REGIONALIZATION OF THE PRECIPATION REGIME IN THE SOUTH-EASTERN PART OF MADAGASCAR

N. RASOLOZAKA [1], Niry A. RAKOTOVAO [2], J. M. RABEHARISOA[3], Adolphe RATIARISON [4]

[1] *PhD student, Atmosphere, Climate and Oceans Dynamics Laboratory (DyACO), Physics and Applications, Sciences and Technologies, University of Antananarivo, Madagascar*
[2] *Maître de conférence, Atmosphere, Climate and Oceans Dynamics Laboratory (DyACO), Physics and Applications, Sciences and Technologies, University of Antananarivo Madagascar*
[3] *Maître de conférence, Atmosphere, Climate and Oceans Dynamics Laboratory (DyACO), Physics and Applications, Sciences and Technologies, University of Antananarivo Madagascar*
[4] *Professeur Titulaire, Atmosphere, Climate and Oceans Dynamics Laboratory (DyACO), Physics and Applications, Sciences and Technologies, University of Antananarivo Madagascar*

ABSTRACT

The amount of rain during a given climatological period, within a specific region, determines the amount of agricultural production and the biodiversity life on that area. To know more about the rainfall class assigned at each point in the region, efficient and valid scientific techniques must be implemented. The main objective of this study is to look for homogeneous rainfall areas in the southeastern part of Madagascar; at 21°S to 24°S of latitude and 46.5°E to 48.5°E of longitude. We've chosen to use two different classification methods that are applied to rainfall data in this region. The first method is to classify linearly the data by using Principal Component Analysis. The second method is to use KOHONEN's competitive artificial neural networks to perform this classification. A single factorial axis, representing 92.693% of the variance of the axes, was selected for the Principal Component Analysis, thus generating three subfields. In the case of the Kohonen networks, composed of 6 by 7 neurons and a hexagonal topology, three distinct subfields could also be identified. In each subfield there are distinct monthly rainfall patterns, and the distribution of this last depends on the topographic position as well as the vegetation of the region considered.

Keyword: *Precipitation, southeastern part of Madagascar, classification, Kohonen neural network, Principal Component Analysis, subfield.*

1. INTRODUCTION

Increasing natural disasters associated with climate change are a greater threat in developing countries than in other countries [1]. Madagascar is one of those countries where climate hazards affect the economy and the ecosystem each year. However, climate conditions ensure both the stability of human life and the terrestrial ecosystem life [2][3][4]. It is therefore essential to design a model for forecasting climate data in each region of the country.
However, dealing with data modeling in a defined geographical area requires a precise classification of these data in the study area to better distribute the areas with similar behaviors. The importance of classifying data according to a certain region remains in its ability to distinguish groups of points or individuals with similar characters. This can assist in decision-making for technical purposes such as the management of hydrological resources, the resolution of water resource problems, and the possible exploitation of water resources. Several studies in Africa focus on this regionalization part because of its importance in agriculture [5][6][7]. Different types of classification can be used in statistical data processing, but most of them are based on Principal Component Analysis [8]. In this article, we have

chosen to compare the results of the Principal Components Analysis with the neural network of KOHONEN, another approach which is the design of a self-organizing map to regionalize precipitation data in the South-East of Madagascar.

2. METHODOLOGIES

2.1. Databases and study area

The data used in this work are precipitation data from the European Climate Medium-Range Weather Forecast (ECMWF) site. There are 117 grid points that are located in the Atsimo-Atsinanana region of Madagascar (Figure 1), at 21 ° -24 ° S latitude and 46.5-48.5 ° E longitude. The Southeast region of Madagascar is mainly characterized by the variability of its altitude (Figure 2).

Fig-1: Map of Atsimo-Atsinanana region, Madagascar

The altimetry data shown in Fig-2 is from the WorldClim site (`www.worldclim.org/`). These are Digital Elevation Model (DEM) that will serve as a benchmark when we get our subfield.

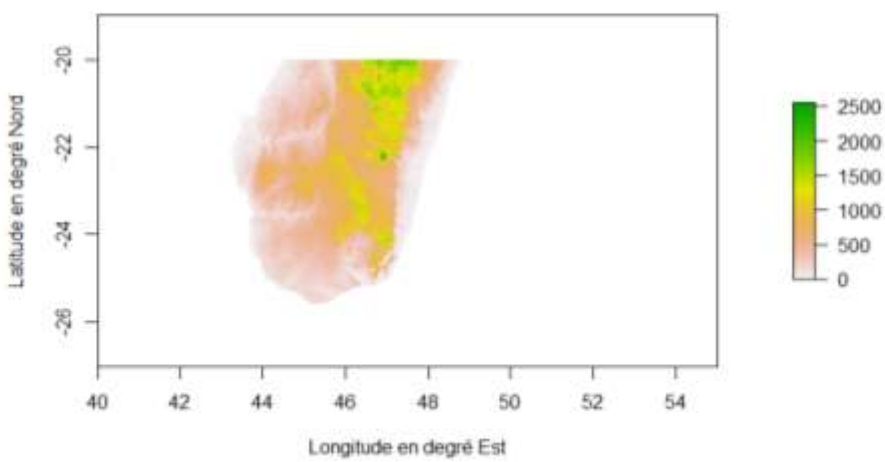

Fig-2: DEM of the Southern Part of Madagascar (DEM units are in m)

For the databases, classification is made using 12 observations which are the 39-years monthly climatological averages (from 1979 to 2017), and 117 variables to be classified which correspond to the grid points. It should be noted that the same data matrix was used for the two types of classifications that were adopted. A map representing these individuals is given by the following figure (Fig-3):

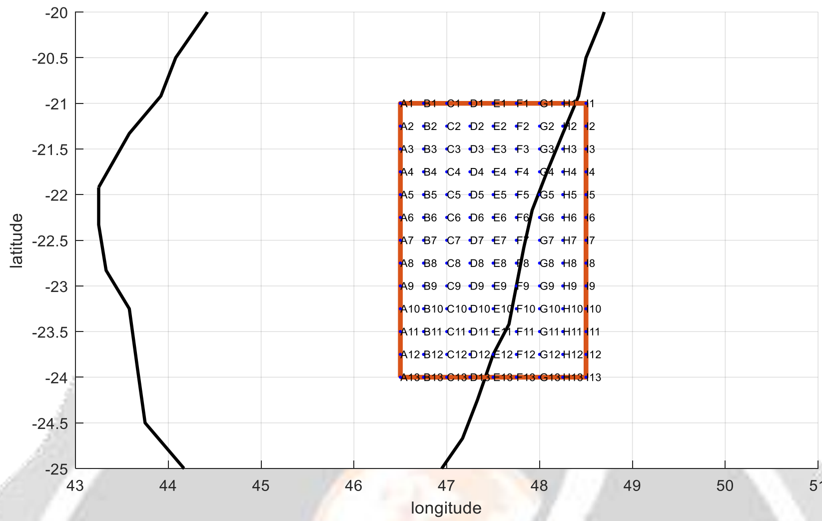

Fig-3: Individuals names on the study area

2.2. Principle Component Analysis

Principle Component Analysis (PCA) is a technique that reduces the size of multivariate data to better group variables of the same character. The data are represented by a matrix with *n* individuals and *p* variables [9][10] where *n* is inferior to *p*.
To get a first idea of the existing associations between the different variables, we use the correlation matrix. The peculiarities of the standard ACP come from the fact that the correlation matrix only has 1 on its main diagonal.
The eigenvalue of a principal component is equal to the sum of the squared correlation coefficients of each input variable with the component. The main components are classified by decreasing eigenvalue. The inertia explained by the i-th main component is associated with the i-th largest eigenvalue.
When working on reduced centered data, we retain the main components corresponding to eigenvalues higher than 1 (Kaiser criterion).
The variables X_j can be located on a sphere of radius 1 centered in 0, initial origin of the axes. The intersection of the sphere and a factorial plane is a circle called a correlation circle.

2.3. Self-organizing map of KOHONEN

Self-Organizing Maps (SOM), also known as the KOHONEN Neural Network, is a type of unsupervised network based on competitive learning [11]. This network classifies learning vectors as groups, each represented by an output neuron. The objective in this type of classification is to reduce intra-class distances and increase inter-class distances, so that elements in a class will be more similar to one another.
To create a SOM map, if we have different variables, we first go to normalizing each observation of each variable. The size of the map depends on the type of classification desired but the optimal dimension is given by the following formula:

$$\mathbf{M5\sqrt{N}}$$

Where M is the number of neurons in a map and N is the number of observations.
We have *n* entries in space, which will then be assigned to *p* classes (Fig-4).

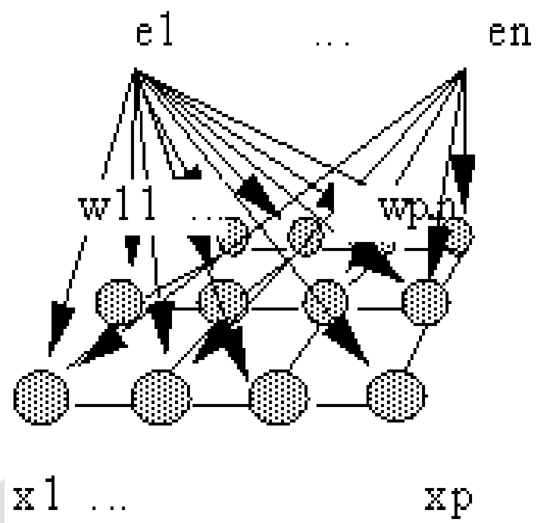

Fig-1. 2-dimensional SOM network with p^2 neurons and n inputs, where each neuron is connected to its 4 closest neighbors

In the interconnect layer, each neuron is strongly connected to its immediate neighbors (physically). The weight then decreases with distance. During the learning of the SOM, this following algorithm [12] is used:
• Initialization of weights to random values, neighborhood radius and step and update period.
• Repeat:
a) For each learning vector
- Calculate the distance squared of each of the neurons j with respect to:
- Select Neuron J which has the shortest distance between and:
- Adapt the weights of neuron J and all its neighbors in the current neighborhood radius: (equation à recopier)

$$w_{ji}(new) = \begin{cases} w_{ji}(prev.) + \alpha[x_i - w_{ji}] & \text{if j is in the neighborhood of J} \\ w_{ji}(prev.) & \text{else} \end{cases}$$

b) Modify the neighborhood radius downward if its update period is reached; change α as needed; as long as the performance is insufficient.

It would also be wise to calculate the average of individuals in each class to determine class characteristics [13]. In practice, the R software is used for learning. Note that the learning data must first be standardized before learning a self-organizing map. The size of the map should then be determined and, for each observation, the winning neuron should be known in the outcome [14].

3. RESULTS

To better understand the credibility of the output classes, we have represented the overall average of the rainfall covering our time interval (Fig-5).

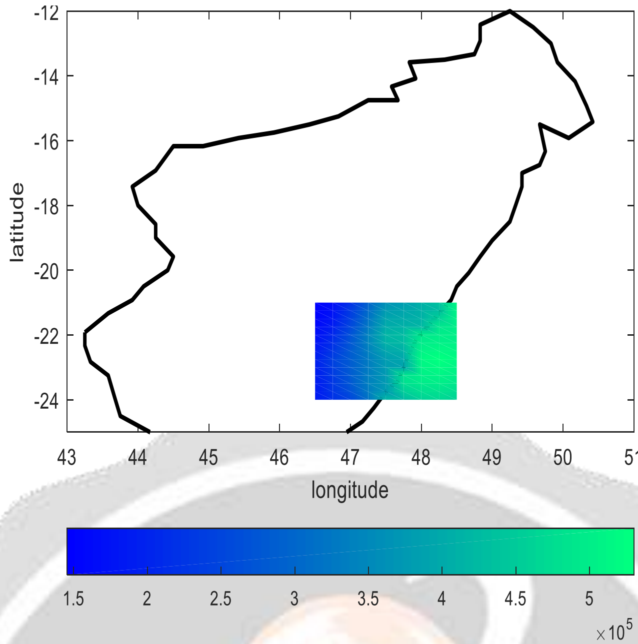

Fig-5: Spatial representation of precipitation data in the study area (precipitation is in mm)

3.1. PRINCIPLE COMPONENT ANALYSIS

The first step in establishing a PCA is to have a look at the explained variance of each principle component. The F1 factorial axis was been retained due to Kaiser criterion because it represents 92.693% of the variance of the factorial axis. Then, F2 was been retained because of the elbow criterion (Fig-6).

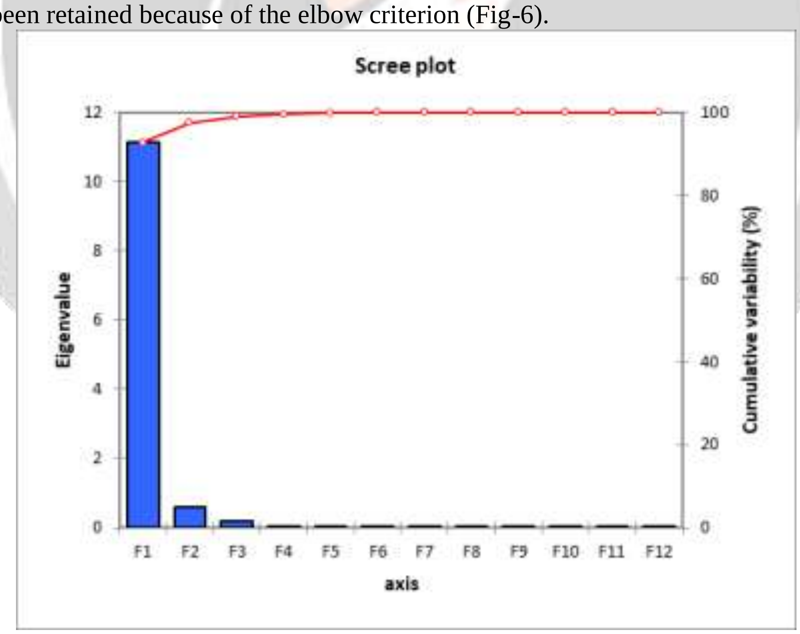

Fig-6 : explained variance of each principle components

Individuals group in the F1 and F2 axis and the correlation circle of the variables are illustrated by the following figure (Fig-7). According to the correlation circle (Fig-7), the monthly variables had been grouped into the F1 factorial axis.

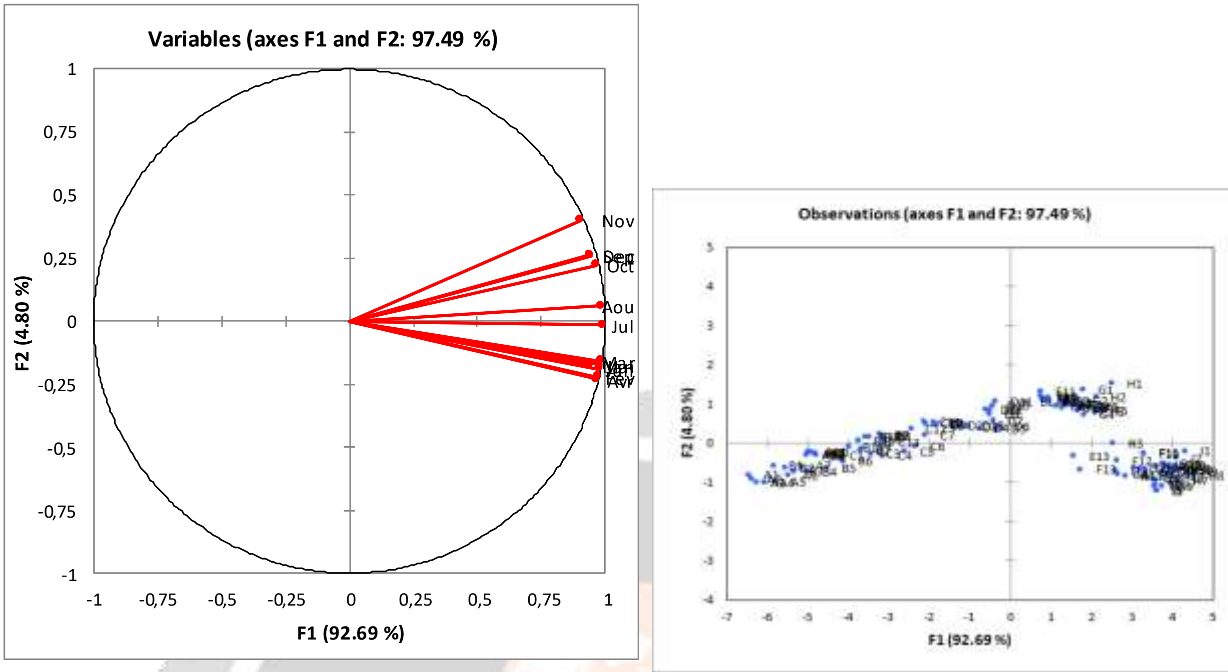

Fig-7 : correlation circle (on left case) - individuals representation in F1 and F2 factorial axis (on right case)

By visualizing those figures and the individual's coordinates, we got 3 distinct classes that we put on a geographical map (Fig-8).

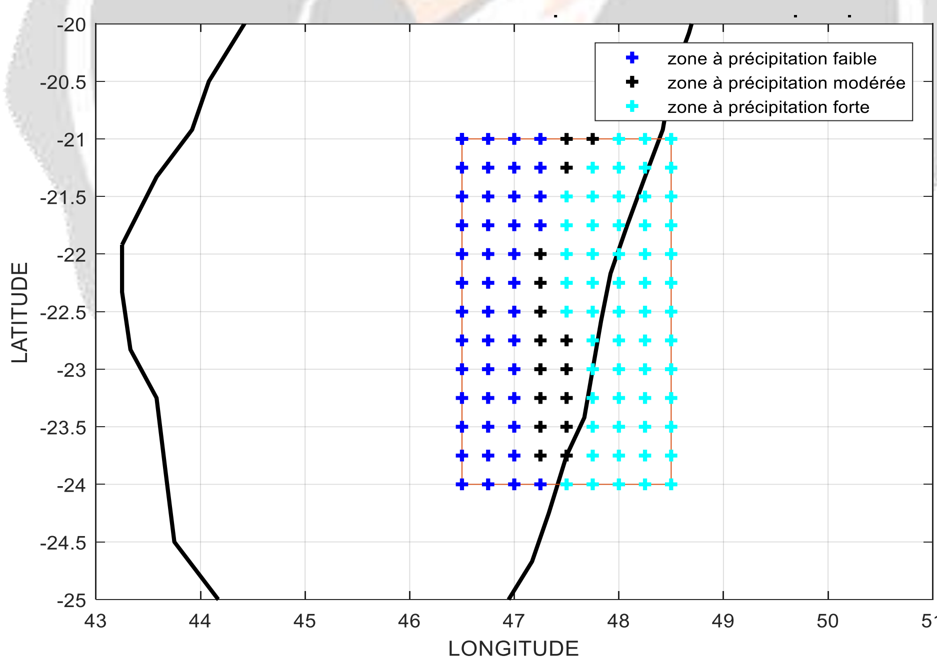

Fig-8: subfields after classification by ACP. Field with low precipitation in blue, field with non-classed or moderate precipitation in black, and field with high precipitation in cyan color

Rainfall quantity is increasing from east to west and as well as the elevation is decreasing (Fig-8, Fig-9).

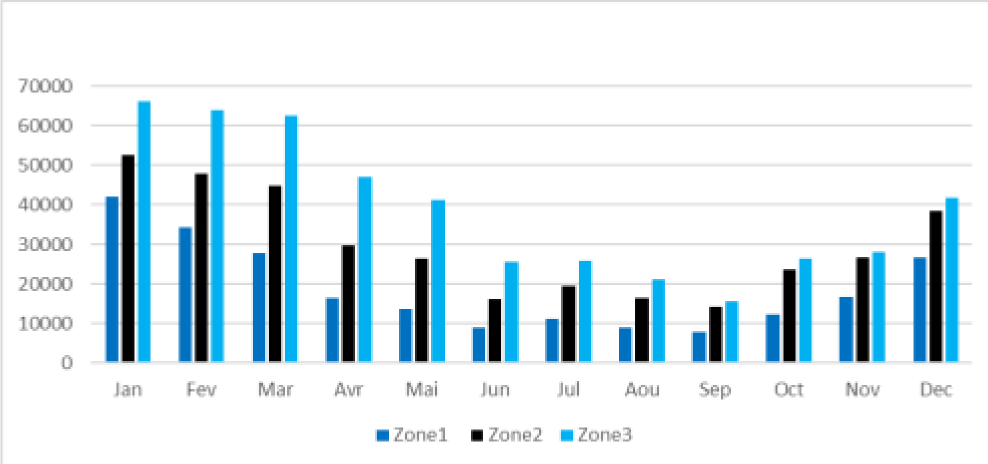

Fig-9: Monthly climatological average of precipitation for each subfield

3.2. Self-Organizing Map

Topology's choice: A network with 6 by 7 Neurons is implemented here. The training progress is given by the next figure (). Convergence is reached within 2500 iterations.

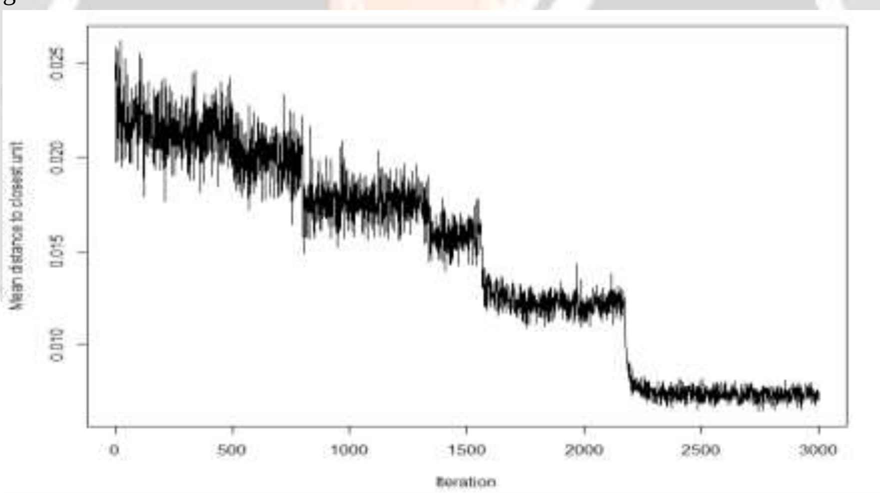

Fig-10: Training progress

The next figures show the classification results. We can see the individual repartitions (grid points) and the variables (month) in the neuron map.

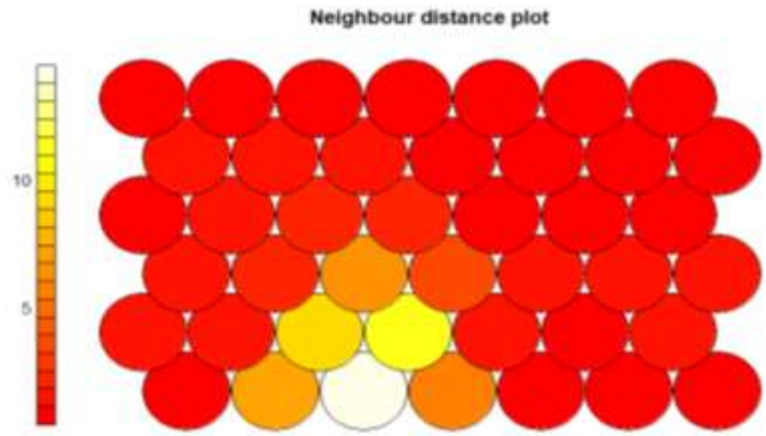

Fig-11 : Neighbour distance plot

Clear colors represent large distances, and darker colors represent neighboring neurons. The Best Matching Unit is shown by the Fig-10. The SOM quality is determined by the repartition of the inputs in the neuron topology. A SOM network is good when the variables are similarly distributed into the neuron map.

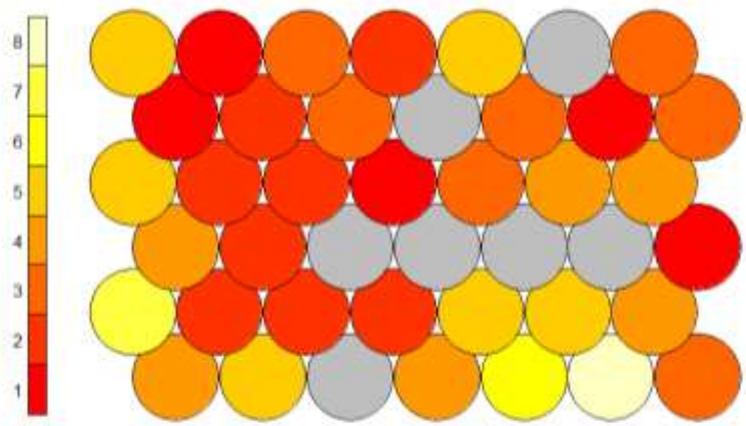

Fig-12 : counts plot

According to the previous figure, the data are similarly distributed into the map. So the map dimension fairly suits in our data.

With the hierarchical ascending classification, we could divide the neuron map into three clusters as shown in the next figure (Fig-13):

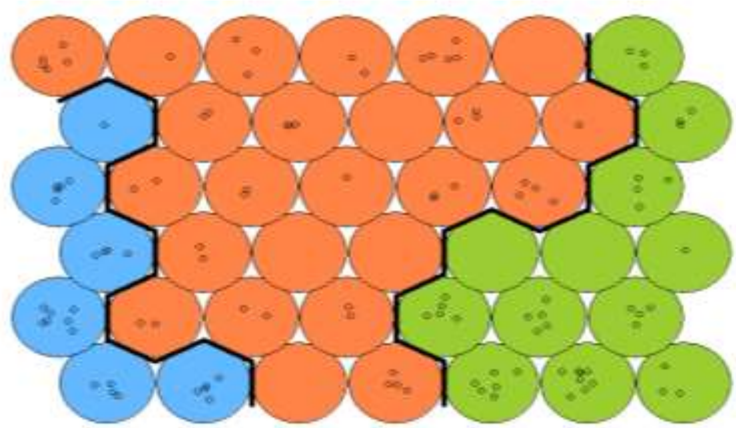

Fig-13 : Mapping plot

Little circles inside the neuron are the individuals earned by each neuron. According to the BMU (figure above), 3 clusters are affected to each individual: C1 in blue, C2 in orange, C3 in green.

In the next table, colored syntaxes every left column indicates the individual's names, and the numbers every right column are the number of the neuron. Note that the neuron number is read from left to right and down to up into the neuron map (Fig-13).

Table 1: SOM output

A1	2	**B1**	2	**C1**	10	**D1**	24	**E1**	25	**F1**	26	**G1**	42	**H1**	42	**I1**	7
A2	2	**B2**	2	**C2**	10	**D2**	24	**E2**	26	**F2**	26	**G2**	35	**H2**	42	**I2**	7
A3	2	**B3**	1	**C3**	16	**D3**	30	**E3**	27	**F3**	27	**G3**	35	**H3**	21	**I3**	7
A4	1	**B4**	9	**C4**	16	**D4**	30	**E4**	27	**F4**	27	**G4**	35	**H4**	13	**I4**	14
A5	1	**B5**	9	**C5**	23	**D5**	31	**E5**	28	**F5**	28	**G5**	13	**H5**	13	**I5**	14
A6	1	**B6**	15	**C6**	23	**D6**	31	**E6**	28	**F6**	28	**G6**	13	**H6**	13	**I6**	14
A7	8	**B7**	15	**C7**	36	**D7**	31	**E7**	33	**F7**	34	**G7**	6	**H7**	6	**I7**	14
A8	8	**B8**	22	**C8**	36	**D8**	38	**E8**	40	**F8**	33	**G8**	6	**H8**	6	**I8**	6
A9	8	**B9**	22	**C9**	36	**D9**	38	**E9**	40	**F9**	33	**G9**	6	**H9**	6	**I9**	6
A10	8	**B10**	22	**C10**	36	**D10**	39	**E10**	40	**F10**	12	**G10**	5	**H10**	5	**I10**	5
A11	8	**B11**	22	**C11**	36	**D11**	39	**E11**	40	**F11**	12	**G11**	5	**H11**	5	**I11**	5
A12	8	**B12**	15	**C12**	29	**D12**	38	**E12**	40	**F12**	4	**G12**	12	**H12**	12	**I12**	12
A13	8	**B13**	15	**C13**	22	**D13**	37	**E13**	11	**F13**	11	**G13**	4	**H13**	4	**I13**	4

The monthly climatological mean of the individual in each cluster are represented by the next figure (Fig-14).

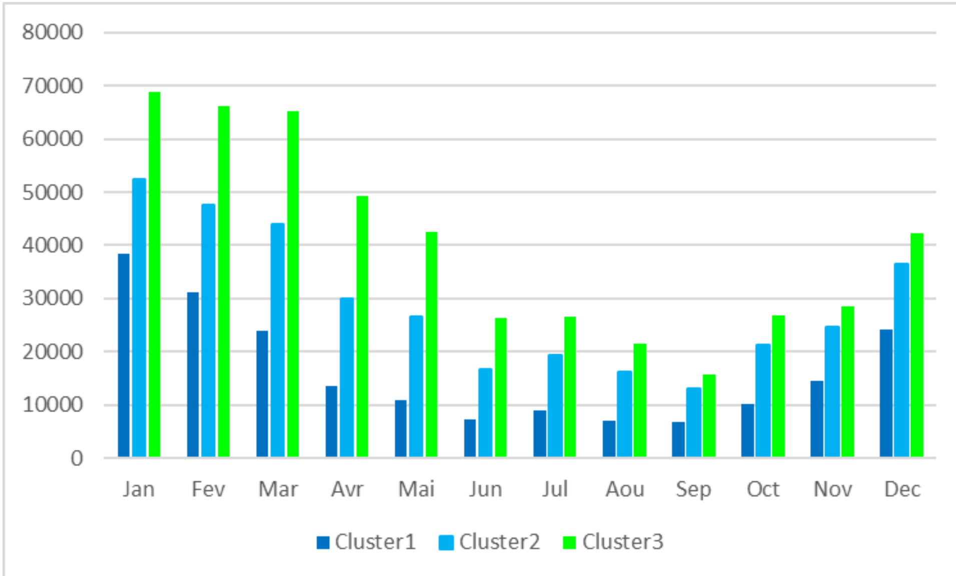

Fig-14 : Monthly climatological average for each cluster

We can then put into the geographical gridded map of Madagascar the individual in each cluster (Fig-14).

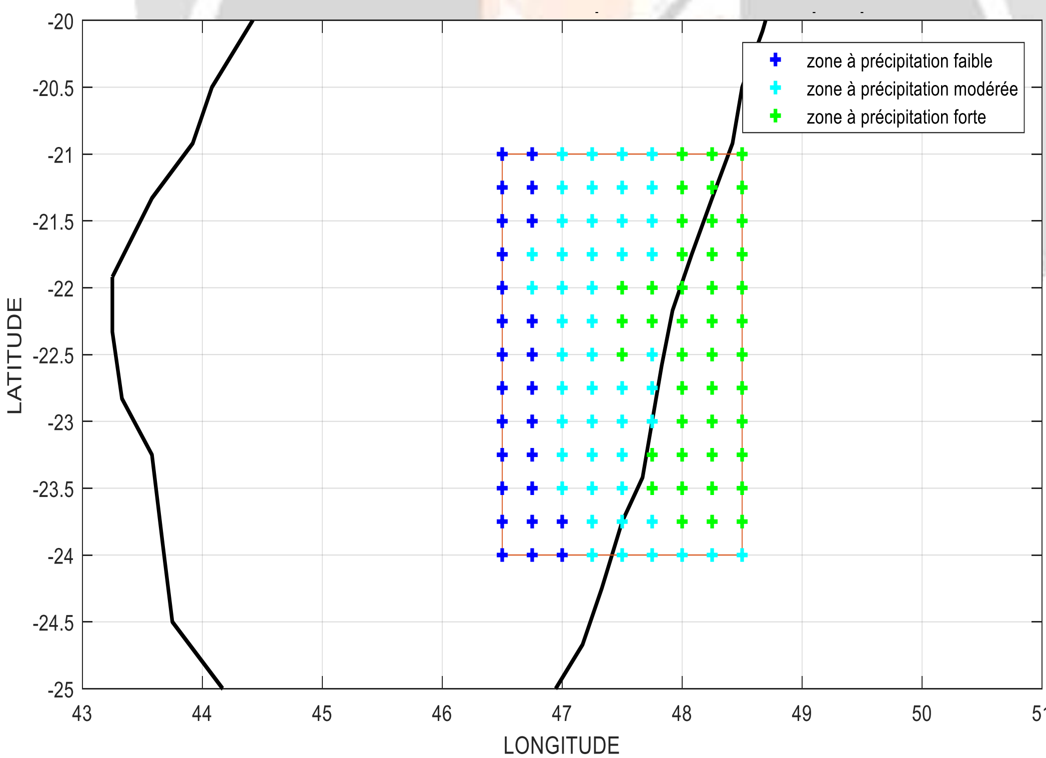

Fig-15 : Subfield by SOM regionalization

As with the Principle Component Analysis, the rainfall pattern is also clustered from east to west. We can remark that the variability caused by the elevation is more viewed by the Kohonen's map than by the PCA.

4. CONCLUSIONS

PCA and SOM are based in different principles. The first one, by reducing the data's dimension, just hold the factorial axis where the variance is highly represented. The individuals are grouped by their distances taking into account their position from the principle component. The second method is based on competitive neuronal network. Each neuron chooses one individual and the neighbor distances define the cluster. The cluster's repartition into the area are more concretely and uniformly determined by the SOM than by the PCA. That is justified by the non-classified grid points in the PCA and the

Regarding to the obtained classes, both for the SOM and for the ACP, the distribution of the homogeneous rainfall zones is affected by the variability of the topology in the extent of the study area. Indeed, it is obvious that precipitation is important in the littoral zones as in the highlands. In addition, the south-eastern region of Madagascar is known for its hyper variability in terms of altitude and its exceptional ecological character.

Once these areas of homogeneous precipitation behaviour have been identified, we can proceed to decision-making according to each subfield, for example to try to model the rainfall and to simulate a long-term forecast for each of them.

5. REFERENCES

[1]. N. Brooks and W. Neil, (January 2013). Country level risk measures of climate-related disaster natural disasters and implication for adaptation to climate change

[2]. Dunham et .al (2010). « Global climate cycles and cyclones: consequences for rainfall patterns and lemur reproduction in southeastern Madagascar"

[3]. Bouma, (2003). "Methodological problems and amendments to demonstrate effects of temperature on the epidemiology of malaria. A new perspective on the highland epidemics in Madagascar, 1972-89."

[4]. Xiaoxu Wu et al, (2015). "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation"

[5]. Frédéric ROSSEL et al., 1998, RÉGIONALISATION DE L'INFLUENCE DU EL NIÑO SUR LES PRÉCIPITATIONS DE L'ÉQUATEUR

[6]. Mustapha Touazi, et al., Régionalisation des débits moyens mensuels en Algérie du nord, Volume 24, numéro 2, 2011. Revue des sciences de l'eau. DOI : 10.7202/1006110ar.

[7]. SALAMA H., REGIONALISATION PLUVIOMETRIQUE DU BASSIN TENSIFT (MAROC) Larhyss Journal, ISSN 1112-3680, n° 09, Décembre 2011, pp. 111-119 © 2011 Faculté des Sciences, Université Hassan II-Aïn Chock, BP 5366 Mâarif, 20100 Casablanca.

[8]. GUIOT J.(1986). Sur la détermination des régions climatiques quasi homogènes, Revue des statistiques appliquées, tome 34, n°2, 15-34

[9]. *R. RAKOTOMALALA*, Analyse en Composantes Principales (ACP), Principes et pratique de l'ACP, Université Lumière Lyon 2, ***disponible sur: http://eric.univlyon2.fr/~ricco/cours/slides/ACP.pdf***

[10]. *C. Duby, S. Robin*, Analyse en Composantes Pricipales, *AgroParisTech. Disponible sur :* ***http://www2.agroparistech.fr/IMG/pdf/AnalyseComposantesPrincipales-AgroParisTech.pdf***

Center for Advanced Life Cycle Engineering (CALCE), University of Maryland, College Park, MD, 20742, U.S.A. EUROPEAN CONFERENCE OF THE PROGNOSTICS AND HEALTH MANAGEMENT SOCIETY

[11]. TEUVO *KOHONEN*, March *1990*, The Self-Organizing Map

[12]. Mia Louise Westerlund, April 24, 2005. Théorie et precessing du SOM. Classification with Kohonen Self-Organizing Maps. Soft Computing, Haskoli Islands.

[13]. Gökhan Akçap_nar, Arif Altun, Erdal Co_gun, Investigating Students' Interaction Profile in an Online Learning Environment with Clustering, 2014 IEEE 14th International Conference on Advanced Learning Technologies

[14]. Jing Tian, Michael H. Azarian, and Michael Pecht, 2014, Anomaly Detection Using Self-Organizing Maps-Based K-Nearest Neighbor Algorithm