Universität für Bodenkultur Wien
University of Natural Resources and Life Sciences, Vienna

Department of Water, Atmosphere and Environment
Institute for Hydrology and Water Management (HyWa)

Estimation of return periods of discharge for the Sio River, Western Kenya

810301 Structural exercises

submitted by:

Benet Bayo, J. A. M.N.: 11726218

Madrigal Sanchez, A. M. M.N.: 11744947

supervised by:

Dipl.-Ing., Dr.nat.techn. Herrnegger, Mathew

16.4.2021

Index

1 Introduction 5
2 Study area and data basis 6
2.1 Study area 6
2.2 Data basis 6
2.2.1 Data availability 9
3 Methods 12
3.1 Temporal interpolation (Data filling) 12
3.2 Spatial interpolation 14
3.3 Hydrological Modelling 15
3.3.1 GR4J 16
3.4 Estimation of Extreme values 21
3.4.1 Generalized Extreme Value (GEV) distribution 22
3.4.2 Parameter estimation 23
3.4.3 Extreme value data set selection 24
3.4.4 Return period and return level 24
4 Results 27
4.1 Temporal interpolation (Data filling) 27
4.1.1 Precipitation 27
4.1.2 Pan Evapotranspiration data 41
4.2 Spatial interpolation 42
4.2.1 Basin precipitation 43
4.3 Hydrological Modelling 44
4.3.1 Input Data 44
4.3.2 Model set Up 46
4.3.3 Calibration results 47
4.3.4 Model performance / discussion 52
4.4 Estimation of Extreme Values of discharge 53
4.4.1 Return period from observed flow 53
4.4.2 Return period from simulated flow with KGE 54
4.4.3 Return period from simulated flow with NSE 55
4.4.4 Discussion of extreme value estimation 56
5 Summary and conclusions 60
6 Bibliography 61
7 Appendix 63
7.1 Long term mean precipitation values per station 63
7.2 GR4J modelling results with Chirps 88

List of Figures

Figure 1. Study area of the Sio River Basin, including precipitation and discharge stations... 6
Figure 2. The Sio River and the Busia-Kisumu road bridge seen from upstream, where the discharge measurement station “1AH01” is located. ... 7
Figure 3. Discharge measurement station “1AH01” at the Sio ... 7
Figure 4. Construction of Thiessen polygon (UTHH 2011) ... 14
Figure 5. Classification of hydrological models proposed by Singh (1995). Source: (Saha and Zeleke 2015, p. 568). ... 16
Figure 6. Schematic diagram of the GR4J model. Source: (eWater Source 4.1.0) ... 18
Figure 7. Extreme values in a normal distribution function (Glen 2016) ... 22
Figure 8. Extreme value distribution functions (Gilleland and Katz 2016) ... 23
Figure 9. Return period and return level representation (Gilleland and Katz 2016). ... 25
Figure 10. Partial series for the Styx River at Jeogla showing model fit and the 90% confidence interval(Institution of Engineers 1987). ... 26
Figure 11. Mean monthly precipitation ... 30
Figure 12. Mean annual precipitation f ... 35
Figure 13. Annual precipitation sums per station ... 40
Figure 14. Long term mean values for annual, monthly and daily evapotranspiration (mm/d) ... 41
Figure 15. Thiessen polygons applied to the study area ... 42
Figure 16. Long term mean values for annual, monthly and daily precipitation in the catchment. ... 44
Figure 17. Daily precipitation time series for Sio River sub-basin ... 45
Figure 18. Pan evapotranspiration data used in this study originating from the pan measurements in Kakamega. ... 45
Figure 19. Observed flow in Sio River sub-basin ... 46
Figure 20. Model output after calibration based on the NSE. ... 48
Figure 21. Model output after calibration based on the KGE ... 49
Figure 22. Observed streamflow and simulated streamflow with KGE and NSE ... 50
Figure 23. Long-term mean values for annual, monthly and daily flow simulated with KGE and NSE and for the observed streamflow ... 51
Figure 24. Pan evapotranspiration used as input and resulting actual evapotranspiration from the hydrological model. ... 52
Figure 25. Extreme value distribution with MLE of the observed streamflow ... 53
Figure 26. Extreme value distribution with L-moments of the observed streamflow ... 54
Figure 27. Extreme value distribution with MLE of the flow simulated with KGE ... 54
Figure 28. Extreme value distribution with L-moments of the flow simulated with KGE ... 55
Figure 29. Extreme value distribution with MLE of the flow simulated with NSE ... 55
Figure 30. Extreme value distribution with L-moments of the flow simulated with NSE ... 56
Figure 31. Maximum annual observed flow ... 57
Figure 32. Maximum simulated flow with NSE ... 57
Figure 33. Maximum simulated flow with KGE ... 57
Figure 34. Return levels for each return period obtained with each method ... 58
Figure 35… Return level boxplot (m^3/s) ... 59

List of Tables

Table 1. Precipitation, discharge and Temperature/Evapotranspiration Stations 8
Table 2. Time series available per station..10
Table 3. percentage of precipitation data missing per station. Stations highlighted in blue was discarded...11
Table 4. Correlation matrix for station X8934113, date 01-01-1980....................................13
Table 5. Model parameters. source: (eWater Source 4.1.0) 17
Table 6. Equations involved. Source: (eWater Source 4.1.0)...18
Table 7. Parameters in GR4J and their default values. Source: (eWater Source 4.1.0)........20
Table 8. Coefficient values with a 80% of confidence (Perrin et al. 2003)............................20
Table 9. Mean monthly values for each station after data imputation (mm/d) 28
Table 10. Mean annual values (mm/d) for each station after data imputation 31
Table 11. Annual precipitation sums for each station after data imputation (mm/a) 36
Table 12. Sub-basin area divided in corresponding Thiessen polygons...............................43
Table 13. Basin annual precipitation sums 43
Table 14. Temporal settings 46
Table 15. Initial selection of parameters 46
Table 16. Parameter value 47
Table 17. Error criterion before and after calibration with NSE and KGE eq. 47
Table 18. Streamflow values obtained with each method for each return period..................58

1 Introduction

Many engineering and management projects rely on meteorological and hydrological data such as rainfall or discharge. This kind of data is required, for instance, for the design of hydraulic structures such as channels and reservoirs, but also for flood studies, irrigation design projects and water resources planning and management.

Therefore, many countries install meteorological stations, and discharge gauges, in order to develop a monitoring network for capturing the hydro-meteorological conditions in specific areas, catchments or regions. The quality of the data collected in such monitoring network is very diverse and strongly depends on the capacities and resources of the country. However, no matter how much data there is, or how good the data is, in the end, engineers must always find solutions to use the information available and to come up with, e.g. as is the case in this report, an estimation of design values and extreme values.

The precipitation ta measured at specific locations are punctual data, which means that they are not necessarily representative for a territory or a long time period. In order to calculate structural design values and extreme values, interpolation must be done in time and space.

In this report, the hydro-meteorological data from the Sio-Malaba-Malakisi river basin in Kenya has been studied. The quality of the measured data in the area is rather poor, due to data gaps and large measuring intervals. The application of hydrological models can be used to extend discharge time series, when observations are too short, in order to enhance the data basis for estimating design and extreme values. In this study, the focus lies on the Sio River Basin.

This project has three main objectives:

- To provide a full time series for each precipitation measuring station in the area by filling the data gaps (temporal interpolation)
- To estimate the areal precipitation for the Sio River basin
- To apply a hydrological model to extend the discharge information for the Sio River
- To calculate extreme values of discharge for different return periods, based on several data inputs

2 Study area and data basis

2.1 Study area

The study area considered for this paper is part of the Sio River basin, which drains into the Lake Victoria in Kenya. As such, the Sio is a tributary to the White Nile. Even though the Sio River Basin covers a basin of 1 402.11 km², this study focuses on the upstream area above the discharge station "*1AH01*" (blue triangle in Figure 1), that covers an area of 1 010.90 km², equivalent to a 72% of the whole Basin. The elevation ranges in the whole sub-basin lies between 1077 to 1671 m.a.s.l.

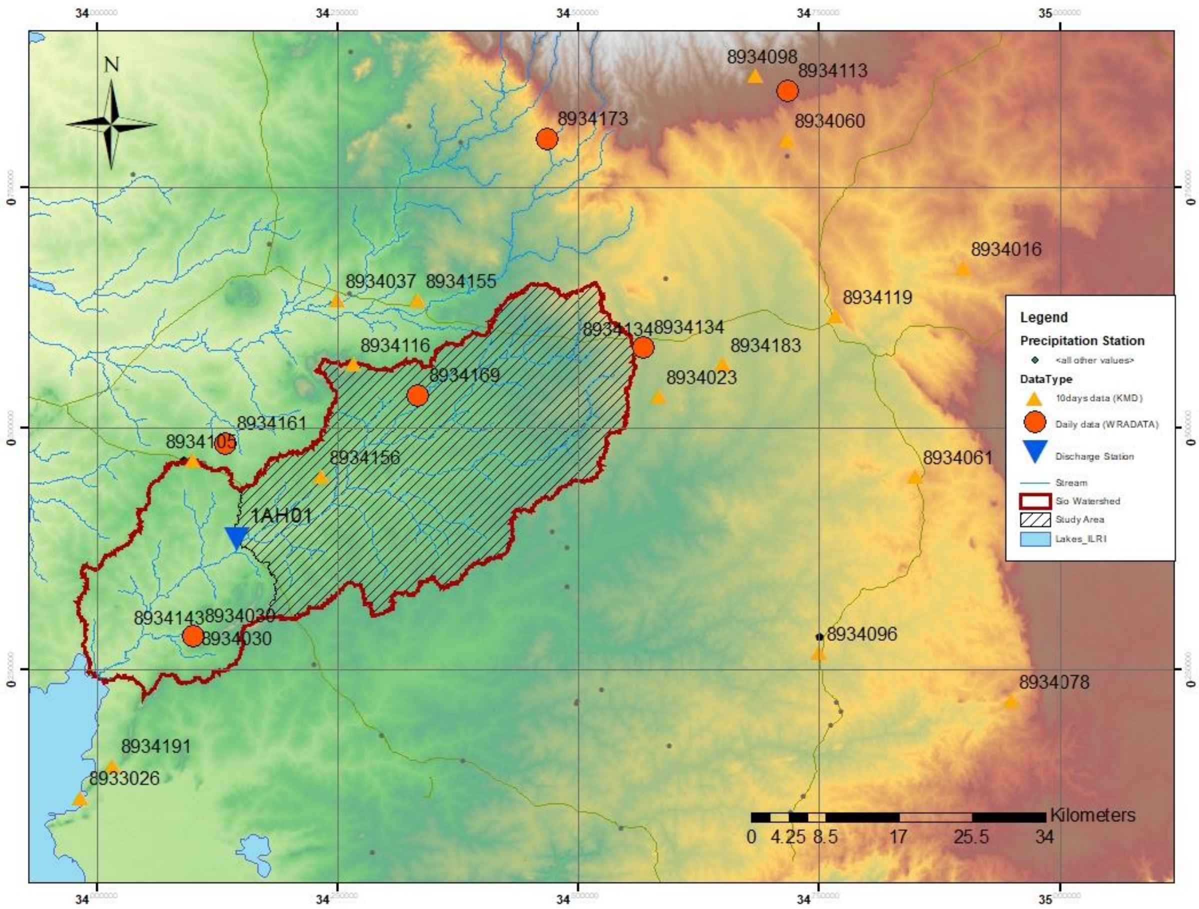


Figure 1. Study area of the Sio River Basin, including precipitation and discharge stations

2.2 Data basis

Precipitation data is available in 25 stations. Besides precipitation, four discharge stations are also available in the near-by area. This study however only covers the catchment upstream of the discharge measurement "1AH01". Details on location, elevation and names of the stations can be seen in Figure 1 and Table 1. .

Figure 2. The Sio River and the Busia-Kisumu road bridge seen from upstream, where the discharge measurement station “1AH01” is located.

Figure 3. Discharge measurement station “1AH01” at the Sio

The data comes from two sources, one is the Kenya Meteorological Department (KMD) and the second one is the Water Resources Authority of Kenya (WRA). Five stations have daily-based data (WRA data), while the other 19 stations are sums of precipitation values for every 10 days (KMD data).

One additional source of daily precipitation data is also considered for this study, namely the “*Climate Hazards Group InfraRed Precipitation with Station* (*CHIRPS*)”. CHIRPS was developed by scientists at the University of California, Santa Barbara and the U.S. Geological Survey Earth Resources Observation and Science Center under the direction of Famine Early Warning Systems Network (FEWS NET) and has over 35 years of quasi global rainfall data. This data source was included in order to support the process of gap filling, which will be explained in detail in the Chapter 3.1.

Table 1. Precipitation, discharge and Temperature/Evapotranspiration Stations

ID	Station name	Measured variable	Source	Longitude	Latitude	Elevation (DEM 90m)
8933026	Port- Victoria Catholic Mission	Precipitation	KMD	33.98	0.12	1134.00
8934016	Lugari Forest Station	Precipitation	KMD	34.90	0.67	1672.81
8934023	Sang'alo Institute Of Science & Technology	Precipitation	KMD	34.58	0.53	1419.73
8934030	Nangina Catholic Mission	Precipitation	KMD	34.10	0.28	1196.22
8934037	Lukolis Dispensary - Busia	Precipitation	KMD	34.25	0.63	1144.17
8934060	Kimilili Agricultural Department	Precipitation	KMD	34.72	0.80	1670.82
8934061	Malava Agiricultural Station	Precipitation	KMD	34.85	0.45	1646.24
8934078	Kaimosi Farmers Training Centre.	Precipitation	KMD	34.95	0.22	1668.48
8934096	Kakamega Meteorological Station	Precipitation	KMD	34.75	0.27	1523.86
8934098	Kimlili Forest Station	Precipitation	KMD	34.68	0.87	2058.22
8934105	Busia Farmers Training Centre	Precipitation	KMD	34.10	0.47	1228.39
8934116	Amukura Chief's Centre	Precipitation	KMD	34.27	0.57	1297.89
8934119	Webuye Agricultural Office	Precipitation	KMD	34.77	0.62	1561.76
8934134	Bungoma Water Supply	Precipitation	KMD	34.57	0.58	1427.25
8934143	Nangina Girls' High School	Precipitation	KMD	34.10	0.28	1196.22
8934155	Amagoro D.O's Office	Precipitation	KMD	34.33	0.63	1224.80
8934156	Nambale Agricultural Office	Precipitation	KMD	34.23	0.45	1233.54
8934183	Nzoia Sugar Factory - Bungoma	Precipitation	KMD	34.65	0.57	1490.14
8934191	Port Victoria Forest Station	Precipitation	KMD	34.02	0.15	1236.23
8934030	Nangina Catholic Mission	Precipitation	WRA	34.10	0.28	1196.22

ID	Station name	Measured variable	Source	Longitude	Latitude	Elevation (DEM 90m)
8934113	Kapsakwony Chief's Office	Precipitation	WRA	34.72	0.85	1938.56
8934134	Bungoma Water Supply	Precipitation	WRA	34.57	0.58	1427.25
8934161	Alupe Cotton Research Station	Precipitation	WRA	34.13	0.48	NA
8934169	Kwangamor	Precipitation	WRA	34.33	0.53	1232.43
8934173	Cheptais Chief's Office	Precipitation	WRA	34.47	0.80	1588.14
1AA01	Malaba river at border bridge	Discharge	WRA	34.27	0.64	NA
1AB01	Malakisi river	Discharge	WRA	34.52	0.84	NA
1AD02	Malakisi river at Bgm-Mlb bridge	Discharge	WRA	34.34	0.63	NA
1AH01	Sio	Discharge	WRA	34.147	0.39	NA
8934096	Kakamega Meteorological Station	Min and max Temperature; Pan Evaporation	KMD	34.75	0.27	NA

2.2.1 Data availability

From the 25 rainfall stations, 19 have more than 30 years of recorded data and 22 have more than 20 years of data (further information in Table 2), but single years exhibit substantial data gaps. Therefore, stations with many gaps were excluded from the analysis.

Table 2. Time series available per station

Station	Start date data	End date data	Years with data	Source
8934161	01.01.1980	31.03.2017	37	KMD
8934016	01.01.1980	03.12.2017	37	WRA
CHIRPS_13	01.01.1981	31.12.2017	36	CHIRPS
8934061	01.01.1980	03.05.2016	36	WRA
8934096	01.01.1980	03.07.2016	36	WRA
8934105	01.01.1980	03.05.2016	36	WRA
8934116	01.01.1980	03.05.2016	36	WRA
8934156	01.01.1980	03.05.2016	36	WRA
8934037	01.01.1980	03.09.2015	35	WRA
8934155	01.01.1980	03.09.2015	35	WRA
8934030	01.01.1980	31.03.2014	34	KMD
8934030	01.01.1980	03.02.2014	34	WRA
8934183	01.04.1980	03.08.2014	34	WRA
8933026	01.01.1980	03.07.2013	33	WRA
8934134	01.01.1980	03.08.2012	32	WRA
8934191	01.01.1984	03.05.2016	32	WRA
8934078	01.01.1980	03.11.2011	31	WRA
8934119	01.01.1980	03.12.2011	31	WRA
8934098	01.01.1980	03.12.2010	30	WRA
8934143	01.01.1984	03.09.2013	29	WRA
8934060	01.01.1980	03.10.2003	23	WRA
8934134	01.01.1980	15.07.2000	20	KMD
8934113	02.01.1985	01.01.2002	17	KMD
8934023	01.01.1980	03.03.1995	15	WRA
8934169	01.01.2012	31.12.2017	5	KMD
8934173	02.01.2001	01.01.2002	1	KMD

In order to select which station should be used within the following procedures, a criterion of maximum 90% of missing values is defined. This means that rainfall stations with less than 10% of values, will be exclude from any further calculations. One of the 25 stations (8934173) was consequently dismissed from the study, due to the short series that it contains (2.6%). The percentage of data missing for each station is given in Table 3 .

Table 3. percentage of precipitation data missing per station. Stations highlighted in blue was discarded

Station / Data basis	% Precip. data not available
8934030	75.2%
8934113	58.2%
8934134	59.1%
8934169	84.2%
8934173	97.4%
8934161	27.5%
CHIRPS_13	2.6%
8933026	73.7%
8934016	19.1%
8934023	63.8%
8934030	58.1%
8934037	54.2%
8934060	39.3%

Station / Data basis	% Precip. data not available
8934061	21.1%
8934078	40.1%
8934096	6.4%
8934098	56.4%
8934105	20.4%
8934116	63.2%
8934119	33.3%
8934134	22.8%
8934143	79.4%
8934155	59.6%
8934156	29.0%
8934183	31.4%
8934191	30.6%

Besides that, out of all the precipitation stations, only 3 are located inside the study area, listed below, and 22 in the surroundings (See Figure 1).

- 8934116: Amukura Chief's Centre (10 days_KMD) - 1297.89 m.a.s.l.
- 8934156: Nambale Agricultural Office (10 days_KMD) - 1233.54 m.a.s.l.
- 8934169: Kwangamor (daily_WRA) - 1232.43 m.a.s.l.

3 Methods

As the aim of this study is the estimation of design or extreme values for different return periods the methods involved must include a hydrological model that can enhance the data availability of discharge from the catchment. This hydrological model requires a substantial amount of rainfall data as an input. Reasons for data gaps and discontinuities in time series and measurements of the data include natural, human, equipment among other reasons.

For all the stations it is necessary to fill the gaps within the period 1980-01-01 to 2017-12-31.Since the rainfall stations contain 2 different measurement intervals, daily and 10 day sums, the data gap filling procedure has to take this into account.

In regions where hydro-meteorological is scarce, or data is of low quality or high measurement uncertainty, hydrological models can be used to enhance information data base. A hydrological model is a simplification of the real system, which helps to understand the processes leading from rainfall to runoff and can help in the prediction and management, with its limitation, of the water resources.

In order to set up a Hydrological model, long times series of precipitation (and temperature, potential evapotranspiration) are required for the calibration and validation. Since the data series obtained from all stations have data gaps, the first step is to fill these missing values. Once the time series is complete, it is necessary to determine how relevant each station is for the area of study. As is mentioned above, some stations are directly located within the sub-basin analyzed; For estimating the areal precipitation of the catchment, the Thiessen polygons method is used. Due to the flatness of the area, a low variability in precipitation is expected, justifying the application of this simple method for areal interpolation.

With the respective polygon enclosing each station, and its proportional relevance towards the total area, the areal precipitation values can be calculated and the be used as input for the hydrological model.

Once the model is calibrated, the estimation of extreme valuescan be done. in the description of the hydrological model is given in chapter 3.4.

It is important to keep in mind that hydrological phenomena have a high variability in space and time Therefore all the estimations done here cannot be considered as an absolute truth, but as a result of statistical analysis, influenced by several assumptions, evident errors in the measurements and in summary large uncertainties.

3.1 Temporal interpolation (Data filling)

Missing data imputation is in some cases crucial to use the data available and not reject records than can be useful for the future analysis. Hydrological models and statistical analysis, such as the case involved here, rely on (long) time series, hence the quality and completeness of time series is essential and the preprocessing of raw data sets is therefore a necessary procedure (Gao et al. 2018).

The collection of environmental data is normally subject to many unplanned incidents throughout time, that compromises the data gathering exercise. Meteorological or generally extreme events, equipment failure, human intervention, among others can hinder the data collection (Baguley 2012).

Several techniques for data imputation exist, including:

(i) *Deletion*, which means to drop or remove dataset with missing values from the analysis; could be explicit or implicit. The implicit method is more often used, with it values are assigned in the missing values, for example NA or a value out of range (i.e. -999).

(ii) *Single imputation*: (ii) *Random substitution*: missing value are imputed from a randomly selected similar record, i.e. Last Observation Carried Forward (or LOCF),

(iii) *Mean substitution*: replacing any missing value with the mean of that variable for all other cases, and

(iv) regression, in which available information for complete and incomplete cases is used to predict the value of a specific variable.

Two methods that improve on single imputation are (vi) maximum likelihood and (vii) multiple imputation (MI).

Mean substitution is not recommended, because this technique inflates the sample size and leads to standard errors that are too small In addition, covariances and variances are underestimated, which also leads to a bias when estimating σ^2 and σ (Baguley 2012). A more refined variation is to employ all the available data to estimate those missing values for instance via regression. The estimations are meant to substitute missing values. This method is known as *conditional mean substitution* or *regression imputation*.

Regression can produce unbiased parameters for intercept and slope but since the estimated value will fall exactly in the regression line, less variables as they would have been with real values are considered.

For this project a simple data filling procedure, based on linear regression models was chosen. The linear regression models are thereby estimated based on the highest correlation among *stations with available data* for the same month. With similar stations is meant not only about physical conditions but also temporary distribution, as it was previously referred, a group of stations have only daily data while the second group has the sums of 10 days. Therefore, it was in the first place compared all the daily data stations between each other, and separately the stations with 10-day data between them.

A correlation matrix (see Table 4) relates, each time step with missing data (NAs) and for every station with all the stations that do have data at this time step and in the same month during the whole time series.

Table 4. Correlation matrix for station X8934113, date 01-01-1980

X8934113	CHIRPS_13	X8934161	X8934134	X8934030	X8934169
1	0.268	0.098	-0.011	-0.015	NA

Table 4 shows an example for a correlation matrix for a specific day (01.01.1980). The station *X8934113* has for the month of January from all the daily records the highest correlation with station *CHIRPS_13* for the given month*.* This means, that the gap in X8934113 will be filled with a linear regression model using the data from CHIRPS_13. The same principle is applied to every missing daily data. Once all the gaps in the daily stations are filled, the same technique is implemented for the other 19 stations with 10-day values. Once the datasets for 10-day is gap-free, a further step is applied to temporally disintegrate the 10-day sums to daily information.

The daily stations were converted to 10-day data, which means that each third part of the month is aggregated, this must be done in order to have comparable data between the 2 categories of the available data sets. Once the daily data is aggregated, a correlation analysis is implemented among all stations, to fill the gaps in every 10-day rainfall stations during the whole time series.

Each of the 10-day stations has been associated with the highest correlated daily station to fill the gaps, with it is calculated the proportion of rain among the 10 days, that were previously summed, so a disaggregation of 10-day data to daily data can be achieved based on the percentage of rain for each of the days, which is applied to the filled value obtained. Here, it is assumed that a pair of stations will have the exact same temporal distribution of rainfall, which could or could not reflect the real distribution of a rainfall event, therefore the results here presented cannot be assumed as a great truth. Now that the temporary imputation is concluded, the calculation of the areal values for the basin can be done.

3.2 Spatial interpolation

Rainfall observations are measured at single locations, which is usually not representative for the overall catchment. There are many methods to transform point values to areal estimates, with different results being obtained, event with the same network data, depending on the method applied (WMO 2008). To obtain the mean areal precipitation from gauge data, methods such as arithmetic average method, Thiessen polygon method and inverse distance-squared method or other interpolation methods exist. For the given case, due to the not very complex topographical conditions, the Thiessen polygon method is applied.

The Thiessen polygon Method is widely used in engineering practice for the estimation of the spatial distribution of rainfall. This method assigns a weight to each gauge station in proportion to the catchment area that is closest to that gauge. The steps of the method is described by UTHH (2011) and is shown below:

- *Gauge network is plotted on a map of the catchment area of interest.*
- *Adjacent stations are connected with lines.*
- *Perpendicular bisectors of each line are constructed (perpendicular line at themidpoint of each line connecting two stations)*
- *The bisectors are extended and used to form the polygon around each gauge station.*
- *Rainfall value for each gauge station is multiplied by the area of each polygon.*
- *All values from step 5 are summed and divided by total basin area.*

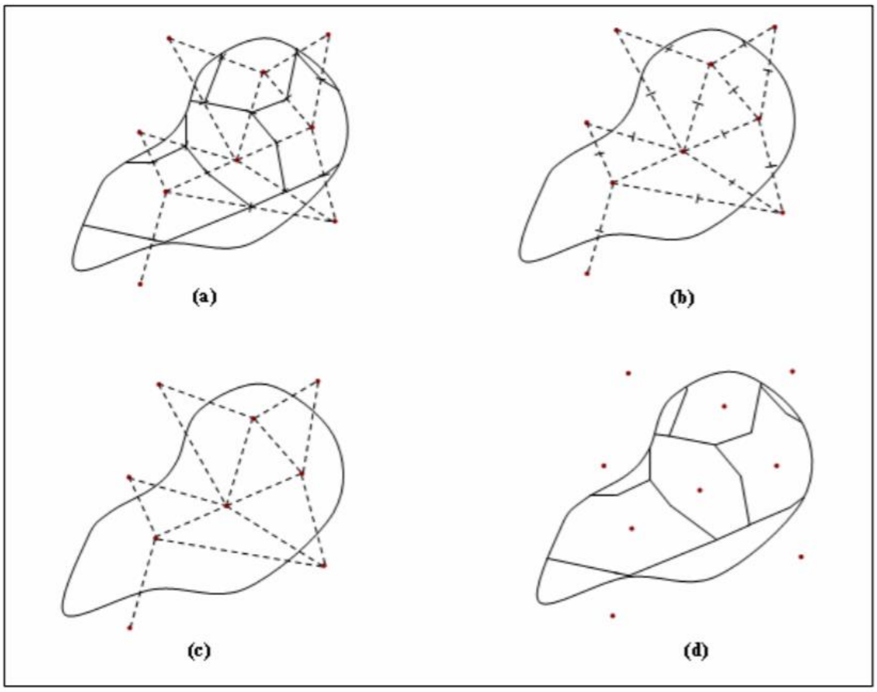


Figure 4. Construction of Thiessen polygon (UTHH 2011)

Although this method is widely used, its application has to be carefully implemented, because the proportionality takes into consideration only a 2-dimension surface, meaning that the z-axis or elevation is neglected. In cases, where the topography shows no significant changes, as it is in this study, the method is appropriate. Otherwise, the isohyets method, can be employed since it takes into account topographical differences within the catchment area (Chavula 2013).

With the weighted area, the average catchment precipitation is calculated with the equation:

$$\overline{P} = \frac{P_1A_1 + P_2A_2 + P_3A_3 + \ldots + P_nA_n}{A_1 + A_2 + A_3 + \ldots + A_n} = \frac{\sum_{i=1}^{n} P_iA_i}{\sum_{i=1}^{n} A_i}$$

where $\overline{P}$ is the weighted average, P_i's are measurements, and A_i's are areas of each polygon.

Once an average precipitation is calculated, the input time series for a hydrological model is completed.

3.3 Hydrological Modelling

Modeling runoff contributes to understanding hydrologic phenomena and the effects involved in the hydrological cycle. Hydrological models are tools used by hydrologists and engineers to answer questions in different areas in water resources, ranging from resource management, to questions related to urban and rural hydrology and groundwater resource management. *Many hydrological models have been developed and refined during the past four decades and it is required to fully understand their characteristics to effortlessly employ them* (Jajarmizad et al. 2012).

The classification of hydrological models is not an easy task since the nature of models is often similar and it is frequent to have models with similar characteristics. A classification introduced by Singh (1995) is shown in Figure 5.

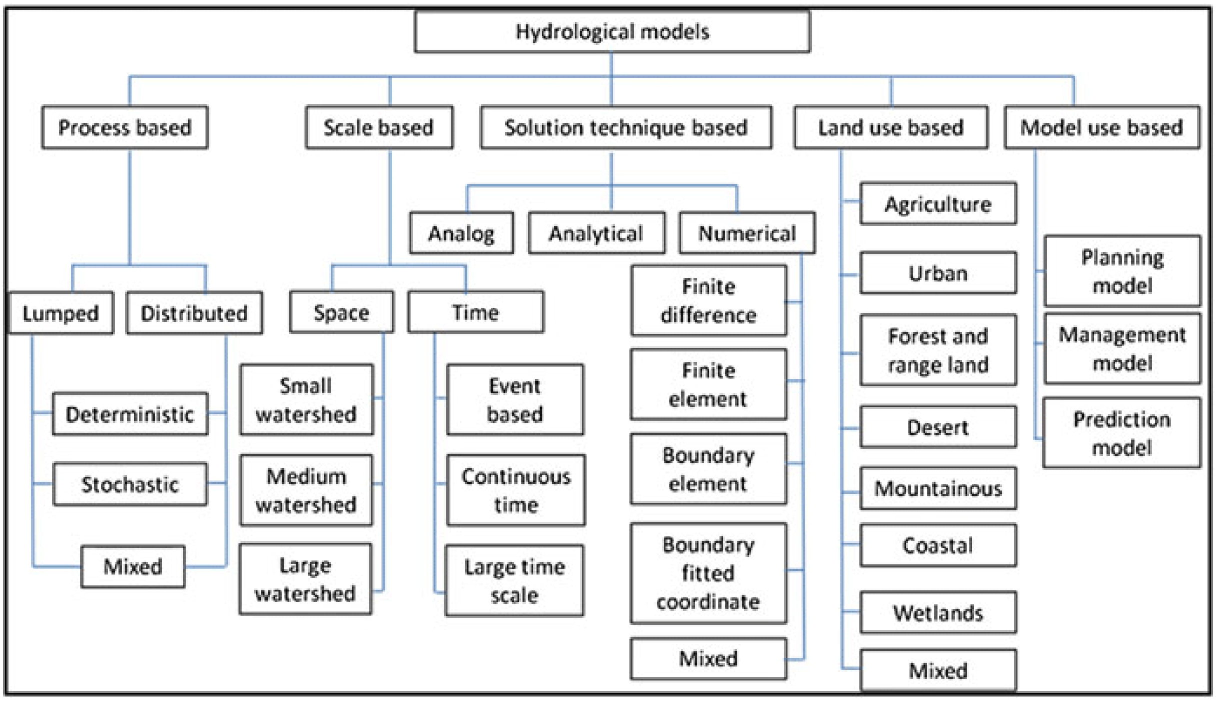


Figure 5. Classification of hydrological models proposed by Singh (1995). Source: (Saha and Zeleke 2015, p. 568).

Conceptual rainfall-runoff models can estimate streamflow with minimum inputs. Several models exist. Anshuman et al. (2019) compiled some of them: "*AWBM (Boughton 2004), IHACRES (Croke et al. 2006), Sacramento (Burnash 1973), SIMHYD (Chiew et al. 2002), SMARG (Goswami et al. 2002), GR4J (Perrin et al. 2003), and SURM (Delgado 2013)".*

From all of them, GR4J has the simplest structure, allowing for a robust calibration process, as well as a package version in R, the programming language selected for this study.

3.3.1 GR4J

The GR4J model was initiated by Claude Michel in the earlies 80s at CEMAGREF, a public research institute in France. At the beginning, the model included one parameter, but the newest versions have up to 5 parameters. *The GR4J modelling approach is mainly empirical* (Michel et al., 2006). GR4J is a **continuous lumped conceptual model**, with a daily time step. It is also capable to include soil moisture in its calculations.

The versions of the GR4J model are: (i) 3-parameter by Edijatno and Michel (1989) and Edijatno (1991); (ii) 4-parameter by Nascimento (1995) and Edijatno et al. (1999) ; (iii) 4-parameter by Perrin et al. (2003) and (iv) 5-parameter by Le Moine (2008).

The main characteristics of the models are (Perrin et al. 2007):

- Process level: simple behavioral relationships at the basin scale, developed empirically and without direct links to the physics of small-scale processes and which can represent an average of several processes
- Spatial approach: GR models are global, a watershed is consider as a whole, an heterogeneous body
- Temporal approach: the GR models are develop for specific operating time steps: annual (GR1A), monthly (GR2M) and daily (GR4J).

Conceptual models express runoff processes with simplification of the whole hydrological process: storage, flow input and output are used to represent an overall idea of the response of a given catchment. This representation is based on the water balance equation that converts rainfall into runoff, evapotranspiration and groundwater recharge, by calculating a distribution of the input towards each component. The model simulates the exchange between atmosphere, storage and hydrological components based on this equation. Storage, both in soil and groundwater are idealized in models. The general governing equations is:

$$\frac{dS}{dt} = P - ET - Q_s \pm GW$$

where

$\frac{dS}{dt}$ is the change in storage, P is precipitation, ET is evapotranspiration,Q_s is surface runoff and GW is groundwater.

Lumped models refer to representations that simplifies a basin as a single homogeneous unit, this means, that the characteristics of the catchment are assigned as equal for the whole area. This type of models are meant to simulate a complete point discharge (runoff and streamflow) and it is not designed to give specific flows within the catchment. Lumped models are suitable for long-term purposes. (EPA 2017). The assumptions of the hydrological process considered inside the model, i.e. land use averaged changes within watershed, tend to under and overestimate runoff values, which leads to uncertainties that have to be kept on mind.

“By assuming homogeneity over the catchment, lumped models lose spatial resolution of the data. For example, rainfall and runoff patterns vary over space and time, but in lumped models they are considered stationary. Due to the many assumptions and averaged conditions that lumped models incorporate, they do not represent large watersheds and catchments accurately (Moradkhani & Sorooshian, 2008)” (EPA 2017).

3.3.1.1 Structure and processes

The model described in Perrin et al. 2003 is represented schematically in the diagram of Figure 6. For a given time step as input: rainfall depth (P) and potential evapotranspiration (E) are given. P is introduced as an estimation of the areal rainfall, reason why areal interpolation was done, and E can be based on monthly or daily data. Input, output and internal fluxes and states are expressed in mm. The model parameters displayed in Figure 6 are given in Table 5:

Table 5. Model parameters. source: (eWater Source 4.1.0)

Parameter	Definition	Unit
E	Potential areal evapotranspiration	mm
$\boldsymbol{E_n}$	Net evapotranspiration capacity	mm
$\boldsymbol{E_s}$	Actual evaporation rate	mm
$\boldsymbol{F(X_2)}$	Groundwater exchange term	mm
P	Areal catchment rainfall	mm
Perc	Percolation leakage	mm
$\boldsymbol{P_n}$	Net rainfall	mm
$\boldsymbol{P_r}$	Total quantity of water to reach routing functions	mm
$\boldsymbol{P_n - P_s}$	Amount of net rainfall that goes directly to the routing functions	mm
$\boldsymbol{P_s}$	Amount of net rainfall that goes directly to the production store	mm
Q	Total stream flow	mm
$\boldsymbol{Q_1}$	Output of uh2	
$\boldsymbol{Q_9}$	Output of uh1	
$\boldsymbol{Q_d}$	Direct flow component	
$\boldsymbol{Q_r}$	Routed flow component	

R	Water content in the routing store	
S	Water content in the production store	
$\boldsymbol{UH_1}, \boldsymbol{UH_2}$	Unit hydrographs	
$\boldsymbol{X_1}$	Capacity of the production soil (sma) store	(mm)
$\boldsymbol{X_2}$	Water exchange coefficient	(mm)
$\boldsymbol{X_3}$	Capacity of the routing store	(mm)
$\boldsymbol{X_4}$	Time parameter for unit hydrographs	(days)

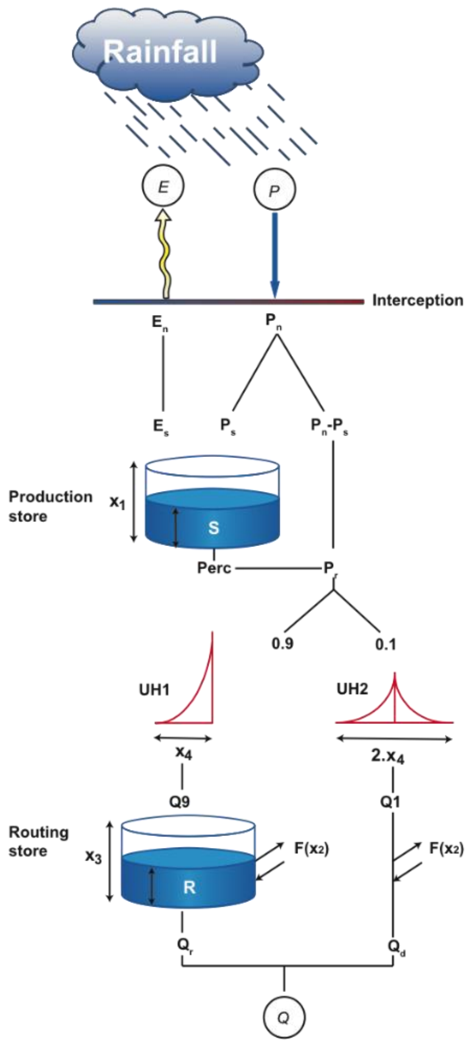


Figure 6. Schematic diagram of the GR4J model. Source: (eWater Source 4.1.0)

Table 6 gives an overview regarding the equations in the hydrological model applied.

Table 6. Equations involved. Source: (eWater Source 4.1.0)

Explanation	Equation	
Determination of net rainfall and PE	If (P≥E) then : $\boldsymbol{P_n} = P - E\ ;\ E_{\boldsymbol{n}} = 0.$ If (P<E) then : $\boldsymbol{P_n} = 0\ ;\ E_{\boldsymbol{n}} = E - P.$	Eq. 1 Eq. 2
Production store: includes soil moisture.	$$\boldsymbol{P_s} = \frac{x_1 * \left(1 - \left(\frac{S}{x_1}\right)^2\right) * \tanh\left(\frac{P_n}{x_1}\right)}{1 + \frac{S}{x_1} * \tanh\left(\frac{P_n}{x_1}\right)}$$	Eq. 3

Explanation	Equation	
	$E_s = \dfrac{S*\left(2-\frac{S}{x_1}\right)*\tanh\left(\frac{E_n}{x_1}\right)}{1+\left(1-\frac{S}{x_1}\right)*\tanh\left(\frac{E_n}{x_1}\right)}$	Eq. 4
Update water content in the production store *Note: S can never exceed x_1	$S = S - E_S + \boldsymbol{P_s}$	Eq. 5
percolation leakage from the production store	$Perc = S\left\{1-\left[1+\left(\frac{4}{9}\frac{S}{x_1}\right)^2\right]^{-1/4}\right\}$	Eq.6
	$S = S - Perc$	Eq. 7
Linear routing with unit hydrographs	$\boldsymbol{P_r} = Perc + (\boldsymbol{P_n} - \boldsymbol{P_s})$	Eq.8
	SH1: For t≤0 SH1 (t)=0	Eq. 9
	SH1: For 0<t<x_4 SH1(t) = $\left(\frac{t}{x_4}\right)^{5/2}$	Eq.10
The ordinates of both unit hydrographs are derived from the corresponding S-curves along time	SH1: For t ≥x_4 SH1(t)=1	Eq.11
	SH2: For t≤0 SH2 (t)=0	Eq.12
	SH2: For 0<t<x_4 SH2(t) =0.5 $\left(\frac{t}{x_4}\right)^{5/2}$	Eq.13
	SH2: For x_4 t <2 $*$ x_4 SH2(t)=1 – ½ $\left(2-\frac{t}{x_4}\right)^{5/2}$	Eq. 14
	SH2: For t ≥ 2 $*$ x_4 SH2(t)=1	Eq. 15
UH1 and UH2 ordinates are:	UH1(j)=SH1(j)-SH1(j-1) UH2(j)=SH(j)-SH2(j-1)	
At each time-step, two unit hydrographs correspond are given by:	$Q9(k) = 0.9\sum_{j=1}^{l} UH1(j)*Pr(k-j+1)$ $l = \mathrm{int}(x_4)+1$ $Q1(k) = 0.1\sum_{j=1}^{m} UH2(j)*Pr(k-j+1)$ $m = \mathrm{int}(2*x_4)+1$	
Inter catchment groundwater exchange *Note: F cannot be larger than x_2	$F = x_2\left(\frac{R}{x_3}\right)^{7/2}$ R: level in routing storage, X_3: reference capacity X_2: water exchange coefficient (positive in case of import, negative in case of export or 0 for no-exchange)	
Non linear routing store	$R = \max(0;\ R + Q9 + F)$	
The outflow of the reservoir is:	$Qr = R\{1-[1+(\frac{R}{x_3})^4]^{-1/4}\}$	
The level in the reservoir is:	$R = R - Q_r$	
Total stream flow:	$Q_d = \max(0;\ Q1 + F)$ $Q = Q_r + Q_d$	

3.3.1.2 Model parameters and calibration

GR4J includes 4 parameters, which were calibrated in this study. x_1, x_2, x_3, x_4,correspond to different parameters that affect the ouput of the model. GR4J includes these parameters with a default value and a range. The range can be modified by the user in the calibration procedure.

Table 7. Parameters in GR4J and their default values. Source: (eWater Source 4.1.0)

Coef	Description	Unit	Default	Range
x1	production store capacity	mm	350	1-5000
x2	intercatchment exchange coefficient	mm/d	0	-10.0 – 5.0
x3	routing store capacity	mm	40	1-500
x4	unadjusted unit hydrograph time constant	days	0.5	0.5 - 0.4

These values are initially chosen by the user but the best fit for the model is optimised during the process of calibration. One approach to define the first set of parameters is assigning values within a confidence interval, for example Perrin et al. (2003), defined a range of 80% confidence for the four parameters (Table 8). These ranges are also used to restrict the values of the parameters during the calibration, in order to get satisfactory and furthermore, realistic parameters.

Table 8. Coefficient values with a 80% of confidence (Perrin et al. 2003)

Coef	Unit	Median value	80% Confidence interval
x1	mm	350	100 - 1200
x2	mm	0	-5 – 3
x3	mm	90	20 – 300
x4	days	1.7	1.1 – 2.9

The initial parameters should be set within the range mentioned to avoid discrepancies at the beginning of the simulation periods. In addition to that, a warm up period is set at the beginning of each simulation, at least one year is recommended.

Once the model runs, the output data obtained are: production store (Prod), net rainfall (Pn), Production store (Ps), Actual evapotranspiration (AE),Percolation (Perc), Linear routing with unit hydrographs (PR), unit hydrographs (Q9, Q1), Nonlinear routing store (Rout), Exchange capacity Exch, outflow of the reservoir (QR), Inter catchment groundwater exchange (QD), Total stream flow Qsim.

The first attempt is here calculated with user-defined parameters, to show the effects of wrongly selected parameters, the results might not be the closest to the true values. Therefore, a parameter calibration process is necessary. In this process, the model parameters (x1 to x4) are automatically changed and the corresponding simulated discharge is compared to the observed values, using an objective function..

For this study, the NSE (Nash and Sutcliffe 1970) and KGE (Gupta et al. 2009) are used as measures to estimate model performance. Both equations are included in the R package of GR4J as efficiency criteria. Based on each proposed formula, all 4 parameter are iteratively adapted in order to find an optimal value, in combination with the other 3, results in a high objective function value.

The NSE criterion is defined with the equation (Nash and Sutcliffe 1970):

$$\boldsymbol{NSE = 1 - \frac{\sum_{t=1}^{T}(Q_m^t - Q_o^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2}}$$

where Qo is the mean of observed discharges,

Q_m^t is the modeled discharge,

Q_o^t is the observed discharge at time t.

The range of the criterion goes from $-\infty$ to 1. A value of *NSE* = 1, means a perfectly fitted model in comparison to observed data. It is also possible to obtain a *NSE* < 0, which means that the observed mean is a better predictor than the model. A model can be qualified as “good quality” with a NSE value between 0.5 and 0.65 (Ritter and Muñoz-Carpena 2013)

The KGE criterion proposed by Gupta et al. (2009) is defined as

$$KGE \ = 1 - \sqrt{(r-1)^2 + (\propto -1)^2 + (\beta - 1)^2}$$

$$r = \ linear\ correlation\ coefficient\ between\ simulated\ and\ observed$$

$$\propto = \frac{\sigma_{sim}}{\sigma_{obs}}$$

$$\beta \ = \frac{\mu_{sim}}{\mu_{obs}}$$

Comparable to the NSE, KGE = 1 indicates perfect agreement between simulations and observations. KGE <0 indicates that the mean of observations provides better estimates than simulations

In this study, optimization of parameters is performed using the NSE and the KGE.

3.4 Estimation of Extreme values

Extreme value analysis consists of studying events with low probability of occurrence. They refer to the values found in the tails of a probability distribution function, as shown in Figure 7. An extreme value can be either a maximum or a minimum, and it represents rare or extreme event such as a natural hazard in environmental sciences.

The main aim of an extreme value analysis is to predict the probability of rare events which have a very low probability, and which may not have been recorded. In other words, it intends to relate the magnitude of an event with its probability of happening. The underlying assumption is that the observed values (=sample) follow a defined distribution function. Given this distribution function, it is possible to estimate the extreme values as a function of probability.

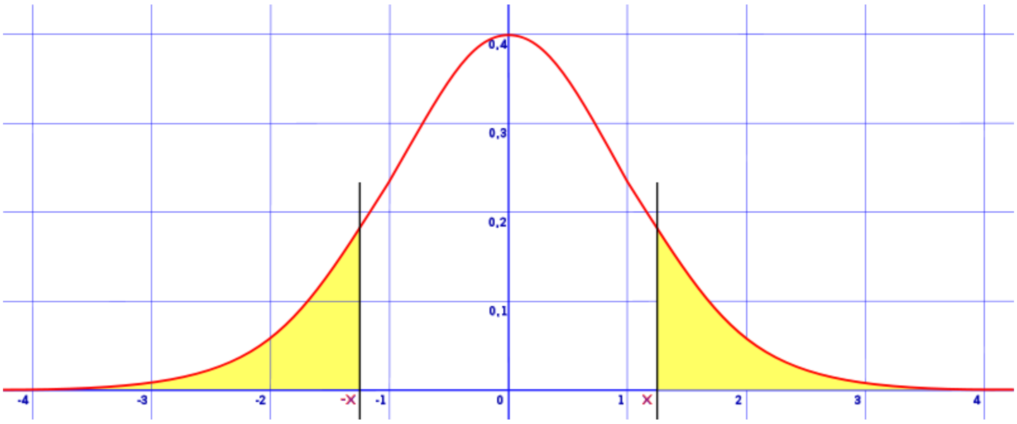


Figure 7. Extreme values in a normal distribution function (Glen 2016)

3.4.1 Generalized Extreme Value (GEV) distribution

In extreme value analysis, an extreme value distribution (EVD) function is fitted to the values of the either higher or lower tails of a normal distribution function. As shown in Figure 88 , there are three types of extreme value distribution functions used in the GEV (Haan 1977):

- Gumbel Distribution
- Fréchet Distribution
- Weibull Distribution

All these distributions are similar and can be used to fit a data set of rare events, as long as the variables are random, independent and identically distributed (Gumbel 2013). The Generalized Extreme Value (GEV) distribution function, is a function that merges the three types of extreme value distributions in one equation.

$$G(z) = \exp\left[-\left\{1+\xi\left(\frac{z-\mu}{\sigma}\right)\right\}_{+}^{-1/\xi}\right]$$

The parameters of the GEV distribution are:

μ : Location parameter

σ : Scale parameter

ξ: Shape parameter

Depending on the sign of the shape parameter (ξ), the GEV assumes one of the three extreme value distribution function types.

If $\xi < 0$, then the GEV is equal to the Weibull distribution function (Type III).

If $\xi > 0$, then the GEV is equal to the Fréchet distribution function (Type II).

If $\xi = 0$, then the GEV is equal to the Gumbel distribution function (Type I).

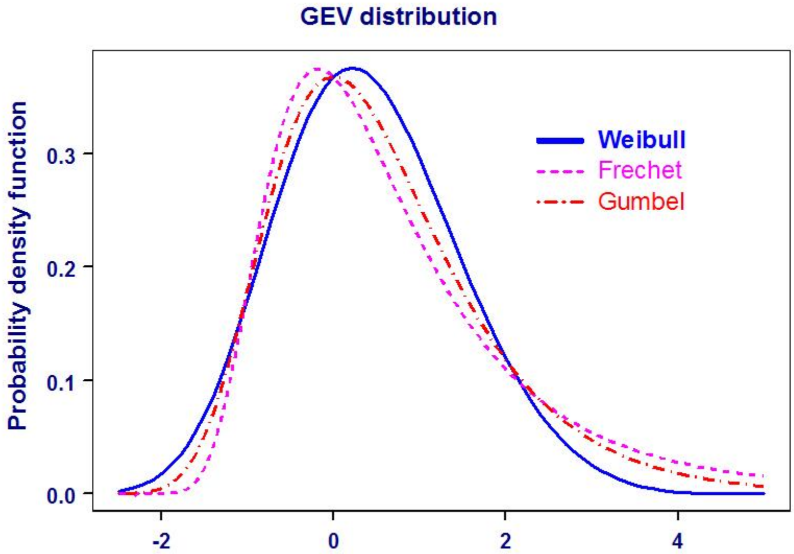


Figure 8. Extreme value distribution functions (Gilleland and Katz 2016)

3.4.2 Parameter estimation

The parameter values of the GEV distribution function are estimated based on observed values. In this case study, two estimators are used: the maximum likelihood estimation (MLE) and the L-moments estimation. ((Gilleland and Katz 2016))

3.4.2.1 L-moments

The method of L-moments was first introduced by Hosking (1990). It is a good estimating method because it is robust regarding outliers and results in good parameter values for small samples. Moreover, it is a clear method and easy to interpret (BMLFUW 2011).

The r^{th} L-moment of a distribution of the X variable is defined as:

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathrm{E} X_{r-k:r}$$

the first four L-moments λ_1, λ_2, λ_3 and λ_4 are used to assess the location, the spread, the loop and the curvature of the distribution function.

3.4.2.2 Maximum Likelihood

The maximum likelihood estimation (MLE) method is based on the idea that, for an assumed distribution function, the chosen parameters must be those that most likely reflect the assumed distribution. From a theoretical point of view, the MLE has numerous favorable properties, but the errors with shorter time series (<500 values) are significantly higher than with other estimation methods. (BMLFUW 2011)

In the MLE, the parameter determination is transformed to an optimization function where the target function L is the following:

$$L(u_1,\ldots,u_m) = f(x_1, u_1,\ldots,u_m) * f(x_2, u_1,\ldots,u_m) * \ldots * f(x_n, u_1,\ldots,u_m)$$

$u_1,\ldots,u_m$: the *m* parameters of the distribution function to be determined.

$f(x_1, u_1, \ldots, u_m)$: value of the distribution function for the observation value x of n values.

For a result that is as accurate as possible, the parameters must be selected so that $L(u_1, \ldots, u_m)$ becomes maximum.

3.4.3 Extreme value data set selection

The extreme value distribution (EVD) function is fitted to the extreme values, meaning either the maximum or the minimum values of the data set, as shown in Figure 7. There are two approaches to determine the extreme values that are used for the extreme value analysis,

i. the block maxima approach and
ii. the peaks over threshold approach.

The first approach takes the maxima of a long data block, for instance, the annual maximum values. The dataset is considerably reduced with this approach; therefore, it is useful for long time series. However, it is important to point out that it could happen that in one year there are two extreme events, and only the greater one would be selected for the extreme values analysis (Gilleland and Katz 2016)

The second approach consists in analyzing the data that exceeds a given threshold. This approach is favorable for shorter time series, as the threshold can be set up at any level. This method has some drawbacks, as a same event can last more than one day, exceeding the threshold a couple of times in a row, and it would be overrepresented. In this case, only the highest value of each event should be selected to guarantee independence. This can be done through declustering, where only the single highest value is retained (Gilleland and Katz 2016).

For this case study, the block maxima approach was chosen. The annual maximum streamflow was calculated for each year with recordings, and this was used to fit a GEV distribution function with the R package extRemes (Gilleland and Katz 2016).

3.4.4 Return period and return level

Once an extreme value distribution function has been defined and fitted with existing data, in this case, the maximum annual streamflow, the probability of occurrence of an event of any magnitude can be extracted, with a confidence interval. The concepts of return level and return period are used to describe the magnitude and probability of an event happening, and are depicted in Figure 9.

The return level is represented by $x(p)$ and refers to the magnitude of an event that has a p probability of being exceeded (Gilleland and Katz 2016).

The return period is represented by T, and it refers to the average waiting time until an event of a similar magnitude happens again. This means that an event of a given magnitude will happen in average every T-years (Gilleland and Katz 2016).

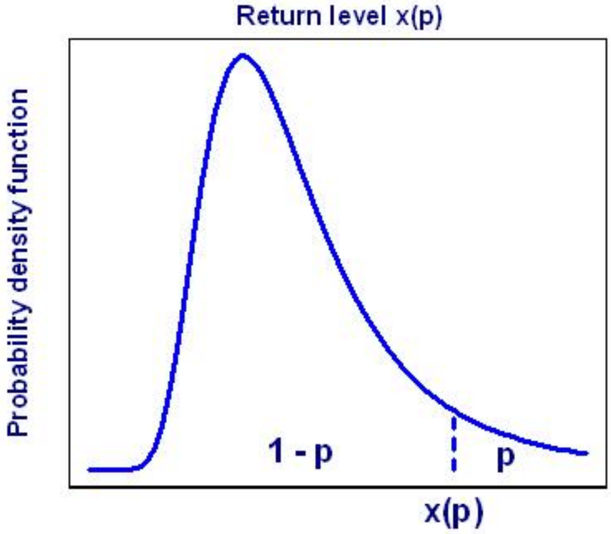


Figure 9. Return period and return level representation (Gilleland and Katz 2016).

The return period is defined as the inverse of the probability of exceedance of an event, so $T = 1/p$. For example, an event with a 1% probability of exceedance $p = 0.01$, will have a return period $T = 1/p = 100$ years.

In order to provide an estimate of the intensity of an event, a flood frequency curve is often used, where discharge is plotted against its corresponding return period, as shown in Figure 10. In the graph, the points are the measured points, the straight line corresponds to the model fit to the measurements, and the discontinuous lines depict the 90% confidence interval.

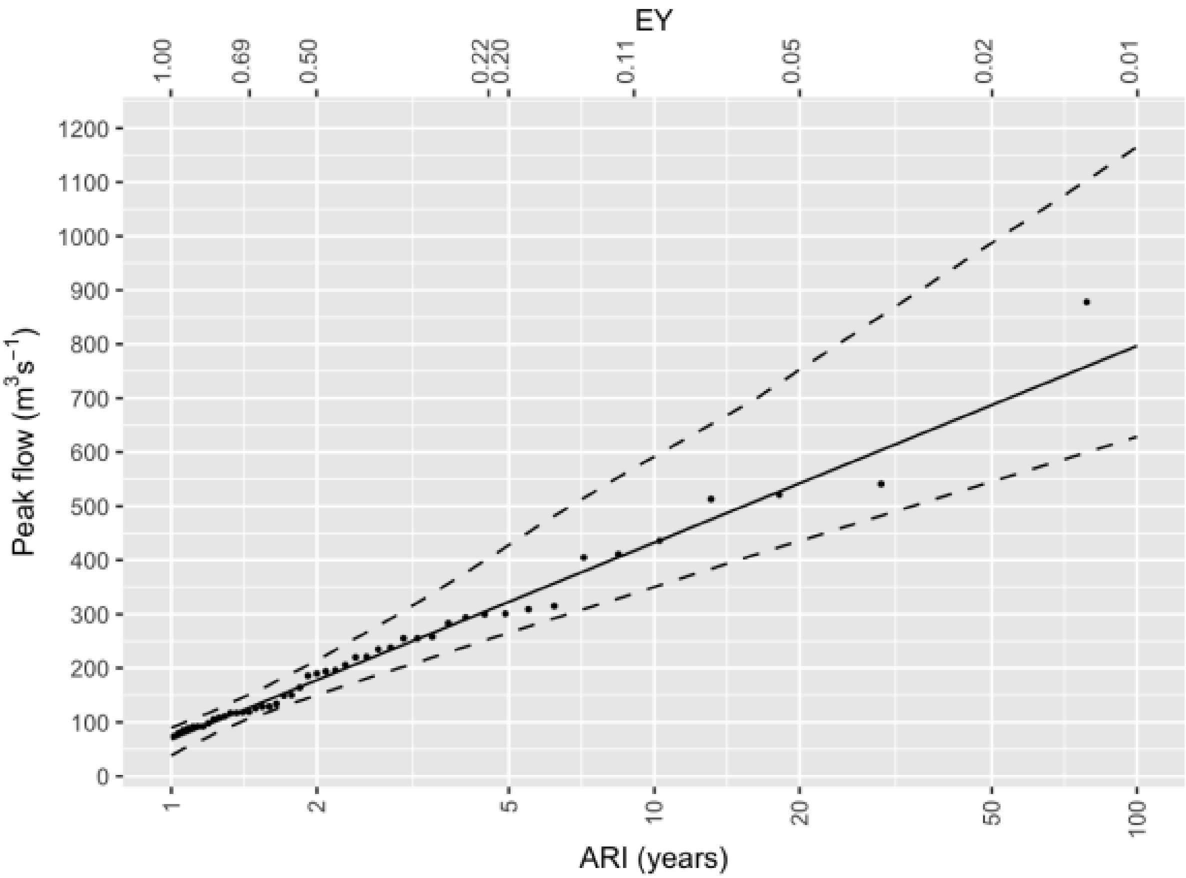


Figure 10. Partial series for the Styx River at Jeogla showing model fit and the 90% confidence interval(Institution of Engineers 1987).

4 Results

In this chapter the results obtained with the input data, described in Chapter 2.2, and with the methodology defined in Chapter 0 are displayed. It includes also the discussion of the analysis done after each procedure and ties it with the study area background. The tables and graphs included are analyzed with three different scopes: (i) temporal analysis for every station, (ii) temporal and spatial analysis in the whole catchment and (iii) extreme value estimation.

4.1 Temporal interpolation (Data filling)

4.1.1 Precipitation

Tables with long-term annual and monthly values for each station are presented here, the results obtained are show hereunder.

Table 9. Mean monthly values for each station after data imputation (mm/d)

Month	X8934030	X8934113	X8934134	X8934169	X8934161	CHIRPS_13	X.8933026	X.8934016	X.8934023	X.8934030
1	0.87	0.76	1.37	3.02	1.78	1.92	1.25	1.08	1.63	0.99
2	1.26	0.77	1.83	2.45	1.41	2.60	2.15	1.44	1.31	1.62
3	2.96	1.68	3.41	2.63	3.75	4.94	3.19	3.06	3.09	3.25
4	4.04	4.17	6.33	2.69	5.93	8.77	4.94	6.31	5.95	6.29
5	3.81	2.77	6.48	3.86	4.88	8.04	3.36	5.69	4.39	4.95
6	1.34	2.34	2.72	2.95	2.24	4.71	1.90	4.13	2.71	2.14
7	1.16	1.54	2.21	1.39	1.93	3.96	1.37	4.32	2.75	1.44
8	2.03	1.99	2.01	2.34	2.84	4.61	0.94	6.01	2.22	4.71
9	1.91	2.81	3.95	2.36	3.88	4.83	1.30	4.86	3.48	5.06
10	2.85	2.82	3.73	1.41	4.34	5.87	2.45	3.86	3.91	5.41
11	3.04	2.76	2.99	1.99	3.74	5.32	1.77	3.00	2.62	3.55
12	2.15	1.13	1.83	1.49	1.87	2.93	1.28	1.58	1.60	2.44

Month	X.8934060	X.8934061	X.8934078	X.8934096	X.8934098	X.8934105	X.8934116	X.8934119	X.8934134	X.8934143
1	1.90	2.00	2.89	2.43	1.57	1.81	1.76	2.19	1.54	0.80
2	1.82	2.37	3.77	3.21	2.35	1.90	1.53	2.45	1.99	1.08
3	3.72	4.39	4.55	6.61	4.09	4.14	5.09	3.85	3.94	4.01
4	5.62	9.18	8.36	8.75	6.03	8.57	7.85	8.10	7.18	7.25
5	7.66	8.07	8.86	8.24	8.00	7.37	8.04	8.33	6.97	4.97
6	4.73	6.91	5.48	5.64	2.91	2.96	2.29	4.85	3.02	2.45
7	2.45	5.58	5.51	5.98	3.09	2.49	3.02	4.30	2.94	1.78
8	3.16	8.54	8.36	7.10	3.24	3.42	3.15	4.95	3.37	3.73
9	4.22	6.16	7.35	5.96	4.24	4.84	3.41	4.97	4.26	3.55
10	5.47	5.72	5.21	5.18	5.10	4.99	4.93	4.75	4.84	4.49
11	4.84	4.37	2.69	6.74	5.05	5.52	3.91	4.74	3.78	3.45
12	2.29	2.52	2.11	4.16	1.30	2.76	1.90	1.82	1.89	1.02

Month	X.8934155	X.8934156	X.8934183	X.8934191
1	2.07	2.31	1.71	1.65
2	2.98	2.27	1.99	1.34
3	4.65	4.74	4.44	3.05
4	6.26	9.47	8.08	4.29
5	7.48	8.79	8.07	3.34
6	3.79	3.67	5.25	1.25
7	2.92	3.36	4.00	1.07
8	2.87	3.52	4.49	1.41
9	3.72	4.73	5.43	1.80
10	4.68	6.09	5.73	2.27
11	5.57	5.19	5.48	3.06
12	1.94	3.04	2.23	1.78

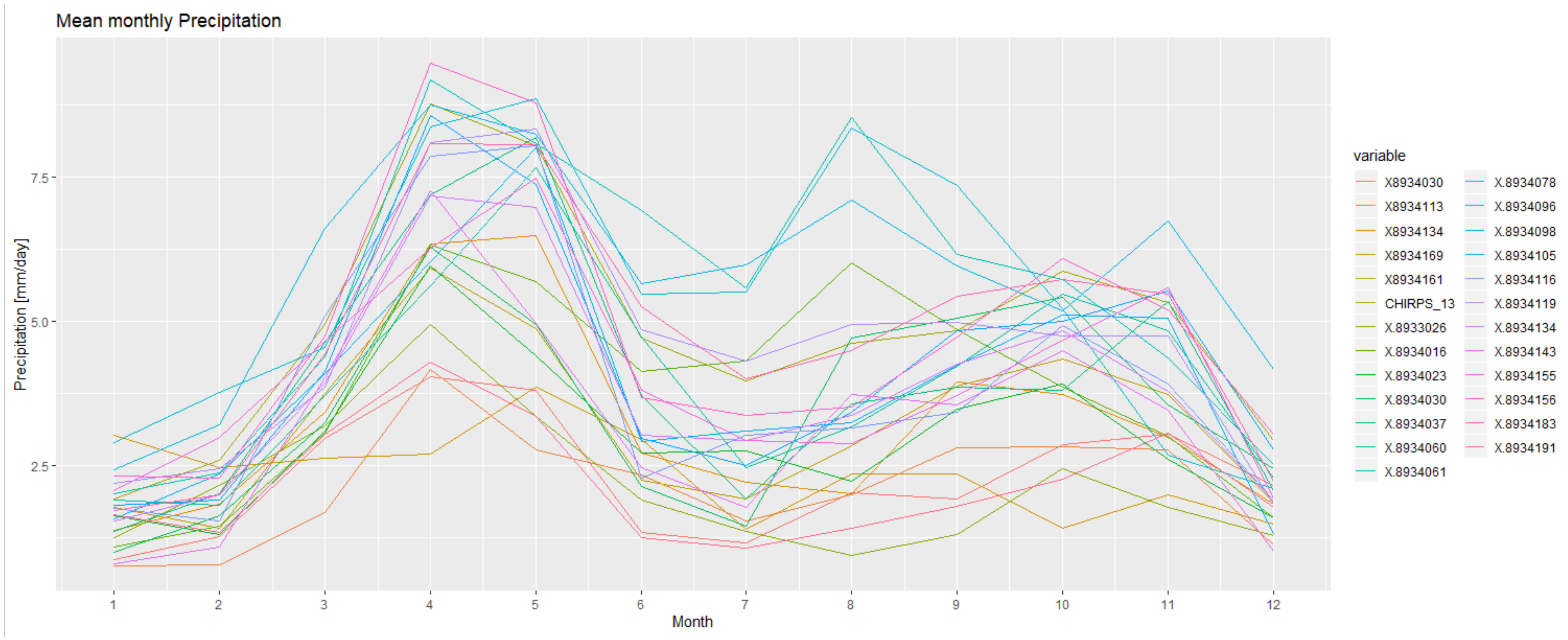


Figure 11. Mean monthly precipitation

Table 10. Mean annual values (mm/d) for each station after data imputation

Year	X8934030	X8934113	X8934134	X8934169	X8934161	CHIRPS_13	X.8933026	X.8934016	X.8934023	X.8934030	X.8934037
1980	2.93	0.91	3.65	0.92	3.63	1.11	2.49	3.42	3.61	2.93	3.09
1981	2.61	1.21	4.67	1.70	4.05	4.78	2.68	4.62	3.70	2.65	3.55
1982	1.91	1.53	4.94	2.18	5.63	5.32	2.58	4.58	5.02	3.12	3.54
1983	1.59	1.29	4.22	1.85	4.71	4.42	2.37	4.51	4.59	2.64	2.49
1984	0.66	1.14	3.96	1.19	4.53	3.94	2.92	3.51	3.77	1.93	2.28
1985	3.65	0.63	5.06	2.08	6.31	4.78	2.43	4.33	4.93	3.65	3.98
1986	3.53	3.50	3.54	1.77	5.90	4.37	3.32	2.86	4.06	3.53	3.14
1987	3.80	3.61	4.48	1.84	5.08	4.65	2.64	4.34	4.10	3.80	3.75
1988	2.43	3.77	4.57	2.52	6.20	5.93	3.41	5.01	4.42	4.03	4.62
1989	1.97	3.82	2.30	1.86	1.71	4.58	1.58	4.38	3.39	3.00	3.48
1990	2.20	4.38	2.29	2.17	1.91	5.04	2.30	3.95	0.66	3.80	3.84
1991	2.10	3.52	2.23	2.17	2.03	5.15	2.67	4.13	0.33	3.55	4.17
1992	1.83	2.06	1.68	1.88	1.44	4.62	1.46	3.53	1.02	2.86	4.63
1993	1.41	1.03	2.11	1.82	0.68	4.11	1.44	3.32	2.62	2.14	3.94
1994	1.63	2.70	3.05	1.93	1.86	5.17	1.58	3.41	5.33	4.08	6.64
1995	1.26	4.99	2.68	1.83	0.13	5.21	1.54	3.07	2.45	3.05	4.86
1996	1.72	4.07	5.76	2.26	4.22	5.26	2.54	2.05	4.30	3.42	4.84
1997	2.07	4.94	5.07	1.94	2.03	4.92	2.04	2.45	3.65	3.02	5.10
1998	2.11	2.94	5.65	2.23	1.89	5.02	2.08	3.95	3.97	3.32	4.85
1999	2.27	3.79	5.54	2.06	2.21	5.06	2.43	3.15	4.31	3.29	5.26
2000	1.62	0.22	2.11	1.60	1.09	4.09	1.44	3.29	1.63	2.63	2.51
2001	1.98	0.14	1.39	2.25	1.62	5.20	1.37	3.37	1.30	3.00	3.47
2002	2.25	1.68	2.76	2.10	4.58	5.31	2.12	3.30	2.90	3.18	3.73
2003	1.96	1.49	2.52	2.18	4.86	4.98	2.00	3.42	2.60	2.20	3.46
2004	1.78	1.57	2.29	2.01	4.92	4.26	1.82	3.32	2.48	2.29	3.01
2005	1.77	1.29	2.18	1.92	3.98	4.19	1.68	3.16	2.22	3.19	2.98
2006	2.75	1.97	3.19	2.75	5.53	6.43	2.44	4.54	2.94	4.28	4.67
2007	1.97	1.41	2.55	2.26	3.21	5.29	1.81	3.47	2.61	2.25	3.55

Year	X8934030	X8934113	X8934134	X8934169	X8934161	CHIRPS_13	X.8933026	X.8934016	X.8934023	X.8934030	X.8934037
2008	1.78	1.22	2.39	1.86	3.23	4.85	1.76	3.03	2.07	5.14	3.23
2009	1.68	1.25	2.35	1.57	0.00	4.75	1.49	4.07	1.86	4.72	3.33
2010	1.64	1.38	2.82	2.07	0.13	5.57	1.84	5.38	2.06	4.78	3.86
2011	4.14	1.49	2.65	1.93	4.86	5.52	2.59	5.06	2.70	4.99	4.57
2012	5.31	1.75	2.96	4.93	4.51	5.89	2.50	4.30	3.13	6.98	3.99
2013	4.27	1.56	2.61	4.75	4.66	4.95	2.56	4.41	2.73	4.46	5.31
2014	2.14	1.50	2.65	3.60	3.89	5.41	2.26	4.04	2.56	3.21	3.80
2015	2.76	2.14	3.21	5.88	2.84	5.75	2.15	4.11	2.97	4.76	4.75
2016	1.61	1.06	2.12	4.42	0.64	4.28	1.59	3.16	1.55	2.51	4.31
2017	1.94	1.97	2.98	4.20	1.70	5.36	1.90	3.99	2.56	4.38	4.42

Year	X.8934060	X.8934061	X.8934078	X.8934096	X.8934098	X.8934105	X.8934116	X.8934119	X.8934134	X.8934143	X.8934155
1980	3.85	4.65	4.16	4.38	3.38	3.01	3.28	3.56	3.65	1.77	2.83
1981	3.71	6.11	6.10	9.53	4.56	4.07	4.66	3.37	4.70	2.82	4.71
1982	4.58	8.68	5.11	6.15	4.68	5.02	4.37	4.76	4.90	2.99	3.97
1983	4.56	7.13	5.18	5.26	4.40	4.65	3.60	4.20	4.22	2.67	3.75
1984	3.83	4.32	3.46	4.48	3.30	4.41	3.67	3.81	3.96	3.07	3.37
1985	3.86	5.85	3.28	6.05	3.93	3.35	4.65	5.60	5.09	3.62	3.92
1986	3.38	4.54	4.05	4.89	3.91	3.01	4.10	3.95	3.51	3.21	3.47
1987	3.99	5.90	4.41	5.49	4.17	4.68	4.24	5.21	4.48	3.76	3.92
1988	4.17	8.12	5.27	7.13	4.54	4.77	4.98	4.31	4.43	3.77	4.39
1989	3.76	4.71	4.98	5.47	4.86	4.56	3.49	4.98	3.50	2.65	3.75
1990	3.68	5.30	5.61	5.50	4.58	4.46	3.29	3.99	4.16	4.58	3.15
1991	4.15	4.57	4.35	5.81	4.47	4.10	4.90	4.00	2.80	3.67	4.27
1992	3.58	4.67	5.54	4.79	4.85	3.99	4.34	4.08	2.51	3.38	3.56
1993	3.87	4.13	4.66	4.11	4.40	3.29	3.05	3.97	2.76	2.34	3.20
1994	4.59	6.46	6.50	5.35	5.91	4.88	4.05	5.55	3.00	2.71	4.42
1995	4.82	5.45	2.85	5.50	4.59	4.27	3.69	5.69	2.85	3.30	3.81
1996	5.00	4.76	4.21	5.73	3.91	4.56	4.79	4.18	5.82	3.08	4.10

Year	X.8934060	X.8934061	X.8934078	X.8934096	X.8934098	X.8934105	X.8934116	X.8934119	X.8934134	X.8934143	X.8934155
1997	4.79	4.90	7.22	4.44	4.04	3.64	4.13	3.67	5.07	2.74	3.58
1998	3.94	5.09	6.23	5.24	4.83	3.75	4.42	3.88	5.54	3.36	3.75
1999	5.17	5.67	5.68	4.79	4.78	3.37	4.43	3.56	5.55	3.62	3.78
2000	4.21	4.46	3.48	3.86	2.14	2.56	3.09	4.38	2.95	2.22	2.90
2001	4.53	5.45	4.58	5.62	2.39	2.77	3.03	5.05	2.42	2.83	3.79
2002	3.72	5.18	5.71	5.88	3.44	5.18	3.56	4.54	2.69	3.21	4.09
2003	3.20	5.60	7.10	5.78	3.25	5.87	4.00	4.19	4.03	2.96	3.75
2004	2.70	4.70	4.52	4.44	2.86	4.59	3.02	3.90	3.15	2.74	3.27
2005	2.67	4.95	4.47	4.72	3.05	3.93	3.09	3.58	2.27	2.80	3.26
2006	4.37	6.47	6.04	7.01	4.51	5.71	4.07	5.51	3.72	4.01	4.97
2007	3.14	5.45	6.24	5.53	3.48	4.94	3.73	5.27	4.14	3.21	3.72
2008	2.93	5.26	5.07	5.40	2.94	5.29	3.39	4.70	4.87	2.87	3.82
2009	2.95	5.03	4.43	4.96	2.81	5.05	3.08	4.23	4.13	2.75	3.35
2010	3.56	6.52	5.62	5.96	3.46	3.67	3.76	5.12	5.58	3.16	4.06
2011	3.60	5.46	6.30	5.51	3.45	5.70	4.72	4.91	4.62	3.90	4.23
2012	5.06	5.06	8.35	6.14	4.77	3.76	3.58	6.86	3.77	5.92	5.44
2013	5.10	5.63	7.56	12.52	4.40	4.14	4.84	6.43	2.70	3.55	6.63
2014	3.83	5.98	4.99	10.14	3.23	4.12	3.89	4.84	2.76	3.49	5.41
2015	4.32	5.86	10.38	6.10	2.95	5.16	5.40	5.07	3.35	3.22	6.15
2016	4.06	5.02	6.11	5.35	2.97	2.99	2.65	4.88	2.22	2.20	5.39
2017	4.63	5.68	6.67	7.05	4.75	3.68	3.94	5.49	3.08	4.22	5.04

Year	X.8934156	X.8934183	X.8934191
1980	4.10	3.69	0.58
1981	4.22	4.47	1.96
1982	7.09	5.75	2.26
1983	6.07	5.01	1.77
1984	6.16	4.64	2.29
1985	5.87	5.86	2.36
1986	5.23	4.24	2.36
1987	5.64	4.47	1.75
1988	5.75	4.78	2.23
1989	3.77	4.11	1.84
1990	4.47	4.78	2.37
1991	3.95	4.35	2.35
1992	3.99	4.60	1.63
1993	3.70	4.82	1.90
1994	4.42	4.88	2.16
1995	3.77	5.18	2.16
1996	4.69	5.07	2.37
1997	4.93	4.65	2.48
1998	4.58	4.99	2.70

Year	X.8934156	X.8934183	X.8934191
1999	5.91	4.70	2.24
2000	2.94	3.38	1.86
2001	3.33	5.19	1.92
2002	5.65	4.82	2.43
2003	4.96	4.47	1.90
2004	4.06	4.30	1.40
2005	3.56	3.95	1.68
2006	5.31	6.04	2.85
2007	6.92	5.61	1.46
2008	5.62	4.82	2.83
2009	5.11	4.65	2.42
2010	4.86	5.06	3.00
2011	4.84	5.79	3.33
2012	4.67	4.64	2.35
2013	5.05	4.96	2.61
2014	4.54	4.58	2.34
2015	4.86	5.02	3.14
2016	2.84	3.58	1.92
2017	3.97	4.51	2.23

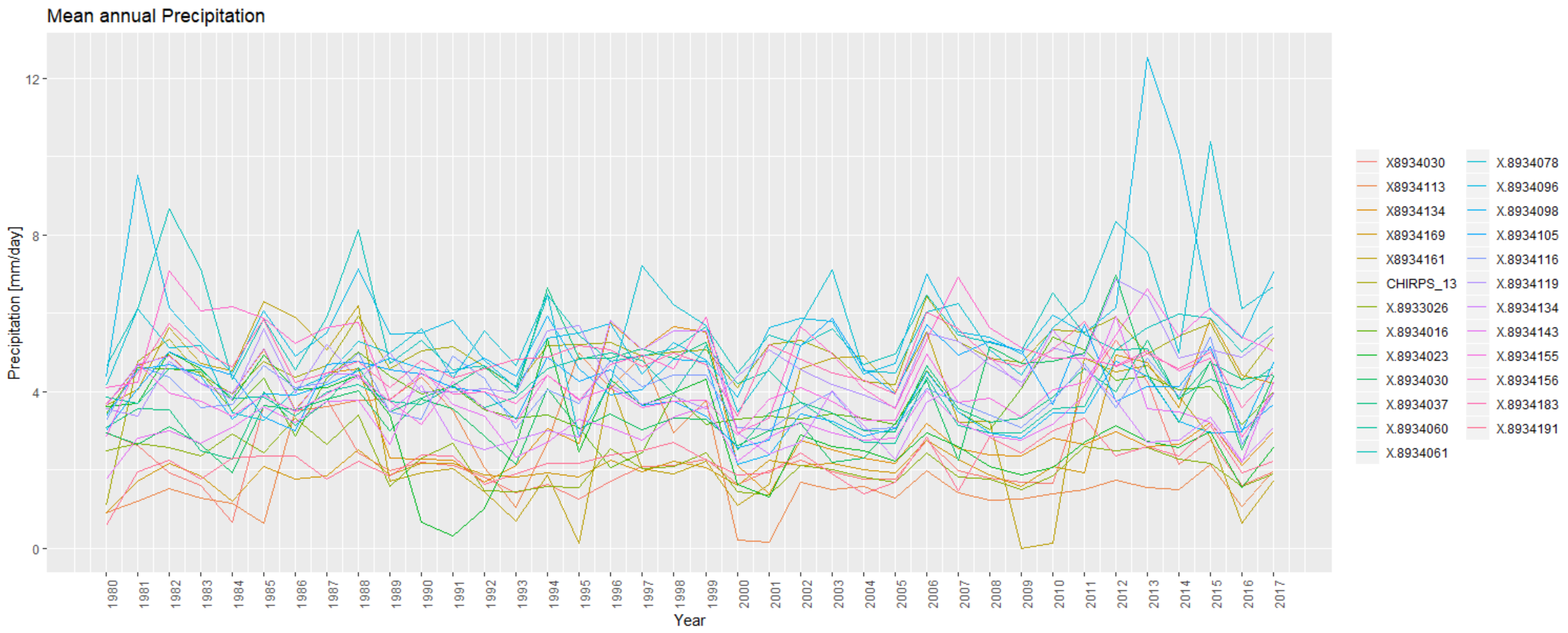


Figure 12. Mean annual precipitation f

Table 11. Annual precipitation sums for each station after data imputation (mm/a)

Year	X8934030	X8934113	X8934134	X8934169	X8934161	CHIRPS_13	X.8933026	X.8934016	X.8934023	X.8934030	X.8934037
1980	1074.10	333.75	1337.10	335.07	1328.50	404.45	909.80	1250.76	1322.90	1074.10	1130.77
1981	954.00	440.41	1703.30	619.81	1477.80	1743.26	977.70	1686.60	1350.80	966.23	1296.82
1982	697.47	559.51	1801.80	794.65	2053.60	1942.51	940.60	1671.30	1833.42	1138.00	1291.70
1983	581.11	469.87	1539.10	674.92	1720.20	1614.06	863.30	1644.70	1674.06	963.76	908.63
1984	242.37	417.42	1449.40	435.19	1659.20	1442.05	1069.50	1283.80	1379.90	704.79	834.10
1985	1331.30	231.10	1845.30	758.46	2301.60	1743.30	887.70	1579.70	1800.50	1332.60	1453.21
1986	1288.80	1276.50	1293.70	646.28	2154.80	1596.46	1212.17	1045.30	1480.50	1287.50	1145.40
1987	1386.70	1319.30	1633.70	671.05	1853.90	1697.50	964.92	1585.60	1494.90	1386.70	1368.00
1988	889.53	1379.00	1674.10	921.45	2270.80	2171.17	1248.47	1833.50	1616.80	1474.93	1691.05
1989	719.80	1393.00	839.00	679.61	622.34	1670.88	576.12	1599.70	1237.20	1096.35	1270.50
1990	803.62	1598.80	834.35	791.88	697.79	1839.50	839.33	1440.54	240.91	1385.32	1402.40
1991	767.82	1283.50	815.61	790.34	739.93	1878.66	972.75	1509.10	118.90	1295.90	1520.30
1992	669.09	753.40	616.29	688.97	526.73	1692.05	536.02	1290.60	374.17	1044.97	1693.18
1993	512.97	375.52	770.55	663.49	248.80	1501.48	525.75	1212.80	957.57	781.89	1439.02
1994	595.69	987.00	1112.50	702.72	679.90	1885.30	577.01	1244.00	1944.90	1487.90	2423.00
1995	460.88	1820.30	976.50	667.36	45.90	1902.79	562.80	1119.00	893.84	1113.89	1774.98
1996	628.31	1488.10	2108.20	825.37	1543.59	1926.26	927.90	750.10	1574.80	1251.76	1772.90
1997	757.26	1802.80	1850.50	708.14	742.12	1794.62	746.07	892.80	1331.29	1103.80	1860.93
1998	770.37	1072.90	2061.80	813.27	688.34	1833.63	760.38	1441.61	1449.46	1213.29	1769.55
1999	828.03	1381.70	2020.40	752.58	805.38	1847.33	886.18	1149.50	1574.73	1199.98	1919.13
2000	591.69	80.10	771.78	584.12	400.06	1495.53	526.75	1205.10	596.46	962.11	919.07
2001	722.39	52.35	508.27	819.74	591.49	1898.56	500.75	1231.04	475.31	1094.52	1265.11
2002	821.27	612.73	1008.78	767.55	1672.92	1938.27	772.40	1205.10	1058.88	1160.42	1360.57
2003	715.49	544.72	918.95	794.65	1774.10	1815.88	728.83	1248.10	947.35	804.28	1262.79
2004	649.73	573.44	838.74	735.73	1798.90	1557.54	665.75	1214.98	909.42	839.89	1100.75
2005	645.13	470.86	795.01	702.24	1453.70	1529.14	612.75	1152.83	810.83	1164.43	1087.17
2006	1004.33	720.37	1165.86	1002.87	2017.30	2346.01	888.89	1655.87	1074.13	1561.42	1705.44

Year	X8934030	X8934113	X8934134	X8934169	X8934161	CHIRPS_13	X.8933026	X.8934016	X.8934023	X.8934030	X.8934037
2007	719.43	516.33	929.60	825.49	1171.60	1929.63	659.09	1267.00	950.95	822.90	1296.27
2008	651.68	447.10	873.21	680.95	1182.40	1776.08	643.51	1108.00	759.34	1879.60	1180.61
2009	613.16	457.97	858.78	573.39	0.00	1732.98	544.34	1485.66	679.81	1721.00	1216.34
2010	600.26	505.46	1030.32	757.24	46.90	2033.27	671.98	1965.50	753.51	1744.37	1408.92
2011	1510.80	543.05	965.70	704.56	1772.60	2013.96	945.46	1846.53	985.26	1822.65	1668.46
2012	1942.50	638.93	1084.06	1803.90	1652.10	2157.00	914.48	1572.22	1146.83	2556.38	1458.74
2013	1557.50	567.72	950.93	1734.90	1700.40	1807.24	934.80	1608.00	997.27	1628.11	1937.87
2014	780.17	548.35	967.59	1312.70	1418.30	1973.08	826.02	1473.52	935.67	1173.30	1385.62
2015	1008.79	779.91	1171.99	2144.70	1035.30	2098.25	785.81	1501.46	1085.37	1737.38	1732.90
2016	589.84	387.57	774.57	1615.90	233.30	1567.23	581.05	1158.36	568.83	920.12	1579.11
2017	708.88	719.49	1086.20	1532.80	619.88	1957.16	692.07	1455.17	933.54	1600.32	1614.07

Year	X.8934060	X.8934061	X.8934078	X.8934096	X.8934098	X.8934105	X.8934116	X.8934119	X.8934134	X.8934143	X.8934155
1980	1410.10	1700.60	1521.50	1602.60	1238.62	1101.37	1200.32	1304.60	1336.40	648.95	1034.75
1981	1353.40	2231.40	2227.00	3477.70	1664.90	1486.62	1701.22	1229.70	1715.90	1030.46	1720.90
1982	1673.50	3167.30	1866.70	2244.80	1707.00	1832.38	1595.72	1737.40	1789.20	1092.88	1449.60
1983	1666.20	2603.00	1891.60	1920.50	1604.20	1698.10	1313.24	1531.63	1539.60	976.08	1369.60
1984	1401.00	1579.60	1268.14	1637.90	1206.80	1614.40	1343.00	1395.30	1448.90	1124.10	1233.70
1985	1410.00	2135.50	1197.20	2210.00	1433.80	1222.00	1698.76	2043.10	1858.50	1321.98	1432.20
1986	1234.30	1658.00	1478.10	1785.90	1427.30	1099.18	1497.15	1443.50	1280.50	1171.12	1266.20
1987	1455.70	2151.91	1608.77	2003.70	1523.10	1709.20	1546.04	1899.95	1633.70	1372.19	1432.60
1988	1525.10	2970.72	1928.50	2608.50	1662.50	1745.70	1820.97	1576.04	1620.30	1380.15	1605.99
1989	1373.70	1719.40	1819.20	1996.00	1774.30	1663.60	1273.60	1816.77	1276.00	965.49	1368.30
1990	1343.71	1934.01	2048.71	2006.60	1671.78	1626.35	1199.74	1454.56	1517.15	1672.49	1149.94
1991	1513.65	1667.10	1588.29	2122.00	1631.00	1495.30	1788.70	1461.24	1020.60	1338.24	1558.10
1992	1310.72	1709.30	2027.90	1754.00	1775.20	1460.24	1587.70	1493.91	919.10	1235.31	1301.30
1993	1411.80	1507.40	1702.18	1498.80	1605.00	1199.87	1112.14	1449.10	1008.20	853.57	1168.20

Year	X.8934060	X.8934061	X.8934078	X.8934096	X.8934098	X.8934105	X.8934116	X.8934119	X.8934134	X.8934143	X.8934155
1994	1673.81	2357.10	2373.25	1954.10	2158.70	1779.50	1476.47	2024.90	1093.70	988.93	1611.70
1995	1758.70	1989.60	1039.50	2007.00	1677.07	1558.70	1346.53	2077.90	1038.80	1206.11	1392.17
1996	1828.50	1743.10	1540.49	2095.60	1432.57	1669.22	1751.74	1528.80	2130.00	1126.48	1500.27
1997	1749.67	1788.50	2634.30	1618.90	1473.75	1328.95	1506.05	1338.60	1848.80	1001.76	1306.23
1998	1438.80	1857.20	2274.80	1914.00	1762.85	1369.90	1614.34	1416.70	2023.10	1225.47	1368.12
1999	1885.50	2069.50	2072.54	1749.70	1743.61	1230.50	1617.86	1301.00	2025.40	1321.65	1381.30
2000	1540.20	1633.20	1273.26	1414.20	781.54	935.50	1129.89	1604.00	1081.10	813.35	1060.86
2001	1653.20	1989.34	1670.07	2050.90	871.81	1011.00	1105.45	1844.90	881.90	1032.03	1383.75
2002	1356.35	1889.80	2084.90	2145.40	1255.06	1890.60	1301.14	1658.46	983.00	1173.09	1492.28
2003	1166.25	2042.54	2591.00	2110.30	1184.69	2142.61	1461.40	1529.45	1469.70	1078.96	1368.89
2004	986.55	1721.24	1654.99	1624.10	1045.89	1678.60	1106.65	1428.83	1151.27	1004.45	1196.02
2005	975.32	1807.63	1631.04	1723.90	1112.17	1436.05	1126.45	1305.16	829.95	1021.27	1190.28
2006	1596.54	2361.28	2205.17	2558.38	1647.03	2085.04	1484.97	2010.23	1358.93	1464.25	1814.03
2007	1145.56	1989.55	2278.40	2020.20	1269.56	1803.60	1360.11	1925.02	1511.40	1170.96	1357.27
2008	1073.42	1923.90	1854.94	1975.50	1075.14	1935.48	1242.14	1720.00	1781.60	1049.75	1398.42
2009	1076.35	1835.95	1615.68	1810.00	1025.39	1843.70	1125.60	1542.24	1507.31	1004.55	1224.10
2010	1300.56	2380.50	2052.34	2175.90	1261.70	1338.67	1372.34	1868.70	2037.00	1153.74	1481.24
2011	1315.31	1994.20	2299.55	2009.90	1259.63	2080.64	1721.78	1793.81	1686.80	1423.97	1542.58
2012	1852.91	1852.46	3057.81	2247.60	1747.33	1375.06	1312.10	2509.01	1378.44	2168.06	1990.25
2013	1860.96	2055.97	2759.60	4569.80	1607.82	1510.24	1766.00	2348.14	985.95	1296.38	2420.57
2014	1399.57	2181.61	1820.46	3701.39	1180.18	1502.58	1418.13	1766.63	1006.43	1273.66	1973.04
2015	1578.48	2140.71	3789.26	2226.20	1076.38	1882.17	1971.60	1849.10	1221.13	1175.90	2244.84
2016	1486.97	1836.38	2235.92	1958.30	1087.36	1094.11	969.45	1787.16	811.77	803.61	1974.49
2017	1689.88	2071.50	2433.36	2572.52	1734.87	1341.56	1437.83	2005.65	1123.72	1540.58	1840.95

Year	X.8934156	X.8934183	X.8934191
1980	1500.20	1349.90	211.58
1981	1542.10	1633.10	715.93
1982	2588.40	2098.80	824.42
1983	2215.30	1828.20	645.77
1984	2253.75	1699.57	839.60
1985	2143.26	2137.80	862.00
1986	1907.30	1546.02	861.40
1987	2059.18	1632.29	640.10
1988	2106.31	1750.59	816.92
1989	1377.18	1499.31	670.23
1990	1633.34	1746.51	864.35
1991	1440.90	1586.70	856.10
1992	1458.70	1682.10	595.10
1993	1351.55	1760.85	692.60
1994	1613.71	1779.56	790.02
1995	1374.58	1891.30	787.06
1996	1714.73	1855.51	868.36
1997	1799.23	1696.72	904.30
1998	1670.32	1822.70	985.30

Year	X.8934156	X.8934183	X.8934191
1999	2156.70	1716.02	817.31
2000	1077.60	1237.47	681.10
2001	1214.69	1894.22	701.10
2002	2063.90	1760.60	885.28
2003	1808.80	1630.50	694.80
2004	1485.93	1573.50	510.71
2005	1297.63	1440.68	612.42
2006	1937.27	2205.10	1039.57
2007	2525.20	2048.53	531.89
2008	2056.75	1763.36	1035.66
2009	1865.30	1695.60	884.86
2010	1775.30	1847.39	1095.80
2011	1766.70	2113.73	1215.00
2012	1709.43	1697.74	860.47
2013	1843.84	1810.40	951.33
2014	1655.29	1672.28	855.91
2015	1772.76	1833.25	1144.93
2016	1040.07	1311.91	704.22
2017	1450.79	1646.19	814.76

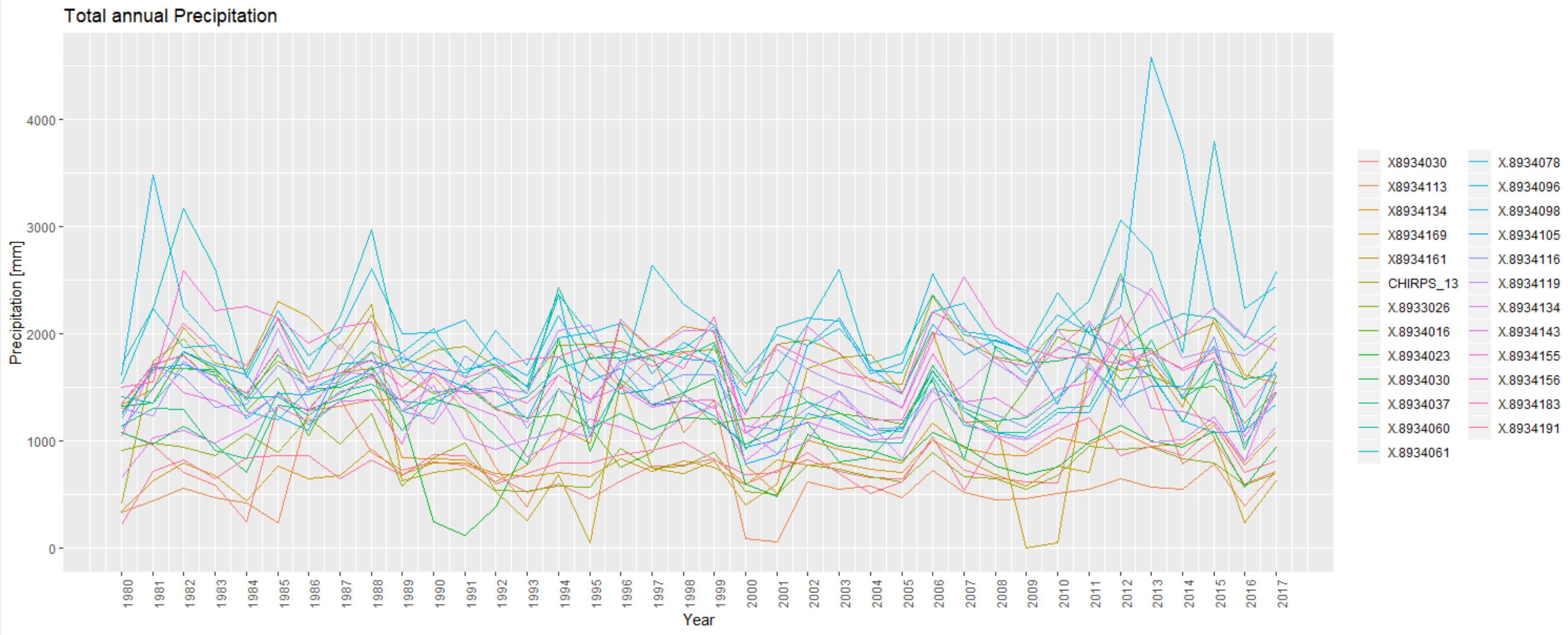


Figure 13. Annual precipitation sums per station

4.1.2 Pan Evapotranspiration data

For the application of the GR4J model, time series of potential evapotranspiration are needed. In our case, pan evaporation data is used, which is available from the KMD station in Kakamega, which is located around 50 km to the east of the Sio catchment.

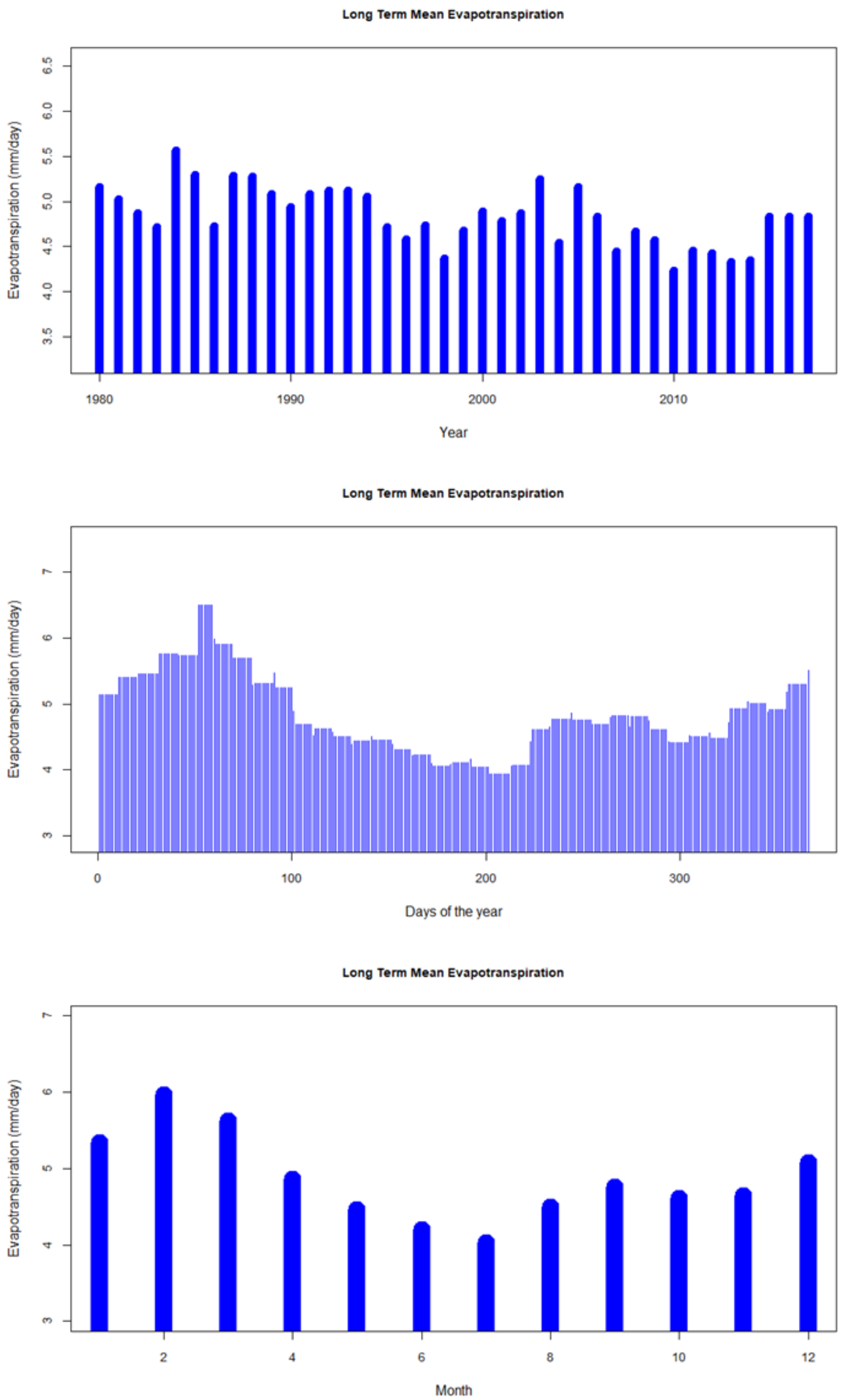


Figure 14. Long term mean values for annual, monthly and daily evapotranspiration (mm/d)

4.2 Spatial interpolation

Following the method explained in chapter 3.2, the Thiessen Polygons were calculated, through ArcMap, for the stations available in the whole area of study. As is shown in Figure 15, only 9 stations are taken into consideration for further calculations.

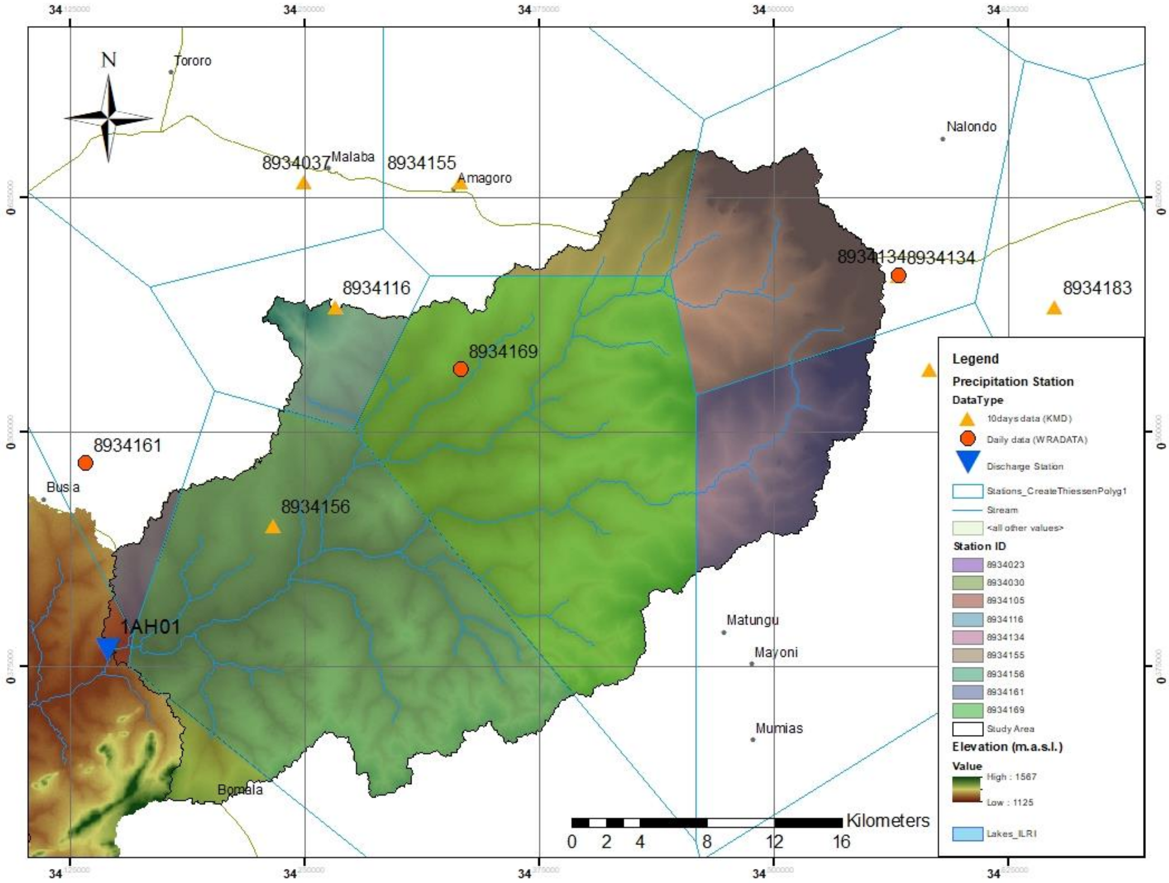


Figure 15. Thiessen polygons applied to the study area

Each station has a percentage of area from the whole catchment area (see Table 12), these percentages are multiplied with the rainfall time series, after the imputation process, to obtain one areal averaged precipitation for the sub-basin. The result of this procedure will end in a unique time series that is required as input for the hydrological model.

Table 12. Sub-basin area divided in corresponding Thiessen polygons

ID	Name	DataType	Area (km²)	%in basin area
8934030	Nangina Catholic Mission	daily_WRADATA	18.01	1.78%
8934134	Bungoma Water Supply	daily_WRADATA	127.29	12.59%
8934169	Kwangamor	daily_WRADATA	344.78	34.11%
8934161	Alupe Cotton Research Station	daily_WRADATA	12.68	1.25%
8934023	Sang'alo Institute Of Science & Technology	10days_KMD	89.37	8.84%
8934105	Busia Farmers Training Centre	10days_KMD	4.23	0.42%
8934116	Amukura Chief's Centre	10days_KMD	38.38	3.80%
8934155	Amagoro DO's Office	10days_KMD	33.36	3.30%
8934156	Nambale Agricultural Office	10days_KMD	342.81	33.91%
		∑area	**1 010.90**	**100%**

4.2.1 Basin precipitation

Based on the areal distribution obtained in the Thiessen polygons and the daily precipitation of the stations involved (See Table 12), the average precipitation inside the whole basin results in the following annual precipitation sums:

Table 13. Basin annual precipitation sums

Year	Annual sums(mm/a)	Year	Annual sums(mm/a)	Year	Annual sums(mm/a)
1980	1028.39	**1994**	1234.76	**2008**	1234.63
1981	1237.90	**1995**	1008.75	**2009**	1109.89
1982	1685.32	**1996**	1420.57	**2010**	1163.60
1983	1458.53	**1997**	1335.44	**2011**	1221.42
1984	1339.46	**1998**	1361.66	**2012**	1611.68
1985	1573.17	**1999**	1521.50	**2013**	1625.08
1986	1287.62	**2000**	810.93	**2014**	1370.37
1987	1426.37	**2001**	910.88	**2015**	1764.08
1988	1555.85	**2002**	1324.19	**2016**	1171.59
1989	1044.41	**2003**	1225.73	**2017**	1375.42
1990	1053.80	**2004**	1063.46	**Long-term mean annual**	**1272.56**
1991	1019.95	**2005**	969.09		
1992	968.22	**2006**	1408.84		
1993	964.39	**2007**	1470.53		

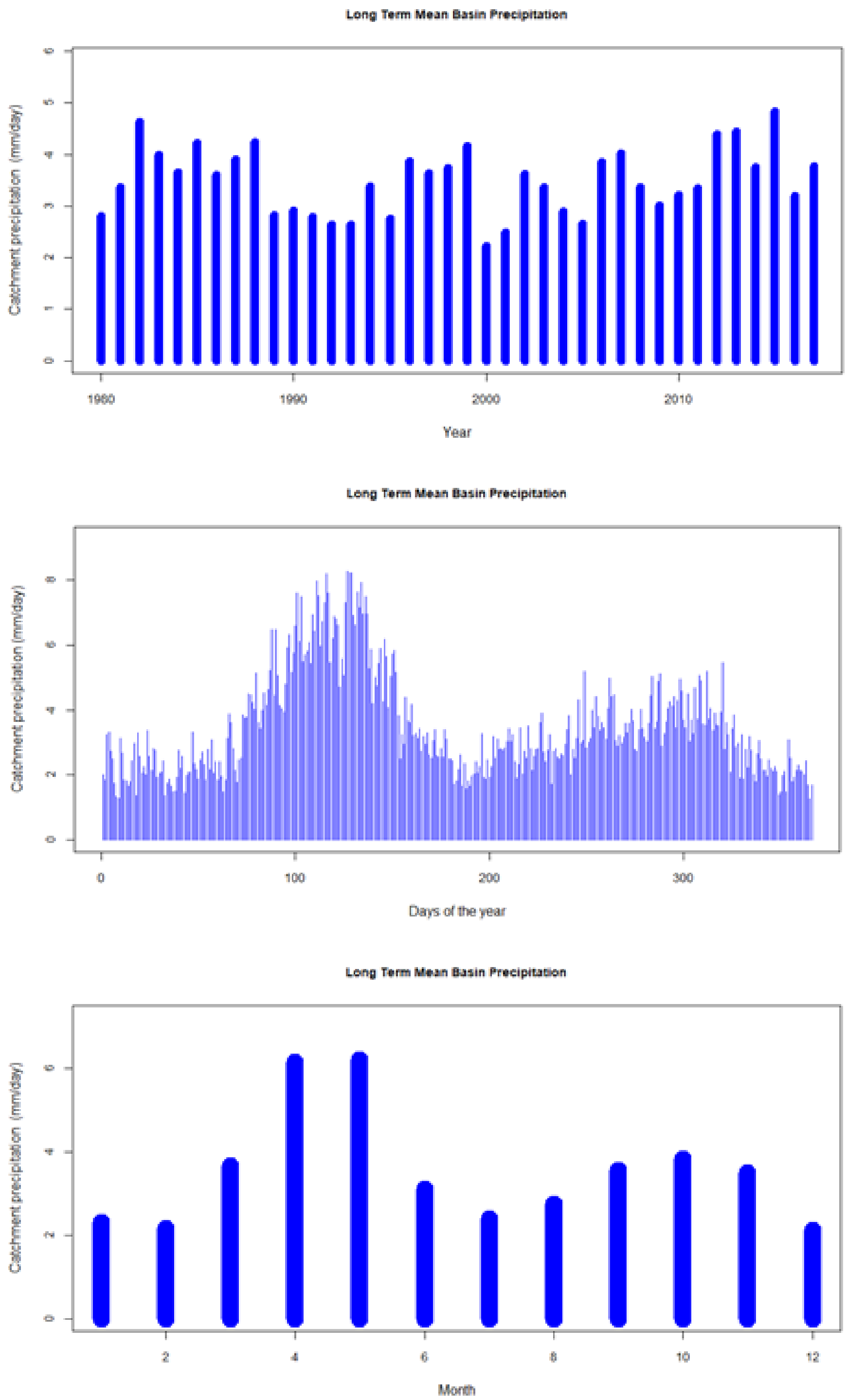


Figure 16. Long term mean values for annual, monthly and daily precipitation in the catchment

4.3 Hydrological Modelling

4.3.1 Input Data

The aerial averaged rainfall time series obtained (Figure 17),the pan evapotranspiration time series (Figure 18) and the observed flow (Figure 19) are required as input. In contrast to precipitation and evapotranspiration, observed flow can include NA values. All time series must be available for the same time period in our case for the period 1980-01-01 to 2017-12-32, a total of 37 years of data.

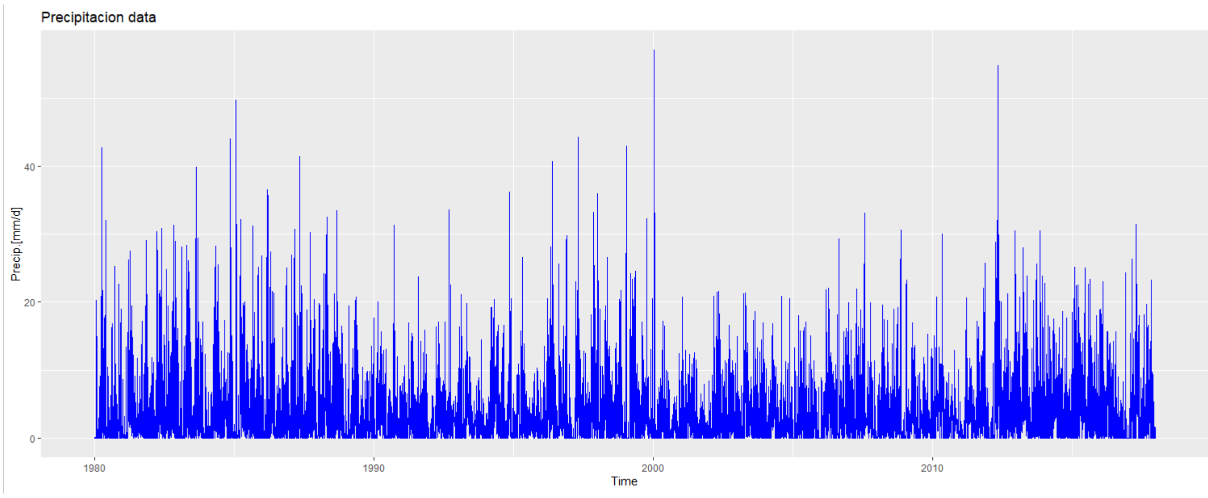


Figure 17. Daily precipitation time series for Sio River sub-basin

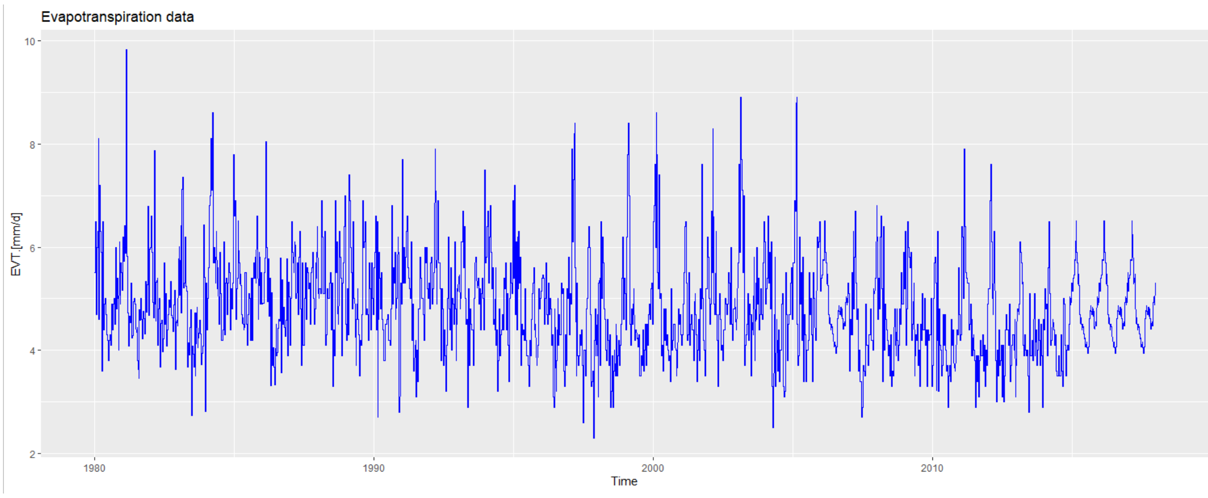


Figure 18. Pan evapotranspiration data used in this study originating from the pan measurements in Kakamega.

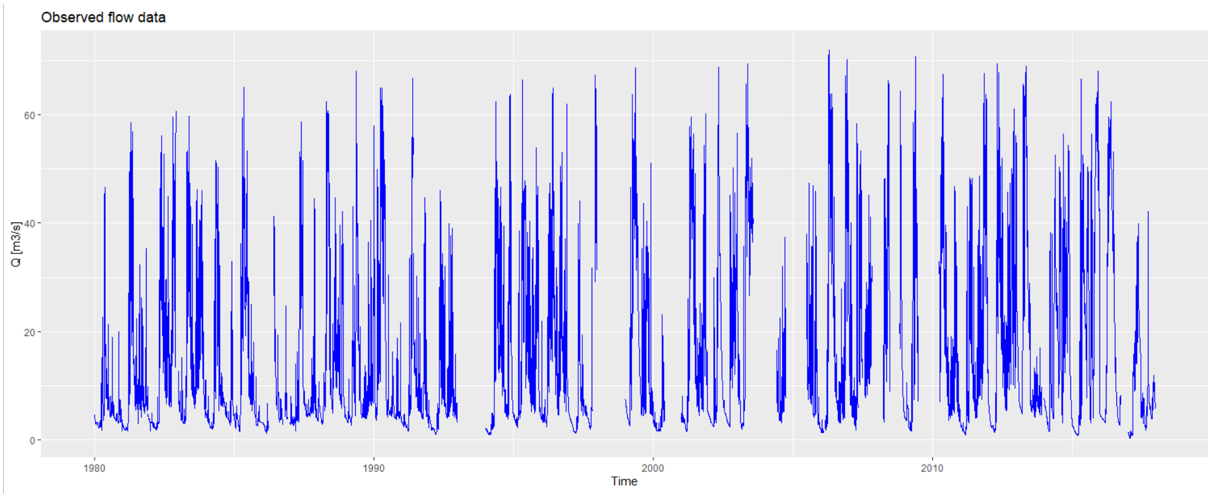


Figure 19. Observed flow in Sio River sub-basin

4.3.2 Model set Up

The model was configured based on three main aspects; (i) temporal extent, (ii) parameter definition and (iii) criterion selection. For the first aspect, 37 years of data are available. Of this 37 years, one year of data is used for model warm up (Table 14).

Table 14. Temporal settings

Period	Start date	End date	Days
Data Sio	1980-01-01	2017-12-31	13880
Model Run	1981-01-01	2017-12-31	13514
Model Warm-Up	1980-01-01	1980-12-31	366

In Chapter 3.3.1.1 the importance of the initial selection of the parameters and its relation to the 80% confidence interval suggested by Perrin et al. (2003) was mentioned, hence the first attempt of the model is defined with the parameters in Table 15. All 4 parameters are selected inside the recommended range

Table 15. Initial selection of parameters

Parameter	Description	Unit	Initial value	80% Confidence interval
x1	production store capacity	mm	70	100-1200
x2	intercatchment exchange coefficient	mm/d	-1	-5 – 3
x3	routing store capacity	mm	100	20-300
x4	unadjusted unit hydrograph time constant	days	1.5	1.1 - 2.9

4.3.3 Calibration results

The calibration function implemented in the R package "airGR" uses the Irstea procedure described by Michel (1991), which optimizes the error criterion selected for the objective function. According to the package description, "*The algorithm combines a global and a local approach. First, a screening is performed using either a rough predefined grid or a list of parameter sets. Then a steepest descent local search algorithm is performed, starting from the result of the screening procedure*. *For this search, since the ranges of parameter values can be quite different, simple mathematical transformations are applied to parameters to make them vary in a similar range and get a similar sensitivity to a predefined search step*."

All these candidates are tested and the best one is kept being the starting point for the next iteration. At the end of each iteration, the search step is either increased or decreased to adapt the progression speed. A composite step can occasionally be done. The calibration algorithm stops when the search step becomes smaller than a predefined threshold.

Table 16. Parameter value

				User defined	After calibration	
Coef	Description	Unit	Range	Initial value	NSE	KGE
x1	Capacity of production soil store (SMA)	mm	100-1200	70	396.52	177.83
x2	Water exchange coefficient	mm	-5 – 3	-1	2.79	2.46
x3	Capacity of the routing store	mm	20-300	100	26.14	24.15
x4	Time parameter for unit hydrographs	days	1.1 - 2.9	1.5	1.25	2.32

With this on mind, after the calibration was done, the parameters obtained, as expected, lay inside the provide range (Table 16), while the error obtained within the last iterations improved towards a better fit of the model. The calibration resulted in a poor NSE of 0.24 (Figure 20) and a KGE of 0.56 (Figure 21). Although the KGE had 0.56 as final value, Rogelis et al. (2016) suggested a "good" model when KGE is larger than 0.5.

This qualification of good or bad model goes further than the value obtained at the end of the calibration, for example in NSE model low flows are greatly overestimated and some periodical events peaks are completely ignored, while other peaks are considerable higher than the measured value. In KGE case, most of the periodical peaks and low flow values are more or less represented but there is a higher frequency of extreme values than what is observed with the measured values. Though many of the misestimation can be attributed to model performance, is also likely to have these discrepancies due to measurement errors.

Table 17. Error criterion before and after calibration with NSE and KGE eq.

	Error criterion based on the NSE formula	Error criterion based on the KGE formula
Initial	-0.216	0.112
After calibration	0.242	0.557

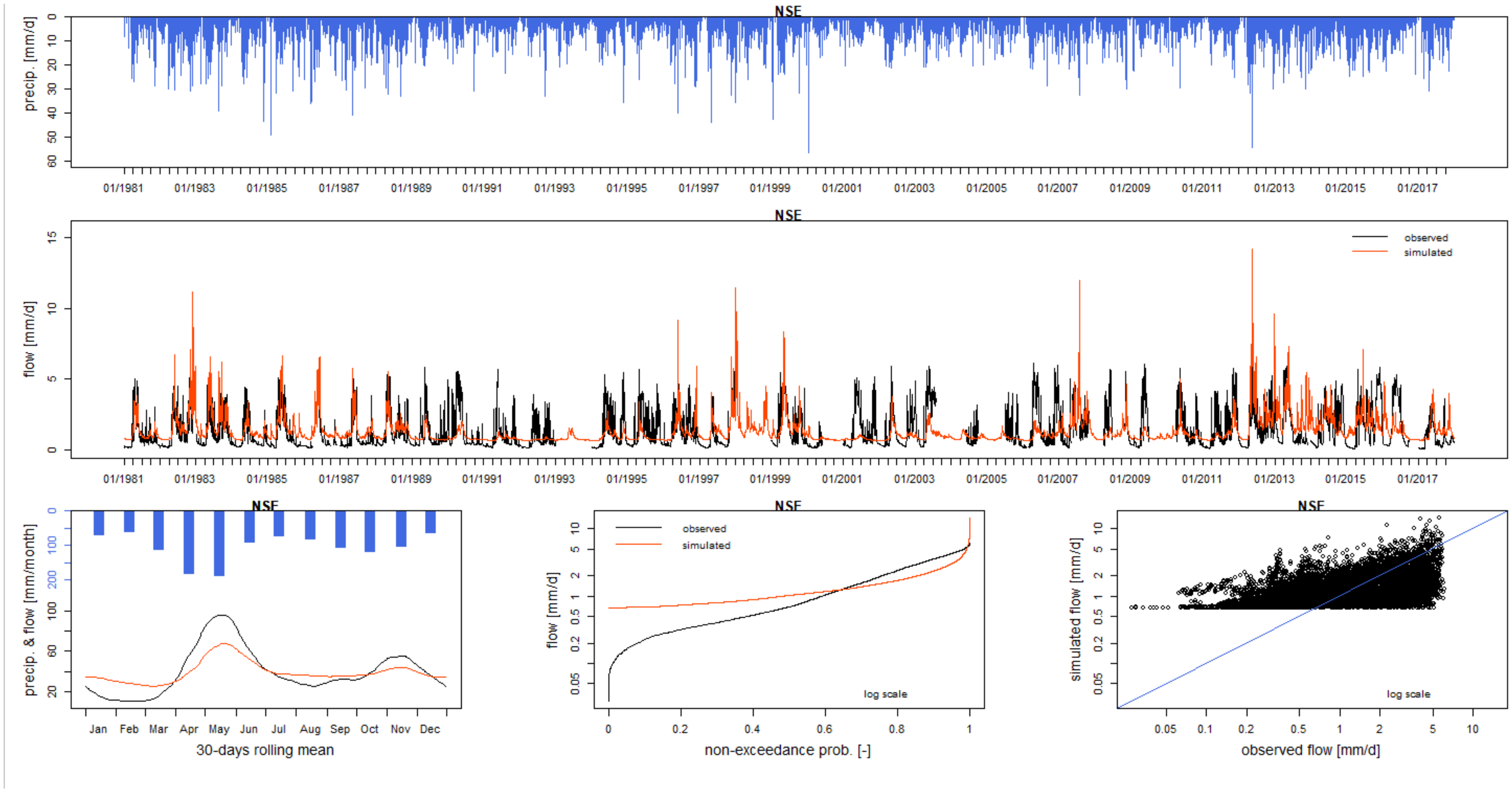


Figure 20. Model output after calibration based on the NSE.

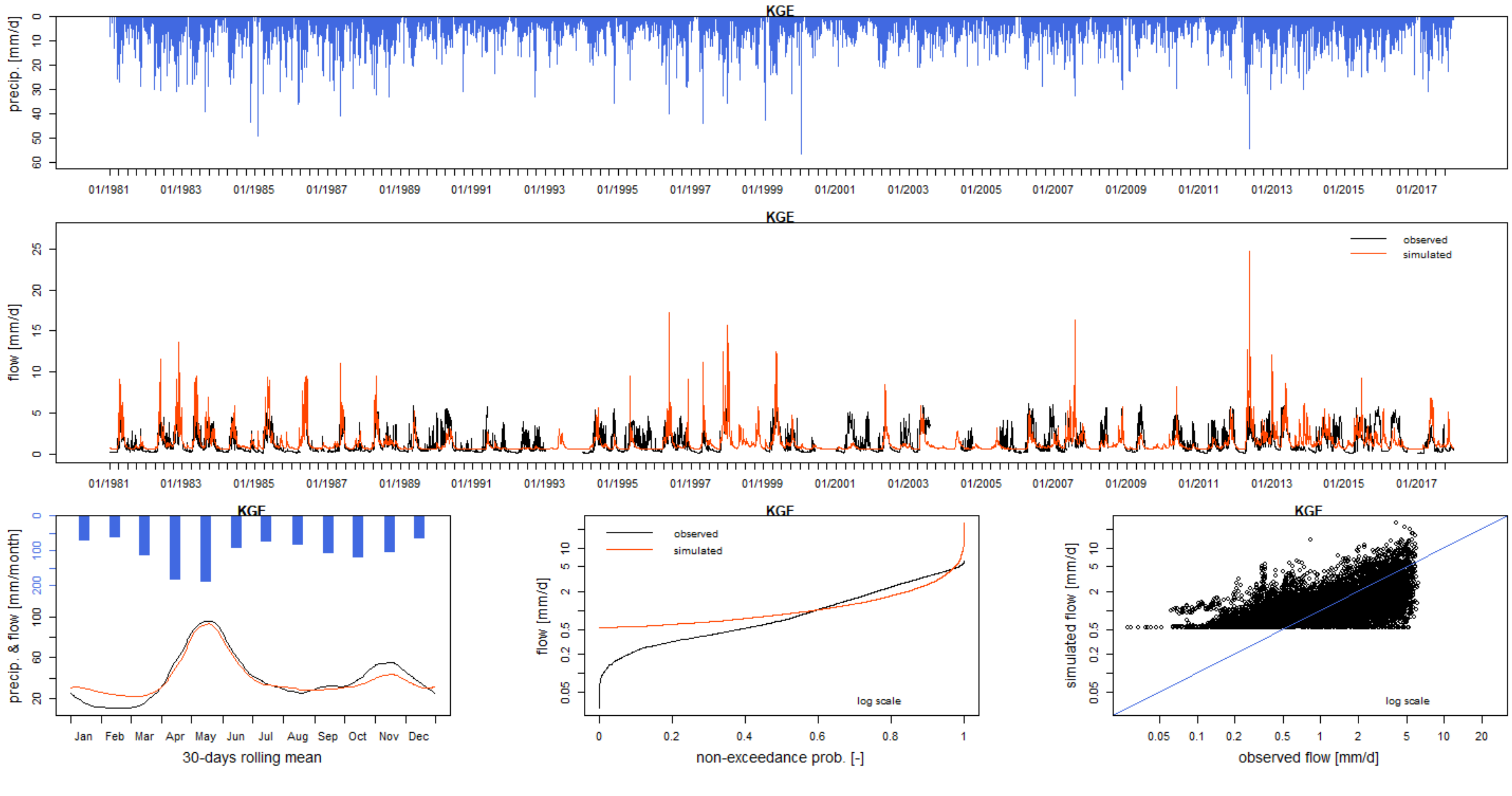


Figure 21. Model output after calibration based on the KGE

The final output of the hydrological model is a simulated flow, which corresponds to the solution that best fits to the observed flow, using the given data and the runoff model structure. In this case, the model calibrated using the KGE the with NSE as optimization functions. Therefore, two different optimal results are found (Figure 22).

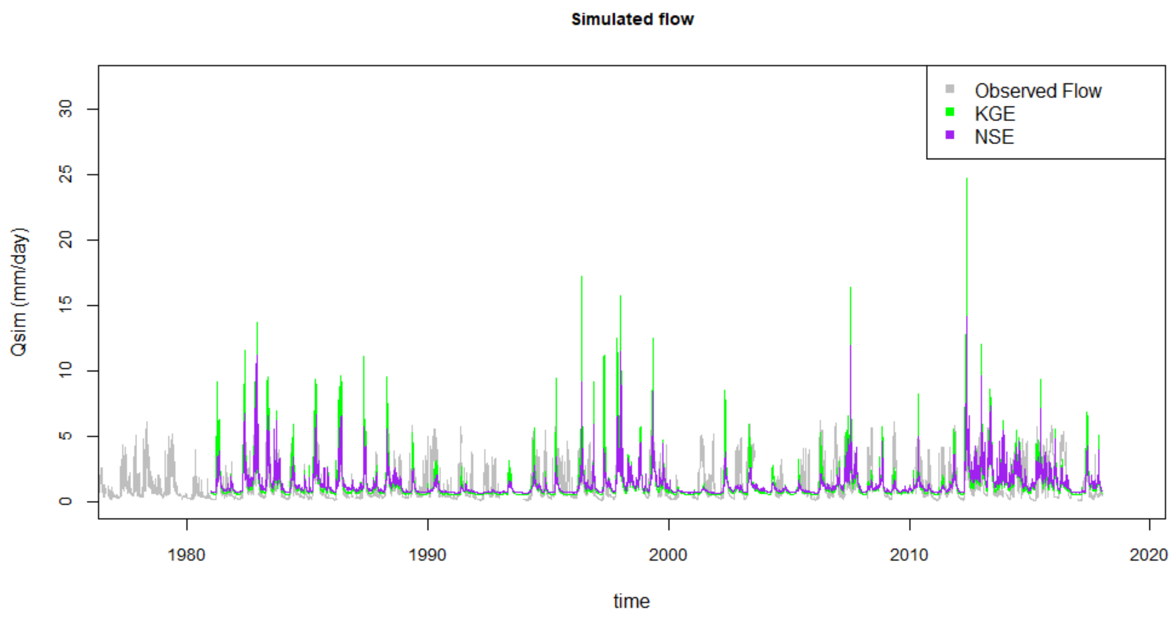


Figure 22. Observed streamflow and simulated streamflow with KGE and NSE

It can be observed that the NSE leads to higher baseflow values and lower peaks, while KGE generates higher flood peaks and lower baseflow.

The annual, monthly and daily long-term mean values for both simulated flows and observed flow are shown in the next figure. This should allow for a better understanding of the results and allow for a comparison.

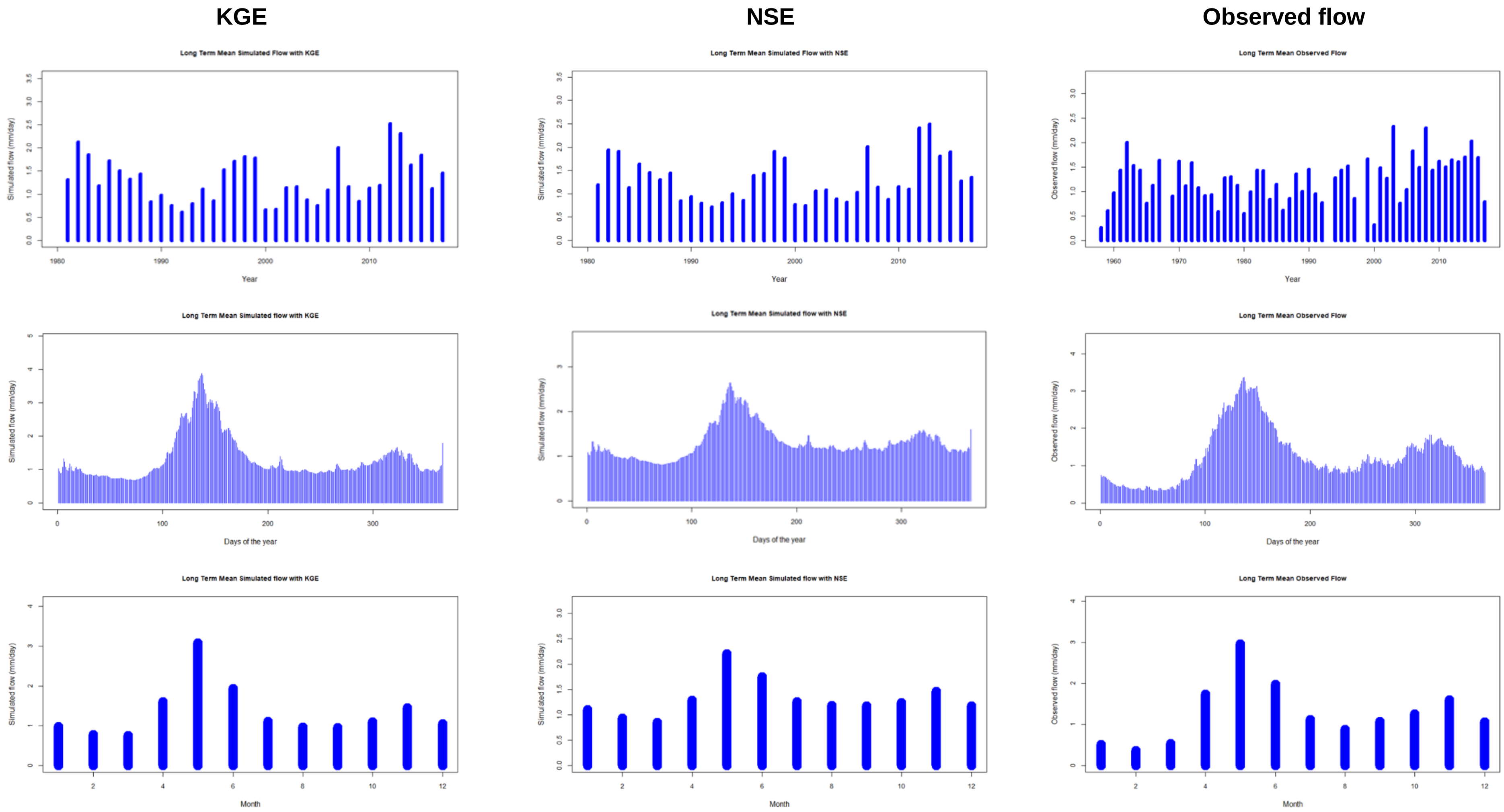


Figure 23. Long-term mean values for annual, monthly and daily flow simulated with KGE and NSE and for the observed streamflow

4.3.4 Model performance / discussion

The model can predict general trends of the flow, if it will increase or decrease, both with KGE and NSE. However, it is not accurate on the quantification of the flow, as there is a constant overestimation of the flow peaks (extreme values) and the base flow is often overestimated as well. Therefore, the model performance should be improved.

The model performance depends on several factors, including the model structure itself and how well it can represent reality, but also other variables such as the quality of the input data, and the parameter ranges used in the model.

In this case study, there is high uncertainty in the input data. The temporal and spatial interpolation allowed for filling the data gaps, but it is not possible to determine how close those interpolations are to the reality. Thus, the precipitation and evapotranspiration used as input parameters for the model are the best available.

If the input data are unprecise, when the hydrological model tries to fit the simulated flow to the measured one, the parameters which define the relationship between the natural processes might be unrealistic.

The hydrological model provides the expected relationship between potential and actual evapotranspiration, as actual evapotranspiration value provided by the model is lower than the measured pan evapotranspiration. This can be observed in Figure 24.

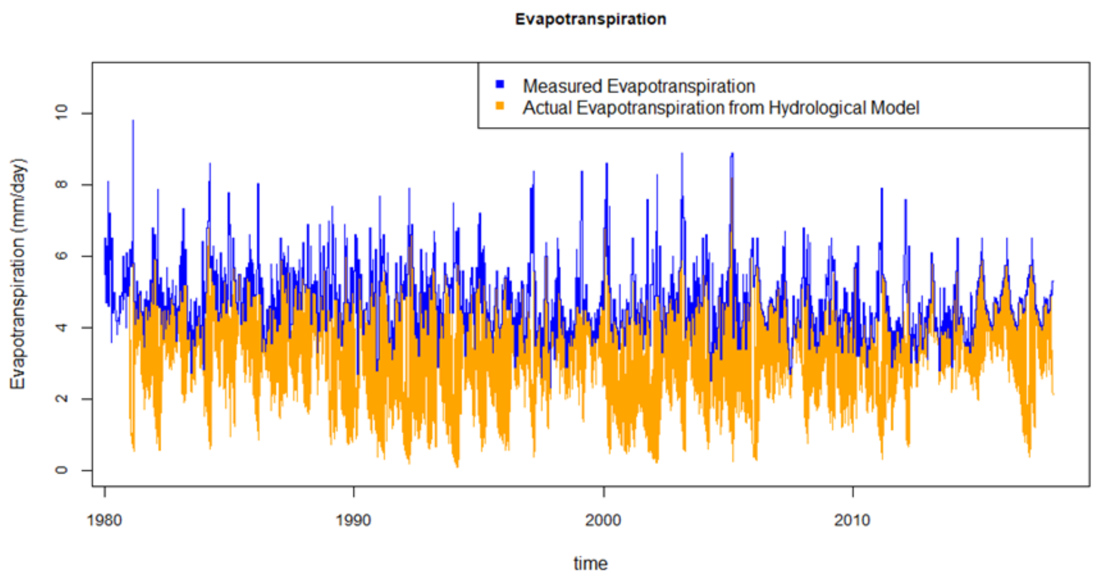


Figure 24. Pan evapotranspiration used as input and resulting actual evapotranspiration from the hydrological model.

An extension of the parameter range would allow finding new optimal solutions. Nonetheless, it is important to note that model parameters represent a natural process, and that the boundaries used are defined in literature. In this example, the parameter values tend to go to extremes, which leads to the idea that the absolute optimum is outside the parameter range. However, this can also mean that there is some natural process happening which is not considered or well defined in the model, and therefore, the model would tend to find a mathematical optimal to fit the reality.

This might very well be the case in the Sio basin, as the studied area has some wetlands and swamps. The processes, for example, that take place in swamps are not reflected in the GR4J model. GR4J is a lumped model, where different land uses within the catchment are ignored.

Moreover, it is important to point out that not only the model input data contain uncertainties, but also the observed flow. It is usually not the streamflow what is measured, but the water level, which is afterwards converted to a discharge rate using a rating curve. The rating curve is specific for each gauging station and should be defined by regular measurements where the water level is measured, and the streamflow is calculated with the velocity-area method. The measurements of water level and discharge to define the rating curve should be done at different flow conditions, ranging from low to high flows, in order to reduce uncertainties.

In this case, it is difficult to determine how realistic the water level peaks in the observed flow are, as there is no information about the rating curve that was used. However, it seems to be an upper limit in the measured peak flows. In this case, the gauging station is under a bridge, and when the water reaches the top level and the are is flooded, the water level would theoretically increase, but it is not contained in the river channel and, therefore leading to an underestimation of peak flows. When the model tries to fit the input data to some flow values which are underestimated, the hydrological model tries to mathematically fit the reality by giving parameter values which do not reflect reality. Therefore, it is important to be aware of the uncertainties in the streamflow values.

4.4 Estimation of Extreme Values of discharge

4.4.1 Return period from observed flow

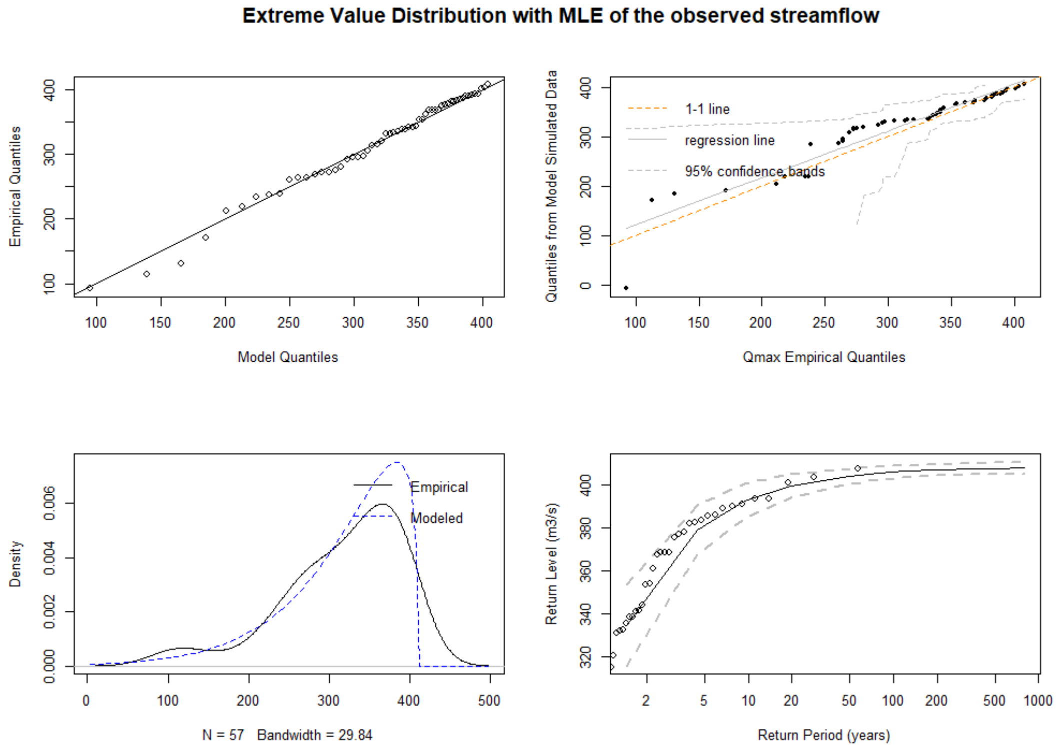


Figure 25. Extreme value distribution with MLE of the observed streamflow

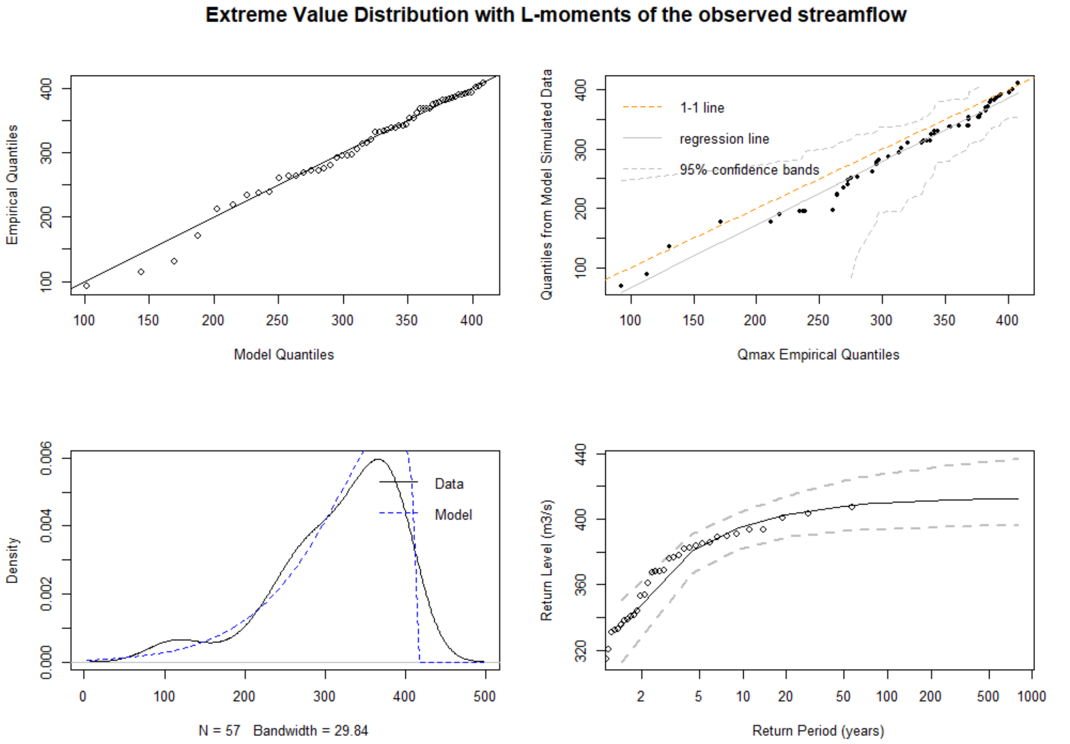


Figure 26. Extreme value distribution with L-moments of the observed streamflow

4.4.2 Return period from simulated flow with KGE

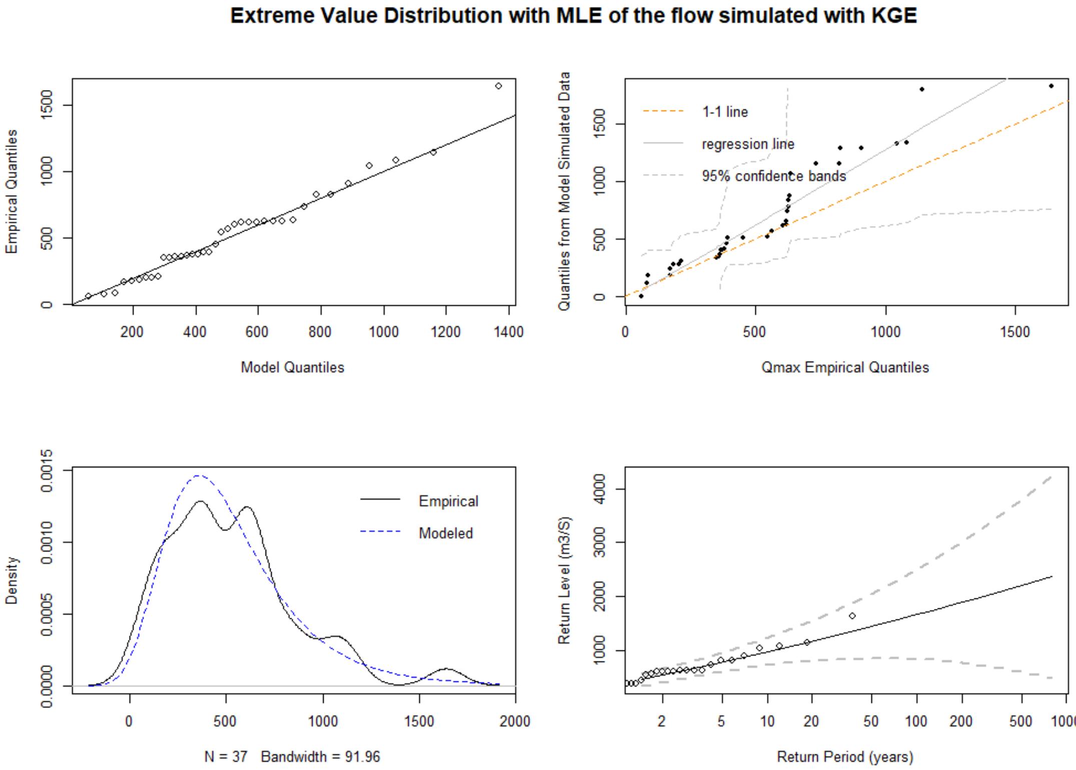


Figure 27. Extreme value distribution with MLE of the flow simulated with KGE

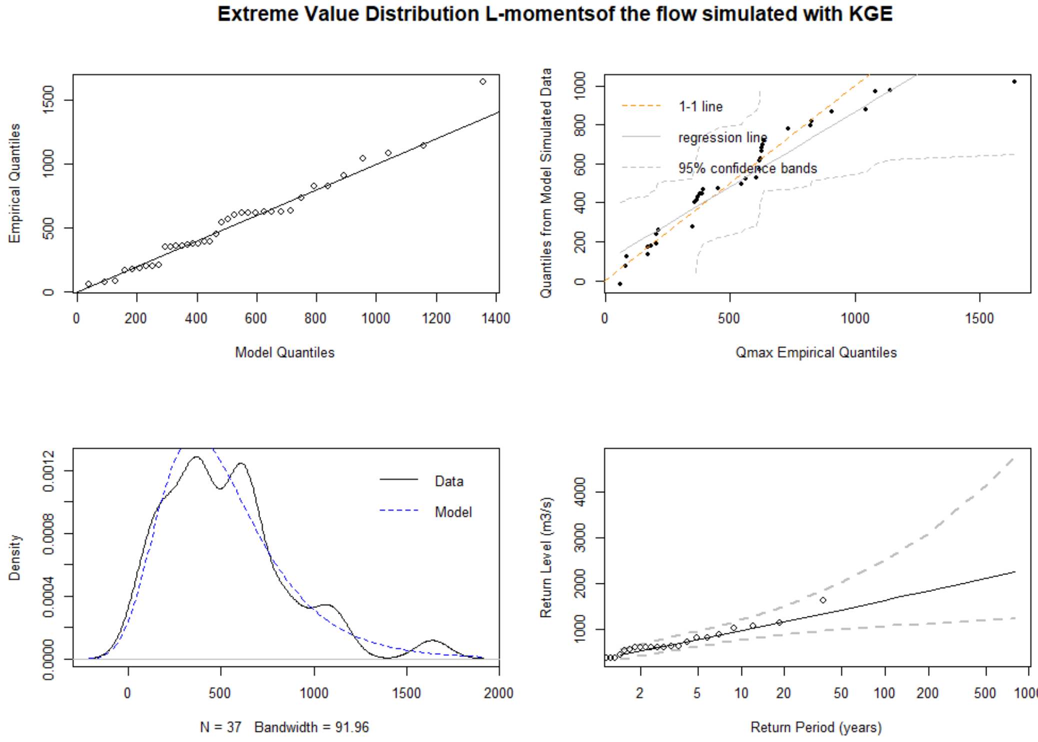


Figure 28. Extreme value distribution with L-moments of the flow simulated with KGE

4.4.3 Return period from simulated flow with NSE

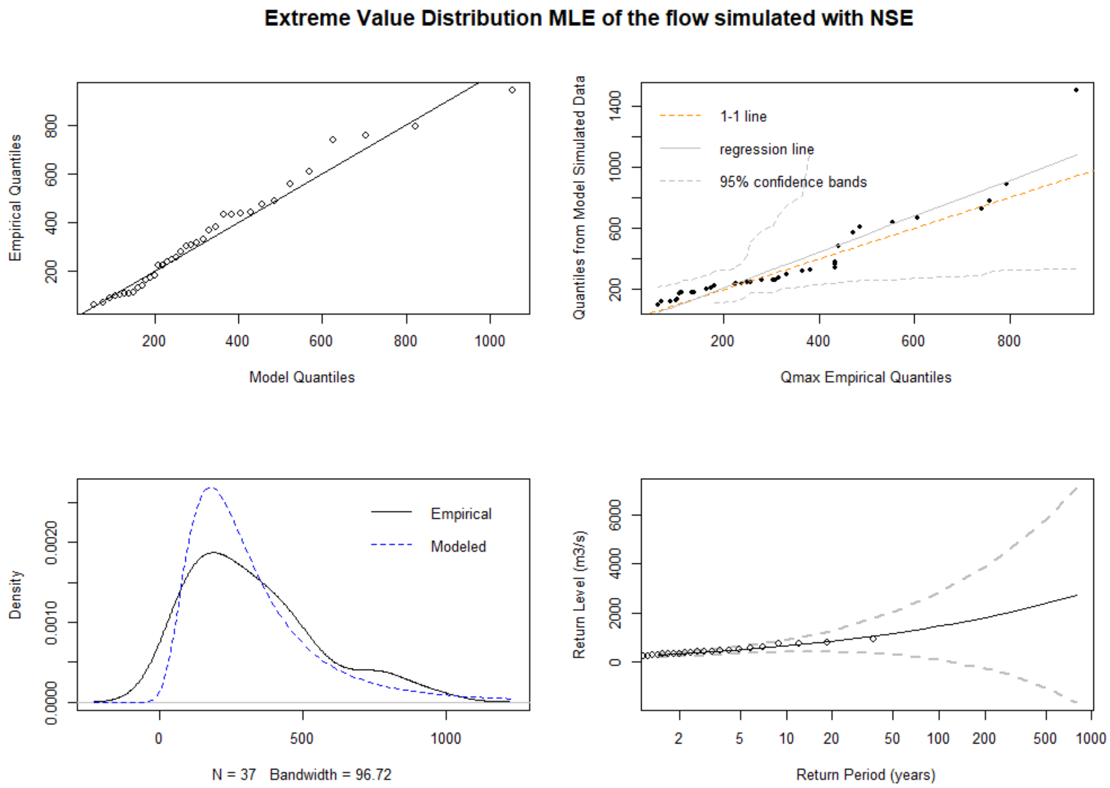


Figure 29. Extreme value distribution with MLE of the flow simulated with NSE

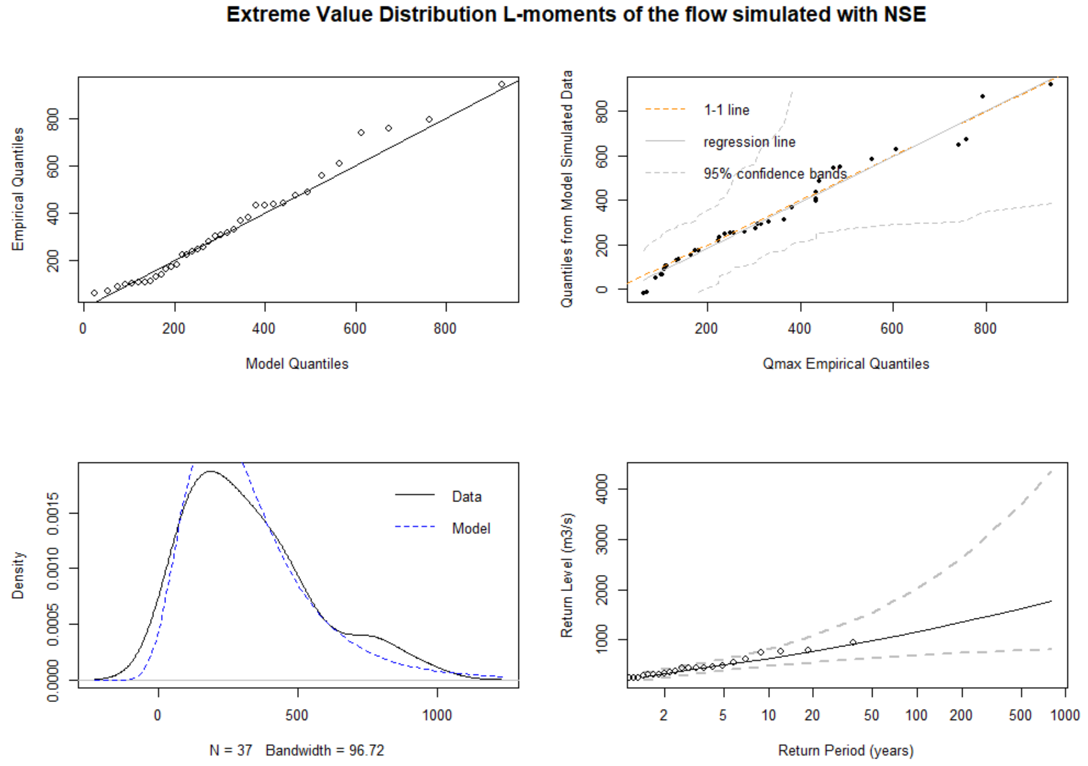


Figure 30. Extreme value distribution with L-moments of the flow simulated with NSE

4.4.4 Discussion of extreme value estimation

The results of the extreme value distribution models have to be interpreted cautiously and with understanding of the error and uncertainty sources. As in any forecasting, its accuracy depends on the quality of the observations or input data, and of the model performance itself.

Regarding the quality of the input data, it is important to keep in mind that only the maximum annual values have been taking into consideration. Figure 31, Figure 32 and Figure 33 show the extreme values obtained with the observed flow and flow simulated with KGE and NSE in the hydrological model.

The return periods obtained with the simulated stream values will probably be too high, as the peaks were much higher than the maximum observed flow values. The observed values will, however, lead to too low values, as some years do not have a full timeseries. Figure 31 shows that some years with uncomplete time series have lower maximum values.

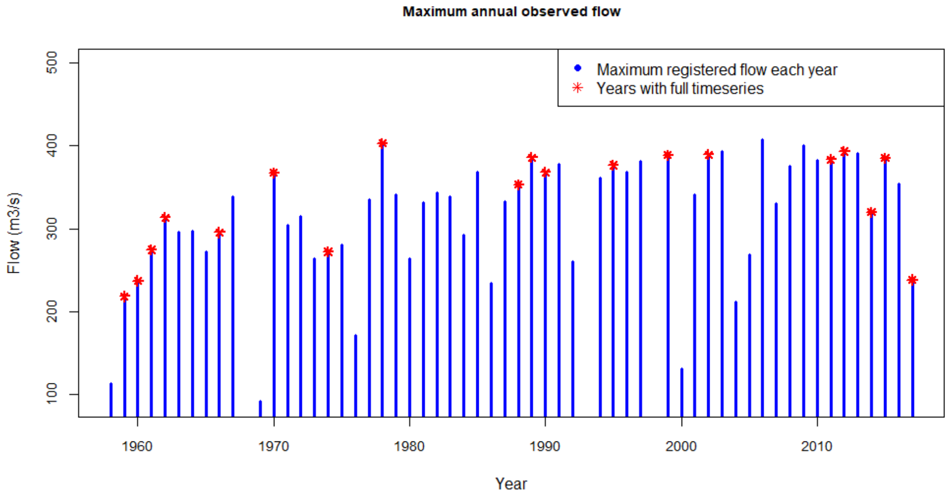


Figure 31. Maximum annual observed flow

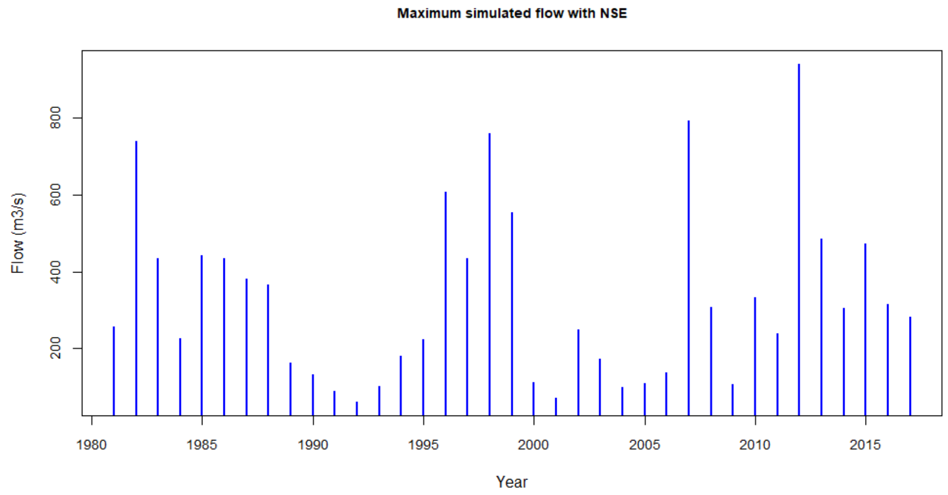


Figure 32. Maximum simulated flow with NSE

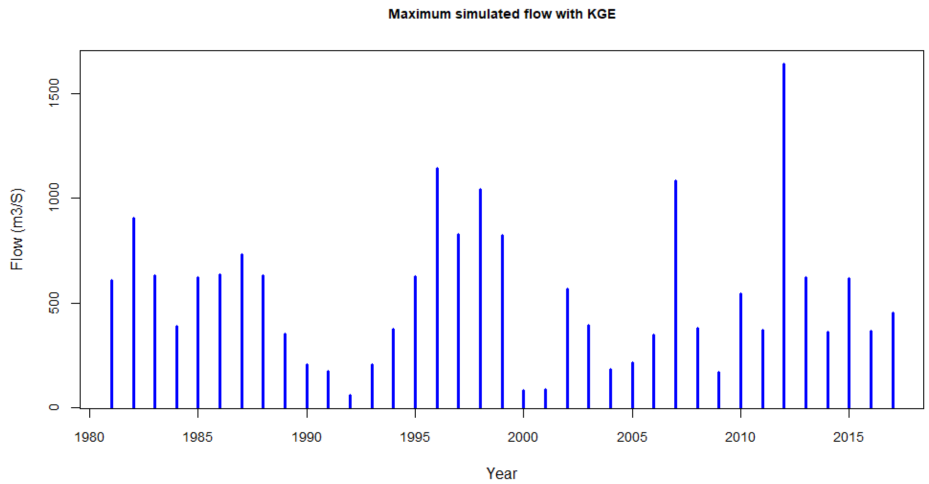


Figure 33. Maximum simulated flow with KGE

Regarding the performance of the extreme value forecasting models, the resulting return levels for the return periods of 10 years, 30 years, 50 years and 100 years, is displayed in Figure 34 and Table 18.

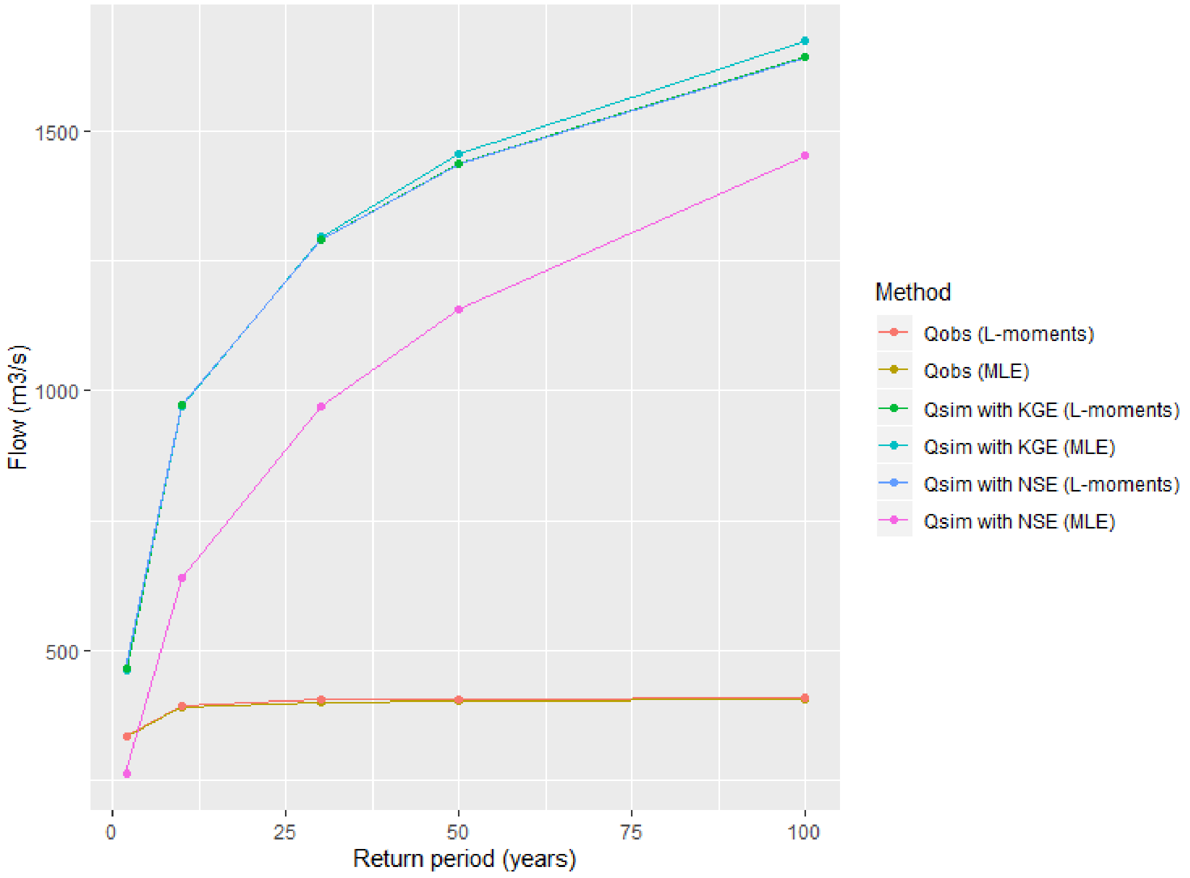


Figure 34. Return levels for each return period obtained with each method

Table 18. Streamflow values obtained with each method for each return period.

	T2		T10		T30		T50		T100	
	mm/d	m^3/s	mm/d	m^3/s	mm/d	m^3/s	mm/d	m^3/s	mm/d	m^3/s
Qobs (MLE)	5.06	335.84	5.92	392.80	6.06	402.09	6.09	404.17	6.12	405.89
Qobs (L-moments)	5.04	334.50	5.95	394.93	6.11	405.55	6.15	408.03	6.18	410.13
Qsim with NSE (MLE)	3.93	260.64	9.60	636.74	14.65	972.06	17.52	1161.98	22.06	1463.07
Qsim with NSE (L-moments)	7.00	464.57	14.65	971.84	19.44	1289.44	21.67	1437.59	24.73	1640.62
Qsim with KGE (MLE)	6.94	460.07	14.52	963.33	19.48	1292.28	21.85	1449.37	25.15	1668.45
Qsim with KGE (L-moments)	7.00	464.57	14.65	971.84	19.44	1289.44	21.67	1437.59	24.73	1640.62

It can be seen, that the extreme values forecast with the observed data leads to a less variable behavior, since higher return levels only lead to a slight increase in discharge. For example, based on the observed discharge, the increase in discharge between a HQ_{50} and HQ_{100} is only

0.4 to 0.5%, which seems unrealistic. At the same time, the simulated streamflow values show a much higher increase of the return levels, ranging between 14 and 26 %, depending on the on model setup and data used.

The extreme values obtained with KGE-based simulations are higher than those obtained with NSE-based simulations. The extreme values obtained with both methods for estimating the parameters of the GEV are similar.

The parameter estimation with L-moments provided higher return level values than the one with MLE in most of the cases. However, in the simulated flow with KGE, the method that has resulted in higher return levels was the MLE, especially for high return levels.

Overall, the main difference is found between the simulated and observed flow. The differences between the simulation data based on the different calibration objective functions (NSE/KGE) or between the parameter estimation methods (MLE/L-moments) are smaller.

It is difficult to interpret the results due to the high level of uncertainty. This uncertainty level increases with the return period. Therefore, the differences between the return level values increase at higher return periods.

With the aim of providing a more visual and understanding overview of the obtained results, Figure 35 shows the return levels obtained with the different techniques plotted as a box-plot for each calculated return period.

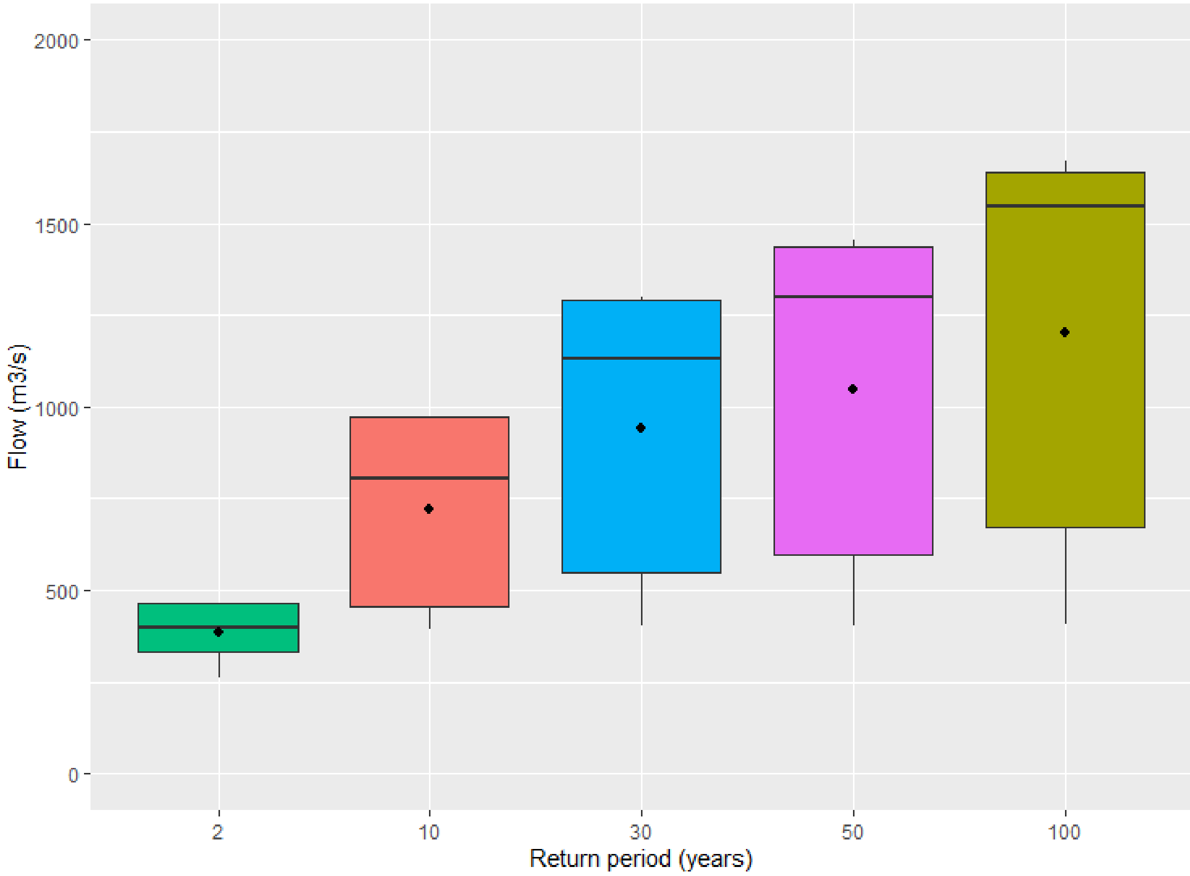


Figure 35… Return level boxplot (m³/s)

5 Summary and conclusions

The aim of this study, conducted in the Sio River Basin (1011 km²), Western Kenya, was to provide

i. complete time series of precipitation for several stations in the region,
ii. the mean catchment precipitation by interpolating the station data,
iii. the set-up and application of a hydrological model to extend information of flow for the Sio and
iv. (the estimation of extreme values of discharge for different return periods based on different data sets.

Data analysis is subject and heavily influenced by the quality and quantity of data. Wrong data will lead to wrong analysis, models and results. This is to say, that all procedures here elaborated are directly linked to the measurements in field. For this analysis was assumed that all data series were correctly gauged.

In addition, the assumption of mean areal rainfall distribution assumes a unique value for the whole catchment and cannot not describe the heterogeneity of rainfall conditions in the basin. In particular, the influence of the elevation on meteorological parameters can be significant. For this study, it is however assumed, that the heterogeneity is not strongly pronounced due to the small elevation differences in the area.

Temporal imputation done by linear regression based on best correlation stations is a robust technique to fill data gaps. It does not exclude data sets, but since the data obtained to fill the gap is calculated with the slope of the linear regression, all the values calculated with it will fall in the same line, which is not able to show the natural fluctuation of the values.

Moreover, the proportionality of rain among 10 days can vary from one station to another, even if they show a high correlation value. Certainly, all assumptions made here were chosen with the best knowledge, despite their limitations.

Consequently, when the hydrological model tries to fit the simulated flow to the measured one, the parameters which define the relationship between the natural processes might be unrealistic.It can be observed that the NSE leads to higher baseflow values and lower peaks, while KGE generates higher flood peak values but lower baseflow. The KGE fits better to the observed values than NSE.

The extreme values forecast with the observed data leads to a less variable behavior, compared to the return levels obtained with the simulated flow values.

The return periods obtained with the simulated streamflow values will probably be too high, as the peaks were much higher than the maximum observed flow values. The return periods obtained with the observed flow will probably be too low, as in many cases the observation time series were not complete and there are indications, that the observations are erroneous in high flow conditions, since the bridge culvert, where the gauging station is located, may be limited in capacity.

It is difficult to interpret the estimated discharge values for different return periods due to the high level of ambiguity. Biases may increase with the application of hydrological and forecasting models. Additionally, uncertainty levels increase with the length of the time projection or forecasting period. Therefore, the dispersion between the return level values increases at higher return periods.

In conclusion, results from this study must be used with caution and reservations.

6 Bibliography

Anshuman, Aatish; Kunnath-Poovakka, Aiswarya; Eldho, T. I. (2019): Performance evaluation of conceptual rainfall-runoff models GR4J and AWBM. In *ISH Journal of Hydraulic Engineering* 30 (4), pp. 1–10. DOI: 10.1080/09715010.2018.1556124.

Baguley, Thomas (2012): Serious stats. A guide to advanced statistics for the behavioral sciences. Basingstoke: Palgrave Macmillan.

BMLFUW (2011): Leitfaden. Verfahren zur Abschätzung von Hochwasserkennwerten. Bundesministeriums für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft Abt.VII3. Available online at https://www.bmlrt.gv.at/dam/jcr:d0aa7fcb-f629-4c92-95e6-b3f1415252a9/Leitfaden_Verfahren_zur_Abschaetzung_von_HWKennwerten.pdf.

Chavula, Geoffrey M. S. (2013): THE EVALUATION OF THE PRESENT AND POTENTIAL WATER RESOURCES MANAGEMENT FOR THE LAKE CHILWA BASIN.

EPA (2017): An Overview of Rainfall-Runoff Model Types. United States Environmental Protection Agency (EPA/600/R-14/152).

eWater Source 4.1.0: GR4J - SRG - Source User Guide 4.1. Available online at https://wiki.ewater.org.au/display/SD41/GR4J+-+SRG, checked on 2/5/2020.

Gao, Yongbo; Merz, Christoph; Lischeid, Gunnar; Schneider, Michael (2018): A review on missing hydrological data processing. In *Environ Earth Sci* 77 (2), p. 337. DOI: 10.1007/s12665-018-7228-6.

Gilleland, Eric; Katz, Richard W. (2016): extRemes 2.0: An Extreme Value Analysis Package in R. In *J. Stat. Soft.* 72 (8). DOI: 10.18637/jss.v072.i08.

Glen, Stephanie (2016): Extreme Value Distribution & the Extreme Value Theory. Available online at https://www.statisticshowto.datasciencecentral.com/extreme-value-distribution/, checked on 2/11/2020.

Gumbel, E. J. (2013): Statistics of extremes. [Place of publication not identified]: Echo Point Books & Media.

Gupta, Hoshin V.; Kling, Harald; Yilmaz, Koray K.; Martinez, Guillermo F. (2009): Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. In *Journal of Hydrology* 377 (1-2), pp. 80–91. DOI: 10.1016/j.jhydrol.2009.08.003.

Haan, Charles T. (1977): Statistical methods in hydrology. Ames: Iowa State Univ. Pr.

Institution of Engineers, Australia (1987): Australian Rainfall and Runoff. With assistance of D. H. Pilgrim. Revised Edition 1987 (Reprinted edition 1998). Barton, ACT (1).

Jajarmizad, Milad; Harun, Sobri; Salarpour, Mohsen (2012): A Review on Theoretical Consideration and Types of Models in Hydrology. In *J. of Environmental Science and Technology* 5 (5), pp. 249–261. DOI: 10.3923/jest.2012.249.261.

Michel, Claude (1991): Hydrologie appliquée aux petits bassins ruraux, Hydrology handbook. Cemagref, Antony, France.

Nash, J. E.; Sutcliffe, J. V. (1970): River flow forecasting through conceptual models part I — A discussion of principles. In *Journal of Hydrology* 10 (3), pp. 282–290. DOI: 10.1016/0022-1694(70)90255-6.

Perrin, Charles; Michel, Claude; Andréassian, Vazken (2003): Improvement of a parsimonious model for streamflow simulation. In *Journal of Hydrology* 279 (1-4), pp. 275–289. DOI: 10.1016/S0022-1694(03)00225-7.

Perrin, Charles; Michel, Claude; Andréassian, Vazken (2007): Modèles hydrologiques du Génie Rural (GR). Cemagref, UR Hydrosystèmes et Bioprocédés: CEMAGREF.

Ritter, Axel; Muñoz-Carpena, Rafael (2013): Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. In *Journal of Hydrology* 480, pp. 33–45. DOI: 10.1016/j.jhydrol.2012.12.004.

Rogelis, María Carolina; Werner, Micha; Obregón, Nelson; Wright, Nigel (2016): Hydrological model assessment for flood early warning in a tropical high mountain basin. In *Hydrol. Earth Syst. Sci. Discuss.*, pp. 1–36. DOI: 10.5194/hess-2016-30.

Saha, Partha Pratim; Zeleke, Ketema (2015): Rainfall-Runoff Modelling for Sustainable Water Resources Management: SWAT Model Review in Australia. In Shimelis Gebriye Setegn, Maria Concepcion Donoso (Eds.): Sustainability of Integrated Water Resources Management. Water Governance, Climate and Ecohydrology, vol. 8. 1st ed. 2015. Cham: Springer, pp. 563–578.

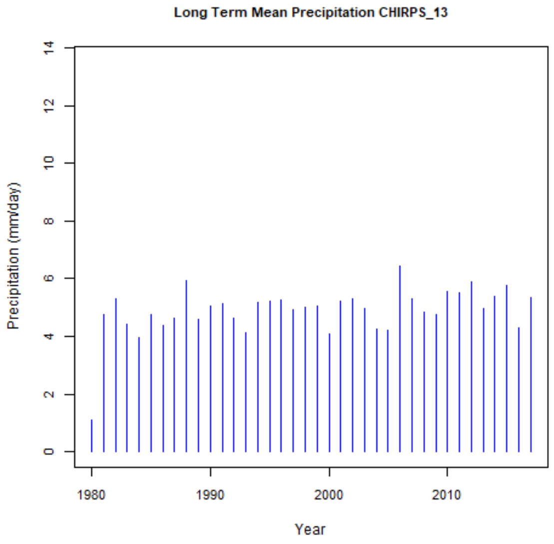
Singh, Vijay P. (Ed.) (1995): Computer models of watershed hydrology. rev. ed. Highlands Ranch, Colo.: Water Resources Publications.

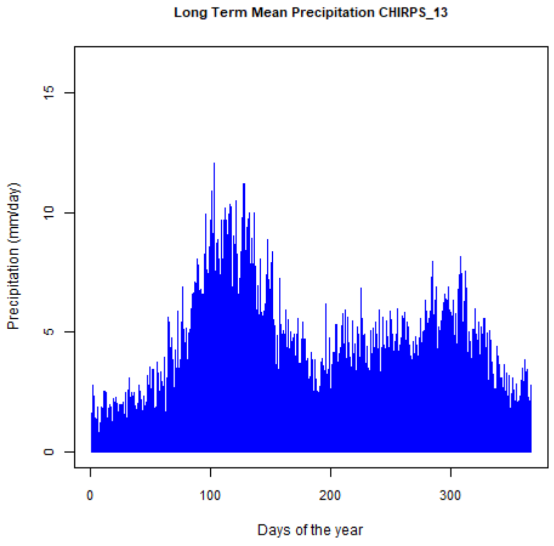
UTHH (2011): Integrated Flood Management. 8-HydrologicalModeling. University of Technology Hamburg. Harburg, 2011. Available online at http://daad.wb.tu-harburg.de/fileadmin/BackUsersResources/Hydrology/8-HydrologicalModeling/additional_info/Thiessen_polygon_Method.pdf.

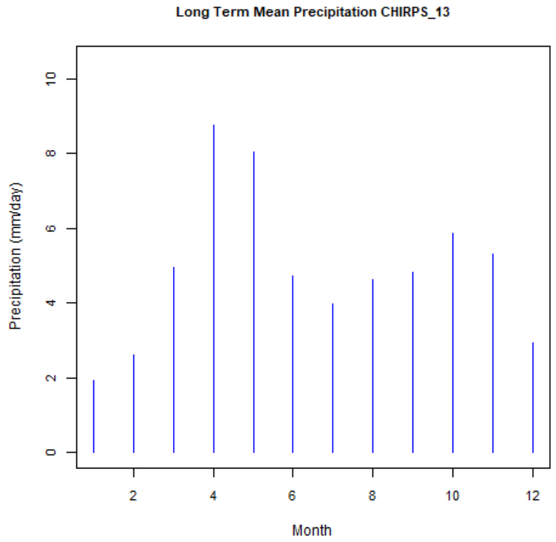
WMO (2008): Guide to hydrological practices. 6th ed. Geneva: World Meteorological Organization.

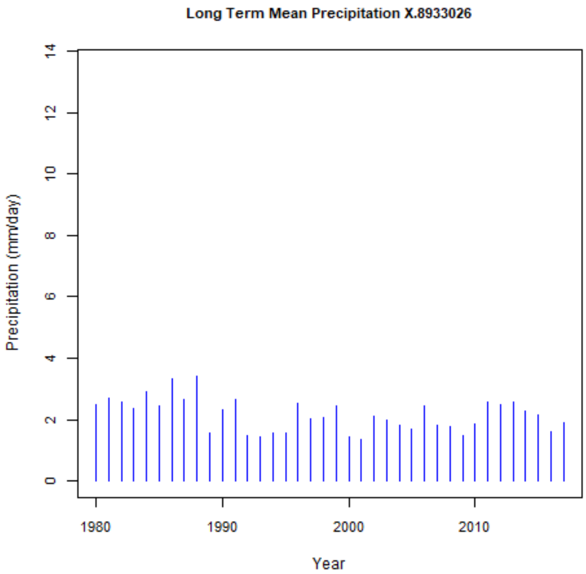
7 Appendix

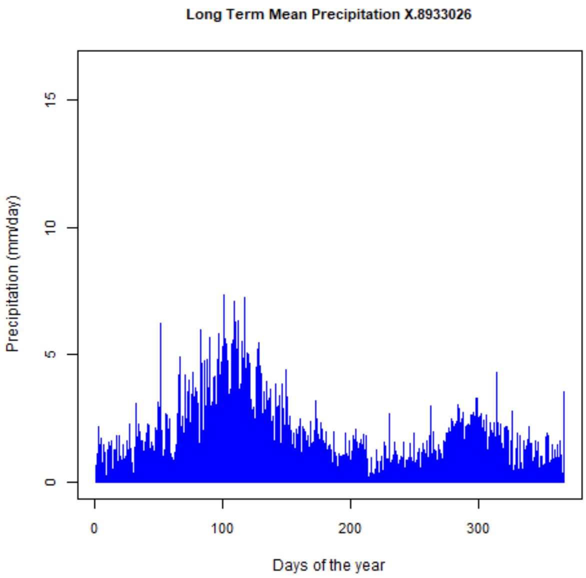
7.1 Long term mean precipitation values per station

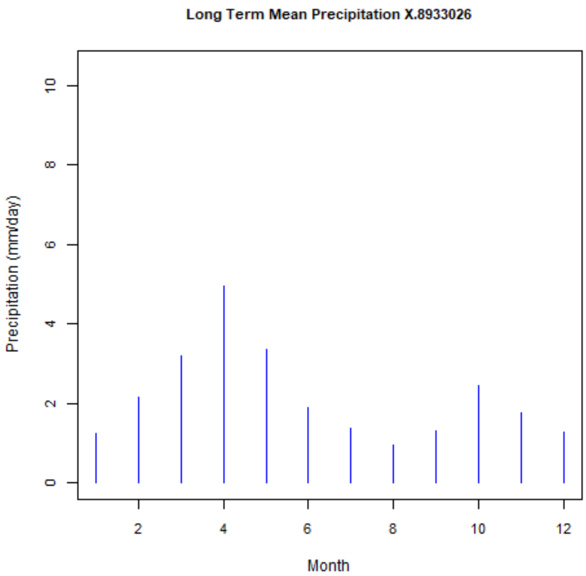


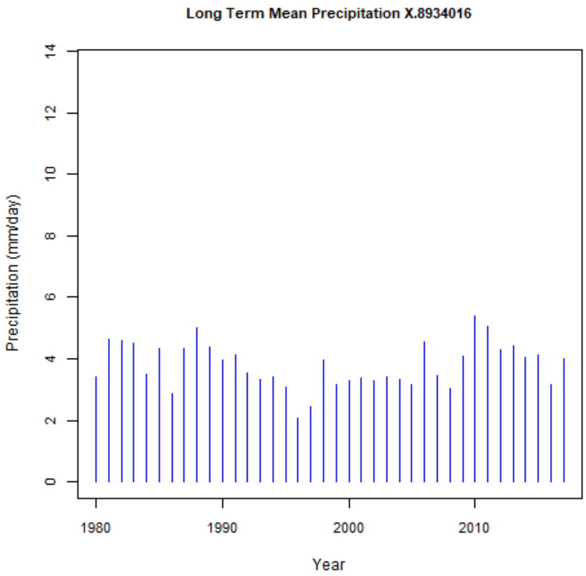


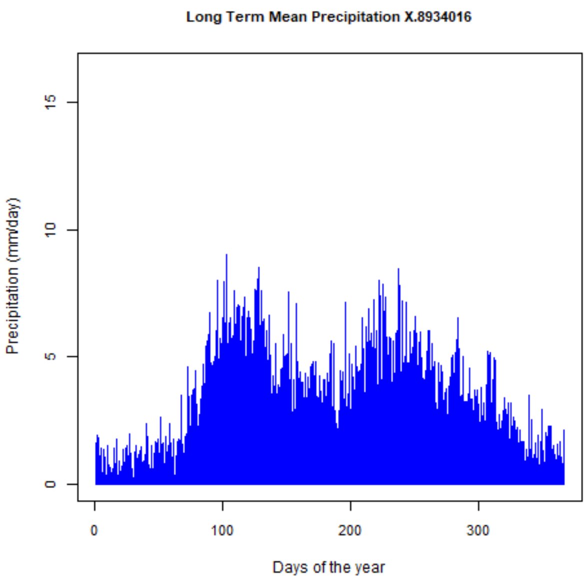

Long Term Mean Precipitation CHIRPS_13
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

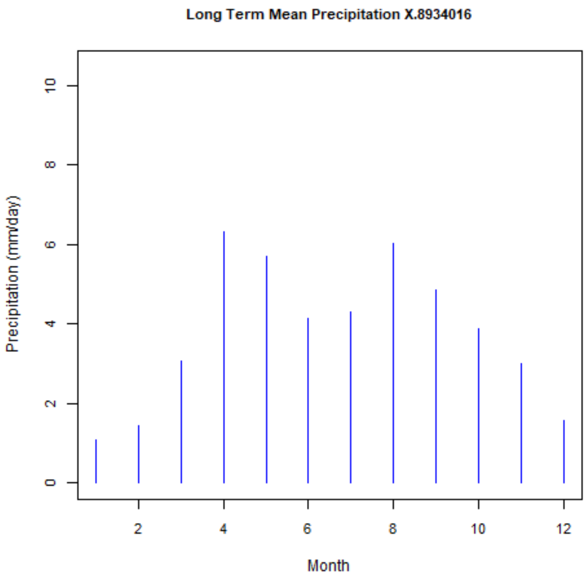

Long Term Mean Precipitation X.8933026
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

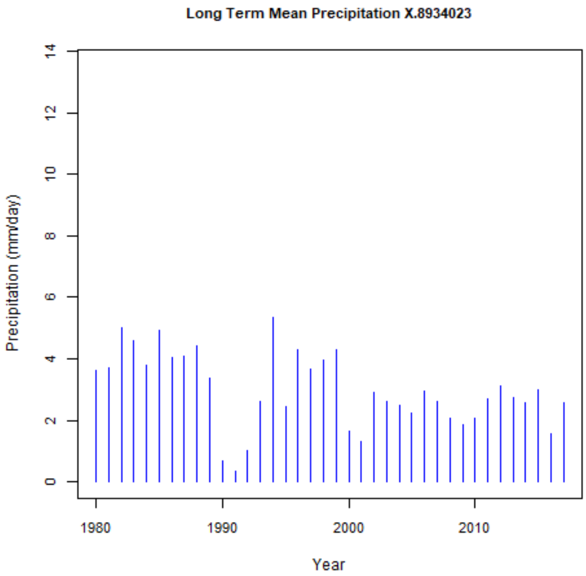

Long Term Mean Precipitation X.8933026
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

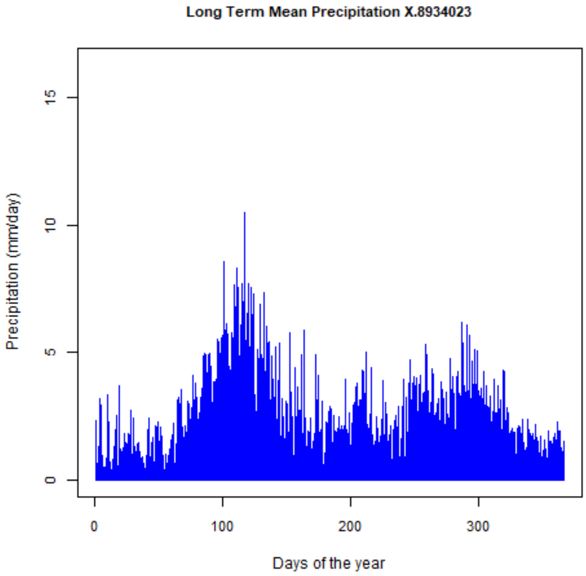

Long Term Mean Precipitation X.8933026
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

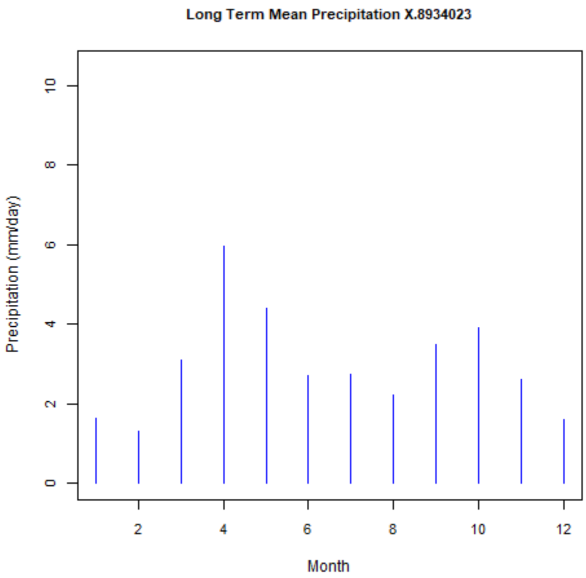

Long Term Mean Precipitation X.8934016
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

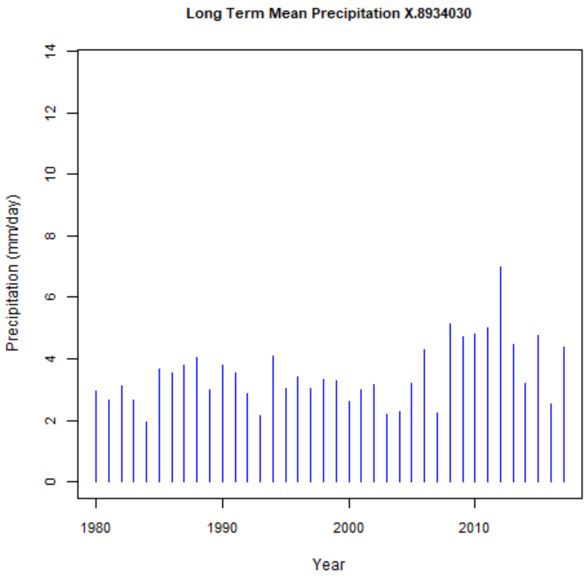

Long Term Mean Precipitation X.8934016
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

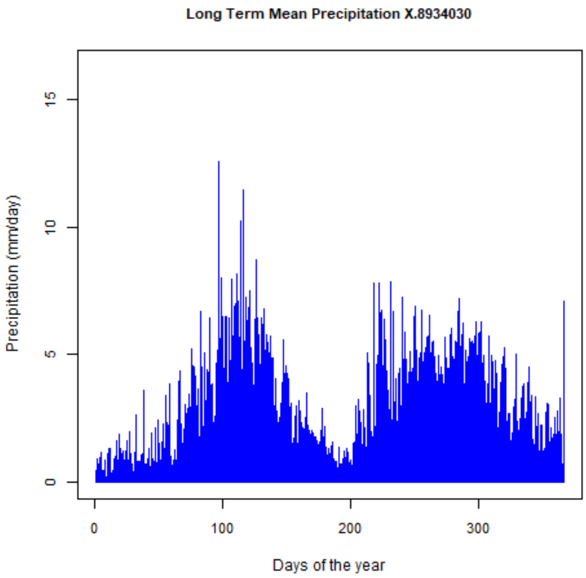

Long Term Mean Precipitation X.8934016
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month


Long Term Mean Precipitation X.8934023
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

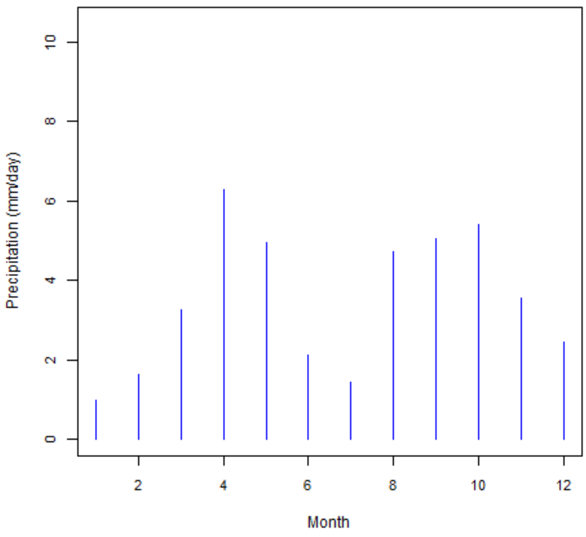

Long Term Mean Precipitation X.8934023
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year


Long Term Mean Precipitation X.8934023
Precipitation (mm/day)
Month

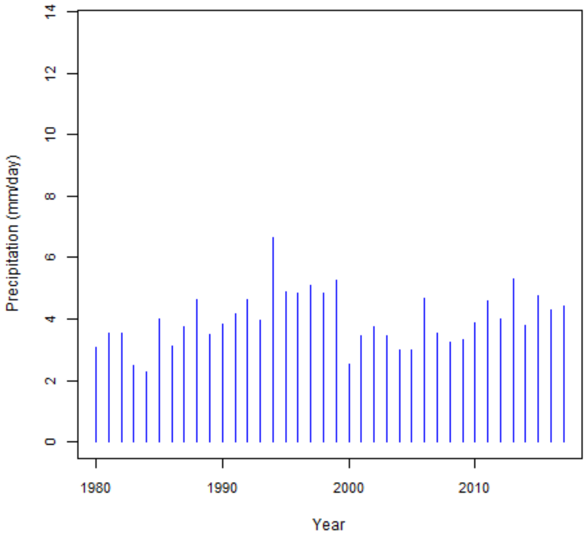

Long Term Mean Precipitation X.8934030
Precipitation (mm/day)
Year


Long Term Mean Precipitation X.8934030
Precipitation (mm/day)
Days of the year

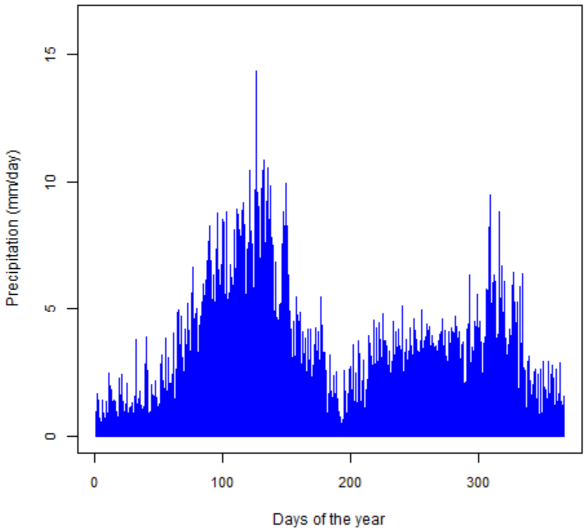
Long Term Mean Precipitation X.8934030

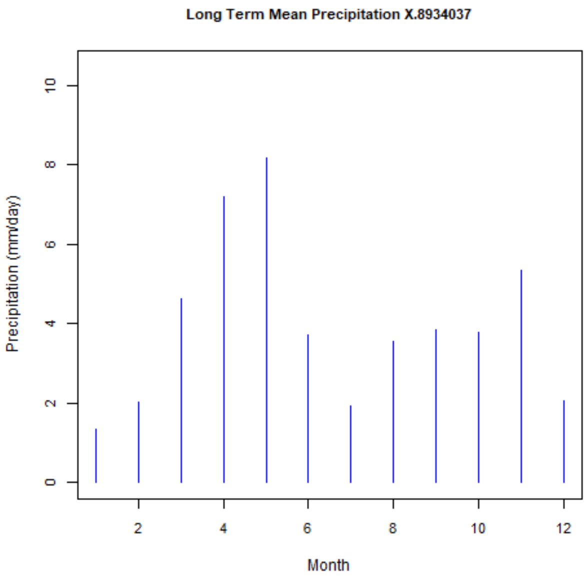


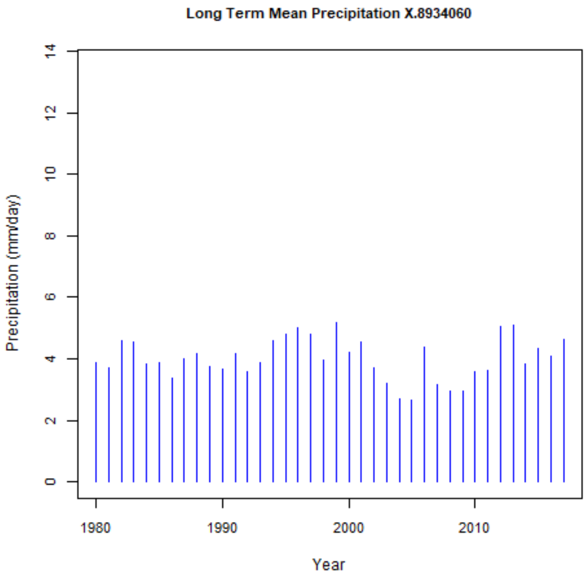
Long Term Mean Precipitation X.8934037

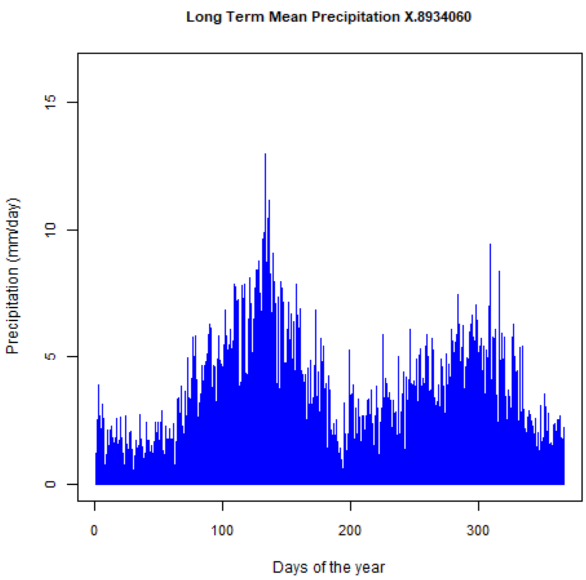


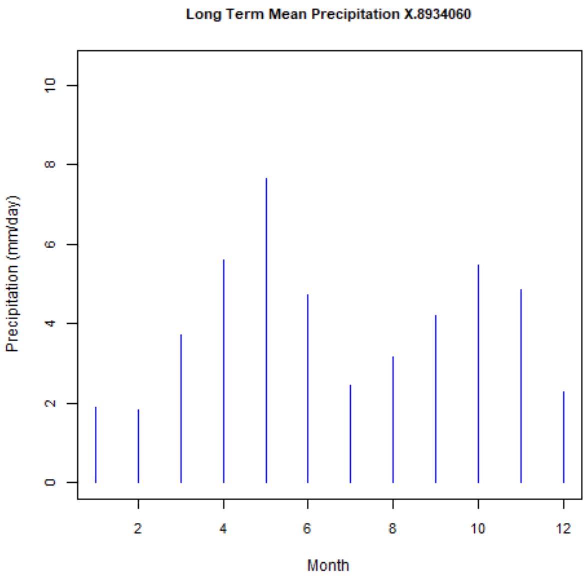
Long Term Mean Precipitation X.8934037

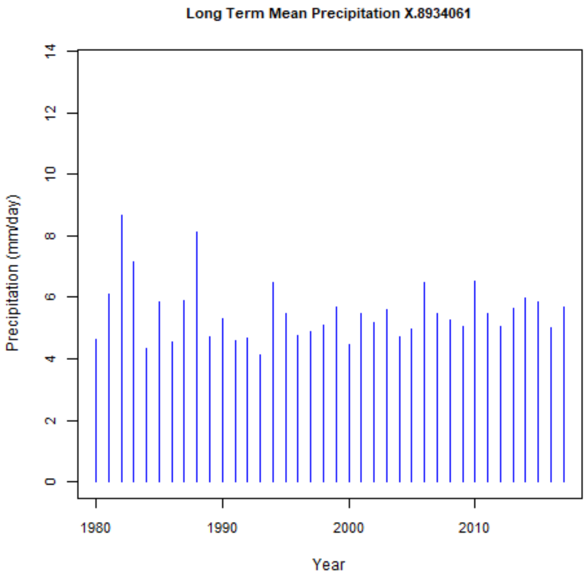


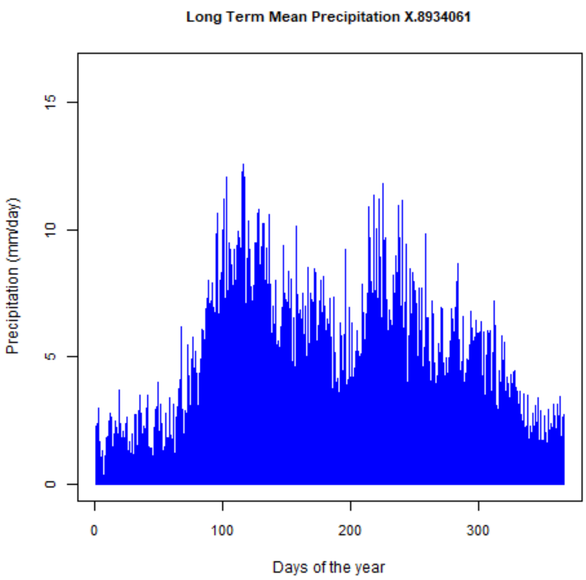

Long Term Mean Precipitation X.8934037
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

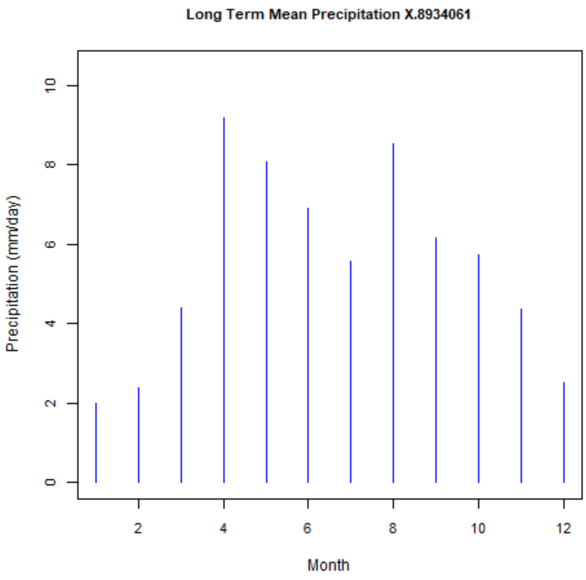

Long Term Mean Precipitation X.8934060
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

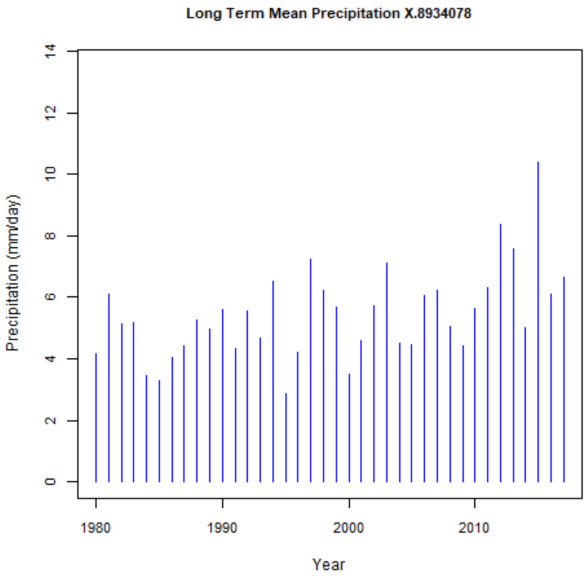

Long Term Mean Precipitation X.8934060
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

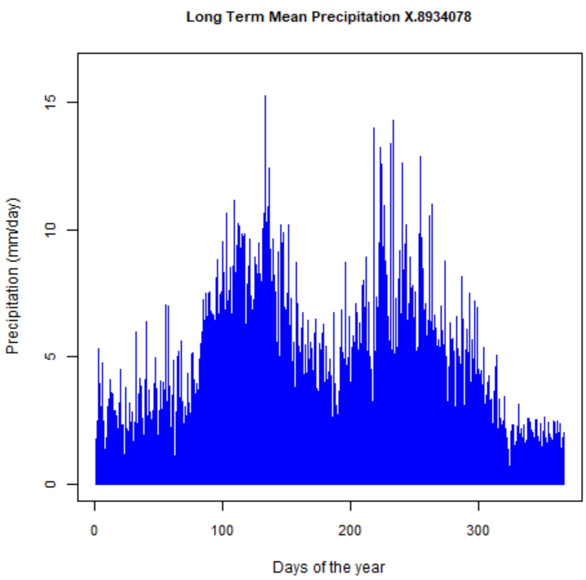

Long Term Mean Precipitation X.8934060
Precipitation (mm/day)
Month
0
2
4
6
8
10
12

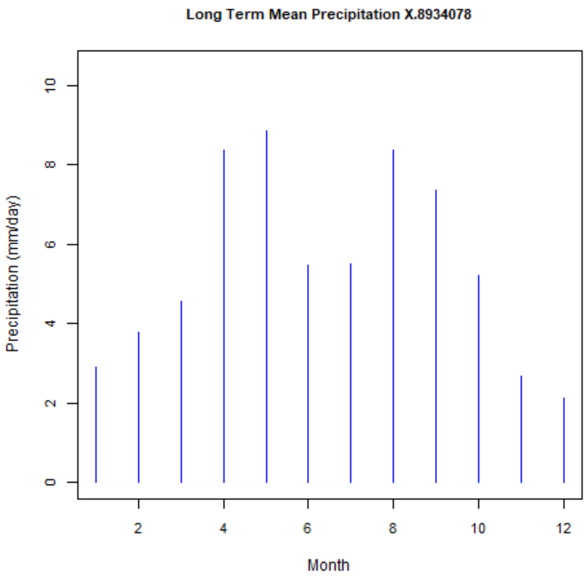

Long Term Mean Precipitation X.8934061
Precipitation (mm/day)
Year
1980
1990
2000
2010
0
2
4
6
8
10
12
14

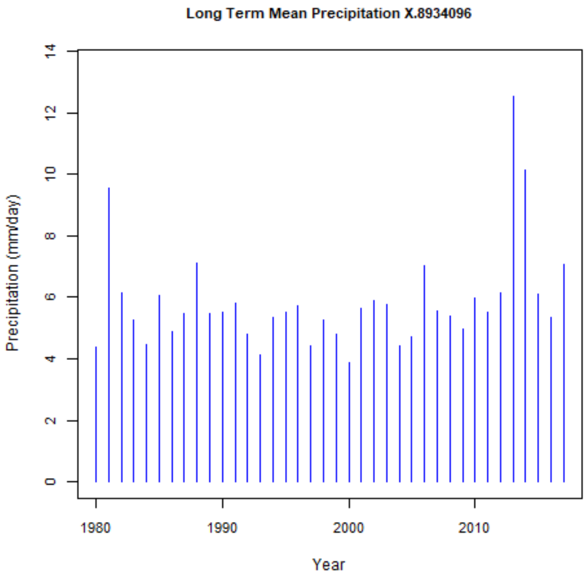

Long Term Mean Precipitation X.8934061
Precipitation (mm/day)
Days of the year
0
100
200
300
5
10
15

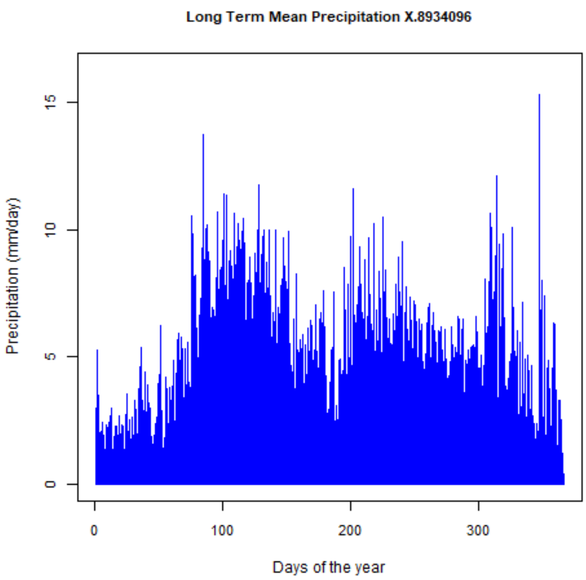

Long Term Mean Precipitation X.8934061
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

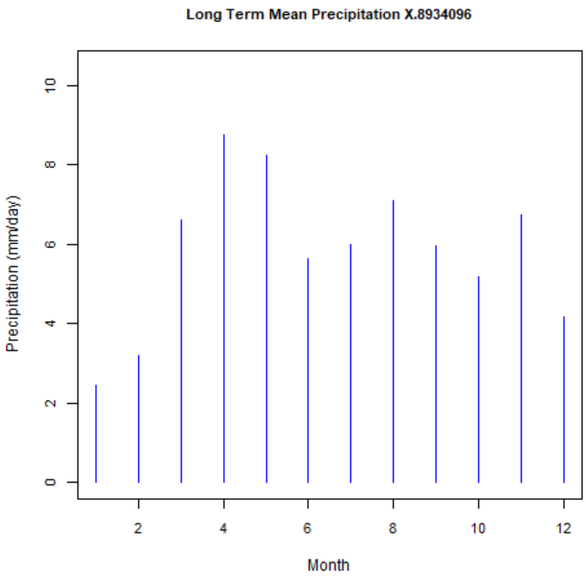

Long Term Mean Precipitation X.8934078
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

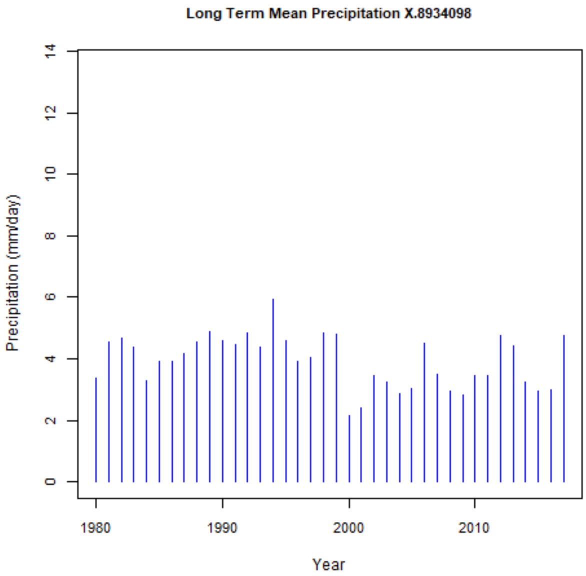

Long Term Mean Precipitation X.8934078
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

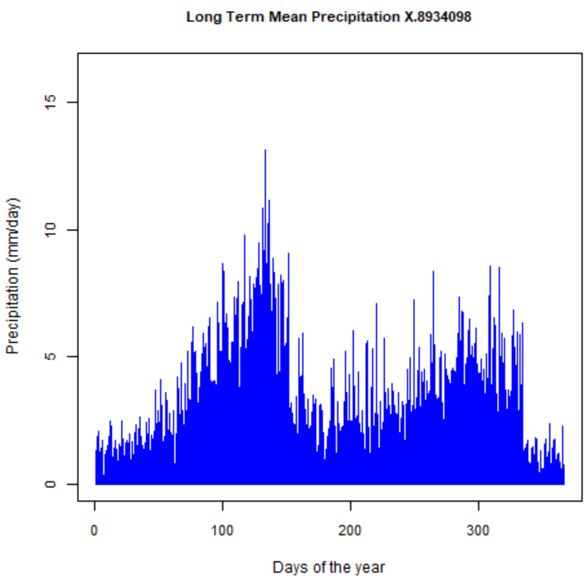

Long Term Mean Precipitation X.8934078
Precipitation (mm/day)
Month
0
2
4
6
8
10
12

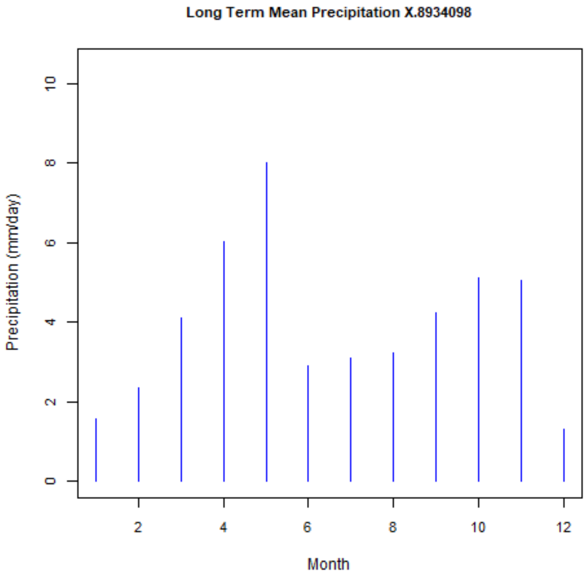

Long Term Mean Precipitation X.8934096
Precipitation (mm/day)
Year
1980
1990
2000
2010
0
2
4
6
8
10
12
14


Long Term Mean Precipitation X.8934096
Precipitation (mm/day)
Days of the year
0
100
200
300
5
10
15

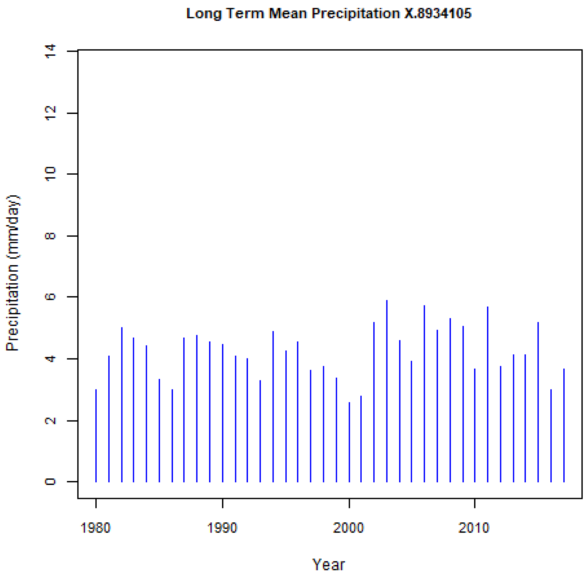

Long Term Mean Precipitation X.8934096
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month


Long Term Mean Precipitation X.8934098
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

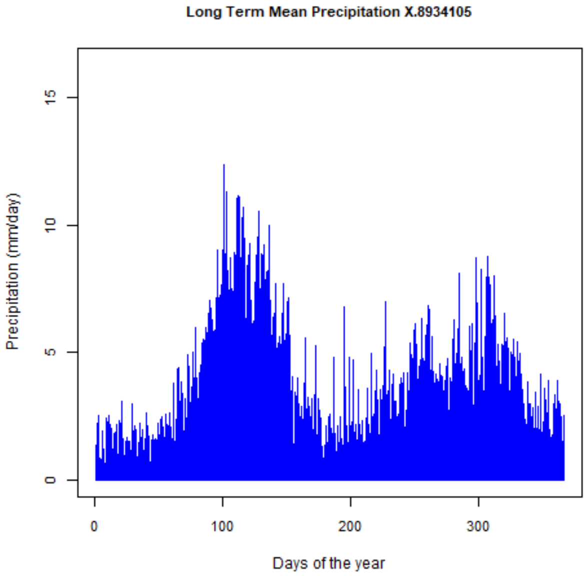

Long Term Mean Precipitation X.8934098
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year



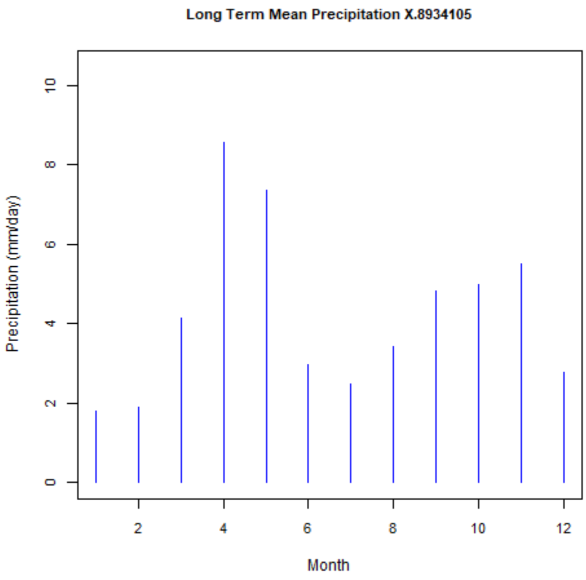
Long Term Mean Precipitation X.8934098
Precipitation (mm/day)
Month

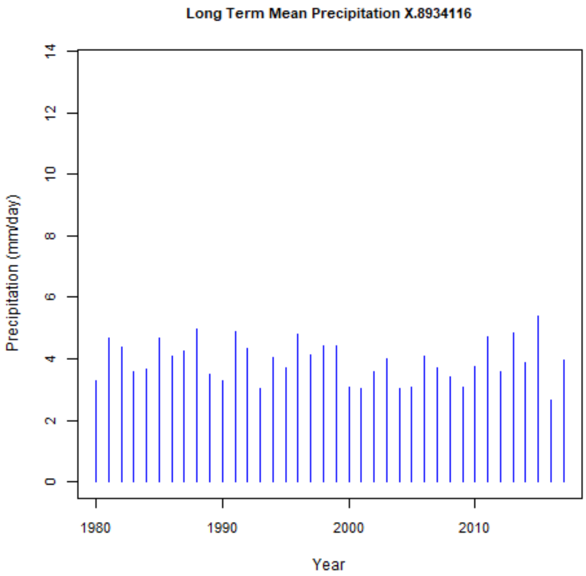


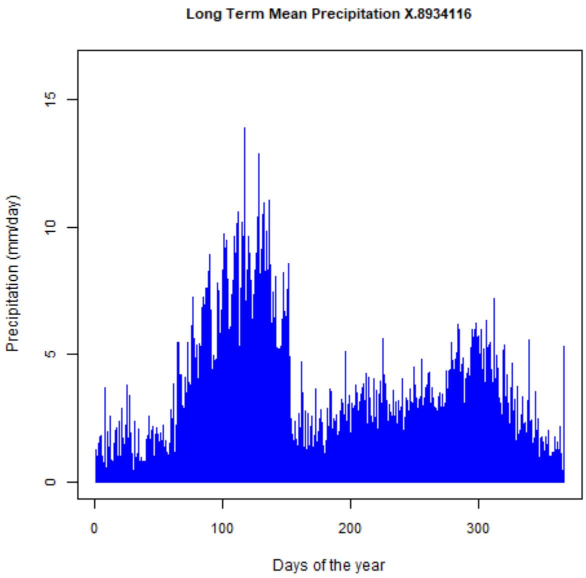
Long Term Mean Precipitation X.8934105
Precipitation (mm/day)
Year



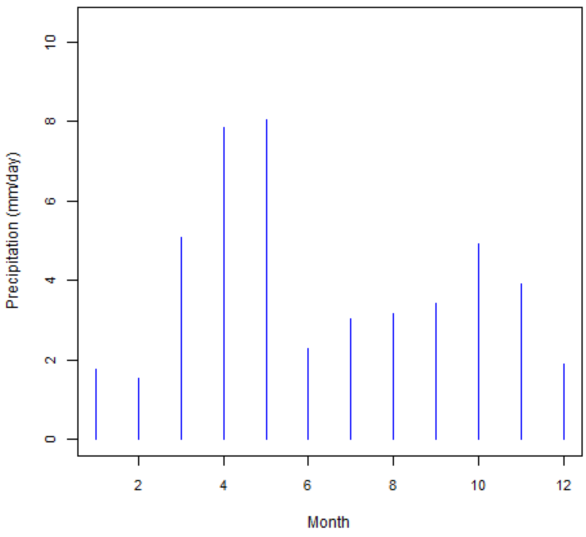
Long Term Mean Precipitation X.8934105
Precipitation (mm/day)
Days of the year


Long Term Mean Precipitation X.8934105
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

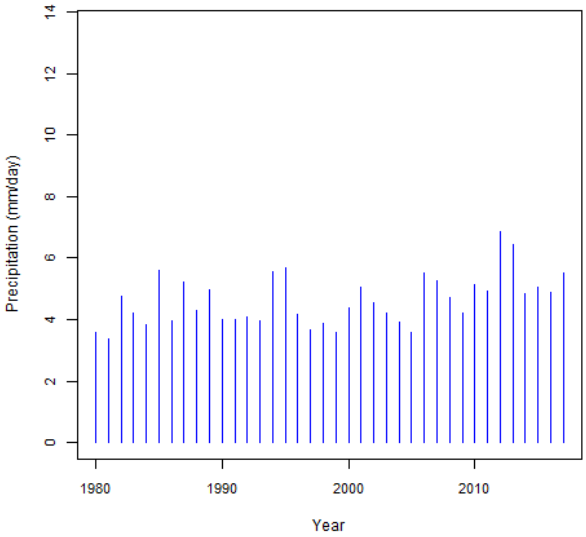

Long Term Mean Precipitation X.8934116
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year


Long Term Mean Precipitation X.8934116
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

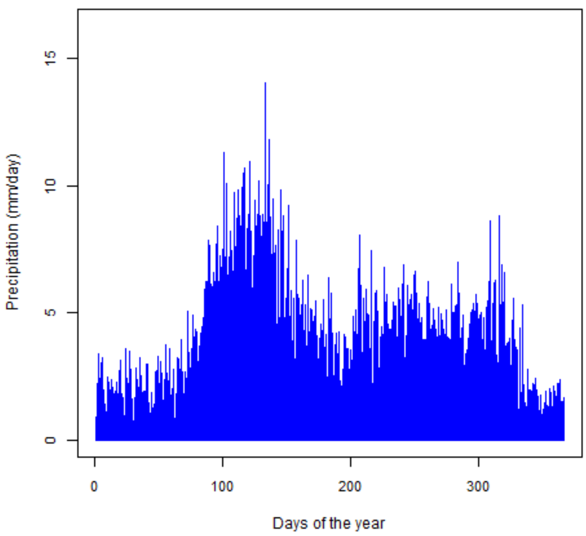
Long Term Mean Precipitation X.8934116

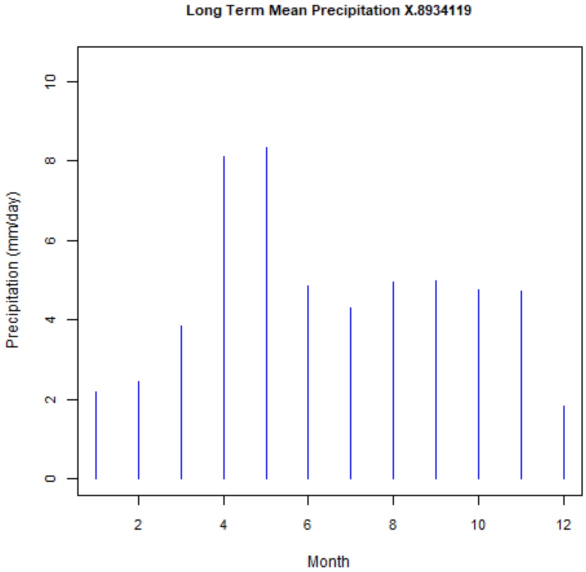


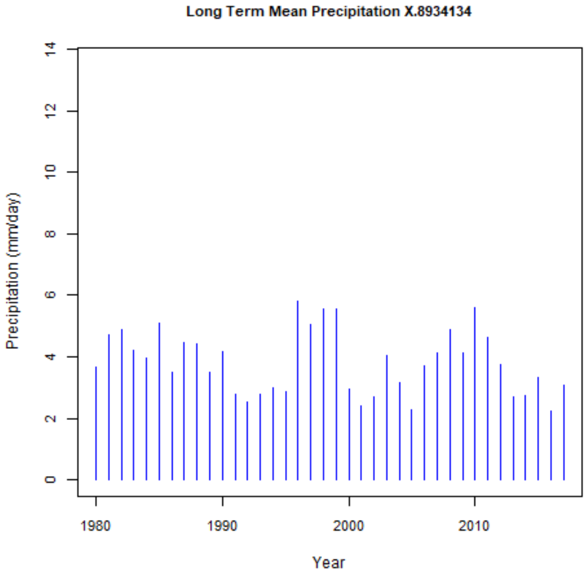
Long Term Mean Precipitation X.8934119

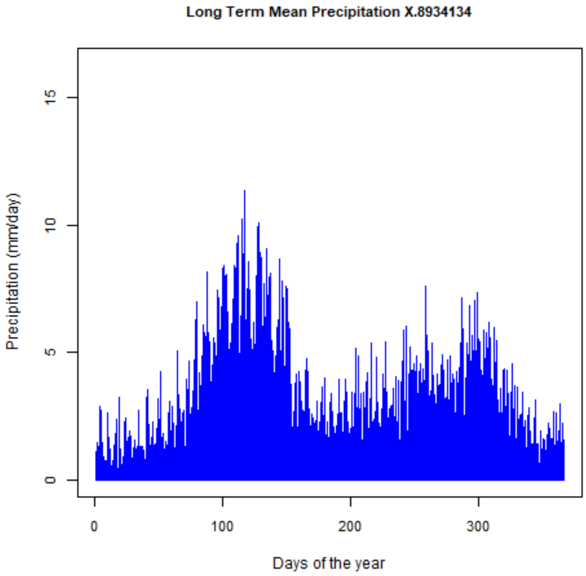


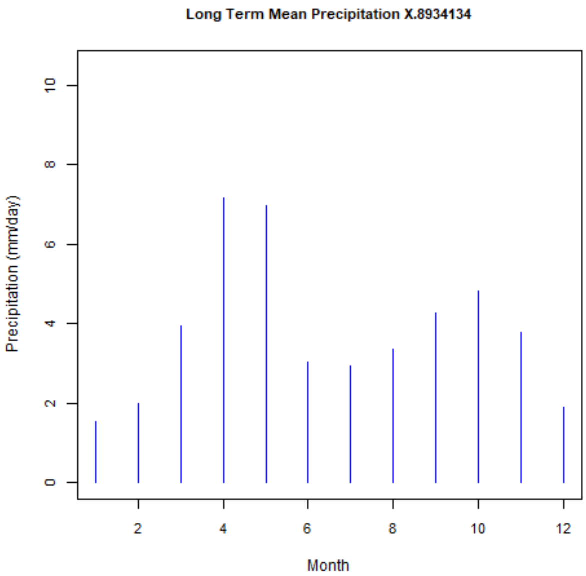
Long Term Mean Precipitation X.8934119

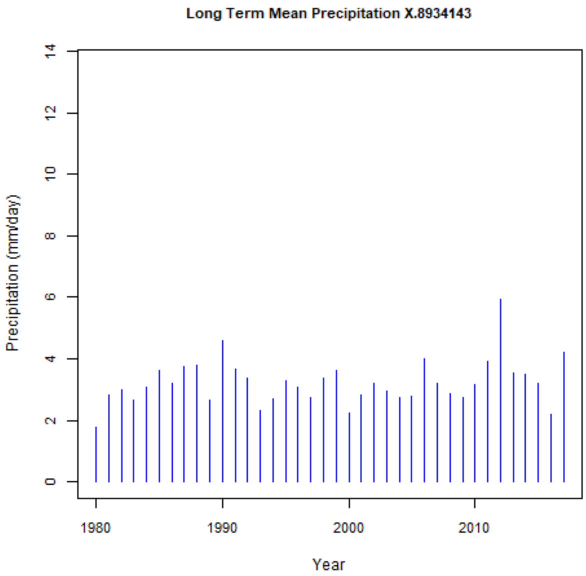


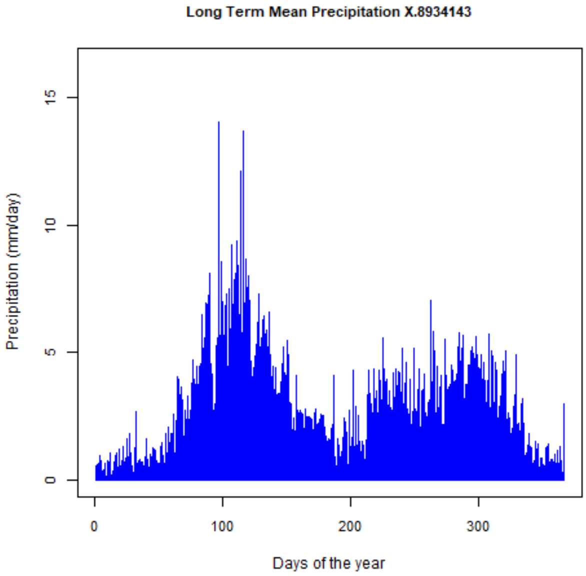

Long Term Mean Precipitation X.8934119
Precipitation (mm/day)
Month
0
2
4
6
8
10
12

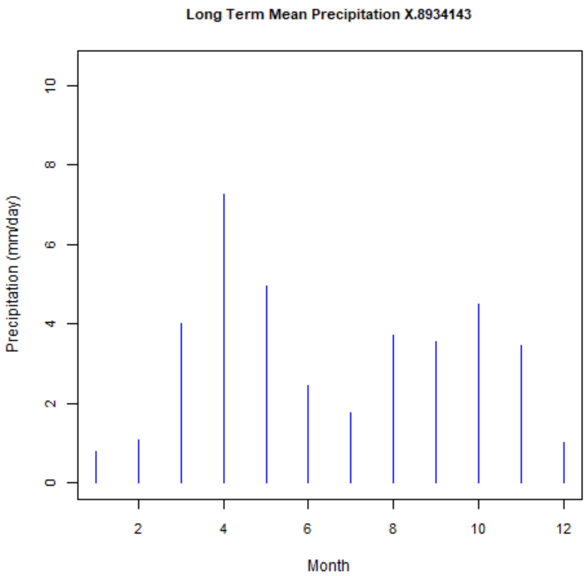

Long Term Mean Precipitation X.8934134
Precipitation (mm/day)
Year
1980
1990
2000
2010

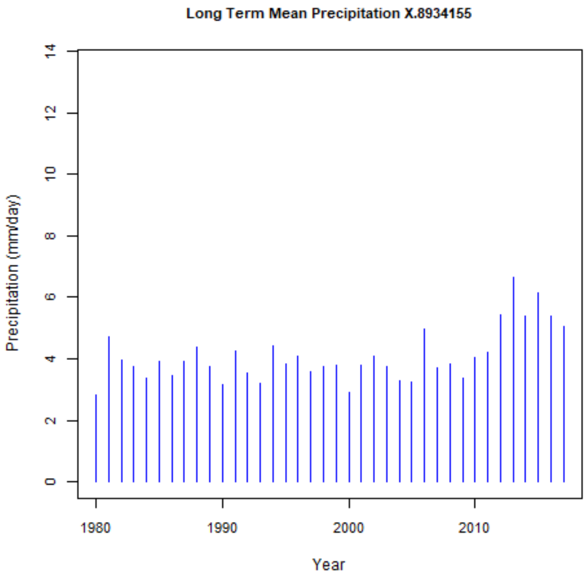

Long Term Mean Precipitation X.8934134
Precipitation (mm/day)
Days of the year
0
5
10
15
100
200
300

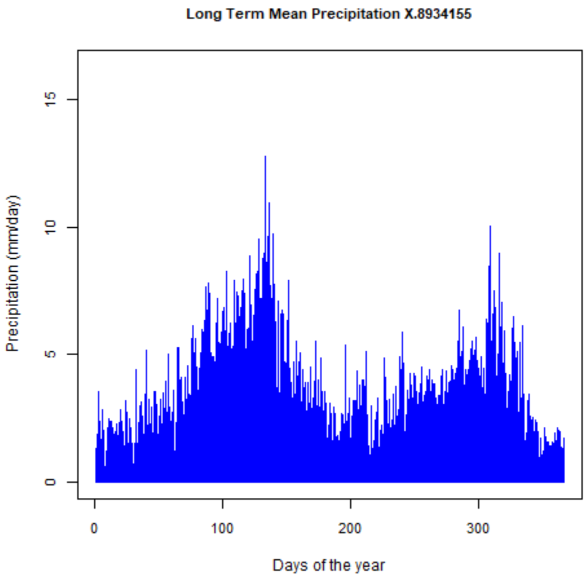

Long Term Mean Precipitation X.8934134
Precipitation (mm/day)
Month

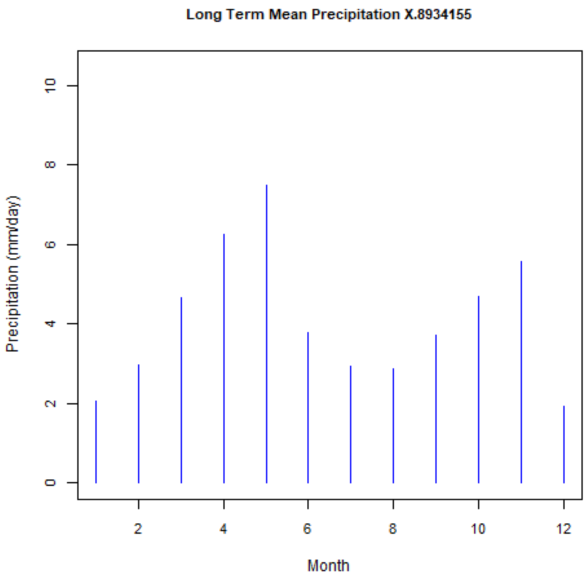

Long Term Mean Precipitation X.8934143
Precipitation (mm/day)
Year

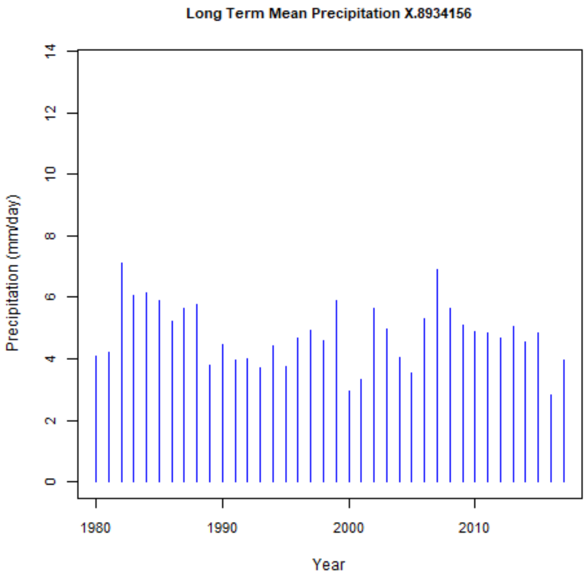

Long Term Mean Precipitation X.8934143
Precipitation (mm/day)
Days of the year

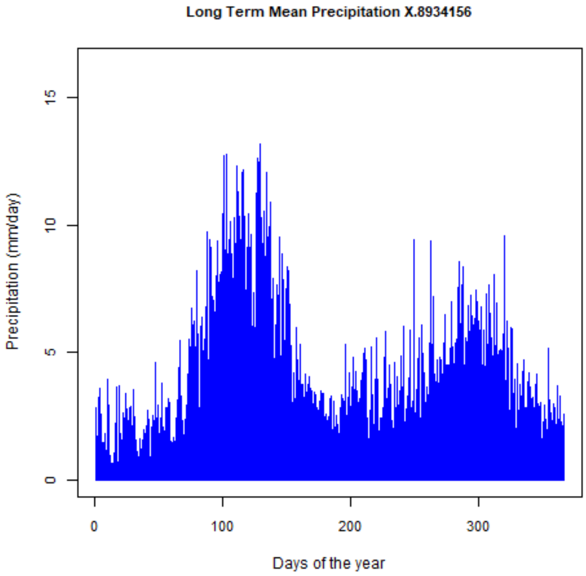

Long Term Mean Precipitation X.8934143
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

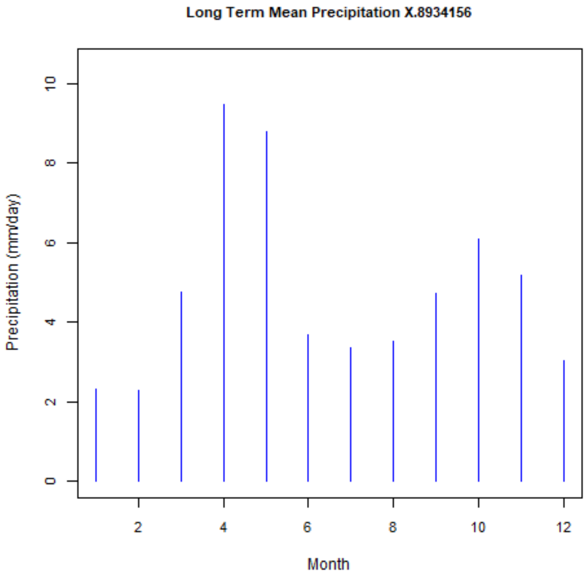

Long Term Mean Precipitation X.8934155
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

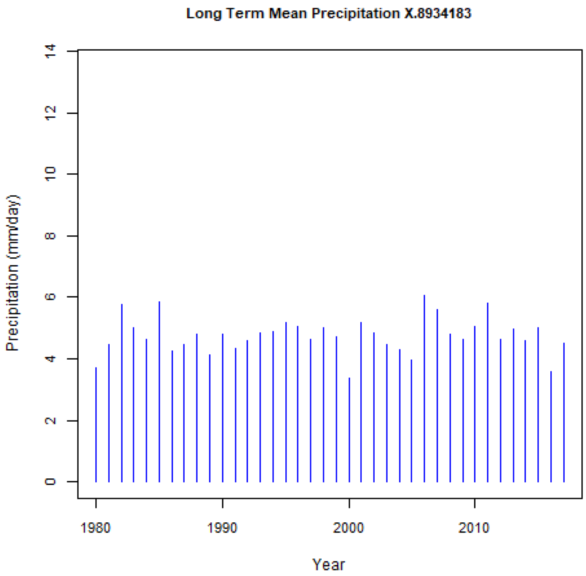

Long Term Mean Precipitation X.8934155
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

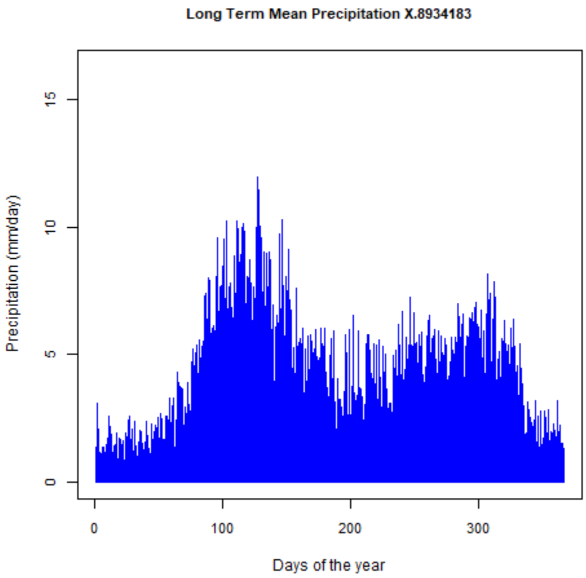

Long Term Mean Precipitation X.8934155
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

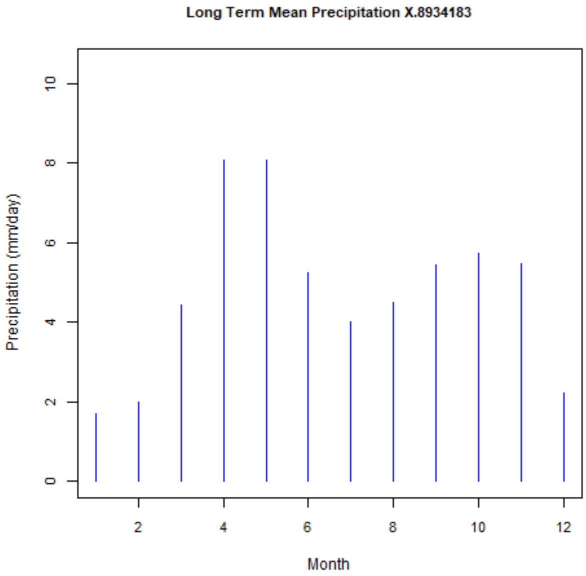

Long Term Mean Precipitation X.8934156
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

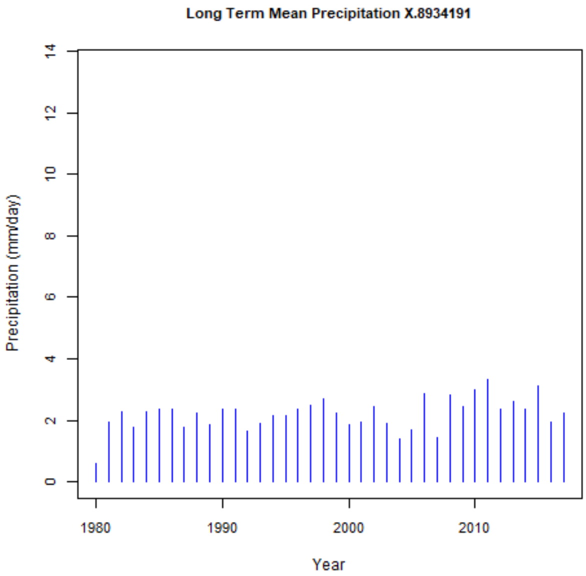

Long Term Mean Precipitation X.8934156
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

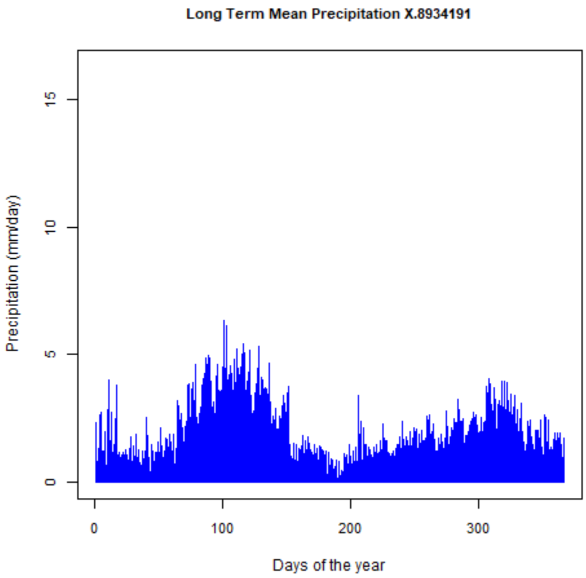

Long Term Mean Precipitation X.8934156
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month


Long Term Mean Precipitation X.8934183
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

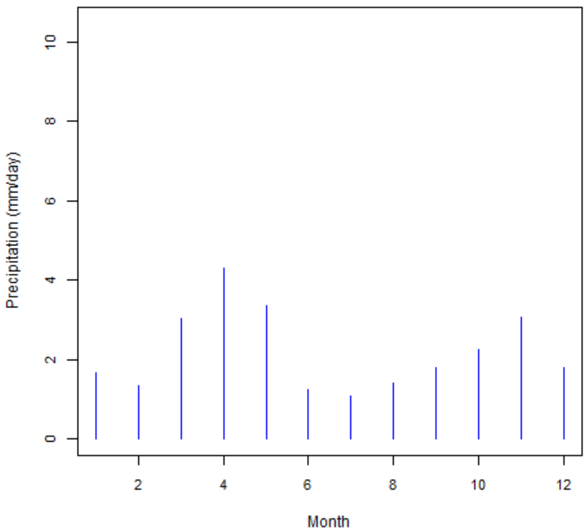

Long Term Mean Precipitation X.8934183
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year


Long Term Mean Precipitation X.8934183
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

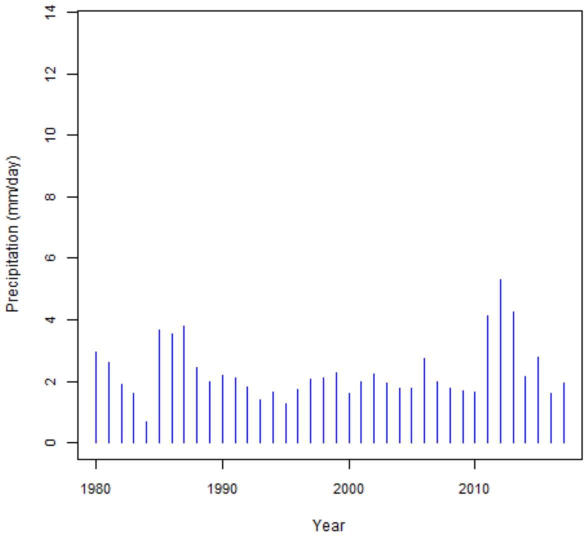

Long Term Mean Precipitation X.8934191
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year


Long Term Mean Precipitation X.8934191
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

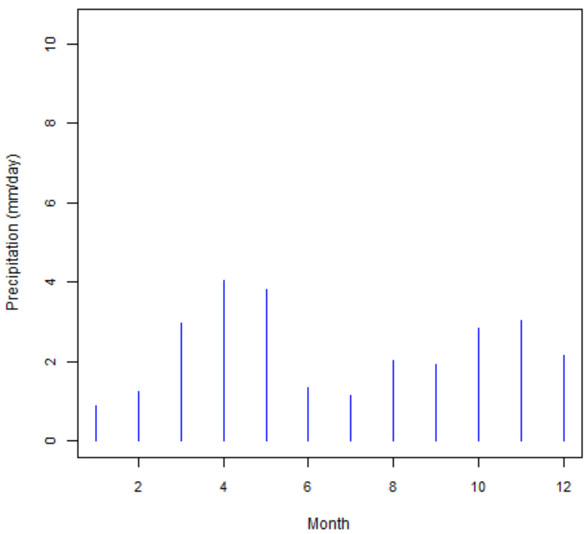
Long Term Mean Precipitation X.8934191

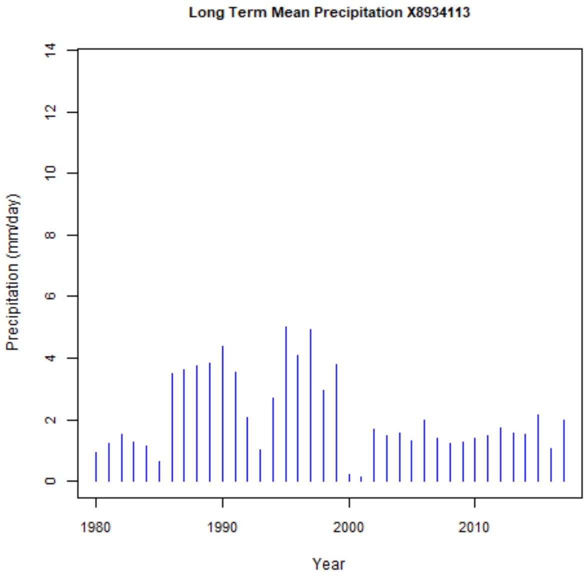


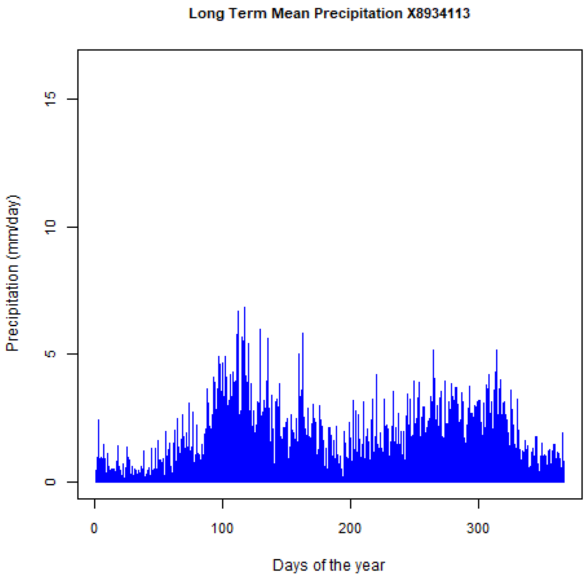
Long Term Mean Precipitation X8934030

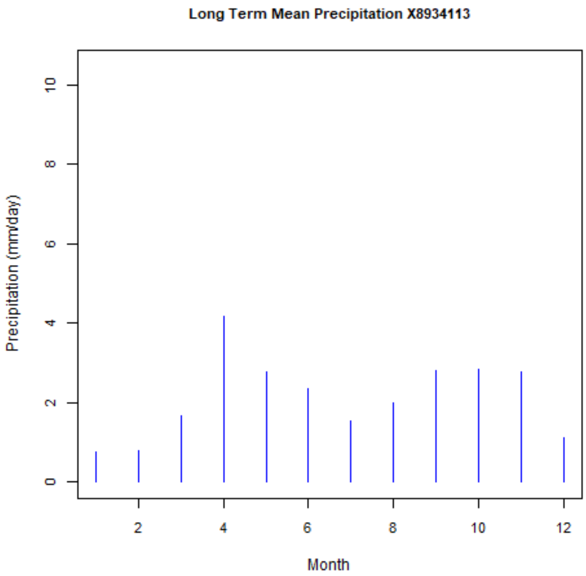


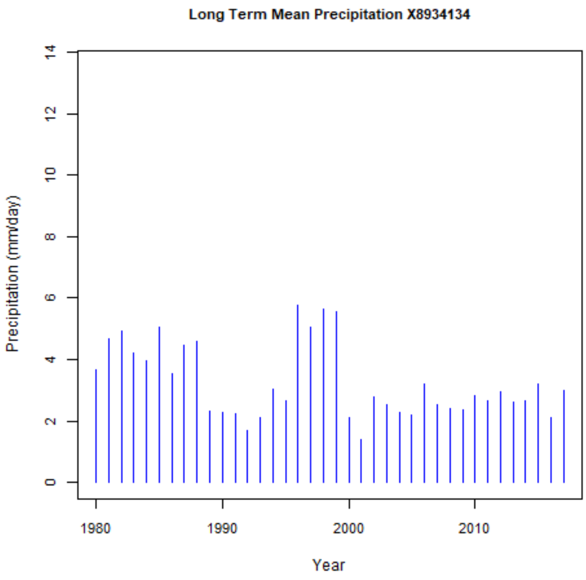
Long Term Mean Precipitation X8934030

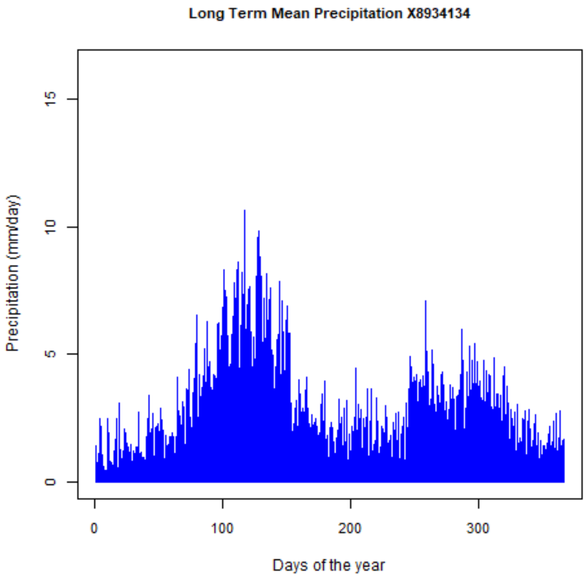


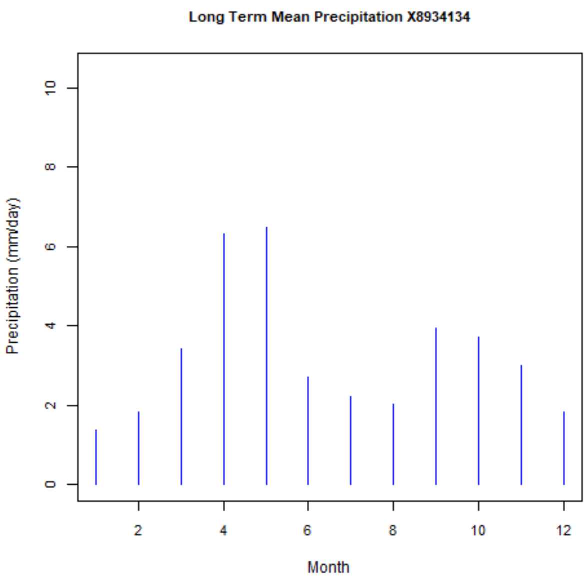

Long Term Mean Precipitation X8934113
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

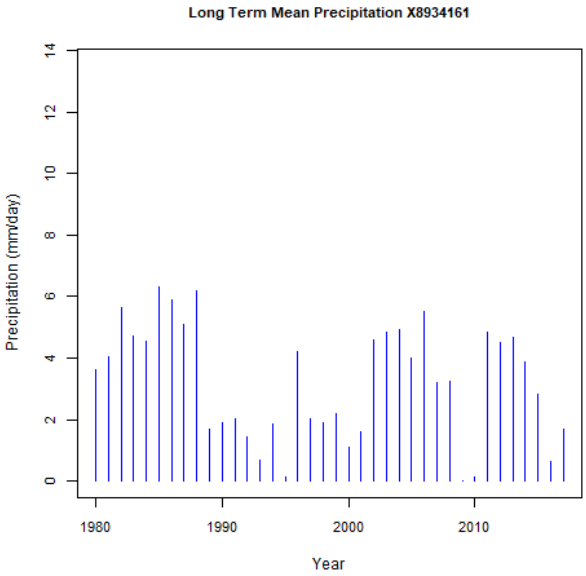

Long Term Mean Precipitation X8934113
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

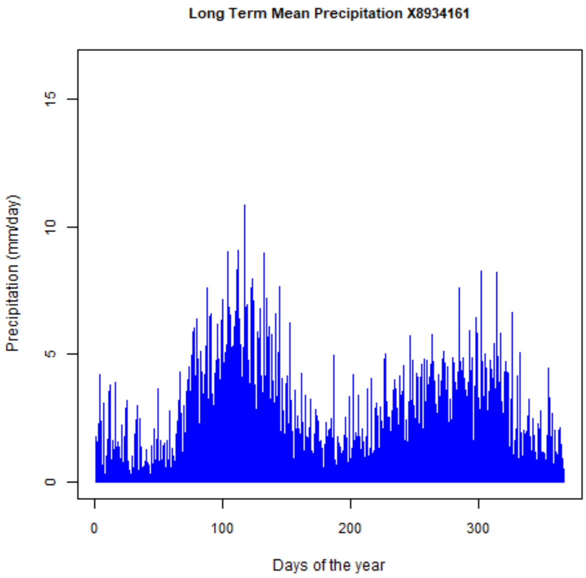

Long Term Mean Precipitation X8934113
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

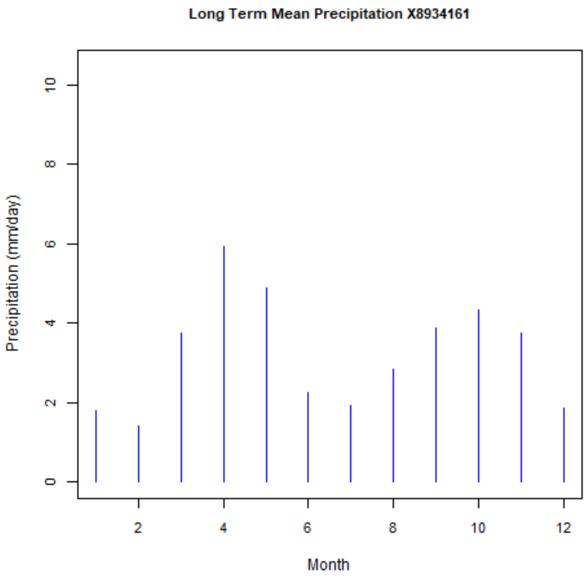

Long Term Mean Precipitation X8934134
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year

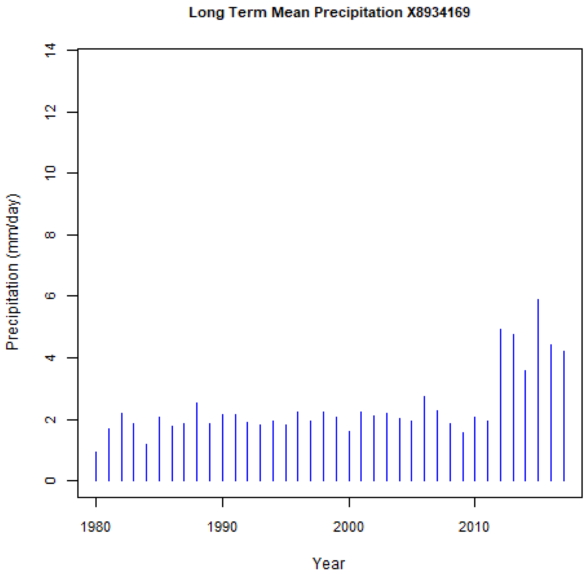

Long Term Mean Precipitation X8934134
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year

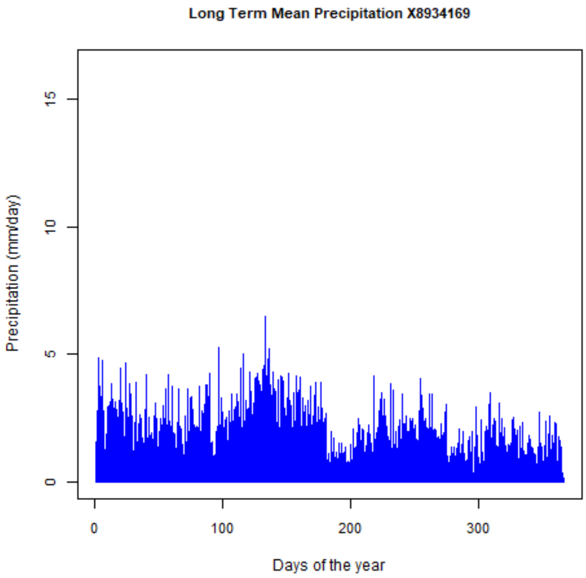

Long Term Mean Precipitation X8934134
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

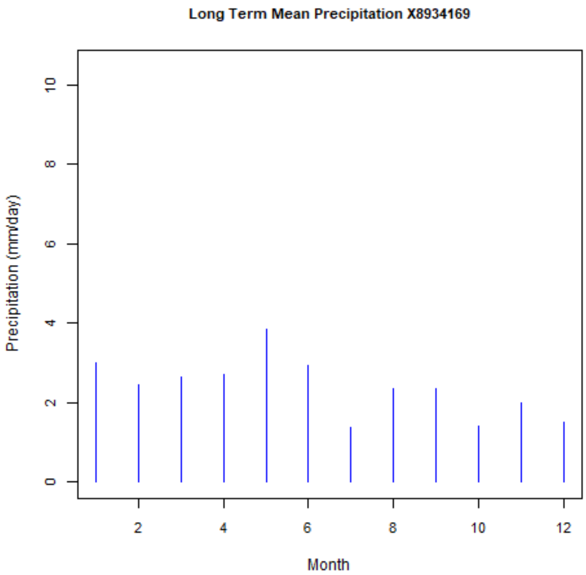

Long Term Mean Precipitation X8934161
Precipitation (mm/day)
Year


Long Term Mean Precipitation X8934161
Precipitation (mm/day)
Days of the year

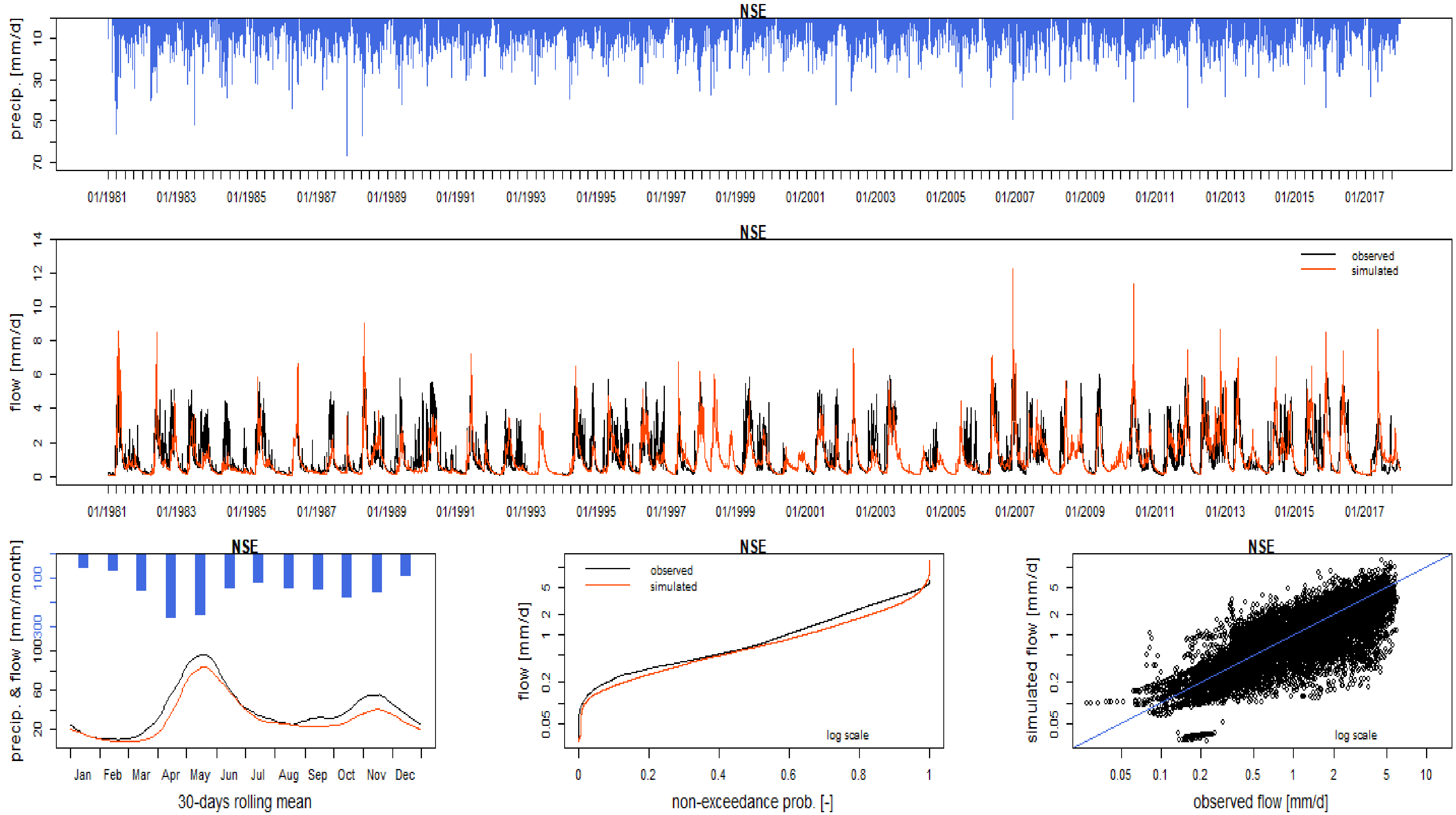

Long Term Mean Precipitation X8934161
Precipitation (mm/day)
Month

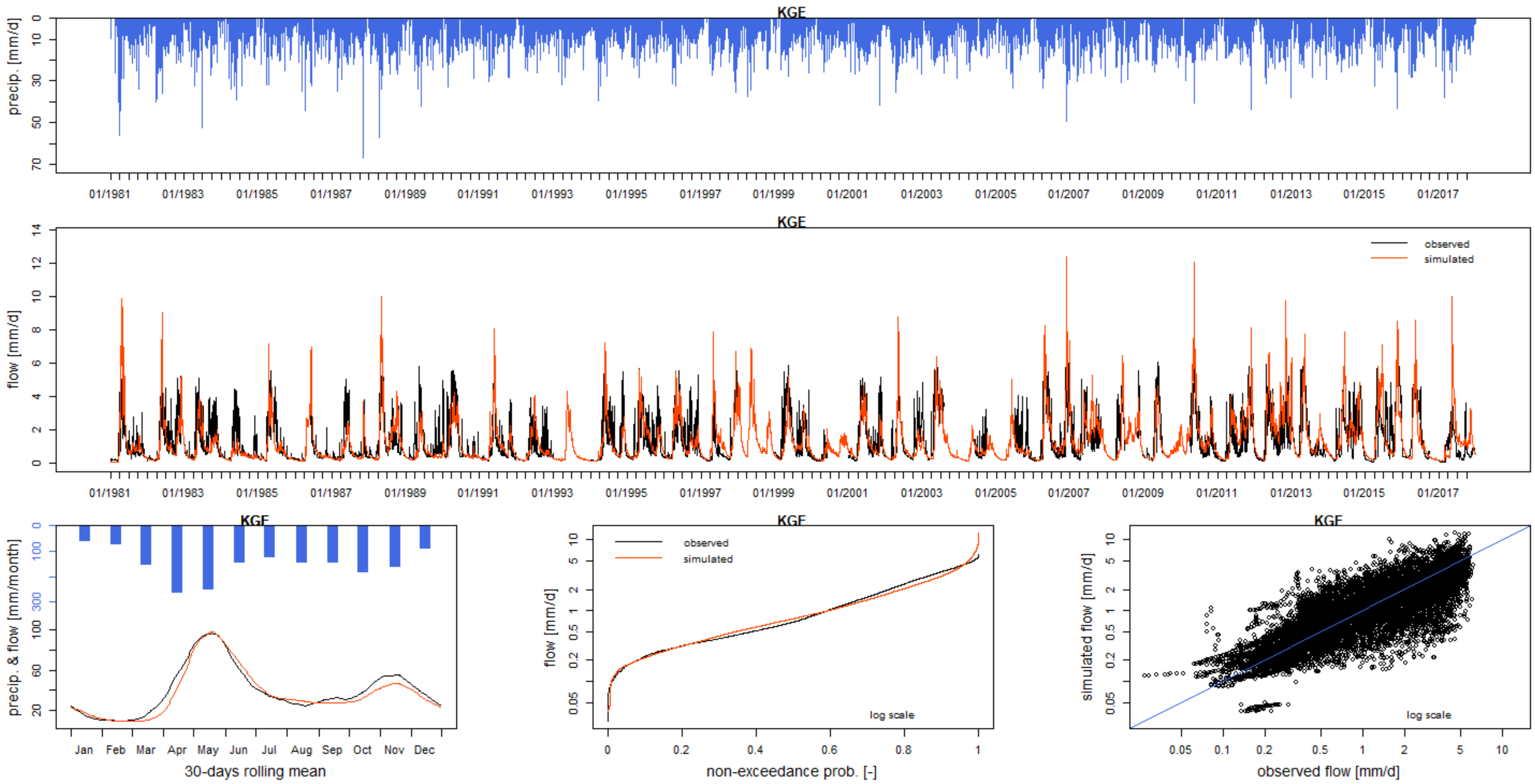

Long Term Mean Precipitation X8934169
Precipitation (mm/day)
0
2
4
6
8
10
12
14
1980
1990
2000
2010
Year


Long Term Mean Precipitation X8934169
Precipitation (mm/day)
0
5
10
15
0
100
200
300
Days of the year


Long Term Mean Precipitation X8934169
Precipitation (mm/day)
0
2
4
6
8
10
2
4
6
8
10
12
Month

7.2 GR4J modelling results with Chirps




KGE
precip. [mm/d]
01/1981
01/1983
01/1985
01/1987
01/1989
01/1991
01/1993
01/1995
01/1997
01/1999
01/2001
01/2003
01/2005
01/2007
01/2009
01/2011
01/2013
01/2015
01/2017
KGE
flow [mm/d]
observed
simulated
KGE
precip. & flow [mm/month]
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
30-days rolling mean
KGE
flow [mm/d]
observed
simulated
log scale
non-exceedance prob. [-]
KGE
simulated flow [mm/d]
log scale
observed flow [mm/d]

				After calibration		ONLY WITH CHIRPS	
Coef	Description	Unit	Range	NSE	KGE	NSE	KGE
x1	Capacity of production soil store (SMA)	mm	100-1200	396.52	177.83	149.97	143.67
x2	Water exchange coefficient	mm	-5 – 3	2.79	2.46	-2.98	-0.32
x3	Capacity of the routing store	mm	20-300	26.14	24.15	300.00	300.00
x4	Time parameter for unit hydrographs	days	1.1 - 2.9	1.25	2.32	1.26	2.17

	Error criterion based on the NSE formula	Error criterion based on the KGE formula	Error criterion based on the NSE formula	Error criterion based on the KGE formula
After calibration	0.242	0.557	0.44	0.71