

UNIVERSITÉ DU QUÉBEC  
INSTITUT NATIONAL DE LA RECHERCHE SCIENTIFIQUE  
CENTRE – EAU TERRE ET ENVIRONNEMENT

**PRÉVISION DES PRÉCIPITATIONS À LONG TERME AU NORD-EST DE  
LA TUNISIE**

Par

FARAH BEN SOUISSI

MÉMOIRE PRÉSENTÉ POUR L'OBTENTION DU GRADE DE MAÎTRISE  
SCIENCES (M.SC.) EN SCIENCES DE TERRE

JURY D'ÉVALUATION

Annie POULIN, examinatrice externe, École de technologie supérieure

Saeid HOMAYOUNI, président, INRS-ETE

Pierre MASSELOT, codirecteur de recherche, London School of Hygiene & Tropical Medicine

Taha B.M.J. OUARDA, directeur de recherche, INRS-ETE

## RÉSUMÉ

L'analyse de la variabilité des précipitations fait l'objet de recherches depuis plusieurs décennies. La Tunisie est caractérisée par une forte variabilité des précipitations qui entraîne des inondations et des sécheresses extrêmes, coûteuses en vies humaines et responsables d'innombrables dégâts matériels. Il est donc important de développer des modèles de prévisions à long terme des précipitations qui peuvent être utilisées efficacement dans les activités de la planification et de la gestion des ressources en eau. L'objectif du présent travail est de développer des modèles de prévision à long terme pour les précipitations annuelles totales et maximales en utilisant des indices d'oscillation climatique à basse fréquence. Ainsi, deux modèles de prévision sont considérés : un modèle de régression linéaire et un modèle fonctionnel linéaire pour la réponse scalaire (FLM-S). Dans cette étude, les données de précipitations journalières collectées à la station pluviométrique "Kélibia" située au nord-est de la Tunisie sont utilisées. L'impact du potentiel des modèles climatiques à grande échelle de l'oscillation de l'Atlantique du Nord (NAO) et l'oscillation décennale du Pacifique (PDO) est évalué dans la modélisation de la variabilité des précipitations. Étant donné que la station est située dans la région méditerranéenne, l'oscillation méditerranéenne (MO) et l'oscillation de la méditerranée occidentale (WeMO) sont également étudiées. Les résultats montrent que les indices NAO, PDO et MO représentent les meilleures variables pour la prévision des précipitations totales annuelles. Les indices MO et WeMO sont les meilleures variables pour la prévision des précipitations maximales annuelles. Les résultats de cette étude peuvent avoir des applications pratiques pour améliorer l'anticipation des catastrophes et la gestion des ressources en eau.

*Mots-clés:* Téléconnections ; indice climatique; Tunisie; variabilité des précipitations; analyse fonctionnelle des données (FDA) ; modèle linéaire fonctionnel pour une réponse scalaire (FLM-S) ; prévision à long terme

## **AVANT-PROPOS**

Ce document présente un mémoire de maîtrise par article. Ce mémoire est composé de deux chapitres. Le premier chapitre présente la synthèse. Cette partie englobe un résumé de la problématique, la méthodologie et les résultats obtenus. Le deuxième chapitre est l'article produit.

Le titre et les auteurs de l'article sont :

“Long term precipitation forecasting for North-eastern Tunisia”

**Auteurs** : Farah Ben Souissi, Pierre Masselot, Taha B.M.J. Ouarda

## **REMERCIEMENTS**

Je tiens à exprimer mes remerciements et ma gratitude à mon directeur de recherche Taha B.M.J. OUARDA pour son encadrement et pour ses précieux conseils durant cette maîtrise en mettant à ma disposition son temps, son expérience et ses compétences.

Je remercie également mon codirecteur Pierre Masselot pour tous ses efforts et sa contribution à l'avancement de mon travail avec ses conseils dynamique et toute l'aide qu'il m'a apporté.

Enfin, je tiens à remercier ma famille et mes amis pour leur support et leur encouragement.

# TABLE DES MATIÈRES

<b>Résumé .....</b>	<b>2</b>
<b>Avant-propos .....</b>	<b>4</b>
<b>Remerciements .....</b>	<b>5</b>
<b>Liste des tableaux .....</b>	<b>8</b>
<b>Liste des figures .....</b>	<b>9</b>
<b>Liste des abréviations.....</b>	<b>10</b>
<b>Partie 1: synthese.....</b>	<b>12</b>
1. Introduction .....	13
1.1 <i>Mise en contexte</i> .....	13
1.2 <i>Problématique</i> .....	13
1.3 <i>Objectif</i> .....	14
1.4 <i>Organisation de la synthèse</i> .....	15
2. Zone d'étude.....	16
3. Contexte théorique de l'analyse des données fonctionnelles .....	16
3.1 <i>L'analyse des données fonctionnelles</i> .....	16
3.2 <i>Le modèle linéaire fonctionnel pour réponse scalaire</i> .....	17
4. Résultats .....	18
5. Conclusions et perspectives .....	21
<b>Partie 2: ARTICLE.....</b>	<b>23</b>
1. Abstract .....	25
2. Introduction and review .....	26
3. Theoretical Background of FDA.....	29
3.1 <i>Data smoothing</i> .....	29
3.2 <i>Functional linear model for a scalar response</i> .....	30

4. Case study .....	31
4.1 Dataset description .....	31
4.2 Modelling strategy .....	33
5. Results and discussion.....	35
6. Summary, conclusions and future work.....	39
7. Bibliography.....	41

## LISTE DES TABLEAUX

Table 1 : Summary of some statistical characteristics of the total and maximum annual precipitation time series.....	47
Table 2 : Sources of climate oscillation data.....	48
Table 3 : Degree of smoothing (df, $\lambda$ ) related to smoothing indices.....	49
Table 4 : Beta smoothing (df, lambda) and performance criteria (GCV, OCV, and RMSE) related to FLM-S for the 1-month scale.....	49
Table 5 : Beta smoothing (df, lambda) and performance criteria (GCV, OCV, and RMSE) related to FLM-S for the 3-month scale.....	50
Table 6 : Significant correlations between climate indices and both total and maximum annual precipitations.....	51
Tableau 7 : Comparison between FLM-S and linear regression models to forecast total and maximum annual precipitation.....	52



## LISTE DES FIGURES

Figure 1 : Emplacement géographique de la zone d'étude « <i>Kélibia</i> ».....	53
Figure 2 : Steps to fit FLM-S .....	54
Figure 3 : estimated $\beta(t)$ values for predicting the logarithm of total annual precipitation from the NAO climate index for both scales: (a) 1-month scale and (c) 3-month scale, as well as the curves of linear correlations between total annual precipitation and the NAO index for both scales: (b) 1-month scale and (d) 3-month scale.....	55
Figure 4 : estimated $\beta(t)$ values for predicting the logarithm of total annual precipitation from the PDO climate index for both scales: (a) 1-month scale and (c) 3-month scale, as well as the curves of linear correlations between total annual precipitation and the PDO index for both scales: (b) 1-month scale and (d) 3-month scale.....	56
Figure 5 : estimated $\beta(t)$ values for predicting the logarithm of total annual precipitation from the MO climate index for both scales: (a) 1-month scale and (c) 3-month scale, as well as the curves of linear correlations between total annual precipitation and the MO index for both scales: (b) 1-month scale and (d) 3-month scale.....	57
Figure 6: estimated $\beta(t)$ values for predicting the logarithm of maximum annual precipitation from the WeMO climate index for both scales: (a) 1-month scale and (c) 3-month scale, as well as the curves of linear correlations between maximum annual precipitation and the WeMO index for both scales: (b) 1-month scale and (d) 3-month scale. ....	58

## LISTE DES ABRÉVIATIONS

AMJ	April – May – June
CPC	Climate Prediction Center
DJF	December – January - February
ESRL	Earth Systems Research Lab
FDA	Functional data analysis, analyse de données fonctionnelles
FLM-S	Functional linear model for scalar response, modèle fonctionnel pour réponse scalaire
FMA	February – March – April
JFM	January – February – March
GCV	Generalized cross-validation
JJA	June – July – August
LOOCV	Leave one out cross validation, validation croisée
MO	Mediterranean Oscillation, oscillation méditerranéenne
MAM	March – April - May
MJJ	May – June - July
NAO	North Atlantic Oscillation, Oscillation de l’atlantique du nord
NOAA	National Oceanic and Atmospheric Administration
NDJ	November - December – January

OND	October – November - December
OCV	Ordinary cross-validation
PDO	Pacific Decadal Oscillation
RMSE	Root mean square error, erreur quadratique moyenne
SON	September – October - November
WeMO	Western Mediterranean Oscillation,
%RMSE	relative root mean square error, erreur quadratique moyenne relative

## **PARTIE 1: SYNTHÈSE**

## **1. Introduction**

### **1.1 Mise en contexte**

Les précipitations sont très importantes pour l'écosystème et les humains puisqu'il s'agit de la principale source d'eau, mais aussi un phénomène météorologique complexe.

Les précipitations en Tunisie suivent un régime Méditerranéen, qui est caractérisé par une importante variabilité (Derouiche et al. 2019). Cette variabilité est principalement due à sa position (entre 30°N et 45°N) qui est directement influencée par les hautes pressions subtropicales et les basses pressions des latitudes moyennes (Rysman et al. 2013). La Tunisie, est plus particulièrement localisée entre la longitude 6-12 E et la latitude 30-38 N. Il s'agit donc d'une zone de transition climatique entre le domaine tempéré européen, le nord de la Méditerranée et le domaine subtropical d'Afrique australe (Derouiche et al. 2019).

La variabilité des précipitations peut causer des catastrophes naturelles extrêmement coûteuses en vies humaines et responsables d'innombrables dégâts matériels. Il est donc nécessaire d'analyser et de prévoir les précipitations pour éviter leurs impacts négatifs sur plusieurs secteurs.

### **1.2 Problématique**

La variabilité importante des précipitations en Tunisie en fait un phénomène difficile à prévoir. Une partie du travail de prévision consiste à rechercher les facteurs explicatifs de cette variabilité. Parmi ceux-ci, les indices climatiques peuvent être un outil intéressant pour comprendre la

variabilité des précipitations. Il a été démontré qu'ils influencent directement la variabilité des précipitations dans le monde (Ouarda et al. 2014; Thiombiano et al. 2018). La région méditerranéenne, due à sa position particulière mentionnée ci-dessus, ne fait pas exception à cette règle. En effet, de nombreuses études ont montré que la variabilité des précipitations en méditerranée est associée à la circulation atmosphérique générale, comme l'oscillation de l'Atlantique Nord (NAO), l'oscillation méditerranéenne (MO) (Conte. 1989) et l'oscillation de la Méditerranée occidentale (WeMO) (Martin-Vide et al. 2006). En Tunisie, Ouachani et al (2011) étudient la variabilité temporelle des précipitations et leurs liens avec les indices climatiques pour le bassin transfrontalier supérieur de Medjerda situé dans le nord de la Tunisie. Les résultats globaux encouragent l'utilisation des téléconnexions pour étudier la variabilité des précipitations en Tunisie.

Les téléconnexions sont définies comme un schéma à grande échelle d'anomalies de pression et de climat qui s'étend sur de vastes zones. Ce phénomène peut être mesuré à l'aide des indices climatiques qui représentent une différence de pression calculée entre deux points caractérisés respectivement, par une pression élevée et une pression basse. Ces indices climatiques peuvent être corrélés significativement avec des variables météorologiques. En effet, pour pouvoir étudier le régime de variabilité des précipitations, il est intéressant d'établir des liens de corrélation entre les indices climatiques et les données pluviométriques.

### **1.3 Objectif**

L'objectif du présent travail se place dans un cadre de prévision. Il consiste à utiliser les indices climatiques pour la prévision à long terme des précipitations.

L'impact des indices climatiques sur les précipitations pouvant être étalé sur plusieurs mois avec un potentiel délai, nous considérons ici le cadre de l'analyse des données fonctionnelles. Ce cadre, qui exprime les données comme des courbes continues au lieu de valeurs scalaires, permet ainsi d'intégrer plus facilement l'ensemble de l'information climatique pour la prévision des précipitations. Dans ce cadre, nous considérons un modèle de régression fonctionnelle pour réponse scalaire.

En plus de ce modèle, nous considérons également un modèle de régression linéaire plus classique pour étudier l'association entre les précipitations et les indices climatiques. La prise en compte des deux modèles permet de renforcer les conclusions tirées de l'étude et de réduire le risque de tomber dans les pièges de l'un ou l'autre des modèles. De plus, cela montre l'intérêt de considérer les modèles de données fonctionnelles dans la recherche hydro-climatologique.

L'objectif primordial de cette recherche est la prévision à long terme des précipitations totales et maximales annuelles en fonction des indices climatiques.

#### **1.4 Organisation de la synthèse**

La présente synthèse est divisée en quatre sections. La section 2 présente la zone d'étude. Par la suite, la section 3 présente le contexte théorique de l'analyse des données fonctionnelles. La section 4 présente un résumé des résultats de l'application à la prévision des précipitations au nord-est de

la Tunisie. En dernier lieu, la section 5 présente les conclusions et les perspectives liées à cette étude.

## **2. Zone d'étude**

Les données de précipitations journalières considérées dans cette étude ont été collectées par l'Institut National de Météorologie de Tunisie à partir de la station "Kélibia" numéro 1282721. "Kélibia" est une ville côtière située au nord-est de la Tunisie. La figure 1 montre une carte de la Tunisie avec l'emplacement de la location de la station.

La station étudiée couvre 31 années de données de précipitations journalières, de 1970-1971 à 2000-2001. Le total et le maximum annuels sont extraits de cette série temporelle.

Dans cette étude, nous évaluons l'impact du potentiel des modèles climatiques à grande échelle de l'oscillation de l'Atlantique du Nord (NAO) et l'oscillation décennale du Pacifique (PDO) pour modéliser la variabilité des précipitations. Par ailleurs, l'oscillation méditerranéenne (MO) et l'oscillation de la Méditerranée occidentale (WeMO) sont également étudiées. Le choix et la définition de ces indices climatiques sont présentés dans la partie article, la section 4.1.

## **3. Contexte théorique de l'analyse des données fonctionnelles**

### **3.1 L'analyse des données fonctionnelles**

L'analyse des données fonctionnelles, notée FDA, a été introduite par Ramsay (1982) puis popularisé par Ramsay et al (2005) ainsi que par Ramsay et al (2009). Cette approche vise à traiter



toute une série de données comme une fonction au lieu d'une série discontinue d'observations. L'idée est donc de convertir les données brutes sous leur forme discrétisée en des données sous forme continue. Il est donc question d'une approche plus représentative du phénomène réel et qui permet une meilleure utilisation de l'ensemble des informations disponibles.

La plupart des techniques statistiques conventionnelles ont été étendues au contexte fonctionnel, comme les modèles linéaires, la détection des valeurs aberrantes et l'analyse en composantes principales (Ramsay et Dalzell. 1991). Cette approche est devenue un outil pertinent utilisé dans de nombreux domaines scientifiques tels que l'écologie (Chiou. 2012), la médecine (Ciarleglio et al. 2016). Récemment, Chebana et al (2012) ont proposé d'utiliser le cadre statistique de la FDA dans le contexte hydrologique.

L'obtention des données fonctionnelles se fait par lissage à partir de la série de données observées. Il s'agit de créer un objet fonctionnel qui s'écrit comme combinaison linéaire de fonction de bases. Les équations mathématiques sont présentées dans la deuxième partie du rapport qui est l'article.

### **3.2 Le modèle linéaire fonctionnel pour réponse scalaire**

Il existe trois méthodes de régressions fonctionnelles :

- ❖ La régression fonctionnelle pour réponse scalaire (scalaire sur fonction) notée FLM-S
- ❖ La régression des réponses fonctionnelles sur un ensemble de prédicteurs scalaires (fonction sur scalaire)

#### ❖ La régression fonctionnelle pour réponse fonctionnelle (fonction sur fonction)

En cas de notre recherche nous nous sommes intéressés au premier modèle FLM-S (appelé en anglais, functional linear model for scalar response) qui considère les données fonctionnelles comme des entrées et des sorties scalaires. L'idée est de relier des courbes climatiques entières à un indicateur annuel de précipitations, à savoir le total et le maximum.

Un tel modèle constitue un moyen pour remédier aux lacunes suivantes. Il permet i) de filtrer le bruit dans la mesure des indices climatiques par lissage et ii) d'intégrer des informations complètes sur les variations des indices au cours de l'année. Il évite ainsi de devoir choisir un sous-ensemble de mois *a priori*.

Grâce à ces avantages, le FLM-S a été utilisé dans diverses études telles que la médecine (Ractliffe et al, 2002a ; 2002b), l'écologie (Bel et al, 2011), l'économétrie (Sood et al, 2009), la spectrométrie (Ferraty et al, 2002), la prévision du débit des cours d'eau (Masselot et al, 2016) et la mortalité liée à la température (Masselot et al, 2018).

#### **4. Résultats**

Les résultats de cette étude sont décrits en détail dans l'article. Cette partie présente un résumé des résultats obtenus de l'application à la prévision des précipitations au nord-est de la Tunisie.

En premier lieu, nous présentons les résultats du modèle FLM-S. Par la suite, nous présentons les résultats du modèle de la régression linéaire. Finalement, nous présentons une comparaison entre ces deux modèles.

Ces deux modèles sont utilisés dans le but de la prévision à long terme des précipitations totales et maximales annuelles au nord-est de la Tunisie. Pour l'évaluation des modèles, Les critères de performances sont calculés. Ces critères sont calculés par validation croisée (RMSE,%RMSE, OCV et GCV) et représentent l'erreur commise par le modèle lorsqu'il s'agit de la prévision des nouvelles valeurs, non utilisées dans l'ajustement. Le F-ratio est également calculé.

Au niveau du FLM-S, il existe deux types de variables d'entrée. Le premier type considère les indices climatiques mensuels (échelle 1 mois) et le deuxième type considère les indices climatiques moyennés sur 3 mois (échelle 3 mois).

Les résultats montrent que, pour l'échelle 1 mois, les modèles ayant comme variables explicatives les indices NAO et MO présentent les erreurs de prévision les plus basses pour la prévision du logarithme des précipitations totales annuelles. Le logarithme a été utilisé pour normaliser les données. Le modèle ayant comme variable explicative l'indice WeMO pour la prévision du logarithme des précipitations maximales annuelles a les erreurs de prévision les moins élevés.

Pour l'échelle 3 mois, les modèles ayant comme variables explicatives les indices NAO et PDO présentent les erreurs les plus basses pour la prévision du logarithme des précipitations totales annuelles. Pour les précipitations maximales, les modèles ayant comme variables explicatives les deux indices méditerranéens MO et WeMO semblent être les meilleurs.

En ce qui concerne les résultats de la régression linéaire simple, nous avons calculé les corrélations linéaires entre les indices climatiques et les précipitations totales et maximales annuelles. Les corrélations statistiquement significatives sont cohérentes avec les résultats trouvés pour le modèle

FLM-S. En effet, pour l'échelle 1 mois, les indices qui corrèlent avec les précipitations totales annuelles sont les indices NAO et MO pour différents mois. L'indice qui corrèle avec les précipitations maximales annuelles est l'indice WeMO.

Pour l'échelle 3 mois, les indices qui corrèlent avec les précipitations totales annuelles sont le NAO et PDO sur différentes périodes. Les indices qui corrèlent avec les précipitations maximales annuelles sont les deux indices méditerranéens MO et WeMO.

Les résultats des coefficients fonctionnels affirment aussi la cohérence entre les résultats obtenus dans le cadre fonctionnel et les résultats trouvés dans le cadre linéaire. En effet, l'interprétation des coefficients fonctionnels, détaillée dans l'article, montre une analogie entre ceux-ci et l'amplitude des corrélations linéaires obtenues.

Par ailleurs, les valeurs du RMSE et du %RMSE calculés pour les modèles FLM-S et les modèles de la régression afin de les comparer, sont très proches. Ceci montre que les deux modèles ont presque le même apport au niveau de la prévision des précipitations totales et maximales annuelles.

C'est vrai que selon des études précédentes, l'approche fonctionnelle permet l'amélioration des prévisions, mais ça n'a pas été le cas pour cette étude. Ceci pourrait être dû à la quantité des données, la série de données étant assez courte pour un modèle aussi complexe que la régression fonctionnelle. Cependant, ce modèle reste pertinent, au vu de sa propension à utiliser toute l'information disponible. De plus, les coefficients fonctionnels présentent aussi un autre aspect intéressant du fonctionnel, permettant une représentation plus compacte de l'information donnée par les corrélations linéaires.

## 5. Conclusions et perspectives

L'objectif de cette recherche était d'étudier l'association entre les précipitations et les indices climatiques au nord-est de la Tunisie afin de fournir des lignes directrices pour une meilleure gestion des ressources en eau et des ressources agricoles. Pour ce faire deux modèles de prévision ont été utilisés : le FLM-S et la régression linéaire. Les résultats de la présente étude fournissent un ensemble de prédicteurs climatiques qui peuvent être utiles pour la prévision à long terme des précipitations. Pour le cas des précipitations totales annuelles, il s'est avéré que les 3 indices NAO, PDO et MO peuvent être des bons prédicteurs et pour le cas des précipitations maximales, les indices Méditerranéens MO et WeMO sont des bons prédicteurs.

En comparant le FLM-S et la régression linéaire, on peut conclure que la régression fonctionnelle donne une performance de prévision légèrement inférieure. En effet, le modèle de régression linéaire est plus performant que le FLM-S. Néanmoins, les deux modèles proposés dans le présent travail ont de bonnes capacités de prévision à long terme et peuvent être utilisés dans la pratique.

Dans la perspective de cette recherche, l'approche fonctionnelle peut être appliquée sur des données de meilleure quantité avec des séries de précipitations plus longues. En effet, la suite du présent travail consiste à prévoir l'ensemble de la série des précipitations au lieu d'une valeur scalaire en utilisant le modèle fonction sur fonction. Une autre perspective serait une étude plus large dans plusieurs bassins d'Afrique du Nord pour des conclusions plus solides et pour tester le potentiel de l'approche proposée dans d'autres études de cas. Les recherches futures pourraient également se

concentrer sur l'intégration d'autres variables dans les modèles de prévision et sur l'examen d'autres approches de prévision telles que le modèle additif généralisé.

**PARTIE 2: ARTICLE**

# Long-term precipitation forecasting for North-eastern Tunisia

Farah Ben Souissi 1, Pierre Masselot 2, Taha B.M.J. Ouarda 1, Emna Ellouze-Gargouri<sup>3</sup>

<sup>1</sup> Canada Research Chair in Statistical Hydro-Climatology, Institut national de la recherche scientifique, Centre Eau Terre Environnement. INRS-ÉTÉ, 490 De la Couronne St, Québec City, QC, Canada

<sup>2</sup> London School of Hygiene & Tropical Medicine, Keppel St, Bloomsbury, London WC1E 7HT, United Kingdom

<sup>3</sup> Laboratoire de Modélisation en Hydraulique et Environnement, École Nationale d'Ingénieurs de Tunis, BP 37, Le Belvédère 1002 Tunis, Tunisie

**Journal:** Ready to submit in Environmental Modeling and Software Journal



## 1. Abstract

The analysis of precipitation variability has been the subject of research for several decades. Tunisia is characterised by a high precipitation variability which leads to extreme floods and droughts that are costly in human lives and responsible for countless material damages. Therefore, it is important to develop efficient precipitation forecasts that can be used effectively in water resources planning and management activities. The present work's objective is to develop long-term forecast models for the total and maximum annual precipitation using low-frequency climate oscillation indices. To this end, two forecasting models are considered: a linear regression model and a linear functional model for scalar response (FLM-S). In this study, daily precipitation data collected from the "Kélibia" rainfall station located in North-eastern Tunisia are used. The association between precipitation and the North Atlantic Oscillation (NAO) and Pacific Decadal Oscillation (PDO) climate indices is evaluated. Since the study region is located in the Mediterranean area, the Mediterranean Oscillation (MO) and the Western Mediterranean Oscillation (WeMO) indices are also considered. Results show that linear and functional regression leads to similar results. Based on cross-validated performance criteria, NAO, PDO, and MO are the best indices to forecast total annual precipitation. On the other hand, MO and WeMO are the best indices to forecast maximum annual precipitation. The results of this study may have practical applications to improve disaster response preparedness and water resource management.

*Keywords:* Teleconnections, climate oscillation index, Tunisia, precipitation variability, functional Data Analysis (FDA), Functional Linear Model for a scalar response (FLM-S), long-term forecasting

## 2. Introduction and review

In the Mediterranean basin, precipitation shows substantial spatial and temporal variability. Several studies attempted to analyse this variability in a many locations in the Mediterranean region, for instance in Algeria (Taibi et al, 2013), Morocco (Singla 2009; Benassi, 2001), Italy (Longobardi and Villani, 2010), Greece (Xoplaki et al. 2000), and in Spain (Sinoga et al. 2011). This variability depends on the location (between 30°N and 45°N) and is clearly influenced by high subtropical pressures and mid-latitude low pressures (Rysman et al., 2013).

In Tunisia, precipitation variability results in strong flood and drought episodes costing human lives and property damage. Therefore, it is important to analyze and forecast precipitation to avoid their adverse social and economic consequences and the severe impact on the water sector, agriculture, and all other water-dependent sectors (Derouiche et al. 2019). A part of the work in forecasting precipitation is to search for explanatory factors explaining this variability.

Climate indices are related to precipitation variability. They have been shown to influence precipitation variability throughout the world directly (Ouarda et al., 2014; Thiombiano et al., 2018, for instance). Several studies have shown that, in the Mediterranean region, precipitation variability is associated with general atmospheric circulation such as the North Atlantic Oscillation (NAO) (Trigo et al., 2004), the Mediterranean Oscillation (MO) (Conte, 1989), and the Western Mediterranean Oscillation (WeMO) (Martin-Vide et al., 2006).

A major difficulty in modeling the association between climate indices and precipitation is that this association can be spread out over several months of the year. It is challenging to integrate consecutive months into a classical linear model because of the important inter-correlation between

them. The main consequence is a decrease in the accuracy of estimates, and ultimately the confidence one can attribute to the resulting conclusion. A typical approach is selecting a subset of months to include in the model, but the risk is then an important loss of information and less forecasting power.

In this context, the functional framework seems to be a promising alternative to classical linear models. Functional Data Analysis (FDA) allows a whole series of data to be treated as a continuous function instead of a discontinuous observation series. It is thus more representative of the real phenomenon and allows better use of the full available information. This approach was introduced by Ramsay (1982) and later widely popularized by Ramsay et al. (2005) as well as Ramsay et al. (2009). FDA has been used in many scientific fields including Paleopathology, Meteorology, Criminology, Economics (Ramsay et al. 2007), ecology (Chiou 2012), medicine (Ratcliffe et al. 2002; Ciarleglio et al. 2016), waste management (Bernardi et al. 2017), biotechnology (Brockhaus et al. 2017), neuroscience (Meyer et al. 2015), and energy (Brockhaus et al. 2015; Chaouch. 2013) .

Recently, (Chebana et al. 2012) proposed using the FDA's statistical framework in the hydrological context. FDA has been applied in several studies for different purposes. Ternynck et al. (2016) used FDA to streamflow hydrograph classification. Brunner et al. 2017 applied FDA to Flood Reactivity Regions identification. Requena et al. (2018) used a functional framework for flow-duration-curve and daily streamflow estimation. Boudreault et al. (2019) applied this approach to Stream Temperature Modeling. FDA was also used to forecast different variables such as air temperature (Curceac et al., 2019). Moreover, Masselot et al. (2016) used functional regression to forecast streamflow. In fact, there are three types of functional regression models: functional

regression model for a scalar response (FLM-S), regression of functional responses on a set of scalar predictors (function on scalar), and functional regression model for functional response (function on function). Readers interested in a detailed review of functional regression models are referred to Morris (2014) and Greven et al. (2017).

The present paper's model of interest is the functional linear model for a scalar response (FLM-S) that considers functional data as inputs and a scalar output. The idea is to relate whole climatic curves to an annual precipitation indicator, namely the total and the maximum. Such a model provides an elegant way to address the shortcomings mentioned above. It allows i) filtering out the noise in climatic indices measurements through smoothing and ii) integrating complete information of the indices variations across the year. It avoids hence the burden of having to choose a subset of months. Using these advantages, the FLM-S has been used in various studies such as medicine (Ractliffe et al. 2002b), ecology (Bel et al. 2011), econometrics (Sood et al. 2009) spectrometry (Ferraty et al. 2002), streamflow forecasting (Masselot et al. 2016) and temperature-related mortality (Masselot et al. 2018).

In this paper, we jointly consider the FLM-S and a more classical linear regression model to study the association between the aforementioned climate indices and precipitation. The purpose is to evaluate whether the indices may offer new insights on precipitation variability and improve our forecasting abilities of the phenomenon. Considering the two models allows strengthening the study's conclusions and reducing the risks associated with either one of the two methods' weaknesses. The linear regression model is selected to explore the potential of using teleconnections in long-term precipitation forecasting through simple, commonly used approaches

in practice. Functional data models are selected to explore if more complex approaches that use data more optimally may significantly improve forecasting skills.

### 3. Theoretical Background of FDA

FDA has usually divided into two parts: i) data smoothing and ii) the application of the relevant method, i.e. FLM-S in the present case.

#### 3.1 Data smoothing

In the FDA, the first step is to convert discrete measurements to functions (curves). This is accomplished by creating a functional object  $x_i(t)$  expressed as a linear combination of  $K$  basis functions (Ramsay, 2006) as shown in equation (1).

$$x_i(t) = \sum_{k=1}^K \theta_k \varphi_k(t) \quad (1)$$

where  $\varphi_k(t)$  is a set of  $K$  basis functions defined. The most commonly used basis functions are a Fourier basis when data are periodic, b-spline basis when data are non-periodic and wavelets.  $\theta_k$  are the coefficients representing the function  $x_i(t)$ . Such an approach reduces the information in  $x_i(t)$  to a small set of coefficients.

The coefficients  $\theta_k$  are estimated by minimizing the penalized sum of square error expressed as (Masselet et al. 2016):

$$PSSE = \sum_{j=1}^T ((z_j - x_i(t_j))^2 + \lambda \int D^2 x_i(t) dt) \quad (2)$$

where  $z_j$  represents the observed discrete measurements of the climate index.  $D^2 x(t)$  is the second derivative of  $x(t)$ .  $\lambda \int D^2 x_i(t)$  is the term penalizing rough functions  $x_i(t)$  through their second derivative. The parameter  $\lambda$  regulates the severity of the penalization. Once created, the functional data are ready to be used in the functional linear model for a scalar response.

### 3.2 Functional linear model for a scalar response

The general formulation of the FLM-S is given by equation (3) (Ramsay, 2005):

$$y_i = \beta_0 + \int_T x_i(t) \beta(t) dt + \varepsilon_i \quad (3)$$

Where  $y_i$ , ( $i=1 \dots n$ ) is a continuous scalar response,  $x_i(t)$  is a functional predictor,  $\beta(t)$  is a functional coefficient,  $\beta_0$  is an intercept,  $\varepsilon_i$  is the residual error,  $n$  is the length of  $y_i$  (number of observations) and  $T$  refers to the limits of the interval for the months.

In this context, the regression coefficient  $\beta(t)$  is functional and gives the influence of  $x_i(t)$  on  $y_i$  for all times  $t \in T$ . Therefore, it is also expressed using a set of basis functions as in equation (1),

*i.e.*  $\beta(t) = \sum_{i=1}^K c_i \phi_i(t)$ . This reduces the fitting of the model (3) to estimate a finite number of scalar

coefficients  $c_i$ . These coefficients are estimated by minimizing the following penalized sum of squares errors (Ramsay, 2009):

$$PENSSE_{\lambda}(\beta) = \sum_{i=1}^n \left[ y_i - \beta_0 - \int_T \beta(t) x_i(t) dt \right]^2 + \lambda \int_T [L\beta(t)]^2 dt \quad (4)$$

Where L is a differential operator used to penalize very complex functions. L typically represents the second derivative of  $\beta$  (.). The coefficient  $\lambda$  is practically chosen by minimizing the ordinary cross-validation (OCV) criterion.

$$OCV(\lambda) = \sum \left( \frac{y_i - \hat{y}_i}{1 - S_{ii}} \right)^2 \quad (5)$$

Where  $S_{ii}$  is the  $i^{\text{th}}$  diagonal element of the hat matrix associated with the FLM-S.

The coefficient  $\lambda$  can be also chosen by minimizing the generalized cross-validation (GCV) criterion. (Ramsay et al. 2005)

$$GCV(\lambda) = \left( \frac{n}{n - df(\lambda)} \right) \left( \frac{SSE}{n - df(\lambda)} \right) \quad (6)$$

Where SSE is the error sums of squares and df is the equivalent degrees of freedom.

## 4. Case study

### 4.1 Dataset description

The National Institute of Meteorology-Tunisia collected this study's daily precipitation data from the "Kélibia" station number 1282721. "Kélibia" is a coastal city situated in the North-eastern part of Tunisia. The studied station is located at the tip of the "Cap Bon" cape and covers 31 years of data, from 1970-1971 to 2000-2001. Figure 1 shows a map of the study area. The annual totals and

maxima are extracted from this time series. Table 1 presents a summary of some statistical characteristics of the total and maximum annual precipitation time series.

We consider the climate indices identified in Ouachani et al. (2011) that concerned a similar study area (northern Tunisia) i.e. NAO, PDO, WeMO, and MO. The NAO is defined as the pressure difference between Ponta Delgada in the Azores and Stykkishólmur in Iceland. According to several studies, NAO influences the Mediterranean basin (Adjez. 2000; Dunkeloh et al. 2003). The correlation analysis reported by Saadaoui et al. (2007) found statistically significant positive correlations between North African precipitation and NAO. The PDO index represents the main component of the sea surface temperature anomalies in the North Pacific Ocean at 20°N poles and is partially correlated to the NAO (Lee et al. 2013).

Since the case study is located in the Mediterranean region (North Africa), MO and the WeMO are also naturally considered. The WeMO index represents the difference between the standardized atmospheric pressure values recorded at Padua (45.40°N, 11.48°E) in northern Italy and San Fernando (Cadiz) (36.28°N, 6.12°W) in southwestern Spain (Martin-Vide et al. 2006). Palutikof et al. (1996) and Conte et al. (1989) defined the MO as a normalized pressure difference between Algiers (36.4°N, 3.1°E) and Cairo (30.1°N, 31.4°E). The MO has positive and negative phases. In the positive phase of MO, cyclogenesis is exceptionally intense while in the negative phase it is exceptionally weak (Martic-Bursac et al. 2017).

Sources of teleconnection information are available on several websites of the National Oceanic and Atmospheric Administration (NOAA): Climate Prediction Center (CPC) and Earth Systems Research Lab (ESRL). The sources of climate indices data used in the current study are mentioned



in Table 2. Climate indices are available monthly. The period considered for each climate index is from 1969-1970 to 1999-2000.

## 4.2 Modelling strategy

Two models are used to study the association between precipitation and climate indices: the FLM-S and the linear regression.

The FLM-S is used to forecast the logarithm of total and maximum annual precipitation based on climate indices (NAO, PDO, WeMO, and MO). The logarithm is used to add symmetry to data. B-spline bases are chosen to smooth climate indices because data are not periodic. Figure 2 shows the steps carried out as part of the FLM-S. There are two types of input variables. The first type considers monthly climate indices (1-month scale) and the second type considers climate indices averaged over 3 months (3-month scale).

For each model, the climate index should be smoothed. For this, the method described in section 3.1 is applied. Using equations (1) and (2),  $j$  is the number of months ( $j \in T = [1, 12]$  (for the case of 1-month scale) or 10 (for the case of 3-month scale)),  $i$  is the number of years ( $i \in [1, 31]$ ),  $K$  is the number of b-splines and  $x_i(t_j)$  represents the functional climate index of the  $i$ th year.

Once the climate index is smoothed, the FLM-S is applied to forecast the total and maximum annual precipitation logarithm. The explanatory variables used here are the smoothed climate indices. The outlined method in section 3.2. is applied throughout the study. Using equations (3) and (4),  $y_i$  represents the response as the  $\log_{10}$  of the total or maximum annual precipitation vectors to add symmetry to data.  $x_i(t)$  is one of the smoothed indices among the four indices (NAO, PDO,

WeMO, and MO). The linear differential operator object L in this study is represented by the second derivative of  $\beta(t)$  since the data are not periodic and we deal with b-splines. After applying FLM-S, climate indices that result in the best forecasting of both total and maximum annual precipitation are retained.

The linear regression model is also used to forecast the logarithm of total and maximum annual precipitation. To do so, correlations between climate indices and total or maximum annual precipitation are calculated to identify a relationship between precipitation series and climate indices at the monthly scale (or average monthly scale). For this purpose, the Pearson Correlation Coefficient is used. Once identified, the months that correlate with precipitation are used as explanatory variables.

To evaluate both models' performances, leave-one-out cross-validation (LOOCV) is used. Given a sample of  $n$  observations,  $n-1$  are used to fit the model, and the remaining one is used for forecasting. This process is repeated for the  $n$  observations. Root mean square error (RMSE) and root mean square error %RMSE is then computed using the  $n$  prediction as defined below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{obs} - y_{pred})^2}{n}}$$

$$\%RMSE = \frac{RMSE}{\bar{y}_{obs}}$$

where  $y_{obs}$  and  $y_{pred}$  are, respectively, the observed and predicted values.  $\bar{y}_{obs}$  is the mean of observed values. In addition, the F-ratio is calculated to assess the significance of the models. The F-ratio is the test statistic we use to decide if the model has a statistically significant forecast.

## 5. Results and discussion

Table 3 presents the results corresponding to the preliminary step which is smoothing the climate indices. It presents both smoothing parameters (degree of freedom (df) and  $\lambda$ ) for each climate index and each scale (1-month and 3-month). The df corresponds to the number of B-spline basis functions. These parameters vary from one index to another and from one scale to another. That explains the existence of indices that are smoother than others. When lambda increases the number of B-spline bases (df) decreases.

Table 4 and 5 show the results related to FLM-S for both the 1-month and 3-month scales. On the one hand, it presents the smoothing parameters (df,  $\lambda$ ) of the beta coefficient that will be interpreted in what follows. On the other hand, it presents the performance criteria calculated by cross-validation (GCV, OCV, and RMSE).

Table 6 presents the statistically significant correlations calculated between climate indices' monthly values and total and maximum annual precipitation at different scales (1-month and 3-month).

According to table 4, models that forecast the logarithm of total annual precipitation using the NAO and MO have the best performance criteria. The model that forecasts the logarithm of maximum annual precipitation using the WeMO climate index has the best performance criteria.

These results are consistent with the linear correlations calculated between climate indices and precipitation related to the 1-month scale presented in Table 6. In fact, the NAO index shows a significant correlation with the total annual precipitation for three different months (January, March, and August). The MO index correlates with the total annual precipitation over May. Finally, the WeMO index correlates with the maximum annual precipitation for March.

According to table 5, models that forecast the logarithm of total annual precipitation using the NAO and PDO lead to the best performance criteria. Models that forecast the logarithm of maximum annual precipitation using the MO and WeMO have the best performance criteria.

This is fully consistent with the correlations found in Table 6. In fact, the NAO index correlates with total annual precipitation over two periods (DJF and JFM). The PDO index also appears with statistically significant correlations with total annual precipitation for different periods (SON, OND, NDJ, and DJF). The MO index correlates with maximum annual precipitation over one period (MAM). Also, the WeMO index correlates maximum annual precipitation over one period (OND). This coherence can also be shown when analysing the functional coefficients. In fact, the functional coefficients  $\beta(t)$ , are analyzed by comparing them to the linear correlations calculated in the scalar case.

Figure 3 illustrates the estimated values of  $\beta(t)$  for forecasting the logarithm of total annual precipitation from the NAO climate index for the 1-month and 3-month scales, as well as the curve of linear correlations between total annual precipitation and the NAO index for both scales (1-month and 3-month).

For the 1-month scale, the  $\beta(t)$  curve shows a positive effect of the NAO index on the total annual precipitation over January, March, and August. This is consistent with the significant positive correlations between total annual precipitation and the NAO index for the same months.

For the 3 months scale, the  $\beta(t)$  curve shows a positive effect of the NAO index on total annual precipitation for both DJF and JFM periods. This is also consistent with significant positive correlations recorded for the same periods.

Figure 4 illustrates the estimated  $\beta(t)$  values for forecasting the logarithm of total annual precipitation from the PDO index for 1-month and 3-month scales, as well as the curve of linear correlations between total annual precipitation and the PDO index for both scales.

For the 1-month scale, there are no significant correlations between the PDO index and total annual precipitation, but both curves have the same shape. For the 3-month scale, the  $\beta(t)$  curve shows an adverse effect of the PDO index on the total annual precipitation during SON, OND and NDJ periods. Similarly, significant negative correlations are recorded between the PDO index and total annual precipitation for the same periods.

Figure 5 illustrates the estimated  $\beta(t)$  values for forecasting the logarithm of total annual precipitation from the MO for the 1-month and 3-month scales, as well as the curve of linear correlations between total annual precipitation and the MO index for both scales.

For the 1-month scale, the  $\beta(t)$  curve is a linear line. However, the same signs are recorded between the functional coefficient and the linear correlations. Indeed, from September to January, the  $\beta(t)$  curve shows the positive effect of the MO index on total annual precipitation. From February to August, this curve shows the negative effect of MO on total annual precipitation. By analogy, the

linear correlations calculated between total annual precipitation and the MO index are generally positive from September to January and are generally negative from February to August.

For the 3-month scale, the  $\beta(t)$  curve shows a positive effect of the MO index on total annual precipitation during the MAM period, which corresponds to the statistically significant negative correlation for the MAM period.

Figure 6 illustrates the estimated  $\beta(t)$  values for forecasting the logarithm of total annual precipitation from the WeMO index for the 1-month and 3-month scales, as well as the curve of linear correlations between annual maximum precipitation and the WeMO index for the 1-month and 3-month scales.

For the 1-month scale, the  $\beta(t)$  curve shows the negative effect of the WeMO index on maximum annual precipitation during March. Similarly, a significant negative correlation is recorded between the WeMO index and the maximum annual precipitation during March.

For the 3-month scale, the beta curve shows the negative effect of the WeMO index on the maximum annual precipitation during the OND period. Similarly, a significant negative correlation is recorded between the WeMO index and the maximum annual precipitation for the same period (OND).

Finally, table 7 presents performance criteria values related respectively to the FLM-S and the linear regression model. Based on the corresponding F-ratio values, the models are significant. In addition, the RMSE and %RMSE values show that these models have good long-term forecasting skills.

The RMSE and %RMSE values calculated for the linear regression are slightly lower than the RMSE and %RMSE calculated for FLM-S. In light of these results, these criteria' values are very close, which means that the two models have almost the same contribution in terms of forecasting total and maximum annual precipitation.

According to previous studies, the functional approach allows the improvement of forecasts, but this was not the case. The linear regression outperforms the functional regression model may be attributable to the data quantity. The data series is relatively short for a model as complex as functional regression. However, this model remains relevant, given its ability to use all the available information. Moreover, the functional coefficients also present another interesting aspect of the functional, allowing a more compact representation of the information given by linear correlations.

## **6. Summary, conclusions and future work**

The present work aims to forecast precipitation based on climate indices using two types of forecasting models which are respectively the FLM-S and the linear regression model. The station considered is situated in North-eastern part of Tunisia and covers 31 years of daily precipitation data.

Both models are applied to forecast total and maximum annual precipitation. Results show that NAO, PDO and MO are the best indices for forecasting total annual precipitation. At the same time, the Mediterranean indices are the best indices for forecasting maximum annual precipitation. This indicates that the total amount of precipitation at the Kelibia station is controlled by global

phenomena occurring at the Atlantic and Pacific Ocean scale in addition to the Mediterranean scale. On the other hand, more regional climatic phenomena occurring at the Mediterranean scale control the station's few largest precipitation events.

By comparing FLM-S and linear regression, it can be concluded that functional regression yields a slightly lower forecasting performance. Indeed, the linear regression model outperforms the FLM-S. This may be due to the quantity of the available data since we dispose of a short dataset. Nevertheless, both models proposed in the present work have good long-term forecasting skills and can be used in practice. The use of low-frequency climate oscillation indices as covariates for the long-term forecasting of Tunisia precipitation is promising. This is so despite the complexity of building such models for Tunisia, where the climate is subject to many climatic oscillation signals occurring in the Atlantic, the Mediterranean and elsewhere. The development of such long-term forecasting models should be more comfortable and probably more promising in other regions of the world where a dominant climate oscillation signal prevails.

From this research, the functional approach can be applied to better quantity data with more extended precipitation series. In fact, the natural follow-up to the present work is to forecast the whole precipitation series instead of a scalar value using the fully function linear model (F-FLM). Another perspective would be a more extensive study in several basins of North-Africa for more robust conclusions and to test the potential of the proposed approach in other case studies. Future research may also focus on integrating other variables in the forecasting models and the consideration of other forecasting approaches such as the Generalised Additive Model (GAM).



## 7. Bibliography

Adjez A. (2000). Changes in the rainfall regime in the North-African region: links with large-scale oscillation phenomena. In: Conference Abstracts' Detection and modeling of recent climate change and its effects on a regional scale', Tarragona; 24

Benassi M., (2001). « Drought and Climate Change in Morocco. Analysis of Precipitation Field and Water Supply ». (80): 5.

Bel, L., Bar-Hen, A., Petit, R J., Cheddadi, R. 2011. «Spatio-temporal functional regression on paleoecological data. » Journal of Applied Statistics 38(4): 695-704.

Bernardi, M. S., Sangalli, L. M., Mazza, G., Ramsay, J. O. 2017. «A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province ». Stochastic Environmental Research and Risk Assessment, 31(1), 23-38.

Boudreault, J., Bergeron, N., St-Hilaire, A., Chebana, F., 2019. « Stream Temperature Modeling Using Functional Regression Models ». JAWRA Journal of the American Water Resources Association 55(6): 1382-1400.

Brockhaus, S., Scheipl, F., Hothorn, T., Greven, S. 2015. «The functional linear array model. Statistical Modelling». 15(3), 279-300.

Brockhaus, S., Melcher, M., Leisch, F., Greven, S. 2017. « Boosting flexible functional regression models with a high number of functional historical effects. Statistics and Computing ». 27(4), 913-926.

Brunner, M.I., Viviroli, D., Furrer, R., Seibert, J., Favre, A.C. .2018. « Identification of Flood Reactivity Regions via the Functional Clustering of Hydrographs ». *Water Resources Research* 54(3): 1852-67.

Chaouch, M. 2013. « Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves». *IEEE Transactions on Smart Grid* 5(1): 411-419.

Chebana, F., Dabo-Niang, S., Ouarda, T.B.M.J. 2012 « Exploratory Functional Flood Frequency Analysis and Outlier Detection ». *Water Resources Research* 48(4).

Chiou, J. 2012. « Dynamical Functional Forecasting and Classification, with Application to Traffic Flow Prediction ». *Annals of Applied Statistics* 6(4): 1588-1614.

Ciarleglio, A., Petkova, E., Tarpey, T., Ogden, R. Todd. 2016. « Flexible Functional Regression Methods for Estimating Individualized Treatment Regimes ». *Stat (International Statistical Institute)* 5(1): 185-99.

Conte, M., Giuffrida, A., Tedesco, S. 1989. « The mediterranean oscillation: impact on precipitation and hydrology in Italy ». In: *Proceedings of the Conference on Climate and Water, Vol. 1. Publications of Academy of Finland, Helsinki*, pp. 121–137.121-37.

Curceac, S., Ternynck, C., Ouarda, T.B.M.J., Chebana, F. 2019. « Short-Term Air Temperature Forecasting Using Nonparametric Functional Data Analysis and SARMA Models ». *Environmental Modelling & Software* 111: 394-408.

- Dünkeloh , A., Jacobeit, J. 2003. «Circulation dynamics of Mediterranean precipitation variability 1948–98». *International Journal of Climatology* 23: 1843–1866.
- Derouiche, S., Mallet, C., Bargaoui, Z. 2019. « Rainfall Event Analysis in the North of Tunisia Using the Self-Organizing Map ». : 6.
- Ferraty, F., Vieu, P. (2002). "The functional nonparametric model and application to spectrometric data." *Computational Statistics* 17(4): 545-564.
- Greven, S., Scheipl, F. 2017. «A general framework for functional regression modelling». *Statistical Modelling*, 17(1-2), 1-35.
- Longobardi, A., Villani, P. 2010. « Trend Analysis of Annual and Seasonal Rainfall Time Series in the Mediterranean Area ». *International Journal of Climatology* 30(10): 1538-46.
- Lee, T., Ouarda, TBMJ., Li, J. 2013. «An orchestrated climate song from the Pacific and Atlantic Oceans and its implication on climatological processes». *Int. J. Climatol.* 33(4): 1015–1020, doi: 10.1002/joc.3488
- Masselot, P., Dabo-Niang, S., Chebana, F., Ouarda, TBMJ. 2016. « Streamflow forecasting using functional regression ». *Journal of Hydrology* 538: 754-66.
- Masselot, P., Chebana, F., Ouarda, T.B.M.J., Bélanger, D., St-Hilaire, A., Gosselin, P. 2018. « A New Look at Weather-Related Health Impacts through Functional Regression ». *Scientific Reports* 8(1): 15241.
- Martin-Vide, J., Lopez-Bustins, J. 2006. « The Western Mediterranean Oscillation and Rainfall in the Iberian Peninsula ». *International Journal of Climatology* 26(11): 1455-75.

- Meyer, M. J., Coull, B. A., Versace, F., Cinciripini, P., Morris, J. S. 2015. « Bayesian function-on-function regression for multilevel functional data». *Biometrics*. 71(3), 563-574.
- Morris, J. S. 2014. «Functional regression». *Annual Review of Statistics and Its Application*, 2, 321-359.
- Ouachani, R., Bargaoui, Z., Ouarda, T. 2011: « Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda Basin», *Int. J. Climatol*, 19
- Ouarda, TBMJ., Charron, C., Niranjana Kumar, K., Marpu, P., Ghedira, H., Molini, A., Khayal, I. 2014. « Evolution of the Rainfall Regime in the United Arab Emirates ». *Journal of Hydrology*, 514: 258-70. 514 (June): 258–270. doi:10.1016/j.jhydrol.2014.04.032.
- Ratcliffe, S.J., Leader, L.R., Heller, G.Z. 2002. « Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression». *Statistics in Medicine* 21(8): 1103-1114.
- Ramsay, J. O. 1982. « When the Data Are Functions ». *Psychometrika* 47(4): 379-96.
- Ramsay, J.O., Dalzell, C.J. 1991. « Some tools for functional data analysis». *J. R. Stat. Soc. Ser. B (Methodol.)* 53, 539–572
- Ramsay, J. O., and Silverman, B. W. 2005, *Functional Data Analysis*, 2nd ed., Springer, New York.
- Ramsay, J. O., and Hooker, G. 2009, *Functional Data Analysis with R and MATLAB*, Springer.
- Ramsay, J.O. 2006. *Functional Data Analysis*. Wiley Online Library.

Ramsay, J. O., Silverman, B. W. 2007. « Applied functional data analysis: methods and case studies». Springer.

Requena, A.I., Chebana, F., Ouarda, T.B.M.J. 2018. « A functional framework for flow-duration-curve and daily streamflow estimation at ungauged sites » *Advances in Water Resources* 113: 328–40.

Rysman, J.-F., S. Verrier, Y. Lemaître, et E. Moreau. 2013. « Space-Time Variability of the Rainfall over the Western Mediterranean Region: A Statistical Analysis ». *Journal of Geophysical Research: Atmospheres* 118(15): 8448-59.

Sâadaoui M, Sakka M. 2007. «Pluviométrie en Méditerranée occidentale et Oscillation Nord Atlantique». *Publication de l'Association Internationale de Climatologie* 20: 17–22

Singla, Stéphanie. 2009 « Impact du changement climatique global sur les régimes hydroclimatiques au Maroc : tendances, ruptures et effets anthropiques sur les écoulements ». : 89.

Sinoga, J.D.R., Marin, R.G., Murillo, J.F.M., Galeote, M.A.G. 2011. « Precipitation Dynamics in Southern Spain: Trends and Cycles ». *International Journal of Climatology* 31(15): 2281-89.

Sood, A., G. M. James et G. J. Tellis 2009. "Functional regression: A new model for forecasting market penetration of new products." *Marketing Science* 28(1): 36-51

Xoplaki, E et al. 2000. « Connection between the Large-Scale 500 hPa Geopotential Height Fields and Precipitation over Greece during wintertime ». *Climate Research* 14: 129-46.

Taibi S., Meddi, M., Souag D., Mahe G. 2013. « Évolution et régionalisation des précipitations au nord de l'Algérie (1936–2009) ». In: *Climate and land surface changes in hydrology*, IAHS Publ. 359, 191-197.

Thiombiano, AN., St-Hilaire, A., El Adlouni, S., Ouarda, TBMJ. 2018. « Nonlinear Response of Precipitation to Climate Indices Using a Non-Stationary Poisson-Generalized Pareto Model: Case Study of Southeastern Canada ». *International Journal of Climatology* 38(S1): e875-88.

Ternynck, C., Ben Alaya, MA., Chebana, F., Dabo-Niang, S., Ouarda, T.B.M.J. 2016. « Streamflow Hydrograph Classification Using Functional Data Analysis ». *Journal of Hydrometeorology* 17(1): 327-44.

Trigo, R M., Pozo-vazquez, D., Osborn,TJ., Castro-diez ,Y., Gamiz-fortis,S., Esteban-parra , MJ. 2004. « North Atlantic Oscillation Influence on Precipitation, River Flow and Water Resources in the Iberian Peninsula ». *International Journal of Climatology* 24(8): 925-44.

Table 1 : Summary of some statistical characteristics of the total and maximum annual precipitation time series

	Total annual precipitation (mm)	Maximum annual precipitation (mm)
Max	1109.7	179.80
Min	328.3	29.00
Mean	545.4	64.08
Standard deviation	156.03	37.56

Table 2 : Sources of climate oscillation data

Climate oscillation	Short name	Source of data (links)
North Atlantic Oscillation	NAO	<a href="https://www.esrl.noaa.gov/psd/data/correlation/nao.data">https://www.esrl.noaa.gov/psd/data/correlation/nao.data</a>
Pacific Decadal Oscillation	PDO	<a href="https://www.esrl.noaa.gov/psd/data/correlation/pdo.data">https://www.esrl.noaa.gov/psd/data/correlation/pdo.data</a>
Western Mediterranean Oscillation	WeMO	<a href="http://www.ub.edu/gc/documents/Web_WeMOi.txt">http://www.ub.edu/gc/documents/Web_WeMOi.txt</a>
Mediterranean Oscillation	MO	<a href="https://crudata.uea.ac.uk/cru/data/moi/moac.dat">https://crudata.uea.ac.uk/cru/data/moi/moac.dat</a>



Table 3 : Degree of smoothing (df,  $\lambda$ ) related to smoothing indices

Climate index	1-month		3-month	
	df	$\lambda$	df	$\lambda$
NAO	3	10	6	$10^{-1}$
PDO	5	1	9	$10^{-3}$
MO	3	10	6	$10^{-1}$
WeMO	2	$10^2$	6	$10^{-1}$

Table 4 : Beta smoothing (df, lambda) and performance criteria (GCV, OCV, and RMSE) related to FLM-S for the 1-month scale

$\log_{10}$ (total annual precipitation)					
Climate indices	df	Lambda	GCV	OCV	RMSE
NAO	4	$10^{1.5}$	0.00043	0.387	0.111
PDO	4	$10^3$	0.00042	0.397	0.113
MO	3	$10^5$	0.00039	0.389	0.112
WeMO	3	$10^5$	0.00047	0.442	0.119
$\log_{10}$ (maximum annual precipitation)					
Climate indices	df	Lambda	GCV	OCV	RMSE
NAO	3	$10^5$	0.00152	1.418	0.214
PDO	4	$10^2$	0.00171	1.71	0.235
MO	5	$10^{-0.5}$	0.00151	1.447	0.216
WeMO	9	10-6.5	0.00138	1.353	0.209

Table 5 : Beta smoothing (df, lambda) and performance criteria (GCV, OCV, and RMSE) related to FLM-S for the 3-month scale

log <sub>10</sub> (total annual precipitation)					
	df	Lambda	GCV	OCV	RMSE
NAO	6	10 <sup>-0.5</sup>	0.00041	0.391	0.112
PDO	7	10 <sup>-1</sup>	0.0004	0.399	0.113
MO	6	10 <sup>-0.5</sup>	0.00047	0.464	0.122
WeMO	3	10 <sup>5</sup>	0.00046	0.422	0.116
log <sub>10</sub> (maximum annual precipitation)					
	df	Lambda	GCV	OCV	RMSE
NAO	3	10 <sup>4</sup>	0.00156	1.47	0.217
PDO	3	10 <sup>4</sup>	0.00144	1.427	0.214
MO	3	10 <sup>3</sup>	0.00138	1.266	0.202
WeMO	8	10 <sup>-2.5</sup>	0.00132	1.347	0.208

Table 6 : Significant correlations between climate indices and both total and maximum annual precipitations

1-month scale		
	total annual precipitation	maximum annual precipitation
NAO_Jan	0.361	-
NAO_March	0.381	-
NAO_Aug	0.371	-
MO_May	-0.377	-
WeMO_March	-	-0.310
3-month scale		
	total annual precipitation	maximum annual precipitation
NAO_DJF	0.339	-
NAO_JFM	0.458	-
PDO_SON	-0.489	-
PDO_OND	-0.536	-
PDO_NDJ	-0.499	-
PDO_DJF	-0.394	-
MO_MAM	-0.339	-
MO_NDJ	-	0.302
WeMO_OND	-	-0.393

Table 7 : Comparison between FLM-S and linear regression models to forecast total and maximum annual precipitation

		log <sub>10</sub> (total annual precipitation)				log <sub>10</sub> (maximum annual precipitation)		
		1-month scale		3-month scale		1-month scale	3-month scale	
		NAO	MO	NAO	PDO	WeMO	MO	WeMO
RMSE	FLM-S	0.111	0.112	0.112	0.113	0.209	0.202	0.208
%RMSE		4.078	4.115	4.115	4.151	11.916	11.517	11.859
F-ratio		6.044	5.451	13.884	17.226	28.110	4.841	25.049
p-value		0.02	0.027	0.0008	0.00027	1.09e-05	0.036	2.5e-05
RMSE	Linear regression	0.097	0.106	0.103	0.107	0.200	0.201	0.196
%RMSE		3.564	3.894	3.784	3.931	11.403	11.460	11.174
F-ratio		5.600	5.73	4.766	3.329	4.156	3.488	5.646
p-value		0.004	0.023	0.017	0.025	0.05	0.07	0.024

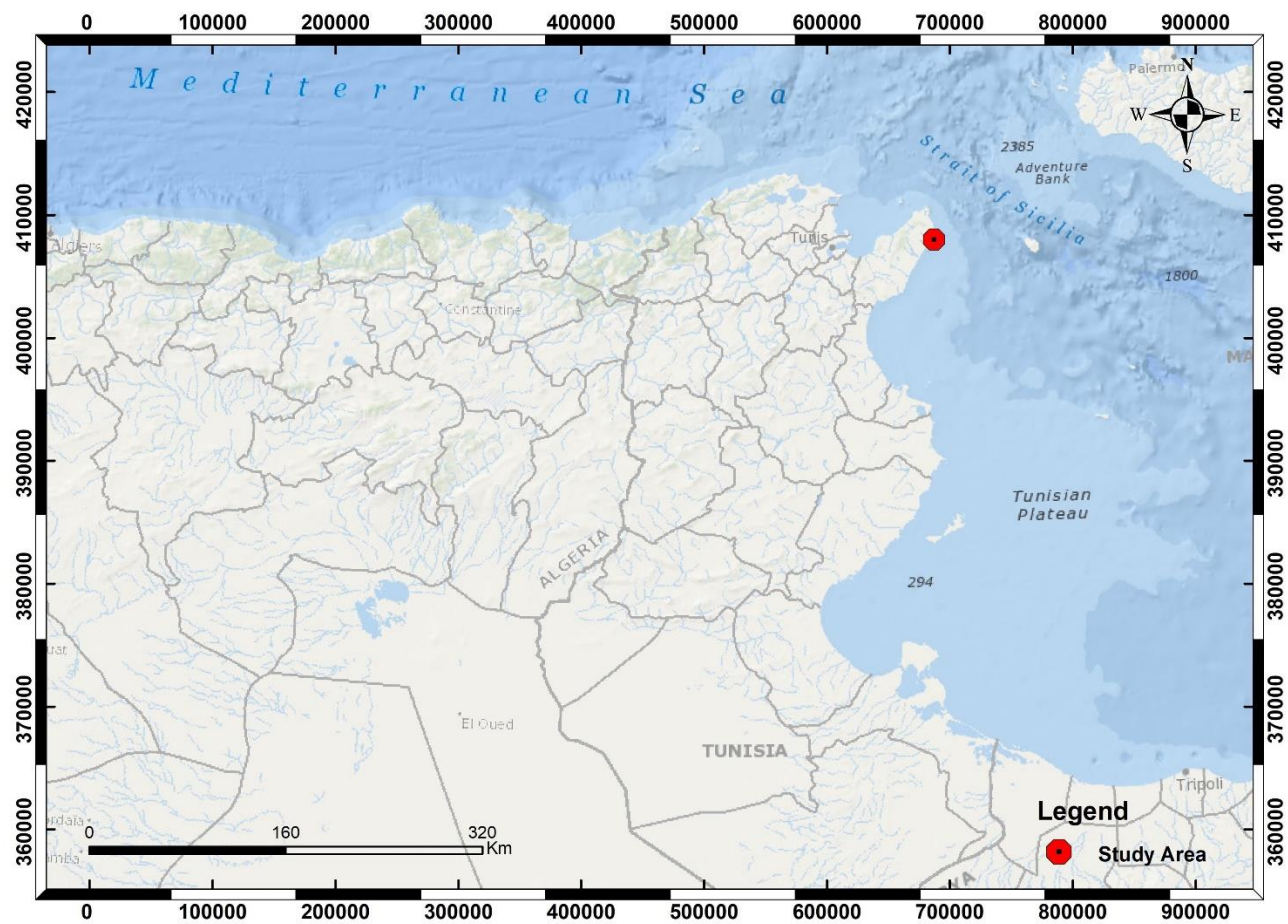


Figure 1 : Emplacement géographique de la zone d'étude « Kélibia »

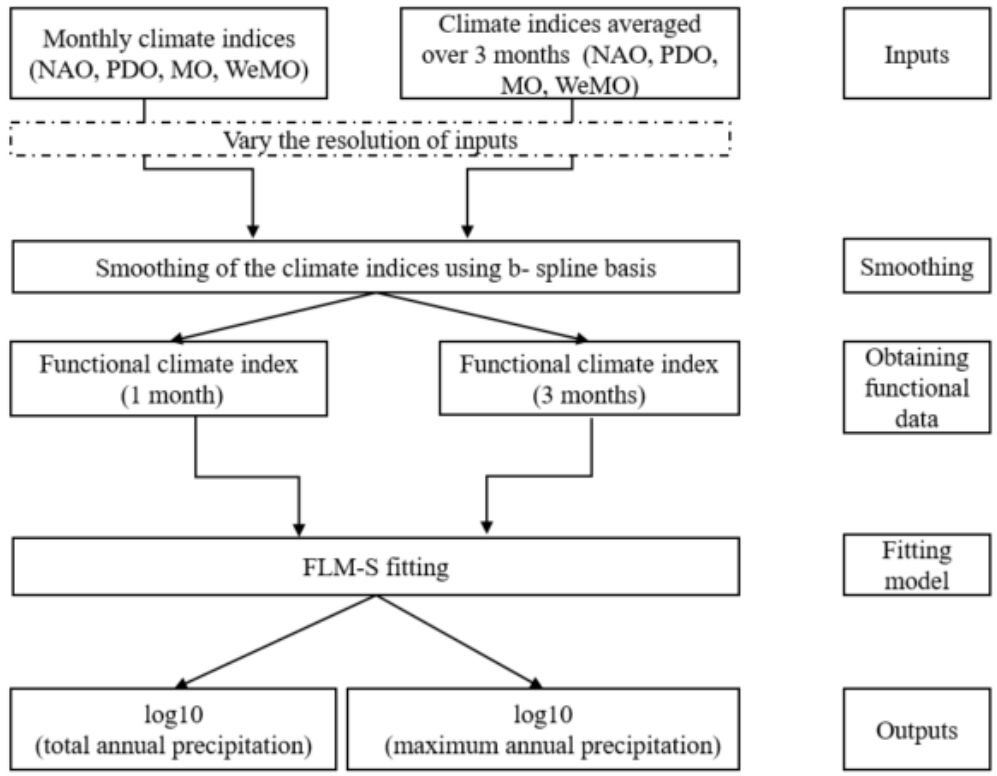


Figure 2 : Steps to fit FLM-S

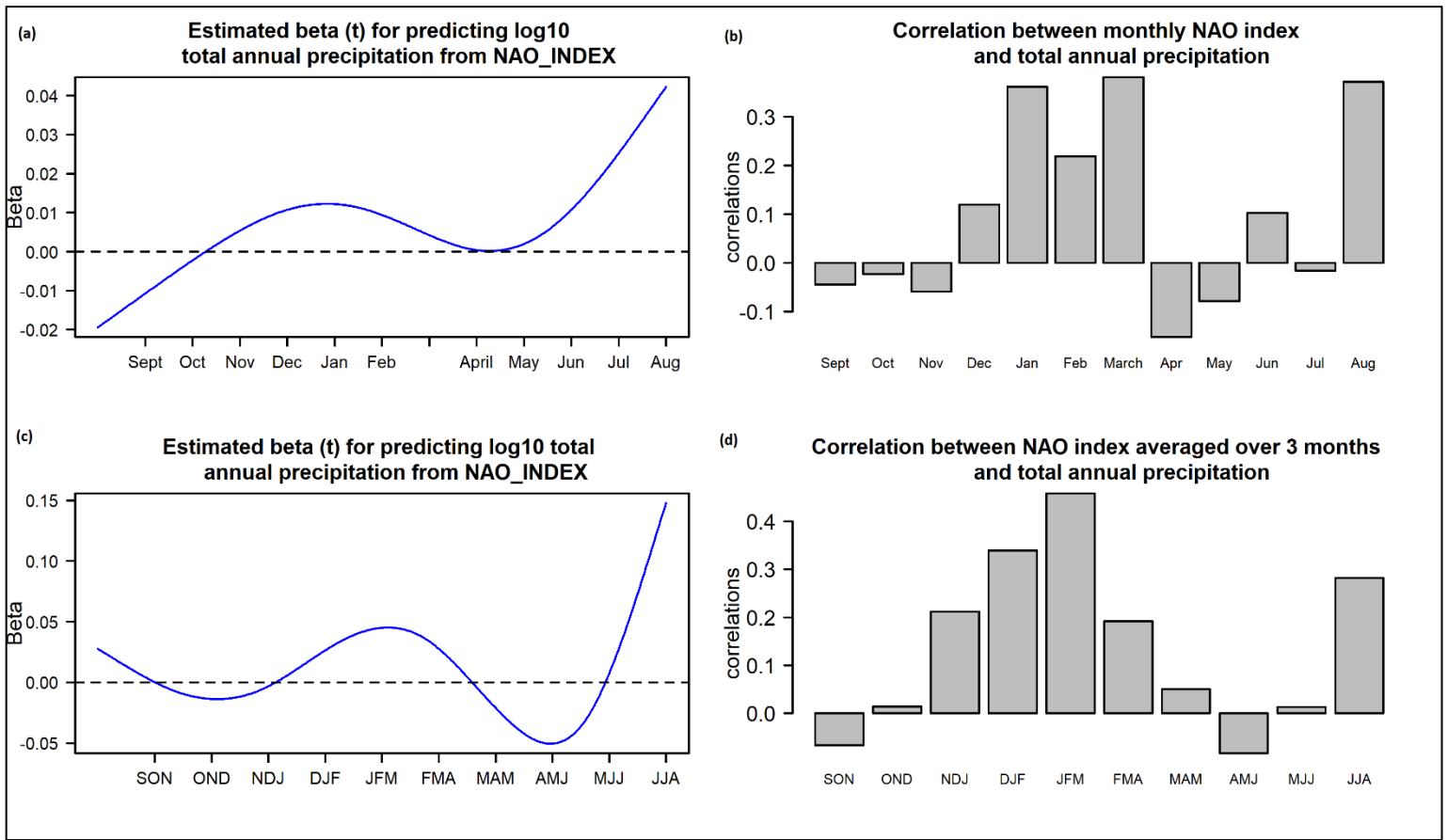


Figure 3 : estimated  $\beta(t)$  values for predicting the logarithm of total annual precipitation from the NAO climate index for both scales: (a) 1-month scale and (c) 3-month scale, as well as the curves of linear correlations between total annual precipitation and the NAO index for both scales: (b) 1-month scale and (d) 3-month scale.

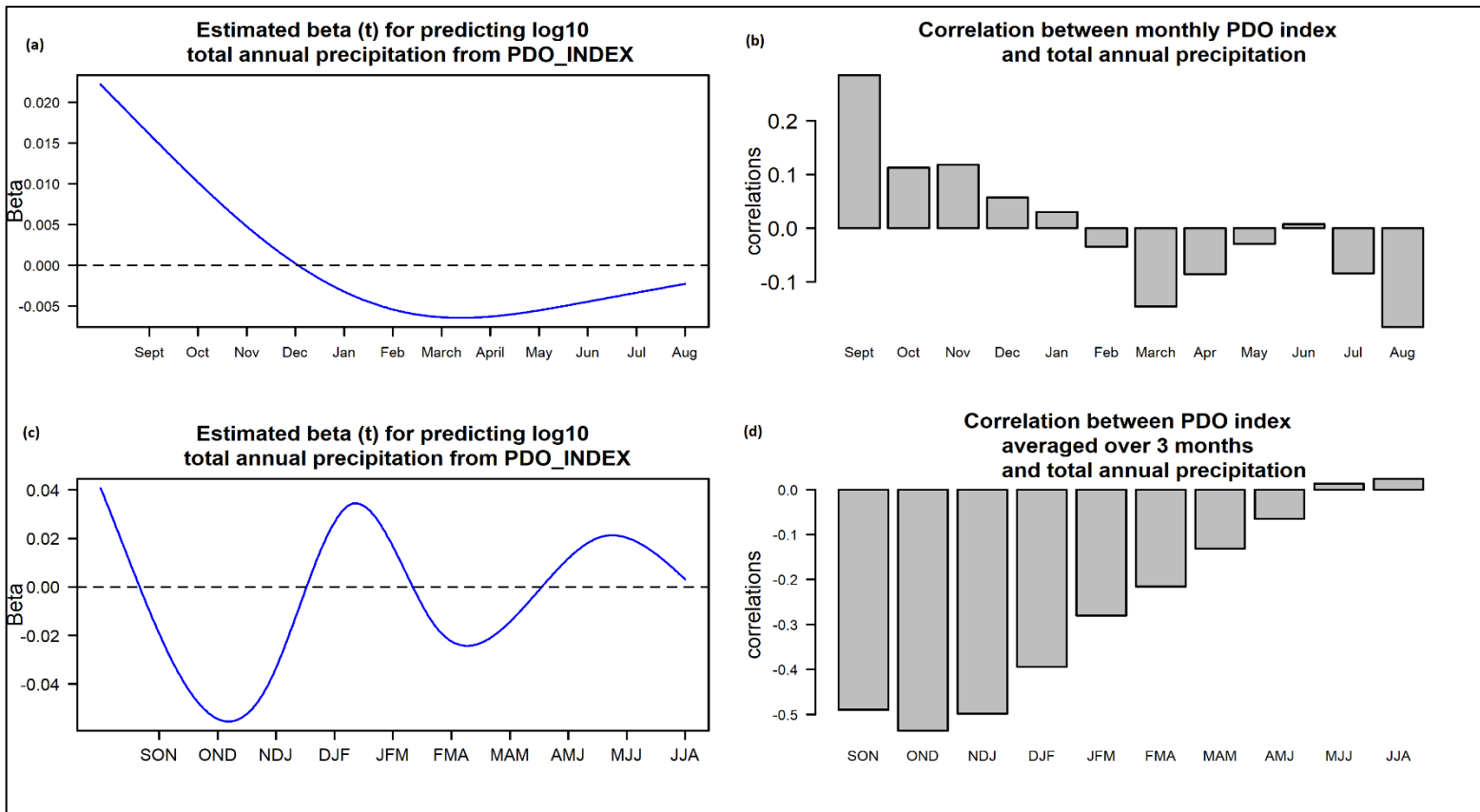


Figure 4 : estimated  $\beta(t)$  values for predicting the logarithm of total annual precipitation from the PDO climate index for both scales: (a) 1-month scale and (c) 3-month scale, as well as the curves of linear correlations between total annual precipitation and the PDO index for both scales: (b) 1-month scale and (d) 3-month scale.



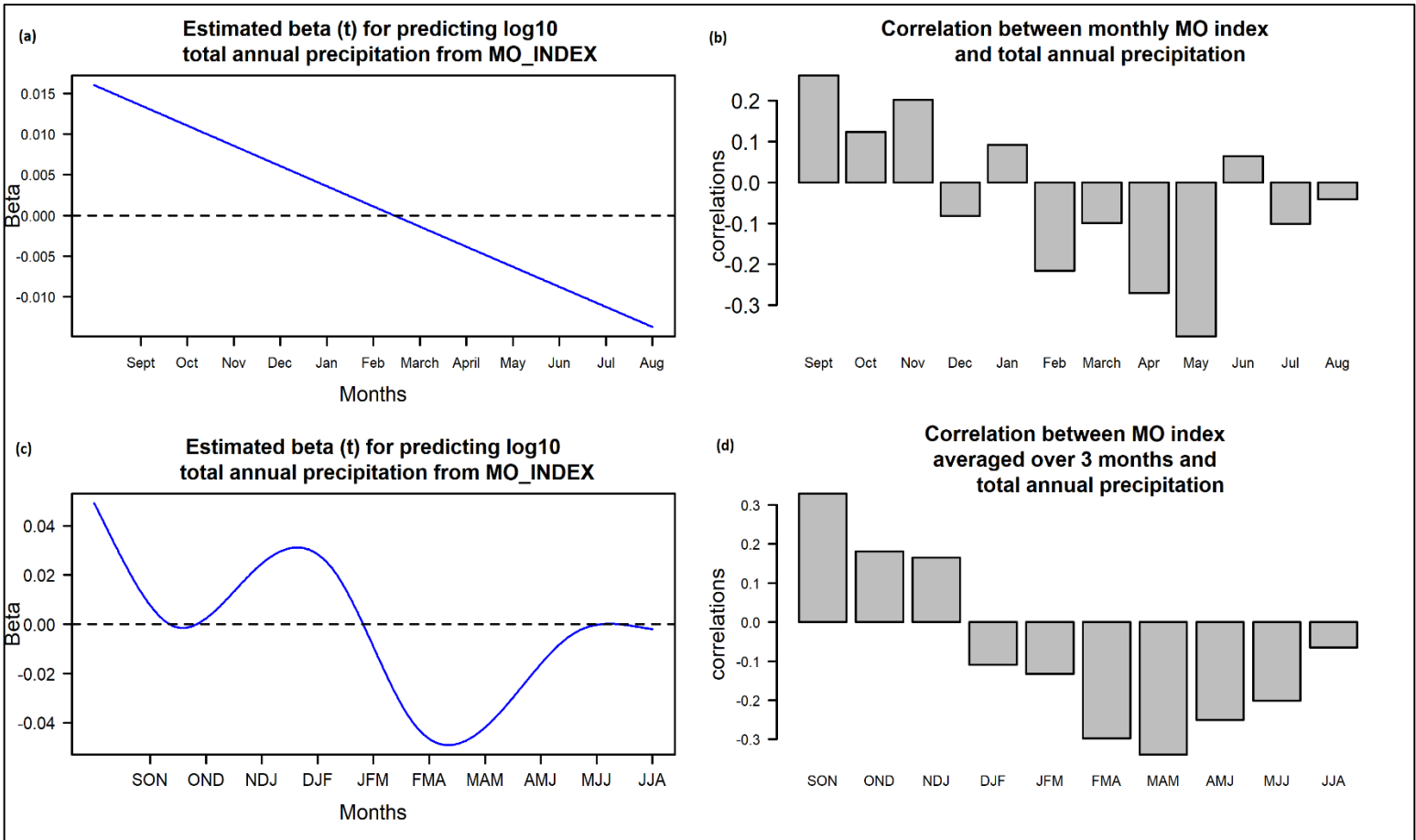


Figure 5 : estimated  $\beta(t)$  values for predicting the logarithm of total annual precipitation from the MO climate index for both scales: (a) 1-month scale and (c) 3-month scale, as well as the curves of linear correlations between total annual precipitation and the MO index for both scales: (b) 1-month scale and (d) 3-month scale.

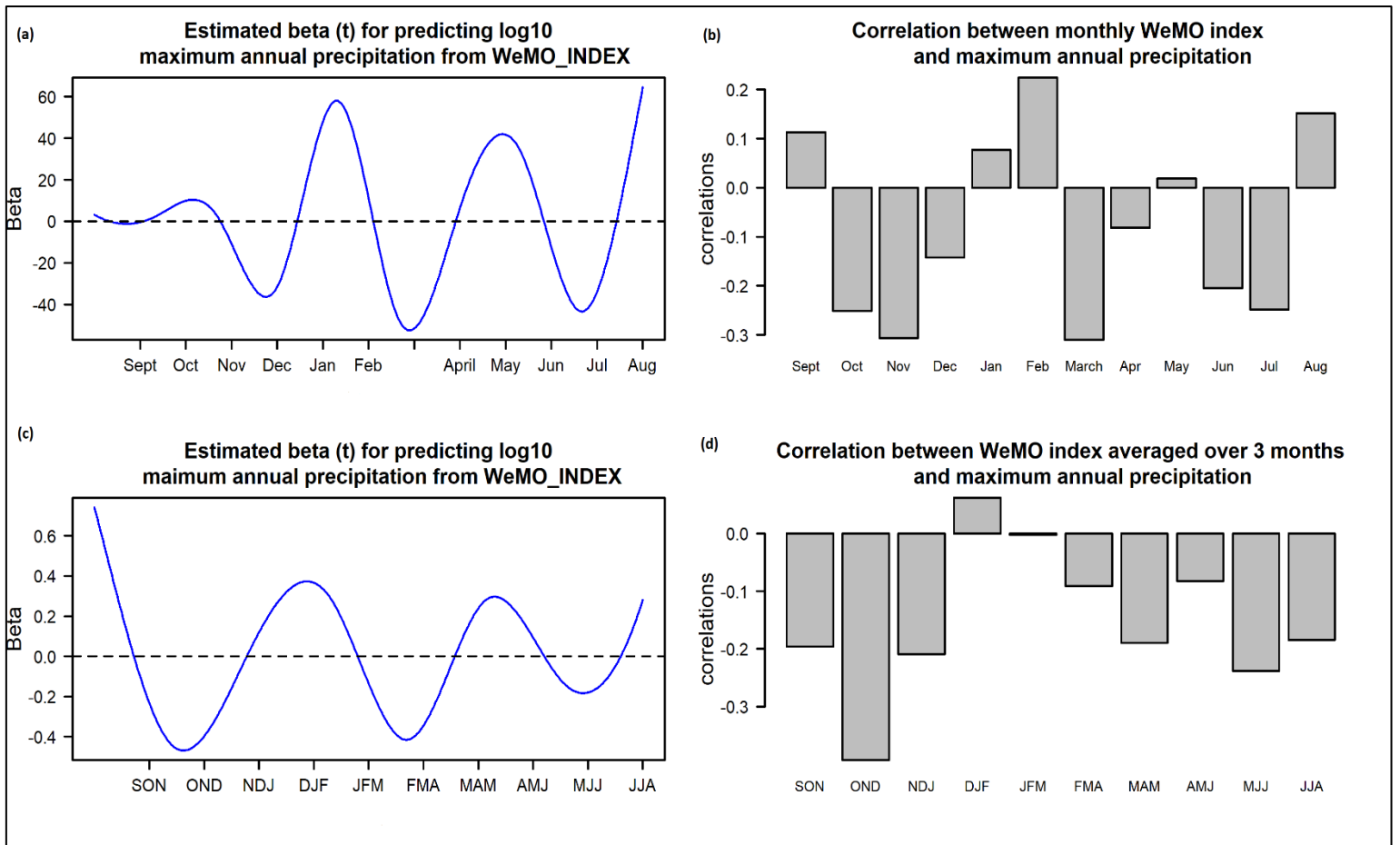


Figure 6: estimated  $\beta(t)$  values for predicting the logarithm of maximum annual precipitation from the WeMO climate index for both scales: (a) 1-month scale and (c) 3-month scale, as well as the curves of linear correlations between maximum annual precipitation and the WeMO index for both scales: (b) 1-month scale and (d) 3-month scale.