République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université du 20 août 1955 de Skikda

Faculté des sciences

Département des sciences agronomique

Spécialité : Aménagement hydro-Agricole

Niveau : Master II (M2)

Cours d'hydrologie statistique appliquée

Préparé par Mr. Hebal Aziz (MCA)

Année universitaire 2024/2025

TABLE OF CONTENTS

CHAPTER I

QUALITY CONTROL, HOMOGENIZATION, AND MAXIMIZATION OF HYDROLOGICAL DATA

1.1 Verification of Hydrological Data

1.1.1 Rainfall Data

Precipitation includes all meteoric waters that fall on the surface of the earth, both in liquid and solid forms (drizzle, sleet, hail), and deposited or occult precipitations (dew, frost, hoarfrost...) caused by temperature or pressure changes. The most commonly used precipitation measurement devices are:

- Pluviometers;
- Pluviographs;
- Snow gauges;
- Radars.

Point analysis concerns the series of measurements from a given point, i.e., a single station. From a series of point measurements, two types of graphs can be constructed: the rainfall height curve and the hyetogram.

A - Rainfall Height Curve

This curve represents the monthly or annual precipitation (in mm).

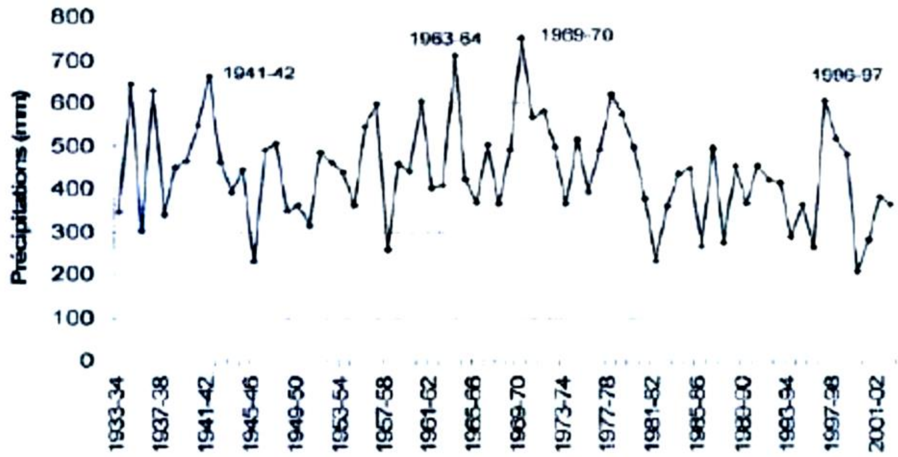


Figure 1.1. Annual precipitation heights recorded at a given station (1933/34-2001/2002)

B – Hyetogram

This is a histogram that represents the intensity of the rain (in mm/h or mm/day) over time. Representing the intensity of the rain requires the acquisition of data over a reduced time step (hours or days).

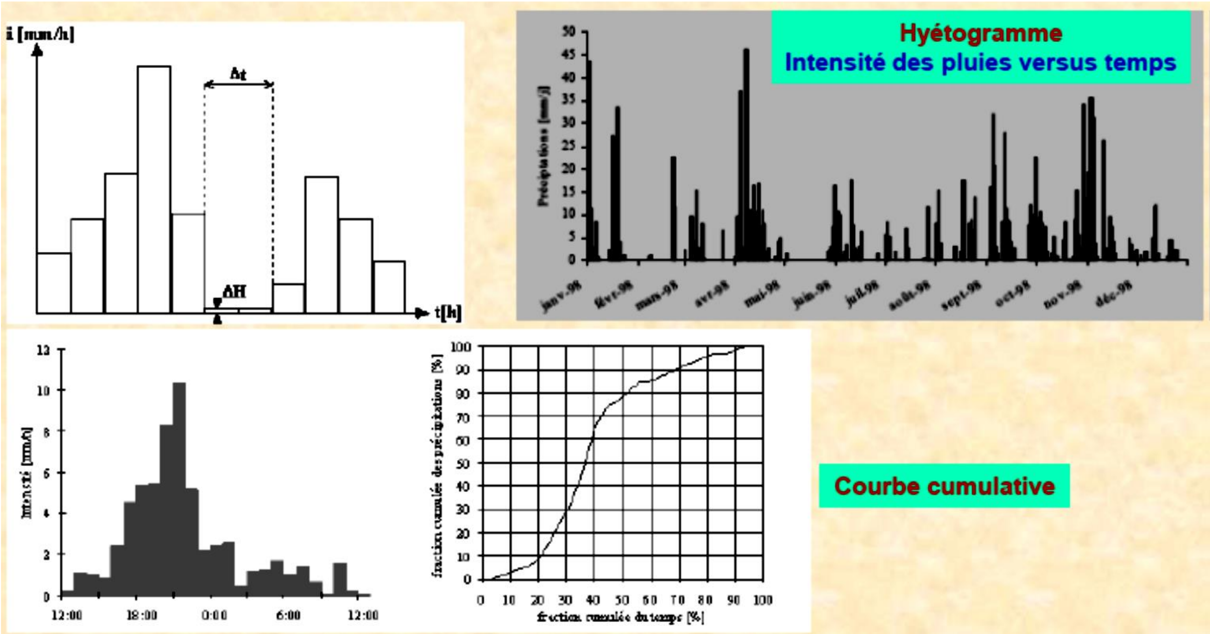


Figure 1.2. Representation of precipitation in the form of a hyetogram

1.1.2 Hydrometric Data

1.1.2.1 Hydrometry

The science that deals with measuring flows and their variation in the bed of a river or stream is called hydrometry. Only the variable flow physically reflects the behavior of the watershed and can be interpreted over time and space. Generally, direct and continuous flow measurements are not available, but rather recordings of water level variations at a given section (hydrometric station). The transition is made from the water level curve over time: H=f(t) (called a limnigram) to the flow curve: Q=f(t) (called a hydrogram) by establishing a rating curve Q=f(H).

A good drawing of a rating curve Q (H) requires a minimum of seven well-distributed gauging's. The determination of this curve is generally carried out through episodic flow measurement campaigns. A gauging is therefore a nearly instantaneous measurement of the flow of a watercourse. The techniques used are numerous and generally complementary; they are based on very different principles depending on the case.

1.1.2.2 Water Level Measurement

Measuring water levels or variations in a water body is carried out at a station that includes at least one height measurement scale (limnimetric scale) and possibly a limnigraph, which provides continuous recording of water level variations in the river over time. The most commonly used tools for measuring water levels are:

- Float limnigraphs;
- Bubble limnigraphs;
- Immersed piezoresistive probes (I.P.P);
- Ultrasonic limnigraphs.

1.1.2.3 Flow Measurement

The techniques used are numerous and generally complementary; they are based on very different principles depending on the case. The most commonly used are:

- Calibrated reservoirs;
- Reel gauging;
- Float gauging;
- Chemical gauging;
- Weir gauging.

The data are provided in the form of instantaneous flows for the study of extreme events such as floods and low flows or as average flows for the study of water supplies.

1.1.3 Types of Errors Encountered in Hydrological Data Series

1.1.3.1 Errors Due to the Sensor

Several errors can be explained by either sensor failure or misuse. Hence the necessity to understand the functioning of the sensor-recorder system used for data acquisition. The sensor may have been modified, either deliberately or not. For example:

- A rain gauge installed 1 meter above the ground was moved to a roof or displaced by a few hundred meters;

- The sensor itself was changed: the transition from a rain gauge or pluviograph receiver cone of 2000 to 1000 or 400 cm^2;
- A limnimetric scale was disassembled and then reattached with a one or two cm discrepancy;
- A radiation sensor was never changed, but the black body aged (suggesting a slight decrease in radiation);
- A limnimeter gradually became clogged, etc.

1.1.3.2 Environmental Condition Changes

In addition to significant sensor displacements:

- Transferring a station from its initial location to another place a few hundred meters away;
- Transferring a station from a low to a higher location (a few 100 m in altitude);
- Vegetation development near the sensor (a curtain of trees near a rain gauge);
- Brush in the river bed under an ultrasonic level sensor;
- Soil condition changes (lawn becoming a parking lot, etc.);
- Recalibration work in the river bed near a station (which itself remains unchanged);
- Urban development around a weather station (temperature, radiation), etc.

1.1.3.3 Errors Related to Certain Measurement Conditions

These are the most difficult to detect because they do not occur systematically but in certain sometimes random situations:

- In a totalizer rain gauge collecting precipitation in liquid or solid form, the collection capacity will depend on this form of precipitation (e.g., 90% of rain but only 50 to 80% of snow due to wind sensitivity). Yet the final data no longer contain information on the form of precipitation or the presence/absence of wind;
- Radiation measurements assume a clean device: however, it may be covered with dew, rain, or even snow (still giving a 'measurement'). The same applies to an anemometer that will be covered with frost but still turn;
- Electronics (or mechanics) may have variable responses depending on the temperature (case of piezometric probes measuring levels...), but this is not recorded in parallel.

1.1.3.4 Errors in Processing and Transcription

These are errors related to data processing and information transfer. Among others, the following examples can be cited:

The case of random accumulations in rainfall series: Due to failure to monitor the device.

Example: Rain that falls on days D and D + 1 is entirely affected to D + 1:

08h	08h	08h	erroneously gives:	08h	08h	08h
← D →	← D+1 →			← D →	← D+1 →	
22 mm	31 mm			0 mm	53 mm!	

Conversely, in a non-heated recording rain gauge, snowfall that occurs once will melt with the return of good weather over the following days, making them appear erroneously as rainy days.

- Case of limnigraphic recordings (levels) to be converted to flow rates using a rating curve:
  - The rating curve has changed over time (modification of the section), but the old rating curve is still used;
  - Or the algorithm for adjusting the rating curve is suddenly changed, which in practice modifies it “strongly”;
  - Or there are different rating curves according to periods, but it is not exactly known when (for which flood?) to switch from one to the other;
- The calibration of a rain gauge is changed (correction according to the measured intensity, especially in high values > 40 mm/h);
- Moreover, and especially, there is a possibility of error with each new transcription:
  - From the device diagram to the slip sent to the central administration;
  - From station reports to monthly or annual reports;
  - From paper document to digital acquisition, etc.
- Protocol changes: transition from average data, integrated over time steps, to “instantaneous" data read at the measurement time (or vice versa).

Examples:

- Case of hourly flow rates: average flow 8h-9h or instantaneous flow read at 9h?
- Case of radiation or wind: measurement for 1 minute at 9h or cumulative from 8h to 9h?

1.2 Homogenization of Hydrological Data

1.2.1 Nature of heterogeneities

- Absence of data: This is the most apparent heterogeneity, remedied by extending data from a failing station B using a correlation between data from this station and those from a reference station A;
- Defective data: If readings have not always been taken under good conditions, they may be eliminated (leading back to the previous case for one or more periods) or corrected to make them usable to a certain extent;
- Data corresponding to two or more non-defective homogeneous series grouped under the same station name: This occurs if the rain gauge has been relocated or the immediate environment has changed.

Consequently, hydrological information :

- May contain erroneous observed values;
- May consist of non-homogeneous series;
- May lack some observed values;
- May be too short to extract significant statistical parameters.

Thus, before any hydrological study, it is recommended to check whether the data series is homogeneous (the sample is indeed part of the same population or not). It is necessary to do:

- A careful “eye” examination of slips and data files;
- Graphic, numerical and statistical tests are essential to highlight the existence of systematic errors.

The objective of any analysis and quality control of rainfall data is to form reliable and homogeneous series over the longest possible period.

Several rainfall data analysis methods can be used :

- Error control: statistical control tests (Median Pettitt, Hubert's segmentation procedure, Buishand, Lee and Heghinian, Wilcoxon, Student, Fisher-Snedecor test, PCA, etc.) and graphical tests (Bois, Double cumulative, Residual cumulative, etc.) that can be internal or relative;
- Correction of heterogeneities: graphical and numerical homogenization methods;
- Reconstitution or filling of missing data: estimation approaches or simple linear regression;
- Extension or maximization of short series: correlation approach.

1.2.2 Graphical Control and Homogenization Methods

The graphical representation of the annual rainfall time series provides an idea of the rainfall trend and highlights the excesses and deficits of rainfall (wet and dry years) recorded at a given station.

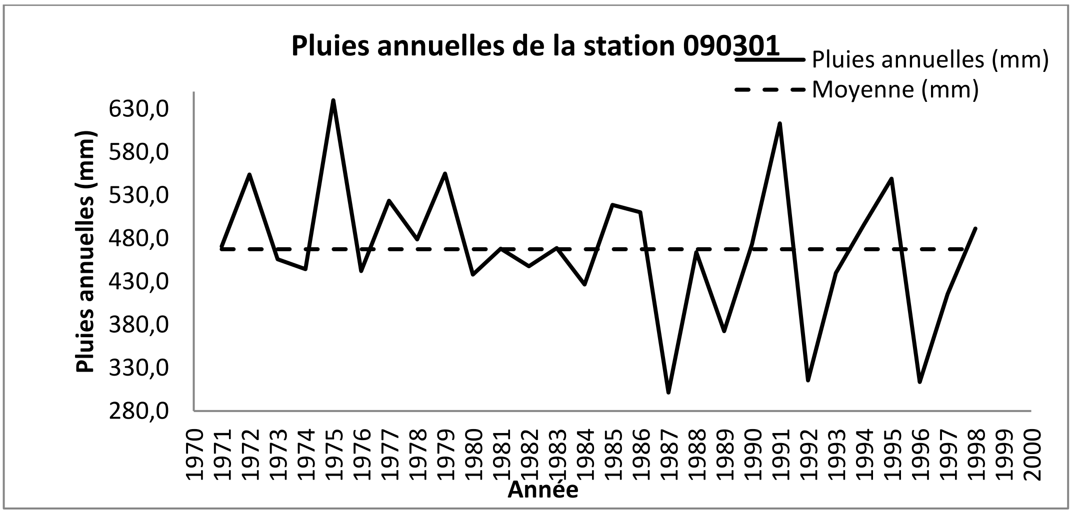


Figure 1.3: Graphical representation of the annual rainfall at station 090301

Graphical techniques exist to verify and confirm the existence of heterogeneity in a rainfall series. Among these methods are:

1.2.2.1 The method of cumulative deviations

By calculating the quantity Z_i as : $Z_i = \sum_{j=1}^{i}(X_i - \bar{X})$, on the observed annual rainfalls (X_i) with arithmetic mean X, and plotting the variation of Z_i over the years i, anomalies in the series can be detected when the Z_i curve does not oscillate regularly around the general mean and deviates too much in a given year or period. This can result from systematic errors or accidental fluctuations. The same process can be applied with respect to the standard deviation. The method does not provide proof but at most an indication that should raise attention.

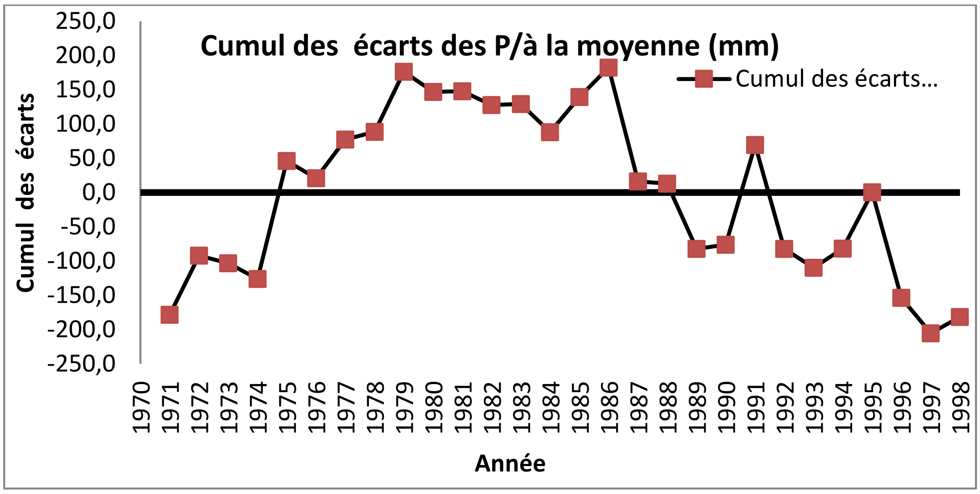


Figure 1.4: Cumulative deviations between annual rainfall (P in mm) and interannual mean (Pm in mm)

1.2.2.2 The method of double cumulative

Homogenization by this powerful and widespread graphical technique requires knowledge of an annual and homogeneous data series observed at a reference station, called a benchmark or base station, neighboring and regional with the station to be corrected.

The double cumulative method is a two-dimensional method used to assess the presence of an anomaly in the series studied to correct it.

A. Principle of the method

It involves comparing the trend of the studied station with that of the benchmark station by plotting the graph of cumulative data at the studied station against cumulative data at the benchmark station.

The method is based on the following principle : In the absence of an anomaly, two neighboring stations A and B measure annual rainfall in a relatively constant ratio from year to year, regardless of whether the year is dry or wet.

That is :

PA (i) /PB (i) = constant and is practically independent of the year i

Consequently, the points M(i) of coordinates, the cumulative rainfalls calculated at each station A and B up to year i, are practically aligned. However, if a systematic error occurs at the studied station, the double cumulative line would show a break in its slope at the year of the error introduction.

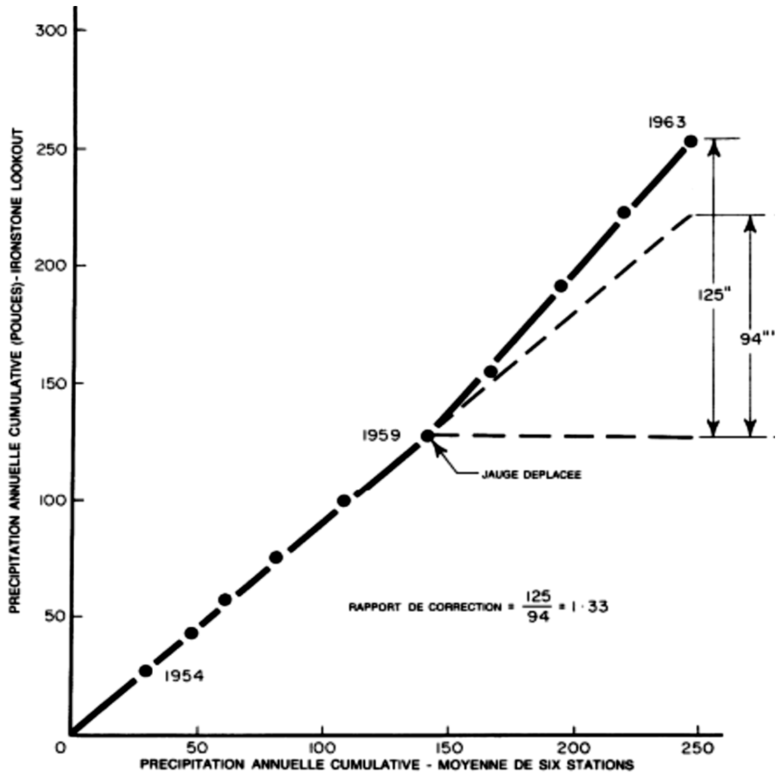


Figure 1.5: Example of applying the double cumulative method

B. Homogenization Procedure

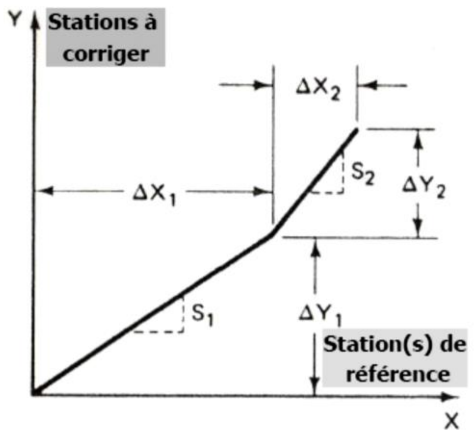
When breaks are present, two questions arise :

- Which portion of the graph should be corrected, i.e., which values of the studied post should be corrected?
- How to proceed with the correction?

For the first question, the station's history is reviewed to find indicative elements that may facilitate decision-making on the period from which errors in measurements are suspected. If no indicative element is available, the most recent data are considered the most reliable.

Regarding the second question, the correction procedure for the unreliable portion of the graph is done by extending the most reliable slope according to the formula:

- $P_{Corrected} = (S_{Adjusted} / S_{Observed}) * P_{Observed}$
- $P_{Observed}$: the measured precipitation.
- $P_{Corrected}$: the corrected precipitation.
- $S_{Adjusted}$: the slope of the reliable graph portion.
- $S_{Observed}$: the slope of the graph portion to be corrected.



$$Y_1' = \frac{S_2}{S_1} \cdot Y_1$$

$$S_i = \frac{\Delta Y_i}{\Delta X_i}$$

If correcting the lower part

$$Y_2' = \frac{S_1}{S_2} \cdot Y_2$$

If correcting the upper part

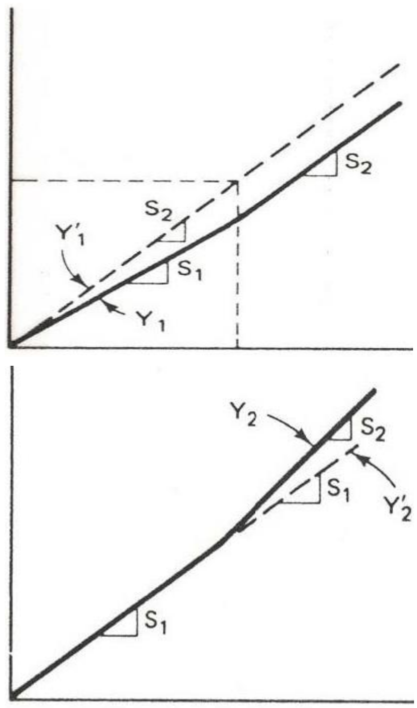


Figure 1.6: Principle of the double cumulative method

C. <u>Choice of a Homogeneous Base Series</u>

It is necessary to find in the region either a post or preferably several posts observed as long as possible to be compared with all the others, among which those that do not seem affected by any heterogeneity will be detected. The average cumulative totals of the selected posts form the homogeneous base series, compared to which the anomalies of other posts in the region are detected and corrected.

Choosing a base group requires preliminary pairwise comparison of all posts likely to be integrated into this group. If no heterogeneity is detectable, the group can be formed. Otherwise, limited groups can be formed in sub-regions where random fluctuations around the connecting lines are smaller, and anomalies are easier to detect. Finally, one base post per region or sub-region can suffice.

The advantage of a group is to attenuate accidental irregularities affecting any given year, thus facilitating the detection of true anomalies.

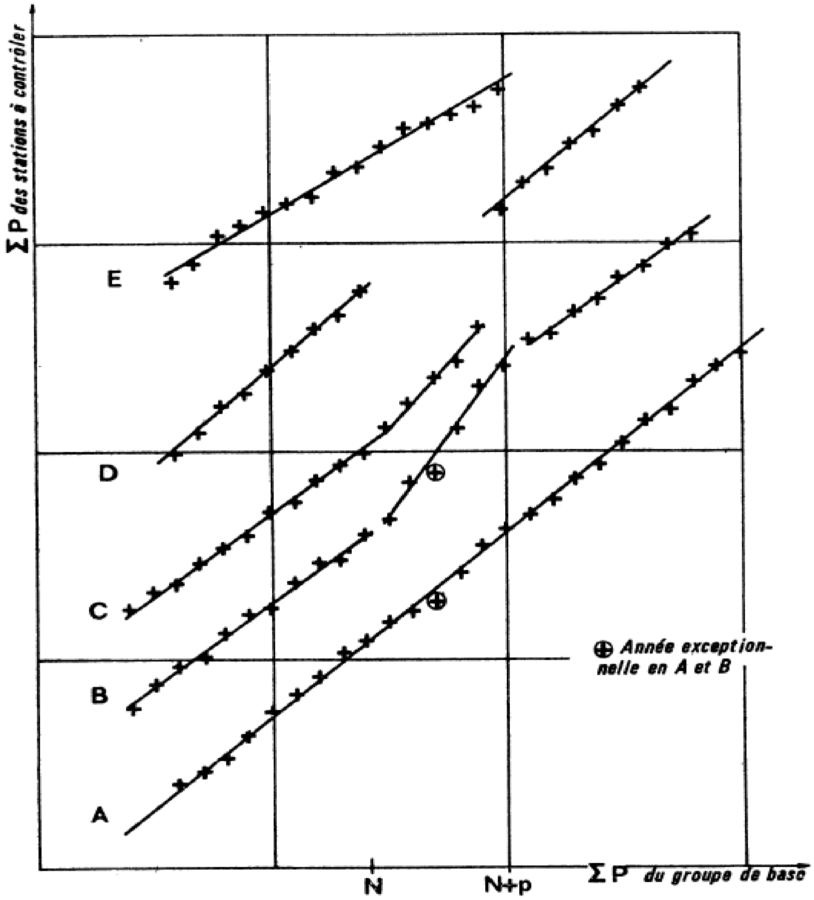


Figure 1.7: Regional application of the method of comparing cumulative annual totals

The base series chosen for the systematic study of all posts in a region can be grouped in a single comparison graph of cumulative totals, with the base series plotted on the x-axis and each vintage indicated opposite the corresponding interval. The graphs can be shifted by changing the origin on the y-axis to avoid crossing their adjustment lines. By judiciously choosing the order of plotting the different graphs and grouping posts by geographical or climatic sub-region, common features will appear in the series of posts in the same sub-region: the same pattern of fluctuations around the adjustment line of the broken line connecting representative points.

This is illustrated in Figure 1.7, where five stations are reported compared to a base group:

- **A** homogeneous station.
- **B** station with two clear breaks, restoring the most frequent slope.
- **C** station with a clear break.
- **D** station with a gap offering two periods of the same slope, hence homogeneous.
- **E** stations with a particular rainfall regime (high point dispersion) but appearing homogeneous.

It is also noted that the years between year N and year N + p are the most disturbed (breaks of B and C, gap of D) due to significant changes in service management. Some exceptional years (circled points) appear here and there without affecting the periods' homogeneity. This general graph for the whole region highlights heterogeneities while maintaining almost the same severity for all posts, considering only the breaks that significantly affect the graph's appearance more than the local and accidental fluctuations mentioned earlier.

In practice, it is advantageous to calculate cumulative totals by going back in time from the last year

1.3 Filling Observation Gaps

It is important to have complete data series for frequency analysis. Data collection is costly. Events may have forced the interruption of measurements.

Table 1.1: Example of a station illustrating the extent of missing data (Station 090301 in the Isser basin (09))

année	sept	oct	nov	déc	janv	févr	mars	avr	mai	juin	juil	août	Annuel
1971	******	3	136.3	35.4	196.8	65.3	57.6	57.3	73.6	21.6	0	0	*****
1972	28.6	47.7	34.8	121.5	81.8	115.9	34	50	0	16.6	0	22.9	553.8
1973	51.1	1.9	7.6	71.2	1.6	135.8	172.5	69.8	14.8	23.5	0	*****	*****
1974	31.3	32.9	82.7	2.8	21.9	73.2	94.1	7.5	62.7	29.8	0	5.6	444.5
1975	42.8	0	127.8	26.5	45.8	125.4	55.6	79.6	63.3	28.7	44.5	0	640.0
1976	57.9	39.4	*****	46.3	63.6	11.5	5.5	68.7	31.1	11.8	0	21.6	*****
1977	*****	*****	88.4	*****	*****	11.1	43.1	72.7	81.7	0	0	0	*****
1978	0	67.3	49.4	14	65.4	121.9	50	41.9	18.1	48.8	0	1.9	478.7
1979	81.3	59.8	62.4	34.2	89.6	22	80.2	98.9	24.3	0	0	2.5	555.2
1980	10	19.3	46.4	147.4	24.2	60.3	34	60.7	13.1	2	0	20.7	438.1
1981	13.3	17.4	8.8	68.8	160.6	50.4	30.3	36.5	43	21.2	0	17.8	468.1
1982	20.8	0	145.3	63.9	0	40.2	12.2	7	5	0	12	*****	*****
1983	0	10.8	21	54.2	71.1	*****	*****	43.1	33.9	24.3	0	1.5	*****
1984	5.5	61.1	37.1	99.3	34	32.1	78.8	16.6	60.9	1.1	0	0	426.5
1985	38	19.6	46.6	49.7	57	39.6	217.3	16.8	9.4	20.7	0	3.9	518.6
1986	34.3	44.2	63.3	106.5	76.5	114.5	8.7	1.5	13	36.7	5.5	5.7	510.4
1987	7.5	34.3	55.2	39.3	14.1	21.2	42.1	53.7	14.2	19.8	0	0	301.4
1988	31	18.1	12.7	152.5	16.6	19.7	31.1	105.1	3.1	25.6	24.8	23.7	464.0
1989	2.6	8	7.2	7.2	61.9	0	25.6	65	96.9	31	42.5	24.5	372.4
1990	7.3	37.3	34.5	77	31.6	102.5	104.9	21.8	18.8	4.9	9.2	23.4	473.2
1991	9.9	62.4	33.6	9.3	137.8	19	50.5	162.9	115.8	0	0	12	613.2
1992	11.8	12.1	82.7	28	43	35.5	16.5	48.8	17.7	4.8	0	14.7	315.6
1993	78.8	14.5	45.4	73.4	83.8	49.1	3	122.7	*****	0	3	0	*****
1994	40.4	59.5	23.2	8	221.9	16	114	0	0	12.5	0	0	495.5
1995	7	*****	74.5	7.1	69.9	129.8	29.2	143.4	33.9	5.7	48.8	0	*****
1996	14	9.5	27.9	23.1	57.2	18	9.1	67.6	31.2	5.8	0	50.1	313.5
1997	40.3	50.3	*****	51.4	10	64.2	23.4	55.2	110.9	2.9	1.2	5.8	*****
1998	37.8	43.1	139.6	55.1	49.5	51.5	75.2	*****	30.5	2	1.2	*****	*****

***** : Missing value.

To use these series, we must first fill the existing gaps in the collected series. Several techniques have been used in hydrology for this purpose. The most commonly used are listed below.

1.3.1 **Method of the Average of Surrounding Stations**

This method is based on the arithmetic average of the known values of the surrounding stations to the one to be filled.

$$\hat{X} = \frac{1}{n_s}\sum_{i=1}^{n} X_i \quad \textbf{(1-1)}$$

Where :

- X : Missing value at the station to be filled ;
- Xi : Known value at station i ;
- n_s : Number of stations around the station to be filled.

The annual average value of the station to be filled must not differ by more than *10%* from the annual average value of each of the other stations.

1.3.2 **The Normal Ratio Method**

Same approach as the previous one, but it is based on a weighted value.

$$\hat{X} = \frac{1}{n_s}\sum_{i=1}^{n} w_i X_i \quad \textbf{(1-2)}$$

Where:

wi: Weighting coefficient for station i, given by :

$$w_i = \frac{A_i}{nA_c} \quad \textbf{(1-3)}$$

Where:

- Ai: Annual average value of station i ;
- Ac: Annual average value of the stations around the station to be filled.

1.3.3 **The Ratio Method**

This method is often used to compensate for gaps in observations or to correct heterogeneities in the series. This method is chosen in the case of a short observation period.

The ratio method uses the ratio between totals or averages of values of two concurrent series as a multiplier for the total or average of the available series. In other words, it is the ratio of the values of a considered deficient month during a series or the month is the same in the different stations for gap filling on a monthly scale, for example.

The equation of this method is:

$$\mathbf{Y=b\ X} \quad \mathbf{(1\text{-}4)}$$

Where :

- Y : Value of a given month at the station to be filled ;
- X : Corresponding value at the reference station ;
- b : Adjustment constant estimated from the ratio:

$$\mathbf{b=V/U} \quad \mathbf{(1\text{-}5)}$$

Where :

- Vs: Sum of values during a homogeneous period at the station to be adjusted ;
- U: Sum of values for the corresponding period at the comparison station.

Subsequently, a matrix of correlation coefficients between stations on a monthly and annual scale is developed. The interest of this matrix highlights affinities for the different stations, generally by couple, so that the two are comparable to each other.

If all conditions are met, we proceed to estimate the missing data for each month by applying equation (1-4).

1.3.4 The Linear Regression Method

In statistics, given a random sample (Y_i, X_i), $i= 1, 2 \dots, n$, a simple regression model assumes the following affine relationship between Yi and Xi:

$$\boldsymbol{Y_i = aX_i + b \quad i = 1, \dots\dots, n} \quad \mathbf{(1\text{-}6)}$$

Linear regression consists of determining an estimate of the values a and b and quantifying the validity of this relationship using the linear correlation coefficient r.

Using the least squares method, we obtain :

$$\boldsymbol{r} = \frac{\sum \boldsymbol{x_i y_i} - \boldsymbol{n\overline{xy}}}{(\boldsymbol{n}-\mathbf{1})\boldsymbol{\sigma_X \sigma_Y}}$$

$$\boldsymbol{a} = \boldsymbol{r}\frac{\boldsymbol{\sigma_Y}}{\boldsymbol{\sigma_X}} \qquad \textbf{(1-7)}$$

$$\boldsymbol{b} = \overline{\boldsymbol{y}} - \boldsymbol{r}\frac{\boldsymbol{\sigma_Y}}{\boldsymbol{\sigma_X}}\overline{\boldsymbol{x}}$$

Where :

- x : Arithmetic mean of x_i ;
- y : Arithmetic mean of y_i ;
- σ_X : Empirical standard deviation of x_i ;
- σ_Y : Empirical standard deviation of y_i ;

It can be demonstrated that r is always between -1 and +1. In practice, its absolute value is rarely equal to 1, but it is generally estimated that the fit is valid as soon as this coefficient is close to +1 or -1. The r2 called determination coefficient is often used. In this case, the fit is valid as soon as this coefficient is close to 1.

1.3.5 Means of Appreciating the Gain Obtained by Extension

The benefit of extending series Y using series X for knowledge of series Y is greater when the correlation coefficient $k^r{}_{xy}$ is high. This benefit was translated by R. VERON into relative efficiency E.

$$\boldsymbol{E} = \mathbf{1} + \left(\mathbf{1} - \frac{\boldsymbol{k}}{\boldsymbol{n}}\right)\left[\frac{\mathbf{1} - (\boldsymbol{k}-\mathbf{2})\boldsymbol{r}^{\mathbf{2}}}{(\boldsymbol{k}-\mathbf{3})}\right] \qquad \textbf{(1-8)}$$

- $r = {}_k r_{xy}$: Correlation coefficient calculated over k years.
- E: Relative efficiency of $\overline{y_k}$ and $\bar{\hat{y}}$ defined by the ratio of the variance of $\bar{\hat{y}}$ to that of y_k.

This benefit is translated using E in the form of a real gain of information expressed using the number of effective or fictitious years n' to which the extended sample y corresponds.

n' varies from k to $\boldsymbol{n'}$(maximum gain functional relationship between x and y and r=1).

$\boldsymbol{n'} = \frac{\boldsymbol{k}}{\boldsymbol{E}}$ with $k \geq 3$ (it is unthinkable to perform a regression with less than 4 values). A new series observed over n' years is obtained.

1.3.6 Conclusion on the Maximization of Data in Hydrology

Depending on the time scale adopted for the study of hydrological variables (rainfall or flows) and the nature of the climate and the local flow regime, the form of cause-effect relationships (between precipitation and flow) or concomitance relationships (two neighboring rainfall stations or flow of two neighboring watercourses) is more or less complex and requires taking into account conditional variables in more or less large numbers.

Whatever the problem, there is a solution method that leads to extending, improving, or creating a new sample of the studied variable. But one must be careful with the use of a correlation between $\boldsymbol{k}$ pairs of values; it is only reliable within the observed interval; beyond that, any extrapolation is hazardous. From this point of view, graphical methods more easily avoid automatic extrapolation than numerical methods. This should not be neglected. Taking into account the (or the) precipitation or flow factor observed over the longest known period makes it possible to affirm that the new sample is indeed the result of maximizing the information. In practice, the extension is mainly in the precipitation-flow direction.

The obtained variable samples are partly fictitious, but the real gain in years of the extension can be estimated exactly if the variables are normal and approximately if they are asymmetric: moving from k to $\boldsymbol{n'}$ years thanks to n years $\boldsymbol{(k < n' < n)}$.

CHAPTER II

STATISTICAL ANALYSIS OF ANNUAL AND MONTHLY RAINFALL AND FLOWS

2.1 Statistical Analysis of Annual Rainfall and runoff (Gaussian Law)

When designing hydraulic structures, the designer is required to search for probabilities of occurrence for well-known values of rainfall and discharge and vice versa, as well as the accuracy of estimating these values. When it is not the average discharge, the choice of probability depends on economic considerations (probable lifespan of the development, cost, risk of destruction, etc.); accuracy is essential as it gives real meaning to the statistical estimation and provides an additional guarantee to the designer since it also depends on a risk probability to be chosen a priori according to comparable criteria.

To solve this type of problem, one generally starts from a sample of the variable defined by the problem, maximized if necessary. Then this sample is subjected to a complete statistical treatment which can be divided into 3 phases:

- Analysis of the sample and choice of the law type ;
- Estimation of the parameters of the law and verification of its adequacy ;
- Estimation of the variable's value for the retained probability P and calculation of the confidence interval at the risk threshold **α** chosen.

2.1.1 Sample Analysis and Law Type Choice

Some elementary notions of statistics must be recalled before analyzing the sample.

2.1.1.1 Calculation of Experimental Frequency

In statistics, the set of flow observations of a river during a year constitutes a test; from each test, various results are drawn, including the annual module. The set of all the realizations of the results m (module) of all possible tests forms a population **M**: an indefinite hypothetical sequence of which only a tiny part, the observed sample **E** of modules over a specified period of **N** years, is ever known.

The realization r in the population **M** occurs **n** times in **N** years; it is called:

- Experimental frequency of r in sample **E**: f = n/N;
- Probability of r in population **M**: prob (r/M) = lim n/N as N → ∞.

The realization is represented by a numerical value, a random variable **X** whose distribution function is:

$$F(x) = Prob\,[X \leq x] \qquad (2.1)$$

The derivative of **F(x)** is the probability density **f(x)** of the random variable, and we can write:

$$\text{Prob}\,[b \leqslant x \leqslant a] = F(a) - F(b) = \int_b^a f(x) \cdot \mathrm{d}x \qquad (2.2)$$

This represents the analogy between the frequency polygon and the probability density curve, as well as between the experimental frequency and the theoretical probability.

For practical calculations of experimental frequency, we sometimes talk about exceedance frequency:

$$F_1(x) = \int_x^{+\infty} f(x) \cdot \mathrm{d}x \qquad (2.3)$$

that is, the probability: **Prob X ≥ x**

or non-exceedance frequency:

$$F(x) = \int_{-\infty}^{x} f(x) \cdot \mathrm{d}x \qquad (2.4)$$

The sum of these frequencies F(x) +F1 (x) is obviously equal to 1.

However, if we calculated the experimental frequency by simply applying n/N, we would arrive at a sum greater than 1. We therefore adopt the formula f = (n - 0.5) /N, which is more consistent than the formula f = n/ (N+ 1) sometimes used, and which diverges especially for extreme values.

In these formulas :

- **n** is, after ranking in descending order, the rank counted from 1 of the realization r ;
- **N** is the number of realizations of the sample.

*2.1.1.2 **Sample characteristics***

It is not always possible to retain all realizations of the variable forming the observed sample in calculations.

This sample can be characterized by certain central and dispersion variables already used in Chapter I.

The mean : $\bar{x} = \frac{\sum x_i}{N}$

The variance: $s^2 = \frac{\sum (x_i - x)^2}{N-1} = \frac{1}{N-1}\left[\sum x_i^2 - N\bar{x}^2\right]$

The standard deviation ***s*** being the square root of the variance, whose second formulation is more conveniently used than the first on electromechanical machines.

The sample only gives a distorted image of the population; its empirical characteristics converge towards those of the population as the sample grows; this distortion represents sampling errors.

The empirical characteristics are realizations of random variables whose probability distributions are sampling distributions more or less dispersed around the theoretical characteristics of the population :

- Mathematical expectation: $E(\bar{x}) = m$
- Theoretical variance: $E(s^2) = \sigma^2$

It is very important to keep in mind the provisional and approximate nature of a sample vis-à-vis the infinite population and to understand that, consequently, the results derived from the statistical analysis of this sample are more or less precise.

This can be illustrated by examples taken from long-term samples, assuming that only fractions of these samples were known.

Thus, the module of a given station could have the following empirical means calculated over samples of 20 years.

years	Flow Q (m^3/s)
1808-1827	1025
1828-1847	1003
1848-1867	971
1868-1887	971

years	flow Q (m^3/s)
1888-1907	990
1908-1927	1058
1928-1947	1024
1941-1960	978

which all oscillate around the observed mean over 153 years: 1005 m^3/s, which is still only an approximation of the true mean.

What are the plausible limits within which an empirical characteristic calculated on the observed sample can vary, i.e., within what limits around this empirical characteristic must the corresponding theoretical characteristic lie? This important question is answered later during the calculation of the confidence interval.

2.1.1.3 Choice of law type

Classified sample, experimental frequency, empirical mean, and variance calculated ; it is a matter of finding a probability distribution that can adequately fit this sample.

From this distribution, with estimated theoretical characteristics, it will be possible to answer any question concerning any possible realization of the variable, and among other things :

- Calculate the estimate of any value of the variable for any chosen probability ;
- Randomly draw as many realizations of the variable as needed.

The choice of the type of distribution that best fits the sample is made using two experimental criteria :

In a given climatic region, a determined hydropluviometric variable generally follows the same distribution at any observation site, hence the interest in systematic regional studies and knowledge of any previous studies;

In the absence of regional information, an attempt is made to graphically plot the observed points on Gaussian scale paper, which allows considering a normal distribution in an asymmetric distribution.

Indeed, the scale anamorphosis allows aligning the points according to a normal distribution, while those following an asymmetric distribution would appear according to a curve with concavity facing upward (hyper-Gaussian distribution) or downward (hypo-Gaussian distribution). The choice must also consider statistical criteria :

According to the central limit theorem, a linear combination of N random variables is normally distributed when N tends to infinity, regardless of the individual distribution of these variables, if they are independent and if their values are not too different;

Apart from the normal distribution, it is hardly possible to address problems of confidence interval, gain in series extensions, correlations between variables without difficulty and without a computer.

All this leads to one conclusion: one always tries to fit a normal distribution, and among asymmetric distributions, if necessary, preference is given to those that allow a return to normality by changing the variable: logarithm (GALTON distribution) or nth root.

In other words, the probability distribution adjustable to the sample is all the closer to normality when :

- The discharge regime is abundant ;
- The variable concerns a long-time scale.

Practically, the module of a river (like that of a rainfall station) follows a normal distribution when the regional precipitation regime is abundant: temperate oceanic and continental region, tropical and equatorial region. This module, on the other hand, follows an asymmetric distribution for sub-desert and desert regimes, and one then uses either a GALTON distribution or an incomplete GAMMA distribution.

The normal distribution or GAUSS distribution has the distribution function:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-1/2\,u^2} \cdot \mathrm{d}u \qquad \textbf{(2.5)}$$

With: $\boldsymbol{u = (x - \overline{x})/s}$, reduced variable.

This second equation in the form : $\boldsymbol{x = x + su}$ is that of the so-called HENRY line, which represents the GAUSS curve on normal probability paper in abscissas.

The table of the GAUSS integral (Table A in the appendix) gives the values of F(x) as a function of those of u.

Plotting the observed values of a given sample on Figure 2.1 shows that fitting a normal distribution is advisable.

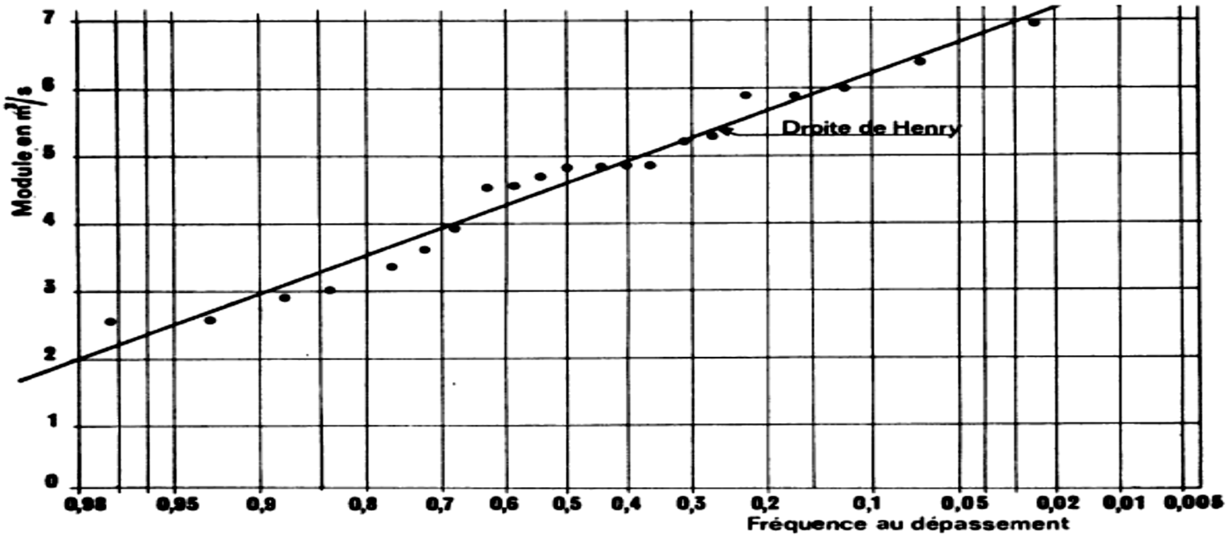


Figure 2.1: Example of normal distribution of annual modules on Gaussian paper

2.1.2 <u>**Estimation of Fitted Distribution Parameters and Goodness-of-Fit Test**</u>

Once a probability distribution has been chosen, the parameters of this distribution must be estimated from the sample. This estimation, which can be done either by the maximum likelihood method or by using moments, yields results that are naturally affected by sampling error.

The estimation methods fall under pure statistics, but the very particular nature of hydrologists' samples requires them to conduct personal research to find the best estimators.

It has been shown that the maximum likelihood method is the most efficient; unfortunately, with few exceptions, its implementation is very delicate. Therefore, for most distributions, parameters are estimated using moments—a consistent but generally inefficient estimation.

The sought estimator must be unbiased, convergent (towards the true value as the sample grows), and have low variance.

It is shown that the best estimators for the parameters of the normal distribution are the mean **(m)** and the variance ***(s^2)*** calculated from the sample.

Slightly exaggerating, one could say that with a small sample size (which is always the case in hydrology, where series rarely exceed a few decades) with significant variance, a probability distribution is always adjustable, and only doubling, tripling, or more of the sample size would verify the fit of this adjustment.

From samples of 10, 20, or 30 years, one tries to draw centennial and millennial conclusions, although prudence in practice requires not seeking a quantile whose probability of occurrence corresponds to a return period greater than three times the sample length. Extrapolation is all the more hazardous as the various distributions generally diverge at low frequencies, making it difficult to choose the most adequate one based on adjustments on small samples, which are poor in rare variable values.

One cannot exclude the hypothesis that the chosen distribution is not the one that truly represents the population; such an error is a fit error. This error adds to the sampling error to explain the divergences between experimental frequencies and theoretical probabilities derived from the fitted distribution.

The fit of a distribution to a sample is generally judged by the chi-square (χ^2) test, a random variable whose distribution was studied by PEARSON. For small samples of hydrological variables, the following rules apply :

- The division into classes must introduce classes of equal theoretical probability and a theoretical frequency of at least 5 values per class;
- The probability of χ^2 in a sample of unknown population represents the fit and thus has a 50% chance of being between 0.25 and 0.75, and a 90% chance of being between 0.05 and 0.95.

In practice, the following procedure is followed :

- Division into k classes according to the above rule ;
- Calculation of :

$$\chi^2 = \sum_{1}^{k} \frac{(n_i - n_{pi})^2}{n_{pi}} \qquad (2.6)$$

where n_i is the observed frequency in class i and n_{pi} is the theoretical frequency in the same class.

- Calculation of the number of degrees of freedom ν, equal to the number of classes minus the number of constraints between the sample and the fitted distribution:
  - 1 for the equality of frequencies $n_i = 2n_{pi}$.
  - 1 per parameter calculated from the distribution.
- The calculation of theoretical class limits is done as follows: Each has a frequency at exceedance
- $\sum_1^c n_{pi}/N$, c being the class number. This frequency introduced into the Gauss table gives the corresponding value u_c of the reduced variable. The class limit x_c can then be calculated with HENRI's equation: $x_c = \bar{x} - s \cdot u_c$

A table (Table B in the appendix) gives the χ^2 value with a certain probability of being exceeded.

- The fit is accepted if $P(\chi^2) \geq \alpha$, the adequacy risk threshold generally set at 0.05; it is rejected if $P(\chi^2) \leq 0.05$. Other threshold values can be chosen; sometimes the fit is not rejected until below the 0.01 threshold;
- Obtaining a high probability above 0.75, for example, should prompt a review of all adjustment and test calculations;
- The meaning of the 0.05 threshold is complex and can be summarized as follows:
  - If $P(\chi^2) > 0.05$, there is a 95% chance of verifying the fit hypothesis, but also a 5% chance of being wrong—this is the Type I error risk;
  - If $P(\chi^2) < 0.05$, the fit is rejected with a 95% chance of being right, but also a 5% chance of being wrong—this is the Type II error risk.

The choice of threshold is not fixed at 0.05. For large samples, one can be more stringent and go, for example, to 0.25.

In any case, the chi-square test is neither powerful enough nor consistent enough, even when applying the rule of equal theoretical probability classes with at least 5 values. Indeed, it has two major flaws :

- Despite the previous rule, the operator retains some freedom in choosing the number of classes, and each division may give very different χ^2 values; a lesser evil would be to calculate all and keep only the minimum value;
- The division into classes leads to neglecting the ends of the range, which are precisely the most important for practical use and fit, and limits the test to central values, which is very restrictive.

2.1.3 Calculation of the Confidence Interval

The confidence interval clarifies the importance of sampling errors. Its calculation and choice result from the study of sampling distributions, i.e., the laws to which empirical characteristics derived from samples are subject.

These empirical characteristics depend on the random variables of the sample, and the sampling errors they introduce are also due to chance, allowing the dispersion of empirical characteristics to be deduced from the properties of the adjusted law on the sample.

Under certain non-restrictive hypotheses, the empirical mean $\hat{\bar{x}}$ of a sample size N of a normal variable is also a normal variable:

- With a mean equal to $\hat{\bar{x}}$.
- With a variance equal to $\frac{S^2}{N}$.
- These equations are valid for $N \geq 30$.

For a small sample $N < 30$, the empirical mean follows a Student's t-distribution with N-1 degrees of freedom, a skewed distribution that approximates a Gaussian distribution for N>30.

Two properties of the confidence interval become immediately apparent: given its sampling distribution, there is an ***α* %** chance of finding the true value of the parameter within the empirical estimate :

- The amplitude is greater when the chosen confidence level **α** is high.
- The amplitude is greater when the sample size **N** is small. Similarly, the empirical variance s^2 :
- For a small sample, follows a χ^2 distribution with N-1 degrees of freedom.
- For a large sample, follows a normal distribution with mean s^2 and standard deviation sqrt $s^2\sqrt{2/N}$. This is also due to the fact that beyond N = 50, the distribution of χ^2 merges with the normal distribution.

The confidence intervals corresponding to empirical means and variances give significance to these calculated parameters.

- **Large Sample (N > 30 for the mean and N > 50 for the variance) :**

$\hat{\bar{x}} \pm u_{\frac{\alpha}{2}} . \frac{s}{\sqrt{N}}$ for the mean.

$s \pm u_{\frac{\alpha}{2}} . \frac{s}{\sqrt{2N}}$ for the standard deviation.

Where ***u*** is the standard normal variable.

- **Small Sample (N < 30 or N < 50) :**

$\hat{\bar{x}} \pm t_{\frac{\alpha}{2}} . \frac{s}{\sqrt{N}}$ for the mean.

Where ***t*** is the Student's t-distribution variable with N-1 degrees of freedom (Table C in the appendix).

$\frac{ns^2}{\chi 1^2} < \sigma^2 < \frac{ns^2}{\chi 2^2}$ For the variance.

Where : $P(x^2_{\alpha/2} \leq n\frac{s^2}{\sigma^2} \leq x^2_{1-\left(\frac{\alpha}{2}\right)})$ =1-$\alpha = IC$

$P[\chi^2 \leqslant \chi_1^2 = P[\chi^2 \leqslant \chi_2^2] = 1 - \frac{\alpha}{2}$ Values are extracted from a chi-squared table with N-1 degrees of freedom.

With : $x_1^2 = x_{\alpha/2}^2 \quad et\ x_2^2 = x_{1-\left(\frac{\alpha}{2}\right)}^2$ since it is asymmetric.

The choice of confidence level $\boldsymbol{\alpha}$ is free and depends on economic considerations regarding the risk the planner is willing to accept. The degree is chosen to be higher if security is sought or if high risk is unacceptable. Commonly accepted values are:

- 95% to 90% for important issues with a high safety margin or for well-known and less dispersed variables;
- 80% for moderately dispersed variables;
- 50% for highly dispersed variables.

The significance of the threshold α is similar to that attributed to the validity threshold of the chi-squared test. In a 95% confidence interval, there is a 95% chance of finding the true value of the estimated parameter within the interval, but there remains a 5% chance that it lies outside the interval without invalidating the assumptions (homogeneity of the variable, validity of the adjusted law).

2.1.4 **Calculation of Quantiles and Significance of the Result**

The value of rainfall or discharge probability **p** is given by the equation of the distribution function of the considered law, where the unknown is the variable for this probability, i.e., x_p or u_p, the reduced variable.

For the Gaussian law :

$$u_p = \frac{x_p - \hat{\bar{x}}}{s} \tag{2.7}$$

From which:

$$x_p = \bar{x} + s u_p, \tag{2.8}$$

The confidence interval of such a quantile ***xp*** is calculated by assuming its variance is the sum of the variances of the mean and the product s * u_p, thus ***S***.

This assumption of independence of the variances of the mean and the standard deviation is not rigorous but can be adopted without significant error. If $\boldsymbol{S_{xp}^2}$ is the variance of the quantile $\boldsymbol{x_p}$, we have :

$$S_{xp}^2 = \frac{s^2}{N} + \frac{s^2}{2N} u_p^2 \qquad S_{xp}^2 = \frac{s^2}{N}\left(1 + \frac{u_p^2}{N}\right) \tag{2.9}$$

The corresponding standard deviation is:

$$s_{xp} = \frac{s}{\sqrt{2N}} \sqrt{u_p^2 + 2} \tag{2.10}$$

Multiplying by $t_{\alpha/2}$ **,** provides the confidence interval at ***α % (IC=1-α).***

Theoretically inapplicable for small samples (N < 30).

2.2 Statistical Analysis of Annual and Monthly Rainfall and runoff (Galton's Law)

Rainfall or hydrometric modules generally do not follow a normal distribution when dealing with highly irregular regimes such as those observed in the Mediterranean region and in arid and semi-arid climates, as is the case in Algeria. In such cases, the series of modules follows a hypergaussian asymmetric distribution, and one can choose from the corresponding arsenal either the GALTON law or the PEARSON III law.

These two laws are also often used to represent samples of monthly discharges, which are practically never normal. They are also used to characterize certain interesting short-term values such as low-flow rates (average over 1, 10, 20, or 30 consecutive days, for example) and flood discharges.

To choose a priori between a normal law and a GALTON law, we have the criteria already mentioned above: regional study and graphical report. We can also look at the value of the coefficient of variation $\boldsymbol{C_v}$ of the sample, equal to the ratio of the empirical characteristics estimated, standard deviation s and mean m: ***Cv= s/m.*** The lower this coefficient of variation, the more likely normality is; from 0.50, the higher this coefficient, the less likely normality is.

2.2.1 **Parameter Calculation**

The GALTON law is also called the log-normal law or gausso-logarithmic law because its expression is close to that of the normal law with a logarithmic anamorphosis of the variable.

Indeed, it can be given the following representation:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{u^2}{2}} \cdot du \qquad \textbf{(2.11)}$$

Identical to that of a normal distribution, but in which:

$$u = a \log(x - x_0) + b \qquad \textbf{(2.12)}$$

The original variable being x and u the transformed variable resembles the reduced variable that follows the normal law.

We can also give another formulation, respecting this time the equation of the reduced

$u = (x - x_0)/s.$

Then we have :

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{u} \frac{1}{u} \cdot e^{-\frac{\log^2 u}{2\sigma^2}} \cdot du \qquad \textbf{(2.13)}$$

In this law, alongside the scale parameter s and shape parameter σ, there is a position parameter x_0, which is a lower bound of the original variable (champ de x_o à $+\infty$).

The system of equation (2.13) can be developed for computer calculation of the three adjustment parameters either by the method of moments or by the method of maximum likelihood. The system of equations (2.11) is more convenient for manual work and refers to the same GAUSS table used for the normal law without any change.

It is more judicious to estimate the value of the parameter x_o graphically than to calculate it. For information and to compare the methods of estimating the parameter x_o, we give here the calculation formula derived from the 2 and 3-order moment equations :

$$\frac{s^4}{\mu_3} = \frac{(\bar{x}-x_0)^3}{s^2+3(x-x_0)^2} \tag{2.14}$$

This equation is derived from the ratio of s^4 (square of the variance s^2 and the third-order centered moment μ_3; it is sufficient by successive approximations to solve it by varying x_o.

Through trial and error. The third-order centered moment (or third-order cumulant) is calculated by :

$$\mu_3 = \frac{1}{(n-1)(n-2)}\left[n\sum_1^n x_i^3 - 3\sum_1^n x_i \cdot \sum_1^n x_i^2 + \frac{2}{n}\left(\sum x_i\right)^3\right] \tag{2.15}$$

The graphical estimation of x_0 is done by plotting the observed points and their experimental frequency on Gaussian probability paper with logarithmic ordinate. The points should align if x_0 is zero; a concavity downwards is corrected with a negative x_0, i.e., by placing $x - x_0$ ordinate values greater than x; a concavity upwards (rarer) is corrected with a positive x_0, i.e., values x-x_0 less than x in ordinates.

Figure 2.2 presents the observed points of a sample of annual runoff volumes that are slightly concave towards the axis of abscissas; a minimum value of x_o = -3 allows a satisfactory linear adjustment by estimation ***Le* +3**. In practice, the smallest absolute value of x_o is adopted, allowing the low values of the sample to align with the high values.

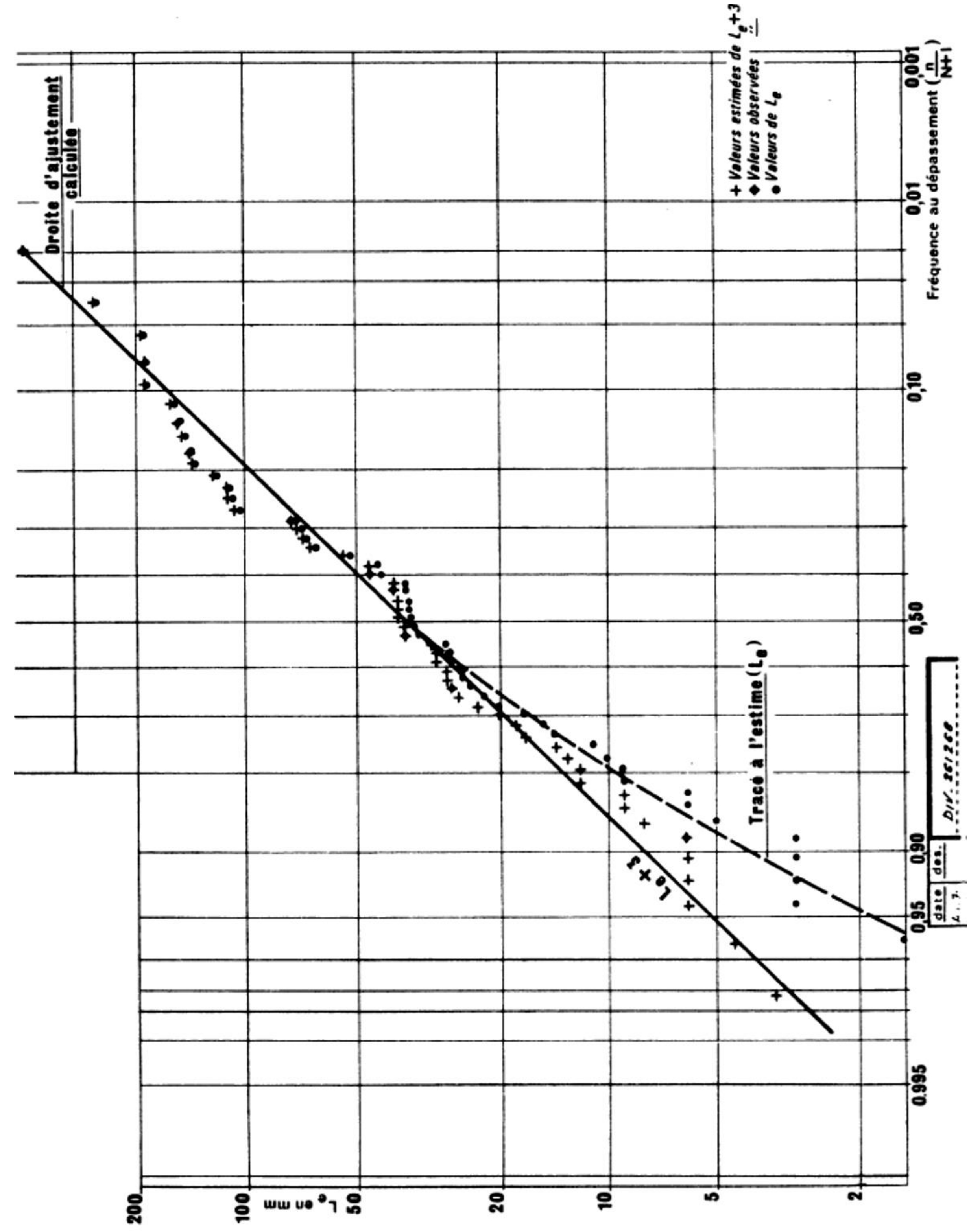


Figure 2.2: Annual Runoff Volumes Galton Law Adjustment

To complete the determination of the GALTON law, it is sufficient to calculate the adjustment line, i.e., its coefficients a and b. Three different methods can be used :

- Using the equations derived from the moments, which give :

$$a = \frac{1{,}517}{\sqrt{\log\left[1 + \frac{s^2}{(\bar{x} - x_0)^2}\right]}} \quad \text{et} \quad b = \frac{1{,}1513}{a} - a\log(\bar{x} - x_0)$$

- Considering the graphical adjustment line and expressing that the transformed variable $\boldsymbol{u}$ is the reduced variable :

$$\boldsymbol{y_i = \log(x_i - x_0)}$$

That is, setting :

$$\boldsymbol{u} = \frac{\boldsymbol{y_i - \hat{\hat{y}}}}{\boldsymbol{s}} \Rightarrow \boldsymbol{u} = \frac{\boldsymbol{1}}{\boldsymbol{s}}\boldsymbol{y_i} - \frac{\boldsymbol{\hat{\hat{y}}}}{\boldsymbol{s}}\ By\ analogy\ with\ \ (\mathbf{2.12})\ o\grave{u}:\ \boldsymbol{u = ay_i + b}$$

$$Where:\ \boldsymbol{a} = \frac{\boldsymbol{1}}{\boldsymbol{s}}\ \boldsymbol{et\ b} = -\frac{\boldsymbol{\hat{\hat{y}}}}{\boldsymbol{s}}$$

And calculating two particular points of the line.

- Using the equations derived from the maximum likelihood (longer calculations as they involve all the logarithms of the deviation to $\boldsymbol{x_0}$ for all sample values):

$$a^2 = \frac{1}{\frac{\sum \log^2(x_i - x_0)}{N} - \frac{\sum^2 \log(x_i - x_0)}{N^2}}$$

$$b = \frac{-a\sum \log(x_i - x_0)}{N}$$

Of course, a $\boldsymbol{X^2}$ goodness-of-fit test must be calculated.

2.2.2 Confidence Interval Calculation

It is the variable $\boldsymbol{y_t = log\ (x_i\text{-}x_o)}$ that follows a normal law. Therefore, confidence intervals can. easily be calculated from the sampling distribution laws of the mean $\hat{\hat{y}}$ and standard deviation $\mathbf{s_y}$ of the series of logarithms $\boldsymbol{log\ (x_i - x_o)}$ as was done for the normal law (§2.1).

The mean $\overline{\boldsymbol{x}}_{\boldsymbol{i}}$ of the variable x_i does not follow a GAUSS law but a GALTON law in which the mode, median, and mean are distinct and placed in this increasing order, whereas they are equal and coincident in a normal law. Consequently, $\overline{\boldsymbol{x}}_{\boldsymbol{i}}$ corresponds to a value y_i that is a simple quantile in the normal series of logarithms. To calculate the confidence interval on the mean $\hat{\overline{\boldsymbol{x}}}_{\boldsymbol{i}}$, first calculate the interval on $\boldsymbol{y_i = \log(\overline{x}_i - x_0)}$considered as a quantile by applying equation (2.12) to the series of logarithms: standard error of $\boldsymbol{y_i}$:

$$\boldsymbol{s_{y_i} = \frac{s_y}{\sqrt{2N}}\sqrt{u^2p+2}} \qquad \textit{\textbf{(2.16)}}$$

Then pass from the logarithmic interval bounds to the mean $\hat{\overline{\boldsymbol{x}}}_{\boldsymbol{i}}$ bounds by extracting the logarithms without forgetting to correct these values for x_o.

Comme on l'a déjà dit plus haut, ce procédé peut être pas très rigoureux est cependant assez précis pour les problèmes que l'on traite.

2.2.3 **Quantile calculation**

Simply use equation (2.12) for a probability corresponding to the value u_p of the reduced GAUSS variable:

$$Log\ (x_p - x_0) = (1/a)\ (u_p - b)$$

The resolution of the logarithmic equation then gives the value of x_p.

The confidence interval calculation is done exactly as for the mean of a GALTON law and with the same formal restrictions, i.e., this interval is:

$$\pm t_{\alpha/2}\frac{s}{\sqrt{2N}}\sqrt{u_p^2+2}$$

According to formula (2.10), α being the chosen confidence level and t the reduced GAUSS variable for the probability ***(1 - α) /2*** and s the standard deviation of the series of logarithms. Very imprecise at central values, the GALTON law (like other asymmetric laws) tends to be not much more so at distant quantiles compared to the normal law.

CHAPTER III

STATISTICAL ANALYSIS OF EXTREME FLOWS (FLOODS AND LOW WATER)

3.1 Characteristics of "Floods" and "Low Water" Information

The annual and monthly average flows studied in Chapter II are averages of daily flows over fairly long periods, with successive realizations rarely linked except for the flows of two successive months. These monthly and annual average flows are represented by variables whose statistical examination is sufficient to completely analyze their characteristics. On the contrary, extreme flows such as floods and low flows can only be fully analyzed and characterized if they are examined as a whole and in their chronological succession. Indeed, a flood or a low flow is a chronicle of instantaneous or daily flows (the latter being sufficient when the variability of the phenomenon is not too great over time, which is generally the case for low flows and for the floods of large rivers) lasting several hours to several days but rarely reaching a month. Finally, the flows constituting this chronicle are closely dependent on those preceding them.

The phenomena of floods and low flows are therefore very different from monthly and annual flows. Their complete study requires various and complex analytical methods; statistical analysis can only provide an incomplete description, which is nevertheless often sufficient for the pursued objective. In many application problems related to hydraulic developments, the statistical knowledge of a variable characteristic of flood or low flow information is enough to answer the questions raised by these problems.

A flood is a phenomenon that can be validly represented by one of the following variables :

- Instantaneous maximum flow (or over a short daily period, for example);
- Flow above a certain threshold (critical flow or base flow).

These variables are related to each other and often to the very shape of the flood hydrograph, which can be roughly reconstructed to be associated with them. This is generally sufficient to size a flood spillway or to calibrate a section of a Wadi for the transit without overflow of a certain flow, these two examples being the most common practical problems motivating the study of floods.

Understanding the low flow phenomenon is much more delicate. Low flow is interesting when water needs can only be met by extracting directly from the river. However, it should be borne in mind that in low flow periods, especially, the effect of multiple withdrawals, diversions, and discharges affecting a section of a watercourse is considerable. This effect can be precisely defined when it involves a major structure like a large diversion dam for a navigation canal, irrigation, etc., but not at all when it involves multiple individual interventions in rural or small industrial environments; and often, on the majority of rivers, these multiple interventions are responsible for low flow disruptions. Under such conditions, the instantaneous flow generally makes no sense, and low flow is preferred to be defined by one of the following variables :

- Lowest average flow over n (10, 20, 30) consecutive days ;
- Lowest average monthly flow ;
- Lowest ranked flow over n (10, 30) days.

3.2 Statistical Laws for Extreme Values

3.2.1 Extreme Value Theory

Extreme Value Theory (EVT) provides tools to estimate the distribution of observations and calculate extreme quantiles using this estimated distribution. It allows extrapolation of the behavior of the distribution tail of data from the largest observed data. It is based on the asymptotic approximation of the distributions of suitably normalized maxima of random vectors whose components are assumed to be I.I.D (Independent and Identically Distributed).

For reconstructing a sample of maximum flood flows, three approaches have been developed :

- The block maxima method, which only retains the maximum values over a given time interval (in hydrology, this interval is the year, and we then speak of annual maxima) ;
- The inflated flow method, which consists of retaining several maximums per year ;

- The peak over threshold (POT) method, which only retains the maximum values of events exceeding a fixed threshold.

For the block maxima method, probabilistic risk analysis related to the occurrence of extreme flood flows is performed from a sample consisting of the highest value measured each year. Von Mises (1954) and Jenkinson (1955) proposed a parametric family of distributions known $\boldsymbol{G(x) = G\mu, \omega, \xi(x)}$ as the Generalized Extreme Value (GEV) distribution, which summarizes all possible limit distributions of the asymptotic law of the maximum of a sample.

$$\boldsymbol{G\mu, \sigma, \dot{\xi}(x)} = \begin{cases} \boldsymbol{exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)_{+}^{\frac{-1}{\xi}}\right) si\xi \neq 0} \\ \boldsymbol{exp\left(-exp\left(-\left(\frac{x-\mu}{\sigma}\right)_{+}\right)\right) si\xi = 0} \end{cases} \quad \textbf{(3.1)}$$

Thus, the shape parameter ξ, through its different possible values, gives great flexibility to the GEV distribution, allowing it to account for the three types of asymptotic behavior represented by the following extreme distributions:

- If $\xi > 0$, G belongs to the Frechet domain of attraction (DAF) ;
- If $\xi = 0$, G belongs to the Gumbel domain of attraction (DAG) ;
- If $\xi < 0$, G belongs to the Weibull domain of attraction (DAW).

3.2.2 **Common Laws in Hydrology**

Asymmetric laws are used, with the first example being the Galton law studied in Chapter II, which is also applicable for extreme flows but will only be mentioned here for reference.

Flood and low flow phenomena represented by the variables defined in the previous paragraph generally have positive skewness, meaning the mode $\boldsymbol{d}$ (or most frequent value) is lower than the median $\boldsymbol{m}$, which is itself lower than the mean $\boldsymbol{\mu}$ (the "bell curve" of probability densities rises faster than it falls, or in the Gaussian paper representation where exceedance frequencies are plotted on the x-axis, the curves have their concavity facing increasing flows: this is also said to be "hypernormal" law).

Non-normal statistical laws are numerous, but if we retain only those applicable to hydrology with no more than 3 or 4 parameters and those that are not too laborious to implement manually, we end up with a limited list that includes almost all the laws used practically by hydrologists, classified into two domains of maximum attraction :

- Gumbel: Exponential, Log-Normal, Gumbel, Gamma, Pearson III, Weibull, Halphen A;
- Frechet: Pareto, Frechet, Halphen B-1, Inverse Gamma, Log-Pearson III.

For their simple analytical presentation, we adopt the distribution function corresponding to the reduced variable u such that $u = (x - x_o)/s$, a dimensionless variable. The parameters are of three orders :

- Position x_o, sometimes the mode (coinciding with the mean in normal and Gumbel laws), sometimes the bound (Galton law case);
- Scale s (standard deviation of the normal law);
- Shape ξ, absent (normal and Gumbel laws) unique in other laws.

The formulation of the most used laws in hydrology is given below :

- GUMBEL or Doubly Exponential Law : Known as the extreme values law, with the distribution function :

$$F(x) = e^{-e^{-u}} \tag{3.2}$$

- INCOMPLETE GAMMA or PEARSON III Law :

$$F(x) = \frac{1}{\Gamma(\gamma)} \int_0^u u^{\gamma-1} e^{-u}\, \mathrm{d}u \tag{3.3}$$

F(x) being the non-exceedance frequency,

Γ(γ) being the complete gamma function equal to $\int_0^\infty u^{\gamma-1} e^{-u}\, \mathrm{d}u$

γ is the positive shape parameter.

- Generalized Exponential Laws: The general expression for the distribution function is :

$$F(x) = \pm e^{-u^{1/d}} \tag{3.4}$$

Depending on the values and signs of the shape and scale parameters ***d*** and ***s***, the following specific laws are obtained :

- ***s and d positive, GOODRICH LAW:***

$$F(x) = 1 - e^{-A} \tag{3.5}$$

With : $A = (1/s)\,(x - x_0)^{1/d}$

- ***s positive, d negative, FRECHET LAW:***

$$F(x) = e^{-A} \tag{3.6}$$

3.2.3 <u>Choosing the best law and its type</u>

When asked why to choose one law over another to represent a certain variable, there is no clear and definitive answer. The choice is complete when one has a computer, limited when one does not. The more or less great adequacy of a law to the considered sample could theoretically be researched :

- Either by using information criteria such as BIC, AIC, and the Schwartz criterion ;
- Or by using graphical criteria such as the moments diagram and the L-moments diagram.

In practice, given the importance of calculations and sampling errors, a common-sense choice is often sufficient, adopting the type of law that has always given the best fits for the studied variable in the considered region. Positive asymmetric laws are almost confused in the central probability interval and only start to diverge, thus individualizing themselves, for extreme probabilities, which unfortunately correspond to very few experimental points; moreover, these extreme points in a short sample may well correspond to more severe occurrences (the probability of a centennial flood occurring in a 30-year sample is far from negligible), significantly limiting the interest of these points in choosing one law over another.

3.2.4 Estimating distribution parameters and quantiles

3.2.5 Annual maximums Series

3.2.5.1 Laws Parameter Estimation

We take as an example in this course the Gumbel law, the most used for flood prediction in Algeria. If Q is adopted as the representation of the flow variable, the probability density of the Gumbel law can be practically written as :

$$f(Q) = e^{-a(Q-Qo)} \tag{3.7}$$

- The scale parameter $s = 1/a$ is different from zero ;
- the position parameter is Q_0.

Their estimation is quite simple using a system of equations formed with the moments of the first three orders :

$$1/a=0.780\ \sigma \tag{3.8}$$

$$Q_0=\overline{Q}-(1/a)\ 0.577 \tag{3.9}$$

$\overline{Q}$ Being the mean of the sample and σ its standard deviation. The great practical interest of the Gumbel law is further emphasized by the ease of its graphical representation: a double logarithmic anamorphosis on the probability scale linearizes the theoretical distribution of the variable, the Gumbel line being :

$$u = a\ (Q - Q_0) \tag{3.10}$$

3.2.5.2 Quantiles Estimation

There is a Gumbel probability scale diagram paper that allows the linear representation of the Gumbel law. This Gumbel paper has on the abscissa scale a frequency scale for exceedance $F(Q)$ and a reduced variable scale u. The fitting line is drawn by calculating 2 or 3 values of Q for 2 or 3 values of u (2.0 and 7.0, for example) using equation (3.10). In the absence of a graduated scale in u, $F(Q)$ can be calculated as $F(Q) = e^{-e^{-u}}$ by entering twice into in a table of the exponential function e^{-u} et $u = -\ln(-\ln(F(Q)))$.

Quantile calculation can be done either by direct reading of the graph or by using the inverse of the e^{-u} table twice and then substituting u in equation (3.10). Important values of u for $F(x)$ can also be determined, as shown in Table 3.1:

Table 3.1: Values of u for some usual values of F(x) for non-exceedance

F(x)	u	Recurrence
0,90	2,25	10 years
0,95	2,97	20 years
0,98	3,90	50 years
0,99	4,60	100 years
0,999	6,91	1000 years

This Gumbel paper also features a logarithmic scale on the ordinate, useful for the Frechet law, which involves substituting log Q for Q in the Gumbel law's expression. The Frechet law is therefore just as easy to use and is employed when a Gumbel adjustment is not sufficiently asymmetric, resulting in an upward concavity for the sample values.

Fitting a law to flood discharge distributions allows estimation of discharge corresponding to a given low exceedance frequency: decadal, centennial, millennial floods. When referring to a millennial flood, the exceedance frequency is around 0.001, but given that observed flood samples rarely exceed a few dozen, the extrapolation is very strong and the precision is affected (if, conversely, the confidence interval of the discharge for a given frequency were calculated, this interval would be very wide for $F1(Q) = 0.001$).

Especially when estimating the discharge of a practically impossible flood ("project flood" when a structure's destruction cannot be accepted), it is arranged so that the estimated frequency $F_1(Q)$ is around 0.00001, but the term "ten-thousand-year flood" is preferably avoided.

3.2.6 **Inflated Series**

When the series to be adjusted includes the k extreme values per year, the following corrections should be applied (for "annualization" of observations):

- Correction of the cumulative non-exceedance probability:

$F_A\left(x_{[r]}\right)=\left\{F_E\left(x_{[r]}\right)\right\}^k$	$F_A(x_{[r]})$: cumulative non-exceedance probability of the annualized series of observations (dimensionless) $F_E(x_{[r]})$: cumulative non-exceedance probability of the annualized series of observations (dimensionless) k : number of observations per year (dimensionless)

- Correction of the fitting line parameters:

The parameter b (slope) is estimated in the same way as for a series composed of annual extremes, while the parameter a (intercept) undergoes the following correction:

$a_A= a_E+b.lin\ (k)$	a_A : annualized parameter in [m³/s] a_E : parameter of the inflated series in [m³/s] b : scale parameter of the Gumbel law in [m³/s] k : number of observations per year (dimensionless)

3.2.7 **Truncated Series**

When the series to be adjusted consists of a selection of independent events chosen from na observation years such that discharge values exceed a threshold x0, the following corrections should be applied (for "annualization" of observations):

- Correction of the cumulative non-exceedance probability according to the Langbein-Takeuchi relation:

$F_A(x)=\exp\left\{-\lambda\cdot\left[1-F_T(x)\right]\right\}$	$F_A(x)$: annualized cumulative non-exceedance probability (dimensionless) $F_T(x)$: cumulative non-exceedance probability of the truncated series (dimensionless) λ : average number of events occurring during the na observation years (λ = n/na with n the number of retained *events) (dimensionless)*

- Correction of the fitting line parameters:

$b_{GUM} = b_{EXP} = b$ $a_{GUM} = a_{EXP} + b \cdot Ln(\lambda)$	b_{GUM} : Gumbel distribution parameter in [m³/s] b_{EXP} : exponential distribution parameter in [m³/s] a_{GUM} : Gumbel distribution parameter in [m³/s] a_{EXP} : exponential distribution parameter in [m³/s] λ : average number of events occurring during the observation period (dimensionless)

For an exponential distribution, the value of $\boldsymbol{b_{EXP}}$ is given by the sample standard deviation σ, while $\boldsymbol{a_{EXP}}$ equals the estimated mean minus the sample standard deviation :

$$\begin{aligned} b_{EXP} &= \hat{\sigma} \\ a_{EXP} &= \hat{\mu} - \hat{\sigma} \end{aligned} \qquad \textbf{\textit{(3.11)}}$$

3.3 Confidence Intervals

In some projects, knowing the decadal flood, for example, is necessary for economic calculations: are the works to be carried out to completely prevent the flood more costly than the insurance premium if the risk of destruction is accepted? It is then necessary to know not only the most probable value but also the confidence interval of the decadal flood discharge.

The method of calculating confidence intervals was presented for a quantile of a random variable distributed according to the GAUSS or GALTON law (chap. II). For other asymmetric laws, the search for sampling distribution laws must be empirical by random drawing in the adjusted law to create a sufficiently large number of fictitious samples of the same size as the observed sample (N or N' depending on its origin) on which the determination of the said distributions is carried out.

Two examples of empirical approximate calculation of the confidence interval on a quantile of a GUMBEL law are presented, with the first formulation by CHOW proposing the following :

$$x_{[r]} \pm s_e \cdot z_\alpha$$

$$s_e = \left[\frac{1}{n} \cdot \left(1 + 1.1396 \cdot K_T + 1.1000 \cdot K_T^2 \right) \right]^{1/2} \cdot s \tag{3.12}$$

$$K_T = -\frac{\sqrt{6}}{\pi} \cdot \left[0.5772 + Ln\left(Ln\left(\frac{T}{T-1} \right) \right) \right]$$

Where :

- $x_{[r]}$: observed discharge of rank r in [m³/s];
- s_e: standard error of estimate in [m³/s];
- z_α: quantile of the standard normal distribution for a confidence index α;
- s: standard deviation of the observations in [m³/s];
- n: number of observations (dimensionless);
- K_T: frequency factor (dimensionless).

The second formulation is due to BERNIER and VERON. Let $\widehat{Q_{10}}$ be the estimation of the decadal flood from a sample of ***N*** values fitted to GUMBEL's law: we have a ***p%*** chance of finding the true value of Q_{10} in the interval $(\widehat{Q_{10}} - T_{2\sigma}, \widehat{Q_{10}} + T_{1\sigma})$, T_1 and T_2 being functions of ***N*** and ***p*** for the frequency considered (here, decennial). In practice, BERNIER and VERON have established charts, each valid for a given confidence interval. For example, for the ***70% CI*** (i.e. ***p = 0.70***), the attached chart (Figure 3.1) gives the values of T_1 and T_2 for the median, decadal and centennial floods as a function of ***N*** (or $\sqrt{N}$ more conveniently).

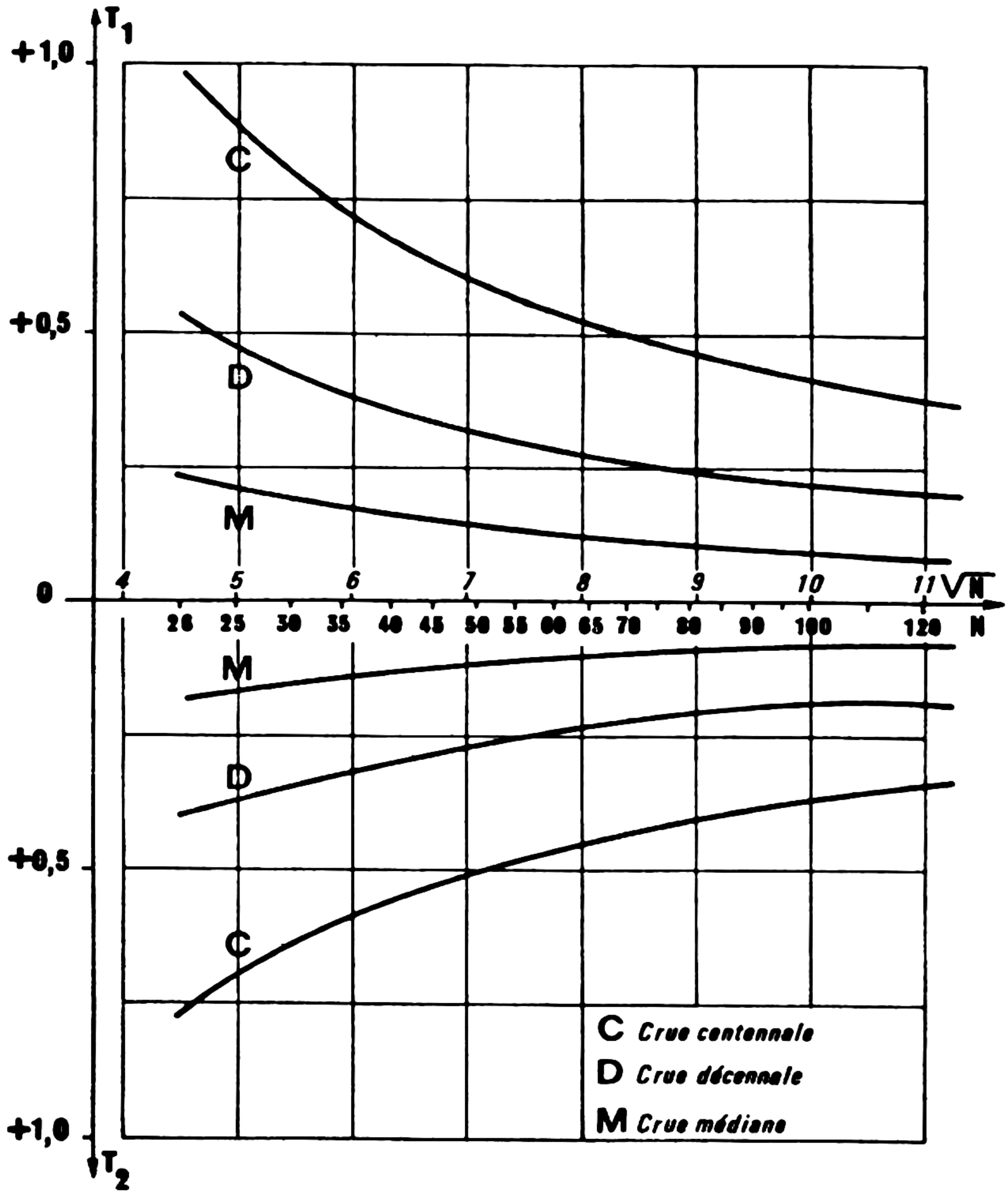


Figure 3.1: GUMBEL Law (70% confidence interval of estimated floods as a function of sample size N)

CHAPTER IV

HYDROMETEOROLOGICAL METHODS FOR ESTIMATING INFLOWS AND FLOODS

4.1 Introduction

Numerous methods exist for estimating flows. The choice of one method over another often depends on the presence or absence of hydrological data and must be adapted to the climatic and regional characteristics. These methods can be classified into three major groups:

- Statistical methods based on the exploitation of past observations of flow values. These methods were discussed in Chapters II and III;
- Hydrometeorological methods based on the exploitation of rainfall information while seeking acceptable rain-flow transformation relationships;
- Summary methods based on the exploitation of a few geographic or climatic parameters characteristic of the studied region.

Each of these methods has its advantages and disadvantages, as well as its application domain and usage conditions. In this chapter, we will discuss the second class of flow estimation methods, which are: Hydrometeorological methods based on rain-flow transformation. These are more suited to sites with little or no gauging.

This family of methods can be subdivided into two subclasses :

- Stochastic methods;
- Deterministic methods.

The deterministic approach, which involves understanding and describing the physical phenomena of hydrodynamics in the soil, is often preferred to a completely blind stochastic approach, the application range of which is necessarily limited in space. The boundary separating "deterministic" models from "stochastic" models is not always easy to define. Almost all "deterministic" models are never purely deterministic, as there are parameters to be calibrated on a case-by-case basis, often using statistical or quasi-statistical procedures.

Conversely, many stochastic models have a conceptual basis or rely on certain fundamental relationships that appear very deterministic.

4.2 Choice Between Models

The choice of a hydrometeorological model is made based on the following elements :

- Objective of the hydrological study : The objectives of the hydrological study are multiple, according to which certain parameters can be neglected or simplified in the modeling.
- Availability of data : Despite the uncertainties that models present compared to others, the lack of data compels us to use them.
- Nature of the model : The robustness and simplicity of the model influence the choice, especially when time or cost is a constraint.

Musy and Higy have provided a table (Table 4.1) that illustrates an example of model selection classification according to the aforementioned criteria (MUSY and HIGY, 1998).

Table 4.1: Flood estimation methods, data type and objectives (Musy and Higy 1998)

Variable of Design	**Necessary data recorded in the catchment area**		**Data step**
	Data type	**Methods**	
Peak flow rate	− Long series of maximum flows − Long series of maximum precipitation (P_max) − Short series of maximum flows (Q_max) − IDF curves	− Frequency Analysis − GRADEX and Derived Methods − Rational Methods and Derived Methods	− Empirical Formulas − Regional Methods − Analogue Methods
Flood hydrograph	− Concomitant series curves of precipitation and flows and IDF curves − Long series of flows	− Simple hydrological method (unit hydrograph, SCS-CN method) − Deterministic method − QDF curves and monofrequency synthetic hydrograph − Flood catalog and frequency analysis	− Synthetic unit hydrograph − Runoff coefficient − Project rain
Historical and/or probable flood scenarios	− Short concomitant series of precipitation and flows and long series of precipitation	− Continuous simulation model. Calibration on the short series of rain-flows then validation on a long series of flows from a long series of precipitation. − Stochastic precipitation model to generate synthetic rain chronicles	.

We'll try to give a few models for each class, by way of illustration and not as an exhaustive list.

4.3 Stochastic Methods

The methods described here can be considered stochastic to the extent that effective rain-flow modeling does not involve any analytical expression derived from a physical analysis of flow processes. Their stochastic nature is reinforced by the fact that either the production function is not explicit, or the calibration of its parameters is a by-product of the overall model calibration. We can cite among these methods :

- The Gradex method;
- The AGREGEE method;
- Correlative methods;
- Rain-flow Duration-Frequency modeling;
- The SHYPRE method (Simulation of Hydrographs for Predetermination).

Here, we will only discuss the Gradex method.

4.3.1 Gradex Method

This method, developed by Guillont and Duband (1967) of the DTG-EDF in Grenoble (French), is a probabilistic hydrometeorological approach for calculating extreme flood flows for return periods ranging from 100 to 10 000 years.

The Gradex method uses hydrometeorological information, i.e., the rain generating the flows. Thus, we have two samples: a sample of fine time step rain (daily, for example) and daily flows supplemented by a few fine time step flood hydrographs.

4.3.1.1 Principles, Hypotheses, and Validity Domain

The basic assumption of the method is that there must be a relationship between the distribution of flows and that of the generating rains since the flows are formed by the rains. Is this relationship simple?

Under certain extreme flow conditions (exceptional floods), yes: the soil is so saturated that any increase in rain will result, in volume terms, in the same increase in flow. In other words, everything that is precipitated runs off.

If the distribution of the random variable rain is exponential, then it can be shown that the distribution of flows under these extreme conditions is asymptotically exponential and that its parameters can be derived from those of the rains. Let us examine the hypotheses that correspond to these conditions:

Hypothesis 1 : The distribution of average precipitation over a basin for a few hours or days is exponential.

We can show that it follows that the distribution of maximum annual average precipitation over a basin for a few hours or days is of the GUMBEL type : $F(P) = \exp\{-\exp[-(P_0 - P) /a]\}$, where P_0 is a constant, as well as **a**, which is called "Gradex."

Hypothesis 2 : If the flow exceeds a certain value (which can vary from the ten-year flow to the fifty-year flow depending on the soils), then the soil is saturated such that during the base runoff time T, any increase in rain equals the same increase in flow, in other words: $dQ = dP$.

The corollary of this hypothesis combined with the previous one is that the distribution of flows will be asymptotically exponential and of the same parameter a as that of the rains.

Hypothesis 3 : The mean ratio, called the shape coefficient, between the maximum instantaneous flow of a flood hydrograph and the maximum average flow over period T is independent of the flow.

The first and third hypotheses are relatively easy to verify : numerous adjustments to Gumbel laws on rain samples have shown, except in rare cases, the validity of this hypothesis.

Indeed, it is easy to imagine that the saturation of the entire basin will be achieved more quickly as the soils are impermeable and the catchment area is small (50 to 1 000 km²) to be homogeneously watered with a rapid runoff response.

Thanks to these three main hypotheses, the principle of the method will consist in carrying out a probabilistic analysis on the rains to identify the Gumbel distribution and then transpose this distribution to that of the flows by extending the estimation of the ten-year flow.

4.3.1.2 Steps of Calculation Using the Gradex Method

1. Description of Phenomena and Data Critique : As with all methods, the first task is the hydroclimatological description and data critique. The Gradex method additionally requires the formation of a rainfall sample, which is actually the result of a study : generally, several rainfall series are available, and through interpolation analyses, a time series of average water depth estimation is reconstructed.

2. Hydrograph Analysis: From this, the time step T to be retained is deduced (rounded to that of the data), and the shape coefficient is established. Furthermore, the limit retention flow, from which saturation is assumed to be reached, is evaluated.

3. Calculation of Rain Gradex: For each season, the Gradex is estimated by fitting the seasonal rains to a Gumbel law, and the annual Gradex is deduced, which is asymptotically the highest of the seasonal Gradex.

4. Calculation of Limit Retention Flow: The ten-year flow (or more if necessary) is estimated by applying the Annual Maximum or Renewal method to the series of observed flows (average over duration T).

5. Calculation of Maximum Instantaneous Flow for a Given Return Period: After calculating the Instantaneous Gradex from the mean Gradex (a' = a.r, where r is the shape coefficient), the distribution of instantaneous flows is established by plotting the Gumbel law of parameter **a'** from the limit retention flow.

In conclusion, this method is well-suited for estimating extreme events (ten-thousand-year floods) as the hypothesis related to saturation is all the more verified as the flow is significant.

4.4 Deterministic Methods

Deterministic methods are based on the detailed study of the transition from precipitation to flow rates. This transition is very complex as it depends on a multitude of parameters: rainfall structure, watershed nature, degree of saturation, vegetation state, evaporation capacity, etc.

Thus, we can better understand the problem of rainstorm-flood relationships :

- A first point is to define "net rain" (what runs off) from useful rain or total rain. An approach to this transition is possible through production functions;
- Knowing the net rain, a second problem is to distribute it over time to get the hydrograph. This transition is done through transfer functions.

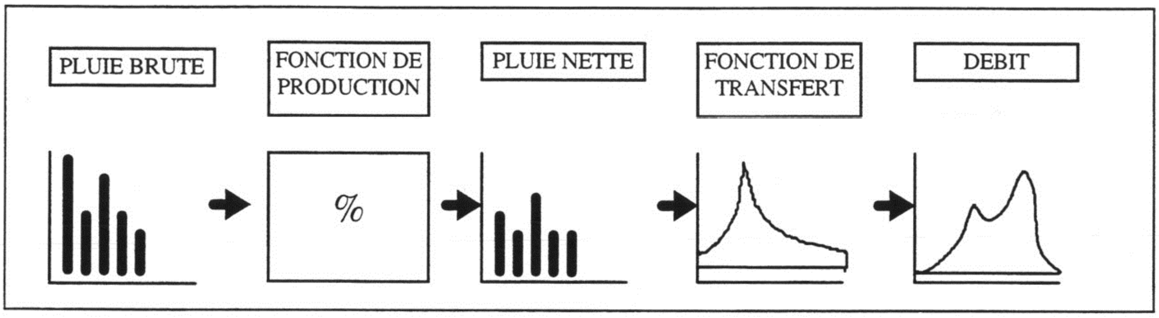


Figure 4.1: Rainfall-Runoff Path Diagram (Production and Transfer Function)

4.4.1 **Production Function**

The transition from total rain to net rain is certainly the decisive step. It is here that the most uncertainties lie and the grossest errors can be made.

The problem is as follows: at time t and during a time interval *dt*, the amount of rain that fell is *I(t) dt = dp(t);* during this same interval, the amount *J(t) dt* was lost. We will call runoff coefficient *kr(t)* the ratio between the runoff rain and the total rain:

$$kr(t) = \frac{I(t) - J(t)}{I(t)} = 1 - \frac{J(t)}{I(t)} \qquad (4\text{-}1)$$

The runoff coefficient will vary over time with the intensity of the rain, with the nature of the soil, and its initial saturation state.

The calculation of runoff water layers is done by subtracting from the total precipitation the amounts of water that are stored, infiltrated, or evaporated on the watershed. Interception, infiltration, storage, and evaporation are considered losses. To calculate these, several models have been implemented, including:

- Initial and constant rate loss model;
- Deficit and constant loss rate model;

- Curve Number (CN) model;
- Green and Ampt model.

For all these models, losses are calculated for each time interval and subtracted from the average areal precipitation of that interval. The remaining water quantity designates the precipitation excess. This quantity is considered uniform over the entire watershed and represents the volume of surface runoff.

4.4.1.1 Initial and Constant Rate Loss Model

This model considers that the potential for maximum loss rate, denoted *fc*, is constant and includes the initial loss rate *Ia,* which represents interception and storage in surface depressions. Interception results from rain absorption by the vegetation cover, and surface storage results from the watershed topography: water stored in surface depressions will either evaporate or infiltrate. As long as *Ia* is not reached, there is no runoff. This operation can be summarized as follows:

$$\begin{aligned} &\boldsymbol{Si} \ \sum_i \boldsymbol{P_i} \leq \boldsymbol{I_a} \ \ \boldsymbol{alors}, \ \ \boldsymbol{P_{et}} = \mathbf{0} \\ &\boldsymbol{Si} \ \sum_i \boldsymbol{P_i} \geq \boldsymbol{I_a} \ \ \boldsymbol{et} \ \ \boldsymbol{P_t} \geq \boldsymbol{f_c} \ \ \boldsymbol{alors}, \ \ \boldsymbol{P_{et}} = \boldsymbol{P_t} - \boldsymbol{f_c} \\ &\boldsymbol{Si} \ \sum_i \boldsymbol{P_i} \geq \boldsymbol{I_a} \ \ \boldsymbol{et} \ \ \boldsymbol{P_t} \leq \boldsymbol{f_c} \ \ \boldsymbol{alors}, \ \ \boldsymbol{P_{et}} = \mathbf{0} \end{aligned} \qquad \textbf{\textit{(4-2)}}$$

Où,

P_t : Average areal precipitation at time t ;

P_{et} : Runoff at time t given by:

$$\boldsymbol{P_{et}} = \begin{cases} \boldsymbol{P_t} - \boldsymbol{f_c} & \boldsymbol{si} \quad \boldsymbol{P_t} > \boldsymbol{f_c} \\ \mathbf{0} & \boldsymbol{ailleurs} \end{cases} \qquad \textbf{\textit{(4-3)}}$$

The constant loss rate corresponding to the soil's absorption power is expressed in mm/h. However, one can use the values given in the following table.

The difficulty of this method lies in determining initial losses and the constant loss rate. The first depends on conditions preceding the rain event under study (for example, if the soil was already saturated with water from previous rains, initial losses would be nearly zero). These losses also depend on the land use and soil type. The second depends on the physical properties of the basin soil and soil type. If the watershed soil is saturated, *Ia* will be close to zero. If the soil is drained, then *Ia* .

The constant loss rate corresponding to the soil's absorption power is expressed in mm/h. However, one can use the values given in the following table.

Tableau 4.2 : Le taux de pertes constant pour les différents types du sol

Group	Soil Type	Loss Rate Range (mm/h)
A	Deep sand, deep loess, aggregated silt	7.5 à 11
B	Shallow loess, sandy soil	3.5 à 7.5
C	Clayey soil, shallow sandy soil, soils with low organic matter content, clay soils	1.2 à 3.5
D	Soils that swell greatly when wet, heavy plastic clays, saline soils	0 à 1.2

4.4.1.2 Deficit and Constant Loss Rate Model

A variant of the previously mentioned model is the quasi-continuous model, which accounts for periods without rain during the event and thus integrates a regeneration (with a fixed rate) of initial losses. This is the "Deficit and Constant Rate" model. To use this model, one must know the initial loss rate, the constant loss rate, and the regeneration rate. The latter can be estimated as the sum of the evaporation rate and the percolation rate.

4.4.1.3 Curve Number (CN) Model

This model estimates precipitation excess as a function of cumulative precipitation, land cover, and initial soil moisture from the following equation:

$$\boldsymbol{P_e = \frac{(P - I_a)^2}{P - I_a + Spr}} \qquad \textit{(4-4)}$$

Où :

Where:

- Pe: Precipitation excess;
- P: Total accumulated precipitation at time t;

- Ia: Initial losses;
- Spr: Maximum retention potential.

There is also the empirical relationship developed from an analysis conducted on a set of small experimental watersheds that gives Ia as a function of Spr:

$$Ia = 0.2S_{pr} \tag{4-5}$$

Thus, we get:

$$P_e = \frac{(P - 0.2Spr)^2}{P + 0.8Spr} \tag{4-6}$$

The maximum retention potential *Spr* and watershed characteristics are linked through the Curve Number (CN):

$$Spr = \frac{25400 - 254CN}{CN} \tag{4-7}$$

CN can be estimated as a function of soil type, land use, and prior hydrological conditions of the watershed.

The NRCS has classified soils into four categories concerning infiltration potential :

Table 4.3: Initial and Final Infiltration Capacities Based on Soil Classes Defined by the SCS

Soil Type	Group	I_0 (mm/h)	I_f (mm/h)
Sandy silt	A	250	12-8
Sandy loam	B	200	8-4
Clay loam	C	130	4-1
Clay, saline soils	D	75	1-0

For a watershed composed of several land use and soil types, a composite CN must be calculated using the formula:

$$CN_{composite} = \frac{\sum A_i * CN_i}{\sum A_i} \tag{4-7-bis}$$

Where, A_i is the elementary surface area.

Note that the values presented in Table (VI.5) are valid for average antecedent moisture conditions (CNII).

Table 4.4: Land Use Type and Associated CN

Land Use	Soil Classes			
	A	B	C	D
Built-up areas	77	85	90	92
Roads	98	98	98	98
Sidewalks	98	98	98	98
Green space	67	78	85	89
Agricultural land	67	78	85	89
Forest	36	60	73	79

To account for antecedent moisture influence, we use the rainfall height during the five days preceding the flood event (NRCS). Then, we define either a dry, normal, or wet CN while respecting the following table.

Table 4.5: Relationship Between Previous Rainfall and CN Type for the NRCS CN Method

CN Types	Rainfall Height (mm) Over the Last 5 Days		
	Annual Basis	Seasonal Basis	
		Growing Season	Winter Season
CN(I): Dry Conditions	H< 12.5	H< 35	H< 12.5
CN(II): Normal Conditions	12.5<H<37.5	35<H<53	12.5<H<27.5
CN(III): Wet Conditions	H>37.5	H>53	H>27.5

4.4.1.4 Green & Ampt Model

It is one of the most used conceptual models for infiltration and precipitation. Its production function is based on Darcy's law generalized to unsaturated media combined with the mass conservation principle. Losses on permeable areas are calculated using the following formula:

$$\boldsymbol{f_t = K_h \left[\frac{1 + (\varphi - \theta_i) S_f}{F_t}\right]} \qquad \textbf{\textit{(4-8)}}$$

Where:

- K_h : Saturated hydraulic conductivity;
- φ : Porosity;
- $(\varphi - \theta_i)$: Volume of the moisture deficit;
- S_f A tabulated parameter representing suction before wetting;
- F_t : Represents cumulative losses at time t.

The precipitation excess is then calculated by subtracting from the average areal precipitation (P_{moy}) for the same time interval the losses calculated by the previous equation.

4.4.2 **Transfer Function**

We have just seen the techniques that allow us to transition from the quantity of rainfall *I(t), dt* fallen during a time interval *dt* to the quantity of ***R(t)*** that will start to run off. Now, we need to know when this runoff will reach the outlet. This transition will be studied using the transfer function. The most well-known methods are :

- The unit hydrograph method;
- The synthetic hydrograph method;
- The rational method;
- The linear reservoir method;
- The first difference of the transfer function (DPFT) method;
- The Gamma distribution or Nash hydrograph;
- The Sokolovski method.

4.4.2.1 Factors Influencing the Choice of Estimation Method

Several factors can play a role in choosing the estimation method, including :

- Climatic factors: Rainfall (intensity, variability), climate (seasonal distribution, etc.), the orientation of the watershed relative to prevailing winds;
- Physical characteristics of the watershed: Altitude, topography, shape, geology, soil, vegetation determine the runoff velocity;
- Flow variability;
- Data availability: Type of data, series size, spatial significance of the observation network, data reliability, recorded seasonal variations;
- Experience from dam breaches caused by floods;
- Regulations and standards: Many countries impose the probability of the design flood depending on various categories of dams;
- Project specifics: Size, importance, and consequences if there is a breach;
- Economic factors: The economic analysis of risks has become very important in project evaluation.

To conduct a detailed study of floods, it is essential to separate surface runoff from hypodermic flow and groundwater flow.

We limit, here, the presentation to the first two methods only.

4.4.2.2 Unit Hydrograph Method

Developed by Sherman, this method aims to study the discharge ***Q*** of a watershed brought by simple runoff. Its purpose is to determine the runoff discharge at the outlet of a watershed for a given time and a probability ***P%***.

To apply this method, proceed as follows :

1. Select homogeneous rainstorms covering the entire watershed over time;
2. Separate runoff from base flow by plotting the recession on semi-logarithmic paper;
3. Retrace the runoff hydrograph of the flood;
4. For each slice, find the runoff depth, discharge \(Q\), and the ratio \(Q/Pn\) by presenting the results in a table. Assign numbers to each slice of runoff on the flood hydrograph;
5. Determine the rainstorm duration. Snyder's formula can be used:

$$t_a = \frac{t_r}{5.5} \qquad (4\text{-}9)$$

With :

- t_a : duration of the rainstorm;
- t_r : response time of the watershed.

4.4.2.3 Synthetic Hydrograph Method or Isochrones Method

Developed by Larrieu, this method is based on the concentration of water in the watershed and the plotting of isochrones.

Certain assumptions are made for its application :

- Uniform distribution of the rainstorm generating the flood over the watershed;
- Constant water transfer speed from upstream to downstream;

- Giandotti's formula remains valid for calculating the concentration time;
- Effective rainfall is calculated by reducing losses.

A. Constructing an Isochrones Network

An isochrone is a line of geometric points where the travel time of a water drop from each point to the mouth is the same. The furthest isochrone represents the watershed's concentration time, meaning the time it takes for the entire watershed area to contribute to the flow at the outlet after a homogeneous rainstorm. The network plot consists of:

- Dividing the flow network into segments of constant length from the mouth to the upstream part of the smallest tributary valleys;
- After numbering, each point is characterized by three parameters :
    - Distance to the mouth;
    - Relative elevation;
    - Water travel time; this parameter appears on the watershed's topographic map;
- The water travel time from point *i*, to the outlet is :

$$t_i = \frac{d_i}{V_m} \qquad (7\text{-}10)$$

With :

d_i : distance measured along the water path;

V_m : vitesse moyenne de l'eau pouvant être estimée par la formule suggérée par l'institut d'hydrologie bavarois :

$$V_m = 20\, sin_{\alpha}^{3/5} \qquad (7\text{-}11)$$

With :

$$sin_{\alpha_i} \cong tg\alpha_i = \frac{\Delta H_i}{d_i} \qquad (7\text{-}12)$$

With :

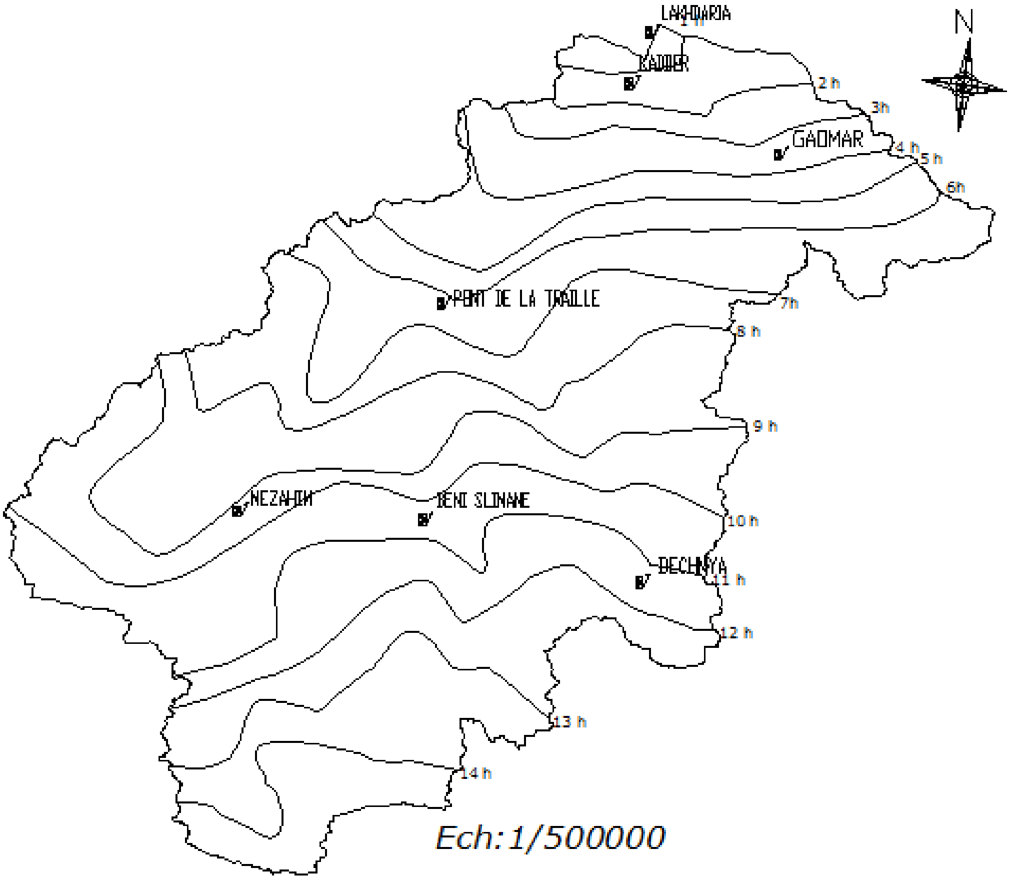
ΔH_i : height difference between point and outlet;

α_i: terrain angle (see average watershed slope).

Table 4.6: Velocity V calculated for different values of slope

Pente %	0.1	0.5	1	5	10	15	20	30
V m/s	0.311	0.838	1.265	3.292	4.999	6.401	7.62	9.601

- Randomly draw a number of points scattered over the basin, from each of which the probable course of the water (i.e. a line of maximum gradient) is traced until it meets the watercourse or one of the talwegs;
- Once all points have been identified by travel time, the isochrone network is interpolated (figure 4.2).



4.3: Isochrones for the Isser watershed

B. <u>Method for Calculating Maximum Discharge</u>

To apply this method, isochrones must be plotted on the watershed. The discharge is given by the following formula:

$$Q_{max} = \frac{S_{max} I_{eff}}{3.6} \quad (4\text{-}13)$$

With :

S_{max} : maximum discharging area (km²) ;

I_{eff} : effective intensity (mm/h), given by :

$$I_{eff} = \frac{P_{eff}}{t_{eff}} \quad (4\text{-}14)$$

With :

P_{eff} : effective rainfall in mm ;

t_{eff} : effective time in hours.

Isochrones are plotted on the watershed after calculating the water transfer speed $\boldsymbol{V_t}$, defined as the ratio between the main river length and the water concentration time, usually expressed in ***km/h***. Discharging areas $\boldsymbol{S_i}$ are determined by planimetry for each time step.

To find the maximum discharge, the effective intensity (see relation 4-14) is needed. This requires knowledge of the effective rainfall given by :

Table 4.7: Method for determining flowing surfaces

S (km²) / teff (h)	A1	A2	A3	A4	A5	Amax flowing
1	A1	A2	A3	A4	A5	A1
2		A1+A2	A2+A3			A1+A2
3			A1+A2+A3	A2+A3+A4		A1+A2+A3
4				A1+A2+…+A4	A2+A3+….+A5	A1+A2+…+A4
5					A1+A2+….+A5	A1+A2+……+A5

To find the maximum discharge, the effective intensity (see relation 4-14) is needed. This requires knowledge of the effective rainfall given by :

$$P_{eff} = P_{net} - D \quad (4\text{-}15)$$

With :

$$P_{net} = P_{maxj,p\,\%} \quad (4\text{-}16)$$

$$P_{max\ t,\ p\%} = P_{max\ j,\ p\%}\left(\frac{t}{24}\right)^{b} \qquad (4\text{-}17)$$

Net maximum daily rainfall at $P_{maxj,\ p\%}$ determined by fitting daily maximum rainfall to a statistical distribution.

The climatic exponent "b" can be calculated by plotting the average maximum storm intensity against time on log-log paper. It represents the slope of the corresponding line, often read on the BODY map (ANRH) for Algeria.

The runoff deficit D is determined by one of the loss calculation models mentioned earlier (see section 4.2.1) if not known in the study region.

CHAPTER V

SUMMARY METHODS

In the absence of data, empirical formulas are used for flow calculation. Several statistical or rational formulas are commonly used in engineering for the sizing of hydraulic structures. As they are closely dependent on the data used for their calibration, the results are often highly variable when they are compared. There are also issues regarding the robustness of these methods.

5.1 Summary Methods for Estimation of Floods

Three types are distinguished :

1. Methods of estimation by analogy between neighboring catchment areas.

2. Summary statistical methods :

- Crupedix Methods;
- Socose Methods.

3. Empirical methods :

- Mallet-Gauthier Formula;
- Turraza Formula;
- Giandotti Formula;
- Francou-Rodier Formula;
- Fuller Formula.

In this course, we will only attempt to present the empirical methods as they are the most used in Algeria. The summary statistical methods and pseudo-deterministic methods depend on a multitude of parameters that are generally unavailable in Algerian basins.

Empirical formulas are generally of three types :

1. Formulas considering the effect of the area (Turraza, Giandotti, and Francou-Rodier formulas);

2. Formulas considering the effect of the return period (Fuller for the calculation of the probable maximum average daily flow);

3. Formulas considering the effect of both parameters at once (Mallet-Gauthier and Fuller for the probable instantaneous maximum flow).

5.1.1 Evolution of Flow with Area

The essential parameters that influence the flood flow of a given frequency for a catchment area are :

- The area;
- The rainfall;
- The geological nature of the basin.

In a restricted region where it can be assumed that "geology" and especially "rainfall" vary little, it is possible to study the evolution of the flood flow of frequency F as a function of the basin area. Most studies show that, as a first approximation, the flow Q_F of frequency F varies as a power function of the area S :

$$Q_F = a_r S^b \qquad (5\text{-}1)$$

Where **b** is less than 1, indicating the attenuation of the peak flood flow as a function of the area. The coefficient *"a_r"* is a regional variable primarily incorporating "rainfall" and "geology."

5.1.2 Evolution of Flow with Frequency

Many authors have proposed linking the variation of flow with the return period T by relationships of the type :

$$Q(T) = Q(1)\,(1 + \beta \log T) \qquad (5\text{-}2)$$

Where :

- Q(T) is the flow with a return period T years;
- Q(1) is the so-called annual flood;
- β is a regional coefficient generally varying from 0.7 to 0.8 but can sometimes reach values greater than 2.

The local character of this value β does not allow us to provide values for each specific case. It can only be observed that the flow becoming a linear function of the return period implies that the distribution laws of flood flows have an asymptotically exponential behavior.

5.1.3 Mallet-Gauthier Formula

In their study on water problems in Algeria, Mallet and Gauthier established a formula expressing the maximum flood flow as a function of precipitation, basin area, and a coefficient K characteristic of the geographical and climatic conditions of the basin.

$$Q_{max,p\%} = 2K_e\, log\left(1 + 20\overline{P}\right)\frac{S}{\sqrt{L_t}}\sqrt{1 + 4\, log\, T - log\, S} \quad \textit{(5-3)}$$

Where :

- $Q_{maxp}\%$ is the peak flow in m^3/s with a return period T;
- *Ke* varies depending on the geographical region where the basin is located. In Algeria, for a normal basin, Ke can be taken as 1, but can reach 5 to 6 for small basins with a steep slope;
- $\overline{P}$ is the average annual precipitation in meters;
- *S* is the basin area in km^2;
- *Lt* is the length of the main talweg in km;
- *T* is the return period (years).

5.1.4 Turraza Formula

This formula uses the maximum average intensity of precipitation determined over a reference interval equal to the basin concentration time. It is slightly more elaborate than other formulas using precipitation as it also incorporates the basin concentration time ***(tc)***, i.e., the time required for a drop of water fallen at the most upstream point of the basin to reach the outlet. It can be expressed as :

$$Q_{max,\, p\%} = \frac{C_{rc}\,\bar{i}_{tc}\,S}{3.6} \quad \textit{(5-4)}$$

Where :

- S is the basin area in km^2 ;

- i_{tc} is the maximum average intensity of precipitation over a duration equal to the concentration time in mm/h ;
- C_{rc} is the runoff coefficient for a probability of exceedance P% :
  - P = 10% $C_{rc} = 0.6$
  - P = 1% $C_{rc} = 0.7$
  - P = 0.1% $C_{rc} = 0.8$
  - P = 0.01% $C_{rc} = 0.9$

5.1.5 **Giandotti Formula**

$$Q_{\max,\,p\%} = \frac{CSP_{tc,\,p\%}\sqrt{h_{moy} - h_{\min}}}{4\sqrt{S} + 1.5L_t} \qquad (5\text{-}5)$$

Where :

- S is the basin area in km^2 ;
- L_t is the length of the main talweg in km;
- H_m is the average altitude in meters;
- H_{min} is the minimum altitude in meters;
- P_{tc} is the layer of water precipitated for a given probability and a duration equal to the concentration time;
- C is a topographic coefficient varying between 66 and 166.

5.1.6 **Francou-Rodier Formula**

Francou and Rodier (1969) classified several hundred floods worldwide in a diagram log Q = f (log A). They observed that in relatively homogeneous regions, the points were more or less aligned. They deduced a general formula of the form:

$$\frac{Q}{Q_0} = \left(\frac{A}{A_0}\right)^{1-\frac{k}{10}} \qquad (5\text{-}6)$$

Where :

- Q is the maximum flow in m^3/s;
- Q_o is the maximum flow of a flood observed in a basin with area A_o (in km^3);
- A is the basin area in km^2;

- k is the Francou-Rodier coefficient varying according to the study region. In the Maghreb, it varies between 4 and 5.

5.1.7 Formule de Fuller

Fuller (1914) was one of the first to introduce the fundamental notion of the variation of the probable maximum flow as a function of the return period T.

$$Q(T) = Q_1\left(1 + 0.8 \log T\right) \qquad (5\text{-}7)$$

- ***Q(T)*** is the maximum daily flood flow in m^3/s for a return period T years;
- ***Q_1*** is the average of the maximum daily flows of each year calculated from available data (it is in some sense an "interannual" flood).

Fuller's formula allows the determination of the most probable flood flows for respective exceedance frequencies 10%, 1%, 0.1%, and 0.01% as follows :

- $Q_{10\%} = 1.8\ Q_1$ for the decadal flood (T=10 years).
- $Q_{1\%} = 2.6\ Q_1$ for the centennial flood (T=100 years).
- $Q_{0.1\%} = 3.4\ Q_1$ for the millennial flood (T=1000 years).
- $Q_{0.01\%} = 4.2\ Q_1$ for the ten-millennial flood (T=10000 years).

To transition from these average maximum daily flows to peak instantaneous flows, Fuller proposes the following formula:

$$Q_{max,\,p\%} = Q_{p\%}\left[1 + \frac{2.66}{S^{0.33}}\right] \qquad (5\text{-}8)$$

Where :

- ***S*** is the basin area in km^2;
- ***$Q_{p\%}$*** is the average maximum daily flow with a probability $P_{\%}$.

5.2 Summary Methods for Estimation of Inflow

5.2.1 Estimation of annual inflow average (A_0)

The determination of these inflows is preferably oriented towards hydrometric observations when they exist on the studied basin or by analogy with a neighboring basin. In their absence, models and empirical formulas based on rainfall should be applied; these models depend on precipitation and deficit.

$$Q = f\,(P-D) \qquad (5\text{-}9)$$

Where :

- Q is the annual flow (contribution).
- P is the average annual precipitation.
- D is the deficit encompassing all losses of the hydrological balance.

5.2.1.1 DERI Formula I

The inflow is given by the following relation :

$$A_0 = 0.915\, P_{moy}^{\ 2.684}\ S^{0.842} \qquad (5\text{-}10)$$

- A_0 : Annual Inflow average (hm^3);
- P_{moy} : Annual rainfall average (mm);
- S : Basin area (Km^2).

5.2.1.2 TURC Formula

The inflow is given by the following relation: $\mathbf{A_0 = Le.\ S}$

$$Le = P_{moy} - D \qquad (5\text{-}11)$$

$$D = \frac{P_{moy}}{\left(0.9 + \left(\frac{P_{moy}}{L}\right)^2\right)^{0.5}} \qquad (5\text{-}12)$$

Where :

- A_0 : Annual Inflow average (hm^3);

- Le : Runoff (mm);
- P_{moy} : Annual rainfall average (mm);
- D : Flow deficit (mm);
- L : Theoretical variable ; $L = 300 + 25\ T_{moy} + 0.05\ (T_{moy})^3$;
- T_{moy} : Annual temperature average (°C).

5.2.1.3 SAMIE Formula

The inflow is given by the following relation : $\mathbf{A_0 = Le.\ S}$

$$Le = P_{moy}^2\left(293 - 2.2S^{\frac{1}{2}}\right) \tag{5-13}$$

Where :

- A_0 : Annual Inflow average (m^3) ;
- L_e : Runoff (mm) ;
- S : Basin area (Km^2) ;
- P_{moy} : Annual rainfall average (m).

5.2.1.4 CHAUMONT Formula

The inflow is given by the following relation : $\mathbf{A_0 = Le.\ S}$

$$Le = 600.P_{moy}\ (1\text{-}10^{-k})$$

$$K = -0.36.P_{moy}^2 \tag{5-14}$$

Where :

- A_0 : Annual Inflow average (m^3) ;
- Le : Runoff (mm) ;
- P_{moy} : Annual precipitation average (m).

5.2.1.5 ANRH Formula

The inflow is given by the following relation :

$$A_0 = 0.513\ P^{2.603}\ D_d^{0.5}\ S^{0.842} \tag{5-15}$$

Where :

- P : Annual rainfall average (mm);
- S : Bassin area (Km^2);
- Dd : Drainage density (Km/Km^2).

5.2.1.6 DERI Formula II

The inflow is given by the following relation:

$$\boldsymbol{A_0 = K\ M_0\ S} \qquad \textit{(5-16)}$$

Where :

- A_0 : Annual Inflow average (hm^3);
- K : Constant given by $K = 31.54 \times 10^3$;
- M_0 : Average specific module ($l/s/Km^2$) given by $M_0 = 11.8\ Pmoy^{2.82}$;
- P_{moy} : Annual precipitation average (m);
- S : Bassin area (Km^2).

5.2.2 Characteristics of Flow

Flow is characterized by modules and coefficients, among which are :

5.2.2.1 Flow Module

The flow module is given by the following formula :

$$\boldsymbol{Me = A_0 / T} \qquad \textit{(5-17)}$$

- Me : Flow module (l/s);
- A_0 : Annual Inflow average (l);
- T : Time of one year in seconds ; T = 31 536 107 s.

5.2.2.2 Relative Flow Module

The relative flow module is given by the following formula:

$$\boldsymbol{q = Me / S} \qquad \textit{(5-18)}$$

Where :

- q : Relative flow module (l/s/Km^2);
- Me : Flow module (l/s);
- S : Basin area (Km^2).

5.2.2.3 Flow Coefficient

It is given by :

$$Ce = Le / P_{moy} \tag{5-19}$$

- Le : Runoff (mm);
- P_{moy} : Annual precipitation average (mm).

5.2.2.4 Coefficient of Variation

To calculate the coefficient of variation Cv, and in the absence of an observation series, empirical formulas must be used.

a) SOKOLOVSKY Formula

$$Cv = 0.78 - 0.29 \log q - 0.063 \log (S + 1) \tag{5-20}$$

Where :

- Cv : Dimensionless coefficient of variation;
- q : Relative flow module;
- S : Basin area (Km^2).

b) OURKGUIPROVODKHOZ Formula

$$Cv = 0.7/q^{0.125} \tag{5-21}$$

Where :

q : Relative flow module.

c) L'ANTONOV Formula

$$Cv = 0.7 / (S + 1000)^{0.097} \tag{5-22}$$

S : Bassin area (Km^2).

d) KRISTEKLY MENKEL Formula

$$Cv = 0.83 / (S^{0.06} \; q^{0.27}) \qquad (5\text{-}23)$$

Where :

- q : Relative flow module;
- S : Basin area (Km^2).

e) Algerian Formula by N. PADOUM

This formula gives good results for estimating the coefficient of variation of the annual inflow, as it is based on a statistical analysis of 42 wadis in northern Algeria, and is written as follows :

$$Cv = 0.93 \times K/q^{0.23} \qquad (5\text{-}24)$$

Where :

- K : Reduction coefficient K = (0.25-1.00), we take K = 0.55;
- q : Relative flow module ($l/s/km^2$).

5.2.3 Frequency Estimation of Inflows

Galton's law allowed for the estimation of annual inflows at different frequencies or return periods. The adjustment parameters of this law, namely the mean, standard deviation, and C_v will be calculated by the formulas mentioned above. The frequency inflow $A_{p\%}$ is calculated by the following formula:

$$A_{(p\%)} = \frac{A_0}{\sqrt{(C_v^2 + 1)}} \times e^{u\sqrt{ln(C_v+1)}} \qquad (5\text{-}25)$$

Where :

- $A_{p\%}$: Frequency inflow for a given probability p% ;
- u : Standard Gaussian variable ;
- A_0 : Average annual inflow (Hm^3) ;
- C_v : Coefficient of variation.

Tableau A. — Valeurs de l'intégrale de Gauss pour $u \geqslant 0$
(Probabilités pour que u soit supérieur ou égal à ...)

u		0	1	2	3	4	5	6	7	8	9
0,0	0,	50000	49601	49202	48803	48405	48006	47608	47210	46812	46414
0,1		46017	45620	45224	44828	44433	44038	43644	43251	42858	42465
0,2		42074	41683	41294	40905	40517	40129	39743	39358	38974	38591
0,3		38209	37828	37448	37070	36693	36317	35942	35569	35197	34827
0,4		34458	34090	33724	33360	32997	32636	32276	31918	31561	31207
0,5		30854	30503	30153	29806	29460	29116	28774	28434	28096	27760
0,6		27425	27093	26763	26435	26109	25785	25463	25143	24825	24510
0,7		24196	23885	23576	23270	22965	22663	22363	22065	21770	21476
0,8		21186	20897	20611	20327	20045	19766	19489	19215	18943	18673
0,9		18406	18141	17879	17619	17361	17106	16853	16602	16354	16109
1,0		15866	15625	15386	15151	14917	14686	14457	14231	14007	13786
1,1		13567	13350	13136	12924	12714	12507	12302	12100	11900	11702
1,2		11507	11314	11123	10935	10749	10565	10383	10204	10027	98525
1,3	0,0	96800	95098	93418	91759	90123	88508	86915	85343	83793	82264
1,4		80757	79270	77804	76359	74934	73529	72145	70781	69437	68112
1,5		66807	65522	64255	63008	61780	60571	59380	58208	57053	55917
1,6		54799	53699	52616	51551	50503	49471	48457	47460	46479	45514
1,7		44565	43633	42716	41815	40930	40059	39204	38364	37538	36727
1,8		35930	35148	34380	33625	32884	32157	31443	30742	30054	29379
1,9		28717	28067	27429	26803	26190	25588	24998	24419	23852	23295
2,0		22750	22216	21692	21178	20675	20182	19699	19226	18763	18309
2,1		17864	17429	17003	16586	16177	15778	15386	15003	14629	14262
2,2		13903	13553	13209	12874	12545	12224	11911	11604	11304	11011
2,3		10724	10444	10170	99031	96419	93867	91375	88940	86563	84242
2,4	$0,0^2$	81975	79763	77603	75494	73436	71428	69469	67557	65691	63872

2,5		62097	60366	58677	57031	55426	53861	52336	50849	49400	47988
2,6		46612	45271	43965	42692	41453	40246	39070	37926	36811	35726
2,7		34670	33642	32641	31667	30720	29798	28901	28028	27179	26354
2,8		25551	24771	24012	23274	22557	21860	21182	20524	19884	19262
2,9		18658	18071	17502	16948	16411	15889	15382	14890	14412	13949
3,0		13499	13062	12639	12228	11829	11442	11067	10703	10350	10008
3,1	$0,0^3$	96760	93544	90426	87403	84474	81635	78885	76219	73638	71136
3,2		68714	66367	64095	61895	59765	57703	55706	53774	51904	50094
3,3		48342	46648	45009	43423	41889	40406	38971	37584	36243	34946
3,4		33693	32481	31311	30179	29086	28029	27009	26023	25071	24151
3,5		23263	22405	21577	20778	20006	19262	18543	17849	17180	16534
3,6		15911	15310	14730	14171	13632	13112	12611	12128	11662	11213
3,7		10780	10363	99611	95740	92010	88417	84957	81624	78414	75324
3,8	$0,0^4$	72348	69483	66726	64072	61517	59059	56694	54418	52228	50122
3,9		48096	46148	44274	42473	40741	39076	37475	35936	34458	33037
4,0		31671	30359	29099	27888	26726	25609	24536	23507	22518	21569
4,1		20658	19783	18944	18138	17365	16624	15912	15230	14575	13948
4,2		13346	12769	12215	11685	11176	10689	10221	97736	93447	89337
4,3	$0,0^5$	85399	81627	78015	74555	71241	68069	65031	62123	59340	56675
4,4		54125	51685	49350	47117	44979	42935	40980	39110	37322	35612
4,5		33977	32414	30920	29492	28127	26823	25577	24386	23249	22162
4,6		21125	20133	19187	18283	17420	16597	15810	15060	14344	13660
4,7		13008	12386	11792	11226	10686	10171	96796	92113	87648	83391
4,8	$0,0^6$	79333	75465	71779	68267	64920	61731	58693	55799	53043	50418
4,9		47918	45538	43272	41115	39061	37107	35247	33476	31792	30190

Pour $u < 0$, il suffit de prendre le complément à 1 des valeurs contenues dans le tableau, les nombres de la première colonne désignant alors les valeurs absolues de u.

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-1/2\, u^2} \cdot du \text{ avec } u = \frac{x - \bar{x}}{s}$$

Tableau B. — Table de distribution de χ^2 (Loi de K. Pearson)
(Valeurs de χ^2 ayant la probabilité P d'être dépassées)

ν \ P	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0158	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,12	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,47	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,13
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,27	19,67	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,84	32,00	39,25
17	6,41	7,56	8,67	10,08	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,80	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	45,32
21	8,90	10,28	11,59	13,24	29,61	32,67	35,48	38,93	46,80
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	43,27
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	49,73
24	10,86	12,40	13,85	15,66	33,20	36,41	39,37	42,98	51,18
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	56,62
26	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64	54,05
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	55,48
28	13,57	15,31	16,93	18,94	37,92	41,34	44,46	48,28	56,89
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	58,30
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	59,70

Lorsque $\nu > 30$ on peut admettre que la quantité $\sqrt{2\chi^2} - \sqrt{2\nu - 1}$ suit la loi normale réduite.

$$\chi^2 = \sum_{1}^{k} \frac{(n_i - n_{pi})^2}{n_{pi}}$$

Tableau C. — Loi de Student-Fisher

Valeur de t qui a la probabilité P d'être dépassée en module (ν, nombre de degrés de liberté)

P / ν	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,955	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819

BIBLIOGRAPHY

1. Akaike, H. 1974. A new look at statistical model identification. IEEE Transactions on Automatic Control. AU, 19. pp 716-722.

2. Bernier, J., and Veron, R. 1963. Sur quelques difficultés rencontrées dans l'estimation d'un débit de crue de probabilité donnée. Communication présentée à la Session n° 74 du Comité Technique de la S.H.F.

3. Bobée, B. 1999. Extreme flood events evaluation using frequency analysis: a critical review. Houille Blanche, 54 (7-8). pp 100-105. doi: 10.1051/lhb/1999090.

4. Buján, César M., Véliz, J., and Manzanares, B. 2004. Hydrologie appliquée : Procédures méthodologiques pour l'exécution d'études des retenues collinaires et petits barrages. ANBT, unite retenues collinaires, 54 p.

5. Dubreuil, P. 1974. Initiation à l'analyse hydrologique. Edition Masson et CIE, ORSTOM, Paris, 216 p.

6. El Adlouni, S., Bobée, B., et Ouarda, T. B.M.J. 2008. On the tails of extreme event distributions in hydrology. Journal of hydrology, 355. pp 16-33. doi :10.1016/j.jhydrol.2008.02.011.

7. Garrido, M. 2002. Modélisation des évènements rares et estimations des quantiles extrêmes, Méthode de sélection de modèles pour les queues de distribution. Thèse de doctorat, Université Grenoble 1.

8. Guillot, P. 1981. The "GRADEX" method, Journal des sciences hydrologiques, 26 (3), pp. 334-336, DOI: 10.1080/02626668109490894

9. Guillot, P. and Duband, D. 1967. La méthode du Gradex pour le calcul de la probabilité des crues à partir des pluies. Dans : Les Crues et leur Evaluation (Actes du colloque de Leningrad), IAHS publication, N° 84, pp. 560-569

10. Gumbel, E.J. 1942. On the Frequency Distribution of Extreme Values in Meteorological Data. Bull. Am. Meteor. Soc. 23. pp. 95-104.

11. Gumbel, E.J. 1955. Statistical theory of extremes values and some practical applications. Journal of the Royal Statistical Society, Serie A, vol. 119, N°.1, p. 106.

12. Gumbel, E.J. 1958. Statistics of Extremes. Columbia University Press, New York.

13. Hebal, A. 2013. Analyse hydrologique de quelques bassins versants du Nord Algerien : Eaux superficielles, crues et aménagements. Thèse de doctorat Es-Science, Université Saad Dahlab de Blida, 314 p.

14. Hingray B., Picouet,C., and Musy, A. 2009. Hydrologie 2 : Une science pour l'ingénieur. Presses polytechniques et universitaires romandes, Lausanne. 640 p.

15. Jenkinson, A.F. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society. 81. pp. 158-171.

16. Larrieu, J. 1956. Méthodes d'analyse de la structure fine des débits. Revue de Statistique Appliquée, IV (2).

17. Larrieu, J. 1957. Evaluation des crues catastrophiques par la méthode des hydrogrammes synthétiques. A.I.H.S. TORONTO, Tome III.

18. Meylan, P., Favre, A-C., and Musy, A. 2008. Hydrologie fréquentielle. Une science prédictive. Presses polytechniques et universitaires romandes. 184 p.

19. Michel, C., and Oberlin, G. 1987. Seuil d'application de la méthode du Gradex (Extrapolating a flood frequency curve using the "Gradex" method). La houille blanche, N°3, pp 199-203.

20. Musy, A., and Higy, C. 1998. Hydrologie appliquée. Edition H*G*A, Bucarest. 365 p.

21. Musy, A., Higy, C., and Reynard, E. 2014. Hydrologie 1. Une science de la nature, Une gestion sociétale. Presses polytechniques et universitaires romandes, 516 p.

22. Nash. J.E. 1957. The form of the instantaneous unit hydrograph. Proc. General Assembly of Toronto. Int. Assoc. Sci. Hydrol. Publ., 45 (3), pp. 114-121.

23. Ouarda, T.B.M.J., Ashkar, F., Bensaid, E., et Hourani, I. 1994. Distributions statistiques utilisées en hydrologie : Transformation et propriétés asymptotiques. Rapport scientifique STAT-13, Département de Mathématiques, Université de Moncton, N.B., Canada.

24. Roche, M. 1963. Hydrologie de surface. Édition Gauthier-Villars, Paris. 431 p.

25. Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6. pp 461-464.

26. Serra, B.G. 1979. Cours d'hydrologie, Cours magistral. Institut national agronomique El Harrach, Alger. 153 p.

27. Snyder, F.F. 1938. Synthetic unit hydrographs. Trans. American Geophysics Union, 19, pp 447-454.

28. Thiery, D. 1989. Eléments d'analysr statistique : application à l'hydrologie, deuxième Edition. Bureau de recherches géologiques et minières, France, 75 p.

29. Touaibia, B. 2004. Manuel pratique d'hydrologie. Edition Madani frères, Guerouaou .163 p.

30. Von Mises, R. 1954. La distribution de la plus grande de n valeurs. In Selected Papers, American Mathematical Society, Providence, RI. II. pp 271-294.