Flood Frequency Analysis through Hydrological Simulation and Regionalization

Ali Reza Matin

Supervisor: Athanasios Loukas
Co-supervisor: Lampros Vasiliades

Master Management
of
Hydrometerological
hasards
August 2013-January 2014

Acknowledgment

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First, my supreme gratitude to Master Management of Hydrometeorological Hasard under the direction of Joseph Fourier and Thessaly Universities.

Foremost, I would like to express my sincere gratitude to my advisor Professor. Athanasios Loukas for the continuous support of my thesis study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of my project and writing of this thesis.

Deepest gratitude is also due to the members of the jury, Professor. Sandrine Anquetin and Professor.Gilles Molinie for their encouragement, insightful comments, knowledge and hard questions that are greatly valued.

Furthermore, I would like to acknowledge the help I received from my Co-Supervisor Lampros Vasiliades and Lab assistant specially Mr.George Papaioannou for providing laboratory facilities. I have benefited from their pieces of advice and much encouragement.

Not forgetting to thanks the assistance provided by Mr. Brice Boudevillain for retrieving and providing the hydro-meteorological data from the SEVnOL and Meteo France databases and the UJF Ph.D. Candidate Olivier Vannier for providing the geographical data.

Last but not least, I would like to thank my parents for their unconditional support, both financially and emotionally throughout my degree. In particular, the patience and understanding shown by my mum, dad and brothers during the honors year is greatly appreciated. I know, at times, my temper is particularly trying.

Finally, I would like to thank the French Government making this facility of my study in Europe.

Table of Contents

List of Figures

List of TABLES:

Abstract

Flow simulation of 26 watersheds of Cevennes area (southern France) have been performed with the GR4J rainfall- runoff hydrological model. The parameters of model acquired through calibration of GR4J model with daily observed meteorological and hydrological data. Areal average precipitation, potential evapotranspiration and flow data were used as input data to GR4J model. Areal average precipitation and potential evapotranspiration for each watershed was estimated by applying Thiessen polygon method and Oudin et al., (2005) formula, respectively. The GR4J model has been calibrated using the Shuffled Complex Evolution (SCE) optimization algorithm and Nash-Sutcliffe Efficiency (NSE) has been used as the objective function. The geomorphological, geological and land use characteristics of the study watersheds were extracted using G.I.S. and used for the estimation of the variation of hydrological model parameters and their regionalization. Flood frequency analysis of the observed flows, the simulated flows using the calibrated hydrological model and the simulated flows using the regionalized hydrological model, has been performed using the Generalized Extreme Value (GEV) probability distribution and the L-Moment parameter estimation method for the maximum annual flows.

Resume

Les simulations de ruissellement de 26 bassins versants de la région des Cévennes (Sud de la France) se déroulent en utilisant le modèle hydrologique GR4J, modèle de ruissellement des précipitations. Les paramètres du modèle sont réalisés par calibration du modèle GR4J avec les données observées quotidiennement. Les précipitations moyennes de surface, L'évapotranspiration potentielle et le débit des données observées étaient nécessaires comme données d'entrée de modèle GR4J. Les précipitations moyennes de surface pour chaque bassin versant ont été estimées grâce à l'application de la méthode des polygones de Thiessen (3) et l'évapotranspiration potentielle est calculée en utilisant la formule d'Oudin et al. , (2005). Le modèle GR4J a été calibré en utilisant 'Shuffled Complex Evolution (SCE)' l'algorithme d'optimisation et 'Nash-Sutcliffe Efficiency (NSE)' a été utilisés pour la fonction objective. Les caractéristiques géomorphologiques, géologiques et d'occupation du sol de l'étude des bassins versant ont été extraites en utilisant S.I.G et ont été utilisée pour l'estimation de la variation des paramètres du modèle hydrologique et de leur régionalisation. L'analyse de la fréquence des crues des flux observés, les flux stimulés utilisent le modèle de calibrage hydrologique et les flux simulés utilisent le modèle de régionalisation hydrologique, ils ont été effectués en utilisant la Valeur Extrême Généralisée (GEV) la distribution des probabilités et la méthode d'estimation des paramètres du Moment -L pour le flux annuel maximum.

Περίληψη

Οι προσομοιώσεις των 26 λεκανών απορροής της περιοχής Cevennes (Νότια Γαλλία) πραγματοποιήθηκαν με την χρήση του υδρολογικού μοντέλου βροχής-απορροής GR4J. Οι παράμετροι του μοντέλου βαθμονομήθηκαν χρησιμοποιώντας ημερήσιες μετεωρολογικές και υδρολογικές μετρήσεις. Η μέση επιφανειακή βροχόπτωση, δυνατική εξατμισοδιαπνοή και παροχή ήταν τα δεδομένα παρατήρησης που χρησιμοποιήθηκαν ως δεδομένα εισόδου στο μοντέλο GR4J. Η μέση επιφανειακή βροχόπτωση για κάθε λεκάνη απορροής υπολογίστηκε εφαρμόζοντας την μέθοδο των πολυγώνων Thiessen και Oudin et al.,(2005) μέθοδος, αντίστοιχα. Το μοντέλο GR4J βαθμονομήθηκε χρησιμοποιώντας τον αλγόριθμο βελτιστοποίησης Shuffled Complex Evolution (SCE) και η Nash-Sutcliffe Efficiency (NSE) χρησιμοποιήθηκε ως αντικειμενική συνάρτηση. Τα γεωμορφολογικά και γεωλογικά χαρακτηριστικά, καθώς και οι χρήσεις γης των υπό μελέτη λεκανών απορροής εκτιμήθηκαν από το πρόγραμμα GIS και χρησιμοποιήθηκαν για την εκτίμηση της μεταβλητότητας των παραμέτρων του υδρολογικού μοντέλου. Η ανάλυση της συχνότητας των μεγίστων παροχών, που έχουν παρατηρηθεί, που έχουν προσομοιωθεί χρησιμοποιώντας το βαθμονομημένο υδρολογικό μοντέλο και που έχουν προσομοιωθεί χρησιμοποιώντας το υδρολογικό μοντέλο με τις εκτιμημένες περιοχικές τιμές των παραμέτρων, πραγματοποιήθηκε με την χρήση της συνάρτησης κατανομής πιθανότητας Generalized Extreme Value (GEV) και της μεθόδου υπολογισμού παραμέτρων L-Moment για τις μέγιστες ετήσιες πλημμύρες.

I. Introduction

I.1. Motivation

Floods are the most destructive of natural hazards. Governments are trying to control and decrease the vulnerability and damage, cause of this hazard. They invest money and time to find out a way to forecast and predict of occurrence of flood and prevent the damage that floods may cause. Flood frequency analysis relates the magnitude of extreme events to their frequency of occurrence through the use of theoretical probability distributions (Chow et al., 1988). One of the most important difficulties for flood frequency analysis is the scarcity of hydrometric data. The aim of this work is to use a conceptual hydrological model to simulate the hydrological response of watersheds under flood conditions and to produce reliable results for the estimation of flood magnitude and frequency in ungauged watersheds.

I.2. General View

This work has important phases that need to follow to achieve the flood frequency analysis on ungauged catchments. Firstly, the database should be compiled. It contains meteorological observed data from our catchments, hydrometric data, and catchments characteristics. Meteorological data include daily precipitation and temperature. Hydrometric data is daily discharge from the study catchments. Geology, geomorphology and land use characteristics are extracted by using GIS software for each separate watershed.

Secondly, the data should be processed in order to be used in the hydrological simulation. Average areal precipitation, temperature and potential evapotranspiration are estimated for all study watersheds. Thiessen polygon method is used for the estimation of average areal precipitation and temperature. Oudin et al., (Oudin et al., 2005) formula is used for the estimation of potential evapotranspiration. Daily areal average rainfall (P), daily potential evapotranspiration (E), and daily discharge (Q) estimated for 26 different watersheds.

The final phase is to simulate the discharge of the study watersheds and extract the annual maxima discharge and using these data for flood frequency analysis. The flow simulation is performed using hydrological models.

Hydrological models are standard tools routinely used today for hydrological investigation in engineering and environmental science. They are applied to extend stream flow time series in space and time. The hydrological models structures usually a combination of linear and nonlinear functions have been developed and implemented in software since the early 1960s. Therefore for more precise view, it is necessary to classify these structures. Probably the most commonly applied classification is one that uses three distinct classes (Wheater et al., 1993).

These are:

1) Metric, also called data-base, empirical or black box which is clear from the name of this model is depend on available time series data and they are completely based on information elicit from input data and do not include any basic information about condition and behavior of catchments.

2) Parametric also called conceptual, explicit soil moisture accounting or grey box that use a storage elements as the main building component and this storage are filled through fluxes such as precipitation, infiltration or percolation and emptied through evapotranspiration, runoff, drainage and etc. However this model type is still based on time series data especially stream flow data in calibration procedure to estimate the model parameter values. Their dependence on particular stream flow measurement makes them difficult to apply in ungauged catchments. In this work, the GR4J rainfall-runoff hydrological model, which is a parametric conceptual model, is used to simulate the stream flows.

3) Physically based models (Freeze and Harlan, 1969; Beven , 2002) are based on the concepts of mass, momentum and energy conservation. Physical realism is the corner-stone of model to relate their parameters such as soil moisture characteristics and unsaturated zone hydraulic conductivity function for subsurface flow, friction coefficients for surface flow of physical characteristics of the catchments (Todini, 1988), thus eliminating the need for model calibration.

Flood Frequency Analysis (FFA) is used to find the relation between flood magnitude and its frequency of occurrence (or return time). Three different methods maybe considered for the extraction of the flood peaks form the observed and/or simulated discharge time series (Cunnane, 1989). These methods are (1) Annual maximum series (AM) method, (2) the partial duration series (PD) or pick over a threshold (POT) method, and (3) the time series (TS) method. However, if the AM series method is used important information may be lost. For example, the second or the third maximum annual flow of a particular year may be greater than the maximum annual flow of some dry years, and they are not used in the AM method (Kite, 1977; Chow et al. 1988). This situation is avoided when the partial duration (PD) or the peaks over a threshold (POT) methods are used. In this study, the AM method is used and the theoretical General Extreme value (GEV) distribution is fitted to the observed and simulated peak flow time series. The Kolmogorov-Smirnov test has been used to test the goodness-of-fit between the empirical and the theoretical probability distribution.

II. Study Area

II.1 Location

The study area is located in the southern France. The Cévennes area is a range of mountains in south-central France, covering parts of the departments of Ardèche, Gard, Hérault and Lozère. The Cévennes are part of the Massif Central region. It runs from southwest (Montagne Noire) to northeast (Monts du Vivarais), with the highest point being the Mont Lozère (1702m). Another notable peak is the Mont Aigoual (1567m). The Loire and Allier Rivers are flowing towards the Atlantic ocean, the Ardèche and its tributaries Chassezac and Cèze Rivers and the Gardons Rivers to the Rhône, Vidourle, Hérault and Dourbie Rivers to the Mediterranean Sea. Figure 1 presents the study area in France.

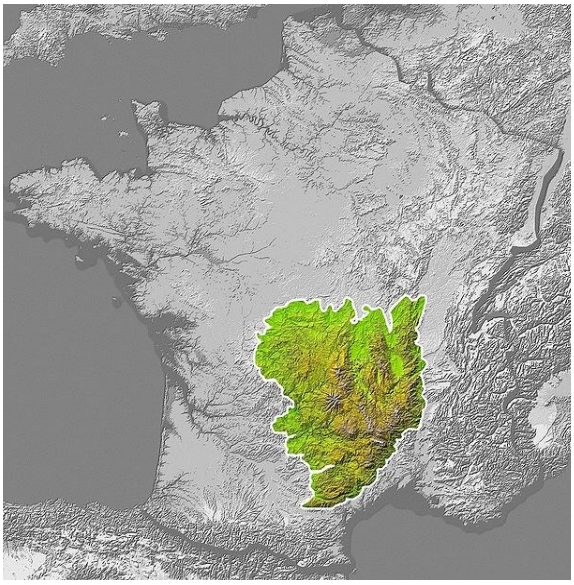

Figure 1: Location of the Cévennes area (Wekipidea.com)

II.2 Data base

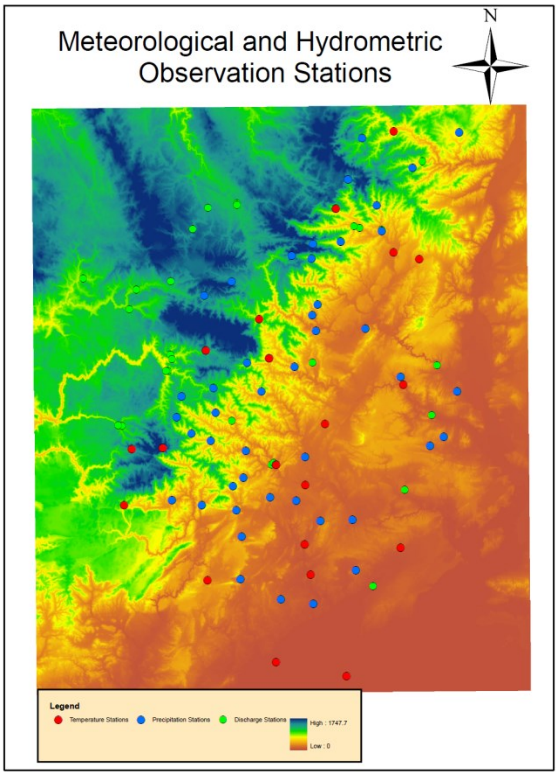

Figure 2: Meteorological and Hydrometric stations of study area case.

Daily precipitation, temperature and discharge are the meteorological and hydrometric data set of case of study. The OHM-CV (Observatoire Hydrométéorologique Méditerranéen Cévennes Vivarais) developed observation devices mainly for flash floods and extreme precipitation analysis in this area. Daily precipitation of 49 different stations and data from 22 meteorological stations were available for precipitation and temperature data, respectively. Discharge stations were important for flow data retrieval and for reference points for watershed delineation. In total, 36 discharge stations were available but because of large data gaps or limited time series length, the data from 26 flow stations have been finally used. The flow time series have been retrieved from the France Hydro-Banque database (http://www.hydro.eaufrance.fr). Figure 2 shows the location of meteorological and hydrometric stations. Red points show the location of temperature stations, blue points show the precipitation stations and green points show the discharge stations.

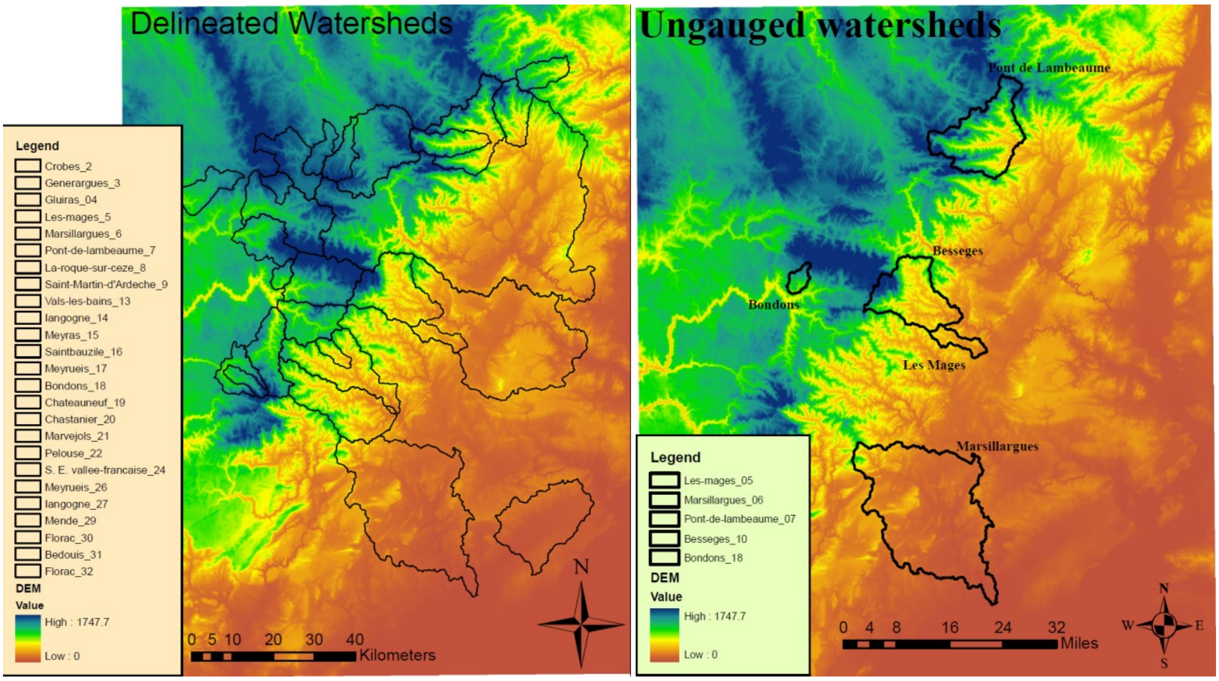

Figure 3: Location of the 26 study watersheds (left) the name of watersheds came from discharge observed station name, also number shows the number of watershed in first total 36 stations, and location of the five (5) selected watersheds for the test of regionalization and flood frequency analysis for ungauged watersheds (right).

Other part of our database is the watersheds geological and geomorphological characteristics. The location of discharge stations was used for automatic delineation of the 26 study watersheds. Five (5) of these watersheds have been selected for testing the regionalization method and flood frequency analysis for ungauged watersheds (Fig. 3). For selection of five watersheds for regionalization analysis was tried to cover the most characteristics of total watersheds. Table 1, represented the maximum and minimum of geomorphology characteristics for all 26 watersheds and five selected watershed. Figure 4 and 5 represented the land use and geology (permeability) characteristics of 26 case study watersheds, respectively.

Table 1: Variation of Watersheds Geomorphology characteristics

Watersheds Characteristics	Total Watersheds		Five Selected Watersheds	
	Max	Min	Max	Min
Area (Sq. Km)	2263.48	25.53	796.16	25.53
MEAN Elevation of W_S (m)	1282.81	79.20	986.91	152.05
MEAN SLOP OF W_S (DEGREE)	23.14	2.41	21.53	6.34
MAIN RIVER LENGTH (KM)	110.88	2.76	110.88	2.76
MEAN SLOP OF MAIN RIVER (DEGREE)	7.28	0.24	7.28	2.24

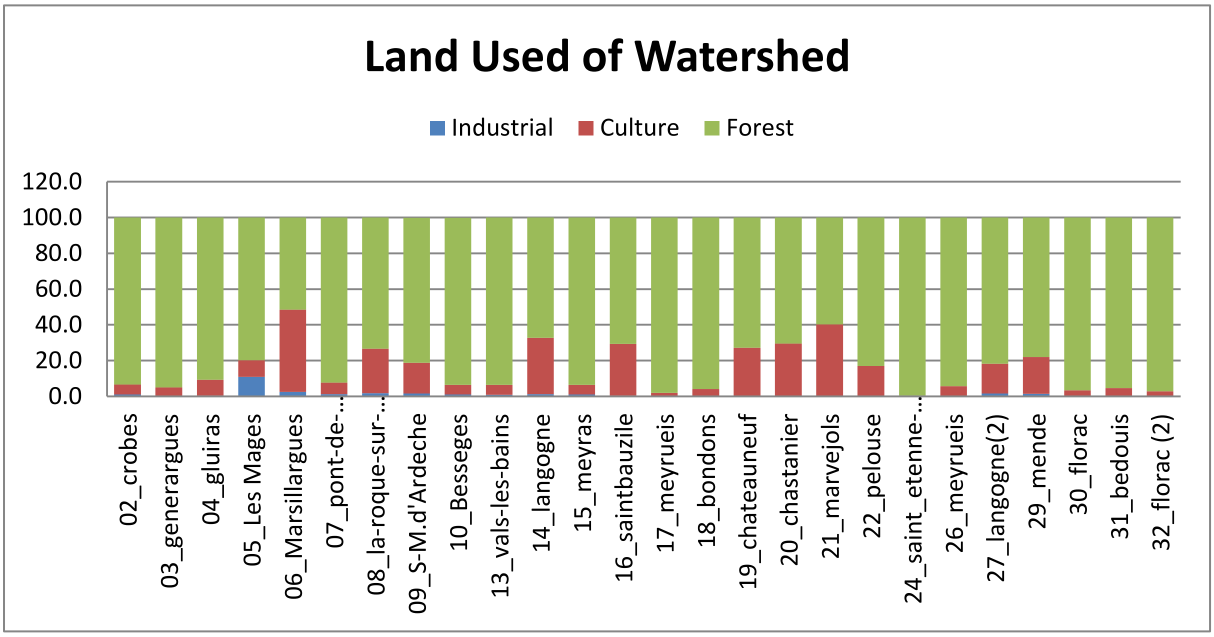

Figure 4: Percentage of Land use in 26 watersheds

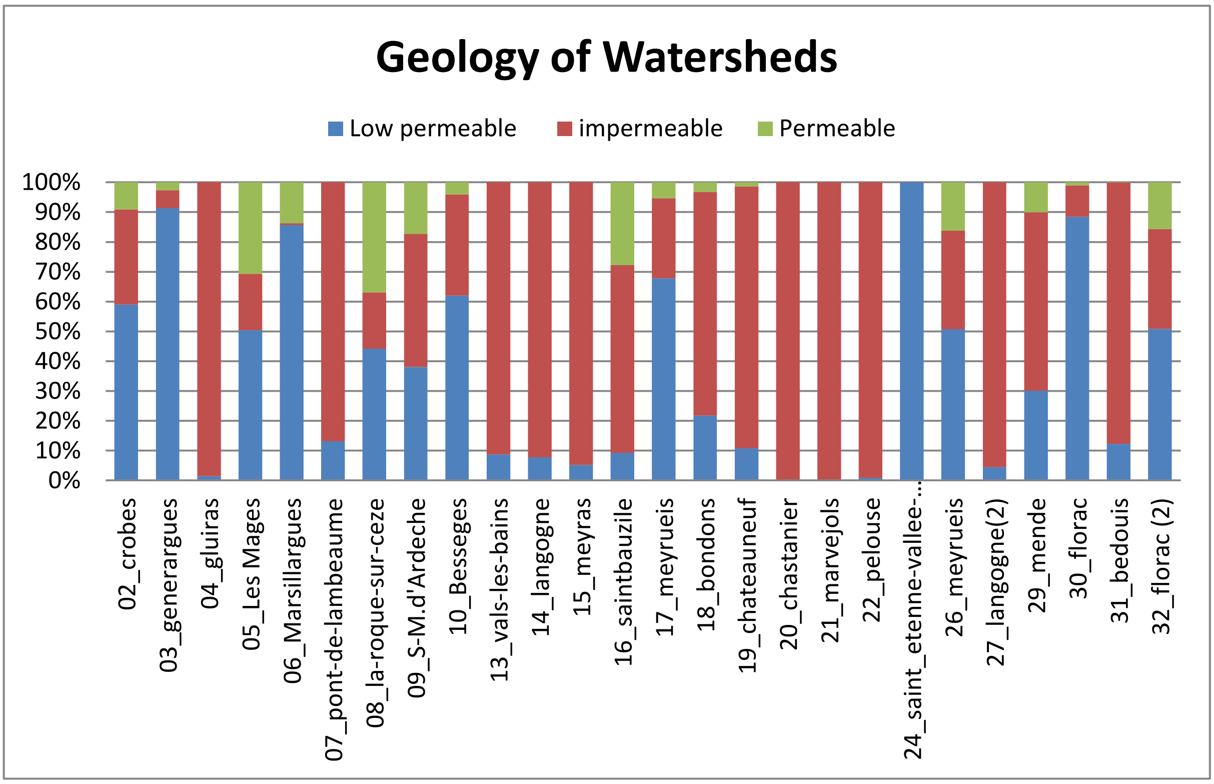

Figure 5: Combination of rocks with geology (permeability) characteristic in all watersheds

II.3 Preprocessing of observation database

Areal average precipitation, potential evapotranspiration, and discharge are necessary data for hydrological simulation. Estimation of the areal average precipitation and temperature for each catchment has done by using of the Thiessen Polygon Method. The method proposed by Oudin et al. (Oudin et al., 2005) is used to estimate the potential evapotranspiration. Areal average precipitation, potential evapotranspiration and discharge of each of 26 study catchments prepared. Land use, soil, elevation, slope and etc. are important characteristics that may affect the occurrence of floods in a watershed. The geomorphological, geological and land use characteristics of watersheds were extracted by using GIS software and digital maps of elevation, geology and land use for the Cévennes area and the watershed characteristic have been acquired.

III Methodology

III.1 Preprocessing

A.1.1 Organize of Database: Daily precipitation, temperature and discharge are the observed meteorological and hydrometric data. These data need preprocess to become ready to use as input data to hydrological analysis. Preprocessing procedure are presented in the next paragraphs.

a. Delineation of the study watersheds was done using GIS using the location of flow stations and digital elevation maps. Digital maps of elevation, geology and land use of the study area were used to extract geomorphological characteristics of study watersheds, such as area, slope, land use etc.

b. Estimation of Areal Average precipitation and Temperature by Thiessen Polygon Method.

The Thiessen method is used to find the areal values of precipitation and temperature by the means of weighing factors for each observation stations. The weighing factor is based on the ratio of the area of the watershed influenced by each station over the total area of the watershed. These areas are created by the bisects of the distance between two neighboring stations and the watershed border and they are irregular polygon. The steps followed in this method are presented below:

1. The location of the stations are located on the watershed map.

2. Adjacent stations are connected with lines.

3. Perpendicular bisectors of each line are constructed (perpendicular line at the midpoint of each line connecting two stations)

4. The bisectors are extended and used to form the polygon around each gauge station.

5. Precipitation value for each gauge station is multiplied by the area of each polygon.

6. All values from step 5 are summed and divided by total basin area.

Average precipitation regions are formulated as follows:

$$\mathrm{P} = \frac{P1A1 + P2A2 + P3A3 + \cdots + PnAn}{A1 + A2 + A3 + \cdots + An} \qquad (1)$$

In this case:

P = precipitation region (mm)

P1, P2, P3,., Pn = precipitation each observation station (mm)

A1, A2, A3,., An = area of each polygon

The above procedure is automatically performed in GIS software with accuracy and in less time. Figure 6 shows and example of Thiessen Polygon methods. Numbers represent the amount of rainfall for each station of the watershed

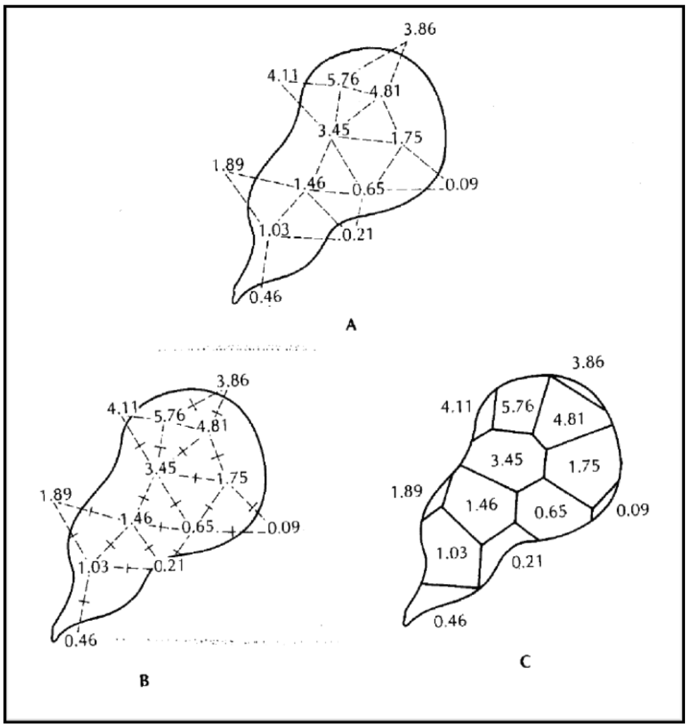

Figure 6: Graphical presentation of Thissen Polygon Method **A**. The stations are connected with lines. **B**. The perpendicular bisector of each line is found. **C**. The bisectors are extended to form the polygons around each station

c. Modification and Filling Gaps of database by linear methods. To clarify more about linear model, I start with a simple equation that shown in below.

$$y = mx + b \tag{2}$$

The standard way to graph this equation place the value of the x on horizontal axis and the y value on the vertical axis and (m) represent the slope of the line and (b) is the intercept of (y) value when the (x) value is equal to zero. This equation can denote a model (figure 7).

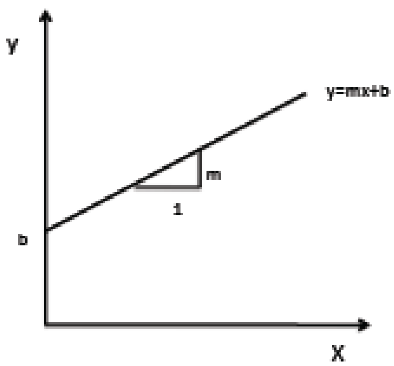

Figure 7: Linear equation

In a linear model, the slope describes how much of an effect x has on y. Elevation is one of the factors that effect on temperature and amount and kind of precipitation. Difference in elevation of precipitation and temperature observation station and mean elevation of watersheds was the reason to use linear model to reduce the error of orographic effects on precipitation and temperature. Modification database was prepared after applying linear model

.

d. Estimation of areal potential evapotranspiration. The areal potential evapotranspiration (PE) has been estimated using the Oudin et al. formula (L. Oudin et al., 2005), which is given in Equation 3.

$$\mathrm{PE} = \begin{cases} \frac{0.408 R_e \ (T+5)}{100} & if\ (T+5) > 0 \\ 0 & otherwise \end{cases} \qquad (3)$$

Where Re is the extraterrestrial solar radiation (MJ m−2 d−1) given by the Julian day and the latitude, and T is the mean air temperature at a 2-m height (°C).

III.2 Processing

III.2.1: Selecting the Hydrology Model,

The GR4J model has been selected for the hydrological analysis of this study. The model has been developed by Perrin and his associates (Perrin et al., 2003). The model has been tested in 429 different catchment located in different climate regions, ranging from semi-arid areas to temperate and tropical humid areas. This test assures the validity of the model and its application in the study area.

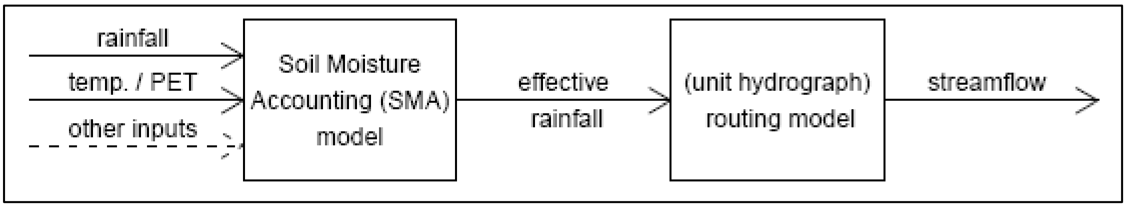

Figure 8: The GR4J model framework (Andrews et al., 2011)

The GR4J is a daily lumped rainfall runoff model that is the improvement version of GR3J during the 15 years process (C. Perrin et al., 2003). The GR4J is a parametric model that based on two component structures: 1) soil moisture accounting (SAM) module; and 2) is a routing or hydrograph module. Figure 8; represent the parametric hydrological model frame work. Precipitation depth (P), potential evapotranspiration (E) and observed flow were input data of four (4) parameter model (GR4J). Where the parameters are; (X1) the maximum capacity of production store (mm), (X2) groundwater exchange coefficient (mm), (X3) one day ahead maximum capacity of routing store (mm), and (X4) time base of unit hydrograph UH1 (days). Figure 9 represent the model diagram of GR4J. All water quantities are expected in millimeter (mm) in this model. Description of physical process in GR4J model that started from rainfall to runoff a river shows in figure 10. Production store (X1) is storage of water on surface of soil that capacity of this storage links more with slope, geology, land use and etc. percolation and evapotranspiration also effect on value of storage. Ground water exchange coefficient (X2) is a function of groundwater exchange which influence routing store. This parameter can have negative and positive value that negative values shows entrance of water to depth aquifer and positive values depends on exiting of water from aquifer to routing storage. Routing storage (X3) is amount of water that soil can keep in its porous. Soil moisture and type of soil can effect on amount of this parameter in GR4J model. And the last parameter is Time peak (X4) is the time of ordinate peak of flood that GR4J model is created from runoff that 90% of runoff is slow flow, which can infiltrates into the soil and other 10% is fast flow, which running on soil surface. Step of calculation of these four parameters explained with details in (Perrin et al., 2003).

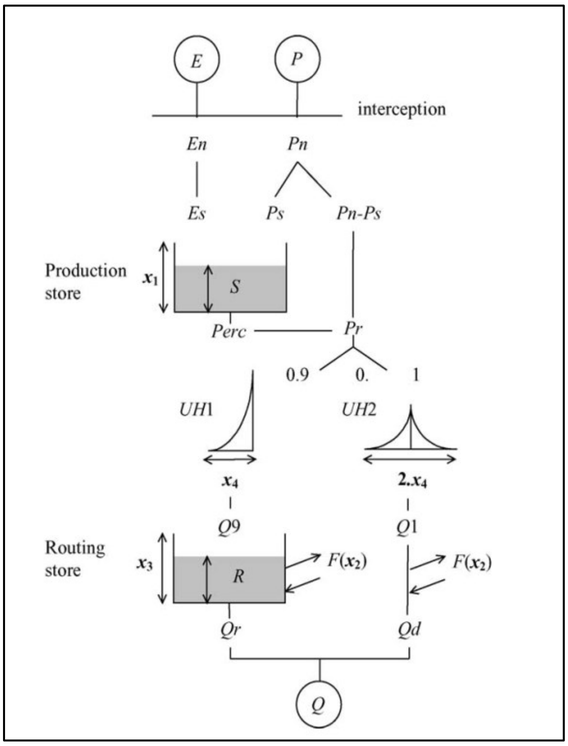

Figure 9: GR4J model diagram (Perrin et al., 2003)

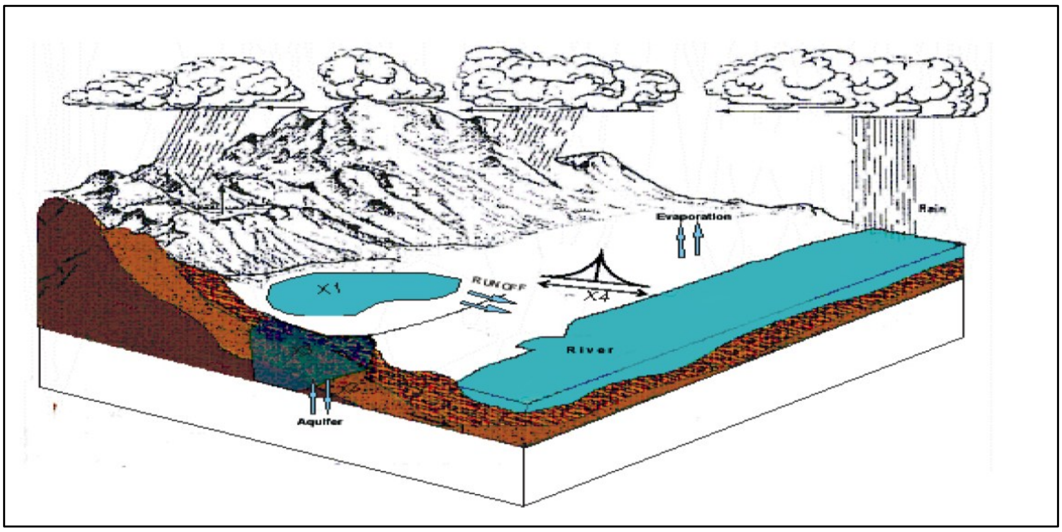

Figure 10: Physic Description of GR4J model (Harlan et al. 2010)

III.2.2 Optimization Algorithm and Objective function methods

A) The GR4J hydrological model of this case of study contains parameters that cannot calculate directly with model. Optimization Algorithm helps to adjust the parameters values to be accordance with input and output behavior of the model. Duan et al. (1992)

found that, on simple model structure and with absence of model structure error or input data error, the parameter estimation problems are not trivial and these problems summarized in table (2).

Table 2: Summary of five major characteristics complicating in optimization problem

Problems	Area of Problems
1. Regions of attraction	More than one main convergence region
2. Minor local optima	Many small ‘pits’ in each region
3. Roughness	Rough response surface with discontinuous derivatives
4. Sensitivity	Poor and varying sensitivity of response surface in region of optimum and non-linear parameter interaction
5. Shape	Non-convex response surface with long curved ridges

Table (1) shows five characterized problem of conceptual watersheds model calibration that are typical of many optimization problems, users face by them in different conditions and an optimization algorithm could deal with those problems to become successful. An optimization algorithm that aims to deal with them must possess the following properties: (1) global convergence in the presence of multiple regions of attraction; (2) ability to avoid being trapped by small pits and bumps on the objective function surface; (3) robustness in the presence of differing parameter sensitivities and parameter interdependence; (4) non-reliance on the availability of an explicit expression for the objective function or the derivatives; (5) capability of handling high-parameter dimensionality. The shuffled complex evolution (SCE) that developed in university of Arizona (UA) specially designed to deal with peculiarities encountered in conceptual watershed model calibration. Four concepts are the base of the method: 1) combination of deterministic and probabilistic approached, 2) systematic evolution of a “complex” of points spanning the parameter space in the direction of global improvement, 3) competitive evolution, 4) complex shuffling. The composition of these four concept, caused that SEC_AU be an effective and robust method.

The Shuffled Complex Evolution (SCE) algorithm finds a global minimum of a function of several variables. Initially, a set of points are drawn randomly from the specified distributions. Each point consists of a set of values of the calibration parameters. For each point, a cost is assigned. These points are then ordered and grouped into "complexes" based on their costs. The next step is an iterative procedure, where the first step is to divide each complex into "simplexes" and propagate each simplex to find a new point with smaller cost using the simplex method. Afterwards, the complexes are merged back; all the points are reshuffled and regrouped into a new set of complexes. After each iteration the points will tend to become neighbors of each other around the global minimum of the cost function.

SCE method has seven steps that there are:

(1) Generate sample--sample **s** points randomly in the feasible parameter space and compute the criterion value at each point. In the absence of prior information on the approximate location of the global optimum, use a uniform probability distribution to generate a sample.

(2) Rank points--sort the **s** points in order of increasing criterion value so that the first point represents the smallest criterion value and the last point represents the largest criterion value (assuming that the goal is to minimize the criterion value).

(3) Partition into complexes--partition the **s** points into **p** complexes, each containing **m** points. The complexes are partitioned such that the first complex contains every **p**(k - 1) + 1 ranked point, the second complex contains every **p**(k - 1) + 2 ranked point, and so on, where k = 1,2 . . . . . m.

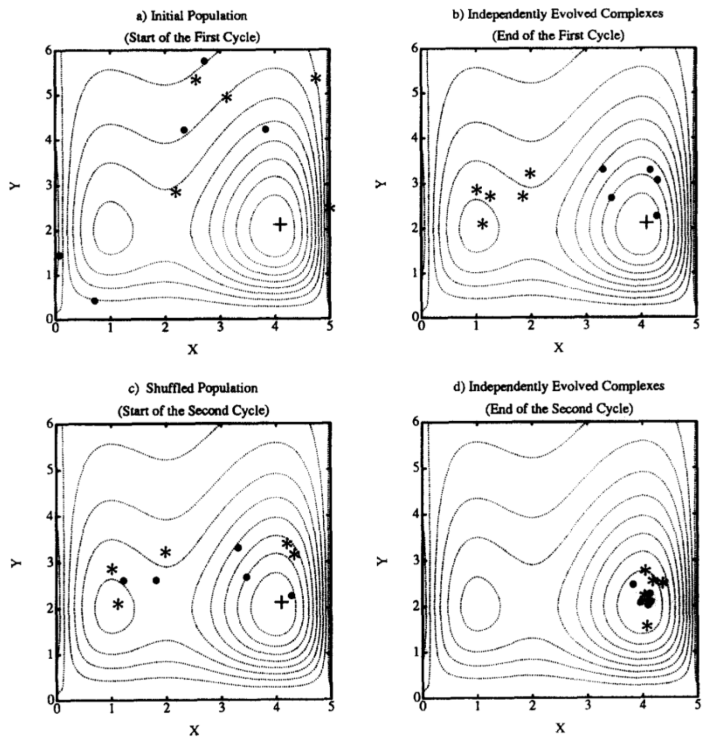

Figure 11: Illustration of the shuffled complex evolution (SCE-UA) method (Duan et al., 1994).

(4) Evolve each complex--evolve each complex according to the competitive complex evolution (CCE) algorithm (which is elaborated below).

(5) Shuffle complexes--combine the points in the evolved complexes into a single sample population; sort the sample population in order of increasing criterion value; the sample population into **p** complexes according to the procedure specified in Step 3.

(6) Check convergence--if any of the pre-specified convergence criteria are satisfied, stop; otherwise, continue.

(7) Check the reduction in the number of complexes--if the minimum number of complexes required in the population, Pmin, is less than p, remove the complex with the lowest ranked points; set p = p - 1 and s = pro; return to Step 4. If Pmin = p, return to step 4.

The SCE-UA method is explained in Fig. 11, by use of a two-dimensional example. The contour lines represent a function surface with a global optimum located at (4, 2) and a local optimum located at (1, 2). Fig. 11 (a) shows that a sample population containing s (in this case, 10) points is divided into p (two) communities (complexes), each containing m (five) members, marked by • and,, respectively. As each community undergoes an independent evolution process, one community (marked by *) is converging toward the local optimum, whereas the other (marked by •) is converging toward the global optimum. The locations of the members in the two evolved communities at the end of the first evolution cycle are illustrated in Fig. 11(b) (to demonstrate clearly the scenario that the two complexes were converging toward two distinct optima, the number of evolution steps taken by each complex, β, was set to a relatively large value of 10. The two evolved communities are shuffled according to the procedure specified in Step 5. The new memberships of the two evolved communities after shuffling are displayed in Fig. 11 (c), and the two communities at the end of the second evolution cycle are shown in Fig. 11 (d). It is clear that both communities are now converging toward the global optimum (Duan et al., 1994).

B) 1.The Nash–Sutcliffe efficiency (NSE), defined by Nash and Sutcliffe, (1970) and mean squared error (MSE) optimization method is the most common criterion that used for calibration and validation of hydrological models with observed data. The NSE optimization method is used to assess the predicative power of hydrological model that it is defined as:

$$MSE = \frac{1}{n} * \sum_{t-1}^{n} \left(x_{s,t} - x_{0,t}\right)^2 \quad (4)$$

$$NSE = 1 - \frac{\sum_{t=1}^{n}\left(x_{s,t} - x_{0,t}\right)^2}{\sum_{t=1}^{n}\left(x_{0,t} - \mu_0\right)^2} = 1 - \frac{MSE}{\sigma_0^2} \quad (5)$$

Where n is the total number of time steps, $x_{s,t}$ is the simulation value at time step t, $x_{0\,,t}$ is the observed value at time step t, and μ_0 and σ_0 are the mean and standard deviation of the observed values. The range number of NSE criterion can be from -∞ to 1. The prefect matched model discharge to the observed data will have E=1 for NSE optimization methods but E=0 means the model prediction are as accurate as the mean of the observation data, however E<0 occurs when the observation mean is better than the model predictor. By evidence of equations 4 and 5 it became clear that these two optimizations are close together but in this case of study we focused on NSE method and also the results of calibration and validation can be generalized with other method like Root Mean Square Error (RMSE) that close to MSE (Gupta et al. 2009). One of the most important characteristic of NSE methods is; covers comparison of both Scale and shape criteria of hydrographs.

2. Also some other statistical method for comparison of calibration and validation simulation was used that we mention bellow.

There are:

Relative Bias (R. bias) as a fraction of the total observed flow, (excluding any time steps with missing values). The best value for relative bias test is zero.

$$\frac{\Sigma(\mathrm{X}-Q)}{\Sigma Q} \tag{6}$$

Root **M**eans **S**quare **E**rror (RMSE) of an estimator $\hat{\theta}$ with respect to an estimated parameter θ is defined as the square root of the mean square error:

$$RMSE(\widehat{\theta}) = \sqrt{MSE}(\widehat{\theta}) \tag{7}$$

The root mean square error (RMSE) is a measure of the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated. Since the RMSE is a good measure of accuracy, it is ideal if it is small.

III.2.3: Regional Analysis

The availability of data is an important aspect in frequency analysis. The estimation of probability of occurrence of extreme flood is an extrapolation based on limited data. In practice, however, data may be limited or in some cases may not be available for a site. In such cases, regional analysis is most useful. The stream-flow predictions in ungauged catchments are a challenge for hydrologist around the word (Sivapalan et al., 2003). Generally predication in ungauged catchments are studied by regionalization approach i.e., transfer of model parameter from gauged catchments to ungauged (Bloschl and Sivapalan, 1995). Three kind of approached are widely used, regression, spatial proximity, and physical similarity (Oudin et al., 2008; Zhang and Chiew, 2009). In this case of study the regression based method was used for regionalization approach. In the regression-based approach the model parameters in ungauged catchments are tried to estimate by linear regression equation (eq. 2) with several catchments characteristics. The key step is to construct relationships between optimization model parameter and catchments characteristics (e.g., soil, vegetation, climate, geomorphology, and etc.) using regression equation. As it mentioned in section (III. 2.1) each parameter of GR4J model can have relationship with some characteristics of catchments that the best one needs to find out. For example Production store (X1) is storage of water on surface of soil that capacity of this storage links more with slope, geology, land use and etc. percolation and evapotranspiration also effect on value of storage. The procedure of regionalization analysis in this case of study is analyzing the classical hydrological model parameter that produced in calibration and validation hydrological model with geomorphology, land-use and geology watersheds characteristics to produce the linear regression model between classical model parameter and watersheds characteristics. The regionalization model parameter for five (5) selected watersheds was produced by using linear regression model. The regional hydrological simulation of discharge for five (5) selected watersheds with regionalized model parameter has done.

III.2.4: Hydrological Simulation and Comparison of Methods

Five watersheds from the study area were selected and treated as ungauged for independently testing the regionalization procedure. These five watersheds were selected to cover the most possible geomorphology, geology and land use characteristics of watershed in case of study that can effect on estimation parameter regionally. In these watersheds classical hydrological simulation with calibrated GR4J model and simulation with GR4J with regionally estimated parameters have been performed. The simulated hydrographs were compared to each other and with the observed hydrograph. Then, the annual maximum discharge time series were extracted from the classical simulation, regional simulation and observed time series and Flood Frequency Analysis (FFA) is performed. Finally, the results of FFA were compared and conclusions were extracted

III.2.5: Flood Frequency Analysis

Flood frequency analysis is used to estimate the magnitude or frequency of flooding for a watershed. Historical data is used in a statistical model to predict how often a flood of a given scale is likely to recur, or to predict the greatest flood likely within a given time period.

III.2.5.1: Extreme value theory

A) Generalized Extreme Value (GEV) Distribution

The probability density function of the GEV distribution is of the form:

$$f(x) = \frac{1}{\alpha}\left[1 - k\left(\frac{x-u}{\alpha}\right)\right]^{\frac{1}{k}-1} e^{-\left[1-k\left(\frac{x-u}{\alpha}\right)\right]^{1/k}} \quad (8)$$

The range of the variable x depends on the sign of the parameter k. When k is negative (Type II extreme value distribution) the variable x can take values in the rang ($u + \alpha/k < x < \infty$) which make it suitable for flood frequency analysis. However when the k is positive, (Type III extreme value distribution) the variable x becomes upper bonded and takes value in the range ($-\infty < x < u + \alpha / k$) which may not be acceptable for analyzing floods unless there is sufficient evidence that such an upper bound does exist. When $k = 0$ the GEV distribution reduces to the type I extreme value distribution discussed in next paragraph. The GEV distribution function is of the form (Jenkinson, 1955) in equation 11.

$$F(x) = \exp\left\{-\left[1 - k\left(\frac{x-u}{\alpha}\right)\right]^{\frac{1}{k}}\right\} \quad (9)$$

These models, along with the Generalized Extreme Value distribution, are widely used in risk management, finance, insurance, economics, hydrology, material sciences, telecommunications, and many other industries dealing with extreme events.

B) Extreme Value I (Gumbel) Distribution

One of the first scientists to apply the theory was a German mathematician Emil Gumbel (1891-1966). Gumbel's focus was primarily on applications of extreme value theory to engineering problems, in particular modeling of meteorological phenomena such as annual flood flows:

"It seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis."

The Gumbel distribution, also known as the Extreme Value Type (I) distribution, is unbounded (defined on the entire real axis), and has the following probability density function (Alves, et al.):

$$f(x) = \frac{1}{\sigma}\exp(-z - \exp(-z)) \quad (10)$$

Where z=(x-μ)/σ, μ is the location parameter, and σ is the distribution scale (σ>0).

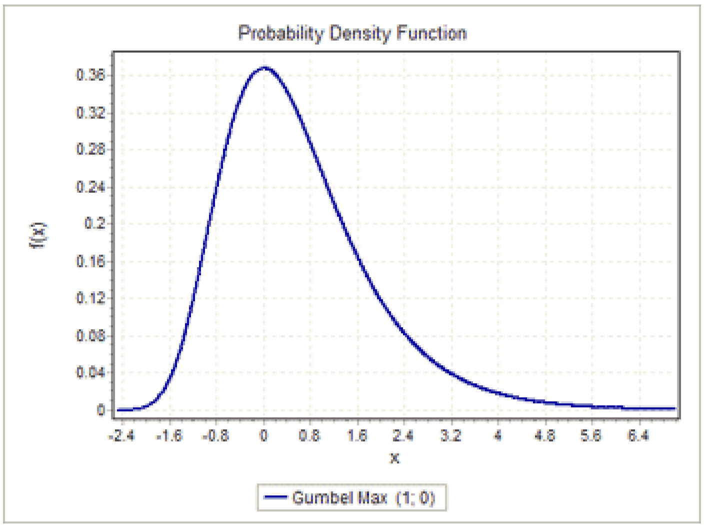

Figure 12: Gumbel PDF curve for σ=1 and μ=0 (Galiatsatou., 2010)

The shape of the Gumbel model does not depend on the distribution parameters: The graph above shows the Gumbel PDF for σ=1 and μ=0.

C) Fréchet Distribution

Maurice Fréchet (1878-1973) was a French mathematician who had identified one possible limit distribution for the largest order statistic in 1927. The Fréchet distribution, also known as the Extreme Value Type (II) distribution, is defined as

$$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1}\exp\left(\left(\frac{\beta}{x}\right)^{\alpha}\right) \quad (11)$$

Where α is the shape parameter (α>0), and β is the scale parameter (β>0). This distribution is bounded on the lower side (x>0) and has a heavy upper tail.

D) Weibull Distribution

Waloddi Weibull (1887-1979) was a Swedish engineer and scientist well-known for his work on strength of materials and fatigue analysis. The Weibull distribution, also known as the Extreme Value Type III distribution, first appeared in his papers in 1939. The two-parameter version of this distribution has the density function

$$\mathrm{f}\,(\mathrm{x}) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right) \tag{12}$$

The Weibull distribution is defined for x>0, and both distribution parameters (α - shape, β - scale) are positive. The two-parameter Weibull distribution can be generalized by adding the location (shift) parameter γ:

$$\mathrm{f}\,(\mathrm{x}) = \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right) \tag{13}$$

In this model, the location parameter γ can take on any real value, and the distribution is defined for x>γ. Even though the Weibull distribution was originally developed to address the problems arising in material sciences, it is widely used in many other areas thanks to its flexibility. When α=1, this distribution reduces to the Exponential model, and when α=2, it mimics the Rayleigh distribution which is mainly used in telecommunications. In addition, it resembles the Normal distribution when α=3.5:

It's worth noting that the Gumbel and Fréchet models described above relate to maxima (largest extreme value), while the Weibull model relates to minima (smallest extreme value). This form of the Weibull distribution is commonly used in practice.

III.2.5.2: Statistical estimation methods

Location, scale and shape are parameters of GEV extreme value distribution that need to compute. Different methods is available like,(Maximum likelihood, method of moments, last square method, weight moment, L-moment and other method that is not mentioned here). L-moment and Maximum likelihood are those methods that in this time they are the most used methods.They are used for computation of GEV and GUM (Gumble) extreme value distribution. Compare the results of two methods (GEV & GUM) and also comparison of results of parameter calculation by two methods of L-moment (LM) and Maximum likelihood (MLE) has a result to choose the better method of extreme value distribution.

A) Method of L-Moment

L-Moments is similar to the method of moments in that we will be solving a system of equations whose order is equal to the number of parameters we are trying to estimate. However, the set of L-Moments equations is instead defined as

$$\beta_r = E(XF(X)^r) \equiv \int_{-\infty}^{\infty} xF(x)^r f(x)dx \tag{14}$$

Where F(X) is the cumulative distribution function of the density function f(x), We will set this equal to an unbiased estimate of β_r, b_r, which is defined as

$$b_r = \frac{1}{n\binom{n-1}{r}} \sum_{i=1}^{n} \binom{i-1}{r} X_{(i)} \tag{15}$$

Where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the sorted values of the observations$X_1, X_2, \dots X_n$. Since we have two parameters in our case, we are concerned with the values b_0 and b_1. We note, however, that in the case of r = 0, the value of b_0 is equal to the observed mean $\overline{X}$. Therefore, the system we must solve to find the estimates of α and β is

$$\begin{bmatrix} \bar{X} = E(X) \\ \frac{1}{n(n-1)} \sum_{i=1}^{n} (i-1) X_{(i)} = E(XF(X)) \end{bmatrix} \tag{16}$$

B) Method of Maximum Likelihood

To begin with, consider the random sample of size n from the population we are looking at, which is presumed to be from the density function f(x) as defined earlier. We denote the sample of size n as

$$x_1, x_2, \dots, x_n.$$

Then we consider the joint density $f(x_1, \dots, x_n)$. If the sample is random, it is true that

$$f(x_1, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n) = \prod_{i=1}^{n} f(x_i) = L(\alpha, \beta). \tag{17}$$

This function L serves as the likelihood that the function has the proper parameters, so the objective of an MLE calculation is to estimate α and β by maximizing this function. However, while calculating this, the numbers in the intermediate steps can become large and computationally intensive from the product operator. Therefore, we instead maximize the natural log of the function, that is, $\ln(L(\alpha, \beta))$, which has the same effect since natural log is monotone increasing above 0. Using the MLE procedure, then, $f(x) = k(\alpha, \beta)(x-c)^{\alpha-1}(d-x)^{\beta-1}$ becomes

$$L(\alpha, \beta) = k(\alpha, \beta)^n \left(\prod_{i=1}^{n} (x_i - c)\right)^{\alpha-1} \left(\prod_{i=1}^{n} (d - x_i)\right)^{\beta-1} \tag{18}$$

Taking the natural log of this and applying rules of logarithms yields

$$\ln(L(\alpha,\beta)) = n\ln k(\alpha,\beta) + (\alpha-1)\left(\sum_{i=1}^{n}\ln(x_i - c)\right) + (\beta-1)\left(\sum_{i=1}^{n}\ln(d - x_i)\right) \quad (19)$$

It is this function that we will maximize to determine the correct α and β.

III.2.5.3: Plots

In statistics probability–probability plot is a probability plot for assessing how closely two data sets agree, which plots the two cumulative distribution functions against each other. Also the quntile plot is more widely used. Quantile plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. First, the set of intervals for the quantiles is chosen. A point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). Thus the line is a parametric curve with the parameter which is the (number of the) interval for the quantile. Quantile plot is used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different in the two distributions. Following the straight line shape is the easiest explanation of good results in these two plots.

IV. Results

After processing the basic database and produce the final data base, the regionalization procedure outlined above is performed. This procedure contains the classical hydrological simulation of the 26 study watersheds, analysis of the hydrological model parameters with the geomorphological parameters of 21 study watersheds and hydrological model parameters regionalization, hydrological simulation with regionalized model parameters at five (5) selected watersheds treated as ungauged, flood frequency analysis at the five (5) watersheds and comparison of the results.

IV.1 Calibration and validation of the conceptual model and discharge simulation

Split sample method for the calibration and validation of the GR4J model has been used. According to this method the database was divided into two different time series, one time series (TS1) data for the calibration of the model and one time series (TS2) for the validation of GR4J rainfall-runoff hydrological model. The GR4J rainfall-runoff model has been automatically calibrated using the Shuffled Complex Evolution (SCE-UA) optimization method and the Nash-Sutcliffe model Efficiency (NSE) has been used as the objective function for the optimization. This study has been performed in the R-Studio environment with Hydromad hydrology package which includes the GR4J rainfall-runoff model.

A) Calibration:

Model parameters have been estimated by optimization of the model parameters for each one of the 26 study watersheds. Some parameters touched the maximum range which may mean that the range of parameter value should be extended and calibration should perform again. The X2 parameter was the parameter that varies widely for most of the 26 study watersheds from 80% confidence interval of parameter (-5 to 3) to a new range (-20 to 5), but other parameters do not vary much. Table 3 presents the calibration results for the 26 study watersheds. This table contains the statistics of the model calibration which compare the observed and the simulated hydrographs, e.g. Nash-Sutcliffe model Efficiency (**NSE**), Relative Bias (**R. bias**), Root Mean Square Error (**RMSE**), the calibration period (**TS1**) and the values of the optimized model parameters. Table 4 compares the median value of the optimized model parameters with the 80% confidence interval of model parameter suggested by the developers of the model (Perrin et al, 2003).

Table 3: Calibration results of 26 selected watersheds

WATERSHEEDS	Calibration (SCE)				Model Parameter (SCE)			
	R. bias	RMSE	NSE	TS1	x2	x3	x4	x1
Crobes	-0.02	2.28	0.85	71-85	-0.17	84.10	1.36	373.61
Generargues	-0.10	2.55	0.81	71-90	-1.95	75.54	1.49	298.57
Gluiras	0.07	1.85	0.81	80-90	0.47	170.97	1.52	494.64
Les-Mages	-0.20	1.89	0.71	88-98	-7.08	68.97	1.33	381.05
Marsillargues	0.07	0.88	0.88	71-91	-4.00	62.26	3.11	190.31
Pont-de-lambeaume	0.02	2.88	0.83	80-95	5.00	220.47	1.14	461.50
La-roque-sur-ceze	0.00	1.92	0.79	71-95	0.16	77.20	2.04	132.00
Saint-Martin-d'Ardeche	-0.09	1.25	0.90	80-95	-1.96	141.99	1.51	234.95
Besseges	0.05	3.31	0.69	73-93	1.84	65.74	1.61	183.22
Vals-les-bains	-0.06	2.39	0.90	99-2004	-3.13	299.82	*1.36*	645.65
Langogne	-0.06	1.99	0.79	88-99	-10.66	230.61	1.73	699.56
Meyras	-0.06	2.91	0.85	88-99	-5.89	354.64	1.28	100.00
Saintbauzile	0.02	0.88	0.83	88-99	-10.60	446.39	1.46	279.20
Meyrueis	-0.05	3.25	0.73	88-99	-1.11	108.48	1.10	357.25
Bondons	-0.05	1.48	0.79	88-99	-1.17	89.66	1.52	973.49
Chateauneuf	0.00	0.84	0.62	88-99	-13.02	189.62	1.57	2255.38
Chastanier	0.03	0.94	0.62	88-99	-13.10	230.19	1.29	2502.54
Marvejols	-0.03	0.96	0.56	88-99	-11.99	148.81	1.95	1661.03
Pelouse	-0.01	1.25	0.55	88-99	-5.27	114.98	1.39	1927.55
Saint_etenne-vallee-francaise	-0.07	3.43	0.55	88-99	-1.38	101.50	2.77	2359.91
Meyrueis	-0.10	2.09	0.76	88-99	-7.13	132.93	1.15	474.63
langogne(Langouroux river)	-0.10	1.79	0.75	88-99	-9.99	387.83	1.35	137.55

Mende	-0.03	0.62	0.69	88-99	-5.49	68.44	1.48	2992.56
Florac	-0.01	1.75	0.92	88-99	-4.12	171.65	1.38	254.17
Bedouis	0.00	1.54	0.48	88-99	-0.01	90.82	2.40	1269.48
Florac (Tarnon river)	-0.04	2.55	0.82	88-99	-1.24	100.73	1.21	269.51

Table 4: Values of median model parameters and approximate 80% confidence

	Median value	80% confidence interval
X1	350	100-1200
X2	0	-5 to 3
X3	90	20-300
X4	1.7	1.1-2.9

B) Validation:

After the calibration of the model, the validation process for another time period has been applied by using the optimized model parameter values for each study watershed. The statistics of hydrological simulation for calibration and validation procedure are presented and compared in Table 5. These results indicate that, overall, the GR4J model performed reasonably well and it is capable to reproduce the observed hydrograph with accuracy, except for the watershed Bedouis.

Table 5: Calibration and validation results of RR model with three different methods of comparison

WATERSHEEDS	Calibration (SCE)				Validation (SCE)			
	R. bias	RMSE	NSE	TS1	R. bias	RMSE	NSE	TS2
Crobes	-0.02	2.28	0.85	71-85	0.05	2.00	0.73	85-93
Generargues	-0.10	2.55	0.81	71-90	0.06	2.94	0.76	90-2007
Gluiras	0.07	1.85	0.81	80-90	0.18	2.04	0.79	90-99
Les-Mages	-0.20	1.89	0.71	88-98	-0.39	3.32	0.60	98-2005
Marsillargues	0.07	0.88	0.88	71-91	0.06	1.30	0.86	91-2007
Pont-de-lambeaume	0.02	2.88	0.83	80-95	0.06	3.01	0.85	95-2007
La-roque-sur-ceze	0.00	1.92	0.79	71-95	0.37	2.38	0.66	95-2007
Saint-Martin-d'Ardeche	-0.09	1.25	0.90	80-95	-0.05	1.32	0.92	95-2007
Besseges	0.05	3.31	0.69	73-93	0.24	3.04	0.70	93-2007
Vals-les-bains	-0.06	2.39	0.90	99-2004	-0.14	1.92	0.80	2004-2007
langogne	-0.06	1.99	0.79	88-99	-0.10	1.38	0.84	99-2007
Meyras	-0.06	2.91	0.85	88-99	0.05	1.78	0.91	99-2007
Saintbauzile	0.02	0.88	0.83	88-99	0.02	0.67	0.84	99-2007
Meyrueis	-0.05	3.25	0.73	88-99	0.01	3.48	0.54	99-2007
Bondons	-0.05	1.48	0.79	88-99	0.03	1.51	0.74	99-2007
Chateauneuf	0.00	0.84	0.62	88-99	-0.07	0.85	0.61	99-2007

Chastanier	0.03	0.94	0.62	88-99	0.02	0.75	0.69	99-2007
Marvejols	-0.03	0.96	0.56	88-99	-0.11	1.18	0.50	99-2007
Pelouse	-0.01	1.25	0.55	88-99	-0.05	1.14	0.57	99-2007
Saint_etenne-vallee-francaise	-0.07	3.43	0.55	88-99	0.51	1.84	0.73	99-2007
Meyrueis	-0.10	2.09	0.76	88-99	-0.03	2.18	0.54	99-2007
langogne(Langouroux river)	-0.10	1.79	0.75	88-99	-0.04	1.21	0.74	99-2007
Mende	-0.03	0.62	0.69	88-99	-0.08	0.60	0.64	99-2007
Florac	-0.01	1.75	0.92	88-99	0.01	1.79	0.90	99-2007
Bedouis	0.00	1.54	0.48	88-99	0.16	1.63	0.26	99-2007
Florac (Tarnon river)	-0.04	2.55	0.82	88-99	-0.01	2.18	0.85	99-2007

The values in Table 4 marked in red color represent the minimum and maximum values of NSE for calibration and validation periods. Also, the NSE took values larger than zero (NSE>0).

Figure 13 presents the hydrographs for two of the study watersheds for which NSE took the smallest and larger value. In this figure (Fig. 13), it is obvious the difference in the simulated and observed hydrographs in these two watersheds. Also Figure 14 presents the comparison of calibration and validation NSE values. The largest variation in NSE values

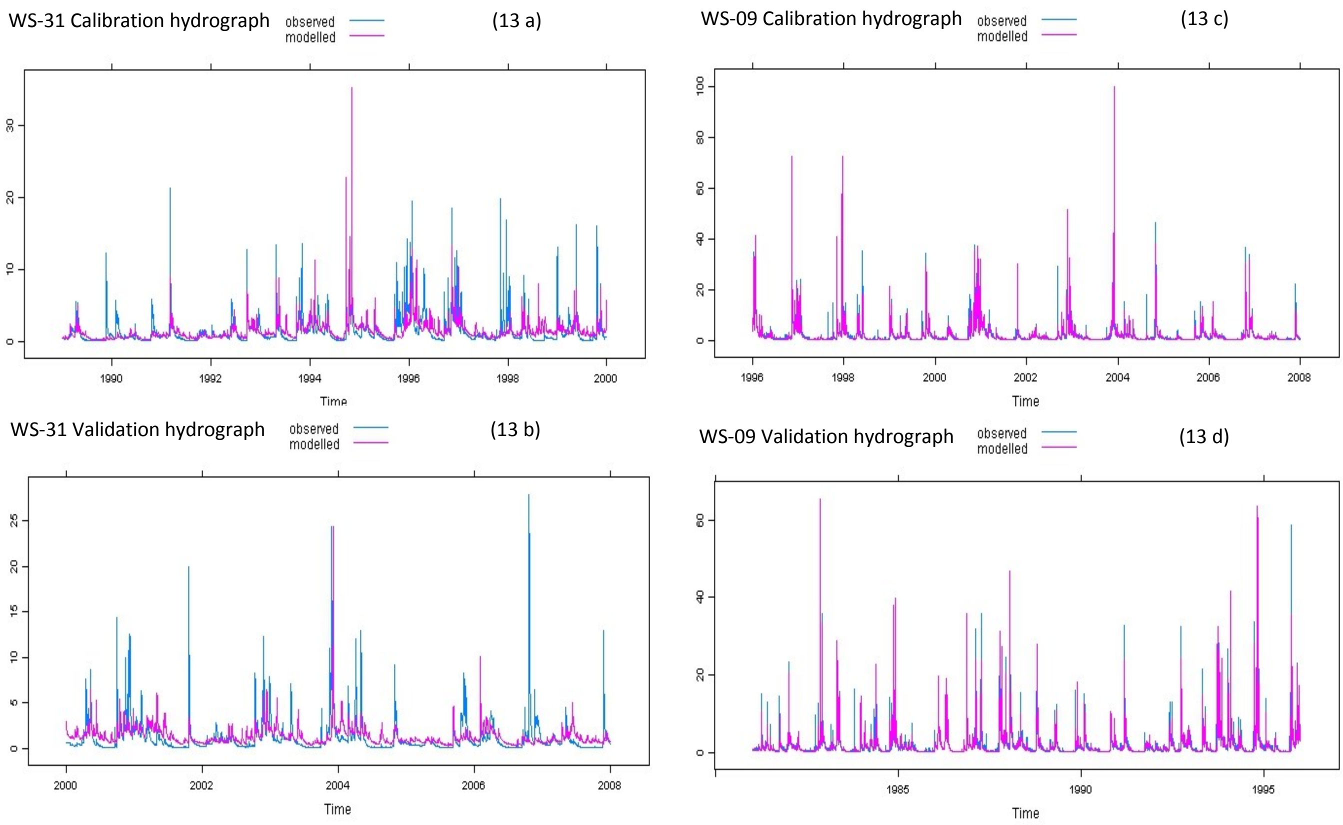

Figure 13: Hydrographs of watershed 31 (13a & 13b) with smallest NSE test values and hydrograph of watershed 9 (13c & 13d) with largest NSE test values

between calibration and validation has been found for the Bedouis watershed (watershed 31) for which the NSE statistic took its smallest value for both calibration and validation.

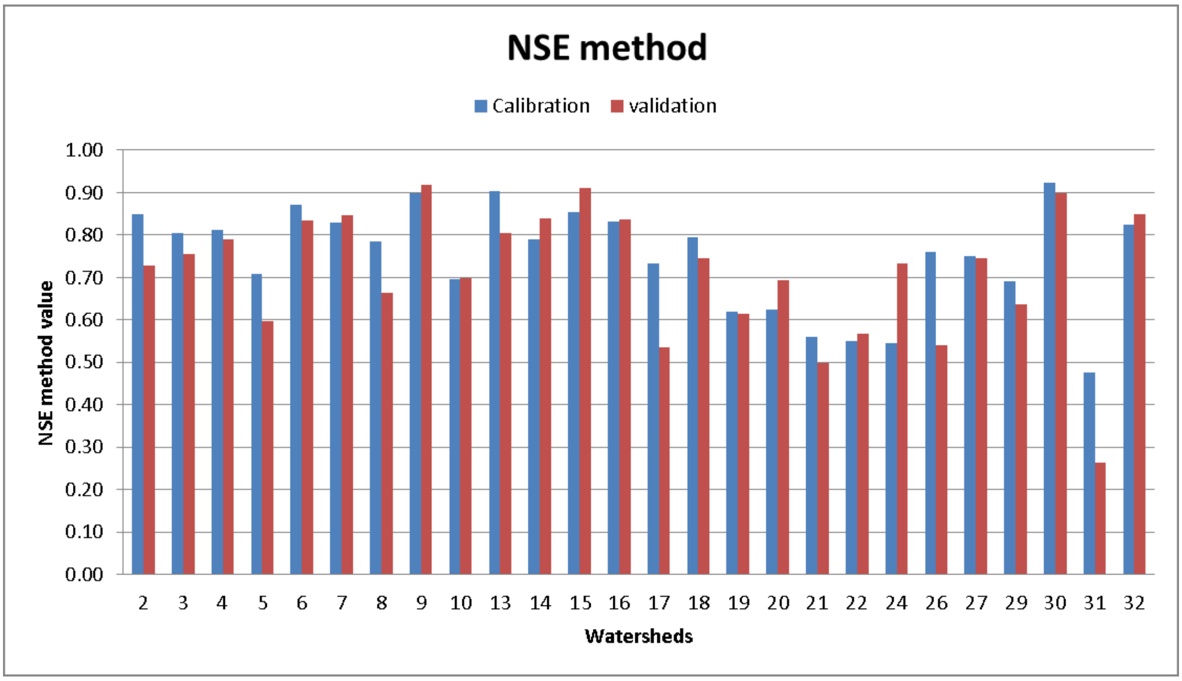

Figure 14: NSE objective function results values for all study's watersheds

IV.2 Analysis and regionalization of the model parameters

The GR4J model is a rainfall-runoff model which has been tested in more than 400 different catchments (Perrin et al, 2003) and thus, it is assumed to be appropriate for the hydrological simulation of the 26 study watersheds. The values of the four model parameters were optimized through the automatic calibration procedure. However, the classical simulation by using calibration-validation procedure is feasible only for gauged watersheds. For ungauged watersheds, within a hydrologically homogeneous region, this classical application is not applicable. In such a case, the model parameters, found through the classical application of the hydrological model in gauged watersheds, are then regionalized by analyzing the model parameter variation with the geomorphological, geological, climatic and land use characteristics of the gauged watersheds and, finally, the hydrological model is applied to ungauged watersheds using the regionally estimated model parameters.

In this part of the study the variation of the model parameter with the characteristics of catchments has been analyzed. Figures 13 to 17 present the best results for the variation of each GR4J model parameter with watershed characteristics. On these figures the linear regression line is also noted and its equation is used for the estimation of the regional value of model parameters for the five (5) selected watersheds treated as ungauged. On these figures (Figs. 13-17) the red line represents the median value of the parameters presented in Table 3. The results of hydrological model regionalization are presented in the next paragraphs.

a) (X1) Production store (mm): The mean watershed slope, the land use, the geology, the area, and the mean annual precipitation of the watershed are the geomorphological and climate model parameters that may affect the value of parameter X1. An analysis has been performed for all the above parameters. This analysis indicated that the mean slope of the watersheds explains better the variation of parameter X1 (Figure 15).

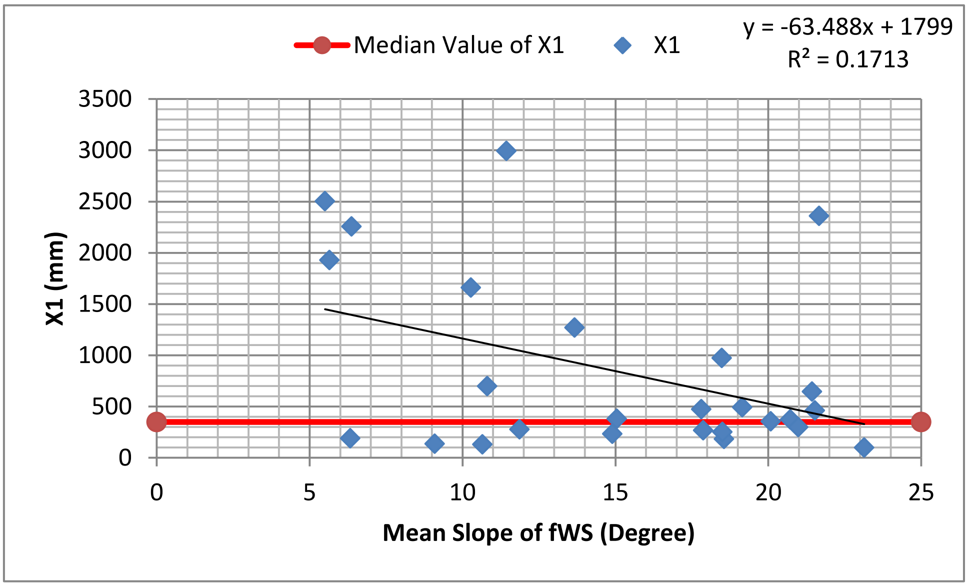

Figure 15: Linear relation of Mean slop of watersheds and X1 model parameter and Median Value of X1

b) (X2) Groundwater exchange coefficient (mm): This parameter take positive and negative values, which means that water is coming to or going out from the routing store (i.e surface water storage), respectively. Mean slope, geological parameter (i.e permeability of watershed rocks), mean elevation, and mean annual precipitation were those characteristics used for linear model comparison. Watershed mean slope better explain the variation of model parameter X2 (Figure 16).

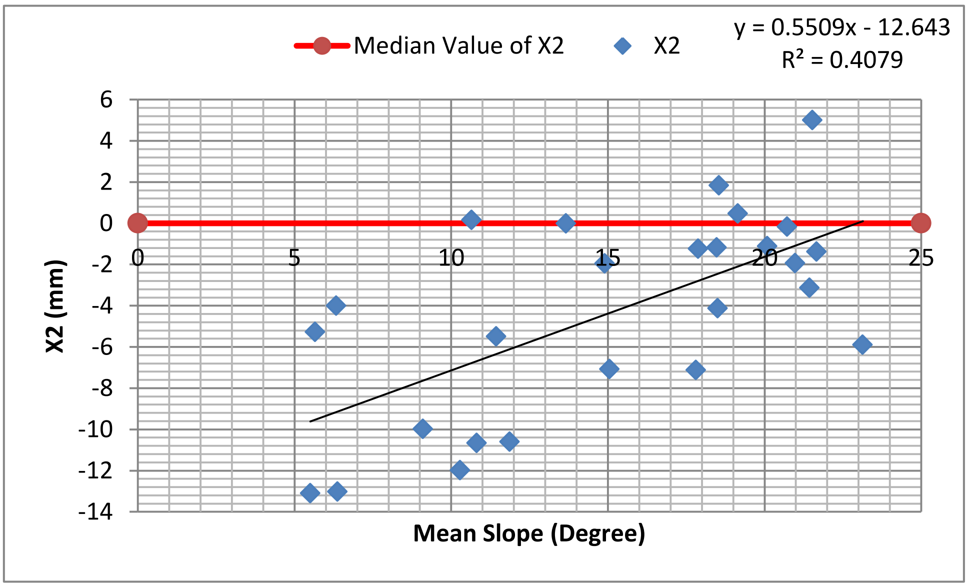

Figure 16: Linear relation of Mean slop of watersheds and X2 model parameter and median value of X2

c) (X3) Maximum capacity of routing store (mm): This parameter depends to the condition of soil porosity and humidity of soil that how much water can be kept in the soil. Mean slope, mean elevation and geological parameters (i.e permeability of watershed rocks) were the watersheds characteristics used in the analysis. Watershed mean elevation was found to better explain the variation of model parameter X3 (Fig. 17).

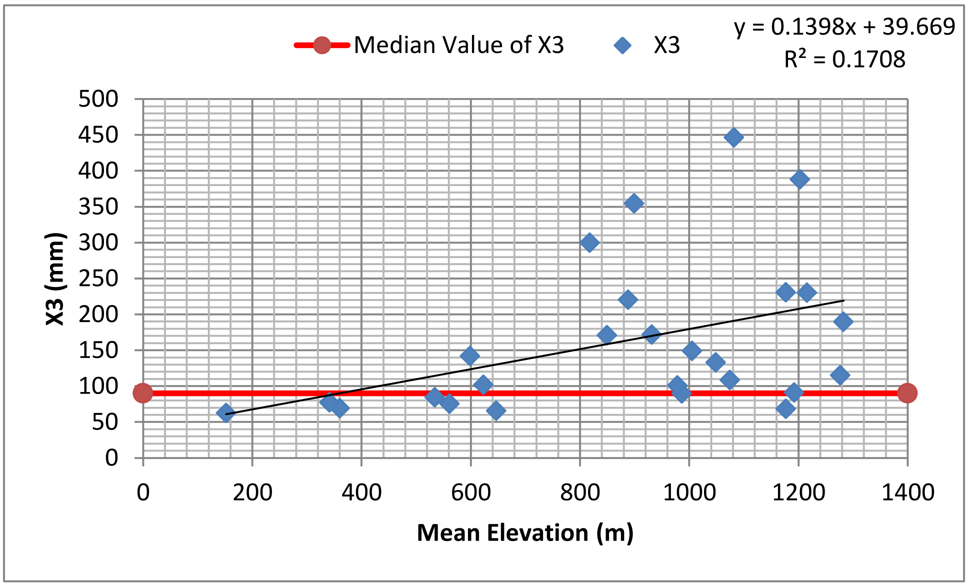

Figure 17: Linear relation of Mean elevation of watersheds and X3 model parameter, and median value of X3

d) (X4) time peak ordinate of hydrograph unit UH1 (Day): The watershed parameters watershed area, mean slope, mean elevation, mean annual precipitation and land use were used in the analysis. Mean annual precipitation was found to better explain the variation of model parameter X4 (Figure 18).

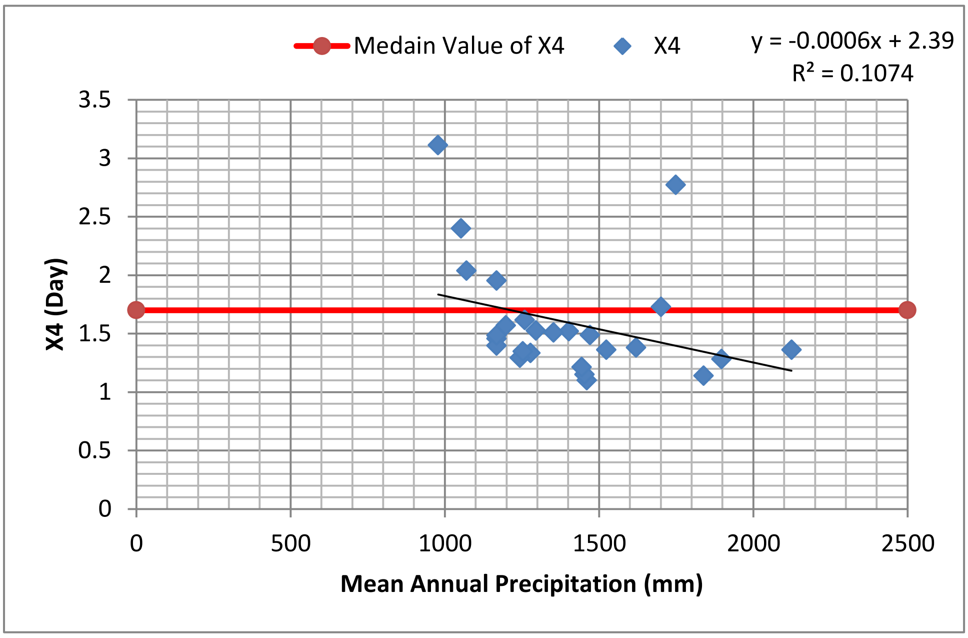

Figure 18: Linear relation of Mean annual precipitation of watersheds and X4 model parameter, and Median value of X4

Table 6: Regional estimated parameters for GR4J model of 5 watersheds

Regional estimated model parameter					
Model Parameters	Les Mages	Marsillargues	pont-de-lambeaume	Besseges	Bondons
X1	843.17	1396.28	432.10	621.11	625.63
X2	-4.75	-5.92	-2.56	-4.82	-4.68
X3	93.30	65.19	164.46	131.95	177.81
X4	1.68	1.83	1.40	1.69	1.67

After the above analysis, the five selected watersheds were treated as ungauged and the hydrological model parameters were estimated using the linear relationships developed in the regional analysis. The values of the model parameters for these five watersheds are presented in Table 6.

IV.3 Hydrological simulation using the regionalized model parameters

The regionally estimated GR4J parameters for the five (5) selected watersheds were used for the regional simulation for ungauged watersheds. The resulting hydrographs were, then, statistically and graphically compared to the simulation results using the optimized GR4J model parameters (classical simulation) and the observed hydrographs. This comprehensive comparison tests the validity of the regional methodology for ungauged watersheds.

Table 7: Regional and validation simulation methods result

WATERSHEEDS	Regional			Validation		
	R. bias	RMSE	NSE	R. bias	RMSE	NSE
05_Les Mages	-0.33	3.74	0.49	-0.39	3.32	0.60
06_Marsillargues	-0.39	2.87	0.30	0.06	1.30	0.86
07_pont-de-lambeaume	-0.31	3.44	0.80	0.06	3.01	0.85
10_Besseges	-0.36	3.55	0.59	0.24	3.04	0.70
18_bondons	-0.11	1.68	0.68	0.03	1.51	0.74

Table 7 presents the statistics of the comparison of two different discharge simulations with the observed discharge. The statistics NSE, RMSE and R.bias are used. Regional simulation had less accuracy than the classical simulation for all five study watersheds, but this is expected since the regional simulation uses regional estimated hydrological model parameters whereas the classical simulation uses optimized model parameters for each particular watershed. However, the regional simulation results are acceptable and only in the Marsillargues watershed the simulation is poor.

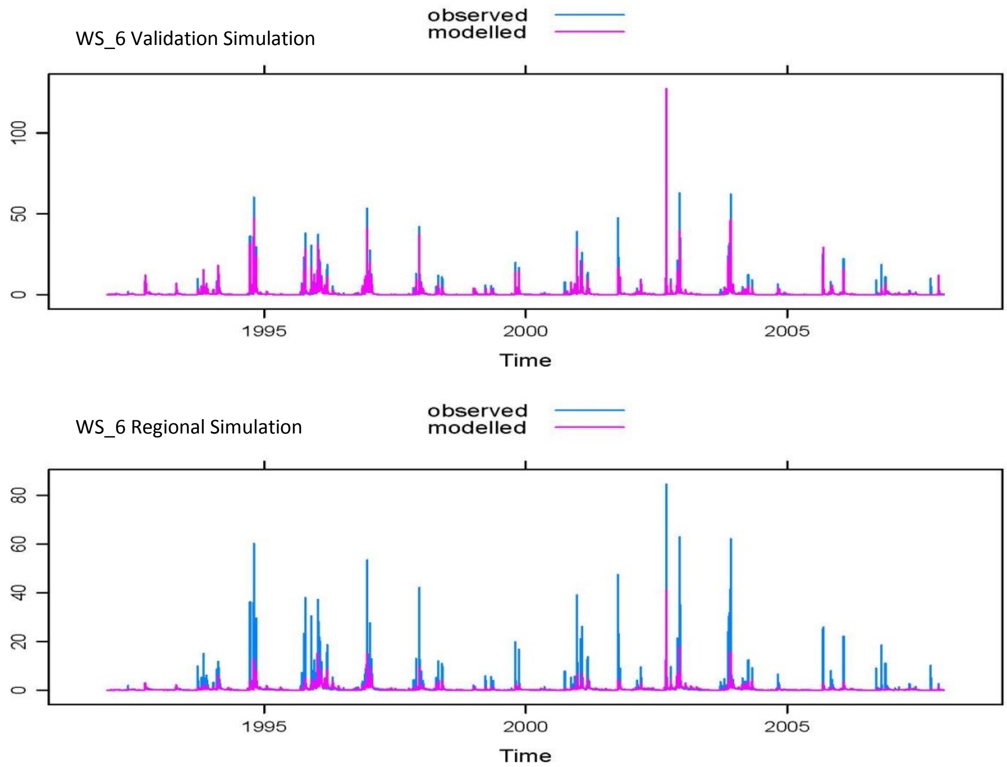

Figure 19: Comparison of classical and regional discharge simulation with observed discharge for Marsillargues watershed (watershed number 6)

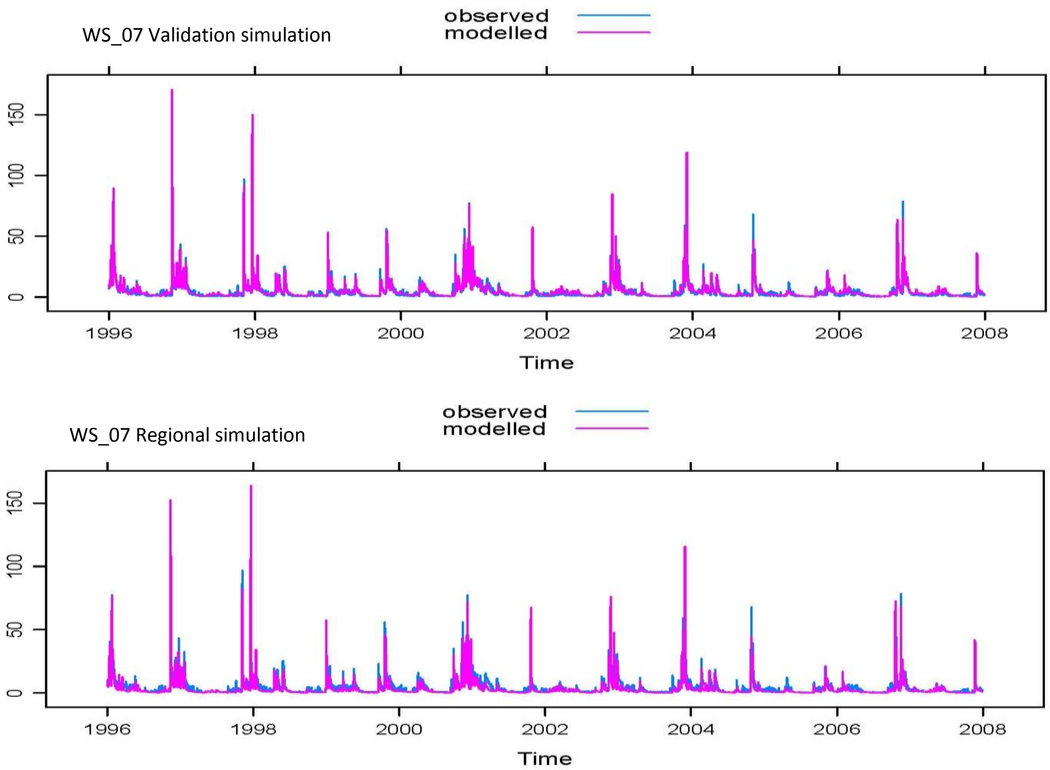

Figure 20: Comparison of classical and regional discharge simulation with observed discharge of Pont de Lambeaume watershed (Watershed number 7).

Figures 19 and 20 present the comparison of the classical and the regional simulation results with the observed discharge for the two watersheds having the worst (Marsillargues watershed) and the best results (Pont de Lambeaume watershed) of the five study watersheds of regionalization method to simulate the discharge in ungauged watersheds.

IV.4 Flood frequency analysis

The final part of this study is flood frequency analysis. Annual maxima flow time series for the two simulated (regional and classical) time series and the observed time series for the five selected study watersheds have been developed. The annual maxima flow time series were manually selected from the simulated and observed hydrographs and used for the flood frequency analysis.

IV.4.1: Flood Frequency Distribution methods

Flood frequency analysis was done by developing the empirical flood frequency distributions and fitting theoretical extreme value frequency distributions. This work has been done in the R-Studio software. The Generalized Extreme Value (GEV) and the Extreme Value I or Gumbel (EVI or GUM) are two theoretical extreme value frequency distributions used in the flood frequency analysis. The Maximum Likelihood (ML) and L-moment (LM) methods were used for the fitting of theoretical extreme value frequency distributions to the empirical flood frequency distributions and theoretical extreme value frequency distributions parameters estimation. Then, the Kolmogorov-Smirnov (KS) test was applied to test the goodness-of-fit of theoretical distributions to the empirical ones.

Table 8: The Kolmogorov-Smirnov test results of two distribution model and two methods of parameter estimator for 5 selected watersheds

Watersheds	KS Test (D-Value)		KS Test (D-Value)		
	GUM_ML	GUM_LM	GEV_ML	GEV_LM	D-Critical
OB_5	0.2328	0.2354	0.2043	**0.203**	0.46799
REG_5	0.3253	0.3084	0.3425	**0.2943**	
CA_5	0.3109	0.2893	na	**0.2665**	
OB_6	0.1473	0.1402	0.1474	**0.1211**	0.29472
REG_6	0.2509	0.2133	**0.1751**	0.2967	
CA_6	**0.1063**	0.1523	0.1092	0.1755	
OB_7	0.1521	**0.1445**	0.1719	0.2124	0.33815
REG_7	0.1839	0.1908	0.1675	**0.1646**	
CA_7	0.1586	0.1354	**0.1214**	0.1849	
OB_10	0.1135	0.1189	**0.0934**	0.1289	0.31417
REG_10	0.1404	0.141	**0.1203**	0.2042	
CA_10	0.1866	0.1838	0.1907	**0.1823**	
OB_18	0.3228	0.3149	0.2588	**0.25**	0.40962
REG_18	**0.2244**	0.2586	0.2409	0.3019	
CA_18	**0.186**	0.1886	0.1875	0.2155	

Table 8 presents the results of the fitting of the theoretical to the empirical flood frequency for the two theoretical flood frequency distributions (i.e. GEV and GUM) and the two methods of fitting (i.e. ML and LM) for the five study watersheds.

The Kolmogorov-Smirnov (KS) test is a goodness-of-fit test to find out the better distribution method. KS test represent the maximum distance between the empirical distribution of the sample and the cumulative distribution of the reference distribution. The KS test value should be smaller than the critical value of the KS statistic at a significance level, usually taken at α=5%, to represent an acceptably fitted frequency distribution. The best fitted distribution is the one with the smaller KS test value. The results of KS test shows that, overall, the GEV distribution is better fitted to the empirical distribution than the Gumbel distribution for all study watersheds for the observed and the regionally simulated flood frequency distributions (Table 8).

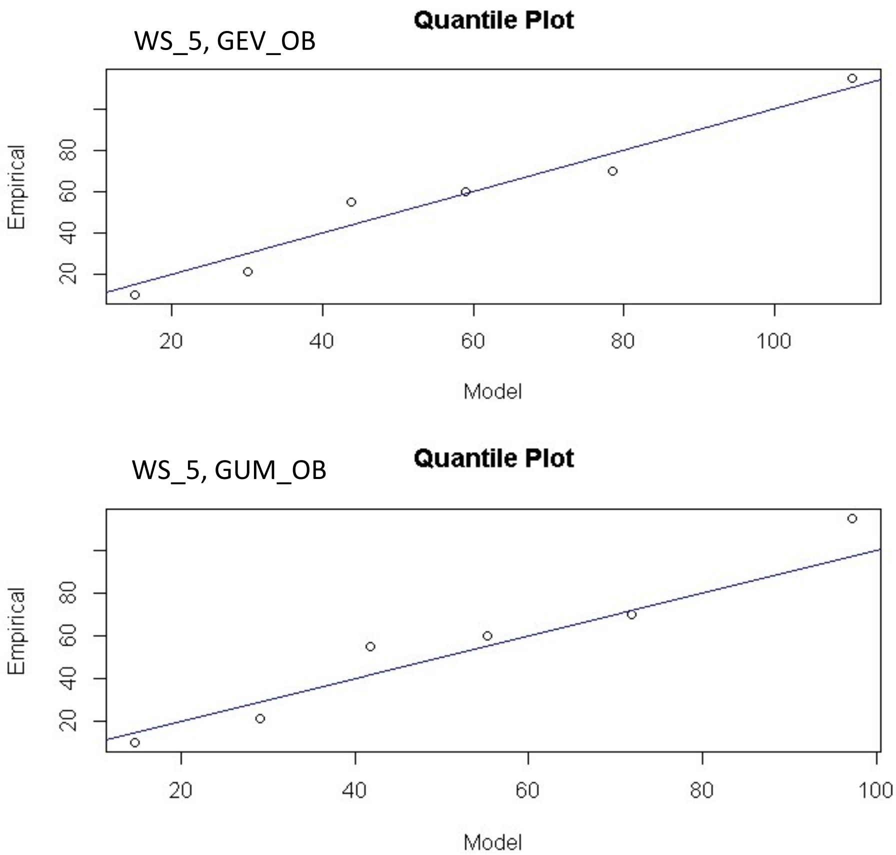

Figure 21: Quantile plots for GEV and GUM distribution models for Les Mages watersheds.

The quantile and probability plots are another way to examine the fitted model distribution. If the GEV is a reasonable distribution for modeling the maxima, the quantile plot should be approximately linear; this is the simple definition of usage of quantile plot. Data modeled with more linearity is reasonable to accept the model. Figure 21 presents the quantile plot for observed annual maxima discharge with GEV and GUM distribution model and indicates that the GEV distribution is better fitted to the observed annual maxima from the GUM distribution for Les Mages watershed.

Any departures from linear distribution are indicative of failure of fitting of the theoretical distribution to the empirical frequency distribution. As an example, Figure 22 shows that the results of the GUM model for the Pont-de-Lambeaume watershed deviate from the line. On the other hand, the results for the GEV and the same watershed indicate better fitting.

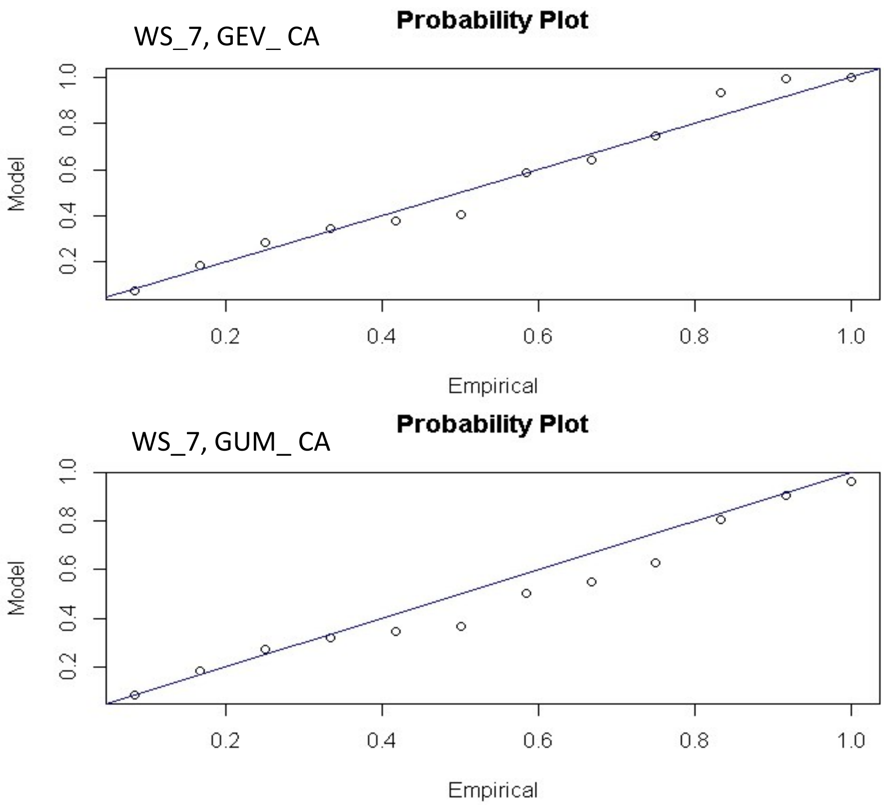

Figure 22: Probability plots for GEV and GUM distribution models for pont-de-lambeaume watersheds.

IV.4.2 Methods for estimation of the distribution model parameters

GEV distribution model for flood frequency analysis of annual maxima discharge was selected, as the best fitted overall model to the empirical frequency distributions for all-time series and watersheds. The Maximum Likelihood (ML) and L-moment (LM) methods were used for the estimation of the fitted GEV parameters.

Figure 23 presents the quantile plot of watershed 6, for observed discharge annual maxima, in which LM method shows more linear distribution than the ML method. The plotted results of ML and LM methods as they are presented in Figures 23 and 24 were not clear enough for the selection of the best method. Therefore, by using the results of KS test presented in Table 7, the LM method is selected as the best overall method for the estimation of the GEV distribution parameters.

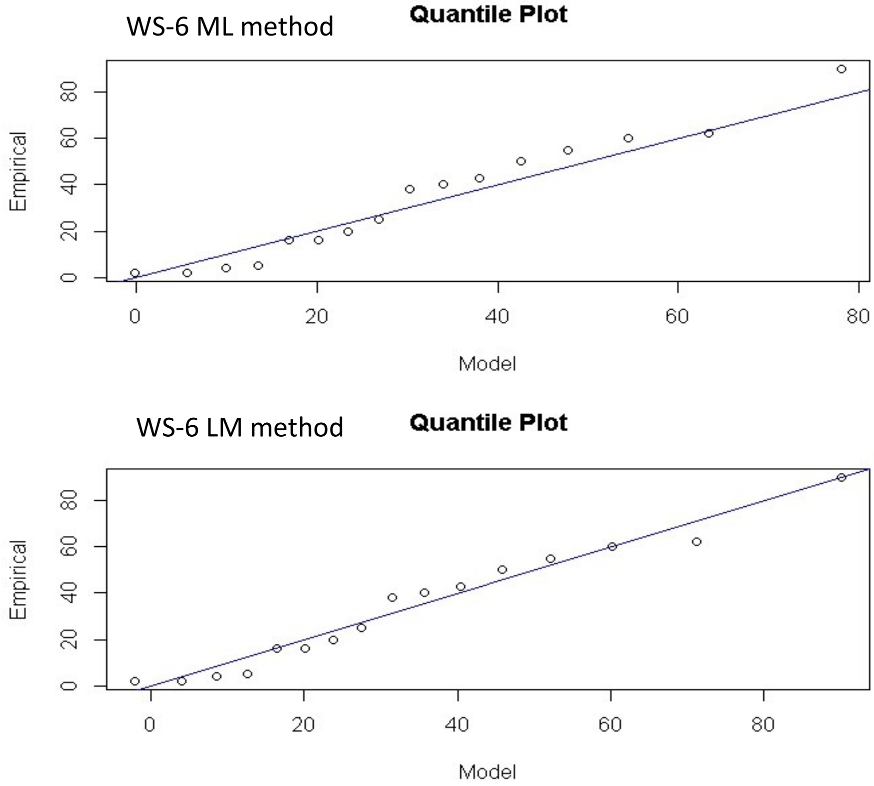

Figure 23: Quantile plots of GEV distribution with estimated parameters with the Maximum Likelihood (ML) and L-moments (LM) methods for Marsillargues watershed.

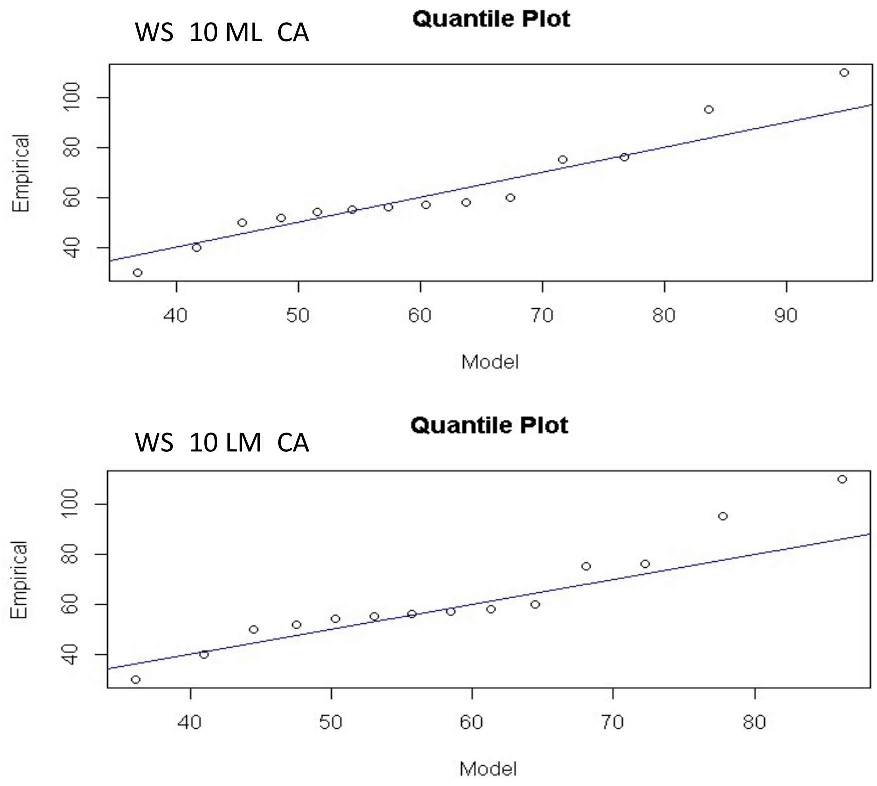

Figure 24: Quantile plot of GEV distribution with estimated parameters using ML and LM methods for Besseges watershed

IV.4.3 Comparison of Flood frequency results for simulations and observed discharge

The Flood Frequency Analysis (FFA) for the two simulated discharges (regional and classical) with the FFA of the observed discharges for the five selected watersheds has been compared. Overall, the classical hydrological simulation results for the annual maxima discharge in all five study watersheds represent better simulation than the regional simulation results of discharge, which is expected since the model parameters are optimized. On the other hand, the regional simulation results vary in the five study watersheds. Figure 25 presents the quantile plots of the FFA with the fitted GEV to the observed and the two simulated time series of annual maxima (regional and classical) for Pont-de-Lambeaume watershed. The first quantile plot (WS-7 REG) indicates that the GEV fitted better the regional simulation results than for the observed discharges. However, the fitted GEV shows better fitting for the classical simulation results.

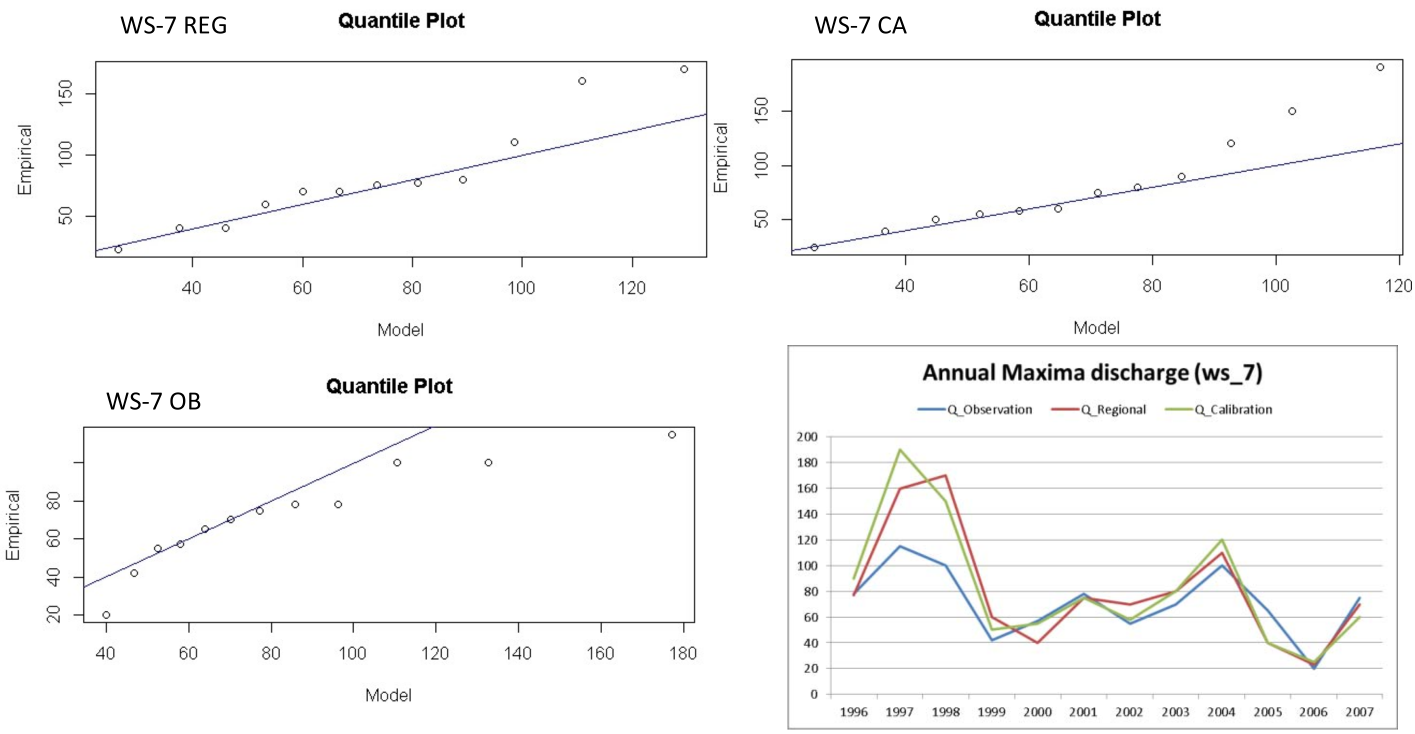

Figure 25: Quantile plot of regional simulation (up left), Classical simulation (up right) and observed (down left) discharge data and graphical annual maxima discharge data for these plots (down right) in watershed 7.

The quantile plot of the FFA with fitted GEV to the observed and two simulation time series of annual maxima (regional and classical) for Les Mages watershed represented in figure 26. The quantile plot of the FFA for observed (WS-5 OB) annual maxima with fitted GEV in this figure (figure 26) illustrated the better fitted results than the two different simulations (regional and classical). The variety exists in the final result for observed and simulations (regional and classical) that could represent some errors in process of case study (figure 27). The results of FFA with fitted GEV for observed and two simulation (regional and classical) indicated better fitted result for observed annual maxima than regional and classical simulation, although the classical simulation had better FFA result than the regional simulation.

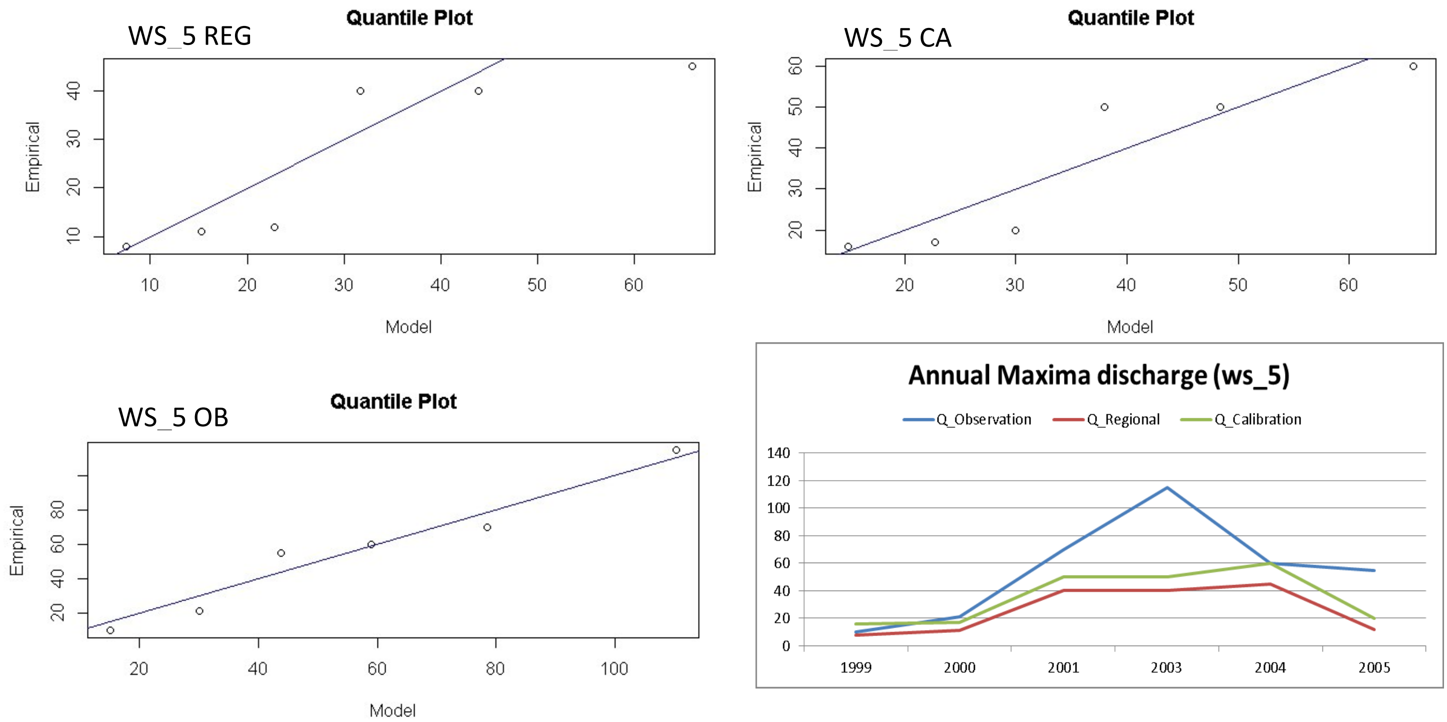

Figure 26: Quantile plot of regional simulation (up left), Classical simulation (up right) and observed (down left) discharge data and graphical annual maxima discharge data for these plots (down right) in watershed 5.

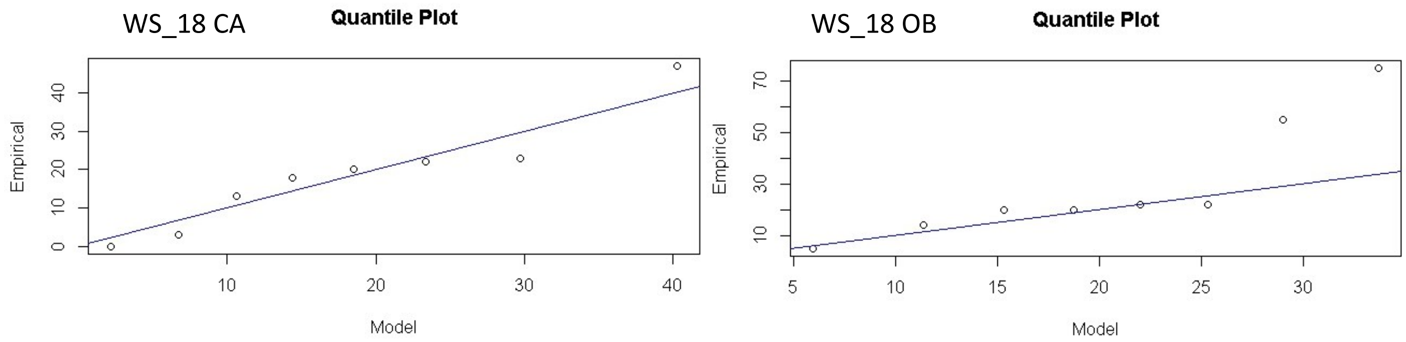

Figure 27: Quantile plot of Classic simulation (left) and observed discharge data (right) for watershed 18.

V. Conclusion and Discussion

Flood frequency analysis through the hydrology simulation and regionalization simulation for Cevennes area (southern France) proposed. This analysis applied to comprehensive more validity of regionalization method for parameter estimation on those watersheds that the observed data can be limitation of flood frequency analysis.

The GR4J hydrological model parameters estimated by optimization of the model parameter for each of the 26 study watersheds. The more variety of X2 model parameter was more than the other parameter and it was the reasons to extend the range of 80% watershed confidence interval of parameter (-5 to 3) to a new range (-20 to 5) and the calibration has done again. The comparison of the statics of model calibration for observed and simulated hydrograph showed the legible result. These results indicate that, overall, the GR4J model performed reasonably well and it is capable to reproduce the observed hydrograph with accuracy, except for the watershed Bedouis. The values of the four model parameters were optimized through the automatic calibration procedure. In regionalization analysis part of the study the variation of the model parameter with the characteristics of catchments has been analyzed. Regionalization analysis indicated that the mean slope, mean elevation and mean annual precipitation of the watersheds explain better the variation of GR4J model parameter. After the regionalization analysis, the five selected watersheds were treated as ungauged and the hydrological model parameters were estimated using the linear relationships developed in the regional analysis. The regionally estimated GR4J parameters for the five (5) selected watersheds were used for the regional simulation discharge for ungauged watersheds. The resulting hydrographs were, then, statistically and graphically compared to the simulation results using the optimized GR4J model parameters (classical simulation) and the observed hydrographs. This comprehensive comparison tests the validity of the regional methodology for ungauged watersheds. Regional simulation had less accuracy than the classical simulation for all five study watersheds, but this is expected since the classical simulation uses optimized model parameters for each particular watershed. However, the regional simulation results are acceptable and only in the Marsillargues watershed the simulation was poor.

The Generalized Extreme Value (GEV) theoretical extreme value frequency distribution and L-moment (LM) methods were used for the Flood frequency analysis that the Kolmogorov-Smirnov (KS) test was applied to test the goodness-of-fit of theoretical distributions to the empirical ones. The Flood Frequency Analysis (FFA) for the two simulated discharges (regional and classical) with the FFA of the observed discharges for the five selected watersheds has been compared. For selected watersheds, which is expected since the model parameters are optimized, the classical hydrological simulation results for the annual maxima discharge represented better simulation of discharge than the regional simulation results of discharge. The result of regional simulation was vary in five selected watersheds and the results of FFA with fitted GEV for observed and two simulation (regional and classical) indicated better fitted result for observed annual maxima than regional and classical simulation, although the classical simulation had better FFA result than the regional simulation.

Because only 26 watersheds were used in this study, there was some uncertainty in the comparison of different regionalization approaches. The linear relation of watersheds characteristics and model parameters, accuracy of annual maximum discharge, and length of time series of basic data need more precise study in the future.

VI. Reference

1. Beven, K. J. (2002). Towards an alternative blueprint for a physical based digitally simulated hydrologic response modelling syste. Hydrological Processes, 16, pp. 189-206
2. Blöschl, G., Sivapalan, M., (1995). Scale issues in hydrological modeling – a review. Hydrol. Process. 9, 251–290.
3. Chow, V. T. Maidment, D. R. and Mays, L. W. (1988). Applied Hydrology McGraw-Hill, New York, NY.
4. Cunnane, C. (1989). Statistical Distribution for Flood Frequency Analysis. Word Meteorological Organization Operational Hydrology, Report No. 33, WMO-No. 718, Geneva, Switzerland.
5. Duan, Q. Sorooshian, S. and Gupta V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol., 158, pp. 465-284
6. Duan, Q., Sorooshian, S. and Gupta V. K. (1992). Effective and efficient global optimization for conceptual Rainfall-Runoff Models. Water Resour. Res., 28(4), pp. 1015-1031
7. Freeze, R. A. and Harlan R. L. (1969). Blueprint for a physically-based digitally simulated hydrologic response model. J. Hydrol., 9, pp. 2161-2173
8. Galiasstou, P., (2010). Statistical Frequency Analysis, Flood Lecture of HydroHasards Master, Volos, Greece.
9. Gan T. Y. and Burges, S. J. (1990). An assessment of a conceptual rainfall-runoff model's ability to represent the dynamics of small hypothetical catchments 2. Hydrologic responses for normal and extreme rainfall. Water Resour. Res., 26(7), pp. 1605-1619
10. Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377, pp. 80-91
11. Harlan, D., Wangsadipura, M., and Munaja, C. M. (2010). Rainfall-Runoff Modeling of Citarum Hulu River Basin by Using GR4J. Proceedings of the World Congress on Engineering 2010 Vol II, WCE 2010, June 30 - July 2, 2010, London, U.K.
12. Jenkinson, A. F., (1955). The Frequency Distribution of the Annual Maximum (or Minimum) values of Meteorological Elements. Quarterly Journal of Royal Meteorological Society, 87, 158.
13. Kite, G. W. (1977). Frequency and Risk Analysis in Hydrology , Water Res. Publications, Fort Collins, Co.
14. Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I – A discussion of principles. J. Hydrol., 10, pp. 282-290
15. Oudin, L., Andréassian, V., Perrin, C., Michel, C., Moine, N.L., (2008). Spatial proximity, physical similarity, regression and ungaged catchments: a comparison of regionalization approaches based on 913 French catchments. Water Resour. Res. 44, W03413. http://dx.doi.org/10.1029/2007WR006240.
16. Oudina, L., Hervieua, F., Michela, C., Perrina, C., Andre´assiana, V., Anctilb, F., Loumagne, C. (2005). Which Potential Evapotranspiration Input for a Lumped Rainfall–Runoff Model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of Hydrology 303 (2005) 290–306.
17. Perrin, C. Michel, C. Andre´assian, V. (2003), Improvement of a Parsimonious Model for Streamflow Simulation, Journal of Hydrology 279 (2003) 275–289.

18. Sefton, C.E. M., and Howarth, S. M (1988). Relationship between dynamic response characteristics and physical descriptors for catchments in England and Wales. J. Hydrol., 211, pp. 1-16.
19. Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E., 2003. IAHS decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. Hydrol. Sci. J. 48 (6), 857–880.
20. Todini, E. (1988). Rainfall runoff modelling-Past, Present and future. J. Hydrol., 100, pp. 341-352.
21. Wheater, H. S., Jakeman, A. J. and Beven, K. J. (1993). Progress and direction in rainfall runoff modelling. In Jakeman, A.J., Beck, M.B and McAleer, M.J. (eds.) Modelling change in environmental systems. John Wiley & Sons, pp. 101-132.
22. Yapo, P. O., Gupta, H. V. and Sorooshian, S. (1966). Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. J. Hydrol., 181, pp. 23-48
23. Zhang, Y., Chiew, H.S., (2009). Relative merits of different methods for runoff predictions in ungauged catchments. Water Resour. Res. 45, W07412. http://dx.doi.org/10.1029/2008WR007504.