

https://dspace.org/

Assessment of daily stochastic rainfall generation in Southern Africa

Tjebane, Wendy

2022
https://hdl.handle.net/10539/34942

Downloaded from a DSpace repository. https://dspace.org/

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG



ASSESSMENT OF DAILY STOCHASTIC RAINFALL GENERATION IN SOUTHERN AFRICA

MSc Research Report (CIVN 7019A)

Wendy Tjebane Student no:604806

604806@students.wits.ac.za

19 September 2022

Table of Contents

Declaration iii

Dedication iv

Acknowledgements v

Abstract vi

List of Figures vii

List of Tables xii

List of Abbreviations xiii

1. INTRODUCTION 0

1.1. Background 0

1.2. Problem Statement 2

1.3. Objectives 2

1.4. Organization of the Research Report 3

2. LITERATURE REVIEW 4

2.1. Review of Stochastic Hydrological Models 4

2.2. WGEN (Weather Generator) Method 8

2.3. Transition Probability Matrix (TPM) Method 9

2.4. Variable Length Bootstrap (VLB) Method 11

2.5. Comparative Review of the methods 13

3. METHODOLOGY 14

3.1. Introduction 14

3.2. Research Approach 14

3.3. Statistical analysis of sourced historic observed data 15

3.3.1. Hebron Station Rainfall Statistical Evaluation 17

3.3.2. Welgevonden Station Statistical Evaluation 19

3.3.3. Nelshoogte Station Statistical Evaluation 21

3.4. GoldSim Software 23

3.5. Transition Probability Matrix (TPM) Procedure 24

3.6. Weather Generator (WGEN) Procedure 30

3.7. Variable Lengths Bootstraps (VLB) Procedure 33

4. PERFORMANCE EVALUATION OF THE GENERATORS 37

4.1. Performance Evaluation of the TPM Method 39

4.1.1. Hebron Rainfall station 39
4.1.2. Welgevonden Rainfall 44
4.1.3. Nelshoogte Rainfall Station 50
4.2. Performance Evaluation of WGEN Method 55
4.2.1. Hebron Rainfall Station 55
4.2.2. Welgevonden Rainfall station 60
4.2.3. Nelshoogte Rainfall station 64
4.3. Performance Evaluation of VLB Method 68
4.3.1. Hebron Rainfall station 68
4.3.2. Welgevonden Rainfall station 72
4.3.3. Nelshoogte Rainfall station 77
5. COMPARISON OF GENERATORS 82
5.1. Hebron Rainfall Station Comparison 82
5.2. Welgevonden Rainfall Station Comparison 91
5.3. Nelshoogte Rainfall station Comparison 98
6. CONCLUSIONS AND RECOMMENDATIONS 107
7. REFERENCES 109

Declaration

I __Wendy Tjebane_____ Student number: ___604806_____I am a student registered for Master of Science in Engineering_____ in the year ___2022_____. I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that **ALL** the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: ___  _____

Date: _____19/09/ 2022_____

Dedication

This is in memory of Mmane, Mme, Mamogolo, Mangwane, Malome and Papa. Although they are no longer part of this world, their memories continue to be source of inspiration and motivation in my life. I am who I am today because of all of you. Thank you.

Acknowledgements

I firstly would like to thank my supervisor, Professor John Ndiritu for the guidance that he provided into completing this research. I am grateful for his constructive criticism, his patience with me and his hard work that added substantial amount of value to this research.

My sincere gratitude is expressed to my family and my partner who have supported my academic journey and constantly checked up on my progress.

I would like to also thank South African Weather Services (SAWS) as they kindly provided the most recent observed rainfall data sets that were used in the research. I also would like to thank GoldSim Technology Group for providing the academic license that was used to carry out parts of the research in GoldSim Software ®.

Special thanks to Gauteng Department of Education for providing funding throughout my studies. Finally, I would like to thank everyone that I have not mentioned above, who have shown me support throughout my studies.

Abstract

This research report sets out to assess daily stochastic rainfall generation applied on Southern African historic rainfall data. The research aims to evaluate the performance of three methods, namely, Transition Probability Matrix (TPM), Weather Generator (WGEN) and Variable Length Bootstrap (VLB). The evaluation was based on the ability of the methods to preserve the daily, monthly, and annual statistics of the observed data, simplicity, and ease of application of the methods. From the evaluation of the performances from all three methods, it can be concluded that all the methods are capable of generating daily data with statistics that preserve the observed historic data, although to different degrees of accuracy. It was also found that utilization of software such as GoldSim makes the application of the methods more convenient and faster.

The evaluations have shown that it is imperative to have a comprehensive understanding of both advantages and disadvantages of the methods and their limitations. Although some of the methods were originally not developed with intention of being applied in South Africa, they can generate data for different climatic conditions that they were not originally developed for.

It is recommended that future research focus more on how the impacts of climate change can be incorporated into the methods. Additional performance assessment for other areas of Southern Africa including the winter rainfall regions of the Western Cape Province and drier areas of South Africa is recommended.

List of Figures

Figure 3-1: Summary of main steps of the Methodology 15
Figure 3-2: Locations of the rainfall stations (Google Earth, 2022) 16
Figure 3-3: Daily rainfall depths measured at Hebron (0594760 W) rainfall station 18
Figure 3-4: Annual rainfall depths measured at Hebron (0594760 W) rainfall station 18
Figure 3-5: Monthly Box plot for the measured rainfall depths at Hebron (0594760 W) rainfall station.. 19
Figure 3-6: Daily rainfall depths measured at Welgevonden (0594764 W) rainfall station 20
Figure 3-7: Annual rainfall depths measured at Welgevonden (0594764 W) rainfall station 20
Figure 3-8: Monthly Box plot for the measured rainfall depths at Welgevonden (0594764 W) rainfall station 21
Figure 3-9: Daily rainfall depths measured at Nelshoogte (0518589) rainfall station 22
Figure 3-10: Annual rainfall depths measured at Nelshoogte (0518589 W) rainfall station 22
Figure 3-11: Monthly Box plot for the measured rainfall depths at Nelshoogte (0518589 W) rainfall station 23
Figure 3-12: Transition Probability Matrix Structure (Zhou et al, 2017) 25
Figure 3-13: Screenshot from GoldSim showing the calculations to generate the daily rainfall data 29
Figure 3-14: Boughton correction calculated using GoldSim software for generating daily rainfall..... 30
Figure 3-15: Typical Screenshot of Generated WGEN Parameters 32
Figure 3-16: Screenshot of WGEN Precipitation generator set up in GoldSim 33
Figure 3-17: Screenshot of the historic daily rainfall.txt 35
Figure 3-18: Screenshot of the RainfilesNames.txt 36
Figure 3-19: Screenshot of the VLBSettingsandfiles.txt file 36
Figure 3-20: Flow chart of the VLB Process 37
Figure 4-1: Box plot legend representing the various quartiles as used in the stochastic analysis 38
Figure 4.1-1: Comparison of the annual probability of exceedance for the generated annual precipitations to the recorded data for Hebron rainfall station 39
Figure 4.1-2: Hebron station rainfall distribution curve from the daily maximum values using TPM Method 40
Figure 4.1-3: Percentiles of daily data generated using TPM Method for Hebron station 40
Figure 4.1-4: Averages of monthly data generated using TPM Method for Hebron station 41
Figure 4.1-5: Maximum values of monthly data generated using TPM Method for Hebron station 42
Figure 4.1-6: Standard deviation values of monthly data generated using TPM Method for Hebron station 42
Figure 4.1-7: Percentiles of monthly data generated using TPM Method for Hebron station 43
Figure 4.1-8: Rainfall distribution curve of the annual totals for data generated using TPM Method for Hebron station 44
Figure 4.1-9: Annual percentiles of data generated using TPM Method for Hebron station 44

Figure 4.1-10: Comparison of the annual probability of exceedance for the generated annual precipitations to the recorded data for Welgevonden rainfall station 45
Figure 4.1-11: Welgevonden station rainfall distribution curve from the daily maximum values using TPM Method 46
Figure 4.1-12: Percentiles of daily data generated using TPM Method for Welgevonden station 46
Figure 4.1-13: Averages of monthly data generated using TPM Method for Welgevonden station 47
Figure 4.1-14: Maximum of monthly data generated using TPM Method for Welgevonden station..... 47
Figure 4.1-15: Standard deviation of monthly data generated using TPM Method for Welgevonden station 48
Figure 4.1-16: Percentiles of monthly data generated using TPM Method for Welgevonden station 48
Figure 4.1-17: Rainfall distribution curve of the annual totals for data generated using TPM Method for Welgevonden station 49
Figure 4.1-18: Annual percentiles of data generated using TPM Method for Welgevonden station 49
Figure 4.1-19:Comparison of the annual probability of exceedance for the generated annual precipitations to the recorded data for Nelshoogte rainfall station 50
Figure 4.1-20: Nelshoogte station rainfall distribution curve from the daily maximum values using TPM Method 51
Figure 4.1-21: Percentiles of daily data generated using TPM Method for Nelshoogte station 51
Figure 4.1-22: Averages of monthly data generated using TPM Method for Nelshoogte station 52
Figure 4.1-23: Maximum values of monthly data generated using TPM Method for Nelshoogte station 53
Figure 4.1-24: Standard deviation values of monthly data generated using TPM Method for Nelshoogte station 53
Figure 4.1-25: Percentiles of monthly data generated using TPM Method for Nelshoogte station 54
Figure 4.1-26: Rainfall distribution curve of the annual totals for data generated using TPM Method for Nelshoogte station 55
Figure 4.1-27: Annual percentiles of data generated using TPM Method for Nelshoogte station 55
Figure 4.2-1: Rainfall distribution curve of the daily maximum values for data generated using WGEN Method for Hebron station 56
Figure 4.2-2: Percentiles of daily data generated using WGEN Method for Hebron station 56
Figure 4.2-3: Averages of monthly data generated using WGEN Method for Hebron station 57
Figure 4.2-4: Maximum values of monthly data generated using WGEN Method for Hebron station..... 58
Figure 4.2-5: Standard deviation values of monthly data generated using WGEN Method for Hebron station 58
Figure 4.2-6: Percentiles of monthly data generated using WGEN Method for Hebron station 59
Figure 4.2-7: Rainfall distribution curve of the annual totals for data generated using WGEN Method for Hebron station 59
Figure 4.2-8: Annual totals of data generated using WGEN Method for Hebron station 60
Figure 4.2-9: Rainfall distribution curve of the daily maximum values for data generated using WGEN Method for Welgevonden station 60
Figure 4.2-10: Percentiles of daily data generated using WGEN Method for Welgevonden station 61
Figure 4.2-11: Averages of monthly data generated using WGEN Method for Welgevonden station 62
Figure 4.2-12: Maximum values of monthly data generated using WGEN Method for Welgevonden station 62

Figure 4.2-13: Standard deviation values of monthly data generated using WGEN Method for Welgevonden station 63
Figure 4.2-14: Percentiles of monthly data generated using WGEN Method for Welgevonden station ... 63
Figure 4.2-15: Annual totals of data generated using WGEN Method for Welgevonden station 64
Figure 4.2-16: Annual percentiles of data generated using WGEN Method for Welgevonden station 64
Figure 4.2-17: Standard deviation values of daily data generated using WGEN Method for Nelshoogte station 65
Figure 4.2-18: Percentiles of daily data generated using WGEN Method for Nelshoogte station 65
Figure 4.2-19: Averages of monthly data generated using WGEN Method for Nelshoogte station 66
Figure 4.2-20: Maximum values of monthly data generated using WGEN Method for Nelshoogte station 66
Figure 4.2-21: Standard deviation of monthly data generated using WGEN Method for Nelshoogte station 67
Figure 4.2-22: Percentiles of monthly data generated using WGEN Method for Nelshoogte station 67
Figure 4.2-23: Rainfall distribution curve of the annual totals for data generated using WGEN Method for Nelshoogte station 68
Figure 4.2-24: Annual percentiles of data generated using WGEN Method for Nelshoogte station 68
Figure 4.3-1: Rainfall distribution curve of the daily maximum values for data generated using VLB Method for Hebron station 69
Figure 4.3-2: Percentiles of daily data generated using VLB Method for Hebron station 69
Figure 4.3-3: Averages of monthly data generated using VLB Method for Hebron station 70
Figure 4.3-4: Maximum values of monthly data generated using VLB Method for Hebron station 70
Figure 4.3-5: Standard deviation values of monthly data generated using VLB Method for Hebron station 71
Figure 4.3-6: Percentiles of monthly data generated using VLB Method for Hebron station 71
Figure 4.3-7: Rainfall distribution curve of the annual totals for data generated using VLB Method for Hebron station 72
Figure 4.3-8: Annual percentiles of data generated using VLB Method for Hebron station 72
Figure 4.3-9: Rainfall distribution curve of the daily maximum values for data generated using VLB Method for Welgevonden station 73
Figure 4.3-10: Extreme percentiles of daily data generated using VLB Method for Welgevonden station 73
Figure 4.3-11: Averages of monthly data generated using VLB Method for Welgevonden station 74
Figure 4.3-12: Maximum values of monthly data generated using VLB Method for Welgevonden station 75
Figure 4.3-13: Standard deviation of monthly data generated using VLB Method for Welgevonden station 75
Figure 4.3-14: Percentiles of monthly data generated using VLB Method for Welgevonden station 76
Figure 4.3-15: Rainfall distribution curve of the annual totals for data generated using VLB Method for Welgevonden station 76
Figure 4.3-16: Annual percentiles of data generated using VLB Method for Welgevonden station 77
Figure 4.3-17: Rainfall distribution curve of the daily maximum values for data generated using VLB Method for Nelshoogte station 77
Figure 4.3-18: Percentiles of daily data generated using VLB Method for Nelshoogte station 78

Figure 4.3-19: Averages of monthly data generated using VLB Method for Nelshoogte station 79
Figure 4.3-20: Maximum values of monthly data generated using VLB Method for Nelshoogte station.. 79
Figure 4.3-21: Standard deviation values of monthly data generated using VLB Method for Nelshoogte station using VLB method 80
Figure 4.3-22: Percentiles of monthly data generated using VLB Method for Nelshoogte station 80
Figure 4.3-23: Rainfall distribution curve of the annual totals for data generated using VLB Method for Nelshoogte station 81
Figure 4.3-24: Annual percentiles of data generated using VLB Method for Nelshoogte station 81
Figure 5.1-1: Comparison of the daily maximum values generated from the methods for the Hebron station 83
Figure 5.1-2: Comparison of the daily average values generated from the methods for the Hebron station 83
Figure 5.1-3: Comparison of the daily skewness values generated from the methods for the Hebron station 84
Figure 5.1-4: Comparison of the daily standard deviation values generated from the methods for the Hebron station 84
Figure 5.1-5: Comparison of the monthly averages from the methods for the Hebron station 85
Figure 5.1-6: Comparison of the monthly maximum values for the methods for the Hebron station 85
Figure 5.1-7: Comparison of the monthly standard deviation for the methods for the Hebron station ... 85
Figure 5.1-8: Comparison of the annual minimum values generated from the methods for the Hebron station 86
Figure 5.1-9: Comparison of the annual average values generated from the methods for the Hebron station 86
Figure 5.1-10: Comparison of the annual maximum values generated from the methods for the Hebron station 87
Figure 5.1-11: Comparison of the annual standard deviation values generated from the methods for the Hebron station 87
Figure 5.2-1: Comparison of the daily maximum generated from the methods for Welgevonden station 91
Figure 5.2-2: Comparison of the daily average generated from the methods for Welgevonden station.. 92
Figure 5.2-3: Comparison of the daily skewness values generated from the methods for Welgevonden station 92
Figure 5.2-4: Comparison of the daily standard deviation values generated from the methods for Welgevonden station 93
Figure 5.2-5:Comparison of the monthly averages from the methods for the Welgevonden station..... 93
Figure 5.2-6: Comparison of the monthly maximum values from the methods for the Welgevonden station 93
Figure 5.2-7: Comparison of the monthly standard deviation from the methods for the Welgevonden station 94
Figure 5.2-8: Comparison for the annual minimum values generated from the methods for the Welgevonden station 94
Figure 5.2-9: Comparison for the annual maximum values generated from the methods for the Welgevonden station 95

Figure 5.2-10: Comparison for the annual average values generated from the methods for the Welgevonden station 95
Figure 5.2-11: Comparison for the annual standard deviation values generated from the methods for the Welgevonden station 96
Figure 5.3-1: Comparison for the daily maximum values generated from the methods for Nelshoogte station 99
Figure 5.3-2: Comparison for the daily average values generated from the methods for Nelshoogte station 99
Figure 5.3-3: Comparison for the daily skewness values generated from the methods for Nelshoogte station 100
Figure 5.3-4: Comparison for the daily standard deviation values generated from the methods for Nelshoogte station 100
Figure 5.3-5: Comparison of the monthly averages for Nelshoogte station using VLB Station 101
Figure 5.3-6: Comparison of the monthly maximum for the Nelshoogte station from the methods 101
Figure 5.3-7: Comparison of the monthly standard deviation for the Nelshoogte station from the methods 101
Figure 5.3-8: Comparison for the annual minimum values generated from the methods for the Nelshoogte station 102
Figure 5.3-9: Comparison for the annual maximum values generated from the methods for the Nelshoogte station 102
Figure 5.3-10: Comparison for the annual average values generated from the methods for the Nelshoogte station 103
Figure 5.3-11: Comparison for the annual standard deviation values generated from the methods for the Nelshoogte station 103

List of Tables

Table 2-1: Comparative review of the methods 13
Table 3-1: Characteristics of the rainfall stations 17
Table 3-2:Example of Basis of states 26
Table 3-3: Maximum state for each month 26
Table 3-4: Example of Rainfall groups of states for the months 27
Table 3-5: Example of probabilities matrix generated for each month 28
Table 5.1-1: Average statistics of the daily data generated using TPM, WGEN and VLB for the Hebron rainfall station 88
Table 5.1-2: Average statistics of the monthly data generated using TPM, WGEN and VLB for the Hebron rainfall station 89
Table 5.1-3: Average statistics of the annual data generated using TPM, WGEN and VLB for the Hebron rainfall station 90
Table 5.2-1: Average statistics of the daily data generated using TPM, WGEN and VLB for the Welgevonden rainfall station 96
Table 5.2-2: Average statistics of the monthly data generated using TPM, WGEN and VLB for the Welgevonden rainfall station 97
Table 5.2-3: Average statistics of the annual data generated using TPM, WGEN and VLB for the Welgevonden rainfall station 98
Table 5.3-1: Average statistics of the daily data generated using TPM, WGEN, and VLB for the Nelshoogte rainfall station 104
Table 5.3-2: Average statistics of the monthly data generated using TPM, WGEN, and VLB for the Nelshoogte rainfall station 105
Table 5.3-3: Average statistics of the annual data generated using TPM, WGEN and VLB for the Nelshoogte rainfall station 105

List of Abbreviations

ARIs	Average Recurrence Intervals
ARMAs	Autoregressive Moving Average models
DMMS	Daily and Monthly Mixed Models
IQR	Interquartile Range
MC	Markov Chain
MAP	Mean Annual Precipitation
PCD	Pollution Control Dam
TPM	Transitions Probability Matrix
TGD	Truncated Geometric Distribution
TNBD	Truncated Negative Binomial Distribution
SAWS	South African Weather Services
SRGs	Stochastic Rainfall Generators
STOMSA	Stochastic Model of South Africa
VLB	Variable Length Bootstrap
WGEN	Weather Generators

1. INTRODUCTION

1.1. Background

Rainfall is one of the main factors that influence hydrologic processes and there have been many studies aimed at predicting future rainfall patterns of various areas in the world (Chapman 1994; Boughton, 1999; Zhao et al. ,2019). Extremely varying rainfall patterns can have a major impact on daily human activities including businesses, mining, and agriculture. Due to climate change, rainfall variability has tended to increase, and it has become even more imperative to reliably predict future rainfall patterns for use in hydrological assessment and design. Water, especially freshwater, is a finite resource and in South Africa as well as globally, there are regions where unsustainable levels of water use have been reached with demand for water exceeding supply.

Stochastic Weather Generators (WGs) have been widely used in many applications including decision support systems for risk-based planning and management of natural resources such as hydrological and groundwater plume seepage simulation models as well as crop studies, electrical power systems and in mining water systems (De Vera & Terra, 2018).

In South Africa, the mining industry is one of the large water users after Irrigation and urban use. What is of concern as well is that mining influences the quality of the water and can also drastically impact the hydrological and topographical conditions of catchments (Punkkinen et al.,2016). Water management is a continuous process that needs to be developed and updated throughout the whole mine life cycle and changes in mine operations require updates to mine water management. Typical mining requires substantial amounts of water for their operations and effective water balance modelling can help improve mining water utilization efficiency, planning and design of infrastructure such as pollution control dams (PCDs) and sumps.

Effective mine water management and design of mine water infrastructure require reliable input of daily rainfall which is typically obtained by stochastic modelling of daily rainfall (Van Niekerk, 1997). Hence it is imperative that the methods used to generate the rainfall stochastics are themselves reliable and as best they can, mimic the rainfall characteristics of the mining area to enhance confidence in the results from the modelling.

Since many hydrological processes such as infiltration and runoff vary considerably within daily (or shorter) time steps, there has been a focus on the realistic stochastic daily rainfall generation (Corporate Research Centre Australia, 2003, De Vera & Terra,2018 and Wilcox, et al., 2020).

Although considerable stochastics rainfall generation research has been conducted globally, little of this has been developed using observed rainfall data in African region (Phakula et al., 2018, Pegram and Clothier, 2002).

Some of research conducted in South Africa include SERGE tool (Eisinger & Wiegand, 2008). SERGE is a stochastic, spatially explicit tool for the simulation of daily rainfall in arid environments. SERGE is a parametric daily rainfall generator that generates rainfall using several randomly positioned rain clouds. It is based on the Zucchini rainfall model (Zucchini et al.,1992) which is intended for investigating the effect of spatial rainfall variability. The Zucchini model comprises 16 parameters and has been calibrated in terms of daily rainfall records throughout South Africa. The model therefore has a reliable representation of temporal rainfall (Eisinger & Wiegand, 2008). SERGE was designed to generate spatially varying rainfall estimates of different spatial correlation, while preserving the temporal rainfall distribution of the Zucchini model in any single cell. In SERGE, spatial variability is introduced sparingly by invoking clouds under the simplifying assumption of constant cloud size on a given day, with the cloud being the area which will receive rain in one of several events each rainy day (Eisinger & Wiegand, 2008).

The planning and system operations of most of the large water resource systems in South Africa have been applying a multi-site monthly streamflow generator since the 1990s (Pegram and McKenzie, 1991; Basson et al., 1994; Van Rooyen and McKenzie, 2004). However, it has recently been recognized that the use of stochastic rainfall generation may hold several advantages over

the stochastic streamflow generation (Ndiritu & Nyaga, 2014). With stochastic rainfall, probabilistic analysis and climate change impacts can be included more realistically and easily in the analysis of catchment hydrological processes and rainfall-dependent activities such as irrigation.

1.2. Problem Statement

Given that few daily stochastic rainfall generation assessments have been carried out in Southern Africa, the focus of this study is therefore to evaluate selected methods for stochastic daily rainfall generation using Southern African historic rainfall data. Southern Africa is characterized by significant rainfall variability on a range of temporal and spatial scales which have an impact on agriculture, water resources and the economy at large. The Southern African region is also vulnerable to extreme climate events (Shongwe et al, 2009). The devastating floods in Mozambique during February/March 2000 and severe droughts of 1991/1992, 2002/2003 and 2003/2004 in South Africa are examples of such extreme events (Cook et al., 2004). In 2015, Cape Town was facing a water shortage that has built up over the past two years (Muller M. , 2017). Concerns were raised when it was realized that dam levels were uncomfortably low after winter rainfall in 2014/15. The combined effect of this prolonged dry period with an ever-growing demand for water culminated in the widely publicized 'Day Zero' water crisis (Sousa et al,2018). It is therefore important that the methods used for generation of stochastic daily rainfall are well understood and reliable.

1.3. Objectives

The research aims to evaluate and compare selected methods of stochastic daily rainfall generation using Southern African historic rainfall data. The evaluation is based on the ability of the methods to preserve the daily, monthly, and annual statistics of the observed data. The assessment is also based on simplicity and ease of application of the methods. The specific objectives of the study are:

(i) To evaluate the stochastic daily rainfall generation performance of the Transition Probability Matrix, the Weather Generator and the Variable-Length Bootstrap

(ii) To compare the performance of the three methods above with each other.

1.4. Organization of the Research Report

This research is presented as follows:

Chapter 2 reviews the available literature on stochastic hydrological models and how the assessed methods were developed. The chapter also highlights significant milestones in the development of stochastic modelling.

Chapter 3 presents the methodology followed for the research which includes the approach and procedures for generating the data using each method.

Chapter 4 presents the performance evaluation of each method and Chapter 5 compares the methods with each other. The performance of each method was firstly presented individually to allow for comprehensive evaluation of performance before comparison. Chapter 6 provides conclusion from the discussions presented in Chapter 4 and 5 and recommendations.

2. LITERATURE REVIEW

Rainfall is the driver for many hydrological processes and lack of reliable rainfall data limits development of hydrologic research and water resources assessments (Zhao et al, 2019). Research on stochastic rainfall generation has been conducted for decades. Haan et al. (1976) studied a stochastic model based on a first-order Markov chain that was developed to simulate daily rainfall at a point. Stochastic rainfall generators (SRGs) aim to simulate realistic rainfall series by reproducing key statistical features of rainfall variability (Wilks and Wilby,1999). As SRGs can generate long-term rainfall sequences at fine resolutions for a given climatology, they are useful in many applications, including risk assessment studies to estimate the return periods of rare events (Wilcox, et al., 2020).

Data generated from SRGs is of importance to planners of water resource or agricultural systems because when used in system simulation, it allows planners to evaluate proposed system designs in a more robust and statistically realistic manner than is possible with historic records alone (Srikanthan and Zhou, 2003, Zhao et al, 2019).

2.1. Review of Stochastic Hydrological Models

Rainfall modelling cuts across the boundaries of hydrology, and incorporates atmospheric physics, meteorology, climatology, and statistics among others. One aspect of rainfall with which hydrologists have been most concerned with is modelling its statistical structure (Waymire & Gupta, 1981). Daily rainfall data generation models can be broadly classified into four main groups; namely two-part models, transition probability matrix models, resampling models, and time series models of the Autoregressive Moving Average model (ARMA) type (Srikanthan and McMahon, 2001). Dam Safety Office (2003) defined four classes for stochastic models as; one-dimensional point process models; multi-station models; space-time models and disaggregation models. Markov chain models and non-parametric models are examples of multi-station models.

Markov chain model can be described as a first order model and second order models' examples are ARMAs (Autoregressive Moving Average models). Stochastic hydrology has been dominated by parametric and non-parametric data generation methods. Ease of applicability influences the choice of the models to apply (Nyaga, 2014). Two-part models include a model for the occurrence of dry and wet days and a model for the generation of rainfall amount on wet days.

Chapman (1994) compared five models, namely, Markov chains of orders 1, 2 and 3 (MC1, MC2 and MC3), truncated negative binomial distribution (TNBD) and the truncated geometric distribution (TGD) with separate parameter values for each month using data from 17 Australian rainfall stations. The study found that different record lengths appear to affect the performance of the models, particularly when wet and dry spells are considered separately. Chapman (1994) concluded that the prospects for regionalisation of parameters are poor unless there is a good sample of long records.

Nathan and Wang (2007) evaluated a method for coupling two different time scales of stochastic hydrological time series models. The key feature of this approach is to first generate two resembling time series: one preserving key statistical properties at a finer time scale and the other at a coarser time scale. Adjustment is then made to the finer time scale series to make it consistent with the coarser time scale series.

Skrikanthan and McMahon (1985) reviewed daily rainfall generating models, and they described the various approaches which had been reported in the literature. They tested Daily and Monthly Mixed (DMMS) models, transition probability matrix models and time series models with daily rainfall data from five stations (Melbourne, Mackay, Perth, Alice Springs, and Darwin) and recommended the Transition Probability Matrix model for generating daily rainfalls in Australia.

Pegram and Clothier (2002) developed the String of Beads Model for stochastic rainfall generation in South Africa. The String of Beads model is a stochastic rainfall model based on the combined observations of a large network of daily rain gauges and an S-band weather radar situated near Bethlehem, South Africa.

The model is complex and is designed to perform two important modelling tasks namely, to simulate long sequences to conduct different possible studies of rainfall over the target area, and secondly, to adapt the model to give short-term forecasts of future rainfall for use in real-time flood forecasting in either rural or urban environments.

Synthetic streamflow generation methods are mostly classified as either parametric or nonparametric (Ndiritu & Nyaga, 2014). Parametric methods include a mathematical expression and an explicitly specified probability distribution that is fitted to the historical data. Nonparametric methods apply resampling techniques, thereby avoiding the need for fitting probability distributions.

Parametric stochastic methods use large numbers of parameters, and the methods applied to preserve cross-correlations and to disaggregate annual flows are complex. Parametric stochastic methods have been widely applied in some regions such as STOMSA (Stochastic Model of South Africa) in South Africa (Ndiritu, 2011). STOMSA was developed in the early 1980s for the then Department of Water Affairs and Forestry (now the Department of Water and Sanitation) as part of the Vaal River System Analysis Study (McKenzie et al., 2004; Pegram, 1986). The primary objective of the model was to serve as a generic streamflow generator that is technically sound and versatile in nature so that it can be incorporated with the models used to analyse the water resource systems of South Africa (Maass, 2017).

STOMSA is applied both in the long-term planning and in the design and short-term operation of complex reservoir systems (Pegram, 1986; Pegram & McKenzie, 1991; Basson et al., 1994; Van Rooyen & McKenzie, 2004). Due to the complexity of STOMSA, some water resources planners prefer to use methods that are easier and more easily understood (Ndiritu, 2011).

A nonparametric model presented in the research by Harrold et al. (2002) and compared to the TPM model by Srikanthan et al. (2005) consists of two parts: a rainfall occurrence model and a rainfall amounts model.

The model reproduces the seasonality of the historical record and the observed characteristics of the historical record at several time scales. Seasonality is modelled using the moving window approach (Rajagopalan et al. 1996; Sharma and Lall, 1999).

It is a data resampling strategy that approximates the random mechanism that produced the historical data. In this methodology, a pattern is chosen from the historical sequence that is similar to the current pattern in the generated sequence, and the successor to the chosen pattern is placed into the generated sequence. The patterns that are examined are those that exist in the predictor set used in the model.

Another feature of the model by Harrold et al. (2002) that was different from the TPM model was the representation of seasonality in the rainfall simulations. A 31-day moving window was used to ensure a smooth transition of generated rainfall characteristics throughout the year, whereas the use of a monthly discretisation in formulating the TPM resulted in an abrupt transition across month boundaries.

The advantage of the non-paramedic model is that it is relatively direct in that it is defined by a very few parameters and achieves remarkable realism in the synthetic sequences it generates. The result is that it is remarkably fast in execution, which is a necessary characteristic with all the detail it presents.

It has been found that there have been few daily stochastic rainfall generation studies using Southern Africa rainfall data and these studies have not focused on assessing and comparing the performance of SRGs. This study aims to fill this gap by assessing and comparing three SRGs.

The SRGs are selected based on ease of applicability, differences in the concepts applied to develop the models and modelling performance reported in the literature. VLB model has been reported to perform well as a monthly stochastic generator (Ndiritu and Nyaga, 2014) and the innovative resampling applied in the method is easily adaptable for daily rainfall generation. The WGEN method is widely used in other regions of the world but has rarely been used in Southern Africa and including it in this study will provide insight of its performance with Southern Africa

climate data. TPM method is a more frequently used method applied for water management in the mining industry. The following is a review of the three selected the modelling methods: WGEN, TPM and VLB.

2.2. WGEN (Weather Generator) Method

Richard and Wright (1984) developed a computer simulation model called WGEN (Weather Generator) that can be used to provide daily values for precipitation, maximum temperature, minimum temperature, and solar radiation in Fortran at the US Department of Agriculture Agricultural Research Service. The model was developed to account for the persistence of each variable, the dependence among the variables, and the seasonal characteristics of each variable. The precipitation component of WGEN is a Markov chain-gamma model, a first order Markov chain is used to generate the occurrence of wet or dry days. When a wet day is generated, the two-parameter gamma distribution is used to generate the precipitation.

Two major generation options are available in WGEN. The user may choose to generate daily values of all four variables or use actual precipitation data and generate the other three variables. In addition to the two major options, the user may choose to apply correction factors to precipitation and/or temperature based on actual mean monthly values.

With the first-order Markov chain model, the probability of rain on a given day is conditioned on the wet or dry status of the previous day. The model was used to generate the daily precipitation that compared well to the observed data. The parameters were defined using 20 years (1951-1970) of daily rainfall data from 139 stations in the USA that were used to define the parameters required. The precipitation amounts and the season distribution of precipitation were accurately represented in the generated data. This method will be evaluated using the South African data. Due to WGEN requiring a long time series of daily weather data to estimate parameters, this limits its use to regions of the world where sufficient data is available.

2.3. Transition Probability Matrix (TPM) Method

The Transition Probability Matrix (TPM) model used to generate synthetic sequences of daily rainfalls is a modification of a method described by Haan et al (1976) and modified for use in Australia by Srikanthan and McMahon (1985). The model is referred to as a transition probability matrix (TPM) or a multi-state Markov chain model. Seasonality of rain is modelled by using 12 TPMs, one for each calendar month.

Although the averages of daily, monthly, and annual rainfalls are reproduced adequately, the simple form of TPM model substantially underestimates the variance of the annual totals, and hence substantially underestimates the severity of severe droughts (Boughton, 1999). McMahon & Skrikanthan (2001) also highlighted that TPM is not suitable for regionalisation and with limited length of data. A markov model incorporating the Wilson-Hilfery transformation was recommended to generate annual rainfall data (Srikanthan and McMahon, 2001).

McMahon & Skrikanthan (2001) reviewed the state of research and practice in the stochastic generation of climate data and recommended the Transition Probability Matrix approach.

Boughton and Hill (1997) brought the techniques together into a composite design flood estimation procedure, particularly for floods with Average Recurrence Intervals (ARIs) of greater than 100 years. Boughton and Hill (1997) introduced a new flood estimation technique which makes use of the available daily rainfall data with a rainfall-runoff model for studies of water yield. The data used for testing the flood estimation procedure were from the 108 km^2 Boggy Creek catchment at Angleside, station number 403226) in Australia. Hydrometric data was provided by the Bureau of Meteorology and the (former) Rural Water Corporation. There were 17 years (1976 to 1992 inclusive) of concurrent daily rainfall and runoff. The probabilities for rain in one state to be followed by rain on the next day in the same or another state are collated into a matrix - the transition probability matrix. Seasonality of rain is modelled by using 12 TPMs, one for each calendar month.

Boughton and Hill (1997) indicated that the frequency distribution of annual totals of generated rainfall had a variance that was substantially less than the variance of annual totals of recorded rainfall. In all tests the variance of the generated annual totals was substantially less than the actual variance. Boughton (1999) describes the transitions probability model which achieves the observed variance of annual totals by using an adjustment factor (F). The difference between the generated annual total (GAT) and the mean of recorded annual totals is increased by an adjustment factor (F). The adjustment factor (F) is obtained as a ratio of the standard deviation of the generated and observed annual rainfall (Boughton, 1999).

This gives an Adjusted Annual Total (AAT) (Boughton, 1999) that is more representative of the historic data. It should be noted that while this adjustment improves the representation of rainfall characteristics at the annual scale, it comes at the cost of inflating the variability at the daily timescale (Srikanthan et al. 2005).

Srikanthan et al. (2002) evaluated the Transition Probability Matrix (TPM) model with Boughton's correction for interannual variability and the simplified Daily and Monthly Mixed (DMMS) model for the generation of daily rainfall data. The study also compared the statistical characteristics of the daily, monthly, and annual streamflow data simulated by a rainfall-runoff model using stochastic daily rainfall obtained using the TPM and DMMS models with the historical streamflow characteristics (Siriwardena, Srikanthan, & McMahon, 2002). It was shown that both models preserved key statistical characteristics of historical rainfall satisfactorily at daily, monthly, and annual levels. However, the DMMS model poorly simulates the rainfall amounts on different types of wet days (solitary wet days, wet days bounded on one side by a wet day and wet days bounded on both sides by wet days) whereas the TPM model preserves these characteristics adequately.

The major drawback of the TPM model over the DMMS model is its inability to preserve correlation between monthly rainfalls; the DMMS model preserves this statistic satisfactorily (Siriwardena, Srikanthan, & McMahon, 2002).

2.4. Variable Length Bootstrap (VLB) Method

Ndiritu (2011) developed a resampling monthly streamflow generator (Variable Length Bootstrap (VLB) method that was later applied for monthly rainfall generation (Ndiritu & Nyaga, 2014). The method applies variable-length bootstrapping of annual flows and a weighted method of fragments for disaggregation and perturbing the bootstrapped annual flows.

Typical resampling methods are unable to generate new data and one of the innovations of the VLB model was an approach that enables synthetic flows other than those found in the historical record to be obtained. An assessment of this is done by comparing the upper and lower bounds of annual flows obtained from VLB. A comparison of the VLB and STOMSA in the study showed that the two methods largely reproduce the mean monthly flows and standard deviations adequately. However, there is a tendency for STOMSA to underestimate the standard deviation.

A 68-year-long monthly data set consisting of five reservoir sites in southern Africa (Van Rooyen & McKenzie, 2004) was used to help develop and test the VLB method.

The main steps of the VLB generator include:

- Generation of variable length blocks of annual time series from the historic time series
- Random sampling of the blocks with replacement to create an annual stochastic time series of the specified length.
- The matching of each of the stochastic time series years with a pair of different years of the historic time series based on the magnitude of the annual flows of the current and the previous year.
- The disaggregation of the stochastic annual values using the monthly distributions of the pair of matching historic years and an appropriate incorporation of perturbations.
- The updating of the stochastic annual values after the disaggregation

The method obtains the initial synthetic sequence of annual flows, the blocks are selected randomly with replacement. Since block creation ensures that the blocks (except the first one) start in a wet rather than a dry period, a procedure is implemented to remove any bias of creating generated sequences that predominantly begin with wet periods (Ndiritu, 2011).

The results showed that there is an inability of STOMSA to model the monthly serial correlation between the end of one year and the next is clear as the 12th value for STOMSA is very low (0.035). However, the VLB method models this correlation reasonably well, capturing 80% of the historical correlation. STOMSA underestimated the annual serial and the annual cross-correlation coefficients, while the VLB reproduced them both adequately.

Ndiritu & Nyaga (2014) developed and tested a monthly nonparametric stochastic rainfall generator that would comprehensively incorporate climate change and changing variability including information from global climate model (GCM) projections. The study also aimed at comparing the non-parametric generator with the parametric PEGRAIM-W generator and it was found that both PEGRAIM-W and VLB are considered suitable for annual and monthly stochastic rainfall generation. VLB however performed better for both annual and monthly generation.

Nyaga (2014) conducted a study to find out how effectively mode decomposition (EMD) can be used as an alternative block termination method to the method used in the VLB rainfall generator and named this generator as the EMD-VLB. Nyaga (2014) found that both generators generally replicate the historic minimum run sums reasonably well. The VLB however locates more of the historic run sums within the inter-quartile range than the EMD-VLB. The EMD-VLB generator has 70.8 % while the VLB has 95% of the historic run sums falling within the inter-quartile range of box plots of the simulated run sums. The variable length bootstrap (VLB) stochastic generator, previously used in streamflow generation was selected and adapted to rainfall generation for a normal, drier, wetter or a more variable climate and was found to provide realistic climate change-impacted stochastic rainfalls that could be matched to the average changes in mean annual rainfall projected by global climate models.

For the current study, the monthly stochastic rainfall generator by Ndiritu &Nyaga (2014) is updated to enable it to generate stochastic daily rainfall.

2.5. Comparative Review of the methods

Based on the literature review presented in the sections above, Table 2-1 compares some of the characteristics of the methods and highlights key elements from the methods. TPM and WGEN are both Markov chain models and VLB method is a non-parametric model which uses resampling techniques instead of fitting to probability distributions.

Table 2-1: Comparative review of the methods

Characteristics of the method	**TPM Method**	**WGEN Method**	**VLB Method**
Type of Model	Multi-State Markov chain model	First order Markov Chain-Gamma Model	Non-parametric Model (Bootstrap)
Limitations of Applicability	No Known Limitations	No Known Limitations	No Known Limitations
Data commonly generated using the model	Precipitation	Precipitation, Temperature and Radiation	Precipitation, Streamflow, Temperature, Radiation, Groundwater levels

3. METHODOLOGY

3.1. Introduction

This chapter outlines the research methodology applied. It describes the process followed in verifying the reliability of the sourced historic rainfall data and the procedures used to generate stochastic daily rainfall data from each method. The chapter also outlines the approach that has been used in evaluating the performance of each method and the statistics used to compare the generated daily rainfall data with the historic rainfall data. In addition, the statistical evaluation of the historic daily rainfall data is presented in this chapter as it forms the basis of assessment for each method.

3.2. Research Approach

Based on the research objectives stipulated in Section 1.3 , the following research approach illustrated in Figure 3-1 was followed. Long records of daily rainfall with at least 50 years of historic observed rainfall data were preferred to reliably assess statistical characteristics of the local rainfall. Statistical evaluation of the acquired historic rainfall data for three stations was carried out to assess the reliability of the observed data and the statistical characteristics for use in the modelling. The reliability of the historic observed daily rainfall data was assessed by calculating the percentage of gaps and errors in the sample size used and was deemed to be reliable if it consisted of less than 5% in gaps and errors. The analysis looked at the rainfall trends and patterns of the historic observed data such as cumulative totals, annual totals, probabilities of exceedance and box and whisker of monthly totals. The statistics that were compared with the historic observed data included the mean, median, standard deviation, mode, minimum, maximum and percentiles values. Once the observed data were deemed reliable for analysis, each method was applied to generate daily stochastic rainfall depths. 100 sequences of daily rainfall data were generated from each method for a period of 50 years. The statistics of the generated data were then compared with the historic rainfall data for each method using box and whisker plots that graphically illustrated the percentiles and variability of the statistics for the 100 sequences of data generated.

The performance of each method was firstly presented individually to allow for comprehensive evaluation before comparison with other methods. Conclusions on the suitability of all three methods as rainfall generators was then made. It should be noted that South Africa's summer rainfall season occurs during the warmer summer months, from October to March. Therefore, the monthly analysis was carried out as per the hydrological cycle from October to September.



Figure 3-1: Summary of main steps of the Methodology

3.3. Statistical analysis of sourced historic observed data

Available rainfall data for three rainfall stations in the vicinity of Mpumalanga Province, South Africa were provided by the South African Weather Stations (SAWS) for use in the research. The data set used was chosen subjectively but taking into consideration that is should be from Southern African region and be long and reliable. The decision to use historic rainfall from Mpumalanga was largely driven by working in a mine water management environment focusing on coal mines which are mainly in Mpumalanga region so the study would be beneficial to our on-going studies in the area, another concerning the other issue is that most of the stations

maintained by south Africa weather services (SAWS) have been discontinued making it difficult to get stations around South Africa with recent long data for use. The rainfall data was collated for analysis of rainfall trends and patterns. The rainfall data included several days whereby rainfall was not measured and the reliability percentage was calculated to be 99 % for the three stations. Table 3-1 summarizes the characteristics of the rainfall stations such as the record periods and the reliability of the observed data. The locations of the rainfall stations are shown in Figure 3-2, the rainfall stations occur within quaternary catchments B60D and X23E. The number of years of observed rainfall data ranges from 58 years to 86 years. Hebron and Welgevonden rainfall stations are approximately 9 km away from each other and Nelshoogte rainfall station is more than 100 km away from the other two stations. As Table 3-1 shows, this region experiences high rainfall that exceeds 1000 mm in Mean Annual Precipitation (MAP). The two stations close to one another were used to assess whether the methods can still replicate the same results for a station with similar characteristics that was close to another one as a way of validating the results obtained.

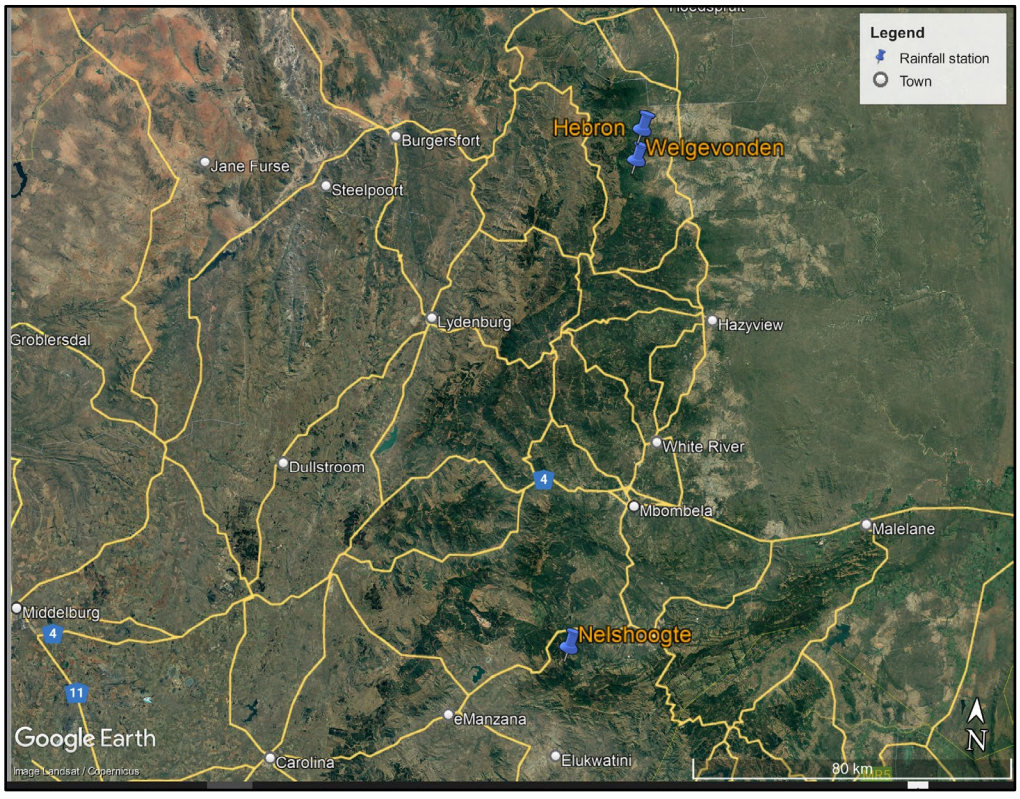


Figure 3-2: Locations of the rainfall stations (Google Earth, 2022)

Table 3-1: Characteristics of the rainfall stations

Station Name	Record of period	No of years	Coordinates (Decimals)	Mean Annual Precipitation (MAP)(mm)	Meters above mean sea level (mamsl)	Percentage of Reliable historic data from the data set
Hebron (0594760 W)	01/06/1949 to 22/12/2019	69	Latitude: -24.67 Longitude: 30.94	1547	968	99
Welgevonden (0594764 W)	02/08/ 1961 to 31/03/2020	58	Latitude: -24.74 Longitude: 30.93	1384	975	99
Nelshoogte (0518589 W)	01/02/1934 to 30/06/2021	86	Latitude: -25.83 Longitude:30.83	1072	1429	99

3.3.1. Hebron Station Rainfall Statistical Evaluation

Figure 3-3 shows the daily rainfall depths measured at Hebron rainfall station, maximum daily rainfall depth observed for the station is 350 mm measured on the 3rd of March 1993. The wettest year for the station was in 1999 with a total rainfall depth of 2485 mm and the driest year was in 1991 with total rainfall of 692.5 mm as shown in Figure 3-4 below. The wet season is between October to March, and the dry season spans between April to September. The wettest month is February as the monthly observed total can reach almost 800 mm/month as illustrated in Figure 3-5.

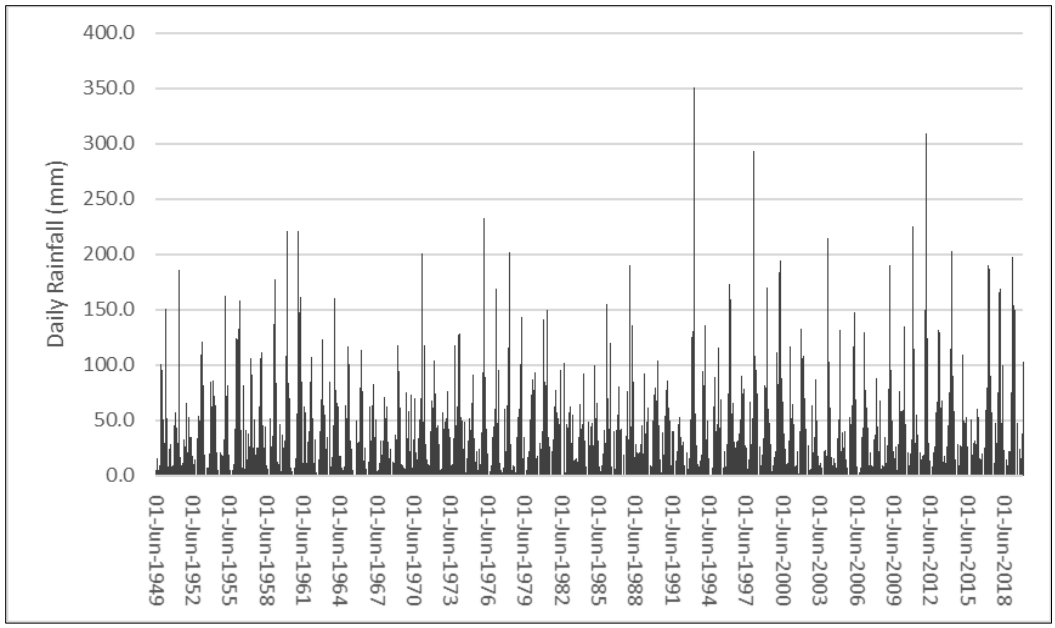


Figure 3-3: Daily rainfall depths measured at Hebron (0594760 W) rainfall station

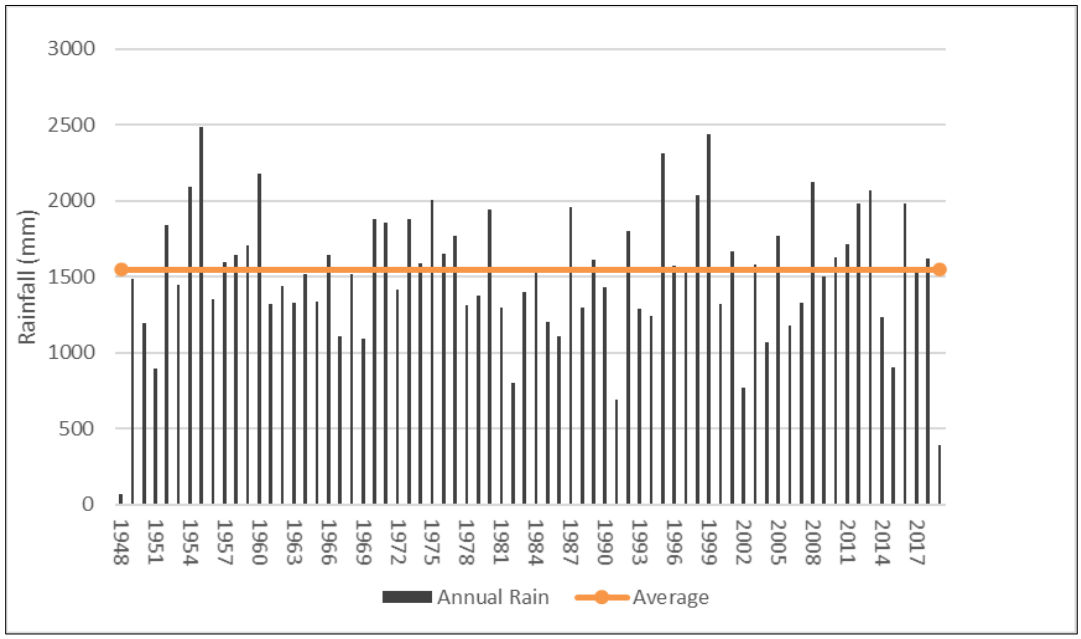


Figure 3-4: Annual rainfall depths measured at Hebron (0594760 W) rainfall station

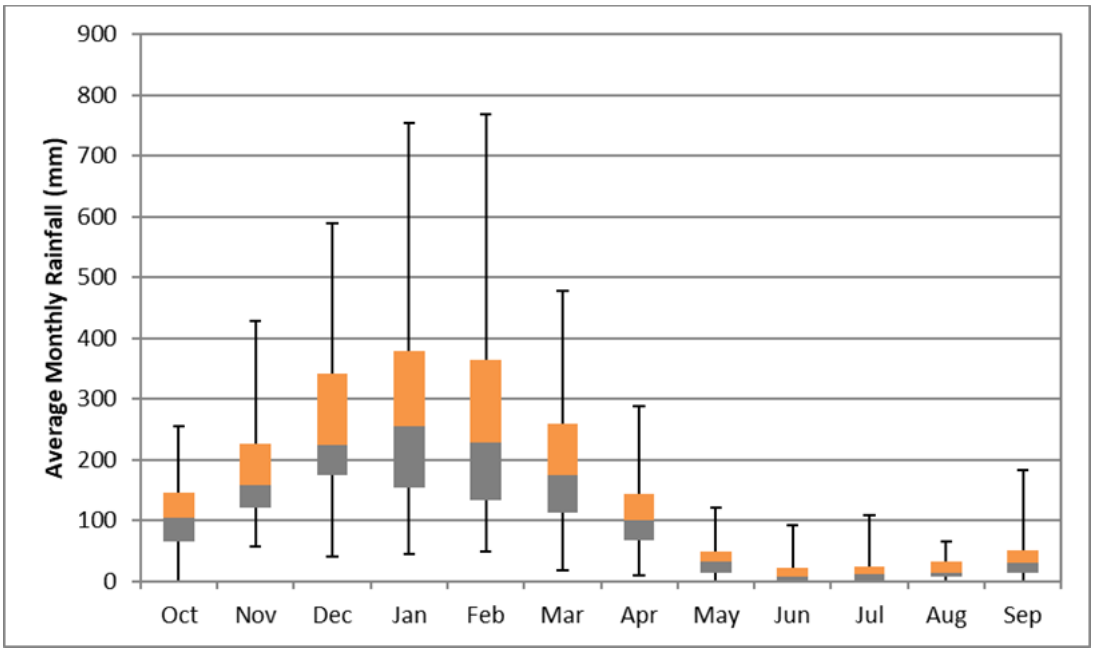


Figure 3-5: Monthly Box plot for the measured rainfall depths at Hebron (0594760 W) rainfall station

3.3.2. Welgevonden Station Statistical Evaluation

Welgevonden rainfall station is in the vicinity of Hebron rainfall station, being approximately 6 km away, and has a similar rainfall pattern.

Figure 3-6 shows the daily rainfall depths measured at Welgevonden rainfall station; the maximum daily rainfall depth observed for the station is 328 mm measured on the 1st of March 1993. The wettest year for the station was also in 1999 with a total rainfall depth of 2996 mm and the driest year was in 1991 with total rainfall of 644 mm as shown in Figure 3-7 below. The wettest month of the year is February with monthly total that can reach almost 700 mm/month as illustrated in Figure 3-8.

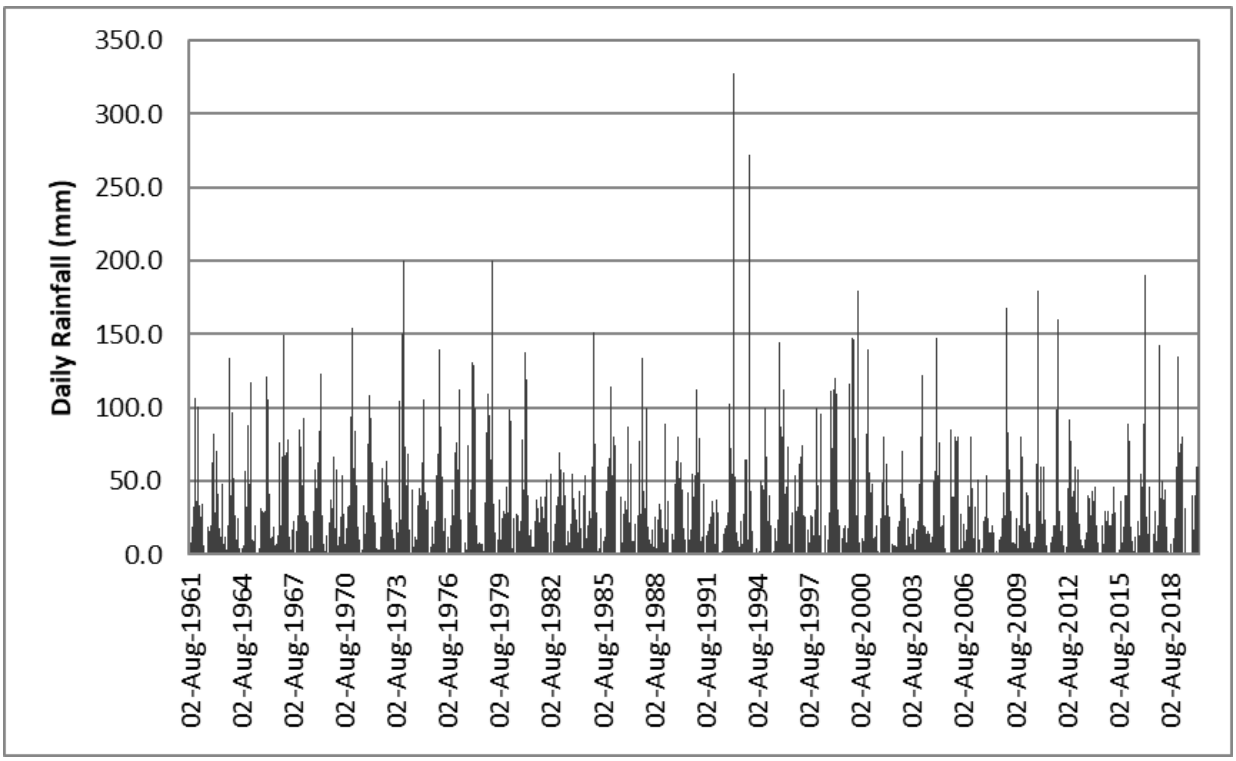


Figure 3-6: Daily rainfall depths measured at Welgevonden (0594764 W) rainfall station

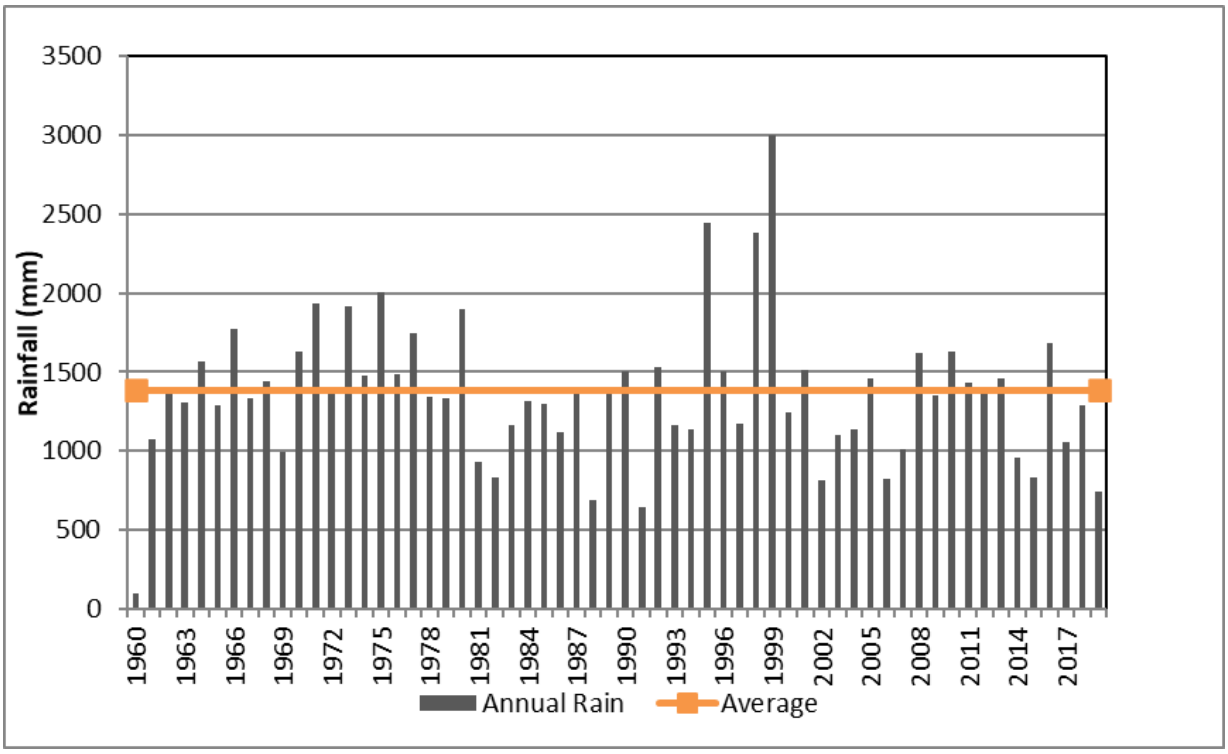


Figure 3-7: Annual rainfall depths measured at Welgevonden (0594764 W) rainfall station

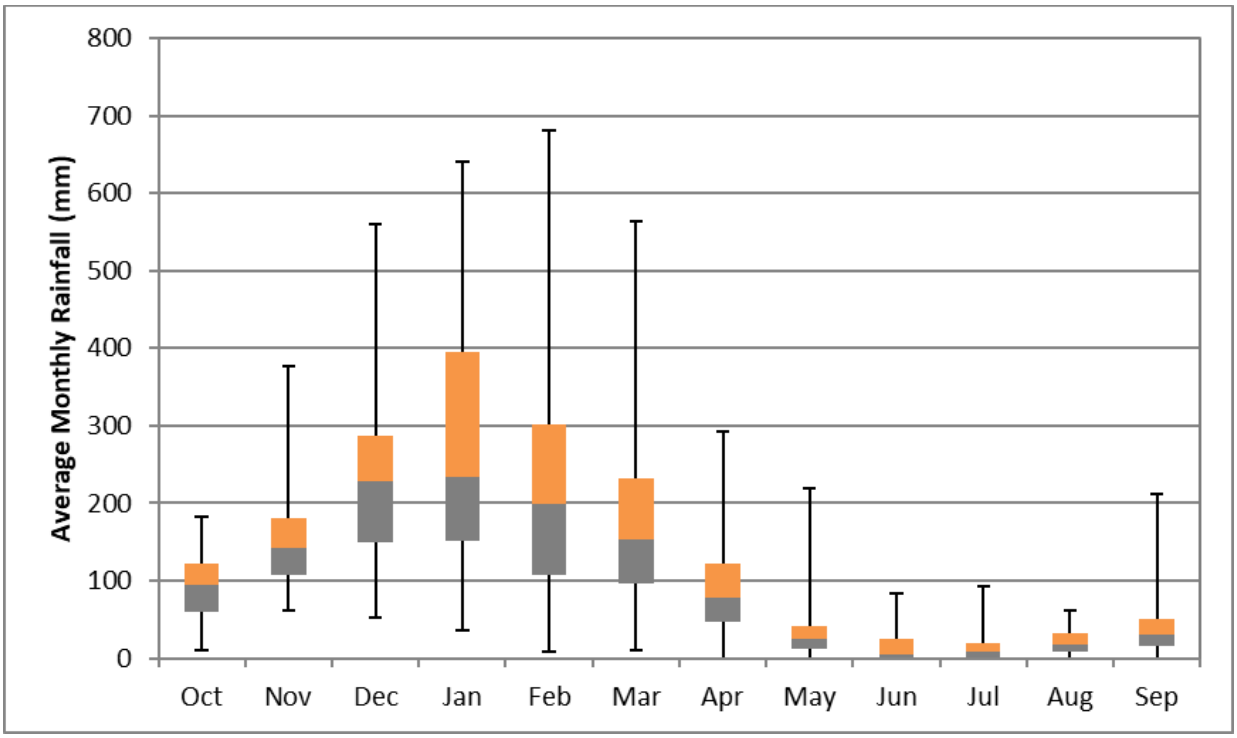


Figure 3-8: Monthly Box plot for the measured rainfall depths at Welgevonden (0594764 W) rainfall station

3.3.3. Nelshoogte Station Statistical Evaluation

Figure 3-9 shows the daily rainfall depths at Nelshoogte rainfall station. Compared to the two stations described above, Nelshoogte has relatively lower rainfall depths. Nelshoogte station has an observed maximum daily rainfall depth of 293 mm on the 5th of February 1939 over the observed period. This has caused the month of February to have a higher historic total compared to other months and can be considered an outlier. Figure 3-10 shows the annual rainfall depths observed for the station. The station has a maximum annual rainfall of 1755 mm and a minimum annual rainfall of 535 mm. Figure 3-11 illustrates that Nelshoogte station has a similar monthly distribution pattern to the other two stations. It can also be noted from Figure 3-11 that there were years in which January and February experienced no rainfall over the observed period.

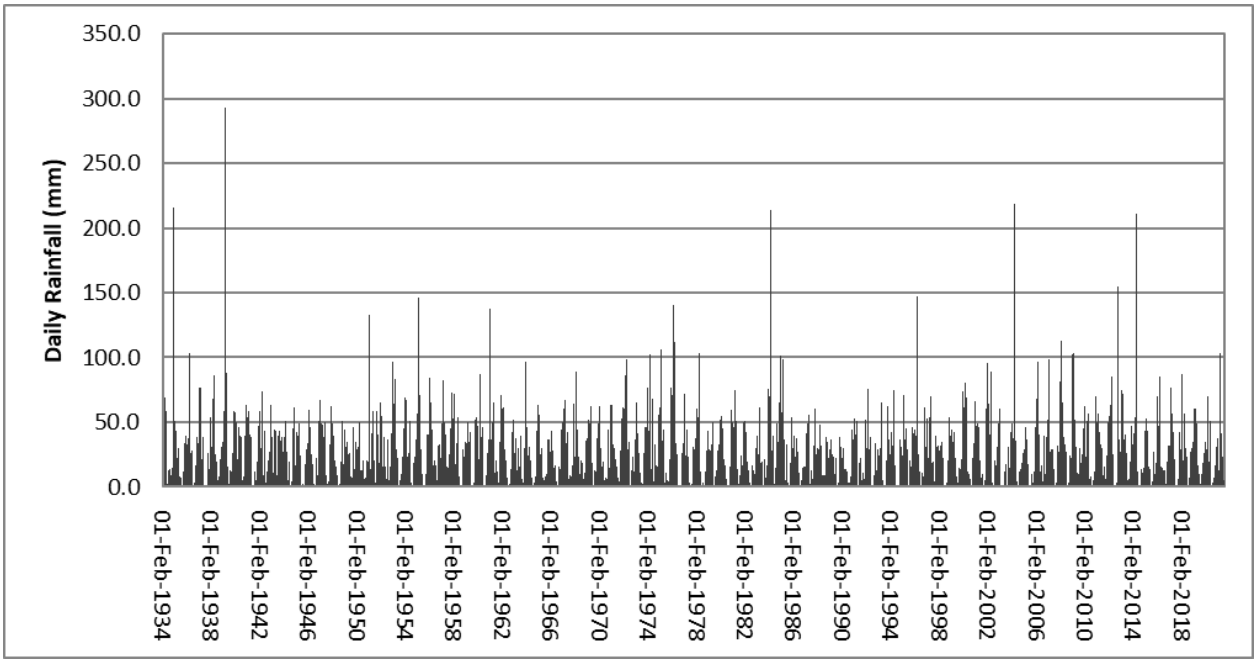


Figure 3-9: Daily rainfall depths measured at Nelshoogte (0518589) rainfall station

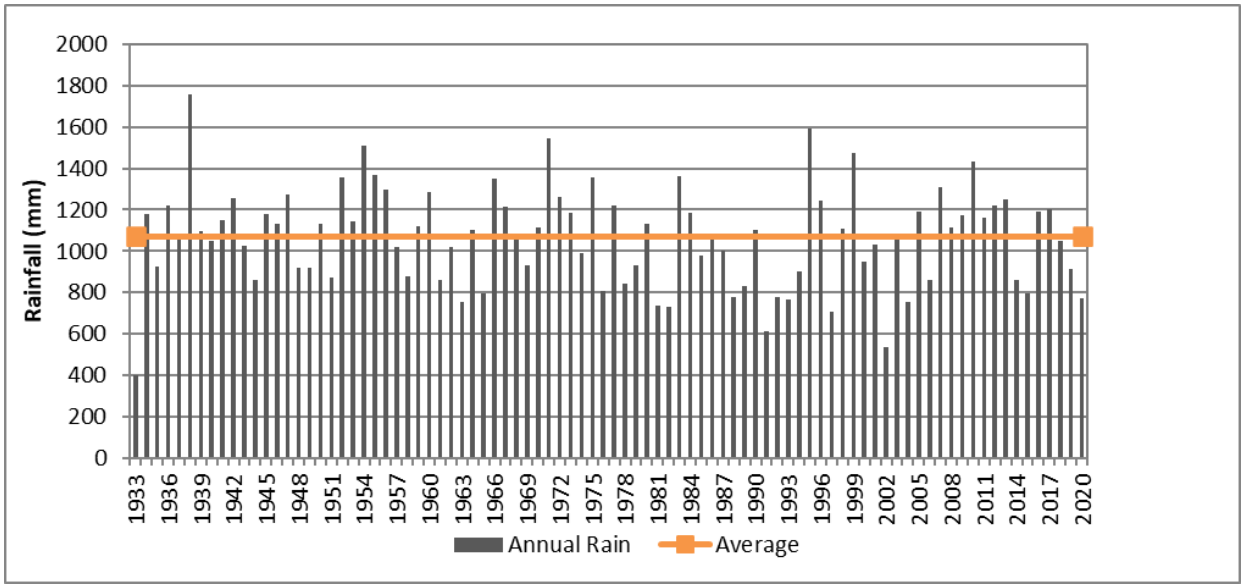


Figure 3-10: Annual rainfall depths measured at Nelshoogte (0518589 W) rainfall station

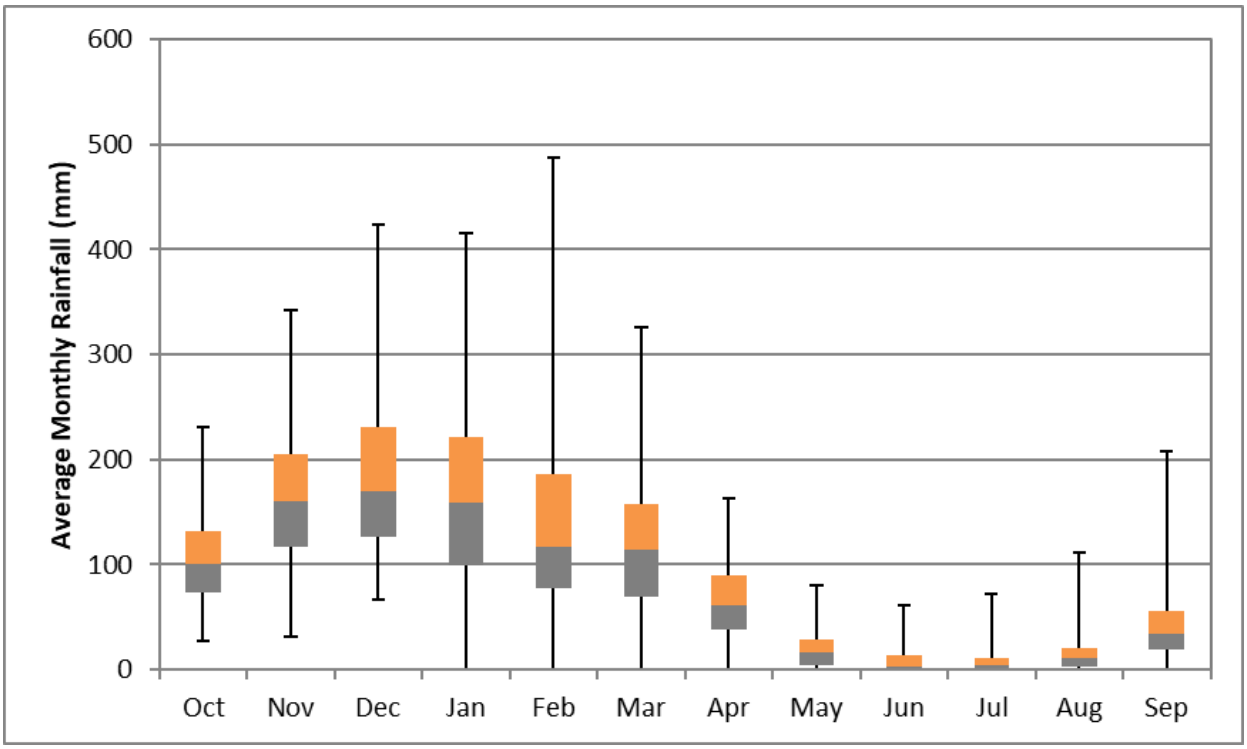


Figure 3-11: Monthly Box plot for the measured rainfall depths at Nelshoogte (0518589 W) rainfall station

3.4. GoldSim Software

GoldSim Software® was used to generate the rainfall stochastics for the TPM and WGEN Methods. GoldSim is a Monte Carlo simulation software solution for dynamically modelling complex systems in engineering, science, and business (GoldSim Technology Group, 2014). GoldSim supports decision-making and risk analysis by simulating future performance while quantitatively representing the uncertainty and risks inherent in all complex systems. The software is user-friendly and highly graphical, such that you can plainly draw and subsequently present an illustration of the system you wish to model in an intuitive way without having to learn a great deal of symbols, notation, and functions

The program runs on personal computers using 32-bit or 64-bit Microsoft Windows operating systems. Specifically, GoldSim requires Windows Vista, Windows 7, Windows 8, Windows 8.1, or Windows Server 2008. An academic license was acquired from GoldSim Technology Group for use in the analysis as part of this research.

3.5. Transition Probability Matrix (TPM) Procedure

The Transition Probability Matrix (TPM) method uses a state-Markov approach to generate the current day's rainfall based on previous day's rainfall state and a random variate. Up to seven rainfall "states" are recognized representing rainfall ranges increasing from state 1 (dry) up to a maximum of state 7 (≥31mm). The current rainfall is generated by linear interpolation between the state probabilities tabulated in the TPM for previous day's rainfall state. A Gamma distributed excess is added in the top state to generate the wettest days.

The number of states for each month can be determined from the guidance provided by McMahon and Srikanthan (1985). If the number of states is less than seven the upper limit of the last state is infinite, the shifted Gamma distribution is used to model rainfall amounts for the last state, while a linear distribution is used for the intermediate states. The TPM is developed by transition counting in the historical record; done separately for each calendar month to impose seasonality.

In Markov chain models, the probability of any future state is determined by the present state and the transition probability, which is the probability of a state switching from one to another, as shown in Equation no 3.1 (Zhou et al, 2017).

$$P\left[X(t_n) = x_n | X\left(t_{n-1}\right) = x_{n-1}, \ldots., X(t_1) = x_1\right] = P[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}] \quad \text{.....Equation (3.1)}$$

Where X(t) is a stochastic process, x_n is the state at $t_n, t_1 < t_2 < \cdots . < t_n$ (Zhou et al, 2017). The size of this matrix is determined by the number of independent states to be modelled; a typical transition probability matrix structure is shown in Figure 3-12.

To implement the Markov chain model, a uniform random number between 0 and 1 is then generated for each time step and the new state is determined by systematically comparing the

generated random number with transition probabilities (Zhou et al, 2017). Microsoft Excel functions and Visual Basic for Applications (VBA Coding) are used to calculate the values to make the process less time consuming.

State	1	2	..	i	..	n
1	$p_{1\to1}$	$p_{1\to2}$	..	$p_{1\to i}$	..	$p_{1\to n}$
2	$p_{2\to1}$	$p_{2\to2}$	..	$p_{2\to i}$	..	$p_{2\to n}$
..	..	..	..	..	..	..
i	$p_{i\to1}$	$p_{i\to2}$	..	$p_{i\to i}$	..	$p_{i\to n}$
..	..	..	..	..	..	..
n	$p_{n\to1}$	$p_{n\to2}$	..	$p_{n\to i}$	..	$p_{n\to n}$

Figure 3-12: Transition Probability Matrix Structure (Zhou et al, 2017)

The following steps are implemented in generating daily rainfall depths (Siriwardena, Srikanthan, & McMahon, 2002):

Step 1: Generate a uniformly distributed random number between 0 and 1. Using the appropriate TPM for the month, determine the state of the next day.

Step 2: If the state is wet, go to step 3. Otherwise, set the rainfall depth to zero and go to step 1.

Step 3: Calculate the rainfall depth by using the linear distribution for the intermediate states and shifted Gamma distribution for the largest state.

Step 4: Repeat steps 1 to 3 until the required length of daily rainfall data is generated.

Table 3-2 to Table 3-5 show the typical tables generated using the TPM Model:

- Daily rainfalls are divided into a number of states as shown in Table 3-2. Based on the observed data, the ranges for each state are determined.

Table 3-2:Example of Basis of states

Basis of States		
State	Top	Bottom
1	0	0
2	0.999	0.0001
3	2.999	0.999
4	6.999	2.999
5	14.999	6.999
6	30.999	14.999
7	1.00E+10	30.999

- Example of maximum states assumed for each month based on the rainfall pattern for a study area are shown below in Table 3-3.

Table 3-3: Maximum state for each month

Month	Maximum states
Jan	7
Feb	7
Mar	7
Apr	7
May	7
Jun	7
Jul	6
Aug	5
Sep	6
Oct	7
Nov	7
Dec	7

- The probabilities for rain in one state to be followed by rain on the next day in the same or another state are collated into a matrix - the transition probability matrix (TPM) - using as many whole years of rainfall record as are available. Seasonality of rain is modelled by using 12 TPMs, one for each calendar month. Table 3-4 shows an example of typical rainfall groups values developed for the months. Table 3-5 shows an example of the typical probability matrix generated.

Table 3-4: Example of Rainfall groups of states for the months

Rainfall group Values:		**Bottom of state**											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rain group of states	1	0	0	0	0	0	0	0	0	0	0	0	0
	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	3	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	4	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
	5	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0
	6	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0
	7	31.0	31.0	31.0	31.0	31.0	31.0	1E+10	1E+10	1E+10	31.0	31.0	31.0
		1E+10	1E+10	1E+10	1E+10	1E+10	1E+10	1E+10	1E+10	1E+10	1E+10	1E+10	1E+10
Bottom of top state		31.0	31.0	31.0	31.0	31.0	31.0	15.0	15.0	15.0	31.0	31.0	31.0

Table 3-5: Example of probabilities matrix generated for each month

		State in the current day						
Jan	1	1	2	3	4	5	6	7
State in the previous day	1	0.674	0.015	0.042	0.077	0.085	0.064	0.043
	2	0.530	0.072	0.096	0.060	0.145	0.084	0.012
	3	0.553	0.059	0.128	0.096	0.096	0.032	0.037
	4	0.504	0.046	0.111	0.130	0.053	0.088	0.069
	5	0.462	0.052	0.120	0.124	0.108	0.084	0.048
	6	0.365	0.069	0.101	0.185	0.116	0.095	0.069
	7	0.313	0.047	0.070	0.148	0.195	0.141	0.086
		0	0	0	0	0	0	0
Feb	2	1	2	3	4	5	6	7
	1	0.732	0.017	0.040	0.061	0.062	0.056	0.032
	2	0.616	0.041	0.082	0.123	0.123	0.014	0.000
	3	0.500	0.049	0.118	0.146	0.111	0.035	0.042
	4	0.520	0.079	0.084	0.123	0.093	0.066	0.035
	5	0.439	0.061	0.096	0.197	0.076	0.071	0.061
	6	0.391	0.034	0.098	0.103	0.149	0.149	0.075
	7	0.314	0.029	0.038	0.133	0.143	0.190	0.152

- Linking the transitions calculations spreadsheet developed to the Gold Sim Software

The calculations of the probability matrix completed in MS Excel are then linked to the GoldSim software to generate the rainfall stochastics because of GoldSim's ability to dynamically generate data into the future. The screenshot below in Figure 3-13 shows the calculations for the TPM model developed in GoldSim once the data is imported into the software. All the calculations and

functions are carried out using GoldSim once the tables of the probability matrices developed in MS Excel have been linked.

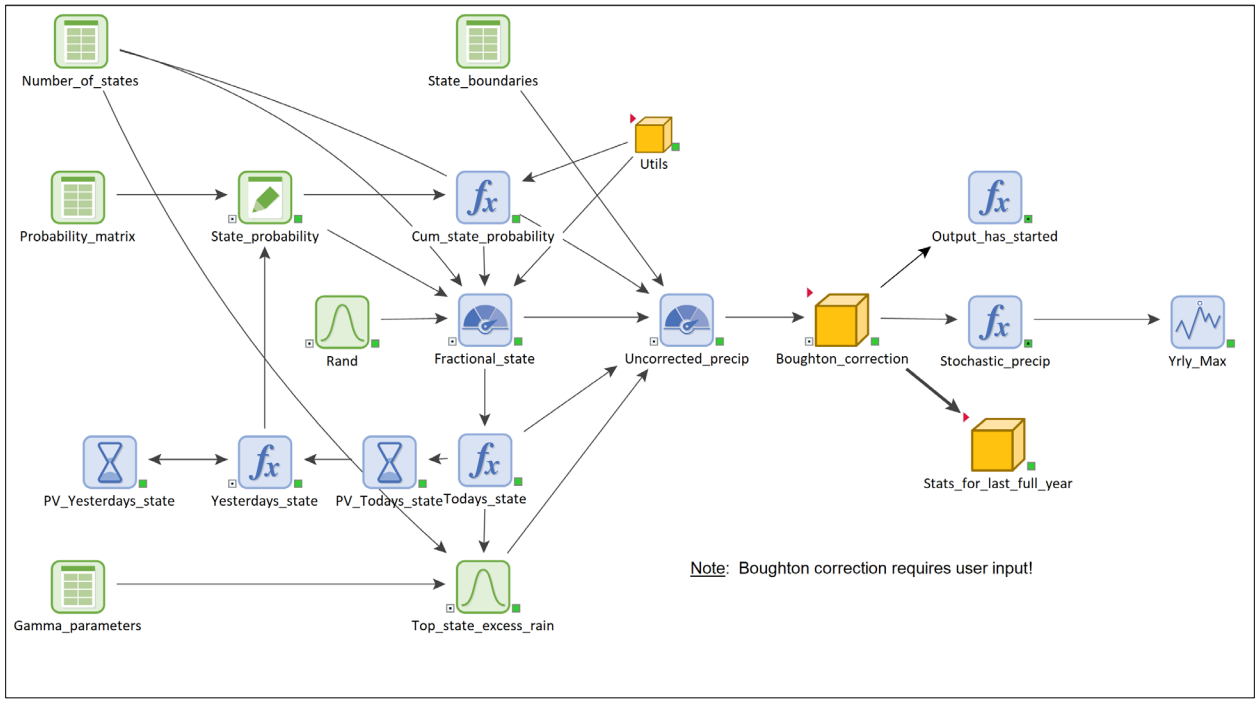


Figure 3-13: Screenshot from GoldSim showing the calculations to generate the daily rainfall data

The model is then improved by adopting an empirical adjustment factor (F) to match the observed standard deviation of the annual rainfall (Boughton, 2009).The adjustment factor is obtained by trial and error until the frequency distribution of the observed and generated annual rainfalls matches. The generated daily rainfall in each year is multiplied by the following ratio:

$$Ratio_i = \frac{\{M + (T_i - M)F\}}{T_i} \quad \text{.....equation (3.2)}$$

where M is the observed mean annual rainfall, and T_i is the generated annual rainfall for year i. Since the slope of the frequency curve is proportional to the standard deviation, the adjustment factor can be directly obtained as a ratio of the standard deviation of the generated and observed annual rainfall (Siriwardena, Srikanthan, & McMahon, 2002). Figure 3-14 shows application of Boughton correction in GoldSim Software.

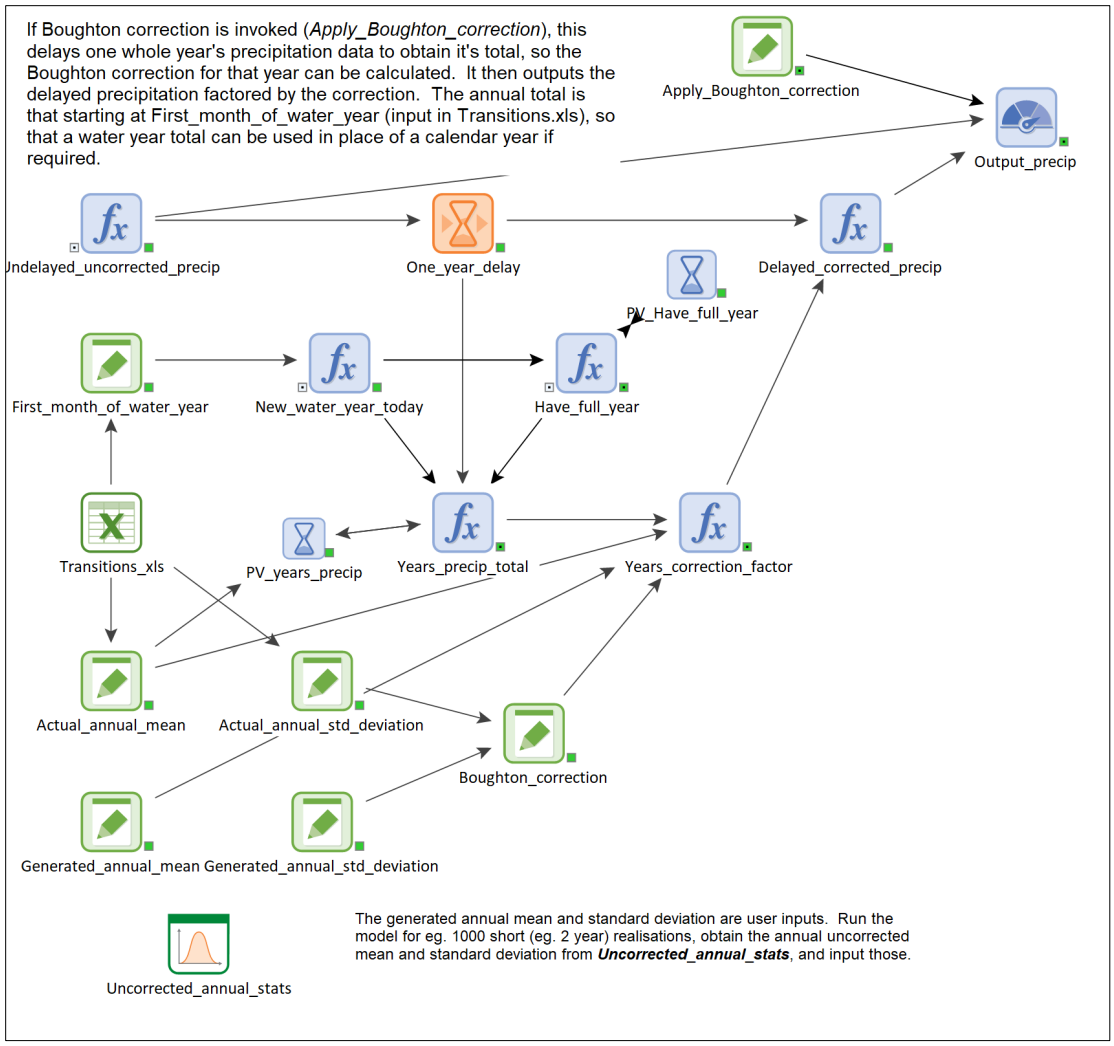


Figure 3-14: Boughton correction calculated using GoldSim software for generating daily rainfall

- Calibration of the Model

Calibration of the model is concerned with preserving the daily, monthly, and annual statistics of the observed data. This is achieved by iteratively changing the limits of each state as well as the number of the states to be considered in each month.

3.6. Weather Generator (WGEN) Procedure

A computer simulation model called WGEN (Weather Generator) has been developed to generate daily values for precipitation, maximum temperature, minimum temperature, and solar

radiation. The model is based on the procedure described by (Richardson & Wright, 1984). For the purposes of this research, only precipitation values were generated.

WGEN provides daily generated values of precipitation (p), maximum temperature (t_{max}), minimum temperature (t_{min}), and solar radiation (r) for a set number of years at a given location. The approach that is used is to generate precipitation for a given day independently of the other variables. Maximum temperature, minimum temperature, and solar radiation are then generated according to whether a wet day or dry day was previously generated. The model is designed to preserve the dependence in time, the correlation between variables, and the seasonal characteristics in actual weather data for the location.

The precipitation component of WGEN is a Markov chain-gamma model. A first order Markov chain is used to generate the occurrence of wet or dry days. When a wet day is generated, the two-parameter gamma distribution is used to generate the precipitation amount. With the first-order Markov chain model, the probability of rain on a given day is conditioned on the wet or dry status of the previous day (Richardson & Wright, 1984). A wet day is defined as a day with 0.01 inch or rain of more. Let P_i (W/W) be the probability of a wet day on day i given a wet day on day i-1 and let P_i(W/D) be the probability of a wet day on day i given a dry day on day i-1.

Then,

P_i (D/ W) =1- P_i(W/W) --------equation (3.3)

P_i(D/D) =1= P_i(W/D) ----------equation (3.4)

where P_i(D/W) and P_i(D/D) are the probabilities of a dry day given a wet day on day i-1 and the probability of a dry day given a dry day on day i-1, respectively. Therefore, the transition probabilities are fully defined by the given Pi(W/W) and Pi(W/D). Several probability densities functions have been used to describe the distribution of rainfall amounts (Smith & Schreiber, 1974), (Woolhiser & Roldan, 1982).

For this application, a distribution with a minimum number of parameters was needed to minimize the problem of defining the parameters for a large number of locations.

Richardson (1982) has shown the two-parameter gamma distribution to be significantly better for describing daily precipitation amounts than the simple one-parameter exponential distribution.

In WGEN, each of the four precipitation parameters are constant for a given month but are varied from month to month. The parameters are used with a Markov chain generation procedure, and the gamma generation procedure described by (Haan, 1977) to generate daily precipitation values.

GoldSim Software was used to set up the WGEN, including parameter generation that are required to simulate generation of daily rainfall depths. A screenshot of the generated parameters exported from GoldSim are shown in Figure 3-15. Figure 3-16 shows the WGEN precipitation generation set up in GoldSim Software once the parameters required have been obtained.

Month	Pi(W/W)	Pi(W/D)	Alpha	Beta	RainObs	Num of Wet Days	Run Wet Days	Rain > 2	Tmax	Max Tmax	Tmax > 95	Tmin	Min Tmin	Tmin < 32	Radiation
Units -->				mm	mm				C	C		C	C		W/m2
Jan	0.411	0.177	0.637	0.310	1.42	7.1	2.6	0.00	38.2	65.1	17.3	19.7	-2.7	2.2	170
Feb	0.403	0.221	0.708	0.310	1.68	7.6	2.7	0.00	43.2	64.9	20.3	24.2	4.7	0.5	233
Mar	0.459	0.275	0.694	0.348	2.51	10.4	3.4	0.00	51.3	77.3	27.4	31.3	13.4	0.2	322
Apr	0.483	0.316	0.773	0.413	3.64	11.4	3.7	0.00	66.1	86.2	30.0	44.5	29.1	0.0	423
May	0.454	0.282	0.784	0.540	4.43	10.4	3.6	0.00	75.7	89.2	31.0	54.3	38.1	0.0	503
Jun	0.471	0.274	0.684	0.589	4.07	10.4	3.9	0.00	83.9	94.1	30.0	63.2	50.6	0.0	536
Jul	0.447	0.236	0.716	0.600	4.12	9.4	3.4	0.00	88.8	98.4	31.0	67.6	56.2	0.0	561
Aug	0.346	0.210	0.659	0.570	2.85	7.6	2.6	0.00	87.5	97.3	31.0	65.6	53.0	0.0	477
Sep	0.427	0.208	0.612	0.821	4.00	8.0	3.1	0.00	80.5	93.2	30.0	57.6	41.6	0.0	381
Oct	0.412	0.191	0.574	0.720	3.07	7.5	2.8	0.00	69.4	86.3	31.0	46.8	30.4	0.0	252
Nov	0.351	0.166	0.700	0.365	1.65	6.3	2.7	0.00	54.4	74.1	27.9	34.4	16.0	0.0	168
Dec	0.426	0.212	0.757	0.266	1.66	8.3	2.9	0.00	42.0	66.6	20.5	24.9	5.6	0.8	139

Temperature and Radiation Inputs:	Abbreviation	Value	Unit	
Daily mean max temp (dry)	TXMD	65.70	C	
Amplitude Max Temp (dry)	ATX	25.58	Cdeg	
Coeff. Var. for Max Temp (dry)	CVTX	0.08	Cdeg	
Amplitude Coeff. Var. for Max Temp (dry)	ACVTX	-0.05	Cdeg	
Daily mean max temp (wet)	TXMW	63.66	C	
Daily mean Min temp	TN	43.59	C	
Amplitude Min Temp	ATN	23.76	Cdeg	74.75933325
Coeff. Var. for Min Temp	CVTN	0.09	Cdeg	
Amplitude Coeff. Var. for Min Temp	ACVTN	-0.06	Cdeg	
Mean Solar Radiation (dry)	RMD	236.7	W/m2	
Amplitude of Solar Radiation (dry)	AR	128	W/m2	
Mean Solar Radiation (wet)	RMW	236.3	W/m2	

Figure 3-15: Typical Screenshot of Generated WGEN Parameters

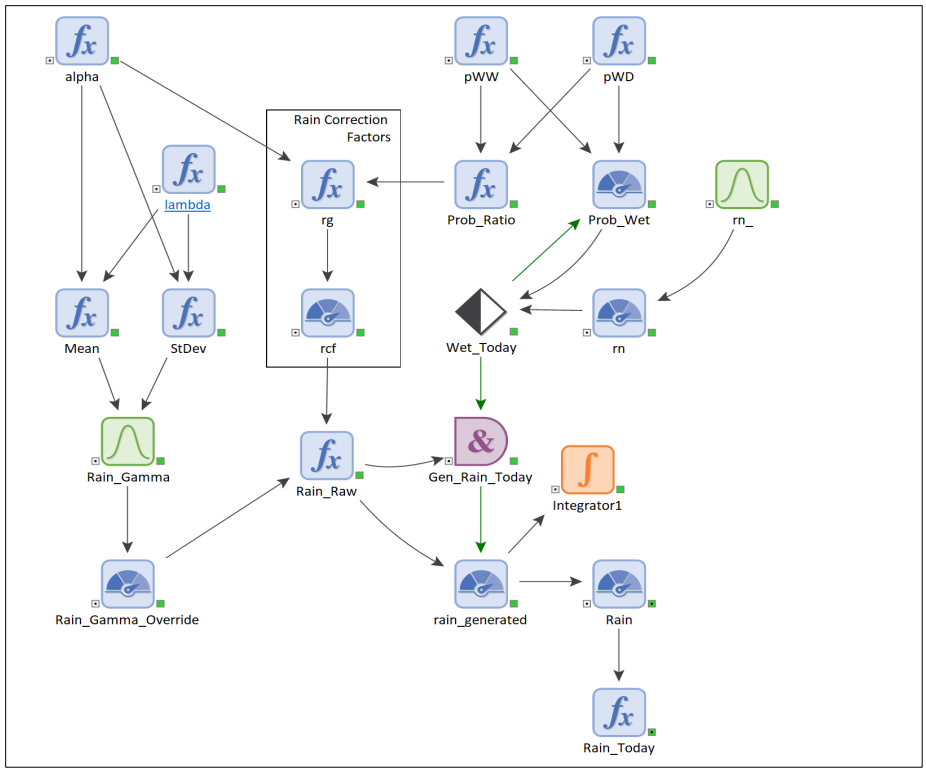


Figure 3-16: Screenshot of WGEN Precipitation generator set up in GoldSim

3.7. Variable Lengths Bootstraps (VLB) Procedure

Fortran programming language was used to code the applications used to generate rainfall stochastics using the VLB Method. Fortran is a general-purpose, compiled high-level programming language that is especially suited to numeric computation and scientific computing. Once the VLB was implemented in Fortran, an executable (.exe) file that is designed to work with text files generated in MS Excel is generated and then using the software does not therefore require a Fortran compiler.

The VLB daily rainfall generator disaggregates the annual stochastic rainfalls from the monthly generator to daily stochastic rainfalls described in detail in WRC Report No.2148/1/13 completed by Ndiritu and Nyaga (2014) and the steps carried out for this are summarized below in this section.

Although the VLB model by Ndiritu and Nyaga (2014) generates both monthly and annual stochastic rainfalls, the observed variation and distribution of daily rainfalls across the year is considered applicable for a direct disaggregation of rainfalls from an annual to a daily time step without the intermediate monthly time step.

Furthermore, the complexity and the computation required for annual to daily disaggregation is assessed as lower than that might be required for an annual to monthly to daily disaggregation. The annual to daily stochastic generator involves the following steps:

i. The nearest neighbour of the annual stochastic rainfall ASi of year i, are identified from the historic rainfall sequence on the basis of annual rainfall magnitude. The nearest neighbors are therefore the annual rainfalls whose values are closest to the annual stochastic rainfall. An equal number of neighbors with lower and with higher rainfall are identified and the number of neighbors selected is a therefore a multiple of two. An exception to this is when the stochastic annual rainfall is either too high or too low for an equal number of higher and lower historic neighbors to be obtained. When this occurs, the nearest neighbors are identified with the rule of an equal proportion of lower and higher neighbors excluded. The number of neighbors to select is a subjective modelling parameter and a value of 8 has been applied in the current modelling.

ii. From the nearest historic neighbors identified in step i, a selected number are randomly selected to supply the initial daily rainfall amounts for set parts of the year. The selected neighbors fill in equal proportion of the year to create the initial daily rainfall sequence for the year. For example, if 5 neighbors are selected, then each neighbor supplies daily rainfalls for a continuous period of 365/5 = 73 days to fill up the year.

iii. As would be expected, the summation of the rainfalls obtained in step ii is highly unlikely to equal the stochastic annual rainfall ASi. Furthermore, it is imperative that the final stochastic daily rainfalls do not simply remain as the resampled historic values (from step ii) since it is certain that the future daily rainfalls will be different

from the historic ones. The daily stochastic rainfalls from step ii are therefore proportioned in a manner that retains the total magnitude of the annual stochastic streamflow ASi. This adjustment helps ensure that new daily rainfall data are generated.

iv. In order to ensure the replication of cross correlations where multiple rainfall stations are involved, the contemporaneous approach described by Ndiritu and Nyaga (2014) and Ndiritu (2011) is applied.

The following files and procedure were followed in generating the daily rainfall depths:

1. Three text files are required to run the applications, namely:

   a. Historic daily rainfall.txt consisting of the daily rainfall data from the rainfall stations, the format of the data is shown below in Figure 3-17.

```
Tzaneen
13879   3               Tz_0594760_W.txt        Tz_0594764_W.txt        Tz_0518589_W.txt
                        1       2       3
year    month   day     mm      mm      mm
62      1       1       0.8     2       2
62      1       2       6.4     7.4     0
62      1       3       0       0       0
62      1       4       0       0       0
62      1       5       0.6     2.3     0
62      1       6       0       0       6.7
62      1       7       0       0       0
62      1       8       0.1     0       0
62      1       9       11.9    0       60.9
62      1       10      3       10.2    0.2
62      1       11      0       0       0
62      1       12      0       0       0
62      1       13      0       0       8
62      1       14      18.4    4.2     0
62      1       15      6.3     0       0
62      1       16      0.3     0       2
62      1       17      18.1    10.5    6
62      1       18      107.2   100.9   24.5
62      1       19      11.1    11.9    0
62      1       20      7.3     18      21
62      1       21      38.4    0.2     0
```

Figure 3-17: Screenshot of the historic daily rainfall.txt

   b. RainfilesNames.txt consisting of names of the rainfall stations and the number of them as shown in Figure 3-18 below.

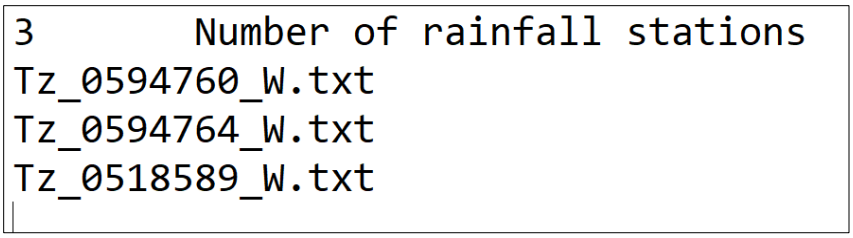


Figure 3-18: Screenshot of the RainfilesNames.txt

c. VLBSettingsandfiles.txt which consist of the length of historic data in years, length of the stochastic sequences to be generated in years and minimum block length among other as shown in Figure 3-19.

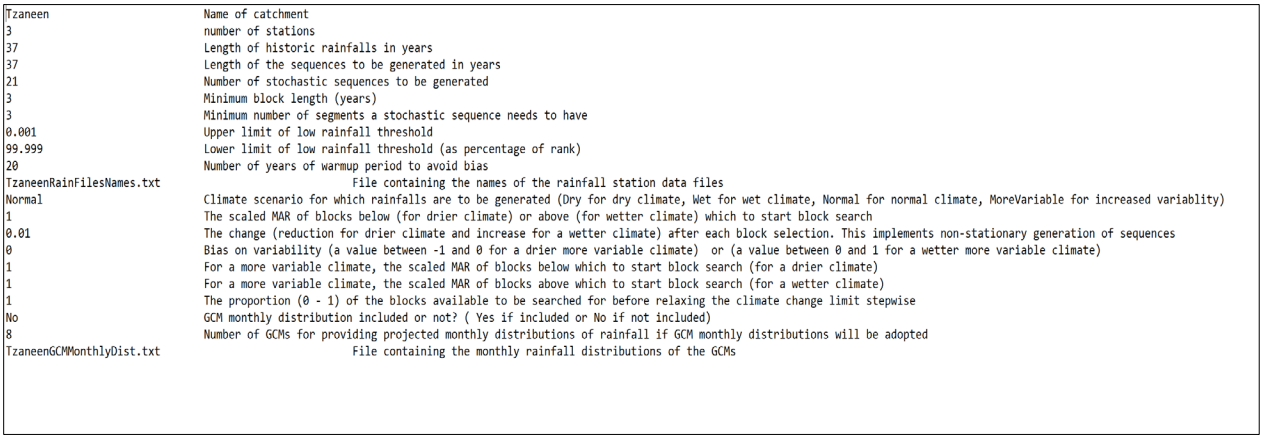


Figure 3-19: Screenshot of the VLBSettingsandfiles.txt file

2. Run the DailyToMonthlyRainfallCreator.exe application, which produces monthly rainfall data files for use by the VLB Monthly Generator.exe. It also carries out an analysis relating annual rainfalls to the number of annual rain days and outputs this as RainAndRainDays.txt.

3. Run the VLB Monthly Generator.exe application. This produces text files of analysis on the stations such as historic annual rainfalls, historic monthly stats, block length and averages and cross relations among others. For the purposes of the daily generation, only the monthly stochastic generation output files are needed. These have names such as Tz-Generated-Tz_0678776_W.txt.

4. Run the AnnualToDailyDisaggregatoe.exe. application. This produces the text files with the generated stochastic daily rainfall data of the requested sequences.

Figure 3-20 shows the summary flow chart of the VLB process followed. The modelling is constrained to 10 sites, 100 years long data and 500 stochastic sequences. The settings can be altered or generalized to allow for any size with ease in the original code.

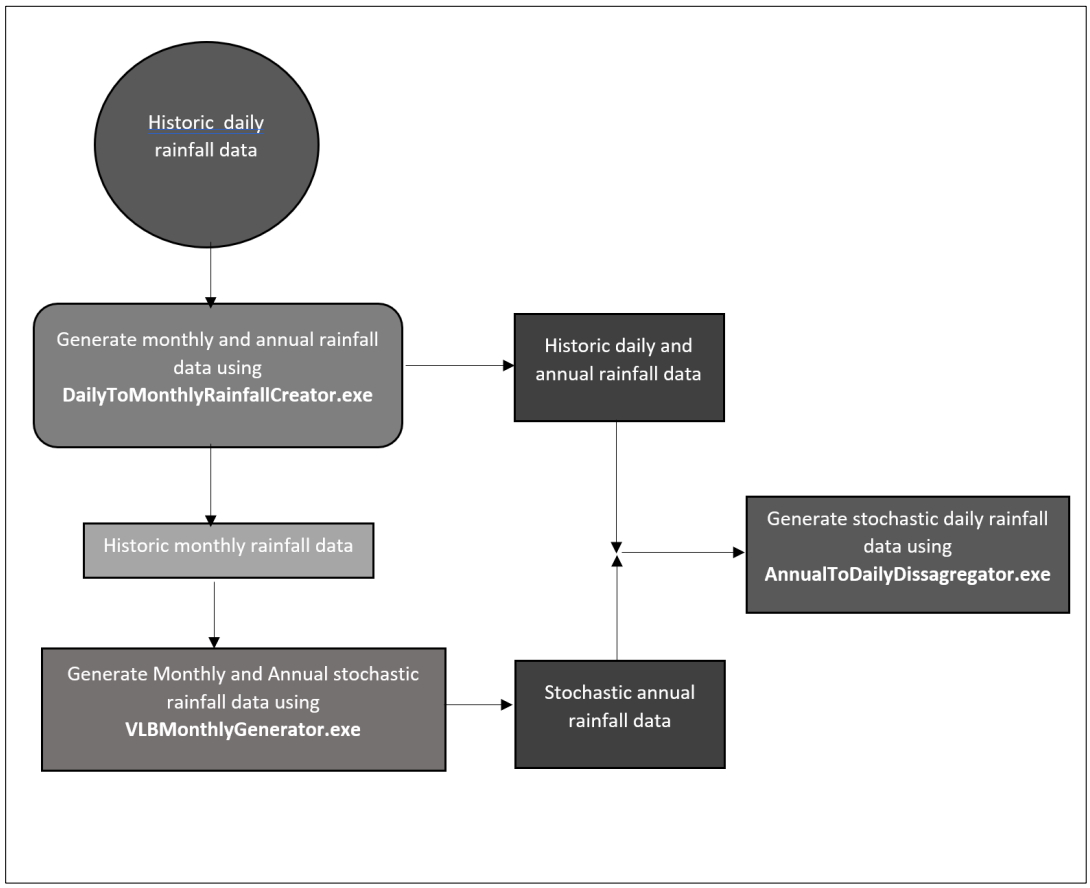


Figure 3-20: Flow chart of the VLB Process

4. PERFORMANCE EVALUATION OF THE GENERATORS

Evaluation of performance and comparison of the generators is mainly done using box and whisker plots. Figure 4-1 defines the percentiles that are used in the box plots. As a general guideline, the replication of a statistic is considered satisfactory if the historic value falls within the box and whisker plot especially the inter-quartile range of the box plot. The box plots help to assess the variability and to identify the range of the generated statistics.

The lower and upper extremes of the box and whisker plot are considered to be the 5th and 95th percentiles respectively.

It is worth noting that models do not provide a perfect representation of reality, and the results can be uncertain (Muller et al, 2009). The predictability of the atmosphere is not perfect, this brings into context the fact that although statistical methods are necessary, results obtained are not totally accurate which is the reason why uncertainties (errors) are there (Agbo, 2021). Daniel S. Wilks (1995) supported the use of these methods on the analyses of different regions that do not necessarily have the same climatic condition because mathematical and statistical laws are true irrespective of the region, the same methods can be applied in different regions to yield accurate results (Agbo, 2021). Models have an intrinsic inability to provide a perfect simulation of the hydrological processes involved in the Rainfall-Runoff transformation (Montanari & Brath, 2004).

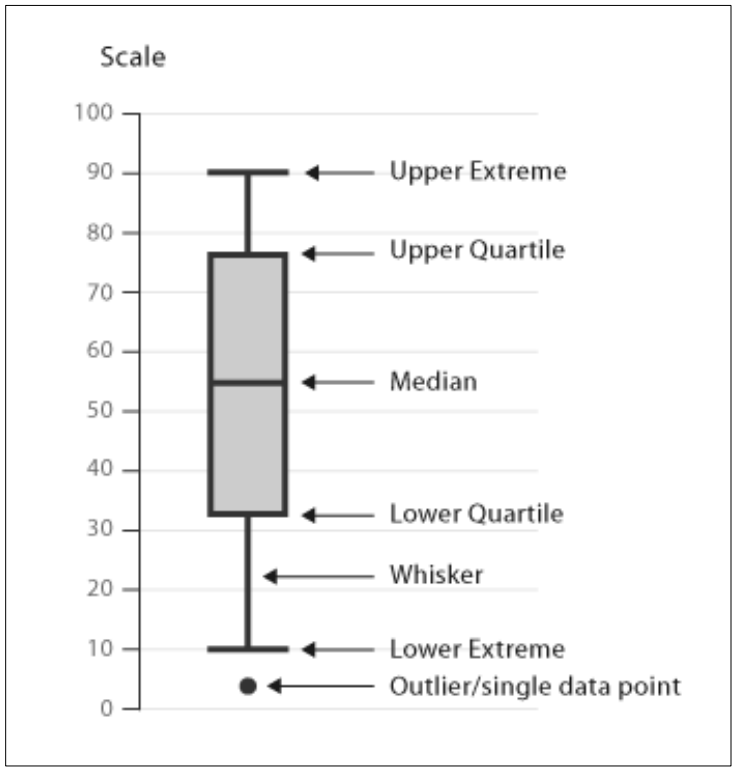


Figure 4-1: Box plot legend representing the various quartiles as used in the stochastic analysis

4.1. Performance Evaluation of the TPM Method

4.1.1. Hebron Rainfall station

Figure 4.1-1 shows the result of the calibration of the TPM model for the Hebron rainfall station from GoldSim. Corrected annual probability of exceedance is obtained by adjusting the standard deviation so that it best fits the recorded annual precipitation.

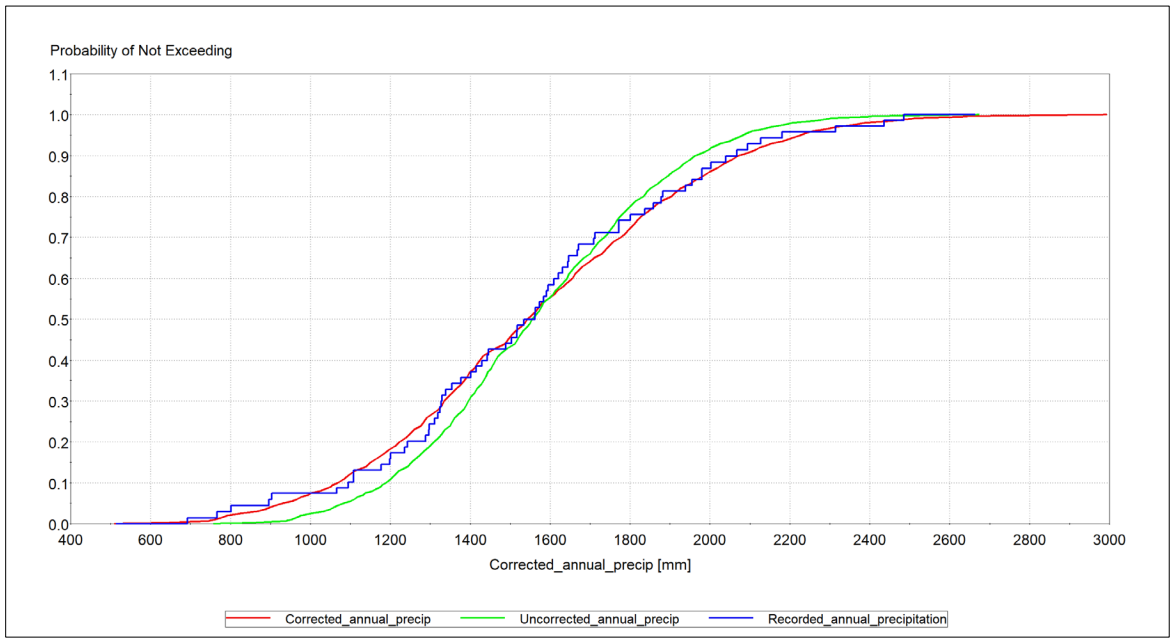


Figure 4.1-1: Comparison of the annual probability of exceedance for the generated annual precipitations to the recorded data for Hebron rainfall station

Daily statistics of the generated data using TPM Method for Hebron rainfall station

Figure 4.1-2 shows rankings from maximum rainfall depths in a day of each year for each sequence, while Figure 4.1-3 shows percentiles from daily rainfall depths of the whole 50-year period generated from each sequence from the 80th percentile to the 98th percentiles using the TPM method for Hebron station.

The daily maximum values were included in the analysis as wet seasons are often the cause for a need of a water management strategies in the mining industry. The box and whisker plots of the rainfall distribution curve show that the historic data falls within the interquartile range of the box and whisker plot.

The values were plotted from 80th percentiles because before the 80th percentile most values for the percentiles are zero and after the 98th percentile the values are high which skews the graph and therefore the 99th and 100th percentiles are not included in the plots. The box and whisker plot show that the percentiles of the historic data fall within the interquartile range except for the 99th percentile, whereby the historic value is lower than the generated values

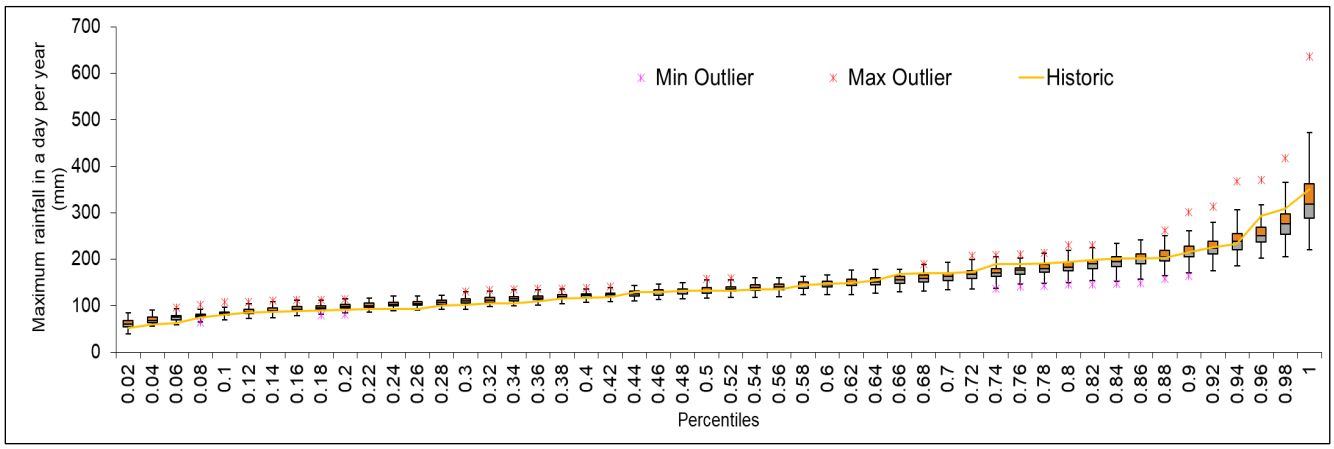


Figure 4.1-2: Hebron station rainfall distribution curve from the daily maximum values using TPM Method

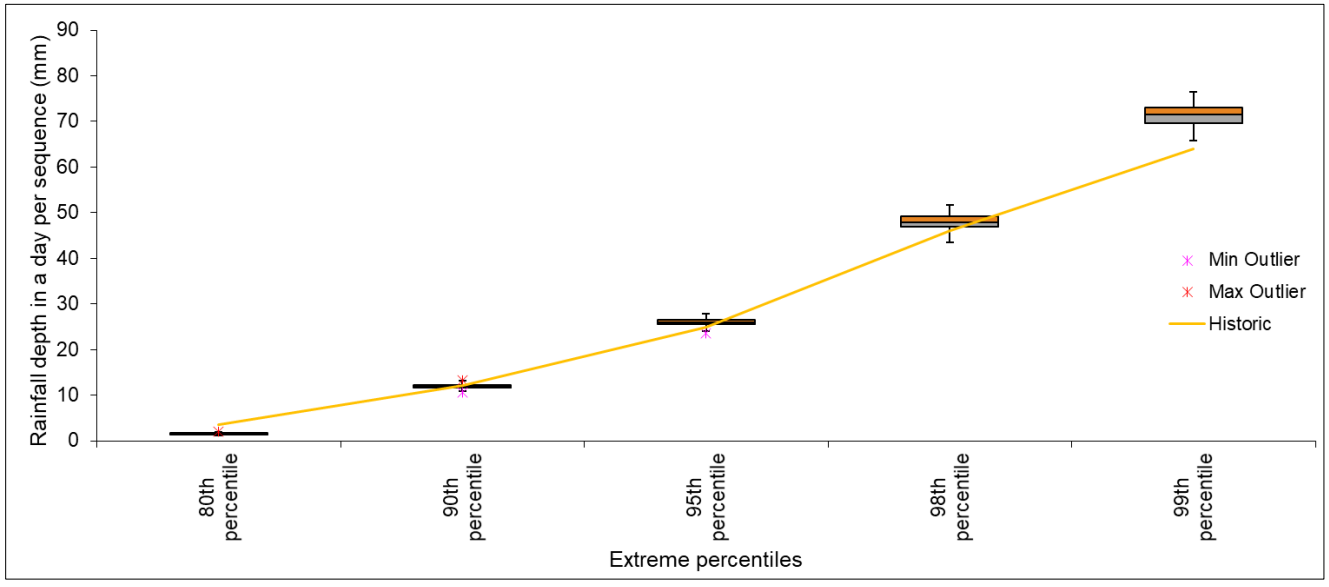


Figure 4.1-3: Percentiles of daily data generated using TPM Method for Hebron station

Monthly statistics of the generated data using TPM Method for Hebron rainfall station

Figure 4.1-4 to Figure 4.1-7 show the monthly statistics of the generated data using TPM Method based on the generated 100 sequences for 50 years. The monthly averages show that it compares well with the historic data as it falls within the interquartile range of the generated data except for July whereby the historic average is outside of the box plot range.

The maximum and standard deviation values shows that generally the data generated for the month of July is lower than the historic value and for the months of June and August are higher than the historic values which could be due to parameter assumptions in the TPM method. The monthly percentiles however compare relatively well to the historic values except for the 75th percentile.

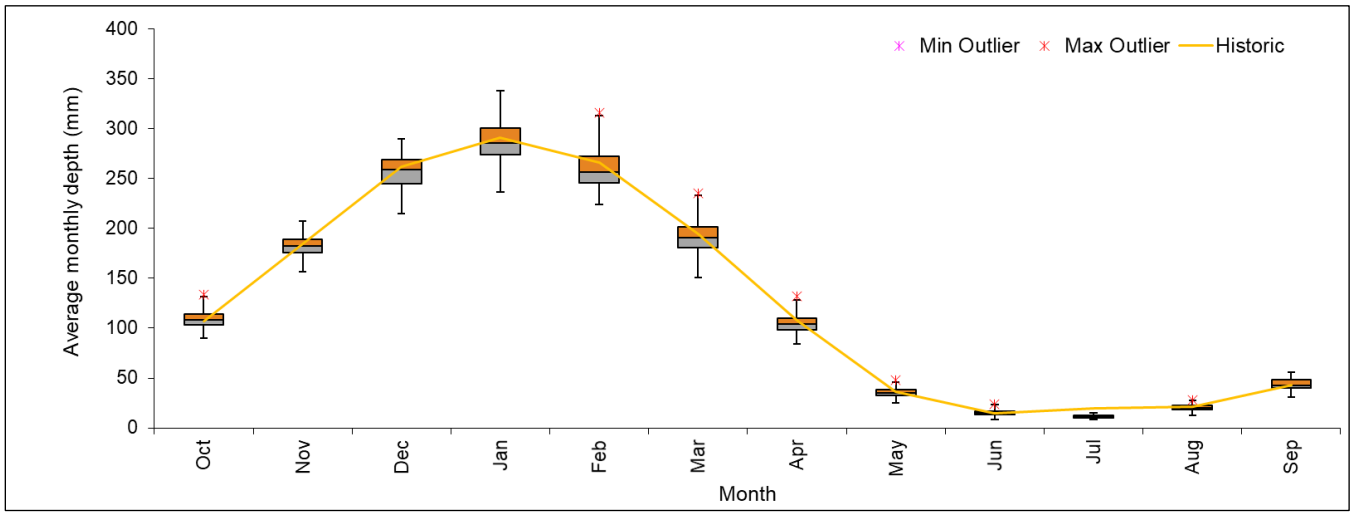


Figure 4.1-4: Averages of monthly data generated using TPM Method for Hebron station

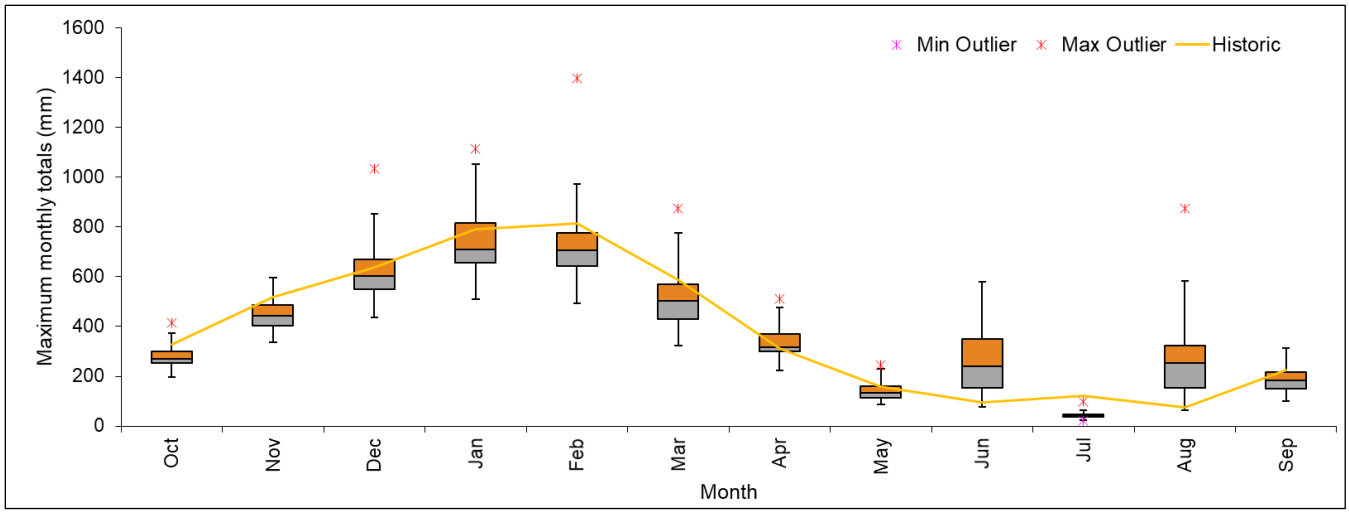


Figure 4.1-5: Maximum values of monthly data generated using TPM Method for Hebron station

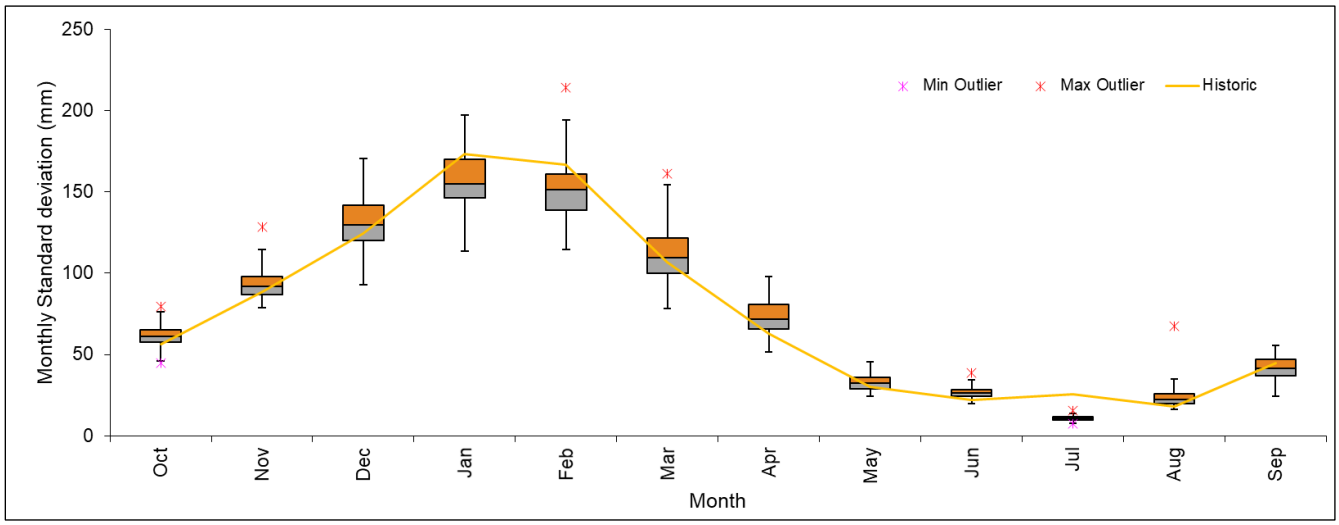


Figure 4.1-6: Standard deviation values of monthly data generated using TPM Method for Hebron station

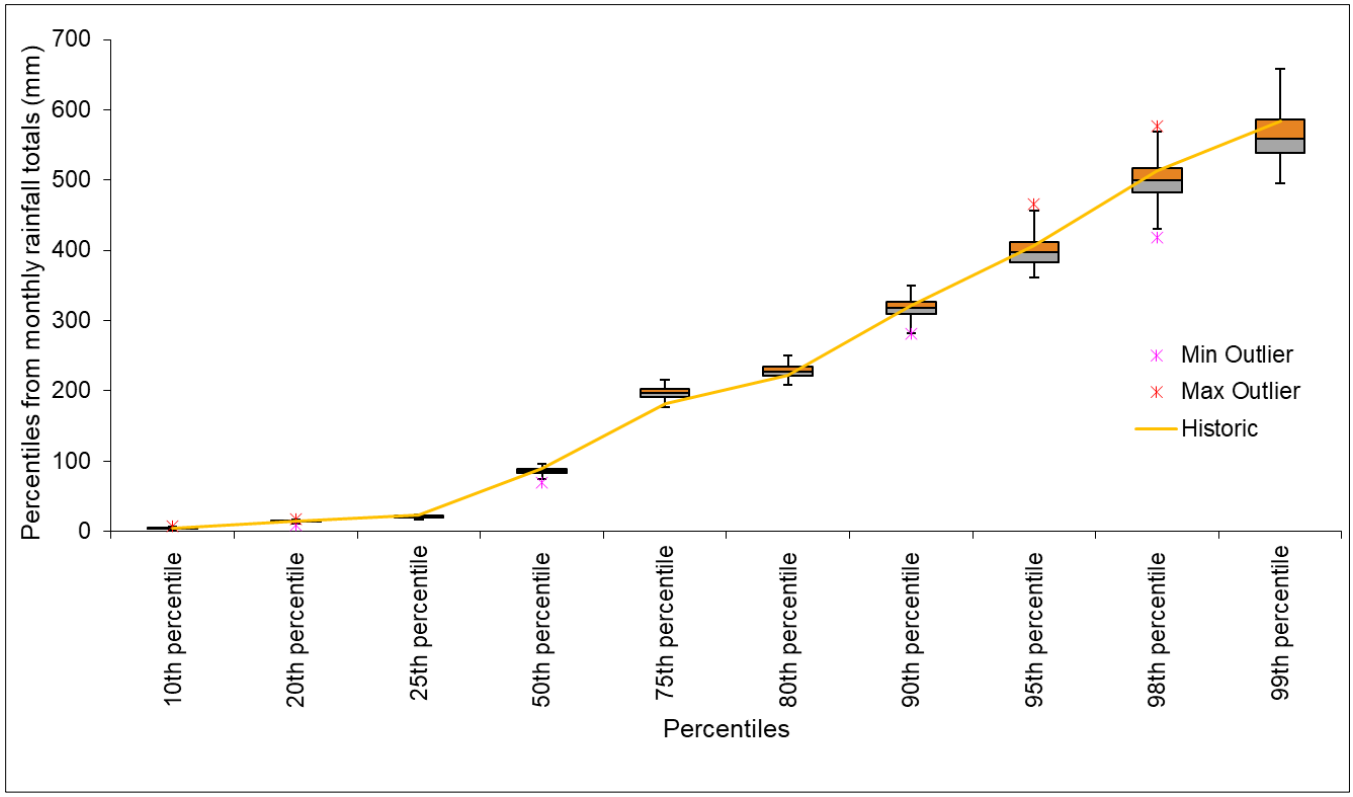


Figure 4.1-7: Percentiles of monthly data generated using TPM Method for Hebron station

Annual statistics of the generated data using TPM Method for Hebron rainfall station

Figure 4.1-8 and Figure 4.1-9 show the annual rainfall distribution and percentiles curve respectively. The historic values fall within the interquartile range of the generated data except for values of less than 8%. The annual percentiles shows that the generated annuals compared well with the historic data except for the low and high percentiles.

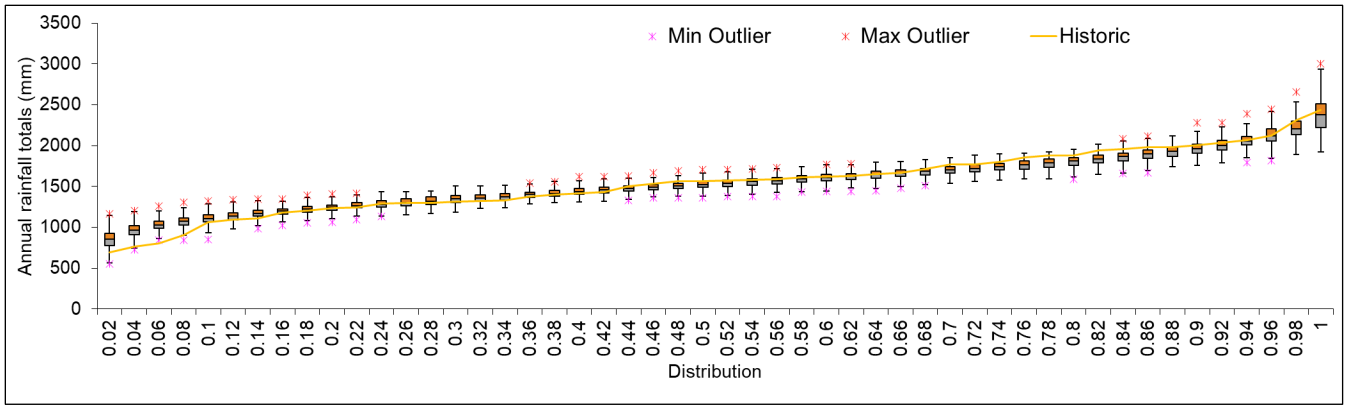


Figure 4.1-8: Rainfall distribution curve of the annual totals for data generated using TPM Method for Hebron station

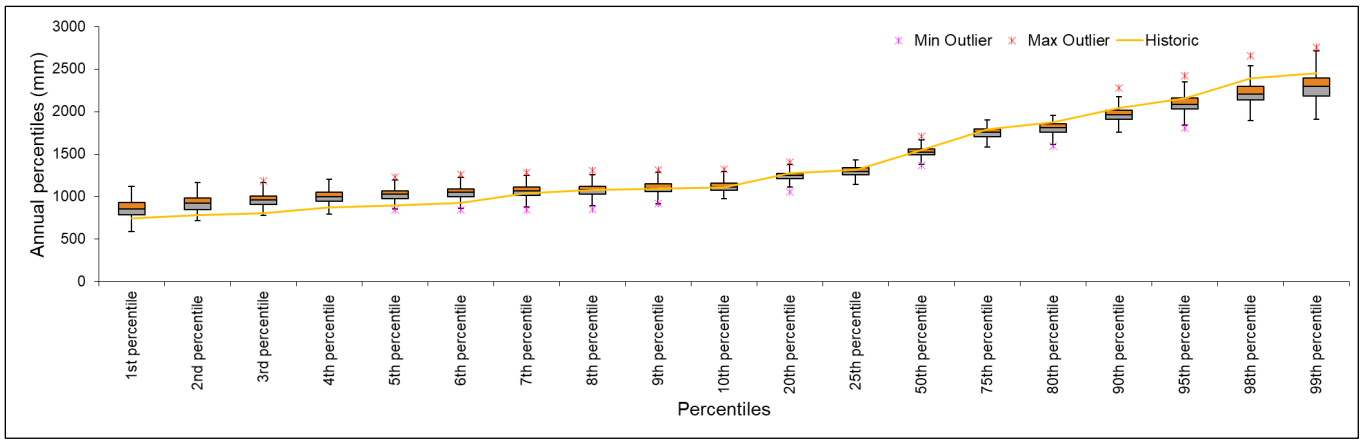


Figure 4.1-9: Annual percentiles of data generated using TPM Method for Hebron station

4.1.2. Welgevonden Rainfall

Figure 4.1-10 shows the result of the calibration of the model for the Welgevonden rainfall station. As can be seen, the uncorrected annual probability exceedance does not fit the recorded annual precipitation while the corrected one fits much better.

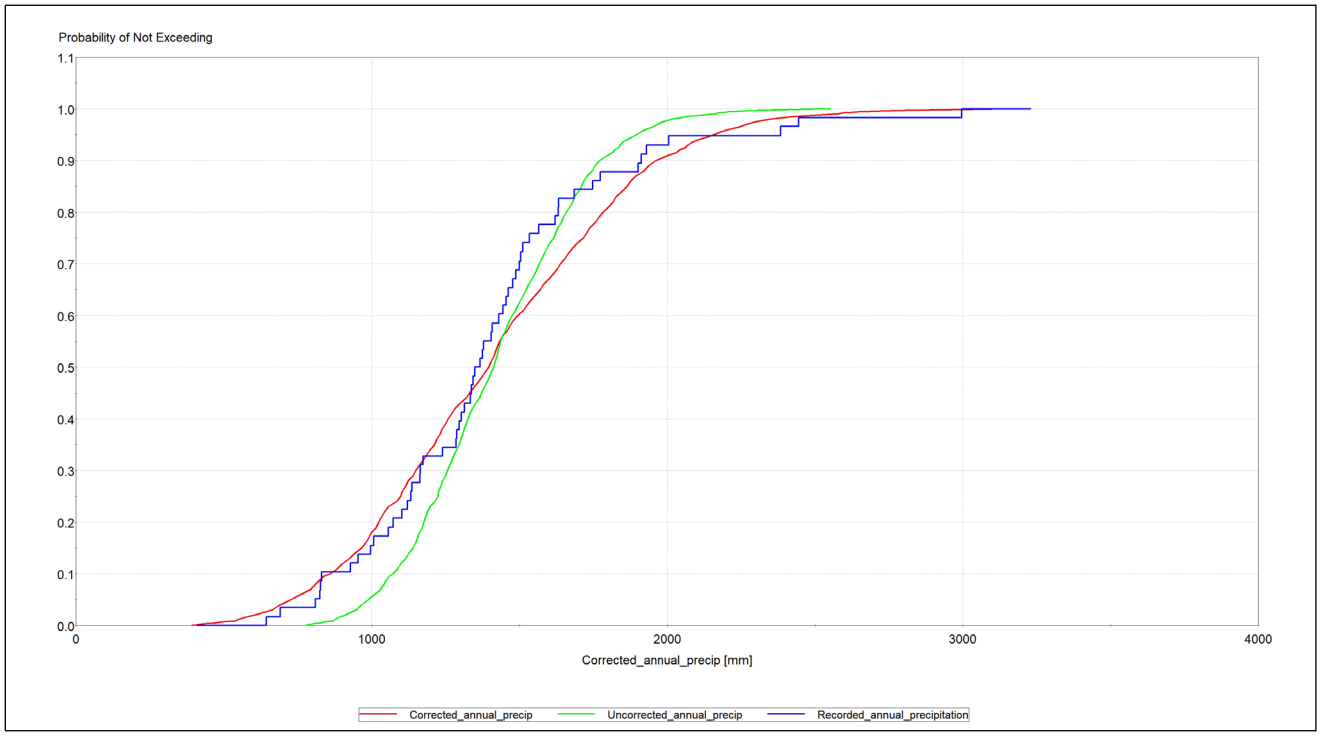


Figure 4.1-10: Comparison of the annual probability of exceedance for the generated annual precipitations to the recorded data for Welgevonden rainfall station

Daily statistics of the generated data using TPM Method for Welgevonden rainfall station

Figure 4.1-11 and Figure 4.1-12 show the box and whisker plots of the rainfall distribution curve from the daily maximum values and the daily percentiles from 100 sequences of 50 years daily generated data using the TPM method for Welgevonden station. The daily maximum rainfall distribution curve shows that the historic data fall within the interquartile range of the box and whisker plots except for the distribution values of less than 14 %. As can be seen in Figure 4.1-12, the daily percentiles of the historic data fall within the box and whisker plot except for the 98th and 99th percentiles.

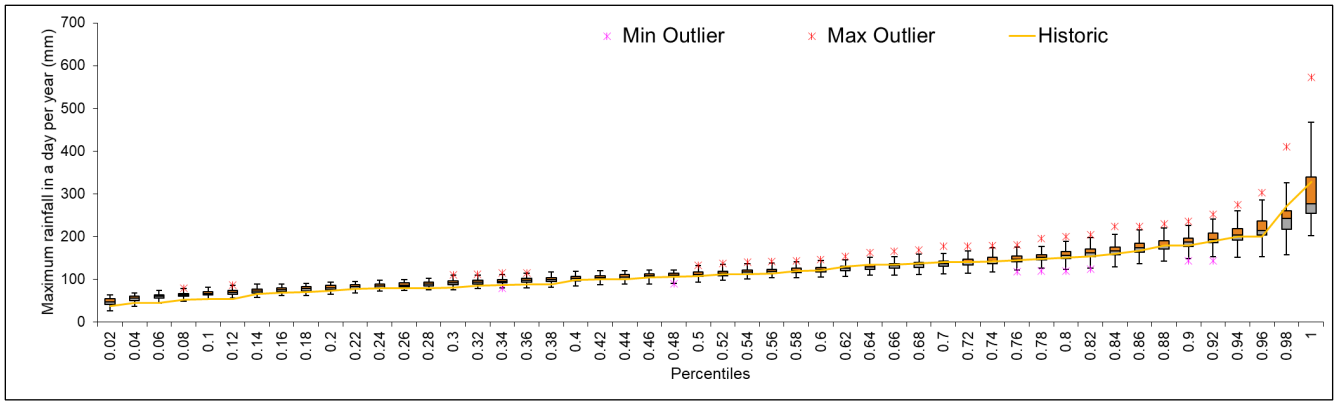


Figure 4.1-11: Welgevonden station rainfall distribution curve from the daily maximum values using TPM Method

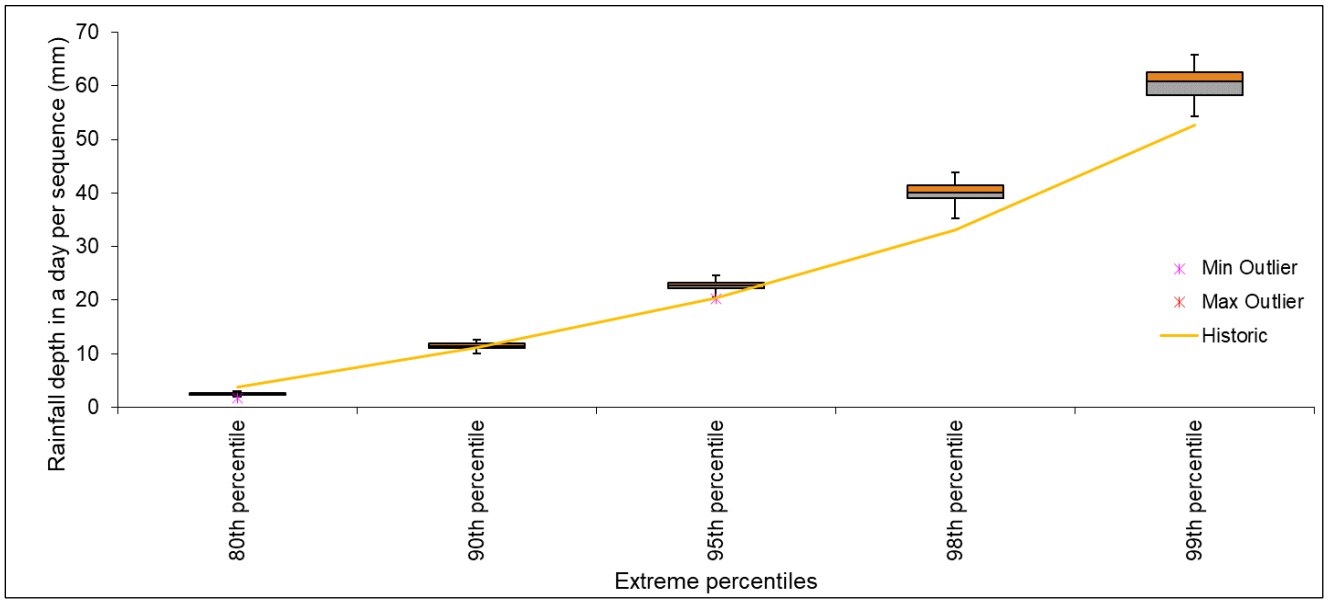


Figure 4.1-12: Percentiles of daily data generated using TPM Method for Welgevonden station

Monthly statistics of the generated data using TPM Method for Welgevonden Rainfall Station

Figure 4.1-13 to Figure 4.1-16 show the monthly statistics of the generated data using TPM method based on the generated 100 sequences of 50 years for Welgevonden station. The monthly averages show that it compares well with the historic data as it falls within the interquartile range of the generated data.

The maximum and standard deviation monthly values indicated that generally the historic values fall within the interquartile range of the box and whisker plots except for January, February, June, July, and August. For the month of February, the values are generally lower. The monthly percentiles however show a similar pattern to the historic values with a slight under estimation for the 98th and 99th percentiles.

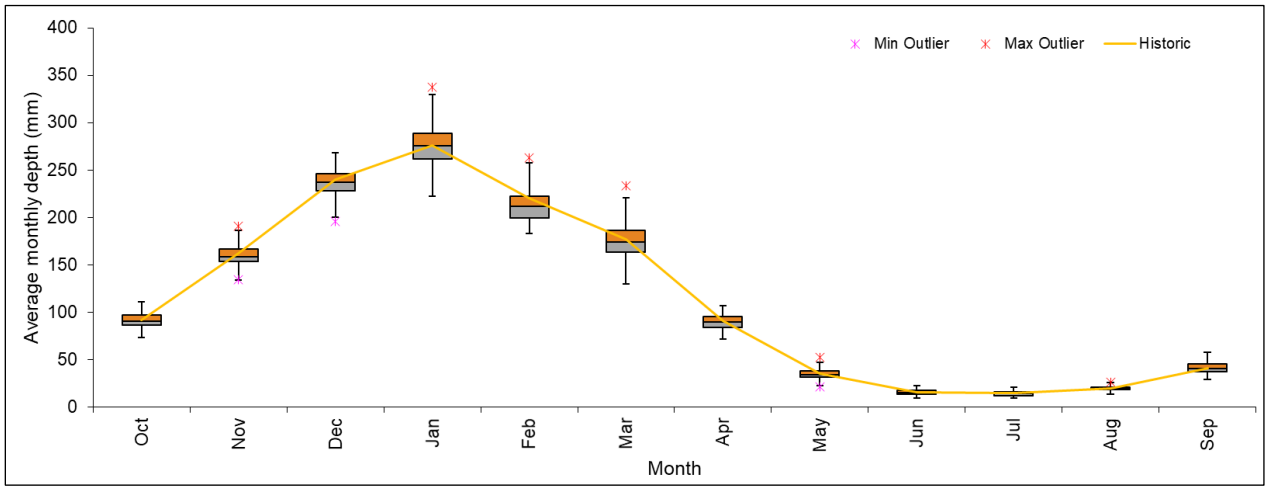


Figure 4.1-13: Averages of monthly data generated using TPM Method for Welgevonden station

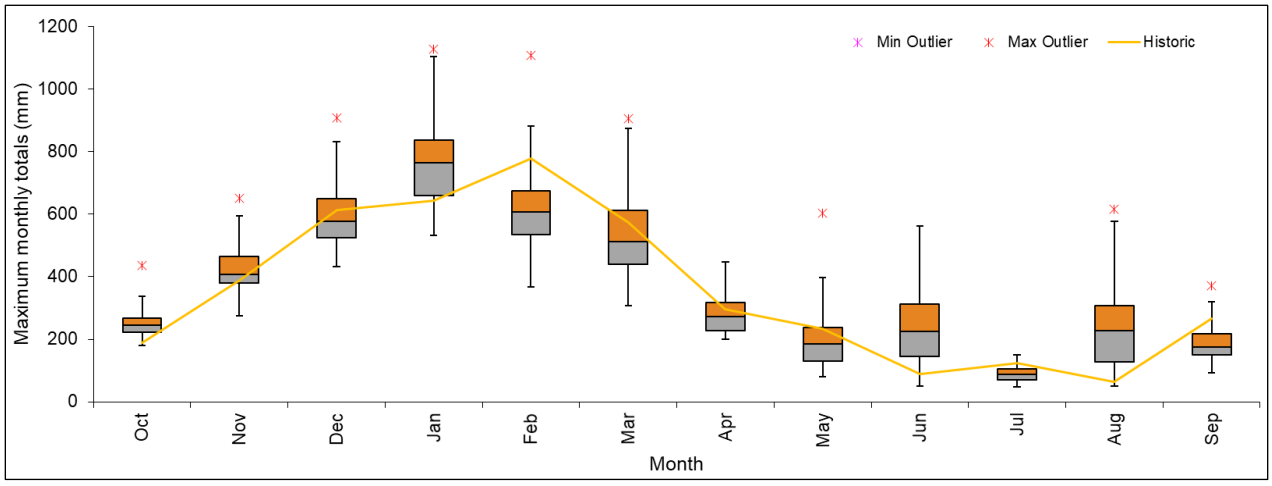


Figure 4.1-14: Maximum of monthly data generated using TPM Method for Welgevonden station

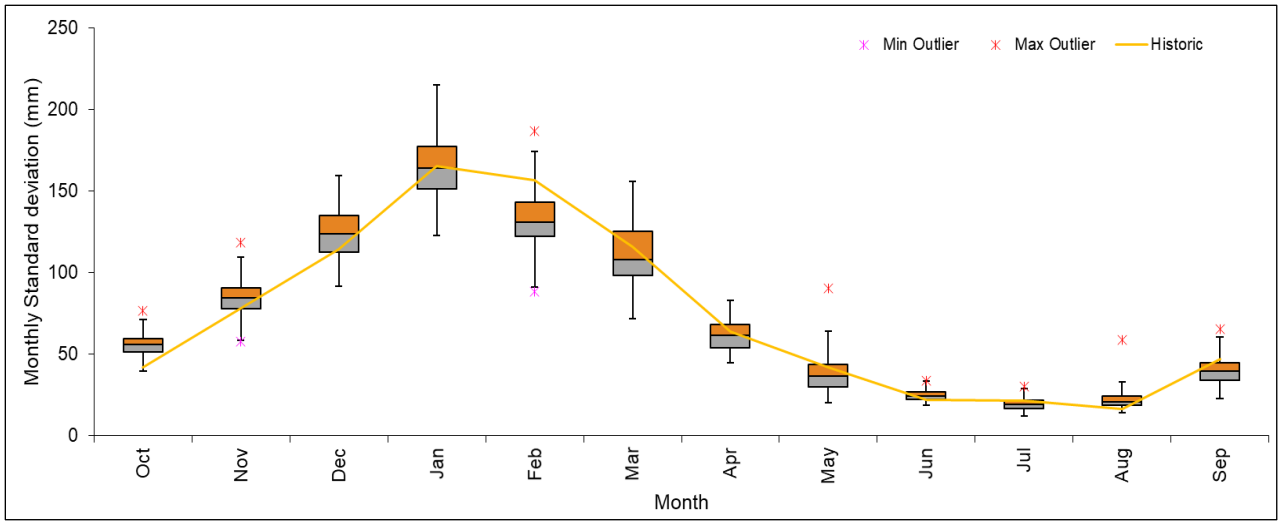


Figure 4.1-15: Standard deviation of monthly data generated using TPM Method for Welgevonden station

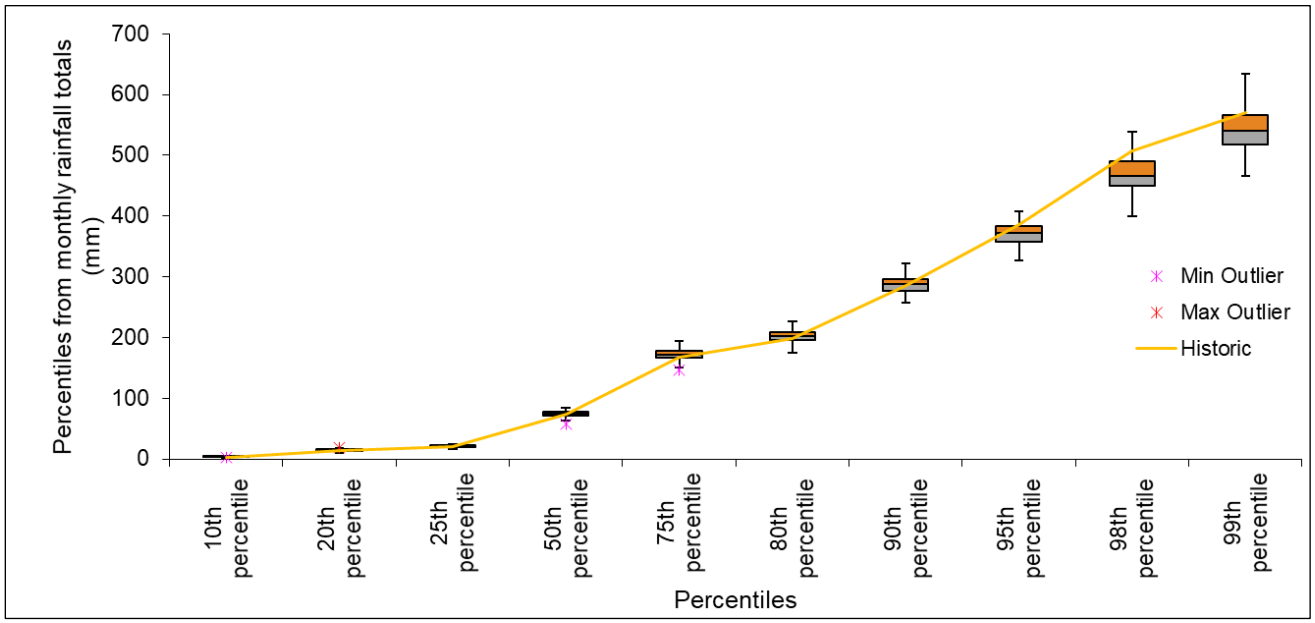


Figure 4.1-16: Percentiles of monthly data generated using TPM Method for Welgevonden station

Annual statistics of the generated data using TPM Method for Welgevonden rainfall station

Figure 4.1-17 and Figure 4.1-18 shows the rainfall distribution curve of the annual totals and percentiles respectively. The historic data mostly fall within the lower and upper extreme values of the box and whisker plot of the rainfall distribution curve. The annual percentiles shows that the generated annuals compare well with the historic data except for the higher percentiles.

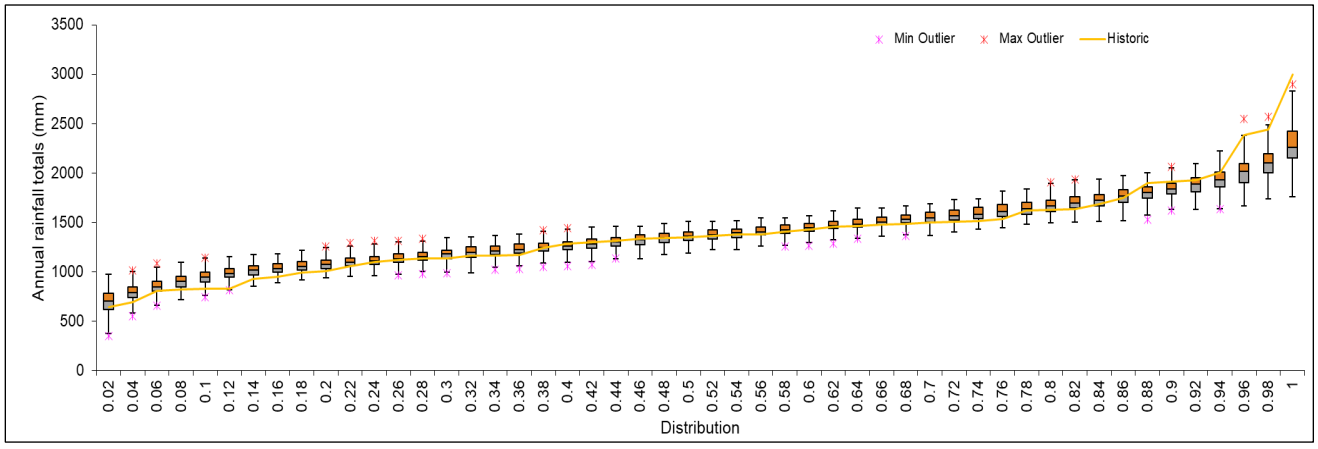


Figure 4.1-17: Rainfall distribution curve of the annual totals for data generated using TPM Method for Welgevonden station

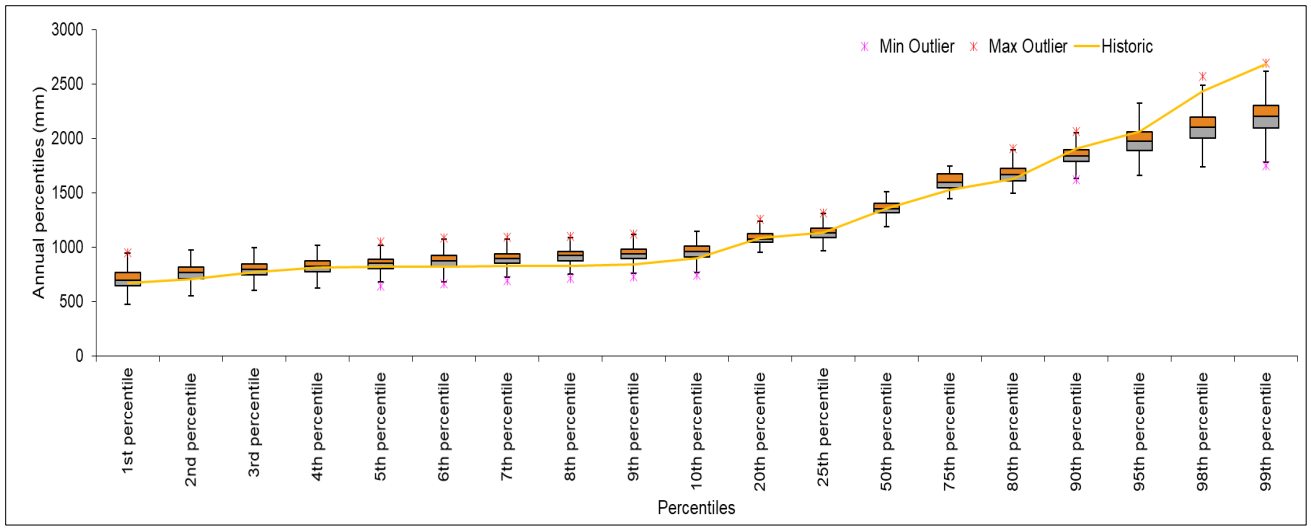


Figure 4.1-18: Annual percentiles of data generated using TPM Method for Welgevonden station

4.1.3. Nelshoogte Rainfall Station

Figure 4.1-19 shows the result of the calibration of the model for the Nelshoogte rainfall station. For below 30% probability, the uncorrected annual probability exceedance underestimated the annual rainfall precipitation, therefore the standard deviation was adjusted in order to fit the data closer to the recorded annual precipitation.

However, Figure 4.1-19 reveals that the corrected curve is worse than the uncorrected one beyond 1000 mm and an improvement to the method of correction may be required for some rainfall data.

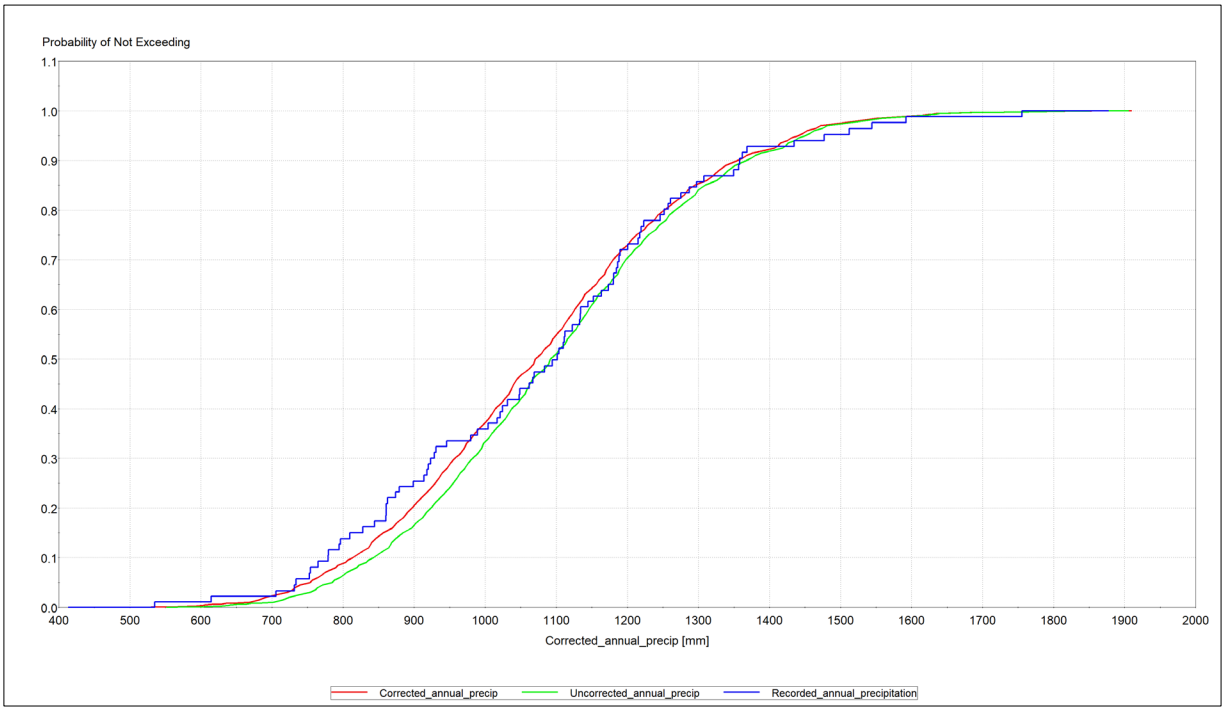


Figure 4.1-19:Comparison of the annual probability of exceedance for the generated annual precipitations to the recorded data for Nelshoogte rainfall station

Daily statistics of the generated data using TPM method for Nelshoogte rainfall station

Figure 4.1-20 and Figure 4.1-21 show the box and whisker plots of the rainfall distribution curve of daily maximum values and the daily percentiles from 100 sequences of 50 years daily generated data using the TPM method for Nelshoogte station.

The plots of the rainfall distribution curve show that for most of the years, the historic data fall within the interquartile range of the box and whisker plots of the generated data. For the daily percentiles, the historic values all fall within the interquartile range indicating that the method does preserve the statistics of the historic data.

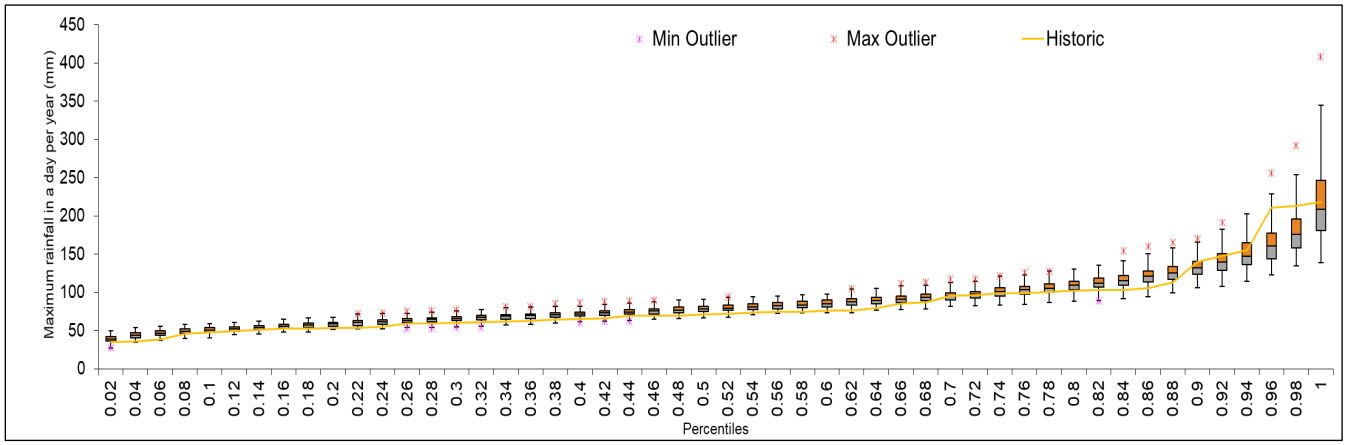


Figure 4.1-20: Nelshoogte station rainfall distribution curve from the daily maximum values using TPM Method

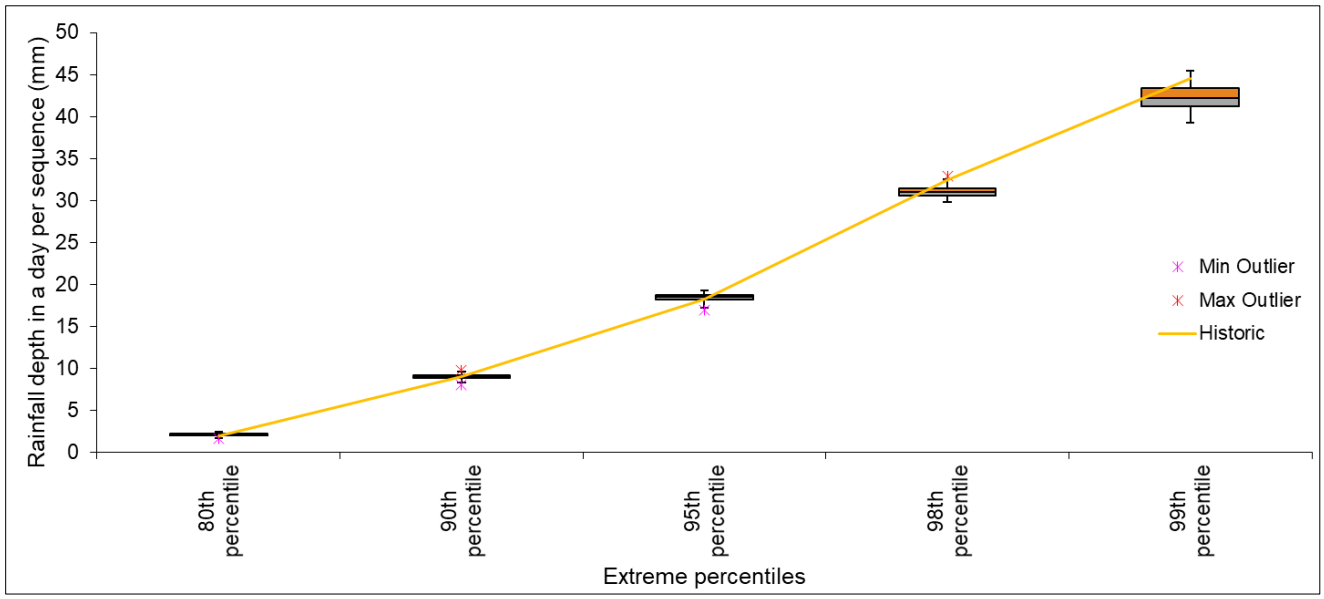


Figure 4.1-21: Percentiles of daily data generated using TPM Method for Nelshoogte station

Monthly statistics of the generated data using TPM method for Nelshoogte rainfall station

Figure 4.1-22 to Figure 4.1-25 show the monthly statistics of the generated data using TPM method based on the generated 100 sequences of 50 years for Nelshoogte station. The monthly averages shows that it compares well with the historic data as it falls within the interquartile range of the generated data. The maximum values show that generally the historic values fall within the lower and upper extreme values of the box and whisker plots with the historic value of February falling outside of the box and whisker plot, this has been common observation for the month of February in other stations as well. The standard deviation historic data mostly falls within the interquartile range. The monthly percentiles however show a similar pattern to the historic data.

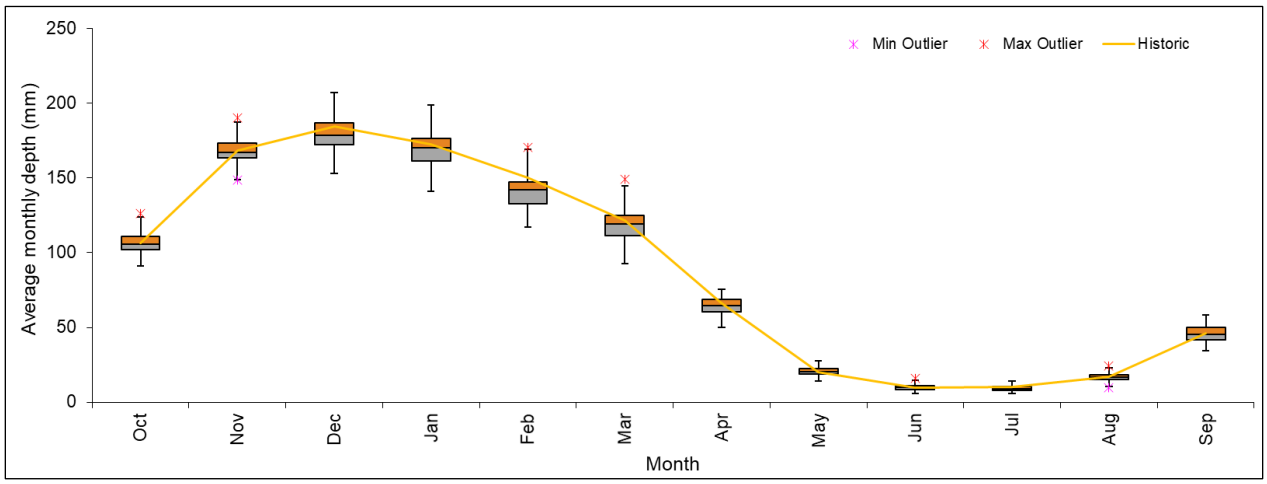


Figure 4.1-22: Averages of monthly data generated using TPM Method for Nelshoogte station

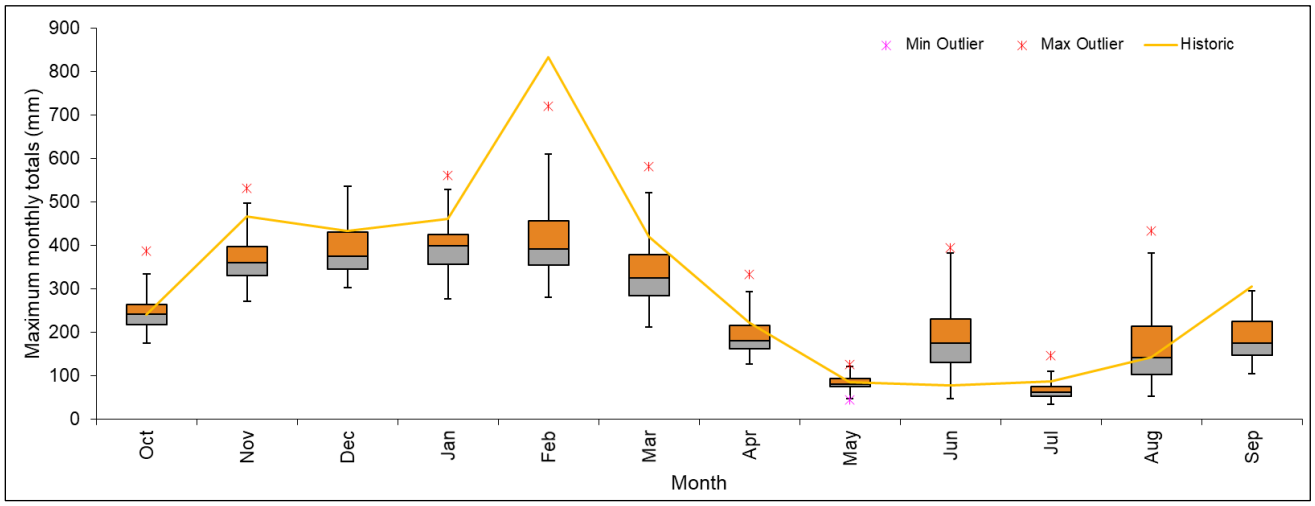


Figure 4.1-23: Maximum values of monthly data generated using TPM Method for Nelshoogte station

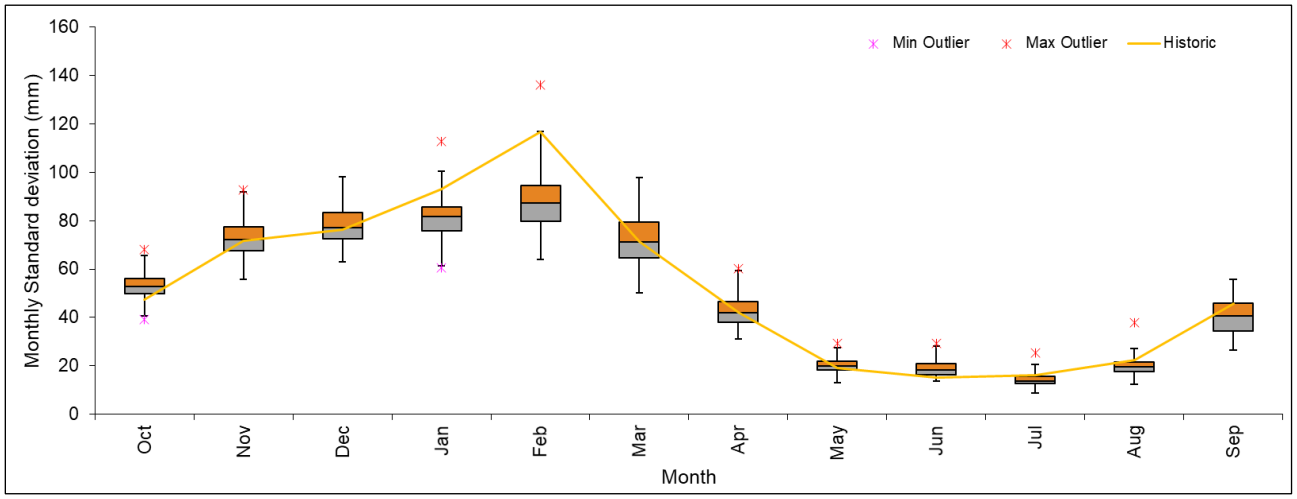


Figure 4.1-24: Standard deviation values of monthly data generated using TPM Method for Nelshoogte station

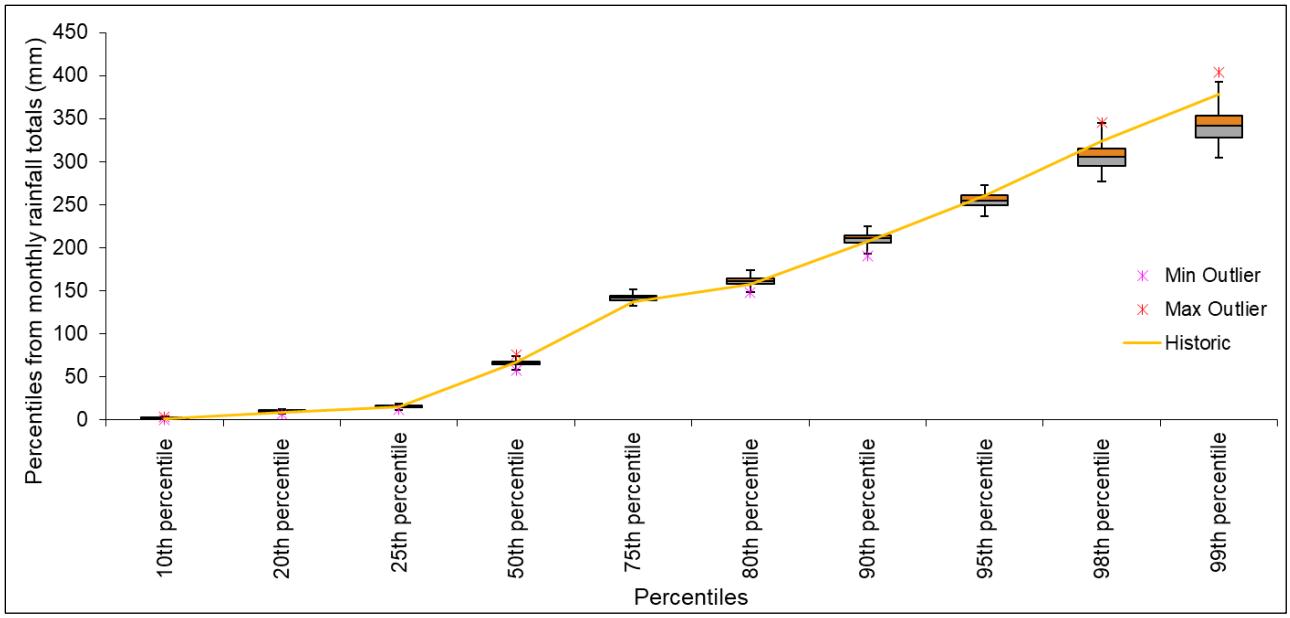


Figure 4.1-25: Percentiles of monthly data generated using TPM Method for Nelshoogte station

Annual statistics of the generated data using TPM method for Nelshoogte rainfall station

Figure 4.1-26 and Figure 4.1-27 shows the rainfall distribution curve of annual totals and percentiles respectively. From 40 % of the distribution to 100 % of the distribution curve, the historic data does fall within the lower and upper extreme values of the box and whisker plots. For less than 40 % of the distribution curve, the data falls within the lower and upper extreme of the box and whisker plots. The annual percentiles show that the generated annuals compared well with the historic data except for the higher percentiles.

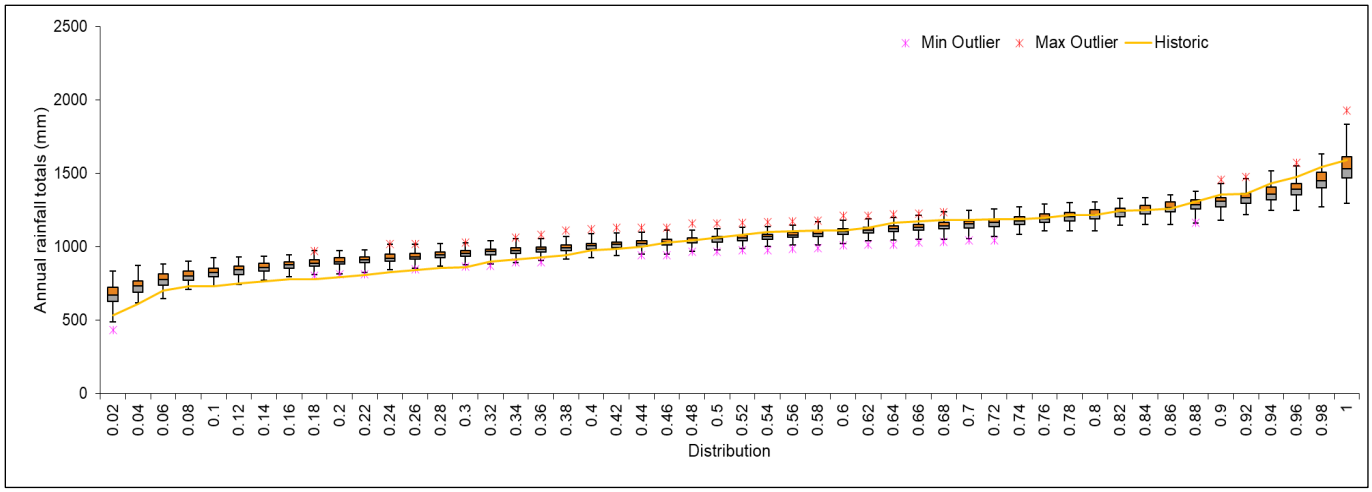


Figure 4.1-26: Rainfall distribution curve of the annual totals for data generated using TPM Method for Nelshoogte station

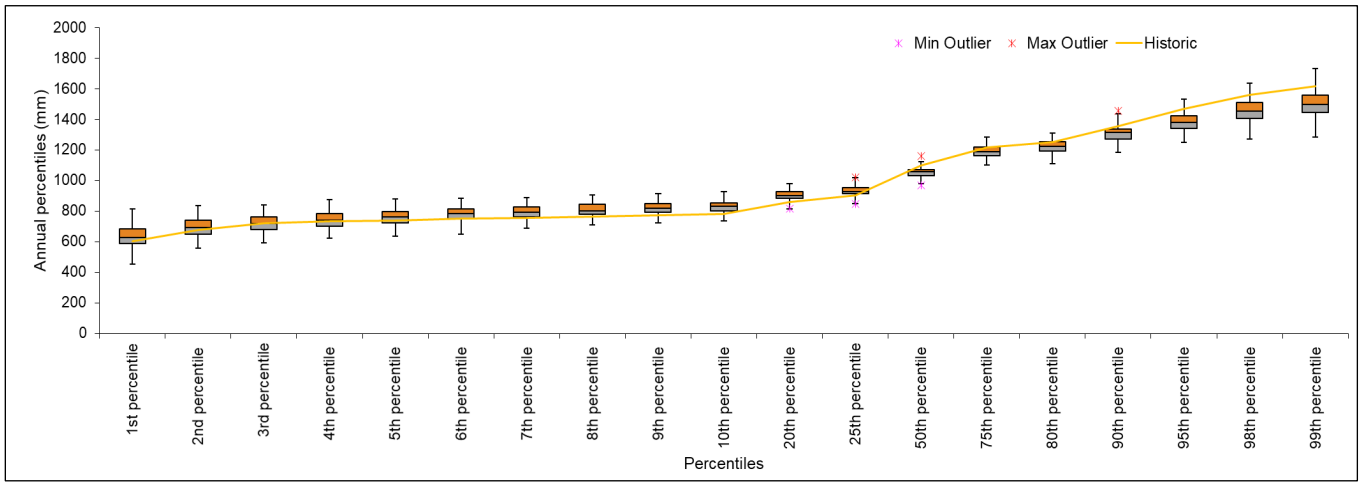


Figure 4.1-27: Annual percentiles of data generated using TPM Method for Nelshoogte station

4.2. Performance Evaluation of WGEN Method

4.2.1. Hebron Rainfall Station

Daily statistics of the generated data using WGEN method for Hebron rainfall station

Figure 4.2-1 and Figure 4.2-2 show the box and whisker plots of the rainfall distribution curve of daily maximum values and daily percentiles from 100 sequences of 50 years daily generated data using the WGEN Method for Hebron station. The rainfall distribution curve clearly shows that the daily maximum values generated using WGEN are lower than the historic values. The 80th and 99th percentiles of the historic values do not fall within the interquartile range of the generated data.

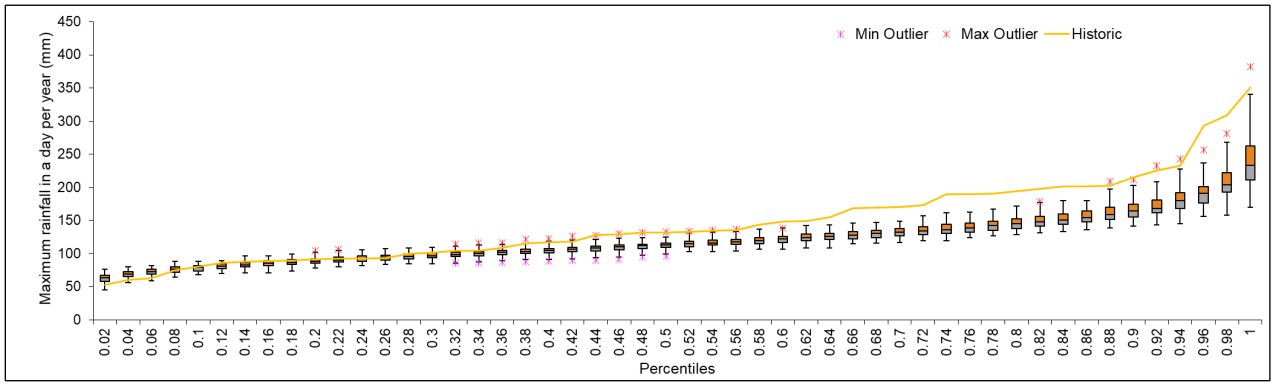


Figure 4.2-1: Rainfall distribution curve of the daily maximum values for data generated using WGEN Method for Hebron station

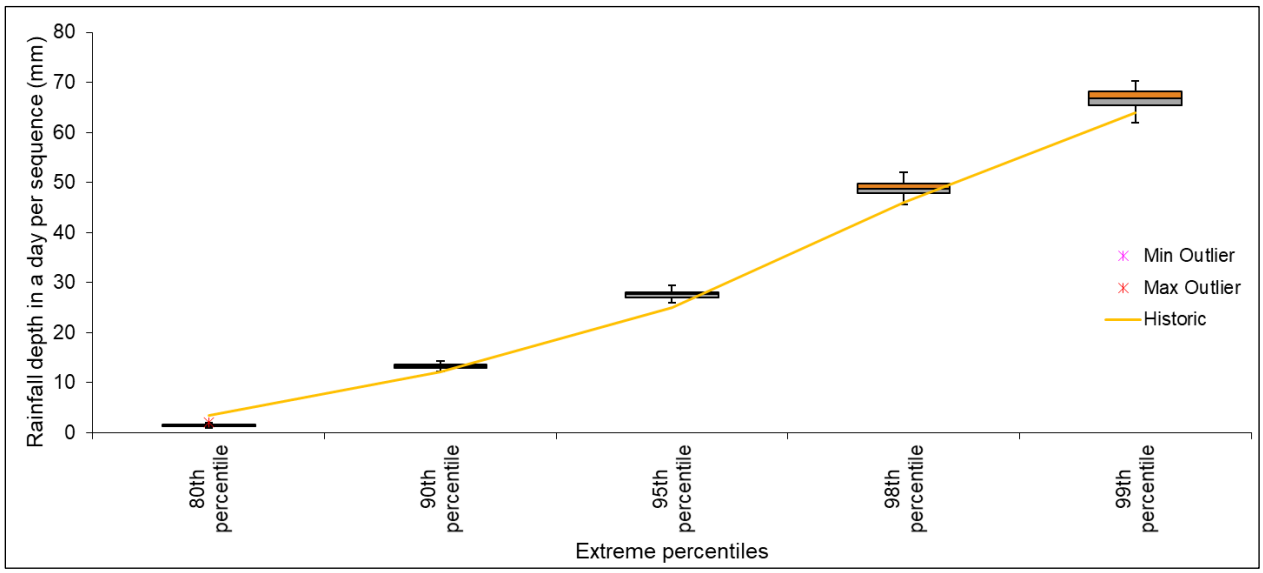


Figure 4.2-2: Percentiles of daily data generated using WGEN Method for Hebron station

Monthly statistics of the generated data using WGEN method for Hebron rainfall station

Figure 4.2-3 to Figure 4.2-6 show the monthly statistics of the generated data using WGEN Method based on the generated 100 sequences for 50 years for Hebron station. The monthly averages show that it compares relatively well with the historic data. The maximum values show that generally the historic values fall mostly within the upper extreme values of the box and whisker plots and the historic value of February falling outside of the box and whisker plot.

For the month of February, June, and August, it is be observed that the generated data of the maximum values is generally much higher than the historic values. The standard deviation historic values mostly fall within the interquartile range. The monthly percentiles however show a similar pattern to the historic data except for the 75th percentile.

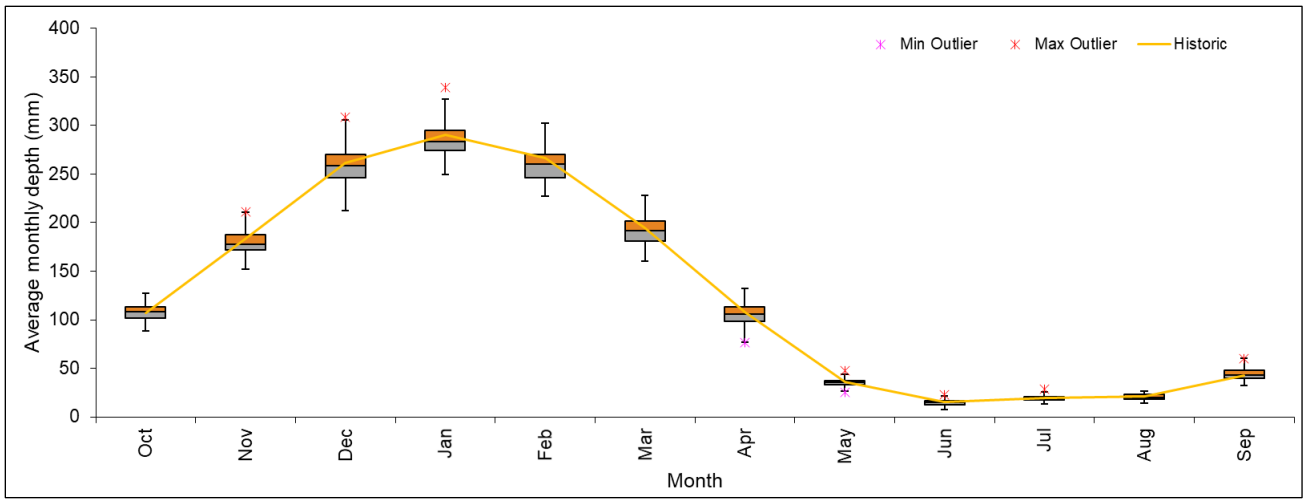


Figure 4.2-3: Averages of monthly data generated using WGEN Method for Hebron station

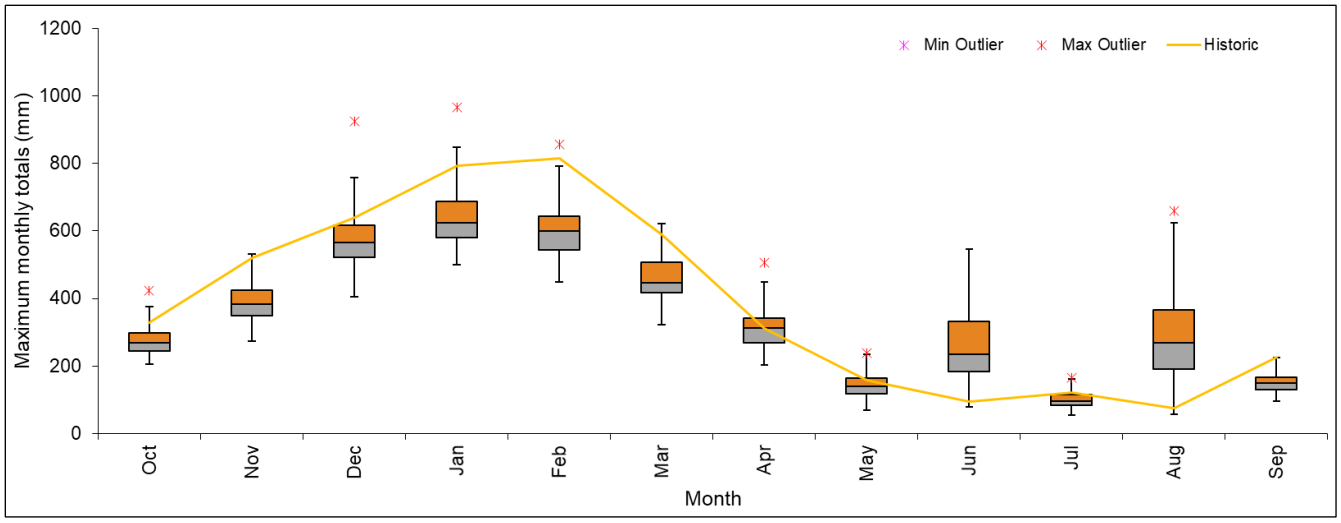


Figure 4.2-4: Maximum values of monthly data generated using WGEN Method for Hebron station

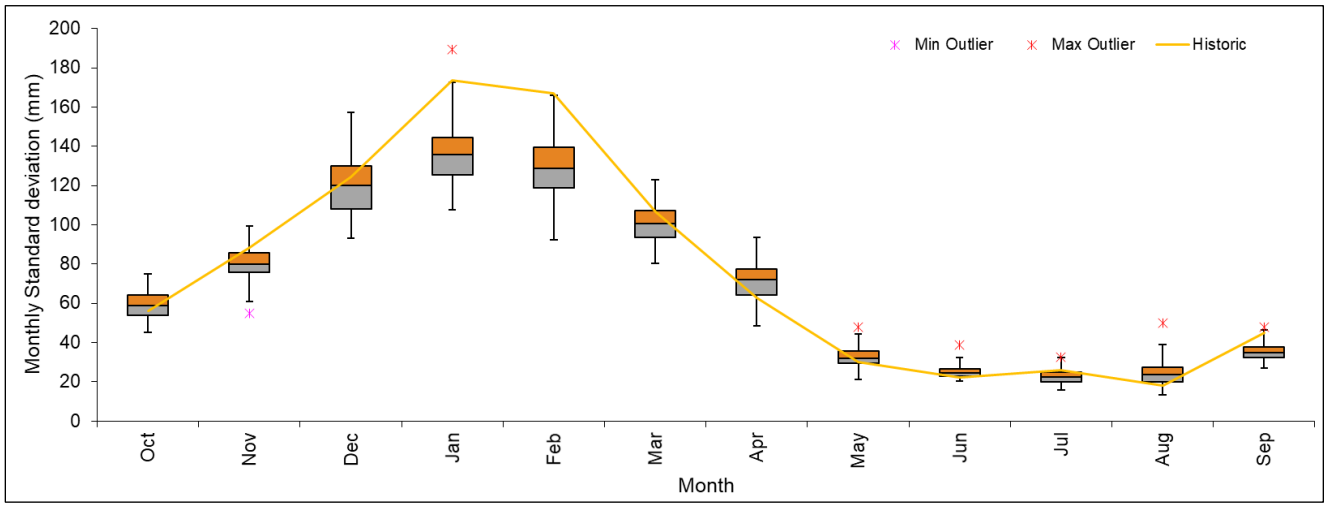


Figure 4.2-5: Standard deviation values of monthly data generated using WGEN Method for Hebron station

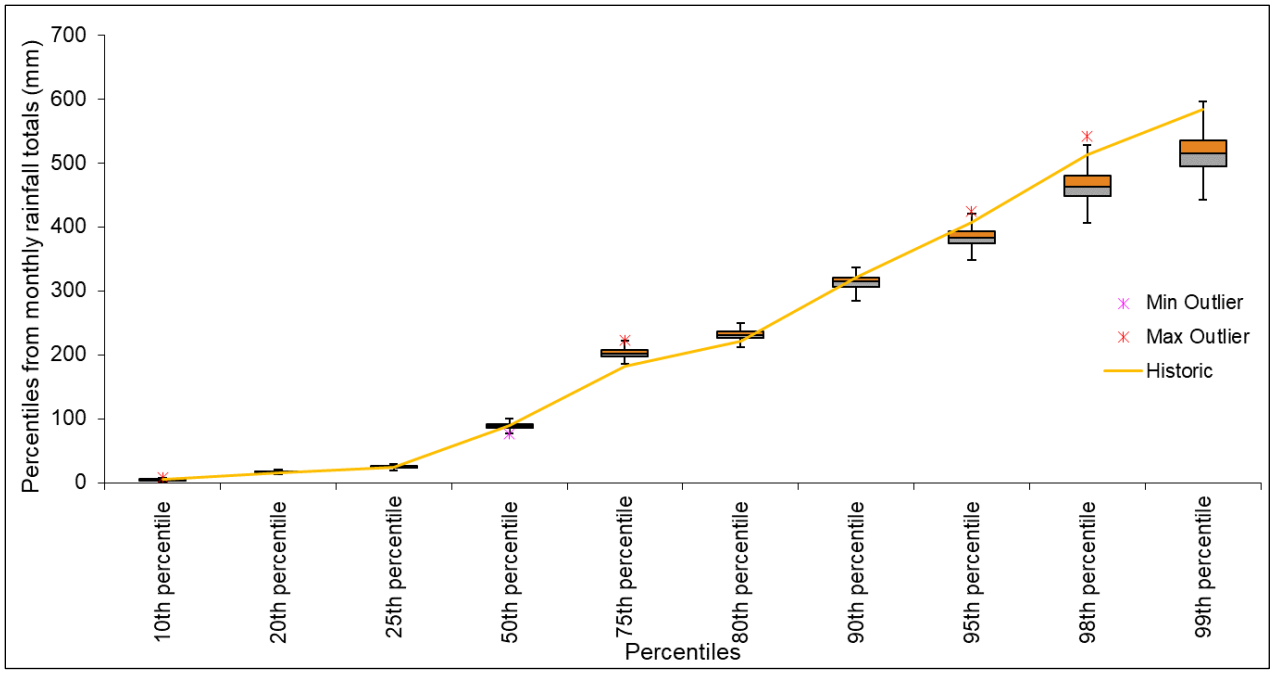


Figure 4.2-6: Percentiles of monthly data generated using WGEN Method for Hebron station

Annual statistics of the generated data using WGEN method for Hebron rainfall station

Figure 4.2-7 and Figure 4.2-8 show the rainfall distribution curve of the annual totals and percentiles respectively. Most of the historic data fall within the interquartile range except for the lower and higher distributions. The annual percentiles indicated that the method overestimated the lower percentiles and underestimated the higher percentiles.

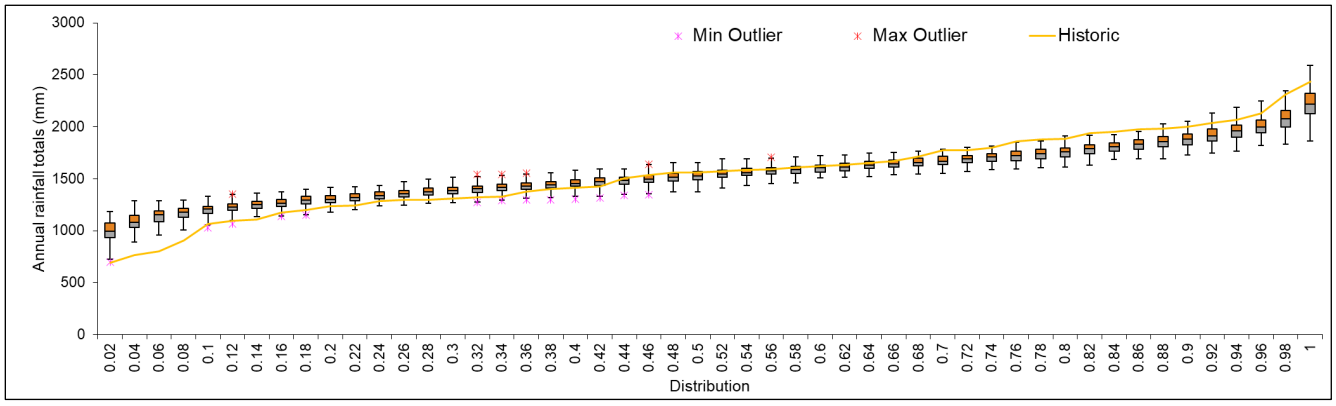


Figure 4.2-7: Rainfall distribution curve of the annual totals for data generated using WGEN Method for Hebron station

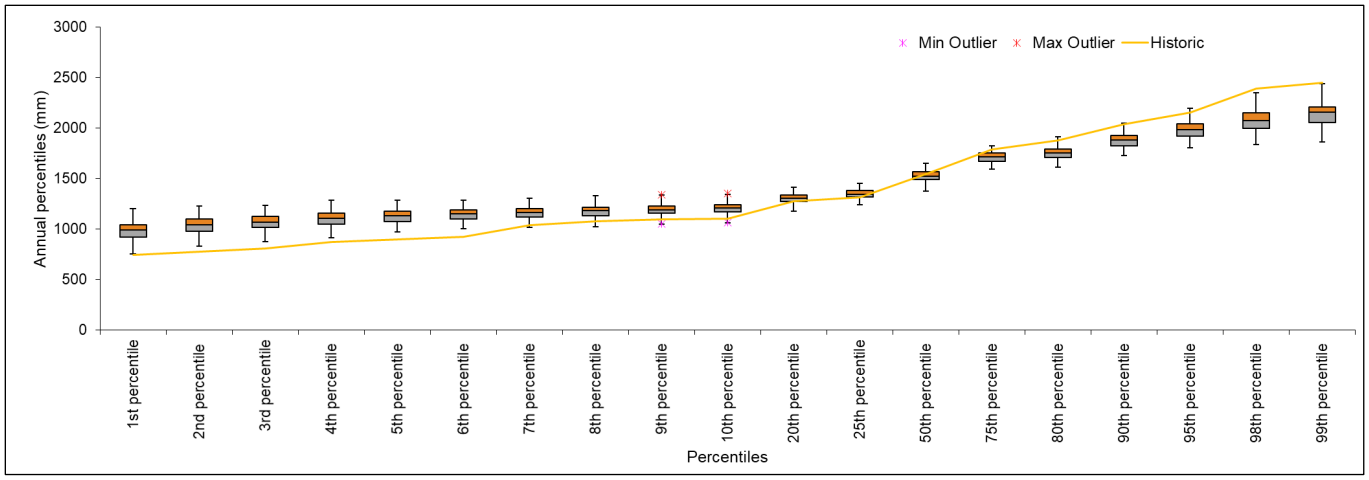


Figure 4.2-8: Annual percentiles of data generated using WGEN Method for Hebron station

4.2.2. Welgevonden Rainfall station

Daily statistics of the generated data using WGEN method for Welgevonden rainfall station

Figure 4.2-9 and Figure 4.2-10 show the box and whisker plots of the rainfall distribution curve of the daily maximum values and the daily percentiles from 100 sequences of 50 years daily generated data using the WGEN Method for Welgevonden station. For most of the box and whisker plots, the historic data falls outside the interquartile range especially for lower and higher distributions indicating that the daily maximum values generated using WGEN method does not correlate with the historic values. The historic value of the daily 98th percentile is lower than for the generated data.

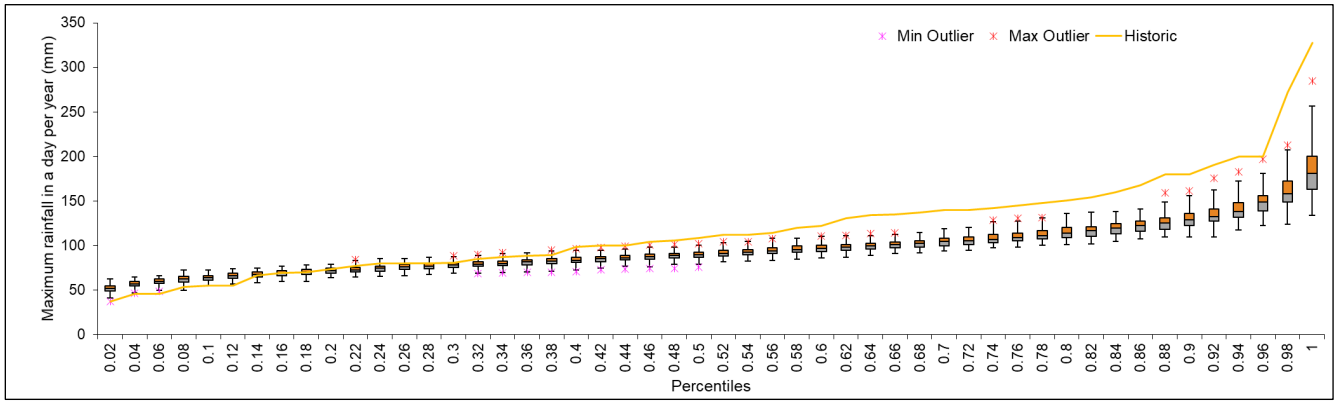


Figure 4.2-9: Rainfall distribution curve of the daily maximum values for data generated using WGEN Method for Welgevonden station

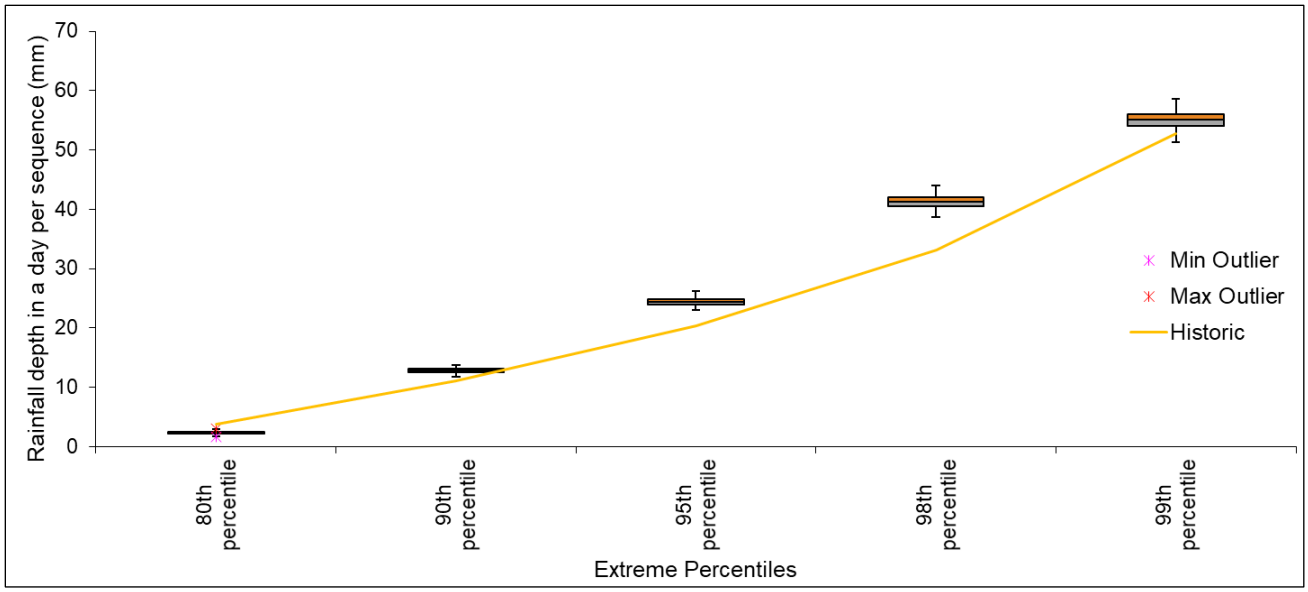


Figure 4.2-10: Percentiles of daily data generated using WGEN Method for Welgevonden station

Monthly statistics of the generated data using WGEN method for Welgevonden rainfall station

Figure 4.2-11 to Figure 4.2-14 show the monthly statistics of the generated data using WGEN Method based on the generated 100 sequences of 50 years for Welgevonden station. The monthly averages shows that it compares well with the historic data as the historic data falls within the interquartile range of the generated data. The maximum and standard deviation values show that in some months the generated data is lower and in other months higher than the historic data. For the month of February, the maximum values show lower generated values compared to the historic, which was also observed with the other stations. The monthly percentiles however show a similar pattern to the historic data except for the higher percentiles.

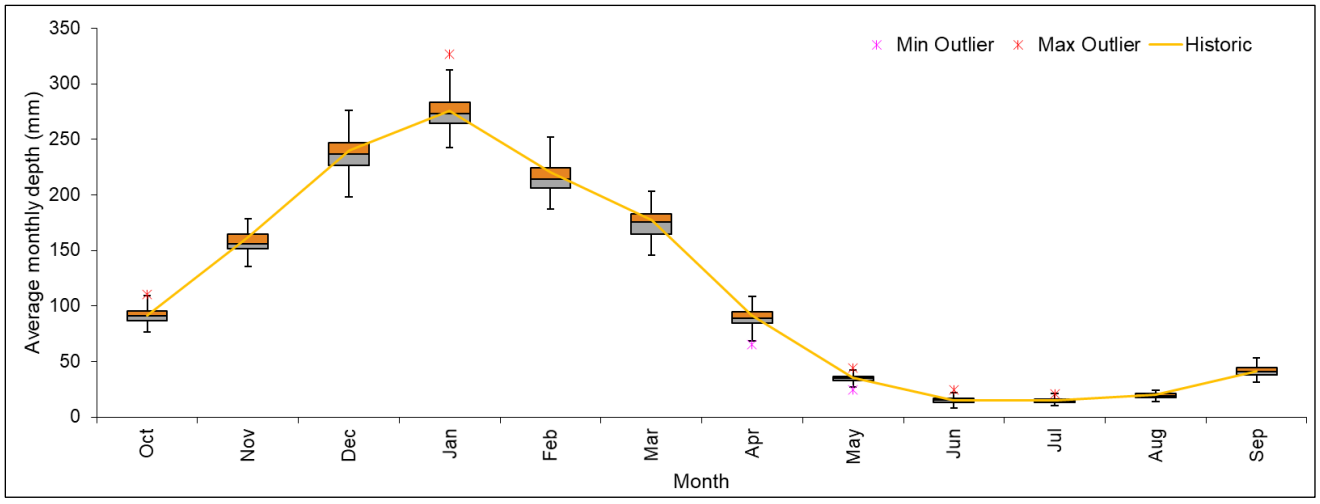


Figure 4.2-11: Averages of monthly data generated using WGEN Method for Welgevonden station

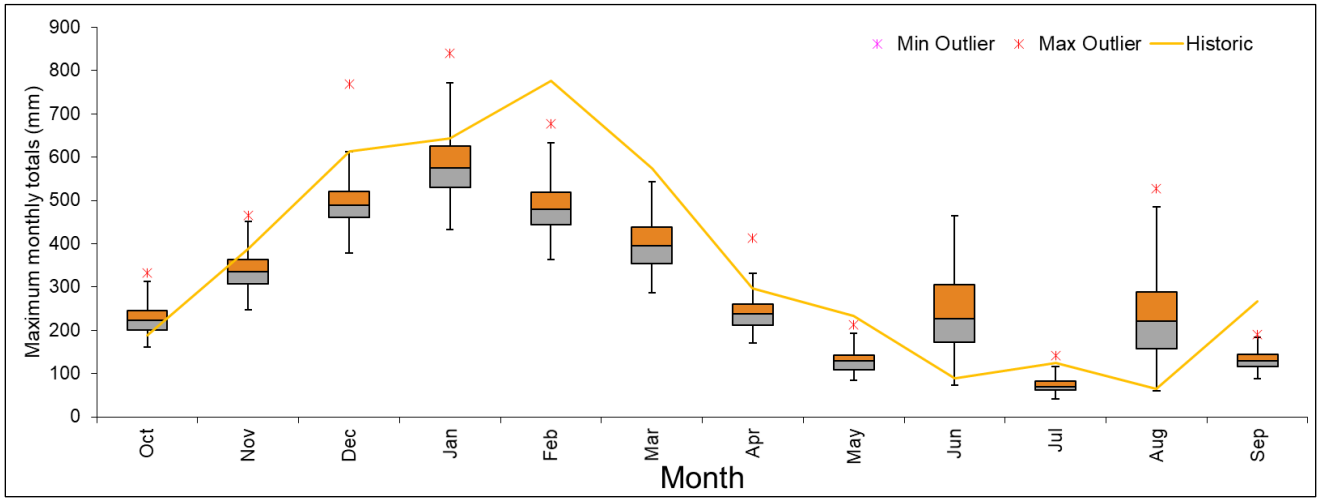


Figure 4.2-12: Maximum values of monthly data generated using WGEN Method for Welgevonden station

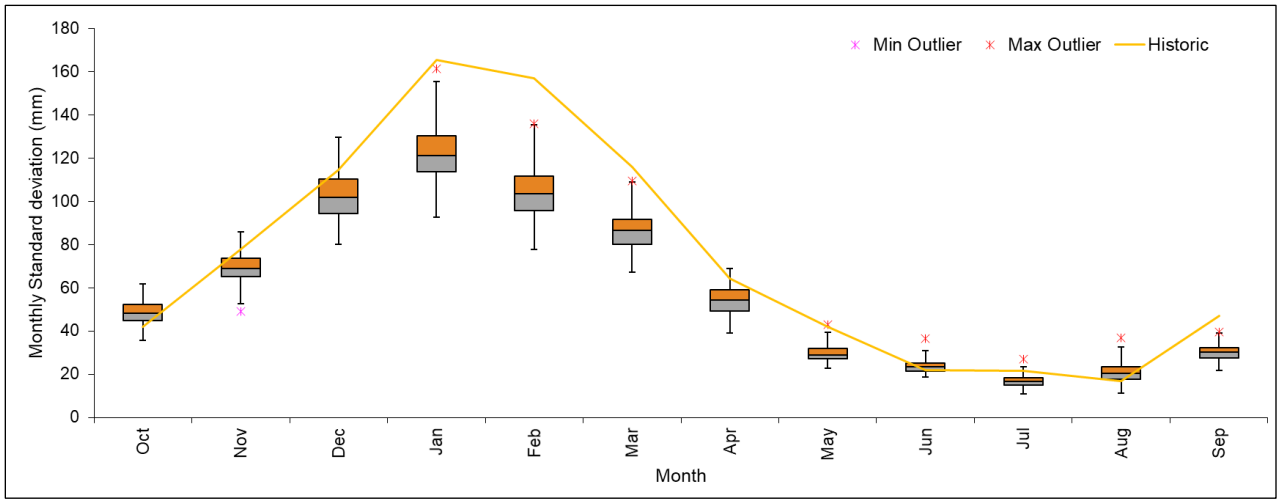


Figure 4.2-13: Standard deviation values of monthly data generated using WGEN Method for Welgevonden station

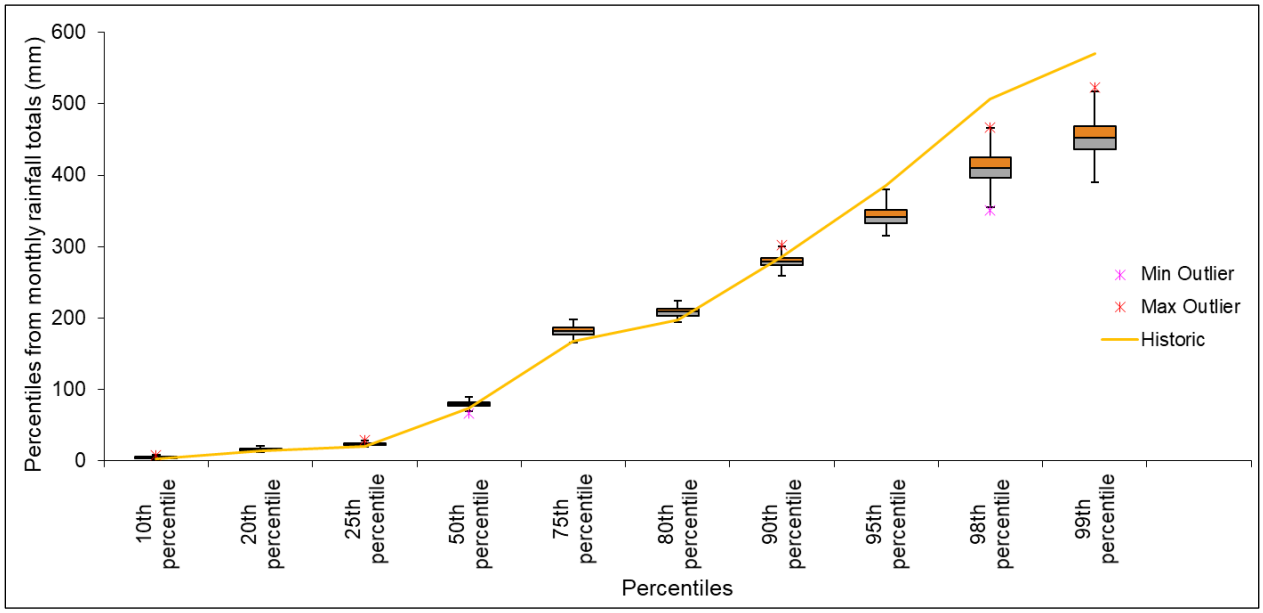


Figure 4.2-14: Percentiles of monthly data generated using WGEN Method for Welgevonden station

Annual statistics of the generated data using WGEN method for Welgevonden rainfall station

Figure 4.2-15 and Figure 4.2-16 show the rainfall distribution of the annual totals and percentiles respectively. The historic values fall within the interquartile range except for the lower and

higher distribution values. The annual percentiles shows that the method overestimated the lower percentiles and underestimated the higher percentiles with only 50th and 75th percentiles being the only ones falling within the interquartile range.

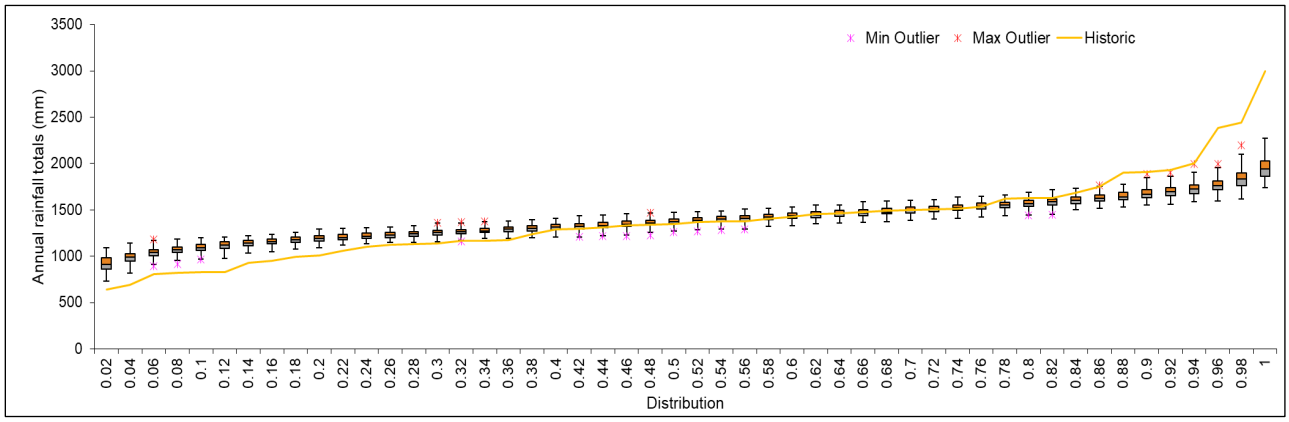


Figure 4.2-15: Annual totals of data generated using WGEN Method for Welgevonden station

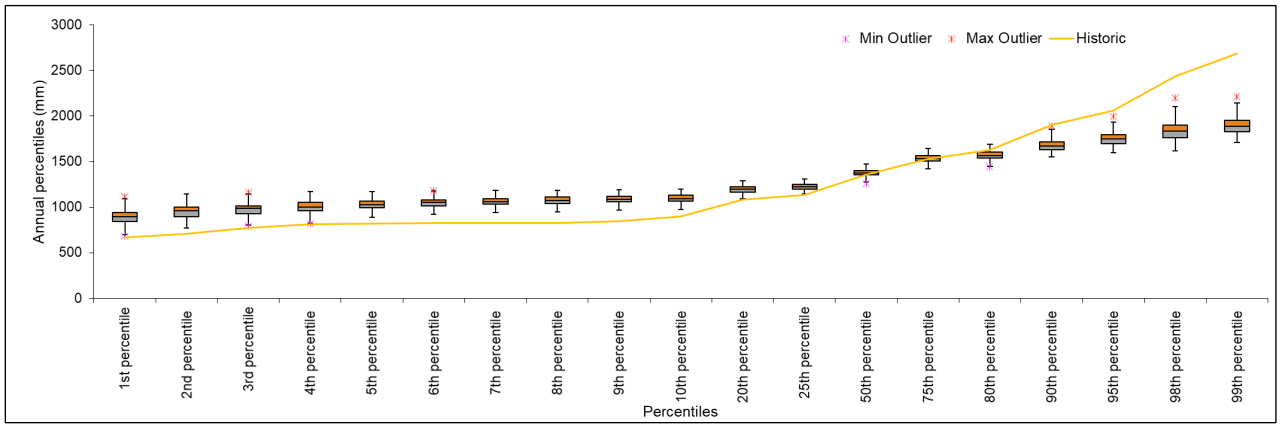


Figure 4.2-16: Annual percentiles of data generated using WGEN Method for Welgevonden station

4.2.3. Nelshoogte Rainfall station

Daily statistics of the generated data using WGEN method for Nelshoogte rainfall station

Figure 4.1-17 to Figure 4.2-18 show the box and whisker plots of the rainfall distribution curve for the daily maximum values and the daily percentiles from 100 sequences of 50 years daily generated data using the WGEN Method for Nelshoogte station. The historic data falls within the interquartile range except for the distribution values above 68 %. Historic daily percentiles fall within the interquartile range except for the 99th of percentile.

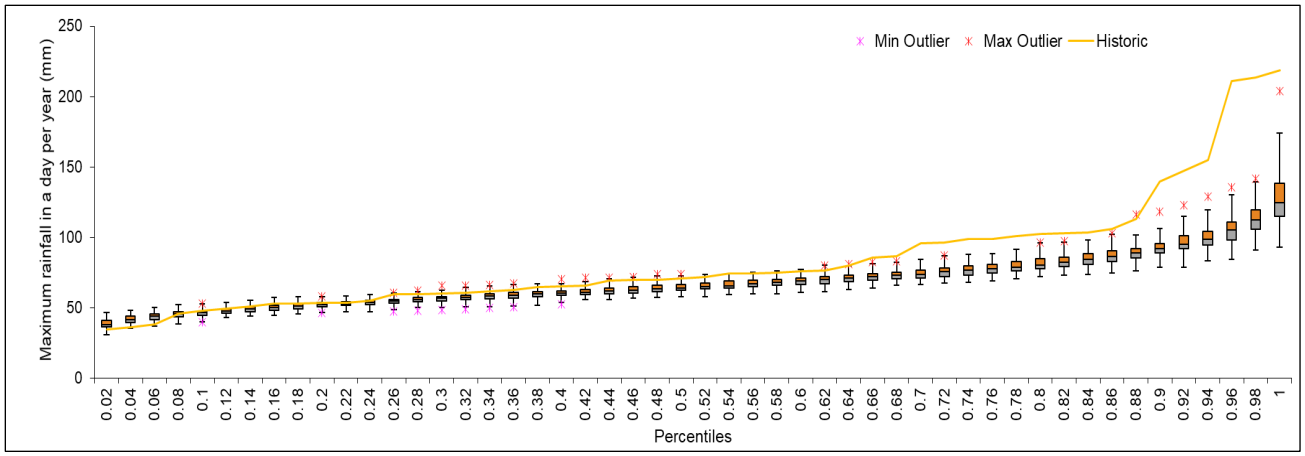


Figure 4.2-17: Nelshoogte station rainfall distribution curve from the daily maximum values using WGEN Method

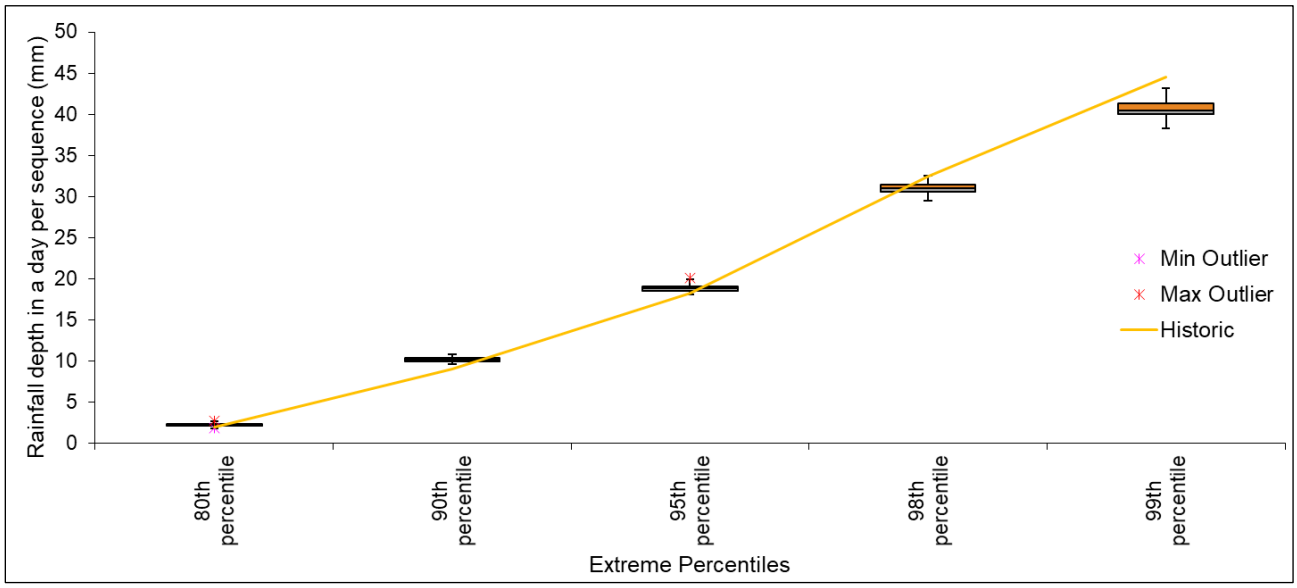


Figure 4.2-18: Percentiles of daily data generated using WGEN Method for Nelshoogte station

Monthly statistics of the generated data using WGEN method for Nelshoogte rainfall station

Figure 4.2-19 to Figure 4.2-22 show the monthly statistics of the generated data using WGEN Method based on the generated 100 sequences for 50 years for Nelshoogte station. The monthly averages show that it compares well with the historic data because the historic data falls within the interquartile range of the generated data. The maximum and standard deviation values show that in some months the generated data is lower and in other months higher than the historic data, with the month of February highlighting that the method underestimated the values. The

monthly percentiles however show a similar pattern to the historic data except for the higher percentiles whereby the historic values are higher than the range.

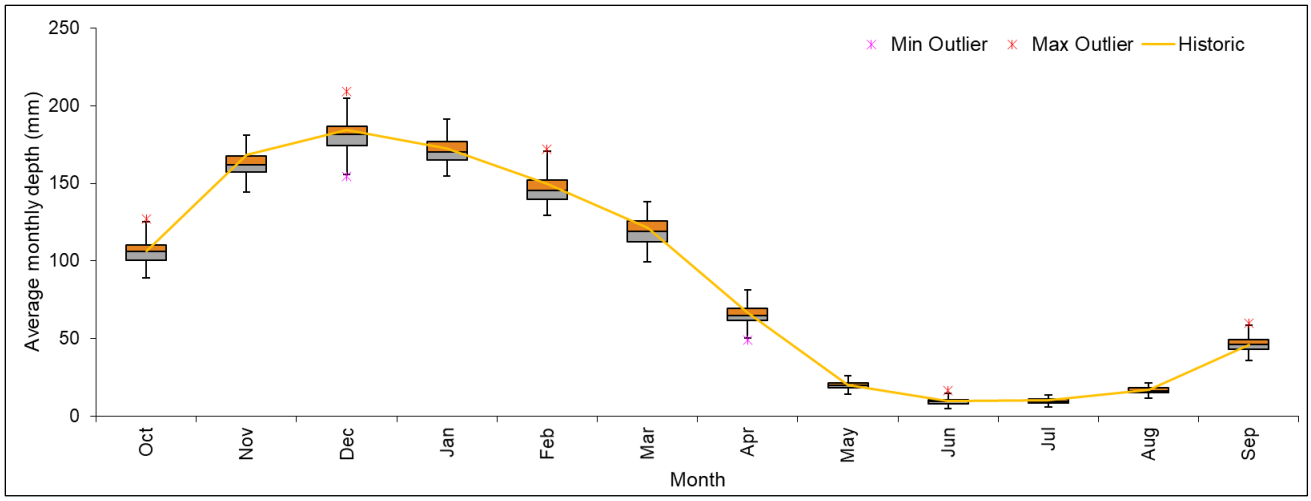


Figure 4.2-19: Averages of monthly data generated using WGEN Method for Nelshoogte station

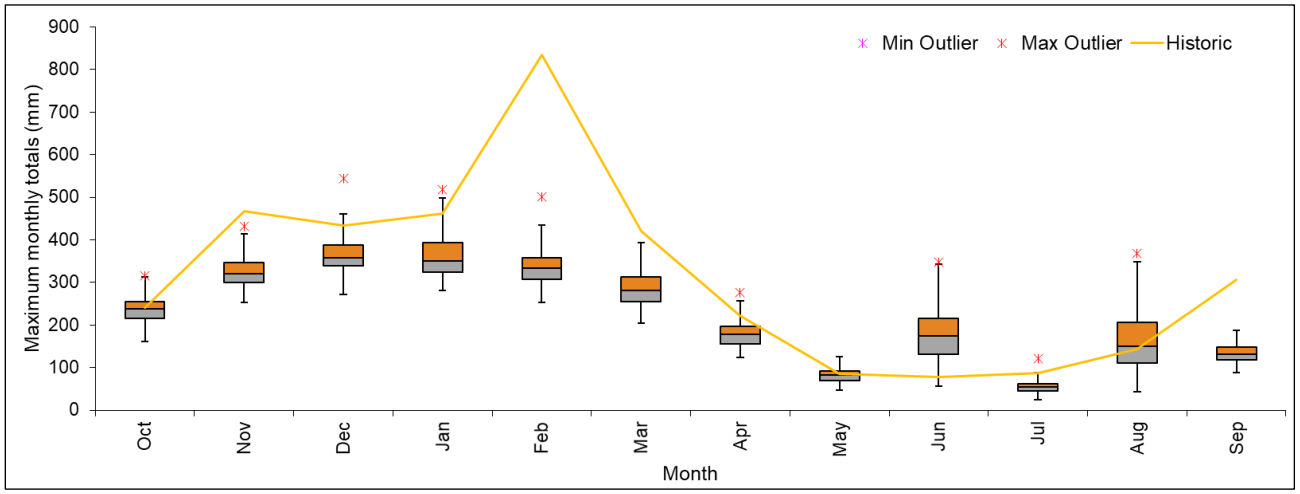


Figure 4.2-20: Maximum values of monthly data generated using WGEN Method for Nelshoogte station

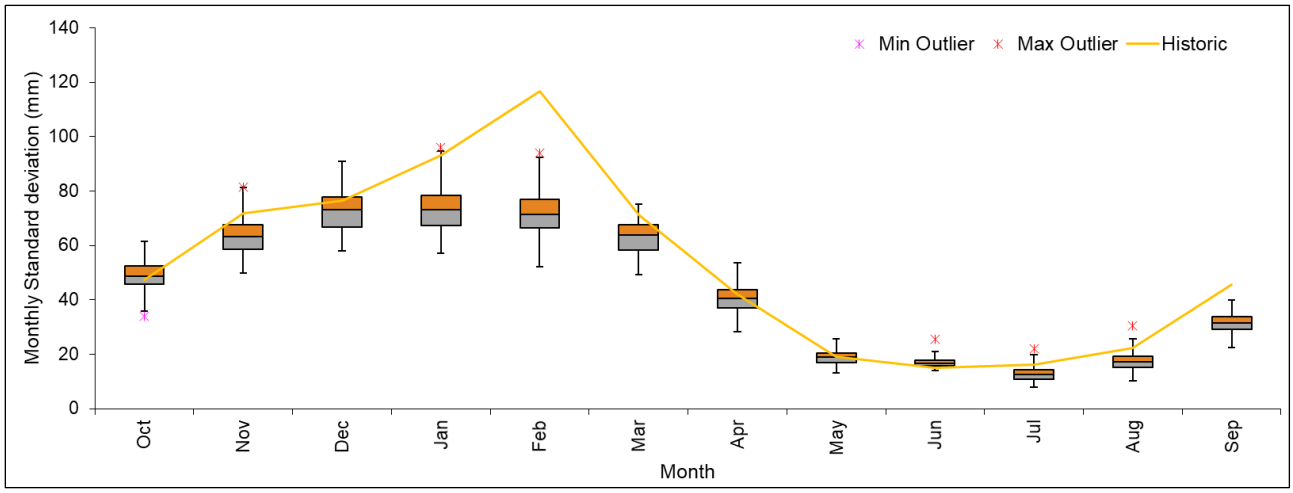


Figure 4.2-21: Standard deviation of monthly data generated using WGEN Method for Nelshoogte station

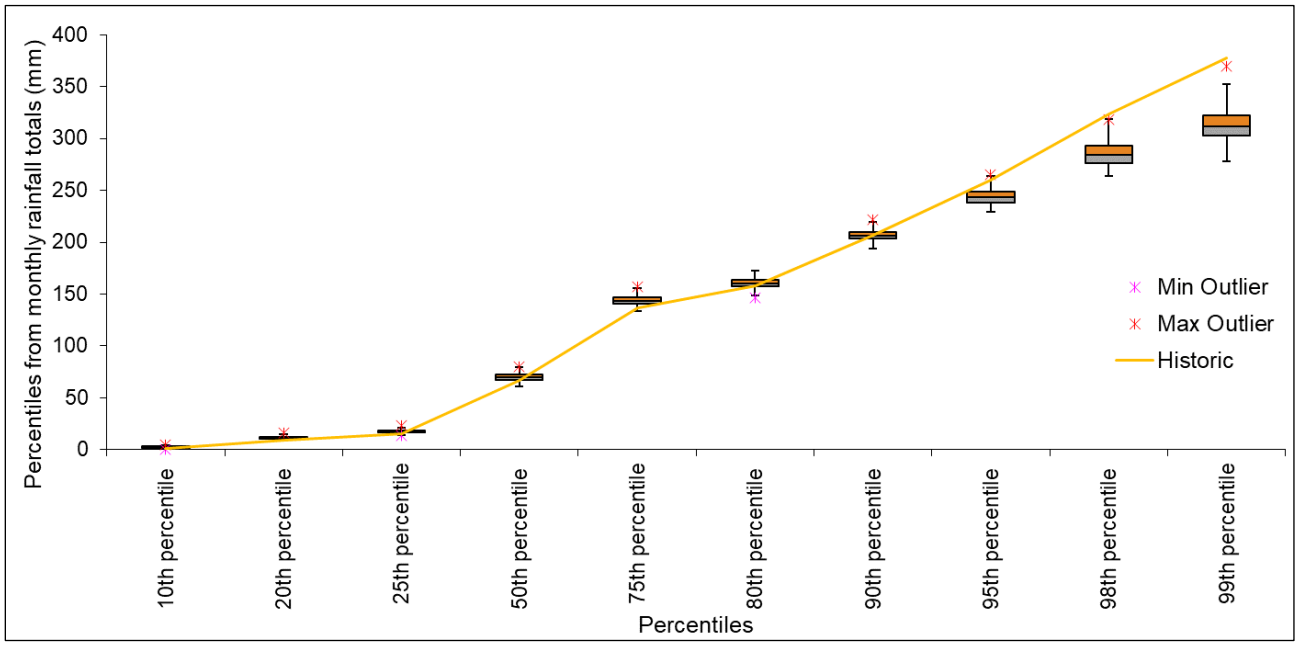


Figure 4.2-22: Percentiles of monthly data generated using WGEN Method for Nelshoogte station

Annual statistics of the generated data using WGEN method for Nelshoogte rainfall station

Figure 4.2-23 and Figure 4.2-24 show the rainfall distribution curve of the annual totals and percentiles respectively. The percentiles show that the method overestimated the lower

percentiles and underestimated the higher percentiles with only 2nd, 50th and 75th percentiles being the only ones falling within the interquartile range.

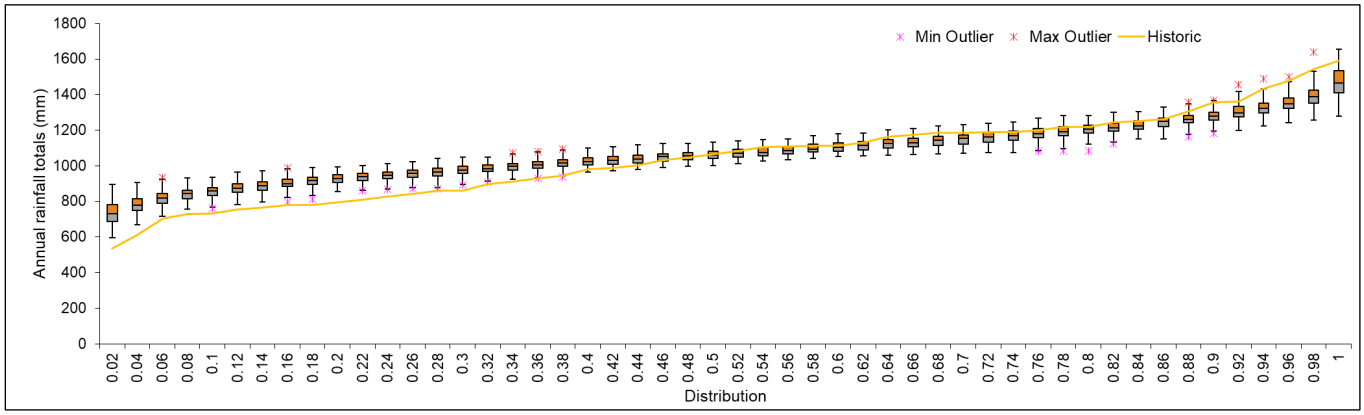


Figure 4.2-23: Rainfall distribution curve of the annual totals for data generated using WGEN Method for Nelshoogte station

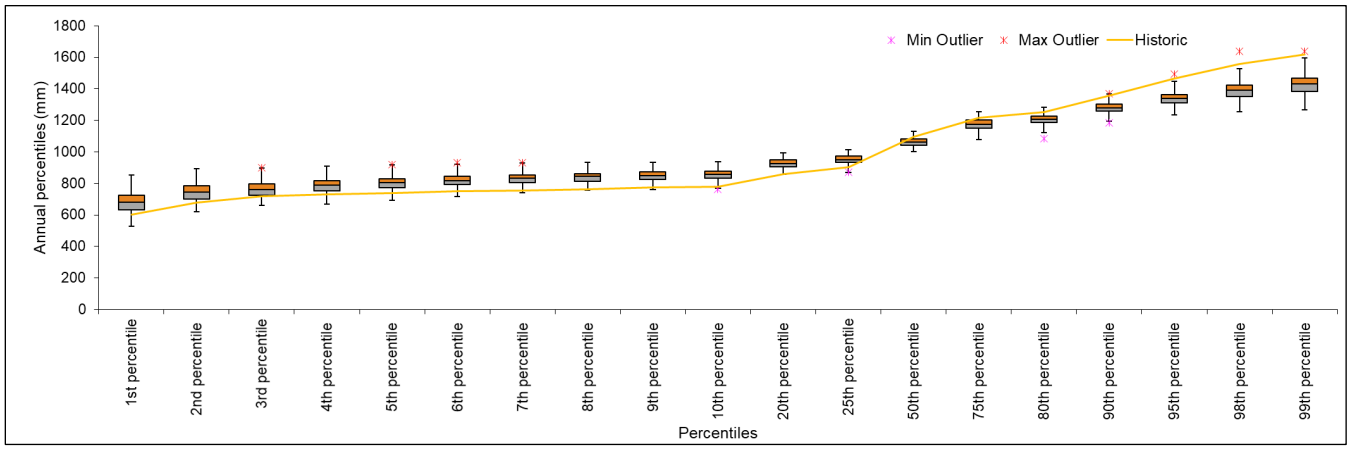


Figure 4.2-24: Annual percentiles of data generated using WGEN Method for Nelshoogte station

4.3. Performance Evaluation of VLB Method

4.3.1. Hebron Rainfall station

Daily statistics of the generated data using VLB Method for Hebron rainfall station

Figure 4.3-1 and Figure 4.3-2 show the box and whisker plots of the rainfall distribution curve of the maximum values and the daily percentiles from 100 sequences of 50 years daily generated data using the VLB Method for Hebron station.

The maximum values of the historic data mostly fall within the interquartile range of rainfall distribution box and whisker plots. The historic values are lower than the range for the generated data for the 98th and 99th percentiles.

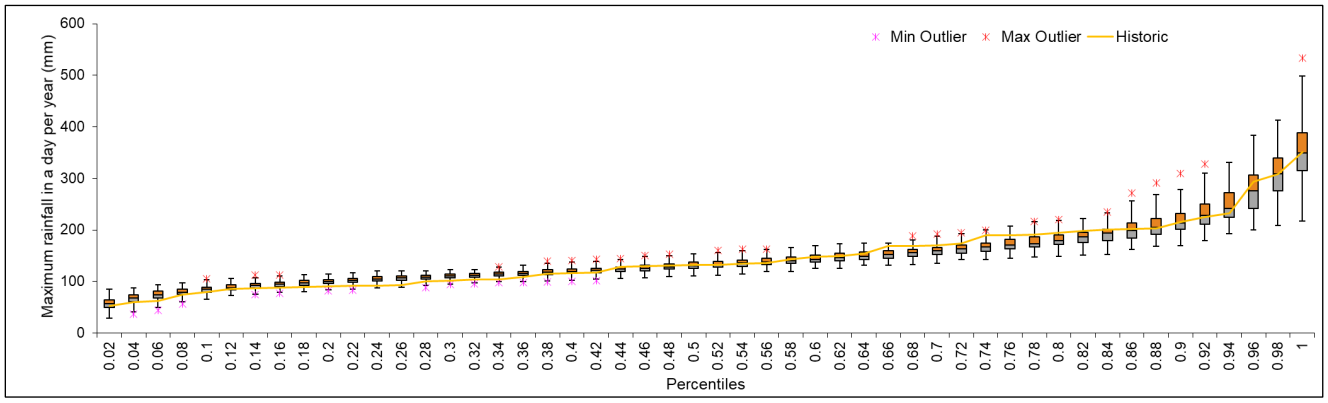


Figure 4.3-1: Rainfall distribution curve of the daily maximum values for data generated using VLB Method for Hebron station

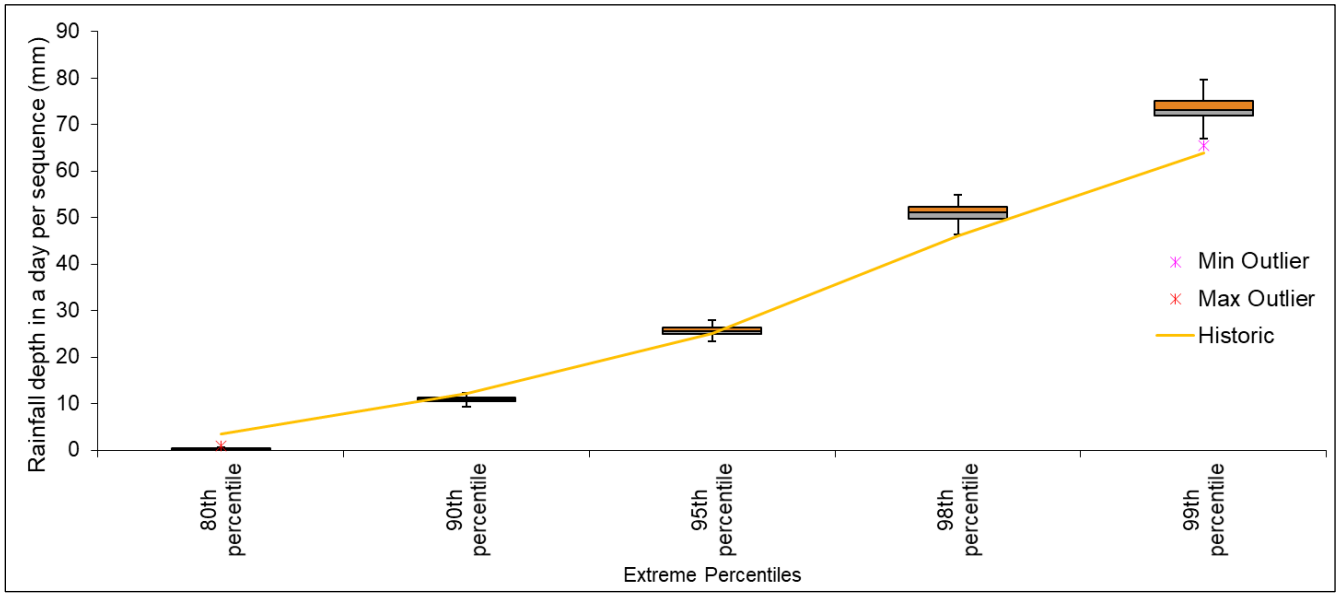


Figure 4.3-2: Percentiles of daily data generated using VLB Method for Hebron station

Monthly statistics of the generated data using VLB Method for Hebron rainfall station

Figure 4.3-3 to Figure 4.3-6 show the monthly statistics of the generated data using VLB Method based on the generated 100 sequences for 50 years for Hebron station.

The monthly averages show that it compares well with the historic data because the historic data falls within the interquartile range of the generated data except February. Most of the maximum values for the historic data falls within the interquartile range as well except for February, July, and August. The monthly percentiles compare relatively well to the historic values.

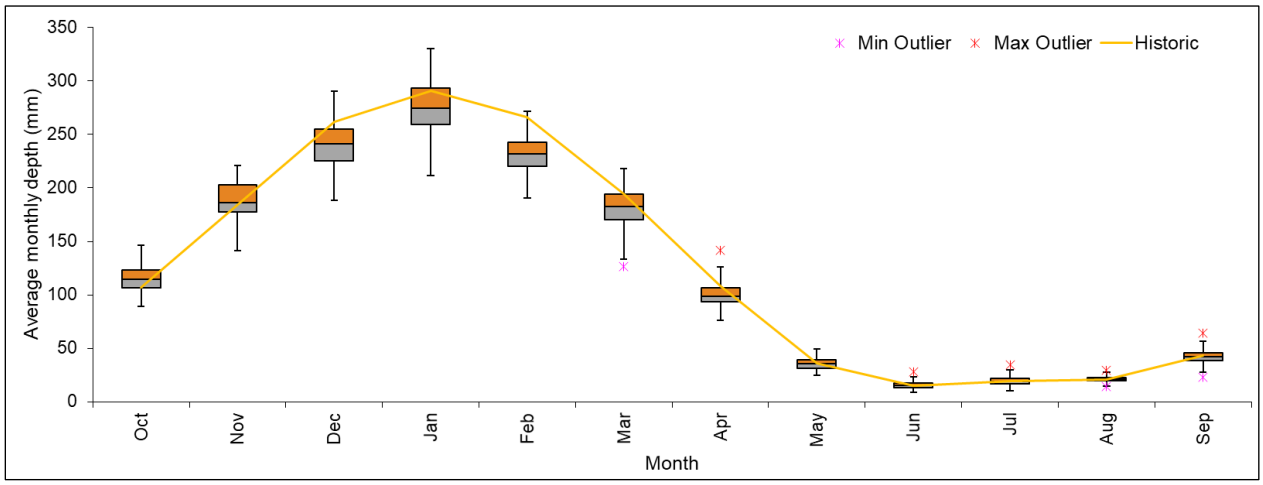


Figure 4.3-3: Averages of monthly data generated using VLB Method for Hebron station

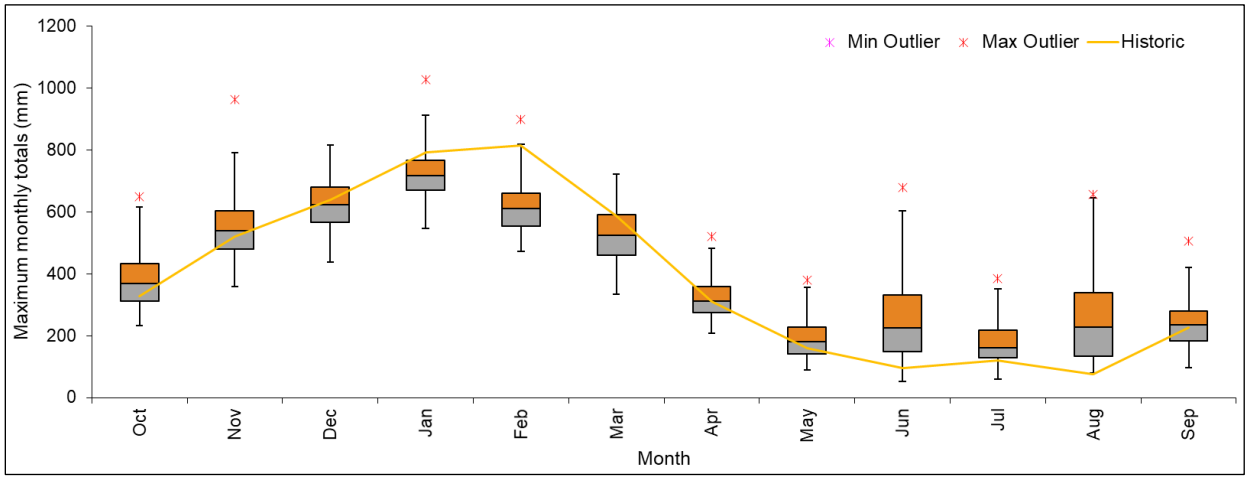


Figure 4.3-4: Maximum values of monthly data generated using VLB Method for Hebron station

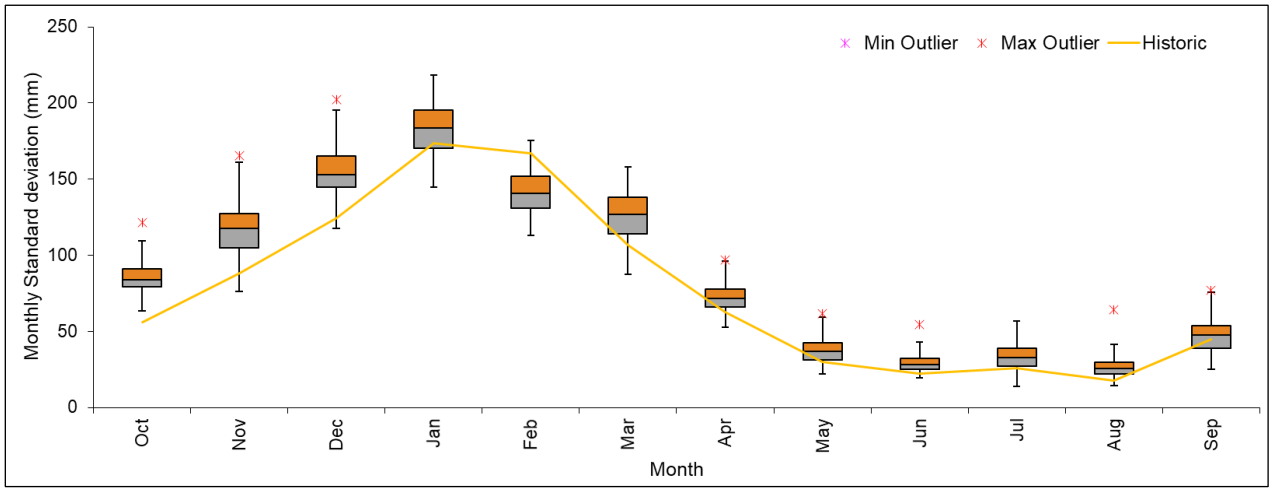


Figure 4.3-5: Standard deviation values of monthly data generated using VLB Method for Hebron station

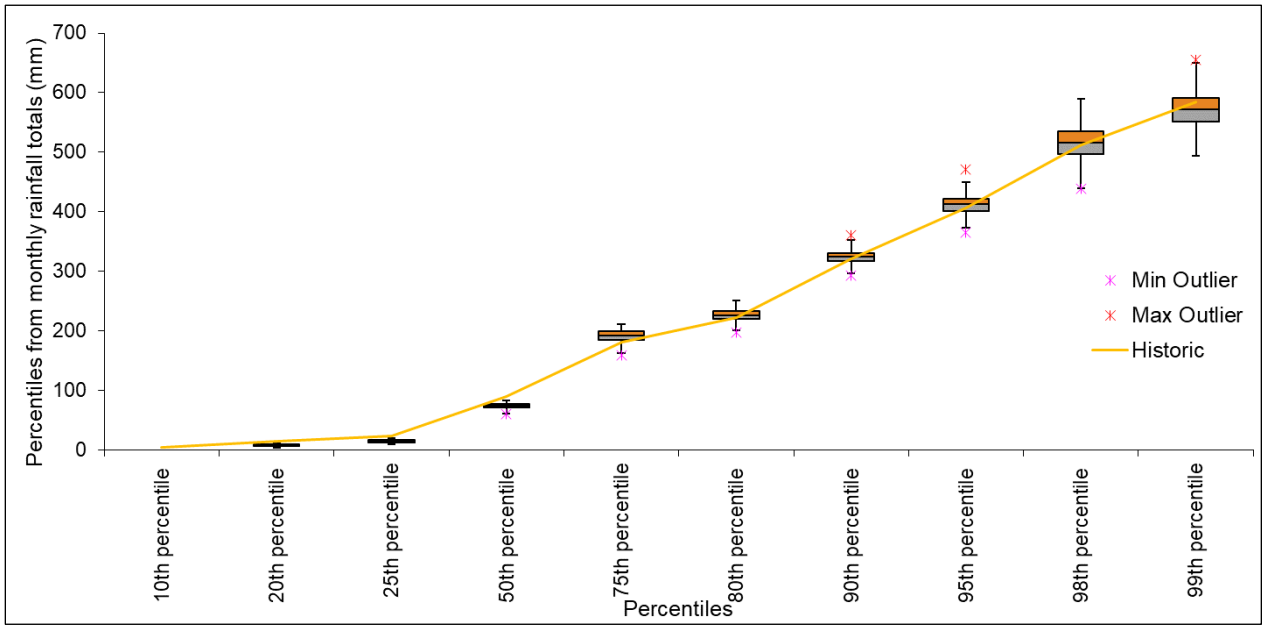


Figure 4.3-6: Percentiles of monthly data generated using VLB Method for Hebron station

Annual statistics of the generated data using VLB Method for Hebron rainfall station

Figure 4.3-7 and Figure 4.3-8 show the annual totals distribution curve and percentiles respectively. The generated data does fall within the lower and upper extreme values of the box and whisker plots with 64% of the historic data falling within the interquartile range and 6% falling outside the plots. The annual percentiles shows that historic data compares relatively well to the generated data.

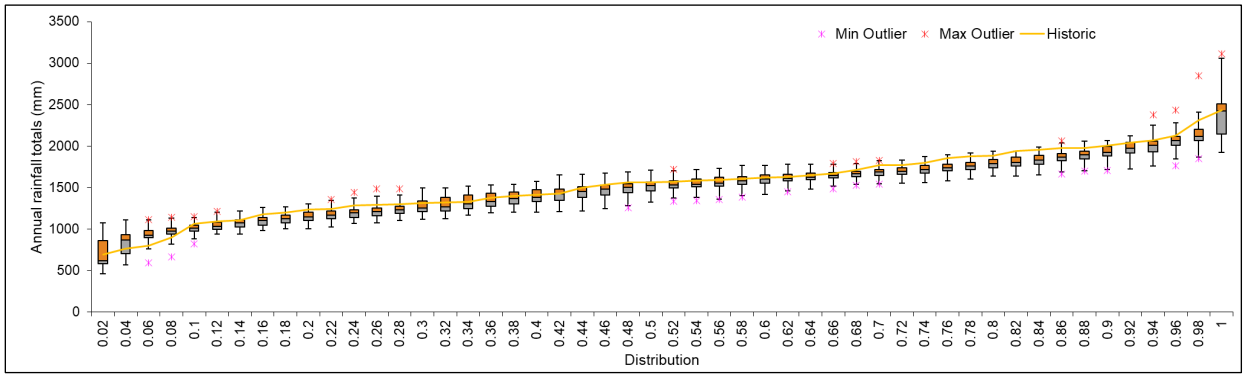


Figure 4.3-7: Rainfall distribution curve of the annual totals for data generated using VLB Method for Hebron station

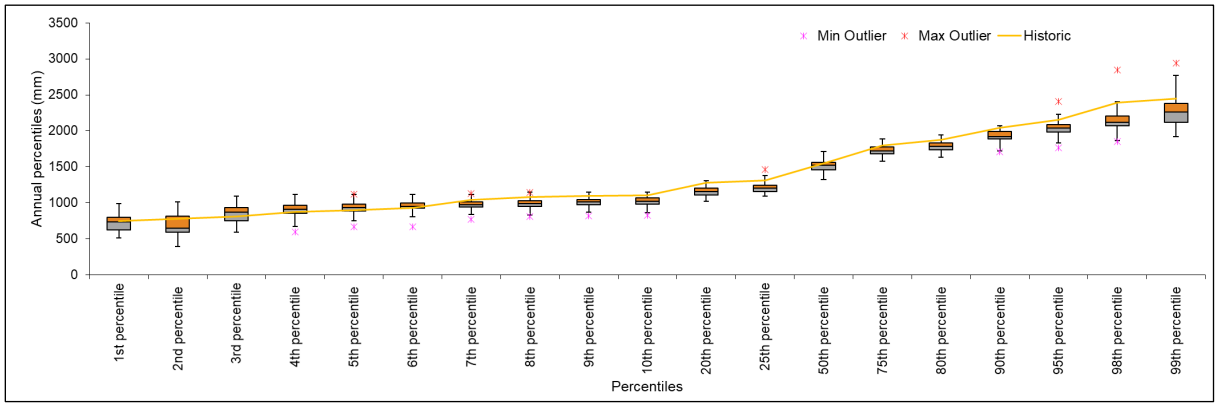


Figure 4.3-8: Annual percentiles of data generated using VLB Method for Hebron station

4.3.2. Welgevonden Rainfall station

Daily statistics of the generated data using VLB Method for Welgevonden rainfall station

Figure 4.3-9 and Figure 4.3-10 show the box and whisker plots of the rainfall distribution curve for the daily maximum values and the percentiles from 100 sequences of 50 years daily generated

data using the VLB Method for Welgevonden station. As can be seen in Figure 4.3-9, the historic values correlate well with the generated daily maximum values from the VLB method. Figure 4.3-10 shows that only the 90th and 95th percentiles of the historic data fall within the interquartile range of the box and whisker plots.

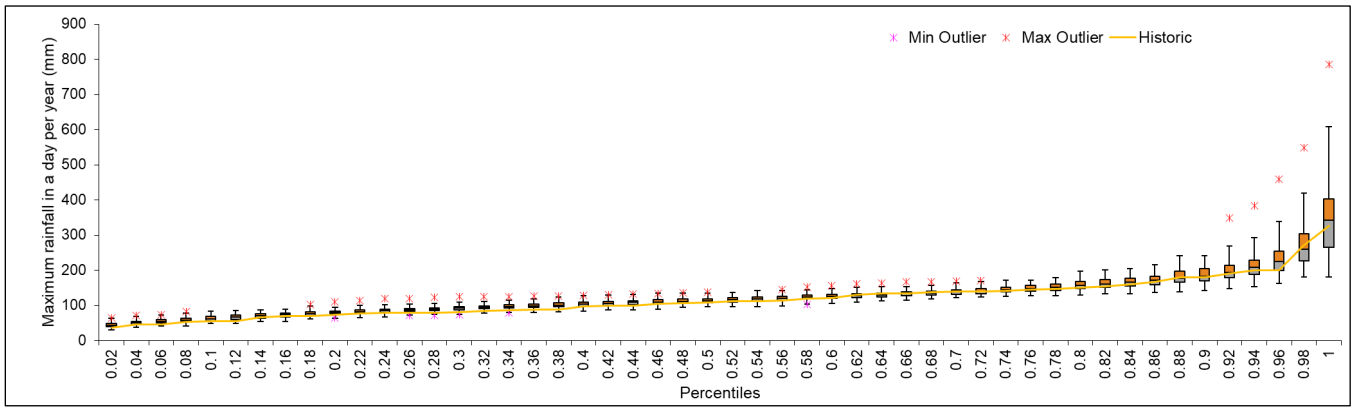


Figure 4.3-9: Rainfall distribution curve of the daily maximum values for data generated using VLB Method for Welgevonden station

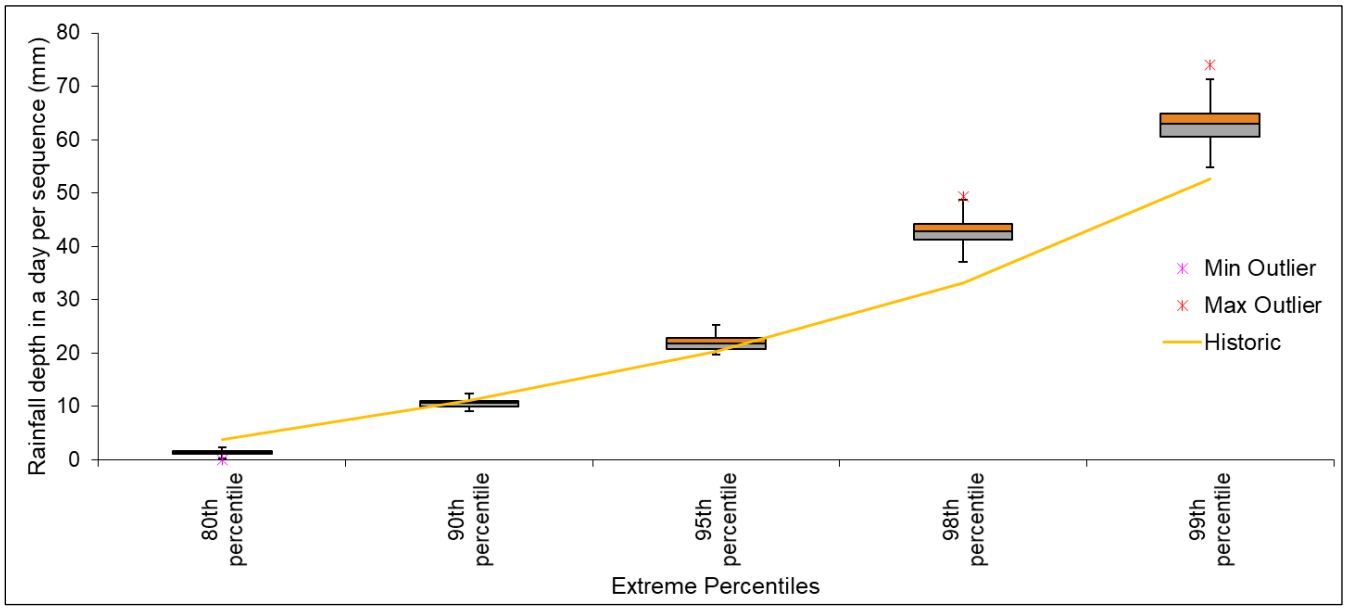


Figure 4.3-10: Extreme percentiles of daily data generated using VLB Method for Welgevonden station

Monthly statistics of the generated data using VLB Method for Welgevonden rainfall station

Figure 4.3-11 to Figure 4.3-14 shows the monthly statistics of the generated data using VLB Method based on the generated 100 sequences for 50 years for Welgevonden station. The monthly averages show that it compares well with the historic data because the historic data falls within the interquartile range of the generated data. Most of the maximum values for the historic data falls within the interquartile range as well except for February, June, July, and August. The generated monthly maximum values are lower than the historic value for February, this was also seen with the WGEN method. The monthly percentiles compare relatively well to the historic values. The standard deviation historic values from October to January falls outside of the box and whisker plots.

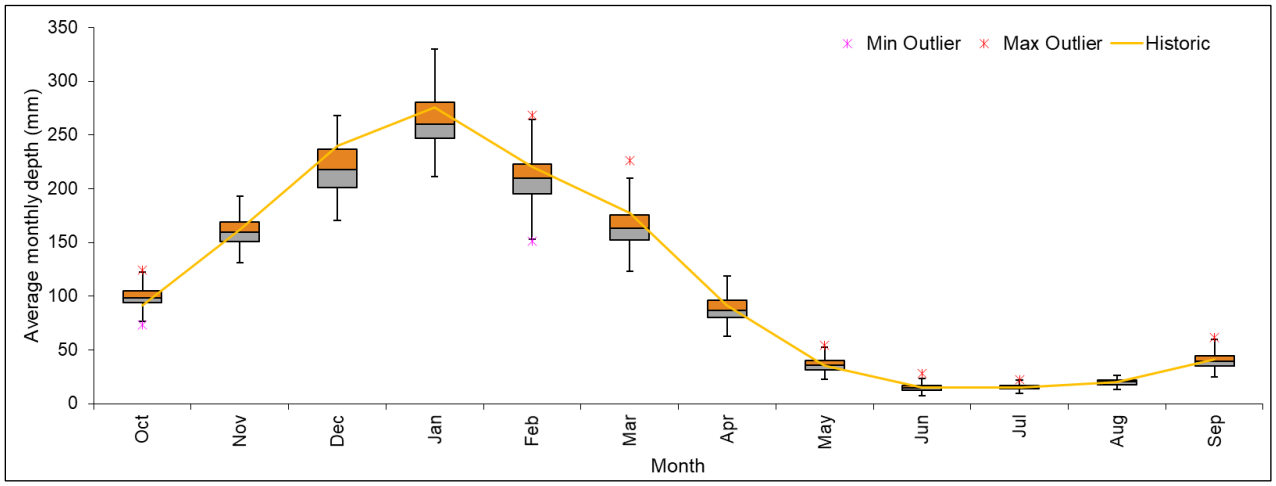


Figure 4.3-11: Averages of monthly data generated using VLB Method for Welgevonden station

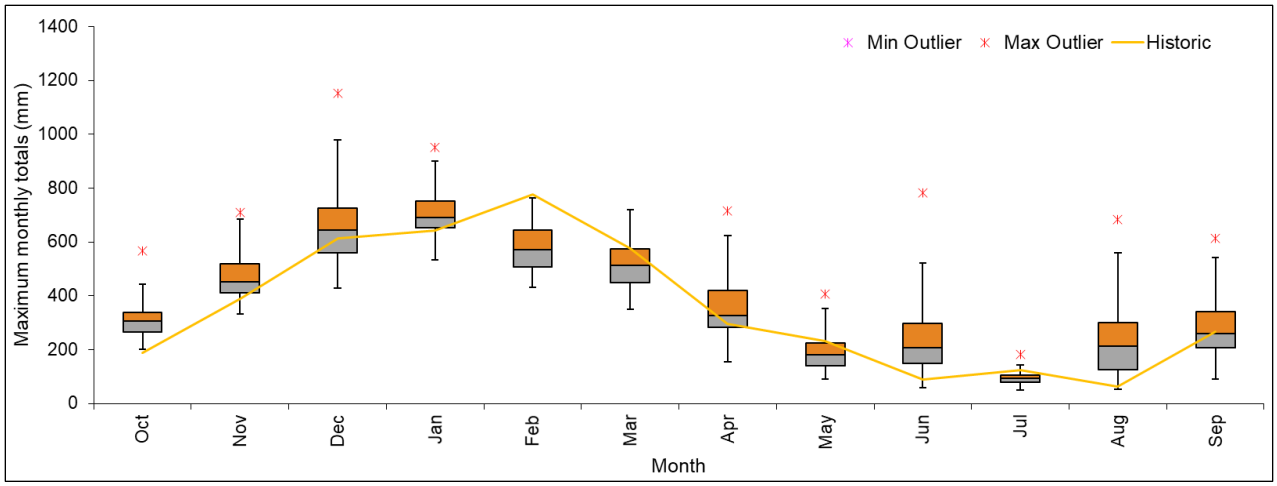


Figure 4.3-12: Maximum values of monthly data generated using VLB Method for Welgevonden station

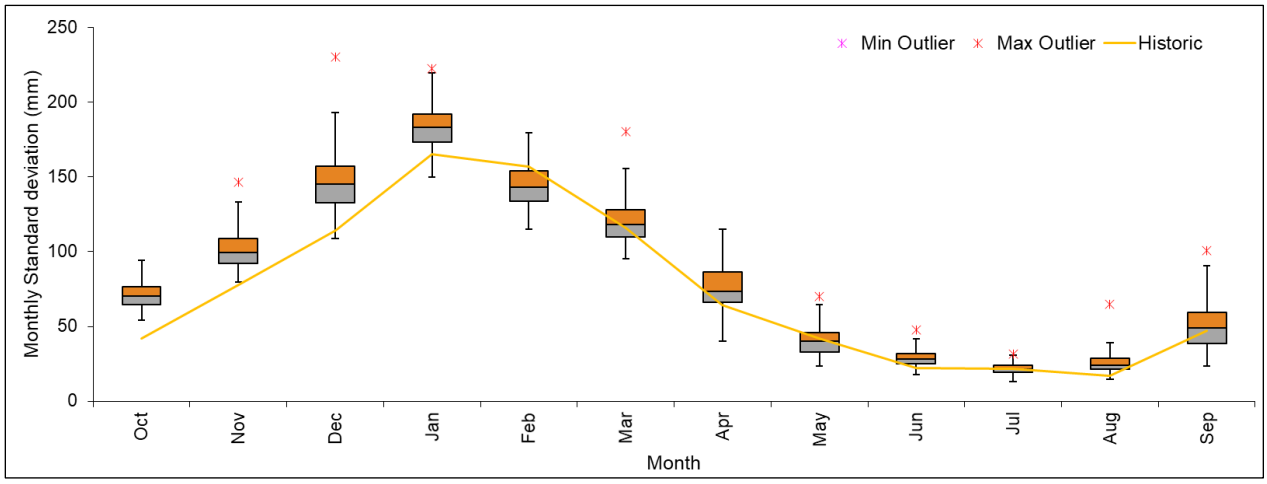


Figure 4.3-13: Standard deviation of monthly data generated using VLB Method for Welgevonden station

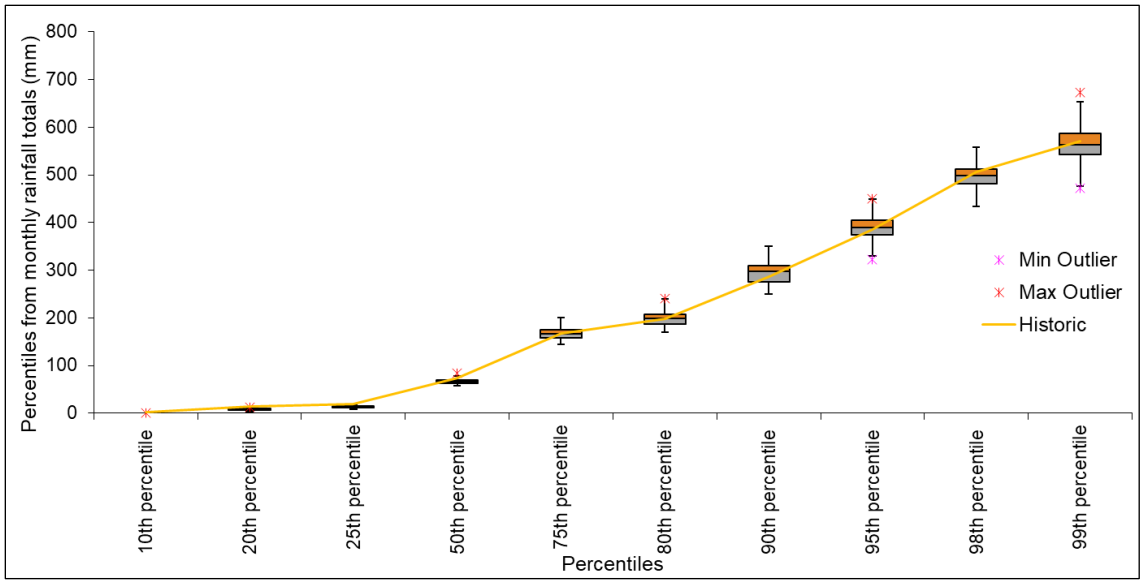


Figure 4.3-14: Percentiles of monthly data generated using VLB Method for Welgevonden station

Annual statistics of the generated data using VLB Method for Welgevonden rainfall station

Figure 4.3-15 and Figure 4.3-16 show the rainfall distribution curve of annual totals and annual percentiles respectively. The historic data compares well to the generated annual rainfall distribution box and whisker plots. The annual percentiles shows that historic data compares relatively well to the generated data.

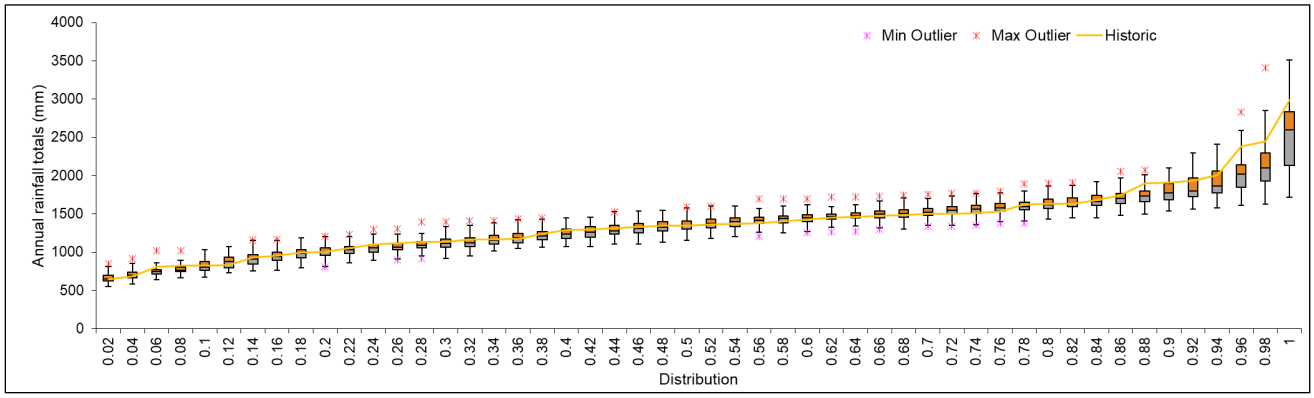


Figure 4.3-15: Rainfall distribution curve of the annual totals for data generated using VLB Method for Welgevonden station

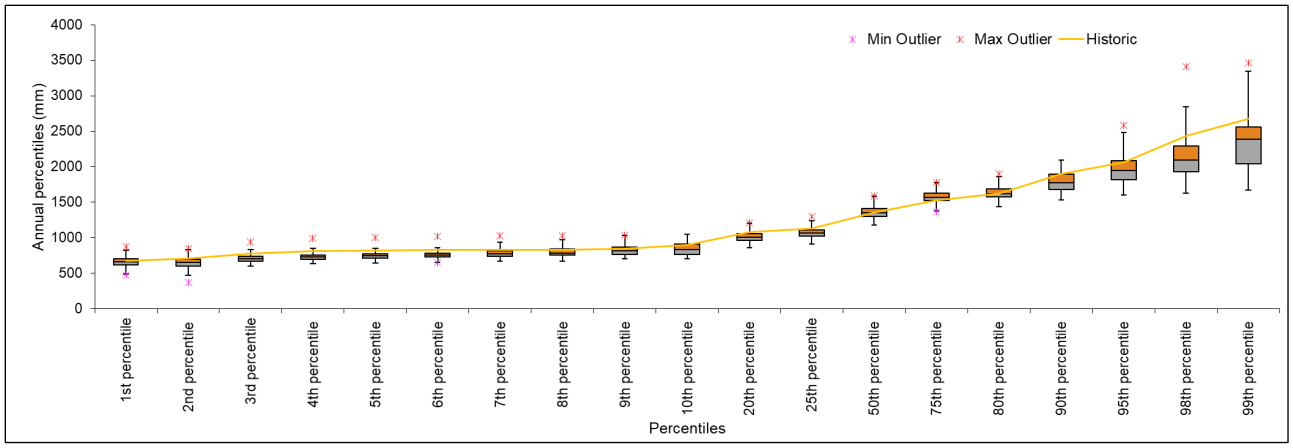


Figure 4.3-16: Annual percentiles of data generated using VLB Method for Welgevonden station

4.3.3. Nelshoogte Rainfall station

Daily statistics of the generated data using VLB Method for Nelshoogte rainfall station

Figure 4.3-17 and Figure 4.3-18 shows the box and whisker plots of the rainfall distribution curve for the daily maximum values and the daily percentiles from 100 sequences of 50 years daily generated data using the VLB Method for Nelshoogte station. As can be seen in Figure 4.3-17, the historic values correlate well with the generated daily maximum values from the VLB method. All percentiles of the historic data fall within the interquartile range of the box and whisker plots.

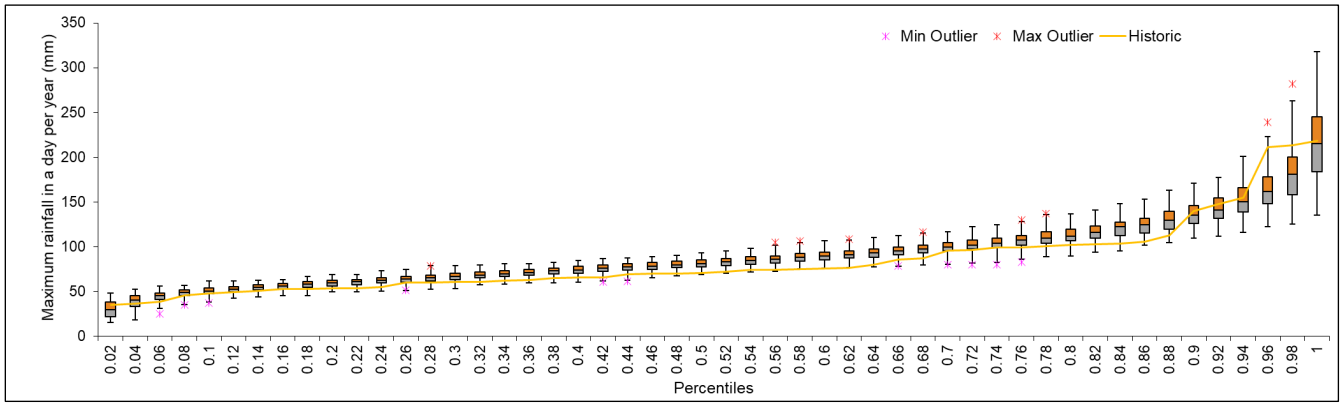


Figure 4.3-17: Rainfall distribution curve of the daily maximum values for data generated using VLB Method for Nelshoogte station

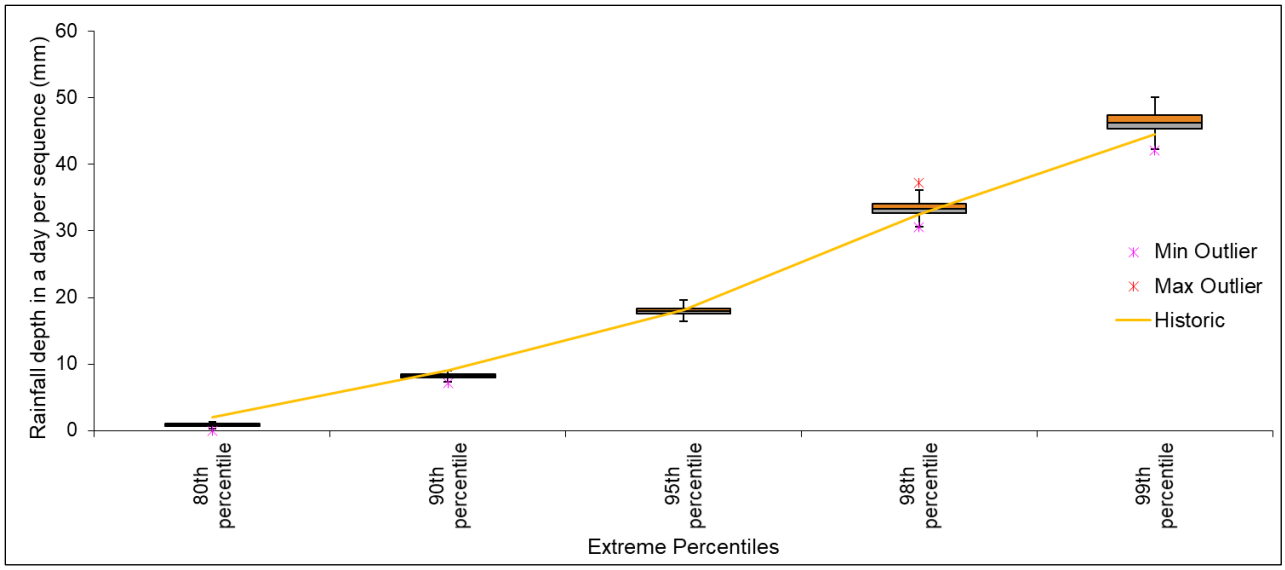


Figure 4.3-18: Percentiles of daily data generated using VLB Method for Nelshoogte station

Monthly statistics of the generated data using VLB Method for Nelshoogte rainfall station

Figure 4.3-19 to Figure 4.3-22 show the monthly statistics of the generated data using VLB Method based on the generated 100 sequences for 50 years for Nelshoogte station. The monthly averages shows that the generated data compares relatively well with the historic data as the historic data falls within the interquartile range of the generated data except for the month of February. The box and whisker maximum values show a similar pattern to the standard deviation values whereby May to September fall within the interquartile range and the rest of the months fall outside of the box and whisker plots.

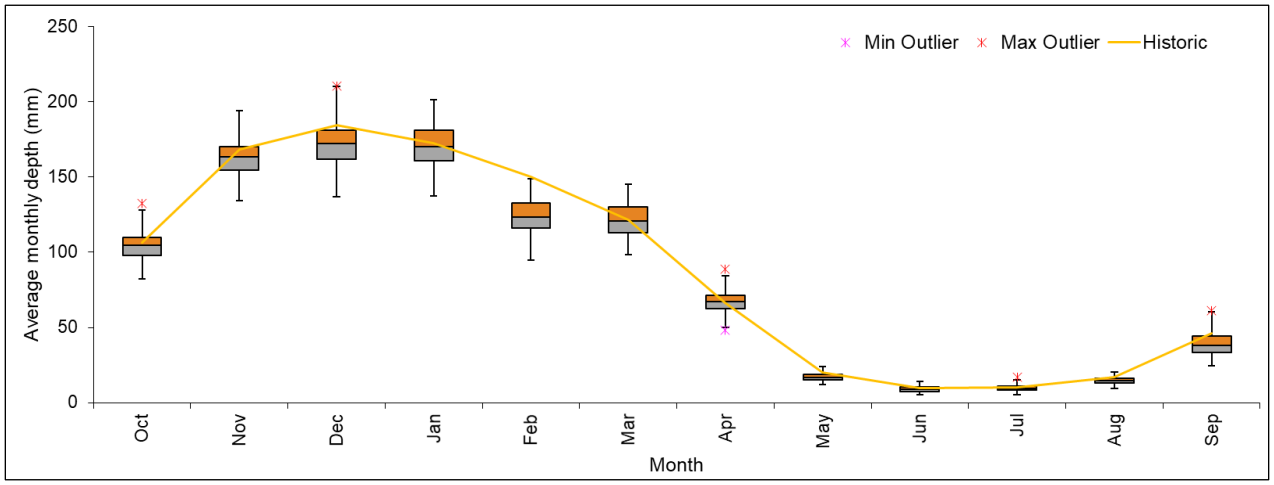


Figure 4.3-19: Averages of monthly data generated using VLB Method for Nelshoogte station

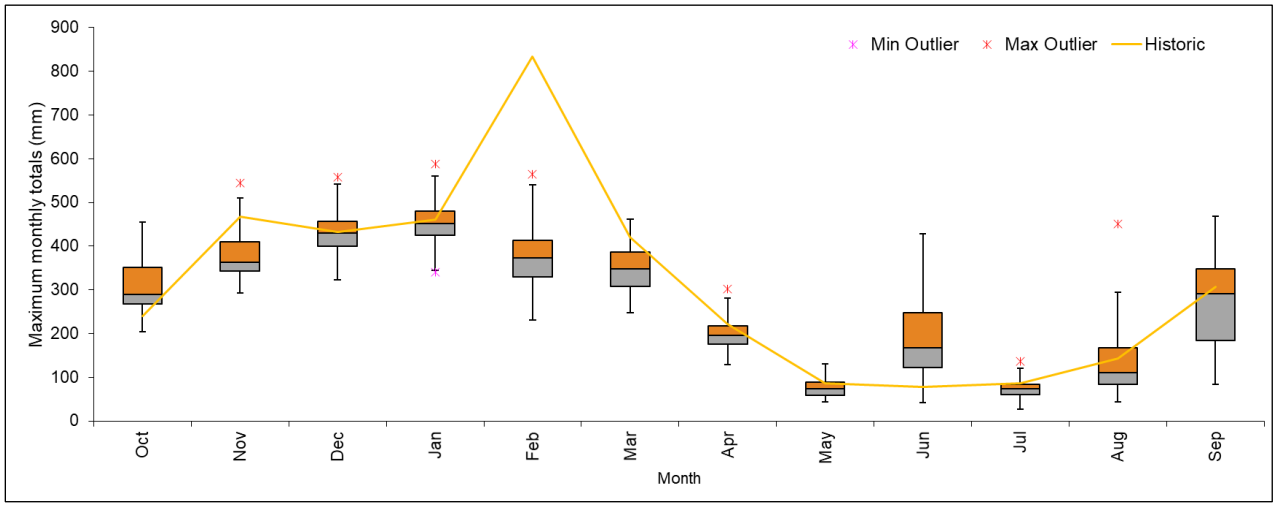


Figure 4.3-20: Maximum values of monthly data generated using VLB Method for Nelshoogte station

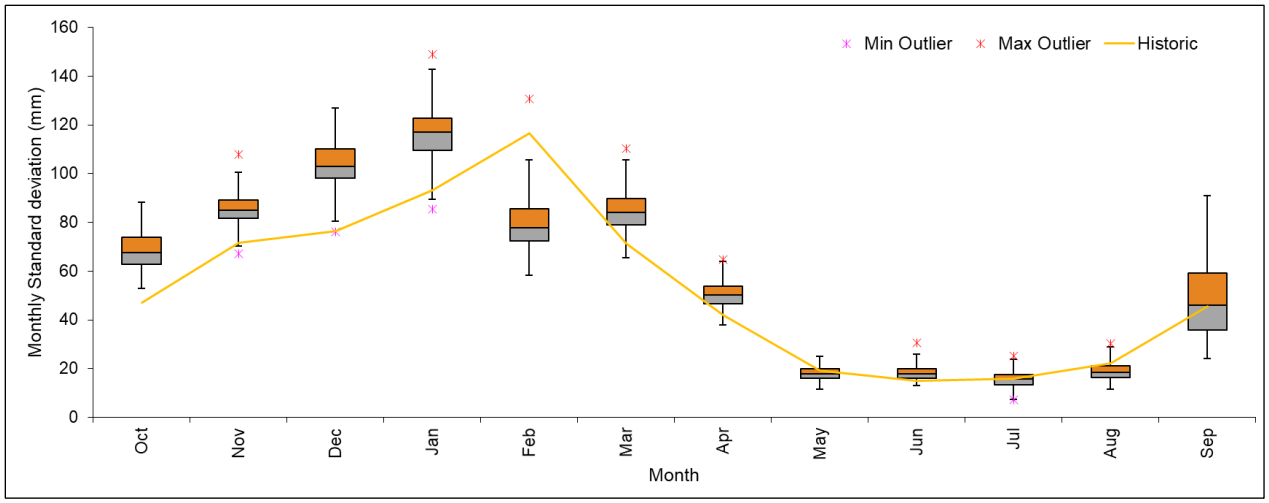


Figure 4.3-21: Standard deviation values of monthly data generated using VLB Method for Nelshoogte station using VLB method

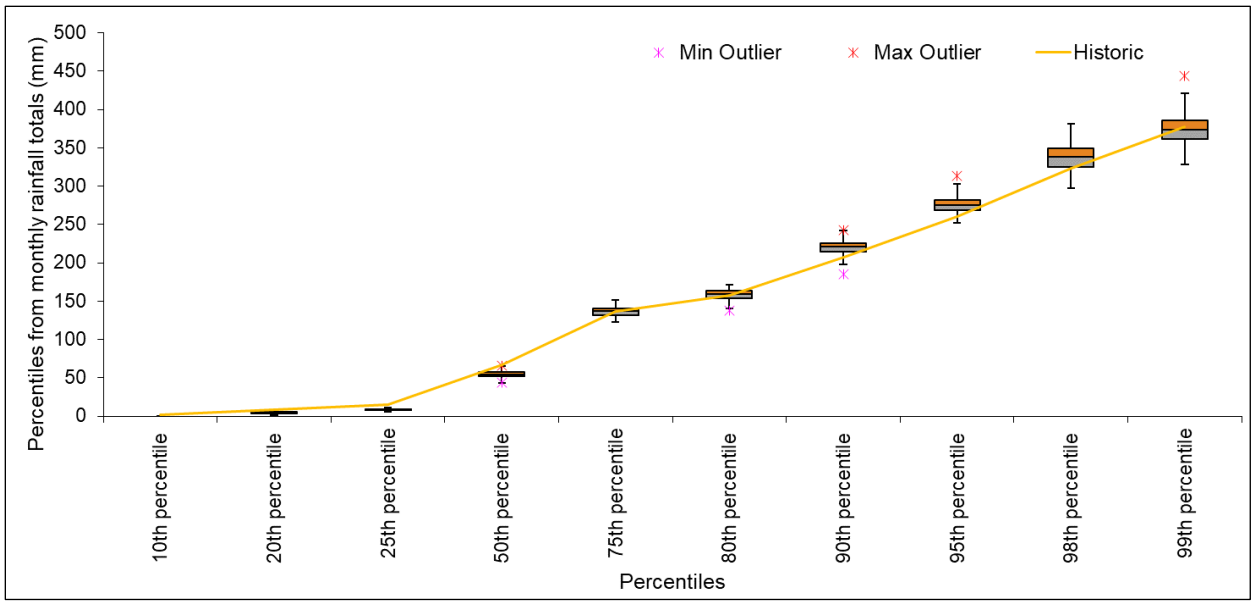


Figure 4.3-22: Percentiles of monthly data generated using VLB Method for Nelshoogte station

Annual statistics of the generated data using VLB Method for Nelshoogte rainfall station

Figure 4.3-23 and Figure 4.3-24 shows the rainfall distribution curve of annual totals and percentiles respectively. The historic values compare well to the generated box and whisker plots of the distribution curves. The annual percentiles shows that historic data compares relatively well to the generated data.

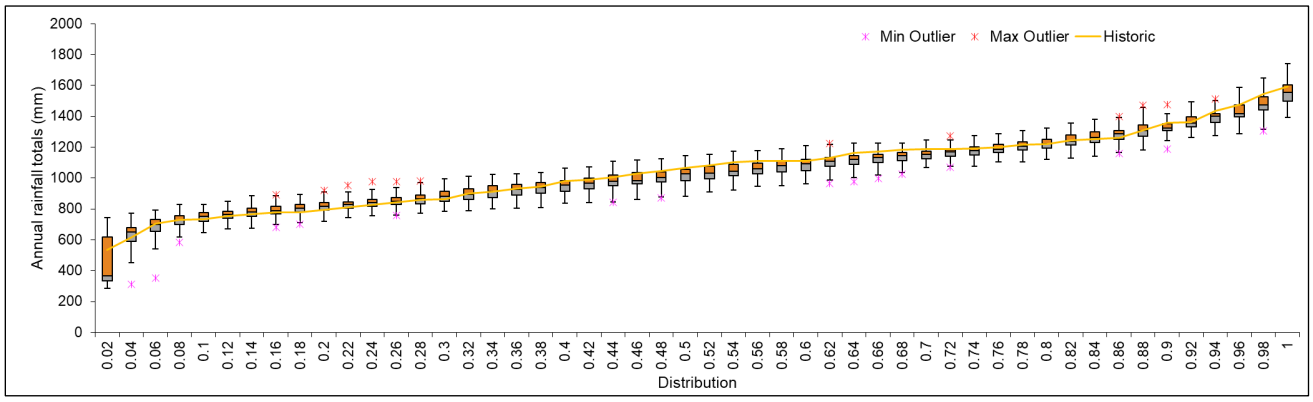


Figure 4.3-23: Rainfall distribution curve of the annual totals for data generated using VLB Method for Nelshoogte station

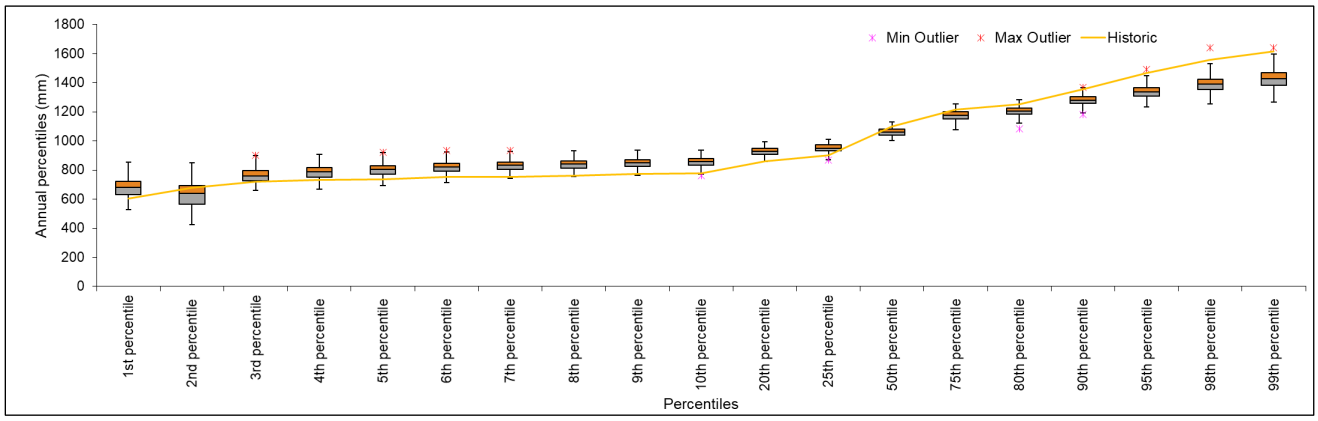


Figure 4.3-24: Annual percentiles of data generated using VLB Method for Nelshoogte station

5. COMPARISON OF GENERATORS

Section 5.1, 5.2 and 5.3 describe the comparison of the three methods for Hebron, Welgevonden and Nelshoogte rainfall stations respectively. For each station, daily, monthly and annual statistics from TPM, WGEN and VLB are compared. The results show that for daily, monthly and annual generated data, WGEN has generally generated lower values than the TPM and VLB Methods.

5.1. Hebron Rainfall Station Comparison

Figure 5.1-1 to Figure 5.1-11 show the comparison of the methods for the daily, monthly, and annual statistics for Hebron station. The results show that both TPM and VLB compared relatively well for the daily statistics except for the daily average in which the historic value does not fall within the interquartile range of the VLB method. WGEN only compared relatively well for the daily average, but for the other statistics, the historic data was outside the box and whisker plots. For the monthly averages shown in Figure 5.1-5, all three methods compared relatively well to the historic monthly averages. With the monthly maximum values, generally in some of the months, the methods underestimated the maximum values compared to the historic values with VLB performing relatively better than the other methods. With the monthly standard deviation values shown in Figure 5.1-7, WGEN and TPM methods performed better than VLB method. The annual averages compared relatively well with the historic value for TPM and WGEN. For the annual maximum values, the TPM and VLB methods performed better. For all the annual statistics from the methods, the historic values still fall within the lower and upper extreme values of the box and whisker plots.

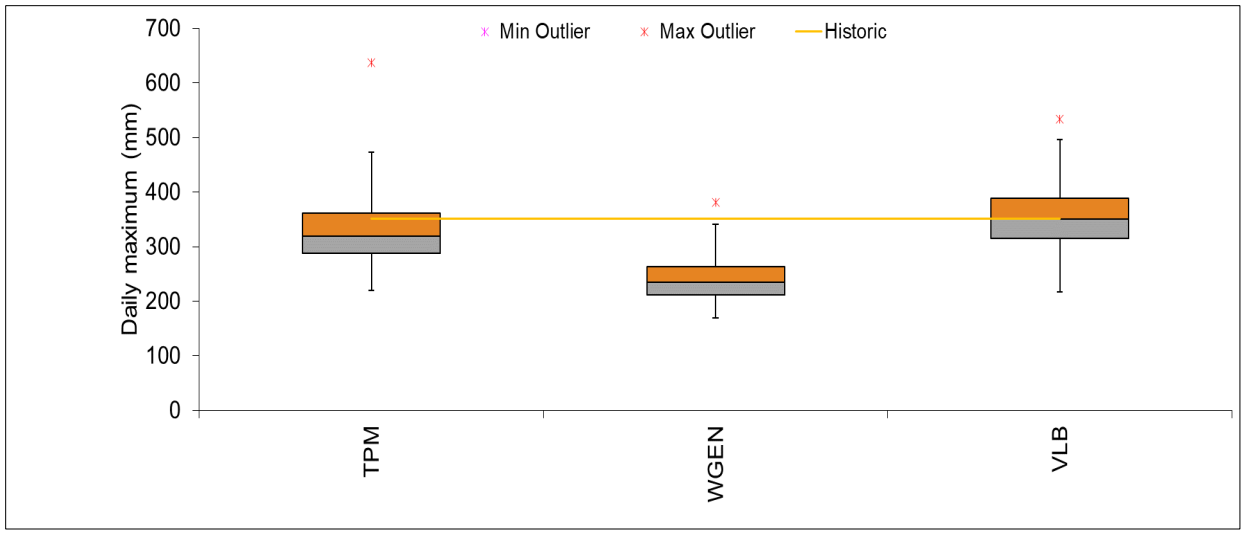


Figure 5.1-1: Comparison of the daily maximum values generated from the methods for the Hebron station

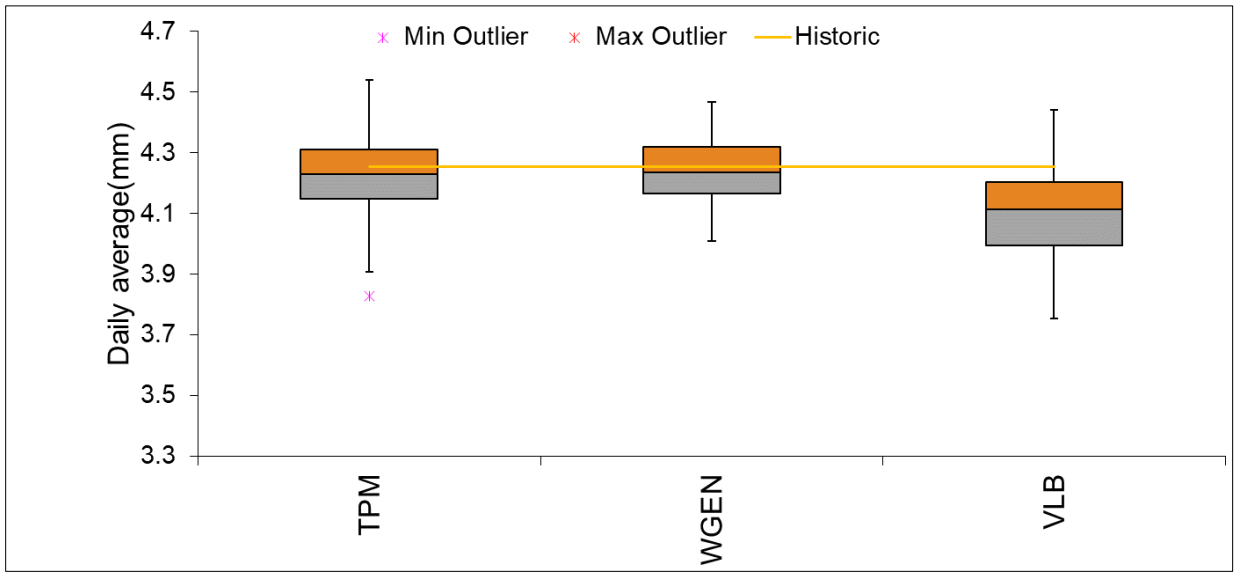


Figure 5.1-2: Comparison of the daily average values generated from the methods for the Hebron station

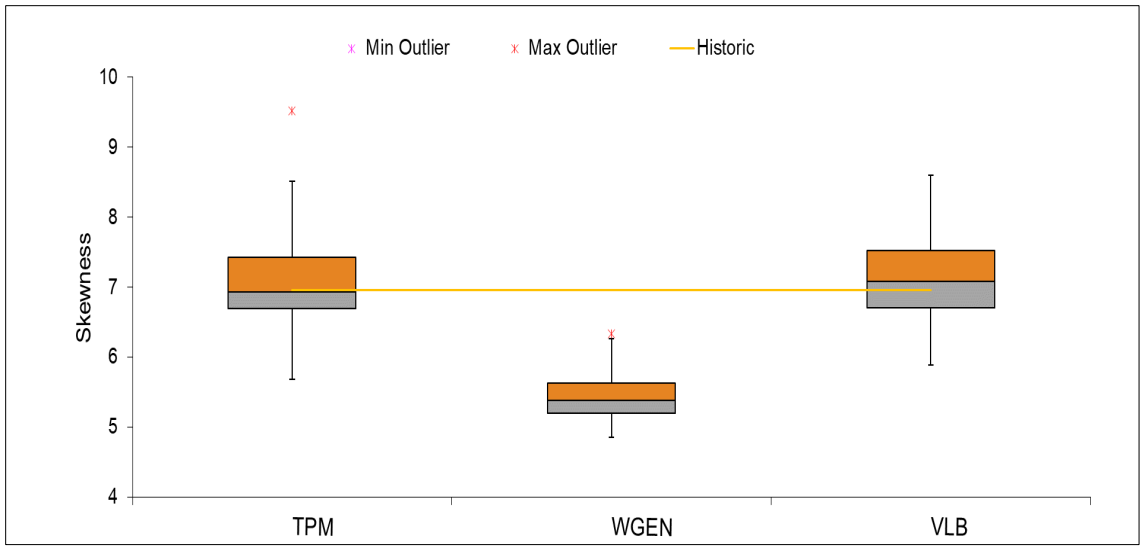


Figure 5.1-3: Comparison of the daily skewness values generated from the methods for the Hebron station

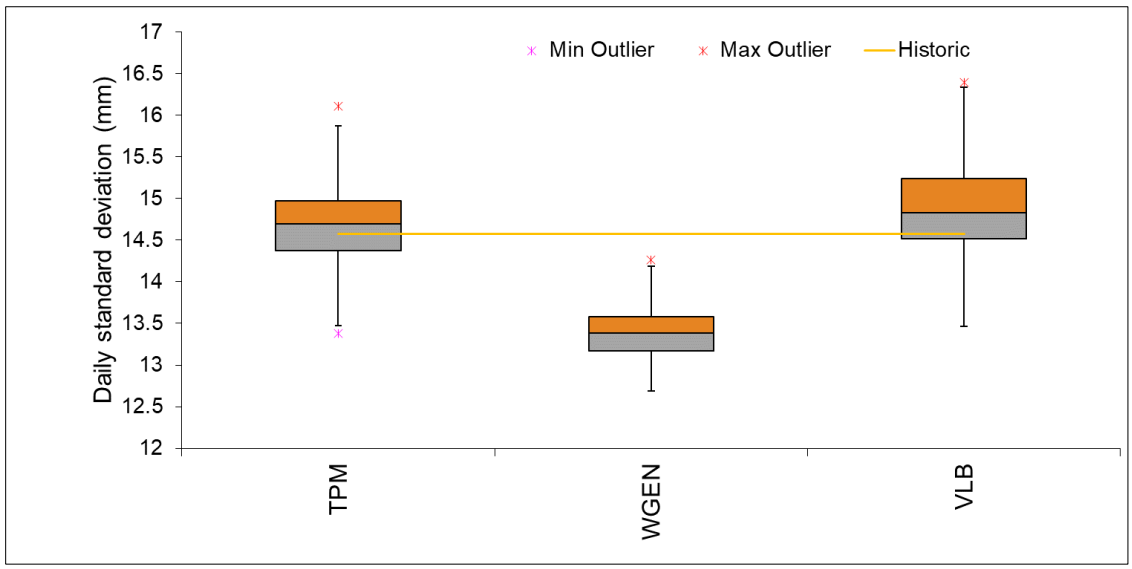


Figure 5.1-4: Comparison of the daily standard deviation values generated from the methods for the Hebron station

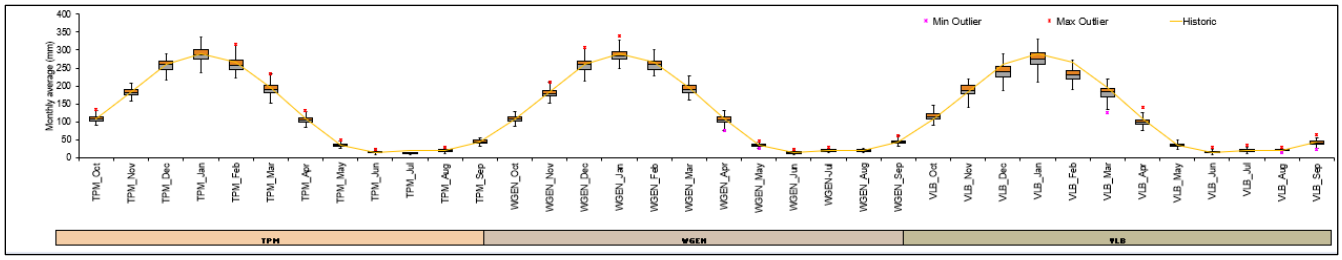


Figure 5.1-5: Comparison of the monthly averages from the methods for the Hebron station

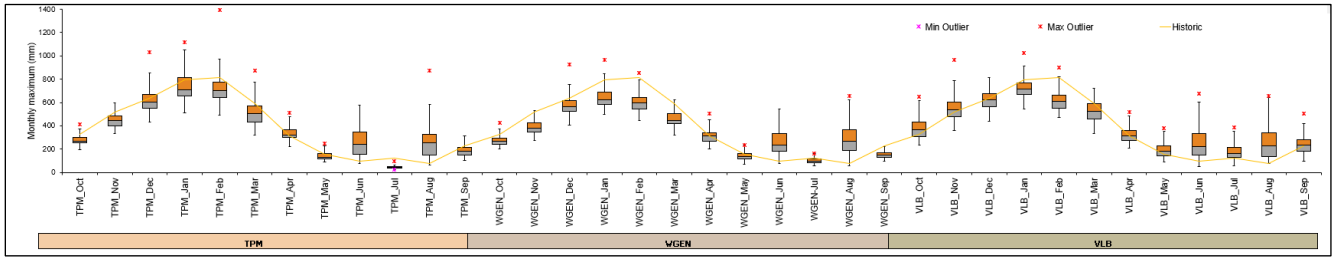


Figure 5.1-6: Comparison of the monthly maximum values for the methods for the Hebron station

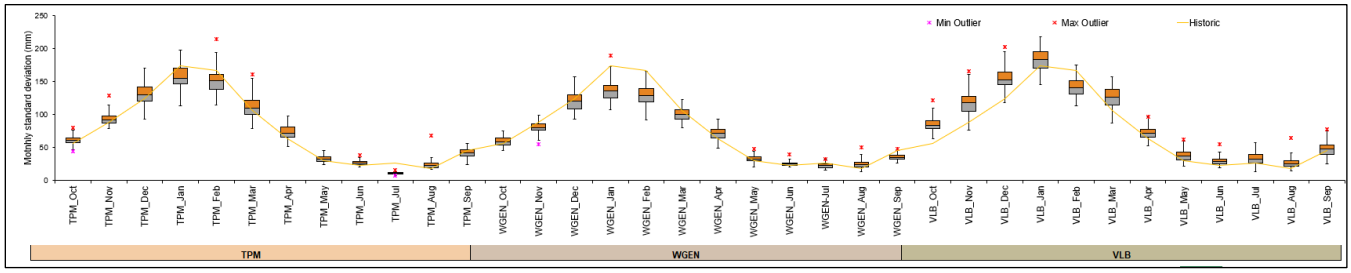


Figure 5.1-7: Comparison of the monthly standard deviation for the methods for the Hebron station

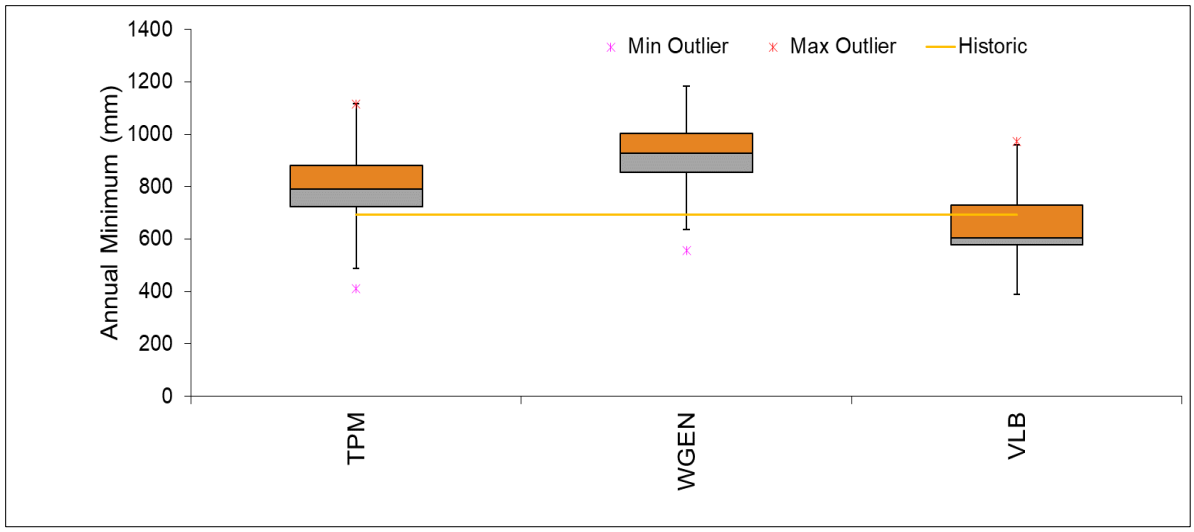


Figure 5.1-8: Comparison of the annual minimum values generated from the methods for the Hebron station

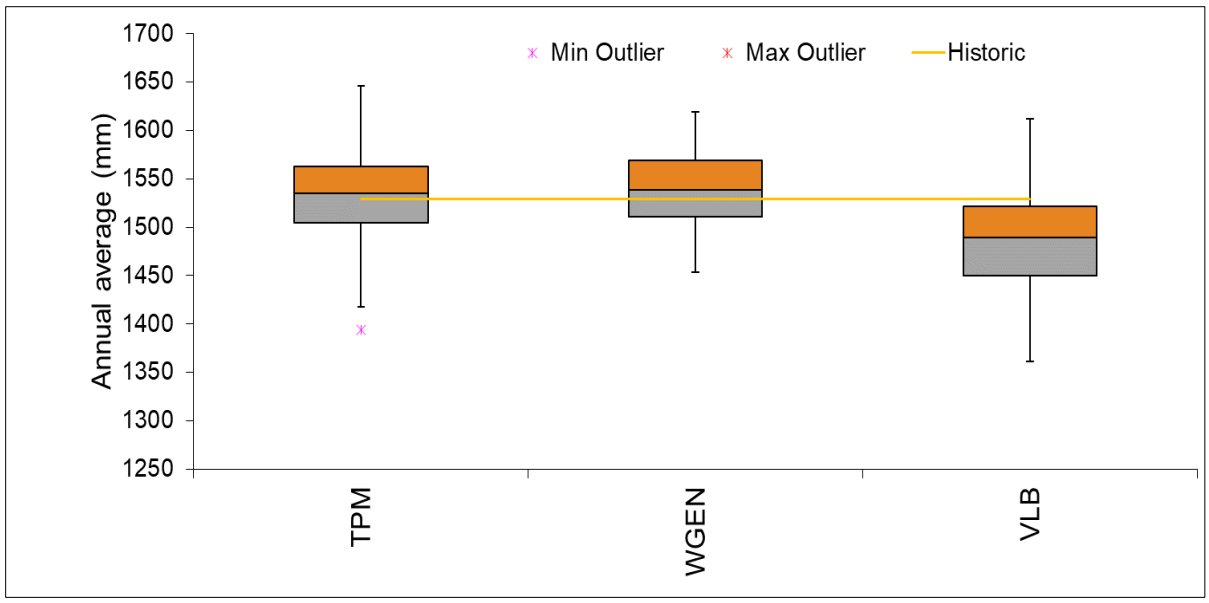


Figure 5.1-9: Comparison of the annual average values generated from the methods for the Hebron station

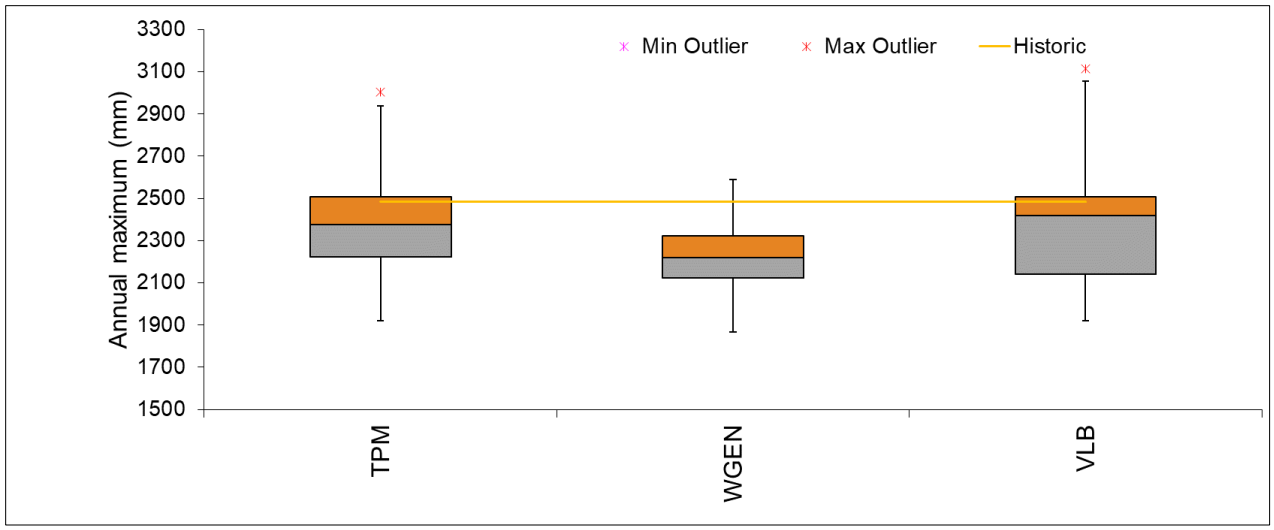


Figure 5.1-10: Comparison of the annual maximum values generated from the methods for the Hebron station

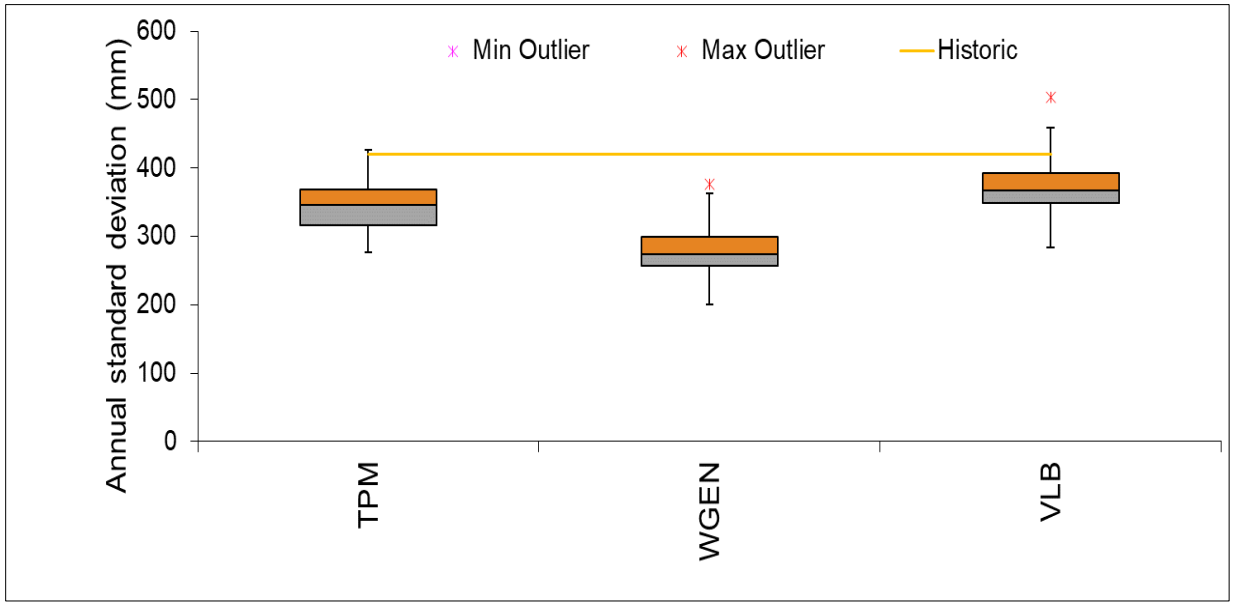


Figure 5.1-11: Comparison of the annual standard deviation values generated from the methods for the Hebron station

Table 5.1-1, Table 5.1-2 and Table 5.1-3 summarizes the average statistics from 100 sequences generated using the three methods as compared to the observed historic data for the Hebron rainfall station. For the daily statistics, the VLB method produced an average daily maximum value that is the closest to the historic value, the average daily maximum value generated from WGEN method has a difference of over 100 mm compared to the historic value. WGEN method also generated data with the lowest standard deviation and skewness average values for the daily statistics compared to the other methods. For the monthly average statistics, TPM methods produced average statistics that were relatively more similar to the historic values. The VLB method generated monthly statistics with the lowest average mean and median values, while WGEN method generated the lowest average standard deviation and skewness values. Similarly with the daily and monthly statistics, the annual average statistics also showed that WGEN tends to generate values that are closer to the mean of the set compared to the other methods. All methods performed relatively well with the annual statistics.

Table 5.1-1: Average statistics of the daily data generated using TPM, WGEN and VLB for the Hebron rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	4.3	4.2	4.2	4.1
Median (mm)	0.0	0.0	0.0	0.0
Minimum (mm)	0.0	0.0	0.0	0.0
Maximum (mm)	350.7	333.5	239	352.1
Standard deviation (mm)	14.6	14.7	13.4	14.9
Sample size	25 614	18 628	18 628	18 628
Mode	0.0	0.0	0.0	0.0
Skewness	7.0	7.1	5.4	7.2
5^{th} Percentile (mm)	0	0	0	0

95th Percentile (mm)	26	26.0	27.6	25.7

Table 5.1-2: Average statistics of the monthly data generated using TPM, WGEN and VLB for the Hebron rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	128.7	128	128.9	124.7
Median (mm)	88.8	85	89.0	74
Minimum (mm)	0.0	0.0	0.0	0.0
Maximum (mm)	815.8	804	685	749.5
Standard deviation (mm)	137.5	135	128.0	139.2
Sample size	844	612	612	612
Mode	0.0	0.0	0.0	0.0
Skewness	1.6	1.5	1.3	1.5
5th Percentile (mm)	0	0	0	0
95th Percentile (mm)	405	399	464	411.6

Table 5.1-3: Average statistics of the annual data generated using TPM, WGEN and VLB for the Hebron rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	1 529.3	1 535	1 537	1 485
Median (mm)	1 534.1	1 523	1 530	1 514
Minimum (mm)	692.5	826.2	921	654
Maximum (mm)	2 485.2	2380.4	2 228	2 378
Standard deviation (mm)	419.7	346	278	372.5
Sample size	71.0	51.0	51.0	51.0
Mode	-	-	-	-
Skewness	-0.38	0.2	0.2	0.03
5^{th} Percentile (mm)	849.0	1047.8	1122	920
95^{th} Percentile (mm)	2153.8	2427.6	1982	2038

*- refers to no value was found as a mode

5.2. Welgevonden Rainfall Station Comparison

Figure 5.2-1 to Figure 5.2-11 show the comparison of the methods for the daily, monthly, and annual statistics for Welgevonden station. The daily statistics show that both TPM and VLB methods compared relatively well for the daily statistics except for the daily average in which the historic value does not fall within the interquartile range of the VLB method. WGEN method only compared relatively well for the daily average, for the other statistics, the historic data was outside of the box and whisker plots. For the monthly averages shown in Figure 5.2-5, all three methods compared relatively well to the historic monthly averages. With the monthly maximum values, generally, TPM method performed better than the other methods. With the monthly standard deviation values shown in Figure 5.2-7, TPM method still performed better than the other methods. The annual averages compared relatively well with the historic value for all the methods. For all the other annual statistics, the historic values did not fall within the interquartile range of the box and whisker plots of the methods.

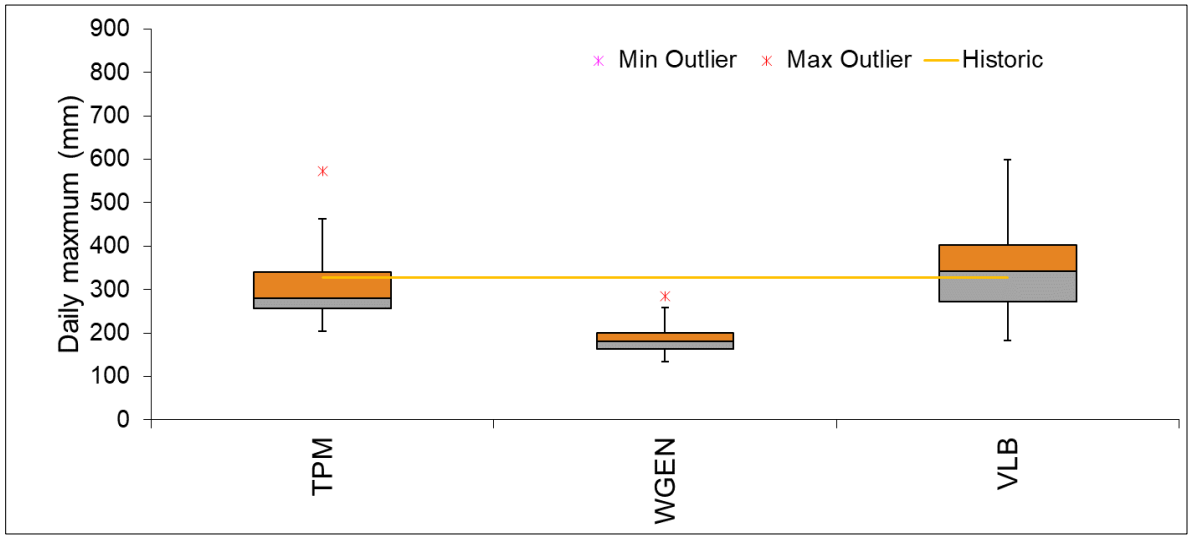


Figure 5.2-1: Comparison of the daily maximum generated from the methods for Welgevonden station

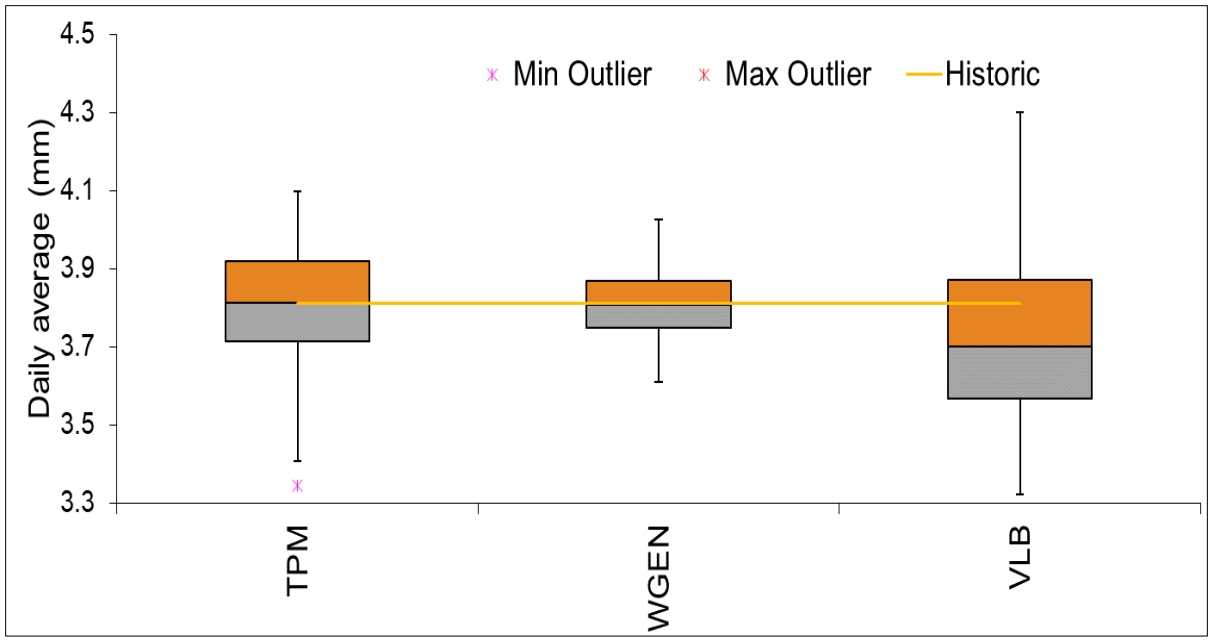


Figure 5.2-2: Comparison of the daily average generated from the methods for Welgevonden station

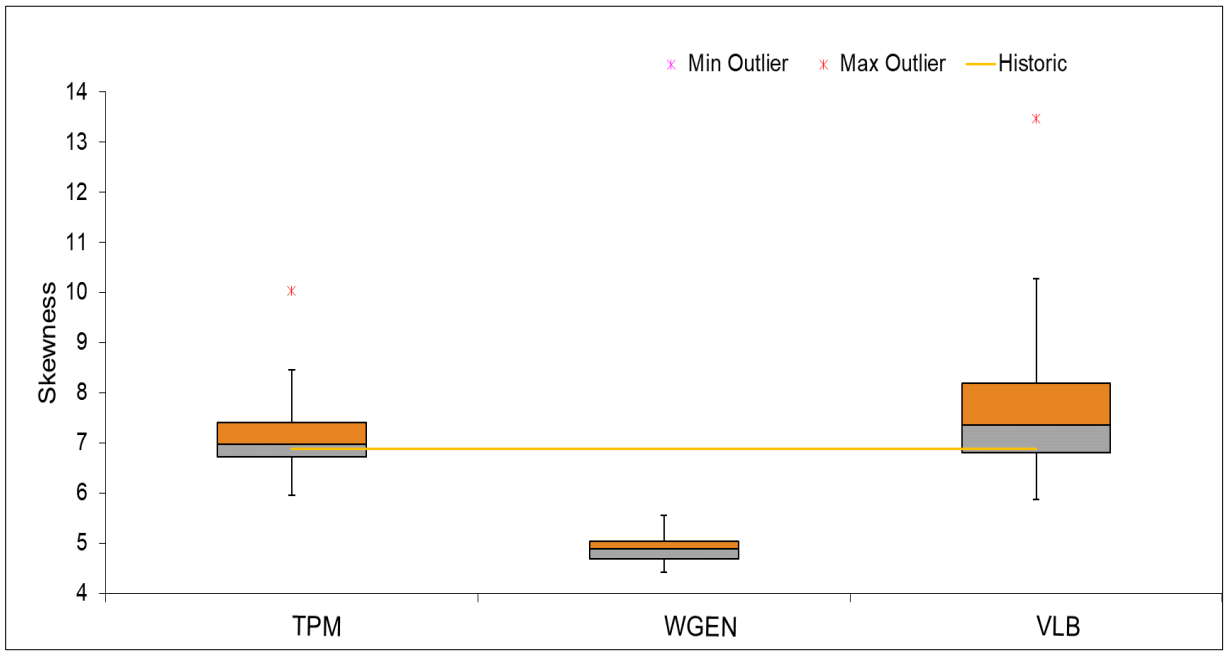


Figure 5.2-3: Comparison of the daily skewness values generated from the methods for Welgevonden station

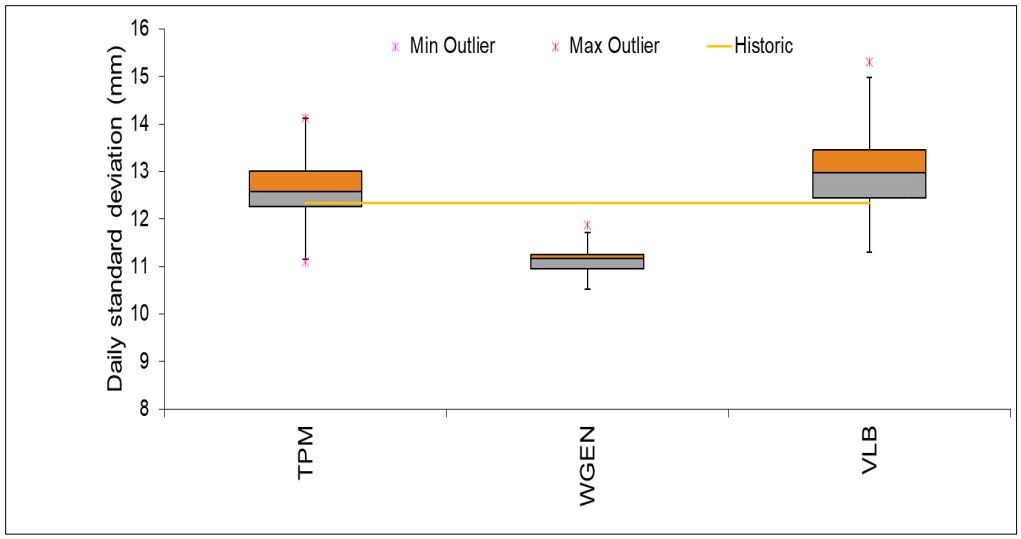


Figure 5.2-4: Comparison of the daily standard deviation values generated from the methods for Welgevonden station

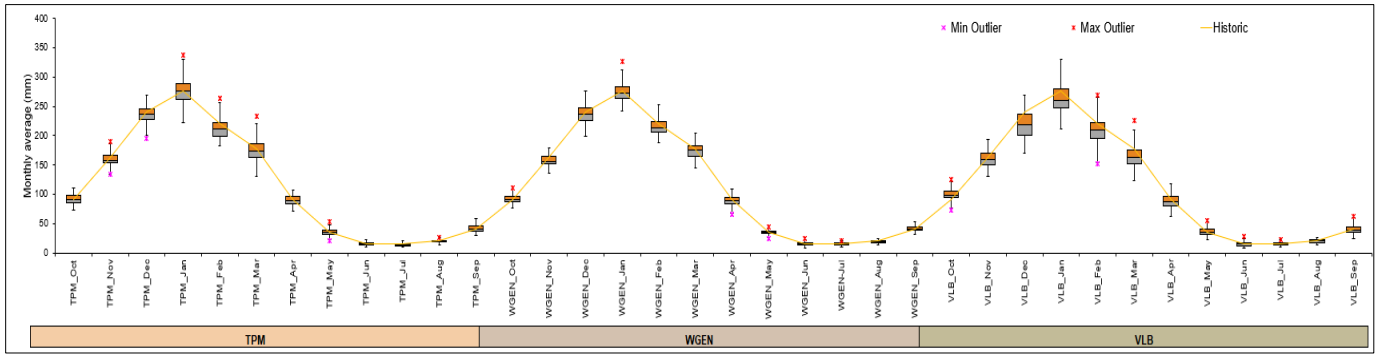


Figure 5.2-5:Comparison of the monthly averages from the methods for the Welgevonden station

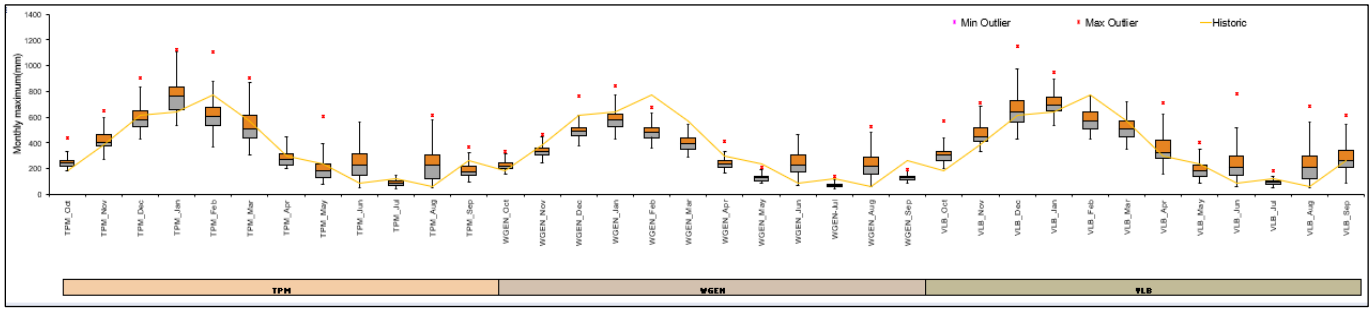


Figure 5.2-6: Comparison of the monthly maximum values from the methods for the Welgevonden station

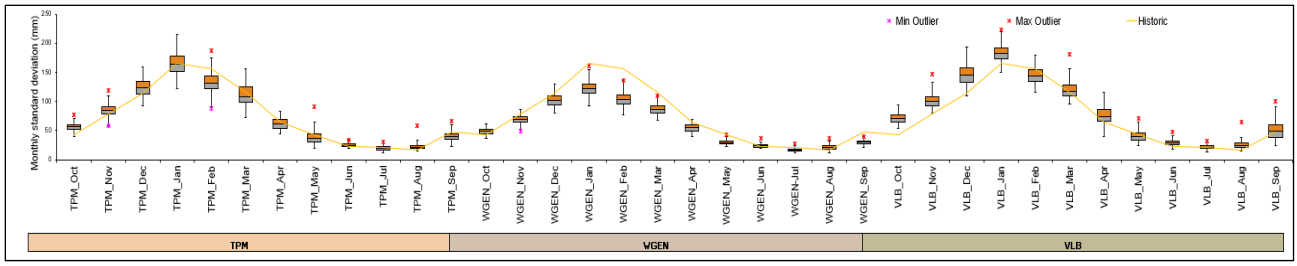


Figure 5.2-7: Comparison of the monthly standard deviation from the methods for the Welgevonden station

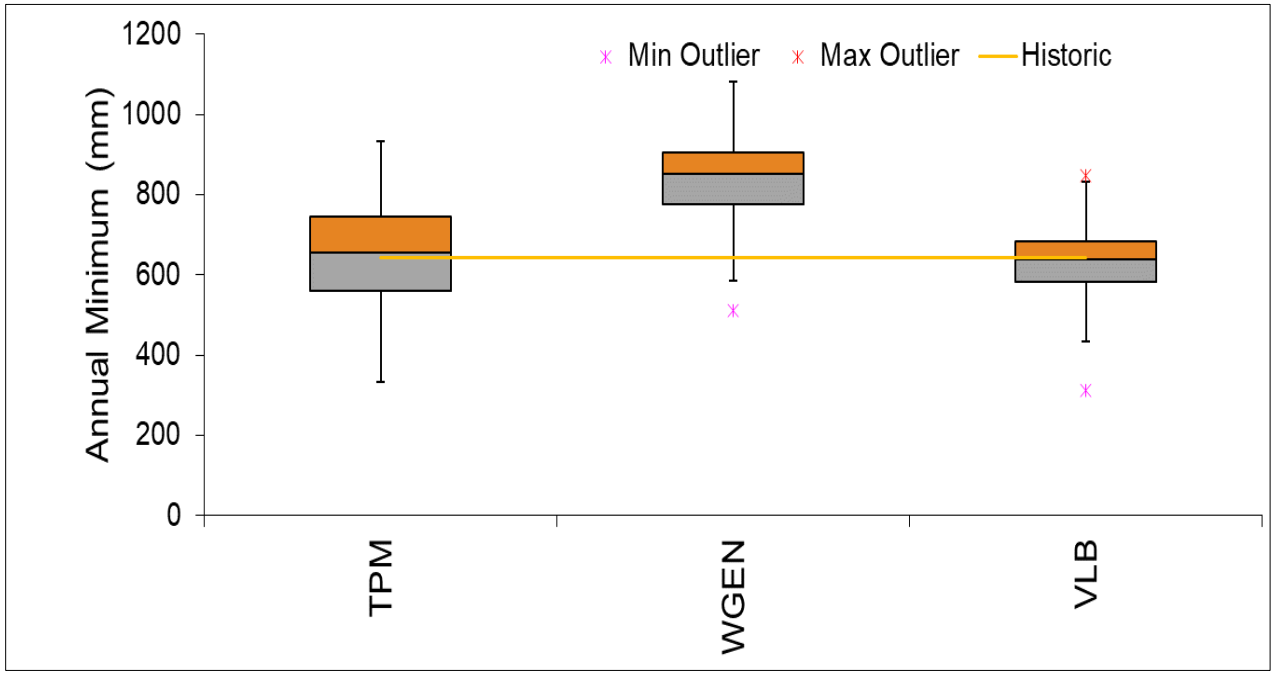


Figure 5.2-8: Comparison for the annual minimum values generated from the methods for the Welgevonden station

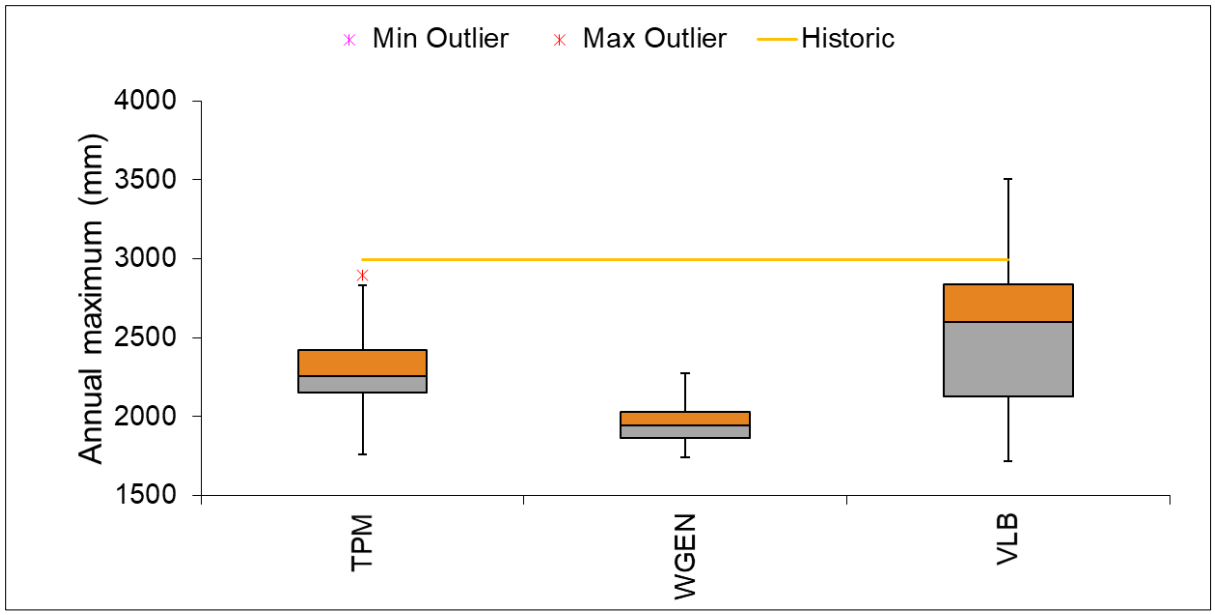


Figure 5.2-9: Comparison for the annual maximum values generated from the methods for the Welgevonden station

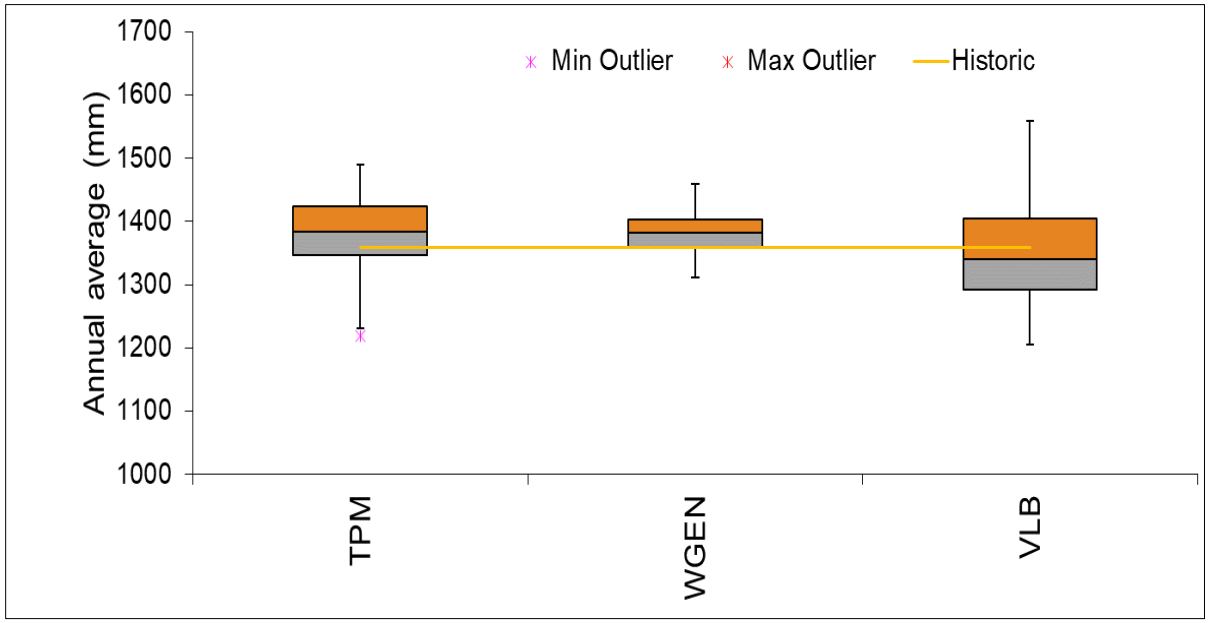


Figure 5.2-10: Comparison for the annual average values generated from the methods for the Welgevonden station

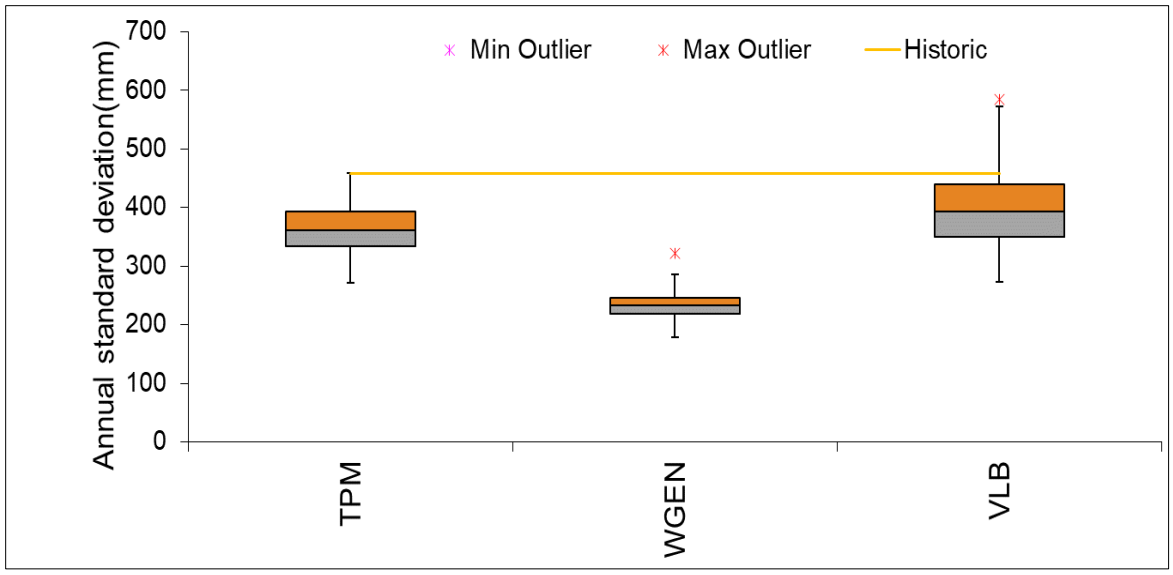


Figure 5.2-11: Comparison for the annual standard deviation values generated from the methods for the Welgevonden station

Table 5.2-1, Table 5.2-2 and Table 5.2-3 summarize the average statistics from 100 sequences generated using the three methods as compared to the observed historic data for the Welgevonden rainfall station. Similarly, to the statistics for the Hebron Station, WGEN method produced the lowest average daily maximum value of more than 100 mm difference with the observed. The average mean values from the daily statistics for the methods matched the historic value except for the value generated from VLB Method. The average monthly statistics also show a similar pattern in which the WGEN method generates data that have lower average maximum and skewness values compared to the other methods. The TPM Method generated average monthly statistics that is relatively similar to the historic data.

Table 5.2-1: Average statistics of the daily data generated using TPM, WGEN and VLB for the Welgevonden rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	3.8	3.8	3.8	3.7
Median (mm)	0.0	0.0	0.0	0.0

Minimum (mm)	0.0	0.0	0.0	0.0
Maximum (mm)	327.8	299.4	184	355.4
Standard deviation (mm)	12.3	12.6	11.1	13.0
Sample size	21 399	18 628	18 628	18 628
Mode	0.0	0.0	0.0	0.0
Skewness	6.9	7.1	4.9	7.7
5^{th} Percentile (mm)	0	0	0	0
95^{th} Percentile (mm)	21.7	22.7	24.5	21.9

Table 5.2-2: Average statistics of the monthly data generated using TPM, WGEN and VLB for the Welgevonden rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	115.8	116	115.8	113
Median (mm)	74.4	75	79.5	66.4
Minimum (mm)	0.0	0.0	0.0	0.0
Maximum (mm)	777.4	791	600	746.7
Standard deviation (mm)	127.4	126	113.4	131.5
Sample size	704	612	612	612
Mode	0.0	0.0	0.0	0.0
Skewness	1.7	1.7	1.2	1.7
5^{th} Percentile (mm)	0	0	0	0
95^{th} Percentile (mm)	384	370	341	388

Table 5.2-3: Average statistics of the annual data generated using TPM, WGEN and VLB for the Welgevonden rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	1 359.0	1 382	1 380	1348.4
Median (mm)	1 346.9	1 359	1 376.6	1 353.1
Minimum (mm)	644.4	644	836.6	623.3
Maximum (mm)	2 996.0	2 310	1951	2 536
Standard deviation (mm)	457.4	361	236	400.0
Sample size	60	51	51	51
Mode	-	-	-	-
Skewness	0.71	0.3	0.1	0.5
5^{th} Percentile (mm)	738.9	852	1023	752.0
95^{th} Percentile (mm)	2024.5	1977	1747	1974.6

*- refers to no value was found as a mode

5.3. Nelshoogte Rainfall station Comparison

Figure 5.3-1 to Figure 5.3-11 show the comparison of the methods for the daily, monthly, and annual statistics for Nelshoogte station. The daily statistics show that both TPM and WGEN compared relatively well for the daily average. With VLB method, the historic value falls within the upper extreme value of the box and whisker plot. The statistics did not fall within the box and whisker plots for WGEN method for all the statistics except for the daily average. For the monthly averages shown in Figure 5.3-5, all three methods compared relatively well to the historic monthly averages. With the monthly maximum values, all the methods performed similarly. The annual averages compared relatively well with the historic value for all the methods. For all the

other annual statistics, the historic values did not fall within the interquartile range of all the methods.

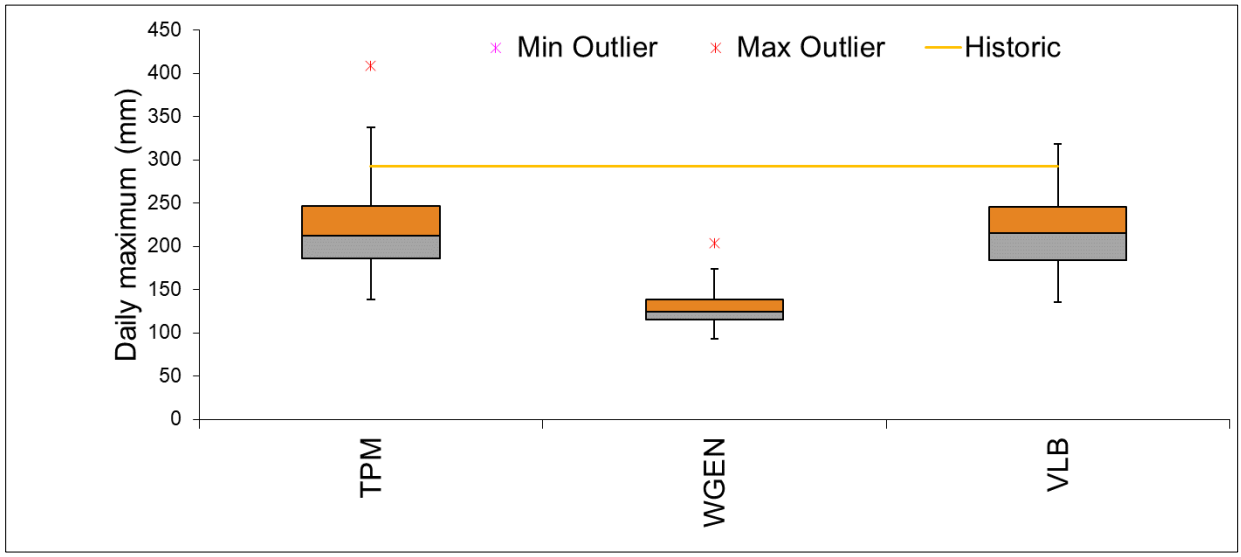


Figure 5.3-1: Comparison for the daily maximum values generated from the methods for Nelshoogte station

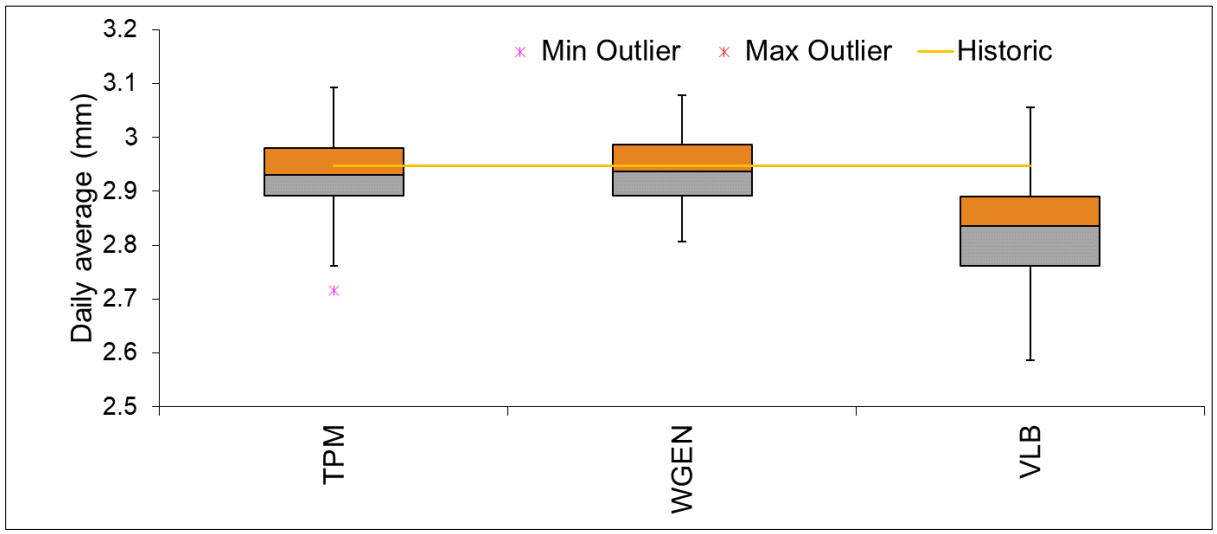


Figure 5.3-2: Comparison for the daily average values generated from the methods for Nelshoogte station

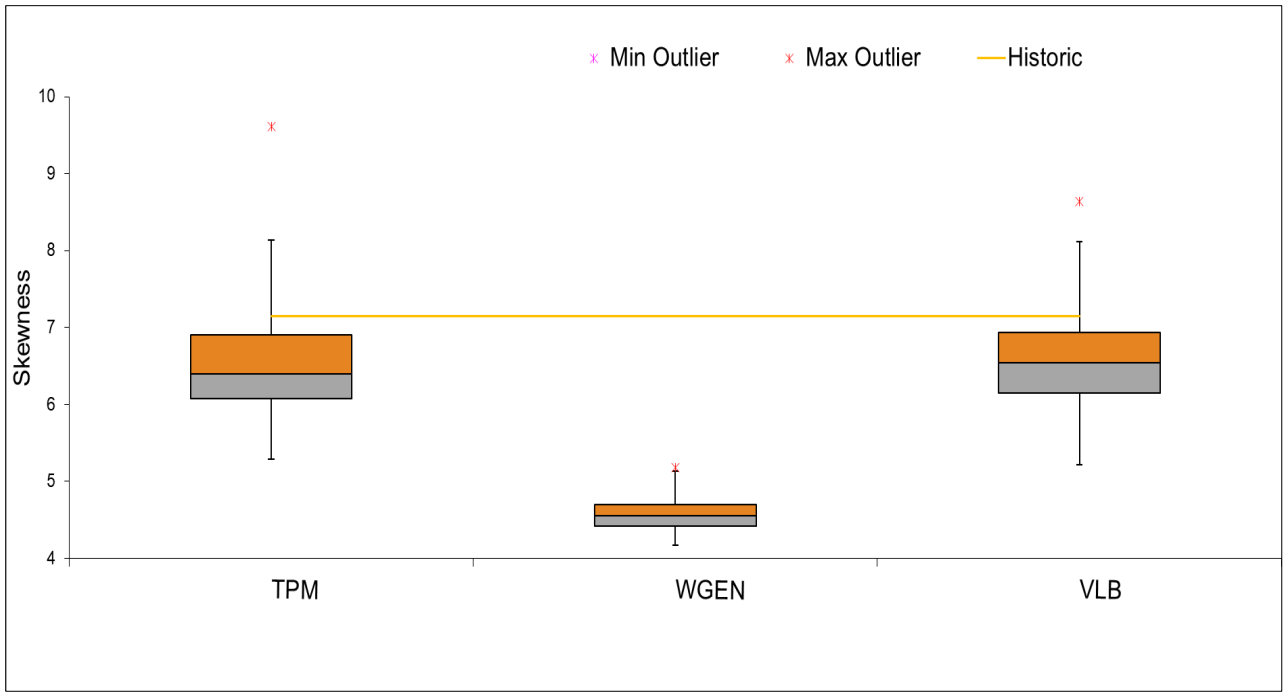


Figure 5.3-3: Comparison for the daily skewness values generated from the methods for Nelshoogte station

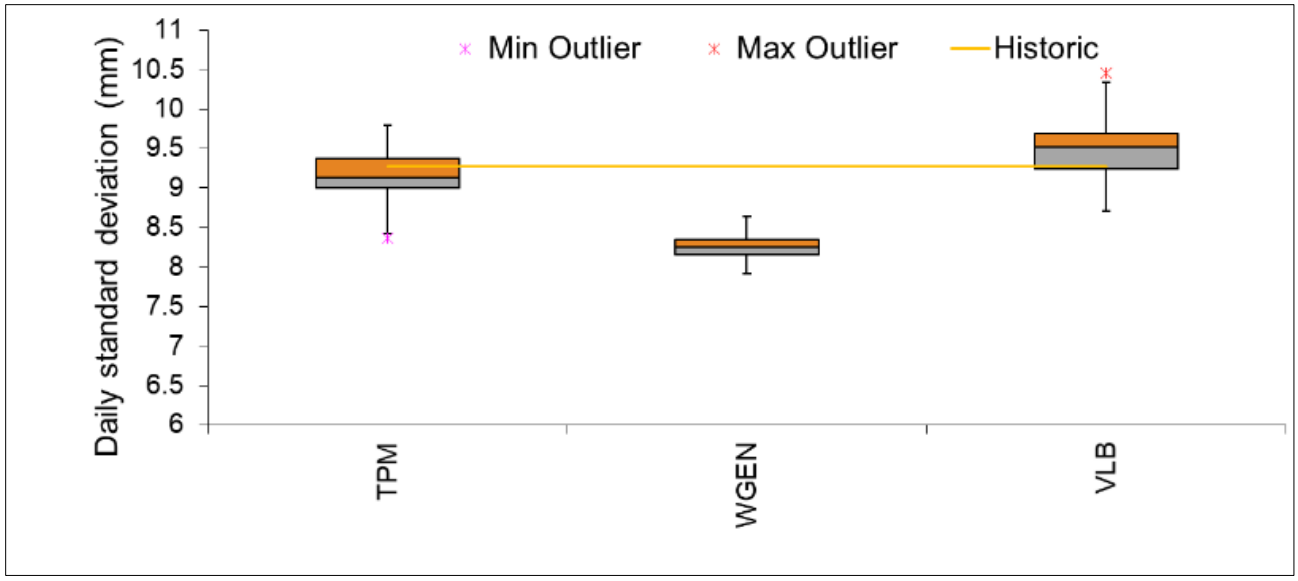


Figure 5.3-4: Comparison for the daily standard deviation values generated from the methods for Nelshoogte station

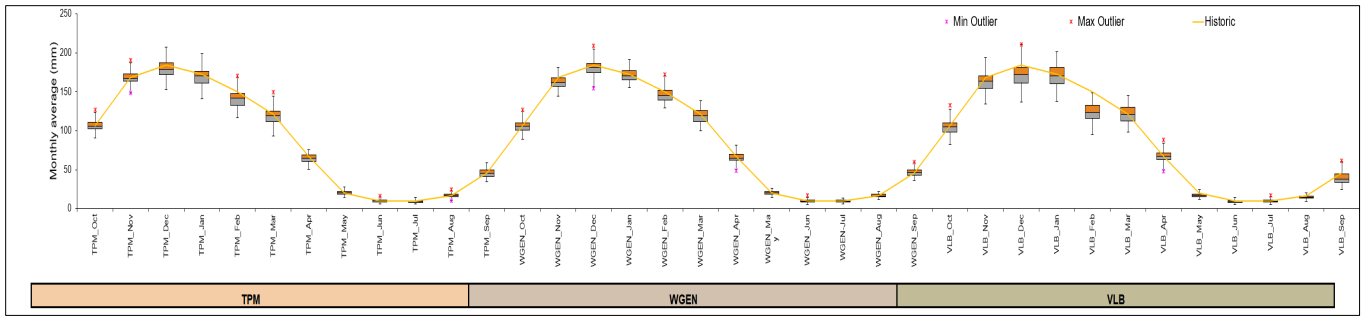


Figure 5.3-5: Comparison of the monthly averages for Nelshoogte station using VLB Station

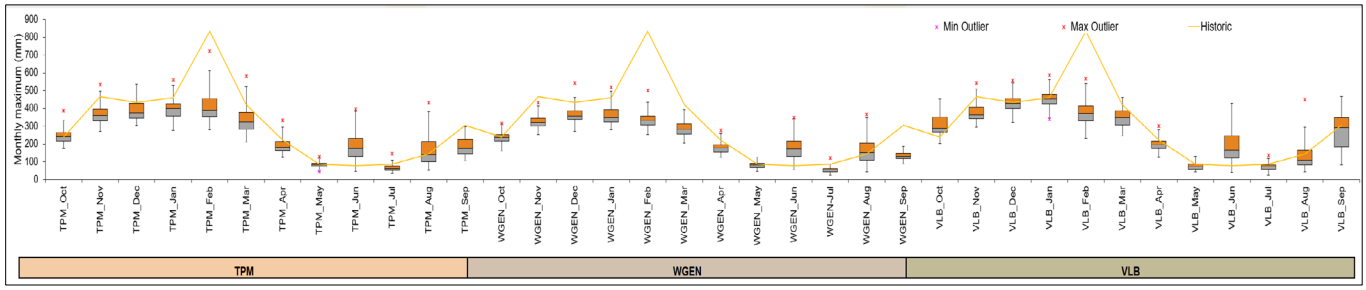


Figure 5.3-6: Comparison of the monthly maximum for the Nelshoogte station from the methods

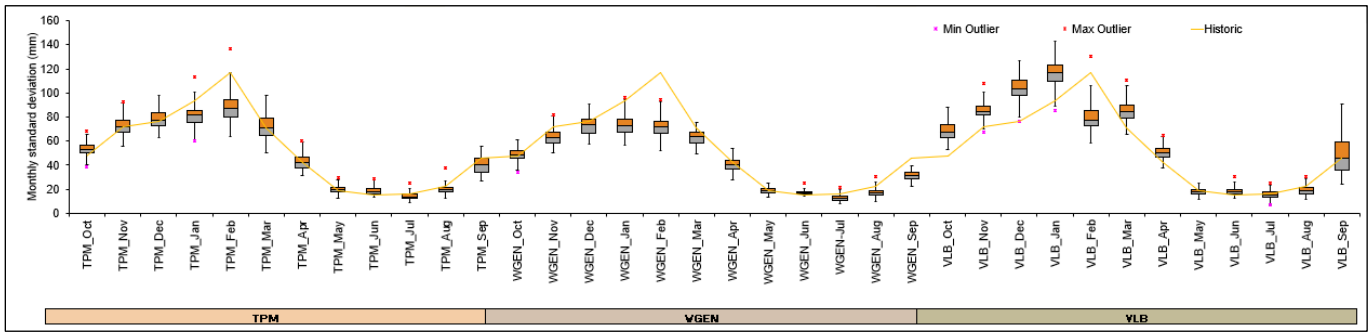


Figure 5.3-7: Comparison of the monthly standard deviation for the Nelshoogte station from the methods

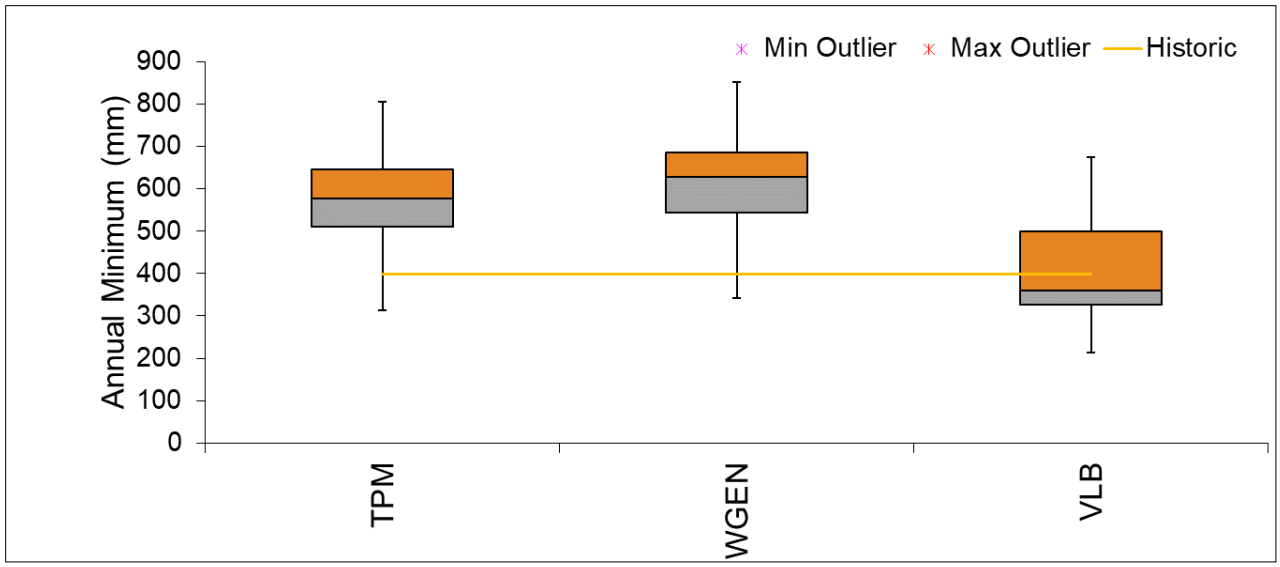


Figure 5.3-8: Comparison for the annual minimum values generated from the methods for the Nelshoogte station

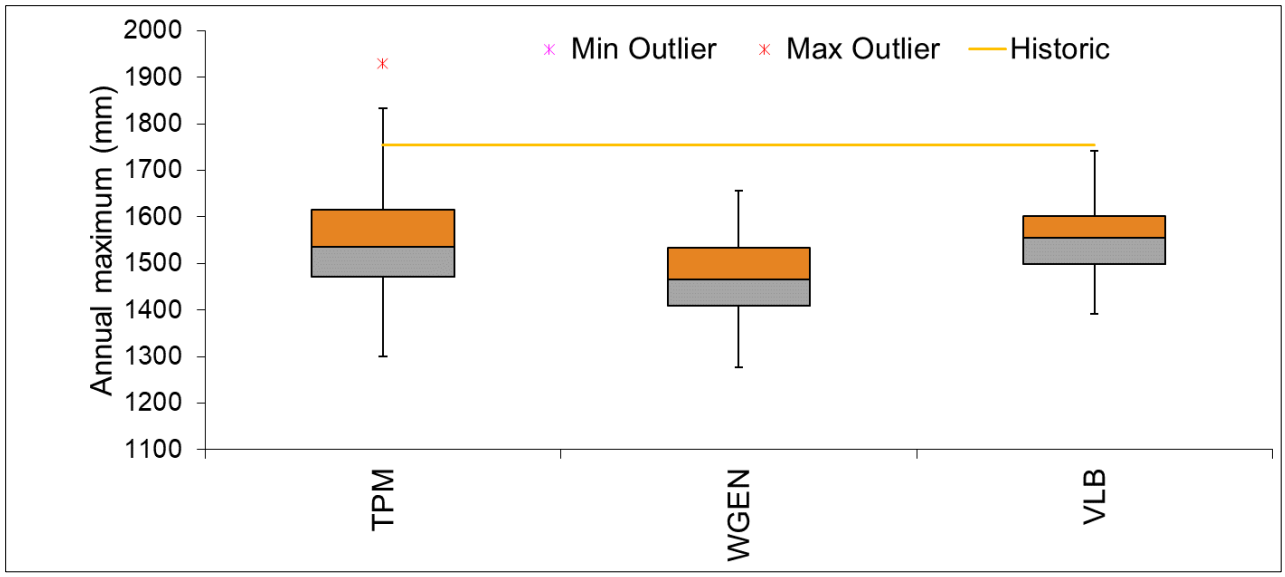


Figure 5.3-9: Comparison for the annual maximum values generated from the methods for the Nelshoogte station

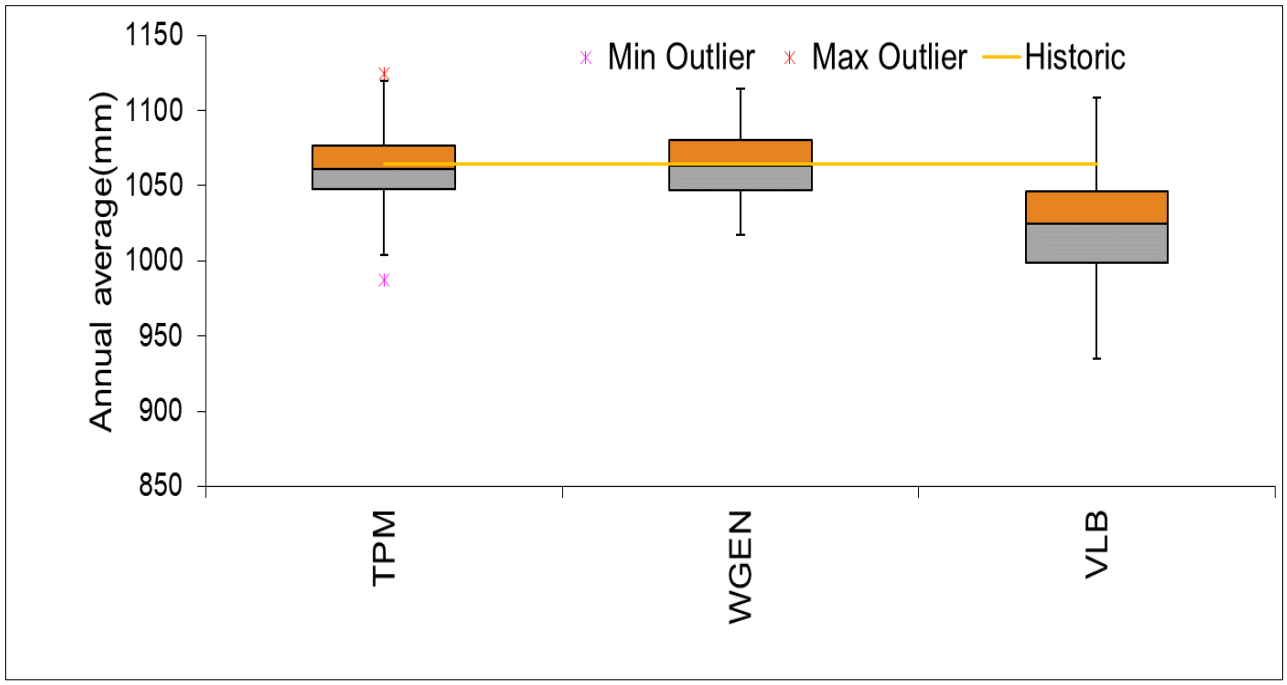


Figure 5.3-10: Comparison for the annual average values generated from the methods for the Nelshoogte station

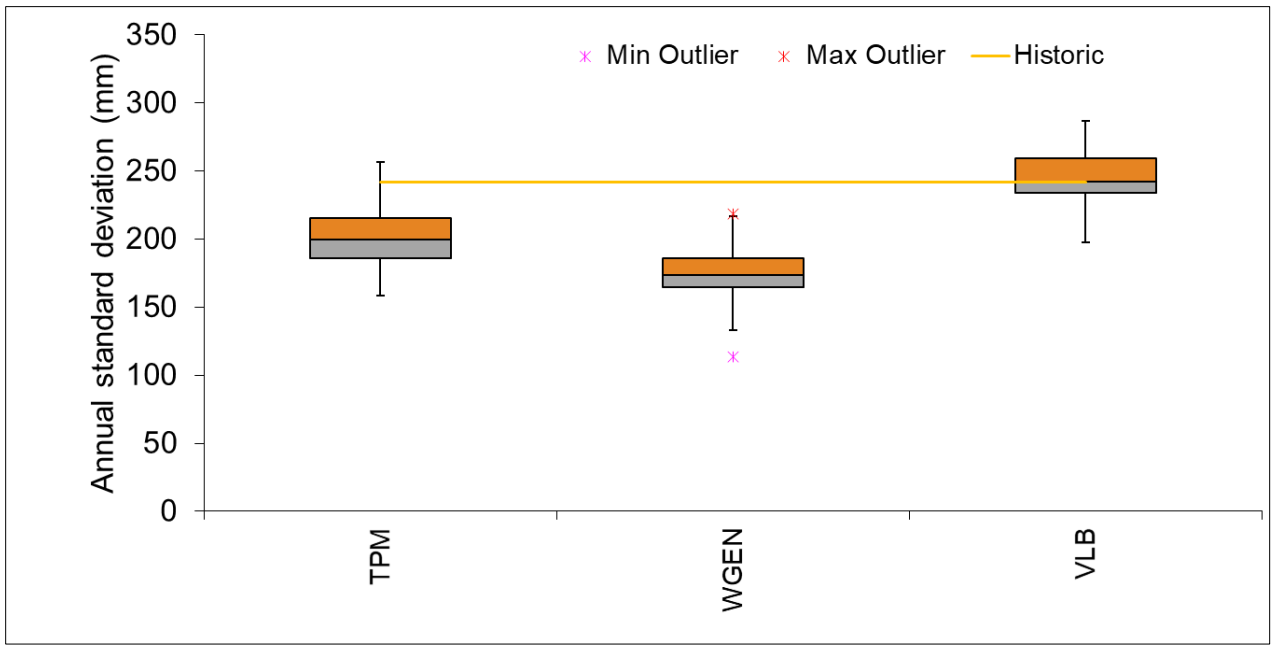


Figure 5.3-11: Comparison for the annual standard deviation values generated from the methods for the Nelshoogte station

Table 5.3-1,Table 5.3-2 and Table 5.3-3 summarize the daily, monthly, annual average statistics from 100 sequences generated using the three methods as compared to the observed historic data for the Nelshoogte rainfall station. TPM and VLB methods generated a similar value for the average daily maximum. TPM and WGEN methods generated a similar daily average mean value to the historic value. As seen in the previous performance evaluations, WGEN produced lower statistic values compared to the historic data for the daily average statistics. The monthly statistics interestingly show that the observed historic maximum is almost twice as high as the generated maximum values from the methods. With the average standard deviation value, VLB method generated value relatively similar to the historic value.

Table 5.3-1: Average statistics of the daily data generated using TPM, WGEN, and VLB for the Nelshoogte rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	2.9	2.9	2.9	2.8
Median (mm)	0.0	0.0	0.0	0.0
Minimum (mm)	0.0	0.0	0.0	0.0
Maximum (mm)	292.6	218.0	129.9	218.2
Standard deviation (mm)	9.3	9.2	8.2	9.5
Sample size	31 792	18 628	18 628	18 628
Mode	0.0	0.0	0.0	0.0
Skewness	7.2	6.6	4.6	6.6
5^{th} Percentile (mm)	0	0	0	0
95^{th} Percentile (mm)	18.2	19.0	18.8	18.0

Table 5.3-2: Average statistics of the monthly data generated using TPM, WGEN, and VLB for the Nelshoogte rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	89.3	89	89.4	86.0
Median (mm)	66.6	66	69.8	54.5
Minimum (mm)	0.0	0.0	0.0	0.0
Maximum (mm)	834.1	464.0	402.0	484.0
Standard deviation (mm)	90.3	86	81.6	93.3
Sample size	1 049	612	612	612
Mode	0.0	0.0	0.0	0.0
Skewness	1.6	1.1	0.9	1.3
5^{th} Percentile (mm)	0	0	0	0
95^{th} Percentile (mm)	260.0	254.0	243.7	275.2

Table 5.3-3: Average statistics of the annual data generated using TPM, WGEN and VLB for the Nelshoogte rainfall station

Description	Observed	TPM	WGEN	VLB
Mean (mm)	1 064.6	1 060	1 064.5	1 022
Median (mm)	1 089.0	1 054	1 062.4	1 020.6
Minimum (mm)	398.2	644.0	615.0	411.4
Maximum (mm)	1 755.6	1 546	1 472	1 553.3
Standard deviation (mm)	241.9	201	175.2	245.6

Sample size	88	51	51	51
Mode	-	-	-	-
Skewness	0.06	0.1	-0.1	-0.1
5^{th} Percentile (mm)	732.2	759.2	801	672.0
95^{th} Percentile (mm)	1462.4	1382	1336.6	1412.0

*- refers to no value was found as a mode

6. CONCLUSIONS AND RECOMMENDATIONS

This study aimed to assess the performance of three daily stochastic rainfall generators; the Transition Probability Matrix (TPM), the Weather Generator (WGEN) and Variable Length Block (VLB) Methods. The performance assessment of all three methods based on preservation of key statistical variables found that all the methods are capable of generating data that preserve the statistics of the observed historic data. The methods however replicate the various statistics to different degrees of accuracy. It can be concluded that the methods can satisfactorily generate daily stochastic rainfall for use in practical hydrological analyses.

It was found that the WGEN method tends to generate values that have standard deviation and skewness values lower than the historic data. TPM and VLB methods have generated daily rainfall depths that have a relatively higher variability than WGEN and that compare relatively better to the historic values. All three methods generate average daily, monthly, and annual data that compare well to the historic values. The generated data from the TPM and VLB methods is more conservative and slightly higher than the observed historic data. Interestingly, the study has also highlighted that on some occasions, the methods did not generate maximum values that are higher than the observed historic data; and this was particularly predominant for the maximum monthly values for February, the month with the highest rainfall. In terms of the monthly patterns of dry and wet seasons of the generated data, all methods were able to preserve the trend of the historic data. The comparison of the methods showed that data generated using WGEN indicated much lower maximum values for daily, monthly, and annual averages. If the modelling requires a high rainfall such as flood related projects, it is recommended that either TPM or VLB methods be applied for the generation of the data as a more conservative approach. And for modeling requiring lower rainfalls such as water supply related projects, then WGEN may be preferable as a more conservative approach.

The study has revealed that it is imperative to have a comprehensive understanding of both advantages and disadvantages of the methods and their limitations. Although some of the methods were originally not developed with intention of being applied in South Africa, they can generate data for different climatic conditions that they were not initially developed for. It is recommended that future research focus more on how climate change can be incorporated into the methods. It was also observed from the methods that for the monthly maximum and standard deviation values, the month of February generated values were lower than the historic values, further analysis is recommended to assess the performance of February in more detail. Additional performance assessment for other areas of Southern Africa including the winter rainfall regions of the Western Cape Province and drier areas of South Africa is recommended. Further studies on the uncertainty aspects of the methods, particularly with WGEN method on Southern African data is also recommended.

7. REFERENCES

Eisinger, D., & Wiegand, K. (2008). SERGE: A spatially explicit generator of local. *South African Journal of Science 104*, 37-42.

GrÄoger, R. (2004). *Introduction to programming in Fortran 77 for students of Science and Engineering.* University of Pennsylvania, Department of Material sciences and Engineering .

Valters, D. (2022, 03 05). *Coding Club*. Retrieved from https://ourcodingclub.github.io/tutorials/fortran-intro

Adams, J. C., Brainerd, W. S., Martin, J. T., Smith, B. T., & Wagener, J. L. (1992). *Fortran 90 Handbook.* New York: McGraw-Hill Book Company.

Agbo, E. P. (2021). The role of statistical methods and tools for Weather Forecasting and Modelling. *IntechOpen*, DOI: 10.5772/intechopen.96854.

Boughton, W. C. (1999). *A daily rainfall generating model for water yield and flood studies.* Australia: Cooperative Research Centre for Catchment Hydrology (99/9).

Boughton, W. C. (2009). Multi-year streamflow drought in eastern Australia. *Australasian Journal of Water Resources 13*, 31 - 42.

Boughton, W. C., & Hill, P. I. (1997). *A Design flood estimation procedure using data generation and a daily water balance model.* Report 97/8. Cooperative Research Centre for Catchment Hydrology.

Centre for Water Resources Research. (2015). *Development and Assessment of an Intergrated Water Resources Accounting Methodology for South Africa.* Pretoria: Water Research Comission (WRC Report No 2205/1/15).

Chapman, T. (1994). Stochastic models for daily rainfall. Hydrology and Water Resources Symposium. *The institution of Engineers,Australia, National Conference Publication 94/15*, 7–12.

Corporate Research Centre Australia. (2003). *Catchment hydrology annual report.* Australia: Cooperative Research Centre for Catchment Hydrology.

Dam Safety Office. (2003). *Stochastic Modeling Methods.* United States Department of Interior Bureau of Reclamation Report No. DSO-03-04.

De Vera, A., & Terra, R. (2018). A Stochastic Precipitation Generator Conditioned by a Climate Index. *15588432 - Journal of Applied Meteorology and Climatology*, 2805-2606.

GoldSim Technology Group. (2014). *The GoldSim Volume 1 User's Guide.* GoldSim Technology Group LLC.

GoldSim Technology Group. (2021, September 13). Retrieved from GoldSim: https://www.goldsim.com/

Google Earth. (2022, March 21). Retrieved from Google Earth: https://www.google.com/earth/index.html

Haan, C. T. (1977). *Statistical Methods in Hydrology.* The Iowa State University. 378 pp. Press, Ames.

Harrorld, T. I., Sharma, A., & Sheather, S. J. (2002). *Representation of long-term variability in daily rainfall generation.* Australia: Hydrology and Water Resources Symposium.

Maass, L. F. (2017). *Comparison of Stochastic streamflow generators and the use thereof within the water resources yield model and Mike Hydro Basin.* Cape Town: MSc Report, Department of Water Engineering, University of Stellenbosch.

McMahon, T. A., & Srikanthan, R. (1985). Stochastic generation of rainfall and evaporation data. *AWRC*, Technical Paper No. 84, 301pp.

Montanari, A., & Brath, A. (2004). A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research* , 40 2003WR002540.

Muller, A., Arnaud, P., Lang, M., & Lavabre, J. (2009). Uncertainties of extreme rainfall quantiles estimated by a stochastic rainfall model and by a generalized Pareto distribution. *Hydrological Sciences Journal*, 54:3, 417-429, DOI: 10.1623/hysj.54.3.417.

Muller, M. (2017). *Understanding the origins of Cape Town's water crisis.* Cape Town: ICE Proceedings Civil Engineering 25(5) 11-26.

Nathan, R. J., & Wang, Q. J. (2007). A method for coupling daily and monthly time scales in stochastic generation of rainfall series. *Journal of hydrology*, Volume 346, Issues 3–4, Pages 122-130.

Ndiritu, J. (2011). A variable-length block bootstrap method for multi-site synthetic streamflow generation. *Hydrological Sciences Journal 56:3*, 362-379.

Ndiritu, J., & Nyaga, J. (2014). *A non-parametric multi-site stochastic rainfall model with applications to climate change.* Johannesburg: Water Research Commission. WRC Report No. 2148/1/13.

Nyaga, J. (2014). *The Use of Emprical Mode Decomposition(EMD) and Variable Length Bootstrap (VLB) for Stochastic Rainfall Generation.* Johannesburg: Unpublished MSc Report , University of the Witwatersrand.

Pegram, G. G., & Clothier, A. N. (2002). *Space time modelling of rainfall using the string of beads model: Integration of radar and rain gauge data.* Durban, South Africa: Water Research Commission Report No 1010/1/02.

Phakula, S., Landman, W. A., & Beraki, A. F. (2018). Forecasting seasonal rainfall characteristics and onset months. *International Journal of Climatology*, 38. 10.1002/joc.5417.

Punkkinen, H., Räsänen, L., Maija Mroueh, U., & Korkealaakso, J. (2016). *Guidelines for mine water management .* Finnish Environment Institute (SYKE).

Richardson, C. W. (1982). *A comparison of three distributions for the generation of daily rainfall amounts.* Colorado: Water Resources Publications, 67-78 .

Richardson, C. W., & Wright, D. A. (1984). WGEN: A model for generating daily weather variables. *U. S. Department of Agriculture*, 8 (83p).

Sausa, P. M., Blamey, R. C., Reason, C. J., Ramos, A. M., & Trigo, R. M. (2018). The 'Day Zero' Cape Town drought and the poleward migration of moisture corridors. *Environmental Research Letters*, Lett. 13 124025.

Siriwardena, L., Srikanthan, R., & McMahon, T. (2002). *Evaluation of two daily rainfall data generation models.* Cooperative research centre for catchment hydrology,Technical Report 02/14.

Smith, R. E., & Schreiber, H. A. (1974). *Point processes of seasonal thunderstorm rainfall. Rainfall depth probabilities.* Water Resource Res. 10(3):418-423.

Srikanthan, R., & McMahon, T. A. (1985). *Stochastic generation of rainfall and evaporation.* AGPS: Australian Water Resources Council, Tech. Pap. no.84,.

Srikanthan, R., & McMahon, T. A. (2001). Stochastic generation of annual, monthly and daily. *Hydrology and Earth System Sciences Discussions, European Geosciences Union*, 5 (4), pp.653-670.

Srikanthan, R., & Zhou, S. (2003). *Stochastic generation of climate data.* Cooperative research centre for catchment hydrology, Technical Report 03/12 .

Srikanthan, R., Harrold, T. I., Sharma, A., & McMahon, T. A. (2005). Comparison of Two Approaches for Generation of Daily. *Stochastic Environmental Research and Risk Assessment*, 215-226.

Thyer, M., & Kuczera, G. (1999). Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms . *Water Resources Research* , 767-773.

Tutorialspoint. (2022, 03 06). *Fortran Overview*. Retrieved from Tutorialspoint: https://www.tutorialspoint.com/fortran/fortran_overview.htm

Van Niekerk, A. M. (1997). *Generic simulation model for opencast Mine Water Systems.* Pretoria: Water Research Commission,Report No. 528/1/97.

Waymire, E., & Gupta, V. K. (1981). The Mathematical Structure of Rainfall Representations: A review of the stochastics rainfall models. *Water Resources Research Vol 17, No 5* , 1261-1272.

Wilcox, C., Aly, C., VISCHE, T., PANTHOU, G., Blanchet, J., QUANTIN, G., & Lebel, T. (2020). Stochastorm: A Stochastic Rainfall Simulator for Convective Storms. *Journal of Hydrometeorology*, 22 (387-403).

Wilks, D. S. (1995). Statistical weather forecasting. *Elsevier : International Geophysics*, 159-232. DOI: 10.1016/S0074-6142(06)80042-2.

Woolhiser, D. A., & Roldan, J. (1982). *Stochastic daily precipitation models. A comparison of distribution of amounts.* Water Resources Res. 18(5): 1461-1468.

Zhao, Y., Nearing, M. A., & Guertin, D. P. (2019). A daily spatially explicit stochastic rainfall generator for a semi-arid climate. *Journal of Hydrology*, 574 (181-192).

Zhou, Y., Yu, Z. J., Li, J., Huang, Y., & Zhang, G. (2017). The Effect of Temporal Resolution on the Accuracy of Predicting Building Occupant Behaviour based on Markov Chain Models. *ScienceDirect: 10th*

International Symposium on Heating, Ventilation and Air Conditioning, ISHVAC2017, Procedia Engineering 205 (2017) 1698–1704.

Zucchini, W., Adamson, P. T., & McNeill, L. (1992). A model of southern African Rainfall. *S. Afr. J. Sci*, 103-109.