



HAL
open science

Characterisation of rainfall events in northern Tunisia using self-organising maps

Sabrina Derouiche, Cécile Mallet, Abdelwaheb Hannachi, Zoubeida Bargaoui

► **To cite this version:**

Sabrina Derouiche, Cécile Mallet, Abdelwaheb Hannachi, Zoubeida Bargaoui. Characterisation of rainfall events in northern Tunisia using self-organising maps. *Journal of Hydrology: Regional Studies*, 2022, 42 (August), pp.101159. 10.1016/j.ejrh.2022.101159 . insu-03741124

HAL Id: insu-03741124

<https://insu.hal.science/insu-03741124>

Submitted on 31 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Characterisation of rainfall events in northern Tunisia using self-organising maps

Sabrina Derouiche^{a,b}, Cécile Mallet^{b,*}, Abdelwaheb Hannachi^c, Zoubeida Bargaoui^a

^a Tunis El Manar University (UTM), National Engineering School of Tunis (ENIT), Laboratoire de Modélisation en Hydraulique et Environnement (LMHE), Le Belvédère, 1002 Tunis, Tunisia

^b LATMOS/IPSL, UVSQ, Paris-Saclay University, Sorbonne University, CNRS, 11 boulevard d'Alembert, 78280 Guyancourt, France

^c Department of Meteorology (MISU), Stockholm University, 106 91 Stockholm, Sweden

ARTICLE INFO

Keywords:

Rain event
Self-Organizing Map
Minimum Inter-Event Time
Hierarchical agglomerative clustering
Data image

ABSTRACT

Study region: The study is carried out for northern Tunisia.

Study focus: Precipitations are often analysed via intensity or accumulation for a specified time-scale (e.g., annual, seasonal, etc). We propose in this study to analyse regional rainfall variability by adopting a variable time step through the rain event concept. This event-based approach, ensures the integration of information related to rain intermittency, which is one of the fundamental properties of precipitations. This study focuses essentially on wet spells characteristics derived from the aggregation of daily winter dataset over a 50 years period (1960–2009). The multivariate analysis, based on the combination of two clustering approaches, i.e., self-organizing map and hierarchical clustering, allows the identification of different rainfall regimes. This study helps to understand rainfall variability patterns and to address rainfall regionalization and water use management issues.

New hydrological insights for the region: The winter precipitations of northern Tunisia are classified into 4 typical situations: Extremely dry seasons with a few short and weak rainfall events, dry seasons, with high frequency of weak events, intermediate seasons with medium amount of rain and intermittent events and rainiest seasons with long and intense events. The regionalization yields two geographical regions: northern sector characterized by rainy seasons, whereas the stations of the southern sector are mostly dry. The temporal variability analysis shows that the dry season classes dominate extending over three consecutive decades from 1970 to 2000.

1. Introduction

The precipitation pattern in the Mediterranean area is characterized by high spatial and temporal variability. This variability, prominent between 30°N and 45°N, is directly controlled by the subtropical high and mid-latitude low-pressure systems (Rysman et al., 2013). These are at the origin of extreme events, e.g., drought or extreme precipitation, and can result in natural disasters, such as loss of human lives and damage to infrastructure, livestock and environment. Tunisia is situated in the northern tip of Africa and extends between latitudes 30–38°N. This latitude band is considered as a climatic transition zone between the temperate mid-latitude

Abbreviations: SOM, Self Organizing Map; HAC, Hierarchical Agglomerative Clustering; MIET, Minimum Inter Event Time; DJF, December January February.

* Corresponding author.

E-mail address: cecile.mallet@latmos.ipsl.fr (C. Mallet).

<https://doi.org/10.1016/j.ejrh.2022.101159>

Received 22 March 2022; Received in revised form 23 June 2022; Accepted 26 June 2022

Available online 2 July 2022

2214-5818/© 2022 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

zone, north of the Mediterranean, and the North African subtropical zone (Slimani et al., 2007). In addition to its latitudinal position, the particular geography of the area, including the topography and the proximity to the Mediterranean Sea, affect the atmospheric circulation and can accentuate the rainfall variability in this region (Jebari et al., 2007).

With the observed current climate change and the increase in water needs, there's a growing concern about the possibility of an increase in the likelihood of drought and/or floods and their increasing and amplifying impacts (Narrant and Douguedroit, 2006; Xoplaki et al., 2006; Trambly et al., 2013, 2020; Schilling et al., 2020). Analysing rainfall variability in Tunisia, potentially exposed to climate variability and change, has been one of the most challenging subjects for hydrologists over the past 30 years. Kebaili-Bargaoui (1989) and Bargaoui (1994) analysed annual rainfall series of the Medjerda basin using the surplus-deficit method (runs) and frequency analysis by means of the Weibull distribution in order to identify drought years. Drought occurrence in Tunisia over a period of 88 years was also investigated using annual rainfall series by Benzarti and Habaieb (2001). Spatial regionalization of monthly precipitation in northern Tunisia was analysed based on the station's topography indices (Jedidi et al., 2008) and univariate frequency analysis (Benzarti, 2003; Merzougui and Slimani, 2012). Trambly et al. (2019) analysed the variability of rainfall on five days resolution in Tunisia with different interpolation methods. A time-dependent two-state Markov chain was calibrated for Grand Tunis region by Bargaoui (1983). Using copula approach, Gargouri-Ellouze and Bargaoui (2009) analysed 5-minute resolution rainfall series to study the relationship between the intensity of the rainfall event and the hydrological response of the watershed as represented by infiltration index. Ghanmi et al., (2013, 2016) analysed rain series with high temporal resolution (5 min) to build intensity-duration-frequency curves for several stations distributed in the north of Tunisia using the multifractal approach.

In this paper, we focus on Northern Tunisia. A large data set of rain gauge measurements, from a relatively large network (70 rain gauges in an area of about 28,000 km²), provides daily accumulation series over sufficiently long periods (50 years) for the proposed study. Daily rainfall data are aggregated into rainy and no rainy spells during a given season. This event-based approach, with a variable time step (or the event duration), ensures the integration of information related to the intermittency of precipitating processes. In the other side, a given monthly or seasonal accumulation may correspond to a multitude of low intensity events or to a single very intense event. Most of regional studies, mentioned earlier, based on rain gauge network observations, provide hourly, daily, monthly, yearly statistics with an estimation of Intensity-duration-frequency curves (IDF) or Intensity-Duration-Area Frequency curves (IDAF). The choice of the time step is a critical point. Indeed, precipitations are characterized by a high variability and intermittency in temporal or spatial domain (Verrier et al., 2011; Akrouf et al., 2015). When large time step is considered, the distinction between the rainy period and the dry period is not possible and the cumulative precipitation values are strongly smoothed and do not reflect the natural variability of rainfall. On the other hand, a too fast sampling rate leads to sampling errors especially when using tipping bucket rain gauges (The bucket volume of the tipping bucket rain gauge corresponds to an equivalent rain height ($h=0.2$ or $h=0.5$ mm)).

Overall, precipitations have a complex structure over a range of time and space scales that need to be equally highlighted. At daily temporal resolution, the day-to-day rainfall variability is characterized by dry and wet periods possessing different properties. The extraction of rainfall event properties (e.g., frequency, duration and the intensity) ensures a better characterization of the rainy season by taking into account additional rainfall aspects equally important rather than the amounts of precipitation. For example, the rainfall event number over a given season reflects the intermittent character which represents a fundamental property of precipitation. In addition, the average duration and average accumulation of rainfall events over the rainy season indicate the behavior of wet spells.

Understanding the variability of rainfall event features is crucial for effective management of water resources especially when the economy of the region is highly dependent on agriculture production such as Tunisia. For the rainfed agriculture and depending to the season, the rain event characteristics (number, intensity and duration) may play an important role in determining the quality and quantity of crops. The importance of event-based rainfall analysis is not limited to the agriculture; however, the identification of the season typology and the characterization of rainfall event can have a direct influence on runoff processes, horizontal flows, infiltration and soil humidity. Moreover, the proposed study may contribute in a future step to the establishment of a synoptic hydro-climatic reference.

The manuscript is organized as follows. Section 2 provides a description of the study area and rain stations characteristics. Section 3 discusses the methodology used for clustering analysis. Section 4 provides results followed by the discussion in Section 5, and a conclusion is given in the last section.

2. Study area and dataset description

Northern Tunisia is located between 35°N and 38°N and covers a total area of about 28,641 km² (Fig. 1). It is characterized by dry, hot summers and moist, cool winters. We adopt Emberger bioclimatic index (Emberger, 1930) which is specific to bioclimatic classification in the Mediterranean region and is based on the total annual rainfall as well as the daily average maximum temperature of the hottest month and the daily average minimum temperature of the coldest month (Mares, 1972). Accordingly, northern Tunisia is divided into three major bioclimatic regions: humid (H) in the extreme northwest, sub-humid (SH) and semi-arid (SA) in the other parts representing respectively 12.5 %, 28 % and 59.5 % of the study area (Fig. 1). In Tunisia, rainfall is influenced by general atmospheric circulation patterns; the most important are: (1) the Icelandic depression, (2) western Mediterranean depressions (the Gulf of Balearic depression and the Gulf of Genoa depression), (3) eastern Mediterranean depressions (Gulfs of Syrtes depression and Gabes depressions), (4) the Azores anticyclone and (5) the central European anticyclone (Merzougui and Slimani, 2012). In addition to regional factors, rainfall is affected by local factors such as marine influences, continentality, topographic conditions and especially the reliefs. Indeed, the general orientation of the reliefs Tell Atlas and Saharan Atlas (Fig. 1), have an important role to weaken or strengthen storms coming from different directions (Slimani et al., 2007). Particularly, the precipitation regime in northern Tunisia is strongly influenced by northerly and especially northwesterly flow bringing winter storms and precipitation from the north (Ben

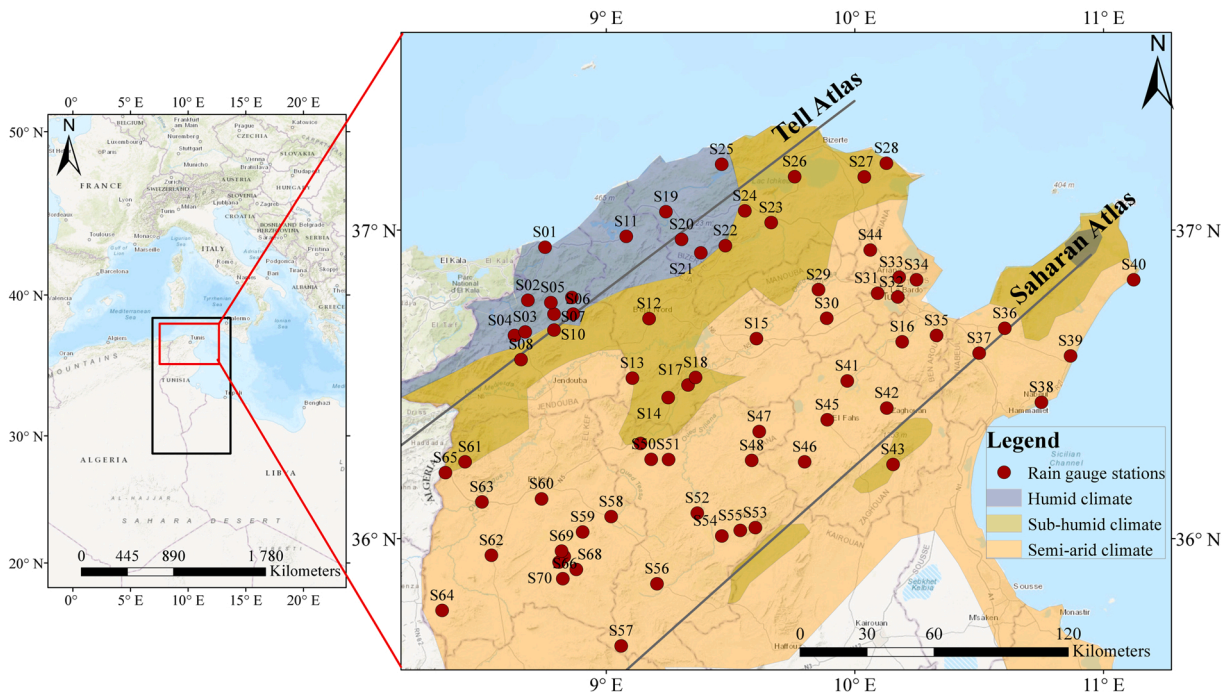


Fig. 1. Rain gauge locations, the Emberger bioclimatic zones and the general orientation of the reliefs (Tell Atlas and Saharan Atlas) in northern Tunisia.

Boubaker, 2003; Xoplaki et al., 2004; Brandimarte et al., 2011 and Jelassi et al., 2016). There is also the northeasterly flow which brings rainfall from the European continent.

Daily rainfall series are used, which are collected from 70 rainfall stations well-distributed in the study region (Fig. 1) over the period 1960 – 2009. The database is obtained from the General Directorate of Water Resources of the Agriculture and Water Resources Ministry. Table 1 displays the rainfall stations characteristics, including the geographic position, elevation, starting date of observations, percentage of missing values, December-January-February (DJF) totals and the bioclimatic region. The dataset consists of 3500 DJF seasons (50 years of observations at 70 stations) corresponding to 31,5910 observed days. For a given year, any missing value detected in the DJF daily observations leads to the exclusion of the whole season from the analysis. The number of considered seasons, for each station, is mentioned in Table 1. In total, the number of seasons without missing values is 3137. Note that DJF for year m corresponds to December of year $m-1$ and January-February of year m .

3. Methodology

The objective of the event-based approach is to understand the rainfall structure and characterize its spatio-temporal variability with respect to event (spells) descriptors. Estimating key descriptors of a rainfall season on the basis of rainfall spells may be prominent for specific applications (e.g., agriculture, ecology, climate change, erosion, water resource management, etc). This study is limited to the winter season, DJF. The choice of this season is motivated by the highest amount of precipitation recorded during the DJF season and also by the cyclonic nature of winter rainfall which allows the definition of independent event in relation with cyclone lifecycle at a synoptic scale.

3.1. Definition of rainfall events and data transformation

The definition of rainfall event is based on the separation of dry and wet spells. There is not a unique way to achieve event delimitation. In the literature, researchers tend in general to choose a minimum inter-event time (MIET) which is the smallest duration of dry period separating two independent rainy periods (Dunkerley, 2008; Molina-Sanchis et al., 2016; Dilmi et al., 2017; Parchure and Gedam, 2019). This approach is adopted in the present study. Some studies consider the hydrological conditions (e.g., the soil surface seals, soil wetness and the event runoff) to estimate the MIET (Cattan et al., 2006; Joo et al., 2014). Moreover, there are other statistical approaches based only on the rainfall time series such as the autocorrelation analysis or the exponential method (Wenzel and Voothees, 1981; Restrepo-Posada, 1982; Asquith et al., 2005). The autocorrelation analysis is based on choosing the time lag on which the autocorrelation coefficient of the rainfall data set converges to zero. This approach is used here to analyse the 50-year DJF daily precipitations for each of the 70 stations. Note that, all days with a rainfall amount greater than zero are considered a rainy day. The minimum rainfall value recorded in the daily timeseries is 0.1 mm/day. Unlike several studies (e.g., Jiang et al., 2011), days with small

Table 1
Rain gauges characteristics.

Map identifiant	Stations	Unique ID	Longitude (°)	Latitude (°)	Altitude (m)	Start of observations	Percentage of daily missing values 1960–2009 (%)	Average DJF rainfall 1960–2009 (mm)	Number of considered seasons	Bioclimatic zone
S01	Tabarka Foret	1483726801	8,75	36,95	10	1959	0.00	385	50	H
S02	Ain Drahim	1483018801	8,68	36,77	715	1889	0.00	692	50	H
S03	Ain Debba	1485013801	8,68	36,7	470	1968	21.48	552	39	H
S04	Ain Baya	1485007801	8,64	36,66	330	1951	18.40	389	40	H
S05	Majen Safi	1483428101	8,78	36,77	530	1942	19.20	492	39	H
S06	Ain Zana	1485053501	8,87	36,73	875	1950	9.86	442	37	H
S07	Ain Sallem	1485042201	8,79	36,73	836	1942	36.72	345	27	SH
S08	Ain Merja	1485035001	8,66	36,58	350	1949	28.90	226	30	SH
S09	Ain Hamraya	1485026001	8,86	36,78	553	1912	28.80	427	28	H
S10	Bouherma	1485140301	8,79	36,68	250	1959	10.00	313	41	SH
S11	Nefza delegation	1483464502	9,08	36,98	30	1964	2.50	406	49	H
S12	Beja Inrat	1485082302	9,17	36,71	234	1959	2.18	237	49	SH
S13	Thibar	1485769002	9,11	36,52	365	1959	1.86	208	49	SH
S14	Teboursouk	1485755802	9,25	36,46	440	1959	1.68	222	48	SH
S15	Mjez el Beb	1485429202	9,6	36,65	142	1959	8.50	157	45	SA
S16	Khelidia	1484372423	10,19	36,64	38	1959	10.47	164	40	SA
S17	Ain Jemmala	1485017702	9,33	36,5	320	1951	29.89	200	35	SH
S18	Ain Tounga SM	1485051102	9,36	36,52	110	1962	12.52	163	43	SH
S19	Sejnen	1483578005	9,25	37,06	135	1966	0.33	360	50	H
S20	Bazina	1483078205	9,3	36,97	460	1960	21.88	449	37	H
S21	Djoumin délégation	1483217005	9,39	36,93	109	1970	1.86	324	48	H
S22	Oued Joumin djantar	1483486105	9,48	36,96	109	1962	9.65	318	44	SH
S23	Sidi Mbarek (Mateur)	1483648105	9,66	37,02	40	1981	0.68	217	49	SH
S24	Ghezalla-barrage	1483287805	9,55	37,07	75	1968	18.11	307	38	SH
S25	Jbel Essema	1483215005	9,46	37,21	210	1959	4.39	295	45	SA
S26	Tinja	1483772405	9,76	37,17	8	1959	0.84	244	48	SA
S27	Alia	1483061705	10,04	37,17	120	1959	1.58	195	47	SA
S28	Ras Djbel	148353005	10,13	37,22	53	1959	8.86	242	43	SA
S29	Bathan	1485079106	9,85	36,81	145	1959	0.78	149	47	SA
S30	Borj el Amri	1485100906	9,89	36,72	55	1959	0.50	149	49	SA
S31	Manouba	1484411924	10,09	36,8	60	1959	5.00	156	46	SA
S32	Tunis Manoubia	1484786307	10,17	36,78	66	1959	0.17	166	49	SA
S33	Ariana	1484066522	10,18	36,85	10	1959	0.00	173	50	SA
S34	Tunis Carthage	1484783207	10,25	36,85	5	1959	0.00	172	50	SA
S35	Morneg	1484177506	10,33	36,66	80	1959	0.84	194	50	SA
S36	Manzel Bouzalfa	1484421708	10,6	36,68	60	1959	0.53	202	46	SA
S37	Grombelia	1484294908	10,5	36,6	50	1959	0.36	175	48	SA
S38	Nabeul	1484462808	10,75	36,44	1	1959	0.36	138	49	SA
S39	Korba	1484373108	10,87	36,59	9	1959	1.00	151	49	SA
S40	Kelibia	1484362908	11,12	36,85	3	1959	0.50	202	50	SA
S41	Bir Mcherga	1484096207	9,97	36,51	155	1959	1.18	141	49	SA
S42	Zaghouen	1484807707	10,13	36,43	230	1959	1.16	163	49	SA
S43	Saouef	1484560707	10,15	36,24	170	1959	3.20	136	44	SH
S44	Cherfech	1485161607	10,05	36,93	59	1959	2.67	183	44	SA
S45	Pond de Fahes	1484524607	9,89	36,39	175	1959	3.00	131	48	SA
S46	Sidi Boubaker	1484623207	9,8	36,25	340	1959	3.18	152	47	SA

(continued on next page)

Table 1 (continued)

Map identifiant	Stations	Unique ID	Longitude (°)	Latitude (°)	Altitude (m)	Start of observations	Percentage of daily missing values 1960–2009 (%)	Average DJF rainfall 1960–2009 (mm)	Number of considered seasons	Bioclimatic zone
S47	Douarada dre	1484130704	9,61	36,35	260	1952	40.90	151	46	SA
S48	Ain Ghesil	1485024404	9,58	36,25	563	1961	12.83	137	43	SA
S49	Krib	1485377804	9,14	36,32	447	1959	2.01	190	48	SA
S50	Ain Tabia	1485046704	9,18	36,28	414	1968	17.33	165	42	SA
S51	Akouat Gare	1485059104	9,26	36,26	140	1932	2.18	139	49	SA
S52	Siliana	1485676404	9,37	36,08	431	1959	0.67	120	50	SA
S53	Saadia Bargou	1486547804	9,6	36,04	524	1959	0.16	214	50	SA
S54	Barrage Lakhmes	1485076704	9,47	36,01	475	1959	4.34	130	47	SA
S55	Ain Zakkar	1485052704	9,54	36,03	676	1932	6.15	139	46	SA
S56	Makthar	1485410204	9,2	35,85	900	1932	0.00	159	50	SA
S57	Rohia	1486543104	9,06	35,65	622	1959	17.01	66	40	SA
S58	Sers	1485361903	9,02	36,07	501	1959	13.32	125	43	SA
S59	Zouarine	1485827203	8,9	36,02	571	1965	9.01	126	46	SA
S60	Kef	1485361903	8,74	36,13	491	1959	1.00	132	47	SA
S61	Gardiamou	1485286403	8,43	36,25	195	1959	5.71	152	46	SA
S62	Tajerouine	1485732803	8,58	35,95	750	1959	10.74	113	41	SH
S63	Oued Mellegue K13	1485499003	8,5	36,12	324	1959	0.84	81	49	SA
S64	Kalaat Lasnem	1485352503	8,34	35,77	623	1959	9.90	101	41	SA
S65	Sakiat Sidi Youssef	1485550203	8,35	36,22	803	1959	10.00	172	45	SH
S66	Ain Zeligua	1485055304	8,83	35,87	853	1959	30.02	117	36	SA
S67	Dehmani municipalité	1485251003	8,83	35,94	622	1957	0.00	131	50	SA
S68	Ksour école	1485383904	8,89	35,9	720	1913	12.7	120	43	SA
S69	Ksour élevage	1485250804	8,82	35,96	600	1959	0.00	116	50	SA
S70	Dehmani élevage	1485250803	8,81	35,92	652	1932	18.03	106	37	SA

rainfall amount or drizzles (<1 mm/day), are not considered as dry days in our study. Indeed, the inclusion of low rainfall events can provide a better characterization of the rainfall intermittency and the events typology over the season. In addition, and for the adopted event-scale, the same rainfall event may include a dry day or days with drizzles, therefore, fixing a threshold may affect the intra-event variability and allows the split of the one single event into separate events.

As shown in Fig. 2, the blue and red curves represent respectively the median and the mean of autocorrelation values for 1-day lag to 5-day lags. The two horizontal lines show the approximate 95 % confidence intervals for a white noise process. The two curves decrease and cut the horizontal line (critical level) between 2 and 3 days. From 3-day lags and beyond, the mean and the median are no longer significant. Therefore, the MIET is of the order of 2–3 days. The proposed analysis is based on the use of a fixed MIET for all stations; thus, we choose to fix MIET to two days. This choice avoids the creation of intensive intra-event gaps of larger MIET.

Fig. 3 illustrates an example in which a fictive 18-day period is separated into three events using the two-day MIET. As shown in Fig. 3, this MIET choice results in one event representing an isolated day (event 1), another with several consecutive days (event 2) and another consisting in several days with a one-day break (event 3).

Once the MIET is defined, six descriptors/features are selected to characterize the analysed DJF season (Table 2). The first three features are: (1) the number of events per season N_e , (2) the number of rainy days per season N_d , and (3) the seasonal cumulative rainfall C_t (Table 2). These three features help to identify the overall rainfall characteristics of the season. Considering the event scale, on the basis of the event number per season, we have defined (4) the average accumulation per event $C_{a,e}$, and (5) the average event duration $D_{a,e}$. Based on the number of rainy days for a given season, we also defined (6) the average accumulation per rainy day per season $C_{a,d}$, which is assumed as an indicator of the "intensity" of the season. Thus, with these six features we enable characterization of the rain events type and the "quality" of rainfall during the season.

The transformed dataset is, then, presented in the form of an observation matrix where each row corresponds to a DJF season in a given station and the columns are the adopted characteristics of rain events over the DJF season.

These six variables have different units and therefore different scales. In order to give equal weights to the data for the further analysis, a normalisation is applied to yield zero-mean and unit variance data.

3.2. Self-Organizing Map (SOM) algorithm and parameters

A multivariate data analysis using SOM method is proposed to divide the sample of DJF rain-event features observed on the stations network during 50 years into clusters in an unsupervised way.

SOM method is defined as a neural network algorithm well-designed for pattern recognition, clustering and dimensionality reduction (Kohonen, 2013). The SOM algorithm has several advantages that make it a valuable tool in data mining. The main one is that it combines the vector quantization and the vector projection. The vector quantization divides a dataset into subsets of homogeneous data represented by a prototype vector. The prototypes projection and visualization aim to understand the structure of the data to identify the possible clusters structure. The SOM can be applied to a large and complex dataset. Moreover, it is robust to missing data (Vatanen et al., 2015).

The SOM map or Kohonen map is a grid of units (neurons), usually a two-dimensional (2D) grid where each neuron is associated to a reference vector (prototype) in the learning data set and thus corresponds to a cluster of rainfall events having common characteristics. The neurons are connected and a topological relationship is defined among neighbourhood neurons. The SOM algorithm preserves the topology of the data, i.e., neighbour neurons on the SOM map are connected to neighbour cluster in the original space (Haykin, 2009; Kalteh et al., 2008; Hannachi, 2021). The structure of the SOM map consists of two layers: the input layer contains the

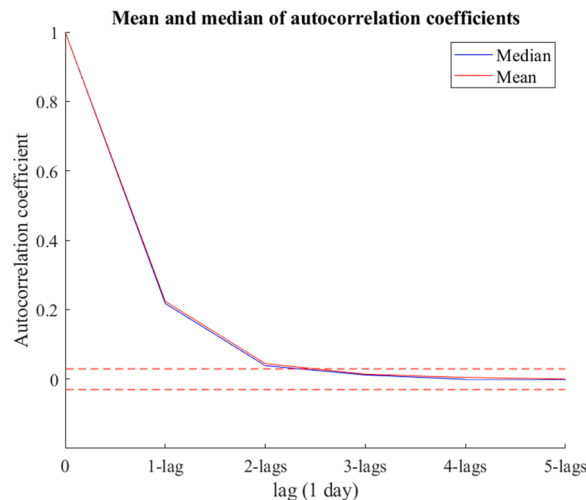


Fig. 2. Median and mean variation of autocorrelation coefficients calculated for 70 stations over the study area from 1-day to 5-day lags. The horizontal dashed lines show the 95 % confidence interval of a white noise process.

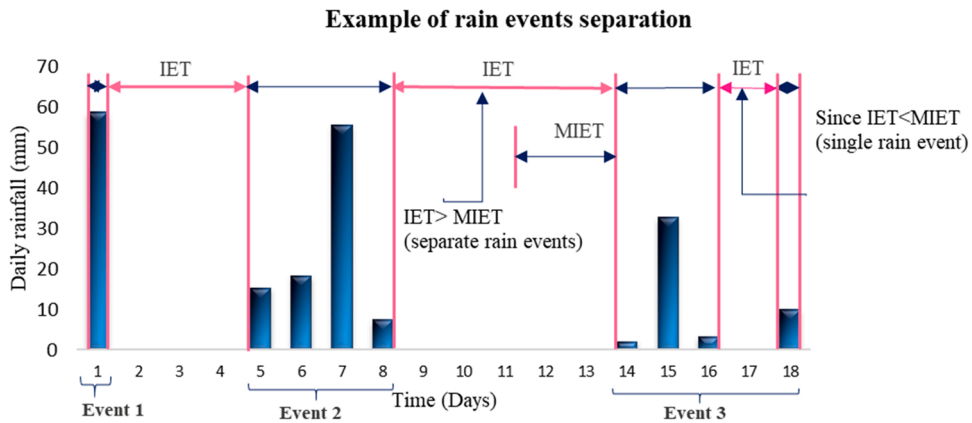


Fig. 3. Illustration of rain event and inter event time (IET).

Table 2

Definition of seasonal rainfall features/descriptors (for individual stations). The notation X_i stands for the total rainfall in i -th day.

No	Descriptor/feature	Symbol	Unit	Definition
1	Number of events	N_e	events/season	
2	Number of rainy days	N_d	days/season	
3	Seasonal cumulative rainfall	C_t	mm	$C_t = \sum_{i=1}^{N_d} X_i$
4	Average accumulation per event	$C_{a,e}$	mm/event	$C_{a,e} = \frac{C_t}{N_e}$
5	Average event duration	$D_{a,e}$	days/event	$D_{a,e} = \frac{N_d}{N_e}$
6	Average accumulation per rainy day	$C_{a,d}$	mm/day/season	$C_{a,d} = \frac{C_t}{N_d}$

prototype (reference vector) in the input space, the output layer is the 2D grid of neurons. To match the input vectors to the map neuron, the SOM uses a competitive learning process. Like the more well-known K-means algorithm, the SOM algorithm is an iterative two-step algorithm: the assignment and the update step. The assignment step consists in associating each observation with the closest weight vector or Best Matching Unit (BMU) then the BMU and its neighbors are updated. SOM model is illustrated in Fig. 4.

After the learning process, the SOM mapping quality needs to be evaluated. Therefore, two quantitative measures that are generally adopted are taken here: the average quantization error (QE) and topographic error (TE). The QE gives the average distance between the rainfall events and the prototype that represents them. TE evaluates the topology preservation. Indeed, it is defined as the average geometric distance between the first and second best matching units in the SOM. If the units are neighbors we consider that the topology has been preserved for the considered input; otherwise, it is an error. TE is given by dividing the total number of errors by the number of inputs.

SOM algorithm parameters (especially map dimension, neighborhood radius σ and iterations number) must be chosen carefully because bad choices can lead to inconsistent results. After studying the effect of changing these parameters on the map quality, the adopted parameters for the training step are listed in Table 3.

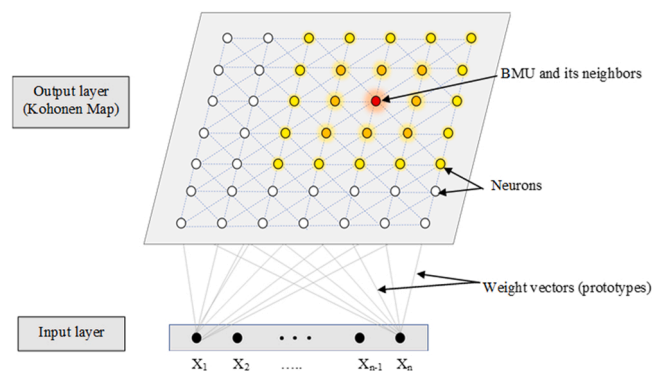


Fig. 4. Illustration of the Kohonen SOM Model. Neurons (or units), shown by small circles, are fully connected to all units in the input layer, the BMU is shown by red color, nodes inside the neighborhood radius are colored in yellow and orange, and nodes outside are colored white. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

Table 3
Adopted parameters for SOM algorithm.

Parameters	
Neurones numbers	320
Map dimension	(20 ×16)
Map lattice	Hexagonal
Neighbourhood function	Gaussian
[σ_{init} σ_{final}] in rough training	[8 3]
[σ_{init} σ_{final}] in fine tuning	[3 0.5]
Iteration number in rough training (T)	1000
Iteration number in fine tuning (T)	5000

In many cases, when a high number of map units is used a second classification step is required to ensure an accurate delimitation of clusters on the SOM map (Vesanto and Alhoniemi, 2000). In the literature, the hierarchical agglomerative clustering (HAC) is commonly combined with SOM in different contexts. For example, Gueye et al. (2010) study the meteorological regime in Senegal for the summer monsoon season on synoptic scales. Yahi et al. (2013) cluster meteorological situations provided by meteorological model outputs during summers in the northern France region, Dilmi et al. (2017) classify rain events observed at very high resolution in Palaiseau region of Paris. El Hourany et al. (2021) provide a robust decomposition of Mediterranean into bio-regions using remote sensing observations. This SOM-HAC procedure exposed in these different studies have shown the relevance of the combined SOM-HAC approach to extract information on spatial and temporal variability from multivariate datasets. These studies in various application domains have in common to deal with data of geophysical nature and which present a significant natural hazard.

This two level-approach SOM-HAC is equally adopted in this study. Indeed, SOM ensures a first clustering and a visual inspection of the rainfall features as a first step. This step reduces the DJF season number where seasons with similar rainfall event properties are grouped and represented by a single, more significant prototype season. The hierarchical clustering is used to cluster the map units into a fewer number of classes, see flowchart (Fig. 5).

3.3. Hierarchical clustering analysis

The hierarchical clustering procedure starts by calculating the interpoint matrix proximity between all singleton clusters (Vesanto and Alhoniemi, 2000). To merge two clusters, different types of linkage criteria can be used. Ward ‘s linkage distance (Ward, 1963) is used here because of its ability to minimize the total within-cluster variance or error sum square (Murtagh and Legendre, 2014). The hierarchical clustering produces a tree-based representation named dendrogram that highlights the hierarchical links between different clusters. The dendrogram is composed of root, merging nodes and branches. The root is linked to the whole dataset, the merging node reflects the distance between the two merged clusters, the height of the branches corresponds to the increase in the

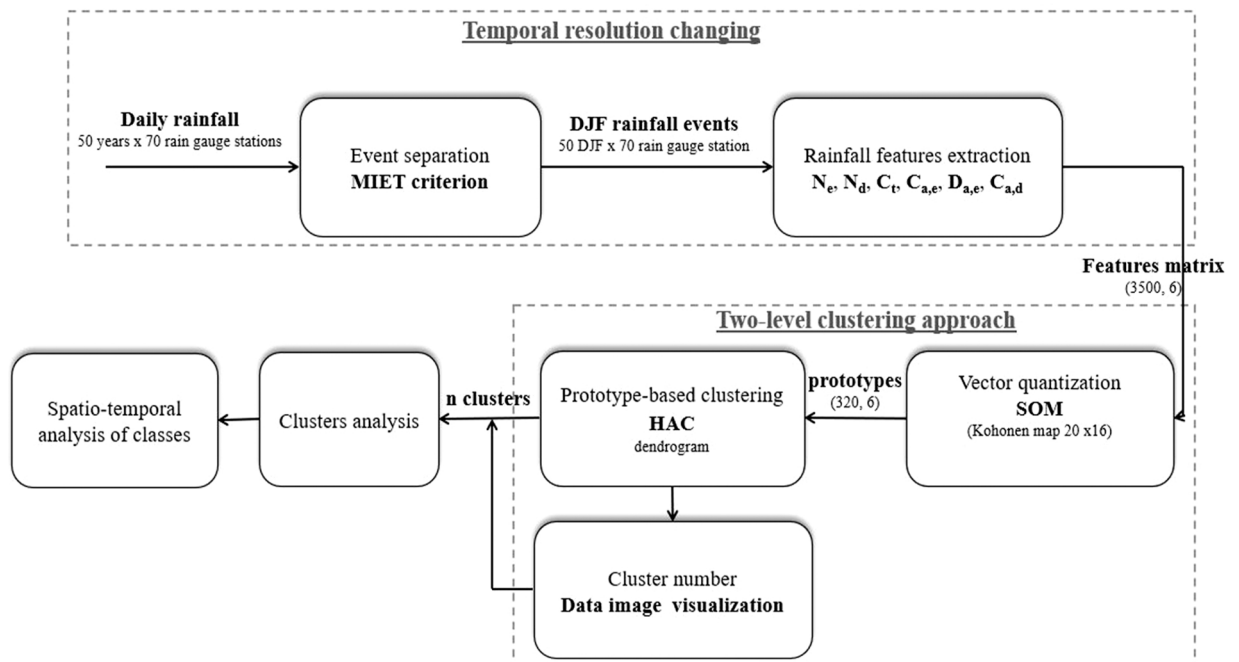


Fig. 5. Flowchart of the adopted methodology.

interclass inertia caused by the merger, and the greater the branches height the more dissimilar are the merged clusters. Cutting the dendrogram at different levels gives a different partitioning (Husson et al., 2010). In general, we tend to cut the tree where the branches are longer. Sometimes, and depending on dendrogram shape, it is not obvious to make a decision about cluster number. To ensure the robustness of the results, we can use other visually based techniques or numerical criteria.

3.4. Data image

Several methods have been proposed in the literature to estimate the number of clusters such as the gap statistic (Tibshirani et al., 2001) and the silhouette (Kaufman and Rousseeuw, 1990). Here, we use a different method based on data visualization, known as data image (Ling, 1973; Minnotte and West, 1999). Data image is a visually-based exploratory tool of clustering. It consists on transforming a multivariate data set (n observations \times p variables) into a two-dimensional image indicating, by colors or gray-scale, the magnitude of each variable for each observation (Martinez and Martinez, 2005). The data can also be visualized based on a similarity/dissimilarity measure, such as Euclidean distance, and transformed into an inter-point distance matrix. This matrix reflects the closeness within the original high-dimensional space and can also be presented as a data image where each pixel represents the distance between any pair of observations. The data can be rearranged in different ways to achieve better contrasts. Minnotte and West (1999) proposed several approaches of hierarchical clustering to reorder the columns and rows of the inter-point distance matrix, enabling clusters to emerge and be discerned along the diagonal of the data image. (Hannachi et al., 2011) used this method for an application to the stratospheric polar vortex, and Hannachi (2012) for jet stream analysis. One of the main characteristic features of data image is that one can use the human natural skill of pattern recognition to determine the likely number of clusters. The methodology proposed in this section is summarized in the flowchart in Fig. 5.

4. Results

4.1. Descriptive statistics of DJF rainfall features

For each of the 3500 DJF (50 years of observations at 70 stations), the six features characteristic of rain events structure, previously described (Table 2), are computed. Year by year boxplot visualization is adopted here to analyse the samples of the generated variables in Fig. 6 and Fig. 7. For a given year, the singular realization (or outlier) is plotted as individual point presenting the observation outside 1.5 times the interquartile range above the 25th percentiles q_1 and below 75th percentile q_3 ($q_1 - 1.5(q_3 - q_1)$ or $q_3 + 1.5(q_3 - q_1)$). For each feature, the median (green line) and the interval $\mu \pm 2\sigma$, where μ is the mean and σ is the standard deviation of the variable, are also plotted to identify the exceptional years (seasons). In Fig. 6a, the boxplots of event number N_e show a negative skewness (median close to the third quartile). The N_e median varies from 5 to 12 events/season, and the lowest value is recorded in 1995 season. Moreover, the outliers (red crosses) are not very frequent, which reflect a low inter-station variability of the event number variable. In Fig. 6b, the boxplots of N_d show prominent positive skewness (median close to the first quartile). For this variable, the significant number of outliers, especially in the last decades, indicates the high inter-station variability. The median of rainy days number shows non-negligible inter-annual variability. Some years have medians particularly high, close to the upper control limit ($\mu + 2\sigma = 40$ days), see, for example, 1973 and 2003. Others have a number of rainy days particularly low, close to $\mu - 2\sigma$ limit (about 15 days), e.g., 1977, 1980, 1988, 1995, 2002. There is also a weak signal of a 5-yr period. However, the analysis of the inter-annual

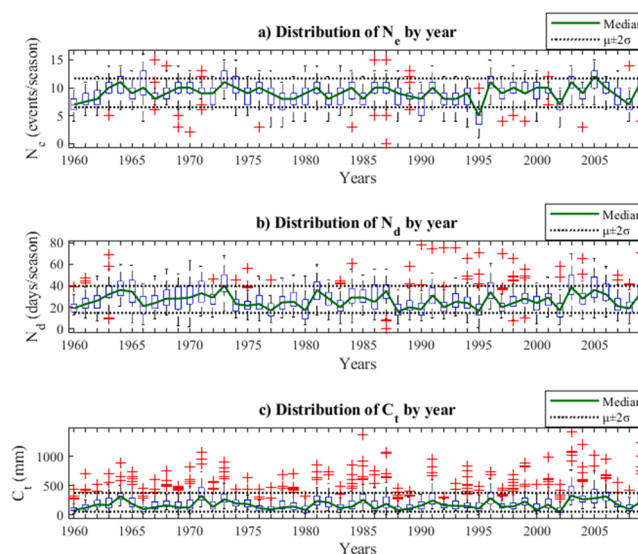


Fig. 6. Boxplot versus time for N_d (a), N_e (b) and C_t (c) across the 70 rainfall stations.

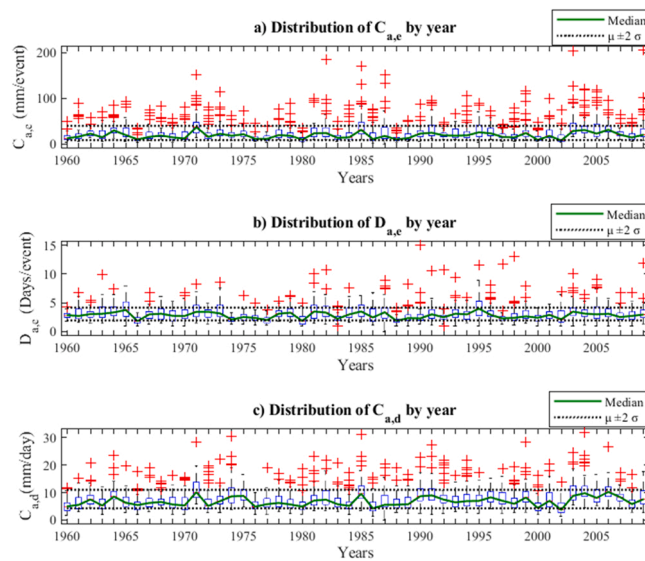


Fig. 7. Boxplot versus time for $C_{a,e}$ (a), $D_{a,e}$ (b) and $C_{a,d}$ (c) across the 70 rainfall stations.

signal and its drivers is beyond the scope of this study. The graph of seasonal accumulation C_t (Fig. 6c) shows a median close to the first quartile for most years reflecting again positive skewness. Overall, the C_t median varies between about 100 and 300 mm. The 25th and 75th percentiles vary respectively from 33 to 478 mm. Extremes (outliers) occur quite frequently in all seasons (years) without exception. The C_t boxplots allow to identify the wettest seasons, e.g., 1964, 1971 and 2003, where the median is close to the $\mu + 2\sigma$ limit (about 400 mm), and also the driest seasons, e.g., 1960, 1980, 2000 and 2002.

The boxplots of event structure variables, the mean accumulation $C_{a,e}$ and duration $D_{a,e}$ per event, are represented respectively in (Fig. 7a) and (Fig. 7b). The mean event accumulation graph (Fig. 7a) shows positive skewness. The median varies between 7 and 36 mm/event. For some stations, average event accumulation exceeds 200 mm/event, see, for example, 2003 and 2009. The graph of the mean duration of event $D_{a,e}$ (Fig. 8b) shows a low inter-annual variability where the median is mostly between 2 and 4 days/event. The singular values refer to stations with particular long events during the season, where the mean event duration can reach 15 days/event, e.g., 1990. In the opposite case, the seasonal mean duration can be very small (1 day/event) in some stations, see, for example,

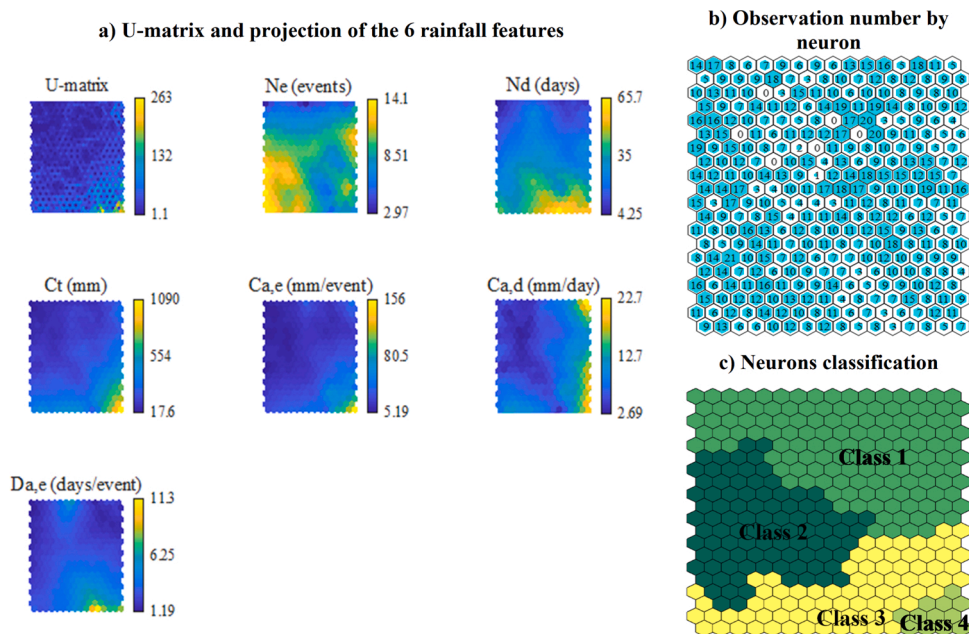


Fig. 8. Projection of the six rainfall event features (variables) on the topological map and the U-matrix (in the upper left) (a), The number of observations represented by each neuron on the SOM map (b), Neurons classification on the topological map (c).

the 1983 and 1992 years. The analysis of the boxplots related to the average accumulation per rainy day $C_{a,d}$ (Fig. 7.c) shows that the median varies from year to year between 3.7 and 10.2 mm/day. The highest medians are recorded in 1971, 1985, 2004 and 2006 with about 10 mm/day. Some seasons have extremely low median, below the lower control limit (5 mm/day) such as 1976, 1986, 2000 and 2002.

One of the main conclusions that can be drawn from Fig. 6 and Fig. 7 is that the wettest seasons seems to result from a few long and intense rainy events rather than accumulation of a large number of weak events. For example, the high seasonal accumulation C_t in 1964 and 2003 corresponds to a relatively high rainy days number and a medium events number. Otherwise, there are a number of years with exceptionally high-values of rainy days N_d , e.g., 1990, which exceeds 70 in a given station, but without particularly exceptional seasonal accumulation C_t , i.e., frequent rains with very small amounts. Moreover, the distribution shown in C_t graph indicates that the majority of stations have low accumulated rainfall amounts, and a large part of the study area is considered a little dry.

4.2. Kohonen map analysis

To analyse the feature matrix previously obtained, we first use SOM map. Fig. 8 displays the mapping obtained by SOM using the parameters of Table 3, and shows the projection of the 6 variables on the topological map and the corresponding U-matrix (Fig. 8a), observation number by units' (Fig. 8b), and the neurons classification (Fig. 8c). The Fig. 8b shows that, after training, the number of observations represented by each neuron is between 0 and 20. The color coverage, in Fig. 8b, reflects the number of observations by each neuron. Only 2 % of the neurons are empty, i.e., they don't represent any data from the feature space. The U-matrix (Fig. 8a, first row and column) enables understanding and interpretation of the structure of the original data, as it represents distances between adjacent neurons. The green- and yellow-colored neurons represent high values, and signify that the weight vectors of the units are distant from each other. Thus, U-matrix (Fig. 8a) shows a distinctive region located in the bottom right corner with higher values than in the other parts of the map.

The color contrasts in (Fig. 8a) reflect the values of the rainfall features at each neuron. The analysis of prototype variables on the topological map shows that SOM map is organized along several lines/directions:

- A diagonal starting from the upper left corner (lowest values) to the bottom right corner (highest values) as shown for the values of C_t and $C_{a,e}$.
- A gradient going from top to bottom for the N_d and $D_{a,e}$.
- A gradient going from left to right according to the average accumulation per rainy day $C_{a,d}$.

The number of events N_e (Fig. 8a, row 1, column 2), on the other hand, has a more complex structure. Note that the seasons with particularly high number of events are situated on the left hand side of the map and correspond to low values of the $C_{a,e}$ (Fig. 8a, row 2, column 2). The average duration $D_{a,e}$ (Fig. 8a, row 3, column 1) of events for those seasons is also low. These seasons are composed of frequent short events with a low rainfall intensity. The seasons with particularly low number of events correspond to low number of rainy days N_d , and are located at the top of the N_d map. The average accumulation and the average duration per event for those seasons are also low. These correspond precisely to dry seasons with few, short and low-intensity rainfall events. The similarity between the seasonal accumulation C_t and the average accumulation per event $C_{a,e}$ indicates that the DJF season is influenced more by the intensity of rainfall events than by their number. The seasons with the highest event number do not necessarily correspond to the seasons with the highest rainfall accumulation.

The very high values for the number of rainy days N_d at the bottom and slightly to the right of N_d map (Fig. 8a, row1, column 3) correspond to rainfall patterns with low seasonal accumulation C_t , low average accumulation per rainy day $C_{a,d}$, and low number of events N_e , but with particularly high average event duration $D_{a,e}$ (Fig. 8a, row 3, column 1). The seasons associated to these neurons are characterized by intermittency (at the event scale) and quite low daily rainfall.

The seasons represented by the neurons in the upper right corner of $C_{a,d}$ map (Fig. 8a, row 2, column 3) are dry with respect to the map of C_t and are characterized by a few, short and intense events. The average accumulation per rainy day $C_{a,d}$ may have significant values in dry or wet seasons because it depends mainly on the number of rainy days N_d . In fact, this part of the map represents seasons with seasonal accumulation C_t between around 100 and 300 mm distributed over a very low number of rainy days N_d (less than about 10 days) per season.

4.3. HAC and data image results

The neurons are aggregated into a number of clusters using the HAC. The obtained dendrogram, based on Ward's algorithm, is presented in Fig. 9a, and shows the links between the 320 neurones. The vertical axis represents the inertia or dissimilarity between vectors. Different numbers of clusters are obtained by cutting through the tree, and the number of clusters corresponds to the number of branches cut. The green, black and red lines show different cutting levels in the dendrogram and yields respectively to 2, 3 and 4 clusters.

The data image is applied to the dissimilarity matrix of the weight vectors (Fig. 9b). The colors refer to the pairwise distance between weight vectors, where the dark blue color represents proximity/closeness and red (and brighter) color represents distance/remoteness. Fig. 9b shows closeness and remoteness over the whole topological domain. The main diagonal line is obvious in the figure and represents the zero-value. Now, the values/positions in the dissimilarity matrix can be reordered after the application of a

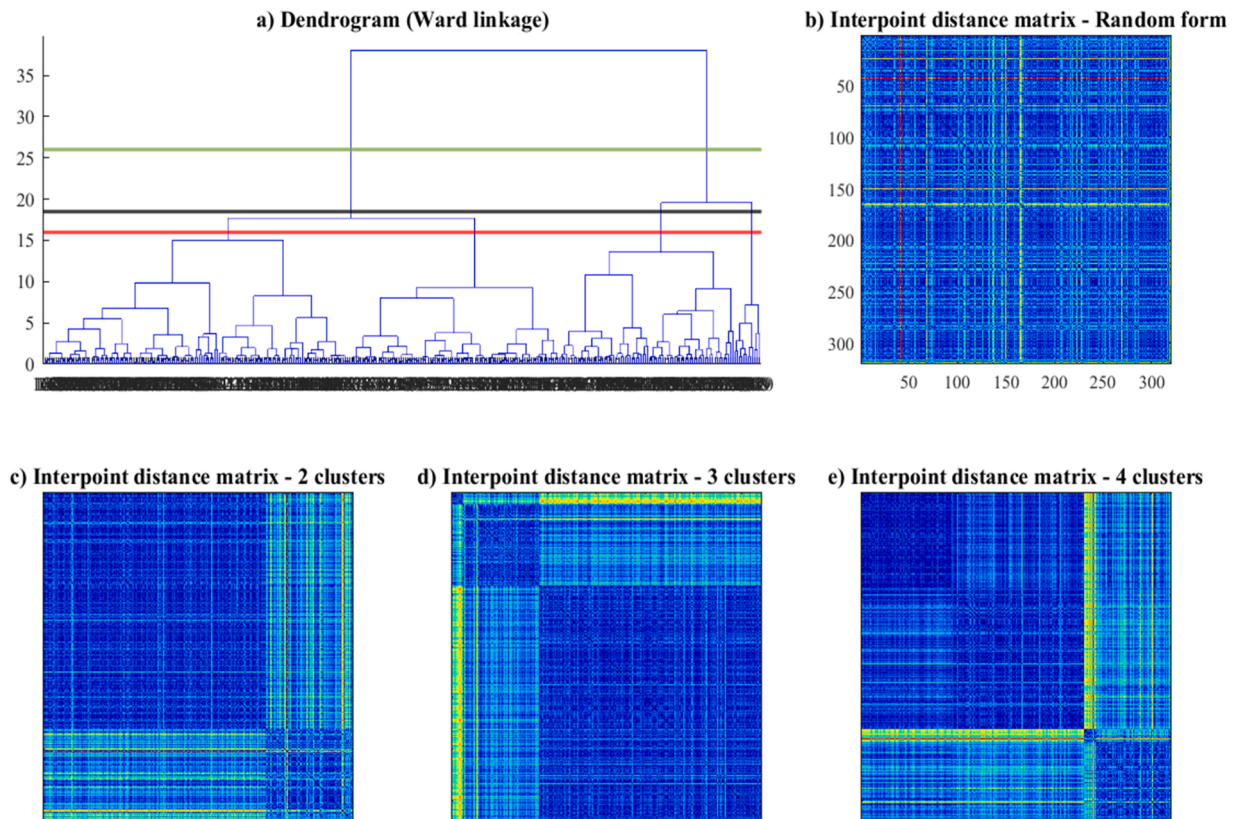


Fig. 9. Dendrogram (a), data image of the pairwise Euclidean distance matrix between weight vectors (b), and as in (b) except that the data are reordered in two- (c), three- (d), and four-class (e) models. The clusters appear as diagonal blocs with darker color than off-diagonal blocs. These classes are obtained through the dendrogram tree.

classification algorithm, e.g., hierarchical clustering here, and the different clusters appear as blocs on the main diagonal. Fig. 9c, Fig. 9d and Fig. 9e show the dissimilarity matrix organized into two, three, and four classes respectively. The blocs observed along the diagonal of each image represent the clusters, and their size reflects the clusters' size. The color contrasts outside the blocs represent the inter-class distances, and therefore the stronger the contrast the better the clustering. Note that in those three figures the axes are not labelled because the order is fixed by the clusters.

Fig. 9c shows two clusters, with two blocs along the main diagonal, with the bigger cluster located in the top left corner. However, for the smaller cluster, a kind of heterogeneity still remains and can be observed when comparing the colour intensity between the two diagonal blocs. This means that this cluster is not homogeneous and may contain elements that could belong to other groups. In the three-cluster hypothesis, the small class in Fig. 9c is now split into two clusters of different size (Fig. 9d). In the assumption of four clusters shown in Fig. 9e, the bigger cluster of Fig. 9c is now split into two clusters. Along the diagonal, the four classes can be clearly distinguished even though the contrast between the two upper diagonal blocs and their off-diagonal blocs is not very strong. With the assumption of an extra cluster, the five-class model (not shown) results in splitting the bigger class of Fig. 9c into three clusters, but these are no longer discerned in the data image. With larger number of clusters, the discernability is completely lost.

Thus, the combination of the dendrogram and the data image analysis favors a partition into three or four clusters. Although the three-class model is equally likely, given the good contrast obtained in Fig. 9d, we will focus the discussion here on the four-class model in order to take into account the nuance that exists in the large class of Fig. 9c. The clusters can be easily identified on the topological map, as the latter is based on proximity. In fact, Fig. 8c shows the repartition of the four clusters on the SOM map obtained by the HAC.

4.4. Class control factors

The conditional histograms of the six rainfall descriptors for each of the four previously selected classes are illustrated in Fig. 10. The mean and the standard deviation (std) of each feature are equally showed in the same figure. Note that the four clusters have not the same observation number. The cluster identification is based on the observation of the colors of clusters neurons on each rainfall feature map mentioned in Fig. 8a in combination of distributions illustrated in Fig. 10 by histogram presentation.

Cluster 1, shown in the top of the topological SOM map (Fig. 8c), is the most populated with about 47 % of all seasons. It is characterized by a low-to-medium number of rainy days N_d , low seasonal rainfall C_t , between about 100 and 200 mm, and low average

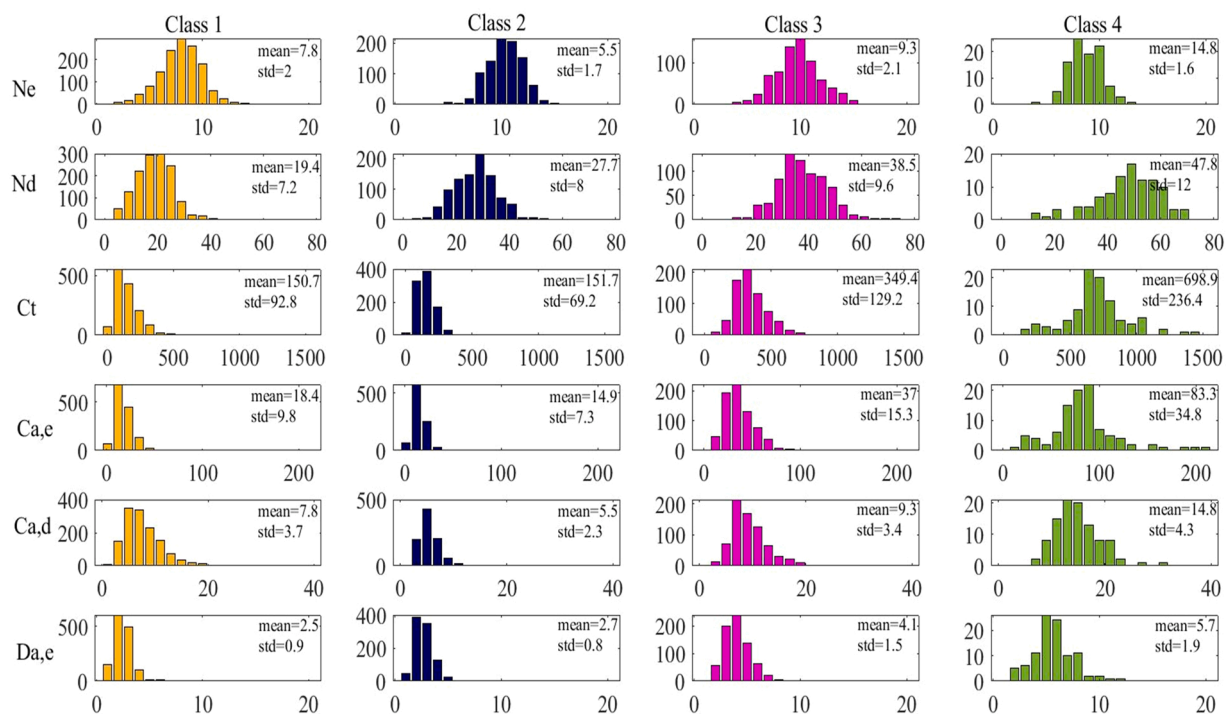


Fig. 10. Conditional histograms of the six DJF rainfall features for the four clusters.

accumulation per event $C_{a,e}$ (Fig. 10, first column). Class 1 has the lowest mode for N_e (8 events), N_d (20 days), C_t (100 mm) and $D_{a,e}$ (2 days/event). This class is typical, it represents dry seasons, with short and weak rainfall events. Cluster 2, shown in the central left part of Fig. 8c, is about 20 % of all seasons (Fig. 8c). This class is characterized by a low average accumulation per rainy day $C_{a,d}$, low number of rainy days N_d and low mean event duration $D_{a,e}$. In addition, it has low seasonal accumulation C_t and low average accumulations per event $C_{a,e}$. Overall, class 2 has almost similar statistics (mean and std in Fig. 10) to class 1 with a difference in the distribution in some features. Indeed, Class 2 has higher mode for N_e (10 events), N_d (30 days), C_t (200 mm) than class 1. The distribution of the features shows that this cluster is a representative of dry seasons, with weak and intermittent events over the season. Cluster 3 (Fig. 8c) is about 25 % of size. The mean values of each features show that this class represents intermediate situations with a medium amount of rain and intermittent events. It is characterized by medium-to-strong seasonal accumulation C_t , between about 200 and 500 mm, and an average accumulation per event $C_{a,e}$ between about 20 and 60 mm/event, which is considered to be medium-to-high, compared to classes 1 and 2 (Fig. 10, third column). The average accumulation per rainy day $C_{a,d}$ has a broad distribution. The event number varies mostly between 8 and 12, but can reach 15 occasionally. The number of rainy days N_d for this class is also high, between about 30 and 60 days, with medium average event duration $D_{a,e}$ in general. This class shares the intermittent character with class 2 which has the same mode for N_e (10 events). However, it shows wetter conditions than the two first classes with higher mode for N_d (35 days), $C_{a,e}$ (50 mm/event), C_t (300 mm), $C_{a,d}$ (8 mm/day) and $D_{a,e}$ (4 days/event). Cluster 4, situated in the lower right corner of the SOM map (Fig. 8c), is the less frequent with about 7 % of all seasons. It is characterized by a very high seasonal rainfall C_t and also high average accumulation per event $C_{a,e}$. The majority of the seasons in this class reaches a record seasonal rainfall C_t between about 500 and 1000 mm, and exceeds 1300 mm in some stations (Fig. 10, last column). It is also characterized by a medium event number and high number of rainy days N_d . The average event duration is high (greater than 4 days per event), and the average accumulation per rainy day is also high, compared to the remaining classes, as it exceeds 30 mm/day. This class represents the rainiest seasons with long and intense events. This class is mainly located in the extreme northwestern part of the domain, represented mostly by station S02 "Ain-Drahem".

If the three clusters hypothesis is adopted, the very dry and the dry seasons (clusters 1 and 2) merge into one single cluster with about 67 % of all seasons. Those clusters share almost the same seasonal accumulation C_t and average event accumulation $C_{a,e}$ with a large difference in event frequency and intensity. The clusters 3 and 4 remain unchanged.

4.5. Spatio-temporal variability of classes

Fig. 11 shows the spatial distribution of the different classes. For each station, a pie chart is used showing the percentages of each class for the observed period. The classes occurrence, for each station, shows a high variability between the north-western part and the remaining parts of the region. Indeed, all stations in humid region (bleu) and about 40 % of stations from the sub-humid region are dominated by classes 3 and 4 (the percentage of wet (class 3) + very wet (class 4) seasons exceeds 50 %). The occurrence of these

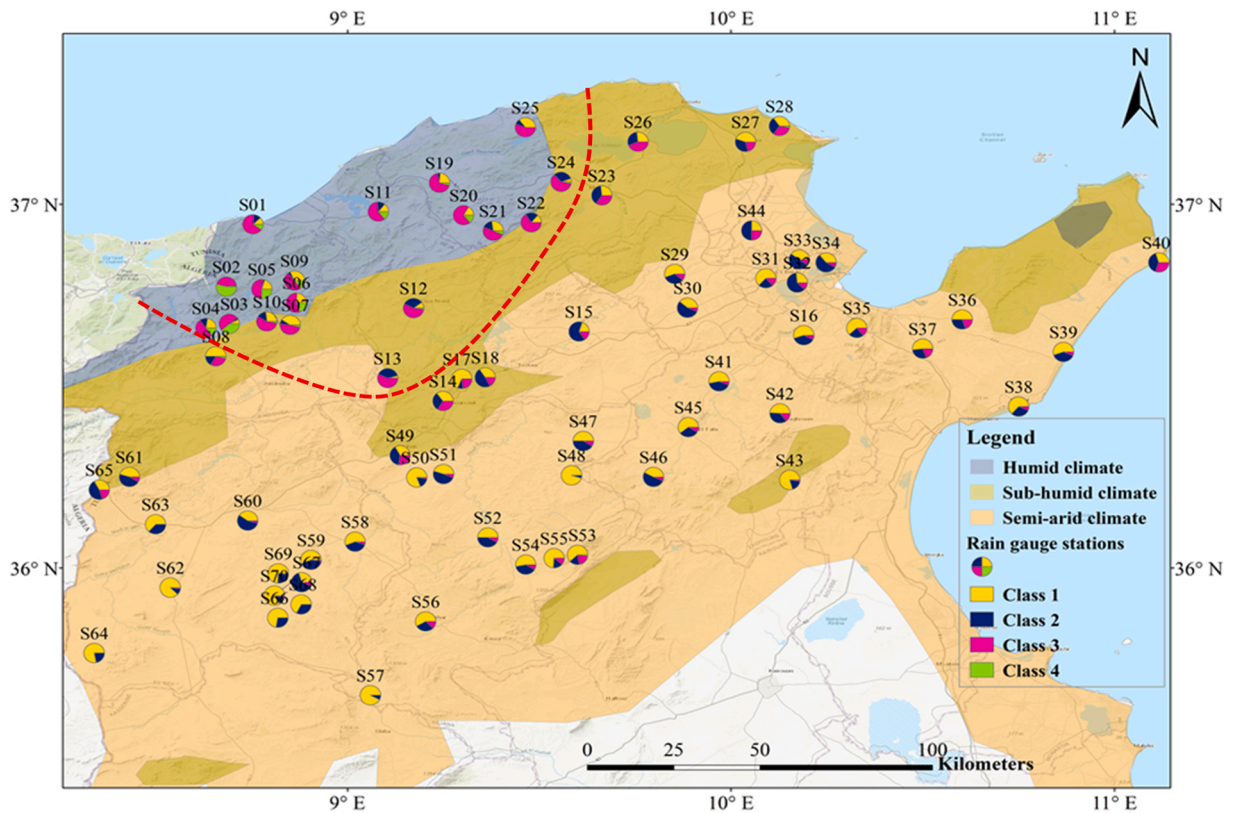


Fig. 11. Percentage of clusters at each station for the entire observational period. The red dashed line separates wet region to dry region.

classes is about 68 % of the time. On the other hand, the southern sector is dominated by the dry and very dry classes, i.e., classes 2 and 1 respectively, approximately 88 % of the time. The very wet seasons with intense events of class 4 occur very rarely in this sector over the 50 years of the record. The two regions are separated by the red dashed line in Fig. 11.

In the remaining part of the paper, we set out to analyse the temporal evolution of the cluster size in the two sectors separately. The northwestern (wet) sector contains 18 stations dominated by classes 3 and 4 (> 50 % of the time (Fig. 11)), whereas the southern (dry) sector contains 52 stations dominated by classes 1 and 2 (> 50 % of the time (Fig. 11)). The temporal variability is analysed using non-overlapping 10-year periods (5 decades). Table 4 shows the percentage of each class by sector in the non-overlapping 10-year periods.

Class 1 shows high percentage during the nineties for both sectors. This period can be classified as a dry decade. Class 2, characterized by frequent events and low rainfall accumulation, was quite frequent in the 1980's in the southern sector, and in the 1970's for the north-western sector. Class 3 populates mostly the northwestern sector. It shows particularly high occurrence frequency in the 1960's. The last cluster, which is the wettest, is exclusively prominent in the north-western sector. This class was a little more frequent during the 2000's

The summarizing result of this cluster size variability shows that over the entire observation period the lack of enough rainfall remains the dominant aspect of the region, which extends over about three consecutive decades from 1970 to 2000. Also, the inter-decadal evolution of the rainfall regime between the two sectors is slightly different.

5. Discussion

In the MIET determination, the significance of autocorrelation is evaluated consistent with the synoptic timescale in the latitudes of the Mediterranean region. Indeed, typical cyclone's life cycle (growth, maturity and decay) of the mid-latitude cyclones and their associated fronts is about 2–3 days. In some favorable and special conditions, cyclone can reach maturity in one day (Zappa et al., 2015). However, this rapid storm intensification is not very frequent. Also, there are cyclones that persist longer than three days (Hakim and Patoux, 2018). Because of the peculiar features of the Mediterranean region such as its complex orography and topography and the high convective and mesoscale activity, the lifecycle of the storm can be affected and reduced. Therefore, smaller MIET (2 days) is more suitable to separate independent rain events in northern Tunisia.

The large difference between the classes of the north and the south sectors could be explained by geographic and orographic factors. Indeed, northern Tunisia is located in the extension of the Atlas Mountain ranges with south-west/north-east orientation. The Tell Atlas extends over the northernmost coast of Tunisia (Fig. 1) and forms a kind of barrier against the westerly and especially the north westerly flow which brings most of the winter precipitation and generates more seasons of class 3 and class 4 (Ben Boubaker, 2003;

Table 4
Cluster size in percentage for the northern and southern sectors in successive decades.

		Class 1 Very dry seasons: low event number	Class 2 Dry seasons: frequent rainfall events, low intensity	Class 3 wet seasons: frequent rainfall event	Class 4 Very wet seasons: long and intense event
Northwestern sector	Decade 1 (1960–1969)	6.7	6.7	78.4	7.8
	Decade 2 (1970–1979)	18.0	23.6	47.2	11.2
	Decade 3 (1980–1989)	22.0	13.0	51.4	13.6
	Decade 4 (1990–1999)	25.7	9.9	57.3	7.0
	Decade 5 (2000–2009)	17.4	10.7	52.8	19.1
	Percentage of classes over 50 years	19.0	13.1	55.6	12.1
	Southern sector	Decade 1 (1960–1969)	49.5	38.3	11.7
Decade 2 (1970–1979)		48.9	37.2	13.8	0
Decade 3 (1980–1989)		47.3	43.9	8.8	0
Decade 4 (1990–1999)		63.6	26.8	9.7	0
Decade 5 (2000–2009)		51.3	31.1	17.3	0.2
Percentage of classes over 50 years		52.3	35.3	12.3	0.1

Merzougui and Slimani, 2012). Thus, the existing reliefs, e.g., Mogods, and Kroumirie mountains, act to shadow the southern sector and reduce precipitation; hence, the dominance of class 1 and class 2. The continentality of the western part of the southern sector combined with its position between the Tell Atlas and Saharan Atlas (Fig. 1) make the driest conditions, with exceptional occurrence of class 1, in this region except some local variability due to high altitude of a number of stations such as S65 (Sakiet Sidi Youssef, 803 m), S66 (Ain Zeligua, 853 m), S53 (Saadia Bargou, 524 m). The stations located near the eastern coast show more variability compared to the western one (class 3 and 2 are more frequent) for the same sector. Indeed, this region is influenced by the northeasterly flow coming from the European continent bringing moderate/weak amount of precipitation (Ben Boubaker, 2003).

Overall, the exposure of stations to different incoming flows combined with some local factors, such as high altitude, existing of forests and proximity to a moisture source, generates more rainfall events hence relatively wet seasons.

6. Conclusions

This paper investigates the temporal and spatial variability of a 50-year long data set of rainfall events in the winter season DJF in northern Tunisia. The data used consist of observed rainfall time series taken from 70 rain gauge stations during DJF 1960–2009. The analysis begins by changing daily rainfall to a sequence of DJF rainfall events and then to seasonal DJF features. These features are clustered using a two level-approach: the SOM combined with hierarchical clustering. The obtained clusters are identified and their space-time regionalization is carried out. Each class characterizes the rain event types and take into account the rain intermittent character at both seasonal and event scales. The rainfall regime was found overall to be dominated by classes describing drought, lack of rain, and intermittent weak events.

Spatially, each station is characterized by a specific distribution of the four clusters over the 50-year period. The spatial regionalization showed that the study area is broadly divided into two sectors, a wet sector in the north (especially in the northwest) and a dry sector in the south. This result is in accordance with previous studies (Jebari et al., 2007; Merzougui and Slimani, 2012; Fathalli et al., 2018).

The analysis carried out of the time variability of the cluster size for both sectors using non-overlapping periods of 10 years shows that dry conditions dominate and extend over three consecutive decades from 1970 to 2000. However, in the last decade (2000–2009), the winter season was particularly very wet, with frequent rainy seasons, especially in the northern sector.

The analysis of rainfall by considering intermittent phenomena and studying events structure enables a better understanding of the rainfall patterns. This can also help identify the various types of seasons by taking into account other rainfall aspects such as the seasonal number of events, their intensity and duration. The method presented here is quite general as it is based on rainfall features, and hence can be applied in other climatic regions, such as the tropics or high latitudes. Further analysis is required to make the method more universal. For instance, the sensitivity to the choice of MIET and equally the choice of threshold to define a rainy day, and its impact are important topics for investigation.

This study must be complemented by a study of the relationship between the rainy season structure analysed here and the climate pattern. The analysis of these complex and non-linear relationships will make it possible to anticipate the impact of large-scale climate

variability on hydrology or agriculture. The preliminary ongoing analysis gives some small but significant correlations between the rainfall features and climatic pattern over such as the arctic oscillation AO and the north Atlantic oscillation NAO. Indeed, NAO pattern is considered one of the recurrent and most prominent atmospheric patterns, especially during the winter season, that affects most of the northern hemisphere, especially Europe, North Africa and the Mediterranean. It is a see-saw that acts to redistribute atmospheric mass between the Icelandic region and the subtropical Atlantic (Wanner et al., 2001). During the negative NAO phase, the jet stream is positioned south of its climatological position and brings warmer and wetter than average conditions over the Mediterranean and northern Africa region. During the positive phase, the conditions are dryer and colder than average (Hurrell, 1995). The correlation analysis is just preliminary results; however, the main objective of the research is to elaborate the relationships between the obtained typology of rainy seasons and the typology of climatic patterns.

CRedit authorship contribution statement

Sabrina Derouiche: Data analysis, SOM and HAC analysis, Writing – original draft. **Cécile Mallet** SOM coding and SOM interpretation, Writing – review & editing. **Abdel Hannachi** event definition, image method's coding, writing – review & editing. **Zoubeida Bargaoui** Database, event definition, statistical analysis, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

ZB would like to thank the General Directorate of Water Resources (DGRE, Tunisia) for providing daily precipitation data. This study was supported by University Tunis El Manar (UTM, Tunisia) with providing travel and accommodation fees for SD to visit LATMOS France and the Department of Meteorology, Stockholm University, MISU, Sweden. This study was supported IMI for inviting ZB to visit MISU. This study was supported by (UTM) for inviting CM to visit ENIT.

References

- Akroun, N., Chazottes, A., Verrier, S., Mallet, C., Barthes, L., 2015. Simulation of yearly rainfall time series at microscale resolution with actual properties: Intermittency, scale invariance, and rainfall distribution. *Water Resour. Res.* 51 (9), 7417–7435.
- Asquith, W., Roussel, M., Thompson, D., Cleveland, T., Fang, X., 2005. Summary of dimensionless texas hyetographs and distribution of storm depth developed for texas department of transportation. Research. Research and Technology Implementation Office, Texas.
- Bargaoui, Z., 1983. Contribution à l'étude statistique de la pluie dans la région de Tunis. Institut National Polytechnique de Toulouse., Toulouse, France (Retrieved from). (http://hydrologie.org/THE/KEBAILI_BARGAOUI_1983).
- Bargaoui, Z., 1994. Comparison of some estimation methods in frequency analysis. *J. Hydraul. Eng.* 120 (2), 229–235.
- Ben Boubaker, H., 2003. Apports pluviométriques et types de flux de surface en Tunisie du Nord. In: Arnould, P., Hotyat, M. (Eds.), *Eau et environnement: Tunisie et milieux méditerranéens*, Vol. 870. ENS Éditions. <https://doi.org/10.4000/books.enseditions.870>.
- Benziarti, Z., 2003. La pluviométrie en Tunisie: Analyse des années très pluvieuses. In: Arnould, Dans P., Hotyat, M. (Eds.), *Eau et environnement: Tunisie et milieux méditerranéens*. ENS Éditions, Lyon, France, p. 869. <https://doi.org/10.4000/books.enseditions>.
- Benziarti, Z., Habaieb, H., 2001. Etude de la persistance de la sécheresse en Tunisie par utilisation des chaînes de Markov (1909-1996). *Sci. Et. Chang. planétaires / Sécheresse* 12 (4), 215–220.
- Brandimarte, L., Di Baldassarre, G., Bruni, G., D'Odorico, P., Montanari, A., 2011. Relation between the North Atlantic oscillation and hydroclimatic conditions in Mediterranean areas. *Water Resour. Manag.* 25 (5), 1269–1279. <https://doi.org/10.1007/s11269-010-97>.
- Cattan, P., Cabidoche, Y., Lcas, J., Colz, M., 2006. Effects of tillage and mulching on runoff under banana (*Musa spp.*) on a tropical Andosol. *Soil Tillage Res.* 86, 38–51.
- Dilmi, D., Mallet, C., Barthès, L., Chazottes, A., 2017. Data-driven clustering of rain events: microphysics information derived from macro-scale observations. *Atmos. Meas. Tech. Eur. Geosci. Union* 10 (4), 1557–1574. <https://doi.org/10.5194/amt-10-1557-2017>.
- Dunkerley, D., 2008. Identifying individual rain events from pluviograph records: a review with analysis of data from an Australian dryland site. *Hydrol. Process.* 5024–5036. <https://doi.org/10.1002/hyp.7122>.
- El Hourany, R., Mejia, C., Faour, G., Crépon, M., Thiria, S., 2021. Evidencing the impact of climate change on the phytoplankton community of the Mediterranean Sea through a bioregionalization approach. *J. Geophys. Res. Oceans* 126. <https://doi.org/10.1029/2020JC016808> e2020JC016808.
- Emberger, L., 1930. La végétation de la région méditerranéenne- Essai d'une classification des groupements végétaux. *Rev. générale De. Bot.* T 42 (461–662), 705–721.
- Fathalli, B., Pohl, B., Castel, T., Safi, M.J., 2018. Errors and uncertainties in regional climate simulations of rainfall variability over Tunisia: a multi-model and multi-member approach. *Clim. Dyn.* 52 (1–2), 335–361. <https://doi.org/10.1007/s>.
- Gargouri-Elouze, E., Bargaoui, Z., 2009. Investigation with Kendall plots of infiltration index–maximum rainfall intensity relationship for regionalization. *Phys. Chem. Earth Parts A/B/C.* 34 (10–12), 642–653. <https://doi.org/10.1016/j.pce.2009.02.00>.
- Ghanmi, H., Bargaoui, Z., Mallet, C., 2013. Investigation of the fractal dimension of rainfall occurrence in a semi-arid Mediterranean climate. *Hydrol. Sci. J.* 58 (3), 483–497. <https://doi.org/10.1080/02626667.2013.775446>.
- Ghanmi, H., Bargaoui, Z., Mallet, C., 2016. Estimation of intensity-duration-frequency relationships according to the property of scale invariance and regionalization analysis in a Mediterranean coastal area. *J. Hydrol.* 514, 38–49. <https://doi.org/10.1016/j.jhydrol.2016.07.002>.
- Gueye, A.J., Niang, A., Sawadogo, S., Sultan, B., DiongueNiang, A., Thiria, S., 2010. Weather regimes over Senegal during the summer monsoon season using self-organizing maps and hierarchical ascendant classification. Part I: synoptic time scal. *Clim. Dyn.* 36 (1–2), 1–18. <https://doi.org/10.1007/s00382-010-0782-6>.
- Hakim, G., Patoux, J., 2018. *Weather: A Concise Introduction*. Cambridge University Press.
- Hannachi, A., 2021. *Pattern Identification and Data Mining in Weather and Climate*. Springer-Verlag, p. 596.
- Hannachi, A., Mitchell, D., Gray, L., Charlton-Perez, A., 2011. On the use of geometric moments to examine the continuum of sudden stratospheric warmings. *J. Atmos. Sci.* 68 (3), 657–674. <https://doi.org/10.1175/2010JAS3585.1>.
- Hannachi, A.W., 2012. The North Atlantic jet stream: a look at preferred positions, paths and transitions. 862–877.

- Haykin, S., 2009. *Neural Networks and Learning Machine*, third ed. Pearson, Hamilton, Canada.
- Hurrell, J., 1995. Decadal trends in the North-Atlantic oscillation: regional temperatures and precipitation. *Science* 269 (5224), 676–679. <https://doi.org/10.1126/science.269.5224.676>.
- Husson, F., Josse, J., Pages, J., 2010. Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data. *Appl. Math. Dep.* 1–17.
- Jebari, S., Berndtsson, R., Uvo, C.Y., Bahri, A., 2007. Regionalizing fine time-scale rainfall affected by topography in semi-arid Tunisia. *Hydrol. Sci. J.* 52, 1199–1215.
- Jedidi, K., Bargaoui, Z., Benzarti, Z., 2008. Identification de zones pluviométriques homogènes d'un bassin versant à partir de l'information topographique. *Rev. Des. Sci. De. l'eau* 22 (3), 383–395.
- Jelassi, M., Gachon, P., Laprise, R., 2016. Occurrence, durée et intensité des précipitations simulées par deux modèles régionaux canadiens du climat sur la région du Maghreb. *Atmos. Ocean* 54 (5), 469–497. <https://doi.org/10.1080/07055900.2016.1228522>.
- Jiang, N., Griffiths, G., N Dirks, K., 2011. Linking New Zealand synoptic weather types to daily rainfall in Auckland. *Weather Clim.* 31, 50–66. <https://doi.org/10.2307/26169717>.
- Joo, J., Lee, J., Kim, J., Jun, H., Jo, D., 2014. Inter-event time definition setting procedure for urban drainagesystems. *Water* 6 (1), 45–58. <https://doi.org/10.3390/w6010045>.
- Kalteh, A., Hiorth, P., Bemdtsson, R., 2008. Review of the self-organizing map (SOM) approach in water resources: analysis, modelling and application. *Environ. Model. Softw.* 23 (7), 835–845. <https://doi.org/10.1016/j.envsoft.2007.10.001>.
- Kaufman, L., Rousseeuw, O., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- Kebaili-Bargaoui, Z., 1989. Occurrence des sécheresses pluviométriques sur le bassin de la Medjerda (Tunisie). *Rev. Des. Sci. De. l'eau* 2 (3), 429–447.
- Kohonen, T., 2013. Essentials of the self-organizing map. *Neural Netw.* 37, 52–65. <https://doi.org/10.1016/j.neunet.2012.09.01>.
- Ling, R.F., 1973. A computer-generated aid for cluster analysis. *Commun. ACM* 16 (6), 355–361.
- Marres, P., 1972. Louis Emberger. 1897-1969. Méditerranée, Deux. série, tome 9 (1–1972), 1–4. (https://www.persee.fr/doc/medit_0025-8296_1972_num_9_1_1414).
- Martinez, W., Martinez, A., 2005. *Exploratory data analysis with MATLAB*. Chapman & Hall/CRC Press, Florida, USA.
- Merzougui, A., Slimani, M., 2012. Régionalisation des lois de distribution des pluies mensuelles en Tunisie. *Hydrol. Sci. J.* 57 (4), 668–685. <https://doi.org/10.1080/02626667.2012.670702>.
- Minnotte, M.C., West, R., 1999. The data image: a tool for exploring high dimensional data set. *Am. Stat. Assoc.* 25–33.
- Molina-Sanchis, I., Lázaro, R., Arnau-Rosalén, E., Calvo-Cases, A., 2016. Rainfall timing and runoff: the influence of the criterion for rain event separation. *J. Hydrol. Hydromech.* 64, 226–236. <https://doi.org/10.1515/johh-2016-0024>.
- Murtagh, F., Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* 31, 274–295.
- Norrant, C., Douguedroit, A., 2006. Monthly and daily precipitation trends in the Mediterranean (1950–2000). *Theor. Appl. Climatol.* 83 (1), 89–106. <https://doi.org/10.1007/s00704-005-0163-y>.
- Parchure, A.S., Gedam, S.K., 2019. Self-organising maps for rain event classification in Mumbai City, India. *ISH J. Hydraul. Eng.* <https://doi.org/10.1080/09715010.2019.1581099>.
- Restrepo-Posada, P., 1982. Identification of independent rainstorms. *J. Hydrol.* 55 (1–4), 303–319. [https://doi.org/10.1016/0022-1694\(82\)90136-6](https://doi.org/10.1016/0022-1694(82)90136-6).
- Rysman, J.F., Verrier, S., Lemaître, Y., Moreau, E., 2013. Space-time variability of the rainfall over the western Mediterranean region: a statistical analysis. *J. Geophys. Res.: Atmospheres* 118 (15), 8448–8459. <https://doi.org/10.1002/jgrd.50656>.
- Schilling, J., Hertig, E., Tramblay, Y., Scheffran, J., 2020. Climate change vulnerability, water resources and social implications in North Africa. *Reg. Environ. Change* 20 (1). <https://doi.org/10.1007/s10113-020-01597-7>.
- Slimani, M., Cudennec, C., Feki, H., 2007. Structure du gradient pluviométrique de la transition Méditerranée-Sahara en Tunisie: déterminants géographiques et saisonnalité. *Hydrol. Sci. J.* 52 (6), 1088–1102. <https://doi.org/10.1623/hysj.52.6.1088>.
- Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B* 63 (2), 411–423. <https://doi.org/10.1111/1467-9868.00293>.
- Tramblay, Y., El Adlouni, S., Servat, E., 2013. Trends and variability in extreme precipitation indices over Maghreb countries. *Nat. Hazards Earth Syst. Sci.* 13 (12), 3235–3248. <https://doi.org/10.5194/nhess-13-3235-2013>.
- Tramblay, Y., Feki, H., Seguí, P.Q., Guijarro, J.A., 2019. The SAFRAN daily gridded precipitation product in Tunisia (1979–2015). *Int. J. Climatol.* 39 (15), 5830–5838. <https://doi.org/10.1002/joc.6181>.
- Tramblay, Y., Ruelland, D., Hanich, L., Bargaoui, Z., Dakhlaoui, H., 2020. Climate change impacts on water resources in North African basins. *EGU Gen. Assem.* 2020 <https://doi.org/10.5194/egusphere-egu2020-7162>.
- Vatanen, T., Osmala, M., Raiko, T., Lagus, K., Sysi-Aho, M., Orešič, M., Lähdesmäki, H., 2015. Self-organization and missing values in SOM and GTM. *Neurocomputing* 147 (1), 60–70. <https://doi.org/10.1016/j.neucom.2014.02.061>.
- Verrier, S., Mallet, C., Barthès, L., 2011. Multiscaling properties of rain in the time domain, taking into account rain support biases. *J. Geophys. Res.: Atmospheres* 116 (D20).
- Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. *IEEE Trans. Neural Netw.* 11, 586–600.
- Wanner, H., Brönnimann, D., Casty, S., Gyalistras, D., Luterbacher, J., Schmutz, C., Stephenson, D., 2001. North Atlantic oscillation-concepts and studies. *Surv. Geophys.* 22 (4), 321–381. <https://doi.org/10.1023/A:1014217317898>.
- Ward, J., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58 (301), 236–244.
- Wenzel, H.G., Voorhees, M.L., 1981. Evaluation of the urban design storm concept. Water Resources Center. University of Illinois at Urbana-Champaign, Illinois. (<http://hdl.handle.net/2142/90287>).
- Xoplaki, E., Gonzalez-Rouco, J.F., Luterbacher, J.U., Wanner, H., 2004. Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends. *Clim. Dyn.* 23 (1), 63–78.
- Xoplaki, E., Luterbacher, J., González-Rouco, J., 2006. Mediterranean summer temperature and winter precipitation, large-scale dynamics, trends. *Il Nuovo Cim.* 29 (1), 45–54. <https://doi.org/10.1393/ncc/i2005-10220-4>.
- Yahi, H., Marticorena, B., Thiria, S., Chatenet, B., Schmechtig, C., Rajot, J.L., Crepon, M., 2013. Statistical relationship between surface PM₁₀ concentration and aerosol optical depth over the Sahel as a function of weather type, using neural network methodology. *J. Geophys. Res. Atmos.* 118 (13) <https://doi.org/10.1002/2013JD019465>, 265–13, 281.
- Zappa, G., Hawcroft, M., Shaffrey, L., Black, E., Brayshaw, D., 2015. Extratropical cyclones and the projected decline of winter Mediterranean precipitation in the CMIP5 models. *Clim. Dyn.* 45 (7), 1727–1738. <https://doi.org/10.1007/s00382-014-2426-8>.