



الجمهورية الجزائرية الديمقراطية الشعبية



République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Ibn Khaldoun –Tiaret–

Faculté des Sciences de la Nature et de la Vie

Département de Biologie

Laboratoire d'Agro-Biotechnologie et Nutrition dans les Zones Semi-Arid

Thèse de Doctorat

En vue de l'obtention du diplôme de Doctorat

Domaine : Sciences de la Nature et de la Vie

Filière : Sciences Biologiques

Spécialité : écologie et environnement

Présentée par : CHADLI SOUHILA

Thème

**Etude comparative de la variabilité de l'indice
NDVI et de la variabilité climatique en Algérie**

Soutenu publiquement le .. / .. /202.

Devant le jury :

Dr AZZAOUI Med	Président	MCA	Université Ibn khaldoun
Dr NACEF Lamri	Examineur	MCA	Université USTHB
Dr. FEKIR youcef	Examineur	MCA	Université de mascara
Pr. REZZOUG waffa	Examinatrice	Pr	Université deIbn khaldoun
Pr DELLAL AEK.	Directeur de thèse	Pr	Université de Ibn khaldoun
Pr. BACHARI Nour El Islam	Co-Directeur de thèse	Pr	Université USTHB

Année universitaire : 2024-2025

Résumé

La prévision précise de la production agricole représente un défi en raison de la nécessité d'identifier et de comprendre les paramètres clés ainsi que leurs interactions. Les modèles efficaces doivent refléter les conditions environnementales régionales, intégrer la variabilité climatique, et considérer la nature dynamique des schémas de consommation agricole, tels que les évolutions de la demande du marché et des tendances de consommation. De plus, les variations dans les variétés de cultures, notamment les différences dans les cycles de croissance, la résistance aux ravageurs, et l'impact de la variabilité de l'NDVI (Normalized Difference Vegetation Index, soit en français : Indice de Végétation par Différence Normalisée), ainsi que les pratiques agricoles variées comme l'irrigation et la gestion des sols, contribuent à la variabilité et aux risques liés à la production. Il est essentiel de prendre en compte l'interaction entre ces paramètres, en particulier la corrélation entre la variabilité du NDVI et celle des conditions climatiques, qui influence directement la précision des prévisions. Dans ce travail, nous introduisons un framework de modélisation basé sur les données, appelé **AFFAC (Advanced Forecasting Framework for Agricultural Cultures)**, pour améliorer la précision des prévisions de production agricole. Ce framework se concentre particulièrement sur l'amélioration des prédictions pour la culture de l'olivier, un élément clé de l'agriculture méditerranéenne. En intégrant des paramètres et des conditions spécifiques à la culture de l'olivier, **AFFAC** vise à fournir des prévisions plus précises et à mieux soutenir la prise de décision dans la production et la gestion des oliveraies. Pour atteindre cet objectif, le framework établit d'abord un processus systématique de collecte et de gestion des données relatives à la production culturelle agricole, en s'appuyant sur des sources de données et des processus diversifiés. Il intègre ensuite les connaissances des experts pour identifier et sélectionner les variables d'entrée qui influencent les facteurs d'impact les plus sensibles de la production culturelle. En outre, le cadre propose des options de personnalisation, permettant aux utilisateurs de sélectionner et d'appliquer facilement différents modèles de prévision en associant des paramètres d'entrée pertinents aux spécifications des modèles. Pour valider l'efficacité de notre framework, nous avons mené des expériences approfondies afin d'évaluer la précision des modèles de prévision générés par **AFFAC**. Les résultats indiquent que le framework **AFFAC** atteint une grande précision dans la prévision des résultats de la production culturelle.

Mots-clés : Prévision des cultures, Modélisation basée sur les données, Modèles prédictifs, Production oléicole, Variabilité climatique, NDVI.

Abstract

Accurately forecasting agricultural production remains a significant challenge due to the need to identify and understand key parameters and their interactions. Effective models must reflect regional environmental conditions, incorporate climate variability, and consider the dynamic nature of agricultural consumption patterns, such as shifts in market demand and consumption trends. In addition, variations in crop varieties—including differences in growth cycles, pest resistance, and the impact of NDVI variability—as well as diverse agricultural practices like irrigation and soil management, contribute to production variability and risk. It is crucial to account for the interaction between these parameters, particularly the correlation between NDVI (Normalized Difference Vegetation Index) variability and climate variability, which directly affects forecast accuracy. In this work, we introduce a data-driven modeling framework called **AFFAC (Advanced Forecasting Framework for Agricultural Cultures)** to enhance the accuracy of agricultural production forecasting. This framework focuses particularly on improving forecasts for olive cultivation, a key component of Mediterranean agriculture. By integrating parameters and conditions specific to olive farming, **AFFAC** aims to provide more accurate forecasts and better support decision-making in the production and management of olive orchards. To achieve this goal, the framework first establishes a systematic process for collecting and managing agricultural production data, relying on diverse data sources and processes. It then incorporates expert knowledge to identify and select input variables that influence the most sensitive impact factors of agricultural production. Furthermore, the framework offers customization options, allowing users to easily select and apply different forecasting models by associating relevant input parameters with model specifications. To validate the effectiveness of our framework, we conducted extensive experiments to assess the accuracy of the forecasts generated by **AFFAC**. The results indicate that the framework **AFFAC** achieves high accuracy in predicting agricultural production outcomes.

Keywords : Crop Forecasting, Data-driven modeling, Predictive models, Olive production, Climate variability, NDVI.

ملخص

تمثل التنبؤات الدقيقة بالإنتاج الزراعي تحديًا كبيرًا نظرًا للحاجة إلى تحديد وفهم العوامل الرئيسية وتفاعلاتها. ويجب أن تعكس النماذج الفعالة الظروف البيئية الإقليمية، وتدمج التغيرات المناخية، وتأخذ في الاعتبار الطبيعة الديناميكية لأنماط الاستهلاك الزراعي، مثل تطورات الطلب في السوق واتجاهات الاستهلاك. بالإضافة إلى ذلك، تسهم الاختلافات في أنواع المحاصيل، وخاصة الفروقات في دورات النمو، ومقاومة الآفات، وتأثير تباين مؤشر الغطاء النباتي (NDVI) إلى جانب تنوع الممارسات الزراعية مثل الري وإدارة التربة، في زيادة التباين والمخاطر المرتبطة بالإنتاج. ومن الضروري مراعاة التفاعل بين هذه العوامل، خاصة الترابط بين تباين مؤشر NDVI وتغير الظروف المناخية، والذي يؤثر مباشرة على دقة التنبؤات. في هذا العمل، نقدم إطارًا نمذجيًا قائمًا على البيانات يُعرف باسم AFFAC (الإطار المتقدم لتوقع إنتاج المحاصيل الزراعية)، بهدف تحسين دقة التنبؤ بالإنتاج الزراعي. يركز هذا الإطار بشكل خاص على تحسين التوقعات الخاصة بإنتاج الزيتون، باعتباره عنصرًا رئيسيًا في الزراعة المتوسطة. ومن خلال دمج المعايير والظروف الخاصة بزراعة الزيتون، يهدف إطار AFFAC إلى تقديم تنبؤات أكثر دقة ودعم أفضل لعملية اتخاذ القرار في إنتاج الزيتون وإدارة مزارعه. ولتحقيق هذا الهدف، يبدأ الإطار بتحديد عملية منهجية لجمع وإدارة البيانات المتعلقة بالإنتاج الزراعي الثقافي، بالاعتماد على مصادر وعمليات بيانات متنوعة. ثم يدمج معارف الخبراء لتحديد واختيار المتغيرات المدخلة التي تؤثر على أكثر العوامل حساسية في إنتاج المحاصيل. علاوة على ذلك، يوفر الإطار خيارات للتخصيص، مما يسمح للمستخدمين باختيار وتطبيق نماذج التنبؤ المختلفة بسهولة، من خلال ربط المعايير المدخلة ذات الصلة بمواصفات النماذج. وللتحقق من فعالية إطارنا، أجرينا تجارب موسعة لتقييم دقة نماذج التنبؤ التي تم إنشاؤها باستخدام AFFAC وتشير النتائج إلى أن الإطار يحقق دقة عالية في التنبؤ بنتائج الإنتاج الزراعي الثقافي.

الكلمات الدالة: التنبؤ بالمحاصيل الزراعية، النمذجة المعتمدة على البيانات، النماذج التنبؤية، إنتاج الزيتون. مؤشر الفرق المعياري للنمو النباتي





Table des matières

Liste des figures	xv
Liste des tableaux	xix
Introduction Générale	7
Partie I État de l'art	15
Chapitre 1 Analyse et Impact de la Variabilité Climatique	17
1.1 Introduction	18
1.2 Le climat	18
1.2.1 Notion d'échelle en climatologie	18
1.2.1.1 Échelle spatiale	19
1.2.1.2 Échelle temporelle	20
1.2.2 Le système climatique	21
1.2.2.1 Définition du système climatique	21
1.2.2.2 Les composantes du système climatique	22
1.2.2.3 Interaction entre les composantes du système climatique	22
1.3 Les caractéristiques climatiques de l'Algérie	23
1.3.1 Saisonnalité et influence de la mer Méditerranéenne	23
1.3.1.1 Répartition des précipitations par région	23
1.3.2 Les températures moyennes	24
1.3.2.1 tendance des augmentation des températures moyennes	24
1.4 Les indices climatiques	26
1.4.1 Indices de température	26
1.4.1.1 Nombre de jours de gel et de chaleur extrême	26
1.4.2 Cumul annuel des précipitations	28

1.4.2.1	Nombre de jours de pluie	28
1.5	La variabilité climatique	28
1.5.1	Définition de la variabilité climatique	28
1.5.2	Les caractéristiques de la variabilité climatique	29
1.5.2.1	Configurations de la variabilité climatique	29
1.5.3	Facteurs de la variabilité climatique	30
1.5.3.1	Oscillations océaniques (ENSO, AMO, PDO)	32
1.5.4	Les tendances de la variabilité et du changement climatique	33
1.6	Impact sur l'agriculture	34
1.6.1	Variabilité des rendements agricoles	34
1.6.2	Besoins en irrigation et en adaptation des cultures	34
1.7	Mesures et indicateurs de la variabilité climatique	35
1.7.1	Indices climatiques	35
1.7.2	Modèles de prédiction climatique	36
1.8	Moyens d'acquisition des paramètres de variabilité climatique	37
1.8.1	Architecture de infrastructure	38
1.8.1.1	La couche des dispositifs :	38
1.8.1.2	La couche de communication :	39
1.8.1.3	La couche de stockage :	39
1.8.1.4	La couche de communication externe :	39
1.8.2	Architecture du système décisionnel	40
1.8.2.1	La collecte des données	40
1.8.2.2	L'exploitation	40
1.9	Conclusion	41
Chapitre 2 Modèles de Préviation de la Production Agricole		43
2.1	Introduction	45
2.2	Principes des Modèles de Préviation	45
2.2.1	Définition des modèles de préviation	45
2.2.2	Caractéristiques des modèles en agriculture	45
2.2.3	Principaux défis en préviation agricole	46
2.2.4	Paradigme de modélisation des coûts de production culturelle	46
2.3	Classification des Approches de Modélisation	47
2.3.1	Modèles empiriques	47
2.3.1.1	Modèle statistiques	48

2.3.1.2	Approches basées sur les séries temporelles	48
2.3.2	Modèles mécanistes	49
2.3.2.1	Modélisation biophysique	50
2.3.2.2	Modèles climatiques intégrés	50
2.3.3	Modèles hybrides	50
2.3.3.1	Combinaison empirique-mécaniste	50
2.3.3.2	Modèles basés sur l'apprentissage automatique	51
2.4	Approches de Modélisation des Cultures	51
2.4.1	Méthodologies employées	51
2.4.1.1	Collecte des données	52
2.4.1.2	Préparation et nettoyage des données	52
2.4.2	Choix des variables explicatives	52
2.4.2.1	Variables climatiques	52
2.4.2.2	Facteurs édaphiques et agronomiques	53
2.4.2.3	Facteurs spatial et indices	53
2.4.3	Techniques de modélisation utilisées	54
2.5	Approches de modèles de prédiction	54
2.5.1	Les fonctions des modèles de prédiction	55
2.6	Différentes formes d'utilisation des modèles	56
2.6.1	Approche analytique dirigée par les données	57
2.7	Définitions préliminaires et notation	57
2.7.1	Transformateurs	57
2.7.2	Clustering	58
2.7.3	Notation	58
2.7.4	Clustering K-means	59
2.7.5	Régressions linéaires et non linéaires	60
2.7.6	ARIMA	60
2.7.6.1	Apprentissage automatique	60
2.7.7	Réseaux de neurones et deep learning	61
2.7.8	Réseaux de Neurones Récurrents (RNNs)	61
2.7.9	Long Short-Term Memory (LSTM) Networks	62
2.7.10	Unités Récurrentes à Porte (GRU)	63
2.7.11	Transformers	64
2.7.11.1	Vanilla Transformer	64
2.7.11.2	Encodage des entrées et encodage positionnel	65

2.7.11.3	Attention Multi-Têtes (Multi-Head Attention)	65
2.7.11.4	Réseau Feed-forward et Réseau Résiduel	66
2.7.11.5	Innovations Récentes dans les Modèles Transformer	67
2.8	Évaluation des Modèles de Prévion	68
2.8.1	classification et comparaisons des travaux	68
2.9	Bilan et Synthèse	70
2.10	intégration des données multi sources	71
2.10.1	intégration des données sources	71
2.10.2	Federated learning	72
2.10.3	Optimisation multicritère	72
2.11	Positionnement de nos recherches	72
2.11.1	Vers la réalisation de nos objectifs	73
2.11.2	Aperçue de notre solution	74
2.12	Conclusion	74

Partie II Contribution 77

Chapitre 3 Cadre fondé sur les modèles pour la modélisation et les tests de prévision dans les systèmes de production agricole 79

3.1	Introduction	80
3.2	AFFAC : Notre cadre de travail	80
3.2.1	Facteurs influençant les coûts de production agricole	81
3.3	Formalisation du problème	82
3.4	Aperçu du cadre AFFAC	83
3.4.1	Collecte des Données	84
3.4.2	Définition de l'Espace des Paramètres pour la Fonction de Coût de Production	86
3.4.3	Définition de $\mathcal{P}rod$ en utilisant des techniques ML/DL	86
3.4.4	Le clustering K-means	86
3.4.5	Modèle ARIMA	87
3.4.6	Réseaux Neuronaux Récurrents (RNNs)	88
3.4.7	Réseaux à Mémoire à Long Terme (LSTM)	88
3.4.8	Unités Récurrentes à Porte (GRU)	89
3.4.9	Transformeurs	89
3.4.10	Evaluation Metrics	93
3.5	Les utilisateurs de notre cadre AFFAC	93

3.5.1	Décideurs	94
3.5.2	Gestionnaires	94
3.5.3	Agriculteurs	94
3.5.4	Scientifiques des données	94
3.5.5	Cas d'optimisation agricole à un niveau managérial	95
3.5.5.1	Données :	96
3.5.5.2	Variables de décision :	96
3.5.5.3	Objectifs :	96
3.5.5.4	Contraintes :	97
3.5.5.5	Exemple de planification :	97
3.5.5.6	Décision finale de localisation variétale	97
3.6	Conclusion	98
Chapitre 4 Zone d'étude : Nord de l'Algérie		99
4.1	Introduction	100
4.2	Etude de cas	100
4.2.1	Région d'étude et données utilisées	100
4.2.2	Étude des tendances des indices climatiques	101
4.2.2.1	Teste de l'Indice d'aridité de De Martonne	101
4.2.2.2	Analyse des précipitation dans le nord de l'algerie	101
4.2.2.3	Regionalisation des precipitations	103
4.2.3	Région d'étude : oleiculture dans le nord de l'Algérie	107
4.2.4	Production d'olives en Algérie	109
4.2.5	Production d'olives dans la region de tiaret	111
4.2.5.1	Présentation de l'exploitation oléicole de Djaïb	111
4.2.5.2	Paramètres morphologiques (feuille) et paramètres physiologiques (réflectance foliaire)de quelques variété d'olives	111
4.2.5.3	Analyse des réflectances foliaires des variétés d'oliviers	112
4.2.5.4	Interprétation et applications	112
4.2.5.5	Indice de végétation normalisé	113
4.2.5.6	Analyse des données saisonière de l'NDVI	114
4.2.5.7	Synthèse sur sur le cas d'étude	117
4.3	Mécanisme d'instanciation de AFFAC sur un cas d'étude	117
4.3.0.1	Expression du manifeste	118
4.3.0.2	Identification des dimensions	119
4.3.0.3	Identification des paramètres	119

4.3.0.4	Modélisation des coûts	120
4.3.0.5	Stockage des <i>Prod</i> dans une base de données	120
4.4	Instanciation de notre Framework AFFAC	120
4.4.1	Aperçu du flux de travail de conception du modèle de prévision	121
4.4.2	Scénario d'application	121
4.4.2.1	Manifest Expression	121
4.4.2.2	Identification des Dimensions et Paramètres	122
4.4.2.3	Construction du Modèle Prévisionnel	123
4.4.2.4	Calibration des poids à l'aide de l'apprentissage automatique	124
4.4.2.5	Test et validation	124
4.4.3	Généricité de notre framework AFFAC	124
4.5	Conclusion	125
Chapitre 5 Évaluation de notre Framework AFFAC		127
5.1	Introduction	128
5.2	Évaluation de notre Cadre	128
5.2.1	Zone d'étude	128
5.2.2	Ensemble de Données	128
5.2.3	Analyse des Paramètres de <i>Prod</i>	131
5.2.3.1	Analyse de Sensibilité	131
5.2.3.2	Analyse des distributions statistiques des données	131
5.2.3.3	Analyse de la corrélation des paramètres	132
5.2.3.4	Tests sur des données normales	134
5.2.4	Évaluation de nos modèles de prévision	135
5.2.4.1	Objectif de l'expérience	136
5.2.4.2	Configuration	136
5.2.4.3	Modélisation Prédictive pour les Olives : Comparaison entre ARIMA et les Réseaux de Neurones	137
5.2.4.4	Modélisation Prédictive pour les Olives : Comparaison des Modèles Transformer	137
5.2.4.5	ARIMA et Réseaux de Neurones vs Modèle Transformer	139
5.2.4.6	Analyse comparative du MAE entre les clusters pour la production d'olives	139
5.3	Implications de la recherche	142
5.3.1	Implications industrielles	142
5.4	Outil prototype d'assistance aux agriculteurs sous forme d'application mobile	143

5.4.1	Présentation de notre outil prototype	143
5.4.2	Architecture de notre application	144
5.4.2.1	Overview of Flask and RESTful API	144
5.4.2.2	Client-Side Interface	145
5.4.2.3	Server-Side Application	145
5.4.2.4	Integration with External Services	145
5.4.3	Processus de conception d'une application mobile	145
5.4.3.1	Sketching	145
5.4.3.2	Wireframing	146
5.4.3.3	Storyboards	146
5.4.4	Présentation de notre outil d'assistance	146
5.4.4.1	Écran d'accueil	146
5.4.4.2	Écran Inscription	147
5.4.4.3	Écran Menu	147
5.4.4.4	Écran de gestion des champs	148
5.4.4.5	Écran de saisie des champs	149
5.4.4.6	Écran de cartographie des champs	150
5.4.4.7	Écran Liste de champs	150
5.4.4.8	Écran de suivi des arbres	151
5.4.4.9	Écran Historique	151
5.4.4.10	Écran de signalement d'anomalies	152
5.4.4.11	Écran Analyse et Rapport	152
5.4.4.12	Écran de contact	152
5.5	Discussion	153
5.6	Discussion	156
5.7	Conclusion	158
Chapitre 6 Conclusion générale et Perspectives		161
6.1	Conclusion	162
6.1.1	État de l'art	162
6.1.2	Contributions	162
6.1.3	Evaluation et Discussion	163
6.2	Perspectives	164
6.2.1	Intégration de la détection d'anomalies	164
6.2.2	Réduction de la complexité du modèle pour améliorer la réactivité	164

6.2.3	Attribution d'un score de confiance aux mesures des capteurs	165
6.2.4	Développement d'une application mobile pour l'invocation du modèle . . .	165
	Bibliographie	167





Liste des figures

1	Schéma global illustrant les interactions entre les paramètres d'entrée génériques (P_1, P_2, \dots, P_n), leur traitement via un modèle de prévision, et un encart détaillant l'exemple classique de la variabilité climatique affectant le NDVI et le rendement agricole.	10
2	Méthodologie adoptée dans cette thèse	13
1.1	Inclusion entre les échelles locale, régionale et globale en climatologie	20
1.2	Relation cause-effet dans le système climatique	22
1.3	Interaction entre les composantes du système climatique	23
1.4	Évolution des températures moyennes entre 2000 et 2020.	25
1.5	Graphique montrant les températures moyennes, maximales et minimales sur une période de 10 jours.	25
1.6	Histogramme du cumul annuel des précipitations sur 12 mois.	28
1.7	Relations causales : Var. Clim. = Variabilité climatique (T : Temp., P : Précip., H : Humidité), Var. NDVI = Variabilité de l'indice NDVI (Végétation), Rend. Agr. = Rendements agricoles (R). Les flèches indiquent l'impact et les influences entre ces variables.	34
1.8	Schéma illustratif de l'architecture d'une solution IOT	38
1.9	L'architecture d'acquisition de données	38
1.10	Couche dispositifs	39
1.11	Les étapes du processus décisionnel	40
1.12	Architecture du système décisionnel	41
2.1	Paradigme de modélisation des coûts de production culturelle.	47
2.2	Techniques de Modélisation.	55
2.3	Schéma du processus de prévision basé sur l'analyse de données climatiques et historiques	56
2.4	Recurrent Neural Networks	62
2.5	Long Short-Term Memory (LSTM) Networks.	62
2.6	Gated Recurrent Units (GRU).	64

2.7	Overview of Transformer Models.	67
2.8	Aperçu de notre solution data-driven.	75
3.1	Aperçu du cadre AFFAC.	81
3.2	Vue d'ensemble des principaux facteurs influençant les coûts de production culturelle.	82
3.3	Processus de collecte de données pour la production agricole.	85
3.4	Flux de travail conceptuel pour la fusion des données.	85
3.5	Les utilisateurs de notre cadre AFFAC.	93
4.1	Répartition des stations météorologique de l'ONM utilisées dans l'étude.	100
4.2	Répartition géographique de l'indice d'aridité de De Martonne en Algérie durant la période 1996-2008.	102
4.3	Répartition géographique de l'indice d'aridité de De Martonne en Algérie durant la période 1982-1996 [96].	103
4.4	Représentation graphique des moyennes annuelles des stations.	104
4.5	Diagramme ombrothermique de walter et lieth de différentes régions du nord.	105
4.6	Représentation des nuages de points des valeurs prises par les deux vecteurs propres correspondant aux deux composantes principales issue d'une rotation oblimin montrant la qualité de la structure simple en prenant comme matrice de dispersion la covariance (a1) et (b1) et la corrélation (a2) et (b2), les données sont celles des précipitations.	106
4.7	Représentation géographique des vecteurs propres correspondant aux trois premières composantes, du champ de précipitations; confirmant le schéma de Buell [27]. Cas de la matrice de corrélation (A1, B1, C1) et matrice de covariance (A2, B2, C2) comme matrice de dispersion.	107
4.8	Représentation géographique des vecteurs propres issue d'une rotation (oblimin) correspondant aux trois premières composantes principales, du champ de précipitations. La matrice de corrélation (A2, B2, C2) ou la matrice de covariance (A1, B1, C1) étant prise comme matrice de dispersion.	108
4.9	Régionalisation des précipitations basée sur le réseau des station de l'ONM	109
4.10	Visualisation des zones oléicoles avec un dégradé de couleurs indiquant la répartition par région dans la zone d'étude.	110
4.11	Distribution de la production oléicole dans le Nord de l'Algérie	110
4.12	Géolocalisation de la parcelle oléicole Djaib	111
4.13	Reflectance foliaire des différentes variété d'olives.	112
4.14	Représentation graphique de l'NDVI (stade de production).	113
4.15	Corrélation de l'NDVI stade de production.	114
4.16	Représentation graphique de l'NDVI (stade de florissant).	114

4.17	Corrélation de l'NDVI stade de florissant.	115
4.18	Diagramme de Walter et Lieth des différentes stations.	115
4.19	Corrélation de l'NDVI stade de nuisant.	116
4.20	Evolution saisonnière de l'olivier en fonction de NDVI Au cours de son cycle annuel de développement.	116
4.21	Diagramme de flux du processus de conception et de test de fonction de coût <i>Prod.</i> . . .	118
4.22	Niveaux de hiérarchisation des <i>What If Design Questions</i>	119
5.1	Visualisation des Zones Oléicoles avec un Gradient de Couleur Indiquant la Répartition par Région dans la Zone d'Étude.	129
5.2	Analyse de Sensibilité des Paramètres de <i>Prod</i>	132
5.3	Extrait des distributions d'échantillons des paramètres choisis	133
5.4	Vue partielle de l'analyse de corrélation pour les paramètres <i>Prod</i>	133
5.5	Représentation de l'ensemble de données dans le plan complexe avec un grand nombre d'échantillons.	134
5.6	Variations de la température.	134
5.7	Valeur de température anormale.	135
5.8	Expérience de Prédiction de Précision des Modèles de Machine Learning.	138
5.9	Expérience de Prédiction de la Précision des Modèles de Deep Learning.	138
5.10	Visualisation du clustering des caractéristiques régionales à l'aide de l'ACP : regroupement basé sur les attributs climatiques et atmosphériques.	140
5.11	Comparaison de l'erreur moyenne par jour entre les clusters.	141
5.12	Présentation de notre outil prototype.	144
5.13	Architecture de notre application.	144
5.14	Processus de conception d'une application mobile.	145
5.15	Écran d'accueil.	146
5.16	Écran d'accueil.	147
5.17	Écran d'accueil.	148
5.18	Écran de gestion des champs.	149
5.19	Écran de saisie des champs.	150
5.20	Écran de cartographie des champs.	151
5.21	Écran Liste de champs.	152
5.22	Écran de suivi des arbres.	153
5.23	Écran Historique.	154

Liste des figures

5.24 Écran de signalement d'anomalies.	155
5.25 Écran Analyse et Rapport.	156
5.26 Écran de contact.	157



Liste des tableaux

1.1	Exemples d'échelles temporelles en climatologie	21
1.2	Configurations de la variabilité climatique	30
1.3	Facteurs de la variabilité climatique avec symboles et exemples	30
1.4	Exemples de règles conditionnelles basées sur des paramètres climatiques	31
1.5	Caractéristiques des oscillations océaniques.	33
1.6	Catégorisation des stratégies d'adaptation des cultures en fonction des besoins en irrigation	35
2.1	Exemples des principales caractéristiques des modèles de prévision pour la production d'olives avec des notations mathématiques.	46
2.2	Étapes et métriques d'évaluation des modèles de prévision en climatologie.	68
3.1	Aperçu contextuel pour le développement de la fonction de production d'olives	83
3.2	Outils et technologies selon les niveaux d'analyse en production agricole	95
3.3	Attribution des ressources et planification des tâches arboricoles	97
4.1	Classification climatique selon l'indice de De Martonne.	101
4.2	Évolution du nombre d'oliviers, des surfaces cultivées et de la production entre 2000 et 2019.	109
4.3	Le périmètre et réflectance foliaire	111
4.4	Corrélation de l'NDVI stade de production	113
4.5	Corrélation du NDVI au stade de floraison	115
4.6	Corrélation du NDVI au stade de nuisibilité	116
4.7	Vue contextuelle pour le développement de la fonction de production d'olives	122
4.8	Vue d'ensemble contextuelle du développement de la fonction de production d'olive	123
5.1	Plage de recherche des hyperparamètres pour les différents composants du Transformer.	136
5.2	RMSE moyenne des différents modèles	140

Glossaire

- **ARIMA** : Auto Regressive Integrated Moving Average (Moyenne mobile intégrée auto-régressive).
- **RNNs** : Recurrent Neural Networks (Réseaux de neurones récurrents).
- **LSTM** : Long Short-Term Memory (Mémoire à long et court terme).
- **RMSE** : Root Mean Square Error (Erreur quadratique moyenne racine).
- **MAE** : Mean Absolute Error (Erreur absolue moyenne).
- **Prod** : Agricultural Production (Production agricole).
- **AFFAC** : Advanced Forecasting Framework for Agricultural Cultures (Cadre avancé de prévision pour les cultures agricoles).
- **NDVI** : Normalized Difference Vegetation Index (Indice de végétation par différence normalisée).
- **FM** : Framework (Cadre de travail).
- **QoS** : Quality of Service attributes (Attributs de qualité de service).
- **IDE** : Integrated Development Environment (Environnement de développement intégré).
- **GPU** : Graphics Processing Unit (Unité de traitement graphique).
- **ML** : Machine Learning (Apprentissage automatique).
- **DL** : Deep Learning (Apprentissage profond).
- **NAO** : North Atlantic Oscillation (Oscillation nord-atlantique).
- **ENSO** : El Niño Southern Oscillation (Oscillation australe El Niño).
- **MQTT** : Message Queuing Telemetry Transport (Protocole de messagerie pour la télémétrie).
- **IoT** : Internet of Things (Internet des objets).

Remerciements

Ce mémoire n'aurait jamais vu le jour sans le concours et les encouragements d'innombrables amis, enseignants, et professeurs. Bien que je ne puisse tous les citer individuellement, je tiens à leur exprimer collectivement toute ma *reconnaissance* et ma profonde *gratitude*. J'espère qu'ils trouveront dans ce travail la justification de leurs efforts et leur *confiance* en moi.

Je tiens tout particulièrement à adresser mes remerciements au *Président du jury*, [Nom du Président], pour avoir honoré cette soutenance de sa présence et pour ses précieux commentaires qui enrichiront davantage ce travail.

Ma gratitude va également à mon *encadreur*, [Nom de l'encadreur], qui a su me guider avec une rigueur scientifique exemplaire et une grande *bienveillance* tout au long de ce parcours. Ses conseils avisés, son écoute, et son dévouement ont joué un rôle crucial dans la réalisation de ce mémoire.

Un grand merci à mon *co-encadreur*, [Nom du co-encadreur], pour son accompagnement attentif, ses suggestions constructives, et ses encouragements constants, qui ont grandement contribué à la qualité de ce travail.

Je suis également profondément reconnaissante envers les membres du *jury*, [Nom des membres du jury], pour le temps qu'ils ont consacré à évaluer ce mémoire, pour leurs remarques pertinentes, et pour leur *expertise*, qui m'a permis de mieux cerner les perspectives de cette étude.

Avant de conclure, je tiens à exprimer ma gratitude de manière très particulière au Dr Ouared AEK, dont le soutien exceptionnel a marqué toutes les étapes de cette aventure. Sa *générosité*, ses conseils éclairés, et son dévouement m'ont profondément inspiré et ont été pour moi une source constante de motivation. Ses encouragements bienveillants et sa confiance inébranlable en mes capacités resteront gravés dans ma mémoire.

Enfin, je tiens à exprimer toute ma *gratitude* à ma famille, qui a été mon pilier tout au long de ce parcours. À mes parents, pour leur amour inconditionnel, leur soutien moral et matériel, et leur foi en mes capacités, qui m'ont toujours porté vers l'avant. Vos sacrifices, votre patience, et vos encouragements constants ont été ma plus grande force, et ce travail est aussi le fruit de votre dévouement.

À mes amis très proches, qui ont partagé mes moments de doute et de joie, je vous remercie du fond du cœur pour votre présence précieuse. Vous avez su m'écouter, m'encourager et me motiver quand j'en avais le plus besoin. Vos paroles réconfortantes et votre soutien indéfectible m'ont permis de garder

confiance et persévérance face aux défis.

À tous, merci d'avoir été là, dans les bons comme dans les moments difficiles. Vous êtes, et resterez toujours, ma plus grande source d'inspiration et de motivation. Ce mémoire est autant le vôtre que le mien.

À toutes et à tous, merci infiniment.

Dédicace

À mes chers parents,

Les mots me manqueront toujours pour exprimer toute ma gratitude envers vous. Vous êtes ma boussole, mon pilier, et ma plus grande source de force et d'inspiration. Vos sacrifices innombrables, votre amour inconditionnel, et votre foi inébranlable en mes capacités m'ont permis de surmonter les obstacles les plus difficiles et de mener à bien ce parcours. Ce travail est le fruit de votre dévouement et de vos prières silencieuses. À vous, chers parents, je dédie humblement cette thèse en signe de mon amour infini et de ma profonde reconnaissance.

À la mémoire de mon regretté encadrant, Docteur Lazreg BENAÏCHATA,

C'est avec une profonde émotion et un immense respect que je dédie également ce travail à votre mémoire. Votre accompagnement patient, vos conseils éclairés, et votre rigueur scientifique ont été pour moi des sources d'inspiration inestimables. Vous avez toujours su allier l'exigence académique à une bienveillance qui m'a profondément marqué. Votre départ a laissé un vide immense, mais votre héritage intellectuel et humain perdure à travers cette thèse. Ce travail est le reflet de votre influence et une modeste tentative de rendre hommage à votre dévouement et à votre passion pour la recherche.

À mes parents et à mon regretté encadrant, je dédie cette œuvre avec toute ma reconnaissance, mon admiration et mon respect éternel.





Introduction Générale

*« Il n'y a rien de plus difficile à prendre en main,
de plus périlleux à conduire, ou de plus incertain dans son succès
, que de prendre les devants dans l'introduction d'un nouvel ordre de choses. »*

— Niccolò Machiavelli. 1469 -1527

Contexte et Motivation

L'agriculture de précision, rendue possible par l'intelligence artificielle (IA) et les approches basées sur les données, constitue une opportunité clé pour le monde académique et l'industrie. Elle optimise les pratiques agricoles grâce aux données et aux outils technologiques, offrant ainsi de nouvelles perspectives en matière de recherche, de financement et d'innovation. Cependant, elle soulève également des défis techniques et logistiques, nécessitant une collaboration interdisciplinaire pour exploiter pleinement son potentiel. Dans ce contexte, les approches basées sur les données sont devenues essentielles en agriculture de précision, offrant des outils pour gérer et optimiser les rendements des cultures majeures comme l'arboriculture. En permettant la spécification précise des besoins et des rendements, ces techniques contribuent à une gestion plus efficace et durable des ressources.

Ces techniques innovantes révolutionnent l'agriculture traditionnelle en utilisant des analyses de données avancées, un suivi précis et des interventions ciblées pour répondre aux besoins spécifiques des cultures (e.g. [12, 4, 69]). En permettant la spécification précise des besoins et des rendements, ces techniques favorisent une gestion plus efficace et durable des ressources [150]. Récemment, l'IA, une approche basée sur les données qui simule l'apprentissage et la prise de décision humaine dans les machines, et l'IoT (Internet des objets), qui connecte les dispositifs pour partager des données en temps réel, se combinent pour améliorer l'agriculture de précision en optimisant les rendements et la gestion des ressources dans les fermes intelligentes [120]. Dans cette dynamique, la variabilité de l'indice de végétation NDVI (Normalized Difference Vegetation Index) et la variabilité climatique (température, précipitations, humidité, etc.) jouent un rôle central. Ces deux dimensions constituent des indicateurs cruciaux pour évaluer l'état des cultures et prédire les rendements. Leur corrélation, souvent mise en évidence dans la littérature, fournit une base pertinente pour construire des modèles prédictifs robustes et contextualisés.

Cadre de questionnement scientifique

Plus spécifiquement, la culture de l'olivier nécessite un contrôle et une surveillance précis. Des algorithmes avancés basés sur les données devraient être utilisés pour l'ajustement des paramètres afin d'améliorer la qualité du produit, d'aligner la production avec la demande du marché, et de réaliser des analyses prédictives. Ces analyses recommandent des améliorations et préviennent les anomalies de production, garantissant ainsi des résultats optimaux dans la production d'olives. Le nord de l'Algérie, région clé pour la production d'huile d'olive, fait face à des défis importants dans la prévision des rendements en raison de la variabilité des conditions environnementales et des pratiques agricoles mises en œuvre¹. Prévoir la production d'huile d'olive dans cette région est essentiel pour améliorer la qualité du produit, adapter la production à la demande du marché, et conduire des analyses prédictives permettant de recommander des améliorations et de prévenir les anomalies de production.

Diverses études menées dans les pays du bassin méditerranéen ont exploré différentes méthodologies et variables pour la prévision des rendements oléicoles, chacune apportant des perspectives uniques tout en se heurtant à des défis spécifiques. Par exemple, [17] a utilisé la régression PLS (Partial Least

1. <https://lecomptoiralgerien.com/collections/huiles-dolive>

Squares) pour développer des modèles de prévision, en prenant la production d'olives comme variable dépendante, et les données météorologiques et aérobiologiques comme variables explicatives. Bien que le modèle ait montré un potentiel intéressant, il a révélé des écarts entre les valeurs observées et celles prédites, dus à la forte variabilité des écosystèmes des cinq zones étudiées. L'étude réalisée par [51] s'est concentrée sur la prévision de la production oléicole en Ombrie, dans le centre de l'Italie, en utilisant un modèle de régression linéaire basé sur des données historiques et des variables multiples. Bien que le modèle ait obtenu un fort pouvoir explicatif, sa précision a diminué au cours des années marquées par des événements climatiques extrêmes. Une autre étude, [54], a développé des modèles fiables de prévision des récoltes d'olives pour la province de Córdoba, en Espagne, en s'appuyant sur 20 ans de données d'émission de pollen, de paramètres météorologiques (température et précipitations), et de données de production. L'étude a proposé trois modèles statistiques permettant de prévoir la production 8, 4 et 2 mois à l'avance. Toutefois, elle a souligné des limites potentielles, telles que l'impact du changement climatique ou les évolutions des pratiques agricoles.

Malgré ces avancées, les modèles existants présentent des succès variables et manquent souvent d'une intégration complète des facteurs pertinents, tels que les maladies, les ravageurs ou la gestion des sols. Il existe un besoin réel de cadres modulaires et adaptables, capables de combiner ces dimensions diverses et de favoriser la réutilisation des connaissances. En effet, la prévision de la production agricole à l'aide d'une fonction de coût permet d'optimiser la planification financière, la gestion des ressources et la prise de décision, tout en favorisant la durabilité et la compétitivité.

Problématique

La fonction de coût constitue une alternative pertinente pour modéliser la production en agriculture. En fournissant une estimation économique de la production, elle permet d'optimiser la planification financière, la gestion des ressources et le processus décisionnel, tout en soutenant les objectifs de durabilité et de compétitivité. Comme nous l'avons mentionné précédemment, la prévision de la production culturelle en agriculture à l'aide d'une fonction de coût optimise la planification financière, la gestion des ressources et la prise de décision, tout en favorisant la durabilité et la compétitivité. Cependant, la conception d'une fonction de coût pour la production en agriculture est une tâche longue et sujette aux erreurs. Premièrement, les systèmes de production culturelle sont souvent complexes, rendant difficile l'automatisation de la collecte des données et la capture de toutes les variables pertinentes. Il faut donc un processus structuré guidé par un protocole rigoureux. Deuxièmement, l'un des plus grands défis est l'identification des paramètres clés et la modélisation des relations entre eux. Ce processus est fondamental pour dériver des modèles prédictifs pertinents. Troisièmement, la variabilité des habitudes de consommation, des cycles de croissance, et des pratiques agricoles accroît les incertitudes et affecte la qualité de la production. Enfin, les modèles doivent continuellement intégrer de nouveaux facteurs pour rester fiables et exploitables.

Le présent travail est guidé par la question de recherche suivante :

RQ : *Comment pouvons-nous aider les concepteurs, tels que les économistes agricoles et les agroécologistes, à créer une fonction de coût efficace pour la prévision de la production agricole ?*

Pour explorer cela, nous nous concentrons sur :

- **RQ1** : Quels sont les facteurs qui influencent les coûts de production culturelle, en tenant compte de dimensions telles que les conditions environnementales (température, humidité, type de sol, etc.) ?
- **RQ2** : Peut-on expliciter les étapes de la modélisation de la production pour les prévisions agricoles ?
- **RQ3** : (À compléter selon le reste du document ou plan prévu...)

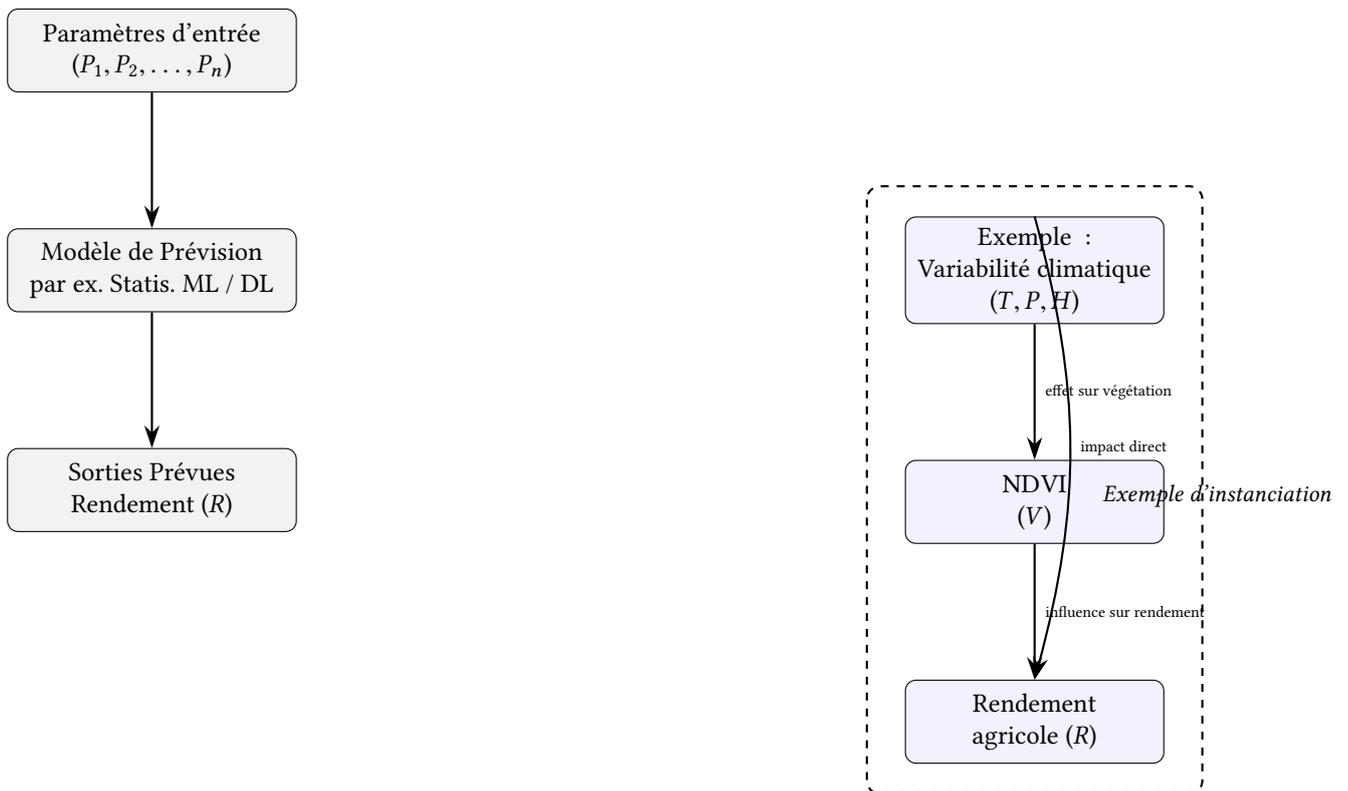


FIGURE 1 – Schéma global illustrant les interactions entre les paramètres d'entrée génériques (P_1, P_2, \dots, P_n), leur traitement via un modèle de prévision, et un encart détaillant l'exemple classique de la variabilité climatique affectant le NDVI et le rendement agricole.

Comme illustré dans la Figure 1, de nombreux travaux ont souligné les corrélations significatives entre la variabilité climatique notamment la température (T), les précipitations (P) et l'humidité (H), la variabilité de l'indice de végétation (NDVI), et les rendements agricoles (R). La partie droite du schéma représente cette configuration typique décrite dans la littérature, où la variabilité climatique agit directement et indirectement (via la végétation) sur la production agricole. Ces interactions ont été largement exploitées dans des modèles de prévision fondés sur l'apprentissage automatique ou l'analyse de séries temporelles. À l'opposé, la partie gauche du schéma traduit notre vision : proposer un cadre méthodologique flexible permettant de modéliser et d'analyser des ensembles de paramètres d'entrée génériques (P_1, P_2, \dots, P_n). Ce cadre vise à guider les concepteurs dans la sélection des variables pertinentes, la structuration des données historiques, et l'instanciation de modèles de prévision adaptés aux spécificités contextuelles. Ainsi, notre contribution repose sur la construction d'un environnement d'aide à la conception, où les chercheurs peuvent explorer de manière systématique l'impact combiné des variables climatiques, environnementales et végétales sur le rendement, afin d'élaborer des modèles

robustes, personnalisables et interprétables pour le *forecasting* agricole.

Cette vision nous a motivés à penser la généricité dans la modélisation, et à concevoir un *framework* adaptable. Ce framework est instancié à travers plusieurs scénarios d'usage, au-delà du seul cas de la prédiction des rendements.

Contributions

Dans cette thèse, nous présentons un cadre conceptuel appelé AFFAC (Advanced Forecasting Framework for Agricultural Crops), conçu pour prédire la production agricole en intégrant une gamme complète de paramètres environnementaux. Ces paramètres incluent les données météorologiques, la qualité du sol, l'utilisation de l'eau pour l'irrigation et les pratiques culturales spécifiques aux cultures. Tout d'abord, le cadre met en œuvre un processus systématique de collecte et de gestion des données relatives à la production agricole en agriculture, en utilisant diverses sources de données et processus. Ensuite, il intègre le savoir-faire des experts pour sélectionner les variables d'entrée pertinentes aux aspects les plus sensibles de la production agricole. Troisièmement, AFFAC est hautement personnalisable, permettant aux utilisateurs d'explorer différents modèles de prévision pour des tâches de conception spécifiques en attachant facilement les paramètres d'entrée pertinents à la spécification du modèle. Tout d'abord, le cadre met en œuvre un processus systématique de collecte et de gestion des données relatives à la production agricole, en utilisant diverses sources de données et processus. Ensuite, il intègre le savoir-faire des experts pour sélectionner les variables d'entrée pertinentes aux aspects les plus sensibles de la production agricole. Troisièmement, AFFAC est hautement personnalisable, permettant aux utilisateurs d'explorer différents modèles de prévision pour des tâches de conception spécifiques en attachant facilement les paramètres d'entrée pertinents à la spécification du modèle.

Les objectifs de recherche de ce travail sont les suivants :

1. *Proposition d'un cadre conceptuel axé sur les données* : Proposer un cadre conceptuel basé sur les données pour la conception de fonctions de coût, afin d'améliorer la gestion de la production agricole et de soutenir la prise de décision dans le secteur de l'agriculture.
2. *Développement de modèles de prévision de conception* : Développer un ensemble de modèles de prévision pour assister les concepteurs dans le processus de conception des fonctions de production, contribuant ainsi à la gestion des cultures, spécifiquement pour la culture de l'olivier en Algérie.
3. *Instanciation dans une étude de cas sur la culture de l'olivier* : Appliquer le cadre conceptuel à une étude de cas centrée sur la culture de l'olivier en Algérie, afin de démontrer son application pratique et sa pertinence dans le contexte local.
4. *Validation du cadre* : Valider le cadre conceptuel proposé à travers une étude de cas sur la culture de l'olivier, afin de démontrer son efficacité et son potentiel d'amélioration dans la prévision des rendements agricoles et le soutien aux producteurs agricoles.

Le modèle AFFAC vise à capturer et à décrire les interactions complexes entre ces variables pour fournir des prévisions précises des rendements agricoles. Notre cadre utilise des techniques d'apprentissage profond et d'apprentissage automatique, permettant le traitement de grands ensembles de données et l'extraction de motifs significatifs. Pour valider notre modèle, nous avons appliqué notre cadre aux

cultures d'olives, qui ont une valeur économique substantielle et de nombreuses applications, les rendant cruciales pour de nombreuses économies. En utilisant des données réelles provenant de ces cultures, nous avons mené des expériences pour évaluer la précision de nos prédictions. Cette étude de cas démontre non seulement la faisabilité de AFFAC, mais met également en lumière les domaines dans lesquels le modèle pourrait être amélioré ou ajusté pour mieux répondre aux besoins spécifiques des producteurs agricoles.

Les contributions de ce travail sont résumées comme suit :

1. Proposition d'un cadre conceptuel axé sur les données : Nous introduisons un cadre conceptuel basé sur les données conçu pour développer des fonctions de coût pour la production culturelle en agriculture.
2. Développement de modèles de prévision : Nous créons un ensemble de modèles de prévision pour aider les concepteurs dans le processus de conception des fonctions de production. Cela contribue au domaine plus large de la gestion des cultures, spécifiquement adapté à la culture de l'olivier, qui est une culture stratégique en Algérie.
3. Instanciation dans une étude de cas sur la culture de l'olivier : Nous appliquons notre cadre à une étude de cas axée sur la culture de l'olivier en Algérie, démontrant son application pratique et sa pertinence.
4. Validation du framework : Nous validons notre framework à travers l'étude de cas, montrant son efficacité et son potentiel d'amélioration dans la prévision des rendements agricoles et le soutien aux producteurs agricoles.

Notre Méthodologie

La Figure 2 illustre la méthodologie suivie pour parfaire notre travail de recherche. Nous adoptons une approche de conception basée sur les principes du génie logiciel pour développer un framework. Le cycle de conception comporte plusieurs étapes majeures (voir la Figure 2) : d'abord, l'identification du problème de recherche, suivie de la proposition de solutions potentielles, puis de la conception de la solution et de ses fondements théoriques, ensuite du développement et de la mise en œuvre de la solution, et enfin, des conclusions. Le cycle commence par l'identification du problème et des lacunes de recherche à travers une revue de la littérature. Le problème est formulé et accompagné de questions de recherche pertinentes. Les solutions proposées pour résoudre ce problème sont définies à partir de l'état de l'art, afin de guider notre contribution originale. L'analyse de la littérature se concentre sur les travaux relatifs aux Modèles de Conception pour la prévision et la production agricole. Lors de la conception de la solution, nous présentons un framework méthodologique qui assure la réutilisation, l'automatisation et la personnalisation des modèles de coûts pour les bases de données. Enfin, dans la phase de validation, nous démontrons la faisabilité et la praticabilité de notre proposition, avant de tirer les conclusions appropriées.

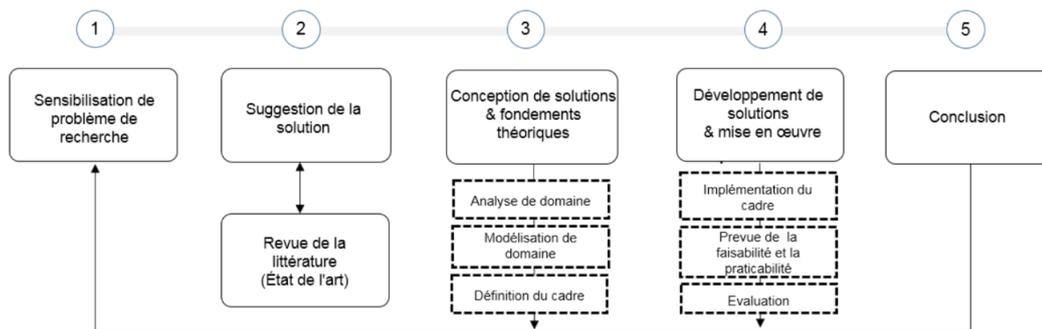


FIGURE 2 – Méthodologie adoptée dans cette thèse

Organisation de thèse

La thèse que nous défendons s’articule autour de deux contributions majeures présentées dans le chapitre 3 et 4, Le chapitre 1 et 2 sont dédiés à l’étude de l’état de l’art.

- *Le chapitre 1* décrit un état de l’art sur la transformation des modèles orientée vers la qualité de service.
- *Le chapitre 2* présente un état de l’art sur les approches et les techniques de modélisation de la fonction de coût de production agricole, ainsi que les principaux concepts et approches de modélisation. Il décrit les techniques de modélisation que nous avons retenues pour cerner notre domaine d’étude. Il vise à déterminer les concepts les plus communs pour décrire un framework d’étude assez générique pour la modélisation et l’évaluation des prévisions dans les systèmes de production agricole. Les chapitres suivants concernent une description de notre framework fondé sur les modèles de l’apprentissage machine.
- *Le chapitre 3* présente la description conceptuelle de notre framework AFFAC pour la modélisation de la fonction de coût de production des rendements agricoles. Notre modèle est basé sur un framework théorique qui repose sur les données historiques. D’abord, nous avons commencé par la description de chaque dimension de l’architecture AFFAC à l’aide des techniques issues de l’apprentissage machine. Une fois construit, cela nous a motivé à proposer une formalisation pour ces dimensions. La deuxième partie de ce chapitre est dédiée à la description de notre framework pour la modélisation et l’évaluation des prévisions dans les systèmes de production agricole, en se centrant sur l’approche de construction d’une fonction de coût de production, en tirant parti des modèles avancés d’apprentissage machine (ML) et d’apprentissage profond (DL) pour optimiser la précision et l’adaptabilité des prévisions des coûts de production agricoles.
- *Le chapitre 4* décrit l’instanciation dans l’étude de cas de la culture de l’olivier : nous appliquons notre framework à une étude de cas centrée sur la culture de l’olivier en Algérie, démontrant son application pratique et sa pertinence.
- *Le chapitre 5* est consacré à la validation de notre framework : nous validons notre framework à travers l’étude de cas, mettant en avant son efficacité et son potentiel d’amélioration dans la prévision des rendements agricoles et le soutien aux producteurs agricoles.
- Enfin, nous présentons une synthèse de nos contributions, ainsi que les limites et les perspec-

tives de notre proposition.

Première partie

État de l'art

Analyse et Impact de la Variabilité Climatique



«Le climat est ce que nous attendons, le temps est ce que nous obtenons.»

— Mark Twain

Sommaire

1.1	Introduction	18
1.2	Le climat	18
1.2.1	Notion d'échelle en climatologie	18
1.2.2	Le système climatique	21
1.3	Les caractéristiques climatiques de l'Algérie	23
1.3.1	Saisonnalité et influence de la mer Méditerranéenne	23
1.3.2	Les températures moyennes	24
1.4	Les indices climatiques	26
1.4.1	Indices de température	26
1.4.2	Cumul annuel des précipitations	28
1.5	La variabilité climatique	28
1.5.1	Définition de la variabilité climatique	28
1.5.2	Les caractéristiques de la variabilité climatique	29
1.5.3	Facteurs de la variabilité climatique	30
1.5.4	Les tendances de la variabilité et du changement climatique	33
1.6	Impact sur l'agriculture	34
1.6.1	Variabilité des rendements agricoles	34
1.6.2	Besoins en irrigation et en adaptation des cultures	34
1.7	Mesures et indicateurs de la variabilité climatique	35
1.7.1	Indices climatiques	35
1.7.2	Modèles de prédiction climatique	36
1.8	Moyens d'acquisition des paramètres de variabilité climatique	37
1.8.1	Architecture de infrastructure	38
1.8.2	Architecture du système décisionnel	40
1.9	Conclusion	41

1.1 Introduction

Dans le cadre de l'étude du climat, il est essentiel de considérer plusieurs échelles d'analyse. Celles-ci permettent de mieux comprendre les phénomènes climatiques à différents niveaux, qu'ils soient globaux, régionaux ou locaux [128]. Nous avons choisi d'illustrer cette hiérarchie des échelles à l'aide d'un diagramme mettant en évidence les relations d'inclusion entre ces différentes échelles. Cette approche permet de visualiser la manière dont les phénomènes à grande échelle, tels que les changements climatiques globaux, influencent les événements à échelles plus petites, comme les variations météorologiques locales. Nous allons également aborder l'échelle temporelle, qui joue un rôle clé dans la compréhension des processus climatiques [116]. En effet, les phénomènes climatiques se manifestent sur des durées variées, allant de quelques jours à plusieurs siècles. Cette diversité temporelle est cruciale pour la planification de stratégies d'adaptation et de prévention face aux risques climatiques [73]. Dans cette optique, l'analyse conjointe de la variabilité climatique et de la variabilité de l'indice de végétation par différence normalisée (NDVI) permet d'approfondir la compréhension des dynamiques environnementales à différentes échelles. Le NDVI, sensible à l'état de la végétation, reflète directement l'influence des conditions climatiques sur les écosystèmes agricoles. Leur corrélation est essentielle pour interpréter les réponses des cultures aux stress environnementaux et pour améliorer les modèles prédictifs liés à la production agricole. Dans cette section, nous allons donc détailler ces deux aspects fondamentaux du climat : les échelles spatiale et temporelle. Nous introduirons ensuite le concept du système climatique, qui englobe l'ensemble des composants interconnectés de la Terre. En explorant la dynamique de ces composantes et leurs interactions, nous chercherons à mieux comprendre les mécanismes du changement climatique et les effets des activités humaines sur le climat mondial et régional [56].

1.2 Le climat

Le climat, caractérisé par sa variabilité, joue un rôle essentiel en agriculture, où il doit être pris en compte pour prévoir les rendements, anticiper les risques et mieux gérer les ressources. Dans cette section, nous explorons les notions d'échelle en climatologie, en distinguant les dimensions spatiales et temporelles, puis nous abordons le système climatique en définissant ses composantes principales et les interactions complexes qui les relient [25, 113].

L'Algorithme 1.1 montre une analyse des données climatiques (T , P , H) en quatre étapes : prétraitement (correction et normalisation), calcul d'indicateurs (moyenne, variance, indices de précipitation), analyse des tendances temporelles avec un test statistique (ex. Mann-Kendall), et visualisation des résultats. Il permet de détecter et représenter les variations climatiques significatives sur une période donnée.

1.2.1 Notion d'échelle en climatologie

En climatologie, il est crucial de comprendre les différentes échelles spatiales et temporelles utilisées pour analyser les phénomènes climatiques

Définition 1 (Notion d'échelle en climatologie)

| La notion d'échelle en climatologie fait référence aux dimensions spatiales et temporelles utilisées pour |

Algorithme 1.1 Analyse de la variation climatique

Entrée : Données climatiques $D = \{T, P, H\}$ où T : températures, P : précipitations, H : humidité.

Sortie : Indicateurs de variation climatique.

- 1: **Étape 1 : Prétraitement des données**
- 2: **pour tout** $d \in D$ **faire**
- 3: Vérifier et corriger les valeurs manquantes dans d .
- 4: Normaliser les données pour garantir une échelle uniforme.
- 5: **fin pour**
- 6: **Étape 2 : Calcul des indicateurs climatiques**
- 7: Moyenne $_T \leftarrow$ Calcul de la moyenne des températures
- 8: Variance $_T \leftarrow$ Variance des températures
- 9: Indice $_P \leftarrow$ Indice de précipitation annuelle
- 10: **Étape 3 : Analyse des tendances temporelles**
- 11: **pour** $t = 1$ à n (nombre d'années) **faire**
- 12: Identifier les variations annuelles $\Delta T, \Delta P, \Delta H$.
- 13: Vérifier l'existence de changements significatifs (ex. *Mann-Kendall Test*).
- 14: **fin pour**
- 15: **Étape 4 : Visualisation des résultats**
- 16: Générer des graphiques pour représenter les tendances (températures, précipitations, humidité).
- 17: **retourner** Résultats des variations climatiques et conclusions.

étudier et analyser les phénomènes climatiques[14]. Elle permet de distinguer les processus locaux (microscale), régionaux (mesoscale) et globaux (macroscale), ainsi que les variations climatiques sur des périodes courtes (jours ou semaines) ou longues (années, décennies, voire siècles). Il est essentiel de distinguer les processus opérant à différentes échelles. Les processus locaux, dits microscale, incluent des phénomènes comme les brises de vallée ou les microclimats urbains. Les processus régionaux, ou mesoscale, englobent les systèmes météorologiques tels que les fronts ou les cyclones régionaux. Enfin, les processus globaux, ou macroscale, concernent des phénomènes à l'échelle planétaire, comme les circulations atmosphériques globales ou les oscillations climatiques (ENSO, NAO).[56]

1.2.1.1 Échelle spatiale

Lors des études climatiques, il est indispensable de considérer la portée géographique des différents paramètres climatiques. Certains éléments présentent des variations significatives à petite échelle, comme les précipitations au sein d'une cellule orageuse ou les variations de température en fonction de l'altitude [6]. Par conséquent, la définition du climat doit être adaptée en fonction de l'échelle d'analyse utilisée[84]. À l'échelle continentale, on parle de macroclimat, caractérisé par une étendue géographique de 100 à 10 000 km. Le mésoclimat correspond à l'échelle régionale ou locale, généralement comprise entre 1 et 100 km, bien que cette notion soit parfois confondue avec celle de microclimat dans le langage courant. Enfin, le microclimat désigne les phénomènes climatiques se manifestant sur de très courtes distances, de l'ordre de 0 à 1000 m [6, 13].

Contexte géographique : Le paragraphe situe la région et met en avant sa *diversité géographique* et climatique. Influence de l'altitude : Intégration du gradient thermique et des effets orographiques

spécifiques au nord de la Méditerranée. Lien échelle-spatial/climatologie : Transition entre les échelles régionale et locale pour expliquer les variations climatiques. Précision scientifique : Termes comme "ombre orographique" et "gradient thermique" pour une approche rigoureuse.

Exemple :

- **Échelle locale** : L'étude du microclimat dans une vallée ou autour d'une zone agricole.
- **Échelle régionale** : L'analyse des variations climatiques dans une région méditerranéenne.
- **Échelle globale** : L'observation des changements climatiques affectant l'ensemble de la planète, comme le réchauffement climatique.

La Figure 1.1 illustre l'inclusion des échelles spatiales en climatologie, où l'échelle locale est incluse dans l'échelle régionale, elle-même incluse dans l'échelle globale. Ce diagramme montre la hiérarchie entre les phénomènes climatiques à différentes échelles, du local au global.

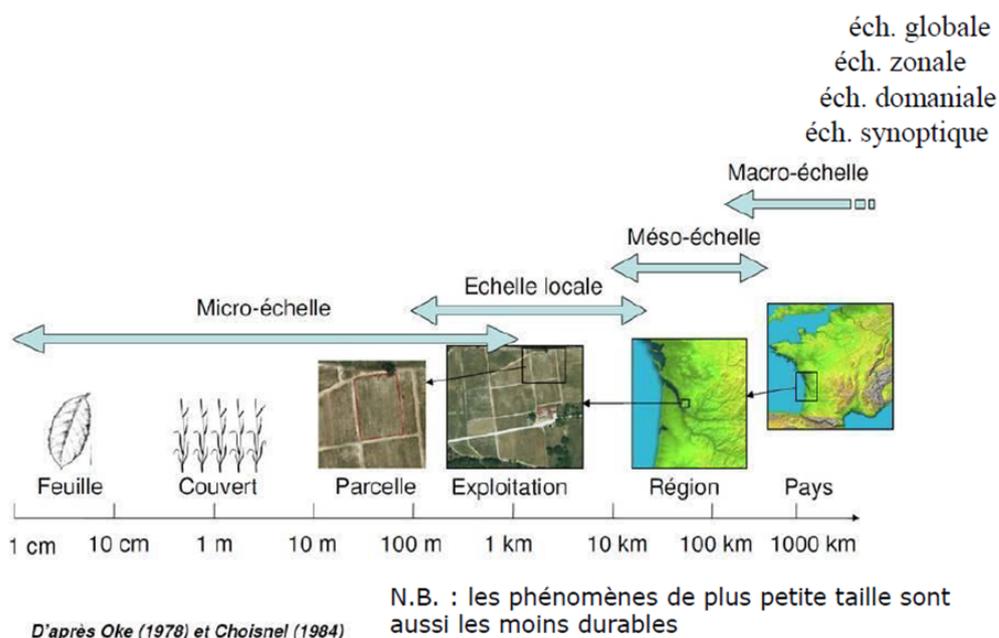


FIGURE 1.1 – Inclusion entre les échelles locale, régionale et globale en climatologie

1.2.1.2 Échelle temporelle

L'échelle temporelle en climatologie représente les conditions météorologiques avec une grande variabilité et changent fréquemment au fil des heures, des jours et des semaines des phénomènes climatiques observés [6]. Cette variabilité résulte de plusieurs facteurs naturels. La première est liée à la rotation de la Terre, qui entraîne l'alternance du jour et de la nuit, responsable des fluctuations diurnes de température. La seconde correspond à la variabilité saisonnière, causée par l'inclinaison de l'axe terrestre. L'inclinaison des rayons solaires incident en pour un point donné, n'est pas la même en été qu'en hiver, de même que la durée du jour et de la nuit change durant les saisons et de ce fait, il en résulte des températures plus basses en hiver qu'en été. Pour le météorologiste les saisons ne coïncident

pas avec les saisons astronomiques. En effet le maximum de chaleur ne se produit pas au moment du solstice d'été mais aux alentours de la mi-juillet[13].

Le Tableau 1.1 illustre quelques exemples associés à chaque échelle :

Type d'échelle temporelle	Exemple
Courte durée (journées à semaines)	Analyse des vagues de chaleur, des tempêtes ou des précipitations intenses.
Moyenne durée (mois à années)	Étude des anomalies climatiques comme El Niño ou La Niña, ou des variations saisonnières.
Longue durée (décennies à siècles)	Observation des changements climatiques globaux, comme le réchauffement climatique ou les variations glaciaires.

Tableau 1.1 – Exemples d'échelles temporelles en climatologie

L'échelle temporelle est fondamentale en climatologie pour analyser les phénomènes climatiques sur des durées variées. Elle permet d'étudier des événements ponctuels tels que les vagues de chaleur ou les tempêtes, ainsi que des tendances à long terme comme le réchauffement climatique. Cette distinction temporelle aide à mieux comprendre les dynamiques climatiques et à orienter les stratégies d'adaptation face aux changements climatiques.

1.2.2 Le système climatique

Le système climatique est composé de plusieurs éléments interconnectés qui interagissent pour réguler les conditions climatiques de la Terre. Dans cette section, nous présentons la relation de cause à effet dans le système climatique, en illustrant l'impact de l'augmentation des gaz à effet de serre sur le réchauffement climatique.

1.2.2.1 Définition du système climatique

Le système climatique représente l'ensemble des composantes principales de la Terre, notamment l'atmosphère, l'hydrosphère, la cryosphère, la biosphère et la lithosphère, qui interagissent de manière dynamique pour réguler le climat [147, 15]. Ce système complexe est soumis à l'influence de facteurs naturels, tels que les variations de l'activité solaire et les éruptions volcaniques, mais également à des facteurs anthropiques, comme les émissions de gaz à effet de serre et les modifications de l'usage des sols [78]. L'évolution du système climatique résulte ainsi d'interactions étroites et parfois non linéaires entre ces différentes composantes, ce qui rend son étude fondamentale pour comprendre les mécanismes des changements climatiques [56].

Les conditions météorologiques sont caractérisées par une forte variabilité qui se manifeste à différentes échelles de temps, changeant d'heure en heure, de jour en jour et de semaine en semaine [5]. Cette variabilité est principalement influencée par deux phénomènes naturels. Le premier est lié à la rotation de la Terre, qui engendre l'alternance entre le jour et la nuit. Ce cycle quotidien est à l'origine des variations diurnes de température. Le second phénomène est la variabilité saisonnière, causée par l'inclinaison de l'axe terrestre. En effet, l'angle d'incidence des rayons solaires diffère selon les saisons, ainsi que la durée du jour et de la nuit. Cela explique pourquoi les températures sont généralement plus

élevées en été, lorsque les rayons solaires sont plus perpendiculaires, et plus basses en hiver, lorsque l'angle est plus faible. Par ailleurs, pour les météorologues, les saisons ne correspondent pas parfaitement aux saisons astronomiques. Par exemple, le pic de chaleur ne survient pas au solstice d'été mais plutôt vers la mi-juillet, en raison de l'inertie thermique du système climatique [52].

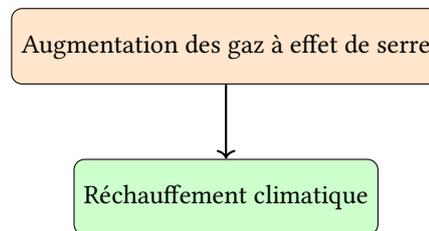


FIGURE 1.2 – Relation cause-effet dans le système climatique

1.2.2.2 Les composantes du système climatique

L'interaction entre les systèmes climatiques, comprenant l'atmosphère, l'hydrosphère, la cryosphère, la lithosphère, la biosphère et l'anthroposphère. Les flèches indiquent les relations et les interactions entre ces éléments essentiels pour comprendre le fonctionnement global du climat.

- **Atmosphère** : C'est la couche de gaz qui entoure la Terre. Elle joue un rôle clé dans la régulation de la température et dans la circulation de l'air et des précipitations.
- **Lithosphère** : Ce composant fait référence aux terres émergées, aux sols et aux roches. La lithosphère joue un rôle dans les échanges de chaleur et la régulation des écosystèmes, influençant ainsi le climat local et global.
- **Biosphère** : C'est l'ensemble des organismes vivants, y compris les végétaux et les animaux. Les processus biologiques (comme la photosynthèse et la respiration) influencent les niveaux de CO₂ et d'autres gaz à effet de serre dans l'atmosphère.

1.2.2.3 Interaction entre les composantes du système climatique

Dans l'étude du système climatique, il est essentiel de comprendre les interactions complexes qui existent entre ses différentes composantes. Ce système peut être divisé en six grandes entités interconnectées : l'atmosphère, l'hydrosphère, la cryosphère, la lithosphère, la biosphère et l'anthroposphère. Chacune de ces composantes joue un rôle fondamental dans les dynamiques climatiques, influençant et étant influencée par les autres. Par exemple, l'atmosphère interagit directement avec l'hydrosphère à travers des phénomènes tels que l'évaporation, les précipitations et la circulation océanique. De même, les activités humaines, regroupées dans l'anthroposphère, modifient considérablement ces interactions naturelles, notamment via les émissions de gaz à effet de serre, l'altération des sols et la destruction des écosystèmes.[141]

La Figure 1.3 illustre ces relations bidirectionnelles entre les composantes du système climatique, mettant en lumière leur interdépendance et leur rôle dans l'équilibre global. Les flèches entre les différents éléments montrent comment ces composantes interagissent et s'influencent mutuellement, créant des boucles de rétroaction parfois positives, parfois négatives. Une telle modélisation systémique est indis-

pensable pour mieux appréhender les mécanismes sous-jacents aux phénomènes climatiques et anticiper les impacts des changements environnementaux, qu'ils soient d'origine naturelle ou anthropique.

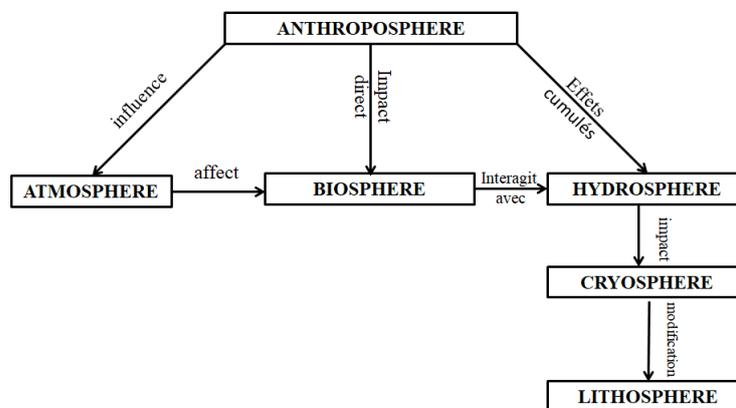


FIGURE 1.3 – Interaction entre les composantes du système climatique

1.3 Les caractéristiques climatiques de l'Algérie

1.3.1 Saisonnalité et influence de la mer Méditerranéenne

La mer Méditerranée est caractérisée par une saisonnalité marquée, influençant fortement les régimes de précipitations et de températures[118]. En hiver, les températures plus basses et les précipitations augmentées, souvent sous forme d'épisodes orageux, favorisent un apport d'eau douce dans le bassin. Ces variations saisonnières impactent non seulement la circulation des courants marins, mais aussi les écosystèmes côtiers et les activités humaines telles que la pêche et le tourisme. La dynamique climatique de cette région est donc un équilibre délicat entre les influences saisonnières de la température et de la pluviométrie. [102, 118]

1.3.1.1 Répartition des précipitations par région

La répartition des précipitations autour de la mer Méditerranée varie considérablement selon les régions, influencée par des facteurs géographiques et climatiques. Les côtes nord, notamment en Espagne, en France et en Italie, reçoivent des précipitations plus abondantes, en particulier durant l'automne et l'hiver, en raison des perturbations venant de l'Atlantique [102]. À l'inverse, les régions situées au sud, comme l'Afrique du Nord, connaissent un climat beaucoup plus aride, avec des précipitations rares et souvent concentrées sur de courtes périodes [122]. L'est méditerranéen, incluant des pays comme la Grèce et la Turquie, reçoit des précipitations modérées, souvent sous forme de pluies orageuses saisonnières [165]. Ces disparités jouent un rôle crucial dans l'équilibre des écosystèmes locaux et l'accès aux ressources en eau dans la région.

1.3.1.1.1 Variabilité interannuelle des précipitations

La variabilité interannuelle des précipitations dans le bassin méditerranéen est un phénomène marquant, influencé par des facteurs climatiques globaux et régionaux. Les années humides alternent avec des années de sécheresse, souvent liées à des phénomènes climatiques tels que l'Oscillation Nord-Atlantique (NAO) et El Niño [148, 131]. Cette variabilité est particulièrement prononcée dans les régions arides du sud, comme l'Afrique du Nord, où une faible variation des précipitations peut avoir des impacts significatifs sur les ressources en eau et l'agriculture [122]. Les régions du nord, bien que moins sensibles à ces fluctuations, peuvent également connaître des événements extrêmes, tels que des inondations ou des sécheresses prolongées [19]. Comprendre cette variabilité est essentiel pour anticiper ses effets sur les écosystèmes et les populations humaines[57].

1.3.1.1.2 Variabilité saisonnière des précipitations

La variabilité saisonnière des précipitations dans le bassin méditerranéen est fortement marquée, reflétant l'influence du climat méditerranéen typique. En hiver, les précipitations sont généralement abondantes, particulièrement dans les régions nord et ouest, en raison des perturbations atlantiques et des systèmes dépressionnaires[102]. En été, en revanche, la majorité des zones connaît une quasi-absence de pluies, avec des périodes prolongées de sécheresse, sauf dans certaines régions montagneuses où des orages locaux peuvent survenir[148]. Le printemps et l'automne jouent un rôle de transition, avec des pluies souvent intenses et localisées, particulièrement durant l'automne, où des épisodes méditerranéens peuvent provoquer de fortes inondations[47]. Cette variabilité saisonnière conditionne les ressources en eau, les pratiques agricoles et les écosystèmes de la région[57].

1.3.2 Les températures moyennes

Dans le contexte climatique, la température moyenne d'une période donnée peut être calculée à partir des températures quotidiennes ou horaires[88]. La formule de la température moyenne pour une période donnée (par exemple, n jours) est exprimée comme suit :

$$T_{\text{moyenne}} = \frac{1}{n} \sum_{i=1}^n T_i$$

où :

- T_{moyenne} est la température moyenne sur la période.
- T_i est la température mesurée à chaque instant i (par exemple, température quotidienne ou horaire).
- n est le nombre total de jours (ou d'instant) dans la période étudiée.

1.3.2.1 tendance des augmentation des températures moyennes

Dans cette section, nous montrons la tendance d'augmentation des températures moyennes sur plusieurs années. Le graphique suivant illustre cette évolution.

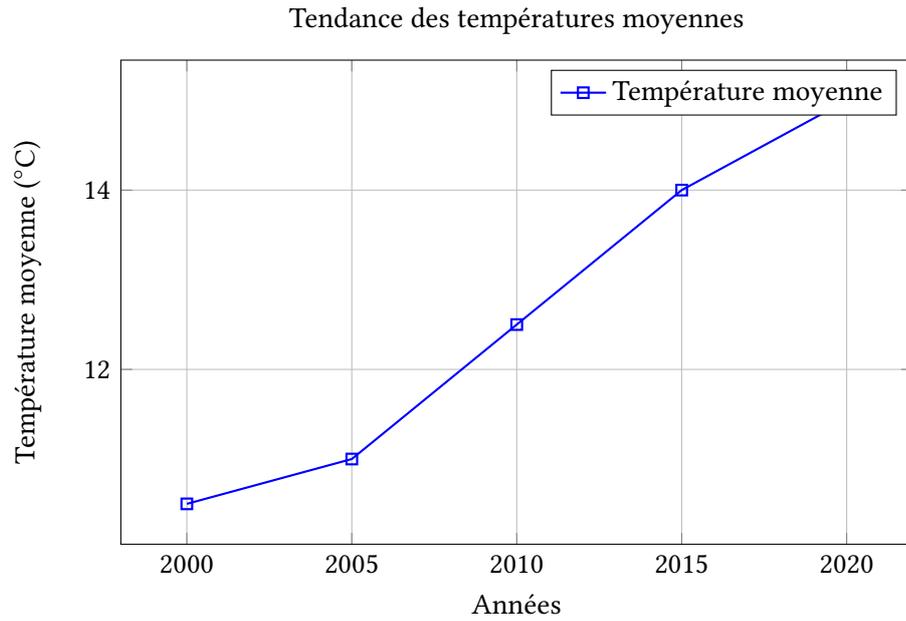


FIGURE 1.4 – Évolution des températures moyennes entre 2000 et 2020.

La 1.5 illustre l'évolution des températures quotidiennes sur une période de 10 jours, en affichant les valeurs moyennes (courbe bleue), maximales extrêmes (courbe rouge) et minimales extrêmes (courbe verte). Elle met en évidence les fluctuations de température et les écarts entre les températures extrêmes observées.

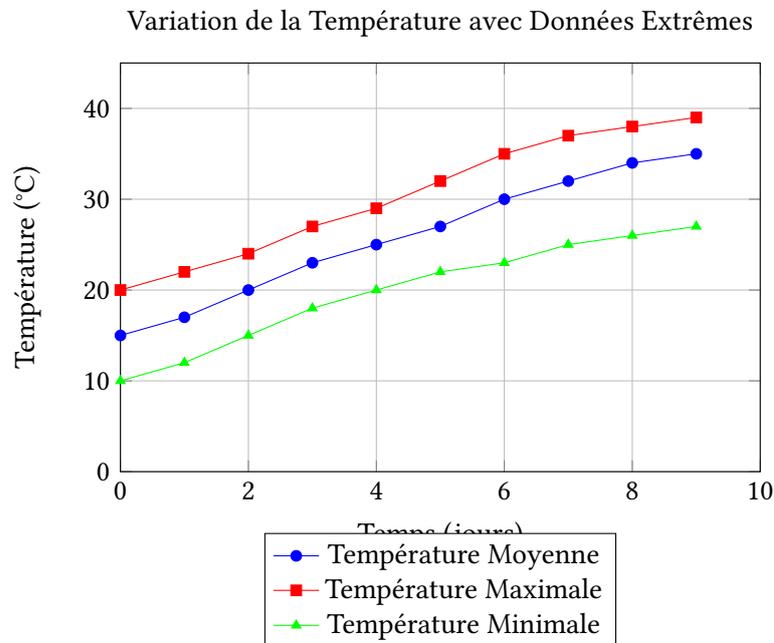


FIGURE 1.5 – Graphique montrant les températures moyennes, maximales et minimales sur une période de 10 jours.

1.4 Les indices climatiques

Cette section examine les tendances du changement climatique en utilisant des règles logiques pour modéliser l'interaction des variables climatiques, telles que la température et les précipitations. Ces indices permettent d'analyser la dynamique climatique et de prévoir les phénomènes extrêmes.

1.4.1 Indices de température

Nous explorons ici la littérature et les travaux de recherche sur les indices de température, qui reposent sur des paramètres climatiques clés. Ces indices servent à évaluer les variations de température, et à comprendre les impacts du changement climatique sur différents écosystèmes et phénomènes environnementaux.

Exemple 1. Exemple : Indices de Température dans un Contexte IoT

Imaginons un capteur IoT mesurant la température pour surveiller le stress thermique dans une zone agricole. L'indice de stress thermique (IST) peut être défini comme suit :

$$IST = T_{air} + H$$

où T_{air} est la température en °C et H l'humidité relative en %. Un IST plus élevé indique un stress thermique accru.

L'indice de température moyen sur 24 heures est calculé à partir des relevés horaires T_i comme suit :

$$T_{moyenne} = \frac{1}{24} \sum_{i=1}^{24} T_i$$

La précision du capteur, par exemple, $\pm 0.5^\circ\text{C}$, affecte directement l'exactitude de ces indices.

1.4.1.1 Nombre de jours de gel et de chaleur extrême

Les jours de gel sont définis comme les jours où la température minimale descend en dessous de 0°C , tandis que les jours de chaleur extrême sont ceux où la température dépasse un seuil de chaleur, par exemple 35°C . Ces journées peuvent avoir un impact significatif sur l'agriculture, affectant les récoltes et la croissance des cultures. Le calcul de ces jours est essentiel pour prévoir les périodes à risque.

L'algorithme suivant montre comment déterminer le nombre de jours de gel et de chaleur extrême à partir d'un ensemble de données de température.

Algorithme : Calcul des jours de gel et de chaleur extrême**Algorithme 1.2** Calcul des jours de gel et de chaleur extrême

```

1: Entrée : Liste des températures journalières  $T = [t_1, t_2, \dots, t_n]$ 
2: Sortie : Nombre de jours de gel et de chaleur extrême
3:  $jours\_gel \leftarrow 0$ 
4:  $jours\_chaleur \leftarrow 0$ 
5: pour chaque température  $t_i \in T$  faire
6:   si  $t_i < 0$  alors
7:      $jours\_gel \leftarrow jours\_gel + 1$ 
8:   sinon si  $t_i > 35$  alors
9:      $jours\_chaleur \leftarrow jours\_chaleur + 1$ 
10:  fin si
11: fin pour
12: Afficher "Jours de gel : ",  $jours\_gel$ 
13: Afficher "Jours de chaleur extrême : ",  $jours\_chaleur$ 

```

Exemple :

Considérons les températures journalières suivantes observées à Tiaret, dans une parcelle agricole, sur une période de 7 jours :

$T = [2, -1, 38, 15, -3, 40, 28]$

- Pour $t_1 = 2^\circ\text{C}$, la température est supérieure à 0 et inférieure à 35, donc aucune action n'est prise.
- Pour $t_2 = -1^\circ\text{C}$, la température est inférieure à 0, donc on incrémente le compteur $jours_gel$.
- Pour $t_3 = 38^\circ\text{C}$, la température est supérieure à 35, donc on incrémente le compteur $jours_chaleur$.
- Pour $t_4 = 15^\circ\text{C}$, la température est dans la plage normale, aucune action n'est prise.
- Pour $t_5 = -3^\circ\text{C}$, la température est inférieure à 0, donc on incrémente le compteur $jours_gel$.
- Pour $t_6 = 40^\circ\text{C}$, la température est supérieure à 35, donc on incrémente le compteur $jours_chaleur$.
- Pour $t_7 = 28^\circ\text{C}$, la température est dans la plage normale, aucune action n'est prise.

Résultat : Après avoir parcouru toutes les températures, on obtient :

$$jours_gel = 2, \quad jours_chaleur = 2$$

Ainsi, il y a 2 jours de gel et 2 jours de chaleur extrême dans cette parcelle à Tiaret, sur la période observée.

1.4.2 Cumul annuel des précipitations

La Figure 1.6 montre le cumul des précipitations sur 12 mois, avec des pics en été (août) et une baisse en hiver (décembre). Cela indique une saison de sécheresse estivale et des mois plus humides au printemps et à l'automne, influençant la gestion des ressources en eau.

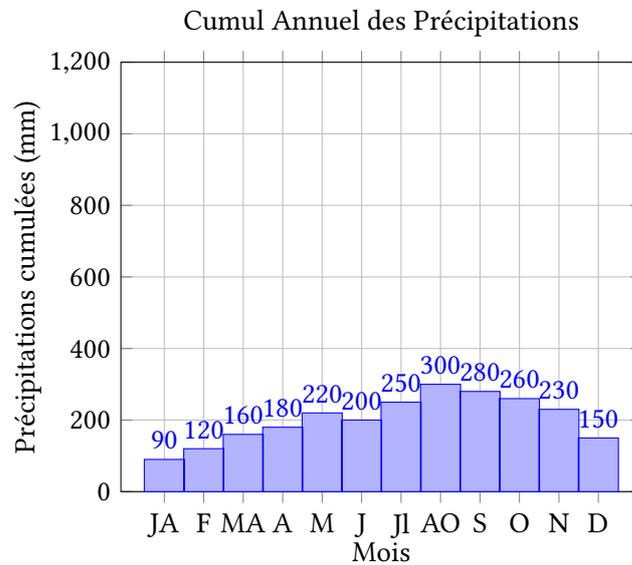


FIGURE 1.6 – Histogramme du cumul annuel des précipitations sur 12 mois.

1.4.2.1 Nombre de jours de pluie

Nous explorons la littérature et les travaux de recherche sur les indices de température et le nombre de jours de pluie, qui reposent sur des paramètres climatiques clés. Ces indices permettent d'évaluer les variations de température et de précipitations [146], ainsi que d'analyser les impacts du changement climatique sur les écosystèmes et les phénomènes environnementaux.[125]

1.5 La variabilité climatique

1.5.1 Définition de la variabilité climatique

En explorant la littérature, il existe plusieurs définitions de la variabilité climatique. Ces définitions sont orientées selon le contexte et les caractéristiques étudiées.

Définition 2 (*Changement de variabilité*)

Le changement de variabilité fait référence à la variation dans les niveaux de dispersion ou de fluctuation d'une donnée au fil du temps ou des conditions. Ce phénomène peut être observé dans des processus stochastiques ou dans des systèmes où les paramètres sous-jacents changent, influençant ainsi la mesure de la variabilité. En termes simples, il s'agit de l'évolution de la variabilité d'un système ou d'un phénomène, qui peut se produire sous l'effet de facteurs internes ou externes.

1.5.2 Les caractéristiques de la variabilité climatique

Nous étudions ici les principales caractéristiques de la variabilité climatique, en analysant des indices comme la température, les précipitations et l'humidité, afin de mieux comprendre les fluctuations climatiques et leur impact environnemental.

Température : La température est l'un des principaux indicateurs de la variabilité climatique. Elle peut fluctuer à des échelles de temps différentes, de l'année à l'échelle décennale, affectant ainsi les écosystèmes. Les vagues de chaleur et les froids extrêmes sont des exemples de phénomènes importants liés à cette variabilité.

Vitesse du vent : Le vent est une composante dynamique du climat, particulièrement influent dans la propagation des incendies de forêt et les tempêtes. Il résulte de la différence de pression entre deux zones.

Précipitations : Les précipitations, qu'elles soient sous forme de pluie, neige ou grêle, jouent un rôle clé dans le climat local et global. Les changements dans les schémas de précipitations peuvent avoir un impact significatif sur la gestion de l'eau, l'agriculture, et la biodiversité.

Humidité : L'humidité affecte la perception de la température et la formation des nuages. Les variations de l'humidité relative influencent la probabilité de formation de pluie et la sévérité des phénomènes climatiques.

Rayonnement solaire : Le rayonnement solaire est essentiel pour l'évaluation de la chaleur reçue par la Terre. Il influe directement sur la température et, par conséquent, sur le climat régional et global.

Variabilité saisonnière : La variabilité saisonnière se réfère aux fluctuations climatiques qui se produisent selon les saisons, affectant les cycles agricoles et l'équilibre des écosystèmes.

1.5.2.1 Configurations de la variabilité climatique

La variabilité climatique se manifeste à travers différentes configurations régionales et globales, influencées par des oscillations océaniques et atmosphériques. Ces phénomènes jouent un rôle central dans la modulation des températures, des précipitations et des extrêmes climatiques.

Configurations de la Variabilité Climatique

[label=☒] **Réchauffement global uniforme.** Anomalies ENSO (El Niño-Southern Oscillation). Oscillations décennales cohérentes.

[label=☐] **Vagues de chaleur localisées.** Événements climatiques extrêmes imprévus. Précipitations irrégulières.

Légende :

- ☒ : Configuration reconnue comme cohérente.
- ☐ : Configuration non reconnue ou incertaine.

Les configurations de la variabilité climatique incluent des phénomènes globaux et locaux. Les **configurations cohérentes** se réfèrent à des tendances régulières ou cycliques bien identifiées. Les **configurations non cohérentes**, en revanche, représentent des événements imprévus ou irréguliers souvent associés à des conditions climatiques extrêmes.

Le Tableau 1.2 résume les principaux types de variabilité climatique : spatiale (différences entre régions), temporelle (fluctuations à court ou long terme), interannuelle (changements d’une année à l’autre) et intrannuelle (variations au sein d’une même année).

Type de Variabilité	Description	Exemple
Variabilité Spatiale	Différences des conditions climatiques entre différentes régions.	Variations climatiques entre zones urbaines et rurales.
Variabilité Temporelle	Fluctuations du climat à court ou long terme.	Changements saisonniers ou réchauffement climatique.
Variabilité Interannuelle	Changements du climat entre différentes années.	Comparaison des températures moyennes d’une année à l’autre.
Variabilité Intrannuelle	Changements climatiques au cours d’une même année.	Les différences de température entre l’hiver et l’été dans une même année.

Tableau 1.2 – Configurations de la variabilité climatique

1.5.3 Facteurs de la variabilité climatique

Les facteurs de la variabilité climatique englobent des mécanismes naturels et anthropiques qui influencent les systèmes atmosphériques et océaniques. Ils déterminent les fluctuations climatiques à différentes échelles temporelles et spatiales.

Le Tableau 1.3 présente les principaux facteurs de la variabilité climatique, avec leurs symboles mathématiques, leur signification, et des exemples concrets de mesures. Il permet de mieux comprendre les éléments clés influençant le climat.

Symbole	Signification	Exemple
\mathcal{T}_{max}	Température maximale quotidienne	Température maximale enregistrée à 14h00.
\mathcal{T}_{min}	Température minimale quotidienne	Température minimale enregistrée à 6h00.
\mathcal{P}	Précipitations (mm)	Pluie cumulée sur une période de 24 heures.
\mathcal{H}	Humidité relative (%)	Humidité de l’air mesurée à 10h00.
\mathcal{V}	Vitesse du vent (m/s)	Vitesse du vent moyenne sur 10 minutes.
\mathcal{S}_R	Rayonnement solaire global (W/m^2)	Énergie reçue par unité de surface à la surface terrestre.
\mathcal{L}	Longueur du jour (heures)	Durée pendant laquelle le soleil est au-dessus de l’horizon.
\mathcal{A}	Altitude (m)	L’altitude d’une station météorologique spécifique.
$\cdot\mathcal{T}$	Amplitude thermique	Différence entre \mathcal{T}_{max} et \mathcal{T}_{min} sur une période donnée.

Tableau 1.3 – Facteurs de la variabilité climatique avec symboles et exemples

La jonction de ces paramètres climatiques repose sur l’application de règles conditionnelles combinant des opérateurs logiques tels que AND et OR. Ces *règles métier*, qui peuvent être explicites ou implicites, reflètent les connaissances et l’expérience des acteurs impliqués dans l’analyse, comme les experts en

climatologie ou les questionnaires de systèmes météorologiques. Ces règles permettent d'associer différents paramètres (**température, précipitations, vent**, etc.) pour parvenir à des décisions adaptées, telles que l'identification d'une situation de sécheresse ou la détection d'une tempête imminente.

Ces relations conditionnelles entre les paramètres peuvent être extraites et affinées à l'aide d'algorithmes de *machine learning*, notamment des modèles tels que les arbres de décision et les forêts d'arbres décisionnels. Ces algorithmes sont capables d'apprendre à partir des données historiques et de détecter des patterns complexes dans les relations entre les paramètres climatiques, permettant ainsi d'améliorer la précision des décisions prises. Les arbres de décision, par exemple, peuvent modéliser de manière explicite les règles conditionnelles sous forme de décisions binaires à chaque nœud, tandis que les forêts d'arbres décisionnels, en combinant plusieurs arbres, permettent de renforcer la robustesse des prédictions et de réduire le risque d'overfitting.

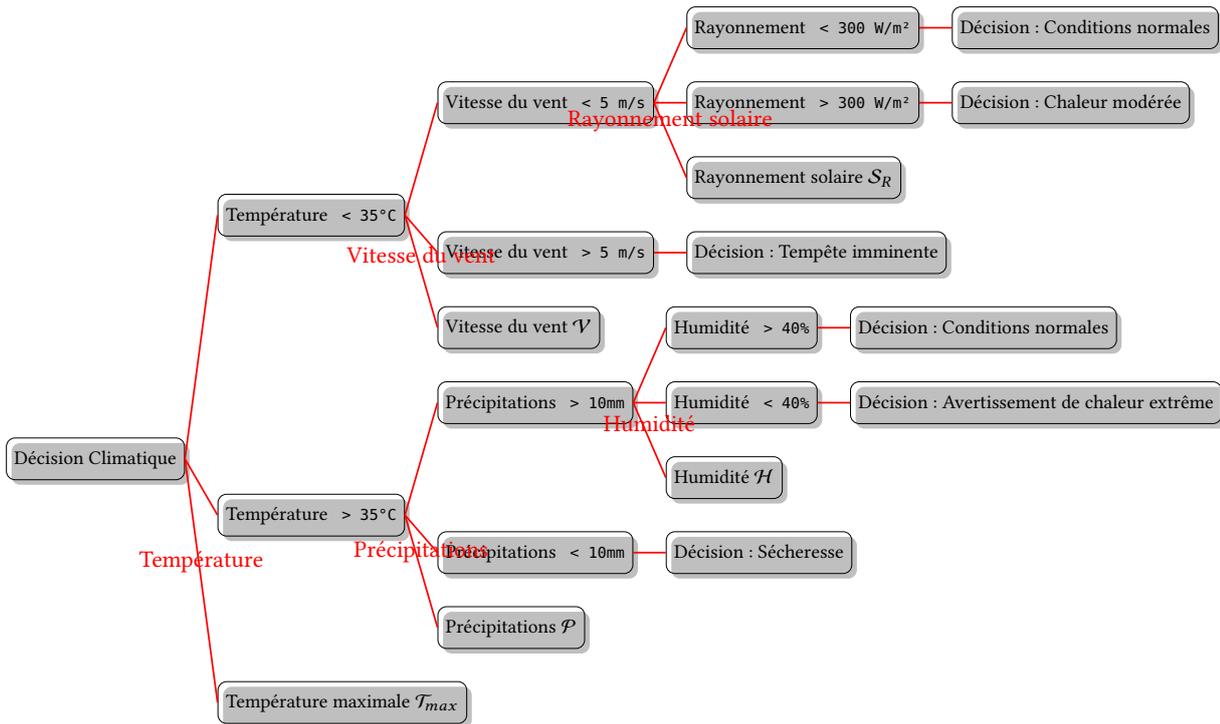
Ainsi, l'intégration des paramètres climatiques à travers ces règles et algorithmes de *machine learning* permet de formaliser les connaissances tacites des experts et de les rendre exploitables pour une prise de décision automatisée et optimisée, en tenant compte des spécificités locales et temporelles des phénomènes climatiques observés.

Le Tableau 1.4 présente des règles conditionnelles basées sur des paramètres climatiques tels que la température maximale, les précipitations, la vitesse du vent, et le rayonnement solaire. Ces règles, définies par des seuils spécifiques, permettent d'analyser et de prédire des conditions climatiques particulières. Elles peuvent être extraites et appliquées à l'aide d'algorithmes de machine learning comme les arbres de décision ou les forêts aléatoires.

Paramètre	Condition	Règle Conditionnelle
\mathcal{T}_{max}	$\mathcal{T}_{max} > 30^\circ C$	Si la température maximale est supérieure à 30C, alors il est probable que la journée soit chaude.
\mathcal{T}_{min}	$\mathcal{T}_{min} < 10^\circ C$	Si la température minimale est inférieure à 10C, alors il est probable qu'il fasse froid la nuit.
\mathcal{P}	$\mathcal{P} > 50 \text{ mm}$	Si les précipitations sont supérieures à 50 mm sur 24 heures, cela peut indiquer des conditions pluvieuses ou orageuses.
\mathcal{V}	$\mathcal{V} > 10 \text{ m/s}$	Si la vitesse du vent dépasse 10 m/s, il pourrait s'agir d'une tempête ou d'un vent fort.
\mathcal{S}_R	$\mathcal{S}_R > 1000 \text{ W/m}^2$	Si le rayonnement solaire est supérieur à 1000 W/m ² , il est probable qu'il y ait un ensoleillement intense.
\mathcal{L}	$\mathcal{L} > 12 \text{ heures}$	Si la durée du jour est supérieure à 12 heures, cela indique une période estivale avec plus de lumière.

Tableau 1.4 – Exemples de règles conditionnelles basées sur des paramètres climatiques

Les décisions climatiques reposent sur des règles de mesure, soit explicites, codées a priori dans un système expert, soit implicites, apprises à partir de données par apprentissage machine. Par exemple, un arbre décisionnel (voir la représentation visuelle sous forme d'un arbre de décision) comme celui illustré analyse des paramètres comme la température maximale \mathcal{T}_{max} , les précipitations \mathcal{P} , l'humidité \mathcal{H} , etc., pour aboutir à des décisions telles que la sécheresse ou des alertes de chaleur extrême. Les systèmes experts utilisent des règles logiques claires, tandis que les modèles d'apprentissage machine découvrent des relations subtiles entre les variables, générant des règles implicites difficiles à formaliser manuellement. Cette combinaison permet une gestion plus précise des risques climatiques.



Exemple 2. Dans le cadre de l'étude de la variabilité climatique d'une région donnée, plusieurs paramètres sont utilisés pour analyser les conditions climatiques quotidiennes. En observant les données collectées sur une période d'une semaine, on remarque que la température maximale quotidienne (T_{max}) varie fortement entre les jours en raison des changements de pression atmosphérique. Par exemple, une température maximale de 35°C a été enregistrée à 14h00 un jour, tandis que le lendemain, elle n'a atteint que 28°C à la même heure.

Par ailleurs, les précipitations (\mathcal{P}) ont joué un rôle clé en influençant l'humidité relative (\mathcal{H}), mesurée à 10h00. Lors d'une journée particulièrement humide, avec une humidité relative atteignant 85%, la température minimale quotidienne (T_{min}) est restée élevée. La vitesse du vent (\mathcal{V}) et le rayonnement solaire (S_R) ont aussi été observés, montrant une corrélation avec l'amplitude thermique ($\cdot\mathcal{T}$), qui a atteint 15°C lors de journées ensoleillées et sans vent. Ces facteurs combinés ont permis d'obtenir une vue d'ensemble des conditions climatiques et de mieux comprendre les causes de la variabilité dans la région.

L'analyse des paramètres climatiques montre une forte corrélation entre la température maximale et la vitesse du vent. Les précipitations et l'humidité relative jouent également un rôle crucial dans la variabilité thermique de la région.

1.5.3.1 Oscillations océaniques (ENSO, AMO, PDO)

Les oscillations océaniques, telles que l'ENSO (El Niño-Southern Oscillation), l'AMO (Atlantic Multidecadal Oscillation) et la PDO (Pacific Decadal Oscillation), jouent un rôle majeur dans la variabilité climatique mondiale. Ces phénomènes périodiques influencent les précipitations, les températures, et les régimes de vent à travers le globe.

Les caractéristiques principales de ces oscillations sont résumées dans le tableau ci-dessous :

Oscillation	Région principale	Impact climatique
ENSO (El Niño-Southern Oscillation)	Pacifique équatorial	Modification des régimes de précipitations (inondations et sécheresses), réchauffement global temporaire.
AMO (Atlantic Multidecadal Oscillation)	Océan Atlantique Nord	Influence sur la fréquence des ouragans dans l'Atlantique, modification des températures en Europe et en Amérique du Nord.
PDO (Pacific Decadal Oscillation)	Pacifique Nord	Variations des températures océaniques et terrestres, impact sur les stocks de poissons et les régimes de précipitations.

Tableau 1.5 – Caractéristiques des oscillations océaniques.

Ces oscillations océaniques sont des indicateurs clés pour comprendre et prédire les variations climatiques à court et long terme. Elles sont surveillées de près pour leurs implications sur l'agriculture, les écosystèmes et les populations humaines.

1.5.4 Les tendances de la variabilité et du changement climatique

Les tendances de la variabilité et du changement climatique peuvent être analysées en termes de relations logiques et mathématiques. L'évolution des paramètres climatiques (comme la température, les précipitations, et l'humidité) au fil du temps peut être modélisée à l'aide de règles conditionnelles, où des opérateurs logiques tels que \wedge (AND), \vee (OR), \oplus (XOR), \Rightarrow (Implique), et \Leftrightarrow (Require) sont utilisés pour exprimer les interactions entre ces variables.

Soit $T(t)$ la température à l'instant t , $P(t)$ les précipitations, et $H(t)$ l'humidité. Une règle de décision pour déterminer si un événement climatique extrême se produit pourrait être formulée comme suit :

$$(T(t) > 35^{\circ}\text{C} \wedge P(t) < 10 \text{ mm}) \Rightarrow \text{Sécheresse}$$

Cela signifie qu'une température supérieure à 35°C ET des précipitations inférieures à 10 mm entraînent une décision de sécheresse.

De même, l'association d'un vent fort avec un faible taux d'humidité pourrait être exprimée par un opérateur XOR, indiquant que l'une ou l'autre des conditions peut suffire à provoquer un avertissement de chaleur extrême :

$$(\mathcal{V}(t) > 5 \text{ m/s}) \oplus (H(t) < 40\%) \Rightarrow \text{Avertissement de chaleur extrême}$$

Les tendances globales du changement climatique peuvent également être modélisées par des

implications conditionnelles complexes, où une condition préalable (par exemple, une augmentation de la température moyenne) implique un ensemble de conséquences :

$$(T(t + 1) > T(t)) \Rightarrow (P(t + 1) \text{ ou } H(t + 1) \text{ modifiés})$$

En utilisant ces notations mathématiques et logiques, on peut exprimer des modèles de variabilité et de changement climatique qui capturent à la fois la complexité des interactions climatiques et les règles qui les régissent.

1.6 Impact sur l'agriculture

Les fluctuations climatiques (T, P, H) influencent la croissance végétale, affectant les rendements agricoles. L'indice NDVI capte cette réponse, offrant un indicateur utile pour anticiper les impacts sur la productivité.

1.6.1 Variabilité des rendements agricoles

La variabilité des rendements agricoles dépend à la fois des fluctuations climatiques et de la dynamique de la végétation, mesurée par l'indice NDVI. Le NDVI reflète l'état de la végétation et son évolution saisonnière, tandis que la variabilité climatique concerne les changements des températures, précipitations et humidité. Une comparaison montre que les baisses du NDVI correspondent souvent à des stress climatiques, comme la sécheresse, affectant les rendements. Toutefois, la variabilité climatique est plus complexe et agit à plusieurs échelles spatiales et temporelles. Combiner les données NDVI et climatiques permet donc une meilleure compréhension et prédiction des variations de la production agricole.

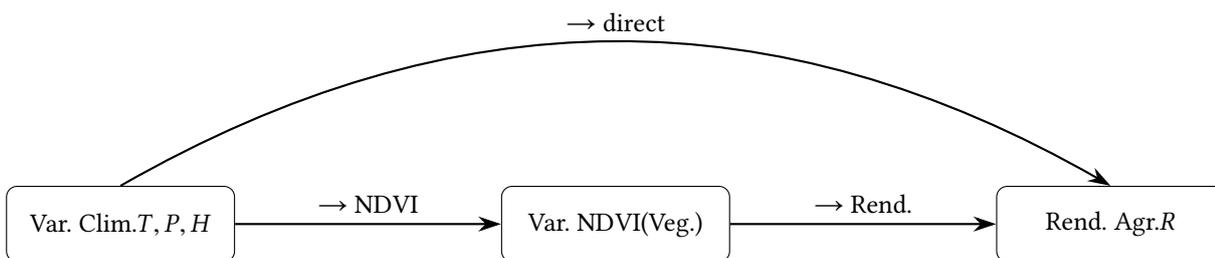


FIGURE 1.7 – Relations causales : Var. Clim. = Variabilité climatique (T : Temp., P : Précip., H : Humidité), Var. NDVI = Variabilité de l'indice NDVI (Végétation), Rend. Agr. = Rendements agricoles (R). Les flèches indiquent l'impact et les influences entre ces variables.

1.6.2 Besoins en irrigation et en adaptation des cultures

Nous avons catégorisé les différents types de stratégies d'adaptation des cultures en fonction de leurs besoins en irrigation, en tenant compte des variations climatiques et des caractéristiques spécifiques de

chaque culture. Cette classification permet d'identifier les méthodes les plus adaptées pour améliorer la résilience des cultures face aux défis liés à l'eau.

Catégorie	Exemple de Culture	Stratégie d'Adaptation et Irrigation
Cultures Résistantes à la Sécheresse	Blé Dur, Millet, Sorgho	Utilisation d'irrigation de subsurface pour limiter l'évaporation, plantations adaptées aux périodes sèches.
Cultures Sensibles à la Sécheresse	Maïs, Riz	Installation de systèmes d'irrigation goutte-à-goutte pour assurer un approvisionnement en eau constant.
Cultures Adaptées à des Climats Humides	Riz en rizières, Canne à sucre	Adaptation des systèmes de drainage pour éviter l'excès d'eau et améliorer la gestion des inondations.
Cultures avec Besoins Modérés en Eau	Tomates, Pomme de terre	Optimisation de l'irrigation par aspersion en fonction des prévisions climatiques pour ajuster les besoins.
Cultures Pérennes avec Résilience Climatique	Vigne, Olive	Systèmes d'irrigation d'appoint pour compléter les faibles précipitations, gestion des variétés résistantes à la chaleur.

Tableau 1.6 – Catégorisation des stratégies d'adaptation des cultures en fonction des besoins en irrigation

1.7 Mesures et indicateurs de la variabilité climatique

Les indices climatiques jouent un rôle crucial dans l'analyse des variations climatiques, en fournissant des informations sur les tendances et les anomalies des paramètres climatiques à différentes échelles spatio-temporelles. Ces indices permettent de mieux comprendre les phénomènes climatiques et de suivre l'évolution du climat à travers des mesures spécifiques. Voici quelques-uns des indices les plus utilisés :

1. La variabilité climatique fait référence aux fluctuations naturelles ou anthropiques des paramètres climatiques sur différentes échelles de temps et d'espace. Les précipitations et les températures sont les deux variables principales utilisées pour évaluer cette variabilité.
2. Les précipitations, par leur répartition spatiale et temporelle, influencent directement les écosystèmes, l'agriculture et les ressources en eau. Les variations des températures affectent, quant à elles, la biodiversité, les cycles de culture et les phénomènes météorologiques extrêmes.
3. Pour mesurer la variabilité climatique, diverses techniques sont utilisées, notamment les observations in situ (stations météorologiques), les données satellitaires et les modèles climatiques.
4. Les indices de variabilité, tels que l'écart-type, le coefficient de variation et les anomalies standardisées, permettent de quantifier les fluctuations des variables climatiques. Ces indices sont essentiels pour comprendre les tendances et les extrêmes climatiques.
5. Les méthodes statistiques comme l'analyse des séries temporelles et les outils géostatistiques complètent ces indices pour une évaluation approfondie de la variabilité climatique.
6. Comprendre la variabilité climatique à travers ces mesures et indicateurs est fondamental pour développer des stratégies d'adaptation face aux changements climatiques et leurs impacts socio-économiques.

1.7.1 Indices climatiques

Les indices climatiques jouent un rôle crucial dans l'analyse des variations climatiques, en fournissant des informations sur les tendances et les anomalies des paramètres climatiques à différentes échelles spatio-temporelles[39]. Ces indices permettent de mieux comprendre les phénomènes climatiques et de suivre l'évolution du climat à travers des mesures spécifiques[146]. Voici quelques-uns des indices les plus utilisés :

- Les indices climatiques sont des outils essentiels pour analyser les variations des paramètres climatiques à différentes échelles spatio-temporelles.
- Parmi les indices globaux, les anomalies de température mesurent l'écart par rapport à une moyenne de référence et permettent de suivre le réchauffement climatique.
- L'indice ENSO (El Niño-Southern Oscillation) reflète les interactions océan-atmosphère dans le Pacifique, influençant le climat à l'échelle mondiale.
- Pour les précipitations, des indices comme l'indice de précipitation standardisé (SPI) et le coefficient de variation (CV) permettent de détecter les sécheresses et les épisodes de forte humidité.
- Les indices de température incluent les maxima et minima extrêmes, ainsi que les degrés-jours, qui analysent les variations saisonnières et les événements climatiques extrêmes.

1.7.2 Modèles de prédiction climatique

Les modèles de prédiction climatique sont des outils mathématiques et informatiques utilisés pour simuler l'évolution des paramètres climatiques à différentes échelles spatio-temporelles. Ils reposent sur des équations physiques représentant les interactions entre l'atmosphère, l'océan, la biosphère et la cryosphère [40, 39]. Les modèles globaux (GCM) permettent d'étudier le climat à l'échelle planétaire, tandis que les modèles régionaux (RCM) offrent une résolution plus fine pour des zones spécifiques. Ces modèles intègrent des données historiques et des scénarios d'émissions de gaz à effet de serre pour prévoir les changements climatiques futurs. Ils sont essentiels pour évaluer les impacts du changement climatique et orienter les stratégies d'adaptation et d'atténuation. Leur précision dépend de la qualité des données et des hypothèses utilisées.

Ces derniers jouent un rôle fondamental dans la compréhension et la prévision des impacts du changement climatique. Leur capacité à simuler les paramètres climatiques globaux et régionaux fournit des informations essentielles pour orienter les politiques d'adaptation et de mitigation.

Exemples d'applications des GCM et RCM :

1. Projections pour le bassin méditerranéen : Les études utilisant des *Global Climate Models* (GCM) ont montré une tendance générale au réchauffement dans la région méditerranéenne, avec une augmentation projetée des températures moyennes de 2 à 5°C d'ici à la fin du 21^e siècle selon les scénarios d'émissions [39]. Par exemple, les GCM révèlent une intensification des sécheresses estivales en raison de la diminution des précipitations estivales, particulièrement marquée dans les zones côtières de l'Afrique du Nord et du sud de l'Espagne [40].

2. Projections des précipitations avec les RCM : Les *Regional Climate Models* (RCM) permettent d'affiner ces projections en intégrant les caractéristiques spécifiques des régions, comme la topographie et les courants marins. Dans le cadre du programme européen *EURO-CORDEX*, les RCM ont identifié une augmentation des précipitations intenses en automne dans le nord-ouest de l'Italie, souvent associée à des épisodes méditerranéens causant des inondations [58]. À l'inverse, les précipitations hivernales dans le sud-est méditerranéen, notamment en Turquie et en Grèce, devraient diminuer de 10 à 20% selon les projections [135].

3. Impacts combinés : Les modèles combinés (GCM et RCM) prévoient également une augmentation de la fréquence et de l'intensité des vagues de chaleur dans le bassin méditerranéen. Ces vagues de

chaleur affecteront non seulement la santé humaine, mais aussi les systèmes agricoles, en réduisant la disponibilité en eau et en augmentant les risques de feux de forêt [40]. Par exemple, des simulations ont montré que le rendement des cultures dans des pays comme l'Espagne pourrait diminuer de 20 à 40% en raison du stress thermique accru [44].

4. Changements dans les écosystèmes marins : Les RCM appliqués au domaine marin montrent que le réchauffement de la mer Méditerranée pourrait atteindre 3°C d'ici 2100, entraînant une redistribution des espèces marines, avec des implications pour la pêche et la biodiversité [2]. Les GCM indiquent également une augmentation de l'acidification des océans, impactant les écosystèmes coralliens et les coquillages [39].

Défis et perspectives : Malgré ces progrès, des incertitudes subsistent, notamment en raison de la variabilité naturelle du climat et des hypothèses concernant les émissions futures. Les efforts actuels se concentrent sur le développement de modèles à très haute résolution et sur l'amélioration de la représentation des processus complexes, comme les interactions entre l'atmosphère et les sols dans les régions arides [58].

1.8 Moyens d'acquisition des paramètres de variabilité climatique

Dans cette section nous présentons l'architecture globale de système composée des deux modules principaux suivants :

1. Une infrastructure acquisition pour chaque systèmes physiques (machines).
2. Un système d'aide à la décision hébergé dans un serveur et utilise les données récoltées à partir des différentes systèmes.

L'infrastructure de système d'acquisition des paramètres de variabilité climatique est composée des éléments suivants :

- Un ensemble de capteurs (humidité, température, lumière, etc.) et d'équipements organisés dans un réseau.
- L'application de contrôle (desktop, application web et mobile), permettant l'exposition des fonctionnalités des composants et des connexions du système.
- Des protocoles de communication permettant la coordination des différentes communications entre les dispositifs intelligents et les applications. Le système décisionnel est composé des éléments suivants :
 - La machine est connectée (système d'étude utilisant des capteurs d'acquisition)).
 - Un serveur contenant les données récoltées des différentes systèmes. Ces données sont d'abord sauvegardées dans une unité de stockage spécifique pour chaque systèmes et envoyées ensuite au serveur.
 - Des méthodes d'analyse et de traitement de données pour extraire de l'information pertinente à partir des différentes données récoltées et stockées dans le serveur.

L'architecture du système est présentée dans le schéma suivant, nous décrivons ensuite en détails tous les composants de cette architecture.

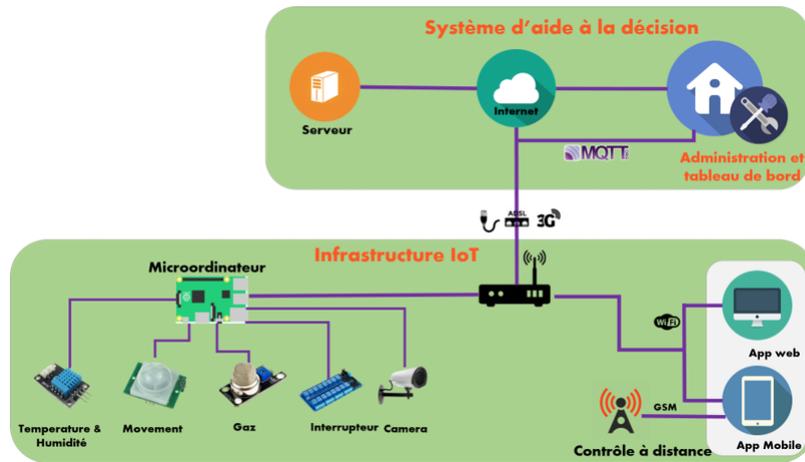


FIGURE 1.8 – Schéma illustratif de l'architecture d'une solution IOT

1.8.1 Architecture de infrastructure

L'architecture se compose de plusieurs couches clés (Figure 1.9) : la couche des dispositifs, chargée de la collecte des données ; la couche de communication, assurant leur transmission ; la couche de stockage, dédiée à la gestion des données ; la couche de communication externe, facilitant les échanges avec l'extérieur ; et enfin, la couche de contrôle d'accès, garantissant la sécurité et la gestion des droits.

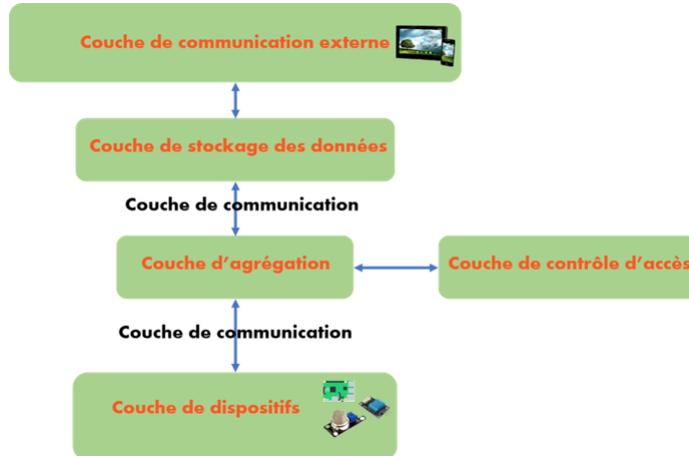


FIGURE 1.9 – L'architecture d'acquisition de données

1.8.1.1 La couche des dispositifs :

Cette couche regroupe les différents objets intelligents (les capteurs et les organes de mouvements et les micros ordinateurs) interconnectés dans un réseau de cumminication. Pour le contrôle de mouvement, nous avons besoin d'un interrupteur pour actionner les organes, le microordinateur commande l'interrupteur en lui fournissant le courant nécessaire à la manipulation des organes. Comme la plupart des microordinateurs ne supportent pas les signaux analogiques, nous utilisons un nœud passerelle comme

convertisseur de signaux analogiques. Cette plateforme se comporte comme un système modulaire où chaque capteur est défini comme un module, ce qui permet au utilisateur de choisir les modules à inclure selon ses besoins. La Figure 1.10 suivante présente les différents dispositifs de l'architecture :

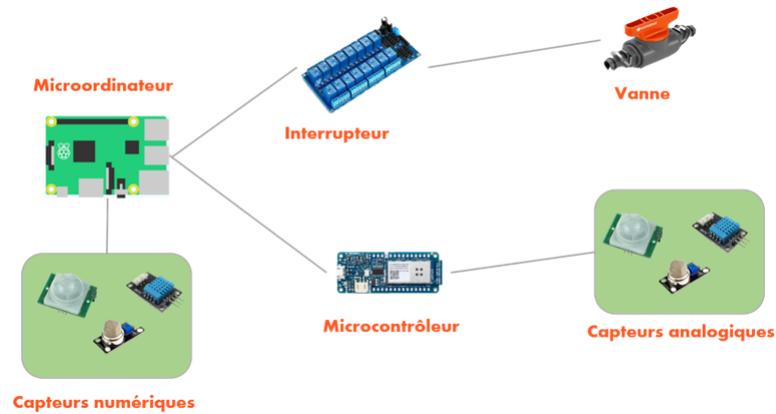


FIGURE 1.10 – Couche dispositifs

1.8.1.2 La couche de communication :

Cette couche gère les échanges entre objets connectés et applications de contrôle (web ou mobile), via une passerelle dédiée. L'architecture étant orientée services, le protocole choisi doit s'y adapter. Le protocole MQTT² a été retenu pour sa légèreté et son efficacité, des aspects détaillés plus loin.

La couche d'agrégation intègre un coordinateur de messages qui fait le lien entre les dispositifs et le réseau d'objets, en assurant la transmission bidirectionnelle des messages.

1.8.1.3 La couche de stockage :

Cette couche s'occupe de la persistance des données en provenance des objets intelligents. La persistance se fera dans des bases de données locales ou distantes, ce qui permettra d'avoir une vue d'ensemble sur les interactions ayant eu lieu avec les objets intelligents. Dans cette architecture, une base de données locale sera intégrée au micro-ordinateur pour conserver l'historique des opérations d'irrigation ainsi que les informations fournies par les capteurs. Ces bases de données serviront ensuite de source d'information pour le système décisionnel [83].

1.8.1.4 La couche de communication externe :

Cette couche assure la connexion des objets intelligents à leur environnement extérieur. Elle repose sur un ou plusieurs dispositifs de contrôle permettant l'interaction depuis des interfaces web ou mobiles. Les données issues des objets peuvent être visualisées sous différentes formes, comme des tableaux de bord ou des graphes. La transmission des commandes peut s'appuyer sur divers canaux de communication

2. MQTT (Message Queuing Telemetry Transport) est un protocole léger de messagerie publié/souscrit, optimisé pour les réseaux contraints et utilisé notamment dans les systèmes IoT.

(3G, GSM, ADSL), en passant par un microordinateur. Le protocole MQTT a été retenu pour faciliter les échanges entre les applications et les objets connectés. La couche de gestion des accès, quant à elle, veille à l'authentification des utilisateurs souhaitant interagir avec les objets, en s'appuyant sur des identifiants et des mots de passe.

1.8.2 Architecture du système décisionnel

Dans la section précédente, nous avons présenté en détails les différents composants du module « Infrastructure d'IoT », dans cette section nous présentons les composants du deuxième module qui est le système décisionnel.

Nous nous sommes inspirés de l'architecture de référence d'un système décisionnel pour mettre en œuvre l'architecture de système. L'architecture décisionnelle de système illustrée dans la Figure 1.11 est passe par les trois étapes suivantes :

- La collecte des données.
- Le stockage de données.
- L'exploitation.



FIGURE 1.11 – Les étapes du processus décisionnel

1.8.2.1 La collecte des données

Dans le cadre de l'aide à la décision, l'exploitation des capteurs repose sur l'intégration des avancées technologiques tant dans l'acquisition que dans l'analyse des données. L'objectif est de garantir la collecte de données précises, actualisées et pertinentes [163]. La collecte constitue en effet la première étape fondamentale du processus décisionnel. Dans ce système, les données seront extraites des bases de données opérationnelles des microordinateurs, principalement à partir des données issues des différents capteurs utilisés [163].

Quant au stockage des données, ces dernières seront transférées et centralisées dans une base de données dédiée, localisée sur un serveur central (Figure 1.11). Toutefois, avant cette phase de stockage, il est primordial que les données soient agrégées et harmonisées, afin d'assurer leur qualité et leur compatibilité pour une exploitation ultérieure [142, 32].

1.8.2.2 L'exploitation

Après les étapes de collecte et de stockage des données, la phase suivante, et probablement la plus cruciale dans un système décisionnel, est celle de l'exploitation des données. C'est à ce stade

que les données récoltées et traitées au cours des phases précédentes sont analysées en profondeur. L'objectif principal de cette phase est de fournir aux décideurs des informations pertinentes, claires et compréhensibles, leur permettant de prendre des décisions éclairées et fondées sur des données fiables. Comme nous l'avons déjà mentionné dans l'état de l'art, il existe plusieurs outils et méthodes d'exploitation des données, comme les tableaux de bord. Dans la Figure 1.12, nous présentons dans le schéma suivant l'architecture de système décisionnel :

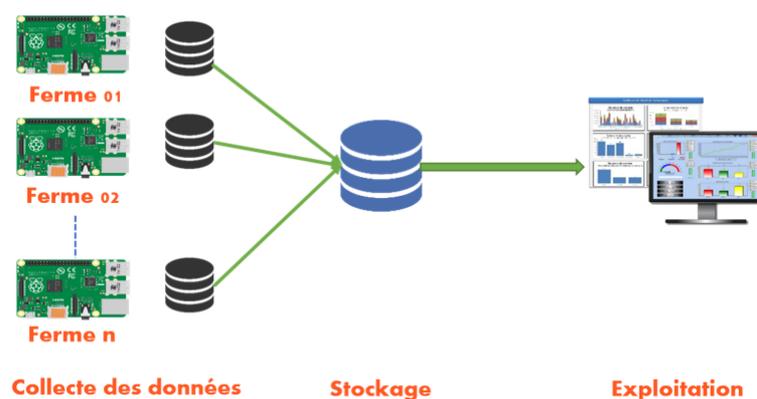


FIGURE 1.12 – Architecture du système décisionnel

1.9 Conclusion

La variabilité climatique, inhérente au système climatique terrestre, se manifeste par des fluctuations des températures et des précipitations à différentes échelles spatio-temporelles. Le climat, caractérisé par des interactions complexes entre l'atmosphère, l'océan, et la biosphère, subit actuellement des modifications significatives, attribuées en grande partie aux activités anthropiques. Les projections climatiques indiquent une augmentation des températures globales pouvant atteindre 3°C d'ici 2050 et 5°C en 2100, accompagnée d'une réduction des précipitations de 10 à 50 %, selon les régions et les saisons. La prise en compte conjointe de la variabilité climatique et de celle de l'indice NDVI s'avère essentielle pour affiner les prévisions et mieux comprendre l'impact des conditions environnementales sur les écosystèmes agricoles. Ce travail a permis d'identifier ces changements dans les paramètres climatiques, notamment sur l'agriculture. En combinant des modèles climatiques, stochastiques et de culture, il est possible d'anticiper les effets du changement climatique et d'orienter les stratégies d'adaptation de manière pertinente et efficace.

Modèles de Prévision de la Production Agricole



« La connaissance s'acquiert par l'expérience, tout le reste n'est que de l'information.. »

– Albert Einstein (Mathématicien, Physicien, Scientifique) (1879 - 1955)

Sommaire

2.1	Introduction	45
2.2	Principes des Modèles de Prévision	45
2.2.1	Définition des modèles de prévision	45
2.2.2	Caractéristiques des modèles en agriculture	45
2.2.3	Principaux défis en prévision agricole	46
2.2.4	Paradigme de modélisation des coûts de production culturelle	46
2.3	Classification des Approches de Modélisation	47
2.3.1	Modèles empiriques	47
2.3.2	Modèles mécanistes	49
2.3.3	Modèles hybrides	50
2.4	Approches de Modélisation des Cultures	51
2.4.1	Méthodologies employées	51
2.4.2	Choix des variables explicatives	52
2.4.3	Techniques de modélisation utilisées	54
2.5	Approches de modèles de prédiction	54
2.5.1	Les fonctions des modèles de prédiction	55
2.6	Différentes formes d'utilisation des modèles	56
2.6.1	Approche analytique dirigée par les données	57
2.7	Définitions préliminaires et notation	57
2.7.1	Transformateurs	57
2.7.2	Clustering	58
2.7.3	Notation	58
2.7.4	Clustering K-means	59
2.7.5	Régressions linéaires et non linéaires	60
2.7.6	ARIMA	60
2.7.7	Réseaux de neurones et deep learning	61
2.7.8	Réseaux de Neurones Récurrents (RNNs)	61
2.7.9	Long Short-Term Memory (LSTM) Networks	62

2.7.10	Unités Récurrentes à Porte (GRU)	63
2.7.11	Transformers	64
2.8	Évaluation des Modèles de Prédiction	68
2.8.1	classification et comparaisons des travaux	68
2.9	Bilan et Synthèse	70
2.10	intégration des données multi sources	71
2.10.1	intégration des données sources	71
2.10.2	Federated learning	72
2.10.3	Optimisation multicritère	72
2.11	Positionnement de nos recherches	72
2.11.1	Vers la réalisation de nos objectifs	73
2.11.2	Aperçue de notre solution	74
2.12	Conclusion	74

2.1 Introduction

La prévision de la production agricole constitue un enjeu majeur dans le contexte de la sécurité alimentaire, de la gestion durable des ressources naturelles et de l'adaptation aux changements climatiques[62, 109]. Ces dernières années, les modèles prédictifs ont évolué grâce à l'intégration de méthodologies analytiques, empiriques et basées sur le machine learning, permettant de mieux comprendre les facteurs influençant les rendements agricoles[35, 121, 104]. Dans ce chapitre, nous avons synthétisé les approches récentes de modélisation, en analysant leurs caractéristiques, avantages, limites et domaines d'application. Une attention particulière a été accordée aux méthodes d'évaluation, notamment via des métriques standard comme la RMSE, la MAE et le coefficient de détermination (R^2), pour juger de leur robustesse et efficacité. Cette synthèse offre une vue d'ensemble des modèles actuels et sert de base à une discussion approfondie, tout en orientant le lecteur vers des solutions adaptées aux enjeux agricoles contemporains.

2.2 Principes des Modèles de Prévision

Dans cette section, nous allons explorer les principes fondamentaux des modèles de prévision utilisés dans le contexte agricole. Nous détaillons les approches théoriques fondant ces modèles et les mécanismes d'estimation des rendements agricoles en fonction des facteurs environnementaux et socio-économiques.

2.2.1 Définition des modèles de prévision

Un *modèle de prévision* est une fonction f qui estime des valeurs futures à partir de paramètres d'entrée $\mathbf{x} \in \mathbb{R}^n$. Ce modèle peut être exprimé par la relation $y = f(\mathbf{x}, \theta)$, où y est la prédiction et θ les paramètres ajustés lors de l'entraînement. La fonction f transforme les données historiques ou contextuelles en une sortie estimée, permettant de prédire des événements futurs.

Comme un exemple d'instanciation de la définition précédente, pour estimer la production agricole, le modèle de prévision peut être formulé comme $y = f(\mathbf{x}, \theta)$, où y est la production estimée, $\mathbf{x} = (T_{moy}, P)$ représente les paramètres climatiques (température moyenne T_{moy} et précipitations P), et θ sont les paramètres du modèle ajustés en fonction des données historiques.

2.2.2 Caractéristiques des modèles en agriculture

Les modèles de prévision agricole intègrent plusieurs caractéristiques essentielles, telles que :

- **Prédiction environnementale** : Utilisation de la température et des précipitations pour estimer la production d'olives.
- **Données historiques** : Intégration des rendements passés pour affiner les prévisions.
- **Adaptabilité** : Ajustement pour différents types d'oliviers et conditions locales.
- **Incertitude** : Prise en compte des variations climatiques dans les prévisions.

Le Tableau 2.1 présente les caractéristiques des modèles de prévision de la production d'olives, incluant des exemples liés à la température, les précipitations et les rendements passés.

Caractéristique	Exemple
Prédiction environnementale	Utilisation de la température T et des précipitations P pour estimer la production d'olives.
Données historiques	Intégration des rendements passés $R_{t-1}, R_{t-2}, \dots, R_{t-n}$ pour affiner les prévisions.
Adaptabilité	Ajustement pour différents types d'oliverais, O_1, O_2, O_3 , et conditions locales (température T et précipitation P).
Incertitude	Prise en compte des variations climatiques ΔT et ΔP dans les prévisions.

Tableau 2.1 – Exemples des principales caractéristiques des modèles de prévision pour la production d'olives avec des notations mathématiques.

2.2.3 Principaux défis en prévision agricole

Dans cette section, nous avons abordé les défis auxquels la prévision agricole est confrontée, notamment la variabilité climatique, les pratiques humaines, ainsi que les caractéristiques morphologiques et physiologiques des cultures. Ces facteurs, souvent interconnectés, influencent directement la production agricole, ce qui nécessite des modèles prédictifs capables d'intégrer ces interactions pour garantir des prévisions précises et une gestion durable des ressources. Par exemple, dans le cadre des capteurs IoT utilisés pour la collecte de données climatiques et environnementales, l'extraction précise des paramètres tels que la température et l'humidité est cruciale pour affiner les prévisions agricoles...ect

2.2.4 Paradigme de modélisation des coûts de production culturelle

En général, deux Paradigme principales sont efficaces pour modéliser la relation entre les intrants et les extrants dans la production culturelle : l'**Apprentissage Automatique (AA)** et les **Modèles Analytiques (MA)**. Les modèles hybrides, souvent appelés **modèles boîte grise**, cherchent à combiner les forces des deux approches, offrant un équilibre entre les connaissances théoriques détaillées fournies par les MA et l'adaptabilité basée sur les données des modèles AA (voir Figure Figure 2.1). Les MA se concentrent sur la compréhension des mécanismes internes influençant les coûts de production culturelle par le biais d'une sélection explicite des paramètres. Cette approche évite le besoin de données d'entraînement et offre une grande efficacité, mais elle peut être complexe et nécessiter une maintenance continue. D'autre part, les modèles AA génèrent des fonctions de coût de production à partir de données expérimentales, les rendant adaptables à différents contextes ; cependant, ils sont souvent spécifiques à une application donnée et nécessitent des expérimentations répétées pour chaque nouvel environnement. Les modèles hybrides combinent ces approches en utilisant des paramètres dérivés des MA tout en exploitant l'AA pour apprendre des caractéristiques inconnues (par exemple, [130, 38]). La Production Culturelle, notée $\mathcal{P}rod$, est représentée comme une fonction mathématique qui évalue des métriques liées aux opérations agricoles, telles que le rendement des cultures et le coût de production, en prenant en compte divers paramètres comme le climat et le sol. Cette approche vise à améliorer l'efficacité et la durabilité agricoles en optimisant les facteurs influençant les coûts de production (par exemple, [11]).

Les modèles hybrides, ou *modèles boîte grise*, combinent les approches boîte blanche et boîte noire. Ils utilisent des paramètres notés $Param = Param_{dérivés} \cup Param_{Inconnus}$, où $Param_{dérivés}$ provient des Modèles Analytiques (MA) et $Param_{Inconnus}$ est appris par des techniques d'Apprentissage Automatique (AA).

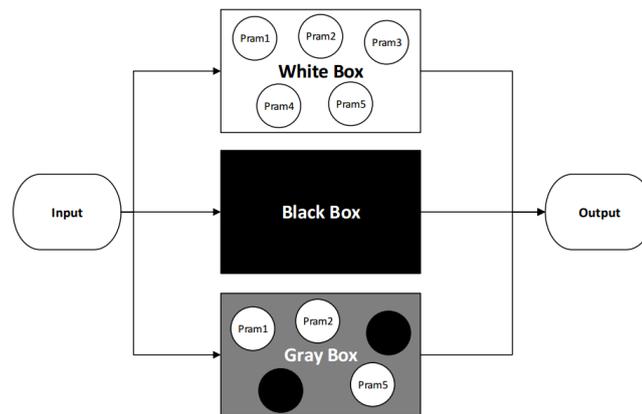


FIGURE 2.1 – Paradigme de modélisation des coûts de production culturelle.

2.3 Classification des Approches de Modélisation

En explorant la littérature, il existe une catégorisation des approches de modélisation utilisées dans la prévision agricole (par ex. [157]). Ces approches peuvent être classées en différentes catégories selon leurs méthodologies, leurs niveaux de complexité, et leur capacité à intégrer des facteurs environnementaux et climatiques. Nous aborderons les principales classifications et discuterons des avantages et des limites de chaque approche.

<ul style="list-style-type: none"> □ Préparation climatique (Tmin, Tmax, Tmoy, précipitations) □ Type de sol (acide, calcaire, etc.) □ Système d'irrigation (goutte-à-goutte, aspersion) 	<ul style="list-style-type: none"> □ Méthodes de fertilisation (engrais organiques, minéraux) □ Protection contre les maladies (fongicides, insecticides) □ Récolte (période optimale, méthodes ma- 	<ul style="list-style-type: none"> nelles ou mécaniques) □ Suivi de la croissance (stades phénologiques des oliviers) □ Analyse des rendements (quantité d'huile, qualité)
---	--	---

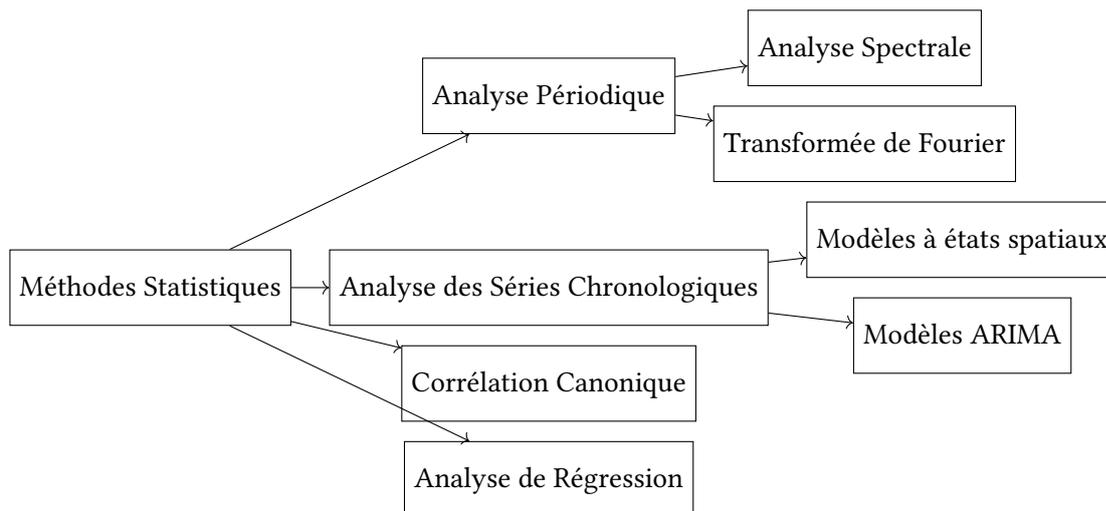
2.3.1 Modèles empiriques

Les modèles empiriques jouent un rôle crucial dans la compréhension des interactions complexes entre le système climatique et les activités humaines. Ces modèles, fondés sur des relations statistiques, utilisent des données historiques pour établir des corrélations entre des variables climatiques (température, précipitations, etc.) et leurs impacts sur des systèmes spécifiques [110, 20]. Dans le contexte climatique, ils permettent d'analyser les réponses des écosystèmes, des cultures agricoles ou des infrastructures aux variations climatiques. Ces modèles intègrent plusieurs composantes, telles que les observations climatiques, les données socio-économiques et les interactions spatiales et temporelles. Bien qu'ils soient simples à implémenter, leur dépendance aux données historiques les rend limités face aux incertitudes futures [74]. Ils sont souvent utilisés comme complément aux modèles mécanistes pour

des analyses prédictives.

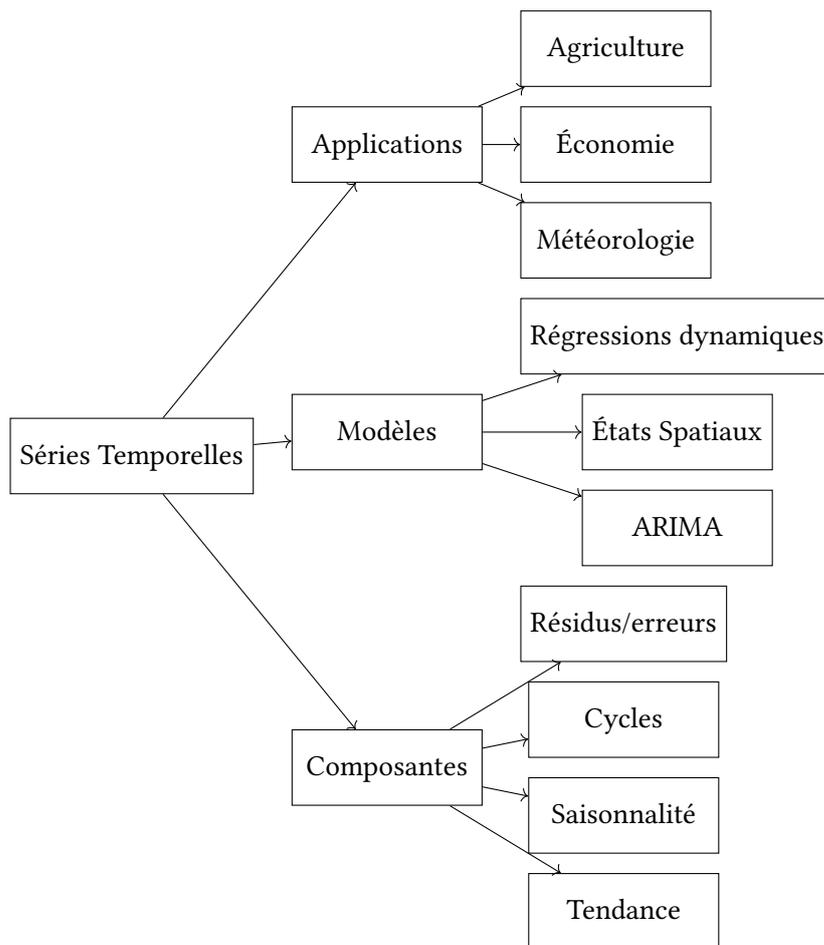
2.3.1.1 Modèle statistiques

Les méthodes statistiques jouent un rôle essentiel dans l'analyse et la prévision des phénomènes complexes, notamment en sciences climatiques. Parmi les approches les plus utilisées, l'analyse de régression permet d'établir des relations fonctionnelles entre une variable dépendante et une ou plusieurs variables indépendantes[22]. La corrélation canonique, quant à elle, identifie les relations entre deux ensembles de variables multidimensionnels, souvent utiles pour explorer les liens entre variables climatiques et socio-économiques[70]. L'analyse des séries chronologiques se concentre sur l'évolution temporelle des données et intègre des techniques comme les modèles ARIMA ou les modèles à états spatiaux pour prévoir les tendances futures[22]. L'analyse périodique, comme la transformation de Fourier, met en évidence les composantes cycliques dans les données, par exemple pour détecter des schémas saisonniers ou des cycles climatiques[127]. D'autres techniques avancées, telles que les méthodes bayésiennes ou les réseaux de neurones, viennent enrichir les outils de prévision[55, 71]. Ces approches, combinées ou appliquées individuellement, permettent une meilleure compréhension et anticipation des phénomènes complexes.



2.3.1.2 Approches basées sur les séries temporelles

Les méthodes statistiques basées sur les séries temporelles jouent un rôle central dans l'analyse et la prévision des phénomènes évolutifs dans le temps. Ces approches permettent d'identifier des tendances, des variations saisonnières et des cycles dans les données chronologiques. Les modèles ARIMA, largement utilisés, combinent les composantes autorégressives et différentielles pour capturer la dynamique sous-jacente[22]. Les modèles à états spatiaux, quant à eux, offrent une flexibilité pour modéliser des systèmes complexes[48]. Ces techniques sont appliquées dans divers domaines tels que la météorologie, l'économie et l'agriculture pour prévoir des événements futurs[67]. Grâce à leur capacité à exploiter les patterns temporels, elles facilitent la prise de décision basée sur des données fiables. Cependant, elles nécessitent souvent des ajustements précis pour gérer les incertitudes et les anomalies dans les données.



2.3.2 Modèles mécanistes

Les modèles mécanistes sont des outils essentiels pour décrire et comprendre les processus sous-jacents à un phénomène, en s'appuyant sur des principes fondamentaux issus des sciences physiques, chimiques ou biologiques. Contrairement aux modèles empiriques, qui reposent sur des corrélations statistiques, les modèles mécanistes s'appuient sur des équations représentant les lois fondamentales de la nature, comme les équations de conservation de la masse, de l'énergie ou de la quantité de mouvement [16]. Ces modèles sont largement utilisés dans divers domaines, tels que l'agriculture [85], la gestion des ressources en eau et les sciences environnementales [31], pour analyser les interactions complexes entre les systèmes naturels et anthropiques. Cependant, leur application nécessite une bonne compréhension des mécanismes sous-jacents et des données précises pour calibrer et valider les simulations. Voici les principaux aspects de ces modèles :

- **Description des processus** : Ils reposent sur des principes physiques, biologiques ou chimiques pour établir des relations causales.
- **Méthode de modélisation** : Utilisation d'équations différentielles ou de simulations numériques pour simuler le comportement du système.
- **Applications** : Employés dans des domaines tels que l'écologie, l'agriculture, et la météorologie.
- **Compréhension des mécanismes** : Permet une compréhension approfondie des mécanismes

sous-jacents à un phénomène.

- **Précision nécessaire** : Nécessitent des données précises et des hypothèses robustes pour fonctionner correctement.
- **Avantage principal** : Leur capacité à expliquer les mécanismes sous-jacents les rend précieux pour la prédiction et l'analyse.

2.3.2.1 Modélisation biophysique

La modélisation biophysique est une approche qui utilise des modèles mathématiques pour simuler et comprendre les interactions entre les systèmes biologiques et les facteurs physiques. Elle permet d'analyser des phénomènes complexes comme la croissance des plantes, la photosynthèse, ou encore la dynamique des écosystèmes. Ces modèles reposent souvent sur des équations différentielles et des principes de la thermodynamique, intégrant des variables environnementales telles que la température, l'humidité et la lumière [143]. En agriculture, cette approche aide à prédire le rendement des cultures en fonction des conditions climatiques [87].

2.3.2.2 Modèles climatiques intégrés

Les modèles climatiques intégrés combinent des simulations de processus physiques, biologiques et socio-économiques pour évaluer les impacts du changement climatique. Ces modèles visent à relier les changements dans les systèmes naturels, comme la température et les précipitations, aux activités humaines telles que les émissions de gaz à effet de serre et l'utilisation des ressources [IPCC2014]. Ils sont utilisés pour explorer différents scénarios d'émissions et d'adaptations, en fournissant des projections de long terme sur l'évolution du climat [126]. En agriculture, ces modèles aident à prédire l'impact des changements climatiques sur les rendements agricoles, tandis qu'en économie, ils permettent d'évaluer les coûts des politiques d'atténuation [75].

2.3.3 Modèles hybrides

Les modèles hybrides combinent différentes approches de modélisation, telles que les modèles empiriques, mécanistes et statistiques, pour améliorer la précision et la robustesse des prédictions. Par exemple, un modèle hybride peut intégrer des données observationnelles avec des modèles physiques pour simuler des phénomènes climatiques complexes [144]. En agriculture, des modèles hybrides combinant les approches de simulation mécanique avec des modèles d'optimisation statistique sont utilisés pour prédire les rendements des cultures en fonction des conditions climatiques [98]. De même, en génie environnemental, des modèles hybrides combinant réseaux de neurones et modèles hydrologiques sont employés pour prévoir les inondations urbaines [159]. Ces modèles permettent ainsi de capturer des dynamiques complexes qui ne peuvent être appréhendées par une seule méthode.

2.3.3.1 Combinaison empirique-mécaniste

La combinaison empirique-mécaniste consiste à intégrer des approches basées sur les données (empiriques) avec des modèles reposant sur des principes fondamentaux (mécanistes) pour améliorer la

précision des prévisions. Par exemple, en agriculture, un modèle hybride peut combiner des équations de croissance des plantes (mécanistes) avec des données historiques sur les rendements pour ajuster les prédictions en fonction des conditions climatiques réelles [65]. De même, en écologie, des modèles de population peuvent intégrer des lois écologiques fondamentales tout en ajustant les paramètres grâce à des observations empiriques pour mieux comprendre les dynamiques des espèces [144]. Cette approche permet de tirer parti de la rigueur théorique des modèles mécanistes et de la flexibilité des modèles empiriques.

2.3.3.2 Modèles basés sur l'apprentissage automatique

Les modèles basés sur l'apprentissage automatique (machine learning) utilisent des algorithmes pour analyser des données, détecter des patterns et faire des prédictions sans nécessiter de règles explicites. Par exemple, en agriculture, des réseaux de neurones peuvent être utilisés pour prédire les rendements des cultures en fonction de facteurs climatiques et environnementaux [93]. De même, en climatologie, des techniques de régression et des forêts aléatoires sont utilisées pour modéliser les changements de température à partir de séries temporelles de données climatiques [98]. Ces modèles permettent de traiter de grandes quantités de données et d'améliorer la précision des prévisions par rapport aux modèles traditionnels.

2.4 Approches de Modélisation des Cultures

Les approches de modélisation des cultures arboricoles, notamment oléicoles, cherchent à intégrer plusieurs paramètres environnementaux pour prédire la croissance et le rendement des plantes. Ces modèles prennent en compte des facteurs climatiques tels que la température, la lumière et les précipitations, ainsi que des paramètres pédologiques comme la texture du sol et la capacité de rétention en eau [111]. Ils intègrent également des variables hydriques, telles que la gestion de l'irrigation, et des aspects variétaux liés aux caractéristiques génétiques des plantes [50]. Par exemple, en oléiculture, les modèles peuvent simuler l'impact du stress hydrique sur la production d'huile, en tenant compte des variations climatiques et de la résistance des différentes variétés d'olivier [7].

2.4.1 Méthodologies employées

Les méthodologies employées pour modéliser les cultures intègrent des aspects du cycle de production, de la production elle-même et du rendement. Le cycle de production inclut les phases de germination, de croissance végétative, de floraison et de maturation, qui sont influencées par des facteurs climatiques et pédologiques [90]. Les modèles de production visent à simuler l'impact de ces facteurs sur la biomasse et la formation de fruits ou graines, tandis que ceux axés sur le rendement intègrent des données historiques et environnementales pour prédire la quantité récoltée [86]. Ces méthodologies utilisent souvent des modèles mathématiques et statistiques pour estimer les effets des pratiques culturales, comme l'irrigation ou la fertilisation, sur le rendement final des cultures [154].

2.4.1.1 Collecte des données

La collecte des données est essentielle pour la modélisation agricole, et elle implique des données numériques, provenant de capteurs et d'outils de télédétection, ainsi que des données sur le terrain collectées directement lors d'observations ou de mesures. L'homogénéisation des données permet de garantir la cohérence et la comparabilité entre différentes sources de données, tout en prenant en compte leur géolocalisation pour l'analyse spatiale [3]. De plus, les données doivent être collectées de manière temporelle afin de suivre l'évolution des phénomènes étudiés sur de longues périodes [41]. Enfin, des données estimées et calculées, telles que les rendements prévisionnels ou les indices de stress hydrique, sont souvent intégrées pour compléter les observations directes et fournir des prédictions fiables [53].

2.4.1.2 Préparation et nettoyage des données

La préparation et le nettoyage des données sont des étapes cruciales dans le traitement des données agricoles. Cela inclut le test de l'homogénéité des données pour s'assurer de leur cohérence et comparabilité, en vérifiant par exemple la constance des unités et des formats [107]. En cas de données manquantes, des techniques de récupération sont employées, comme l'imputation des valeurs manquantes à l'aide de méthodes statistiques ou de machine learning [140]. Le remplacement des données manquantes peut également être effectué par des interpolations ou en utilisant des valeurs moyennes, selon le contexte et la distribution des données [162].

2.4.2 Choix des variables explicatives

Dans le domaine agricole, le choix des variables explicatives joue un rôle crucial pour garantir la pertinence et la précision des modèles prédictifs. Par exemple, dans cette section, nous mettons en avant l'impact de ce choix sur la qualité des prévisions, en réduisant les biais et en améliorant la capacité d'interprétation des résultats. Cette opération sert notamment à sélectionner les facteurs les plus influents, tels que les conditions climatiques, le type de sol ou les pratiques culturales, afin d'optimiser les décisions agricoles. Ce processus peut être réalisé manuellement, basé sur l'expertise des agronomes, ou automatiquement grâce à des méthodes avancées comme la régression pas à pas, l'analyse de corrélation ou encore les algorithmes d'apprentissage automatique [66, 92]. En résumé, combiner ces approches permet de maximiser la fiabilité des modèles tout en tenant compte des spécificités locales et des contraintes pratiques.

2.4.2.1 Variables climatiques

Les variables climatiques telles que la température, les précipitations, l'humidité relative et l'ensoleillement sont parmi les facteurs les plus influents sur le rendement et la production agricole. L'identification de ces variables est cruciale pour prédire la production et optimiser les rendements. Des scénarios d'apprentissage supervisé, utilisant des techniques comme les forêts aléatoires et les réseaux de neurones, peuvent être déployés pour analyser l'impact de ces variables climatiques sur les cultures [41]. En intégrant des données historiques et des prévisions climatiques futures, ces modèles peuvent fournir des prévisions optimales sur les rendements, en tenant compte de différents scénarios climatiques futurs

et des stratégies d'adaptation possibles [164]. Ces approches permettent aux agriculteurs de mieux anticiper les variations de production et de prendre des décisions plus informées pour maximiser les rendements.

2.4.2.2 Facteurs édaphiques et agronomiques

La détermination des variables édaphiques les plus importantes influençant le rendement et la production agricole constitue une étape cruciale dans l'élaboration de modèles prédictifs robustes. Ces paramètres incluent généralement la texture du sol, la capacité de rétention en eau, la teneur en matière organique, le pH, ainsi que la disponibilité des éléments nutritifs essentiels tels que l'azote, le phosphore et le potassium [24]. Ces variables doivent être calibrées selon le contexte spécifique et les caractéristiques des régions étudiées, afin de tenir compte des variations locales des conditions pédologiques et climatiques. L'intégration de scénarios d'apprentissage automatique, tels que les forêts aléatoires ou les réseaux neuronaux profonds, permet d'identifier les facteurs les plus influents en combinant des jeux de données hétérogènes provenant de multiples sources [92]. De plus, des approches de prévision optimales, intégrant ces variables clés, améliorent non seulement la précision des modèles, mais aussi leur capacité à fournir des recommandations adaptées pour chaque région. Cette méthodologie contextuelle est particulièrement pertinente dans le cadre de l'agriculture de précision, où les décisions doivent être prises à des échelles spatio-temporelles fines pour maximiser le rendement tout en préservant les ressources naturelles [114].

2.4.2.3 Facteurs spatial et indices

L'extraction des paramètres géoréférencés, tels que l'altitude, la latitude et la longitude, ainsi que les indices de végétation (comme le NDVI), constitue une étape cruciale dans l'analyse des données de télédétection. Ces paramètres permettent de combiner des facteurs spatiaux, comme la distribution géographique et les caractéristiques topographiques, avec des informations sur l'état des surfaces végétalisées. En utilisant des images issues de satellites ou de drones, cette approche offre un moyen efficace pour étudier les interactions entre les facteurs environnementaux et les dynamiques spatiales, apportant des solutions innovantes dans des domaines tels que l'agriculture de précision, la gestion des ressources naturelles et la surveillance environnementale.

```
1 import numpy as np
2 from sentinelhub import SHConfig, SentinelHubRequest, MimeType, CRS, BBox, DataSource
3
4 config = SHConfig()
5 config.sh_client_id = 'votre_client_id'
6 config.sh_client_secret = 'votre_client_secret'
7
8 latitude, longitude = 40.7128, -74.0060
9
10 # Extraction des données géographiques
11 altitude, coordinates = 100, (latitude, longitude)
12 print(f"Coordonnées : {coordinates}, Altitude : {altitude} m")
13
14 # Calcul du NDVI partir de Sentinel-2
15 def get_ndvi(lat, lon) :
16     bbox = BBox(bbox=[lon-0.01, lat-0.01, lon+0.01, lat+0.01], crs=CRS.WGS84)
17     request = SentinelHubRequest(data_source=DataSource.SENTINEL2_L1C, bbox=bbox, time
18         =('2024-01-01', '2024-01-31'), config=config, bands=['B4', 'B8'])
19     data = request.get_data()
20     red, nir = data[0]['data'][..., 0], data[0]['data'][..., 1]
21     return np.mean((nir - red) / (nir + red))
22
23 ndvi = get_ndvi(latitude, longitude)
24 print(f"NDVI : {ndvi :.2f}")
```

Listing 2.1 – Extraction des paramètres géographiques et calcul du NDVI

2.4.3 Techniques de modélisation utilisées

Les techniques de modélisation jouent un rôle essentiel dans l'analyse des données complexes. Elles peuvent être classées en trois axes principaux : les régressions linéaires et non linéaires, les réseaux de neurones, et le deep learning. L'organigramme suivant illustre cette classification en mettant en évidence les relations et les caractéristiques de chaque approche (voir Figure 2.2).

2.5 Approches de modèles de prédiction

Les approches de modèles de prédiction jouent un rôle central dans l'évaluation des impacts des variations climatiques, notamment sur les systèmes agricoles. Ces modèles se déclinent en plusieurs catégories, telles que les modèles climatiques globaux (GCM), les modèles régionaux (RCM), les modèles stochastiques et les modèles de culture. Les modèles climatiques, en particulier les GCM, permettent de simuler l'évolution future des paramètres climatiques, comme les températures et les précipitations, à l'échelle globale [giorgi1991approaches]. En complément, les modèles régionaux offrent une résolution plus fine adaptée aux spécificités locales [72], tandis que les modèles de culture sont utilisés pour évaluer les effets des changements climatiques sur les rendements agricoles [85]. Ces approches méthodologiques sont souvent couplées pour maximiser la précision des prévisions et élaborer des scénarios d'adapta-

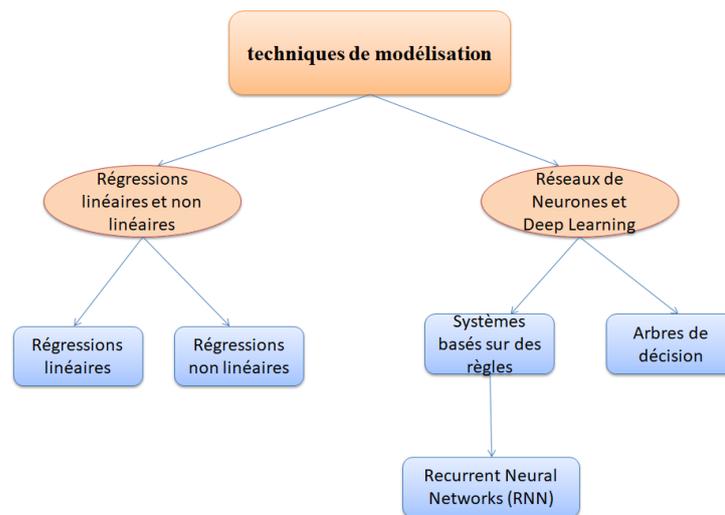


FIGURE 2.2 – Techniques de Modélisation.

tion[144]. Cependant, leur application est confrontée à des défis, notamment l'intégration de données fiables, la gestion des incertitudes et la complexité des interactions climatiques[132]. Malgré ces limites, l'utilisation conjointe de modèles climatiques, stochastiques et de culture constitue une base solide pour anticiper les impacts du changement climatique et proposer des solutions d'adaptation pertinentes et efficaces[34].

2.5.1 Les fonctions des modèles de prédiction

Ce diagramme illustre un processus de prévision basé sur l'analyse de données historiques et climatiques. Il commence par la collecte et le prétraitement des données, suivis par l'application d'algorithmes prédictifs pour simuler des scénarios futurs. Une étape décisionnelle évalue la fiabilité des résultats obtenus : si les résultats ne sont pas satisfaisants, un affinage des modèles est effectué. Enfin, une fois validés, les résultats servent à produire des prévisions et à évaluer les impacts.

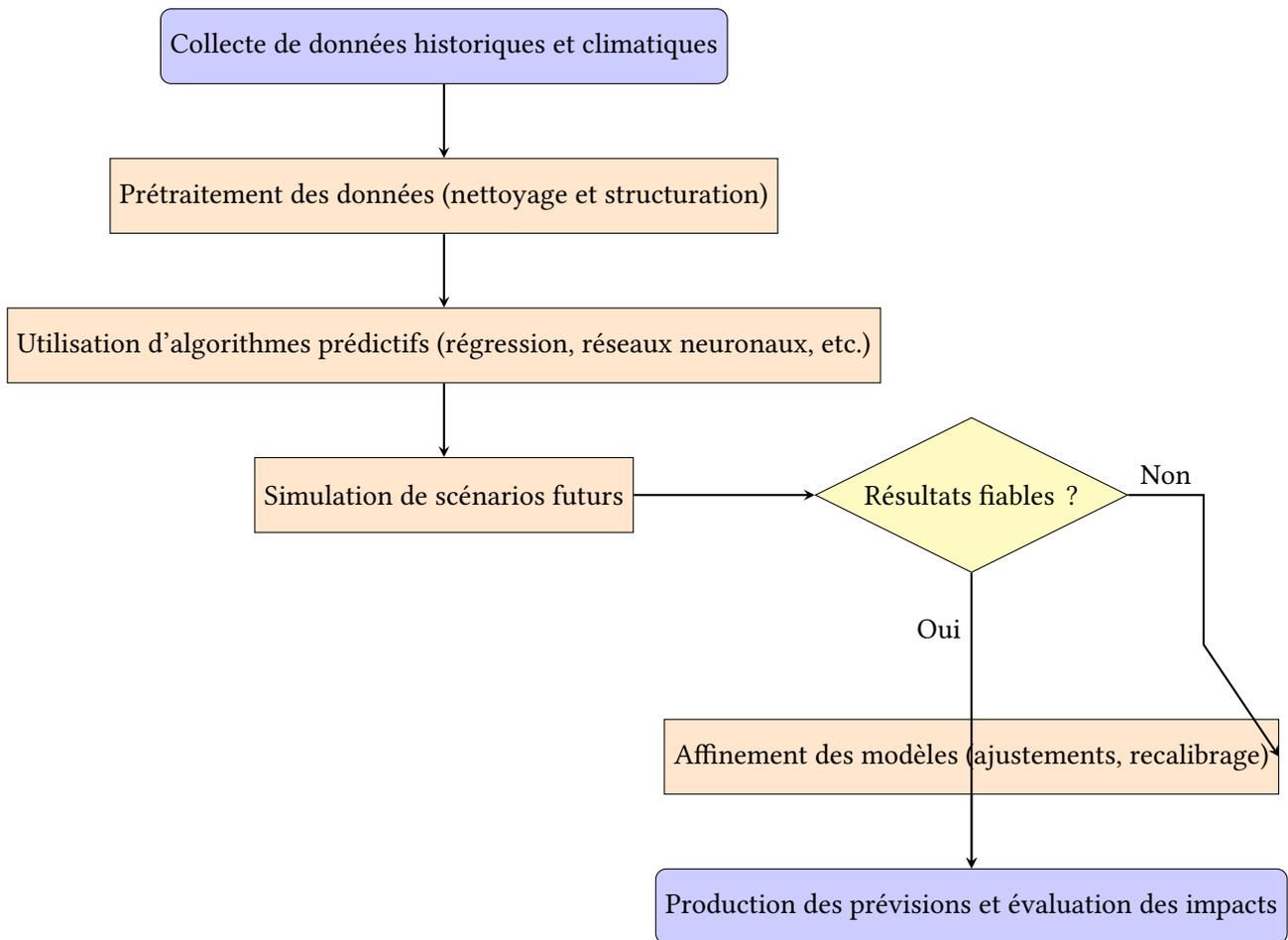


FIGURE 2.3 – Schéma du processus de prévision basé sur l'analyse de données climatiques et historiques

2.6 Différentes formes d'utilisation des modèles

Les modèles prédictifs peuvent être utilisés sous diverses formes selon les besoins et les contextes d'application. Le transfer learning consiste à adapter un modèle préalablement entraîné sur un domaine donné à une nouvelle tâche, permettant de réduire les besoins en données et en calculs[61, 26]. La calibration des modèles, quant à elle, vise à ajuster les paramètres du modèle pour améliorer sa précision sur des ensembles de données spécifiques[63]. Les approches dites « boîte noire » exploitent des modèles complexes, tels que les réseaux neuronaux profonds, dont le fonctionnement interne est difficilement interprétable mais qui offrent souvent une grande performance prédictive[60]. Enfin, le développement de modèles à partir de zéro implique la conception et la formation d'un modèle entièrement nouveau, nécessitant des données abondantes et une compréhension approfondie du problème étudié. Chacune de ces approches présente des avantages et des limitations selon le contexte d'utilisation.

2.6.1 Approche analytique dirigée par les données

L'approche analytique dirigée par les données repose sur l'exploitation massive d'informations quantitatives pour comprendre et prédire les phénomènes complexes. Dans le domaine climatique, ces données incluent les séries temporelles des températures, des précipitations, des niveaux de CO₂, des vitesses de vent et des indices climatiques tels que l'ENSO (El Niño Southern Oscillation)[45, 28]. En agroécologie, les données utilisées englobent les rendements agricoles, les types de cultures, l'humidité des sols, la phénologie des plantes, et les interactions entre les pratiques agricoles et les conditions climatiques. Ces informations sont recueillies via des stations météorologiques, des capteurs au sol, des satellites et des bases de données géospatiales. L'analyse dirigée par les données permet d'identifier des schémas cachés, de tester des hypothèses et de générer des prédictions précises grâce à des outils tels que l'apprentissage automatique et les modèles statistiques, offrant ainsi des solutions pour la gestion durable des ressources naturelles et l'adaptation au changement climatique[9].

2.7 Définitions préliminaires et notation

Cette section offre une vue d'ensemble intégrative des transformateurs (transformers) et du regroupement (clustering). Nous discutons de leurs cadres théoriques et fournissons des formulations mathématiques détaillées et les notations essentielles pour comprendre leur application.

2.7.1 Transformateurs

Les transformateurs sont des architectures avancées de réseaux neuronaux conçues pour traiter efficacement des données séquentielles dans diverses applications[151]. Contrairement aux modèles traditionnels qui traitent les données séquentiellement, les transformateurs utilisent un mécanisme appelé auto-attention (*self-attention*), leur permettant de traiter toutes les parties des données simultanément. Cette capacité de traitement parallèle améliore considérablement l'évolutivité et l'efficacité du modèle en fonction de la longueur et de la complexité des séquences[43].

Pour utiliser les transformateurs, les données doivent être formatées en tokens, qui sont des unités discrètes représentant les composantes essentielles des données :

- **Données temporelles** : Considérées comme des séquences de tokens dans le temps, où chaque point de données est transformé en un vecteur capturant les dynamiques temporelles.

Le transformateur traite les données d'entrée sous forme de matrice de séquences \mathbf{X} , qui subit des transformations via des mécanismes d'auto-attention. L'objectif de l'auto-attention est de calculer une sortie tenant compte de la relation entre chaque token de la séquence d'entrée et tous les autres, en pondérant leur influence en fonction de la tâche donnée.

1. **Matrices de requêtes, clés et valeurs** : Les matrices \mathbf{Q} , \mathbf{K} et \mathbf{V} sont dérivées de la même matrice d'entrée \mathbf{X} , mais transformées à l'aide de différentes matrices de poids :

$$\mathbf{Q} = \mathbf{XW}^Q, \quad \mathbf{K} = \mathbf{XW}^K, \quad \mathbf{V} = \mathbf{XW}^V$$

Ici, \mathbf{W}^Q , \mathbf{W}^K et \mathbf{W}^V sont les matrices de poids spécifiques aux requêtes, clés et valeurs respecti-

vement.

2. **Calcul de l'auto-attention** : Le mécanisme d'auto-attention ajuste l'attention reçue par chaque token en fonction de l'affinité requête-clé, puis utilise cette attention pour former une somme pondérée des vecteurs de valeur. La formule de calcul est la suivante :

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}$$

La fonction *softmax* garantit que les scores d'attention, représentant les poids d'influence, s'additionnent à 1.

3. **Scores d'attention** : Le score d'attention α_{ij} pour la i -ème requête et la j -ème clé est calculé par la formule :

$$e_{ij} = \frac{(\mathbf{x}_i \mathbf{W}^Q)(\mathbf{x}_j \mathbf{W}^K)^T}{\sqrt{d_k}}$$

où e_{ij} est le produit scalaire normalisé entre la i -ème requête et la j -ème clé, et d_k est la dimension des vecteurs clé. Le facteur d'échelle $\sqrt{d_k}$ aide à stabiliser les gradients pendant l'entraînement. Le score final α_{ij} est obtenu en normalisant ces valeurs à l'aide de la fonction *softmax* :

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})}$$

2.7.2 Clustering

Le *clustering* est une technique utilisée pour regrouper un ensemble d'objets de manière à ce que les objets d'un même groupe (appelé cluster) soient plus similaires entre eux qu'avec ceux d'autres groupes. C'est une méthode d'apprentissage non supervisé particulièrement utile pour identifier des modèles distincts à partir de grands ensembles de données sans étiquetage préalable. Dans le contexte de la consommation d'électricité, les algorithmes de clustering peuvent identifier et regrouper des profils d'utilisation similaires parmi différents consommateurs. Cela permet, par exemple, aux compagnies d'électricité d'adapter leurs stratégies d'engagement client et d'optimiser la production et la distribution d'énergie [82, 80].

L'objectif du clustering est de minimiser la variance intra-cluster, souvent formulée comme suit :

$$\min_C \sum_{i=1}^k \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \mu_i\|^2$$

où μ_i représente le centroïde du cluster C_i :

$$\mu_i = \frac{1}{|C_i|} \sum_{\mathbf{x} \in C_i} \mathbf{x}$$

2.7.3 Notation

Voici un tableau résumant les notations utilisées dans les descriptions.

Symbole	Description
X	Matrice de séquences d'entrée pour les transformateurs, chaque colonne représentant un vecteur token.
W^Q	Matrice de poids pour les requêtes dans le mécanisme d'auto-attention des transformateurs.
W^K	Matrice de poids pour les clés dans le mécanisme d'auto-attention des transformateurs.
W^V	Matrice de poids pour les valeurs dans le mécanisme d'auto-attention des transformateurs.
Q	Matrice des requêtes dérivée de la matrice d'entrée X dans les transformateurs.
K	Matrice des clés dérivée de la matrice d'entrée X dans les transformateurs.
V	Matrice des valeurs dérivée de la matrice d'entrée X dans les transformateurs.
d_k	Dimension des vecteurs clé dans les transformateurs, utilisée pour normaliser les produits scalaires.
e_{ij}	Logit d'attention brut calculé entre la i -ème requête et la j -ème clé dans les transformateurs.
α_{ij}	Score d'attention pour la i -ème requête par rapport à la j -ème clé dans les transformateurs.
μ_i	Centroïde du cluster C_i dans le clustering.
C	Ensemble de tous les clusters dans le clustering.
$ C_i $	Nombre d'éléments dans le cluster C_i dans le clustering.
x	Notation générale pour un point de données ou un vecteur dans le contexte des transformateurs et du clustering.

2.7.4 Clustering K-means

L'analyse de clustering K-means est un processus d'apprentissage non supervisé qui identifie des motifs potentiellement similaires dans un ensemble de données sans information préalable [81]. L'algorithme vise à partitionner un ensemble de données en k clusters, où chaque cluster est défini par son centroïde. L'objectif est de minimiser la somme des distances quadratiques (erreurs) entre les points de données et leurs centroïdes respectifs [108]. Cela est mathématiquement représenté comme suit :

$$SSE(C) = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2$$

où $C = \{C_1, C_2, \dots, C_k\}$ représente l'ensemble des clusters, x_j sont les points de données, et μ_i sont les centroïdes des clusters. L'objectif est de trouver la configuration des clusters C^* qui minimise la SSE :

$$C^* = \arg \min_C SSE(C)$$

L'algorithme K-means suit un processus itératif composé de deux étapes principales :

1. **Étape d'affectation** : Chaque point de données est affecté au centroïde de cluster le plus proche en fonction de la distance euclidienne.
2. **Étape de mise à jour** : Les centroïdes des clusters sont recalculés comme la moyenne de tous les points de données affectés.

Ces étapes sont répétées jusqu'à convergence, généralement atteinte lorsque les centroïdes ne changent plus significativement entre les itérations. L'algorithme peut converger vers un minimum local ; il est donc souvent exécuté plusieurs fois avec des initialisations différentes pour trouver le meilleur clustering [64].

Le K-means est efficace sur le plan computationnel et adapté aux grands ensembles de données, bien que la qualité du clustering dépende de la position initiale des centroïdes. La complexité temporelle de l'algorithme est $O(tnkd)$, où n est le nombre de points de données, k est le nombre de clusters, d est le nombre de dimensions, et t est le nombre d'itérations [161].

2.7.5 Régressions linéaires et non linéaires

2.7.6 ARIMA

Le modèle Auto-Regressive Integrated Moving Average (ARIMA), développé par Box et Jenkins, est une technique populaire pour la prévision des séries temporelles. Les modèles ARIMA modélisent les données observées x_t en les considérant comme une combinaison de leurs valeurs passées et des erreurs de prévision passées [23]. Cette méthode suppose que la série temporelle peut être décomposée en composantes de tendance, saisonnières et irrégulières.

Plutôt que de modéliser ces composantes séparément, ARIMA transforme la série temporelle à travers des différenciations pour la rendre stationnaire, supprimant ainsi efficacement les tendances et la saisonnalité [76]. Une fois la stationnarité atteinte, le modèle ARIMA utilise une autoregression (AR) et des moyennes mobiles (MA) pour décrire la dynamique des séries temporelles. Le modèle est paramétré par trois entiers (p, d, q) , où p est le nombre d'observations retardées dans le modèle (partie autorégressive), d est le nombre de différenciations appliquées aux observations brutes, et q est la taille de la fenêtre de moyenne mobile.

Le modèle ARIMA est largement utilisé en raison de sa flexibilité et de son efficacité à gérer une variété de modèles de séries temporelles, ce qui en fait un outil fondamental pour l'analyse des séries temporelles.

2.7.6.1 Apprentissage automatique

Soit $\mathcal{D} = (X_i, Yield_i)_1^N$ un ensemble de taille N , où $X_i = (X_{i_1}, \dots, X_{i_n})$ représente l'ensemble des caractéristiques avec n paramètres décrivant les conditions de production agricole, et $Yield_i$ représente le rendement ou la production cible associée.

Dans ce contexte, X_i inclut des paramètres liés à la production agricole : *paramètres climatiques*, *paramètres agronomiques*, *paramètres de gestion des cultures*, et *paramètres des ressources*. Nous appelons

P_i l'espace vectoriel des caractéristiques. Soit Y_{i_m} l'ensemble des structures physiques possibles des cultures (par exemple, ++ = production élevée, + = production moyenne, = = production stable, - = production faible) correspondant à une configuration agricole donnée.

Notre objectif est de prédire la production agricole (*yield*) en fonction des caractéristiques de la configuration agricole. Pour cela, nous utilisons un modèle de machine learning M défini comme une fonction $f(X_i)$, qui prédit la catégorie de rendement (Y_{i_m}) pour une configuration donnée X_i . L'objectif est également d'extraire des règles généralisées permettant de mieux comprendre les facteurs influençant la production. Les règles de prévision sont définies à partir de $(X_{Climate}, X_{Crop}, X_{Manag}, X_{Res})$, où :

1. $X_{Climate}$ est l'ensemble des paramètres climatiques, tels que les précipitations, la température minimale/moyenne/maximale, l'humidité, la durée d'ensoleillement, etc.,
2. X_{Crop} est l'ensemble des caractéristiques agronomiques liées aux cultures, comme le type de culture, la densité de plantation, le stade phénologique, les types de semences utilisées, et les variétés cultivées,
3. X_{Manag} représente les paramètres de gestion des cultures, comme les pratiques agricoles (*labour, irrigation, fertilisation, traitements phytosanitaires*), les techniques de récolte, et les rotations culturales,
4. X_{Res} représente les paramètres liés aux ressources, tels que la qualité du sol (pH, matière organique), l'accès à l'eau, les équipements agricoles, ainsi que les infrastructures logistiques.

Le modèle M est entraîné à partir des données historiques contenant les relations entre X_i et $Yield_i$. Une fois le modèle construit, il peut être utilisé pour :

- Prédire le rendement agricole pour différentes configurations climatiques et pratiques agricoles,
- Identifier les combinaisons optimales de paramètres $(X_{Climate}, X_{Crop}, X_{Manag}, X_{Res})$ pour maximiser la production,
- Aider les agriculteurs à planifier et à prendre des décisions éclairées basées sur les prévisions obtenues.

2.7.7 Réseaux de neurones et deep learning

2.7.8 Réseaux de Neurones Récurrents (RNNs)

Les réseaux de neurones récurrents (*Recurrent Neural Networks*, RNNs) sont un type spécialisé de réseau neuronal conçu pour traiter des données séquentielles, telles que la prévision de séries temporelles, la modélisation du langage et la reconnaissance vocale [60]. Contrairement aux réseaux de neurones à propagation directe (*feedforward neural networks*), les RNNs comportent des boucles dans leur architecture, ce qui leur permet de conserver une mémoire des entrées précédentes. Cette capacité est particulièrement efficace pour les tâches impliquant des séquences temporelles, où l'ordre et le contexte des points de données sont essentiels.

Comme illustré à la Figure Figure 2.4, un RNN traite une séquence d'entrée $x = (x_1, x_2, \dots, x_T)$, où chaque x_t représente un vecteur de dimension d à l'instant t . Le réseau calcule un état caché h_t pour chaque instant, mis à jour de manière récursive à l'aide de l'équation suivante :

$$h_t = f_h(W_x x_t + W_h h_{t-1} + b_h),$$

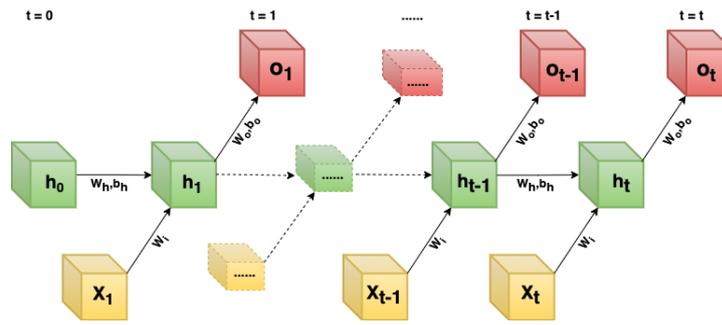


FIGURE 2.4 – Recurrent Neural Networks .

où W_x et W_h sont des matrices de poids, b_h est un vecteur de biais, et f_h est une fonction d'activation non linéaire telle que tanh ou ReLU. La sortie à chaque instant o_t est ensuite dérivée de l'état caché à l'aide de :

$$o_t = f_o(W_o h_t + b_o),$$

où W_o est la matrice de poids de sortie et b_o le vecteur de biais de sortie.

Cette architecture permet aux RNN de capturer les dépendances entre différents instants, ce qui les rend adaptés à l'apprentissage de motifs complexes dans les séquences. Cependant, les RNN standards rencontrent souvent des difficultés avec les dépendances à long terme en raison de problèmes tels que l'atténuation ou l'explosion des gradients. Pour atténuer ces problèmes, des architectures avancées comme les réseaux Long Short-Term Memory (LSTM) et les Gated Recurrent Units (GRU) ont été développées. Ces architectures introduisent des mécanismes de portes permettant de mieux gérer le flux d'information et de maintenir les dépendances à long terme de manière plus efficace.

2.7.9 Long Short-Term Memory (LSTM) Networks

Pour répondre aux limitations des RNN mentionnées précédemment, les réseaux Long Short-Term Memory (LSTM) ont été développés [68]. Les LSTM améliorent la structure de base des RNN en intégrant une série de portes qui régulent le flux d'information, leur permettant ainsi de maintenir plus efficacement les dépendances à long terme.

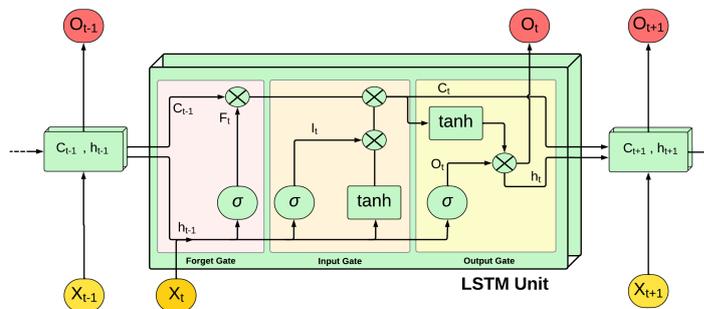


FIGURE 2.5 – Long Short-Term Memory (LSTM) Networks.

Les LSTM introduisent trois portes principales : la porte d'entrée, la porte d'oubli et la porte de sortie. Ces portes modulent les informations ajoutées, conservées ou sorties de l'état de la cellule à chaque instant :

1. **Porte d'entrée (Input Gate)** : Contrôle la quantité de nouvelle information provenant de l'entrée actuelle et de l'état caché précédent qui doit être ajoutée à l'état de la cellule.
2. **Porte d'oubli (Forget Gate)** : Détermine dans quelle mesure l'information de l'état de la cellule précédent doit être conservée ou oubliée.
3. **Porte de sortie (Output Gate)** : Décide quelle partie de l'état de la cellule doit être sortie comme l'état caché à l'instant actuel.

Ces portes permettent aux LSTM de surmonter les problèmes d'atténuation et d'explosion des gradients, les rendant plus robustes pour le traitement des longues séquences par rapport aux RNN traditionnels, comme illustré dans la représentation schématique d'une unité LSTM à la Figure Figure 2.5. L'état de la cellule c_t dans un LSTM est mis à jour comme suit :

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t,$$

où f_t est la porte d'oubli, i_t est la porte d'entrée, et \tilde{c}_t est l'état de cellule candidat. L'état caché h_t est calculé à l'aide de la porte de sortie o_t :

$$h_t = o_t \odot \tanh(c_t),$$

où o_t représente la porte de sortie. Ces mécanismes permettent aux LSTM de maintenir et de modifier efficacement l'état de la cellule, capturant ainsi les dépendances à long terme dans les données.

2.7.10 Unités Récurrentes à Porte (GRU)

Les Unités Récurrentes à Porte (Gated Recurrent Units) (GRU) sont un type de réseaux de neurones récurrents (RNN) à portes, introduit pour la première fois par Cho et al. [36]. Les GRU, comme illustré à la Figure Figure 2.6, sont une version simplifiée des réseaux Long Short-Term Memory (LSTM), conçue pour résoudre certains des mêmes défis, tels que l'apprentissage des dépendances à long terme dans les données séquentielles.

La principale différence entre les GRU et les LSTM est que les GRU combinent les fonctions de la porte d'oubli et de la porte d'entrée en une seule porte, ce qui aboutit à une architecture plus simplifiée avec seulement deux portes : la porte de mise à jour et la porte de réinitialisation.

1. **Porte de mise à jour** : Cette porte détermine la quantité d'information provenant du passé qui doit être transmise vers l'avenir. Elle fonctionne de manière similaire à l'effet combiné des portes d'oubli et d'entrée dans un LSTM, décidant quelles informations conserver et quelles informations ignorer.
2. **Porte de réinitialisation** : Cette porte contrôle la quantité d'information passée à oublier. Elle permet au modèle de réinitialiser l'état caché lorsque cela est nécessaire, garantissant que les informations non pertinentes sont supprimées de la mémoire du réseau.

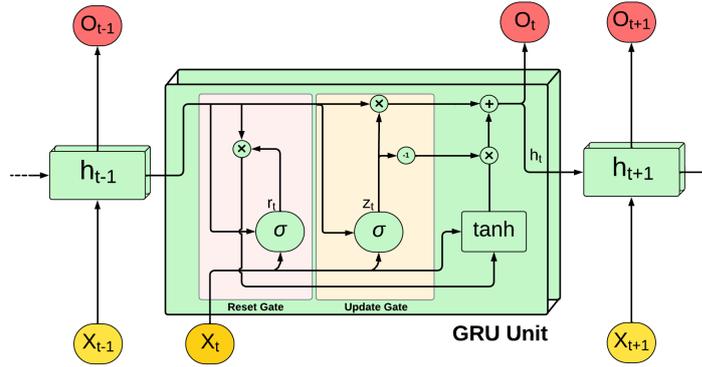


FIGURE 2.6 – Gated Recurrent Units (GRU).

En raison de cette structure simplifiée, les GRU sont computationnellement plus efficaces que les LSTM, tout en offrant de puissantes capacités pour capturer les dépendances dans les données séquentielles. L'état de la cellule et l'état caché dans un GRU sont combinés en un seul état, ce qui simplifie l'architecture et réduit le nombre de paramètres à entraîner.

Les équations de mise à jour pour un GRU sont les suivantes :

$$z_t = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z)$$

$$r_t = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{r}_t \odot \mathbf{U}_h \mathbf{h}_{t-1} + \mathbf{b}_h)$$

$$\mathbf{h}_t = (1 - z_t) \odot \mathbf{h}_{t-1} + z_t \odot \tilde{\mathbf{h}}_t$$

où z_t est la porte de mise à jour, r_t est la porte de réinitialisation, \mathbf{h}_t est l'état caché à l'instant t , \mathbf{x}_t est l'entrée à l'instant t , σ est la fonction sigmoïde, et \odot désigne la multiplication élément par élément.

Les GRU ont montré de bonnes performances dans une variété de tâches impliquant des données séquentielles, telles que la modélisation de langage, la reconnaissance vocale et la prédiction de séries temporelles, offrant un équilibre entre la complexité du modèle et la performance.

2.7.11 Transformers

2.7.11.1 Vanilla Transformer

Le modèle Transformer de base, introduit par deghani et al. (2018)[42], représente une avancée significative dans les tâches de modélisation de séquences. Contrairement aux réseaux de neurones récurrents (RNN) traditionnels ou aux réseaux Long Short-Term Memory (LSTM), le Transformer

ne repose pas sur la récursion pour traiter les informations séquentielles. Au lieu de cela, il utilise une architecture novatrice basée sur des mécanismes d'auto-attention, lui permettant de gérer les dépendances indépendamment de leur distance dans la séquence d'entrée.

Le Transformer comprend un encodeur et un décodeur, chacun composé de plusieurs couches identiques. Chaque couche de l'encodeur inclut un mécanisme d'auto-attention multi-têtes suivi d'un réseau de neurones feed-forward positionnel. En revanche, chaque couche du décodeur intègre un mécanisme d'attention croisée supplémentaire entre l'auto-attention multi-têtes et le réseau feed-forward. Ce mécanisme d'attention double permet au modèle de capturer et d'utiliser efficacement l'information provenant de l'ensemble de la séquence d'entrée [151].

2.7.11.2 Encodage des entrées et encodage positionnel

Pour conserver l'information sur l'ordre séquentiel, le modèle Transformer utilise un encodage positionnel ajouté aux embeddings d'entrée. Cette approche est cruciale car le mécanisme d'auto-attention lui-même est invariant par permutation, ce qui signifie qu'il ne capture pas intrinsèquement l'ordre de la séquence.

1. Encodage positionnel absolu (Encodage Positionnel Absolu) : Dans le Transformer de base, les vecteurs d'encodage positionnel sont générés à l'aide de fonctions sinus et cosinus de fréquences différentes pour chaque dimension, comme défini par les équations suivantes :

$$PE_{(pos,i)} = \begin{cases} \sin\left(\frac{pos}{10000^{\frac{i}{d}}}\right) & \text{if } i\%2 = 0 \\ \cos\left(\frac{pos}{10000^{\frac{i}{d}}}\right) & \text{if } i\%2 = 1 \end{cases}$$

où pos est la position, i est la dimension, et d est le nombre total de dimensions. Cet encodage permet au modèle d'incorporer des informations sur la position de chaque token dans la séquence, lui permettant ainsi de distinguer les tokens situés à différentes positions.

2. **Encodage positionnel relatif (Encodage Positionnel Relatif)** : Certaines méthodes se concentrent sur les relations positionnelles par paires entre les éléments d'entrée, en ajoutant des embeddings positionnels relatifs apprenables aux clés du mécanisme d'attention [138]. Il existe également des approches hybrides qui combinent à la fois les encodages positionnels absolus et relatifs [133].

2.7.11.3 Attention Multi-Têtes (Multi-Head Attention)

Le mécanisme d'attention multi-têtes est un composant clé de l'architecture Transformer, permettant au modèle de se concentrer simultanément sur différentes parties de la séquence d'entrée [151]. Chaque tête d'attention apprend à se focaliser sur différents aspects de l'entrée, améliorant ainsi la capacité du modèle à capturer des relations complexes et contextuelles [138]. En combinant les résultats de ces têtes, le Transformer peut mieux gérer les dépendances à longue distance et capturer des informations à différentes échelles.

Le mécanisme d'attention utilise le modèle Query-Key-Value (QKV), et l'attention par produit scalaire mis à l'échelle est définie comme suit :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_k}}\right)V$$

où $Q \in \mathbb{R}^{N \times D_k}$, $K \in \mathbb{R}^{M \times D_k}$, et $V \in \mathbb{R}^{M \times D_v}$. Ici, N et M représentent respectivement les longueurs des requêtes et des clés (ou des valeurs), et D_k et D_v représentent les dimensions des clés (ou des requêtes) et des valeurs, respectivement.

L'attention multi-têtes étend le mécanisme d'auto-attention en utilisant plusieurs ensembles de projections apprenables :

$$\text{MultiHeadAttn}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

où chaque tête d'attention head_i est calculée comme suit :

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Dans cette formulation, W_i^Q , W_i^K , et W_i^V sont les matrices de poids apprenables pour la i -ème tête, et W^O est une matrice de poids apprenables qui projette les sorties concaténées des multiples têtes d'attention vers la dimensionnalité d'origine.

Cette structure permet au modèle de prêter attention simultanément à l'information provenant de différents sous-espaces de représentation à différentes positions, améliorant ainsi de manière significative sa capacité d'apprentissage.

2.7.11.4 Réseau Feed-forward et Réseau Résiduel

Le réseau feed-forward dans une couche Transformer est un réseau entièrement connecté appliqué à chaque position de manière séparée et identique :

$$\text{FFN}(H_0) = \text{ReLU}(H_0W_1 + b_1)W_2 + b_2$$

where H_0 is the output of the previous layer, $W_1 \in \mathbb{R}^{D_m \times D_f}$, $W_2 \in \mathbb{R}^{D_f \times D_m}$, $b_1 \in \mathbb{R}^{D_f}$, and $b_2 \in \mathbb{R}^{D_m}$ are trainable parameters.

Chaque couche dans un Transformer inclut également des connexions résiduelles et une normalisation de couche pour garantir un meilleur flux de gradient et une convergence plus rapide :

$$H_0 = \text{LayerNorm}(\text{SelfAttn}(X) + X)$$

$$H = \text{LayerNorm}(\text{FFN}(H_0) + H_0)$$

where $\text{SelfAttn}(\cdot)$ denotes the self-attention mechanism and $\text{LayerNorm}(\cdot)$ denotes layer normalization.

2.7.11.5 Innovations Récentes dans les Modèles Transformer

Le modèle Transformer, introduit par Vaswani et al. [151], représente une avancée majeure dans la modélisation des séquences, utilisant un cadre encodeur-décodeur. Ce modèle a obtenu des résultats exceptionnels dans divers domaines, y compris le traitement du langage naturel (NLP) [43], la reconnaissance vocale [46], et la vision par ordinateur [106]. Plus récemment, les modèles basés sur Transformer sont devenus de plus en plus populaires pour l'analyse de séries temporelles [155]. La figure Figure 2.7 présente tous les modèles phares qui s'attaquent à la tâche complexe de prévision de séries temporelles (TSF).

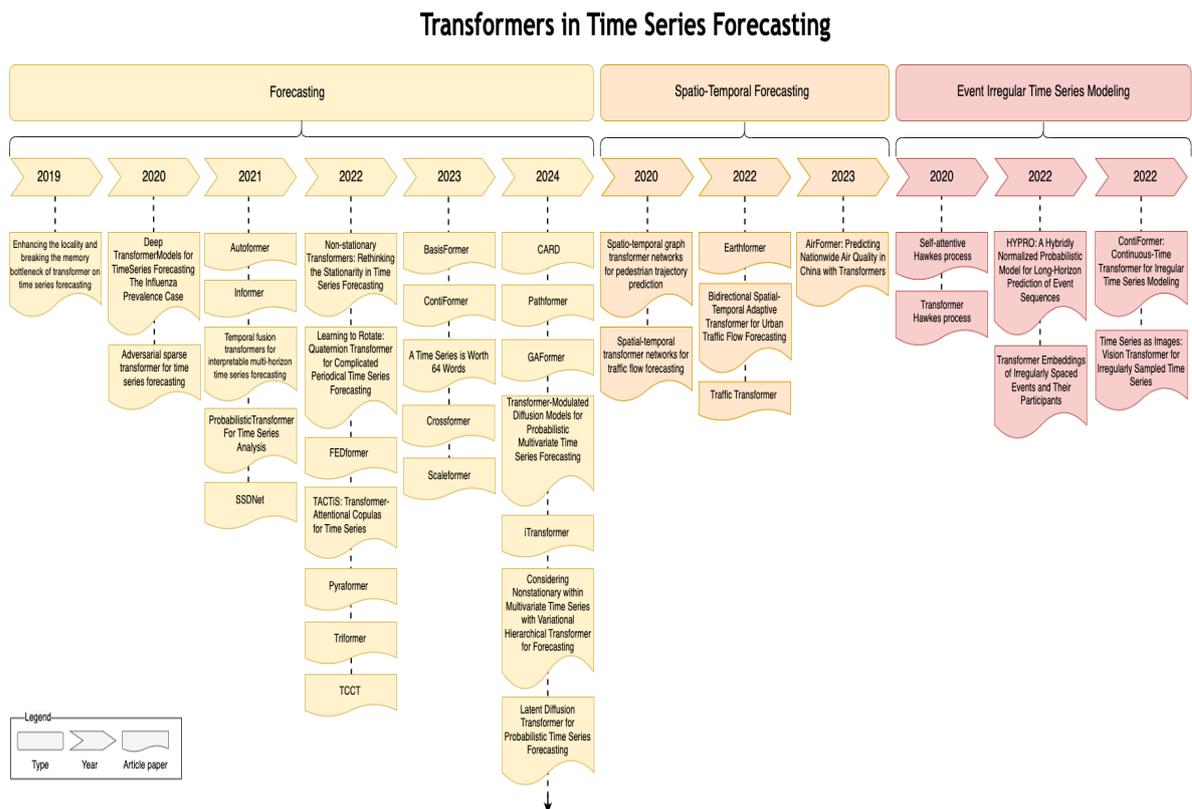


FIGURE 2.7 – Overview of Transformer Models.

En 2024, plusieurs nouveaux modèles Transformer ont émergé, montrant des résultats prometteurs dans la prévision de séries temporelles. Ces modèles incluent :

- **CARD** [153] : Utilise l'alignement des canaux pour améliorer de manière robuste la précision

des prévisions.

- **Pathformer** [33] : Utilise des chemins adaptatifs pour gérer efficacement les données multi-échelles.
- **GAFormer** [158] : Améliore la compréhension des séries temporelles grâce à des embeddings prenant en compte les groupes.
- **Transformer-Modulated Diffusion Models for Probabilistic Multivariate Time Series Forecasting** [101] : Combine les architectures Transformer avec des modèles de diffusion pour des prévisions probabilistes.
- **iTransformer** [105] : Propose une architecture Transformer inversée pour de meilleures performances de prévision.

Ces avancées illustrent l'évolution continue et l'adaptation des modèles Transformer à diverses tâches complexes dans la prédiction de séries temporelles.

2.8 Évaluation des Modèles de Prédiction

L'évaluation des modèles de prédiction en climatologie consiste à quantifier la précision des prédictions des variables climatiques (températures, précipitations, etc.) à l'aide de métriques adaptées. Cette évaluation repose sur la comparaison entre les valeurs prédites (\hat{y}) et les valeurs observées (y). Les principales étapes incluent la division des données, le choix des métriques, et l'interprétation des résultats.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2, \quad \text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Étape	Description	Formules
Division des données	Séparer les données en ensembles d'entraînement, validation et test.	-
Évaluation des erreurs	Mesurer la différence entre les prédictions et les observations.	$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2, \quad \text{MAE} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i - y_i $
Analyse de la variance expliquée	Évaluer la proportion de variance capturée par le modèle.	$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$
Validation croisée	Vérifier la robustesse du modèle sur plusieurs sous-ensembles de données.	-

Tableau 2.2 – Étapes et métriques d'évaluation des modèles de prédiction en climatologie.

2.8.1 classification et comparaisons des travaux

La prédiction précise de la production agricole reste un défi majeur en raison de la nécessité d'identifier et de comprendre les paramètres clés ainsi que leurs interactions complexes. Des modèles efficaces

doivent refléter les conditions environnementales régionales et tenir compte de la nature dynamique des modèles de consommation agricoles, tels que les changements dans la demande du marché et les tendances de consommation. De plus, les variations dans les types de cultures, comprenant les différences dans les cycles de croissance, la résistance aux maladies et les pratiques agricoles diverses comme l'irrigation et la gestion des sols, contribuent encore à la variabilité de la production et aux risques associés.

Ainsi, les modèles de prévision doivent être constamment mis à jour et améliorés en intégrant de nouvelles données et en anticipant les facteurs émergents qui pourraient influencer les résultats futurs. Dans ce cadre, nous introduisons un cadre de modélisation basé sur les données appelé AFFAC (Advanced Forecasting Framework for Agricultural Cultures) pour améliorer l'exactitude des prévisions de la production agricole.

Le processus d'élaboration de ce cadre inclut une analyse semi-automatique, où le concepteur et/ou les utilisateurs interviennent à des étapes clés, lorsque l'expertise du concepteur et la validation par les utilisateurs sont nécessaires. Ce processus est formalisé par plusieurs algorithmes. Nous estimons que l'annotation multidimensionnelle, dans le contexte agricole, doit être réalisée dès la phase conceptuelle, en s'appuyant sur des modèles définis indépendamment de toute contrainte d'implémentation. Ces modèles incluent des données sur les besoins en termes de production agricole et les sources de données relatives aux conditions environnementales, aux pratiques agricoles, et aux demandes du marché.

Nous plaçons ensuite notre approche dans le cadre des phases du cycle de conception des systèmes de prévision. Lors de la phase de "définition des besoins", plusieurs éléments doivent être pris en compte, notamment la collecte de données orientée vers les besoins des utilisateurs, qui s'effectue souvent en fonction des objectifs spécifiques de la prévision. Les utilisateurs finaux, tels que les agriculteurs et les décideurs, expriment naturellement leurs besoins sous forme de buts à atteindre pour optimiser la production et la consommation.

Une représentation formelle des besoins est cruciale pour la validation des modèles de prévision. Nous proposons d'utiliser des ontologies pour structurer ces besoins de manière indépendante des contraintes d'implémentation. Cela permet une validation plus rigoureuse des besoins et facilite l'inférence de nouvelles connaissances relatives à la gestion des cultures.

Lorsque la collecte des besoins est orientée vers les sources de données agricoles, les travaux antérieurs se sont souvent limités à l'utilisation de bases de données classiques, souvent relationnelles. Dans notre approche, nous proposons une méthode mixte qui confronte les données des sources à priori aux besoins des utilisateurs dès le début du processus de conception. Nous considérons ici différentes sources de données, y compris les données climatiques, les pratiques agricoles et les tendances de marché.

La phase suivante, la "modélisation conceptuelle", est essentielle pour structurer les relations entre les différents paramètres agricoles (climat, sol, culture, marché). Nous intégrons dans notre méthode une approche ontologique qui permet de mieux capturer les interactions complexes entre ces paramètres. Ce formalisme ontologique présente un pouvoir expressif plus important que les formalismes conceptuels classiques (par exemple, E/A, UML), et permet ainsi une meilleure modélisation des dynamiques agricoles.

Enfin, dans la phase de "modélisation logique" et "physique", nous proposons un cadre flexible permettant de déployer des modèles de prévision à la carte, selon les besoins spécifiques des utilisateurs et les différentes architectures de stockage des données. Cette approche offre une représentation plus

détaillée et plus dynamique du processus de prévision, en permettant une intégration efficace des schémas conceptuels et logiques dans une architecture multicouche, favorisant ainsi la persistance des données et des modèles utilisés pour la prévision de la production agricole.

2.9 Bilan et Synthèse

En passant en revue la littérature et les différents usages des modèles de prévision de la production agricole (MdP), nous avons remarqué qu'ils sont soit produits ou consommés par la communauté scientifique et les praticiens agricoles. Les producteurs de MdP sont généralement des chercheurs, des étudiants en agronomie, et des professionnels de l'industrie agroalimentaire. D'autre part, les consommateurs de MdP sont des agriculteurs, des conseillers agricoles, et des étudiants.

Dans la Section 3.2.1, nous avons mis en évidence trois types de consommation des MdP :

- Le premier type correspond à l'utilisation d'un MdP comme une boîte noire, fournie par des outils de prévision comme les conseillers agricoles automatisés.
- Le second type se réfère à la réutilisation des MdP existants, où ces modèles peuvent être adaptés ou étendus pour tenir compte des spécificités du contexte agricole, comme le type de culture ou les conditions climatiques locales.
- Le troisième type est un mode d'apprentissage, où des étudiants ou des chercheurs peuvent consulter un MdP pour étudier sa structure, ses composants, et ses résultats dans un but pédagogique.

Cependant, en analysant la littérature, nous constatons que le processus de développement de MdP nécessite une connaissance approfondie de plusieurs domaines : agronomie, gestion des ressources naturelles, apprentissage machine pour calibrer les paramètres, statistiques pour estimer des facteurs comme les rendements des cultures, etc. En effet, l'utilisation actuelle des MdP reste complexe, longue, et sujette à des erreurs pour plusieurs raisons :

- Pour savoir si un modèle de prévision existe pour une configuration agricole spécifique (par exemple, un type de culture ou une région géographique), il faut rechercher dans une vaste littérature un modèle adapté à ses besoins.
- Nous avons constaté qu'il existe une grande variété de modèles de prévision, parfois bien expliqués dans des articles scientifiques, mais souvent noyés dans des discussions peu claires.
- Pour réutiliser un MdP existant pour produire ou reproduire des résultats de prévision, il est souvent nécessaire d'extraire manuellement les paramètres et les formules mathématiques des documents de recherche. Une autre possibilité consiste à modifier le code d'un outil de gestion agricole (quand cela est possible), mais cela nécessite une expertise technique dans le système utilisé.

Cette situation constitue un défi majeur pour les concepteurs de MdP et est en contradiction avec l'ère du partage dans laquelle nous vivons, où la plupart des données et des modèles sont accessibles publiquement. Les MdP ne doivent pas être une exception à cette règle de partage et de réutilisation. Ainsi, nous pensons que la création d'un référentiel ouvert encouragera les chercheurs et les praticiens agricoles à partager, réutiliser, étendre et comparer les modèles de prévision.

Nos travaux dans ce cadre permettent de répondre aux questions suivantes :

1. Comment trouver un modèle de prévision adapté à un besoin spécifique dans un contexte agricole donné ?
2. Comment exprimer ce besoin de manière claire et précise ?
3. Comment obtenir et partager des modèles de prévision à travers un référentiel ouvert ?
4. Comment exprimer les modèles de prévision sous des formats interprétables par machine pour faciliter leur utilisation ?

À notre connaissance, il n'existe pas de travaux dans la littérature traitant spécifiquement de ces questions pour guider les utilisateurs de systèmes de prévision agricoles dans leur utilisation des MdP, ni de référentiel permettant aux chercheurs de comparer leurs modèles avec ceux déjà existants. L'objectif de cette recherche n'est pas seulement de réduire la difficulté de l'utilisation des MdP dans l'évaluation des rendements agricoles, mais aussi de fournir un outil éducatif pour les chercheurs et les étudiants dans le domaine de l'agronomie ou l'agroécologie.

2.10 intégration des données multi sources

Un modèle prédictif pour optimiser les rendements agricoles en intégrant des données provenant de diverses sources (parcelles agricoles, météorologie, géolocalisation) et en appliquant quatre types d'analyse : descriptive, diagnostique, prédictive et prospective. L'analyse descriptive identifiera les tendances historiques, tandis que l'analyse diagnostique explorera les facteurs influençant les rendements[124]. L'analyse prédictive utilisera des modèles de machine learning pour anticiper les rendements futurs, et l'analyse prospective simulera des scénarios afin de guider les agriculteurs dans l'adoption de stratégies adaptées[149, 77]. Le modèle tiendra compte des contraintes spécifiques des exploitations agricoles (disponibilité des ressources, saisonnalité, limites financières) et des préférences des agriculteurs (types de culture, approches durables[99]). L'approche Data-Driven Approach Automation (DDAA) intégrera le machine learning et le deep learning pour automatiser des tâches en imitant les actions humaines. L'acquisition de données provenant de multiples sources, y compris des journaux de données et de processus, permettra une compréhension plus complète des systèmes agricoles. Les objectifs sont de fournir un outil analytique pour aider les agriculteurs à mieux diagnostiquer, prévoir et anticiper les tendances de production, tout en tenant compte de leurs contraintes et préférences pour optimiser la gestion agricole[97].

2.10.1 intégration des données sources

L'intégration des données provenant de sources multiples est essentielle pour la création d'un modèle prédictif robuste dans le domaine agricole. Les paramètres doivent être collectés à partir de différentes sources en utilisant une variété de moyens de collecte. Cela inclut des capteurs IoT pour recueillir des données en temps réel sur les conditions environnementales, les équipements agricoles et les récoltes, ainsi que des outils manuels pour les relevés effectués par les agriculteurs sur le terrain[136]. De plus, des technologies telles que les connexions Wi-Fi ou les réseaux sans fil peuvent être utilisées pour

transmettre les données collectées vers des systèmes centralisés où elles peuvent être analysées[91]. Ces données proviennent de divers objets connectés tels que des stations météorologiques, des capteurs de sol, des drones ou des véhicules agricoles. L'intégration efficace de ces données provenant de différentes sources permet une vision globale et précise des conditions de production, tout en tenant compte des spécificités locales et des paramètres contextuels de chaque exploitation[79].

2.10.2 Federated learning

Dans le domaine agricole, l'apprentissage fédéré (Federated Learning) constitue une approche prometteuse pour la gestion et l'analyse des données décentralisées tout en garantissant la confidentialité des informations des exploitations agricoles[100]. L'intégration de l'apprentissage fédéré au niveau stratégique nécessite la mise en place d'un modèle centralisé qui coordonne l'apprentissage de multiples modèles locaux, chacun étant formé sur les données spécifiques à une exploitation ou à un groupe d'exploitations[160]. Ce modèle pivot, ou modèle global, combine les résultats des modèles locaux sans nécessiter le transfert des données sensibles vers un serveur central. Cela permet aux agriculteurs de conserver leurs données locales tout en bénéficiant des avantages d'un apprentissage collaboratif. En pratique, les agriculteurs peuvent ainsi partager des informations sur les tendances de production, les conditions météorologiques ou les rendements agricoles tout en respectant la confidentialité de leurs pratiques agricoles et de leurs informations sensibles. L'apprentissage fédéré offre une solution scalable pour optimiser la gestion des ressources, prédire les rendements, et adapter les stratégies agricoles, tout en respectant les contraintes de confidentialité et de sécurité des données, essentielles dans le contexte agricole[89].

2.10.3 Optimisation multicritère

L'optimisation multicritère (MCDM) est une approche essentielle dans le domaine agricole pour résoudre des problèmes complexes de décision où plusieurs critères doivent être pris en compte simultanément[134]. Dans le contexte de la culture des oliviers, ces critères peuvent inclure le rendement des oliviers, les coûts de production, les exigences en eau, et l'impact environnemental des différentes pratiques agricoles. Par exemple, un exploitant oléicole pourrait être amené à choisir entre plusieurs types de traitements phytosanitaires ou différentes méthodes d'irrigation, chacun ayant des effets variés sur le rendement et l'environnement[8].

2.11 Positionnement de nos recherches

L'une des principales caractéristiques des modèles d'optimisation dans le domaine agricole est qu'ils évoluent en parallèle avec les avancées des technologies agricoles et des systèmes de gestion des données. Pour créer un modèle d'optimisation pour un système agricole, les chercheurs prennent en compte plusieurs aspects, tels que les types de cultures, les conditions climatiques, les systèmes de gestion des ressources (eau, fertilisation, etc.), les plateformes de gestion des exploitations agricoles, les dispositifs de collecte de données (capteurs IoT, images satellites, etc.), et bien sûr les objectifs spécifiques des agriculteurs (rendement, durabilité, rentabilité, etc.). En fait, d'après une revue des approches existantes,

les modèles d'optimisation actuels ne sont pas toujours génériques et ne prennent pas en compte toutes les dimensions de la gestion agricole.

Nous proposons que pour élaborer un modèle d'optimisation agricole, toutes ces dimensions doivent être intégrées, de manière à fournir une solution plus cohérente et globale. Nous avons décidé de proposer notre approche comme un référentiel de modèles d'optimisation afin d'aider les agriculteurs (producteurs) et les chercheurs à réutiliser et à développer leurs propres modèles d'optimisation adaptés aux réalités locales et aux objectifs spécifiques de chaque exploitation.

Le choix d'un modèle d'optimisation est une décision cruciale pour améliorer l'efficacité des systèmes agricoles. L'objectif de ce travail est de définir un modèle d'optimisation prenant en compte les différentes dimensions de la gestion agricole. Ce modèle sera la brique principale du développement d'algorithmes d'optimisation des processus agricoles, dont les composants pourront être adaptés en fonction des besoins des utilisateurs. Cela signifie que le processus d'optimisation pourra être configuré en fonction des spécificités de chaque exploitation et des priorités des agriculteurs (réduction des coûts, augmentation des rendements, réduction de l'impact environnemental, etc.).

Notre vision est de valoriser tous les efforts entrepris jusqu'à présent pour définir des modèles d'optimisation dans l'agriculture. Nous proposons, dans le chapitre suivant, un méta-modèle pour décrire les processus d'optimisation dans les systèmes agricoles, permettant de rassembler les approches existantes et de créer une solution plus intégrée.

2.11.1 Vers la réalisation de nos objectifs

Une fois que l'idée de créer le référentiel a commencé à mûrir progressivement, nous avons constaté l'absence d'un formalisme dédié à l'expression unifiée des modèles de prévision de la production agricole. Nous avons également identifié la nécessité d'un formalisme pour exprimer le contexte du système agricole sur lequel l'utilisation d'un modèle de prévision donné peut être précisée. Un autre point important sur lequel nous devons attirer l'attention est la manière d'exprimer les besoins des utilisateurs des modèles de prévision de la production agricole, qui chercheront une solution pour optimiser le système agricole.

Par conséquent, dans la suite, nous appellerons "Manifeste" les caractéristiques qui correspondent à une demande spécifique d'un praticien ou chercheur. La première étape vers la réalisation de ce référentiel est donc l'implémentation d'un formalisme des concepts des modèles de prévision de la production agricole, qui doit être suffisamment riche et évolutif pour être compatible avec les modèles existants, afin d'homogénéiser les modèles de prévision. Ce même formalisme servira de support pour homogénéiser les "manifestes".

Comme nous le verrons dans les chapitres suivants, notre approche repose principalement sur les approches dirigées par les données. Alors que cette dernière aide à atteindre la capacité persistante que nous visons, l'ingénierie dirigée par les données offre l'opportunité de raccourcir le cycle de vie du développement, en apportant des moyens d'exprimer des systèmes et de les analyser de manière plus structurée et efficace.

2.11.2 Aperçue de notre solution

Notre solution est dédiée à la définition d'une méthodologie de conception de modèles de calcul (MdC) orientée vers la réutilisation dans le domaine de l'agriculture. Celle-ci repose sur les techniques de ML/DL, car elles offrent différents services (par exemple, extraction de données, prétraitement, apprentissage et test de modèles). Le framework résultant permet la modélisation d'une fonction de prévision et leur manipulation. En effet, nous allons présenter une vue globale de notre framework AFFAC, et les chapitres suivants traiteront des détails de chaque contribution.

La ?? montre l'aperçu général de notre approche appelée AFFAC. Les piliers de cette approche sont :

- (i) **Les données sources** permettant de construire des dérivés des données sources possibles qui doivent être considérées comme des entrées dans la fonction de prévision des cultures agricoles.
- (ii) **La préparation des données** permettant de préparer les données suivant les contraintes et les exigences des modèles demandés en termes de type de données, format, etc.
- (iii) **Apprentissage et test de modèles** : ce service découvre des patterns et des modèles dans les données afin de trouver des modèles d'interprétation, ainsi que le test suivant un jeu de données pour mesurer leur efficacité.

Notre contribution est basée sur une approche dérivée de données et est destinée à apporter une réutilisation flexible dans le domaine de la modélisation des fonctions de production agricole. Notre cadre AFFAC repose sur les quatre dimensions AMTE. AMTE est utilisé comme ligne directrice pour un ensemble d'étapes afin de développer notre framework (AFFAC) : l'analyse de domaine (A), la modélisation de domaine (M), le test des modèles (T) et l'exploitation des modèles (E). La vue d'ensemble de la méthodologie AFFAC est présentée dans la ?? et repose sur les étapes suivantes :

1. **L'analyse de domaine** : Analyse du domaine par une étude ciblée et approfondie des concepts relatifs aux fonctions de production.
2. **La modélisation de domaine** : Explication des paramètres et des facteurs d'impact qui sont utilisés ou définis pour la clarification et la conceptualisation du domaine de la conception de modèles de prévision des cultures agricoles.
3. **La construction des modèles** : L'utilisation de l'apprentissage machine pour la modélisation du système ou pour affiner les modèles importés.
4. **L'exploitation** : Enfin, l'exploitation d'un modèle de prévision pour définir et manipuler les modèles.

La Figure 2.8 montre les scénarios d'utilisation du cadre AFFAC. Ce dernier aide les concepteurs à manipuler et exploiter les fonctions de production persistées en appliquant différentes opérations.

2.12 Conclusion

La prévision de la production agricole constitue un enjeu central pour la sécurité alimentaire, la gestion durable des ressources naturelles et l'adaptation aux changements climatiques. À travers ce chapitre, nous avons exploré les différentes approches récentes de modélisation utilisées pour prédire les rendements agricoles, en mettant en lumière leurs caractéristiques, avantages et limites. L'intégration

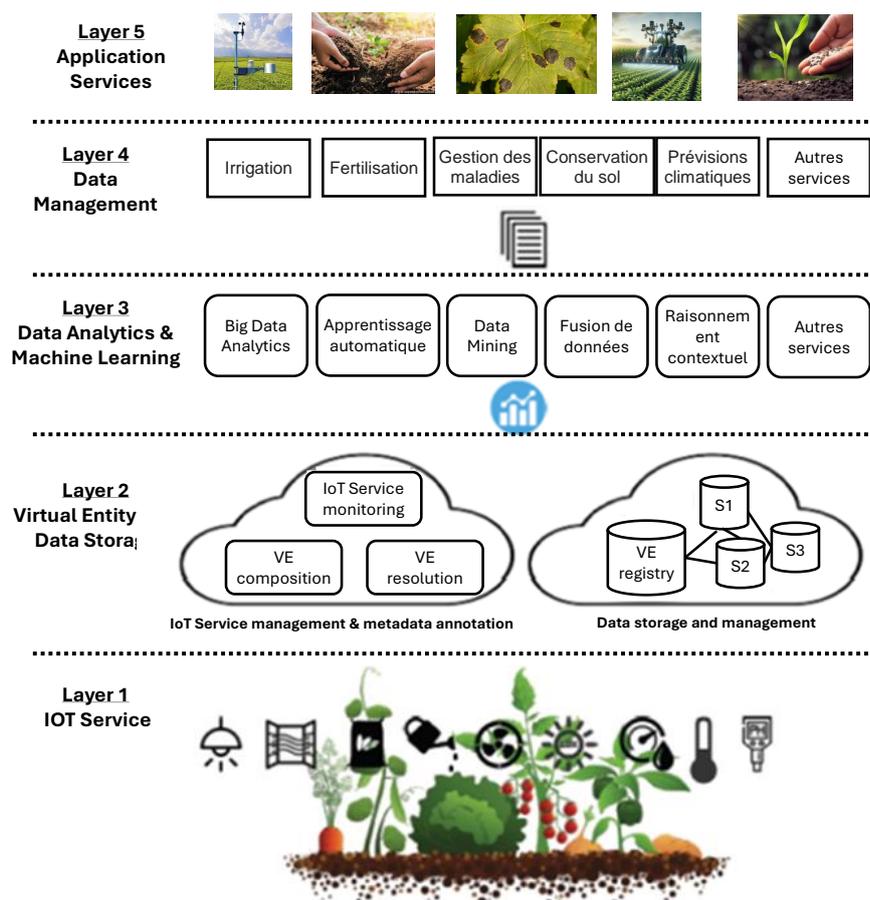


FIGURE 2.8 – Aperçu de notre solution data-driven.

de méthodologies analytiques et empiriques, ainsi que l'essor des modèles basés sur le machine learning, a permis des avancées significatives, offrant des outils plus précis et plus adaptés pour comprendre et anticiper les variations des rendements agricoles sous diverses conditions.

Les modèles prédictifs, en particulier ceux utilisant l'apprentissage automatique, ont montré un potentiel considérable pour améliorer la précision des prévisions, notamment en prenant en compte des facteurs complexes tels que la variabilité climatique. Cependant, la comparaison de ces modèles avec des approches plus traditionnelles souligne la nécessité d'une évaluation rigoureuse pour mesurer leur efficacité. L'utilisation de métriques standards, telles que la RMSE, la MAE et le coefficient de détermination (R^2), reste essentielle pour garantir la robustesse des modèles et leur capacité à s'adapter à des données nouvelles et diverses.

Cette synthèse des modèles de prévision agricole a permis d'offrir une vue d'ensemble des solutions disponibles tout en soulignant les défis actuels. En conclusion, les modèles modernes, notamment ceux basés sur l'apprentissage automatique, représentent une voie prometteuse pour améliorer la gestion des rendements agricoles. Ils offrent ainsi des perspectives intéressantes pour répondre aux défis contemporains de l'agriculture, en particulier face aux risques climatiques, et pour aider les acteurs du secteur à prendre des décisions éclairées et adaptées aux conditions changeantes.

Deuxième partie

Contribution

Cadre fondé sur les modèles pour la modélisation et les tests de prévision dans les systèmes de production agricole



« Chaque individu apporte au monde sa contribution unique. »

— Jack Kornfield

Sommaire

3.1	Introduction	80
3.2	AFFAC : Notre cadre de travail	80
3.2.1	Facteurs influençant les coûts de production agricole	81
3.3	Formalisation du problème	82
3.4	Aperçu du cadre AFFAC	83
3.4.1	Collecte des Données	84
3.4.2	Définition de l'Espace des Paramètres pour la Fonction de Coût de Production	86
3.4.3	Définition de <i>Prod</i> en utilisant des techniques ML/DL	86
3.4.4	Le clustering K-means	86
3.4.5	Modèle ARIMA	87
3.4.6	Réseaux Neuronaux Récurrents (RNNs)	88
3.4.7	Réseaux à Mémoire à Long Terme (LSTM)	88
3.4.8	Unités Récurrentes à Porte (GRU)	89
3.4.9	Transformeurs	89
3.4.10	Evaluation Metrics	93
3.5	Les utilisateurs de notre cadre AFFAC	93
3.5.1	Décideurs	94
3.5.2	Gestionnaires	94
3.5.3	Agriculteurs	94
3.5.4	Scientifiques des données	94
3.5.5	Cas d'optimisation agricole à un niveau managérial	95
3.6	Conclusion	98

3.1 Introduction

Dans ce chapitre, nous introduisons un cadre de modélisation basé sur les données appelé AFFAC (Advanced Forecasting Framework for Agricultural Crops) pour améliorer la précision des prévisions de la production agricole. Ce cadre se concentre particulièrement sur l'amélioration des prévisions pour la culture de l'olivier, un élément clé de l'agriculture méditerranéenne. En intégrant des paramètres et des conditions spécifiques à la culture de l'olivier, AFFAC vise à fournir des prévisions plus précises et à mieux soutenir la prise de décision dans la production et la gestion des oliveraies. Pour ce faire, le cadre établit d'abord un processus systématique pour collecter et gérer les données relatives à la production agricole, en tirant parti de diverses sources de données et processus. Il intègre ensuite les connaissances d'experts pour identifier et sélectionner les variables d'entrée qui influencent les facteurs d'impact les plus sensibles de la production culturelle. De plus, le cadre offre des options de personnalisation, permettant aux utilisateurs de sélectionner et d'appliquer facilement différents modèles de prévision en associant les paramètres d'entrée pertinents aux spécifications du modèle.

3.2 AFFAC : Notre cadre de travail

Dans cette section, nous présentons notre cadre de travail, AFFAC (voir Figure Figure 3.1). Ce cadre a été conçu pour relever le défi de modélisation du coût de production *Prod*, que nous définissons comme un problème d'apprentissage. Nous décrivons ensuite notre approche pour construire une fonction de coût de production, en exploitant des modèles avancés d'apprentissage automatique (ML) et d'apprentissage profond (DL) afin d'optimiser la précision et l'adaptabilité dans la prévision des coûts de production agricole.

Le cadre AFFAC intègre des données relatives aux observations climatiques et météorologiques, en commençant par les **Sources de Données** telles que les paramètres du sol, les données SIG, les champs d'oliviers, l'irrigation et la météo. La collecte de ces données s'effectue via des méthodes d'**Acquisition de Données**, comme les API, les saisies manuelles et des options de connectivité (Wi-Fi, Ethernet, Bluetooth). La phase de **Prétraitement des Données** inclut le stockage, l'analyse statistique et la transformation des données. Enfin, les **Services d'IA** exploitent l'intelligence artificielle et le data mining pour fournir des analyses, en tirant parti de diverses sources de données et technologies dédiées à l'analyse agricole et environnementale.

Notre objectif est de proposer un cadre d'assistance pour la conception et l'évaluation d'un modèle de prévision axé sur la production agricole. Un aspect essentiel pour y parvenir est la modélisation des processus à travers des couches d'abstraction, ce qui permet de structurer efficacement le modèle. En intégrant plusieurs couches d'abstraction, nous améliorons la complexité et la précision du modèle. La Figure 3.1 illustre la conception structurée de notre cadre.

Dans cette architecture multicouche illustrée dans la Figure 3.1, nous commençons par le niveau manifeste, qui fournit les sources de données utilisées comme entrées pour la fonction mathématique de coût. Cette approche par couches nécessite de définir les relations ou correspondances entre les différentes couches. En utilisant des relations d'instanciation intra-couche, nous montrons comment la formule mathématique globale est dérivée de divers facteurs influents.

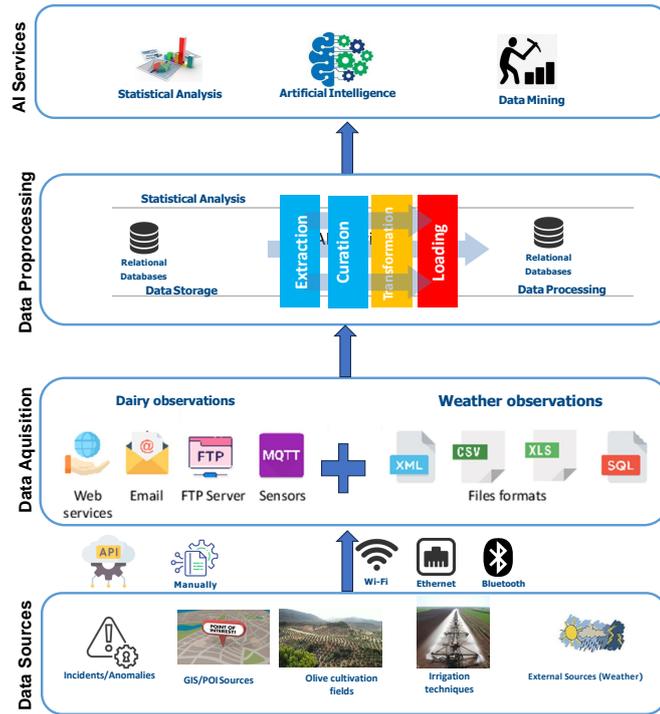


FIGURE 3.1 – Aperçu du cadre AFFAC.

3.2.1 Facteurs influençant les coûts de production agricole

Dans cette étude, un modèle de prévision de la production agricole, noté $\mathcal{P}rod$, est une fonction mathématique essentielle pour estimer et simuler la production agricole. Le modèle $\mathcal{P}rod$ correspondant à une métrique spécifique d'une culture élémentaire est défini par la fonction suivante :

$$\mathcal{P}rod_{m_i}^{Culture} : \mathbb{P}^n \longrightarrow Value \in \mathbb{R} \quad (3.1)$$

où :

- m_i : est un facteur mesurable utilisé pour analyser la performance et l'efficacité des opérations agricoles, tels que le rendement des cultures, la productivité, l'indice de récolte, le coût de production, etc.
- *Culture* : fait référence à la culture de plantes destinées à l'alimentation, à l'industrie, aux fibres ou à d'autres produits (par exemple, *olives*, *céréales*, *dattes*).
- \mathbb{P} : représente le domaine des paramètres couvrant l'ensemble de l'environnement de la culture en termes de climat, sol et topographie (c.-à-d. *température*, *humidité*, *type de sol*, *élévation* et *pente*) [11, 29].
- n : est la cardinalité des paramètres d'entrée pris en compte par la fonction mathématique de production $\mathcal{P}rod_{m_i}^{Culture}$.
- *Value* : représente le résultat de sortie. Ce résultat est un nombre réel qui est typé en fonction de l'attribut de qualité de service (QoS) ciblé (c.-à-d. rendement des cultures, productivité, indice de récolte, coût de production).

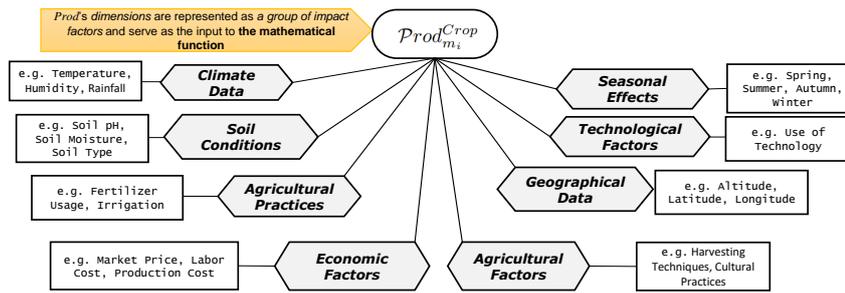


FIGURE 3.2 – Vue d'ensemble des principaux facteurs influençant les coûts de production culturelle.

La fonction mathématique de production ci-dessus dépend de plusieurs paramètres contextuels agricoles [21], tels que les *facteurs environnementaux, biologiques, agronomiques, technologiques, économiques, et socio-économiques*. En comprenant et en optimisant ces paramètres, il est possible d'améliorer la production des cultures, augmentant ainsi l'efficacité et la durabilité des pratiques agricoles, comme illustré dans la Figure 3.2. Par exemple, les fonctions de coût de production peuvent être dérivées des **données agronomiques**, telles que la *fertilisation du sol* et la *densité des cultures*, ainsi que du **contrôle des ravageurs**, comme l'*efficacité de l'irrigation* et la *consommation énergétique des machines*. En analysant ces facteurs, les agriculteurs et les planificateurs agricoles peuvent optimiser l'utilisation des ressources, réduire les coûts et améliorer le rendement et la qualité des cultures.

Exemple 1. Expression des besoins *Considérons les exigences des concepteurs (par exemple, économistes agricoles, développeurs de logiciels) et les contraintes liées à la culture des olives à l'aide des directives structurées suivantes :*

- ① *Type de culture : Production d'olives, avec des contraintes telles que la disponibilité en eau et les coûts d'irrigation.*
- ② *Paramètres du système :*
 - *Facteurs agricoles : Fertilité des sols, stockage orienté par rangées.*
 - *Conditions environnementales : Type de sol, Température, humidité, disponibilité des données (non limitées par la compression).*
 - *Métriques technologiques : Efficacité de l'irrigation, consommation énergétique des machines.*
- ③ *Coûts de production : Prévion des coûts associés à la production de cultures d'olives.*

Le Tableau 4.8 fournit une description du contexte pour le modèle de prévision à développer pour la production d'olives. La conception d'une fonction de coût pour ce contexte nécessite une compréhension approfondie des paramètres et la création d'un modèle capturant leurs interactions afin de dériver les métriques souhaitées.

3.3 Formalisation du problème

Le principal défi de notre approche a été de définir la prévision des performances à l'aide d'une fonction de prédiction des coûts, appelée $Prod_{m_i}^{Culture}(P_i)$, au sein du cadre AFFAC. Nous proposons que l'intégration de plusieurs modèles pour générer cette fonction de coût offre des avantages significatifs

Tableau 3.1 – Aperçu contextuel pour le développement de la fonction de production d’olives

Problème	Type de culture	Variétés d’olives
	Attributs QoS	Densité des cultures
	Contraintes	Disponibilité en eau, coûts d’irrigation
Paramètres du système	Facteurs agricoles	Fertilité des sols Densité des cultures Techniques d’égagage
	Conditions environnementales	Température Humidité Type de sol
	Métriques technologiques	Efficacité de l’irrigation Consommation énergétique des machines
	<i>Prod</i>	Prévision des coûts de production des olives

pour les agroécologistes et les économistes agricoles. Pour ce faire, nous introduisons deux approches distinctes : (i) les modèles d’apprentissage automatique (MLM) et (ii) les modèles d’apprentissage profond (DLM). Ces modèles offrent flexibilité et adaptabilité dans la construction de prévisions de coûts précises.

La métrique m_i , ou paramètre cible, peut être calculée en fonction de différentes combinaisons de paramètres (\oplus). Pour un ensemble de paramètres d’entrée $P_i = p_{i_1}, \dots, p_{i_k}$, où $P_i \in cxt$ (c’est-à-dire un espace de paramètres), les valeurs de la métrique peuvent être dérivées comme suit :

$$Prod_{m_i}^{Culture}(P_i) = p_{i_1} \oplus p_{i_2} \oplus \dots \oplus p_{i_k} \quad (3.2)$$

Enfin, la prévision de la fonction de coût de production $Prod_{m_i}^{Culture}$, pour les productions agricoles telles que les olives, peut être modélisée comme un processus d’entrée-sortie comme suit :

Prévision des coûts de production

Entrée (cxt_i) :

- Facteurs agricoles (par exemple, fertilité du sol, densité des cultures).
- Conditions environnementales (par exemple, température, humidité, type de sol).
- Indicateurs technologiques (par exemple, efficacité de l’irrigation, consommation énergétique des machines).
- La région d’incertitude (par exemple, conditions climatiques futures, demande du marché).

Sortie :

$Prod_{m_i}^{Culture}$: la fonction de coût de production prédite pour les paramètres agricoles donnés.

3.4 Aperçu du cadre AFFAC

Le cadre AFFAC, illustré dans la Figure 3.1, utilise des approches basées sur les données pour développer des fonctions de coût de production des cultures. L’approche centrale de AFFAC consiste à découvrir les facteurs d’impact cachés dans la fonction de coût de production à travers des modèles d’apprentissage

automatique (*Machine Learning*, ML) et d'apprentissage profond (*Deep Learning*, DL). Le cadre AFFAC, présenté dans la Figure 3.1, suit quatre étapes principales : (i) Exécution des benchmarks et collecte des données, (ii) Génération et échantillonnage de l'espace des paramètres, (iii) Construction d'une fonction de coût de production, et (iv) Évaluation de la fonction de coût de production à l'aide de diverses métriques pertinentes.

AFFAC est défini par cinq composants clés : *Dataset* (jeu de données), *Parameters* (paramètres), *Cost Function* (fonction de coût), *Sensitivity Analysis Methods* (méthodes d'analyse de sensibilité), et *Feature Selection* (sélection des caractéristiques).

$$\text{AFFAC} = \langle (\mathcal{D}, \text{cxt}, \text{Prod}_{m_i}^{\text{Culture}}, \lambda, \mathbf{F}(p), \text{Clust} \rangle$$

Où :

- **D** : Un jeu de données, organisé sous la forme $(X_{ij}; Y_i)$, avec $i \in [1; n]$ et $j \in [1; J]$. \mathcal{D} est divisé en les instances d'entraînement $\mathcal{D}_{\text{train}}$ et les instances de test $\mathcal{D}_{\text{test}}$; chaque instance est un vecteur de caractéristiques. L'entrée P doit garantir une couverture plus large de \mathcal{D} .
- **cxt** : Espace des paramètres P_1, \dots, P_n couvrant l'ensemble de l'environnement de la production agricole dérivé d'un ensemble de dimensions telles que les conditions environnementales (par exemple, *température, humidité, type de sol*).
- **Prod_{m_i}^{culture}** : La fonction de coût correspondant à un métrique donné m_i produite par des modèles d'apprentissage automatique (ML) ou d'apprentissage profond (DL) ; par exemple, *modèle de régression, modèle DNN*.
- λ : Méthodes d'analyse de sensibilité pour analyser les paramètres de sensibilité ; par exemple, les distributions statistiques des données et l'importance cumulative des caractéristiques.
- **F** : Techniques de sélection de caractéristiques, telles que *Analyse en Composantes Principales* (ACP) et *Analyse de Corrélation* (CA), couramment utilisées.
- **Clust** : Algorithme de clustering appliqué à chaque jeu de données $\mathcal{D}_{\text{train}}$.

3.4.1 Collecte des Données

Dans les *approches d'apprentissage basées sur les données*, un ensemble de données de haute qualité \mathcal{D} est essentiel pour un entraînement réussi du réseau. Notre approche se concentre sur l'automatisation des tâches et la collecte des données pertinentes pour prédire la fonction de coût de production agricole, en garantissant précision et efficacité. Le processus d'automatisation de ces tâches et de collecte des données est détaillé dans la Figure 3.3.

1. **Collecte des données** : Collecter des données sur des paramètres tels que les conditions environnementales et les pratiques agricoles, qui seront utilisées à la fois pour l'entraînement et l'évaluation du modèle. Cette étape garantit que toutes les données pertinentes sont recueillies à partir de diverses sources pour soutenir une analyse complète.
2. **Extraction des données** : Extraire les données nécessaires des ensembles de données brutes collectées. Cela implique d'identifier et de récupérer des informations spécifiques qui seront cruciales pour l'analyse ultérieure et le développement du modèle.
3. **Transformation des données** : Transformer les données extraites dans un format approprié pour l'analyse. Cette étape comprend le nettoyage, la normalisation et la conversion des données en une structure cohérente qui peut être facilement intégrée dans le système d'analyse agricole.

4. **Fusion des données** : Fusionner les données traitées dans un ensemble de données unifié. L'objectif de la fusion des journaux est de créer un journal fusionné dans un format unique en combinant un journal de données avec un journal des ressources informatiques et en incluant les attributs pertinents (c'est-à-dire les caractéristiques). Cela consiste à collecter des journaux de données provenant de diverses sources et à les analyser dans un ensemble de données structuré et unifié, comme illustré dans la Figure 3.4.
5. **Chargement des données** : Charger les données générées à partir de diverses collections de données dans le système d'analyse agricole. Cette dernière étape garantit que les données sont prêtes pour l'analyse et l'entraînement du modèle, permettant ainsi une exploration et une prise de décision ultérieures.

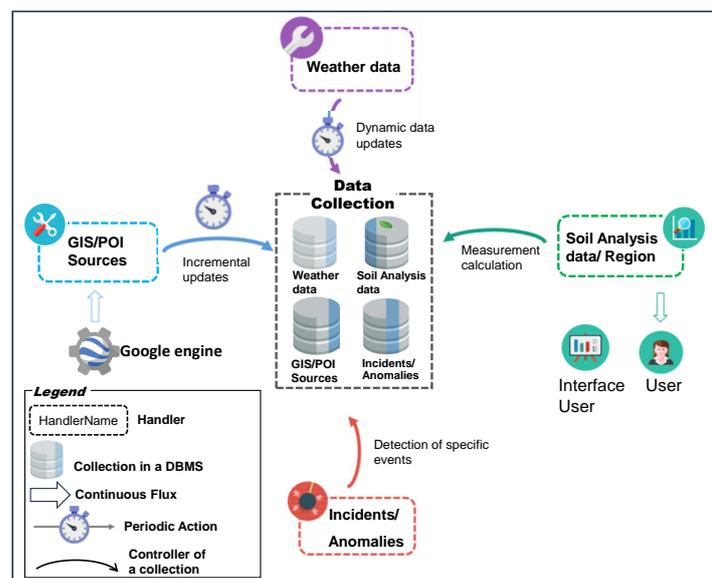


FIGURE 3.3 – Processus de collecte de données pour la production agricole.

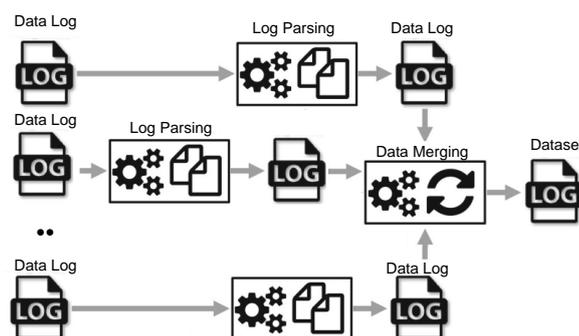


FIGURE 3.4 – Flux de travail conceptuel pour la fusion des données.

La Figure 3.3 présente une approche structurée de la collecte et de la gestion des données liées à la production agricole en agriculture, intégrant plusieurs sources de données et processus. Elle offre une vue d'ensemble détaillée du flux de travail de collecte et de gestion des données, mettant en évidence l'intégration de sources diverses telles que les données météorologiques, les systèmes GIS/POI, Google

Engine, l'analyse du sol et les rapports d'incidents. Ces ensembles de données sont analysés pour fournir des informations exploitables, avec un accent particulier sur le traitement des données en temps réel, les mises à jour continues et l'interaction avec les utilisateurs. Le diagramme souligne le rôle crucial de la technologie dans l'optimisation des pratiques agricoles, en mettant en avant les interfaces utilisateur et les points d'interaction qui facilitent la prise de décisions efficaces.

3.4.2 Définition de l'Espace des Paramètres pour la Fonction de Coût de Production

L'ensemble de données $\mathcal{D} = (P_i, Value_i)_{i=1}^N$ représente une collection de taille N , où $P_i = (P_{i_1}, \dots, P_{i_n})$ désigne l'espace des paramètres avec n paramètres, et $Value_i$ indique le paramètre cible.

L'ensemble P_i comprend des paramètres liés à la fonction de coût de production, tels que les *paramètres environnementaux*, les *caractéristiques du sol*, les *pratiques de gestion des cultures* et les *facteurs technologiques*. Le $Value_i$ reflète le métrique lié à la production m_i (par exemple, le *rendement*, le *coût de culture*) pour une instance de culture donnée.

L'objectif est de prédire cette valeur de métrique à l'aide de la fonction de coût de production, en garantissant une généralisation efficace des données d'entraînement vers de nouveaux scénarios agricoles.

Voici la description du processus de génération de la fonction de prédiction du coût en utilisant ces différents modèles.

3.4.3 Définition de \mathcal{P}_{prod} en utilisant des techniques ML/DL

Dans cette section, nous présentons la fonction de coût de production en utilisant des modèles ML et DL pour illustrer son application à travers diverses combinaisons de paramètres (\oplus). Initialement, notre cadre utilise des techniques de clustering, telles que K-means. Après la phase de clustering, nous modélisons la fonction de coût en utilisant un ensemble d'algorithmes issus de AFFAC.

Le clustering exploite la diversité et similarité des paramètres climatiques des régions agricoles pour adapter chaque modèle de prévision à un cluster spécifique.

3.4.4 Le clustering K-means

L'analyse de *clustering K-means* est un processus d'apprentissage non supervisé qui identifie des modèles potentiellement similaires dans un jeu de données sans information préalable [81]. L'algorithme vise à diviser un jeu de données en k clusters, où chaque cluster est défini par son centroïde. L'objectif est de minimiser la somme des distances au carré (erreurs) entre les points de données et leurs centroïdes respectifs [108]. Cela est représenté mathématiquement par :

$$SSE(C) = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (3.3)$$

où $C = \{C_1, C_2, \dots, C_k\}$ est l'ensemble des clusters, x_j sont les points de données, et μ_i sont les

centroïdes des clusters. L'objectif est de trouver la configuration des clusters C^* qui minimise la SSE (somme des erreurs au carré) :

$$C^* = \arg \min_C SSE(C) \quad (3.4)$$

L'algorithme K-means suit un processus itératif composé de deux étapes principales :

1. **Étape d'assignation** : Chaque point de donnée est assigné au centre de cluster le plus proche en fonction de la distance euclidienne.
2. **Étape de mise à jour** : Les centroïdes des clusters sont recalculés comme la moyenne de tous les points de données qui leur sont assignés.

Ces étapes sont répétées jusqu'à convergence, qui est généralement atteinte lorsque les centroïdes ne changent plus de manière significative entre les itérations. L'algorithme peut converger vers un minimum local ; il est donc souvent exécuté plusieurs fois avec des initialisations différentes afin de trouver le meilleur regroupement [64].

K-means est efficace sur le plan computationnel et convient aux grands ensembles de données, bien que la qualité du regroupement dépende de l'initialisation des centroïdes. La complexité temporelle de l'algorithme est $O(nkd)$, où n représente le nombre de points de données, k le nombre de clusters, d le nombre de dimensions, et t le nombre d'itérations []

3.4.5 Modèle ARIMA

Le coût peut être modélisé à l'aide d'un modèle ARIMA (AutoRégressif Intégré à Moyenne Mobile) :

$$Prod_{m_i,t}^{Culture} = \sum_{i=1}^p \phi_i Prod_{m_i,t-i}^{Culture} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t \quad (3.5)$$

où ϕ_i sont les coefficients autorégressifs, θ_j sont les coefficients de moyenne mobile, et ϵ_t représente le terme d'erreur [23]. L'ordre du modèle ARIMA est (p, d, q) .

Le modèle ARIMA est utile pour les séries temporelles où le coût dépend des valeurs passées et de facteurs stochastiques. Les paramètres ϕ_i et θ_j sont estimés et ajustés pour garantir la précision. Les performances du modèle sont évaluées à l'aide du critère d'information d'Akaike (AIC) :

$$AIC = 2k - 2 \ln(L) \quad (3.6)$$

où k est le nombre de paramètres et L est la vraisemblance du modèle.

3.4.6 Réseaux Neuronaux Récurrents (RNNs)

Les RNNs peuvent être utilisés pour modéliser la fonction de coût en capturant les dépendances entre les étapes temporelles comme suit [60] :

$$h_t = \sigma(W_h \cdot x_t + U_h \cdot h_{t-1} + b_h) \quad (3.7)$$

$$Prod_{m_i,t}^{Culture} = W_o \cdot h_t + b_o \quad (3.8)$$

où h_t est l'état caché au temps t , x_t est l'entrée au temps t , W_h et U_h sont les matrices de poids pour l'entrée et l'état caché respectivement, et b_h est le terme de biais. La fonction σ est une fonction d'activation telle que tanh ou ReLU. La couche de sortie est définie par les poids W_o et le biais b_o , produisant le coût prédit $Prod_{m_i,t}^{Culture}$.

Les RNNs sont particulièrement efficaces pour les données séquentielles, car ils peuvent apprendre des motifs temporels en utilisant l'état caché h_t pour transmettre des informations à travers les étapes temporelles. Le modèle de coût est entraîné en minimisant la fonction de perte, telle que l'erreur quadratique moyenne (MSE), calculée comme suit :

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (3.9)$$

où Y_i est le coût réel et \hat{Y}_i est le coût prédit au temps i .

3.4.7 Réseaux à Mémoire à Long Terme (LSTM)

Les réseaux LSTM, un type de RNN, modélisent les données séquentielles et capturent les dépendances à long terme [68]. Les équations du LSTM sont :

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3.10)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3.11)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3.12)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (3.13)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (3.14)$$

$$h_t = o_t \odot \tanh(c_t) \quad (3.15)$$

La fonction de production est :

$$Prod_{m_i,t}^{Culture} = W_y h_t + b_y \quad (3.16)$$

où W_y est la matrice de poids de sortie, h_t l'état caché à l'instant t , et b_y le biais associé.

Par ailleurs, f_t , i_t , et o_t désignent respectivement les portes d'oubli, d'entrée et de sortie, \tilde{c}_t l'état candidat, c_t l'état de la cellule, et h_t l'état caché. Les fonctions σ et \tanh représentent la sigmoïde et la tangente hyperbolique, tandis que \odot indique la multiplication élément par élément.

3.4.8 Unités Récurrentes à Porte (GRU)

Les GRU (Unités Récurrentes à Porte) sont des RNN (Réseaux de Neurones Récurrents) qui modélisent efficacement les données séquentielles en utilisant des mécanismes de porte [36]. Les équations des GRU sont :

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (3.17)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (3.18)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \quad (3.19)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (3.20)$$

La fonction de production est :

$$\text{Prod}_{m_i,t}^{\text{Culture}} = W_y h_t + b_y \quad (3.21)$$

où z_t est la porte de mise à jour, r_t la porte de réinitialisation, \tilde{h}_t l'état candidat, et h_t le nouvel état caché. Les fonctions σ et \tanh sont respectivement la fonction sigmoïde et la tangente hyperbolique, avec \odot représentant la multiplication élément par élément.

Les GRU sont efficaces pour les tâches nécessitant des temps d'entraînement plus courts et des modèles plus simples. La fonction de coût est optimisée en utilisant l'erreur quadratique moyenne (MSE) :

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (3.22)$$

où Y_i est le coût réel et \hat{Y}_i est le coût prédit au temps i .

3.4.9 Transformeurs

Dans cette section, nous présentons deux types différents de modèles Transformer qui exploitent le mécanisme de self-attention pour traiter efficacement les séquences de données.

Transformeur Vanilla

Les Transformeurs classiques capturent les dépendances de séquences avec le mécanisme de self-attention [151] :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.23)$$

où Q , K , et V sont les matrices de requêtes, de clés et de valeurs. L'attention multi-tête améliore cela :

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (3.24)$$

with

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3.25)$$

Les couches feed-forward ajoutent de la complexité :

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (3.26)$$

Les connexions résiduelles stabilisent l'entraînement :

$$\text{LayerNorm}(x + \text{Sublayer}(x)) \quad (3.27)$$

Représentation globale du bloc :

$$H_{\text{attn}} = \text{LayerNorm}(x + \text{MultiHead}(x, x, x)) \quad (3.28)$$

$$H_{\text{ffn}} = \text{LayerNorm}(H_{\text{attn}} + \text{FFN}(H_{\text{attn}})) \quad (3.29)$$

Incorporation du contexte agricole :

$$\mathcal{P}_{m_i, t}^{\text{Culture}} = \text{Transformer}(x_t) \quad (3.30)$$

Le modèle est optimisé en utilisant l'erreur quadratique moyenne (MSE) :

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (3.31)$$

Encodage d'entrée et encodage positionnel

Dans les transformeurs, les encodages d'entrée et positionnels représentent les données et l'ordre des séquences. La séquence d'entrée $X = (x_1, \dots, x_n)$ est mappée vers des vecteurs :

$$E = \text{Embed}(X) \quad (3.32)$$

où $E \in \mathbb{R}^{n \times d_{\text{model}}}$. Pour intégrer le contexte culturel :

$$E^{\text{Culture}} = \mathcal{P}\text{rod}_{m_i}^{\text{Culture}} \cdot E \quad (3.33)$$

où $\mathcal{P}\text{rod}_{m_i}^{\text{Culture}}$ ajuste les embeddings en fonction de la culture Culture_i .

L'encodage positionnel est ajouté pour capturer les positions des tokens :

$$PE_{(pos, 2i)} = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right) \quad (3.34)$$

$$PE_{(pos, 2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right) \quad (3.35)$$

où pos est l'indice de position et i est l'indice de dimension. L'entrée finale est :

$$Z = E^{\text{Culture}} + PE \quad (3.36)$$

Cela combine les informations sémantiques, positionnelles et culturelles, améliorant ainsi la performance du modèle.

Attention multi-tête

L'attention multi-tête permet aux transformeurs de se concentrer simultanément sur différentes parties de l'entrée. Elle intègre le contexte culturel avec $\mathcal{P}\text{rod}_{m_i}^{\text{Culture}}$ pour une spécificité améliorée. Définie comme :

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (3.37)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3.38)$$

où W^O , W_i^Q , W_i^K , et W_i^V sont des matrices apprises, et h est le nombre de têtes. Le mécanisme d'attention est :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.39)$$

Pour le contexte culturel :

$$\text{head}_i^{\text{Culture}} = \mathcal{P}\text{rod}_{m_i}^{\text{Culture}} \cdot \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3.40)$$

La sortie finale est :

$$\text{MultiHead}^{\text{Culture}}(Q, K, V) = \text{Concat}(\text{head}_1^{\text{Culture}}, \dots, \text{head}_h^{\text{Culture}})W^O \quad (3.41)$$

Cette méthode améliore la capacité du modèle à capturer les dépendances spécifiques à l'agriculture

Innovations récentes dans les modèles Transformer

Les avancées récentes dans les transformeurs, en particulier avec la notation $\mathcal{P}\text{rod}_{m_i}^{\text{Culture}}$, améliorent la performance du modèle en intégrant les nuances agricoles. Définie comme :

$$\mathcal{P}\text{rod}_{m_i}^{\text{Culture}} = \prod_{j=1}^n f(x_j, \text{Culture}_i) \quad (3.42)$$

où $f(x_j, \text{Culture}_i)$ adapte les caractéristiques d'entrée x_j au contexte agricole Culture_i . Implémenté par :

$$E^{\text{Culture}} = \text{Embed}(X) + \text{CultureEmbed}(\text{Culture}_i) \quad (3.43)$$

Les embeddings agricoles améliorent le traitement du langage dans des tâches telles que la traduction et l'analyse des sentiments, rendant les modèles plus efficaces dans des contextes multilingues diversifiés.

Réseau feed-forward et réseau résiduel

Le réseau feed-forward dans les transformeurs applique une transformation d'activation ReLU sur deux couches, ce qui permet de modéliser des relations non linéaires entre les caractéristiques d'entrée et d'améliorer la capacité d'apprentissage du modèle.

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (3.44)$$

Pour intégrer le contexte culturel :

$$\text{FFN}^{\text{Culture}}(x) = \mathcal{P}\text{rod}_{m_i}^{\text{Culture}} \cdot \text{FFN}(x) \quad (3.45)$$

où $\mathcal{P}\text{rod}_{m_i}^{\text{Culture}}$ ajuste la sortie en fonction des caractéristiques culturelles, permettant ainsi d'adapter la prédiction aux spécificités de chaque culture.

Les connexions résiduelles et la normalisation des couches stabilisent l'entraînement :

$$\text{LayerNorm}(x + \text{Sublayer}(x)) \quad (3.46)$$

Pour les modèles conscients du contexte culturel :

$$\text{LayerNorm}(x + \text{FFN}^{\text{Culture}}(x)) \quad (3.47)$$

3.4.10 Evaluation Metrics

Nous évaluons la qualité des productions agricoles $\mathcal{P}_{\text{Prod}}$ en utilisant l'erreur quadratique moyenne (RMSE) entre la valeur estimée du coût et la valeur réelle du coût :

$$\text{RMSE}(Y, \hat{Y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2} \quad (3.48)$$

3.5 Les utilisateurs de notre cadre AFFAC

Notre cadre AFFAC de modélisation et d'analyse de la production agricole peut être exploré et exploité par plusieurs acteurs suivant leurs couches, qui commencent par la couche d'extraction des données sources, puis passent à la couche des services qui offre des services fonctionnels en fonction des types d'utilisateurs.

Considérons le cas d'un comité de pilotage dans le domaine agricole, où divers acteurs cherchent à évaluer et améliorer la production agricole. La difficulté réside dans la manière dont la production agricole est analysée sous différentes perspectives, chacune nécessitant des niveaux d'abstraction, des granularités de données et des objectifs distincts. La Figure 3.5 illustre les utilisateurs de notre cadre AFFAC. Les informations sont collectées via questionnaires, capteurs et géolocalisation, puis stockées dans une base de données AFFAC. Un scientifique utilise des techniques d'apprentissage automatique de l'intelligence artificielle pour détecter les patterns et prédire des résultats à partir de données.

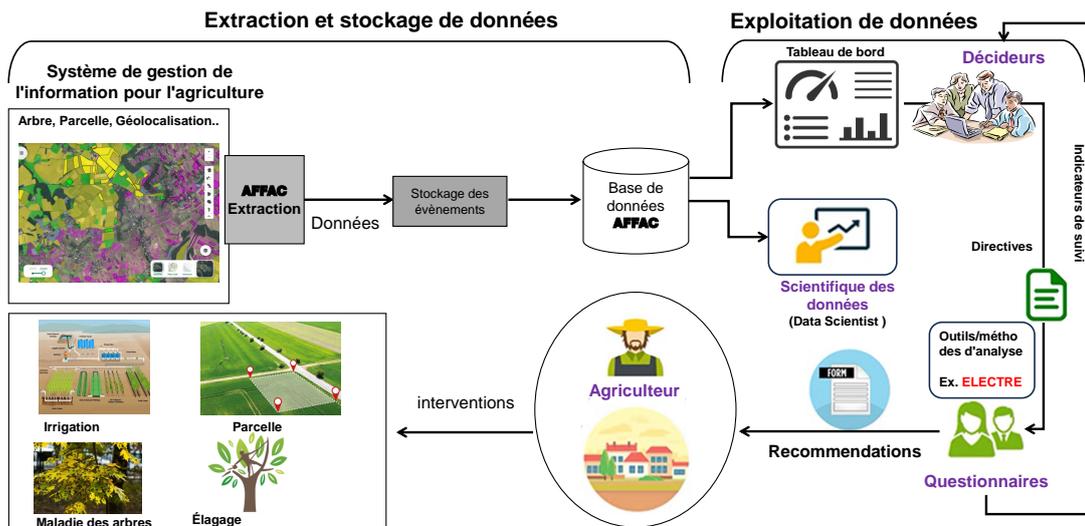


FIGURE 3.5 – Les utilisateurs de notre cadre AFFAC.

Les gestionnaires appliquent des méthodes comme l'optimisation multi-critères, notamment la

méthode ELECTRE III (La méthode d'analyse multicritère) [112], pour analyser les tendances et générer des recommandations. Cette approche permet de prendre en compte plusieurs critères dans les décisions agricoles, optimisant ainsi les interventions. Les agriculteurs reçoivent des directives pour améliorer l'efficacité de leurs actions. Les décideurs utilisent des tableaux de bord contenant des indicateurs clés pour la gestion stratégique. Ce système vise à améliorer la prise de décision en agriculture grâce à l'analyse et à l'exploitation des données collectées, tout en intégrant des méthodes d'optimisation avancées comme ELECTRE III pour garantir des choix éclairés et adaptés aux objectifs multiples. Ce cadre offre un espace pour les travaux futurs pour les étudiants en Master et PhD, qui pourront intégrer des méthodes et des services dans l'architecture de notre cadre proposé.

3.5.1 Décideurs

De manière générale, les décideurs (responsables agricoles, économistes) se concentrent sur les objectifs stratégiques globaux en exploitant des indicateurs de haut niveau (ex. : tendances de production, rendement moyen par hectare, impact des politiques agricoles). Ces indicateurs permettent d'identifier les profils de production, de prendre des décisions éclairées sur les politiques agricoles et d'orienter les stratégies d'investissement et d'affectation des ressources.

3.5.2 Gestionnaires

Au niveau managérial, les gestionnaires jouent un rôle clé dans la planification de la production agricole en utilisant des méthodes dédiées, telles que l'optimisation des ressources, la gestion des stocks et la prévision des rendements. Ils traduisent les stratégies définies par les décideurs en actions concrètes et assurent la coordination des moyens humains, techniques et financiers pour maximiser l'efficacité et la durabilité de la production.

3.5.3 Agriculteurs

D'un autre côté, les agriculteurs se focalisent sur les indicateurs et mesures opérationnels, en utilisant des métriques dérivées (ex. : efficacité des intrants, fréquence d'irrigation, rendement variétal) pour optimiser leurs pratiques agricoles et maximiser leur productivité. En définissant un profil de production, ils expriment leurs besoins en termes de techniques agronomiques et d'optimisation des ressources.

3.5.4 Scientifiques des données

Les scientifiques des données et les chercheurs en agro-informatique analysent les données brutes et appliquent des techniques avancées (ex. : corrélations entre les conditions climatiques et les rendements, modélisation prédictive, indicateurs de stress hydrique) pour comprendre les dynamiques de production et expliquer les causes cachées des variations de rendement. Les résultats fournis aux décideurs et aux agriculteurs sont justifiés, détaillant les processus sous-jacents et les chemins de raisonnement menant aux recommandations d'optimisation.

Tableau 3.2 – Outils et technologies selon les niveaux d’analyse en production agricole

Niveau / Rôle	Outils et applications	Objectifs
Opérationnel (Agriculteurs)	Applications mobiles dédiées, capteurs connectés, drones agricoles	Suivi des cultures, gestion des intrants, détection des maladies
Managérial (Gestionnaires)	Outils de planification agricole, SIG, blockchain	Optimisation de la production, allocation des ressources
Décisionnel (Décideurs)	Tableaux de bord interactifs, intelligence artificielle pour les prédictions	Prise de décisions stratégiques, élaboration de politiques agricoles
Scientifique (Chercheurs)	Algorithmes d’apprentissage automatique, big data	Analyse prédictive, optimisation des rendements

Exemple 2. Analyse de la production agricole *L’évaluation de la production agricole se fait selon différents niveaux d’acteurs avec des critères spécifiques :*

Niveau Décisionnel (Décideurs)

- **Entrées** : Tendances de production, indicateurs économiques et environnementaux, prévisions climatiques.
- **Sorties** : Politiques agricoles, stratégies d’investissement, gestion des crises agricoles.
- **Objectif** : Optimiser l’impact des politiques agricoles et assurer la durabilité du secteur.
- **Contraintes** : Budget limité, incertitude climatique, contraintes réglementaires.

Niveau Opérationnel (Agriculteurs)

- **Entrées** : Conditions météorologiques, disponibilité des intrants, indicateurs agronomiques.
- **Sorties** : Pratiques agricoles optimisées, augmentation des rendements, meilleure gestion des ressources.
- **Objectif** : Maximiser la productivité en minimisant l’impact environnemental.
- **Contraintes** : Variabilité climatique, coût des intrants, accès aux infrastructures.

Niveau Managérial (Gestionnaires)

- **Entrées** : Ressources disponibles, SIG, prévisions de production, stratégies des décideurs.
- **Sorties** : Plans de culture, allocation des ressources, optimisation de la chaîne d’approvisionnement.
- **Objectif** : Maximiser l’efficacité et la rentabilité en assurant une planification optimale.
- **Contraintes** : Contraintes logistiques, réglementations agricoles, accès aux technologies.

Niveau Scientifique (Scientifiques des données)

- **Entrées** : Données issues de capteurs et satellites, historique des rendements, modèles climatiques.
- **Sorties** : Modèles prédictifs, détection des maladies, outils d’aide à la décision.
- **Objectif** : Améliorer la précision des prévisions et fournir des recommandations optimisées.
- **Contraintes** : Disponibilité et qualité des données, complexité des modèles, adaptation aux réalités locales.

3.5.5 Cas d’optimisation agricole à un niveau managérial

La planification des variétés agricoles selon les spécifications régionales et climatiques est un défi en raison de la variabilité des paramètres. Nous proposons un cas d’optimisation agricole managérial combinant analyse spatio-temporelle et optimisation multi-critères pour identifier les zones propices à la production à grande échelle. Ce scénario utilise des données environnementales (climat, sol, paramètres phénologiques) pour élaborer un modèle de prévision régional. Nous analysons la production d’olives dans les régions steppiques agro-pastorales avec l’NDVI et des enquêtes de terrain pour évaluer les dynamiques de production.

Dans cette section, nous modélisons le problème de localisation afin d'optimiser la production arboricole.

3.5.5.1 Données :

- M : ensemble des ressources physiques nécessaires à la production arboricole (ex. équipements agricoles, machines de culture), où $m = |M|$.
- N : ensemble des tâches liées à la production arboricole (ex. plantation, récolte, irrigation), où $n = |N|$.
- J_i : ensemble des opérations associées à la tâche $i \in N$, tel que l'opération $j \in J_i$ doit être réalisée avant l'opération $j + 1 \in J_i$.
- $M_{ij} \subset M$: ensemble des ressources pouvant réaliser l'opération $j \in J_i$ de la tâche $i \in N$.
- P_{ijk} : durée de l'opération $j \in J_i$ de la tâche $i \in N$ sur la ressource $k \in M_{ij}$.
- B : un entier représentant une valeur très grande utilisée pour les contraintes ou limites.
- C : ensemble des contraintes liées à la faisabilité des alternatives agricoles, en intégrant des aspects spécifiques tels que les ressources énergétiques disponibles ou les limitations d'usage propres aux systèmes de production agricole.
- V : ensemble des variétés oléicoles considérées (ex. Chemlal, Sigoise, Arbequina), où $v = |V|$.
- L : ensemble des localisations géographiques potentielles pour la culture.
- R_{lv} : rendement estimé de la variété $v \in V$ dans la localisation $l \in L$, basé sur des critères agro-climatiques.
- Compat_{lv} : indicateur binaire valant 1 si la variété v est compatible avec la localisation l , et 0 sinon.

3.5.5.2 Variables de décision :

- X_{ijk} : variable binaire valant 1 si l'opération $j \in J_i$ de la tâche $i \in N$ est attribuée à la ressource $k \in M_{ij}$, et 0 sinon.
- Z_{ijhkg} : variable binaire valant 1 si l'opération $j \in J_i$ de la tâche $i \in N$ précède l'opération $g \in J_h$ de la tâche $h \in N$ sur la ressource $k \in M_{ij} \cap M_{hg}$, et 0 sinon.
- S_{ijk} : date de début de l'opération $j \in J_i$ de la tâche $i \in N$ sur la ressource $k \in M_{ij}$.
- C_{ijk} : date de fin de l'opération $j \in J_i$ de la tâche $i \in N$ sur la ressource $k \in M_{ij}$.
- C_i : date de fin de la tâche $i \in N$.
- C_{max} : makespan, soit la durée totale de traitement de toutes les tâches.
- $ELECETRE III_use$: variable binaire valant 1 si des ressources électriques de l'ensemble $ELECETRE III$ sont utilisées pour l'opération j de la tâche i , et 0 sinon.
- Y_{lv} : variable binaire valant 1 si la variété $v \in V$ est implantée dans la localisation $l \in L$, et 0 sinon.

3.5.5.3 Objectifs :

L'objectif principal de ce modèle est d'optimiser la planification et la localisation des opérations agricoles (ex. plantation, entretien, récolte) en tenant compte :

- des contraintes écologiques propres à chaque région (ex. sol, climat, phénologie),
- des ressources matérielles disponibles,
- de la gestion énergétique via l'ensemble *ELECETRE III*,
- du rendement agronomique selon le choix des variétés et leur implantation géographique.

Formellement, les deux objectifs peuvent être définis comme :

$$\min C_{max} \quad \text{et} \quad \max \sum_{l \in L} \sum_{v \in V} R_{lv} \cdot Y_{lv}$$

3.5.5.4 Contraintes :

- Les tâches doivent être exécutées dans un ordre spécifique.
- Une ressource ne peut être affectée qu'à une seule opération à la fois.
- Respect des délais de production (irrigation, récolte...).
- Intégration des conditions écologiques dans les durées d'opérations.
- Une seule variété par localisation : $\sum_{v \in V} Y_{lv} \leq 1 \quad \forall l \in L$.
- Compatibilité obligatoire : $Y_{lv} \leq \text{Compat}_{lv} \quad \forall l \in L, v \in V$.

3.5.5.5 Exemple de planification :

Étape	Tâche	Opération	Ressource	Durée	Début	Fin
1	Plantation	Ouverture de potet	Mise en potet (m_1)	3 h	2025-02-21 08 :00	2025-02-21 11 :00
2	Plantation	Plantation	Cueilleur ou secoueur (m_1)	4 h	2025-02-21 11 :00	2025-02-21 15 :00
3	Récolte	Récolte	Cueillette (m_2)	5 h	2025-02-22 08 :00	2025-02-22 13 :00
4	Récolte	Post-récolte	Cueillette (m_2)	3 h	2025-02-22 13 :00	2025-02-22 16 :00

Tableau 3.3 – Attribution des ressources et planification des tâches arboricoles

Utilisation des ressources électriques (*ELECETRE III*) : - *ELECETRE III_use₁* = 1 pour l'irrigation lors de la plantation. - *ELECETRE III_use₂* = 1 pour le traitement post-récolte.

Calcul du makespan :

$$\text{Makespan} = 2025-02-22 16:00 - 2025-02-21 08:00 = 32 \text{ heures}$$

3.5.5.6 Décision finale de localisation variétale

À la lumière de cette optimisation, une **décision stratégique peut être formalisée** pour recommander l'implantation des variétés les plus performantes dans les régions les plus compatibles. Cette décision est guidée par le critère de rendement R_{lv} tout en respectant la compatibilité agro-climatique Compat_{lv} .

Localisation	Variété choisie	Compatibilité	Rendement estimé (q/ha)
Laghouat	Chemlal	Oui	22
El Bayadh	Sigoise	Oui	18
Naâma	Arbequina	Non	0 (non retenue)

Ainsi, le modèle propose une ****localisation optimale**** de la variété *Chemlal* à Laghouat et *Sigoise* à El Bayadh, en maximisant la production tout en respectant les spécificités régionales.

3.6 Conclusion

L'objectif principal de ce chapitre est de proposer un ensemble de principes de conception pour développer des fonctions de coût mathématiques dans un cadre de modélisation basé sur les données. En général, le développement de modèles de production (*Prods*) est freiné par l'absence de processus de conception structuré. Souvent, ces modèles sont soit esquissés de manière informelle, directement implémentés dans des outils sans conception préalable, soit conçus mentalement sans documentation formelle de conception.

Pour résoudre ce problème, il est nécessaire de repenser le processus de développement des *Prod*, en mettant l'accent sur la conception et les tests afin de créer des solutions efficaces pour les *Prod*.

Notre principale contribution est l'introduction d'un cadre basé sur un modèle de coût spécifiquement destiné à la conception et au développement des *Prods* de prédiction mathématique des coûts. Ce cadre définit un processus systématique pour collecter et gérer les données relatives à la production agricole, en se concentrant sur plusieurs sources de données et processus. Il intègre les connaissances des concepteurs pour sélectionner les variables d'entrée liées aux facteurs d'impact les plus sensibles de la production culturelle. Il est facilement personnalisable et permet aux utilisateurs de parcourir différents modèles de prévision en fonction de leurs tâches de conception, simplement en attachant les paramètres d'entrée aux spécifications du modèle.

Nous qualifions cette contribution de cadre de modélisation basé sur les données pour la prévision de la production agricole. Le processus de conception proposé comprend les phases suivantes : (1) Expression des besoins, (2) Identification des dimensions, (3) Identification des facteurs impactant les coûts de production agricole, (4) Construction du modèle de prévision, et (5) Test et évaluation.

Le cadre AFFAC propose une approche intégrée qui soutient différents niveaux d'utilisateurs, allant des agriculteurs aux décideurs, en passant par les gestionnaires et les scientifiques. Au niveau managérial, des méthodes comme ELECTRE III permettent d'optimiser les décisions multi-critères liées à la planification agricole. Cette complémentarité d'usages renforce la prise de décision fondée sur les données pour une production agricole durable et efficiente.

Zone d'étude : Nord de l'Algérie



« Chaque individu apporte au monde sa contribution unique. »

— Jack Kornfield

Sommaire

4.1	Introduction	100
4.2	Etude de cas	100
4.2.1	Région d'étude et données utilisées	100
4.2.2	Étude des tendances des indices climatiques	101
4.2.3	Région d'étude : oleiculture dans le nord de l'Algérie	107
4.2.4	Production d'olives en Algérie	109
4.2.5	Production d'olives dans la region de tiaret	111
4.3	Mécanisme d'instanciation de AFFAC sur un cas d'étude	117
4.4	Instanciation de notre Framework AFFAC	120
4.4.1	Aperçu du flux de travail de conception du modèle de prévision	121
4.4.2	Scénario d'application	121
4.4.3	Généricité de notre framework AFFAC	124
4.5	Conclusion	125

4.1 Introduction

Le climat joue un rôle déterminant dans le mode de vie des populations et des cultures, tout en impactant profondément les économies des nations. Les projections climatiques, influencées par les activités humaines récentes et les importantes émissions de gaz à effet de serre d'origine anthropique, entraînent des conséquences significatives sur les productions agricoles [115]. Ce chapitre présente une instanciation de AFFAC dans une étude de cas sur la culture de l'olivier. Nous appliquons notre cadre à une étude de cas axée sur les oleicultures en Algérie, démontrant son application pratique et sa pertinence.

4.2 Etude de cas

4.2.1 Région d'étude et données utilisées

L'étude se base sur une cinquantaine de stations d'observation météorologique du réseau de l'Office National de la Météo (ONM), réparties sur la partie Nord de l'Algérie (32°N, 38°N; 2°W, 9°E), comme illustré à la Figure 4.1. La période d'étude considérée s'étale de 1980 à 2017. Les données ont été téléchargées à partir du site de la *National Oceanic and Atmospheric Administration* (<https://www7.ncdc.noaa.gov/CD0/cdo>).

La Figure 4.1 illustre la répartition des stations météorologique de l'ONM utilisées dans l'étude.

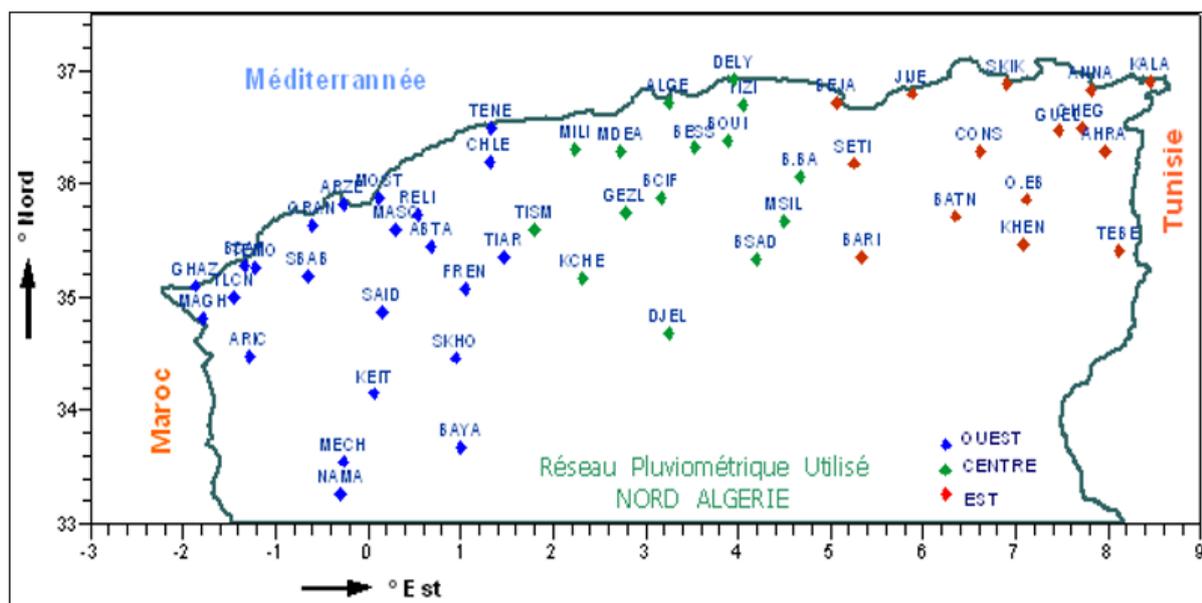


FIGURE 4.1 – Répartition des stations météorologique de l'ONM utilisées dans l'étude.

4.2.2 Étude des tendances des indices climatiques

Dans cette section, nous allons étudier les tendances des indices climatiques.

4.2.2.1 Teste de l'Indice d'aridité de De Martonne

L'indice d'aridité est censé représenter le degré de sécheresse du climat à un lieu donné, Ces indices ont été calculés pour la première fois au début du XXe siècle par Wladimir Köppen et Rudolf Geiger, plus tard d'autre scientifique comme Charles Warren Thornthwaite puis Mikhaïl Ivanovitch Budyko ont proposés les leurs. On en cite l'indice annuel d'aridité De Martonne (1926 et 1941) qui utilise un indice regroupant des données hydrométriques et thermiques, en particulier l'expression analytique $IA = P/(T+10)$ Où P et T sont respectivement les valeurs annuelles des précipitations et des températures[37]. Les climats sont classés selon cet indice en :

Indice de De Martonne	Climat
0-10	Aride à extrêmement aride
10-20	Semi-aride
20-30	Semi-humide
30-60	Humide
>60	Très humide

Tableau 4.1 – Classification climatique selon l'indice de De Martonne.

Les gradients de ces indices peuvent également servir à délimiter des zones en fonction de leur pluviométrie.

La cartographie de l'indice de De Martonne (voir Figure 4.3 et Figure 4.2) sur les deux dernières périodes détectées par les tests non paramétriques indique un déplacement significatif des types du climat dans un gradient Sud-Nord et Est Est-Ouest, le type de climat aride à extrêmement aride s'est avancé pour atteindre les régions d'intérieur jusqu'à la station de Ghardaïa et les régions voisines.

Le type de climat semi-aride a dominé le Nord-Ouest de l'Algérie allant de Mechria, AinSafra et El Bayadh jusqu'à la station de Tlemcen. On peut remarquer aussi le même type de climat dans la région de Chlef. De ce fait, que les stations du Nord Est qui n'ont pas subi vraiment de changements vers l'aridité Figure 4.3.

4.2.2.2 Analyse des précipitation dans le nord de l'algerie

La moyenne arithmétique Le graphique Figure 4.4 représente l'évolution de la moyenne des précipitations annuelles aux près des 45 stations étudiées ; On remarque que cette dernière augmente de l'Ouest à l'Est et diminue en fur et a mesure en dirigeant vers le sud.

– **La région Nord-Ouest** : Cette région présente une faible portion de précipitations par rapport

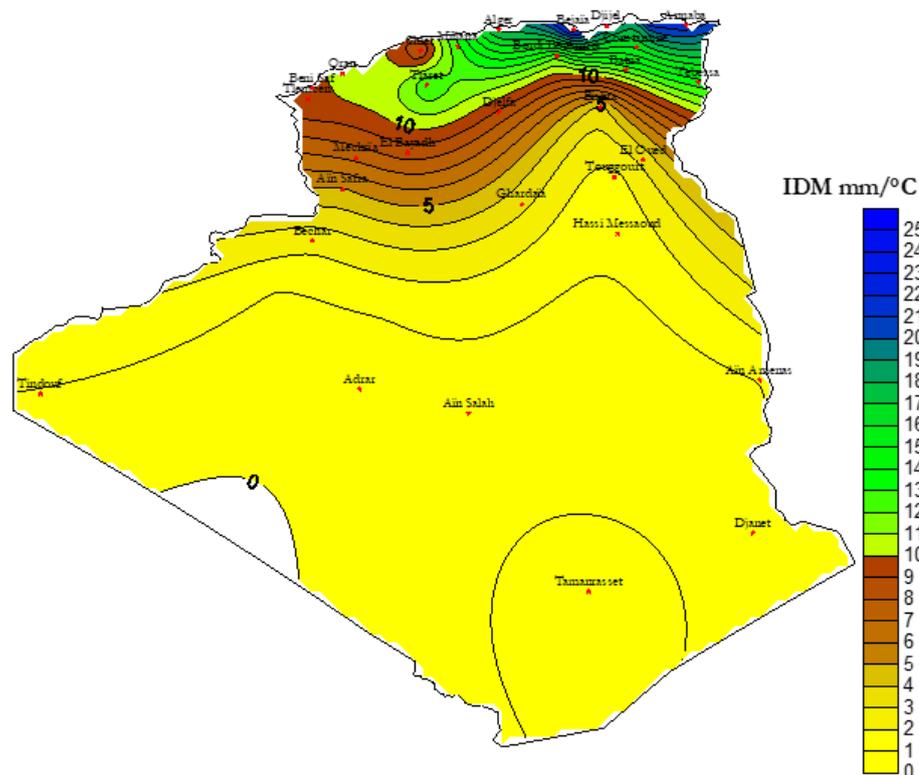


FIGURE 4.2 – Répartition géographique de l'indice d'aridité de De Martonne en Algérie durant la période 1996-2008.

à la moyenne des moyennes annuelles (M Moy) des stations, qui est de 421.36 mm.

- 76,19 % des stations enregistrent des quantités moyennes annuelles de précipitations inférieures à cette moyenne. La station Kheiter (197,66 mm) présente la quantité la plus faible.
- 23,81 % des stations enregistrent des quantités supérieures à cette moyenne, notamment :
 - Ténès (531,49 mm),
 - Tlemcen (513,82 mm),
 - Tiaret (503,59 mm),
 - Ain Témouchent (450,72 mm),
 - Freneda (439,69 mm).
- **La région Nord-Centre** : Dans cette région, les stations sont réparties en deux groupes homogènes :
 - 50 % des stations enregistrent des précipitations moyennes annuelles supérieures à 421.36 mm, avec un maximum de 746.85 mm à la station Miliana.
 - 50 % des stations enregistrent des précipitations inférieures, les plus faibles étant :
 - Msila (214,33 mm),
 - Boughzoul (218,24 mm).
- **La région Nord-Est** : Cette région enregistre les plus grandes quantités moyennes annuelles de précipitations, avec un maximum de 928.09 mm à la station Jijel.
 - 69,23 % des stations enregistrent des niveaux de précipitations supérieurs à 421.36 mm.

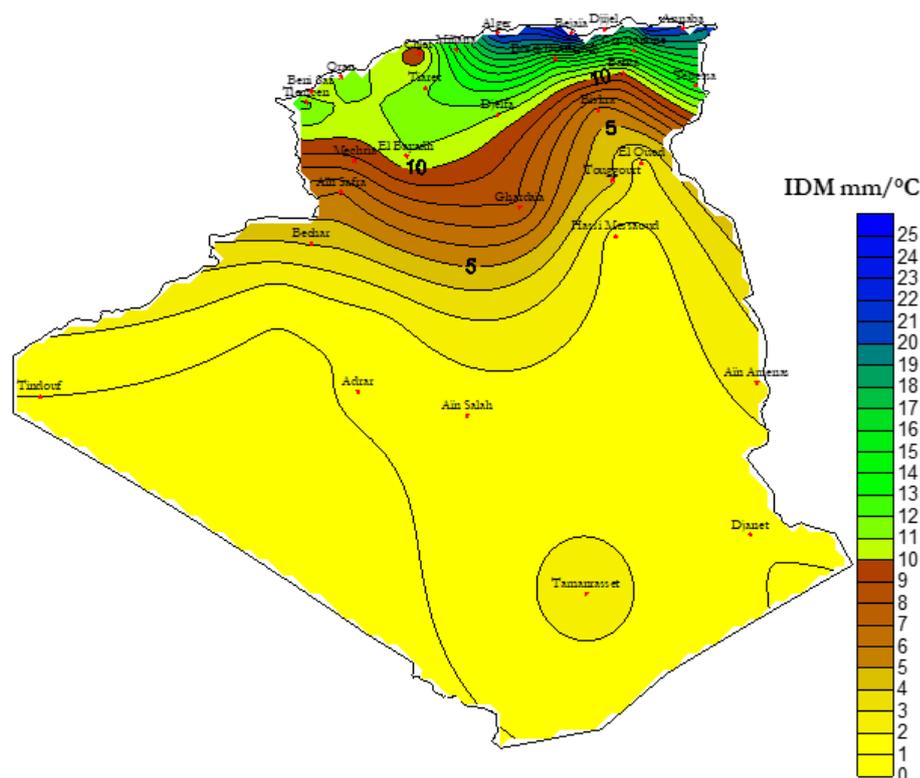


FIGURE 4.3 – Répartition géographique de l'indice d'aridité de De Martonne en Algérie durant la période 1982-1996 [96].

- 30,77 % des stations enregistrent des précipitations moyennes annuelles inférieures à cette valeur, avec un écart relativement réduit.

Ces résultats peuvent être expliqués par la situation géographique dont les fortes quantités moyennes annuelles de précipitations enregistrées sont localisées au niveau des stations situées entre la méditerranée et la chaîne montagneuse de l'Atlas Tellien qui restreint la circulation des nuages vers le sud ce qui est confirmé par l'enregistrement de la plus faible quantité moyenne annuelle de précipitations à la station Kheiter (197.66 mm) située au Sud Ouest et la plus grande à La station Jijel (928.09 mm) située au Nord Est de la région étudiée.

Après l'analyse des précipitations et des températures, il est impératif de présenter un climagramme de Bagnols et Gausse de quelques régions du nord, ce qui nous permet d'avoir un aperçu des périodes sèches et humides d'une part et de voir la tendance des moyennes de températures et des cumuls de précipitation d'autre part.

4.2.2.3 Régionalisation des précipitations

La régionalisation du paramètre précipitation est réalisée dans le but de l'identification des régions ayant des caractéristiques pluviométriques homogènes. Afin de réaliser la régionalisation, nous nous sommes basés sur l'analyse en composantes principales avec rotation des vecteurs propres en optant pour

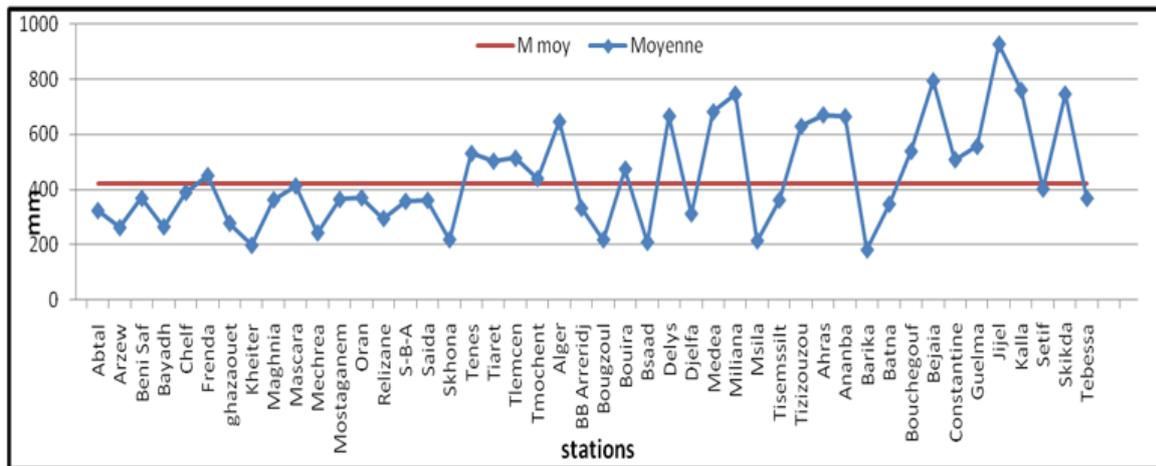


FIGURE 4.4 – Représentation graphique des moyennes annuelles des stations.

Oblimin comme critère de rotation [139, 145]

La majorité des cas d'application de l'ACP aux données atmosphériques se retrouve dans l'analyse des champs de données dont les dimensions sont élevées. Les sorties de l'analyse en composantes principales sont cartographiées, notamment les vecteurs propres (loading) qui peuvent décrire la variabilité spatiale ou temporelle du paramètre étudié. Ces cartes sont utilisées après analyse pour en déduire la régionalisation. Dans une première étape et dans le but de sélectionner les premières composantes les plus significatives, une Analyse en Composantes Principales sans rotation est appliquée à l'ensemble des données. Le fichier de données est préparé sous forme de matrice où les colonnes représentent les stations et les lignes représentent le temps. Ensuite une ACP avec rotation (Oblique) est appliquée aux composantes retenus.

Plusieurs investigations de ce genre ont été conduites en analysant les champs des vecteurs propres [152]. Vu que les quantités de précipitations dans les différentes régions sont très hétérogènes, la matrice de dispersion utilisée dans le calcul des valeurs propres et des vecteurs propres est la matrice de corrélation. Toutefois une comparaison entre les utilisations des deux matrices de dispersion est faite Figure 4.7 et Figure 4.8 .

La Figure 4.7 et Figure 4.8 permettent de comparer une structure simple utilisant la matrice des variances-covariances [Figure 4.7 (a1) et (b1)] et la matrice des corrélations [Figure 4.7 (a2) et (b2)]. L'analyse de ces figures montre bien que les Figure 4.8 (a2) et (b2) présentent plus de variables dans les hyperplans des composantes principales et moins de variables dites complexes c'est-à-dire celles qui n'appartiennent à aucun hyperplan. La solution utilisant la matrice de corrélation donne une meilleure classification. D'après [156], « si la matrice de variance-covariance est utilisée, les postes pluviométriques ayant une grande variabilité auront plus d'influence sur la première composante principale ». Dans ce cas précis, le champ des vecteurs propres permet le partage du domaine en régions homogènes. La structure des champs illustrée dans la Figure 4.7, démontre que les champs basés sur la matrice de corrélation sont plus structurés et les limites sont mieux discernables pour la valeur de l'isoplette ou d'iso-saturation 0,4. Cependant, [139] a utilisé comme limite des régions la valeur 0.6. A l'échelle de l'hémisphère, Buell (1975, 1979) avait utilisé la valeur 0,2.

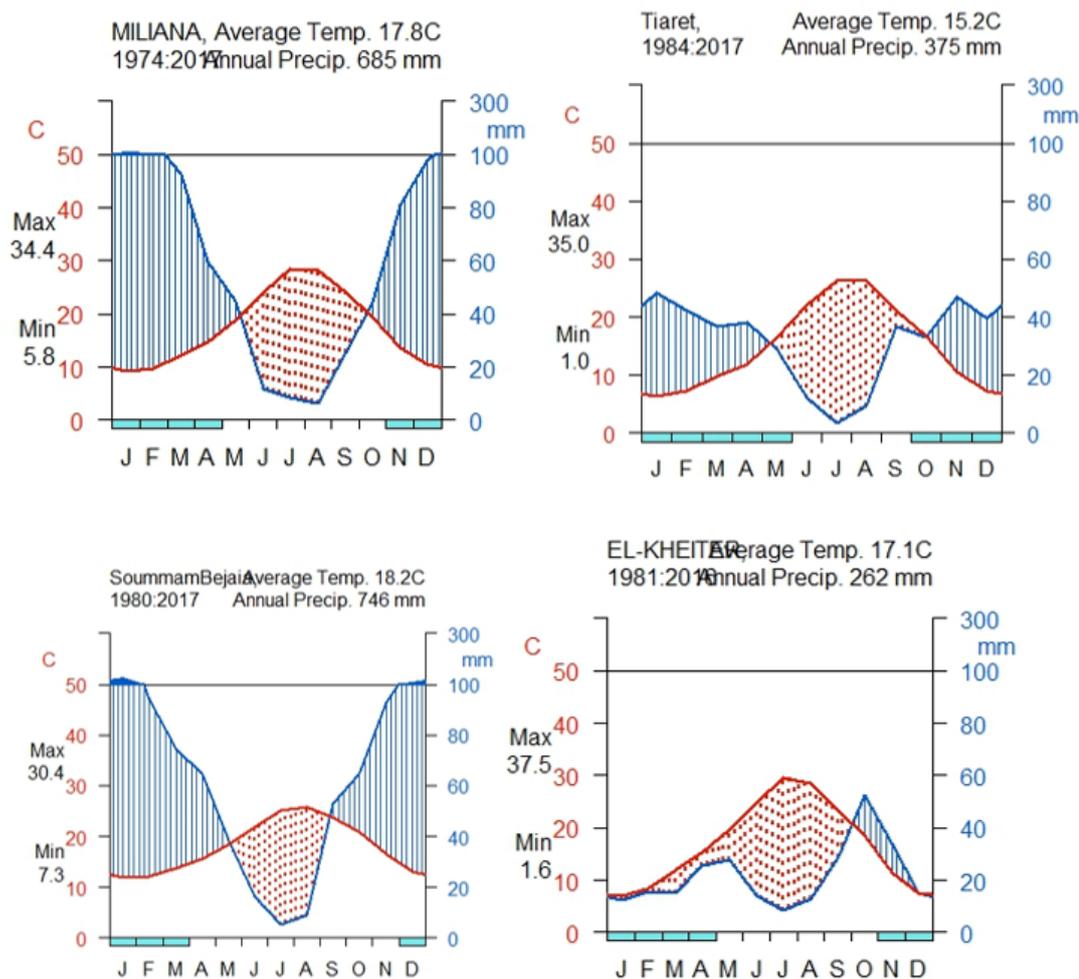


FIGURE 4.5 – Diagramme ombrothermique de walter et lieth de différentes régions du nord.

Les vecteurs propres présentés dans la Figure 4.6 issus de la rotation sont représentés géographiquement à partir de l'isoplète 0,4, frontière de la région expliquée par l'hyperplan d'une composante principale Figure 4.6. Ceci est réalisé pour n'importe quel champ pluviométrique que nous voulons régionaliser. Cette analyse permet le partage du domaine d'étude en un nombre k de régions homogènes. Les limites géographiques des régions, ainsi que leur nombre, étant déterminés par l'ACP.

Le résultat de la régionalisation est illustré dans la Figure 4.9. Cette régionalisation est réalisée par saison (Septembre-Décembre ; Janvier-Mars ; Avril-Juin. Ce choix est motivé par les différentes utilisations des résultats notamment la prévision saisonnière. La Figure 4.9 (a) représente la régionalisation de la saison allant de septembre à décembre où cinq régions homogènes apparaissent :

- La région Ouest (W)
- La région Centre Nord (CN)
- La région Nord-Est (NE) :
- La région Centre Sud (CS) :
- La région Sud-Est (SE) :

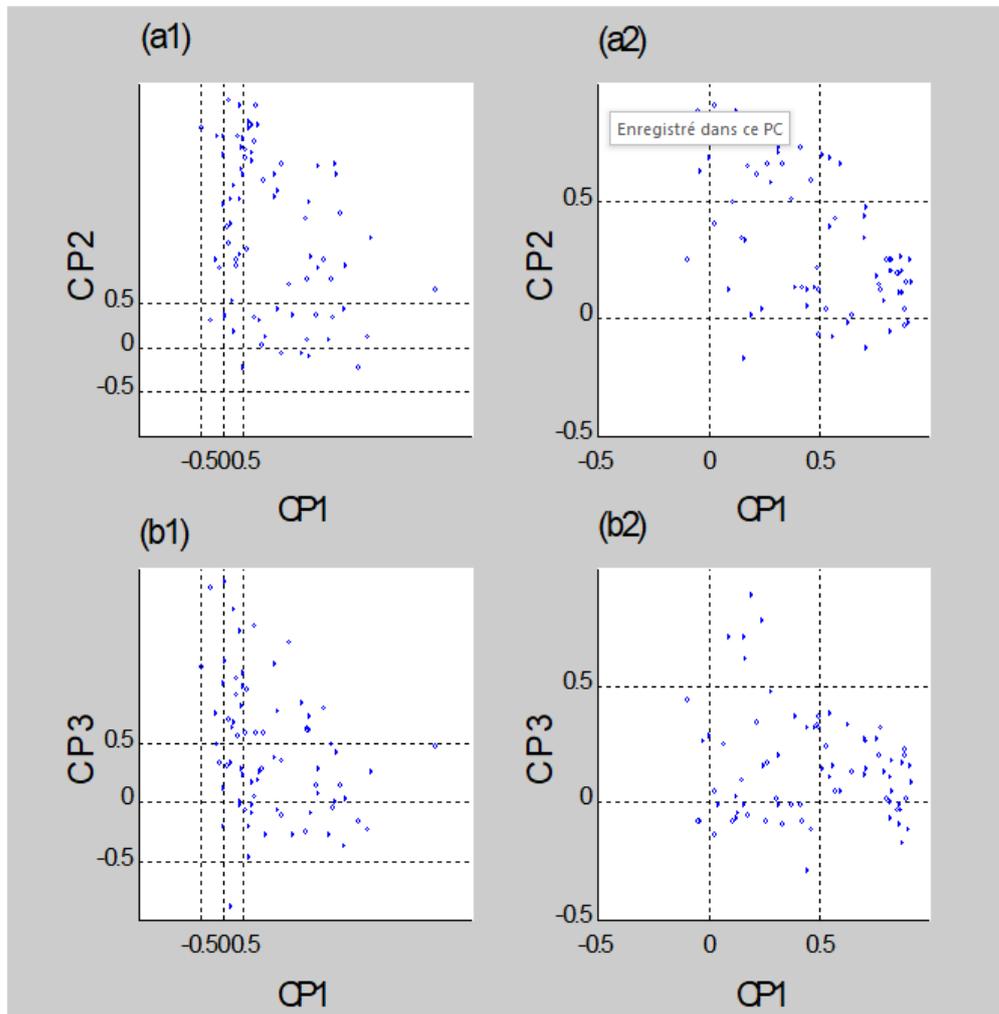


FIGURE 4.6 – Représentation des nuages de points des valeurs prises par les deux vecteurs propres correspondant aux deux composantes principales issue d'une rotation oblimin montrant la qualité de la structure simple en prenant comme matrice de dispersion la covariance (a1) et (b1) et la corrélation (a2) et (b2), les données sont celles des précipitations.

La Figure 4.8 représenté la régionalisation de la saison des mois allant de janvier à mars où quatre régions sont observées notées W (Ouest), NE (Nord-Est), CS (Centre Sud) et SE Sud-Est. Entre la Figure 4.8 (a) et la Figure 4.8 (b), nous constatons que les régions W et NE s'élargissent pour partager le domaine pris par la région CN ; ainsi le Nord de l'Algérie durant l'automne et le début d'hiver est subdivisé en trois grandes régions homogènes Ouest, Centre et Est, pour évoluer durant l'hiver jusqu'au début du printemps en seulement deux régions W et NE. Les régions représentant les parties Sud de notre zone d'étude évoluent aussi durant les deux saisons (Sep.-Déc. et Jan.-Mars) : la région CS s'élargit tandis que la région SE rétrécit. Enfin, en analysant la Figure 4.9. (c) durant la saison allant du printemps jusqu'en été, nous constatons la subdivision du Nord de l'Algérie en trois grandes régions côtières NW, CN et NE Figure 4.9 (e) où les pluies sont d'origines classique qui proviennent du nord et du nord Ouest , alors que les régions Sud du domaine sont subdivisées en deux régions notées SW et CS qui peuvent être regroupées en une seule représentant les hauts plateaux ; l'origine des pluies dans ces régions est locale

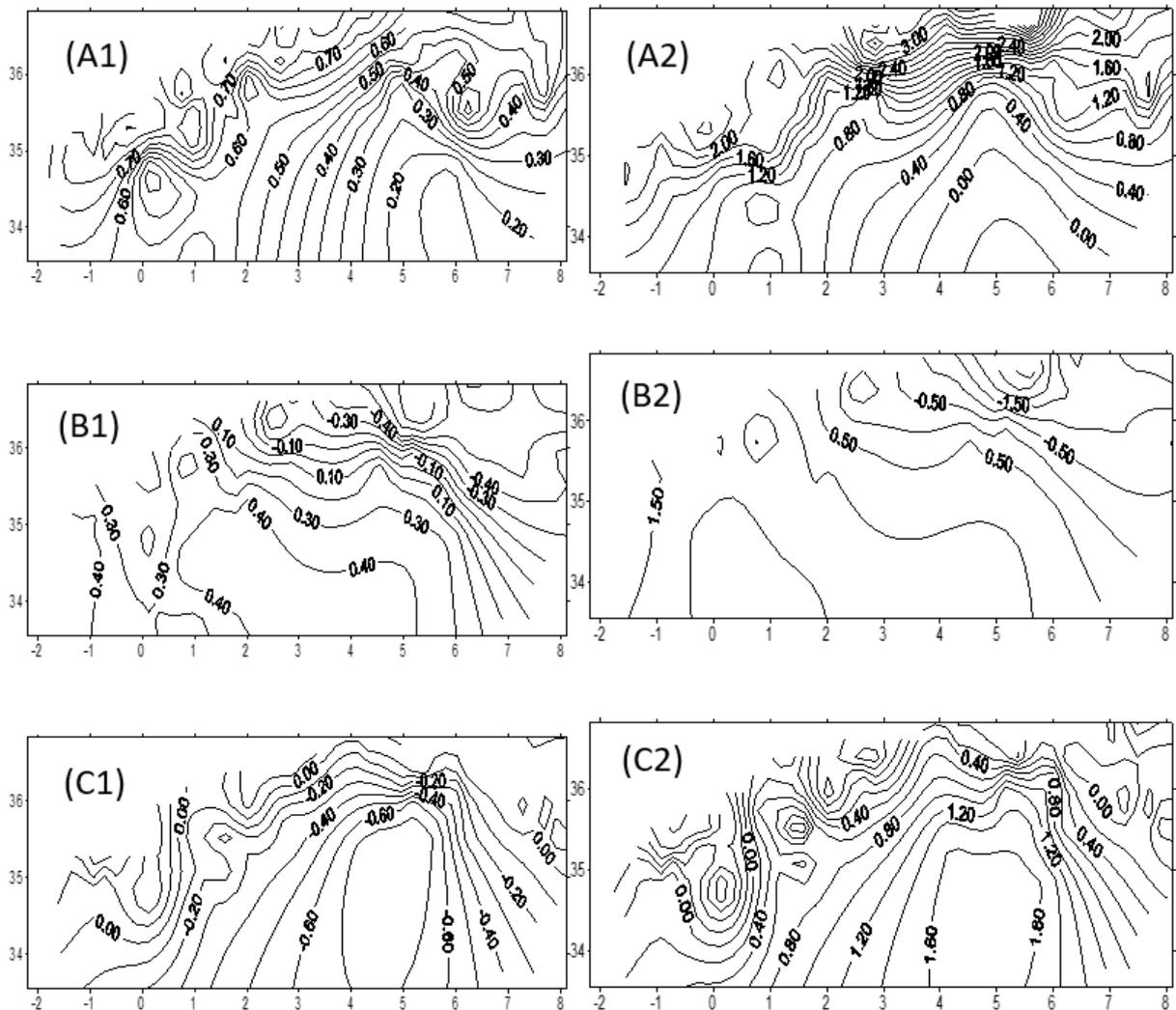


FIGURE 4.7 – Représentation géographique des vecteurs propres correspondant aux trois premières composantes, du champ de précipitations; confirmant le schéma de Buell [27]. Cas de la matrice de corrélation (A1, B1, C1) et matrice de covariance (A2, B2, C2) comme matrice de dispersion.

due au reliefs.

Les figures ainsi réalisées montrent que les régions se chevauchent, nous estimons que ce résultat est logique car les limites des régions ne sont pas figées et peuvent se superposer dans les frontières.

4.2.3 Région d'étude : oleiculture dans le nord de l'Algérie

Le nord de l'Algérie, une région clé pour la production oleicole, fait face à des défis dans la prévision des rendements agricoles en raison des conditions environnementales fluctuantes et des pratiques agricoles³.

Prévoir la production d'olive dans cette région est essentiel pour améliorer la qualité des produits,

3. <https://lecomptoiralgerien.com/collections/huiles-dolive>

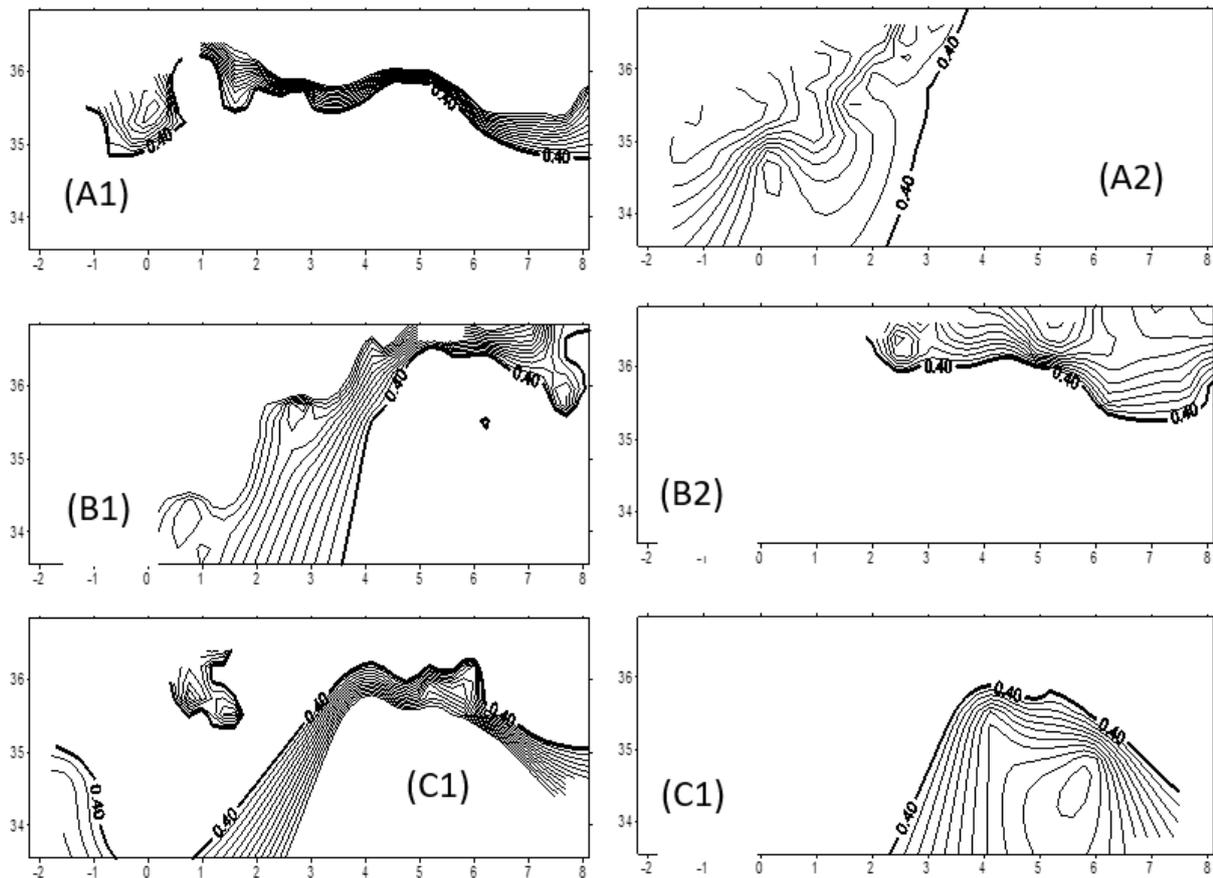


FIGURE 4.8 – Représentation géographique des vecteurs propres issue d'une rotation (oblimin) correspondant aux trois premières composantes principales, du champ de précipitations. La matrice de corrélation (A2, B2, C2) ou la matrice de covariance (A1, B1, C1) étant prise comme matrice de dispersion.

aligner la production sur la demande du marché, et mener des analyses prédictives afin de recommander des améliorations et d'éviter les anomalies de production.

La région nord de l'Algérie s'étend sur une superficie de 382 000 km², soit environ 20% du territoire national, où sont concentrées presque toutes les populations et activités. Cette région complexe se compose d'une succession de chaînes montagneuses, de zones côtières et sublittorales, ainsi que de plaines. Elle se divise en deux sous-ensembles : le Tell et les Hautes Plaines Steppiques, répartis sur 36 communes.

Le verger oléicole en Algérie couvre une superficie de 431 506 hectares, comptant 60 632 901 oliviers [1].

Au cours des 20 dernières années, les surfaces cultivées en oliviers ont connu une croissance significative (Figure 4.10.), passant de 168 080 hectares en 2000 à 294 200 hectares en 2010, atteignant finalement 431 506 hectares en 2019, soit une augmentation de 75%.

Environ 80% de ces oliveraies sont situées dans des zones montagneuses, sur des terres accidentées et marginales avec une faible fertilité. Les 20% restants se trouvent dans les plaines occidentales du pays[117].

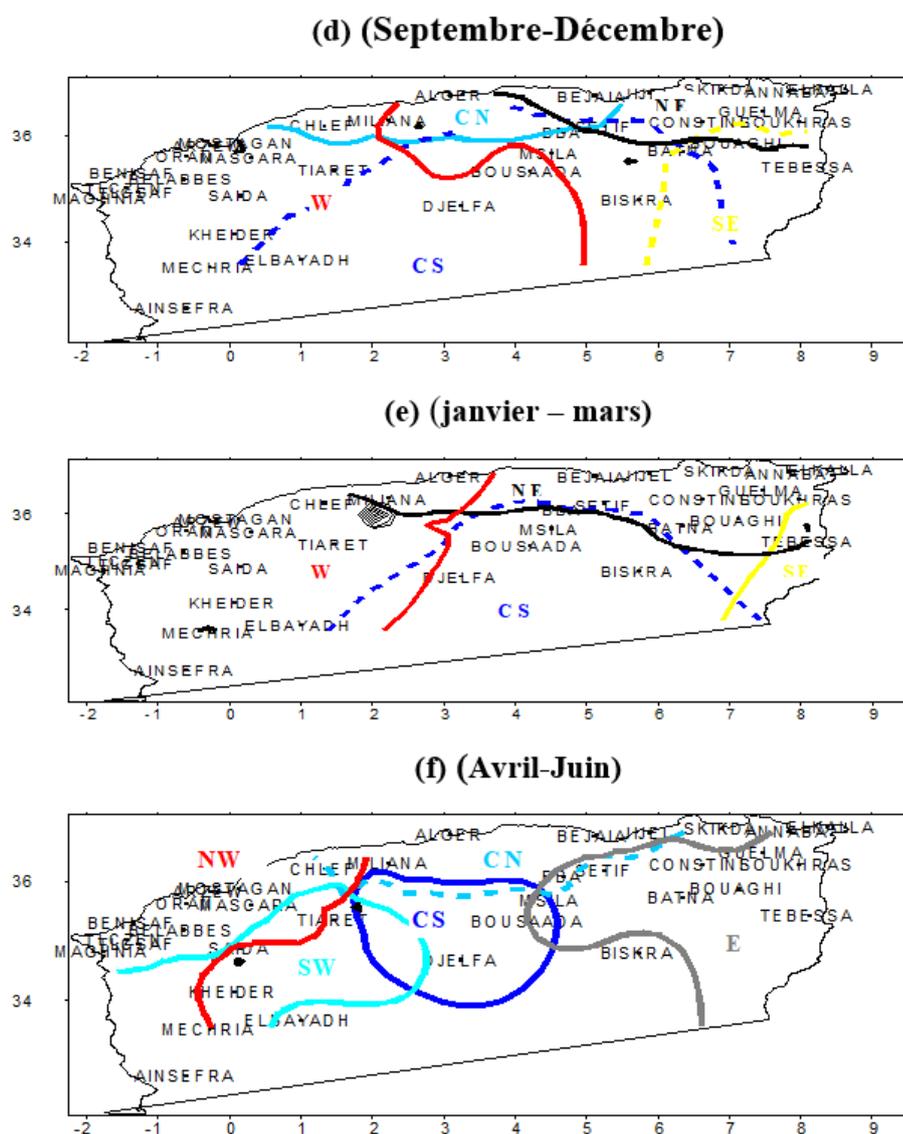


FIGURE 4.9 – Régionalisation des précipitations basée sur le réseau des station de l'ONM .

Année	Nombre d'oliviers	Nombre d'hectares	Production (qx)
2000	16 702 610 arbres	168 080 ha	2 171 124
2010	36 335 782 arbres	294 200 ha	3 112 520
2019	60 632 901 arbres	431 506 ha	8 687 541

Tableau 4.2 – Évolution du nombre d'oliviers, des surfaces cultivées et de la production entre 2000 et 2019.

4.2.4 Production d'olives en Algérie

Le verger oléicole en Algérie s'étend sur 431 506 ha avec 60 632 901 oliviers [10]. Durant ces 20 dernières années les surfaces cultivées en oliviers ont connu une grande croissance (voir le tableau n°

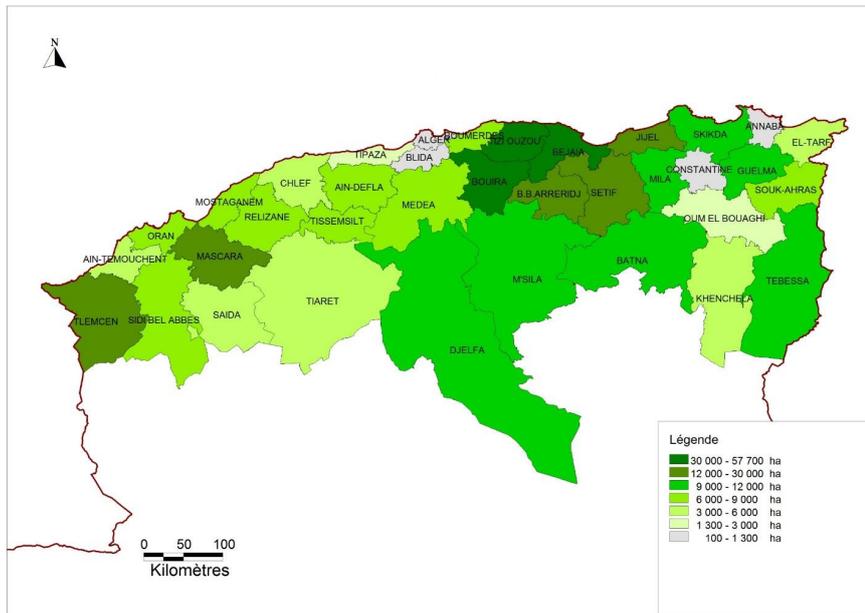


FIGURE 4.10 – Visualisation des zones oléicoles avec un dégradé de couleurs indiquant la répartition par région dans la zone d'étude.

05), passant de 168 080 ha en 2000 à 294 200 ha en 2010 et à 431 506 ha en 2019. Soit une augmentation de 75 par rapport à 2010 et de 156,8 par rapport à 2000 [10, 18]. Cette dernière a entraîné l'augmentation des rendements en huile d'olive, notamment avec la création des nouvelles zones productrices au sud et aux régions septentrionales du pays. Cependant, le nombre d'oliviers a connu également une grande évolution, il est de 16 702 610 arbres en 2000, il passe à 36 335 782 arbres en 2010 pour atteindre les 60 632 901 arbres en 2019. Soit une hausse égale à 117,5 par rapport à 2010 et 263 par rapport à 2000.

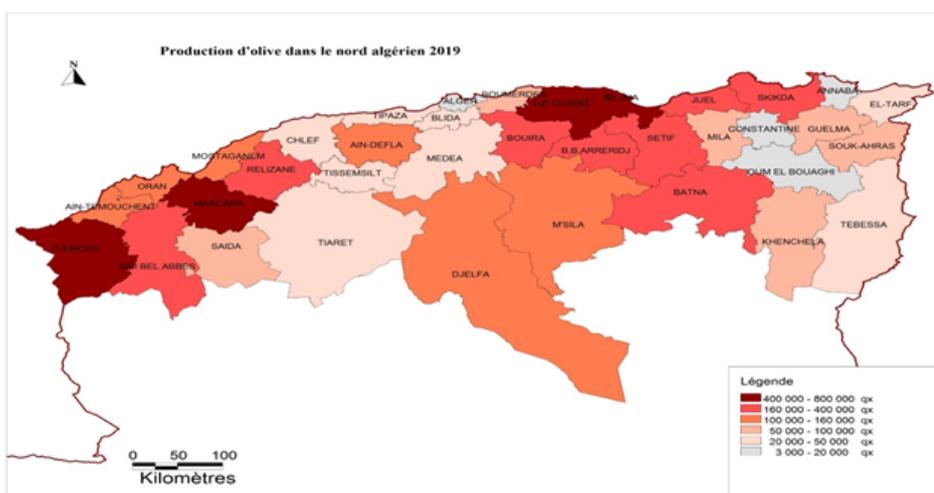


FIGURE 4.11 – Distribution de la production oléicole dans le Nord de l'Algérie

4.2.5 Production d'olives dans la region de tiaret

Afin d'évaluer la production oléicole dans la région de la wilaya de Tiaret, il est essentiel de prendre en compte la variabilité des différents paramètres écologiques des cultures oléicoles. La superficie oléicole de Tiaret a connu une évolution significative au cours des dernières années, atteignant aujourd'hui environ 3964.83ha, grâce aux efforts de plantation et d'extension des vergers. Cette progression témoigne de l'intérêt croissant pour la culture de l'olivier dans cette région, notamment dans les zones au climat favorable. Par ailleurs, des traitements spécifiques sont envisagés dans la partie sud-est de la wilaya pour optimiser le rendement des cultures et s'adapter aux conditions locales. Ces initiatives visent à renforcer la contribution de Tiaret à la production nationale d'huile d'olive.

4.2.5.1 Présentation de l'exploitation oléicole de Djaïb

Dans cette exploitation de 113 ha se localisant dans le sud est de la wilaya de Taret plus exactement commune de rechaiga, existe 61 parcelles arboricoles dont 33 sont de différentes variétés oléicoles .

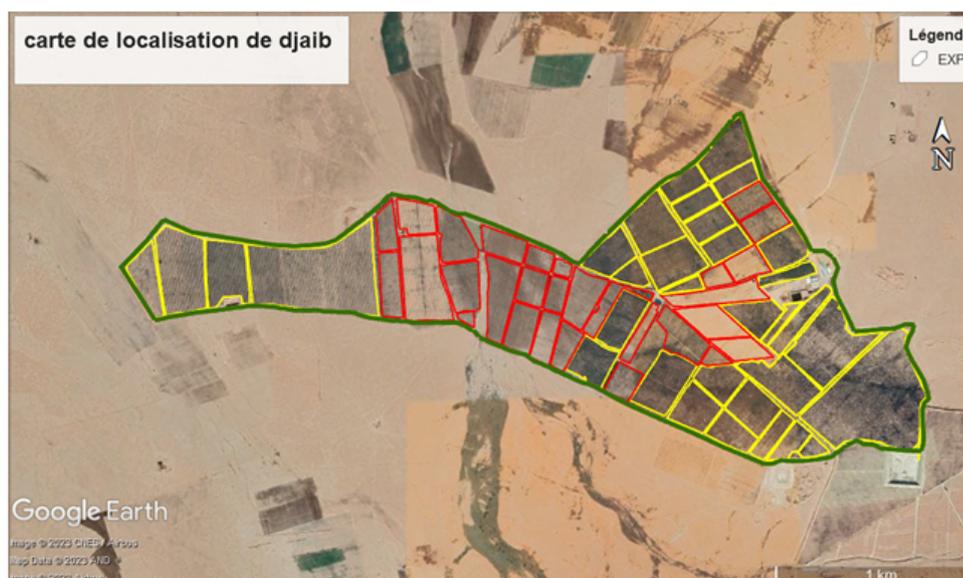


FIGURE 4.12 – Géolocalisation de la parcelle oléicole Djaïb

4.2.5.2 Paramètres morphologiques (feuille) et paramètres physiologiques (réflectance foliaire) de quelques variété d'olives

Tableau 4.3 – Le périmètre et réflectance foliaire

Variétés	SF (cm ²)	Per (cm)	LF (cm)	R (%)	V (%)	B (%)	T (%)
Picual	5.918	14.362	6.97	22.50	30.85	11.00	21.45
Chemllale	9.026	17.034	8.22	19.65	32.83	6.71	19.73
Sigoise	8.3	17.546	8.546	24.71	61.71	7.52	31.31

Le tableau présente les paramètres physiologiques liés aux réflectances foliaires de trois variétés d'oliviers : *Picual*, *Chemllale* et *Sigoise*. Les données sont exprimées en pourcentage (%) pour les différentes bandes spectrales : rouge (R), verte (V), bleue (B) ainsi que la réflectance totale (T). Voici une analyse comparative détaillée des données.

4.2.5.3 Analyse des réflectances foliaires des variétés d'oliviers

La figure ci-dessous illustre les différences de réflectance foliaire des variétés *Picual*, *Chemllale*, et *Sigoise* dans les bandes spectrales rouge, verte, bleue, ainsi que la réflectance totale. On observe que la variété *Sigoise* présente une réflectance verte particulièrement élevée (61,71 %), contrairement à *Chemllale* et *Picual*. Par ailleurs, *Chemllale* montre la plus faible réflectance dans la bande bleue (6,71 %), indiquant une absorption plus importante de cette longueur d'onde. Ces différences reflètent les caractéristiques physiologiques et morphologiques propres à chaque variété.

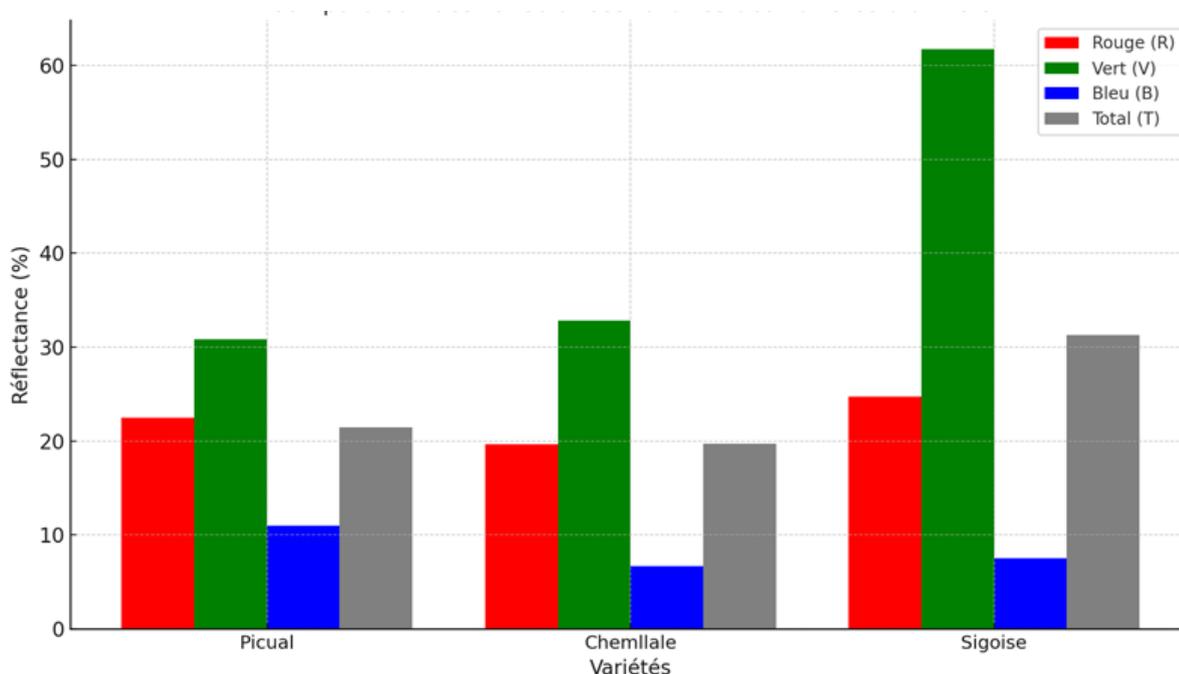


FIGURE 4.13 – Reflectance foliaire des diferentes variété d'olives.

4.2.5.4 Interprétation et applications

Les variations de réflectance observées entre les trois variétés d'oliviers traduisent des différences dans leurs caractéristiques physiologiques, telles que la teneur en chlorophylle, l'épaisseur des feuilles ou leur adaptation à des conditions environnementales spécifiques. Ces résultats sont utiles pour :

- **La sélection variétale** : Identifier les variétés les mieux adaptées aux conditions lumineuses spécifiques.
- **La gestion agricole** : Optimiser les pratiques culturales en fonction des besoins en lumière et des performances photosynthétiques.

- **La télédétection** : Discriminer les variétés à l'aide d'images satellitaires ou de drones en se basant sur leurs signatures spectrales distinctes.

4.2.5.5 Indice de végétation normalisé

L'olivier passe par les phases suivantes : (1) Janvier, Février : induction, initiation et différenciation florale ; (2) au courant de mois de Mars : croissance et développement des inflorescences à l'aisselle des feuilles que portent les rameaux de l'année précédente ; (3) Avril : pleine floraison ; (4) Fin Avril-début Mai : fécondation et nouaison des fruits ; (5) Juin : début de développement et grossissement des fruits ; (6) Septembre : véraison ; (7) Octobre : maturation du fruit et son enrichissement en huile (8) Mai -Novembre à Janvier : récolte des fruits. Donc La période la plus intense du cycle annuel se déroule en été. Au cours de cette phase, les besoins en eau et en nutriments de l'arbre sont les plus intenses. Mais en ce qui concerne l'NDVI généralement il est lié aux conditions climatiques soit les précipitations ou les températures.

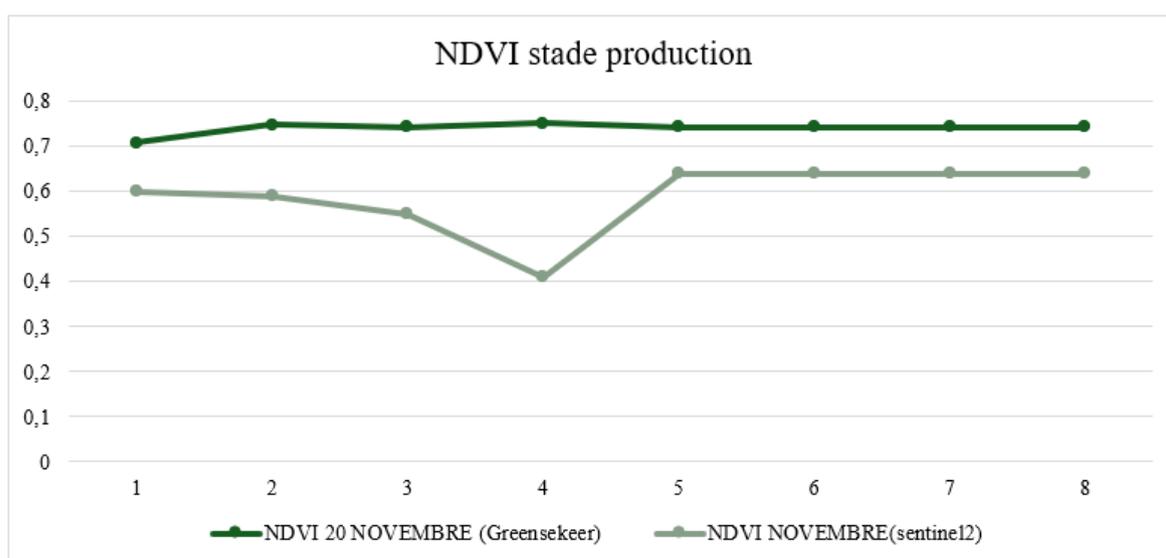


FIGURE 4.14 – Représentation graphique de l'NDVI (stade de production).

Tableau 4.4 – Corrélation de l'NDVI stade de production

Mesures	Green Seeker Novembre
Sentinel2 Novembre	
Green Seeker Novembre	0.51671
Sentinel2 Novembre	-0.27068

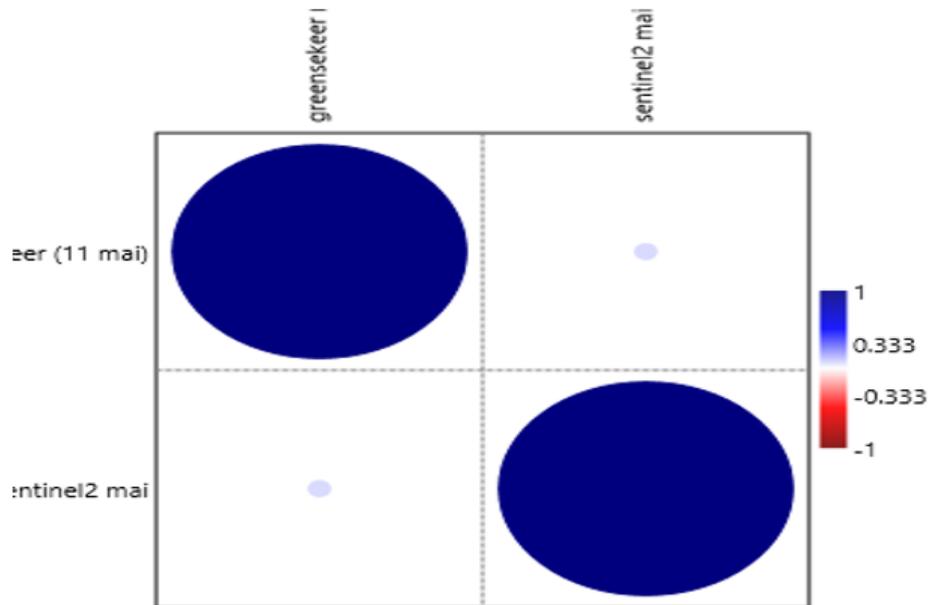


FIGURE 4.15 – Corrélation de l'NDVI stade de production.

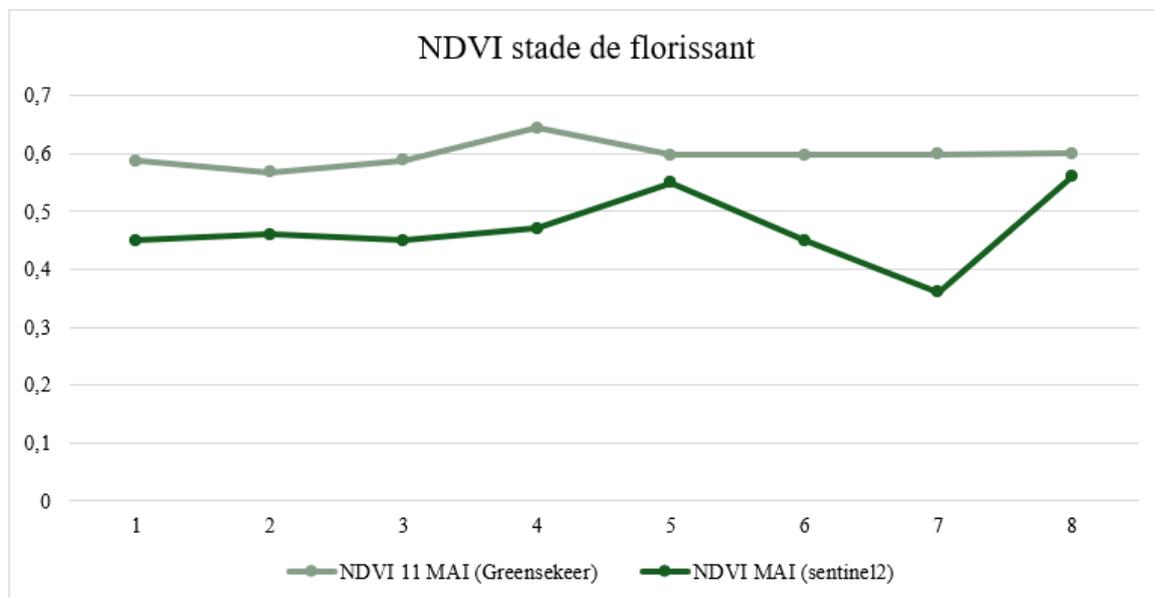


FIGURE 4.16 – Représentation graphique de l'NDVI (stade de florissant).

4.2.5.6 Analyse des données saisonnière de l'NDVI

D'après le tableau l'analyse des données saisonnières de l'NDVI on a remarqué une évolution des valeurs de l'indice de végétation du printemps à l'automne passant par l'été avec une régression remarquable en hiver donc on peut dire que les valeurs de l'INDVI saisonnières en une relation avec le cycle végétatif de l'olivier. L'analyse des données saisonnières de l'Indice de Végétation par Différence

Tableau 4.5 – Corrélation du NDVI au stade de floraison

Mesures	Green Seeker (11 Mai)	Sentinel-2 (Mai)
Green Seeker (11 Mai)	0.86452	–
Sentinel-2 (Mai)	0.072511	–

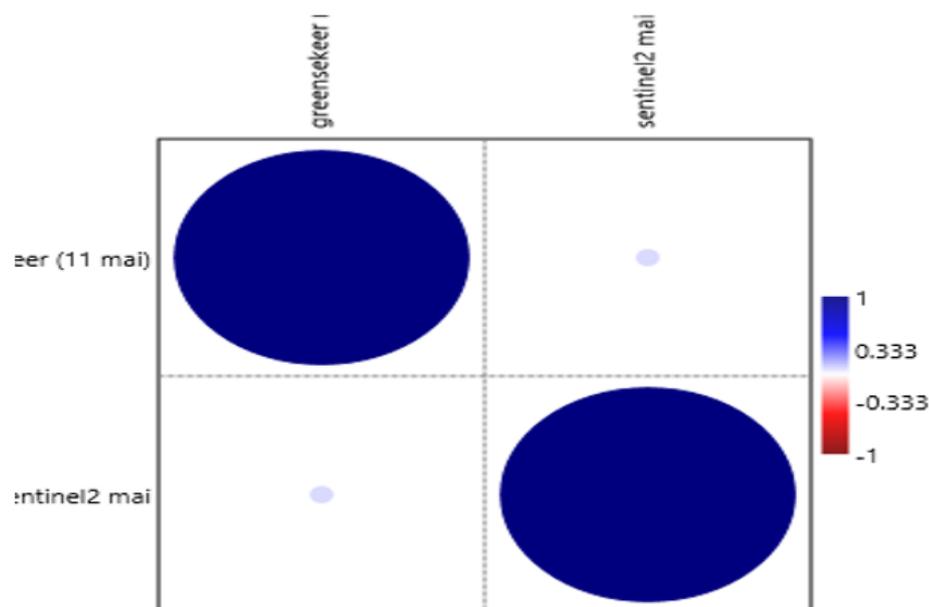


FIGURE 4.17 – Corrélation de l'NDVI au stade de florissant.

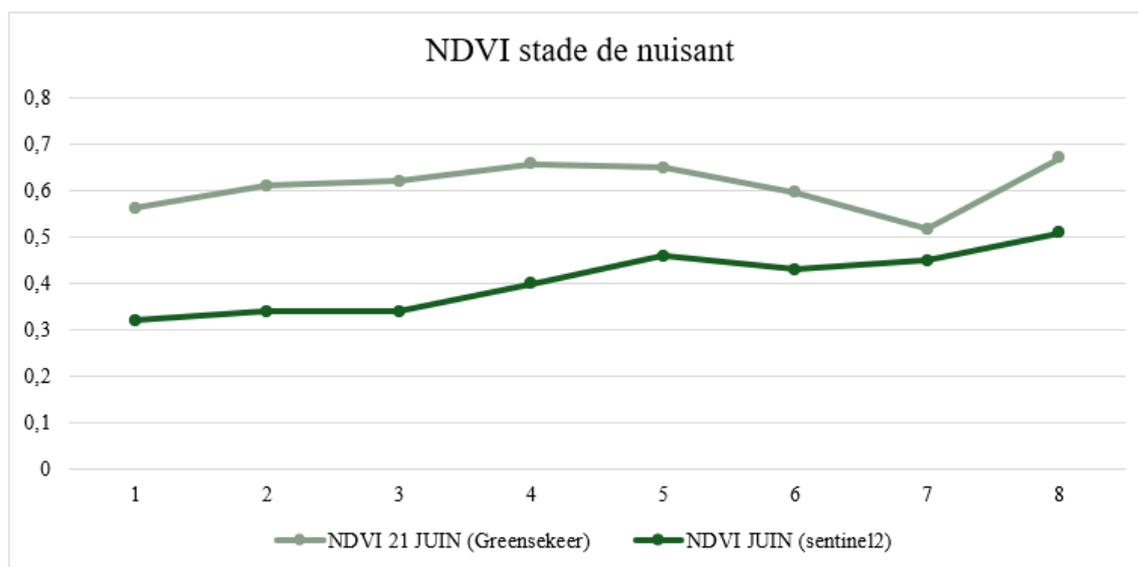


FIGURE 4.18 – Diagramme de Walter et Lieth des différentes stations.

Normalisée (NDVI) a permis de mettre en évidence une variation notable des valeurs de cet indice en fonction des saisons, reflétant étroitement le cycle végétatif de l'olivier. En effet, les valeurs du NDVI

Tableau 4.6 – Corrélation du NDVI au stade de nuisibilité

Mesures	Green Seeker (21 Juin)	Sentinel-2 (Juin)
Green Seeker (21 Juin)	0.43543	-
Sentinel-2 (Juin)	0.32284	-

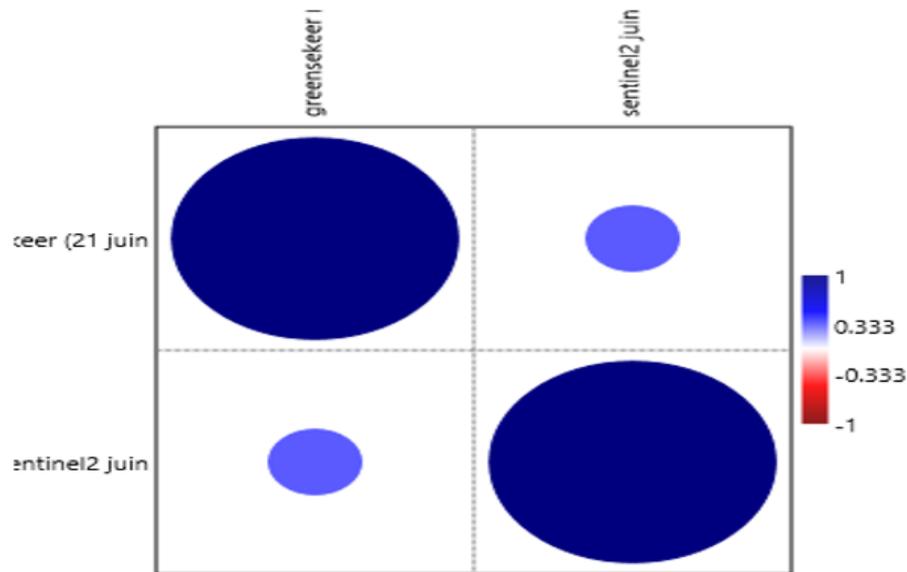


FIGURE 4.19 – Corrélation de l'NDVI stade de nuisant.

augmentent progressivement du printemps à l'automne, marquant une activité végétative accrue, avant de diminuer de manière significative en hiver, période de repos physiologique de l'arbre.

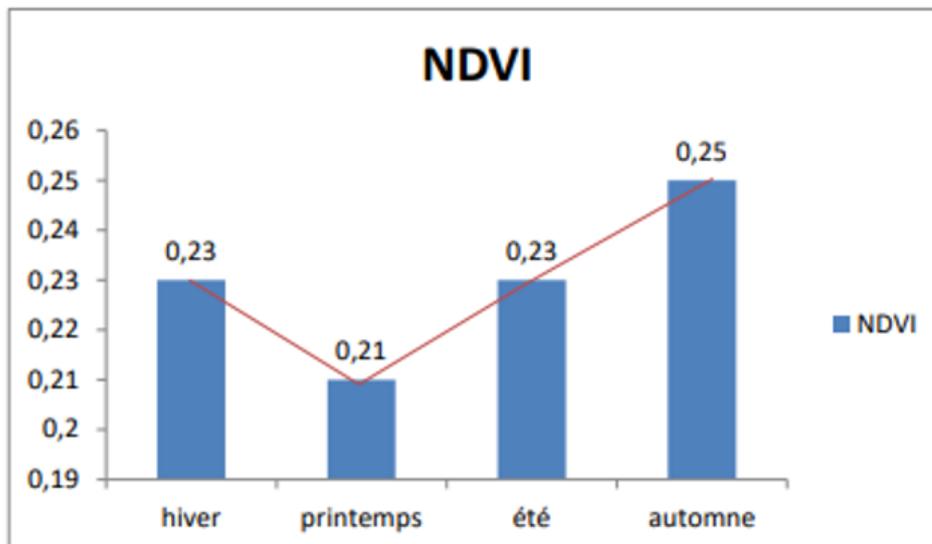


FIGURE 4.20 – Evolution saisonnière de l'olivier en fonction de NDVI Au cours de son cycle annuel de développement.

Au printemps, la reprise de l'activité photosynthétique, associée à la floraison et au développement des jeunes pousses, se traduit par une hausse notable des valeurs de l'NDVI. Pendant l'été, malgré les conditions parfois contraignantes (notamment en cette zones semi-arides), les valeurs de l'indice se maintiennent à des niveaux élevés, reflétant la résilience de l'olivier et son adaptation à ces environnements (voir la Figure 4.20). L'automne, période marquée par la maturation des fruits, voit également des valeurs relativement élevées de l'NDVI, bien que celles-ci puissent légèrement diminuer en fin de saison avec la préparation de l'arbre à la dormance hivernale. En revanche, l'hiver est caractérisé par une régression importante des valeurs du NDVI, correspondant au ralentissement de l'activité biologique de l'olivier, typique de son cycle végétatif.

4.2.5.7 Synthèse sur sur le cas d'étude

L'analyse du cas d'étude constitue une étape cruciale pour la maîtrise et l'analyse des variables nécessaires à la conception de notre framework. Cette étude permet de mieux comprendre les interactions entre différents facteurs influençant les performances du système agricole. Parmi ces variables, nous identifions les paramètres climatiques (température, précipitations, humidité, etc.), les caractéristiques du sol (texture, structure, teneur en matière organique, capacité de rétention d'eau, etc.), ainsi que les indices de réflectance obtenus à partir de l'imagerie multispectrale et hyperspectrale. Ces indices, tels que le NDVI (Normalized Difference Vegetation Index), offrent des informations essentielles sur l'état de santé et la vigueur des cultures. En complément, les paramètres morphologiques des plantes (hauteur, biomasse, densité foliaire, etc.) sont intégrés pour affiner les modèles et améliorer la précision des prédictions. L'intégration cohérente de ces différentes variables au sein de notre framework vise à fournir une approche systématique pour optimiser les pratiques agricoles, tout en tenant compte des particularités du cas d'étude.

Après avoir présenté le détail de notre cas d'étude, nous nous concentrons sur l'analyse des variables clés nécessaires à l'instanciation de notre framework. Ces variables incluent le climat, les caractéristiques du sol, les indices de réflectance et les paramètres morphologiques des cultures. Notre objectif est d'instancier le framework proposé dans le chapitre précédent sur ce cas d'étude, appliqué spécifiquement.

4.3 Mécanisme d'instanciation de AFFAC sur un cas d'étude

Nous rappelons que le processus de conception de la fonction de coût de production ($\mathcal{P}rod$) proposé est guidé par un ensemble de lignes directrices pour la conception de cette fonction. Typiquement, ce processus se compose des phases suivantes, comme illustré dans la Figure Figure 4.21 :

1. Expression du manifeste,
2. Identification des dimensions,
3. Identification des paramètres,
4. Construction du modèle ($\mathcal{P}rod$), et
5. Test du $\mathcal{P}rod$.

Étant donné que les tests continus sont une pratique importante de DevOps que nous avons appliquée à la conception et au développement de la fonction $\mathcal{P}rod$, nous fournissons une vérification précoce de tous

les éléments de la fonction (par exemple, les paramètres, les dimensions, les formules de composition des coûts et les performances) à toutes les étapes du cycle de vie du développement de la fonction *Prod*. Les modifications et améliorations apportées par un concepteur à chaque étape de la conception seront probablement testées avant de passer à l'étape suivante. Examinons maintenant en détail les étapes du processus de conception et les tests de chacune d'elles.

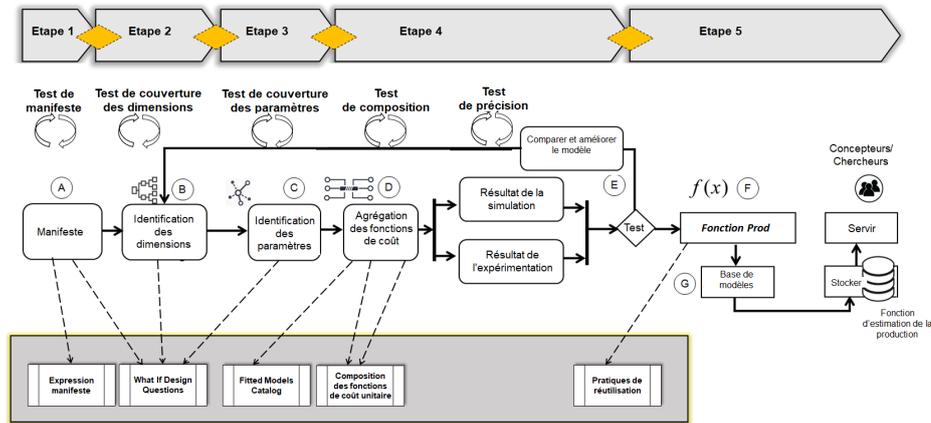


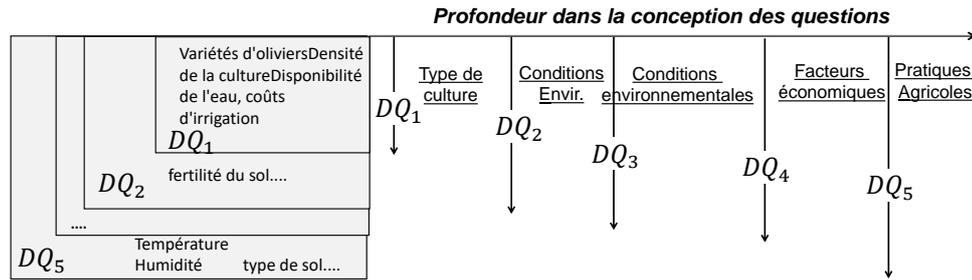
FIGURE 4.21 – Diagramme de flux du processus de conception et de test de fonction de coût *Prod*.

4.3.0.1 Expression du manifeste

Le manifeste représente les spécifications du concepteur qui se concentrent sur deux points principaux, à savoir le Contexte de la fonction *Prod* (par exemple, Base de données, Requête, Architecture de déploiement) et les attributs QoS (par exemple, performance et énergie). À cette première étape, le concepteur est assisté pour exprimer explicitement le manifeste en utilisant la ligne directrice de conception d'expression du manifeste, qui contient un ensemble de questions de conception de la *Prod*. Cette assistance vise à combler le fossé entre la conception de la *Prod* de haut niveau et la plateforme d'exécution de la base de données. Notre approche illustre les niveaux de hiérarchisation des *What If Design Questions* en fonction des dimensions de la *Prod*. Le niveau de détail des questions de conception dépend de l'expérience du concepteur en matière de conception. Le test du manifeste est basé sur notre langage de conception pour contrôler l'expression du manifeste du concepteur grâce à la variabilité fournie par notre espace de variabilité des paramètres pour instancier le manifeste du concepteur. Par exemple, si l'agriculteur sélectionne *l'irrigation goutte-à-goutte* comme méthode d'irrigation, un *réservoir d'eau* est *nécessaire* comme source principale d'approvisionnement en eau. S'il opte pour l'utilisation de *drones agricoles* pour surveiller les cultures, *l'utilisation de capteurs au sol sera réduite* comme méthode complémentaire de collecte de données, etc.

Exemple 1. Exemple de Questions de Conception *Prod* dans la Production Agricole d'Oliviers

Dans le contexte de la production agricole d'olives, un exemple de description du contexte *Prod* (Production Agricole) est donné à travers l'évaluation de : ${}^{\text{Prod}}DQ$, les questions de conception $DQ_1, DQ_2, DQ_3 \dots DQ_m$. Ces questions portent sur les données climatiques, les caractéristiques du sol et les processus de stockage. Le résultat de l'évaluation d'une règle peut être : \surd : vrai, \boxtimes : faux, ou \square : indéfini (ce qui signifie qu'il peut être vrai ou faux).

FIGURE 4.22 – Niveaux de hiérarchisation des *What If Design Questions*

$\mathcal{D}Q_1$	Les données sur la température moyenne des régions oléicoles sont-elles disponibles ?	✓
$\mathcal{D}Q_2$	Les caractéristiques du sol (pH, type de sol) sont-elles enregistrées pour chaque parcelle ?	□
$\mathcal{D}Q_3$	Les précipitations annuelles sont-elles utilisées pour prévoir les rendements ?	☒
$\mathcal{D}Q_4$	Les données historiques de production sont-elles accessibles pour une analyse comparative ?	✓

4.3.0.2 Identification des dimensions

Sur la base du manifeste du concepteur, notre cadre propose un certain nombre de dimensions de la *Prod*, telles que “base de données”, “requête” et “périphérique de stockage”, comme étant obligatoires ou optionnelles. La sélection d’une dimension obligatoire/optionnelle est guidée par notre espace de variabilité des paramètres en fonction de ses contraintes d’intégrité *require/exclude*.

À cette étape, notre cadre teste la couverture des dimensions d’entrée liées à la fonction *Prod* en cours de conception, y compris les combinaisons de matériel, de logiciel, de données, etc. Ces dimensions contiennent les paramètres qui ont un impact sur la métrique de performance donnée m_i (par exemple, le temps de réponse, la consommation d’énergie).

Par exemple, les dimensions “Agricultural factors” et “Environmental conditions” sont obligatoires pour estimer le coût ou le rendement de la production. Cependant, la dimension “Technological metrics” peut être optionnelle si le concepteur considère comme valide la *question de conception* “*what if*” suivante : “efficacité de l’irrigation ?”. Cette question implique un scénario hypothétique pouvant être évalué sans métriques technologiques. Le concepteur peut alors se baser sur des paramètres agronomiques jugés suffisants.

4.3.0.3 Identification des paramètres

La conception d’une *Prod* des cultures agricoles repose sur un ensemble de paramètres appelés espace des paramètres, parmi lesquels il est difficile de choisir les bons paramètres à considérer comme entrées de la fonction *Prod*.

À cette étape, nous fournissons au concepteur l’espace des paramètres de conception de la fonction *Prod* de manière structurée, afin de faciliter l’extraction des paramètres.

Grâce à une analyse automatique des *What If Design Questions* $\mathcal{D}Q^{Prod} = \mathcal{D}Q^{undef} \cup \mathcal{D}Q^{true} \cup \mathcal{D}Q^{false}$, les paramètres importants et non importants peuvent être identifiés en fonction des dimensions d’entrée. En conséquence, tous les paramètres non inclus dans les dimensions sélectionnées seront

supprimés de l'espace des paramètres de conception de la fonction *Prod*.

De plus, l'identification des paramètres est effectuée à l'aide des contraintes considérées dans notre espace de paramètres. Plusieurs types de contraintes sont inclus dans notre \mathcal{FM} : (i) *require/exclude* pour indiquer si la fonctionnalité est présente ou non dans le produit (c'est-à-dire dans le contexte de la fonction *Prod*), (ii) *obligatoire/optionnel*, et (iii) *Or/alternatif* pour indiquer les relations entre une fonctionnalité parente et une fonctionnalité enfant.

4.3.0.4 Modélisation des coûts

L'étape de modélisation des coûts consiste à traduire tous les paramètres en unités de coût et à établir un modèle qui exprime les interrelations entre les différents paramètres à l'aide d'équations mathématiques. À ce stade, nous proposons une série de formules pour calculer les coûts de la fonction *Prod* en fonction des modèles physiques associés.

Notre approche de conception de la fonction *Prod* permet une évaluation précise des coûts en calculant l'intégralité des coûts de performance sur la base de chaque paramètre de conception individuel, ce qui est réalisé grâce à notre processus de modélisation des coûts.

Les formules de modélisation des coûts de base que nous utilisons couvrent trois aspects principaux, à savoir : l'ensemble de l'environnement de la culture en termes de climat, sol et topographie (c'est-à-dire la température, l'humidité, le type de sol, topographie locale).

4.3.0.5 Stockage des *Prod* dans une base de données

Une fois que la *Prod* est formée et testée, elle est stockée dans une base de données. Ce dépôt peut être consulté à tout moment pour la récupération et l'analyse de la fonction *Prod*, ce qui permet aux concepteurs de tester de nouveaux scénarios ou d'ajuster les paramètres existants.

4.4 Instanciation de notre Framework AFFAC

À travers cette étude de cas, nous espérons fournir des éclairages précieux sur les avantages et les défis de la conception d'une nouvelle fonction de production pour l'optimisation de la prévision. Nous décrivons d'abord les principales fonctionnalités de l'outil de support, puis nous présentons un scénario d'application détaillant les différentes phases de conception du modèle de prévision.

Pour mettre en avant notre framework AFFAC et démontrer son utilité, nous avons développé un mécanisme d'installation permettant aux concepteurs de créer, réutiliser ou adapter des Prods. Les fonctionnalités de l'outil sont conçues sur la base des exigences suivantes des utilisateurs :

- Fournir tous les paramètres utilisés dans un modèle de prévision particulier, avec la possibilité de visualiser les modèles.
- Vérifier l’analyse de sensibilité des paramètres du modèle.
- Fournir un catalogue de modèles ajustés pouvant être appliqués au modèle de prévision.
- Tester les modèles affinés et visualiser l’historique de toutes les modifications apportées au modèle de prévision, ainsi que l’erreur de prédiction.

De plus, nous avons développé un mécanisme d’installation de AFFAC qui fournit un espace de travail de conception divisé en quatre parties : (1) Contexte et Question de Conception, (2) Analyse des Paramètres, (3) Construction du Modèle et (4) Résultats-Rapports et Exploration des Coûts. Nous détaillons dans la section suivante un scénario d’application qui présente toutes les étapes de la conception du modèle de prévision réalisées à travers l’outil de conception.

4.4.1 Aperçu du flux de travail de conception du modèle de prévision

La conception du modèle agricole *Prod*, guidée par des recommandations en modélisation prédictive, suit un processus structuré en cinq phases : formulation du besoin, définition des dimensions, sélection des paramètres, construction du modèle, puis test des modèles de prévision (voir Figure Figure 4.21). Chaque phase intègre une vérification systématique des éléments du modèle (paramètres, dimensions, formules de coût et de performance). Les ajustements effectués sont testés avant de passer à l’étape suivante du développement des *Prods*. Les sections suivantes détaillent ce processus ainsi que les tests associés.

4.4.2 Scénario d’application

Afin de développer le modèle de prévision à utiliser pour la prédiction de la production d’olives, nous avons choisi de concevoir un modèle de fonction de prévision. Toutes les étapes décrites dans la section sur le flux de travail de conception du modèle de prévision sont détaillées en fonction du scénario d’application considéré.

Le processus de conception du modèle de prédiction des rendements repose sur plusieurs paramètres liés au climat, à la culture et à la géolocalisation. Une description détaillée du contexte pour le modèle de prévision de la production d’olives est fournie. La création d’une fonction de coût dans ce contexte nécessite une compréhension approfondie des différents paramètres et la construction d’un modèle qui capte leurs interactions afin d’obtenir les résultats souhaités.

4.4.2.1 Manifest Expression

Dans cette étape, nous définissons les exigences du concepteur et les contraintes de l’application de base de données en utilisant la ligne directrice de l’expression manifeste (①) : (1) Le type de culture est *variétés d’olives*, les attributs de qualité de service (QoS) incluent la *densité de culture*, et les contraintes sont la *disponibilité de l’eau* et les *coûts d’irrigation*. (2) Les facteurs agricoles incluent la *fertilité du sol* et la *densité de culture*. (3) Les conditions environnementales comprennent la *température* et l’*humidité*, tandis

Tableau 4.7 – Vue contextuelle pour le développement de la fonction de production d'olives

Problématique	Type de culture	<i>Variétés d'olives</i>
	Attributs de QoS	<i>Densité de plantation</i>
	Contraintes	<i>Disponibilité en eau, Coûts d'irrigation</i>
Paramètres du système	Facteurs agricoles	<i>Fertilité du sol</i>
		<i>Densité de plantation</i>
	Conditions environnementales	<i>Technique de taille</i>
		<i>Température</i>
		<i>Humidité</i>
Indicateurs technologiques	<i>Type de sol</i>	
	<i>Efficacité d'irrigation</i>	
		<i>Consommation énergétique des machines</i>
<i>Prod</i>	Prévision des coûts de production des olives	

que les métriques technologiques incluent le *type de sol*, l'*efficacité de l'irrigation* et la *consommation d'énergie des machines*.

So, here's an example of a user manifest.

"Un utilisateur souhaite construire une fonction de coût de production (Prod) pour estimer la production d'olives, en tenant compte des coûts d'irrigation, de fertilisation et des équipements utilisés, avec des informations détaillées sur l'environnement de production agricole".

4.4.2.2 Identification des Dimensions et Paramètres

La ligne directrice de l'explorateur aide les concepteurs à naviguer, éditer et sélectionner les éléments du design de la fonction de coût. Notre cadre crée un espace de paramètres, permettant d'estimer la production, qui est affiné en fonction de l'expression manifeste et des questions de conception (②). Les paramètres proposés sont catégorisés en fonction des dimensions du Prod comme suit :

1. **Paramètres des données agricoles** : Nombre d'objets à stocker et relations entre les facteurs environnementaux et agricoles (irrigation, fertilisation, etc.).
2. **Paramètres des opérations agricoles** : Taille des parcelles, quantité d'eau utilisée, et outils employés dans les différentes étapes de la production.
3. **Paramètres des ressources matérielles** : Coûts d'infrastructure, d'exploitation et de maintenance des équipements agricoles.

Ces paramètres permettent d'estimer les coûts liés à la production agricole, comme pour la culture de l'olive.

Pour guider l'extraction des paramètres avec précision et garantir l'identification correcte des paramètres pertinents dans le contexte actuel de la fonction de production, nous définissons l'espace de variabilité comme une description de l'architecture du système et de ses paramètres. Nous exploitons les connaissances du domaine, qui incluent un ensemble de fonctionnalités et un ensemble de contraintes d'intégrité, couramment utilisées pour orienter le processus de conception. Cette approche facilite

Tableau 4.8 – Vue d’ensemble contextuelle du développement de la fonction de production d’olive

Problème	Type de culture	Variétés d’Olives
	Attributs QoS	densité de culture
	Contraintes	Disponibilité en eau, Coûts d’irrigation
Paramètres du système	Facteurs agricoles	fertilité du sol densité de culture technique de taille
	Conditions environnementales	température humidité type de sol
	Métriques technologiques	efficacité de l’irrigation consommation d’énergie des machines
<i>Prod</i>	Prévision des coûts de production des récoltes d’olives	

l’identification correcte des types de paramètres et la dissimulation des paramètres non pertinents. Grâce à la variabilité offerte par l’espace de variabilité, il devient possible d’instancier les paramètres dans le contexte de *Prod*.

4.4.2.3 Construction du Modèle Prévisionnel

La construction du modèle prévisionnel repose sur l’identification des fonctions de coût unitaire associées aux paramètres d’entrée liés à la production d’olive (③). À cette étape, le concepteur est assisté par un catalogue de modèles ajustés pour sélectionner la signature du modèle de prévision qui correspond au manifeste de l’utilisateur, basé sur les paramètres de culture identifiés dans le tableau.

Le degré de granularité du modèle prévisionnel dépend de l’objectif de la prévision : coût de production, rendement de culture, ou gestion des ressources.

- Un modèle de production \mathcal{P} est un ensemble de l prévisions : $\mathcal{P} = \{P_1, P_2, \dots, P_l\}$,
- Chaque prévision P_i ($1 \leq i \leq l$) est un ensemble de paramètres influençant la production : $\mathcal{P} = \{f_1, f_2, \dots, f_m\}$, et
- Le paramètre f_i est une variable liée à l’agriculture, à l’environnement ou à la technologie (par exemple, la fertilité du sol, les conditions climatiques ou l’efficacité d’irrigation).

Le concepteur commence par combiner les paramètres de base pour décrire la fonction de prévision de la production d’olive. Par exemple, dans le contexte de la culture d’olive, pour une prévision P_i influencée par la température et l’humidité, le modèle de prévision peut être calculé en utilisant les paramètres de température et d’humidité collectés.

Ensuite, les prévisions sont combinées pour estimer les coûts globaux de production ou les rendements. Finalement, la prévision globale est calculée en agrégeant les résultats des différentes prévisions de production. L’agrégation des prévisions de production (*Prod*) qui représente la somme des prévisions des différents paramètres de production. Nous notons que *Prod* est composé de trois parties à chaque niveau de prévision (agriculture, environnement et technologie), à savoir : $Prod_1$, $Prod_2$ et $Prod_3$, tels que :

- $Prod_1$ est la prévision du rendement en fonction des facteurs agricoles (ex. : densité de culture,

fertilité du sol),

- $Prod_2$ est la prévision des coûts liés aux conditions environnementales (ex. : température, humidité, disponibilité en eau),
- $Prod_3$ est la prévision des coûts associés aux technologies utilisées (ex. : irrigation, consommation énergétique des machines).

4.4.2.4 Calibration des poids à l'aide de l'apprentissage automatique

La calibration des poids dans les fonctions de prévision des cultures agricoles à l'aide de l'apprentissage automatique consiste à ajuster les paramètres du modèle pour améliorer la précision des prédictions. En utilisant des données telles que les conditions climatiques, du sol, et les historiques de production, l'apprentissage automatique permet d'ajuster les poids associés à chaque facteur influent, optimisant ainsi la fonction de prévision. Ce processus, basé sur des techniques comme la régression ou les réseaux neuronaux, repose sur la minimisation des erreurs de prévision via des méthodes telles que la validation croisée. Cela permet de rendre les modèles plus fiables et adaptés aux spécificités locales des cultures agricoles.

4.4.2.5 Test et validation

Pour évaluer la qualité du $Prod$ obtenu, nous avons mené plusieurs expériences concernant l'*Exactitude*, la *Robustesse*, et les caractéristiques *non-disruptives*. Étant donné que la caractéristique de réponse rapide est liée à la complexité du $Prod$.

4.4.3 Généricité de notre framework AFFAC

Sur la base des conclusions de cette revue de la littérature, pour créer un modèle de prévision (MP) pour la production agricole, les chercheurs dans le domaine des sciences agricoles prennent en compte plusieurs aspects (*type de culture, qualité du sol, variables climatiques, méthodes d'irrigation, périodes de plantation et de récolte, pratiques agricoles, gestion des ravageurs et attributs de la QoS*). À notre connaissance, les modèles de prévision existants ne prennent pas en compte tous ces facteurs. Nous soutenons que, pour développer un MP pertinent pour la production agricole, tous ces facteurs doivent être intégrés.

Il existe une vaste quantité de travaux explorant le développement et l'adaptation des modèles de prévision pour les systèmes agricoles. Cependant, les travaux précédents ne se sont pas concentrés sur la facilitation de la réutilisation de cette expérience pour des applications pratiques par les agriculteurs et les chercheurs. Nous avons identifié un manque de cadres dédiés pour définir explicitement les primitives de conception fondamentales, permettant ainsi une expression facile de la fonction de coût globale de manière composée.

De plus, les recherches antérieures n'ont pas fourni d'outils de support pour aider les agriculteurs et chercheurs à réduire le temps de développement nécessaire pour créer leurs propres modèles de prévision. Globalement, bien que des travaux substantiels aient été réalisés pour adapter les modèles de prévision agricoles, aucun d'eux ne s'est concentré sur la fourniture d'outils permettant d'enrichir, d'étendre

et d'améliorer ces modèles de prévision à travers une documentation structurée et des processus de raffinement.

4.5 Conclusion

Dans ce chapitre, nous avons présenté en détail une étude de cas qui illustre l'application pratique du cadre AFFAC (Advanced Forecasting Framework for Agricultural Crops) dans le domaine de la prévision de la production agricole. Cette étude a permis d'instancier le cadre AFFAC en tenant compte des spécificités de la culture de l'olive, un pilier majeur de l'agriculture méditerranéenne. À travers ce cas d'étude, nous avons démontré comment AFFAC facilite la collecte et la gestion des données, l'intégration des connaissances d'experts et la sélection des variables d'entrée pour améliorer les prévisions. En appliquant ce cadre, nous avons non seulement affiné les prévisions de production, mais aussi fourni un outil décisionnel précieux pour la gestion des oliveraies, ouvrant la voie à une meilleure prise de décision et à une gestion optimisée des ressources agricoles.

Évaluation de notre Framework AFFAC



« *The true method of knowledge is experiment.* »

— William Blake

Sommaire

5.1	Introduction	128
5.2	Évaluation de notre Cadre	128
5.2.1	Zone d'étude	128
5.2.2	Ensemble de Données	128
5.2.3	Analyse des Paramètres de <i>Prod</i>	131
5.2.4	Évaluation de nos modèles de prévision	135
5.3	Implications de la recherche	142
5.3.1	Implications industrielles	142
5.4	Outil prototype d'assistance aux agriculteurs sous forme d'application mobile	143
5.4.1	Présentation de notre outil prototype	143
5.4.2	Architecture de notre application	144
5.4.3	Processus de conception d'une application mobile	145
5.4.4	Présentation de notre outil d'assistance	146
5.5	Discussion	153
5.6	Discussion	156
5.7	Conclusion	158

5.1 Introduction

Après avoir décrit notre approche dans le chapitre précédent, nous aborderons dans ce chapitre une étude de cas portant sur les oliécultures, qui nécessite un contrôle et un suivi précis. Des algorithmes avancés basés sur les données doivent être utilisés pour l'ajustement des paramètres afin d'améliorer la qualité du produit, d'aligner la production avec la demande du marché et de réaliser des analyses prédictives. Ces analyses recommandent des améliorations et préviennent les anomalies de production, garantissant ainsi des résultats optimaux dans le rendement et la production d'olives.

Dans ce chapitre, nous allons montrer l'utilité de notre approche en présentant un scénario d'utilisation global du Framework AFFAC. Nous mettrons également en évidence les implémentations techniques et nous détaillerons ces éléments étape par étape. Pour faciliter la compréhension, nous utiliserons des captures d'écran afin de présenter les principaux résultats. Nous commencerons par présenter les outils et technologies utilisés pour le développement, puis nous aborderons l'architecture technique de la solution, ainsi que l'architecture du code.

5.2 Évaluation de notre Cadre

Pour évaluer AFFAC, nous avons mené plusieurs expériences. Avant de les décrire en détail, nous commencerons par discuter des objectifs de nos expériences. Ensuite, nous présenterons la machine expérimentale utilisée pour calculer le temps de réponse ainsi que l'ensemble de données exploité.

5.2.1 Zone d'étude

La zone d'étude couvre une superficie de 382 000 km², soit environ 20% du territoire national, où se concentrent presque toute la population et les activités. Il s'agit d'un ensemble complexe comprenant une succession de chaînes montagneuses, de zones côtières et sous-littorales, ainsi que de plaines. Cet ensemble est constitué de deux sous-ensembles : le Tell et les Hautes Plaines Steppiques, répartis sur 36 communes.

L'oliveraie en Algérie s'étend sur une superficie de 431 506 hectares, regroupant 60 632 901 oliviers[119]. Au cours des 20 dernières années, les surfaces cultivées en oliviers ont connu une croissance significative, passant de 168 080 hectares en 2000 à 294 200 hectares en 2010, pour atteindre 431 506 hectares en 2019. Cela représente une augmentation de 75%.

Par ailleurs, 80% des oliveraies sont situées dans des zones montagneuses, sur des terres accidentées et marginales à faible fertilité. Les 20% restants se trouvent dans les plaines occidentales du pays.

5.2.2 Ensemble de Données

Nous avons obtenu la base de données sur la production et le rendement des olives auprès des services du Ministère de l'Agriculture et du Développement Rural, répartis sur l'ensemble du territoire algérien, pour les années 2010 à 2019. Ces données sont disponibles sur GitHub⁴. Après analyse de cette

4. <https://github.com/ouared14/Prod-CULT>

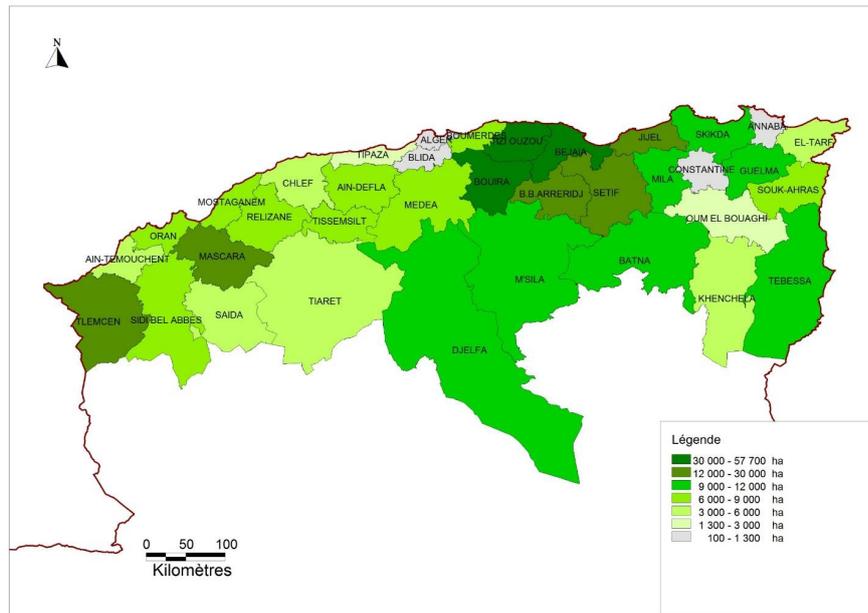


FIGURE 5.1 – Visualisation des Zones Oléicoles avec un Gradient de Couleur Indiquant la Répartition par Région dans la Zone d’Étude.

base de données, nous avons sélectionné 34 régions situées dans le nord de l’Algérie.

La deuxième étape a consisté à acquérir des données climatiques (T_{min} , T_{max} , T_{moy} , $Précipitations$) issues des stations choisies, sur une base mensuelle, sur une période de 30 ans pour chaque station.

Dans le cadre de la prévision de la production d’olive, les stations électriques ont été stratégiquement équipées de capteurs sophistiqués pour mesurer l’humidité et la température. Ces capteurs fournissent des données cruciales permettant de surveiller les conditions environnementales influençant directement la croissance des oliviers et la qualité des olives.

En complément, nous avons intégré des données géographiques supplémentaires, extraites automatiquement à partir de Google Maps Engine⁵. Le processus d’extraction des données a été mis en œuvre grâce à un script qui collecte et compile les informations géographiques pertinentes.

Le Listing Liste 5.1 présente un extrait de code JavaScript utilisé pour l’intégration des données géographiques supplémentaires extraites automatiquement de Google Maps Engine. Ci-dessous se trouve un exemple de snippet de code JavaScript démontrant ce processus.

5. <https://earthengine.google.com/>

```

1 function initialize() {
2     var map = new google.maps.Map(document.getElementById('map'),{
3         center : { lat : -34.397, lng : 150.644 },
4         zoom : 8 }); }
5 function extractData(location) {
6     var geocoder = new google.maps.Geocoder();
7     geocoder.geocode({ 'address' : location },
8
9     function(results, status) {
10        if (status === 'OK') {
11            console.log(results[0].formatted_address);
12            console.log(results[0].geometry.location);
13        } else {
14            console.error('Geocode .... ' + status);
15        } }); }
16 var locations = ['Algiers, Algeria', 'Oran, Algeria'];
17 locations.forEach(function(location) {
18     extractData(location);});

```

Listing 5.1 – Extrait de code JavaScript pour l’intégration automatique de données géographiques supplémentaires extraites de Google Maps Engine.

Feature	Description	Type	Data Collection Method	Dimension
Region	The region where olives are cultivated	Categorical	Manual / Database	Geographical Data
Year	Year of harvest	Numerical	Manual / Database	Seasonal Effects
Month	Month of harvest	Categorical	Manual / Database	Seasonal Effects
Area_Planted	Planted area (hectares)	Numerical	Manual / Database	Agricultural Practices
Olive_Variety	Type of olives grown	Categorical	Manual / Database	Agricultural Practices
Irrigation_Method	Irrigation method used	Categorical	Manual / Survey	Agricultural Practices
Soil_Type	Type of soil	Categorical	Manual / Sensor	Soil Conditions
Fertilizer_Used	Amount of fertilizer used (kg/ha)	Numerical	Manual / Database	Agricultural Practices
Pesticides_Used	Amount of pesticides used (kg/ha)	Numerical	Manual / Database	Agricultural Practices
Tmin	Average minimum temperature (°C)	Numerical	Automatic / Sensor	Climate Data
Tmax	Average maximum temperature (°C)	Numerical	Automatic / Sensor	Climate Data
Tmean	Average temperature (°C)	Numerical	Automatic / Sensor	Climate Data
Precipitation	Total monthly precipitation (mm)	Numerical	Automatic / Sensor	Climate Data
Sunshine_Hours	Monthly sunshine hours	Numerical	Automatic / Sensor	Climate Data
Labor_Cost	Labor cost per hectare	Numerical	Manual / Database	Economic Factors
Market_Price	Selling price per kg (local and international)	Numerical	Manual / Market Survey	Economic Factors
Target (Production)	Olive production (tonnes)	Numerical	Calculated	Economic Factors

La Table Section 5.2.2 présente l'ensemble de données sur la production d'olives, en mettant en évidence la description de chaque caractéristique, son type de données et sa méthode de collecte. Cet ensemble de données couvre divers aspects de la culture de l'olivier, notamment la région géographique, la surface plantée, la variété d'olives, les conditions climatiques et les facteurs économiques. Il précise également la méthode de collecte des données, indiquant si celles-ci ont été obtenues manuellement, automatiquement via des capteurs, ou issues de bases de données et d'enquêtes.

5.2.3 Analyse des Paramètres de $\mathcal{P}rod$

Pour identifier les paramètres clés, nous avons utilisé une analyse de sensibilité afin d'évaluer si les résultats sont influencés par des paramètres d'entrée significatifs. Cela implique d'examiner les valeurs des domaines des paramètres, y compris leur plage de variation et leur distribution de probabilités.

Cette section évalue deux caractéristiques essentielles de *AFFAC* : l'analyse de sensibilité des paramètres de la base de données et les graphiques d'effets pour la visualisation. Nous avons identifié les paramètres importants et non importants à l'aide des trois techniques suivantes : **(i)** Analyse de Sensibilité, **(ii)** Analyse de la Distribution Statistique des Données, et **(iii)** Analyse de Corrélation.

5.2.3.1 Analyse de Sensibilité

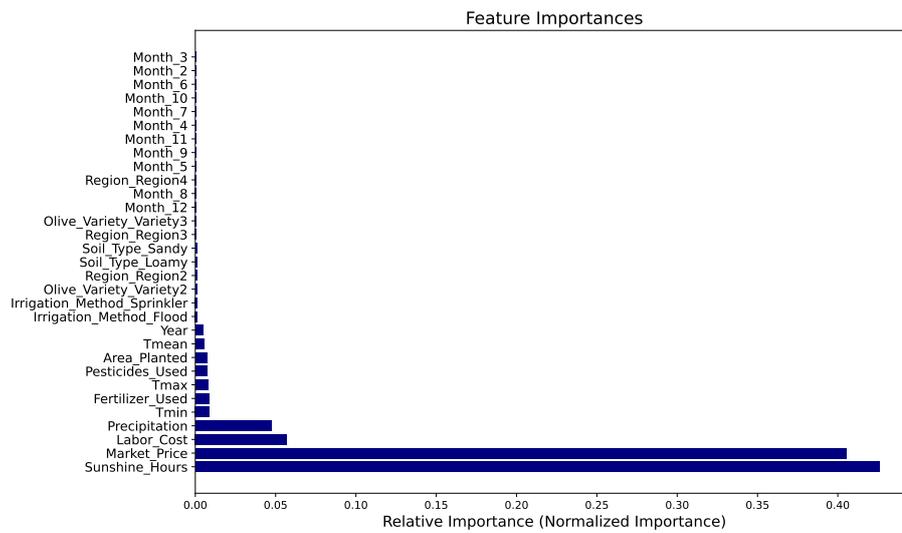
Pour identifier les paramètres les plus significatifs, nous avons utilisé des techniques de sélection de caractéristiques, validées par des avis d'experts. Nous avons veillé à ce que l'importance de chaque prédicteur soit déterminée de manière indépendante au modèle, en utilisant une approche par filtre. Le résultat est une liste classée des paramètres de $\mathcal{P}rod$ selon leur importance. La Figure 3.3 illustre l'importance de tous les paramètres ainsi que leur force prédictive cumulative.

En utilisant l'algorithme PCA, nous avons réduit la dimensionnalité de notre ensemble de données aux 17 caractéristiques les plus critiques. Ces dernières sont représentées en fonction de leur importance normalisée, dont la somme totale est égale à 1. Sur le côté droit de la même figure, nous affichons l'importance cumulative en fonction du nombre de caractéristiques. Une ligne verticale indique le seuil d'importance cumulative, fixé à 95% dans ce cas.

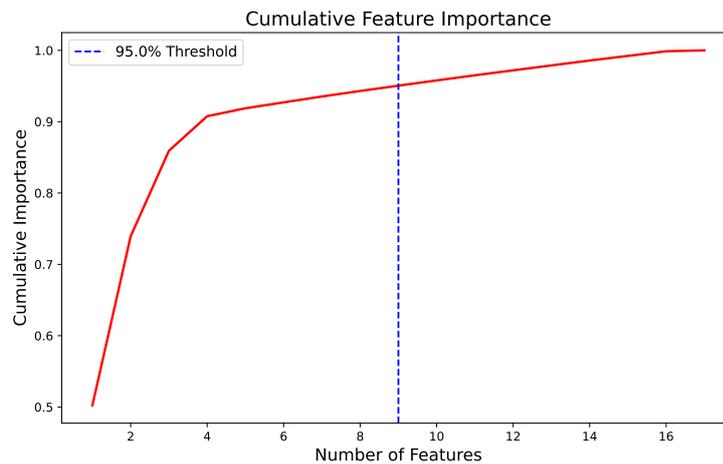
5.2.3.2 Analyse des distributions statistiques des données

La Figure 5.3 illustre le processus de vérification de la linéarité des variables. Dans cette figure, nous présentons un exemple comportant plusieurs paramètres qui tracent la forme de la distribution afin d'élucider la variance des données. Les paramètres incluent les **facteurs agricoles** tels que *la fertilité du sol* et *la densité des cultures* ; les **conditions environnementales** comme *la température*, *l'humidité*, et *le type de sol* ; ainsi que les **mesures technologiques** telles que *l'efficacité de l'irrigation* et *la consommation d'énergie des machines*. De plus, nous considérons *cxti*, représentant la région d'incertitude, qui englobe les conditions climatiques futures et la demande du marché.

Ces paramètres sont utilisés comme entrées pour notre modèle. Il est évident qu'ils ne prennent pas en compte les relations non linéaires entre les paramètres de $\mathcal{P}rod$ et le résultat.



(a) Importance des Caractéristiques



(b) Importance Cumulative des Caractéristiques

FIGURE 5.2 – Analyse de Sensibilité des Paramètres de $\mathcal{P}rod$

5.2.3.3 Analyse de la corrélation des paramètres

Pour chaque paramètre d'entrée, un boxplot distinct est fourni. Ces boxplots offrent des informations sur les valeurs d'impact des paramètres de $\mathcal{P}rod$ et illustrent comment les variations des paramètres influencent la performance. Dans un boxplot, les limites supérieure et inférieure de la boîte correspondent respectivement aux 25% et 75% percentiles, englobant ainsi les 50% centraux des valeurs observées, y compris la médiane.

Le choix des valeurs de domaine des paramètres pour l'analyse est crucial, car il affecte la dispersion des valeurs utilisées dans le modèle de coût. La Figure 5.4a montre la dispersion des données pour les paramètres influents, tels que *fertilité du sol*, *densité des cultures*, *température*, *humidité*, *type de sol*, *efficacité de l'irrigation* et *consommation énergétique des machines*. Initialement, il est évident que les médianes pour ces paramètres sont assez similaires, voire identiques, dans différents cas. Par exemple,

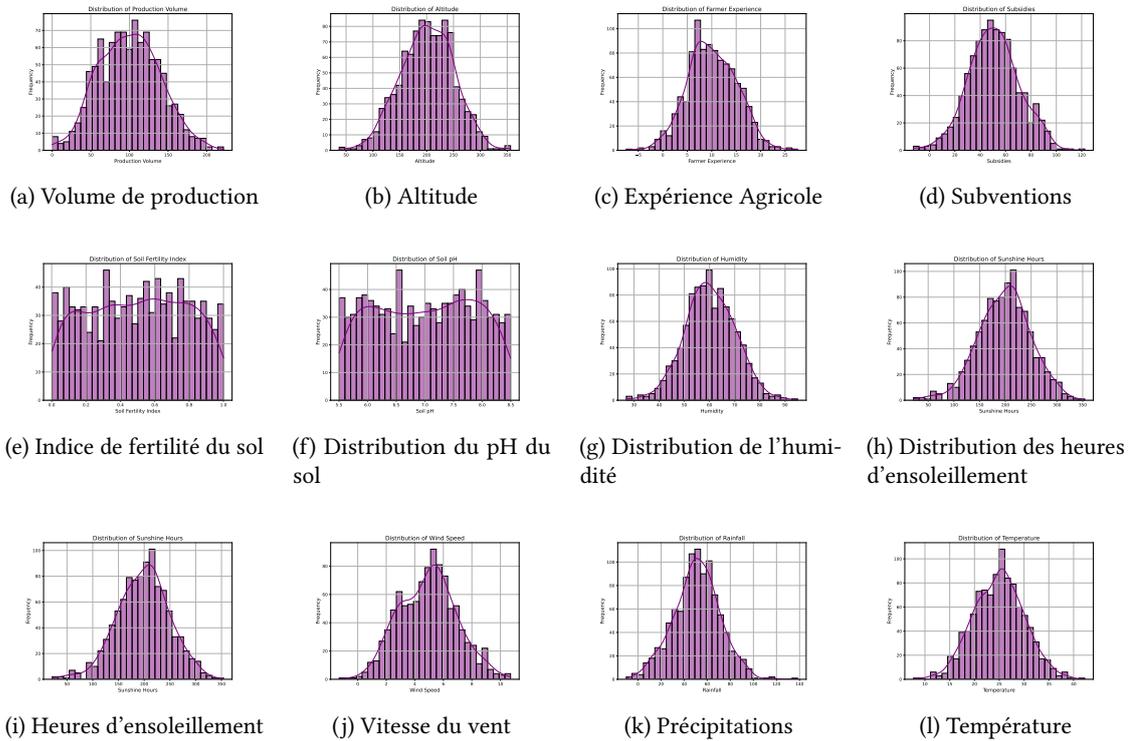
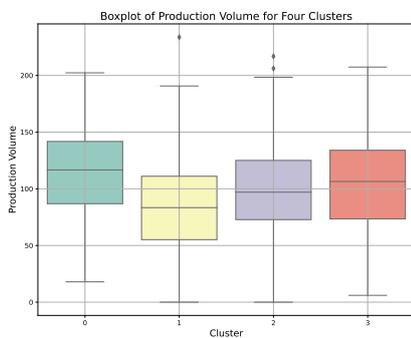


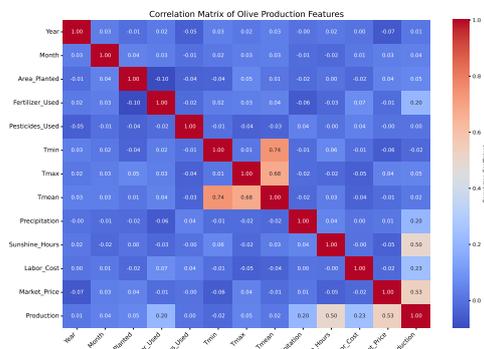
FIGURE 5.3 – Extrait des distributions d'échantillons des paramètres choisis

fertilité du sol et *densité des cultures* montrent des valeurs médianes similaires dans leurs graphiques respectifs. De plus, *efficacité de l'irrigation* a un impact significatif sur les résultats par rapport à d'autres facteurs.

Une matrice de corrélation, présentée dans la Figure 5.4b, résume les relations entre les variables en affichant les coefficients de corrélation. Chaque cellule de cette matrice indique la corrélation entre les caractéristiques et la sortie. Le graphique de corrélation visualise les corrélations positives et négatives, qui sont observées dans les blocs hors diagonale.



(a) Boxplot (Production d'Olives)



(b) Coefficients de Corrélation des Paramètres

FIGURE 5.4 – Vue partielle de l'analyse de corrélation pour les paramètres $\mathcal{P}rod$

5.2.3.4 Tests sur des données normales

Pour réaliser les tests, nous échantillons à partir d'une distribution normale et projetons les données en deux dimensions : nous nous plaçons dans le plan complexe où les observations suivent une distribution normale avec une moyenne et une variance données, tout en suivant une distribution uniforme sur un intervalle donné (voir Figure 5.5). Typiquement, la règle des 3-sigmas est utilisée pour déterminer la séparation entre les données normales et anormales. Si nous appliquons cette méthode ici, l'élimination des données les plus aberrantes devrait réduire, à chaque itération, le nombre d'éléments situés au-delà de la moyenne. Cependant, l'évolution du modèle n'est pas suffisamment rapide, et nous nous retrouvons rapidement sans aucune anomalie en utilisant cette méthode.

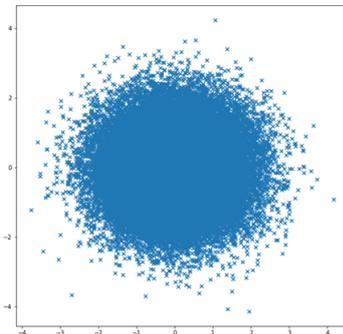


FIGURE 5.5 – Représentation de l'ensemble de données dans le plan complexe avec un grand nombre d'échantillons.

Différents phénomènes ont été observés dans les données, en fonction du capteur utilisé. Jusqu'à quatre capteurs ont été fournis selon la station observée : la température de la station, la température de la fondation, l'humidité de la station et l'humidité de la fondation. Ces capteurs ont enregistré des mesures à des fréquences d'échantillonnage variables, chacune générant une série temporelle univariée. La Figure 5.6 illustre les variations de température. Les mesures de température et d'humidité présentent parfois des pics réguliers similaires aux signaux observés au niveau de la station. Dans la Figure 5.6, nous présentons les variations d'amplitude.

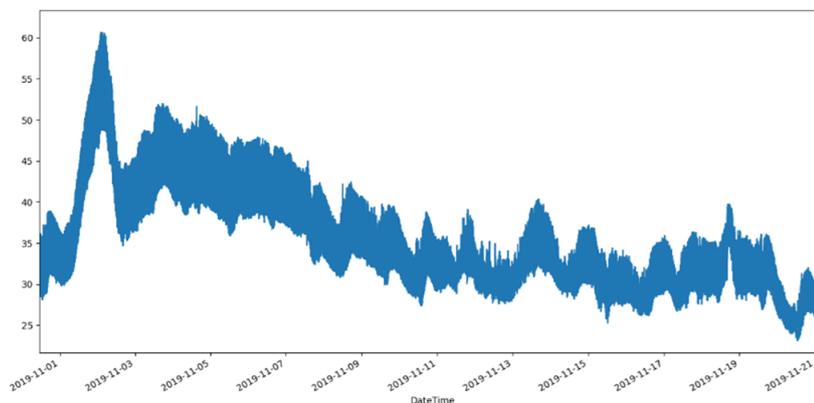


FIGURE 5.6 – Variations de la température.

Les valeurs manquantes sont un problème fréquent avec les données issues des capteurs et peuvent

être détectées comme des anomalies lorsqu'elles interrompent la continuité du signal. Une seule valeur manquante n'est pas toujours significative, mais cela peut devenir problématique si elle persiste pendant plusieurs heures ou jours. Les valeurs anormales de température peuvent compliquer davantage ce problème ; par exemple, dans la Figure 5.7, il est clair qu'une défaillance est survenue vers la fin du mois d'octobre. En raison des températures élevées durant cette période, cette défaillance n'aurait pas été détectée par des seuils fixes. En revanche, une défaillance beaucoup plus facile à détecter apparaît dans la figure suivante.

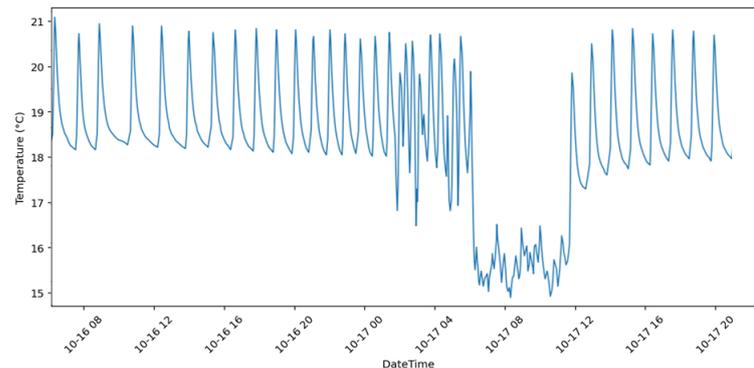


FIGURE 5.7 – Valeur de température anormale.

Résumé :

Dans AFFAC, l'analyse des paramètres de la fonction *Prod* facilite l'exploration des paramètres existants, permet l'ajout ou la modification de nouveaux paramètres et soutient leur validation. Cette fonctionnalité est particulièrement bénéfique pour les développeurs et les concepteurs. De plus, elle constitue une ressource importante pour ces professionnels. Comme indiqué précédemment, AFFAC propose diverses méthodes analytiques pour évaluer la pertinence des paramètres de la fonction *Prod* en fonction de leur influence sur le paramètre cible.

Comprendre l'ensemble des paramètres d'entrée dans le contexte de *Prod* permet d'étendre le système pour prendre en charge des scénarios auparavant non envisagés. De nouveaux paramètres sont incorporés en réponse à des situations émergentes.

Le climat est le facteur le plus significatif dans la détermination de la croissance et de la productivité des plantes, notamment pour la culture des olives dans la zone climatique méditerranéenne. Cette région se caractérise par un temps variable, en particulier au printemps, où les températures fluctuent rapidement d'un jour à l'autre. Les résultats obtenus confirment qu'il existe une corrélation entre les températures et la production. D'un point de vue écologique, la répartition géographique de cette espèce d'arbre n'est pas limitée par le type de sol. Cela en fait une plante bien adaptée aux sols pauvres et la place parmi les principaux protecteurs de l'environnement, un facteur crucial pour l'économie du pays.

5.2.4 Évaluation de nos modèles de prévision

Pour évaluer AFFAC, nous avons réalisé une série d'expériences. Avant de détailler les spécificités, nous présenterons d'abord les objectifs de ces expériences, suivis d'un aperçu de l'installation expérimentale, y compris la machine utilisée pour mesurer les temps de réponse et le jeu de données employé.

5.2.4.1 Objectif de l'expérience

Cette expérience a pour but de comparer la performance de divers modèles prédictifs—ARIMA, RNN, LSTM, GRU, et les composants Transformer—dans la prévision de la production d'olives. En évaluant leur Erreur Quadratique Moyenne (RMSE) sur différents groupes de données, l'étude cherche à identifier l'approche la plus efficace pour traiter les tendances saisonnières et les modèles complexes dans les données agricoles. L'objectif final est d'améliorer la précision des prévisions et de fournir des informations sur les techniques de modélisation optimales pour la prévision agricole.

5.2.4.2 Configuration

Les expériences ont été réalisées en utilisant Python 3.8, en exploitant les bibliothèques Keras et TensorFlow pour la mise en œuvre des modèles d'apprentissage profond, disponibles sur Github ⁶. Le processus de développement a été facilité par l'environnement de développement intégré Spyder (IDE), et le support GPU a été activé pour TensorFlow afin d'optimiser les performances de calcul. L'environnement système comprenait un système d'exploitation Windows 11, fonctionnant sur un processeur Intel Core i7-10700K cadencé à 3,8 GHz, garantissant une gestion efficace des tâches intensives d'apprentissage automatique et offrant une vitesse d'exécution améliorée pour les processus d'entraînement et d'inférence.

L'optimisation des hyperparamètres est essentielle pour la performance des algorithmes d'apprentissage profond. Dans cette étude, nous avons sélectionné les hyperparamètres en fonction de l'erreur du modèle par rapport aux résultats réels dans le jeu de données de validation. Le processus d'ajustement commence par la recommandation de réglages initiaux, qui sont ensuite affinés par des tests empiriques. Après un ensemble de tests, les valeurs d'hyperparamètres appropriées sont choisies. Le tableau Tableau 5.1 illustre les hyperparamètres optimaux choisis parmi les plages rapportées pour chaque modèle, y compris Vanilla Transformer, Positional Encoding, Multi-head Attention, Recent Innovations, et Feed-forward & Residual Network. Ces paramètres ont été optimisés pour garantir la meilleure performance du modèle et minimiser l'erreur dans les prédictions.

Nous avons appliqué l'Analyse en Composantes Principales (PCA) et l'Analyse de Corrélation Canonique (CCA) pour simplifier les données de dimensions élevées en dimensions réduites.

Hyperparamètre	Vanilla Transformer	Positional Encoding	Multi-head Attention	Recent Innovations	Feed-forward & Residual Network
Iterations	20	20	20	20	20
Taux d'apprentissage	0.01, 0.001, 0.0001	0.01, 0.001, 0.0001	0.01, 0.001, 0.0001	0.01, 0.001, 0.0001	0.01, 0.001, 0.0001
Taille de minibatch	64, 128, 256	64, 128, 256	64, 128, 256	64, 128, 256	64, 128, 256
Unités cachées	64, 128, 256	64, 128, 256	64, 128, 256	64, 128, 256	64, 128, 256
Nombre de têtes	8, 16, 32	8, 16, 32	8, 16, 32	8, 16, 32	8, 16, 32
Probabilité de Dropout	0.1, 0.2, 0.3	0.1, 0.2, 0.3	0.1, 0.2, 0.3	0.1, 0.2, 0.3	0.1, 0.2, 0.3
Unités Feed-forward Network	128, 256, 512	128, 256, 512	128, 256, 512	128, 256, 512	128, 256, 512
Probabilité de connexion résiduelle	0.1, 0.2, 0.3	0.1, 0.2, 0.3	0.1, 0.2, 0.3	0.1, 0.2, 0.3	0.1, 0.2, 0.3

Tableau 5.1 – Plage de recherche des hyperparamètres pour les différents composants du Transformer.

6. <https://github.com/ouared14/Prod-CULT>

Notre implémentation est basée sur Keras TensorFlow 2.0. Les paramètres d'entraînement ont été soigneusement sélectionnés, en tenant compte des plages définies dans le Tableau 5.1. Pour l'entraînement, nous avons utilisé les paramètres suivants : Itérations = 8500, Taille du batch = 256, Taux d'apprentissage de base = 0.0009, Politique de taux d'apprentissage = Step, Taille du step = 1100, Gamma = 0.12, et Momentum = 0.88. Ces réglages ont été optimisés pour améliorer la performance du modèle et réduire les erreurs pendant l'entraînement.

5.2.4.3 Modélisation Prédictive pour les Olives : Comparaison entre ARIMA et les Réseaux de Neurones

La Figure 5.8 présente les performances, en termes d'erreur quadratique moyenne (RMSE), de quatre modèles : ARIMA, les réseaux de neurones récurrents (RNN), les réseaux à mémoire longue à court terme (LSTM) et les unités récurrentes à portes (GRU), appliqués à quatre clusters de données de production d'olives. ARIMA montre une performance supérieure dans les clusters 0 et 2, ce qui indique sa capacité à gérer des données présentant des tendances saisonnières claires et des croissances linéaires. Cela suggère que ces clusters possèdent des caractéristiques linéaires sous-jacentes, adaptées à l'approche de modélisation plus simple mais efficace d'ARIMA.

En revanche, les clusters 1 et 3 bénéficient de meilleures performances avec les modèles RNN et LSTM, qui excellent dans la capture des dépendances temporelles complexes et des motifs non linéaires, typiques des jeux de données agricoles influencés par de nombreuses variables telles que la variabilité climatique et les pratiques agricoles. Bien que la performance du GRU soit légèrement inférieure à celle des RNN et LSTM, il offre néanmoins un avantage compétitif grâce à son apprentissage efficace des données séquentielles.

Ces observations suggèrent que le choix du modèle en fonction des caractéristiques des données peut significativement améliorer la précision des prévisions. Ainsi, ARIMA pourrait être privilégié pour les régions présentant des rendements saisonniers prévisibles, tandis que les RNN et LSTM seraient plus adaptés aux régions soumises à des impacts environnementaux variables.

Les résultats globaux soulignent l'importance d'adapter les approches de modélisation aux spécificités des patrons de production d'olives, facilitant ainsi une meilleure prise de décision et allocation des ressources en prévision agricole. Les recherches futures pourraient explorer des modèles hybrides combinant les forces d'ARIMA et des réseaux neuronaux afin de fournir des prévisions complètes et précises, améliorant davantage les prédictions de rendement agricole.

5.2.4.4 Modélisation Prédictive pour les Olives : Comparaison des Modèles Transformer

La Figure 5.9 présente une comparaison des différentes composantes des modèles Transformer — Vanilla Transformer, codage positionnel, Multi-head Attention, Recent Innovations, et Feed-forward & Residual Network — en évaluant leurs performances en termes de RMSE sur quatre clusters de données de production d'olives. L'analyse révèle que la composante « Recent Innovations » obtient les valeurs de RMSE les plus faibles dans les clusters 0 et 3, démontrant ainsi sa capacité à capturer des motifs complexes et dynamiques dans les données de manière plus efficace que les autres composantes.

Le Feed-forward & Residual Network affiche une performance compétitive, en particulier dans le

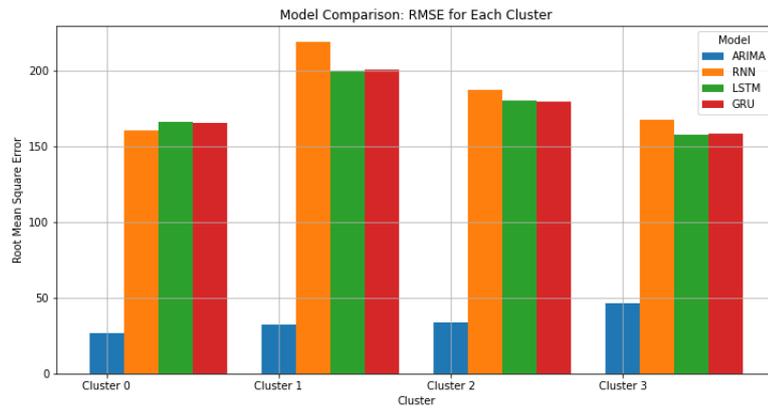


FIGURE 5.8 – Expérience de Prédiction de Précision des Modèles de Machine Learning.

cluster 2, suggérant une robustesse dans la gestion des transformations non linéaires et du traitement de l'information, ce qui le rend précieux pour des jeux de données présentant des dépendances complexes. Le codage positionnel performe constamment bien, indiquant son efficacité dans l'encodage des données séquentielles, un aspect crucial dans la modélisation de la production agricole où les séries temporelles jouent un rôle central.

Par ailleurs, le Vanilla Transformer et la composante Multi-head Attention présentent une variabilité plus importante, cette dernière excellant dans le cluster 1, mettant en lumière sa capacité à traiter en parallèle différentes représentations d'entrée. Cette figure souligne l'importance de sélectionner la composante appropriée du modèle Transformer en fonction des caractéristiques spécifiques de chaque cluster, contribuant ainsi à une meilleure précision des prévisions en production agricole.

Des recherches futures pourraient approfondir l'intégration de ces composantes afin de développer des modèles hybrides qui maximisent les forces de chacune, améliorant ainsi la précision et l'efficacité des modèles de prévision agricole.

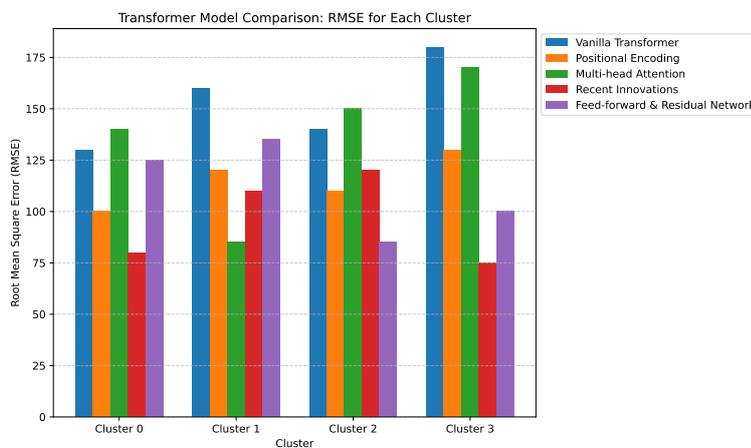


FIGURE 5.9 – Expérience de Prédiction de la Précision des Modèles de Deep Learning.

5.2.4.5 ARIMA et Réseaux de Neurones vs Modèle Transformer

Comme l'illustre le Tableau 5.2, les valeurs de RMSE obtenues par différentes techniques d'apprentissage automatique et de réseaux de neurones pour la prédiction de la production d'olives sont présentées. La comparaison inclut des modèles classiques de séries temporelles tels que ARIMA, ainsi que des architectures avancées de réseaux de neurones comme les réseaux récurrents (RNN), les réseaux à mémoire longue à court terme (LSTM), les unités récurrentes à portes (GRU), et diverses composantes des modèles Transformer, incluant Vanilla Transformer, codage d'entrée et positionnel, Multi-head Attention, Recent Innovations, et Feed-forward & Residual Networks.

Le modèle ARIMA présente la RMSE la plus élevée avec 12,5, ce qui témoigne de sa relative faiblesse à capturer des dépendances complexes dans le jeu de données. En revanche, les modèles avancés basés sur Transformer affichent généralement de meilleures performances. Les composantes Recent Innovations, avec une RMSE de 7,4, et Feed-forward & Residual Network, avec une RMSE de 7,6, se montrent particulièrement efficaces, soulignant leur capacité à gérer des interactions complexes entre caractéristiques et à fournir des prédictions précises. Multi-head Attention, avec une RMSE de 7,8, démontre une forte performance, probablement grâce à sa capacité à se concentrer simultanément sur différents aspects de l'entrée, ce qui est crucial dans des jeux de données multifacettes comme ceux de la production agricole.

À l'inverse, les modèles traditionnels de réseaux de neurones tels que RNN (10,2), LSTM (9,7) et GRU (9,5) affichent des performances modérées, probablement limitées par leur nature séquentielle, moins adaptée à la gestion des relations complexes présentes dans les données comparé aux modèles basés sur Transformer.

Cette analyse met en évidence le potentiel des architectures Transformer dans la prévision de la production agricole et suggère leur applicabilité dans des scénarios où la capture de motifs temporels et spatiaux complexes est essentielle. Des recherches futures pourraient explorer l'intégration synergique de ces techniques afin d'améliorer encore la précision et la robustesse des modèles dans des contextes agricoles variés.

En résumé, nous avons évalué divers modèles d'apprentissage automatique, incluant ARIMA, RNN, LSTM, GRU, ainsi que des architectures Transformer, pour la prédiction de la production d'olives à travers différents clusters. Les résultats montrent que les modèles avancés, tels que l'Attention Multi-tête et les Innovations Récentes, surpassent les méthodes traditionnelles, offrant une meilleure précision en termes de RMSE sur l'ensemble des clusters. Notamment, le modèle Attention Multi-tête atteint systématiquement le RMSE le plus faible, ce qui témoigne de son efficacité à capturer les motifs complexes présents dans les données de rendement oléicole. Ces résultats soulignent le potentiel des modèles d'apprentissage automatique de pointe pour améliorer la prévision agricole.

5.2.4.6 Analyse comparative du MAE entre les clusters pour la production d'olives

Dans la dernière expérimentation, nous avons utilisé une technique de clustering pour améliorer la prévision. Nous avons regroupé les régions selon leurs propriétés climatiques et leurs informations géographiques afin de comparer les performances des modèles durant les mois de production : octobre, novembre, décembre et janvier. La méthode du coude (Elbow method) a été utilisée pour déterminer

Tableau 5.2 – RMSE moyenne des différents modèles

Modèle	RMSE moyenne
Innovations Récentes	7,4
Réseau Feed-forward & Résiduel	7,6
Attention Multi-tête	7,8
Encodage d'Entrée & Positionnel	8,3
Transformer Classique	8,2
GRU	9,5
LSTM	9,7
RNN	10,2
ARIMA	12,5

le nombre optimal de clusters pour l'algorithme k-means. La Figure 5.10 présente le graphique du coude, indiquant que quatre clusters sont optimaux, car ajouter davantage de clusters n'améliore pas significativement la variance expliquée par le modèle. La visualisation des données groupées révèle des regroupements distincts de régions aux schémas de production similaires, simplifiant ainsi le processus de modélisation et améliorant l'efficacité des prévisions.

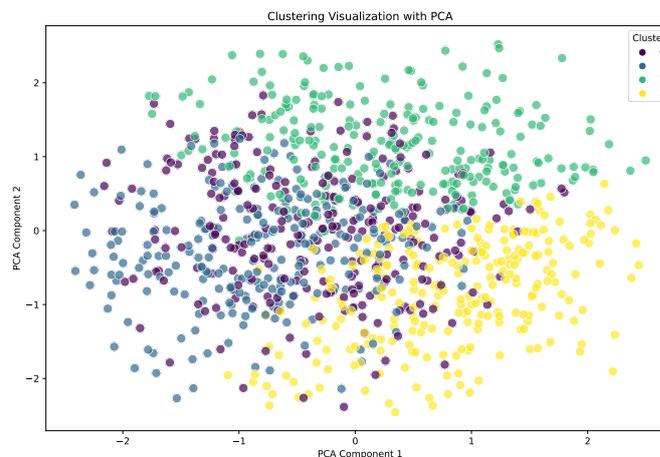


FIGURE 5.10 – Visualisation du clustering des caractéristiques régionales à l'aide de l'ACP : regroupement basé sur les attributs climatiques et atmosphériques.

La Figure 5.11 présente la comparaison de l'erreur moyenne par mois selon les clusters. Elle illustre clairement les profils moyens de production pour chaque cluster, mettant en évidence la variance des variations de production selon les modèles. Par exemple, dans le Cluster 3, on observe des valeurs de MAE (Mean Absolute Error) nettement plus faibles tout au long des mois par rapport au Cluster 2, où la variation de production est sensiblement différente. Ces variations dans les profils d'erreur reflètent des modes d'utilisation distincts de la production entre les clusters, ce qui aide à affiner la conception du modèle.

La forte variabilité observée entre les clusters influence directement la conception du modèle. Le

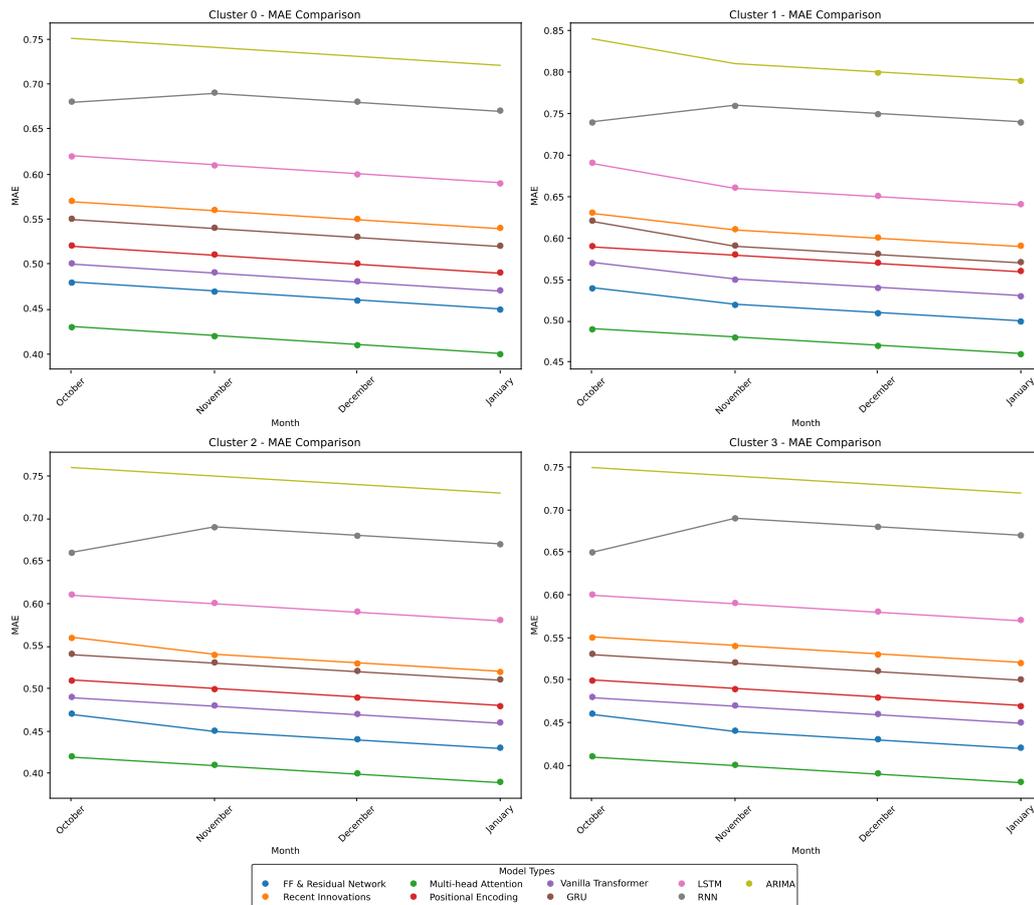


FIGURE 5.11 – Comparaison de l'erreur moyenne par jour entre les clusters.

clustering nous a permis d'adapter le modèle de prévision pour mieux gérer les différentes variations de production.

Par exemple, le modèle Feed-forward & Residual Network, appartenant au Cluster 0, affiche une MAE de 0,48 en octobre, 0,47 en novembre, et reste à 0,45 en janvier. En revanche, des modèles comme ARIMA (également dans le Cluster 0) présentent des valeurs d'erreur légèrement plus élevées au cours des mois, suggérant une marge d'optimisation dans certaines régions.

Ces différences dans le profil d'erreur ont directement orienté la conception du modèle, garantissant que l'approche de prévision soit robuste et adaptable aux différents schémas de production d'olives. Le processus de clustering a significativement amélioré l'efficacité du modèle de prévision proposé, le rendant plus flexible et capable de gérer la diversité des profils observés selon les régions.

L'évaluation des modèles d'apprentissage automatique (ARIMA, RNN, LSTM, GRU et architectures Transformer) pour la prédiction de la production d'olives montre que les approches avancées comme Multi-head Attention et Recent Innovations surpassent systématiquement les méthodes traditionnelles sur l'ensemble des clusters. Le modèle Recent Innovations obtient des résultats particulièrement remarquables, affichant la plus faible RMSE grâce à sa capacité à capturer efficacement les motifs complexes des données de rendement oléicole. Ces résultats soulignent le potentiel transformateur des techniques

de pointe en apprentissage automatique pour la prévision agricole.

5.3 Implications de la recherche

Nous avons identifié trois façons dont notre cadre a des implications pour la recherche :

- Notre cadre introduit une approche novatrice en offrant une solution flexible et évolutive visant à améliorer la productivité dans divers systèmes de production culturelle. Nous affirmons que la modélisation détaillée des paramètres environnementaux dans la production culturelle permet aux chercheurs et aux praticiens d’adapter le cadre à leurs besoins spécifiques. Cette personnalisation permet aux professionnels, tels que les économistes agricoles et les développeurs de logiciels, de concevoir des fonctions de coût plus précises et efficaces, optimisant ainsi la production culturelle en agriculture en fonction des contextes opérationnels et environnementaux spécifiques. Cette approche repose sur le principe que l’efficacité d’un modèle de coût de production culturelle dépend de paramètres clairement définis à travers plusieurs couches : environnementale, biologique, agronomique, technologique, économique, sociopolitique et culturelle. La valeur tirée du modèle est façonnée par les interactions complexes entre ces facteurs, mettant en évidence la nécessité de comprendre de manière approfondie comment chaque couche influence les coûts de production globaux.
- Notre cadre ouvre la voie à l’exploration de la possibilité de capitaliser sur les résultats des tests des praticiens et chercheurs en les stockant dans un référentiel basé sur des modèles.
- Notre cadre peut être utilisé dans les environnements d’enseignement et d’apprentissage, permettant aux étudiants de se concentrer sur les composants logiques de leurs applications et sur les différentes signatures des fonctions de coût de production, tout en comprenant les facteurs résumant l’impact des coûts de production culturelle.

En conclusion, cette étude fournit des approches de modélisation des coûts de production culturelle et des éléments de conception basés sur un cadre de modélisation axé sur les données.

5.3.1 Implications industrielles

En termes d’implications industrielles, nous avons identifié trois façons dont notre cadre peut être appliqué :

- Notre cadre peut automatiquement générer la documentation pour le modèle de production culturelle, facilitant ainsi sa réutilisation et manipulation par les praticiens. D’un point de vue pratique, nous envisageons de développer un système web dédié à la reproductibilité des productions culturelles modèles, comprenant une gamme de techniques pour explorer les modèles de production culturelle.
- En définissant efficacement le contexte du modèle de production culturelle, notre cadre aide les chercheurs et les praticiens à maintenir leur concentration lors des tests de performance, évitant ainsi une focalisation excessive sur les préoccupations liées aux graphiques.
- Des processus automatisés peuvent être réalisés en incorporant des pratiques itératives et incrémentielles dans le déploiement automatisé. Ces pratiques permettent aux scientifiques de répéter

rapidement les expériences jusqu'à ce que les erreurs soient minimisées.

- Notre cadre facilite la collaboration entre les concepteurs de modèles de production culturelle et leur permet d'automatiser et de rationaliser le flux de travail traditionnel pour le calibrage des paramètres du modèle à travers les collectes de données. Lorsqu'un développement de modèle nécessite l'expertise de plusieurs domaines, notre cadre permet aux concepteurs de travailler ensemble de manière synchrone. Il prend également en charge la fusion automatique, la résolution des conflits et le contrôle des versions.

Les chercheurs et les professionnels de l'industrie dans le domaine de la production culturelle sont encouragés à partager leurs expériences, ce qui contribuera à combler le fossé entre la recherche académique et les pratiques industrielles. À mesure que le domaine de la conception physique des bases de données continue de mûrir, il existe un mouvement croissant pour établir un référentiel de connaissances open-source basé sur le web. Cette initiative vise à résoudre divers défis socio-techniques, notamment la science ouverte, les logiciels open-source, la collaboration multidisciplinaire et la reproductibilité des expériences de performance des modèles.

5.4 Outil prototype d'assistance aux agriculteurs sous forme d'application mobile

L'application mobile vise à proposer un service d'assistance aux agriculteurs en leur fournissant des analyses basées sur leurs données. Notre cadre est constitué de trois modules principaux. Le premier module est consacré au stockage de données, où toutes les informations pertinentes sont collectées et organisées. Le deuxième module se concentre sur l'analyse, regroupant des analyses descriptives, prédictives et diagnostiques. Ce module est particulièrement utile pour les analystes de données et les gestionnaires, qui envoient des rapports détaillés aux agriculteurs. Enfin, le troisième module est axé sur la recommandation et l'intervention. Il permet aux agriculteurs d'agir en fonction des recommandations issues des rapports, facilitant ainsi une gestion proactive et efficace de leurs cultures. Grâce à ces trois modules, OLIVNET offre une approche intégrée pour améliorer la productivité et la durabilité des exploitations agricoles.

5.4.1 Présentation de notre outil prototype

Pour la conception UI/UX, nous avons utilisé Figma : Figma est un outil vectoriel basé dans le cloud, permettant aux utilisateurs de travailler de n'importe où via un navigateur. Il s'agit d'un outil de conception et de prototypage conçu pour les designers numériques, suffisamment accessible pour être utilisé par des non-spécialistes tout en facilitant le partage ou l'extraction de fichiers. Pour l'application mobile, nous avons utilisé WinDev Mobile : c'est un environnement de développement intégré (EDI) qui permet de créer des applications multiplateformes (Android, iOS, etc.) à partir d'un code unique, avec des fonctionnalités avancées pour la gestion des interfaces et des services connectés. Pour la base de données, nous avons utilisé le SGBD HyperFileSQL de WinDev avec le langage WLangage : HyperFileSQL est un système de gestion de base de données (SGBD) relationnelle intégré à WinDev, qui permet de stocker, gérer et récupérer des données efficacement, tout en utilisant WLangage pour simplifier les opérations sur les données.

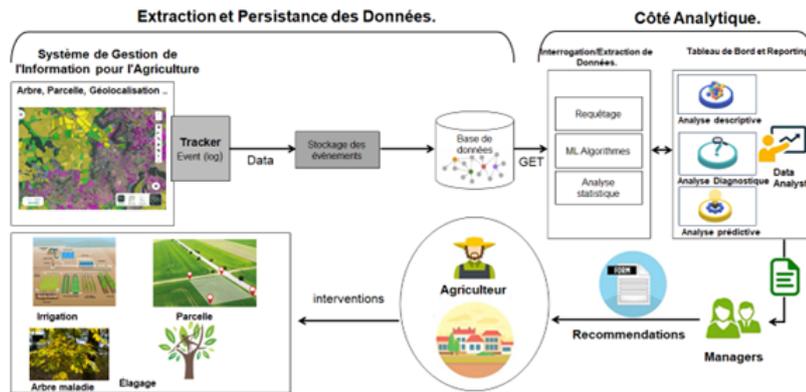


FIGURE 5.12 – Présentation de notre outil prototype.

5.4.2 Architecture de notre application

La mise en œuvre de notre application mobile suit une approche basée sur des règles, en utilisant un dictionnaire de réponses stocké dans un système de base de données comme par exemple MongoDB. L'API Flask invoque soit le système basé sur des règles, soit le modèle d'IA en fonction de la requête de l'utilisateur. De plus, la section FAQ est dédiée à répondre aux questions fréquemment posées, tandis que l'historique des interactions des utilisateurs peut être exploité par les services du tableau de bord dans le centre utilisateur. La Figure 5.13 illustre l'architecture de notre application.

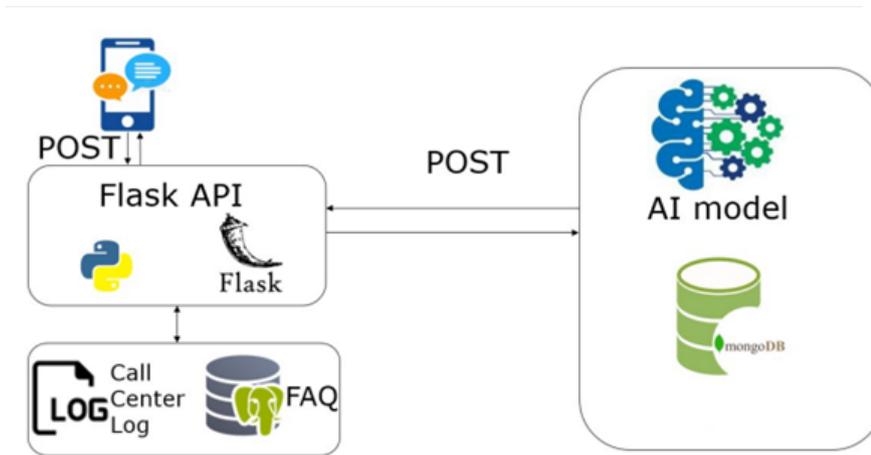


FIGURE 5.13 – Architecture de notre application.

5.4.2.1 Overview of Flask and RESTful API

At this part of the project, we will talk about deploy our model ai using Flask as illustrated in the following figure.

5.4.2.2 Client-Side Interface

The client-side interface can invoke the AI model or model-based chatbot from any user interface, whether it's a tablet, mobile device, or website. This invocation is done simply by introducing the text of the query and invoking the system remotely.

5.4.2.3 Server-Side Application

The server-side application handles the chatbot model, whether it is an AI model or a dictionary of responses, along with the similarity metrics to generate the appropriate response.

5.4.2.4 Integration with External Services

L'intégration avec des services externes est une technique qui connecte notre système à un écosystème plus large. Par exemple, nous pouvons intégrer notre système avec des services météorologiques et de géolocalisation pour explorer des propriétés atmosphériques telles que la localisation de la température. Cela permet à notre système de suivre et d'analyser en temps réel la situation actuelle, fournissant des informations précieuses et améliorant ainsi la fonctionnalité globale ainsi que l'expérience utilisateur.

5.4.3 Processus de conception d'une application mobile

In designing this application, we followed a set of stages. The Figure 5.15 illustrates the Mobile app design process

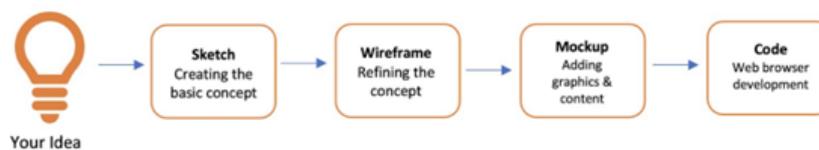


FIGURE 5.14 – Processus de conception d'une application mobile.

5.4.3.1 Sketching

Le terme « Sketching » dans la conception d'applications désigne la phase préliminaire de production d'esquisses brutes et de faible fidélité pour recueillir des idées d'inspiration et de mise en page pour votre application mobile. À la suite des discussions de l'équipe, nous avons élaboré cette conception préliminaire. L'esquisse de notre application mobile réalisée lors d'une séance de brainstorming avec notre superviseur.

5.4.3.2 Wireframing

Les wireframes sont un outil fondamental dans le processus de conception d'applications mobiles, servant de représentations visuelles simplifiées de la mise en page et des fonctionnalités d'une application.

5.4.3.3 Storyboards

Un storyboard sert de plan visuel, présentant les idées clés et la structure d'une expérience prévue. Il utilise une séquence d'illustrations, d'images ou d'écrans pour fournir une prévisualisation et établir l'ordre des événements. Cela permet d'obtenir des retours précoces et d'affiner le projet avant la création de l'expérience réelle.

5.4.4 Présentation de notre outil d'assistance

Dans cette section, nous présentons l'interface UI/UX de notre solution mobile.

5.4.4.1 Écran d'accueil

L'interface d'accueil de OLIVINTEL est conçue de manière à refléter visuellement les aspects écologiques et environnementaux. Chaque élément de design, des couleurs aux icônes, vise à sensibiliser les utilisateurs à la durabilité et à l'importance de la préservation de la nature, tout en offrant une expérience utilisateur harmonieuse et intuitive.

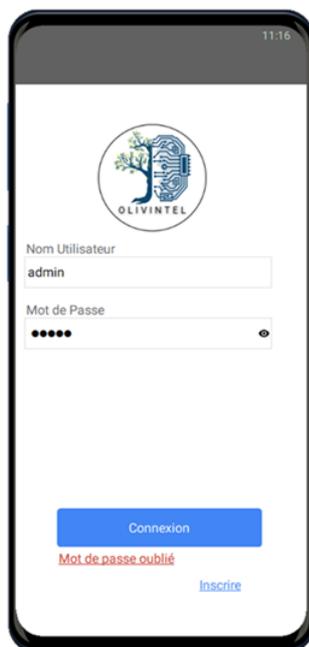


FIGURE 5.15 – Écran d'accueil.

5.4.4.2 Écran Inscription

Les utilisateurs qui représentent les agriculteurs peuvent s'inscrire dans notre système via un formulaire, en spécifiant la wilaya, daïra, commune, et l'adresse. Ils doivent également créer un compte en choisissant un nom d'utilisateur et un mot de passe. Cela leur permettra d'accéder à notre plateforme et d'exploiter les services disponibles, tout en garantissant la sécurité et la personnalisation de leur expérience utilisateur (Figure 5.16).



FIGURE 5.16 – Écran d'accueil.

5.4.4.3 Écran Menu

L'interface du menu principal de l'application OLIVNET est conçue pour offrir un accès rapide et intuitif aux principales fonctionnalités destinées aux agriculteurs. Comme illustre la Figure 5.17, le menu présente les sections suivantes :

1. **Mes Champs** : Cette section permet aux utilisateurs de gérer leurs parcelles d'oliviers, d'ajouter des détails sur les cultures, et de suivre l'évolution des plantations.
2. **Suivi des Arbres** : Ici, l'agriculteur peut surveiller l'état de ses arbres, enregistrer des observations sur la santé des oliviers, et consulter des alertes liées à leur croissance ou à des événements climatiques.
3. **Statistiques** : Cette fonction affiche des graphiques et des données analytiques sur la production, la santé des arbres, et d'autres indicateurs clés pour aider à prendre des décisions éclairées.

4. **Manifestes** : Dans cette section, les utilisateurs peuvent consigner et gérer les différentes activités réalisées sur leurs parcelles (traitements, récoltes, etc.), facilitant ainsi le suivi des interventions.
5. **Rapports** : Les agriculteurs peuvent générer des rapports personnalisés sur l'état des cultures, les rendements, et les événements marquants, offrant une vue d'ensemble claire de la situation de leurs exploitations.
6. **À propos de l'application** : Une section informative qui présente OLIVNET, ses objectifs, et les services qu'elle propose aux utilisateurs pour améliorer la gestion de leurs exploitations agricoles.

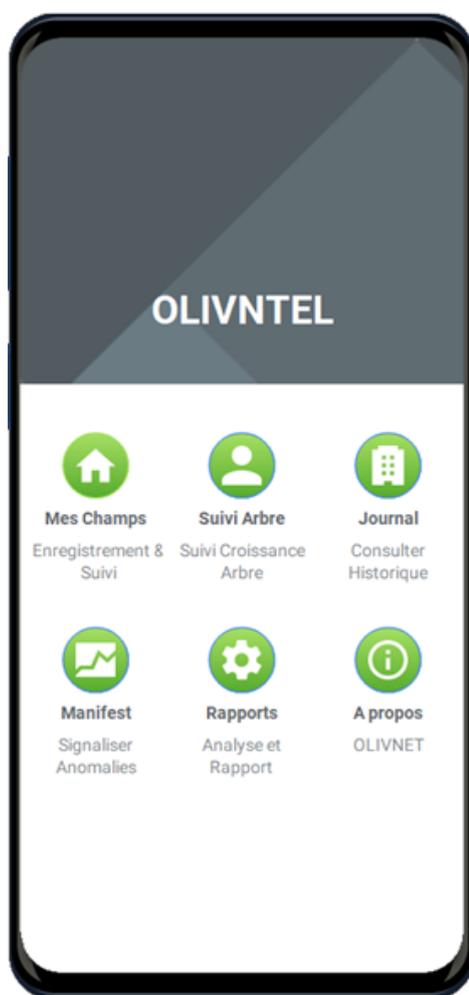


FIGURE 5.17 – Écran d'accueil.

L'interface d'OLIVNET est simple et conviviale, facilitant ainsi la navigation et l'accès aux différentes fonctionnalités pour optimiser la gestion des oliveraies.

5.4.4.4 Écran de gestion des champs

L'interface de gestion des champs de l'application OLIVNET propose une vue intuitive sur les activités liées à la gestion des parcelles d'oliviers. Elle regroupe plusieurs fonctionnalités essentielles.

Les utilisateurs peuvent (Figure 5.18) :

- enregistrer de nouvelles parcelles en précisant la superficie, le type de sol et la variété d'olivier, avec la possibilité d'importer des données existantes ;
- gérer les systèmes d'irrigation, programmer les horaires d'arrosage et configurer des alertes pour les besoins en eau ;
- enregistrer les types de fertilisants, les quantités et les dates avec des recommandations basées sur l'analyse du sol ;
- planifier et suivre les activités d'élagage, notant les méthodes et les résultats ;
- consigner les opérations de labour et gérer les traitements de désherbage avec des alertes pour les interventions à venir.

L'interface est claire et ergonomique, avec des icônes facilement identifiables et des graphiques pour suivre les performances en temps réel, permettant ainsi aux agriculteurs d'optimiser la gestion de leurs oliveraies.

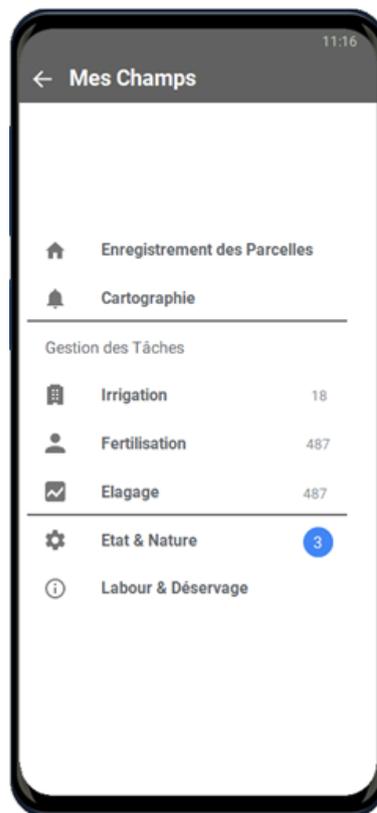


FIGURE 5.18 – Écran de gestion des champs.

5.4.4.5 Écran de saisie des champs

L'interface d'enregistrement des parcelles permet aux utilisateurs de saisir des informations clés sur chaque parcelle d'oliviers (Figure 5.19). Le formulaire inclut des champs pour :

- le nom ;
- la superficie ;

- le type de sol ;
- l'adresse ;
- le type d'irrigation ;
- et une option pour télécharger une photo.

Des validations assurent la saisie correcte des données avant de cliquer sur "Enregistrer". Un message de confirmation informe l'utilisateur de l'ajout réussi de la parcelle. Cette interface intuitive facilite la gestion des exploitations agricoles.

The screenshot shows a mobile application interface titled "Enregistrement Parcelles". At the top right, the time is 11:16. Below the title is a back arrow and a link "Liste Parcelles". The form contains the following fields and values:

- Nom de Parcelle: Parcelle 1
- Superficie: 12 500,00
- Type: Olivier
- Latitude: 12,00
- Longitude: 33,00
- Nombre Arbres: 67
- Adresse: Tiaret
- Type Irrigation: Goutte à goutte

Below the form is a photo upload button (three dots) and a satellite image of a field. At the bottom is a blue "Enregistrer" button.

FIGURE 5.19 – Écran de saisie des champs.

5.4.4.6 Écran de cartographie des champs

L'interface de cartographie des champs permet de visualiser la localisation des parcelles (Figure 5.20). Elle intègre des services en ligne, comme ceux de Google, pour récupérer des informations et des images satellites enrichies par des données spatio-temporelles. Cette fonctionnalité permet aux agriculteurs de prendre des décisions basées sur des informations actualisées et contextuelles.

5.4.4.7 Écran Liste de champs

L'interface Liste de champs affiche pour chaque parcelle des informations essentielles : nom, superficie, type de sol, type d'irrigation et statut. Les options de filtrage et de recherche, ainsi que des boutons



FIGURE 5.20 – Écran de cartographie des champs.

d'action, facilitent la navigation et la gestion efficace des parcelles (Figure 5.20).

5.4.4.8 Écran de suivi des arbres

L'interface de suivi des arbres permet de gérer les cultures. Elle présente une liste d'arbres avec des informations clés, telles que le numéro d'identification, le nom, la variété, la date de plantation et l'état de santé. Les utilisateurs peuvent ajouter des observations et consulter des graphiques pour suivre les performances des arbres (Figure 5.22).

5.4.4.9 Écran Historique

L'interface Historique affiche des graphiques qui décrivent la situation des arbres et aident à la prise de décision. Ces visualisations rendent les données compréhensibles et offrent des insights pour optimiser la gestion des cultures (Figure 5.23).

11:16

← Liste Parcelles

ID de Parcelle

Nom

Superficie

Adresse

ID de Parcelle

Nom

Superficie

Adresse

ID de Parcelle

Nom

Superficie

FIGURE 5.21 – Écran Liste de champs.

5.4.4.10 Écran de signalement d'anomalies

L'interface Écran de Signalement d'Anomalies permet de signaler facilement des problèmes (Figure 5.24). Les agriculteurs peuvent :

- décrire les anomalies ;
- envoyer des images pour analyse ;
- et exploiter un traitement basé sur des modèles de réseaux de neurones convolutifs (CNN).

Cela facilite une gestion rapide et efficace des anomalies.

5.4.4.11 Écran Analyse et Rapport

L'interface Analyse et Rapport fournit des mesures et métriques concernant les parcelles, comme le rendement et la qualité du sol. Ces informations aident les agriculteurs à optimiser leurs pratiques agricoles grâce à des recommandations basées sur des données en temps réel (Figure 5.25).

5.4.4.12 Écran de contact

La section À Propos de l'application présente ses fonctionnalités principales et met en avant son objectif : améliorer la productivité agricole avec des technologies modernes comme l'intelligence

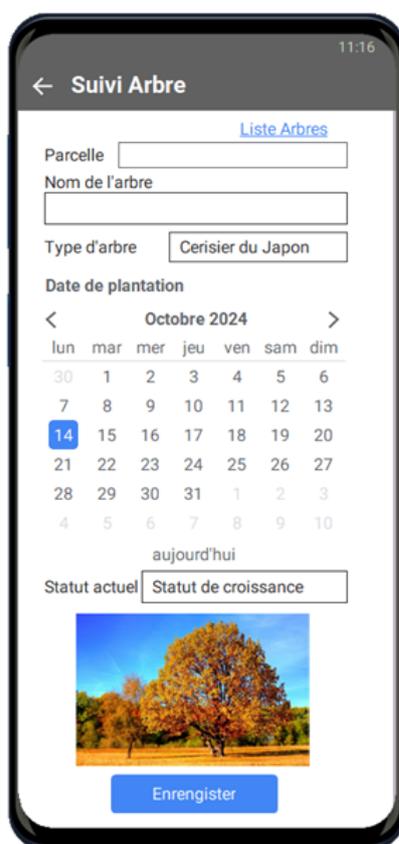


FIGURE 5.22 – Écran de suivi des arbres.

artificielle (Figure 5.26).

En conclusion, dans cette section, nous avons proposé une application mobile dédiée aux agriculteurs pour les assister dans leurs activités quotidiennes de gestion des parcelles et des arbres. Notre outil offre des fonctionnalités de gestion telles que la saisie des parcelles, le suivi des manifestations, ainsi qu'un module de recommandations permettant de recevoir des alertes, des rapports et des analyses automatisées. Ces fonctionnalités visent à faciliter la prise de décision et à optimiser la gestion des ressources agricoles, contribuant ainsi à une meilleure productivité et efficacité dans le secteur.

5.5 Discussion

En combinant une large gamme de facteurs environnementaux et spécifiques aux cultures et en utilisant des modèles de pointe basés sur des transformateurs tels que LSTM, CNN-LSTM et Multi-head Attention, le cadre AFFAC améliore considérablement la prévision agricole. Contrairement aux méthodes conventionnelles comme la méthode ARMA dans l'étude de Kurtouglu et al. [95], qui se limitait aux séries chronologiques stables, ou les modèles de séries chronologiques dans [103], qui ont démontré la supériorité des réseaux neuronaux profonds mais étaient limités par un manque de généralisation à travers les environnements, notre étude propose un cadre modulaire et très flexible capable de capturer des relations complexes et non linéaires dans une variété de contextes agricoles.

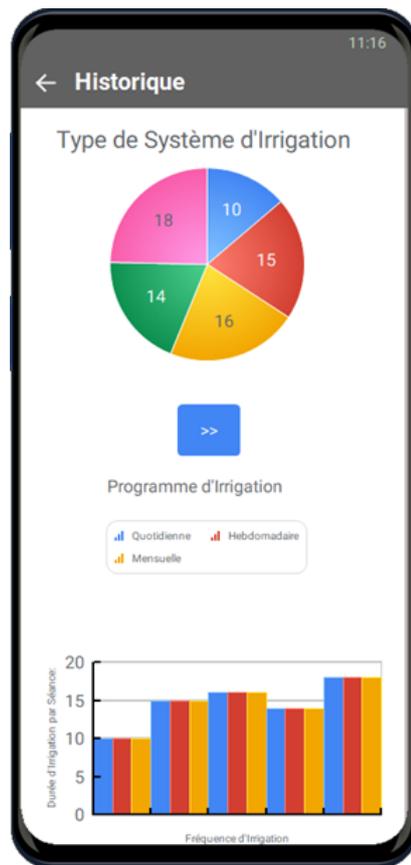


FIGURE 5.23 – Écran Historique.

Le potentiel de l'apprentissage automatique et des modèles hybrides pour la prévision des rendements des cultures a été montré par [30], [123] et [129] ; cependant, leurs résultats étaient fréquemment limités par de petits ensembles de données, la spécificité régionale ou une concentration sur des cultures particulières. Par exemple, les modèles Recent Innovations de notre cadre ont systématiquement obtenu des valeurs RMSE plus faibles et ont mieux capturé les motifs complexes des rendements que les modèles de régression logistique et d'arbres de décision dans [30], bien qu'ils aient été efficaces pour prédire les rendements des cultures en Afrique de l'Ouest. De manière comparable, notre cadre est applicable à une variété de cultures et de zones géographiques, tandis que le modèle hybride CNN-DNN utilisé dans [123] a réussi à prédire le rendement du soja dans la ceinture de maïs des États-Unis. Bien que l'étude [129] ait été révolutionnaire dans son utilisation de la télédétection pour prévoir la production d'olives dans ses premières étapes, elle n'a pas exploré l'intégration de nombreuses caractéristiques environnementales aussi complètement que AFFAC, qui prend en compte les pratiques spécifiques aux cultures, les conditions météorologiques, la qualité du sol et l'irrigation.

AFFAC n'est cependant pas sans limitations. Contrairement aux modèles plus accessibles dans [30], la complexité du cadre peut poser des difficultés pour les utilisateurs ayant peu de connaissances techniques, bien que sa conception modulaire permette des modifications. De plus, le modèle ARMA de [95], qui nécessite seulement des données historiques de séries chronologiques, présente une limitation moins sévère que le cadre AFFAC, qui dépend principalement de jeux de données diversifiés et de haute

FIGURE 5.24 – Écran de signalement d’anomalies.

qualité qui ne sont pas toujours disponibles dans des environnements à ressources limitées. En outre, bien que AFFAC soit assez performant pour effectuer des prévisions, il n'utilise pas encore de stratégies d'explicabilité comme le modèle XAI-CROP dans [137], qui émet des suggestions faciles à comprendre grâce à LIME. Son manque d'interprétabilité peut empêcher les parties prenantes, qui ont besoin d'outils de prise de décision clairs, de l'utiliser.

D'autres recherches, cependant, présentent des avantages spécifiques. Par exemple, le système de détection d'objets basé sur YOLO dans [49] offre des solutions extrêmement précises en temps réel pour des tâches d'agriculture de précision telles que le comptage des fruits et des fleurs, une application spécialisée que AFFAC ne traite pas directement. De même, la méthode de détection de maladies basée sur l'IA proposée par [59] montre comment l'apprentissage profond peut être utilisé pour résoudre des problèmes agricoles particuliers, comme la reconnaissance des maladies des haricots communs, ce qui pourrait améliorer les capacités de prévision plus globales de AFFAC. De plus, l'importance d'incorporer la dynamique du marché dans la prise de décision agricole est mise en évidence par l'accent mis par [94] sur la prévision des prix en utilisant des modèles ARIMA et RNN. Il s'agit d'un domaine qui pourrait être exploré dans les générations futures du cadre AFFAC.

En conclusion, il est important de reconnaître les limitations du cadre AFFAC, notamment sa complexité, sa dépendance aux données et son manque d'explicabilité, même s'il représente un progrès considérable dans la prévision agricole grâce à ses méthodologies de modélisation sophistiquées et à

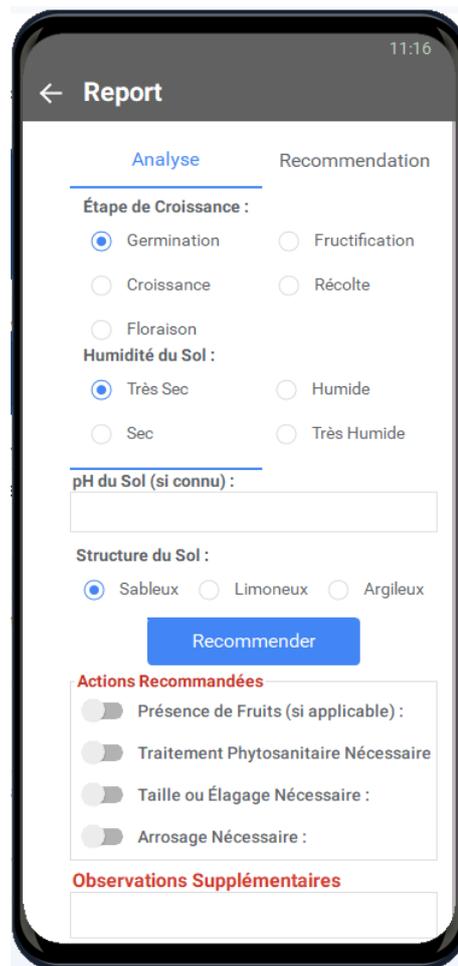


FIGURE 5.25 – Écran Analyse et Rapport.

son intégration complète des données. Les travaux futurs sur le cadre AFFAC pourraient aborder ces limitations et améliorer davantage son applicabilité et son impact dans divers contextes agricoles, en tirant parti des points forts d'autres études, comme l'interprétabilité de [137], la simplicité de [30] et la précision en temps réel de [49].

5.6 Discussion

En combinant une large gamme de facteurs environnementaux et spécifiques aux cultures, et en utilisant des modèles de pointe basés sur les transformers tels que LSTM, CNN-LSTM et l'Attention Multi-tête, le cadre AFFAC améliore significativement la prévision agricole. Contrairement aux méthodes classiques comme la méthode ARMA utilisée dans l'étude de Kurtouglu et al. [95], qui se limite aux séries temporelles stables, ou aux modèles de séries temporelles de [103], qui ont démontré la supériorité des réseaux de neurones profonds mais étaient limités par un manque de généralisation entre environnements, cette étude propose un cadre modulaire et très flexible capable de capturer des relations complexes et non linéaires dans une variété de contextes agricoles. Le potentiel de l'apprentissage



FIGURE 5.26 – Écran de contact.

automatique et des modèles hybrides pour la prédiction des rendements a été montré par [30], [123] et [129] ; néanmoins, leurs résultats étaient souvent limités par de petits jeux de données, une spécificité régionale ou une focalisation sur certaines cultures. Par exemple, les modèles Innovations Récentes du cadre ont systématiquement obtenu des valeurs de RMSE plus faibles et ont mieux capturé les schémas complexes de rendement que les modèles de régression logistique et d'arbre de décision de [30], bien que ces derniers soient efficaces pour prédire le rendement des cultures en Afrique de l'Ouest. De même, le cadre proposé est applicable à diverses cultures et zones géographiques, tandis que le modèle hybride CNN-DNN utilisé dans [123] a réussi à prédire le rendement du soja dans la Corn Belt américaine. Bien que l'étude [129] ait été révolutionnaire par son utilisation de la télédétection pour prévoir la production d'olives à ses débuts, elle n'a pas exploré l'intégration aussi complète de nombreuses caractéristiques environnementales que AFFAC, qui prend en compte les pratiques culturelles spécifiques, la météo, la qualité des sols et l'irrigation.

Cependant, AFFAC présente certaines limitations. Contrairement aux modèles plus accessibles présentés dans [30], la complexité du cadre peut poser des difficultés aux utilisateurs peu expérimentés, bien que sa conception modulaire permette des modifications. De plus, le modèle ARMA de [95], qui ne nécessite que des données historiques de séries temporelles, présente une limitation moins importante que le cadre AFFAC, qui dépend principalement de jeux de données diversifiés et de haute qualité, souvent indisponibles dans des environnements aux ressources limitées. Par ailleurs, même si AFFAC

est performant en prédiction, il n'intègre pas encore de stratégies d'explicabilité comme le modèle XAI-CROP de [137], qui fournit des explications facilement compréhensibles grâce à LIME. Son manque d'interprétabilité pourrait freiner son adoption par les acteurs nécessitant des outils décisionnels clairs.

D'autres travaux présentent cependant des avantages spécifiques. Par exemple, le système de détection d'objets basé sur YOLO dans [49] offre des solutions très précises et en temps réel pour des applications d'agriculture de précision, telles que le comptage de fruits et de fleurs, une application spécialisée que AFFAC ne couvre pas directement. De même, la méthode de détection des maladies basée sur l'intelligence artificielle proposée par [59] illustre comment le deep learning peut résoudre des problèmes agricoles spécifiques, comme la reconnaissance des maladies du haricot commun, ce qui pourrait compléter les capacités de prévision plus globales de AFFAC. En outre, l'importance d'intégrer les dynamiques du marché dans la prise de décision agricole est mise en lumière par [94], qui se concentre sur la prévision des prix à l'aide de modèles ARIMA et RNN. Ce domaine pourrait être exploré dans les prochaines versions du cadre AFFAC.

En conclusion, il est essentiel de reconnaître les limites du cadre AFFAC, notamment sa complexité, sa dépendance aux données et son manque d'explicabilité, même s'il constitue une avancée majeure dans la prévision agricole grâce à ses méthodologies sophistiquées de modélisation et à son intégration approfondie des données. Les travaux futurs sur AFFAC pourraient lever ces limitations et améliorer son applicabilité ainsi que son impact dans divers contextes agricoles, en s'inspirant des points forts d'autres études, telles que l'interprétabilité de [137], la simplicité de [30] et la précision en temps réel de [49].

5.7 Conclusion

Le climat est le facteur le plus significatif dans la détermination de la croissance et de la productivité des plantes, en particulier pour la culture des olives dans la zone climatique méditerranéenne. Cette région se caractérise par des conditions météorologiques variables, notamment au printemps, où les températures fluctuent rapidement d'un jour à l'autre. Les résultats obtenus confirment qu'il existe une corrélation entre les températures et la production. D'un point de vue écologique, la distribution géographique de cette espèce d'arbre n'est pas limitée par le type de sol. Cela en fait une plante bien adaptée aux sols pauvres, la plaçant parmi les principaux protecteurs de l'environnement et un facteur crucial pour l'économie du pays.

L'objectif principal de ce chapitre est de proposer un ensemble de principes de conception pour le développement de fonctions de coût mathématiques au sein d'un cadre de modélisation basé sur les données. En général, le développement de modèles de production (*Prods*) est freiné par l'absence d'un processus de conception structuré. Souvent, ces modèles sont soit esquissés de manière informelle, directement implémentés dans des outils sans conception préalable, soit conçus mentalement sans documentation formelle de la conception.

Pour aborder cette question, il est nécessaire de repenser le processus de développement des *Prod*, en mettant l'accent sur la conception et les tests afin de créer des solutions *Prod* efficaces.

Notre principale contribution est l'introduction d'un cadre basé sur des modèles spécifiquement destiné à la conception et au développement des *Prods* de prévision mathématique des coûts. Ce cadre définit un processus systématique pour collecter et gérer les données relatives à la production

culturelle en agriculture, en se concentrant sur les multiples sources de données et processus. Il intègre les connaissances des concepteurs pour la sélection des variables d'entrée liées aux facteurs d'impact les plus sensibles de la production culturelle. Il peut être facilement personnalisé et permet aux utilisateurs de naviguer parmi différents modèles de prévision en fonction de leurs tâches de conception en attachant simplement les paramètres d'entrée à la spécification du modèle.

Nous désignons cette contribution sous le nom de Cadre de Modélisation Basé sur les Données pour la Prévision de la Production Culturelle en Agriculture. Le processus de conception proposé comprend les phases suivantes : (1) Expression des exigences, (2) Identification des dimensions, (3) Identification des facteurs impactant les coûts de production culturelle, (4) Construction du modèle de prévision, et (5) Tests et évaluation.

Pour évaluer les attributs de qualité du *Prod* développé, nous avons mené plusieurs expériences évaluant ses capacités de précision. Les résultats ont démontré que notre *Prod* prédit avec précision les performances des requêtes, incitant ainsi les concepteurs à faire moins d'hypothèses lors de la phase de conception. Ce succès est attribué aux lignes directrices de conception complètes de notre cadre.

En résumé, les modèles avancés d'apprentissage automatique, notamment la Multi-head Attention et les innovations récentes, surpassent largement les méthodes traditionnelles pour prédire la production d'olives. Le modèle Multi-head Attention présente systématiquement la plus faible erreur quadratique moyenne (RMSE), soulignant son efficacité à capter des motifs complexes de rendement et mettant en évidence la valeur des techniques modernes dans la prévision agricole.

Notre travail présente plusieurs pistes de recherche pour l'avenir, offrant des opportunités pour de futures études. D'un point de vue théorique, nous prévoyons d'incorporer la détection d'anomalies dans les données provenant des capteurs. Bien que le contexte semble similaire au nôtre, les données sont suffisamment spécifiques pour permettre l'application de certaines méthodes qui ne se généralisent pas à notre étude de cas. Nous envisageons également d'étudier comment notre cadre peut réduire la complexité du *Prod* afin d'améliorer sa réactivité rapide. D'un point de vue méthodologique, nous prévoyons également d'attribuer un score de confiance aux mesures des capteurs montés sur des équipements industriels, afin d'utiliser ces mesures à des fins prédictives ou diagnostiques (par exemple, prévision des défaillances, détection ou étude de comportements spécifiques, etc.). D'un point de vue pratique, nous envisageons de développer une application mobile pour invoquer le modèle et recevoir des recommandations en fonction des entrées de l'utilisateur.

Conclusion générale et Perspectives



*« Pour accomplir de grandes choses, il faut non seulement agir, mais aussi rêver ;
non seulement planifier, mais aussi croire. »*

— Anatole France (1844-1924)

Sommaire

6.1 Conclusion	162
6.1.1 État de l'art	162
6.1.2 Contributions	162
6.1.3 Evaluation et Discussion	163
6.2 Perspectives	164
6.2.1 Intégration de la détection d'anomalies	164
6.2.2 Réduction de la complexité du modèle pour améliorer la réactivité	164
6.2.3 Attribution d'un score de confiance aux mesures des capteurs	165
6.2.4 Développement d'une application mobile pour l'invocation du modèle	165

6.1 Conclusion

Ce chapitre présente un bilan de notre travail ainsi qu'un ensemble d'ouvertures et de perspectives.

6.1.1 État de l'art

Notre travail a été guidé par une étude approfondie des travaux existants sur la prévision et le forecasting des cultures agricoles, ainsi que sur les principales dimensions à prendre en compte lors de la modélisation dirigée par les données. Nous avons structuré notre étude en trois axes : **(i)** Premier axe : Nous avons d'abord explicité les paramètres d'impact, étudié le formalisme et les méthodes utilisées, en mettant un accent particulier sur les outils dédiés. Nous avons également décrit le processus de collecte des données et des modèles, en tenant compte de la qualité de service (QoS) comme une contrainte imposée par les utilisateurs ou les systèmes en cours de conception. **(ii)** Deuxième axe : Ce volet porte sur les approches de modélisation de la fonction de prévision. Nous avons présenté les différents aspects de la modélisation dans le cadre des approches dirigées par les données, telles que l'apprentissage automatique (machine learning et deep learning). Nous avons également passé en revue les travaux intégrant des données historiques pour améliorer la précision des modèles et la qualité des services. Une attention particulière a été accordée à la variabilité de l'indice NDVI, indicateur clé de l'activité végétative, et à sa corrélation avec la variabilité climatique (température, précipitations, humidité), dans le but de capturer les relations causales entre conditions climatiques et rendements agricoles. **(iii)** Troisième axe : Enfin, nous avons détaillé les travaux de modélisation appliqués au domaine agricole, en étudiant les principales approches adoptées par la communauté. À partir de cette analyse, nous avons identifié les lacunes des approches existantes et proposé nos contributions pour les rendre plus efficaces. En conclusion, nous avons positionné nos contributions, exposé notre motivation, et donné un aperçu de la solution envisagée.

6.1.2 Contributions

L'objectif principal de ce travail était de concevoir un framework appelé AFFAC pour Modeling and Testing Forecasting in Agricultural Production Systems, basé sur l'ingénierie dirigée par les modèles. Fournir un processus collaboratif pour ce type de transformation de modèle est difficile et crucial. Par conséquent, l'intégration de techniques de machine learning/deep learning dans une architecture de modèles nous aide à créer une décision finale combinée à partir de vues de conception individuelles et différentes. Les croyances des modèles sont basées sur les connaissances d'experts dans le domaine de Modeling and Testing Forecasting in Agricultural Production Systems et sont structurées en paires d'exemples (paramètres des cultures, fonction de production).

Bien que, dans notre expérimentation, tous les modèles utilisent la même méthode pour produire la fonction de production, l'idée de produire la fonction de production finale à partir de modèles avec des stratégies différentes mérite d'être approfondie. Une première étape de la démarche consiste à envoyer toutes les données nécessaires aux modèles afin qu'ils produisent leurs propres solutions. Dans la deuxième étape, le MA récupère toutes les propositions des modèles et utilise des heuristiques liées à un ensemble recommandé d'attributs QoS pour générer la meilleure fonction de production correspondant

aux paramètres des cultures initiaux.

Une autre caractéristique de notre modèle est qu'il est conçu de manière générique, de sorte que son architecture modulaire permet de remplacer facilement certains composants sans affecter la fonctionnalité globale, tels que : (1) le type de données utilisées par les modèles selon un cas spécifique de Modeling and Testing Forecasting in Agricultural Production Systems, (2) l'algorithme utilisé par les modèles pour produire la fonction de production, et (3) le composant plug-gable dans le modèle de prevision qui intègre des heuristiques pour affiner la fonction de production finale.

Notre première contribution consiste en la proposition d'un cadre pour la modélisation et le test de la prévision dans les systèmes de production agricole, basé sur des modèles d'apprentissage automatique (ML) et d'apprentissage profond (DL), afin d'exploiter la diversité des points de vue dans la modélisation de la fonction de production. En effet, nous avons proposé un langage pour décrire les composantes conceptuelles de cette solution en tenant compte de l'exigence de la qualité de service. Nous avons indiqué les composantes de MonTransBDI qui pourraient être considérées dans la conception d'un système de transformation de modèle collaborative. La logique de résolution pour obtenir un modèle cible est construite sur la base des paramètres des composants obtenus. Nos résultats montrent que notre composante peut couvrir les paramètres qui impactent la transformation dans un contexte de la transformation des modèles.

Dans la deuxième contribution, nous avons démontré comment instancier notre framework en spécialisant l'algorithme de recherche. Ainsi, nous avons proposé d'appliquer un algorithme génétique. Cet algorithme génétique permet d'évoluer et de reproduire des solutions même en l'absence d'un exemple identique dans la base de données d'exemples.

Pour illustrer cette approche, nous avons effectué une instance spécifique à la production d'olives dans le nord de l'Algérie. En tenant compte des particularités régionales (climat, type de sol, variétés d'oliviers, etc.), l'algorithme génétique a été adapté pour optimiser la prédiction des fonctions de production en combinant des paramètres tels que les données climatiques, les rendements passés, et les contraintes liées à la qualité du produit.

L'algorithme a été implémenté sous forme d'un outil support, permettant de générer des fonctions de production adaptées aux spécificités locales et contribuant ainsi à une meilleure prise de décision pour les agriculteurs et les responsables agricoles.

6.1.3 Evaluation et Discussion

Nous avons évalué notre modèle sur un cas d'étude concernant la production d'olives dans le nord de l'Algérie. Nous avons d'abord comparé les fonctions de production générées par les modèles ARIMA et les modèles de machine learning (ML) et deep learning (DL). Ensuite, nous avons évalué les attributs de qualité du modèle développé en menant plusieurs expériences pour évaluer ses capacités de précision.

Les résultats ont démontré que notre modèle prédit avec précision les performances de la requête, permettant ainsi aux concepteurs de faire moins d'hypothèses durant la phase de conception. Ce succès est attribué aux directives de conception complètes de notre framework.

En résumé, les modèles avancés de machine learning, en particulier Multi-head Attention et les innovations récentes, surpassent largement les méthodes traditionnelles dans la prévision de la pro-

duction d'olives. Le modèle Multi-head Attention présente constamment le plus bas RMSE, soulignant son efficacité à capturer les schémas complexes de rendement et mettant en évidence la valeur des techniques modernes dans les prévisions agricoles.

Nous avons évalué divers modèles de machine learning, y compris ARIMA, RNN, LSTM, GRU, et les architectures Transformer, pour prédire la production d'olives à travers différents clusters. Les résultats révèlent que les modèles avancés, comme Multi-head Attention et les innovations récentes, surpassent les méthodes traditionnelles, offrant une précision supérieure en RMSE dans tous les clusters. Notamment, le modèle Multi-head Attention atteint systématiquement le RMSE le plus bas, ce qui indique son efficacité à capturer les schémas complexes dans les données de rendement des olives. Ces résultats soulignent le potentiel des modèles de machine learning à la pointe de la technologie pour améliorer les prévisions agricoles.

En conclusion, le cadre AFFAC propose une approche innovante pour la prévision des cultures d'olives, en utilisant des modèles d'apprentissage profond basés sur des transformateurs. Il améliore les modèles traditionnels en collectant, gérant et analysant les données agricoles, réduisant ainsi les erreurs de prédiction et identifiant les facteurs clés des coûts de production. Le succès du cadre se reflète dans son processus de conception, qui inclut l'expression des exigences, l'identification des dimensions et des facteurs impactant les coûts, la construction du modèle de prévision, ainsi que les tests et l'évaluation. Les tests de la fonction $\mathcal{P}rod$ confirment sa capacité à prédire avec précision la performance des requêtes, réduisant les hypothèses des concepteurs grâce à des directives de conception solides. Les modèles d'apprentissage automatique avancés surpassent les méthodes traditionnelles, obtenant des RMSE plus bas et capturant efficacement les motifs complexes de rendement, ce qui renforce la précision des prévisions agricoles et soutient des solutions durables.

6.2 Perspectives

Les travaux réalisés dans le cadre de cette thèse ouvrent plusieurs perspectives de recherche à court et à long terme.

6.2.1 Intégration de la détection d'anomalies

D'un point de vue théorique, une piste intéressante consiste à intégrer la détection d'anomalies dans les données provenant des capteurs. Bien que le contexte soit similaire à celui de notre étude de cas, les données peuvent être suffisamment spécifiques pour permettre l'application de méthodes particulières qui ne seraient peut-être pas généralisables à d'autres types de systèmes agricoles. Cette approche pourrait permettre d'identifier des comportements inhabituels dans les données de production agricole, ce qui offrirait une opportunité d'améliorer la gestion des processus de production en détectant des anomalies en temps réel.

6.2.2 Réduction de la complexité du modèle pour améliorer la réactivité

En parallèle, nous prévoyons d'étudier comment notre cadre peut réduire la complexité des prédictions tout en améliorant la rapidité de la réponse du modèle. Cette approche vise à optimiser l'usage des

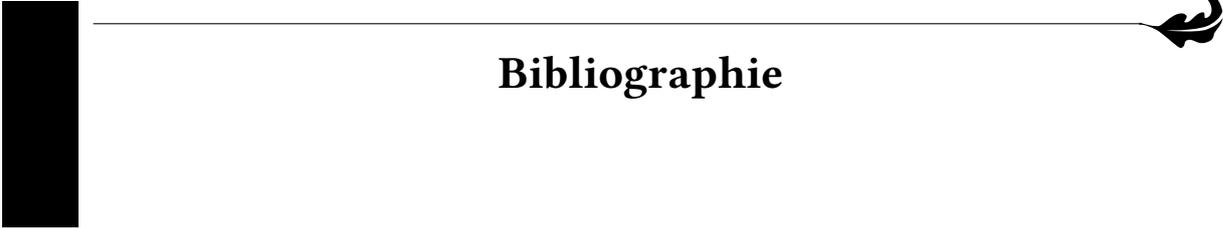
ressources computationnelles et à accélérer la prise de décision, ce qui est crucial dans des contextes où la réactivité est primordiale. En allégeant le processus de calcul, nous pourrions améliorer l'efficacité du système, notamment dans les applications nécessitant des prévisions en temps réel.

6.2.3 Attribution d'un score de confiance aux mesures des capteurs

D'un point de vue méthodologique, une autre voie de recherche concerne l'attribution d'un score de confiance aux mesures issues des capteurs montés sur les équipements industriels, dans le but d'utiliser ces mesures pour des fins prédictives ou diagnostiques. Cela pourrait inclure la prévision de défaillances, la détection ou l'étude de comportements spécifiques, comme les variations dans les rendements agricoles en fonction des conditions climatiques ou des paramètres environnementaux. L'intégration de scores de confiance permettrait d'améliorer la fiabilité des décisions prises à partir de ces données et de mieux gérer les incertitudes liées aux capteurs.

6.2.4 Développement d'une application mobile pour l'invocation du modèle

Du point de vue pratique, nous avons l'intention de développer une application mobile permettant d'invoquer le modèle et de recevoir des recommandations basées sur les entrées des utilisateurs. Cette application offrirait une interface conviviale pour les agriculteurs et les décideurs afin qu'ils puissent accéder aux prévisions de production et aux conseils en fonction de paramètres spécifiques, comme les prévisions climatiques ou les conditions de sol. Cela pourrait grandement améliorer l'accessibilité et l'utilité du modèle développé, en facilitant son adoption sur le terrain par des utilisateurs non experts.



Bibliographie

- [1] P. ADAIR et al. « L'agriculture algérienne : l'héritage du passé et les défis contemporains ». In : *les cahiers du cread* 38.3 (2022), p. 413–440 (cf. p. 108).
- [2] F. E. A. ADLOFF. « Mediterranean Sea response to climate change in an ensemble of twenty first century scenarios ». In : *Climatic Change* 130 (2015), p. 241–258 (cf. p. 37).
- [3] A. AHMAD et al. « Understanding the Agricultural Systems Modeling : A Review ». In : *International Journal* 10.3 (2021) (cf. p. 52).
- [4] I. AHMED et P. K. YADAV. « Predicting Apple Plant Diseases in Orchards Using Machine Learning and Deep Learning Algorithms ». In : *SN Computer Science* 5.6 (2024), p. 1–17 (cf. p. 8).
- [5] C. D. AHRENS. *Meteorology Today : An Introduction to Weather, Climate, and the Environment*. Brooks/Cole Publishing Company, 2000 (cf. p. 21).
- [6] C. D. AHRENS. *Meteorology today : an introduction to weather, climate, and the environment*. Cengage Learning Canada Inc, 2015 (cf. p. 19, 20).
- [7] S. ALI et al. « Interactive effects of planting models with limited irrigation on soil water, temperature, respiration and winter wheat production under simulated rainfall conditions ». In : *Agricultural Water Management* 204 (2018), p. 198–211 (cf. p. 51).
- [8] K. M. AMIR et al. « Integrating TRIZ and MCDM for innovative and sustainable decision-making : a case study in the life cycle of olive oil ». In : *International TRIZ Future Conference*. Springer. 2023, p. 226–236 (cf. p. 72).
- [9] C. R. ANDERSON et al. *Agroecology now ! : Transformations towards more just and sustainable food systems*. Springer Nature, 2021 (cf. p. 57).
- [10] L. ATTALLAOUI. « La filière oléicole en Algérie, état des lieux, opportunités et stratégies d'acteurs ; cas de la wilaya Djelfa ». Thèse de doct. UNIVERSITE MOHAMED KHIDER BISKRA, 2022 (cf. p. 109, 110).
- [11] S. ATTARAN DOWOM, S. REZAEIAN et H. R. POURIANFAR. « Agronomic and environmental factors affecting cultivation of the winter mushroom or Enokitake : Achievements and prospects ». In : *Applied microbiology and biotechnology* 103.6 (2019), p. 2469–2481 (cf. p. 46, 81).
- [12] V. BALAFAS et al. « Machine learning and deep learning for plant disease classification and detection ». In : *IEEE Access* (2023) (cf. p. 8).
- [13] H. W. BARKER et al. « Assessing 1D atmospheric solar radiative transfer models : Interpretation and handling of unresolved clouds ». In : *Journal of Climate* 16.16 (2003), p. 2676–2699 (cf. p. 19, 21).
- [14] R. G. BARRY et R. J. CHORLEY. *Atmosphere, weather and climate*. Routledge, 2009 (cf. p. 19).
- [15] R. G. BARRY et R. J. CHORLEY. *Atmosphere, Weather, and Climate*. 9th. Routledge, 2010 (cf. p. 21).
- [16] M. B. BECK. « Water quality modeling : A review of the analysis of uncertainty ». In : *Water Resources Research* 23.8 (1987), p. 1393–1442 (cf. p. 49).

- [17] A. BEN DHIAB et al. « Modeling olive-crop forecasting in Tunisia ». In : *Theoretical and applied climatology* 128 (2020), p. 541–549 (cf. p. 8).
- [18] L. N. BEN YELLES. « L’arboriculture fruitière dans la région de sabra : état des lieux et perspective pour son développement. » Thèse de doct. 2022 (cf. p. 110).
- [19] M. BENISTON et D. B. STEPHENSON. « Extreme climatic events and their evolution under changing climatic conditions ». In : *Global and Planetary Change* 44.1-4 (2006), p. 1–9 (cf. p. 24).
- [20] K. J. BEVEN. *Rainfall-Runoff Modelling : The Primer*. 2nd. Wiley-Blackwell, 2012. ISBN : 978-0470714591 (cf. p. 47).
- [21] E. BORSATO, P. TAROLLI et F. MARINELLO. « Sustainable patterns of main agricultural products combining different footprint parameters ». In : *Journal of cleaner production* 179 (2018), p. 357–367 (cf. p. 82).
- [22] G. E. P. BOX et al. *Time Series Analysis : Forecasting and Control*. 5th. Wiley, 2015. ISBN : 978-1118675021 (cf. p. 48).
- [23] G. E. BOX et al. *Time series analysis : forecasting and control*. John Wiley & Sons, 2015 (cf. p. 60, 87).
- [24] N. C. BRADY, R. R. WEIL et R. R. WEIL. *The nature and properties of soils*. T. 13. Prentice Hall Upper Saddle River, NJ, 2008 (cf. p. 53).
- [25] G. P. BRASSEUR et al. « Klimawandel in Deutschland : Einführung ». In : *Klimawandel in Deutschland : Entwicklung, Folgen, Risiken und Perspektiven*. Springer, 2017, p. 1–4 (cf. p. 18).
- [26] J. C. BROWN et al. « Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data ». In : *Remote Sensing of Environment* 130 (2013), p. 39–50 (cf. p. 56).
- [27] C. E. BUELL. « On the physical interpretation of empirical orthogonal functions. » In : *Preprints Sixth Conf. on Prob. and Stats. in Atmos. Sci. Bauff. Alta., Amer. Meteor. Soc.* 112. 1979 (cf. p. 107).
- [28] A. H. BUTLER, L. M. POLVANI et C. DESER. « Separating the stratospheric and tropospheric pathways of El Niño–Southern Oscillation teleconnections ». In : *Environmental Research Letters* 9.2 (2014), p. 024014 (cf. p. 57).
- [29] S. ÇAKMAKÇI et R. ÇAKMAKÇI. « Quality and nutritional parameters of food in agri-food production systems ». In : *Foods* 12.2 (2023), p. 351 (cf. p. 81).
- [30] L. S. CEDRIC et al. « Crops yield prediction based on machine learning models : Case of West African countries ». In : *Smart Agricultural Technology* 2 (2022), p. 100049 (cf. p. 154, 156–158).
- [31] S. C. CHAPRA. *Surface Water-Quality Modeling*. McGraw-Hill, 1997. ISBN : 978-0070113643 (cf. p. 49).
- [32] H. CHEN, X. LI et S. XU. *Data Storage and Management Technologies in Cloud Computing*. Springer, 2021. ISBN : 9783030684415 (cf. p. 40).
- [33] P. CHEN et al. « Pathformer : Multi-scale transformers with adaptive pathways for time series forecasting ». In : *arXiv preprint arXiv :2402.05956* (2024) (cf. p. 68).
- [34] N. CHINDARKAR. « Gender and climate change-induced migration : proposing a framework for analysis ». In : *Environmental Research Letters* 7.2 (2012), p. 025601 (cf. p. 55).
- [35] A. CHLINGARYAN, S. SUKKAR et A. RANI. « A review of machine learning in agricultural yield prediction ». In : *Computers and Electronics in Agriculture* 151 (2018), p. 16–32. DOI : [10.1016/j.compag.2018.06.021](https://doi.org/10.1016/j.compag.2018.06.021) (cf. p. 45).
- [36] K. CHO et al. « On the properties of neural machine translation : Encoder-decoder approaches ». In : *arXiv preprint arXiv :1409.1259* (2014) (cf. p. 63, 89).
- [37] A. CHOLLEY. « Emmanuel de Martonne ». In : *Annales de géographie*. T. 65. 347. Persée-Portail des revues scientifiques en SHS. 1956, p. 1–14 (cf. p. 101).

- [38] P. CHRISTIAS et M. MOCANU. « A machine learning framework for olive farms profit prediction ». In : *Water* 13.23 (2021), p. 3461 (cf. p. 46).
- [39] I. P. ON CLIMATE CHANGE (IPCC). *Climate Change 2021 : The Physical Science Basis*. 2021. URL : <https://www.ipcc.ch/report/ar6/wg1/> (cf. p. 35–37).
- [40] W. COLLINS, R. KNUTTI et J. E. A. ARBLASTER. « Long-term Climate Change : Projections, Commitments and Irreversibility ». In : *Earth's Climate System (ECS)* 6 (2019), p. 1023–1054. URL : <https://www.cambridge.org/core> (cf. p. 36, 37).
- [41] S. DAVIS et AL. « Impact of Climatic Variables on Crop Yield : A Machine Learning Approach ». In : *Field Crops Research* 210 (2017), p. 34–45 (cf. p. 52).
- [42] M. DEGHANI et al. « Universal transformers ». In : *arXiv preprint arXiv :1807.03819* (2018) (cf. p. 64).
- [43] J. DEVLIN. « Bert : Pre-training of deep bidirectional transformers for language understanding ». In : *arXiv preprint arXiv :1810.04805* (2018) (cf. p. 57, 67).
- [44] F. E. A. DOBLAS-REYES. « The impact of climate change on Mediterranean agriculture ». In : *Environmental Research Letters* 13.11 (2018), p. 113002 (cf. p. 37).
- [45] D. I. DOMEISEN, C. I. GARFINKEL et A. H. BUTLER. « The teleconnection of El Niño Southern Oscillation to the stratosphere ». In : *Reviews of Geophysics* 57.1 (2019), p. 5–47 (cf. p. 57).
- [46] L. DONG, S. XU et B. XU. « Speech-transformer : a no-recurrence sequence-to-sequence model for speech recognition ». In : *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2018, p. 5884–5888 (cf. p. 67).
- [47] P. DROBINSKI, S. SOMOT, A. ORR et al. « Modelling the Mediterranean climate system ». In : *Progress in Oceanography* 124 (2014), p. 1–27 (cf. p. 24).
- [48] J. DURBIN et S. J. KOOPMAN. *Time Series Analysis by State Space Methods*. 2nd. Oxford University Press, 2012. ISBN : 978-0199641178 (cf. p. 48).
- [49] Y. EGI, M. HAJYZADEH et E. EYCEYURT. *Drone-Computer Communication Based Tomato Generative Organ Counting Model Using YOLO V5 and Deep-Sort*. *Agriculture* 2022, 12, 1290. 2022 (cf. p. 155, 156, 158).
- [50] I. FERNÁNDEZ GARCÍA et al. « Trends and challenges in irrigation scheduling in the semi-arid area of Spain ». In : *Water* 12.3 (2020), p. 785 (cf. p. 51).
- [51] M. FORNACIARI, F. ORLANDI et B. ROMANO. « Yield forecasting for olive trees : a new approach in a historical series (Umbria, Central Italy) ». In : *Agronomy Journal* 97.6 (2005), p. 1537–1542 (cf. p. 9).
- [52] Y. FOUQUART. *Climat et environnement*. Ellipses, 2003 (cf. p. 22).
- [53] R. L. DE FREITAS CUNHA et B. SILVA. « Estimating crop yields with remote sensing and deep learning ». In : *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*. IEEE. 2020, p. 273–278 (cf. p. 52).
- [54] C. GALÁN et al. « Forecasting olive (*Olea europaea*) crop yield based on pollen emission ». In : *Field Crops Research* 86.1 (2004), p. 43–51 (cf. p. 9).
- [55] A. GELMAN et al. *Bayesian Data Analysis*. 3rd. Chapman et Hall/CRC, 2013. ISBN : 978-1439840955 (cf. p. 48).
- [56] M. GHIL et V. LUCARINI. « The physics of climate variability and climate change ». In : *Reviews of Modern Physics* 92.3 (2020), p. 035002 (cf. p. 18, 19, 21).
- [57] C. GIANNAKOPOULOS et al. « Climatic changes and associated impacts in the Mediterranean resulting from a 2 °C global warming ». In : *Global and Planetary Change* 68.3 (2005), p. 209–224 (cf. p. 24).
- [58] F. GIORGI et L. O. MEARNs. « Introduction to the special section : Regional climate modeling revisited ». In : *Journal of Geophysical Research : Atmospheres* 114.D11 (2009), p. D11 (cf. p. 36, 37).

- [59] D. GOMEZ et al. « Advancing common bean (*Phaseolus vulgaris* L.) disease detection with YOLO driven deep learning to enhance agricultural AI ». In : *Scientific Reports* 14.1 (2024), p. 15596 (cf. p. 155, 158).
- [60] I GOODFELLOW. *Deep Learning. I. Goodfellow, Y. Bengio, A. Courville*. 2016 (cf. p. 56, 61, 88).
- [61] I. GOODFELLOW et Y. BENGIO. « Aaron Courville ». In : *Deep learning* 11 (2016), p. 18 (cf. p. 56).
- [62] P. J. GREGORY, J. S. INGRAM et M. BRKLACICH. « Climate change and food security ». In : *Philosophical Transactions of the Royal Society B : Biological Sciences* 360.1463 (2005), p. 2139–2148. DOI : [10.1098/rstb.2005.1745](https://doi.org/10.1098/rstb.2005.1745) (cf. p. 45).
- [63] H. V. GUPTA et al. « Decomposition of the mean squared error and NSE performance criteria : Implications for improving hydrological modelling ». In : *Journal of hydrology* 377.1-2 (2009), p. 80–91 (cf. p. 56).
- [64] J. A. HARTIGAN et M. A. WONG. « Algorithm AS 136 : A k-means clustering algorithm ». In : *Journal of the royal statistical society. series c (applied statistics)* 28.1 (1979), p. 100–108 (cf. p. 60, 87).
- [65] A HARTLEY, R JONES et T JANES. « Projections relatives aux changements des services écosystémiques face au changement climatique ». In : *UNEP-WCMC Technical Report* (2015) (cf. p. 51).
- [66] T. HASTIE. *The elements of statistical learning : data mining, inference, and prediction*. 2009 (cf. p. 52).
- [67] K. W. HIPEL et A. I. MCLEOD. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, 2005. ISBN : 978-0444507574 (cf. p. 48).
- [68] S. HOCHREITER et J. SCHMIDHUBER. « Long short-term memory ». In : *Neural computation* 9.8 (1997), p. 1735–1780 (cf. p. 62, 88).
- [69] S. C. A. HOUETOHOSSOU et al. « Deep learning methods for biotic and abiotic stresses detection and classification in fruits and vegetables : State of the art and perspectives ». In : *Artificial Intelligence in Agriculture* (2023) (cf. p. 8).
- [70] W. K. HÄRDLE et L. SIMAR. *Applied Multivariate Statistical Analysis*. 4th. Springer, 2019. ISBN : 978-3662585023 (cf. p. 48).
- [71] W. W. HSIEH. *Machine Learning Methods in the Environmental Sciences : Neural Networks and Kernels*. Cambridge University Press, 2009. ISBN : 978-0521791922 (cf. p. 48).
- [72] F. M. HUGHES. « Floodplain biogeomorphology ». In : *Progress in physical geography* 21.4 (1997), p. 501–529 (cf. p. 54).
- [73] M. HULME. « Why we disagree about climate change ». In : *Zygon* 50.4 (2015) (cf. p. 18).
- [74] C. HUNTINGFORD et al. « Machine learning and artificial intelligence to aid climate change research and preparedness ». In : *Environmental Research Letters* 14.12 (2019), p. 124007. DOI : [10.1088/1748-9326/ab4e55](https://doi.org/10.1088/1748-9326/ab4e55) (cf. p. 47).
- [75] Y.-S. HWANG et K. M. REYNOLDS. « Economic evaluation of mitigation strategies for climate change : A case study ». In : *Environmental Economics and Policy Studies* 17.4 (2015), p. 573–592. DOI : [10.1007/s10018-015-0108-2](https://doi.org/10.1007/s10018-015-0108-2) (cf. p. 50).
- [76] R. J. HYNDMAN et G. ATHANASOPOULOS. *Forecasting : principles and practice*. OTexts, 2018 (cf. p. 60).
- [77] S INIYAN, V. A. VARMA et C. T. NAIDU. « Crop yield prediction using machine learning techniques ». In : *Advances in Engineering Software* 175 (2023), p. 103326 (cf. p. 71).
- [78] IPCC. *Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Available at : <https://www.ipcc.ch/report/ar6/wg1/>. 2021 (cf. p. 21).
- [79] A. H. IVALE, I. DE ALENCAR NÃÃS et M. DE CAMARGO JANI. « A Multi-Method Approach to Assess the Adoption of Precision Agriculture Technology in Brazil. » In : *Agris On-Line Papers in Economics & Informatics* 16.3 (2024) (cf. p. 72).

- [80] A. K. JAIN. « Data clustering : 50 years beyond K-means ». In : *Pattern recognition letters* 31.8 (2010), p. 651–666 (cf. p. 58).
- [81] A. K. JAIN et R. C. DUBES. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988 (cf. p. 59, 86).
- [82] A. K. JAIN, M. N. MURTY et P. J. FLYNN. « Data clustering : a review ». In : *ACM computing surveys (CSUR)* 31.3 (1999), p. 264–323 (cf. p. 58).
- [83] Y. JAME et H. CUTFORTH. « Crop growth models for decision support systems ». In : *Canadian Journal of Plant Science* 76.1 (1996), p. 9–19 (cf. p. 39).
- [84] E. JEROME et al. « Etablissement du bilan de carbone d’une exploitation agricole wallonne pratiquant le système allaitant : effets du climat et de la gestion du pâturage. Rapport de synthèse. Janvier 2012–Décembre 2013. » In : (2013) (cf. p. 19).
- [85] J. W. JONES, G. HOOGENBOOM, C. H. PORTER et al. « The DSSAT cropping system model ». In : *European Journal of Agronomy* 18.3-4 (2003), p. 235–265 (cf. p. 49, 54).
- [86] M. JONES et R. L. DAVIS. « Yield Prediction Models Incorporating Environmental and Historical Data ». In : *Computers and Electronics in Agriculture* 135 (2017), p. 120–130. DOI : [10.1016/j.compag.2017.02.011](https://doi.org/10.1016/j.compag.2017.02.011) (cf. p. 51).
- [87] P. A. JONES et L. B. SMITH. « Crop Yield Prediction Using Mechanistic Models ». In : *Agricultural Systems* 123 (2014), p. 12–25. DOI : [10.1016/j.agsy.2014.05.001](https://doi.org/10.1016/j.agsy.2014.05.001) (cf. p. 50).
- [88] P. D. JONES, M. HULME et K. R. BRIFFA. « Surface air temperature changes : A review of the data ». In : *Reviews of Geophysics* 37.2 (1999), p. 173–199 (cf. p. 24).
- [89] P. KAIROUZ et al. « Advances and open problems in federated learning ». In : *Foundations and trends® in machine learning* 14.1–2 (2021), p. 1–210 (cf. p. 72).
- [90] D. A. KANE. « Does Soil Carbon support Climate Resilient Agricultural Systems ? Searching for Evidence and Developing New Measurement Tools ». Thèse de doct. Yale University, 2022 (cf. p. 51).
- [91] P. KOUTROUMPIS et al. « The internet of things in agriculture : A survey ». In : *Computers and Electronics in Agriculture* 114 (2015), p. 80–88 (cf. p. 72).
- [92] M KUHN. *Applied predictive modeling*. 2013 (cf. p. 52, 53).
- [93] R. KUMAR et N. SHARMA. « Predicting Crop Yields Using Neural Networks Based on Climate and Environmental Factors ». In : *Agricultural Systems* 103.2 (2019), p. 134–145. DOI : [10.1016/j.agsy.2019.05.006](https://doi.org/10.1016/j.agsy.2019.05.006) (cf. p. 51).
- [94] P. KUMARI et al. « Recurrent neural network architecture for forecasting banana prices in Gujarat, India ». In : *Plos one* 18.6 (2023), e0275702 (cf. p. 155, 158).
- [95] S. KURTOĞLU, A. S. UZUNDUMLU et E. GÖVEZ. « Olive Oil Production Forecasts for a Macro Perspective during 2024–2027 ». In : *Applied Fruit Science* 66.3 (2024), p. 1089–1100 (cf. p. 153, 154, 156, 157).
- [96] B. LAZREG. « Impact du changement climatique sur la production agricole en Algérie ». Thèse de doct. Université IBN-KHALDOUN-Tiaret-Faculté des Sciences de la Nature et de la Vie, 2016 (cf. p. 103).
- [97] Y. LECUN, Y. BENGIO et G. HINTON. « Deep learning ». In : *nature* 521.7553 (2015), p. 436–444 (cf. p. 71).
- [98] M. LEE et K. P. JOHNSON. « Hybrid Models for Crop Yield Prediction : Combining Mechanistic and Statistical Approaches ». In : *Agricultural Systems* 178 (2019), p. 102–115. DOI : [10.1016/j.agsy.2019.07.003](https://doi.org/10.1016/j.agsy.2019.07.003) (cf. p. 50, 51).
- [99] H. LI, Y. ZHAO et F. ZHENG. « The framework of an agricultural land-use decision support system based on ecological environmental constraints ». In : *Science of the total environment* 717 (2020), p. 137149 (cf. p. 71).

- [100] T. LI et al. « Federated learning : Challenges, methods, and future directions ». In : *IEEE signal processing magazine* 37.3 (2020), p. 50–60 (cf. p. 72).
- [101] Y. LI et al. « Transformer-Modulated Diffusion Models for Probabilistic Multivariate Time Series Forecasting ». In : *The Twelfth International Conference on Learning Representations* (cf. p. 68).
- [102] P. LIONELLO, P. MALANOTTE-RIZZOLI et R. BOSCOLO. *The Climate of the Mediterranean Region : From the Past to the Future*. Amsterdam : Elsevier, 2006 (cf. p. 23, 24).
- [103] G. LIU et al. « A state of art review on time series forecasting with machine learning for environmental parameters in agricultural greenhouses ». In : *Information Processing in Agriculture* (2022) (cf. p. 153, 156).
- [104] Y. LIU, L. ZHANG et J. YANG. « A review of machine learning applications in precision agriculture ». In : *Computers, Environment and Urban Systems* 77 (2020), p. 101350. DOI : [10.1016/j.compenvurbsys.2019.101350](https://doi.org/10.1016/j.compenvurbsys.2019.101350) (cf. p. 45).
- [105] Y. LIU et al. « itransformer : Inverted transformers are effective for time series forecasting ». In : *arXiv preprint arXiv :2310.06625* (2023) (cf. p. 68).
- [106] Z. LIU et al. « Swin transformer : Hierarchical vision transformer using shifted windows ». In : *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, p. 10012–10022 (cf. p. 67).
- [107] Z. LIU et AL. « Data Cleaning and Homogeneity Testing for Agricultural Models ». In : *Agricultural Data Science* 34 (2017), p. 122–135 (cf. p. 52).
- [108] S. LLOYD. « Least squares quantization in PCM ». In : *IEEE transactions on information theory* 28.2 (1982), p. 129–137 (cf. p. 59, 86).
- [109] D. B. LOBELL, W. SCHLENKER et J. COSTA-ROBERTS. « Climate change and crop production : Impacts and adaptation ». In : *Environmental Research Letters* 6.3 (2011), p. 034012. DOI : [10.1088/1748-9326/6/3/034012](https://doi.org/10.1088/1748-9326/6/3/034012) (cf. p. 45).
- [110] D. B. LOBELL et al. « Prioritizing climate change adaptation needs for food security in 2030 ». In : *Science* 319.5863 (2008), p. 607–610. DOI : [10.1126/science.1152339](https://doi.org/10.1126/science.1152339) (cf. p. 47).
- [111] C. M. LOPES et A. MONTEIRO. « Modeling Olive Tree Growth and Yield Under Mediterranean Climate Conditions ». In : *Agricultural and Forest Meteorology* 237 (2017), p. 104–115. DOI : [10.1016/j.agrformet.2017.01.004](https://doi.org/10.1016/j.agrformet.2017.01.004) (cf. p. 51).
- [112] C. MARTIN et M. LEGRET. « La méthode multicritère ELECTRE III : définitions, principe et exemple d’application à la gestion des eaux pluviales en milieu urbain ». In : *Bulletin des laboratoires des ponts et chaussées* 258-259 (2005) (cf. p. 94).
- [113] V. MASSON-DELMOTTE. « Réchauffement climatique : état des connaissances scientifiques, enjeux, risques et options d’action ». In : *Comptes Rendus. Géoscience* 352.4-5 (2020), p. 251–277 (cf. p. 18).
- [114] A. McBRATNEY et al. « Future directions of precision agriculture ». In : *Precision agriculture* 6 (2005), p. 7–23 (cf. p. 53).
- [115] J. J. McCARTHY et al. « Impacts, Adaptation and vulnerability ». In : *Third Assessment Report of the Intergovernmental panel on climate change, working Group 2* (2001), p. 85 (cf. p. 100).
- [116] D. SALAS Y MÉLIA. « Les principaux enseignements du 6 e rapport du groupe I du GIEC ». In : *Annales des Mines-Responsabilité et environnement*. 2. Cairn/Softwin. 2022, p. 11–16 (cf. p. 18).
- [117] M MENDIL et A SEBAI. « Catalogue des variétés Algériennes de l’olivier ». In : *Ministere de l’agriculture et du développement rural, ITAF Alger, Algeria* 98 (2006) (cf. p. 108).
- [118] C. MILLOT. « Circulation in the western Mediterranean Sea ». In : *Journal of Marine Systems* 20.1-4 (1999), p. 423–442 (cf. p. 23).

- [119] MINISTÈRE DE L'AGRICULTURE ET DU DÉVELOPPEMENT RURAL. *Rapport sur la variabilité climatique et la production agricole*. Consulté en 2021. 2021 (cf. p. 128).
- [120] E. S. MOHAMED et al. « Smart farming for improving agricultural management ». In : *The Egyptian Journal of Remote Sensing and Space Science* 24.3 (2021), p. 971–981 (cf. p. 8).
- [121] M. S. MOHAMMED, M. AL-BADAWI et R. AL-AHMAD. « Recent advances in machine learning for crop yield prediction ». In : *Computers and Electronics in Agriculture* 160 (2019), p. 22–39. DOI : [10.1016/j.compag.2019.02.006](https://doi.org/10.1016/j.compag.2019.02.006) (cf. p. 45).
- [122] S. E. NICHOLSON. *Dryland Climatology*. Cambridge : Cambridge University Press, 2011 (cf. p. 23, 24).
- [123] A. OIKONOMIDIS, C. CATAL et A. KASSAHUN. « Hybrid deep learning-based models for crop yield prediction ». In : *Applied artificial intelligence* 36.1 (2022), p. 2031822 (cf. p. 154, 157).
- [124] S. A. OSINGA et al. « Big data in agriculture : Between opportunity and solution ». In : *Agricultural Systems* 195 (2022), p. 103298 (cf. p. 71).
- [125] C. PARMESAN et G. YOHE. « A globally coherent fingerprint of climate change impacts across natural systems ». In : *Nature* 421.6918 (2003), p. 37–42 (cf. p. 28).
- [126] M. L. PARRY et al. *Climate Change 2007 : Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK : Cambridge University Press, 2007. ISBN : 978-0-521-70597-4 (cf. p. 50).
- [127] D. B. PERCIVAL et A. T. WALDEN. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 2006. ISBN : 978-0521685085 (cf. p. 48).
- [128] R. PIELKE. « Mesoscale Meteorological Modeling Academic ». In : *San Diego, CA* (2002) (cf. p. 18).
- [129] M. I. RAMOS et al. « Improving early prediction of crop yield in Spanish olive groves using satellite imagery and machine learning ». In : *PloS one* 20.1 (2025), e0311530 (cf. p. 154, 157).
- [130] M. RASHEDI et al. « Machine learning-based model predictive controller design for cell culture processes ». In : *Biotechnology and Bioengineering* 120.8 (2023), p. 2144–2159 (cf. p. 46).
- [131] C. F. ROPELEWSKI et M. S. HALPERT. *El Niño and its global impact*. Geneva : World Meteorological Organization, 2009 (cf. p. 24).
- [132] C. ROSENZWEIG et al. « Assessment of observed changes and responses in natural and managed systems ». In : (2007) (cf. p. 55).
- [133] E. ROSSI et al. « Temporal graph networks for deep learning on dynamic graphs ». In : *arXiv preprint arXiv :2006.10637* (2020) (cf. p. 65).
- [134] B. ROY. *Multicriteria methodology for decision aiding*. T. 12. Springer Science & Business Media, 2013 (cf. p. 72).
- [135] W. E. A. SAMUEL. « Projected precipitation changes in the Mediterranean Basin under global warming ». In : *Climate Dynamics* 54 (2020), p. 2751–2767 (cf. p. 36).
- [136] U. SHAFI et al. « Precision agriculture techniques and practices : From considerations to applications ». In : *Sensors* 19.17 (2019), p. 3796 (cf. p. 71).
- [137] M. Y. SHAMS, S. A. GAMEL et F. M. TALAAT. « Enhancing crop recommendation systems with explainable artificial intelligence : a study on agricultural decision-making ». In : *Neural Computing and Applications* 36.11 (2024), p. 5695–5714 (cf. p. 155, 156, 158).
- [138] P. SHAW, J. USZKOREIT et A. VASWANI. « Self-attention with relative position representations ». In : *arXiv preprint arXiv :1803.02155* (2018) (cf. p. 65).

- [139] T. SIEW-YAN-YU et al. « Régionalisation du régime des précipitations dans la région des Bois-francs et de l'Estrie par l'analyse en composantes principales ». In : *Canadian Journal of Civil Engineering* 25.6 (1998), p. 1050–1058 (cf. p. 104).
- [140] R. SINGH et AL. « Imputation of Missing Data in Agricultural Datasets Using Machine Learning Techniques ». In : *Journal of Data Science and Analytics* 14 (2018), p. 57–70 (cf. p. 52).
- [141] B. J. SKINNER et S. C. PORTER. « The blue planet : an introduction to earth system science ». In : *New York : John Wiley* (1995) (cf. p. 22).
- [142] J. SMITH et R. PATEL. « Techniques for data aggregation and storage in IoT systems for real-time applications ». In : *Journal of Cloud Computing* 8 (2019), p. 33. DOI : [10.1186/s13677-019-0167-5](https://doi.org/10.1186/s13677-019-0167-5) (cf. p. 40).
- [143] J. D. SMITH. *Principles of Environmental Modeling*. Springer, 2012. ISBN : 978-3-642-23456-7 (cf. p. 50).
- [144] J. D. SMITH et S. L. TAYLOR. « Hybrid Climate Models : Integrating Observational Data and Physical Models ». In : *Journal of Climate Research* 34.2 (2020), p. 123–140. DOI : [10.1016/j.jcr.2020.05.001](https://doi.org/10.1016/j.jcr.2020.05.001) (cf. p. 50, 51, 55).
- [145] S. TAIBI et al. « Évolution et régionalisation des précipitations au nord de l'Algérie (1936–2009) ». In : *Climate and land surface changes in hydrology, IAHS Publ* 359 (2013), p. 191–197 (cf. p. 104).
- [146] A. M. K. TANK, F. W. ZWIERS et X. ZHANG. « Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation ». In : *World Meteorological Organization (WMO)* 1500 (2009), p. 1–52 (cf. p. 28, 35).
- [147] K. E. TRENBERTH. *Climate System Modeling*. Cambridge University Press, 1992 (cf. p. 21).
- [148] R. M. TRIGO, T. J. OSBORN et J. M. CORTE-REAL. « The North Atlantic Oscillation influence on Europe : climate impacts and associated physical mechanisms ». In : *Climate Research* 30.1 (2006), p. 27–41 (cf. p. 24).
- [149] T. VAN KLOMPENBURG, A. KASSAHUN et C. CATAL. « Crop yield prediction using machine learning : A systematic literature review ». In : *Computers and electronics in agriculture* 177 (2020), p. 105709 (cf. p. 71).
- [150] M. VASILEIOU et al. « Transforming weed management in sustainable agriculture with artificial intelligence : A systematic literature review towards weed identification and deep learning ». In : *Crop Protection* (2023), p. 106522 (cf. p. 8).
- [151] A. VASWANI. « Attention is all you need ». In : *Advances in Neural Information Processing Systems* (2017) (cf. p. 57, 65, 67, 89).
- [152] X. L. WANG et al. « New techniques for the detection and adjustment of shifts in daily precipitation data series ». In : *Journal of Applied Meteorology and Climatology* 49.12 (2010), p. 2416–2436 (cf. p. 104).
- [153] X. WANG et al. « CARD : Channel aligned robust blend transformer for time series forecasting ». In : *The Twelfth International Conference on Learning Representations*. 2024 (cf. p. 67).
- [154] Z. WANG, Q. CAO et Y. SHEN. « Modeling light availability for crop strips planted within apple orchard ». In : *Agricultural Systems* 170 (2019), p. 28–38 (cf. p. 51).
- [155] Q. WEN et al. « Transformers in time series : A survey ». In : *arXiv preprint arXiv :2202.07125* (2022) (cf. p. 67).
- [156] D. WHITE, M. RICHMAN et B. YARNAL. « Climate regionalization and rotation of principal components ». In : *International Journal of Climatology* 11.1 (1991), p. 1–25 (cf. p. 104).
- [157] F. D. WIHARTIKO et al. « Agricultural price prediction models : a systematic literature review ». In : *International Conference on Industrial Engineering and Operations Management Singapore, March, 2021*, p. 7–11 (cf. p. 47).

-
- [158] J. XIAO, R. LIU et E. L. DYER. « GAFormer : Enhancing Timeseries Transformers Through Group-Aware Embeddings ». In : *The Twelfth International Conference on Learning Representations* (cf. p. 68).
- [159] L. YANG et W. ZHOU. « Neural Network and Hydrological Model Integration for Urban Flood Prediction ». In : *Environmental Modelling Software* 104 (2018), p. 144–158. DOI : [10.1016/j.envsoft.2018.02.010](https://doi.org/10.1016/j.envsoft.2018.02.010) (cf. p. 50).
- [160] Q. YANG et al. « Federated machine learning : Concept and applications ». In : *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.2 (2019), p. 1–19 (cf. p. 72).
- [161] M. J. ZAKI et W. MEIRA. *Data mining and analysis : fundamental concepts and algorithms*. Cambridge University Press, 2014 (cf. p. 60).
- [162] W. ZHANG et AL. « Replacing Missing Data in Agricultural Datasets : A Comparative Study of Methods ». In : *Agronomy Journal* 111 (2019), p. 23–35 (cf. p. 52).
- [163] Z. ZHANG, Q. LI et Y. WANG. « Recent advances in data acquisition systems and sensors for precision agriculture ». In : *Sensors and Actuators B : Chemical* 305 (2020), p. 127586. DOI : [10.1016/j.snb.2020.127586](https://doi.org/10.1016/j.snb.2020.127586) (cf. p. 40).
- [164] W. ZHAO et AL. « Optimizing Agricultural Yield Forecasting with Climate Change Scenarios ». In : *Agricultural Systems* 180 (2020), p. 23–37 (cf. p. 53).
- [165] B. ZIV et al. « Regional and global atmospheric patterns governing rainfall in the southern Levant ». In : *International Journal of Climatology* 24.1 (2004), p. 55–73 (cf. p. 23).