Al-Azhar University Civil Engineering Research Magazine (CERM)
Vol. (40) No. (4) October, 2018

# Optimal Selection of Rainfall Gauges for Safe Extreme Events Estimation Using a Geostatistical Approach

**Abdelkhalek. A. M [1], A. G. Awadallah [1,2], Nabil Awadallah[1]**
1 Civil Engineering Dept., Faculty of Engineering, Fayoum University, Fayoum, Egypt
2 Corresponding Author

**ملخص البحث**

يهدف هذا البحث إلى عرض طريقة لتحديد العدد الأمثل لمحطات رصد الأمطار ومواقع هذه المحطات مع تحقيق خطأ نسبي مقبول في تقدير عمق الأمطار عند فترات تكرارية عالية. وتعتمد المنهجية على المقارنة بين أعماق الأمطار عند فترات تكرارية عالية مقدرة باستخدام العدد الكلي لمحطات رصد الأمطار في مقابل تلك المقدرة باستخدام عدد أقل من المحطات لحين الوصول إلى العدد الأمثل المناظر لخطأ نسبي مقبول. وتستخدم طريقة أخذ العينات ((Latin Hyper Cube Sampling (LHS) لتكوين عدد كبير من العينات من المحطات ثم يتم إجراء تحليل تكراري لكل عينة لتحديد المجموعة المثلى منها ذات أقل خطأ. ويتم عمل التحليل التكراري للأمطار باستخدام طريقتين: الأولى هي طريقة التحليل التكراري الإقليمي والثانية طريقة التحليل التكراري لبيانات الأمطار لكل موقع على حدة. وتم تطبيق المنهجية على محطات الأمطار الموجودة بحوض تصريف وولنت جالش (WGEW) والذي يقع في ولاية أريزونا بالولايات المتحدة الأمريكية. وفي النهاية أكدت النتائج التى تم التوصل إليها أنه في حدود خطأ معين مقبول عند فترات تكرارية عالية، فإن عدد المحطات المطلوب لتحقيق خطأ مناظر أو أقل باستخدام طريقة التحليل التكراري الإقليمي أقل من ذلك المستنتج باستخدام طريقة التحليل التكراري لكل موقع على حدة. ويوضح البحث كذلك الأماكن المثلى لمحطات رصد الأمطار.

## Abstract

This paper objective is to present a method for optimizing rain gauge network aiming to determine the optimal number of stations and their locations, to achieve an acceptable error in extreme rainfall estimation. The optimization is based on the comparison between the maximum daily rainfall depths at high return periods deduced using the entire rain gauges networks and that deduced using an "optimal" number of rain gauges. A Latin Hyper Cube Sampling (LHS) method is used to generate samples of stations. Each generated sample is analyzed to obtain the optimum set of gauges locations. The optimal number and locations of rain gauges is obtained in two cases; the first one using a regional frequency analysis technique and the second using an at-site frequency analysis technique. The methodology is applied on the existing rain gauges network of the Walnut Gulch Experimental Watershed (WGEW), Arizona, USA. The results showed that, a lower number of rain gauges is required based on the regional frequency analysis technique compared to the at-site frequency analysis technique to achieve the same relative error at the high return periods. The study also suggests optimum locations for the rain gauges.

**Keywords:** Optimization; Regional frequency analysis; LHS; Arizona;

## 1. Introduction

Accurate estimations of hydrologic variables are essential in the field of science and engineering because of their environmental and social impacts on water management. Water management is not only limited to flood control but also water supply, water quality improvements and environmental restoration/preservation. Consequently, optimal design of rain gauge networks is essential in rainfall amount estimation and hence in all related water activities (e.g. flood simulation, flood protection design, water power generation, expansion in agriculture, etc.).

In many cities around the world, rainfall networks suffer either from a reduced number or redundant stations, and also other stations that are not located in their appropriate locations. All of these issues cause uneconomic networks and/or inaccurate rainfall estimates. Rainfall networks optimization procedures vary in their methodologies, their considered variables, their optimization indicator as well as the considered rainfall characteristics. The Kriging approach and the Standard Error Method, which were developed by Matheron (1971) and Caffey (1965) are widely used. Many variables are commonly considered as optimization indicators (e.g. rainfall depth measurements, elevations on which rain gauges are placed, rain gauges spacing, cost of rainfall network, etc.). Adhikary et al. (2015), Khairul et al. (2016) and Al-abadi and Al-aboodi (2014) considered rainfall data as the only variable and the indicator of optimality was the decrease of the Kriging error after removing the redundant stations. Pardo-Igúzquiza (1998) obtained the relative errors of rainfall depth estimation and considered also the cost of rain gauges in the network. A combination between the rainfall data and secondary sources of information (e.g. elevation, humidity, and temperature etc.) was also considered in networks optimization (Putthividhya and Tanaka, 2012). However, optimization of an existing network can be made by relocating the existing rain gauges without increasing their number (Haggag et al. 2016). Minimizing the coefficient of variation of the resulting error corresponding to the number of rain gauges and their location could also be taken as an optimizing criterion (Bras and Rodriguez-Iturbe, 1976; Hughes and Lettenmaier, 1981; Bastin et al., 1984; Bogardi and Bardossy, 1985; Rouhani,1985).

In this study, we considered the error in rainfall frequency analysis estimates at high return periods as an optimization criterion. The impact of reducing the number of rain gauges on the estimates of rainfall depths at high return periods is studied for the first time in this research. The regional rainfall frequency analysis (RRFA) method and the at-site rainfall frequency analysis (STRFA) method are also used along with three error criteria for the rainfall estimates at high return periods.

The paper comes in five sections. After the current paper introduction, the research methodology is detailed in section 2, followed by the description of the study area and the available rainfall data (section 3). Section 4 presents the results and discussion of the application of the methodology; while the final section (5) summarizes the research conclusions and recommendations.

## 2. Methodology

The framework of the developed methodology and its sequential steps are summarized in Figure (2) and detailed hereafter:

1. *Remove of outliers from the existing data*

The main purpose of this step is to exclude the inconsistent data from the available dataset to decrease error and increase the power of statistical tests. Thus, after obtaining the maximum daily rainfall over the record duration, the U.S. Water Resources Council (1981) proposed a test was applied to exclude the measurements that are extremely large or small relative to the rest of the data and, therefore, could be removed from the dataset.

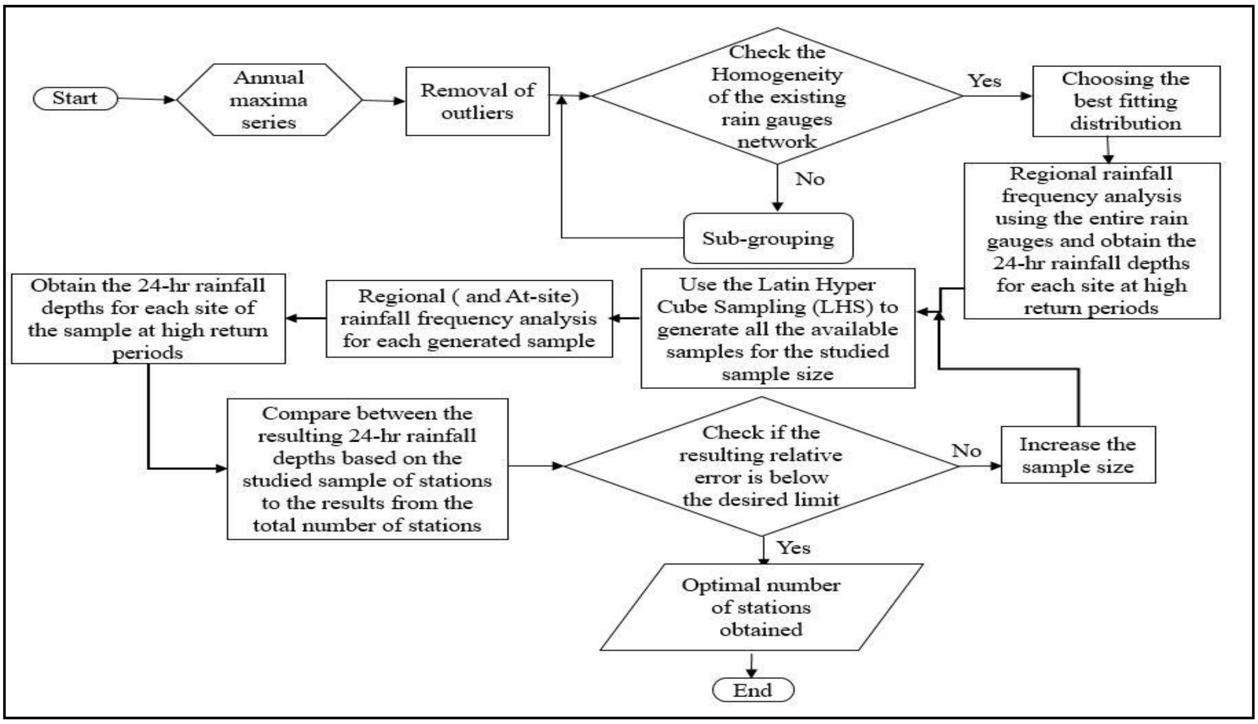



**Fig 1**: Flow chart describing the framework of the methodology

2. *Undertake a homogeneity check to investigate whether the study zone is homogeneous or not*

Homogeneity check should be applied to all the rain gauges records to ensure that all the stations belong to the same population. If the region isn't homogeneous, it should be divided to sub-groups and then choose the suitable distribution to fit the data of each one of these groups. Among the methods to check homogeneity, the Wiltshire (1986a and b) method is used. This method expresses the homogeneity of a region in terms of a rainfall statistic such as the coefficient of variation (Cv).

3. *Choose the best fitting distribution for each group*

To fit the rainfall data of each homogenous group, several distributions are available (e.g. Normal, Lognormal, Gamma, and Gumbel …. etc.). Distributions are prioritized for the studied data and the distribution with the highest priority is the one used to analyze the data. To select the appropriate distributions to fit the data, several techniques are used: the Akaike Information Criterion (Akaike, 1973, 1974) (AIC), the Bayesian Information Criterion (Schwarz, 1978) (BIC). The AIC and BIC are calculated as function of the sample size, the number of parameters, and the likelihood function. The distribution which has the minimum AIC and BIC is considered the best one to fit the data (Vogel and Fennessey, 1993).

4. *Obtain the 24-hr rainfall values at high return periods using rainfall frequency analysis undertaken for all 90 available stations.*

Rainfall frequency analysis uses the historical records from the available rain gauge stations to estimate the design rainfall depth at any desired return period using the appropriate distribution. The estimation of the design rainfall is based on the probability of occurring and the parameters of the used distribution. In this study, the method of moments is selected for its simplicity and since it is the most commonly used method in frequency analysis (Ayyub and McCuen, 2011). Regional frequency analysis uses the

data from different locations to compensate for short records at a single site (National Research Council, 1988; Stedinger et al., 1993). It is based on the concept of regional homogeneity which assumes that annual maximum rainfall populations at several sites in a region are similar from a statistical perspective. The regional parameters for the used distribution are obtained using the regional weighted average moments (skewness coefficient, standard deviation and kurtosis coefficient). The procedure of this method is described as follows:
First, data of each site are divided by their mean:

$$q_{ij} = Q_{ij} / \mu_j \tag{1}$$

where $q_{ij}$ is the standardized observation (i) at site (j) and $Q_{ij}$ is the original record (i) at site (j) and μj is the mean of all data at site (j). Then, the standard deviation $\sigma_{(j)}$, skewness $\gamma_{3(j)}$, and kurtosis $\gamma_{4(j)}$ coefficients are calculated for each site using the standardized observations. The regional weighted moments: weighted standard deviation (wstdvr), weighted skewness (wskewr), and weighted kurtosis (wkurtr) are calculated from equations 2, 3 and 4, respectively, as follows:

$$\text{wstdvr} = \frac{\sum_{j=1}^{n} \sigma_{(j)} * Nj}{\sum_{j=1}^{n} Nj} \tag{2}$$

$$\text{wskewr} = \frac{\sum_{j=1}^{n} \gamma_{3(j)} * Nj}{\sum_{j=1}^{n} Nj} \tag{3}$$

$$\text{wkurtr} = \frac{\sum_{j=1}^{n} \gamma_{4(j)} * Nj}{\sum_{j=1}^{n} Nj} \tag{4}$$

where Nj is the number of observations at site j and n is the total number of sites in the region. Using the weighted moments of the selected distribution, the regional frequency depths (X(T)) at T return periods are calculated. Then the regional frequency analysis rainfall depths at 24-hr duration are obtained at each site:

$$\text{TRFA}_j(T) = X(T) * \mu_j \tag{5}$$

where $\text{TRFA}_j$ (T) is the regional frequency analysis rainfall depths at site j based on the entire (Total) number of stations using the selected distribution.

5. *Using the Latine Hypercube Sampling (LHS) method, the available number of samples were generated.*

A sampling method is used to generate randomly drawn samples. Each generated sample represents a possible combination which could occur. To generate samples in this study, the orthogonal sampling technique, which is a special case of the Latin Hypercube Sampling (LHS) (McKay et al. 1979), is used. First, the sample space is divided into subspaces. Latin Hypercube sampling is applied to each one of the subspaces where one sample is taken in each subspace. Finally, all the subspaces must have the same density. The orthogonal samples technique creates orthogonal samples of 1: k*m in N dimensions. where N is the number of dimensions (N≥ 1. N integer), k is the number of large subdivisions (subspaces) per dimension (k≥ 1. k integer), and m is the number of bins per dimension in one subspace. For N=2, k=2, and m=4, the available number of samples is 1: k*m (8). If k is reduced to 1 then the method reduces to the Latin Hypercube sampling which is used in this study. The orthogonal LHS is implemented using a Matlab code for sample generation.

6. *Step No. 4 is repeated using the available stations in each generated sample to obtain the sample regional frequency analysis depths using regional analysis at different return periods. These rainfall depths are hereafter named SRFA depths, where S stands for sample.*

7. *The obtained TRFA depths based on the total number of stations are compared to the SRFA depth obtained based on a sample of stations.*

The comparison was made using three criteria as follows:

a. The relative error between the obtained TRFA depths at a certain return period and the SRFA calculated at each site. This criterion is to be applied to make sure that the error at each location is less than a specified /desired error. It could be useful when interest is for a specified location. It is calculated as follows:

$$E(j) = \left[\frac{abs(diff(j))}{TRFA_j}\right] * 100 \quad (6)$$

where diff(j) at a certain return period is calculated as follows:

$$diff(j) = TRFA_j - SRFA_j \quad \text{(at a certain return period)} \quad (7)$$

Then, the mean of the relative errors ($E_{av}$) at a certain return period for the $m^{th}$ sample is calculated as follows:

$$E_{av}(m) = \frac{\sum_{j=1}^{n} E(j)}{n} \quad (8)$$

where n is the number of stations per sample (sample size). Finally, the mean of the mean relative errors for all generated samples with n number of stations per-sample is calculated as:

$$Mean\ (E_{av})(n) = \frac{\sum_{m=1}^{N} E_{av}(m)}{N} \quad (9)$$

where: N is the total number of samples generated using n stations.

b. The Relative error between the mean of TRFA depths and the mean of SRFA depths based on the generated sample (m). This criterion could be applied to limit the error in the mean rainfall depth estimation over a whole catchment of interest to the specified / desired error. As such, it can be used to specify the required number of stations which gives satisfying results when the application uses an average rainfall at a certain return period. It is calculated as follows:

$$E_{mean}(m) = \left[\frac{abs(mean\_diff\ (m))}{TRFA_{mean}}\right] * 100 \quad (10)$$

where mean diff (m) is calculated as follows:

$$mean\_diff\ (m) = TRFA_{mean} - SRFA_{mean} \quad (11)$$

where $TRFA_{mean}$ and $SRFA_{mean}$ is the mean rainfall depths based on the total number of rain gauges and the generated sample (m), respectively. The mean of relative mean errors is based on the mean rainfall depth ($E_{av}(R_{mean})$ for all the generated samples with n number of stations per sample is calculated as follows:

$$E_{av}(R_{mean})\ (n) = \frac{\sum_{m=1}^{N} E_{mean}(m)}{N} \quad (12)$$

where N is the total number of samples generated using n stations.

c. The Relative error between the maximum of TRFA depths and the maximum of SRFA depths based on the generated sample (m). This criterion could be applied to limit the error in the maximum rainfall depth estimation in the catchment regardless of where it occurred, which means that it is the most conservative criterion. It is calculated as follows:

$$E_{max}(m) = \left[\frac{abs(max\_diff\ (m))}{TRFA_{max}}\right] * 100 \quad (13)$$

where max_diff (m) is calculated as follows:

$$\text{max\_diff}(m) = TRFA_{max} - SRFA_{max} \qquad (14)$$

where $TRFA_{max}$ and $SRFA_{max}$ is the maximum rainfall depths based on the total number of rain gauges and the generated sample (m), respectively. The average of relative errors is based on the maximum rainfall depth ($E_{av}(R_{max})$ for all generated samples with n number of stations per sample is calculated as follows:

$$E_{av}(R_{max})(n) = \frac{\sum_{m=1}^{N} E_{max}(m)}{N} \qquad (15)$$

where N is the total number of samples

8. *Step No. 7 is repeated using the at-site rainfall frequency analysis depth at each site to obtain the at-site relative errors for the same above mentioned error criteria.*

Finally, the specific methods used in the study are listed in Table (1):

Table (1): Used methods in each step of this research

| Subject | The used method | Reason |
|---|---|---|
| Homogeneity check | Wiltshire | Based on Coefficient of variation which is better suited for short records |
| Frequency analysis distribution | Gamma distribution | Selected distribution is determined based on AIC and BIC and Moment ratio diagrams. |
| Sampling techniques | Orthogonal sampling | Less number of iterations, short run time, higher number of generated samples |

## 3. Study Area and Data Description

The Walnut Gulch Experimental Watershed (WGEW), Arizona, USA is selected as the study area to apply the methodology. WGEW is located in the southeastern Arizona, between latitudes 110$^{o}$ 0' 0" West and 31$^{o}$ 45' 0" North. Its size is approximately 149 km$^2$ (Heilman et al. 2008).Figure (2) shows the location of the watershed in Arizona State and the 90 existing rain gauges in it. An extensive precipitation database has been developed over the past 62 years starting in August, 1953 and continuing to December 2015. The study area data are operated and managed by the U.S. Department of Agriculture, Agricultural Research Service, Southwest Watershed Research Center (SWRC) in Tucson. WGEW is recognized.

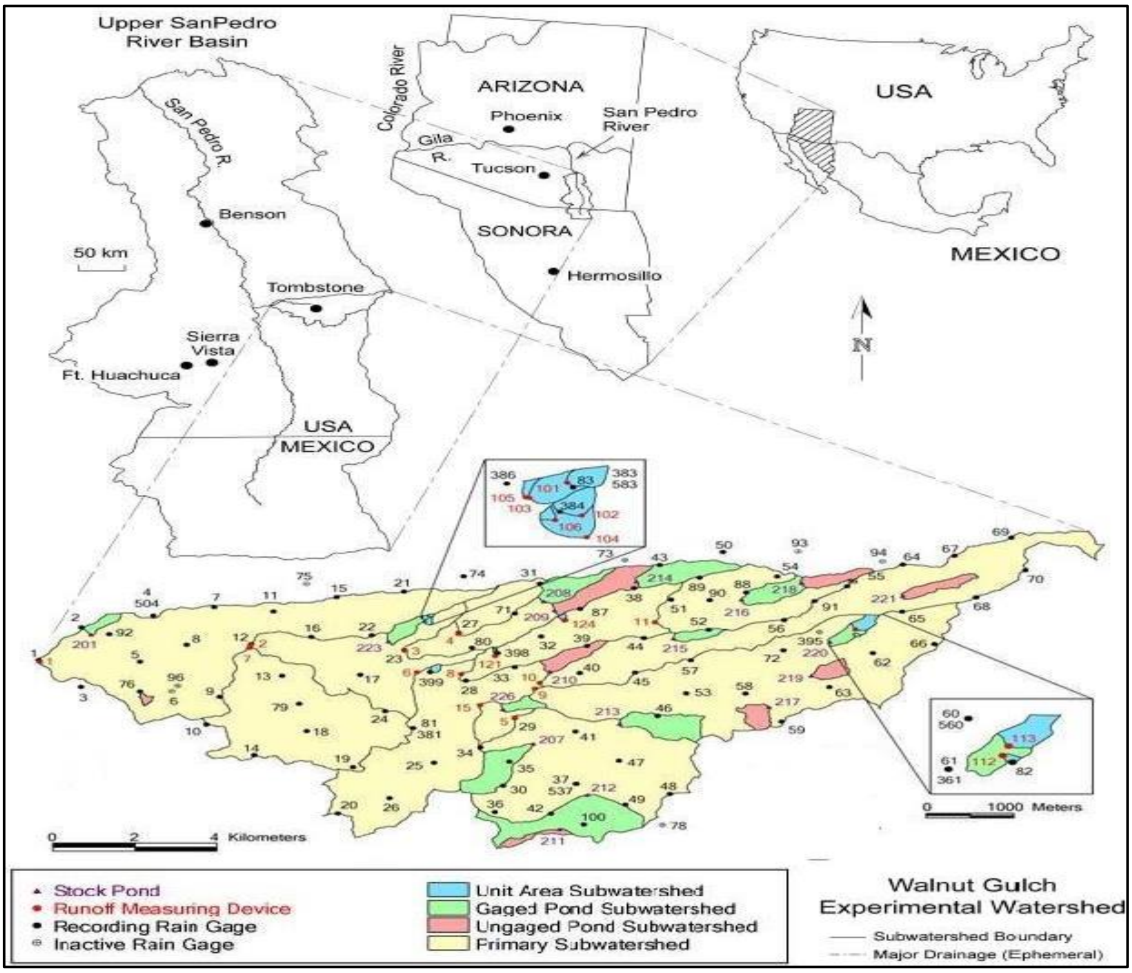



**Fig 2:** Location of the WGEW and locations of the existing rain gauges (after SWRC, 2007).

as the most densely instrumented semiarid experimental watershed in the world (0.6 gauges/km$^2$), and a premier outdoor laboratory for semiarid watershed hydrology studies (Renard et al., 2008). All data are available at the website: (http://www.tucson.ars.ag.gov/dap/).

## 4. Results and Discussion

*4.1 Removal of outliers from the existing data.*

After the maximum daily rainfall depths are obtained for all available stations, the outlier statistical test (U.S. Water Resources Council, 1981) was applied and the results of the test indicated that there are no outliers.

*4.2 Homogeneity check to investigate whether the study zone is homogeneous or not.*

The Wiltshire test (1986a and b) was used to check the homogeneity of the data and the p-value of the statistic was 0.64 much higher than the limit value of 0.05. The region is thus considered homogeneous and only one frequency distribution can be used for rainfall frequency analysis.

*4.3 Choose the best fitting distribution.*

The two goodness of fit indicators (AIC, BIC criteria) and the moment ratio diagrams were applied to all the stations. Ordinary moment ratio diagram and L-moment diagram were used to select the best fitting distribution as shown in Fig (3). As shown in the figure, the P-III distribution is the nearest one to the simulated average point of all stations. As the two parameter Gamma distribution has the same Cs-CK relationship as the Pearson type III distribution, it could also be. Furthermore, to choose between distributions, AIC (Akaike, 1973, 1974) and BIC (Schwarz, 1978) are used.

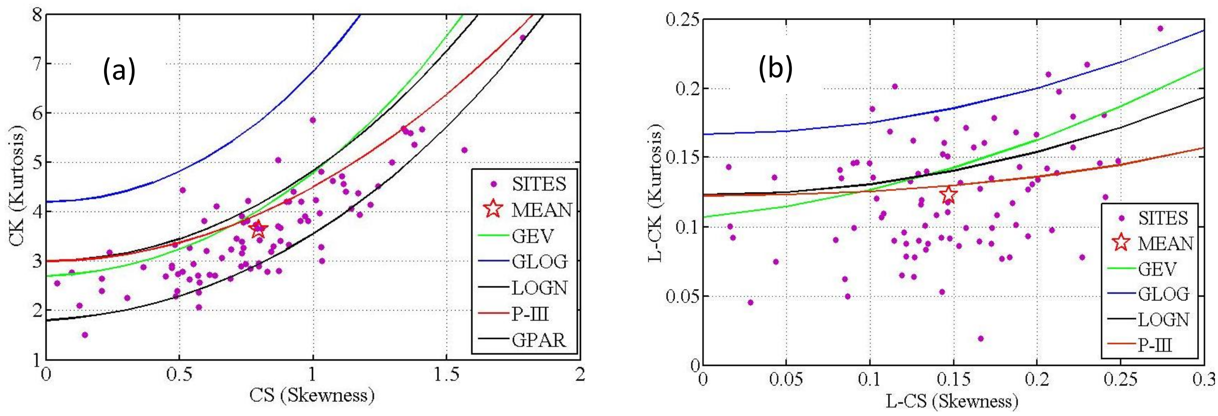



**Fig 3:** Moment ratio diagrams based on (a) ordinary moments and (b) L- moments.

The AIC and BIC results also gave the highest priority to the Gamma distribution where nearly half of the stations are appropriate to be fitted using the Gamma distribution. The Gamma distribution was the one used in this research; its parameters (Beta and Alpha) are obtained as follows:

$$\mathrm{Beta} = 1/(\mathrm{wavgr}/\mathrm{wstdvr}^2) \quad (16)$$

$$\mathrm{Alpha} = 1/(\mathrm{Beta} * \mathrm{wavgr}) \quad (17)$$

where wstdvr, wskewr, and wkurtr are calculated from equations 2, 3 and 4 respectively.

*4.4. Error Results for the Regional Rainfall Frequency Analysis Case (RRFA).*

The daily regional rainfall depths at high return periods were obtained for the entire number of rain gauges using the Gamma distribution. These rainfall depths for each gauging station are the reference of the comparison with the regional and at-site rainfall depths calculated based on a sample of stations. The available number of samples were generated using LHS for sample sizes of 3 to 10 and 15, 20, 25, 30, and 35 stations per sample. The RRFA and STRFA were applied to each sample. The relative errors between the regional rainfall depths calculated based on the entire number of stations and based on a sample of stations were obtained for the three relative error criteria previously presented in the methodology section. The results of the three criteria are shown in Fig (4-a, b and c) for the Mean ($E_{av}$) (Eq (9)), $E_{av}$ ($R_{mean}$) (Eq (12)) and $E_{av}$ ($R_{max}$) (Eq (15)), respectively.

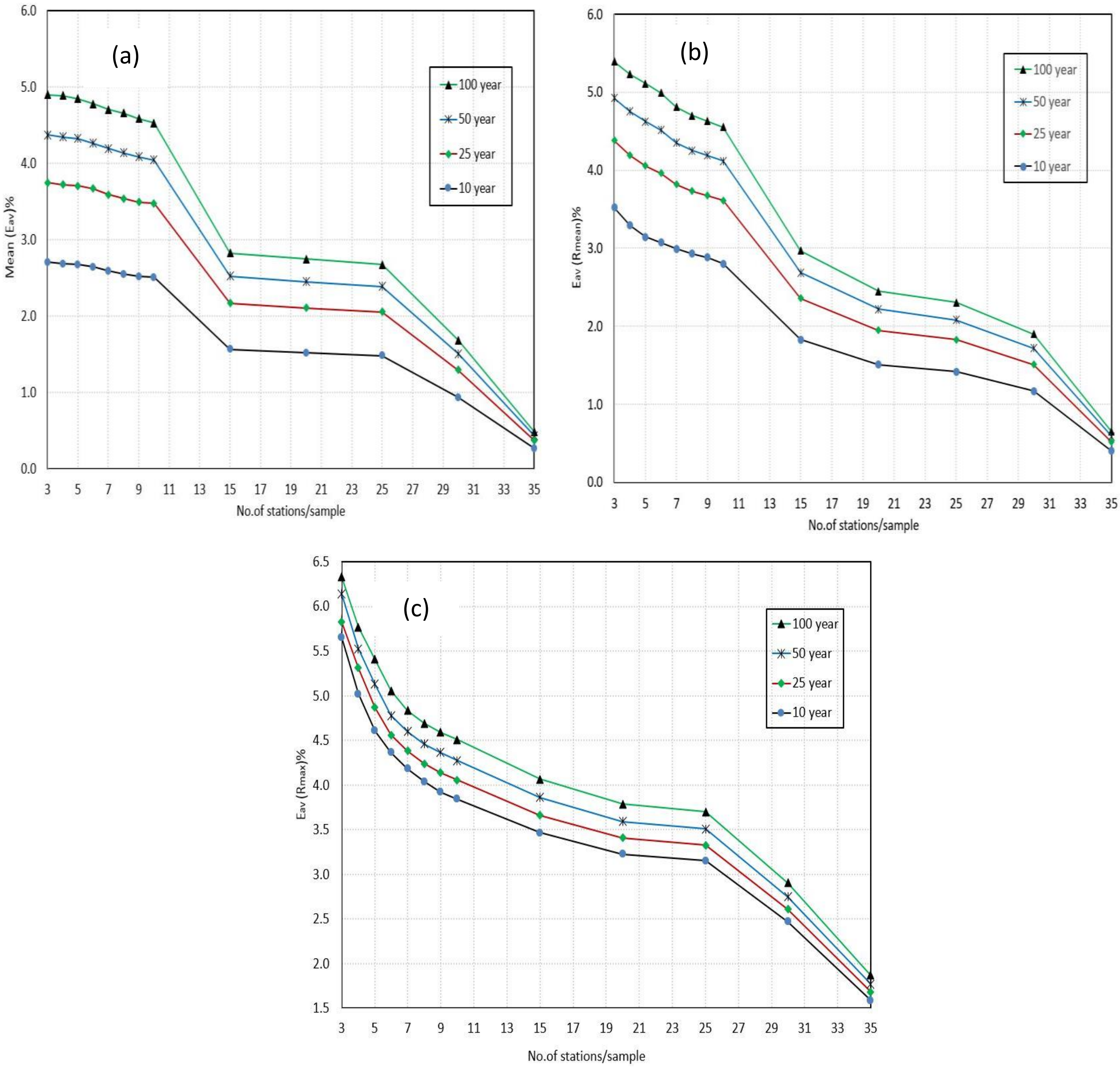



**Fig 4:** Relative error of the RRFA case (a) Mean (Eav) & (b) Eav (Rmean) and (c) Eav (Rmax).

In Fig (4-a), it is shown that, to achieve a Mean $E_{av}$ of 5 % for the 100-year return period, 3 stations are required. Furthermore, it is noted that the slopes of the curves are relatively steep till 15 stations per sample. However, they are nearly flat between 15 to 30 stations per sample, which indicates that the increase of the number of stations per sample is more effective for sample sizes less than 15 stations per sample. The curves for various return periods converge as the number of stations increases, which is clearly noted at 35 stations per sample.

Fig (4-b) shows that, to achieve $E_{av}$ ($R_{mean}$) of 5 %, 6 stations and 3 stations are required, for the 100- and the 50-year return periods, respectively. The slopes of the curves show the same previously described patterns.

Fig (4-c) shows that, to achieve $E_{av}$ ($R_{max}$) of 5 %, 6 stations and 5 stations are required, for the 100- and the 50-year return periods, respectively. The slopes of the curves show the same previously described patterns. The differences in the relative errors between the various high return periods are less important than those of Fig (4-a) and Fig (a-b).

*4.5. Error Results for the At-site Rainfall Frequency Analysis Case (STRFA).*

The same three relative error criteria are studied for the STRFA case. However, in equations Eq 7, 11, and 14, the SRFA depth based on the generated sample of stations ($SRFA_j$) is replaced by the STRFA depth of the site j.

Fig (5-a) shows that all sample sizes in the Mean ($E_{av}$) criterion of the STRFA case produce equal relative errors at the same return period, because the value of the rain "diff (j)" (in Eq (7)) for each site is not affected by the sample size. Furthermore, it is shown from Fig (5-a) that the resulting error at the 100- and 50-year return periods is 5.3% and 4.7%, respectively, for any number of stations.

Fig (5-b) shows the obtained relative error in the $E_{av}$ ($R_{mean}$) criterion of STRFA case. In this criterion it is shown that, to achieve $E_{av}$ ($R_{mean}$) of 5%, 7 and 6 stations are required for the 100- and 50-year return period respectively, compared to 6 and 3 stations in the same error criterion based on the RRFA case.

Fig (5-c) shows that, to achieve an Eav (Rmax) of 5%, 15 and 13 stations are required at the 100- and 50-year return period, respectively, while are six and five stations were needed for the same error criterion based on the RRFA case.

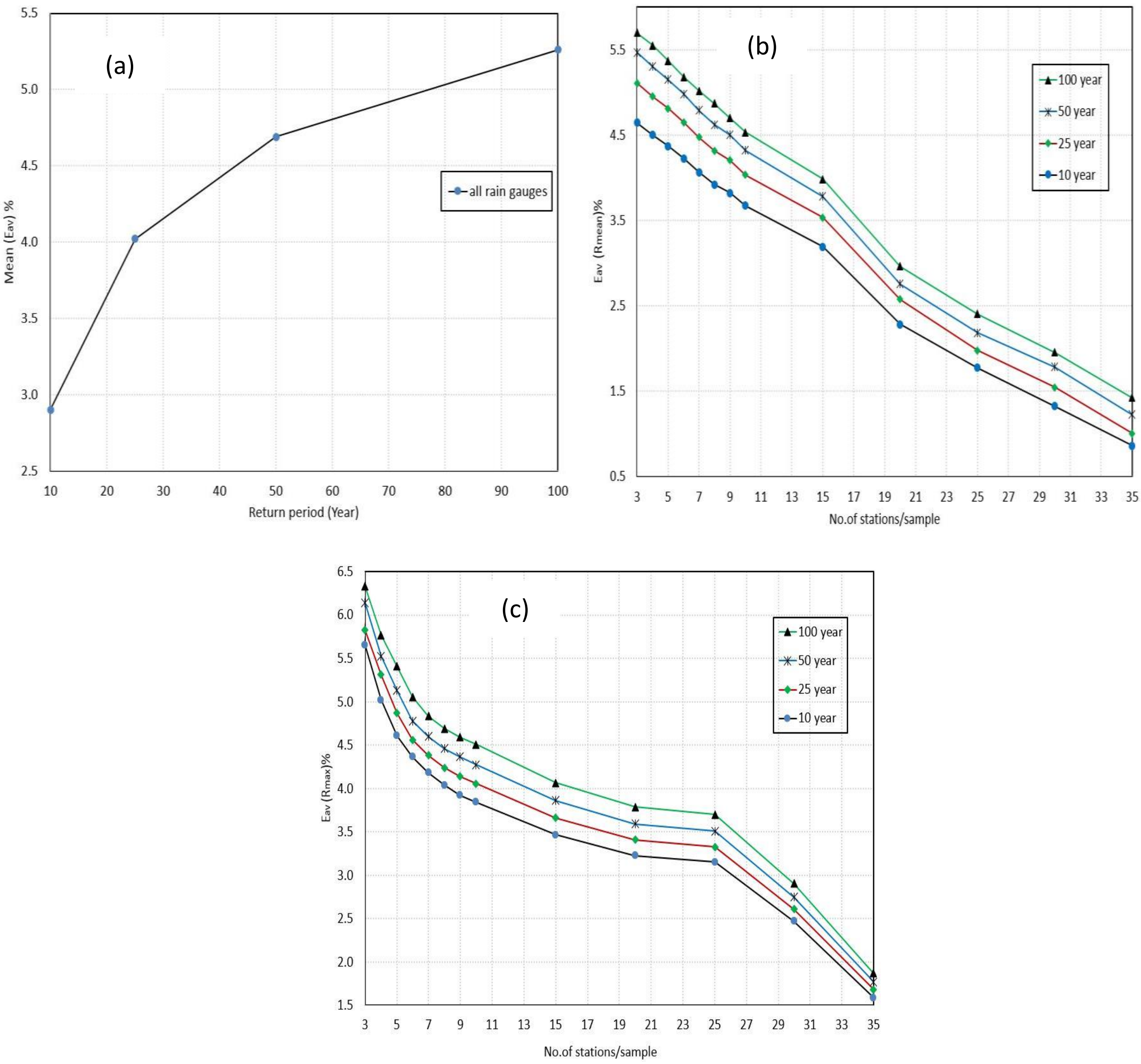



**Fig 5:** Relative error of the STRFA case (a) Mean (Eav) & (b) Eav (Rmean) and (c) Eav (Rmax).

Table (2) shows the required number of stations corresponding to the accepted relative error (Mean ($E_{av}$)). The results confirm that the obtained relative error based on the RRFA case is always less than the same error type based on the STRFA case.

**Table (2):** Mean ($E_{av}$) criteria for the RRFA and STRFA

| Average of realtive errors based the individual error of each site ( Mean Eav) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| No. of Stations | RRFA case | | | | STRFA case | | | |
| | 10 year | 25 year | 50 year | 100 year | 10 year | 25 year | 50 year | 100 year |
| 3 | 2.71 | 3.75 | 4.37 | 4.90 | 2.9 | 4.02 | 4.69 | 5.26 |
| 4 | 2.69 | 3.73 | 4.35 | 4.89 | | | | |
| 5 | 2.68 | 3.71 | 4.33 | 4.85 | | | | |
| 6 | 2.65 | 3.67 | 4.27 | 4.78 | | | | |
| 7 | 2.59 | 3.59 | 4.18 | 4.68 | | | | |
| 8 | 2.55 | 3.54 | 4.12 | 4.61 | | | | |
| 9 | 2.52 | 3.49 | 4.06 | 4.55 | | | | |
| 10 | 2.51 | 3.48 | 4.04 | 4.53 | | | | |
| 20 | 1.52 | 2.11 | 2.46 | 2.75 | | | | |
| 90 | 0.0 | 0.0 | 0.0 | 0.0 | | | | |

Table (3) shows a high difference in the required number of stations between the RRFA case and the STRFA case for the $E_{av}$ ($R_{max}$) criterion. However, the difference in the required number of stations is low in the $E_{av}$ ($R_{mean}$) criterion. It is also clear that the difference in the required number of stations obtained based on the RRFA case and based on the STRFA case increases as the return period increases in case of the $E_{av}$ ($R_{max}$) criterion. As in Table (2), the results of these two error criteria indicate that the required number of stations is less in the case of the RRFA especially in the $E_{av}$ ($R_{max}$) criterion.

**Table (3):** Required stations for the Eav ($R_{mean}$) and Eav ($R_{max}$) criteria.

| Acceptable error | The required number of stations based on the maximum rainfall depth in the watershed (Eav (Rmax)) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | RRFA case | | | | STRFA case | | | |
| | 10 year | 25 year | 50 year | 100 year | 10 year | 25 year | 50 year | 100 year |
| 2.0% | 32 | 33 | 35 | 35 | 33 | >35 | >35 | >35 |
| 3.0% | 26 | 27 | 29 | 30 | 24 | 31 | 33 | 35 |
| 4.0% | 8 | 10 | 13 | 15 | 10 | 17 | 25 | 31 |
| 5.0% | 5 | 6 | 7 | 8 | 7 | 9 | 12 | 15 |
| | The required number of stations based on the mean rainfall depth in the watershed (Eav (Rmean)) | | | | | | | |
| 1.0% | 31 | 33 | 33 | 34 | 33 | 35 | >35 | >35 |
| 2.0% | 14 | 19 | 26 | 29 | 22 | 25 | 27 | 30 |
| 3.0% | 7 | 13 | 14 | 15 | 15 | 17 | 19 | 20 |
| 4.0% | 3 | 6 | 11 | 12 | 8 | 10 | 13 | 15 |
| 5.0% | 3 | 3 | 3 | 6 | 3 | 4 | 6 | 7 |

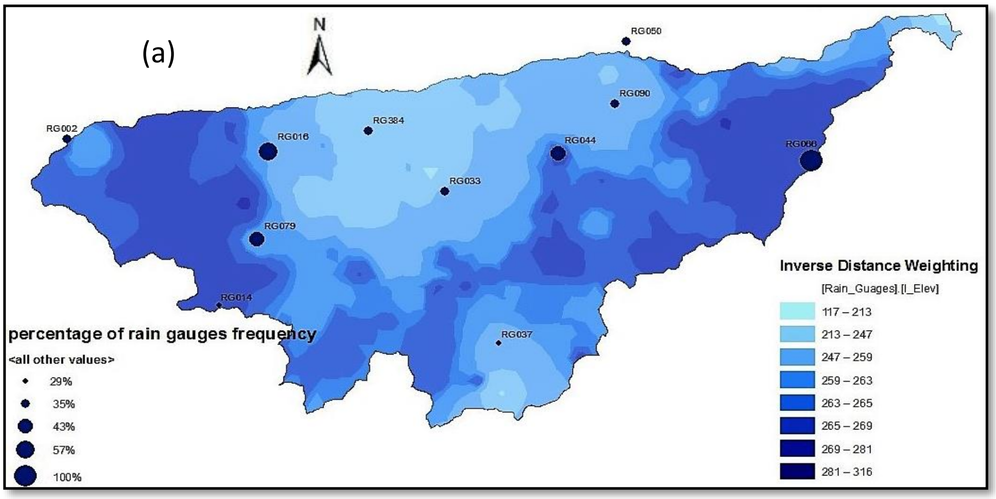
*4.6. Recommended locations to allocate the required optimum number of stations.*

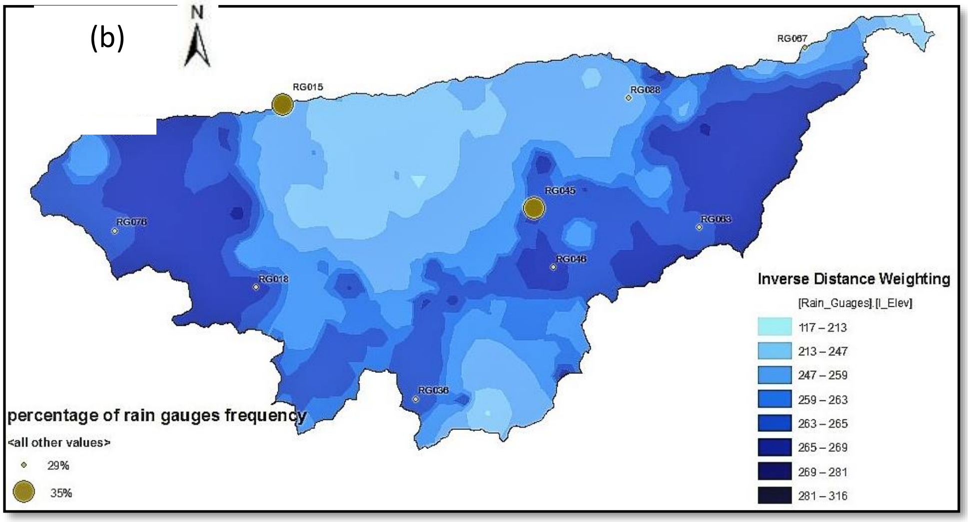
Since the results of the three studied error criteria indicate that the RRFA case gives always a lower number of required stations to achieve the desired error, the optimum sample which gives the minimum error in each criterion was determined for each sample size. The top 10 highly repeated stations in all the studied sample sizes are determined based on the percentage of appearing in the optimum samples compared to the total number of studied sample sizes. These 10 stations are shown with the isohyetal map of the WGEW in the background.

Fig (6-a) illustrates the locations of the top 10 highly repeated stations for the Mean ($E_{av}$) criterion based on the RRFA case. RG066 station repeated percentage is 100 %, which means that it appeared in all the optimum samples. It is also noted that the stations with high appearance percentage are located in the relatively high rainfall zones. There are nearly no stations among the top 10 located in the lower rainfall zone (Upper central part of WGEW).

Fig (6-b) presents the locations of the top 10 highly repeated stations for the $E_{av}$ ($R_{mean}$) criterion based on the RRFA case. It is clear that no stations have exceptional high appearance percentage in the optimum samples and the selected stations are nearly regularly distributed over the catchment. Thus, in this criterion there is no preference to any zone to allocate the selected station in it. However, the chosen stations must cover all the catchment.

Fig (6-c) illustrates the locations of the top 10 highly repeated stations for the $E_{av}$ ($R_{max}$) criterion based on the RRFA case. RG066, RG053, and RG047 have the maximum appearance percentage (57 %) of the optimum samples. In spite that, this percentage isn't high to determine recommended zones to allocate the selected stations, it is shown from the spatial distribution of the high frequently appearing stations that the lower rainfall zone is not a priority zone to allocate stations. The above findings suggest that it is better to allocate rainfall gauging stations in high rainfall zones, based on the isoheytal map, if one is interested in getting safe estimation of the daily rainfall events at high return periods.





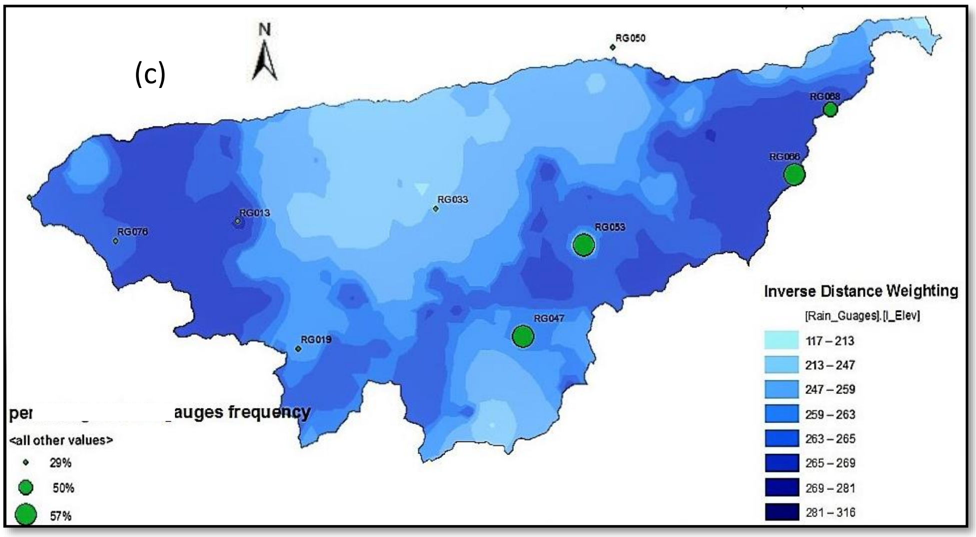




**Fig 6:** Top 10 high repeated stations based on the RRFA (a) Mean ($E_{av}$) & (b) $E_{av}$ ($R_{mean}$) and (c) $E_{av}$ ($R_{max}$).

## 5. Conclusion and recommendations

To optimize a rainfall network based on the rainfall depths calculated at high return periods, the regional frequency analysis (RRFA) and the at-site frequency analysis (STRFA) methods are used. The comparison between the RRFA and STRFA was applied using three criteria at the high return periods. The required number of stations to achieve an acceptable error are obtained for the three criteria. The three relative error criteria targeted determining the required number of stations corresponding to, an individual error at each site, satisfying the desired error in the mean rainfall depth, and the relative error in the maximum rainfall depth in the catchment, respectively. To achieve a 5 % error at the 100-year return period, the required number of stations for the three described criteria are 3, 6 and 8 stations respectively based on the RRFA method. Based on the STRFA, the least error is 5.26 % based on the entire number of stations and the required number of stations for the other two criteria are 7 and 15 stations to achieve an error of 5 % at the 100-year return period respectively. Thus it is clear that the RRFA must be applied if the region is homogeneous. The required number of stations based on the RRFA are less compared to the STRFA especially for the $E_{av}$

($R_{max}$) criterion. The relative error shows an important error decrease from 3 to 15 stations which means that increasing the number of stations is effective in reducing errors within these sample sizes. The best zones to allocate the selected rainfall stations are specified for the three criteria. It is clear that, the selected rain gauges must be allocated in the relatively high rainfall zones. Moreover, it is preferred to allocate them on the relatively high altitude sites.

## References


Akaike, H. (1973). Information theory and extension of the maximum Likelihood principle. In: B.N. Petrov & F. Csaki, eds. 2nd international symposium on information theory. Budapest,Hongry: Akadémiai Kiado, 267 281.

Akaike, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Ann I Stat Math, 26, 363–387.

Adhikary, S. K., Yilmaz, A. G., & Muttil, N. (2015). Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. Hydrological Processes, 29(11), 2582–2599. https://doi.org/10.1002/hyp.10389

Al-abadi, A. M., & Al-aboodi, A. H. D. (2014). Optimum Rain-Gauges Network Design of Some Cities in Iraq : ةصلاخلا, (4), 946–958.

Ayyub, B. M., & McCuen, R. H. (2011). Probability, statistics, and reliability for engineers and scientists. CRC press.

Bogardi, I., Bardossy, A., (1985). Multicriterion network design using geostatistics, Water Resour. Res., 21 (2), 199–208.

Caffey, J. E. (1965) Inter-station correlations in annual precipitation and annual effective precipitation. Hydrology Paper 6, Colorado State Univ., Fort Collins, Colorado, USA.

Fisher, Ronald Aylmer. (1912). "001: On an Absolute Criterion for Fitting Frequency Curves.". Guide to Hydrological Practices, Volume-I: “Hydrology – From Measurement to Hydrological Information”, WMO No. 168.

Haggag, M., Ali, A., & Awadallah, A. (2016). Evaluation of Rain Gauge Network in Arid Regions Using Geostatistical Approach : Case Study In Northern Oman, 1–22.

Heilman, P., Nichols, M. H., Goodrich, D. C., Miller, S. N., & Guertin, D. P. (2008). Geographic information systems database , Walnut Gulch Experimental Watershed , Arizona , United States, 44(November 2007), 1–6. https://doi.org/10.1029/2006WR005777

Hughes, J. P. & Lettenmaier, D. P. (1981). Data requirements for kriging: estimation and network design. Wat. Resour. Res. 17(6), 1641-165

Khairul, M., Mohd, B., Yusof, F., Daud, Z. M., Yusop, Z., Afif, M., … Technology, A. (2016). Optimal Design of Rain Gauge Network in Johor By, 11(25), 2422–2428.

Matheron, G. (1971). The Theory of Regionalized Variables and its Applications Ecole Des Mines, Fontainebleau, France.

McKay, M. D., W. J. Conover, & R. J. Beckman (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245.

National Research Council (1988). Committee on Techniques for Estimating Probabilities of Extreme Floods, “Estimating Probabilities of Extreme Floods, Methods and Recommended Research”, National Academy Press,Washington, D.C.

Schwarz G. (1978). Estimating the dimension of a model. Ann Stat, 6, 461–464.

Stedinger, J.R., Vogel, R.M. & E. Foufoula-Georgiou, E. (1993). Frequency Analysis of Extreme Events, in “Handbook of Hydrology”, ed. D.R. Maidment, McGraw-Hill, New York, NY, pp. 18.1-18.66.

Stol, PT. (1972). The relative efficiency of the density of rain gauge net- works. J Hydrol 15:193–208s

Pardo-Igúzquiza, E. (1998). Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. Journal of Hydrology 210(1–4), 206–220. https://doi.org/10.1016/S0022-1694(98)00188-7

Patel, Anant D. and Project Manager. (2016). “Analysis of Optimum Number of Rain Gauge in Shetrunji River Basin , Gujarat - India.” 2(11):380–84.

Pearson, Karl. (1894). Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London, Vol. 185, 71-110

Putthividhya, A., & Tanaka, K. (2012). Optimal Rain Gauge Network Design and Spatial Precipitation Mapping Based on Geostatistical Analysis from Colocated Elevation and Humidity Data, 3(2).

Rodriguez-Iturbe, I., Megia, J.M., (1974). The design of rainfall networks in time and space, Water Resour. Res., 10 (4), 713–728.

Rouhani, S., (1985). Variance reduction analysis, Water Resour. Res., 21 (6), 837–846.

U.S. Water Resources Council (1981). “Guidelines for determining flood flow frequency”, Bulletin 17B, Hydrology Committee, Water Resources Research Council, Washington.

Vogel, R.M. & Fennessey, N.M. L moment diagrams should replace product moment diagrams.Water Resour Res 1993, 29, 1745–1752.

Wiltshire, S.E. (1986a). “Regional Flood Frequency Analysis I: Homogeneity Statistics.” Hydrological Sciences Journal 31(3):321–33. Retrieved (http://dx.doi.org/10.1080/02626668609491051).

Wiltshire, S.E. (1986b). “Identification of Homogeneous Regions for Flood Frequency Analysis”, Journal of Hydraulics, Vol. 84, pp. 287-307.