SPATIAL AND TEMPORAL VARIABILITY OF PRECIPITATION IN SOUTHERN OF MADAGASCAR

Harimino Andriamalala RAJAONARISOA[1*], Irrish Parker RAMAHAZOSOA[2], Adolphe Andriamanga RATIARISON[3]

[1] PhD student, Dynamic laboratory of Atmosphere, Climate and Oceans (DyACO), Physics and Applications, Sciences and Technologies Domain, University of Antananarivo, Madagascar
[2] Lecturer, Dynamic laboratory of Atmosphere, Climate and Oceans (DyACO), Physics and applications, Sciences and Technologies Domain, University of Antananarivo, Madagascar
[3] Professor Emeritus and laboratory manager, Dynamic laboratory of Atmosphere, Climate and Oceans (DyACO), Physics and applications, Sciences and Technologies Domain, University of Antananarivo, Madagascar

ABSTRACT

This research aims to analyze the spatial and temporal variability of precipitation in Southern of Madagascar. Precipitation data is reanalysis data in grid points (54 points), between January 01st, 1979 and December 31st, 2017. Various technics are used to characterize this variability. The Principal Component Analysis method makes it possible to represent individuals in the factorial plan. The first factor axis corresponds to precipitation and the second to the seasons of the year (winter and Southern summer). Then, the individuals are grouped according to the signs of their coordinates on the factorial plan. As a result, we get a subdivision into four zones: a zone of weak precipitation in Southern winter, a zone of weak precipitation in Southern summer, a zone of heavy precipitation in Southern winter and a zone of heavy precipitation in Southern summer. To optimize the regionalization, the representation on the factorial plan is combined with the K-Means and Fuzzy C-Means clustering methods. Then, another representation on the t-sne plan with perplexity parameter 5 also followed by the two previous clustering methods. All classifications are evaluated by Dunn's index. The best classifications are those of t-sne representation combined with the Fuzzy C-Means or K-Means clustering method. The final result of the spatial and temporal variability analysis of precipitation shows an East-West gradient of precipitation whether in summer or in Southern winter.

Keywords: *Principal Component Analysis, Fuzzy C-Means, K-Means, Precipitation, t-sne, Variability.*

1. INTRODUCTION

The analysis of the spatial and temporal variability of precipitation is the subject of numerous studies [9], [27]. This work generally lead to a zoning of the distribution of precipitation. Research are also conducted on the impacts of this variability to the vegetation [2], hydrological models [7] and underground water resources [31]. The majority of these studies relate to an arid or semi-arid zone such as the Sahelian West Africa zone [6]. In these areas, water is a main issue. Southern of Madagascar is also a victim of this lack of water. Limited access to clean water and poor sanitation and hygiene practices are of particular concerns which is especially linked with chronic malnutrition [34]. In addition, South Madagascar is also a victim of climate change and is vulnerable to the phenomenon of drought. International organizations such as UNICEF are seeking to assess the current situation of groundwater resources in Southern of Madagascar. As a result, according to the bulletin N ° 12 (from October 21 to

November 20, 2019) [33], concerning the drought in the South Madagascar, UNICEF reports great variabilities of the aquifers in the Great South Madagascar.

But it is important to know about the spatial and temporal variability of precipitation in the South Madagascar to address those issues.

Objectives of this study are to find the best zoning for the spatial and temporal distribution of precipitation in Southern of Madagascar using various mathematical technics.

2. METHODS

2.1 Experimentation data

We use precipitation data from ECMWF (European Center for Medium-Range Weather Forecast) [12]. These data are reanalysis data for a spatial resolution of 1 ° x1 °, between 22 ° S - 30 ° S and 43 ° E - 48 ° E. Thus, at each grid point (in total 54 grid points), we have series of daily precipitation data (expressed in meters then converted to millimeters). The temporal coverage of the data is between January 01st, 1979 and December 31st, 2017, which is over 39 years. Figure 1 shows the grid points with their classifications to which the precipitation data relate.

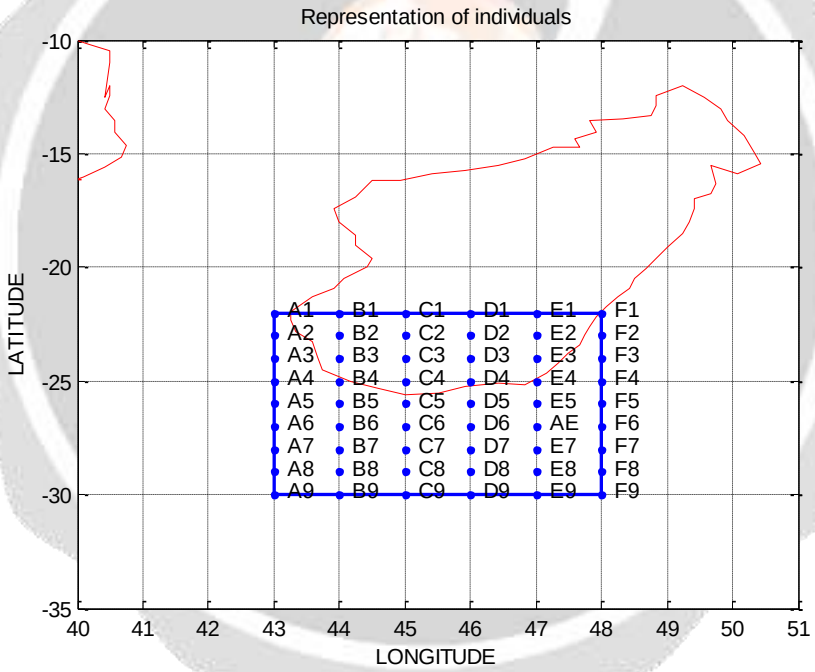

Fig-1: Representation of individuals on grid points.

2.2 Dataset

The cumulative monthly precipitation data is calculated from the daily precipitation data. For each month, the average of these accumulations over the 39 years forms the average monthly precipitation accumulation data. We consider the grid points as individuals and the months of the year as variables. Thus, we obtain Table 1 formed by 54 rows and 12 columns, each cell of which contains the average monthly precipitation cumulative value at the corresponding point. This Table constitutes the data set for all the following mathematical treatments.

Table -1: Dataset Table (Precipitation in millimeter).

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
A1	197.8	156.3	77.3	25.8	11.1	6.1	3.4	3.2	3.5	5	19	66.8
A2	153.5	123.5	64.8	22.5	13	7.5	5.4	3.8	4.4	6.3	19	63.3
A3	114.8	101.4	55.3	28.3	19.7	11.2	9.4	5.4	6	7.8	18.3	58.8
A4	88.4	87.1	55.1	36.6	30	17.3	15.4	9	7.7	11.5	20.6	54.3
A5	74	79.5	63.5	46.5	44.7	27.9	24.8	13.4	10.5	16.6	23.7	54.3
A6	69.4	71.3	76	59	62.6	41.6	37.1	21.5	16.3	24.1	26.3	53.4
A7	73.2	69.5	77.9	65.8	72.7	49.2	41.8	31.4	22	28.7	27.2	53.3
A8	67.4	69.2	78.4	72.8	76.6	53.3	46.4	39.2	26.7	34.6	30.9	50.7
A9	61.2	63.3	78.1	68.7	73.3	54.3	48.4	42.8	31.5	38.1	36.7	53.4
B1	330.7	250.9	139.6	34.2	9.9	2.7	1.8	2.9	8.3	28.1	72.4	181.6
B2	284.6	214.2	124.3	39.4	16.1	4.6	3.7	4.0	9.4	32.6	74.4	172.5
B3	206.9	167.6	90.9	32.5	15.5	5.9	4.8	4.9	8.5	26.1	51	137.3
B4	91.3	92.1	50.8	34.1	29	17.5	14.5	9.4	8.5	14.7	22.4	61.1
B5	78.6	81.4	59.1	50.6	48.5	29.4	27.1	14.9	10.8	18.2	25.5	56.1
B6	79.1	76.9	79.4	67	70.6	44.7	40	24.3	14.7	24	28	56.3
B7	81.2	75.2	81.2	74.2	80.2	51.3	43.3	33.8	22	30.4	26.6	53.7
B8	75.9	74	79.1	81.1	80.5	52.2	45.9	41.3	27.2	35.3	29.3	52.6
B9	66.5	67.9	79.8	75.5	76.8	53.4	48.9	45.3	33.1	39.2	34.3	52.7
C1	252.8	187.6	102.8	27.7	9.7	2.2	2	2.3	8.2	31.8	72.4	171.5
C2	231.4	170	100.7	36.4	15	4.2	4	4.3	10.6	38.5	82.2	172.3
C3	197.2	154.2	89.4	38.4	17.8	7.6	6.6	6.6	12	35.7	73.2	164.9
C4	158.1	132.3	83	47.5	29.1	15.3	13.7	12.7	15.3	33.8	56.7	135.5
C5	86.9	85.8	67.2	61.2	64.7	38.3	35.5	20.6	14.4	21.7	29.1	63.8
C6	89.7	87.5	88.6	79.6	92.8	51.2	44.3	31	15.9	26.1	31.5	65.5
C7	90.4	86.4	87.8	85.9	88.4	55.9	47	38.8	23.3	33.6	28.8	57.2
C8	86	85	85.1	87.8	82.7	54.5	46.5	42.6	30.1	38.6	31.50	56.6
C9	73.4	78.9	85	77.9	75.5	53.2	49.2	44.3	33	42.1	34.5	55.7
D1	202	147.9	88.2	26	8.5	3.1	3.1	2.6	6	23	57.3	130.3
D2	195.2	144.3	93.6	38.8	15.8	6.8	7.4	6.4	10.8	33.4	76.3	149.5
D3	178.7	141.8	91.9	48.3	19.9	10.2	11.3	9.3	14.3	36.5	81.5	153.3
D4	146	129.3	86.5	58.5	31.5	19.2	18.9	15.9	19.5	36.1	68.4	139.8
D5	85.7	92.8	72	70.1	68.8	43.7	38.4	25.6	17.9	22.6	31.7	67
D6	101.3	107.4	101.3	94.5	106.6	61	50.1	41.2	21.1	30	35.1	78.4
D7	96.2	99.6	97.8	95.5	93.2	59.7	48.7	46.3	27.1	37.9	34.4	65.8
D8	92.9	94.1	93.9	88.1	82.1	57.7	47.8	46	32.6	44.6	37.2	60.6
D9	84.7	87.4	95.8	77.5	77.8	56	48.9	46.5	33.1	46.6	38.7	63.4
E1	269.4	222.7	195.1	105.3	79.9	43.3	51.6	43.1	40	82.5	131.5	202.5
E2	257.4	220.1	197.2	119.6	89.7	48.2	57.4	50.3	49.9	96.5	143.6	205.7
E3	241.8	218.7	191.3	128.1	87.4	44.7	53.2	42	45.9	88.7	141.1	206.5
E4	169.1	170.9	141.4	110.8	72.4	41.3	43.6	33.2	34.6	57.2	95.4	160.7
E5	114.2	131.7	113.7	109.1	101.7	65.9	57.7	44.1	27.4	33.9	45.6	86.4
E6	118.5	128.2	115.6	112.8	110.6	67	57.4	52.3	30	38.6	46	94.4
E7	101.6	113.9	105.8	106.2	92.9	63	49.9	51.6	32.8	43.7	44.5	83.5
E8	100	104.8	102.2	97.6	85.2	61.5	49.4	49.3	36.2	50.3	46.4	72.4
E9	98.2	93.5	103.5	82.3	80.5	59.6	49.7	48.2	38.1	51.6	45.1	70.9
F1	275.7	266.9	266.8	192.7	168.6	97.8	98.2	73.8	51.8	91.3	98	155.9
F2	253.1	254.1	250.2	188.8	164.3	98.3	93.1	73.7	51.5	89.5	91.5	150.6
F3	207.7	226.8	223.1	181.1	161.5	96.4	89.2	71.7	50.2	81.1	86.8	143.2
F4	166.4	193.4	182	159.6	146.8	89.1	79.7	65.9	41.9	61.3	76.4	122.7
F5	147.5	159.8	144.2	135.2	125.1	82.1	68.3	60.1	35.6	48.7	65.4	106.2

F6	130.6	134.3	119.4	118.6	102.2	70.5	56.8	52.7	36.4	48.4	62.5	98.2
F7	105.5	120.9	110.1	113.3	89	63.4	49.6	49.6	36.7	50.2	56.3	95.7
F8	104.9	112.8	112.3	101.2	85.6	62.5	49.5	48.5	38.7	54.4	54.2	84.3
F9	108.5	104.6	111.1	91.9	83.2	61.9	50.9	48.1	44.5	59.4	51.9	81.8

2.3 Methods

Four main steps adopted for data processing:

- Step 1: Reduction of the dimension of the data in two dimensions,
- Step 2: Visualization of individuals in a plan,
- Step3: Regrouping of individuals using a cluster method,
- Step 4: Evaluation of the classification obtained by the Dunn index.

a) Step 1: Reduction of the dimension of the data in two dimensions

We start from the data shown in Table 1. Each individual (or each row) has 12 variables (the 12 months of the year). The initial dimension of the data is therefore equal to 12. The first step is to reduce this dimension to 2. For this, we use two different methods: Principal Component Analysis [8] [16] [17] [18] [25] [30] and t-sne (t-distributed Stochastic Neighbor Embedding) [19] [20] [23] [35]. For the Principal Component Analysis, we are looking for a representation of the 54 individuals, in a 2-dimensional subspace. We are then trying to define 2 new linear combination variables of the 12 initial variables which will cause the least information loss. These two new variables are called "Principal components". The axes which they determine are the "Principal axes" and the plan thus formed is the "factorial plan". The t-sne method is another method for reducing the size of the data. Unlike Principal Component Analysis, it is an unsupervised nonlinear technics. The t-sne algorithm constructs a new representation of the data so that close data in 12-dimensional space has a high probability of having close representations in the new 2-dimensional space. Conversely, data that is distant in the original space has a low probability of having close representations in the new space [24]. The algorithm begins by converting the large Euclidean distances between the data points into conditional probabilities that represent similarities. The similarity between point x_i and point x_j is the conditional probability $p_{j|i}$ that x_i would choose as its neighbor x_j if the neighbors were chosen in proportion to their probability density under a Gaussian centered on x_i. The equation (1) gives us the mathematical formula of $p_{j|i}$:

$$p_{j|i} = \frac{exp(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2})}{\sum_{k \neq i} exp(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2})}. \quad (1)$$

The probability that x_i and x_j are close in high dimensional space is p_{ij} such that

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}. \quad (2)$$

For our case, n = 12.

Probability that the representation of x_i and x_j in 2-dimensional space is close is q_{ij} such that

$$q_{ij} = \frac{(1+\|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1+\|y_k - y_l\|^2)^{-1}}. \quad (3)$$

For nearby data points, the probability is relatively high, while for widely separated data points, it will be minimal.

Another feature of t-SNE, perplexity, is an adjustable parameter that indicates how to balance the local and global aspects of the data. The parameter is in a way an estimate of the number of neighbors close to each point.

The value of perplexity has an influencing effect on the resulting images. It is defined by

$$Perp(P_i) = 2^{H(P_i)}, \quad (4)$$

$$where \ H(P_i) = -\sum p_{i|j} \log_2 p_{i|j}. \quad (5)$$

Optimizing the representation in the t-sne plane consists in minimizing the Kullback - Leibler divergence between the two distributions by varying the location of the points on the low dimensional map, which means minimizing it by using the gradient descent method, of the following elements

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{ji} \log \frac{p_{ji}}{q_{ji}}. \quad (6)$$

The t-sne algorithm is as follows:
Data: dataset X = {x_1, x_2,…, x_{12}} ;
Cost function parameters: perplexity ***perp*** ;
Optimization parameters: number of iterations T, learning rate η, momentum α (t).

Result: Low-dimensional data representation $Y^{(T)} = \{y_1, y_2\}$
Begin

Compute pairwise affinities $p_{j|i}$ with perplexity ***perp*** using equation

$$p_{j|i} = \frac{\exp(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2})}{\sum_{k \neq i} \exp(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2})}.$$

Set $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$

Simple initial solution $Y^{(0)} = \{y_1, y_2\}$ from $\mathbb{N}(0, 10^{-4}I)$ where I denotes the unit matrix.

For t = 1 **to** T **do**

Compute low-dimensional q_{ij} using the equation

$$q_{ij} = \frac{(1+\|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1+\|y_k - y_l\|^2)^{-1}},$$

Compute gradient $\frac{\delta C}{\delta y_i}$ using the equation:

$$\frac{\delta C}{\delta y_i} = 4 \sum_j \frac{(p_{ij} - q_{ij})(y_i - y_j)}{(1+\|y_i - y_j\|^2)},$$

Set $Y^{(t)} = Y^{(t-1)} + \eta \frac{\delta C}{\delta Y} + \alpha(t)(Y^{(t-1)} - Y^{(t-2)})$

End

End

b) Step 2: Visualization of individuals in a plan

With the Principal Component Analysis, the representation of individuals in the factorial plan (F_1, F_2) allows to visualize the data in 2 dimensions. For t-sne, at each value of perplexity, we can visualize the individuals in the t-sne plane. As part of our study, we take whole values of perplexity up to 50. Among the 50 possible visualizations, the one which offers a better apparent grouping of 2-dimensional individuals will be retained.

c) Step3: Regrouping of individuals using a cluster method

Once we have the representation of individuals in 2 dimensions, the next step is clustering. Two methods are used in our research: the K-Means method [13] [22] and the Fuzzy C-Means [3] [5] [14] [21] [26] [28] [36].

- **K-Means method**

The k-Means method seeks to minimize the sum of the squares of errors

$$J(C) = \sum_{c \in C} \sum_{w \in c} d(w, m(c)), \quad (7)$$

where $m(c)$ is the center of gravity of class c.
The k-Means method does the following steps:

i. Initialize k-centers m_1, m_2,…, m_k,
ii. Assign each individual to the nearest class $C = \{c_1, c_2, .., c_k\}$:
$$w \in c_l \; si \; d(w, m_l) = \min_{l'=1,...,K} d(w, m_{l'}),$$
iii. Recalculate the centers of gravity of the new classes,
iv. Repeat ii) and iii) until convergence.

➢ Fuzzy C-Means (FCM) method

Fuzzy C-Means (FCM) is a grouping method that allows each data point to belong to several clusters with different degrees of membership. FCM is based on the minimization of the following objective function

$$J_m = \sum_{i=1}^{D}\sum_{j=1}^{N} \mu_{ij}^{m} \left\| x_i - c_j \right\|^2, \quad (8)$$

where:
D: Number of data points,
N : Cluster name,
x_i is the i^{th} data point,
c_j is the center of the j^{th} cluster,
μ_{ij} is the degree of membership of x_i in j th group ($0 \leq \mu_{ij} \leq 1$). For a given data point x_i, the sum of the membership values of all the clusters is equal to one,
m: exponent of the fuzzy partition matrix to control the degree of fuzzy overlap (m> 1).

Fuzzy overlap refers to the degree of confusion of boundaries between clusters, that is, the number of data points with significant membership in multiple clusters. The possible values of the exponent m must be greater than 1. The smaller values indicate a lower degree of overlap. In other words, as m approaches 1, the boundaries between the groups become sharper.
Fuzzy C-Means performs the following steps during clustering:

i. Randomly initialize cluster membership values μ_{ij} ,
ii. Calculate cluster centers c_j such that $c_j = \frac{\sum_{i=1}^{D} \mu_{ij}^{m} x_i}{\sum_{i=1}^{D} \mu_{ij}^{m}}$,
iii. Update μ_{ij} according to the formula : $\mu_{ij} = \frac{1}{\sum_{k=1}^{N} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}}$,
iv. Calculate the objective function $J_m = \sum_{i=1}^{D}\sum_{j=1}^{N} \mu_{ij}^{m} \left\| x_i - c_j \right\|^2$,
v. Repeat steps ii through iv until J_m is improved by less than a specified minimum threshold (default, 0.01 between two successive iterations) or after a specified maximum number of iterations (default, 25 iterations).

Optimizing clustering involves adjusting fuzzy overlaps. Indeed, for each overlap exponent m:

- ✓ Cluster data,
- ✓ Classify each data point in the cluster to which it has the highest degree of membership,
- ✓ Find data points with maximum membership values less than 0.6. These points have more indistinct classification,
- ✓ To quantify the degree of fuzzy overlap, calculate the average maximum membership value for all data points. A higher average maximum membership value indicates that there is less blurry overlap,
- ✓ Plot the results of the grouping.

d) Step 4: Evaluation of the classification obtained by the Dunn index

Dunn index [4] [11] [10] [15] [29] [32] [37] searches for the minimum distance between two classes in the partition while taking into account the distribution of the elements inside of class. The Dunn index or score, noted S_D , is based on the average points of each group μ_k with

$$\mu_k = \frac{1}{|I_k|} \sum_{i \in I_k} x^i, \quad (9)$$

where
x^i : individual or point,
I_k : the unity of points belonging to k-group.
The diameter of the group Δ_k is such that

$$\Delta_k = max_{i,i' \in I_k} d(x^i, x^{i'}), \quad (10)$$

where $d(x^i, x^{i'})$ is the Euclidean distance between the two individuals x^i and $x^{i'}$.

Dunn index or score has the expression

$$S_D = \frac{\min_{1\le k\le k'\le K} d(\mu_k,\mu_{k'})}{\max_{1\le k\le K} \Delta_k}, \quad (11)$$

where $K \geq 2$, the number of groups we want to form. (For our case, $K = 4$).

Dunn index is a positive or zero real number. More the Dunn index is large, the partition is better.

3. RESULTS

3.1 Data visualization results

The Figure 2-a shows the representation of the individuals in the factorial plan (F_1, F_2). The first factorial axis represents 61.92% of the total inertia and corresponds to precipitation. The second factorial axis (33.13% of total inertia) relates to the seasons of the year (winter and southern summer).

For the visualization of the data by t-sne, the perplexity parameter retained is worth 5. In fact, each perplexity parameter corresponds to a representation. We can vary the perplexity parameter from 1 to 50. Thus among the obtained 50 representations, only from the value of perplexity 5 gives groupings obviously visual according to the Figure 2-b.

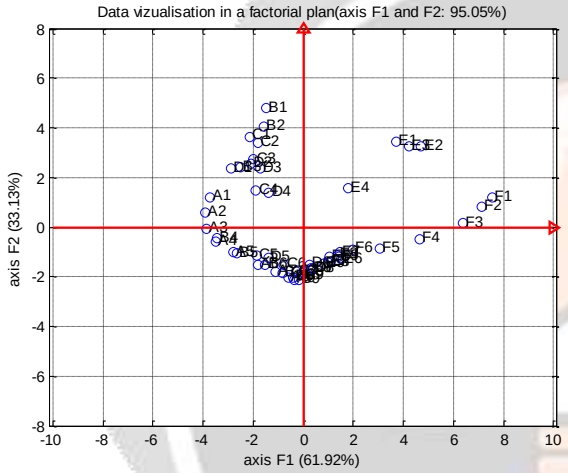

Fig-2-a Factorial representation of individuals.

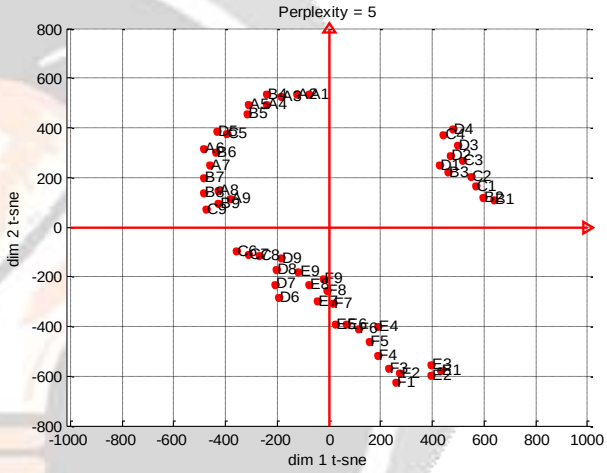

Fig-2-b Representation of individuals on the t-sne plan.

3.2 Results of grouping individuals by clustering

In the context of our study, the number of clusters is fixed at four. Indeed, within the framework of the Analysis in Principal Components, we apply the criterion of Kaiser [37] for the principal axis choice to be retained. Only the axis whose eigenvalues are greater than or equal to 1 are retained. As a result, we retain the first two main axes F_1 and F_2. By grouping the individuals according to the sign of the coordinates on the factorial plane, we have four zones: the first, a zone of weak precipitation in southern winter; the second, a zone of weak precipitation in southern summer; the third, a zone of heavy precipitation in southern winter and the last, a zone of heavy precipitation in southern summer.

- **Representation in the factorial plan and K-Means and Fuzzy C-Means**

Figures 3 show us the results of the representation of individuals on the factorial level with the different clustering methods. Figure 3-a shows us the result of clustering with k-Means. The Figures 3-b, 3-c and 3-d illustrate the results of clustering with the Fuzzy C-Means method with different values of the parameter m. For m = 1.1 and 1.2, notice the similarity of the results. From m = 1.3, points with fuzzy overlap have appeared. The parameter m used is m = 1.1.

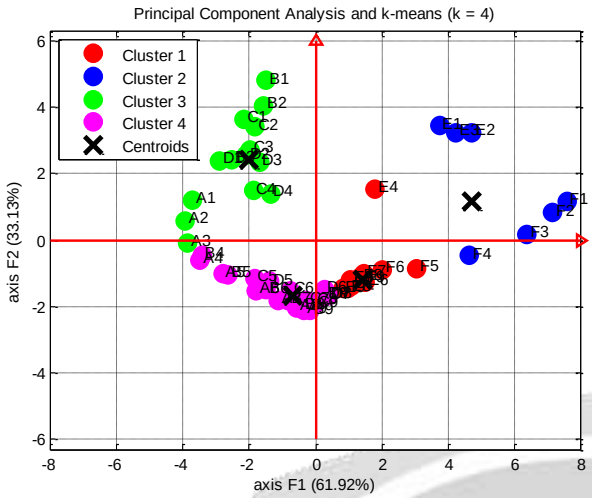

Fig-3-a: Representation of the individuals on the factorial plan and clustering by K-Means (k = 4).

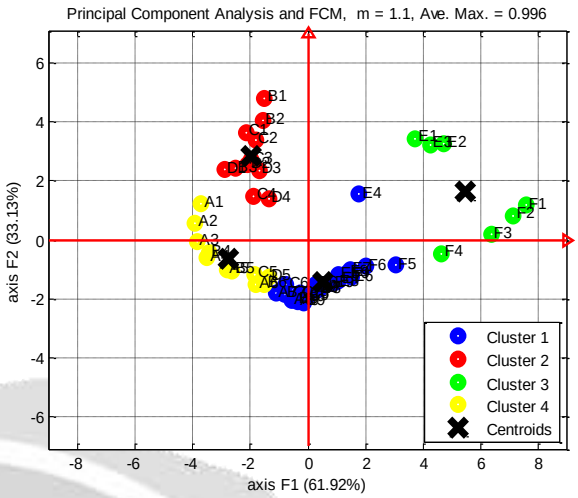

Fig-3-b: Representation of the individuals on the factorial plan and clustering by Fuzzy C-Means (m = 1.1; C = 4).

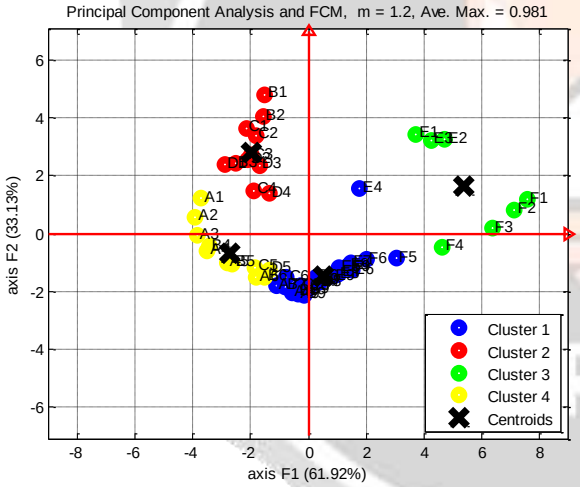

Fig-3-c: Representation of the individuals on the factorial plan and clustering by Fuzzy C-Means (m = 1.2; C = 4).

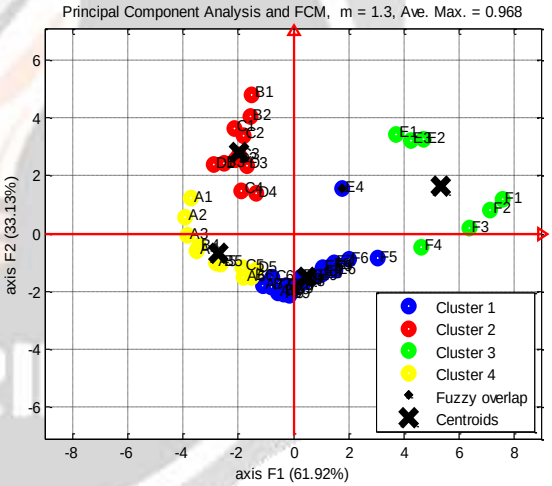

Fig-3-d: Representation of the individuals on the factorial plan and clustering by Fuzzy C-Means (m = 1.3; C = 4).

➢ **Representation on the plan t-sne (perplexity = 5) with K-Means and Fuzzy C-Means**

Figures 4 show the results of the representation of individuals on the t-sne plan with the different clustering methods. Keeping the same process, we obtain the same result whatever the method of clustering used.

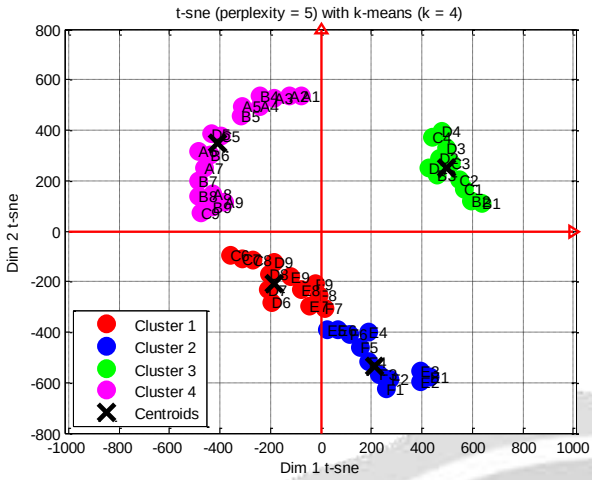

Fig-4-a: Representation of the individuals on the t-sne plan and clustering by K-Means (k = 4).

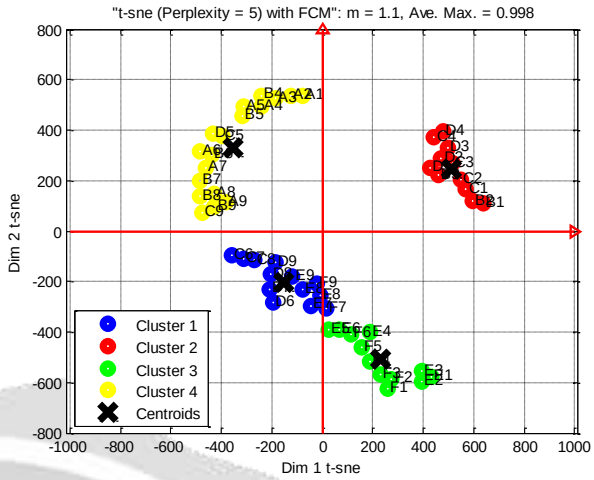

Fig-4-b: Representation of the individuals on the t-sne plan and clustering by Fuzzy C-Means (perplexity =5; m = 1.1; C = 4).

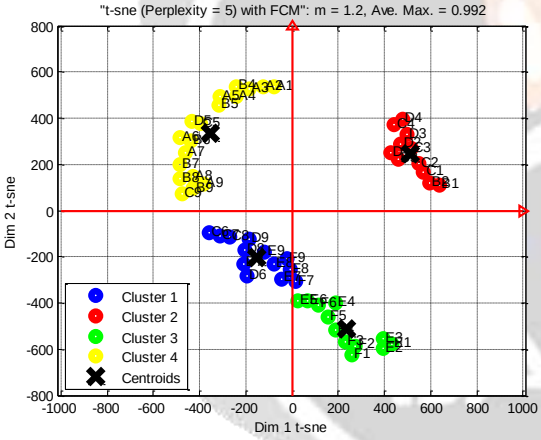

Fig-4-c: Representation of the individuals on the t-sne plan and clustering by Fuzzy C-Means (perplexity = 5; m = 1.2; C = 4).

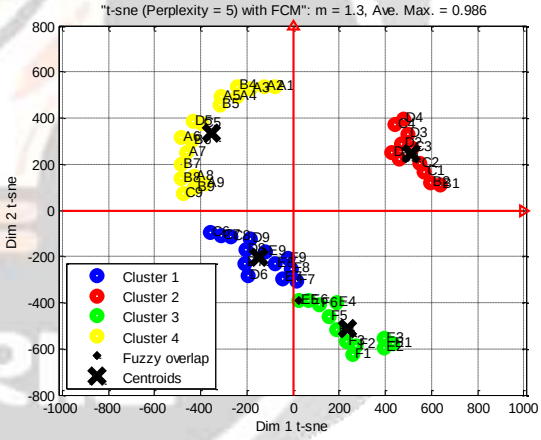

Fig-4-d: Representation of the individuals on the t-sne plan and clustering by Fuzzy C-Means (perplexity =5; m = 1.3; C = 4).

3.3 Classification assessment results

The following Table summarizes the Dunn index values for each classification. The best classification is the one with the highest Dunn index value. This is the visualization of data by t-sne with perplexity parameter = 5 accompanied by the Fuzzy C-Means clustering method (C = 4) with parameter m = 1.1 or K-Means (k = 4).

Table 2: Classification assessment results.

Data visualization method	Clustering method	Dunn index
Representation on the factorial plan	K-Means (k = 4)	0.08269694
Representation on the factorial plan	FCM (m =1.1 or 1.2; C = 4)	0.1391
Representation on the t-sne plan (perplexity = 5)	k-means (k = 4)	0.1411823
Representation on the t-sne plan (perplexity = 5)	FCM (m =1.1 or 1.2; C = 4)	0.1411823

3.4 Result of the spatial and temporal variability of precipitation in the South Madagascar

Figure 5 shows the result of the analysis of the spatial and temporal variability of precipitation in Southern of Madagascar. In yellow, we have zone 1 which corresponds to a zone of weak precipitation in Southern winter. In blue, we have zone 2 which corresponds to a zone of weak precipitation in Southern summer. In red, we have zone 3 which is a zone of heavy precipitation in Southern winter and in green, zone 4 which is a zone of heavy precipitation in Southern summer.

Whether in summer or austral winter, the precipitation gradient is always East-West.

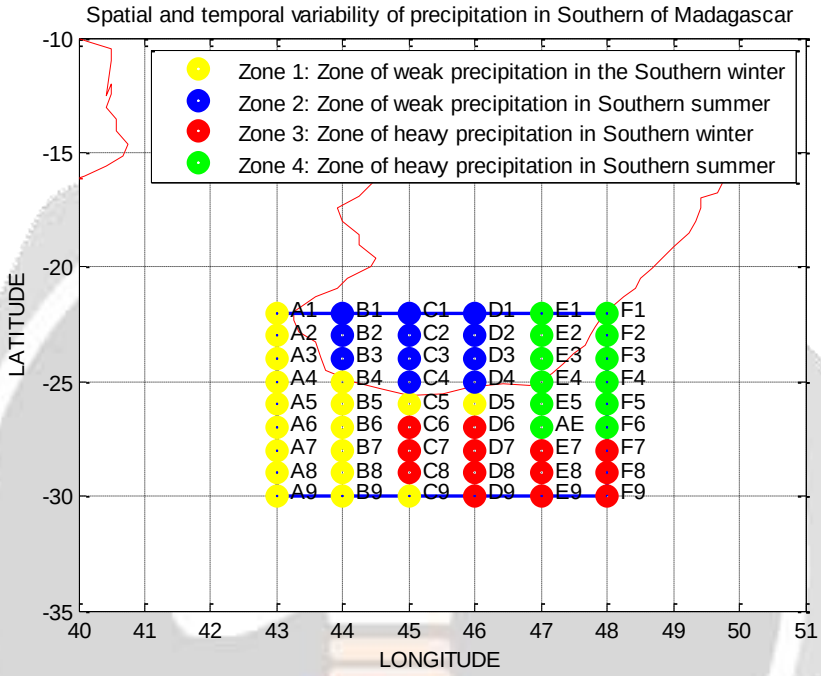

Fig-5: Spatial and temporal variability of precipitation in Southern of Madagascar.

4. DISCUSSIONS

The analysis of the spatial and temporal variability of precipitation is done in four stages: reduction of the size of the data, visualization of the data, clustering and evaluation of the classifications. Principal Component Analysis and t-sne are used for the reduction of the size of the data as well as the visualization of the data in a plan. The associated clustering methods are K-Means and Fuzzy C-Means. The Dunn index assesses each classification obtained.

Our results allow us to say that within the framework of this research, the new method of visualization of the data t-sne is much better than the representation in the factorial plane. Likewise, the Fuzzy C-Means clustering method is also more promising than the K-Means method. The new methods of data dimension reduction and clustering are therefore more interesting than the old ones.

Fuzzy C-Means and t-sne are generally used in machine learning. Our research has successfully applied these methods in the field of climatology.

Note that this research is limited to a fixed number of clusters (four clusters).

5. CONCLUSION

In conclusion, the importance of this research is to demonstrate that the new technics of t-sne data reduction dimension is more efficient than Principal Component Analysis. Likewise, the recent Fuzzy C-Means clustering method is also more interesting than K-Means. The analysis of the spatial and temporal variability of precipitation brings us back to the regionalization of the study area. We have four zones: zone 1 which corresponds

to a zone of weak precipitation in Southern winter. Zone 2 which corresponds to a zone of weak precipitation in Southern summer. Zone 3 which is a zone of heavy precipitation in Southern winter and zone 4 which is a zone of heavy precipitation in Southern summer. The results show that the precipitation in the South Madagascar presents an East-West gradient whether in summer or austral winter.

6. REFERENCES

[1] Ajar, D. (1982). Le problème de la détermination du nombre de facteurs en analyse factorielle. *Revue des sciences de l'éducation*, *8*(1), 45-62. Retrieved from https://www.erudit.org/fr/revues/rse/1982-v8-n1-rse3482/900356ar/.

[2] André, Mathilde. (2016). Variabilité spatio-temporelle des précipitations à La Réunion et impact sur la végétation. Retrieved from https://pdfs.semanticscholar.org/20d7/004ca25e3bdd828060eab3d079ea48a81e7b.pdf?_ga=2.208532631.40781815.1581088240-194701805.1581088240.

[3] Balafar, M. A. (2014). Fuzzy C-mean based brain MRI segmentation algorithms. *Artificial intelligence review*, *41*(3), 441-449. Retrieved from https://link.springer.com/article/10.1007/s10462-012-9318-2 .

[4] Bezdek, J. C., & Pal, N. R. (1995, November). Cluster validation with generalized Dunn's indices. In *Proceedings 1995 Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems* (pp. 190-193). IEEE. Retrieved from https://www.researchgate.net/publication/220886783_Cluster_Validation_with_Generalized_Dunn's_Indices/link/56e9d48808aec8bc078138fe/download.

[5] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2-3), 191-203. Retrieved from http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/RO/TemeProiecte/Biblio/FuzzyCMeans/FCM%20-%20The%20Fuzzy%20c-Means%20Clustering%20Algorithm.pdf .

[6] Bielders, P. C., Gérard, B., Bogaert, P. P., & Vanclooster, P. M.(2006).Caractérisation de la variabilité spatio-temporelle de la pluie au Fakara, Niger. Retrieved from http://docplayer.fr/18541499-Caracterisation-de-la-variabilite-spatio-temporelle-de-la-pluie-au-fakara-niger.html.

[7] Bourqui, M. (2008). Impact de la variabilité spatiale des pluies sur les performances des modèles hydrologiques. *PhD diss., Cemagref-Antony, Ecole Nationale du Génie Rural, des Eaux et des Forêts*. Retrieved from http://hydrologie.org/THE/BOURQUI.pdf.

[8] Chen, X., Chen, C., & Jin, L. (2011). Principal component analyses in anthropological genetics. *Advances in Anthropology*, *1*(02), 9. Retrieved from https://m.scirp.org/papers/8685.

[9] Dacosta, H., Kandia, K. Y., & Malou, R. A. Y. M. O. N. D. (2002). La variabilité spatio-temporelle des précipitations au Sénégal depuis un siècle. *IAHS PUBLICATION*, 499-506. Retrieved from https://iahs.info/uploads/dms/iahs_274_499.pdf .

[10] Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics.* ***3*** (3): 32–57. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/01969727308546046 .

[11] Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, *4*(1), 95-104. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/01969727408546059.

[12] ECMWF. (2019, December 29). ERA Interim, Daily. Retrieved from https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/.

[13] Ganga, S., & Meyyappan, T. (2014). Performance of Students Evaluation in Education Sector Using Clustering K-Means Algorithms. *International Journal of Computer Science and Mobile Computing*, *3*(7), 579-584. Retrieved from https://s3.amazonaws.com/academia.edu.documents/34287464/V3I7201499a11.pdf?response-content-disposition=inline%3B%20filename%3DPerformance_of_Students_Evaluation_in_Ed.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20200207%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20200207T161234Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=062787c2b395923887e1cfa48b97024beff9ea835c2d1bab97e14183afded7c8.

[14] Gokten, P. O., Baser, F., & Gokten, S. (2017). Using fuzzy c-means clustering algorithm in financial health scoring. *The Audit Financiar journal*, *15*(147), 385-385. Retrieved from http://revista.cafr.ro/temp/Article_9540.pdf.

[15] Havens, T. C., Bezdek, J. C., Keller, J. M., & Popescu, M. (2008, December). Dunn's cluster validity index as a contrast measure of VAT images. In *2008 19th international conference on pattern recognition* (pp. 1-4). IEEE. Retrieved from https://pages.mtu.edu/~thavens/papers/ICPR_2008_Havens.pdf.

[16] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, *24*(6), 417. Retrieved from https://psycnet.apa.org/record/1934-00645-001.

[17] Jackson, J. E. (1990). A User's Guide to Principal Components, chapter 3. Retrieved from https://books.google.fr/books?hl=fr&lr=&id=f9s6g6cmUTUC&oi=fnd&pg=PR7&dq=Jackson,+J.+E.+(1990).+A+User%E2%80%99s+Guide+to+Principal+Components,+chapter+3&ots=Lcs6cN0mS-&sig=PnoBZcKx-2oQi_4FszrAYkfsYcE#v=onepage&q=Jackson%2C%20J.%20E.%20(1990).%20A%20User%E2%80%99s%20Guide%20to%20Principal%20Components%2C%20chapter%203&f=false.

[18] Jolliffe, I. T. (2002). Principal components in regression analysis. *Principal Component Analysis*, 167-198. Retrieved from https://link.springer.com/chapter/10.1007%2F0-387-22440-8_8.

[19] Kobak, D., & Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nature communications*, *10*(1), 1-14. Retrieved from https://www.nature.com/articles/s41467-019-13056-x.pdf.

[20] L.J.P. van der Maaten. (2009). Learning a Parametric Embedding by Preserving Local Structure. In *Proceedings of the Twelfth International Conference on Artificial Intelligence & Statistics (AI-STATS), JMLR W&CP* 5:384-391. Retrieved from https://lvdmaaten.github.io/publications/papers/AISTATS_2009.pdf .

[21] Li, X., Lu, X., Tian, J., Gao, P., Kong, H., & Xu, G. (2009). Application of fuzzy c-means clustering in data analysis of metabolomics. *Analytical chemistry*, *81*(11), 4468-4475. Retrieved from https://pubs.acs.org/doi/abs/10.1021/ac900353t.

[22] Li, Y., & Wu, H. (2012). A clustering method based on K-means algorithm. *Physics Procedia*, *25*, 1104-1109. Retrieved from https://reader.elsevier.com/reader/sd/pii/S1875389212006220?token=0A016AAD2A40C9ADDC48C06F83348FD89D8D518418F1B49B6037C243D6E4BCF3386B68357C33910E9D4FE0A7912CC405.

[23] Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(Nov), 2579-2605. Retrieved from http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf .

[24] Mamadou, L. (2019, December 8). Réduction de dimension non linéaire. Retrieved from https://bookdown.org/content/2957/reduction-de-dimension-non-lineaire.html#.

[25] Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D. P., Saikhom, R., & Laishram, M. (2017). Multivariate statistical data analysis-principal component analysis (PCA). *Int J Liv Res. 2017c*, *7*(5), 60-78. Retrieved from

https://s3.amazonaws.com/academia.edu.documents/54658288/1_Multivariate_Statistical_Data_Analysis-_Principal_Component_Analysis_PCA.pdf?response-content-disposition=inline%3B%20filename%3DMultivariate_Statistical_Data_Analysis-P.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20200207%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20200207T155447Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=f53e55b018f37080d2a0c6b459db1b590a8eeb9354a9dc894c25c7ba3927fb2d.

[26] Nayak, J., Naik, B., & Behera, H. S. (2015). Fuzzy C-means (FCM) clustering algorithm: a decade review from 2000 to 2014. In *Computational intelligence in data mining-volume 2* (pp. 133-149). Springer, New Delhi. Retrieved from https://books.google.mg/books?id=PdupDwAAQBAJ&pg=PA747&lpg=PA747&dq=Nayak,+J.,+Naik,+B.,+%26+Behera,+H.+S.+(2015).+Fuzzy+C-means+(FCM)+clustering+algorithm:+a+decade+review+from+2000+to+2014.+In+Computational+intelligence+in+data+mining-volume+2+(pp.+133-149).+Springer,+New+Delhi.&source=bl&ots=iS-QaGarxT&sig=ACfU3U35_6ZbkeNgS09kcNcOFeqgOXgLEw&hl=fr&sa=X&ved=2ahUKEwjDqZWK67_nAhV3UBUIHWaLCe0Q6AEwAHoECAcQAQ#v=onepage&q=Nayak%2C%20J.%2C%20Naik%2C%20B.%2C%20%26%20Behera%2C%20H.%20S.%20(2015).%20Fuzzy%20C-means%20(FCM)%20clustering%20algorithm%3A%20a%20decade%20review%20from%202000%20to%202014.%20In%20Computational%20intelligence%20in%20data%20mining-volume%202%20(pp.%20133-149).%20Springer%2C%20New%20Delhi.&f=false.

[27] Niel, H., Leduc, C., & Dieulin, C. (2005). Caractérisation de la Variabilité Spatiale et Temporelle des Précipitations Annuelles sur le Bassin du Lac Tchad au Cours du 20ème Siècle/Spatial and temporal variability of annual rainfall in the Lake Chad Basin during the 20th century. *Hydrological sciences journal*, *50*(2). Retrieved from https://www.tandfonline.com/doi/pdf/10.1623/hysj.50.2.223.61800?needAccess=true .

[28] Öztürk, A., Lallich, S., & Darmont, J. (2018, May). A visual quality index for fuzzy c-means. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 546-555). Springer, Cham. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-92007-8_46 .

[29] Pakhira, M. K., Bandyopadhyay, S., & Maulik, U. (2004). Validity index for crisp and fuzzy clusters. *Pattern recognition*, *37*(3), 487-501. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0031320303002838?via%3Dihub .

[30] Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, *49*(4), 974-997. Retrieved from https://www.researchgate.net/publication/222533061_How_Many_Principal_Components_Stopping_Rules_for_Determining_the_Number_of_Non-Trivial_Axes_Revisited.

[31] Saley, M. B., Tanoh, R., Kouamé, K. F., Oga, M. S., Kouadio, B. H., Djagoua, E. V., & Savane, I. (2009, May). Variabiltié spatio-temporelle de la pluviométrie et son impact sur les ressources en eaux souterraines: cas du district d'Abidjan (sud de la Côte d'Ivoire). In *14e colloque International en évaluation environnementale, Niamey* (pp. 26-29). Retrieved from https://www.sifee.org/static/uploaded/Files/ressources/actes-des-colloques/niamey/simultanee-1/3_SALEY_comm.pdf .

[32] Trauwaert, E. (1988). On the meaning of Dunn's partition coefficient for fuzzy clusters. *Fuzzy sets and systems*, *25*(2), 217-242.Retrieved from https://www.sciencedirect.com/science/article/abs/pii/0165011488901893.

[33] UNICEF. (2019). Bulletin du monitoring de la sècheresse dans le grand Sud de Madagascar, bulletin n°13, (du 21 Octobre au 20 Novembre 2019). Retrieved from https://www.unicef.org/madagascar/media/2706/file/Bulletin%20WASH%20secheresse%20sud%20-%20Numero%2013%20-%20Decembre%202019.pdf.

[34] UNICEF. (2019, July 14). Water, sanitation and hygiene. Retrieved from https://www.unicef.org/madagascar/programme/wash.

[35] Van Der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research*, *15*(1), 3221-3245. Retrieved from http://jmlr.org/papers/volume15/vandermaaten14a/vandermaaten14a.pdf.

[36] Wang, C. J. (2015). A new integrated fuzzifier evaluation and selection (nifes) algorithm for fuzzy clustering. *Journal of Applied Mathematics and Physics*, *3*(07), 802. Retrieved from https://www.scirp.org/pdf/JAMP_2015063016233085.pdf .

[37] Z Ansari, MF Azeem, A Waseem, AV Babu. (2011). Quantitative evaluation of performance and validity indices for clustering the web navigational sessions, *World of Computer Science and Information Technology Journal 1* (5), 217-226. Retreived from https://www.researchgate.net/publication/268376916_Quantitative_Evaluation_of_Performance_and_Validity_Indices_for_Clustering_the_Web_Navigational_Sessions/link/562f491808ae22b170370c40/download .