Annual Peak Rainfall Data Augmentation – A Bayesian Joint Probability Approach for Catchments in Lesotho

Khahiso Kanetsi

A research report submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, in fulfilment of the requirements for the degree of Master of Science in Engineering

Johannesburg, 2017

DECLARATION

I declare that this research report is my own unaided work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other University.

____

Signature of Candidate

___________________ day of _________________________ 20___________________

ABSTRACT

The main problem to be investigated is how short duration data records can be augmented using existing data from nearby catchments with data with long periods of record.

The purpose of the investigation is to establish a method of improving hydrological data using data from a gauged catchment to improve data from an ungauged catchment. The investigation is undertaken using rainfall data for catchments in Lesotho.

Marginal distributions describing the annual maximum rainfall for the catchments, and a joint distribution of pairs of catchments were established. The parameters of these distributions were estimated using the Bayesian – Markov Chain Monte Carlo approach, and using both the single-site (univariate) estimation and the two-site (bivariate) estimations.

The results of the analyses show that for catchments with data with short periods of record, the precision of the estimated location and scale parameters improved when the estimates were carried out using the two-site (bivariate) method. Rainfall events predicted using bivariate analyses parameters were generally higher than the univariate analyses parameters.

From the results, it can be concluded that the two-site approach can be used to improve the precision of the rainfall predictions for catchments with data with short periods of record. This method can be used in practice by hydrologists and design engineers to enhance available data for use in designs and assessments.

In memory of my grandmother

‘Mamotebang Rosina Lebokollane

1931 – 2015

ACKNOWLEDGEMENTS

I could not have finished this work if it was not for the grace of God.

I owe an inexpressible debt to my academic supervisor – Professor John Ndiritu – whose patience and academic acumen were the driving force behind my completion of this work.

A special thank you is also extended to the University of the Witwatersrand, in particular the academic and support staff in the School of Civil and Environmental Engineering and the Faculty of Engineering and Built Environment. I would like to express gratitude to Mafube Consulting (PTY) Ltd, Lesotho Meteorological Services and the Water Commission, Lesotho for assistance with information and rainfall data.

It would clearly be remiss of me if I would forget the unwavering support of my son Mr. Teboho Kanetsi who has been very understanding and forgiving of my lack of parental presence and affection on many occasions. I must also express my deepest gratitude to my parents, Mr. Khotso Kanetsi and Mrs. ‘Makhahiso Kanetsi, for instilling in me a sense of direction and resilience, and for their continuous support and encouragement throughout my years of study.

Thank you.

CONTENTS

LISTS OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

In this section, a brief background of Lesotho, the country's meteorology and rainfall data is outlined. The problem statement, hypothesis, purpose, scope and plan of the study are also outlined.

1.1. Background

The Kingdom of Lesotho is a landlocked country, completely surrounded by three provinces of the Republic of South Africa. The country covers an area of approximately 30 300 square kilometres (km^2) with altitude varying from 1500 m to 3482 m above sea level; approximately 25% of the area is in the lowlands while the remaining 75% is in the highlands. The population of Lesotho is just over 1.9 million people.

The headwaters of two of the largest rivers in Southern Africa, the Orange (Senqu) River and the Tugela River are located in the highlands of Lesotho and the country is located within the Orange-Senqu River Basin. The three major catchments in the country are the Senqu catchment, the Makhaleng catchment and the Mohokare catchment. The mean annual rainfall in Lesotho is just under 800 mm and approximately 85% of the total rainfall can fall between October and April.

The Ministry of Energy and Meteorology and the Ministry of Water are responsible for the management of water resources data through establishments such as the Lesotho Meteorological Services (LMS) and the Department of Water Affairs (DWA). The water resources data collection network comprises climate stations, surface water gauging stations, ground water stations and wetlands discharge stations.

There are 50 climate stations, 8 of which measure precipitation only, while the other 42 also measure other parameters such as soil moisture, temperature and wind speed. The LMS has a number of challenges with regard to climate data collection and one of the challenges is access to stations in remote areas. One of the possible improvements that they look forward to is the construction of automatic climate stations in remote areas (The Office of the Commissioner of Water, 2013).

Rainfall data is used by hydrologists and engineers to estimate extreme rainfall events such as floods for purposes of design of hydraulic structures in order to limit risk of damage to property and loss of life. This is challenging when the available hydrological data are limited, have significant gaps or when data in the relevant catchment area are not available. Unavailability of data can result in ad hoc methods of estimating rainfall events which can result in inaccurate and inconsistent estimations of hydrological events and hydraulic designs which consequently have negative economic, social and environmental impacts.

There is a need for improved data records in order to reduce uncertainties in event predictions; therefore it is necessary to explore methods in which the existing data can be enhanced such that more reliable hydrological event predictions and hydraulic designs can be prepared.

1.2. Problems Statement

1.2.1. Description of the Main Problems to be investigated

The main problem to be investigated is how data with short periods of record can be augmented using statistical formulations and existing data from other nearby catchments with longer periods of data record.

In the context of catchments in Lesotho, there are climate stations in place to collect rainfall data; however some of the stations are currently not operational, some can go for years without operating and others may be inaccessible from time to time. This means the data that is collected varies in record length and often has significant gaps thereby increasing the likelihood of inaccurate and distorted estimates of hydrological patterns, erroneous designs by engineers and flawed decisions by planners.

1.2.2. Significance of the Specified Problems

Rainfall-based flood estimation methods make use of historical rainfall data. This data is often analysed using frequency analysis, which requires a relatively long periods of record (Chadwick et. al., 2004:313).

The problem with data that has short periods of record is that they can mislead the user when used in statistical analysis. This problem is worth investigating because data plays a vital role in hydrological designs and assessments. For a designer, poor quality data or data with short periods of record could result in inaccurate estimates and incorrect decisions which may do more harm than good, often resulting in undesirable economic, environmental and social impacts e.g. frequent flooding due to under-designed stormwater drainage systems, loss of lives and property etc.

A solution to the problems stated above will give designers and planners the benefit of carrying out more accurate assessments and designs despite current challenges in the data collection network.

1.3. Hypotheses to be investigated

Hydrological variables are highly correlated in space i.e. a relationship exists between two variables that are taken at two different points in space.

If such relationships exist, then data from one catchment can be correlated to data from a nearby catchment and if this relationship can be established, it can be used jointly to augment hydrological data for an ungauged[1] catchment using hydrological data from a gauged[2] catchment.

[1] An ungauged catchment may be a catchment with short periods of data record, records with gaps or no records at all.

[2] A gauged catchment may be a catchment with long periods of data record.

1.4. Purpose of the Study

The purpose of the investigation is to establish a method of improving hydrological data using data from a gauged catchment to improve data from an ungauged catchment.

This method can be used in practice by hydrologists and design engineers to enhance available data for use in designs and assessments in order to reduce uncertainties caused by insufficient or poor quality data. This would ensure improved accuracy in designs and hence better informed decisions.

1.5. Scope and Limits of the Study

The investigation was undertaken using rainfall data for catchments in Lesotho. The data was primarily sourced from Lesotho Meteorological Services.

Data with short periods of record (from what is considered as ungauged catchments or catchments with poor quality data) and data with long periods of record (from what is considered as gauged catchments or catchments with good quality data) was used in the analysis. Focus was on pairs of adjacent catchments where one catchment had poor quality data and the other had good quality data.

Statistical methods were used to analyse the data and the aim of the analysis was to determine whether or not there was a possibility of improving poor quality data from one catchment using good quality data from an adjacent catchment.

1.6. The Plan of Development of the Research Report

In the chapters to follow, the theories that form the basis of the studies previously undertaken by other researchers are outlined in the literature review. This is followed by a detailed description of the research procedure (methodology) and a factual account of the results and findings. Finally, the results are discussed, conclusions drawn and recommendations made.

2. LITERATURE REVIEW

This literature review outlines findings from research done and reports published on the subject of maximising the value of available data with the purpose of generating improved predictions, in line with the study's problem statement and hypothesis. The review covers the importance of catchment hydrological data and the possibility of transfer of information between catchments. In an effort to establish a suitable approach and methodology for the study, flood prediction techniques, including widely used probability distribution models and methods of estimating parameters of the probability distribution models, are also discussed. The Bayesian Markov Chain Monte Carlo (MCMC) method of estimating parameters is scrutinized, particularly in light of univariate and multivariate probability distributions, copula functions and the frequency analysis.

"The IAHS Decade on Predictions in Ungauged Basins (PUB) is an initiative of the International Association of Hydrological Sciences (IAHS). It is aimed at formulating and implementing appropriate science programs to engage and energise the scientific community in a coordinated manner, towards achieving major advances in the capacity to make reliable predictions in ungauged basins" (Sivapalan et. al., 2003:860). One of the key science questions for PUB is "*How can we maximise the scientific value of available data in generating improved predictions?"*, which is in line with one of PUB's scientific objectives of increasing the awareness of the value of data for water resources management. This study's problem statement is aligned to PUB's initiative in that there is a need to determine how data with short periods of record can be augmented using statistical formulation and existing data from other nearby catchments with data that has longer periods of record.

The study seeks to investigate a hypothesis that hydrological variables are highly correlated in space; that a relationship exists between two variables that are taken at two different points in space. This means that data from one catchment can be correlated to data from a nearby catchment, i.e. hydrological data can be used jointly to augment data with short periods of record. Chadwick et. al. (2004:309) explain that there are two types of catchments; gauged and ungauged catchments and that for ungauged catchments, transferring data from a nearby donor gauged catchment or a catchment with similar catchment descriptors to the subject site is recommended (Chadwick et. al., 2004:328).

The purpose of this study is to establish a method of improving hydrological data using data from a gauged catchment to improve data from an ungauged catchment. Sivapalan et. al. (2003:860) defines an ungauged basin as one with insufficient (in quality and quantity) records of hydrological observations to enable computation of hydrological variables of interest at the appropriate spatial and temporal scales, and to the accuracy acceptable for practical applications. Sivapalan et. al. (2003:859) highlights the importance of data by explaining that decisions required for preventing and managing natural disasters can only be made with the widest possible information being made available based on accurate and reliable predictions. It is further highlighted that most widely used predictive tools are data driven, and their application is based on the principle that the past is a reasonable guide to the future and that data from any one basin is a useful guide to estimating hydrological responses at another basin.Shortage of data decreases the reliability of design values and increases the risks to engineering works. Therefore hydrologists are looking for ways data from various sources can be used together to ensure more reliable estimates (Liang, et. al.,2011).

With the need for improved data identified, it is important to establish appropriate methods that can be used to achieve this improvement. According to Hrachowitz et. al. (2013:1201), the primary research objectives of the PUB initiative were to improve the ability of existing hydrological models to predict in ungauged basins with reduced uncertainty and to develop new innovative models representing the space-time variability of hydrological processes, thus improving the confidence in predictions in ungauged basins. Although PUB is a very recent initiative, the need for better quality data was recognized much earlier than that. Matalas and Jacobs (1964:E1) investigated a correlation procedure for augmentation of hydrological data where a linear regression for a short and long sequence of hydrological events was used to lengthen the short sequence. Similarly, Vogel and Kroll (1990:259) explored how streamflow record augmentation procedures made use of the cross-correlation between stream flows at two or more stream gauges to obtain improved estimates of the mean and variance of the flows at short-record gauge. The findings of their investigation were that the gain resulting from record augmentation procedures represented a 39-62% increase in total information. The key issues from these investigations were the use of correlation procedures for hydrological data augmentation and how the mean and variance could be used as indicators of the accuracy of a record. There has been a shift from reliance on generalised regression equations to applications of techniques for transferring hydrological data from gauged to ungauged catchments (Chadwick et. al., 2004:309). From the above findings, it can be said that data augmentation results in an increase in total information and improved estimates of parameters such as mean and variance.

As this study looks at data augmentation with the purpose of improving flood predictions, the review focuses on techniques and probability distribution models used to analyse rainfall data in flood prediction. According to Chadwick, et al. (2004:309) there are two types of flood prediction techniques, namely statistical methods (e.g. frequency analysis) and the unit hydrograph rainfall-runoff model. In engineering practice, the objective of a hydrological frequency analysis is to provide design values with an assigned return period (Liang et. al., 2011:1183). The frequency analysis approach is commonly used by design engineers in Lesotho to estimate rainfall or river flow design values for specific return periods (e.g. 2, 5, 10, 20, 50 and 100 years). This technique is often used when designing stormwater structures, and even though most designers are aware that the shorter the record the less accurate the estimations, this technique is still used in its conventional form.

One of the key decisions required before undertaking hydrological frequency analyses is selection of a suitable probability distribution for the data to be analysed. Chung and Kim (2013) explain that in rainfall frequency analysis, a probability distribution should be selected to estimate the value for the given return period and that the most widely used models are the Gumbel distribution and the GEV distribution. In their comparative study with the Gumbel and the GEV distribution performed to evaluate the efficiency of the GEV, they found that using the GEV distribution could be applied effectively to the rainfall frequency analysis. With these findings in mind, the GEV can be considered for use as the probability distribution of the selected data in this study. Its suitability would however need to be assessed using goodness of fit tests.

Chung and Kim (2013) further explain that once a distribution is selected to fit the rainfall data, the parameters of the selected distribution should be estimated using parametric methods such as Maximum Likelihood Estimation (MLE) and L-Moments. A comparison of the Bayesian MCMC method and the MLE method for parameter estimation, Chung and Kim (2013) found that the Bayesian MCMC had no advantage over the MLE but, with respect to uncertainty analysis, the Bayesian MCMC could significantly reduce the range of uncertainty.

Using regional floods data with at-site flood extremes to estimate parameters of a distribution is a frequently adopted approach and the Bayesian formula provides a theoretical framework for combining both prior and sample information (Liang et. al., 2011:1191). In their application of the Bayesian approach to hydrological frequency analysis, Liang, et. al. (2011) applied a Bayesian frequency analysis method to estimate parameters for a Pearson type III probability distribution. They used the Adaptive Metropolis (AM) sampling technique to generate parameter sets from Bayesian parameter posterior distributions.

Bayesian methods allow a richer and more complete representation of flood records and information together with their uncertainty (Reis and Stedinger, 2005:97). Reis and Stedinger (2005) employ the Bayesian MCMC method to derive empirical approximations of the posterior distribution of parameters of a log-normal distribution and a log-Pearson type 3 distribution. They compare the Bayesian-MCMC approach to three other methods (Adjusted-moment estimator, Expected moment estimator and the Maximum likelihood estimator) used to estimate parameters of a distribution. The

Bayesian method was found to have an advantage over the other methods, particularly the adjusted-moments and the expected-moments estimators, as it used the full likelihood function which was an effective way to represent information for a site.

The use of the MCMC technique is also illustrated in Campbell et. al.'s (1999) where Bayesian procedures and the MCMC were used for parameter estimation in a Bayesian framework. Campbell et. al. (1999) present the Bayesian procedure for parameter estimation in a non-linear flood event model. In their approach, a large sample of the posterior distribution is generated using the MCMC method. They argue that an MCMC approach based on a general Metropolis-Hastings has more modelling flexibility and is easier to code than the Gibbs sampler.

Gaume, et. al. (2010) proposed the likelihood formulation and a Bayesian MCMC algorithm to infer parameter values of a regional distribution. The method is applied to two cases and the results were reported to show a potential of using the Bayesian framework with unconventional data.

Wang (2001:1707) introduced a procedure for augmenting flood peak records by adopting a bivariate extreme value distribution for data at two gauging stations and using the Bayesian framework to formulate and solve the inference problem. Wang (2001:1708) employs the Markov Chain Monte Carlo (MCMC) simulation technique, using the Metropolis-Hasting algorithm, to draw a sample of parameters from the distribution. Data from the two stations has different record lengths and information is transferred between the two data sets to improve the statistical parameters that can be derived from the distribution. Findings from Wang's approach revealed that when using, jointly, all the available data at the two stations, the two-site approach proved to be highly efficient.

Escalante-Sandoval (2006) proposed the use of the Logistic model for bivariate extreme value distributions for the case of flood frequency analysis. Using the maximum likelihood method to estimate the parameters, there was a significant improvement when the parameters were estimated using the bivariate distribution instead of the univariate distribution.

Because adopting a multivariate distribution approach requires a mathematical function that describes the dependence of the different data sets, copula functions have been explored. A copula function is a mathematical technique which offers a flexible way of describing nonlinear dependence among multivariate data in isolation from their marginal probability distributions (Mitková, et. al. 2014). Balakrishnan and Lai (2009) define copulas as functions that join multivariate distributions to their one—dimensional marginal distribution functions.

According to Mitková (2014), commonly used copulas include Elliptical copulas (e.g. Normal, Gaussian and Student-t copulas), Archimedean copulas (e.g. Clayton, Gumbel-Hougaard and Frank copulas) and Extreme-Value copulas (e.g. Husler-Reiss, Galambos, Tawn and t-EV copulas).The Archimedean copula is the popular class, used in hydrological application (De Michele et. al., 2005) mainly because of its flexibility and easy construction.

Yang, et. al. (2013) investigated the joint probability distribution of wind speed and significant wave height in the Bohai Bay by comparing the Gumbel logistic model, the Gumbel-Hougaard (GH) copula function and the Clayton copula function. Their main findings were that the GH copula function was optimal and the joint design values for the wave height and wind speeds were larger than the marginal values.

Xu, et. al. (2015) carried out a bivariate hydrological risk analysis framework based on the bivariate flood frequency analysis through copula methods and one of their conclusions was that bivariate hydrological risk values (characterised based on the joint return period) can provide decision support for hydraulic facility design as well as actual flood control and mitigation.

From this literature review, it is evident that researchers have been investigating methods through which hydrological data records can be improved using probability methods such as the Bayesian formulation. In most of the literature reviewed, the Bayesian MCMC approach is used for runoff data and not for rainfall data.

The Bayesian framework can be used to infer model parameters and a sample of these parameters can be drawn using the MCMC technique, with sampling algorithms such as the Metropolis-Hasting algorithm, the Adaptive Metropolis sampling technique and the Gibbs sampler. There seems to be an overall consensus that the advantages of the Bayesian approach are that it makes use of prior information (based on expert knowledge, experience or historical data) and sample information to infer parameters for probability distributions and allows modelling of the estimation uncertainties.

Some research has also been undertaken to determine the advantages of augmentation of data records with focus on bivariate analyses versus univariate analyses. The basis of comparing single (univariate) and joint (bivariate) analyses is the precision of the resulting estimation of parameters of the assumed distributions (as outlined in the chapters to follow).

3. METHODOLOGY

The main objective of the study was find out if the estimation of extreme daily rainfall in a catchment with a short data period could be improved by using data from a catchment with long periods of record. The specific objectives of the study were: i) to establish suitable statistical methods for univariate and bivariate analyses of the data, ii) to compare the results from univariate and bivariate analyses and iii) to determine if the bivariate approach would lead to improved extreme rainfall estimation . These specific objectives would assist in testing the hypothesis that data could be used jointly to augment hydrological data for an ungauged catchment using hydrological data from a gauged catchment, and therefore meet the main objective of the study. In the literature review it was shown that hydrological data could be used jointly to augment data with short periods of record and that there exists techniques and models for carrying out the augmentation. The methodology in this section outlines: i) the selection of data for the study, ii) selection and description of the statistical methods used, iii) the estimation of model parameters for use in estimating extreme rainfalls and iv) the approach taken to compare univariate and bivariate analysis and to determine whether data with long periods of record has improved the quality of extreme rainfall estimate in the ungauged catchment.

Annual maximum rainfall data was collected from selected catchments, then statistical methods were used to determine how hydrological data could be improved through transfer of information between two catchments.

Rainfall data from three (3) catchments was obtained from the Lesotho Meteorological Services. Pairs of catchments and their corresponding rainfall data were selected for use in the investigation such that:

i. Catchment (Rainfall Station) 1 had a long period data record
ii. Catchment (Rainfall Station) 2 had a short period data record
iii. The catchments are in the same region and adjacent to each other.

The marginal distributions describing the annual maximum rainfall for Catchments 1 and 2, and a joint distribution of both catchments were established. The parameters of these distributions were estimated using the Bayesian – MCMC approach, which is outlined in Section 3.3 below. For each catchment, statistical location and scale parameters were determined using both the single-site (univariate) estimations (where data from each catchment was used separately) and the two-site (bivariate) estimations (where data from the two catchments combined was used). These parameters characterize the population (annual maximum daily rainfall) as they are used to determine the probability density functions of the sample variables, therefore enabling the estimation of the occurrence of a hydrological event. A comparison was made between the univariate analyses parameter values and the bivariate analyses parameter values in order to establish if the bivariate analyses parameter values had higher precision than the univariate analyses values.

Two pairs of data were analysed; both pairs comprised a long period of data set and a short period of data set.

The approach that was used in this investigation is similar to Wang's (2001) approach to flood record augmentation where he used annual maximum discharge values. The choice to use rainfall data instead of river flow data in this investigation was based on the fact that the design of hydraulic structures used flows obtained by models that use extreme rainfall data.

3.1. Probability Models – Marginal and Joint Probability Distributions

3.1.1. The Marginal Probability Distribution: Generalised Extreme Value (GEV) Distribution

The generalised extreme value (GEV) probability distribution is a continuous probability distribution usually used to model the maxima or high points of a sequence of random variables (Chow, 1988). The Kolmogorov-Smirnov goodness of fit test was used to test the GEV as a hypothesised distribution for annual daily maximum rainfall for the selected catchments. Using the test, the deviation of the observed cumulative probability from the hypothesised (GEV) cumulative distribution function was determined, and the hypothesis was accepted for two (2) out of three (3) data sets. This probability distribution was found suitable and was selected for use in modelling annual daily maximum rainfall data as this type of data was a composition of rainfall maxima data. Equation 3-1 defines the GEV cumulative distribution function (F(x)) and Equations 3-2 and 3-3 define the GEV probability density function (f(x)).

$$F(x) = \begin{cases} exp\left\{-\left[1+\left(\frac{x-\mu}{\sigma}\right)\xi\right]^{-1/\xi}\right\} for\ \xi \neq 0 \\ exp\left[-exp-\left(\frac{x-\mu}{\sigma}\right)\right] for\ \xi = 0 \end{cases} \tag{3-1}$$

$$f(x) = \begin{cases} F(x)\ \frac{1}{\sigma}\left[1+\left(\frac{x-\mu}{\sigma}\right)\xi\right]^{-\frac{1}{\xi}-1}\ for\ \xi \neq 0 \\ F(x)\ \frac{1}{\sigma}\left[exp-\left(\frac{x-\mu}{\sigma}\right)\right] for\ \xi = 0 \end{cases} \tag{3-2}$$

$$f(x) = \begin{cases} exp\left\{-\left[1+\left(\frac{x-\mu}{\sigma}\right)\xi\right]^{-\frac{1}{\xi}}\right\}\frac{1}{\sigma}\left[1+\left(\frac{x-\mu}{\sigma}\right)\xi\right]^{-\frac{1}{\xi}-1}\ for\ \xi \neq 0 \\ exp\left[-exp-\left(\frac{x-\mu}{\sigma}\right)\right]\frac{1}{\sigma}\left[exp-\left(\frac{x-\mu}{\sigma}\right)\right] for\ \xi = 0 \end{cases} \tag{3-3}$$

where μ is the Location parameter, σ is the Scale parameter, ξ is the Shape parameter and x is the rainfall maxima data point (Chow, 1988).

The collective symbol for all three parameters (μ, σ *and* ξ) is θ.

The Type I (Gumbel) extreme value distribution ($\xi = 0$) was chosen as the marginal distribution mainly because this distribution is useful in predicting the chance that an extreme event will occur, and unlike the generalised extreme value (GEV) distribution, the Gumbel extreme value distribution had fewer parameters and was therefore more parsimonious. Equations 3-4 to 3-6 describe this distribution.

$$F(x) = exp\left[-exp - \left(\frac{x-\mu}{\sigma}\right)\right] \quad \text{3-4}$$

$$f(x) = F(x)\ \frac{1}{\sigma}\left[exp - \left(\frac{x-\mu}{\sigma}\right)\right] \quad \text{3-5}$$

$$f(x) = \ exp\left[-exp - \left(\frac{x-\mu}{\sigma}\right)\right]\frac{1}{\sigma}\left[exp - \left(\frac{x-\mu}{\sigma}\right)\right]\ for\ \xi = 0 \quad \text{3-6}$$

where μ is the Location parameter, σ is the Scale parameter and x is the rainfall maxima data point (Chow, 1988).

3.1.2. The Joint Probability Distribution

A bivariate distribution function was constructed using the Gumbel-Hougaard copula (Mitková, 2014)). The Gumbel-Hougaard copula forms part of the Archimedean copulas class, which were commonly used in hydrological applications because of their flexibility and easy construction. Equations 3-7 to 3-14 describe the formulation of the Gumbel-Hougaard copula for the analysis.

$$C(u,v) = exp\left\{-[(-\ln u)^m + (-\ln v)^m\ ]^{\frac{1}{m}}\right\} \quad \text{3-7}$$

where

C(u, v) = the copula

$u = F(x)$ = marginal distribution of rainfall maximum data at station 1

$v = F(y)$ = marginal distribution of rainfall maximum data at station 2

$m =$ generator of the copula (m = 1 when u and v are independent of each other and $m = \infty$ when u and v are dependent) (Mitková, 2014).

$$C(F(x), F(y)) = exp\left\{-[(-\ln F(x))^m + (-\ln F(y))^m]^{\frac{1}{m}}\right\} \qquad 3\text{-}8$$

Therefore, the bivariate cumulative distribution function and the bivariate probability density function are given by:

$$F(x, y) = exp\left\{-[(-\ln F(x))^m + (-\ln F(y))^m]^{\frac{1}{m}}\right\} \qquad 3\text{-}9$$

$$f(x, y) = \frac{f(x)f(y)}{F(x)F(y)} \times [-lnF(x)]^{m-1} \times [-lnF(y)]^{m-1}$$

$$\times\ exp\left\{-[(-\ln F(x))^m + (-\ln F(y))^m]^{\frac{1}{m}}\right\}$$

$$\times\ [(-\ln F(x))^m + (-\ln F(y))^m]^{\frac{1}{m^{-2}}}$$

$$\times\ \left\{(m-1) + [(-\ln F(x))^m + (-\ln F(y))^m]^{\frac{1}{m}}\right\} \qquad 3\text{-}10$$

where

$$F(x) = exp\left[-exp - \left(\frac{x-\mu_x}{\sigma_x}\right)\right] \qquad 3\text{-}11$$

$$f(x) = F(x)\ \frac{1}{\sigma}\left[exp - \left(\frac{x-\mu_x}{\sigma_x}\right)\right] \qquad 3\text{-}12$$

$$F(y) = exp\left[-exp - \left(\frac{y-\mu_y}{\sigma_y}\right)\right] \qquad 3\text{-}13$$

$$f(y) = F(y)\ \frac{1}{\sigma}\left[exp - \left(\frac{y-\mu_y}{\sigma_y}\right)\right] \qquad 3\text{-}14$$

3.2. Parameter estimation: Bayesian Formulation, MCMC Simulation and the Metropolis Hastings Algorithm

3.2.1 Bayesian Formulation

According to Baye's Theorem,

$$Posterior \propto Likelihood \times Prior \qquad \text{3-15}$$

where:

Posterior is the probability density of θ with given observations (sample) of *x:* $p(\theta|x)$*;*

Likelihood is the likelihood function of sample x (probability density function of *x)*, conditional on the parameters θ*.* This provides the chances of each value of θ having led to that observed value of *x:* $f(\mathrm{x}|\theta)$*;*

Prior is prior probability density of the parameters θ*,* prior to the analysis of sample *x.* This is used to incorporate information about the θ into the analysis: $p(\theta)$ (Box, 1973).

Therefore the Bayesian formula is given by:

$$p(\theta|x) = \ [f(x|\theta)p(\theta)]/\int[f(x|\theta)p(\theta)d\theta] \qquad \text{3-16}$$

The denominator $\int f(x|\theta)p(\theta)d\theta$ is the normalisation constant to obtain a unit area under the probability density function of $p(\theta|x)$ (Box, 1973).Without the normalisation constant, the posterior is unscaled and although it has all the shape information, it is not the exact posterior density.

3.2.2 Markov Chain Monte Carlo (MCMC)

The Markov Chain Monte Carlo approach is one of the methods that can be used to draw a sample from a posterior. The technique is based on stochastic simulation.

Simulation of values from the distributions can be undertaken using different techniques, like the Gibbs sampler, Metropolis-Hastings algorithm or a hybrid which is a combination of the Gibbs sampler and the Metropolis-Hastings algorithm methods.

Either the Gibbs sampling or the Metropolis-Hasting algorithm could have been used to simulate values of the unknown parameter (θ). The main difference between the two approaches is that the Gibbs sampler is automatic in that the transition kernel is formed by the full conditional distribution (Garmerman, 2006) while the proposal distribution and its properties (e.g. variance) for the transition kernel must be chosen. The Metropolis-Hasting was desirable in this case as it provides options; the transition kernel is user defined and it also employs a resampling technique such as the rejection method (Garmerman, 2006).

For this study, the Metropolis-Hastings algorithm was selected for use in simulating values of the unknown parameter (θ) from the posterior density function. The Metropolis-Hastings algorithm was used because it had more modelling flexibility that other samplers like the Gibbs sampler.

3.2.3 Metropolis-Hastings Algorithm

The algorithm was used to create a sequence of data points for θ which converge to the posterior distribution $p(\theta|x)$, forming a numerical representation of the posterior distribution.

The Metropolis-Hastings Algorithm consisted of the following steps:

1. Initialization: Arbitrary values of θ were chosen, such that $\theta^0 = (\mu^0, \sigma^0, \xi^{\ 0})$ for t=1;
2. A proposal distribution or transition matrix $q(\theta'|\theta^{t-1})$ was selected; this distribution was arbitrary and was used as a rule for iterative simulation of successive values in the chain;

3. Proposal: The next candidates for θ were generated using the proposal distribution;
4. Acceptance: an acceptance ratio was established which was used to decide whether or not to accept the proposed candidate:

$$r = \frac{p(\theta' | x)/q(\theta' | \theta^{t-1})}{p(\theta^{t-1} | x)/q(\theta^{t-1} | \theta')} \qquad \textit{3-17}$$

If $r \geq 1$, then the candidate was accepted ($\theta^t = \theta'$). Otherwise the candidate was rejected ($\theta^t = \theta^{t-1}$).

5. Steps 2 to 4 were repeated until the sequence of numbers converged to $p(\theta|x)$.

The generated data points represented the posterior distribution (Wang, 2001).

3.3. Methods of Selecting Sources of Data

The rainfall data used in this research was obtained from the Lesotho Meteorological Services (LMS). There was no method of selecting the historical source of data used as the LMS was the main custodian of rainfall data in Lesotho. Other establishments or institutions collect rainfall data for specific periods of time, usually to serve a specific project or study; therefore the length of data collected was usually limited.

The data was selected based on its length and the possibility of it being correlated with data from a neighbouring catchment; i.e. pairs of adjacent catchments and their corresponding rainfall data were selected such that the data for each pair could be correlated.

3.4. Data Collection

Rainfall data from three rainfall stations was collected from the Lesotho Meteorological Services. The rainfall stations were located at three adjacent catchment areas.

Figure 3-1 and Table 3-1 show the location of the rainfall stations, including coordinates and elevations, together with the rainfall stations' data period and length of data record for each set of data collected from the respective stations.

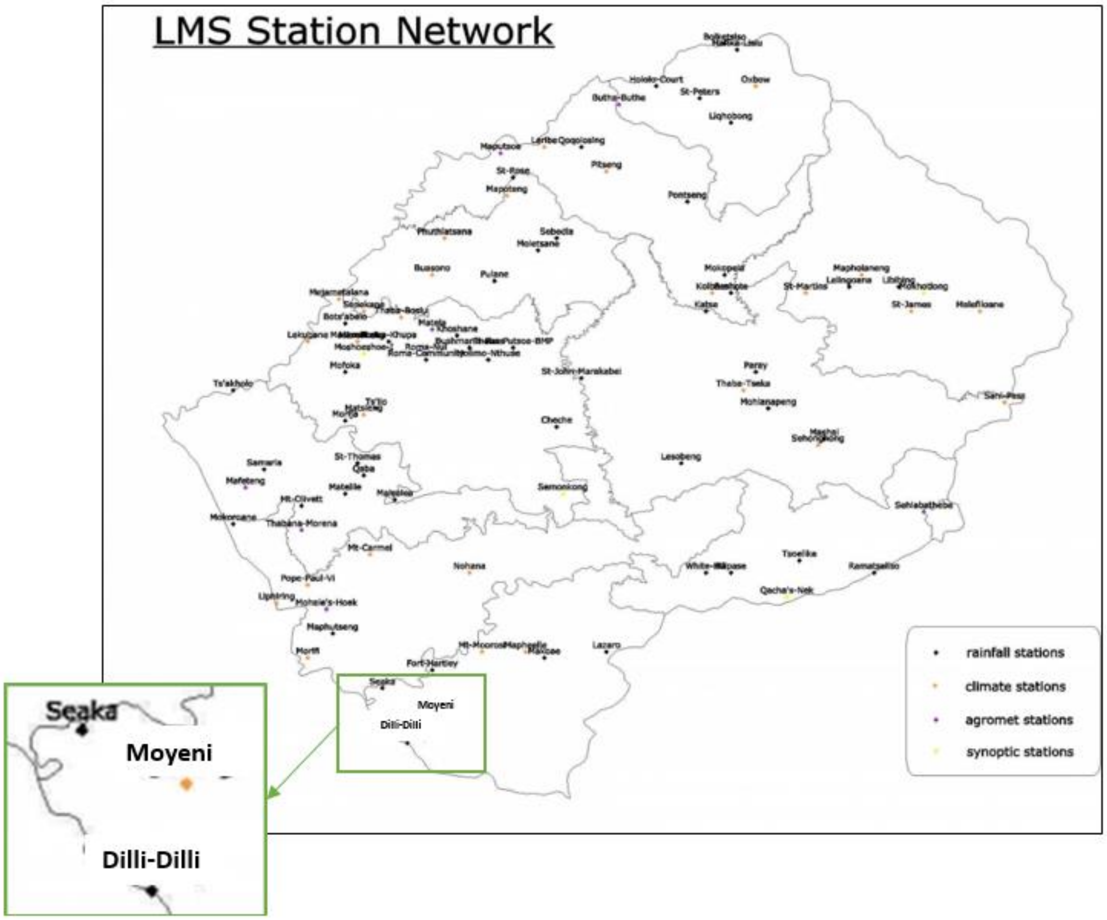


Figure 3-1: Location of Rainfall Stations: Dilli, Seaka and Moyeni. Image from Lesotho Meteorological Services © 2013, (http://www.lesmet.org.ls/sites/default/files/climatology/lms_stations)

Table 3-1: Rainfall Stations, data periods and data lengths for daily rainfall at Dilli-Dilli, Seaka and Moyeni

Rainfall Station/Catchment	Coordinates	Distance from Moyeni (km)	Elevation (m)	Data Period	Length of data record
Dilli-Dilli	30º 31´ 43.23´´ S 27º 40´ 14.02´´ E	20	1518	1997 – 2011	15 years
Seaka	30º 21´ 15.21´´ S 27º 36´ 27.83´´ E	15	1421	1978 – 2011	34 years
Moyeni	30º 24´ 01.12´´ S 27º 42´ 01.50´´ E	0	1518	1950 – 2012	63 years

From Figure 3-1 and Table 3-1, it can be seen that the data lengths of the catchments varied, with Moyeni having the longest record, and Dilli-Dilli having the shortest. Figure 3-2 below shows the data from the three rainfall stations.

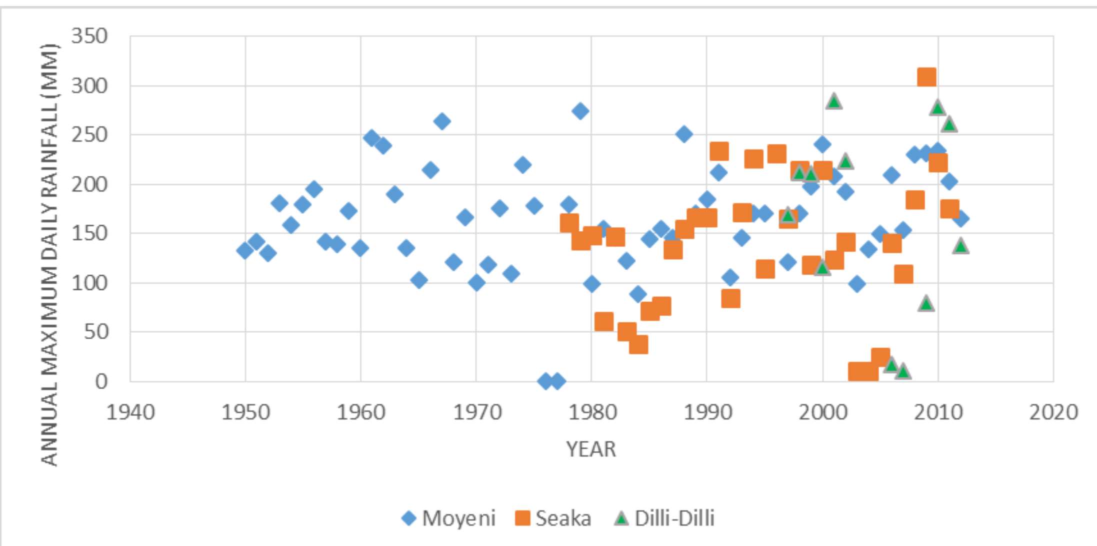


Figure 3-2: Annual Maxima daily rainfall for Dilli-Dilli, Moyeni and Seaka Rainfall Stations from 1950 to 2012

3.5. Data Analysis

Marginal probability distributions for each catchment were established and univariate analyses were carried out to estimate statistical location and scale parameters. Joint probability distributions for pairs of catchments were also established and bivariate analyses were carried out to estimate statistical locations and scale parameters. Parameters of the distributions were inferred using a spreadsheet (Microsoft ® Excel ® 2013), and the precision of the inferred parameters was determined based on the variability of their distribution. The resultant parameters were used in a frequency analysis to estimate probability distributions of rainfall events.

3.5.1 Bayesian Analysis in Microsoft ® Excel ® 2013

Microsoft ® Excel ® 2013 is a computer software manufactured by Microsoft Corporation, which, through use of a spreadsheet system, enables the use of formulae to format, organise and calculate data. The software has a variety of features, including statistical analyses and allowance for graphical representation of data. Steps I to IV inform how the analysis was set up.

I. Setting Up the Posterior Model

The posterior model included a likelihood model and prior distributions for the model parameters. It was the probability density of θ (parameters of the probability distribution of the rainfall data) with given observations (sample) of x, $p(\theta|x)$.

The likelihood function of sample x (probability density function of x), conditional on the parameters $\theta = (\mu, \sigma)$, was set up using Equation 3-6 for univariate analyses and 3-10 for the bivariate analyses. It is important to note that in the bivariate analyses, Equation 3-10 only used data for the years where there was a record for both rainfall stations in the pair. Data points for years where there is a record for one rainfall station only were incorporated into the bivariate analysis as follows:

$$f(x_k, y_k) \times f(x_i) \times f(y_j) \tag{3-18}$$

where

$x_k = data\ for\ Station\ 1\ for\ years\ where\ there\ is\ a\ record\ for\ both\ rainfall\ stations$

$y_k = data\ for\ Station\ 2\ for\ years\ where\ there\ is\ a\ record\ for\ both\ rainfall\ stations$

$x_i = data\ for\ Station\ 1\ for\ years\ where\ there\ is\ a\ record\ for\ \ rainfall\ station\ 1\ only$

$x_j = data\ for\ Station\ 2\ for\ years\ where\ there\ is\ a\ record\ for\ rainfall\ stations\ 2\ only$

The likelihood was determined using rainfall data (*x*) and the proposed values of μ and σ; i.e. for each iteration a different value of $f(x|\theta)$ was used. The prior probability density of the parameters θ, prior to the analysis of sample *x,* which in this analysis was given by a normal distribution with mean = 0 and variance = 100, is shown below:

$$p(\theta) = \frac{1}{\sqrt{(2\pi \times Variance)}}\ [\,e^{-\left((\theta - mean)^2 / (2\times Variance)\right)}] \tag{3-19}$$

Meaning

$$p(\mu) = \frac{1}{\sqrt{(2\pi \times Variance)}}\ [\,e^{-\left((\mu - mean)^2 / (2\times Variance)\right)}] \tag{3-20}$$

$$p(\sigma) = \frac{1}{\sqrt{(2\pi \times Variance)}}\ [\,e^{-\left((\sigma - mean)^2 / (2\times Variance)\right)}] \tag{3-21}$$

II. Performing MCMC Simulation

The random-walk Metropolis-Hastings (MH) algorithm was used to obtain Markov Chain Monte Carlo (MCMC) samples of model parameters.

A. Arbitrary values of $\theta\ (\mu\ and\ \sigma)$ were chosen and for the chosen set of $\mu\ and\ \sigma$, the prior probabilities and the likelihood of x were determined and used to calculate the posterior;

B. Random numbers ($\mathcal{E}$) were generated and used to calculate proposed values of θ for each iteration; i.e. the following proposal distribution was used:

$$\theta' = \theta_0 + \varepsilon_0 \qquad 3\text{-}22$$

C. For each pair of θ *and* θ', acceptance ration r (ratio of the posterior value of θ' *and* θ) was calculated to decide whether or not to accept the proposed candidate:

$$r = \frac{p(\theta\prime|x)}{p(\theta_0|x)} \qquad 3\text{-}23$$

If $r \geq 1$, then the candidate was accepted ($\theta_t = \theta'$). Otherwise the candidate was rejected ($\theta^t = \theta^{t-1}$).

D. Steps B and C were repeated 500 times as this number of iterations was found adequate to achieve convergence to the posterior distribution.

An efficient MH sampler has an acceptance rate of between 15% and 50% (StataCorp, 2015). The choice of a proposal distribution was important for the rate at which the Markov chain explored its stationary distribution. Different initial values were tried to verify that the convergence of the MCMC was unaffected by starting values. Appendix I presents the steps and calculations involved for a single iteration.

III. Convergence diagnostics

Trace plots are the most accessible convergence diagnostics and are easy to inspect visually. The trace plots of a well-mixing parameter should traverse the posterior domain rapidly and should have nearly constant mean and variance.

Samples simulated using MCMC methods are correlated; the smaller the correlation the more efficient the sampling process.

Trace plots and correlation plots were used to diagnose convergence of the simulated results of the iterations.

IV. Summarising and Reporting Results

The summaries after simulation included posterior mean, standard deviation, coefficient of variation and 95% confidence interval. The inference was only valid if the Markov chain had converged and the acceptance rate was equal or greater than 15%.

3.5.2 Single-Site Estimations

For each rainfall station, a univariate analysis was carried out as described in Sections 3.1.1 and 3.2, where the posterior distribution of the location and scale parameters of the marginal distribution were inferred. The Type I GEV probability density function as described in Equation 3-6 was used as a likelihood model for the distribution of the rainfall data in the univariate (single-site) analysis.

The Normal distribution, with mean = 0 and variance = 100, was chosen as the prior distribution for all the parameters. The initial values for μ and σ were estimated simultaneously using the method of L-Moments. Parameters μ and σ can be expressed in terms of L-moments, which are in turn expressed in terms of probability weighted moments as shown in Equations 3-24 to 3-29 (Hann, 1977).

$$\mu = \lambda_1 - \sigma\gamma \tag{3-24}$$

$$\sigma = \frac{\lambda_2}{log\,2} \tag{3-25}$$

where γ = Euler constant = 0.5772

The 1st and 2nd L-Moments (λ_1 *and* λ_2) are expressed in terms of the probability weighted moments as shown in Equations 3-26 and 3-27 below.

$$\lambda_1 = M_{100} \tag{3-26}$$

$$\lambda_2 = 2 \times M_{110} - M_{110} \tag{3-27}$$

The probability weighted moments are expressed in terms of the sample size (N), sample data (x_i) and data rank number (i) in ascending order.

$$M_{100} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \textit{3-28}$$

$$M_{110} = \frac{1}{N}\sum_{i=1}^{N} \frac{(i-1)}{(N-1)} x_i \qquad \textit{3-29}$$

Steps I to V describe the implementation of the analysis.

I. Using Microsoft ® Excel ® 2013 for Bayesian analysis

The Type I GEV probability density function and the normal distribution function were programmed into Excel as the likelihood model of the data and prior distribution of the parameters, respectively. Simulation was performed with 500 iterations.

II. Simulation Results

The resulting values of μ *and* σ from the 500 iterations were summarised to determine the mean, standard deviations, coefficient of variation and 95% confidence intervals. Graphical representations of the distribution of the results also formed part of the results.

III. Convergence diagnostics

Graphical representations (Trace and Auto-correlation plots) were assessed in order to diagnose convergence of each simulation.

IV. Sensitivity Analysis

Five (5) additional simulations were undertaken using different initial values of μ and σ, in order to assess the sensitivity of the simulation results to different initial values. This was done by comparing the posterior values of mean from all the six (6) simulations.

V. Frequency Analysis

The final results of the simulation were used in the frequency analysis of the rainfall data. The frequency analysis was undertaken with the aim of relating the magnitude of extreme rainfall events to their frequency of occurrence through the use of their assumed probability distribution.

The magnitude of extreme rainfall events (x) were determined using pairs of μ and σ at different significance levels and for different return periods (T), as shown in Equation 3-30 and 3-31 (Chadwick, 2004).

$$T = \frac{1}{1-F(x)}, where\ 0 < F(x) < 1 \qquad \textit{3-30}$$

$$x = \mu - (\sigma \times u), where\ \ u = -ln\left[ln\left(\frac{1}{F(x)}\right)\right] f(x) = F(x)\ \frac{1}{\sigma}[exp - u] \qquad \textit{3-31}$$

3.5.3 Two-Site Estimations

For two pairs of rainfall stations, bivariate analyses were carried out as described in Sections 31.2 and 3.2, where the posterior distribution of the location and scale parameters of the joint distributions were inferred. Table 3-2 lists how the rainfall stations were paired.

Table 3-2: Rainfall Stations in Pairs for the Two-Site (Bivariate) Analysis

	Rainfall Station 1 (Variable X)	Rainfall Station 2 (Variable Y)
Pair 1	Moyeni	Seaka
Pair 2	Moyeni	Dilli-Dilli

The bivariate probability density function as described in Equation 3-10 was used as the likelihood model for the distribution of the rainfall data in the two-site analyses.

The Normal distribution, with mean = 0 and variance = 100, was chosen as the prior distribution for all the parameters. The same initial values for μ and σ, used in the univariate analyses, were used in the bivariate analyses. Steps I to V describe the implementation of the analysis.

I. Using Microsoft ® Excel ® 2013 for Bayesian analysis

The bivariate probability density function and the normal distribution function were programmed into Excel as the likelihood model of the data and prior distribution of the parameters, respectively. Simulation was performed with 500 iterations.

II. Simulation Results

The resulting values of μ and σ from the 500 iterations were summarised to determine the mean, standard deviations, coefficient of variation and 95% confidence intervals. Graphical representations of the distribution of the results also formed part of the results.

III. Convergence diagnostics

Graphical representations (Trace and Auto-correlation plots) were assessed in order to diagnose convergence of each simulation.

IV. Sensitivity Analysis

Four (4) additional simulations are undertaken using different initial values of μ and σ, in order to check the sensitivity of the simulation results to different initial values. Posterior values of mean were compared.

V. Frequency Analysis

The final results of the simulation were used in the frequency analysis of the rainfall data. The frequency analysis was undertaken with the aim of relating the magnitude of extreme rainfall events to their frequency of occurrence through the use of their assumed probability distribution.

The magnitude of extreme rainfall events (*x, y*) were determined using pairs of μ and σ at different significance levels and for different return periods (T_x, T_y), as shown in Equations 3-32 to 3-35 (Chadwick, 2004).

$$T_x = \frac{1}{1-F(x)}, where\ 0 < F(x) < 1 \quad \text{3-32}$$

$$x = \mu_x - (\sigma_x \times u), where\ \ u = -ln\left[ln\left(\frac{1}{F(x)}\right)\right] f(x) = F(x)\ \frac{1}{\sigma_x}[exp -u] \quad \text{3-33}$$

$$T_y = \frac{1}{1-F(y)}, where\ 0 < F(y) < 1 \quad \text{3-34}$$

$$y = \mu_y - \left(\sigma_y \times v\right), where\ \ v = -ln\left[ln\left(\frac{1}{F(y)}\right)\right] f(y) = F(y)\ \frac{1}{\sigma_y}[exp -v] \quad \text{3-35}$$

4. RESULTS AND ANALYSIS

This chapter outlines results of the Bayesian MCMC analyses for both the single-site (univariate) and two-site (bivariate) analyses. Simulation results (estimated parameters using the Bayesian MCMC approach) and corresponding predictions of rainfall events for each rainfall station are presented.

A comparison was also made of results from univariate vs bivariate analyses in order to establish whether there was an improvement in precision of parameters when using the bivariate approach. An improvement in this precision was considered a reduction in the uncertainty and therefore an improvement of rainfall estimation.

4.1. Single-Site Results

Estimated values of GEV parameters were used in the single-site simulation to determine parameters using the Bayesian MCMC analysis. The results of the simulation were used to predict rainfall magnitude for various return periods.

4.1.1 Estimation of initial values of GEV parameters

The estimated values of μ and σ were determined using the method of L-moments (as described in Section 3.5.2). Table 4-1 shows the estimated values of μ and σ.

Table 4-1: Estimated values of μ and σ

Parameter	**Moyeni**	**Seaka**	**Dilli-Dilli**
M_{100}	164.849	139.828	166.737
M_{110}	97.472	90.025	111.737
λ_1	164.849	139.828	166.737
λ_2	30.096	40.223	56.431
Estimated μ	**99.976**	**133.617**	**187.459**
Estimated σ	**107.143**	**62.704**	**58.535**

4.1.2 Simulation Results

In the simulations, the estimated values of parameters μ and σ were used as initial values.

Table 4-2 shows summary of results of the simulation; the posterior mean, standard deviation, coefficient of variation and 95% confidence interval for each inferred parameter (μ and σ) in the single-site analysis for Moyeni.

Table 4-2: Summary Results of the Univariate Simulation - Posterior Mean, Standard Deviation, Coefficient of Variance and 95% Confidence Intervals - Moyeni

Summary Results	**Moyeni**			
	μ		σ	
Mean	117.034		83.592	
Standard Deviation	5.696		5.553	
Coefficient of Variation	0.049		0.066	
95% Confidence Interval	116.5	117.5	83.1	84.1

Figure 4-1 and Figure 4-2 show distributions of posterior values of μ and σ for the single-site analyses of Moyeni.

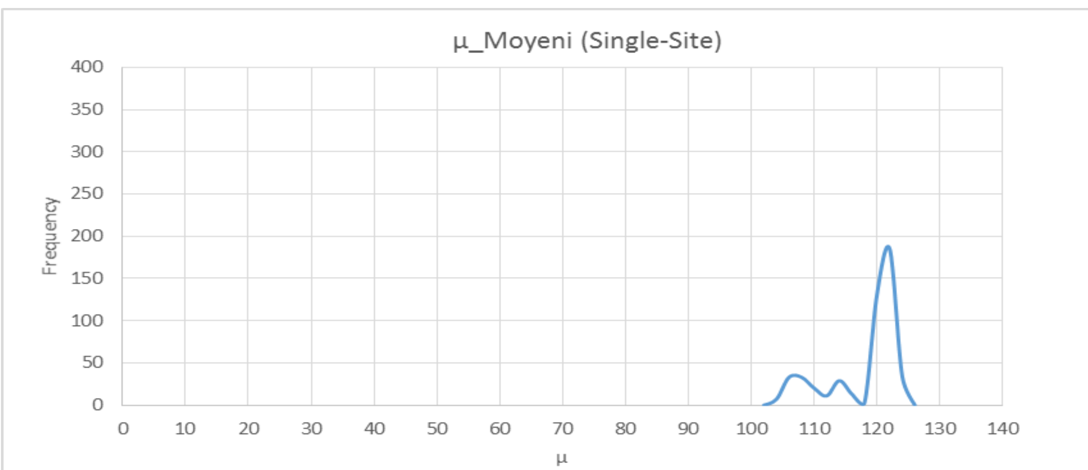


Figure 4-1: Single-Site - μ Distribution for Moyeni Rainfall Station

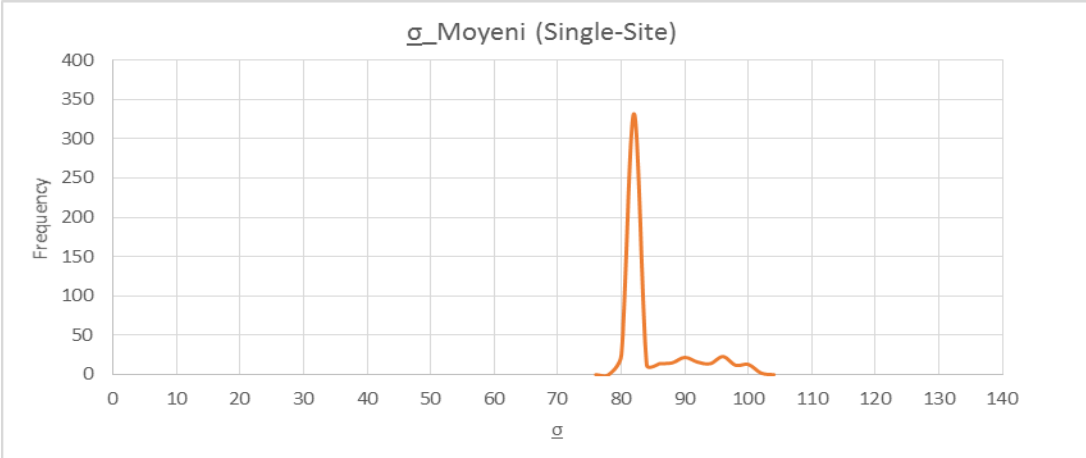


Figure 4-2: Single-Site - σ Distribution for Moyeni Rainfall Station

Table 4-3 shows summary results of the simulation; the posterior mean, standard deviation, coefficient of variation and 95% confidence interval for each inferred parameter (μ and σ) in the single-site analysis for Seaka.

Table 4-3: Summary Results of the Univariate Simulation - Posterior Mean, Standard Deviation, Coefficient of Variance and 95% Confidence Intervals for Seaka

Summary Results	Seaka			
	μ		σ	
Mean	21.376		43.320	
Standard Deviation	14.418		14.993	
Coefficient of Variation	0.675		0.346	
95% Confidence Interval	20.1	22.6	42.0	44.6

Figure 4-3 and Figure 4-4 show distributions of posterior values of μ and σ for the single-site analyses of Seaka.

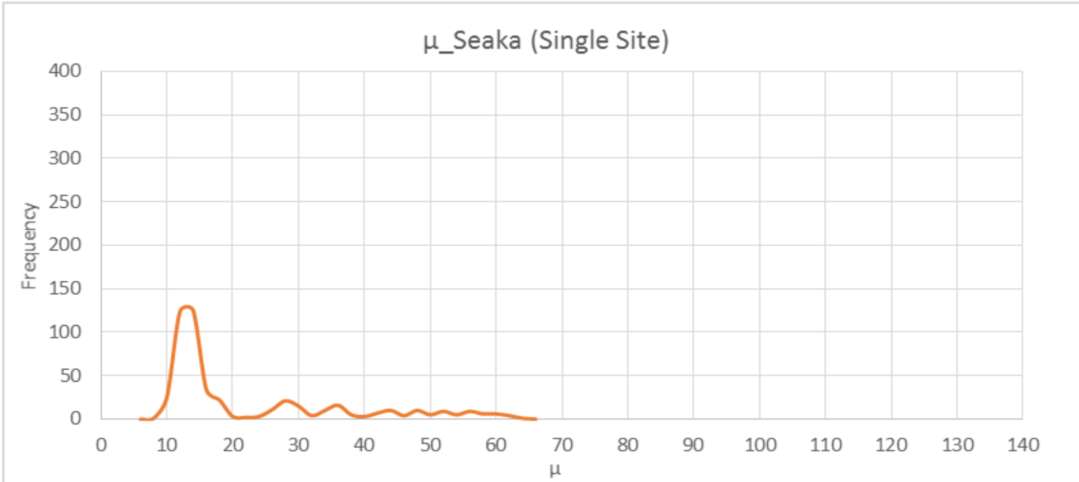


Figure 4-3: Single-Site - μ Distribution for Seaka Rainfall Station

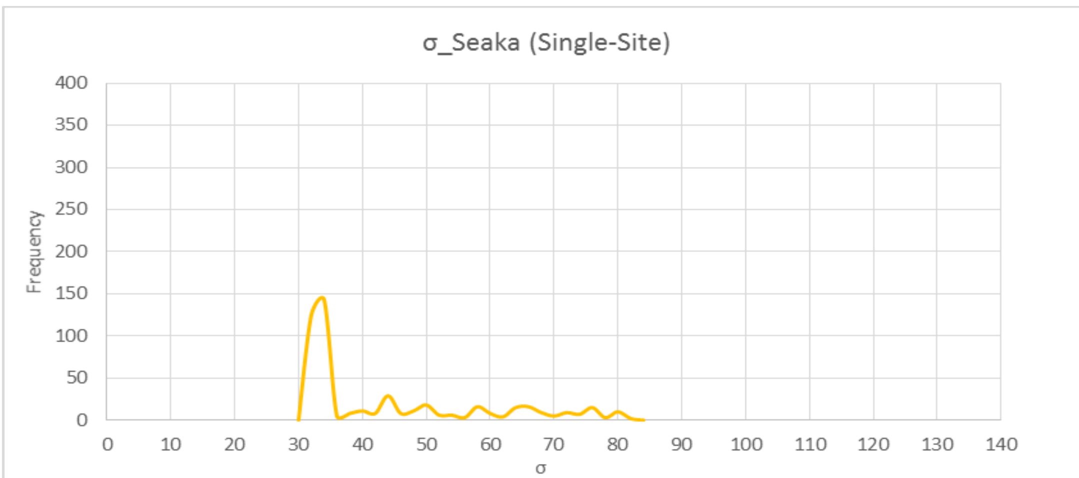


Figure 4-4: Single-Site - σ Distribution for Seaka Rainfall Station

Table 4-4 shows summary results of the simulation; the posterior mean, standard deviation, coefficient of variation and 95% confidence interval for each simulated parameter (μ and σ) in the single-site analysis for Dilli-Dilli.

Table 4-4: Summary Results of the Univariate Simulation - Posterior Mean, Standard Deviation, Coefficient of Variance and 95% Confidence Intervals – Dilli-Dilli

Summary Results	Dilli-Dilli			
	μ		σ	
Mean	10.778		39.513	
Standard Deviation	11.245		22.343	
Coefficient of Variation	1.043		0.565	
95% Confidence Interval	9.8	11.8	37.6	41.5

Figure 4-5 and Figure 4-6 show distributions of posterior values of μ and σ for the single-site analyses of Dilli-Dilli.

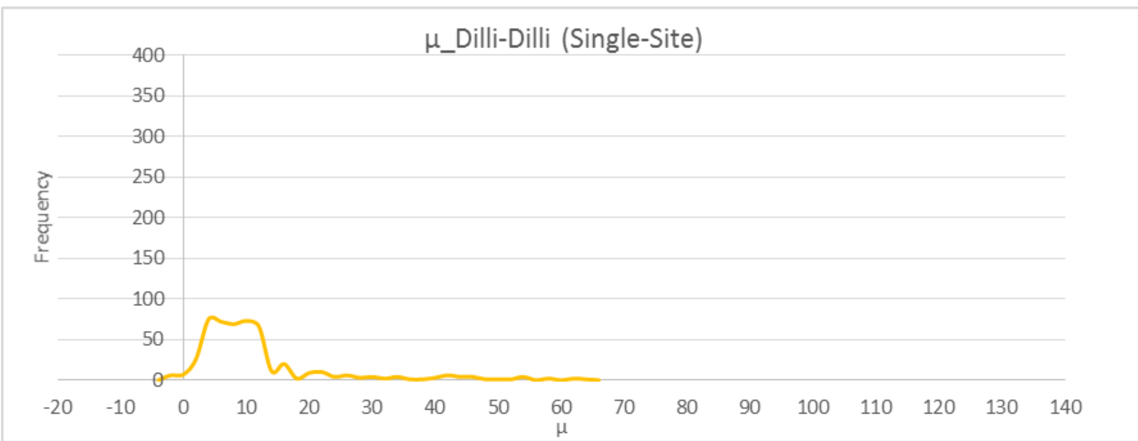


Figure 4-5: Single-Site - μ Distribution for Dilli-Dilli Rainfall Station

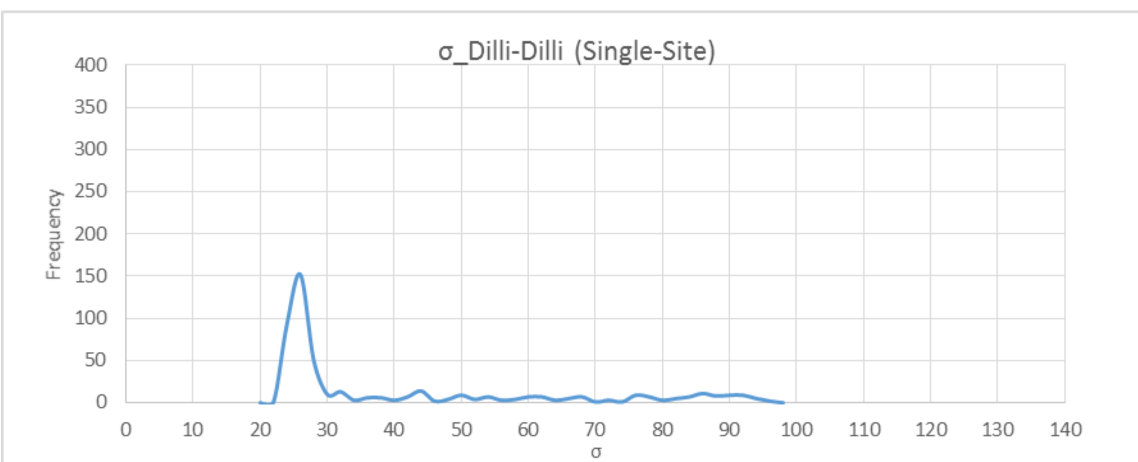


Figure 4-6: Single-Site - σ Distribution for Dilli-Dilli Rainfall Station

The results above showed a general increase in variability (decrease in precision) of the inferred values of μ and σ, with decrease in data record length. This is depicted in Figure 4-7 which shows an increase in the coefficient of variation with decrease in the length of the data record.

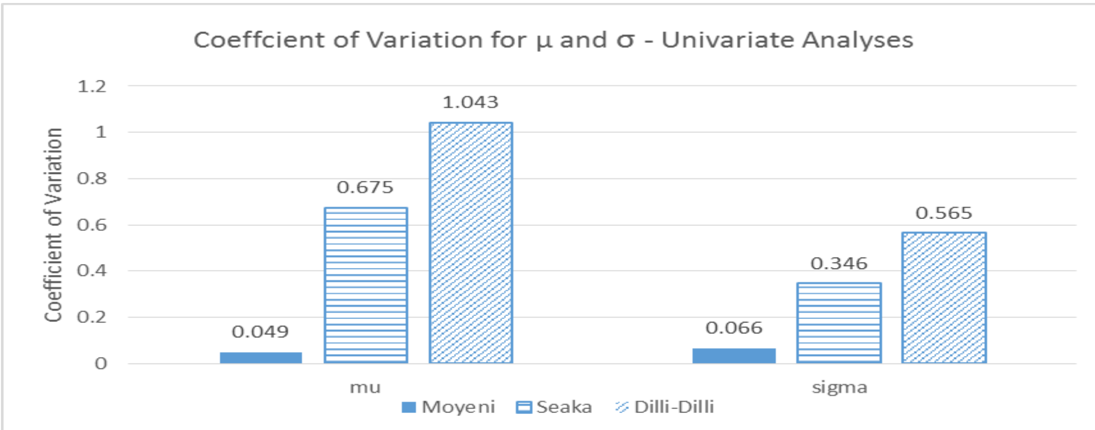


Figure 4-7: Coefficients of variation for μ and σ – Univariate Analyses

The variability of μ in the case of Seaka and Dilli-Dilli was 14 and 21 times higher than variability for Moyeni, respectively and variability of σ in the case of Seaka and Dilli-Dilli was 5 and 9 times higher than variability for Moyeni, respectively. The results of Moyeni were more precise compared to Seaka and Dilli-Dilli and Seaka's results were more precise than those of Dilli-Dilli.

4.1.3 Convergence Diagnostics

Trace plots and autocorrelation plots, as shown in Appendix 2, were used to assess convergence. From the trace and autocorrelation plots, there seemed to be overall good mixing and presence of convergence in the simulations. This was evident in the nearly constant mean & variance in the trace plots and the small values of correlations (<0.05) in the autocorrelation plots, particularly after the first 200 iterations.

4.1.4 Sensitivity Analysis

Five (5) additional simulations were undertaken, where different initial parameter values of μ and σ were used in order to assess the sensitivity of the results to the changing initial parameter values.

Table 4-5 and Figure 4-8 show summary results from 6 simulations of Moyeni single-site analysis.

Table 4-5: Initial values and Posterior mean for 6 simulations for Moyeni Rainfall Station

Rainfall Station	**MOYENI**											
	1		**2**		**3**		**4**		**5**		**6**	
Simulation	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Initial Value	108	100	119	110	130	120	140	130	151	140	162	150
Posterior Mean	118	84	117	93	113	105	114	116	110	116	112	126
			Analysis A		Analysis B		Analysis C		Analysis D		Analysis E	
Change in Initial Value			10%	10%	20%	20%	30%	30%	40%	40%	50%	50%
Change in Posterior Mean			-1%	12%	-4%	25%	-3%	39%	-6%	39%	-5%	51%

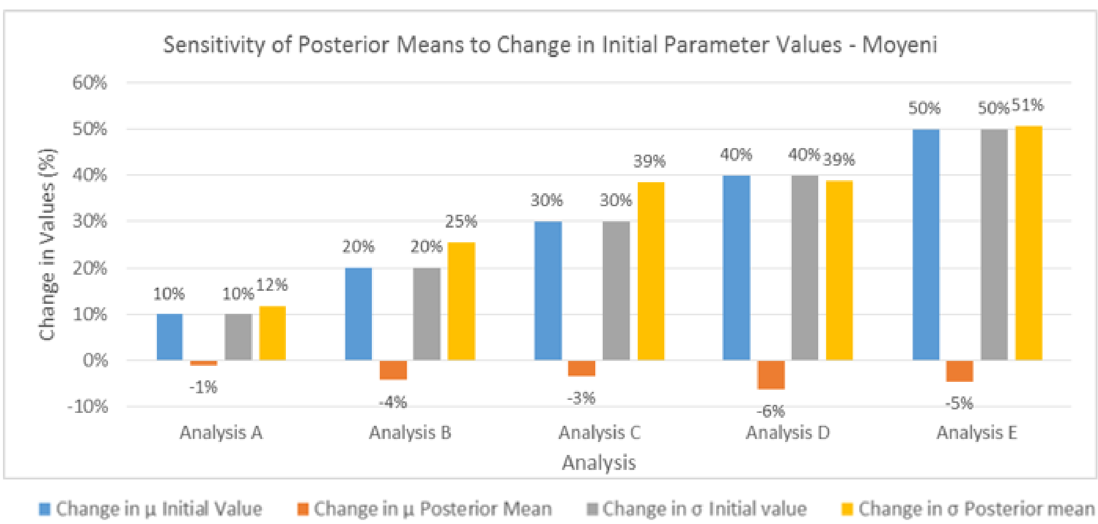


Figure 4-8: Sensitivity of Posterior Mean to Change in Initial Parameter Values – Moyeni Rainfall Station

Overall, changes in the initial values of µ resulted in lower changes in posterior mean while changes in initial values of σ resulted in generally equal changes in posterior mean. This meant for Moyeni, σ was more sensitive to changes in initial values than µ.

Table 4-6 and Figure 4-9 show summary results from 6 simulations of Seaka single-site analysis.

Table 4-6: Initial values and Posterior mean for 3 simulations – Seaka Rainfall Station

Rainfall Station	**SEAKA**											
	1		**2**		**3**		**4**		**5**		**6**	
Simulation	µ	σ	µ	σ	µ	σ	µ	σ	µ	σ	µ	σ
Initial Value	63	134	57	121	52	111	48	103	45	95	42	89
Posterior Mean	21	43	19	41	21	39	19	37	20	39	17	36
			Analysis A		Analysis B		Analysis C		Analysis D		Analysis E	
Change in Initial Value			-9%	-9%	-17%	-17%	-23%	-23%	-29%	-29%	-33%	-33%
Change in Posterior Mean			-10%	-4%	0%	-11%	-10%	-16%	-8%	-9%	-22%	-18%

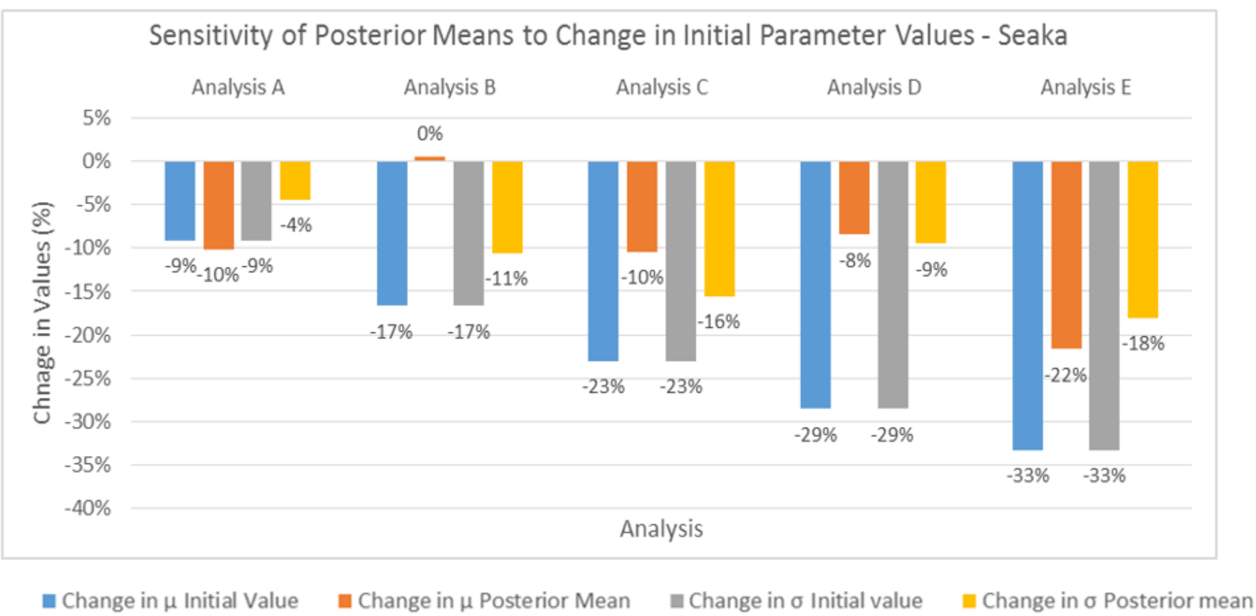


Figure 4-9: Sensitivity of Posterior Mean to Change in Initial Parameter Values for Seaka Rainfall Station

Overall, changes in the initial values of μ and σ resulted in lower changes in posterior mean. This meant for Seaka, both μ and σ were not sensitive to changes in initial values.

Table 4-7 and Figure 4-10 show summary results from 6 simulations of Dilli-Dilli single-site analysis.

Table 4-7: Initial values and Posterior mean for 3 simulations for Dilli-Dilli Rainfall Station

Rainfall Station	**DILLI-DILLI**											
	1		**2**		**3**		**4**		**5**		**6**	
Simulation	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Initial Value	59	187	53	170	49	156	45	144	42	134	39	125
Posterior Mean	11	39	10	36	9	33	9	32	9	30	7	32
			Analysis A		Analysis B		Analysis C		Analysis D		Analysis E	
Change in Initial Value			-9%	-9%	-17%	-17%	-23%	-23%	-29%	-29%	-33%	-33%
Change in Posterior Mean			-10%	-9%	-13%	-16%	-16%	-20%	-17%	-23%	-36%	-19%

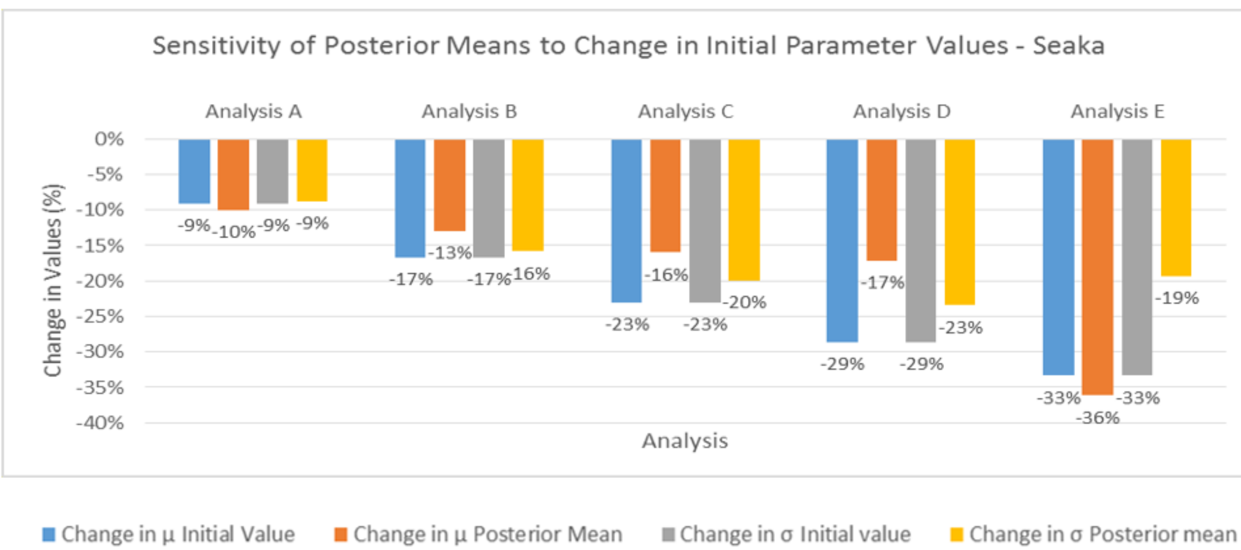


Figure 4-10: Sensitivity of Posterior Mean to Change in Initial Parameter Values – Dilli-Dilli Rainfall Station

Overall, changes in the initial values of μ and σ resulted in lower changes in posterior mean. This meant for Dilli-Dilli, both μ and σ were not sensitive to changes in initial values.

The resulting posterior mean revealed that the analysis was not significantly sensitive to the initial parameter values, with the exception of the σ parameter for Moyeni. The parameter σ was more sensitive to the initial parameter values than μ and sensitivity increased with decrease in data length, i.e. Moyeni showed the least overall sensitivity while Dilli-Dilli showed the most sensitivity.

4.1.5 Frequency Analysis

Since the sensitivity analysis showed that the analyses were not significantly sensitive to the initial values, results from simulation 1 (where the initial parameters were obtained from the historic data using the method of L-moments) were used in the frequency analyses. For each pair of μ and σ a frequency analysis was undertaken to determine the probability distribution, the cumulative distribution and the magnitude of rainfall at various return periods.

The resulting rainfall events for Moyeni are shown in Table 4-8 and the corresponding cumulative density plots and probability density plots are shown in Figure 4-11 and Figure 4-12.

Table 4-8: Single-Site Rainfall Events and Return Periods at various Significance Levels - Moyeni

		Moyeni Univariate Analysis		
		2.5% CL	**Mean**	**97.5% CL**
μ		116.534	117.034	117.533
σ		83.106	83.592	84.079
		Rainfall (mm)		
Return Period	**2 Years**	146.994	147.671	148.349
	5 Years	241.188	242.417	243.647
	10 Years	303.552	305.147	306.742
	20 Years	363.374	365.319	367.264
	50 Years	440.807	443.206	445.604
	100 Years	498.832	501.571	504.309
	200 Years	556.646	559.723	562.800

There was an average difference of 0.50% in values of μ and σ between the 2.5% and 97.5% confidence interval and the mean values. The resulting average difference in values of rainfall determined from the μ and σ values was 0.52%.

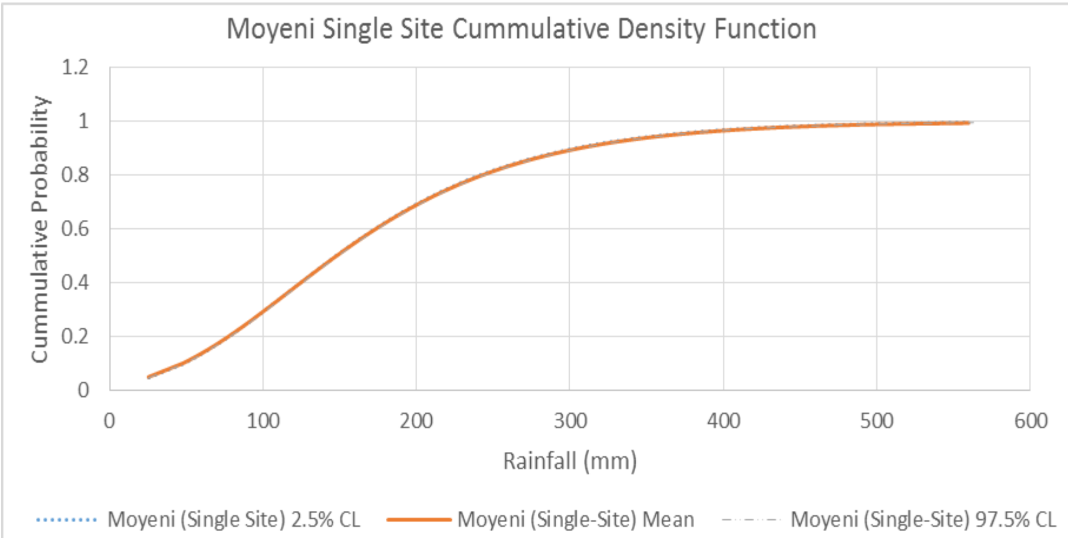


Figure 4-11: Moyeni Single-Site: Cumulative Probability

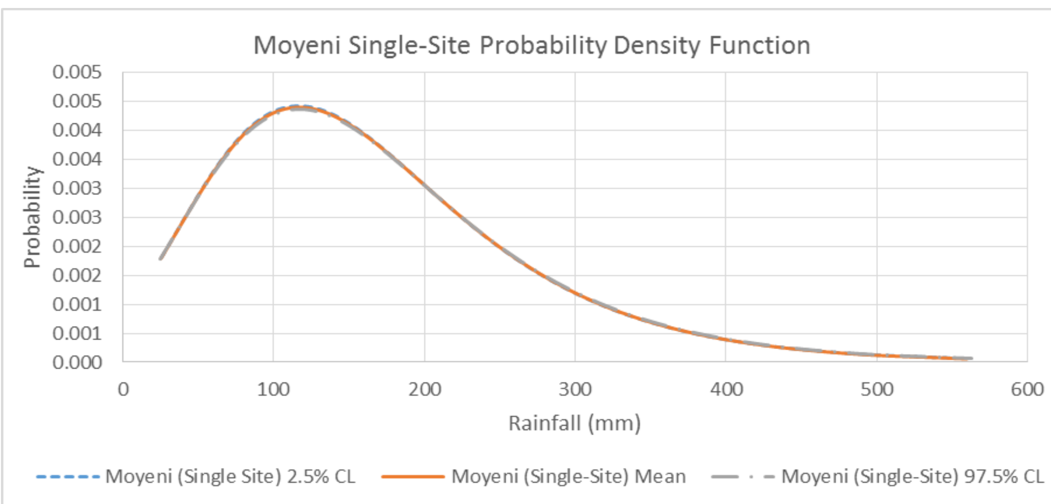


Figure 4-12: Moyeni Single-Site: Probability

The resulting rainfall events for Seaka are shown in Table 4-9 and the corresponding cumulative density plots and probability density plots are shown in Figure 4-13 and Figure 4-14.

Table 4-9: Single-Site Rainfall Events and Return Periods at various Significance Levels - Seaka

		Seaka Univariate Analysis		
		Seaka 2.5% CL	**Seaka Mean**	**Seaka 97.5% CL**
μ		65.213	65.547	65.882
σ		69.512	69.838	70.164
		Rainfall (mm)		
Return period	**2 Years**	90.690	91.144	91.598
	5 Years	169.477	170.300	171.124
	10 Years	221.641	222.709	223.777
	20 Years	271.678	272.980	274.283
	50 Years	336.445	338.052	339.658
	100 Years	384.979	386.813	388.647
	200 Years	433.337	435.397	437.458

There was an average difference of 4.5% in values of μ and σ between the 2.5% and 97.5% confidence interval and the mean values. The resulting average difference in values of rainfall determined from the μ and σ values was 3.6%.

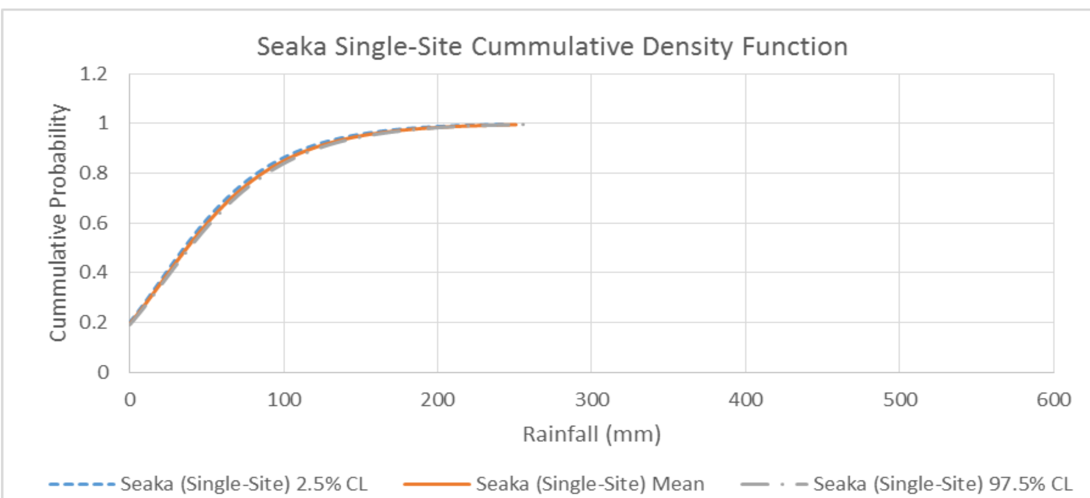


Figure 4-13: Seaka Single-Site: Cumulative Probability

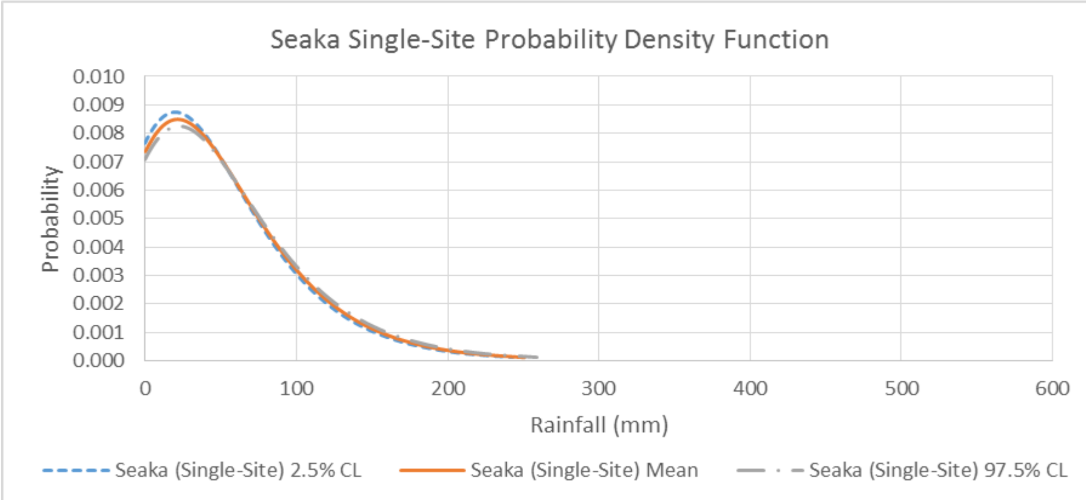


Figure 4-14: Seaka Single-Site: Probability

The resulting rainfall events for Seaka are shown in Table 4-10 and the corresponding cumulative density plots and probability density plots are shown in Figure 4-15 and Figure 4-16.

Table 4-10: Single-Site Rainfall Events and Return Periods at various Significance Levels - Moyeni

		Dilli-Dilli Univariate Analysis		
		Dilli-Dilli 2.5%	**Dilli-Dilli Mean**	**Dilli-Dilli 97.5%**
μ		9.792	10.778	11.763
σ		37.555	39.513	41.472
		Rainfall (mm)		
Return period	**2 Years**	23.556	25.260	26.963
	5 Years	66.122	70.045	73.968
	10 Years	94.304	99.697	105.089
	20 Years	121.337	128.139	134.942
	50 Years	156.328	164.955	173.583
	100 Years	182.549	192.544	202.538
	200 Years	208.674	220.032	231.389

There was an average difference of 7.1% in values of μ and σ between the 2.5% and 97.5% confidence interval and the mean values. The resulting average difference in values of rainfall determined from the μ and σ values was 5.5%.

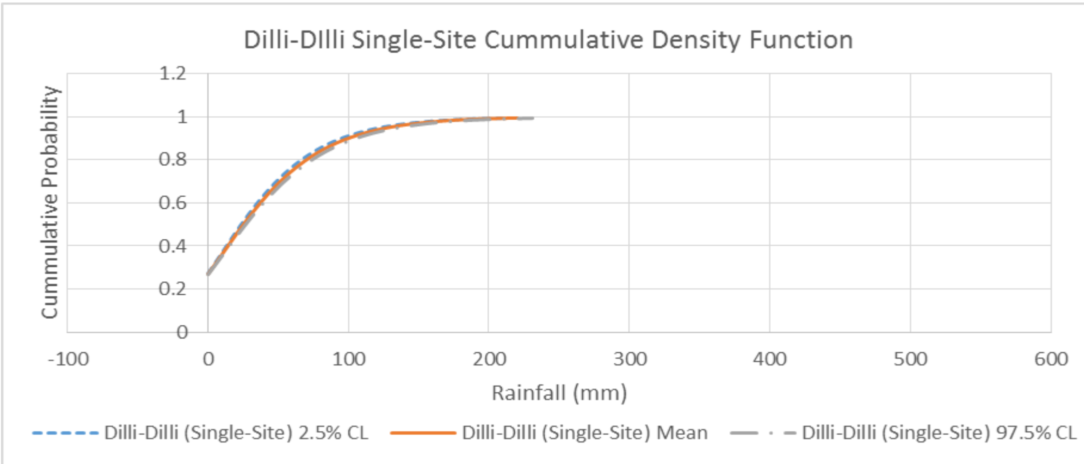


Figure 4-15: Dilli-Dilli Single-Site: Cumulative Probability

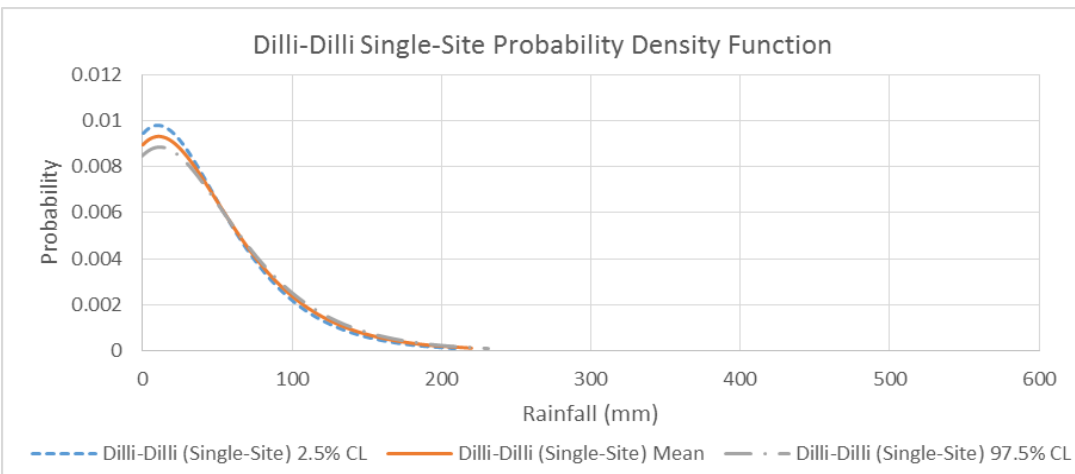


Figure 4-16: Dilli-Dilli Single-Site: Probability

The values predicted using values of μ and σ at the 2.5% confidence level, mean and the 97.5% confidence level did not vary significantly for all three rainfall stations. However the deviation was smallest for Moyeni and biggest for Dilli-Dilli. This is consistent with the observed decrease in precision with decrease in record length highlighted in Section 4.1.2.

4.2. Two-Site Results

Estimated values of GEV parameters were used in the two-site simulation to determine parameters using the Bayesian MCMC analysis. The results of the two-site simulation were used to predict rainfall magnitude for various return periods.

4.2.1. Initial values of GEV parameters

The same initial values used in the single-site (univariate) analyses were used in the two-site (bivariate) analyses. These values are presented in Table 4-11.

Table 4-11: Two-Site Initial Values of σ and μ

Parameter	**Moyeni**	**Seaka**	**Dilli-Dilli**
Estimated μ	99.976	133.617	187.459
Estimated σ	107.143	62.704	58.535
m	1	1	1

4.2.2. Simulation Results

In the simulations, the estimated values of parameters μ and σ were used as initial values. Table 4-12 shows summary results of the simulation; the posterior mean, standard deviation, coefficient of variation and 95% confidence interval for each simulated parameter of Moyeni in the two-site analyses of Moyeni and Seaka.

Table 4-12: Summary Results of the Bivariate (Moyeni and Seaka Rainfall Stations) Simulation – Moyeni Posterior Mean, Posterior Standard Deviation and 95% Confidence Intervals

Summary Results	**Moyeni & Seaka Two-site Analysis**			
	μ _Moyeni		**σ _Moyeni**	
Mean	109.986		101.66	
Standard Deviation	3.817		0.75	
Coefficient of Variation	0.04		0.01	
95% Confidence Interval	109.65	110.32	101.59	101.72

Figure 4-18 and Figure 4-17 below show distributions of Moyeni posterior values of μ and σ for the two-site analyses of Moyeni and Seaka.

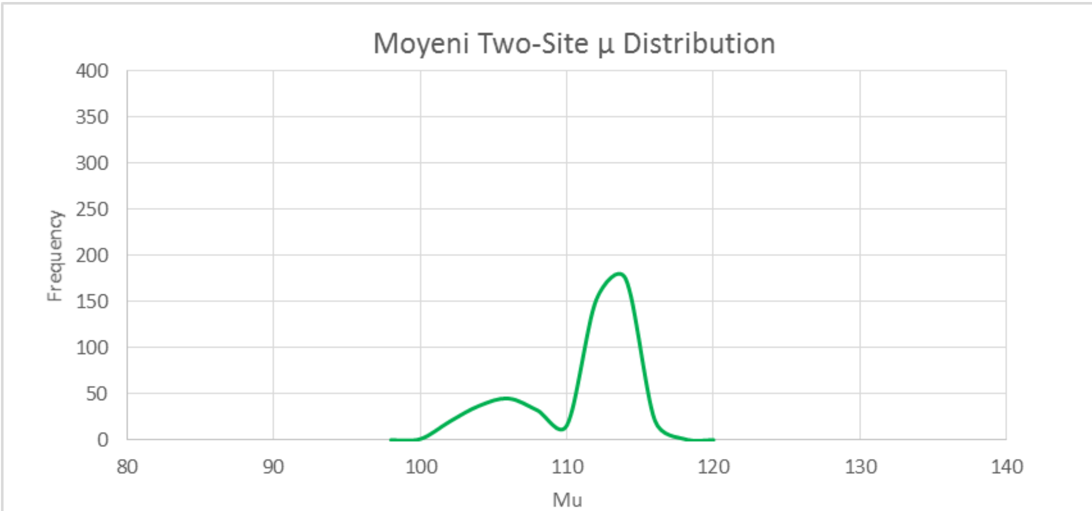


Figure 4-17: Two-Site (Moyeni and Seaka) - μ Distribution for Moyeni Rainfall Station

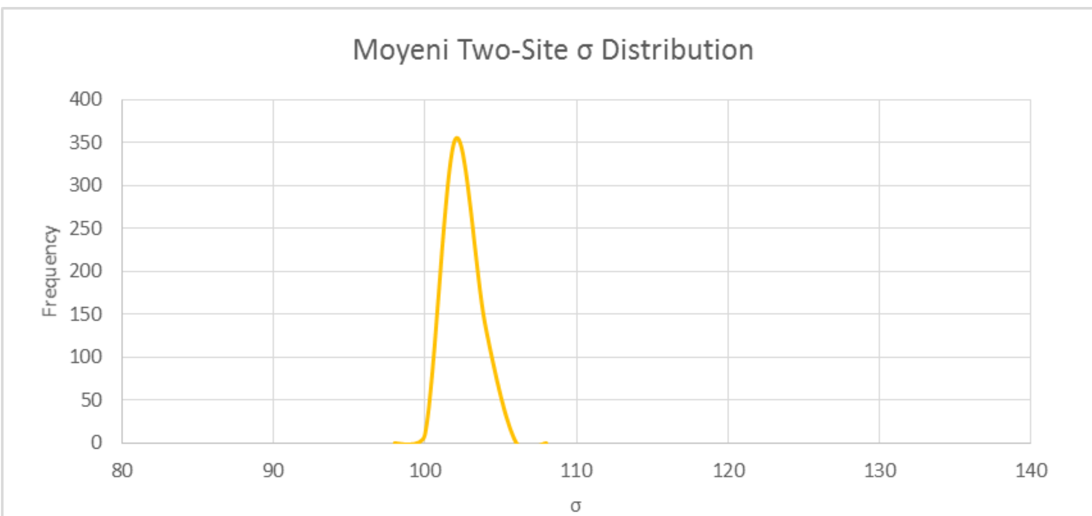


Figure 4-18: Two-Site (Moyeni and Seaka) - σ Distribution for Moyeni Rainfall Station

Table 4-13 shows summary results of the simulation; the posterior mean, standard deviation, coefficient of variation and 95% confidence interval for each simulated parameter of Seaka in the two-site analyses of Moyeni and Seaka.

Table 4-13: Summary Results of the Bivariate (Moyeni and Seaka Rainfall Stations) Simulation – Seaka Posterior Mean, Posterior Standard Deviation and 95% Confidence Intervals

Summary Results	Moyeni & Seaka Two-site Analysis			
	μ _Seaka		σ _Seaka	
Mean	65.547		69.838	
Standard Deviation	3.817		3.719	
Coefficient of Variation	0.06		0.05	
95% Confidence Interval	65.21	65.88	69.51	70.16

Figure 4-20 and Figure 4-19 show distributions of Seaka posterior values of μ and σ for the two-site analyses of Moyeni and Seaka.

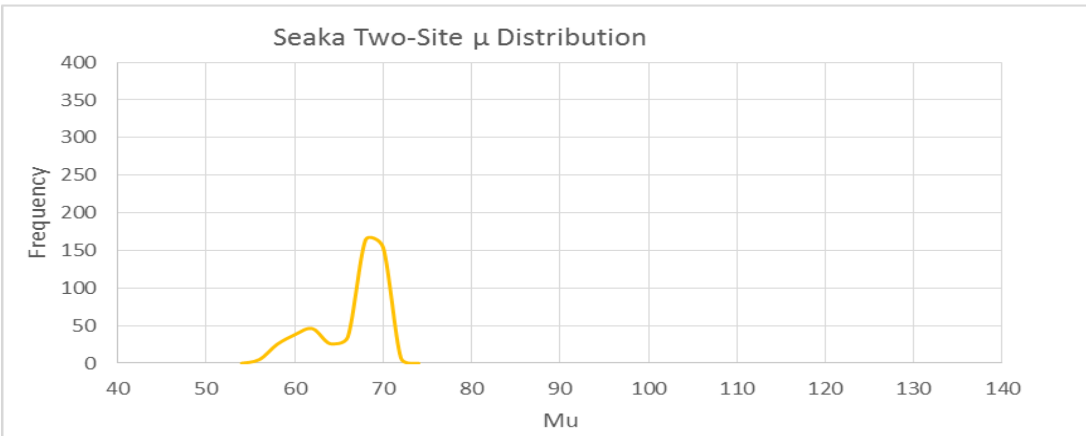


Figure 4-19: Two-Site (Moyeni and Seaka) - μ Distribution for Seaka Rainfall Station

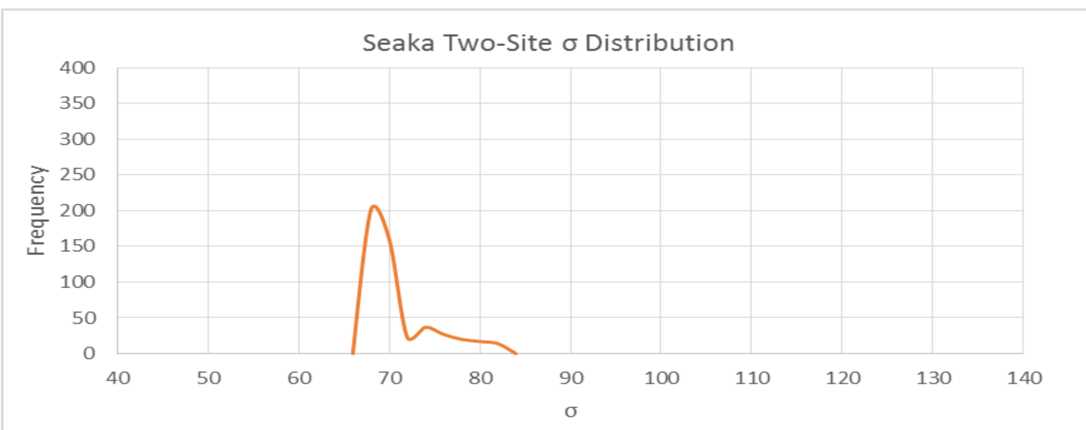


Figure 4-20: Two-Site (Moyeni and Seaka) - σ Distribution for Seaka Rainfall Station

Table 4-14 below summary results of the simulation; the posterior mean, standard deviation, coefficient of variation and 95% confidence interval for each simulated parameter of Moyeni in the two-site analyses of Moyeni and Dilli-Dilli.

Table 4-14: Summary Results of the Bivariate (Moyeni and Dilli-Dilli Rainfall Stations) Simulation – Moyeni Posterior Mean, Posterior Standard Deviation and 95% Confidence Intervals

Summary Results	Moyeni & Dilli-Dilli Two-site Analysis			
	μ _Moyeni		**σ _Moyeni**	
Mean	113.67		101.87	
Standard Deviation	4.64		0.67	
Coefficient of Variation	0.04		0.01	
95% Confidence Interval	113.27	114.08	101.82	101.93

Figure 4-22 and Figure 4-21 show distributions of posterior values of μ and σ for the two-site analyses.

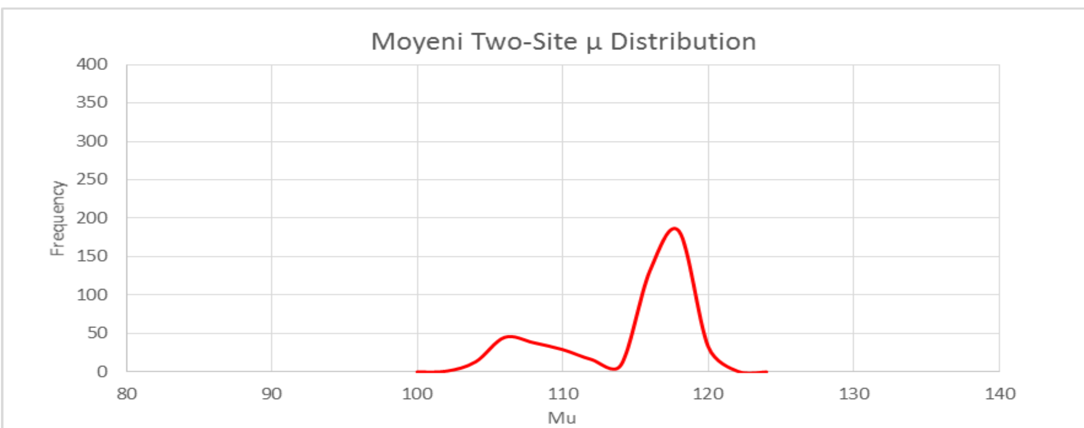


Figure 4-21: Two-Site (Moyeni and Dilli-Dilli) - μ Distribution for Moyeni Rainfall Station

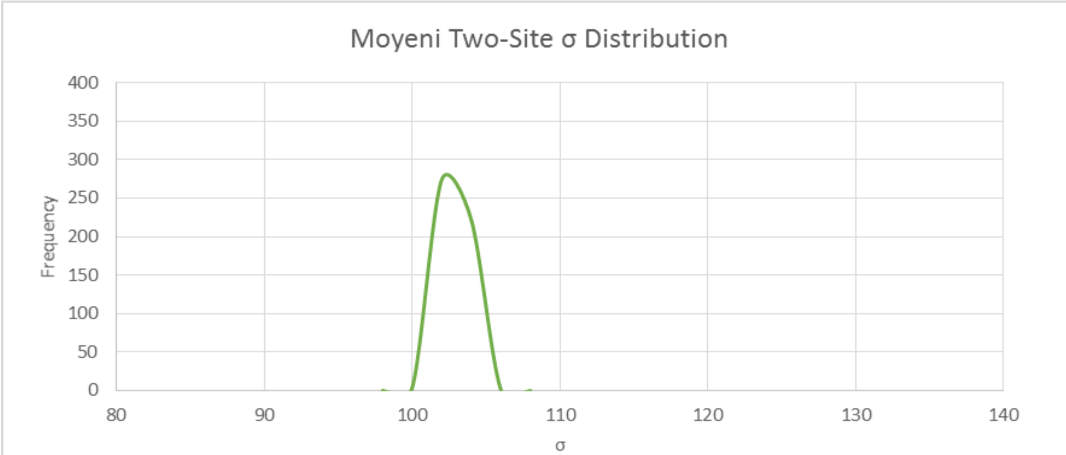


Figure 4-22: Two-Site (Moyeni and Dilli-Dilli) - σ Distribution for Moyeni Rainfall Station

Table 4-15 shows summary results of the simulation; the posterior mean, standard deviation, coefficient of variation and 95% confidence interval for each simulated parameter of Dilli-Dilli in the two-site analyses of Moyeni and Dilli-Dilli.

Table 4-15: Summary Results of the Bivariate (Moyeni and Dilli-Dilli Rainfall Stations) Simulation – Dilli-Dilli Posterior Mean, Posterior Standard Deviation and 95% Confidence Intervals

Summary Results	Moyeni & Dilli-Dilli Two-site Analysis			
	μ _Dilli-Dilli		σ _Dilli-Dilli	
Mean	86.90		101.53	
Standard Deviation	10.66		4.64	
Coefficient of Variation	0.12		0.05	
95% Confidence Interval	85.96	87.83	101.12	101.94

Figure 4-24 and Figure 4-23 show distributions of posterior values of μ and σ for the two-site analyses.

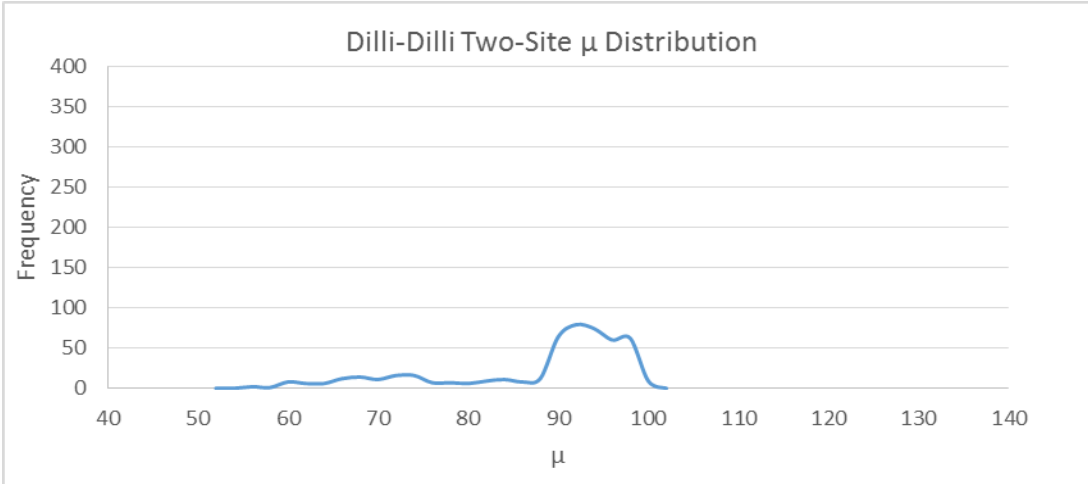


Figure 4-23: Two-Site (Moyeni and Dilli-Dilli) - μ Distribution for Dilli-Dilli Rainfall Station

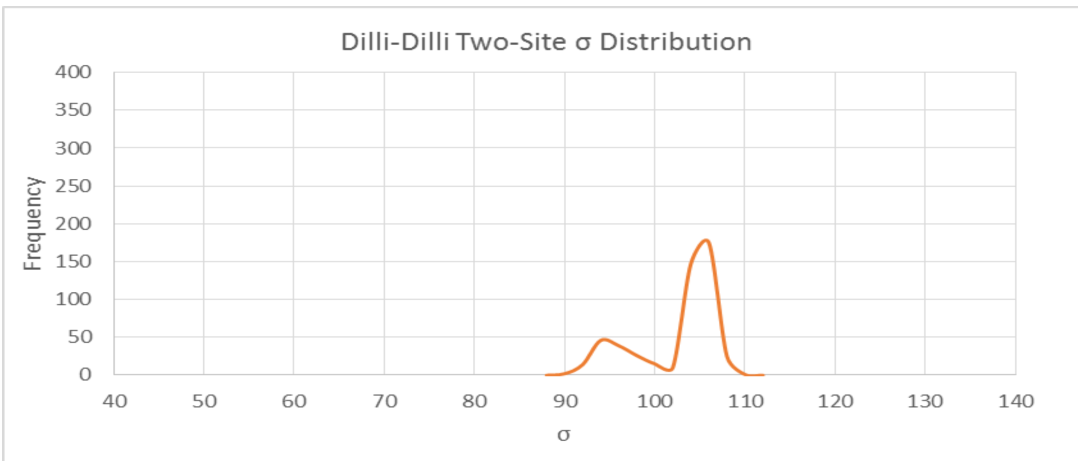


Figure 4-24: Two-Site (Moyeni and Dilli-Dilli) - σ Distribution for Dilli-Dilli Rainfall Station

The results showed a general increase in variability (decrease in precision) of the inferred values of μ and σ, with decrease in data record length. This is depicted in Figure 4-25 which shows the increase in the coefficient of variation with decrease in the length of the data record.

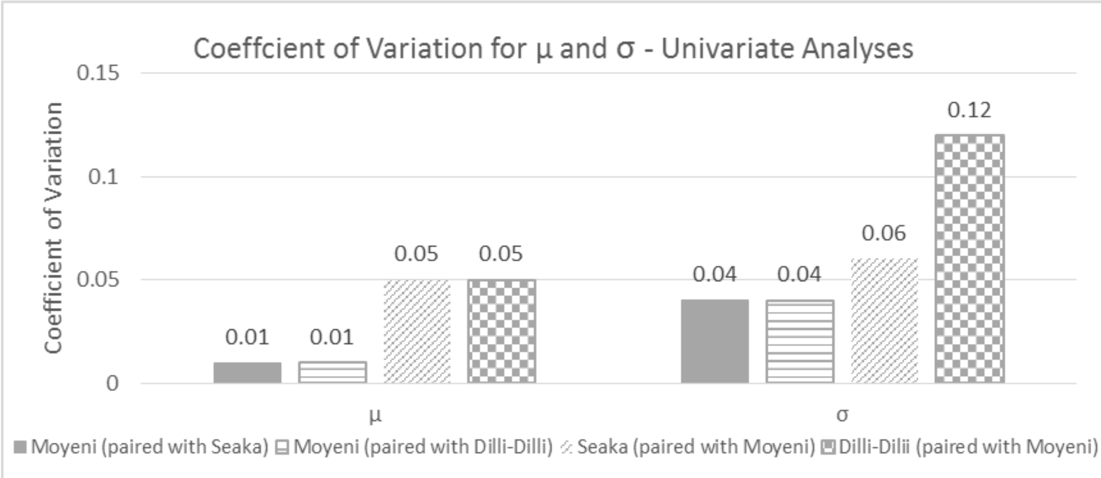


Figure 4-25: Coefficients of variation for μ and σ – Univariate Analyses

The variability of μ in the case of Seaka and Dilli-Dilli was 5 times higher than variability for Moyeni and variability of σ in the case of Seaka and Dilli-Dilli was 1.5 and 3 times higher than variability for Moyeni, respectively. The results of Moyeni were more precise compared to Seaka and Dilli-Dilli.

4.2.3. Convergence Diagnostics

Trace plots and autocorrelation plots, as shown in Appendix 2, were used to assess convergence.

From the trace and autocorrelation plots, there seemed to be overall good mixing and presence of convergence in the simulations. This was evident in the nearly constant mean & variance in the trace plots and the small values of correlations (<0.05) in the autocorrelation plots, particularly after the first 200 iterations.

4.2.4. Sensitivity Analysis

Four (4) additional simulations were undertaken, where different initial parameter values of μ and σ were used in order to assess the sensitivity of the results to the changing initial parameter values.

Table 4-16 and Figure 4-26 show summary results from 5 simulations of Moyeni (paired with Seaka) two-site analysis.

Table 4-16: Initial values and Posterior mean for 5 simulations: Moyeni (paired with Seaka) Rainfall Station

Rainfall Station	**MOYENI (paired with Seaka)**									
	1		**2**		**3**		**4**		**5**	
Simulation	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Initial Value	107	100	118	110	129	120	139	130	150	140
Posterior Mean	110	102	99	92	94	81	91	71	92	61
			Analysis A		Analysis B		Analysis C		Analysis D	
Change in Initial Value			10%	10%	20%	20%	30%	30%	40%	40%
Change in Posterior Mean			-10%	-10%	-15%	-20%	-17%	-30%	-16%	-40%

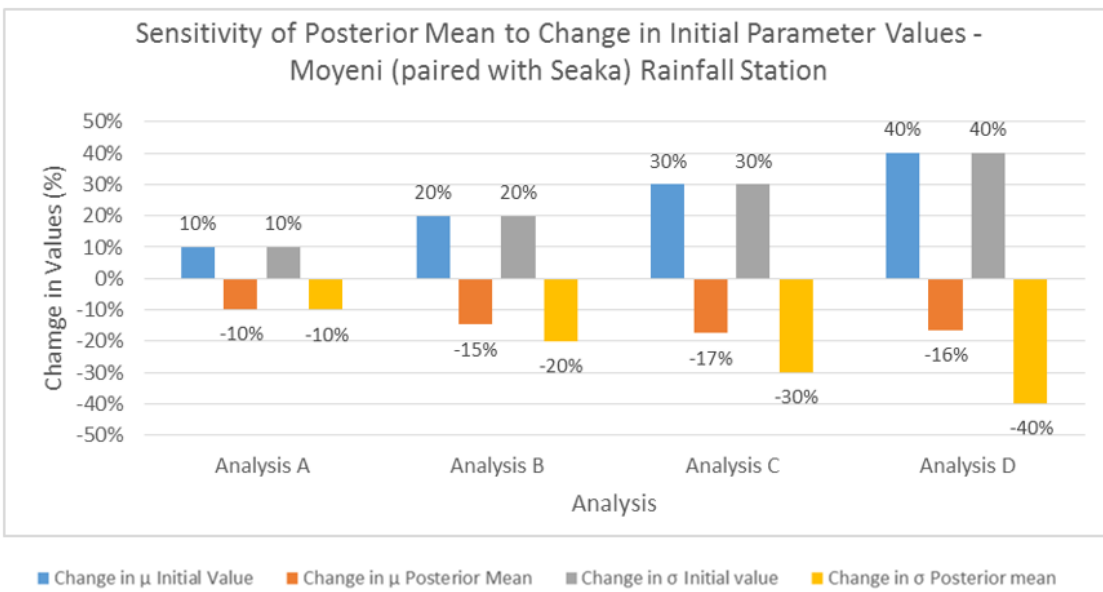


Figure 4-26: Sensitivity of Posterior Mean to Change in Initial Parameter Values – Moyeni (paired with Seaka) Rainfall Station

Overall, changes in the initial values of μ resulted in lower changes in posterior mean values, while changes in initial values of σ resulted in generally equal changes in posterior mean values. This meant σ was more sensitive to changes in initial values than μ.

Table 4-17 and Figure 4-27 show summary results from 5 simulations of Moyeni (paired with Dilli-Dilli) two-site analysis.

Table 4-17: Initial values and Posterior mean for 5 simulations: Moyeni (paired with Dilli-Dilli) Rainfall Station

Rainfall Station	**MOYENI (paired with Dilli)**									
Simulation	**1**		**2**		**3**		**4**		**5**	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Initial Value	107	100	118	110	129	120	139	130	150	140
Posterior Mean	114	102	104	92	98	82	92	73	76	62
			Analysis A		Analysis B		Analysis C		Analysis D	
Change in Initial Value			10%	10%	20%	20%	30%	30%	40%	40%
Change in Posterior Mean			-5%	-9%	-11%	-19%	-16%	-28%	-31%	-39%

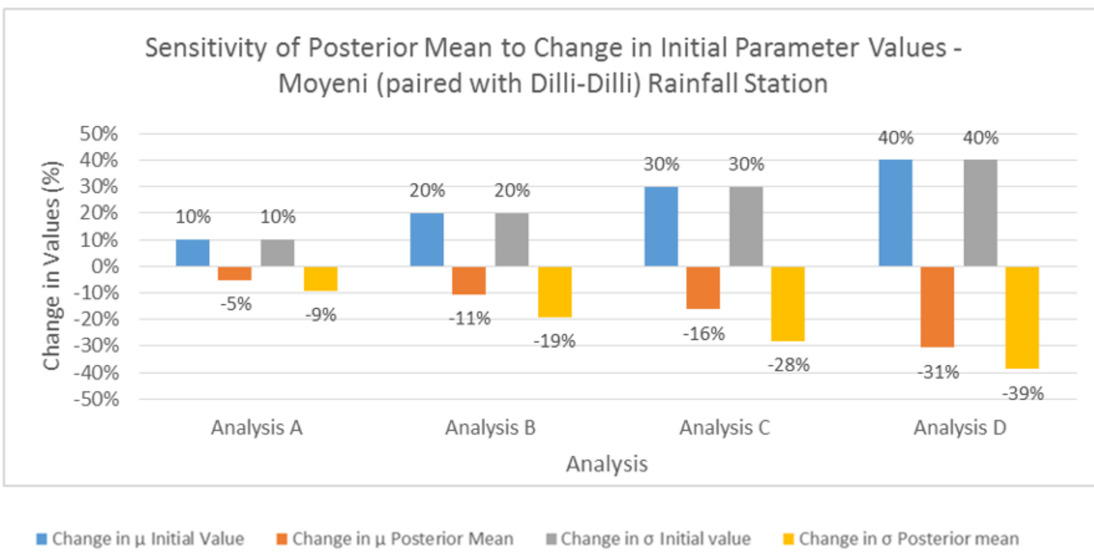


Figure 4-27: Sensitivity of Posterior Mean to Change in Initial Parameter Values – Moyeni (paired with Dilli-Dilli) Rainfall Station

Overall, changes in the initial values of μ resulted in lower changes in posterior mean values, while changes in initial values of σ resulted in generally equal changes in posterior mean values. This meant σ was more sensitive to changes in initial values than μ.

Table 4-18 and Figure 4-28 show summary results from 5 simulations of Seaka (paired with Moyeni) two-site analysis.

Table 4-18: Initial values and Posterior mean for 5 simulations: Seaka (paired with Moyeni) Rainfall Station

Rainfall Station	SEAKA (paired with Moyeni)									
	1		2		3		4		5	
Simulation	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Initial Value	63	134	57	121	52	111	48	103	45	95
Posterior Mean	66	70	59	62	58	56	59	51	65	47
			Analysis A		Analysis B		Analysis C		Analysis D	
Change in Initial Value			-9%	-9%	-17%	-17%	-23%	-23%	-29%	-29%
Change in Posterior Mean			-10%	-11%	-11%	-20%	-10%	-27%	0%	-32%

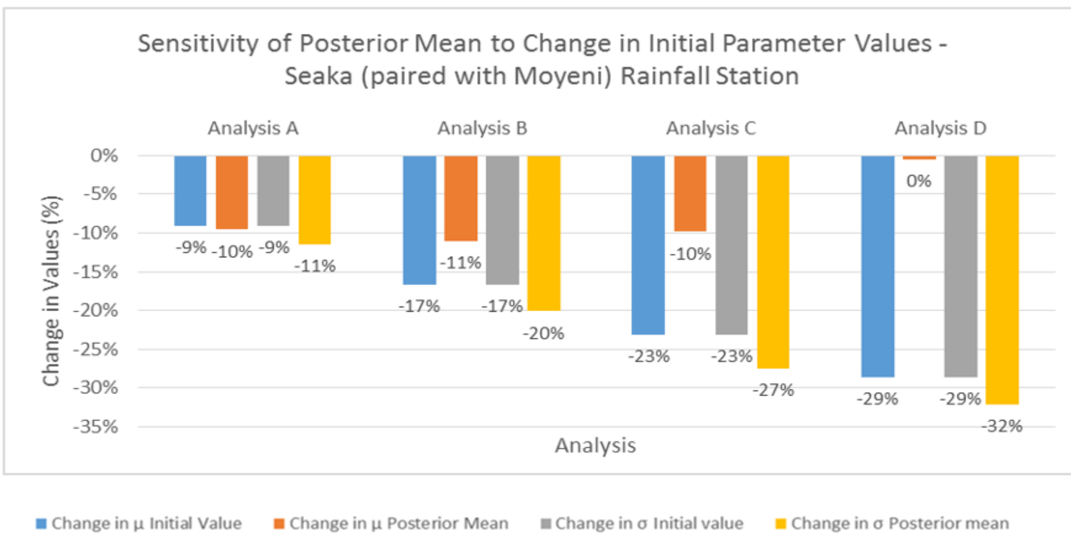


Figure 4-28: Sensitivity of Posterior Mean to Change in Initial Parameter Values – Seaka (paired with Moyeni) Rainfall Station

Overall, changes in the initial values of µ resulted in lower changes in posterior mean values, while changes in initial values of σ resulted in higher changes in posterior mean values. This meant σ was more sensitive to changes in initial values than µ.

Table 4-19 and Figure 4-29 show summary results from 5 simulations of Dilli-Dilli (paired with Moyeni) two-site analysis.

Table 4-19: Initial values and Posterior mean for 5 simulations: Dilli-Dilli (paired with Moyeni) Rainfall Station

Rainfall Station	**DILLI-DILLI (paired with Moyeni)**									
	1		**2**		**3**		**4**		**5**	
Simulation	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Initial Value	59	187	53	170	49	156	45	144	42	134
Posterior Mean	87	102	85	93	72	89	74	84	69	69
			Analysis A		Analysis B		Analysis C		Analysis D	
Change in Initial Value			-9%	-9%	-17%	-17%	-23%	-23%	-29%	-29%
Change in Posterior Mean			-3%	-8%	-17%	-13%	-15%	-18%	-21%	-32%

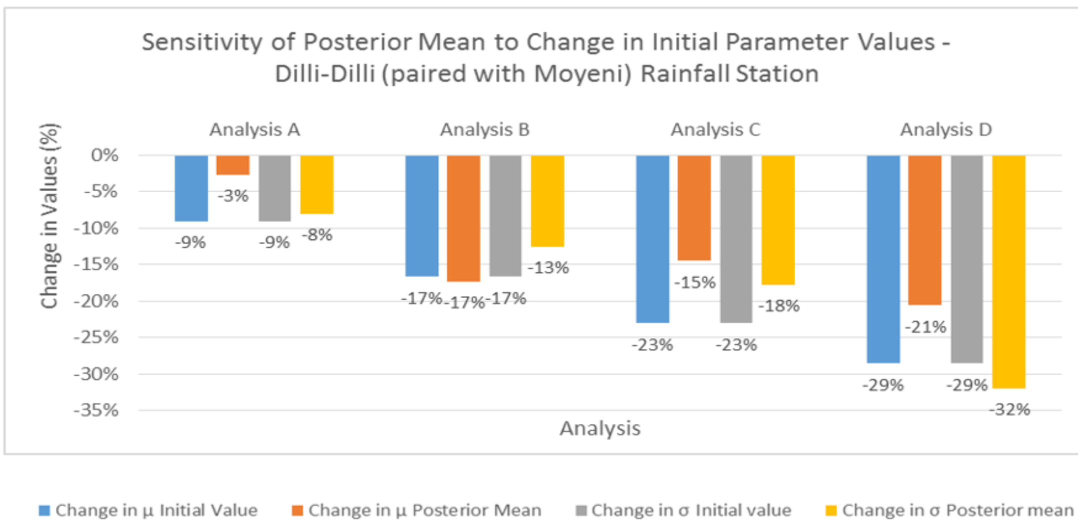


Figure 4-29: Sensitivity of Posterior Mean to Change in Initial Parameter Values – Dilli-Dilli (paired with Moyeni) Rainfall Station

Overall, changes in the initial values of μ resulted in generally lower or equal changes in posterior mean values, while changes in initial values of σ resulted in equal changes in posterior mean values. This meant σ was more sensitive to changes in initial values than μ.

The resulting posterior mean revealed that the analysis for μ was not significantly sensitive to the initial parameter values, while the analysis for σ was more sensitive to the initial parameter values.

4.2.5. Frequency Analysis

Since the sensitivity analysis shows that the analyses were not significantly sensitive to the initial values, results from Simulation 1 (where the initial parameters were obtained from the historic data using the method of L-moments) were used in the frequency analyses.

For each pair of rainfall stations and the corresponding μ_x, σ_x, μ_y and σ_y, a frequency analysis was undertaken to determine the probability distribution, the cumulative distribution and the magnitude of rainfall at various return periods.

The resulting rainfall events from the two-site analysis of Moyeni with Seaka are shown in Table 4-20 and the corresponding cumulative density plots and probability density plots are shown in Figure 4-30 and Figure 4-31.

Table 4-20: Two-Site Rainfall Events and Return Periods at various Significance Levels – Moyeni (paired with Seaka)

Simulation Results		Moyeni Two-Site (Bivariate) Analysis with Seaka		
		2.5% CL	**Mean**	**97.5% CL**
mu		109.652	109.986	110.321
sigma		101.592	101.658	101.723
Return period	**2 Years**	146.886	147.245	147.604
	5 Years	262.033	262.467	262.900
	10 Years	338.271	338.753	339.236
	20 Years	411.400	411.929	412.459
	50 Years	506.057	506.648	507.239
	100 Years	576.990	577.627	578.263
	200 Years	647.664	648.346	649.028

There was an average difference of 0.18% in values of μ and σ between the mean and the values at 2.5% and 97.5% confidence interval. The resulting average difference in values of rainfall determined from the μ and σ values was 0.14%.

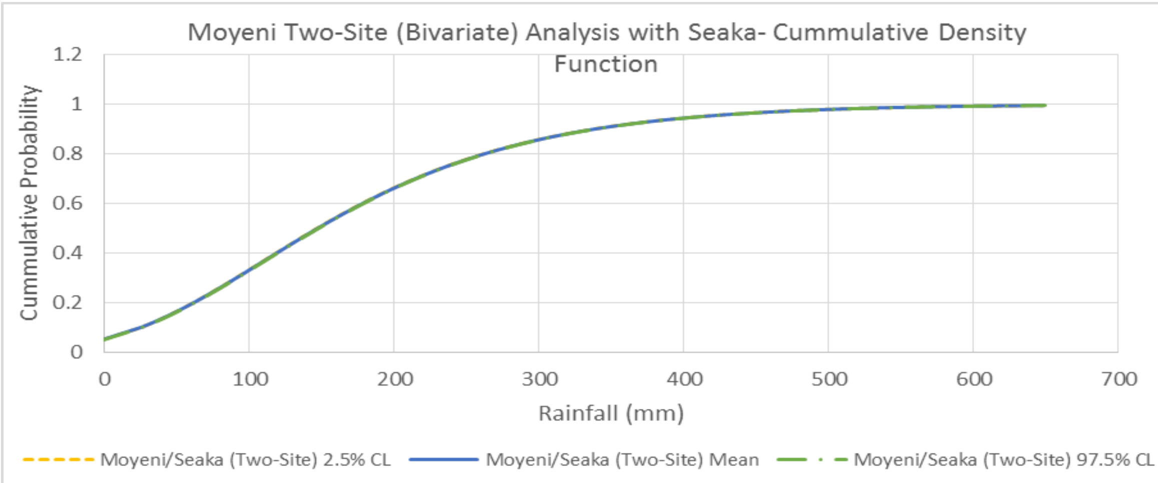


Figure 4-30: Moyeni Two-Site (Bivariate) Analysis with Seaka: Cumulative Probability

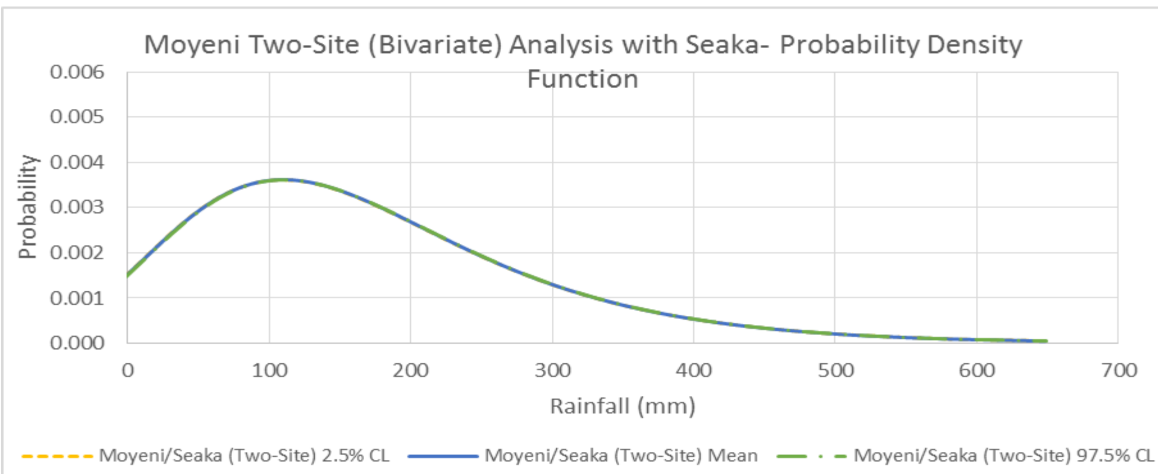


Figure 4-31: Moyeni Two-Site (Bivariate) Analysis with Seaka: Probability

The resulting rainfall events from the two-site analysis of Moyeni with Dilli-Dilli are shown in Table 4-21 and the corresponding cumulative density plots and probability density plots are shown in Figure 4-32 and Figure 4-33.

Table 4-21: Two-Site Rainfall Events and Return Periods at various Significance Levels – Moyeni (paired with Dilli-Dilli)

Simulation Results		Moyeni Two-Site (Bivariate Analysis) with Dilli-Dilli		
		2.5% CL	Mean	97.5% CL
mu		113.266	113.673	114.080
sigma		101.815	101.874	101.933
Return period	2 Years	150.583	151.011	151.440
	5 Years	265.983	266.478	266.973
	10 Years	342.387	342.927	343.466
	20 Years	415.677	416.258	416.840
	50 Years	510.542	511.178	511.815
	100 Years	581.631	582.308	582.985
	200 Years	652.460	653.178	653.896

There was an average difference of 0.21% in values of μ and σ between the mean and the values at 2.5% and 97.5% confidence interval. The resulting average difference in values of rainfall determined from the μ and σ values was 0.16%.

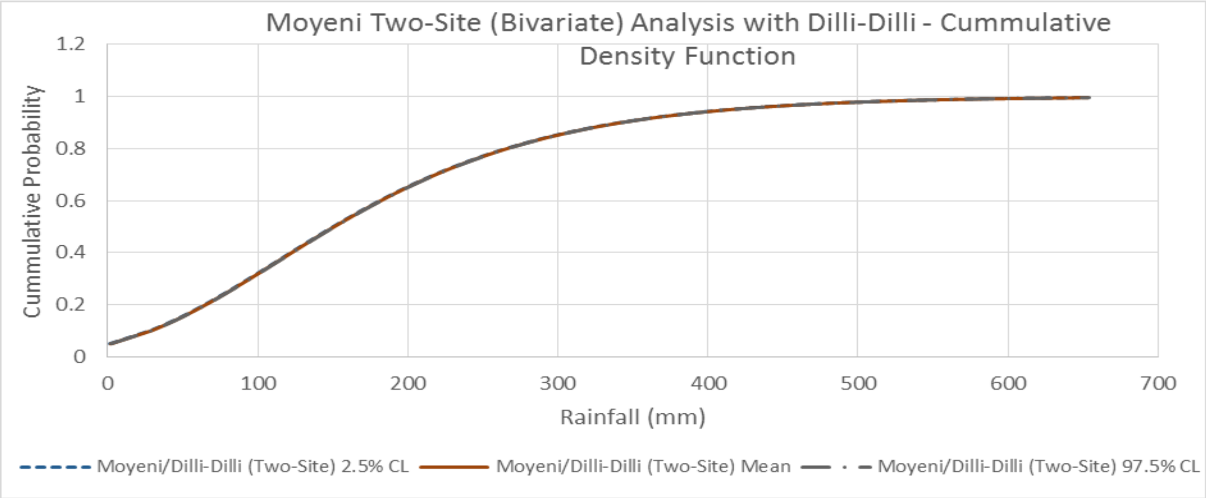


Figure 4-32: Moyeni Two-Site (Bivariate) Analysis with Dilli-Dilli: Cumulative Probability

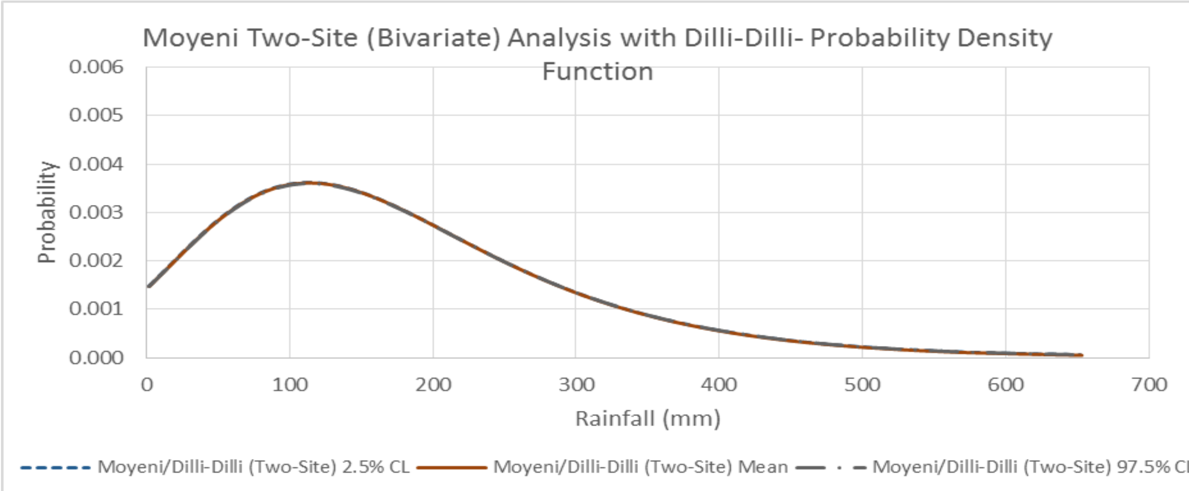


Figure 4-33: Moyeni Two-Site (Bivariate) Analysis with Dilli-Dilli: Probability

The resulting rainfall events from the two-site analysis of Seaka with Moyeni are shown in Table 4-22 and the corresponding cumulative density plots and probability density plots are shown in Figure 4-34 and Figure 4-35.

Table 4-22: Two-Site Rainfall Events and Return Periods at various Significance Levels – Seaka (paired with Moyeni)

		Seaka Two-Site (Bivariate) Analysis with Moyeni		
		2.5% CL	**Mean**	**97.5% CL**
mu		65.213	65.547	65.882
sigma		69.512	69.838	70.164
Return period	**2 Years**	90.690	91.144	91.598
	5 Years	169.477	170.300	171.124
	10 Years	221.641	222.709	223.777
	20 Years	271.678	272.980	274.283
	50 Years	336.445	338.052	339.658
	100 Years	384.979	386.813	388.647
	200 Years	433.337	435.397	437.458

There was an average difference of 0.49% in values of μ and σ between the 2.5% and 97.5% confidence interval and the mean values. The resulting average difference in values of rainfall determined from the μ and σ values was 0.48%.

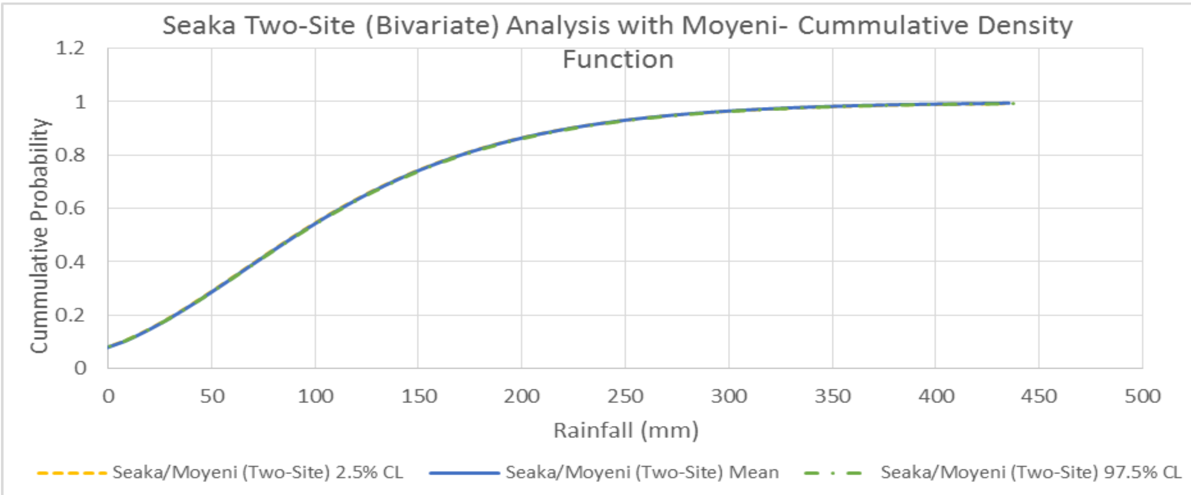


Figure 4-34: Seaka Two-Site (Bivariate) Analysis with Moyeni: Cumulative Probability

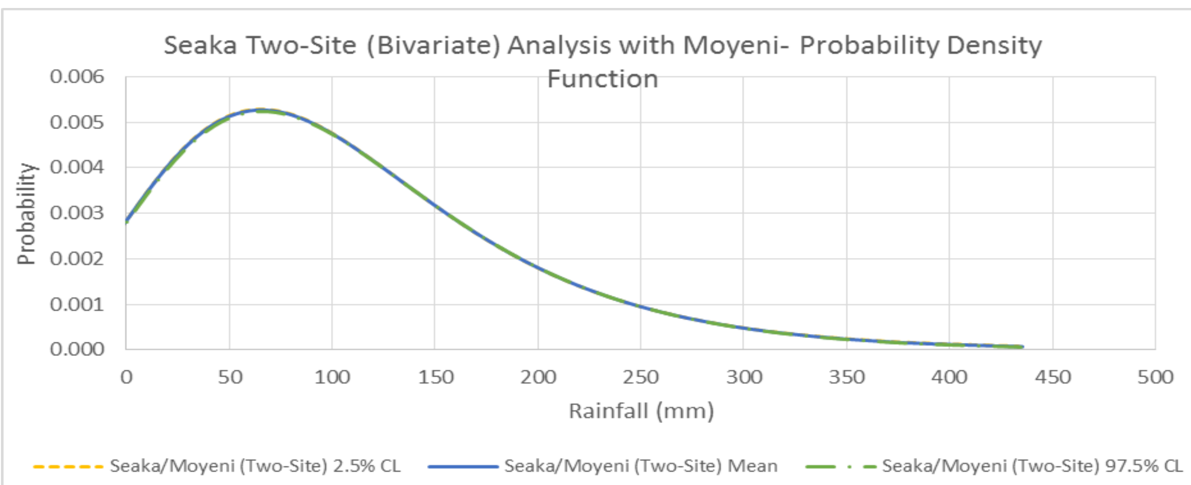


Figure 4-35: Seaka Two-Site (Bivariate) Analysis with Moyeni: Probability

The resulting rainfall events from the two-site analysis of Dilli-Dilli with Moyeni are shown in Table 4-23 and the corresponding cumulative density plots and probability density plots are shown in Figure 4-36 and Figure 4-37.

Table 4-23: Two-Site Rainfall Events and Return Periods at various Significance Levels – Dilli-Dilli (paired with Moyeni)

		Dilli-Dilli Two-Site (Bivariate Analysis) with Moyeni		
		2.5% CL	**Mean**	**97.5% CL**
mu		85.960	86.895	87.829
sigma		101.123	101.530	101.937
Return period	**2 Years**	123.023	124.107	125.190
	5 Years	237.639	239.184	240.728
	10 Years	313.524	315.375	317.225
	20 Years	386.316	388.459	390.602
	50 Years	480.536	483.059	485.581
	100 Years	551.142	553.948	556.754
	200 Years	621.489	624.579	627.668

There was an average difference of 0.74% in values of μ and σ between the 2.5% and 97.5% confidence interval and the mean values. The resulting average difference in values of rainfall determined from the μ and σ values was 0.60%.

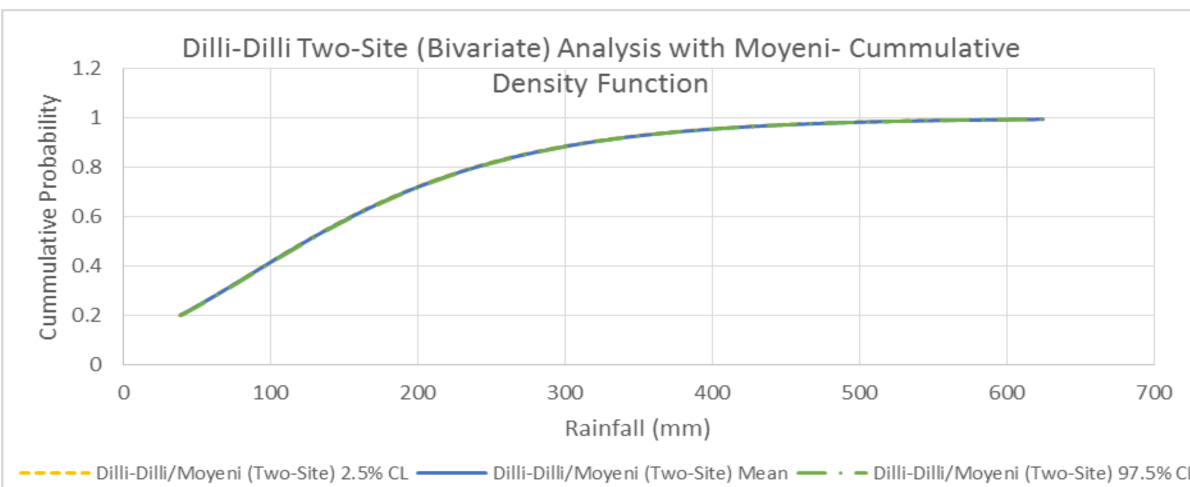


Figure 4-36: Dilli-Dilli Two-Site (Bivariate) Analysis with Moyeni: Cumulative Probability

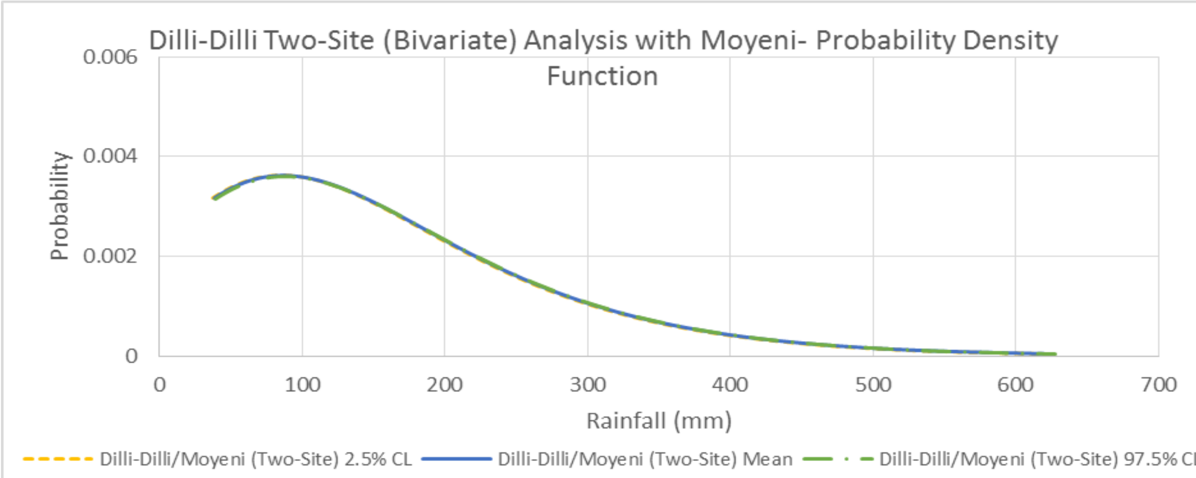


Figure 4-37: Dilli-Dilli Two-Site (Bivariate) Analysis with Moyeni: Probability

The difference between the values predicted using values of μ and σ at the 2.5% confidence level, mean and the 97.5% confidence level did not vary significantly for all three rainfall stations. However the deviation is smallest for Moyeni and biggest for Dilli-Dilli. This was consistent with the observed decrease in precision with the decrease in record length highlighted in section 4.2.2.

4.3. Comparison of Single-Site Analysis results with Two-Site Analyses results

This section shows comparisons made of single-site (univariate) and two-site (bivariate) analyses results; these are results of inferred values of μ and σ and results of predicted rainfall events are compared.

4.3.1. Comparison of Inferred values of μ and σ: Single-Site vs Two-Site Approach

Figure 4-38 and Figure 4-39 show the distribution of posterior values μ and σ for Moyeni as determined from both the univariate and the bivariate analyses, while Figure 4-40 and Figure 4-41 show the values of the coefficient of variation and the posterior mean for both the univariate and the bivariate analyses.

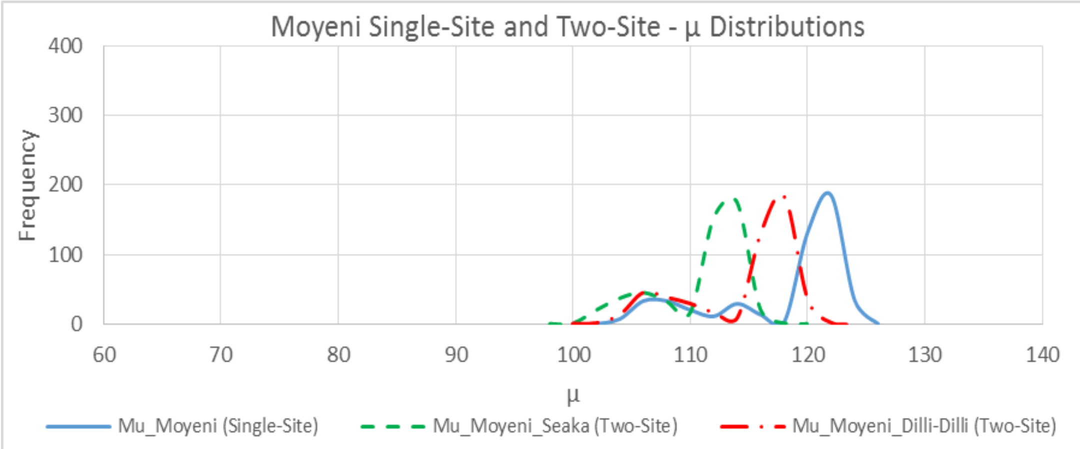


Figure 4-38: Moyeni Single-Site and Two-Site: µ Distributions

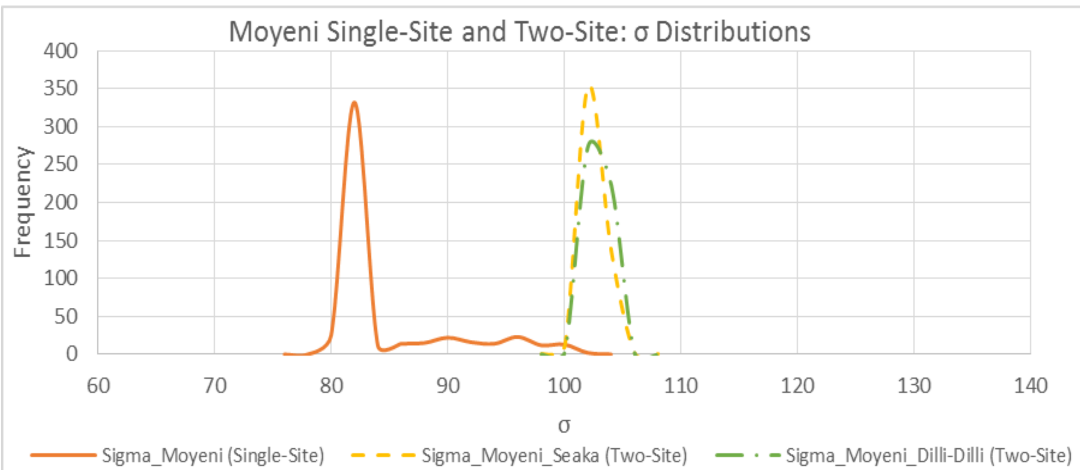


Figure 4-39: Moyeni Single-Site and Two-Site: σ Distributions

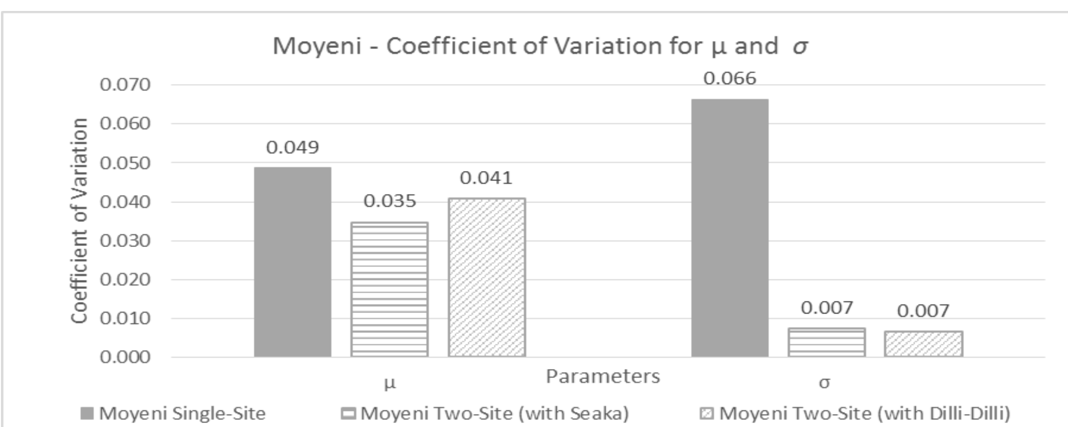


Figure 4-40: Single-Site and Two-Site Coefficient of Variation for Moyeni

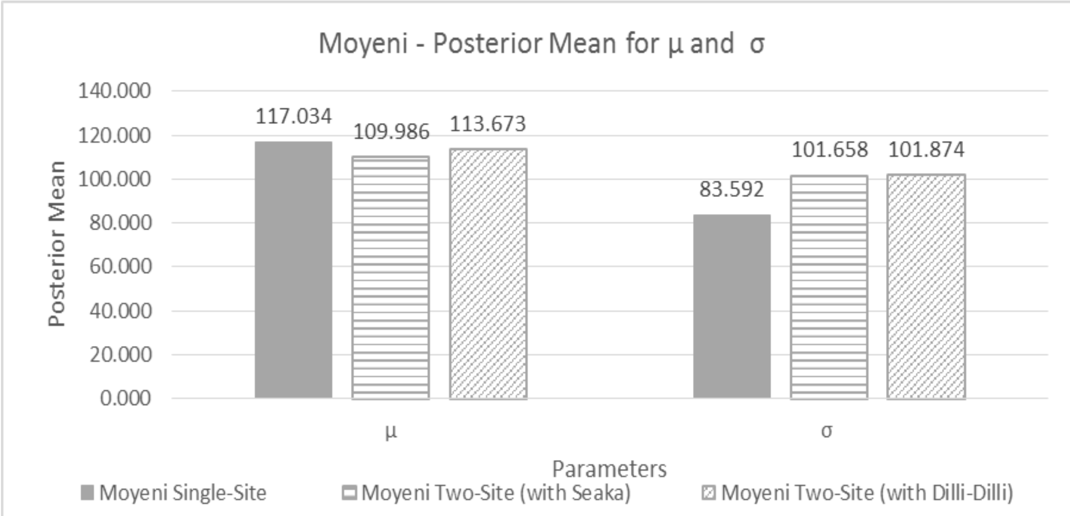


Figure 4-41: Single-Site and Two-Site Posterior Mean for Moyeni

The coefficient of variation for μ decreased by 29% and 14% for Moyeni Two-Site (with Seaka) and Moyeni Two-Site (with Dilli-Dilli) respectively when determined from the bivariate analyses, while the coefficient of variation for σ decreased by 89% for both Moyeni Two-Site (with Seaka) and Moyeni Two-Site (with Dilli-Dilli) when determined from the bivariate analyses. The posterior mean for μ decreased (as seen in the shift of the frequency distributions to the left in Figure 4-38) by 6% and 2.9% for Moyeni Two-Site (with Seaka) and Moyeni Two-Site (with Dilli-Dilli) respectively when determined from the bivariate analyses, while the posterior mean of σ increases (as seen in the shift of the frequency distributions to the right in Figure 4-39) by 22% for both Moyeni Two-Site (with Seaka) and Moyeni Two-Site (with Dilli-Dilli) when determined from the bivariate analyses.

Figure 4-42 and Figure 4-43 below show the distribution of posterior values μ and σ for Seaka as determined from both the univariate and the bivariate analyses, while Figure 4-44 and Figure 4-45 show the values of the coefficient of variation and the posterior mean for both the univariate and the bivariate analyses.

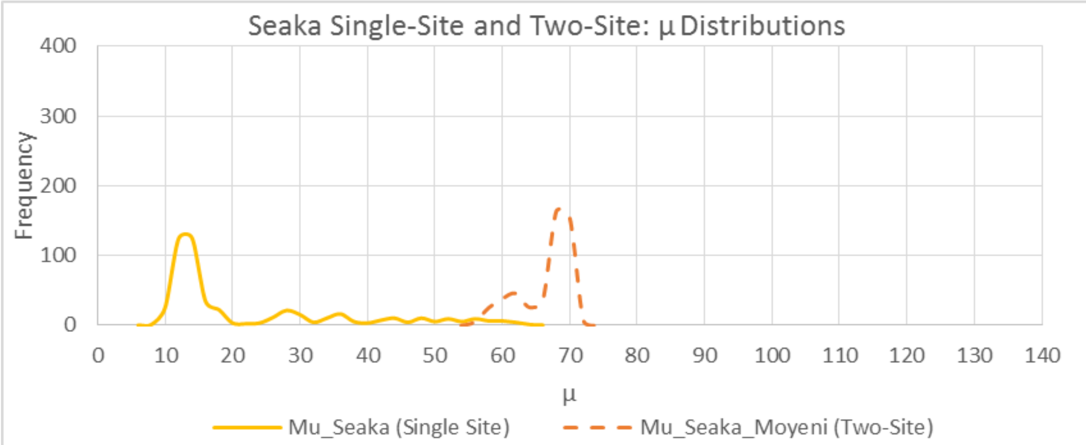


Figure 4-42: Seaka Single-Site and Two-Site: µ Distributions

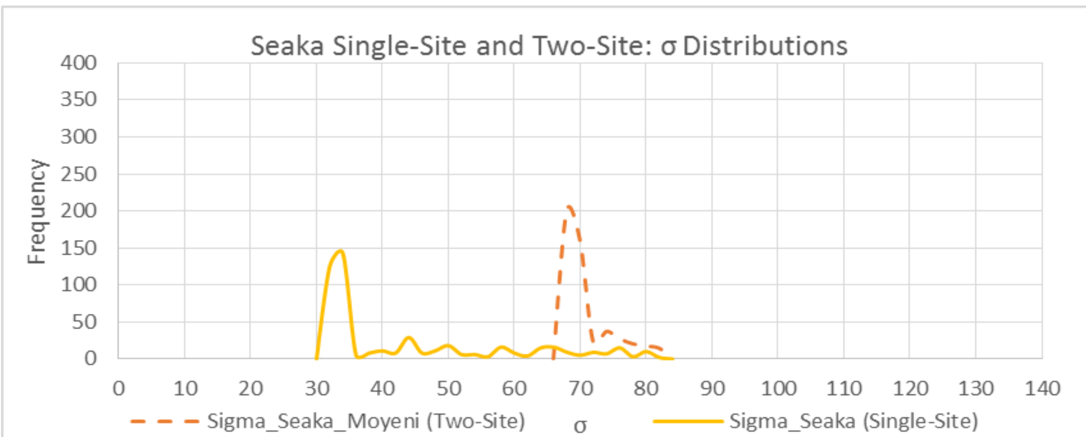


Figure 4-43: Seaka Single-Site and Two-Site: σ Distributions

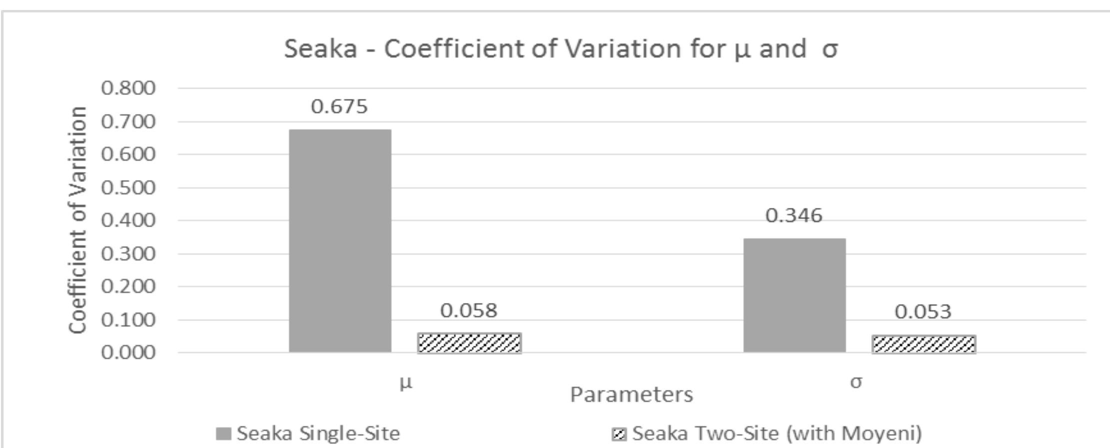


Figure 4-44: Single-Site and Two-Site Coefficient of Variation for Seaka

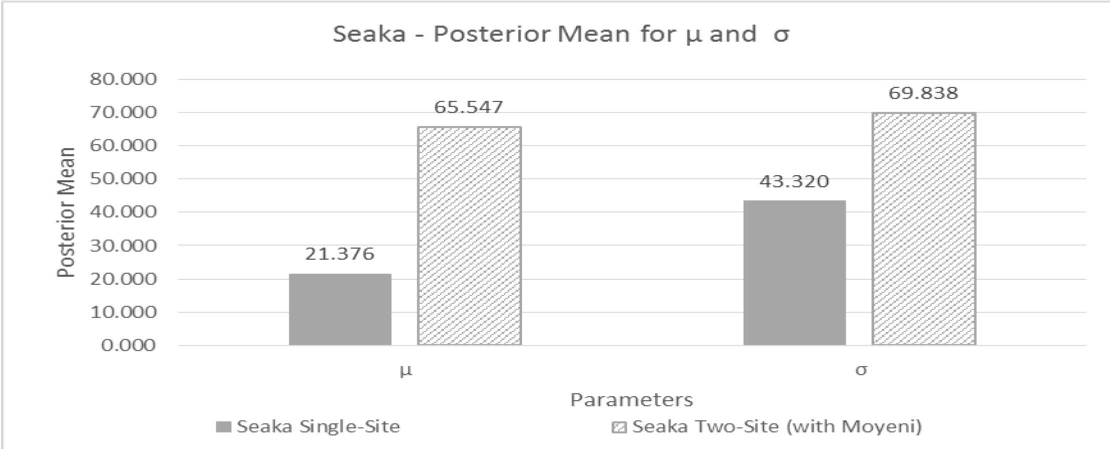


Figure 4-45: Single-Site and Two-Site Posterior Mean for Seaka

The coefficient of variation for µ and σ decreased by 91% and 85%, respectively, for Seaka Two-Site when determined from the bivariate analyses. The posterior mean for µ and σ increased (as seen in the shift of the frequency distributions to the right in Figure 4-42 and Figure 4-43) by 207% and 61%, respectively, for Seaka Two-Site when determined from the bivariate analyses.

Figure 4-46 and Figure 4-47 below show the distribution of posterior values µ and σ for Dilli-Dilli as determined from both the univariate and the bivariate analyses, while Figure 4-48 and Figure 4-49 show the values of the coefficient of variation and the posterior mean for both the univariate and the bivariate analyses.

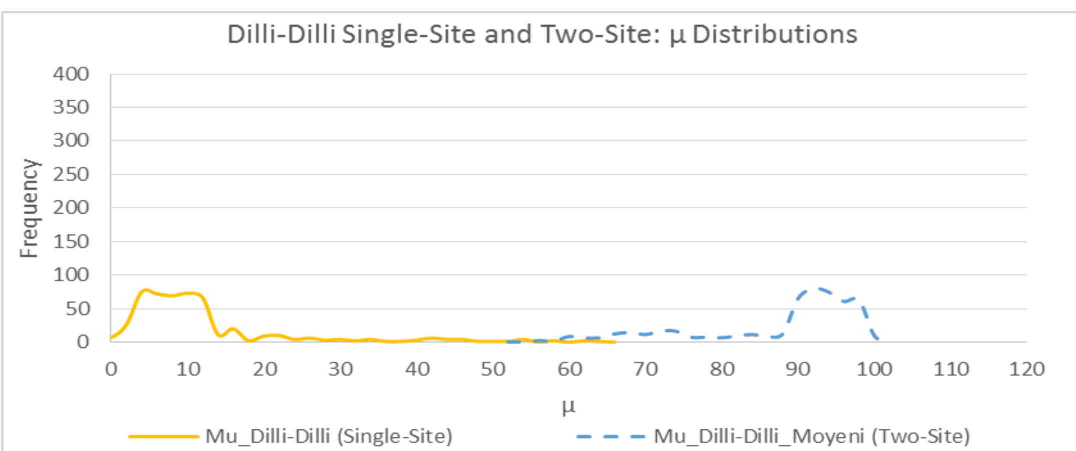


Figure 4-46: Dilli-Dilli Single-Site and Two-Site: µ Distributions

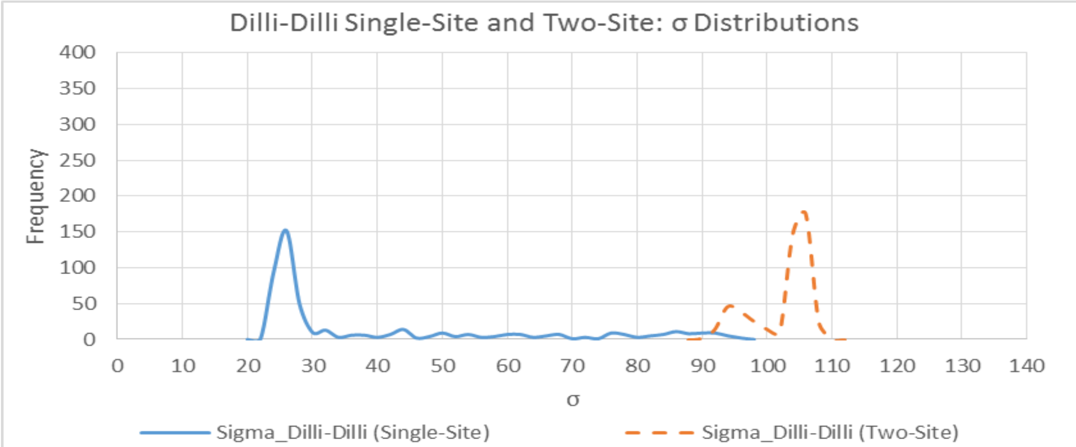


Figure 4-47: Dilli-Dilli Single-Site and Two-Site: σ Distributions

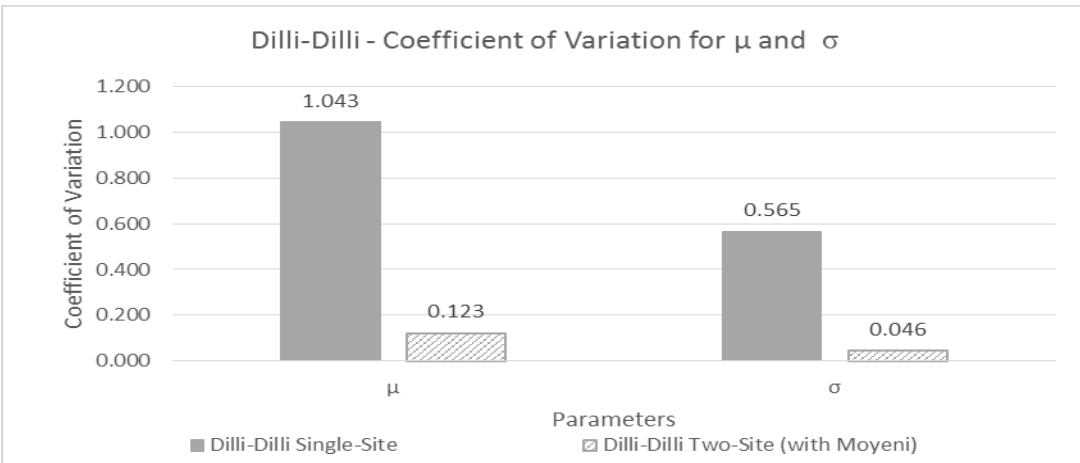


Figure 4-48: Single-Site and Two-Site Coefficient of Variation for Dilli-Dilli

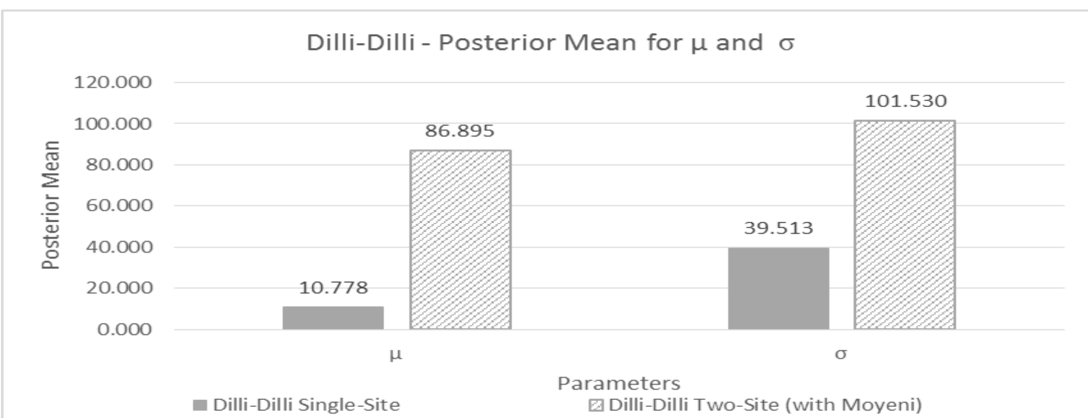


Figure 4-49: Single-Site and Two-Site Posterior Mean for Dilli-Dilli

The coefficient of variation for μ and σ decreased by 88% and 92%, respectively, for Dilli-Dilli Two-Site when determined from the bivariate analyses. The posterior mean for μ and σ increases (as seen in the shift of the frequency distributions to the right in Figure 4-46 and Figure 4-47) by 706% and 157%, respectively, for Dilli-Dilli Two-Site when determined from the bivariate analyses.

4.3.2. Comparison of Simulated Extreme Rainfall Events: At-Site vs Two-Site Approach

Table 4-24 shows the mean values of μ and σ and results of the frequency analyses carried out using mean values of the parameters determined from the univariate and bivariate analyses, for Moyeni rainfall station.

Table 4-24: Single-Site and Two-Site mean values of parameter σ and μ, and predicted rainfall events for Moyeni

		Moyeni Single-Site	Moyeni Two-Site (Bivariate) Analysis with Seaka	Moyeni Two-Site (Bivariate Analysis) with Dilli-Dilli
μ		117.03	109.99	113.67
σ		83.59	101.66	101.87
Return Period (Years)	**2**	147.67	147.24	151.01
	5	242.42	262.47	266.48
	10	305.15	338.75	342.93
	20	365.32	411.93	416.26
	50	443.21	506.65	511.18
	100	501.57	577.63	582.31
	200	559.72	648.35	653.18

Figure 4-50 is a graphical representations of predicted rainfall depths for various return periods, determined through univariate and bivariate analyses, for Moyeni rainfall station.

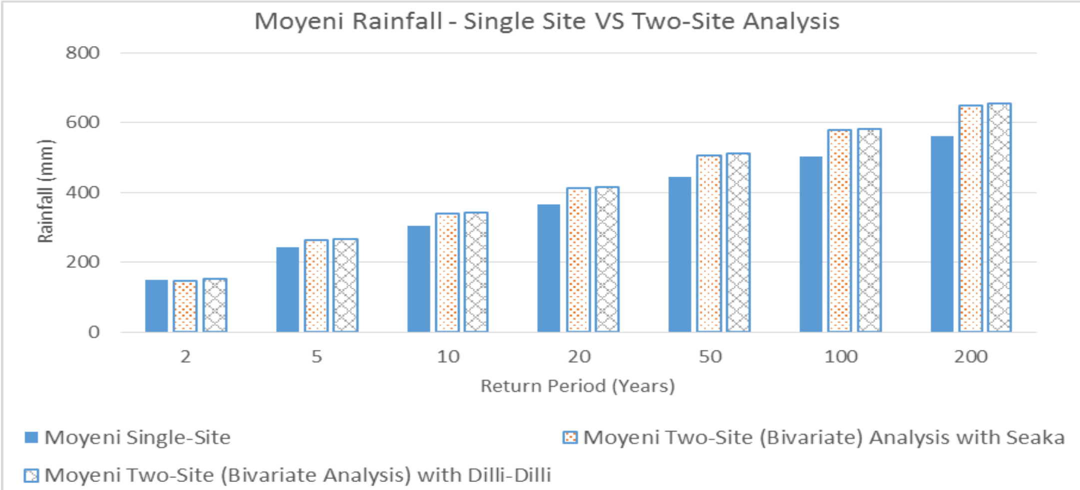


Figure 4-50: Single-Site and Two-Site values of predicted rainfall events for Moyeni

Table 4-25 shows the mean values of μ and σ and results of the frequency analyses carried out using mean values of the parameters determined from the univariate and bivariate analyses, for Seaka rainfall station.

Table 4-25: Single-Site and Two-Site mean values of parameter σ and μ, and predicted rainfall events for Seaka

		Seaka Single-Site	Seaka Two-Site (Bivariate) Analysis with Moyeni
μ		21.38	65.55
σ		43.32	69.84
Return Period (Years)	**2**	37.25	91.14
	5	86.35	170.30
	10	118.86	222.71
	20	150.04	272.98
	50	190.41	338.05
	100	220.65	386.81
	200	250.79	435.40

Figure 4-51 is a graphical representations of predicted rainfall depths for various return periods, determined through univariate and bivariate analyses, for Seaka rainfall station.

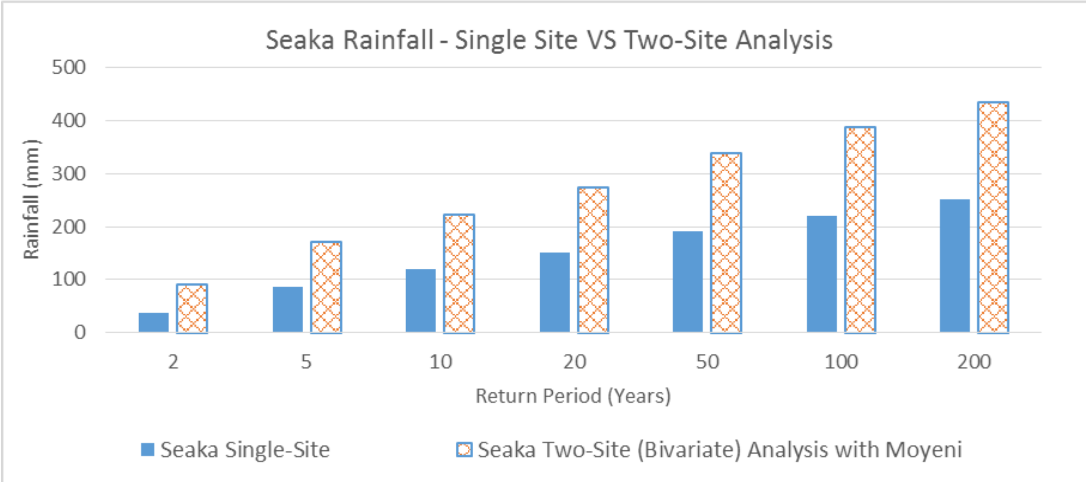


Figure 4-51: Single-Site and Two-Site values of predicted rainfall events for Seaka

Table 4-26 show the mean values of μ and σ and results of the frequency analyses carried out using mean values of the parameters determined from the univariate and bivariate analyses, for each rainfall station.

Table 4-26: Single-Site and Two-Site mean values of parameter σ and μ, and predicted rainfall events for Dilli-Dilli

<table>
<tr><th colspan="2"></th><th>Dilli-Dilli Single-Site</th><th>Dilli-Dilli Two-Site (Bivariate Analysis) with Moyeni</th></tr>
<tr><td colspan="2">μ</td><td>10.78</td><td>86.89</td></tr>
<tr><td colspan="2">σ</td><td>39.51</td><td>101.53</td></tr>
<tr><td rowspan="7">Return Period (Years)</td><td>2</td><td>25.26</td><td>124.11</td></tr>
<tr><td>5</td><td>70.04</td><td>239.18</td></tr>
<tr><td>10</td><td>99.70</td><td>315.37</td></tr>
<tr><td>20</td><td>128.14</td><td>388.46</td></tr>
<tr><td>50</td><td>164.96</td><td>483.06</td></tr>
<tr><td>100</td><td>192.54</td><td>553.95</td></tr>
<tr><td>200</td><td>220.03</td><td>624.58</td></tr>
</table>

Figure 4-52 is a graphical representations of predicted rainfall depths for various return periods, determined through univariate and bivariate analyses, for Dilli-Dilli rainfall station.

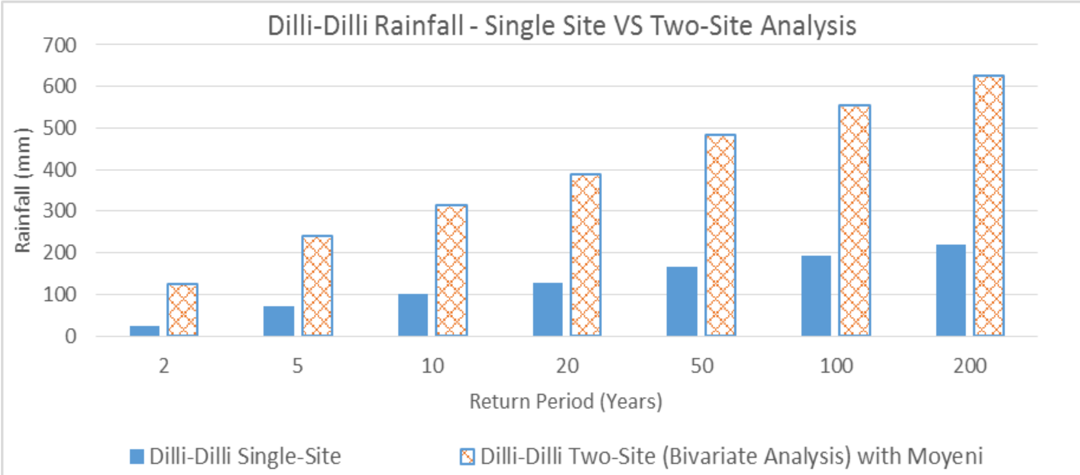


Figure 4-52: Single-Site and Two-Site values of predicted rainfall events for Dilli-Dilli

Figure 4-53, Figure 4-54 show the probability and cumulative probability plots as generated from parameter mean values from the Moyeni univariate and the bivariate analyses.

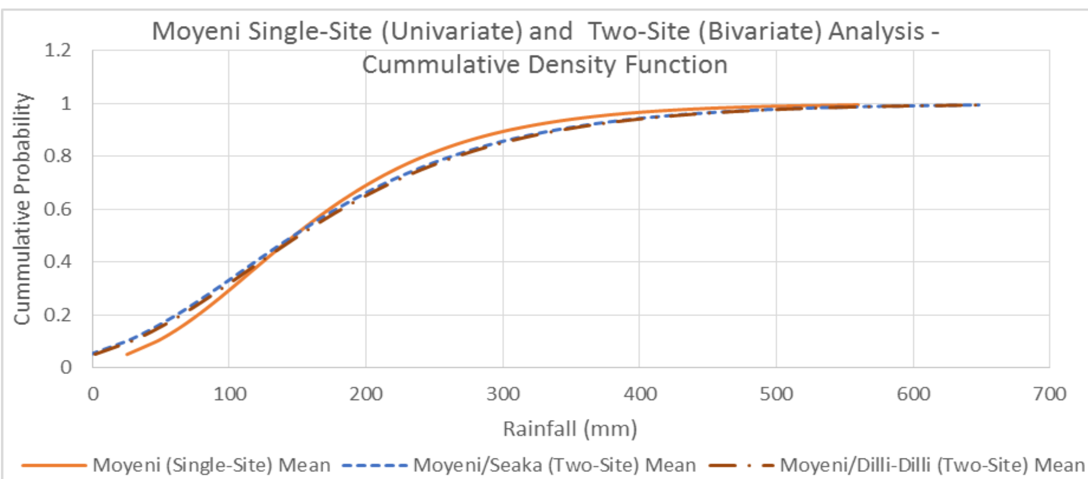


Figure 4-53: Single-Site vs Two-Site Analysis: Cumulative Probability - Moyeni Rainfall Station

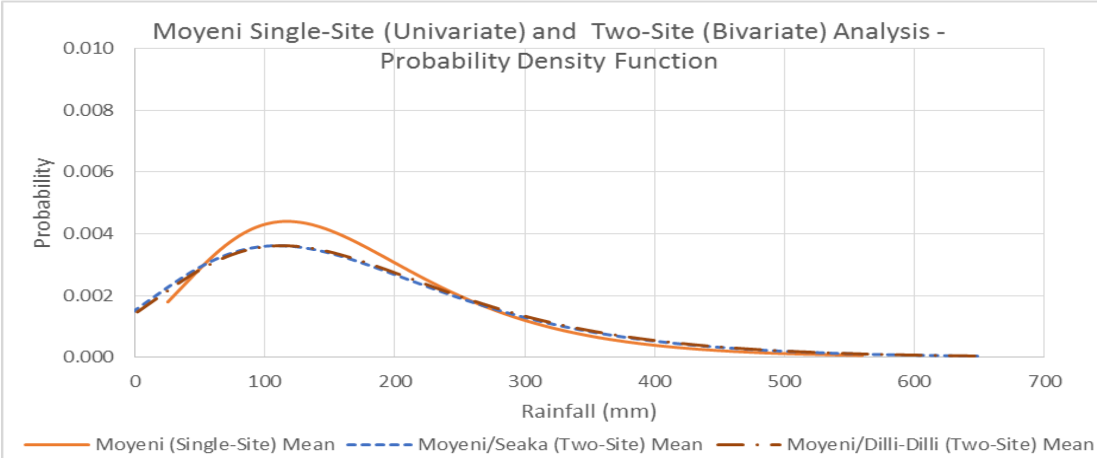


Figure 4-54: Single-Site vs Two-Site Analysis: Probability - Moyeni Rainfall Station

Figure 4-55 and Figure 4-56 show the probability ad cumulative probability plots as generated from parameter mean values from the Seaka univariate and the bivariate analyses.

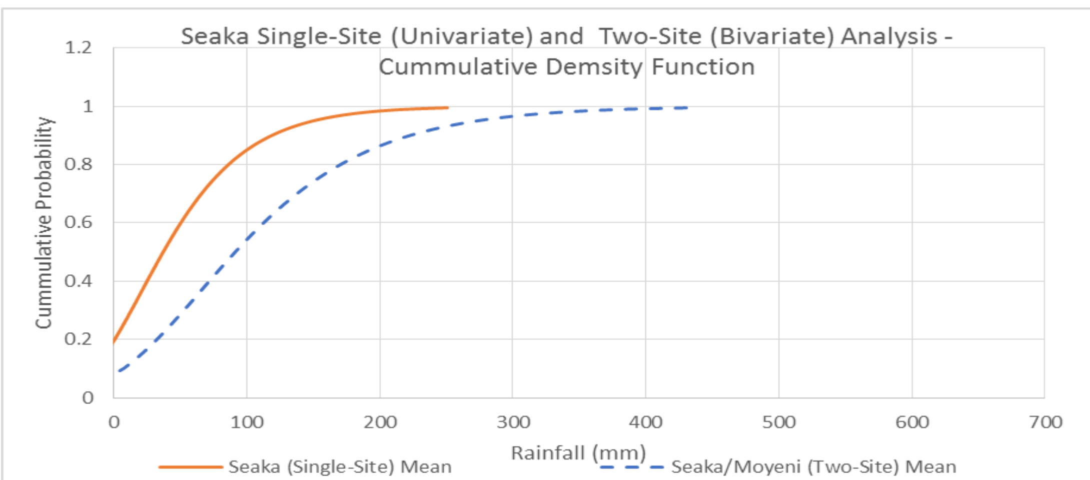


Figure 4-55: Single-Site vs Two-Site Analysis: Cumulative Probability - Seaka Rainfall Station

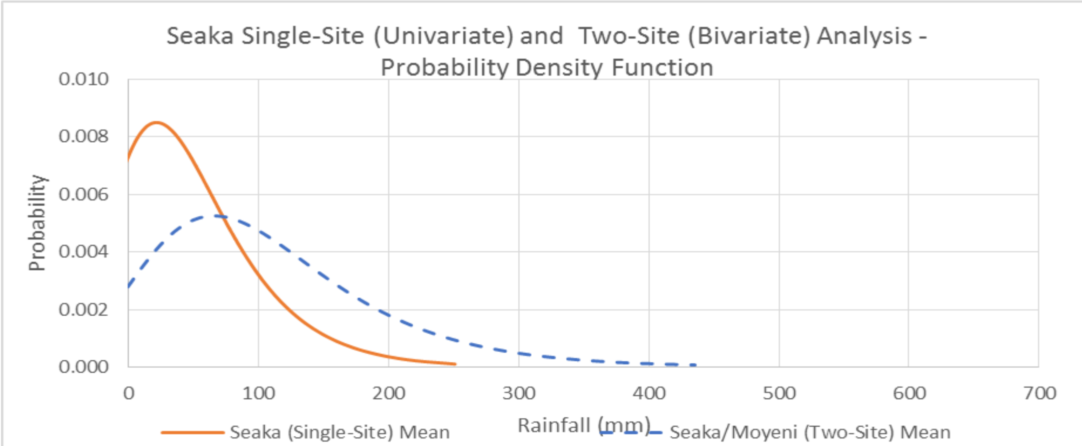


Figure 4-56: Single-Site vs Two-Site Analysis: Probability - Seaka Rainfall Station

Figure 4-57 and Figure 4-58 show the probability ad cumulative probability plots as generated from parameter mean values from the univariate and the bivariate analyses.

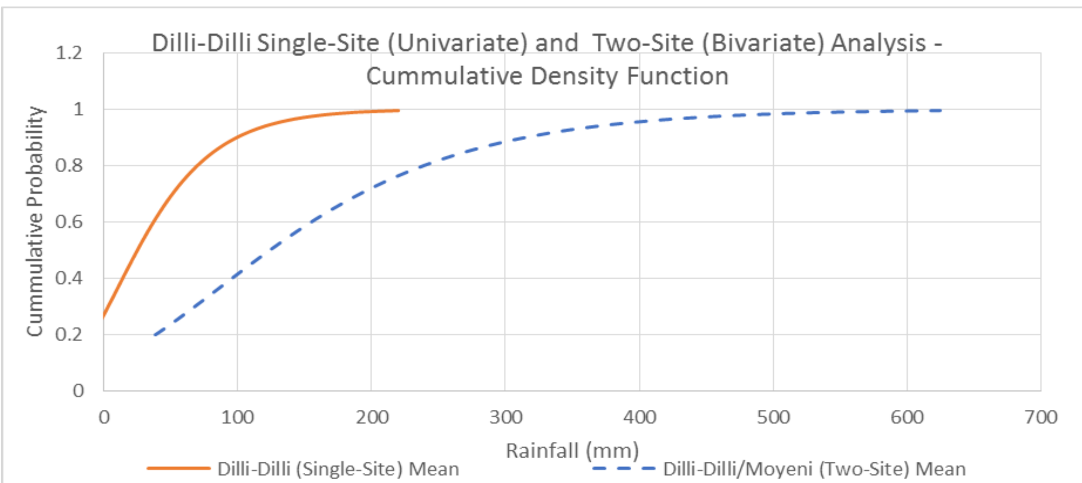


Figure 4-57: Single-Site vs Two-Site Analysis: Cumulative Probability – Dilli-Dilli Rainfall Station

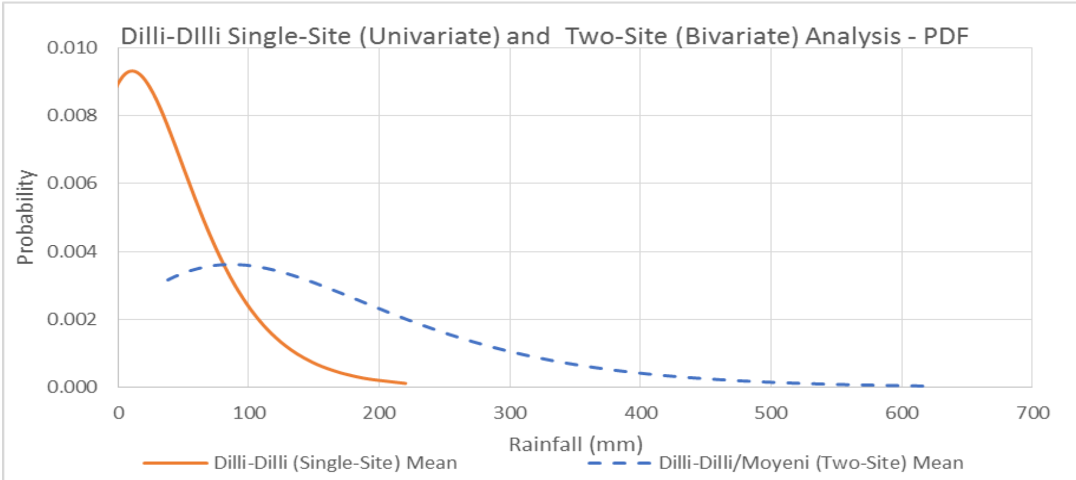


Figure 4-58: Single-Site vs Two-Site Analysis: Probability - Seaka Rainfall Station

In the PDFs shown in Figure 4-54, Figure 4-56 and Figure 4-58 above, the bivariate PDFs were flatter than the univariate PDF; this was typical of GEV distributions where an increase in σ (scale parameter) resulted in a flatter PDF. The figures also showed a slight shift to the left by Moyeni's bivariate PDF and a shift to the right by Seaka and Dilli-Dilli's bivariate PDFs; the shift to the left indicated a decrease in the mean rainfall while a shift to the right indicates an increase in rainfall.

There was an overall increase in the values of predicted rainfall depths when the parameters determined from the bivariate analyses were used. Moyeni rainfall depth values showed the least increase, while the Dilli-Dilli rainfall depths values show the greatest increase.

5. DISCUSSION

Results, which include inferred values of the GEV distribution parameters, together with the frequency analyses (using the inferred parameters) outputs, from the single-site (univariate) and two-site (bivariate) analyses are summarised and discussed in this section. A discussion of the findings from the comparison of results from univariate versus bivariate analyses is also undertaken around the issue of possible improvement in precision of parameters when using the bivariate approach.

5.1. Single-Site Results

The estimated values of μ and σ, determined from the method of L-moments, were used as initial values in the Bayesian univariate analyses of each rainfall station.The summary results of the simulation, which included the posterior mean, standard deviation, coefficient of variation and 95% confidence interval, and the distribution of posterior values of the parameters showed a general increase in variability (decrease in precision) of the estimated values of μ and σ, with decrease in data record length.

For each pair of μ and σ, a frequency analysis was undertaken to predict the magnitude of rainfall at various return periods. The difference between the values predicted using values of μ and σ at the 2.5% confidence level, mean and the 97.5% confidence level did not vary significantly (maximum 7.1% deviation) for all three rainfall stations. The deviation was smallest for Moyeni and biggest for Dilli-Dilli. This was consistent with the observed decrease in precision with the decrease in record length.

5.2. Two-Site Results

The same initial values used in the Bayesian single-site (univariate) analyses were used in the two-site (bivariate) analyses. The summary results of the simulation, which included the posterior mean, standard deviation, coefficient of variation and 95%

confidence interval for each simulated parameter in the two-site analyses showed a general increase in variability (decrease in precision) of the estimated values of μ and σ, with decrease in data record length, as in the single-site analyses. For each pair of rainfall stations and the corresponding μ and σ for each station, a frequency analysis was undertaken to determine the magnitude of rainfall at various return periods. The difference between the values predicted using values of μ and σ at the 2.5% confidence level, mean and the 97.5% confidence level was not significant (maximum 0.74%) for all three rainfall stations.

Sensitivity analyses were undertaken to check sensitivity of simulation results to the initial parameter values used. The results showed that μ was less sensitive than sigma.

5.3. Comparison of Single-Site Analysis results with Two-Site Analysis results

The comparison revealed that values for μ for Moyeni decreased by 3-6% when determined from the bivariate analyses, with the most decrease realised when Moyeni was paired with Seaka. The values of σ for Moyeni increased by 22% when determined from the bivariate analyses. There was an overall increase in the values of μ and σ for Seaka and Dilli-Dilli when the parameters were determined using the bivariate analyses, with the most significant average increase realised for Dilli-Dilli. The significance of the type of analysis on the value of σ for Moyeni was not anticipated but the significant impact of the type of analysis on the parameter values of Seaka and Dilli-Dilli was expected; the anticipation was that impact would be realised mostly for rainfall stations that had short duration data.

The precision of parameters (as measured by the value of the coefficient of variation) increased when the parameters were determined from the bivariate analyses. The precision of the value of μ for Moyeni was least affected (14-29% decrease in variability) by the type of analysis, compared to the 88-92% decrease in variability for

Seaka and Dilli-Dilli. This was in line with the anticipation that combining data in a bivariate analysis would improve the precision of the estimated parameter values.

There was an overall increase in the values of predicted rainfall depths when the parameters determined from the bivariate analyses were used. Moyeni rainfall values showed the least increase (2-16%), while the Dilli-Dilli rainfall values showed the greatest increase (183-391%). Seaka rainfall depth values increased by 73-144%. This was in line with the observed increase in μ values when using the bivariate analysis.

6. CONCLUSIONS

The objective of the study was to establish a method of improving hydrological data using data from a gauged (with long periods of record) catchment to improve data from an ungauged (with short periods of record) catchment. The parameters of the assumed distribution (Type I GEV distribution) were estimated using the Bayesian – Markov Chain Monte Carlo approach. For each catchment, statistical location (μ) and scale parameters (σ) were determined using both the single-site (univariate) estimation and the two-site (bivariate) estimation and the results were compared in order to establish if there was improvement in the estimate of the parameters when the bivariate approach was used.

Results of the analyses revealed that the precision of the parameters was higher for the catchment with the longest period of record, and lower for catchments with the shorter periods of record. The precision of the parameter estimation also improved when the bivariate approach was used, with the biggest improvement observed for the catchments with shorter periods of record. In addition to the precision of the parameters improving when using the bivariate approach, the values the parameters generally increased which in turn increased the predicted rainfall depths. Predicted rainfall values increased, with considerable increases (greater than 300%) realised for Dilli-Dilli, revealing the large uncertainties in design rainfall estimation that could occur especially when short records were used.

It can be concluded that the objective of the study has been achieved and the findings revealed the anticipated improvement in precision when short length data was augmented with longer length data in bivariate analysis. In more simple terms, the bivariate approach enabled more precise predictions of extreme rainfall applicable to flood design.

7. RECOMMENDATIONS

As outlined in the Section 6, the objectives of the study were achieved; and it was found that the bivariate approach can be used in practice by hydrologists and design engineers to enhance precision of extreme rainfall estimation for use in flood design and assessments in order to reduce uncertainties caused by insufficient or poor quality data.

There is however need for additional investigations to be undertaken along the lines of the impact of this approach on the catchment with long duration data. It may also be useful to carry out more rigorous investigations of how the initial value the Monte Carlo Markov Chain impacts the analyses and how the prior knowledge of the distribution of the parameters impacts the results of the Bayesian analysis. Additionally, it may also be worth using data from other rainfall stations with a wider range in data lengths, using various probability distributions to thoroughly investigate the phenomenon of data correlation among catchments in the same area.

8. REFERENCES

Ali, M. (2013). *Markov Chain Monte Carlo (MCMC).* Notes. University of the Witwatersrand, Johannesburg. (Unpublished).

Balakrishnan, N., Lai, C., (2009). *Continuous Bivariate Distributions.* 2nd Edition. New York: Springer Science + Business Media, p. 33-38.

Beven, K.J. (2000). *Rainfall-runoff Modelling.* New York: John Wiley.

Bolgov, M.V.et.al. (eds). (2002). *Hydrological Models for Environmental Management*. Netherlands: Kluver Academic Publishers, p. 1-15.

Bolstad, W.M. (2010). *Understanding Computational Bayesian Statistics.* Hoboken, N.J.: Wiley.

Box, G.E.P. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Mass: Addison-Wesley.

Campbell, E.P., Fox, D.R. and Bates, B.C. (1999). *A Bayesian Approach to Parameter Estimation and Pooling in Non-linear Flood Event Models*. Water Resources Research, 35(1), p. 211-220.

Chadwick, A., Morfett, J. and Borthwick, M. (2004). *Hydraulics in Civil and Environmental Engineering*. 4th Edition. London: Spon Press, p. 309-328.

Chow, V.T., Maidment, D.R., and Mays, L.W. (1988). *Applied Hydrology.* Singapore: McGraw-Hill Book Co., p. 380-405.

Chung, E., Kim, S.U. (2013). *Bayesian Rainfall Frequency Analysis with Extreme Value using the Informative Prior Distribution.* KSCE Journal of Civil Engineering, 17(6), p. 1502-1514.

De Michele, C., Slavadori, G., Canossi, M., Petaccia, A. and Rosso, R. (2005). *Bivariate Statistical Approach to Check Adequacy of Dam Spillway*. Journal of Hydrologic Engineering 10(1).

Escalante-Sandoval, C. (2006). *Application of Bivariate Extreme Value Distribution to Flood Frequency Analysis: A Case Study of North-Western Mexico.* Nat. Hazards, 42, p. 37-46.

Gamerman, D. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Boca Raton: Taylor & Francis.

Gaume, E., Gaal, L., Viglione, A., Szolgay, J., Kohnova, S. and Bloschl, G. (2010). *Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites*. Journal of Hydrology, 394, p 101-117.

Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Science*. Cambridge: Cambridge University Press.

Haan, C.T. (1977). *Statistical Methods in Hydrology.* Iowa: State University Press.

Hrachowitz, M, Savenije, HHG, Blöschl, G, McDonnell, JJ, Sivapalan, M, Pomeroy, JW, Arheimer, B, Blume, T, Clark, MP, Ehret, U, Fenicia, F, Freer, JE, Gelfan, A, Gupta, HV, Hughes, DA, Hut, RW, Montanari, A, Pande, S, Tetzlaff, D, Troch, PA, Uhlenbrook, S, Wagener, T, Winsemius, HC, Woods, RA, Zehe, E and Cudennec, C 2013. *A decade of Predictions in Ungauged Basins (PUB)-a review*. Hydrological Sciences Journal, 58(6), p. 1198-1255.

Kotz, S. (2000). *Extreme Value Distributions: Theory and Applications.* London: Imperial College.

Liang, Z.M., Li, B.Q., Yu, Z.B. et.al. (2011). *Application of Bayesian Approach to Hydrological Frequency Analysis*. Science China Technological Sciences, 54, p. 1183-1192.

Matalas, N.C. and Jacobs, B., (1964). *A Correlation Procedure for Augmenting Hydrological Data*. U.S. Geological Survey Paper 434-E, p. E1.

Martins, E.S. and Stedinger, J.R. (2000). *Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data*. Water Resources Research, 36(3), p 737-744.

McCuen, R.H. (2003). *Modelling Change: Statistical Methods*. Florida: Lewis Publishers.

Mitkova, V.B., Halmova,D. (2014). *Joint Modelling of Flood Peak Discharges, Volume and Duration: A Case Study of the Danube River in Bratislava*. J. Hydrol. Hydromech. 62(3), p. 186-196.

The Office of the Commissioner of Water, Lesotho (2013). *State of Water Resources Report, April 2011 – March 2013.* Report. Ministry of Energy, Meteorology and Water Affairs, Lesotho. (Unpublished).

Press, S.J. (1989). *Bayesian Statistics: Principles, Models and Applications.* New York: Wiley.

Reis, D.S. and Stedinger, J. R. (2005). *Bayesian MCMC flood frequency analysis with historical information*. Journal of Hydrology, 313, p 97-116.

Reiss, R.D. (2007). *Statistical Analysis of extreme Value: with applications to insurance, finance, hydrology and other fields.* Basel Boston: Birkhauser.

Signh, V.P, Vijay, P. (1998). *Entropy-Based Parameter Estimation in Hydrology*. Dodrecht: Kluwer.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshimi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S. and Zehe, E. (2003). *IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an Exciting Future for the Hydrological Sciences*. Hydrological Sciences-Journal, 48(6), p. 860.

StataCorp. (2015). *Stata release 14: Statistical Software.* College Station, Texas: StataCorp LP.

Vogel, R.M. and Kroll, C.N. (1990). *The Value of Streamflow Record Augmentation Procedures in Low-Flow and Flood-Flow Frequency Analysis*. Journal of Hydrology, 125, p. 259-276.

Vujica, Y. (1972). *Probability and Statistics in Hydrology.* Water Resources.

Wang, Q.J. (2001). *A Bayesian Joint-Probability Approach for Flood Record Augmentation*. Water Resources Research, 37(6), p. 1707-1712.

Xu, Y., Huang, G. and Fan, Y. (2015). *Multivariate Flood Risk Analysis of Wei River.* Stoch Environ Res Risk Assess. DOI 10.1007/s00477-015-1196-0.

Yang, X., Zhang, Q. (2013). *Joint Probability Distribution of Winds and Waves from Wave Simulation of 20 years (1989-2008) in Bohai Bay.* Water Science and Engineering, 6(3), p. 296-307.

9. APPENDICES

I. APPENDIX 1: USING EXCEL FOR BAYESIAN ANALYSIS

Baye's Theorem:

$$Posterior \propto Likelihood \times Prior$$

Posterior - the probability density of θ (parameters of the probability distribution of the rainfall data) with given observations (sample) of *x:*

$$p(\theta|x)$$

Likelihood – the likelihood function of sample x (probability density function of *x)*, conditional on the parameters $\theta = (\mu, \sigma)$*,* which is given by a Type I Generalised Extreme Value (GEV) probability distribution:

$$f(x|\theta) = \frac{1}{\sigma}\left[\mathrm{e}^{-\left({(\mathrm{x}-\mu)}/{\sigma}\right)}\right] e^{-\left(\mathrm{e}^{-{(\mathrm{x}-\mu)}/{\sigma}}\right)} = \frac{1}{\sigma}\prod_{n=1}^{63}\left[\mathrm{e}^{-\left({(\mathrm{x}-\mu)}/{\sigma}\right)}\right] e^{-\left(\mathrm{e}^{-{(\mathrm{x}-\mu)}/{\sigma}}\right)}$$

The likelihood is determined using rainfall data (*x*) and the proposed values of μ and σ; i.e. for each iteration a different value of $f(x|\theta)$ will be used.

Prior – prior probability density of the parameters θ, prior to the analysis of sample x, which in this analysis is given by a normal distribution with mean = 0 and variance = 100.

$$p(\theta) = \frac{1}{\sqrt{(2\pi \times Variance)}} \left[e^{-\left((\theta - mean)^2 / (2 \times Variance)\right)} \right]$$

Meaning

$$p(\mu) = \frac{1}{\sqrt{(2\pi \times Variance)}} \left[e^{-\left((\mu - mean)^2 / (2 \times Variance)\right)} \right]$$

$$p(\sigma) = \frac{1}{\sqrt{(2\pi \times Variance)}} \left[e^{-\left((\sigma - mean)^2 / (2 \times Variance)\right)} \right]$$

Therefore the posterior distribution is given by:

$$p(\mu, \sigma | x) = \frac{f(x|\mu, \sigma)p(\mu)p(\sigma)}{\int f(x|\mu, \sigma)p(\mu)p(\sigma)d\mu d\sigma}$$

The denominator $\int f(x|\mu, \sigma)p(\mu)p(\sigma)d\mu d\sigma$ is the normalisation constant to obtain a unit area under the probability density function of $p(\theta|x)$. Without the normalisation constant, the posterior is unscaled; it has all the shape information, but it is not the exact posterior density.

- **Metropolis-Hastings Algorithm:** *(example using Moyeni data)***Initialization**:

Arbitrary values of θ_{t-1} are chosen for *t-1*: $\mu_0 = 49.7$ *and* $\sigma_0 = 42.3$

Prior probability of μ_0 *and* σ_0:

$$p(\mu_o) = \frac{1}{\sqrt{(2\pi \times 100)}} \left[e^{-\left((49.7-0)^2 / (2\times100) \right)} \right] = 1.7E-07$$

$$p(\sigma_o) = \frac{1}{\sqrt{(2\pi \times 100)}} \left[e^{-\left((42.3-0)^2 / (2\times100) \right)} \right] = 5.2E-06$$

Likelihood of x given μ_0 *and* σ_0:

n	**Year**	**Moyeni Rainfall Maxima (mm) (x)**	$f(x_n \mid \theta_{t-1})$
1	1950	132.8	0.0038
2	1951	141.5	0.0030
3	1952	130.3	0.0041
…	…	…	…
63	2012	165.0	0.0017
$f(x \mid \theta)$			$\mathbf{2.4\,E-172}$

Posterior is given by

$$p(\mu_0, \sigma_0 | x) = f(x|\mu_0, \sigma_0)p(\mu_0)p(\sigma_0) = (2.4E-172) \times (1.7E-07) \times (5.19\,E-06) = \mathbf{2.2\,E-184}$$

- Using the chosen proposal distribution (Gaussian distribution with mean = θ_{t-1}, variance = 10 and probability = 0.1) $q(\theta' | \theta_{t-1})$ to simulate a proposed value θ':

$$q(\theta' | \theta_{t-1}) = \frac{1}{\sqrt{2\pi \times Variance}} \left[e^{-\left((\theta' - mean)^2 / (2 \times Variance) \right)} \right]$$

$$q(\mu' | 49.7) = \frac{1}{\sqrt{2\pi \times 10}} \left[e^{-\left((\mu' - 49.7)^2 / (2 \times 10) \right)} \right] = 0.1$$

$$\mu' = 51.86$$

$$q(\sigma' | 42.3) = \frac{1}{\sqrt{2\pi \times 10}} \left[e^{-\left((\sigma' - 42.3)^2 / (2 \times 10) \right)} \right] = 0.1$$

$$\sigma' = 44.46$$

Prior probability of μ' $and\,\sigma'$:

$$p(\mu') = \frac{1}{\sqrt{(2\pi \times 100)}}\left[e^{-\left((51.56-0)^2/(2\times100)\right)}\right] = 5.8E-08$$

$$p(\sigma') = \frac{1}{\sqrt{(2\pi \times 100)}}\left[e^{-\left((44.46-0)^2/(2\times100)\right)}\right] = 2.0E-06$$

Likelihood of x given μ' $and\,\sigma'$:

n	Year	Moyeni Rainfall Maxima (mm) (x)	$f(x_n\|\theta_{t-1})$
1	1950	132.8	0.00428
2	1951	141.5	0.00342
3	1952	130.3	0.00457
...	...	...	
63	2012	165.0	0.00190
$f(x\|\theta)$			2.18 x 10^{-168}

Posterior is given by

$$p(\mu', \sigma'|x) = f(x|\mu', \sigma')p(\mu')p(\sigma') = (2.2EE - 168) \times (5.8E - 08) \times (2.0E - 06) = \mathbf{2.18E - 168}$$

t	$\boldsymbol{\mu_{t-1}}$	$\boldsymbol{\sigma_{t-1}}$	$\boldsymbol{p(\mu_{t-1})}$	$\boldsymbol{p(\sigma_{t-1})}$	$\boldsymbol{f(x|\mu_{t-1}, \sigma_{t-1})}$	$\boldsymbol{p(\mu_{t-1}, \sigma_{t-1}|x)}$	$\boldsymbol{\mu'}$	$\boldsymbol{\sigma'}$	$\boldsymbol{p(\mu')}$	$\boldsymbol{p(\sigma')}$	$\boldsymbol{f(x|\mu', \sigma')}$	$\boldsymbol{p(\mu', \sigma'|x)}$
1	49.7	42.3	1.7E-07	5.2E-06	2.4E-172	2.2E-184	51.9	44.6	5.8E-08	2.0E-06	2.2E-168	2.6E-181

- Acceptance: an acceptance ratio is established which is used to decide whether or not to accept the proposed candidate:

$$r = \frac{p(\theta'|x)/q(\theta'|\theta_{t-1})}{p(\theta_{t-1}|x)/q(\theta_{t-1}|\theta')}$$

Since $q(\theta_{t-1}|\theta') = q(\theta'|\theta_{t-1}) = 1$, the above equation is reduced to

$$r = \frac{p(\theta'|x)}{p(\theta_{t-1}|x)} = \frac{p(\mu', \sigma'|x)}{p(\mu_0, \sigma_0|x)} = 1.18E03$$

If $r \geq 1$, then the candidate is accepted ($\theta_t = \theta'$). Otherwise the candidate is rejected ($\theta^t = \theta^{t-1}$). Since *r>1*, then the proposal is accepted:

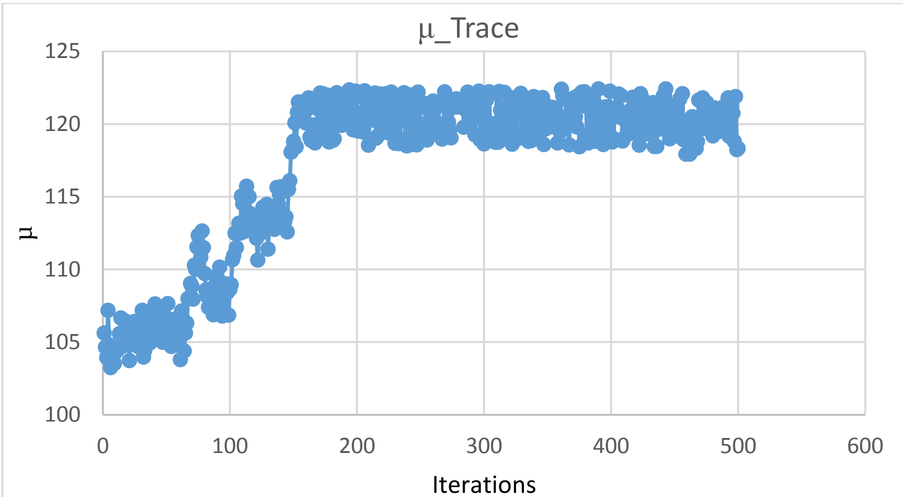
$$\boldsymbol{\mu_t = 51.9 \; and \; \sigma_t = 44.6}$$

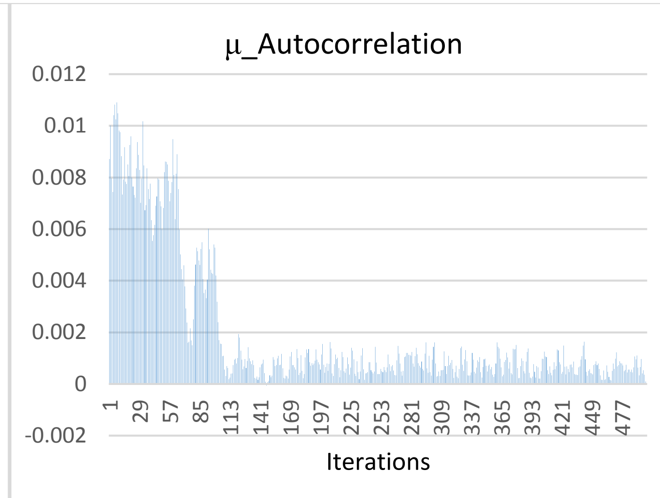
- Steps 2 and 3 are repeated until the sequence of numbers converges to $p(\theta|x)$:

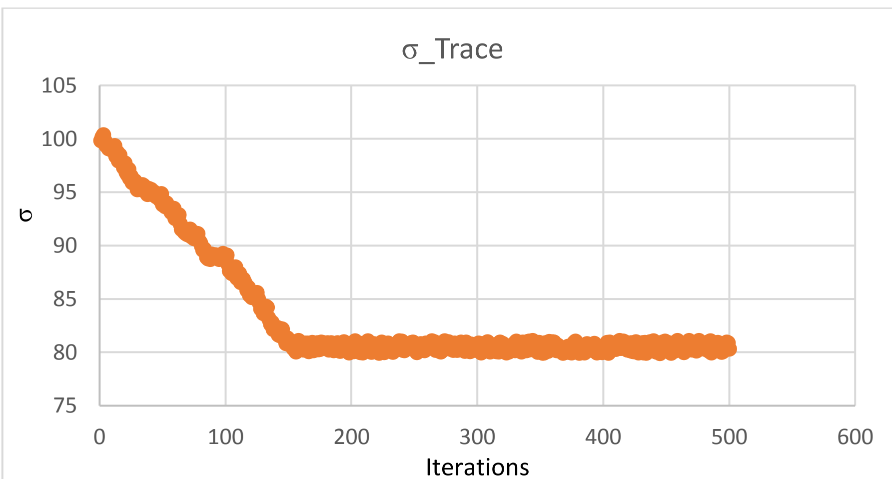
t	$\boldsymbol{\mu_{t-1}}$	$\boldsymbol{\sigma_{t-1}}$	$\boldsymbol{p(\mu_{t-1})}$	$\boldsymbol{p(\sigma_{t-1})}$	$\boldsymbol{f(x\|\mu_{t-1},\sigma_{t-1})}$	$\boldsymbol{p(\mu_{t-1},\sigma_{t-1}\|x)}$	$\boldsymbol{\mu'}$	$\boldsymbol{\sigma'}$	$\boldsymbol{p(\mu')}$	$\boldsymbol{p(\sigma')}$	$\boldsymbol{f(x\|\mu',\sigma')}$	$\boldsymbol{p(\mu',\sigma'\|x)}$	r	
1	49.7	42.3	1.7E-07	5.2E-06	2.4E-172	2.2E-184	51.9	44.6	5.8E-08	2.0E-06	2.2E-168	2.6E-181	1180	Accept
2	51.9	44.6	5.8E-08	2.0E-06	2.2E-168	2.6E-181	54.0	46.6	1.9E-08	7.6E-07	8.0E-165	1.1E-178	439	Accept
3	54.0	46.6	1.9E-08	7.6E-07	8.0E-165	1.1E-178	56.2	48.8	5.6E-09	2.73E-07	1.5E-161	2.3E-176	200	Accept
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14	77.7	70.3	3.1E-15	7.3E-13	9.4E-140	2.1E-166	79.9	72.5	5.6E-16	1.6E-13	3.0E-138	2.6E-166	1.3	Accept
15	79.9	72.5	5.6E-16	1.6E-13	3.0E-138	2.6E-166	82.0	74.6	9.7E-17	3.2E-14	7.9E-137	2.5E-166	0.94	Reject
16	79.9	72.5	5.6E-16	1.6E-13	3.0E-138	2.6E-166	90.0	20.0	1.0E-19	5.4E-03	5.1E-104	2.8E-125	1.08E41	Accept
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

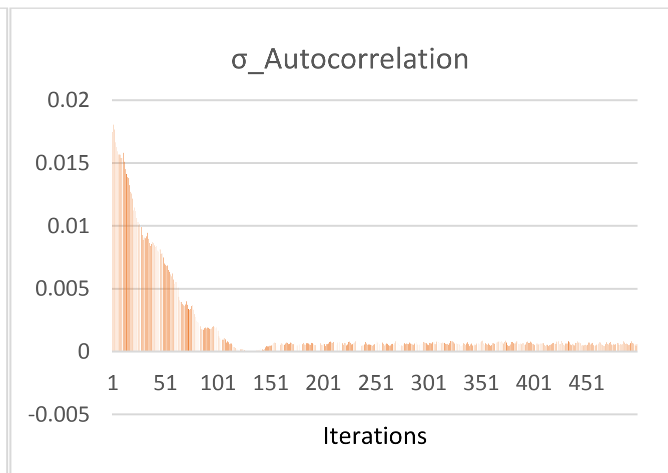
II. APPENDIX 2: CONVERGENCE DIAGNOSITICS

Moyeni Single-Site

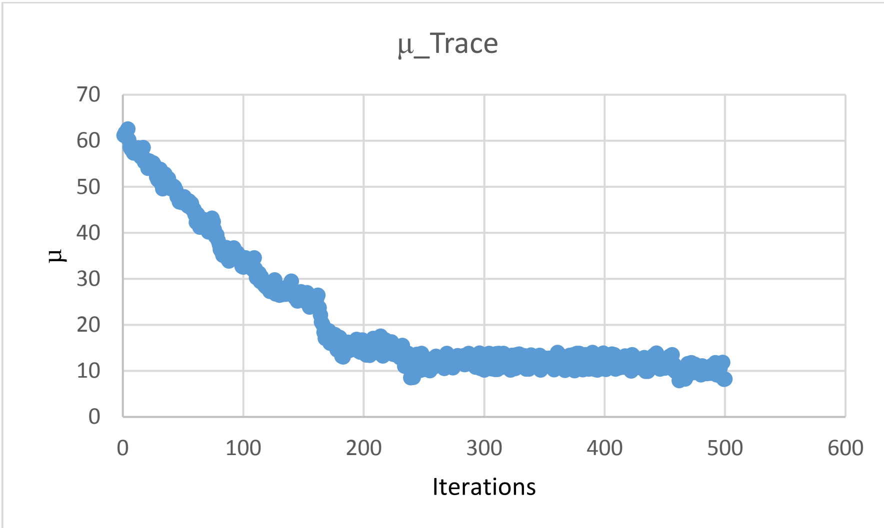


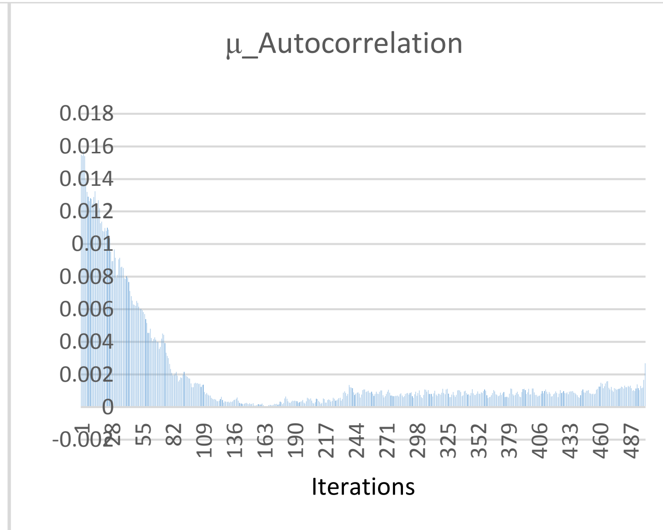


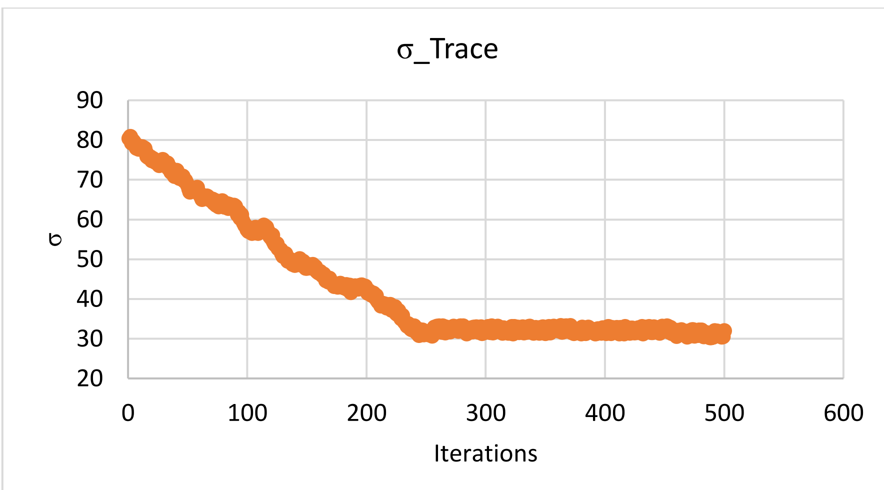


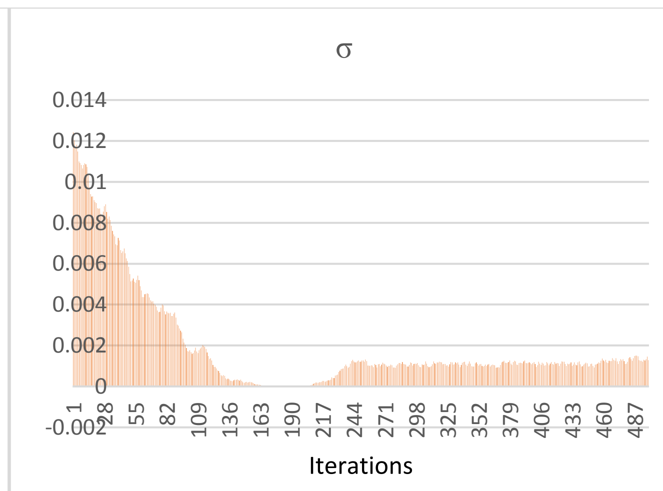


Seaka Single-Site

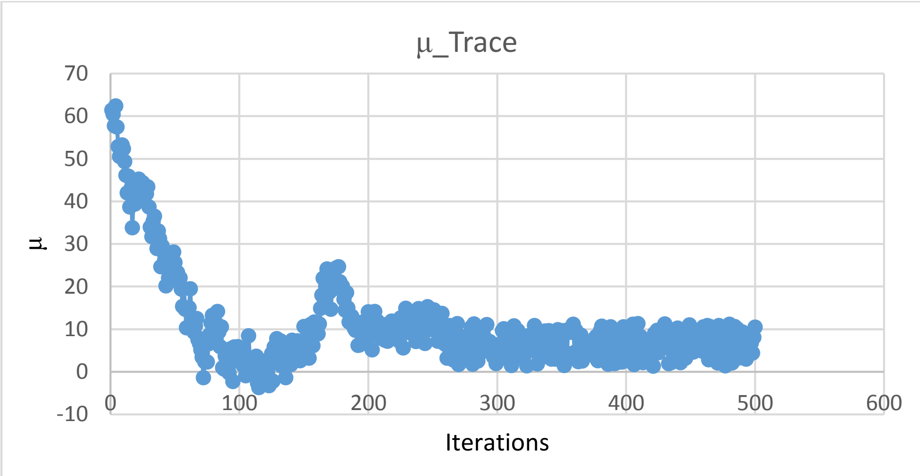


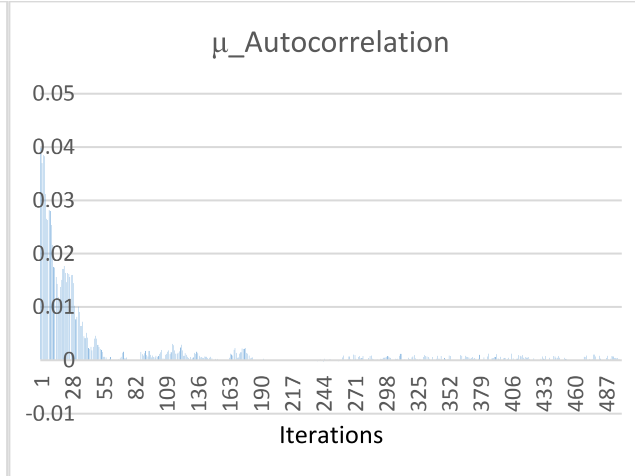


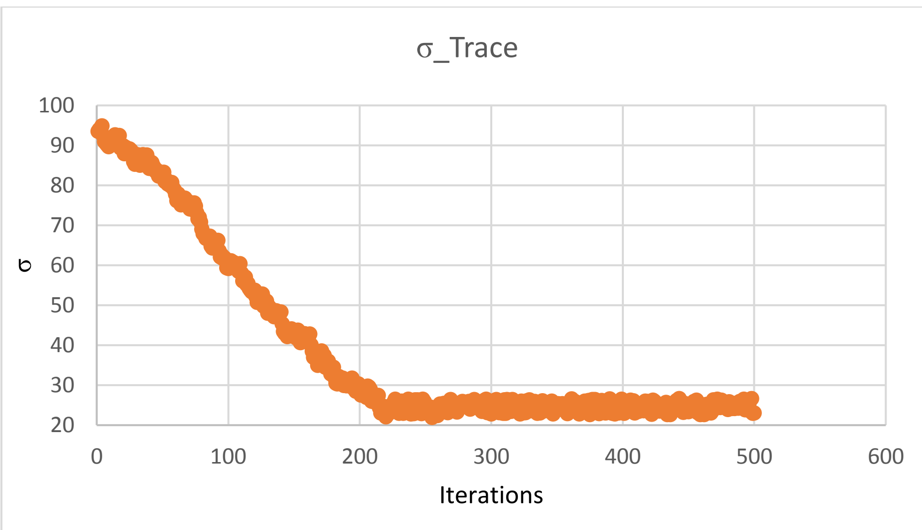


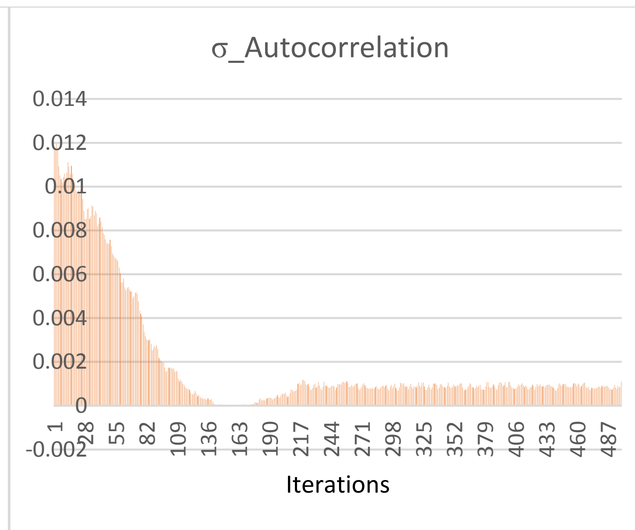


Dilli-Dilli Single Site

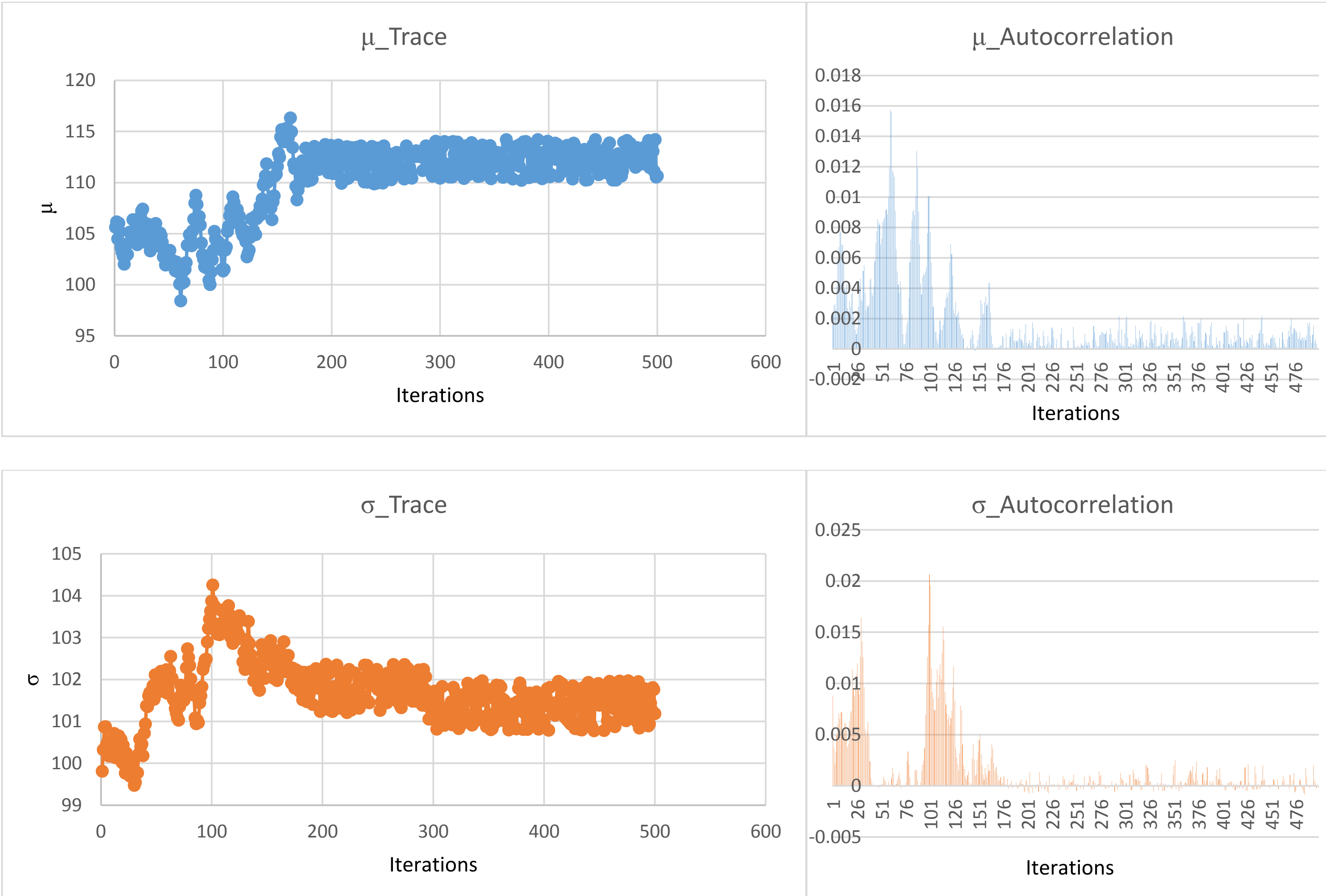




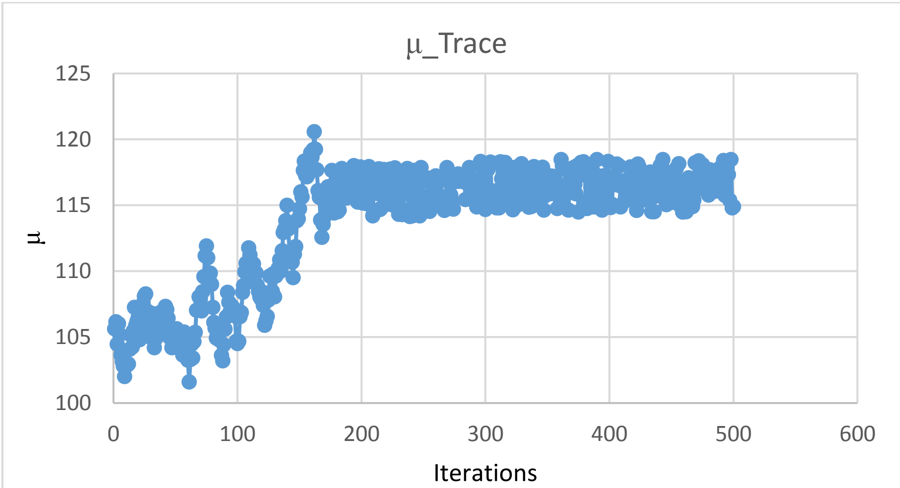


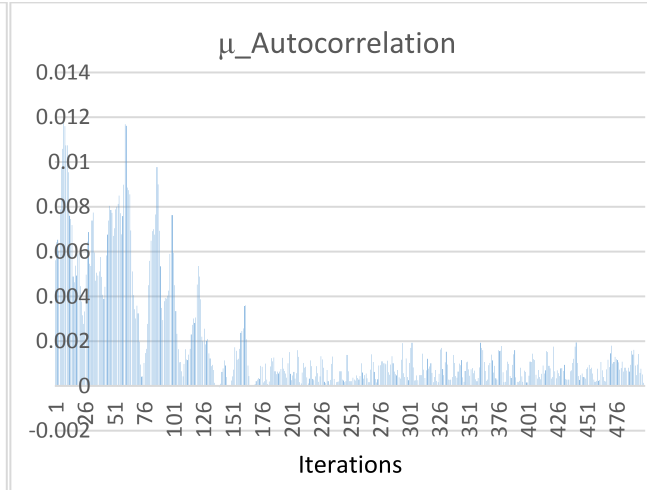


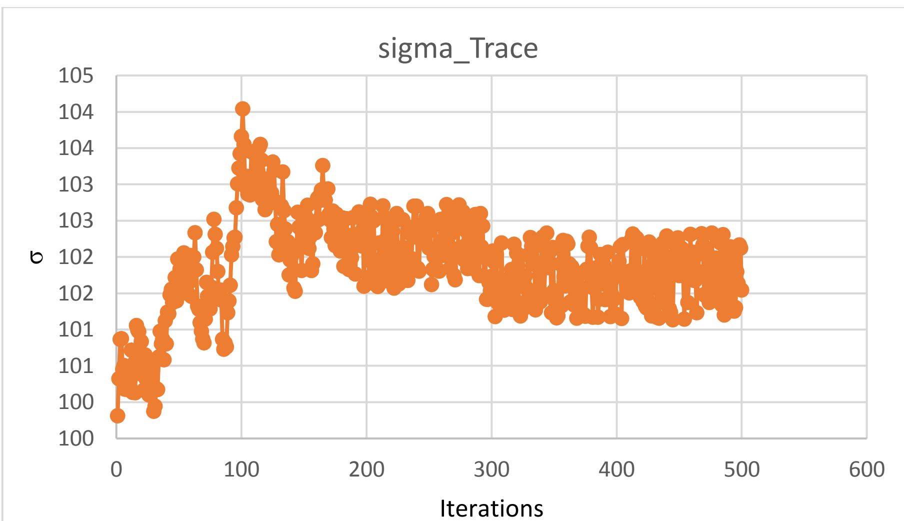
Moyeni Two-Site (with Seaka)

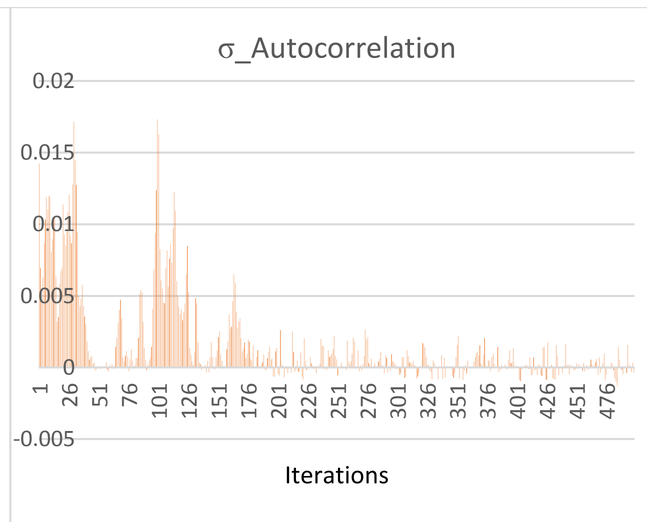


Moyeni Two-Site (with Dilli-Dilli)

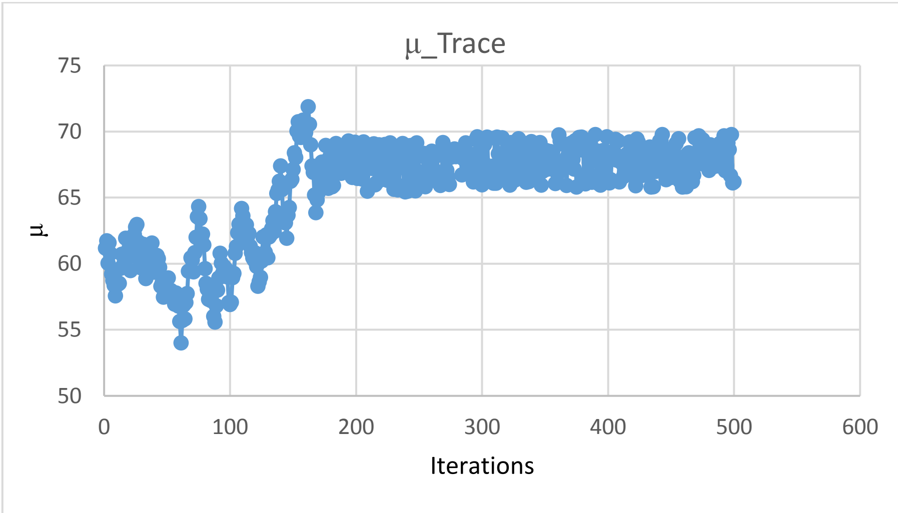


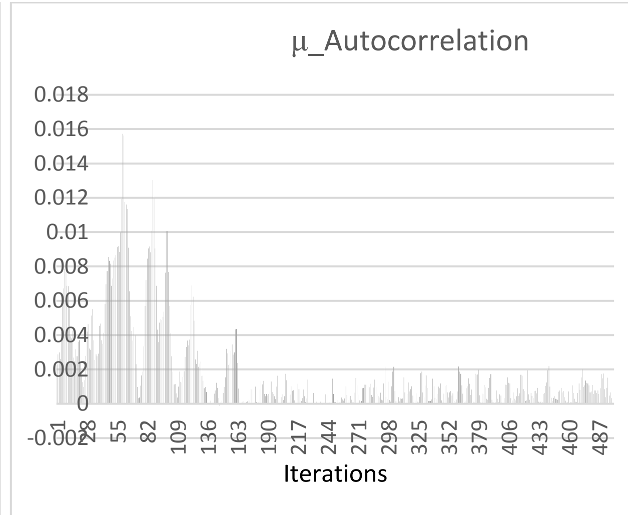


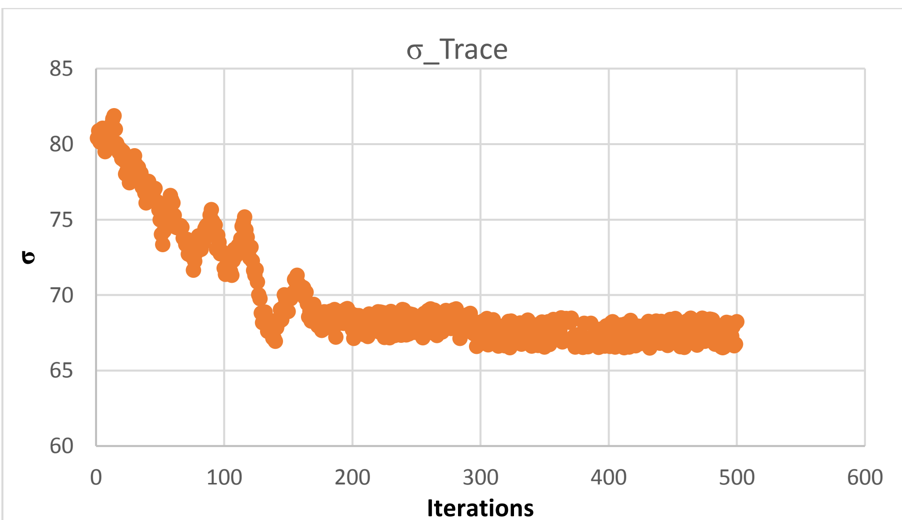


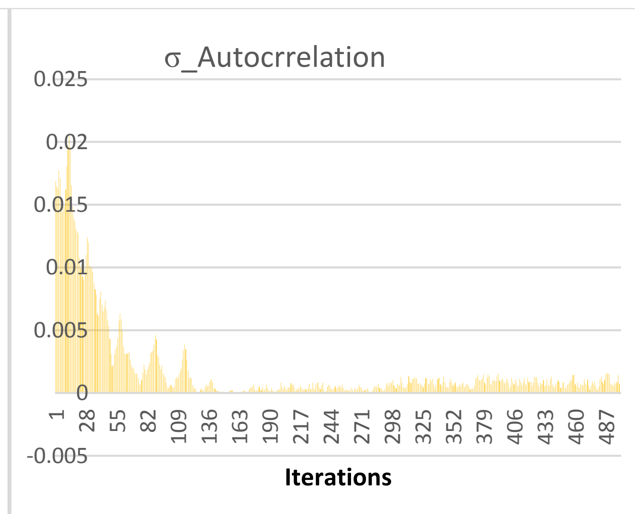


Seaka Two-Site (with Moyeni)









Dilli-Dilli two-Site (with Moyeni)