A NON-PARAMETRIC MULTI-SITE STOCHASTIC RAINFALL MODEL WITH APPLICATIONS TO CLIMATE CHANGE

Report to the
Water Research Commission

by

John Ndiritu and Job Nyaga
School of Civil and Environmental Engineering
University of the Witwatersrand

WRC Report No. 2148/1/13
ISBN 978-1-4312-0515-8

March 2014

Obtainable from

Water Research Commission
Private Bag X03
Gezina, 0031

orders@wrc.org.za or download from www.wrc.org.za

DISCLAIMER

This report has been reviewed by the Water Research Commission (WRC) and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the WRC, nor does mention of trade names or commercial products constitute endorsement or recommendation for use

Executive Summary

The planning and operation of most of the large water resource systems in South Africa has been applying a multi-site monthly streamflow generator since the 1990s but it has recently been recognized that the use of stochastic rainfall generation may hold several advantages over the stochastic streamflow generation. Since rainfall is the main input into the hydrological cycle, applying stochastics on rainfall rather than streamflow is naturally more inclusive. With stochastic rainfall, probabilistic analysis can be included more realistically and easily in the analysis of catchment hydrological processes and rainfall-dependent activities such as irrigation. The impacts of climate change and increasing variability on basin hydrology and water resources can also be studied with more ease with a rainfall rather than a streamflow stochastic generator. Consequently, the Department of Water Affairs (DWA) commissioned the development of a parametric stochastic monthly rainfall generator (PEGRAIM-W). The PEGRAIM-W generator is currently being tested in another WRC project (WRC Project K5/2155).

Although parametric approaches have dominated stochastic hydrology, they exhibit several limitations in comparison to non-parametric approaches. They require the data to be fitted to specific probability distributions while non-parametric methods do not. Parametric methods also typically use large numbers of parameters unlike most non-parametric methods. The ease of use and simplicity of some parametric methods has often led to their preference to parametric methods. The development of an effective and efficient non-parametric rainfall generator is therefore likely to add significant value to water resources modelling in South Africa.

Like in other parts of the world, climate change is a major concern in South Africa and climate change is currently a very active field of research. In South Africa, climate change impacts have however not specifically been incorporated into the comprehensive probabilistic approach that is applied by DWA and its consultants in long-term and operational planning of water resource systems.

Given the background, this Water Research Commission project contracted to the University of the Witwatersrand (School of Civil and Environmental Engineering) set out to develop and test a monthly non-parametric stochastic rainfall generator that would comprehensively incorporate climate change and changing variability including information from global climate model (GCM) projections. The project also aimed at comparing the non-parametric generator with the parametric PEGRAIM-W generator. These objectives were achieved as follows:

i) A literature review of stochastic hydrologic data generation (mainly rainfall and streamflow) that enabled the selection of the non-parametric method to use. The Variable Length Bootstrap (VLB) approach that had previously been applied to streamflow generation was selected out of a group of 5 non-parametric methods.

ii) A literature review on climate change and variability research aimed at finding an appropriate manner of incorporating climate change/variability into stochastic rainfall generation. This revealed that global climate models (GCMs), the only physically-based approaches for long-term climate predictions fail to replicate long-term persistence and other important hydro-climatic characteristics.

iii) The adaptation of the VLB stochastic model to rainfall generation based on the observed temporal and spatial characteristics of rainfall. A multiple rainfall generation problem consisting of 10 sites spread out in South Africa was used for this.

iv) An assessment of the performance of the VLB generator by comparing 11 annual and 10 monthly statistics obtained from the generator to historic ones.

v) The comparison of the VLB rainfall generator with PEGRAIM-W, a parametric rainfall generator currently being tested in an on-going WRC project (WRC Project K5/2155). The VLB was found to perform slightly better than PEGRAIM-W at the annual and the monthly time scale.

vi) The development and testing of climate variability modelling by appropriately biasing block selection of the VLB generator to obtain a drier, a wetter or a more variable climate. This approach produced stochastic sequences of highly varied characteristics depending on the settings used. It was easy to generate rainfalls that match the long-term average shifts in mean annual precipitation (MAP) and seasonal characteristics predicted by multiple GCMs while maintaining (or increasing) the expected inter-decadal variability. Given that GCMs fail to model inter-decadal variability, the approach developed here effectively complements these projections. This part of the project applied another 10-site rainfall generation problem located in the Western Cape of South Africa.

vii) The development of a user-friendly GUI for the rainfall generator. This is included separately in the accompanying CD to this report.

For capacity building, Mr Job Nyaga, currently an MSc student in the School of Civil and Environmental Engineering of the University of the Witwatersrand has been working on this project and has acquired considerable knowledge in stochastic hydrology and skills for conducting and reporting research. The main objective of his MSc research is to find out how effectively mode decomposition (EMD) can be used as an alternative block termination method to the method used in the current VLB generator. Preliminary results from this analysis are presented in Appendix A.

Acknowledgements

The project team is very grateful to the Water Research Commission for providing financial and administrative support for the project. The project team also thanks the members of the Reference Group for their valuable contributions that added significantly to the success of the project. The Reference Group members are:

Mr W Nomquphu	:	Water Research Commission (Chairperson)
Prof G G S Pegram	:	Pegram and Associates
Prof D Hughes	:	Rhodes University
Prof J Smithers	:	University of KwaZulu-Natal
Prof F Ilunga	:	University of South Africa
Dr W Nyabeze	:	Washington Nyabeze and Associates
Dr B Mwaka	:	Department of Water Affairs
Mr P van Rooyen	:	WRP Consulting
Mr E Nel	:	Department of Water Affairs
Mr B Haasbroek	:	Hydrosol Consulting
Mr N Malose	:	Department of Water Affairs

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACC	Annual Cross Correlation
ASC	Annual Serial Correlation
CMIP5	Coupled Model Intercomparison Group Project Phase 5
CSAG	Climate Systems Analysis Group
DWA	Department of Water Affairs
EMD	Empirical Mode Decomposition
GCM	Global Climate Model (or General Circulation Model)
k-NN	k Nearest Neighbour
MAP	Mean Annual Precipitation
RCM	Regional Climate Model
RCP	Representative Concentration Pathways
STOMSA	Stochastic Model of South Africa
TM	Transition Matrix
VLB	Variable Length Bootstrap

1 INTRODUCTION

The planning and operation of most of the large water resource systems in South Africa has been applying a multi-site monthly streamflow generator since the 1990s (Pegram and McKenzie, 1991; Basson et al., 1994; Van Rooyen and McKenzie, 2004) but it has recently been recognized that the use of stochastic rainfall generation may hold several advantages over the stochastic streamflow generation. Since rainfall is the main input into the hydrological cycle, applying stochastics on rainfall rather than streamflow is naturally more inclusive. With stochastic rainfall, probabilistic analysis can be included more realistically and easily in the analysis of catchment hydrological processes (e.g. sediment generation, pollutant transport) and rainfall-dependent activities such as irrigation. The impacts of climate variability/change on basin hydrology and water resources can also be studied with more ease with a rainfall rather than a streamflow stochastic generator. Consequently, the Department of Water Affairs (DWA) has commissioned the development of a stochastic monthly rainfall generator that is based on a recently developed daily rainfall generator (Srikanthan and Pegram, 2009). A review of the literature indicates that most stochastic rainfall generators are either daily or sub-daily and may therefore not be computationally efficient for the monthly-time step water resources assessment typical to South Africa.

Although parametric approaches have dominated stochastic hydrology, they exhibit several limitations in comparison to non-parametric approaches. They require the data to be fitted to a specific probability distribution while non-parametric methods do not. Parametric methods also typically use large numbers of parameters unlike most non-parametric methods. The ease of use and simplicity of some parametric methods (especially the bootstrap) has often led to their preference to parametric methods (Vogel and Shallcross, 1996). The Variable Length Block (VLB) bootstrap, a non-parametric stochastic streamflow generator developed recently (Ndiritu, 2011) has been found to match (and out match in some aspects) the performance of the STOMSA parametric model (Van Rooyen and McKenzie, 2004) that is routinely applied in water resource yield and planning analysis in South Africa.

Like in other parts of the world, climate change is a major concern in South Africa and climate change is currently a very active field of research in South Africa. Climate change has however not specifically been incorporated into the comprehensive probabilistic approach that is applied by DWA and its consultants in long-term and operational planning of water resource systems (Basson et al., 1994; Basson and Van Rooyen, 2001). Global climate models (GCMs) are the most common methods applied for long-term climate forecasting although they are mostly not validated before application (Kundzewicz and Stakhiv, 2010) and have generally perform poorly in validation tests (Anagnostopoulos et al., 2010). The common approach of using multiple GCMs as a means of incorporating uncertainty is found to grossly under-estimate the uncertainty (Koutsoyiannis, 2011).

Given the background, this Water Research Commission project contracted to the University of the Witwatersrand (School of Civil and Environmental Engineering) set out to develop and test a monthly non-parametric stochastic rainfall generator that would comprehensively incorporate climate change/variability including information from GCM projections. The project also aimed at comparing the non-parametric generator with a parametric one. This is the final report of the project.

2 LITERATURE REVIEW

2.1 Introduction

This chapter reports the review of stochastic hydrologic generation and climate variability modeling. The literature review includes both rainfall and streamflow generation methods as many of the methods applicable to streamflow generation could be used for stochastic rainfall generation (e.g. Srikanthan et al., 2002; Mehrotra and Sharma, 2007). A review of the relevant studies on stochastic hydrologic generation is presented in Section 2.2 from which the non-parametric approaches to be considered for stochastic rainfall generation are identified. The identified non-parametric methods are then reviewed to more detail in Section 2.3 in order to select a single non-parametric single method to apply. Section 2.4 reviews climate change/variability with the aim of finding an appropriate approach to incorporate climate change/variability in the rainfall generation. A summary of the literature review then follows in Section 2.5.

2.2 Analysis of Studies on Stochastic Rainfall and Streamflow Generation

A search of studies on stochastic rainfall and streamflow generation was conducted and those applying daily or/and monthly time intervals were selected for analysis. The analysis includes both non-parametric, parametric and hybrids of the two for completeness and to broaden the researchers' scope of knowledge on issues of hydrologic data generation other than those related to the process of generation only. For the studies selected, Table 1 describes in brief; i) the main objective/s, ii) the case study area/s data, iii) the main methods applied and iv) the main findings/outcomes of the study. It is recognized that this selection of 34 studies in Table 1 is in no way exhaustive but is considered to be adequately comprehensive. It is also intended that the review of the literature will continue for the duration albeit at a slower pace.

Table 2.1 Analysis of stochastic rainfall and streamflow generation studies

References	Objectives	Study area and/or data	Main methods involved	Main findings/outcomes
Bayazit et al. (2001)	To compare the relative performances of wavelet and Fourier analysis for annual stochastic streamflow generation.	40 year long annual flow series at Homa station on Manavgat river in Turkey.	Decomposition of the observed series into components at various frequencies and then random reconstruction to obtain a synthetic series by use of wavelet and Fourier analysis.	The mean and standard deviation is preserved but skewness of non-normal series is not replicated. Wavelet analysis is more suitable than Fourier analysis in the simulation of streamflow series because of better analysis of signals with sharp spikes as typical streamflow series.
Brissete et al. (2007).	To develop a simple, fast, efficient and easy to implement algorithm for the stochastic generation of daily and monthly multi-site precipitation. To deal with the spatial intermittence problem.	Synthetically generated data for 8 stations near Periboka river basin, Quebec, Canada.	A variant of the Wilks (1998) approach that effectively deals with the intermittence problem.	An algorithm performs adequately and deals satisfactorily with the spatial intermittence problem.
Clark et al. (2004)	To develop a method for reconstructing daily space-time variability in forecasted precipitation and temperature fields using a reordering approach.	Rainfall and temperature series from 2307 stations across the USA.	Re-ordering of the ensemble outputs in order to recover the space-time variability of precipitation and temperature fields in the historical data. The reordering is based on the order observed from controlled samples of the historic data.	The method, referred to as the Schaake Shuffle, almost entirely recovers the inter-variable correlations, inter-site correlations and the observed temporal persistence.
Eisinger and Wiegand (2008)	Description of a stochastic and spatially explicit generator of local daily rainfall in Southern Africa's arid environments.	Daily rainfall records throughout South Africa used to calibrate rainfall generator (Zucchini and Adamson (1984)) that is improved in this study.	A parametric daily rainfall generator (SERGE) that generates rain using several randomly positioned rain clouds	The SERGE model preserves the long term characteristics at each point and also the spatial autocorrelation of variable length. SERGE can be used in nature conservation and range management.
Ilich and Despotovic (2008)	To develop a simple method for effective multi-site generation of stochastic hydrologic time series at a monthly time scale and the construction of a covariance	Data from 4 sites each consisting of 84 years of historic weekly flows from Southern Alberta, Canada.	Simulation of random variables with arbitrary covariance structure and; (1) generation of weekly flows (2) Reordering the weekly flows to construct the desired correlations, (3)	A close match of the important statistics between the historic and synthetic flow series' is obtained.

	structure that maintains annual autocorrelation.		reordering the simulated years to conform to remaining statistics. The generation of an algorithm for re-ordering the generated subsets of each synthetic year is done by the use of a correlation matrix	
Kim et al. (2008)	To develop and evaluate a daily multi-site stochastic rainfall generator that incorporates spatial temporal dependences that explicitly reduce the complexity in the simulation of daily rainfall amounts.	Multiple gauging locations in South Florida USA.	A Markovian method to represent the temporal dependence and direct acyclic (DAG) graph to encode the spatial dependence of daily rainfall among stations.	The method performed better than the conventional Markov type model in the representation of daily spatial-temporal dependencies and is thus effective in reducing the complexity in the generation of rainfall.
Lall and Sharma (1996)	To resample scalar or vector valued hydrological time series by the use of the nearest neighbour bootstrap in order to preserve the dependence structure.	Monthly flows from Weber River near Orkney, USA.	A nearest neighbour bootstrapping method of stochastic data generation	The method preserved historic statistics adequately. Effective resampling from k-nearest neighbours in which the kennel decreases with distance and adapts to the local sampling density.
Lin and Lee (1992)	To develop and test a method that applies both aggregation and disaggregation approaches and thereby replicate over-year seasonal correlations.	47 years of streamflows from Tanshui River at Gueishan, Taiwan.	Two autoregressive moving average (ARMA) models.	The model preserves over year seasonal correlations that the model of Mejia and Rousselle (1976) and its many improvements are not able to do. The methodology is considered to harmonize aggregation and disaggregation approaches in parametric stochastic generation.
Maheepala and Perera (1996)	To provide an improved disaggregation method that explicitly preserves the over-year monthly serial and cross correlations as well as other monthly and annual parameters.	Single site data o five Australia streams that have variable annual flows and multi-site data from an Australian river basin.	Disaggregation by the use of modified synthetic fragments; 1) Generation of representative monthly series monthly series 2) generation of an annual series that preserves the annual parameters and 3) disaggregation of the annual series in (2) using the monthly fragments in (1).	The method preserves the auto and cross correlations as well as cross-yearly correlations. A good disaggregation approach results if the historic data that is not considered of good quality is excluded from the disaggregation process.
Mehrotra and Sharma	To formulate and assess a semi-	A network of 30 rain gauge	A two-state, first order Markov model	The model reproduces daily and longer time-scale key

(2007)	parametric daily rainfall stochastic generator that replicates local spatial-temporal dependence and also longer term variability and features such as droughts.	stations around Sydney Australia	for rainfall occurrence and a non-parametric kernel density approach for generating rainfall amounts. A moving window is used to allow smooth transition from one month to another.	spatial and temporal characteristics of rainfall adequately. The model is able to replicate temporal attributes such as the distribution of wet and dry spells and the number of rain days.
Mehrotra and Sharma (2009)	To compare three multi-site daily rainfall generators to replicate historic spatial-temporal characteristics.	A network of rain gauge stations around Sydney, Australia	Modified Markov Model (MMM), a re-ordering method for reconstruction of space-time variability and the k-nearest neighbour (K-NN).	All three methods replicate spatial-temporal dependence statistics reasonably well and the MMM produces the best overall results.
Mehrotra et al. (2006)	A comparison of three multi-site precipitation generators to model temporal-spatial dependence in the simulation of daily point rainfall occurrences.	30 rain gauge stations around Sydney Australia	Hidden Markov Model(HMM), the Wilks (1998) model and K-nearest neighbour (K-NN)	All models are able to replicate spatial and temporal dependence. Wilks model offers a better way of modelling the serial dependence at each location.
Ndiritu J.(2011b)	To find out if fragment based perturbations would solve the problem of over-estimating minimum flows of the Variable Length Bootstrap (VLB) streamflow generator.	68 years of historic data applied on 5 sites in the Vaal, South Africa.	Generation of the yearly and monthly data by the VLB method, weighting the averages of the historic fragments obtained from historic matching years and superimposing the perturbations to counter the effects of smoothing.	The disaggregation preserves other historic statistics but grossly under-estimates the minimum monthly flows. This is considered to arise from inadequate matching of the historic years.
Ndiritu. (2011a)	To develop a simple non-parametric monthly streamflow generator that obtains new data and not just resampling historic data (which is a major limitation of many non-parametric methods) and preserves historic statistics effectively.	68 years of monthly streamflows from 5 sites in the Vaal catchment, South Africa	A variable length block bootstrap with a modified weighted and perturbed method of fragments. A parametric model (STOMSA) that is widely applied in practice in South Africa.	The model (VLB) replicated most annual and monthly statistics and obtained new data to a level matching the parametric STOMSA model. The mode l preserved the correlation between the last month of the year and the first month of the following year unlike STOMSA. The non-parametric mode however over-estimates the minimum monthly flows more than STOMSA.
Pegram and McKenzie (1991)	To analyse, fit, verify and validate a synthetic streamflow generator that produces correctly cross-correlated flow sequences at a large number of sites	Multi-site streamflow sequences from the Vaal catchment, South Africa.	A parametric ARMA model for establishing the marginal distribution of the annual totals, estimation of a time-series structure of the normalized variates. A key station approach for	The model preserves the temporal and spatial historic statistics satisfactorily.

	simultaneously in the Vaal River System study.		disaggregating annual flows to a monthly time step.	
Prairie et al. (2006)	The use of modified k-nearest neighbour (k-NN) for monthly streamflow generation and a comparison with a parametric model.	90 year long monthly streamflows from Colorado river at Lees Ferry, Arizona, USA.	The K-NN and an auto regressive (AR1) parametric model.	The k-NN generator captured all the features present in the historic data while the AR1 model could not capture multimodality of some historic probability density functions and also failed to adequately replicate skewness. The k-NN failed to capture the temporal correlations between the last month of the year and the first month of the following year. Although this variant of the k-NN extrapolates beyond the historic values, this ability is limited.
Rajagopalan and Lall (1999)	To construct a multivariate non-parametric time series generator by the use of a K-nearest neighbour (K-NN) simulator for daily precipitation and other weather variables.	30 years of weather data (rainfall, solar radiation, minimum and maximum temperature, average wind speed and average dew point temperature) at Salt Lake City, Utah, USA	A multi-variate Markovian k-NN with lag -1 dependence. A vector of the weather variables is resampled from the historical data (k- nearest neighbours) by the reconditioning of a vector of the same variables on the preceding day.	The k-NN bootstrap preserves the cross-dependence and frequency better than a lag-1 auto regressive model. The ability to simultaneously generate all the weather variables is a notable advantage of the method. The K-NN however does not generate extremes beyond historical record.
Serinaldi and Kilsby (2012)	The use of a modular class of multi-site monthly rainfall generators in water resource management applications and impact studies that adapts to the complexity of generation problem.	Six 150 years long catchment areal average rainfall series from 15 basins of UK and Wales.	Classical parametric (seasonal standardization) model, non-parametric time series analysis (bootstrap resampling) and a more complex and adaptable parametric generator for more complex parts of the historic time series.	The modular approach was successful and the model adapted successfully to the complexity of the time series with the season and the need to incorporate climate change/variability impacts.
Sharma and Mehrotra (2006)	To improve the K-nearest neighbour (K-NN) resampling method by incorporating weights that quantify the influence of the predictor variables.	Synthetic linear data and a rainfall-downscaling problem over 15 stations near Sydney, Australia.	A weighted K-NN and three versions of the unweighted K-NN method.	The influence of weight assists in dealing with non-linearity in hydrologic and meteorological processes since they exhibit non-linear behaviour. The weighted K-NN is found to perform better than the unweighted k-NN approaches.
Sharma and O'Neill(2001)	To formulate and test a non-parametric approach that replicates short-term and longer	84 years of monthly streamflows from Beaver River near Beaver, Utah, USA	A kernel density approach that applies variable kernel to enable the replication of the occurrence of zero	The model is capable of modelling both short and longer term dependencies as well as nonstandard probability density functional forms. Including the sum

	(seasonal and inter-annual) dependence and also complex aspects of monthly distributions such as large asymmetry and multimodality.	and the Burrendong dam inflows, New South Wales, Australia.	flows.	of flows in the past 12 months (aggregate flow) is found to improve the generation substantially.
Sharma et al. (2003)	To develop a non-parametric model for the stochastic generation of daily rainfall amounts that accommodates seasonality and reproduces important distributional and dependence properties of observed rainfall.	14 locations in Australia, 6 locations in South Africa, 24 locations in North America and 22 in Pacific Islands.	Use of Kennel density estimation with a moving window approach to represent the seasonal variations (daily, annual and inter-annual timescales).	The method reproduces seasonal variations in the rainfall record. Because the shape of the kennel PDF is determined by the observed data, the effective use of this method requires data that is adequately long.
Srikanthan and McMahon (2001)	To review stochastic annual, monthly and daily weather generators.	Data from the numerous studies reviewed.	Numerous methods are included in the review.	Inter -annual variability and long term persistence are some of the major aspects not (adequately) modelled by most generators. The review proposes the simple lag-1 Markov model for annual flows, the disaggregation method of Mejia and Rousselle (1976) for monthly flows and the transition probability matrix method for daily and also monthly and annual generation. The review also indicates that daily generation is much more developed than monthly and annual generation.
Srikanthan and Pegram (2009)	To evaluate the performance of a nested multi-site daily rainfall stochastic generation model.	30 rainfall stations around Sydney, new South Wales, Australia.	A two part model based on Wilks' (1998) to generate rainfall occurrence and then rainfall amounts. Nesting daily flow generation within the monthly and annual models in order to replicate monthly and annual characteristics.	The model preserved the important characteristics of daily, monthly and annual time scales but not the skewness of the monthly rainfall.
Srikanthan et al. (2002)	To compare the method of fragments with the kernel density-based method in the stochastic generation of monthly rainfall.	10 rainfall stations that are located in several locations over Australia	Kernel density based method (Sharma and O'Neil (2001)) and the method of fragments (Meheepala and Perera, 1996).	Both methods adequately preserved the monthly and the annually characteristics. The kernel density based method was marginally better in disaggregation than the method of fragments.

Srikanthan et al. (2005)	To compare a parametric with a non-parametric stochastic daily rainfall generator.	125 years of rainfall data at Sydney and Melbourne, Australia	The Transition probability matrix (TPM) and a kernel density-based non-parametric model that generates rainfall occurrence and then rainfall amounts (ROG+RAG model).	Both models perform well in replicating most daily, monthly and annual statistics. The ROG+RAG however performs slightly better than the TPM although it is more complex.
Srinivas and Srinivasan (2000)	The use of the strengths of parametric and non-parametric methods in the form of a hybrid monthly stochastic streamflow generator.	Single-site annual streamflow data from different parts of the world.	Hybrid model, parametric model and a non-parametric model.	The hybrid performs better in preservation of summary statistics, dependence structure, marginal distribution and drought characteristics than the stand-alone parametric and non-parametric approach
Srinivas and Srinivasan (2005)	Extension of the method by Srinivas and Srinivasan (2000) to multiple sites by applying the contemporaneous approach to maintain cross correlation	A 3-streamflow sequences in the Cauvery river basin, India.	Two multi-site hybrid models (Srinvasan and Srinivasan (2000) and Tasker and Dunne (1997)) that have varied features.	The hybrid model of Srinvasan and Srinivasan (2000) preserves historic statistics better than the model of Tasker and Dunne (1997).
Ünal et al. (2004)	To compare several methods of stochastic generation of monthly and annual rainfall.	64 years of monthly rainfall from Goztepe station, Turkey.	Auto regressive (AR) method, method of fragments, modified method of fragments, Thomas-Fiering model, a wavelet based approach and a hybrid of Thomas-Fiering and the wavelet method. .	All models generally preserve long term characteristics of time series with each performing better in particular aspects. The wavelet model preserves best the mean value but has the limitation of not preserving skewness and requiring the sequences to have a normal distribution.
Vogel and Shallcross (1996)	To assess the performance of a moving block bootstrap for stochastic streamflow generation and compare this with a parametric ARMA generator.	A Large number of synthetic annual streamflows for Monte Carlo experimentation.	The moving block bootstrap and an auto regressive (AR1) lognormal parametric generator.	The moving block bootstrap obtained lower errors in reservoir storage tests than the AR1 model. The simple bootstrap was considered a credible alternative to the more complex parametric models.
Wang and Ding (2007)	To formulate and assess the performance of a kernel density based method for synthetic generation of daily streamflow.	Two river gauge stations-Pingshan and Yi-Ping, in Jinsha River, China.	A multivariate kennel density model.	The model replicated daily streamflow sequences adequately well to allow its use in practice. The determination of the model order, optimization of bandwidth coefficient and the choice of kennel functions are outstanding issues.
Wang and Nathan (2007)	To develop and test a rainfall generator that simultaneously	Rainfall at Lake Eppalock catchment and 6 other sites	A parametric model that couples stochastic generation at daily and	The Coupled model satisfactorily replicated the mean, Coefficient of Variation (CV), and skewness of daily and

	preserves historic statistics at the daily and monthly time scale.	of varied characteristics across Australia	monthly time steps.	monthly rainfall total and also the CV and skewness of annual rainfalls.
Wang et al. (2011)	To formulate and evaluate the performance of a wavelet transform based daily rainfall generator that is more robust than that based on Haar wavelet (Bayazit et al. (2001)).	53 years of daily streamflow data at Pingshan on Jinsha river, China	Decomposition of historic streamflow and controlled reconstruction/generation of daily streamflows using an adaptive wavelet transform that acts as a low pass filter.	The method replicates the complete range of daily streamflow statistics satisfactorily. This wavelet approach is able to preserve skewness and to handle non-normally distributed historic data unlike the generation based on the Haar wavelet.
Wilks (1998)	To extend a parametric daily rainfall occurrence and amounts model for simultaneous multisite generation of daily rainfall.	A network of 25 rain gauge stations in New York state, USA	Markov chain for rainfall occurrence generation and exponential distribution for rainfall amounts generation. Spatially correlated but temporally independent random numbers are used to preserve cross-correlations.	The method replicates historic statistics at the daily and monthly time step reasonably. The exponential distribution is found to replicate inter-annual variability better than the Gamma distribution.
Yates et al. (2003)	To formulate and assess the performance of a k-nearest neighbour method that uses Mahalanobis instead of Euclidian distance for the stochastic generation of monthly rainfall.	Rainfall from multiple sites from two regions in the USA (the Rocky Mountains and the north central USA)	K-nearest neighbour that uses Mahalannobis distances and correlations of variables to weight the distances in selection of neighbours.	The method replicated historic statistics sufficiently. It is however noted that the modified k-NN does not obtain values other than those in the historic data and methods to enable this are proposed as potential improvements.

The analysis of the studies on stochastic hydrologic generation in Table 2.1 reveals that research on non-parametric methods has picked up substantially in recent years after decades of domination by parametric approaches. Notable hybrids of the two approaches have also been formulated (e.g. Srinivas and Srinivasan, 2000, 2005). The analysis on Table 2.1 also reveals the numerous ways that stochastic generation can be modified to try and deal with specific characteristics of the data and to meet specific objectives. Most of the studies analysed indicate that the research findings can be used to help improve analysis/decision making but hardly any inform if this is actually being done. A separate study could specifically indicate the level of uptake of hydrologic research in practice but the indications are that this is quite low (Hughes, 2004).

2.3 Comparison of non-parametric stochastic rainfall and streamflow generators

The analysis in Table 2.1 identified the following five non-parametric approaches for the stochastic generation of rainfall, streamflow and other hydrometric data.

- Wavelet approaches (Bayazit et al., 2001); Wang et al., 2011; Ünal et al., 2004).
- Reordering (Schaake Shuffle) methods (Clark et al., 2004a, b; Mehrotra and Sharma, 2009).
- Nearest neighbour approaches (k-NN) (Rajagopalan and Lall, 1999; Lall and Sharma, 1996, Mehrotra et al., 2006, Sharma and Mehrotra, 2006, Yates et al., 2003; Prairie et al., 2006).
- Kernel density based methods (Sharma et al., 2003; Mehrotra and Sharma, 2007, 2009; Srikanthan et al., 2002; Srikanthan et al., 2005; Wang and Ding, 2007).
- Bootstrap methods (Ndiritu, 2011a, 2011b; Serinaldi and Kilsby, 2012; Vogel and Shallcross, 1996).

The five approaches are compared in Table 2.2 based on criteria that are considered important for stochastic hydrologic generation and the performance of the methods reported in Table 2.1. The one comparative analysis of non-parametric methods in the studies included in Table 2.1 (Mehrotra and Sharma, 2009) found the reordering method to replicate historic statistics better than the k-NN. The analysis in Table 2.2 places the generalized wavelet (Wang et al., 2011) and the bootstrap with the ability to extrapolate beyond the historic data (Ndiritu, 2011a) as the methods that may be best to apply for this study. In addition, the simplicity and the observed performance of the reordering method (Clark et al., 2004; Mehrotra and Sharma, 2009) are also appealing in spite of the method's inability to extrapolate beyond the historic data. The complexity of the kernel density-based methods prevents their consideration in this study while the computation intensity of the nearest neighbour approach diminishes its preference.

Table 2.2 A comparison of five approaches for non-parametric stochastic hydrologic generation

Criterion	Stochastic generation approach.				
	Wavelet	Reordering	Nearest neighbour	Kernel density	Bootstrap
Ability to preserve historic characteristics	If the simple Haar wavelet is used, then skewness is not preserved. If a more generalized wavelet is used, then within-year historic statistics are preserved. Long-term variability and persistence may not be preserved sufficiently	Preserves within-year statistics satisfactorily but does not include replication of long-term variability and persistence in its currently used forms.	Preserves within-year statistics adequately and have also been modified to replicate inter-annual dependence. Current forms of this method are not designed to replicate inter-decadal variability and persistence.	Preserve within-year statistics well and formulations for replicating inter-annual dependence have been formulated. Longer-term dependence is however not modeled in these approaches.	Preserve within-year statistics if the simple method of fragments is used in disaggregation. Use of a pair of weighted and perturbed fragments preserves most within-year statistics but over-estimates the minimum flow. The methods can preserve inter-annual and longer term statistics easily by the selection of long building blocks.
Ability to extrapolate beyond the range of historic data (to generate new data).	Has full ability to extrapolate beyond the historic values	Does not have this ability	Most of the formulations do not have this ability. A formulation with limited extrapolation ability has been developed.	Has full ability to extrapolate beyond the historic values	Most bootstrap methods do not have this ability but a bootstrap that has the full ability to extrapolate has been developed.
Limitations of applicability	If the Haar wavelet is used, the historic data is required to possess a normal distribution. More generalized wavelets do not require this.	May not generate effectively if there are many historic values that take on similar values (e.g. daily rainfall with many zeroes).	No known limitation.	No known limitation.	No known limitation.
Possibility of generating negative values	The structure of the approach enables this possibility although this is not mentioned in the studies cited.	Not possible	Not possible	It is possible and an effective approach for dealing with this problem has been devised.	Not possible
Ease of use	The fundamentals of the approach are easy to grasp. The Haar wavelet is easy to understand but understanding more generalized wavelets may be more involving.	Method is easy to understand and set up. The length of the moving window for reordering is subjectively selected.	Generally easy to understand although the approach could be computation intensive. The number of neighbours and method of computing distance between data points are subjective	The method is complex and computationally intensive. There is subjectivity in the selection of the bandwidth and the type of kernel to use.	Bootstrap methods are generally easy to understand and apply. The selection of the minimum block length is subjective. Where weighting and perturbation is done, the selection of the form of weighting and level of perturbation is also subjective.

2.4 Climate Change and Variability

Climate change/variability related research is currently an area of great interest in southern Africa (Green ,2008; Schulze, 2011, 2012; Schulze et al., 2011; Steynor et al., 2009; Lumsden et al., 2011; Hughes et al., 2011; Hughes, 2012) as in most parts of the world. Most climate change research has however not been applied in practice as the uncertainties related to future climate change are still enormous (Ghil et al., 2002; Hughes, 2012) and the research is not often presented in a form easily usable by the practitioner.

Global climate models (General circulation models) (GCMs) are the only physically-based approaches currently available method of attempting to model long-term climate change (Aleix et al., 2013) but different GCMs typically obtain highly variable and uncertain projections (Mujumdar and Ghosh, 2008; Johnson and Sharma, 2009; Buytaert et al., 2009; Aleix et al., 2013) and generally produce large biases in precipitation (Weiland et al., 2010). To try and address the consequent uncertainties, climate studies usually apply results from many GCMs (Dyer et al., 2014; Schulze et al., 2011; Vase et al., 2011; Weiland et al., 2010). This practice is however still greatly inadequate since the ability of GCMs to model climate is mostly not verified/validated (Koutsoyiannis et al., 2009) and the few validation tests reveal that GCMs miss out important hydro-climatic characteristics (Kundzewicz and Stakhiv, 2010). GCMs are found to be inadequate in replicating inter-annual (Weiland et al 2010) and inter-decadal variability (Koutsoyiannis et al., 2007 2008; Anagnostopoulos et al., 2010) which has been found to exist in hydro-climatic and many other natural time series (Hurst, 1951). Consequently, GCM simulations result in considerably smaller variability than those of observed time series (Aleix et al., 2013). Koutsoyiannis (2011) found the uncertainty bounds of streamflow projections based on rainfall from multiple GCMs to be much narrower that those from stochastic methods that adequately incorporate Hurst's phenomenon. Vase et al. (2011) found 15 GCMs unable to simulate both annual rainfall magnitudes and their trends for South East Australia. Kundzewicz and Stakhiv (2010) highlight that GCMs were designed for assessment of the impact of greenhouse gases (GHG) on global climate for developing policies for slowing down the GHG emissions and not for water resources planning adaptation measures. Based on a review of tests of the ability of GCMs to simulate observed rainfall and temperatures, they suggest that GCM and downscaled data based on GCM results is not yet ready for practical water resources planning. On a more positive note, GCMs as the only physically-based tools available for estimating the long-term changes to the central measure (averages) (Aleix et al., 2013) could be used for this task while applying other approaches to realistically simulate long-term persistence. It is possible that GCM projections can also be used for indicating the possible shifts in seasonal rainfall although Dessu and Malesse (2013) reveal an unsatisfactory GCM replication of monthly rainfall patterns on a study in the Mara River basin in East Africa. The Climate Systems Analysis Group (CSAG) at the University of Cape Town has downscaled rainfalls from multiple GCMs to over 2600 rainfall stations in South Africa (CSAG, 2008) and the changes in monthly rainfall for each station are available (http://cip.csag.uct.ac.za/webclient2/app/#datasets).

Sharif and Burn (2006) used a non-parametric nearest neighbour approach to model monthly rainfalls and temperatures to match specific climate change scenarios. This approach did not specifically seek to model inter-annual variability but graphical plots of the simulated series' indicated reasonable existence of this variability. Chen et al. (2010) used Fast Fourier Transforms to incorporate long-term climate variability in their daily stochastic weather generator. Koutsoyiannis (2011) demonstrates the need to incorporate long-term persistence for predicting hydro-climatic time series and illustrates how incorporating apparent non-stationarities whose physical basis is not known in data could grossly underestimate uncertainties. Koutsoyiannis (2011) further shows that it is likely to be much safer to use a stationary model that adequately accounts for persistence. Using empirical mode decomposition, Lee and Ouarda (2011) show how apparent trends that could be considered as non-stationarity might simply be parts of longer term oscillations in the observed records.

2.5 Summary of Literature Review

The literature review has analysed studies on stochastic rainfall and streamflow generation (Section 2) and identified five non-parametric methods that could potentially be applied to meet the objectives of this research (Section 3). The review in Section 3 favours the generalized wavelet approach and the variable length bootstrap (VLB). Out of the two, the VLB has the advantage of being simpler to understand in addition to having an easy method of incorporating long-term variability in comparison with the generalized wavelet. The application of the generalized wavelet to multiple rainfall sites is also likely to be much more complex than the simple contemporaneous approach (Ndiritu, 2011) that the VLB applies very effectively. The VLB will therefore be used as the non-parametric generator for this study. If it is however found that the VLB has serious limitations and/or deficiencies, the generalized wavelet will be included as an additional non-parametric rainfall generator. The review of climate-related variability reveals that the climate projection technology based on GCM is still largely unsatisfactory and applying GCM projections alone could lead to large under-estimation of uncertainties of future rainfall. The literature review also informs that it is safer to use a stationary model that takes in the variability in the existing data fully than to use a model that considers some of the observed variability as non-stationary and therefore incorporates it in generating plausible projections (Koutsoyiannis, 2011). A variation of the VLB to enable the generation of stationary stochastic rainfalls for a drier, more variable or a wetter climate will therefore be sought. This will be done whilst maintaining the ability of the VLB to incorporate observed long-term persistence in the data. Koutsoyiannis (2011) used Hurst's coefficient to quantify this persistence while approaches minimum run sums (Peel et al., 2004, 2005) have also be applied. Since GCMs are still the only physically-based climate models in existence, the shift in overall average rainfall from multiple GCM projected could be used as a guide for how dry or wet the climate could be. In addition an allowance will be made for use of projections of monthly shifts in monthly rainfall patterns from multiple GCMs in disaggregating annual rainfalls.

3 DEVELOPMENT OF VARIABLE LENGTH BLOCK (VLB) GENERATOR

3.1 Introduction

The VLB rainfall generator is a variation of the VLB streamflow generator that had been found to perform at par with the widely used STOMSA streamflow generator (Ndiritu, 2011 a). The VLB streamflow generator was found to over-estimate the lowest flows and an attempt to improve the replication of the lowest flows was found to grossly underestimate them (Ndiritu, 2011 b). In adapting the VLB for rainfall generation, the concepts and logic applied in the development of the VLB streamflow generator are re-evaluated in consideration of the temporal and spatial dependence characteristics of rainfall rather than those of streamflow. This is done with the objective to maximize both simplicity and effectiveness.

An appropriate data set is required to acquire the required knowledge of the temporal and spatial dependence structure of rainfall and to also develop and test the generator adequately. For this, the extensive rainfall database by Lynch (2003) and Kunz (2009) was used and a group of 10 stations with a concurrent 94 years long rainfall record was selected for the study. The 10 rainfall stations are widely spread out in South Africa and required very little patching (averaging 3.5%). Figure 3.1 shows the locations while Table 3.1 shows the percentage of patching and the basic annual statistics of the rainfall measured at the stations. For 9 of the 10 stations that are located in the summer rainfall region, the hydrologic year is assumed to start in July while for the single station (0020866 W) located in the winter rainfall region, the hydrologic year is assumed to start in January. This ensures that the year begins and ends in the driest months as seen in Figure 3.2. Table 3.2 presents the annual serial and cross correlations coefficients of the stations while the monthly cross correlation coefficients are presented graphically in Figure 3.3. Figure 3.4 presents the monthly serial correlation coefficients and the average monthly cross and serial correlations are shown on Figure 3.5.

Out of the 10 stations, 0020866W is located in the winter rainfall region of the Western Cape while the other 9 are in the summer rainfall region of South Africa. Station 0020866W therefore has very low annual cross correlation coefficients with the other stations that average to 0.05. The average annual cross correlation coefficient among the other stations is high at 0.4. Table 3.2 also reveals very low annual serial correlations for the 10 stations. The monthly cross correlation plots on Figure 3.3 reveal a large variation of the correlations between different stations although these variations are fairly similar for all months. No distinct pattern is observed with the monthly serial correlation plots on Figure 3.4. Like for the annual time-step, the cross correlations are substantially higher than the serial correlations at the monthly time step as seen on Figure 3.5. The monthly cross and serial correlation coefficients average to 0.22 and 0.07 respectively.

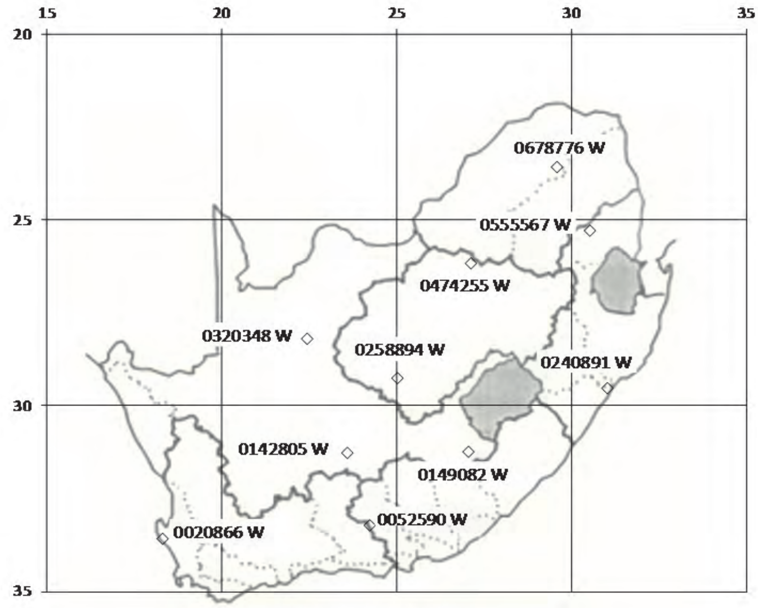


Figure 3.1 Location of selected rainfall stations

Table 3.1 Basic statistics of rainfall stations

Station	0020866W	0555567W	0474255W	0320348W	0258894W	0678776W	0052590W	0142805W	0149082W	0240891W
Mean	605	830	579	325	394	843	238	320	588	995
Stdev	115	221	151	138	140	285	90	103	141	218
CV	0.19	0.27	0.26	0.42	0.35	0.34	0.38	0.32	0.24	0.22
Skewness	0.31	0.93	0.36	1.53	0.84	0.86	0.85	0.43	-0.01	0.6
Minimum	349	556	209	104	159	405	69	113	247	549
Maximum	857	1501	1061	959	793	1577	607	627	990	1741
% patching	0.5	2	7.1	0.6	4.4	6.3	8	0.3	2.4	3.4

CV: Coefficient of variation

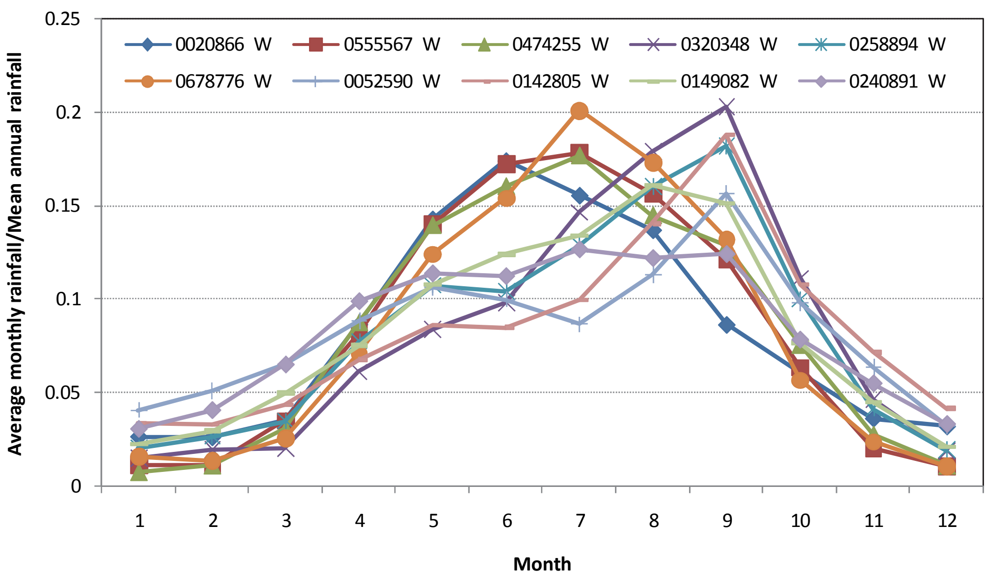


Figure 3.2 Monthly rainfall distributions of selected rainfall stations

Table 3.2 Cross correlation and serial correlation coefficients of annual rainfalls

	ACC										ASC
Station	0020866W	0555567W	0474255W	0320348W	0258894W	0678776W	0052590W	0142805W	0149082W	0240891W	
0020866W	1										0.09
0555567W	0	1									-0.06
0474255W	0	0.44	1								0.05
0320348W	-0.12	0.46	0.3	1							-0.08
0258894W	0	0.39	0.37	0.73	1						-0.15
0678776W	-0.05	0.76	0.32	0.46	0.42	1					-0.07
0052590W	-0.07	0.2	0.13	0.39	0.37	0.31	1				0.1
0142805W	-0.03	0.37	0.32	0.66	0.66	0.41	0.55	1			-0.07
0149082W	0	0.42	0.31	0.52	0.64	0.43	0.42	0.6	1		-0.16
0240891W	0.16	0.27	0.22	0.23	0.35	0.23	0.19	0.34	0.31	1	0.01

ACC – Annual cross correlation coefficient; ASC – Annual serial correlation coefficient

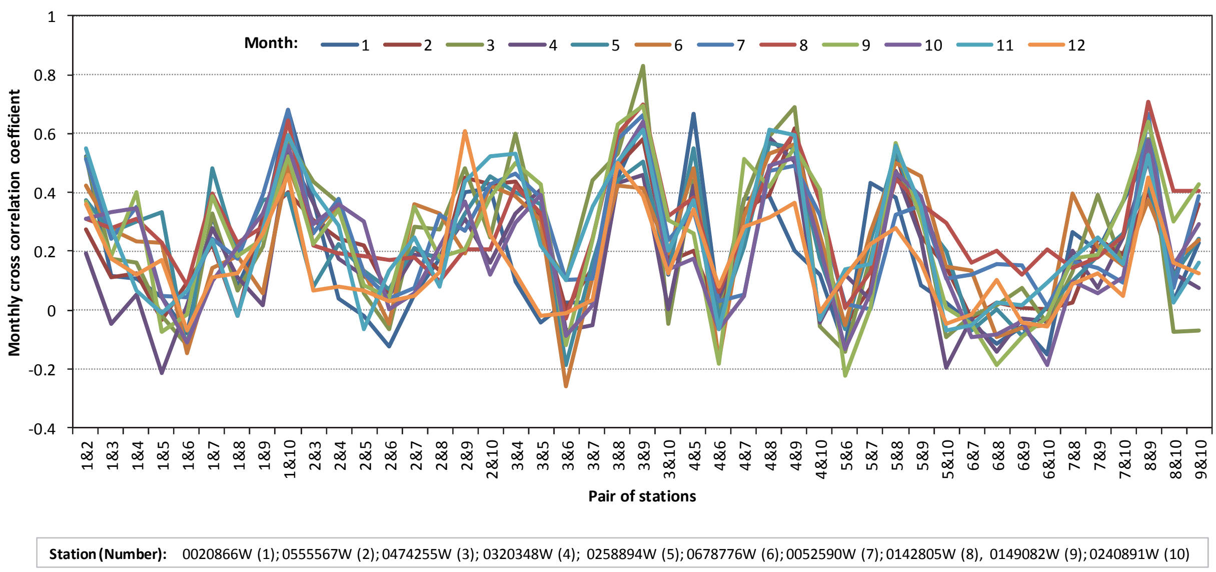


Figure 3.3 Monthly cross correlation coefficients for rainfall stations

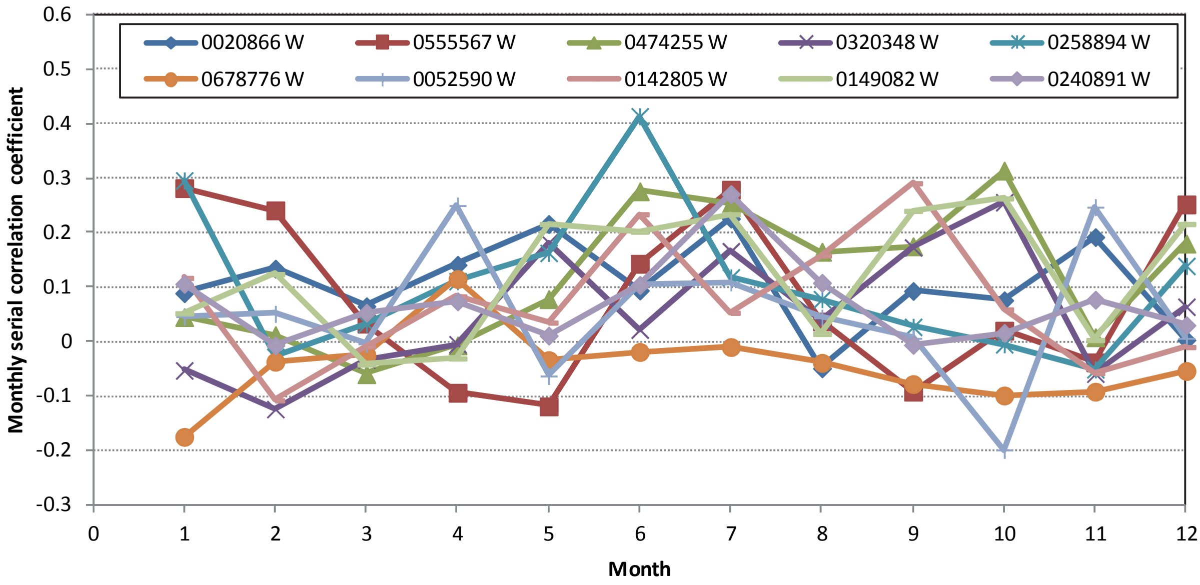


Figure 3.4 Monthly serial correlation coefficients for rainfall stations

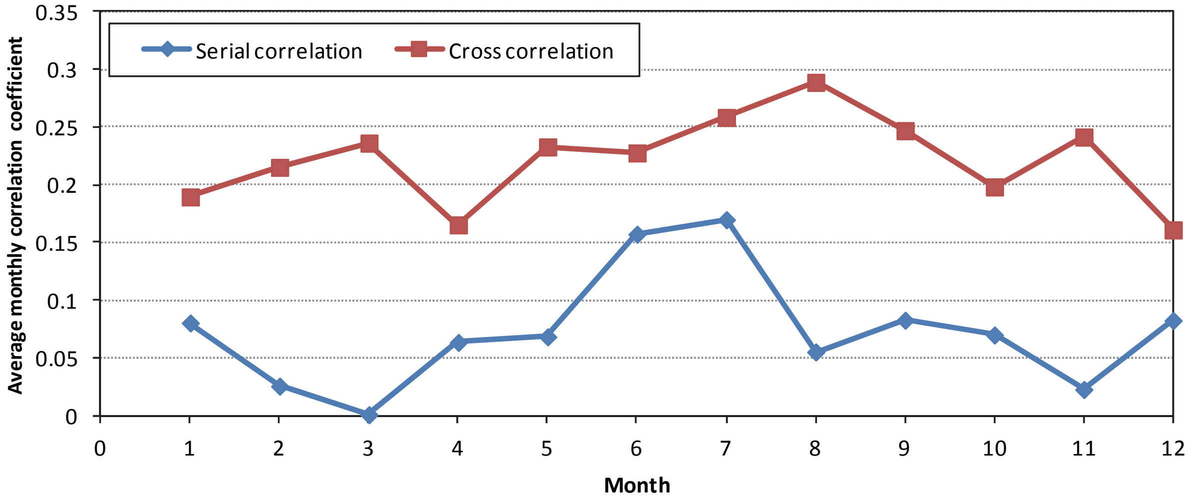


Figure 3.5 Average monthly cross and serial correlation coefficients of rainfall stations

3.2 Evaluating the VLB Streamflow Generator and adapting it for Rainfall Generation

The VLB streamflow generator is described in detail by Ndiritu (2011a, b) and a detailed description will therefore not be provided here. The main steps of this VLB generator are:

i. Generation of variable length blocks of annual time series from the historic time series

ii. Random sampling of the blocks with replacement to create an annual stochastic time series of the specified length.

iii. 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.
iv. The disaggregation of the stochastic annual values using the monthly distributions of the pair of matching historic years and an appropriate incorporation of perturbations.
v. The updating of the stochastic annual values after the disaggregation.

An assessment of the 5 steps revealed that step i) was formulated in a manner that encourages the generation of variability of the wetter rather than drier climatic periods. In addition, since the annual serial correlation of rainfalls was found to be negligible (Table 3.2), the value of matching stochastic rainfalls to historic ones in step iii) needed to be reconsidered. The details of the assessment and the consequent modifications made now follow.

3.2.1 Generation of variable length blocks

The traditional bootstrapping is often applied using blocks of constant length (e.g. Vogel and Shallcross, 1996) but the use of a constant length does not reflect the observed multi-annual variability of climate especially the occurrence of extended dry periods. Using four of the ten rainfall series selected for this study, Figure 3.6 illustrates the occurrence of the dry periods. The VLB streamflow generator was designed to allow for this variability by creating blocks that start and terminate during wet periods. This helped to locate the historic dry periods within the blocks and it was reasoned that the generated series would then replicate the historic low flow characteristics adequately. This argument however did not recognize the need to specifically create variability in the possible low flow patterns as it ends up generating low flow patterns that are fairly similar to the historic ones irrespective of how the blocks are resampled. Conceptually, the approach followed ends up creating more variability for the higher flow rather than the lower flow periods while the lower flow periods are the ones of more interest for water resources assessment.

An approach to stochastic rainfall generation that favours the creation of more variability for the lower rainfall (drought) periods is therefore adopted here. In this approach, the blocks are terminated if the annual rainfall is lower that a low rainfall threshold specified as a percentile and not as a ratio of the average rainfall (as done in the original VLB generator). With this method, the blocks link up during low flow periods and therefore generate other plausible low period rainfall patterns in addition to those similar to the historic ones. The use of a percentile rather than a ratio of the mean rainfall ensures that the threshold is realistic as there will be some lower annual rainfalls in the time series while using a ratio of the mean rainfall could lead to a threshold that is lower than the lowest annual rainfall in the time series. Using a percentile is also more consistent for multi-site generation because each site is alternately selected as the lead sequence whereas the annual rainfall variability amongst the sites may be highly variable as demonstrated on Figure 3.7 and seen on Table 3.1.

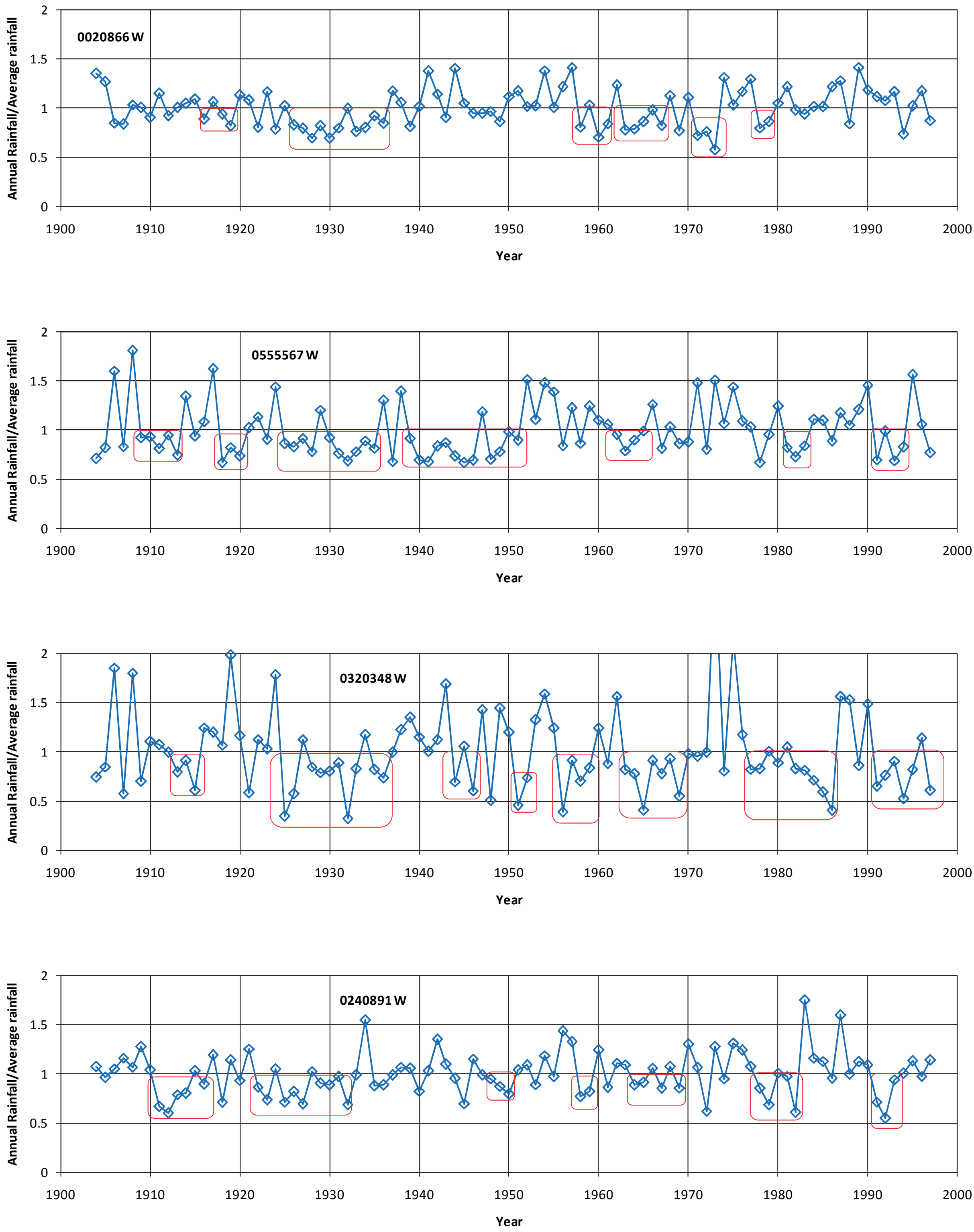


Figure 3.6 Annual rainfall time series highlighting low rainfall periods

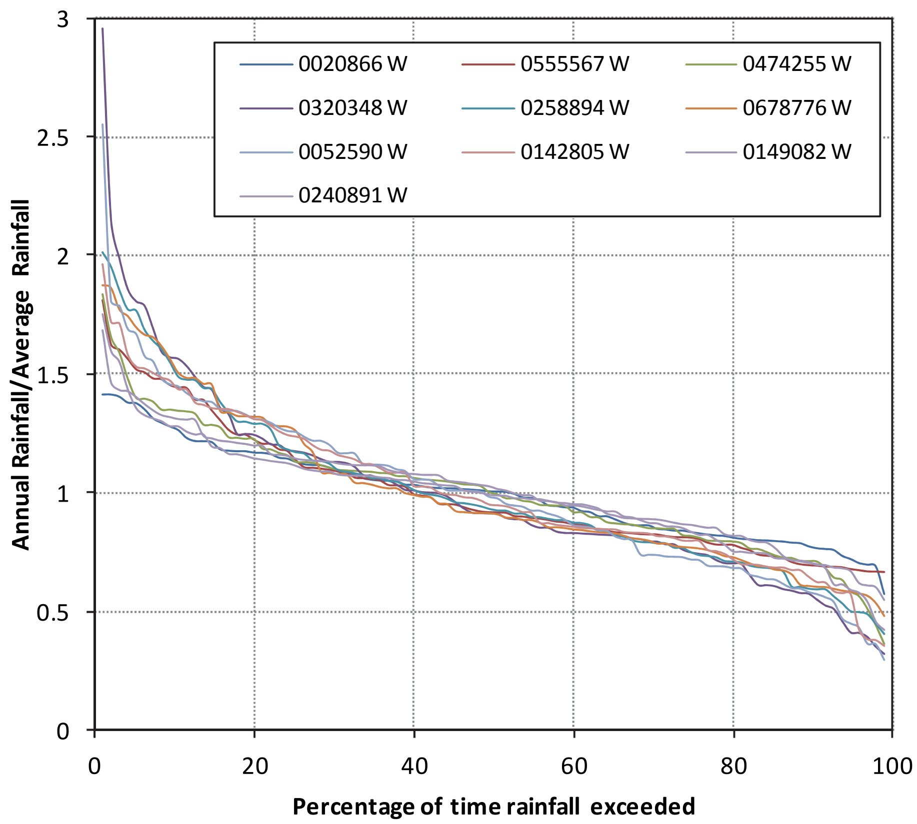


Figure 3.7 Percentile plots of annual rainfall of the selected stations

3.2.2 Matching historic years to stochastic years.

The matching of historic years to stochastic years in the VLB streamflow generator was incorporated to model the observed dependence of monthly flow distribution on annual flow magnitudes (Ndiritu, 2011a). An analysis of the dependence of monthly rainfall distribution on annual rainfall was carried out by trying to establish relationships between the coefficient of variation and the skewness of the 12 monthly values with the corresponding annual rainfall. Table 3.3 shows the linear relationships obtained for the 10 rainfall time series. Although a correlation is found to exist between skewness and annual rainfall magnitude given the large slopes on the 5th column of Table 3.3, the correlation coefficients (R^2) on column 7 are very low. The low values on columns 2 and 4 of Table 3.3 reveal the absence of correlation between coefficient of variation of monthly rainfalls and annual rainfall magnitude. Considering the need to keep the generation simple, it was decided not to carry out any matching of historic and stochastic years and to simply assume that the historic monthly rainfall distribution in any year could be applied in the disaggregation of any stochastic annual rainfall.

A detailed description of the resulting VLB rainfall generator now follows.

Table 3.3 The variation of the coefficient of variation and skewness of monthly rainfalls with annual rainfall.

Station	y=mx + c y =CV, x = Annual Rainfall			y=mx + c y =Skewness, x = Annual Rainfall		
	m	c	R^2	m	c	R^2
0020866 W	0.0023	0.0735	0.0009	-0.1933	1.2433	0.0033
0555567 W	0.0046	0.0804	0.0077	0.4902	0.4086	0.0668
0474255 W	-0.018	0.1075	0.1058	-0.478	1.4029	0.0454
0320348 W	-0.0056	0.1107	0.0134	-0.0805	1.5727	0.0027
0258894 W	-0.0005	0.0931	0.0001	0.0265	1.2117	0.0002
0678776 W	0.0185	0.074	0.0965	0.5988	0.5764	0.073
0052590 W	-0.0021	0.091	0.0015	-0.0074	1.2828	0.00002
0142805 W	-0.0053	0.0958	0.0072	-0.3271	1.6086	0.0231
0149082 W	-0.0089	0.0868	0.0233	-0.1223	1.1081	0.0025
0240891 W	0.0054	0.0625	0.0075	0.7228	0.184	0.06

3.3 The VLB Rainfall Generator

3.3.1 Block generation

Following are the steps applied to generate blocks of variable length from the historic time series.

i. Define a low-rainfall year as that having an annual rainfall lower than that exceeded for a set proportion of time and find the corresponding rainfall value by a simple plotting position approach (e.g. the Weibull method). To encourage increased variability in block generation, a range of the proportion of exceedance is set and the specific value to apply for a given generation is obtained randomly within the range.
ii. Set the minimum length of the block in years.
iii. Starting with the beginning of the series, shift forwards by a length equal to the minimum block length and then move sequentially and locate the first low-rainfall year.
iv. Define the beginning of the series to this year as the first block and obtain the other blocks in a similar manner starting with the following year. Check that the last block also meets the minimum block length requirement.

For the 94 year-long time series for rainfall station 014902 W, defining low rainfall as that exceeded 60% of the time and using a minimum block length of 4 years gives 16 blocks as illustrated on Figure 3.8.

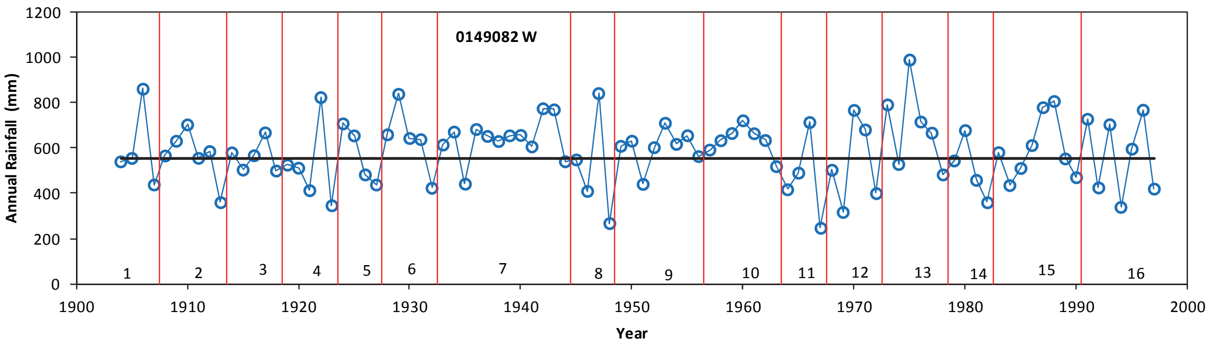


Figure 3.8 The generation of variable length blocks. The black horizontal line defines the low rainfall threshold and the vertical red lines the termination of the blocks. The blocks are numbered as 1 to 16 above the x-axis.

3.3.2 Generation of initial annual stochastic rainfall series

Because the blocks are generated so as to terminate in low rainfall years, there is a tendency for the blocks to also start in a dry rather than a normal rainfall year. Resampling these blocks to create the stochastic rainfall series would therefore lead to sequences that predominantly start in a dry period. In order to correct this bias, a warm-up period subjectively selected as 20 years is applied. If stochastic sequences of length ***N*** are required, random resampling of the blocks with replacement is done until a length equal to or exceeding ***N***+20 is achieved. The stochastic rainfall sequence of length ***N*** is then obtained starting at a point selected randomly within the first 20 years. Figure 3.9 illustrates the generation of a 100 year long sequence by sampling the blocks from Figure 3.8 randomly with replacement. The numbers of the resampled blocks are placed above the x-axis on Figure 3.8. After a total length of 100+20 = 120 years is reached, a warm-up period of 16 years is randomly obtained between 1 and 20 and the 100 year long sequence is then obtained from year 17 to year 116.

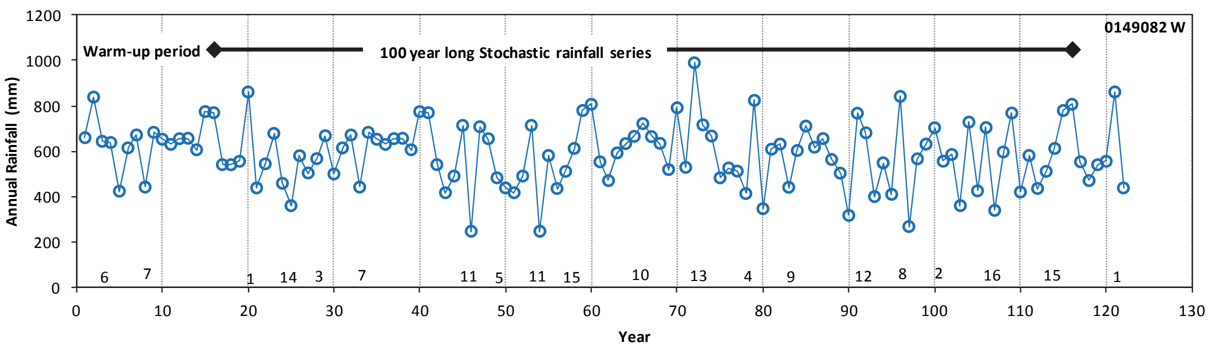


Figure 3.9 The generation of a 100 year long stochastic annual rainfall sequence.

3.3.3 Perturbing stochastic annual rainfalls

The original VLB streamflow generator (Ndiritu, 2011a) applied weighted pairs of fragments from the historic time series to enable the generation of annual and monthly rainfalls other than those found in the historic sequences. The weighting had the effect of smoothing the monthly distribution and perturbations having a triangular distribution were imposed on the resulting fragments to counter the smoothing. There was however no evidence provided to show that this would recover the original monthly distribution structure and the VLB was also found to over-estimate the lowest monthly streamflows. Ndiritu (2011b) derived a fragment-based perturbations approach that was theoretically capable of recovering the original monthly distribution structure but this resulted in under-estimation of the lowest monthly flows. The probable reason for this was considered to be bias in the matching of historic and synthetic years that was part of the generator. Since it has been decided not to carry out any matching of historic and synthetic years (Section3.2.2) and to consider the historic fragments from any year to be applicable for any other, it is possible that bias is unlikely to occur and fragments-based perturbation approach was therefore been considered for perturbing the annual rainfalls and for their disaggregation into monthly rainfalls. Although this approach performed very well for perturbing the annual rainfall and in most respects for disaggregation, it was found to i) over-estimate the 25th percentile, ii) to under-estimate the standard deviations and the highest rainfalls and iii) to under-estimate and over-estimate some of the lowest rainfalls. It was therefore decided to use another approach for disaggregation but still use the fragment-based perturbations method for perturbing the initial annual rainfalls. A description of the fragment-based perturbations and the disaggregation method is now given in current and the following sections respectively.

Suppose we want to perturb the stochastic annual rainfall obtained for year ***i***. Two years ***k*** and ***l*** are selected randomly from the historic record and their monthly fragments defined as the ratio of the monthly rainfall to the annual rainfall are computed. The fragments $\boldsymbol{f^1_{i,j}}$ and $\boldsymbol{f^2_{i,j}}$ are obtained as:

$$f^1_{i,j} = \frac{m_{k,j}}{A_k} \quad and \quad f^2_{i,j} = \frac{m_{l,j}}{A_l} \qquad for \quad j = 1,2,\ldots\ldots,12 \tag{3.1}$$

Where $\boldsymbol{m_{k,j}}$ and $\boldsymbol{m_{l,j}}$ are the monthly rainfalls for month ***j*** of year ***k*** and ***l*** and $\boldsymbol{A_k}$ and $\boldsymbol{A_l}$ are the annual rainfalls for year ***k*** and ***l*** respectively.

Figure 3.10a illustrates typical fragment values and distributions obtained from rainfall station 0240891W for three years,

For each month ***j***, an average weighted stochastic fragment $\boldsymbol{f^{1,s}_{i,j}}$ is obtained using the linearly varying $\boldsymbol{w^1_j}$ and $\boldsymbol{w^2_j}$ as shown on Figure 3.10b. Any weights that add up to unity for each month could be applied and the linear variation is chosen because there is no basis for a more complex form. The stochastic fragments are thus obtained as;

$$f^{s,1}_{i,j} = \left(\frac{12-j}{11}\right) f^1_{i,j} + \left(\frac{j-1}{11}\right) f^2_{i,j} \qquad for \quad j = 1,2,\ldots\ldots,12 \tag{3.2}$$

The averaging of the fragments leads to a smoothing of the monthly distribution and therefore need to be perturbed in a manner that can recover the original distribution. This is done using the fragment-based method first presented by Ndiritu (2011 b).

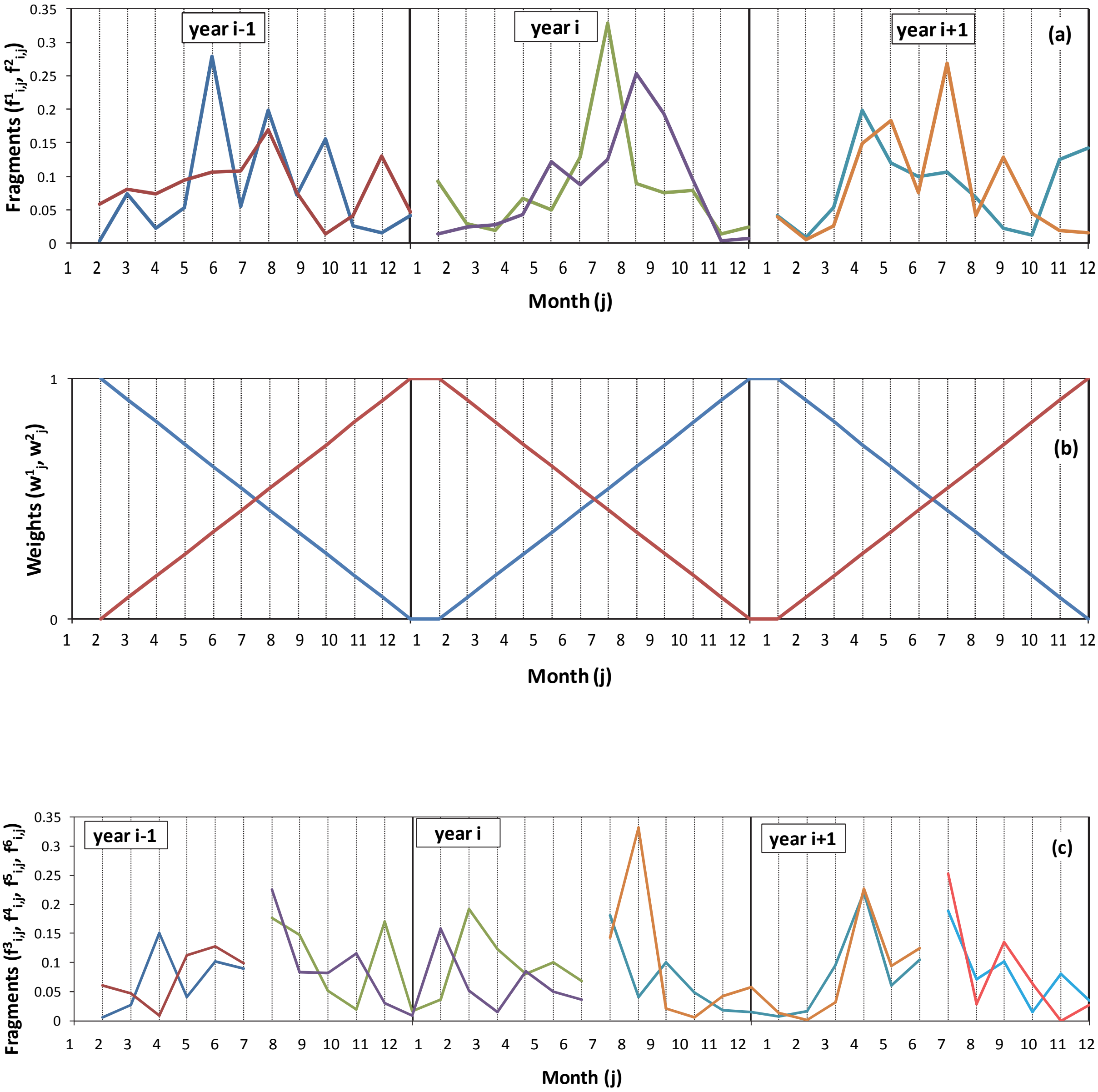


Figure 3.10 Fragments for disaggregating and perturbing stochastic annual rainfalls

The weights applied to obtain the stochastic fragment in equation 3.2 are greater for $f^1_{i,j}$ for the first half of the year and $f^2_{i,j}$ for the second half. The changes that would be needed to obtain $f^1_{i,j}$ from $f^{s,1}_{i,j}$ for the first half and $f^2_{i,j}$ from $f^{s,1}_{i,j}$ for the second half of the year would recover the original variability monthly distribution structure. The respective differences $\Delta p^1_{i,j}$ and $\Delta p^2_{i,j}$ are obtained as;

$$\Delta p^{1}_{i,j} = f^{1}_{i,j} - f^{s,1}_{i,j} = \left(1 - \frac{12-j}{11}\right) f^{1}_{i,j} - \left(\frac{j-1}{11}\right) f^{2}_{i,j} \; = \; \left(\frac{j-1}{11}\right)\left(f^{1}_{i,j} - f^{2}_{i,j}\right) \; for\; j = 1, 2, \dots, 6 \quad (3.3)$$

$$\Delta p^{2}_{i,j} = f^{2}_{i,j} - f^{s,1}_{i,j} = \left(1 - \frac{j-1}{11}\right) f^{2}_{i,j} - \left(\frac{12-j}{11}\right) f^{1}_{i,j} \; = \; \left(\frac{12-j}{11}\right)\left(f^{2}_{i,j} - f^{1}_{i,j}\right) \; for\; j = 7, 8, \dots, 12 \quad (3.4)$$

Since using the same years as years ***k*** and ***l*** used to obtain $f^{s,1}_{i,j}$ would lead to stochastic fragments that are identical to the historic ones, the fragments to apply in obtaining $\Delta p^{1}_{i,j}$ and $\Delta p^{2}_{i,j}$ need to come from other historic years. Since the fragments for obtaining $f^{s,1}_{i,j}$ have been obtained considering the year to start in month 1, it is considered more appropriate to obtain the perturbations with years considered to start from month 7 so that the discontinuities imposed on the time series do not both locate at the same point (month 1). The perturbations are therefore obtained from two pairs of historic years say ***n*** and ***o*** and ***p*** and ***q*** such that;

$$\Delta p^{1}_{i,j} = \left(\frac{j-1}{11}\right)\left(f^{3}_{i,j} - f^{4}_{i,j}\right) \quad for\; j = 1, 2, \dots, 6 \qquad (3.5)$$

$$\Delta p^{2}_{i,j} = \left(\frac{12-j}{11}\right)\left(f^{5}_{i,j} - f^{6}_{i,j}\right) \; for\; j = 7, 8, \dots, 12 \qquad (3.6)$$

where

$$f^{3}_{i,j} = \frac{m_{n,j}}{A_n}, \quad f^{4}_{i,j} = \frac{m_{o,j}}{A_o}, \quad f^{5}_{i,j} = \frac{m_{p,j}}{A_p} \quad and \quad f^{6}_{i,j} = \frac{m_{q,j}}{A_q} \qquad for \quad j = 1,2, \dots\dots, 12 \qquad (3.7)$$

and $m_{n,j}$, $m_{o,j}$, $m_{p,j}$ and $m_{q,j}$ are the monthly rainfalls for month ***j*** of year ***n***, ***o***, ***p*** and ***q*** and A_n, A_o, A_p and A_q are the annual rainfalls for year ***n***, ***o***, ***p*** and ***q*** respectively.

Figure 3.10c demonstrates how fragments selected for generating perturbations span across synthetic years.

Including the perturbations $\Delta p^{1}_{i,j}$ and $\Delta p^{2}_{i,j}$ obtained in equations 3.5 and 3.6 to the stochastic fragment $f^{s,1}_{i,j}$ as computed in equation 3.2 gives the final stochastic perturbations $f^{s,2}_{i,j}$ as;

$$f^{s,2}_{i,j} = \left(\frac{12-j}{11}\right) f^{1}_{i,j} + \left(\frac{j-1}{11}\right)\left(f^{2}_{i,j} + f^{3}_{i,j} - f^{4}_{i,j}\right) \qquad for \quad j = 1, 2, \dots\dots, 6 \qquad (3.8)$$

$$f^{s,2}_{i,j} = \left(\frac{12-j}{11}\right)\left(f^{1}_{i,j} + f^{5}_{i,j} - f^{6}_{i,j}\right) + \left(\frac{j-1}{11}\right)\left(f^{2}_{i,j}\right) \quad for \quad j = 7, 8, \dots\dots, 12 \qquad (3.9)$$

In case the final stochastic fragments $f^{s2}_{i,j}$ turns out to be less than zero, the proportion by which the perturbation $\Delta p^1_{i,j}$ or $\Delta p^2_{i,j}$ needs to be scaled down to achieve a stochastic fragment of zero is determined and all the perturbations of the complete year are scaled by this value. The scaling down is applied for all perturbations across the year to prevent bias towards generating higher rainfalls. The scaling down was found to be required about 2% of the time.

The final stochastic fragments are used to obtain the monthly rainfalls $m^s_{i,j}$ from the initial annual stochastic rainfall A^s_i for year I using equation 3.10.

$$m^s_{i,j} = f^{s,2}_{i,j} A^s_i \quad for\ j = 1, 2, \ldots\ldots, 12 \tag{3.10}$$

The final annual stochastic rainfall A^{sf}_i is then obtained as the summation of the monthly stochastic rainfalls.

$$A^{sf}_i = \sum_{j=1}^{12} m^s_{i,j} = A^s_i \sum_{j=1}^{12} f^{s,2}_{i,j} \tag{3.11}$$

The rest of this section demonstrates how the formulation above generates unbiased perturbations.

If in equation 3.11, $\sum_{j=1}^{12} f^{s,2}_{i,j} = 1$, then $A^{sf}_i = A^s_i$ and there would be no perturbation on the initial annual stochastic rainfall. It therefore needs to be shown that $\sum_{j=1}^{12} f^{s,2}_{i,j}$ is highly unlikely to equal unity (1.0) and that that the perturbations are not biased towards lower or higher values.

From equations 3.8 and 3.9,

$$\begin{aligned}\sum_{j=1}^{12} f^{s,2}_{i,j} = \quad & \sum_{j=1}^{6}\left[\left(\frac{12-j}{11}\right) f^1_{i,j} + \left(\frac{j-1}{11}\right)\left(f^2_{i,j} + f^3_{i,j} - f^4_{i,j}\right)\right] \\ & + \sum_{j=7}^{12}\left[\left(\frac{12-j}{11}\right)\left(f^1_{i,j} + f^5_{i,j} - f^6_{i,j}\right) + \left(\frac{j-1}{11}\right)\left(f^2_{i,j}\right)\right]\end{aligned} \tag{3.12}$$

Defining the differences between the pairs of fragments as;

$$\Delta f^{1-2}_{i,j} = f^1_{i,j} - f^2_{i,j},\ \Delta f^{3-4}_{i,j} = f^3_{i,j} - f^4_{i,j},\ and\ \Delta f^{5-6}_{i,j} = f^5_{i,j} - f^6_{i,j} \tag{3.13}$$

and considering that

$$f^2_{i,j} = f^1_{i,j} - \Delta f^{1-2}_{i,j} \tag{3.14}$$

Equation 3.12 becomes

$$\sum_{j=1}^{12} f_{i,j}^{s,2} = \sum_{j=1}^{6}\left[\left(\frac{12-j}{11}\right)f_{i,j}^{1} + \left(\frac{j-1}{11}\right)\left(f_{i,j}^{1} - \Delta f_{i,j}^{1-2} + \Delta f_{i,j}^{3-4}\right)\right] + \sum_{j=7}^{12}\left[\left(\frac{12-j}{11}\right)\left(f_{i,j}^{1} + \Delta f_{i,j}^{5-6}\right) + \left(\frac{j-1}{11}\right)\left(f_{i,j}^{1} - \Delta f_{i,j}^{1-2}\right)\right] \tag{3.15}$$

Placing common terms appropriately and simplifying equation 3.15 in 4 steps below leads to equation 3.19.

$$\sum_{j=1}^{12} f_{i,j}^{s,2} = \sum_{j=1}^{6}\left[\left(\frac{12-j+j-1}{11}\right)f_{i,j}^{1} + \left(\frac{j-1}{11}\right)\left(-\Delta f_{i,j}^{1-2} + \Delta f_{i,j}^{3-4}\right)\right] + \sum_{j=7}^{12}\left[\left(\frac{12-j+j-1}{11}\right)f_{i,j}^{1} + \left(\frac{12-j}{11}\right)\Delta f_{i,j}^{5-6} + \left(\frac{j-1}{11}\right)\left(-\Delta f_{i,j}^{1-2}\right)\right] \tag{3.16}$$

$$\sum_{j=1}^{12} f_{i,j}^{s,2} = \sum_{j=1}^{6}\left[f_{i,j}^{1} + \left(\frac{j-1}{11}\right)\left(-\Delta f_{i,j}^{1-2} + \Delta f_{i,j}^{3-4}\right)\right] + \sum_{j=7}^{12}\left[f_{i,j}^{1} + \left(\frac{12-j}{11}\right)\Delta f_{i,j}^{5-6} + \left(\frac{j-1}{11}\right)\left(-\Delta f_{i,j}^{1-2}\right)\right] \tag{3.17}$$

$$\sum_{j=1}^{12} f_{i,j}^{s,2} = \sum_{j=1}^{12} f_{i,j}^{1} - \sum_{j=1}^{12}\left(\frac{j-1}{11}\right)\Delta f_{i,j}^{1-2} + \sum_{j=1}^{6}\left(\frac{j-1}{11}\right)\Delta f_{i,j}^{3-4} + \sum\nolimits_{j=7}^{12}\left(\frac{12-j}{11}\right)\Delta f_{i,j}^{5-6} \tag{3.18}$$

Since $\sum_{j=1}^{12} f_{i,j}^{1} = 1$, then

$$\sum_{j=1}^{12} f_{i,j}^{s,2} = 1 - \sum_{j=1}^{12}\left(\frac{j-1}{11}\right)\Delta f_{i,j}^{1-2} + \sum_{j=1}^{6}\left(\frac{j-1}{11}\right)\Delta f_{i,j}^{3-4} + \sum\nolimits_{j=7}^{12}\left(\frac{12-j}{11}\right)\Delta f_{i,j}^{5-6} \tag{3.19}$$

The three summations on the right hand side of equation 3.19 therefore express the perturbation $\boldsymbol{p^{ert}_i}$ that is imposed on the initial annual stochastic rainfall.

$$p_i^{ert} = -\sum_{j=1}^{12}\left(\frac{j-1}{11}\right)\Delta f_{i,j}^{1-2} + \sum_{j=1}^{6}\left(\frac{j-1}{11}\right)\Delta f_{i,j}^{3-4} + \sum\nolimits_{j=7}^{12}\left(\frac{12-j}{11}\right)\Delta f_{i,j}^{5-6} \tag{3.20}$$

It is obvious that the value of $\boldsymbol{p^{ert}_i}$ is highly unlikely to equal zero for all possible combinations of years (***k***, ***l***, ***n***, ***o***, ***p*** and ***q***) as the differences in fragments $\Delta f^{1-2}_{i,j}$, $\Delta f^{3-4}_{i,j}$, and $\Delta f^{5-6}_{i,j}$ are themselves only likely to equal zero when both months for the pair of years have zero rainfall. Applying the perturbation however does not bias the rainfall because the expected value of the summation of $\boldsymbol{p}^{ert}_i$ over the total period of data generation (say $\boldsymbol{N}$ years long) is zero (equation 3.21) since the differences in fragments are equally likely to take positive or negative values. Figure 3.11 presents a probability distribution plot of the perturbations from 100 94-years long stochastic generations of the 10 rainfall stations selected for this study. The overall average of the perturbations was 0.008 indicating the generation was practically unbiased.

$$E\left[\sum_{i=1}^{N} p_i^{ert}\right] = E\left[-\sum_{i=1}^{N}\sum_{j=1}^{12}\left(\frac{j-1}{11}\right)\Delta f_{i,j}^{1-2} + \sum_{i=1}^{N}\sum_{j=1}^{6}\left(\frac{j-1}{11}\right)\Delta f_{i,j}^{3-4} + \sum_{i=1}^{N}\sum\nolimits_{j=7}^{12}\left(\frac{12-j}{11}\right)\Delta f_{i,j}^{5-6}\right] = 0$$

(3.21)

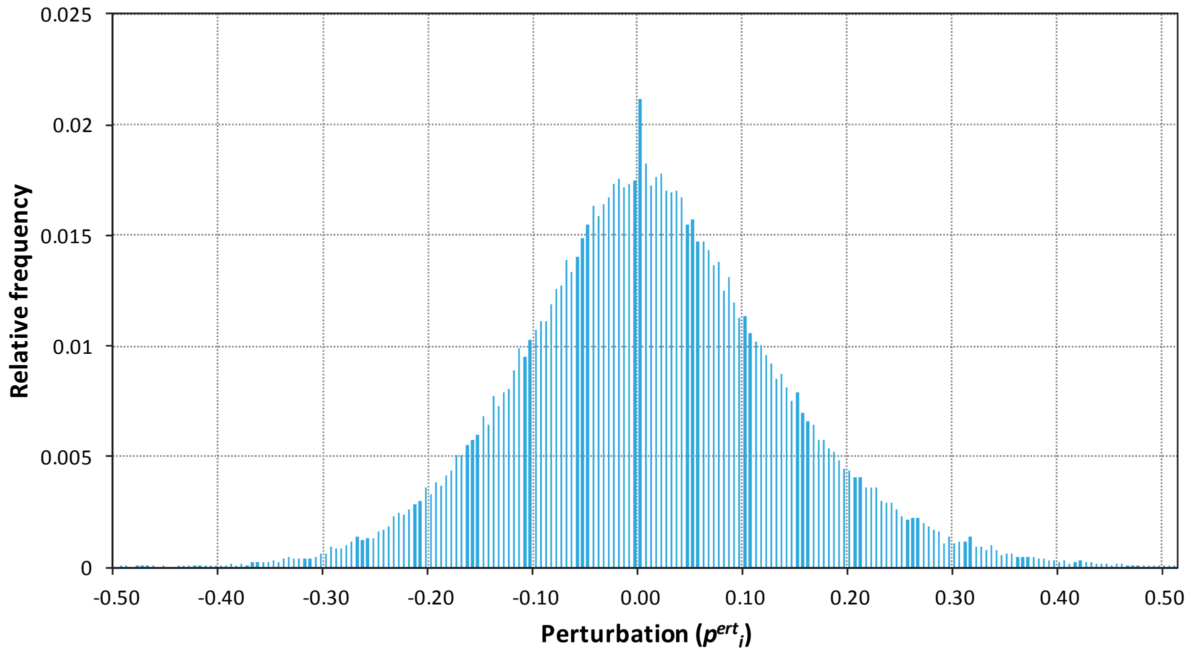


Figure 3.11 Probability density plot of fragment-based perturbations.

3.3.4 Disaggregating stochastic annual rainfalls

In aiming to keep the method applied as simple as possible yet robust, it was decided to disaggregate the final annual stochastic rainfall $\boldsymbol{A}^{sf}_i$ using the monthly fragments selected from any year of the sequence. This was considered realistic because: i) the observed monthly serial correlations were negligible (Figure 3.5), ii) no strong dependence of monthly rainfall distributions on annual rainfall characteristics was detected (Table 3.3), and iii) there is a large number of possible monthly fragments that would be available and the actual monthly flows would be highly varied as they would be obtained from perturbed stochastic annual rainfalls. The final stochastic monthly rainfall $m_{i,j}^{sf}$ for month *j* of year *i* is obtained as

$$m_{i,j}^{sf} = f_{i,j}^{s,7} A_i^{sf} \quad for\ j = 1, 2, \ldots\ldots, 12 \tag{3.22}$$

Where $f_{i,j}^{7} = \frac{m_{r,j}}{A_r} \quad for\ j = 1, 2, \ldots\ldots, 12$ is the fragment for month ***j*** of a randomly selected year ***r*** and $A_i^{s,f}$ is the final perturbed annual rainfall obtained from equation 3.11.

This approach was found to get rid of the biases observed with the weighted fragment-based disaggregation. The STOMSA stochastic streamflow generator (Van Rooyen and Mckenzie, 2004) that is widely used in South Africa also uses historic fragments to disaggregate annual streamflows although this leads to gross under-estimation of the monthly serial correlation between the end of one year and the beginning of the next (Ndiritu, 2011a).

3.3.5 Modelling spatial cross correlations

In order to preserve spatial cross correlations in generating the annual stochastic rainfalls and in their disaggregation, the contemporaneous approach (Srinivas and Srinivasan, 2001, 2005) is used. Each rainfall station is used alternately as the lead sequence and the resampling order of blocks obtained using this sequence is adopted for all the other rainfall stations. Similar alternate lead sequencing is also applied in the selection of the years to use for disaggregating and perturbing the initial annual stochastic rainfalls.

4 PERFORMANCE OF NON-PARAMETRIC GENERATOR AND COMPARISON WITH A PARAMETRIC GENERATOR

4.1 Introduction

The performance of the VLB rainfall generator is assessed using the 10 rainfall station dataset described in Section 3.1. For this multi-site problem, 101 stochastic sequences that are 94 years long (as the historic ones) are generated. Generation performance is assessed by finding out how closely the single statistics of historic dataset locate within box plots of the 101 annual and monthly statistics obtained from the generated series. The statistics applied are the mean, the median, the 25th and the 75th percentile, the lowest and the highest rainfall, the standard deviation, the skewness, and the serial and cross correlation coefficients. In addition, the minimum run sums test that compares the minimum historic and stochastic annual rainfalls lengths ranging from 1-24 years is included to assess the replication of long-term dependence characteristics. As indicated in Section 3.3, the generator requires the minimum block length and the range of proportion of exceedance that defines a dry year. After some trial tests, a minimum block length of 3 years and a proportion exceedance range of 60 to 90% was adopted.

The comparison of the VLB generator with a parametric generator required the choice of the parametric generator. As mentioned in Chapter 1 and 2 of the literature review, stochastic rainfall generation has focused more on daily rainfall generation while there is an on-going WRC project (WRC Project K5/2155) that is assessing the PEGRAIM-W monthly multi site stochastic rainfall model (Pegram, 2011) that was commissioned by the Department of Water Affairs (DWA). Given these reasons, the PEGRAIM-W parametric model was the natural choice for comparison with the VLB non-parametric generator. The 10 rainfall station generation problem used to test the VLB model was used for the comparison. The PEGRAIM-W model was therefore used to generate 101 stochastic rainfall sequences and the 10 annual and monthly statistics used to assess the VLB were computed for these sequences. Box plots of the two generators form the main basis of the comparison.

Section 4.2 presents the performance and comparison of the two generators using the annual statistics while performance and comparison using monthly statistics is given in Section 4.3. Section 4.4 then summarizes the Chapter.

4.2 Comparison of Parametric and Non-Parametric Generator using Annual Statistics

The statistics applied for the comparison are the mean, the median, the 25th and the 75th percentile, the lowest and the highest rainfall, the standard deviation, the skewness, and the serial and cross correlation coefficients. In addition, the minimum run sums test that compares the minimum historic and stochastic annual rainfalls lengths ranging from 1-24 years is included to assess the replication of long-term dependence characteristics. Box plots of these statistics with the respective historic statistics superimposed are presented in Figure 4.1 to Figure 4.20.

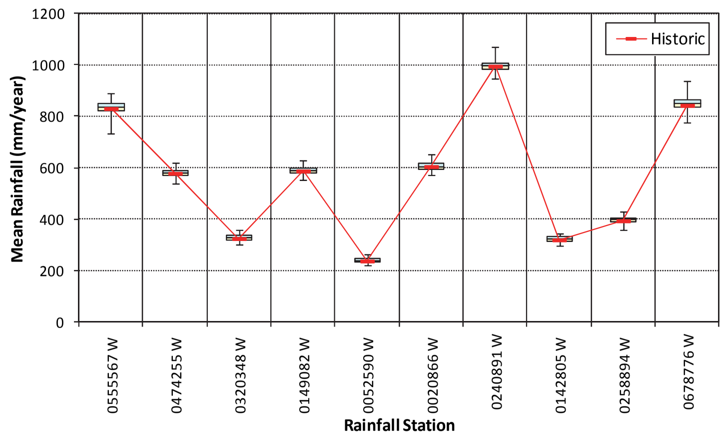


Figure 4.1a Box plots of annual mean rainfall from VLB generator

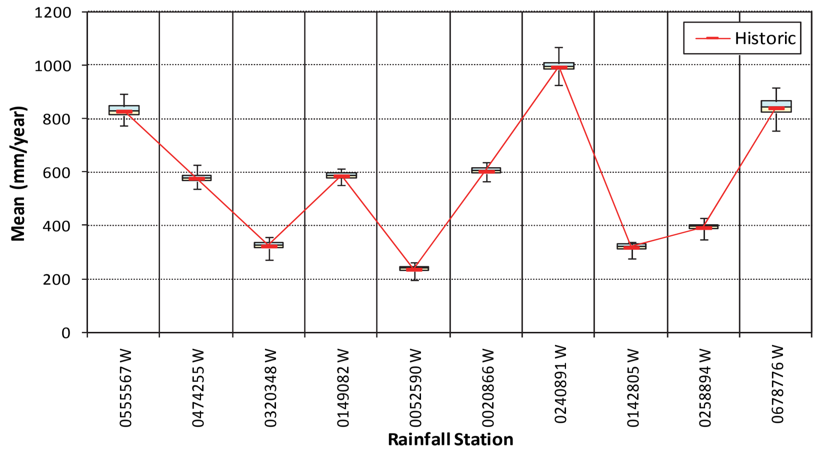


Figure 4.1b Box plots of annual mean rainfall from PEGRAIM-W

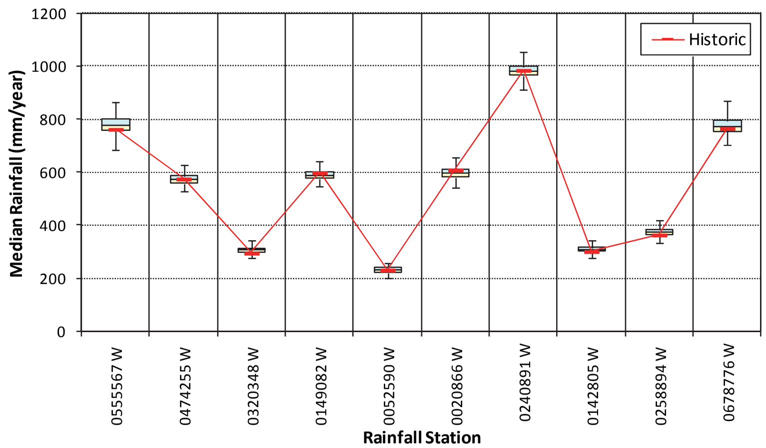


Figure 4.2a Box plots of annual median rainfall from VLB generator

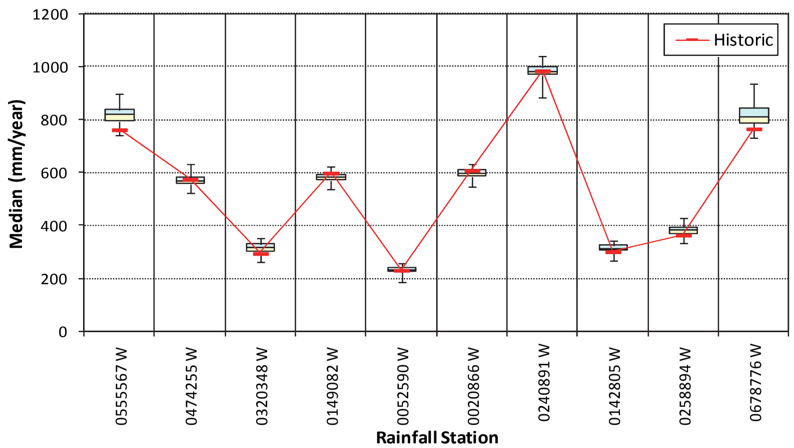


Figure 4.2b Box plots of annual median rainfall from PEGRAIM-W

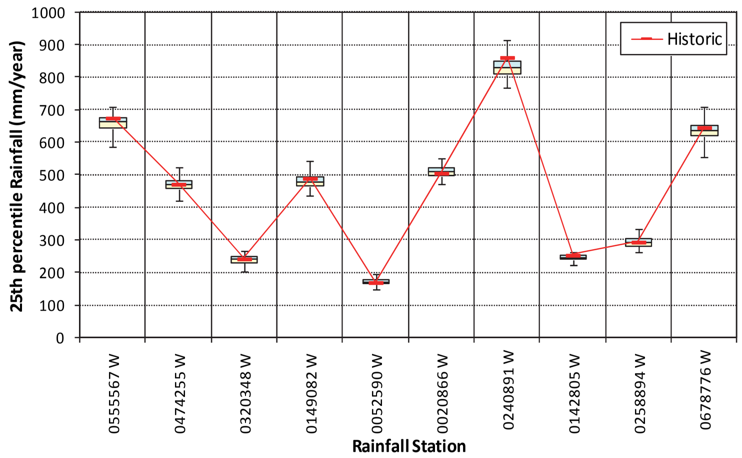


Figure 4.3a Box plots of annual 25th percentile rainfalls from VLB generator

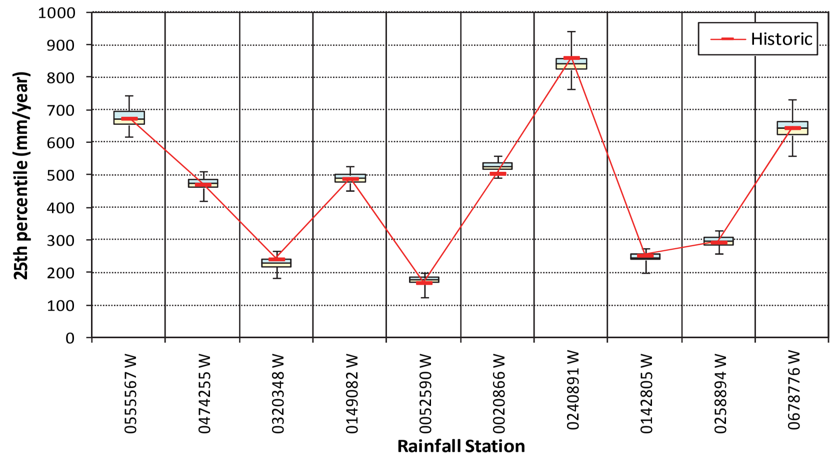


Figure 4.3b Box plots of annual 25th percentile rainfalls from PEGRAIM-W

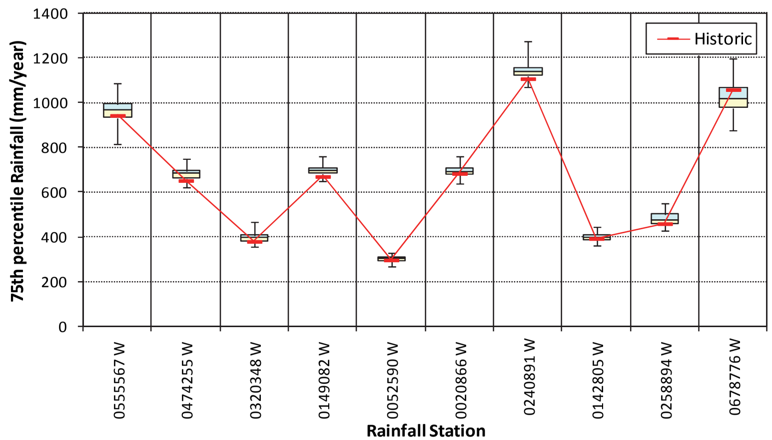


Figure 4.4a Box plots of annual 75th percentile rainfalls from VLB generator

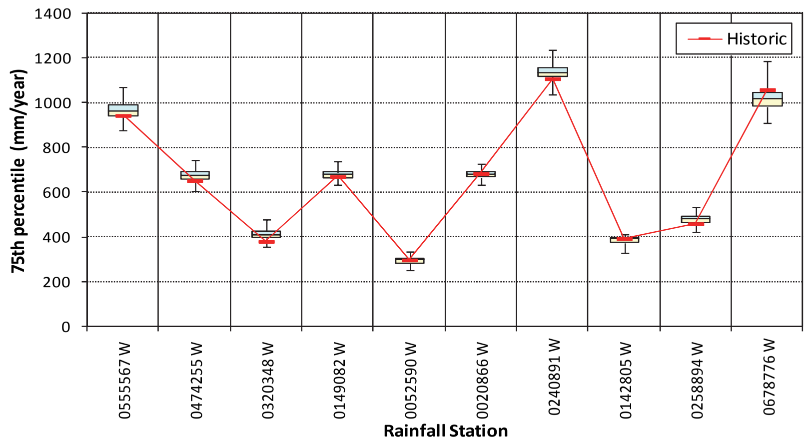


Figure 4.4b Box plots of annual 75th percentile rainfalls from PEGRAIM-W

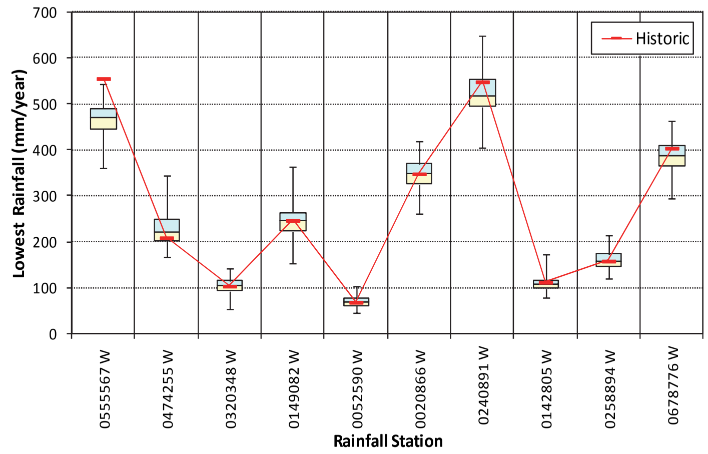


Figure 4.5a Box plots of lowest annual rainfalls from VLB generator

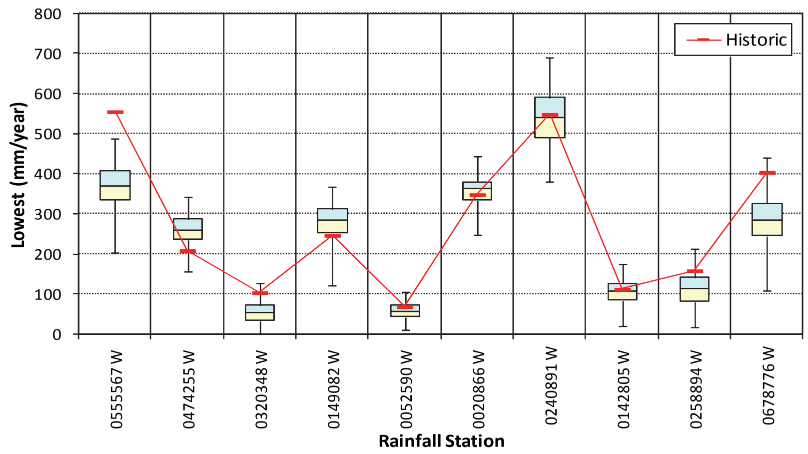


Figure 4.5b Box plots of lowest annual rainfalls from PEGRAIM-W

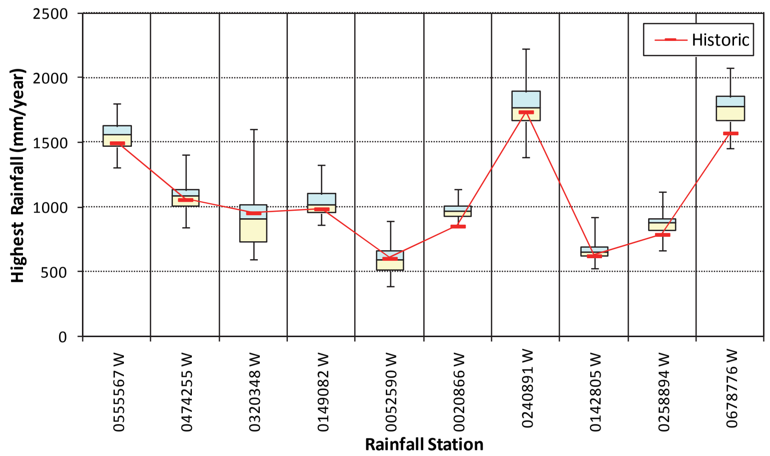


Figure 4.6a Box plots of highest annual rainfalls from VLB generator

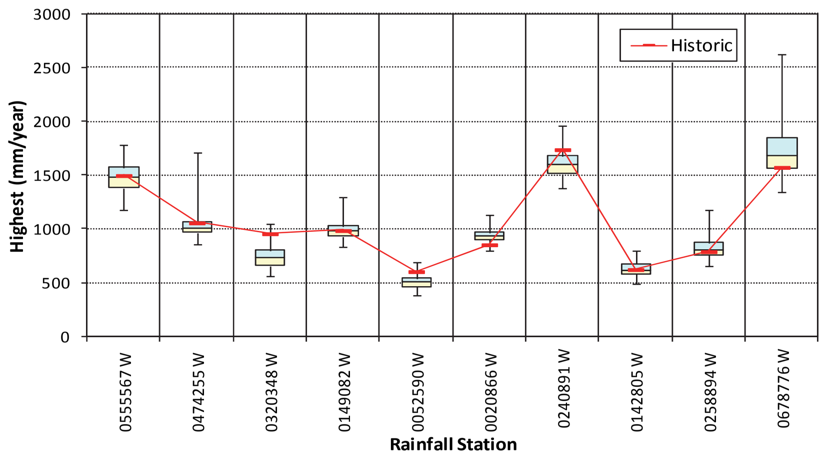


Figure 4.6b Box plots of highest annual rainfalls from PEGRAIM-W

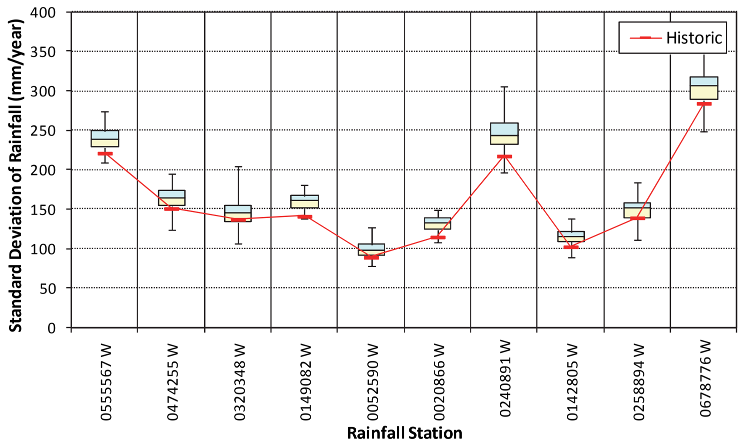


Figure 4.7a Box plots of standard deviation of annual rainfalls from VLB generator

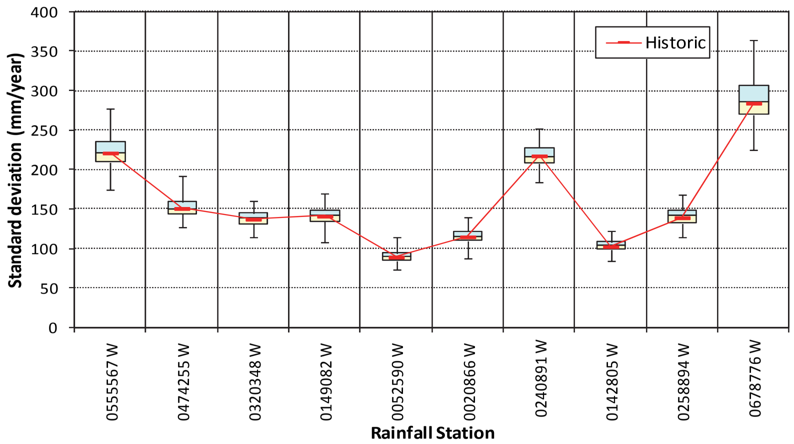


Figure 4.7b Box plots of standard deviation of annual rainfalls from PEGRAIM-W

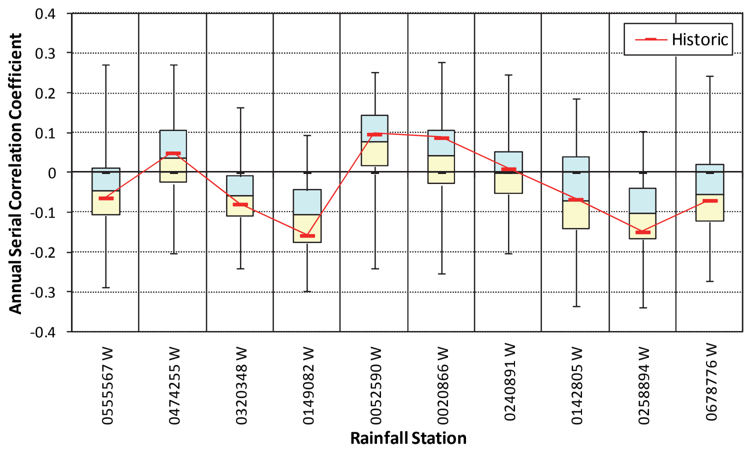


Figure 4.8a Box plots of annual serial correlation coefficient of rainfalls from VLB generator

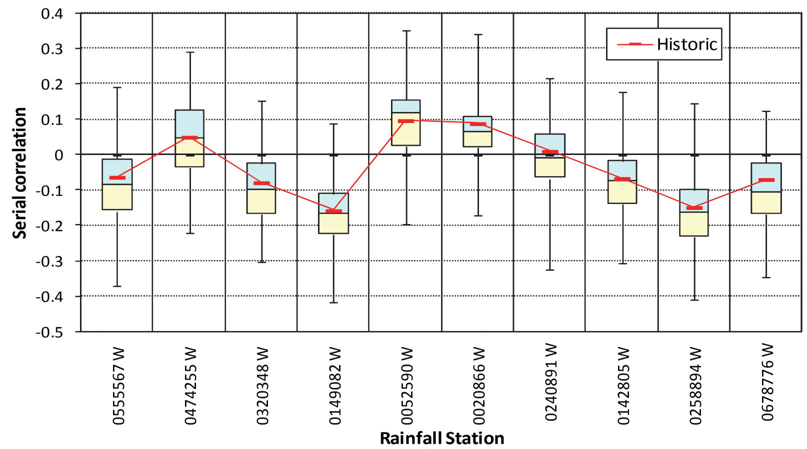


Figure 4.8a Box plots of annual serial correlation coefficient of rainfalls from PEGRAIM-W

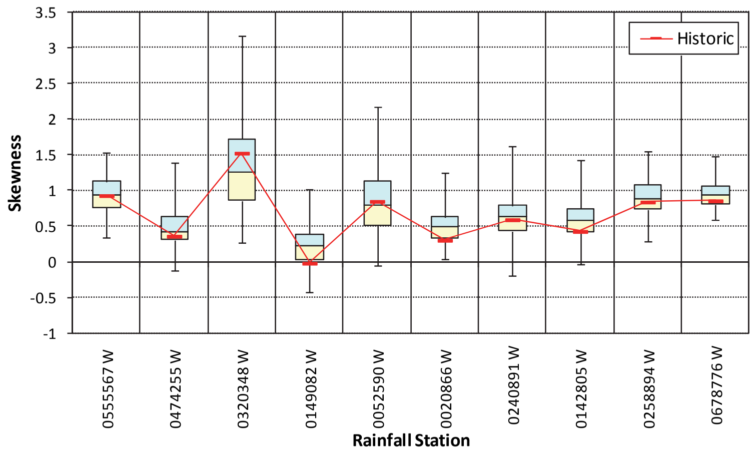


Figure 4.9a Box plots of skewness of annual rainfalls from VLB generator

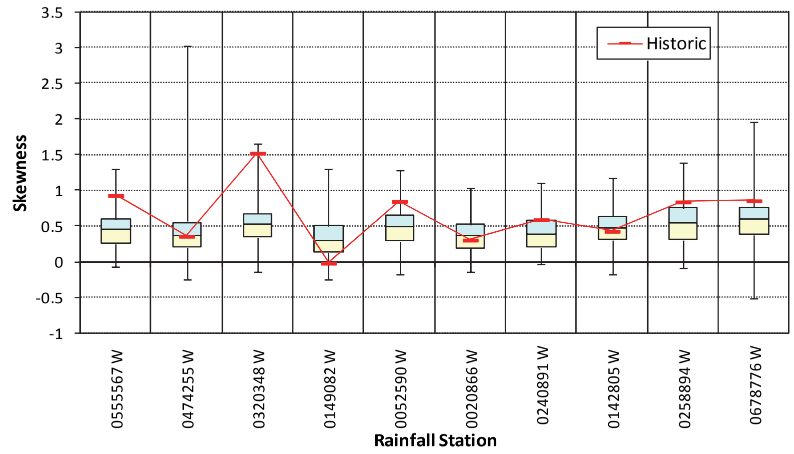


Figure 4.9b Box plots of skewness of annual rainfalls from PEGRAIM-W

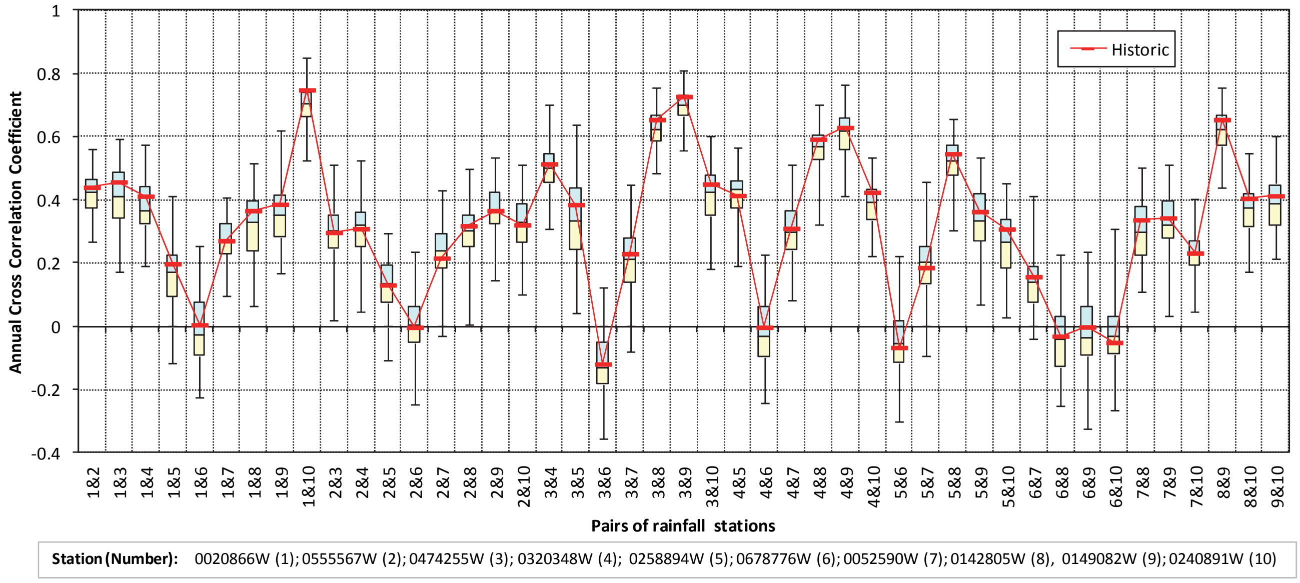


Figure 4.10a Box plots of annual cross correlation coefficient from VLB generator

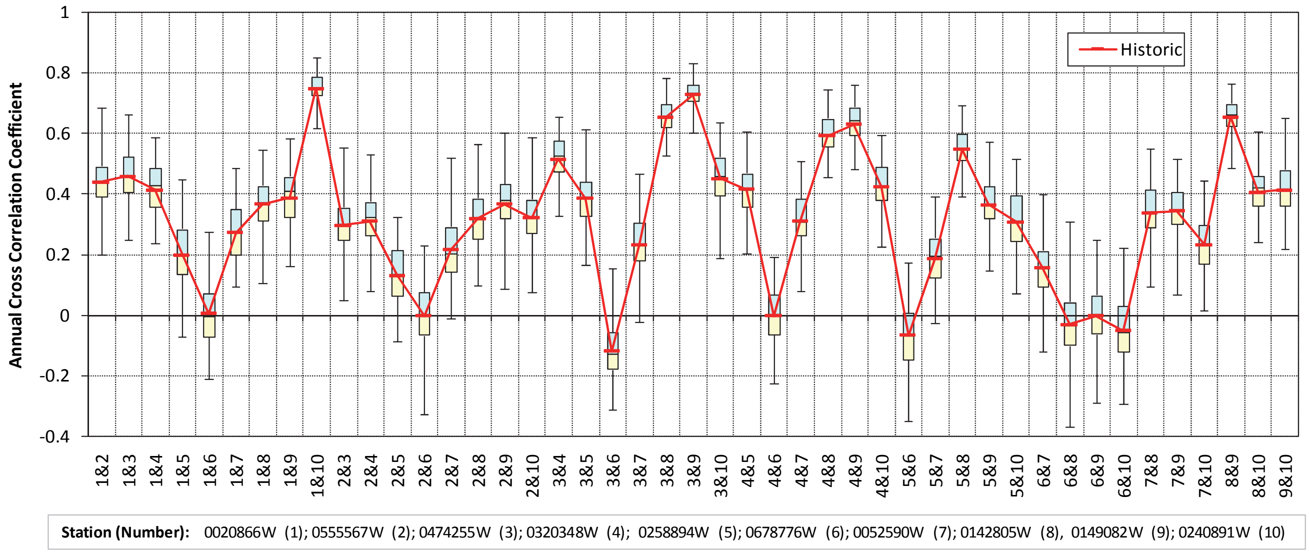


Figure 4.10b Box plots of annual cross correlation coefficient from PEGRAIM-W

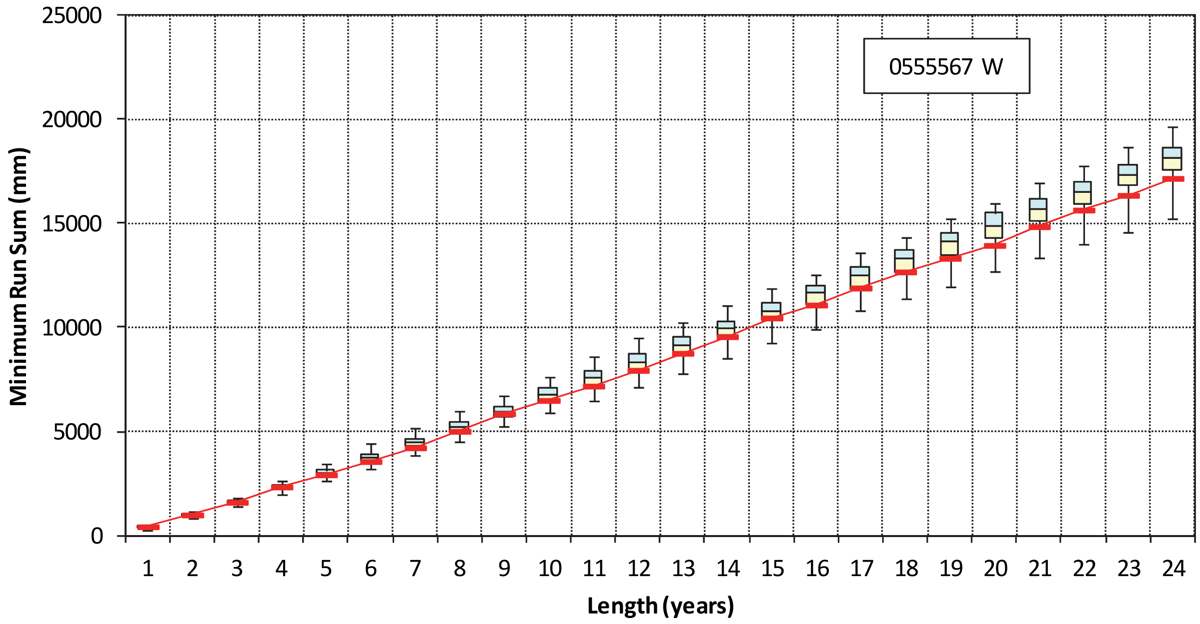


Figure 4.11a Box plots of minimum run sums for station 0555567 W from VLB generator

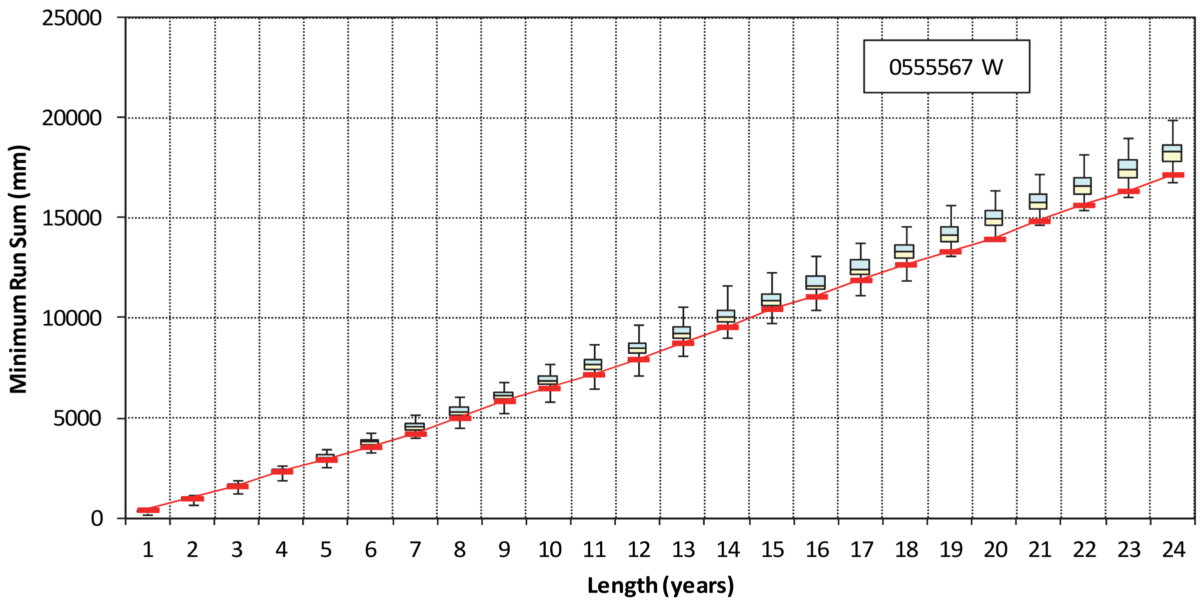


Figure 4.11b Box plots of minimum run sums for station 0555567 W from PEGRAIM-W

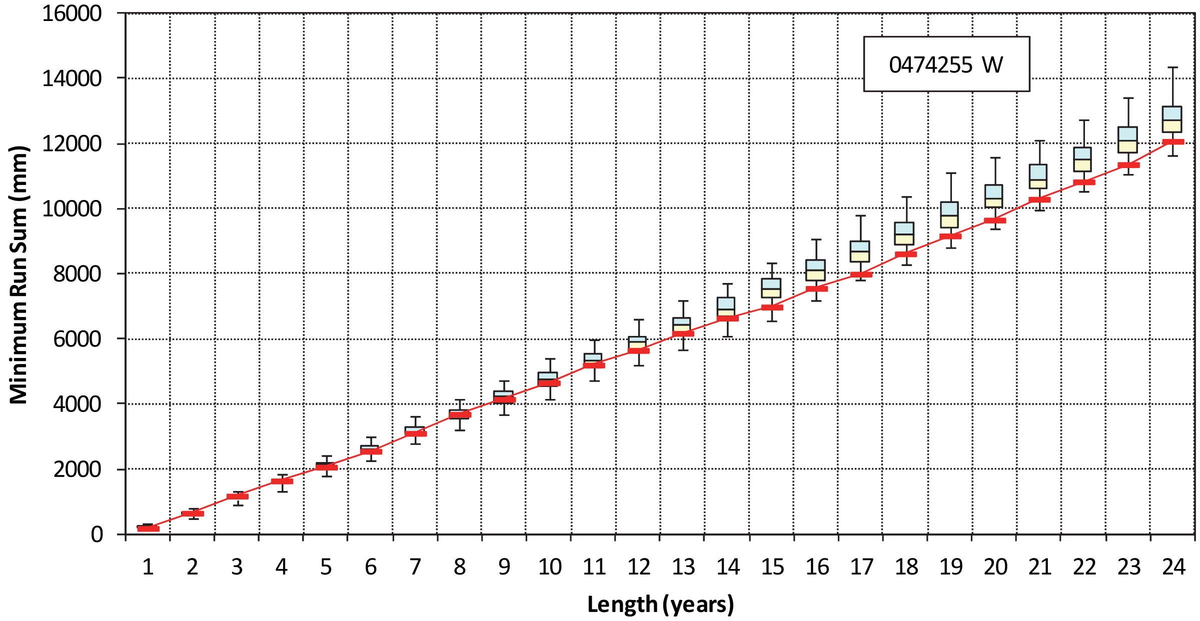


Figure 4.12a Box plots of minimum run sums for station 0474255 W from VLB generator

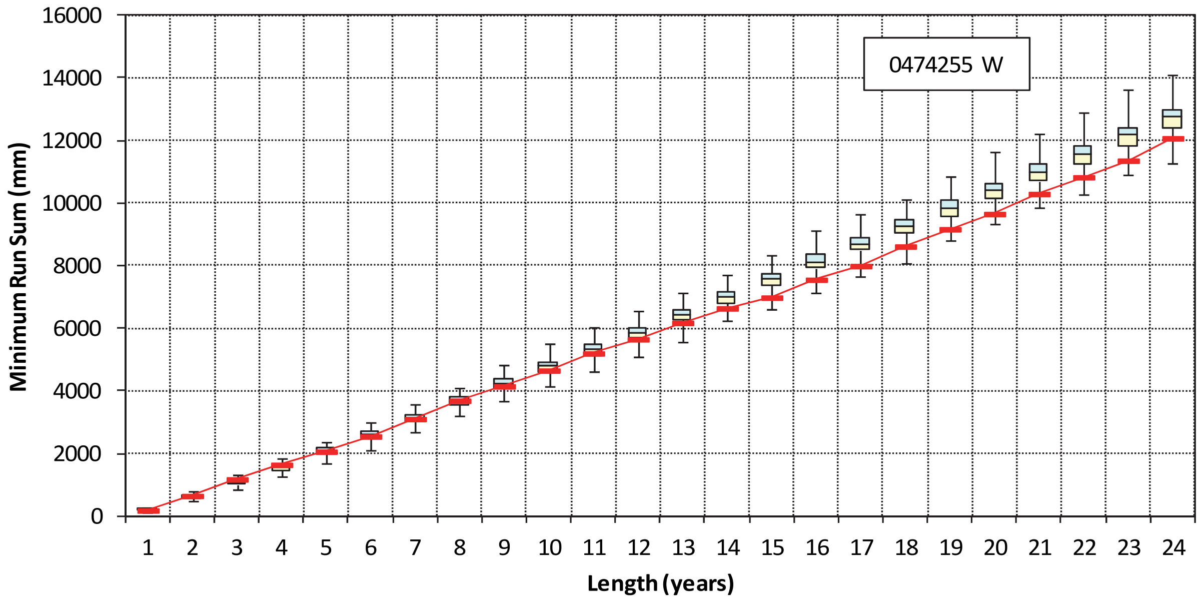


Figure 4.12b Box plots of minimum run sums for station 0474255 W from PEGRAIM-W

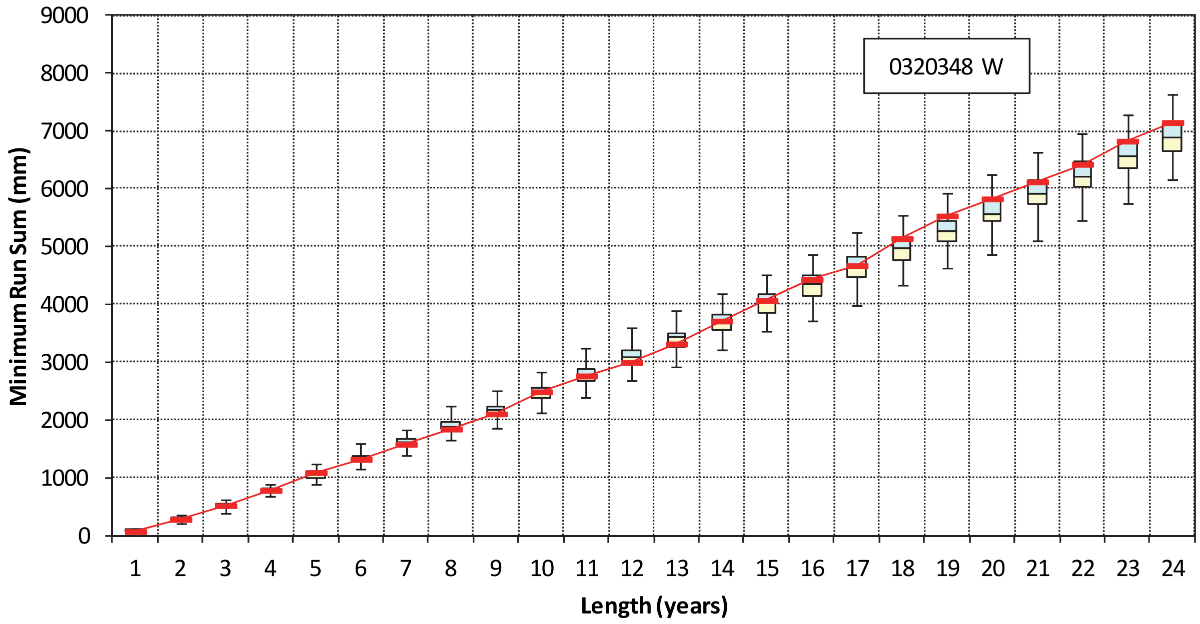


Figure 4.13a Box plots of minimum run sums for station 0320348 W from VLB generator

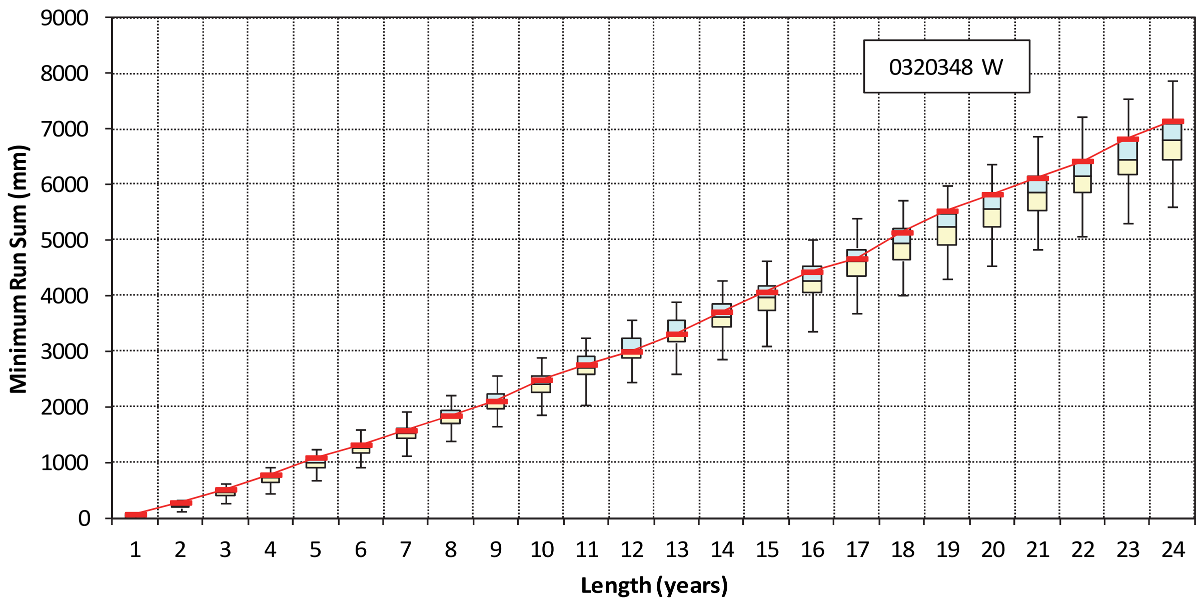


Figure 4.13b Box plots of minimum run sums for station 0320348 W from PEGRAIM-W

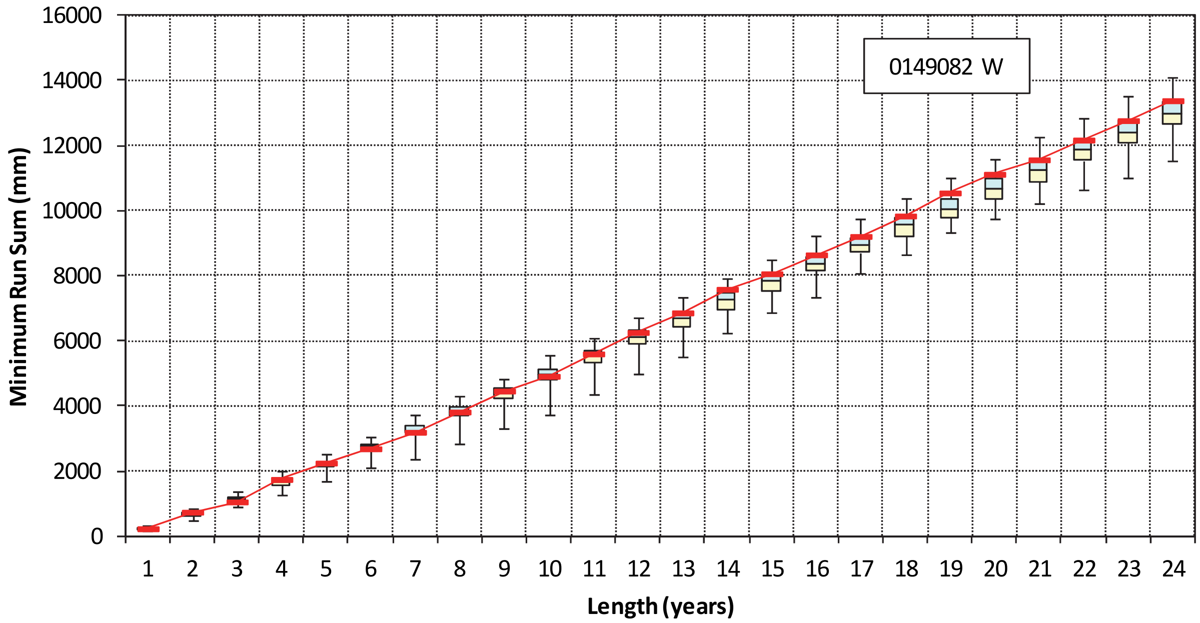


Figure 4.14a Box plots of minimum run sums for station 0149082 W from VLB generator

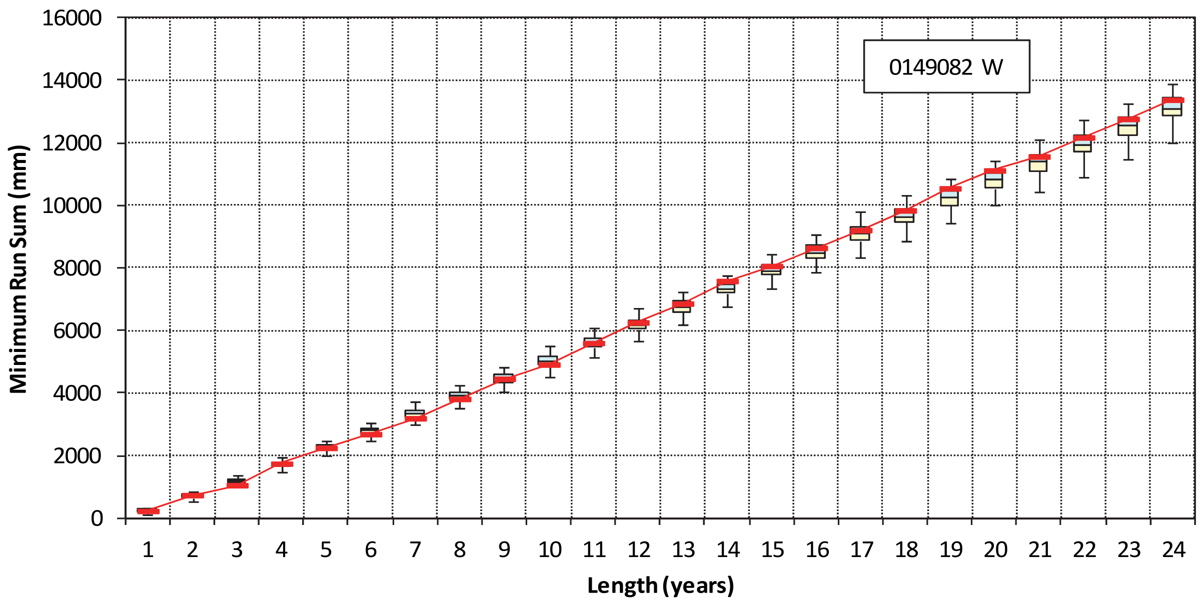


Figure 4.14b Box plots of minimum run sums for station 0149082 W from PEGRAIM-W

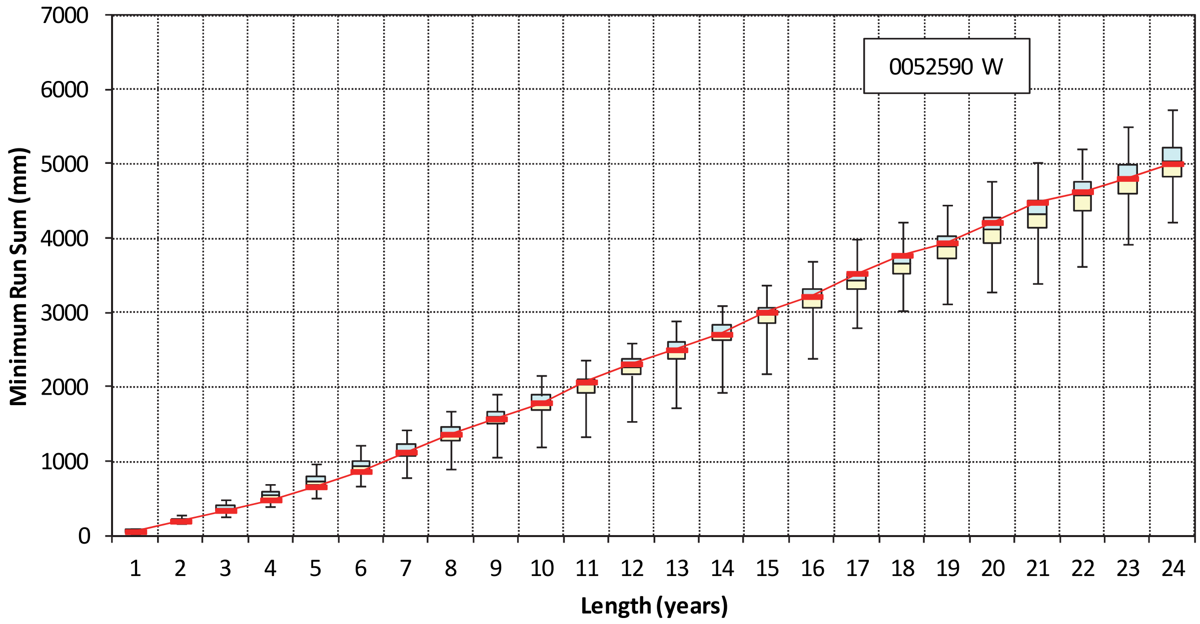


Figure 4.15a Box plots of minimum run sums for station 0052590 W from VLB generator

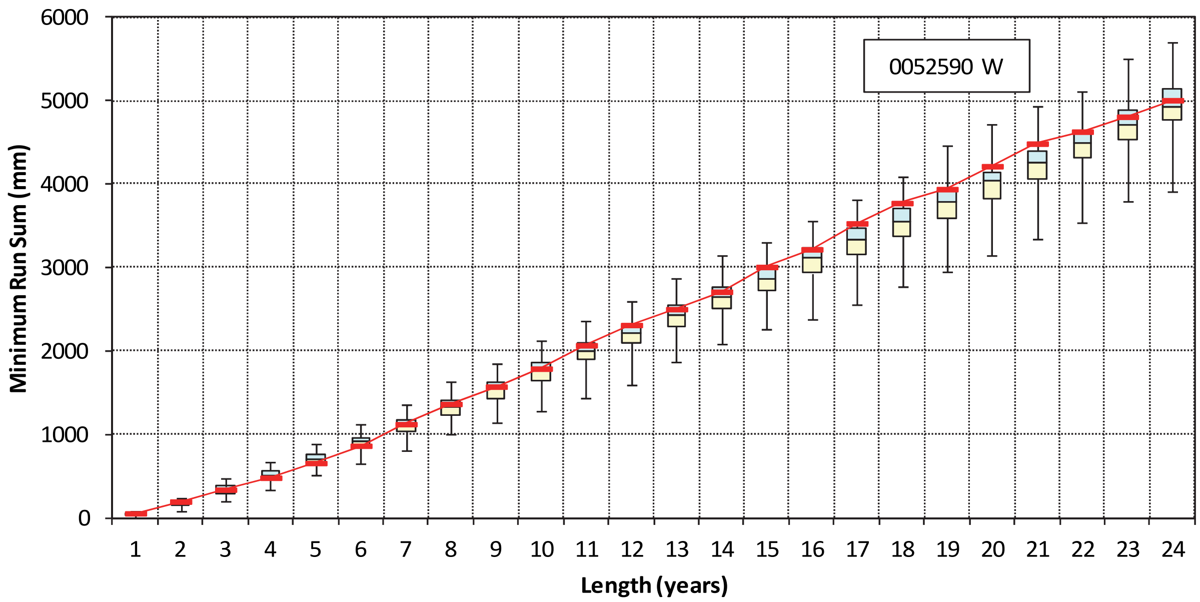


Figure 4.15b Box plots of minimum run sums for station 0052590 W from PEGRAIM-W

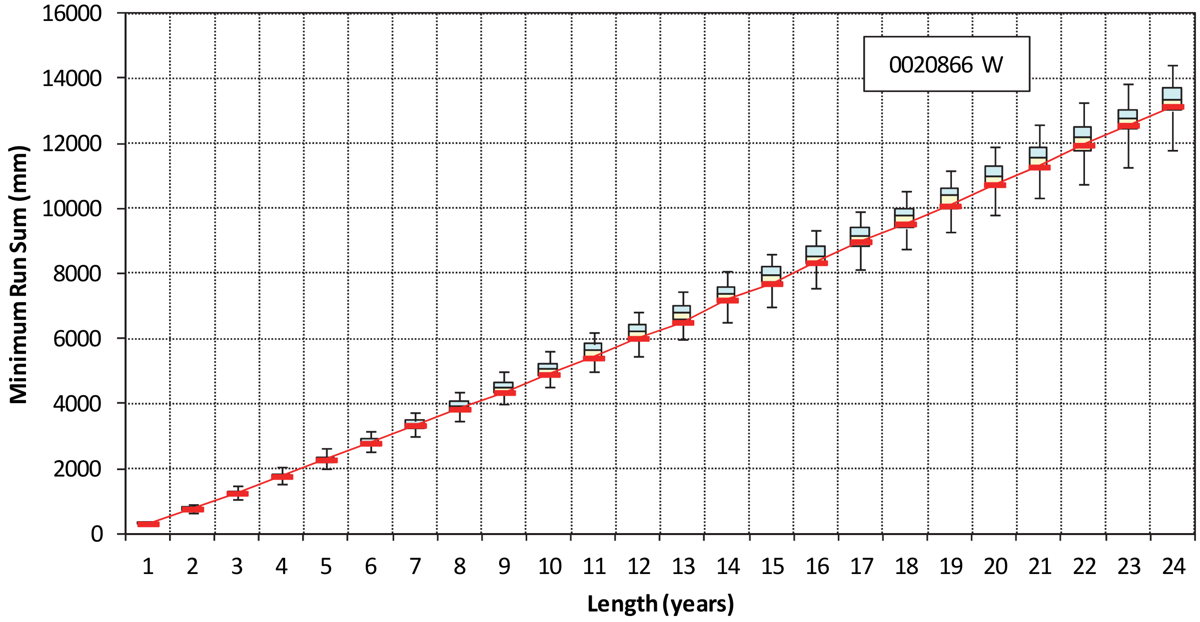


Figure 4.16a Box plots of minimum run sums for station 0020866 W from VLB generator

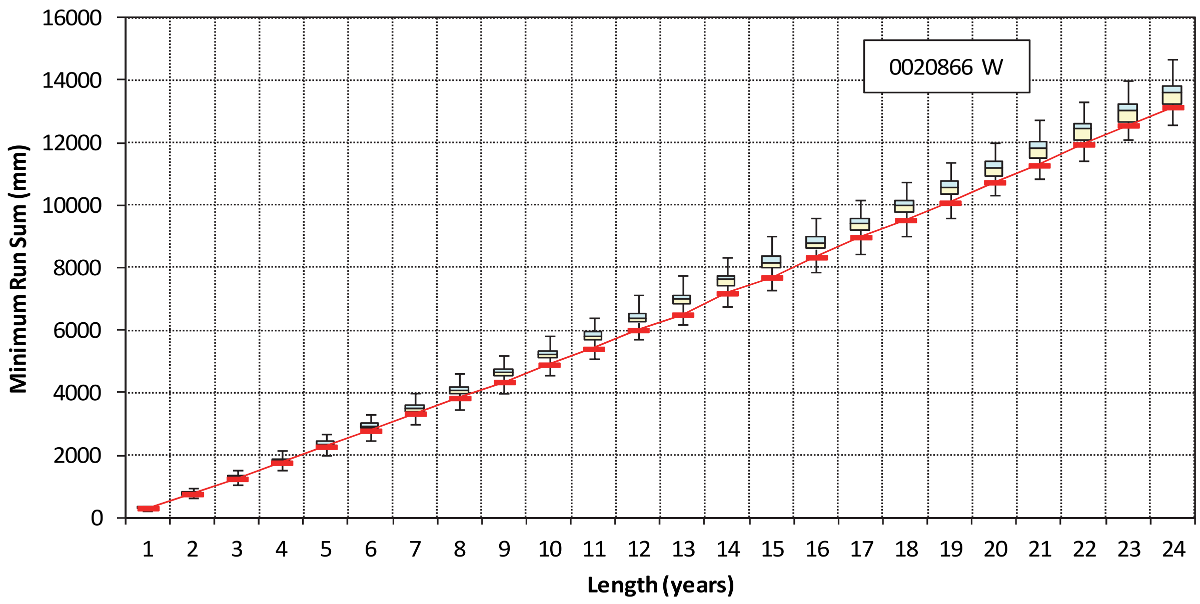


Figure 4.16b Box plots of minimum run sums for station 0020866 W from PEGRAIM-W

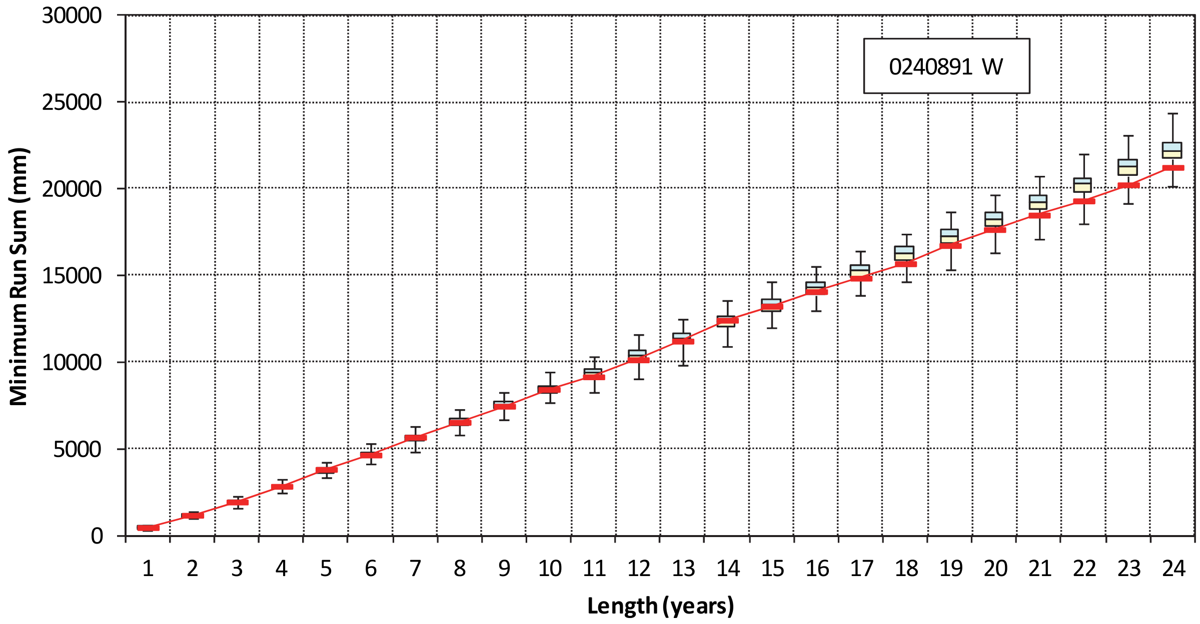


Figure 4.17a Box plots of minimum run sums for station 0240891 W from VLB generator

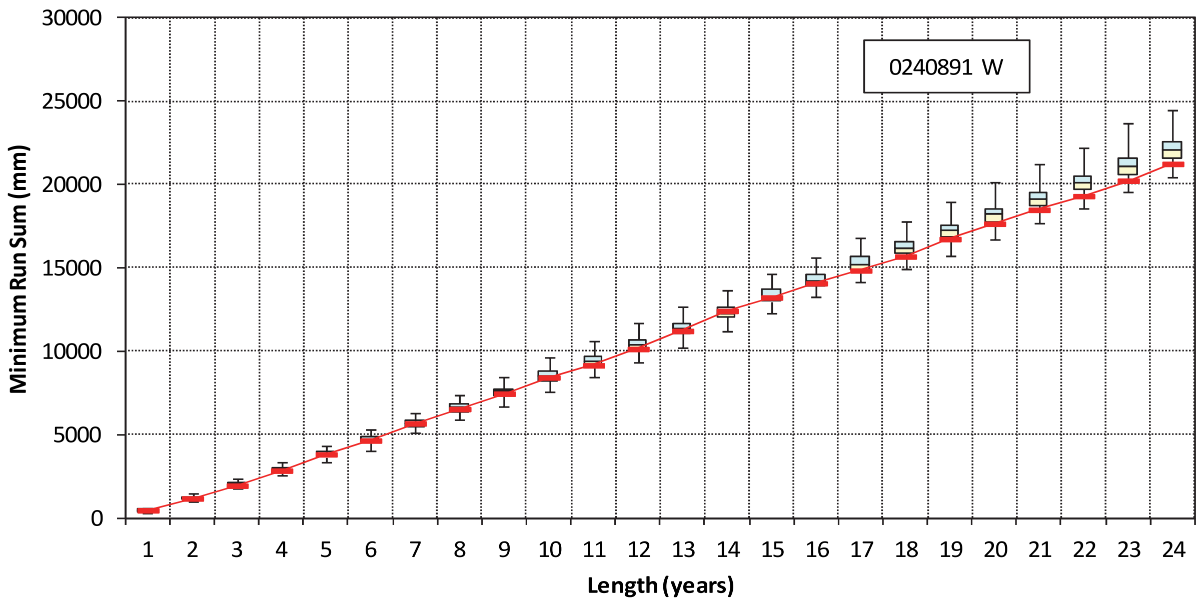


Figure 4.17b Box plots of minimum run sums for station 0240891 W from PEGRAIM-W

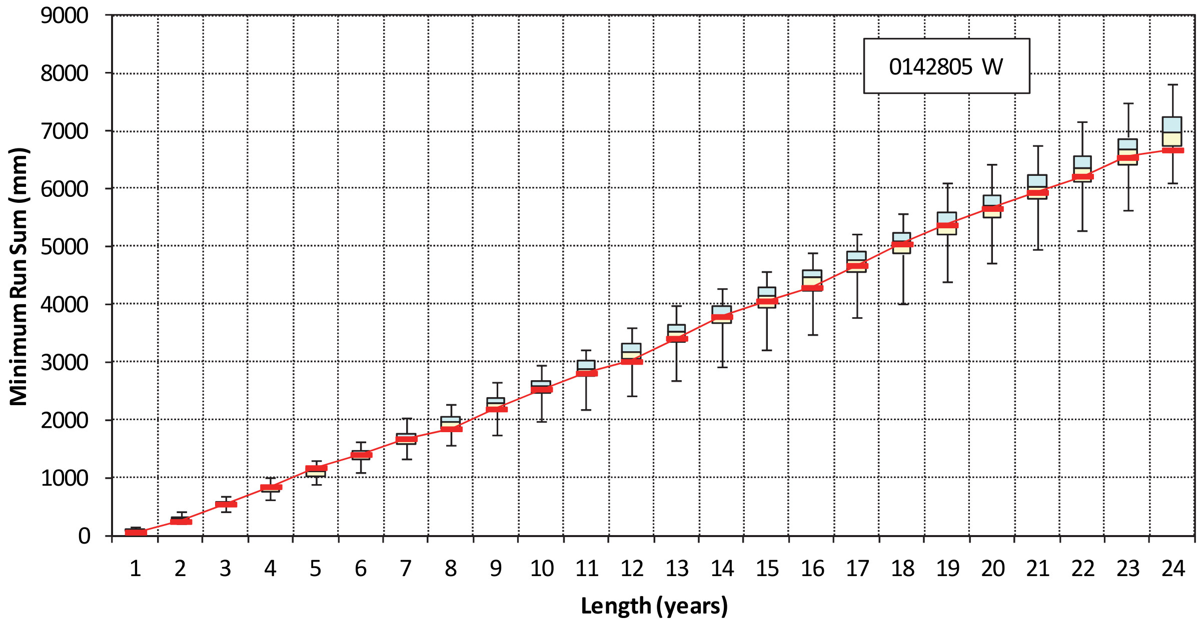


Figure 4.18a Box plots of minimum run sums for station 0142805 W from VLB generator

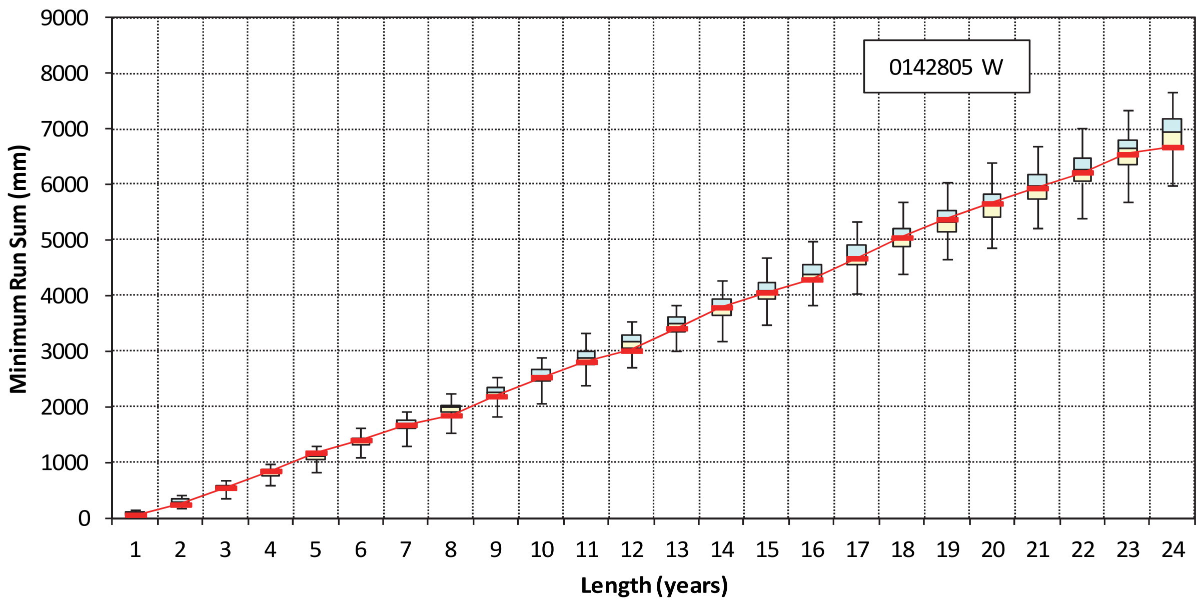


Figure 4.18b Box plots of minimum run sums for station 0142805 W from PEGRAIM-W

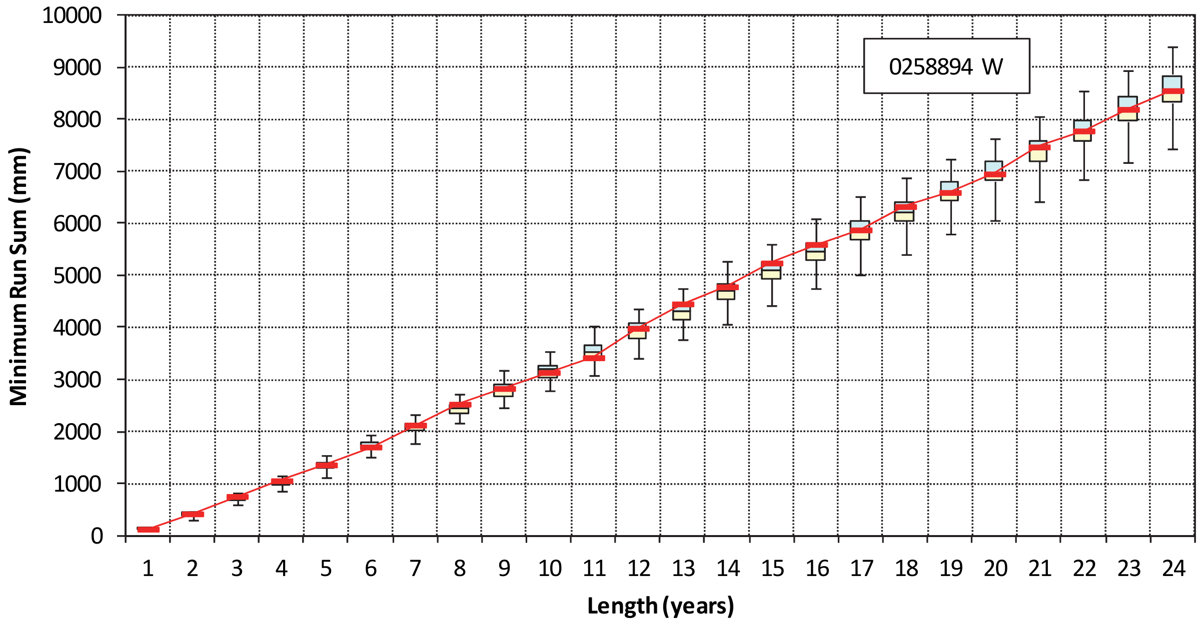


Figure 4.19a Box plots of minimum run sums for station 0258894 W from VLB generator

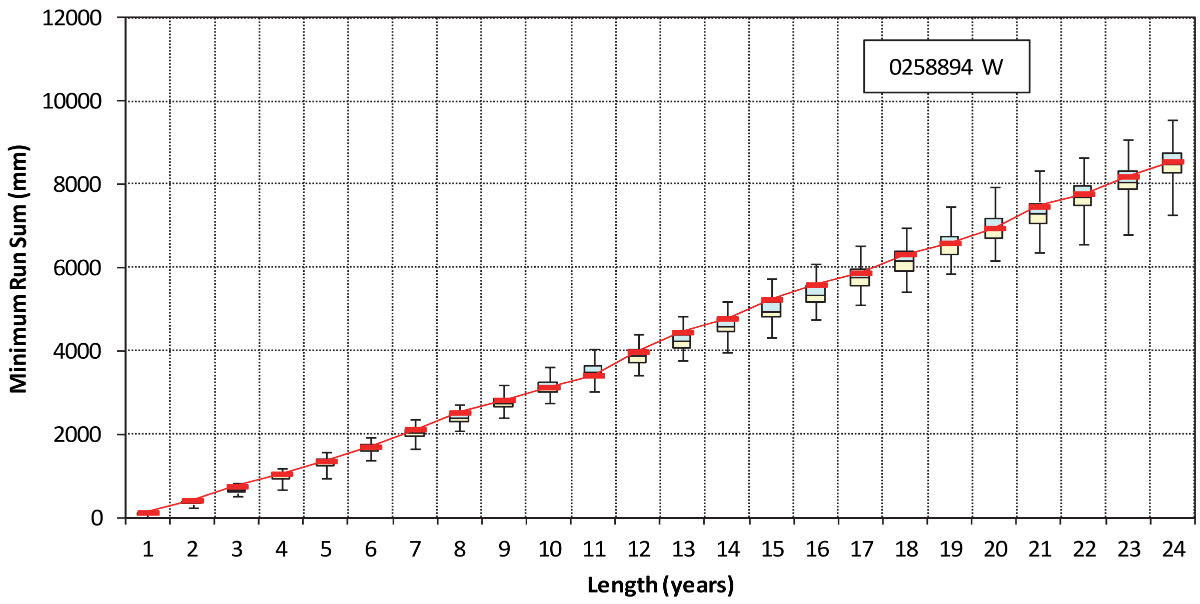


Figure 4.19b Box plots of minimum run sums for station 0258894 W from PEGRAIM-W

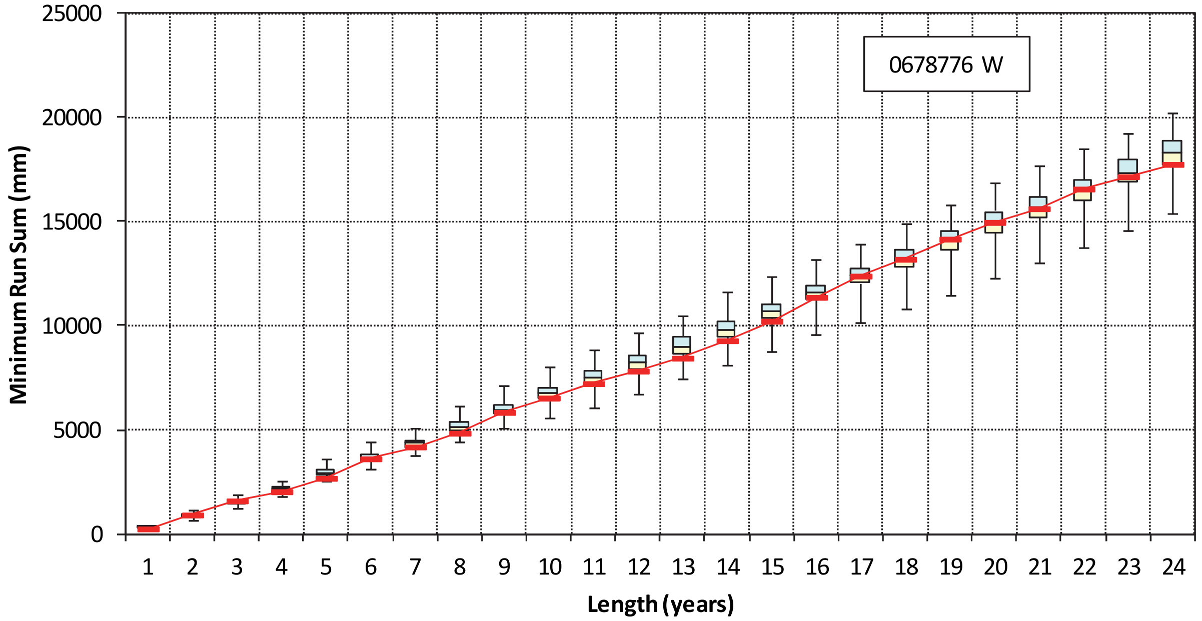


Figure 4.20a Box plots of minimum run sums for station 0678776 W from VLB generator

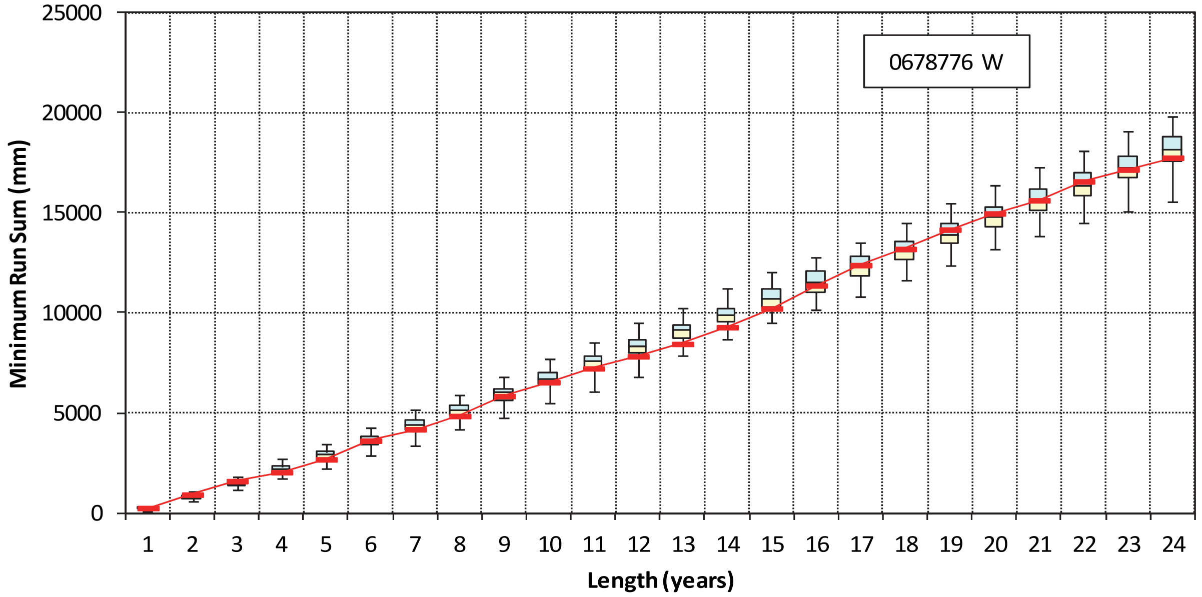
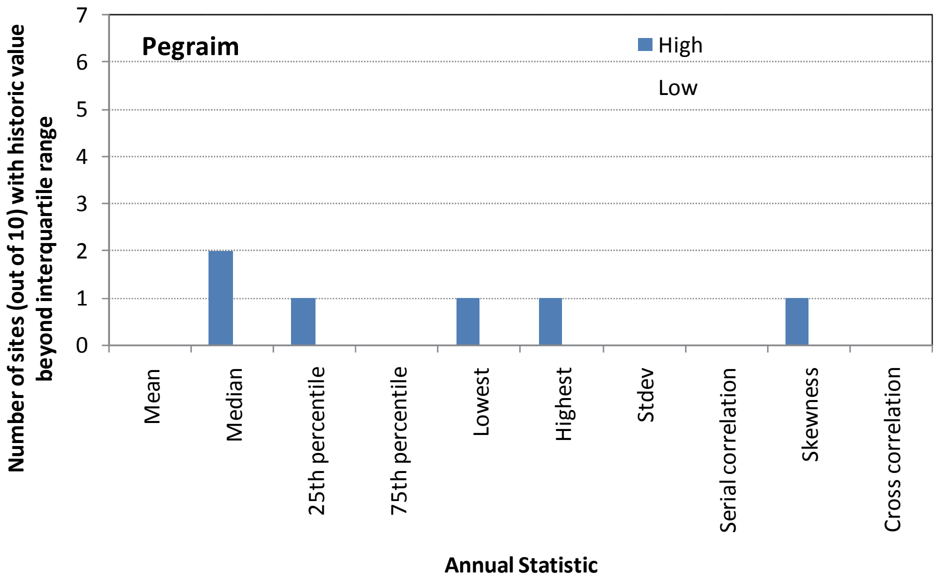


Figure 4.20b Box plots of minimum run sums for station 0678776 W from PEGRAIM-W

The box plots in Figures 4.1 to 4.20 reveal that both PEGRAIM-W and the VLB generators are able to replicate most of the historic statistics well with 82 and 90% of the historic statistics falling within the inter-quartile range of the box plots for PEGRAIM-W and VLB respectively. Figure 4.21 shows the number of times that the historic annual statistics fall beyond the inter-quartile ranges for PEGRAIM-W and VLB respectively with an indication whether more of the stochastic rainfalls fall higher or lower than the historic values.



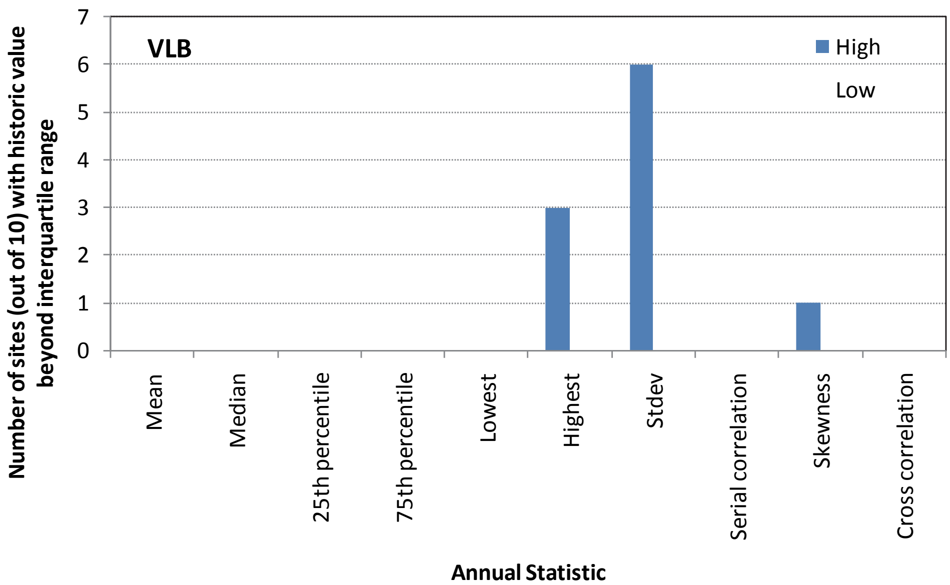


Figure 4.21 The number of times that annual historic statistics fall beyond the inter-quartile range of the box plots of stochastic sequences for PEGRAIM-W and VLB generator

Figure 4.21 reveals that VLB has 3 of the 10 statistics falling beyond the inter-quartile range for some of the stations while PEGRAIM-W has 5. For the cases where there are more than 2 stations with historic values beyond the inter-quartile range, the potential significance of the deviations is assessed by the percentages of the deviation of the median of the stochastic values from the respective historic ones. For PEGRAIM-W, the average deviations for the Lowest annual rainfall, the highest annual rainfall and skewness are 38, 15 and 43 percent respectively. For VLB, the average deviations for the standard deviation of annual rainfall and the highest annual rainfall are 11 and 14 percent respectively.

Both PEGRAIM-W and VLB replicate the minimum run sums reasonably well (Figure 4.11 to 4.20) and are therefore considered to capture the long-term dependencies adequately. There is also a striking similarity of the locations of the historic plots on the box plots for all the 10 sites although the ranges of the box plots are not as closely similar.

4.3 Comparison of Parametric and Non-Parametric Generator using Monthly Statistics

Figures 4.22 to 4.41 are box plots of the same 10 monthly statistics applied to compare annual rainfall generation in Section 4.2. To ease visualization, the monthly box plots are presented in two groups of 5 rainfall stations each.

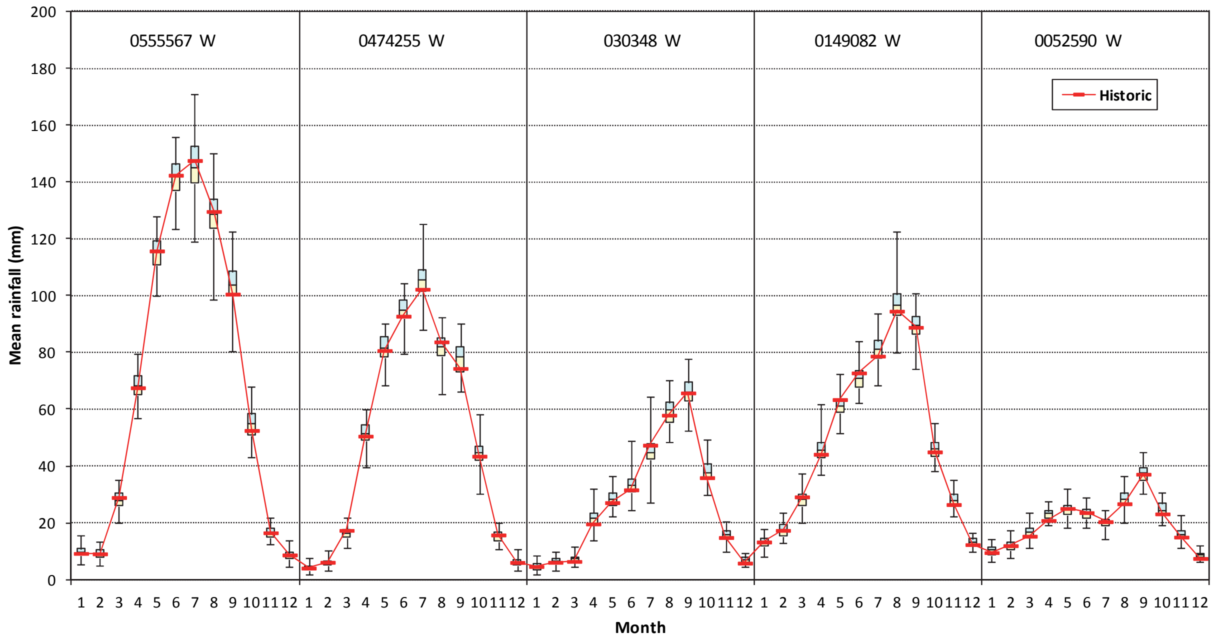


Figure 4.22a Box plots of monthly mean rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

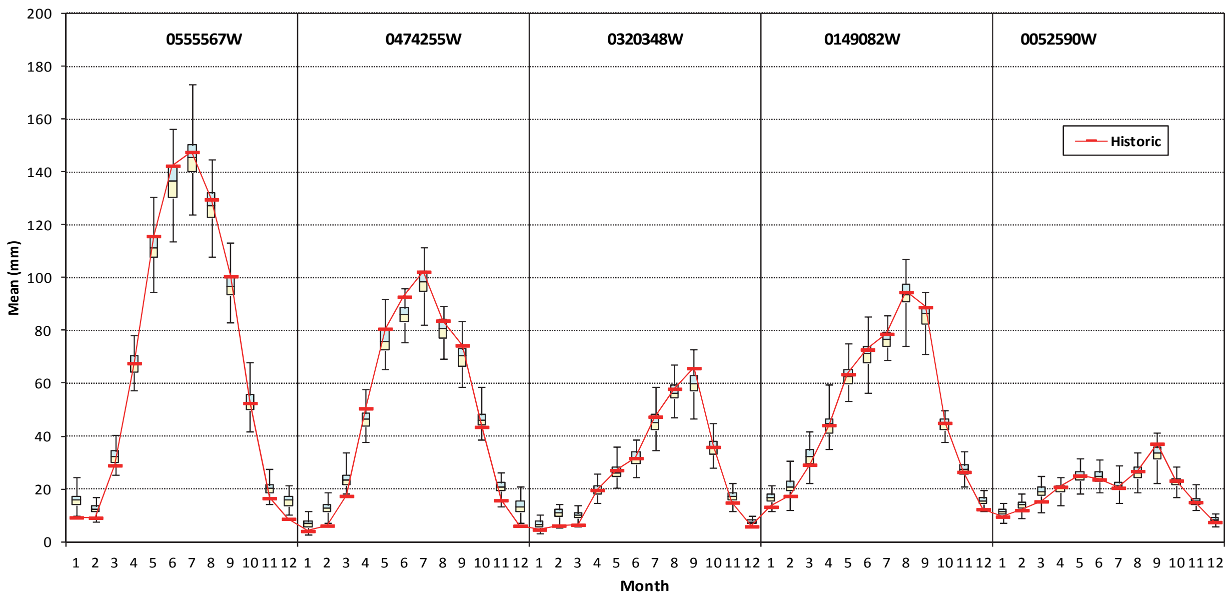


Figure 4.22b Box plots of monthly mean rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

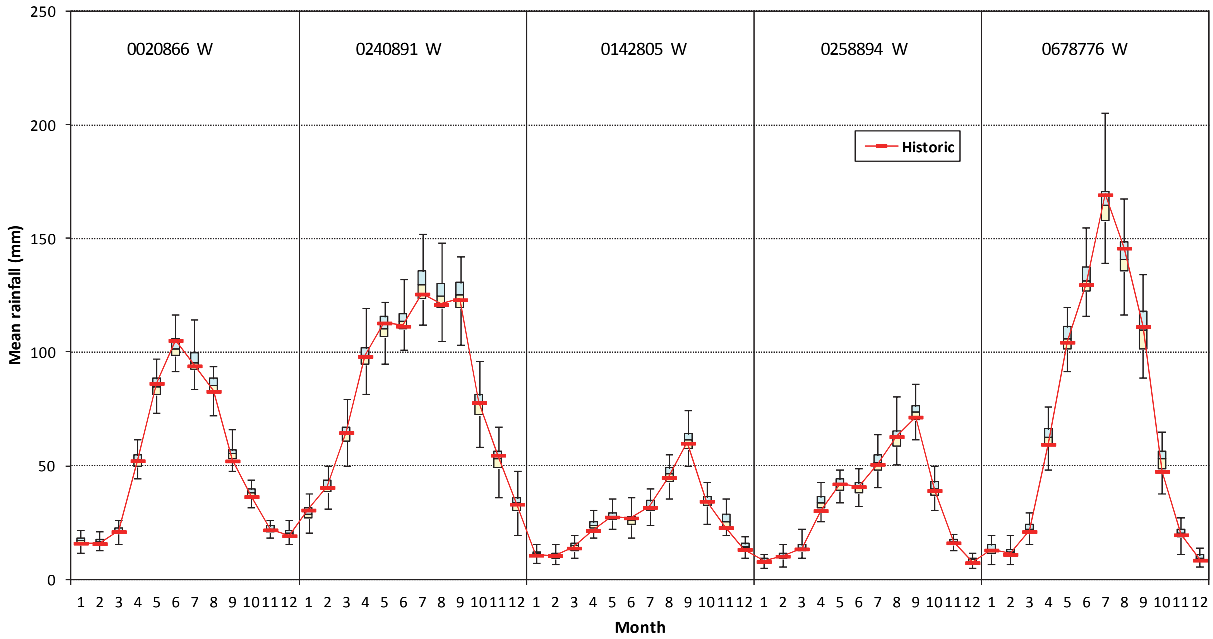


Figure 4.23a Box plots of monthly mean rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

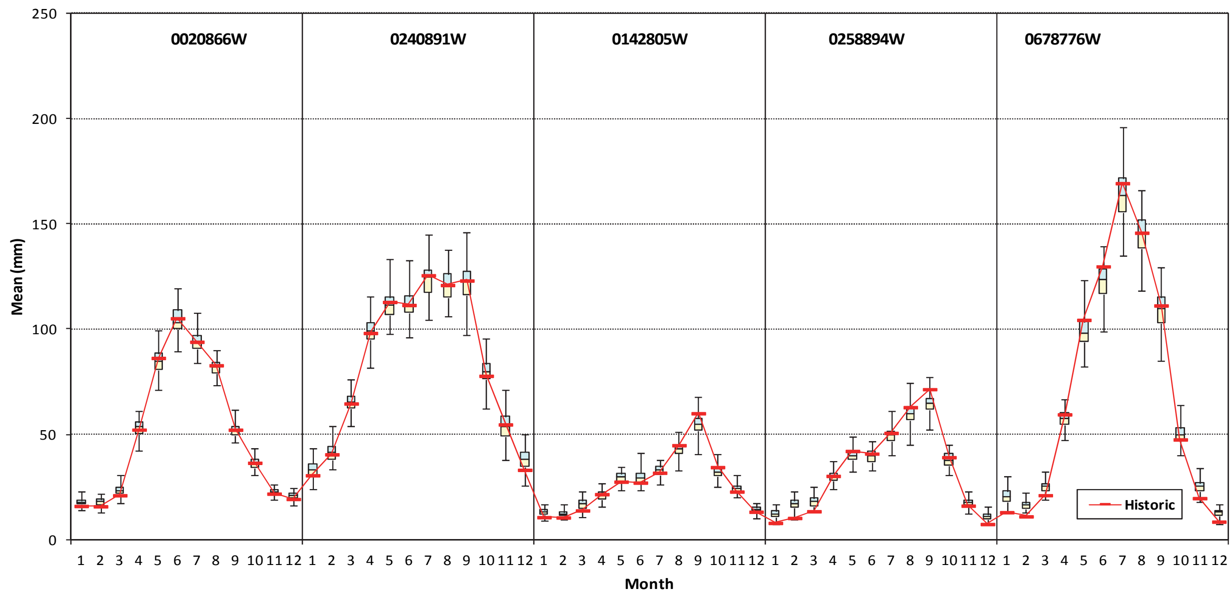


Figure 4.23b Box plots of monthly mean rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

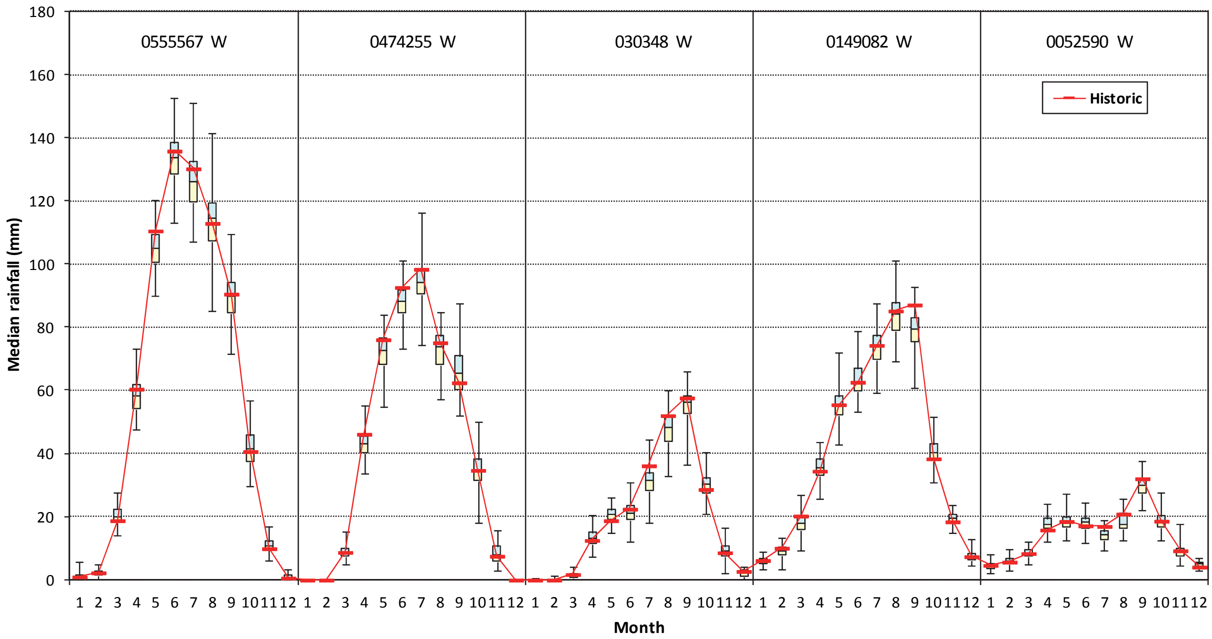


Figure 4.24a Box plots of monthly median rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

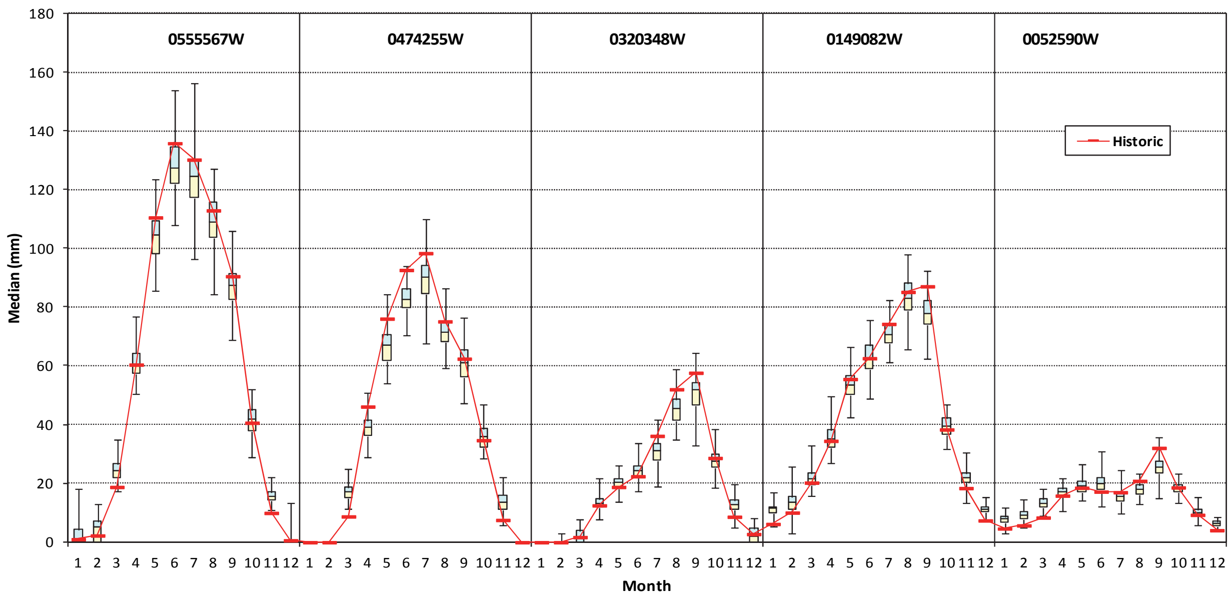


Figure 4.24b Box plots of monthly median rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

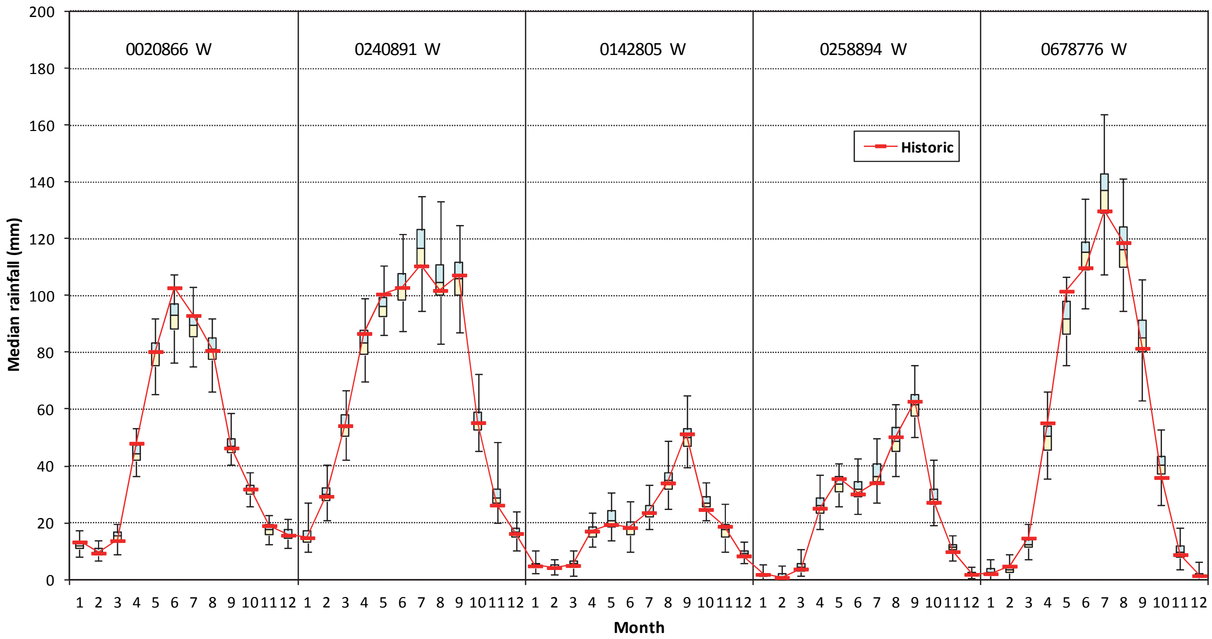


Figure 4.25a Box plots of monthly median rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

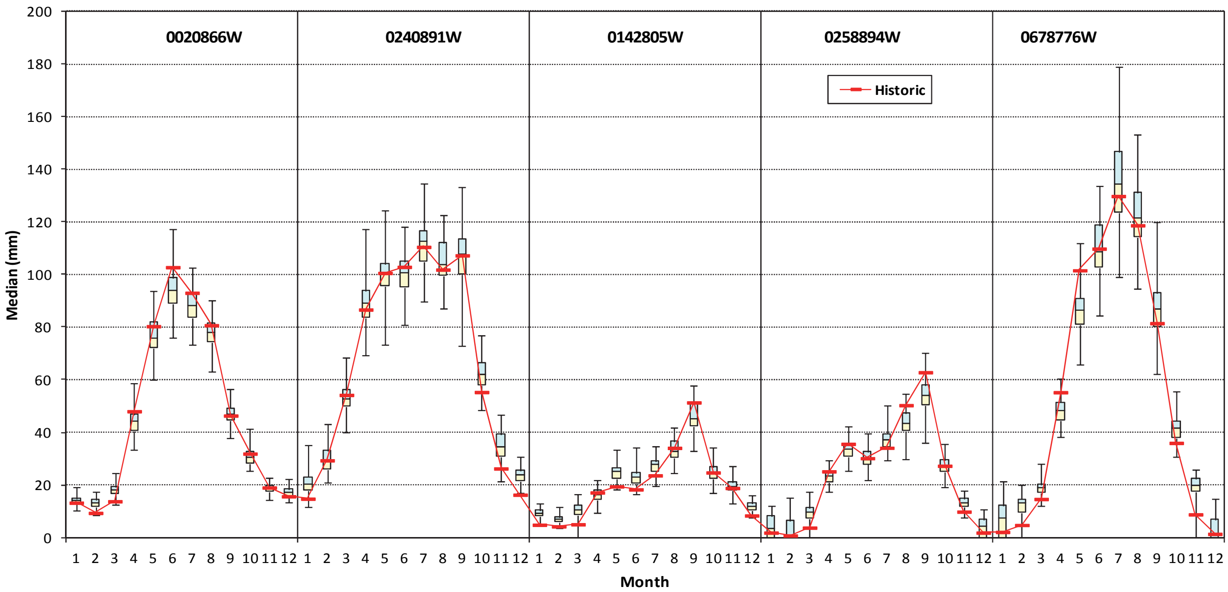


Figure 4.25b Box plots of monthly median rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

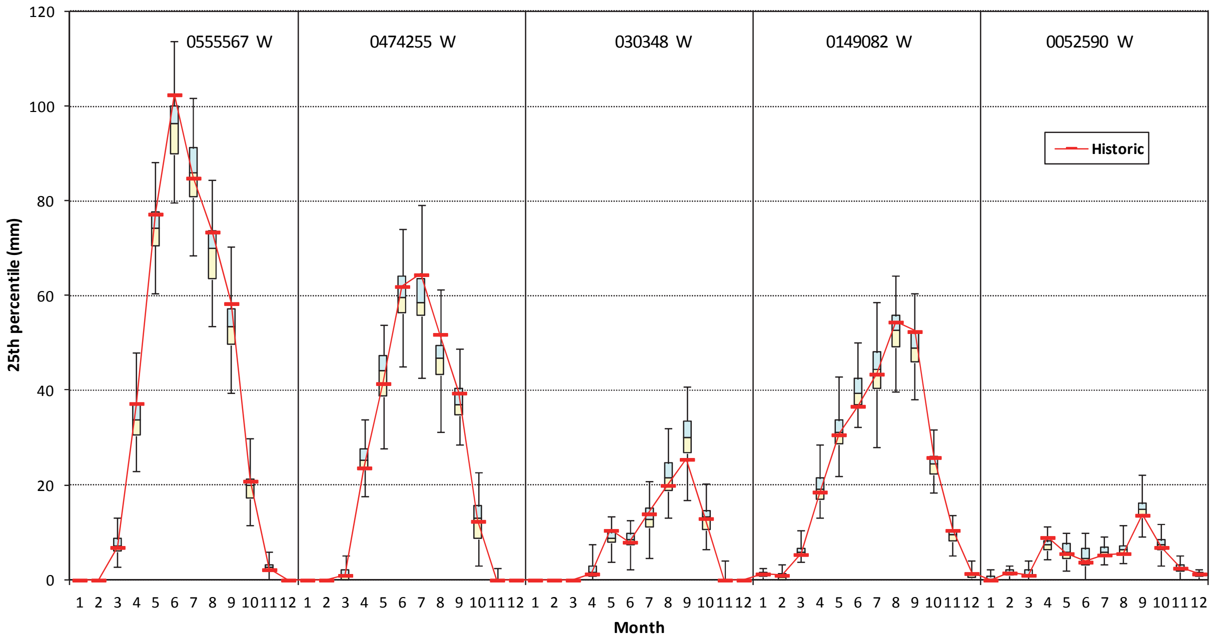


Figure 4.26a Box plots of monthly 25th percentile rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

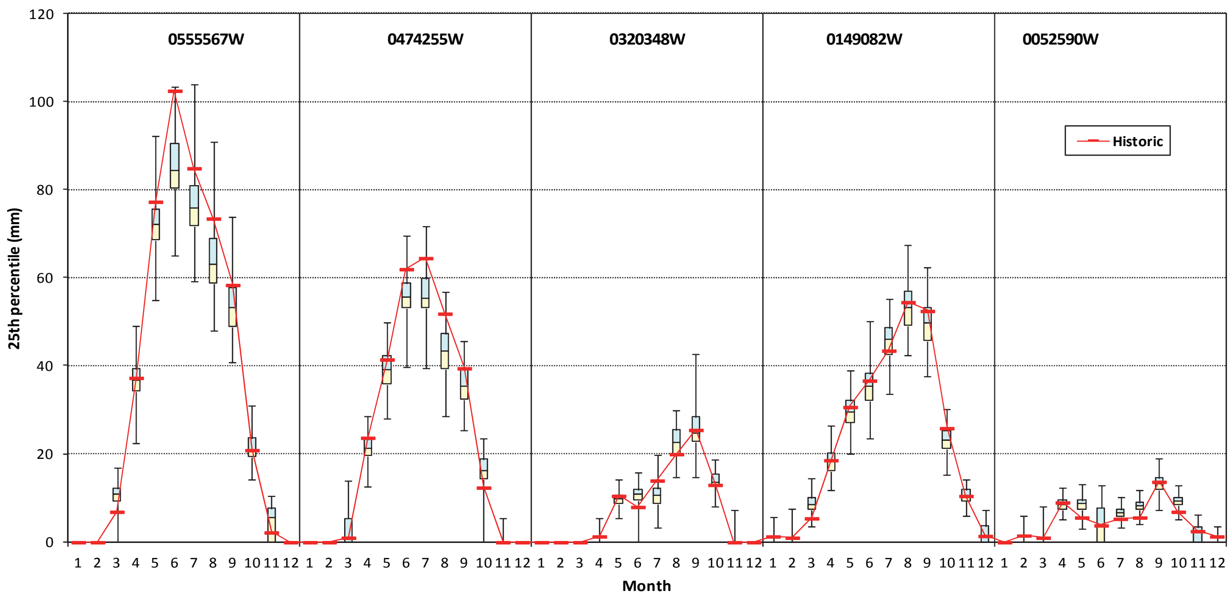


Figure 4.26b Box plots of monthly 25th percentile rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

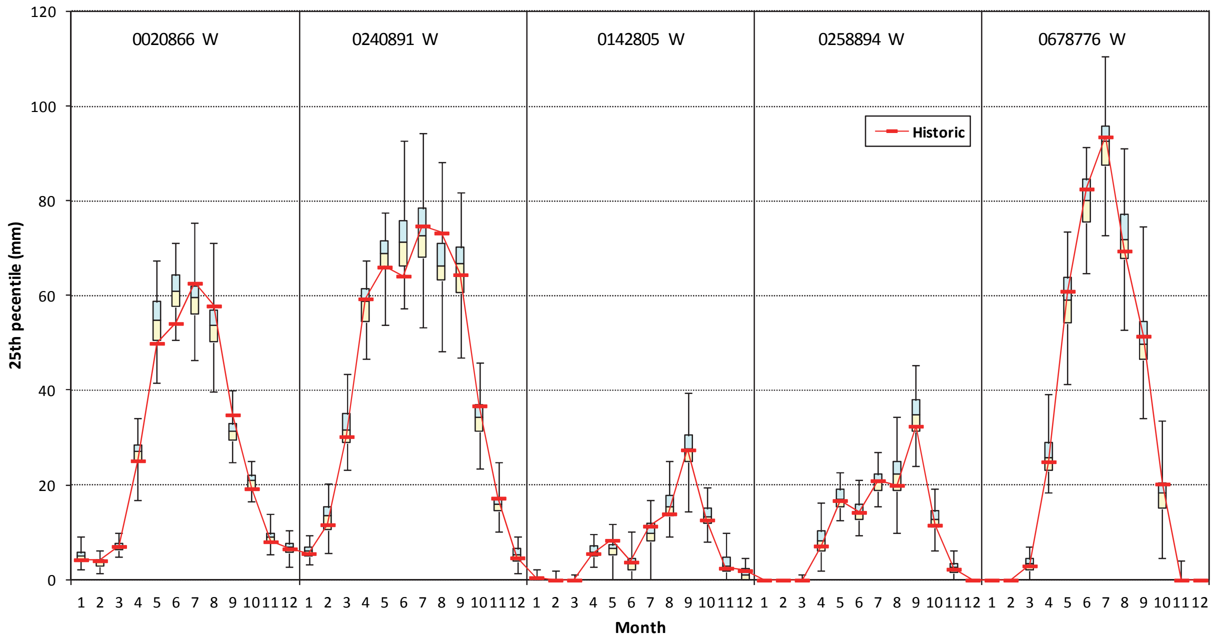


Figure 4.27a Box plots of monthly 25th percentile rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

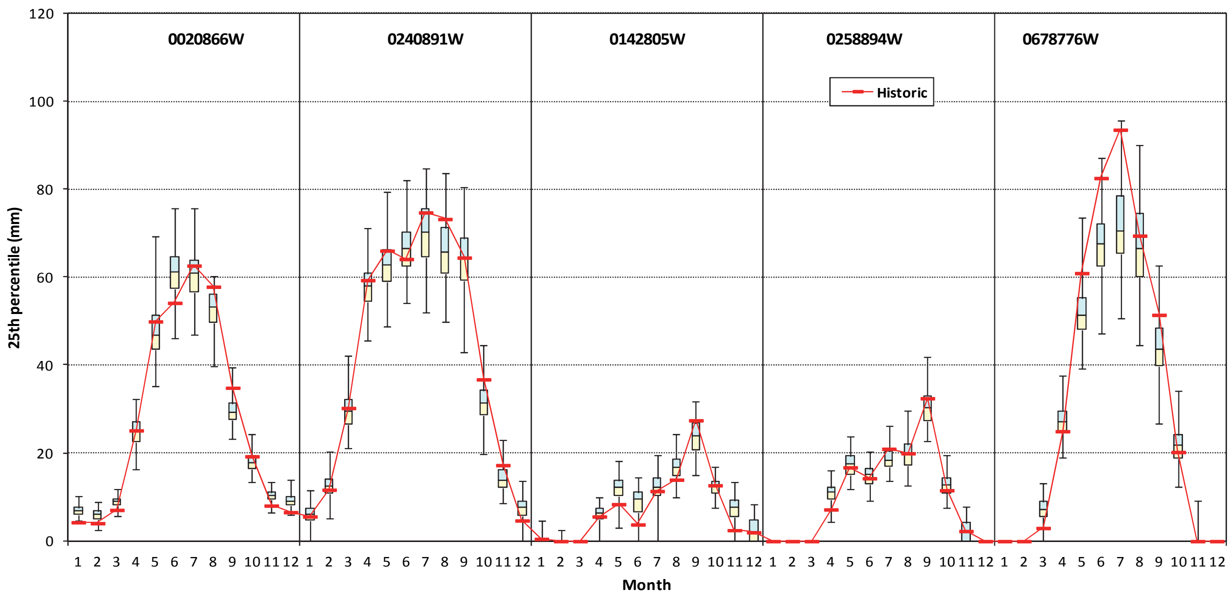


Figure 4.27b Box plots of monthly 25th percentile rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

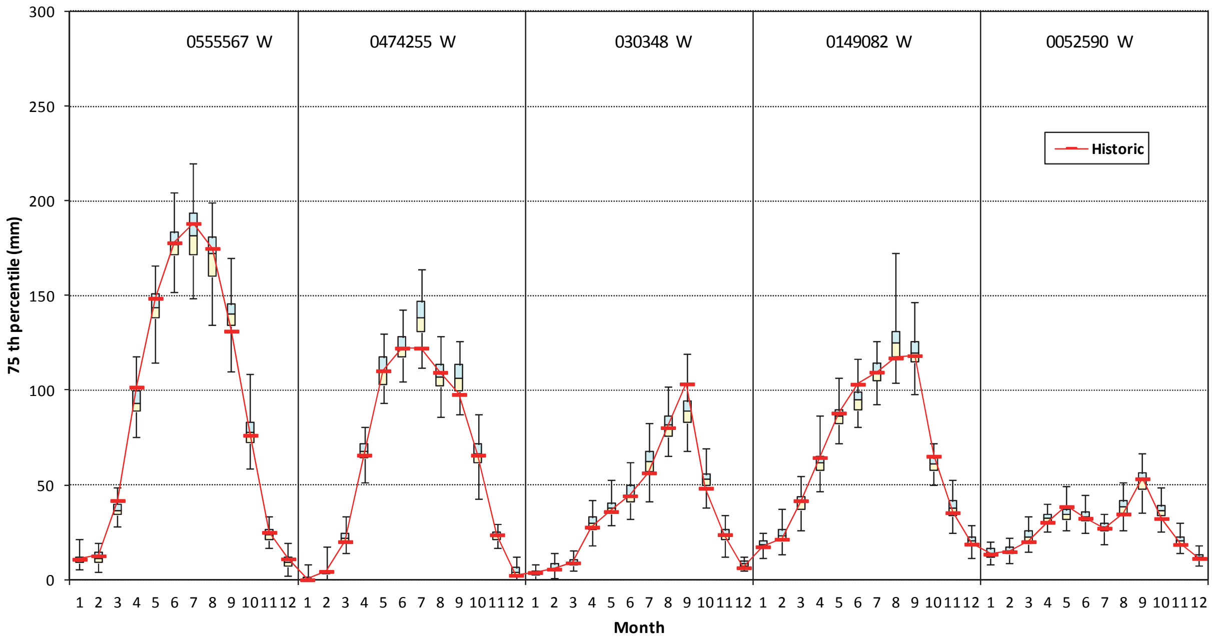


Figure 4.28a Box plots of monthly 75th percentile rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

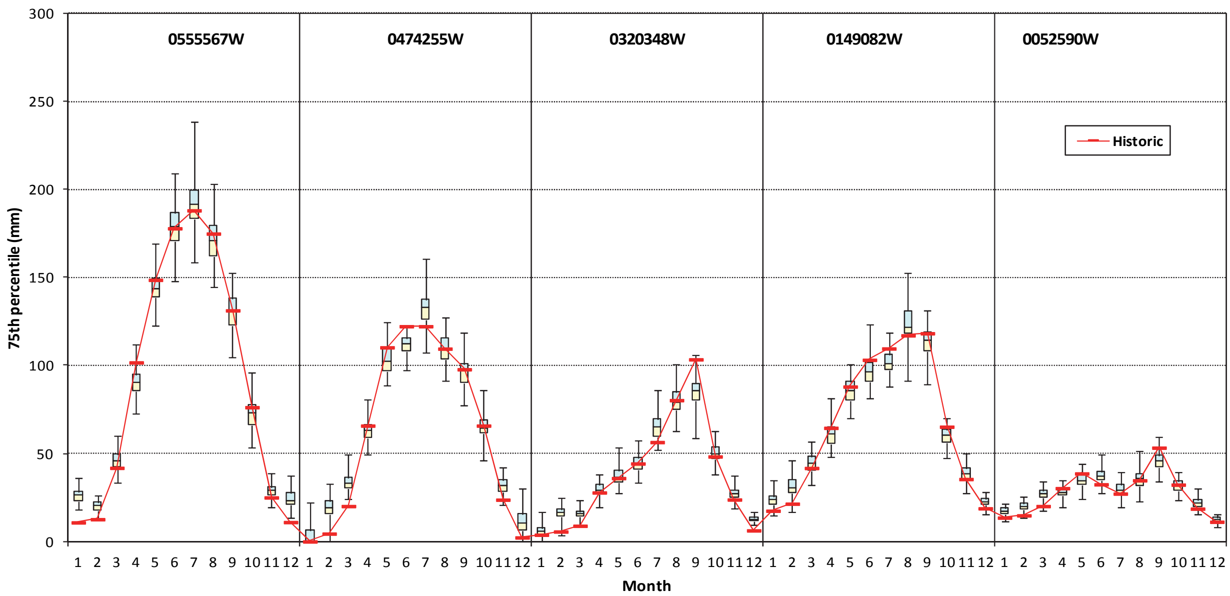


Figure 4.28b Box plots of monthly 75th percentile rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

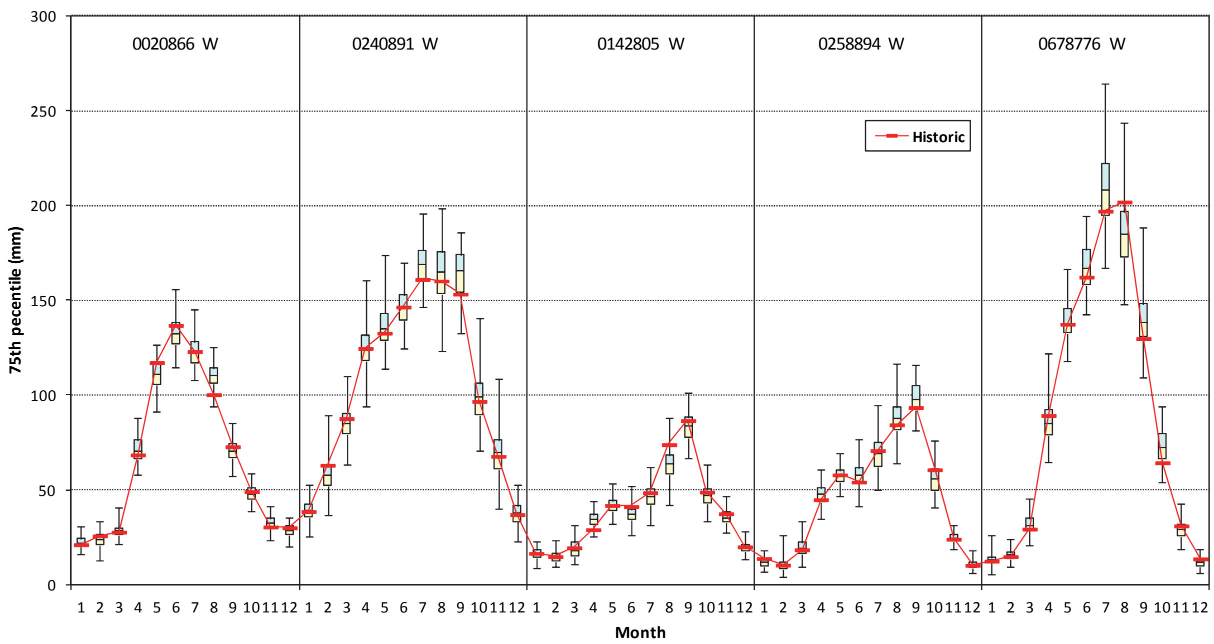


Figure 4.29a Box plots of monthly 75th percentile rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

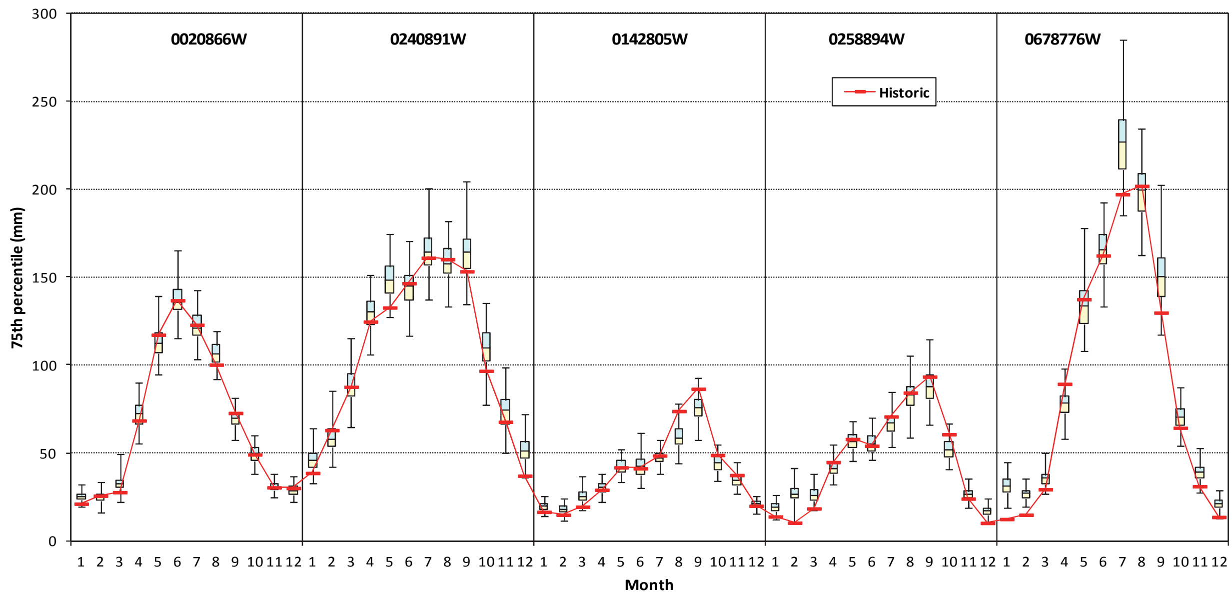


Figure 4.29b Box plots of monthly 75th percentile rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

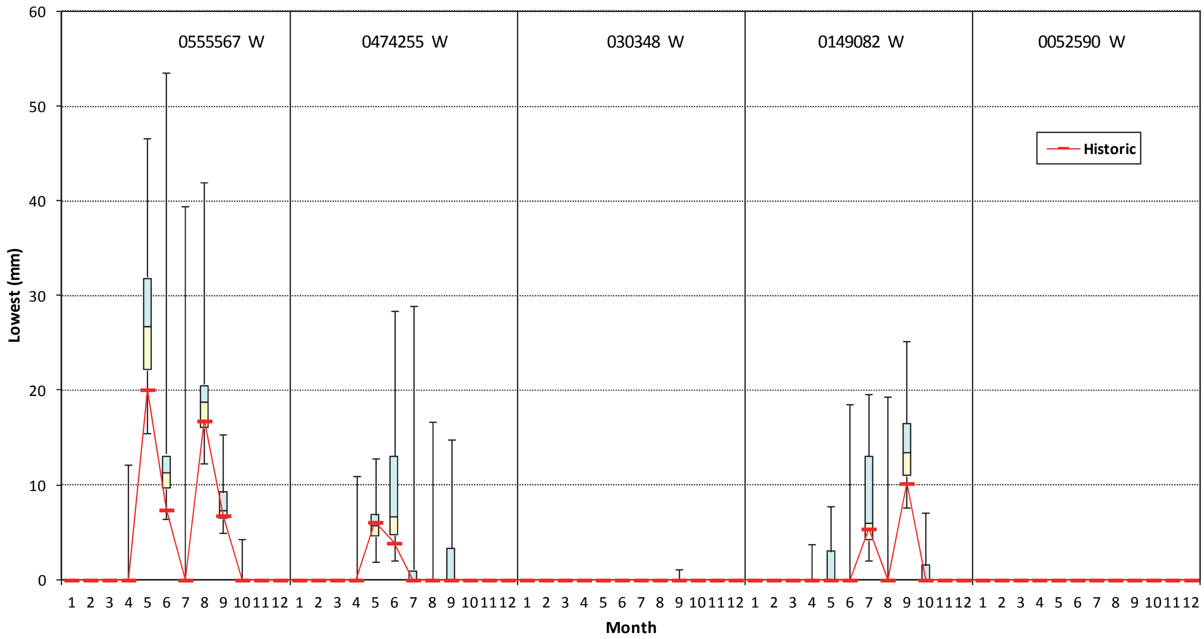


Figure 4.30a Box plots of monthly lowest rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

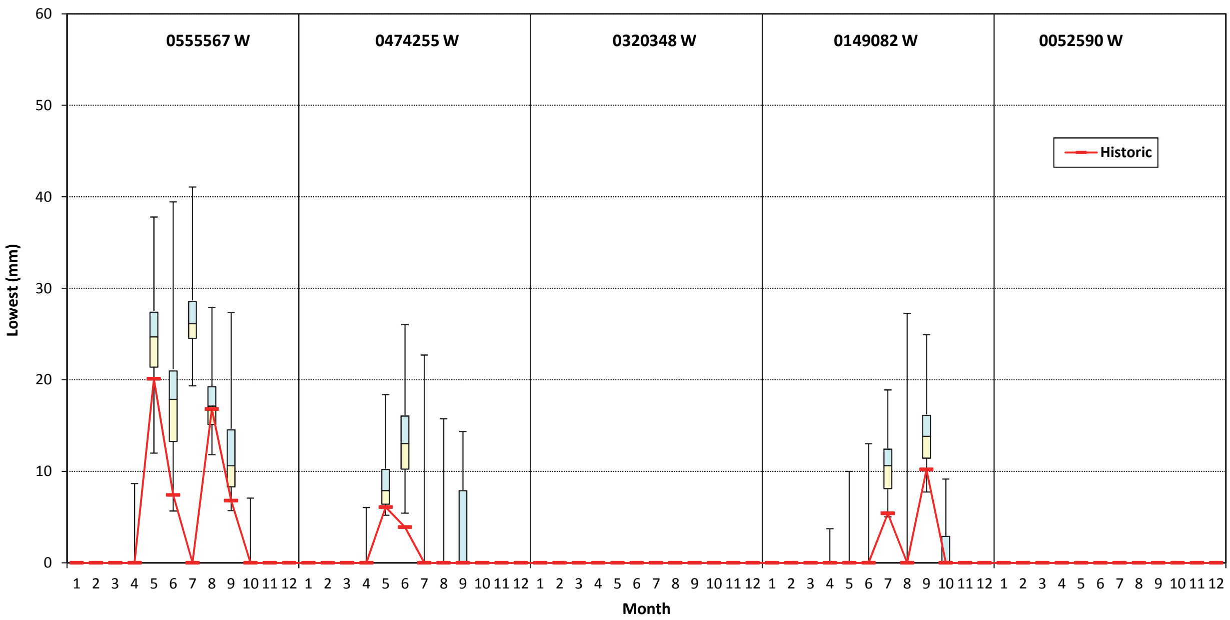


Figure 4.30b Box plots of monthly lowest rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

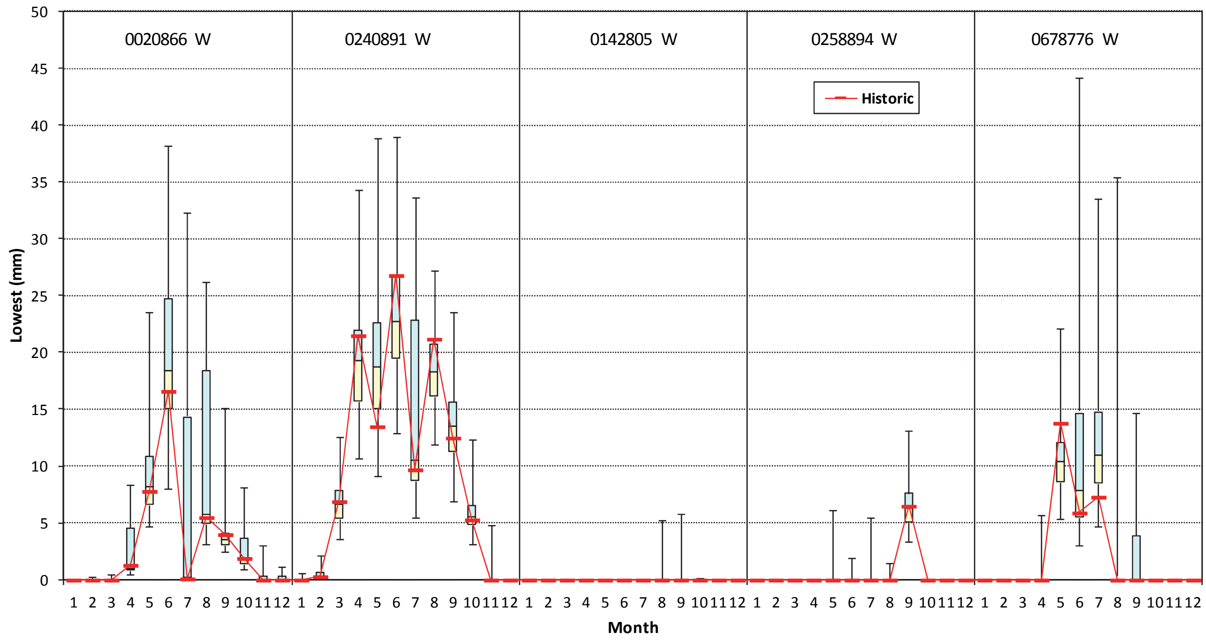


Figure 4.31a Box plots of monthly lowest rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

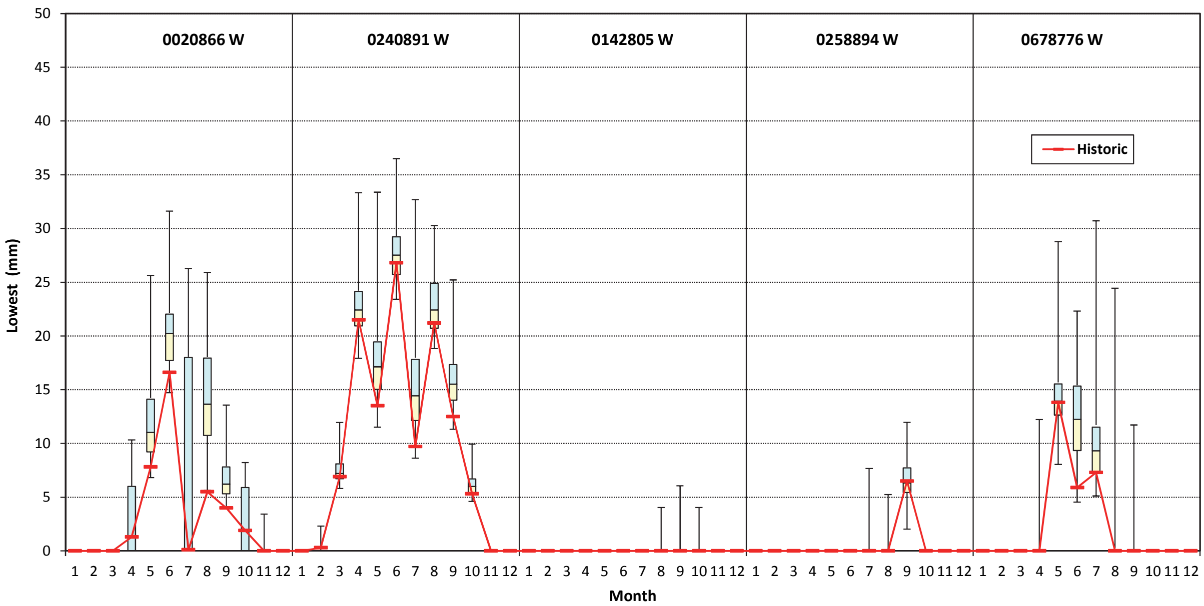


Figure 4.31b Box plots of monthly lowest rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

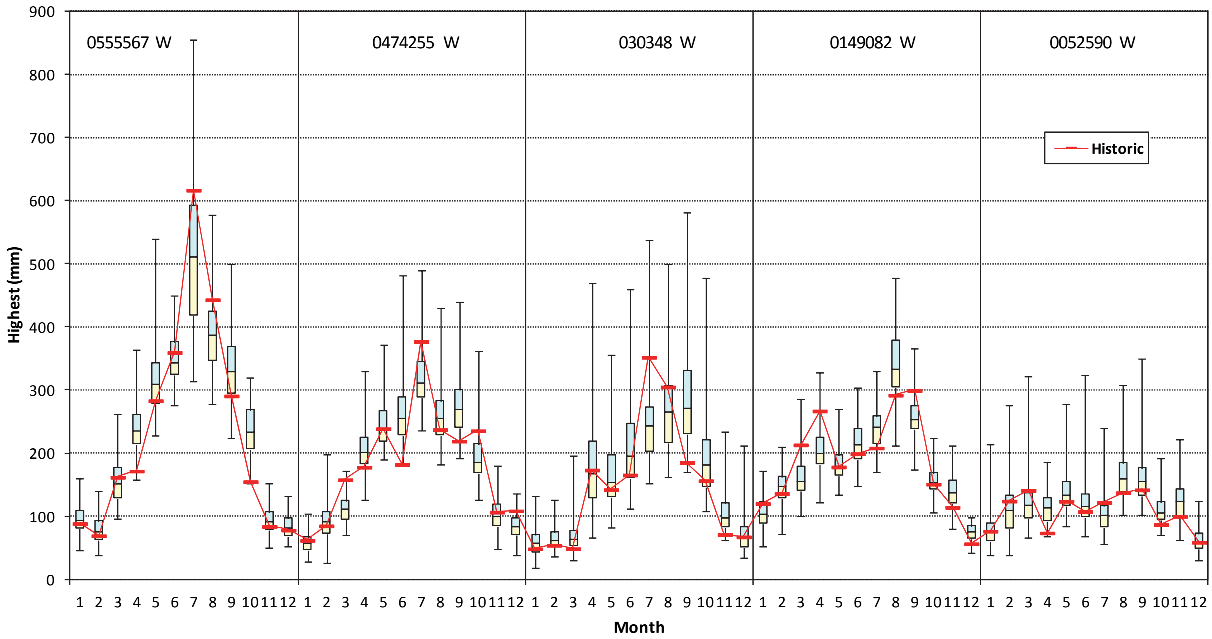


Figure 4.32a Box plots of monthly highest rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

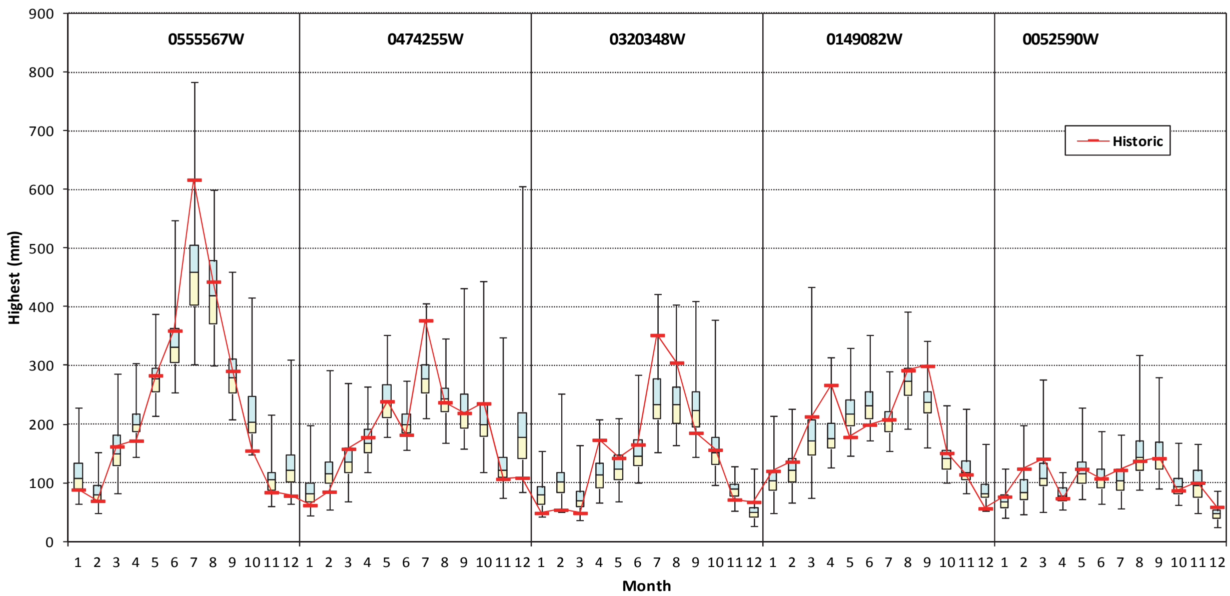


Figure 4.32b Box plots of monthly highest rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

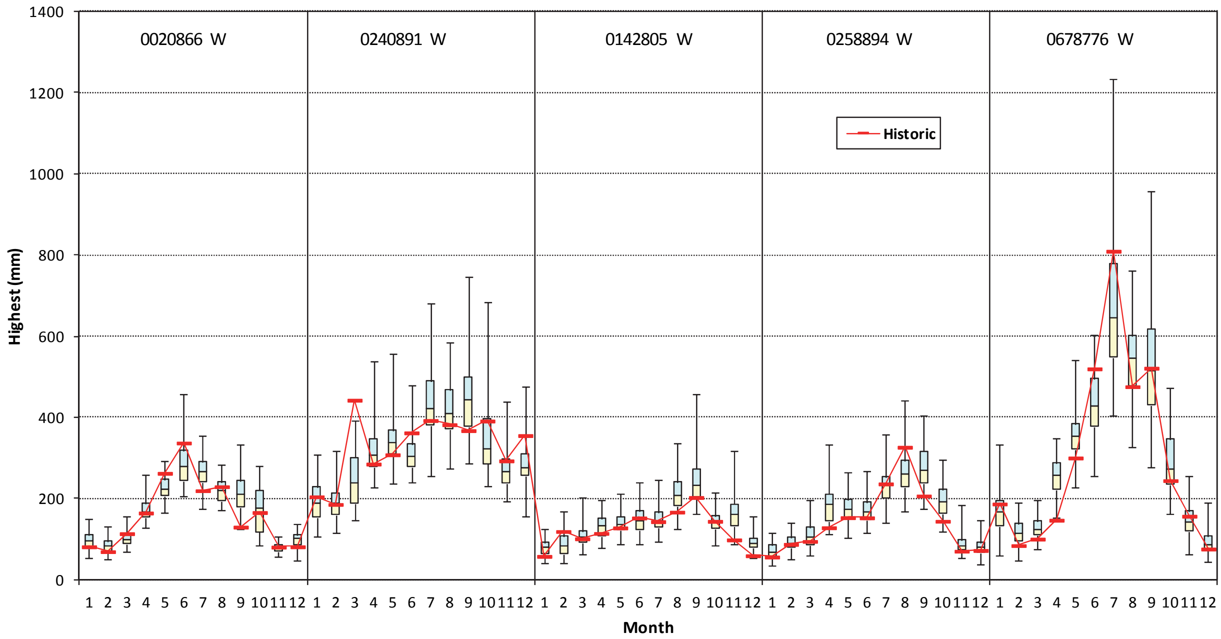


Figure 4.33a Box plots of monthly highest rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

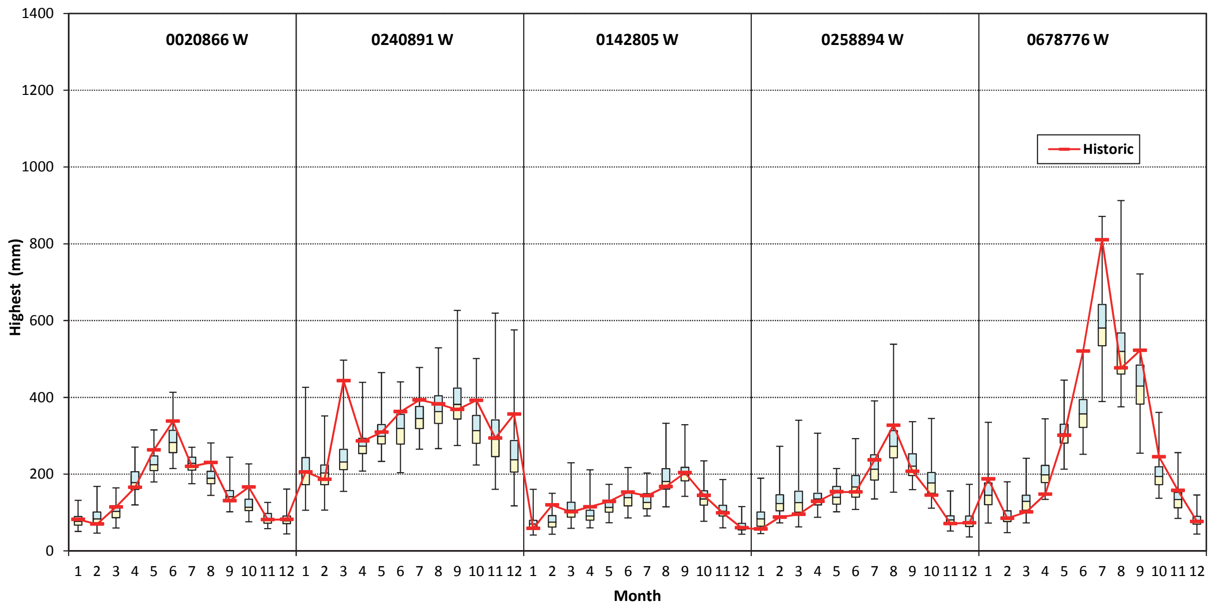


Figure 4.33b Box plots of monthly highest rainfalls for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

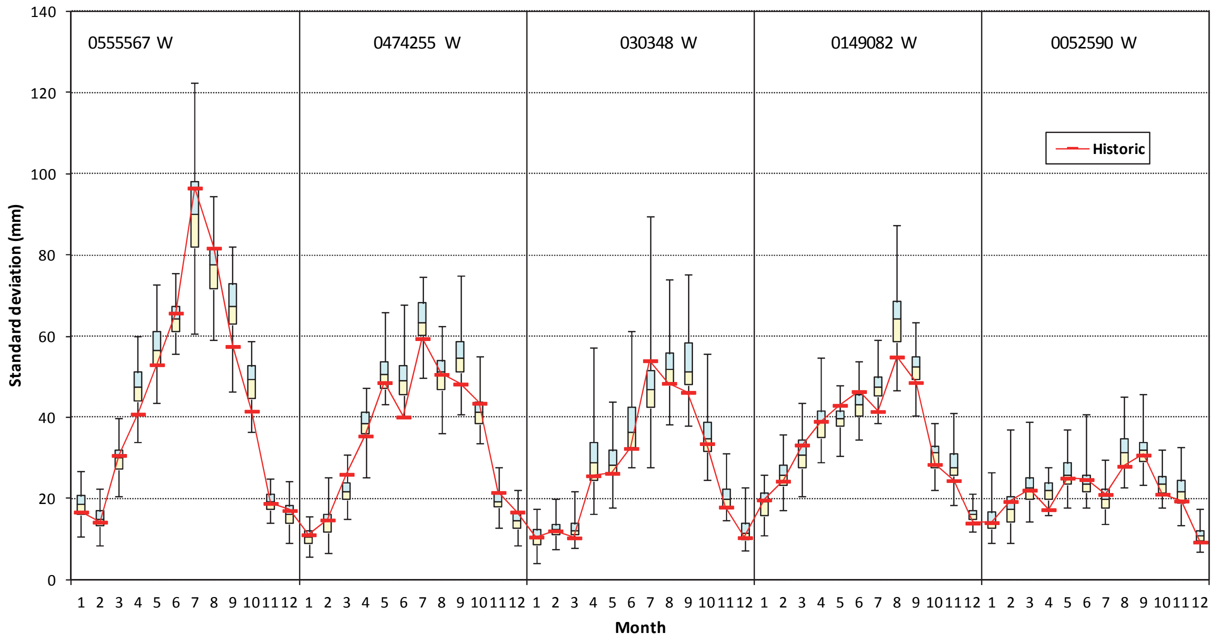


Figure 4.34a Box plots of monthly standard deviations for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

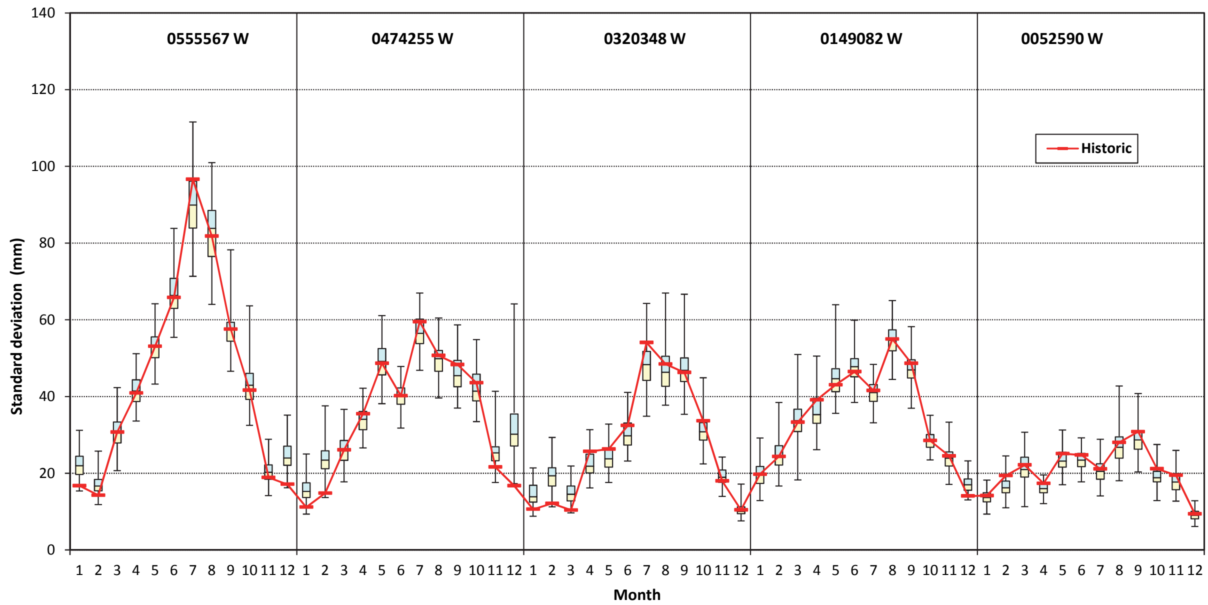


Figure 4.34b Box plots of monthly standard deviations for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

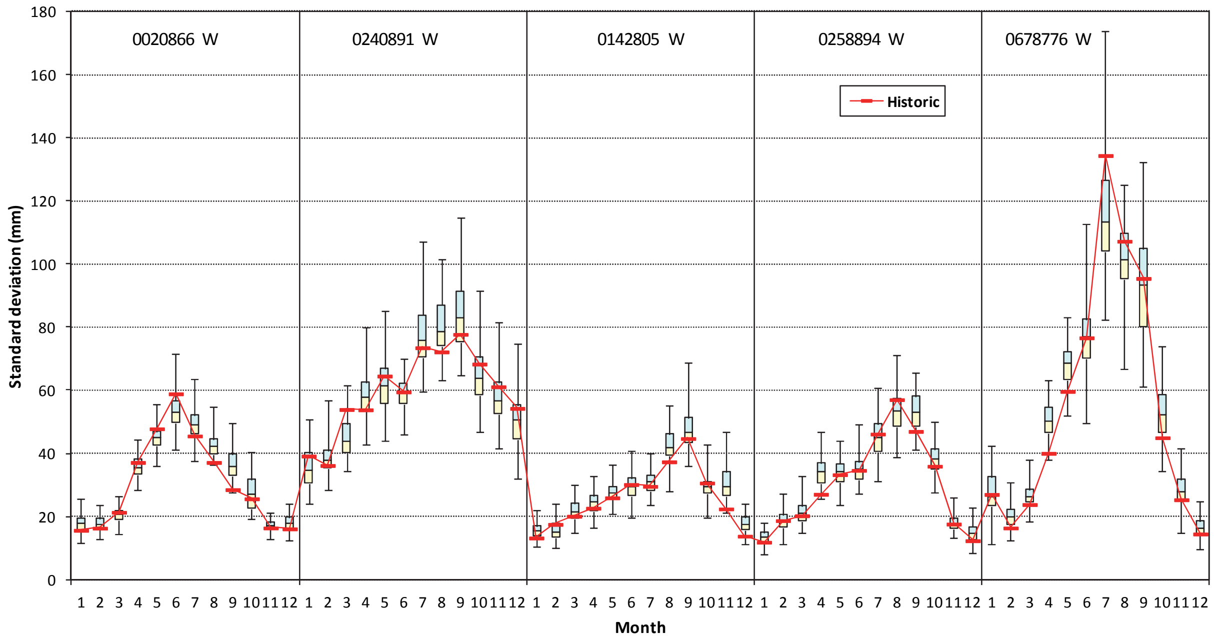


Figure 4.35a Box plots of monthly standard deviations for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

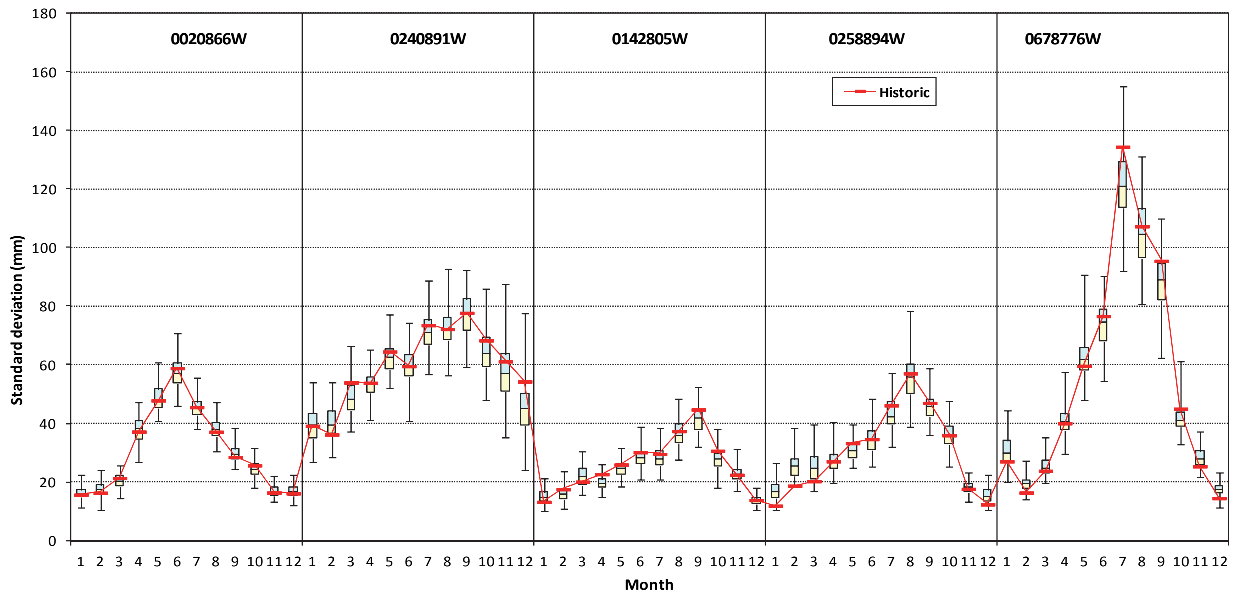


Figure 4.35b Box plots of monthly standard deviations for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

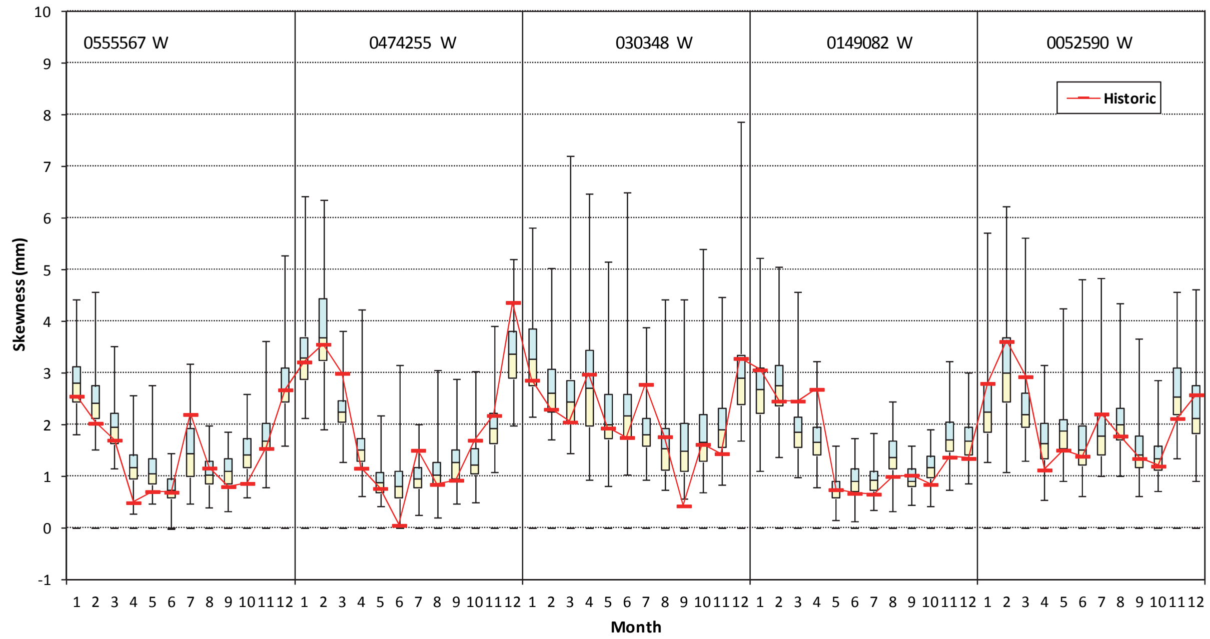


Figure 4.36a Box plots of monthly skewness of rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

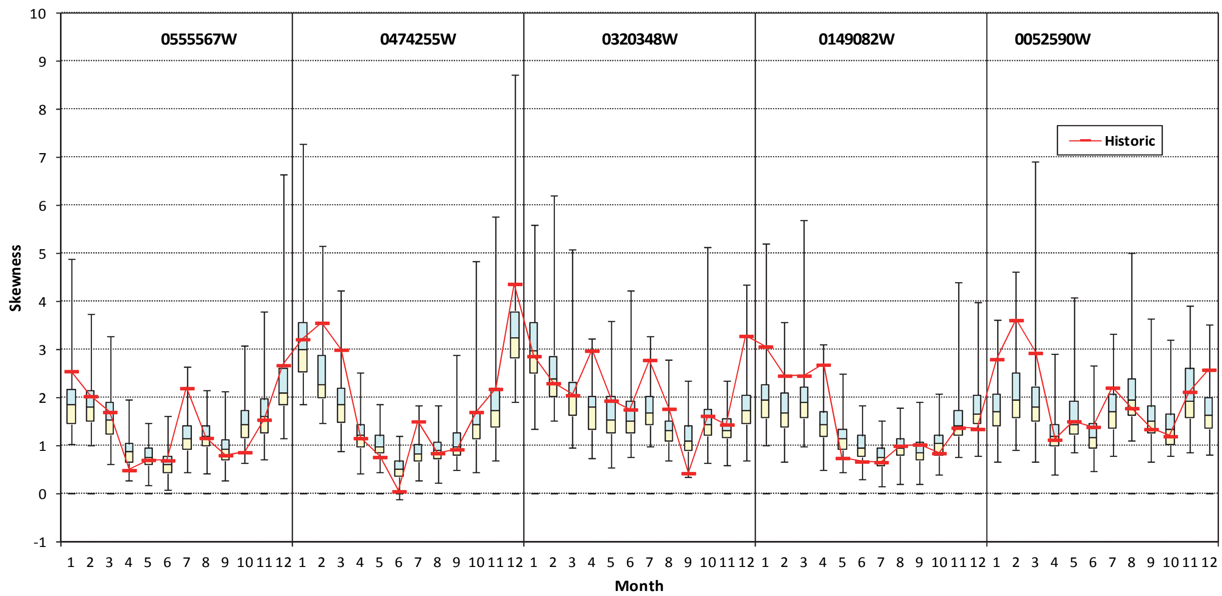


Figure 4.36b Box plots of monthly skewness of rainfalls for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

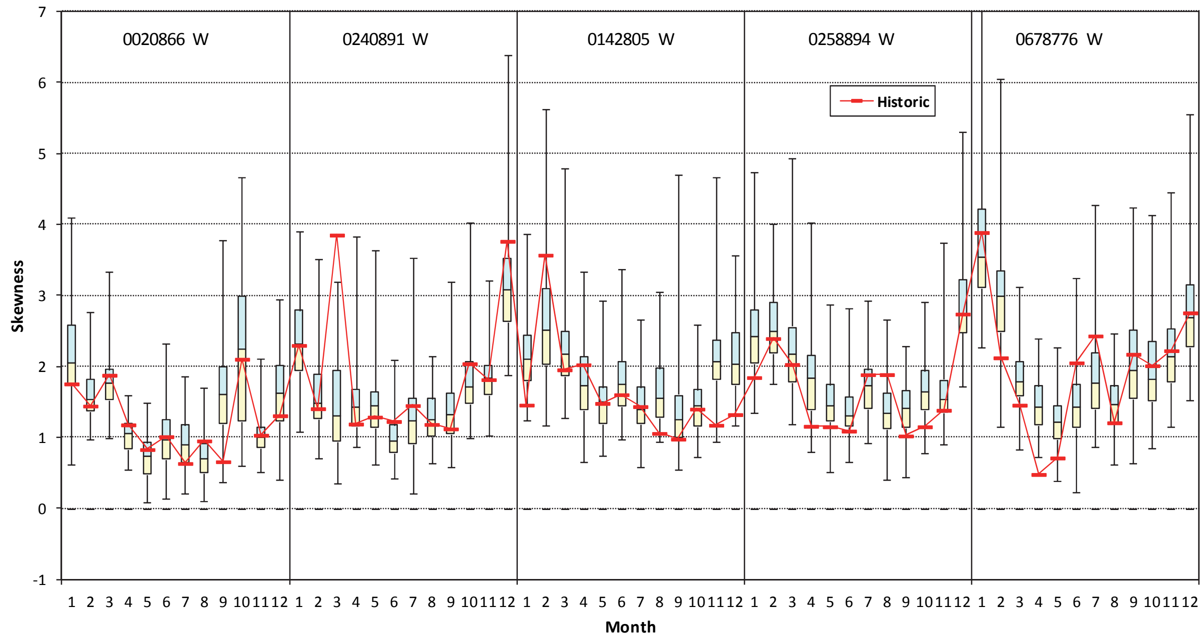


Figure 4.37a Box plots of monthly skewness for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

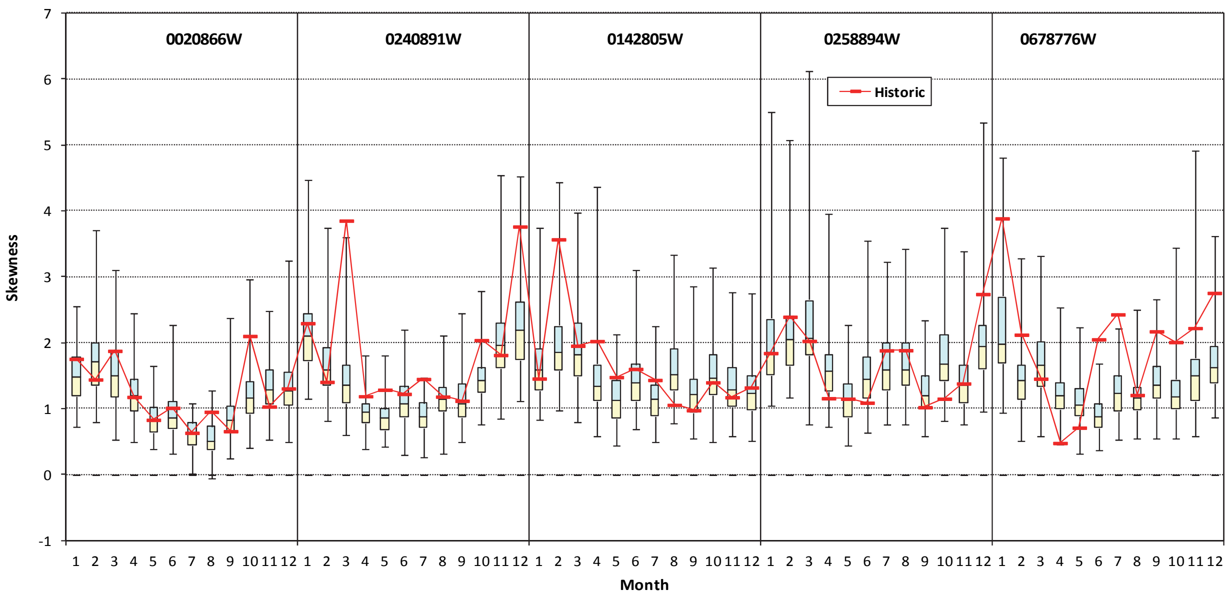


Figure 4.37b Box plots of monthly skewness for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

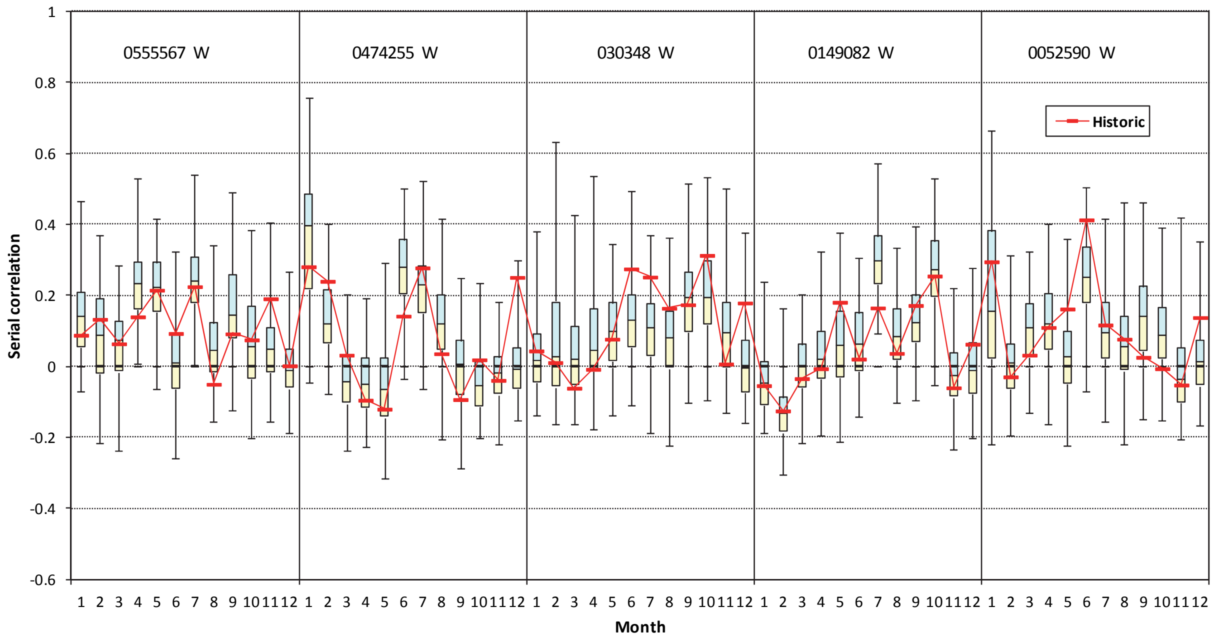


Figure 4.38a Box plots of monthly serial correlation coefficients for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from VLB generator

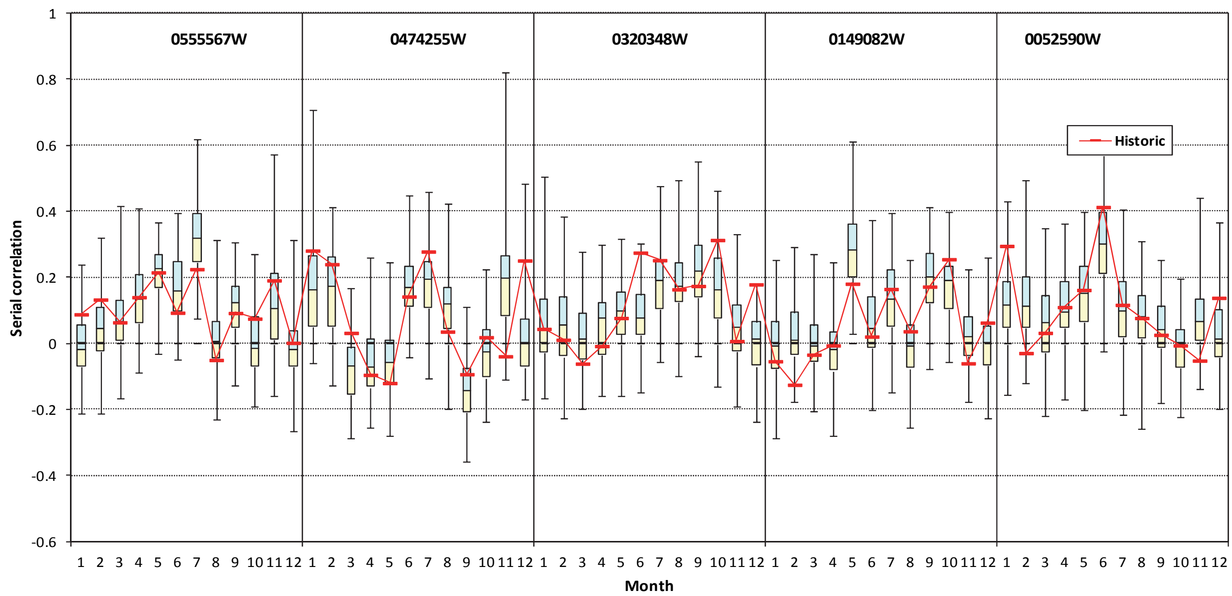


Figure 4.38b Box plots of monthly serial correlation coefficients for stations 0555567 W, 0474255 W, 0320348 W, 0149082 W and 0052590 W from PEGRAIM-W

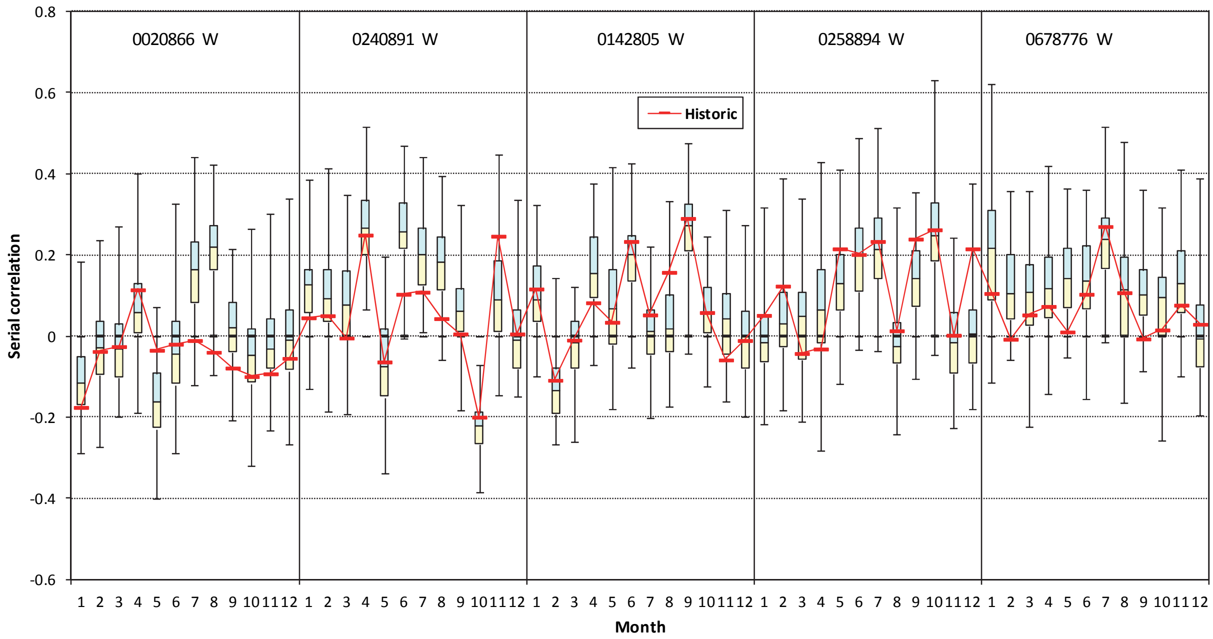


Figure 4.39a Box plots of monthly serial correlation coefficient for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from VLB generator

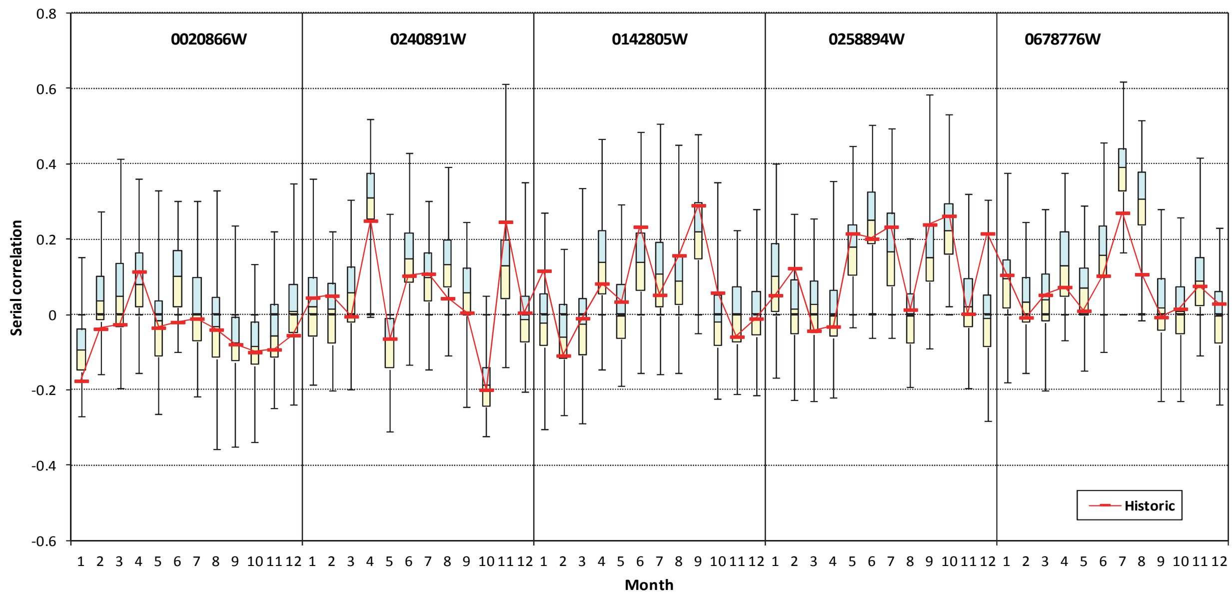


Figure 4.39b Box plots of monthly serial correlation coefficient for stations 0020866 W, 0240891 W, 0142805 W, 0258894 W and 0678776 W from PEGRAIM-W

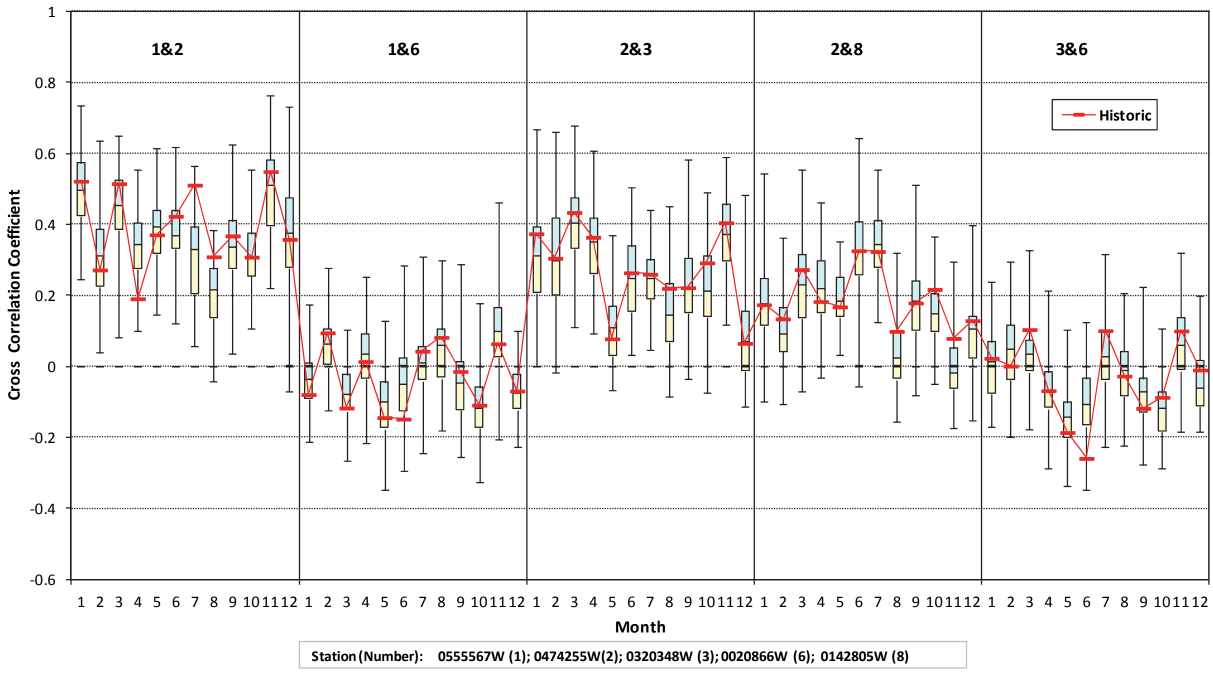


Figure 4.40a Box plots of monthly cross correlation coefficient for selected pairs of rainfall stations from VLB generator

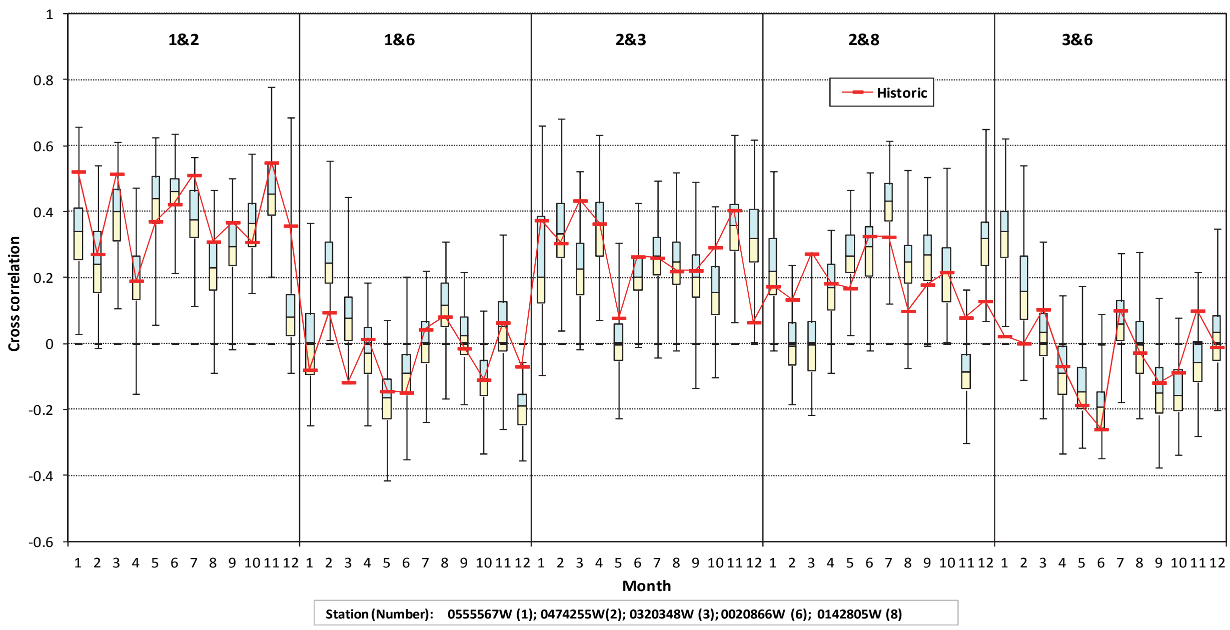


Figure 4.40b Box plots of monthly cross correlation coefficient for selected pairs of rainfall stations from PEGRAIM-W

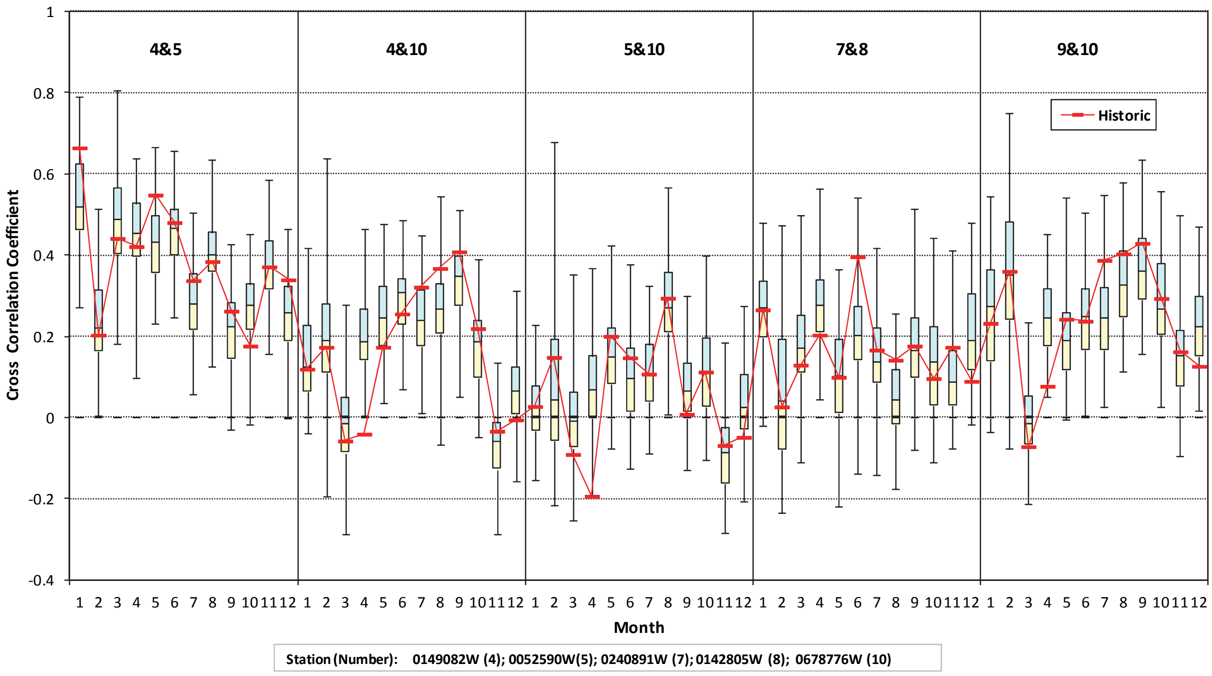


Figure 4.41a Box plots of monthly cross correlation coefficient for selected pairs of rainfall stations from VLB generator

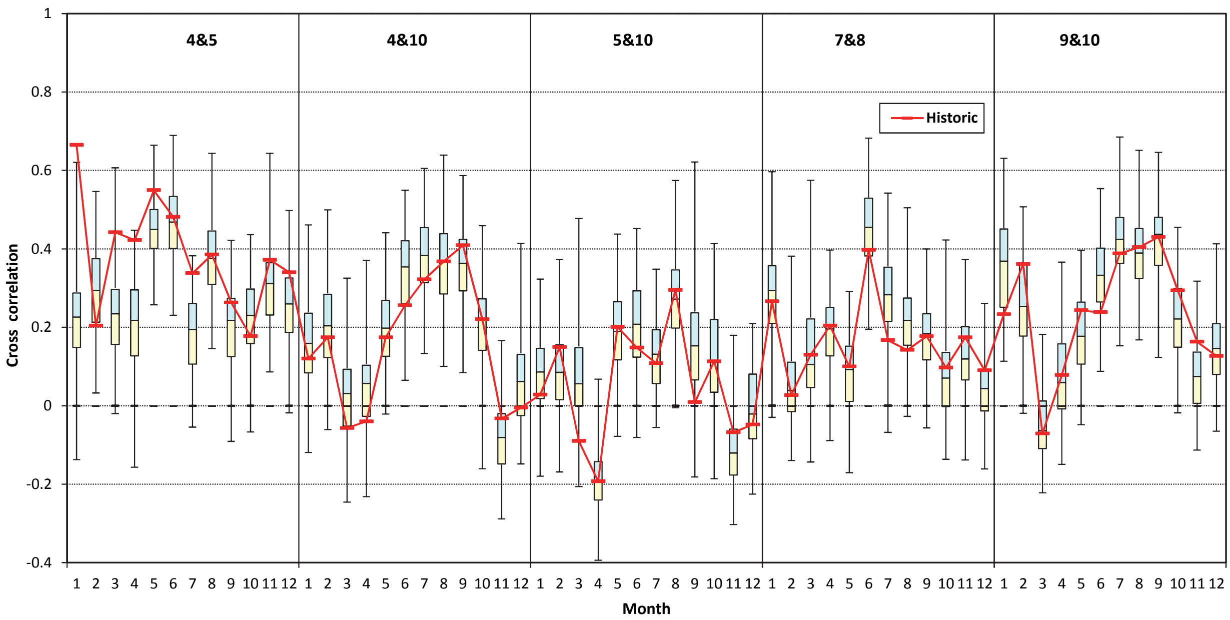
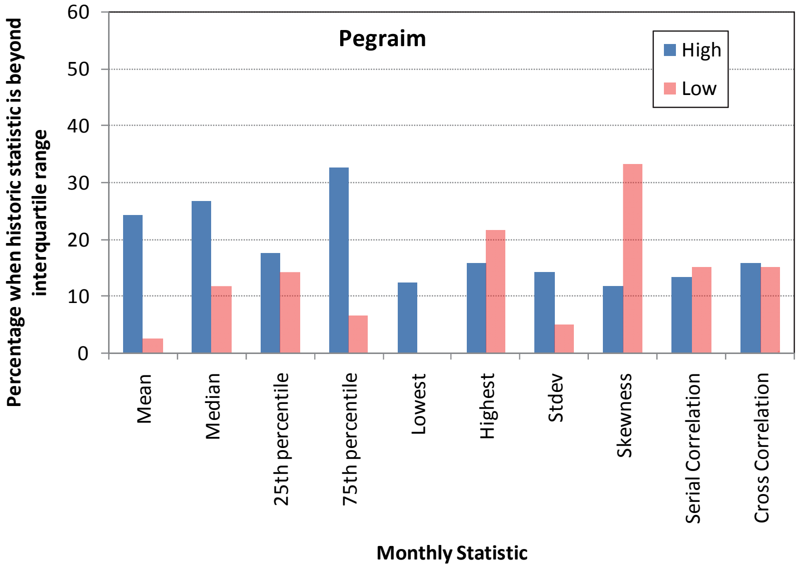


Figure 4.41b Box plots of monthly cross correlation coefficient for selected pairs of rainfall stations from PEGRAIM-W

The box plots for the monthly statistics reveal that both PEGRAIM-W and VLB replicate the historic statistics reasonably well. For PEGRAIM-W and VLB, 69.1 and 79.2 % of the historic values fall within the inter-quartile ranges of the box plots respectively. There are however some statistics where a substantial proportion of the historic values fall beyond the inter-quartile range of the box plots as seen in Figure 4.42.



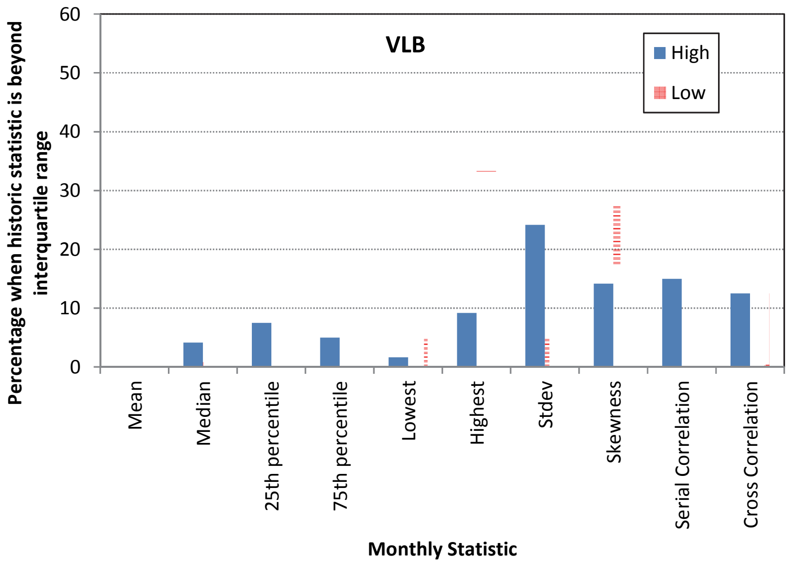


Figure 4.42 The Percentages that monthly historic statistics fall beyond the inter-quartile range of the box plots of stochastic sequences for PEGRAIM-W and VLB generator

On Figure 4.42 the blue columns denote the cases where the historic value is lower than the lower quartile of the stochastic values while a red column denotes cases where the historic value is higher than the upper quartile of the stochastic sequences.

Figure 4.42 reveals that all the historic monthly means and most of the medians fall within the inter-quartile range for the VLB while a considerable proportion of the historic means and medians fall below the lower quartile for PEGRAIM-W. Based on Figure 4.42, VLB replicates 7 of the monthly statistics better than PEGRAIM-W (Mean, Median, 25th percentile, 75th percentile, lowest, Skewness, and cross correlation) while PEGRAIM-W replicates 3 of the statistics (Highest, Standard deviation and Serial correlation) better and VLB has therefore performed better than PEGRAIM-W. The significance of the biases observed would however depend on the specific analyses for which the generation is being done.

4.4 Summary of VLB Performance and its Comparison with PEGRAIM-W

This Chapter presents the performance of the non-parametric VLB generator and its comparison with the PEGRAIM-W parametric monthly stochastic rainfall generator using a 10-site stochastic rainfall generation problem and ten statistics. For PEGRAIM-W and the VLB generator 82 and 90% annual historic statistics locate within the inter-quartile ranges of the stochastic annual statistics. PEGRAIM-W underestimates the lowest annual rainfalls by an average of 38% for 3 of the 10 stations while VLB overestimates the standard deviation for 6 of the 10 sites by an average of 11%. The VLB therefore replicates annual statistics marginally better than PEGRAIM-W. Both methods replicate the long-term minimum run sums reasonably well suggesting they adequately model any long-term dependencies of the rainfall sequences.

For the monthly simulation, PEGRAIM-W and VLB locate 69 and 79% of the historic statistics within the inter-quartile range respectively. VLB replicates 7 of the monthly statistics better while PEGRAIM-W performs better for the other three.

Based on the analysis in this Chapter, 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. The analyses here subjectively applied 101 stochastically generated sequences and this could be varied in additional analyses. Other statistical measures in addition to the 10 applied here could also be used.

5 CLIMATE CHANGE AND VARIABILITY MODELLING

5.1 Introduction

The literature review (Chapter 2) informed that quantification of climate change and variability is very uncertain and climate change and variability research is sometimes not presented in a form that is easy for practical water resources decision making. Section 4 of the literature review proposed that climate variability and change will be incorporated into the rainfall generator in a way that incorporates inter-decadal persistence in the data while modelling for a drier, a more variable or a wetter climate. An analysis of the blocks generated by the VLB indicated (as expected) that there is substantial variation of their mean annual precipitation (MAP) above and below the overall MAP of the complete historic record from which they are derived. Different scenarios of the possible states of future climate (drier, more variable, or wetter) can therefore be obtained by appropriately biasing the sampling of blocks based on their MAP. If no bias is implemented (as in Chapters 3 and 4), the generated rainfalls are then for the normal climate. Climate change projections suggest that different changes will happen for different regions of South Africa and the rainfall generation problem used in Chapters 3 and 4 with 10 rainfall stations widely spread throughout South Africa may not be appropriate for climate change and variability modelling. The Reference Group meeting for this project held in August 13 2013 also proposed that the generator be tested on a more closely spaced rainfall network. A data set of 10 rainfall stations located in the Western Cape of South Africa (Figure 5.1) is therefore used in this Chapter. This data was obtained from Lynch (2003) and consisted of 99 years of concurrent monthly rainfalls. Sixty percent (60%) of this data was patched but this was considered realistic as practical water resources analysis in South Africa applies patching extensively.

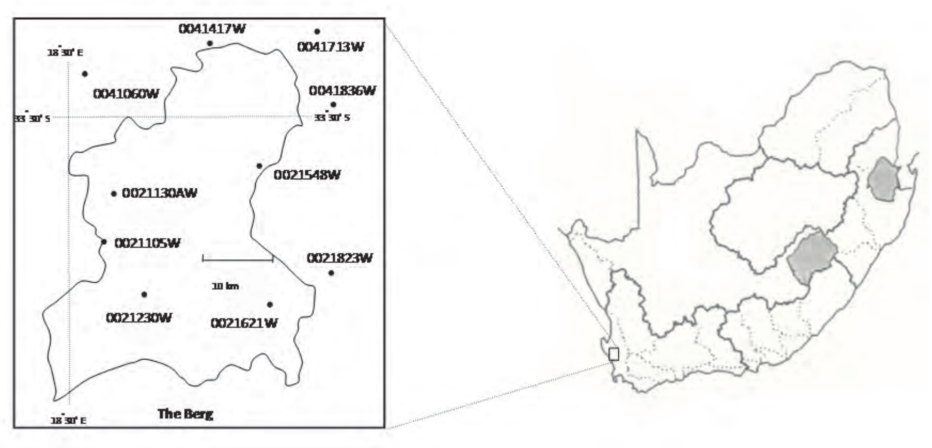


Figure 5.1 Rainfall stations used to assess the climate variability VLB stochastic rainfall generator

As stated in the summary of the literature review (Section 3.4), the climate variability modelling needed to allow for the projected shifts in seasonality of rainfall obtained from multiple GCMs. Since the use of GCM projections is widespread in spite of their poor validation performance (Section 3.4), it is intended that this Chapter demonstrates the generation of stochastic rainfalls that match the overall change in MAP as projected by GCMs. The generated rainfalls are then likely to possess reasonable levels of uncertainty because they are derived from a realistic stochastic generator while meeting the average shifts projected by GCMs. From the University of Cape Town Climate Systems Analysis Group (CSAG) website (http://cip.csag.uct.ac.za/webclient2/app/#datasets), the Cape Town airport rainfall station was selected for obtaining plausible monthly shifts in monthly rainfall for the study catchment. It is possible that other rainfall stations (or average changes in rainfall from raw GCM data for the grid in which the study area falls) could be more representative but the main objective here is to demonstrate the method. The changes to average monthly rainfalls were read off from plots of projected future changes across 10 different statistically downscaled CMIP5 GCMs for pathways RCP 4.5 and RCP 8.5. These plots were downloaded from the CSAG website. The changes were calculated relative the historical period 1980-2000. Figure 5.2 shows the projected changes from the 10 GCMs with a general agreement amongst most of them of a reduction in rainfalls in the winter months (May to August). A plot of the average monthly rainfalls based on Figure 5.2 is presented on Figure 5.3. A comparison of this with the average monthly rainfalls for the selected historical period clearly shows the projected reduction in the winter rainfall. This results in an overall reduction of 6.5% in the MAP.

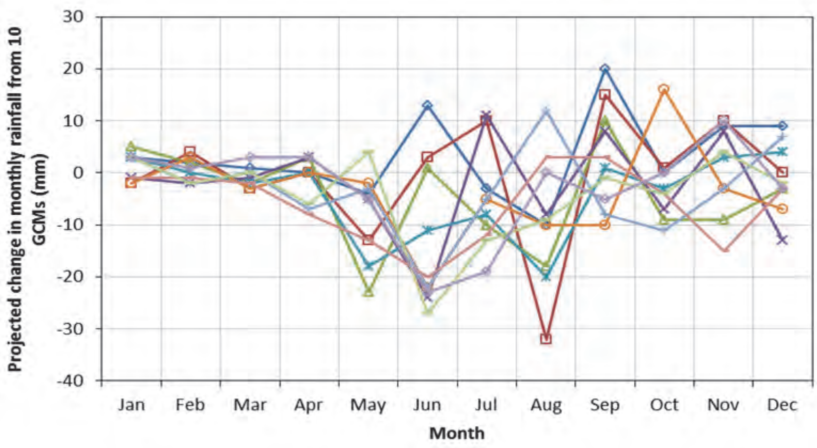


Figure 5. 2 Projected changes to average monthly rainfall from 10 GCMs (adapted from http://cip.csag.uct.ac.za/webclient2/app/#datasets), 1980-2000 at Cape Town International airport

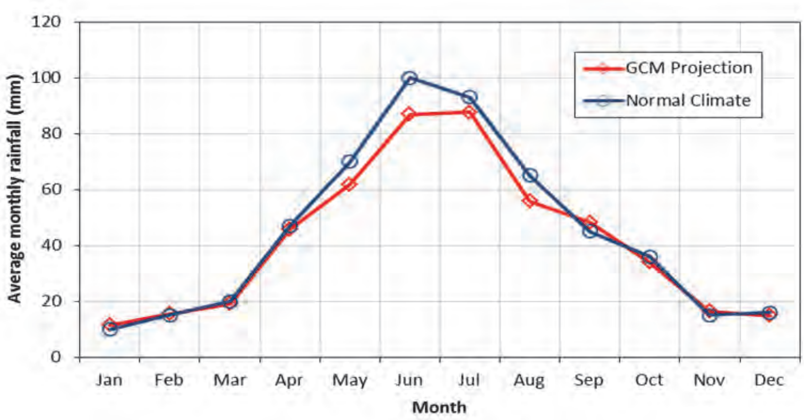


Figure 5.3 GCM Projection and average monthly rainfalls for 1980-2000 at Cape Town International airport

5.2 Modelling Annual and Longer-term Climate Change by Biasing Block Selection

The generation of stochastic rainfall sequences for a climate other than the normal one as reflected in the historic time series is implemented by biasing the selection of blocks in a manner that reflects the required climate change scenario. For example, if the user needs to analyse a drier climate scenario, block selection is carried out to favour the selection of drier rather than wetter blocks in generating the stochastic rainfalls. This variation to the random block selection of the standard VLB generator (Section 3.3.1 of Chapter 3) enables the VLB to generate a wide range of possible climate change scenarios because it is easy to obtain large variations in block dryness or wetness. To demonstrate this, the block lengths and their respective scaled MAP (MAP of block/MAP of complete historic sequence) obtained in the generation of 101 ninety nine (99) years long rainfall sequences of the 10 rainfall station problem were analysed. For this, a minimum block length of 3 years a range of proportion of exceedance for defining a dry year of 60 to 90% (as applied in Chapter 3) was adopted. Block lengths ranging from 3 to 30 years were obtained in different proportions and numbers as seen in Figures 5.4 and 5.5. It was observed that shorter blocks were generally generated in higher percentages than longer ones. Each selected block takes a length equal to its length and weighting the percentages in direct proportion to the block length obtains a better estimate of block significance. The weighted relative frequency on Figure 5.4 and the product of block length with number of times that blocks are selected on Figure 5.5 reveals that some long blocks (some exceeding 20 years) are as significant as the shorter ones. Figure 5.4 reveals (as expected) that the scaled MAP of short blocks will be lower than unity (1.0) since many of these will be located entirely in dry years while longer blocks will generally have scaled MAPs close to but slightly higher than 1.0 (since the shorter blocks tend to have lower MAPs).

An assessment of the variability of the scaled MAP of the blocks using cumulative probability density plots on Figures 5.6a and 5.6b reveal a large variability that reduces as the block length increases. This variability is applied to obtain stochastic rainfalls for drier, wetter or more variable climate by biasing block selection appropriately. The following steps are followed to implement the bias for obtaining stochastic rainfalls for a drier climate.

i) Set an initial upper limit of the highest scaled MAP (hereafter denoted as parameter $\boldsymbol{U_L}$) of the block that can be selected for generating stochastic rainfalls. This effectively reduces the search domain for blocks to those with scaled MAPs lower than or equal to the set limit ($\boldsymbol{U_L}$).

ii) Search for a block within the reduced search domain to a limiting number of times defined as the number of blocks available ($\boldsymbol{n_b}$) which is known at this stage (see Sections 3.3.1 and 3.3.2) multiplied by another parameter called level of search $\boldsymbol{L_s}$. The upper limit of $\boldsymbol{L_s}$ is 1.0 (which ensures that all the blocks are searched) and the lower limit is any value that allows for at least a single search.

iii) If no block is found in step ii), increase the upper limit to a value higher than $\boldsymbol{U_L}$ by a relaxation step $\boldsymbol{R_S}$ (which is the third parameter of the method) and go to step i) using this updated $\boldsymbol{U_L}$ instead of the initial one. Keep repeating this until a block is selected.

The generation of stochastic rainfalls for a wetter climate follows the same steps with the following variations: i) In step i), an initial lower limit ($\boldsymbol{L_L}$) is set and only blocks with a higher scaled MAP are selected. ii) In step iii), the relaxation step R_S is a reduction and not an increase in the lower limit L_L. Parameter R_s specifies how refined the search for blocks needs to be and a value of 0.01 or 0.02 was found reasonable. Figures 5.7 and 5.8 are flow charts describing the generation of stochastic rainfalls for a drier and a wetter climate respectively.

A more variable climate is considered to be one that has a higher proportion of more extremes on the two tails of the distribution of the MAP of the generated rainfalls since rainfall MAP is known to cluster around the middle of the distribution. A more variable rainfall could therefore be obtained by combining the generation of rainfalls for a dry climate and for a wet climate. Depending on the proportions of wet and dry climate rainfalls and the parameters (U_L, L_L, L_S and R_S) applied, the more variable climate could also be dry, normal or wet. A parameter termed as the variability bias (V_B) that varies between -1 (totally biased to dry climate) and 1 (totally biased to a wet climate) is therefore used when generating rainfall for a more variable climate.

Trial runs of the block selection showed that the approach adopted was effective and efficient. Figure 5.9 shows the probability density plots of the scaled MAPs of blocks selected in generating 101 ninety nine (99) years long stochastic rainfalls for a normal climate and three climate change scenarios.

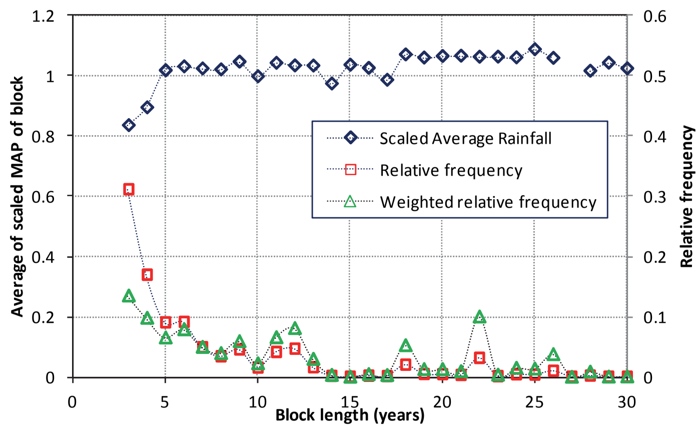


Figure 5.4 Scaled average rainfall, relative frequency of blocks and relative frequency weighted by block length

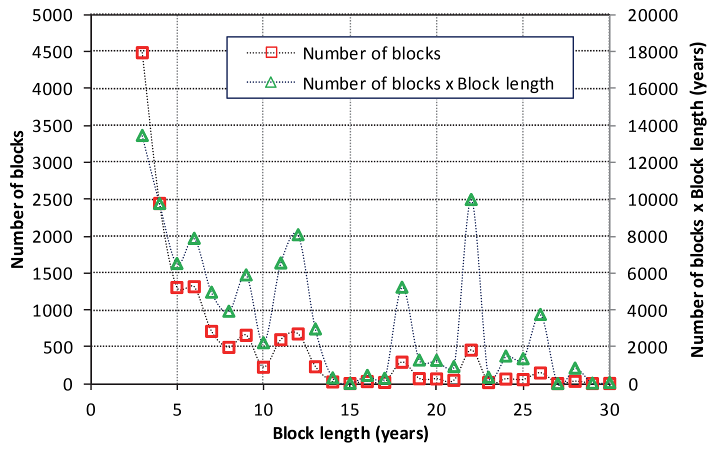


Figure 5.5 Number of blocks generated and the total length of blocks of specified length

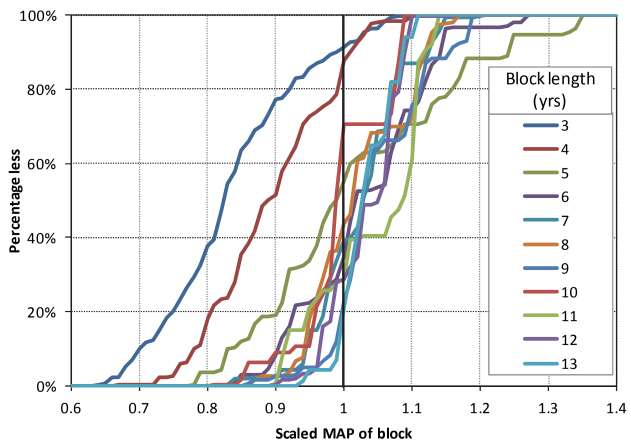


Figure 5.6a Cumulative density plots of scaled MAP for block lengths of 3 to 13 years

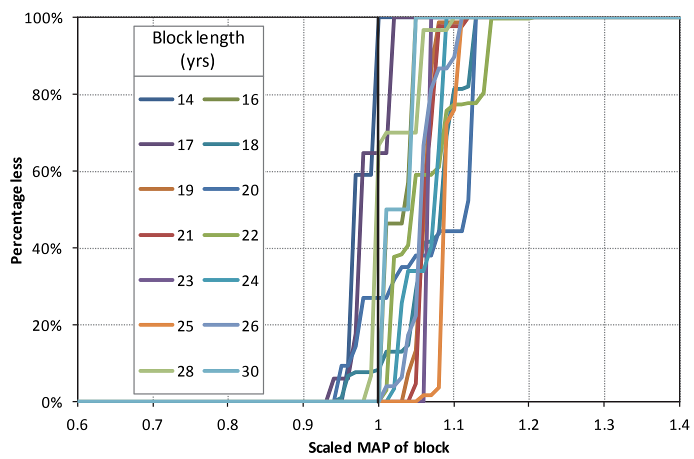


Figure 5.6b Cumulative density plots for scaled MAP for block lengths of 14 to 30 years

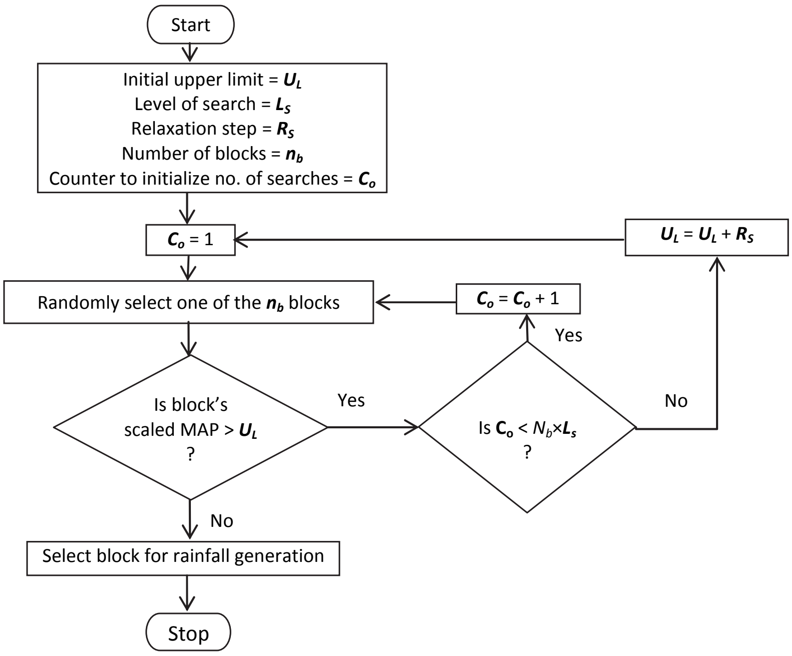


Figure 5.7 Block selection for a drier climate

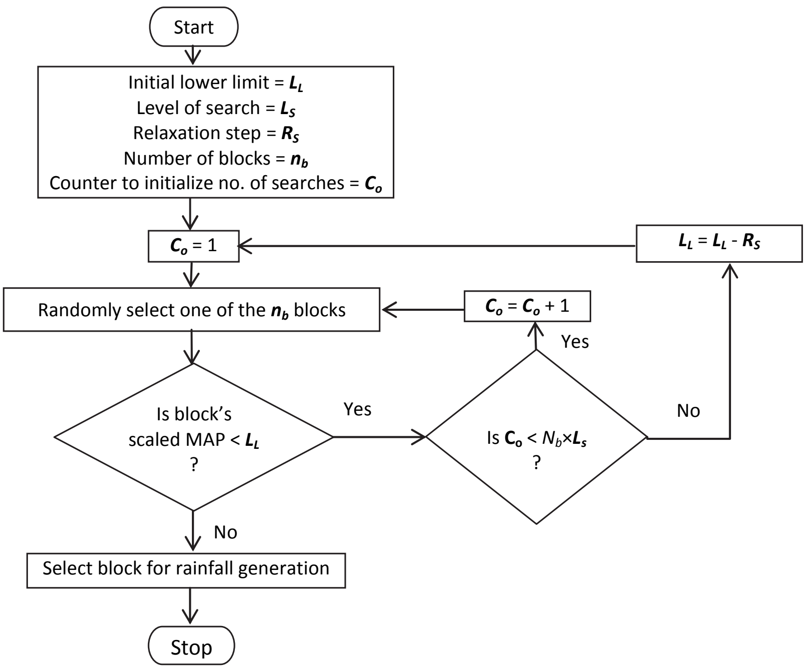


Figure 5.8 Block selection for a wetter climate

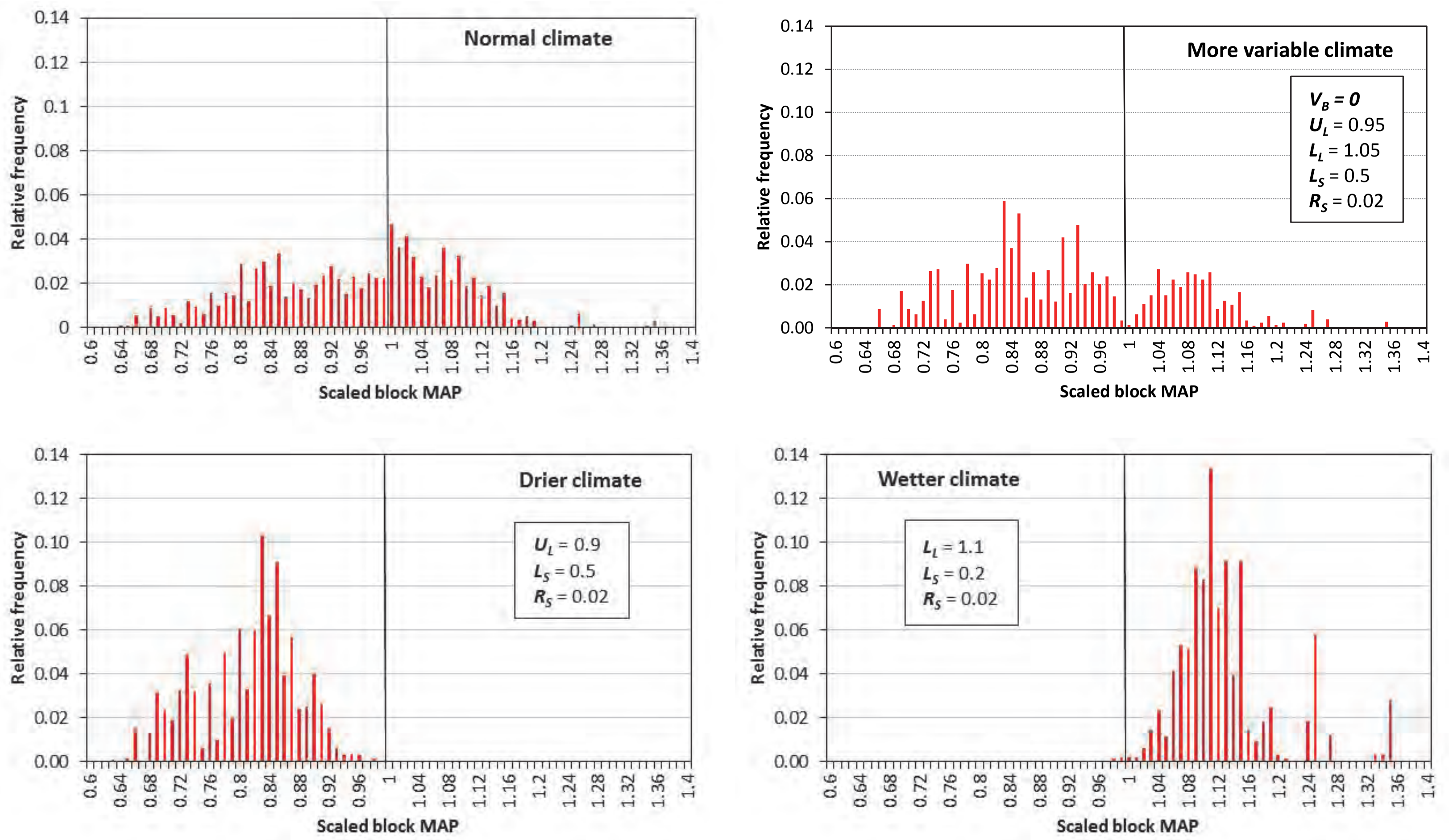


Figure 5.9 Probability density plots of scaled MAP of blocks selected in generating stochastic rainfalls for various climate scenarios

5.3 Annual Rainfall Statistics for Various Climate Change Scenarios

The effect of imposing climate change and variability on the generated rainfalls is assessed using five statistics; the mean, median, lowest rainfall, highest rainfall and the standard deviation of annual rainfalls. In addition, the minimum run sums test is also applied in the assessment. The assessment uses 101 99 years long simultaneous generations of rainfall at the 10 stations. Figures 5.10 to 5.14 show box plots of the five statistics for the normal climate, an extremely dry climate (using $\boldsymbol{U_L}$ = 0.6, $\boldsymbol{L_s}$ =0.99 and $\boldsymbol{R_S}$ = 0.02), an extremely wet climate (using $\boldsymbol{L_L}$ = 1.4, $\boldsymbol{L_s}$ = 0.99 and $\boldsymbol{R_S}$=0.02) and a moderately more variable climate (using $\boldsymbol{V_B}$ = 0, $\boldsymbol{U_L}$ = 0.95, $\boldsymbol{L_L}$ = 1.05, $\boldsymbol{L_s}$ = 0.5 and $\boldsymbol{R_S}$=0.02 with drier and wetter rainfalls in equal proportion). Figures 5.15 to 5.19 are minimum run sum box plots for the same scenarios for five of the 10 stations. The plots for the other five not presented here exhibit similar behaviour. The expected behaviour is obtained with all the statistics and the minimum run sums. It is also observed that the wetter climate (and a drier climate to a lesser extent) also increases the variability of most of the statistics.

Normal climate

More variable climate

Drier climate

Wetter climate

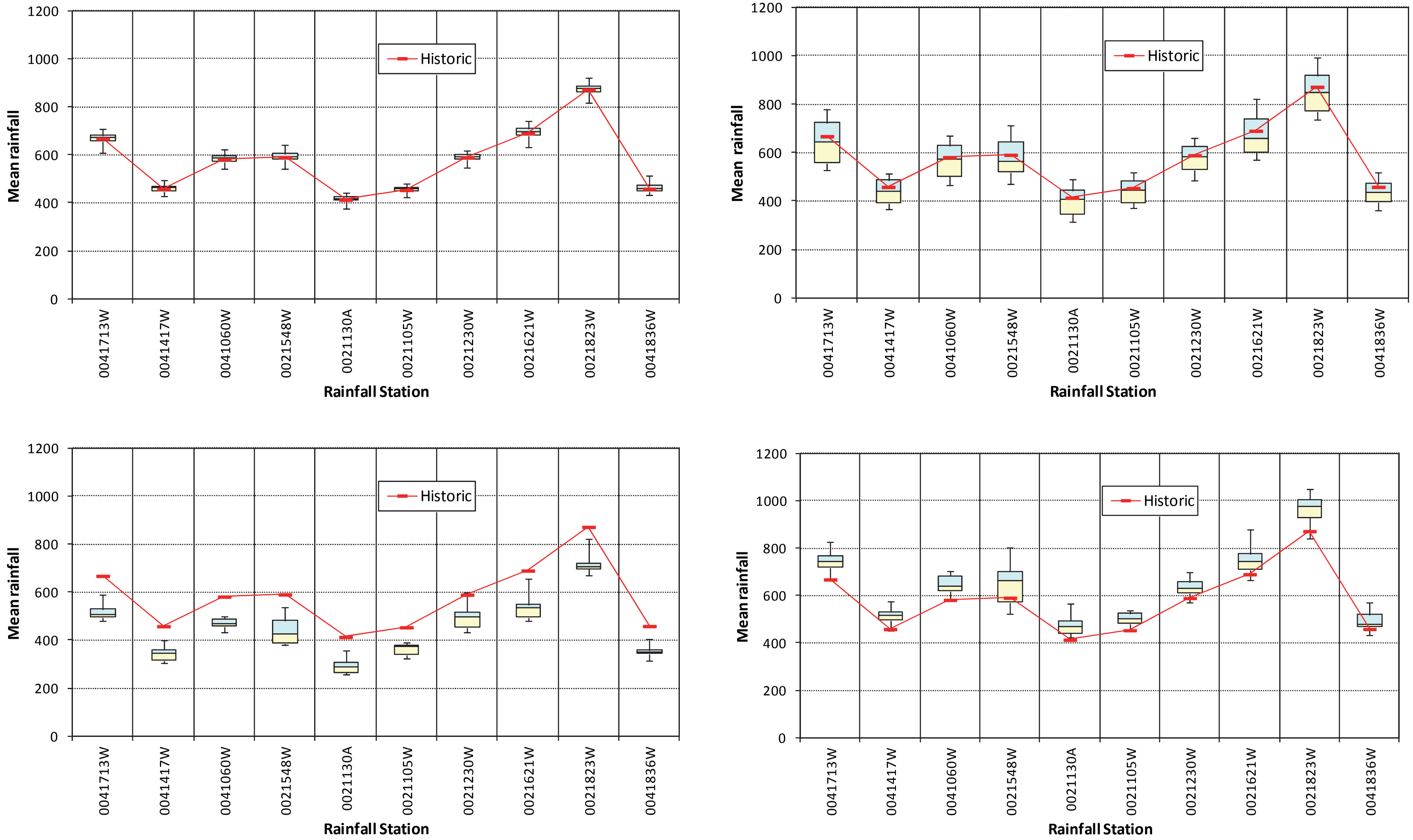


Figure 5.10 Box plots of mean of annual rainfalls for various climate scenarios

Normal climate

More variable climate

Drier climate

Wetter climate

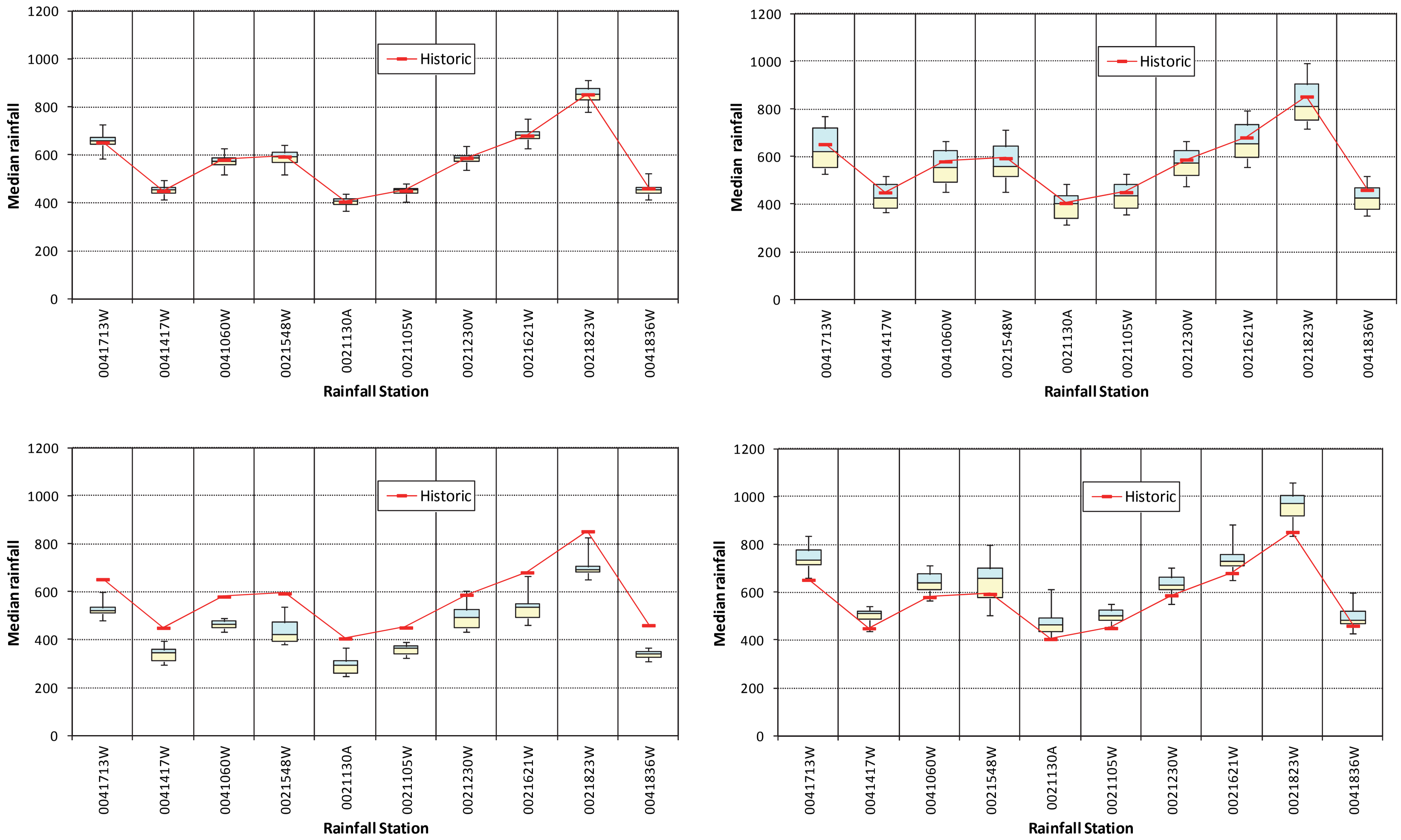


Figure 5.11 Box plots of median of annual rainfalls for various climate scenarios

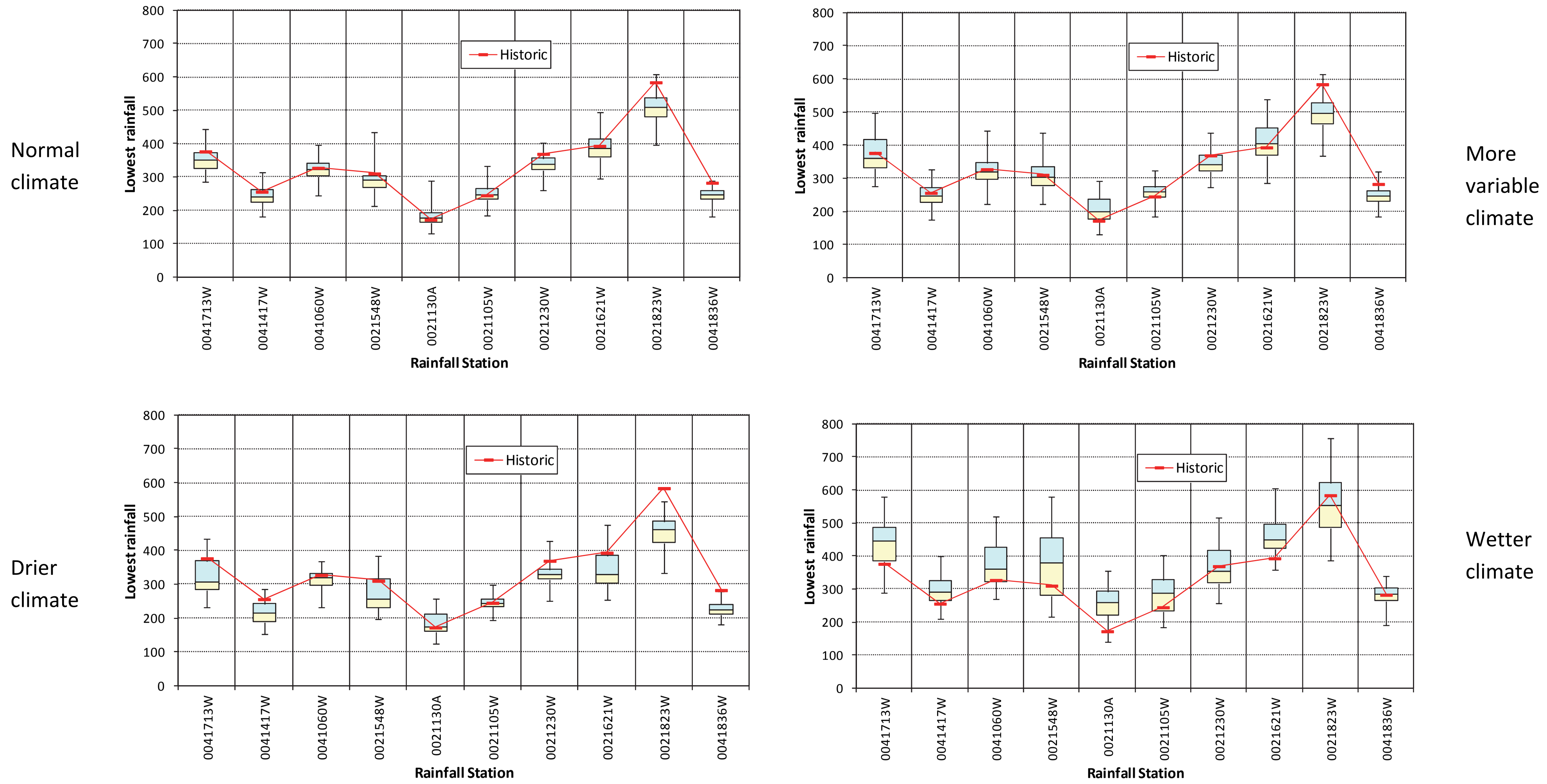
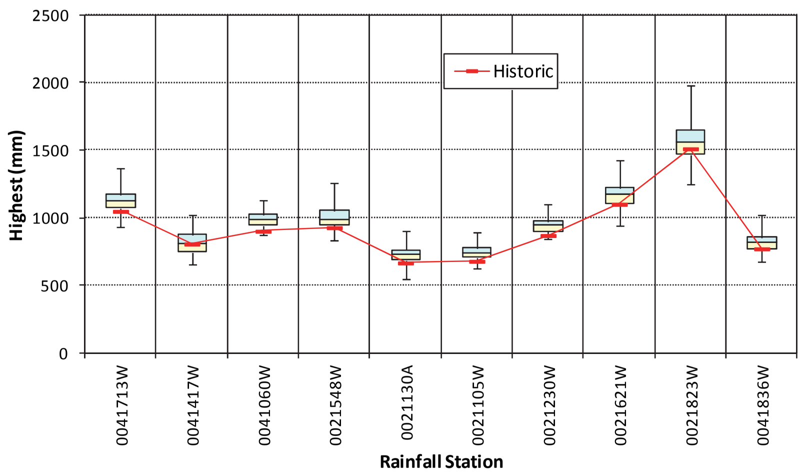
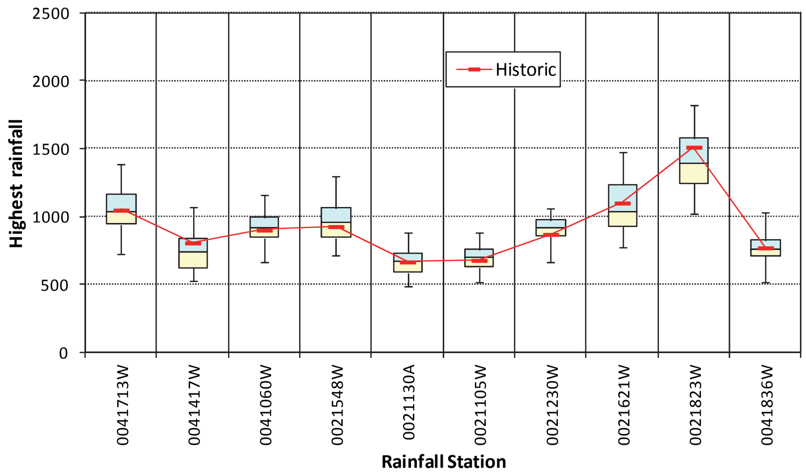


Figure 5.12 Box plots of the lowest annual rainfalls for various climate scenarios

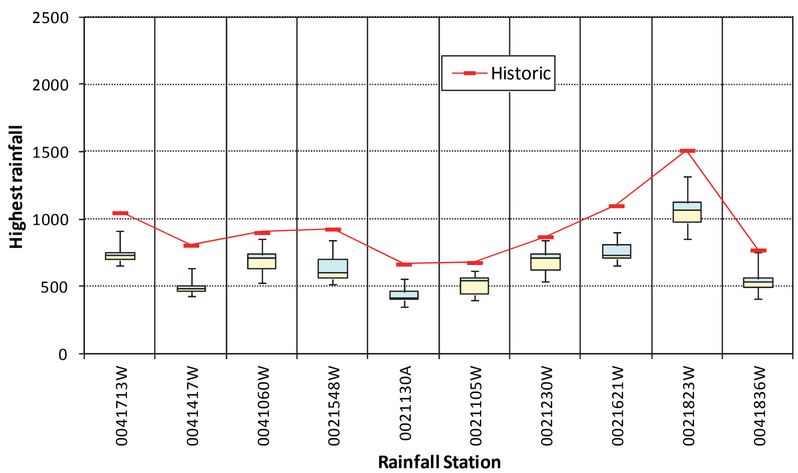
Normal climate

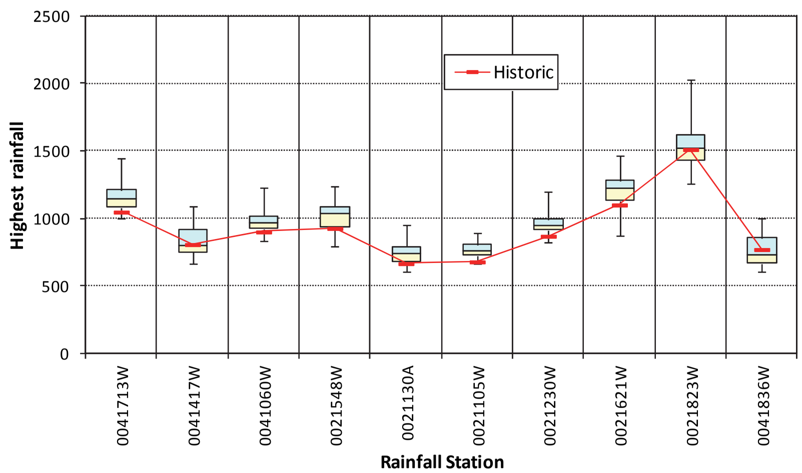




More variable climate

Drier climate





Wetter climate

Figure 5.13 Box plots of the highest annual rainfalls for various climate scenarios

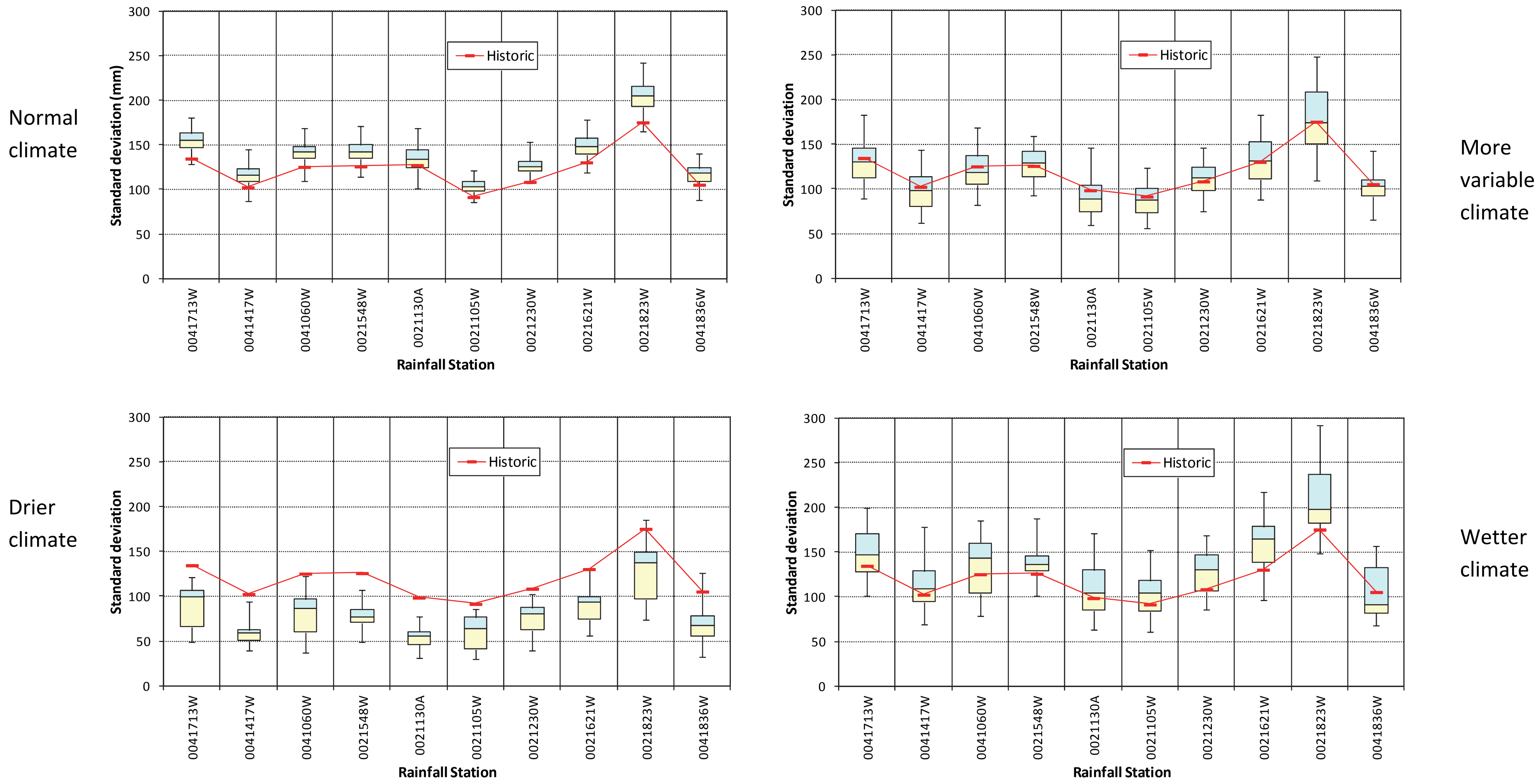


Figure 5.14 Box plots of the standard deviation of annual rainfalls for various climate scenarios

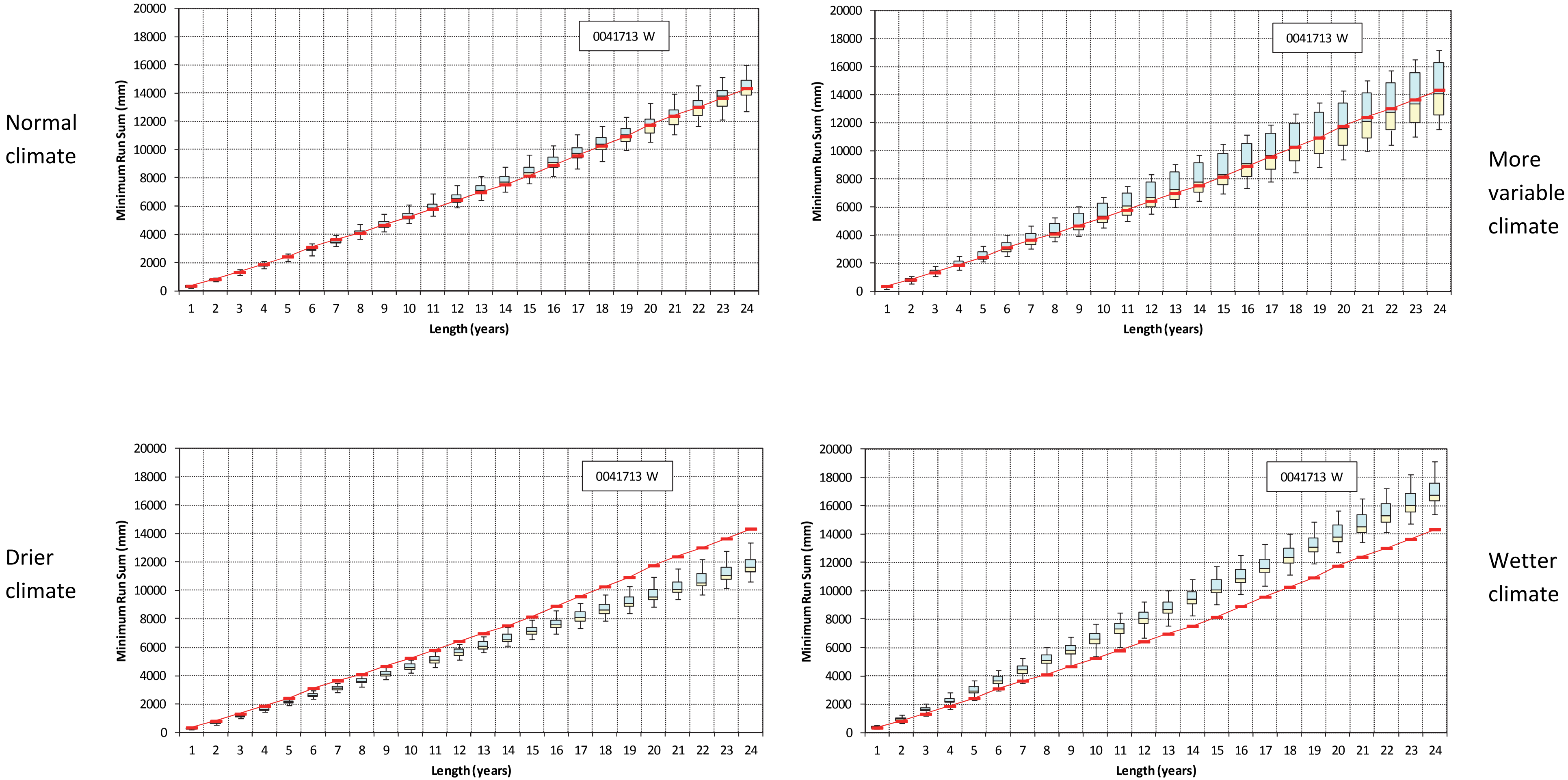


Figure 5.15 Box plots of minimum run sums for various climate scenarios for rainfall station 0041713 W

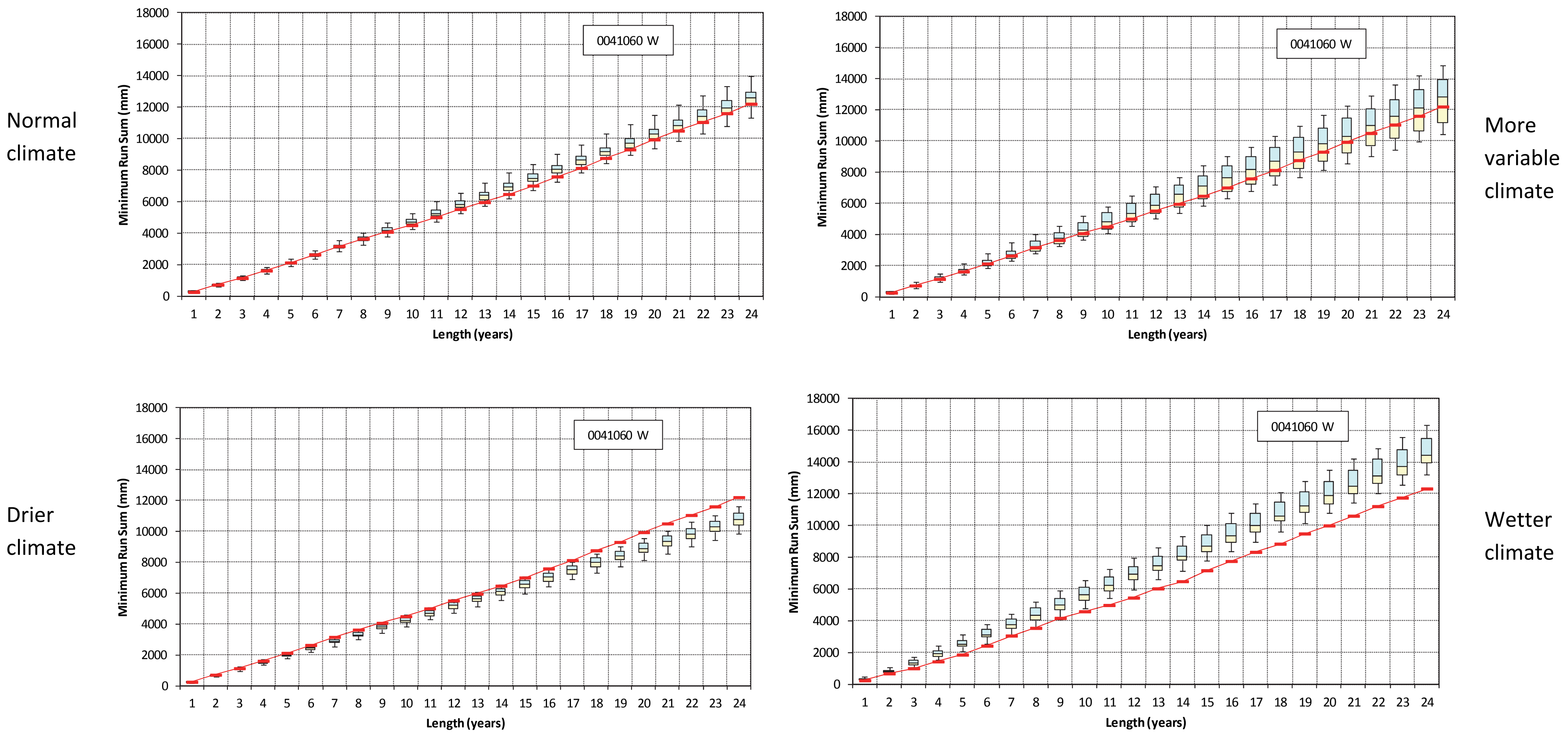


Figure 5.16 Box plots of minimum run sums for various climate scenarios for rainfall station 0041060 W

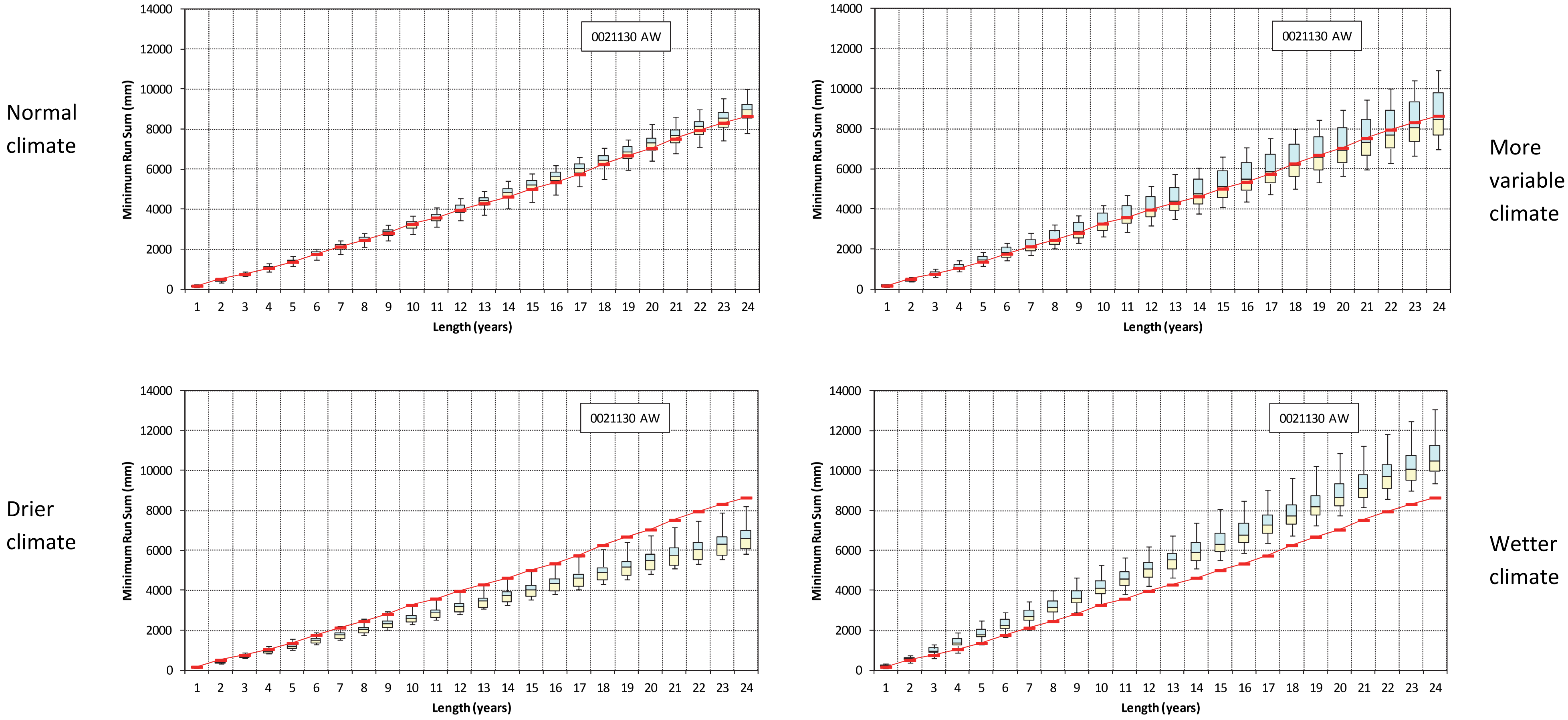


Figure 5.17 Box plots of minimum run sums for various climate scenarios for rainfall station 0021130 AW

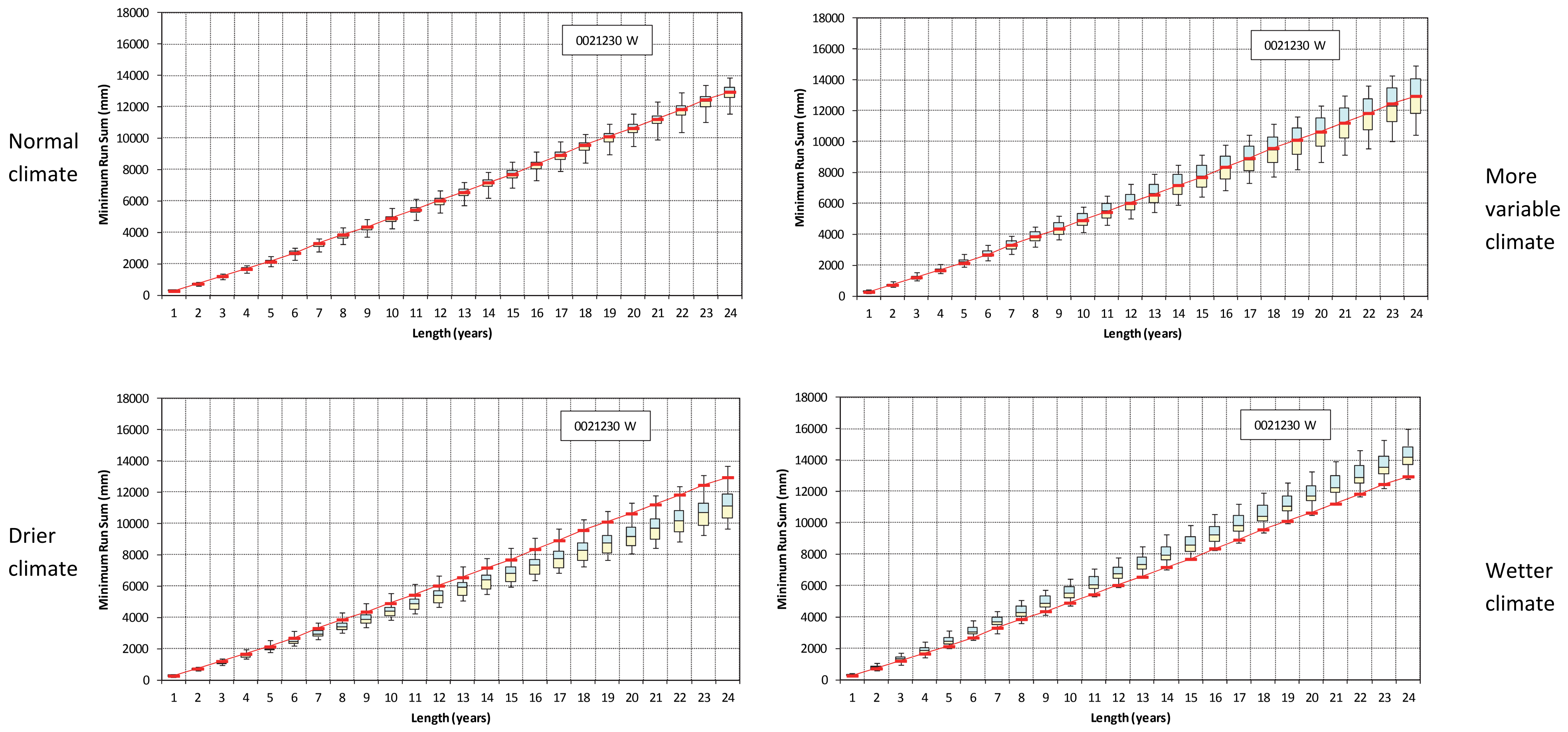


Figure 5.18 Box plots of minimum run sums for various climate scenarios for rainfall station 0021230 W

Normal climate

More variable climate

Drier climate

Wetter climate

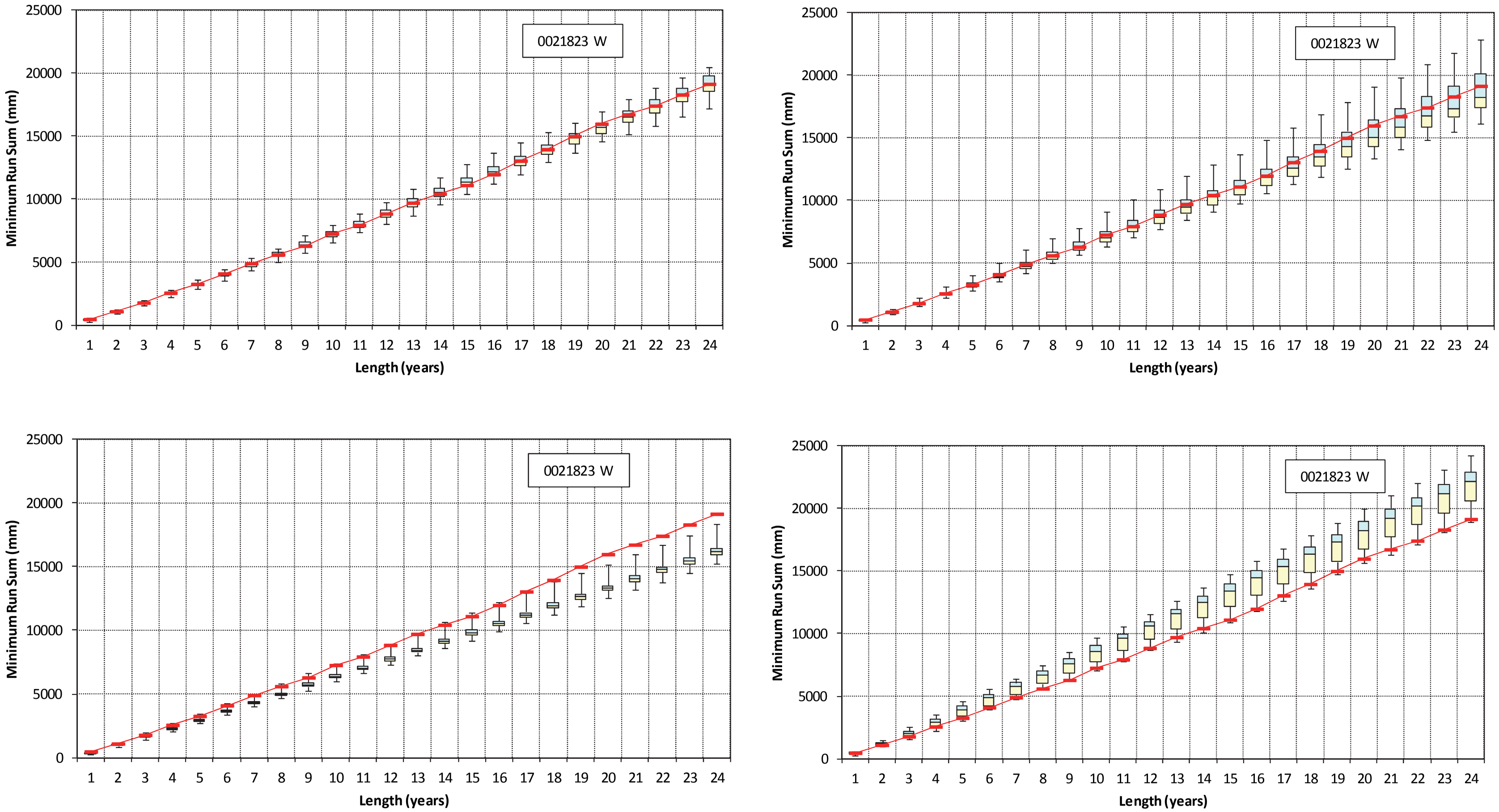


Figure 5.19 Box plots of minimum run sums for various climate scenarios for rainfall station 0021823 W

5.4 Incorporating GCM Projections into Rainfall Generator

5.4.1 *Matching average MAP of stochastic rainfalls to GCM projections*

Section 5.2 reveals the large change in mean annual rainfall and other statistics that can be obtained by biasing block selection. To obtain the projected change in MAP by a single or multiple GCMs, initial search limits $\boldsymbol{U_L}$ and $\boldsymbol{L_L}$ are varied iteratively (setting $\boldsymbol{L_S}$ and $\boldsymbol{R_S}$ to reasonable values; $\boldsymbol{L_S}$=1.0 and $\boldsymbol{R_S}$ =0.01 were used here) to obtain the required change in MAP. For the example problem of 10 rainfall stations where an overall reduction in MAP of 6.5 % is forecast, a $\boldsymbol{U_L}$ of 1.02 was found to obtain the required reduction as seen on Figure 5.20. Figure 5.20 shows the variation of the change in MAP with initial search limits for a dry and a wet climate. For this example, 501 99 years- long sequences were generating using the same VLB parameters as in Section 2. Figure 5.20 reveals the expected reduction in the percentage change in MAP as the bias reduces ($\boldsymbol{U_L}$ becomes higher for a dry climate and $\boldsymbol{L_L}$ becomes lower for a wetter climate).

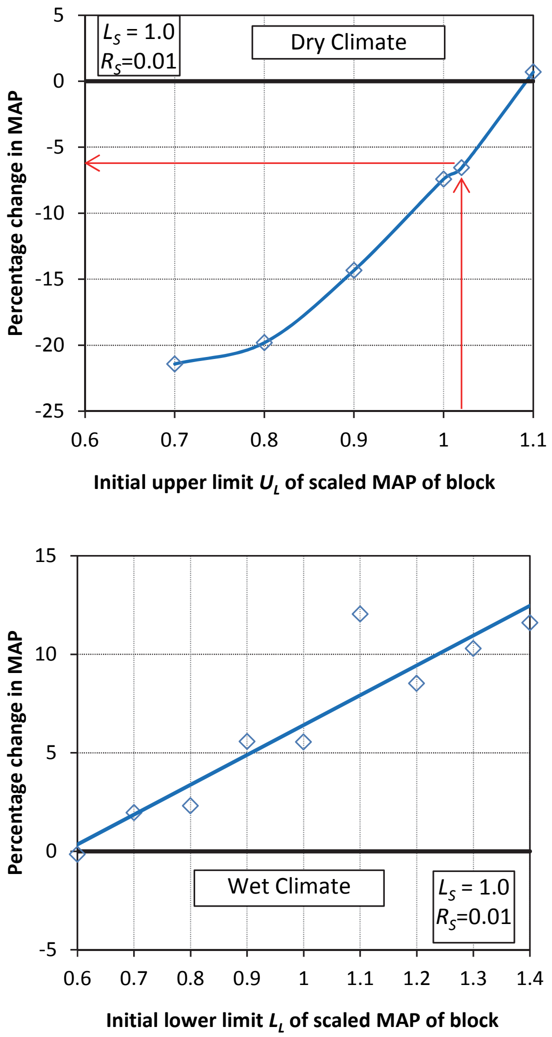


Figure 5.20 The percentage change in mean annual rainfall (MAP) for a range of upper and upper limits of the scaled MAP of blocks

For completeness, box plots of the annual rainfalls and the minimum run sums for the 501 stochastic sequences are presented in Figure 5.21 showing the expected reduction in mean rainfall for all the 10 sites. Figure 5.22 shows the expected reduction in the minimum run sums for one of the 10 sites.

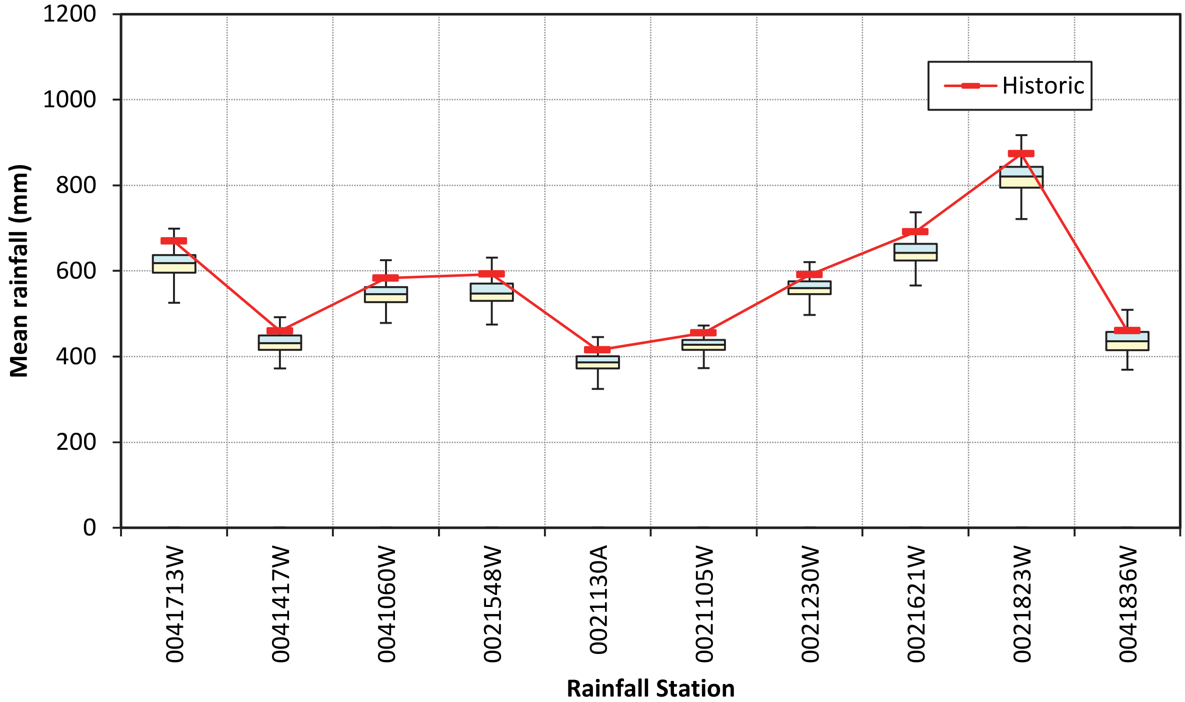


Figure 5.21 Box plot of GCM-projected mean annual rainfall compared with historic means

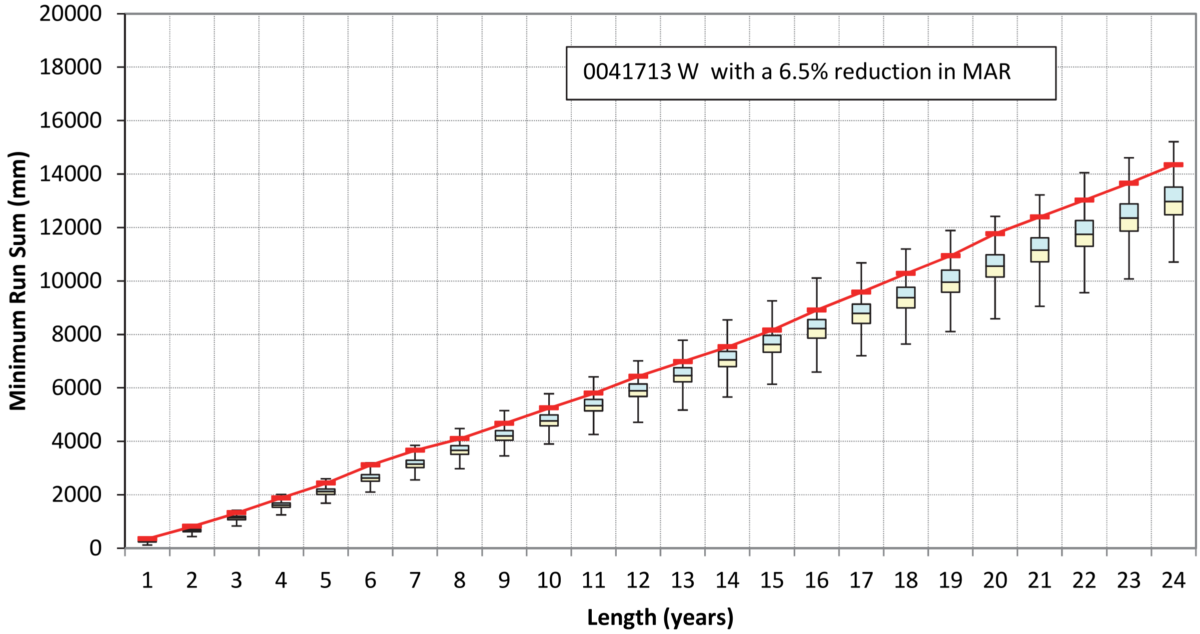


Figure 5.22 Box plots of GCM-projected minimum run sums compared with historic minimum run sums.

5.4.2 *Matching stochastic monthly rainfall distribution to seasonal GCM projections*

This section describes the computation of monthly rainfalls to match the average monthly rainfalls forecast by a multi model ensemble of GCMs. Figure 5.3 provides the average monthly rainfall distribution from 10 GCM projections for a representative station in the study area. In order to obtain an overall monthly rainfall distribution as projected by the GCMs whilst aiming to maintain the variability already existing in the monthly rainfalls, the following steps are followed:

i) Compute the grand monthly and the grand total rainfall from all stations and all stochastic sequences. From equation 3.22 in Chapter 3, $m_{i,j}^{sf}$ is the monthly rainfall for month *j* of year *i* for a stochastic sequence at a given site. If the site is denoted by *k* (*k*=1 to n_s where n_s is the number of sites) and the stochastic sequence by *p* (*p*=1 to n_p where n_p is the number of stochastic sequences), then the monthly stochastic rainfall for month *j* of year *i* of site *k* for sequence *p* can be expressed as $m_{i,j,k,p}^{sf}$. The grand monthly total for month *j* $M_{grand\ total,j}$ is then obtained as

$$M_{grand\ total,j} = \sum_{p=1}^{n_p}\sum_{k=1}^{n_s}\sum_{i=1}^{n_y} m_{i,j,k,p}^{sf} \qquad for\ j = 1,2,\quad 12 \qquad 5.1$$

The grand total rainfall $A_{grand\ total}$ is then obtained as

$$A_{grand\ total} = \sum_{j=1}^{12} M_{grand\ total,j} \qquad 5.2$$

ii) Using this total rainfall and the average monthly distribution from the GCMs $f_{GCM\ av,j}$, for month *j*=1 to 12, obtain the 12 grand monthly GCM totals $M_{grand\ total,j}^{GCM}$.

$$M_{grand\ total,j}^{GCM} = A_{grand\ total} \times f_{GCM\ av,j} \qquad 5.3$$

iii) Multiply each individual stochastic monthly rainfall by the ratio of the grand monthly GCM total to the grand monthly total to obtain the final GCM projection-modified monthly rainfall $m_{i,j,k,p}^{GCM}$. for month *j* of year *i* of site *k* for sequence *p*

$$m_{i,j,k,p}^{GCM} = m_{i,j,k,p}^{sf} \times \frac{M_{grand\ total,j}^{GCM}}{M_{grand\ total,j}} \qquad 5.4$$

iv) Aggregate the final monthly rainfalls to obtain new annual stochastic rainfalls.

In step ii), the individual GCM distributions that lead to the average monthly distribution could be used as a means of using multiple GCMs projections to obtain additional variability on the stochastic sequences. For this, the individual monthly distributions from each GCM, $f_{GCM,j}^{q}$ for GCM number *q* (*q* = 1 to n_{GCM} where n_{GCM} is the total number of GCMs used) are applied to obtain different grand monthly GCM totals.

$$M_{grand\ total,j,q}^{GCM} = A_{grand\ total} \times f_{GCM,j}^{q} \qquad 5.5$$

The final GCM-modified monthly rainfalls are then obtained as

$$m_{i,j,k,p}^{GCM} = m_{i,j,k,p}^{sf} \times \frac{M_{grand\ total,j,q}^{GCM}}{M_{grand\ total,j}} \qquad \begin{array}{c} for\ q = 1, 2, \quad n_{GCM} \qquad and \\ p = \dfrac{n_p}{n_{GCM}}(q-1)+1 \quad to \quad \dfrac{n_p}{n_{GCM}}q \end{array} \qquad 5.6$$

The approach that uses the monthly distributions of the individual GCMs (using equations 5.1, 5.2, 5.5 and 5.6) is applied and a total of 501 stochastic sequences are generated using the parameters applied in Section 3.1. Each GCM distribution is therefore used 50 times since projections from 10 GCMs were applied.

The box plots of the mean monthly rainfalls for 5 of the 10 sites of the generations are shown on Figure 5.23 while Figure 5.24 compares the mean monthly rainfalls of the normal climate to the medians of the mean monthly rainfalls of the GCM projection stochastic monthly rainfalls. Both Figures reveal the reduction in the monthly rainfalls and also shifts in the seasonal rainfall patterns as projected by the GCMs. It is noted that the historic rainfalls on Figure 5.21 and 5.22 are for the entire 99 years of historic data while the historic data used to obtain the normal climate on Figure 5.3 is based on a 20 year period. Differences in the change in rainfall are evident especially for August where Figures 5.21 and 5.22 display a much bigger reduction than Figure 5.3.

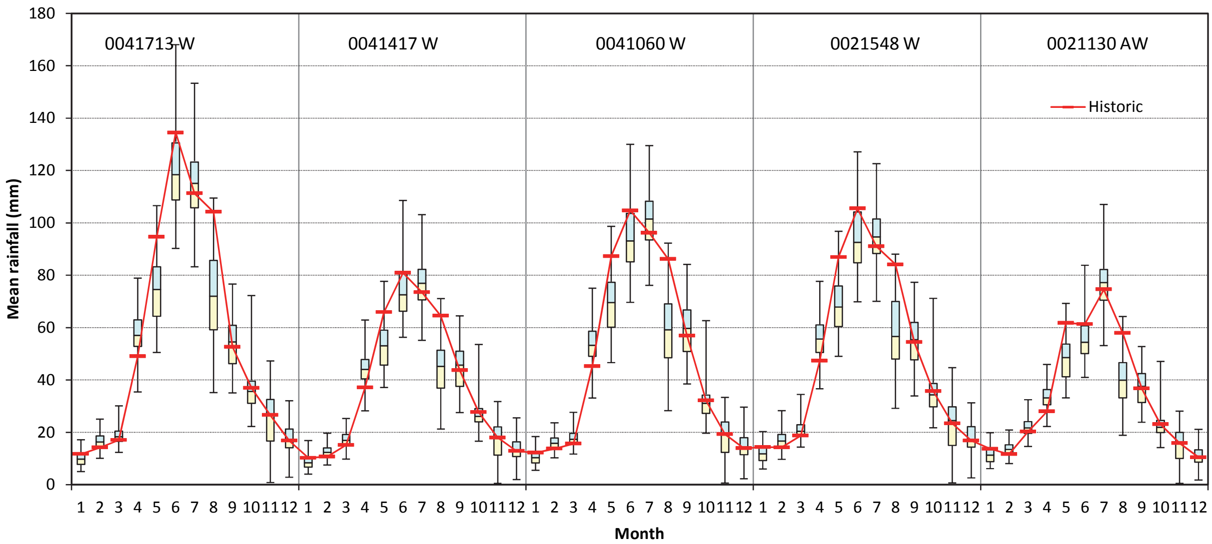


Figure 5.23 Box plots of GCM projected monthly mean rainfall compared to historic monthly means

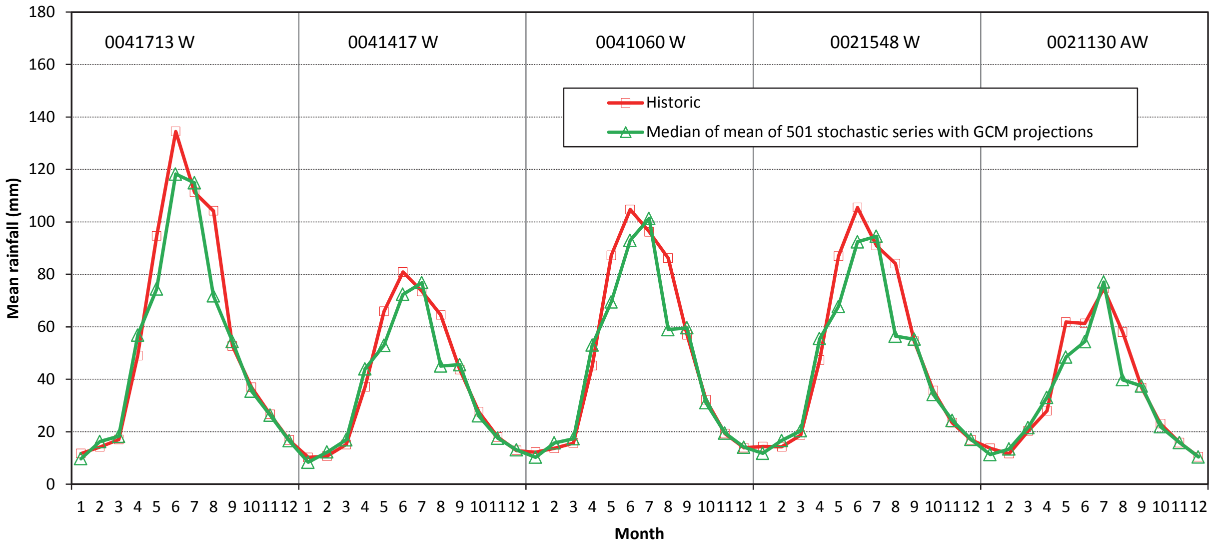


Figure 5.24 Comparison of the median of GCM projected rainfalls with historic mean monthly rainfalls

5.5 Summary of Climate Change and Variability Modelling

A simple and effective method of modelling climate change and variability has been developed by biasing VLB block selection based on the average block mean annual precipitation (MAP). The method applies parameters that can be realistically set to obtain stochastic rainfall sequences of a wetter, drier or a more variable climate. In addition, it has been demonstrated how the approach can be used to generating stochastic rainfalls that match the overall shifts in MAP obtained from GCM projections and how the monthly distributions of the stochastic rainfalls can be made to match seasonal rainfall shifts projected by multiple GCMs. Box plots obtained from the climate variability modeling indicate that the generator can be used to obtain a wide range of realistic climate change/variability scenarios and could therefore be very useful for assessments of the possible impacts of climate change and/or variability on water resources systems.

6 CONCLUSIONS AND RECOMMENDATIONS

This project set out to develop and test a monthly non-parametric stochastic rainfall generator that would comprehensively incorporate climate change/variability including information from global climate model (GCM) projections. The project also included a comparison of the non-parametric generator with a parametric one. Two multisite rainfall generation problems were applied; one with 10 stations widely spread across South Africa with negligible patching and the other with 10 stations in a catchment located in the Western Cape with about 60% of the data patched. The variable length bootstrap (VLB) stochastic generator, previously used in streamflow generation was selected and adapted to rainfall generation for a normal climate and for a drier, a wetter or a more variable climate. To assess the performance of the generator, comparison of boxplots with historic values of several statistics was done. The statistics applied are the mean, the median, the 25^{th} and the 75^{th} percentile, the lowest and the highest rainfall, the standard deviation, the skewness, and the serial and cross correlation coefficients. In addition, the minimum run sums test that compares the minimum historic and stochastic annual rainfalls lengths ranging from 1-24 years is included to assess the replication of long-term dependence characteristics. The parametric PEGRAIM-W monthly stochastic rainfall generator that is currently being used in WRC Project K5/2155 was selected for the comparison.

The VLB generator adapted very well to rainfall generation as rainfall data all the statistical measures were replicated reasonably well at the annual and monthly time scale. The VLB model performed better than PEGRAIM-W at the annual and monthly time scale although both VLB and PEGRAIM-W models were found to perform reasonably well for practical application.

The generation of stochastic rainfalls for a wetter, a drier or a more variable climate was achieved by appropriately biasing block selection based on the mean annual precipitation (MAP) of individual blocks. This approach was found to be effective and capable of generating sequences of highly varied characteristics (as quantified by the statistics). It was demonstrated how the change on average MAP from GCMs can be achieved by simple iteration involving a single parameter of the block selection model. An approach for matching the monthly rainfall patterns to those of GCMs was developed and also demonstrated. The literature however revealed that GCM projections are still highly uncertain and the methods developed here therefore allow for the generation of rainfalls for drier, wetter or more variable climate without the use of GCM projections.

Following are some notable highlights from this project.

i. Bootstrap methods are robust stochastic generators as they cause low levels of distortion to the basic characteristics of the data and make assumption that are easy to understand and therefore implement.
ii. Monthly rainfalls lend themselves easily to bootstrap (resampling) stochastic generation as monthly and annual serial correlations are usually negligible.
iii. The weighted fragments method of perturbing annual rainfalls (Section 3.3.3 and Figure 3.11) is very successful in overcoming the failure of the basic bootstrap to generate any new data.

iv. The use of variable block lengths obtains blocks of highly variable characteristics (Section 5.2 and Figures 5.4, 5.5, 5.6a and 5.6b)and biasing block selection based on these characteristics can makes data generation very versatile. This was demonstrated here in the generation of stochastic rainfalls for a drier, a wetter or a more variable climate.

It is recommended that the VLB generator be tested for practical water resources systems studies as in the on-going WRC Project K5/2155 that is testing the PEGRAIM-W generator. It is also recommended that the climate change and variability modeling developed in this study be tested for climate change studies as it can effectively complement GCM and/or Regional circulation model (RCM) rainfall outputs for a wide range of hydrological and water resources analyses.

It is likely that an effective daily stochastic rainfall generator can be developed by applying appropriate disaggregation to the monthly rainfalls generated by the VLB model and this is proposed for future development.

The block termination approach of the current VLB generator (Section 3.2.1) is realistic but could probably be improved by using more comprehensive methods of identifying the underlying patterns in the historic time series. The possibility of using Empirical Mode Decomposition (EMD) is therefore being investigated and will be reported in the MSc research report of J Nyaga.

LIST OF REFERENCES

Aleix S-C, Valdes, JB, Dominguez F. and Rajagopal S (2013). Characterizing the water extremes of the new century in the US South-west: a comprehensive assessment from state-of-the-art climate model projections, International Journal of Water Resources Development, 29:2, 152-171.

Anagnostopoulos GG, Koutsoyiannis D, Christofides A, Efstratiadis A and Mamassis N (2010). A comparison of local and aggregated climate model outputs with observed data. Hydrol. Sci. J. 55(7), 1094-1110.

Basson MS and Van Rooyen J A (2001). Practical application of probabilistic approaches to the management of water resource systems, Journal of Hydrology 241 (2001) 53-61.

Basson MS, Allen RB, Pegram GGS and Rooyen JA (1994). Probabilistic Management of Water Resource and Hydropower Systems. Water Resour. Publ., Colorado, USA. 424 pp.

Bayazit M, ÖnÖz B, Aksoy H (2001). Non-parametric streamflow simulation by wavelet or Fourier analysis. Hydrological Sciences Journal 46 (4).

Brissete FP, Khalili M, Leconte R (2007). Efficient stochastic generation of multi-site synthetic precipitation data. Journal of Hydrology 345, 121-133.

Buishand TA, Brandsma T (2001). Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbour resampling. Water Resources Research, Vol. 37(11), 2761-2776.

Buytaert W, Ce´lleri R, and Timbe L (2009). Predicting climate change impacts on water resources in the tropical Andes: Effects of GCM uncertainty, Geophysical Research Letters, Vol. 36, L07406, doi:10.1029/2008GL037048, 2009.

Chen J, Brissette F P and Leconte R (2010). A daily stochastic weather generator for preserving low-frequency of climate variability, Journal of Hydrology 388, 480-490.

Clark MP, Gangopadhyay S, Hay LE, Rajagopalan B and Wilby RL (2004). The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields, J. Hydrometeorol., 5, 243- 262.

Clark MP, Gagopadhyay S, Brandon D, Werner LH, Rajagopalan B and Yates D (2004). A resampling procedure for generating conditioned daily weather sequences. Water Resources Research, Vol. 40, W04304.

CSAG (Climate Systems Analysis Group) (2008) Downscaled GCMs for different time slices. Report prepared for WRC Research Project No. K5/1843. Climate Systems Analysis Group, University of Cape Town, RSA.

Dessu SB and Melesse AM (2013). Evaluation and Comparison of Satellite and GCM Rainfall Estimates for the Mara River Basin, Kenya/Tanzania. Younos T.and Grady C.A. (eds.), Climate Change and Water Resources, Hdb Env Chem (2013) 25: 29-46, DOI 10.1007/698_2013_219.

Dyer F, ElSawah S, Croke B, Griffiths R, Harrison E, Lucena-Moya P and Jakeman A (2014). The effects of climate change on ecologically-relevant flow regime and water quality attributes, Stoch. Environ. Res. Risk. Assess (2014) 28:67-82.

Eisinger D and Wiegand K (2008). SERGE: A spatially explicit generator of local rainfall in Southern Africa. South African Journal of Science 104(1-2), 37-42.

Engelbrecht FA, Landman WA, Engelbrecht CJ, Landman S, Bopape MM, Roux B, McGregor JL and Thatcher M (2011). Multi-scale climate modelling over Southern Africa using a variable-resolution global model, Water SA Vol. 37 No. 5 647-658.

Faucher D, Rasmussen PF, Bobée B, (2001). A distribution function based bandwidth selection method for kennel quantile estimation. Journal of Hydrology 250, 1-11.

Ghil M (2002). Natural Climate Variability. Vol. 1, The Earth system: physical and chemical dimensions of global environmental change, Eds., MacCracken M C and Perry J S., in Encyclopedia of Global Environmental Change, Ed. -in-Chief, Munn T. John Wiley. P 544-549.

Green G C (2008). Towards Defining the WRC Research Portfolio on Climate Change for 2008-2013, WRC Report No. KV 207/08.

Harrold TI, Sharma A, Sheather SJ (2003). A non-parametric model for stochastic generation of daily rainfall amounts. Water Resources Research, vol. 39(12) 1343.

Hughes DA (2004). Three decades of hydrological modelling research in South Africa. South African Journal of Science 100, 638-642.

Hughes DA (2012). Hydrological impacts of climate change in southern Africa – what do we really know? Proceedings,16th SANCIAHS Symposium, Pretoria, South Africa.

Hughes DA, Kingston DG and Todd MC (2011). Uncertainty in water resources availability in the Okavango River Basin as a result of climate change. Hydrol. Earth Syst. Sci. 15, 931-941.

Hurst HE (1951). Long term storage capacities of reservoirs. Trans. ASCE 116, 776-808.

Ilich N and Despotovic J (2008). A simple method for effective multisite generation of stochastic hydrological time series. Stochastic Environment Resources Risk Assessment 22,265-279.

Johnson F and Sharma A (2009). Measurement of GCM Skill in Predicting Variables Relevant for Hydroclimatological Assessments. J. Climate, 22, 4373-4382.

Kenabatho PK, McIntyre NR, Chandler RE, Wheater HS (2012). Stochastic simulation of rainfall in the semi-arid Limpopo basin, Botswana. International journal of climatology 32(7), 1113-1127.

Kim T, Ahn H, Chung G and Yoo C (2008). Stochastic multi-site generation of daily rainfall occurrence in South Florida. Stochastic Environmental Risk Assess 22, 705-717.

Koutsoyiannis D, Efstratiadis A, Mamassis N and Christofides A (2008). On the credibility of climate predictions, Hydrological Sciences Journal, 53: 4, 671- 684.

Koutsoyiannis D (2011). Hurst-Kolmogorov Dynamics and Uncertainty. Journal of the American Water Resources Association (JAWRA) 47(3):481-495. DOI: 10.1111/j.1752-1688.2011.00543.x.

Koutsoyiannis D, Efstratiadis A and Georgakakos K (2007). Uncertainty assessment of future hydroclimatic predictions: a comparison of probabilistic and scenario-based approaches. J. Hydromet. 8(3), 261-281.

Koutsoyiannis D, Efstratiadis A, Mamassis N and Christofides A (2008). On the credibility of climate predictions. Hydrol. Sci. J. 53(4), 671-684.

Koutsoyiannis D, Montanari A, Lins H F and Cohn TA (2009). Climate, hydrology and freshwater: towards an interactive incorporation of hydrological experience into climate research Discussion of "The implications of projected climate change for freshwater resources and their management, Hydrol.Sci. J., 54(2). 394-405.

Kundzewicz ZW, Mata lJ, Arnell NW, Döll P, Jimenez B, Miller K, Oki T, Şen Z and Shiklomanov I (2008). The implications of projected climate change for freshwater resources and their management, Hydrol.Sci. J., 53:1, 3-10.

Kundzewicz ZW and Stakhiv EZ (2010). Are climate models “ready for prime time” in water resources management applications, or is more research needed?, Hydrol.Sci. J., 55(7), 1085 —1089.

Kunz R (2009). Rainfall Data Extraction, Version Number 1.2, ICFR, PMB, South Africa.

Lall U and Sharma A (1996). A nearest neighbour bootstrap for resampling hydrological time series. Water Resources Research, Vol. 32(3), 679-693.

Lee T and Ouarda TBMJ (2011). Prediction of climate nonstationary oscillation processes with empirical mode decomposition, J. Geophys. Res., 116, D06107, doi: 10.1029/2010JD015142.

Lin G and Lee F (1992). An aggregation-disaggregation approach for hydrologic time series modelling. Journal of Hydrology 138, 543-557.

Lloyd P (2009). On the determination of trends in rainfall. Water SA 35(3).

Lumsden TG, Schulze RE and Hewitson BC (2011). Evaluation of potential changes in hydrologically relevant statistics of rainfall in Southern Africa under conditions of climate change, Water SA Vol. 35 No. 5, 649-656.

Maheepala S and Perera BJC (1996). Monthly data generation by disaggregation. Journal of Hydrology 178, 277-291.

Mehrotra R and Sharma A (2006). Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach. Advances in Water Resources 29, 987-999.

Mehrotra R and Sharma A (2007). A semi parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. Journal of Hydrology 335, 180-193.

Mehrotra R and Sharma A (2009). Evaluating spatio-temporal representations in daily rainfall sequences from three stochastic multi-site weather generation approaches. Advances in Water Resources 32,948-962.

Mehrotra R, Srikanthan R and Sharma A (2006). A comparison of three multi-site precipitation occurrence generator. Journal of Hydrology 331, 280-292.

Mejia JM and Rouselle J (1976). Disaggregation models in hydrology revisited. Water Resources Research, 12, 185-186.

Mujumdar PP and Ghosh S (2008). Modeling GCM and scenario uncertainty using a possibilistic approach: Application to the Mahanadi River, India, Water Resour. Res., 44, W06407, doi:10.1029/2007WR006137.

Murray C, Peel MC and McMahon TA (2006). Recent frequency component changes in interannual climate variability, Geophysical Research Letters, 33, 1-5, L16810, doi:10.1029/2006 GL025670.

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

Ndiritu J (2011b). Effect of fragment-based perturbations on the disaggregation performance of a VLB streamflow generator. Physics and Chemistry of the Earth 36, 823-829.

Parlange MC and Katz WK (1999). An Extended Version of the Richardson Model for Simulating Daily Weather Variables Environmental and Societal Impacts Group, National Centre for Atmospheric Research, Boulder, Colorado.

Peel MC, Pegram GGS and McMahon TA (2004). Global analysis of runs of annual precipitation and runoff equal to or below the median. run length. Int. J. Climatol. 24: 807-822.

Peel MC, McMahon TA and Pegram GGS (2005). Global analysis of runs of annual precipitation and runoff equal to or below the median: run magnitude and severity. Int. J. Climatol. 25: 549-568.

Pegram GGS and McKenzie RS (1991). Synthetic streamflow generation in the Vaal River System Study, The Civil Engineer in South Africa, 15-24.

Pegram GGS (2011). PEGRAIM-W Users Manual: Report to Department of Water Affairs

Piantadosi J, Howlett P, Borwein J and Henstridge J (2012). Maximum entropy methods for generating simulated rainfall. Numerical algebra, control and optimization (2) 2, 233-256.

Prairie JR, Rajagopalan B, Fulp TJ and Zagona EA (2006). Modified K-NN Model for Stochastic Streamflow Simulation. Journal of Hydrologic Engineering, 11(4), 371-378.

Rajagopalan B and Lall U (1999). A K-nearest neighbour simulator for daily precipitation and other weather variables. Water Resources Research Vol. 35(10) 3089-3101.

Rao RM and Bopardikar AJ (1998). Wavelet Tansforms, Introduction to Theory and Applications. Addison-Wesley, Reading, Massachusetts, USA.

Schulze RE (2011). Approaches towards practical adaptive management options for selected water-related sectors in South Africa in a context of climate change, Water SA Vol. 37 No. 5, 621-646.

Schulze RE (2012). Evaluation of Uncertainties when Assessing Climate Change Impacts on Hydrological Responses in South Africa, Proceedings 16th SANCIAHS Symposium, Pretoria, South Africa.

Schulze RE, Hewitson BC, Barichievy KR, Tadross MA, Kunz RP, Horan MJC and Lumsden TG (2011). Methodological Approaches to Assessing Eco-Hydrological Responses to Climate Change in South Africa, WRC Report No. 1562/1/10.

Scott DW (1992). Multivariate Density Estimation: Theory, Practice and Visualization, John Wiley & Sons Inc

Serinaldi F and Kilsby CG (2012). A modular class of multisite monthly rainfall generators for water resource management and impact studies, Journal of Hydrology 464-465,528-540.

Sharif M and Burn D H (2006). Simulating climate change scenarios using an improved K-nearest neighbor model, Journal of Hydrology 325, 179-196.

Sharma A and O'Neill R (2002). A non-parametric approach for representing interannual dependence in monthly streamflow sequences. Water Resources Research, 38(7), 1100.

Sharma A (2000). Seasonal to inter-annual rainfall probabilistic forecasts for improved water supply management: Part 3 – A non-parametric probabilistic forecast model. Journal of Hydrology 239, 249-258.

Sharma A, Harrold TI and Sheather SJ (2003). A non-parametric model for the stochastic generation of daily rainfall amounts. Water Resources Research, 39(12), 1343.

Sharma A, Srikanthan R and McMahon TA (2002). Stochastic generation of monthly rainfall data. Cooperative research centre for catchment hydrology, Technical Report 02/8.

Srikanthan R, Harrold TI, Sharma A and McMahon TA (2005). Comparison of two approaches for generation of daily rainfall data. Stochastic Environmental Resources Assessment, 19, 215-226.

Srikanthan R and McMahon TA (2001). Stochastic generation of annual, monthly and daily climate data: A review. Hydrology and Earth System Sciences, 5(4), 653-670.

Srikanthan R and Pegram GGS (2009). A nested multisite daily rainfall stochastic generation model. Journal of Hydrology 371 (2009) 142-153.

Srikanthan R and McMahon TA (2006). Comparison of two non-parametric alternatives for stochastic generation of monthly rainfall. Journal of Hydrologic Engineering, vol. 111 (3), 222-229.

Srinivas VV and Srinivasan K (2000). Post blackening approach for modelling dependent annual streamflows. Journal of hydrology 230, 86-126.

Srinivas VV and Srinivasan K (2005). Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows. Journal of hydrology 302, 307-330.

Srinivas VV and Srinivasan K (2006). Hybrid matched-block bootstrap for stochastic simulation of multi-season streamflows. Journal of hydrology 329, 1-5.

Steynor AC, Hewitson BC and Tadross MA (2009). Projected future runoff of the Breede River under climate change, Water SA Vol. 35 No. 4, 433-440.

Tasker GD and Dunne P (1997). Bootstrap position analysis for forecasting low flow frequency. Journal of Water Resources Planning and Management ASCE 123 (6), 359-367.

Ünal NE, Aksoy H and Akar T (2004). Annual and monthly rainfall data generation schemes. Stochastic Environmental Resources Assessment, vol. 18,245-257.

Van Rooyen P and McKenzie R (2004). Monthly Multi-Site Stochastic Streamflow Model, STOMSA User Guide, WRC Report No. 909/1/04.

Vaze J, Teng J and Chiew FHS (2011). Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia, Hydrol. Process. 25, 1486-1497.

Vogel RM and Shallcross AL (1996). The moving blocks bootstrap versus parametric time series models. Water Resources Research, vol. 32 (6) 1875-1882.

Wang QJ and Nathan RA (2007). A method for coupling daily and monthly time scales in stochastic generation of rainfall series. Journal of hydrology, 346, 122-130.

Wang W and Ding JA (2007). A multivariate non-parametric model for synthetic generation of daily streamflow. Hydrological Processes 21, 1764-1771.

Wang W, Hu S and Li Y (2011). Wavelet transform method for synthetic generation of daily streamflow. Water Resources Management 25, 41-57.

Weiland SFC, van Beek LPH, Kwadijk JCJ and Bierkens MFP (2010). The ability of a GCM-forced hydrological model to reproduce global discharge variability, Hydrol. Earth Syst. Sci., 14, 1595-1621.

Wilks DS (1998). Multi-site generalization of daily stochastic precipitation generation model. Journal of Hydrology 210, 178-191.

Zucchini W and Adamson PT (1984). The occurrence and severity of droughts in South Africa. WRC Report No. 91/1/84. Water Research Commission, Pretoria.

APPENDIX A: APPLICATION OF EMPIRICAL MODE DECOMPOSITION FOR VLB BLOCK TERMINATION

This is a summary of some of the results obtained from the MSc research project by J Nyaga.

1 EMPIRICAL MODE DECOMPOSITION

Figures A1 shows the 94 year long historic time series for rainfall station 0020866 W while Figure A2 to A6 show the Intrinsic Mode Functions (IMFs) obtained for rainfall station using Empirical Mode Decomposition (EMD). Figure A7 shows the residual from the EMD for the same station.

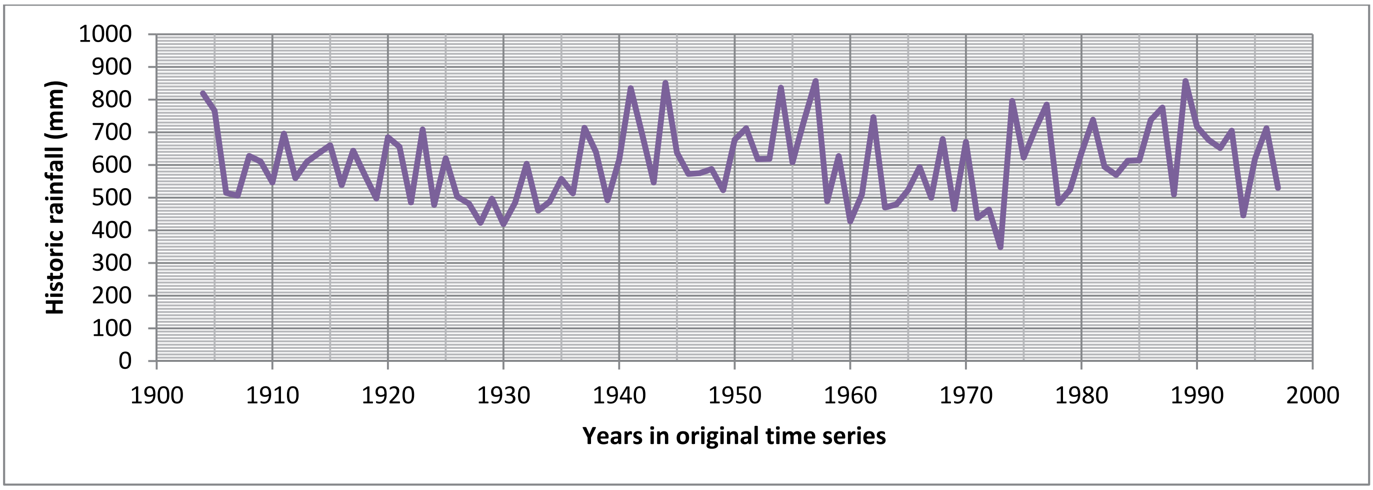


Figure A1- Annual historic time series of rainfall station 0020866W

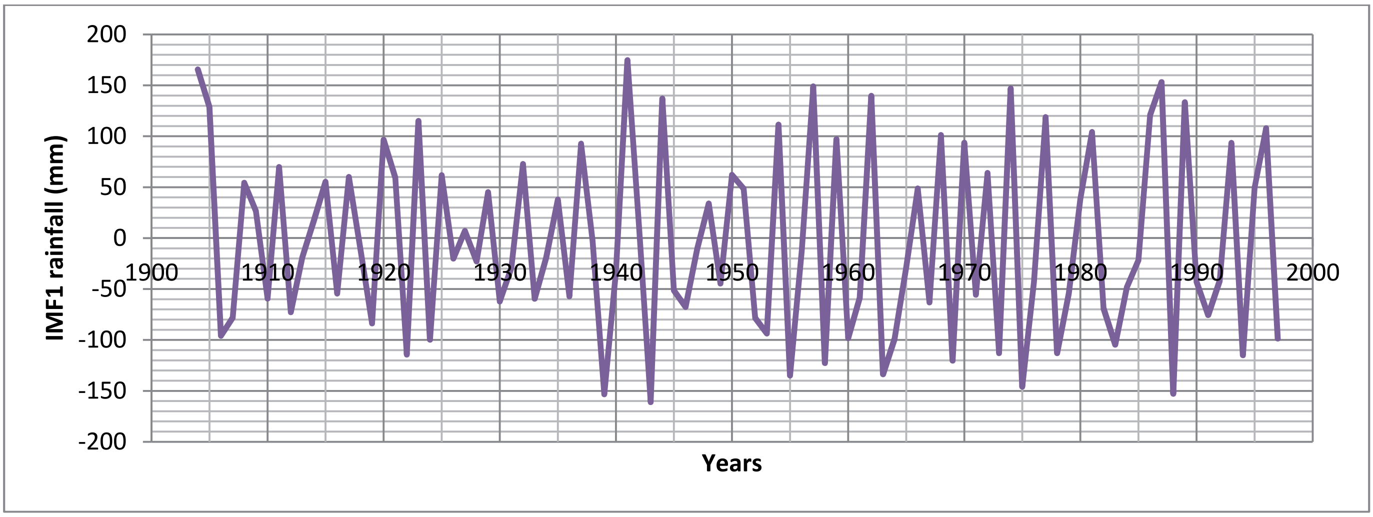


Figure A2- Intrinsic mode function (IMF1) of the decomposed time series for station 0020866W

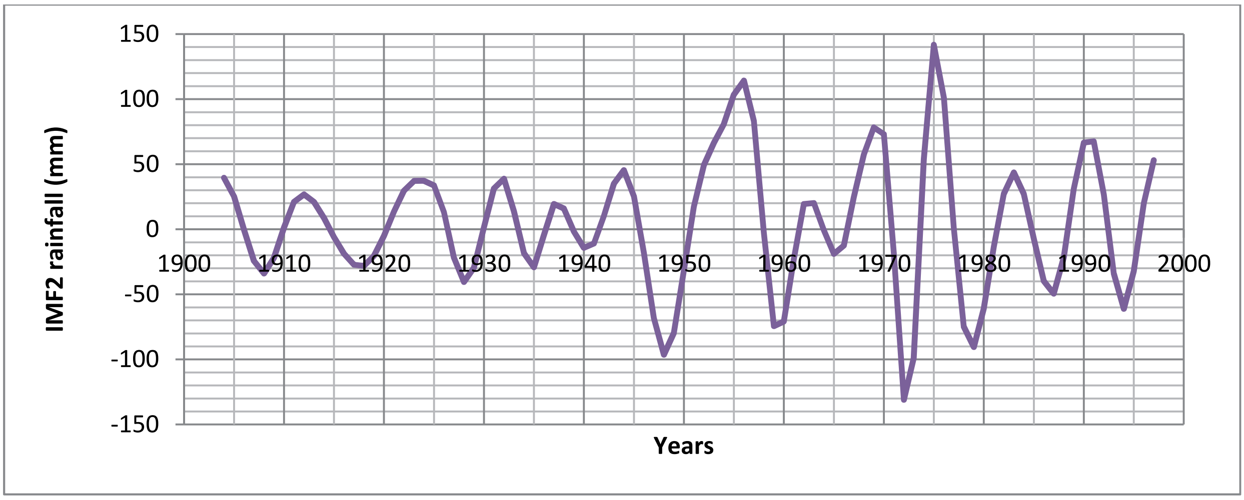


Figure A3- Intrinsic mode function IMF2 of the decomposed times series for station 0020866W

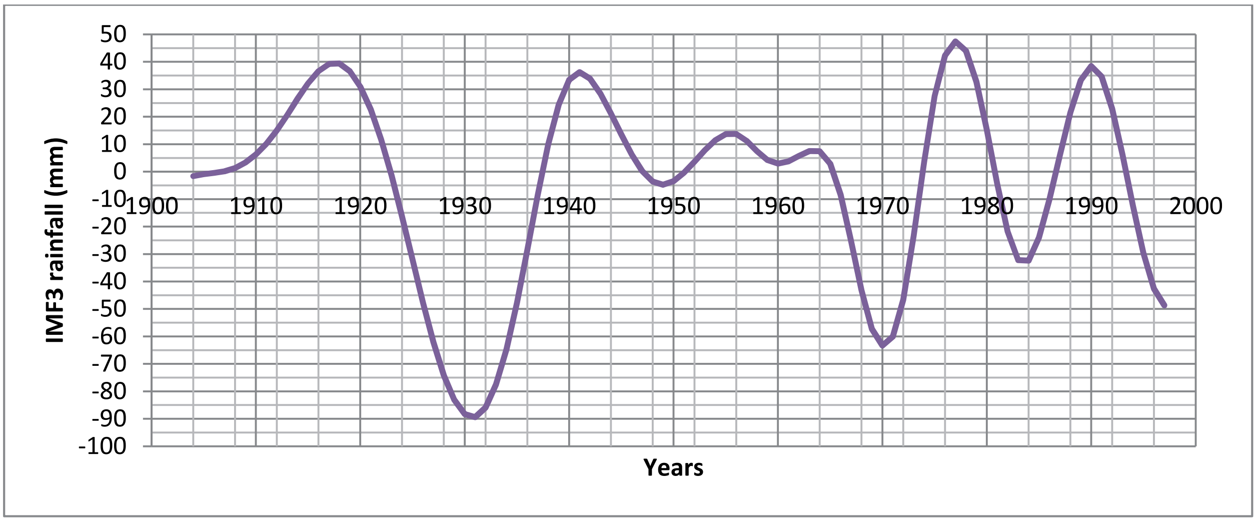


Figure A4- Intrinsic mode function IMF3 of the decomposed time series for station 0020866W

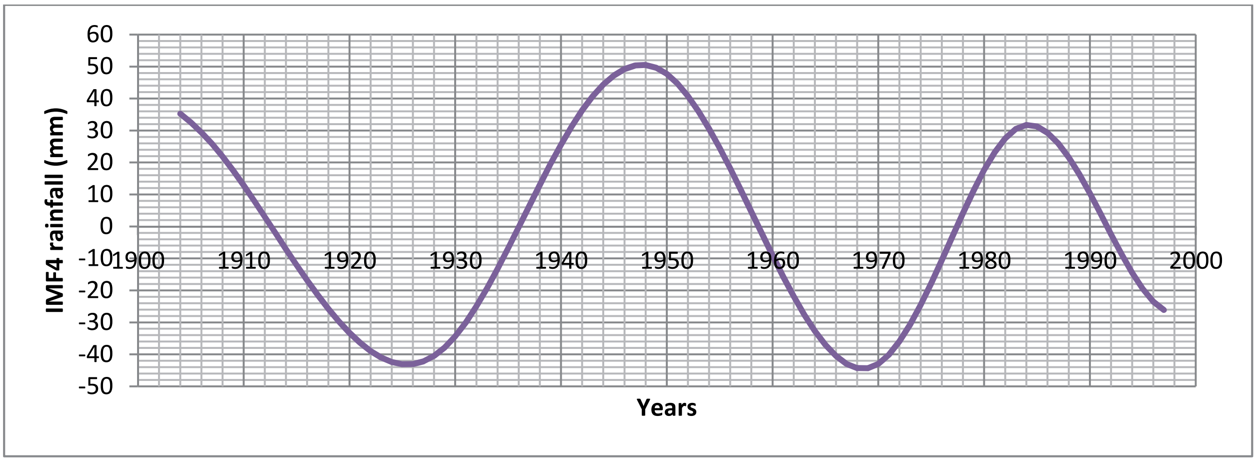


Figure A5- Intrinsic mode function IMF4 of the decomposed time series for station 0020866W

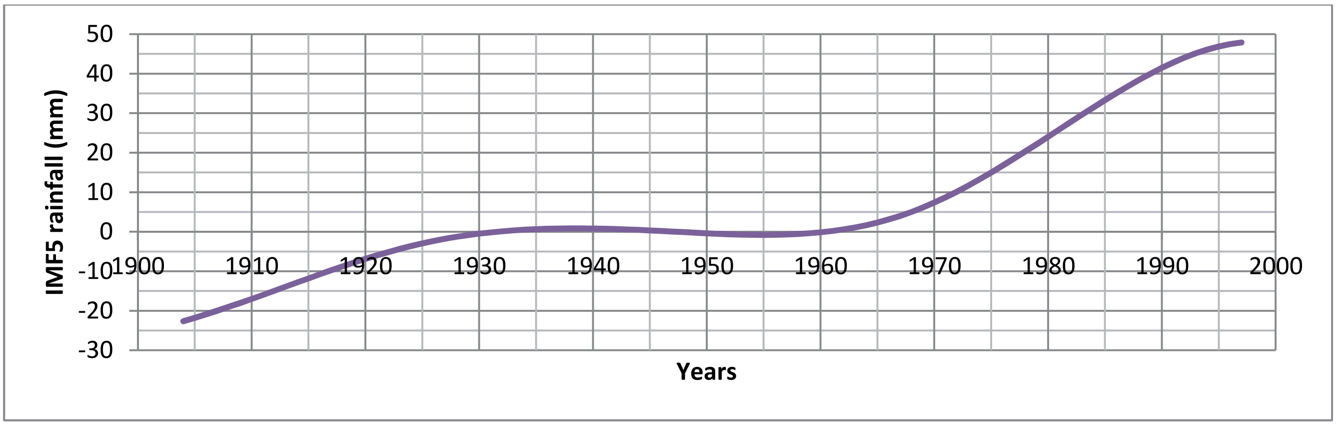


Figure A6- Intrinsic mode function IMF5of the decomposed time series of station 0020866W

.

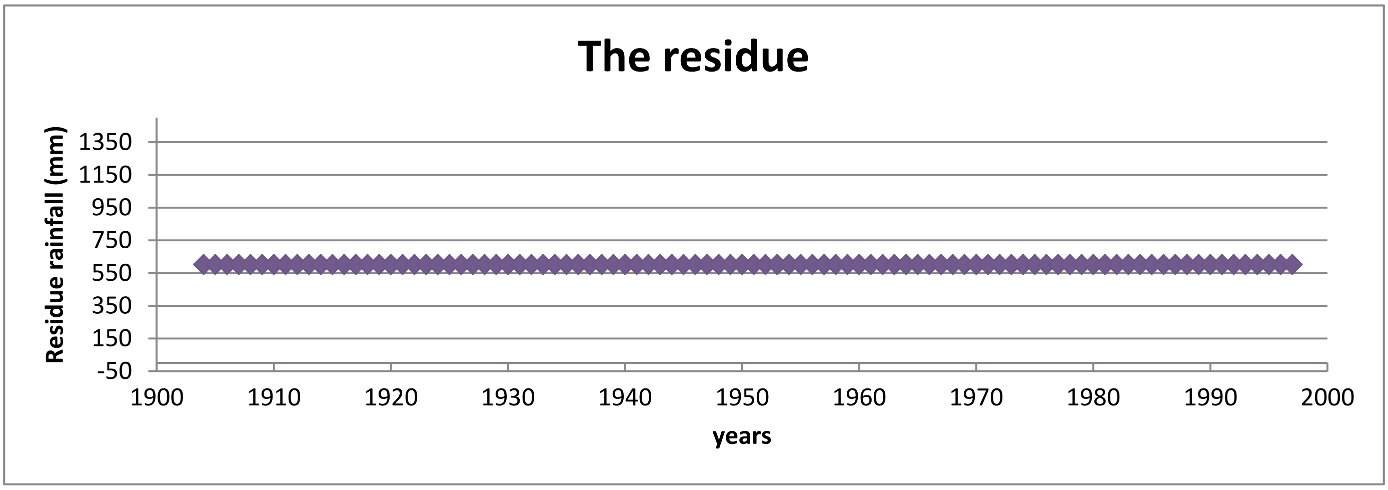


Figure A7 The residual trend of the times series from station 0020866W

2 Generation of blocks

The minima and maxima of all the IMFs are used as block termination locations as illustrated on Figure A8. This is based on the expectation that EMD identifies locations at which the time series behaviour changes at various time scales. The resulting blocks therefore have multiple lengths and these create the population of blocks for random resampling by the VLB generator. The selection of the IMF to obtain a block from is done in inverse proportion to the average length of the blocks of the IMF because longer blocks take up longer lengths than shorter ones. It is important to note that EMD is being used to only locate block termination and the actual blocks are generated from the historic time series and not the IMFs.

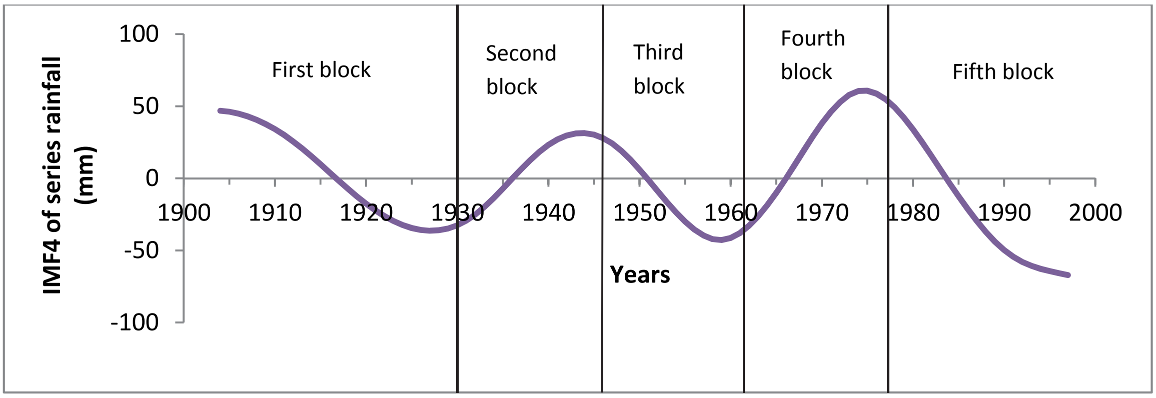


Figure A8 shows the method adopted to identify the locations to terminate the

Based on this method, the IMF1 to IMF5 obtained for station 0020866 W (Figures A2 to A6) would respectively produce 65, 24, 12, 5 and 3 blocks each resulting in a total of 109 blocks of a large variety of lengths.

This is done for all the rainfall stations resulting in a large number of block terminations. Just as in the original VLB, the different stations are alternately used as lead sequences in order to preserve cross correlation among the stations.

3 Preliminary Results obtained using the EMD-VLB Hybrid.

Figure A9 shows box plots of the mean annual rainfall showing reasonable generation except for one rainfall station

Figure A10 shows a typical box plot of minimum run sums that shows reasonable validation of the approach.

The observed large variability from these box plots is an interesting preliminary result as it considerably larger than those obtained with the VLB and PEGRAIM-W (see Chapter 4).

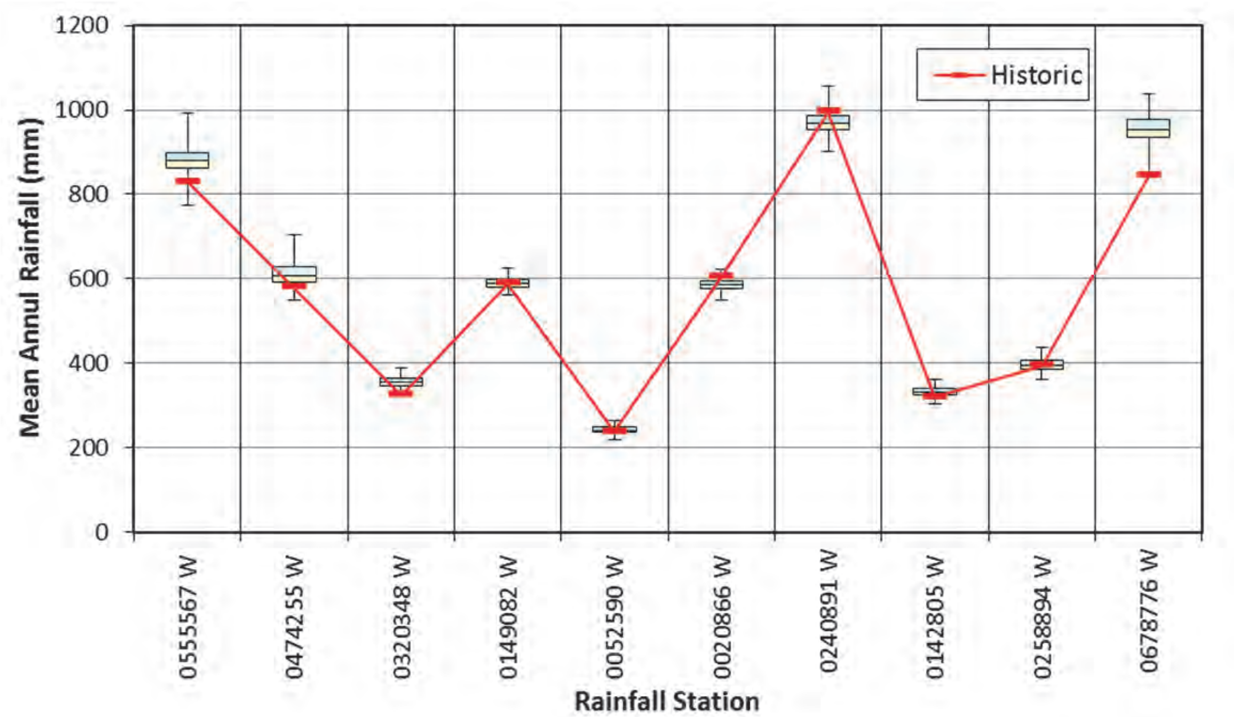


Figure A9 Box plots of mean annual rainfall from hybrid EMD- VLB model

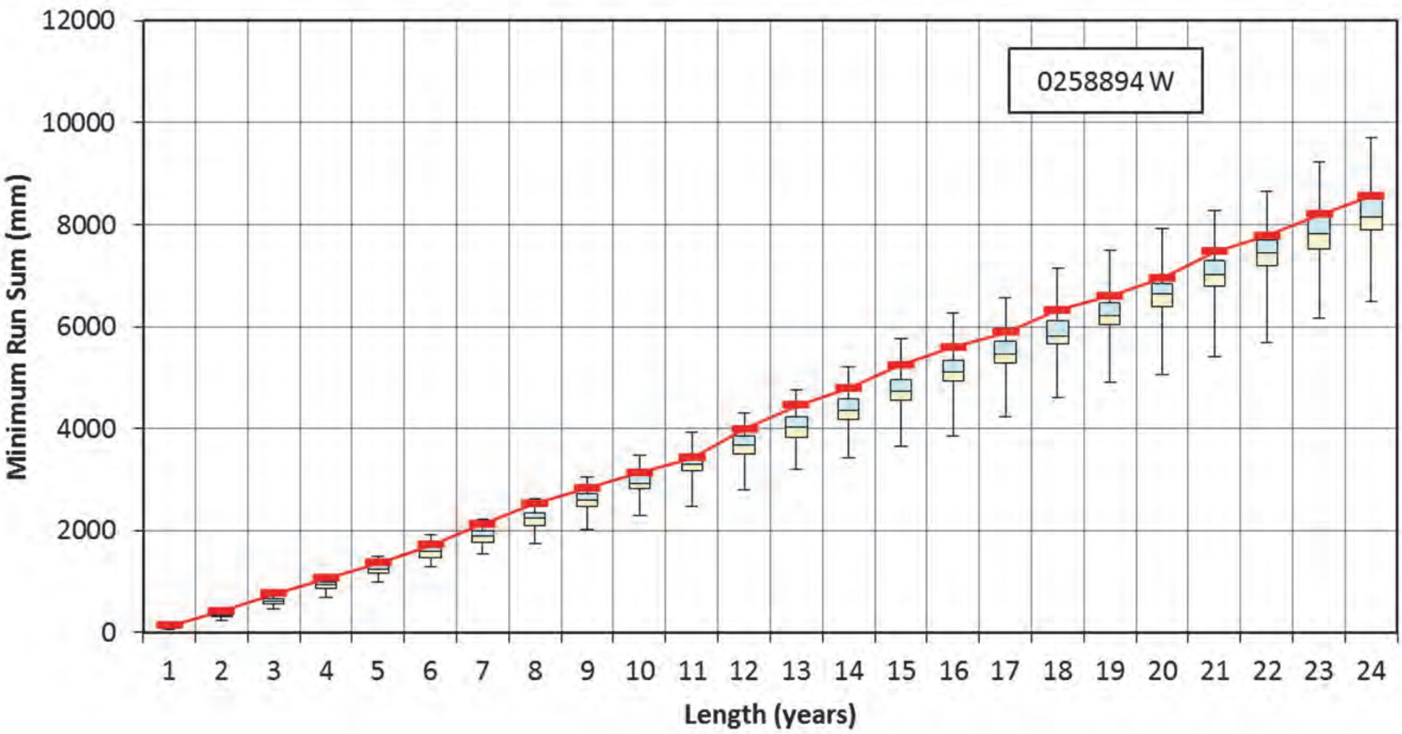


Figure A10 Box plots of minimum run sums of rainfall from hybrid EMD- VLB model