REGIONALIZATION OF PRECIPITATION DATA IN THE SOFIA REGION

JAKELY Vertu [1], HARY Jean [2], RANDRIANARIVELO Eddy Flocaudel [3], JOHANESA Fernand [4], HANTA Tina Olga [5], RANDRIANANDRASANARIVO Raphaëlson Jacques [6], MAXWELL Djaffard [7]

[1] PhD student , Doctoral School of Life Engineering and Modeling , University of Mahajanga, Madagascar
[2] Lecturer, Higher Institute of Science and Technology of Mahajanga- Faculty of, Science Technology and Environment , University of Mahajanga, Madagascar
[3] PhD student , Doctoral School of Life Engineering and Modeling , University of Mahajanga, Madagascar
[4] PhD student , Doctoral School of Life Engineering and Modeling , University of Mahajanga, Madagascar
[5] PhD student , Doctoral School of Life Engineering and Modeling , University of Mahajanga, Madagascar
[6] PhD student , Doctoral School of Life Engineering and Modeling , University of Mahajanga, Madagascar
[7] Professor, Higher Institute of Science and Technology of Mahajanga -Faculty of Science, Technology and Environment, University of Mahajanga , Madagascar

ABSTRACT

The objective of this work is to assess the state of drought in our study area delimited between latitudes from 14°S to 17°S and longitudes from 47°E to 49°E. In this study, we used precipitation data collected from ECMWF. According to the Standardized Precipitation Index (SPI) method, the SPI value ranges from -1 to 5 in this area. The results showed that drought increases as one moves away from the west coast (the Mandritsara District being most affected). In general, the Sofia Region has a drought close to normal, with a value of 76.19% . The results of the Principal Component Analysis (PCA) confirmed this situation and showed the existence of four different sub-zones of rainfall regimes.

Keywords : *disturbance, drought, SPI, ACP, rain.*

1. INTRODUCTION

Global warming is increasing the frequency and severity of natural disasters around the world. He never stops not evolve and causes dramatic consequences on planet earth, among others, the disruption of the date of the rainy season and intensification of rain [1]. However, the main activity of the population in Sofia Region is agriculture which depends on precipitation. Our goal is to assess the drought situation in this region.

2. METHODOLOGIES

2.1 Data used

The data used are precipitation data from the European Center for Medium range Weather Forecasts (ECMWF). They have a netcdf or " nc " extension and are available online at "http: //www.ecmwf.int" at the synoptic scale. These data are taken from 1979 until 2020 with a spatial resolution step of 0.5*0.5.

2.2 Study area

OUR The study area focuses on the north-western part of Madagascar, east of the Mozambique Channel, particularly in the Sofia region which is delimited by the geographical coordinates between longitudes 47° to 49° East and latitudes - 14° to -17°. It is shown in figure-1 below . This area is made up of seven districts, such as Mampikony , Port-Bergé, Antohihy which is the capital of the region, Analalava , Bealanana , Befandriana -North and Mandritsara . In this figure, they are colored green, blue, turquoise, black, red, yellow and purple respectively. This zone constitutes a vast territory covering an area of 52,503 km ^{2}.

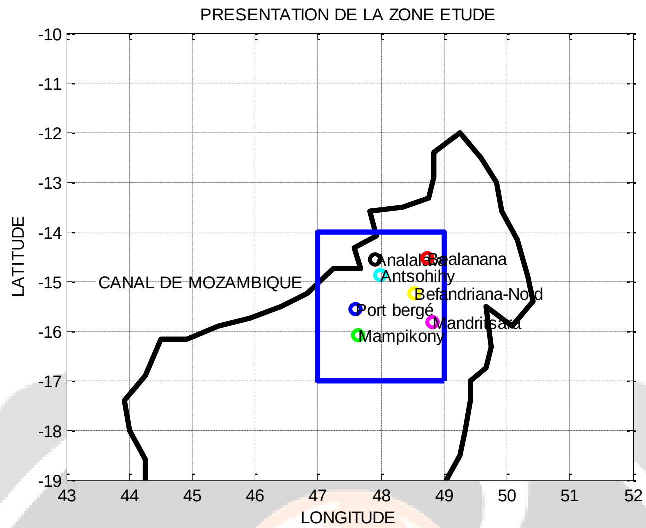


Figure- 1 : Presentation of the study area

2.3 Standardized Precipitation Index (SPI)

The Standardized Precipitation Index is commonly used to characterize local and regional droughts. It is capable of quantifying the deviation of precipitation over a given period, the deficit or the surplus compared to historical average precipitation. It is also based on a long-term precipitation history, of a minimum of thirty years to obtain reliable results. Practically, we use the following formula giving this index:

$$SPI = \frac{(P_i - P_m)}{\sigma} \quad (1)$$

$avec:$

$P_i = pluie\ de\ l'année\ i$

$P_m = pluie\ moyenne\ de\ la\ série\ sur\ l'échelle\ temporelle\ considérée$

$\sigma = écart-type\ de\ la\ série\ sur\ l'échelle\ temporelle\ considérée$

Dryness or wetness depends on the SPI values. Table -1 shows the correspondence between SPI and states [2]:

Table -1: State of drought and humidity

Standardized Precipitation Index (SPI)	Drought Category
SPI ≤ - 2.0	Extreme drought
]-2.0 -1.5]	Strong drought
]-1.5 -1.0]	Moderate drought
]-1.0 1.0[	Close to normal
[1.0 1.5[	Moderate humidity
[1.5 2.0[	High humidity
SPI ≥ 2.0	Extreme humidity

2.4 Principal Component Analysis

2.4.1 General presentation of the ACP

Principal Component Analysis denoted PCA is a so-called factorial method, dimension reduction for the statistical exploration of complex quantitative data: graphic representations of individuals, variables; quality of representation. This process, widely used in climate data analysis, simplifies complex datasets [3].

2.4.2 Objectives of the ACP

- From the point of view individuals: the ACP seeks if we can distinguish among all individuals, groups which are similar or which are different from others
- From the point of view of variables: analogously, PCA can study the structures of their connections. We prospect those which present a strong correlation, or contrary, those which have none with the others [4], [5], [6].

2.4.3. Data

- Let p: real statistical variable X_j, for j *(j=1,...,p)*, observed on n individuals i *(i=1,…,n)*.
- X: data table
- X_{ij} : represents the measurement of the variable X_j on the i-th individual
- X_i : th observation of the table

These measurements are grouped into a matrix X of order *(n*p)* [3] .

$$
\begin{array}{c} \\ X = \end{array}
\begin{array}{c}
\begin{matrix} X_1 & X_2 & \dots & X_4 & \dots & X_p \end{matrix} \\
\begin{bmatrix}
x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\
x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\
. & . & \dots & . & \dots & . \\
x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\
. & . & \dots & . & \dots & . \\
x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np}
\end{bmatrix}
\end{array}
$$

Mathematical calculation of PCA:

- **Climatological average:** allows the rate of daily, monthly and annual precipitation to be determined [7] .

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij} \quad (2)$$

- **It should be noted that:**

The mean of the reduced centered variables is zero and their standard deviation is 1 [8].

Average : ensures the reduction of an entire sample of values to a single one.

$$moy(X_j) = \sum \{X_{ij} | i \in I\} \quad (3)$$

Variance: it is the mean of the squares minus the square of the mean. It makes it possible to characterize the dispersion of values compared to the average.

$$\operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}j)^2 \quad (4)$$

Standard deviation: it is a measure of dispersion of X_j, [8].

$$\sigma\left(x_j\right) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2} \quad (5)$$

2.4.4 Correlation matrix

It gives the first idea of the relationships that exist between the different variables.

2.4.5 Eigenvalues

These are the sum of the squared correlation coefficients of each input variable with the component [9].

2.4.6 Choice of the main axes to retain

The choice of the number of axes to retain is an essential point even if we do not have a rigorous solution. The best-known solution consists of applying *Kaiser 's empirical criterion* : by centering and reducing the data, we retain the principal components corresponding to eigenvalues greater than 1 [8], [9].

2.4.7 Quality of representation of individuals

Quality of representation of individual i on the k axis [9]:

$$qlt_k(\mathrm{e}_i) = \cos^2(\theta) = \frac{c_{ik}^{\ 2}}{\left\|e_i\right\|^2} \text{ and } \left\|e_{ik}\right\|^2 = \sum_{k=1}^{p} c_{ik}^{\ 2} \quad (6)$$

We use the parameter $\cos^2\theta$ to characterize the quality of representation (*qlt*) on an axis.

- The closer qlt_i is to 1, the more well it is represented.
- The closer qlt_i is to 0, the more poorly it is represented [10].

2.4.8 Quality of representation of variables

Remember that the coordinates of the variables form the elements of the matrix ϕ and that these projections are placed in a circle of unit radius (correlation circle). On a factorial plan defined by two main axes, the interpretation of the projection of a variable translates as follows:

- a variable close to the correlation circle is well represented in this plane;
- a variable close to the origin of the correlation circle is poorly represented in this plane [4], [5], [6].

3. RESULTS AND DISCUSSION

3.1 Results on the standardized precipitation index

To assess the drought situation in a given region, a suitable index must be chosen. In this study, we used the “Standardized Precipitation Index”. Because it is easy to handle and does not require many other variables. It is interesting to analyze the variation of the Standardized Precipitation Index as a function of time. Figure-2 shows the distribution of drought in the Sofia region according to the value of the standardized precipitation index (SPI). The Analalava district is the wettest compared to the others. The districts of Port-Bergé, Antsohihy, Bealanana and Befandriana-Nord have a drought category close to normal. Mandritsara and Mampikony districts are the most affected by drought. The latter increases as we move away from the West Coast.

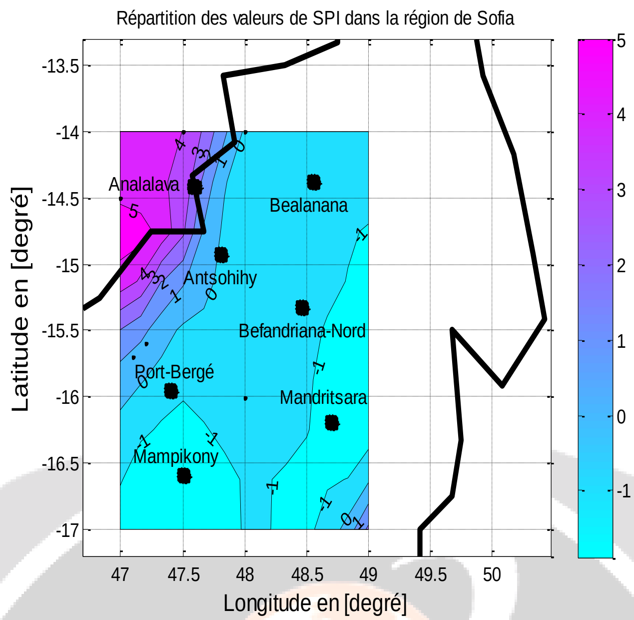
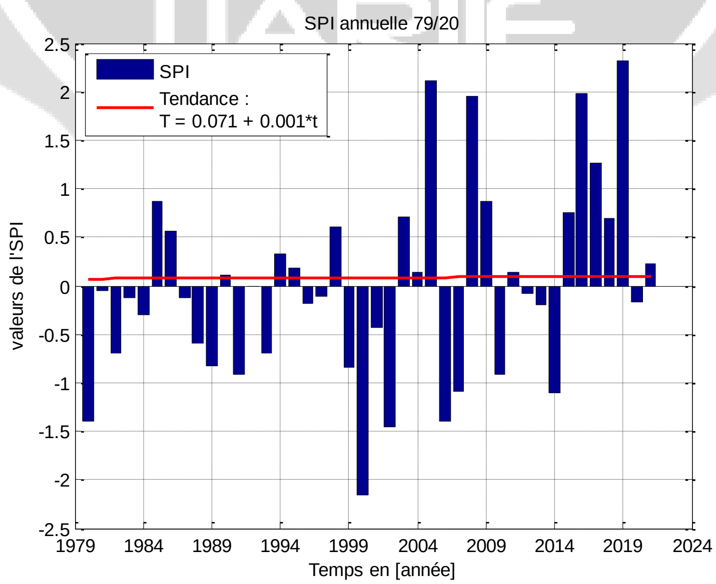


Figure- 2: Distribution of SPI value

3.2 Study of the state of drought

Charte-1 represents the annual evolution of the drought index in our area. In this figure, negative indices correspond to dry periods and positive indices represent wet periods. In this figure the trend (right in red) is increasing with a positive slope of 0.001. According to the drought classification in Table-2, the year 1999 experienced extreme drought, the years 1979, 2001, 2005, 2006 suffered moderate drought, the years 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 2000, 2002, 2003, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2017, 2019, 2020 experienced near-normal drought. Moderate humidity was observed in 2016 and high humidity in 2007 and 2015, while periods of extreme humidity marked the years 2004 and 2018.
Thus, near-normal drought dominates in this region. And this result is confirmed in table-3, of which 76.19% for drought close to normal, that of moderate drought reaches the value of 9.52% and the others are negligible.



Charte-1: Histogram of the drought index

Table- 2 : State of drought in the Sofia region from 1979-2020

Years	SPI	Drought category
1979	-1.4	Moderate drought
1980	-0.06	Close to normal
1981	-0.70	Close to normal
1982	-0.13	Close to normal
1983	-0.29	Close to normal
1984	0.87	Close to normal
1985	0.57	Close to normal
1986	-0.12	Close to normal
1987	-0.59	Close to normal
1988	-0.82	Close to normal
1989	0.11	Close to normal
1990	-0.92	Close to normal
1991	0	Close to normal
1992	-0.70	Close to normal
1993	0.32	Close to normal
1994	0.18	Close to normal
1995	-0.20	Close to normal
1996	-0.10	Close to normal
1997	0.60	Close to normal
1998	-0.84	Close to normal
1999	-2.15	Extreme drought
2000	-0.43	Close to normal
2001	-1.45	Moderate drought
2002	0.71	Close to normal
2003	0.14	Close to normal
2004	2.11	Extreme humidity
2005	-1.4	Moderate drought
2006	-1.09	Moderate drought
2007	1.95	High humidity
2008	0.88	Close to normal
2009	-0.91	Close to normal
2010	0.14	Close to normal
2011	-0.08	Close to normal
2012	-0.20	Close to normal
2013	-1.1	Close to normal
2014	0.75	Close to normal
2015	1.98	High humidity
2016	1.27	Moderate humidity
2017	0.69	Close to normal
2018	2.32	Extreme humidity
2019	-0.16	Close to normal
2020	0.22	Close to normal

Table-3 : Drought category and their percentage

Drought Category	Number of years	Percentage of time (year)
Extreme drought	1	2.38%
Severe drought	0	0%
Moderate drought	4	9.52%
Drought close to normal	32	76.19%
Moderate humidity	1	2.38%
High humidity	2	4.76%
Extreme humidity	2	2.38%
Total	42	97.64%

3.2 General PCA results

3.2.1 Choice of areas to retain regarding rainfall

Table-4 illustrates the eigenvalues, inertia and cumulative inertia of rainfall. By applying the Kaiser criterion (we must retain the factors which have eigenvalues greater than 1) and according to the analysis of this table, we can retain the first two axes. They represent respectively 69.23% and 25.80% of the total inertia. Figure-3 shows the eigenvalues of rainfall.

Table-4: Eigenvalues of, inertia and cumulative percentage of rainfall.

Factors	Eigenvalues	Inertia (% variance)	Cumulative (%)
1	**8,308**	**69,232**	69,212
2	**3,096**	**25,801**	95,005
3	0.318	2,646	97,650
4	0.116	0.965	98,614
5	0.057	0.473	99,087
6	0.039	0.321	99,408
7	0.023	0.191	99,599
8	0.018	0.147	99,745
9	0.014	0.117	99,862
10	0.005	0.041	99,903
11	0.004	0.037	99,940
12	0.004	0.030	99,970

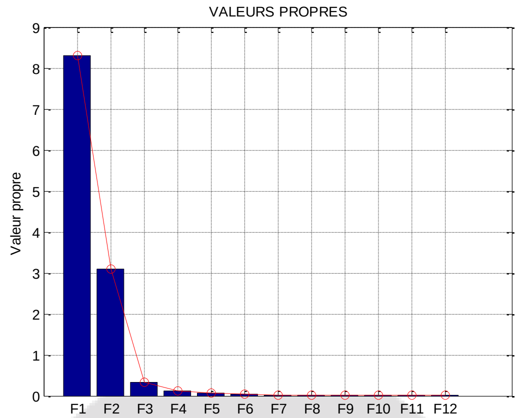


Figure- 3 : Eigenvalues of rainfall

3.2.2 Results on the variables

In Table-5 of the coordinates of the variables below, the linear correlation coefficients between the variables and the factors retained are classified in the coordinate column. Based on these coefficients we can project the variables into the factorial plan. As a result, the months of: January, February, March, April, May, September, October, November and December are variables well represented on the $F_{1\ axis,}$ because they have coordinates close to 1. The months of: June, July and August are variables well represented on the $F_{2\ axis}$, because they have eigenvalues close to minus one (-1).

Table-5 : Coordinates of variables

Variables	Contact details	
	F1	F2
January	0.794	0.561
FEBRUARY	0.807	0.533
March	0.935	0.279
April	0.960	0.077
May	0.923	-0.340
June	0.642	-0.744
July	0.511	-0.827
August	0.687	-0.716
September	0.897	-0.343
October	0.917	0.020
November	0.883	0.427
December	0.859	0.473

3.2.3 Projection of variables in the factorial plan F1-F2

From the projection of the variables in the factorial plan F1-F2 (figure- 4), the months of: January, February, March, April, May, September, October, November and December are variables positively correlated with the axis F1. The months of November, December, January and February have a strong positive correlation on this axis. These are the months of the summer season, so we can say that the F1 axis represents the summer season in our area. On the other hand, the month of June, the month of July and the month of August are variables which have a strong correlation to the F2 axis. These are the months of the winter season, so we can say that the F2 axis shows the winter season in our area.

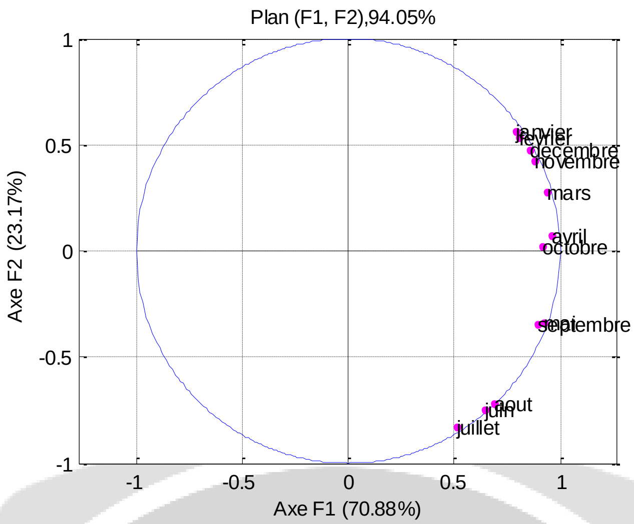


Figure-4 : Variable correlation circle

3.3 Results on individuals

3.3.1 Projection on individuals

From the classification of individuals, we can subdivide our study area into eight different sub-zones: four sub-zones of F1 and four sub-zones on F2. Figure-5 below shows the grouping of 117 points of individuals in the factorial space F1 and F2.

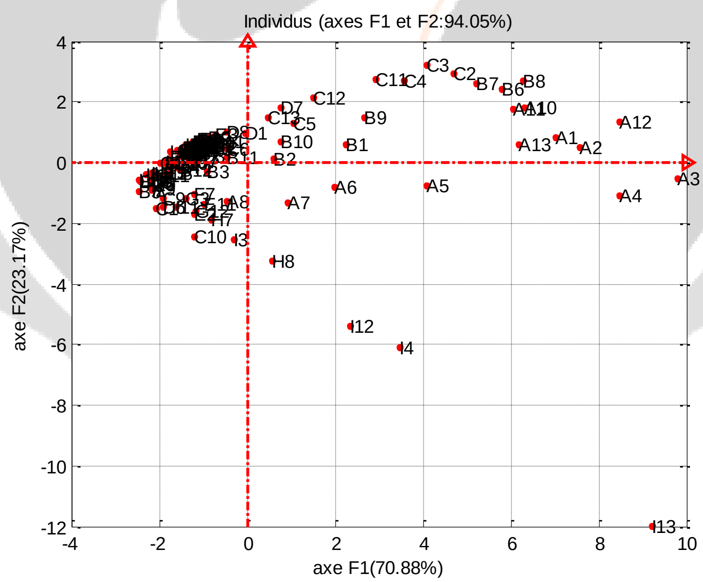


Figure-5 : Classification of individuals

According to the results of our analysis, our study area can be subdivided into eight regions: four sub-zones carry information from the summer season and four sub-zones carry information from the winter season (figure-6).
Four sub-zones have information for the **summer season** :
Region I : area with high rainfall during the summer season which is marked by legends in green hexagons such as: A1, A2, A3, A4, A12 and A13;
Region II : moderately rainy area during the summer season which is marked by blue hexagon legends such as: A10, A11, B6, B7, B8, C2, C3, C4 and C11;

Region III : area with low rainfall during the summer season shown by legends in purple hexagons such as: B1, A5, and B9

Region IV : non-rainy zone during the summer season noted by the symbol and legends in turquoise hexagon such as: B4, B5, B12, B13, C1, C7, C8, C9, D2, D3, D4, D5, D6, D9, D13, E3, E4, E5, E6, E7, E8, E9, E11, E12, E13, F1, F2, F3, F4, F5, F7, F8, F9, F10, F11, F12, F13, G10, G11, G13, H1, H2, H3, H4, H5, H6, H9, H10, H11, H12, H13, I1, I2, I5, I6, I7, I8, I9, I10 and I11.

Four sub-zones carry information for the winter season:

Region V : area with high rainfall during the winter season noted by the legend in green circles such as: I3;

Regions VI : area with average precipitation during the winter season which is represented by legends in blue circles such as: I4 and I12;

Regions VII : area with low precipitation in winter represented by legends in purple circles such as: C10, E2, H7, H8, and I3;

Regions VIII : non-rainy zone during the winter season which is marked by legends in turquoise circles such as: C5, C12, C13, D1, D7 and D8.

This result confirms the previous result (figure 2). Drought increases as we move away from the West Coast towards the East.

Those who are in the red hexagon, are the unclassified areas.

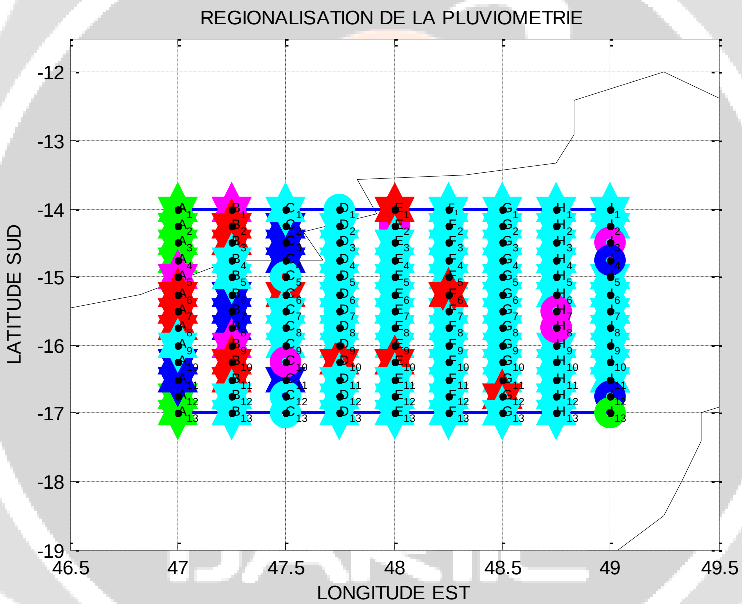


Figure-6 : Classification of rainfall data in the study area

4. CONCLUSION

This work consists of studying the state of drought and the distribution of rainfall in the Sofia Region. The values of the standardized precipitation index (SPI) give us the state of drought. The results showed that near-normal drought dominates in this region, with a value of 76.19%, and it increases as one moves away from the west coast. The PCA results confirmed this situation and showed that the zone is subdivided into four sub-zones with different characteristics: sub-zone of high rainfall, sub-zone of average rainfall, sub-zone with low rainfall and non-rainy sub-zone during the rainy season. winter and summer.

5. REFERENCES

[1]: SITUATION OF THE MAIN ENVIRONMENTAL INDICATORS OF MADAGASCAR IN 2007.

[2]: Geographical Space Review and Moroccan Society n°32, February 2020.

[3]: P.besse www.math. univ-toulouse.fr/besse, M2 MASS, TP4: Introduction to SAS software Multivariate statistical procedures: Component Analysis.

[4]: C. DUBY, S ROBIN; Principal component analysis, July 2006. Pages 4,5.
[5]: Aimé KOUDOU et al, CONTRIBUTION OF PRINCIPAL COMPONENT ANALYSIS TO THE REGIONALIZATION OF THE RAINFALL OF BSSIN VESRSANT DU N'ZI, CENTRAL IVORY COAST, 2015. Page 162.
[6]: Ali Kouani, S. El Jamali and M.Talbi, Principal Component Analysis. A factorial method for processing educational data, February 2007. Page 3.
[7]: J. Gergaud, Fundamental Unit: Linear Algebra: an application to Principal Component Analysis, October 2006. Page 27.
[8] RAFENOMANANJARA Nomena Barisata, Study of rainfall in the South-West region of Madagascar: Approach by Principal Component Analysis and Anal, its Spectra, November 4, 2011. Pages 12, 13, 22.
[9]: ADD3-MAB, D-interpretation of an ACP, 2011. Papes 4, 29.
[10]: Jean-Marc Labatte, jean-marc. Labatte@univ-angers.fr.