Chapter 7 Methodology for processing meteorological and hydrometric data at basin level

Capítulo 7 Metodología para el tratamiento de datos meteorológicos e hidrométricos a escala de cuenca

SÁNCHEZ-QUISPE, Sonia Tatiana†*, NAVARRO-FARFÁN, María del Mar and GARCÍA-ROMERO, Liliana

Universidad Michoacana de San Nicolás de Hidalgo, Postgraduate course of the School of Civil Engineering, Mexico

Universidad Michoacana de San Nicolás de Hidalgo; Postgraduate studies at the School of Chemical Engineering, Mexico

ID 1st Author: *Sonia Tatiana, Sánchez-Quispe* / **ORC ID:** 0000-0002-8394-495X, **CVU CONACYT ID:** 202363

ID 1st Co-author: *María del Mar, Navarro-Farfán* / **ORC ID:** 0000-0002-8423-3092, **CVU CONACYT ID:** 859151

ID 2nd Co-author: *Liliana, García-Romero* / **ORC ID:** 0000-0002-6537-7736, **CVU CONACYT ID:** 557195

DOI: 10.35429/H.2021.16.107.145

S. Sánchez, M. Navarro and L. García

* pvieyrar@uaemex.mx

A. Marroquín, J. Olivares, D. Ventura, L. Cruz. (Coord.) CIERMMI Women in Science TXVI Engineering and Technology. Handbooks-©ECORFAN-México, Querétaro, 2021.

Abstract

Meteorological and hydrometric data recorded by stations require analysis and processing before being used in any study as recommended by the World Meteorological Organization to ensure the reliability of the results obtained from these series. Currently, there is a vast number of tests that can be used for this purpose, generally applied in isolation. However, there is no clear methodology that specifies the conditions of application of the tests, the order, and the conditions of application for meteorological and hydrometric series. This research proposes a methodology for the selection and validation of meteorological and hydrometric data according to the characteristics of the available information and the study area. This methodology is the result of the work carried out for several years, where through the analysis of different series it has been concluded that the methodology presented here allows the efficient treatment and discretization of meteorological and hydrometric information, where it has been verified that the results obtained from the different studies where this information has been used have given reliable results. An application case has been selected for the description of the methodology and the analysis of the results. This chapter will be developed through 6 sections. Section 7.1 gives a brief introduction to the treatment of meteorological and hydrometric stations. Section 7.2 shows the development of the methodology proposed in this work. Section 7.3 describes the application case and the characteristics of the implemented data. Sections 7.4 and 7.5 describe the development of the tests used for meteorological and hydrometric data processing, respectively. Finally, section 7.6 concentrates on the conclusions obtained from the application of this work.

Data processing, Meteorological, Hydrometric, Selection criteria

Resumen

Los datos meteorológicos e hidrométricos registrados por las estaciones, requieren de un análisis y tratamiento antes de ser utilizados en cualquier estudio según recomienda la Organización Meteorológica Mundial para garantizar la fiabilidad de los resultados obtenidos a partir de estas series. Actualmente, existe un vasto número de pruebas que pueden ser utilizadas para tal fin, aplicadas generalmente de forma aislada. Sin embargo, no hay una metodología clara que especifique las condiciones de aplicación de las pruebas, el orden y las condiciones de aplicación para series meteorológicas e hidrométricas. Este trabajo propone una metodología para la selección y validación de los datos meteorológicos e hidrométricos de acuerdo a las características de la información disponible y la zona de estudio. Esta metodología es el resultado del trabajo realizado por varios años, donde a través del análisis de distintas series se ha llegado a la conclusión que la metodología que aquí se presenta permite el tratamiento y discretización eficiente de la información meteorológica e hidrométrica, dónde se ha podido constatar que los resultados obtenidos de los distintos estudios donde esta información se ha utilizado, han dado resultados fiables. Se ha seleccionado un caso de aplicación para la descripción de la metodología y el análisis de resultados. Este Capítulo se desarrollará a través de 6 secciones. En la sección 7.1 se da una breve introducción sobre el tratamiento de estaciones meteorológicas e hidrométricas. La sección 7.2 muestra todo el desarrollo de la metodología propuesta en este trabajo. En la sección 7.3 se describe el caso de aplicación y las características de los datos implementados. En las secciones 7.4 y 7.5 se describe el desarrollo de las pruebas utilizadas para el tratamiento de datos meteorológicos e hidrométricos, respectivamente. Finalmente, en la sección 7.6 se concentran las conclusiones obtenidas de la aplicación de este trabajo.

Tratamiento de datos, Meteorológicos, Hidrométricos, Criterios de selección

7.1 Introduction

The reliability of meteorological information is substantial in any research related to hydrology and water resources systems at the basin scale; therefore, it is important to implement tools and methodologies that allow the identification of data containing errors, whether due to data collection, instrument failures, sensor and/or communication failures, transcription of information, or relocation of the stations where these variables are measured. Precipitation is the component of the hydrological cycle that represents the main input of hydrological models, since it represents the amount of water that falls on the earth's surface, and, therefore, the starting point for the evaluation of available water resources, the analysis of extreme events, or the evaluation of climate change. In Mexico, there are precipitation records through 3266 meteorological stations distributed throughout the country (Salvador-González et al., 2018).

However, the spatial and temporal coverage of the available information is often deficient due to instrument failures or errors in data collection. These errors cause the meteorological series available to be incomplete and unreliable records.

Precipitation information recorded at meteorological stations must be representative and accurate of the place where it is measured, to be used reliably in hydrological analysis and water resource management, the series used must comply with two properties: they must be homogeneous and independent (Cao & Yan, 2012). This will provide greater reliability in the results obtained from the investigations where it is intended to use such series (flood events, frosts, droughts or climate change), reducing the uncertainty associated with hydrological modelling.

Similarly, the time series of flow measurements recorded at hydrometric stations is another basic information for studies related to hydrology or water resource systems. It is common that, like precipitation records, these series contain errors and missing data, generated by technical failures in the instruments, errors in the handling of the equipment or recording of the information. In Mexico, the measurement of these data is performed through 861 points for the measurement of flows that give rise to the network of piezometric stations in our country (CONAGUA, 2016). To ensure the reliability of data from hydrometric stations, as well as precipitation series, they must comply with at least two properties: they must be homogeneous and persistent (Merlos et al., 2014).

Both meteorological and hydrometric data may not represent the actual weather variation due to failures in the measuring instruments, errors caused accidentally by the person responsible for data collection, the location of the station, among others, resulting in variations in real data, and causing the user of this climatic information to obtain erroneous results or to make erroneous inferences. Having meteorological and hydrometric data series under homogeneous conditions is currently of interest to the scientific community (Costa & Soares, 2006), so it must go through a validation process before being used in other applications. To achieve this objective, it is necessary to apply verification and processing methodologies to identify the stations that provide users with reliable data.

Currently, there are several methods that allow evaluating the properties of precipitation or flow time series, i.e., evaluating the consistency of the series (Salas, 1980). Authors such as Campos Aranda (1998) have implemented some tests to evaluate the properties of the series, but the procedure is oriented to the isolated application of the tests. Currently, there is no joint methodology that frames the process to be followed for the evaluation and validation of precipitation and flow time series in water-related studies. The current literature, shows the consistency tests par excellence: to test the independence of meteorological series, the Anderson Limits test is used; for homogeneity, Helmert statistical test, sequence test, double mass curve, or some more specific ones such as Student's t or Cramer; for flow series, tests such as the Runoff Coefficient or Relative Modulus are used (Sanchez, 2017; Campos, 2007). These methods can be a powerful tool in the data management stage, if they are applied correctly and timely under a methodology that allows the orderly and comprehensive analysis of precipitation and flow series, which also shows the guidelines to consider a time series as valid.

In this sense, this article proposes an integral methodology for the treatment and validation of time series, both of precipitation and flow rates. The methodology is based on existing statistical methods that have proven to be valid and reliable, classifying them in different stages of application according to the characteristics of the series and the case under study; so that, from the data provided by official institutions, users can evaluate the series provided, to ensure that they are reliable, or otherwise allow them to be discarded if they do not meet the minimum characteristics for the results obtained to be valid.

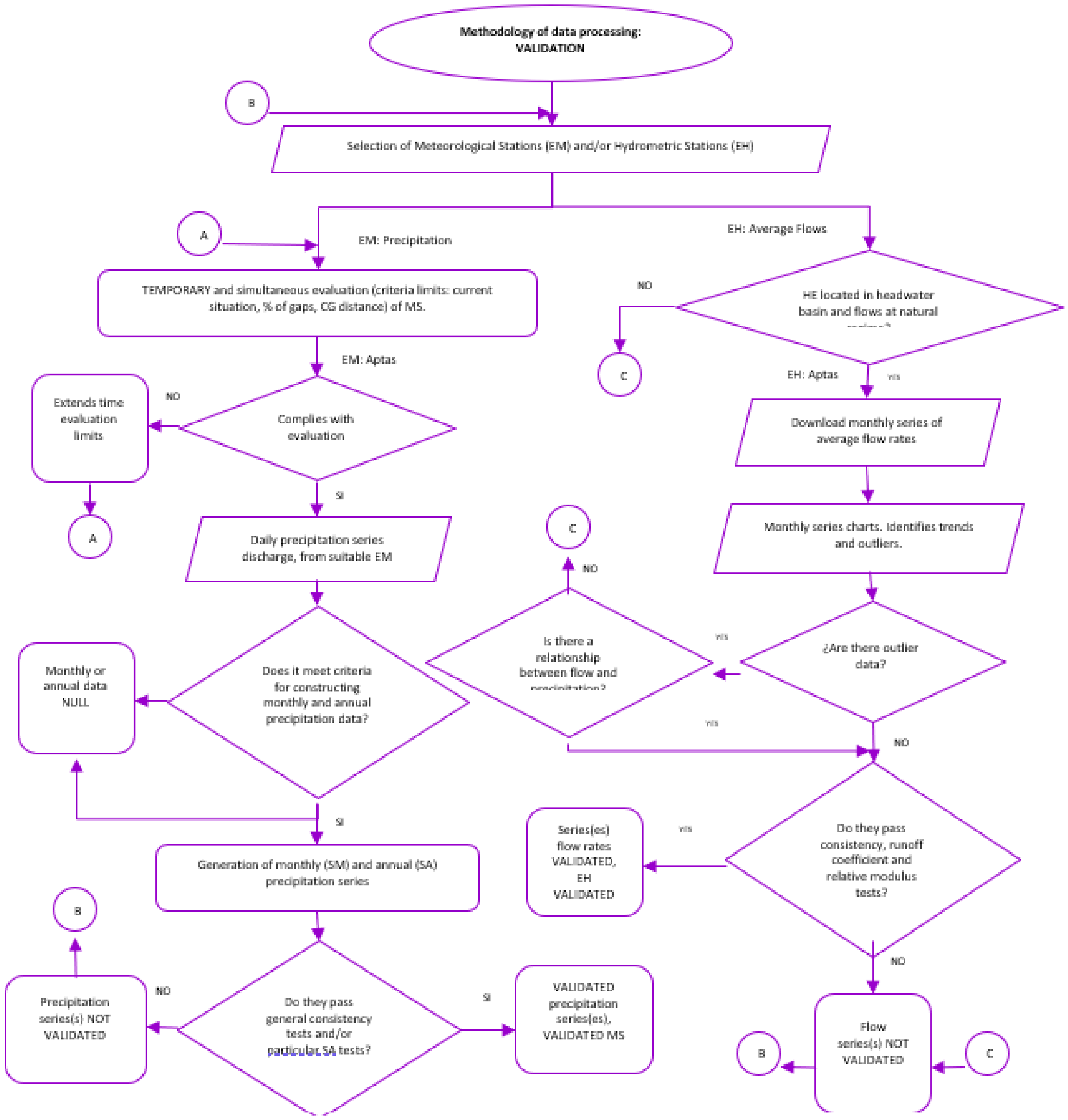
The methodology proposes a set of ordered tests to evaluate the consistency in the precipitation and flow series, in such a way that it complies with the conditions of randomness, homogeneity, independence (persistence for the hydrometric series) and seasonality (Merlos et al., 2014), providing users of this information with less uncertainty associated with these series, which are often the basis for the study of other variables such as floods, droughts or climate change.

7.2 Methodology

The methodology proposed in this work can be divided into two main aspects. On the one hand, the proposal for the treatment of meteorological data is presented, and on the other hand, the analysis of data from hydrometric series is shown. In most studies related to hydrology and in research related to water resources and the environment, both types of data are always involved: precipitation and hydrometry (Guajardo-Panes et al., 2017; Walker, 2000). However, their implementation and reliability depend entirely on the quantity, quality and characteristics of their data, which generally present errors and failures due to various factors.

Figure 7.1 shows the general scheme of the methodology proposed in this work, two branches can be clearly distinguished. The one on the left side represents the treatment of the meteorological series or stations (EM), and the branch on the right side shows the stages for the treatment of the flow series (mean flows) or hydrometric stations (EH). Three stages can be distinguished on both sides: selection of the stations suitable for analysis, obtaining or generating the series, and analysis of the consistency of the series to determine their reliability as a time series of precipitation or flow rates.

Figure 7.1 General methodological scheme



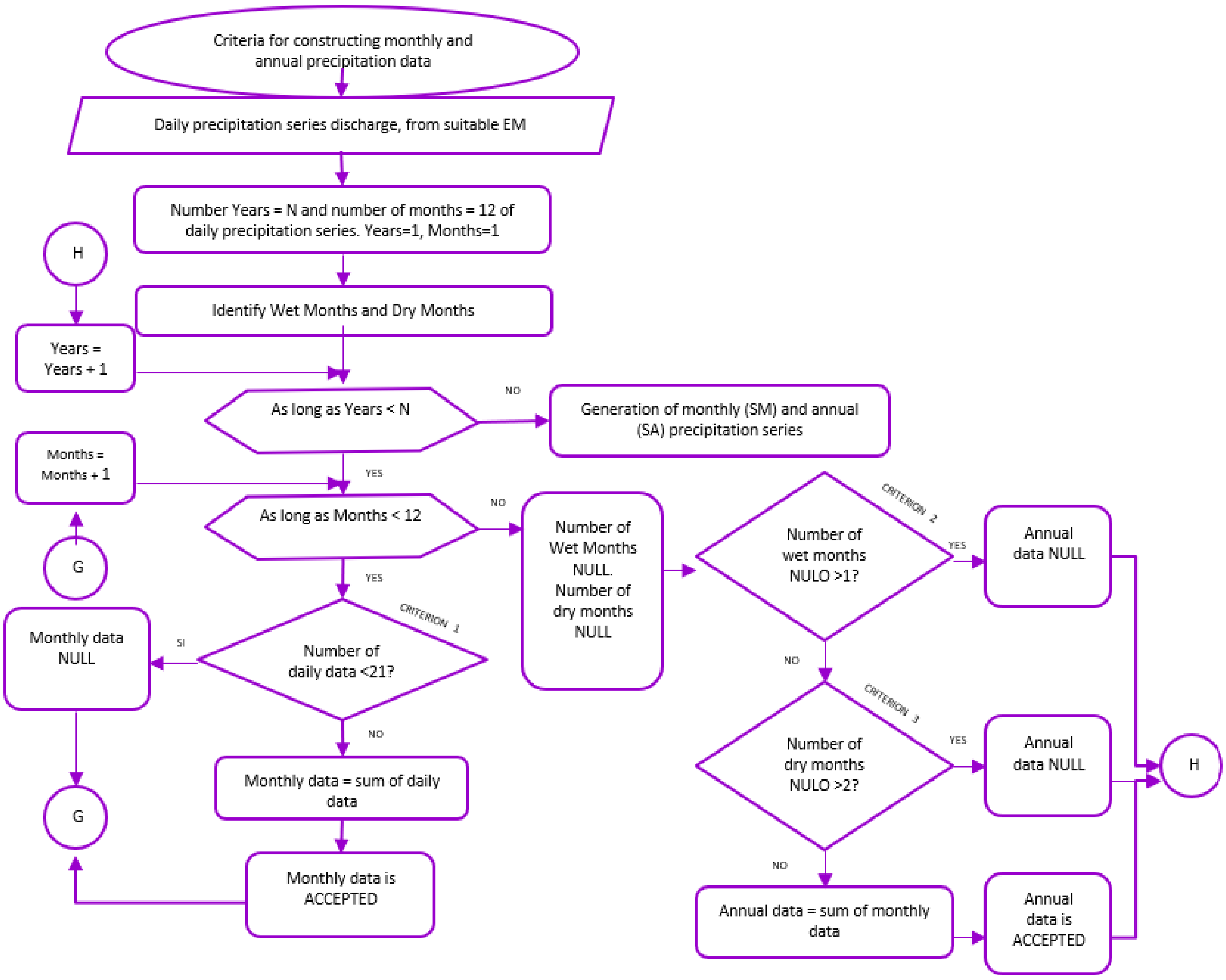
Source: Own elaboration

The left branch (Figure 7.1) allows the meteorological data to be analyzed. The first stage of the methodology shows the process to be followed for the *selection of weather stations*, and subsequently obtain the precipitation time series. The first step for the selection of the meteorological stations consists of carrying out an exploration of all the existing stations in the study area (spatial distribution). Subsequently, a classification of the available stations should be made in order to select the best ones based on five criteria: i) number of years of service, ii) percentage of missing data, iii) distance of the stations from the center of gravity of the area under study (or hydrological basin), iv) degree of data updating and v) geographical location of the station with respect to the rest of the available stations (generally a uniform distribution of the stations is sought, so that they cover the area). From the above classification, only the stations considered as suitable should be selected, those with the highest number of years of service, the lowest percentage of missing data, the shortest distance to the center of gravity of the area under study, those with the most updated information and those with a useful geographic location. Generally, this information is concentrated in a table, where the stations should be classified based on the selection criteria described above. Subsequently, weights are assigned by intervals of each criterion and weights for each criterion of the station, allowing the first group of weights, assigning a weight to each criterion interval of the station and to determine the overall weight of each station the second group of weights is applied (this second group of weights can be one or more scenarios of combination of weights for the criteria), it can be decided by a single scenario, or by the average of weights of the scenarios. The stations with a high overall weight are the best rated and will in turn be the selected stations.

Once the best stations have been selected, we will then proceed with the *obtaining and generating series for analysis*. The first step is to access the precipitation time series (only from the selected stations) through the official databases. Generally, the data available to the user are on a daily scale (CICESE, 2018). Starting from the data obtained, generally incomplete and discontinuous series, it is necessary to *the generation of the series* which will be used in the last stage of the meteorological series analysis (series consistency stage).

The generation of the series makes it possible to obtain the data on the indicated time scale (daily, monthly or annual) for the application of each of the tests necessary to evaluate the two indispensable properties of the precipitation series: homogeneity and independence. To generate these series, it is necessary to perform a previous analysis of the available data in each one, since they are series, usually incomplete, it is necessary to establish criteria to select the data that will be part of the series to be analyzed in the data consistency stage. Due to the variability that characterizes precipitation, the annual precipitation obtained from a year in which the missing data were wet months is not considered with the same degree of uncertainty as that in which the missing data correspond to dry months. The diagram in Figure 7.2 establishes the criteria proposed in this methodology for deciding which data are suitable for consideration in the data consistency stage and which should be left out due to the degree of uncertainty associated with the availability of the information.

Figure 7.2 Criteria for the generation of monthly and yearly series from daily data



Source: Own elaboration

The criteria establish that monthly series are generated from daily data, and annual series from monthly data. Basically, three criteria are established: i) if a month has less than 21 days of information, then the month cannot be considered in the data consistency stage and is considered null. ii) If a year lacks two or more wet months, the year is considered null. iii) Finally, if a year lacks 3 or more dry months, then the year must be considered null. These criteria established in the methodology may vary and in some cases more risky or more conservative criteria may be selected, especially according to the study to be carried out. The important thing at this stage is to establish and apply criteria with knowledge of the risk assumed in the generation of the monthly and annual series.

The time series (annual scale) obtained at this point will be evaluated through the data consistency stage, where the objective is to evaluate the reliability of the precipitation data through the application of general and specific tests to prove the homogeneity and independence of the series, characteristics of a precipitation series. Here begins the third stage, series consistency analysis.

Homogeneity and independence are two properties that precipitation series must have to be considered valid (Guajardo-Panes et al., 2017). If any of these characteristics is not met in a series, it is advisable that it not be used for other studies. For homogeneity, two types of tests have been proposed: general and specific. The former should be applied in all cases, and when there is a discrepancy between the results of these, then at least one specific test should be applied. To the first group belong: the Sequence test, Helmert statistical test and Double Mass Curve. As specific tests, the proposed methodology suggests:

Cramer's statistical test, Student's t-test and the Wald-Wolfowitz statistical test; there is the possibility of performing more than one specific test and if there is discrepancy between the results, homogeneity is accepted or rejected by establishing the criterion of the number of tests that pass homogeneity. Regarding the independence of the series, the Anderson Limits test has been selected with 95% reliability. If the application of the tests selected and proposed in this phase shows that the analyzed precipitation series comply with the properties of homogeneity and independence, then the reliability of the series can be guaranteed for use in other studies.

Note the right branch of the diagram in Figure 7.1. This phase shows the process for the analysis of the series of mean flows recorded at the hydrometric stations. Again, three stages can be distinguished: selection of the stations suitable for analysis, obtaining or generating the series, and analyzing the consistency of the series.

The selection of a suitable hydrometric station depends mainly on the nature of the data recorded by the station, since the station should be in a natural regime. As a result of water management and watershed regulation, some of the data measured by the stations are altered. In this case, prior to analyzing the reliability of the hydrometric data, they must undergo a process called restitution of flows to the natural regime (Solera, 2003). According to the methodology proposed in this work, the flow series (in natural regime) must comply with two characteristics or properties: homogeneity and persistence.

If it is a station in natural regime, generally located in the headwaters of the basins (areas where regulation is practically null), the suitable hydrometric station has been selected, and the corresponding information will be downloaded, accessing the time series of average flows through the official databases (CONAGUA, 2016). These are generally available to the user on a monthly scale.

At this stage it is established that the generation of annual series is made from monthly scale data. Basically, a criterion is established: if a year lacks three or more months, the year is considered null. This criterion established in the methodology may vary, selecting in some cases riskier or more conservative criteria, especially according to the study to be carried out. It can also be considered to establish two criteria instead of one criterion: one for dry months and another for wet months, as indicated in the precipitation. The important thing at this stage is to establish and apply criteria with knowledge of the risk assumed in the generation of the annual series.

In the consistency analysis stage, general and specific procedures are identified. Within the general procedures, it is recommended to perform a graphic analysis of the data obtained in order to identify the trend and outliers that may represent Sánes in the records or the measuring instruments. To determine if it is an erroneous value, the first step should be to compare the flow recorded at a given time with the precipitation values recorded for the same date, then determine if the data comes from an extreme event or if it is an error in the data obtained. In this case, it is suggested not to consider the hydrometric station in the data consistency analysis and therefore reject the station for the study.

Once the hydrometric stations have been visually purified, three specific procedures are performed on the annual series of mean flows; first, two procedures are applied: the Runoff Coefficient (Ce) and Relative Modulus (Mr). Finally, consistency tests are applied to the data: the Sequences test, the Helmert statistical test (both evaluate homogeneity); and the Anderson limits test (evaluates persistence). If the series meet the conditions established in these tests, then it can be said that the hydrometric data analyzed are validated, and therefore can be used with certain reliability in other studies or research. It should be noted that in the case of the consistency tests, if only one test meets the homogeneity, it is at the modeler's discretion to reject; and with respect to the persistence test, the test is not necessarily met in all the average flow series.

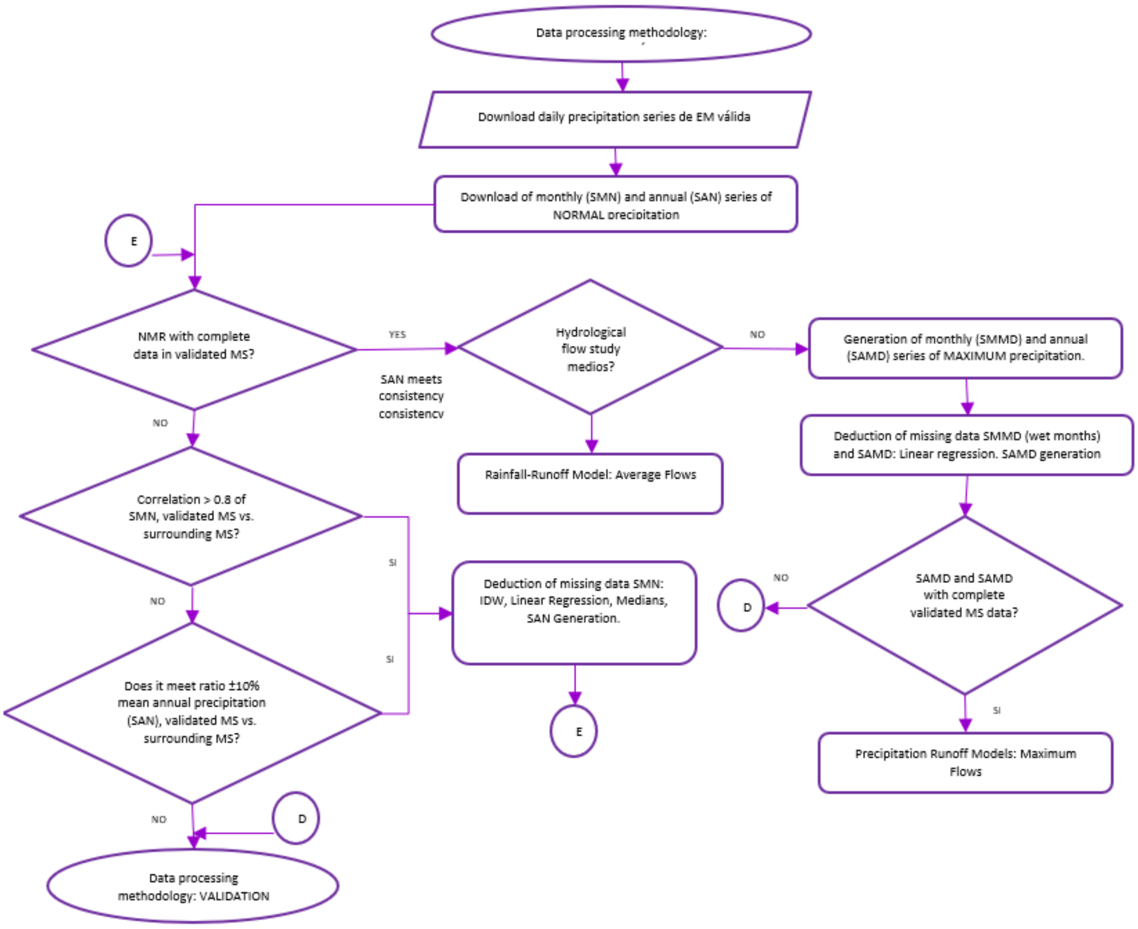
The mathematical description and the theoretical basis of the statistical tests selected in this methodology can be found in the following sections. The methodology was applied in a zone of the metropolitan area of the city of Guadalajara in Mexico. The theoretical development of the tests is accompanied by the result obtained for the application case, in order to achieve a better understanding of the methodology proposed in this work.

The meteorological and hydrometric stations thus validated in this methodology contain the precipitation and mean flow series, respectively, with gaps at the monthly and annual scales. There are several methods for the deduction, estimation or filling of missing data in order to make the series robust, in quantity and quality of information. Once the missing data have been estimated, it is recommended to go to the data generation and consistency stage of the proposed methodology, either in the left or right branch of the methodology in Figure 7.1, in order to validate the deduced data.

A methodological proposal for the deduction of missing data for precipitation series is presented in Figure 7.3. This is done on a monthly scale, for Normal and Maximum Daily precipitation data. This deduction of missing data is limited to Normal precipitation when the hydrological study is of average flows; and is extended to Maximum Daily precipitation if the hydrological study is of maximum flows. The deduction of Normal precipitation requires surrounding meteorological stations, and the deduction of Maximum Daily precipitation requires the Normal precipitation deduced from the same station.

The deduction of missing data from the Monthly Normal Series (NMSS) of precipitation considers 2 methods of selection of surrounding stations: spatial correlation and +/- 10% mean annual precipitation. It also considers the following missing data deduction methods: IDW, simple or double linear regression, averages and natural neighbor method (Mejía et al., 2019). The deduction of missing data, differentiating or not, wet months from dry months, is at the discretion of the user. The generation of missing data of the Normal Annual Series (SAN) follows Criterion 2 and Criterion 3 established in Figure 7.2. The deduction of data from the Monthly Maximum Daily Series (MMDS) of precipitation considers the simple linear regression method, and is only performed for the wet months because these are the ones that have an impact on the annual maximum daily precipitation. The generation of the missing data of the Annual Maximum Daily Series follows Criterion 2 established in Figure 7.2.

Figure 7.3 Deduction, estimation or filling of missing precipitation data in the data measured at weather stations



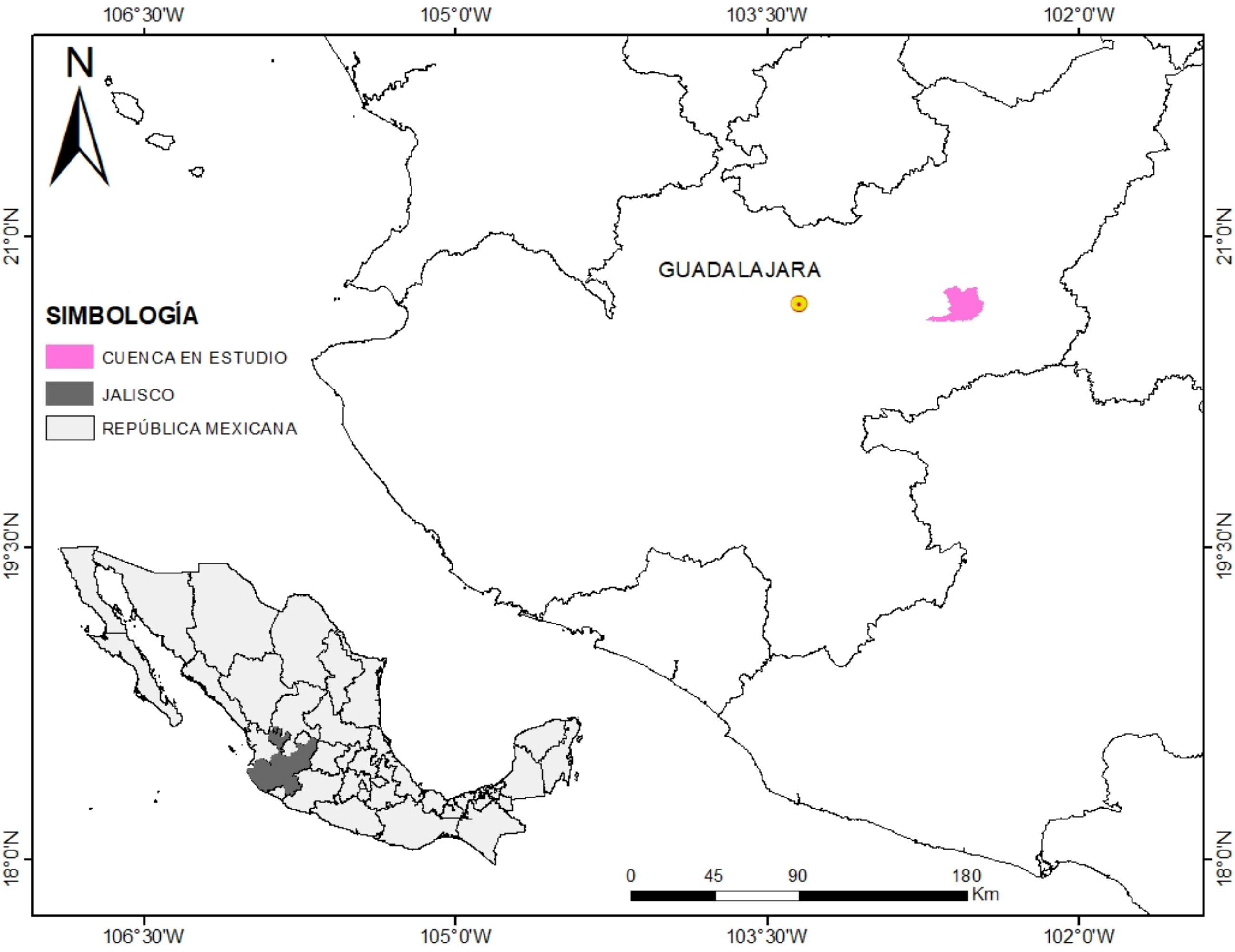
Source: Own elaboration

7.3 Application case

The Guadalajara Metropolitan Area (ZMG) in the state of Jalisco is the second largest urban center in the country. (Some of the cities that make up the region were formerly agricultural towns (i.e. Zapopan), but accelerated population growth in urban areas, inadequate planning of water infrastructure and inefficient regulation of pollutants have become some of the causes that affect the supply, distribution and quality of the water resources available in the area. The search for alternative water supply systems and the reuse of water are nowadays imperative functions for cities and municipalities. Taking care of water and giving it an efficient use with a focus on sustainability is fundamental for the development of the state of Jalisco (Lugo, 2014). The inputs for any study focused on the conservation of water resources are mainly time series of precipitation and flow rates.

7.4). Some of the cities that make up the region were formerly agricultural towns (i.e. Zapopan), but accelerated population growth in urban areas, inadequate planning of water infrastructure and inefficient regulation of pollutants have become some of the causes that affect the supply, distribution and quality of the water resources available in the area. The search for alternative water supply systems and the reuse of water are nowadays imperative functions for cities and municipalities. Taking care of water and giving it an efficient use with a focus on sustainability is fundamental for the development of the state of Jalisco (Lugo, 2014). The inputs for any study focused on the conservation of water resources are mainly time series of precipitation and flow rates.

Figure 7.4 Macro-localization of the application case (hydrological basin)



Source: Own elaboration

Figure 7.5 shows the location of the application case or study basin, it is located at the intersection of the municipalities: Atotonilco el Alto, San Ignacio Cerro Gordo, Tepatitlán de Morelos and Tototlán; with a population of approximately 226,536 inhabitants. The economy of this zone depends mainly on cattle raising, agriculture, fishing and commerce. The main watercourse in the area is the Los Morales stream (IIEG, 2020).

An initial exploration of all meteorological and hydrometric stations available in the study area was carried out. In order to perform a first discretization of the stations, the closest surrounding stations with a low percentage of missing data were selected. This ensures that the sample is representative for the analysis and, therefore, the application of the consistency tests is simpler and more reliable.

Selection of weather stations (temporal and spatial analysis)

In this sense, Table 7.1 shows information on the 5 stations closest to the basin under study. Criteria have been applied to qualify the stations taking as a reference the temporal selection criteria established in Section 2 (years of service, percentage of missing data, distance to the center of gravity of the area under study), where it is observed that one of the most influential parameters is the Euclidean distance between the station and the center of gravity of the watershed, since this ensures that the information is representative of the area under study.

Table 7.1 General information on weather stations

Code	Name	Period of Years Service	Cash	% Voids	D.E.* (m)
14076	Jesús María, Jal	70.9	67.2	5.3%	38,932
14080	La Cuña, Jal	65.6	63.2	3.7%	44,161
14086	La Manzanilla de la Paz, Jal.	64.3	54.3	15.6%	99,177
14087	La Red, Jal.	53.6	52.2	2.6%	24,609
14121	Guadalajara (SMN), Jal	42.0	38.0	9.5%	86,934
near the basin under study. * Euclidean distance to the center of gravity of the basin.					

Source: Own elaboration

The calculation of the Euclidean distance corresponds to the distance from the meteorological station to the center of gravity of the basin and is obtained with the expression of Equation 1. To apply it, it is necessary to know the geographic location of each station in UTM coordinates and the coordinates of the center of gravity of the basin. This information is summarized in Table 7.2.

$$\boldsymbol{D.E.} = \sqrt{(\boldsymbol{x_2} - \boldsymbol{x_1})^2 + (\boldsymbol{y_2} - \boldsymbol{y_1})^2 + (\boldsymbol{z_2} - \boldsymbol{z_1})^2} \qquad (1)$$

Where:

- x_i is the x-coordinate of station i; x_j represents the x-coordinate of station j (in this case the center of gravity of the basin).
- y_i is the y-coordinate of station i; y_j represents the y-coordinate of station j (in this case the center of gravity of the basin).
- z_i is the z-coordinate of station i; z_j represents the z-coordinate of station j (in this case the center of gravity of the basin).

Table 7.2 UTM coordinates of the meteorological stations and center of gravity of the basin

Code	Coordinates UTM X (m)	Y (m)	Z (m)
14076	791857	2280390	2129
14080	728685	2323722	1490
14086	688313	2212493	2050
14087	729140	2290502	1746
14121	666632	2289769	1567
C.G.*	753528	2287213	1893
* Center of gravity of the basin.			

Source: Own elaboration

Since the problem is to identify the most suitable meteorological station or stations, the decision is based on a discrete multi-criteria structure, very useful when more than 3 stations are to be evaluated. The criteria analyzed in the case study are: effective years of precipitation, percentage of gaps and Euclidean distance to the center of gravity of the basin.

For each criterion, criterion intervals are established, for which the maximum and minimum value of each criterion is identified, and the number of intervals is defined, assigning 5 criterion intervals in our case, this value is higher or lower, according to the number of stations under analysis. Table 7.3 presents the weights per criterion interval, where the highest weight is assigned to the highest number of effective years, the lowest percentage of voids, and the smallest Euclidean distance to the center of gravity.

In the table 7.3, for each station, weights are assigned to each criterion, following Table 3.4, establishing different scenarios, which can be two or three; for the example, two scenarios were established; the first scenario considers equal weight to the three criteria, assigning 0.33 to each one. The second scenario considers 0.5 for the percentage of voids criterion, 0.3 for the Euclidean distance criterion and 0.2 for the effective years of precipitation criterion.

The result per scenario is an overall weight per station, which allows assigning a priority to the station, with 1 corresponding to the highest priority and therefore the most suitable. For the two scenarios, the priority results do not coincide with the same stations as the best. In this situation, the most representative scenario of the weight of the criteria is selected, selecting scenario 2. The best stations are those that correspond to the lowest priority order, thus selecting stations 14076 (priority 2) and 14087 (priority 1).

It is important to point out that the number of stations to be selected depends on whether there are stations within the basin and the area of the basin under study (if there are no stations within the basin), in the case of the study corresponds to 2 stations (Aparicio, 1989).

Table 7.3 Assignment of weights to criterion and criterion intervals for case studies

Stations	Criteria						Scenarios			
	Effective years Precipitation		Percentage of voids Precipitation		Euclidean Distance Station with Basin (m)		E1		E2	
	C1		C2		C3		0.33C1+0.33C2+ 0.33C3		0.20C1+0.50C2+ 0.30C3	
	Value	Weight	Value	Weight	Value	Weight	Overall weight	Priority	Overall weight	Priority
14076	67.2	5	5.3	4	38,932	5	4.7	1	4.5	2
14080	63.2	5	3.7	4	44,161	4	4.3	2	4.2	3
14086	54.3	3	15.6	1	99,177	1	1.3	4	1.4	5
14087	52.2	3	2.6	5	24,609	5	4.3	2	4.6	1
14121	38.0	1	9.5	3	86,934	1	1.7	3	2	4

Source: Own elaboration

Table 7.4 Assigning weights to criterion and criterion intervals

	Effective Years of Precipitation. C1		Percentage of voids. C2		Euclidean distance. C3	
Maximum Value	68		1		100,000	
Minimum Value	38		16		25,000	
Interval	Intervalo	Weight	Intervalo	Weight	Intervalo	Weight
1	38-44	1	1-4	5	25,000-40,000	5
2	44-50	2	4-7	4	40,000-55,000	4
3	50-56	3	7-10	3	55,000-70,000	3
4	56-62	4	10-13	2	70,000-85,000	2
5	62-68	5	13-16	1	85,000-100,000	1

Source: Own elaboration

From the analysis of Table 7.3, it is concluded that giving equal weight to the criteria may not be correct (scenario 1), and it is advisable to give different weights (scenario 2) when the interaction of the criteria is understood, for the case study, more weight is given to the percentage of voids, and the second in importance is the distance of the station to the center of gravity, leaving in third place the effective years. In both scenarios, stations 14076 and 14087 have a high priority to be selected.

It can be seen that the stations with the shortest distance and the lowest percentage of voids out of the 5 have been selected. Station 14080 (priority 3) would be the next useful station, if spatially evaluating the selected stations does not meet the spatial coverage of the basin, this third station would be used. According to the Thiessen polygon plot, it was determined that stations 14076 and 14087 are the ones that cover the entire basin, and therefore it is not necessary to select more support stations.

Precipitation generation and series generation

Table 7.5 and Table 7.6 present the annual precipitation series for stations 14076 and 14087, respectively. The generation of these series was performed based on the criteria for series generation shown in the Methodology section of this work. These data are basic for the application of the tests selected in this methodology. Each test has specific considerations that will be specified in each case, but without exception they must be continuous series (except for the Double Mass Curve test) with at least 12 years of data.

Table 7.5 Annual precipitation series for station 14076

# Data	Year	PMA	# Data	Year	PMA	# Data	Year	PMA
1	1944	659.1	25	1968	995.9	49	1992	NULL
2	1945	665.5	26	1969	675.9	50	1993	NULL
3	1946	999.1	27	1970	613.8	51	1994	583.5
4	1947	964.5	28	1971	1079.6	52	1995	1019
5	1948	862	29	1972	764.2	53	1996	630.5
6	1949	851	30	1973	980.7	54	1997	798
7	1950	774.5	31	1974	884.2	55	1998	941.5
8	1951	640.5	32	1975	997.1	56	1999	515.5
9	1952	1109	33	1976	1078.3	57	2000	703
10	1953	998.5	34	1977	935.9	58	2001	842.5
11	1954	838.5	35	1978	640.3	59	2002	814
12	1955	1213	36	1979	608.6	60	2003	1126
13	1956	818.5	37	1980	856.4	61	2004	1033.4
14	1957	645	38	1981	881.5	62	2005	817.3
15	1958	1261	39	1982	676.5	63	2006	923.5
16	1959	1211.5	40	1983	454.7	64	2007	921.5
17	1960	716.5	41	1984	762.9	65	2008	808.5
18	1961	1022.5	42	1985	1015.8	66	2009	948.3
19	1962	854	43	1986	984.3	67	2010	985.7
20	1963	1202.3	44	1987	533.7	68	2011	651
21	1964	846.3	45	1988	757.5	69	2012	738.6
22	1965	943.4	46	1989	802.7	70	2013	1150.5
23	1966	767.8	47	1990	835.8	71	2014	1145.5
24	1967	997.6	48	1991	NULL			

Source: Own elaboration

Table 7.6 Annual precipitation series for station 14087

# Data	Year	PMA	# Data	Year	PMA	# Data	Year	PMA
1	1961	802	19	1979	625.5	37	1997	869.6
2	1962	823.9	20	1980	896.1	38	1998	910.8
3	1963	871.	21	1981	898.4	39	1999	700
4	1964	865.4	22	1982	696.6	40	2000	668
5	1965	1121.9	23	1983	913.5	41	2001	514.9
6	1966	1070.7	24	1984	867.8	42	2002	931.2
7	1967	1360.7	25	1985	833.5	43	2003	904.3
8	1968	915.2	26	1986	943	44	2004	991.5
9	1969	37.3	27	1987	873.4	45	2005	829.3
10	1970	849	28	1988	822.5	46	2006	847
11	1971	884	29	1989	605.9	47	2007	893.1
12	1972	1011.4	30	1990	990.2	48	2008	1103.4
13	1973	1060.8	31	1991	960.8	49	2009	893.3
14	1974	864.4	32	1992	1128.7	50	2010	971.4
15	1975	948.8	33	1993	638.6	51	2011	471
16	1976	1081.4	34	1994	745.5	52	2012	775.7
17	1977	739.9	35	1995	928	53	2013	1114.8
18	1978	878.5	36	1996	829.2	54	2014	731.2

Source: Own elaboration

Selection of hydrometric stations

For the selection of hydrometric stations, it is suggested to generate a buffer of influence over the study area, in order to determine the stations that can be modeled, and in this case, to perform the modeling, it is necessary to consider that the basins must be in a natural regime.

For the case of application, the hydrometric station that generates the basin under study has the code 12607 "La Yerbabuena", located in the state of Jalisco, which has information from August 1965 to November 1992. The annual data series used for the review of the station in question consists then of 27 years and is presented in Table 7.7. The hydrometric station in question generates a basin of 299.20 km^2 of surface area.

Obtaining and generating series of average flow rates

To obtain the information for each year, the hydrometric station was checked to ensure that it had at least 75% of the monthly information (9 months minimum).

Table 7.7 Annual series of hydrometric station 12607 "La Yerbabuena"

Year	Annual Volume (hm^3)	Year	Annual Volume (hm^3)
1966	67.40	1980	50.48
1967	177.48	1981	72.42
1968	167.50	1982	15.02
1969	21.59	1983	92.17
1970	94.29	1984	133.05
1971	103.60	1985	91.37
1972	34.39	1986	48.36
1973	170.79	1987	29.55
1974	41.07	1988	87.55
1975	127.08	1989	12.73
1976	189.41	1990	54.02
1977	92.64	1991	146.46
1978	38.57	1992	120.55
1979	18.15		

Source: Own elaboration

Also shown is the monthly mean flow information (Table 7.8) from hydrometric station 12607, which will be useful for the calculation of the relative modulus.

Table 7.8 Monthly average flows from hydrometric station 12607

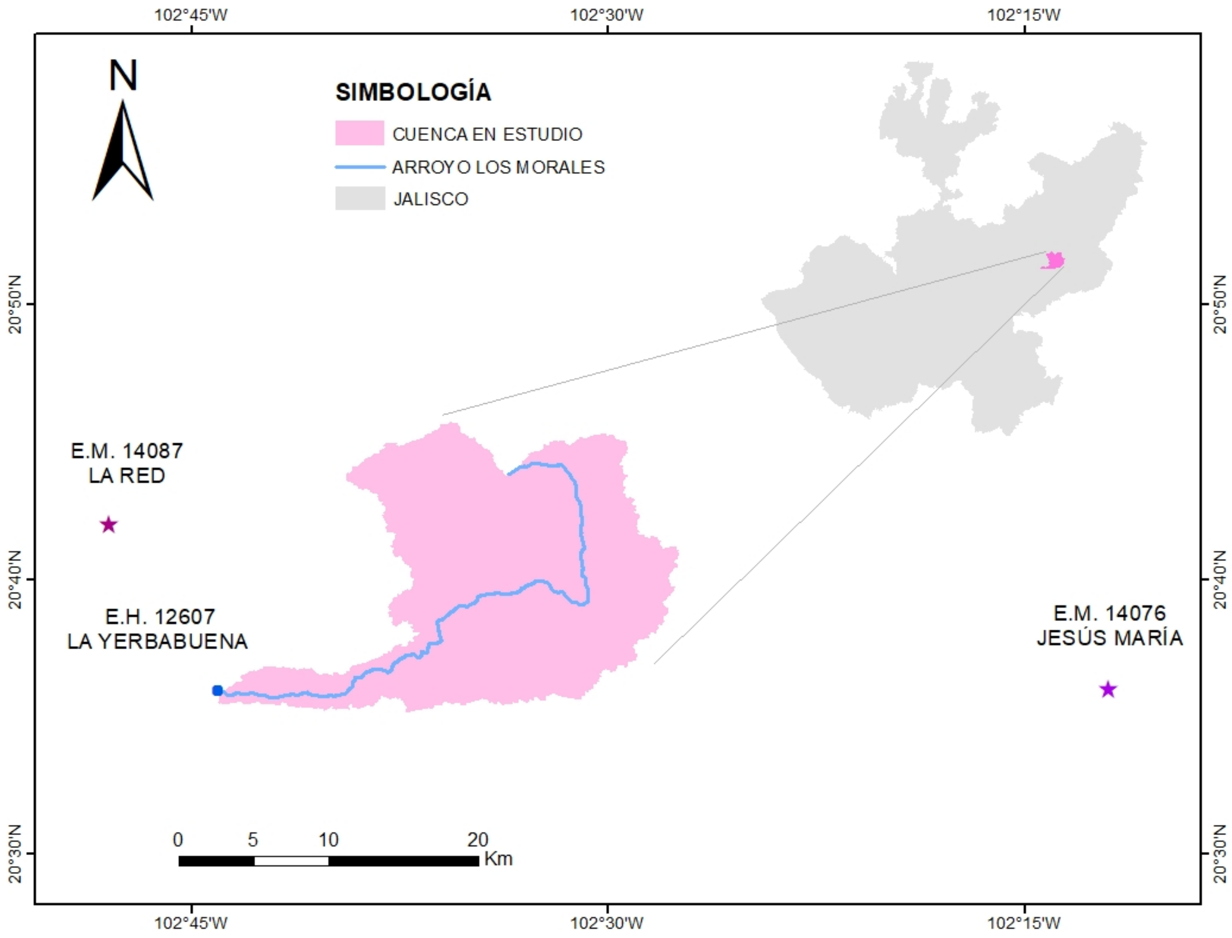
Month	Average monthly flow rate (hm^3)
October	9.57
November	2.11
December	0.76
January	0.94
February	0.54
March	0.38
April	0.32
May	0.36
June	2.10
July	20.97
August	26.00
September	22.85

Source: Own elaboration

Validation of precipitation and average flow stations

For the case under study, two meteorological stations and one hydrometric station are available (Figure 7.5). According to the established methodology, all the consistency tests proposed in the methodology will be applied to the corresponding stations, and the results and analysis obtained in each case will be shown in the following sections.

Figure 7.5 Location of the basin under study, hydrometric station and meteorological stations



Source: Own elaboration

7.4 Validation of Meteorological Stations

This section describes in detail each of the tests applied to the series from the meteorological stations. In order to facilitate the understanding of each test, the theoretical basis of each one will be developed, and at the same time the results of its application will be presented.

The importance of meteorological stations in a region should be emphasized, since precipitation as an element that defines the climatic conditions of any area is unquestionable. Precipitation data obtained by instruments require, however, specific treatments in order to make them more reliable. Several techniques have been developed over time to process this climatic variable. This section proposes tests that can be applied to precipitation data to ensure their reliability for use in other studies.

For Salas et al. (1980), non-homogeneity in data is common in hydrological time series; it is either human-induced or produced by significant, evolutionary or sudden natural disturbance factors (such as natural disasters or system regulation). In addition, hydrologic data can have significant systematic errors that produce inconsistent series, meaning that they do not meet any of the properties of homogeneity or independence.

Precipitation time series should demonstrate homogeneity in their data. This is achieved through the implementation of the Sequences test (Mather, 1975), the Helmert test (Doorembos, 1976), or Double Mass Curve (Martinez et al., 2006), or Wald-Wolfowitz (Siegel, 2015); in addition to specific tests such as Student's t-test (WMO, 1966), or Cramer's t-test (WMO, 1966). A series is said to be homogeneous when the tests show that the elements present in the sample come statistically from the same population. Similarly, the rainfall series must demonstrate independence. This property is evaluated through the Anderson Bounds test (Salas, 1980). It is said that a series is independent when it is demonstrated that the probability that the occurrence of any precipitation data present in the sample does not depend on the occurrence of the subsequent or preceding precipitation value in time or space.

Several characteristics of the time series, such as the mean, standard deviation and serial correlations, can be affected when a trend and/or a positive or negative jump (slip) occur in hydrological series due to lack of homogeneity and independence, generating greater uncertainty associated with the data. In addition, it should be added that the longer a precipitation series is, the greater the probability that the homogeneity of the series, produced by human activities or by an accidental interruption of nature, or the non-independence of the series, is incurred, attributing this problem to systematic errors (inconsistency).

Table 7.9 shows the meteorological stations selected for this analysis, in addition to the information and characterization of each one (station code, name, period of years of service, percentage of gaps, mean annual precipitation (MAP) and coordinates).

Table 7.9 General data of the weather stations to be used

Code	Name	Period of years		% gaps	PMA (mm)	Coordinates UTM		
		Service	Cash			X (m)	Y (m)	Z (m)
14076	Jesús María, Jal	70.9	67.2	5.3%	864.34	791857	2280390	2129
14087	La Red, Jal	53.6	52.2	2.6%	861.20	729140	2290502	1746

Source: Own elaboration

7.4.1 Homogeneity

Homogeneity tests can be classified into two groups, parametric and nonparametric. The latter are less rigorous than the former, but much simpler to perform. Based on the above, one can speak of general tests and specific tests. Among the general (or non-parametric) tests are: Sequences test, Helmert test and Double Mass Curve; and specific (or parametric) tests are: Student's t, Cramer and Wald-Wolfowitz (Campos, 1998).

It is advisable to apply, as a first approximation, the general tests. If there are discrepancies in the results obtained (one indicates homogeneity and the other does not), then we proceed with the application of the specific tests in order to clarify whether the station is homogeneous or does not comply with this characteristic. The particular tests are generally probability-based tests, where the homogeneity of the series is determined from a null hypothesis (H0) and a rule to accept or reject H0 based on an associated probability.

Additionally, homogeneity tests can be classified into two groups:

- Tests that do not require an additional station to determine the homogeneity of their data, where the analysis is performed with the station's own data: Helmert, Sequences, Student's t-test and Cramer's test, which are explained in the following sections.

- Tests that require at least one auxiliary station nearby to perform the analysis: Double Mass Curve and Wald-Wolfowitz test.

In the following sections, the theoretical basis of each test and an example of application are developed.

Sequence testing

This test consists of analyzing the sign of the deviations of each data with respect to the sample median, and comparing the number of allowed changes (u) based on the sample size *(n)*. There is a number of allowed changes depending on the sample size. If the number of changes recorded is between the values established in the ranges presented in Table 7.10, then the series is said to be homogeneous (Mather, 1975), otherwise the series is non-homogeneous.

Table 7.10 Ranges of changes allowed for the Sequence test, according to the number of data

n	u	n	u	n	u	n	u
12	5 – 8	22	9 – 14	32	13 – 20	50	22 – 30
14	5 – 10	24	9 – 16	34	14 – 21	60	26 – 36
16	6 – 11	26	10 – 17	36	15 – 22	70	31 – 41
18	7 – 12	28	11 – 18	38	16 – 23	80	35 – 47
20	8 – 13	30	12 – 19	40	16 – 25	100	45 – 57

Source: Own elaboration

For the Sequence test, we will then need to obtain the median of the continuous series according to the number of data we have. In this sense, we continue comparing the chronologically ordered series with the value of the mean or median, so if the value in the series is less than the median we place an L (lower, lower) or M (major, higher); then, we continue marking the sequences that are formed with the L/M column, taking into account that when we have a change from L to M or vice versa, the number of sequences that are formed is increased by one.

For the test of sequences in station 14076 (Table 7.10); having a series of 21 years of continuous data (odd number) it is required to obtain the median of the series, which is 842.5 and, we have that the number of sequences that are formed according to the series is 12; This value is within the range established as adequate in Table 7.10, where for a series of 21 data, a number of changes from 8 to 14 are allowed (values rounded to the lower and upper limits for 20 and 22 data), which means that the station is homogeneous.

Helmert test

This test consists of a simple procedure where the series must be ordered chronologically and the sign of the deviations of each data with respect to the arithmetic mean of the series is analyzed. If a deviation with a certain sign is followed by another of the same sign, then it is said that there is an "S" sequence, otherwise it is considered a "C" change. Once the entire series has been analyzed, the number of changes and the number of sequences are counted, and the inequality of Equation 2 is applied. If the inequality is satisfied, the station can be considered as homogeneous (Doorembos, 1976).

$$-\sqrt{n-1} \leq (S-C) \leq \sqrt{n-1} \qquad (2)$$

For the Helmert test, a procedure quite similar to the Sequence test is used, however, the value with which the precipitation series will be compared is the mean (in the same way, marking with M and L; however, now the number of changes between L and M will be considered; in this sense, as long as the series remains at L (M) an S (sequence) is written and, if this value changes from L to M (or vice versa), a C (change) will be written. At the end, we count the S and C and apply the formula proposed in Equation 2.

It is recommended that the Sequence and Helmert homogeneity tests be performed simultaneously. Table 7.11 shows the results obtained from the application of both tests for station 14076 and Table 4.4 for station 14087.

For the Helmert test, it is compared with the mean value of 871.8 mm, which gives a total of 11 changes and 9 sequences, so the range of application is as shown below:

$$-\sqrt{21-1} \leq 9-11 \leq \sqrt{21-1}$$
$$-4.47 \leq -2 \leq 4.47$$

Therefore, according to Helmert's test, the station is homogeneous.

Table 7.11 Sequence and Helmert test for station 14076

Year	PMA	Sequence Testing		Helmert test	
		Comparison	Sequence	Comparison	Changes
1994	583.5	L	1	L	
1995	1019	M	2	M	C
1996	630.5	L	3	L	C
1997	798	L	3	L	S
1998	941.5	M	4	M	C
1999	515.5	L	5	L	C
2000	703	L	5	L	S
2001	842.5	L	5	L	S
2002	814	L	5	L	S
2003	1126	M	6	M	C
2004	1033.4	M	6	M	S
2005	817.3	L	7	L	C
2006	923.5	M	8	M	C
2007	921.5	M	8	M	S
2008	808.5	L	9	L	C
2009	948.3	M	10	M	C
2010	985.7	M	10	M	S
2011	651	L	11	L	C
2012	738.6	L	11	L	S
2013	1150.5	M	12	M	C
2014	1145.5	M	12	M	S

Source: Own elaboration

The results obtained for station 14087 are shown in Table 7.12 The series has 33 years of consecutive data. The median value for this series is 873.4 mm. Analyzing the deviations of each data with respect to the median value, 20 sequences were counted. According to the information in Table 7.10, for a 33-year series, between 13 and 20 sequences are allowed (taking the value for 32 data), so the series can be considered as homogeneous.

For the Helmert test, the mean value for the series is 851.8 mm. The analysis of the deviations from the mean value accounted for a total of 17 changes and 15 sequences. Applying Equation 2, it is observed that the inequality is fulfilled, so the station can be considered as homogeneous.

$$-\sqrt{33-1} \leq 15-17 \leq \sqrt{33-1}$$

$$-5.66 \leq -2 \leq 5.66$$

Table 7.12 Sequence and Helmert test for station 14087

Year	PMA	Sequence Testing		Helmert test	
		Comparison	Sequence	Comparison	Changes
1970	849	L	1	L	
1971	884	M	2	M	C
1972	1011.4	M	2	M	S
1973	1060.8	M	2	M	S
1974	864.4	L	3	M	S
1975	948.8	M	4	M	S
1976	1081.4	M	4	M	S
1977	739.9	L	5	L	C
1978	878.5	M	6	M	C
1979	625.5	L	7	L	C
1980	896.1	M	8	M	C
1981	898.4	M	8	M	S
1982	696.6	L	9	L	C
1983	913.5	M	10	M	C
1984	867.8	L	11	M	S
1985	833.5	L	11	L	C
1986	943	M	12	M	C
1987	873.4	L	13	M	S
1988	822.5	L	13	L	C
1989	605.9	L	13	L	S
1990	990.2	M	14	M	C
1991	960.8	M	14	M	S
1992	1128.7	M	14	M	S
1993	638.6	L	15	L	C
1994	745.5	L	15	L	S
1995	928	M	16	M	C
1996	829.2	L	17	L	C
1997	869.6	L	17	M	C
1998	910.8	M	18	M	S
1999	700	L	19	L	C
2000	668	L	19	L	S
2001	514.9	L	19	L	S
2002	931.2	M	20	M	C

Source: Own elaboration

Double mass curve test

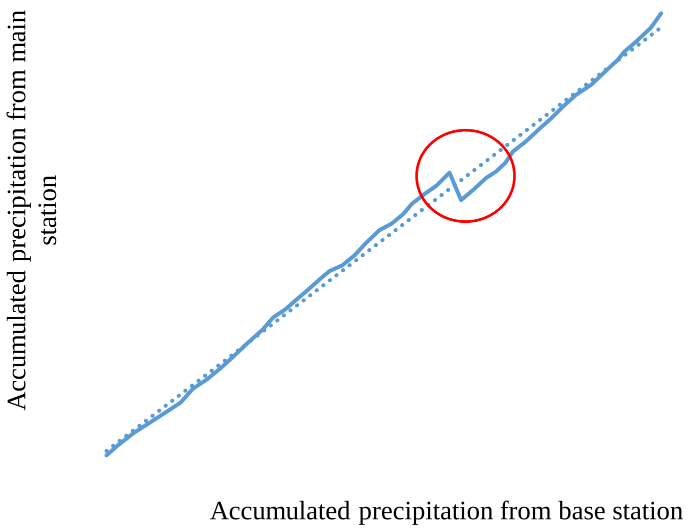
The Double Mass Curve method (Martinez et al., 2006) consists of checking whether the records of a rainfall station have suffered variations that lead to erroneous values. These variations may be due to a change in the instrumental location, a variation in the peripheral conditions of the measurement site or a change in the operator of the equipment of the observer taking the readings.

The mass curve method considers that, in a homogeneous meteorological area, the precipitation values occurring at different points of that area in annual or seasonal periods, have a relationship of proportionality that can be represented graphically. This representation consists of identifying the station to be monitored (main station) and obtaining the annual precipitation value. For the contrast, it will be necessary to have at least one base station whose annual data series must coincide with that of the station to be monitored.

For each station (main and base station), in each year, starting from the first year with a record, the accumulated value of the base station is obtained (if there is more than one, then the values of the base stations are averaged and accumulated for successive years), and the accumulated value of the station to be monitored.

Then, on a system of orthogonal axes, the cumulative annual precipitation values of the station to be monitored are plotted in ordinates and the cumulative mean annual precipitation values of the base station in abscissae. If the records have not undergone variations, the points are aligned in a line with a single and uniform slope, therefore, it will not be necessary to make corrections. If, on the other hand, there are variations in the slope of the line, it means that part of the series contains erroneous values and the data record must be corrected from the year in which the slope of the line changes before it can be used. For this case, it is necessary to obtain a Correction Factor that is proportional to the variation of the slope of the line (Graphic 7.1).

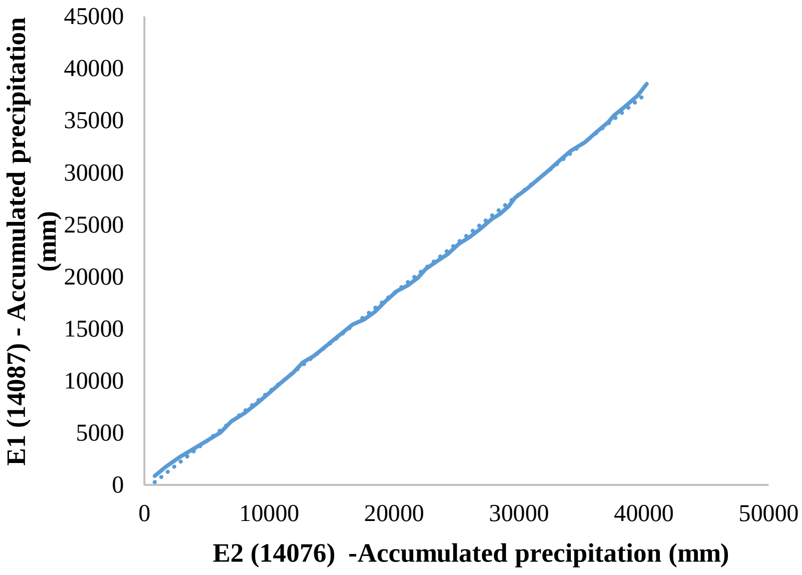
Figure 7.1 Representation of the double mass curve test



Source: Own elaboration

For the application case, station 14087 has been considered as the main station (E1) and station 14076 as the base station (E2). The coincident period for both stations is from 1962 to 2014. Once the accumulated precipitation has been obtained for the main station and the base station, the station homogeneity was determined from the graph in Graphic 7.2.

Graphic 7.2 Results of the double mass curve test, performed with data from stations 14076 and 14087



Source: Own elaboration

In order to make the graphic, the information on mean annual precipitation (MAP) presented in Table 7.13 is used.

Table 7.13 Annual rainfall series for the double mass curve test

Year	PMA (mm)		Accumulated (mm)		Year	PMA (mm)		Accumulated (mm)	
	14087	14076	14087	14076		14087	14076	14087	14076
1962	823.9	854	823.9	854	1988	822.5	757.5	21931.1	19915.1
1964	865.4	846.3	1689.3	1700.3	1989	605.9	802.7	22537	20717.8
1965	1121.9	943.4	2811.2	2643.7	1990	990.2	835.8	23527.2	21553.6
1966	1070.7	767.8	3881.9	3411.5	1994	745.5	583.5	24272.7	22137.1
1967	1360.7	997.6	5242.6	4409.1	1995	928	1019	25200.7	23156.1
1970	849	613.8	6091.6	5022.9	1996	829.2	630.5	26029.9	23786.6
1971	884	1079.6	6975.6	6102.5	1997	869.6	798	26899.5	24584.6
1972	1011.4	764.2	7987	6866.7	1998	910.8	941.5	27810.3	25526.1
1973	1060.8	980.7	9047.8	7847.4	1999	700	515.5	28510.3	26041.6
1974	864.4	884.2	9912.2	8731.6	2000	668	703	29178.3	26744.6
1975	948.8	997.1	10861	9728.7	2001	514.9	842.5	29693.2	27587.1
1976	1081.4	1078.3	11942.4	10807	2002	931.2	814	30624.4	28401.1
1977	739.9	935.9	12682.3	11742.9	2004	991.5	1033.4	31615.9	29434.5
1978	878.5	640.3	13560.8	12383.2	2005	829.3	817.3	32445.2	30251.8
1979	625.5	608.6	14186.3	12991.8	2006	847	923.5	33292.2	31175.3
1980	896.1	856.4	15082.4	13848.2	2007	893.1	921.5	34185.3	32096.8
1981	898.4	881.5	15980.8	14729.7	2008	1103.4	808.5	35288.7	32905.3
1982	696.6	676.5	16677.4	15406.2	2009	893.3	948.3	36182	33853.6
1983	913.5	454.7	17590.9	15860.9	2010	971.4	985.7	37153.4	34839.3
1984	867.8	762.9	18458.7	16623.8	2011	471	651	37624.4	35490.3
1985	833.5	1015.8	19292.2	17639.6	2012	775.7	738.6	38400.1	36228.9
1986	943	984.3	20235.2	18623.9	2013	1114.8	1150.5	39514.9	37379.4
1987	873.4	533.7	21108.6	19157.6	2014	731.2	1145.5	40246.1	38524.9

Source: Own elaboration

Statistical test t of Student

When the cause of the loss of homogeneity of the series is due to an abrupt change in the mean, the parametric Student's t-test is especially useful. So, first of all, what is recommended is to make a graph of the annual rainfall, in which the behavior of the series with respect to time can be observed, in this way, we can delimit the periods of time in which there is a jump (change in the trend of the mean of the series) and that, makes the mean of the rainfall increase or decrease; so, we will have two periods n_1 and n_2, each one with the calculation of the value of the mean X_1 and X_2 respectively.

This test is powerful for detecting inconsistency in the means, and it is a robust test (except when the length of the two periods selected for comparison of their means is unequal, because then the distribution of the data may not be skewed).

It is understood that a test is robust when it is insensitive to the shape of the probability distribution of the series. Due to the above, it is recommended to apply the test of the *t* that the values of n1 and n2 of each mean to be compared $\bar{x}_1$ y $\bar{x}_2$, no sean similares (Campos, 1998).

Student's t statistic is defined by Equation 3 (WMO, 1966):

$$\boldsymbol{t_d} = \frac{\bar{x}_1 - \bar{x}_2}{\left[\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]^{1/2}} \tag{3}$$

Being S_1^2 y S_2^2 the variances of x_i in the two periods of record, respectively. Then, $n_1 S_1^2$ can be calculated with Equation 4.

$$n_1 S_1^2 = \sum_1^{n_1} x_i^2 - \frac{1}{n_1}\left(\sum_1^{n_1} x_i^2\right)^2 \tag{4}$$

And similarly it can be calculated $n_2 S_2^2$.

The absolute value of t_d is generally compared to the value of *t* of the distribution *t* two-tailed Student's method with $\nu = (n_1 + n_2 - 2)$ degrees of freedom and a 5% significance level. The values of *t* are concentrated in Table 7.14.

Table 7.14 Significance values for the values of the distribution of the t de Student and Cramer

Degrees of Freedom	Level of Significance		Degrees of Freedom	Level of Significance	
	5% *	5% **		5% *	5% **
1	6.314	12.706	18	1.734	2.101
2	2.920	4.303	19	1.729	2.093
3	2.353	3.182	20	1.725	2.086
4	2.132	2.776	21	1.721	2.080
5	2.015	2.571	22	1.717	2.074
6	1.943	2.447	23	1.714	2.069
7	1.895	2.365	24	1.711	2.064
8	1.860	2.306	25	1.708	2.060
9	1.833	2.262	26	1.706	2.056
10	1.812	2.228	27	1.703	2.052
11	1.796	2.201	28	1.701	2.048
12	1.782	2.179	29	1.699	2.045
13	1.771	2.160	30	1.697	2.042
14	1.761	2.145	40	1.684	2.021
15	1.753	2.131	60	1.671	2.000
16	1.746	2.120	120	1.658	1.980
17	1.740	2.110	∞	1.645	1.960

* One-tailed test
** Two-tailed test

Source: Own elaboration

Cramer's statistical test

Sometimes, it may be more convenient to compare the mean of the whole series and the mean of a certain part of the record, to investigate the homogeneity; for such purpose Cramer's test is quite useful, when the periods of the series are different. n_1 y n_2 are so similar that the *t* de Student loses validity (Campos, 1998).

The Cramer's test requires the arithmetic mean values $\bar{x}$ (Equation 4) and standard deviation *S* (Equation 5) of the total log of *n* values:

$$\bar{x} = \frac{\sum x_i}{n} \tag{4}$$

$$S = \frac{\sum (x_i - \bar{x})^2}{n-1} \tag{5}$$

On the other hand, $\bar{x}_k$ is the average of the period of n' values (subperiod of n), that has the greatest difference with respect to the average ($\bar{x}$) of the complete series (n); so that the value t_k is calculated according to the formulation shown in Equation 6 to Equation 8 (WMO, 1966):

$$\bar{x}_k = \frac{\sum_{i=k-1}^{i=k+n} x_i}{n'} \tag{6}$$

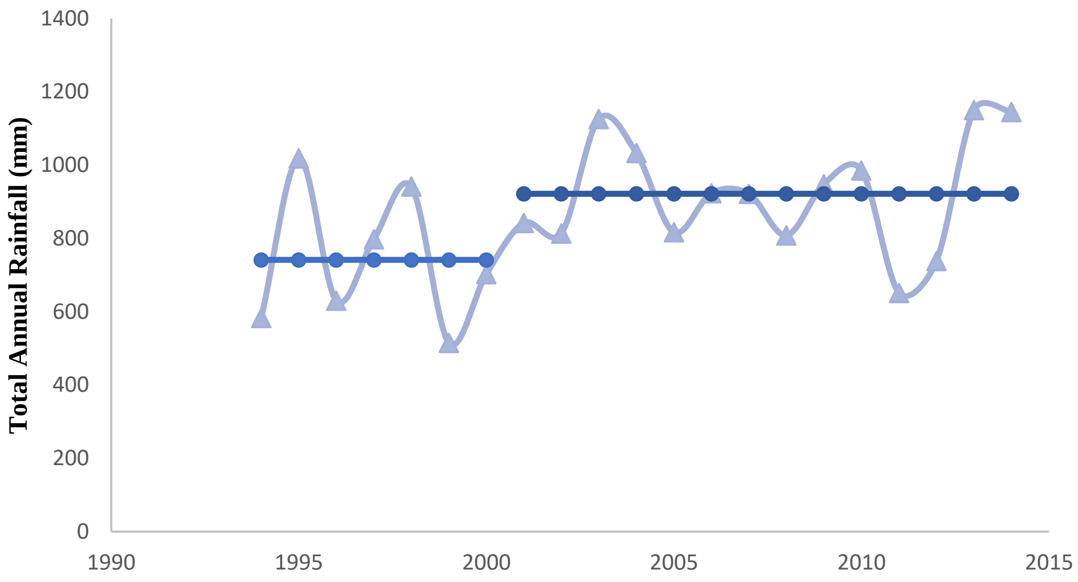
$$\tau_k = \frac{(\bar{x}_k - \bar{x})}{S} \tag{7}$$

$$t_k = \left[\frac{n'(n-2)}{n-n'[1+(\tau_k)^2]}\right]^{1/2} \cdot (\tau_k) \tag{8}$$

As with the Sequences and Helmert tests, it is recommended that the Studen and Cramer's t-tests be performed at the same time, since: i) they are complementary, ii) they are based on the understanding of the graph of the annual total rainfall series and, iii) the significance value with which they are compared is the same (Table 7.15).

Graphic 7.3 represents the annual precipitation values for station 14076, such that two well-marked time periods can be distinguished, with N_1 being the period from 1994 to 2000 and N_2 from 2001 to 2014.

Graphic 7.3 Behaviour of total annual rainfall through time at station 14076



Source: Own elaboration

Once the periods have been delimited, the basic statistics for each period should be obtained, such as mean, variance and number of data (Table 7.15).

Table 7.15 Values for the test of *t* de Student for station 14076

$\overline{x_1}$	741.57	N_1	7	S_1^2	30005.96
$\overline{x_2}$	921.88	N_2	14	S_2^2	22131.95

Source: Own elaboration

These values are substituted into Equation 2, so that the following value of t_d.

$$t_d = \frac{741.57-921.88}{\left[\frac{(7*30005.96)+(14*22131.95)}{7+14-2}\cdot\left(\frac{1}{7}+\frac{1}{14}\right)\right]^{1/2}}$$

$$t_d = -1.43$$

In this case, the absolute value obtained from t_d is 1.43 and, according to the values in Table 4.6, a value of 1.729 is allowed for the 19 degrees of freedom. In this case the series can be considered as homogeneous, since the allowed value is less than the value calculated by the test statistic *t* de Student. Likewise, the Cramer's test is applied, which makes use of the second period of data from the original series (Graphic 7.3) of the Cramer's test (Graphic 7.2). *t* de Student, Therefore, as shown in Table 7.16, we have the information required to apply the formula described in Equation 8.

Table 7.16 Station 14076 values for Cramer's test

Values of the complete series		Secondary series values with jump	
$\overline{x}$	861.78	$\overline{x_k}$	921.88
S	178.83	**n’**	14
n	21		

Source: Own elaboration

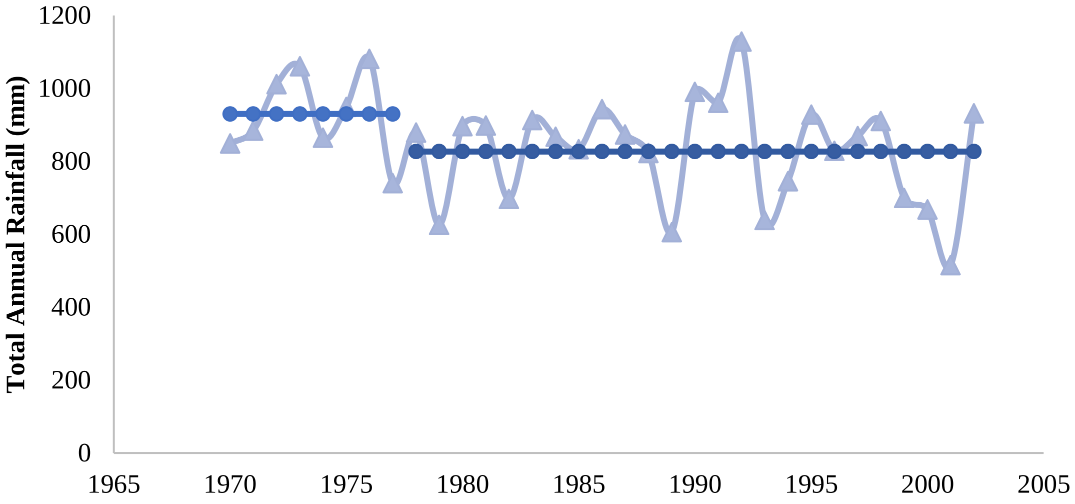
By applying the values shown in Table 7.16 for Equation 7 and Equation 8, it is obtained:

$$\tau_k = \frac{(921.88 - 861.78)}{178.83} = 0.34$$

$$t_k = \left[\frac{14 * (21 - 2)}{21 - 14 * [1 + (0.34)^2]}\right]^{1/2} \cdot (0.34) = 2.35$$

The calculated value of Cramer's statistic ($\boldsymbol{t_k}$) is 2.35. According to the data shown in Table 4.6, the allowed value for 19 degrees of freedom is 1.729 for a significance value of 5% (one-tailed test). Since the value calculated with Cramer's formula is greater than the allowed value, the series cannot be considered homogeneous. If the analysis is performed for a two-tailed test, the allowed value is 2.093, but likewise the value of $\boldsymbol{t_k} = 2.35$ is still higher, so the series cannot be considered as homogeneous.De forma similar, tal como se realizó para la estación 14076, se tiene la gráfica de las precipitaciones totales anuales de la estación 14087 (Figura 4.4). En la cual se muestran los tramos de estudio de N_1 (1970 - 1977) y N_2 (1978 - 2002), con los cuales se puede obtener la información mostrada en la Tabla 7.17.

Graphic 7.4 Behaviour of total annual rainfall through time at station 14087



Source: Own elaboration

Table 7.17 Values for the test of *t* de Student for station 14087

$\overline{x_1}$	929.96	N_1	8	S_1^2	12016.04
$\overline{x_2}$	826.81	N_2	25	S_2^2	19609.90

Source: Own elaboration

With the above values, we obtain the value of t_d, such that:

$$\boldsymbol{t_d} = \frac{\mathbf{929.96 - 826.81}}{\left[\frac{\mathbf{(8 * 12016.04) + (25 * 19609.90)}}{\mathbf{8 + 25 - 2}} \cdot \left(\frac{\mathbf{1}}{\mathbf{8}} + \frac{\mathbf{1}}{\mathbf{25}}\right)\right]^{\mathbf{1/2}}}$$

$$\boldsymbol{t_d} = \mathbf{094}$$

The value obtained from t_d is 0.94 and, for the 31 degrees of freedom that the series has, it is necessary to have a maximum value of 1.697, which is fulfilled; therefore, it is said that the series is homogeneous by means of the test of *t* de Student.

Now, the same is done for Cramer's test, so the results of the test in question are presented, considering the information shown in Table 7.18.

$$\boldsymbol{\tau_k} = \frac{\mathbf{(826.81 - 851.82)}}{\mathbf{140.44}} = \mathbf{-0.18}$$

$$\boldsymbol{t_k} = \left[\frac{\mathbf{25 * (33 - 2)}}{\mathbf{33 - 25 * [1 + (-0.18)^2]}}\right]^{\mathbf{1/2}} \cdot \mathbf{(-0.18) = -1.85}$$

Table 7.18 Station 14087 values for Cramer's test

Values of the complete series		Secondary series values with jump	
$\bar{x}$	851.82	$\overline{x_k}$	826.81
S	140.44	n’	25
n	33		

Source: Own elaboration

Now, we have that, for Cramer's test, the absolute value of t_k is 1.85 and, according to Table 4.6, the maximum allowable value for the 31 degrees of freedom and with a 5% significance level for a one-tailed test is 1.692, so the series is considered to be inhomogeneous; however, when using the 5% significance level for a two-tailed test, the maximum allowable value is 2.042, which is not exceeded and therefore, station 14087 is considered to be homogeneous by the Cramer's test with a 5% significance level for a two-tailed test. 042, which is not exceeded and therefore, station 14087 is homogeneous by means of Cramer's test with a 5% significance level and for a two-tailed test.

Thus, we can determine that, by performing the exercise for the two stations presented in the case study, both are homogeneous by means of the test station. t de Student but, not homogeneous according to Cramer's test; in this sense, we also know that Cramer's test was not necessary to apply because the length of the series proposed for the Cramer's test is not homogeneous according to the Cramer's test. *t* de Student (N_1 y N_2) are different, which indicates, in principle, that it is only necessary to apply this test.

Wald - Wolfowitz Statistical Test

This test makes it possible to determine whether there is any difference between two annual rainfall series of size N_1 y N_2 (from two different weather stations). To apply the test, a single series should be generated by mixing the data coming from the two stations and sorting them in an increasing order. Subsequently, the number of sequences or spells of the ordered series is determined. A sequence is defined as any succession of values of the same series, which are indicated with *X* for the season in which homogeneity is being investigated and with *Y* for the auxiliary station (Campos, 1998).

When samples are small ($N_1, N_2 \leq 20$) Table 7.19 presents the critical values of the number of sequences, so that if a number of spurts were found to be (r) equal to or less than the tabulated value, the series will be different due to a certain cause, at a significance level of 5%.

This test is complementary to any other statistical test for homogeneity of a meteorological station, such as Cramer or t de Student. Then, if the probability obtained for this test exceeds the significance percentage (5% or 10%), one series is homogeneous and the other is non-homogeneous; on the contrary, if the probability value is within the range of the significance level, the two series will be homogeneous or non-homogeneous. This is determined by knowing the homogeneity or non-homogeneity of the series through the results of the test of *t* de Student o Cramer.

Table 7.19 Critical values of the number of sequences in the Wald-Wolfowitz test, for small samples

N_1/N_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

Source: Own elaboration

When N_1 o N_2 > 20, Table 4.11 cannot be used, then the statistic is evaluated. *z*, according to Equation 9 (Siegel, 2015).

$$\boldsymbol{z} = \frac{\left| r - \left(\frac{2N_1N_2}{N_1+N_2} + 1 \right) \right| - 0.50}{\sqrt{\frac{2N_1N_2(2N_1N_2 - N_1 - N_2)}{(N_1+N_2)^2(N_1+N_2-1)}}} \tag{9}$$

If the calculated value of *z* in Equation 4.9 has an associated probability '*p*', read directly in Table 7.20, equal to or less than the adopted significance level (5% and 10%) the series will be different and, therefore, if one of them is known to be homogeneous, the other will be non-homogeneous. Therefore, according to the values in Table 4.12, it is recommended that the absolute value of z be greater than or equal to 1.65 for a significance value of 5% and greater than or equal to 1.29 for a significance value of 10%.

Table 7.20 Associated probabilities '*p*' auxiliary in the Wald-Wolfowitz test

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007

Source: Own elaboration

For the application of the Wald-Wolfowitz test, the precipitation series of the two selected weather stations 14076 and 14087 shown in Table 7.5 and Table 7.6, respectively, are required.

For this test, it is not necessary that the periods be equal or consecutive; it is sufficient that they be representative series (it is recommended to have at least 25 years of data) for each of the stations under study. The data from both stations are combined in the same series and are ordered from lowest to highest, always taking into account which value belongs to each station; henceforth, the values of station 14076 will be referred to as the series *X* and for those of station 14087 such as the series *Y*.

Table 7.21 Number of sequences obtained for the Wald-Wolfowitz test

P (mm)	Serie	# sec.	P (mm)	Serie	# sec.	P (mm)	Serie	# sec.	P (mm)	Serie	# sec.	P (mm)	Serie	# sec.
454.7	X	1	716.5	X	13	842.5	X	21	921.5	X	29	1015.8	X	41
471	Y	2	731.2	Y	14	846.3	X	21	923.5	X	29	1019	X	41
514.9	Y	2	738.6	X	15	847	Y	22	928	Y	30	1022.5	X	41
515.5	X	3	739.9	Y	16	849	Y	22	931.2	Y	30	1033.4	X	41
533.7	X	3	745.5	Y	16	851	X	23	935.9	X	31	1060.8	Y	42
583.5	X	3	757.5	X	17	854	X	23	941.5	X	31	1070.7	Y	42
605.9	Y	4	762.9	X	17	856.4	X	23	943	Y	32	1078.3	X	43
608.6	X	5	764.2	X	17	862	X	23	943.4	X	33	1079.6	X	43
613.8	X	5	767.8	X	17	864.4	Y	24	948.3	X	33	1081.4	Y	44
625.5	Y	6	774.5	X	17	865.4	Y	24	948.8	Y	34	1103.4	Y	44
630.5	X	7	775.7	Y	18	867.8	Y	24	960.8	Y	34	1109	X	45
638.6	Y	8	798	X	19	869.6	Y	24	964.5	X	35	1114.8	Y	46
640.3	X	9	802.7	X	19	871.7	Y	24	971.4	Y	36	1121.9	Y	46
640.5	X	9	808.5	X	19	873.4	Y	24	980.7	X	37	1126	X	47
645	X	9	814	X	19	878.5	Y	24	984.3	X	37	1128.7	Y	48
651	X	9	817.3	X	19	881.5	X	25	985.7	X	37	1145.5	X	49
659.1	X	9	818.5	X	19	884	Y	26	990.2	Y	38	1150.5	X	49
665.5	X	9	822.5	Y	20	884.2	X	27	991.5	Y	38	1211.5	X	49
668	Y	10	823.9	Y	20	893.1	Y	28	995.9	X	39	1213	X	49
675.9	X	11	829.2	Y	20	893.3	Y	28	997.1	X	39	1261	X	49
676.5	X	11	829.3	Y	20	896.1	Y	28	997.6	X	39	1360.7	Y	50
696.6	Y	12	833.5	Y	20	898.4	Y	28	998.5	X	39			
700	Y	12	835.8	X	21	910.8	Y	28	999.1	X	39			
703	X	13	838.5	X	21	913.5	Y	28	1011.4	Y	40			

Source: Own elaboration

As shown in Table 4.13, we have a total of 50 sequences (r), which are obtained considering the change of the series (from *X* to *Y* or vice versa). In addition to the number of sequences, it is necessary to know the number of data of each series, being 67 and 50 data for stations 14076 and 14087 respectively; given that the number of data per series exceeds 20 values, it is not possible to use the **¡Error! No se encuentra el origen de la referencia.**4.11, so that the value of z is obtained according to the formulation described in Equation 9, such that:

$$\boldsymbol{z = \frac{\left|50 - \left(\frac{2 * 67 * 50}{67 + 50} + 1\right)\right| - 0.50}{\sqrt{\frac{(2 * 67 * 50)\left((2 * 67 * 50) - 67 - 50\right)}{(67 + 50)^2(67 + 50 - 1)}}}}$$

$$\boldsymbol{z = 1.473}$$

Calculating the z value, the probability is obtained according to the information shown in Table 4.12, which results in 0.0708, which exceeds the 5% significance value, it is known that one of the two series is not homogeneous. In this sense, and when reviewing the information obtained with the other consistency tests, it is known that station 14076 did not pass the homogeneity test by Cramer (with 5% significance for one tail), so it is then considered that station 14076 would also be inhomogeneous by means of the Wald-Wolfowitz test and, thus, station 14087 would be homogeneous.

However, when using a significance level of 10%, it is required that the value of the probability obtained by means of the value of *z* is less than 0.10, which is true and, thus, both stations 14076 and 14087 are homogeneous by means of the Wald-Wolfowitz test for a significance level of 10%.

Independence

Anderson limits

It is said that the data in a sample are independent when the value of one of them does not affect the value of the next data in the same series. To determine the probability limits of independent series, the Anderson Limits test with 95% confidence is used (Anderson, 1941).

The independence of the series is determined by means of a graph named *correlogram.* This is constructed through the estimation of the confidence limits, called Anderson limits, hence the name of the test, and the determination of the autocorrelation coefficients. (r) which are plotted on the ordinates, while on the abscissa axis are plotted the time delays of the series or lags. (k). The number of lags depends on the number of data in the series, i.e., the greater the number of data, the greater the number of lags needed to evaluate the independence of the series, and can be calculated with Equation 4.10. Both elements, Anderson bounds and the autocorrelation coefficients (r_k), dependen de los tiempos de retraso de la serie (Salas et al., 1980).

$$\mathbf{k} = \frac{n}{3} \qquad (10)$$

To calculate the correlogram, it should be considered that from the original data series (X) a modified series is generated (Y) which depends on the time lag (k) that applies to the series. Thus, for the same k, you have a series of X and a series Y.

Then, according to the number of lags (k), will be the number of values you have in the correlogram (ρ), represented in Equation 11 and Equation 12; where σ_x y σ_y are the standard deviations of the series X y Y, respectively and n represents the number of data in the series.

$$r_k = \beta \frac{\sigma_x}{\sigma_y} \qquad (11)$$

Where:

$$\boldsymbol{\beta} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} \qquad (12)$$

To determine the probability limits of independent (or persistent for flow series), i.e. the upper and lower Anderson limits (Anderson, 1941) with a 95% confidence level, they are calculated using Equation 13.

$$l_{r(95\%)} = \frac{-1 \pm 1.96\sqrt{n-k-1}}{n-k} \qquad (13)$$

If, when applying the test with a significance level of 5%, it is found that the series is dependent (not independent), it is recommended to use a significance level of 1% (Equation 14).

$$l_{r(99\%)} = \frac{-1 \pm 2.326\sqrt{n-k-1}}{n-k} \qquad (14)$$

As for the meteorological stations, if and only if less than 10% of the values of the calculated correlogram exceed the confidence limits, the data series is said to be independent.

When this test is applied to the hydrometric stations, it does not seek independence of the series, but rather evaluates its dependence due to the nature of the information.

In the series of runoff volumes or flow rates there must be a dependence of the data (due to the regulation of the systems) that translates into a relationship of the data evaluated with that which precedes it. Such dependence increases as the sampling interval of a series is reduced, so that there is more dependence between successive monthly values than between annual magnitudes (Campos, 2007).

Therefore, the interpretation of the dependence of a flow series requires that at least 90% of the values of the correlogram are outside the Anderson Limits, then the series is said to be persistent (dependent).

For the application case, we will refer to the data in Table 3.5 and Table 3.6 for the annual precipitation series for stations 14076 and 14087 respectively, whose number of data in each series are 21 and 30. Using Equation 4.13 the Anderson Limits were calculated with 7 lags for station 14076 and 10 lags for station 14087.

Table 7.22 shows the value of the calculated autocorrelation coefficients, and the value of the Anderson Limits for station 14076.

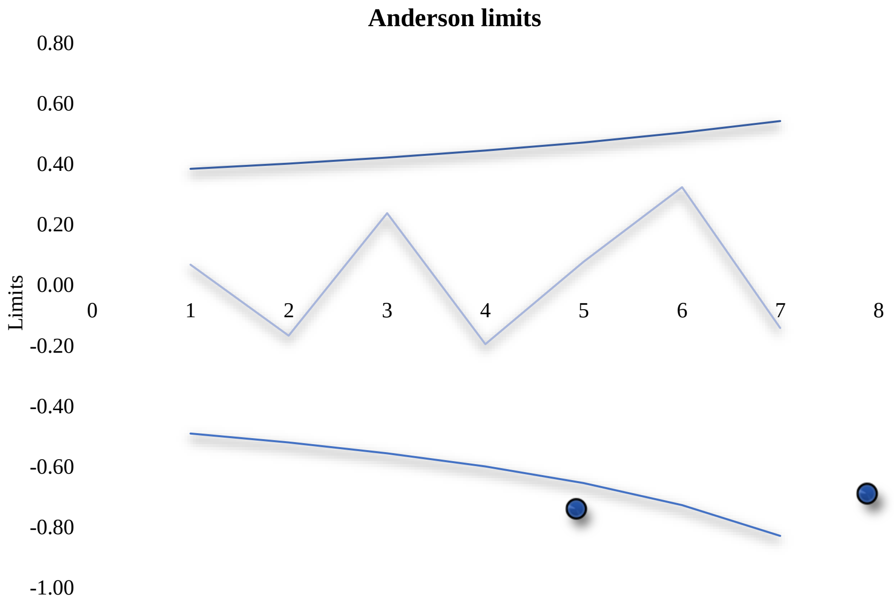
Table 7.22 Autocorrelation coefficients for station 14076

k	X	X^2	Y	XY	n
1	16951.8	14955257.4	17513.8	14884535.2	20
2	15801.3	13631607.2	16499.8	13633640.8	19
3	15062.7	13086077.2	15869.3	13398143.7	18
4	14411.7	12662276.2	15071.3	12684408.7	17
5	13426	11690671.7	14129.8	11892451.5	16
6	12477.7	10791398.9	13614.3	11449383.1	15
7	11669.2	10137726.6	12911.3	10711077	14
β	**σx**	**σy**	r_k	l_r **inf**	l_r **sup**
0.07	175.78	176.26	0.07	-0.49	0.39
-0.18	165.08	178.18	-0.17	-0.52	0.40
0.25	168.27	173.50	0.24	-0.56	0.42
-0.21	166.73	177.54	-0.19	-0.60	0.45
0.08	168.24	182.78	0.08	-0.65	0.47
0.30	171.52	159.68	0.32	-0.73	0.50
-0.12	177.87	154.95	-0.14	-0.83	0.54

Source: Own elaboration

Graphic 7.5 shows the correlogram for station 14076. In this plot it is observed that there are no values of the autocorrelation coefficients that are outside the Anderson Limits, so the series can be considered as independent.

Graphic 7.5 Correlogram and Anderson Limits for station 14076



Source: Own elaboration

Table 7.23 concentrates the results obtained for the autocorrelation coefficients and values of the Anderosn Limits for station 14087. Thirty years of data corresponding to the period (1973 - 2002) were analyzed.

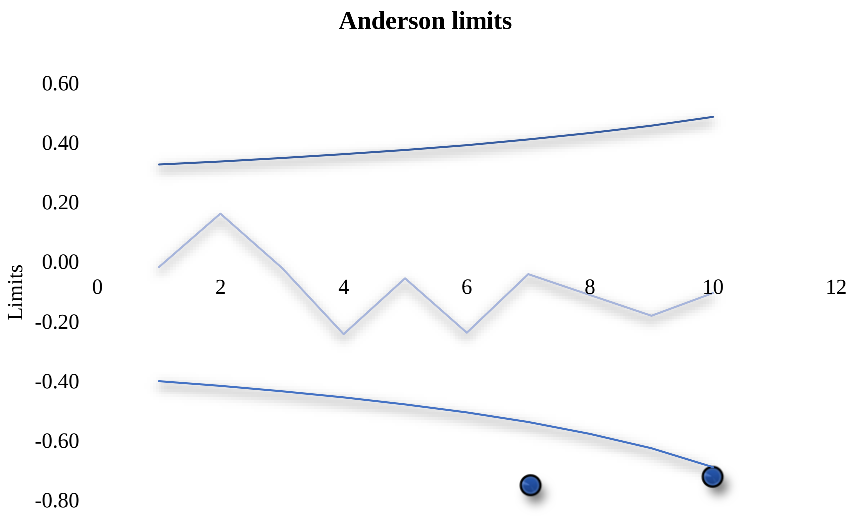
Figure 7.6 shows the resulting correlogram for station 14087. It is observed that all the autocorrelation coefficients are within the Anderson Limits, which means that the series can be considered as independent.

Table 7.23 Autocorrelation coefficients for station 14087

K	X	X^2	Y	XY	n
1	24434.3	21202984.2	24304.7	20468555.3	29
2	23919.4	20937862.2	23637.8	20286289.8	28
3	23251.4	20491638.2	23302.8	20057288.1	27
4	22551.4	20001638.2	22682.1	19550941.4	26
5	21640.6	19172081.6	22370.9	19337523.4	25
6	20771	18415877.4	21710.3	18663447.5	24
7	19941.8	17728304.8	21343.9	18485558.5	23
8	19013.8	16867120.8	20747.3	17874067.2	22
9	18268.3	16311350.5	20049.6	17347303.9	21
10	17629.7	15903540.6	19736.2	17350902.9	20
β	**σx**	**σy**	**r_k**	**l_r inf**	**l_r sup**
-0.02	148.27	143.33	-0.02	-0.40	0.33
0.19	136.68	155.07	0.16	-0.42	0.34
-0.02	134.22	152.06	-0.02	-0.43	0.35
-0.28	132.88	152.81	-0.24	-0.45	0.36
-0.06	135.32	156.74	-0.05	-0.48	0.38
-0.29	138.23	167.01	-0.24	-0.50	0.39
-0.05	141.11	165.15	-0.04	-0.54	0.41
-0.13	143.79	172.45	-0.11	-0.58	0.43
-0.22	144.81	180.82	-0.18	-0.63	0.46
-0.13	138.26	170.63	-0.10	-0.69	0.49

Source: Own elaboration

Graphic 7.6 Correlogram and Anderson Limits for station 14087



Source: Own elaboration

Summary of results. Validation of weather stations

Table 7.24 shows the summary results of the consistency tests applied to meteorological stations 14076 and 14087. With the exception of the Cramer's test, all the others allowed homogeneity to be demonstrated for all cases. The non-homogeneity of Cramer's test, considered as specific, was verified again through the Wald-Wolfowitz test, since this is a complementary test for any of the statistical homogeneity tests shown in this work, it was determined that the series can be considered as homogeneous. In addition, the rest of the tests applied showed that the data can be considered homogeneous, which is an indication that gives robustness to the results presented.

Table 7.24 Consistency test results for weather stations 14076 and 14087

Test	**Station**	
	14076	**14087**
Sequences	Homogeneous	Homogeneous
Helmert	Homogeneous	Homogeneous
Double Mass Curve	Homogeneous	Homogeneous
t de Student	Homogeneous *	Homogeneous *
Cramer	Non-homogeneous **	Homogeneous **
Wald - Wolfowitz	Homogeneous ***	Homogeneous ***
Anderson limits	Independent *****	Independent
* 5% significance level for one-tailed. ** Two-tailed significance level of 5%. *** 10% significance level **** Significance level of 5%.		

Source: Own elaboration

The application of at least two homogeneity tests is recommended in all cases. Generally, it is recommended to start with the general tests, and if there is a discrepancy of results, then the application of at least a third specific test that can help in the validation of the homogeneity of the series is recommended.

The test used to verify the independence of the series, Anderson Limits, is a very robust test, in this case the series analyzed were found to be independent.

7.5 Validation of hydrometric stations

7.5.1 Natural regime stations

The natural regime hydrometric series is given by the historical series of flows that would have flowed through that place if there were no human intervention in the basin. Anthropogenic actions are all works of regulation or use of surface or groundwater that alter the amount of flow that would have flowed through the river (Solera, 2003).

Therefore, if a mathematical model is configured to simulate its operation and is fed with the natural regime series, the result will be the series also simulated in natural regime.

In Mexico, the necessary information is not available to carry out a restitution to a natural regime; for this reason, it is proposed to work with undisturbed hydrometric stations, with which high or headwater basins are created. These basins are areas adjacent to the water divide or watershed in the highest altimetric portion of the basin and, in this zone, the first runoff is formed after the soil has retained or absorbed the water according to its capacity (Cotler et al., 2013).

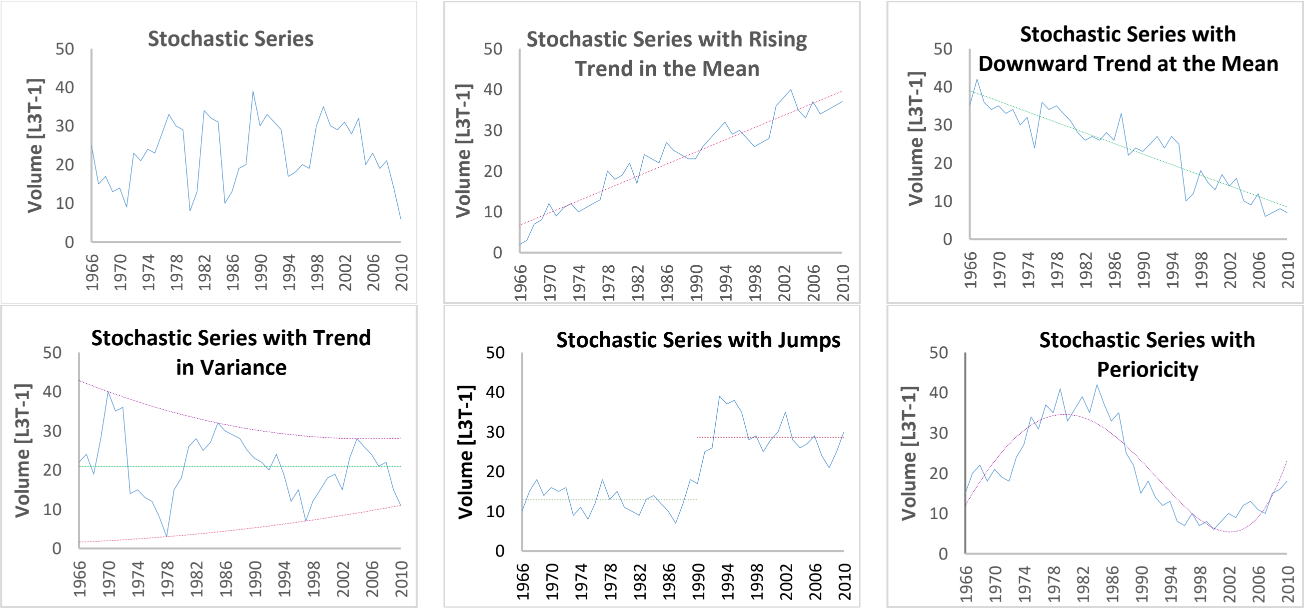
When starting from hydrometric series in which there are no anthropogenic actions, it is necessary to check that there are no important works in the basin; in addition to the behavior of the hydrometric series itself. Therefore, in addition to homogeneity and persistence tests, a visual review of the flow series over time is performed; thus, for the basins generated from the hydrometric stations, the runoff coefficient is obtained, and the relative modulus is calculated.

Visual Review of Runoff Series

As mentioned by Campos Aranda in his book: Estimación y Aprovechamiento del Escurrimiento (2007), hydrology defines a chronological series or time series as a succession of observations that measure the variation over time of some aspect of a phenomenon, such as the flow or volume of a watercourse, the water level in a lake or reservoir, etc. In hydrology only two components are accepted: deterministic and random or stochastic.

The deterministic component is that which can be evaluated for prediction purposes and consists mainly of trend-like behavior and cyclic or periodic form, as well as sudden changes, called jumps, which are inhomogeneities of a particular type. On the other hand, the stochastic component consists of irregular oscillations and random effects that cannot be strictly explained physically and require probabilistic concepts for their description. Graphic 7.6 schematizes time series with various types of deterministic components.

Graphic 7.6 Time series with various types of deterministic components

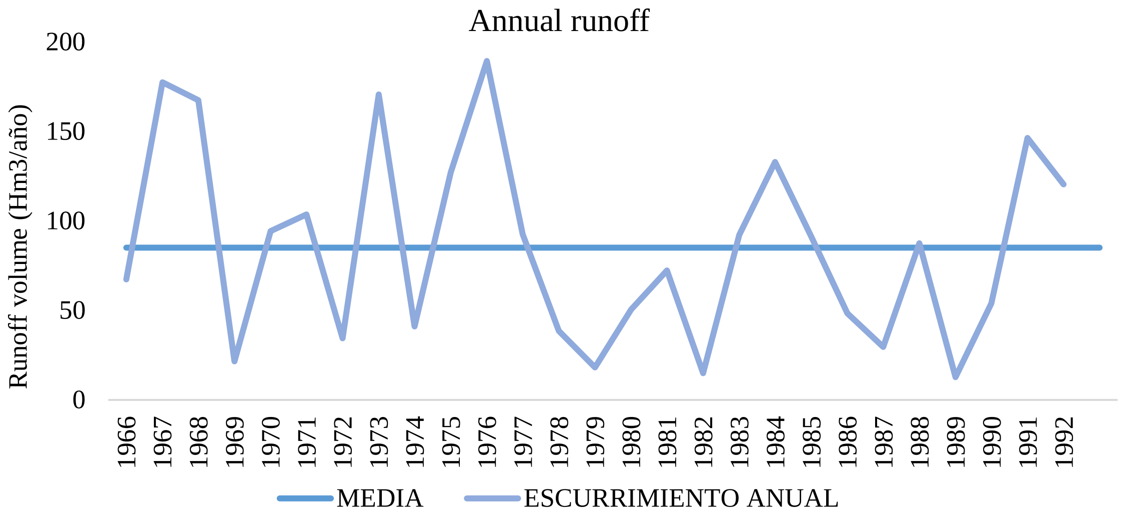


Source: Own elaboration; adapted from Campos (2007).

In general, trends in annual runoff time series can result from changes in the hydrologic environment that produces the series or from alterations that come from natural or human-induced gradual variations. Whether the trend in the time series is due to changes in the watershed or to errors in measurement, the fact is that it causes the series to be labeled as inconsistent. On some occasions, the trend in the mean may be quite obvious, however, in most cases there is some doubt as to whether the suspected systematic error effects are significant or not, which is why numerical tests such as persistence are imperative.

Apart from the trend, sudden changes called jumps can occur, which can be the result of catastrophic natural events such as earthquakes or forest fires, or consequences of hydraulic works built in the basin. In general, the presence of a jump in the series indicates that somehow the homogeneity of the record has been lost, i.e., now the observations that integrate it come from two populations, perhaps statistically different, and therefore, it will be necessary to test whether or not the homogeneity was lost. However, as shown in Graphic 7.7, the series does not show marked jumps over time; in the period from 1977 to 1983, there is a very slight decrease in runoff volume. However, despite the fact that we have old information, it is worth mentioning that there are no control works in the basin, so the stream bed flows without alterations, which indicates that the basin is in a natural regime. In addition, the basin does not receive surface contributions from other basins, which means that the basin under study is a headwater basin.

Graphic 7.7 Annual runoff volumes for hydrometric station 12607 "La Yerbabuena"



Source: Own elaboration

However, comparing the series shown in Figure 7.7 with the graphs shown in Figure 7.6 the annual volumes of the hydrometric station are representative of a normal stochastic series, with no tendency to rise or fall according to the mean and without significant jumps or periodicity, which indicates that the flow was maintained over time and that there is no significant anthropogenic alteration that could alter the results of future modeling, so it is concluded as part of the visual review that the hydrometric station is suitable for hydrological modeling. Antes de obtener los parámetros que permiten entender la estación In order to determine the hydrometric characteristics with respect to the basin under study (runoff coefficient and relative modulus), the homogeneity (Sequences and Helmert) and persistence (Anderson Limits) tests described above must be performed.

With this, the Helmert and Sequences test procedure (Table 7.25) is shown using the series presented in Table 7.7.

Table 7.25 Sequence and Helmert test for station 14076

Year	PMA	Sequence Testing		Helmert test	
		Comparison	Sequence	COMPARISON	CHANGES
1966	**67.40**	L	1	L	
1967	**177.48**	M	2	M	C
1968	**167.50**	M	2	M	S
1969	**21.59**	L	3	L	C
1970	**94.29**	M	4	M	C
1971	**103.60**	M	4	M	S
1972	**34.39**	L	5	L	C
1973	**170.79**	M	6	M	C
1974	**41.07**	L	7	L	C
1975	**127.08**	M	8	M	C
1976	**189.41**	M	8	M	S
1977	**92.64**	M	8	M	S
1978	**38.57**	L	9	L	C
1979	**18.15**	L	9	L	S
1980	**50.48**	L	9	L	S
1981	**72.42**	L	9	L	S
1982	**15.02**	L	9	L	S
1983	**92.17**	M	10	M	C
1984	**133.05**	M	10	M	S
1985	**91.37**	M	10	M	S
1986	**48.36**	L	11	L	C
1987	**29.55**	L	11	L	S
1988	**87.55**	L	11	M	C
1989	**12.73**	L	11	L	C
1990	**54.02**	L	11	L	S
1991	**146.46**	M	12	M	C
1992	**120.55**	M	12	M	S

Source: Own elaboration

Since the continuous series consists of 27 years (Table 7.24), the comparison value for the sequence test will be the median with a value of 87.55.Hm^3. With this, a total of 12 sequences is obtained and when reviewing Table 4.2 and taking the value for 26 data, it is found that the number of sequences allowed is from 10 to 17 sequences), so, according to the sequences test, it is found that the series is homogeneous.

Similarly, for the Helmert test, the value of the mean is 85.10. Hm^3, this gives a total of 13 changes and 13 sequences, so that when Equation 4.1 is applied, it is shown that the station is homogeneous, with a difference between sequences and changes of zero, as shown below:

$$\boldsymbol{-\sqrt{27-1} \leq 13-13 \leq \sqrt{27-1}}$$

$$\boldsymbol{-5.10 \leq 0 \leq 5.10}$$

It is important to mention that the application of this test is only an indicator of what happens in the hydrometric station and serves to understand the behavior of the series in question; however, it is not a limiting factor for the use of the station in question.

On the other hand, since it is a persistence test, the correlogram should be outside the Anderson limits, which indicates the non-independence of the series.

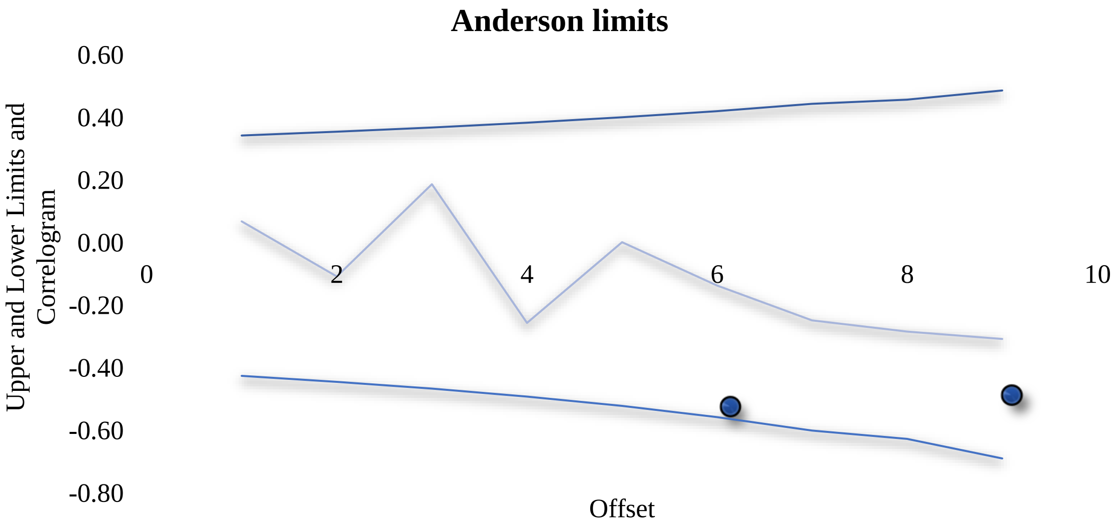
Table 7.26 shows the correlogram results and the limits for the 9 lags generated according to the annual series (Table 7.7).

Table 7.26 Result of the Anderson limit test for hydrometric station 12607

K	X	X²	Y	XY	n
1	2177.14801	256623.771	2230.30216	191939.931	26
2	2030.69198	235174.402	2023.63976	157164.842	25
3	1976.67025	232256.055	1801.05438	159727.409	24
4	1963.9355	232093.881	1816.80983	140810.03	23
5	1876.38378	224428.577	1801.14906	153811.856	22
6	1846.83041	223555.175	1701.36788	142613.568	21
7	1798.46954	221216.402	1728.7964	143600.841	20
8	1707.1015	212868.283	3399.48329	163028.703	20
9	1574.05427	195166.718	3267.04777	143198.67	19
β	**σx**	**σy**	**r_k**	**l_r inf**	**l_r sup**
0.07	54.52	54.88	0.07	-0.42	0.34
-0.10	54.09	52.09	-0.11	-0.44	0.36
0.16	54.95	47.83	0.19	-0.47	0.37
-0.22	54.10	47.30	-0.25	-0.49	0.39
0.00	55.37	46.07	0.00	-0.52	0.40
-0.11	55.29	46.98	-0.13	-0.56	0.42
-0.20	55.96	45.21	-0.25	-0.60	0.45
-1.89	57.49	385.38	-0.28	-0.63	0.46
-1.97	58.17	373.86	-0.31	-0.69	0.49

Source: Own elaboration

Graphic 7.7 Correlogram and Anderson Limits for hydrometric station 12607



Source: Own elaboration

On the other hand, Figure 5.3 shows the correlogram with the Anderson limits, which is within the Anderson limits; this shows that the station is not persistent (since it is within the limits and not outside), this indicates that greater care should be taken with respect to the simulations; however, this is not an impediment to continue using the information provided by the hydrometric station.

Once the consistency tests are performed at the hydrometric station, the runoff coefficient and relative modulus values are obtained, which require information not only from the station itself, but also from meteorological information and from the basin under study.

Runoff Coefficient

Chow et al. (1994) defined the runoff coefficient as the ratio of direct runoff to the average precipitation intensity of a storm. However, because of the variability of precipitation intensity, this value is difficult to determine using observed information, so it can also be defined as the ratio of the volume of direct runoff to the volume of precipitation in the basin, in a given time period, such that Equation 15 is obtained:

$$\boldsymbol{Ce} = \frac{\boldsymbol{V_E}}{\boldsymbol{V_P}} \tag{15}$$

Where V_E is the annual volume and V_P is the annual volume precipitated in the area?

The runoff coefficient is an imprecise variable, because it implies a fixed relationship between runoff and rainfall in the basin, which is not actually true. The proportion of total rainfall that will flow as surface runoff depends on the permeability of the soil and the slope of the area.

Another way to understand it is to obtain the parameter K with the relationship between the runoff sheet in mm (E) and the precipitation in mm (P), as shown in Equation 16.

$$\boldsymbol{K} = \frac{\boldsymbol{E}}{\boldsymbol{P}} \tag{16}$$

For both Ce and K, it must be fulfilled that: K<1 and Ce<1; this is due to the fact that the runoff of a basin must be less than the present precipitation.

For the case of application, we have the information of the annual flow rates presented in Table 7.7, which shows that the average annual flow rate at the hydrometric station is 85.10 Hm^3; the value of the surface area of the watershed is also recalled, which is 299.20 km^2.

The precipitation in the basin is of 903.23 mm, which is obtained by means of a fictitious station at the center of the basin generated by hydrometric station 12607, according to the monthly information obtained from meteorological stations 14076 and 14087.

With the values described in the previous paragraphs, Equation 15 and Equation 16 are developed, which must be identical, since the volume of annual contribution (V_E) and the annual precipitated volume (PV) are the sheets are the runoff sheet and precipitation multiplied by the basin area, such that:

$$\boldsymbol{V_E = 85.10\ Hm^3}$$

$$\boldsymbol{V_P = (903.23\ mm) * (299.20\ km^2)} \quad \rightarrow \quad \boldsymbol{V_P = 270.26\ Hm^3}$$

$$\boldsymbol{Ce = \frac{85.10\ Hm^3}{270.26\ Hm^3}} \quad \rightarrow \quad \boldsymbol{Ce = 0.3149}$$

Likewise, the same is done for the runoff sheets (284.42 mm) and that of precipitation, so that the value of K is obtained.

$$\boldsymbol{K = \frac{284.42\ mm}{903.23\ mm}} \quad \rightarrow \quad \boldsymbol{K = 0.3149}$$

It is interesting to note from the above formulation that the runoff coefficient can be obtained as follows (Ce) or the parameter K in either of the two ways. The most important thing is to verify that the runoff volume is less than the precipitated volume. If this is not the case, it is clear that there is a problem in the basin, since it is very likely that it is in an altered regime and has a contribution from an external source.

Modes of Contributions

Sánchez (2017) mentions that different modes can be presented in the gauging data, such as: daily, monthly or annual flows, contribution, equivalent water sheet and specific flow (Figure 5.4). These last two allow relating runoff and precipitation to the study area and that is why, it is used as a parameter for understanding the hydrometric stations under review.

Daily flow rates: which may correspond to the daily reading of a limnimetric scale or correspond to the mean ordinate of the daily graph of a limnigraph.

Monthly or average monthly flows: for a given year, it is the average of all the days of that month. For a series of years, it refers to the average of all October values, all November values, etc. for the entire series studied.

For a given year, the annual or mean annual flow (modulus) is the average of all the days of that year, for the series of years it refers to the average of all the years of the series under consideration.

The contribution is normally referred to a year (annual contribution), although it is sometimes referred to a month (monthly contribution). It is the volume of water contributed by the watercourse at the point considered during a year or a month (Hm^3).

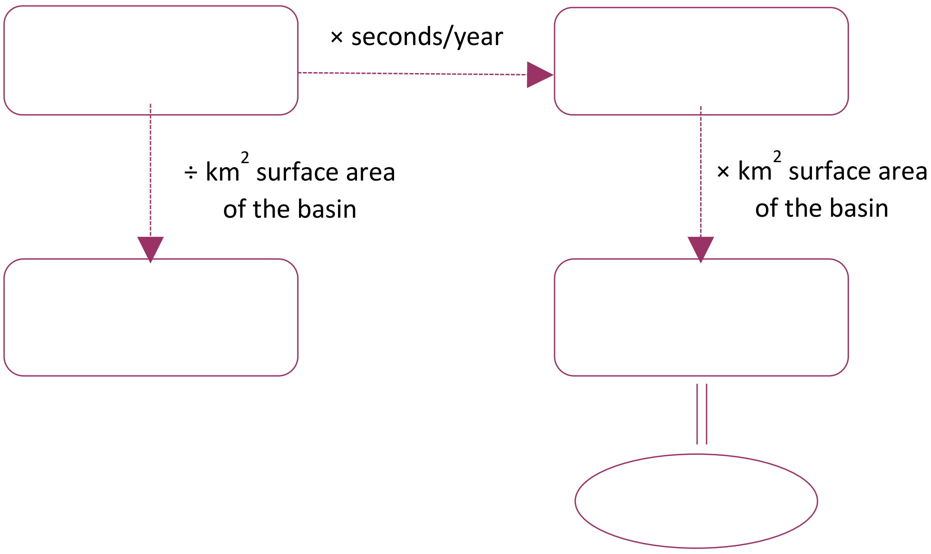
The equivalent sheet of water is the thickness of water that would be obtained by distributing over the entire basin the volume of the annual inflow (en mm). It is obtained by dividing the annual contribution by the surface area of the basin. It is useful especially when we want to compare runoff with precipitation. If the basin is hydrologically closed and the data come from more than 20 years, this value should be similar to the non-evapotranspired precipitation. (P-ETR). The specific flow rate is the flow rate per unit area. It represents the flow rate provided by each km^2 basin. It is calculated by dividing the flow (normally annual average flow by the surface area of the basin or sub-basin considered (liters /seg·km^2); this parameter is also known as relative modulus and as shown in Equation 17 is obtained with the modulus or flow rate (M) in liters per second and with the basin area (S) in km^2.

$$M_r = \frac{M}{S} \tag{17}$$

The relative module allows comparing the flow of various basins, being their surfaces different (Sanchez, 2017). Mountain areas provide more than 20 liters/sec-km^2, while, in the lower parts of the same basin only 4 or 5 liters/sec-km2 are generated. Likewise, we know that:

- If the relative modulus is less than 5 liters/sec-km^2, there is a water shortage.
- If the relative modulus is in the range of 5 to 15 liters/sec-km^2, it is said to be in average values and if the relative modulus is greater than 15 liters/sec-km^2, the values are high and it is possible that a parameter is being poorly considered.

Figure 7.6 Report on parameters obtained from gauging in the basins



Source: own elaboration; adapted from Sánchez (2017)

Now, the relative modulus of the basin generated from hydrometric station 12607 is obtained, for this purpose, it is convenient to obtain the monthly flow information, so Table 3.8 shows the monthly series of the hydrometric station under study, with which an average monthly value of 7.24 is obtained Hm^3.

Thus, Equation 17 is developed, considering that special care must be taken with the change of units; in such a way that it is obtained:

$$M_r = \frac{7.24\,\frac{Hm^3}{mes}}{299.20\ km^2}$$

$$M_r = 9.34\,\frac{l}{s \bullet km^2}$$

Thus, the relative modulus is within the range of average values, which does not indicate the scarcity of water or the contribution to the basin from any other external source, which shows that the hydrometric station has not been altered and can be used without problem to be modelled.

Summary of results. Validation of hydrometric stations

Table 7.27 shows the summary results of the consistency tests and basin parameters with respect to the hydrometric series applied to hydrometric station 12607. As can be seen, the station is not persistent, however, it is convenient to remember that this is not an impediment to continue using the station. With respect to the values obtained for the runoff coefficient (parameter K) and the relative modulus, they are within the recommended ranges, since $Ce < 1$ y $5 < M_r < 15$.

Table 7.27 Consistency test results for weather stations 12607

Test	Meteorological Station 12607
Sequences	Homogeneous
Helmert	Homogeneous
Anderson limits	Non Persistent
Ce & K	0.3149
$M_r\ (l/s \bullet km^2)$	9.34

Source: own elaboration

7.6 Conclusions

The development of this work allowed us to generate a methodology for the treatment of precipitation series and average flows from meteorological and hydrometric stations, as well as to provide reliability to the data so that they can provide certainty in their application for modeling purposes or other objectives.

This work allowed the analysis of a set of tests to evaluate the consistency of precipitation and mean flow series. Although the suggested tests are already existing methods, this work shows a clear procedure for the evaluation of the properties that characterize each type of series, such as homogeneity and independence.

The methodology provides a set of established criteria that allow the selection of the best meteorological and hydrometric stations. The current literature does not reflect criteria that show an established procedure to make such selection, therefore, in this work a complete process has been presented to execute this selection of stations considering specific aspects of each series and compare it with the rest; in this way and through the weighting of such criteria, it was possible to assign a numerical value to each station, compare it with other stations and determine which of the available stations are the best.

Within the framework of the methodology, the procedure for generating monthly and annual precipitation data when there are gaps was also proposed, and the corresponding procedure for generating annual average flow data when there are gaps was also proposed. Criteria were established to consider a data as null, a valuable criterion as a starting point in the treatment of data with an understanding of risk.

There is a complementary contribution to the methodology, and that is the deduction of missing data necessary to complete some gaps identified in the generation. Although data deduction methodologies exist, this work proposes criteria, and based on this proposal, it is possible to be risky or conservative in the deduction of missing data.

7.7 Acknowledgments

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) for the economic stimulus to the National System of Researchers research granted to the first author and the current financing of the doctoral scholarship granted to the second author. As well as for the concluded financing of the doctoral scholarship granted to the third author.

To the Universidad Michoacana de San Nicolás de Hidalgo, Coordination of Scientific Research for the support of research project 2021.

7.8 References

Anderson, R. (1941). Distribution of the Serial Correlation Coefficient. The Annals of Mathematical Statistics.

Aparicio Mijares, Francisco J. (1989). Fundamentos de Hidrología Superficial. Ed. Limusa. México D.F.

Campos Aranda, D. F. (1998). Procesos del Ciclo Hidrológico. México: Editorial Universitaria Potosina.

Campos Aranda, D. F. (2007). Estimación y Aprovechamiento del Escurrimiento. México: UNAM.

Cao, L. J., & Yan, Z. W. (2012). Progress in research on homogenization of climate data. Advances in Climate Change Research, 3, 59-67.

Costa, A. C., & Soares, A. (2006). Identification in inhomogeneities in precipitation time series using SUR models and the Ellipse test (pp. 419-428). In: Proceedings of Accuracy 2006. Caetano, M., & Painho, M. (eds.). 7th International Syposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Instituto Geográfico Portugués. Lisbon.

Chow, V., Maidment, D. R., y Mays, L. (1994). Hidrología aplicada. Colombia: McGraw Hill.

CICESE. (2018, febrero). Base de Datos Climatológica Nacional. CLICOM. http://clicom-mex.cicese.mx/mapa.html

CONAGUA. (2016, 31 julio). Banco Nacional de Datos de Aguas Superficiales. BANDAS. https://app.conagua.gob.mx/bandas/

Cotler Ávalos, H., Galindo Alcántar, A., González Mora, I., Pineda López, R., y Ríos Patrón, E. (2013). Cuencas hidrográficas. Fundamentos para su manejo y gestión. México: Secretaría de Medio Ambiente y Recursos Naturales.

Doorembos, J. (1976). Agro-meteotrological field stations: Vol. Irrigation and Drainage Paper No. 27. Food And Agiculture Organization of the United Nations.

Gobierno del Estado de Jalisco. (2019, 5 enero). Municipios de Jalisco. Jalisco. https://www.jalisco.gob.mx/es/jalisco/municipios

Guajardo-Panes, R. A., Granados-Ramírez, G. R., Sánchez-Cohen, I., Díaz-Padilla, G., & Barbosa-Moreno, F. (2017). Validación espacial de datos climatológicos y pruebas de homogeneidad: Caso Veracruz, México. Tecnologia y Ciencias Del Agua, 8(5), 157–177. https://doi.org/10.24850/j-tyca-2017-05-11

IIEG. Instituto de Información Estadística y Geográfica de Jalisco (IIEG) (2020). Principales Ríos y Cuerpos de Agua en el Estado de Jalisco. Instituto de Información Estadística y Geográfica de Jalisco. https://iieg.gob.mx/ns/wp-content/uploads/2020/04/RIOS.pdf

Lugo Arias, F. T. (2014). Agua y Reservas Hidrológicas. Jalisco, México: Programa Sectorial.

Martínez P.E., Martínez P., & Castaño S. (2006). Fundamentos de Hidrogeología. Ed. Mundi-Prensa, México D.F.

Mather, J.R. 1975: Estimation of areal average precipitation using different network densities and averaging techniques. Publication in Climatology, vol. XXVIII, No. 2, C.W. Thornthwaite Associates, Laboratory of Climatology, Elmer, New Jersey

Mejía, J., Vega Méndez, F., Man Eddin Abou Asaad, S., & Mora, E. (2019). Hidrogeografía de una cuenca de usos múltiples ubicada en la cordillera de la Costa Venezuela. Cuadernos de Geografía: Revista Colombiana de Geografía, 30(1), 217-238.

Merlos Villegas, F., Sánchez Quispe, S. T., y Almanza Campos, J. A. (2014). Creación de un Sistema de Información Hidrológico para el Cálculo de Intensidades Máximas y Gestión de Datos Meteorológicos. Jalisco, México: XXIII Congreso Nacional de Hidráulica.

Salas, J. D., Delleur, J. W., Yevjevich, V., y Lane, W. L. (1980). Applied Modeling of Hydrologic Time Series. Colorado: Water Resources Publications.

Sánchez San Román, F. (2017). Hidrología Superficial y Subterránea. España: Createspace Independent Pub.

Salvador-González, M. C., Canul-Reich, J., & Corona-Ferreira, A. (2018). Análisis del Dataset de CLICOM de precipitaciones del Estado de Tabasco, México. IV Conferencia Internacional en Ciencias Computacionales e Informática.

Sandoval Erazo, W., y Aguilera Ortiz, E. (2014). Determinación de Caudales en cuencas con poca información Hidrológica. Ciencia UNEMI, 100 - 110.

Siegel, S. & TRILLAS, EDITORIAL. (2015). Estadistica No Parametrica (4.a ed.). Trillas.

Solera Solera, A., y Andreu Álvarez, J. (2003). Herramientas y métodos para la ayuda a la decisión en la gestión sistemática de recursos hídricos. Aplicación a las cuencas de los ríos Tajo y Júcar. Valencia, España: Universidad Politécnica de Valencia.

Walker, S. (2000). The value of hydrometric information in water resources management and flood control. Meteorological Applications, 7(4), 387–397. https://doi.org/10.1017/S1350482700001626

World Meteorological Organization (WMO). (1966). Climate change (N.o 79). Secretariat of the WMO.