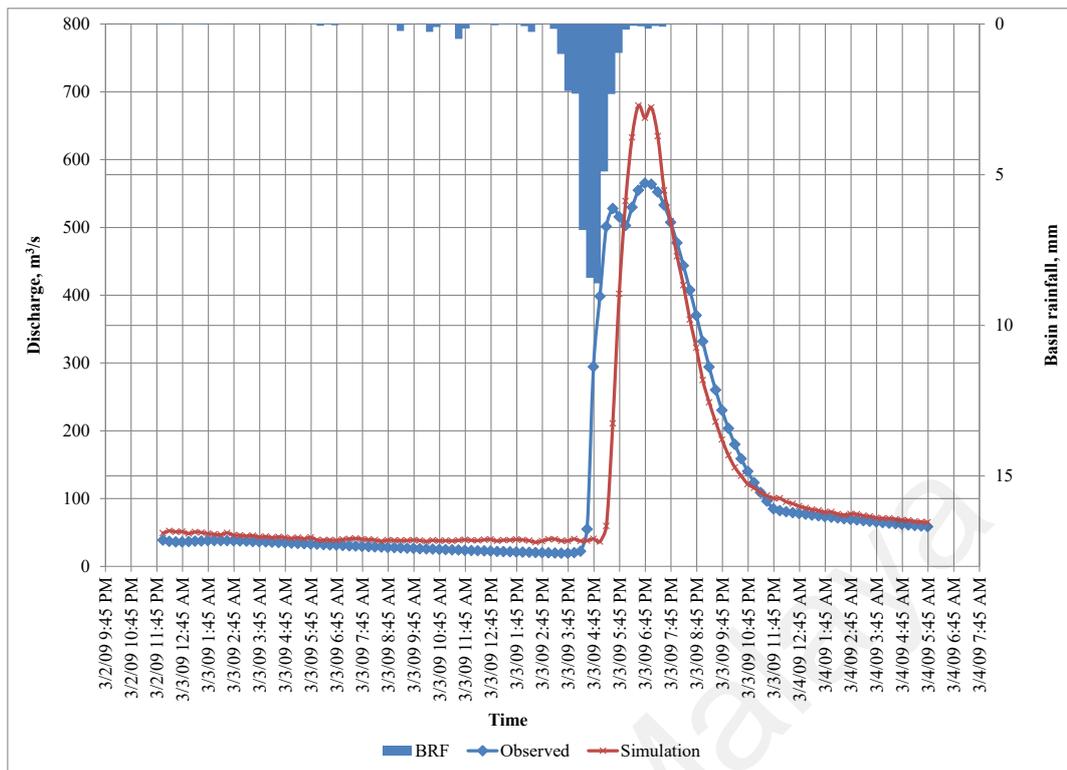
The hydrograph of the simulation results are plotted in Figure 4.26 and the performance indicators values of the simulation results are tabulated in Table 4.18. Based on the result, all datasets have NSE, r and R^2 values more than 0.75, which are efficient enough to simulate a flood hydrograph, except for the datasets on 18 April 2012, where the results of NSE_{opt} recorded less than 0.5. For this dataset, the four redundant stations deteriorated the model efficiency. This dataset also recorded the highest Absolute Error (AE) values of 47.4%, in contrast, the rest of the results have AE

value less than 10%. These results are possibly due to the characteristics of the rainfall events (e.g. convective or stratified), which are not included in the scope of this study.

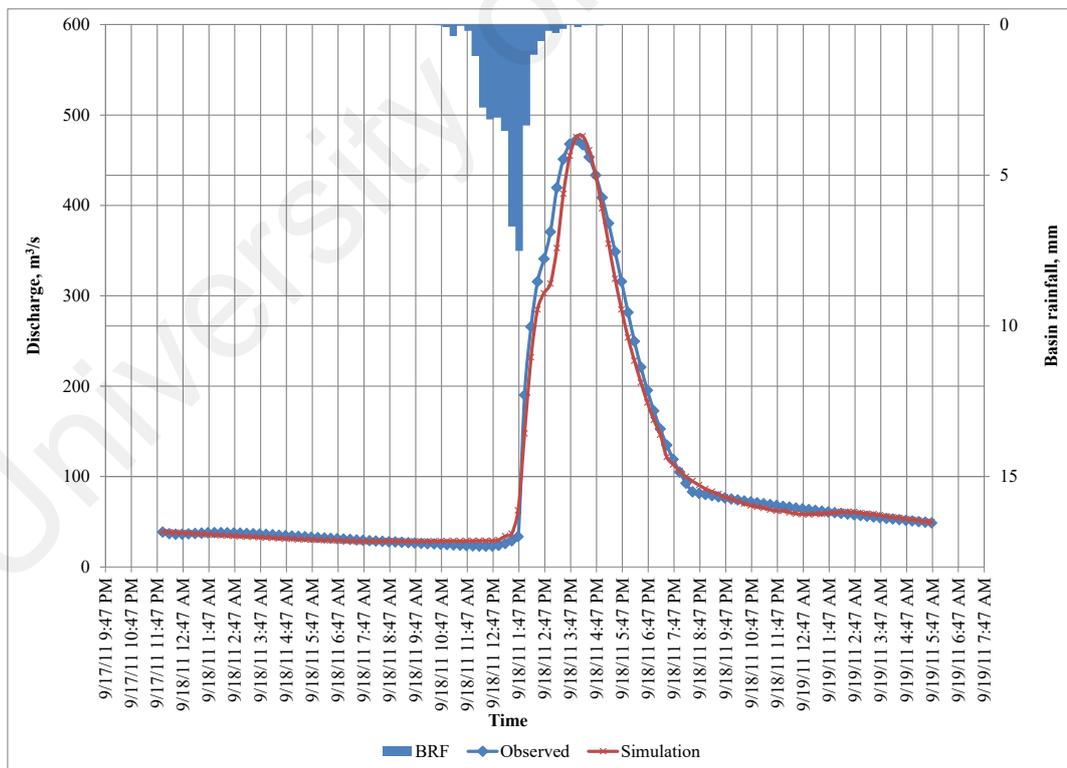
Even the performance indicators showed that the optimal network has a good performance to simulate the flood hydrograph; however, the comparison of the observed and simulated hydrographs showed that two datasets have similar hydrograph shapes. The rest of the datasets dated in the year 2012 did not simulate the flood very well at the high flow level, especially dataset on 18 April 2012. However, the dataset on 21 August 2012 has good rising and falling hydrograph part but did not reach a similar observed high flow. Based on the literature, a river transformation project called ‘river of life’ was implemented in early 2012 (A. R. Othman & Majid, 2016). One of the scopes of the project is to redevelop the river channel for beautification purpose. This is possibly the main reason why the model can not simulate the flood hydrograph for the datasets. Nonetheless, the model can be improved by re-calibrating it with the optimal rainfall network, as suggested by Bardossy and Das (2008).

Table 4.18: Validation of the optimal rain gauge network

Date	NSE_{ref}	NSE_{opt}	<i>r</i>	<i>R</i>²	AE (%)
3 March 2009	0.8597	0.794	0.9015	0.8127	7.64
18 September 2011	0.9937	0.9862	0.9960	0.9920	0.75
18 April 2012	0.9323	0.4904	0.9229	0.8517	47.40
21 August 2012	0.9406	0.8555	0.9753	0.9511	9.05

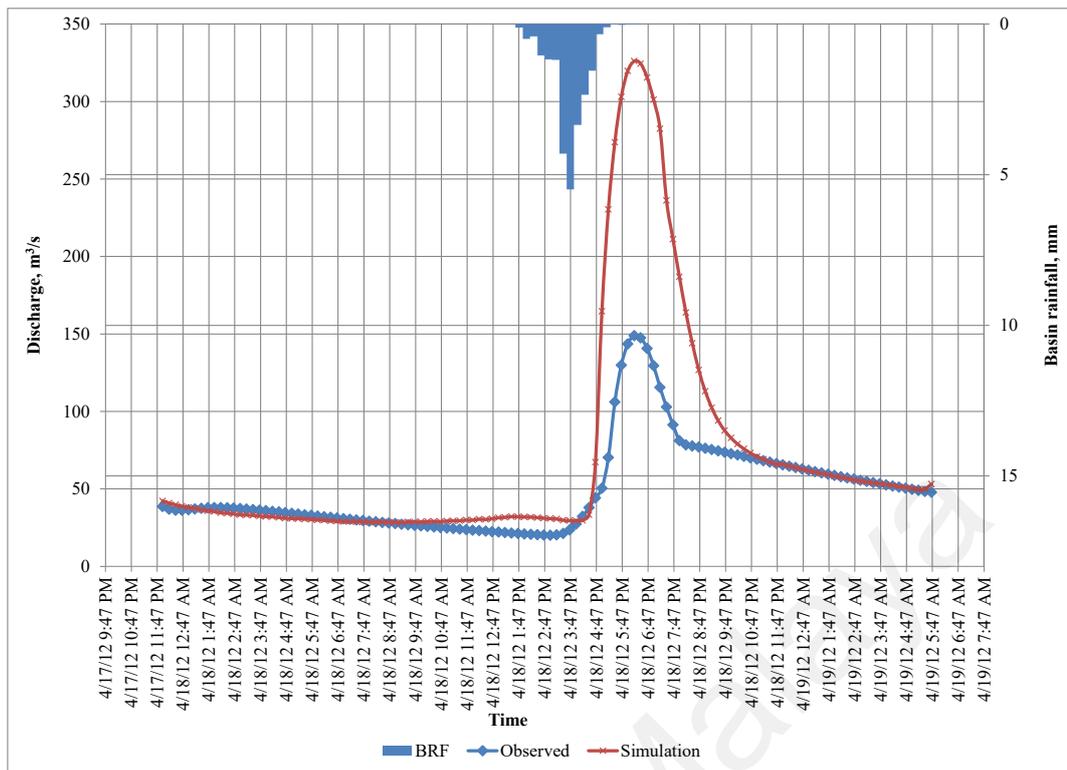


(a) 3 March 2009

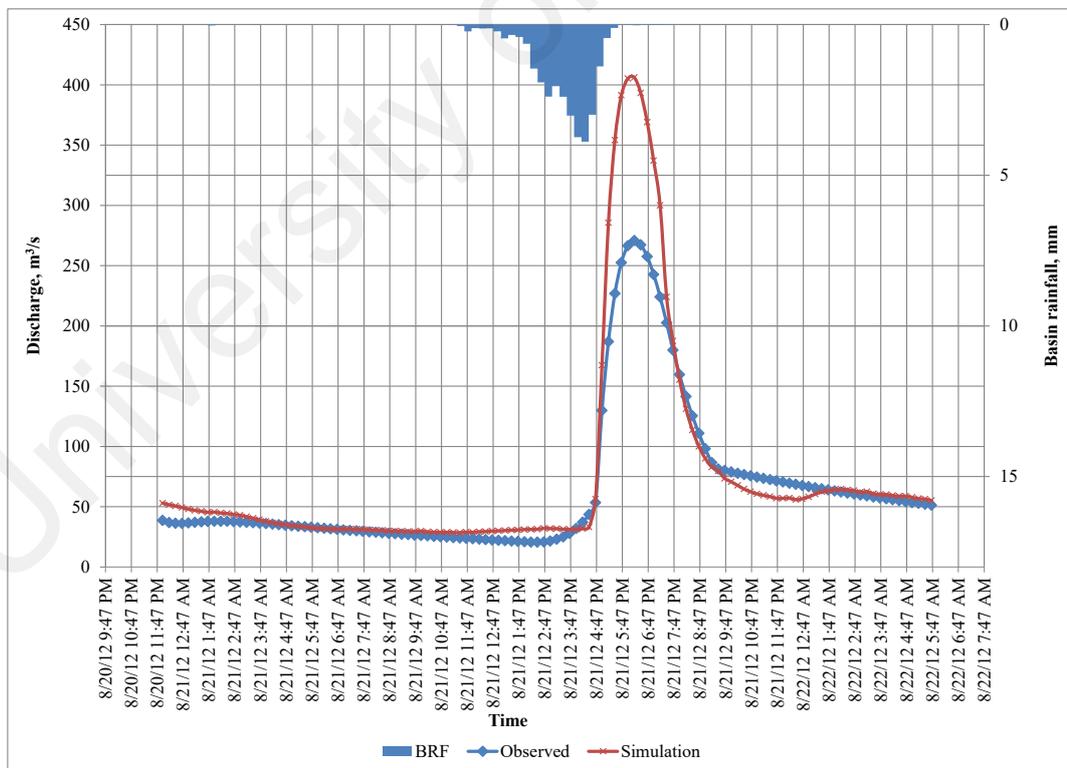


(b) 18 September 2011

Figure 4.26, continued



(c) 18 April 2012



(d) 21 August 2012

Figure 4.26: Comparison of the observed and simulated flood hydrograph of the optimal rain gauge network.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter briefly summarizes the findings and highlighting the contributions of this thesis. Recommendations are also presented for the new direction of future research.

The aim of this thesis is to optimize the number of rain gauges for effective rainfall data acquisition in a specific catchment. For this purpose, the geostatistical approach is adopted in the optimization process where two different tools of optimization, the Cross-Validation technique, and MPSO method are applied. To achieve this aim, four objectives have been determined. The selection of the semivariogram model for spatial interpolation process has been executed using the AHP method to achieve the first objective. The second objective, the PSO algorithm is modified (distinguished as MPSO) for a rain gauge network optimization problem. Then, the third objective, the existing rain gauge network is optimized for an optimal number of stations to gain an optimized network. The last objective is to validate the optimized rain gauge network by simulating flood events using a hydrological Tank Model. The conclusions of each objective are discussed in the next subchapter.

5.2 The selection of the appropriate semivariogram model

A Multi-criteria Decision-making method, namely the AHP is adopted to evaluate the performance of a variogram model to produce the best spatial rainfall map. Geostatistical analysis was applied to study the spatial structure of the cumulative daily rainfall data in the upper Klang River basin. Four criteria were determined as performance indicators of a spatial rainfall map: root-mean-square error, average standard error, means standardized error and root-mean-square standardized error, to

assess the five variogram model candidates, i.e. Spherical, Tetraspherical, Pentaspherical, Exponential, and Gaussian.

The AHP results indicate that the Spherical and Gaussian models indicate effective performance to produce a spatial rainfall map that fulfills the performance indicators. Two out of three sets of priority weights used in this study had both these models ranked in top place with scores of at least 3.56. Moreover, the Spherical model was found to be slightly better than the Gaussian model because all priority weight sets had similar results that ranked the Spherical model in the first place. The smoothness of the contour map represents the spatial rainfall distribution generated by the Spherical model, justifying its performance.

5.3 The performance of MPSO

The PSO algorithm was modified and distinguished as MPSO to solve the integer-based optimization problem such as a rain gauge network optimization problem. The modifications of the original PSO were made. However, the simplicity of the algorithm was preserved. The algorithms have been tested for two scenarios, a single run test for the controlled and uncontrolled Random Number (r_1 and r_2) value and multi-run test for the uncontrolled Random Number (r_1 and r_2) value. Based on the tests conducted, the conclusions on MPSO performance could be outlined as follows:

1. For both test scenarios, the MPSO had a better convergence rate (cr) compared to the SPSO, regardless of the maximum number of iterations and particles swarm size (N). The MPSO has recorded the average cr value as lower as 0.852 meanwhile the SPSO only recorded the average cr value as higher as 0.523. The MPSO was also found iteration independent since the algorithm used fully randomized parameters value. Furthermore, the MPSO was able to converge to

the lowest best value for the test functions similar to SPSO. However, both algorithms have the potential to trap the global value.

2. The MPSO has been observed to have an equal ability as the SPSO to give reliable optimization results. A statistical t-test of the paired two samples between the MPSO and the SPSO results failed to reject the null hypothesis (the mean optimized value by MPSO was equal to the mean optimized value by SPSO).

5.4 The rain gauge network optimization

In this study, two optimization processes have been developed by coupling cross-validation with geostatistical analysis and MPSO to prioritize the rain gauge stations in a network through optimization. The method was applied to optimize the number of stations in a rain gauge network in a tropical urban area. The total daily rainfall data from 55 rain gauge stations were used to perform the optimization process for 7 flood events. The aim of optimization was to reduce the number of rain gauge stations in the existing network for an optimal network. Four important points are summarized as follows:

1. By using the two new methods, CV-Geo and MPSO, the number of stations in the existing rain gauge network could be optimized based on the lowest E_{rms} value of the spatial interpolation error. However, at the lowest E_{rms} value, both approaches resulted in a different total number of stations in the optimized rain gauge network.
2. The optimized rain gauge network exhibited a better semivariogram structure, especially in the nugget value that has been drastically improved by both methods. However, MPSO had shown a slightly better nugget value since it has recorded the lowest value of the nugget.

3. The MPSO is the modified version of SPSO which has been purposely developed to solve the rain gauge network optimization problem. The MPSO performance is acknowledged as the application to optimize the number of rain gauge network. This is one of the novelties in this study, which is to contribute to the hydrological and swarm intelligent applications to solve a practical problem of rain gauge network optimization.
4. In the current study area, the rain gauge stations were prioritized based on their importance in the network (represented by F_r value). Four stations, namely T06, N08, N15, and N20 have been considered ineffective and could, therefore, be relocated within the study area or eliminated from the existing network.
5. A preliminary evaluation of the optimized network without the four mentioned stations showed satisfactory results in flood simulation using a lump hydrologic model. Three out of four flood simulations have yielded the NSE, r , and R^2 values more than 0.75, which have indicated that the optimized network is efficient enough to produce rainfall data to simulate a flood hydrograph.

5.5 Recommendations for future research

The recommendation for future research is primarily to improve the optimization method developed in this study for a better result, and the recommendations are as follows:

1. Despite the fact that the MPSO method has been found to have good optimization performance, the application of the method in the optimization of the rain gauge network has indicated the possibility of the method to trap at global value. It is essential to explore the other improvement needed by the method. Moreover, it is important to highlight that the value of the parameter of the algorithm was an ensemble from the previous study. Thus, it is suggested in the future to adopt an

advanced method to do the calibration process to enhance the algorithm's parameters value. To this end, the MPSO-IP has been applied in a real optimization problem, and it is also an opportunity to apply it to solve other similar problems in future studies.

2. This study adopted a single-objective of the optimization problem where the spatial interpolation error as the objective function was minimized to gain the optimal size of the rain gauge network. The reason why a single objective case was adopted was because both novel optimization tools developed in this study are new in hydrological network design, to the best of knowledge. However, an essential fact regarding the application of both methods in other fields is that their ability to solve a multi-objective optimization problem. Thus, it is a great contribution to apply the hydrological network design using a multi-objective optimization approach in future studies. For instance, incorporating the NSE of the hydrologic simulation as another objective function to optimize the rain gauge network.
3. A new age hydrological research will allow the application of radar rain measurement as input data. Currently, the advancement of radar technology and the increased computer capacity have encouraged its application in hydrological research. One of the advantages of radar rain data is that the availability of the spatial rain information that can be utilized to enhance the spatial interpolation process by the geostatistical method. This offers a great perspective to utilize the radar rain and ground rain measurements as input data for the optimal design of a rain gauge network. In this case, the optimum rain gauge network would be more robust for both conditions.

REFERENCES

- Abdel-Kader, R. F. (2011). Fuzzy Particle Swarm Optimization with Simulated Annealing and Neighborhood Information Communication for Solving TSP. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2(5), 15-21.
- Adhikary, S. K., Yilmaz, A. G. , & Muttill, N. (2015). Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrological Processes*, 29(11), 2582-2599. doi: 10.1002/hyp.10389
- Al-Abadi, A. M. A., & Al-Aboodi, A. H. D. (2014). Optimum Rain-Gauges Network Design of Some Cities in Iraq. *Journal of Babylon University, Engineering Sciences*, 22(4), 946-958.
- Alfonso, L., Lobbrecht, A., & Price, R. (2010). Optimization of water level monitoring network in polder systems using information theory. *Water Resources Research*, 46(12), p. W12553. doi: 10.1029/2009wr008953
- Alfonso, L., Ridolfi, E., Gaytan-Aguilar, S., Napolitano, F., & Russo, F. (2014). Ensemble Entropy for Monitoring Network Design. *Entropy*, 16(3), 1365-1375. doi: 10.3390/e16031365
- Attar, M., Abedini, M. J., & Akbari, R. (2018). Optimal prioritization of rain gauge stations for areal estimation of annual rainfall via coupling geostatistics with artificial bee colony optimization. *Journal of Spatial Science*, 64(2), 257-274. doi: 10.1080/14498596.2018.1431970
- Awadallah, A. G. (2012). Selecting Optimum Locations of Rainfall Stations Using Kriging and Entropy. *International Journal of Civil & Environmental Engineering IJCEE*, 12, 36-41.
- Aziz, M. K. B. M., Yusof, F., Daud, Z. M., Yusop, Z., & Kasno, M. A. (2016). Optimal design of rain gauge network in Johor by using geostatistics and particle swarm optimization. *International Journal of GEOMATE*, 11(25), 2422-2428. doi: 10.21660/2016.25.5137
- Bakhtiari, B., Kermani, M. N., & Bordbar, M. H. (2013). Rain Gauge Station Network Design for Hormozgan Province in Iran. *DESERT*, 45-52.
- Barca, E., Passarella, G., & Uricchio, V. (2008). Optimal extension of the rain gauge monitoring network of the Apulian Regional Consortium for Crop Protection. *Environ Monit Assess*, 145(1-3), 375-386. doi: 10.1007/s10661-007-0046-z
- Bardossy, A., & Das, T. (2008). Influence of rainfall observation network on model calibration and application. *Hydrology and Earth System Sciences*, 12(1), 77-89.
- Bastin, G., Lorent, B., Duque, C., & Gevers, M. (1984). Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. *Water Resources Research* 20, 463-470.

- Bras, R. L., & Colon, R. (1978). Time-averaged areal mean of precipitation: Estimation and network design. *Water Resources Research*, 14(5), 878-888.
- Bras, R. L., & Rodríguez-Iturbe, I. (1976). Network design for the estimation of areal mean of rainfall events. *Water Resources Research*, 12(6), 1185–1195. doi: 10.1029/WR012i006p01185
- Chao, L., Singh, V. P., & Mishra, A. K. (2012). Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information minimum redundancy. *Water Resources Research*, 48(5), W05521. doi: 10.1029/2011wr011251
- Chen, Y.C., Wei, C., & Yeh, H.C. (2008). Rainfall network design using kriging and entropy. *Hydrological Processes*, 22(3), 340-346. doi: 10.1002/hyp.6292
- Cheng, K. S., Lin, Y. C. , & Liou, J. J. (2008). Rain-gauge network evaluation and augmentation using geostatistics. *Hydrological Processes*, 22(14), 2554-2564. doi: 10.1002/hyp.6851
- Chiles, J. P., & Delfiner, P. (1999). *Geostatistics: Modelling Spatial Uncertainty*. New York: JohnWiley & Sons.
- Coulter, E. D., Coakley, J., & Sessions, J. (2006). The Analytic Hierarchy Process: A Tutorial for Use in Prioritizing Forest Road Investments to Minimize Environmental Effects. *International Journal of Forest Engineering*, 17(2), 51-69.
- Eberhart, R. C., & Shi, Y. (1995, Nov. 27 -Dec. 1 1995). *Particle Swarm Optimization*. Paper presented at the Proceedings of IEEE International Conference on Neural Networks.
- Eberhart, R. C., & Shi, Y. (2001, May 27 - 30, 2001). *Tracking and Optimizing Dynamic Systems with Particle Swarms*. Paper presented at the Proceedings of IEEE Congress on Evolutionary Computation, COEX, Seoul, Korea.
- Fan, H. (2010). Discrete Particle Swarm Optimization for TSP based on Neighborhood. *Journal of Computational Information Systems*, 6(10), 3407-3414.
- Feki, H., Slimani, M., & Cudennec, C. (2010). *A comparison of three geostatistical procedures for rainfall network optimization*. Sousse, Tunisia. <http://irec.cmerp.net/irec10/papers/REA/Paper%20ID51.pdf>
- Feki, H., Slimani, M., & Cudennec, C. (2012). Incorporating elevation in rainfall interpolation in Tunisia using geostatistical methods. *Hydrological Sciences Journal*, 57(7), 1294-1314. doi: 10.1080/02626667.2012.710334
- Feki, H., Slimani, M., & Cudennec, C. (2016). Geostatistically based optimization of a rainfall monitoring network extension: case of the climatically heterogeneous Tunisia. *Hydrology Research*. doi: 10.2166/nh.2016.256

- Foresti, L., Pozdnoukhov, A., Tuia, D., & Kanevski, M. (2010). Extreme Precipitation Modelling Using Geostatistics and Machine Learning Algorithms. *16*. doi: 10.1007/978-90-481-2322-3
- Garcia, M., Peters-Lidard, C. D. , & Goodrich, D. C. (2008). Spatial interpolation of precipitation in a dense gauge network for monsoon storm events in the southwestern United States. *Water Resources Research*, *44*(5), p. W05S13. doi: 10.1029/2006wr005788
- Goldberg, E. F. G., Goldberg, M. C., & Souza, G. R. (2008). Particle Swarm Optimization Algorithm for the Traveling Salesman Problem, *Travelling Salesman Problem* (pp. 202). Vienna, Austria: I-Tech.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.
- Goovaerts, P. (1999, 25 - 28 July 1999). *Performance Comparison Of Geostatistical Algorithms For Incorporating Elevation Into The Mapping Of Precipitation*. Paper presented at the Proceedings of the 4th International Conference on GeoComputation, Mary Washington College Fredericksburg, Virginia, USA.
- Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, *228*(1–2), 113-129. doi: [http://dx.doi.org/10.1016/S0022-1694\(00\)00144-X](http://dx.doi.org/10.1016/S0022-1694(00)00144-X)
- Grimes, D. I. F., & Pardo-Igúzquiza, E. (2010). Geostatistical Analysis of Rainfall. *Geographical Analysis*, *42*(42), 136–160.
- Hamby, D. M. (1995). A Comparison of Sensitivity Analysis Techniques. *Health Physics Journal* *68*(2), 195-204.
- Hoffman, F. O., & Gardner, R. H. (1983). Evaluation of uncertainties in environment radiological assessment model In J. E. Till & H. R. Meyer (Eds.), *Radiological Assessment: A Textbook on Environmental Dose Assessment*. Washington: US Nuclear Regulatory Commission.
- Hongliang, X., Chong-Yu, X., Sælthun, N. R., Youpeng, X., Bin, Z., & Hua, C. (2015). Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modelling – A case study of humid area in southern China. *Journal of Hydrology*, *525*, 138-151. doi: 10.1016/j.jhydrol.2015.03.034
- Ishizaka, A., & Lusti, M. (2006). How to derive priorities in AHP: a comparative study. *Central European Journal of Operations Research*, *14*(4), 387-400.
- Jabatan Pengairan dan Saliran. (2009). DID Manual, Volume 4 - Hydrology and Water Resources. Kuala Lumpur, Malaysia.
- Jabatan Pengairan dan Saliran. (2017). *Kompendium - Data dan Maklumat Asas*. Kuala Lumpur.
- Jin, W. (2011). Particle Swarm Optimization with Adaptive Parameter Control and Opposition. *Journal of Computational Information Systems*, *7*(12), 4463-4470.

- Johnston, K., Hoef, J. M. V., Krivoruchko, K., & Lucas, N. (2003). *ArcGIS 9 - Using ArcGIS Geostatistical Analyst*. United State of America: ESRI.
- Journel, A. G., & Huijbregts, C. J. (1978). *Mining Geostatistics*. New York: Academic Press.
- Kar, A. K., Lohani, A. K., Goel, N. K., & Roy, G. P. (2015). Rain gauge network design for flood forecasting using multi-criteria decision analysis and clustering techniques in lower Mahanadi river basin, India. *Journal of Hydrology: Regional Studies*, 4, 313-332. doi: 10.1016/j.ejrh.2015.07.003
- Kassim, A.H. M., & Kottegoda, N. T. (1991). Rainfall Network Design Through Comparative Kriging Methods. *Journal of Hydrological Science*, 36(3), 223-240.
- Kennedy, J., & Eberhart, R. C. (1997). *A discrete binary version of the particle swarm algorithm*. Paper presented at the IEEE International Conference on Systems, Man, and Cybernetics, Hyatt Orlando, Orlando, Florida, USA
- Krstanovic, P. F., & Singh, V. P. (1992a). Evaluation of Rainfall Networks Using Entropy: I. Theoretical Development. *Water Resources Management*, 6, 279-293.
- Krstanovic, P. F., & Singh, V. P. (1992b). Evaluation of Rainfall Networks Using Entropy: II. Application. *Water Resources Management*, 6, 295-314.
- Leach, J. M., Kornelsen, K. C., Samuel, J., & Coulibaly, P. (2015). Hydrometric network design using streamflow signatures and indicators of hydrologic alteration. *Journal of Hydrology*, 529, 1350-1359. doi: 10.1016/j.jhydrol.2015.08.048
- Lenhart, T., Eckhardt, K., Fohrer, N., & Frede, H.-G. (2002). Comparison of two different approaches of sensitivity analysis. *Physics and Chemistry of the Earth*, 27, 645-654.
- Ly, S., Charles, C., & Degré, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, 15(7), 2259-2274. doi: 10.5194/hess-15-2259-2011
- Mishra, A. K., & Coulibaly, P. (2009). Developments in hydrometric network design: A review. *Reviews of Geophysics*, 47(2). doi: 10.1029/2007RG000243View/save
- Mishra, A. K., & Coulibaly, P. (2010). Hydrometric network evaluation for Canadian watersheds. *Journal of Hydrology*, 380(3-4), 420-437. doi: 10.1016/j.jhydrol.2009.11.015
- Mishra, A. K., & Coulibaly, P. (2014). Variability in Canadian Seasonal Streamflow Information and Its Implication for Hydrometric Network Design. *Journal of Hydrologic Engineering*, 19(8), 05014003. doi: 10.1061/(ASCE)HE.1943-5584.0000971

- Mu, A. Q., Cao, D. X., & Wang, X. H. (2009). A Modified Particle Swarm Optimization Algorithm. *Natural Science*, 01(02), 151-155. doi: 10.4236/ns.2009.12019
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology* 10(3), 282-290. doi: 10.1016/0022-1694(70)90255-6
- Nemec, J., & Askew, A. J. (1986). *Mean And Variance In Network-Design Philosophies*. Paper presented at the Integrated Design of Hydrological Networks (Proceedings of the Budapest Symposium, July 1986). IAHS Publ. no. 158,1986.
- Othman, A. R., & Majid, N. H. A. (2016). Urban River and Its Heritage Value: A river of life at Precinct 7, Kuala Lumpur. *Environment-Behaviour Proceedings Journal*, 1(1), 58. doi: 10.21834/e-bpj.v1i1.272
- Othman, F., Akbari, A., & Samah, A. A. (2011). Spatial rainfall analysis for an urbanized tropical river basin. *International Journal of the Physical Sciences*, 6(20), 4861-4868. doi: 10.5897/ijps11.754
- Papamichail, D. M., & Metaxa, I. G. (1996). Geostatistical analysis of spatial variability of rainfall and optimal design of a rain gauge network. *Water Resources Management*, 10(2), 107-127. doi: 10.1007/BF00429682
- Pardo-Igu'zquiza, E. (1998). Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *Journal of Hydrology*(210), 206–220.
- Poli, R. (2008). Analysis of the Publications on the Applications of Particle Swarm Optimisation. *Journal of Artificial Evolution and Applications*, 2008, 1-10. doi: 10.1155/2008/685175
- Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization. *Swarm Intelligence*, 1(1), 33-57. doi: 10.1007/s11721-007-0002-0
- Putthividhya, A., & Tanaka, K. (2012). Optimal Rain Gauge Network Design and Spatial Precipitation Mapping Based on Geostatistical Analysis from Colocated Elevation and Humidity Data. *International Journal of Environmental Science and Development*, 3(2).
- Ravizi, M. K. (2012). *Modified Particle Swarm Optimization and Gravitational Search Algorithm for Slope Stability Evaluation* (Doctor of Philosophy), Universiti Kebangsaan Malaysia, Bangi, Malaysia.
- Ridolfi, E., Montesarchio, V., Russo, F., & Napolitano, F. (2011). An entropy approach for evaluating the maximum information content achievable by an urban rainfall network. *Natural Hazards and Earth System Science*, 11(7), 2075-2083. doi: 10.5194/nhess-11-2075-2011

- Ridolfi, E., Yan, K., Alfonso, L., Baldassarre, G. D., Napolitano, F., Russo, F., & Bates, P. D. (2012). *An entropy method for floodplain monitoring network design*. Paper presented at the AIP Conference, Gdańsk, Poland.
- Ruiz-Cárdenas, R., Ferreira, M. A. R., & Schmidt, A. M. (2010). Stochastic search algorithms for optimal design of monitoring networks. *Environmetrics*, *21*, 102-112. doi: 10.1002/env.989
- Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York, NY, U.S.A.: McGraw-Hill International.
- Saaty, T. L. (1990). How to make a decision: The Analytic Hierarchy Process. *European Journal of Operational Research*, *48*, 9-26.
- Saaty, T. L. (2008). Decision Making With The Analytic Hierarchy Process. *Int. J. Services Sciences*, *1*(1), 83-98.
- Saaty, T. L., & Vargas, L. G. (1991). *Prediction, projection, and forecasting: Applications of the analytic hierarchy process in economics, finance, politics, games, and sports*. Boston: Kluwer Academic Publishers.
- Safaei, N., Tavakkoli-Moghaddam, R., & Kiassat, C. (2012). Annealing-based particle swarm optimization to solve the redundant reliability problem with multiple component choices. *Applied Soft Computing*, *12*(11), 3462-3471. doi: 10.1016/j.asoc.2012.07.020
- Shaghaghian, M. R., & Abedini, M. J. (2013). Rain gauge network design using coupled geostatistical and multivariate techniques. *Scientia Iranica*, *20*(2), 259-269. doi: 10.1016/j.scient.2012.11.014
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, Ill: University of Illinois Press.
- Shi, X. H., Liang, Y. C., Lee, H. P., Lu, C., & Wang, Q. X. (2007). Particle swarm optimization-based algorithms for TSP and generalized TSP. *Information Processing Letters*, *103*(5), 169-176. doi: 10.1016/j.ipl.2007.03.010
- Shi, Y., & Eberhart, R. C. (1998). *A modified particle swarm optimizer*. Paper presented at the Proceedings of IEEE International Conference on Evolutionary Computation, Anchorage, AK.
- Sorman, U., & Balkan, G. (1983). An Application Of Network Design Procedures For Redesigning Kizilirmak River Basin Raingauge Network. *Hydrological Sciences Journal* *28*(2), 233-246.
- Tasgetiren, M. F., Suganthan, P. N., & Pan, Q. K. (2007). *A discrete particle swarm optimization algorithm for the generalized traveling salesman problem*. Paper presented at the Genetic and Evolutionary Computation Conference, GECCO'07, London, England, United Kingdom.

- Vivekanandan, N., & Jagtap, R. S. (2013a). Evaluation and Selection of Rain Gauge Network using Entropy. *Journal of The Institution of Engineers (India): Series A*, 93(4), 223-232. doi: 10.1007/s40030-013-0032-0
- Vivekanandan, N., & Jagtap, R. S. (2013b). Review and Analysis of Stream Gauge and Rain Gauge Networks Using Spatial Regression Approach. *Wyno Journal of Engineering & Technology Research*, 1(2), 10-21.
- Vivekanandan, N., Roy, S. K., & Chavan, A. K. (2012). Evaluation of Rain Gauge Network using Maximum Information Minimum Redundancy Theory. *International Journal of Scientific Research and Reviews*, 1(3), 96-107.
- Volkman, T. H. M., Lyon, S. W., Gupta, H. V., & Troch, P. A. (2010). Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain. *Water Resources Research*, 46(11). doi: 10.1029/2010wr009145
- Wei, C., Chiang, J. L., Wey, T. H., Yeh, H. C., & Cheng, Y. C. (2010). Rain Gauges Network Design using Discrete Entropy and Kriging. *Geophysical Research Abstracts*, 12(EGU2010-10421).
- World Meteorological Organization. (2008). Guide to Hydrological Practices *Volume I (Hydrology – From Measurement to Hydrological Information)* (6 ed.). Geneva 2, Switzerland: World Meteorological Organization.
- Xu, H., Xua, C. Y., Chen, H., Zhang, Z., & Li, L. (2013). Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. *Journal of Hydrology*, 505, 1-12. doi: 10.1016/j.jhydrol.2013.09.004
- Xuedan, L., Qiang, W., Haiyan, L., & Lili, L. (2009). *Particle Swarm Optimization with Dynamic Inertia Weight and Mutation*. Paper presented at the Third International Conference on Genetic and Evolutionary Computing, Guilin, China.
- Xuesong, Y., Can, Z., Wenjing, L., Wei, L., Wei, C., & Hanmin, L. (2012). Solve Traveling Salesman Problem Using Particle Swarm Optimization Algorithm. *International Journal of Computer Science and Network Security*, 9(6), 264-271.
- Xuesong, Z., & Srinivasan, R. (2009). GIS-Based Spatial Precipitation Estimation: A Comparison of Geostatistical Approaches. *Journal Of The American Water Resources Association*, 45(4), 894-906. doi: 10.1111/j.1752-1688.2009.00335.x
- Yang, X., Yuan, J., Yuan, J., & Mao, H. (2007). A modified particle swarm optimizer with dynamic adaptation. *Applied Mathematics and Computation*, 189(2), 1205-1213. doi: 10.1016/j.amc.2006.12.045
- Yeasmin, D., & Pasha, M. F. K. (2008). *Runoff Prediction by GIS using Optimal Number of Rain Gauges*. Paper presented at the World Environmental and Water Resources Congress, Honolulu, Hawaii, United States.

- Yeh, H. C., Chen, Y. C., Wei, C., & Chen, R. H. (2011). Entropy and kriging approach to rainfall network design. *Paddy and Water Environment*, 9(3), 343-355. doi: 10.1007/s10333-010-0247-x
- Yong, J., Hyunglok, K., Jongjin, B., & Minha, C. (2014). Rain-Gauge Network Evaluations Using Spatiotemporal Correlation Structure for Semi-Mountainous Regions. *Terr. Atmos. Ocean. Sci.*, 25(2), 267-278. doi: 10.3319/TAO.2013.10.31.01(Hy
- Yoo, C., Jung, K., & Lee, J. (2008). Evaluation of Rain Gauge Network Using Entropy Theory: Comparison of Mixed and Continuous Distribution Function Applications. *Journal Of Hydrologic Engineering*(13), 226-235. doi: 10.1061//asce/1084-0699/2008/13:4/226
- Yoon, D. L., & Lahiri, S. N. (2002). Least squares variogram fitting by spatial subsampling. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(4), 837-854.
- Zhang, J., & Yao, N. (2008). The geostatistical framework for spatial prediction. *Geospatial Information Science*, 11(3), 180-185. doi: 10.1007/s11806-008-0087-7

LIST OF PUBLICATIONS AND PAPERS PRESENTED

A. Published Journal Paper

1. Mohd Zaharifudin Muhamad, Ali, & Faridah Othman, (2017). Selection of Variogram Model for Spatial Rainfall Mapping Using Analytical Hierarchy Procedure (AHP), *Scientia Iranica*, 24(1), 28-39. (ISI, IF=0.475)
2. Mohd Zaharifudin Muhamad Ali, & Faridah Othman, (2018). Rain Gauge Network Optimization in A Tropical Urban Area by Coupling Cross Validation with the Geostatistical Technique, *Hydrological Sciences Journal*, 63(3), 474-491. (ISI, IF=2.061)

B. Journal Paper to be / ready for submission

1. Mohd Zaharifudin Muhamad Ali, & Faridah Othman,
Modified Particle Swarm Optimization Method for Rain Gauge Network Optimization Problem.