RÉPUBLIQUE DÉMOCRATIQUE POPULAIRE D'ALGÉRIE
Ministère de l'Enseignement supérieur et
Recherche scientifique
ÉCOLE NATIONALE POLYTECHNIQUE

Département Hydraulique
Laboratoire de recherche en sciences de l'eau

End-of-studies project dissertation for obtaining
the state engineer diploma in hydraulics

Using artificial intelligence techniques and standardised precipitation evapotranspiration index for meteorological drought forecasting

LYRAA Yasser

Under the direction of TACHI Salah Eddine and BENZIADA Salim

Presented and publicly supported on 03/07/2024

Composition of the Jury :

President	Mrs. TCHEKIKEN Chahinez	MCB	ENP
Promoter	Mr. TACHI Salah Eddine	MCA	Annaba University
Promoter	Mr. BENZIADA Salim	MAA	ENP
Examiner	Mr. BOUGUERRA Hamza	MCA	Annaba University
Examiner	Mr. HASNAOUI Yacine	PHD Student	ENP
Invited	Mr. SZCZEPANEK Robert	MCA	Jagiellonian University
Invited	Mrs. HATZAKI Maria	MCA	Athenes University

ENP 2024

RÉPUBLIQUE DÉMOCRATIQUE POPULAIRE D'ALGÉRIE
Ministère de l'Enseignement supérieur et
Recherche scientifique
ÉCOLE NATIONALE POLYTECHNIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

LRS EAU
ENP - ALGERIE

Département Hydraulique
Laboratoire de recherche en sciences de l'eau

End-of-studies project dissertation for obtaining
the state engineer diploma in hydraulics

USING ARTIFICIAL INTELLIGENCE TECHNIQUES AND STANDARDISED PRECIPITATION EVAPOTRANSPIRATION INDEX FOR METEOROLOGICAL DROUGHT FORECASTING

LYRAA Yasser
Under the direction of TACHI Salah Eddine and BENZIADA Salim

Presented and publicly supported on 03/07/2024

Composition of the Jury :

President	Mrs. TCHEKIKEN Chahinez	MCB	ENP
Promoter	Mr. TACHI Salah Eddine	MCA	Annaba University
Promoter	Mr. BENZIADA Salim	MAA	ENP
Examiner	Mr. BOUGUERRA Hamza	MCA	Annaba University
Examiner	Mr. HASNAOUI Yacine	PHD Student	ENP
Invited	Mr. SZCZEPANEK Robert	MCA	Jagiellonian University
Invited	Mrs. HATZAKI Maria	MCA	Athenes University

ENP 2024

RÉPUBLIQUE DÉMOCRATIQUE POPULAIRE D'ALGÉRIE
Ministère de l'Enseignement supérieur et
Recherche scientifique
ÉCOLE NATIONALE POLYTECHNIQUE

Département Hydraulique
Laboratoire de recherche en sciences de l'eau

Mémoire de Projet de Fin d'Étude en vue de l'obtention de diplôme ingénieur d'état en hydraulique

UTILISATION DES TECHNIQUES D'APPRENTISSAGE AUTOMATIQUE ET L'INDICE STANDARDISÉ DE PRÉCIPITATION ET D'ÉVAPOTRANSPIRATION POUR LA PRÉVISION DE SÉCHRESSE MÉTÉOROLOGIQUE

LYRAA Yasser
Sous la direction de TACHI Salah Eddine et BENZIADA Salim
Présenté et soutenu publiquement le 03/07/2024

Composition du Jury :

Présidente	Mrs. TCHEKIKEN Chahinez	MCB	ENP
Promoteur	Mr. TACHI Salah Eddine	MCA	Université d'Annaba
Promoteur	Mr. BENZIADA Salim	MAA	ENP
Examinateur	Mr. BOUGUERRA Hamza	MCA	Université d'Annaba
Examinateur	Mr. HASNAOUI Yacine	Doctorant	ENP
Invité	Mr. SZCZEPANEK Robert	MCA	Université Jagellonne
Invité	Mrs. HATZAKI Maria	MCA	Université d'Athènes

ENP 2024

الملخص : تبحث هذه الأطروحة تطبيق الاستخبارات الصناعية لرصد الجفاف والتنبؤ به في المنطقة الشمالية الشرقية من الجزائر، باستخدام مؤشرات الدوران الجوي ومؤشر .(SPEI) والهدف الرئيسي هو وضع نماذج تنبؤ دقيقة وتحديد مؤشرات دوران الغلاف الجوي الأكثر تأثيرا في هذه المنطقة.فالجفاف، الذي يشكل خطراً طبيعياً متكرراً في الجزائر، يؤثر تأثيراً كبيراً على الزراعة والموارد المائية والأنشطة الاجتماعية - الاقتصادية. وكثيراً ما تقصر أساليب الرصد التقليدية في التنبؤ ببداية الجفاف وشدته. وتستفيد هذه الدراسة من تقنيات الذكاء الإصطناعي، وتحديداً نموذج غابة العشوائيات، لتحليل وتفسير مجموعات البيانات الكبيرة التي تشمل متغيرات مناخية ومؤشرات دوران الغلاف الجوي وتشمل منهجية البحث جمع البيانات المناخية التاريخية ومؤشرات الدوران الجوي ذات الصلة بمنطقة شمال شرق الجزائر. ويُستخدم مؤشرSPEI ، الذي يتضمن بيانات عن الهطول والحرارة على حد سواء، لقياس أحوال الجفاف. ويجري تدريب نموذج الغابات العشوائية والتحقق منه للتنبؤ بقيم SPEI استناداً إلى مؤشرات الغلاف الجوي المختارة.وتبين النتائج أن نموذج غابة العشوائيات يحقق دقة عالية في التنبؤ بالجفاف، حيث ثبت أن بعض مؤشرات الدوران في الغلاف الجوي هي عوامل تنبؤ هامة. وتسلط النتائج الضوء على إمكانات المبادرة في تعزيز نظم رصد الجفاف، وتوفير المعلومات الموثوقة في الوقت المناسب لصنع القرار وإدارة الموارد.ولا تسهم هذه الدراسة في فهم ديناميات الجفاف في الجزائر فحسب، بل توفر أيضاً إطاراً لإدماج منظمة العفو الدولية في نظم الرصد البيئي. ويؤدي النجاح في تحديد مؤشرات الغلاف الجوي المؤثرة إلى زيادة إثراء المعارف العلمية اللازمة لوضع نماذج قوية للتنبؤ.

الكلمات الرئيسية : الجفاف, مؤشرات دوران الغلاف, المنطقة الشمالية الشرقية من الجزائر.

Résumé : Cette thèse étudie l'application de l'intelligence artificielle (IA) pour surveiller et prévoir la sécheresse dans la région du Nord-Est d'Algérie, en utilisant les indices de circulation atmosphérique et l'indice normalisé de précipitation-évapotranspiration (SPEI). L'objectif principal est d'élaborer des modèles de prévision précis et d'identifier les indices de circulation atmosphérique les plus influents dans cette région. La sécheresse, un danger naturel récurrent en Algérie, a des répercussions importantes sur l'agriculture, les ressources en eau et les activités socioéconomiques. Les méthodes de surveillance traditionnelles sont souvent insuffisantes pour prédire l'apparition et la gravité de la sécheresse. Cette étude utilise des techniques d'IA, en particulier le modèle Random Forest, pour analyser et interpréter de grands ensembles de données comprenant des variables climatiques et des indices de circulation atmosphérique. La méthodologie de recherche comprend la collecte de données climatiques historiques et d'indices de circulation atmosphérique pertinents pour la région du Nord-Est de l'Algérie. L'indice SPEI, qui intègre les données de précipitation et de température, est utilisé pour quantifier les conditions de sécheresse. Le modèle Random Forest est formé et validé pour prédire les valeurs SPEI en fonction des indices atmosphériques sélectionnés. Les résultats démontrent que le modèle Random Forest atteint une grande précision dans la prévision de la sécheresse, avec certains indices de circulation atmosphérique qui se révèlent être des prédicteurs importants. Les résultats soulignent le potentiel de l'IA pour améliorer les systèmes de surveillance de la sécheresse, offrant des informations opportunes et fiables pour la prise de décisions et la gestion des ressources. Cette étude contribue non seulement à la compréhension de la dynamique de la sécheresse en Algérie, mais fournit également un cadre pour l'intégration de l'IA dans les systèmes de surveillance environnementale. L'identification réussie d'indices atmosphériques influents enrichit davantage les connaissances scientifiques nécessaires à l'élaboration de modèles prédictifs solides.

Mots-clés : les indices de circulation atmosphérique, SPEI, sécheresse, Nord-Est d'Algérie.

Abstract : This thesis investigates the application of Artificial Intelligence (AI) to monitor and forecast drought in the Northeast region of Algeria, utilizing atmospheric circulation indices and the Standardized Precipitation-Evapotranspiration Index (SPEI). The primary objective is to develop accurate forecasting models and identify the atmospheric circulation indices most influential in this region. Drought, a recurrent natural hazard in Algeria, significantly impacts agriculture, water resources, and socio-economic activities. Traditional monitoring methods often fall short in predicting drought onset and severity. This study leverages AI techniques, specifically the Random Forest model, to analyze and interpret large datasets comprising climatic variables and atmospheric circulation indices. The research methodology includes collecting historical climate data and atmospheric circulation indices relevant to the Northeast region of Algeria. The SPEI index, which incorporates both precipitation and temperature data, is used to quantify drought conditions. The Random Forest model is trained and validated to predict SPEI values based on the selected atmospheric indices. Results demonstrate that the Random Forest model achieves high accuracy in forecasting drought, with some atmospheric circulation indices proving to be significant predictors. The findings highlight the potential of AI to enhance drought monitoring systems, offering timely and reliable information for decision-making and resource management. This study not only contributes to the understanding of drought dynamics in Algeria but also provides a framework for integrating AI into environmental monitoring systems. The successful identification of influential atmospheric indices further enriches the scientific knowledge required for developing robust predictive models.

Keywords : atmospheric circulation indices, SPEI, drought, AI, Northeast of Algeria.

Acknowledgements

First and foremost, I would like to express my gratitude to God Almighty, who has given us the strength and patience to complete this humble work. Our heartfelt thanks go to ; our beloved parents for their support and patience.

Our thesis supervisors, Mr. TACHI Salah Eddine and Mr. BENZIADA Salim, for their invaluable support throughout our project.

We also thank the president and members of the jury who will have the honor of evaluating our work.

Our thanks also extend to all our teachers throughout the years of study. We appreciate every teacher who has aided us with their knowledge from the early stages of school until this moment.

We would like to extend our sincerest gratitude to all those who have contributed, directly or indirectly, to the development of this thesis. To all our close friends and the entire hydraulic class of 2024, To my friend Mohamed, Zakaria and Fawzi we express our gratitude.

Table of Contents

List of Figures

List of Tables

GENERAL INTRODUCTION

GENERAL INTRODUCTION

Drought is a recurrent and severe climatic event that poses significant threats to ecosystems, water resources, and agricultural productivity, particularly in regions such as the northern lands of Algeria. Characterized by prolonged periods of deficient rainfall, droughts can lead to substantial economic losses and adverse social impacts. Effective monitoring and prediction of drought events are essential for mitigating their effects and formulating adaptive strategies. Traditional methods of drought prediction, often reliant on historical climatic data and statistical techniques, have shown limitations in accuracy and adaptability to changing climatic patterns [1].

In recent years, the advent of machine learning (ML) has revolutionized the field of environmental modeling and prediction [2]. Machine learning algorithms, with their ability to handle large datasets and uncover complex, non-linear relationships, offer promising advancements in the predictive modeling of drought events. This study leverages the power of machine learning models to predict drought occurrences in the northern lands of Algeria, utilizing atmospheric circulation indices as predictive features and the Standardized Precipitation Evapotranspiration Index (SPEI) as the target label.

The northern region of Algeria, characterized by its Mediterranean climate, is particularly susceptible to drought due to its variability in precipitation and high evapotranspiration rates [3] . Atmospheric circulation indices, such as the North At- lantic Oscillation (NAO), the Mediterranean Oscillation (MO), and the El Nin˜o Southern Oscillation (ENSO), have been identified as influential factors affecting re- gional climatic conditions. These indices represent large-scale climate patterns that influence weather and climate variability on a global scale, thereby serving as crucial indicators for drought prediction.

The primary objective of this research is to develop and evaluate machine lear- ning models that can accurately predict drought events in the northern lands of Algeria. By integrating atmospheric circulation indices as explanatory variables and utilizing the SPEI as a measure of drought severity, this study aims to enhance the understanding of the underlying climatic drivers of drought and improve predictive capabilities. The SPEI, which incorporates both precipitation and potential evapo- transpiration, provides a comprehensive measure of drought conditions, making it an ideal target for this modeling endeavor.

This thesis is structured as follows : the first section provides a detailed review of the literature on drought prediction methods and the application of machine learning in environmental modeling. Following this, the methodology section describes the data sources, preprocessing steps, and the machine learning algorithms employed in this study. Then, Then, The results section presents the performance of the developed models, highlighting the most influential atmospheric circulation indices. Finally, the discussion and conclusion sections interpret the findings, discuss the implications for drought management, and suggest directions for future research.

By harnessing the capabilities of machine learning and utilizing relevant climatic indices, this research contributes to the growing body of knowledge aimed at improving drought prediction and management in vulnerable regions. The findings of this study are expected to provide valuable insights for policymakers and stakeholders involved in water resource management and agricultural planning in Algeria.

Chapter I

STATE OF THE ART AND LITERATURE REVIEW

Introduction

This chapter delves into state of the art on drought, providing a comprehensive exploration of the various definitions, classifications, types, causes, and consequences of drought. Understanding the multifaceted nature of drought is essential for effectively modeling and predicting its occurrence, particularly in regions susceptible to climatic variability such as the northern lands of Algeria. The chapter begins by discussing different definitions of drought, highlighting the variability in how drought is conceptualized across disciplines and contexts.

Following this, the classification of these definitions is presented, offering a structured framework to understand the diverse perspectives on drought. The types of drought—meteorological, agricultural, hydrological, and socio-economic—are then examined, each with distinct characteristics and impacts. The causes of drought, both natural and anthropogenic, are explored to provide insight into the complex interplay of factors that contribute to drought conditions. The consequences of drought are discussed, emphasizing the far-reaching effects on ecosystems, economies, and societies.

Finally, the chapter reviews the latest works similar to our study, focusing on research that utilizes atmospheric circulation indices and drought indices for drought analysis. This review highlights the relevance and significance of integrating these indices in drought prediction models, setting the stage for our own methodological approach.

I.1 Drought definition

Drought is not a term that is agreed upon by all. Definitions must to be particular to the application and pertinent to the area being studied. There are two types of drought indicators : quantitative (needing statistical analysis) and qualitative (typically descriptive, language definitions of drought severity).[4].

I.1.1 Drought definition according to Wilhite et Glantz (1985)

Wilhite and Glantz have developed more than 150 published definitions. They describe drought conceptually (as an idea or concept) and operatively (by how drought works or is done so that it can be measured). But generally speaking, drought can be defined as a prolonged period of insufficient rainfall, over one or more seasons or even years, causing water shortages in certain sectors of a country's economy.[5].

I.1.2 Definition according to National Weather Service

The National Weather Service defines drought as a sufficiently large and prolonged shortage of rainfall that could have adverse consequences for the plant and animal life of a location and exhaust water reserves, both for domestic needs and for the operation of power plants, especially in areas where rainfall is normally sufficient for these purposes.[5].

I.2 Classification of drought definition

There are different definitions of drought depending on the influence factor used. In literature, they can be grouped as follows :

- Definition of drought based on precipitation.
- Definitions of drought based on evapotranspiration.
- Definition of weather drought.
- Definition of drought by flow.
- Definition of drought based on soil humidity.
- Definition of vegetation-based drought

I.2.1 Definition of drought based on precipitation.

I.2.1.1 Blumenstock (1942)

defines drought as a period in which precipitation is less than a small amount of 0.1 inch (2.54mm) in 48 hours [6].

I.2.1.2 Palmer (1957)

defines drought as monthly or annual rainfall below a certain percentage of normal [6].

I.2.1.3 The National Agricultural Commission of the United States (1965)

defines drought as a time when rainfall is lower than normal. Once again, the National Committee on Agriculture (1976) classified weather drought as a situation where there is a significant decrease (more than 25 percent) from the normal period of 4 consecutive weeks from mid-May to mid-October [6].

I.2.1.4 Thiruvengadachari (1988)

confirmed that the efficiency of using rainfall varies in both time and space, thus limiting the use of rain as a single or primary indicator of drought [6].

I.2.2 Definitions of drought based on evapotranspiration

I.2.2.1 Thornthwaite (1948)

defines drought as a condition in which the amount of water needed for sweating and direct evaporation (i.e. potential evapotranspiration) exceeds the available soil moisture[6].

I.2.3 Definition of weather drought

I.2.3.1 Condra (1944)

defines drought as a period of strong wind, low precipitation, high temperature and low humidity[6].

I.2.3.2 Palmer (1965)

defines drought as a situation where the actual precipitation is lower than the climate precipitation suitable for existing conditions [6].

I.2.3.3 The World Meteorological Organization (1975)

classified atmospheric drought as drought involving precipitation, temperature, humidity and wind speed [6].

I.2.4 Definition of drought by flow

I.2.4.1 Yevjevich (1967)

suggested that drought constituted a deficit of flow relative to long-term average flow [6].

I.2.4.2 Joseph (1970)

defines drought as a period characterized by the lowest average flow at a specified measurement point in a stream of rivers for 14 consecutive days during a climate year beginning on April 1st.[6].

I.2.5 Definition of drought based on soil moisture

Agricultural drought is generally defined as the lack of soil moisture to meet crop water requirements.

I.2.5.1 The American Meteorological Society

defines drought as the prolonged and abnormal deficit of soil moisture [6].

I.2.5.2 Shantz (1970)

defines drought as a situation in which the humidity of the available soil decreases, so that vegetation can no longer absorb soil water quickly enough to meet the transpiration needs [6].

I.2.5.3 The World Meteorological Organization (1975

defines drought as the lack of soil moisture in terms of plant behavior, for a given crop [6].

I.2.5.4 The National Commission for Agriculture (1976)

defines drought as the period when soil moisture is insufficient during the crop season to allow the type of crop growth to reach its grain yield potential [6].

I.2.5.5 Smith (1978)

defines drought as a condition in which a plant does not develop or ripen properly due to insufficient humidity [6].

I.2.6 Vegetation-related definitions of drought

I.2.6.1 Thiruvengadachary (1988)

assesses drought over 15-day periods, based on analysis of the vegetation index map and green maps, as well as statistics on vegetation Index [6].

I.3 Types of drought

According to the following operational definitions (Wilhite and Glantz, 1985), there are four main types of drought : meteorological, hydrological, agricultural and socio-economic [5].

I.3.1 Weather drought

Weather drought is based on the degree of dryness of a dry period compared to normal (median or average) and the duration of that dry period [7]. It is characterized by the absence of rainfall in a given region over a certain period of time. It is often defined, for an area characterized by seasonal rainfall, as the measurement of the deviation of cumulative rainfall over a given period from the normal of that period, calculated over at least 30 years. In regions where rain is received throughout the year, the definition of drought is based on the number of days when rainfall is below a given critical level [8]. Some definitions of meteorological drought have been developed to be applied in various countries around the world [5] :

1. Less than 2.54 mm of rainfall in 48 hours (United States).
2. 15 days, none of which reached 0.25 mm (Grande-Bretagne).
3. When annual rainfall is less than 180 mm (Libye).
4. Effective seasonal precipitation is more than twice the average gap (Inde).
5. A period of 6 days without rain (Indonesia).

I.3.2 Hydrological drought

Hydrological drought is associated with the effects of periods of precipitation, snowfall on surface or groundwater reserves (weather drought, if prolonged, leads to hydrologic drought). The frequency and severity of hydrological droughts are often defined at the scale of a spilling basin or a basin. Although all droughts are caused by a rainfall deficit, hydrologists are more concerned about how this deficit affects the hydrological system. Hydrological droughts are generally lagging or lagging behind weather and agricultural drought. It takes more time for insufficient precipitation to appear in components of the hydrological system such as soil moisture, stream flow and freatic watercourses with marked surface water depletion and drying of inland water masses such as lakes, dams, etc. Consequently, these impacts are phased out from those of other economic sectors [9].

I.3.3 Agricultural drought

Agricultural drought associates various characteristics of weather (or hydrological) drought with impacts on agriculture including rainfall shortages, differences between actual and potential evapotranspiration, soil water deficits, reduced groundwater or reservoir levels, etc. The demand for plant water depends on weather conditions, the biological characteristics of the plant, its growth stage and the physical and biological properties of the soil. A proper definition of agricultural drought should be able to take into account the variable sensitivity of crops during the different stages of crop development, from raising to ripening. Inadequate moisture in the arable layer during planting can impair germination, resulting in a low plant population per hectare and a reduction in final yield [9].

I.3.4 The socio-economic drought

It occurs when water shortages begin to affect people and their lives. It combines economic assets and elements of meteorological, agricultural and hydrological droughts. This type of definition differs from others by the fact that this drought is based on the supply and demand process. Socio-economic drought occurs when the supply of an economic good or commodity (e.g. drilling, hydroelectric power...) can no longer satisfy the demand for that product and the cause of this deficit is related to the climate, and specifically to the lack of precipitation [1]. In most cases, demand for economic goods is increasing as a result of population growth and per capita consumption. Supply may also increase as a result of improved efficiency of production and the technology used [9].

I.4 Causes and consequences of drought

I.4.1 Causes of drought

Insufficient rainfall

In arid and semi-arid regions, drought is when a region experiences long periods without rain, especially for more than one season [10].

Human Causes

Human activities play a relatively important role in the management of the water cycle, such as deforestation, construction and agriculture, which have a negative impact on the water Cycle. Trees and plant coverings are essential to the water cycle as they help limit evaporation, store water and attract rainfall [10].

Drying of surface water flow

Lakes, rivers and streams are the main suppliers of downstream surface water in various geographical regions of the world. During extremely hot seasons or due to certain human activities, open water surfaces may dry downwards, contributing to significant drought and the demand for water becoming higher than available water. Irrigation systems and hydroelectric dams are among the human activities that can significantly reduce the amount of water flowing downstream to other areas [10].

Global warming

Human actions are contributing to increased greenhouse gas emissions into the atmosphere, thereby resulting in a continuous rise in average global temperatures. As a result, evaporation and evapotranspiration levels have increased and higher temperatures have resulted in forest fires and prolonged drought periods [10].

I.4.2 Consequences and impacts of drought

Environmental impacts of drought

Drying of water masses : surface waters such as lakes, rivers, ponds, streams, and lagoons dry out during prolonged dry conditions that destroy natural habitats. Decrease in soil quality : it is reduced due to the reduction in organic activity and drought that kills soil organisms. Inappropriate conditions for the survival of plants and vegetation : drought leads to the loss of fertile land and consequent desertification. Migration and even death of wildlife : animals migrate and find themselves in new places where they can be vulnerable and threatened. This leads to loss of biodiversity and disruption of natural ecosystems [10].

Economic impact of drought

Increased budgetary expenditure by farmers : in times of drought, farmers spend more money on irrigation of crops in order to maintain yields. Industrial and governmental losses and higher energy costs for hydropower-dependent economies. Outbreak of water-borne diseases, migration of people and anxiety, and health consequences (Hunger, anemia, malnutrition and death) [10].

I.4.3 Latest works and researches on drought phenomena

Because of its detrimental impacts on the ecosystem, drought evaluation is crucial to the management of water resources. Many developed indices are available to determine the level of drought in every given region of the world, and each one needs to be validated by a research before it can be used in that particular area. The Standardized Precipitation Index (SPI) and the Reconnaissance Drought Index (RDI) are two assessment instruments used to determine the severity of the drought episode.[11].

The World Meteorological Organization advised that the SPI be applied globally as a drought assessment index with consideration of the region's meteorological circumstances and limitations. [12].

The assesement of climate change in a specific area is based on two factors which are precepitation and air temperature, based on these parameters we can get a significant drought indices that has a good estimation of scarcity of water in that region by taking measurement of these two parameters on differnt points in that region and using GIS interpolation methods like IDW to get spatial maps of that index that reflects the state of the region everywhere [13].

Drought indices that take into account two meteorological parameters—precipitation and air temperature—allow us to assess the extent of climate change in a given area. These parameters allow us to generate a significant drought index that provides a reliable estimate of the region's water scarcity. [14].

The Palmer Drought Index (PDI) is a tool used to gauge the severity of droughts ; a specific index value is frequently used to determine when to start or stop specific drought contingency plan components. To measure a precipitation shortfall at various time periods, the Standardized Precipitation Index (SPI) was recently created. It was intended to serve as a drought indicator that acknowledges the significance of temporal scales in the examination of water supply and use. [15].

Richard and Michael invented "Standardized Anomaly Index" To track drought occurrences in two specific regions—the Northeast of Brazil and the West African Sahel [16].

The drought indices can be used for more than only determining the availability of water resources. For example, they can be used to control forest fires, which provide a useful method of identifying areas that are susceptible to fire outbreaks and aid in putting out these deadly incidents. [17].

El Niño is a meteorological phenomenon characterized by atypically high water temperatures in the southern Pacific Ocean. This occurrence has a significant impact on drought at the global warming scale ; Lyon has provided insightful information on this topic. [18].

In a comparison of the SPI and RDI, Jamshidi et al. found that while both are useful indices, the RDI shows more evidence of the Iranian drought than the SPI index. [19].

Due to their effectiveness, predictability, and strong correlation with actual conditions, the Palmer Drought Severity Index (PDSI) and the Crop Moisture Index (CMI) are the most frequently used drought indices in the agricultural sector. However, there are other indices that are developed for specific regions, such as the Crop-Specific Drought Index (CSDI) model, which was developed by Meyer et al using data from an 8-year study conducted on certain USA lands. [20].

Since evapotranspiration varies depending on elevation and climate, evapotranspiration formulations greatly influence the recognition of drought indices and influence the outcome.In their study, Mohamed and Scholz concentrated on evaluating drought using the Penman-Monteith reference technique of the Food and Agriculture Organization, Thornthwaite, Hargreaves, and Blaney-Criddle techniques to calculate the RDI and demonstrate the index's sensetivity to the evapotranspiration formula used to calculate any drought index. [21].

Temperature trends have opposite relationships with SPI and SPEI trends ; temperature trends have a direct relationship with SPI trends and an inverse relationship with SPEI trends, according to a comparison study conducted by Nwayor and Robeson. Continued rises in temperature and the vapor-pressure deficit are probably going to lead to a global decoupling of the two indicators. [22].

Berhail and Katipoğlu made a study on the Mekerra watershed to determine the correlatoin between the SPI and SPEI , the study gave a high correlation between these two indices, the researchers observed a positive linear correlation ($r > 0.75$, $p < 0.0001$) [23].

Ortiz-Gómez et al conducted a study by using data of 14 meteorological stations in North Central Mexico to check the sensitivity of RDI and SPEI to type of evatranspiration methods (Thornthwaite, Hargreaves – Samani, Droogers – Allen, Allen, Dorji, Priestley – Taylor, Makkink and Irmak.) [24].

Xu et al made a study in China using the ARIMA-LSTM model on the SPEI index, the model was found suitable for the forecasting of long-term drought in China [25].

Based on SPI and SPEI, Mousavi et al. conducted research on historical drought conditions in the South Saskatchewan River Watershed in Alberta, Canada. It was discovered that SPEI performs better at identifying dry situations. [26].

As an exemple of studies made to get reliable prediction in order to reduce vulnerability and improve management of drought-dependent businesses. It involves the prediction of SPEI values using MP5 and support vector machine (SVR) in southern Italy. It shows to be particularly suitable for drought forecasting in areas with long and severe drought events [27].

Also a study of drought forcasting in central India used the SPEI coupled with machine learning model called Multilayer Perceptron Artificial Neural Network Model. The results indicate that the forecasted SPEI values are suitable for drought prediction [28].

I.5 The developped research

In the present study, we aim to find a relation between The atmospheric circulation patterns and drought in East Northern Algeria by using SPEI index and this by using

Recent Machine learning models to get better results. There were different studies that have been published in the same context, and we enumerate some of these works : LUP-PICHINI et al., 2021, A studiedy the relationship between NAO and rainfall regime in a key area of the Mediterranean basin by using advanced statistical methods that made it possible to investigate a specific area of the Mediterranean and then to extend and contextualize to more geographical locations.

Another study on the Maharloo Basin in Iran by KHALIGHI et al., 2013, about long-term precipitation fore- cast for drought relief using atmospheric circulation factors using an artificial neu- ral network (ANN) and multi-regression stepwise methods,It was found that NAO (North Atlantic Oscillation), PNA (Pacific North America) and El Nin˜o are the main indices to forecast drought in the study area.

PARK et al., 2016 investigated the different regions of the united states of America, where 16 multi-sensor indices from the Moderate Resolution Imaging Spectroradio- meter(MODIS) and Tropical Rainfall Measuring Mission (TRMM) satellite sensors were used in addition with machine learning models like random forest, boosted regres- sion trees, and Cubist. the results showed a strong visual agreement between the real data and predicted data from this model PARK et al., 2016.

SAHA et al., 2021, constructed machine learning models. Namely, Bagging and Artificial Neural Network (ANN), and that based on spatial drought vulnerability index in Karnataka state of India. At the local level, A study were developed by Bouguerra et al, 2023 in order to study the meteorological drought variability over northern Algeria using advanced atmospheric circulation patterns as variables that affect drought with SPI index.

There is a similar study made by chinese researchers on the Huai River basin, China to investigate meteorological drought in summer time by taking on considera- tion just atmospheric circulation patterns and SPEI index to build a self organizing map neural network (SOM) YIXING YIN et JIAO, 2024.

Conclusion

In this chapter, we have provided a thorough research on drought, encompassing its definitions, classifications, types, causes, and consequences. The diverse definitions of drought underscore the complexity of this phenomenon and the necessity of a multi-dimensional approach to its study. By classifying these definitions, we established a clearer framework for understanding the different ways drought can be conceptualized and measured.

The examination of drought types—meteorological, agricultural, hydrological, and socio-economic—revealed the various manifestations and impacts of drought across different sectors and scales. The causes of drought, ranging from natural climatic variability to human-induced changes, highlighted the multifactorial nature of drought conditions. The discussion on the consequences of drought underscored its profound effects on environmental, economic, and social systems, reinforcing the importance of effective drought prediction and management strategies.

The review of recent works in the field demonstrated the growing importance of using atmospheric circulation indices and drought indices in drought studies. These indices provide valuable insights into the drivers and patterns of drought, enhancing the accuracy and reliability of predictive models. Our study builds on this body of research, leveraging machine learning techniques to integrate atmospheric circulation indices with the SPEI drought index, aiming to improve drought prediction in the northern lands of Algeria.

This chapter establishes a solid foundation for our research, situating it within the broader context of drought studies and highlighting the significance of our methodological approach. By integrating biographical insights with cutting-edge research, we aim to contribute to the ongoing efforts to better understand and mitigate the impacts of drought.

Chapter II

STUDY AREA

Introduction

This chapter provides a comprehensive overview of the study area, focusing on the northern lands of Algeria, which are particularly susceptible to drought conditions. Understanding the geographical, geological, and climatic characteristics of this region is crucial for accurately modeling and predicting drought events. To this end, various maps have been utilized to present a detailed description of the study area, including the elevation map, slope map, geology map, rainfall map, and land use map. Each of these maps offers insights into different aspects of the region, from its topographical features to its climatic patterns and land utilization, which collectively influence drought occurrences.

The elevation map delineates the varying altitudes across the region, while the slope map highlights the gradient and steepness of the terrain. The geology map provides information on the underlying geological formations, which can affect soil moisture retention and runoff patterns. The rainfall map illustrates the spatial distribution of precipitation, a key factor in drought modeling. Finally, the land use map categorizes the area based on human activities and natural vegetation, indicating the proportion of land dedicated to agriculture, forests, urban areas, and other uses. Each map is discussed in detail, with particular emphasis on the proportions of different categories within the study area, providing a holistic understanding of the region's physical and climatic context.

Content

Geographical location and delimitation

Our study area concerns the east northern region of Algeria, with a surface of 149808 Km^2 bordered with the miditeranian sea from the north, the Algerian desert from the south, the Algerian western lands from the west and Tunisian and Libian borders from the east. It includes (11) different watersheds

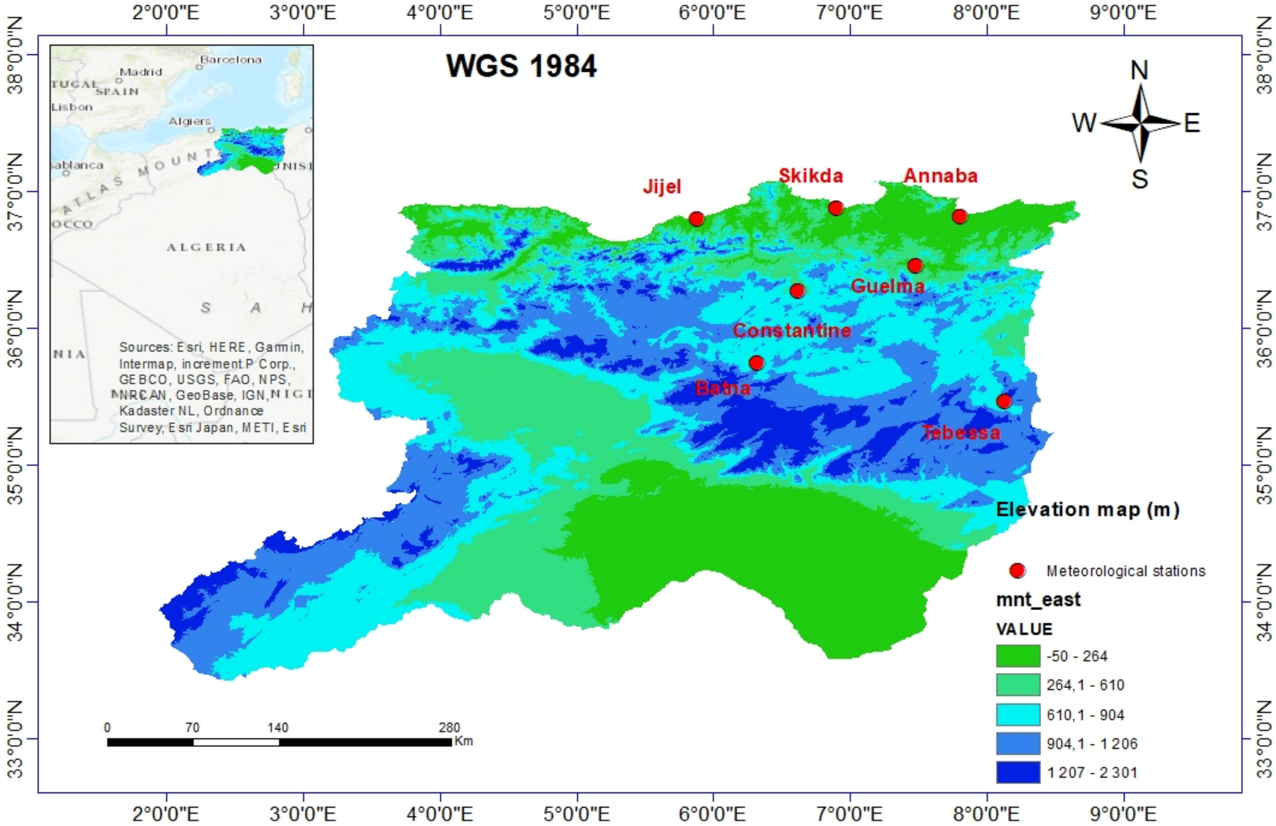

FIGURE II.1 — Elevation map

Stream network map

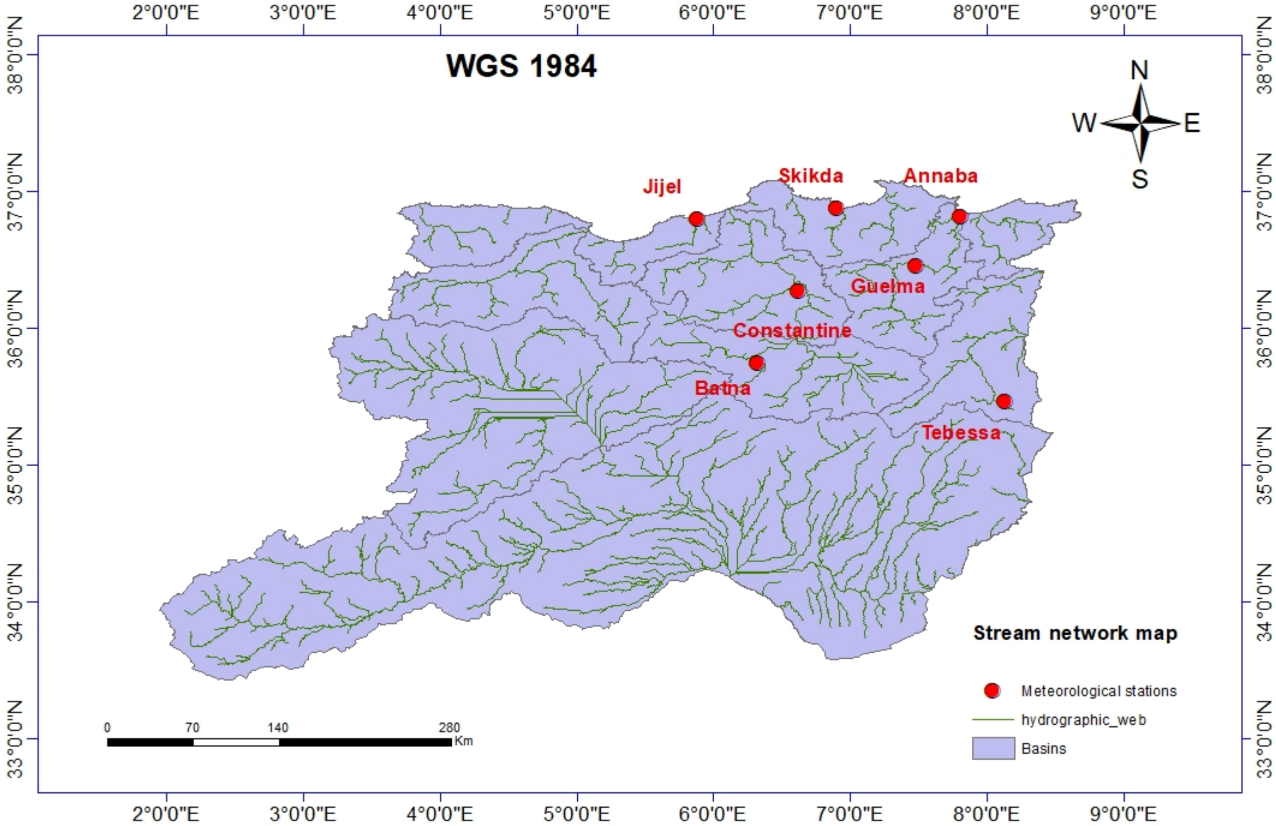

FIGURE II.2 — Stream network map

Slope map

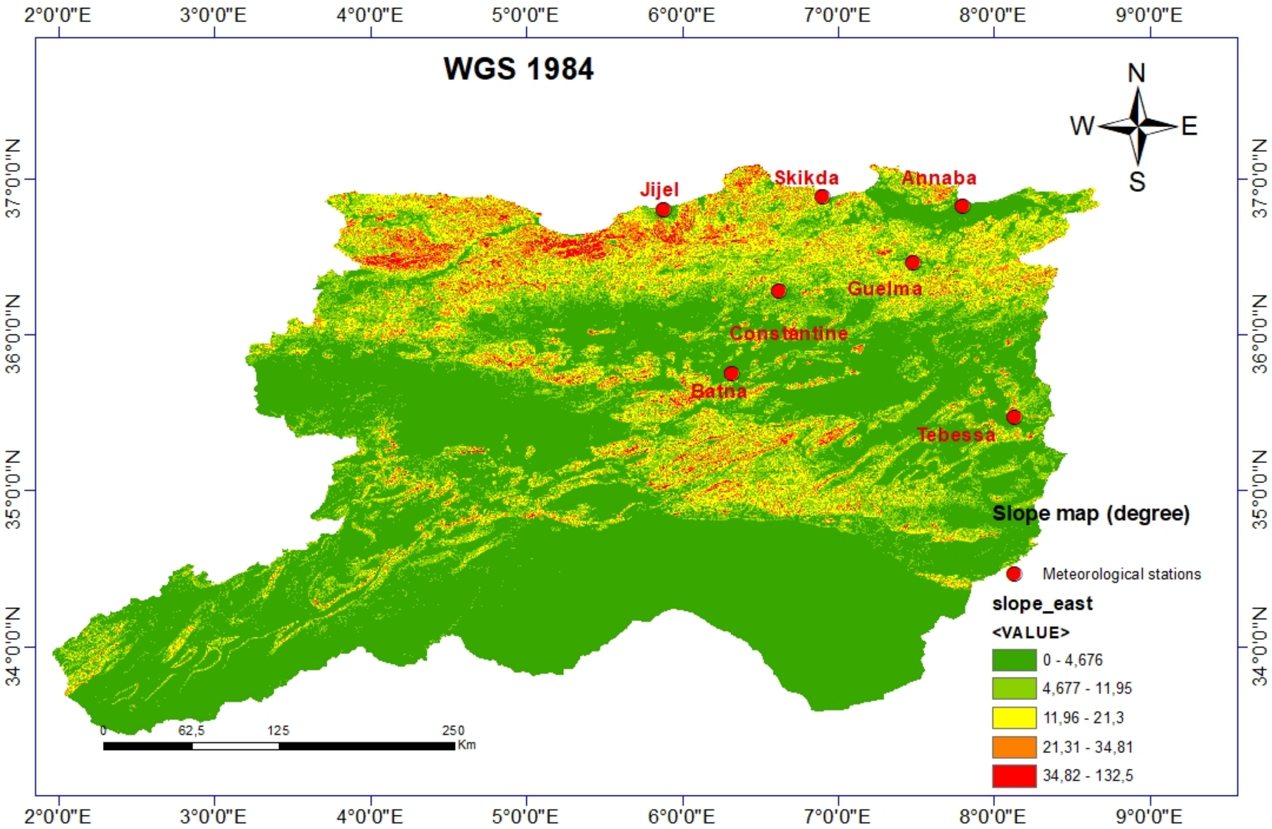

Figure II.3 — Slope map

this map represent the gradient of slope in our study area . We notice that in the south regions from our area (Ouled djellal, Biskra, Khenchla ...) the slope belongs to category 1 which means that we have almost flat surfaces . and the most intense slope we found them in the north beside the coastal lands where we have mountains and high hills most of them are in the region of Bejaia.

Geology map

the following map represent the geological structure of the present study area.

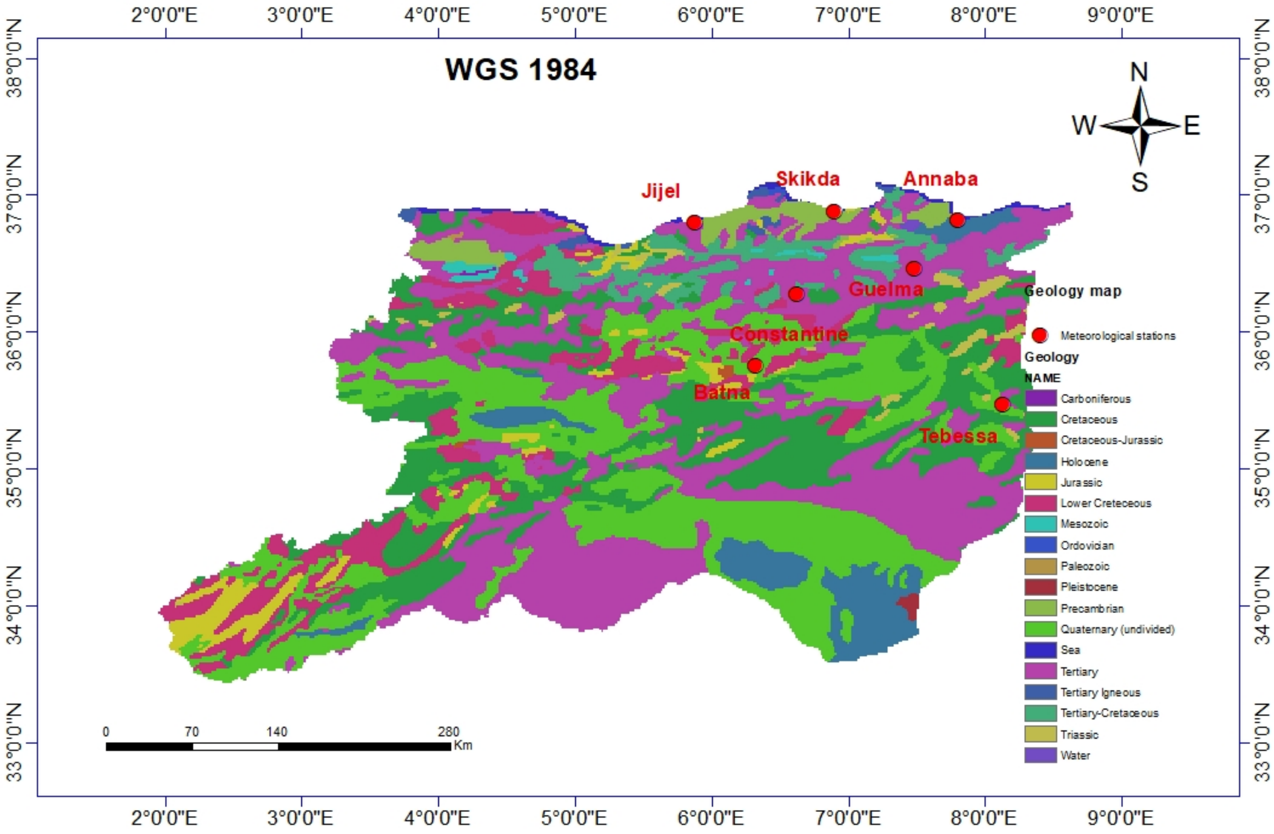

FIGURE II.4 — Geology map

Rainfall map

the next map represent the average annual precepitation in our region for years of our rainfall data serie the maximum value is 741 mm and the minimum is 397 mm. We can notice that the precepetation is high in the north , the highest values are in Bejaia city Jijel city and Skikda city. the more we go down the more we have a decrease in the amount of precepetation. the lowest values are in Oum El Bouaghi city and Batna city.

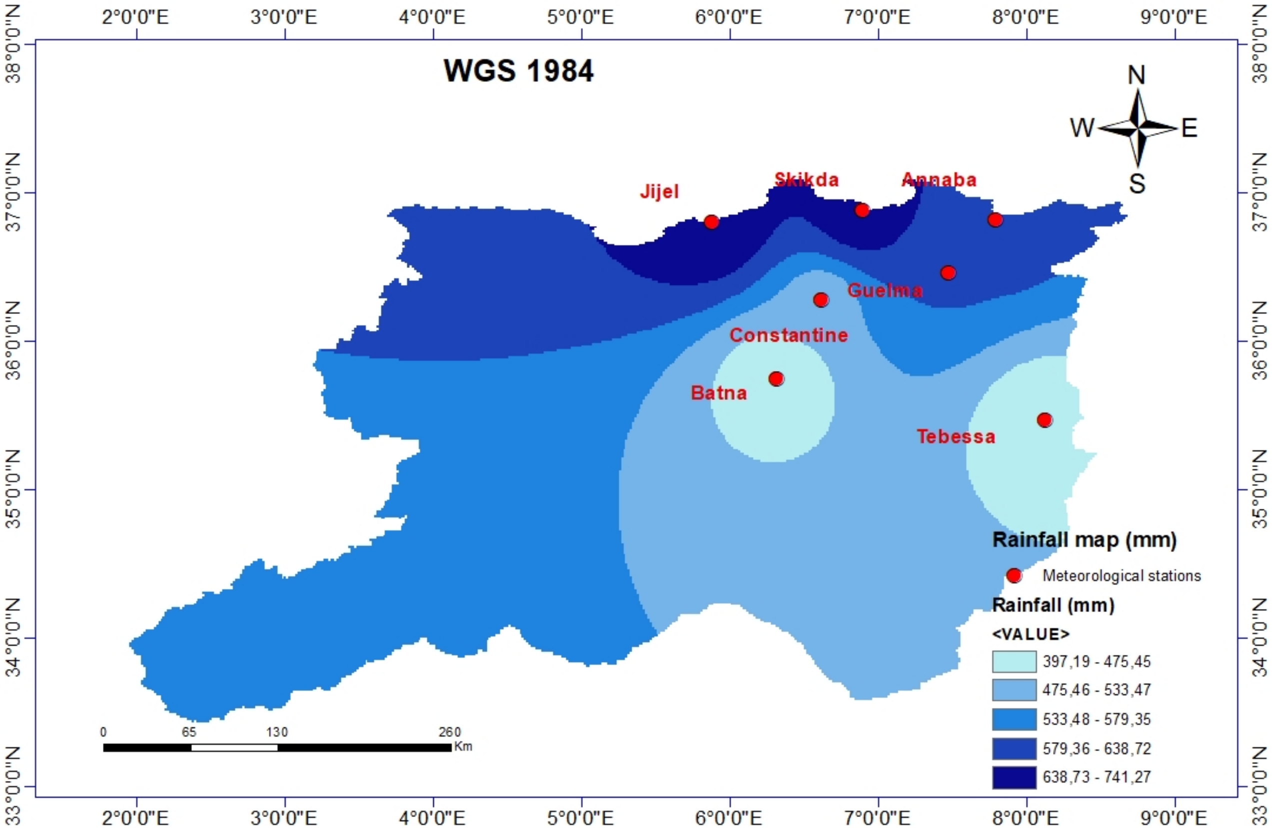

Figure II.5 — Rainfall map

Landuse map

the following map has been extracted from the data provided by the satellite Sentinel 2 ,10 meter resolution.

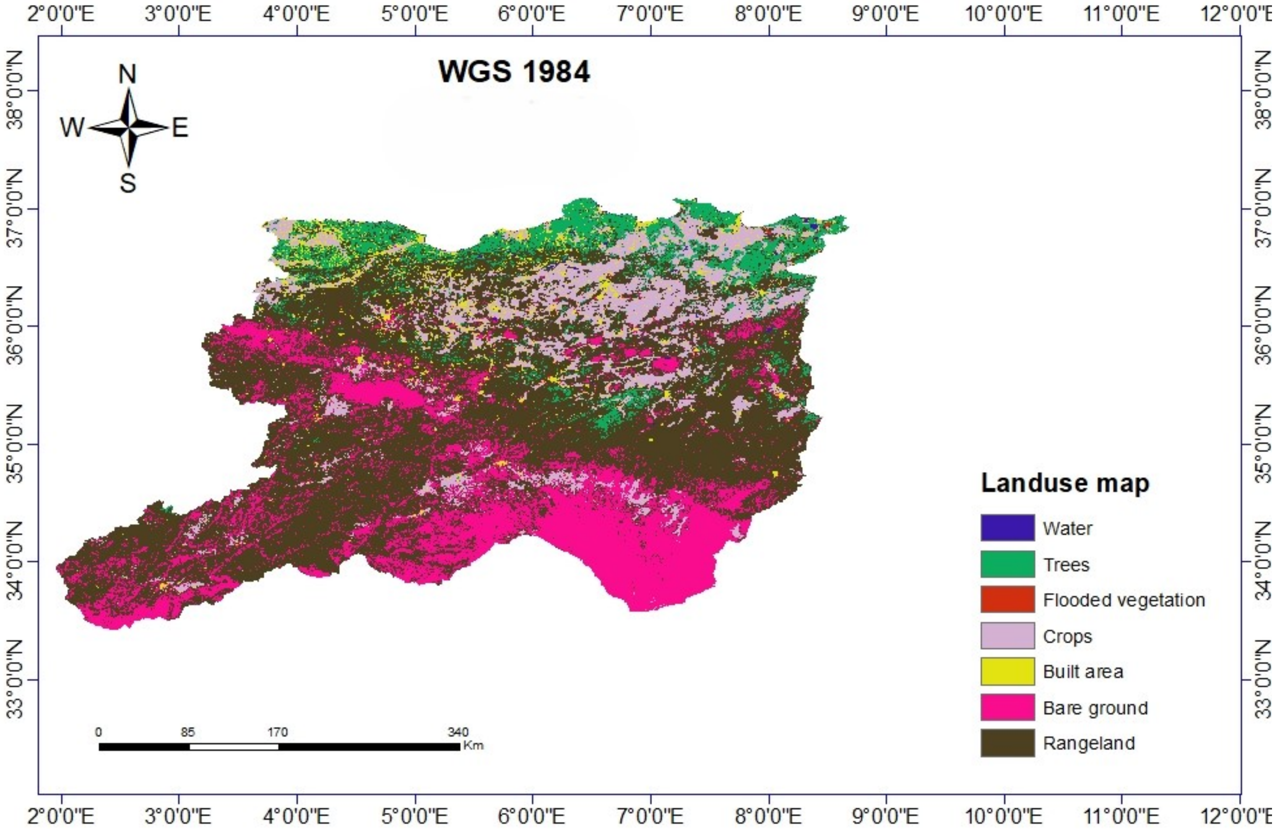

Figure II.6 — Landuse map

— Water : Areas where water was predominantly present throughout the year ; may not cover areas with sporadic or ephemeral water ; contains little to no sparse vegetation, no rock outcrop nor built up features like docks ; examples : rivers, ponds, lakes, oceans, flooded salt plains.

— Trees : Any significant clustering of tall (15 feet or higher) dense vegetation, typically with a closed or dense canopy ; examples : wooded vegetation, clusters of dense tall vegetation within savannas, plantations, swamp or mangroves (dense/tall vegetation with ephemeral water or canopy too thick to detect water underneath).

— Trees : Any significant clustering of tall (15 feet or higher) dense vegetation, typically with a closed or dense canopy ; examples : wooded vegetation, clusters of dense tall vegetation within savannas, plantations, swamp or mangroves (dense/tall vegetation with ephemeral water or canopy too thick to detect water underneath).

— Crops : Human planted/plotted cereals, grasses, and crops not at tree height ; examples : corn, wheat, soy, fallow plots of structured land.

— Built Area : Human made structures ; major road and rail networks ; large homogenous impervious surfaces including parking structures, office buildings and residential housing ; examples : houses, dense villages / towns / cities, paved roads, asphalt.

— Bare ground : Areas of rock or soil with very sparse to no vegetation for the entire year ; large areas of sand and deserts with no to little vegetation ; examples : exposed rock or soil, desert and sand dunes, dry salt flats/pans, dried lake beds, mines.

— Rangeland : Open areas covered in homogenous grasses with little to no taller vegetation ; wild cereals and grasses with no obvious human plotting (i.e., not a plotted field) ; examples : natural meadows and fields with sparse to no tree cover, open savanna with few to no trees, parks/golf courses/lawns, pastures.

the following table represent the proportion and surface of each type of structure :

Table II.1 — Landuse map proportions

Categorie	Proportions	Area (Km^2)
Water	0.2	280.28
Forest land	7.3	10285.15
Crops land	12.5	17596.17
Built area	3.2	4457.58
Bare ground	22.6	31717.04
Range land	54.2	76033.79

Conclusion

In this chapter, we have presented a detailed overview of the study area in the northern lands of Algeria through the use of various maps, each contributing to a comprehensive understanding of the region's characteristics. The elevation map revealed the altitudinal variations, which are crucial for understanding climate and vegetation patterns. The slope map highlighted the terrain's steepness, influencing water runoff and erosion processes. The geology map provided insights into the region's geological framework, essential for understanding soil properties and water retention capacities.

The rainfall map showcased the spatial distribution of precipitation, underscoring areas prone to higher or lower rainfall, which directly impacts drought conditions. Lastly, the land use map detailed the human and natural utilization of the land, indicating the proportions dedicated to agriculture, forests, urbanization, and other uses, which are critical for assessing the region's vulnerability to drought.

By integrating these diverse geographical, geological, and climatic perspectives, this chapter establishes a foundational understanding of the study area, setting the stage for the subsequent analysis of drought patterns and the application of machine learning models. The detailed mapping and categorization provide essential context for interpreting the study's findings and for developing targeted drought mitigation strategies tailored to the region's unique characteristics.

Chapter III

METHODOLOGY

Introduction

This chapter presents the methodological framework utilized in predicting drought events in the northern lands of Algeria through the application of machine learning techniques. The study leverages the Random Forest algorithm, integrating atmospheric circulation indices as predictive features and the Standardized Precipitation Evapotranspiration Index (SPEI) as the target variable. The methodology is meticulously designed to encompass data collection, preprocessing, the computation of the SPEI, an examination of the atmospheric circulation indices, and the implementation of the Random Forest model.

We begin by detailing the data sources, including meteorological and climatic datasets, followed by general statistics to provide an overview of the data landscape. The process of calculating the SPEI is then outlined, highlighting its importance as a comprehensive measure of drought conditions. Subsequently, a general overview of the atmospheric circulation indices is provided, emphasizing their relevance in influencing regional climatic patterns. The chapter culminates with an overview of machine learning models, focusing particularly on the Random Forest algorithm, and elucidating the rationale behind its selection for this study.

III.1 Meteorological Data

III.1.1 Data source

For this study, the primary data sources include historical records of rainfall and temperature, which are crucial for calculating the Standardized Precipitation Evapotranspiration Index (SPEI). Rainfall and temperature data were obtained from the National Meteorological Office (ONM)of Algeria, providing high-resolution, long-term records essential for accurate drought assessment. To address any missing data, supplementary information was sourced from the CRU (Climatic Research Unit) dataset [29], the National Centers for Environmental Information (NCEI) [30], and Tutiempo.net [31]. These additional sources allowed us to fill gaps using imputation techniques such as linear interpolation for short gaps and regression-based methods for longer gaps, ensuring robust and continuous datasets. This comprehensive approach to data collection and preprocessing ensures the integrity of the analysis and the accuracy of the resulting SPEI calculations and machine learning models.

We have 7 meteorological stations that we worked on their meteorological data :

— Annaba

— Batna

— Jijel

— Guelma

— Constantine

— Skikda

— Tebessa

III.1.2 Statistical analysis

III.1.2.1 Rainfall

the following table shows a brief statistical analysis for annual precipitation data serie for different stations :

TABLE III.1 — Rainfall data statistics

	Annaba	Batna	Jijel	Tebessq	Guelma	Constantine	Tebessa
Count	46	46	46	46	46	46	46
Mean	629.27	440.68	641.1	411.59	592.86	517.56	730.65
Std	155.07	97.4	171.8	108.05	138.49	143.83	154.32
Min	275.7	219.5	347.5	217	364.4	257.9	413.9
25%	521.85	369.85	518.95	339.1	503.48	419.6	645.58
50%	613.2	444.45	607.55	407.15	574.7	499.95	761.95
75%	729.13	510.3	736	456.05	675.28	586.93	818.93
Max	1126.6	652.7	1091.6	680.5	982.5	862.4	1148.9

the following chart represent the variation of annual precipitation data series from 1960 to 2005 :

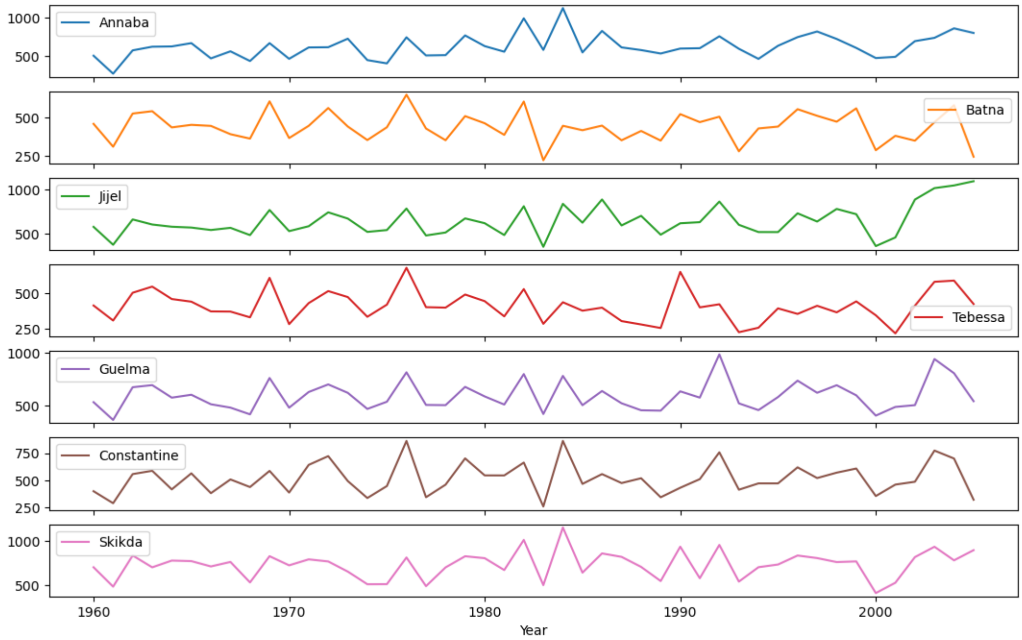

FIGURE III.1 — Variation of annual precipitation from 1960 to 2005

III.1.2.2 Temperature

the following table shows a brief statistical analysis for average inter-annual Temperature data serie for different stations :

TABLE III.2 — Temperature data statistics

	Annaba	Batna	Constantine	Jijel	Skikda	Guelma	Tebessa
Count	12	12	12	12	12	12	12
Mean	17.45	14.09	15.26	16.07	15.5	15.57	14.72
Std	5.18	7.43	6.9	5.79	5.93	6.56	7.4
Min	11.09	4.78	6.86	9.04	8.33	7.59	5.31
25%	12.75	7.99	9.31	10.86	10.21	9.8	8.5
50%	16.8	13.28	14.43	15.06	14.58	14.76	14
75%	21.59	20.57	21.48	21.24	20.76	21.61	21.23
Max	25.28	24.98	25.42	24.66	24.4	25.14	25.42

and these are histograms of those data series also from 1960 to 2005 of 12 months of the year :

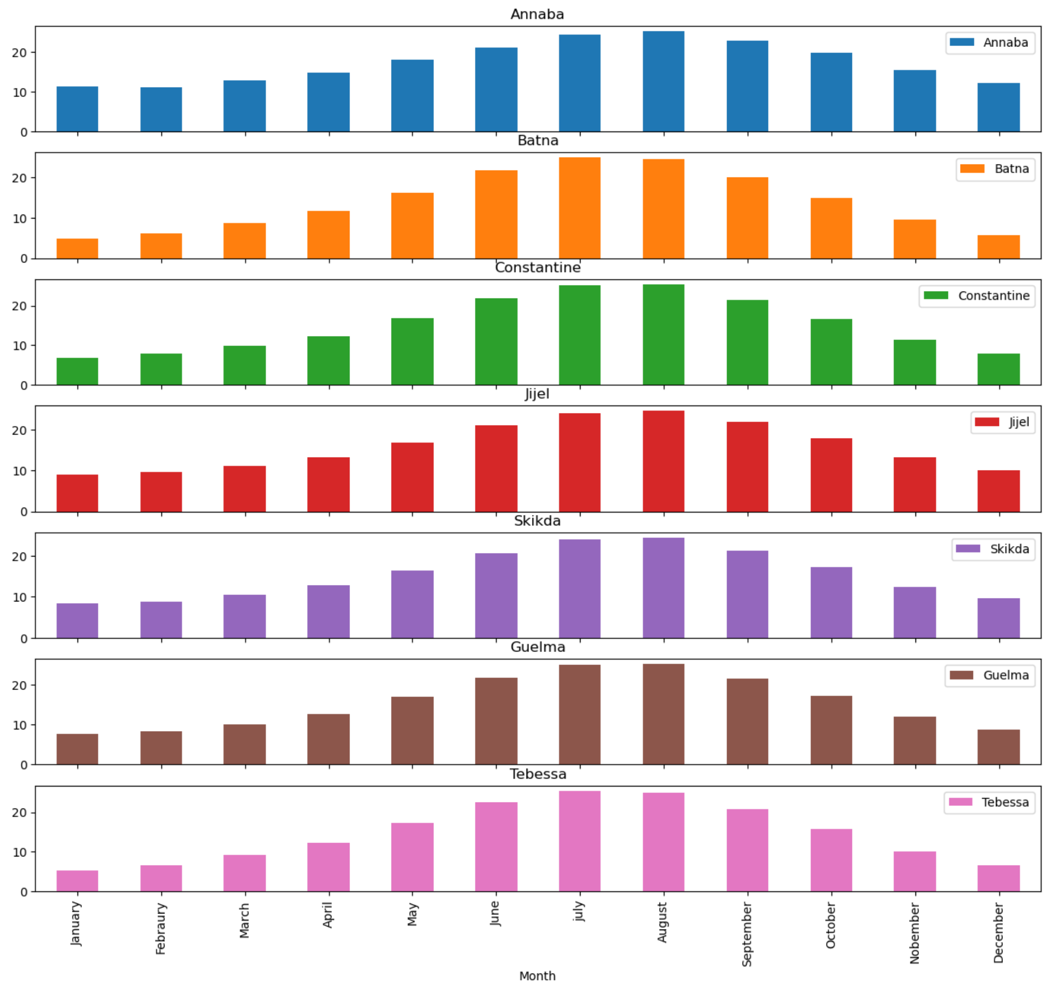

FIGURE III.2 — Average inter-annual temperature from 1960 to 2005

III.1.2.3 Atmospheric circulation indices

the following table shows a brief statistical analysis for Atmospheric circulation indices data series for different stations :

TABLE III.3 — Atmospheric circulation indices statistical analysis table

	NAO	SOI	TPI	WeMOi	NCP	WeI	EMP	MOI1	MOI2
count	552,00	552,00	552,00	552,00	552,00	552,00	552,00	552,00	552,00
mean	0,02	-0,21	0,12	0,24	0,02	-0,01	-182,92	-0,01	0,00
std	1,73	1,11	1,48	1,08	0,93	0,38	74,32	0,47	0,47
min	-4,70	-3,46	-4,38	-3,21	-2,52	-1,26	-395,30	-1,39	-1,64
25%	-1,13	-0,88	-0,85	-0,48	-0,57	-0,25	-230,88	-0,32	-0,31
50%	0,04	-0,16	0,30	0,27	0,03	0,01	-183,47	0,00	0,02
75%	1,20	0,56	1,09	0,95	0,65	0,25	-130,98	0,34	0,36
max	5,26	2,85	4,20	3,96	2,42	1,08	76,42	1,24	1,04

the following chart represent the variation of Atmospheric circulation indices data series from 1960 to 2005 :

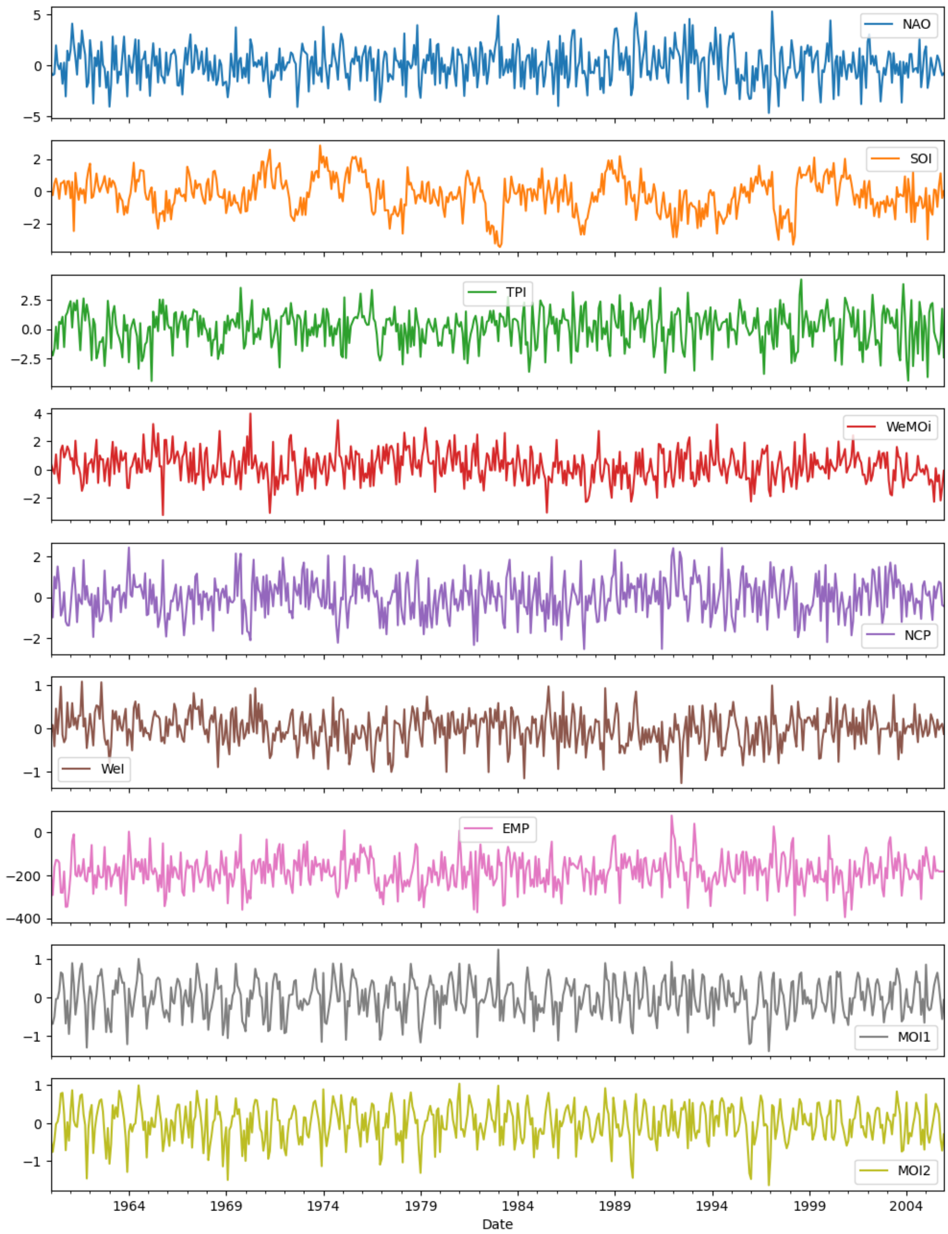

Figure III.3 — Variation of atmospheric circulation indices

III.2 Drought index

Generally speaking, the drought index is used to measure the frequency of drought. An assortment of drought indexes have been proposed by various scholars. Among these indices are the SPEI, the standardized precipitation index (SPI), and the Palmer drought severity index (PDSI). It is widely used to analyze how vegetation responds to drought because, among the many drought indices, the SPEI not only takes precipitation into account but also combines the sensitivity of the PDSI to realistic evapotranspiration variations and the benefit of the SPI at multiple time scales [32]. Vicente-Serrano et al developed the SPEI index under the standard of A Multiscalar Drought Index Sensitive to Global Warming [33]. A climatic water balance, the buildup of deficit/surplus at various time scales, and adjustment to a log-logistic probability distribution are all steps in the intricate process of calculating the index. The standardized precipitation index (SPEI) and the SPI are comparable mathematically, but the SPEI takes temperature into account [33]. The SPI's primary detractors point out that its computation solely uses precipitation data. The indicator does not take into account additional factors like temperature, evapotranspiration, wind speed, and soil water retention capacity that can affect droughts[33]. The SPI and other precipitation-based drought indices are predicated on two premises : 1) that precipitation is significantly more variable than other variables, such temperature and potential evapotranspiration (PET), and 2) that other variables are stationary, meaning they do not exhibit a temporal pattern. In this case, the temporal variability in precipitation controls droughts, and the significance of these other variables is minimal[33]. Therefore, it is better to employ drought indices (such the PDSI) that incorporate temperature data into their derivation, particularly for applications incorporating future climate scenarios[33].

Calculation

Precipitation on a monthly (or weekly) basis is used as input data to compute the SPI. The monthly (or weekly) difference between precipitation and PET is used by the SPEI. This illustrates a straightforward climatic water balance (Thornthwaite 1948), from which the SPEI is derived by calculating it at various time scales [33]. The initial phase, calculating the PET, is challenging because to the multiple factors involved, such as air humidity, soil incoming radiation, surface temperature, water vapor pressure, and sensible and latent heat fluxes between the ground and atmosphere [33]. In this study we used the hargreaves method for the calculation of potential evapotranspiration (PET)[34] :

$$ET_{HS} = k_{RS}.R_a.(Tmax - Tmin)^{HE}(\frac{Tmax + Tmin}{2} + HT)$$

— ETHS is daily PET in mm/day

— Ra is extraterrestrial radiation in mm/day

— Tmax and Tmin are daily maximum and minimum air temperature in °C

— kRS is the empirical radiation adjustment coefficient

— HE is empirical Hargreaves exponent

— HT is empirical temperature coefficient

With a value for PET, the difference between the precipitation P and PET for the month i is calculated using :

$$D_i = P_i - PET_i$$

which provides a simple measure of the water surplus or deficit for the analyzed month.

The calculated Di values are aggregated at different time scales, following the same procedure as that for the SPI. The difference in a given month j and year i depends on the chosen time scale k. For example, the accumulated difference for one month in a particular year i with a 12-month time scale is calculated using

if j<k :

$$X_{i,j}^{k} = \sum_{l=13-k+j}^{12} D_{i-1,l} - \sum_{l=1}^{j} D_{i,l}$$

if j>=k :

$$X_{i,j}^{k} = \sum_{l=j-k+1}^{j} D_{i,l}$$

We tested the most suitable distribution to model the Di values calculated at different time scales. For this purpose, L-moment ratio diagrams were used because they allow for comparison of the empirical frequency distribution of D series computed at different time scales with a number of theoretical distributions (Hosking 1990). The L moments are analogous to conventional central moments, but they are able to characterize a wider range of distribution functions and are more robust in relation to outliers in the data.

To create the L/moment ratio diagrams, L moment ratios (L skewness Tho3 and L kurtosis Tho4) must be calculated. Here Tho3 and Tho4 are calculated as follows :

$$\tau_3 = \frac{\lambda_3}{\lambda_2}$$

and :

$$\tau_4 = \frac{\lambda_4}{\lambda_2}$$

where Lambda2, Lambda3, and Lambda4 are the L moments of the D series, obtained from probability-weighted moments (PWMs) using the formulas

$$\lambda_1 = \omega_0$$

$$\lambda_2 = \omega_0 - 2\omega_1$$

$$\lambda_3 = \omega_0 - 6\omega_1 + 6\omega_2$$

and :

$$\lambda_4 = \omega_0 - 12\omega_1 + 30\omega_2 - 20\omega_3$$

The PWMs of order s are calculated as

$$\omega_s = \frac{1}{N}\sum_{i=1}^{N}(1-F_i)^s D_i$$

where Fi is a frequency estimator calculated following the approach of Hosking (1990) :

$$F_i = \frac{i-0.35}{N}$$

The probability density function of a three-parameter log-logistic distributed variable is expressed as

$$f(x) = \frac{\alpha}{\beta}(\frac{x-\gamma}{\alpha})(1+(\frac{x-\gamma}{\alpha})^{\beta})^{-2}$$

where alpha, beta, and gamma are scale, shape, and origin parameters, respectively, for D values in the range (gamma > D < inf).

$$\beta = \frac{2\omega_1-\omega_0}{6\omega_1-\omega_0-6\omega_2}$$

$$\alpha = \frac{(\omega_0-2\omega_1)\beta}{\chi(1+\frac{1}{\beta})\chi(1-\frac{1}{\beta})}$$

and :

$$\gamma = \omega_0 - \alpha\chi(\frac{l+1}{\beta})\chi(\frac{l-1}{\beta})$$

The probability distribution function of the D series, according to the log-logistic distribution, is given by

$$F(x) = (1+(\frac{\alpha}{x-\gamma})^{\beta})^{-1}$$

With F(x) the SPEI can easily be obtained as the standardized values of F(x). For example, following the classical approximation of Abramowitz and Stegun (1965)

$$SPEI = W - \frac{C_0+C_1W+C_2W^2}{1+d_1W+d_2W^2+d_3W^3}$$

with

$$W = \sqrt{-2ln(P)}$$

for P $<=$ 0.5

III.3 Atmospheric circulation indices

Regional and global climate conditions are mostly determined by global climatic circulation indicators. These patterns of atmospheric circulation, which span a large geographic area and impact weather-related events, show significant variability [35]. Oceanic heat sources cause disturbances in the atmosphere, allowing waves to disperse and revert to their initial state. These waves affect the climate in far-off places, defining patterns of teleconnection. The Intergovernmental Panel on Climate Change (IPCC) states that large-scale, quasistationary atmospheric Rossby waves are the primary mechanism through which climate indices link neighboring regions. As a result, some regions experience higher temperatures or more precipitation than what is indicated by the dominant global-scale changes [35].

The North Atlantic Oscillation (NAO)

The NAO was identified by Sir Gilbert Walker in the 1920s as the differential in surface pressure between the tropics and northern regions of the NAO. Although the NAO is one of the most prevalent and prominent patterns of atmospheric variability, its temporal and geographical structure are not well understood [35]. The NAO (20° N80° N, 80° W30° E) is normally described as the difference between the station pressures on Iceland and the Azores. Monthly variations in the daily NAO index are consistent with NAO patterns [35].

The Southern Oscillation Index (SOI)

Based on the recorded differences in seal level pressure (SLP) between Tahiti and Darwin, Australia, the SOI was calculated. The air pressure that varies widely between the eastern and western tropical Pacifics is measured by the SOI. The positive phase of the SOI shows an above-normal pressure over Darwin coinciding with above-normal cold ocean water (La Niña), while negative values of the SOI suggest below-normal air pressure in Tahiti along with abnormally warm ocean waters over the eastern tropical Pacific typical of El Niño [35].

Trans Polar Index (TPI)

The TPI is defined as the normalized pressure difference between Hobart and Stanley. The index was first suggested by Pittock and has been updated and analysed further by Jones et al[36].

Mediterranean Oscillation Indices (MOI)

The MOI is defined by Palutikof et al. (1996) and Conte et al. (1989) as the normalized pressure difference between Algiers (36.4°N, 3.1°E) and Cairo (30.1°N, 31.4°E) A second version of the index can be calculated from Gibraltar's Northern Frontier (36.1°N, 5.3°W) and Lod Airport in Occupied Palestine (32.0°N, 34.5°E) (Palutikof et al., 2003). Here data is presented for both variants, using pressure interpolated (16 point Bessel) from NCEP/NCAR reanalysis data[37].

Western Mediterranean Oscillation (WeMOI)

Paris Londonn westerly index(WeI)

Cornes et al. (2012a) recovered, quality controlled and homogenized a >300 year daily series of sea-level pressure (SLP) for the city of London (1692-2007). Cornes et al. (2012b) digitized and corrected a number of records to create another very long daily series of SLP for Paris (1670-2007). Both series contain some data gaps, notably 1726-1747 for Paris. Cornes et al. (2012a,b) also assessed the homogeneity of the final series, and data users should refer carefully to these papers when using these data, to take residual concerns into account. Cornes et al. (2013) used monthly means of the Paris and London SLP series to construct a monthly Paris-London Westerly index from 1692 onwards. This can be used as an estimate of the North Atlantic Oscillation (NAO). See Cornes et al. (2013) for a description of how the Paris London index is calculated and for a comparison with NAO indices[38].

North Sea Caspian Pattern (NCP)

The NCP dataset presented here is calculated from the normalised 500hpa pressure difference between averages of North Sea (0°E, 55°N and 10°E, 55°N) and North Caspian (50°E, 45°N and 60°E, 45°N) centres of action. This is the same formulation used as Kutiel and Benaroch (2002) who selected these locations by use of linear correlation between pressure grid points and a GIS approach[39].

The Eastern Mediterranean teleconnection Pattern (EMP)

made to investigate possible teleconnection patterns of atmospheric circulation, centered over eastern Mediterranean (EM) with the aid of gridded NCEP/NCAR daily values of geopotential heights for the period 1958–2003[40].

III.4 Machine learning model selection

Definition

AI refers to the simulation of human intelligence processes by computer systems. These processes include learning (the acquisition of information and rules for using it), reasoning (using rules to reach approximate or definite conclusions), problem-solving, perception (recognizing patterns in data), and language understanding. AI systems are designed to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and language translation [41].

ML is a subset of AI that focuses on the development of algorithms and statistical models that allow computer systems to learn from and improve based on data without being explicitly programmed. ML algorithms use patterns and inference to make predictions or decisions. The key idea behind ML is to enable computers to learn from data and adapt their behavior or performance over time, often achieving tasks with accuracy and efficiency that surpass human capabilities [41].

Random forest

The random forest, developed by Breiman, uses randomization to create a large number of decision trees. The output of these trees is aggregated into a single output using voting for classification problems or averaging for regression problems. Randomization is implemented in 2 ways. First, the data set is sampled with replacement (bootstrap sampling). The process of aggregating a new sample this way is called “bootstrap aggregation” or “bagging” [42].

In our study we choose the random forest as our selected machine learning model

Data normalization and division

we used the atmospheric circulatin indices as features and values of SPEI calculated as label, this has been done in all meteorological stations within our region Prior to model training, the dataset is normalized to eliminate discrepancies in scale among features.

At the same time, it accelerates the convergence speed of the model and can help avoid overfitting to some extent. The normalization function is :

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where xmax and xmin represent the maximum and minimum values of the sample data, respectively.

The data are randomly divided into 80 percent,and 20 percent for training, and test purposes, respectively. The training set is used for model parameter optimization, the test set for the final assessment of model performance [43].

Conclusion

In this chapter, we have systematically outlined the methodological approach employed to predict drought events in the northern lands of Algeria. The data sources and their corresponding general statistics were thoroughly discussed, providing a foundational understanding of the datasets utilized in this study. The computation of the SPEI was detailed, underscoring its role as a critical measure of drought conditions by accounting for both precipitation and potential evapotranspiration.

An overview of the atmospheric circulation indices was presented, highlighting their significance as influential predictors in the model. The methodological chapter also provided a comprehensive overview of machine learning models, with a particular focus on the Random Forest algorithm. The choice of the Random Forest model was justified based on its robustness, ability to handle large datasets, and proficiency in capturing complex, non-linear relationships.

By integrating these diverse methodological components, this chapter sets the stage for the subsequent analysis and results, ensuring a rigorous and systematic approach to drought prediction. The methodologies discussed herein provide a solid framework for leveraging machine learning techniques to enhance the predictive capabilities and understanding of drought events in the region.

Chapter IV

RESULTS AND DISCUSSION

Introduction

This chapter presents the findings of our study on drought prediction in the northern lands of Algeria, detailing the process of calculating the Standardized Precipitation Evapotranspiration Index (SPEI), analyzing the correlation between atmospheric circulation indices, and evaluating the performance of the Random Forest model. The chapter begins by outlining the methodology for calculating the SPEI, including the software used and the step-by-step procedure to derive the final results. Subsequently, it explores the correlations between various atmospheric circulation indices, providing insights into their interrelationships and their influence on drought conditions.

The chapter then evaluates the performance of the Random Forest model using several metrics, including accuracy score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC) score, and the feature importance of each atmospheric index across different meteorological stations. The results are interpreted and discussed in detail, highlighting the model's predictive capabilities and the significance of various atmospheric indices in forecasting drought events.

IV.1 Drought index

using the data we mentioned at the beginning of the previous chapter rainfall precepetation, minimum and maximum temperature values on every station, for every month , we were able to calculate the SPEI drought index for 1 month, 3 months and 6 months. our calculation has been done with the help of R studio program and SPEI package embedded inside that program.

IV.1.1 SPEI

the following table shows a brief statistical analysis for SPEI data serie for different stations :

Table IV.1 — SPEI statistical analysis table

	Annaba	**Batna**	**Constantine**	**Guelma**	**Jijel**	**Skikda**	**Tebessa**
count	552,000	552,000	552,000	552,000	552,000	552,000	552,000
mean	0,029	0,276	0,175	0,155	-0,060	-0,048	0,155
std	0,994	0,884	0,907	0,941	0,908	0,968	0,928
min	-3,768	-2,089	-2,162	-1,996	-3,411	-2,249	-2,056
25%	-0,682	-0,348	-0,479	-0,523	-0,756	-0,863	-0,562
50%	-0,020	0,258	0,144	0,137	-0,090	-0,071	0,097
75%	0,803	0,961	0,802	0,878	0,642	0,695	0,830
max	3,266	2,426	2,479	2,452	2,250	2,138	2,386

— SPEI values of Annaba station varies between -3.76 and 3.26 with mean value equal to 0.028

— SPEI values of Batna station varies between -2.08 and 2.42 with mean value equal to 0.27

— SPEI values of Jijel station varies between -3.41 and 2.25 with mean value equal to -0.06

— SPEI values of Guelma station varies between -1.99 and 2.45 with mean value equal to 0.15

— SPEI values of Constantine station varies between -2.16 and 2.47 with mean value equal to 0.17

— SPEI values of Skikda station varies between -2.24 and 2.13 with mean value equal to -0.04

— SPEI values of Tebessa station varies between -2.05 and 2.38 with mean value equal to 0.15

the following chart represent the variation of SPEI data serie from 1960 to 2005 :

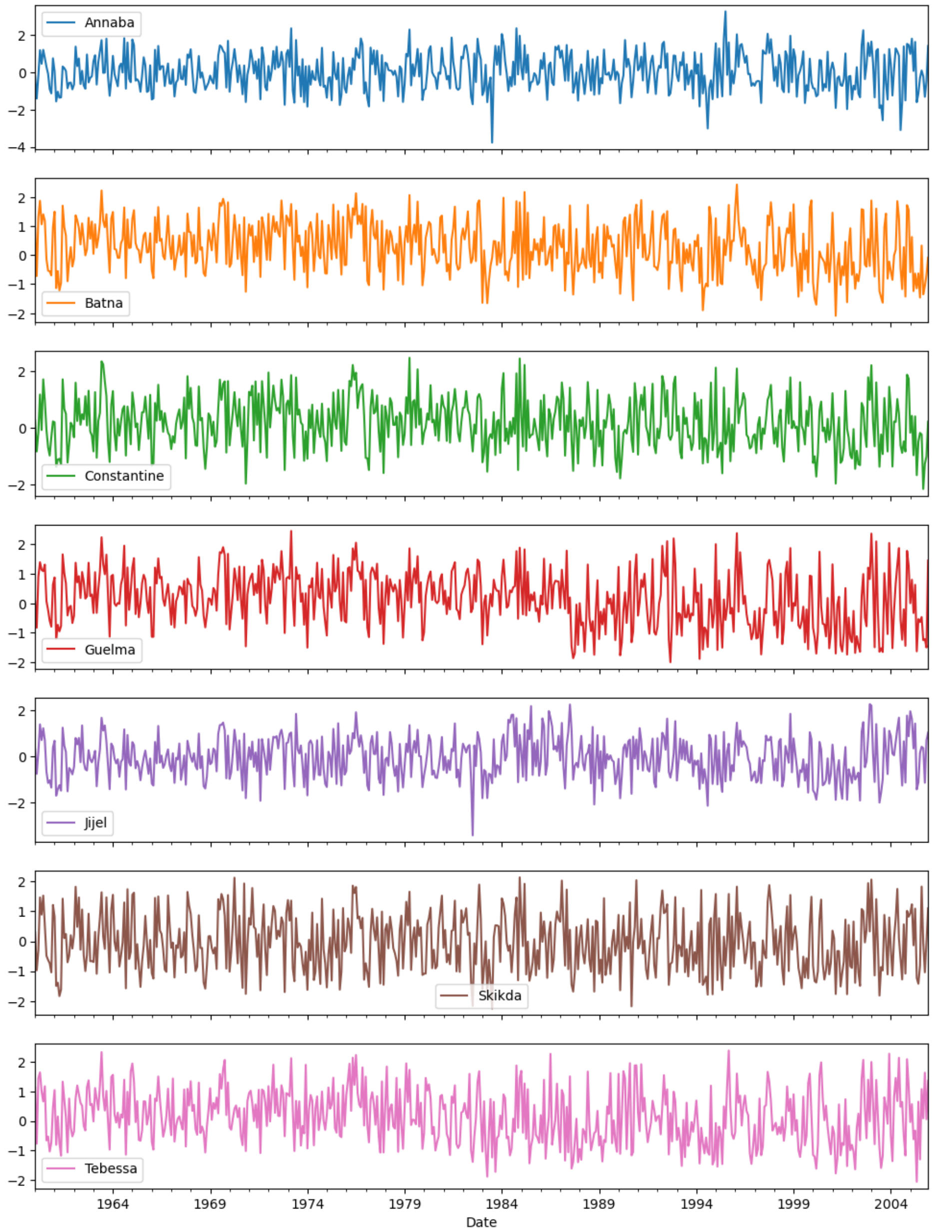

Figure IV.1 — Variation of SPEI data serie from 1960 to 2005

IV.1.2 SPEI3

the following table shows a brief statistical analysis for SPEI of 3 months data serie for different stations :

Table IV.2 — SPEI 3 statistical analysis table

	Annaba	**Batna**	**Constantine**	**Guelma**	**Jijel**	**Skikda**	**Tebessa**
count	550,000	550,000	550,000	550,000	550,000	550,000	550,000
mean	0,039	0,355	0,231	0,186	-0,155	-0,065	0,208
std	0,991	0,828	0,901	0,938	0,886	0,966	0,941
min	-2,887	-1,834	-2,080	-1,984	-2,500	-2,132	-2,211
25%	-0,720	-0,218	-0,396	-0,519	-0,818	-0,788	-0,507
50%	-0,013	0,334	0,205	0,248	-0,172	-0,179	0,236
75%	0,777	0,935	0,917	0,886	0,469	0,647	0,868
max	2,778	2,688	2,814	2,366	2,208	2,226	2,561

— SPEI of 3 months values of Annaba station varies between -3.76 and 3.26 with mean value equal to 0.028

— SPEI of 3 months values of Batna station varies between -2.08 and 2.42 with mean value equal to 0.27

— SPEI of 3 months values of Jijel station varies between -3.41 and 2.25 with mean value equal to -0.06

— SPEI of 3 months values of Guelma station varies between -1.99 and 2.45 with mean value equal to 0.15

— SPEI of 3 months values of Constantine station varies between -2.16 and 2.47 with mean value equal to 0.17

— SPEI of 3 months values of Skikda station varies between -2.24 and 2.13 with mean value equal to -0.04

— SPEI of 3 months values of Tebessa station varies between -2.05 and 2.38 with mean value equal to 0.15

the following chart represent the variation of SPEI of 3 months data serie from 1960 to 2005 :

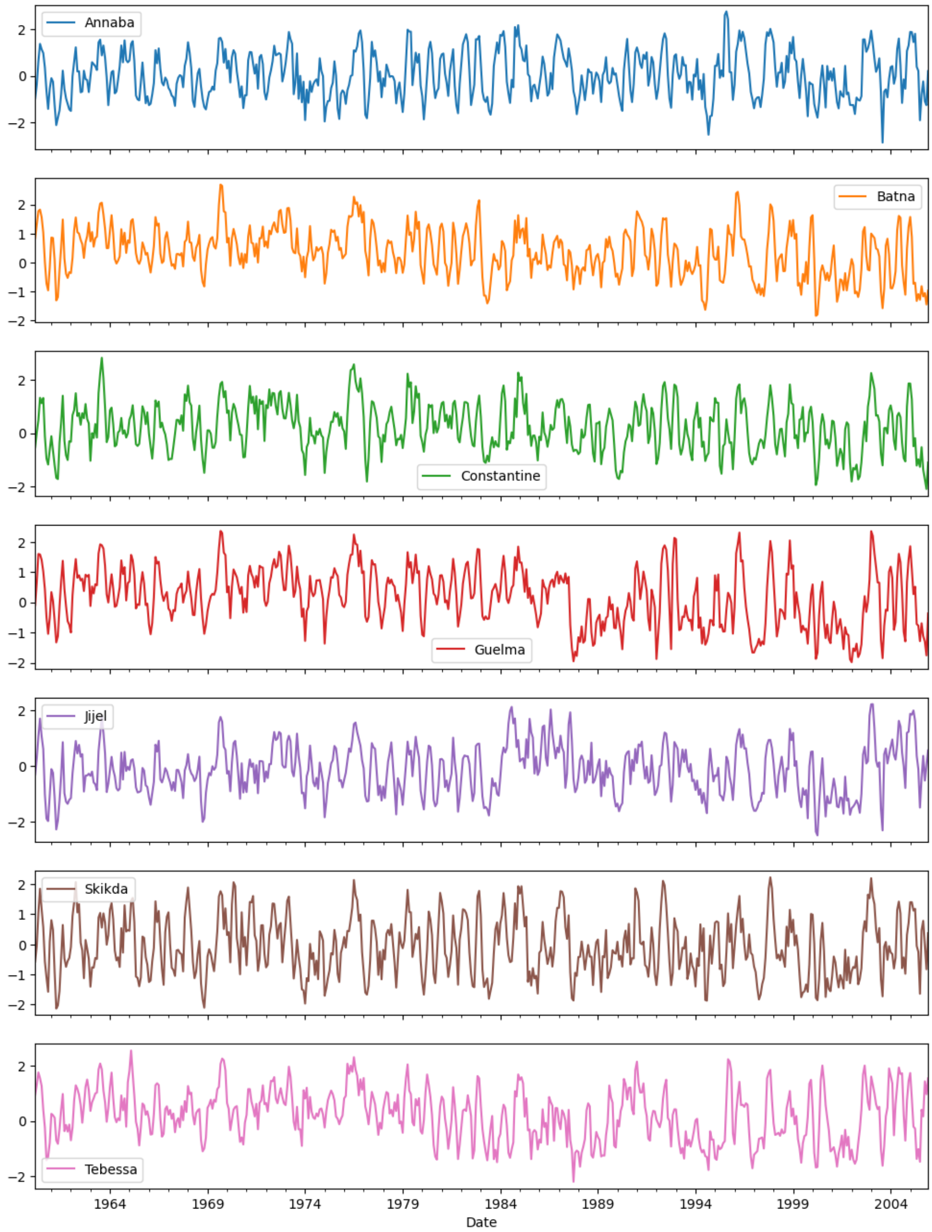

Figure IV.2 — Variation of SPEI 3 data serie from 1960 to 2005

IV.1.3 SPEI6

the following table shows a brief statistical analysis for SPEI of 6 months data serie for different stations :

Table IV.3 — SPEI 6 statistical analysis table

	Annaba	**Batna**	**Constantine**	**Guelma**	**Jijel**	**Skikda**	**Tebessa**
count	547,000	547,000	547,000	547,000	547,000	547,000	547,000
mean	0,036	0,414	0,289	0,228	-0,247	-0,063	0,255
std	1,003	0,772	0,879	0,943	0,860	0,987	0,922
min	-2,186	-1,451	-1,841	-2,144	-2,051	-2,389	-1,859
25%	-0,722	-0,117	-0,355	-0,483	-0,842	-0,800	-0,434
50%	-0,020	0,383	0,269	0,312	-0,284	-0,184	0,314
75%	0,822	0,924	0,933	0,901	0,364	0,723	0,938
max	2,392	2,760	2,831	2,753	2,073	2,138	2,392

— SPEI of 6 months values of Annaba station varies between -3.76 and 3.26 with mean value equal to 0.028

— SPEI of 6 months values of Batna station varies between -2.08 and 2.42 with mean value equal to 0.27

— SPEI of 6 months values of Jijel station varies between -3.41 and 2.25 with mean value equal to -0.06

— SPEI of 6 months values of Guelma station varies between -1.99 and 2.45 with mean value equal to 0.15

— SPEI of 6 months values of Constantine station varies between -2.16 and 2.47 with mean value equal to 0.17

— SPEI of 6 months values of Skikda station varies between -2.24 and 2.13 with mean value equal to -0.04

— SPEI of 6 months values of Tebessa station varies between -2.05 and 2.38 with mean value equal to 0.15

the following chart represent the variation of SPEI of 6 months data serie from 1960 to 2005 :

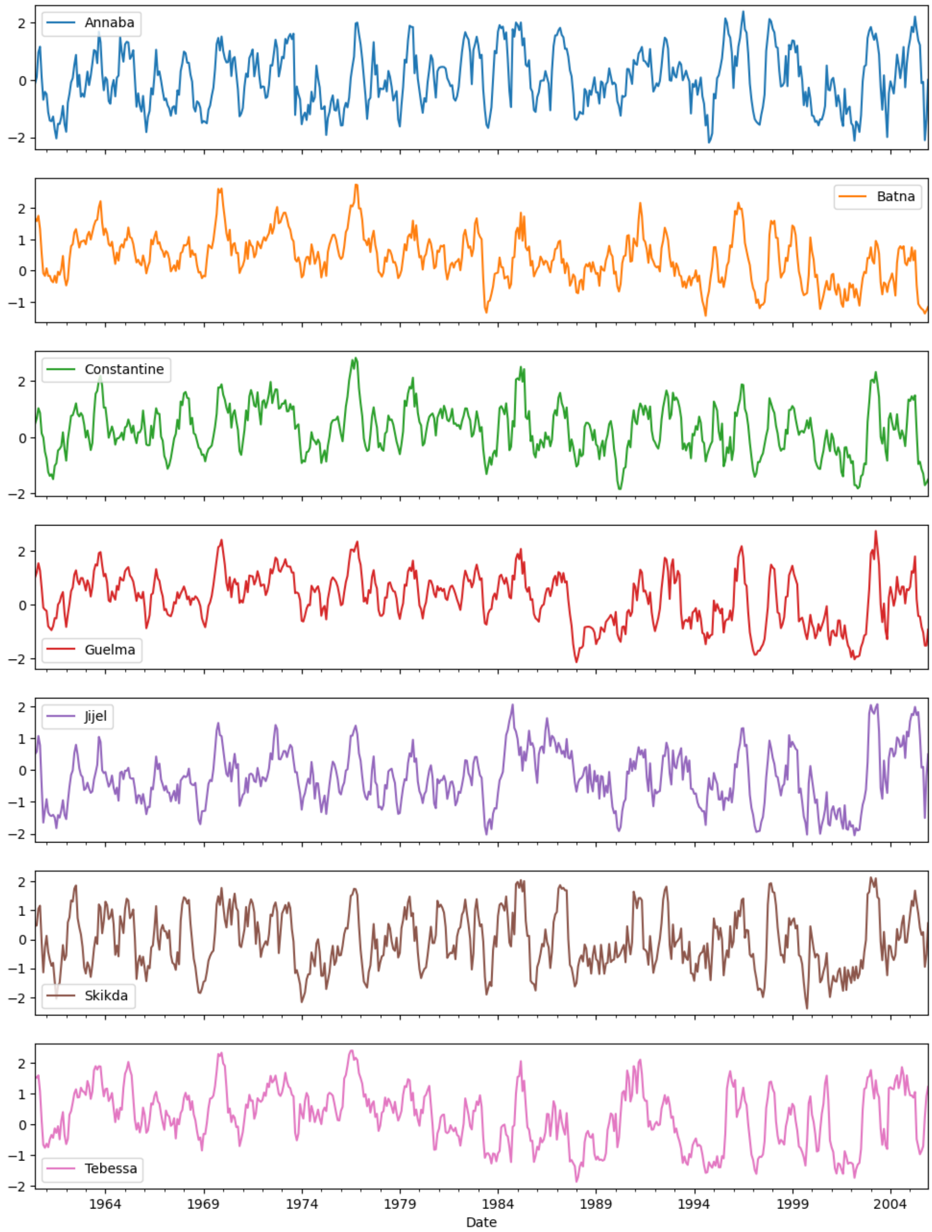

Figure IV.3 — Variation of SPEI 6 data serie from 1960 to 2005

IV.2 Atmospheric circulation indices

In this section, it is important to discuss the correlation of the Atmospheric circulation indices and how far the similarity of these indices to each other. We have derived the following correlation matrices from our model script to get an insight about this point :

Atmospheric circulation indices correlation matrix for 1 month

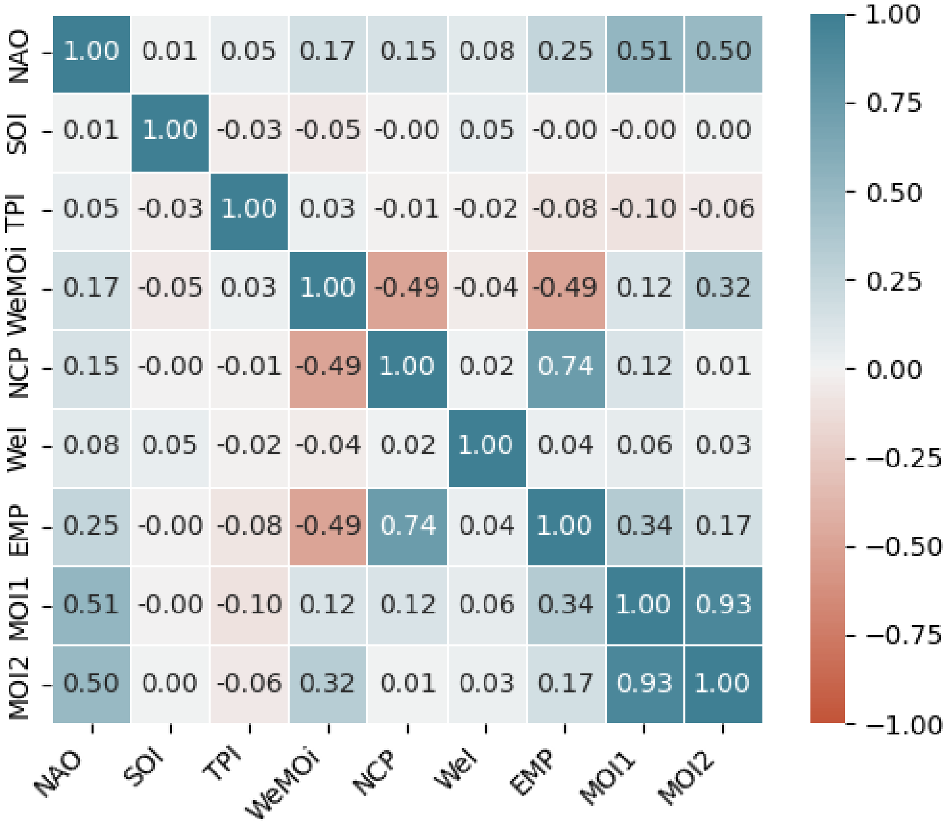

Figure IV.4 — Atmospheric circulation indices correlation matrix for 1 month

there are some remarques on the matrix above :

— high direct correlation with 74 percent between the NCP and EMP

— high correlation with 93 percent between the MOI1 and MOI2

— high correlation between NAO and both MOI1 and MOI2 with 50 percent

Although the correlation between MOI1 and MOI2 is too big, we can't delete one of them and leave the other due to the necessicity of the presence of both of them in the inputs of our model.

Atmospheric circulation indices correlation matrix for 3 month

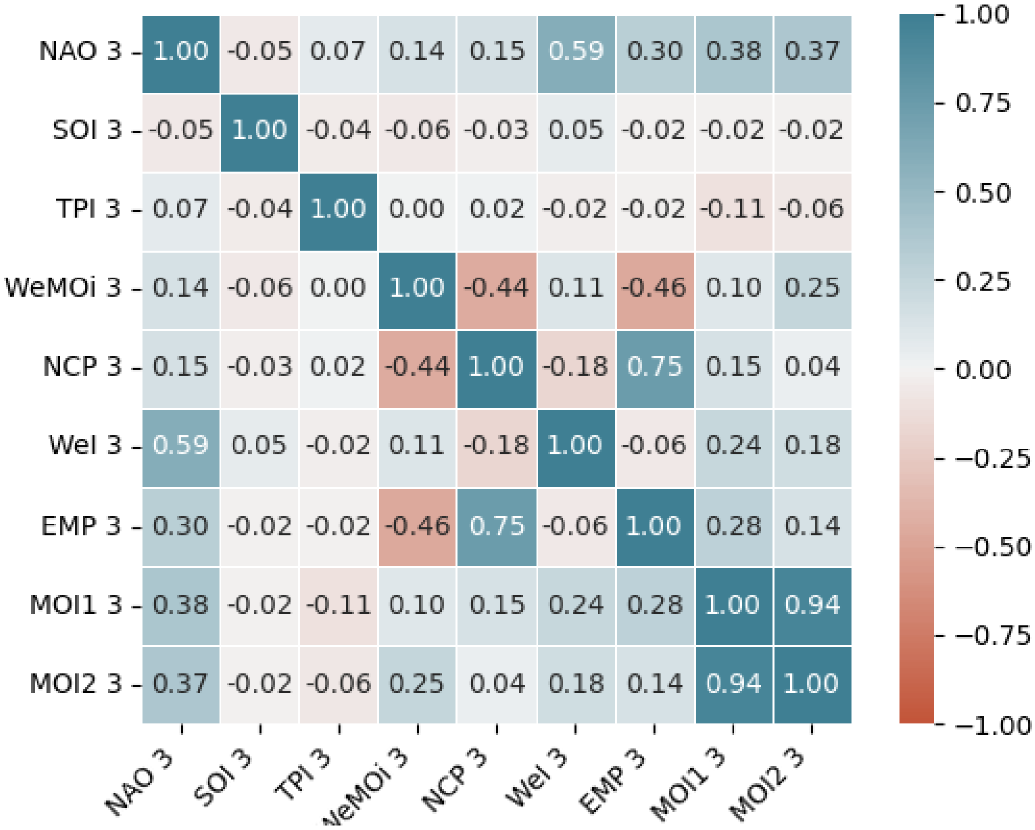

Figure IV.5 — Atmospheric circulation indices correlation matrix for 3 months

compared to the previous matrix the values didn't differe a lot

Atmospheric circulation indices correlation matrix for 6 month

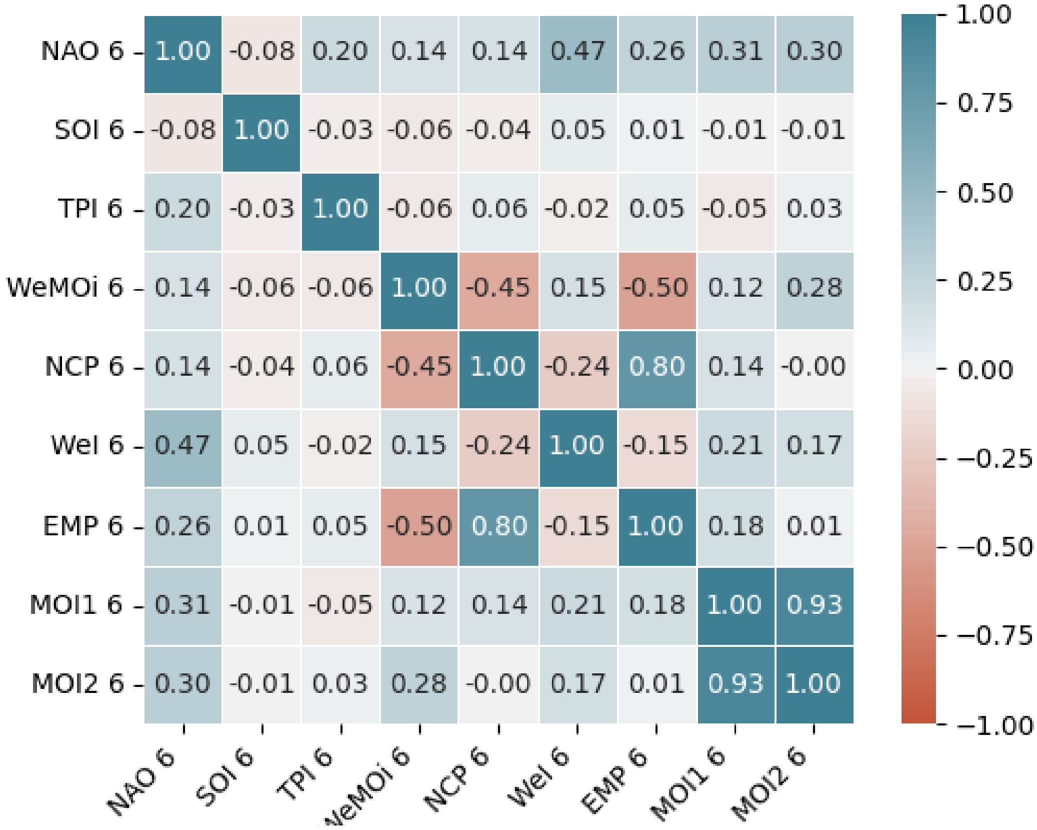

Figure IV.6 — Atmospheric circulation indices correlation matrix for 6 months

compared to the previous matrices the correlation between NCP and EMP got higher in the scale of 6 months

Correlation matrix of Atmospheric circulation indices with SPEI index 1 month

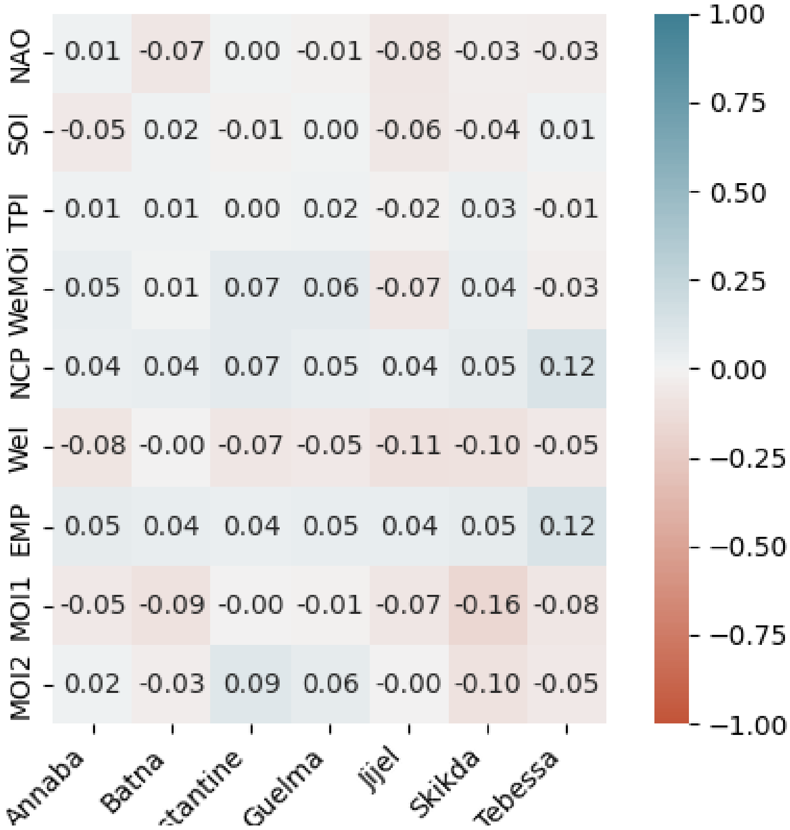

Figure IV.7 — Correlation matrix of Atmospheric circulation indices with SPEI index of 1 month

Correlation matrix of Atmospheric circulation indices with SPEI index 3 month

for this model we calculated the mean of three months of any atmospheric index and give it to that month as a representative of that month and 2 previous months

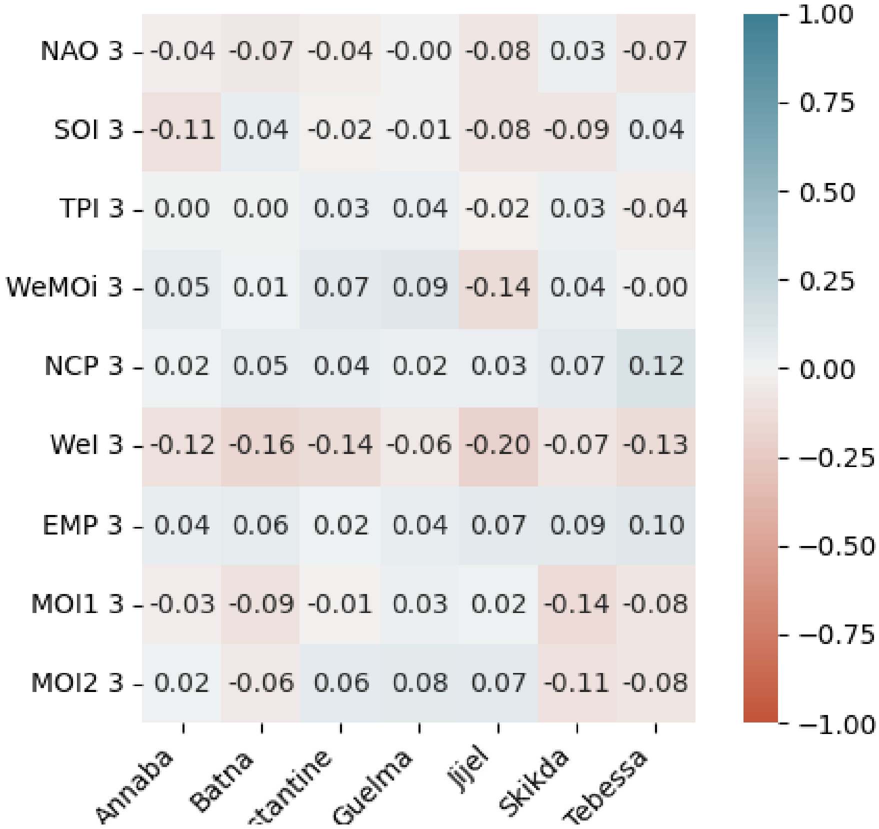

Figure IV.8 — Correlation matrix of Atmospheric circulation indices with SPEI index of 3 months

Correlation matrix of Atmospheric circulation indices with SPEI index 6 month

for this model we calculated the mean of six months of any atmospheric index and give it to that month as a representative of that month and 5 previous months

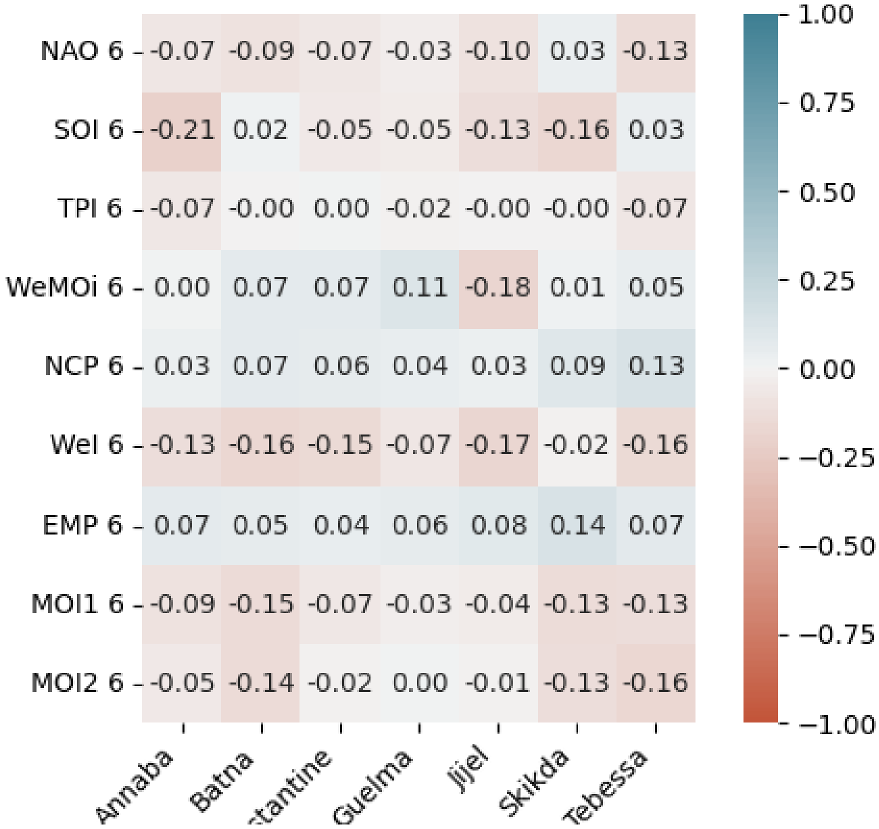

FIGURE IV.9 — Correlation matrix of Atmospheric circulation indices with SPEI index of 6 months

In all these matrices we can notice that the label of our model whether it is SPEI of 1 month, 3 months or 6 months is weakly correlated to the features of our model (the atmospheric circulation indices)

IV.3 Model validation criterias and results

Accuracy

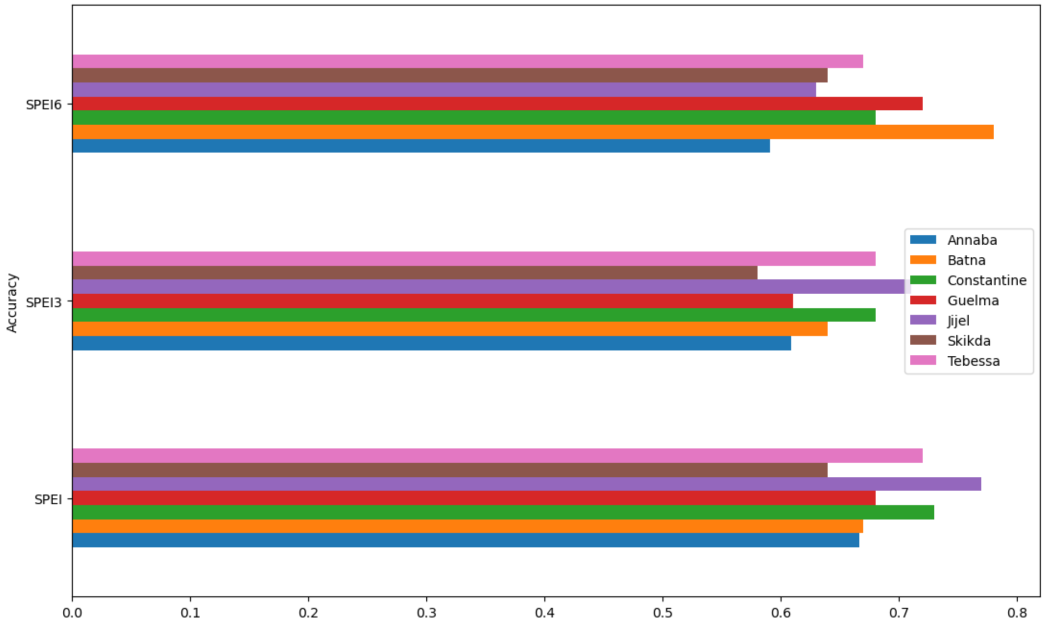

FIGURE IV.10 — Accuracy bar chart

Discussion

the previous figure shows the accuracy of random forest model for each station and for different SPEI time scales calculations, the more accurate results go to SPEI of 1 month and 6 months.

— In Annaba station , the accuracy for SPEI 1 month model is 0.66 , for 3 months model is 0.6 and for 6 months model is 0.59. we notice that SPEI of 1 month is the most accurate.

— In Batna station , the accuracy for SPEI 1 month model is 0.67, for 3 months model is 0.64 and for 6 months model is 0.78. we notice that SPEI of 6 months is the most accurate.

— In Constantine station , the accuracy for SPEI 1 month model is 0.73, for 3 months model is 0.68 and for 6 months model is 0.68. we notice that SPEI of 1 month is the most accurate.

— In Guelma station , the accuracy for SPEI 1 month model is 0.68, for 3 months model is 0.61 and for 6 months model is 0.72. we notice that SPEI of 6 month is the most accurate.

— In Jijel station , the accuracy for SPEI 1 month model is 0.77, for 3 months model is 0.71 and for 6 months model is 0.63. we notice that SPEI of 1 month is the most accurate.

— In Skikda station , the accuracy for SPEI 1 month model is 0.64, for 3 months model is 0.58 and for 6 months model is 0.64. we notice that SPEI of 1 month is the most accurate.

— In Tebessa station , the accuracy for SPEI 1 month model is 0.72, for 3 months model is 0.68 and for 6 months model is 0.67. we notice that SPEI of 1 month is the most accurate.

Table IV.4 — Accuracy table

Accuracy	Annaba	Batna	Constantine	Guelma	Jijel	Skikda	Tebessa
SPEI	0,6666	0,67	0,73	0,68	0,77	0,64	0,72
SPEI3	0,609	0,64	0,68	0,61	0,71	0,58	0,68
SPEI6	0,5909	0,78	0,68	0,72	0,63	0,64	0,67

we generally notice that the accuracy of the model in all stations is high in the SPEI of 1 month and slightly decrease in the 6 months, whereas the highest acccuracies were related to Jijel and Constantine. the highest accuracy value was for Batna in the scale of 6 months with value of 78 percent in accuracy.

Feature importance

SPEI of 1 month :

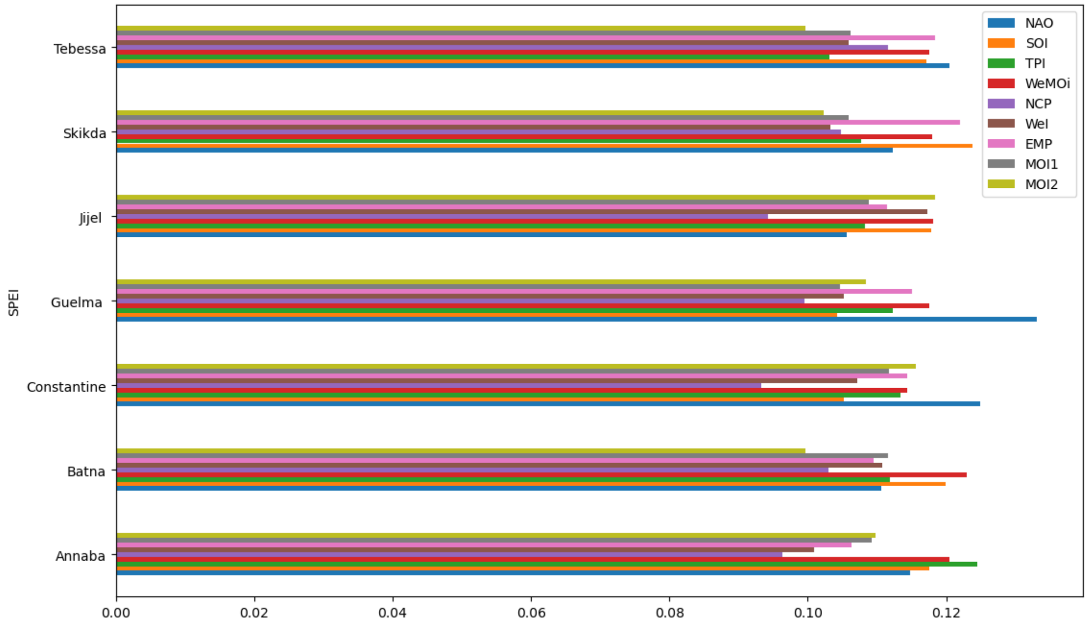

Figure IV.11 — Feature importances for models related to SPEI of 1 month

Discussion

the previous chart shows the importance of each atmospheric circulation index for each random forest model proper to each meteorological station for SPEI time scale of 1 month

Table IV.5 — Feature importances for models related to SPEI of 1 month table

SPEI	NAO	SOI	TPI	WeMOi	NCP	WeI	EMP	MOI1	MOI2
Annaba	0,115	0,118	0,124	0,120	0,096	0,101	0,106	0,109	0,110
Batna	0,111	0,120	0,112	0,123	0,103	0,111	0,110	0,112	0,100
Constantine	0,125	0,105	0,113	0,114	0,093	0,107	0,114	0,112	0,116
Guelma	0,133	0,104	0,112	0,118	0,100	0,105	0,115	0,105	0,108
Jijel	0,106	0,118	0,108	0,118	0,094	0,117	0,111	0,109	0,118
Skikda	0,112	0,124	0,108	0,118	0,105	0,103	0,122	0,106	0,102
Tebessa	0,120	0,117	0,103	0,118	0,112	0,106	0,118	0,106	0,100

— In Annaba station , the atmospheric circulation indices that highly influence our model are TPI, WeMOI and SOI

— In Batna station , the atmospheric circulation indices that highly influence our model are WeMOI, SOI and TPI

— In Constantine station , the atmospheric circulation indices that highly influence our model are NAO, WeMOI and EMP

— In Guelma station , the atmospheric circulation indices that highly influence our model are NAO, WeMOI and EMP

— In Jijel station , the atmospheric circulation indices that highly influence our model are MOI2, WeMOI and SOI

— In Skikda station , the atmospheric circulation indices that highly influence our model are SOI, EMP and WeMOI

— In Tebessa station , the atmospheric circulation indices that highly influence our model are NAO, EMP and SOI

SPEI of 3 months :

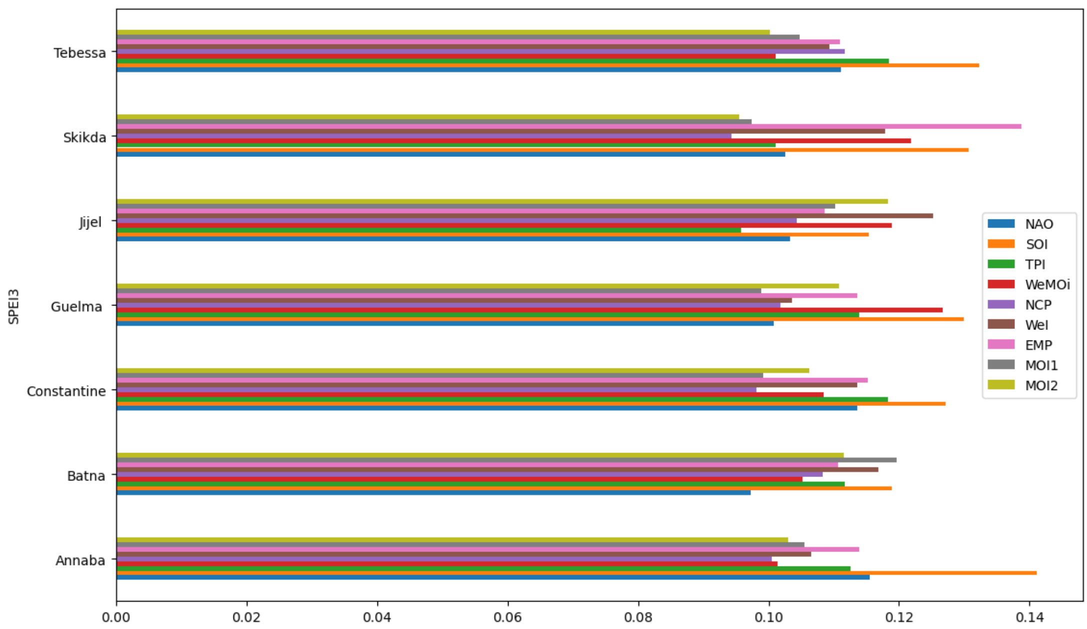

FIGURE IV.12 — Feature importances for models related to SPEI of 3 months

Discussion

the previous chart shows the importance of each atmospheric circulation index for each random forest model proper to each meteorological station for SPEI time scale of 3 months

Table IV.6 — Feature importances for models related to SPEI of 3 months table

SPEI3	NAO	SOI	TPI	WeMOi	NCP	WeI	EMP	MOI1	MOI2
Annaba	0,115	0,141	0,113	0,101	0,100	0,107	0,114	0,106	0,103
Batna	0,097	0,119	0,112	0,105	0,108	0,117	0,111	0,120	0,112
Constantine	0,114	0,127	0,118	0,108	0,098	0,114	0,115	0,099	0,106
Guelma	0,101	0,130	0,114	0,127	0,102	0,104	0,114	0,099	0,111
Jijel	0,103	0,115	0,096	0,119	0,104	0,125	0,109	0,110	0,118
Skikda	0,103	0,131	0,101	0,122	0,094	0,118	0,139	0,097	0,096
Tebessa	0,111	0,132	0,118	0,101	0,112	0,109	0,111	0,105	0,100

— In Annaba station , the atmospheric circulation indices that highly influence our model are SOI, NAO and EMP

— In Batna station , the atmospheric circulation indices that highly influence our model are MOI1, SOI and WeI

— In Constantine station , the atmospheric circulation indices that highly influence our model are SOI, TPI and EMP

— In Guelma station , the atmospheric circulation indices that highly influence our model are SOI, WeMOI and TPI

— In Jijel station , the atmospheric circulation indices that highly influence our model are WeI, WeMOI and MOI2

— In Skikda station , the atmospheric circulation indices that highly influence our model are EMP, SOI and WeMOI

— In Tebessa station , the atmospheric circulation indices that highly influence our model are SOI, TPI and NCP

SPEI of 6 months :

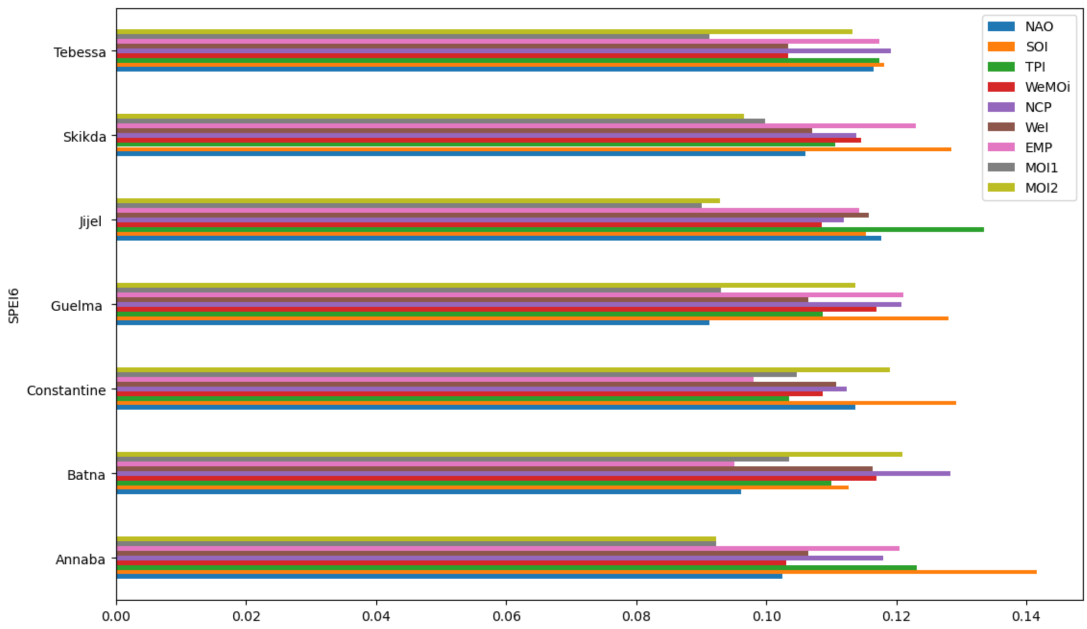

Figure IV.13 — Feature importances for models related to SPEI of 6 months

Discussion

the previous chart shows the importance of each atmospheric circulation index for each random forest model proper to each meteorological station for SPEI time scale of 6 months

Table IV.7 — Feature importances for models related to SPEI of 6 months table

SPEI3	NAO	SOI	TPI	WeMOi	NCP	WeI	EMP	MOI1	MOI2
Annaba	0,115	0,141	0,113	0,101	0,100	0,107	0,114	0,106	0,103
Batna	0,097	0,119	0,112	0,105	0,108	0,117	0,111	0,120	0,112
Constantine	0,114	0,127	0,118	0,108	0,098	0,114	0,115	0,099	0,106
Guelma	0,101	0,130	0,114	0,127	0,102	0,104	0,114	0,099	0,111
Jijel	0,103	0,115	0,096	0,119	0,104	0,125	0,109	0,110	0,118
Skikda	0,103	0,131	0,101	0,122	0,094	0,118	0,139	0,097	0,096
Tebessa	0,111	0,132	0,118	0,101	0,112	0,109	0,111	0,105	0,100

— In Annaba station , the atmospheric circulation indices that highly influence our model are SOI, TPI and EMP

— In Batna station , the atmospheric circulation indices that highly influence our model are NCP, MOI2 and WeMOI

— In Constantine station , the atmospheric circulation indices that highly influence our model are SOI, MOI2 and NAO

- In Guelma station , the atmospheric circulation indices that highly influence our model are SOI, EMP and NCP
- In Jijel station , the atmospheric circulation indices that highly influence our model are TPI, NAO and WeI
- In Skikda station , the atmospheric circulation indices that highly influence our model are SOI, EMP and WeMOI
- In Tebessa station , the atmospheric circulation indices that highly influence our model are NCP, SOI and TPI

Conclusion

In this chapter, we have detailed the results and discussion of our drought prediction study, focusing on the calculation of the SPEI, the correlation between atmospheric circulation indices, and the evaluation of the Random Forest model.

We began by explaining the process of calculating the SPEI, utilizing software such as R or Python, and outlined the steps from data collection to the final computation of the index. This comprehensive approach ensured accurate and reliable SPEI values, which served as the target variable in our predictive model.

The analysis of correlations between atmospheric circulation indices revealed significant interrelationships, indicating how these indices collectively influence drought conditions. Understanding these correlations was crucial for selecting relevant features for the model and interpreting their impact on drought prediction.

The evaluation of the Random Forest model demonstrated its strong predictive performance, with high accuracy and AUC-ROC scores across various meteorological stations. The feature importance analysis provided valuable insights into the relative influence of each atmospheric index, highlighting key predictors of drought events in the region. The interpretation of these results underscored the robustness and reliability of the Random Forest model in capturing the complex dynamics of drought conditions.

Overall, the results and discussion presented in this chapter confirm the efficacy of using atmospheric circulation indices in combination with machine learning techniques to predict drought events. The findings provide a solid foundation for further research and practical applications in drought management and mitigation strategies in the northern lands of Algeria.

CONCLUSION AND PERSPECTIVES

Conclusion and perspectives

This thesis set out to explore the potential of using machine learning models, specifically the Random Forest algorithm, to predict drought events in the northern lands of Algeria. By incorporating atmospheric circulation indices as predictive features and the Standardized Precipitation Evapotranspiration Index (SPEI) as the target label, the study aimed to enhance the understanding and predictive capabilities regarding drought occurrences in this climatically vulnerable region.

The Random Forest model demonstrated robust performance across various meteorological stations, indicating its efficacy in handling the complexities and non-linearities inherent in climatic data. The model successfully captured the intricate relationships between atmospheric circulation patterns and drought conditions, as evidenced by its high predictive accuracy and consistency across different locations. Key atmospheric indices such as the North Atlantic Oscillation (NAO), Mediterranean Oscillation (MO), and El Niño Southern Oscillation (ENSO) were found to be significant predictors, highlighting their influence on regional drought dynamics.

Several critical findings emerged from this research :

Model Performance : The Random Forest model achieved strong predictive accuracy, effectively identifying drought conditions as measured by the SPEI. The model's performance was consistent across different meteorological stations, underscoring its reliability and generalizability.

Influential Factors : Atmospheric circulation indices played a crucial role in the model's predictions, confirming their significance in influencing drought patterns in northern Algeria. These indices provided valuable insights into the large-scale climatic drivers of regional drought events.

Utility of SPEI : The SPEI proved to be an effective target variable, capturing both precipitation deficits and potential evapotranspiration, thus offering a comprehensive measure of drought severity. Its use enhanced the model's ability to predict drought conditions accurately.

Implications for Drought Management : The successful application of the Random Forest model in predicting drought events holds significant implications for drought management and mitigation strategies. The model's predictive capabilities can aid policymakers and stakeholders in making informed decisions regarding water resource management, agricultural planning, and disaster preparedness.

This study's findings contribute to the growing body of knowledge on drought prediction and the application of machine learning in environmental modeling. By demonstrating the effectiveness of the Random Forest algorithm, this research provides a valuable tool for improving drought prediction in regions prone to climatic variability.

Future Research Directions Building on the success of this study, future research could explore several avenues to further enhance drought prediction models :

Incorporating Additional Features : Including other relevant climatic and environmental variables, such as soil moisture content, vegetation indices, and land surface temperature, could improve model performance and offer a more holistic understanding of drought dynamics.

Model Comparison and Hybrid Approaches : Comparing the Random Forest model with other machine learning algorithms and exploring hybrid modeling approaches could yield even more accurate and reliable predictions.

Long-Term Predictions : Extending the models to provide long-term drought forecasts could be beneficial for strategic planning and resource management.

Climate Change Scenarios : Investigating the impact of different climate change scenarios on drought predictions could provide valuable insights for future climate adaptation strategies.

In conclusion, the application of the Random Forest model in this study has proven to be a powerful approach for predicting drought events in northern Algeria. The findings underscore the potential of machine learning in enhancing environmental predictive models, thereby contributing to more effective drought management and mitigation efforts in the face of increasing climatic variability.

Bibliographie

[1] X. LIU, X. Zhu, Y. Pan, S. Li, Y. Liu, and Y. Ma, "Agricultural drought monitoring : Progress, challenges, and prospects," *Journal of Geographical Sciences*, vol. 26, pp. 750–767, 06 2016.

[2] A. Wani, F. Rahayu, I. Ben Amor, M. Quadir, M. Murianingrum, P. Parnidi, A. Ayub, S. Supriyadi, S. Sakiroh, S. Saefudin, A. Kumar, and E. Latifah, "Environmental resilience through artificial intelligence : innovations in monitoring and management," *Environmental Science and Pollution Research*, vol. 31, 02 2024.

[3] S. Ajjur and S. G. Al-Ghamdi, "Evapotranspiration and water availability response to climate change in the middle east and north africa," *Climatic Change*, vol. 166, 06 2021.

[4] E. L. Tate and A. Gustard, *Drought Definition : A Hydrological Perspective*, pp. 23–48. Dordrecht : Springer Netherlands, 2000.

[5] D. Wilhite and M. Glantz, "Understanding : the drought phenomenon : The role of definitions," *Water International - WATER INT*, vol. 10, pp. 111–120, 01 1985.

[6] G. Ravikumar, "Land use based integrated drought assessment." http ://hdl.handle.net/10603/74559, 1997. Accessed : June 4, 2024.

[7] A. S.Djouane, "Etude de la sécheresse climatique dans les hauts plateaux de l'algérie, par deux approches : Analyse statistique et images satellitaires," master's thesis, El oued, 2015.

[8] A. Hirche, "Synthèse Écologique régionale afrique du nord," 01 2013.

[9] "National drought mitigation center." `https://drought.unl.edu/`. Accessed : June 5, 2024.

[10] S. Madaan, "Causes and effects of drought." https ://eartheclipse.com/environment/natural-disaster/causes-and-effects-of-drought.html. Accessed : June 5, 2024.

[11] N. Arain, A. Akber, S. Shah, and L. Ahmad, "Analysis of historical drought using standardized precipitation index and reconnaissance drought index (rdi) in tharparkar, sindh," *Environ. Ecol*, vol. 39, no. 1, pp. 65–70, 2021.

[12] *Handbook of Drought and Water Scarcity : Principles of Drought and Water Scarcity (1st ed.)*. CRC Press. https ://doi.org/10.1201/9781315404219, 2017.

[13] E. Baltas, "Spatial distribution of climatic indices in northern greece," *Meteorological Applications : A journal of forecasting, practical applications, training techniques and modelling*, vol. 14, no. 1, pp. 69–78, 2007.

[14] R. Zhang, Y. Qu, X. Zhang, X. Wu, X. Zhou, B. Ren, J. Zeng, and Q. Wang, "Spatio-temporal variability in annual drought severity, duration, and frequency from 1901 to 2020," *Climate Research*, vol. 87, pp. 81–97, 2022.

[15] N. B. Guttman, "Comparing the palmer drought index and the standardized precipitation index 1," *JAWRA Journal of the American Water Resources Association*, vol. 34, no. 1, pp. 113–121, 1998.

[16] R. W. Katz and M. H. Glantz, "Anatomy of a rainfall index," *Monthly Weather Review*, vol. 114, no. 4, pp. 764 – 771, 1986.

[17] B. G. M. Keetch, John J., "A drought index for forest fire control," p. 35, 1968.

[18] B. Lyon, "The strength of el niño and the spatial extent of tropical drought," vol. 31, 2004.

[19] H. Jamshidi, D. Khalili, M. Rezaeianzadeh, and E. Hosseinipour, "Assessment and comparison of spi and rdi meteorological drought indices in selected synoptic stations of iran," 05 2011.

[20] S. Meyer, K. Hubbard, and D. Wilhite, "A crop-specific drought index for corn : I. model development and validation," *Agronomy Journal - AGRON J*, vol. 85, 03 1993.

[21] R. Mohammed and M. Scholz, "Impact of evapotranspiration formulations at various elevations on the reconnaissance drought index," *Water Resources Management*, vol. 31, pp. 531–548, 2017.

[22] R. S. Nwayor, I.J., "Exploring the relationship between spi and spei in a warming world," *Theor Appl Climatol*, vol. 155, pp. 2559—-2569, 2024.

[23] K. O. Berhail, S., "Comparison of the spi and spei as drought assessment tools in a semi-arid region : case of the wadi mekerra basin (northwest of algeria)," *Theor Appl Climatol*, vol. 154, pp. 1373—-1393, 2023.

[24] F.-C. R. Ortiz-Gómez, R. and G. Medina-García, "Sensitivity of the rdi and spei drought indices to different models for estimating evapotranspiration potential in semiarid regions," *Water Resour Manage*, vol. 36, pp. 2471—-2492, 2022.

[25] Z. Q.-D. Y. e. a. Xu, D., "Application of a hybrid arima-lstm model based on the spei for drought forecasting," *Environ Sci Pollut Res*, vol. 29, pp. 4128—4144, 2022.

[26] J. D.-K. R. e. a. Mousavi, R., "Analysis of historical drought conditions based on spi and spei at various timescales in the south saskatchewan river watershed, alberta, canada," *Theor Appl Climatol*, vol. 153, pp. 873–887, 2023.

[27] G.-F. Di Nunno, F., "patio-temporal analysis of drought in southern italy : a combined clustering-forecasting approach based on spei index and artificial intelligence algorithms," *Stoch Environ Res Risk Assess*, vol. 37, pp. 2349—-2375, 2023.

[28] K. R.-B. P. e. a. Shrivastava, S., "Estimation of the standardized precipitation evapotranspiration index (spei) using a multilayer perceptron artificial neural network model for central india," *Pure Appl. Geophys*, vol. 179, pp. 1461—-1473, 2022.

[29] U. o. E. A. Climatic Research Unit, "Cru ts4.04 : Climatic research unit (cru) time-series (ts) version 4.04 of high-resolution gridded data of month-by-month variation in climate," 2020. Accessed : 2023-06-15.

[30] N. C. for Environmental Information (NCEI), "Global historical climatology network," 2023. Accessed : 2023-06-15.

[31] S. Tutiempo Network, "Climate data : Weather history and climate data," 2023. Accessed : 2023-06-15.

[32] S. Wang, X. Xing, Y. Wu, J. Guo, M. Li, and B. Fu, "Seasonal response of the ndvi to the spei at different time scales in yinshanbeilu, inner mongolia, china," *Land*, vol. 13, no. 4, 2024.

[33] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, "A multiscalar drought index sensitive to global warming : The standardized precipitation evapotranspiration index," *Journal of Climate*, vol. 23, no. 7, pp. 1696 – 1718, 2010.

[34] Z. Li, Y. Yang, G. Kan, and Y. Hong, "Study on the applicability of the hargreaves potential evapotranspiration estimation method in crest distributed hydrological model (version 3.0) applications," *Water*, vol. 10, no. 12, 2018.

[35] H. M. Hasanean, A. K. Almaashi, and A. H. Labban, "Variability in global climatic circulation indices and its relationship," *Atmosphere*, vol. 14, no. 12, 2023.

[36] P. Jones, "Trans polar index (tpi)." https ://https ://crudata.uea.ac.uk/cru/data/tpi/. Accessed : June 6, 2024.

[37] I. Harris, "Mediterranean oscillation indices (moi))." https ://https ://crudata.uea.ac.uk/cru/data/moi/. Accessed : June 6, 2024.

[38] T. Osborn and R. Cornes, "Paris-london westerly index)." https ://https ://crudata.uea.ac.uk/cru/data/parislondon/. Accessed : June 6, 2024.

[39] A. Harding, "North sea caspian pattern (ncp)." https ://https ://cru-data.uea.ac.uk/cru/data/ncp/. Accessed : June 6, 2024.

[40] M. Hatzaki, H. A. Flocas, D. N. Asimakopoulos, and P. Maheras, "The eastern mediterranean teleconnection pattern : identification and definition," *International Journal of Climatology*, vol. 27, no. 6, pp. 727–737, 2007.

[41] X. Yao and Y. Liu, *Machine Learning*, pp. 341–373. Boston, MA : Springer US, 2005.

[42] S. J. Rigatti, "Random Forest," *Journal of Insurance Medicine*, vol. 47, pp. 31–39, 01 2017.

[43] J. Zhao, W. Yang, and F. Zhu, "A cnn-lstm-attention model for near-crash event identification on mountainous roads," *Applied Sciences*, vol. 14, no. 11, 2024.