Statistical descriptive analysis of three climate variables; Precipitation, temperature and relative humidity. Study cases (Innaouene watershed; Morocco)

Zineb Moumen[1] *, Ismail* Elhassnaoui[2], and *Fatima* Daid[1]

[1]Laboratory of Functional Ecology and Environment, Faculty of Sciences and Technologies, University of Sidi Mohammed Ben Abdellah, Fès B.P. 2202 – Route d'Imouzzer, Fes, Morocco.
[2]Mohammed V University, Mohammadia School of Engineers, Rabat, Morocco

Abstract. Hydrological data are mandatory for the elaboration of studies related to water management. Statistical analysis is a crucial step for the grasping of the distribution of data range. The presented study consists on applying statistical descriptive analysis on three climate variable; precipitation, temperature and relative humidity. Two gages stations were used; Bab Marzouka and Idriss First in Innaouène watershed. The obtained results underlines that the temporal variation exposes an overall rising trend in the temperature and a decreasing trend in the rainfall and relative humidity over the fours studied decades, it also uncovers the intra-seasonal fluctuation, humid and rainy in the winter and sec and dry in the summer, the autumn and the spring are considered as transition season where the temperature is moderated. The spatial variation is marked by a slight decreasing in precipitation and increasing of temperature moving from the middle part of the watershed to the downstream, which could be explained by the topographic variation and its impact on the climate. High altitude are generally marked by high precipitation and lower temperature comparing to lowlands areas.

1 Introduction

Hydrological data are crucial for carrying out studies related to water resources issues, for example to assess the level of water scarcity on crop yields, precipitation and evapotranspiration are required.
These data could be extracted from several sources (eg observed climate data recorded by weather stations, or open access data gateways such as global weather data, IPCC, GMIP). [1]. The climatic variables vary in space (latitude, longitude, altitude) and in time scale (hours, day, month, etc.), the acquisition of a set of data in a defined area during a determined period defined by the spatio-temporal resolution.
The statistic descriptive is usually used to describe the form of the distribution of the data range with some common parameters such as central tendency, skewness, and taildness. [2]
The central tendency is measured using three parameters, the mean, the mode, and the median, this later correspond to the value of the studied variable at which the values on both sides are equal, considering a data series with n observation arranged in ascending order the median is the observation that half the values on two, the dispersion is then concluded by measuring how closely the random variables are distributed, clustered or spread out around the central value.

1.1 Data Symmetry

1.1.1 Skewness Coefficient

The symmetry of data describes the data distribution in respect to the mean, the data may be tailed to the left or the right, differently said skewed to the left or the right, depending on if the skewness coefficient is positive or negative, it is expressed by the following equation:

$$Cs = \frac{n \sum_{i=1}^{n}(x_i - \overline{x})^3}{(n-1)(n-2)S^3} \quad (1)$$

-The skewed right data or positive skewness has Cs>0
-The skewed left data or the negative skewness has Cs<0

1.1.2. Tailedness coefficient

Concerning the tailedness (**Fig**.1), this parameter is referred to the coefficient of Kurtosis value, this later points out the effect of existing outliers and measure the tailedness of distribution, it is expressed by the following equation

$$K = \frac{n \sum_{i=1}^{n}(x_{i-}\overline{x})^4}{(n-1)(n-2)(n-3)S^4} \quad (2)$$

Based on this measure the distribution could be:
-Mesokurtic: k=3 the distribution is referred to as normal.
-Leptokurtic: k>3 the distribution is more concentrated near the mean.
-Platykurtic: k<3, a smaller concentration of probability near the mean.

1.1.3. The Kurtosis test.

Kurtosis illustrates how the tails of distribution differ from the normal distribution. In general, normally distributed data establishes the reference point for kurtosis. The kurtosis coefficient that varies from 0 may indicate that the data are not normally distributed.
-Positive kurtosis: A distribution with a positive kurtosis value implies that the distribution has heavier tails than the normal distribution.
Negative kurtosis: A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.
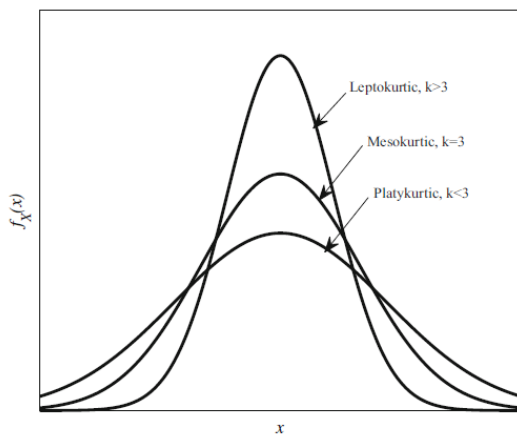

Fig.1. Plot of Frequency distribution in function of a random variable that shows the difference between Mesokurtic, Leptokurtic, Plakurtic Zone.
Source: MAITY, R.,. Springer Transactions in Civil and Environmental Engineering Statistical Methods in Hydrology and Hydroclimatology [en línea]. S.l.: s.n. ISBN 9789811087783. Disponible en:
http://www.springer.com/series/13593.

In this study case, the statistic descriptive were carried out under Minitab, using two Hydrometric station of Innaouene watershed (Bab Marzouka and Idriss First Stations). Central tendency parameters were defined and described for the three climatic variables (Precipitation, temperature and relative humidity) to define the variation and fluctuation trends.

2 Study Area

The Innaouene watershed is located between the cities of Fez and Taza (region known geographically as the Fez-Taza corridor). It covers an area of 3640 km² and a perimeter of 268 km. It joins the Idriss Fisrt dam 20 km northeast of Fez. It is limited to the east by the

middle Moulouya watershed to the north-west by the Ouergha and to the south-west by the upper Sebou. (**Fig**.2)

The basin is positioned in a mountainous area (between Prerif and Moyen Atlas) with a maximum altitude of 1970 m at the northeast and southeast and 57 m towards its outlet to the west. It disposes a Mediterranean climate, marked by strong seasonal contrasts and rainfall irregularities.

3 Statistic descriptive

3.1. Data and methodology

The methodology of this study consists on analysing the central tendency of three climate variable; Precipitation, temperature and relative humidity. The first variable was recorded from observed gauge stations (Bab Marzouka and Idriss First) provided by the watershed hydraulic agency of Sebou, the two second variables are generated from The Climate Forecast System Reanalysis database CFSR. Since the gauge station did not dispose of a temperature and relative humidity data.
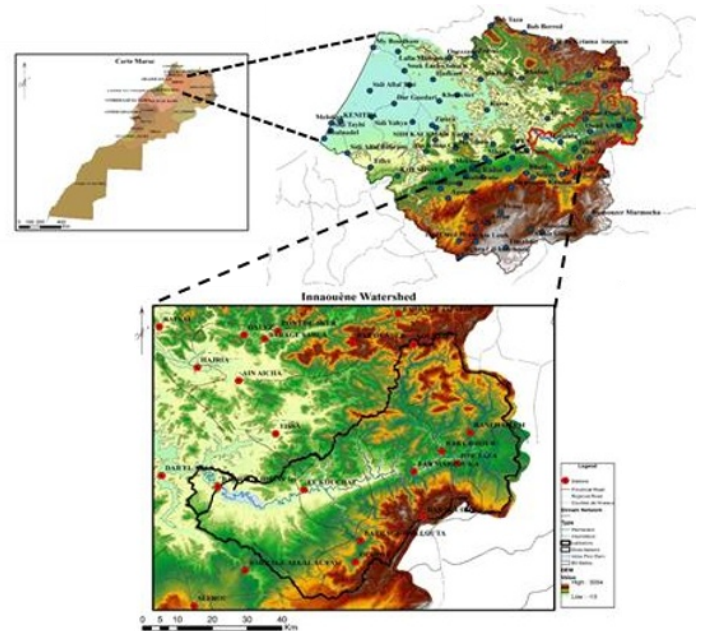

Fig. 2. Study Area

The CFSR platform disposes of a range of data series over 36 year (from 1979 to 1936) with a spatial resolution of 38 Km. It was successfully applied for hydrological modelling. The studies carried out by this database applied to the Moroccan watersheds show satisfactory results [3], [4] on the Beht basin the results were mediocre [5]. To process the Thiessen polygon was applied to the observed gauge station in order to define the influence area of every station, the mean of the nearest surrounding three stations was used for the descriptive statistical analysis.

A measure of central parameter tendency was realised over four different decade (1998-2010).

4 Results and discussion

4.1. Kurtosis and Skewness Test

Kurtosis illustrates how the tails of distribution differ from the normal distribution. In general, normally distributed data establishes the reference point for kurtosis. The kurtosis coefficient that varies from 0 may indicate that the data are not normally distributed. (Fig.3. & Tab.1.)
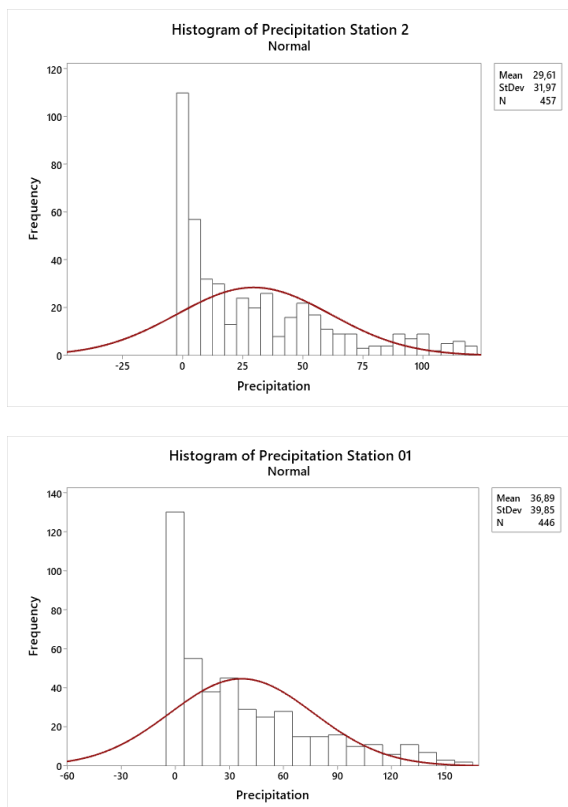

Fig.3. Frequency plot of a data set of Bab Marzouka station and Idriss First Station for the period ranging from 1979 to 2014.

Tab.1. Skewness and kurtosis coefficient for the two stations Bab Marzouka and Idriss First.

Variable	Decade	Skewness	Kurtosis
Precipitation Station 01	1980	0,95	-0,11
	1990	1,22	0,69
	2000	1,19	0,65
	2010	1,02	-0,00
Precipitation Station 01	1980	0,98	-0,17
	1990	1,13	0,44
	2000	1,23	0,75
	2010	1,12	0,49

4.2. Central tendency Analysis

Starting by relative humidity, for the station 01 Bab Marzouka the mean value is in the order of 0.55% in autumn, 0.7 in winter, and 0.59 in spring and around 0.3 in summer, in decade 2010 a slight increase is marked by 0.1 % for spring, the minimum value is recorded in summer and the maximum in winter. The standard deviation benchmark how data are spread to their mean value, the high the value is the bigger the data span, for the relative humidity this value is in the order of 0.006 more or less 0.04 for the all season, with a slight rise for the last two decades. The variance corresponds to the standard deviation squared, a low variance traduce a high precision and a low marge of defects, its value range in the round of the 0.01, the maximum value corresponds to the fall of the decade 2010.

The median is barely similar to the mean for example for the fall of the 1980 decade it equals 0.53 when the mean takes a value of 0.54, which implies that the data distribution is symmetric.

The temperature has mean values closes to the median with a slight difference of plus or − 2 °C, contrarily to the relative humidity the temperature data has shown a high variance for all the recorded decade, the interquartile interval is high and the max is recorded in the summer with a value of around 28°C. The large variance could be explained by the error linked to the data generated while the high interquartile is due to the nature of the climate and diurnal cycle.

The precipitation records different value during the four seasons, the maximum value is recorded in the winter, for instance for the decade of 1980 the mean precipitation value during the winter has reached 72.1 mm, and 5.68 in the summer, moving through the year this value increase remarkably, in decade 2000 no precipitation was recorded in the summer, for all the decades the median differ to the mean, In 1980 the mean value of the winter is 73.1mm, and it is greater than the median that takes a value of 49 mm, that implies that the data are asymmetric and 50 % of the values are positioned under the mean, from where the presence of extreme value and outliers. The interquartile interval is high, especially in the winter season.

Comparing to the first station the second one is characterized by a slight increase in the temperature recorded value, for instance for the decade 1980 the average temperature for fall, spring summer winter are successively 19.6°C,16.26°C, 28.473°C, 9.64°C; versus 18.96°C, 14.85°C, 26.66°C, 9.19°C.

The relative humidity shows a slight variance for the four decade during the four seasons, the lowest value is recorded during the summer, where the humidity is low, also during the summer, the lowest interquartile interval is recorded for example in the summer of the decade 1980 the value is 0.07554, not to mention that during 2010 relative humidity marks a slight increase, the maximum value is recorded during the winter with 0.7754.

As mentioned in the first station the precipitation vary widely from summer to the winter in which the maximum value is recorded the standard deviation and the variance is very high, this could be explained first by the presence of outlier that maybe due to the uncertainties and also the occurrence of an extreme

event and sudden precipitation that increase the difference.

The upper graphs show the total average precipitation, average temperature, relative humidity according to four decades; 1980, 1990, 200 and 2010, the lefts graphs, the seasonal variation of the three mentioned variables according to the decades. The graphs have been established for the studied stations.
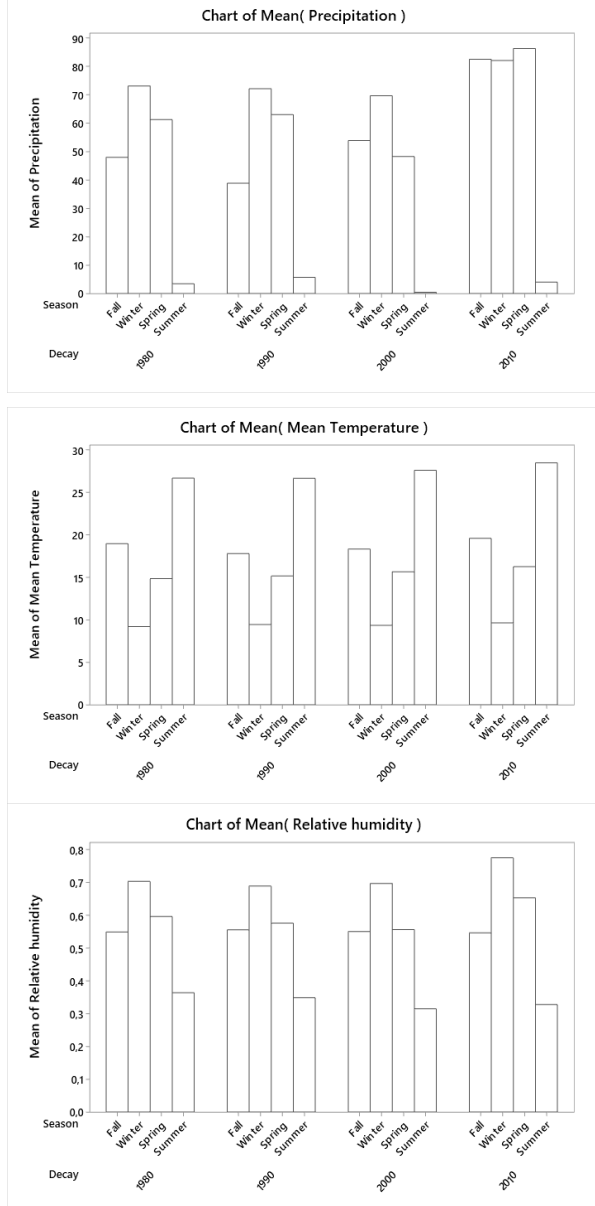

Fig.4. The chart of the mean Precipitation, Temperature and relative humidity for Bab Marzouka Station
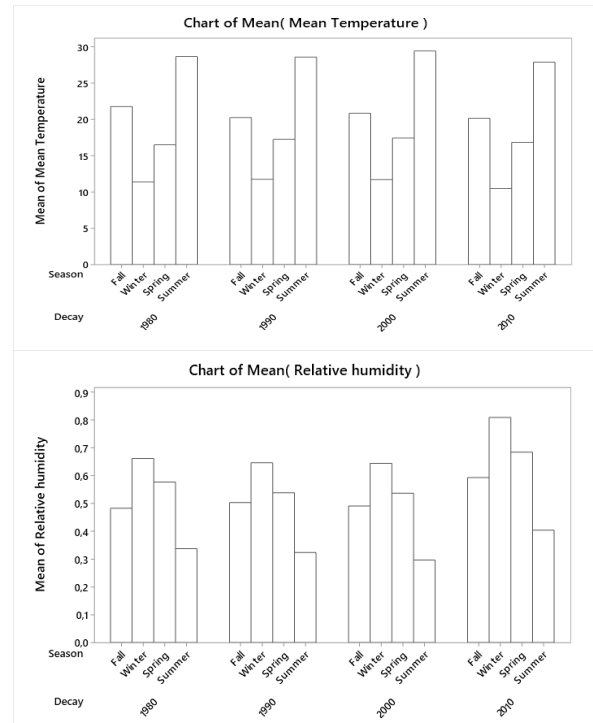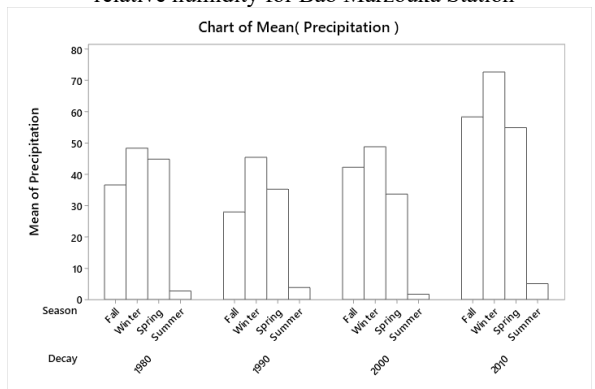

Fig.5. The chart of the mean Precipitation, Temperature and relative humidity for Idriss First Station

4.3. Temporal variation of the studied climate variables
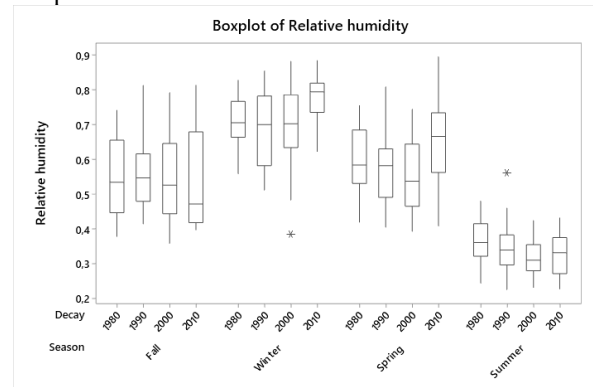
4.3.1. Bab Marzouka Station

Temperature:

An overall rising trend is noted from 1980 to 1990, during the last four decade the temperature has increased from around 17 °C to 19 °C, the maximum mean is recorded in the summer of decade 2010. The seasonal variation exposes a noticeable variation, the lowest value is recorded during the winter with an average of around 9°C, followed by spring fall and the summer.

Precipitation

This later also score a hike over the decades from an average of 45 mm in 1980 to 65 mm in 1990. The seasonal variation of the precipitation goes contrarily to the temperature, the high value is recorded in the winter where in the summer the precipitation is barely null.

Relative humidity

The relative humidity displays a slight rising in the last decade by around 0.05%, the variation goes with the precipitation variation, increase in the winter when high rainfall is recorded and decrease in the summer when the temperature increase.
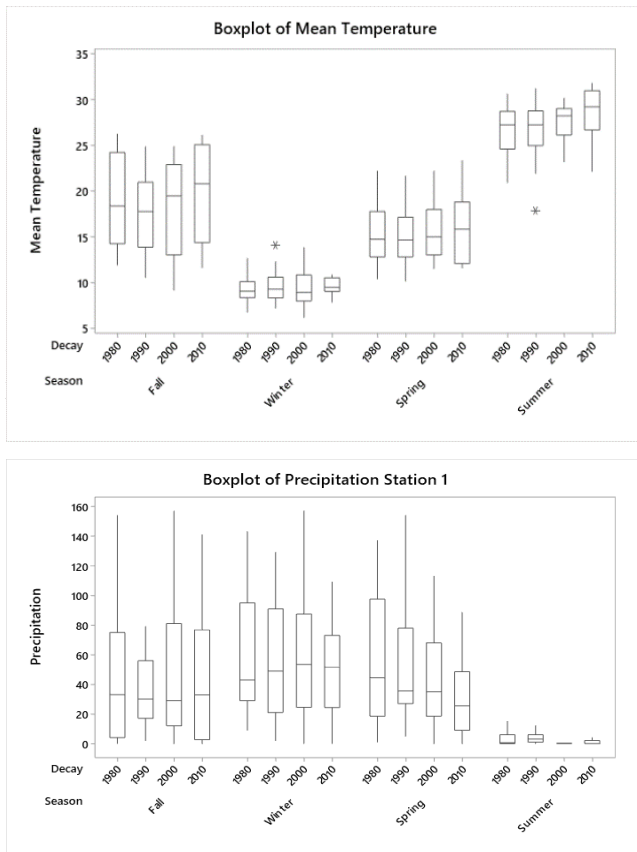
Fig.6. bar chart and box plot of the variables; precipitation, relative humidity and temperature in the station 1 over the decade 1980 to 2010

4.3.2. Idriss First Station

The average temperature records a slight decrease in the decade 201 (around 2°C), contrarily to the rainfall and relative humidity that display in noticeable increase, the intra-seasonal variation goes in concordance with the station 01.

The box plot underlines the dispersion of the dataset, the length of the whiskers measure the outer range from 10 to 25 percentile, and from 75 to 90 percentile, How noticeably different the extremes are. The IQR is small compared to the whiskers, the middle clustering of data about the median represents a large dispersion of the outlier, ability to visualize data skewness. The study of the skewness shows whether the deflection from the median is negative or positive. In upward skewness, the median is shifted toward the lower position with a wider range of data in the upper quartiles compared to the lower, the opposite shows a downward skewness

The plots display the dispersion of the variables; precipitation, temperature and relative humidity, data in both station bab Marzouka and Idriss First, throughout four decades

For the precipitation data, the upper whiskers are large, for fall, winter, and spring. The data are spanning in the first middle

50% of the total rainfall is varying from 0mm to around 62 mm, the maximum recorded value is in the round of 120 mm, for most of the decades in the four seasons the median is skewed left or right, for example during the winter of the decade 1980 the medial is skewed left which mean that 50% of the precipitation

varies between 50 mm and 60 mm, moreover, the fill weights of the plots in the summer is smaller than the rest of the seasons, a large dispersion is also recorded ducting the spring of the decade 2010, the temperature shows a high inter-seasonal variation, during the maximum value are recorded during the summer, in 1980 50% of the temperature vary between 30 and 34 °C, in 2000 the median is barley situated in the middle of the box, less dispersion is marked compared to the fall. Some outlier appears in the winter of 2010, the summer of 1990 and 2000.

The relative humidity vary according to the precipitation rate, most of the ranges have an asymmetric distribution since the median is not in the middle, in the winter of 2010 the relative humidity is too close to the limit of the box, which mean that 50% is close to the value of Q75%, the station 2 marks the same variation trend as the first station with slight variability in the value.
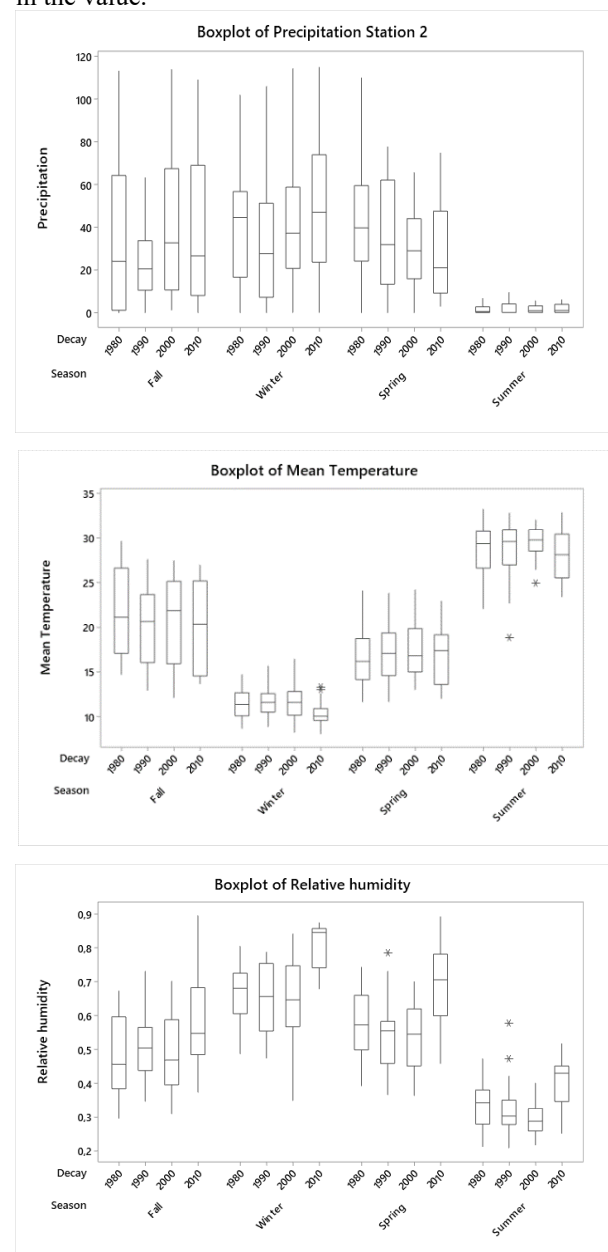

Fig.7. bar chart and box plot of the variables; precipitation, relative humidity and temperature in the station of Idriss First over the decade 1980 to 2010

4.4. Temporal variation of the studied climate variables

The plot underlines the average temperature, the average relative humidity, and the total precipitation for the two stations: bab marzouka and Idriss First throughout four decades.

The Station 2 marks a high value regarding the station, the appearance of the boxes shapes is almost similar, in term of fill weight and whiskers length, the difference is noted in the positioning of the values, and the median,

The box dispersion is high and this indicates that there is a variation in the data set, the median is quietly skewed. For the temperature, most of the value varies between the minimal value (in the order of 6 °C) and the median which vary in the surrounding of 17°C. The relative humidity rises slightly over the decades.

The difference recorded between the two stations could be due to the geographic position, the Idriss First Station where the value of temperature is high and the precipitation and the relative humidity is higher is located near Saiss lowland, the Bab Marzouka station is located in higher altitude (Digital Elevation Model).
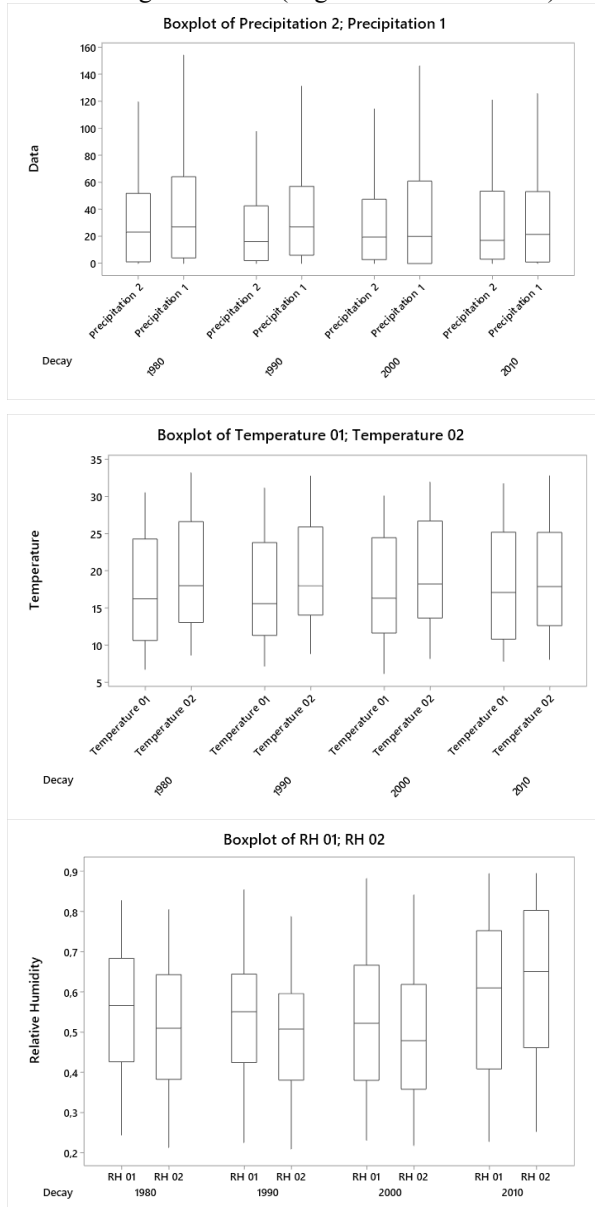

Fig.8. bar chart and box plot of the variables; precipitation, relative humidity and temperature in the station of Idriss First and Bab Marzouka over the decade 1980 to 2010

5 Conclusion

The statistical Analysis is mandatory for the grasping of the hydrological data. The carried out study focused on elaborating a descriptive statistic by means of central tendency and other dispersion parameters, like the test of kurtosis, the skewness and tailedness coefficient. Those parameters were chosen because they provide a clear understanding of the data distribution around the mean. The boxplots displays the distribution of the different quartiles and also the outliers. Two gauge stations were picked for these study Bab Marzouka and Idriss First. The obtaining results shows that

Concerning the Kurtosis test, 1 negative values were recorded during the decade of 1980, then the data distribution has a lighter tail than the normal, thing that could be explained by the low variation of the precipitation frequency.

The climate of the basin is marked by four different seasons; and noticeable contrast in term of temperature and precipitation.

The temporal variation exposes an overall rising trend in the temperature and a decreasing trend in the rainfall and relative humidity over the fours studied decades, it also uncovers the intra-seasonal fluctuation, humid and rainy in the winter and sec and dry in the summer, the autumn and the spring are considered as transition season where the temperature is moderated.

The spatial variation is marked by a slight decreasing in precipitation and increasing of temperature moving from the middle part of the watershed to the downstream, which could be explained by the topographic variation and its impact on the climate. High altitude are generally marked by high precipitation and lower temperature comparing to lowlands areas.

Reference

1. IPCC, climate Change 2014: Synthesis Report. Contribution of working Groups I, II and III to the Fifth Assessment Report of the intergovernmental Panel on Climate change [Core writing Team, R.K. Pachauri and L.A Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp. Available at: (http://www.ipcc.ch/report/ar5/syr/) (2014)

2. MAITY, R.. Springer Transactions in Civil and Environmental Engineering Statistical Methods in Hydrology and Hydroclimatology [en línea]. S.l.: s.n. ISBN 9789811087783. Disponible en: http://www.springer.com/series/13593L. (2006)

3. Brirhet H, 'Intégration des modèles globaux et distribués dans la modélisation hydrologique du bassin versant d'Issan, région de Souss.', PhDThesis, Thèse de Doctorat, Université Sidi Mohamed Ben ABDELLAH, Fes (2016)

4. Schilling, J., Hertig, E., Tramblay, Y., & Scheffran, J. Climate change vulnerability , water

resources and social implications in North Africa (2020).

5. Yang, Z., Zhang, Q., & Hao, X.Evapotranspiration Trend and Its Relationship with Precipitation over the Loess Plateau during the Last Three Decades. *Advances in Meteorology*, *2016*. https://doi.org/10.1155/2016/6809749 (2016)