9th International Workshop on Climate Informatics
October 02-04, 2019
Hosted by École Normale Supérieure, Paris, France

RAINFALL EVENT ANALYSIS IN THE NORTH OF TUNISIA USING THE SELF-ORGANIZING MAP

Derouiche Sabrine[1,2], Mallet Cécile[1], Bargaoui Zoubeida[2]

Abstract— **during the last decade, scientific research has shown a growing interest in quantifying the impact of climate variability on rainfall and water resources. In Tunisia, the variability of rainfall is at the origin of natural disasters extremely expensive in human lives and responsible for innumerable material damages. So, the analysis of rain events characteristics is an essential step for improving our understanding of spatial and temporal variation in precipitation. For this study, 70 raingauge stations in Northern Tunisia are used over 50 years (1959-2008). It is proposed to adopt a seasonal analysis (December-January- February) with a separation of daily rainall time series into events after the determination of the minimum inter event time MIT. This data transformation give us 6 rainfall variables: (1) rainfall event number, (2) total event duration, (3) Average precipitation, (4) total precipitation, (5) Average intensity, (6) Average duration. Those variables are clustered by using the method of Self Organizing Map SOM and give 4 type of seasons.**

I. MOTIVATION

The rainfall pattern in the Mediterranean is characterized by an important spatial and temporal variability. This variability is mainly due to its position (between 30°N and 45° N) which is directly influenced by subtropical high pressures and low mid-latitude pressures [1]. Tunisia, is located between the longitude 6-12 E and the latitude 30-38 N. This is a climatic transition zone between the temperate European domain, north of the Mediterranean and the southern African subtropical domain. Studies of rainfall patterns in Tunisia are often based on annual, seasonal or monthly rainfall averages [2], [3] and [4]. Since precipitation is an intermittent phenomenon that appears in the form of events, the study of the variability of events characteristics, proposed in this study, is well suited to the analysis of precipitation variability. The variability of the events of a single one season (December-January- February) is analysis. The DJF is considered the wet season in Tunisia.

[1] LATMOS/CNRS/UVSQ/Université Paris-Saclay, 11 boulevard d'Alembert, 78280 Guyancourt, France. [2]LMHE/ENIT/University of Tunis El Manar, Farhat HACHED EL MANAR BP 37, LE BELVEDERE 1002 Tunis, Tunisia.

II. DATA AND METHOD

This study is based on a daily rainfall database from 1959 to 2008 collected from the General Direction of Water Resources of the Ministry of Agriculture and Water Resources over 70 rain gauge stations distributed on the northern part of the Tunisian territory (Fig.1).

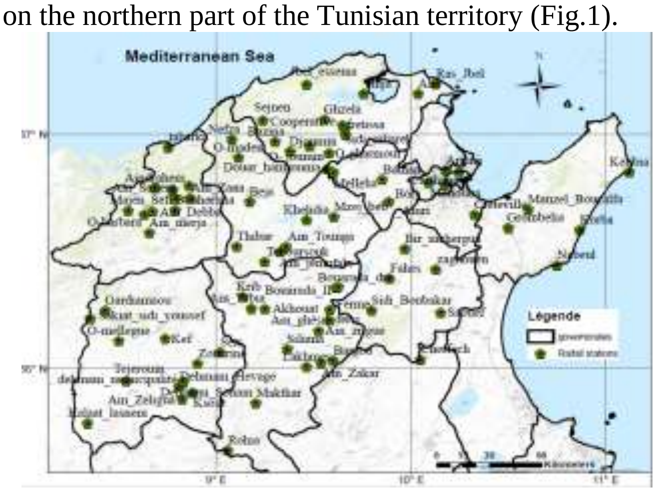

Fig.1: Rainfall stations distribution

The time series in rain gauge stations is broken down into a separate rain event by a dry period called Minimum Inter event Time MIT. In the literature, there are several statistical method to determinate the MIT such as the autocorrelation analysis and the coefficient of variation analysis [5] and [6]. The method adopted in this study is the autocorrelation analysis. This technic is based on the statistical independence of rain events [7]. Once the MIT is estimated. Two rain events are considered independent when the non-rain or the Inter Event Time IET in the time series is greater than the MIT. If it is smaller than the MIT the rain event is considered as a single event

Descriptive variables of precipitation events calculated for the same season for each station during 50 years are described in Tab.1.

Tab.1: Variables are characterizing rain events (DJF)

Nom	Symbol	Unit	Formula
Event number	EN	Events	
Total duration	TD	days	$TD = \sum_{i}^{NE} TD_i$
Precipitation	P	mm	
Average precipitation	MP	mm/event	$MP = \frac{P}{EN}$
Average duration	MD	days/event	$MD = \frac{TD}{EN}$
Average intensity	MI	mm/day	$MI = \frac{P}{TD}$

In order to classify the data, to analyze the link between those variables and to understand the structure of the rainfall variables the Self Organizing Map (SOM) is used. A SOM is an unsupervised learning algorithm based on artificial neural networks that produce a low-dimensional representation of a high-dimensional input dataset [8] and [9]. SOMs can be used for a variety of operations in exploratory data analysis, such as clustering, data compression, non-linear projection and pattern recognition. In this paper, we run the SOM tool in MATLAB using the SOM algorithm as described by Vesanto et al. [10]. The training of Kohonen map is done starting from the matrix of input data constituted of 3500 observations (50 years * 70 stations) and six variables described in (Tab.1). Many training have been performed for different SOM parameters in order to have a better quality of the map and minimal quantification and topological errors. SOM parameters finally obtained are described in (Tab.2). The learning of the map is very sensitive to neighborhood radius. If the radius is very small there is a risk of losing the data structure and also we can get big topological error. Also, the choice of the neuron number influences the results. If neurone number is very large or close to the number of initial data, there is a risk of overfitting but if it is small the quantification error will be important. It has some empirical techniques to introduce an optimal number of referent vectors [8]. Usually, the SOM is used combined with another classification technique especially the Hierarchical Agglomerative clustering HAC. This second classification is applied to the referents vectors trained by SOM algorithm to reduce the number of cluster, to allow a better understanding of the data and also to extract the information in a relevant way. The principle of HAC is the agglomeration of all data starting with individual elements. The algorithm of HAC is iterative, it merges at each stage the closest classes based on a criterion of similarity or dissimilarity (usually Euclidean distance). There are several agglomeration or linkage strategies used, the most known are single, complete, centroid, average, and Ward strategy [11]. In this study the approach of Ward is adopted .Ward's method minimizes the increase in total within-cluster sum of squared error. This increase is proportional to the squared Euclidean distance between cluster centers [12]

Tab.2: SOM parameters

parameters	
Neuron number	320
Neighborhood function	Gauss
Fine-tuning	
Epoch number T	5000
Initial and Final radius of training	[3 0.5]

III. Evaluation and Discussion

As explain previously we have firstly to define the right value of the MIT to avoid merging two independent events. The autocorrelation analysis is done for all stations. The 4 stations used in (Fig.3) are chosen randomly, they are distributed so that they cover the studied area. The correlograms (Fig.3) give the autocorrelation coefficient of daily rainfall data for the season (DJF) over 50 years from 1959 to 2008 with the approximate 95%-confidence intervals for a white noise process. Outside the blue lines, the autocorrelation coefficient are considered significant.

Generally, for a four days lag or more the correlations are no longer significant. The significance of autocorrelation values with a three day lag differs from rain gauge station to another. The coefficient for a lag of 2 days is usually small (less than 0.1) and significant. For a 1 day step, the autocorrelation values vary between 0.2 and 0.45 (a high coefficient). This statistical study showed that the daily rainfall time series are decorrelated after two days. The six variables (Tab.1) corresponding to each season are thus computed in using a MIT equal to two days to define the events.

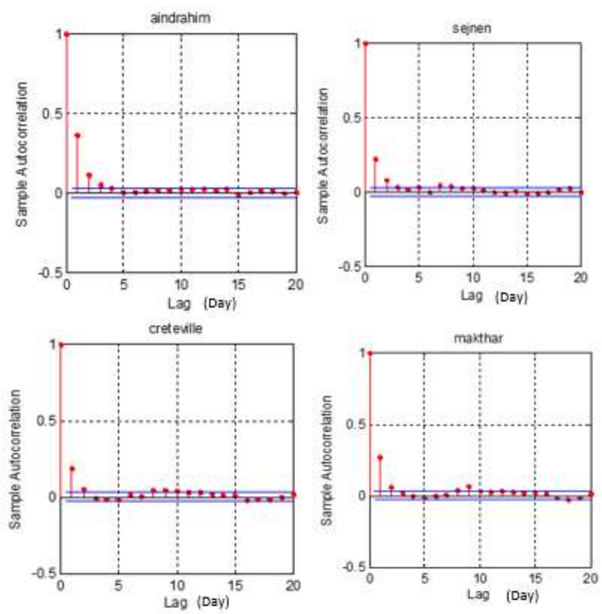

Fig.3: Autocorrelation Analysis

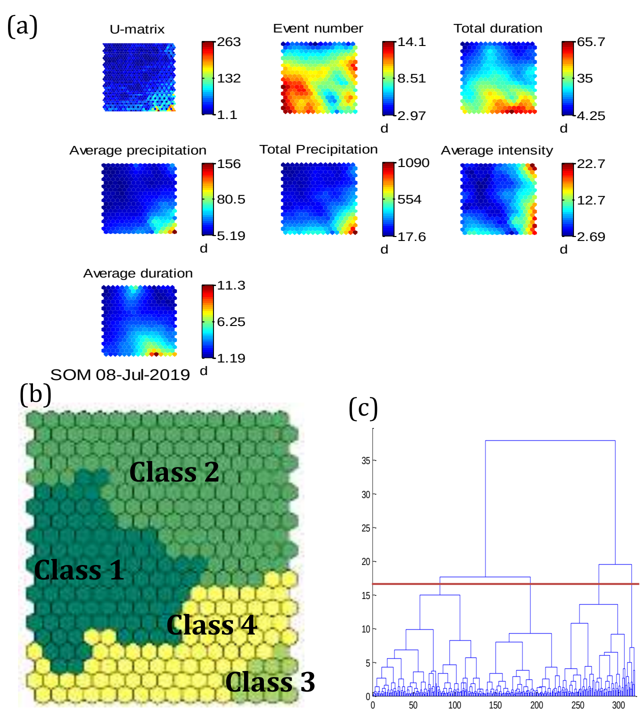

Fig.4: (a) Variables projection in topological Map. (b) Classes delimitation in topological map. (c) Dendrogram

The visualization in space of the topological map (Fig. 4a) of the values of the rainfall variables, corresponding to each neurone, makes it possible to study their relations. The U matrix in (Fig.4a) is a very important visualization for the interpretation of structures of the data. It presents the distance (the similarity) between neurons pairwise. It allowed us to distinguish that in the lower right of the map there is a group of observation less dense, which are very dissimilar compared to other data and which represent extreme situations. The outcome of the HAC is called a dendrogram (Fig.4c). This visualization shows the cluster and sub-cluster relationships and the order in which the clusters were consolidated. The closeness of the clusters can be depicted by lengths of the limbs, and the data items can be clustered by cutting the dendrogram (Fig.4c). So, we can group the corresponding seasons into four clusters and the difference between clusters is visualized in the distribution of (Fig.5). Using the delimitation of clusters in topological map (Fig.4b) and the observation the of the variables projection (Fig.4a) the four clusters are characterized.

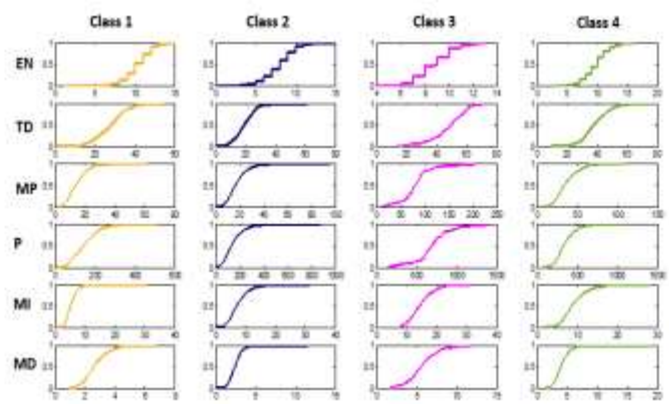

Fig.5: Empirical distribution of classes for the 6 rainfall variables

Class 1 is characterized by a low amount of precipitation (the total of the season, the average of events, or the average per day are weak). However, the event number and the total duration is strong. The class 1 represents the dry seasons with intermittent low rainfall. Class 2 is characterized by a weak values of all variables compared to other classes. So, this class represent the very dry seasons with very low amount of precipitation and very low number of event or rainy days. Class 3 groups very exceptional seasons, characterized by some strong and long events, with an important quantity of rainfall and a strong intensity. This class represent the very wet seasons with extreme events. Class 4 is characterized by a good quantity of precipitation but less than the class 3, also the mean intensity and the total duration is also strong. But the event number is the most important compared to other clusters. This class represent the wet seasons with intermittency of precipitation in the season.

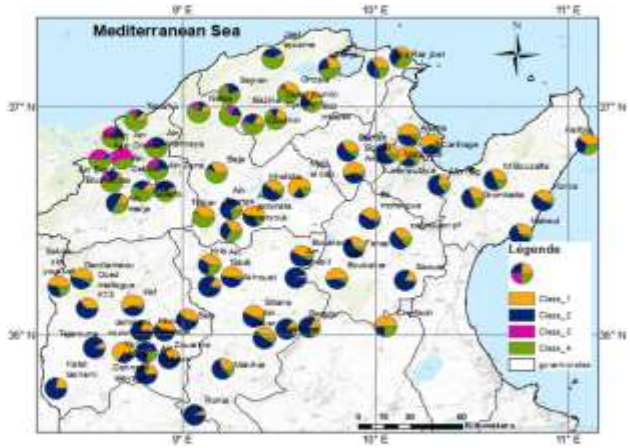

Fig.6: Spatial distribution of classes

The (Fig.6) shows the percentages of each class for each rain gauge station over 50 years. It shows that the rainy seasons, classes 3 and 4, are generally located in the northern part of the region (pink and green colors) and the dry seasons, classes 1 and 2 (yellow and blue colors), are located in the southern part of the study area. The wet area is directly influenced by North West flux coming from the Atlantic during the winter season. The exceptional stations with pink color are located in a forest area. In the south of the studied area the dry seasons (blue) dominates and this region is known by a warm desert climate. For the temporal distribution of classes, the (Fig.7) represent the percentage of stations in each cluster over the years.

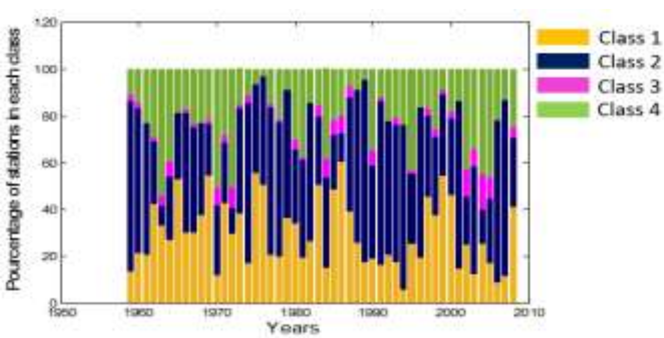

Fig.7: Annual distribution of classes

This figure allows to identify some exceptional seasons like 1963, 1970, 1973, 2002, 2003, 2004 and 2005 seasons where the class 3 and 4 dominate. We observe also the predominance of class 1 and 2 for more than 90 % of stations in 1975, 1976, 1988 and 1989 seasons.

The SOM combined with the HAC clustering for 6 rainfall descriptors allow to identify 4 rainfall situations in DJF seasons. The frequency of each class differ from a rain gauge station to another (Fig.6) and depends on the topography of the area, the proximity to the sea, local meteorological conditions and Mediterranean and Atlantic flux.

The definition of rain event from daily rain gauge data allow the analysis of event characteristics, and the intermittence of precipitation that can be useful to study the teleconnection with global atmospheric circulation and also the variables extracted from event can be used to study various topics like meteorology and hydrology. This classification may be useful in agriculture field, to predict the production especially the cereal and olive.

REFERENCES

[1] J-F Rysman, S. Verrier, Y. Lemaître, E. Moreau. “Space-time variability of the rainfall over the western Mediterranean region: A statistical analysis”. Journal of Geophysical Research: Atmospheres, *American Geophysical Union*, 118 (15), pp.8448-8459, 2013.

[2] A.Merzougui and M.Slimani, ‘’Régionalisation des lois de distribution des pluies mensuelles en Tunisie’’, *Hydrological Sciences Journal*, 57:4, 668-685, DOI: 10.1080/02626667.2012.670702, 2011.

[3] A.Douguedroit. ‘’Precipitation in Tunisia (1951-1980)’’. In: Méditerranée, troisième série, tome 66. Recherches climatiques en régions méditerranéennes II, pp. 23-33, 1988.

[4] L.Hénia, ‘’Les précipitations pluvieuses dans la Tunisie tellienne’’. Publ. de l’Université de Tunis, Deuxième série Géographie, vol. 14, 1980.

[5] A. Sharad Parchure, S. Kumar Gedam. “Precipitation Regionalization Using Self-Organizing Maps for Mumbai City, India”. *Journal of Water Resource and Protection, 10, 939-956.* DOI: 10.4236/jwarp.2018.109055,2018.

[6] J.G. Joo, J.H. Lee, H.D Jun, J.H .Kim, D.J. Jo.” Inter-Event Time Definition Setting Procedure for Urban Drainage Systems”. *Water*, 6, 45-58. DOI: 10.3390/w6010045, 2014.

[7] N. Akrour, A. Chazottes, S .Verrier, C. Mallet and L.Barthes . Simulation of yearly rainfall time series at microscale resolution with actual properties: Intermittency, scale invariance, and rainfall distribution, Water Resour. Res., 51, 7417– 7435, doi:10.1002/2014WR016357, 2015.

[8] T.Kohonen, ‘’Self-Organizing Maps’’. Third Edition, Springer, Berlin. https://doi.org/10.1007/978-3-642-97610-0, 1995.

[9] T.Kohonen, Essentials of the Self-Organizing Map. Neural Networks,37,5265.https://doi.org/10.1016/j.neunet.2012.09.01,2013.

[10] J.Vesanto, J.Himberg, E.Alhoniemi and J.Parhankagas. SOM Toolbox for Matlab 5, Report A57. http://www.cis.hut.fi/projects/somtoolbox/, 2000.

[11] F.Murtagh, P. Contreras, Methods of hierarchical clustering. Comput.Res.Repository.abs/1105.0121(2011).http://arxiv.org/abs/1105.0121. 2011

[12] Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? Journal of classification31, 274–295, 2014.