



HAL
open science

Impact du changement climatique dans les modèles numériques à l'échelle régionale

Sabrina Derouiche

► **To cite this version:**

Sabrina Derouiche. Impact du changement climatique dans les modèles numériques à l'échelle régionale. Milieux et Changements globaux. Université Paris-Saclay; Université de Tunis El Manar, 2022. Français. NNT : 2022UPASJ025 . tel-04582470

HAL Id: tel-04582470

<https://theses.hal.science/tel-04582470>

Submitted on 22 May 2024

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.

Impact du changement climatique dans les modèles numériques à l'échelle régionale

Climate change impacts on regional scale models

Thèse de doctorat de l'université Paris-Saclay et de l'université de Tunis El Manar

École doctorale n°129 : sciences de l'environnement d'Île-de-France (SEIF)

Spécialité de doctorat : Hydrologie

Graduate School : Géosciences, climat, environnement et planètes

Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans les unités de recherche **LATMOS (Université Paris-Saclay, UVSQ, CNRS)** et **LMHE (Université Tunis El Manar, ENIT)** sous la direction de **Cécile MALLET**, professeure, la co-direction de **Zoubeida BARGAOUI**, professeure

Thèse soutenue à Tunis, le 22 décembre 2022, par

Sabrina DEROUICHE

Composition du Jury

Membres du jury avec voix délibérative

Khelifa MAALEL

Professeur, École nationale d'ingénieurs de Tunis (ENIT)

Président

Habib ABIDA

Professeur, Faculté des sciences de Sfax (FSS)

Rapporteur et Examineur

N'deye NIANG

Maîtresse de conférences, Conservatoire National des Arts et Métiers (CNAM)

Rapportrice et Examinatrice

Yadh ZAHAR

Professeur, École Nationale d'Architecture et d'Urbanisme (ENAU)

Examineur

Titre : Impact du changement climatique dans les modèles numériques à l'échelle régionale

Mots clés : Pluies journalières, réseau pluviométrique, modèle de réanalyse, événement pluvieux, analyse multivariée, variabilité climatique.

Résumé : La région méditerranéenne est considérée une des zones les plus vulnérables au changement climatique vue sa richesse socio-culturelle et sa biodiversité. Plusieurs études élaborées sur l'évolution temporelle de différents paramètres météorologiques ont trouvé une variabilité climatique significative à la fin du XXe siècle. En plus, d'après les modèles de projection climatique, il est attendu que les changements climatiques s'intensifient sur toute la région vers la fin du XXIe siècle. Par conséquent, leurs impacts deviennent de plus en plus dangereux et coûteux.

La pluie est considérée comme la signature la plus sensible du climat pour l'Homme. Ainsi, son analyse et la caractérisation des régimes pluviométriques dans la région permet d'appréhender son évolution future. Cette recherche s'appuie surtout sur une base de données journalières des observations sur 70 stations pluviométriques sur une période de 50 ans (1960-2009) et sur une échelle régionale concernant tout le nord de la Tunisie. D'autre part, les pluies journalières calculées par des modèles de réanalyse ERA-Interim sont également considérées dans cette thèse. Ces données estimées ont l'avantage d'être complètes dans le temps et dans l'espace. Elles peuvent jouer un rôle important dans la compréhension de la variabilité climatique d'où il est primordial d'évaluer leur qualité par rapport aux observations.

Le traitement des données pluviométriques est également inédit puisqu'il s'agit d'analyser la variabilité spatiale et temporelle à l'échelle de structures pluvieuses organisées en événements pluvieux.

Bien que les précipitations, en Tunisie, ont été analysées par plusieurs hydrologues et géographes à différentes échelles allant de quelques minutes à des années, le découpage en épisodes pluvieux et épisodes secs proposé dans ce travail est original. Cette approche vise à prendre en considération le caractère intermittent de la pluie, qui est une des propriétés fondamentales des précipitations. L'agrégation des jours de pluie a conduit à envisager six descripteurs des événements pluvieux pour chaque point de mesure sur une période de 50 ans. L'espace multidimensionnel ainsi créé est analysé en première partie par une méthode factorielle classique (l'analyse en composantes principales ACP). Puis par la méthode de classification non-linéaire SOM (Self-Organizing Map) associée à une classification hiérarchique ascendante (CHA). Les deux approches (ACP et SOM+CHA) ont permis de comprendre les structures des données pluviométriques et d'en définir une typologie.

L'analyse en composantes principale ACP a conduit à résumer les descripteurs de pluie adoptés en trois composantes principales : La première représente un indicateur de quantité de pluie, la seconde l'intermittence des pluies pendant la saison et la troisième est un indicateur de structure (des événements longs de faible intensité contre des courts événements intenses). L'interprétation spatiale a divisé la zone d'étude en trois régions d'orientation NE-SO, avec une opposition entre la façade Nord-Ouest (Bizerte-Tabarka) et son arrière-pays et la façade Sud-Est (Nabeul-Hammamet) et son arrière-pays avec une zone intermédiaire située entre ces deux régions.

Par ailleurs, la thèse investigate les corrélations entre les composantes principales issues des ACPs et des indices climatiques dérivant les modes de variabilité climatiques. Des corrélations significatives ont été mises en évidence pour les indices Oscillation nord Atlantique (NAO) et Oscillation Arctique (AO). Aussi, des liens entre les anomalies de pression à la surface de la mer et les composantes principales ont été montrés par une analyse composite.

Les méthodes combinées (SOM et CHA) sont appliquées aux descripteurs de pluie associés au réseau de mesure et mettent en évidence 4 classes ayant des typologies différentes pour les structures d'épisodes pluvieux. Leur variabilité spatiale et temporelle a ensuite été analysée. Ces classes sont utilisées comme une référence pour l'analyse des données estimées par le modèle de réanalyse.

Les analyses univariées et multivariées des pluies du modèle et la comparaison avec les observations ont montré que le nombre de jours de pluie et la durée des événements sont fortement surestimés dans les réanalyses. En plus, les distributions statistiques n'ont pas la même asymétrie. Par contre, le modèle montre une bonne cohérence des structures temporelle des classes de pluie avec les observations à une échelle régionale. Ces résultats représentent une importante contribution à l'étude critique des données de réanalyse. Un prétraitement de pluie du modèle est primordial, particulièrement, dans l'étude des réponses hydrologiques.

Title : Climate change impacts on regional scale models

Keywords : Daily rainfall, rain gauge network, reanalysis model, rainfall event, multivariate analysis, climate variability

Abstract : The Mediterranean region is considered one of the most vulnerable areas to climate change because of its socio-cultural wealth and its biodiversity. Several studies about the evolution of different parameters found a significant climate variability at the end of the 20th century. Moreover, according to climate projection models, an intensification of this climate change is expected, over the region, to the end of the 21th century. Consequently, their impacts become more dangerous and expensive.

Rain is considered to be the most sensitive signature of climate for humans. Thus, its analysis and the characterization of rainfall regimes over the region allow to apprehend its future evolution. This study is mainly based on daily rainfall observations collected from 70 rain gauge stations over 50-year period (1960-2009) on a regional scale covering all of northern Tunisia. On the other hand, daily precipitations produced by ERA-Interim reanalysis model, equally at regional scale, are also considered in this study. These estimated data have the advantage of being complete over the time and the space. They can have an important role in understanding the climate variability, hence it is essential to assess their quality in relation to observations. The rainfall data processing is novel. Indeed, the analysis of spatial and temporal variability analysis was carried out on the rain event scale. Although rainfall in Tunisia has been analysed by several hydrologists and geographers at different scales ranging from a few minutes to years, the division into rainy episodes and dry episodes proposed in this analysis is original. This approach aims to take into account the intermittent nature of the rain which is one of the fundamental properties of the precipitations. The aggregation of rainy days led to consider six descriptors of rainy events for each measurement point over a period of 50 years.

The multidimensional space, thus, created was analysed in the first part by a classic factorial method PCA (Principal Component Analysis), then by the non-linear classification method SOM (Self-Organizing Map) combined with Hierarchical agglomerative clustering (HAC). The two approaches (ACP and SOM+CHA) allowed to understand the rainfall data structures and to define a typology.

The principal component analysis summarized the six rainfall descriptors adopted into three main components: the first one is an indicator of the rainfall quantity, the second one represents the intermittent character of rain over the season and the third one is a structure indicator (long and weak events against short and intense events).

The spatial interpretation divided the study area into three regions of NE-SWest orientation, with an opposition between the North-West facade (Bizerte-Tabarka) and its hinterland and the South-East facade (Nabeul-Hammamet) and its hinterland with an intermediate zone located between these two regions. Moreover, the thesis investigated the correlations between the principal components of PCAs and climate patterns indices. Significant correlations were found for the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) indices. In addition, relationships between sea surface pressure anomalies and principal components were shown by a composite analysis.

The combined methods (SOM and HAC) were applied to rainfall descriptors produced by the rain gauge stations network and highlighted 4 classes with different typologies of wet spells structure. Their spatial and temporal variability was, then, analysed. These classes were used as a reference for the analysis of the reanalysis data.

The univariate and multivariate analysis of the model data and the comparison with the observations showed that the number of rainy days and the duration of the events are significantly overestimated in the reanalyses. Moreover, the statistical distributions didn't have the same asymmetry. On the other hand, the model showed a good coherence of the temporal structures of the rainfall classes with the observations on a regional scale. These results represent an important contribution to the critical study of reanalysis data. A pre-processing of model rainfall data is essential, particularly, in the study of hydrological responses.

Remerciements

Je tiens à exprimer ma profonde reconnaissance à ma directrice de thèse Madame Cécile Mallet, Professeur à l'Université de Versailles Saint-Quentin-en-Yvelines, Chercheur au Laboratoire Atmosphères, Milieux, Observations Spatiales, équipe : Statistique Processus Aérosols Cycle de l'Eau (SPACE), pour son aide, suggestions et conseils qui m'ont été précieux pour que je puisse bien mener ce travail de recherche. J'ai pris un grand plaisir à travailler avec elle.

J'adresse ma profonde gratitude à ma co-directrice de thèse Madame Zoubeida Bargaoui, Professeur au département génie civil de l'Ecole Nationale d'Ingénieurs à Tunis, pour m'avoir proposé un sujet de thèse aussi intéressant qui vient dans le cadre d'un projet de coopération Franco-Tunisienne. Durant le temps de cette thèse, Mme Bargaoui m'a transmis ses connaissances et ses visions de recherche. Je suis très reconnaissante pour le temps qu'elle m'a accordé, ses qualités pédagogiques et scientifiques, sa compréhension et sa sympathie.

Je remercie ensuite l'ensemble des membres du jury, qui m'ont fait l'honneur de bien vouloir étudier avec attention mon travail : M. Habib Abida et Mme N'deye Niang pour avoir accepté d'être rapporteurs de cette thèse et pour m'avoir fourni une multitude de pistes de recherche ; M. Yadh Zahar pour avoir accepté d'examiner cette thèse; M. Safouane Moulelhi pour avoir accepté l'invitation à participer à ce jury ; et enfin Mr. Khelifa Maalel pour m'avoir fait l'honneur d'accepter de présider ce jury.

Je tiens aussi à remercier Monsieur Abdelwahab Hannachi, Professeur à l'Université de Stockholm et directeur des études doctorales au département de la météorologie (MISU) pour son accueil chaleureux à MISU, son aide et ses précieux conseils.

Je remercie également toute l'équipe SPACE du laboratoire LATMOS, spécialement Monsieur Laurent Barthès, Maître de conférences à l'Université de Versailles Saint-Quentin en Yvelines pour son aide et support pendant mon séjour au LATMOS. Je remercie particulièrement Dilmi Djalel du LATMOS pour sa patience, ces conseils et son support. Je remercie Mme Solène Turquety et Mme Marjolaine Chiriaco du LATMOS pour leur temps.

Je tiens à remercier les enseignants et les administratives du laboratoire LMHE pour leurs qualités et professionnalisme. Je remercie également les jeunes chercheurs du LMHE, particulièrement, Cyrine Abid, Aymen Ben Jaafar, Saoussen Dhib, Amel Bjeoui et Ines Boussetta pour les beaux moments que nous avons vécus ensemble.

Je tiens aussi à remercier la Ministère de l'Enseignement Supérieur et de Recherche Scientifique Tunisienne qui m'a accordé une bourse durant 3 ans (2017-2019) et une bourse d'un stage à l'université de Stockholm dans le cadre du programme bourse d'encouragement des jeunes chercheurs.

Je saisis cette occasion aussi pour remercier profondément la Direction Générale des Ressources en Eau (DGRE) pour sa collaboration et la fourniture des données pluviométriques.

A tous mes amis de l'ENIT ou d'ailleurs pour leur soutien moral et leurs encouragements. Enfin, à tous ceux qui ont participé de près ou de loin à la réalisation de ce travail je vous dis un grand merci.

Enfin, mes chaleureuses pensées pour toute ma famille de m'avoir soutenue et encouragée tout au long de ce parcours. Merci à mes parents Hedi Derouiche et Samia Mami, qui m'ont accompagné par leur soutien et leur affection au long de mon travail de thèse et depuis toujours. Un merci tout particulier à mon mari Mouhamed Amine Souilem pour sa patience, son encouragement, et son soutien qui m'a permis de surmonter les obstacles dans les situations les plus difficiles. Un merci particulier à mes sœurs Khaoula et Khouloud d'être toujours présentes et à mes frères Khaled et Fraj. Et un merci spécial à ma petite fille Salma d'être ma source de force, d'inspiration et de bonheur.

Merci à Tous !

Table des matières

INTRODUCTION GENERALE	1
I CHAPITRE I : REVUE BIBLIOGRAPHIQUE	9
<i>I.1 Introduction</i>	<i>10</i>
<i>I.2 Echelle spatio-temporelle des processus hydrologiques.....</i>	<i>10</i>
<i>I.3 L'analyse par événement pluvieux.....</i>	<i>12</i>
I.3.1 Définition d'un événement pluvieux et des seuils pluviométriques.....	12
I.3.2 Approches de détermination de la durée minimale entre deux évènements pluvieux indépendants (MIET) 13	
<i>I.4 Analyses descriptives multivariées en hydrologie</i>	<i>14</i>
<i>I.5 Analyses de téléconnexions dans la région Méditerranéenne et en Tunisie</i>	<i>17</i>
<i>I.6 Conclusion</i>	<i>19</i>
II CHAPITRE II : BASE DE DONNEES HYDRO-CLIMATIQUES	20
<i>II.1 Introduction</i>	<i>21</i>
<i>II.2 Caractéristiques climatiques de la zone d'étude.....</i>	<i>21</i>
<i>II.3 Flux pluvieux et influence des reliefs.....</i>	<i>22</i>
II.3.1 Flux pluvieux.....	22
II.3.2 Influence des reliefs sur la propagation des flux pluvieux.....	24
<i>II.4 Base de données pluviométriques (in Situ)</i>	<i>24</i>
II.4.1 Stations retenues pour l'étude	24
II.4.2 Choix de la saison d'étude	28
<i>II.5 Simulations numériques et modèle climatiques.....</i>	<i>31</i>
II.5.1 Principe de simulation numérique et modèles climatiques globaux GCM et régionaux RCM	31
II.5.2 Modèle de réanalyse météorologique	33
II.5.3 Données fournies par les ré-analyses météorologiques : ERA-Interim	34
<i>II.6 Données de la variabilité climatique.....</i>	<i>35</i>
II.6.1 Choix des indices climatique	35
II.6.2 Données climatiques de large échelle : Pression à la surface de la mer (SLP)	37
<i>II.7 Conclusion</i>	<i>37</i>
III CHAPITRE III : METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX	39
<i>III.1 Introduction</i>	<i>40</i>
<i>III.2 Transformation des données journalières en événement pluvieux.....</i>	<i>40</i>
III.2.1 Délimitation des événements pluvieux par l'analyse d'autocorrélation	40
III.2.2 Choix de descripteurs de pluie par saison.....	42
<i>III.3 Analyse statistique univariée des descripteurs de pluie</i>	<i>43</i>
III.3.1 Analyse statistique des descripteurs des observations (pluviomètres).....	43
III.3.1.1 Distribution des descripteurs des stations pluviométriques	43
III.3.1.2 Variabilité temporelle des descripteurs des observations	44
III.3.1.3 Variabilité spatiale des descripteurs des observations.....	47
III.3.2 Analyse statistique des descripteurs produits par le modèle de réanalyse	53
III.3.2.1 Distribution des descripteurs de réanalyse ERA-Interim.....	53

III.3.2.2	Variabilité spatiale des descripteurs du modèle ERA-Interim	55
III.3.2.3	Variabilité temporelle des descripteurs ERA- Interim	57
III.4	<i>Méthodologie d'analyse multivariée des descripteurs d'événements pluvieux</i>	58
III.4.1	Approche factorielle linéaire : Analyse en composantes principales	58
III.4.2	Approche non linéaire SOM et CAH	59
III.5	<i>Conclusion</i>	60
IV	CHAPITRE IV : OUTILS UTILISE.....	62
IV.1	<i>Introduction</i>	63
IV.2	<i>Analyse des données par une méthode factorielle : Analyse en composantes principales (ACP)</i> ..	63
IV.3	<i>Analyse par classification non supervisées : aperçu sur les méthodes de clustering utilisées</i>	63
IV.3.1	Classification non paramétrique et mesure de similarité.....	63
IV.3.2	Les cartes auto-organisatrices (Self-Organizing Map : SOM)	64
IV.3.2.1	Les étapes d'apprentissage.....	65
IV.3.2.2	La fonction de voisinage.....	66
IV.3.2.3	Évaluation de la qualité de la carte	67
IV.3.2.4	Les paramètres d'apprentissage.....	68
IV.3.2.5	Projection des variables et individus supplémentaires sur la carte topologique	69
IV.3.2.6	Récapitulation des étapes de la SOM.....	70
IV.3.3	La classification hiérarchique ascendante (CHA)	72
IV.3.3.1	Critères d'agrégation.....	72
IV.3.3.2	Récapitulation des étapes de l'algorithme de CHA.....	73
IV.3.3.3	Choix du nombre de classes.....	74
IV.3.3.4	Dendrogramme	74
IV.3.4	Data image.....	75
IV.4	<i>Téléconnexions</i>	76
IV.4.1	Analyse de corrélation linéaires : corrélation de Pearson et significativité.....	76
IV.4.2	Analyse composite	77
IV.5	<i>Conclusion</i>	77
V	CHAPITRE V : ANALYSE FACTORIELLE DES DESCRIPTEURS ET TELECONNEXIONS CLIMATIQUES	79
V.1	<i>Introduction</i>	80
V.2	<i>Analyse des descripteurs de pluie par une analyse en composantes principales ACP</i>	80
V.2.1	Comblement des valeurs manquantes.....	80
V.2.2	Lecture de la matrice de corrélation	81
V.2.3	Part de la variance expliquée par les facteurs principaux	82
V.2.4	Analyse de corrélation variables-composantes principales	83
V.2.5	Analyse spatio-temporelle des composantes principales retenues.....	85
V.2.5.1	Analyse spatio-temporelle de la première composante principale CP1 : quantité de pluie.....	85
V.2.5.2	Analyse spatio-temporelle de la deuxième composante principale CP2 : intermittence de pluie..	88
V.2.5.3	Analyse spatio-temporelle de la troisième composante principale CP3 : conditions météorologiques	91
V.2.6	Récapitulation de l'analyse en composantes principales	95
V.3	<i>Téléconnexion climatiques</i>	95

V.3.1	Corrélations des indices climatiques avec les composantes principales des descripteurs pluviométriques	95
V.3.2	Analyse composite de SLP pour CP1, CP2 et CP3.....	99
V.3.2.1	Analyse composite de SLP avec les axes principaux de CP1 : quantité de pluie.....	103
V.3.2.2	Analyse composite de SLP avec les axes principaux de CP2 : Intermittence d'événements	103
V.3.2.3	Analyse composite de SLP avec le premier axe de CP3 : conditions météorologiques	104
V.4	<i>Conclusion</i>	104
VI	CHAPITRE VI : REGIONALISATION SPATIALE ET TEMPORELLE DES DESCRIPTEURS DE PLUIE PAR CLASSIFICATION ET EVALUATION DU MODELE NUMERIQUE DE REANALYSE ERA-INTERIM	107
VI.1	<i>Introduction</i>	108
VI.2	<i>Application de la SOM sur les descripteurs de pluie observés et choix des paramètres d'apprentissage de la carte</i>	108
VI.3	<i>Classification des saisons DJF dans les stations pluviométriques</i>	108
VI.3.1	Classification des cartes auto-organisatrice SOM.....	108
VI.3.2	Classification hiérarchique et détermination du nombre de classes.....	111
VI.3.3	Identification et analyse des typologies des saisons DJF par CAH	113
VI.3.4	Variabilité spatio-temporelle des classes	116
VI.3.4.1	Variabilité spatiale des classes dans les stations pluviométriques	116
VI.3.4.2	Variabilité temporelle des classes	118
VI.4	<i>Évaluation de modèle de réanalyse Era-Interim</i>	120
VI.4.1	Affectation des saisons DJF des mailles ERA-Interim sur la carte auto-organisatrice.....	120
VI.4.2	Comparaison de la variabilité temporelle des classes	121
VI.4.3	Comparaison de la variabilité spatiale des classes	123
VI.5	<i>Conclusion</i>	126
	CONCLUSIONS ET PERSPECTIVES	127
	REFERENCES	131
	ANNEXES	

Annexes

Annexe A : les indices climatiques.....	1
Annexe B : Analyse en composantes principales ACP.....	12
Annexe C : Nuage des points et corrélation entre les indices climatiques et les composantes des précipitations	20
Annexe D : Test de Student de différence entre les descripteurs des classes (p-value)	22

Liste des figures

Figure I-1 Echelles spatiales et temporelles des processus hydrologiques correspondants (Tavakolifar et al, 2017)	11
Figure II-1 Localisation de la zone d'étude et les étages bioclimatiques	22
Figure II-2 Schéma de synthèse pour les différents points d'actions influençant le climat tunisien. (A: Anticyclone, D: Dépression, Flèches: Direction des Flux)	23
Figure II-3 L'orientation des principaux reliefs de la zone d'étude	24
Figure II-4 Localisation des stations pluviométriques dans la zone d'étude	28
Figure II-5 Moyenne mensuelle de précipitations sur 50 ans (1959-2008) dans 8 stations pluviométriques réparties dans la zone d'étude	29
Figure II-6 Période couverte par des séries journalière DJF complètes dans les 70 stations	30
Figure II-7 Anomalies moyenne du cumul saisonnier DJF dans les 70 stations entre 1960-2009	30
Figure II-8 Illustration des différentes composantes du système climatique et leurs interactions (IPCC, 2007)	32
Figure II-9 Schématisation de la grille tridimensionnelle du modèle climatique (source : https://news.climate.columbia.edu/2018/05/18/climate-models-accuracy/)	32
Figure II-10 Représentation de la zone d'étude dans le modèle ERA-Interim	34
Figure II-11 Séries chronologiques des indices climatiques saisonniers DJF, SON, JJA et MAM entre 1959 et 2008	37
Figure III-1 (a) Moyenne et médiane des coefficients d'autocorrélation de pluie journalière à 70 stations pour 5 pas du temps. (b) Distribution des coefficients d'autocorrélation journalière à 70 stations pour 5 pas de temps	41
Figure III-2 Principe de séparation des événements pluvieux. IET (inter-event time) est la période sans pluie	42
Figure III-3 Histogrammes de distribution des descripteurs de pluie (Nombre d'observations=3500) et illustration des valeurs moyenne et médiane	44
Figure III-4 Variabilité temporelle de distribution des variables (a) N_e , (b) N_j et (c) C_t dans 70 stations entre 1960 et 2009 et sur trois périodes (1960-2009 ; 1960-1990 ; 1991-2009)	46
Figure III-5 Variabilité temporelle de distribution des variables (a) $C_{m,e}$, (b) $D_{m,e}$ et (c) $C_{m,j}$ dans 70 stations entre 1960 et 2009 et sur trois périodes (1960-2009 ; 1960-1990 ; 1991-2009)	47
Figure III-6 Variabilité spatiale des moyennes des descripteurs (a) N_e et (b) N_j , (c) C_t , $C_{m,e}$, $D_{m,e}$ et $C_{m,j}$ entre 1960-2009	49
Figure III-7 Variabilité spatiale des moyennes des descripteurs (a) N_e et (b) N_j , (c) C_t , (d) $C_{m,e}$, (e) $D_{m,e}$ et (f) $C_{m,j}$ entre 1960-1990	51
Figure III-8 Variabilité spatiale des anomalies des descripteurs (a) N_e et (b) N_j , (c) C_t , (d) $C_{m,e}$, (e) $D_{m,e}$ et (f) $C_{m,j}$ entre 1960-1990 et 1991-2009	52
Figure III-9 Histogrammes de distribution des descripteurs de pluie (DJF) des données de modèle de réanalyse ERA-Interim entre 1980 et 2009. Nombre d'individus : 270 (9 mailles x 30 ans)	54
Figure III-10 Variabilité spatiale de distribution des descripteurs (a) N_e , (b) N_j et (c) C_t sur 30 ans (1980-2009)	56
Figure III-11 Variabilité spatiale de distribution des descripteurs (a) $C_{m,e}$, (b) $D_{m,e}$ et (c) $C_{m,j}$ sur 30 ans (1980 et 2009)	56
Figure III-12 Variabilité temporelle de distribution des descripteurs (a) N_e , (b) N_j et (c) C_t dans les 9 mailles de ERA-Interim entre 1980 et 2009	57

Figure III-13 Variabilité temporelle de distribution des descripteurs (a) $C_{m,e}$, (b) $D_{m,e}$ et (c) $C_{m,j}$ dans les 9 mailles de ERA-Interim entre 1980 et 2009.	58
Figure III-14 Schéma de la méthodologie d'analyse multivariée par l'analyse en composantes principales.....	59
Figure III-15 Schéma de la méthodologie d'analyse multivariée des saisons DJF par l'approche non linéaire.....	60
Figure IV-1 Architecture de la carte auto-organisatrice	65
Figure IV-2 Illustration du principe de projection des observations sur la carte de kohonen (Hajjar, 2014, modifiée)	66
Figure IV-3 Illustration de la fonction de voisinage gaussienne	67
Figure IV-4 Type d'architecture de la carte	69
Figure IV-5 les étapes de l'algorithme de la SOM.....	71
Figure IV-6 Les différentes étapes de l'algorithme de la classification hiérarchique ascendante	74
Figure IV-7 Schéma illustratif du dendrogramme des données fictives	75
Figure V-1 Répartition des stations retenues pour l'analyse en composantes principales. Les stations en rouge sont retenues (44 stations) et les stations en noir sont éliminées	81
Figure V-2 (a) Projection des variables (descripteurs pluviométriques) sur le plan factoriel 1-2, (b) sur le plan factoriel 1-3	84
Figure V-3 Séries chronologiques de la première composante principale CP1 (quantité de pluie) dans 44 stations entre 1960 et 2009. Le trait noir en gras représentent la moyenne des 44 stations par an. d'étude. Les boîtes à moustache représente la variabilité des composantes des 44 stations sur trois périodes (1960-2009 ;1960-1990 ;1991-2009).....	85
Figure V-4 Pourcentage d'inertie et d'inertie cumulée expliquées par les axes principaux de l'analyse de CP1(50 ans x 44 stations)	86
Figure V-5 Corrélation des stations avec les axes principaux de CP1 (à gauche) et répartition des stations corrélées et anti-corrélées avec l'axe 2 (à droite). Les stations en gris sont faiblement corrélées avec le premier plan factoriel.....	87
Figure V-6 Projection des années (50 ans) sur le premier plan factoriel de CP1. La taille de triangle représente la qualité de représentation. Les couleurs sont des indicateurs des années de plus anciennes (bleu) au plus récentes (rouge)	88
Figure V-7 Séries chronologiques de la deuxième composante principale CP2 (intermittence de pluie) dans 44 stations entre 1960 et 2009. Le trait noir en gras représente la moyenne des 44 stations par an. Les boîtes à moustache représentent la variabilité des composantes des 44 stations sur trois périodes (1960-2009 ;1960-1990 ;1991-2009).....	89
Figure V-8 Pourcentage d'inertie et d'inertie cumulée expliquées par les axes principaux de l'analyse de CP2 (50 ans x 44 stations)	90
Figure V-9 Corrélation des stations avec les axes principaux de CP2 (à gauche) et répartition des stations corrélées et anti-corrélées avec l'axe 2 (à droite). Les stations en gris sont faiblement corrélées avec le premier plan factoriel.....	90
Figure V-10 Projection des années (50 ans) sur le premier plan factoriel de CP2. La taille de triangle représente la qualité de représentation. Les couleurs sont des indicateurs des années de plus anciennes (bleu) au plus récentes (rouge)	91
Figure V-11 Séries chronologiques de la première composante principale CP3(conditions météorologiques) dans 44 stations entre 1960 et 2009. Le trait noir en gras représente la moyenne des 44 stations par an. Les	

boîtes à moustache représentent la variabilité des composantes des 44 stations sur trois périodes (1960-2009 ;1960-1990 ;1991-2009).....	92
Figure V-12 Pourcentage d'inertie et d'inertie cumulée expliquées par les axes principaux de l'analyse de CP3 (50 ans x 44 stations)	93
Figure V-13 Corrélation des stations avec les axes principaux de CP3 (à gauche) et répartition des stations corrélées et anti-corrélées avec l'axe 2 (à droite). Les stations en gris sont faiblement corrélées avec le premier plan factoriel.....	94
Figure V-14 Projection des années (50 ans) sur le premier plan factoriel de CP3. La taille de trinagle représente la qualité de représentation. Les couleurs sont des indicateurs des années de plus anciennes (bleu) au plus récentes (rouge)	94
Figure V-15 Diagramme synthétique des résultats des deux ACP et la variance expliquée des axes retenus	95
Figure V-16 Analyse composite de SLP et les facteurs principaux (CP1-1 et CP1-2) de CP1. (a) et (b) Identification des fortes (+) et des faibles (-) anomalies de CP1-1 et CP1-2. (c) et (d) SLP moyenne correspondant aux années de fortes anomalies de CP1-1 et CP1-2 noté SLP (+). (e) et (f) SLP moyenne correspondant aux années de faibles anomalies de CP1-1 et CP1-2 noté SLP (-). (g) et (f) les anomalies de SLP (SLP(+)-SLP(-)) de CP1-1 et CP1-2 (seulement les anomalies significatives à 90 %, selon le test de Student, sont visualisées)	100
Figure V-17 Analyse composite de SLP et les facteurs principaux (CP2-1 et CP2-2) de CP2. (a) et (b) Identification des fortes (+) et des faibles (-) anomalies de CP2-1 et CP2-2. (c) et (d) SLP moyenne correspondant aux années de fortes anomalies de CP2-1 et CP2-2 noté SLP (+). (e) et (f) SLP moyenne correspondant aux années de faibles anomalies de CP2-1 et CP2-2 noté SLP (-). (g) et (f) les anomalies de SLP (SLP(+)-SLP(-)) de CP2-1 et CP2-2 (seulement les anomalies significatives à 90 %, selon le test de Student, sont visualisées)	101
Figure V-18 Analyse composite de SLP et les composantes CP3-1. (a) les forte et les faibles anomalies du premier facteur principal de CP3. (b) SLP moyenne correspondant aux années de fortes anomalies de CP3-1 noté SLP(+). (c) SLP moyenne correspondant aux années de faibles anomalies de CP3-1 respectivement noté SLP(-). (d) les anomalies de SLP (SLP(+)-SLP(-)) de CP3-1 (seulement les anomalies significatives à 90 %, selon le test de Student, sont visualisées)	102
Figure VI-1 (a) Projection des six variables sur la carte topologique et matrice des distances entre neurones (U-matrix). (b) Nombre des individus (année/station) représentés par chaque neurone de la carte.....	110
Figure VI-2 (a) Dendrogramme. Les traits horizontaux de couleur verte, noire et rouge sont les niveaux de coupure à 2,3 et 4 classes respectivement. (b) data image de matrice de distance entre les vecteurs référents deux à deux (c) matrice de distance réorganisée en deux classes, (d) matrice de distance réorganisée en trois classes, (e) matrice de distance réorganisée en quatre classes	112
Figure VI-3 (a) classification des neurones dur la carte topologique. (b) Délimitation des classes sur les cartes de projection des variables de prototypes.....	113
Figure VI-4 Etude de distribution des descripteurs étudiés pour les 4 classes (très sèche, sèche, humide et très humide).	114
Figure VI-5 Distribution des quatre classes (très sèche , sèche, humide et très humide) dans les 70 stations pluviométriques par des graphiques en camembert	117
Figure VI-6 Distribution des quatre classes par année (dans les 70 stations).	119

Figure VI-7 Anomalies de pourcentage d'occurrence de chaque classe sur 5 décades non chevauchantes (1960-1969 ; 1970-1979 ; 1980-1989 ; 1990-1999 ; 2000-2009) . (b) dans la région humide du nord-ouest.(c) dans la région sèche (nord-est et sud).....	120
Figure VI-8 Projection des saisons ERA-Interim (individus supplémentaires) sur la carte topologique. La ligne noire en gras indique les limites des 4 classes	121
Figure VI-9 Répartition des 4 classes pour chaque année (1980-2009)	122
Figure VI-10 Variabilité décennale des anomalies d'occurrence de la classe 1 (a), classe 2 (b), classe 3 (c) et classe 4 (d) pour les données de réanalyse ERA-interim et pour les observations	123
Figure VI-11 (a) Répartition des classes de ERA Interim pour chaque maille (1980-2009).(b) Répartition des classes des stations pluviométriques dans chaque maille (1980-2009).....	125

Liste des tableaux

Tableau II-1	Caractéristiques des stations pluviométriques utilisées	25
Tableau II-2	Répartition des stations dans les mailles	35
Tableau II-3	Tableau récapitulatifs des indices climatiques utilisés (description et source)	35
Tableau III-1	Définition des descripteurs pluvieux saisonniers (pour chaque station pluviométrique), Xi représente le cumul d'un jour i et Dk représente la durée de l'événement k	43
Tableau III-2	Comparaison entre les descripteurs de pluie des observations (70 x 30 = 2100 saisons DJF) et des données ERA Interim (9 x 30 = 270 saisons DJF) entre 1980 et 2009. (vert : approximativement symétrique, jaune : distribution modérément dissymétrique, rose : distribution fortement dissymétrique)	54
Tableau V-1	Matrice de corrélation des six variables en jeu (50 ans x 44 stations=2200 individus). Les coefficients en gras représentent des corrélations significatives à 95 %	82
Tableau V-2	Valeurs propres et pourcentage de la variance expliquée par chaque axe factoriel (nombre d'observations : 2200, nombre de variables : 6).....	82
Tableau V-3	Contribution et qualité de représentation de chaque variable à la formation des trois premiers axes factoriels (en %).....	84
Tableau VI-1	Paramètres adoptés pour l'entraînement de la carte	108
Tableau VI-2	Caractéristiques des quatre classes (TFa : très faible, Fa : faible, M : moyen, Fo : fort, TFo : très fort)	114
Tableau VI-3	Moyenne et écart type de descripteurs pour les 4 classes	115
Tableau VI-4	Pourcentage des quatre classes sur la période de 50 ans (1960-2009) et sur les sous périodes (1960-1990 et 1991-2009).....	118

Liste des abréviations

ACP	Analyse en composantes principales
AO	Arctic oscillation
CHA	Classification hiérarchique ascendante
DGRE	Direction générale des ressources en eaux
ENSO	El Niño southern oscillation
GES	Gaz à effet de serre
MIET	Minimum inter-event time
MO	Mediterranean oscillation
NAO	North Atlantic oscillation
NOAA	Agence américaine d'observation océanique et atmosphérique
PDO	Pacific decadal oscillation
MCG	Modèles climatiques globaux
MCR	Modèles climatiques régionaux
SLP	Sea level pressure
SOI	Southern oscillation index
SOM	Self-Organizing Map
WeMO	Western Mediterranean oscillation

Introduction générale

La thèse s'intitule "**impact du changement climatique dans les modèles numériques à l'échelle régionale**". Il s'agit d'un titre générique que nous essayons de motiver dans ce qui suit.

Notre point de départ sont les émissions de gaz à effet de serre. On trouve dans le 4^{ème} rapport de synthèse (2007) de l'IPCC (Intergovernmental Panel on Climate Change) que « l'essentiel de l'élévation de la température moyenne du globe observée depuis le milieu du XX^{ème} siècle est très probablement attribuable à la hausse des concentrations des gaz à effet de serre (GES) anthropiques ». L'atmosphère terrestre produit naturellement un effet de serre parce que les gaz qui la composent (vapeur d'eau H₂O, dioxyde de carbone CO₂, méthane CH₄, protoxyde d'azote N₂O et ozone O₃) absorbent une partie du rayonnement infrarouge émis par la Terre. Toutefois les activités humaines accentuent l'effet de serre naturel. En effet, la combustion de combustibles fossiles génère des émissions de gaz à effet de serre (dioxyde de carbone notamment comptant pour 65% de l'effet de serre anthropique). Le défrichement des terres et des forêts libère du dioxyde de carbone. Les activités des cimenteries et de l'élevage des ruminants sont responsables de l'augmentation des émissions de méthane qui est le deuxième gaz (15%) responsable de l'augmentation de l'effet de serre additionnel (Jancovici, 2007). Le changement climatique est ainsi d'abord une question liée à la trop forte émission des gaz à effet de serre. Les émissions mondiales de gaz à effet de serre (GES) ont doublé entre 1970 et 2017. Entre 1990 et 2017, elles ont augmenté de plus de 40%. En 2017 les émissions de carbone qui constituent les trois quarts des émissions de GES étaient estimées à 53.5 milliards de tonnes équivalent CO₂ (INSEE, 2020). Face à cette urgence mondiale et en vue de lutter contre le changement climatique, l'accord de Paris (2015) sur le climat vise à réduire les émissions de gaz à effet de serre des pays signataires de la Convention.

Bien que la Tunisie ne soit pas un pays particulièrement émetteur, nous subissons les impacts des excès d'émissions de GES. Notre thèse s'inscrit dans le cadre de l'analyse des impacts du changement climatique en Tunisie. Il est très important de souligner que de telles études font nécessairement appel à l'analyse des observations météorologiques y compris les données pluviométriques sur lesquelles nous nous focaliserons.

En effet, comme le souligne Guillemot (2009) « les modèles de météorologie comme de climat n'auraient pas pu se développer sans les puissants réseaux d'observations terrestres, océaniques et spatiales mis en place par des institutions scientifiques internationales puissantes... ». Les données pour observer le climat sont réparties en général dans trois catégories : celles provenant

d'instruments terrestres et maritimes, d'instruments à bord de satellites et données collectées lors de campagnes de terrain ciblées. Comme le souligne Guillemot (2009) « une caractéristique essentielle des données d'où qu'elles proviennent est qu'elles sont éminemment indirectes et reconstruites ». Nous traiterons dans cette thèse de données pluviométriques, publiées par la Direction Générale des Ressources en eau. Ces données proviennent de l'observation des pluies par pluviomètres, qui est une observation indirecte vu que le contenu d'eau collecté par le pluviomètre est versé dans une burette graduée, étalonnée pour lire directement en mm le volume recueilli. Nous étudierons une longue série constituée de 5 décennies de pluviométries journalières observée sur un réseau de 70 stations entre 1959 et 2009. Il s'agit de données aux stations pluviométriques, qui ont fait l'objet de contrôle d'homogénéité avant leur diffusion publique. Elles possèdent une cohérence spatiale car provenant de systèmes précipitants ainsi une dynamique liée à une situation synoptique donnée. Nous étudierons les structures internes des événements pluvieux pour voir si elles ont été impactées par le changement climatique. Ces données contiennent forcément des lacunes comme toute donnée faisant partie d'un protocole expérimental. Nous adopterons deux procédures d'analyse, avec et sans prise en compte des lacunes. D'autre part, nous avons pour objectif d'analyser statistiquement ces données, et d'en faire un jeu de référence pour évaluer les données de réanalyse.

Quel est l'intérêt des données de réanalyse ? Les données de réanalyse sont constituées de séries de données « retraitées et complétées » à partir de modèles du climat. Il s'agit de données modèles. Ces données ont l'avantage d'être complètes (sans lacunes) dans le temps et dans l'espace. Pour cette raison, elles sont largement utilisées en climatologie et en hydrologie. Toutefois, comme elles proviennent de modèles, elles ont besoin d'être validées. A notre connaissance, il n'y a pas eu de validation de données de réanalyse pluviométrique pour le cas de la Tunisie, ce qui représente une motivation supplémentaire de notre étude.

L'étude des impacts du changement climatique est essentiellement basée sur la modélisation. Selon le 4^{ième} rapport de l'IPCC « seuls les modèles qui tiennent compte des forçages anthropiques parviennent à simuler les configurations du réchauffement observé et leurs variations. Aucun modèle couplé du climat mondial ne tenant compte que des seuls forçages naturels n'a reproduit les tendances au réchauffement... ». Ainsi, la conséquence de l'augmentation des émissions de gaz à effet de serre est l'augmentation des températures à la surface de la Terre. Ces changements de température ont un effet sur les variables des cycles hydrogéochimiques c'est à dire sur l'environnement. Selon le 4^{ième} rapport de l'IPCC (2007) « les observations effectuées depuis 1961 montrent que la température moyenne des océans du monde a augmenté et que les océans absorbent plus que 80% de la chaleur ajoutée au système

climatique. Un tel réchauffement entraîne la dilatation de l'eau de mer, contribuant à la montée du niveau des mers ». L'océan joue ainsi un rôle primordial, dans les changements dus à l'augmentation des émissions des GES. C'est pourquoi, nous envisageons dans le cadre de cette thèse d'utiliser des indices de variabilité climatique issus de l'observation du géopotential, de la température et de la pression à la surface des océans afin de voir si les données pluviométriques analysées sont en cohérence (en téléconnexion) avec les indices de variabilité climatique. Nous utiliserons des pluies journalières et à notre connaissance il n'y a pas eu d'étude de téléconnexions entre les descripteurs de la structure des pluies journalières en Tunisie.

Les modèles intégrés du changement climatique cherchent à « capturer la chaîne causale qui conduit des déterminants des émissions de gaz à effet de serre aux impacts du changement climatique et à leurs répercussions socio-économiques » (Ambrosi et Courtois, 2004). On trouve dans Tol (2006) un historique de l'évaluation (assessment) intégrée des impacts du changement climatique, dont les travaux pionniers datent du début des années 1990. Les modèles d'évaluation intégrée (integrated assessment models) combinent des connaissances de plusieurs disciplines pour l'aide à la décision. Il y a une distinction entre deux grands types de modèles : policy-optimisation et policy-evaluation models. Policy evaluation models sont des exercices de type « what-if » qui modélisent les conséquences de politiques concernant l'émission des GES sur l'environnement et le bien-être. Les modèles de type policy-optimisation sélectionnent les meilleures politiques issues d'exercices de type what-if. Selon Tol (2006) les modèles de type policy-evaluation sont élaborés par des scientifiques des sciences physiques et biologiques alors que les modèles d'optimisation le sont par des économistes. Dans cette thèse, nous nous intéressons aux modèles géophysiques qui représentent les processus physiques des cycles hydro géophysiques et de la circulation atmosphérique générale et du climat.

La modélisation du cycle du carbone est au cœur de l'étude des impacts du changement climatique. Ce qu'on appelle les « Earth system models », ont pour objectif de représenter (modéliser) le système Terre. Ils utilisent des couplages de modèles de circulation générale atmosphérique et océanique avec la modélisation du cycle du carbone sur les terres et les océans (Döscher et al., 2022). Ces modèles sont forcés par les émissions de CO₂. Les couplages concernent la modélisation de la banquise, et de la végétation et la chimie de l'atmosphère (Gillemot, 2009). Les modèles simulent les réponses en termes de bilan hydrique (contenu en eau du sol, évapotranspiration réelle), production nette primaire (NPP) et phénologie. Les modèles sont « évalués », c'est-à-dire qu'on évalue leur performances ou capacité à reproduire

un phénomène climatique particulier (Guillemot, 2009) pour la période historique. Puis, ils sont utilisés pour simuler sous différentes hypothèses (scénarios) les impacts globaux et régionaux (les Tropiques et l'hémisphère Sud, les latitudes moyennes et hautes dans l'hémisphère nord) sur leurs écosystèmes. Tout cela pour insister sur le fait que la variabilité régionale est analysée par modélisation et qu'elle est déterminée par la variabilité globale. Nous étudierons les observations pluviométriques journalières et les réanalyses de pluie, en nous focalisant sur le Nord de la Tunisie, c'est-à-dire dans un contexte, qui est la région nord Afrique-Sud Méditerranée. C'est pour cela que l'intitulé mentionne « les modèles numériques à l'échelle régionale ».

Les données de réanalyse sont très importantes d'un point de vue opérationnel. Pour réaliser l'évaluation de l'état du climat l'Organisation Météorologique Mondiale (OMM) et l'IPCC utilisent largement les données de réanalyse. Comme indiqué plus haut, nous allons utiliser des données pluviométriques de réanalyse, en plus des données observées in-situ et publiées par la Direction Générale des ressources en eau. Les données de réanalyses se basent sur un modèle de prévision (forecasting model), une méthode d'assimilation des données (data assimilation) et des observations qui sont assimilées. On trouve dans Dee et al. 2011 une synthèse des données et modèles utilisés dans le cas des données de réanalyse ERA-Interim du Centre européen pour les prévisions météorologiques à moyen terme (Center for Medium-range Weather Forecasts ECMWF). Dans ce type de modèles de réanalyse, la prévision utilise un modèle physique de l'atmosphère (nuages et convection, couche limite, effets orographiques et de surface dus à la rugosité de la végétation, ozone, transferts radiatifs, aérosols, etc...). Il y a également un modèle de bilan radiatif et hydrique à la surface du sol et un modèle d'océan. Les observations utilisées dans ces modèles proviennent de nombreux capteurs mis en place dans les avions et les bateaux et aussi des stations météorologiques des différents pays (in-situ data) incluant des observations de radiosondages, des données satellitaires et des climatologies (statistiques des observations conventionnelles au sol). Parmi les observations satellitaires, on peut citer le rayonnement solaire par ciel clair, profils d'ozone, couverture neigeuse, données altimétriques etc. Parmi les observations conventionnelles, on cite la température de l'air, le vent, l'humidité spécifique, pression à la surface, humidité relative (Dee et al., 2011). Les émissions de GES sont introduites dans les modèles de réanalyse comme forçages selon les historiques observés.

Le centre européen ECMWF améliore continuellement ses produits de réanalyse. Ils sont passés de EAR-15 à ERA-40 (1957-2002) (Uppala et al., 2005) puis à Era-Interim en améliorant la représentativité temporelle des champs réanalysés et en changeant de modèle de Méthode

d'assimilation (de 3D-Var à 4D-Var) (Dee et al. 2011). La section 3.4.1.3 consacrée aux observations de l'atmosphère et faisant partie du chapitre observations : surface and atmospheric climate change, du 4^{ème} rapport de l'IPCC se réfère aux données de réanalyses ERA-40. Toutefois ERA-40 surestiment les précipitations sur les océans tropicaux (Uppala et al., 2005). ERA-Interim a pris fin en 2019 (1979-2019) mais ont été utilisées pour l'évaluation de l'IPCC. Elles sont récemment remplacées par ERA5 en vue d'améliorer par exemple la représentation du bilan hydrique et radiatif. Le modèle H-Tessel a remplacé le modèle TESSEL utilisé dans ERA-40 et ERA-Interim (Dee et al., 2011). D'autres sources offrent des données de réanalyse disponibles à l'échelle globale tels que MERRA-2 de la NASA, JRA-55 de la JMA et CFSR version 2 provenant du NCEP (Hersbach et al., 2020). Cependant, dans ce travail, nous adopterons les données ERA-Interim plus couramment utilisées pour la région d'étude.

Jones (2016), propose une analyse critique des bases de données de points de grille disponibles pour le cas de la température de l'air au-dessus des terres et de la température de l'océan (SST). Puis, il évalue les données de réanalyse en comparaison de ces données de grille. Il mentionne que dans le cas de Era-Interim la comparaison est excellente mais c'est selon lui un résultat attendu vu que ce sont justement les températures de surface qui sont assimilées. Comme le mentionne Maurer et al. (2010) les données de réanalyses sont plus adéquates comme entrées de modèles hydrologiques que les données des modèles climatiques régionaux (regional climate models RCM) pour l'estimation des débits des cours d'eau à partir des forçages atmosphériques. Toutefois, les données de réanalyse présentent des biais pour la variable pluie moyenne d'un jour pluvieux. Maurer et al. (2010) trouvent une surestimation des événements de très faible intensité et de la probabilité d'occurrence des valeurs extrêmes à droite de la distribution statistique (dans les valeurs fortes). Nous quantifierons les biais dans le cadre de notre région d'étude basés sur trois décennies d'observations Era-Interim (1979-2009).

Les précipitations représentent une variable clé dans le cycle hydrologique. Les précipitations dans les données de réanalyse sont le résultat du modèle de prévision basé sur les variables température et humidité de l'air qui ont été assimilées. Selon Dee et al. (2011), il y a eu une amélioration de la représentation de la variabilité spatiale et temporelle des précipitations en passant de ERA-40 à Era-Interim surtout à l'échelle des continents (hors océans). Ces comparaisons se font par rapport aux échelles régionales (et en utilisant des produits de précipitations estimées dans des mailles de calcul) provenant de l'interpolation des observations pluviométriques in-situ. Par exemple Simmons et al. (2010) ont comparé les estimations de Era-Interim aux évaluations de GPCC (Global Precipitation Climatology Center) (résolution $0.1^\circ \times 0.1^\circ$). Era-Interim surestime les précipitations dans l'hémisphère nord (période 1989-2009).

Pour le cas de l'Afrique du Nord, une sous-estimation de 1 mm/jour est observable sur les cartes de la Figure 25 dans Dee et al. (2011). Il y a aussi sous-estimation de la décroissance des moyennes estimée entre 1990-1999 et 2000-2009 (même figure) évaluée à 1 mm selon GPCC, par opposition à zéro (pas de croissance) selon ERA-Interim. D'autres évaluations de ERA-Interim ont été effectuées à l'échelle régionale. Citons les travaux de Balsamo et al. (2010) et Bromwich et al. (2011) pour la région Antarctique. Plusieurs recherches comparent les observations aux réanalyses (Croce et al., 2021) et mentionnent qu'il y a des biais dans les réanalyses et qu'il faut les corriger si on veut étudier l'impact local du changement climatique sur l'hydrologie (Cucchi, et al., 2021). Santer et al. (2004) comparent les anomalies de plusieurs variables dans les réanalyses et dans leur modèle de simulation. Poursuivant les efforts de la littérature pour quantifier les biais des données Era-Interim, notre thèse évaluera les données de précipitation de Era-Interim pour la région Nord Tunisie. C'est un travail original, qui n'a pas été effectué à notre connaissance jusqu'ici. Nous focaliserons seulement sur la saison d'hiver, tout en veillant à ce que les méthodes soient reproductibles pour d'autres saisons. Nous analyserons les biais dans les données de réanalyse des précipitations journalières et dans leurs caractéristiques structurelles. En effet, dans la thèse on compare les caractéristiques statistiques dans les observations de pluie au sol et dans les réanalyses de Era-Interim. Comme le mentionne Guillemot (2009) nous « construirons » le phénomène à valider. Il s'agit ici de la structure spatio-temporelle des événements pluvieux de la saison hivernale en nous basant sur un historique de 5 décennies (1959-2009) d'observations pluviométriques au sol. Nous privilégions l'analyse multidimensionnelle des séries pluviométriques journalières et de réanalyse Era-Interim pour dégager s'il y a des changements dans le temps.

L'analyse de la variabilité spatiale et temporelle de la pluviométrie en Tunisie a été menée par différents auteurs. Nous les citons dans ce qui suit, sans prétendre à l'exhaustivité, en commençant seulement par les années 1990 et en les mentionnant selon le pas de temps de l'étude, en allant de l'échelle annuelle à l'échelle la plus fine. Kebaili-Bargaoui (1989) et Bargaoui (1994) ont analysé des séries pluviométriques annuelles du bassin de la Medjerda en utilisant la méthode des surplus-déficits (runs) et l'analyse fréquentielle par la loi statistique de Weibull pour identifier les périodes de sécheresse. A cette époque-là, les questions de manque d'eau liées à la sécheresse étaient d'actualité. La sécheresse et son occurrence en Tunisie sur une période 88 ans est également investiguées en se basant sur les séries pluviométriques annuelles par Benzarti et Hbaieb (2001). L'étude des structures spatiales des pluies mensuelles a été analysée en se basant sur des indices topographiques des stations au nord de la Tunisie (Jedidi et al., 2008) et par une méthode fréquentielle univariée

sur tout le territoire tunisien (Benzarti, 2003 ; Merzougui et Slimani, 2012). La technique par d'ondelette continue est utilisée pour analyser les pluies mensuelles au sud de la Tunisie par Jemai et al. (2017). L'étude de Tramblay et al. (2019) a analysé la variabilité des pluies d'une résolution de 5 jours, en Tunisie, avec différentes méthodes d'interpolation. La variabilité des pluies journalières a été étudiée par chaînes de de Markov dans la région de grand Tunis (Bargaoui, 1983). Également, les événements pluvieux extrêmes à l'échelle journalière ont été analysés par Zahar (2015) et Zahar et al. (2000 et 2007). Parmi les études aux pas de temps les plus fins, nous pouvons mentionner une approche de copules, adopté par Gargouri-Ellouze et Bargaoui (2009) pour des séries pluviométriques d'une résolution de 5 minutes pour étudier la relation entre l'intensité des événements pluvieux et la réponse hydrologique du bassin versant. Ghanmi et al. (2013 et 2016) ont étudié les séries temporelles de pluie également à une forte résolution (5 minutes) et les courbes intensité-durée-fréquence pour plusieurs stations réparties au nord de la Tunisie en utilisant une approche multifractale. Ainsi les études précédentes n'ont pas analysé la structure temporelle des épisodes pluvieux, sur laquelle se focalisera cette thèse en vue d'examiner s'il y a un impact statistiquement décelable du changement climatique sur la variabilité de la structure spatio-temporelle des pluies journalières. La thèse visera aussi à analyser s'il y a des téléconnctions dans ces structures. En effet, plusieurs recherches ont associé la variabilité du régime pluviométrique à la circulation atmosphérique générale (Irannezhad et al., 2017 ; Marshall et al., 2020). Ces études s'intéressent à la réponse du système pluvieux d'une région particulière aux fluctuations climatiques à l'échelle planétaires. En effet, ces fluctuations sont caractérisées par des organisations préférentielles à basse fréquence dans la circulation atmosphérique connues par les modes de variabilité climatique (en anglais 'teleconnection pattern') et ils sont décrits par des séries chronologiques des indices. Dans la région méditerranéenne, les études comme celles de López-Moreno et al. (2011) et Tramblay et al. (2013) ont mis en évidence des relations significatives entre la variabilité pluviométrique et certains modes de variabilité climatiques comme le NAO (North Atlantic oscillations), l'ENSO (El niño southern oscillation) et WeMO (Western mediteranean oscillation), etc.

En conclusion, cette thèse "**Impact du changement climatique dans les modèles numériques à l'échelle régionale**" traitera des modèles de réanalyse des précipitations et d'un modèle en particulier : les données Era-Interim. L'impact dans ces modèles sera traité en cherchant s'il y a des anomalies persistantes identifiables d'après les séries de pluviométrie enregistrées au sol, sur un réseau de mesures de 70 pluviomètres, au cours des dernières décennies (1959-2009). Si des anomalies sont mises en évidence dans les séries de données sol, on cherchera si les modèles (les réanalyses) permettent de les retrouver. Comme point de départ, nous adoptons une

démarche statistique multidimensionnelle pouvant refléter la complexité de variabilité spatio-temporelle des pluies. La plus grande partie de l'étude sera consacrée à l'élaboration du modèle de structure spatio-temporelle des pluies, sans lequel l'étude des anomalies sur les 5 décennies observées n'aurait pu se faire.

Ce manuscrit comprend six chapitres en plus de l'introduction et de la conclusion. Le chapitre I est consacré à la revue de bibliographie. Le Chapitre II contient une présentation des deux jeux de données de précipitations journalières utilisées dans cette étude. Une base des données pluviométriques collectées par un réseau des pluviomètres et des précipitations obtenues par un modèle de réanalyse. Les anomalies des cumuls saisonniers y sont quantifiées en considérant la moyenne des 50 ans. La deuxième partie de ce chapitre présente les indices climatiques utilisés qui décrivent les modes de variabilité climatiques à une échelle régionale et globale. Le chapitre III présente la méthodologie de transformer les données pluviométriques de l'échelle journalière, à celle d'événements. Le chapitre présente le critère d'agrégation d'événements pluvieux choisi et les six descripteurs, d'événements et de saisons, retenus. Les statistiques élémentaires des descripteurs d'événements pluvieux pour les données du réseau de pluviomètres et pour celles du modèle de réanalyse sont analysées pour la période entière et en la divisant en deux sous-périodes. De plus les cartes d'anomalies entre les deux sous-périodes sont établies pour chaque sous-période. La fin de ce chapitre présente les stratégies multivariées envisagées pour l'analyse des descripteurs. Le Chapitre IV fait un bref rappel des méthodes d'analyse multivariée linéaires et non linéaires adoptées pour l'analyse de la structure des événements pluvieux : analyse en composante principale, analyse par cartes auto-organisatrices, par classification hiérarchique, et par la méthode 'data image'. Le chapitre V présente les étapes de synthétisation des descripteurs de pluie en des axes principaux par une méthode factorielle linéaire (ACP). Une investigation des liens entre les facteurs principaux des données pluviométriques et les modes de variabilité climatique dominants sont élaborés par une analyse de corrélation avec les indices climatiques puis par une analyse composite de la pression à surface de la mer (SLP) globale (données en grille). Dans le Chapitre VI, les résultats de la typologie des événements et des saisons pluvieux identifiés par la méthode neuronale SOM combinée avec une classification hiérarchique sont présentés et interprétés. Les classes obtenues sont utilisées pour l'évaluation des données issues du modèle de réanalyse en s'appuyant sur la même méthode pour étudier les tendances des séries temporelles trouvées. De plus les anomalies d'occurrence des classes identifiées sont étudiées par décennie. La conclusion fait la synthèse des anomalies répertoriées.

Chapitre I : Revue Bibliographique

I.1	INTRODUCTION	10
I.2	ECHELLE SPATIO-TEMPORELLE DES PROCESSUS HYDROLOGIQUES	10
I.3	L'ANALYSE PAR EVENEMENT PLUVIEUX	12
I.3.1	<i>Définition d'un événement pluvieux et des seuils pluviométriques.....</i>	<i>12</i>
I.3.2	<i>Approches de détermination de la durée minimale entre deux évènements pluvieux indépendants (MIET) 13</i>	
I.4	ANALYSES DESCRIPTIVES MULTIVARIEES EN HYDROLOGIE.....	14
I.5	ANALYSES DE TELECONNEXIONS DANS LA REGION MEDITERRANEENNE ET EN TUNISIE	17
I.6	CONCLUSION	19

I.1 Introduction

Ce chapitre est consacré à revue bibliographique qui porte sur le concept de l'évènement pluvieux et les différentes approches de définition de cet événement. En second lieu, elle porte sur l'analyse multivariée en hydrologie. La revue concerne aussi l'influence des oscillations de températures et des pressions dans les océans sur les conditions météorologiques dans la région méditerranéenne et en Tunisie.

I.2 Echelle spatio-temporelle des processus hydrologiques

Les précipitations n'ont pas une échelle caractéristique, elles doivent être étudiées dans une gamme d'échelles pertinentes pour chaque objectif de recherche. En effet, les processus physiques du système pluvieux s'étendent sur une gamme d'échelles spatiales et temporelles. D'après la classification des échelles de pluie par Orlanski (1975), les pluies peuvent s'étendre de l'échelle des cellules convectives qui se déroulent en quelque minutes entre 2 et 20 Km à l'échelle synoptique et planétaire associée au front ouragan où les précipitations durent de 1 jour à quelques jours et s'étendent de 20 à 2000 Km. Bloschl & Sivapalan (1995) ont proposé une synthèse de relation entre l'échelle temporelle et l'échelle spatiale en hydrologie tout en précisant les processus hydrologiques associés à chaque échelle spatio-temporelle. Par conséquent, les impacts des pluies sont très dépendants de l'échelle considérée. Par exemple, l'impact d'un orage isolé se limite à une échelle locale (ou un bassin versant de petite taille), mais il n'aura probablement pas d'effet à une échelle régionale de plus de 1000 Km (Bloschl & Sivapalan, 1995). Les différentes échelles des processus hydrologiques sont illustrées dans la figure I-1. Dans cette thèse, nous nous intéresserons aux précipitations liées aux systèmes frontaux qui correspondent à une échelle spatiale synoptique et une échelle temporelle qui dure de 1 jour à quelque jours (varie selon la durée de l'évènement pluvieux généré).

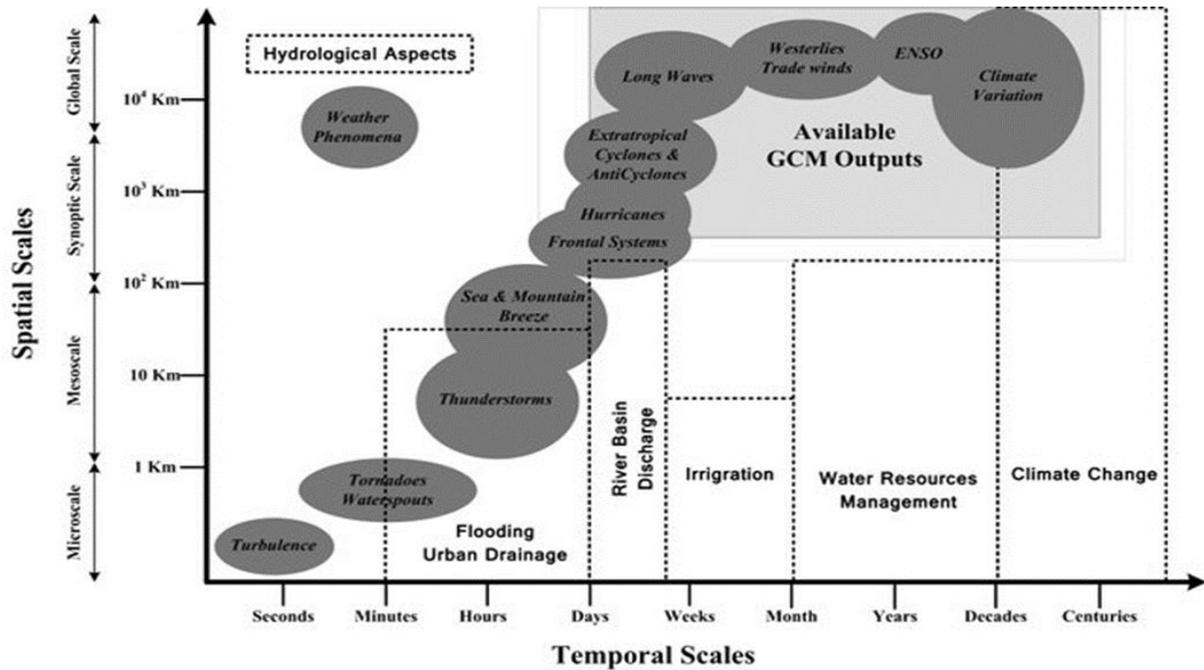


Figure I-1 Echelles spatiales et temporelles des processus hydrologiques correspondants (Tavakolifar et al, 2017)

La plupart des études régionale, basées sur un réseau d'observations, donnent des statistiques horaires, journalières, mensuelles et annuelles avec une estimation des courbes IDF (intensité-durée-fréquence) ou bien des courbes IDAF (intensité-durée-aire-fréquence). Le choix du pas du temps dans les études des précipitations est un point crucial. Les précipitations sont caractérisées par une forte variabilité et intermittence dans le temps et dans l'espace (Verrier et al., 2011 ; Akrouf et al., 2015). Quand un large pas de temps est considéré, la distinction entre les épisodes pluvieux et les épisodes secs devient impossible et les valeurs des cumuls de précipitations seront fortement lissées de façon qu'elles ne reflètent plus la variabilité naturelle des pluies. Dans le cas contraire, un taux d'échantillonnage très rapide conduit à des erreurs d'échantillonnage, spécialement quand un pluviomètre à auget basculant est utilisé.

Il est mentionné dans la littérature que le changement climatique aura un impact sur l'intensité des pluies : des pluies de plus en plus intenses et des épisodes secs de plus en plus longs sont projetés pour la zone méditerranéenne. Ainsi, pour l'analyse de l'impact du changement climatique sur les modèles estimant les précipitations doit être analysé à des échelles adéquates. A une échelle temporelle spécifique (mensuelle ou annuelle), les chercheurs se focalisent surtout sur les aspects quantitatifs de la pluie. Or, un cumul pluvieux à un pas de temps fixe (année, saison, mois, etc) peut correspondre à une multitude d'événements pluvieux de faible intensité ou à un seul événement très intense. Par conséquent, les différents types d'événements ne présentent pas, certainement, le même impact sur le sol, l'agriculture, les zones urbaines,

etc. L'identification des événements pluvieux suite à la séparation des épisodes secs et humides à partir des pluies journalières et la caractérisation de ces événements permettent d'appréhender la question de l'intensification des pluies dont il est question dans les projections climatiques dans le cadre du changement climatique.

I.3 L'analyse par événement pluvieux

I.3.1 Définition d'un événement pluvieux et des seuils pluviométriques

Le principe de l'approche de l'événement pluvieux est de découper une série chronologique de pluies en des événements pluvieux séparés les uns des autres. Or, il n'existe pas de consensus préétabli pour réaliser ce découpage. La définition d'un événement pluvieux peut dépendre de plusieurs facteurs et particulièrement la résolution spatio-temporelle choisie pour l'étude, ainsi que les bruits instrumentaux (dans les séries à forte résolution temporelle) et le caractère intermittent de pluie complique la délimitation des événements pluvieux. Par exemple, dans certaines études, les conditions hydrologiques, comme la trace au sol, l'humidité du sol et le débit de l'événement, sont prises en considération pour délimiter un événement pluvieux (Cattan et al., 2006 ; Joo et al., 2014). D'autres études adoptent l'approche météorologique qui considère qu'un événement pluvieux est associé à une perturbation atmosphérique localisée dans le temps et dans l'espace. En général, les chercheurs tendent à choisir une durée minimale de non-pluie, connue dans la littérature par 'minimum inter-event time : MIET' qui permet de considérer l'événement antérieur indépendant de l'événement suivant (Restrepo-Posada and Eagleson, 1982 ; Dunkerley, 2008 ; Molina-Sanchis et al., 2016). Une telle approche est quelque part subjective, car la notion d'indépendance des événements peut changer d'une étude à une autre selon le contexte et l'objectif de l'étude. En effet, dans plusieurs études, le choix de la MIET ne s'appuie pas sur les propriétés de la pluie même, mais sur d'autres considérations. Par exemple, Bracken et al. (2008) adoptent une MIET de 12 h pour étudier les débits en Espagne juste pour réduire l'effet de l'humidité antécédente. Lloyd (1990) a estimé qu'une MIET de 3 h est suffisante pour le séchage du couvert végétal dans une étude d'interception dans la forêt amazonienne centrale. Certaines autres études combinent le choix de la MIET avec d'autres critères, comme fixer une hauteur minimale de pluie (Fornis et al., 2005). Dans les études mentionnées, l'extraction des événements pluvieux s'appuie souvent sur des séries pluviométriques de faible résolution temporelle (échelle horaire, échelle sous horaire) et la délimitation des événements liés aux systèmes frontaux (à l'échelle des jours) est nouveau. En s'appuyant sur une base de données journalière, la définition d'un jour sec ou humide et l'adoption des seuils de pluies est une question primordiale. En effet, selon l'objectif de la recherche, plusieurs chercheurs tendent à fixer un seuil qui varie de 1 à 10 mm/jour pour définir

un jour humide (Rivoire et al., 2019). Par exemple, Douguedroit (1987) utilise le seuil de 1 mm /jour dans la région méditerranéenne, parce qu'il considère qu'au-dessous de cette quantité la pluie ne sera pas absorbé par le sol vu la forte évapotranspiration dans la région. Ce seuil de 1 mm/jour était largement utilisé (Jiang et al., 2011 ; Polade et al., 2014 ; Raymond et al., 2018) même si cette valeur est arbitraire et n'a pas été prouvée par aucune étude expérimentale (Rivoire et al., 2019). Contrairement aux études mentionnées, aucun seuil n'est fixé à priori dans cette recherche, c-à-dire un jour qui enregistre un cumul pluvieux supérieur à 0 mm/jour est considéré un jour pluvieux. Un tel seuil (0 mm) permettra de prendre en considération les événements pluvieux les plus faibles des pluies légères (bruines). D'autre part, la fixation d'un seuil dans une analyse qui se focalise à la structure interne de l'événement pluvieux risque de modifier considérablement les résultats. En effet, cela peut agir sur le nombre et la structure d'événement et empêche la détection de la variabilité intra-événement en le découpant un seul événement en des événements séparés.

I.3.2 Approches de détermination de la durée minimale entre deux événements pluvieux indépendants (MIET)

Plusieurs méthodes statistiques sont proposées pour déterminer une MIET adéquates (Dunkerley, 2008 ; Joo et al. 2014). Les approches les plus utilisées sont :

L'analyse d'autocorrélation : Cette approche est appliquée sur des séries pluviométriques dans plusieurs études (par exemple : Asquith et al., 2005 ; Wenzel & Voorhees 1981 ; Gaál et al., 2014). Il s'agit de choisir le pas du temps nécessaire pour que le coefficient d'autocorrélation, de la série chronologique de pluies, diminue par rapport un certain seuil et converge vers zéro.

Le coefficient de variation : Cette technique consiste à calculer le coefficient de variation de la IET (la durée des jours secs entre deux événements pluvieux). La MIET optimale est obtenue lorsque le coefficient de variation de IET est égal à 1 (Restrepo-Posada & Eagleson, 1982).

Le choix d'une MIET adéquate est une étape très délicate parce qu'elle influence ensuite les caractéristiques de l'évènement comme la durée et l'intensité ainsi que le nombre d'évènements. Par exemple avec une petite MIET, on risque de diviser un même événement et le considérer comme deux événements différents. Par contre, si on choisit une grande MIET, on risque de fusionner deux événements indépendants dans un même événement et de cette manière, le nombre de jours secs au niveau d'un événement augmente ce qui modifie la variabilité intra-événement.

Pour des événements pluvieux qui présentent des propriétés très différentes dans le temps et dans l'espace, des outils d'analyse multivariée peuvent être appliqués pour détecter les structures dominantes des précipitations ainsi que leurs modes de variabilité.

I.4 Analyses descriptives multivariées en hydrologie

Les approches descriptives multivariées sont l'ensemble des méthodes de la statistique exploratoire qui permettent de traiter simultanément un nombre quelconque de variables (il s'agit d'aller au-delà d'analyse d'une seule ou deux variables).

L'hydrologie, la météorologie, et la climatologie sont, en général, des sciences basées sur l'exploration et l'analyse des données massives. Les techniques les plus classiques de l'analyse descriptive multivariée sont les méthodes factorielles. Elles consistent à trouver les facteurs en nombre réduits en synthétisant le mieux possible les données en jeu. La méthode la plus connue est l'analyse en composantes principales ACP, adoptée dans cette étude. Elle s'agit d'une méthode linéaire qui dépend d'un modèle géométrique et ne s'appuie pas sur un modèle probabiliste ou des hypothèses de nature statistique (Lebart et al., 2000). Cette technique est créée par Person en 1901 et développée par Hotelling (1933) (Jolliffe, 2002). C'est un outil puissant dans la synthèse et la réduction de l'information quand il s'agit d'analyser de grandes bases de données (Escofier et Pagès, 2008). L'ACP est une technique largement utilisée dans le domaine de la météorologie et l'hydrologie pour identifier les relations de cause à effet, pour détecter les variables redondantes, pour mieux identifier les zones aux caractéristiques climatiques ou hydrologiques homogènes, etc. Citons par exemple l'étude de Comrie et Glenn (1988) qui ont utilisé l'ACP sur des pluies mensuelles de 309 stations pluviométriques réparties sur la frontière entre les États-Unis et le Mexique entre 1960 et 1990 pour une étude de régionalisation. Jayawardene et al. (2005) ont appliqué l'ACP sur les pluies journalières dans le but d'identifier les zones humides et les zones sèches à Sri Lanka. Darand et Mansouri (2014) ont utilisé l'ACP combiné avec la classification hiérarchique ascendante CHA pour classifier des données pluviométriques journalières en grille couvrant le territoire iranien obtenues de la base de données APHRODITE. Martínez et al. (2007) ont analysé des indices pluviométriques calculés à partir des données journalières par l'ACP pour une analyse profonde de régime de pluie à Catalogne (Espagne). En Tunisie, Jebari et al. (2007) ont utilisé l'ACP pour une classification spatiale des précipitations de la Dorsale tunisienne à des échelles temporelles fines (de 30 min à 24 h). Ellouze et al. (2009) ont adopté cette analyse aussi pour une classification spatiale des pluies annuelles des stations réparties au sud tunisien. Les études sont très nombreuses dans les domaines de la météorologie et climatologie (Wigley et al., 1982 ;

White and Richman, 1991 ; Yu et al., 1997 ; Giannini et al., 2003, Hannachi, 2004 ; Astel et al., 2004 et Jhajharia et al., 2013).

Il existe également d'autres approches multivariées, non linéaires, comme l'analyse par classification. Le principe de clustering consiste à organiser les données en des groupes homogènes selon un critère de similarité choisi. On peut distinguer deux types de classification principaux : la classification supervisée et la classification non supervisée. Le processus de la classification supervisée comprend deux phases. La phase d'apprentissage consiste à créer un classifieur ou un modèle à partir des classes connues. Ensuite, dans la phase de classification, on peut prédire les classes des nouvelles données avec le classifieur créé au préalable (Mephu and Njiwoua, 2005 ; Caruana and Niculescu-Mizil, 2006). Dans la littérature, on trouve différentes méthodes de classification supervisée : les plus proches voisins (Cunningham and Delany, 2007), les arbres de décisions (Xu et al., 2019), la régression logistique (Lee, 2018), les réseaux de neurones et d'autres. En revanche, la classification non supervisée, connue par le terme anglais « clustering », fait un partitionnement des données sans aucune connaissance au préalable en se basant sur un critère de ressemblance entre les données à classifier (Greene et al, 2008).

Il existe diverses techniques de clustering non paramétriques utilisées dans les domaines de l'hydrologie et la climatologie. Elles sont divisées en trois catégories principales :

- La classification par partitionnement : dans cette catégorie, l'ensemble des données est divisé en des sous-ensembles et chaque sous-ensemble représente une classe. Cette classification donne une seule partition des données en minimisant itérativement une fonction coût. Les classes obtenues sont représentées par un prototype qui est le centre de gravité du sous-ensemble dans la plupart des méthodes par partitionnement. Parmi les caractéristiques de cette classification, c'est que le nombre de classes est toujours fixé à l'avance. La classification la plus connue et la plus utilisée est l'algorithme des centres mobiles (K- moyennes). Dans cette méthode, les prototypes sont initialisés d'une manière aléatoire, ensuite, à chaque itération, ces prototypes font des mises à jour en minimisant une fonction coût jusqu'à atteindre une partition optimale des données (Morissette & Chartier, 2013). La méthode des cartes auto-organisatrices (Self Organizing Map SOM) partage la même démarche que celle des k-moyennes avec l'option qu'elle préserve la topologie des données, c'est-à-dire les éléments proches dans l'espace des données seront dans la même classe ou dans des classes voisine à l'espace des sorties (Kohonen, 2001). Contrairement à l'algorithme de k- moyennes qui aboutit à une partition fixe, il existe des algorithmes qui

se basent sur le même principe sauf qu'il donne une partition floue où les individus peuvent appartenir à plusieurs classes à la fois comme le cas de la méthode c-moyennes floues connu par « fuzzy c-means » (Malhotra et al., 2014). Il existe aussi la méthode de nuées dynamiques qui représente une forme généralisée du K-moyennes où le prototype des classes se trouve sous différentes formes, autre que le centre de gravité comme une loi de probabilité ou un histogramme, etc.

- La classification hiérarchique : Cette classification appartient à la famille des classifications non supervisées. Elle identifie les relations hiérarchiques de l'ensemble des individus d'une manière ascendante ou descendante. Si la classification hiérarchique est ascendante (CHA), on admet au début que chaque individu représente une classe séparée, ensuite à chaque itération les classes similaires se fusionnent suivant un critère de similarité choisi pour former une nouvelle classe. À la fin, on obtient une seule classe contenant tous les individus. En revanche, dans la classification hiérarchique descendante, on admet au début que tous les individus forment une seule classe et à chaque itération le nombre de classes augmente, suivant un critère de dissimilarité, pour avoir à la fin des classes formées d'un seul individu. La version ascendante dite aussi agglomérative est la plus utilisée et les relations hiérarchiques sont visualisées par une structure arborescente appelée dendrogramme. Contrairement à la classification par partitionnement, le nombre de classes n'est pas fixé à priori. Il est plutôt choisi selon la structure du dendrogramme obtenue et le niveau de coupure des branches du dendrogramme qui donne des différentes partitions en changeant le niveau (Murtagh and Contreras, 2011).
- Classification à base de densité et de grille : Ces deux types de classification se ressemblent dans leur principe. Dans la classification basée sur la densité, les zones de forte densité dans l'ensemble des données sont considérées des classes et les zones de faibles densités sont considérées des bruits. La méthode DBSCAN est la plus connue dans cette famille de classification (Ester et al., 1996). Quant à la classification à base de grille, l'espace des observations est divisé par une grille contenant des unités de même dimension. Les unités de la grille caractérisées par une forte densité des données (dépassant un certain seuil) forment les classes.

Dans cette étude, une méthode neuronale de la carte auto-organisatrice, connue par self-organizing map (SOM) combinée avec une classification hiérarchique, est proposée pour l'analyse des précipitations.

La carte auto-organisatrice est une technique statistique développée par le mathématicien Teuvo Kohonen en 1984. Cette méthode est utilisée dans plusieurs contextes comme l'analyse

exploratoire des données, la classification, la réduction des bases de données de grande dimension ou reconnaissance de formes (Alhoniemi et al, 1999). Les cartes auto-organisatrices font partie de la famille des réseaux de neurones artificiels (RNA) avec un modèle d'apprentissage non supervisé (Kohonen, 1988 ; Vesanto & Alhoniemi, 2000). La méthode SOM est de plus en plus utilisée dans le domaine climatique (Malmgren & Winter, 1999 ; Hewitson, & Crane, 2002 ; Liu & Weisberg, 2011 ; Nasteski, 2017). Dans le domaine de la météorologie et l'analyse des précipitations, les études sont nombreuses. Schuenemann et al. (2009) ont étudié les tendances des précipitations journalières dans le Greenland entre les années 1961 et 1999 par une analyse SOM. Puis, ils ont lié les tendances identifiées à celles des séries de pression à la surface de la mer SLP dans l'Océan Atlantique. Cette méthode de classification a été adoptée par Natita et al. (2016) pour identifier les tendances dans les séries d'humidité spécifique au sud de Thaïlande. Parchure et Gedam (2019) ont utilisé l'algorithme SOM pour la classification de plus de 23 000 événements précipitant observés par 47 stations météorologiques de la région de Mumbai en Inde. La SOM combinée à une méthode de classification hiérarchique ascendante (CHA) a permis l'étude du régime météorologique au Sénégal pour la saison de la mousson d'été à l'échelle synoptique (Gueye et al., 2010). Dilmi et al. (2017) ont utilisé cette même approche pour la classification des événements précipitants observés à très fine résolution dans un site à Palaiseau en France. Les liens non linéaires entre la structure macrophysique des événements et les processus microphysiques sous-jacents ont ainsi pu être mis en évidence.

I.5 Analyses de téléconnexions dans la région Méditerranéenne et en Tunisie

Le nord de la Tunisie appartient au sud du bassin méditerranéen et cette zone est considérée comme une zone à risque « Hot spot » dans le contexte de l'évolution du climat, d'après plusieurs chercheurs (Giorgi, 2006 ; Giorgi and Lionello, 2008) et d'après le sixième rapport du GIEC (2022) (Groupe d'experts intergouvernemental sur l'évolution du climat) qui classifie cette région parmi les zones les plus vulnérables et les plus exposées à la variabilité climatique vue sa richesse socio-culturelle et sa biodiversité (IPCC, 2022). Les perturbations climatiques influençant le climat méditerranéen sont liées d'une part aux oscillations à grande échelle (échelle globale) et d'autre part à des oscillations régionales et locales (Ulbrich et al., 2012). À ce propos, plusieurs études ont investigué l'impact des différentes téléconnexions de l'hémisphère nord sur les conditions climatiques méditerranéennes. Par exemple, López-Moreno et al., 2011, ont étudié les influences de NAO sur la température et la précipitation des régions montagneuses de la Méditerranée. Skliris et al. (2012) ont analysé le changement de la

température à la surface de la mer de la Méditerranée par rapport aux oscillations de l'Atlantique à une échelle interannuelle le NAO, et une échelle multidécennale comme de L'oscillation multidécennale atlantique (AMO : Atlantic Multidecadal Oscillation) et le mode de variabilité Atlantic est (EA : East Atlantic). Xoplaki et al. (2004) ont trouvé que le mode de variabilité NAO est le premier responsable à la variabilité des précipitations méditerranéennes hivernales à une échelle décennale tout le long du 20e siècle. Trambly et al. (2013) ont analysé les liens entre les précipitations des pays du Maghreb (Maroc, Algérie et Tunisie) et les indices MO, NAO, WeMO et l'ENSO et ils ont trouvé une certaine corrélation entre la quantité des pluies et le nombre de jours secs et humides et les indices NAO et MO. Quadrelli et al., 2001 ont investigué les liens entre les anomalies de précipitations hivernales mensuelles dans la région méditerranéenne et les anomalies dans la circulation atmosphérique à grande échelle et ont détecté une corrélation remarquable avec le NAO et une corrélation plus faible avec l'EA (Est Atlantic Pattern). Criado-Aldeanueva and Soto-Navarro (2013), se sont focalisés sur l'impact des oscillations régionales, en se basant sur l'indice MO (Mediterranean Oscillation), sur les précipitations et l'évapotranspiration de la région méditerranéenne. La plupart des études citées ont prouvé que le phénomène NAO représente le mode de variabilité le plus corrélé avec le climat méditerranéen et nord-africain. Le phénomène ENSO a des impacts économiques, écologiques, sociales et climatiques sur tout le globe qui touchent aussi le bassin méditerranéen malgré l'éloignement géographique (Power et al. 1998 ; Chiang and Sobel, 2002 ; Gutierrez, 2017). Deux études de Ropelewski and Halpert en 1987 et 1989 ont analysé les relations entre le phénomène El Nino et la variabilité des précipitations sur tout le globe. Ils ont trouvé que les relations des précipitations des régions nord-africaine et sud européenne sont instables dans le temps. Une étude de Wang et al. (2014) a montré que la région méditerranéenne devient plus humide pendant la phase El Nino se l'ENSO et El Nino combiné avec une phase chaude du PDO peut provoquer une sévère sécheresse en Afrique.

En Tunisie, Ouachani et al. (2013) ont étudié la téléconnexions entre des précipitations mensuelles dans le bassin de Medjerda au nord de la Tunisie et plusieurs indices climatiques (AO, NAO, WeMOI, SOI...) par la technique d'ondelette et ils ont trouvé que cette région est sous l'influence d'une combinaison des oscillations climatiques ENSO, NAO, WeMO, etc. Le ENSO a donné une forte corrélation négative avec les pluies mensuelles d'une région EL Feija à l'extrême nord-ouest tunisien. Une étude plus ancienne élaborée par Douguedroit (1988) dans laquelle les relations entre les précipitations mensuelles d'hiver en Tunisie et des indices climatiques ont été investiguées où il a trouvé une forte corrélation avec l'indice NAO. Une

forte corrélation positive entre pluies mensuelles de la Tunisie et l'indice de NAO est aussi démontré dans l'étude de Horst-Joachim Lüdecke et al. (2021) pendant les mois d'octobre et février. Ainsi, les indices régionaux WeMO et MO sont les modes de variabilité dominants dans la région méditerranéenne. En effet, le MO a montré une forte corrélation positive avec les pluies mensuelles hivernales au niveau du bassin trans frontalier de Medjerda entre la Tunisie et l'Algérie (Boulmaiz et al., 2022). Également, l'indice MO a montré une bonne cohérence entre les pluies mensuelles de la région sud de la Tunisie par la méthode des ondelettes continues (Jemai et al., 2017). Dans cette étude, la réponse des précipitations à la variabilité climatique à grande échelle sera analysée par des corrélations linéaires, en s'appuyant sur une base de données des indices climatiques. En outre, nous avons utilisé une analyse composite. C'est une technique complémentaire à l'analyse des corrélations linéaires qui est largement utilisée en climatologie pour investiguer les téléconnexions entre des variables climatiques et des paramètres météorologiques (Poccard-Leclercq, 2000 ; Ning et Bradley, 2014 ; Fu et al, 2020 ; Lai et al., 2021).

I.6 Conclusion

Ce chapitre a donné, en premier lieu, un aperçu sur l'échelle de l'événement pluvieux proposée ainsi que les différentes approches de définition. En deuxième lieu, les différentes méthodes multivariées adoptées dans les domaines de l'hydrologie, la climatologie et la météorologie sont présentées. Ce chapitre de revue bibliographique a fini par la présentation de l'influence de la variabilité climatique globale sur les régimes pluviométriques à la région méditerranéenne.

Chapitre II : Base de données hydro-climatiques

II.1	INTRODUCTION	21
II.2	CARACTERISTIQUES CLIMATIQUES DE LA ZONE D'ETUDE	21
II.3	FLUX PLUVIEUX ET INFLUENCE DES RELIEFS	22
II.3.1	<i>Flux pluvieux</i>	22
II.3.2	<i>Influence des reliefs sur la propagation des flux pluvieux</i>	24
II.4	BASE DE DONNEES PLUVIOMETRIQUES (IN SITU)	24
II.4.1	<i>Stations retenues pour l'étude</i>	24
II.4.2	<i>Choix de la saison d'étude</i>	28
II.5	SIMULATIONS NUMERIQUES ET MODELE CLIMATIQUES	31
II.5.1	<i>Principe de simulation numérique et modèles climatiques globaux GCM et régionaux RCM</i>	31
II.5.2	<i>Modèle de réanalyse météorologique</i>	33
II.5.3	<i>Données fournies par les ré-analyses météorologiques : ERA-Interim</i>	34
II.6	DONNEES DE LA VARIABILITE CLIMATIQUE	35
II.6.1	<i>Choix des indices climatique</i>	35
II.6.2	<i>Données climatiques de large échelle : Pression à la surface de la mer (SLP)</i>	37
II.7	CONCLUSION	37

II.1 Introduction

L'étude présentée dans le manuscrit s'intéresse principalement à l'analyse de la variabilité spatiale et temporelle des précipitations dans la région nord de la Tunisie. Ce chapitre est consacré à la description de la zone d'étude et les données utilisées dans cette recherche. Deux bases de données pluviométriques sont présentées : les pluies journalières collectées d'un réseau de pluviomètres et les pluies journalières produite par le modèle de réanalyse ERA-Interim. Dans un second temps, les téléconnexions et les données climatiques utilisées dans l'investigation des liens entre l'information pluviométrique et la circulation atmosphérique générale sont présentées.

II.2 Caractéristiques climatiques de la zone d'étude

La zone d'étude présentée dans la figure II-1 est située au nord de la Tunisie entre 35°N - 38° N et couvre une superficie de 28 641 km². Cette zone appartient à la rive sud de la mer Méditerranée, délimitée par la mer du côté nord et est, l'Algérie du côté ouest et du Sahara Tunisienne du côté Sud. Dans les annuaires pluviométriques de la Tunisie et pour les études hydrologiques, le nord tunisien groupe les quatre gouvernorats du nord-ouest (Béjà, Jendouba, Kel et Siliana) et les sept gouvernorats du nord-est (Tunis, Manubah, Ariana, Ben Arous, Bizerte, Zaghouen et Nabeul). La zone d'étude est délimitée, dans la figure II-1, par un trait gras.

La Tunisie se situe dans une zone de transition entre le climat tempéré des moyennes latitudes et le climat subtropical aride (Estienne et Godard, 1992 ; Slimani et al., 2007 ; Rysman et al., 2013). La partie nord du pays présente majoritairement un climat méditerranéen caractérisé par un hiver doux et humide et un été sec et chaud. En adoptant le critère d'Emberger de classification des étages bioclimatiques basé sur les pluies annuelles, la température moyenne journalière du mois le plus chaud et la température moyenne journalière du mois le plus froid (Emberger, 1930, Marres, 1972), la région d'étude est considérée comme une nuance entre le climat humide, sub-humide et semi-aride délimitées dans la figure II-1(Smaoui, 2015). L'extrême nord-ouest présente une zone montagneuse avec un climat humide, caractérisé un cumul annuel moyen entre 800 et 1200 mm (Hénia, 1986). Une grande partie de la zone d'étude appartient au climat semi-aride caractérisé par un cumul pluviométrique annuel entre 400 et 600 mm. Il existe aussi une zone intermédiaire appartenant au climat subhumide (Selmi et Nasri, 1997). La zone d'étude est considérée la région la mieux arrosée du pays, de ce fait, les ressources en eau et le potentiel agricole de cette région sont cruciaux pour l'économie tunisienne. Ainsi, l'information pluviométrique fiable, à différentes échelles spatiales et

temporelles, ne peut être que d'un grand intérêt dans des domaines aussi dépendants de l'agriculture, du climat et des ressources en eau en générale.

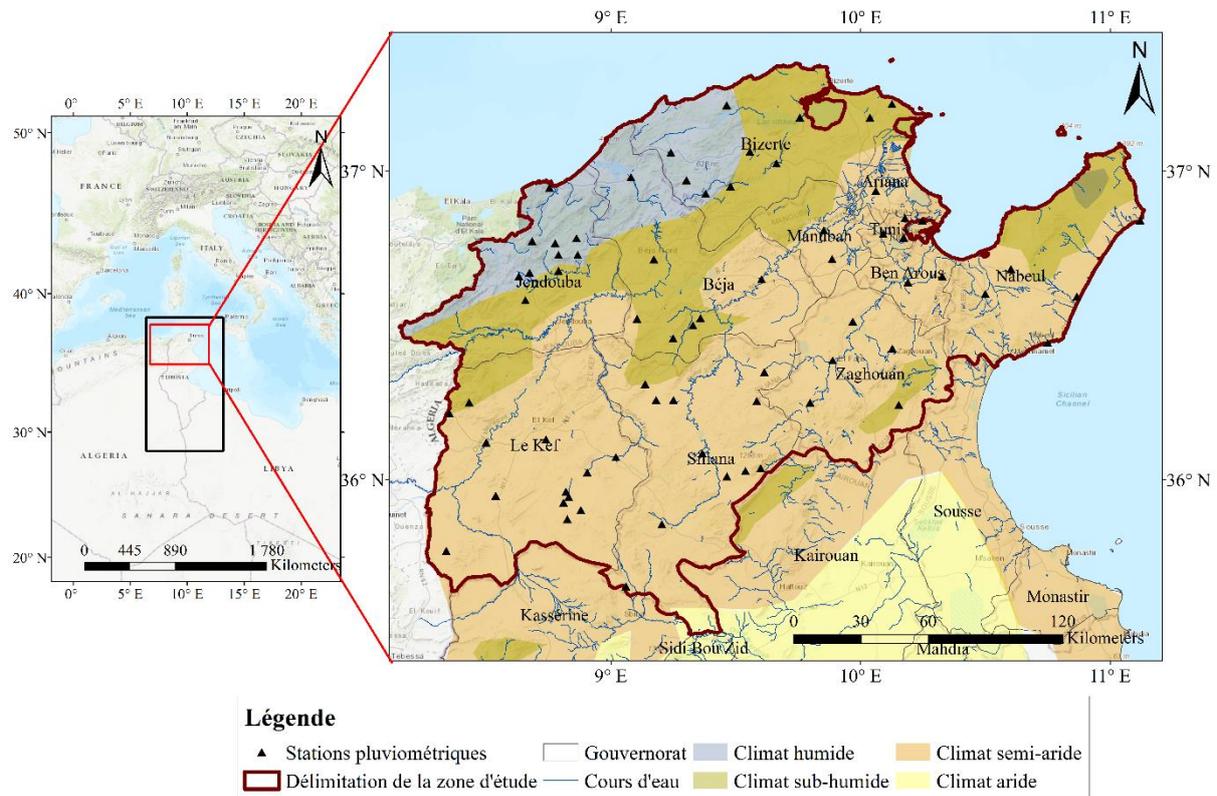


Figure II-1 Localisation de la zone d'étude et les étages bioclimatiques

II.3 Flux pluvieux et influence des reliefs

II.3.1 Flux pluvieux

Les flux pluvieux qui frappent le nord de la Tunisie viennent de diverses directions :

- Les flux ouest à nord-ouest : ces flux sont considérés les plus fréquents. Ils sont très abondants pendant la saison d'hiver et occupent environ les 2/3 des jours de l'année d'après l'étude de Ben Boubaker (2003) sur les flux pluvieux en Tunisie entre les années 1994 et 1997. Les apports pluviométriques liés à ces flux sont dominants. Les flux ouest et nord-ouest arrivent directement de l'Atlantique, ils sont attirés au Nord du pays par une dépression à la méditerranée occidentale appelée la dépression du golfe des Baléares (Merzougui et Slimani, 2012).
- Les flux nord à nord-est : ce type de flux a des origines européennes et continentales. Ils sont plus fréquents pendant la saison d'été. Ils apportent plus de précipitations au printemps et en automne (Ben Boubaker, 2003).
- Les flux d'origine méridionale : ces flux sont les moins abondants. Ils ont deux trajectoires soit est à sud-est ou bien sud à sud-ouest. Il est plus probable que les flux est à sud-est soient accompagnés par des pluies vu leur trajectoire maritime.

D'après des études précédentes (par exemple Slimani et al, 2007 ; Merzougui et Slimani, 2012 ; Xoplaki et al, 2004), les principaux centres d'action de la circulation atmosphérique qui influencent le climat et la pluie en Tunisie et génèrent des perturbations climatiques sont :

- La dépression¹ de l'Islande et les dépressions de la méditerranée occidentale (la dépression du golfe des Baléares en Espagne et la dépression du golfe de Gênes en Italie) du côté occidental.
- La dépression du golfe du Gabès et la dépression du golfe des Syrtes du côté de la méditerranée orientale.
- L'anticyclone² des Açores et du nord Atlantique de l'ouest.
- L'anticyclone de l'Europe centrale et orientale de l'est.

Le mouvement des masses d'air de diverses origines polaires, tropicales et arctiques sont aussi des facteurs régionaux qui influencent le climat Tunisien (Berndtsson, 1989 ; Jebari et al., 2007). Les caractéristiques orographiques et la proximité de la mer jouent un rôle important dans la propagation des flux pluvieux et influencent fortement le climat dans la zone d'étude. La figure II-2 schématise la localisation et les types des points d'actions qui influencent le climat Tunisien et la trajectoire des différents flux qui frappent le nord de la Tunisie.

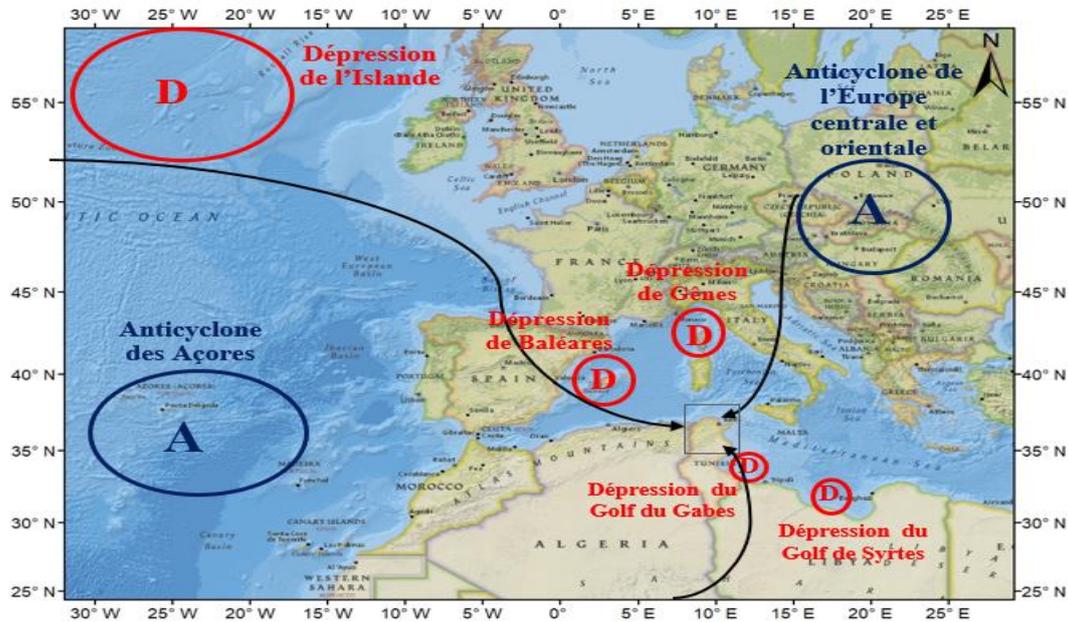


Figure II-2 Schéma de synthèse pour les différents points d'actions influençant le climat tunisien. (A: Anticyclone, D: Dépression, Flèches: Direction des Flux)

¹ Une dépression (cyclone) est une zone de basse pression. La dépression résulte de l'ascension des masses d'air vers l'atmosphère et cela favorise la formation des nuages et des précipitations. Elle est toujours associée à un mauvais temps.

² L'anticyclone est une zone de haute pression atmosphérique. Il est caractérisé par le mouvement descendant de masse d'air froide qui empêche la formation des nuages et augmente la pression au niveau de sol. L'anticyclone est souvent associé à un temps sec avec un ciel clair et ensoleillé.

II.3.2 Influence des reliefs sur la propagation des flux pluvieux

Le nord de la Tunisie est situé dans l'extension des massifs montagneux de l'Atlas d'orientation sud-ouest/nord-est. L'Atlas prend source au Maroc et se subdivise en deux chaînons majeurs séparés par des plaines (l'Atlas Tellien et la Dorsale). Comme le montre la figure II-3, le premier s'étend sur les côtes les plus septentrionales de la Tunisie et forme un véritable écran face aux flux provenant du nord qui sont les plus pluvieux. Par contre les flux d'est et nord-est entre sur le territoire parallèlement à l'axe de ces reliefs. Une grande partie de la zone d'étude située à la vallée de Medjerda est considérée à l'abri pluvial surtout face aux flux nord-ouest. Les reliefs qui bordent cette zone, les monts de Kroumerie et Mogods au nord et la haute Tell au sud, affaiblissent le potentiel pluviométrique des flux (Fig. II-3) (Ben Boubaker, 2003).

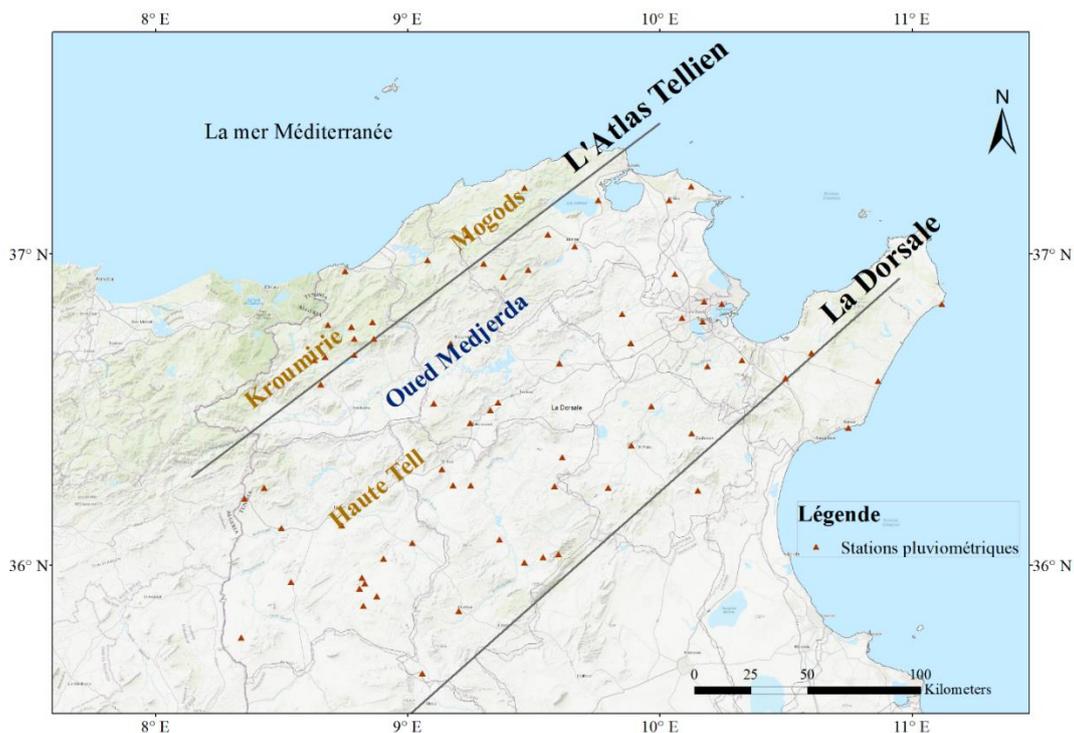


Figure II-3 L'orientation des principaux reliefs de la zone d'étude

II.4 Base de données pluviométriques (in Situ)

II.4.1 Stations retenues pour l'étude

Le nord de la Tunisie est caractérisé par un réseau de suivi pluviométrique assez dense par rapport aux autres régions du Pays. Le tableau II-1 représente les caractéristiques des stations pluviométriques retenues après le traitement préliminaire des données comme la localisation géométrique, l'altitude et le début d'observation. Certaines stations sont en service depuis un siècle comme celles d'Ain Drahim à Jendouba et Ksour-école au gouvernorat du Kef. Cette étude est basée sur une base de données journalière collectés auprès de la Direction Générale des Ressources en Eau (DGRE) du Ministère de l'agriculture et des ressources hydrauliques.

BASE DE DONNEES HYDRO-CLIMATIQUES

Dans certaines stations, les séries journalières ne sont pas suffisamment longues et continues. Pour assurer une meilleure qualité de résultats, les années contenant un nombre important des lacunes sont éliminées de la base de données utilisée. Par conséquent, plusieurs stations sont aussi éliminées. Dans la plupart des stations, les données sont disponibles entre Septembre 1959 et Aout 2009. Cette période de 50 ans est choisie pour toute l'étude. Le tableau II-1 présente également les pluies annuelles moyennes sur la période 1959 et 2009 ainsi que l'apport (en %) des 3 saisons (septembre-octobre-novembre (SON), décembre-janvier-février (DJF), mars-avril-mai (MAM)) par rapport au cumul pluviométrique total.

Concernant la qualité des données des stations pluviométriques, la direction générale des ressources en eau (DGRE) fait plusieurs vérifications et corrections avant leur publication. A la réception du bulletin pluviométrique, une étude de comparaison aux relevés des postes voisins se réalise dans le but d'éliminer toute observation qui semble aberrante. Une deuxième vérification se fait à la préparation de l'annuaire par une comparaison des bulletins avec leur double et avec les souches afin de noter les rectifications éventuelles. Des vérifications très minutieuses sont également effectuées sur les dates et l'heure de collecte des données pour éviter un décalage au niveau des observations (Annuaire pluviométrique, 2017).

Tableau II-1Caractéristiques des stations pluviométriques utilisées

Id	Stations	ID unique	Longitude (°)	Latitude (°)	Altitude (m)	Début observation	Pourcentage des lacunes entre 1959-2009	Pluie annuelle moyenne entre 1959-2009 (mm)	Apport des pluies SON (%) 1959-2009	Apport des pluies DJF (%) 1959-2009	Apport des pluies MAM (%) 1959-2009	Nombre de saisons considérées
S01	Tabarka Foret	1483726801	8,75	36,95	10	1959	0	930	32,5	43	21,7	50
S02	Ain Drahim	1483018801	8,68	36,77	715	1889	0	1544	27	44,5	24,9	50
S03	Ain Debba	1485013801	8,68	36,7	470	1968	21.48	950	25,7	46,2	25,4	39
S04	Ain Baya	1485007801	8,64	36,66	330	1951	18.4	712	25,3	45,3	25,8	40
S05	Majen Safi	1483428101	8,78	36,77	530	1942	19.2	841	27,5	45,4	24,3	39
S06	Ain Zana	1485053501	8,87	36,73	875	1950	9.86	820	26,5	45	29,6	37
S07	Ain Sallem	1485042201	8,79	36,73	836	1942	36.72	522	29,2	45,9	22,2	27
S08	Ain Merja	1485035001	8,66	36,58	350	1949	28.9	388	27,1	42	26	30
S09	Ain Hamraya	1485026001	8,86	36,78	553	1912	28.8	642	26,9	46,5	23	28
S10	Bouhertma	1485140301	8,79	36,68	250	1959	10	616	26,9	45,1	24,1	41
S11	Nefza delegation	1483464502	9,08	36,98	30	1964	2.5	898	31	44,5	21,4	49
S12	Beja Inrat	1485082302	9,17	36,71	234	1959	2.18	560	28,1	41,3	25,2	49
S13	Thibar	1485769002	9,11	36,52	365	1959	1.86	538	27	39,9	27,4	49
S14	Teboursouk	1485755802	9,25	36,46	440	1959	1.68	551	25,7	40,5	26,9	48
S15	Mjez el Beb	1485429202	9,6	36,65	142	1959	8.5	369	29,7	38,7	24,7	45
S16	Khelidia	1484372423	10,19	36,64	38	1959	10.47	392	34	37,7	23,1	40
S17	Ain Jemmala	1485017702	9,33	36,5	320	1951	29.89	335	25,6	41,8	26,4	35
S18	Ain Tounga SM	1485051102	9,36	36,52	110	1962	12.52	360	27,3	39,8	25,1	43

BASE DE DONNEES HYDRO-CLIMATIQUES

Id	Stations	ID unique	Longitude (°)	Latitude (°)	Altitude (m)	Début observation	Pourcentage des lacunes entre 1959-2009	Pluie annuelle moyenne entre 1959-2009 (mm)	Apport des pluies SON (%) 1959-2009	Apport des pluies DJF (%) 1959-2009	Apport des pluies MAM (%) 1959-2009	Nombre de saisons considérées
S19	Sejnen	1483578005	9,25	37,06	135	1966	0.33	803	31,3	45,8	19,9	50
S20	Bazina	1483078205	9,3	36,97	460	1960	21.88	711	26,6	48,8	21,5	37
S21	Djoumin délégation	1483217005	9,39	36,93	109	1970	1.86	691	27,3	46,8	22,5	48
S22	Oued Joumin djantar	1483486105	9,48	36,96	109	1962	9.65	635	27,5	30,3	31,5	44
S23	Sidi Mbarek(Mat eur)	1483648105	9,66	37,02	40	1981	0.68	520	31,4	42,7	22,5	49
S24	Ghezalla-barrage	1483287805	9,55	37,07	75	1968	18.11	560	30,9	43,6	21,4	38
S25	Jbel Essema	1483215005	9,46	37,21	210	1959	4.39	681	33,7	43,9	19,7	45
S26	Tinja	1483772405	9,76	37,17	8	1959	0.84	569	31	45	20,8	48
S27	Alia	1483061705	10,04	37,17	120	1959	1.58	466	31	44	21,8	47
S28	Ras Djbel	148353005	10,13	37,22	53	1959	8.86	521	33	43,6	20,7	43
S29	Bathan	1485079106	9,85	36,81	145	1959	0.78	379	31	40	23,3	47
S30	Borj el Amri	1485100906	9,89	36,72	55	1959	0.5	412	30,8	37,7	25,1	49
S31	Manouba	1484411924	10,09	36,8	60	1959	5	393	33	39,1	23,4	46
S32	Tunis Manoubia	1484786307	10,17	36,78	66	1959	0.17	455	32,2	39,3	23,5	49
S33	Ariana	1484066522	10,18	36,85	10	1959	0	463	32,9	39,9	22,2	50
S34	Tunis Carthage	1484783207	10,25	36,85	5	1959	0	460	32,6	39,5	22,8	50
S35	Morneg	1484177506	10,33	36,66	80	1959	0.84	467	30,2	42,6	3,3	50
S36	Manzel Bouzalfa	1484421708	10,6	36,68	60	1959	0.53	502	33,4	41,2	20,8	46
S37	Grombelia	1484294908	10,5	36,6	50	1959	0.36	483	34,5	38,8	21,4	48
S38	Nabeul	1484462808	10,75	36,44	1	1959	0.36	399	35,7	37,6	21,2	49
S39	Korba	1484373108	10,87	36,59	9	1959	1	412	38	38,4	19,8	49
S40	Kelibia	1484362908	11,12	36,85	3	1959	0.5	517	35,9	41,5	19,6	50
S41	Bir Mcherga	1484096207	9,97	36,51	155	1959	1.18	411	30,2	36	26,5	49
S42	Zaghouen	1484807707	10,13	36,43	230	1959	1.16	460	32,3	36	25,9	49
S43	Saouef	1484560707	10,15	36,24	170	1959	3.2	364	33,9	35	24,1	44
S44	Cherfech	1485161607	10,05	36,93	59	1959	2.67	445	31,3	41,6	22,2	44
S45	Pond de Fahes	1484524607	9,89	36,39	175	1959	3	380	32,1	33,7	27,7	48
S46	Sidi Boubaker	1484623207	9,8	36,25	340	1959	3.18	452	29,3	34,03	27,5	47
S47	Bouarada dre	1484130704	9,61	36,35	260	1952	40.9	221	27,3	36,4	27,5	46
S48	Ain Ghesil	1485024404	9,58	36,25	563	1961	12.83	316	28	36,5	25,5	43
S49	Krib	1485377804	9,14	36,32	447	1959	2.01	491	25,5	39,7	28,6	48
S50	Ain Tabia	1485046704	9,18	36,28	414	1968	17.33	343	28,3	40	26,6	42
S51	Akouat Gare	1485059104	9,26	36,26	140	1932	2.18	392	26,9	34,6	32,2	49
S52	Siliana	1485676404	9,37	36,08	431	1959	0.67	405	29,3	30,9	29,7	50
S53	Saadia Bargou	1486547804	9,6	36,04	524	1959	0.16	599	26	36,7	28,8	50
S54	Barrage Lakhmes	1485076704	9,47	36,01	475	1959	4.34	425	28,6	32,3	27,7	47
S55	Ain Zakkar	1485052704	9,54	36,03	676	1932	6.15	370	28,03	35,6	24 ,6	46
S56	Makthar	1485410204	9,2	35,85	900	1932	0	503	27,5	32,2	29	50

BASE DE DONNEES HYDRO-CLIMATIQUES

Id	Stations	ID unique	Longitude (°)	Latitude (°)	Altitude (m)	Début observation	Pourcentage des lacunes entre 1959-2009	Pluie annuelle moyenne entre 1959-2009 (mm)	Apport des pluies SON (%) 1959-2009	Apport des pluies DJF (%) 1959-2009	Apport des pluies MAM (%) 1959-2009	Nombre de saisons considérées
S57	Rohia	1486543104	9,06	35,65	622	1959	17.01	240	29,9	24,2	31	40
S58	Sers	1485361903	9,02	36,07	501	1959	13.32	351	28,2	32,2	30,5	43
S59	Zouarine	1485827203	8,9	36,02	571	1965	9.01	349	27,07	33,2	29,9	46
S60	Kef	1485361903	8,74	36,13	491	1959	1	442	24,3	32,5	31,9	47
S61	Gardiamou	1485286403	8,43	36,25	195	1959	5.71	415	26,2	37,8	29,6	46
S62	Tajerouine	1485732803	8,58	35,95	750	1959	10.74	355	27,3	29,1	30,7	41
S63	Oued Mellegue K13	1485499003	8,5	36,12	324	1959	0.84	306	28,4	28,1	30,9	49
S64	Kalaat Lasnem	1485352503	8,34	35,77	623	1959	9.9	343	28	26	31,1	41
S65	Sakiat_Sidi_Youssef	1485550203	8,35	36,22	803	1959	10	454	24,3	36,1	29,8	45
S66	Ain Zeligua	1485055304	8,83	35,87	853	1959	30.02	286	27,4	29,4	32,2	36
S67	Dehmani municipalité	1485251003	8,83	35,94	622	1957	0	445	28,2	29,1	29,8	50
S68	Ksour école	1485383904	8,89	35,9	720	1913	12.7	342	27,5	30,3	31,5	43
S69	Ksour élevage	1485250804	8,82	35,96	600	1959	0	389	28,7	28,2	29,7	50
S70	Dehmani élevage	1485250803	8,81	35,92	652	1932	18.03	304	28	30,1	30,3	37

La répartition des stations retenues dans l'étude est illustrée dans la figure II-4 par leurs identifiants. Sur les trois régions bioclimatiques, les stations sont réparties comme suit : 15,7 % des stations dans la région humide, 20% dans la région sub-humide et 64,3 % dans la zone semi-aride. En raison de cette répartition géographique, les pluies moyennes annuelles (Tab. II-1) montrent une variabilité importante entre les stations.

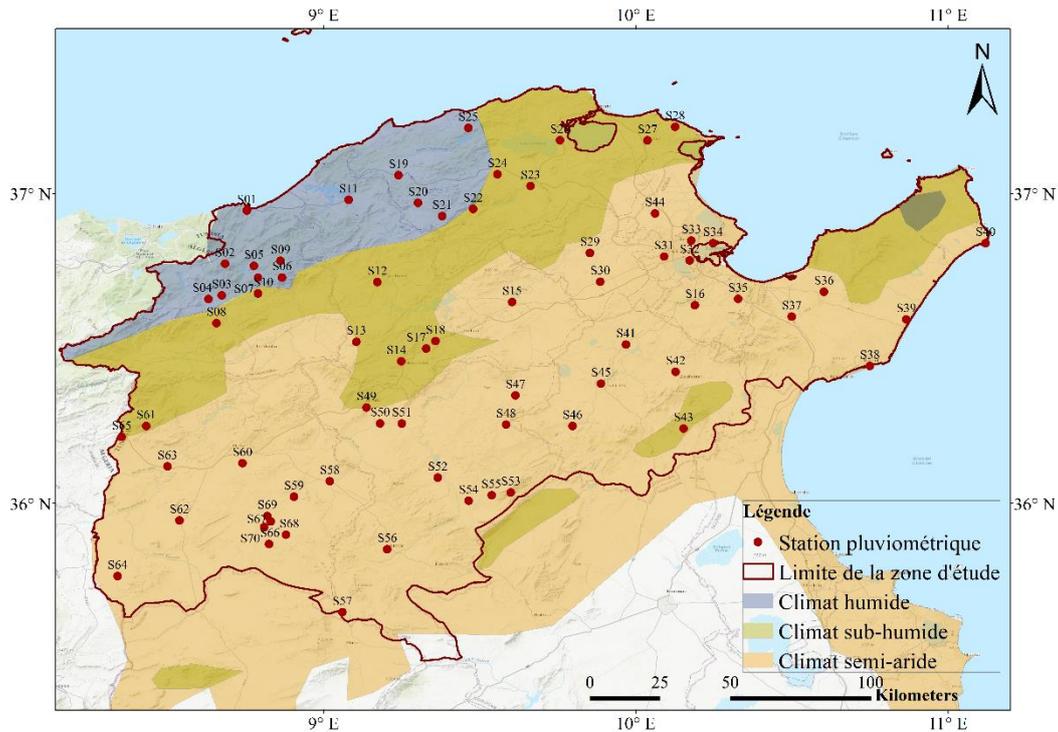


Figure II-4 Localisation des stations pluviométriques dans la zone d'étude

II.4.2 Choix de la saison d'étude

La figure II-5 représente les pluies moyennes mensuelles sur 50 ans (1959 - 2009) de septembre à aout dans huit stations de la zone d'étude. Dans la majorité des stations, les mois les plus pluvieux sont entre décembre (Ain Drahim, Nefza délégation, Sejnen et Tunis Carthage) et janvier (Saadia Bargou et Kef) sauf pour les stations de Zaghouen et Nabeul le mois le plus pluvieux est octobre. Les mois Juin, juillet et aout enregistrent les cumuls les plus faibles entre 1959 et 2009. Trenberth. (1983) a présenté une définition météorologique des saisons en se basant sur la variabilité mensuelle de la température en moyennes latitudes. Selon cette définition, les 12 mois de l'année se divisent en quatre saisons : (1) l'hiver : décembre-janvier-février, (2) le printemps : mars-avril-mai, (3) l'été : juin-juillet-aout, (4) l'automne : septembre-octobre-novembre. Comme la Tunisie est située dans les moyennes latitudes, la même répartition des saisons est adoptée dans cette étude. Les précipitations hivernales (DJF) sont considérées. Ce choix est corroboré par les observations des réseaux étudiés. En effet, comme le montre le tableau II-1, pour la plus grande partie des stations (86 %), les mois DJF apportent le plus grand cumul comparativement aux mois (SON) et (MAM).

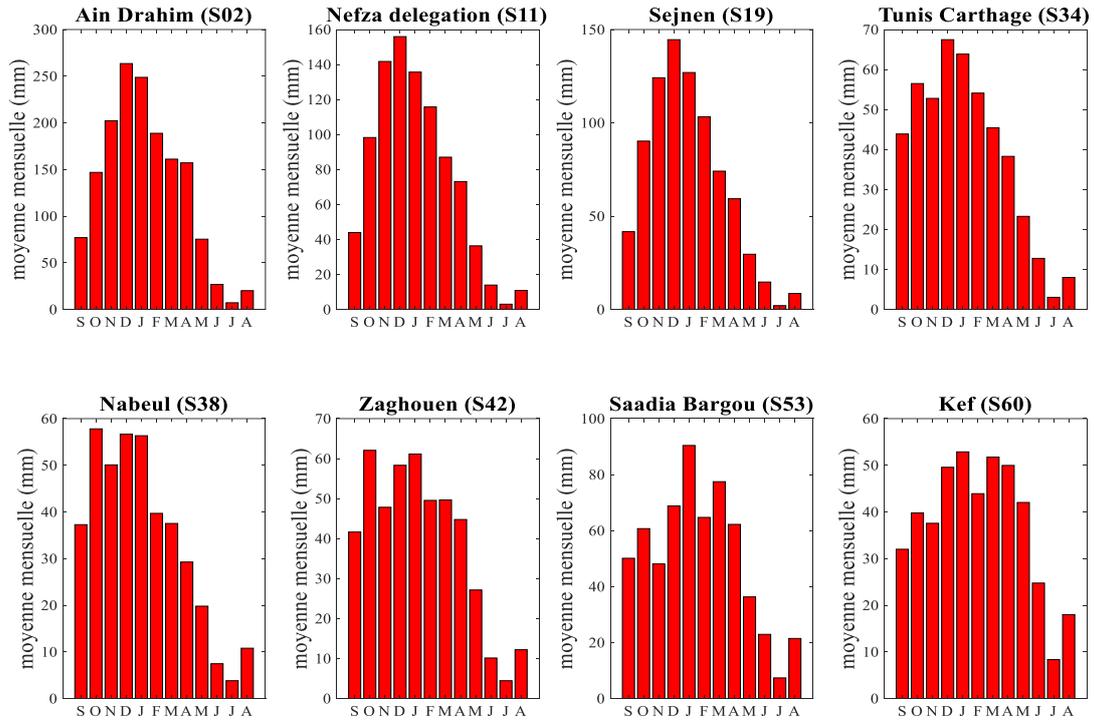


Figure II-5 Moyenne mensuelle de précipitations sur 50 ans (1959-2008) dans 8 stations pluviométriques réparties dans la zone d'étude

Sur le plan national, cette saison est d'une grande importance, spécialement, pour l'agriculture pluviale et pour la mobilisation des eaux. Sur le plan régional, les pluies hivernales sont fortement influencées par les flux nordiques et particulièrement les flux nord-ouest et cela permet de détecter les téléconnexions avec les phénomènes climatiques à grande échelle qui sont très actifs pendant l'hiver (Ben Boubaker, 2003 ; Xoplaki et al., 2004 ; Brandimarte et al., 2011 et Jelassi et al., 2016). Notons qu'une saison DJF d'une année m correspond à décembre de l'année $m-1$, janvier de l'année m et février de l'année m à la suite du manuscrit.

L'existence des valeurs manquantes dans les séries journalières DJF entraîne l'élimination de la saison entière de l'analyse. La figure II-6 visualise les saisons couvertes par des séries complètes sur la période de 50 ans et dans les 70 stations de la zone d'étude. Seulement 11 stations possèdent des données complètes d'où 50 saisons DJF. Alors que les 59 stations restantes contiennent un nombre variable des vides, varie de 1 à 23, sur la période d'étude (Tab.II-1, dernière colonne). Réellement, le nombre total des saisons considérées est de 3 137 mais dans l'analyse qui suit, nous avons considéré une période commune de 50 ans (1960-2009) pour toutes les stations.



Figure II-6 Période couverte par des séries journalière DJF complètes dans les 70 stations

Une analyse des séries chronologiques des anomalies du cumul saisonnier (DJF) moyenne dans les 70 stations pluviométriques (Fig.II-7). Cette anomalie est définie comme l'écart entre le cumul saisonnier annuel (moyenne de toutes les stations) et le cumul saisonnier moyen sur 50 ans d'observations. La courbe des anomalies a montré une persistance d'anomalies de même signe sur deux périodes. En effet, des fortes anomalies positives ont été enregistrées successivement, sur toutes la zones d'étude, entre les années 2003 et 2006,. On constate aussi que des anomalies négatives ont persisté 7 années successives entre 1974 à 1980.

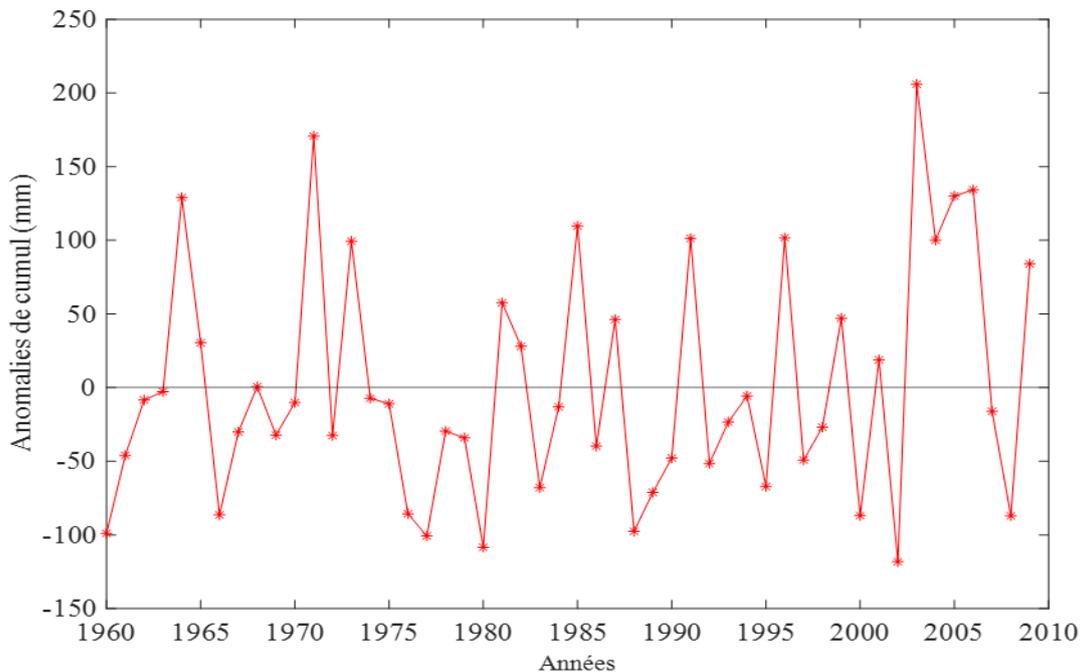


Figure II-7 Anomalies moyenne du cumul saisonnier DJF dans les 70 stations entre 1960-2009

II.5 Simulations numériques et modèle climatiques

II.5.1 Principe de simulation numérique et modèles climatiques globaux GCM et régionaux RCM

Le système climatique est composé majoritairement de cinq composantes : (1) l'atmosphère, (2) les océans, (3) la cryosphère (glace et neige), (4) la lithosphère (la surface terrestre), (5) la biosphère (les organismes vivants) (Fig. II-8). Les interactions entre ces différentes composantes déterminent la circulation et l'échange de l'énergie, principalement reçue du Soleil (IPCC, 2007). La complexité du système climatique et la limitation des observations, dans le temps et dans l'espace, rendent une analyse analytique ou expérimentale de ce système dans sa globalité très difficile d'où vient la nécessité de modéliser les processus physiques en jeu. Les modèles climatiques sont basés sur un ensemble d'équations mathématiques qui décrivent les lois physiques qui régissent le système climatique. Ils sont utilisés pour comprendre le système du globe terrestre, tester des scénarios et tirer des conclusions sur les conditions climatiques passées et futures et pour détecter les anomalies les événements extrêmes et les signaux du changement climatiques. Historiquement, La modélisation climatique a commencé par un modèle conceptuel simple et cette description a été améliorée à la fin du 17^e siècle en prenant en compte la circulation atmosphérique générale dominante et spécialement la cellule de Hadley (Hadley, 1735), la cellule de Ferrel (Ferrel, 1856) et la cellule polaire. Le premier modèle mathématique a lieu au début de 19^e siècle par Fourier (Fourier, 1822). Ce modèle se base sur l'équilibre entre l'énergie solaire reçu par le globe terrestre et les radiations réémises vers l'espace. Au début du 20^e siècle, l'évolution de l'atmosphère a été décrit par un ensemble d'équations : l'équation de conservation de la masse, de la quantité de mouvement et l'énergie avec l'équation des gaz parfait (Bjerknes, 1904 et 1995). Cette théorie était la base pour plusieurs autres recherches sur la résolution numérique de ces équations dans l'absence des outils de calcul puissants à cette époque. Pendant les années 1960, les premières simulations numériques du climat ont été élaborées seulement avec la composante de l'atmosphère. Avec la progression successive des outils de calcul, les modèles climatiques sont améliorés en faisant des couplages avec les autres composantes (Froidurot, 2015).

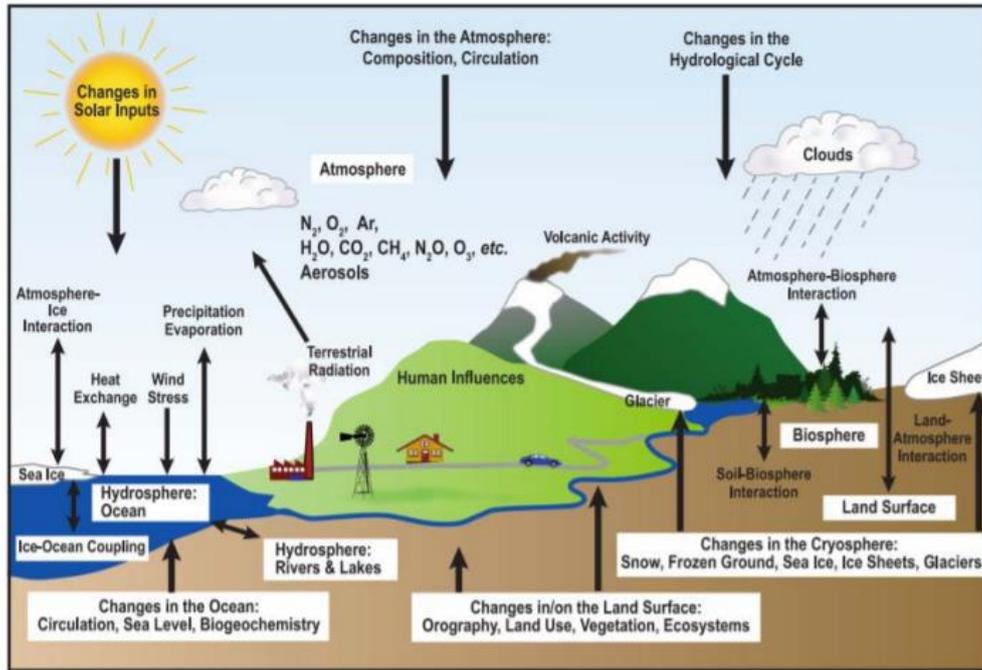


Figure II-8 Illustration des différentes composantes du système climatique et leurs interactions (IPCC, 2007)

Dans les modèles climatiques, le globe terrestre est divisé en une grille tridimensionnelle de cellules (Fig.II-9). Les processus sont modélisés dans chaque cellule et les résultats trouvés sont ensuite transmis aux cellules voisines pour modéliser l'échange d'énergies et de matières dans le temps. La taille de la cellule de la grille représente la résolution spatiale du modèle. Plus les cellules de la grille sont de petites tailles, plus le modèle est détaillé, mais cela exige plus de puissance de calcul. Dans les modèles, il y a aussi le facteur du temps, appelée le pas de temps ou résolution temporelle, qui s'exprime en minute, en heure, en jour, en mois et en année. Un faible pas du temps donne des résultats plus détaillés, mais le calcul est extrêmement coûteux.

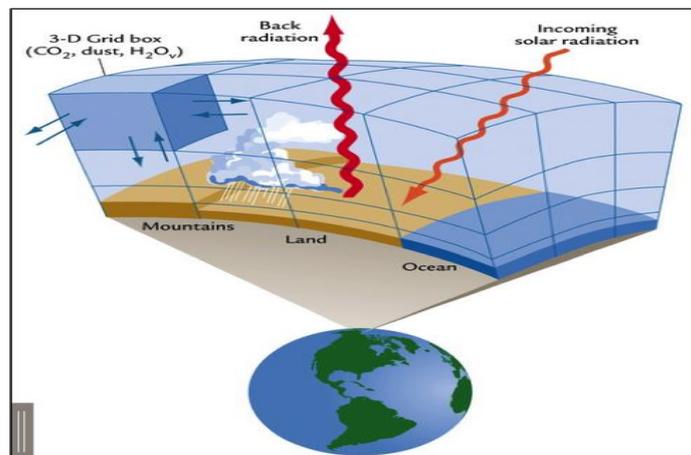


Figure II-9 Schématisation de la grille tridimensionnelle du modèle climatique (source : <https://news.climate.columbia.edu/2018/05/18/climate-models-accuracy/>)

En se basant sur la taille des cellules de grille, on peut distinguer les modèles climatiques globaux (GCM : global climate model) et les modèles climatiques régionaux (RCM : Regional climate model). L'assimilation des données d'observation n'est pas utilisée dans les GCM et les RCM pour cela ces modèles sont exécutés sur des longues périodes temporelles. La différence entre les deux modèles, c'est que le GCM possède une résolution spatiale relativement faible entre 100 et 300 km. Cette résolution n'empêche pas le modèle de donner de bonnes descriptions de la circulation atmosphérique générale. Par contre, il se trouve moins performant dans la description des processus à petite échelle comme la convection, les vents locaux et autres. Les RCM simule le climat sur une surface limitée ce qui permet d'adopter une forte résolution spatiale de l'ordre de 10 à 50 km. Ces modèles régionaux sont pilotés à leur condition aux limites par des autres modèles de résolution plus faible : des modèles globaux, des réanalyses ou d'autres RCM. Pour ce travail, il était prévu d'utiliser les données du modèle régional MED-CORDEX (Coordinated Regional Climate Downscaling Experiment) mais comme l'accès à ces données était difficile, nous avons choisi d'étudier les données du modèle de réanalyse ERA- Interim utilisé comme conditions aux limites de modèle MED-CORDEX (Ruti et al., 2016).

II.5.2 Modèle de réanalyse météorologique

Une analyse, en météorologie, est définie comme une représentation instantanée de l'état de l'atmosphère. Ces données résultent d'une combinaison entre les modèles (numériques et physiques) et les observations. Ces deux sources de données sont incomplètes. Les observations apportent des informations réelles sur la physique, mais d'un caractère localisé. Par contre, les modèles donnent des interprétations de l'évolution de l'état de l'atmosphère bien réparties dans le temps et dans l'espace, mais ces informations restent des simplifications à une réalité physiquement plus compliquée. À ce propos, les méthodes d'assimilation des données viennent pour combiner les observations et les données des modèles d'une manière optimale. Il s'agit d'introduire les observations dans le modèle et chercher la trajectoire la plus proche des observations (distance minimale) dans une fenêtre d'assimilation typiquement de 6 heures (Blayo, 2002). Les analyses produites quotidiennement par les centres de prévisions sont stockées et mises en commun pour former des conditions météorologiques historiques. Ces données sont corrigées et ajustées par des données d'observations de différentes sources qui étaient disponibles et n'étaient pas utilisées lors des analyses précédentes. Cela assure une meilleure représentation du climat du passé.

II.5.3 Données fournies par les ré-analyses météorologiques : ERA-Interim

ERA-Interim (European Retrospective Analysis Interim) est une réanalyse atmosphérique globale disponible du 1er janvier 1979 au 31 août 2019. ERA-Interim est produite par le Centre européen pour les prévisions météorologiques à moyen terme (ECMWF : European Centre for Medium-Range Weather Forecasts) (Dee et al., 2011). Ces données sont fournies par la communauté scientifique IPSL de la plateforme ClimServ (<https://climserv.ipsl.polytechnique.fr/fr/les-donnees/era-interim.html>). La résolution horizontale est $0.75^{\circ} \times 0.75^{\circ}$ (environ 80 Km) sur 60 niveaux à la verticale entre la surface et 0.1 hPa. La résolution temporelle est de 6 h. la production de ERA-Interim est basée sur le cycle 31r2 du système de prévision IFS (Integrated Forecasting System) de ECMWF (version 2006). Ce système adopte une assimilation variationnelle quadridimensionnelle (4D-VAR) comme technique d'assimilation de données (Fréville, 2015).

Comme indiqué dans la figure II-10, la zone d'étude est répartie dans 12 mailles de ERA-Interim dont 9 mailles contiennent des stations pluviométriques et elles sont notées de M1 à M9 aux centres des mailles. Les Mailles M3, M4, M5 et M7 contiennent le nombre le plus important des stations, ce qui permet de couvrir, spatialement, la quasi-totalité de chacune de ces mailles.

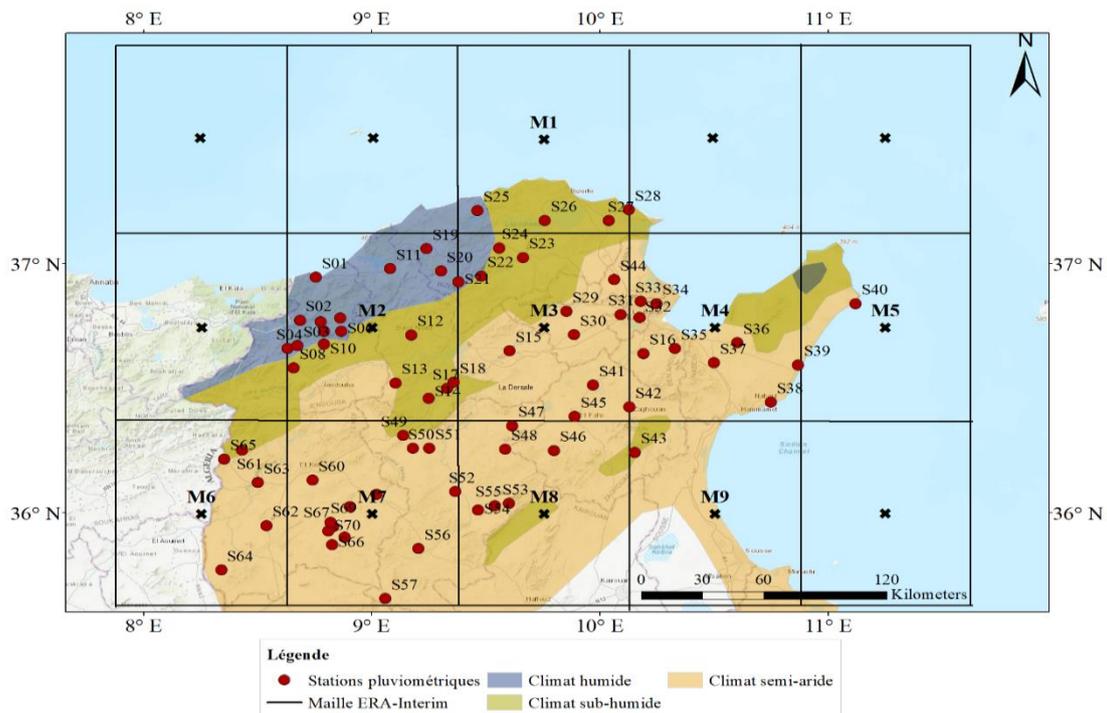


Figure II-10 Représentation de la zone d'étude dans le modèle ERA-Interim

Le tableau II-2 illustre le nombre et la liste des stations pour chaque maille du modèle ERA-Interim.

Tableau II-2 Répartition des stations dans les mailles

Identifiant de la maille	Nombre de stations par maille	Stations dans la maille
M1	4 (5.7 %)	S25- S26- S27- S28
M2	19 (27,1 %)	S01- S02- S03- S04- S05- S06- S07- S08- S09- S10- S11- S12- S13- S14- S17- S18- S19- S20- S21
M3	10 (14.3 %)	S15- S22- S23- S24- S29- S30- S31- S41- S44- S45
M4	10 (14.3 %)	S16- S32- S33- S34- S35- S36- S37- S38- S39- S42
M5	1 (1.4 %)	S40
M6	5 (7.1 %)	S61- S62- S63- S64- S65
M7	14 (20 %)	S49- S50- S51- S52- S56- S57- S58- S59- S60- S66- S67- S68- S69- S70
M8	6 (8.6 %)	S46- S47-S48- S53- S54- S55
M9	1 (1.4%)	S43

II.6 Données de la variabilité climatique

II.6.1 Choix des indices climatique

Le tableau II-3 récapitule les indices des oscillations climatiques utilisés dans cette étude. Les indices de AO, NAO, SOI et PDO représentent les modes de variabilité dominants à une échelle globale où leur influence sont prouvées sur une grande étendue spatiale. Les modes de variabilité et les phénomènes climatiques correspondants ainsi que leurs séries chronologiques sont présentés en détails dans l'annexe A.

Tableau II-3 Tableau récapitulatifs des indices climatiques utilisés (description et source)

Nom	Période	Description et source
AO (oscillation arctique)	1950-2020	L'indice de AO est le premier mode d'analyse EOF (la fonction empirique orthogonale) des anomalies de la hauteur géopotentielle de 1000 millibars à l'hémisphère nord ente les latitudes 20°N et 90°N. Source : https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/ao.long.data
NAO (oscillation nord atlantique)	1821-2015	Le NAO est la différence de pression normalisée entre deux stations (une près des Açores et l'autre près de l'Islande). Source : http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/NAO/
PDO (oscillation décennale du Pacifique)	1900-2018	La PDO est défini comme la première composante principale de la variabilité des SSTs mensuelles du Pacifique Nord polarisé à 20° N. Source : http://research.jisao.washington.edu/pdo/PDO.latest
SOI (indice d'oscillation australe) (ENSO)	1866-2019	Le SOI est la différence de pression normalisée entre Tahiti et Darwin (au Pacifique) Source : https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/SOI/

WeMOI (l'indice d'oscillation Méditerranéenne de l'Ouest)	1850-2018	Le WeMOI est la différence entre les valeurs de pression de surface standardisée enregistrée entre Padua (45.40 °N, 11.48 °E) au nord de l'Italie et San Fernando (Cádiz) (36.28 °N, 6.12 °W) au Sud-Ouest de l'Espagne. Source : http://www.ub.edu/gc/documents/Web_WeMOi.txt
MO (l'oscillation méditerranéenne)	1940-2020	L'indice de MO est obtenu de la différence de pression normalisée entre Alger (36.4 °N, 3.1 °E) et le Caire (30.1 °N, 31.4 °E) Source : https://crudata.uea.ac.uk/cru/data/moi/

Les indices sont utilisés dans l'étude de corrélation avec les descripteurs pluviométriques sur une période de 50 ans et sous forme des moyennes saisonnières DJF (décembre-janvier-février), SON (septembre-octobre-novembre), JJA (juin-juillet-août) et MAM (mars-avril-mai). Les indices climatiques représentés dans la figure II-11 sont standardisés et ils reflètent des différentes situations du climat à l'échelle globale. Une forte amplitude d'un mode de variabilité climatique se traduit par des valeurs élevées de l'indice correspondant (en valeur absolue).

Bien que certains modes de variabilité soient liés à un mois ou une saison, où leurs indices peuvent atteindre des grandes valeurs (par exemple : l'hiver pour l'ENSO et le NAO), le comportement des océans et les anomalies de leurs pression et température commencent des mois à l'avance. À ce propos, plusieurs études ont prouvé que la longue persistance des anomalies atmosphériques est capable d'influencer le régime des précipitations avec un décalage temporel (Tabari et Willems, 2018 ; Wang et al., 2018 ; Aamir & Hassan, 2020 ; Qiao et al., 2021 ; Yang Xing, 2022). Dans cette étude, les relations synchrones entre les précipitations et les indices climatiques sont élaborées pour la saison d'hiver DJF alors que les relations asynchrones se basent sur l'étude de corrélation entre les pluies hivernales et les indices climatiques décalés par une, deux et trois saisons qui précèdent la saison pluviométrique hivernale. Les séries chronologiques représentées par la figure II-11 montrent que la variabilité inter saisonnière est moins importante pour les indices des oscillations du Pacifique. Les valeurs élevées des indices peuvent apparaître généralement dans les quatre saisons pour le AO, le NAO, WeMOI et le SOI. Par contre, le PDO est plus prononcé pendant la saison d'été JJA et la saison d'automne (SON) et le MO a un comportement exceptionnel pendant la saison SON en le comparant avec les autres saisons.

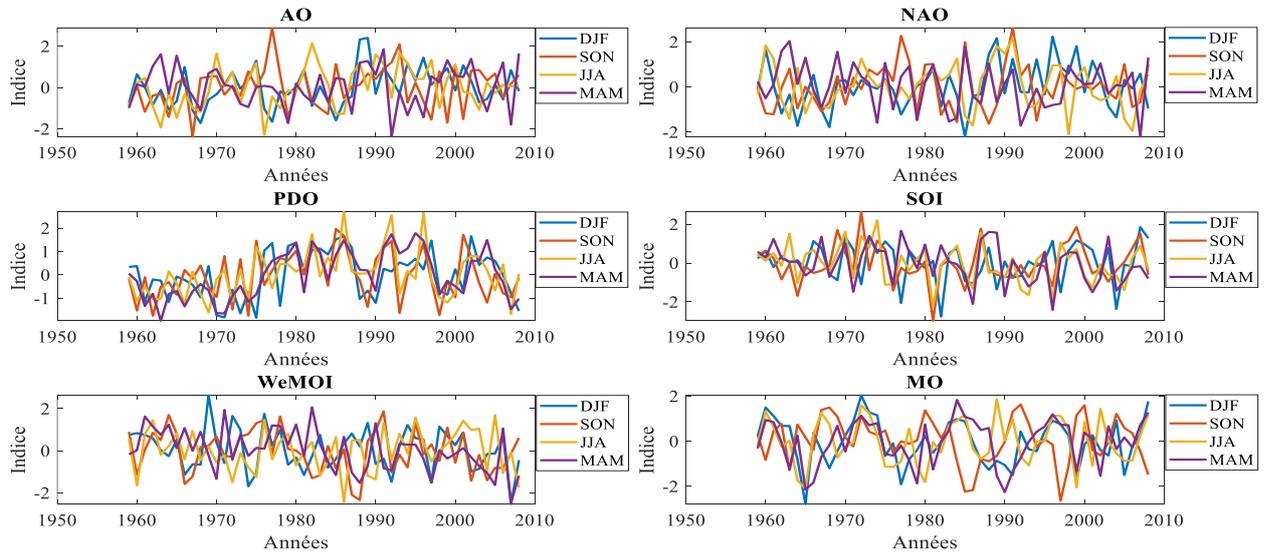


Figure II-11 Séries chronologiques des indices climatiques saisonniers DJF, SON, JJA et MAM entre 1959 et 2008

II.6.2 Données climatiques de large échelle : Pression à la surface de la mer (SLP)

La pression à la surface de la mer (SLP) est un indicateur climatique qui reflète les changements climatiques à une échelle globale et régionale. Cette variable climatique est utilisée pour les calculs de flux air-mer et l'assimilation dans les modèles océaniques et atmosphériques (Høyer and Karagali, 2016). La variabilité de la pression à la surface de la mer à grande échelle affecte les conditions météorologiques et spécialement la température et les précipitations. Plusieurs études sont menées pour étudier les corrélations des SLP avec les précipitations. Par exemple, Thompson and Green. (2004) ont étudié les relations par des techniques statistiques entre le régime de précipitations hivernales dans le bassin méditerranéen et les SLP au nord Atlantique et à la Méditerranée.

Les SLPs de cette étude sont des réanalyses NCEP/NCAR téléchargées du site du NOAA (<https://www.psl.noaa.gov/data/gridded/index.html>) sous forme des moyennes mensuelles caractérisées par une couverture spatiale globale d'un pas de 2.5° en longitude [0E à 357.5 E] et 2.5° en latitude [-90S à 90N] où la dimension de la grille est (144 x 73).

II.7 Conclusion

Les données (in situ) utilisées sont collectées des stations pluviométriques réparties sur toute la zone d'étude à une échelle journalière. Les différentes caractéristiques de ces stations sont identifiées, y compris la position géographique, l'altitude, le pourcentage des données manquantes et la date d'ouverture et les apports saisonniers des précipitations. Après le prétraitement des données et l'élimination des stations avec un pourcentage très élevé des lacunes, 70 stations sont retenues avec une période de 50 ans (1960-2009).

La zone d'étude se répartit sur 9 mailles du modèle avec une répartition des stations pluviométriques par maille très déséquilibrée. En effet, seulement 4 mailles bénéficient d'un réseau relativement dense (plus de 10 stations par maille).

Les indices climatiques avec une probable influence sont identifiés et choisis pour investiguer les liens entre les précipitations de la région et la variabilité climatique à grande échelle.

Chapitre III : Méthodologie d'analyse de précipitations par évènement pluvieux

III.1	INTRODUCTION	40
III.2	TRANSFORMATION DES DONNEES JOURNALIERES EN EVENEMENT PLUVIEUX	40
III.2.1	<i>Délimitation des événements pluvieux par l'analyse d'autocorrélation</i>	<i>40</i>
III.2.2	<i>Choix de descripteurs de pluie par saison</i>	<i>42</i>
III.3	ANALYSE STATISTIQUE UNIVARIEE DES DESCRIPTEURS DE PLUIE	43
III.3.1	<i>Analyse statistique des descripteurs des observations (pluviomètres)</i>	<i>43</i>
III.3.2	<i>Analyse statistique des descripteurs produits par le modèle de réanalyse</i>	<i>53</i>
III.4	METHODOLOGIE D'ANALYSE MULTIVARIEE DES DESCRIPTEURS D'EVENEMENTS PLUVIEUX	58
III.4.1	<i>Approche factorielle linéaire : Analyse en composantes principales.....</i>	<i>58</i>
III.4.2	<i>Approche non linéaire SOM et CAH</i>	<i>59</i>
III.5	CONCLUSION	60

III.1 Introduction

Ce chapitre présente, en première partie, l'événement pluvieux et l'approche du changement d'échelle temporelle des pluies journalières à des événements pluvieux puis à des descripteurs pendant la saison choisie (DJF). En deuxième partie, une analyse descriptive univariée est appliquée aux descripteurs des pluviomètres et du modèle de réanalyse également. Dans la dernière partie, les approches d'analyses multivariées et la démarche à suivre sont présentées et schématisées par des organigrammes.

III.2 Transformation des données journalières en événement pluvieux

III.2.1 Délimitation des événements pluvieux par l'analyse d'autocorrélation

La détermination de la MIET, dans cette étude, est basée sur l'analyse d'autocorrélation des séries chronologiques des précipitations journalières DJF pour chacune des 70 stations sur une période de 50 ans. Comme indiquée dans la figure III-1a et la figure III-1b les deux traits horizontaux représentent l'intervalle de confiance de 95 % pour un processus de bruit blanc. Les valeurs d'autocorrélation à l'extérieur de cet intervalle sont considérées significatives. Les courbes rouges et bleues dans la figure III-1a illustrent les valeurs d'autocorrélation moyenne et médiane, pour toutes les stations pluviométriques. Elles diminuent en fonction de décalage temporel jusqu'à couper la ligne horizontale entre 2 et 3 jours. À partir de 3 jours, les valeurs moyenne et médiane des coefficients d'autocorrélation ne sont plus significatives.

Dans la figure III-1b, les boîtes à moustache représentent les coefficients d'autocorrélation dans les 70 stations. Les coefficients d'autocorrélation pour un jour de décalage sont significatives dans toutes les stations et relativement élevés, où la valeur médiane est supérieure à 0.2, ce qui prouve la dépendance des pluies journalières à un pas de temps d'un jour. Pour le décalage de deux jours, les coefficients d'autocorrélation sont non significatifs (entre les lignes rouges horizontales) dans environ 40 % des stations et très faibles (médiane égale à 0.04) en dehors de l'intervalle de confiance à 95 %. L'autocorrélation est non significative dans 75 % des stations à un pas de temps de trois jours. À un pas de temps de quatre jours, la corrélation est non significative à 89 % des stations. À un pas de temps plus grand, la corrélation est non significative dans 100 % des stations.

Même s'il existe des stations avec une faible corrélation significative à quatre et à cinq jours de décalage, la durée minimale de non-pluie pendant la saison hivernale varie, en moyenne, entre deux et trois jours.

Ce résultat est cohérent avec le régime pluviométrique aux latitudes de la région méditerranéenne à l'échelle synoptique. En effet, le cycle de vie d'un cyclone typique aux

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

moyennes latitudes (développement, maturité et déclin) est de deux ou trois jours. Dans des conditions spéciales et favorables, un cyclone peut atteindre la maturité dans un seul jour. Cependant, cette intensification rapide reste peu fréquente. En outre, les cyclones peuvent persister plus de trois jours selon la nature de front et les facteurs locaux (Hakim and Patoux, 2018). La région méditerranéenne est caractérisée par des conditions topographiques et orographiques particulières et une convection élevée, ce qui affecte et réduit le cycle de vie des cyclones (Fita et al., 2007). En combinant les résultats statistiques avec les aspects physiques de la cyclogenèse, en moyennes latitudes, et sous la condition d'avoir un critère commun pour le découpage les séries chronologiques journalières, nous avons choisi d'adopter la MIET la plus petite qui est deux jours. Autrement dit, on considère qu'une période de deux jours sans pluie est suffisante pour le développement d'un nouvel événement.

La MIET, de deux jours, choisie pour découper les séries chronologiques des observations est également appliquée aux données produites par le modèle de réanalyse.

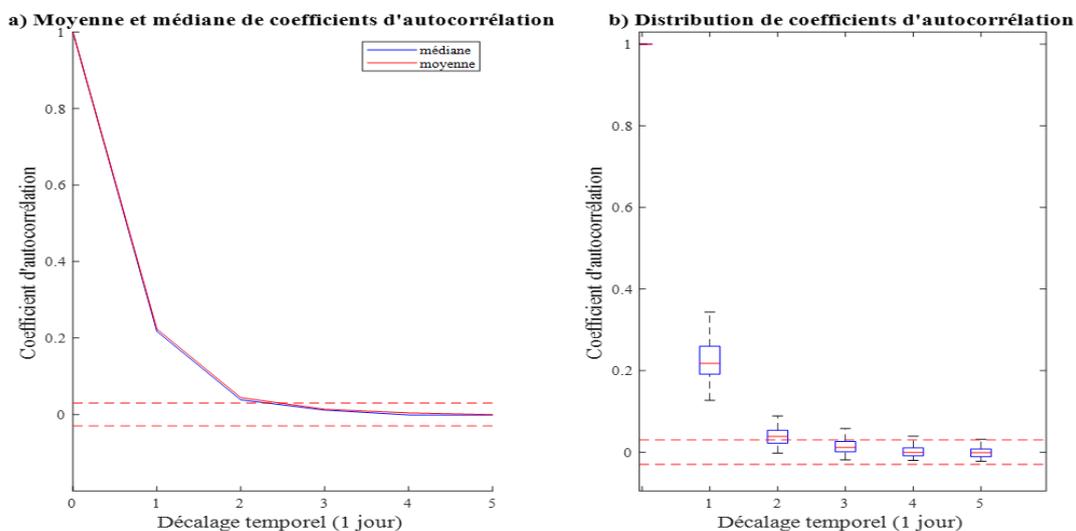


Figure III-1 (a) Moyenne et médiane des coefficients d'autocorrélation de pluie journalière à 70 stations pour 5 pas du temps. (b) Distribution des coefficients d'autocorrélation journalière à 70 stations pour 5 pas de temps

La figure III-2 montre un exemple illustratif dans lequel une période fictive de 18 jours est séparée en des événements pluvieux en considérant un MIET de deux jours. Comme indiqué dans la figure III-2, de découpage de cette période donne un événement représenté par un jour isolé (événement n°1), un deuxième événement de plusieurs jours pluvieux consécutifs (événement n°2) et un troisième événement de plusieurs jours contenant un jour de pause (événement n°3).

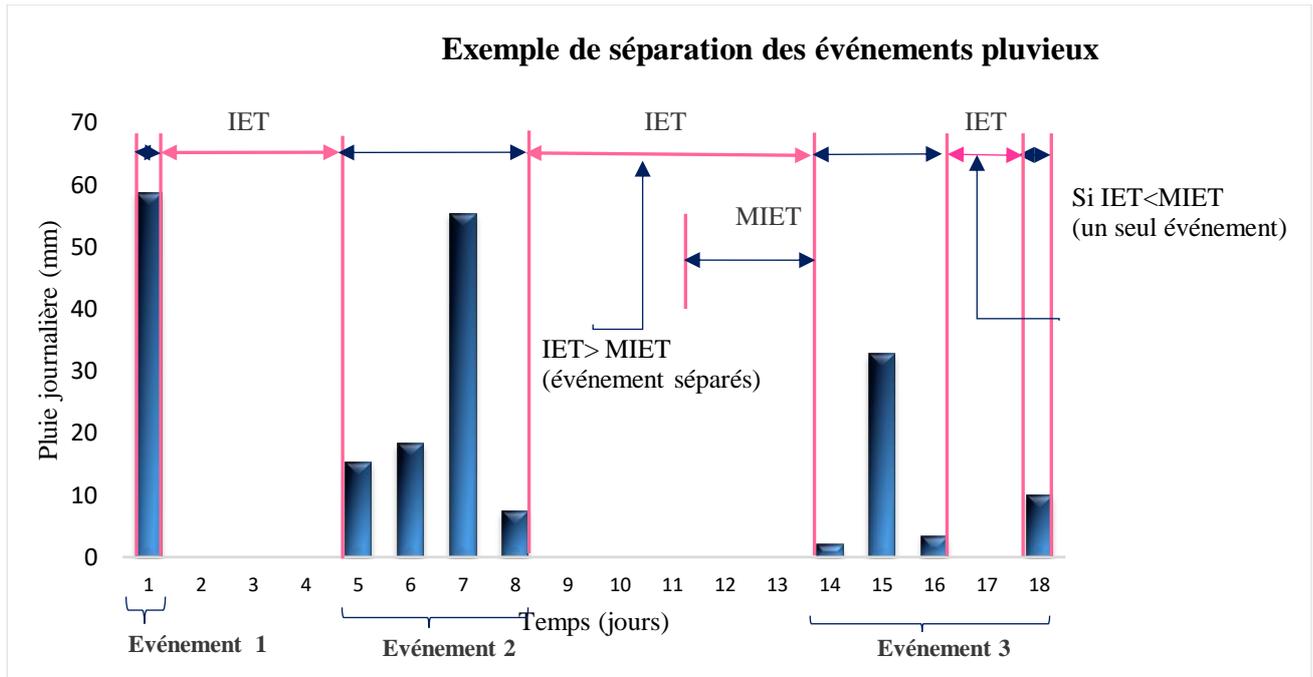


Figure III-2 Principe de séparation des événements pluvieux. IET (inter-event time) est la période sans pluie

III.2.2 Choix de descripteurs de pluie par saison

Une fois la MIET est fixée à deux jours et les événements ainsi définis, les informations, sur la nature et la structure d'événements, seront extraites en s'appuyant sur un nombre de paramètres ou propriétés bien choisies. Dans la littérature, les propriétés sont fortement dépendantes de l'échelle des séries chronologiques et le contexte de la recherche. Par exemple, Dilmi et al. (2017) ont choisi un nombre optimal de propriétés des événements pluvieux extraits des séries chronologiques de précipitations à résolution fine (5 min) qui consistent à la durée d'événement, l'écart type des taux précipitants, le taux de précipitations maximal, la variation absolue du taux précipitant, la hauteur d'eau cumulé de l'événement. Dans le même contexte, Parchure et Gedam (2019), ont analysé des précipitations à l'échelle sub-horaire (15 min) par différentes propriétés des événements comme la durée, le cumul, le taux de précipitations maximal, temps de pic et le temps sec intra-événement, etc. Dans les études mentionnées, les propriétés d'événements choisies sont pertinentes à l'échelle choisie pour l'étude et particulièrement adaptées aux précipitations à résolution temporelle fine. Dans notre étude menée sur des séries chronologiques journalières de la saison DJF, les indicateurs du caractère d'alternance pluie/non-pluie les plus adaptés à cette échelle sont (Tab. III-1) : (1) le nombre d'événements par saison N_e , (2) le nombre de jours pluvieux par saison N_j . Ainsi l'intensité et la hauteur d'eau cumulées de la saison sont définies par deux descripteurs : (3) le cumul total saisonnier C_t , (6)

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

le cumul moyen par jours pluvieux $C_{m,j}$. Ces derniers représentent des caractéristiques globales de la saison DJF indépendamment de la qualité des événements. La structure des événements pluvieux pendant la saison DJF est définie par deux descripteurs qui sont : (4) le cumul moyen par événement $C_{m,e}$ et (5) la durée moyenne d'événement par saison $D_{m,e}$. Les six descripteurs caractérisant les épisodes pluvieux DJF sont illustrés dans le tableau III-1.

Tableau III-1 Définition des descripteurs pluvieux saisonniers (pour chaque station pluviométrique), X_i représente le cumul d'un jour i et D_k représente la durée de l'événement k

N°	Nom	Symbole	Unité	Quantification
1	Nombre d'événements	N_e	Événements/saison	
2	Nombre de jours pluvieux	N_j	Jours/saison	
3	Cumul total saisonnier	C_t	mm	$C_t = \sum_i^{N_j} X_i$
4	Cumul moyen par événement	$C_{m,e}$	mm/événement	$C_{m,e} = \frac{C_t}{N_e}$
5	Durée moyenne d'événement	$D_{m,e}$	jours/événement	$D_{m,e} = \sum_k^{N_e} \frac{D_k}{N_e}$
6	Cumul moyen par jour pluvieux	$C_{m,j}$	mm/jour/saison	$C_{m,j} = \frac{C_t}{N_j}$

En général, les analyses exploratoires et les approches de classification sont souvent précédées par des analyses statistiques descriptives comme une phase préliminaire. Cela permet de mieux comprendre les données en présentant une synthèse numérique des données (minimum, maximum, moyenne, médiane, percentile, etc). Une analyse univariée des aspects spatial et temporel des descripteurs (du modèle et des données du terrain) est réalisée dans la section suivante.

III.3 Analyse statistique univariée des descripteurs de pluie

III.3.1 Analyse statistique des descripteurs des observations (pluviomètres)

III.3.1.1 Distribution des descripteurs des stations pluviométriques

La figure III-3 illustre les histogrammes de distribution des indices pluviométriques calculés pour la saison DJF pendant 50 ans d'observations dans 70 stations pluviométriques (3500 saisons en total). L'histogramme du nombre d'événements N_e est symétrique et les valeurs varient entre 2 et 15 avec un mode égale à 10 événements par saison. Le nombre de jours pluvieux N_j est caractérisé par une distribution bimodale où le premier mode est environ 20 jours et le deuxième est égal à 30 jours pluvieux pendant la saison DJF. La distribution du cumul total saisonnier C_t est asymétrique. Les valeurs sont inférieures à 500 mm avec un mode

égal à 117 mm. Les saisons avec un total pluviométrique supérieur à 500 mm sont peu fréquentes. La distribution de cumul moyen par événement $C_{m,e}$ est asymétrique à gauche, c-à-dire les faibles valeurs sont plus fréquentes. L'histogramme de la durée moyenne par événement $D_{m,e}$ montre que les valeurs les plus fréquentes sont entre 2 et 3 jours /événement. Dans quelques situations particulières, $D_{m,e}$ peut dépasser le 10 jours/événement. La distribution de cumul moyen par jour pluvieux $C_{m,j}$ est asymétrique à gauche avec un mode égal à 6 mm/jour ce qui signifie que les saisons de faible intensité de pluie sont dominantes.

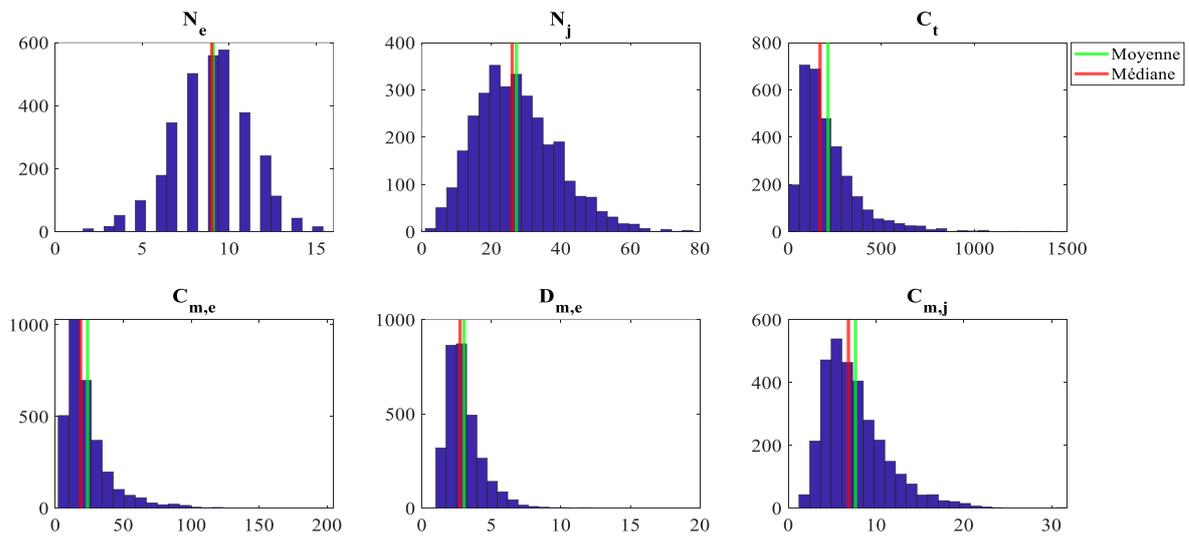


Figure III-3 Histogrammes de distribution des descripteurs de pluie (Nombre d'observations=3500) et illustration des valeurs moyenne et médiane

Dans le tableau III-2, le minimum, le maximum, la médiane et la moyenne des échantillons des descripteurs d'évènements ainsi que le coefficient d'asymétrie sont représentés. Une distribution normale a un coefficient d'asymétrie égale à 0. Si ce coefficient est inférieur à 0.5 (en valeur absolue), la distribution reste approximativement symétrique. On admet que si le coefficient d'asymétrie est compris entre 0.5 et 1 en valeur absolue, la distribution est considérée modérément dissymétrique. Si le coefficient en valeur absolue est supérieur à 1, la distribution est considérée fortement dissymétrique. Pour les observations, la distribution du nombre d'évènement N_e est symétrique et la distribution du nombre de jours N_j est modérément dissymétrique. Par contre, la distribution des quatre autres descripteurs est fortement dissymétrique.

III.3.1.2 Variabilité temporelle des descripteurs des observations

Les six descripteurs choisis pour caractériser la saison sont analysés par la visualisation des boîtes à moustache et représentés dans (Fig.III-4) et (Fig.III-5). Pour chaque année (saison), la forme de la boîte à moustache reflète la distribution de la variable dans les 70 stations. L'allure

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

de la courbe verte (médiane) est un indicateur de l'évolution temporelle globale de la région. La longueur des boîtes à moustache (interquartile) est un indicateur de la variabilité spatiale à l'intérieur de la région d'étude pour une année donnée. Les valeurs considérées comme aberrantes (outliers) représentées par des croix rouges se basent sur la comparaison aux seuils ($S1 = q_1 - 1.5(q_3 - q_1)$ et $S2 = q_3 + 1.5(q_3 - q_1)$) où q_1 et q_3 sont respectivement le premier et le troisième quartile. L'intervalle $\mu \pm 2\sigma$, où μ est la moyenne et σ est l'écart type de la variable, qui est généralement adopté dans le cadre de la loi normale (symétrie) sont ajouté comme limites aux descripteurs qui peuvent être considérés comme symétriques et modérément symétriques alors que l'intervalle $\mu \pm 3\sigma$ est utilisée pour les descripteurs fortement dissymétriques. Par la suite, la distribution de chaque descripteur dans la zone d'étude est représentée par des boîte à moustache par période. La première période est de 1960-2009, puis sur des sous périodes de 1960-1990 et de 1991-2009.

Dans la figure III-4a, la distribution du nombre d'événements N_e est négativement asymétrique (médiane proche du troisième quartile) dans la distribution interannuelle. La médiane de N_e varie entre 5 et 12 événements/saison et les valeurs les plus faibles sont enregistrées en 1995. Durant la période récente (1991-2009) le nombre d'événements N_e a plus de valeurs aberrantes vers la gauche de la distribution (pour les faibles valeurs). Aussi, la médiane et la dispersion de ce descripteur ont baissé.

Dans la figure III-4b, les boîtes à moustache du N_j , de la distribution interannuelle, montrent une asymétrie positive (médiane proche du premier quartile). En se basant sur le nombre important des valeurs dans la probabilité au non dépassement de 0.25 et 0.75, la variabilité entre les stations est importante pour cette variable. La fluctuation de la médiane entre 15 et 40 jours reflète une variabilité interannuelle non négligeable. Les années 1973 et 2003 ont connu des jours pluvieux fréquents durant la saison traduit par une médiane particulièrement élevée (proche de la limite de contrôle supérieur ($\mu + 2\sigma = 40$ jours)). Par contre, les années 1977, 1980, 1988, 1995 et 2002 ont enregistré une faible fréquence des jours pluvieux (proche de la limite de contrôle inférieur $\mu - 2\sigma = 15$ jours/saisons). Par période, le descripteur du nombre de jour N_j ne montre pas des changements entre les deux sous périodes.

Le graphe du cumul saisonnier C_t présenté par (Fig.III-4c) montre aussi une asymétrie positive. Les valeurs médianes varient entre 100 et 300 mm. Les 25^e percentiles et les 75^e percentiles varient respectivement entre 33 à 478 mm. Les valeurs extrêmes apparaissent fréquemment dans toutes les années (saisons) sans exception. Les distributions du cumul saisonnier par

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

période montrent une augmentation de la médiane pendant la sous période la plus récente (1991-2009)

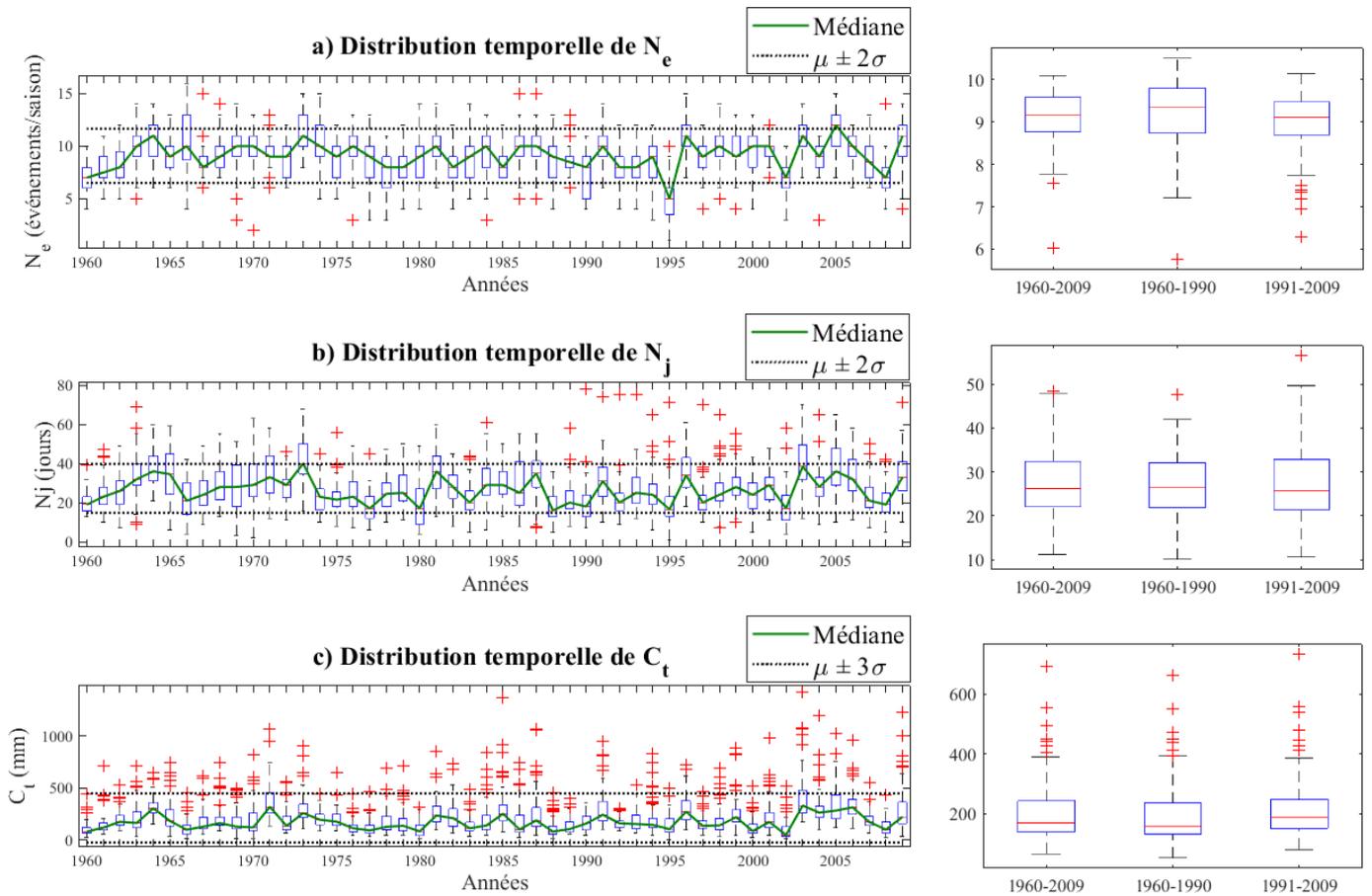


Figure III-4 Variabilité temporelle de distribution des variables (a) N_e , (b) N_j et (c) C_t dans 70 stations entre 1960 et 2009 et sur trois périodes (1960-2009 ; 1960-1990 ; 1991-2009)

Les graphes des variables caractérisant la structure d'événement pluvieux, le cumul moyen par événement $C_{m,e}$ et la durée moyenne par événement $D_{m,e}$, sont respectivement les figures (Fig.III-5a) et (Fig.III-5b). Le cumul moyen par événement $C_{m,e}$ dans (Fig.III-5a) montre une distribution positivement asymétrique dans les distributions interannuelles. En général, la médiane varie entre 7 à 36 mm/événement. Pour certaines saisons comme celles des années 2003 et 2009, $C_{m,e}$ dépasse 200 mm/ événement dans certaines stations ce qui prouve l'existence des événements exceptionnels. Par période, la médiane et la dispersion du cumul moyen par événement augmente dans la sous période (1991-2009).

Le graphe de la durée moyenne par événement $D_{m,e}$ (Fig.III-5b) montre une faible variabilité interannuelle où la médiane varie généralement entre 2 et 4 jours/événement. Les valeurs extrêmes représentent les stations caractérisées par des événements particulièrement longs pendant la saison où $D_{m,e}$ peut atteindre 19 jours/événement (par exemple en 1990). Dans le cas

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

contraire, la durée moyenne par événement peut être très faible (1 jour/événement), voir par exemple les années 1983 et 1992. La distribution par période ne montre pas une différence importante de la durée moyenne par événement.

L'analyse des boîtes à moustache interannuelle du cumul moyen par jour pluvieux $C_{m,j}$ (Fig.III-5c) montre que la médiane varie d'une année à une autre entre 3.7 et 10.2 mm/jour. Les médianes les plus élevées sont enregistrées pendant les années 1971, 1985, 2004 et 2006 qui sont autour de 10 mm/jour. Par période, le 75^e percentile de la $C_{m,j}$ a augmenté considérablement ce qui a donné moins des valeurs aberrantes pendant la période la plus récente (1991-2009).

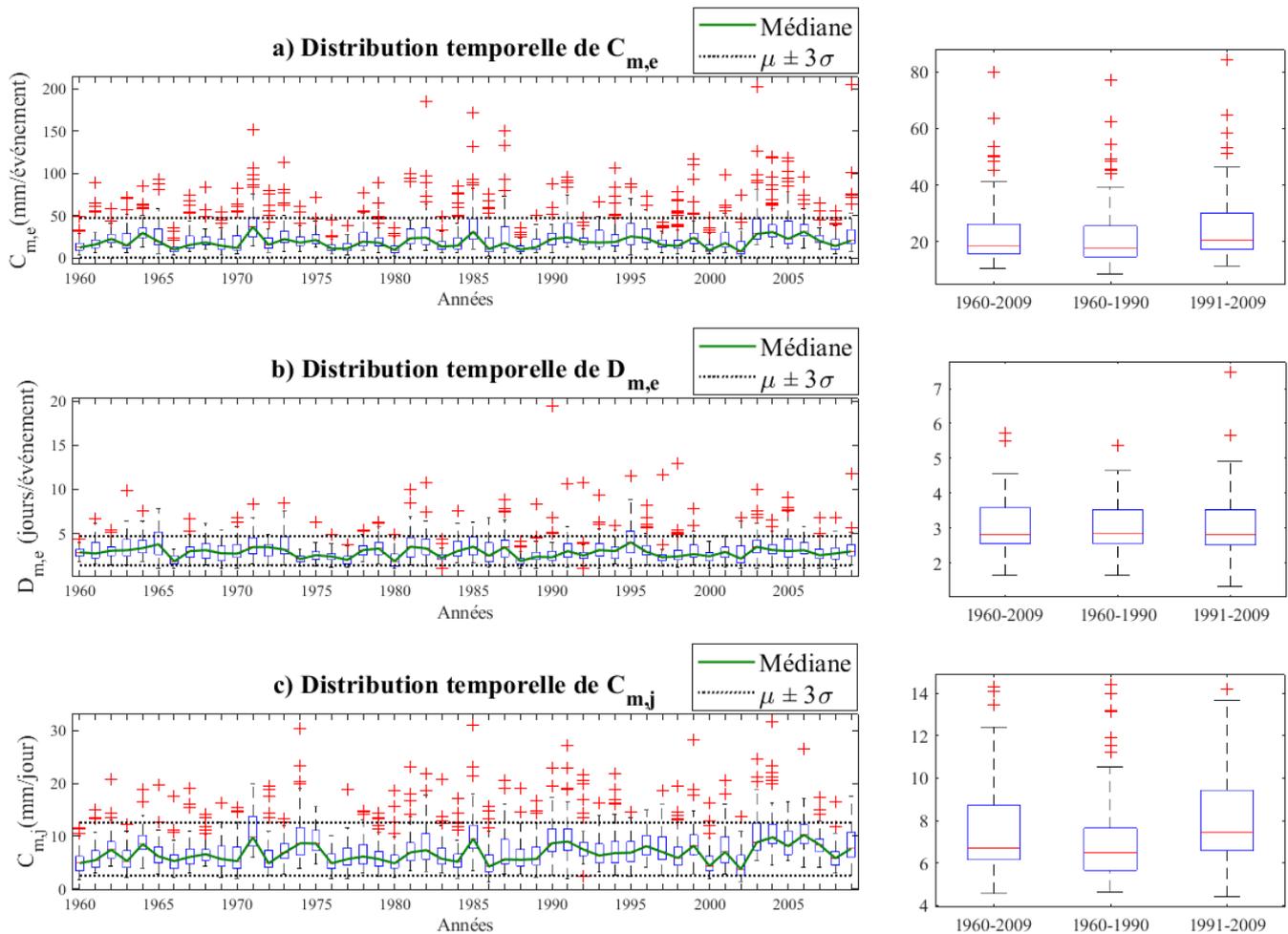


Figure III-5 Variabilité temporelle de distribution des variables (a) $C_{m,e}$, (b) $D_{m,e}$ et (c) $C_{m,j}$ dans 70 stations entre 1960 et 2009 et sur trois périodes (1960-2009 ; 1960-1990 ; 1991-2009).

III.3.1.3 Variabilité spatiale des descripteurs des observations

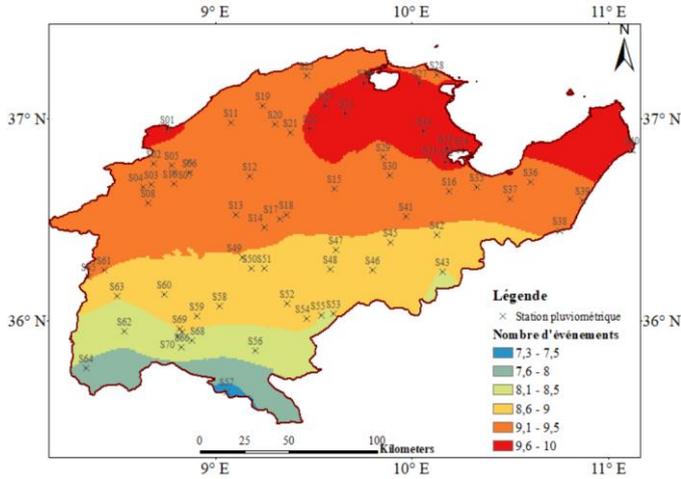
La variabilité spatiale des valeurs moyennes de chaque descripteur de pluie entre 1960-2009 est illustré dans des cartes de la figure (Fig.III-6a, b, c, d, e et f). L'interpolation spatiale des descripteurs est effectuée par la technique de krigeage ordinaire. Le nombre d'évènements N_e (Fig.III-6a) augmente du sud vers le nord et enregistre les valeurs les plus fortes dans la région

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

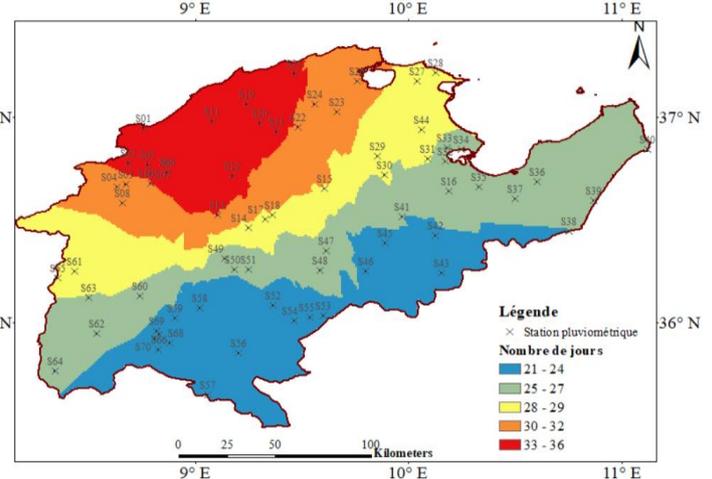
nord-est de la zone d'étude où le nombre d'événements moyen est entre 9.5 et 10 événements par saison. Les valeurs de N_e dominantes dans la zone d'étude sont entre 8 et 9 événements. La variabilité spatiale du nombre de jours N_j (Fig.III-6b) est caractérisée par une orientation NO-SE. L'occurrence de la pluie est plus importante dans la région nord-ouest avec plus de 30 jours de pluie par saison en moyenne. La partie sud est caractérisée par des pluies moins récurrentes (moins de 25 jours par saison). La variabilité spatiale du cumul total DJF C_t a globalement une orientation NO-SE (Fig.III-6c). Les valeurs moyennes les plus fortes, entre 500-700 mm par saison, sont localisées au nord-ouest et d'une faible extension spatiale (dans la station S02 (Ain Drahem) et une station voisine S03 (Ain Debba)). Les stations les moins pluvieuses qui enregistrent moins de 100 mm par saison sont localisées au sud-ouest de la zone d'étude, plus précisément dans les stations S57 (Station Rohia) et S63 (Station Oued Mellegue K13). La plus grande partie de la zone d'étude est caractérisée par un cumul saisonnier moyen entre 100 et 200 mm par saison, représentée en bleu clair (Fig.III-6d). Certaines stations montrent un comportement différent par rapport aux stations voisines, voir par exemple, les stations S11 (Nefza délégation), S20 (Bazina), S36 (Menzel Bouzalfa), S40 (Kélibia) et S53 (Saadia Bargou) qui sont particulièrement plus pluvieuses. Les événements pluvieux les plus intenses sont enregistrés à l'extrême nord-ouest (entre 41 et 60 mm/événement) en moyenne (Fig.III-6d). La région sud est caractérisée par la dominance des événements de faible cumul entre 11 et 20 (mm/événement). Entre les deux régions, il existe une zone intermédiaire où le cumul moyen par événement varie entre 21 et 40 (mm/événement). La durée moyenne d'événement $D_{m,e}$ (Fig.III-6e) varie entre 2,4 et 4,5 (jours/événement) dans une direction NO-SE. Les événements des stations du nord durent, en moyenne, entre 3 et 4 jours et exceptionnellement plus de 4 jours/ événements dans la station S11 (Nefza délégation). La durée moyenne des événements des stations au sud de la région varie entre 2,4 et 3 (jours/événement). La longueur de l'événement ne reflète pas forcément son intensité, par exemple, l'extrême nord-ouest de la zone d'étude qui est le plus pluvieux avec les événements les plus intenses ne possède pas les événements les plus longs. La figure III-6f illustre la répartition spatiale du cumul moyen par jour pluvieux $C_{m,j}$ qui enregistre une variabilité significative sur la zone d'étude par rapport aux autres descripteurs. Dans la station S11 (Nefza), les jours pluvieux sont les plus intenses (entre 10 et 11 mm/jour pluvieux) avec la durée d'événement la plus longue. Dans la zone sud, la région la plus sèche avec des événements pluvieux peu récurrents, il existe des stations avec un cumul moyen par jours pluvieux relativement importants comme celle de S43 (Saouef), S45 (Pond de Fahes) et S53 (Saadia Bargou) qui sont des zones essentiellement forestières.

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

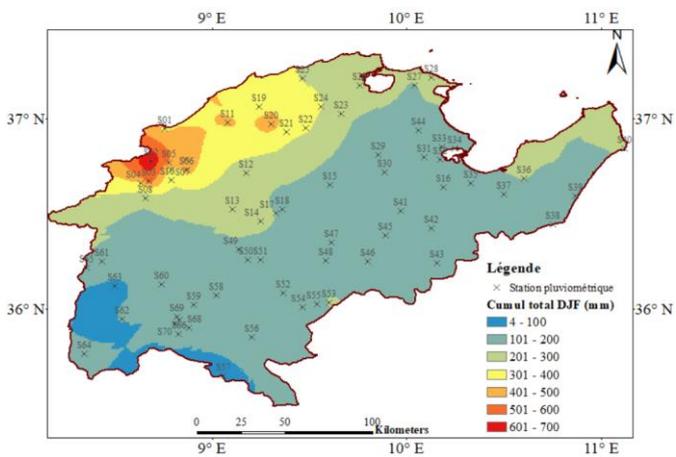
(a) Variabilité spatiale du nombre d'événements N_e entre 1960-2009



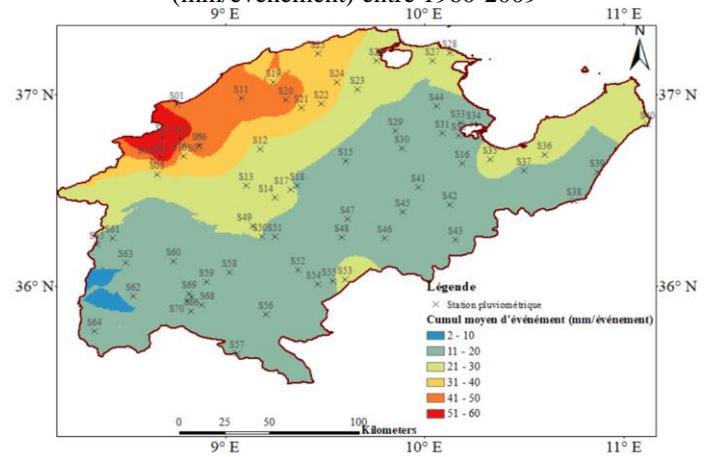
(b) Variabilité spatiale du nombre de jours N_j entre 1960-2009



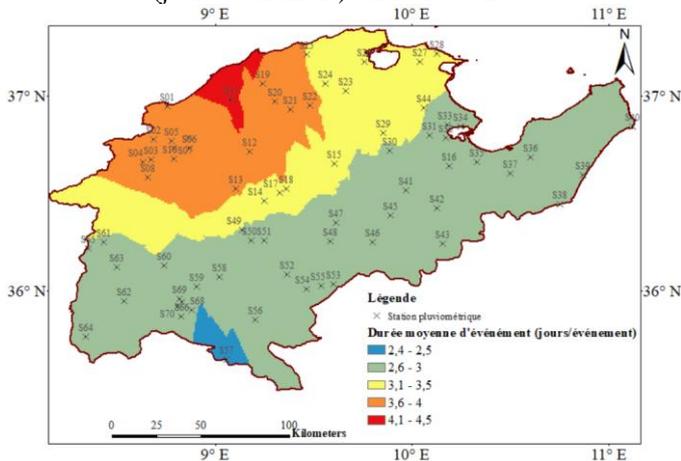
(c) Variabilité spatiale du cumul total C_t (mm) entre 1960-2009



(d) Variabilité spatiale du cumul moyen par événement $C_{m,e}$ (mm/événement) entre 1960-2009



(e) Variabilité spatiale de la durée moyenne par événement $D_{m,e}$ (jours/événement) entre 1960-2009



(f) Variabilité spatiale du cumul moyen par jour pluvieux $C_{m,j}$ (mm/jour) entre 1960-2009

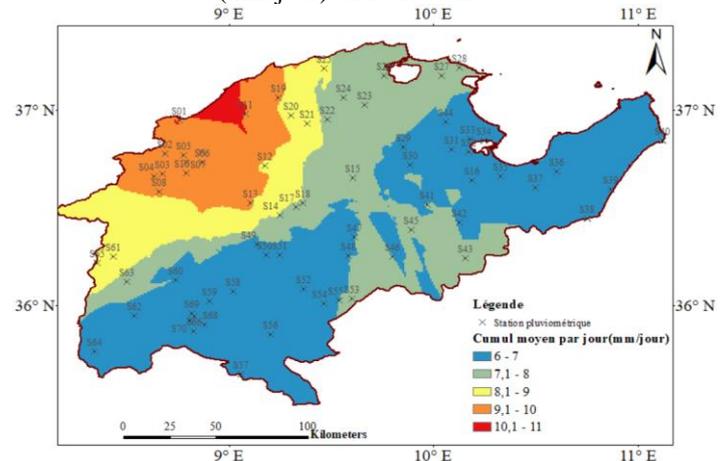


Figure III-6 Variabilité spatiale des moyennes des descripteurs (a) N_e et (b) N_j , (c) C_t , $C_{m,e}$, $D_{m,e}$ et $C_{m,j}$ entre 1960-2009

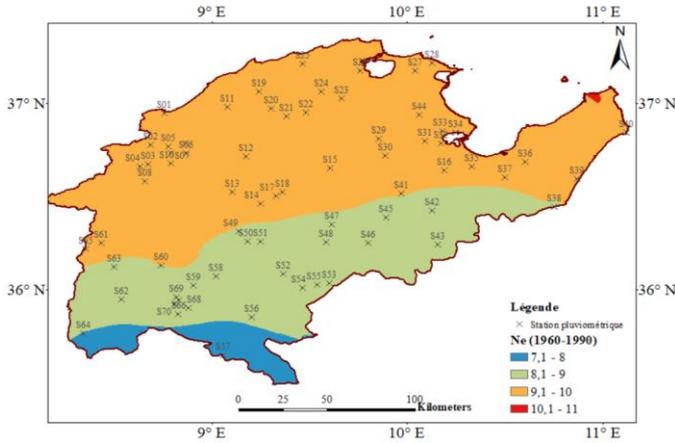
METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

Pour la sous période 1996-1990 (Fig.III-7), la structure spatiale de la majorité des descripteurs ne montre pas une grande différence avec la répartition des descripteurs sur la période 1960-2009 à l'exception du nombre d'évènements N_e (Fig.III-7a) et le nombre de jours N_j (Fig.III-7b).

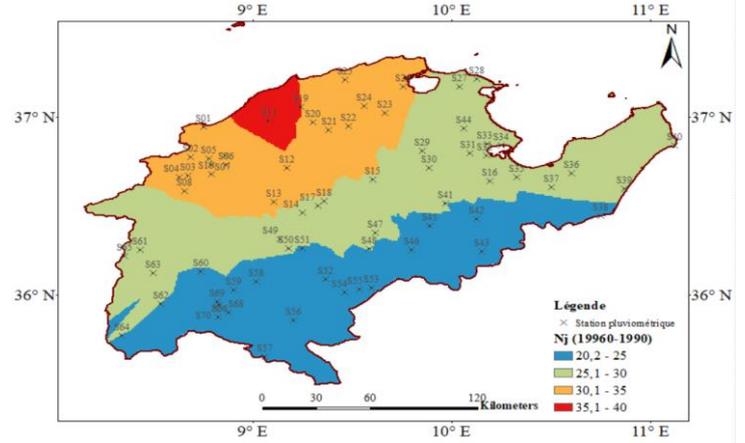
Les cartes de la figure III-8, illustre les écarts des valeurs de chacun des six descripteurs entre les deux sous-périodes adoptées 1960-1990 et 1991-2009 (cartes des anomalies). En observant les structures spatiales trouvées, on remarque que de descripteurs du nombre d'évènement N_e montre un changement entre la partie nord et la partie sud de la zone d'étude entre les deux périodes. Pour le cumul total saisonnier C_t (Fig.III-8c), on observe une structure bipolaire centrée sur la station S13 (Teboursouk). Ce même dipôle apparait dans la carte du nombre de jours N_j (Fig.III-8b) mais avec moindre extension spatiale. Les autres descripteurs ne montrent pas des structures d'anomalies claires.

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

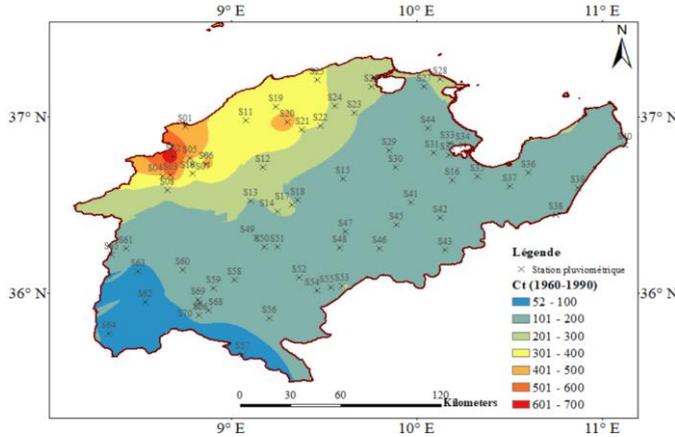
(a) Variabilité spatiale du nombre d'événements (N_e) entre 1960-1990



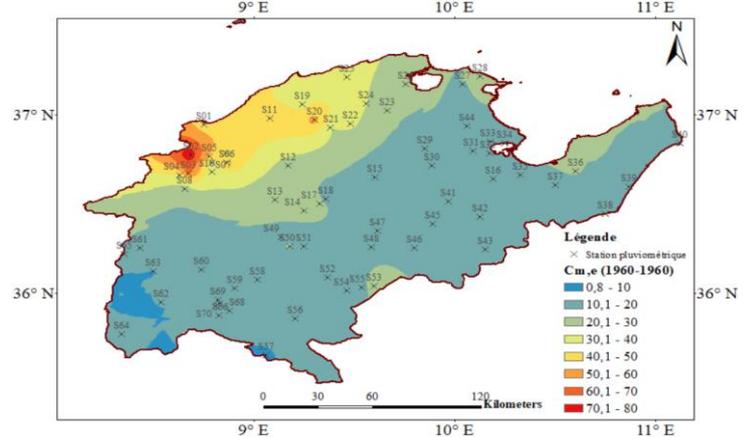
(b) Variabilité spatiale du nombre de jours (N_j) entre 1960-1990



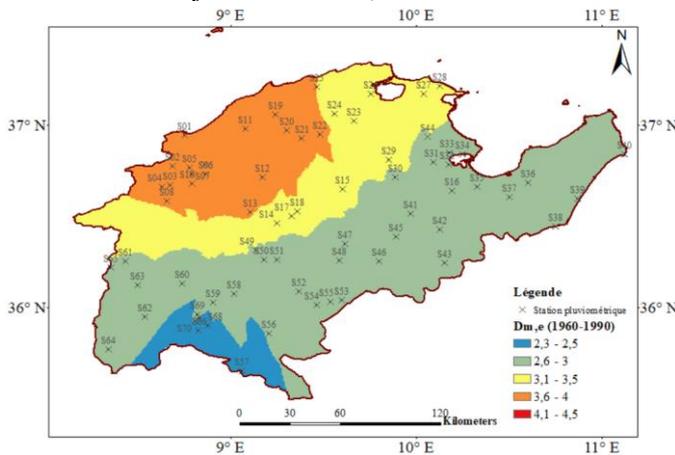
(c) Variabilité spatiale du cumul total C_t (mm) entre 1960-1990



(d) Variabilité spatiale du cumul moyen par événement $C_{m,e}$ (mm/événement) entre 1960-1990



(e) Variabilité spatiale de la durée moyenne par événement $D_{m,e}$ (jours/événement) entre 1960-1990



(f) Variabilité spatiale du cumul moyen par jour pluvieux $C_{m,j}$ (mm/jour pluvieux) entre 1960-1990

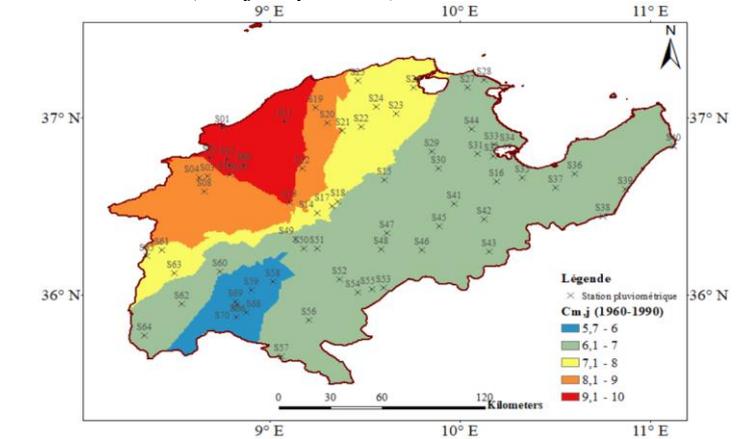
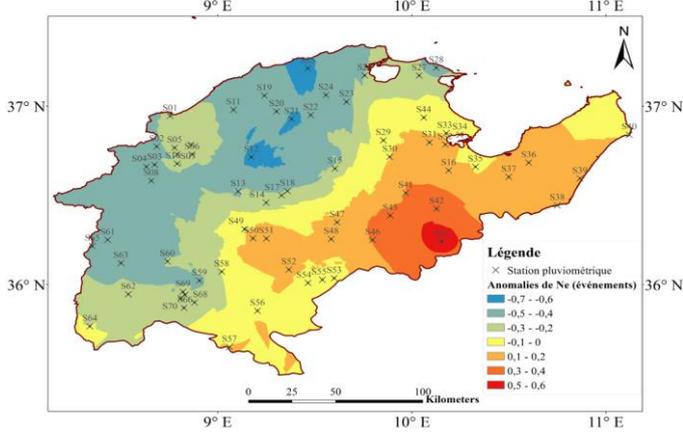


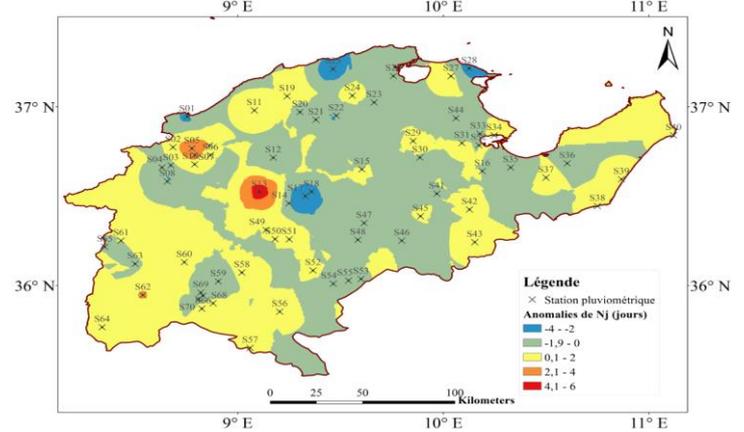
Figure III-7 Variabilité spatiale des moyennes des descripteurs (a) N_e et (b) N_j , (c) C_t , (d) $C_{m,e}$, (e) $D_{m,e}$ et (f) $C_{m,j}$ entre 1960-1990

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

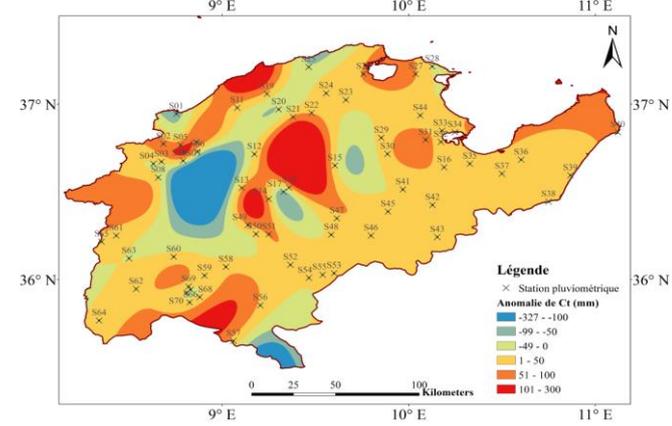
(a) Variabilité spatiale des anomalies du nombre d'événements N_e entre 1960-1990 et 1991-2009



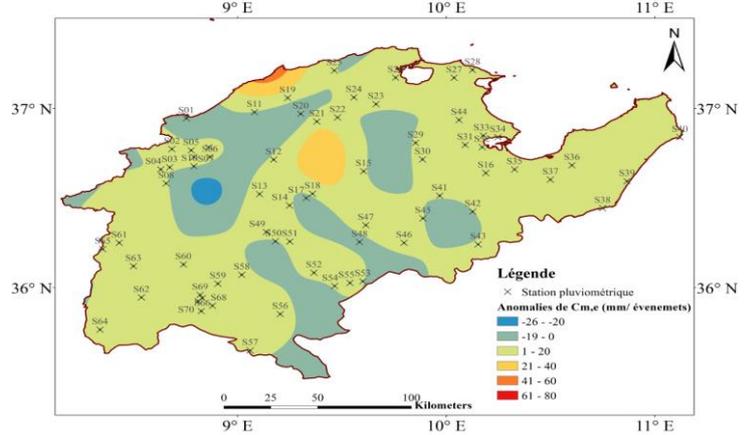
(b) Variabilité spatiale des anomalies du nombre de jours N_j entre 1960-1990 et 1991-2009



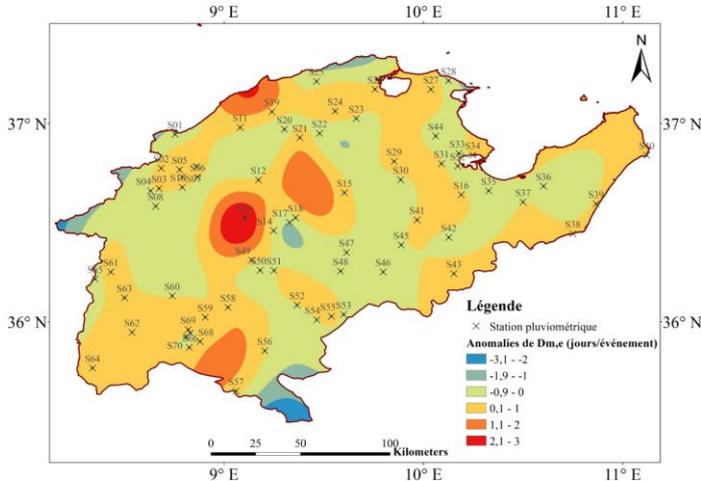
(c) Variabilité spatiale des anomalies du cumul total C_t (mm) entre 1960-1990 et 1991-2009



(d) Variabilité spatiale des anomalies du cumul moyen par événement $C_{m,e}$ (mm/événement) entre 1960-1990 et 1991-2009



(e) Variabilité spatiale des anomalies de la durée moyenne par événement $D_{m,e}$ (jours/événement) entre 1960-1990 et 1991-2009



(f) Variabilité spatiale des anomalies du cumul moyen par jour $C_{m,j}$ (mm/jour pluvieux) entre 1960-1990 et 1991-2009

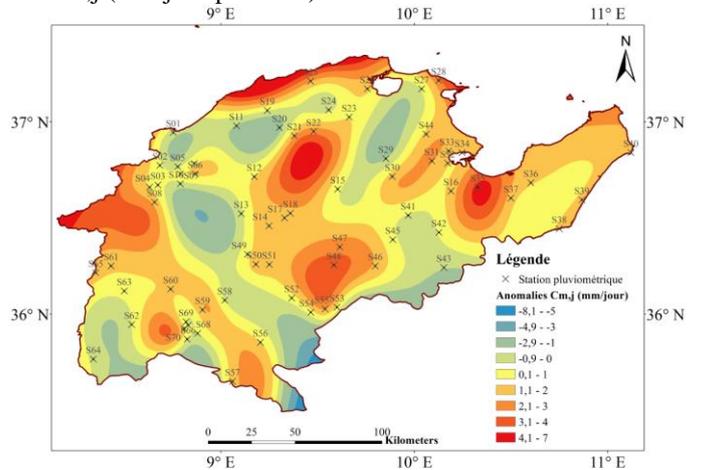


Figure III-8 Variabilité spatiale des anomalies des descripteurs (a) N_e et (b) N_j , (c) C_t , (d) $C_{m,e}$, (e) $D_{m,e}$ et (f) $C_{m,j}$ entre les deux sous-périodes 1960-1990 et 1991-2009

III.3.2 Analyse statistique des descripteurs produits par le modèle de réanalyse

III.3.2.1 Distribution des descripteurs de réanalyse ERA-Interim

Les données journalières produites par le modèle de réanalyses contiennent un nombre important des valeurs extrêmement faibles (jusqu'à l'ordre de 10^{-6} mm). Nous avons considéré que ces valeurs sont des jours sans pluie et elles ont remplacées par 0 mm/jour dans la phase de prétraitement des données et avant le découpage des séries pluviométriques journalières du modèle et l'extraction des événements. Par conséquence, la valeur non nulle minimale existant dans la série du modèle est devenue 0.1 mm/jour. La même démarche de découpage des séries pluviométriques journalières en des événements pluvieux est effectué pour les 9 mailles représentant le nord de la Tunisie dans le modèle de réanalyse ERA-Interim. En adoptant la même MIET que les observations, les descripteurs des saisons DJF sont calculés pour une période de 30 ans de 1980 à 2009. Les histogrammes dans la figure III-9 représentent la distribution de chaque descripteur de pluie et le tableau III-2 récapitule les statistiques élémentaires entre les descripteurs des données observées et les données simulées de réanalyse entre 1980 et 2009 qui représentent la période commune entre les deux sources de données. Le nombre d'événements N_e varie entre 2 et 13 événements par saison avec une moyenne de 7 événements/saison. Le nombre de jours N_j est caractérisé par une distribution bimodale le premier mode est égal à 55 jours et le deuxième est de 70 jours. Le nombre de jours N_j est multiplié par 2.3 dans le modèle par rapport aux observations. Le nombre minimal de jours de pluie est très surévalué par le modèle (31 jours). Le cumul saisonnier C_t varie entre 42 mm et 605 mm/saison. La valeur moyenne est égale à 216 mm qui est presque la même que les observations avec une différence dans la distribution de C_t . En effet, le mode de C_t observé est presque 100 mm contre 200 mm pour les cumuls estimés par le modèle. Les valeurs moyenne et médiane du cumul moyen par événement $C_{m,e}$ sont respectivement 34.9 et 27.8 mm/événement. Le modèle donne des événements pluvieux moins fréquents, mais plus intenses pendant la saison DJF. La distribution de la durée moyenne par événement $D_{m,e}$ montre la dominance des longs événements avec une valeur moyenne égale à 9.6 jours/événements. Cela peut être expliqué par la surestimation du nombre de jours pluvieux N_j . Le descripteur de cumul moyen par jour pluvieux varie entre 1 et 7.2 mm/jours pluvieux, et la valeur médiane est environ 3.4 mm/jour pluvieux ce qui décrit une faible intensité journalière de pluie.

En général, le modèle fourni un cumul saisonnier comparable avec celui des observations, mais, la nature d'événements et la récurrence de pluie sont différentes. En effet, ce même cumul est

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

produit par beaucoup plus des jours de pluies (par rapport aux observations) mais de faible intensité.

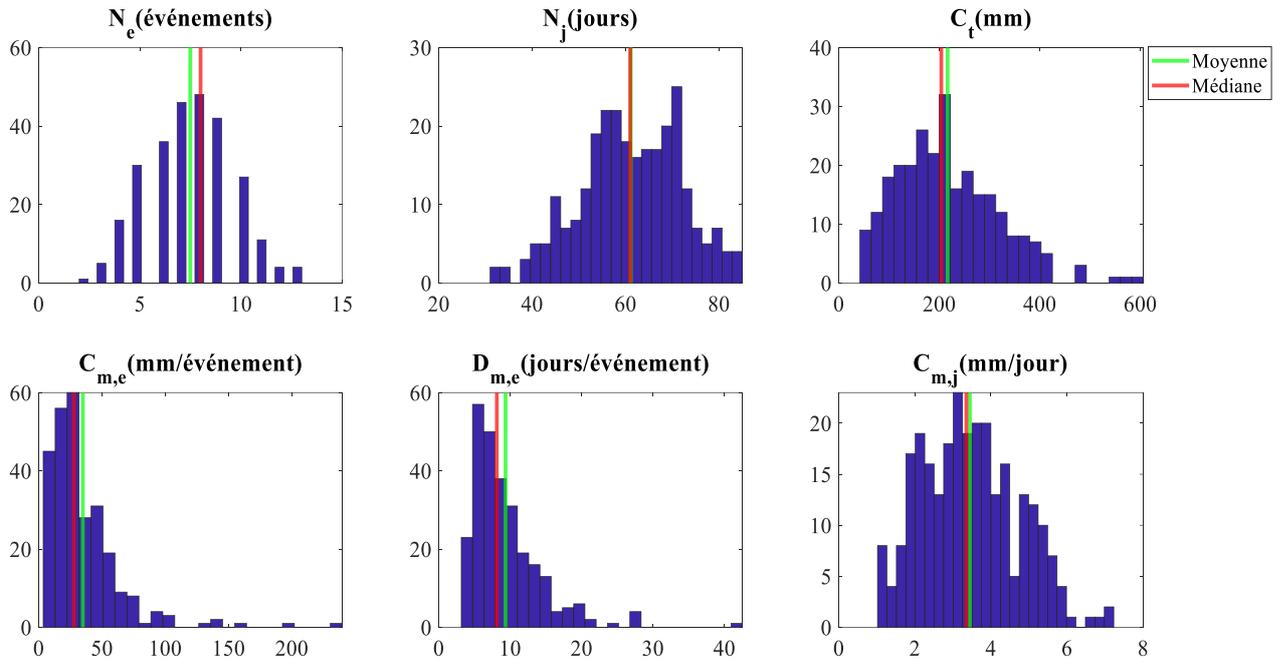


Figure III-9 Histogrammes de distribution des descripteurs de pluie (DJF) des données de modèle de réanalyse ERA-Interim entre 1980 et 2009. Nombre d'individus : 270 (9 mailles x 30 ans)

Tableau III-2 Comparaison entre les descripteurs de pluie des observations (70 x 30 = 2100 saisons DJF) et des données ERA Interim (9 x 30 = 270 saisons DJF) entre 1980 et 2009. (vert : approximativement symétrique, jaune : distribution modérément dissymétrique, rose : distribution fortement dissymétrique)

	Observations					ERA Interim				
	Min	Max	Médiane	Moyenne	Coefficient d'asymétrie	Min	Max	Médiane	Moyenne	Coefficient d'asymétrie
N_e (événements)	1	16	9	8.9	-0.2	2	13	8	7.5	0.08
N_j (jours)	1	78	27	26.8	0.67	31	85	61	61.1	-0.2
C_t (mm)	3	1419	173.3	215.2	2.1	42	605.4	204	216.5	0.7
$C_{m,e}$ (mm/événement)	2.2	204.8	19.3	24.3	2.8	3.6	239.8	27.8	34.9	3
$D_{m,e}$ (jours/événement)	1	19.5	2.8	3	2.1	3.1	42.5	8.1	9.6	2.3
$C_{m,j}$ (mm/jour)	1.2	31.8	7	7.8	1.5	1	7.2	3.4	3.4	0.3

En comparant les données modèles aux observations, on trouve une discordance sur l'interprétation du descripteur cumulé moyen par jour pluvieux. En effet, il est considéré fortement dissymétrique dans les observations et symétrique dans le modèle. Une autre discordance mais moindre est trouvée pour le cumul total qui est considéré fortement

dissymétrique dans les données d'observation mais modérément dissymétrique dans le modèle. Une troisième discordance est trouvée pour le descripteur nombre de jours qui est considéré modérément dissymétrique dans les observations mais symétrique dans le modèle. Au total, on trouve une discordance pour trois descripteurs sur six ce qui justifie amplement cette étude.

III.3.2.2 Variabilité spatiale des descripteurs du modèle ERA-Interim

La variabilité spatiale des descripteurs de pluie dans les 9 mailles de modèle de réanalyse ERA-Interim est illustrée dans les figures III-10 et III-11. Chaque boîte à moustache représente la distribution d'un descripteur donné sur 30 ans. N_e se comportent inversement aux autres descripteurs (Fig.III-10a) c-à-dire les saisons les plus humides sont composé d'un faible nombre d'événements trop longs et inversement. Les cinq autres descripteurs (N_j , C_t , $C_{m,e}$, $D_{m,e}$ et $C_{m,j}$) partagent presque la même allure. Selon la position latitudinale des mailles, la maille M1 se caractérise par les conditions les plus humides, c-à-dire des événements longs et plus intenses que les autres mailles ($C_{m,e}$ et $D_{m,e}$), un cumul saisonnier (C_t) plus élevé que les autres, des jours pluvieux très fréquents et de faible intensité (N_j et $C_{m,j}$). Cette maille M1 couvre la partie la plus septentrionale de la zone d'étude et s'étend à la Méditerranée. Les mailles M2, M3, M4 et M5 sont moins pluvieuses que M1 et elles présentent aussi une variabilité selon la position longitudinale de la maille. En effet, les courbe de la médiane (ligne verte), pour tous les descripteurs à l'exception de N_e , diminuent dans la maille M3 puis remontent considérablement vers la maille M4 et M5. Cette maille M5 couvre en grande partie la mer (> 75 %). La maille M4 couvrent le golfe de Tunis et le golfe de Hammamet avec le continent. Entre les mailles M2 et M3, qui couvrent le continent, un gradient de quantité et de fréquence de précipitations est décrit par le modèle. En allant vers le sud, les mailles M6, M7, M8 et M9 deviennent de plus en plus sèches et les pluies sont devenues moins abondantes. Selon les longitudes des mailles, les cumuls totaux C_t (Fig.III-10c) et par événement $C_{m,e}$ (Fig.III-11a) et par jour $C_{m,j}$ (Fig.III-11c) tendent à diminuer de l'ouest vers l'est. Suivant la position latitudinale des mailles, le modèle de réanalyse ERA-Interim a montré une différence dans le régime pluviométrique entre le nord et le sud. Ce gradient est aussi montré avec les descripteurs observés. Sur la même latitude, les mailles couvrant une partie la mer avec le continent ont un comportement relativement différent.

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

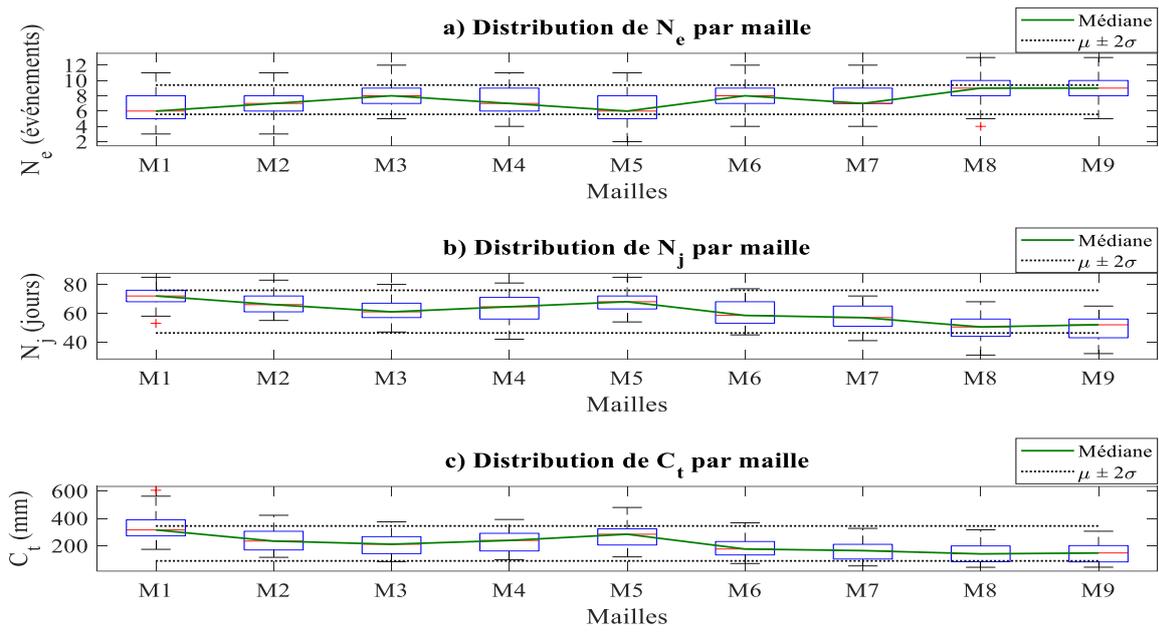


Figure III-10 Variabilité spatiale de distribution des descripteurs (a) N_e , (b) N_j et (c) C_t sur 30 ans (1980-2009).

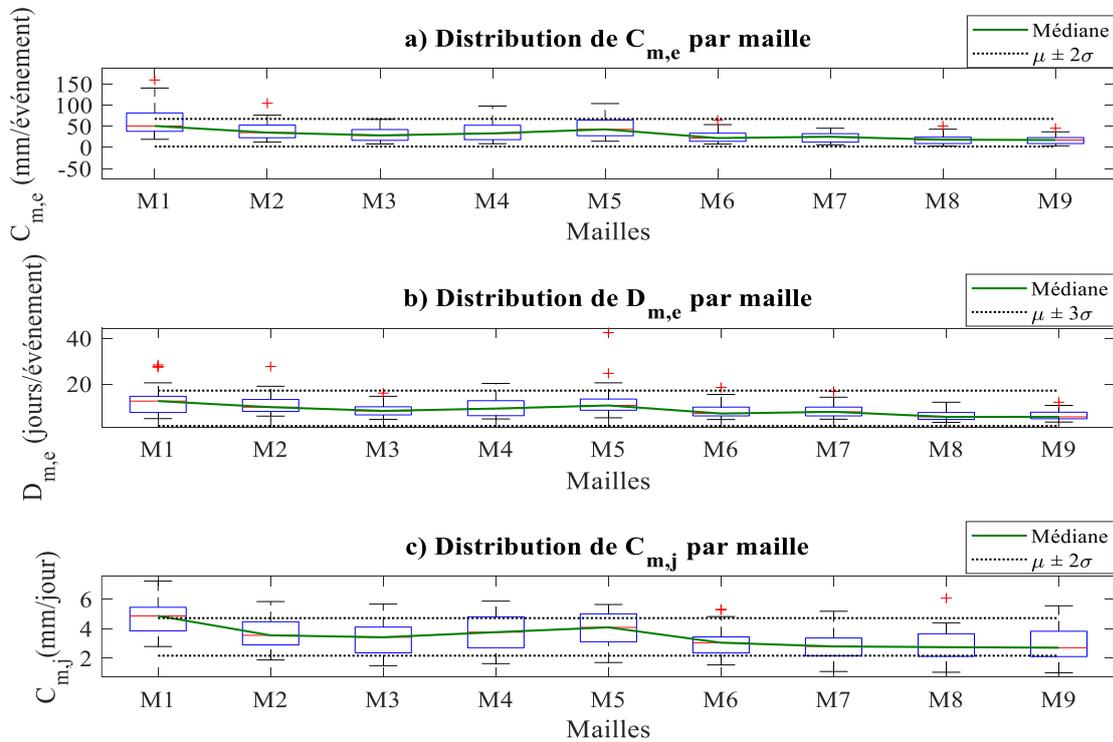


Figure III-11 Variabilité spatiale de distribution des descripteurs (a) $C_{m,e}$, (b) $D_{m,e}$ et (c) $C_{m,j}$ sur 30 ans (1980 et 2009).

En général, les mailles couvrant le nord de la zone d'étude ont des conditions plus humides que celles du sud ce qui est cohérent avec les données des stations. Par contre, le gradient de la quantité et de la fréquence de la pluie entre la région nord-ouest (correspond à la maille M2) et le reste de la zone d'étude n'est pas détecté par le modèle de réanalyse.

III.3.2.3 Variabilité temporelle des descripteurs ERA- Interim

Les graphes dans les figures III-12 et III-13 représentent la variabilité temporelle de distribution de chaque descripteur dans les 9 mailles de ERA-Interim entre 1980 et 2009. La médiane (ligne verte) du descripteur N_e varie entre 4 et 10 événements et il dépasse le seuil de contrôle en 1983, 1988 et 2000. Dans le modèle, les conditions les plus humides sont enregistrées pendant les années 1996, 2003 et 2005. Ces conditions se manifestent par des événements intenses $C_{m,e}$ et longs, un cumul saisonnier C_t élevé et un nombre de jours élevé. Les conditions les plus sèches apparaissent dans les saisons DJF de 1980, 1988 et 2002. Bien que le modèle ERA-Interim surestime certains descripteurs (comme le N_j , $C_{m,e}$, $D_{m,e}$) et sous-estime d'autres, la structure de variabilité temporelle simulée par ce modèle est cohérente avec celle des observations, spécialement pour le cumul total C_t (voir les figures III-4 et III-5).

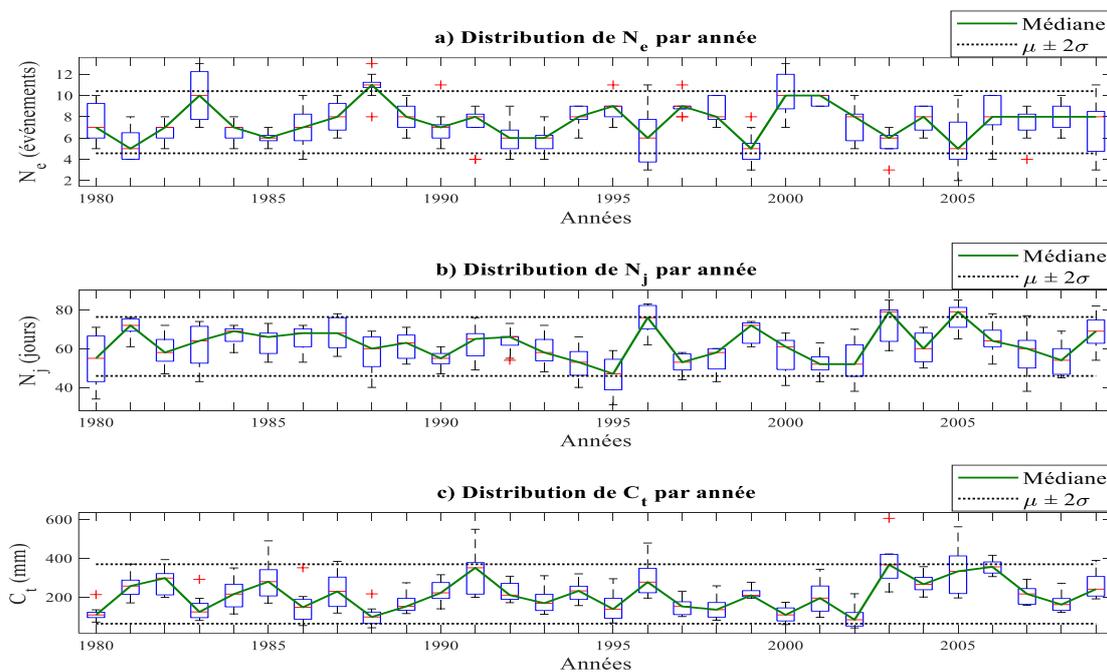


Figure III-12 Variabilité temporelle de distribution des descripteurs (a) N_e , (b) N_j et (c) C_t dans les 9 mailles de ERA-Interim entre 1980 et 2009

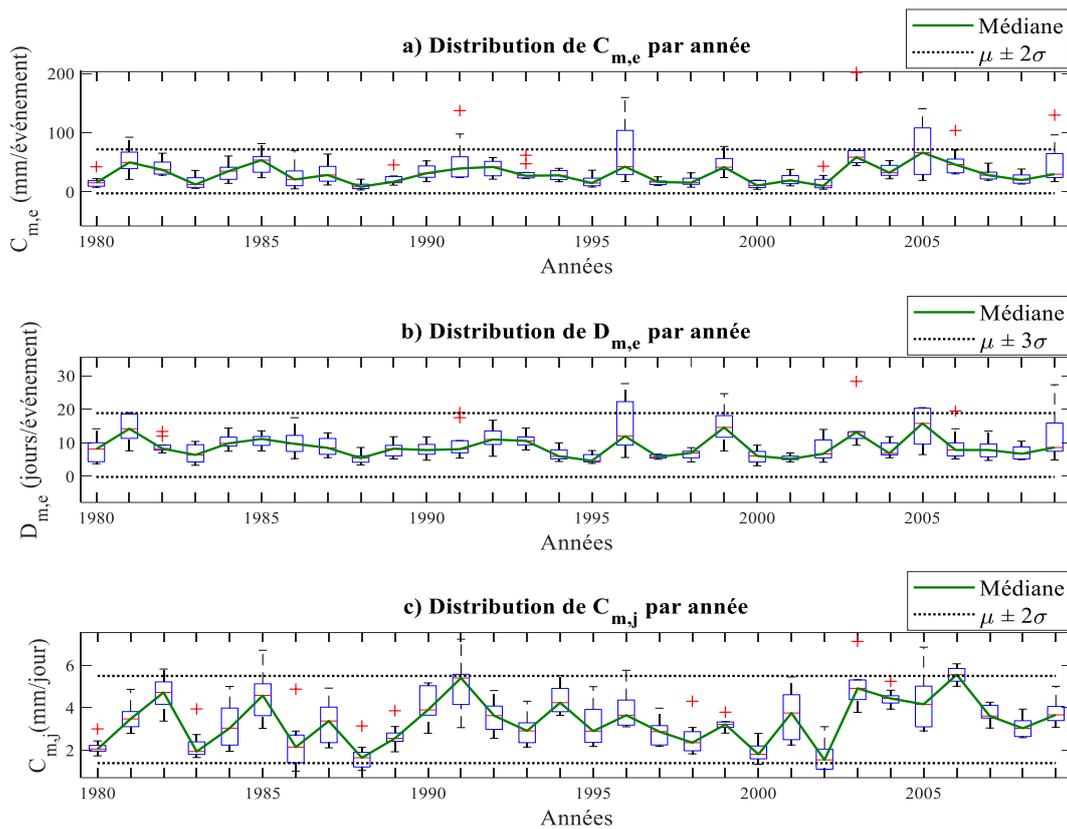


Figure III-13 Variabilité temporelle de distribution des descripteurs (a) $C_{m,e}$, (b) $D_{m,e}$ et (c) $C_{m,j}$ dans les 9 mailles de ERA-Interim entre 1980 et 2009.

III.4 Méthodologie d'analyse multivariée des descripteurs d'événements pluvieux

La définition des descripteurs caractéristiques d'une saison et les analyser par une méthode d'analyse multivariée, nous a semblé une approche pertinente pour résoudre la complexité de comparer entre elles des séries temporelles de précipitations correspondant à des résolutions spatiales différentes ou observées dans des lieux différents (Derouiche et al., 2019 ; Derouiche et al., 2022). La méthodologie développée correspond donc à une approche par caractéristique où l'objectif est de réaliser une typologie des saisons observées au pas de temps journalier. L'étude proposée repose sur deux axes : le premier exploite des approches linéaires et le second des approches non linéaires.

III.4.1 Approche factorielle linéaire : Analyse en composantes principales

Une première étude du jeu de données obtenu en dimension 6 repose sur l'analyse en composante principale. L'ACP va nous permettre de réduire les dimensions du jeu de données et d'étudier la variabilité spatiale et temporelle des composantes principales après les avoir interprétées. Après avoir concentré l'information initiale (série chronologique journalière observée pendant une saison) en 6 descripteurs puis en nombre inférieur des composantes

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

principales, nous étudierons les relations entre ces composantes et les indices de climat. La méthodologie proposée est résumée par le diagramme dans la figure III-14.

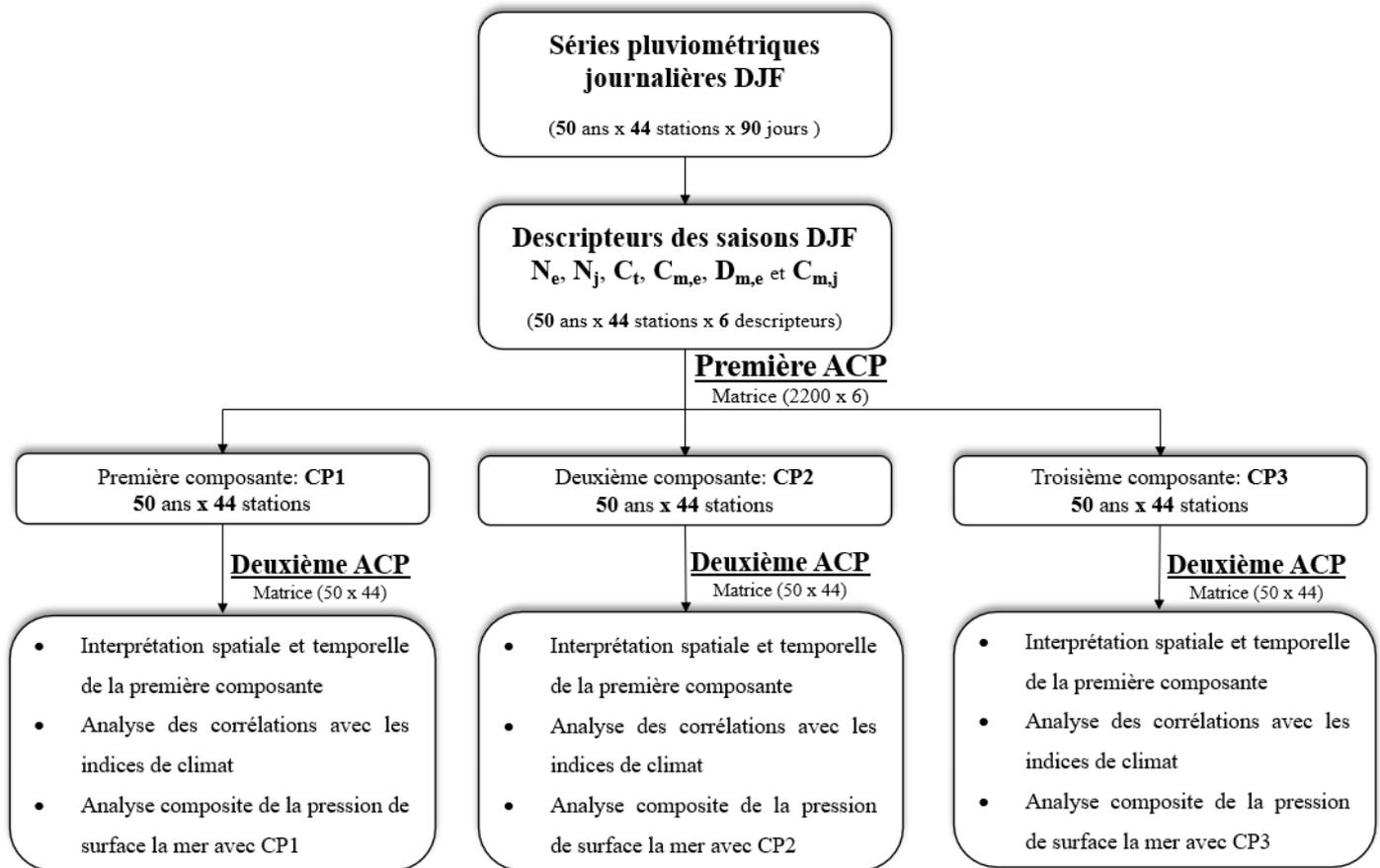


Figure III-14 Schéma de la méthodologie d'analyse multivariée par l'analyse en composantes principales

Cette première approche qui exploite les dépendances linéaires entre les descripteurs présente l'avantage de la simplicité mais a l'inconvénient d'exploiter une quantité d'information relativement faible. Ainsi, l'existence des données manquantes dans la base de données entraîne l'élimination d'un nombre de stations de cette analyse.

III.4.2 Approche non linéaire SOM et CAH

Dans un second temps, une classification non supervisée de l'échantillon de caractéristiques d'événements de pluie DJF observés sur le réseau de stations est réalisée avec une procédure qui couple SOM et CHA (Fig. III-15). La combinaison de SOM et CHA tire parti des deux types d'algorithmes. SOM est adopté pour assurer une quantification vectorielle et une inspection visuelle des caractéristiques des précipitations dans un premier temps. Cette étape réduit le nombre de saisons DJF en regroupant les saisons de propriétés d'événements pluviométriques similaires par une seule saison prototype plus significative. Le nombre

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

relativement important de prototypes fournis par la carte SOM permet de prendre en compte la complexité du jeu de données. Puis, la classification hiérarchique ascendante CHA combinée avec la méthode data image sont utilisées pour définir le nombre de classes de saisons DJF partageant des caractéristiques similaires. La méthodologie est résumée par le diagramme de la figure III-15. Les résultats de cette classification sont exploités pour évaluer la qualité et la fiabilité des données de réanalyse et leur capacité à détecter la variabilité spatio-temporelle des descripteurs pluviométriques, dans la zone d'étude.

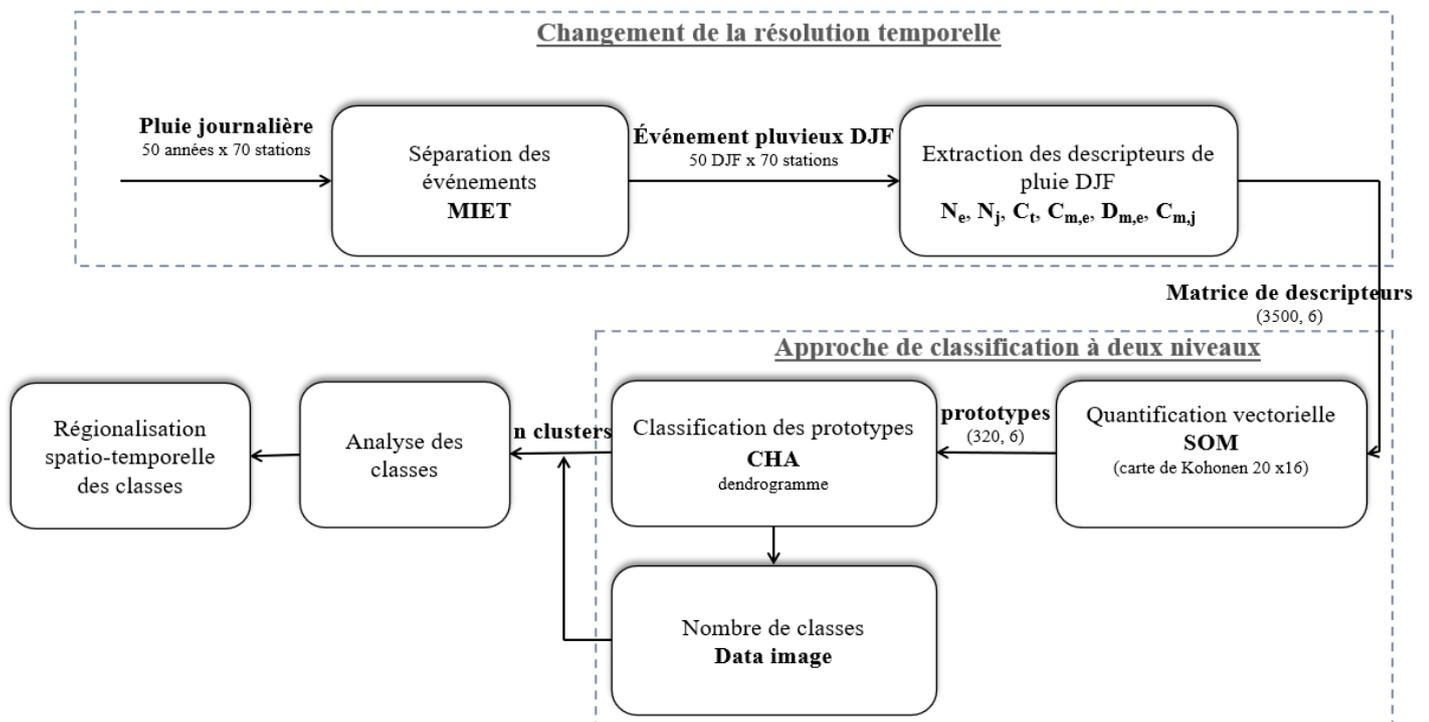


Figure III-15 Schéma de la méthodologie d'analyse multivariée des saisons DJF par l'approche non linéaire

III.5 Conclusion

Dans ce chapitre, les séries journalières ont été réduites en considérant seulement les épisodes pluvieux dans les séries journalières et en éliminant les épisodes secs. L'extraction des épisodes humides a été basé sur la détermination d'une durée minimale de non-pluie (MIET) jugée suffisante pour considérer que les deux événements séparés sont indépendants. Les données journalières des stations ont donné un MIET nuancée entre deux et trois jours. Après avoir tenu compte des caractéristiques physiques du cycle de vie des cyclones dans la région méditerranéenne et la spécificité de cette région, la durée minimale la plus petite (2 jours) est adoptée. La séparation des événements en se basant sur ce critère a permis une première réduction de la base de données journalière DJF en des événements de différents cumuls, durée et intensité. Certaines caractéristiques des événements pluvieux sont totalisées et d'autres sont

METHODOLOGIE D'ANALYSE DE PRECIPITATIONS PAR EVENEMENT PLUVIEUX

moyennées sur la saison DJF et construisent des descripteurs des événements pluvieux par saison. Le même critère et la même démarche sont appliqués aux données journalières du modèle de réanalyse.

L'étude des changements des descripteurs in situ par sous périodes (1990-1990 ; 1991-2009) a permis de mettre en évidence une diminution des statistiques du nombre d'événements pluvieux pendant les années les plus récentes (1991-2009) contre une augmentation considérable de 25e, 50e et 75e percentiles du cumul moyen par jour pluvieux sur la même période.

L'analyse statistique univariée comparative entre les descripteurs du modèle de réanalyse et les descripteurs des observations a montré que le volume saisonnier DJF total des eaux apportées par la pluie à une échelle régionale est presque similaire aux valeurs d'observations. Néanmoins, ce volume n'est pas toujours réparti également dans le temps et dans l'espace. À une échelle temporelle plus fine (l'échelle d'un événement pluvieux et l'échelle d'un jour pluvieux), le modèle a donné un volume moyen de précipitations biaisé. L'occurrence des jours pluvieux dans le modèle est la caractéristique la plus surestimée. Par conséquence, la durée des événements pluvieux a été trouvée longue par rapport aux durées observées.

À l'exception du nombre d'évènements, la variabilité temporelle (1980-2009) des descripteurs a montré une bonne cohérence avec les observations. En effet, les années avec des conditions exceptionnellement sèches (faible cumul, faible intensité d'évènements) comme celle des années 2000 et 2002 sont bien simulées par le modèle ERA-Interim. Les années avec des conditions extrêmement humides sur le terrain (par exemple 2003), sont également bien représentées par le modèle. Par contre, l'évolution spatiale des descripteurs dans les deux sources de données montre une incohérence surtout dans la région nord de la zone d'étude. Par exemple, la distinction de la zone nord-ouest dans le cumul et la fréquence des pluies d'observation n'a pas été simulée par le modèle.

L'analyse univariée des descripteurs a permis d'analyser la variabilité spatio-temporelle de chaque descripteur séparément et d'extraire quelques différences existantes entre les observations et la simulation. Cependant, l'analyse des bases des données multidimensionnelles nécessite un traitement par des approches d'analyse multivariée pour mettre en évidence les liens entre les différents descripteurs d'événements pluvieux pendant la saison DJF (durée, amplitude, fréquence, intensité ...). Pour ce faire, deux stratégies d'analyse multivariée sont proposées. Les détails des algorithmes des méthodes de classifications adoptées et les aspects mathématiques font l'objet du chapitre suivant

Chapitre IV : Outils utilisé

IV.1	INTRODUCTION	63
IV.2	ANALYSE DES DONNEES PAR UNE METHODE FACTORIELLE : ANALYSE EN COMPOSANTES PRINCIPALES (ACP).....	63
IV.3	ANALYSE PAR CLASSIFICATION NON SUPERVISEES : APERÇU SUR LES METHODES DE CLUSTERING UTILISEES.....	63
IV.3.1	<i>Classification non paramétrique et mesure de similarité</i>	63
IV.3.2	<i>Les cartes auto-organisatrices (Self-Organizing Map : SOM)</i>	64
IV.3.3	<i>La classification hiérarchique ascendante (CHA).....</i>	72
IV.3.4	<i>Data image.....</i>	75
IV.4	TÉLÉCONNEXIONS	76
IV.4.1	<i>Analyse de corrélation linéaires : corrélation de Pearson et significativité.....</i>	76
IV.4.2	<i>Analyse composite</i>	77
IV.5	CONCLUSION	77

IV.1 Introduction

Les approches envisagées dans cette étude font appel à différentes méthodes comme l'analyse en composantes principales, les carte auto-organisatrices, la classification hiérarchique et data image. Ce chapitre présente les aspects mathématiques de chaque méthode et décrit les algorithmes et les paramètres de chacune. Il s'adresse plus particulièrement aux lecteurs pour lesquels ces méthodes d'analyse ne sont pas familières.

IV.2 Analyse des données par une méthode factorielle : Analyse en composantes principales (ACP)

L'analyse en composantes principale est une méthode de synthèse et de réduction multivariée linéaires très utilisées et bien développée par plusieurs mathématiciens et chercheurs citons, par exemple, Lebart et al. (2000) et Escofier et Pagès (2008). Cette méthode fait appel à des notions mathématiques d'algèbre linéaire qui sont présentées dans l'annexe B.

IV.3 Analyse par classification non supervisées : aperçu sur les méthodes de clustering utilisées

IV.3.1 Classification non paramétrique et mesure de similarité

Dans l'apprentissage non supervisé, on dispose uniquement un ensemble des données à partir desquelles on opte à extraire et comprendre des structures et des connaissances naturellement présentes. L'objectif est de trouver ces structures sans aucune information au préalable sur l'ensemble des données. La classification non supervisée s'applique souvent sur des tableaux des données formées par un ensemble des observations, ou individus, décrites par une ou plusieurs caractéristiques qui sont les variables. Les classifications automatiques non paramétriques (adoptées dans cette thèse) sont basées généralement sur la mesure de ressemblance/dissemlance entre les individus. Cette ressemblance est déterminée en mesurant la proximité entre les individus à partir du calcul des distances (Hajjar, 2014).

Plusieurs distances sont proposées pour les classifications, mais la distance euclidienne et le carré de la distance euclidienne reste la plus populaire et la plus utilisée (Xu and Wunsch, 2005 ; Saxena et al. 2017).

- La distance euclidienne entre deux vecteurs x_i et x_j de dimension p est égale à (Eq. IV-1) :

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad IV-1$$

- La distance de Manhattan entre deux vecteurs x_i et x_j de dimension p est égale à (Eq. IV-2) :

$$d(x_i, x_j) = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad IV-2$$

- La distance de Chebyshev entre deux vecteurs x_i et x_j de dimension p est égale à (Eq. IV-3) :

$$d(x_i, x_j) = \max_{k=1}^p |x_{ik} - x_{jk}| \quad IV-3$$

- La distance de Minkowski entre deux vecteurs x_i et x_j de dimension p est égale à (Eq. IV-4) qui est une généralisation des trois distances précédentes (avec q est un entier positif):

$$d(x_i, x_j) = \left(\sum_{k=1}^p |x_{ik} - x_{jk}|^q \right)^{\frac{1}{q}} \quad IV-4$$

Les distances utilisées doivent vérifier certains critères :

- Symétrie : $d(x_i, x_j) = d(x_j, x_i)$
- Positivité : $d(x_i, x_j) > 0$
- Inégalité triangulaire : $d(x_i, x_j) < d(x_i, x_m) + d(x_m, x_j)$
- $d(x_i, x_i) = 0$

Le choix de la distance influence la forme des classes obtenues. Par exemple, la distance euclidienne donne des classes hypersphérique et la distance de Chebyshev forment des classes hypercubique (Hajjar, 2014).

IV.3.2 Les cartes auto-organisatrices (Self-Organizing Map : SOM)

La SOM est une méthode neuronale de quantification vectorielle. Chaque zone de l'espace multidimensionnel initial discrétisé est représentée par un neurone référent (prototype). L'algorithme proposé par Kohonen, permet ainsi de projeter d'une façon non linéaire les observations multidimensionnelles par des neurones référents (Kalteh et al., 2008). Ces neurones référents sont dans un espace de faible dimension (couramment 2D), comparativement à l'espace initial. Ils constituent la carte topologique ou bien carte de Kohonen. Cette carte est un estimateur de la densité des observations de l'espace initial (les zones denses sont représentées plus finement sur la carte). Des neurones voisins sur la carte de Kohonen représentent des zones proches dans l'espace initial (Peeters et al., 2006). Cette représentation 2D de données multidimensionnelles permet une analyse visuelle et facilite la compréhension de la structure de variabilité dans l'espace d'observation. Cette visualisation permet notamment d'étudier les dépendances entre les variables (Kohonen, 2001). La figure VI-1 illustre l'architecture de la carte auto-organisatrice. Cette figure montre une connexion totale entre les observations de l'espace d'entrée et les neurones de la carte par des vecteurs référents m .

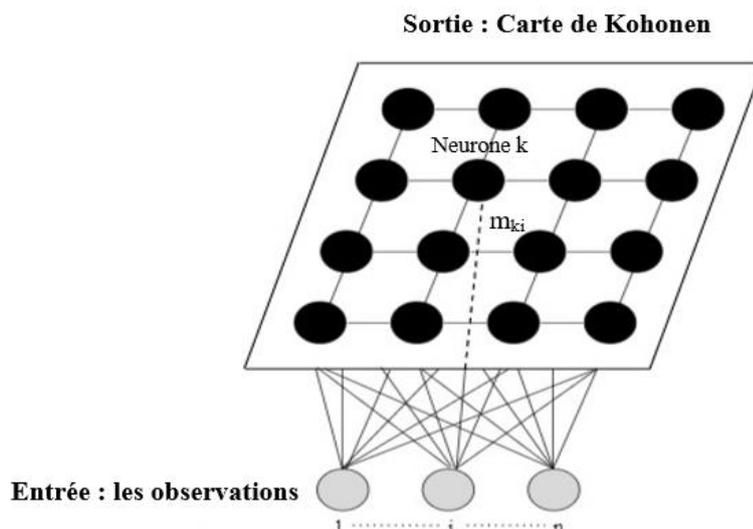


Figure IV-1 Architecture de la carte auto-organisatrice

IV.3.2.1 Les étapes d'apprentissage

L'algorithme de la SOM fait projeter les données de l'espace d'entrée X de dimension p vers un espace de sortie sous forme d'une carte 2D. Cette carte est constituée d'un ensemble de neurones organisés d'une manière significative. Chaque neurone est associé à un vecteur poids (prototype) $m_i \in \mathbb{R}^p$. Dans une première étape, les vecteurs référents m_i sont initialisés par des vecteurs ayant la même dimension aux vecteurs de l'espace initiale (Kohonen, 2001). Dans chaque itération t de l'apprentissage de la carte, il existe deux phases. La première est une phase d'affectation qui consiste à affecter un vecteur de l'espace initiale x_j (pris au hasard) à son vecteur gagnant c connue par BMU (Best Matching Unit) en suivant l'équation IV-5 (Vesanto, 2000) :

$$c = \arg \min_i \{d(x_j - m_i)\} \quad IV-5$$

Comme indique (Eq. IV-5) la phase d'affectation consiste à associer chaque observation au référent le plus proche dans l'espace de représentation initial (BMU) en se basant sur une distance euclidienne d comme indice de similarité. La deuxième phase de l'apprentissage est la phase de minimisation dans laquelle on détermine les prototypes. Dans la version d'optimisation global de la SOM (version batch), adoptée dans cette étude, le calcul des nouvelles coordonnées des référents m_i suit l'équation IV-6. Cette équation est la moyenne pondérée des observations x pondérées par la valeur de la fonction de voisinage $h_{i,c}(t)$ avec son BMU noté c (Vesanto, 2000) :

$$m_i(t + 1) = \frac{\int_{j=1}^n h_{i,c(j)}(t)x_j}{\int_{j=1}^n h_{i,c(j)}(t)} \quad IV-6$$

$h_{i,c}(t)$ est la fonction de voisinage qui indique la proximité entre le neurone gagnant c et un neurone i .

Dans la version séquentielle de la SOM (la plus connue), la phase de minimisation suit l'équation IV-7 :

$$m_i(t + 1) = m_i(t) + \alpha(t) h_{i,c}(t)(x - m_i(t)) \quad IV-7$$

c est le neurone gagnant du vecteur de l'espace d'entrée $x(t)$ à l'itération t , $h_{i,c}(t)$ est la fonction de voisinage. $\alpha(t)$ est le pas d'apprentissage. Ce paramètre α est un régulateur de vitesse d'apprentissage. C'est une fonction décroissante en fonction du temps d'apprentissage. Elle est utilisée pour ralentir le processus d'apprentissage à la fin. La figure IV-2 illustre le principe de projection des données sur la carte.

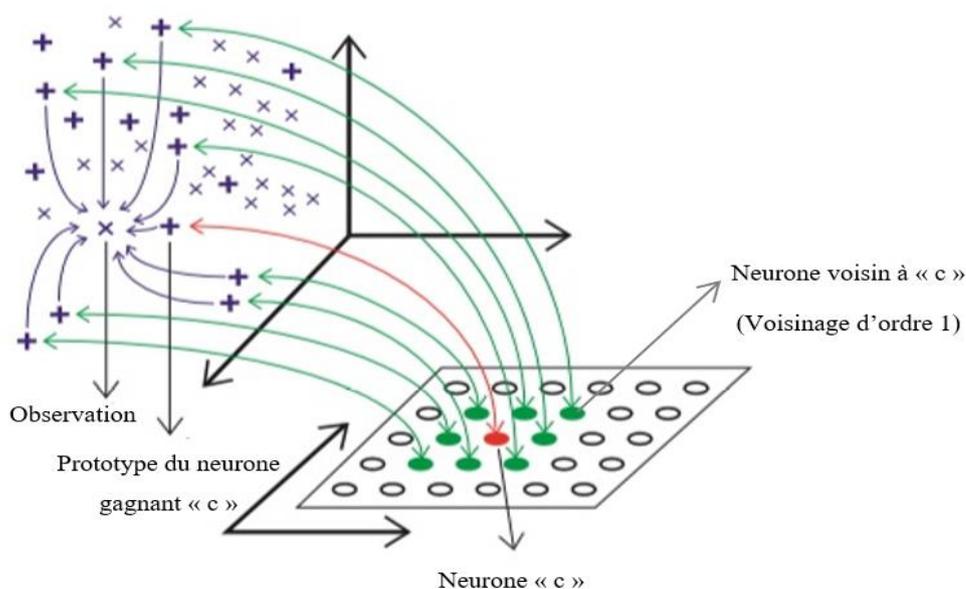


Figure IV-2 Illustration du principe de projection des observations sur la carte de kohonen (Hajjar, 2014, modifiée)

IV.3.2.2 La fonction de voisinage

L'introduction d'une fonction de voisinage dans le processus d'apprentissage de la carte garantit la préservation de la topologie des données de l'espace initial. Il existe plusieurs types de fonction de voisinage (fonction gaussienne, bubble, Epanechicov, etc). La fonction gaussienne reste la plus commune et la plus flexible et elle assure une convergence plus rapide de l'algorithme SOM (Haykin, 2009). Cette fonction de voisinage, donnée par l'équation IV-8, est centrée sur le neurone gagnant c et elle indique l'amplitude d'appartenance au voisinage à ce dernier :

$$h_{i,c} = e^{\left(\frac{-\|r_c - r_i\|^2}{2\sigma^2(t)}\right)} \quad IV-8$$

r_c et r_i sont respectivement la position du neurone gagnant c et le neurone i sur la carte. $\sigma(t)$ est le rayon de voisinage est une fonction décroissante en fonction du temps d'apprentissage (Eq.IV-9).

$$\sigma(t) = \sigma_{init} + \left(\frac{t}{T}\right)(\sigma_{final} - \sigma_{init}) \quad IV-9$$

σ_{init} et σ_{final} représentent respectivement le rayon de voisinage initial et final, T est le nombre d'itérations et t l'itération actuelle. La fonction de voisinage gaussienne en fonction de la distance entre le neurone gagnant et un neurone i est illustrée par la figure IV-3.

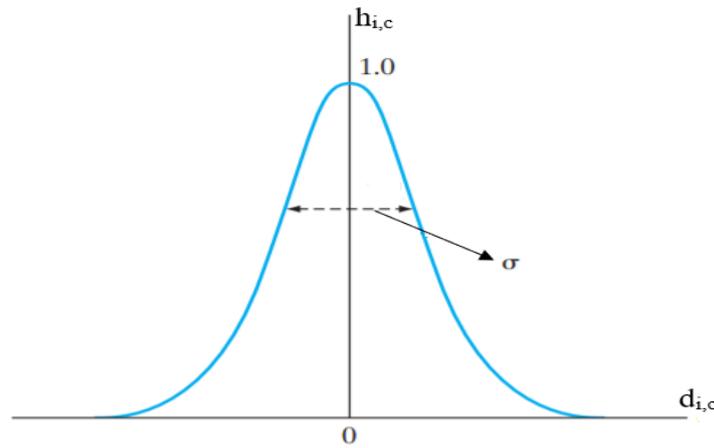


Figure IV-3 Illustration de la fonction de voisinage gaussienne

Au début de l'apprentissage, avec un rayon d'apprentissage assez large, la mise à jour des prototypes touche un grand nombre de neurones voisins au neurone gagnant c . Au cours des itérations, la mise à jour se limite au neurone gagnant et ces voisins immédiats. À la fin de l'apprentissage, le rayon de voisinage devient très petit et l'ajustement se limite aux prototypes où chacun est le centre de gravité des données qu'il représente et l'algorithme est alors identique à celui des K-moyennes.

IV.3.2.3 Évaluation de la qualité de la carte

La qualité de la carte est évaluée par deux statistiques majeures. La première est la capacité de quantification qui se traduit par le calcul de la moyenne des erreurs de quantification connue par (Mean quantization error) et donnée par l'équation IV-10 :

$$Qe = \frac{\sum_{i=1}^n \|x_i - m_{ci}\|}{n} \quad IV-10$$

x_i est l'observation i , m_{ci} est le vecteur gagnant de l'observation i , n le nombre d'observations

Le deuxième critère permet la quantification de préservation de la topologie par le calcul d'une erreur topographique. Le principe est de comparer les positions des prototypes dans l'espace des données par rapport aux positions de leurs neurones correspondants sur la carte. L'erreur topographique T_e (Eq. IV-11) calcul la proportion d'observations dont leur premier neurone gagnant c et deuxième neurone gagnant c' ne sont pas voisins sur la carte. Un deuxième neurone gagnant d'une observation x_i est celui lié au prototype le plus proche après celui de neurone gagnant c . Si cette erreur est égale à 0 alors la préservation de la topologie est considérée parfaite (Breard, ,2017).

$$T_e = \frac{\sum_{i=1}^n E}{n} \quad IV-11$$

$$E = \begin{cases} 1 & \text{si } \|r_c - r_{c'}\| \neq 1 \quad (\text{les neurones ne sont pas voisins}) \\ 0 & \text{si } \|r_c - r_{c'}\| = 1 \quad (\text{les neurones sont voisins}) \end{cases}$$

r_c et $r_{c'}$ sont respectivement les positions du premier et du deuxième neurone gagnant d'une observation x_i .

IV.3.2.4 Les paramètres d'apprentissage

L'algorithme de la SOM exige un nombre de paramètres à fixer à priori tels que la dimension de la carte, le rayon de voisinage, la structure des neurones, etc. cette phase est très délicate, car un mauvais choix peut donner à des résultats incohérents.

- Le nombre de neurones :

Le nombre de neurones peut être choisi en se basant sur une formule approximative $M \approx 5 \times \sqrt{n}$ (n est le nombre d'observations et M le nombre de neurones) dans le but de donner une carte 2D (Vesanto et al., 2000). Ce nombre n'est pas final, il peut être modifié après plusieurs lancements de l'algorithme d'apprentissage pour assurer une meilleure qualité des résultats.

- Initialisation des prototypes :

Il existe deux manières pour initialiser les prototypes soit aléatoirement ou bien linéairement. Pour l'initialisation aléatoire des prototypes, le résultat varie d'une exécution à une autre. Dans ce cas, on peut lancer le programme plusieurs fois et garder la version de résultats avec la plus faible erreur de quantification et topographique. L'initialisation linéaire est plus stable, il s'agit d'appliquer une analyse en composante principale sur la matrice des données initiale et positionner les prototypes sur les deux premiers axes principaux. L'initialisation linéaire évite le risque de la torsion de la carte surtout dans la version batch de l'algorithme d'apprentissage (Vesanto et al., 2000).

- Choix des rayons de voisinage initial et final :

Dans les cartes de grande dimension, l'apprentissage se fait en deux étapes. La première est une phase d'organisation dans laquelle le rayon de voisinage initial et final sont assez grands. Cette phase vise à organiser les neurones et les orienter. La deuxième phase est une phase de convergence. Elle permet de mettre les prototypes à leurs positions finales en se basant sur l'ordre pris à la phase d'organisation. Dans cette phase, le rayon de voisinage doit être petit. Pour avoir une meilleure préservation de la topologie, ce rayon diminue pour atteindre 1. Mais si la SOM est utilisée dans le but de classification, on peut choisir un rayon de voisinage inférieur à 1 pour assurer un bon déploiement de la carte par rapport aux observations (Vesanto et al., 2000).

- **La forme des neurones :**

Les neurones dans la carte peuvent être de forme hexagonale ou bien rectangulaire. Dans la topologie hexagonale chaque neurone possède six voisins immédiats et dans la topologie rectangulaire chaque neurone possède 4 voisins (Vesanto et al., 2000). Les deux architectures possibles de la carte sont représentées dans la figure IV-4.

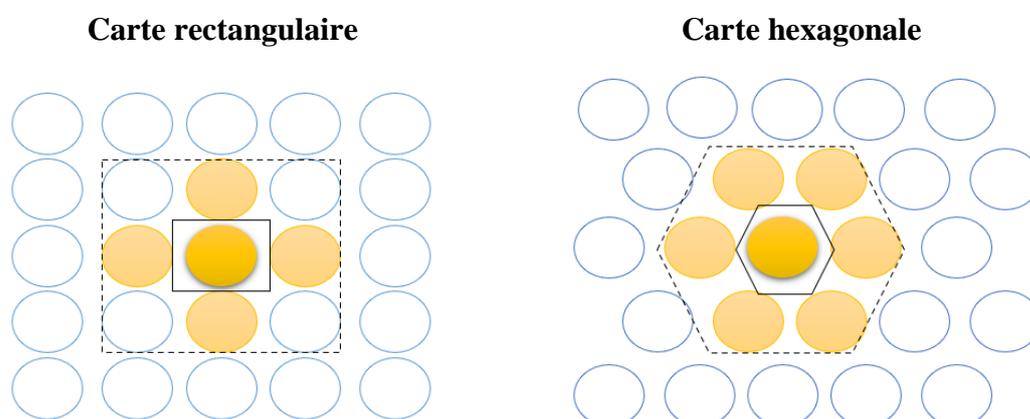


Figure IV-4 Type d'architecture de la carte

IV.3.2.5 Projection des variables et individus supplémentaires sur la carte topologique

Une fois l'entraînement de la carte est terminé et les neurones sont bien organisés et structurés sur la carte, il est possible de croiser la partition obtenue avec d'autres individus supplémentaires qui n'ont pas participé à la formation de la carte et qui dispose la même dimension des individus initiaux. Il s'agit d'affecter chaque individu supplémentaire au prototype, résultant de la SOM, le plus proche en sens de distance (distance euclidienne). Les individus supplémentaires doivent subir le même prétraitement que les individus initiaux (centrage, réduction ou normalisation).

Il est possible également d'interpréter les classes de la SOM selon des variables supplémentaires (illustratives) non utilisées dans l'apprentissage de la carte de Kohonen. Il

s'agit dans ce cas d'affecter les données des variables supplémentaires aux individus classées par chaque neurone. La distribution de ces variables peut être visualisée par une projection sur la carte de Kohonen (2-D) identiquement aux variables initiales et cela permet d'établir les liens entre les anciennes et les nouvelles variables.

IV.3.2.6 Récapitulation des étapes de la SOM

Dans cette étude, l'algorithme d'optimisation globale (version batch) est adopté. L'outil de calcul utilisé est la SOM toolbox Matlab développé par Vesanto et al. (2010). Cette version de l'algorithme de Kohonen est couramment utilisée parce qu'elle utilise beaucoup moins de mémoire lors de l'exécution (Vesanto et al, 2000) et donne une meilleure qualité de la carte avec des erreurs de quantification et des erreurs topologiques inférieures à celles données par l'apprentissage séquentiel de la SOM (Hua et al., 2009). L'espace d'entrée utilisé pour la carte auto-organisatrice est sous forme d'une matrice de n individus (observations) et p variables. Le choix des paramètres est une phase très importante parce qu'elle influence considérablement les résultats. Par exemple, si le nombre de neurones est faible, des données hétérogènes et éloignées dans l'espace d'entrée peuvent être représentées par le même neurone. Par contre, si le nombre de vecteurs référents est trop grand ou bien proche du nombre des données initiales, on risque d'avoir un sur-apprentissage connu par 'overfitting'. Pareil pour le rayon de voisinage, s'il est très petit (proche de 0) on risque de perdre le concept de la préservation de la topologie qui est l'atout de cette méthode. Pour éviter les problèmes d'échelle, on transforme les données par centrage et réduction en des données normalisées. Les différentes étapes de l'algorithme SOM sont décrites dans la figure IV-5. Les étapes 4 et 5 se répètent pour T itérations.

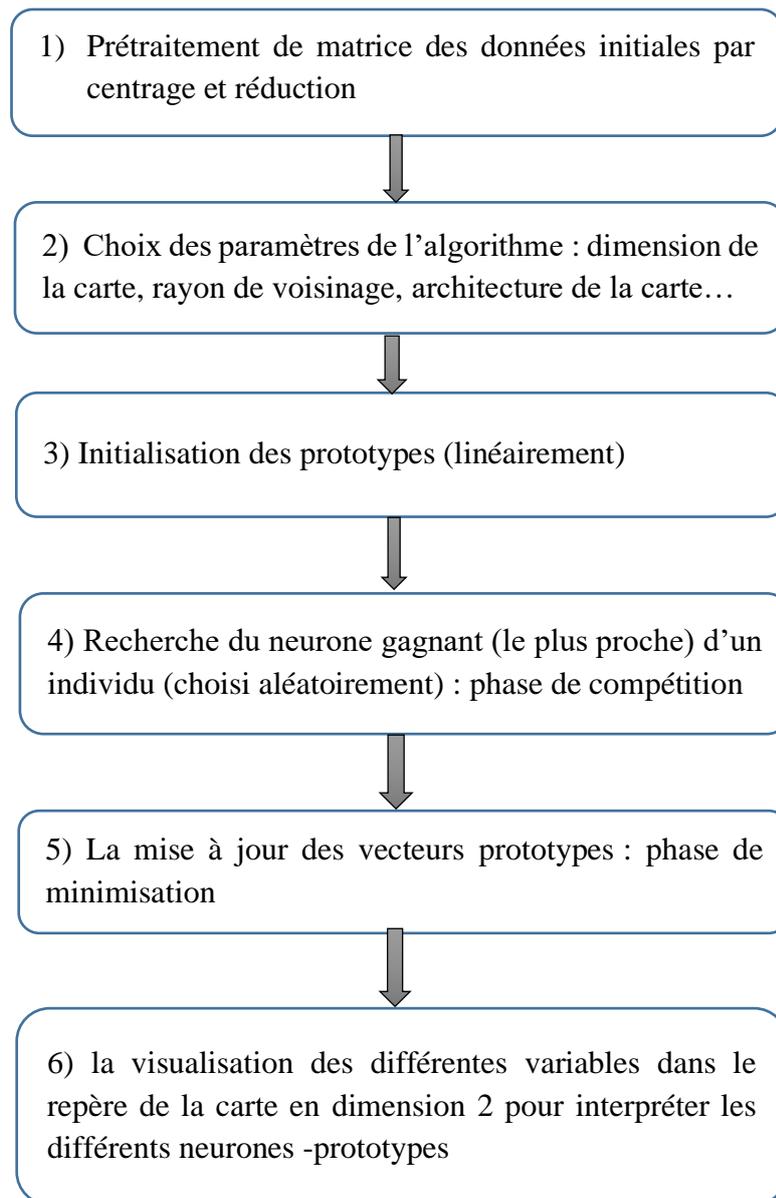


Figure IV-5 les étapes de l'algorithme de la SOM

Pour une carte de faible dimension, la structure des données et la répartition des classes peuvent être interprétées visuellement en se basant sur la U-matrice qui visualise la distance entre les neurones deux à deux (Kohonen, 2001). Cependant, la classification n'est pas toujours évidente pour une carte de grande dimension. Dans ce cas, on recourt à une deuxième classification automatique appliquée aux vecteurs prototypes. Le plus souvent, on utilise la classification hiérarchique (Chon, 2011 ; Parchure & Gedam, 2019, Mignot et al., 2020, Ioannou et al., 2021) ou bien la méthode des centres mobiles (Perez-Hoyos et al., 2014 ; Brentan et al., 2018). En adoptant cette approche, le résultat de la carte auto-organisatrice devient intermédiaire et cela

fait diminuer le risque d'un mauvais choix des paramètres sur la qualité des classes finales) (Vesanto et Alhoniemi, 2000).

IV.3.3 La classification hiérarchique ascendante (CHA)

La classification hiérarchique ascendante est une méthode très utilisée avec la SOM dans le but de classifier les prototypes de la carte auto-organisatrice et de réduire le nombre de classes (Ambroise et al., 2000). La classification hiérarchique ascendante définit les relations hiérarchiques entre les individus d'une matrice des données et les visualise par une structure arborescente appelée dendrogramme. Au début d'algorithme, chaque individu est considéré une classe. Ensuite d'autres classes sont formées d'une manière itérative en fusionnant à chaque itération les deux classes les plus similaires en se basant une mesure de similarité entre les individus et sur un critère d'agrégation fixés au préalable. À la fin de l'algorithme, on obtient une seule grande classe contenant tous les individus (Rui & Ii, 2005 ; Murtagh & Contreras, 2012).

IV.3.3.1 Critères d'agrégation

Le critère d'agrégation est la mesure de distance entre deux classes. Il existe plusieurs manières de calculer cette distance et les plus connus sont (Murtagh, 1983 ; Murtagh & Legendre, 2014) :

- **Critère du saut maximal (complete linkage)** : c'est la distance la plus grande qui sépare un élément x_i de la classe $C1$ et un élément x_j de la classe $C2$ donnée par l'équation IV-12.

$$D(C1, C2) = \max_{x_i \in C1; x_j \in C2} d(x_i, x_j) \quad IV-12$$

- **Critère de Saut minimal (single linkage)** : c'est la distance la plus petite qui sépare un élément x_i de la classe $C1$ et un élément x_j de la classe $C2$ donnée par l'équation IV-13.

$$D(C1, C2) = \min_{x_i \in C1; x_j \in C2} d(x_i, x_j) \quad IV-13$$

- **Critère de la moyenne (average linkage)** : c'est la distance moyenne ente tous les individus de la classe $C1$ et tous les individus de la classe $C2$ donnée par l'équation IV-14.

$$D(C1, C2) = \frac{1}{n_{c1}} \frac{1}{n_{c2}} \sum_{x_i \in C1} \sum_{x_j \in C2} d(x_i, x_j) \quad IV-14$$

n_{c1} et n_{c2} sont respectivement le nombre d'individus dans la classe $C1$ et $C2$

- **Critère du centre de gravité (centroid linkage)** : c'est la distance entre le centre de gravité $G1$ de la classe $C1$ et le centre de gravité $G2$ de la classe $C2$ donnée par l'équation IV-15.

$$D(C1, C2) = d(G1, G2) \quad IV-15$$

n_{c_1} et n_{c_2} sont respectivement le nombre d'individus dans la classe C1 et C2

$$G1 = \frac{1}{n_{c_1}} \sum_{x_i \in C1} \sum_i^{n_{c_1}} x_i \quad \text{et} \quad G2 = \frac{1}{n_{c_2}} \sum_j^{n_{c_2}} x_j$$

- **Critère de Ward** : c'est une distance pondérée entre les barycentres des classes C1 et C2. Elle vise à maximiser la variance inter-classe ou bien à minimiser la variance intra-classe (Ward,1963). Ce critère exige l'utilisation du carré de la distance euclidienne (Eq. IV-16).

$$D(C1, C2) = \frac{n_{c_1} n_{c_2}}{n_{c_1} + n_{c_2}} d^2(G1, G2) \quad IV-16$$

n_{c_1} et n_{c_2} sont respectivement le nombre d'individus dans la classe C1 et C2, G1 de la classe C1 et le centre de gravité G2 de la classe C2. $d^2(G1, G2)$ est le carré de distance euclidienne entre les centres des classes C1 et C2.

Le choix du critère d'agrégation peut être adapté aux données. Il faut être conscient que l'application de la classification hiérarchique ascendante peut conduire à des partition et structure des classes différentes pour le même ensemble des données en fonction du critère d'agrégation adopté. La stratégie du saut minimal a tendance à produire des classes déséquilibrées et de forme allongée, surtout pour les grandes bases de données. La stratégie du saut maximal aboutit à des classes compactes, mais ce critère est sensible aux points aberrants. Le critère de la moyenne a tendance de fusionner les classes avec des faibles variance et donne des classes plus au moins compactes et il est moins sensible aux points aberrants. Ce critère est considéré comme l'intermédiaire entre les deux premiers critères du saut maximal et minimal. La stratégie de distance entre les centres de gravité donne des classes de forme semblable à celle de la distance moyenne entre les classes. Le critère de Ward est le plus utilisé, il a tendance à donner des classes sphériques de taille similaire et il est aussi sensible aux points extrêmes.

IV.3.3.2 Récapitulation des étapes de l'algorithme de CHA

L'algorithme de la CHA se résume en quatre étapes décrites dans la figure IV-6. Les deux étapes 3 et 4 sont itératives. À la fin des itérations tous les individus initiaux se regroupe dans une seule classe.

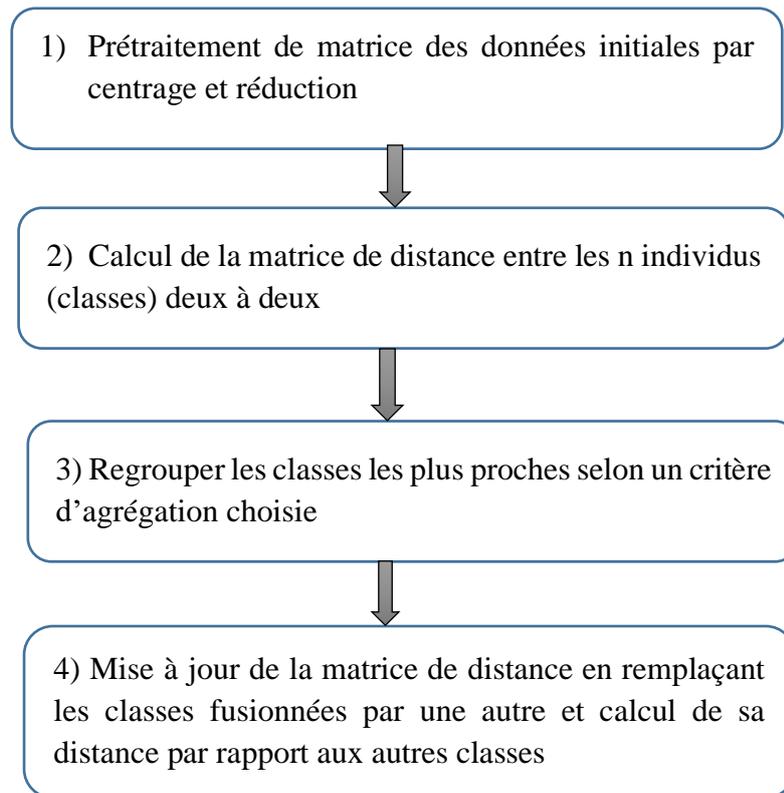


Figure IV-6 Les différentes étapes de l'algorithme de la classification hiérarchique ascendante

IV.3.3.3 Choix du nombre de classes

Le choix du nombre de classes est connu comme la partie la plus délicate dans toutes les approches de classification non supervisées. Dans la littérature, plusieurs techniques sont proposées pour trouver le nombre de classes optimal (Fischer, 2014). Parmi ces techniques, on trouve par exemple la « gap statistic » élaborée par Tibshirani et al. (2001) ou la technique de la statistique Silhouette (Kaufman & Rousseeuw, 1990) qui sont des méthodes basées sur des approches statistiques et des hypothèses. Dans cette étude, des techniques basées sur la visualisation sont adoptées pour déterminer le nombre probable de classes comme le dendrogramme résultant de la classification hiérarchique ascendante ou bien par la méthode de data image.

IV.3.3.4 Dendrogramme

Le dendrogramme est une structure arborescente qui visualise les relations hiérarchiques entre les différentes classes et les classes fusionnées en chaque étape de l'algorithme CHA. La figure IV-7 montre un exemple de dendrogramme de classifications de 20 individus par le critère de Ward. Les individus occupent l'axe horizontal et l'axe vertical présente les distances entre classes fusionnées. Chaque paire de classes fusionnées est visualisée par la forme d'un U renversé. Cet arbre sert à définir des différentes partitions des données et à détecter le nombre

de classes tout en choisissant un niveau de coupure. La figure IV-7 montre 3 niveaux de coupure différents et le nombre de classes n'est que le nombre de branches coupées à ce niveau-là. Par exemple, le niveau de coupure 1 (Bleu) divise les données en deux classes, le deuxième niveau (en rouge) donne trois classes et le troisième niveau (vert) donne six classes. Afin de définir une bonne partition des, il faut que les individus d'une même classe soient proches et les individus des classes différentes soient éloignés. En pratique, on a tendance à couper l'arbre là où les branches se trouvent assez longues, car elles reflètent l'hétérogénéité (la dissemblance) entre les différentes classes (Husson et al., 2010).

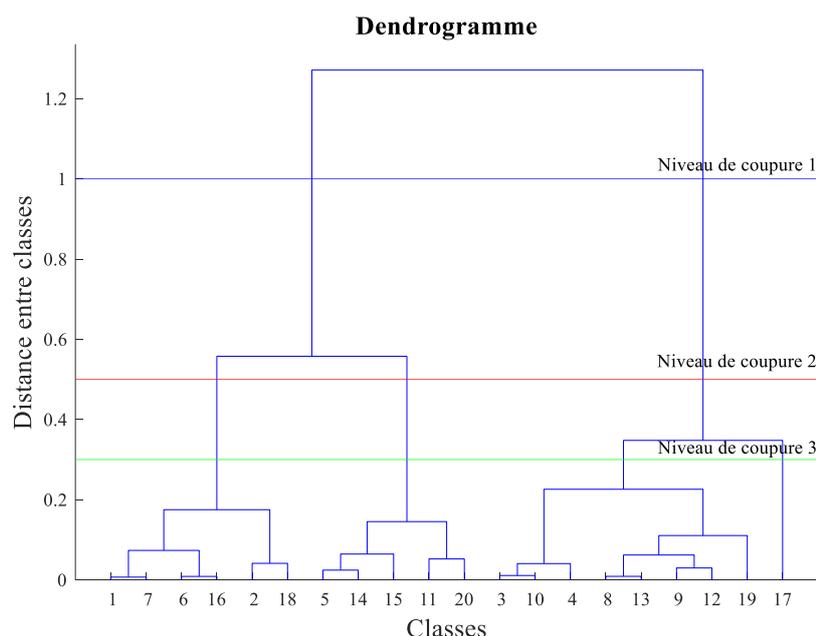


Figure IV-7 Schéma illustratif du dendrogramme des données fictives

IV.3.4 Data image

Data image est une des méthodes exploratoires multidimensionnelles basée sur la visualisation. Elle est utilisée dans cette étude dans le but de détecter le nombre probable de classes (Ling, 1973 ; Wegman, 1990 et Minnotte and West, 1999). Son principe est de transformer des données multivariées en une image bidimensionnelle d'une variable donnée, les contrastes des couleurs reflètent la magnitude de chaque observation (Martinez & Martinez, 2005). Les données peuvent être aussi visualisées en se basant sur les matrices de similarité /dissimilarité obtenues en calculant la distance entre les observations deux à deux (ex. la distance euclidienne). La matrice de dissimilarité reflète la proximité dans l'espace des données originales. Elle peut être présentée comme une image où chaque pixel représente une distance d'une paire d'observation. Ces pixels peuvent être réorganisés de plusieurs manières pour mettre en valeur la différence entre les données et pour rendre cette différence visible dans

l'image. À ce propos, Minotte et West (1999) ont montré que l'application d'une des stratégies de la classification hiérarchiques (Eq. IV-12, Eq. IV-13, Eq. IV-14, Eq. IV-15, Eq. IV-16) à la matrice de dissimilarité permet aux classes d'apparaître claires et discernable le long de la diagonale de l'image. Le nombre de classes est donné à priori et après plusieurs lancements de l'algorithme de la *data image* avec différents nombres de classes on devient capable de choisir la meilleure partition des données. Notons que cette méthode donne de meilleurs résultats avec une grande base des données et le réarrangement de la matrice des données dans l'image peut être établi au moyen d'autres approches de classification (par exemple : K-moyennes). Cette technique est utilisée dans plusieurs recherches, par exemple Hannachi et al. (2010) ont appliqué le *data image* pour l'étude du phénomène du vortex polaire et Hannachi et al. (2012) pour l'étude des courants jet.

IV.4 Téléconnexions

IV.4.1 Analyse de corrélation linéaires : corrélation de Pearson et significativité

L'analyse de corrélation est une approche mathématique largement utilisée dans les analyses climatiques (DeCastro et al., 2006 ; Aamir et Hassan, 2020 ; Yang et Xing, 2022). Le coefficient de corrélation est un indicateur statistique qui mesure l'intensité de relation linéaire entre deux (ou plusieurs) variables. Le coefficient de corrélation de Person a été calculé dans cette étude pour investiguer les relations linéaires entre les séries chronologiques des composantes de pluie et les indices climatiques. Le calcul de coefficient de corrélation, r , entre deux variables X et Y suit l'Eq.IV-17 :

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad \text{IV-17}$$

X_i et Y_i représentent les deux variables, i est l'indice des données, \bar{X} et \bar{Y} représente les moyennes des variables X et Y respectivement. La valeur de r varie entre -1 et 1. La corrélation est évaluée pour sa magnitude (fortes et faibles valeurs) et sa direction (corrélation positive ou négative). Les résultats d'analyse de corrélation doivent être interprétés avec prudence, ainsi, il est important de donner un sens aux résultats en les rattachant des mécanismes météorologique ou climatique. En effet, dans certains cas, un faible coefficient de corrélation peut cacher des relations fortes, mais de nature non linéaire. Contrairement, un coefficient de corrélation fort ne reflète pas toujours une relation de causalité entre les variables étudiées. L'analyse de corrélation aussi est très sensible à la taille de l'échantillon, c'est-à-dire, un faible coefficient de corrélation d'un échantillon de grande taille peut être plus significatif qu'un coefficient élevé d'un échantillon de petite taille. Pour cela, l'évaluation de la significativité est nécessaire dans

chaque analyse de corrélation. La significativité de corrélation est déterminée par un test de Student connue par t-test. Le t-test est donné par l'Eq.IV-18 :

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} \quad IV-18$$

r et n sont le coefficient de corrélation et le nombre d'observations des séries chronologiques respectivement. Pour les deux variables étudiées (composante de précipitation et indice climatique), la corrélation est considérée significative pour un niveau de significativité α à n-2 degré de liberté si $|t|$ est supérieur à la valeur critique $t(1-\alpha/2)$ ou si la valeur, p, est inférieur au niveau de significativité α . L'hypothèse nulle (il n'existe pas de corrélation entre les variables) est rejetée pour retenir l'hypothèse alternative (il existe une corrélation significative entre les variables). La valeur critique et la valeur p existent dans le tableau de distribution de Student.

IV.4.2 Analyse composite

Dans ce travail, l'analyse composite sert principalement à étudier les relations des anomalies des pressions à la surface de la mer (SLP) et leur impact sur la variabilité pluviométrique. Cette technique consiste à comparer les valeurs moyennes de deux échantillons extraits de la même série chronologique entre elle. Les deux échantillons correspondent aux valeurs extrêmes de la série chronologique. Les échantillons sont choisis selon un critère extérieur fixé à priori (par exemple les valeurs entre l'intervalle (moyenne \pm 0.5 écart type, moyenne \pm 1 écart type, moyenne \pm 2 écarts type, ou bien en fixant un nombre (n) de plus fortes valeurs et de plus faibles valeurs de la série chronologique). Ensuite, la différence entre les deux échantillons choisis est calculée, puis validée par un test de significativité de la différence à un pourcentage choisi de confiance. Le test de Student est appliqué, dans cette étude, pour l'étude de significativité. La variable SLP choisi est à la base de calcul de la majorité des indices climatiques décrivant les modes de variabilité climatique à grande échelle. Cette analyse permet de vérifier les corrélations trouvées avec les indices et d'investiguer s'ils existent d'autres modes de variabilité, à l'échelle globale, qui ne sont pas prises en considération par les indices. Il est possible également d'évaluer d'autre variables climatiques comme la hauteur géopotentielle, les flux d'humidité, les vents horizontaux et verticaux pour évaluer les configurations atmosphériques correspondant aux anomalies de différents aspects de pluie.

IV.5 Conclusion

Ce chapitre a mis l'accent sur les méthodes et les approches utilisées en décrivant leurs aspects mathématiques et algorithmes. Plusieurs méthodes d'analyse multivariée ont été présentées,

particulièrement, les cartes auto-organisatrices, la classification hiérarchique, data image. A la fin de ce chapitre, les méthodes d'analyse de téléconnexions entre la variabilité climatique globale et les précipitations ont été présentées.

Chapitre V : Analyse factorielle des descripteurs et téléconnexions climatiques

V.1	INTRODUCTION	80
V.2	ANALYSE DES DESCRIPTEURS DE PLUIE PAR UNE ANALYSE EN COMPOSANTES PRINCIPALES ACP	80
V.2.1	<i>Comblement des valeurs manquantes</i>	80
V.2.2	<i>Lecture de la matrice de corrélation</i>	81
V.2.3	<i>Part de la variance expliquée par les facteurs principaux</i>	82
V.2.4	<i>Analyse de corrélation variables-composantes principales</i>	83
V.2.5	<i>Analyse spatio-temporelle des composantes principales retenues</i>	85
V.2.6	<i>Récapitulation de l'analyse en composantes principales</i>	95
V.3	TELECONNEXIONX CLIMATIQUES	95
V.3.1	<i>Corrélations des indices climatiques avec les composantes principales des descripteurs pluviométriques</i>	95
V.3.2	<i>Analyse composite de SLP pour CP1, CP2 et CP3</i>	99
V.4	CONCLUSION	104

V.1 Introduction

L'objectif de ce chapitre est de traiter simultanément les variables en jeu (descripteurs), par une application de l'ACP, dans le but de d'extraire des informations synthétiques pertinentes. Ces informations extraites sous forme de composantes sont exploitées pour une étude de téléconnexions entre les fluctuations climatiques à une échelle globale et les régimes de précipitation aux différentes sous régions de la zone d'étude.

V.2 Analyse des descripteurs de pluie par une analyse en composantes principales ACP

V.2.1 Comblement des valeurs manquantes

L'existence des valeurs manquantes dans une base de données est considérée comme un problème incontournable dans les études statistiques et parmi lesquels l'analyse en composantes principales ACP classique adoptée dans cette étude. Généralement, il existe deux approches pour résoudre ce problème : soit par l'élimination des individus (ou variables) contenant des lacunes de la matrice (individus x variables), ou bien par le comblement des vides.

Dans notre étude et comme indiquée dans la figure II-6, chaque station pluviométrique contient un nombre des valeurs manquantes pour des années différentes sur la période de 50 ans. L'élimination de tous les individus (ou toutes les variables) contenant des valeurs manquantes est la solution la plus radicale. Mais, elle est considérée la méthode la moins recommandée dans la littérature parce qu'elle entraîne la suppression d'un grand pourcentage des données d'observation. En effet, l'adoption de cette approche sur notre base de données a entraîné la perte de 84 % de l'information pluviométrique observée (équivalent à 59 sur 70 stations).

Le comblement des lacunes est l'approche qui permet de garder le maximum des observations. À ce propos, il existe plusieurs techniques d'imputation des données comme remplacer une valeur manquante par une valeur unique (moyenne, médiane, mode...), imputation par centre de groupe, par le plus proche voisin, etc. Il existe également des techniques plus compliquées comme les approches itératives (par exemple : Espérance-Maximisation EM), les méthodes d'imputation par régression : régression linéaire (Hastie et al., 1999), forêts aléatoires (Stekhoven et Buehlmann, 2012), décomposition en des valeurs singulières SVD (Hastie et al., 1999), etc. Chacune de ces techniques présente des avantages et des inconvénients et un pourcentage d'incertitude. Cela dépend de plusieurs facteurs en jeu comme : la nature physique des données, la nature des valeurs manquantes (MCAR : Missing Completely At Random ; MAR : Missing At Random ; NMAR : Not Missing At Random), le pourcentage des valeurs manquantes, le nombre des variables, la distribution des individus.

L'imputation par la moyenne est l'une de méthode qui a montré son efficacité dans le comblement des lacunes dans la base de données pluviométriques par une erreur d'estimation

inférieur à d'autres techniques (Addi et al., 2022). Cette méthode est adoptée dans ce travail avec la condition que le pourcentage des valeurs manquantes pour chaque station soit inférieur à 10 % (Nor et al., 2020). Cette limite fixée pour le taux des lacunes empêche une diminution considérable de la variance des descripteurs pluviométriques pour chaque station et empêche également la distorsion des données et de modifier les liaisons entre les variables en jeu. Alors, les stations contenant moins de 10 % de lacunes, sur une période de 50 ans, sont retenues dans lesquelles les vides ont été remplacés par la valeur moyenne du descripteur de la station. Pour assurer la fiabilité de l'imputation, environ 37 % des stations sont exclues de l'ACP. Par conséquent, l'analyse en composantes principales se fonde alors sur 50 ans d'indices pluviométriques répartis sur les 44 stations qui restent. Les stations retenues pour l'ACP sont illustrées dans la figure V-1.

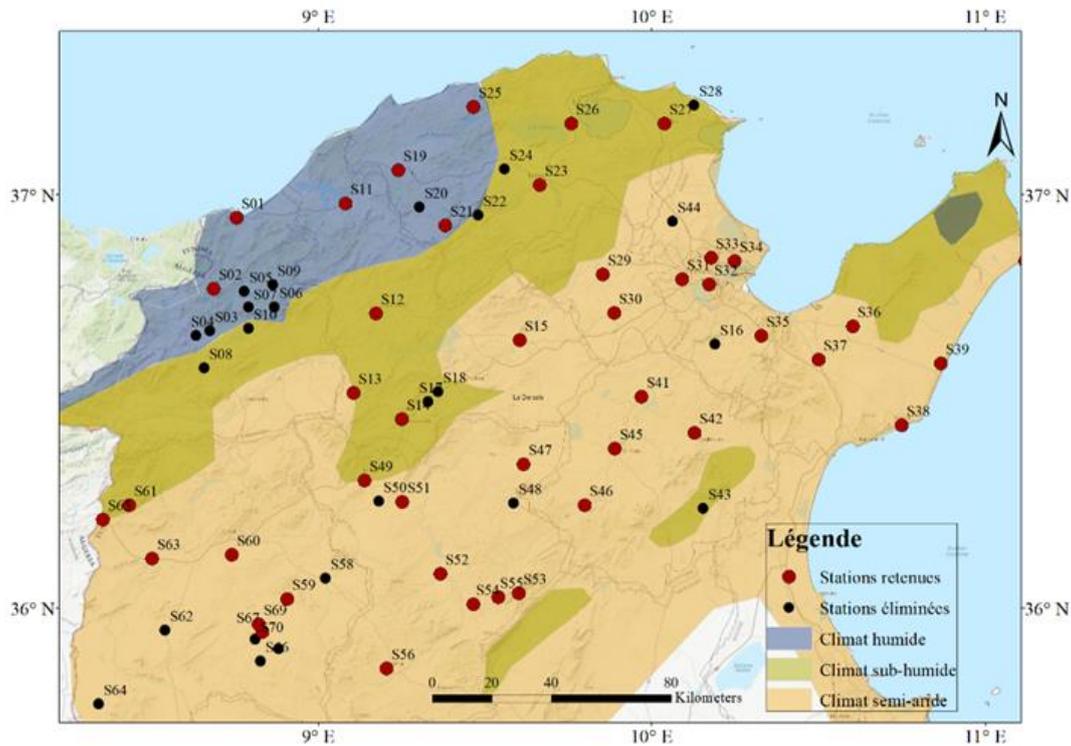


Figure V-1 Répartition des stations retenues pour l'analyse en composantes principales. Les stations en rouge sont retenues (44 stations) et les stations en noir sont éliminées

La base des données multidimensionnelle (6-D dans notre cas) décrivant les séries de 50 ans des saisons DJF dans les stations pluviométriques est analysée dans une première étape où les 6 descripteurs de pluie correspondent aux variables et le nombre de saisons pour toutes les stations correspondent aux individus (50 saisons DJF x 44 stations).

V.2.2 Lecture de la matrice de corrélation

Le tableau V-1 présente les coefficients de corrélations entre les variables choisies. Cette matrice permet d'étudier les relations bilatérales entre les variables en jeu. Le nombre

d'événements par saison N_e montre une faible corrélation avec les autres variables. Les variables qui mesurent la quantité de pluie à différentes échelles temporelles (par saison C_t , par événement $C_{m,e}$, par jour pluvieux $C_{m,j}$) sont positivement et fortement corrélées avec des coefficients de corrélation linéaire supérieur à 0.7. Le nombre de jours pluvieux N_j et le cumul saisonnier C_t sont aussi fortement corrélés.

Tableau V-1 Matrice de corrélation des six variables en jeu (50 ans x 44 stations=2200 individus). Les coefficients en gras représentent des corrélations significatives à 95 %

Variables	N_e	N_j	C_t	$C_{m,e}$	$D_{m,e}$	$C_{m,j}$
N_e	1.00					
N_j	0.37	1.00				
C_t	0.20	0.70	1.00			
$C_{m,e}$	-0.13	0.58	0.91	1.00		
$D_{m,e}$	-0.21	0.79	0.57	0.69	1.00	
$C_{m,j}$	-0.01	0.13	0.72	0.71	0.12	1.00

V.2.3 Part de la variance expliquée par les facteurs principaux

La diagonalisation de la matrice de corrélation (Tab.V-1) donne les résultats affichés dans le tableau V-2. La première colonne représente le numéro (k) des facteurs principaux (ou composantes principales). Le nombre des facteurs est égal au nombre de variables et ces derniers sont hiérarchisés de façon que la part de la variance expliquée par les facteurs principaux diminue en allant du premier facteur au dernier comme indiqué dans la deuxième et troisième colonnes du tableau V-3. Généralement, les premières composantes contiennent l'essentiel de l'information permettant d'expliquer les corrélations de la matrice de départ. Le premier facteur correspond à la première valeur propre décrit à lui seul 57.32 % de la variance totale des données en jeu (Tab.V-2, colonne 3). Le deuxième et le troisième facteur regroupent respectivement 21,11% et 18.88% de la variance totale. Selon le critère de Kaiser, les trois premiers axes sont retenus dans l'analyse puisqu'ils ont des variances (valeurs propres) supérieur à variance moyenne (égale à 1 dans l'ACP normée). Ces trois composantes décrivent 97.31 % (Tab.V-2, colonne 4) de l'information existant dans les données initiales et l'information décrite par les trois autres axes (numéro 4, 5 et 6) peuvent être considérée comme résiduelle.

Tableau V-2 Valeurs propres et pourcentage de la variance expliquée par chaque axe factoriel (nombre d'observations : 2200, nombre de variables : 6)

N° du facteur (k)	Valeur propre (λ_k)	Variance expliquée par l'axe k (en %)	Variance cumulée (en %)
1	3.44	57.32	57.32

2	1.27	21.11	78.43
3	1.13	18.88	97.31
4	0.10	1.59	98.89
5	0.05	0.90	99.79
6	0.01	0.21	100

V.2.4 Analyse de corrélation variables-composantes principales

Les cercles de corrélation existant dans les figures V-2a et V-2b permettent d'étudier les corrélations entre les variables initiales et les trois facteurs principaux retenus. Les points de projection d'une variable sur un axe k correspondent au coefficient de corrélation entre la variable et le facteur principal k. La distance entre le centre du cercle de corrélation et la variable projetée reflète la qualité de représentation de cette variable, c-à-dire les variables proches du cercle sont considérées bien représentées par le plan factoriel selon le critère de qualité de représentation adopté pour évaluer la qualité de projection des variables (voir annexe B). Le tableau V-3 présente la contribution et la qualité de représentation de chaque variable dans la définition des trois facteurs principaux (voir annexe B).

Sur le plan factoriel 1-2 (Fig.V-2a), le cumul saisonnier C_t et le cumul moyen par événement $C_{m,e}$ sont fortement corrélés avec l'axe 1 où le coefficient de corrélation se trouve proche de 1. Ils ont également plus 50 % de contribution dans la formation de cet axe (Tab.V-3). Ainsi, le nombre des jours pluvieux N_j , la durée moyenne par événement $D_{m,e}$ et le cumul moyen par jours pluvieux $C_{m,j}$ ont une corrélation supérieure à 0.5 avec l'axe 1. À l'exception du nombre d'événements N_e , toutes les variables varient dans la même direction par rapport à l'axe 1. Par rapport au deuxième axe factoriel, le nombre d'événements N_e est très fortement anti-corrélé et contribue à lui seul à 58 % de la définition de deuxième axe factoriel (Tab.V-3). Le descripteur du nombre de jours N_j a également une corrélation significative avec l'axe 2. La position de N_e dans le cercle de corrélation montre qu'une forte intermittence d'événements pendant une saison donnée ne correspond pas forcément à une saison pluvieuse.

La durée moyenne par événement $D_{m,e}$ et le cumul moyen par jour pluvieux $C_{m,j}$ ont la plus grande participation à la formation du facteur 3 (66 % de contribution). L'axe 3 oppose deux variables : $C_{m,j}$ est positivement corrélé avec le facteur 3 contrairement à $D_{m,e}$ (Fig.V-2b).

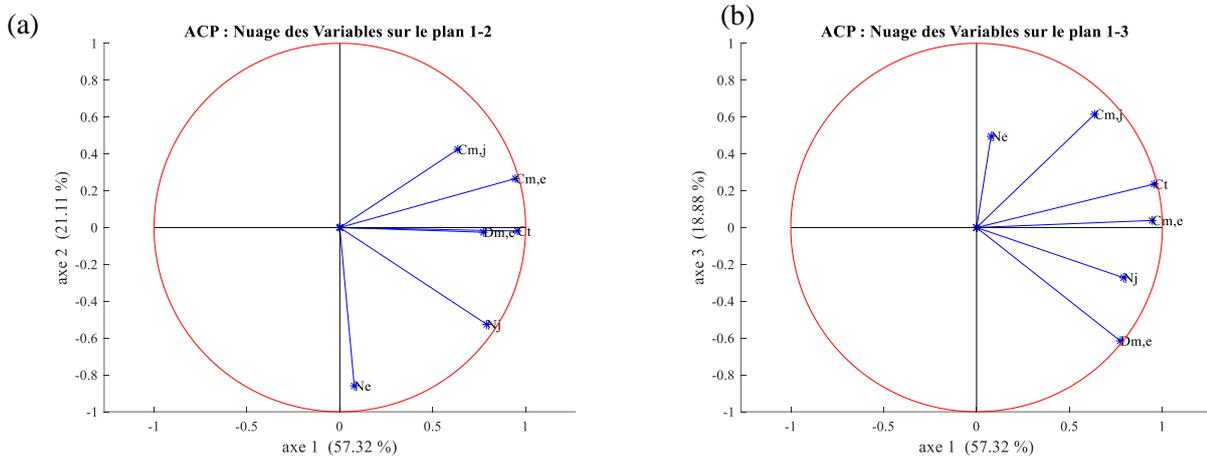


Figure V-2 (a) Projection des variables (descripteurs pluviométriques) sur le plan factoriel 1-2, (b) sur le plan factoriel 1-3

Tableau V-3 Contribution et qualité de représentation de chaque variable à la formation des trois premiers axes factoriels (en %)

	Contribution de la variable (%)			Qualité de représentation (%)		
	Axe 1	Axe 2	Axe 3	Axe 1	Axe 2	Axe 3
N_e	0	58	22	0	74	25
N_j	18	22	7	67	28	7
C_t	27	0	5	91	0	6
$C_{m,e}$	26	6	0	89	0	0
$D_{m,e}$	17	0	33	60	0	38
$C_{m,j}$	12	14	33	40	18	38

La lecture de projection des variables sur le plan factoriel et les corrélations avec les composantes principales permet de donner une signification aux axes. Dans notre cas, le premier axe s'avère être globalement un indicateur de quantité de pluie à l'échelle de la saison et à l'échelle de l'évènement. Il oppose les saisons pluvieuses et les saisons sèches. L'axe 2 reflète le caractère intermittent de la saison, anti-corrélé avec le nombre d'évènements N_e ou dans une moindre mesure avec le nombre de jours pluvieux N_j et corrélé avec les cumuls moyens par jour et par évènements. Il oppose les saisons avec beaucoup d'évènements à celles au cours desquelles peu d'évènements ont eu lieu. L'axe 3 qui contient moins d'information est plus difficile à interpréter. Il oppose la durée moyenne par évènement $D_{m,e}$ et le cumul moyen par jours pluvieux $C_{m,j}$, c'est-à-dire des saisons avec les évènements longs et des jours peu intenses ou bien évènements courts et des jours intenses. Ces types d'évènements sont liés essentiellement aux conditions météorologiques : pluie engendrée par un front froid ou un front chaud.

V.2.5 Analyse spatio-temporelle des composantes principales retenues

Après la synthétisation des descripteurs pluviométriques dans trois composantes principales notées dans la suite du manuscrit (CP1, CP2 et CP3), chaque composante sera étudiée séparément dans le but de d'extraire les modes de variabilité spatiale sur une période de 50 ans. Pour ce faire, une deuxième Analyse en composantes principales est réalisée à chaque composante principale retenue. Les composantes sont réorganisées dans une matrice (50 x 44) où les individus sont les 50 ans d'observations et les variables sont les stations pluviométriques en jeu.

V.2.5.1 Analyse spatio-temporelle de la première composante principale CP1 : quantité de pluie

La figure V-3 montre l'évolution temporelle de CP1 sur une période de 50 ans dans les 44 stations et le trait noir en gras illustre la moyenne des composantes principales. Les boîtes à moustaches (à droite) représentent la distribution de la composante CP1 sur toute la zone d'étude sur trois périodes différentes. La première est la période totale, la deuxième est une sous période de 1960-1990 et la troisième est une sous période de 1991-2009. L'analyse interannuelle de CP1 permet de distinguer des années exceptionnelles avec des extrêmes positives (1971, 1973 et 2003) et négatives (2002 et 2008). Les valeurs positives de CP1 correspondent aux saisons pluvieuses et les valeurs négatives aux saisons sèches. Par période, une augmentation de la médiane de composante de la quantité de pluie CP1 dans la période la plus récente (1991-2009).

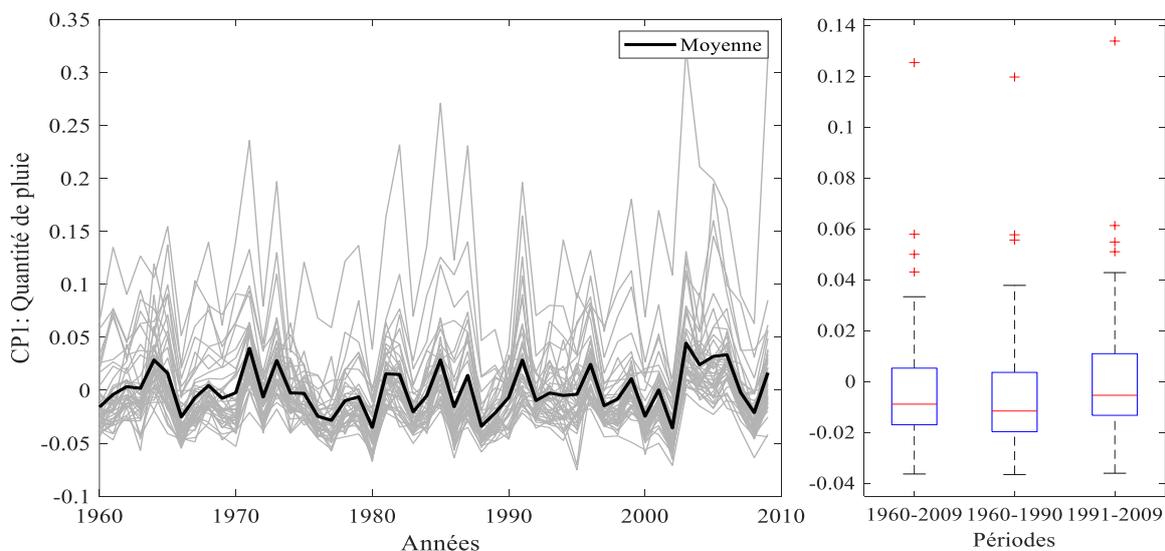


Figure V-3 Séries chronologiques de la première composante principale CP1 (quantité de pluie) dans 44 stations entre 1960 et 2009. Le trait noir en gras représente la moyenne des 44 stations par an d'étude. Les boîtes à moustache représentent la variabilité des composantes des 44 stations sur trois périodes (1960-2009 ; 1960-1990 ; 1991-2009)

La part de la variance expliquée par les axes principaux est illustrée dans la figure V-4. La première composante et la deuxième décrivent respectivement 63.1 % et 9 % de l'inertie totale, ce qui donne en totale 72.1 %. Ce diagramme d'inertie montre un décrochement au troisième axe suivi d'une décroissance régulière d'inertie pour les axes restants. Dans l'analyse qui suit, les deux premiers axes (premier plan factoriel) sont retenus.

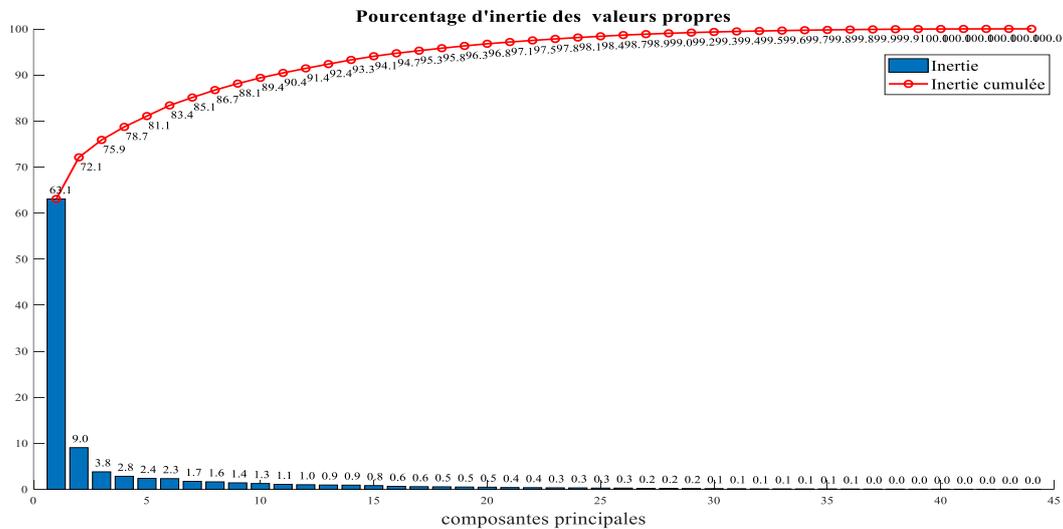


Figure V-4 Pourcentage d'inertie et d'inertie cumulée expliquées par les axes principaux de l'analyse de CPI(50 ans x 44 stations)

La projection des variables (stations) dans le cercle de corrélation illustrée dans la figure V-5 (à gauche) montre que toutes les variables sont corrélées positivement entre elles et corrélées fortement avec l'axe 1. La première composante définit un facteur taille, cet axe correspond à une synthèse de l'effet commun de ces variables, L'axe 1 est un indicateur du cumul global à l'échelle de toute la région. L'axe 2 oppose deux groupes de variables (stations). Certaines stations contribuent positivement à la formation de l'axe et les autres négativement (Fig.V-5). Cet axe est connu par le facteur forme. Les stations qui sont fortement et positivement corrélées avec l'axe 2 (dans le cercle de corrélations) couvrent la partie sud-est de la zone d'étude et les stations négativement et fortement corrélées avec l'axe 2 couvrent la partie nord-ouest (Fig.V-5). L'axe 2 est un indicateur d'écart de pluviométrie entre la région nord-ouest et la partie sud et particulièrement le sud-est de la zone d'étude. Cet écart résulte de l'exposition de ces deux groupes de stations aux flux pluvieux. En effet, les stations situées au nord-ouest sont directement exposées aux flux d'ouest et du nord-ouest qui sont les flux dominants pendant l'hiver. Les stations localisées au nord-est sont influencées par les flux d'est (voir la section II.3).

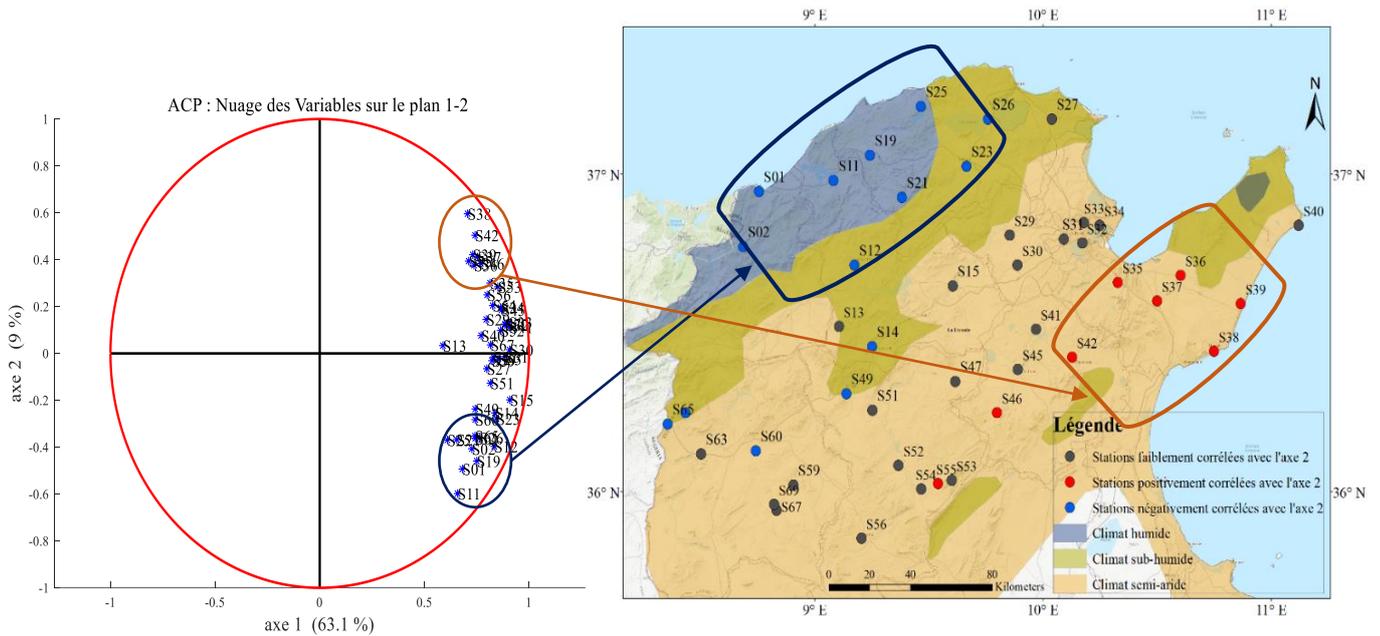


Figure V-5 Corrélation des stations avec les axes principaux de CP1 (à gauche) et répartition des stations corrélées et anti-corrélées avec l'axe 2 (à droite). Les stations en gris sont faiblement corrélées avec le premier plan factoriel

La figure V-6 visualise la projection des individus (années) sur le premier plan factoriel. Les marqueurs des individus sont deux triangles proportionnés par les deux composantes et leur taille reflète la qualité de représentation (voir annexe B). Les triangles sont orientés selon l'abscisse et l'ordonnée (vers la droite, c'est l'abscisse et vers le haut, c'est l'ordonnée). Selon les tailles des triangles (Fig.V-6), plus de 50 % des années sont bien représentées sur ce plan factoriel. Par une inspection visuelle, le nuage des individus peut être divisé en cinq groupes. En se référant de l'analyse univariée des indices pluviométriques (Fig.III-4 et Fig.III-5) et de la variabilité de composante principale CP1 (Fig.V-3) les différents groupes sont identifiés. Le groupe 1 rassemble les années (les saisons DJF) avec les conditions les plus sèches que la moyenne qui se manifestent par un faible cumul saisonnier C_t (Fig.III-4c), un faible cumul par événement $C_{m,e}$ (Fig.III-5a) et un faible nombre de jours pluvieux N_j (Fig.III-4b) avec un gradient normal entre la région nord-ouest et la partie sud, car ces années sont au centre du nuage de point par rapport à l'axe 2. Les années du groupe 2 ont une pluviométrie proche de la moyenne (origine du repère dans l'ACP normé) ou bien, elles sont mal représentées par ce plan factoriel. Les années du groupe 3 correspondent les années les plus humides traduit par des forts cumuls par événement pluvieux d'où d'un cumul saisonnier important également avec un faible gradient entre la région nord-ouest et la partie sud. Pendant les années appartenant au groupe 4, l'écart pluviométrique entre les stations du nord et celles du sud diminue par rapport à la moyenne où l'apport pluviométrique des stations du nord-ouest diminuent exceptionnellement

contrairement aux années du groupe 5 où l'écart de l'apport entre le nord-ouest et le sud-est devient plus grand.

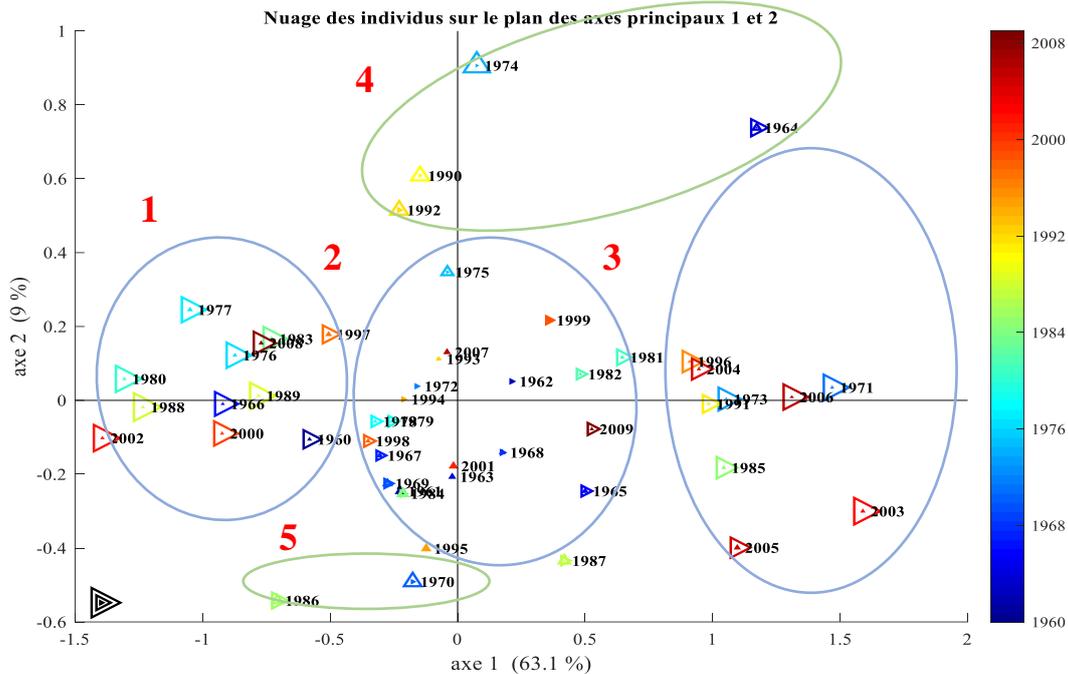


Figure V-6 Projection des années (50 ans) sur le premier plan factoriel de CP1. La taille de triangle représente la qualité de représentation. Les couleurs sont des indicateurs des années de plus anciennes (bleu) au plus récentes (rouge)

En résumé, l'axe 1 reflète l'impact des phénomènes globaux à grande échelle qui induit un comportement identique pour toutes les stations. L'axe 2 reflète l'influence des flux provenant de plusieurs directions qui agissent exceptionnellement sur la pluviosité des stations directement exposées à ces flux.

V.2.5.2 Analyse spatio-temporelle de la deuxième composante principale CP2 : intermittence de pluie

La deuxième composante principale CP2 retenue de l'analyse et la synthèse des six descripteurs pluviométriques représente l'intermittence de pluie pendant la saison DJF avec un pourcentage 21 % de la variance totale des données (voir Tab.V-2). Le nombre d'événements N_e est l'indice qui contribue majoritairement à la formation de ce facteur suivi par le nombre de jours par saison N_j (Tab.V-3). Une analyse des séries chronologiques de CP2 est effectuée pour les 44 stations de la zone d'étude (Fig V-7) (courbes en gris). Sur le plan temporel, il existe des années exceptionnelles qui se distinguent par une chute considérable de la moyenne des valeurs de CP2 (1973, 1996 et 2005) ou bien par une augmentation importante (1975, 2002 et 2008). Les chutes de CP2 correspondent à une forte intermittence d'événements pluvieux pendant la saison et les valeurs élevées de CP2 correspondent à une faible intermittence

d'événements (voir Fig.III-4a et Fig.III-4b). L'analyse de cette composante d'intermittence par période, par boîte à moustache, montre que la dispersion devient plus faible durant la période récente (la différence X75%-X25% devient plus petite) avec l'apparition des valeurs aberrantes.

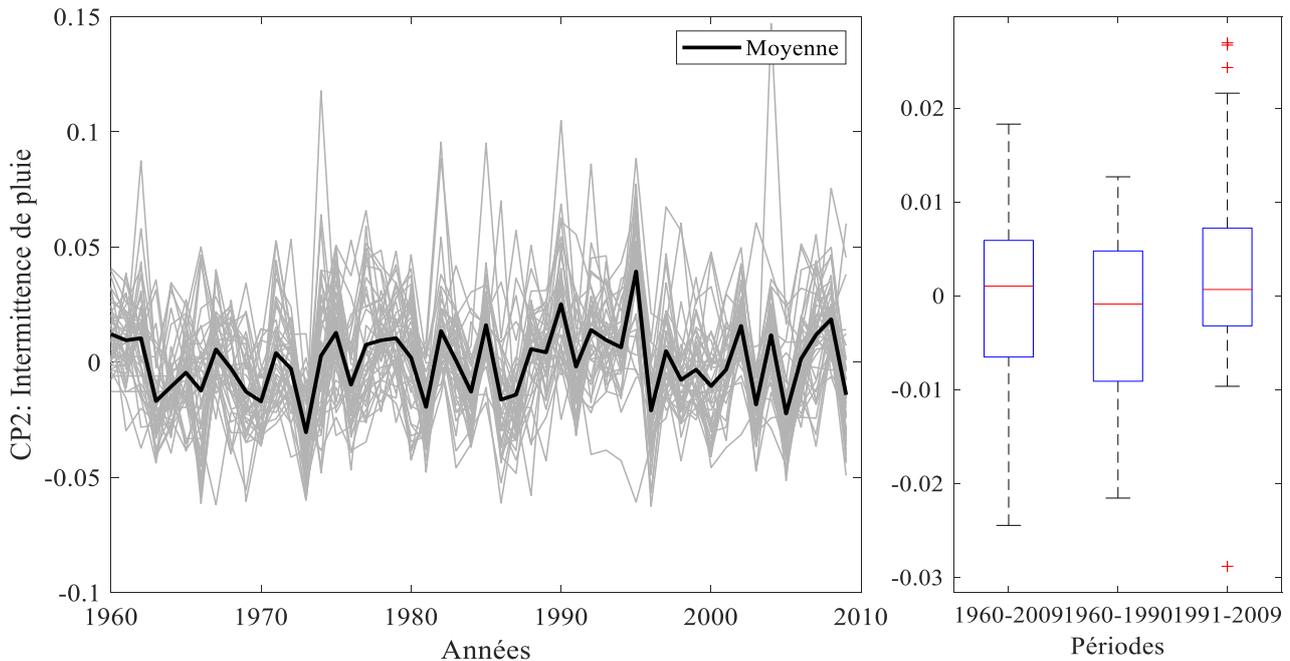


Figure V-7 Séries chronologiques de la deuxième composante principale CP2 (intermittence de pluie) dans 44 stations entre 1960 et 2009. Le trait noir en gras représente la moyenne des 44 stations par an. Les boîtes à moustache représsentent la variabilité des composantes des 44 stations sur trois périodes (1960-2009 ;1960-1990 ;1991-2009)

L'analyse en composantes principales de la matrice (50 x 44) de CP2 donne le diagramme d'inertie visualisé dans la figure V-8. Les deux premiers axes principaux reflètent 50% de la variance. Il est retenu pour la suite d'analyse et il représente le mode de variabilité dominant de cette série. Le 50 % d'inertie restante est répartie sur les autres axes principaux avec des faibles pourcentages et qui reflète une variabilité très localisée dans le temps et dans l'espace.

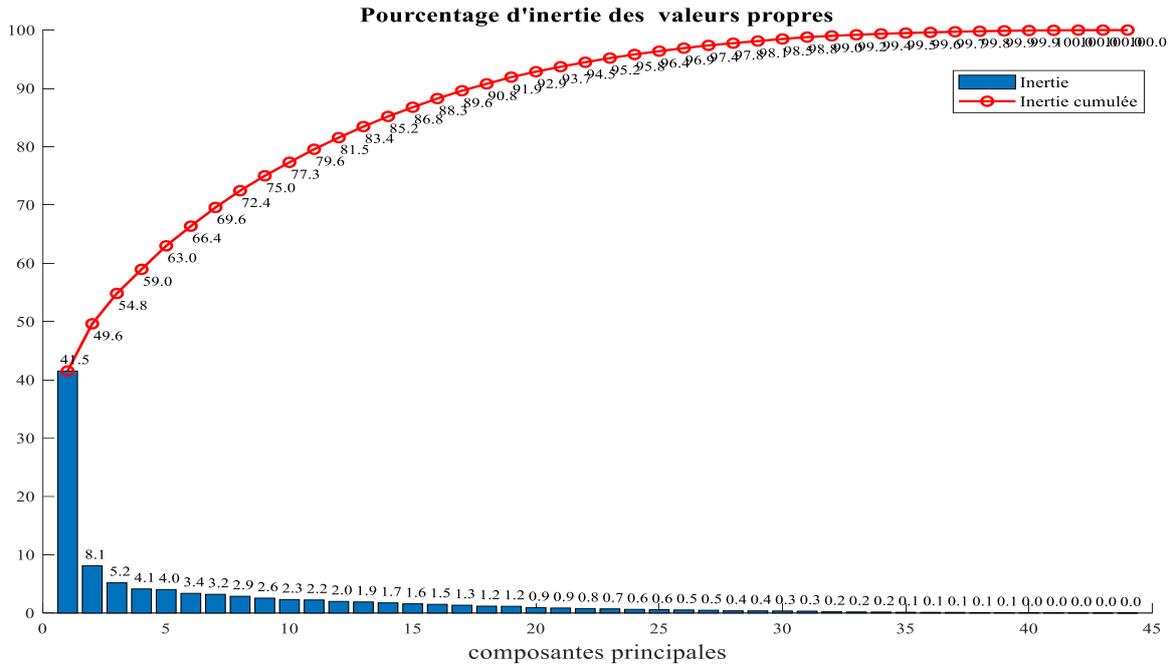


Figure V-8 Pourcentage d'inertie et d'inertie cumulée expliquées par les axes principaux de l'analyse de CP2 (50 ans x 44 stations)

La projection des stations (variables) sur le premier plan factoriel dans la figure V-9 montre que la plupart des variables sont positivement et assez fortement corrélées avec l'axe 1 à l'exception de la station S02 (Ain Drahem) qui est fortement corrélée avec l'axe 2. D'autres stations comme S11 (Nefza délégation) et S13 (Thibar) sont mal représentées par ce premier plan factoriel. D'après la figure V-9, l'écart NO-SE persiste encore pour la variable caractérisant l'intermittence de pluie.

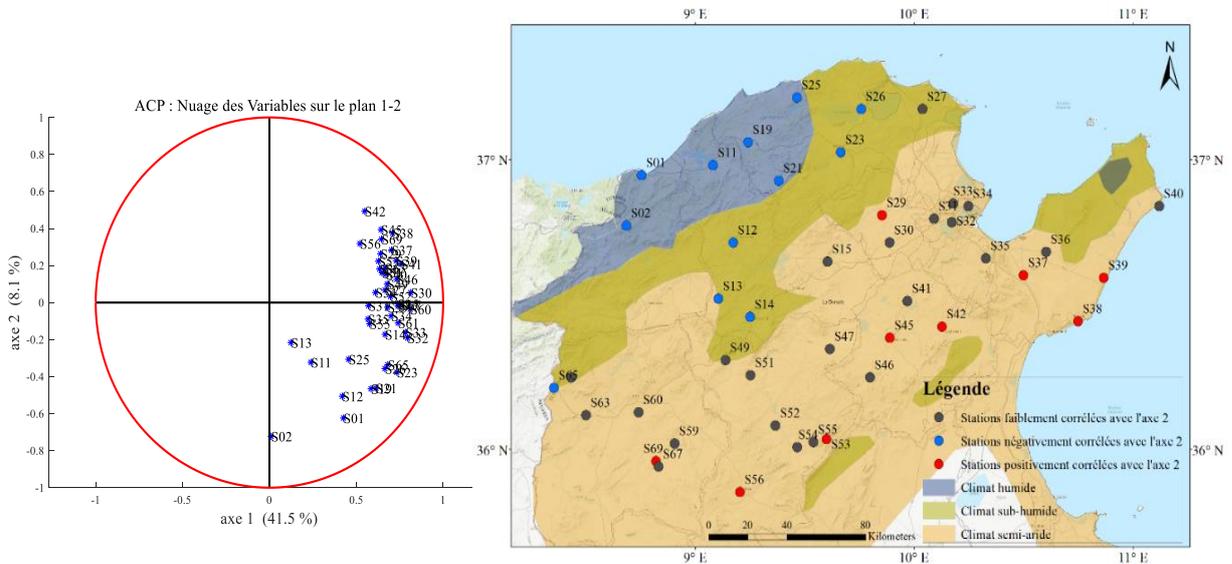


Figure V-9 Corrélation des stations avec les axes principaux de CP2 (à gauche) et répartition des stations corrélées et anti-corrélées avec l'axe 2 (à droite). Les stations en gris sont faiblement corrélées avec le premier plan factoriel

Dans la figure V-10, les années projetées sur le premier plan factoriel peuvent être divisé en 5. Les années du groupe 1 correspondent aux saisons avec une intermittence d'événements importante (voir Fig.III-4a). Le groupe 2 rassemble les années avec un nombre d'événements moyen et les années mal représentées par ce plan factoriel. Le groupe 3 correspond aux saisons avec faible nombre d'événements pluvieux. Le groupe 4 rassemble les années dont l'écart entre les stations de la zone semi-aride (sud) et les stations de la zone humides (nord-ouest) devient plus élevé que la moyenne et inversement pour le groupe 5.

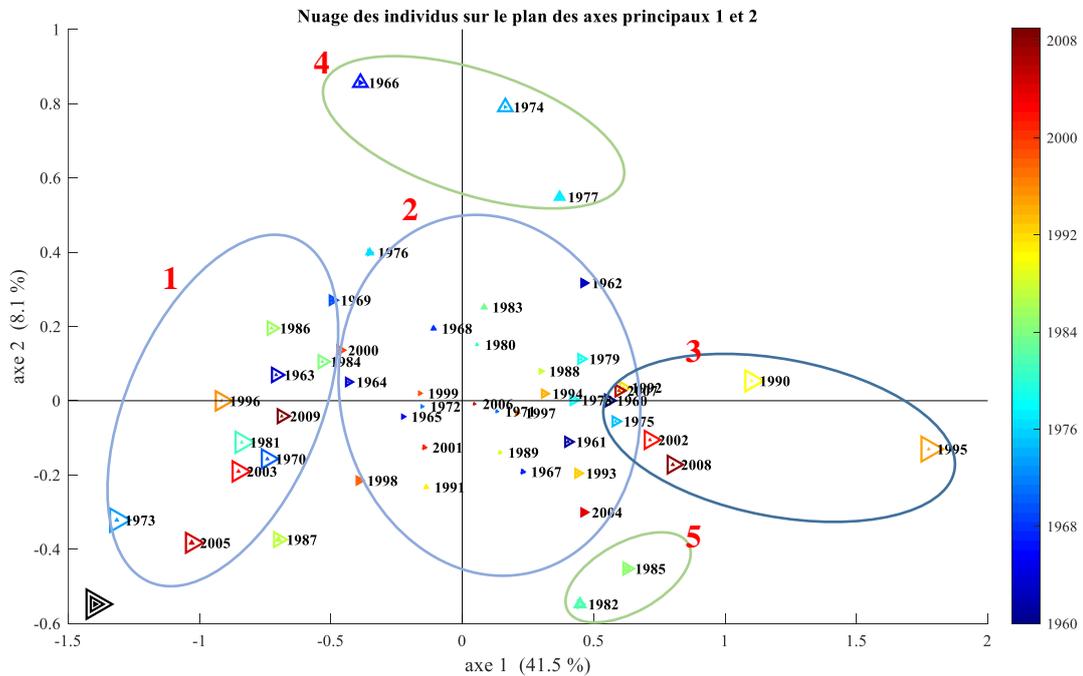


Figure V-10 Projection des années (50 ans) sur le premier plan factoriel de CP2. La taille de triangle représente la qualité de représentation. Les couleurs sont des indicateurs des années de plus anciennes (bleu) au plus récentes (rouge)

L'année 1974 se distingue par un écart entre les stations de la zone semi-aride (sud) et les stations de la zone humides (nord-ouest) aussi bien pour ce qui concerne les cumuls d'eau que les nombre d'évènements. L'année 1990 qui se distingue également (figure V-6) par un écart important des cumuls d'eau entre les deux zones se caractérise par un nombre d'évènements particulièrement faible pour toute la région. Inversement, l'année 1966 se distingue par un écart important du nombre d'évènements, entre les deux zones, se caractérise par un cumul d'eau particulièrement faible pour toute la région (figure V-6).

V.2.5.3 Analyse spatio-temporelle de la troisième composante principale CP3 : conditions météorologiques

La troisième composante principale CP3, retenue de la synthétisation des indices pluviométriques, décrit 18.88 % de la variance (Tab.V-2). Ce facteur est un indicateur de type

d'événement pluvieux qui s'appuie sur deux variables : la durée moyenne par événement $D_{m,e}$ et le cumul moyen par jours pluvieux $C_{m,j}$. Physiquement, ce caractère est lié essentiellement à la nature de rencontre entre les masses d'air de différentes températures qui ont engendré les précipitations (front chaud ou front froid). En effet, le front froid donne des événements pluvieux brusques, courts et intenses contrairement au front chaud qui donne des pluies peu intenses sur une longue durée. L'analyse des séries chronologiques de CP3 dans les 44 stations de la zone d'étude sur une période de 50 ans est visualisée dans la figure V-11. Les grandes valeurs de CP3 (par exemple en 1974) correspondent à une saison DJF avec un cumul moyen par jour pluvieux $C_{m,j}$ extrêmement faible et une grande durée moyen par événement $D_{m,e}$ (voir Fig.III-5b et c), c-à-dire des saisons dominées par des pluies stratiformes. Par contre, une faible valeur de CP3 (par exemple en 1995) correspond à cumul moyen par jour pluvieux $C_{m,j}$ exceptionnellement fort et à une $D_{m,e}$ moyen et même faible (voir Fig.III-5b et c), c-à-dire pluie de nature convective. Les séries chronologiques de CP3 dans les 44 stations (en gris) montrent une grande variabilité interannuelle et entre les stations. A l'échelle des périodes, la dispersion (écart entre 25^e et 75^e percentiles) de cette composante a augmenté dans la zone d'étude dans le période la plus récente. En outre, les trois statistiques (25^e percentile, la médiane et 75^e percentiles) ont augmenté considérablement (dans le même sens) durant la dernière période de 1990-2009.

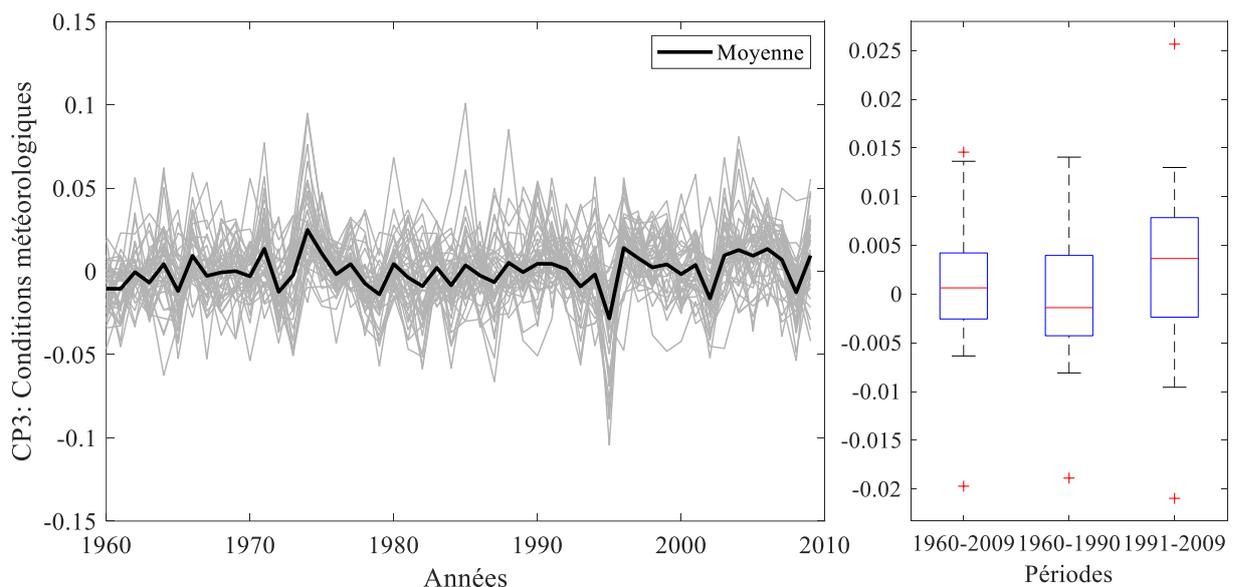


Figure V-11 Séries chronologiques de la première composante principale CP3(conditions météorologiques) dans 44 stations entre 1960 et 2009. Le trait noir en gras représente la moyenne des 44 stations par an. Les boîtes à moustache reprsésentent la variabilité des composantes des 44 stations sur trois périodes (1960-2009 ;1960-1990 ;1991-2009)

Une analyse par ACP de la matrice (50 x 44) de CP3 a été réalisée pour extraire et comprendre le mode de variabilité de cet aspect de pluie pendant la saison DJF. Le diagramme d'inertie (Fig.V-12) montre que le premier axe ne présente que 26.2 % de la variance et que le premier plan factoriel cumule 34.7 %. Le reste de l'inertie est répartie sur plusieurs axes avec un faible pourcentage pour chacun. Seulement le premier plan est retenu pour la projection des variables et individus.

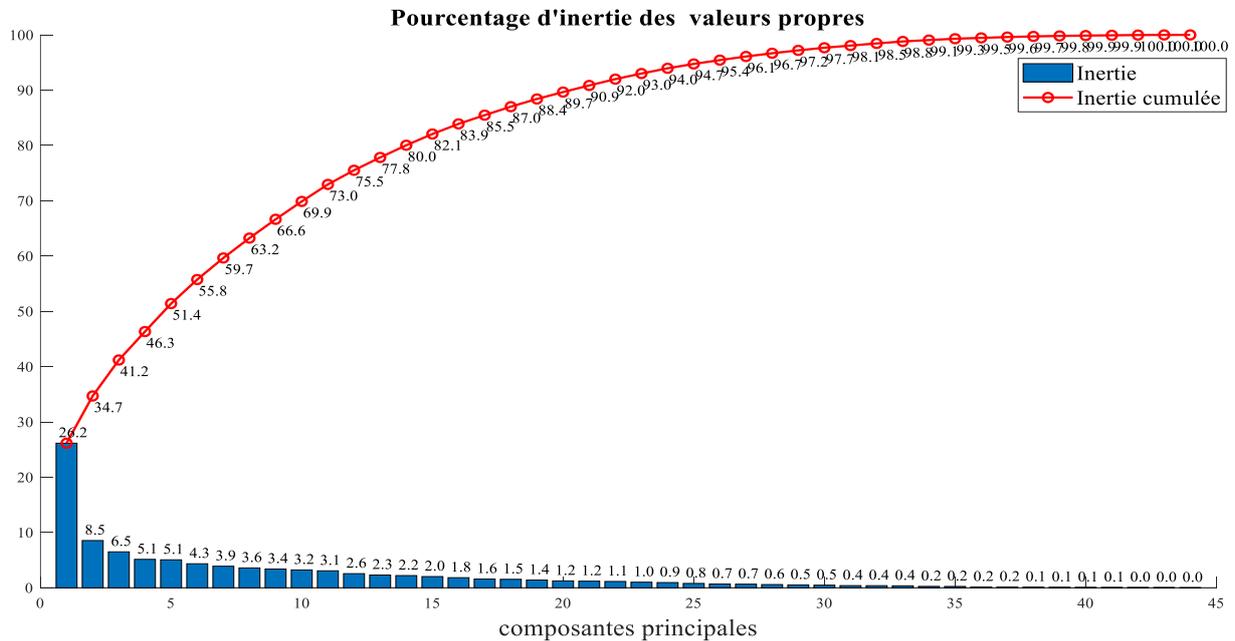


Figure V-12 Pourcentage d'inertie et d'inertie cumulée expliquées par les axes principaux de l'analyse de CP3 (50 ans x 44 stations)

Le cercle de corrélation des stations (Fig.V-13 à gauche) montre que les stations sont positivement corrélées avec l'axe 1 à l'exception de S13 (Thibar). La plupart des stations sont moyennement et assez fortement corrélées avec l'axe 1. La position de quelques stations par rapport à l'origine comme S13 (Thibar), S02 (Ain Drahem), S19 (Sejnen), S12 (Béja Inrat) et S15 (Mjez el Beb) indique qu'elles sont mal représentées par ce plan factoriel. Pour la suite, seulement l'axe 1 sera interprété et exploité de l'analyse de la composante CP3. L'axe 2 de CP3 présente très peu de variance dans la base de données qui ne dépasse pas 1.5%.

Dans (Fig.V-13 à droite), les stations les plus corrélées avec l'axe 1 sont colorées en rouge et les stations faiblement corrélées avec le même axe sont en gris. Les stations en rouge sont localisées dans la partie sud de la zone d'étude avec une orientation SE-NE. Les conditions météorologiques spécifiques (ou les flux) induisant des pluies convectives ou stratiformes agissent seulement sur le couloir contenant les stations en rouge.

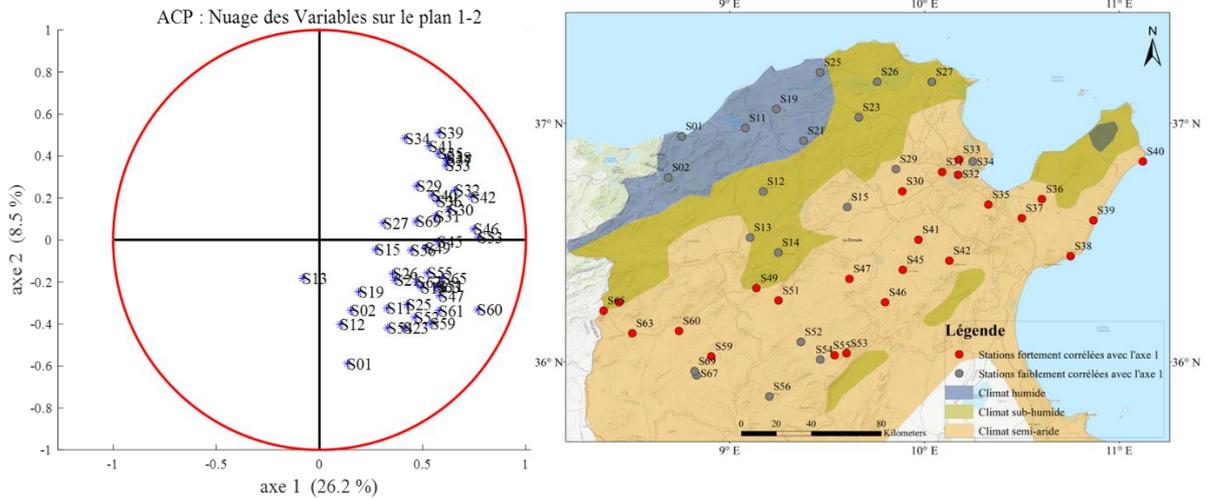


Figure V-13 Corrélation des stations avec les axes principaux de CP3 (à gauche) et répartition des stations corrélées et anti-corrélées avec l'axe 2 (à droite). Les stations en gris sont faiblement corrélées avec le premier plan factoriel.

Par rapport à l'axe 1 (Fig.V-14), les années peuvent être divisées en 3 classes. Les années de la classe 1 (ex. 2002 et 2008) correspondent aux saisons avec des longs événements ($D_{m,e}$ élevée) et pluies peu intenses ($C_{m,j}$ faible) par rapport à la moyenne dans les stations du secteur sud (Fig.V-13). La classe 2, qui est la plus grande, contient les années mal représentées par le plan factoriel. Les années de la classe 3 (ex. 1974, 1996), elles se caractérisent par des événements pluvieux courts et intenses dans les stations bien représentées par l'axe 1.

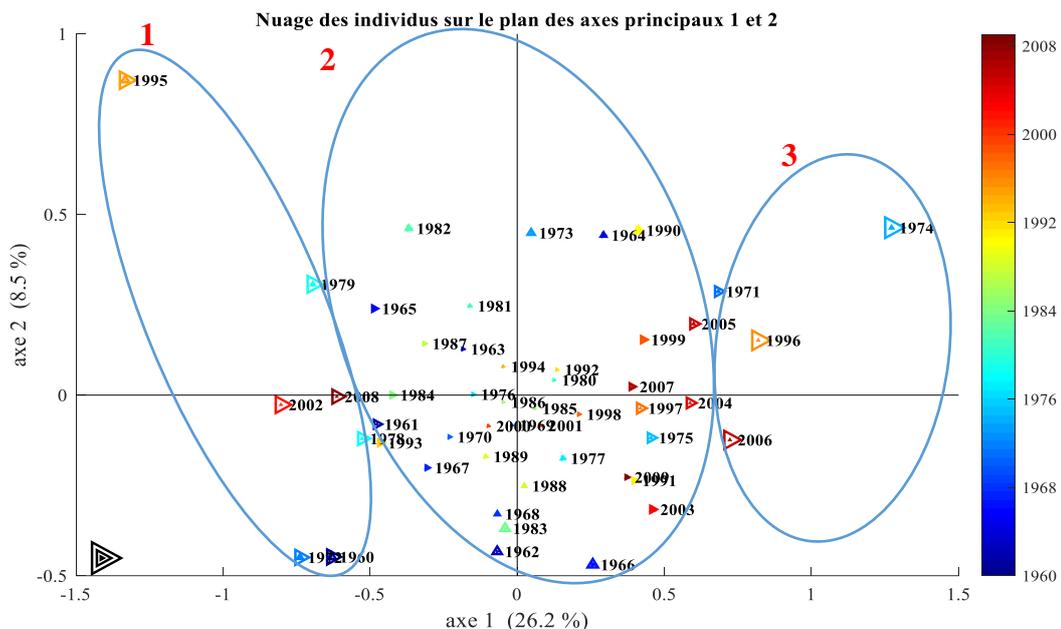


Figure V-14 Projection des années (50 ans) sur le premier plan factoriel de CP3. La taille de triangle représente la qualité de représentation. Les couleurs sont des indicateurs des années de plus anciennes (bleu) au plus récentes (rouge)

V.2.6 Récapitulation de l'analyse en composantes principales

Comme indiqué dans le diagramme de la figure (Fig.V-15), l'analyse en composantes principales s'est effectuée sur deux phases : la première ACP réduit la base de données de dimension six à une espace tridimensionnelle formé de trois axes principaux CP1, CP2 et CP3. Chaque composante représente des informations synthétiques sur la période de 50 ans dans les 44 stations en jeu. Ensuite, la deuxième ACP, appliqué aux composantes, a permis d'extraire les modes de variabilité spatiale dominants représentés par les deux premiers axes principaux. Les deux premiers axes de CP1 et CP2 et le premier axe de CP3 sont exploitées dans la partie suivante pour investiguer les probables liaisons avec la variabilité climatique à grande échelle.

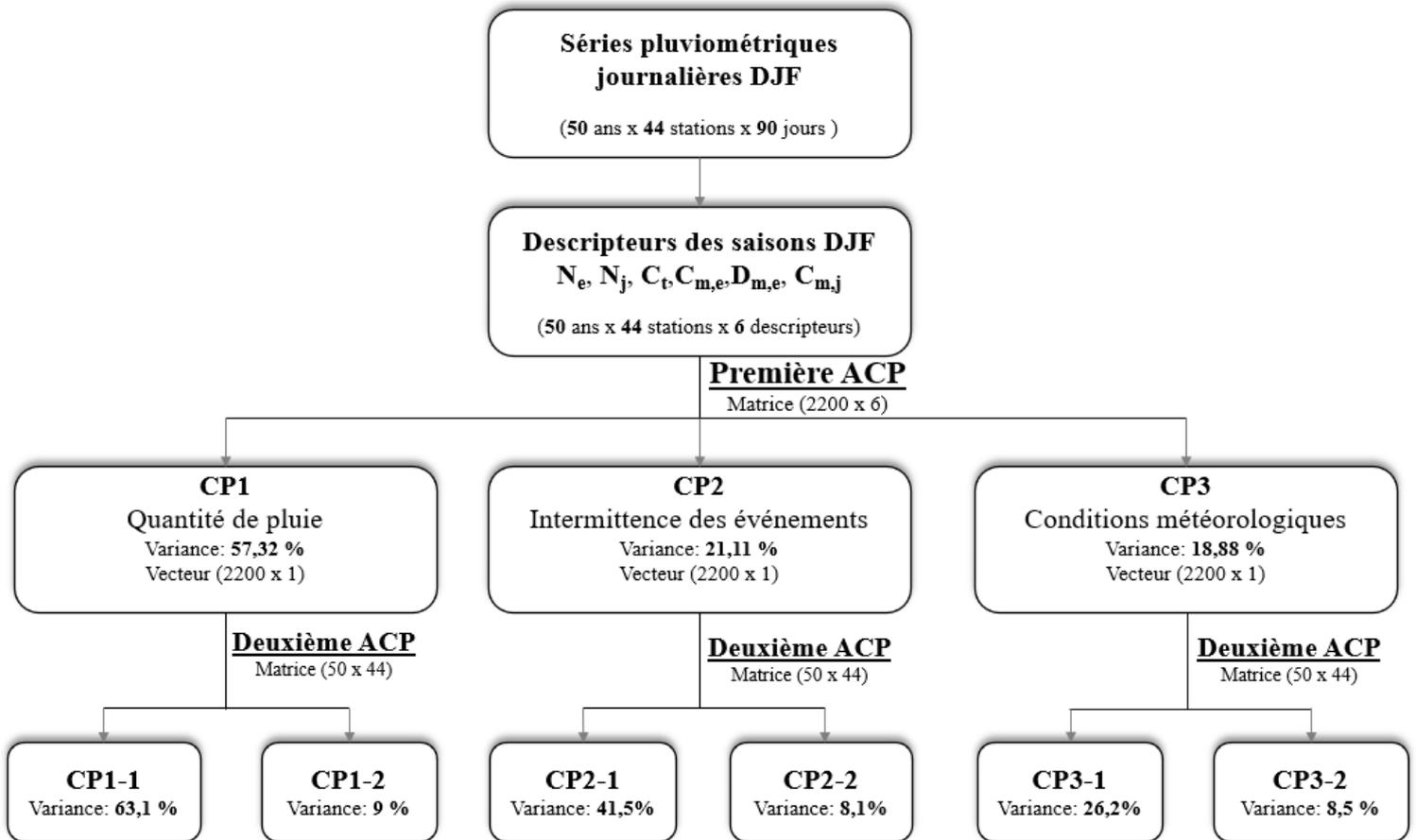


Figure V-15 Diagramme synthétique des résultats des deux ACP et la variance expliquée des axes retenus

V.3 Téléconnexionx climatiques

V.3.1 Corrélations des indices climatiques avec les composantes principales des descripteurs pluviométriques

Les modes de variabilité climatiques sont exprimés par des indices calculés de manières différentes et décrits en détails dans l'annexe A. Les indices climatiques IC considérés correspondent aux oscillations suivantes : (1) oscillation nord atlantique NAO, (2) oscillation arctique AO, (3) oscillation australe SO, (4) oscillation décennale du Pacifique PDO, (5)

oscillation méditerranéenne de l'ouest WeMO, (6) oscillation méditerranéenne MO. Chaque indice est un indicateur de phase et d'amplitude du mode de variabilité climatique en jeu. Le NAO est une oscillation de pression au nord de l'Atlantique entre la région de dépression subpolaire (Islande) et la région de l'anticyclone des Açores. Le AO est une anomalie de la hauteur géopotentielle de 1000 millibars à l'hémisphère nord (de 20°N à 90°N). Le SO est une oscillation de pression entre l'est et l'ouest du Pacifique du sud. Le PDO est une anomalie de température centrée sur l'océan Pacifique aux latitudes moyennes. Le WeMO et le MO sont des modes de variabilité climatiques régionales, ils décrivent des oscillations de pressions entre deux points dans la région ouest de Méditerranée et la région est respectivement. Dans le but d'investiguer l'existence et la nature des liens entre ces indices et la synthèse l'information pluviométrique issue de l'ACP, une analyse de corrélation a été élaborée dont ses principaux résultats sont récapitulés dans le tableau V-4. Les séries chronologiques des IC sont utilisés à une échelle saisonnière : les indices climatiques DJF décrivant les oscillations qui se déroulent simultanément avec les pluies hivernales étudiées sont indiquées par (0) dans le tableau V-4, les indices décalés dans le temps d'une saison (septembre-octobre-novembre SON), de deux saisons (juin-juillet-août JJA) et de trois saisons (mars-avril-mai MAM) sont indiqués dans le tableau V-4 par (-1), (-2) et (-3) respectivement. L'étude de corrélation avec un décalage est proposée pour investiguer les liaisons asynchrones entre les précipitations, dans la zone d'étude, et la circulation atmosphérique générale les nuages des points des composantes pluviométriques et indices climatiques sont présentés dans l'annexe C.

Le tableau V-4 montre que la plupart des coefficients linéaires de corrélation entre les différentes composantes de pluie et les indices climatiques sont faibles et peu significatives. Pour 6 indices et 5 composantes et 4 saisons de calcul d'indices (120 cas) on trouve 14 coefficients significatifs. Toutefois leur pouvoir prédictif est faible vu que le coefficient ne dépasse pas 0.4 en valeur absolue. Les corrélations positives et négatives, statistiquement significatives au niveau de 95%, sont indiquées par des couleurs (Tab.V-4). Pour les axes CP1-1 et CP1-2 issues de l'ACP de CP1, le tableau V-4 montre que les indices sont faiblement corrélés avec l'axe 1 (CP1-1) mais présentent quelques corrélations significatives avec l'axe CP1-2. Pendant la saison DJF, l'oscillation nord Atlantique NAO et l'oscillation Arctique AO agissent sur l'écart entre la pluviométrie des stations de la partie sud et les stations du nord-ouest (l'axe CP1-2).

En effet, la phase positive du NAO (et de même pour AO), traduit par un indice positif, coïncide avec des conditions plus froides et sèches qui touchent la région nord-ouest de la Tunisie. Par conséquence, l'apport pluviométrique des stations du nord-ouest diminue exceptionnellement

et parfois au-dessous de la pluviométrie des stations de sud qui sont habituellement les plus sèches. L'indice PDO décrivant l'oscillation décennale du pacifique montre une corrélation négative significative avec l'axe CP1-2, c-à-dire les valeurs négative de PDO, décrivant la phase froide de PDO, coïncide avec des conditions plus sèches que la moyenne au nord-ouest ou bien plus humide au sud-est de la zone d'étude. L'oscillation de la méditerranée de l'ouest exprimée par l'indice de WeMO est corrélée avec la pluviosité de la saison (DJF) avec un décalage d'une saison, c-à-dire l'oscillation de pression en méditerranée pendant la saison d'automne (SON). En effet, la phase positive de WeMO correspond à une baisse de pluie au nord-ouest/augmentation de pluie au sud-est. L'indice d'oscillation méditerranéenne MO montre une corrélation significative positive avec l'axe CP1-1 avec un retard de deux saisons (JJA). Une phase positive de MO à la saison JJA correspond à des conditions plus humides dans toute la zone d'étude. L'indice SOI exprimant le mode de variabilité (oscillation australe El Nino) est positivement corrélé avec l'axe CP1-2 avec le décalage temporel le plus grand (3 saisons). L'indice SOI positif, correspondant à la situation climatique El niño. Cette phase (en MAM) correspond à des conditions plus sèches à la région nord-ouest/ou plus humide à la région sud-est de la zone, pendant la saison DJF de la même année, de manière que l'écart pluviométrique existant entre le nord et le sud de la zone d'étude diminue.

Les facteurs principaux CP2-1 et CP2-2 de l'analyse spatio-temporelle de la composante CP2, décrivant l'intermittence d'événements pluvieux, sont également exploités pour investiguer les liens avec les différents modes de variabilité climatique. Pendant la saison DJF, les indices d'oscillation nord atlantique (NAO) et de l'oscillation arctique (AO) sont positivement et significativement corrélés avec l'axe CP2-1. En effet, les phases positives de NAO et AO coïncident avec des événements pluvieux moins récurrents sur toute la zone d'étude (plus précisément au niveau des stations bien et assez bien représentées par l'axe CP2-1 de CP2 (Fig.V-9) pendant la saison DJF. Avec un décalage de trois mois, l'indice de AO et NAO sont négativement corrélés avec l'axe CP2-1, les phases négatives de NAO et AO au printemps (MAM) coïncident avec d'événements pluvieux plus fréquents à l'hiver de la même année. L'indice de WeMO, calculée en DJF et avec un décalage d'une saison (SON) et de deux saisons (JJA), est significativement corrélé avec l'axe CP2-2. Une phase positive de WeMO correspond à une augmentation d'occurrence des événements pluvieux au nord/ diminution d'occurrence des événements au sud la zone d'étude selon la direction des flux engendrés par cette oscillation climatique.

Le premier facteur CP3-1 de la composante CP3 ne montre aucune corrélation significative avec les indices climatiques.

Tableau V-4 Coefficients de corrélation entre les indices climatiques et les facteurs principaux retenus de CP1, CP2 et CP3 sur 50 ans d'observations. La couleur rose correspond à une corrélation significative au niveau de 95%

Coefficients de corrélation			CP1 : quantité de pluie		CP2 : intermittence de pluie		CP3 : conditions météorologiques
Indice climatique	Décalage (saison)	Saison de calcul de l'indice	CP1-1	CP1-2	CP2-1	CP2-2	CP3-1
NAO	0	DJF	-0.17	0.28 (p=0.04)	0.36 (p=0.008)	-0.13	-0.05
	-1	SON	0.07	0.19	-0.2	0.07	0.11
	-2	JJA	0.00	0.15	0.1	-0.21	-0.002
	-3	MAM	0.12	-0.07	-0.36 (p=0.008)	0.13	-0.002
WeMOI	0	DJF	-0.1	-0.03	-0.03	0.27 (p=0.05)	0.07
	-1	SON	-0.01	0.29 (p=0.03)	-0.036	0.27 (p=0.05)	0.07
	-2	JJA	-0.20	0.06	0.12	0.29 (p=0.03)	0.06
	-3	MAM	-0.02	0.04	-0.05	0.26	-0.22
SOI	0	DJF	-0.06	0.18	0.2	0.11	0.006
	-1	SON	-0.07	-0.03	0.02	-0.08	0.01
	-2	JJA	0.07	0.18	-0.05	0.11	0.17
	-3	MAM	-0.03	0.33 (p=0.02)	0.12	0.22	0.12
AO	0	DJF	-0.09	0.32 (p=0.02)	0.28 (p=0.04)	-0.2	-0.05
	-1	SON	0.16	0.09	0.02	-0.08	0.15
	-2	JJA	0.02	-0.03	0.12	-0.28 (p=0.04)	-0.02
	-3	MAM	0.17	0.04	-0.29 (p=0.03)	0.09	0.20
PDO	0	DJF	-0.02	0.33 (p=0.02)	-0.11	-0.22	-0.03
	-1	SON	-0.21	-0.08	-0.06	0.09	0.1
	-2	JJA	-0.04	0.06	0.11	-0.04	0.01
	-3	MAM	-0.09	-0.17	-0.01	-0.10	0.004
MO	0	DJF	0.23	0.02	-0.18	-0.023	0.08
	-1	SON	0.08	0.05	-0.06	0.09	-0.02
	-2	JJA	0.28 (p=0.04)	0.09	0.03	-0.017	0.019
	-3	MAM	0.09	-0.19	-0.08	-0.021	-0.04

En général, les indices climatiques sont majoritairement corrélés avec le deuxième facteur de CP1 et CP2 qui décrivent des écarts des apports pluviométriques et de l'intermittence entre les régions NO et SE. Par conséquent, les régions exposées aux flux engendrés par ces oscillations, pendant leurs différentes phases, sont plus sensibles aux modes de variabilité

climatiques. MO et WeMO sont des oscillations méditerranéennes et régionales qui n'ont pas le même potentiel. Bien que le mode de variabilité climatique MO montre une très bonne cohérence avec des séries pluviométriques mensuelles dans différentes régions de la Tunisie, spécialement la partie sud (Jemai et al., 2017), dans cette étude la quantité des précipitations et surtout l'occurrence d'évènement pluvieux sont plus corrélées avec la variabilité et l'alternance des deux phases de WeMO surtout pendant l'automne (SON). Le mode de variabilité le plus éloigné, l'ENSO (indice SOI, au sud du pacifique), est significativement corrélé avec un écart temporel considérable (3 saisons) avec CP1-2. Par contre, l'oscillation décennale du pacifique (indice PDO, au nord du Pacifique) est corrélé sans aucun décalage temporel avec le même axe. L'influence du mode de variabilité ENSO a été également prouvée dans l'étude Ouachani et al. (2013) avec les précipitations à l'échelle annuelle avec un décalage de 2 ans. Les oscillations de l'Atlantique représentées par AO et NAO ont des corrélations significatives sur plusieurs composantes (CP1-2, CP2-1 et CP2-2) en DJF et avec un décalage temporel de deux et trois saisons.

V.3.2 Analyse composite de SLP pour CP1, CP2 et CP3

L'analyse des séries chronologiques de CP1, CP2 et CP3 a synthétisé le mode de variabilité spatiale dominant (de précipitations hivernales) au cours du temps dans deux facteurs principaux. Les composantes de ces deux axes varient dans le temps en enregistrant des années (saisons DJF) de fortes valeurs (en valeur absolue). Un seuil de $\mu \pm 1\sigma$ (μ est la moyenne de la composante et σ est l'écart type), avec une hypothèse de linéarité, a été fixé pour chaque axe pour identifier les années avec des indices relativement forts qui correspondent à des situations pluviométriques particulières. Dans les figures [V-16a](#) et b, [V-17a](#) et b et [V-18a](#), les années avec un indice élevé sont représentées par des points rouges et les années de faibles indices sont représentées par des points bleus. Ces années enregistrent des anomalies spatio-temporelles significatives dans les différents aspects de précipitations (les composantes CP1, CP2 et CP3) et elles sont retenues pour une analyse composite. Cette analyse permet d'investiguer s'il existe des véritables liens entre la circulation atmosphérique représentée par la variable climatique SLP (la pression à la surface de la mer) pendant les années de fortes/faibles anomalies de précipitations. Les premiers facteurs CP1-1, CP2-1 et CP3-1 de CP1, CP2 et CP3 respectivement représentent un mode de variabilité commun sur toute la région. Par contre, les axes CP1-2 et CP2-2 représentent les modes de variabilité de précipitations par région et reflètent généralement un gradient entre la partie nord et sud de la zone d'étude (voir la section V.2.5).

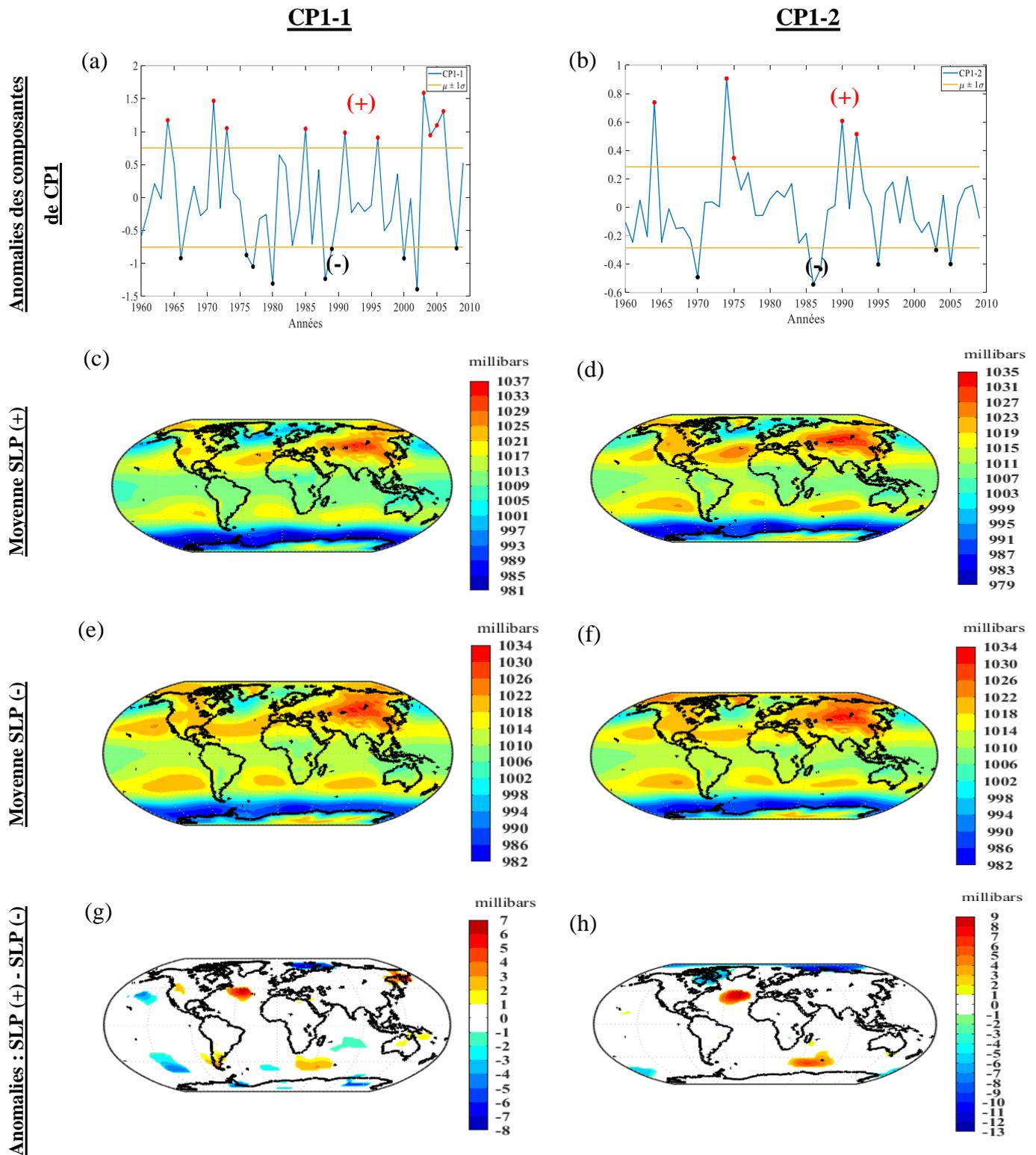


Figure V-16 Analyse composite de SLP et les facteurs principaux (CP1-1 et CP1-2) de CP1. (a) et (b) Identification des fortes (+) et des faibles (-) anomalies de CP1-1 et CP1-2. (c) et (d) SLP moyenne correspondant aux années de fortes anomalies de CP1-1 et CP1-2 noté SLP (+). (e) et (f) SLP moyenne correspondant aux années de faibles anomalies de CP1-1 et CP1-2 noté SLP (-). (g) et (f) les anomalies de SLP (SLP(+)-SLP(-)) de CP1-1 et CP1-2 (seulement les anomalies significatives à 90 %, selon le test de Student, sont visualisées)

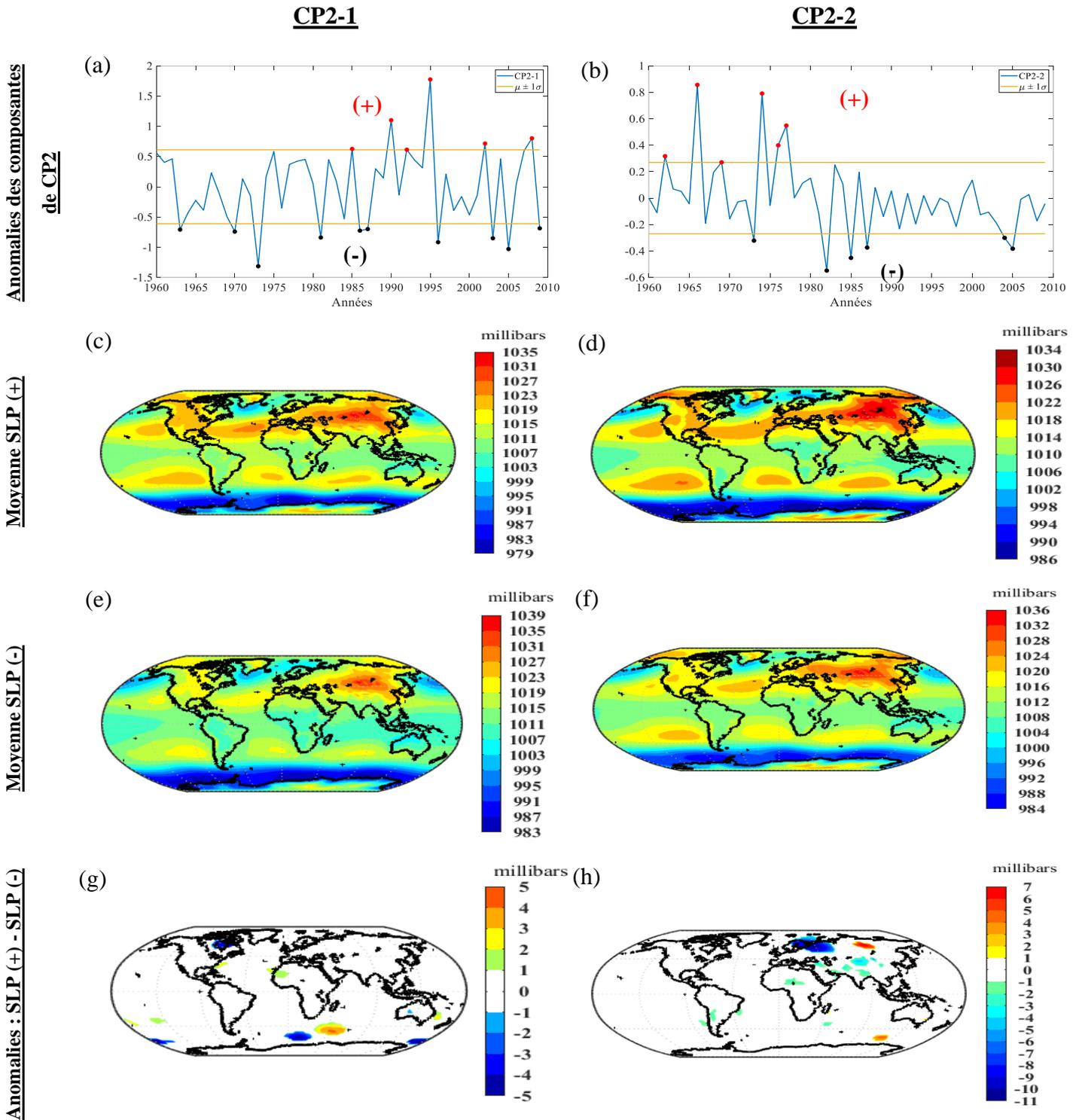


Figure V-17 Analyse composite de SLP et les facteurs principaux (CP2-1 et CP2-2) de CP2. (a) et (b) Identification des fortes (+) et des faibles (-) anomalies de CP2-1 et CP2-2. (c) et (d) SLP moyenne correspondant aux années de fortes anomalies de CP2-1 et CP2-2 noté SLP (+). (e) et (f) SLP moyenne correspondant aux années de faibles anomalies de CP2-1 et CP2-2 noté SLP (-). (g) et (f) les anomalies de SLP (SLP(+)-SLP(-)) de CP2-1 et CP2-2 (seulement les anomalies significatives à 90 %, selon le test de Student, sont visualisées)

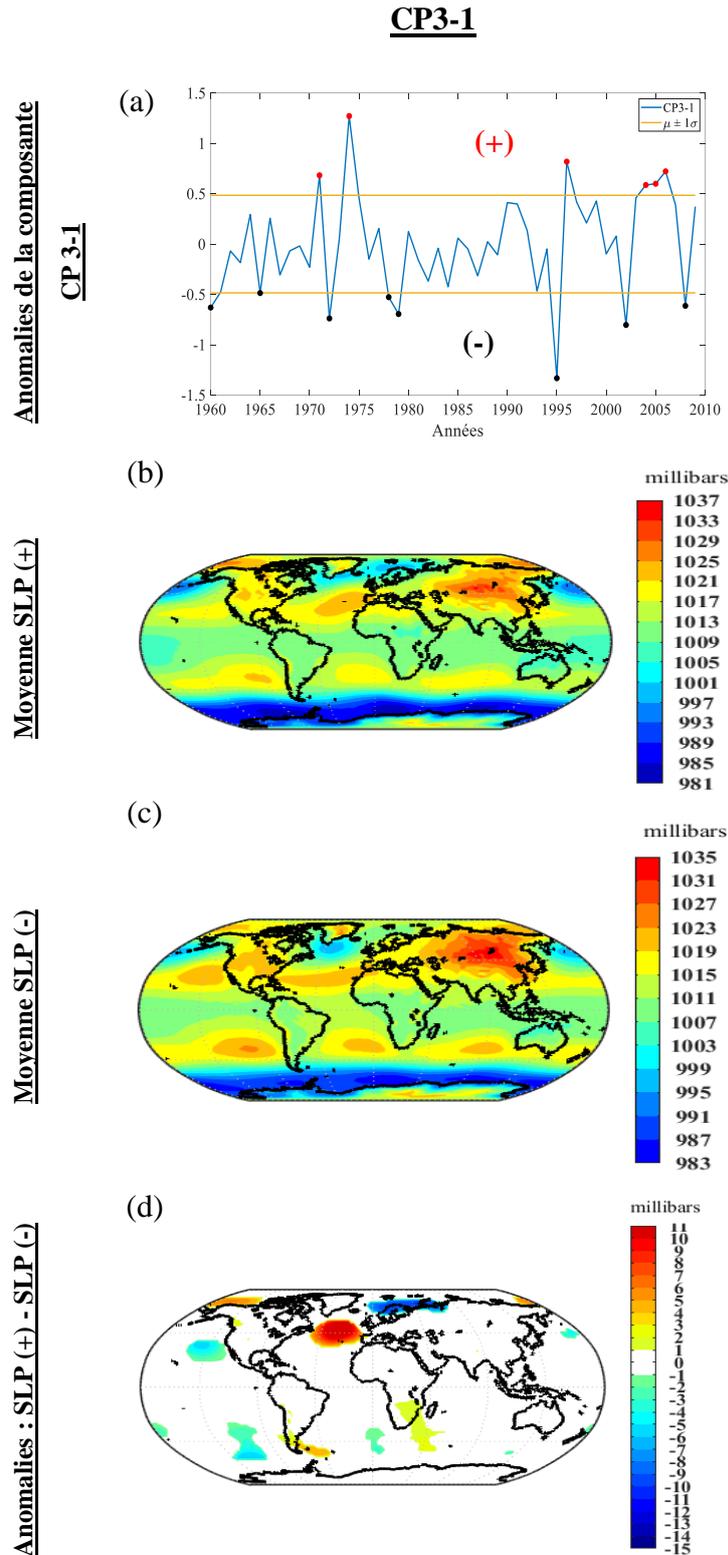


Figure V-18 Analyse composite de SLP et les composantes CP3-1. (a) les fortes et les faibles anomalies du premier facteur principal de CP3. (b) SLP moyenne correspondant aux années de fortes anomalies de CP3-1 noté SLP(+). (c) SLP moyenne correspondant aux années de faibles anomalies de CP3-1 respectivement noté SLP(-). (d) les anomalies de SLP (SLP(+)-SLP(-)) de CP3-1 (seulement les anomalies significatives à 90 %, selon le test de Student, sont visualisées)

V.3.2.1 Analyse composite de SLP avec les axes principaux de CP1 : quantité de pluie

Le facteur CP1-1 de la composante CP1 décrit la variabilité de la quantité de précipitations sur toute la zone d'étude. Les signes (+) et (-) reflètent les années humides et sèches respectivement (Fig.V-16a) selon un seuil fixé ($\mu \pm 1\sigma$) avec une hypothèse de linéarité. La différence composite de SLP (Fig.V-16g) montre une forte anomalie positive au niveau des Açores au nord de l'atlantique et forte anomalie négative dans la région subpolaire. Dans l'hémisphère sud, il existe des anomalies significatives moins élevées à l'océan antarctique glacial et au sud du pacifique. L'axe CP1-2 décrit l'écart de la quantité de précipitations entre le nord-ouest et le sud-est de la zone d'étude. Il enregistre 5 ans avec des valeurs positives avec une amplitude variable (Fig.V-16b). Durant ces années, les stations du nord-ouest enregistrent une chute remarquable au cumul de précipitations. Simultanément, l'anticyclone de la zone subtropicale se trouve plus prononcé, en général, et particulièrement l'anticyclone des Açores au nord Atlantique (Fig.V-16d). Les années avec des composantes négatives de CP1-2 coïncident avec une baisse de pression dans l'Atlantique ou l'anticyclone se positionne à un niveau latitudinal plus bas (Fig.V-16f). La différence composite de SLP dans la figure V-16h montre une forte anomalie positive au niveau des Açores et des fortes anomalies négatives dans la région de Groenland qui a une structure semblable à l'oscillation nord Atlantique NAO. Il existe également une anomalie positive significative à l'océan Antarctique glaciale.

V.3.2.2 Analyse composite de SLP avec les axes principaux de CP2 : Intermittence d'événements

La même démarche a été suivie pour extraire les modes de variabilité climatiques liés à la composante CP2, c-à-dire à l'intermittence des événements pluvieux. Les valeurs CP2-1 sont positives pour les années de faible intermittence d'événements et négative pour de forte intermittence (Fig.V-17a). Globalement, une baisse de pression (Fig.V-17e) engendre plus d'événements pluvieux pendant la saison DJF et contrairement (Fig.V-17c). La différence composite (Fig.V-17g) à un niveau de significativité de 90 % montre quelques anomalies significatives de faible étendue spatiale. L'axe CP2-2 décrit une différence d'intermittence entre le nord-ouest et sud-est de la zone d'étude. En effet, les valeurs positives de l'axe 2 reflètent un écart plus élevé traduit par plus d'événements pluvieux que la moyenne à la région nord-ouest et contrairement pour les valeurs négatives (Fig.V-17b). La différence composite (Fig.V-17h) montre une forte anomalie négative à l'Europe du nord qui correspond à moins d'événements au nord de la zone d'étude.

V.3.2.3 Analyse composite de SLP avec le premier axe de CP3 : conditions météorologiques

La troisième composante CP3 décrit la nature des événements pluvieux. En effet, dans l'axe CP3-1, les valeurs au-dessus du seuil $\mu + 1\sigma$ (Fig.V-18a), avec une hypothèse de linéarité, sont des saisons caractérisées par des événements pluvieux de longue durée et de faible intensité (pluies stratiformes) dans la plupart de la zone d'étude et contrairement pour les valeurs au-dessous du seuil $\mu - 1\sigma$ qui sont caractérisées par des événements pluvieux courts et intenses pendant la saison DJF. La différence composite de SLP à 90 % de confiance (Fig.V-18d) montre une structure bipolaire d'une anomalie positive au nord de l'Atlantique et une forte anomalie négative à l'ouest de la Russie. Cette structure des anomalies est semblable au mode de variabilité connu par l'Atlantique Est/Russie ouest (EA/WR). En outre, il existe des anomalies négatives moins importantes au Pacifique nord et sud.

V.4 Conclusion

L'analyse factorielle des données multidimensionnelles est une analyse multivariée qui sert à comprendre la structure des données, à extraire l'essentiel de l'information par des projections dans un espace de faible dimension (2-D ou 3-D) et à réduire et synthétiser la base des données. Cette analyse a permis également de vérifier la qualité des données d'observations, collectées des stations pluviométriques, en évaluant les corrélations entre les différentes stations et la cohérence avec les conditions bioclimatiques, topographiques et orographiques de la zone d'étude.

Cette analyse, qui s'appuie sur une base de données pluviométriques multidimensionnelle (6-D), a été effectuée à deux niveaux. En Premier niveau, l'ACP a synthétisé les 6 descripteurs de pluie en trois axes principaux : 1) l'apport pluviométrique CP1, 2) l'intermittence des pluies CP2 et 3) les types d'événements et leurs conditions météorologiques CP3. En deuxième niveau, une deuxième ACP a été réalisée sur les axes retenus et a synthétisé les modes de variabilité spatiales (des composantes) au cours de 50 ans d'observations.

L'interprétation spatiale a mis en évidence trois régions orientées NE-SO, avec une opposition entre la façade Nord-Ouest du domaine d'étude (Bizerte-Tabarka) et son arrière-pays et la façade Sud-Est (Nabeul-Hammamet) et son arrière-pays avec une zone intermédiaire géographiquement située entre ces deux régions. Cette répartition dépend des types des flux pluvieux et ses différentes origines. Elle dépend également des conditions locales de chaque pluviomètre. Temporellement, saisons les plus humides sont devenues de plus en plus fréquentes après les années 2000.

De plus les indices de variabilité climatiques au nombre de six ont été analysés par l'intermédiaire de coefficients de corrélation linéaire de Pearson avec les séries chronologiques des composantes principales. Cette analyse a donné, globalement, des faibles corrélations (faible pouvoir prédictif) avec les indices climatiques. Les corrélations statistiquement significatives existantes nous ont conduit aux interprétations suivantes : l'apport pluviométrique (CP1) dans la zone d'étude répond simultanément (même saison) aux oscillations de pression nord atlantique NAO, aux oscillations Arctique AO et aux oscillations au nord du Pacifique (PDO). Cependant, ces modes de variabilité climatiques ont montré des liens avec la zone d'étude par région. En effet, ils coïncident avec des conditions exceptionnelles soit au nord/nord-ouest, soit de la partie sud-est. Cela dépend des directions des flux engendrés par ces oscillations ainsi que de leurs amplitudes. Les apports de pluie pendant la saison DJF répondent par un retard temporel de trois mois à l'oscillation du sud de Pacifique ENSO, par un retard d'une et de deux saisons pour les oscillations régionales méditerranéennes MO et WeMO. L'occurrence des événements CP2 (intermittence) au cours de la saison hivernale a montré des corrélations significatives avec les oscillations de l'atlantique NAO et AO, et exceptionnellement avec l'oscillation régionale WeMO d'une manière synchrone et asynchrone. La composante CP3 des conditions météorologiques d'événements n'a pas montré des corrélations significatives avec les indices climatiques.

Comme une analyse complémentaire aux calculs des corrélations, une analyse composite de la variable climatique SLP (la pression à la surface de la mer) a été réalisée pour investiguer des liens avec les fluctuations de pressions non soulignées par les indices et pour valider également les corrélations trouvées. Cette analyse a mis en évidence les liens entre les apports pluviométriques et les anomalies de pressions au nord de l'Atlantique trouvés également par l'indice de NAO. La composante qui décrit les conditions météorologiques des événements pluvieux (fronts froids ou fronts chauds) a montré des anomalies significatives avec les fluctuations de pression (DJF) sur une période de 50 ans au nord de l'Atlantique, au nord du Pacifique et au niveau de l'océan Antarctique à l'hémisphère sud.

L'utilisation de la méthode factorielle dans l'analyse des données pluviométriques présente différentes limites :

- Parfois des difficultés dans l'interprétation des axes principaux.
- Ne détecte pas les structures complexes et non linéaires dans les données.
- Perte d'information importante, car seules les dépendances linéaires sont analysées.

Ainsi, l'algorithme de l'ACP n'accepte pas les valeurs manquantes. Dans cette étude, nous avons proposé d'imputer les valeurs manquantes par la valeur moyenne de la variable contenant les lacunes. Néanmoins, cette technique n'est fiable qu'avec un pourcentage de lacunes inférieur à 10 %. La fixation d'une telle limite de valeurs manquantes a empêché la diminution de la variance des données, mais cela a entraîné l'élimination de certaines stations pluviométriques de cette analyse. Il existe plusieurs méthodes d'imputation plus compliquées qui n'exigent pas un seuil des valeurs manquantes : des méthodes itératives et des méthodes basées sur des hypothèses statistiques, etc. Ces techniques possèdent des erreurs d'estimation qui varient de l'une à l'autre et leur efficacité dépend fortement de la climatologie de la zone d'étude, la densité du réseau de mesure et la position géographiques des stations. Dans cette analyse, nous avons choisi d'extraire des modes de variabilité pluviométriques en se basant sur les données mesurées par les pluviomètres, pour cela, nous avons toléré un faible pourcentage de lacune pour chaque station. Mais l'imputation des données pluviométriques et l'évaluation des différentes méthodes d'imputation pour la zone d'étude reste l'objet d'une étude ultérieure. Ce sujet est considéré d'une grande importance pour les études hydrologiques et la modélisation qui dépendent fortement de la bonne qualité et la continuité des séries chronologiques de pluie. Dans la partie qui suit, nous proposons une analyse des données pluviométriques multidimensionnelles par classification en utilisant une approche non linéaire.

Chapitre VI : Régionalisation spatiale et temporelle des descripteurs de pluie par classification et évaluation du modèle numérique de réanalyse ERA-Interim

VI.1	INTRODUCTION	108
VI.2	APPLICATION DE LA SOM SUR LES DESCRIPTEURS DE PLUIE OBSERVES ET CHOIX DES PARAMETRES D'APPRENTISSAGE DE LA CARTE	108
VI.3	CLASSIFICATION DES SAISONS DJF DANS LES STATIONS PLUVIOMETRIQUES.....	108
VI.3.1	<i>Classification des cartes auto-organisatrice SOM.....</i>	108
VI.3.2	<i>Classification hiérarchique et détermination du nombre de classes.....</i>	111
VI.3.3	<i>Identification et analyse des typologies des saisons DJF par CAH.....</i>	113
VI.3.4	<i>Variabilité spatio-temporelle des classes.....</i>	116
VI.4	ÉVALUATION DE MODELE DE REANALYSE ERA-INTERIM.....	120
VI.4.1	<i>Affectation des saisons DJF des mailles ERA-Interim sur la carte auto-organisatrice.....</i>	120
VI.4.2	<i>Comparaison de la variabilité temporelle des classes.....</i>	121
VI.4.3	<i>Comparaison de la variabilité spatiale des classes.....</i>	123
VI.5	CONCLUSION	126

VI.1 Introduction

Dans ce chapitre, on s'intéresse à classifier les saisons DJF dans la zone d'étude par une approche non linéaire basée sur la méthode SOM « Self-Organizing Map » et la classification hiérarchique. La classification des 3500 saisons DJF des stations pluviométriques selon leurs caractéristiques en un nombre réduit de classes permet d'identifier les situations pluviométriques dominantes sur 50 ans d'observations (1960-2009) ainsi que la typologie des saisons et d'événements pluvieux. L'évaluation des données simulées par le modèle de réanalyse s'appuie sur les classes et les régimes pluviométriques résultant de la classification des observations.

VI.2 Application de la SOM sur les descripteurs de pluie observés et choix des paramètres d'apprentissage de la carte

Cette analyse s'appuie une base de données multidimensionnelle où les descripteurs pluviométriques de la saison (voir Tableau III-1) sont les variables calculées sur une période de 50 ans à 70 stations (3500 individus). Un seul individu (saison/station) représente les 6 descripteurs d'une saison i à une station j (vecteur de six dimensions). Par exemple, la station S02 (Ain drahem) à l'année 2000, la station S13 (Thibar) à l'année 1960 sont deux individus de 3500. Les valeurs manquantes sont gérées automatiquement par l'algorithme de la SOM. Une carte de bonne qualité, avec des erreurs de quantification et topologique minimales, s'obtient après plusieurs lancements de l'algorithme en jouant sur les différents paramètres mentionnés dans le tableau VI-1.

Tableau VI-1 Paramètres adoptés pour l'entraînement de la carte

Paramètres	
Nombre de neurones	320
Dimension de la carte	(20 x 16)
Grille	Hexagonale
Fonction de voisinage	Gaussienne
[σ_{init} σ_{final}] dans la phase d'organisation	[8 3]
[σ_{init} σ_{final}] dans la phase de convergence	[3 0.5]
Nombre d'itérations : phase d'organisation	1000
Nombre d'itérations : phase de convergence	5000

VI.3 Classification des saisons DJF dans les stations pluviométriques

VI.3.1 Classification des cartes auto-organisatrice SOM

Après apprentissage, les individus sont affectés aux neurones de la carte. Comme il est illustré dans la figure VI-1b, chaque neurone représente un nombre d'individus à l'exception de 2 % de neurones qui sont vides, c-à-dire ne représentent aucune donnée de l'espace des individus. Généralement, les individus par neurone varient entre 0 et 20 avec une moyenne de 10

saison/station par neurone. La densité des i est indiquée par la taille des hexagones en bleue. Cette représentation est importante dans l'interprétation de la carte où il ne faut pas se focaliser sur les neurones qui ne représentent pas des données.

La U-matrice (ou la matrice de distance) (Fig.VI-1a, 1^{er} ligne et colonne) visualise la distance euclidienne entre les neurones adjacents. La couleur bleue reflète les faibles distances, d'où une zone homogène sur la carte, c-à-dire les neurones sont peu différents. Les neurones en vert et jaune représentent les valeurs élevées de distance et signifient que les vecteurs référents associés aux neurones sont aussi éloignés. Cette représentation permet de comprendre et interpréter la structure des données, de détecter la dissimilarité entre les observations ainsi que les données exceptionnelles. Dans notre cas, la carte U-matrice (Fig.VI-1a) présente en bas et à droite une région distincte reflétant des saisons avec un comportement pluviométrique significativement dissimilaire par rapport aux autres saisons.

La figure VI-1a représente également la projection des variables des prototypes sur la carte topologique. D'après la structure de ces variables, l'analyse non supervisée par l'algorithme SOM montre que la carte est organisée suivant plusieurs axes/direction :

- une diagonale allant de l'extrémité en haut à gauche (valeurs les plus faibles) à l'extrémité en bas à droite (valeurs les plus fortes) suivant les valeurs de C_t et $C_{m,e}$.
- un gradient de haut (valeurs faibles) en bas (valeurs fortes) suivant le N_j et la $D_{m,e}$.
- un gradient de gauche (valeurs faibles) à droite (valeurs fortes) suivant le cumul moyen par jour pluvieux $C_{m,j}$.

La variable nombre d'événements N_e (Fig.VI-1a, ligne 1, colonne 2) a une structure plus complexe, les saisons ayant un nombre d'évènements particulièrement important se situant à gauche et correspond à un $C_{m,e}$ faible (Fig.VI-1a, ligne 2, colonne 2). Pour ces saisons, la durée moyenne des évènements est également faible. Il s'agit des saisons caractérisées par des évènements récurrents, mais courts et peu intenses. Les saisons ayant un nombre d'évènements particulièrement faible se situent en haut et elles correspondent à un nombre de jours pluvieux N_j également faible. Pour ces saisons, le cumul saisonnier et le cumul par évènement sont faibles, la durée moyenne des évènements est faible également. Il s'agit des saisons les plus sèches avec peu d'évènements courts et peu intenses.

La similitude entre la structure du cumul saisonnier et du cumul moyen par évènement indique que la saison DJF est influencée par l'intensité des évènements pluvieux et pas par leur nombre. Les saisons ayant les nombres d'évènements les plus importants ne sont pas les saisons les plus pluvieuses. Les valeurs exceptionnelles du nombre de jours pluvieux N_j en bas de la carte

légèrement vers la droite (Fig.VI-1a, ligne1, colonne 3) correspondent à un cumul saisonnier faible, un cumul moyen par jour pluvieux très faible ainsi qu'à un nombre d'événements faible ayant une durée moyenne particulièrement importante. Ces neurones correspondent des saisons caractérisées par des pluies très abondantes avec un cumul journalier extrêmement petit. La variable cumul moyen par jour pluvieux $C_{m,j}$ (Fig. VI-1a, ligne 2, colonne 3) peut avoir des valeurs importantes dans des saisons sèches ou des saisons humides parce qu'elle dépend essentiellement de N_j . Par exemple, les saisons représentées par les neurones en haut à droite (Fig. VI-1a, ligne 2, colonne 3) dans la carte de variable $C_{m,j}$ sont des saisons sèches caractérisées par quelques événements courts et intenses. Cette partie de la carte représente des saisons où le cumul saisonnier est entre 100 et 300 mm réparti sur un nombre de jours pluvieux par saison très faible (inférieur à 10 jours par saison).

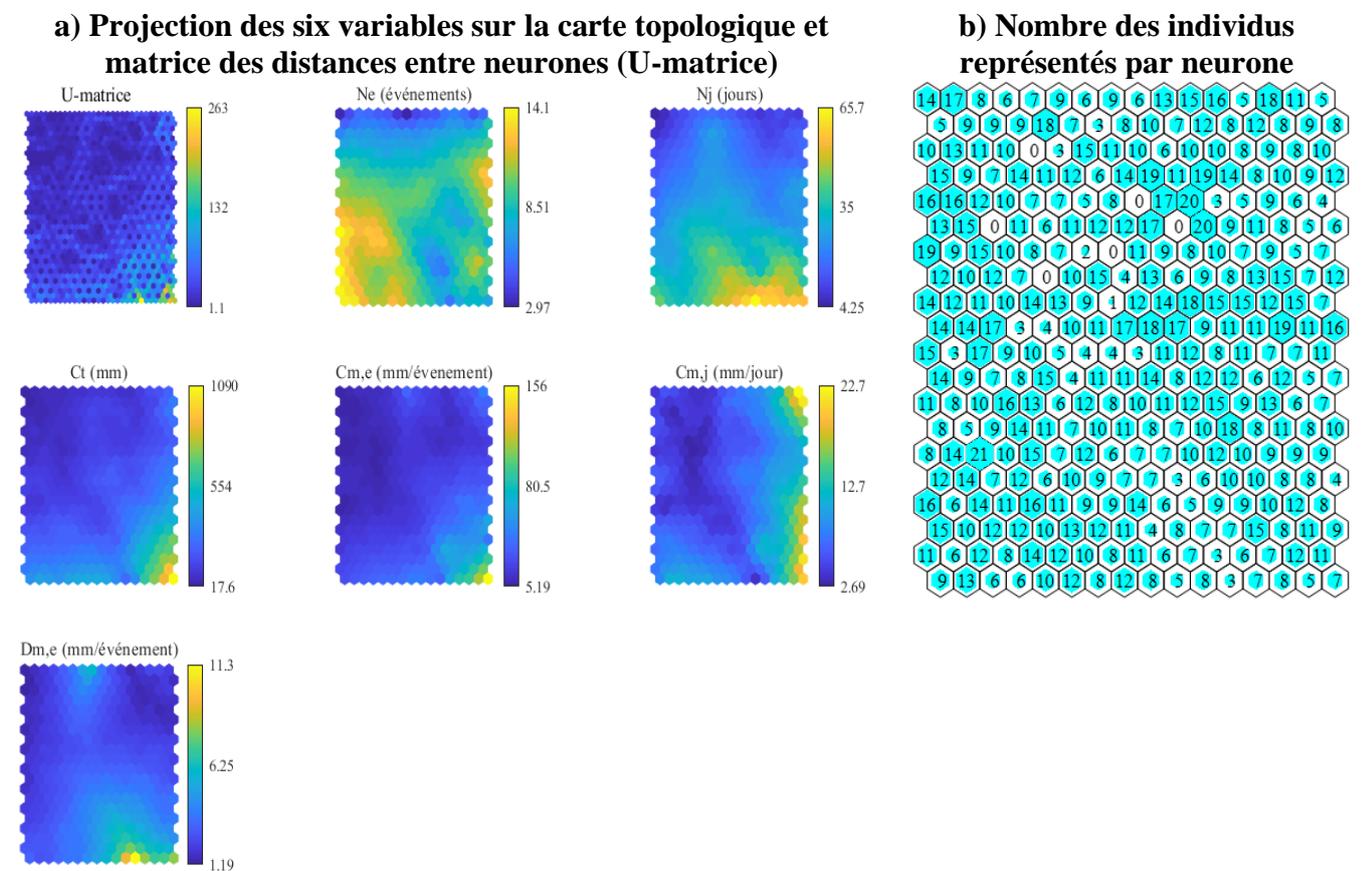


Figure VI-1 (a) Projection des six variables sur la carte topologique et matrice des distances entre neurones (U-matrix). (b) Nombre des individus (année/station) représentés par chaque neurone de la carte.

La projection et la visualisation des référents sur une carte de faible dimension assure une inspection visuelle des structures des données, mais le nombre de prototypes (des classes de la

SOM) reste élevé. Dans le but d'avoir des informations synthétiques, on regroupe les neurones représentant des données similaires sur la carte. Ce regroupement peut se faire d'une manière manuelle en utilisant la U-matrice (la matrice de distance des prototypes). Il s'agit de détecter les grandes distances séparant les zones homogènes (de faible distance). Ces zones de grande distance sont appelées des zones frontières. Mais l'utilisation de cette approche manuelle n'est pas toujours évidente, car les zones frontières ne se trouvant pas toujours claires. Pour assurer une bonne délimitation des classes et une meilleure partition des prototypes, une classification automatique est privilégiée.

VI.3.2 Classification hiérarchique et détermination du nombre de classes

La classification hiérarchique appliquée aux référents des neurones de la carte topologique permet de donner une vision plus globale à la carte en regroupant les neurones similaires. Le dendrogramme obtenu, basé sur l'algorithme de Ward, est présenté dans la figure VI-2a montre les liens entre les 320 vecteurs référents. L'axe vertical représente l'inertie ou la dissemblance entre les groupes fusionnés à chaque étape de l'algorithme. Différents nombres de classes sont obtenus en coupant l'arbre, et le nombre de classes correspond au nombre de branches coupées. Les lignes verte, noire et rouge montrent différents niveaux de coupe dans le dendrogramme et donnent respectivement 2, 3 et 4 classes.

La méthode data image est appliquée à la matrice de distance des 320 vecteurs référents. Dans la figure VI-2b, les couleurs font référence à la distance entre les vecteurs référents deux à deux. Les couleurs bleues et rouges (et plus lumineuses) représentent respectivement la proximité et l'éloignement (la distance). La figure VI-2b illustre la proximité et l'éloignement sur l'ensemble du domaine topologique. La ligne diagonale sur la même figure représente la valeur zéro. Maintenant, les valeurs/positions dans la matrice de distance peuvent être réordonnées après l'application d'un algorithme de classification, par exemple, la classification hiérarchique ici, et les différents groupes obtenus apparaissent sous forme de blocs sur la diagonale principale. Les Fig. VI-2c, Fig.VI-2d et Fig.VI-2e montrent la matrice de distance organisée en deux, trois et quatre classes respectivement. Les blocs observés le long de la diagonale de chaque image représentent les classes, et leur taille reflète la taille des classes. Les contrastes de couleurs à l'extérieur des blocs représentent les distances inter-classes, et donc plus le contraste est fort, plus la classification est bonne. Notez que dans ces trois figures, les axes ne sont pas étiquetés, car l'ordre est fixé par les classes.

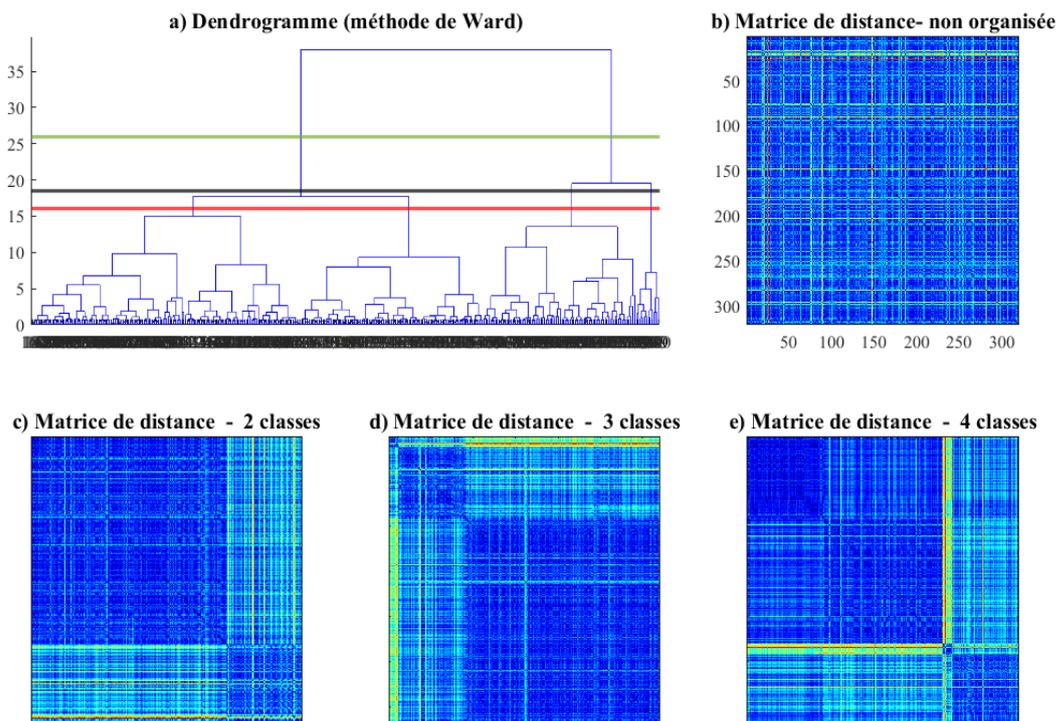


Figure VI-2 (a) Dendrogramme. Les traits horizontaux de couleur verte, noire et rouge sont les niveaux de coupure à 2,3 et 4 classes respectivement. (b) data image de matrice de distance entre les vecteurs référents deux à deux (c) matrice de distance réorganisée en deux classes, (d) matrice de distance réorganisée en trois classes, (e) matrice de distance réorganisée en quatre classes

La figure VI-2c montre deux classes, avec deux blocs le long de la diagonale principale, la plus grande classe est située dans le coin supérieur gauche. Cependant, pour la plus petite classe, une sorte d'hétérogénéité existe et peut être observée en comparant l'intensité de la couleur entre les deux blocs sur la diagonale. Cela signifie que cette classe n'est pas homogène et peut contenir des éléments qui pourraient appartenir à d'autres classes. Dans l'hypothèse des trois classes, la petite classe de la Fig. VI-2c est divisée en deux classes de taille différente (Fig. VI-2d). Dans l'hypothèse de quatre classes illustrées sur la figure VI-2e, la plus grande classe de la figure VI-2c est maintenant divisée en deux. Le long de la diagonale, les quatre classes se distinguent clairement même si le contraste entre les deux blocs supérieurs et leurs blocs hors diagonale n'est pas très fort. Maintenant, avec l'hypothèse d'une classe supplémentaire, le modèle à cinq classes (non illustré) aboutit à diviser la plus grande classe de la figure VI-2c en trois classes, mais ceux-ci ne sont plus facilement discernables. Avec un plus grand nombre de classes, la visibilité des limites des blocs est complètement perdue.

Ainsi, la combinaison du dendrogramme et la technique de « data image » favorise une partition en trois ou quatre clusters. Bien que le modèle à trois classes soit également probable, en se basant sur le bon contraste obtenu sur la Fig. VI-2d, nous concentrerons la discussion ici sur le

modèle à quatre classes afin de prendre en compte la nuance existante dans la grande classe de la Fig.VI-2c et pour assurer une analyse assez fine.

VI.3.3 Identification et analyse des typologies des saisons DJF par CAH

La classification des prototypes en quatre classes donne la partition des neurones représentée par la figure VI-3a. Les classes sont organisées sur la carte du haut vers le bas. Les caractéristiques de chaque classe peuvent être identifiées en délimitant les classes sur les cartes de projection des variables des prototypes (Fig.VI-3b).

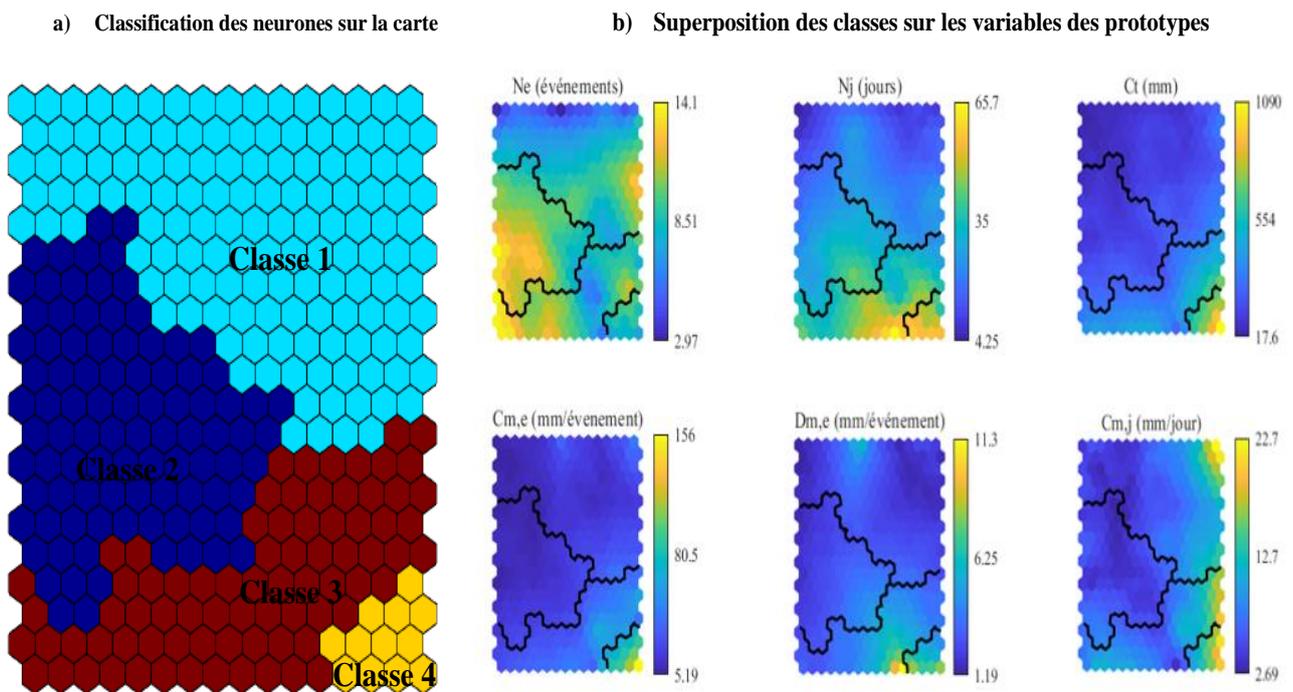


Figure VI-3 (a) classification des neurones sur la carte topologique. (b) Délimitation des classes sur les cartes de projection des variables de prototypes.

En identifiant l'intervalle des valeurs pour chaque variable sur la carte, les caractéristiques des classes sont synthétisées par des attributs qualitatifs (Tableau VI-2) : très faible, faible, moyen, fort, très fort. Cette étape préliminaire pour l'identification des classes fait ressortir différentes situations pluviométriques des conditions les plus faibles dans la classe 1 vers les conditions les plus humides dans la classe 4.

Tableau VI-2 Caractéristiques des quatre classes (TFa : très faible, Fa : faible, M : moyen, Fo : fort, TFo : très fort)

	Classe 1	Classe 2	Classe 3	Classe 4
N_e	TFa à Fo	Fo à TFo	M à TFo	Fa à M
N_j	Fa	Fa à M	M à TFo	Fo
C_t	TFa	TFa	M	TFo
C_{m,e}	TFa	TFa	Fa	M à TFo
D_{m,e}	Fa	TFa	M à TFo	M
C_{m,j}	TFa à TFo	TFa	M à Fo	M à TFo

Les histogrammes des six descripteurs pour chacune des quatre classes retenues sont illustrés dans la figure VI-4.

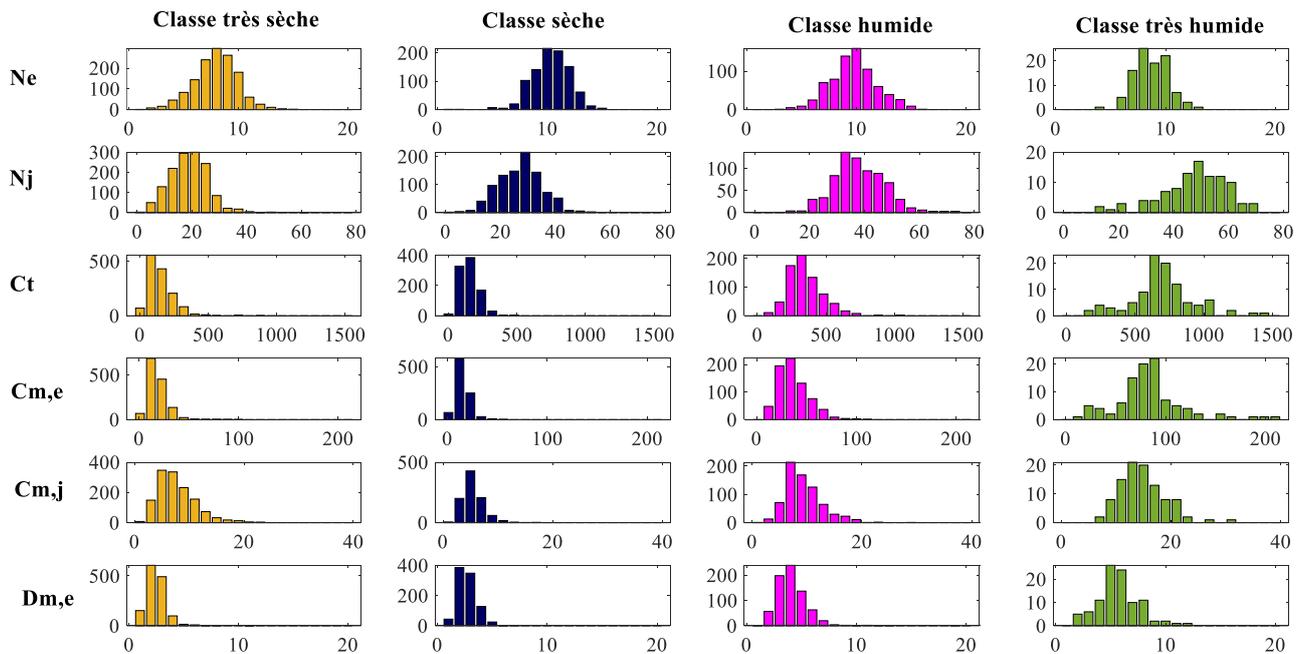


Figure VI-4 Etude de distribution des descripteurs étudiés pour les 4 classes (très sèche, sèche, humide et très humide).

La moyenne et l'écart type pour chaque descripteur sont également illustrés dans le tableau VI-3. Le nombre de saisons pour chaque classe est illustré dans le même tableau. La caractérisation des classes est basée sur l'observation des couleurs de neurone des classes dans la carte de chaque descripteur (Fig. VI-3b) combinée avec les distributions (Fig-VI-4) et les statistiques (Tab.VI-3). La significativité de différence entre les variables de chaque classe a été testé par un test de Student à un niveau de 95 %. Pour chaque descripteur, la différence entre les classes est testée significative à l'exception du cumul total C_t entre la classe 1 et la classe 2. Les résultats du test sont présentés dans l'annexe D.

Tableau VI-3 Moyenne et écart type de descripteurs pour les 4 classes

Statistiques des classes	Classe C1 (très sèche)		Classe C2(sèche)		Classe C3 (humide)		Classe C4 (très humide)	
	1378 saisons		933 saisons		727 saisons		99 saisons	
	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type
N_e (événements)	7.8	2.0	5.5	1.7	9.3	2.1	14.8	1.6
N_j (jours)	19.4	7.2	27.7	8.0	38.5	9.6	47.8	12.0
C_t (mm)	150.7	92.8	151.7	69.2	349.4	129.2	698.9	236.4
$C_{m,e}$ (mm/événement)	18.4	9.8	14.9	7.3	37.0	15.3	83.3	34.8
$D_{m,e}$ (jours/événement)	2.5	0.9	2.7	0.8	4.1	1.5	5.7	1.9
$C_{m,j}$ (mm/jours pluvieux)	7.8	3.7	5.5	2.3	9.3	3.4	14.8	4.3

La classe C1 représente la situation dominante qui contient 47.4 % des cas. Elle occupe tout le haut de la carte (Fig.VI-3a). Elle est caractérisée par un nombre total de jours de pluie N_j faible à moyen, des cumuls saisonniers faibles en majorité entre 100 et 200 mm (Fig.VI-4) et un cumul moyen par évènements faibles également. Cette classe représente des saisons sèches avec des évènements très courts et peu intenses.

La classe 2 qui représente 20.4 % des saisons (Fig.VI-3a) est à gauche et au centre sur l'axe vertical de la carte. Elle est caractérisée par un cumul moyen par jour pluvieux $C_{m,j}$ faible, un nombre total de jours de pluie moyen N_j et une durée moyenne d'événement $D_{m,e}$ faibles. Elle est caractérisée également par des cumuls saisonniers et des cumuls moyens par événement faibles. Par contre, le nombre d'événements N_e est relativement élevé. En général, cette classe a des statistiques similaires (moyenne et écart-type dans Tab.V3) avec la classe 1 avec une différence dans la distribution de quelques descripteurs. En effet, la classe 2 possède un mode plus élevé pour le N_e (10 événements), N_j (30 jours) et C_t (200 mm) qui correspondent respectivement à (7 événements), (20 jours) et (100 mm) dans la classe 1. Cette classe est représentative des saisons sèches, avec des faibles événements fréquents.

La classe 3 qui représente 24.9 % est située entre la classe 4 et les deux classes de saisons sèches (Fig.VI-3a). Les valeurs moyennes des descripteurs (Tab. VI-3) montre qu'elle représente des situations intermédiaires avec une quantité de pluie moyenne et avec une forte intermittence d'événements. Cette classe est caractérisée par un cumul saisonnier C_t moyen et relativement fort entre 200 et 500 mm. Le cumul moyen par événement $C_{m,e}$ est généralement entre 20 et 60 mm/événement qui sont des valeurs moyennes, voire fortes en les comparant avec la classe 1 et 2 (Fig.VI-4). Les valeurs de cumul moyen par jour pluvieux $C_{m,j}$ sont

distribuées entre les faibles, les moyennes et les fortes valeurs. Le nombre d'événements est élevé entre 8 et 12 et il atteint 15, ce qui montre une forte intermittence de pluie pendant les saisons de cette classe. Le nombre de jours pluvieux N_j dans cette classe est également élevé entre 30 et 60 jours et la durée moyenne des événements $D_{m,e}$ est globalement moyenne.

La classe 4 qui contient 7.2 % des saisons est située en bas à droite de la carte topologique (Fig.VI-3a). Elle est clairement caractérisée par des cumuls saisonniers et cumuls moyens par événement importants. La majorité des saisons de cette classe enregistre des pluies saisonnières entre 500 et 1000 mm/saison et le C_t a dépassé le 1300 mm dans certaines saisons (Fig.VI-4). Elle est aussi caractérisée par un nombre d'événements moyen, mais un nombre de jours pluvieux N_j^i important et une durée moyenne d'événement forte (supérieur à 4 jours/événement). L'intensité moyenne par jours pluvieux est également forte (entre 10 et 20 mm/jours) et elle atteint 30 mm/jour. Cette classe est la moins fréquente et elle est représentative des saisons extrêmement pluvieuses avec des longs événements intenses.

VI.3.4 Variabilité spatio-temporelle des classes

VI.3.4.1 Variabilité spatiale des classes dans les stations pluviométriques

La figure VI-5 représente la répartition spatiale des classes dans la zone d'étude. Pour chaque station pluviométrique, le graphique en camembert représente le pourcentage de chaque classe sur 50 ans d'observation. Pour classifier les différentes régions de la zone d'étude, nous avons considéré que les stations dominées par les classes humide et très humide (pourcentage de la classe humide + pourcentage de la classe très humide dépasse 50 %) sont des stations humides. De même, les stations dominées par les classes très sèche et sèche (pourcentage de la classe très sèche + pourcentage de la classe sèche dépasse 50 %) sont considérées sèches. En se basant sur ce critère, on distingue deux régions : (1) région humide au nord-ouest contenant 18 stations et (2) une région relativement sèche de 52 stations réparties sur le reste de la zone d'étude (nord-est et sud (Fig.VI-5). L'occurrence des classes 3 et 4 au nord-ouest est de l'ordre de 68 %. Par contre, l'occurrence de ces deux classes ne dépasse pas le 11% pour le reste de la zone d'étude. La classe 4 représentant des saisons avec des événements très intenses qui apparaissent très rarement dans le secteur sud sur 50 ans d'observations.

La station S02 (Ain Drahem) est une station très particulière dans la région nord-ouest, car l'occurrence des saisons DJF de la classe 4 dépasse le 50 % entre 1960 et 2009. Dans la partie sud, il existe des stations exceptionnellement sèches, comme S48 (Ain Ghésil) et S57 (Rohia), ou la quasi-totalité des saisons sont de la classe 1 représentant les conditions les plus sèches.

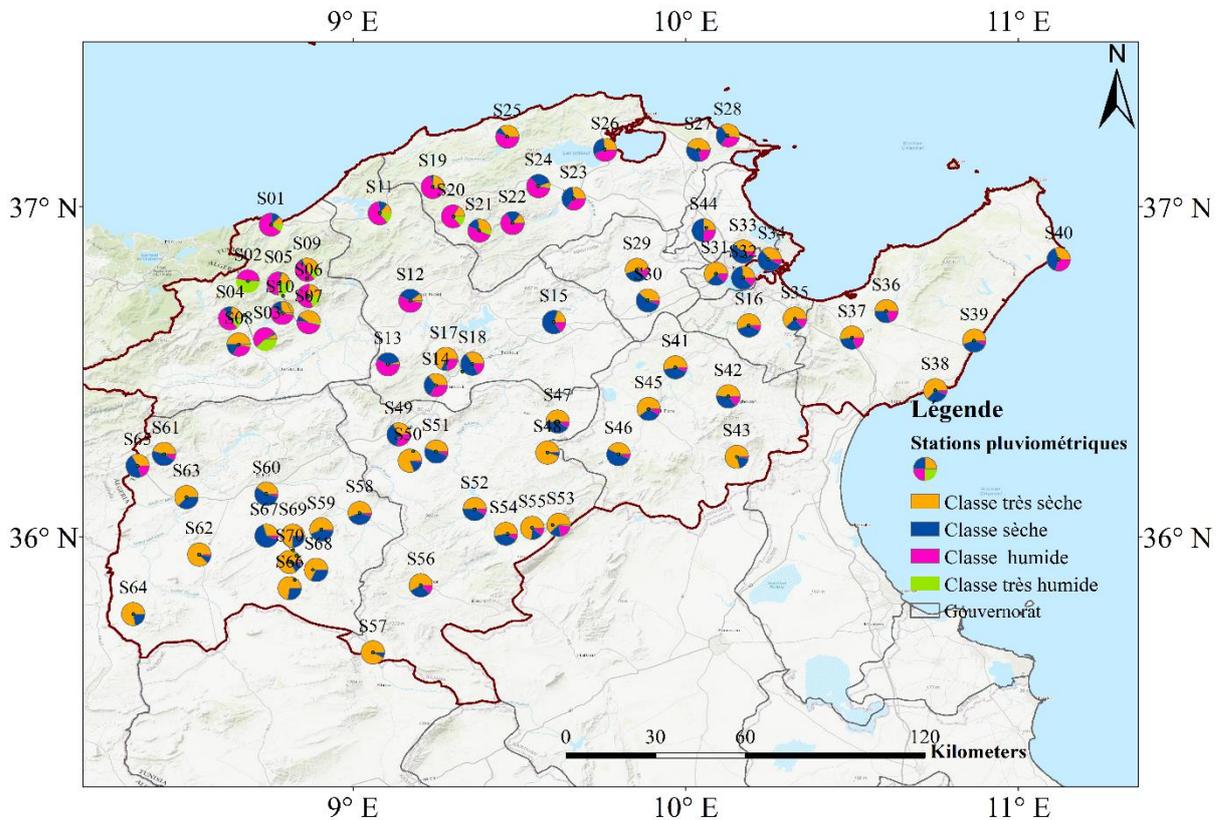


Figure VI-5 Distribution des quatre classes (très sèche , sèche, humide et très humide) dans les 70 stations pluviométriques par des graphiques en camembert .

La différence entre le régime pluviométrique des deux zones peut être expliquée par les conditions orographiques. En effet, le nord de la Tunisie est situé à l'extension des massifs montagneux de l'Atlas avec une orientation sud-ouest/nord-est comme indiqué dans la figure II-3. L'Atlas Tellien s'étend sur la partie septentrionale de la zone d'étude et forme une barrière contre les flux d'ouest et particulièrement les flux du nord-ouest qui sont à l'origine des précipitations hivernales. Donc les montagnes existant (comme Mogods et Kroumirie (Fig.II-3) mettent les stations de la zone sud dans une ombre pluviométrique. Les stations humides sont majoritairement situées au versant exposé au flux de l'Atlas Tellien. Ainsi la proximité de ces stations à la Méditerranée, qui est une source d'humidité, agit sur le régime pluviométrique de toute la zone en favorisant des événements pluvieux plus fréquents et plus intenses.

La continentalité de la partie sud-ouest de la zone d'étude combiné avec sa position entre l'Atlas Tellien et l'Atlas Saharien (Fig.II-3) donne les conditions les plus sèches à cette région à l'exception de d'une variabilité locale due aux fortes altitudes pour quelque stations comme les stations S65 (Sakiet Sidi Youssef, 803 m), S66 (Ain Zeligua, 853 m), S53 (Saadia Bargou, 524 m). Les stations localisées à l'est du secteur sud montrent plus de variabilité comparant à la partie ouest. Cette région est exposée aux flux provenant du continent européen qui apportent une quantité modérée de précipitations.

VI.3.4.2 Variabilité temporelle des classes

Le tableau VI-4 illustre la distribution des 4 classes obtenues sur la période d'étude de 1960-2009 et sur les sous périodes. Les pourcentages d'apparition de la classe très sèches et la classe très humides (les classes extrêmes) augmentent considérablement durant la sous période la plus récentes de 1991-2009 en les comparants avec les deux classes intermédiaires (sèches et humides).

Tableau VI-4 Pourcentage des quatre classes sur la période de 50 ans (1960-2009) et sur les sous périodes (1960-1990 et 1991-2009)

Pourcentage d'occurrence de chaque classe (%)	Classe C1 : Très sèche Saisons très sèches, peu d'évènements	Classe C2 : Sèche Saisons sèches, évènements fréquents, faible intensité.	Classe C3 : Humide Saisons moyennes, évènements fréquents,	Classe C4 : Très humide Saisons très pluvieuses, événement long et intense
Période entière (1960-2009)	43.9	29,7	23.2	3.1
Sous période 1 (1960-1990)	40.8	33.7	22.6	2.8
Sous période 2 (1991-2009)	48	24	24.3	3.5

A une échelle interannuelle, la distribution des quatre classes par année (saison DJF) est illustrée dans la figure VI-6. Chaque barre représente la distribution des classes dans les 70 stations. Globalement, on remarque une dominance des classes 1 et 2 à l'exception de quelques années comme : 1964, 1971, 1973, 2003, 2005 et 2006. En considérant que les années dominées par classe 3 et 4 (pourcentage en classe humide + pourcentage en classe très humide > 50 %) des années humides sur toute la région, on trouve que ces années ne présentent que 12 % de la période d'étude (50 ans). Dans le cas contraire, il existe des années avec des saisons DJF extrêmement sèches comme 1976, 1977 et 1990 où le pourcentage des classes humides est 0 %.

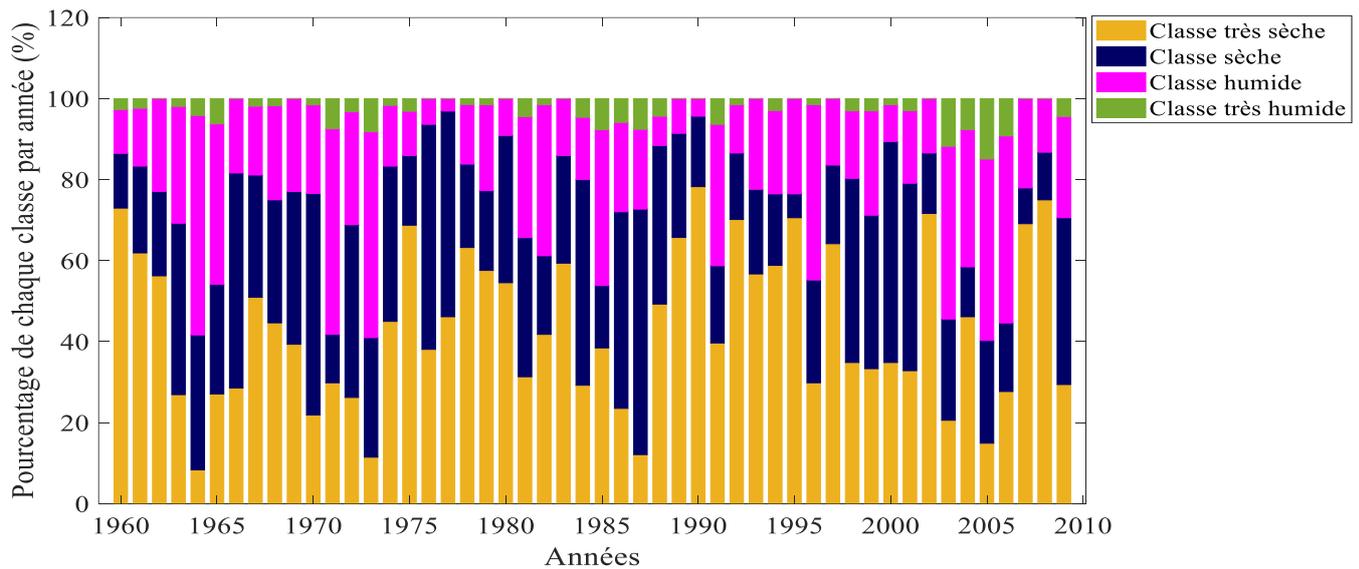


Figure VI-6 Distribution des quatre classes par année (dans les 70 stations).

L'étude de l'évolution temporelle des classes est élaborée également par périodes de 10 ans non chevauchantes (5 décades) sur les deux zones géographiques séparément. La région humide du nord-ouest (18 stations) et la région relativement sèche du nord-est et du sud de la zone d'étude (52 stations). La figure VI-7 représente les anomalies de pourcentage d'occurrence de chacune des classes sur 5 décades non chevauchantes dans toute la zone d'étude (ligne 1). Cette anomalie est l'écart entre le pourcentage d'occurrence de la classe C_i ($i=1,2,3$ et 4) par la décade et le pourcentage d'occurrence de la même classe sur 50 ans d'observations. Ainsi, la figure VI-7 illustre les anomalies d'occurrence de chaque classe dans la région humide du nord-ouest (ligne 2) et dans la région sèche (ligne 3) séparément.

Dans la région humide du nord-ouest (FigVI-7, ligne 2, colonne1), l'anomalie du pourcentage d'occurrence de la classe très sèche est plutôt positive et elle persiste sur 4 décades successives commençant des années 1970 aux année 2000. Cette même classe a enregistré une anomalie de pourcentage d'occurrence très forte dans toute la zone d'étude (FigVI-7, 1 ligne 1, 2 et 3, colonne 1) pendant les années 1990.

L'anomalie de pourcentage d'occurrence de la classe humide dans la région humide du nord-ouest reste négative sur quatre décades successives commençant des années 1970 à la décade 2000-2009 (Fig.VI-7, ligne 2, colonne 3). Au niveau de cette même région, l'anomalie positive la plus forte est enregistré pendant la décade 2000-2009 (Fig.VI-7, ligne 2, colonne 4).

Dans la région nord-ouest, les classes qui représentent les situations extrêmes, spécialement, les conditions les plus sèches tendent à augmenter au cours du temps (à l'échelle de décennies),

et les conditions les plus humides sont devenues plus fréquentes durant les années les plus récentes.

Pour le reste de la zone d'étude, il n'existe presque pas une anomalie, du même signe, qui persiste au cours des décennies à part l'anomalie de la classe très humide qui reste négatives et faible pour 4 décades successives (Fig VI-7, ligne 3, colonne 4).

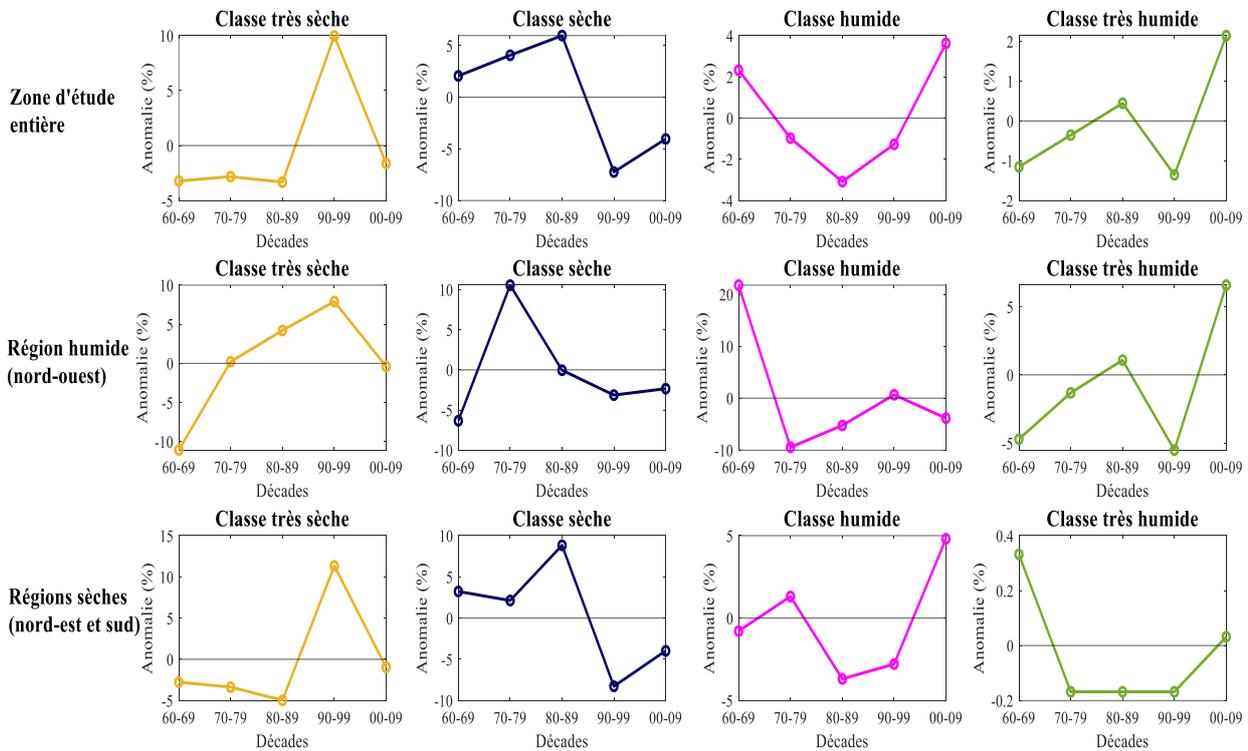


Figure VI-7 Anomalies de pourcentage d'occurrence de chaque classe sur 5 décades non chevauchantes (1960-1969 ; 1970-1979 ; 1980-1989 ; 1990-1999 ; 2000-2009) . (b) dans la région humide du nord-ouest.(c) dans la région sèche (nord-est et sud)

VI.4 Évaluation de modèle de réanalyse Era-Interim

VI.4.1 Affectation des saisons DJF des mailles ERA-Interim sur la carte auto-organisatrice

Après avoir classifié les saisons et caractériser chaque classe, la carte obtenue est utilisée dans cette partie comme référence pour évaluer la capacité du modèle à reproduire les caractéristiques des saisons ainsi que leur variabilité spatio-temporelle. Comme il a été montré dans la section III.3.2, les descripteurs observés et estimés ne sont pas du même ordre de grandeur. Alors le centrage et la réduction des deux sources des données permettent de comparer leurs évolutions spatiale et temporelle. Les vecteurs des descripteurs centrés réduits du modèle de réanalyse ERA-Interim (30 saisons x 9 mailles = 270 saisons) sont affectés aux neurones les plus proches de la carte obtenue dans la figure (Fig.VI-1) de l'apprentissage des données d'observations dans les stations. Il s'agit de calculer une distance entre les vecteurs

des saisons (6-D) ERA-Interim et les vecteurs prototypes liés aux neurones. Cette affectation permet par la suite d'identifier la classe à laquelle appartient chaque saison estimée par le modèle. La figure VI-8 représente la répartition des saisons de ERA-Interim sur la carte qui sont réparties comme suit : **33.3 %** des saisons appartiennent à la classe très sèche, **37 %** à la classe sèche, **27.4 %** à la classe humide et **2.6 %** à la classe très humide.

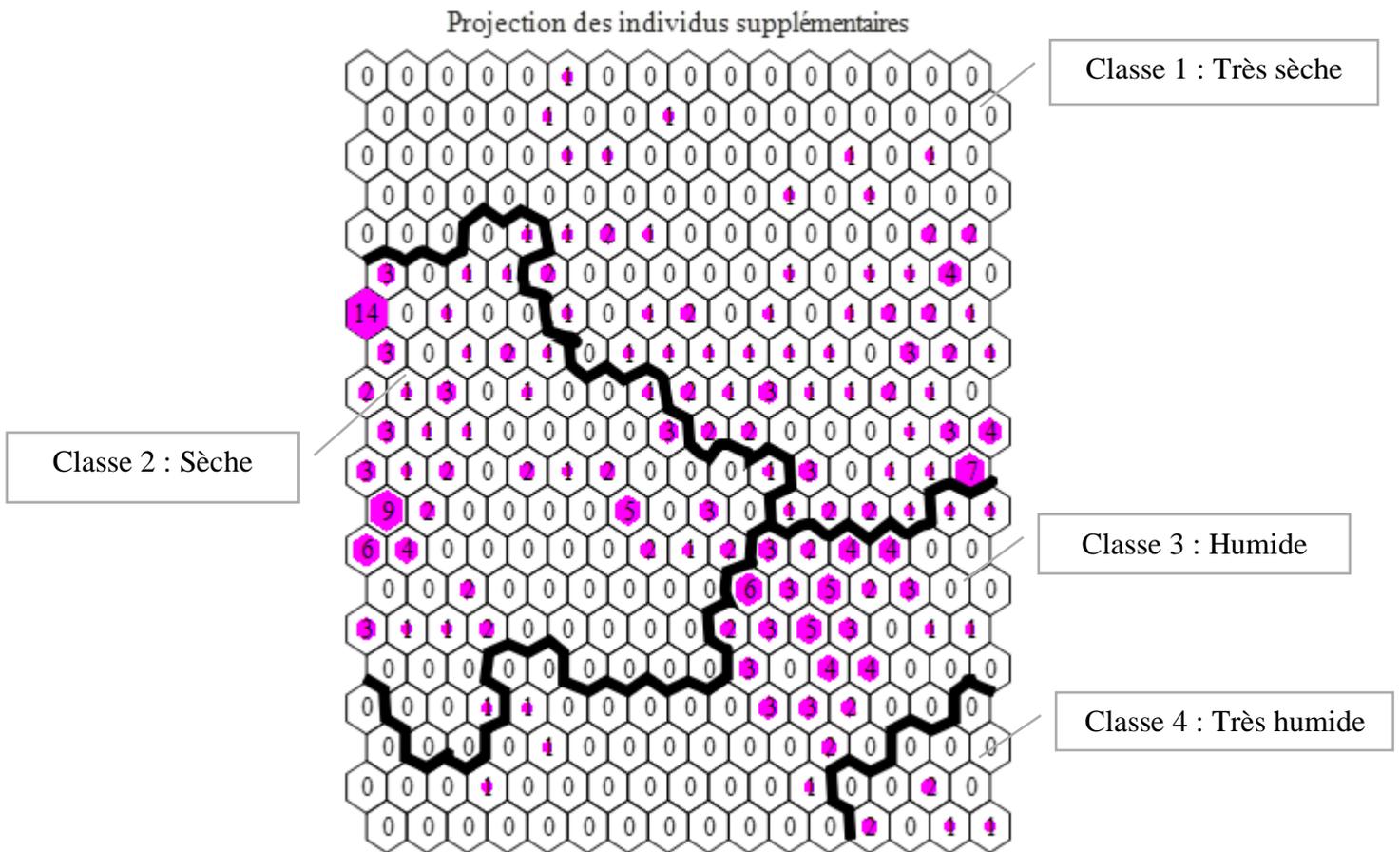


Figure VI-8 Projection des saisons ERA-Interim (individus supplémentaires) sur la carte topologique.
La ligne noire en gras indique les limites des 4 classes

Sur une période de 30 ans, presque 70 % des saisons DJF simulées par le modèle sont affectées aux classes très sèche et sèche.

VI.4.2 Comparaison de la variabilité temporelle des classes

À une échelle interannuelle entre 1980 et 2009, les classes de très sèche au très humide sont réparties comme indique la figure VI-9. Il existe presque 27 % des années humides sur la période de 30 ans (1980-2009) au cours desquelles les pourcentages des classes humide + classe très humide dépassent 50 %. Également, on distingue des années avec saisons DJF extrêmement sèches comme celles des années 1994, 1997, 1998, 2000 et 2002.

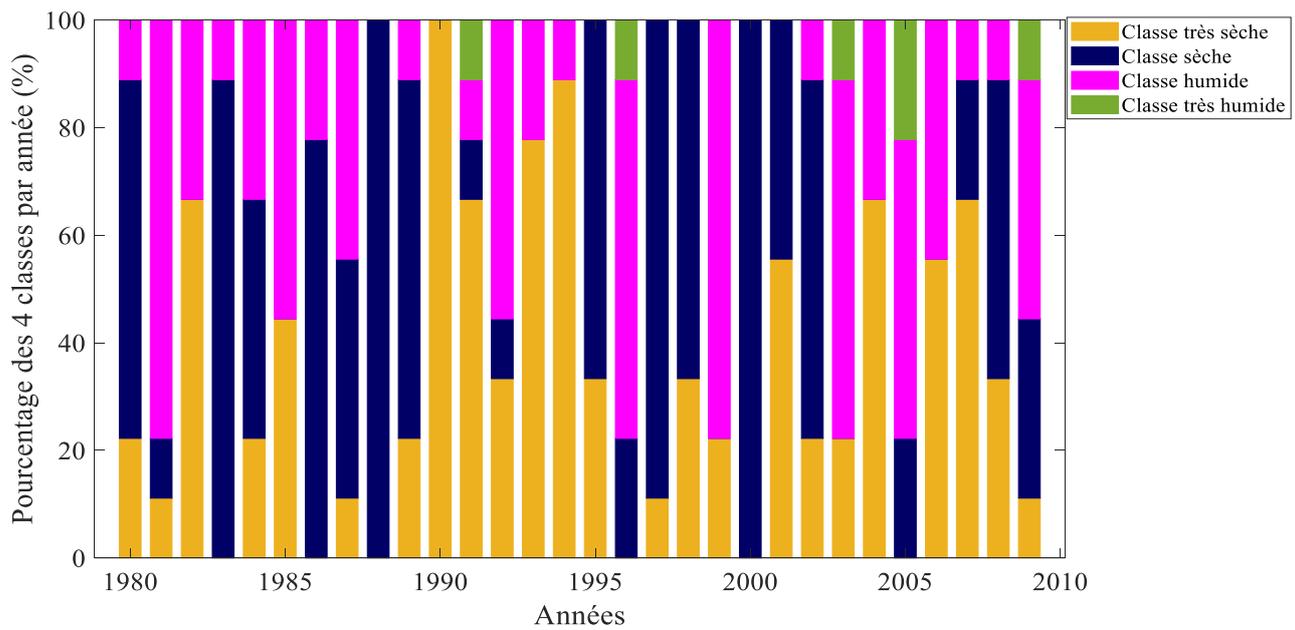


Figure VI-9 Répartition des 4 classes pour chaque année (1980-2009)

À une échelle décennale, la figure VI-10 illustre la variabilité temporelle des anomalies d'occurrence des 4 classes dans le modèle de réanalyse ERA-Intérim sur trois décades de 1980 à 2009 et des données des observations sur 5 décades non chevauchantes de 1960-2009. Les anomalies de pourcentage d'occurrence est l'écart entre l'occurrence de chaque classe à chaque décade pourcentage et l'occurrence à la période entière (50 ans pour l'observations et 30 ans pour les réanalyses). En comparant la période commune de trois décades (1980-1989 ; 1990-1999 ; 2000-2009) entre les deux sources des données, le modèle de réanalyse donne une variabilité temporelle d'anomalies relativement cohérente avec celle des observations pour les deux décades 1990-1999 et 2000-2009 (Fig.VI-10). Par contre, l'écart entre les anomalies des observations et le modèle de réanalyse est plus important pour la décade 1980-1989 qui est environ 15% pour la classe très sèche, 8 % pour la classe sèche, 6% pour la classe humide et 2% pour la classe très humide (environ 8% en moyenne). La structure temporelle des anomalies est conservée par le modèle ERA-Interim surtout pour les classes des saisons très sèches (Fig.VI-10a) et des saisons sèches (Fig.VI-10b) où les courbes des observations et de modèle ont la même allure. Pour les classes des saisons humides (Fig.VI-10c) et des saisons très humides (Fig.VI-10d), il existe une opposition des anomalies pendant la décade 1980-1989, c-à-dire une surestimation d'occurrence des saisons humides et une sous-estimation de d'occurrence de la classe très humide. La décade 2000-2009 a la meilleure représentation des typologies des saisons par le modèle de réanalyse (1.1% d'écart en moyenne) suivi par la décade 1990-1999 d'environ 2.6 % d'écart moyen.

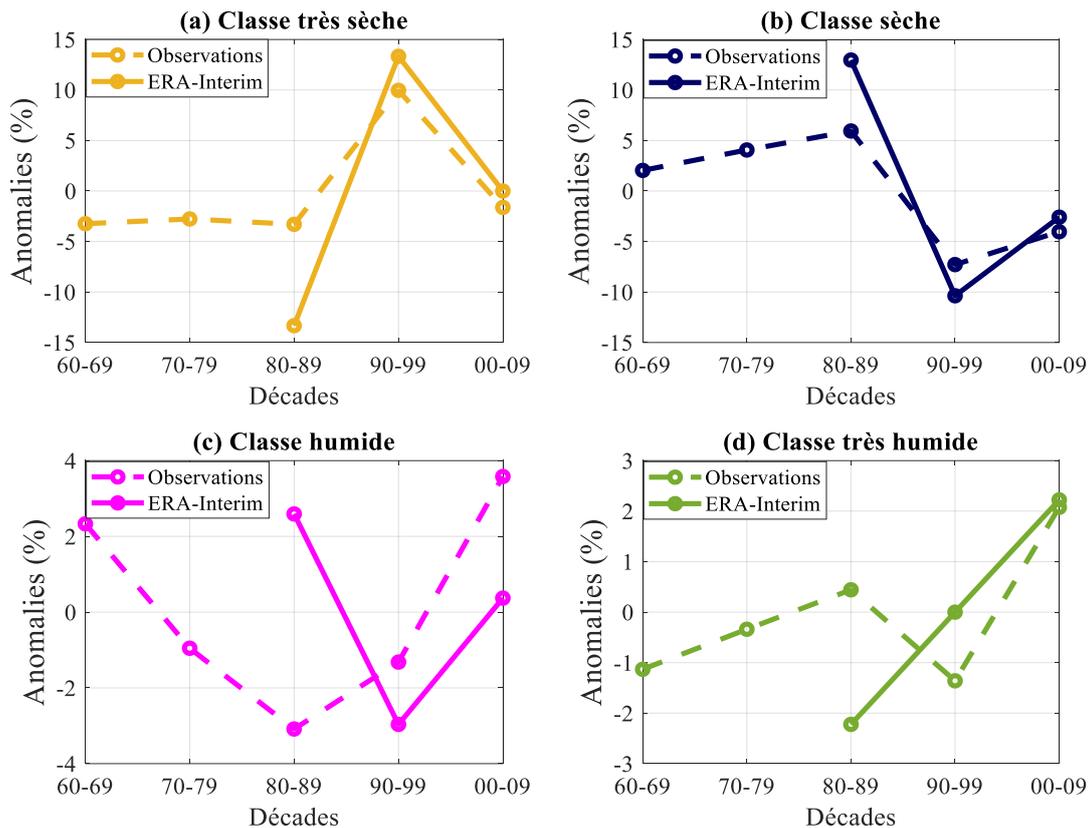


Figure VI-10 Variabilité décennale des anomalies d'occurrence de la classe 1(très sèche) (a), classe 2 (sèche) (b), classe 3 (humide) (c) et classe 4 (très humide)(d) pour les données de réanalyse ERA-interim et pour les observations

VI.4.3 Comparaison de la variabilité spatiale des classes

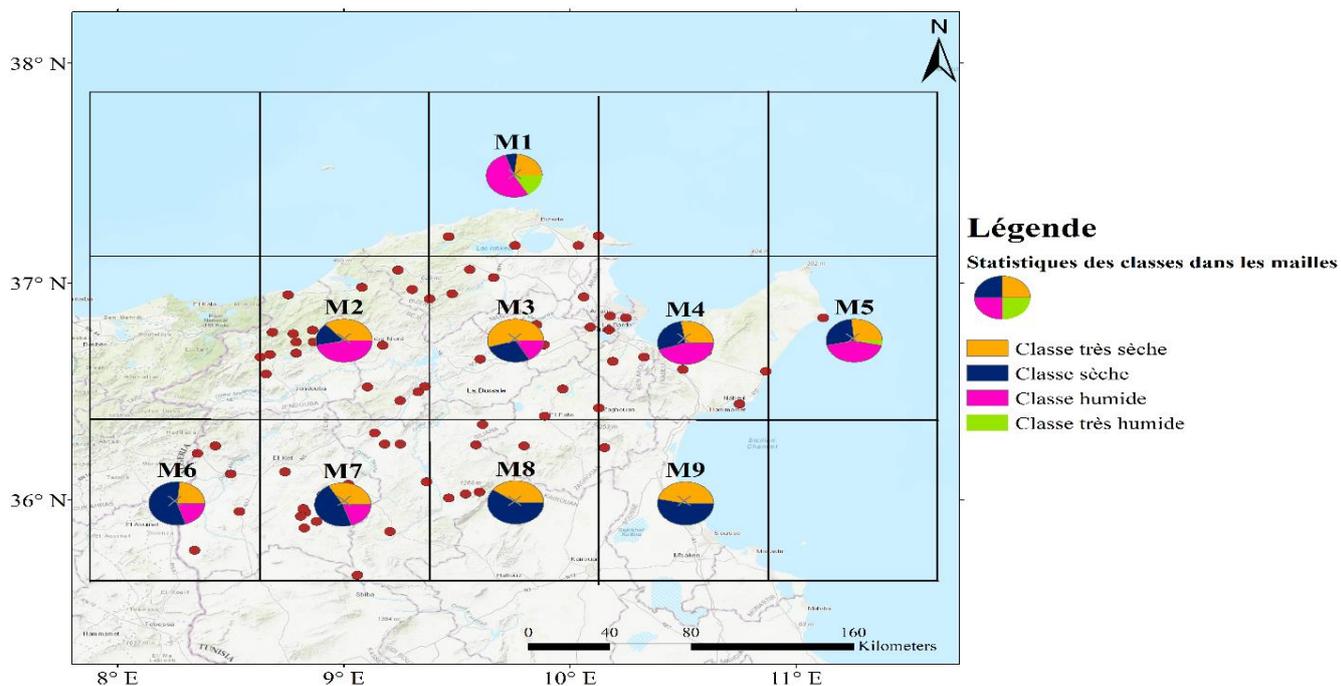
La figure VI-11a représente la répartition des classes dans les 9 mailles représentant le nord de la Tunisie. Chaque camembert décrit les pourcentages d'occurrence des quatre classes par maille. En général, les classes très sèches et sèches deviennent de plus en plus fréquentes en allant vers le sud. La classe très humide décrivant des événements intenses et longs apparaît dans la maille M1, M4 et M5. La caractéristique commune entre ces trois mailles, c'est qu'elles couvrent en grande partie la mer. La maille M8 et M9 sont les mailles les plus sèches, où sur 30 ans les saisons varient entre la classe très sèche et la classe sèche. Dans la figure VI-11b, la répartition des quatre classes est élaborée pour les stations pluviométriques existantes dans chaque maille.

À l'échelle de stations, les informations fournies sont très locales voire ponctuelles par contre une maille des données moyennée sur une superficie de plusieurs Kilomètres. Par conséquent, les mailles M1, M5 et M9 contenant très peu de stations sont éliminées de la comparaison. En outre, elles couvrent en grande partie la mer. Seulement les mailles M2, M3, M4, M6, M7 et M8 sont considérées (Fig.VI-11b). Les mailles M4, M7 et M8 ont une répartition des classes

similaires entre modèle et observations. La maille M6 est plutôt sèche dans les deux sources des données sauf que les faibles événements pluvieux de la classe sèche sont plus fréquents dans le modèle. Par contre, la distribution des quatre classes dans les mailles M2 et M3 du modèle et dans les stations pluviométriques dans chacune de ces mailles diverge. En effet, les 19 stations dans la maille M2 sont dominées par des saisons des classes humide et très humide (Fig.VI-11b) alors qu'elle montre des conditions beaucoup moins humides dans le modèle (Fig.VI-11b). La maille M3 est dominée par la classe très sèche dans le modèle alors que les 10 stations localisées dans la même maille sont dominées par les classes humides et sèches décrivant une forte intermittence d'événements pendant la saison DJF.

Globalement, les mailles existant dans la région semi-aride et dans le continent sont mieux représentées dans le modèle ERA-Interim que les régions côtières et les régions avec des conditions topographiques et géographiques complexes (nord-ouest).

(a) Répartition des quatre classes de ERA-Interim pour chaque maille (1980-2009)



(b) Répartition des quatre classes dans les stations pluviométriques par maille (1980-2009)

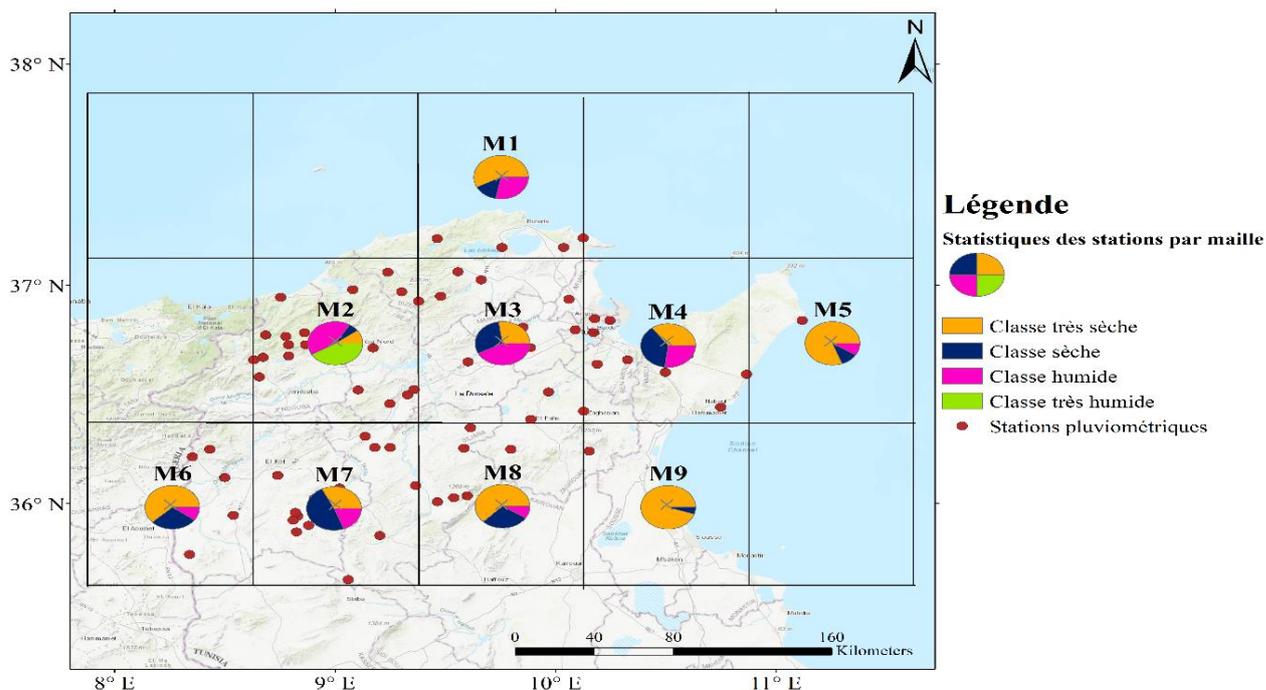


Figure VI-11 (a) Répartition des classes de ERA Inetrim pour chaque maille (1980-2009).(b) Répartition des classes des stations pluviométriques dans chaque maille (1980-2009).

VI.5 Conclusion

Dans ce chapitre, l'analyse multivariée d'une base de données pluviométrique multidimensionnelle (6-D), composée de 50 ans d'observations dans chacune des 70 pluviomètres, a été effectuée en se basant sur une méthode de classification non linéaire à deux niveaux. Le premier niveau consiste à classer toutes les saisons DJF par une quantification vectorielle (SOM) et les représenter par des vecteurs référents. Dans le deuxième niveau de classification, les prototypes sont classés par une classification hiérarchique ascendante pour synthétiser la typologie des saisons DJF en un nombre réduit des classes. Les données spatio-temporelles des pluviomètres sont regroupées en 4 classes typiques : (1) la classe 1 groupe 47 % des cas et représente les saisons extrêmement sèches, munies d'un faible nombre d'événements courts et peu intense, (2) la classe 2 représente 20 % des saisons et se caractérise par des saisons globalement sèches composées par des faibles événements fréquents, (3) la classe 3 est composée par des événements de pluviométrie moyenne et de forte intermittence pendant la saison, (4) la classe 4 représente 7 % des cas et composée par d'événements de longue durée et de très forte intensité.

L'évaluation du modèle de réanalyse s'effectue en projetant les vecteurs (6-D) décrivant les saisons DJF du modèle sur la carte topologique des données d'observations de manière d'affecter chaque saison à la classe de caractéristique similaire. Et cela donne un partitionnement des données de réanalyse.

La régionalisation spatio-temporelle des classes a souligné les spécificités pluviométriques de la zone d'étude et la variabilité interannuelle et inter-décadaire des différentes situations pluviométriques dans les données observées et les données simulées. Les résultats de comparaison des anomalies d'occurrence de chaque classe sur une période de 3 décennies de 1980 à 2009 ont montré une bonne représentation des typologies des saisons par le modèle de réanalyse ERA Interim sur les deux décades 1980-1989 et 2000-2009 contre une différence considérable sur la décade 1980-1989 et particulièrement pour les deux classes humide et très humide. La comparaison des structures spatiales des classes à l'échelle de la maille du modèle a montré une certaine surtout dans les zones de conditions topographiques et géographiques complexes.

Conclusions et perspectives

Les modèles numériques du climat sont des outils de laboratoire précieux pour faire avancer notre compréhension du système climatique à différentes échelles spatiales et temporelles. En développant ces modèles, les défis lancés étaient de pouvoir restituer les caractéristiques du climat. Ces modèles sont donc évalués sur leur capacité à reproduire de nombreux processus à toutes ces échelles spatiales et temporelles et de détecter les tendances existant dans les observations. Dans cette thèse, le modèle de réanalyse Era-Interim du Centre Européen de prévision météorologique à moyen terme a été adopté et plus particulièrement ses résultats concernant l'estimation de la pluviométrie journalière à l'échelle régionale du nord de la Tunisie. Les pluies journalières de la saison Décembre-Janvier-Février (DJF) qui représente la saison d'hiver dans l'hémisphère Nord ont été étudiées. Toutefois, la démarche est reproductible pour les autres saisons.

L'évaluation des biais dans le modèle de réanalyse adopté est une étape importante si on souhaite l'utiliser pour les études d'impact du changement climatique sur l'hydrologie de la région méditerranéenne et nord-africaine. Pour ce faire, une base de données pluviométrique journalière de 70 stations, couvrant l'échelle régionale du nord de la Tunisie et publiées par la Direction générale des ressources en eau, a été utilisée.

La pluie journalière de la saison d'hiver dans la région méditerranéenne se distingue par la succession d'épisodes secs et d'épisodes pluvieux. En adoptant un intervalle sec d'au moins deux jours pour considérer deux épisodes pluvieux distincts, nous avons défini six descripteurs pour caractériser une saison. Il s'agit de (1) le nombre d'événements, (2) le nombre de jours pluvieux, (3) le cumul saisonnier, (4) le cumul moyen par événement, (5) la durée moyenne par événement et (6) le cumul moyen par jour pluvieux. Dans l'ensemble de la thèse, l'évolution spatio-temporelle de ces descripteurs est analysée par différentes approches.

La méthode multivariée, SOM et la classification hiérarchique, ont été appliquées à ces descripteurs des 70 stations sur la période 1960-2009 afin de faire ressortir des typologies des régimes pluviométriques (classes) tout en soulignant l'aspect intermittent des pluies ainsi que la structure des événements pluvieux à l'échelle synoptique. L'interprétation des statistiques des descripteurs par classe a permis de les identifier comme suit : classe C1 représentant les saisons DJF où les saisons sont extrêmement sèches munies d'un faible nombre d'événements courts et peu intense (très sèche), classe C2 pour celles des saisons globalement sèches composées par des faibles événements fréquents, classe C3 qui inclut les saisons avec des

CONCLUSIONS ET PERSPECTIVES

événements de pluviométrie moyenne et de forte intermittence (humide) et classe C4 celles qui représente des événements de longue durée et de forte intensité (très humide). Les résultats de l'approche multivariée a permis de classifier la matrice des descripteurs issus des données de réanalyse au cours de la période 1979-2009 en quatre classes également.

Pour les données pluviométriques au sol, les anomalies du pourcentage d'occurrence de chaque classe sur cinq décades ont été étudiées. Pour l'ensemble de la zone d'étude, une anomalie négative a duré trois décennies 1960-1969 à 1980-1989 pour le pourcentage d'occurrence de la classe C1 (saisons très sèches) en même temps qu'une anomalie positive sur les trois décennies 1960-1969 à 1980-1989 pour le pourcentage d'occurrence de la classe C2 (saisons sèches). Pour la classe C3 (saisons humides), une anomalie négative a persisté sur trois décennies successives de 1970-1979 à 1990-1999. Dans les données pluviométriques au sol, la persistance et le signe des anomalies des deux classes C1 (très sèche) et C2 (sèche) sur les trois décennies 1960-1969 à 1980-1989 se retrouvent également, comme pour la région entière, à l'est et au sud de la zone d'étude.

Le pourcentage d'anomalie d'occurrence de chaque classe par décennie sur les trois décennies communes entre les observations et les réanalyses 1980-1989, 1990-1999 et 2000-2009 a été évalué. Le but était d'étudier la cohérence dans l'occurrence des classes dans le temps et dans l'espace par les deux jeux de données. Les résultats ont montré une bonne représentation des classes par le modèle ERA-Interim sur les deux décades 1990-1999 et 2000-2009 contre une représentation biaisée d'occurrence des classes pendant la décade 1980-1989. En effet, pendant cette décennie, l'écart d'anomalies du modèle par rapport aux observations est le plus élevé. De plus entre 1980-1989, les anomalies la classe C3 (humide) et la classe C4 (très humide) divergent où le modèle a surestimé l'occurrence des saisons humides et a sous-estimé l'occurrence des classes très humides.

Au nord-ouest de la zone d'étude, une anomalie positive du pourcentage d'occurrence de la classe C1 (saisons très sèches) a duré 4 décennies de 1970-1979 à 2000-2009 en même temps qu'une anomalie négative du pourcentage d'occurrence de la classe C3 (saisons humides). Vu le faible nombre de mailles des données Era-Interim, l'étude avec des sous-régions n'a pas été menée.

D'autre part, la prolongation d'une anomalie négative de la classe C2 (saisons sèches) sur 3 décennies de 1980-1989 à 2000-2009 (avec un retard d'une décennie sur la période 1970-1979 à 2000-2009) au nord-ouest de la zone d'étude a été mis en évidence. Ceci ne peut pas être étudié pour les données Era-Interim vu que la décennie 1970-1979 n'est pas couverte.

CONCLUSIONS ET PERSPECTIVES

A partir des observations des pluviométries au sol, nous avons étudié la variabilité spatiale des anomalies des descripteurs en divisant la période d'étude en deux sous-périodes : 1960-1990 et 1991-2009. Les anomalies du nombre d'événements pluvieux entre 1960-1990 et 1991-2009 sont négatives au nord de la zone d'étude et positives dans sa partie sud. Les anomalies du cumul pluviométrique des événements pluvieux entre 1960-1990 et 1991-2009 présentent un dipôle avec des anomalies négatives à l'ouest et positives à l'est.

Des études antérieures ont montré un lien entre les anomalies des pluies mensuelles de la saison hivernale dans la région méditerranéenne et les anomalies dans la circulation atmosphérique à grande échelle. Pour cette raison, nous avons étudié à travers une analyse composite de la variable climatique SLP (Sea Level Pressure) avec les facteurs principaux issus de l'analyse en composantes principales des observations. Cette analyse a permis de mettre en évidence des liens entre les années de fortes anomalies positives et négatives des composantes principales des observations pluviométriques et les anomalies de pression à une échelle globale. La visualisation des anomalies significatives montre des structures bipolaires d'anomalies de pression localisées au nord Atlantique qui coïncident avec des anomalies de la quantité de pluie apportée au nord de la Tunisie et des anomalies du facteur conditions météorologiques qui porte sur la structure de l'événement pluvieux.

Les perspectives de travail seraient d'abord d'étendre la période d'observation au-delà de 2009, ensuite d'étudier les bais dans les données ERA-5 qui est la version la plus récente des réanalyses du centre européen de prévision météorologique à moyen terme caractérisée par une forte résolution spatiale et temporelle. Enfin, vu le nombre de lacunes dans les observations sol, il serait important de procéder à des méthodes de comblement pour aller vers la réalisation d'une base de données de points de grille qui pourraient servir de base dans les études hydrologiques.

Dans des travaux futurs et en se basant sur les résultats d'évaluation du modèle, nous suggérons d'analyser les causes et les facteurs des différences trouvées et de s'approfondir sur les forçages des modèles.

Nous suggérons également d'étudier les crues résultantes des différentes classes d'événements pluvieux trouvées à partir des observations, puis, d'investiguer l'existence d'une typologie des crues en liaison avec celle de pluie et l'intervention du bassin. Dans le même contexte, des liens peuvent être établis entre les typologies des événements pluvieux et des saisons, produites par la méthode SOM, et la production agricole, la variabilité climatique, la végétation et un nombre important de paramètres directement et indirectement liés à la variabilité de précipitations.

CONCLUSIONS ET PERSPECTIVES

En perspective, il serait intéressant d'analyser la variabilité de la durée sèche entre les événements pluvieux (durée de l'inter-événement). En effet, dans la littérature, les périodes sèches sont devenues de plus en plus longues sous l'influence du changement climatique.

Références

- Aamir, E., & Hassan, I. (2020). The impact of climate indices on precipitation variability in Baluchistan, Pakistan. *Tellus A*, 72(1), 1-46. doi:10.1080/16000870.2020.1833584
- Addi, M., Gyasi-Agyei, Y., & Obuobie, E. (2022). Evaluation of imputation techniques for infilling missing daily rainfall records on river basins in Ghana. *Hydrological Sciences Journal*, 67(4), 613-627. Récupéré sur <https://doi.org/10.1080/02626667.2022.2030868>
- Akrour, 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 Resources Research*, 51(9), 7417-7435.
- Alhoniemi, E., Hollmen, J., Simula, O., & Vesanto, J. (1999, January 1). Process monitoring and modeling using the self-organizing map. *Integrated Computer-Aided Engineering*, 6(1), 3-14. doi:10.3233/ICA-1999-6102
- Ambrosi, P. & Courtois, P. (2004). Impacts du changement climatique et modélisation intégrée, la part de l'arbitraire. *Natures Sciences Sociétés*, 12, 375-386. <https://www.cairn.info/revue--2004-4-page-375.htm>.
- Ambroise, C., Seze, G., Badran, F., & Sylvie, T. (2000, January). Hierarchical clustering of self-organizing maps for cloud classification. *Neurocomputing*, 30, 47-52. doi:10.1016/S0925-2312(99)00141-1
- Annuaire pluviométrique de Tunisie 2016-2017. (2017). Ministère de l'Agriculture et des Ressources Hydrauliques - Direction Générale des Ressources en eaux.
- 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*. Texas : Research and Technology Implementation Office.
- Astel, A., Mazerski, J., Polkowska, Ż., & Namieśnik, J. (2004, March). Application of PCA and time series analysis in studies of precipitation in Tricity (Poland). *Advances in Environmental Research*, 8(3-4), 337-349. doi:10.1016/S1093-0191(02)00107-7
- Balsamo, G., Boussetta, S., Lopez, P., and Ferranti, L.(2010). Evaluation of ERA-Interim and ERA-Interim-GPCP-rescaled precipitation over the U.S.A., ERA Report Series No. 5, ECMWF, Reading, UK, 10 pp.
- 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. *Journal of Hydraulic engineering*, 120(2), 229-235. doi:10.1061/(ASCE)0733-9429(1994)120:2(228)
- Ben Boubaker, H. (2003). Apports pluviométriques et types de flux de surface en Tunisie du Nord. Dans P. Arnould, & M. Hotyat, *Eau et environnement: Tunisie et milieux méditerranéens* (Vol. 870). ENS Éditions. doi:10.4000/books.enseditions.870

REFERENCES

- Benzarti, Z. (2003). La pluviométrie en Tunisie : Analyse des années très pluvieuses. Dans P. Arnould, & M. Hotyat, *Eau et environnement: Tunisie et milieux méditerranéens* (p. 869). Lyon, France: ENS Éditions. doi:10.4000/books.enseditions
- Benzarti, Z., & Habaieb, H. (2001). Etude de la persistance de la sécheresse en Tunisie par utilisation des chaînes de Markov (1909-1996). *Science et changements planétaires / Sécheresse*, 12(4), 215–220.
- Berndtsson, R. (1989, July). Topographical and coastal influence on spatial precipitation patterns in Tunisia. *International Journal of Climatology*, 9(4), 357 - 369. doi:10.1002/joc.3370090403
- Bjerknes, V. (1904). *Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik [The problem of weather prediction as seen from the standpoint of mechanics and physics]* (Vol. 21). Wien: Ed. Hölzel.
- Bjerknes, V. (1995). Le problème de la prévision du temps du point de vue de la mécanique et de la physique. *La Météorologie*, 8(9), 55–62.
- Blayo, E. (2002). *Modélisation et assimilation de données en océanographie*. Mémoire en vue de l'obtention du diplôme d'Habilitation à Diriger des Recherches, Université Joseph Fourier, Grenoble I.
- Bloschl G. et Sivapalan M. (1995). Scale issues in hydrological modelling: a review. *Hydrological processes* 9: 251-290.
- Boulmaiz, T., Boutaghane, H., Abida, H., Saber, M., Kantoush, S.A., Trambly, Y. (2022). Exploring the Spatio-Temporal Variability of Precipitation over the Medjerda Transboundary Basin in North Africa. *Water*, 14, 423. <https://doi.org/10.3390/w14030423>
- Bourque, J., Poulin, N., & Cleaver, A. F. (2006). Évaluation de l'utilisation et de la présentation des résultats d'analyses factorielles et d'analyses en composantes principales en éducation. *Revue des sciences de l'éducation*, 32(2), 325–344. doi:10.7202/014411ar
- Bracken, L.J., Cox, N.J., Shannon, J., 2008. The relationship between rainfall inputs and flood generation in south-east Spain. *Hydrol. Process.*, 22, 683–696.
- 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 Resources Management*, 25(5), 1269–1279. doi:10.1007/s11269-010-97
- Breard, G. T. (2017). *Evaluating Self-Organizing Map Quality Measures as Convergence Criteria*. Master of science thesis, University of Rhode Island. doi:10.23860/thesis-breard-gregory-2017
- Brentan, B., Meirelles, G., Luvizotto Jr, E., & Izquierdo, J. (2018). Hybrid SOM+ k -Means clustering to improve planning, operation and management in water distribution systems. *Environmental Modelling and Software*, 106. doi:10.1016/j.envsoft.2018.02.013
- Bromwich, D.H., Nicolas, J.P., Monaghan, A.J. (2011). An assessment of precipitation changes over Antarctica and the Southern Ocean since 1989 in contemporary global reanalyses. *J. Climate*, DOI:10.1175/2011JCLI4074.1.

REFERENCES

- Caruana, R., & Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning*, 161-168. doi:10.1145/1143844.1143865
- 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 and Tillage Research*, 86, 38–51.
- Chiang, J. C., & Sobel, A. H. (2002). Tropical Tropospheric Temperature Variations Caused by ENSO and Their Influence on the Remote Tropical Climate. *Journal of climate*, 15(18), 2616-2631. doi:10.1175/1520-0442(2002)015<2616:TTTVCB>2.0.CO;2
- Chon, T. (2011). Self-Organizing Maps applied to ecological sciences. *Ecological Informatics*, 6(1), 50–61. doi:10.1016/j.ecoinf.2010.11.002
- Comrie, A. C., & Glenn, E. C. (1998, December). Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. *Climate Research*, 10 (3), 201-215. Récupéré sur <https://www.jstor.org/stable/24865968>
- Criado-Aldeanueva, F., & Soto-Navarro, J. (2020). Climatic Indices over the Mediterranean Sea: A Review. *Applied Sciences*, 10(17). doi:10.3390/app10175790
- Croce, P., Formichi, P., Landi, F.(2021). Evaluation of Current Trends of Climatic Actions in Europe Based on Observations and Regional Reanalysis. *Remote Sensing*. 13(11):2025. <https://doi.org/10.3390/rs13112025>
- Cucchi, M., Weedon, G P., Amici, A.,; Bellouin, N., Lange, S., et al...(2020). Earth System Science Data; *Katlenburg-Lindau* Vol. 12, N° 3, 2097-2120. DOI:10.5194/essd-12-2097-2020
- Cunningham, P., & Delany, S. (2007). k-Nearest Neighbour Classifiers. *ACM Computing Surveys*, 54(6). doi:10.1145/3459665
- Darand, M., & Mansouri Daneshvar, M. R. (2014). Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. *Environmental Processes*, 1(4), 517-532. doi:10.1007/s40710-014-0039-1
- Decastro, M., Lorenzo, N., Taboada, J. J., Sarmiento, M., Alvarez, I., & Gómez-Gesteira, M. (2006). Influence of teleconnection patterns on precipitation variability and on river flow regimes in the Miño River basin (NW Iberian Peninsula). *Climate Research*, 32(1), 63–73. doi:10.3354/cr032063
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553-597. doi:10.1002/qj.828
- Derouiche S., Mallet C., Bargaoui Z. (2019). Rainfall event analysis in the north of Tunisia using the self-organizing map. Proceedings of the 9th International Workshop on Climate Informatics, Oct 2019, Paris, France. 5 p - insu-02314437
- Derouiche, S., Mallet, C., Hannachi, A., & Bargaoui, Z. (2022). Characterisation of rainfall events in northern Tunisia using self-organising maps. *Journal of Hydrology: Regional Studies*, 42, 101159. doi: 10.1016/j.ejrh.2022.101159
- Dilmi, D., Mallet, C., Barthès, L., & Chazottes, A. (2017). Data-driven clustering of rain events: microphysics information derived from macro-scale observations Atmospheric

REFERENCES

- Measurement Techniques. *European Geosciences Union*, 10(4), 1557-1574. doi:10.5194/amt-10-1557-2017
- Douguedroit, A. (1987). The variations of dry spells in marseilles from 1865 to 1984. *Journal of Climatology*, 7(6), 541-551. Récupéré sur <https://doi.org/10.1002/joc.3370070603>
- Douguedroit, A. (1988). Les précipitations en Tunisie (1951-1980). *Méditerranée*, 66(4), 23-33. doi:10.3406/medit.1988.2578
- Döscher, R., Acosta, M., Alessandri, A., et al. (2022). The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geosci Model Dev* 15:2973–3020. <https://doi.org/10.5194/gmd-15-2973-2022>
- Dunkerley, D. (2008). Identifying individual rain events from pluviograph records: A review with analysis of data from an Australian dryland site. *Hydrological Processes*, 5024–5036. doi:10.1002/hyp.7122
- Ellouze, M., Azri, C., & Abida, H. (2009). Special variability of monthly and annual rainfall data over Souther Tunisia. *Atmospheric Research*, 93(4), 832-839. doi:10.1016/j.atmosres.2009.04.005
- Emberger, L. (1930). La végétation de la région méditerranéenne- Essai d'une classification des groupements végétaux. *Revue générale de botanique*, T 42, 461–662; 705- 721.
- Escofier, B., & Pagès, J. (2008). *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*. Sciences Sup. Paris: Dunod. Récupéré sur <https://hal.archives-ouvertes.fr/hal-00382085>
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231. doi:10.5555/3001460.3001507
- Estienne, P., & Godard, A. (1992). *Climatologie*. Paris: Armand Colin, ISBN 2200219180.
- Ferrel, W. (1856). An essay on the winds and currents of the ocean. *Nashville Journal of Medicine and Surgery*, 11, 287–301.
- Fischer, A. (2014). Deux méthodes d'apprentissage non supervisé: synthèse sur la méthode des centres mobiles et présentation des courbes principales. *Journal de la Société Française de Statistique*, 155(2), 2-35.
- Fita, L., Romero, R., Luque, A., Emanuel, K., & Ramis, C. (2007). Analysis of the environments of seven Mediterranean tropical-like storms using an axisymmetric, nonhydrostatic, cloudresolving model. *Natural Hazards and Earth System Sciences*, 7(1), 41–56. doi:10.5194/nhess-7-41-2007
- Fourier, J. (1822). *Théorie de la chaleur*, Chez Firmin Didot, Père et Fils, Paris, France.
- Fornis, R.L., Vermeulen, H.R., Niewenhuis, J.D., 2005. Kinetic energy-rainfall intensity relationship for central Cebu, Philippines for soil erosion studies. *J. Hydrol.*, 300, 20–32.
- Fréville, H. (2015). *Observation et simulation de la température de surface en Antarctique : application à l'estimation de la densité superficielle de la neige*. Toulouse III: Océanographie. Université Paul Sabatier. Récupéré sur <https://tel.archives-ouvertes.fr/tel-01512722>

REFERENCES

- Froidurot, S. (2015). *Approche multi-échelle pour l'évaluation de la pluie dans les modèles climatiques régionaux. Etude dans le sud-est de la France*. Université Grenoble Alpes. Océan, Atmosphère. Récupéré sur <https://tel.archives-ouvertes.fr/tel-01311155>
- Fu, J., Zhang, F., & D. Hewson, T. (2020). Object-Oriented Composite Analysis of Warm-Sector Rainfall in North China. *Monthly Weather Review*, 148(7), 2719–2735. doi:10.1175/MWR-D-19-0038.1
- Gaál, L., Molnar, P., & Szolgay, J. (2014). Selection of intense rainfall events based on intensity thresholds and lightning data in Switzerland. *Hydrology and Earth System Sciences*, 18(5), 1561–1573. doi:10.5194/hess-18-1561-2014
- Gargouri-Ellouze, E., & Bargaoui, Z. (2009). Investigation with Kendall plots of infiltration index–maximum rainfall intensity relationship for regionalization. *Physics and Chemistry of the Earth Parts A/B/C*, 34(10-12), 642-653. doi: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. *Hydrological Sciences Journal*, 58(3), 483-497. doi: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. *Journal of Hydrology*, 514, 38–49. doi:10.1016/j.jhydrol.2016.07.002
- Giannini, A., Saravanan, R., & Chang, P. (2003). Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales. *Science*, 302(5647), 1027–1030. doi:10.1126/science.1089357
- Giorgi, F. (2006). Climate change hot-spots. *Geophysical Research Letters*, 33(8), 1-4. doi:10.1029/2006GL025734
- Giorgi, F., & Lionello, P. (2008). Climate change projections for the Mediterranean region. *Global and Planetary Change*, 63(2-3), 90-104. doi:10.1016/j.gloplacha.2007.09.005
- Greene, D., Cunningham, P., & Mayer, R. (2008). *Unsupervised Learning and Clustering. Cognitive Technologies*. Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-75171-7_3
- Gueye, A. J., Niang, A., Sawadogo, S., Sultan, B., DiongueNianga, 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 scale. *Climate Dynamics*, 36(1-2), 1–18. doi:10.1007/s00382-010-0782-6
- Guillemot, H. (2009). Comment évaluer un modèle numérique de climat ? Circulations et dialogues entre simulations et observations dans les pratiques des modélisateurs. *Revue d'anthropologie des connaissances*, 3, 2, 273-293.
- Gutierrez, L. (2017). Impacts of El Niño-Southern Oscillation on the wheat market: A global dynamic analysis. *PLOS ONE*, 12(6). doi:10.1371/journal.pone.0179086
- Hadley, G. (1735). Concerning the Cause of the General Trade-Winds. *Philosophical Transactions of the Royal Society of London (1683-1775)*, 39, 58–62. Récupéré sur <http://www.jstor.org/stable/103976>

REFERENCES

- Hajjar, C. (2014). *Cartes auto-organisatrices pour la classification de données symboliques mixtes, de données de type intervalle et de données discrétisées*. thèse de doctorat: Supélec. Récupéré sur <https://tel.archives-ouvertes.fr/tel-01142849>
- Hakim, G., & Patoux, J. (2018). *Weather: A concise Introduction*. Cambridge University Press.
- Hannachi, A. (2004). A Primer for EOF Analysis of Climate Data. *Department of Meteorology, University of Reading, UK*.
- Hannachi, A. W. (2012). The North Atlantic jet stream: a look at preferred positions, paths and transitions. 862-877.
- Hannachi, A., Mitchell, D., Gray, L., & Charlton-Perez, A. (2010). On the use of geometric moments to examine the continuum of sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, 68(3), 657–674. Récupéré sur <https://doi.org/10.1175/2010JAS3585.1>
- Hastie, T., Tibshirani, R., Sherlock, G., Michael, E., & Botstein, D. (1999). *Imputing Missing Data for Gene Expression*. Technical report, Stanford
- Haykin, S. (2009). *Neural Networks and Learning Machine. 3rd ed.* Hamilton, Canada: Pearson.
- Henia, L. (1986). La variabilité spatiale des pluies en Tunisie (Spatial variability of rainfall in Tunis). *Bulletin de l'Association de Géographes Français*, 141-147. doi:10.3406/bagf.1986.1334
- Hersbach, H. et al. (2020). The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999-2049. <https://doi.org/10.1002/qj.3803>
- Hewitson, B., & Crane, R. G. (2002). Self-Organizing Maps: Applications to synoptic climatology. *Climate Research*, 22(1), 13-26. doi:10.3354/cr022013
- Horst-Joachim Lüdecke, H., Lüdecke, Gisela Müller-Plath, G., Müller-Plath, Michael G., Wallace, M. G. Wallace., & Sebastian Lüning, S. Lüning. (2021). Decadal and multidecadal natural variability of African rainfall. *Journal of hydrology*, 34, 100795. doi: [10.1016/j.ejrh.2021.100795](https://doi.org/10.1016/j.ejrh.2021.100795)
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441. doi:10.1037/h0071325
- Høyer, J. L., & Karagali, I. (2016). Sea Surface Temperature Climate Data Record for the North Sea and Baltic Sea. *Journal of Climate*, 29(7), 2529–2541. doi:10.1175/JCLI-D-15-0663.1
- Hua, G., Skalesky, M., & Westermann, K. (2009). Exploratory analysis of CIA Fact book data using Kohonen self-organizing maps. *Case Studies in Business, Industry and Government Statistics (CSBIGS)*, 3(1), 48–59.
- Husson, F., Josse, J., & Pages, J. (2010). Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data. *Applied Mathematics Department*, 1-17.
- INSEE. (2020). *Changement climatique et effet de serre*. Montrouge: institut nationale de la statistique et des études économiques.
- Ioannou, A., Creaco, E., & Laspidou, C. (2021). Exploring the Effectiveness of Clustering *Sustainability*, 2603(13). Récupéré sur <https://doi.org/10.3390/su13052603>

REFERENCES

- IPCC. (2007). *Climate Change 2007: The Physical Science Basis; Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change - Summary for Policymakers*. Cambridge University Press. Récupéré sur <https://wedocs.unep.org/20.500.11822/30763>
- IPCC, (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Lösche, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. In Press.
- Irannezhad, M., Chen, D., Kløve, B. and Moradkhani, H. (2017). Analysing the variability and trends of precipitation extremes in Finland and their connection to atmospheric circulation pat-terns. *International Journal of Climatology*, 37, 1053–1066.
- Jancovici, J.-M. (2007, aout 1). Quels sont les gaz à effet de serre ? Récupéré sur changement climatique: <https://jancovici.com/changement-climatique/gaz-a-effet-de-serre-et-cycle-du-carbone/quels-sont-les-gaz-a-effet-de-serre-quels-sont-leurs-contribution-a-leffet-de-serre/>
- Jayawardene, H. K., Sonnadara, D. U., & Jayewardene, D. R. (2005). Spatial interpolation of weekly rainfall depth in the dry zone of Sri Lanka. *Climate Research*, 29(3), 223-231. doi:10.3354/cr029223
- Jebari, S., Berndtsson, R., Uvo, C. Y., & Bahri, A. (2007). Regionalizing fine time-scale rainfall affected by topography in semi-arid Tunisia. *Hydrological Sciences Journal*, 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. *Revue des Sciences 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. *Atmosphere-Ocean*, 54(5), 469-497. doi:10.1080/07055900.2016.1228522
- Jemai, S., Ellouze, M., & Abida, H. (2017). Variability of Precipitation in Arid Climates Using the Wavelet Approach: Case Study of Watershed of Gabes in South-East Tunisia. *Atmosphere*, 8, 178. Récupéré sur <https://doi.org/10.3390/atmos8090178>
- Jhajharia, D., Chattopadhyay, S., Choudhary, R. R., Dev, V., Singh, V. P., & Lal, S. (2013). Influence of climate on incidences of malaria in the Thar Desert, northwest India. *International Journal of Climatology*, 33(2), 312–325. doi:10.1002/joc.3424
- Jiang, N., Griffiths, G., & Dirks, K. N. (2011). Linking synoptic weather types to daily rainfall in Auckland. *Weather and Climate*, 31, 50–66. Récupéré sur <https://doi.org/10.2307/26169717>
- Jolliffe, I. T. (2002). *Principal Component Analysis, Second Edition*. New York: Springer-Verlag.
- Jones, P. (2016). The reliability of global and hemispheric surface temperature records. *Adv. Atmos. Sci.* 33, 269–282. <https://doi.org/10.1007/s00376-015-5194-4>
- Joo, J., Lee, J., Kim, J., Jun, H., & Jo, D. (2014). Inter-Event Time Definition Setting Procedure for Urban Drainage Systems. *Water*, 6(1), 45–58. doi:10.3390/w6010045

REFERENCES

- Kalteh, A. M., Hjorth, P., & Berndtsson, R. (2008). Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling & Software*, 23(7), 835–845. doi: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). *Revue des Sciences de l'eau*, 2(3), 429-447.
- Kohonen, T. (1988). *Self-Organization and Associative Memory*. Berlin, Heidelberg: Springer. Récupéré sur <https://doi.org/10.1007/978-3-642-88163-3>
- Kohonen, T. (2001). *Self-Organizing Maps*. Berlin: Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-56927-2
- Lai, H.-W., Chen, H. W., Kukulies, J., Ou, T., & Chen, D. (2021). Regionalization of seasonal precipitation over the Tibetan Plateau and associated large-scale atmospheric Systems. *Journal of Climate*, 34(7), 2635–2651. doi:10.1175/JCLI-D-20-0521.1
- Lebart, L., Morineau, A., & Piron, M. (2000). *Statistique exploratoire multidimensionnelle* (éd. 3ème édition). Dunod.
- Lee, K., Chung, J., Cha, Y., & Suh, C. (2018). *Machine learning approaches for learning analytics: Collaborative filtering or regression with experts?* Korea.
- Ling, R. F. (1973). A computer-generated aid for cluster analysis. *Communications of the ACM*, 16(6), 355–361.
- Liu, Y., & Weisberg, R. H. (2011). *A review of self-organizing map applications in meteorology and oceanography*. (J. I. Mwasiagi, Éd.) Rijeka, Croatia: InTech. doi:10.5772/13146
- Lloyd, C.R., 1990. The temporal distribution of Amazonian rainfall and its implications for forest interception. *Q. J. Roy. Meteor. Soc.*, 116, 1487–1494.
- Lopez-Moreno, J., Vicente-Serrano, S., Morán-Tejeda, E., Lorenzo-Lacruz, J., Kenawy, A., & Beniston, M. (2011). Effects of the North Atlantic Oscillation (NAO) on combined temperature and precipitation winter modes in the Mediterranean mountains: Observed relationships and projections for the 21st century. *Global and Planetary Change*, 77(1-2), 62-76. doi:10.1016/j.gloplacha.2011.03.003
- Malhotra, V. K., Kaur, H., & Alam, M. (2014). An analysis of fuzzy clustering methods. *International Journal of Computer Applications*, 94(19), 9-12. doi:10.5120/16497-6578
- Malmgren, B. A., & Winter, A. (1999). Climate zonation in Puerto Rico based on principal components analysis and an artificial neural network. *Journal of Climate*, 12(4), 977–985. doi:10.1175/1520-0442(1999)012<0977:CZIPRB>2.0.CO;2
- Marres, P. (1972). Louis Emberger. 1897-1969. In: *Méditerranée*, deuxième série, tome 9, 1-1972. pp. 1-4; https://www.persee.fr/doc/medit_0025-8296_1972_num_9_1_1414
- Marshall, G.J., Jylhä, K., Kivinen, S., Laapas, M. and Dyrddal, A.V. (2020) The role of atmospheric circulation patterns in driving recent changes in indices of extreme seasonal precipitation across Arctic Fennoscandia. *Climatic Change*, 162, 741–759.
- Martínez, M. D., Lana, X., Burgueño, A., & Serra, C. (2007). Spatial and temporal daily rainfall regime in Catalonia (NE Spain) derived from four precipitation indices, years 1950–2000. *International Journal of Climatology*, 27(1), 123-138. doi:10.1002/joc.1369

REFERENCES

- Martinez, W., & Martinez, A. (2005). *Exploratory data analysis with MATLAB*. Florida, USA: Chapman & Hall/CRC Press.
- Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., and Cayan, D. R. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrol. Earth Syst. Sci*, 14, 1125–1138, <https://doi.org/10.5194/hess-14-1125-2010>, 2010.
- Mephu Nguifo, E., & Njiwoua, P. (2005). Treillis de concepts et classification supervisée. *Revue des Sciences et Technologies de l'Information - Série TSI : Technique et Science Informatiques, Lavoisier*, 24(4), 449-488. doi:10.3166/tsi.24.449-488
- Merzougui, A., & Slimani, M. (2012). Régionalisation des lois de distribution des pluies mensuelles en Tunisie. *Hydrological Sciences Journal*, 57(4), 668-685. doi:10.1080/02626667.2012.670702
- Mignot, J., Mejia, C., Sorrow, C., Sylla, A., Crépon, M., & and Thiria, S. (2020). Towards an objective assessment of climate multi-model ensembles – a case study: the Senegalo-Mauritanian upwelling region. *Geoscientific Model Development*, 13, 2723–2742. Récupéré sur <https://doi.org/10.519>
- Minnotte, M. C., & West, R. (1999). The data image: A tool for exploring high dimensional data set. *American Statistical Association*, 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. *Journal of Hydrology and Hydromechanics*, 64, 226–236. Récupéré sur <https://doi.org/10.1515/johh-2016-0024>
- Morissette, L., & Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, 9(1), 15-24. doi:10.20982/tqmp.09.1.p015
- Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *Computer Journal*, 24(4), 354–359. doi:10.1093/comjnl/26.4.354
- Murtagh, F., & Contreras, P. (2011). Methods of hierarchical clustering. *ArXiv11050121 Cs Math Stat*. doi:10.48550/arXiv.1105.0121
- Murtagh, F., & Contreras, P. (2012). Methods of Hierarchical Clustering, in Data Mining and Knowledge Discovery. *Wiley-Interscience*, 2, 86-97.
- Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification*, 31, 274–295.
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*, 4, 51-62.
- Natita, W., Wiboonsak, W., & Dusadee, S. (2016). Appropriate Learning Rate and Neighborhood Function of Self-organizing Map (SOM) for Specific Humidity Pattern Classification over Southern Thailand. *International Journal of Modeling and Optimization*, 6(1), 61-65. doi:10.7763/IJMO.2016.V6.504
- Ning, L., & Bradley, R. S. (2014). Winter precipitation variability and corresponding teleconnections over the northeastern United States. *Journal of Geophysical Research: Atmospheres*, 119(13), 7931-7945. doi:10.1002/2014JD021591

REFERENCES

- Nor, S. M., Shaharudin, S. M., Ismail, S., Zainuddin, N. H., & Tan, M. L. (2020). A comparative study of different imputation methods for daily rainfall data in east-coast Peninsular Malaysia. *Bulletin of Electrical Engineering and Informatics*, 9(2), 635-643. doi:10.11591/eei.v9i2.2090
- Orlanski, I.(1975). A rational subdivision of scales for atmospheric pro-cesses. *B. Am. Meteorol. Soc.* 56, 527–530.
- Ouachani, R., Bargaoui, Z., & Ouarda, T. (2013). Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda Basin. *International journal of Climatology*, 33(1), 58–76. doi:10.1002/joc.3407
- Parchure, A. S., & Gedam, S. K. (2019). Self-organising maps for rain event classification in Mumbai City, India. *ISH Journal of Hydraulic Engineering*. doi:10.1080/09715010.2019.1581099
- Peeters, L., Bacao, F., & Lobo, V. D. (2006). Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM, a variant of Kohonen's Self-Organizing Map. *Hydrology and Earth System Sciences*, 11(4), 1309–1321. doi:10.5194/hess-11-1309-2007
- Perez-Hoyos, A., Martinez, B., Javier Garcia-Haro, F., Moreno, A., & Amparo Gilabert, M. (2014). Identification of Ecosystem Functional Types from Coarse Resolution Imagery Using a Self-Organizing Map Approach: A Case Study for Spain. *Remote Sensing*, 6(11), 11391-11419.
- Poccard-Leclercq, I. F. (2000). *Etude diagnostique de nouvelles données climatiques : les réanalyses. Exemples d'application aux précipitations en Afrique tropicale*. Thèse de doctorat, Université de Bourgogne, Géographie. Récupéré sur <https://tel.archives-ouvertes.fr/tel-00012042>
- Polade, S., Pierce, D., Cayan, D., Gershunov, A., & Dettinger, M. D. (2014). The key role of dry days in changing regional climate and precipitation regimes. *Scientific Reports*, 4(1). doi:10.1038/srep04364
- Power, S., Tseitkin, F., Torok, S., Lavery, B., Dahni, R., & McAvaney, B. (1998). Australian temperature, Australian rainfall and the Southern Oscillation, 1910-1992: Coherent variability and recent changes. *Australian Meteorological Magazine*, 47(2), 85-101.
- Qiao, P., Liu, W., Zhang, Y., & Gong, Z. .. (2021). Complex Networks Reveal Teleconnections between the Global SST and Rainfall in Southwest China. *Atmosphere*, 12(1). doi:10.3390/atmos12010101
- Quadrelli, R., Pavan, V., & Molteni, F. (2001). Wintertime variability of Mediterranean precipitation and its link with large-scale circulation anomalies. *Climate Dynamics*, 17, 457–466.
- Raymond, F., Ullmann, A., Camberlin, P., Oueslati, B., & Drobinski, P. (2018). Atmospheric conditions and weather regimes associated with extreme winter dry spells over the Mediterranean basin. *Climate Dynamics*, 50, 4437–4453. Récupéré sur <https://doi.org/10.1007/s00382-017-3884-6>
- Restrepo-Posada, P. (1982). Identification of independent rainstorms. *Journal of Hydrology*, 55(1-4), 303-319. doi:10.1016/0022-1694(82)90136-6

REFERENCES

- Rivoire, P., Trambly, Y., Neppel, L., Hertig, E., & Vicente-Serrano, S. (2019). Impact of the dry-day definition on Mediterranean extreme dry-spell analysis. *Natural Hazards and Earth System Sciences*, 19(8), 1629–1638. Récupéré sur <https://doi.org/10.5194/nhess-19-1629-2019>
- Ropelewski, C. F., & Halpert, M. S. (1989). Precipitation patterns associated with high index phase of Southern Oscillation. *Journal of Climate*, 2(3), 268–284. doi:10.1175/1520-0442(1989)002<0268:PPAWTH>2.0.CO;2
- Ropelewski, C., & Halpert, M. (1987). Global and regional scale precipitation patterns associated with El Niño/Southern Oscillation. *Monthly Weather Review*, 115(8), 1606–1626. doi:10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2
- Rui, X., & Ii, D. C. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645–678.
- Ruti, P., Somot, S., Giorgi, F., Dubois, C., Flaouanas, E., Obermann, A., . . . Vervatis, V. (2016). Med-CORDEX Initiative for Mediterranean Climate Studies. *Bulletin of the American Meteorological society*, 97(7), 1187–1208. doi:10.1175/BAMS-D-14-00176.1
- Rysman, J. F., Verrier, S., Lemaître, Y., & Moreau, E. (2013). Space-time variability of the rainfall over the western Mediterranean region: A statistical analysis. *Journal of Geophysical Research: Atmospheres*, 118 (15), 8448–8459. doi: 10.1002/jgrd.50656
- Santer, B. D., et al. (2004). Identification of anthropogenic climate change using a second-generation reanalysis, *J. Geophys. Res.*, 109, D21104, doi:10.1029/2004JD005075.
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., . . . Chin-Teng, L. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267, 664–681. doi:10.1016/j.neucom.2017.06.053
- Schuenemann, K. C., Cassano, J. J., & Finnis, J. (2009). Synoptic forcing of precipitation over Greenland: Climatology for 1961–99. *Journal of Hydrometeorology*, 10(1), 60–78. Récupéré sur <https://doi.org/10.1175/2008JHM1014.1>
- Selmi, S., & Nasri, S. (1997). Les lacs et retenues collinaires en Tunisie. *Fonds Documentaire ORSTOM*. Récupéré sur http://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_6/griseli1/010013332
- Simmons, AJ., Willett, KM., Jones, PD., Thorne, PW., Dee, DP. (2010). Lowfrequency variations in surface atmospheric humidity, temperature and precipitation: Inferences from reanalyses and monthly gridded observational datasets. *J. Geophys. Res.* 115: D01110.
- Skliris, N., Sofianos, S., Gkanasos, A., Mantziafou, A., Vervatis, V., Axaopoulos, P., & Lascaratos, A. (2012). Decadal scale variability of sea surface temperature in the Mediterranean Sea in relation to atmospheric variability. *Ocean Dynamics*, 62(1), 13–30. doi:10.1007/s10236-011-0493-5
- Slimani, M., Cudennec, C., & Feki, H. (2007). Structure du gradient pluviométrique de la transition Méditerranée–Sahara en Tunisie: déterminants géographiques et saisonnalité. *Hydrological Sciences Journal*, 52(6), 1088–1102. doi:10.1623/hysj.52.6.1088
- Smaoui, A. (2015). Bioclimat et Végétation de la Tunisie et des régions prospectées pendant le 12ème ITER Mediterraneum de OPTIMA., 27, pp. 13-20.

REFERENCES

- Stekhoven, D. J., & Buehlmann, P. (2012). Missforest - non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118. Récupéré sur <https://doi.org/10.1093/bioinformatics/btr597>
- Tabari, H., & Willems, P. (2018). Lagged influence of Atlantic and Pacific climate patterns on European extreme precipitation. *Scientific Reports*, 8(1), 5748. doi:10.1038/s41598-018-28111-1
- Tavakolifar, H., Shahghasemi, E. & Nazif, S. (2017). Evaluation of climate change impacts on extreme rainfall events characteristics using a synoptic weather typing-based daily precipitation downscaling model. *Journal of Water and Climate Change*, 8 (3), 388–411.
- Thompson, R., & Green, D. N. (2004). Mediterranean precipitation and its relationship with sea level pressure patterns. *Annals of geophysics*, 47(5), 1617-1633.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society Series B*, 63(2), 411–423. doi:10.1111/1467-9868.00293
- Tramblay, Y., El Adlouni, S., & Servat, E. (2013). Trends and variability in extreme precipitation indices over Maghreb countries. *Natural Hazards and Earth System Sciences*, 13(12), 3235–3248. doi: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). *International Journal of Climatology*, 39(15), 5830-5838. doi:10.1002/joc.6181
- Trenberth, K. E. (1983). What are the seasons? *Bulletin of the american meteorological society*, 64(11), 1276–1282. Récupéré sur [https://doi.org/10.1175/1520-0477\(1983\)064<1276:WATS>2.0.CO;2](https://doi.org/10.1175/1520-0477(1983)064<1276:WATS>2.0.CO;2)
- Tol, Richard S.J. (2006). "Integrated Assessment Modelling," Working Papers FNU-102, Research unit Sustainability and Global Change, Hamburg University, revised May 2006.
- Ulbrich, U., Lionello, P., Belušić, D., Jacobeit, J., Knippertz, P., Kuglitsch, F., . . . Ziv, B. (2012). Climate of the Mediterranean: Synoptic Patterns, Temperature, Precipitation, Winds, and Their Extremes. Dans P. Lionello (Éd.), *The Climate of the Mediterranean Region* (pp. 301 - 346). Amsterdam, The Netherlands: Elsevier. Récupéré sur <https://www.researchgate.net/publication/250928164>
- Uppala, S.M., et al. (2005). The Era-40 Re-Analysis. *Quarterly Journal of the Royal Meteorological Society*, 131, 2961-3012. <http://dx.doi.org/10.1256/qj.04.176>
- Verrier, S., Mallet, C., & Barthès, L. (2011). Multiscaling properties of rain in the time domain, taking into account rain support biases. *Journal of Geophysical Research: Atmospheres*, 116(D20)
- Vesanto, J. (2000). Neural Network Tool for Data Mining: Som Toolbox. *Proceedings of Symposium on Tool Environments and Development Methods for Intelligent Systems (TOOL-MET2000)*, 184–196.
- Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11, 586–600.

REFERENCES

- Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankagas, J. (2000). *SOM Toolbox for Matlab 5*. Helsinki University of Technology, Finland. Récupéré sur <http://www.cis.hut.fi/projects/somtoolbox/>.
- Wang, S., Huang, J., He, Y., & Guan, Y. (2014). Combined effects of the Pacific decadal oscillation and El Ninosouthern oscillation on global land dry–wet changes. *Scientific reports*(6651). doi:10.1038/srep06651
- Wang, Z., Yang, S., Lau, N.-C., & Duan, A. (2018). Teleconnection between Summer NAO and East China Rainfall Variations: A Bridge Effect of the Tibetan Plateau. *Journal of Climate*, 31(16), 6433–6444. doi:10.1175/JCLI-D-17-0413.1
- Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244.
- Wegman, E. (1990). Hyperdimensional Data Analysis Using Parallel Coordinates. (L. Taylor & Francis, Éd.) *Journal of the American Statistical Association*, 85(411), 664-675. doi:10.2307/2290001
- Wenzel, H. G., & Voorhees, M. L. (1981). *Evaluation of the urban design storm concept*. Illinois: University of Illinois at Urbana-Champaign. Water Resources Center. Récupéré sur <http://hdl.handle.net/2142/90287>
- White, D., Richman, M., & Yarnal, B. (1991). Climate regionalization and rotation of principal components. *International Journal of Climatology*, 11(1), 1-25. doi:10.1002/joc.3370110102
- Wigley, T. M., Lough, J. M., & Jones, P. D. (1984). Spatial patterns of precipitation in England and Wales and a revised, homogeneous England and Wales precipitation series. *Journal of Climatology*, 4(1), 1-25. doi:10.1002/joc.3370040102
- Xoplaki, E., Gonzalez-Rouco, J. F., Luterbacher, J. U., & Wanner, H. (2004). Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends. *Climate dynamics*, 23(1), 63-78.
- Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645 - 678. doi:10.1109/TNN.2005.845141
- Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internetusage behaviors using machine learning algorithms. *Computers in Human Behavior*, 98, 166–173. doi:10.1016/j.chb.2019.04.015
- Yang, R., & Xing, B. (2022). Teleconnections of Large-Scale Climate Patterns to Regional Drought in Mid-Latitudes: A Case Study in Xinjiang, China. *Atmosphere*, 13(2), 230. doi:10.3390/atmos13020230
- Yu, Z.-P., Chu, P.-S., & Schroeder, T. (1997, October 1). Predictive Skills of Seasonal to Annual Rainfall Variations in the U.S. Affiliated Pacific Islands: Canonical Correlation Analysis and Multivariate Principal Component Regression Approaches. *Journal of climate*, 10(10), 2586–2599. doi:10.1175/1520-0442(1997)010<2586:PSOSTA>2.0.CO;2
- Zahar, Y. (2015). Une hydrologie des extrêmes, inondations et sécheresses : éléments choisis de caractérisation en Tunisie. *3ème journée scientifique de l'académie tunisienne . La mesure du risque hydrologique : avancées et applica-tions scientifiques dans le secteur de l'eau*. Beit EL Hikma.

REFERENCES

- Zahar, Y., & Laborde, J. (2007). Modélisation statistique et synthèse cartographique des précipitations journalières extrêmes de Tunisie. *Revue Des Sciences De L'Eau*, 20(4), 409–424. doi:10.7202/016914ar
- Zahar, Y., Laborde, J., & Benzarti, Z. (2000). Les pluies journalières extrêmes de Tunisie : modélisation statistique par la combinaison de deux populations de distributions exponentielles et poissonniennes. *International Journal of Climatology*(12), 159–167.

Annexes

Annexe A : les indices climatiques

1. L'Oscillation Nord Atlantique (NAO)

Le NAO, acronyme de 'North Atlantic Oscillation', est représenté comme le mode de variabilité climatique régional qui domine les hivers de l'hémisphère nord (Lamb and Pepler, 1987 ; Wooling et al, 2010). La découverte de ce phénomène remonte à l'époque de Vikings où les téléconnexions de l'Atlantique étaient évoquées dans un livre norvégien intitulé 'Kings Mirror' vers l'année 1235 (Feldstein & Franzke, 2017). Après quelques siècles, le NAO a devenu un centre d'intérêt des météorologistes et des études d'investigation ont commencé depuis plusieurs années (Walker, 1924 ; Walker et Bliss, 1932). Le NAO est dû à une bascule des masses d'air (pression) au nord de l'Atlantique entre la région de dépression sub-polaire près de l'Islande et la région de l'anticyclone subtropical près des Açores. Cette bascule de pression est connue par le terme oscillation (Antoniadou et al, 2001 ; Wanner et al, 2001 ; Hurrell et al, 2002). L'activité de cette oscillation est mesurée par un indice. Cet indice mesure la différence entre les anomalies de pression normalisée entre la station météorologique, Reykjavik ou Stykkisholmur, proche de la dépression de l'Islande et la station, Delgada ou Gibraltar, proche de la zone des hautes pressions des Açores (Rogers, 1984, Hurrell, 1995). La figure 1 montre l'évolution temporelle de des moyennes annuelles de l'indice NAO entre 1825 et 2019. Les séries sont prises du site de NOAA (Agence américaine d'observation océanique et atmosphérique). A l'échelle interannuelle, cet indice montre une variabilité remarquable mais il possède aussi une certaine persistance à l'échelle décennale par exemple une phase presque négative sur les décennies 1870-1880 et 2000-2010 et un indice assez positif sur la décennie 1900-1910.

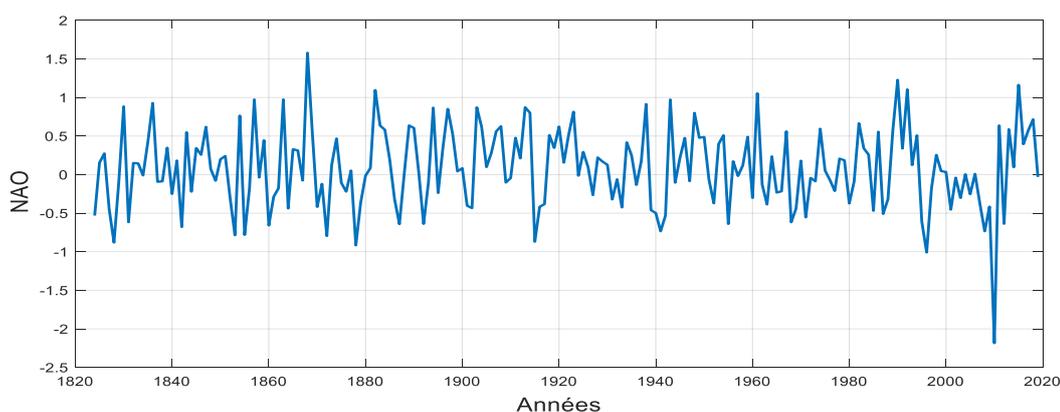


Figure 1 : Série chronologique des moyennes annuelles de l'indice NAO entre 1820 et 2020

Chaque phase reflète une situation climatique puisque cette oscillation agit sur le courant jet connu par 'jet-stream' traversant l'Atlantique. Le jet-stream est un courant atmosphérique rapide, sa vitesse varie entre 200 et 300 Km/h et il peut atteindre le 400 km/h. Il souffle de l'ouest vers l'est et se situe au-dessus de la surface de la terre d'environ 10 à 15 Km (au niveau de la tropopause³). Il résulte de la rencontre entre les masses d'air chaudes d'origine tropicale et les masse d'air froides d'origine polaire (Baray et al., 2017). Les fluctuations de la pression nord atlantique (NAO) ont une forte influence sur l'Amérique du nord, Groenland, l'Europe, l'Afrique du nord et l'Asie du nord spécialement pendant la saison hivernale entre décembre et mars (Hurrell, 1996 ; Benner, 1999). Le NAO est dans la phase positive (noté NAO⁺) si l'anticyclone des Açores et la dépression sub-polaire sont plus forts que la moyenne. Cette phase positive se traduit par un courant-jet plus fort qui passe entre les deux cellules de pressions et se dirige vers le nord de l'Europe comme montre la figure 2 (la figure à droite). L'Europe du nord connaît alors des précipitations et plus fortes que la moyenne, une élévation de température liée aux masses d'air plus chaudes provenant des basses latitudes et une augmentation des tempêtes. Simultanément, la région méditerranéenne et la région nord-américaine connaissent des conditions plus sèches et plus froides. La phase négative de la NAO (noté NAO⁻) résulte d'une dépression subpolaire et un anticyclone tropical plus faibles que la moyenne. Le faible gradient de pression au niveau de l'Atlantique du nord entraîne un ralentissement dans les vents d'ouest et la faible dépression subpolaire permet à l'Europe du nord d'aspirer des masses d'air polaires froides et cela résulte une diminution des précipitations. Le faible anticyclone tropical de la phase négative redirige le courant-jet vers le nord (Groenland) puis il redescend vers l'Europe de sud et le nord de l'Afrique et influence les hivers de ces régions qu'ils deviennent plus humides et chauds (Rogers and van Loon, 1979 ; [Marshall et al, 2001](#)). L'influence de NAO- est schématisée dans la figure 2.

³ La tropopause est une mince couche atmosphérique qui sépare la troposphère (la couche inférieure de l'atmosphère) et la stratosphère (la deuxième couche de l'atmosphère). Cette couche est caractérisée par une chute remarquable dans le profil vertical de la température de la masse d'air.

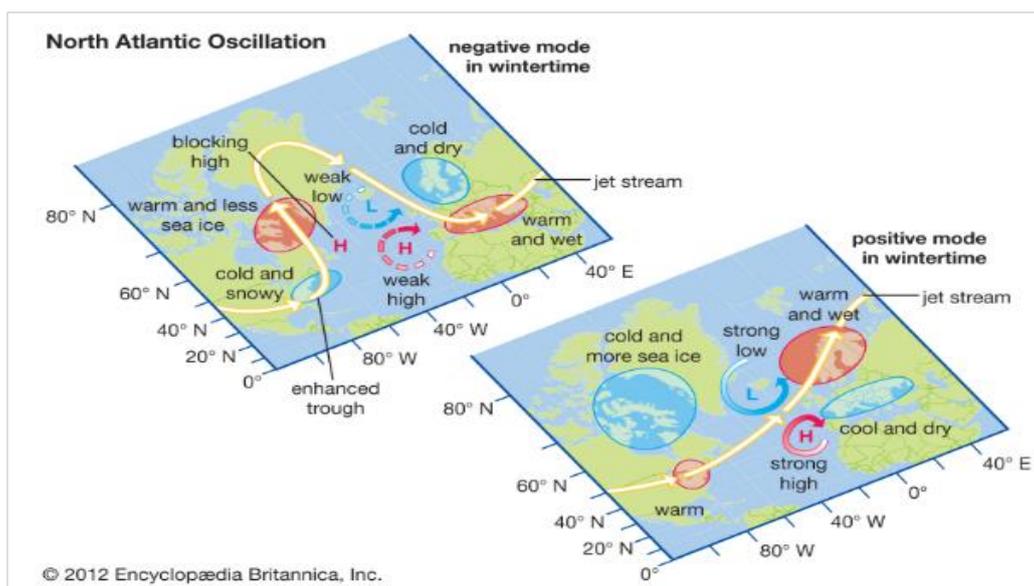


Figure 2 : Le trajet typique du jet-stream du front polaire pendant les phases négatives et positives de l'oscillation nord Atlantique (source : <https://www.britannica.com/science/North-Atlantic-Oscillation>).

2. L'Oscillation Arctique (AO)

L'Oscillation Arctique est un mode de variabilité à grande échelle appelé aussi mode annulaire. L'indice AO est obtenu à partir de premier mode d'analyse EOF (la fonction empirique orthogonale) des anomalies de la hauteur géopotentielle de 1000 millibars à l'hémisphère nord entre les latitudes 20°N et 90°N. Cet indice est fortement corrélé avec le NAO (Thompson et Wallace, 1998, 2000). La figure 3 montre la variation temporelle de l'indice AO entre 1950 et 2020 fournie par l'agence américaine d'observation océanique et atmosphérique.

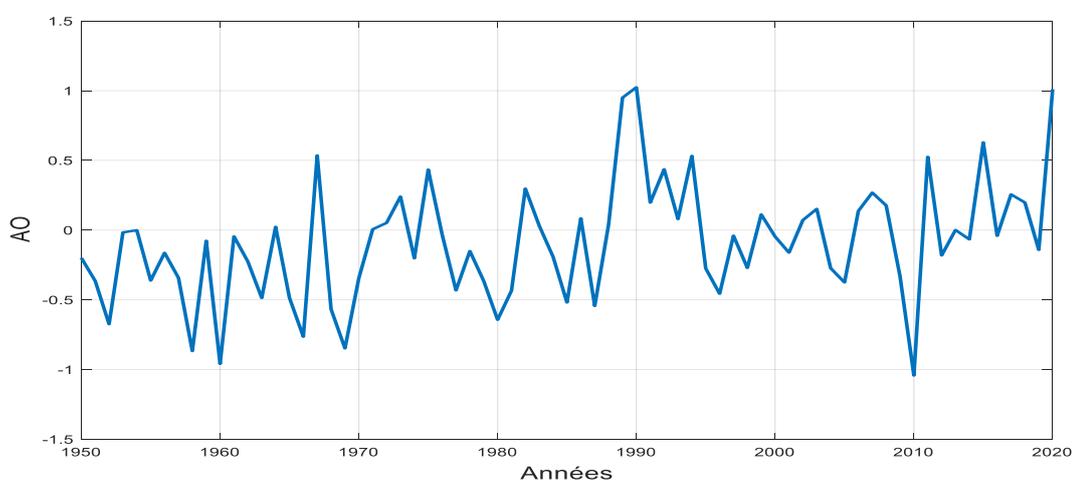


Figure 3 : Série chronologique des moyennes annuelles de l'indice AO entre 1950 et 2020

De même l'oscillation arctique comprend une phase positive et une phase négative. Pendant la phase positive, un anneau de vent de forte vitesse tourne dans le sens contraire des aiguilles d'une montre autour du pôle nord environ 55°N de latitude et les masses d'air les plus froides restent dans la région polaire. Pendant la phase négative, la vitesse d'anneau du vent diminue et permet aux masses d'air polaires de pénétrer plus facilement aux moyennes latitudes et d'augmenter les tempêtes. Plusieurs études ont analysé l'effet de basculement de cet indice entre les phases positives et négatives sur l'hémisphère nord incluant le Nord de l'Afrique ([Wang and Chen, 2010](#) ; [Zuo et al., 2015](#)).

3. Les oscillations méditerranéennes de l'ouest (WeMO)

L'Oscillation Méditerranéenne de l'ouest (WeMO) est un mode de variabilité régionale. Il est défini dans le cadre synoptique de la méditerranée occidentale. L'indice de WeMO est la différence de pression de surface standardisée enregistrée entre deux points d'observations spécifiques le Padua (45.40°N , 11.48°E) au nord de l'Italie et San Fernando, Cádiz (36.28°N , 6.12°W) au sud-ouest de l'Espagne. Le premier point est influencé par l'anticyclone de l'Europe centrale et le deuxième par l'anticyclone des Açores (Martín-Vide and López-Bustins, 2006). La figure 4 présente la série chronologique des moyennes annuelles de WeMO entre les années 1821 et 2018 (Source : http://www.ub.edu/gc/documents/Web_WeMOi.txt).

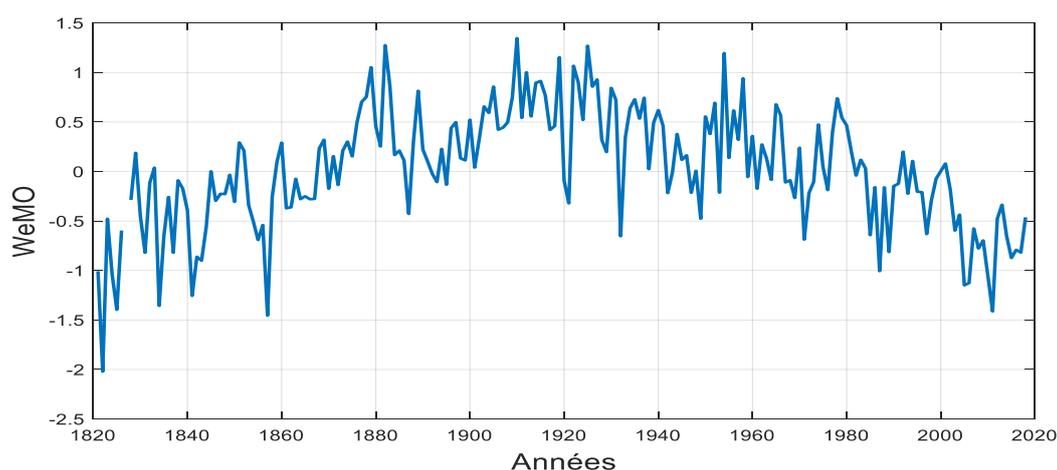


Figure 4 : Série chronologique des moyennes annuelles de l'indice WeMO entre 1821 et 2018

La phase positive du WeMO coïncide avec l'anticyclone des Açores et à une basse pression dans le Golf de Gênes. Pendant cette phase, les vents dominants en méditerranée occidentale sont typiquement ouest et nord-ouest d'origine de nord Atlantique. La phase négative coïncide avec l'anticyclone de l'Europe centrale et une dépression dans le golfe de Cadix (Portugal).

Le vent dominant est de direction est-ouest et des flux humides traversent la Méditerranée accompagnés des précipitations parfois torrentielles (López-Bustins et al., 2020, Fig.5).

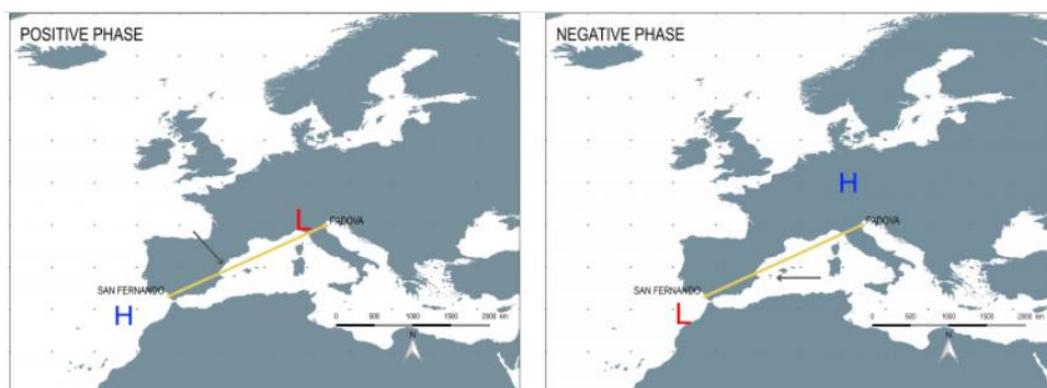


Figure 5 : Influence de WeMO pendant la phase positive et négative (<http://www.ub.edu/gc/wemo/>). (H : haute pression, L : basse pression)

4. L'oscillation méditerranéenne (MO)

L'oscillation méditerranéenne (MO) représente aussi un mode de variabilité régional caractéristique du bassin méditerranéen. L'indice MO est défini comme la différence de pression normalisée entre deux points de mesures l'un situé à l'est du méditerranée et l'autre dans et l'autre à l'ouest (Alger / Le Caire, Padoue / Cadix et autres). L'indice MO considéré dans cette étude est obtenu de la différence de pression normalisée entre Alger (36.4 °N, 3.1 °E) et le Caire (30.1 °N, 31.4 °E) (Conte et al., 1989 ; Palutikof et al., 1996 ; Criada-Aldeanueva and Soto-Navarro, 2013). La figure 6 présente les moyennes annuelles de l'indice MO entre 1948 et 2019 calculé à partir des données journalières existant dans ce site (<https://crudata.uea.ac.uk/cru/data/moi/>).

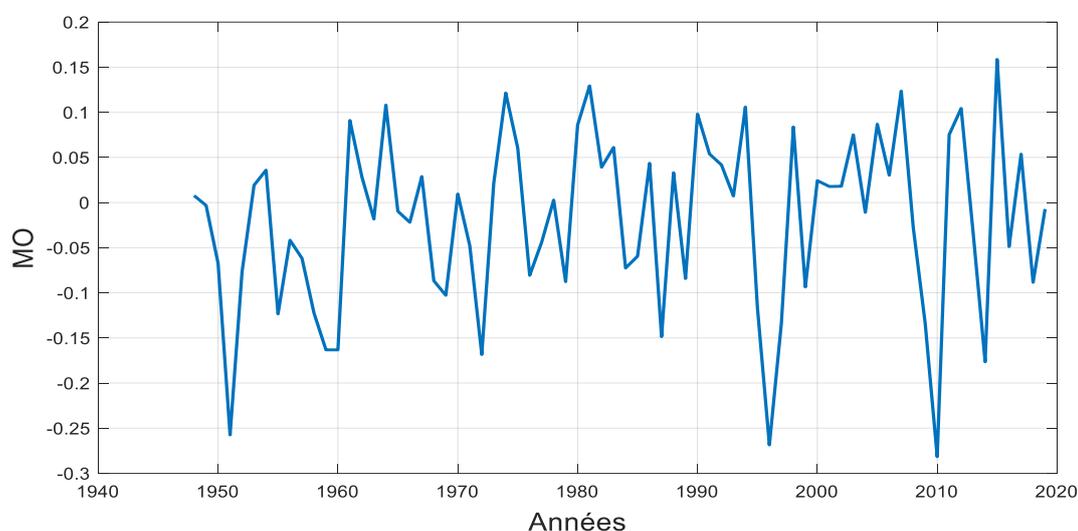


Figure 6 : Série chronologique des moyennes annuelles de l'indice MO entre 1948 et 2019

5. Oscillation Australe El Niño (ENSO)

El Niño est un mot d'origine espagnol qui signifie "l'Enfant Jésus". Il s'agit d'un phénomène climatique qui apparaît tous les trois à six ans pendant la période de Noël au sud de l'océan pacifique et s'étend jusqu'au mois d'avril. Ce phénomène résulte d'une interaction entre l'atmosphère et l'océan (Philander, 1990). Pendant les années normales, une différence de pression atmosphérique existe dans le Pacifique de Sud entre les côtes est (Equateur, Pérou) de haute pression et les côtes ouest (Indonésie et l'Australie) de basse pression. Les vents alizés⁴ soufflent d'est en ouest (des hautes pressions vers les basses pressions). En traversant le Pacifique, les alizées se charge d'humidité et les eaux chaudes (température > 28 °) s'accumule à l'ouest du pacifique et provoque de fortes précipitations sur toute la région (L'Indonésie, Philippines, la Nouvelle Guinée, le nord de l'Australie, etc). Dans l'autre côté sud-américaines, les eaux froides remontent de la profondeur et amènent de la sécheresse sur les côtes du Chili, du Pérou et de l'Equateur. Lors d'un phénomène El Niño, les alizées perdent de leur force sur le centre et l'ouest du pacifique suite à une augmentation de pression près de l'Australie et l'Indonésie. Cette perturbation a pour conséquence le déplacement des pluies de l'ouest vers le Pacifique centrale et des cyclones se forment. De même, les eaux chaudes ne s'accumulent plus à la côte ouest et les eaux froides ne remontent plus vers la surface au niveau de la côte est du Pacifique. Au cours de cette perturbation, L'Australie et ses environs connaissent des périodes de grande sècheresse avec une augmentation des risques d'incendies. Simultanément, la côte Est du Pacifique connaît des fortes précipitations suite au réchauffement des eaux au large du Pérou. Aux deux côtes du Pacifique Sud, si la pression augmente à l'ouest elle diminue à l'est et inversement pour cela on parle d'une oscillation australe (Southern Oscillation) et si elle est associée à un phénomène El Niño on l'appelle ENSO (El Niño Southern Oscillation). El Niño est considérée la phase chaude de l'ENSO. La phase froide est appelée La Niña (Iwakiri and Watanabe, 2020). L'ENSO est le mode de variabilité climatique le plus étudié dans la littérature (Horel and Wallace, 1981 ; Rasmussen and Wallace, 1983 ; Lau, 1997 ; Dijkstra, 2006) La figure 7 illustre la situation normale du Pacifique, la phase chaude de l'ENSO (El Niño), la phase froide (La Niña) et la circulation océanique pendant les trois phases.

⁴ Les alizés sont des vents réguliers qui soufflent le long de l'année de la part et d'autre de l'équateur dans la zone intertropicale. Ce sont des vents à vitesse modérées qui ne dépasse pas 20 Km/h qui soufflent dans les couches basses de l'atmosphère.

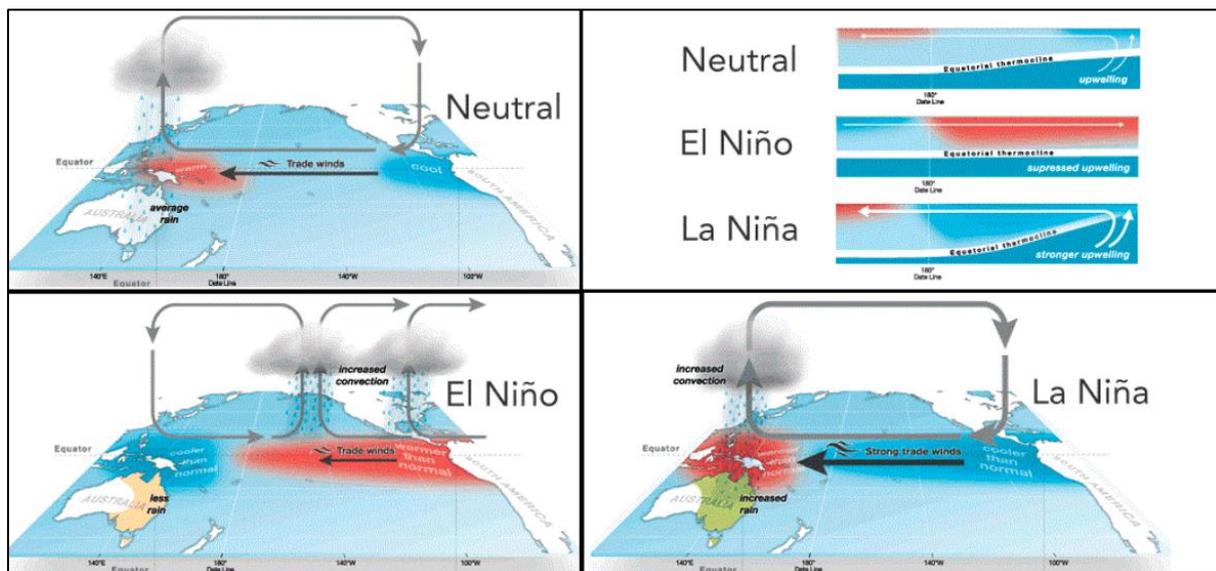


Figure 7 : Illustration des changements au Pacifique associés aux différentes phases du phénomène ENSO (source : <https://iri.columbia.edu/our-expertise/climate/enso/enso-essentials/>)

Plusieurs indices sont utilisés pour étudier et analyser l'ENSO. Le plus ancien est créé par Walker et Bliss en 1932 appelé l'indice d'oscillation australe (SOI). Cet indice mesure la différence de pression standardisée entre deux stations météorologiques. L'une à Darwin en Australie du côté ouest du Pacifique et l'autre à Tahiti en Polynésie française au Pacifique centrale. Cet indice existe à l'échelle mensuelle sur le site de l'agence américaine d'observation océanique et atmosphérique. La figure 8 présente les moyennes annuelles de SOI entre les années 1866 à 2019. Les valeurs positives de SOI (Fig.8) sont associées aux années normales et les valeurs négatives représentent les années El Niño. Depuis le 1980, la fréquence d'apparition de la phase el Niño a augmenté considérablement.

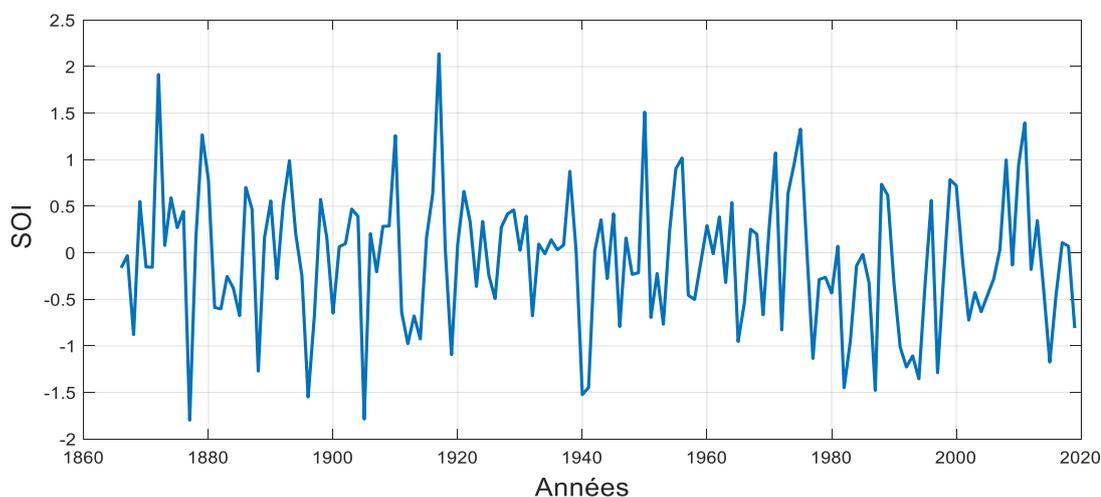


Figure 8 : Série chronologique des moyennes annuelles de l'indice SOI entre 1866 et 2019

6. Oscillation Décennale du Pacifique (PDO)

L'oscillation décennale du pacifique (PDO : Pacific Decadal Oscillation) est souvent définie comme un mode de variabilité à long terme de type El Niño du nord pacifique. Malgré la ressemblance, les deux phénomènes ont des caractéristiques temporelles et spatiales différentes. La PDO a deux phases chaude et froide parallèlement à l'ENSO (Mantua & Hare, 2002). La PDO est à sa phase chaude (PDO positive), si la température à la surface de la mer SST est anormalement fraîche à l'intérieur du Pacifique de Nord et chaude le long de l'Amérique du Nord occidentale (Fig. 9a). Cette phase de PDO accompagnée par une élévation de la SST du pacifique équatoriale et peut rehausser un épisode El Nino en amplifiant ses conséquences comme il peut affaiblir un épisode La Niña. La phase froide (PDO négative) est l'opposée de la précédente et semblable à La Niña (Fig. 9a). L'indice de l'oscillation décennale du Pacifique est défini comme la première composante principale de la variabilité des SSTs mensuelles du Pacifique Nord polarisé à 20° N. La PDO est un phénomène qui persiste dans le temps et montre une certaine périodicité. La figure 9b présente les moyennes annuelles de l'indice PDO entre 1900 et 2018 fourni par JIDAO ('Joint Institute for the Study of the Atmosphere and Ocean'). La période allant de 1900 à 1945 est caractérisée par la dominance de la phase chaude de PDO cependant la phase froide du PDO se trouve dominante entre 1945 et 1975 et à partir de 1975 un deuxième inversement de phase est observé.

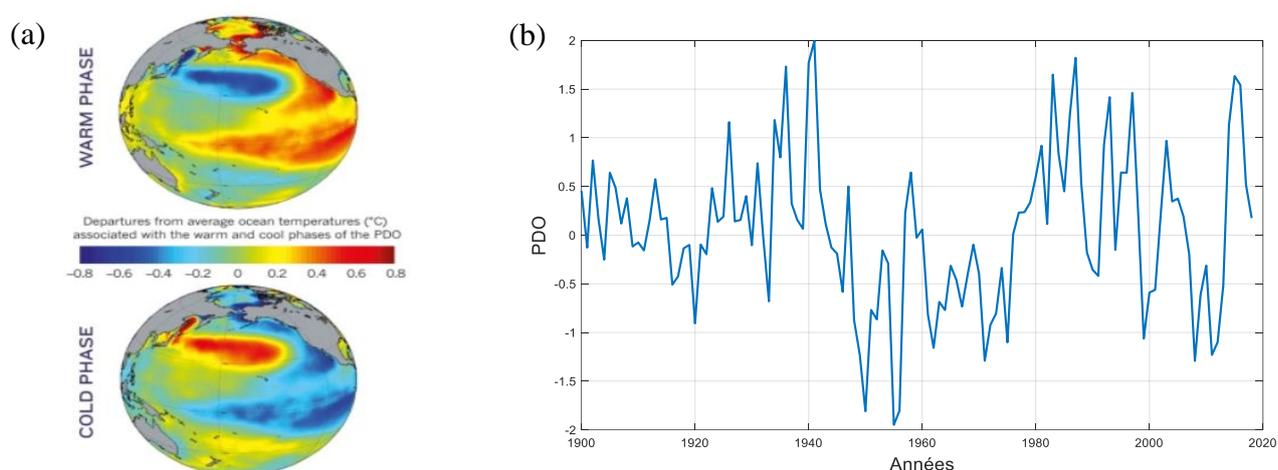


Figure 9 : (a) Illustration de la phase chaude et la phase froide de PDO (source : <https://ddecoverley.wordpress.com/2015/07/12/pdo-flip-may-signal-onset-of-plutonic-era/>).

(b) Série chronologique des moyennes annuelles du PDO entre 1900 et 2018 (source : <http://research.jisao.washington.edu/pdo/PDO.latest>)

Références :

- Baray, J.-L., Pointin, Y., Baelen, J. V., Lothon, M., Campistron, B., Cammas, J.-P., . . . Tridon, F. (2017, November). Case Study and Climatological Analysis of Upper-Tropospheric Jet Stream and Stratosphere–Troposphere Exchanges Using VHF Profilers and Radionuclide Measurements in France. *Journal of applied meteorology and climatology*, 56(11), 3081–3097. doi:10.1175/JAMC-D-16-0353.1
- Benner, T. C. (1999, March). Central England temperatures: long-term variability and teleconnections. *International Journal of Climatology*, 19(4), 391-403. doi:10.1002/(SICI)1097-0088(19990330)19:4<391::AID-JOC365>3.0.CO;2-Z
- Conte, M., Giuffrida, A., & Tedesco, S. (1989). Mediterranean Oscillation: Impact on Precipitation and Hydrology in Italy. Dans C. o. Water (Éd.), *Publications of the Academy of Finland: Helsinki*.
- Criado-Aldeanueva, F., & Soto-Navarro, F. (2013). The Mediterranean Oscillation Teleconnection Index: Station-Based versus Principal Component Paradigms. *Advances in Meteorology*. doi:10.1155/2013/738501
- Feldstein, S. B., & Franzke, C. L. (2017). Atmospheric Teleconnection Patterns. Dans C. L. Franzke, & T. J. O'Kane (Éds.), *Nonlinear and Stochastic Climate Dynamics* (pp. 54-104). Cambridge: University Press. doi:10.1017/9781316339251.004
- Horel, J. D., & Wallace, J. M. (1981). Planetary-Scale Atmospheric Phenomena Associated with the Southern Oscillation. *Monthly Weather Review*, 109(4), 813-829. doi:10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2
- Hurrell, J. (1995). Decadal trends in the North-Atlantic Oscillation: regional temperatures and precipitation. *Science*, 269(5224), 676–679. doi:10.1126/science.269.5224.676
- Hurrell, J. W. (1996). Influence of variations in extratropical wintertime teleconnections on northern hemisphere temperature. *Geophysical Research Letters*, 23(6), 665-668. doi:10.1029/96GL00459
- Hurrell, J., Kushnir, Y., Ottersen, G., & Visbeck, M. (2002). An overview of the North Atlantic Oscillation. Dans W. Hurrell, Y. Kushnir, G. Ottersen, & M. Visbeck, *The North Atlantic Oscillation: Climate Significance and Environmental Impact* (Vol. 134, pp. 1-36).
- Iwakiri, T., & Watanabe, M. (2020). Multiyear La Niña Impact on Summer Temperature over Japan. *Journal of the Meteorological Society of Japan Ser II*, 98(6), 1245-1260. doi:10.2151/jmsj.2020-064
- Lamb, P. J., & Pepler, R. A. (1987). North Atlantic Oscillation: concept and an application. *Bulletin of the American Meteorological Society*, 86(10), 1218-1225. doi:10.1175/1520-0477(1987)068<1218:NAOCAA>2.0.CO;2
- Lau, N.-C. (1997). Interactions between global SST anomalies and the midlatitude atmospheric circulation. *Bulletin of the American Meteorological Society*, 78(1), 21–34. doi:10.1175/1520-0477(1997)078<0021:IBGSAA>2.0.CO;2
- López-Bustins, J. A., Arbiol-Roca, L., Martín-Vide, J., Barrera-Escoda, A., & Prohom, M. (2020). Intra-annual variability of the Western Mediterranean Oscillation (WeMO) and occurrence of extreme torrential precipitation in Catalonia (NE Iberia). *Natural*

- Hazards and Earth System Sciences*, 20(9), 2483-2501. doi:10.5194/nhess-20-2483-2020
- Mantua, N. J., & Hare, S. R. (2002). The Pacific Decadal Oscillation. *Journal of Oceanography*, 58, 35-44. doi:10.1023/A:1015820616384
- Marshall, J., Kushnir, Y., Battisti, D., Chang, P., Czaja, A., Dickson, R., . . . Visbeck, M. (2001). North Atlantic climate variability: phenomena, impacts and mechanisms. *International Journal of Climatology*, 21(15), 1863–1898. doi:10.1002/joc.693
- Martin-Vide, J., & Lopez-Bustins, J. (2006). The western Mediterranean oscillation and rainfall in the Iberian Peninsula. *International Journal of Climatology*, 26(11), 1455-1475. doi:10.1002/joc.1388
- Palutikof, J., Conte, M., Casimiro Mendes, J., Goodess, C., & Espirito Santo, F. .. (1996). Climate and climate change. Dans C. J. Brandt, & J. B. Thornes (Éds.), *Mediterranean Desertification and Land Use* (pp. 43-86). London: John Wiley and Sons.
- Philander, S. (1989). El Nino, La Nina, and the Southern Oscillation. *International Geophysics Series*, 46, 293.
- Rasmussen, E. M., & Wallace, J. M. (1983). Meteorological Aspects of the El Niño/Southern Oscillation. *Science*, 222(4629), 1195-1202. doi:10.1126/science.222.4629.1195
- Rogers, J. (1984). The Association between the North Atlantic Oscillation and the Southern Oscillation in the Northern Hemisphere. *Monthly Weather Review*, 112(10), 1999–2015. doi:10.1175/1520-0493(1984)112<1999:TABTNA>2.0.CO;2
- Rogers, J., & Van Loon, H. (1979). The seesaw in winter temperatures between Greenland and northern Europe. Part II: some oceanic and atmospheric effects in middle and high latitudes. *Monthly Weather Review*, 107(5), 509–519. doi:10.1175/1520-0493(1979)107<0509:TSIWTF>2.0.CO;2
- Thompson, D. W., & Wallace, J. M. (2000). Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability. *Journal of Climate*, 13(5), 1000–1016. doi:10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2
- Trenberth, K. E., Caron, J. M., Stepaniak, D. P., & Worley, S. (2002). Evolution of El Niño–Southern Oscillation and global atmospheric surface temperatures. *Journal of Geophysical Research: Atmospheres*, 107(D8), AAC 5-1-AAC 5-17. doi:10.1029/2000JD000298
- Walker, G. (1924). Correlation in Seasonal Variations of Weather, IX. A Further Study of World Weather. *Memoirs of the India Meteorological Department*, 24(9), 275-332.
- Walker, G., & Bliss, E. (1932). World Weather V. *Memoirs of the Royal Meteorological Society*. 4, 53-84.
- Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, 109(4), 784–812. doi:10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2
- Wang, L., & Chen, W. (2010). Downward Arctic Oscillation signal associated with moderate weak stratospheric polar vortex and the cold December 2009. *Geophysical Research Letters*, 37(9). doi:10.1029/2010GL042659

- Wanner, H., Brönnimann, D., Casty, S., Gyalistras, D., Luterbacher, J., Schmutz, C., & Stephenson, D. (2001). North Atlantic Oscillation-concepts and studies. *Surveys in Geophysics*, 22(4), 321-381. doi:10.1023/A:1014217317898
- Woollings, T., Hannachi, A., Hoskins, B., & Turner, A. (2010). A Regime View of the North Atlantic Oscillation and Its Response to Anthropogenic Forcing. *Journal of climate*, 23(6), 1291–1307. doi:10.1175/2009JCLI3087.1
- Zuo, J., Ren, H.-L., & Li, W. (2015). Contrasting Impacts of the Arctic Oscillation on Surface Air Temperature Anomalies in Southern China between Early and Middle-to-Late Winter. *Journal of Climate*, 28(10), 4015-4026. doi:10.1175/JCLI-D-14-00687.1

Annexe B : Analyse en composantes principales ACP

1. Description des données

L'ACP s'applique sur un tableau rectangulaire de mesure où les colonnes représentent les variables quantitatives (valeurs numériques) et les lignes figurent les unités appelées aussi les individus sur lesquels les mesures sont effectuées ([Lebart et al., 2000](#)).

Soit le tableau X de données de n lignes et p colonnes et x_{ik} est la valeur de la variable k pour l'individu i .

$$X = \begin{pmatrix} x_{11} & \dots & x_{1k} & \dots & x_{1p} \\ \vdots & & & & \\ x_{i1} & & x_{ik} & & x_{ip} \\ \vdots & & & & \\ x_{n1} & \dots & x_{nk} & \dots & x_{np} \end{pmatrix}$$

Cette matrice X forme un nuage de n individus dans un espace de p dimension ou bien un nuage de p variables dans un espace de n dimension. L'ACP se fait en projetant ces nuages, des points dans un sous-espace de dimension s tel que $s \leq p$ (par exemple droite, plan ou autres). Mais ce sous-espace doit satisfaire certain critère afin de donner une bonne visualisation au nuage des points en préservant la forme de nuage de l'espace initiale ([Escofier et Pagès, 2008](#)).

On admet que la matrice X subi un centrage au préalable pour que le centre de gravité des individus coïncide avec l'origine du repère.

2. Projection de nuage des individus

L'étude des individus se base sur l'évaluation de leur ressemblance deux à deux c'est-à-dire deux individus semblables possèdent des valeurs proches pour l'ensemble des variables. Cette ressemblance est mesurée couramment par la distance euclidienne classique. Le calcul de cette distance sert à préciser la forme du nuage et à détecter les directions d'allongement remarquables. Dès que la dimension p de l'espace dépasse le 3, il devient impossible de présenter le nuage des individus d'où vient l'idée de l'ACP. Le principe est de trouver un système d'axes unitaires u_k ($\|u_k\| = 1$) de projection des individus où la moyenne des carrées de distances des points projetés soit maximales. Cela signifie que ces axes restituent fidèlement la forme de nuage de manière qu'on peut distinguer la variabilité entre les données. Les vecteurs u_k de la nouvelle base sont appelés des axes principaux en ACP et les nouvelles coordonnées sont des composantes principales C_k qui remplacent les anciennes variables ([Lebart et al., 2000](#)).

Les projections des individus sur un axe u_k sont définis par le produit scalaire (Eq.1) :

$$C_k = Xu_k \quad \text{VI-1}$$

La somme des carrés des projections est (Eq. 2) :

$$C'_k C = u'_k X' X u_k \quad \text{VI-2}$$

La détermination de u_k est basée sur la maximisation de la somme des carrés des points projetés (Eq. 3) :

$$\begin{cases} \max(u'_k X' X u_k) \\ u'_k u_k = 1 \end{cases} \quad \text{VI-3}$$

Ce genre de problème d'optimisation (une fonction à plusieurs variables liées à une contrainte ($u'_k u_k = 1$) où les inconnues sont les composantes de vecteur u_k) peut être résolu par la méthode de multiplicateurs de Lagrange (Eq. 4).

$$l = u'_k X' X u_k - \lambda(u'_k u_k - 1) \quad \text{VI-4}$$

En appliquant une dérivation partielle par rapport à u_k et par rapport au multiplicateur de Lagrange λ , on trouve après simplification l'Equation 5 :

$$\begin{aligned} X' X u_k &= \lambda_k u_k \\ u'_k u_k &= 1 \end{aligned} \quad \text{VI-5}$$

La matrice $X' X$ avec un facteur de pondération $\frac{1}{n}$ ou bien $\frac{1}{n-1}$ n'est que la matrice S (p×p) de variance-covariance (variance au diagonal et covariance hors diagonale) des p variables étudiées.

$$S = \begin{pmatrix} \text{var}(X_1) & \dots \text{cov}(X_1, X_k) \dots & \text{cov}(X_1, X_p) \\ \vdots & & \\ \text{cov}(X_k, X_1) & \text{var}(X_k) & \text{cov}(X_k, X_p) \\ \vdots & & \\ \text{cov}(X_p, X_1) & \dots \text{cov}(X_p, X_k) \dots & \text{var}(X_p) \end{pmatrix}$$

Avec Eq.6 et Eq.7 sont respectivement les équations de variance d'un vecteur X_k et de covariance des vecteurs X_k et X_p . \bar{x}_k et \bar{x}_p sont les moyennes des vecteurs X_k et X_p

$$\text{var}(X_k) = \frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \quad \text{VIVI-6}$$

$$\text{cov}(X_p, X_k) = \frac{1}{n} \sum_{i=1}^n (x_{ip} - \bar{x}_p)(x_{ik} - \bar{x}_k) \quad \text{VIVI-7}$$

Le premier vecteur propre u_1 de la matrice variance covariance S est le vecteur qui donne la variance maximale des données initiales X. Un vecteur propre représente la direction dans laquelle la matrice agit et la valeur propre λ est un facteur multiplicatif associé à cette direction. Notons que la matrice variance-covariance S est symétrique et positive. Par conséquent, ces valeurs propres seront toujours positives ou nulles et ces vecteurs propres sont orthogonaux deux à deux ($u_k u_{(k' \neq k)} = 0$). La variance de u_k déterminée par l'équation 8 :

$$\frac{1}{n}(u_k' X' X u_k) = \frac{1}{n}(u_k' \lambda_k u_k) = \frac{1}{n} \lambda_k \quad \text{VIVI-8}$$

La valeur propre λ_k représente la variance maximale des projections sur l'axe k. L'ordre d'un axe u_k se fait selon l'importance de sa valeur propre λ_k associée. Les valeurs propres sont triées d'un ordre décroissant ($\lambda_1 \geq \lambda_2 \geq \dots \lambda_k \geq \dots \geq \lambda_p$). Le plan qui maximise la variance des projections est constitué des deux premiers vecteurs propres. Le plus souvent la part de variance expliquée par un axe k est exprimé en pourcentage (Eq. 9) :

$$\frac{100 \lambda_k}{\sum_1^p \lambda_k} \% \quad \text{VIVI-9}$$

La figure 1 schématise les deux premiers axes principaux u_1 et u_2 d'un nuage des individus projeté par rapport aux variables 1 et 2. C_1 et C_2 sont les composantes principales. Ces composantes sont des combinaisons linéaires des variables initiales.

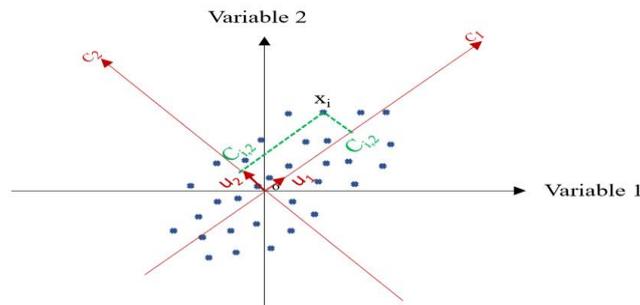


Figure VIVI-12 Schématisation des deux premiers axes principaux d'un nuage des individus

3. Projection de nuage des variables

Dans la projection du nuage de variables, on cherche un sous-espace dans lequel la somme des carrées des projections des variables est maximale. Le même concept de la projection des individus est appliqué dans l'espace des variables dans le but de trouver les axes unitaires v_i ($\|v_i\| = 1$). On trouve alors l'équation 10 qui représente la projection des variables sur un axe principal d'ordre i :

$$\begin{aligned} XX' v_i &= \gamma_i v_i & \text{VI-10} \\ v_i' v_i &= 1 \end{aligned}$$

v_i Représente le vecteur propre de la matrice XX' associé à la valeur propre γ_i

4. Relations duales et formules de transition en ACP

En réalité, le nuage des individus et le nuage des variables sont deux représentations à une même base de données. Il existe une forte relation entre les deux espaces appelées relation duale. Les valeurs propres des deux matrices XX' et $X'X$ sont égaux $\gamma = \lambda$ (Lebart et al., 2000). Les vecteurs propres des individus et des variables sont liés par des formules de transition exprimées par Eq.11 et Eq. 12 :

$$u_i = \frac{1}{\sqrt{\lambda_i}} X'v_i \quad \text{VI-11}$$

$$v_i = \frac{1}{\sqrt{\lambda_i}} Xu_i \quad \text{VI-12}$$

En se basant sur ces relations, il est inutile de calculer les vecteurs et les valeurs propres dans les deux espaces (des individus et des variables). Il suffit de les calculer soit pour la matrice XX' ou $X'X$ ([Lebart et al., 2000](#)). Sous la forme matricielle, les formules de transition deviennent Eq. 13 et Eq. 14 :

$$U = X'V\Lambda^{-1/2} \quad \text{VI-13}$$

$$V = XU\Lambda^{-1/2} \quad \text{VI-14}$$

U et V sont les matrices des vecteurs propres des deux espaces d'individus et des variables, Λ est une matrice diagonale des valeurs propres.

5. Choix du nombre des facteurs principaux

Le nombre des composantes principales à retenir pour l'interprétation dépend du contexte de l'étude. Le plus souvent les composantes ayant des faibles valeurs propres sont éliminées pour réduire le nombre des variables. Il existe dans la littérature plusieurs critères pour choisir le nombre optimal des facteurs principaux les plus connus sont :

- Critère de coude : [Cattell \(1966\)](#) propose d'étudier la courbe décroissante des valeurs propres λ_k (scree plot) et détecter le décrochement (coude) d'une stabilisation de la variation de λ_k . Cette technique est nuancée. Généralement, on retient les composantes principales jusqu'à la valeur du coude. Dans le cas où cette valeur est très faible, on peut l'éliminer.
- Critère de Kaiser – Guttman : selon ce critère, les axes retenus sont ceux qui ont une variance supérieure à la variance moyenne ([Conway and Huffcutt, 2003](#)).
- Proportion de variance expliquée : Il s'agit de fixer à priori une proportion de variance à expliquer. Dans la littérature, certains proposent un seuil de 80 % et d'autres 70% ([Bourque et al, 2006](#)). En utilisant cette approche, on peut dans certains cas retenir des facteurs qui expliquent un faible pourcentage de variance.

L'interprétabilité des axes peut également être un élément à prendre en compte pour retenir ou non axe ayant une proportion de variance faible.

6. Qualité de représentations des individus et des variables

Cette mesure nous indique la proximité d'un individu ou d'une variable par rapport au vecteur ou plan factoriel retenu. La qualité de représentation est le cosinus au carré de l'angle θ entre le vecteur de l'individu (ou variable) et sa projection sur l'axe ou le plan considéré. Les deux

équations 15 et 16 représentent respectivement la qualité de représentation d'un individu i sur l'axe k et d'une variable j sur l'axe k .

$$q_{lt}(\text{individu } i) = \cos^2(\theta_{ik}) = \frac{c_{ik}^2}{x_i^2} \quad \text{VI-15}$$

$$q_{lt}(\text{variable } j) = \cos^2(\theta_{jk}) = \frac{c_{jk}^2}{x_j^2} \quad \text{VI-16}$$

Si l'angle θ est proche de 0, la qualité de représentation de l'élément étudié sera proche de 1. Cet élément, individus ou variable, est considéré bien représenté par rapport à l'axe k (parce qu'il est proche de l'axe k) et inversement.

7. Contribution des individus et des variables

La mesure de contribution permet de quantifier l'importance de chaque individu et variable à la formation d'un vecteur propre ([Abdi et al, 2013](#)). La contribution 'ctr' est définie par la coordonnée au carré d'un individu ou d'une variable sur un vecteur propre k divisée par la valeur propre λ_k associée à ce vecteur propre (Eq. 17).

$$ctr_k(\text{individu } i) = \frac{p_i c_{ik}^2}{\lambda_k} \quad \text{VI-17}$$

p_i est le poids de l'individu i , c_{ik} est la projection de l'individu i sur le facteur principal k . En pratique, on retient les individus dont la contribution est supérieure à la contribution moyenne dans l'interprétation des individus.

La contribution d'une variable j dans la construction du facteur principal k est égale à (Eq. 18) :

$$ctr_k(\text{variable } j) = \frac{c_{jk}^2}{\lambda_k} \quad \text{VI-18}$$

8. Analyse en composantes principales normée

L'application de l'ACP sur des variables qui ne disposent pas des mêmes ordres de grandeurs risque de biaiser les résultats étant donné que cette technique est très sensible à la variance des variables initiales. Pour cette raison, les variables de la matrice X doivent être transformées, à priori, à des échelles comparables. Donc en plus du centrage, les variables de la matrice X subissent aussi une réduction en divisant les données centrées par l'écart type suivant l'équation 19 et les données deviennent entrées-réduites. Dans ce cas et pour chaque variable, la moyenne est égale à 0 et l'écart-type est égale à 1 :

$$x_{ik}^{normé} = \frac{(x_{ik} - \bar{x}_k)}{\sigma_k} \quad \text{VI-19}$$

Avec $\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$ est la moyenne de la variable k et $\sigma_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}$ est l'écart type de la variable k.

Cette transformation, appelée normalisation, n'a pas d'incidence sur la variabilité des individus et des variables. Cette approche qui est couramment utilisée est appelée analyse en composante principale normée ([Duby and Robin, 2006](#)). Les particularités de cette approche sont :

- La matrice de variance-covariance des données normée S est égale une matrice des corrélations R .
- La variance totale qui est la trace de la matrice de corrélation est égale à p (dimension de l'espace de représentation initial) est également égale à la somme des valeurs propres λ_k de S
- Les coordonnées des variables sur les facteurs principaux sont égales aux corrélations des variables avec les composantes principales correspondantes.

9. Récapitulation des étapes de l'ACP et interprétations

Pour récapituler, l'application d'un ACP suit quelques étapes décrites dans la figure 2.

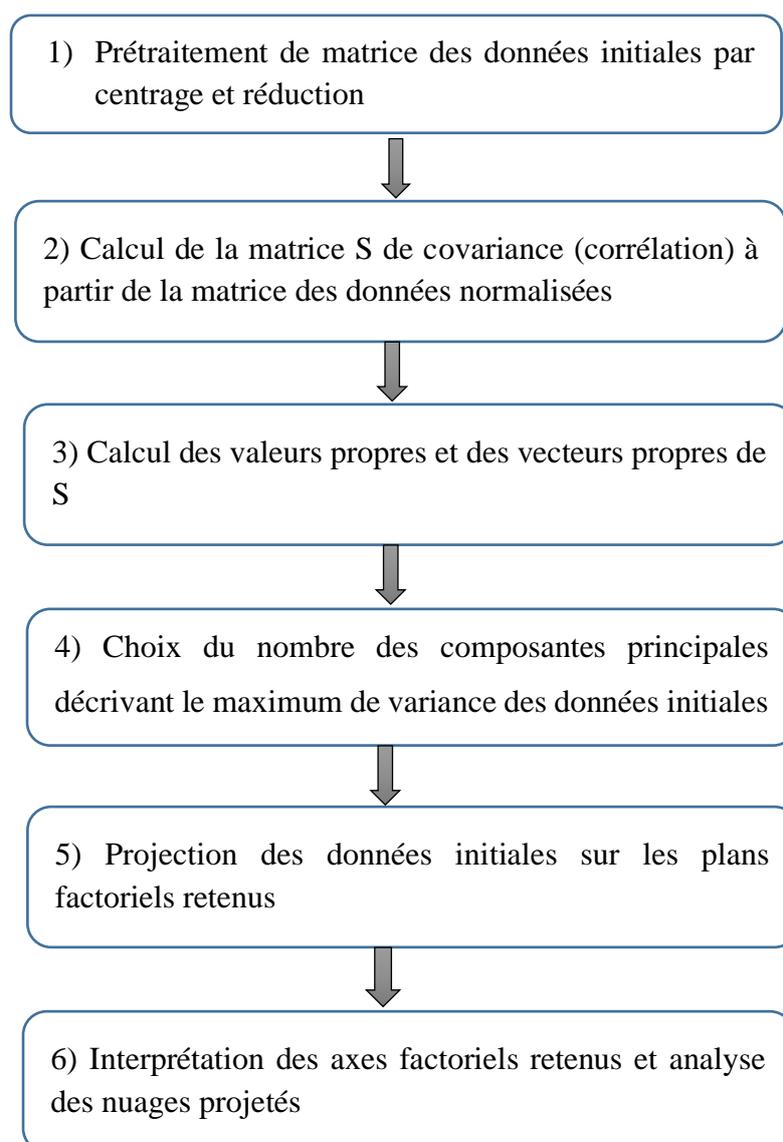


Figure VI-13 les étapes d'application de l'analyse en composantes principales

L'application de l'ACP commence par le prétraitement de la matrice des données initiales soit par un centrage ou le plus souvent par une normalisation des variables.

La deuxième étape est le calcul de la matrice de covariance ou de corrélation (dans l'ACP normée). Cette matrice permet d'étudier les relations entre les variables deux à deux et de vérifier l'existence des informations redondantes dans le cas d'une forte corrélation.

La troisième étape consiste à calculer les vecteurs et les valeurs propres de la matrice de corrélation (covariance) pour calculer les composantes principales par la suite. Les vecteurs propres sont les directions des axes principaux où il y a le maximum de variance et les valeurs propres sont les coefficients associés à ces axes. Le calcul de ces derniers est basé sur un concept d'algèbre linéaire.

Ces composantes sont des nouvelles variables, décorrélées entre elles, formées des combinaisons linéaires des variables initiales. La première composante explique de maximum possible d'informations des données initiales (exprimé par le maximum de variance), la deuxième explique le maximum de variances restantes et ainsi de suite jusqu'à la dernière composante qui explique la plus faible variance. D'un point de vue géométrique, on peut considérer les composantes principales comme de nouveaux axes qui donnent le meilleur angle pour visualiser et évaluer les données de façon que la différence entre les observations soit claire et visible.

En se basant sur des critères empiriques, nous choisissons les composantes à retenir qui préserve une variance significative dans le but de réduire la dimensionnalité de l'ensemble des données. Dans une dernière étape, on projette les individus et les variables sur les plans factoriels choisis afin d'interpréter la signification des axes et étudier les regroupements d'observations dans cet espace de dimensions réduites. Il est possible également de projeter des variables et des individus externes sans les introduire dans la formation des facteurs principaux.

Références

- Abdi, H., Williams, L. J., & Valentin, D. (2013, February 14). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *WIREs Computational Statistics*, 5(2), 149-179. doi:10.1002/wics.1246
- Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1(2), 245-276. doi:10.1207/s15327906mbr0102_10
- Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147-168. doi:10.1177/1094428103251541
- Duby, C., & Robin, S. (2006). *Analyse en Composantes Principales*. Institut National Agronomique Paris – Grignon.

Annexe C : Nuage des points et corrélation entre les indices climatiques et les composantes des précipitations

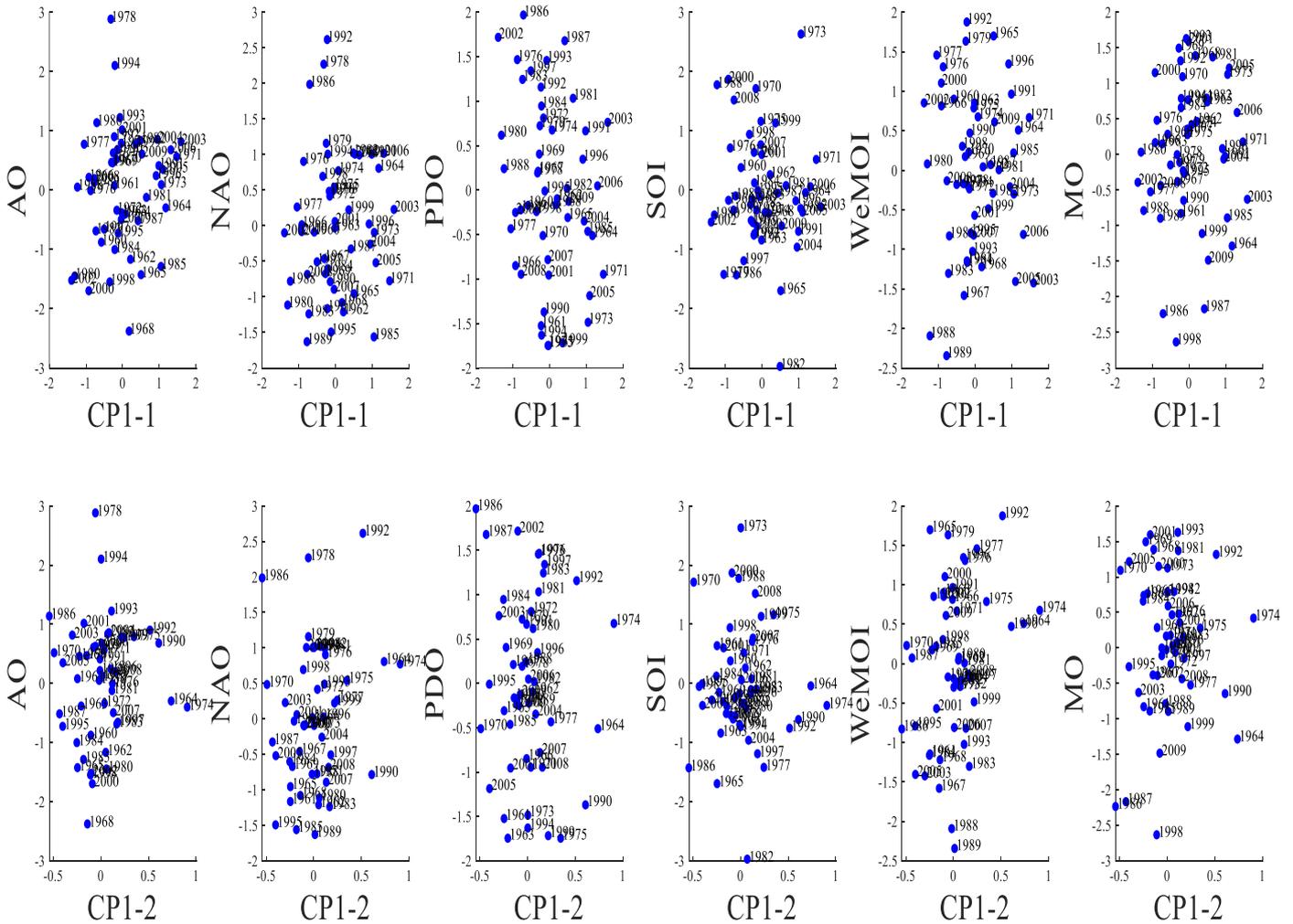


Figure 1 : Nuage des points des composantes CPI-1 et CPI-2 (qui représentent la variabilité spatiale de la quantité de pluie) et des indices climatiques

ANNEXES

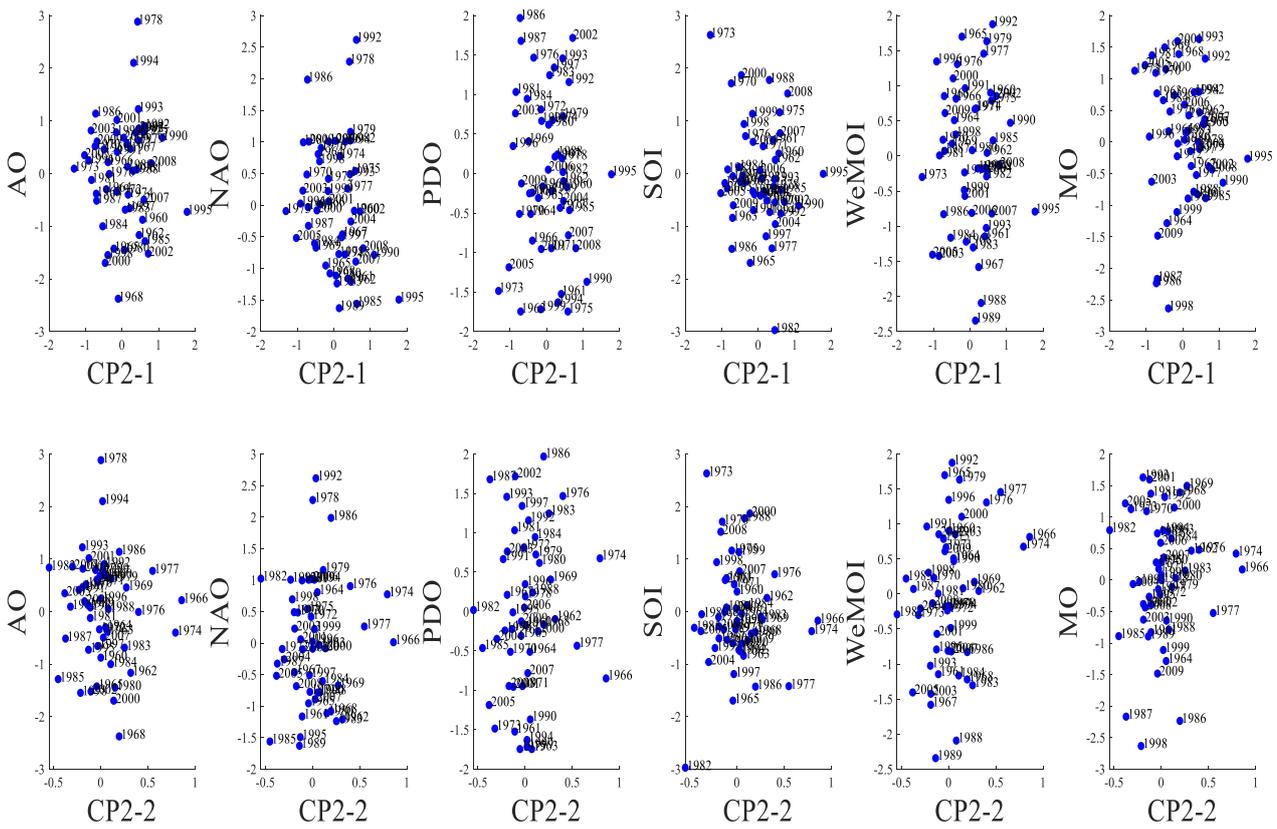


Figure 2 : Nuage des points des composantes CP2-1 et CP2-2 (qui représentent la variabilité spatiale de l'intermittence de pluie) et des indices climatiques

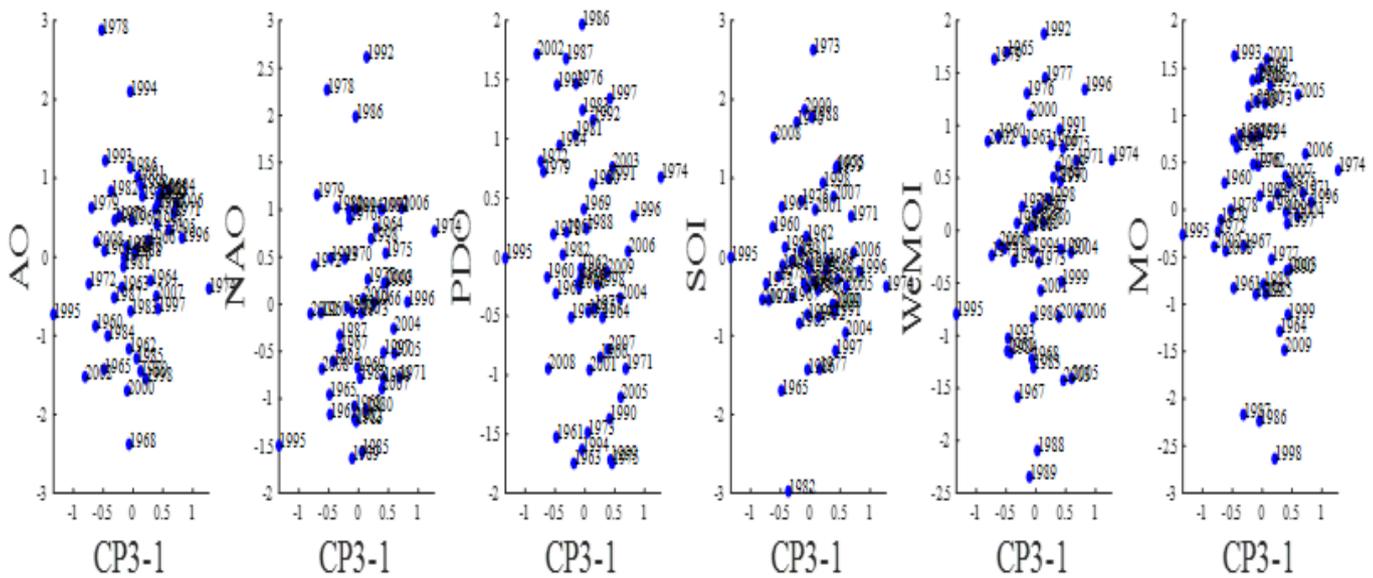


Figure 3 : Nuage des points des composantes CP3-1 (qui représente la variabilité spatiale de la composante des conditions météorologiques) et des indices climatiques

ANNEXES

Annexe D : Test de Student de différence entre les descripteurs des classes (p-value)

Ne				
	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	1	2.25E-165	1.19E-76	0.00012857
Classe 2		1	8.97E-10	2.94E-18
Classe 3			1	4.65E-06
Classe 4				1
Nj				
	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	1	3.63E-130	0	8.32E-204
Classe 2		1	3.33E-118	9.84E-92
Classe 3			1	1.70E-17
Classe 4				1
Ct				
	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	1	0.7871956	1.29E-266	0.00E+00
Classe 2		1	3.24E-245	1.06E-294
Classe 3			1	1.03E-86
Classe 4				1
Cm,e				
	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	1	1.64E-19	9.38E-201	5.93E-302
Classe 2		1	2.11E-234	1.09E-281
Classe 3			1	1.40E-91
Classe 4				1
Dm,e				
	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	1	6.12E-59	2.46E-21	1.15E-66
Classe 2		1	6.64E-136	2.33E-174
Classe 3			1	3.05E-42
Classe 4				1
Cm,j				
	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	1	1.35E-09	2.24E-183	1.36E-162
Classe 2		1	1.31E-116	2.40E-131
Classe 3			1	8.15E-21
Classe 4				1+J1:N36