



Stage de découverte

Introduction à l'analyse des données géospatiales
appliquée au contexte agricole

Daoudi Adam

Travail encadré par : M. Reda Trankil

Élève ingénieur en Data Science

INSEA

addaoudi@insea.ac.ma

11 août 2025

Remerciements

Je tiens à exprimer ma profonde gratitude à toute l'équipe du **Pôle Digital de l'Agriculture et des Forêts** pour leur accueil chaleureux, leur disponibilité et leur accompagnement tout au long de ce stage.

Un grand merci à **M. Reda Trankil**, mon encadrant, pour sa bienveillance, ses conseils avisés et la richesse de ses explications.

Je remercie également l'**INSEA**, et tout particulièrement la **Direction de la Formation Continue, des Stages et des Relations avec les Entreprises**, pour m'avoir offert cette opportunité de découverte professionnelle.

Enfin, merci à toutes les personnes qui ont contribué, de près ou de loin, à la réussite de cette expérience.

Résumé

Ce rapport de stage explore l'intégration des données géospatiales pour une agriculture durable dans les régions semi-arides du Maroc, face aux défis de la sécheresse et de la rareté de l'eau. En se concentrant sur la région de Béni mellal Khenifra, l'étude combine l'Indice de Précipitations Standardisées (SPI, via CHIRPS) et l'Indice de Végétation par Différence Normalisée (NDVI, via MODIS) pour détecter et prédire les épisodes de sécheresse. La méthodologie inclut l'analyse des données raster et temporelles, leur prétraitement à l'échelle saisonnière, la visualisation des corrélations SPI-NDVI, et l'application de techniques de machine learning pour classifier et anticiper les sécheresses. Aligné avec les initiatives du Pôle Digital de l'Agriculture, de la Forêt et Observatoire de la Sécheresse et la Stratégie Génération Green, ce travail met en évidence le potentiel des données géospatiales pour surveiller les cultures, optimiser l'irrigation et renforcer la résilience agricole. Les résultats soulignent leur rôle clé dans la transition vers une agriculture 4.0, favorisant des décisions éclairées pour une gestion durable des écosystèmes agricoles marocains.

Mots-clés : Pôle Digital, Agriculture 4.0, Sécheresse, Données géospatiales, SPI, NDVI, Machine-Learning

Table des matières

Remerciements	1
Abstract	3
1 Introduction	7
1.1 Agriculture marocaine : défis et solutions	7
1.2 Le Pôle Digital : Missions et Enjeux	7
1.3 Projets du Pôle Digital	8
1.4 Le Pôle Digital : Moteur de l'agriculture intelligente au Maroc	8
1.5 Le Pôle Digital : Objectifs pour une Agriculture Durable	8
2 Analyse des données géospatiales au service de l'agriculture	9
2.1 Introduction	9
2.2 Structure des données géospatiales	9
2.3 Types de données géospatiales	10
3 Synthèse de bibliographie	10
3.1 Introduction générale au contexte de la sécheresse	10
3.2 Bibliographie commentée	10
3.3 Problématique	12
4 Methodology	12
4.1 Présentation des données	12
5 Méthodologie Générale	13
5.1 Collecte des Données	13
5.2 Prétraitement des Données	13
5.3 Analyse Exploratoire des Données (EDA)	14
5.4 Modélisation et Prédiction	14
5.5 Résultats et Interprétation	14
6 Partie 1 : Collecte et prétraitement des données géospatiales	15
6.1 Sources de Données	15
6.2 Téléchargement et Extraction des Données	16
6.3 Prétraitement des Données	16
7 Partie 2 : Analyse Exploratoire des Données (EDA)	17
7.1 Visualisation des cartes NDVI et SPI	17
7.2 Visualisation Temporelle	19

7.3	Analyse de la Saisonnalité	20
7.4	Corrélation entre SPI et NDVI	22
7.5	Détection des Événements de Sécheresse et classification	23
8	Partie 3 : Modélisation et Prédiction	25
8.1	Préparation des Features	25
8.2	Modèle de Classification de la Sécheresse	25
8.3	Modèle de Prédiction du NDVI	26
9	Partie 4 : Résultats et Interprétation	28
10	Conclusion	29
A	Annexes	30

Table des figures

1	Logo du Pôle Digital de l'Agriculture, de la Forêt et Observatoire de la Sécheresse	8
2	Variation mensuelle du NDVI – Année la plus verte (2018)	18
3	SPI mensuel – Année la plus sèche (2023)	19
4	Évolution temporelle du SPI et du NDVI moyens – Région BMK (2000–2025)	20
5	Boxplot mensuels de NDVI	21
6	Boxplot mensuels de SPI	21
7	Matrice de corrélation SPI - NDVI	22
8	Correlation croisée	23
9	Carte thermique SPI	24
10	Matrice de confusion	26
11	Prévisions du NDVI à l'aide d'un modèle de régression basé sur l'indice SPI	27
12	Prévision des tendances futures du NDVI selon le modèle de prédiction. . .	28

Liste des tableaux

1	Résumé des caractéristiques des jeux de données MODIS et CHIRPS . . .	13
2	Extrait des moyennes mensuelles de SPI par date	16
3	Extrait des valeurs moyennes de NDVI par date	17
4	Épisodes de sécheresse détectés dans la région de Béni Mellal-Khénifra (2001–2024)	24
5	Rapport de classification	25
6	Statistiques SPI et NDVI	28
7	Répartition des conditions	29

1 Introduction

1.1 Agriculture marocaine : défis et solutions

L'agriculture représente un secteur clé pour l'économie marocaine, contribuant significativement à la sécurité alimentaire, à l'emploi rural et au développement socio-économique du pays. Cependant, ce secteur fait face à de nombreux défis majeurs tels que la rareté des ressources en eau, les aléas climatiques, la fragmentation des exploitations agricoles, et la nécessité d'améliorer la productivité tout en préservant l'environnement.

Pour relever ces enjeux, plusieurs solutions innovantes sont envisagées, notamment la digitalisation et l'intégration des technologies avancées comme l'intelligence artificielle, le Big Data et l'analyse des données spatiales. Ces approches permettent une gestion plus précise et efficace des ressources agricoles, facilitent la prise de décision et favorisent une agriculture durable et résiliente.

Dans ce contexte, le Pôle Digital de l'Agriculture et des Forêts joue un rôle central en animant des réseaux collaboratifs (stations météo connectées, fermes digitales) et en exploitant les données collectées pour soutenir les décisions territoriales, politiques et techniques. Par son approche innovante centrée sur la donnée et la technologie, le Pôle Digital contribue activement à construire une agriculture marocaine intelligente, capable de s'adapter aux défis actuels et futurs.

1.2 Le Pôle Digital : Missions et Enjeux

Le Pôle Digital de l'Agriculture, de la Forêt et Observatoire de la Sécheresse est une structure nationale innovante, placée sous l'égide du ministère de l'Agriculture, de la Pêche Maritime, du Développement Rural et des Eaux et Forêts. Ce pôle a pour mission principale de digitaliser les chaînes de valeur agricoles et forestières du Maroc, de l'amont à l'aval, en s'appuyant sur les technologies de l'information, les données géospatiales et les outils de la data science et du machine learning.

S'inscrivant dans une logique de transition agro-digitale, le Pôle Digital joue un rôle clé dans la modernisation du secteur agricole à travers le suivi de la sécheresse, l'analyse des données environnementales, la prise de décision intelligente, et la création de plateformes et services numériques dédiés à l'agriculture, à la forêt, et à la résilience climatique.

Ses projets sont structurés autour de programmes de recherche-développement et d'innovation (RDI) et couvrent l'ensemble de la chaîne agricole : depuis la planification et la production, jusqu'à la logistique et la commercialisation.



FIGURE 1 – Logo du Pôle Digital de l'Agriculture, de la Forêt et Observatoire de la Sécheresse

1.3 Projets du Pôle Digital

Parmi les projets emblématiques portés par le Pôle, on peut citer :

- **E-Rchadata** : plateforme fournissant des services agrométéorologiques intelligents.
- **Satellites pour l'irrigation et la data agricole** : exploitation d'images satellitaires pour un pilotage raisonné de l'eau.
- **Drones imageurs et pulvérisateurs pour les traitements phytosanitaires**.
- **E-Khsab** : élevage connecté.
- **Agribot 4.0** : robotisation de certaines tâches agricoles.
- **Frigo et transport connectés** : pour la chaîne logistique agroalimentaire.
- **E-ntaj Market Place** : plateforme de e-commerce agricole.

1.4 Le Pôle Digital : Moteur de l'agriculture intelligente au Maroc

En parallèle, le Pôle anime des réseaux nationaux collaboratifs tels que les stations météo connectées ou encore les fermes digitales, pour enrichir ses bases de données et appuyer les décisions territoriales, politiques ou techniques.

Par son approche centrée sur la donnée, la technologie et l'intelligence artificielle, le Pôle Digital se positionne aujourd'hui comme un acteur majeur dans la construction d'une agriculture marocaine durable, résiliente et intelligente.

1.5 Le Pôle Digital : Objectifs pour une Agriculture Durable

- **Promouvoir l'innovation numérique** : Utiliser des technologies comme les drones et les capteurs pour moderniser l'agriculture et la gestion forestière.

- **Surveiller la sécheresse** : Mettre en place des systèmes de suivi pour anticiper et atténuer les effets de la sécheresse.
- **Assurer la durabilité** : Protéger les ressources naturelles face aux défis climatiques.
- **Faciliter l'accès aux informations** : Fournir des données fiables aux agriculteurs et décideurs pour une meilleure planification.

2 Analyse des données géospatiales au service de l'agriculture

2.1 Introduction

Dans le cadre de l'étude du sujet de stage, l'analyse des données géospatiales émerge comme un outil puissant pour révolutionner le secteur agricole marocain, confronté à des défis comme la sécheresse et la rareté de l'eau. Ces données, issues de technologies telles que les satellites, les drones et les systèmes d'information géographique (SIG), permettent de cartographier les ressources naturelles, de surveiller les cultures en temps réel et d'optimiser la gestion de l'irrigation. Leur utilité réside dans la capacité à fournir des informations précises sur les variations climatiques, les sols et les besoins en eau, soutenant ainsi une agriculture 4.0 résiliente. Intégrées au Pôle Digital de l'Agriculture, de la Forêt et Observatoire de la Sécheresse, ces données géospatiales offrent une base solide pour des décisions éclairées, alignées avec la Stratégie Génération Green, et constituent un levier clé pour cette analyse appliquée au contexte marocain.

2.2 Structure des données géospatiales

Les données géospatiales sont des données temporelles liées à un emplacement spécifique sur la surface de la Terre. Elles peuvent fournir des informations sur les relations entre variables et révéler des modèles et des tendances. Les données géospatiales se composent de trois éléments fondamentaux qui définissent leur structure et leur utilisation.

- **Informations sur l'emplacement** : Désigne la référence géographique d'un objet ou d'un événement, généralement représentée par des coordonnées de latitude et de longitude
- **Informations sur les attributs** : Fournit des détails descriptifs sur un lieu définissant ses caractéristiques.
- **Informations temporelles** : Indique la date de collecte des données géospatiales ou la date à laquelle un événement s'est produit.

2.3 Types de données géospatiales

- **Vecteurs et attributs** : des informations descriptives sur un emplacement, exprimées sous forme de points, de lignes et de polygones.
- **Imagerie satellite raster** : des images haute résolution de notre monde, prises du dessus.
- **Nuages de points** : un ensemble de points cartographiés avec colocalisation qui peuvent être recontextualisés sous forme de modèles 3D.

3 Synthèse de bibliographie

3.1 Introduction générale au contexte de la sécheresse

La sécheresse est l'un des phénomènes climatiques les plus redoutés dans les régions semi-arides telles que le Maroc, en raison de ses impacts majeurs sur l'agriculture, les ressources en eau et les écosystèmes. Contrairement à d'autres catastrophes naturelles, elle se caractérise par une lenteur d'apparition et une difficulté à être détectée rapidement. Dans le contexte du changement climatique, les épisodes de sécheresse deviennent plus fréquents, plus intenses et plus durables, ce qui renforce la nécessité d'outils d'analyse fiables pour leur suivi et leur anticipation. Les données satellitaires librement accessibles comme celles de MODIS (pour le NDVI) et CHIRPS (pour les précipitations) offrent aujourd'hui des possibilités considérables pour observer la sécheresse sur le long terme, même dans des zones à faible couverture stationnelle.

Plusieurs types de sécheresse sont identifiés dans la littérature (McKee et al., 1993 ; Wilhite et Glantz, 1985). La sécheresse météorologique correspond à un déficit prolongé des précipitations par rapport à la moyenne. La sécheresse agricole se manifeste par un manque d'humidité dans le sol affectant la croissance des cultures, souvent à la suite d'une sécheresse météorologique. La sécheresse hydrologique concerne la diminution des débits des cours d'eau et des réserves en eaux souterraines. Enfin, la sécheresse socio-économique survient lorsque les effets du manque d'eau impactent directement les activités humaines (production agricole, approvisionnement en eau, etc.). Dans le cadre de ce travail, nous nous focaliserons sur la sécheresse agricole, qui est la plus directement observable à travers la végétation et les précipitations.

3.2 Bibliographie commentée

Le suivi de la sécheresse repose sur des indicateurs pertinents. Le SPI (Standardized Precipitation Index), proposé par McKee et al. en 1993, est un indice probabiliste basé uniquement sur les précipitations. Il est calculé à différentes échelles temporelles (1, 3,

6, 12 mois...) et permet de détecter différents types de sécheresse : courte durée pour l'agriculture, longue durée pour l'hydrologie. Un SPI inférieur à -1 signale un épisode de sécheresse. Les données CHIRPS (Climate Hazards Group InfraRed Precipitation with Stations), disponibles depuis 1981, permettent de générer des SPI avec une résolution spatiale de 0.05° .

Le NDVI (Normalized Difference Vegetation Index), introduit par Rouse et al. en 1974, est un indice spectral dérivé des bandes rouge et infrarouge proche. Il mesure la vigueur de la végétation : un NDVI élevé (> 0.5) indique une végétation saine, alors qu'une baisse du NDVI reflète un stress hydrique ou une perte de biomasse. Les données MODIS (MOD13Q1) fournissent des images NDVI tous les 16 jours avec une résolution de $250m$, ce qui est largement suffisant pour une analyse à l'échelle régionale.

L'intérêt de combiner SPI et NDVI réside dans leur complémentarité : le SPI détecte le déficit pluviométrique, tandis que le NDVI capte la réponse de la végétation. De nombreux travaux (Habitou et al., 2020 ; Ahmad et al., 2020 ; Gao et al., 2018) ont démontré que la corrélation entre SPI et NDVI est significative à l'échelle mensuelle ou saisonnière, notamment dans les régions agricoles. En analysant les deux indicateurs simultanément, il est possible de caractériser les sécheresses agricoles et de développer des modèles de prédiction.

L'échelle mensuelle est donc retenue comme la plus appropriée dans cette étude, car elle permet de suivre la dynamique de la sécheresse avec une fréquence suffisante et une disponibilité optimale des données SPI et NDVI. L'indicateur SPI sera calculé à partir des données CHIRPS mensuelles, tandis que le NDVI sera extrait des images MODIS ou récupéré depuis la plateforme marocaine CGMS. La méthodologie adoptée dans ce projet se base sur plusieurs étapes successives. Tout d'abord, une zone d'étude située dans une région semi-aride du Maroc sera sélectionnée en fonction de sa vulnérabilité aux épisodes de sécheresse [1]. Ensuite, les données nécessaires seront collectées : les précipitations issues de la base CHIRPS seront utilisées pour le calcul de l'indice SPI [2], tandis que les données de végétation NDVI seront extraites à partir du capteur MODIS [3]. Un prétraitement sera effectué afin d'harmoniser les données satellites et, si disponibles, les données issues des stations météorologiques. À cette étape, les données seront également agrégées à l'échelle saisonnière, particulièrement pour le SPI, afin de tenir compte du caractère cyclique des précipitations et de renforcer la corrélation observée durant les saisons pluvieuses [4] [5]. Des analyses visuelles seront ensuite menées à travers des cartes SPI et NDVI, l'étude de leur corrélation et l'identification des périodes de sécheresse ($SPI < -1$) [6]. La phase suivante intègre des algorithmes de machine learning pour classer les situations en sécheresse ou non-sécheresse, prédire les valeurs futures de SPI et NDVI, et détecter les anomalies extrêmes [7]. L'évaluation des modèles sera réalisée à l'aide de métriques pertinentes : coefficient de corrélation CC, précision, F1-score, ainsi que (RMSE, ME) et un coefficient de détermination R^2 satisfaisant [6]. Enfin, une visualisation

finale des résultats permettra de représenter l'évolution spatio-temporelle de la sécheresse et d'interpréter les performances des modèles prédictifs.

3.3 Problématique

Comment combiner les indices SPI (précipitations) et NDVI (santé de la végétation) avec des méthodes de machine learning pour détecter, caractériser et prédire les épisodes de sécheresse dans la région de Béni mellal Khenifra ?

4 Methodology

4.1 Présentation des données

Dans le cadre de cette étude, deux sources principales de données géospatiales ont été utilisées : les données **MODIS** (Moderate Resolution Imaging Spectroradiometer) pour le suivi de la végétation, et les données **CHIRPS** (Climate Hazards Group InfraRed Precipitation with Station data) pour l'analyse des précipitations. Le produit MODIS sélectionné est *MOD13Q1*, qui fournit l'indice de végétation par différence normalisée (**NDVI**) tous les *16 jours* avec une résolution spatiale de *250 mètres*. Ces données sont issues des capteurs embarqués sur les satellites **Terra** (actif depuis 2000, soit *25 ans* de données disponibles) et **Aqua** (actif depuis 2002, soit *23 ans* de données). Quant à **CHIRPS**, il s'agit d'un produit combinant données satellitaires et observations au sol, offrant des estimations pluviométriques depuis *1981 jusqu'en 2025*, soit *44 années* de données disponibles, avec une résolution spatiale de *0,05°* (environ 5 km) et une résolution temporelle *quotidienne ou mensuelle*. Ces deux sources sont reconnues pour leur *fiabilité scientifique* et sont largement utilisées dans les études climatiques et agricoles. Afin de garantir la cohérence temporelle entre NDVI et précipitations, une période d'analyse commune de **2003 à 2023** est retenue dans ce projet. Cette période assure à la fois la *stabilité des capteurs MODIS* et une *longueur suffisante de séries temporelles* pour des calculs robustes d'indices comme le **SPI** et l'analyse croisée avec la végétation.

TABLE 1 – Résumé des caractéristiques des jeux de données MODIS et CHIRPS

Caractéristique	MODIS (MOD13Q1)	CHIRPS
Type de données	Indice de végétation NDVI	Précipitations-SPI (satellite + stations au sol)
Fournisseur	NASA LP DAAC (capteur MODIS sur Terra/Aqua)	Climate Hazards Center, UCSB
Période disponible	2000–2025 (Terra : 25 ans, Aqua : 23 ans)	1981–2025 (44 ans)
Résolution spatiale	250 mètres	0,05° (environ 5 km)
Résolution temporelle	16 jours	Quotidienne, pentade, mensuelle
Format des fichiers	HDF, GeoTIFF	GeoTIFF, NetCDF, CSV
Système de projection	Sinusoidal (MODIS)	WGS84 (EPSG :4326)
Unité de mesure	NDVI \times 10 000 (valeurs de -2000 à +10000)	Millimètres (mm)
Produit Earth Engine	MODIS/006/MOD13Q1	UCSB-CHG/CHIRPS/DAILY ou /MONTHLY
Utilisations principales	Suivi de la végétation, agriculture, sécheresse	Analyse pluviométrique, sécheresse, calcul du SPI
Période recommandée pour l'étude	2003–2023 (période commune, stable et complète)	

5 Méthodologie Générale

Le projet vise à analyser et combiner les indices SPI (précipitations) et NDVI (santé de la végétation) pour détecter, caractériser et prédire les épisodes de sécheresse dans la région de Béni Mellal-Khénifra. La méthodologie adoptée suit une approche systématique en plusieurs étapes clés :

5.1 Collecte des Données

Les données sont collectées à partir de sources satellites fiables :

- **MODIS** : utilisé pour obtenir les données NDVI, qui reflètent la santé de la végétation.
- **CHIRPS** : employé pour les données de précipitations, nécessaires au calcul de l'indice SPI.

Les données sont récupérées automatiquement à l'aide de scripts **Python**, en exploitant des outils comme `urllib.request` pour CHIRPS et des techniques de scraping pour MODIS via le site CGMS-Maroc.

5.2 Prétraitement des Données

Les données brutes passent par un processus de prétraitement rigoureux :

- **Nettoyage** : élimination des valeurs aberrantes et traitement des données manquantes.

- **Reprojection** : uniformisation des systèmes de coordonnées avec des bibliothèques telles que `rasterio` et `geopandas`.
- **Découpage** : restriction des données à la région d'étude de Béni Mellal-Khénifra.
- **Calcul des indices** :
 - **SPI** : calculé à partir des précipitations avec une fonction dédiée `calculer_spi`.
 - **NDVI** : extraction des valeurs moyennes avec ajustement d'échelle.

5.3 Analyse Exploratoire des Données (EDA)

Une analyse exploratoire est réalisée pour comprendre les caractéristiques des données :

- **Visualisation temporelle** : création de séries chronologiques pour SPI et NDVI avec `matplotlib`.
- **Analyse de la saisonnalité** : utilisation de boxplots mensuels et saisonniers pour détecter les motifs récurrents.
- **Corrélation** : évaluation des relations entre SPI et NDVI à l'aide du coefficient de Pearson et de la corrélation croisée.
- **Détection des sécheresses** : identification des périodes de sécheresse en fonction de seuils prédéfinis pour le SPI.

5.4 Modélisation et Prédiction

Des modèles de machine learning sont développés pour analyser et prédire les épisodes de sécheresse :

- **Préparation des features** : génération de variables explicatives comme les décalages temporels (lags), moyennes mobiles et tendances.
- **Modèle de classification** : utilisation de `RandomForestClassifier` pour classer les niveaux de sécheresse.
- **Modèle de prédiction** : emploi de `RandomForestRegressor` pour prévoir les valeurs futures de NDVI.
- **Évaluation** : mesure des performances avec des métriques telles que l'accuracy, le MSE (Mean Squared Error) et le coefficient R^2 .

5.5 Résultats et Interprétation

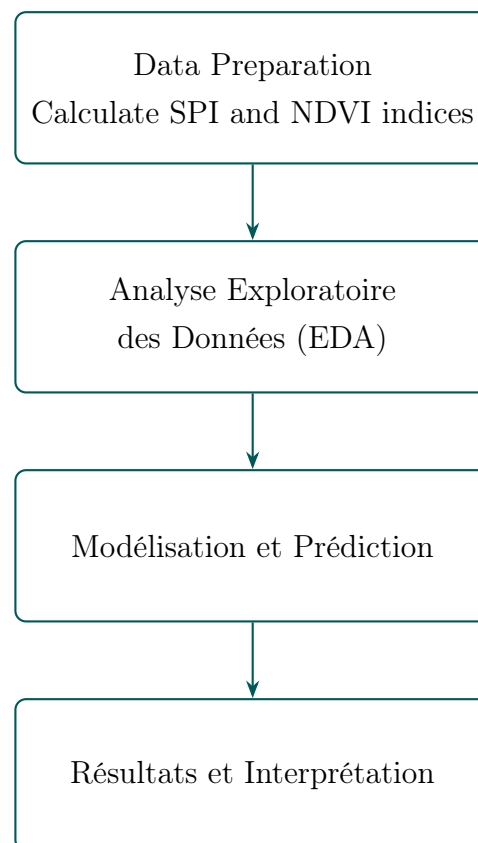
Les résultats sont synthétisés et interprétés pour une communication claire :

- **Rapport d'analyse** : présentation des statistiques clés et résumé des épisodes de sécheresse identifiés.

- **Tableau de bord interactif** : visualisation dynamique des données avec `plotly`.
- **Interprétation** : discussion des implications des résultats pour la gestion des sécheresses dans la région.

Cette méthodologie garantit une analyse complète et rigoureuse des données géospatiales, en intégrant des techniques de traitement, d'analyse statistique et de machine learning pour atteindre les objectifs du projet.

Diagramme en Bloc - Surveillance de la Sécheresse à Béni Mellal-Khénifra



6 Partie 1 : Collecte et prétraitement des données géospatiales

6.1 Sources de Données

Les données utilisées dans cette étude proviennent de deux sources principales :

- **MODIS** : Fournit les données de l'indice NDVI (Normalized Difference Vegetation Index), qui mesure la santé et la densité de la végétation.
- **CHIRPS** : Fournit les données de précipitations nécessaires au calcul de l'indice SPI (Standardized Precipitation Index), un indicateur de la sécheresse.

6.2 Téléchargement et Extraction des Données

Le téléchargement des données a été automatisé à l'aide de scripts Python. Pour les données CHIRPS, la bibliothèque `urllib.request` a été utilisée pour récupérer les fichiers depuis leur serveur officiel. Pour les données MODIS, une approche de scraping a été mise en œuvre en accédant au site CGMS-Maroc pour extraire les fichiers nécessaires.

6.3 Prétraitement des Données

Le prétraitement des données a été effectué en plusieurs étapes clés :

- **Nettoyage des données** : Les valeurs aberrantes ont été supprimées et les données manquantes ont été interpolées ou exclues selon leur impact.

```
1 data = np.where(data == src.nodata | np.isnan(data), np.nan, data)
```

Listing 1 – Nettoyage des données

- **Reprojection** : Les données ont été uniformisées dans un même système de coordonnées à l'aide des bibliothèques `rasterio` et `geopandas`.
- **Découpage géographique** : Les données ont été découpées selon les limites de la région de Béni Mellal-Khénifra à l'aide d'un fichier GeoJSON, `beni_mellal_khenifra.geojson`, permettant de filtrer uniquement les observations géospatiales pertinentes.
- **Calcul des indices** :
 - **SPI** : Calculé à partir des données de précipitations avec une fenêtre temporelle ajustable.
 - **NDVI** : Moyennes mensuelles extraites avec un ajustement d'échelle pour refléter les variations réelles.

Ainsi que l'extraction des valeurs moyenne de SPI :

```
1 spi_df = extract_mean_values("chirps_beni_mellal", "SPI", shapefile_path="beni_mellal_khenifra.geojson")
```

Listing 2 – Calcul des valeurs moyennes SPI

TABLE 2 – Extrait des moyennes mensuelles de SPI par date

Date	Mean SPI	Mois	Année
1981-01-01	-1.052647	1	1981
1981-02-01	-0.441742	2	1981
1981-03-01	0.077146	3	1981
1981-04-01	0.208871	4	1981
1981-05-01	-0.147592	5	1981

Et ceux de NDVI :

```
1 ndvi_df = extract_mean_values("ndvi_beni_mellal", "NDVI", shapefile_path
    ="beni_mellal_khenifra.geojson")
```

Listing 3 – Calcul des valeurs moyennes NDVI

TABLE 3 – Extrait des valeurs moyennes de NDVI par date

Date	Mean NDVI	Mois	Année
2000-02-18	0.337062	2	2000
2000-03-05	0.313214	3	2000
2000-03-21	0.296915	3	2000
2000-04-06	0.300845	4	2000
2000-04-22	0.290360	4	2000

Afin de croiser les informations climatiques issues de l'indice SPI et celles relatives à la végétation via l'indice NDVI, les deux jeux de données ont été fusionnés dans un seul `DataFrame` nommé `merged_df`. Cette fusion a été effectuée sur la base de la date commune aux deux jeux de données, permettant ainsi de regrouper, pour chaque observation temporelle, les valeurs correspondantes de SPI et de NDVI. Le `DataFrame` résultant contient notamment les colonnes suivantes : `date`, `mean_spi`, `mean_ndvi`, `month`, et `year`. Cette structure unifiée a permis de réaliser des analyses conjointes et des visualisations croisant les deux indices, ainsi que d'appliquer des méthodes de classification et de modélisation prédictive basées sur ces deux variables.

```
1 merged_df = pd.merge(spi_df, monthly_ndvi, on='date', how='inner')
2 merged_df = merged_df.dropna()
```

Listing 4 – Données NDVI et SPI fusionnées

7 Partie 2 : Analyse Exploratoire des Données (EDA)

7.1 Visualisation des cartes NDVI et SPI

Afin d'identifier les années les plus représentatives des conditions extrêmes en termes de végétation et de sécheresse, un traitement statistique a été réalisé sur les données fusionnées `merged_df`. Le code suivant permet de calculer les moyennes annuelles des indices SPI et NDVI, puis de détecter automatiquement :

- L'année la plus verte, correspondant à la valeur maximale de NDVI (indice de végétation)
- L'année la plus sèche, correspondant à la valeur minimale de SPI (indice de sécheresse).

Grâce à ce traitement, l'année 2018 a été identifiée comme l'année la plus verte, tandis que l'année 2023 s'est révélée être la plus marquée par la sécheresse.

```

df = merged_df.copy()

df['date'] = pd.to_datetime(df['date'])

df['year'] = df['date'].dt.year

annual_means = df.groupby('year')[['mean_ndvi', 'mean_spi']].mean().reset_index()

year_greenest = annual_means.loc[annual_means['mean_ndvi'].idxmax(), 'year']
year_driest = annual_means.loc[annual_means['mean_spi'].idxmin(), 'year']

print(f" Année la plus verte (NDVI max): {year_greenest}")
print(f" Année la plus sèche (SPI min): {year_driest}")

```

Output:

- Année la plus verte (NDVI max) : 2018
- Année la plus sèche (SPI min) : 2023

Ces résultats ont permis de guider le choix des cartes thématiques présentées ci-après, représentant la répartition spatiale des indices SPI et NDVI pour les deux années cibles. Ces cartes offrent une visualisation intuitive des zones les plus touchées ou les plus favorables selon les conditions climatiques, facilitant ainsi l'analyse géospatiale de la sécheresse et de la couverture végétale au Maroc.

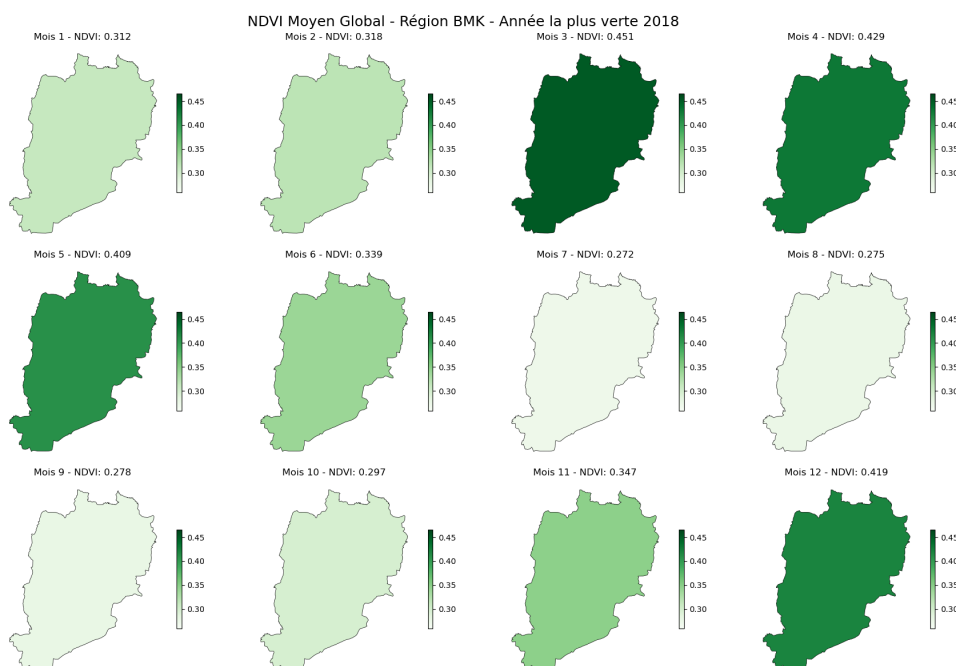


FIGURE 2 – Variation mensuelle du NDVI – Année la plus verte (2018)

La figure montre l'évolution mensuelle du NDVI en 2018, l'année la plus verte. Les pics de végétation apparaissent en mars, avril, mai et décembre, tandis que juillet et août affichent les valeurs les plus faibles.

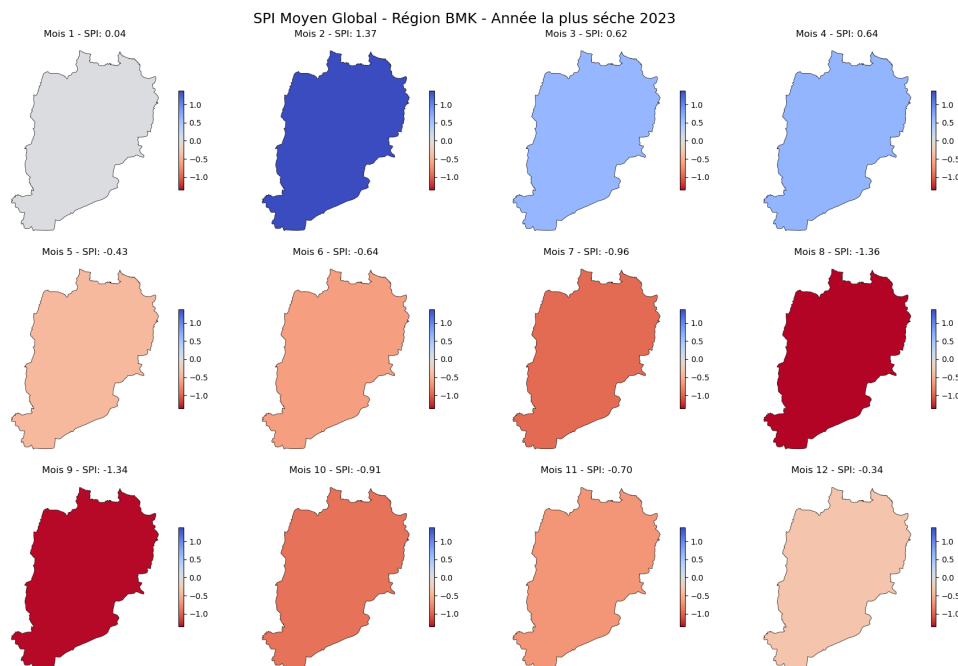


FIGURE 3 – SPI mensuel – Année la plus sèche (2023)

Cette figure montre l'évolution mensuelle du SPI en 2023, identifiée comme l'année la plus sèche. On observe des valeurs fortement négatives à partir de juillet, avec un pic de sécheresse en août ($\text{SPI} = -1.36$), indiquant une sécheresse sévère affectant presque toute la région. Les premiers mois montrent des conditions plus humides, surtout en février ($\text{SPI} = 1.37$). Cette figure illustre bien la transition d'un début d'année humide vers une sécheresse marquée en été.

7.2 Visualisation Temporelle

Les séries chronologiques des indices SPI et NDVI ont été tracées à l'aide de la bibliothèque `matplotlib`. Ces visualisations ont permis d'observer les tendances à long terme et les variations saisonnières dans la région.

La figure ci-dessous illustre l'évolution temporelle conjointe du SPI et du NDVI moyens pour la région de Beni Mellal-Khénifra entre 2000 et 2025. La courbe du SPI révèle une forte variabilité interannuelle avec l'alternance de périodes humides et sèches, marquée notamment par des épisodes de sécheresse prononcés observés autour de 2023. En parallèle, la courbe du NDVI met en évidence une dynamique saisonnière régulière jusqu'en 2020, traduisant la vigueur de la couverture végétale durant les périodes de croissance. Cependant, à partir de 2020, on note une baisse marquée des valeurs du NDVI, reflétant

une dégradation notable de la végétation. Cette tendance semble coïncider avec une diminution générale du SPI sur la même période, ce qui suggère une relation directe entre les conditions climatiques plus sèches et la détérioration de la végétation. L'année 2023, identifiée comme la plus sèche, montre une chute importante du NDVI, confirmant l'impact négatif des déficits pluviométriques sur la santé végétale de la région.

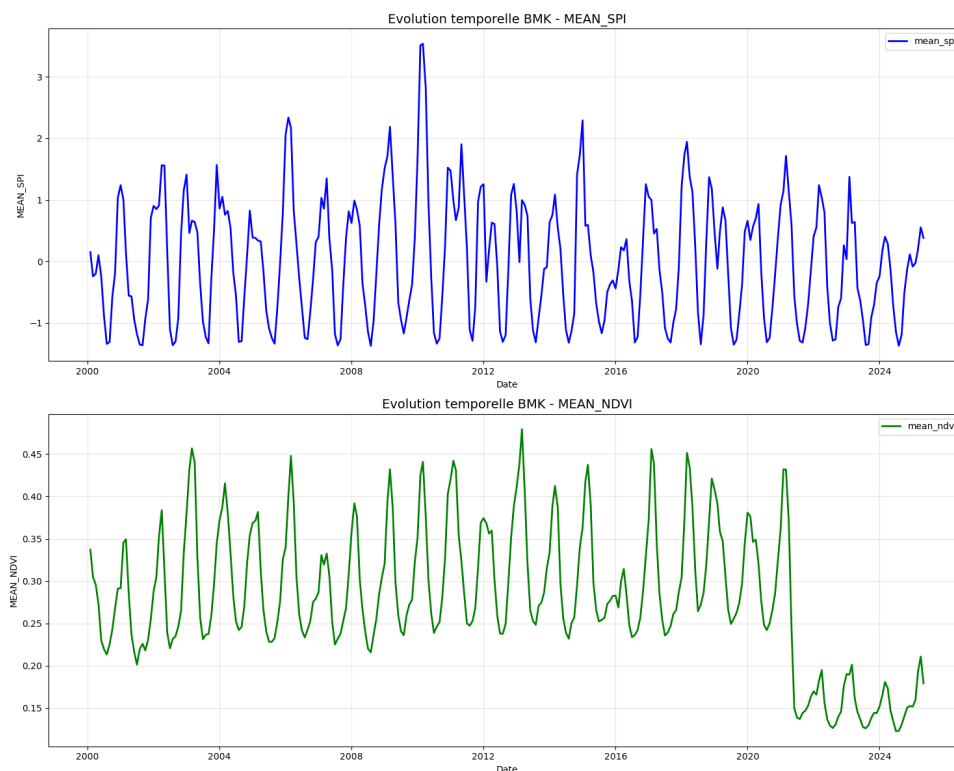


FIGURE 4 – Évolution temporelle du SPI et du NDVI moyens – Région BMK (2000–2025)

7.3 Analyse de la Saisonnalité

Des boxplots mensuels et saisonniers ont été générés pour identifier les motifs récurrents dans les données NDVI et SPI, révélant des cycles liés aux saisons agricoles.

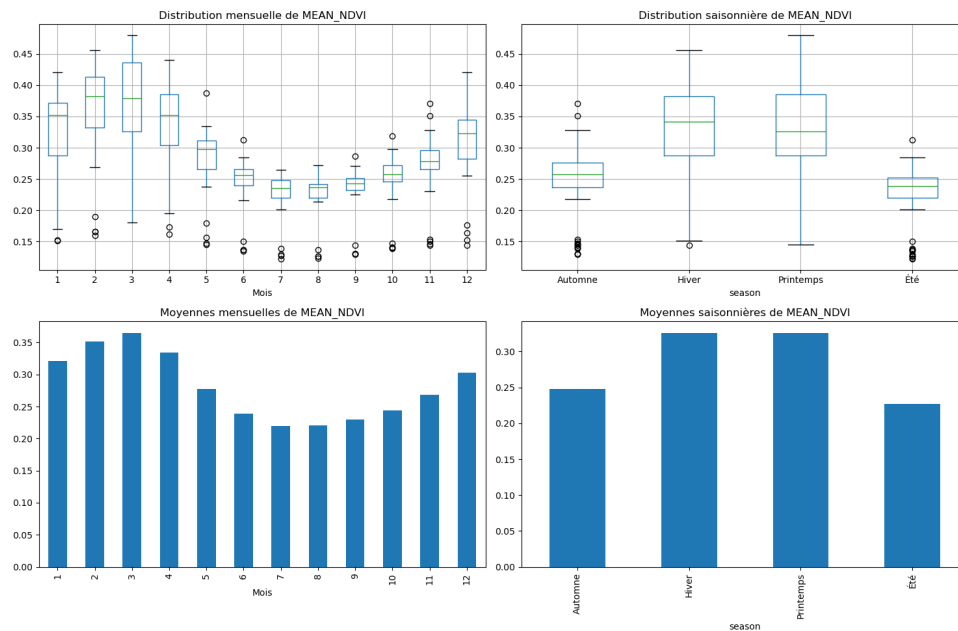


FIGURE 5 – Boxplot mensuels de NDVI

Analyse du NDVI : L'évolution mensuelle du NDVI montre un maximum en hiver et au printemps (0,35–0,45, de janvier à avril), période correspondant probablement à la saison des pluies et à une couverture végétale dense. Les valeurs chutent nettement en été (0,20–0,25, de juin à septembre), traduisant un stress hydrique marqué et une faible activité photosynthétique. L'analyse saisonnière confirme ce cycle : un NDVI élevé en hiver-printemps, un minimum estival et une reprise modérée en automne, reflet des conditions climatiques semi-arides de la région étudiée.

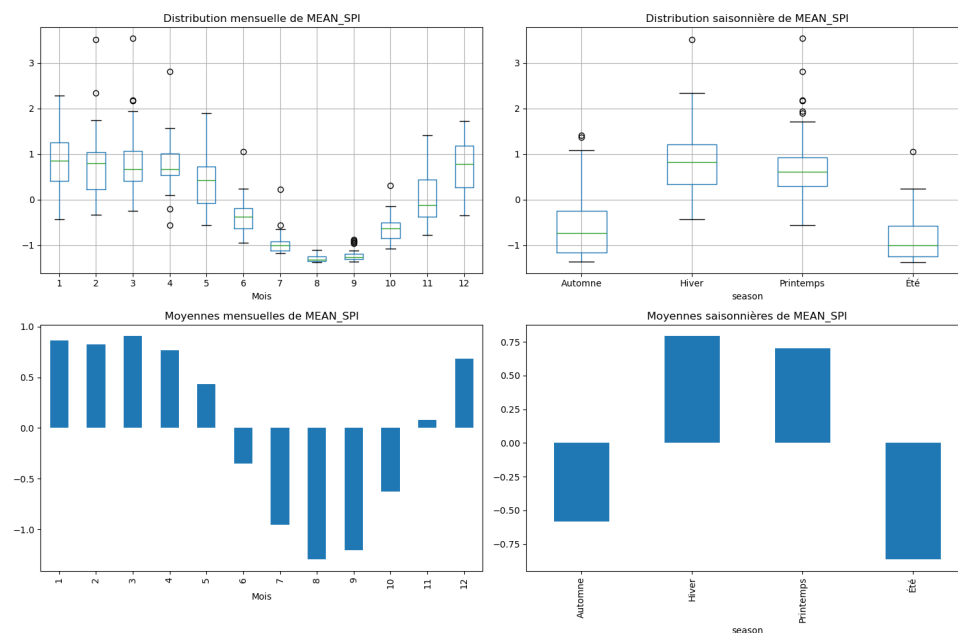


FIGURE 6 – Boxplot mensuels de SPI

Analyse du SPI et dépendanc NDVI–SPI : Le SPI présente un profil similaire, avec des valeurs positives en hiver et au printemps (jusqu'à +0,9 en mars) indiquant un excédent pluviométrique, puis une chute marquée en été (jusqu'à -1,3 en août) caractérisant une sécheresse sévère. La concordance temporelle entre SPI positif et NDVI élevé confirme la dépendance directe de la végétation à l'humidité disponible, tandis que les périodes de SPI négatif correspondent à une réduction du couvert végétal. Ce schéma, typique des zones méditerranéennes et tropicales à saison sèche, met en évidence l'importance des précipitations hivernales pour soutenir la productivité agricole.

7.4 Corrélation entre SPI et NDVI

Une analyse de corrélation a été réalisée avec le coefficient de Pearson et une corrélation croisée pour évaluer la relation entre les précipitations (SPI) et la végétation (NDVI). Les résultats ont montré une corrélation positive significative avec un décalage temporel de 1 à 3 mois.

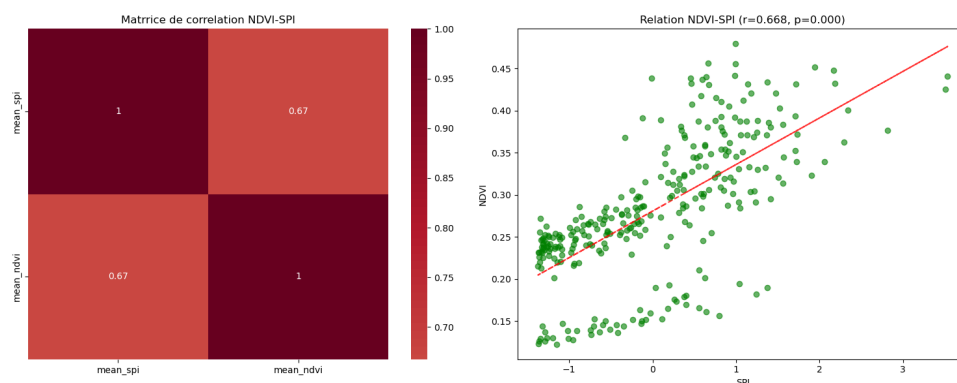


FIGURE 7 – Matrice de corrélation SPI - NDVI

Corrélation NDVI–SPI : La matrice de corrélation indique un coefficient de 0,67 entre le NDVI moyen et le SPI moyen, traduisant une relation positive modérée à forte entre l'humidité climatique et l'état de la végétation. Le nuage de points accompagné de la droite de régression confirme cette tendance : les valeurs élevées du SPI (conditions humides) sont généralement associées à un NDVI plus élevé, tandis que les valeurs négatives du SPI (sécheresse) correspondent à un NDVI plus faible. Cette relation statistiquement significative ($p < 0,001$) met en évidence l'influence directe de la disponibilité en eau sur la dynamique saisonnière de la végétation.

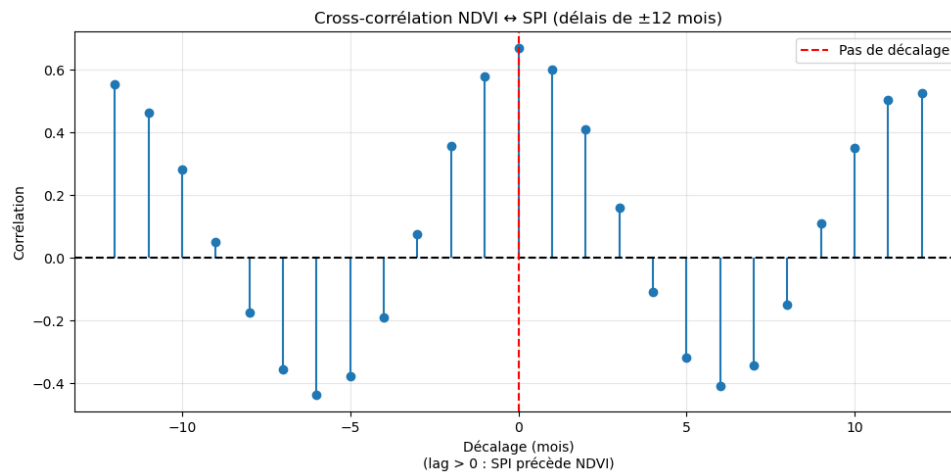


FIGURE 8 – Corrélation croisée

La figure suggère une corrélation croisée où les variations du SPI précèdent celles du NDVI, indiquant un effet retardé de la sécheresse sur la végétation.

7.5 Détection des Événements de Sécheresse et classification

Afin d'interpréter les conditions hydrologiques et végétales de la région, une classification des périodes de sécheresse a été établie en croisant les valeurs du SPI et du NDVI, permettant d'identifier les épisodes extrêmes, modérés ou normaux selon des seuils définis.

```
def classify_drought(spi_series, ndvi_series):
    result = []
    for spi, ndvi in zip(spi_series, ndvi_series):
        if pd.isna(spi) or pd.isna(ndvi):
            result.append(np.nan)
        elif spi < -2 and ndvi < 0.2:
            result.append('Sechresse extrême')
        elif spi < -1.5:
            result.append('Sechresse sévère')
        elif spi < -1.0:
            result.append('Sechresse modérée')
        elif spi > 0.5:
            result.append('Humide')
        else:
            result.append('Normale')
    return pd.Series(result, index=spi_series.index)
```


Les périodes de sécheresse ont été identifiées en appliquant des seuils au SPI (ex. $SPI < -1$ pour une sécheresse modérée). Voici un tableau récapitulatif des événements détectés :

Début	Fin	Durée (mois)	SPI Min	SPI Moyen	NDVI Min
2001-07	2001-09	3	-1.36	-1.30	0.201
2002-07	2002-09	3	-1.36	-1.26	0.221
2005-07	2005-09	3	-1.34	-1.22	0.228
2007-07	2007-09	3	-1.36	-1.27	0.225
2010-07	2010-09	3	-1.34	-1.25	0.239
2012-07	2012-09	3	-1.31	-1.21	0.238
2014-07	2014-09	3	-1.32	-1.19	0.232
2017-07	2017-09	3	-1.32	-1.22	0.236
2019-07	2019-09	3	-1.35	-1.24	0.249
2021-07	2021-10	4	-1.32	-1.17	0.137
2022-07	2022-09	3	-1.29	-1.19	0.126
2024-07	2024-09	3	-1.37	-1.24	0.123

TABLE 4 – Épisodes de sécheresse détectés dans la région de Béni Mellal-Khénifra (2001–2024)

Événements de sécheresse détectés : 12

Afin de visualiser la distribution temporelle et l'intensité des sécheresses, une carte thermique a été générée à partir des valeurs moyennes mensuelles de l'indice SPI. Celle-ci permet d'identifier rapidement les mois et années les plus touchés, ainsi que les périodes excédentaires en humidité.

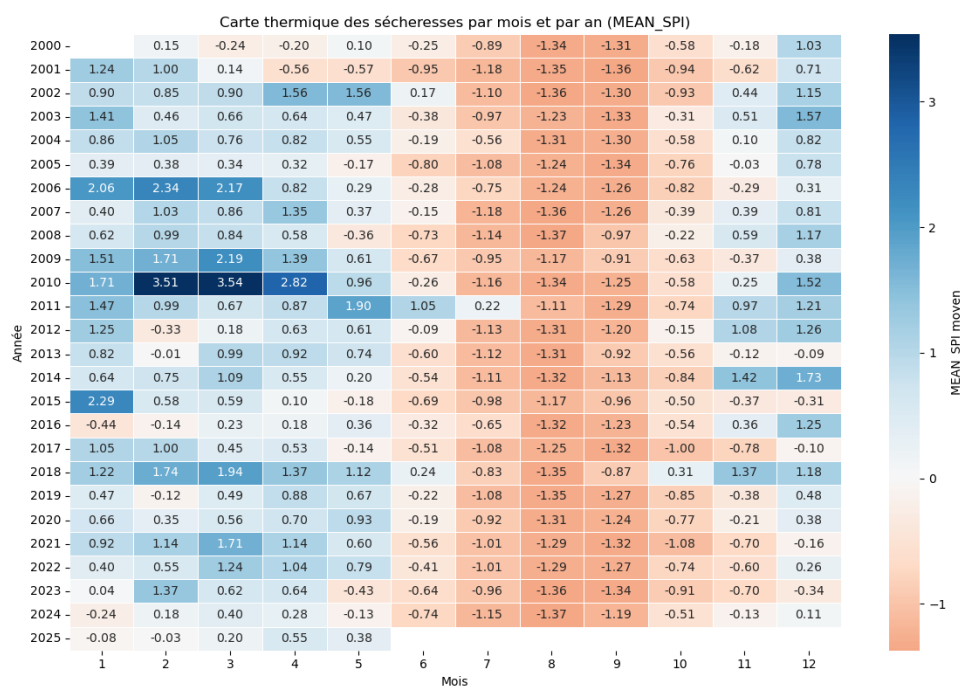


FIGURE 9 – Carte thermique SPI

Le graphique a été produit à l'aide de la fonction `generate_drought_heatmap`, qui calcule la moyenne mensuelle de l'indice SPI pour chaque année, puis la représente sous forme d'une matrice colorée. Les teintes bleues traduisent des conditions humides, tandis que les teintes rouges indiquent des périodes de sécheresse, offrant ainsi une lecture claire et synthétique de la variabilité spatio-temporelle des événements

8 Partie 3 : Modélisation et Prédiction

8.1 Préparation des Features

Afin d'optimiser la performance des modèles prédictifs, la fonction `prepare_features` a été conçue pour enrichir la base de données initiale avec des variables explicatives supplémentaires. Elle extrait d'abord les variables temporelles (mois, année, saison) pour capturer la saisonnalité, puis crée des variables de décalage (lags) pour les indices NDVI et SPI sur 1, 2, 3, 6 et 12 mois, intégrant ainsi la mémoire des conditions passées. Des moyennes mobiles sur 3, 6 et 12 mois sont ensuite calculées pour lisser les variations et dégager les tendances, complétées par des indicateurs de tendance instantanée (variation mensuelle) des deux indices. Ces transformations fournissent au modèle des informations temporelles, saisonnières et dynamiques essentielles à la prédiction des sécheresses.

8.2 Modèle de Classification de la Sécheresse

Un modèle `RandomForestClassifier` a été entraîné à l'aide de la fonction `train_drought_prediction_model`, qui prend en entrée les données enrichies par `prepare_features` et retourne les performances du modèle, pour classer les niveaux de sécheresse (ex. normale, modérée, sévère). Les performances ont été évaluées avec une accuracy de 98% et une matrice de confusion montrant une bonne séparation des classes.

TABLE 5 – Rapport de classification

Classe	Précision	Rappel	F1-score	Support
0	1.00	1.00	1.00	20
1	0.96	1.00	0.98	27
2	1.00	0.92	0.96	12
Exactitude	0.98 (sur 59 échantillons)			
Moyenne macro	0.99	0.97	0.98	59
Moyenne pondérée	0.98	0.98	0.98	59

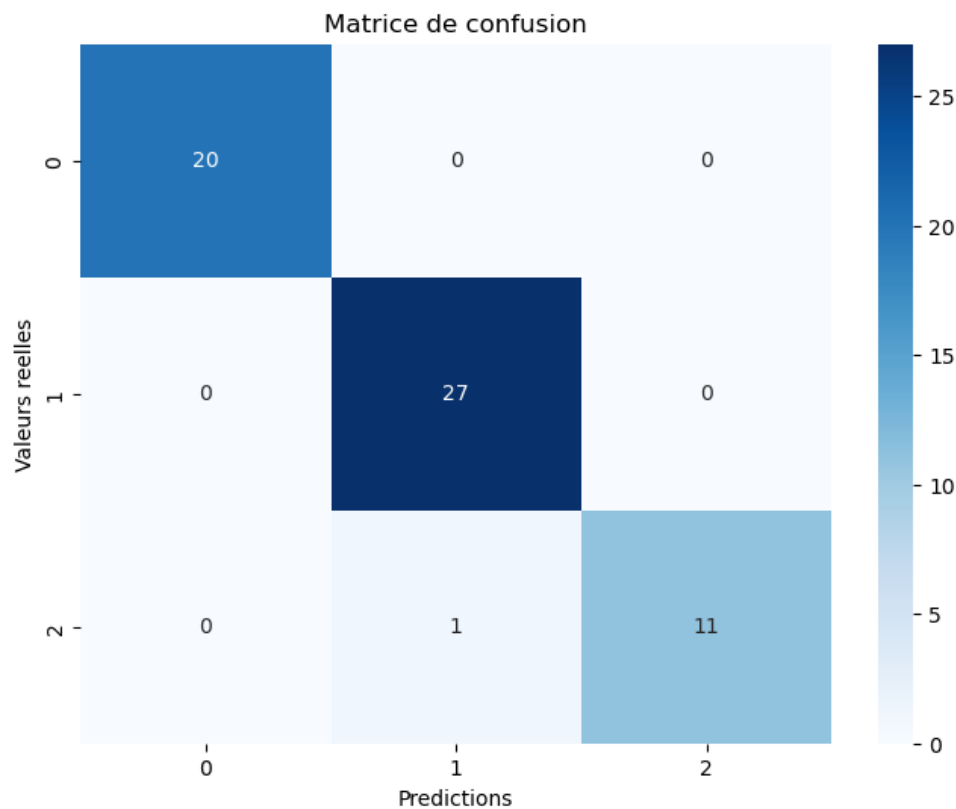


FIGURE 10 – Matrice de confusion

La matrice de confusion (Figure 10) évalue le `RandomForestClassifier` pour la classification des sécheresses (classes : 0 = normale, 1 = modérée, 2 = sévère). Avec une précision de 98,31% (58/59 instances correctes), le modèle excelle : **F1-scores** de 1,00 (classe 0), 0,981 (classe 1) et 0,957 (classe 2). Une unique erreur (classe 2 prédite comme 1) suggère une similarité SPI/NDVI entre modérée et sévère. Dans le monitoring de la sécheresse à Béni Mellal-Khénifra, cette fiabilité aide à la détection précoce, mais un rééquilibrage pour la classe 2 pourrait optimiser le rappel des événements sévères.

8.3 Modèle de Prédiction du NDVI

Un modèle `RandomForestRegressor` a été utilisé pour prédire les valeurs futures de NDVI à partir des données SPI et des features temporelles. Les métriques obtenues incluent un MSE de 0.000556 et un R^2 de 0.93, indiquant une bonne capacité prédictive.

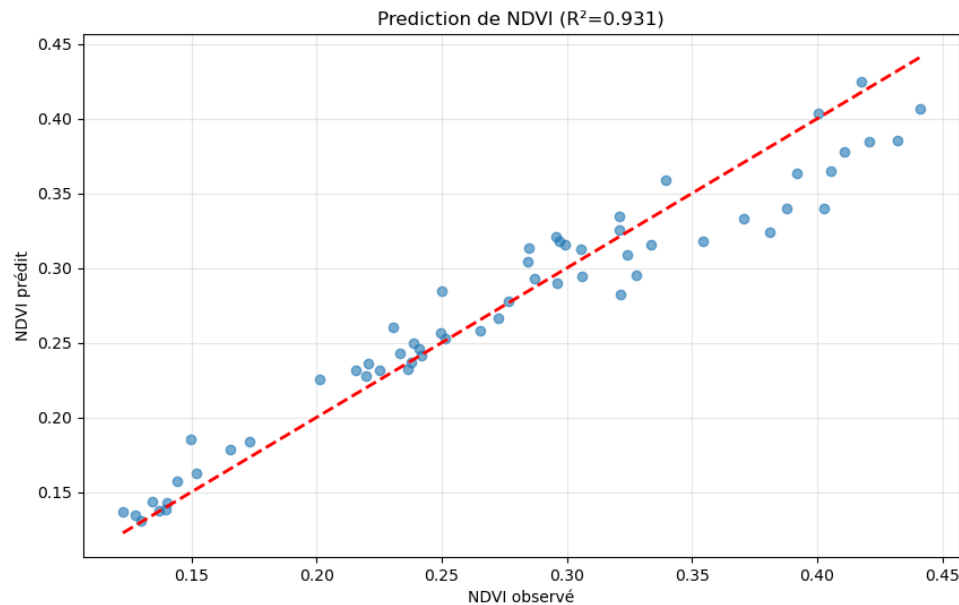


FIGURE 11 – Prévisions du NDVI à l’aide d’un modèle de régression basé sur l’indice SPI

Les points alignés sur la ligne de régression (Figure 11) indiquent une excellente capacité prédictive, particulièrement utile pour anticiper les impacts de la sécheresse à Béni Mellal-Khénifra, bien que des outliers mineurs suggèrent une sensibilité aux variations extrêmes de SPI. Des optimisations comme l’ajout de features climatiques pourraient encore améliorer la robustesse.

Afin d’anticiper l’évolution de la végétation, des prévisions du NDVI ont été réalisées sur une période de 12 mois en utilisant un modèle de régression entraîné sur les données historiques du NDVI et de l’indice SPI. Les scénarios futurs de SPI ont été simulés à partir de la moyenne et de l’écart-type observés, permettant ainsi de projeter les valeurs potentielles du NDVI dans diverses conditions climatiques. La Figure 12 présente la comparaison entre les valeurs historiques et les valeurs prévues, illustrant la continuité et la tendance générale de l’évolution du NDVI.

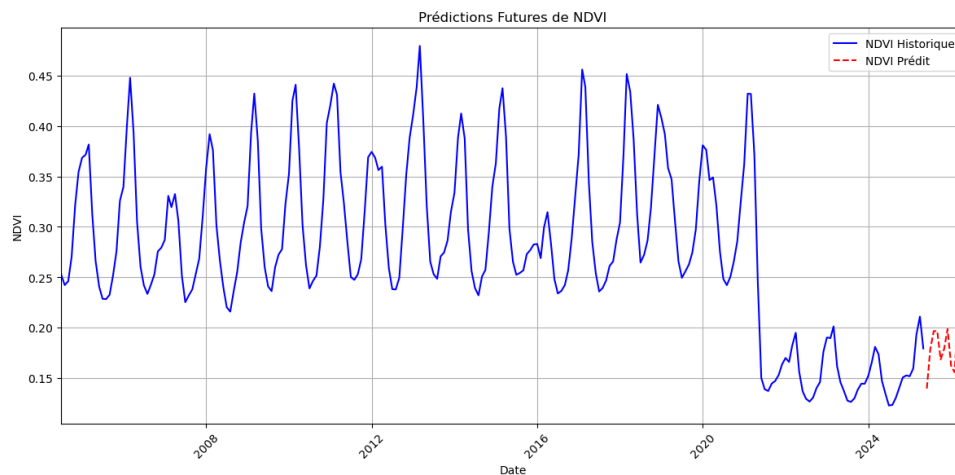


FIGURE 12 – Prévion des tendances futures du NDVI selon le modèle de prédiction.

9 Partie 4 : Résultats et Interprétation

1. Période d'analyse

- Début : 1er février 2000
- Fin : 1er mai 2025
- Nombre d'observations : 304

2. Statistiques SPI et NDVI

Indicateur	Moyenne	Écart-type	Min	Max
SPI	0.021	0.989	-1.372	3.536
NDVI	0.282	0.082	0.123	0.479

TABLE 6 – Statistiques SPI et NDVI

3. Événements de sécheresse détectés

- Nombre total : 12
- Durée moyenne : 3.1 mois
- Durée maximale : 4 mois
- SPI minimum enregistré : -1.370

Événements les plus sévères :

- 2024-07 à 2024-09 (SPI min : -1.370)

- 2007-07 à 2007-09 (SPI min : -1.363)
- 2001-07 à 2001-09 (SPI min : -1.361)

4. Répartition des conditions

Condition	Durée (mois)	Pourcentage
Normale	141	46.4%
Humide	103	33.9%
Sécheresse modérée	60	19.7%

TABLE 7 – Répartition des conditions

5. Corrélation NDVI - SPI

- Coefficient de corrélation : 0.668
- P-value : 0.000000
- Significativité : Significative

10 Conclusion

L'analyse réalisée sur la région de Béni Mellal-Khénifra, couvrant la période de février 2000 à mai 2025, a permis de caractériser avec précision les épisodes de sécheresse à partir des indices SPI et NDVI. Les résultats montrent une bonne corrélation (0.668) entre ces deux indicateurs, confirmant leur complémentarité dans la surveillance de la sécheresse.

Au total, 12 épisodes de sécheresse ont été détectés, avec une durée moyenne de 3.1 mois, dont certains épisodes particulièrement sévères, notamment celui de juillet à septembre 2024. La répartition temporelle met en évidence que près de 20% des mois analysés correspondent à des conditions de sécheresse modérée, soulignant l'importance de la problématique dans cette région.

Ces conclusions soulignent la nécessité d'un suivi continu et d'une gestion proactive des ressources hydriques afin d'atténuer les impacts socio-économiques et environnementaux liés à la sécheresse.

Références

- [1] Gao, F., Zhang, Y., Ren, X., Yao, Y., Hao, Z., & Cai, W. (2020). Evaluation of CHIRPS and its application for drought monitoring over the Haihe River Basin, China.
- [2] Climate Hazards Center. *CHIRPS : Climate Hazards Group InfraRed Precipitation with Station data*. Disponible sur : <https://www.chc.ucsb.edu/data/chirps> (consulté le 3 juillet 2025).
- [3] NASA. *MODIS Vegetation Indices (MOD13) Product*. Disponible sur : <https://modis.gsfc.nasa.gov/data/dataproduct/mod13.php> (consulté le 3 juillet 2025).
- [4] McKee, T.B., Doesken, N.J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. Department of Atmospheric Science, Colorado State University, Fort Collins, CO.
- [5] Spinoni, J., Naumann, G., Carrao, H., Barbosa, P., & Vogt, J. (2014). World drought frequency, duration, and severity for 1951–2010.
- [6] Habitou, N., Morabbi, A., Ouazar, D., Bouziane, A., Hasnaoui, M.D., & Sabri, H. (2020). *CHIRPS precipitation open data for drought monitoring : application to the Tensift basin, Morocco*. 28 septembre 2020.
- [7] Spatiotemporal characterization of current and future droughts in the High Atlas basins (Morocco).

A Annexes

Code source

L'intégralité du code Python utilisé pour cette application est disponible sur le dépôt GitHub suivant : <https://github.com/adam04-D/Drought-monitoring.git>

FIN DU RAPPORT