

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE



Université 20 Août 1955 Skikda

Faculté des Sciences

Département des Sciences Agronomiques



Filière : Sciences Agronomiques

Mémoire :

Présentée pour l'obtention du diplôme de Master (LMD)

SPÉCIALITÉ : AMÉNAGEMENT HYDRO-AGRICOLE

Thème :

**Contribution à la prévision des débits du bassin
versant de la Seybous à l'aide de modèles
d'intelligence artificielle.**

Présenté par :

- Debehi Maroua
- Djouama Chaima
- Chalabi Wafa
- Babouche Aya

Membres de Jury:

Mr : HEDDAM Salim	(prof)	Président	Université du 20 Août 1955 – Skikda
Mr : MERABET Khaled	(MAB)	Examineur	Université du 20 Août 1955 – Skikda
Mr : HEBAL Aziz	(MCA)	Promoteur	Université du 20 Août 1955 – Skikda
Mr : DAIF Noureddine	(Dr)	Co-Promoteur	Université du 20 Août 1955 – Skikda

Année universitaire : 2024-2025

Remerciement

Nous tenons à exprimer notre profonde gratitude à notre promoteur **Mr. HEBAL Aziz**, pour ses précieuses orientations, ses conseils éclairés et sa disponibilité constante tout au long de ce travail. Son accompagnement rigoureux et bienveillant a été essentiel à la réalisation de ce mémoire.

Nous remercions également **Mr. DAIF Noureddine**, Co-promoteur, pour son aide, ses remarques constructives et son soutien tout au long de cette étude.

Nos sincères remerciements vont aux **membres du jury** qui ont accepté de consacrer de leur temps pour évaluer ce travail, ainsi que pour leurs observations et suggestions enrichissantes.

Nous remercions les ingénieurs et techniciens de la **Direction de l'Hydraulique de la Wilaya de SKIKDA**, pour nous avoir fournis les données nécessaires et nous avoir permis d'effectuer notre stage dans des conditions favorables à l'apprentissage.

Nos remerciements s'adressent également à l'ensemble **des enseignants du département d'agronomie** pour la qualité de leur enseignement, leur accompagnement durant tout notre parcours universitaire, et les connaissances précieuses qu'ils nous ont transmises.

Dédicaces

Je tiens à dédier ce travail

À celui qu'Allah a couronné de Prestige et de dignité... à celui qui m'a appris à donner sans rien attendre ... Mon cher père.

À mon ange dans la vie ... au sens de l'amour, au sens de la tendresse et du dévouement ... au sourire de la vie et au secret de l'existence.... Ma mère bien-aimée.

À mes frères, Mon soutien et ma côte stable ... Aux mains qui se tendent vers moi quand je trébuche.

À tous mes amis et proches, merci pour votre aide, votre soutien et vos encouragements durant ces moments difficiles.

Sans oublier tous les professeurs que ce soit du primaire, du moyen, du secondaire ou de l'enseignement supérieur.

Debeli Maroua

Dédicaces

Je dédie le fruit de cet humble effort à ceux qui ont été favorisés par Dieu

À mon père et à ma mère, battements de cœur et chaleur de l'âme, à mes frères, compagnons de route et soutien silencieux, et à tous ceux qui m'ont apporté leur soutien par des mots et des prières.

Pour le peuple palestinien, vous êtes la véritable signification de la patience et le symbole le plus élevé de la dignité.

Je vous dédie ce travail comme un gage de loyauté et une prière incessante pour une aube libre tant attendue.

Djouana Chaïma

الاهداء

الحمد لله حبًا وشكرًا وامتنانًا، فما كنت لأصل إلى ما أنا عليه اليوم لولا فضل الله، فله الحمد على البداية والخاتمة

أهدي ثمرة تخرجي هذه إلى حبيبي الأول، من كلل العرق جبينه وعلمني أن النجاح لا يأتي إلا بالصبر والإصرار، إلى من كان دائماً سندي ومصدر قوتي، إلى والدي العزيز

أهدي هذه الفرحة لتلك الإنسانية العظيمة، صاحبة أحنّ روح وأجمل قلب، من كانت تتمنى دومًا أن تقر عينها بي في مثل هذا اليوم، إلى جنتي في الدنيا، أُمي الغالية

اهدي هذا التخرج إلى إخوتي، ضلعي الثابت وأمان أيامي، إلى من كانوا لي نبغًا لا ينضب من الحب والدعم، قرّة عيني وصفوة أيامي

اهدي تخرجي لكل من كانوا عونًا وسندًا في هذا الطريق، صديقاتي الأوفياء ورفيقات الدرب، ممن شاركوني الشدائد قبل الأفراح

أهديكم هذا الإنجاز وتلك الثمرة الأولى من حلم طالما راودني، وها أنا اليوم أقطفها بفضل الله تعالى وكرمه

شلابي وفاء

Dédicaces

Louange à Dieu, le Très-Haut, le Très-Généreux. C'est par Sa grâce, Sa lumière et Sa miséricorde que ce chemin a été possible. À Lui reviennent la première reconnaissance et la gratitude éternelle.

À mes parents, source d'amour et de sacrifice. De vous, j'ai appris que le plus grand trésor est un cœur sincère et un espoir inébranlable.

À mes frères et sœurs, mon refuge dans les hauts et les bas de la vie.

À mon frère Islam, symbole d'honneur et de dévouement.

À ma sœur Amira, étoile du savoir et voix d'inspiration.

À Monsef, le frère et le compagnon, le silence qui comprend et le soutien fidèle.

À toute ma famille, petite et grande, vous êtes l'origine et la force qui m'ont porté.

À la mémoire de ceux qui habitent la lumière : ma tante bien-aimée, mes chers grands-parents — vous êtes mes racines cachées. À mes chers enseignants, des bancs de l'école primaire aux amphithéâtres de l'université, merci pour chaque mot, chaque regard, chaque leçon.

Au Professeur Aziz Habbal et au Dr Nouredidine Daif, vous avez été des phares sur mon chemin.

À mon amie fidèle Imane, ta présence a été une bénédiction inestimable.

Je dédie ce diplôme à tous ceux qui rêvent, qui croient que la connaissance est une arme noble, et que la lumière naît quand l'effort rencontre la foi.

Babouche Aya

Résumé

Cette étude s'inscrit dans le cadre de la modélisation hydrologique à l'aide de l'intelligence artificielle, et porte spécifiquement sur la prévision du débit journalier dans le bassin versant de la Seybouse, situé dans le nord-est de l'Algérie. La variabilité hydrologique dans les zones semi-arides, couplée à des pressions croissantes sur les ressources en eau, rend nécessaire le développement d'outils prédictifs fiables pour une gestion durable et anticipative. La problématique centrale de ce travail réside dans l'identification du ou des modèles les plus performants capables de restituer avec précision les dynamiques journalières des débits, à partir de variables hydrométéorologiques facilement disponibles. L'objectif principal de ce travail est donc de comparer l'efficacité de plusieurs architectures de réseaux de neurones artificiels dans la modélisation des écoulements journaliers. Pour cela, les données ont été collectées à partir de trois stations hydrométriques du bassin de la Seybouse, incluant les précipitations, l'évapotranspiration potentielle et les débits journaliers. Après une phase de prétraitement rigoureux, des analyses statistiques (ACF, PACF et XCF) ont été appliquées afin de déterminer les retards temporels pertinents. Quatre modèles d'intelligence artificielle ont ensuite été développés et testés : les réseaux de neurones classiques (ANN), les réseaux récurrents à mémoire courte (GRU), les réseaux à mémoire longue (LSTM) et les réseaux convolutifs (CNN). Les performances ont été évaluées via cinq métriques : le coefficient de corrélation (R), le coefficient NSE (Nash-Sutcliffe Efficiency), l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (RMSE) et l'indice d'efficacité de Kling-Gupta (KGE).

Les résultats obtenus montrent que les modèles CNN2 et LSTM2 se démarquent particulièrement. Le modèle CNN2 a donné les meilleures performances sur les stations 140301 (NSE = 0.937, KGE = 0.954) et 140501 (NSE = 0.900), tandis que le modèle LSTM2 s'est montré plus adapté à la station 140602 (NSE = 0.898, KGE = 0.908). Ces performances traduisent la capacité des réseaux profonds à capter la dynamique non linéaire des écoulements, même dans des contextes hydrologiques complexes. En conclusion, ce travail met en évidence l'efficacité des approches par apprentissage profond pour la prévision hydrologique, et confirme l'intérêt de les intégrer dans les outils de gestion des ressources en eau, en particulier dans les zones soumises à un stress hydrique croissant comme le bassin de la Seybouse.

Mots-clés

Prévision hydrologique ; Apprentissage profond ; Réseaux de neurones ; Débit journalier ; Bassin versant de la Seybouse

Abstract

This study falls within the framework of hydrological modeling using artificial intelligence, specifically focusing on daily streamflow forecasting in the Seybouse watershed, located in northeastern Algeria. Hydrological variability in semi-arid areas, coupled with increasing pressure on water resources, necessitates the development of reliable predictive tools for sustainable and anticipatory management. The central issue of this work lies in identifying the most efficient models capable of accurately reproducing daily discharge dynamics based on readily available hydrometeorological variables. The main objective is to compare the effectiveness of different artificial neural network architectures in modeling daily flows. For this purpose, data were collected from three hydrometric stations in the Seybouse basin, including precipitation, potential evapotranspiration, and daily streamflow. Following a rigorous preprocessing phase, statistical analyses (ACF, PACF, and XCF) were applied to determine the most relevant time lags. Four artificial intelligence models were developed and evaluated: classical neural networks (ANN), Gated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), and Convolutional Neural Networks (CNN). Model performance was assessed using five key metrics: correlation coefficient (R), Nash-Sutcliffe Efficiency (NSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Kling-Gupta Efficiency index (KGE).

The results show that CNN2 and LSTM2 models stand out. CNN2 achieved the best performance for stations 140301 (NSE = 0.937, KGE = 0.954) and 140501 (NSE = 0.900), while LSTM2 was better suited for station 140602 (NSE = 0.898, KGE = 0.908). These outcomes demonstrate the ability of deep learning networks to capture nonlinear flow dynamics, even in complex hydrological contexts. In conclusion, this study confirms the relevance of deep learning approaches in hydrological forecasting and supports their integration into water resource management tools, particularly in water-stressed regions like the Seybouse basin.

Keywords

Hydrological forecasting; Deep learning; Neural networks; Daily streamflow; Seybouse watershed

الملخص

تندرج هذه الدراسة في إطار النمذجة الهيدرولوجية باستخدام تقنيات الذكاء الاصطناعي، وتركز تحديداً على التنبؤ بالتدفق اليومي في حوض وادي سيبوس الواقع في شمال شرق الجزائر. إن التباين الهيدرولوجي في المناطق شبه الجافة، إلى جانب الضغوط المتزايدة على الموارد المائية، يفرض ضرورة تطوير أدوات تنبؤية موثوقة لدعم إدارة مستدامة واستباقية للمياه. وتتمثل الإشكالية المركزية في هذا العمل في تحديد النموذج أو النماذج الأكثر كفاءة في تمثيل ديناميكيات التصريف اليومية بدقة، اعتماداً على متغيرات هيدرولوجية متاحة بسهولة. الهدف الرئيسي هو مقارنة فعالية عدة بنى معمارية للشبكات العصبية الاصطناعية في نمذجة التدفق اليومي ولتحقيق ذلك، تم جمع البيانات من ثلاث محطات هيدرومترية ضمن حوض سيبوس، وتشمل الأمطار، التبخر-النتح المحتمل، والتدفق اليومي. وبعد معالجة البيانات تصحيحياً ومعياريًا، تم تطبيق تحاليل إحصائية ACF ، PACF ، XCF لتحديد التأخيرات الزمنية المناسبة. تم تطوير أربعة نماذج بالذكاء الاصطناعي واختبارها، وهي: الشبكات العصبية التقليدية (ANN) ، و شبكات الذاكرة القصيرة (GRU) ، و شبكات الذاكرة الطويلة (LSTM) ، و الشبكات العصبية الالتفافية (CNN). تم تقييم أداء النماذج باستخدام خمس مقاييس إحصائية: معامل الارتباط (R) ، معامل ناش-سوتكليف (NSE) ، متوسط الخطأ المطلق (MAE) ، متوسط الجذر التربيعي للخطأ (RMSE) ، ومؤشر كفاءة كلينغ-غوبتا (KGE).

أظهرت النتائج أن نموذجي CNN2 و LSTM2 تفوقا بشكل واضح. حيث حقق CNN2 أفضل أداء في المحطتين 140301 (NSE = 0.937, KGE = 0.954) و 140501 (NSE = 0.900) ، بينما تميز LSTM2 في المحطة 140602 (NSE = 0.898, KGE = 0.908). وتدل هذه النتائج على قدرة نماذج التعلم العميق على تمثيل السلوك غير الخطي للتدفق المائي بدقة، حتى في البيانات الهيدرولوجية المعقدة. وبناءً عليه، تؤكد هذه الدراسة فعالية التعلم العميق في التنبؤ الهيدرولوجي وأهميته في أدوات إدارة الموارد المائية، خاصة في المناطق التي تعاني من الإجهاد المائي مثل حوض السيبوس.

الكلمات المفتاحية

التنبؤ الهيدرولوجي؛ التعلم العميق؛ الشبكات العصبية؛ الصبيب اليومي؛ حوض وادي سيبوس

SOMMAIRE

Dédicace	I
Remerciements	II
Résumé	IV
Sommaire	IX
Liste des figures	XII
Liste des tableaux	XIV
Liste des abréviations	XV
INTRODUCTION.....	1
CHAPITRE I : LA MODELISATION HYDROLOGIQUE	
I.1 Introduction.....	3
I.2 Fondements de la modélisation hydrologique	3
I.2.1 Le cycle hydrologique.....	4
I.2.2 Bilan hydrique.....	5
I.2.3 Concepts clés en modélisation hydrologique.....	5
I.2.3.1 Définition du bassin versant.....	5
I.2.3.2 Les caractéristiques d'un bassin versant	5
I.2.4 Facteurs influençant le débit dans un bassin hydrologique.....	7
I.2.4.1 Facteurs climatiques.....	7
I.2.4.2 Facteurs physiques	7
I.2.4.3 Facteurs anthropiques	7
I.3 Typologie des modèles hydrologiques.....	8
I.3.1 Selon la nature de la variable	8
I.3.1.1 Modèles stochastiques	8
I.3.1.2 Modèles déterministes	9
I.3.2 Selon le traitement de l'espace (Homogénéité du bassin)	9
I.3.2.1 Modèles globaux	9
I.3.2.2 Modèles semi-distribués	9
I.3.2.3 Modèles distribués	10
I.3.3 Selon le fondement théorique et approche de développement.....	10

I.3.3.1 Modèles à base physique	19
I.3.3.2 Modèles empirique.....	10
I.3.3.3 Modèles conceptuels	11
I.4 Choix et mise en œuvre d'un modèle hydrologique	11
I.4.1 Choix du modèle	11
I.4.2 Données nécessaires pour la modélisation pluie-débit	12
I.4.3 Étapes de la mise en œuvre	13
I.4.3.1 Calibration du modèle (Calage)	13
I.4. 3.1 Validation du modèle	13
I.5. Problématique et limites de la modélisation hydrologique	13
I.5 Conclusion	14

CHAPITRE II : REGION D'ETUDE ET DONNEES UTILISEES

II .1 Introduction.....	15
II.2 Situation géographique	15
II.3 Aperçu topographique	16
II.3.1 Les plaines littorales	16
II.3.2 L'Atlas tellien	17
II.3.4 Les hautes plaines	17
II.3.5 L'Atlas saharien.....	17
II.4 Végétation et sols.....	17
II.5 Aperçu climatique.....	18
II.6 Caractéristiques générales du bassin de la Seybouse	19
II.6.1 Situation géographique	20
II.6.2 Sous-bassins sélectionnés	20
II.6.2.1 Le sous-bassin d'Ain Berda.....	20
II.6.2.2 Le sous-bassin de Lezadj Ammar	20
II.6.2.3 Le sous-bassin de Bouchegouf	21
II.7. Données utilisées	21
II.7.1 Collecte des données	21
II.7.2 Stations sélectionnées	22

II.7.3 Caractéristiques statistiques des séries	23
II.8 Conclusion	25

CHAPITRE III : MATERIELS ET METHODES

III .1 Introduction	26
III .2 Étapes de mise en œuvre de l'étude	26
III .3 Modèles d'intelligence artificielle appliqués à la prévision du débit	27
III .3.1 Réseaux de neurones artificiels (ANN)	28
III .3.2 Réseaux de neurones à mémoire longue (LSTM)	30
III .3.3 Réseaux de neurones à unité récurrente à portes (GRU).....	32
III .3.4 Réseaux de neurones convolutifs (CNN)	34
III .4 Évaluation des performances des modèles	35
III .4.1 Critères numériques de performance.....	36
III .4.2 Critères graphiques de visualisation des résultats	37
III .5 Outils et logiciels utilisés.....	38
III .6 Conclusion.....	39

CHAPITRE IV : RÉSULTATS ET DISCUSSIONS

V.1 Introduction	40
IV.2 Analyse et Interprétation des Résultats	40
IV. 3 Comparaison et généralisation des performances des modèles	49
IV. 4 Évaluation des Performances des Modèles en Prévision Hydrologique : Comparaison avec la Littérature.....	55
IV.5 Conclusion	57
CONCLUSION GENERALE	58
Référence bibliographique	60

LISTE DES FIGURES

CHAPITRE I : LA MODELISATION HYDROLOGIQUE

Figure I.1. Le cycle hydrologique.	4
Figure I.2. Schéma conceptuel d'un bassin versant et de ses composantes principales.....	6
Figure I.3. Classification des modèles hydrologiques (Berthier, 2005).....	11

CHAPITRE II : REGION D'ETUDE ET DONNEES UTILISEES

Figure 1 : Situation géographique de la région d'étude dans l'Est algérien.....	16
Figure II.5. Situation géographique du bassin versant de la seybouse en Algérie (ANRH, 2014)	20
Figure II.4 : Matrices de corrélation entre les variables climatiques et hydrologiques pour les stations d'Ain Berda, Bouchegouf et Mezadj Ammar	24

CHAPITRE III : MATERIELS ET METHODES

Figure.III.1 Schéma de la démarche méthodologique pour la prévision du débit journalier à l'aide de modèles d'intelligence artificielle	27
Figure.III.2 Architecture d'un Perceptron Multicouche (MLP).....	29
Figure.III.3 Structure interne d'une cellule LSTM (Long Short-Term Memory).....	31
Figure.III.4 Architecture interne d'une cellule GRU (Gated Recurrent Unit).....	33
Figure.III.5 Architecture typique d'un réseau de neurones convolutif (CNN 1D) pour la modélisation de séries temporelles.....	35

CHAPITRE IV : RÉSULTATS ET DISCUSSIONS

Figure IV.1. Comparaison des Prédictions de Débit avec les Observations pour la Station 140301 à l'Aide de Différents Modèles.....	46
Figure IV.2.Comparaison des Prédictions de Débit avec les Observations pour la Station 140501 à l'Aide de Différents Modèles.....	47
Figure IV.3.Comparaison des Prédictions de Débit avec les Observations pour la Station 140602 à l'Aide de Différents Modèles.....	48

Figure IV.4 : Comparaison des distributions des erreurs des modèles pour les trois stations (Boxplot et Violin plot).....	52
Figure IV.5 : Diagramme de Taylor pour les trois stations – Comparaison des modèles de prévision du débit.	53
Figure IV.6 : Matrice de corrélation des performances des modèles pour la station 140301 – Comparaison des métriques	53
Figure IV.7 : Radar plot des performances des modèles pour les trois stations – Visualisation des métriques de performance.....	53
Figure IV.8 : Bar plot des performances des modèles pour les trois stations – Visualisation des métriques de performance	54

LISTE DES TABLEAUX

CHAPITRE I : LA MODELISATION HYDROLOGIQUE

Tableau.I.1. Exemple de caractéristiques des modèles pluie-débit (Berthier, 2005)8

CHAPITRE II : REGION D’ETUDE ET DONNEES UTILISEES

Tableau II.1. Caractéristiques des bassins versants étudiés (Source ANRH, 2021)21

Tableau II.2 : Stations choisies dans cette étude (*Sources : ANRH, 2021 ; ONM, 2021*)23

Tableau II.3. Paramètres statistiques des données utilisées au pas de temps journalier23

Tableau II.4. Les combinaisons entrées pour différents modèles 25

CHAPITRE IV : RÉSULTATS ET DISCUSSIONS

Tableau.IV.1. Performances des différents modèles pour la station MEDJZ AMAR40

Tableau.IV.2. Performances des différents modèles pour la station BOUCHEGHOUF .. 42

Tableau.IV.3. Performances des différents modèles pour la station AIN BERDA44

Tableau IV.4 : Synthèse des performances des meilleures combinaisons de modèles pour les trois stations.....49

LISTE DES ABREVIATIONS

ACF = Autocorrelation Function = Fonction d'autocorrélation

PCA = Principal Component Analysis = Analyse en composantes principales

AI = Artificial Intelligence = Intelligence artificielle

ANFIS = Adaptive Neuro-Fuzzy Inference System = Système adaptatif d'inférence neuro floue

ANN = Artificial Neural Networks = Réseaux de neurones artificiels

ANRH = Agence Nationale des Ressources Hydraulique

DWT = Discrete Wavelet Transform = Ondelettes de type transformation discrète

ELM = Extreme Learning Machine = Machine d'apprentissage extrême

EMD = Empirical Mode Decomposition = Décomposition en mode empirique

GA = Genetic Algorithm = Algorithme génétique

GEP = Gene Expression Programming = programmation d'expression génique

GP = Genetic Programming = Programmation génétique

GR2m = Modèle conceptuel Génie rural à 2 paramètres Mensuel

GR4J = Modèle conceptuel Génie rural à 4 paramètres journalier

GR5J = Modèle conceptuel Génie rural à 5 paramètres journalier

GR6J = Modèle conceptuel Génie rural à 6 paramètres journalier

GRNN= Generalized Regression Neural Networks= Réseaux de neurones de régression généralisée

GSA = Gravitational Search Algorithm = Algorithme de recherche gravitationnelle

GWO = Grey Wolf Optimizer = Algorithme d'optimisation du loup gris

LSSVM = Least-Squares Support Vector Machines = Machines à vecteurs de support en moindres carrés

M5Tree = Model Tree = Modèles d'arbres

MAE = Mean Absolute Error = Erreur moyenne absolue

MARS = Multivariate Adaptive Regression Splines = spline de régression adaptative multivariée

MLP = Multi-Layer Perceptron = Perceptron multicouche

MLR = multiple linear regression = Régression linéaire multiple

MODWT = Maximum Overlap Discrete Wavelet Transform = Transformée en Ondelette Discrète à Chevauchement Maximum

NSE = Nash Sutcliffe criterion = Critère de Nash Sutcliffe

ONM = Office Nationale de la Météorologie

PACF = Partial Autocorrelation Function = Fonction d'autocorrélation partielle PSO = Particle swarm optimization = Optimisation par essais des particules

Q = Débits

R^2 = Coefficient de détermination

R = Coefficient de corrélation

RBF = Radial Basis Function = Fonction de base radiale

RF = Random Forest = Forêt aléatoire

RMSE = Root Mean Square Error = Erreur moyenne quadratique

SVM = Support Vector Machine = Machines à vecteurs de support

SVR = Support Vector Regression = Régression à vecteurs de support

WANN = Wavelet-Artificial Neural Networks models = Modèles ondelettes-réseaux de neurones artificiels

WT = wavelet Transform = transformation en ondelettes

WSVR = Wavelet Support Vector Regression models = Modèles Ondelettes- régression à vecteurs de support

VMD = Variational Mode Decomposition = La décomposition en modes variationnels

INTRODUCTION

INTRODUCTION

L'eau constitue une ressource vitale pour l'environnement, l'économie et les sociétés humaines. Elle intervient dans tous les secteurs d'activité, de l'agriculture à l'énergie, en passant par l'approvisionnement en eau potable et la préservation des écosystèmes. Dans un contexte de croissance démographique, de changement climatique et d'augmentation des pressions anthropiques, la gestion durable des ressources en eau devient un enjeu stratégique à l'échelle mondiale (Falkenmark & Rockström, 2004 ; Vörösmarty et al., 2000 ; Gleick, 1993). Dans les régions semi-arides comme le nord-est de l'Algérie, la rareté et l'irrégularité des précipitations accentuent la vulnérabilité des bassins versants et rendent la gestion hydrologique encore plus complexe.

Dans ce cadre, la prévision des débits constitue un outil fondamental pour la planification et la gestion des ressources en eau. Elle permet d'anticiper les pénuries ou les excès d'eau, de gérer les ouvrages hydrauliques et de prévenir les risques d'inondation (Shamseldin, 1997). Or, la mesure directe des débits dans les rivières reste coûteuse, techniquement contraignante et souvent sujette à des erreurs dues aux incertitudes dans les courbes de tarage, à l'accès difficile aux sites, ou encore aux phénomènes extrêmes (McMillan et al., 2012). Face à ces contraintes, les chercheurs et ingénieurs ont recours à la modélisation hydrologique pluie-débit, qui permet de simuler le comportement du bassin versant à partir de variables d'entrée accessibles comme les précipitations et l'évapotranspiration.

Depuis deux décennies, les modèles classiques ont été enrichis par les avancées en intelligence artificielle, qui offrent des outils puissants pour modéliser les relations non linéaires et complexes entre les variables hydrologiques. Les réseaux de neurones artificiels (ANN), les modèles récurrents comme LSTM et GRU, ou encore les réseaux convolutifs (CNN), ont montré des performances prometteuses dans la prévision du débit journalier, même dans des bassins soumis à une forte variabilité climatique (Kratzert et al., 2019).

C'est dans ce contexte que s'inscrit le présent travail, qui vise à évaluer les performances comparées de plusieurs modèles d'intelligence artificielle (ANN, LSTM, CNN, GRU) pour la prévision du débit journalier sur trois stations hydrométriques situées dans le bassin versant de la Seybouse. Les variables d'entrée utilisées sont les précipitations,

l'évapotranspiration potentielle et le débit décalé dans le temps, avec une sélection du décalage optimal à l'aide des fonctions ACF, PACF et XCF.

Les objectifs principaux de cette étude sont les suivants :

- Identifier les combinaisons temporelles optimales des variables d'entrée hydrométéorologiques (précipitations, évapotranspiration, débit) en s'appuyant sur des outils d'analyse statistique tels que les fonctions d'autocorrélation (ACF, PACF, XCF) afin d'améliorer la qualité des prévisions ;
- Évaluer et comparer les performances prédictives de plusieurs modèles d'intelligence artificielle (ANN, LSTM, CNN, GRU) en termes de précision, de stabilité et de capacité de généralisation dans le contexte d'un bassin semi-aride ;
- Développer un outil décisionnel fiable pour appuyer la gestion durable des ressources en eau dans le bassin versant de la Seybouse, en fournissant des prévisions de débit adaptées aux contraintes locales et aux enjeux climatiques futurs.

Pour atteindre les objectifs de cette étude, notre travail a été structuré en quatre chapitres, encadrés par une introduction et une conclusion générale. Le premier chapitre, intitulé « *Modélisation hydrologique* », présente les fondements théoriques de la modélisation pluie-débit, les différents types de modèles hydrologiques, ainsi que les concepts-clés nécessaires à la compréhension de la dynamique des bassins versants. Le deuxième chapitre est consacré à la *région d'étude* et *données utilisées*. Il décrit les caractéristiques géographiques, climatiques et hydrologiques du bassin versant de la Seybouse, ainsi que les stations sélectionnées et les séries de données exploitées. Dans le troisième chapitre, « *Matériel et méthodes* », nous exposons la démarche méthodologique adoptée, en détaillant les outils logiciels utilisés, les étapes de préparation des données, les fonctions statistiques appliquées (ACF, PACF, XCF), ainsi que les architectures des modèles d'intelligence artificielle retenus (ANN, LSTM, CNN, GRU). Le quatrième chapitre est consacré aux *résultats et à leur discussion*. Il présente les performances des différents modèles appliqués à la prévision du débit journalier, en comparant leurs précisions respectives, et discute des avantages et limites de chaque approche dans le contexte du bassin étudié. Ce mémoire s'achève par une conclusion générale récapitulant les principaux enseignements de l'étude et proposant des perspectives pour de futurs travaux.

CHAPITRE I
LA MODELISATION
HYDROLOGIQUE

CHAPITRE I : LA MODELISATION HYDROLOGIQUE

I.1 Introduction

La modélisation hydrologique consiste à représenter de manière simplifiée le fonctionnement du cycle de l'eau dans un bassin versant, à travers des outils mathématiques et numériques. Elle permet de simuler et de prévoir les principaux processus hydrologiques, en transformant notamment des données climatiques (pluie, température) en débits. En contexte aride ou semi-aride, comme en Algérie, cette approche est particulièrement précieuse pour optimiser la gestion des ressources en eau et appuyer les décisions en matière d'aménagement.

Ce travail s'intéresse plus spécifiquement à la modélisation pluie-débit, qui simule la transformation des précipitations en écoulements à l'échelle d'un bassin. Elle repose sur deux étapes clés : la production (liée à l'infiltration et à l'évapotranspiration) et le transfert (ou routage) vers l'exutoire. Ces modèles sont utilisés pour combler des lacunes de données, prévoir les crues et étiages, ou encore évaluer les effets des aménagements et du changement climatique.

Dans cette partie du travail ; nous rappellerons ce qu'est la modélisation hydrologique et nous évoquerons également les types de modèles utilisés dans ce domaine. Mais, avant d'entamer cette revue, il est utile de donner un bref rappel sur le cycle hydrologique qui est la base de toute modélisation hydrologique.

I. 2. Fondements de la modélisation hydrologique

La modélisation hydrologique repose sur une compréhension approfondie des processus qui gouvernent le cycle de l'eau. Avant d'aborder les types de modèles et leurs applications, il est essentiel de rappeler les bases physiques qui sous-tendent toute représentation du fonctionnement hydrologique d'un bassin versant. Cette section présente les principaux éléments du cycle hydrologique, ainsi que la notion de bilan hydrique, socle fondamental de toute modélisation.

I.2.1. Le cycle hydrologique

Le cycle hydrologique désigne l'ensemble des transferts et transformations de l'eau dans la nature. Il débute par les précipitations, qui représentent la principale source d'apport en eau au niveau d'un bassin versant. Ces précipitations peuvent prendre différentes formes (pluie, neige, grêle) et leur variabilité influence fortement les écoulements hydrologiques (Brutsaert, 2005). Une part de cette eau est renvoyée vers l'atmosphère sous forme de vapeur, par l'intermédiaire de l'évapotranspiration, processus combinant l'évaporation des surfaces (sols, plans d'eau) et la transpiration des végétaux. L'infiltration, quant à elle, permet à une fraction de l'eau de s'introduire dans le sol, alimentant les nappes et participant à l'écoulement souterrain. Lorsque l'intensité des précipitations dépasse la capacité d'infiltration, l'eau excédentaire génère un ruissellement de surface, qui alimente directement les rivières et les oueds, souvent de manière rapide et brutale en milieu aride.

À cela s'ajoutent d'autres composantes telles que le stockage temporaire (dans les sols, les nappes phréatiques, ou sous forme de neige) et le drainage souterrain, qui assure un écoulement de base vers les cours d'eau, même en absence de précipitations. Ces différentes composantes interagissent selon des échelles spatiales et temporelles variables, ce qui complexifie leur représentation dans les modèles hydrologiques (Chow et al., 1988 ; Musy & Higy, 2004), la **Figure I.1** illustre de manière schématique le cheminement de l'eau à travers les différents réservoirs naturels.

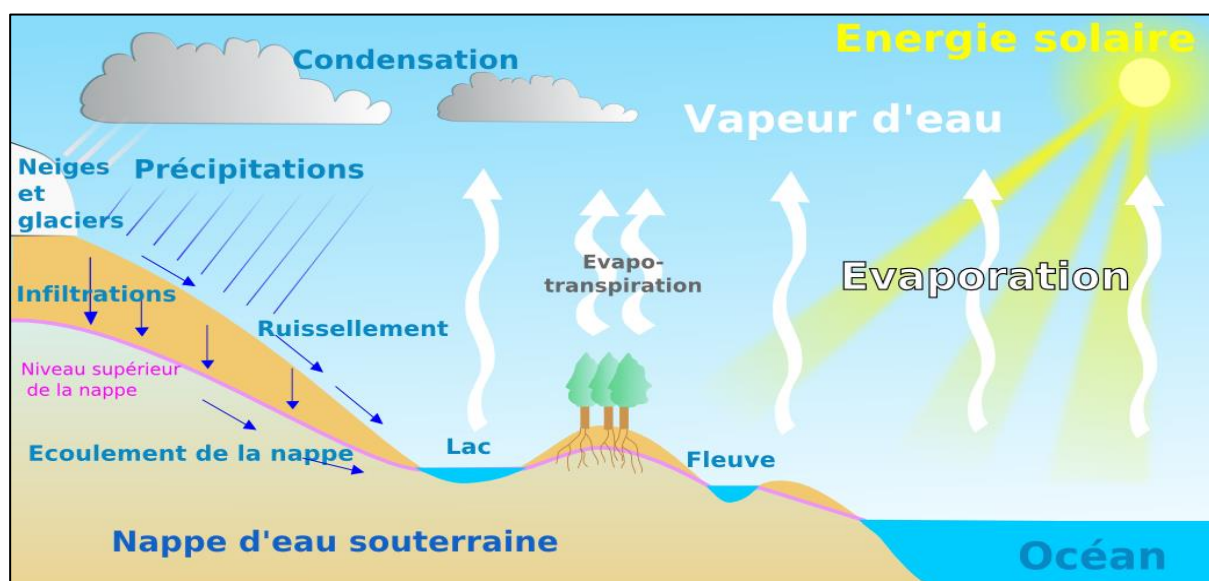


Figure I.1. Le cycle hydrologique

I.2.2 Bilan hydrique

Le **bilan hydrique** permet de quantifier les entrées, sorties et variations de stockage d'eau dans un bassin versant. Il repose sur le principe fondamental de conservation de la masse. À l'échelle d'un bassin, l'équation générale du bilan hydrique s'écrit :

$$P=ET+R+I+\Delta S \qquad \mathbf{I.1}$$

Où P représente les précipitations, ET l'évapotranspiration, R le ruissellement, I l'infiltration profonde, et ΔS la variation de stockage dans le sol ou les nappes. Cette équation permet d'estimer la disponibilité en eau dans un système donné, de détecter des déséquilibres, et de valider les résultats issus des modèles hydrologiques. Elle constitue ainsi un outil fondamental pour l'analyse et la gestion intégrée des ressources en eau (UNESCO, 1978).

I.2.3 Concepts clés en modélisation hydrologique

La modélisation hydrologique repose sur plusieurs concepts fondamentaux, dont la maîtrise est indispensable pour construire, calibrer et interpréter un modèle. Parmi ces concepts, le bassin versant constitue l'unité spatiale de base, tandis que des notions temporelles comme le temps de concentration ou le temps de réponse permettent de caractériser le comportement dynamique du système.

I.2.3.1 Définition du bassin versant

Le bassin versant est l'unité spatiale de base en hydrologie. Il correspond à une surface topographiquement délimitée, au sein de laquelle toutes les eaux de ruissellement convergent vers un même point de sortie, appelé exutoire. Sa forme et ses caractéristiques physiques influencent fortement la réponse hydrologique du système aux apports pluviométriques (Chow et al., 1988).

I.2.3.2 Les caractéristiques d'un bassin versant

Les caractéristiques d'un bassin versant influencent fortement la quantité, la vitesse et la distribution spatiale des écoulements générés par les précipitations. Elles se répartissent en plusieurs grandes catégories.

Les caractéristiques morphométriques incluent la surface du bassin, son périmètre, la longueur hydraulique (distance entre le point le plus éloigné et l'exutoire), et des indices de forme comme le coefficient de compacité, l'indice de Gravelius, et le rapport d'élongation, qui permettent de qualifier la forme du bassin (circulaire, allongée, etc.).

Les caractéristiques topographiques, ou de relief, comprennent l'altitude minimale et maximale, la pente moyenne, la pente du cours principal, ainsi que des indices de déclivité. Ces éléments influencent directement la vitesse du ruissellement, la capacité d'infiltration et la concentration des écoulements. On y associe également des paramètres temporels comme le temps de concentration, qui correspond au temps nécessaire à une goutte d'eau tombée à l'endroit le plus éloigné du bassin pour atteindre l'exutoire. Ce paramètre, essentiel en modélisation hydrologique, permet d'estimer la réactivité d'un bassin aux précipitations ; plus le temps de concentration est court, plus la réponse du bassin est rapide (Kirpich, 1940 ; Viessman & Lewis, 2003).

Enfin, les caractéristiques du réseau hydrographique décrivent la manière dont les cours d'eau structurent le bassin. On considère notamment la densité de drainage (longueur totale des cours d'eau rapportée à la superficie du bassin), la sinuosité, et l'ordre des rivières selon la classification de Strahler. Ces paramètres permettent de comprendre la manière dont les eaux collectées se dirigent vers l'exutoire, ce qui est crucial dans les modèles distribués. Temps de concentration, défini comme le temps nécessaire à une goutte d'eau tombée au point le plus éloigné du bassin pour atteindre l'exutoire. Ce temps reflète la rapidité de réponse du bassin à un événement pluvieux et dépend à la fois de la morphologie du terrain et de la structure du réseau hydrographique.



Figure I.2. Schéma conceptuel d'un bassin versant et de ses composantes principales

I.2.4 Facteurs influençant le débit dans un bassin hydrologique

Le débit observé à l'exutoire d'un bassin est le résultat de nombreux facteurs agissant sur la quantité, la vitesse et la temporalité des écoulements. Ces facteurs peuvent être regroupés en trois grandes catégories : climatiques, physiques et anthropiques.

I.2.4.1 Facteurs climatiques

Les précipitations sont le principal moteur du cycle hydrologique. Leur intensité, fréquence et répartition conditionnent directement les volumes d'eau générés par ruissellement ou infiltration (Beck et al., 2017). La température influence l'évapotranspiration. Une élévation thermique augmente les pertes par évaporation, réduisant la quantité d'eau disponible pour l'alimentation des cours d'eau (Huntington, 2006).

I.2.4.2 Facteurs physiques

La topographie du bassin (altitude, pente) détermine la vitesse d'écoulement. Des pentes fortes génèrent une réponse hydrologique plus rapide que des reliefs modérés. La géologie et la nature des sols affectent l'infiltration. Les terrains perméables favorisent la recharge souterraine, tandis que les substrats imperméables amplifient le ruissellement (Rawls, Brakensiek, & Miller, 1983). La végétation ralentit les écoulements, limite l'érosion et augmente l'infiltration. Elle joue aussi un rôle dans l'évapotranspiration, modulant le bilan hydrique.

I.2.4.3 Facteurs anthropiques

L'urbanisation augmente les surfaces imperméables, réduisant l'infiltration et intensifiant les débits de pointe (Shuster et al., 2005). Les barrages et ouvrages hydrauliques modifient les régimes naturels. Ils régulent les crues, soutiennent les étiages mais perturbent les flux sédimentaires et les écosystèmes (Graf, 2006).

I.3 Typologie des modèles hydrologiques

Les classifications de modèles pluie-débit sont nombreuses et peuvent varier en fonction de plusieurs caractéristiques, comme mentionné par Berthier en 2005. Pour plaisanter, Perrin (2002) disait qu'il existe presque autant de classifications que d'hydrologues. Le tableau suivant propose une classification des différents types de modèles selon leurs caractéristiques, en s'appuyant sur les travaux : d'Ambroise (1998), Andrassian (2002), Clarke (1973), Oudin (2004) et Perrin (2002).

Tableau.I.1. Exemple de caractéristiques des modèles pluie-débit (Berthier, 2005)

Caractéristiques	Type de modèle
Selon la nature de la variable	Modèle déterministe (variables non aléatoires) Modèle stochastique (variables aléatoires)
Traitement de l'espace (Homogénéité du bassin)	Modèle global Modèle semi-distribué (ou semi-global) Modèle spatialisé
Fondement théorique et approche de développement	Modèle empirique Modèle conceptuel Modèle à base physique

I.3.1 Selon la nature de la variable

I.3.1.1 Modèles stochastiques

Un modèle stochastique a, au moins, une composante de caractère aléatoire qui n'est pas explicite dans l'entrée du modèle, mais seulement implicite ou "cachée" (Le Clerc, 2004). Par conséquent, des entrées identiques se traduiront généralement par des sorties différentes si elles sont exécutées dans le modèle dans des conditions identiques. Cette notion peut être étendue aux modèles dans lesquels l'entrée a un caractère stochastique direct. Les modèles stochastiques utilisent des variables aléatoires, pour représenter l'incertitude du processus et génèrent des résultats différents, à partir d'un ensemble de données d'entrée et de valeurs de paramètres lorsqu'ils s'exécutent dans des conditions identiques (Pechlivanidis et al., 2011).

I.3.1.2 Modèles déterministes

Un modèle est dit déterministe si aucune de ses grandeurs n'est considérée comme aléatoire (Maftai, 2002). Un tel modèle associe à chaque jeu de variables d'entrée, de variables d'état et de paramètres une valeur déterminée ou déterminable, réalisation unique des variables de sortie (Le Clerc, 2004). La majorité des modèles hydrologiques sont déterministes (Gnouma, 2006).

Les modèles déterministes sont complexes et nécessitent beaucoup de données et de temps de calcul. Ils utilisent des équations aux dérivées partielles (EDP) non linéaires pour décrire les processus hydrologiques, et il est souvent impossible de résoudre ces équations de manière analytique. Cependant, l'un des avantages des modèles déterministes est qu'ils permettent une meilleure compréhension du système hydrologique en fournissant une vue détaillée du processus (Jajarmizadeh et al., 2012).

I.3.2 Selon le traitement de l'espace (Homogénéité du bassin)

I.3.2.1 Modèles globaux

Dans les modèles tels que GR (Edijatno et Michel, 1989), HEC-1 (Hydrologic Engineering Center, 1998) ou PDM Probability Distributed Moisture (Moore, 2007), le bassin versant est considéré comme une seule entité réagissant dans son ensemble. Les caractéristiques physiques et les grandeurs météorologiques sont supposées uniformes à l'échelle du bassin, assimilé à une unité ponctuelle. Ces valeurs moyennes s'appliquent à l'ensemble du bassin versant, et les processus hydrologiques sont simulés pour l'ensemble du bassin.

I.3.2.2 Modèles semi-distribués

Ces modèles prennent en compte la variabilité spatiale des processus hydrologiques modélisés, des variables d'entrée, des conditions aux limites et des caractéristiques du bassin (Yahi, 2010). Dans les modèles semi-distribués, ces quantités peuvent varier partiellement dans l'espace en divisant le bassin en petits sous-bassins, qui sont ensuite traités comme une seule unité. On peut aussi choisir de spatialiser uniquement les facteurs dont la répartition spatiale est jugée cruciale dans la réponse hydrologique d'un bassin versant. Ces types de modèles effectuent des prévisions distribuées dans l'espace en utilisant des variables d'état moyennées sur le carré de quadrillage. Cette approche de distribution offre une meilleure représentation de l'hétérogénéité pouvant exister sur un bassin. Cependant, la spatialisation entraîne une complexification des modèles.

I.3.2.3 Modèles distribués

Les modèles distribués, généralement associés aux modèles à base physique, présentent plusieurs avantages dans les études relatives à l'effet du changement des conditions du terrain sur les processus hydrologiques, la modélisation de la qualité de l'eau et la prévision sur les bassins versants non-jaugés. Une des raisons pour lesquelles les modèles de paramètres distribués n'ont pas été largement utilisés est l'indisponibilité de bases de données détaillées. Cependant, les améliorations futures dans l'acquisition de données, y compris l'application des systèmes d'information géographique (SIG) et des images satellitaires, vont probablement conduire à une utilisation plus étendue des modèles de paramètres distribués.

I.3.3 Selon le fondement théorique et approche de développement

On peut classer les modèles pluie-débit en trois groupes principaux : les modèles empiriques (modèles boîte noire, régression linéaire ...etc), conceptuels et les modèles à base physique.

I.3.3.1 Modèles à base physique

Les modèles physiques, également appelés modèles basés sur les processus ou modèles mécanistes, sont basés sur la compréhension de la physique liée aux processus hydrologiques. Des équations basées sur la physique régissent le modèle pour représenter plusieurs parties des réponses hydrologiques réelles dans le bassin versant (Knights, 2017 ; Vaze et al., 2012). On peut citer les modèles HYDROTEL et ATHYS.

Ne suivant que les équations de la physique, et ne comportant idéalement aucun paramètre. Il n'existe pas de modèle à base physique au sens strict en hydrologie. Certains modèles comme le système hydrologique européen (SHE) tendent à s'en rapprocher. L'importance de l'hétérogénéité spatiale dans la réponse hydrologique des bassins versants rend cependant difficile, voire impossible, l'utilisation de tels modèles (Henine, 2005).

I.3.3.2 Modèles empirique

Les modèles empiriques reposent sur les relations observées entre les entrées et les sorties de l'hydro système considéré. Ils expriment la relation entre variables d'entrée et de sortie du système (relation pluie débit) à l'aide d'un ensemble d'équations développées et ajustées sur la base des données obtenues sur le système. Un modèle empirique ne cherche pas à décrire les causes du phénomène hydrologique considéré ni de ni à expliquer le fonctionnement du

système, le système est vu comme une boîte noire. Divers outils sont utilisés pour l'élaboration des formules ou modèles reliant les entrées et les sorties : la régression, les réseaux de neurones, SVR (Sector Vector Machine for Regression) , ANFIS (Adaptive Neuro Fuzzy Inference System) , etc.

I.3.3.3 Modèles conceptuels

Les modèles conceptuels hydrologiques, représentent une approche simplifiée pour décrire les processus hydrologiques dans un bassin versant. Les modèles conceptuels, tentent à représenter le bassin versant par des analogies et des concepts. L'analogie la plus réputée en hydrologie est la représentation des fonctionnements des sols et des nappes, par des réservoirs d'humidité (d'où l'appellation modèle à réservoir) interconnecté suivant une dimension verticale, dont le débit de vidange de chaque réservoir dépend du taux de remplissage (Ambroise, 1999). Les modèles de génie rural GR sont des exemples de modèles conceptuels pour des applications d'ingénierie.

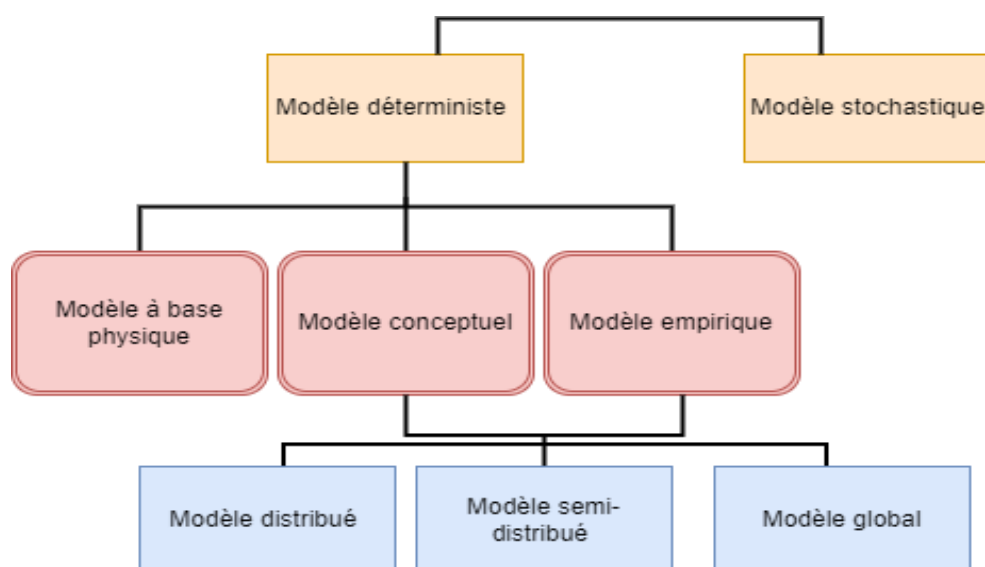


Figure I.3. Classification des modèles hydrologiques (Berthier, 2005)

I.4 Choix et mise en œuvre d'un modèle hydrologique

I.4.1 Choix du modèle

Le choix du modèle hydrologique à utiliser constitue une étape cruciale dans toute étude de simulation pluie-débit. Il conditionne non seulement la qualité des résultats obtenus, mais également leur pertinence par rapport aux objectifs de l'étude. Ce choix repose sur une analyse

multicritère tenant compte de plusieurs aspects : la disponibilité des données, la complexité physique du bassin versant, l'échelle spatio-temporelle de l'étude, ainsi que les objectifs spécifiques de la modélisation.

D'un point de vue général, les modèles hydrologiques peuvent être classés selon trois critères : leur structure (globale, semi-distribuée, distribuée), leur nature (empirique, conceptuelle, physique), et leur finalité (diagnostique, prédictive, opérationnelle). Chaque catégorie présente des avantages et des limites. Les modèles conceptuels, tels que GR4J ou NAM, sont adaptés à des bassins pour lesquels seules des données hydrologiques et pluviométriques sont disponibles. Ils offrent un bon compromis entre simplicité de mise en œuvre et capacité de reproduction des débits (Perrin et al., 2003). En revanche, les modèles physiques distribués comme MIKE SHE ou SWAT exigent une base de données plus riche et une connaissance fine des paramètres du bassin, mais permettent une meilleure représentation spatiale des processus hydrologiques (Abbaspour, 2015).

En pratique, le choix du modèle dépend également du niveau de complexité souhaité, des ressources techniques disponibles (logiciels, puissance de calcul), ainsi que de la finalité de l'étude : gestion des crues, évaluation des impacts du changement climatique, gestion de l'eau agricole, etc. Une approche itérative est souvent recommandée, partant d'un modèle simple vers des configurations plus complexes, afin de valider progressivement les hypothèses de modélisation (Beven, 2012).

I.4.2 Données nécessaires pour la modélisation pluie-débit

La modélisation hydrologique pluie-débit nécessite l'intégration de trois grandes catégories de données : météorologiques, hydrologiques et géographiques. Les données météorologiques, principalement les précipitations et la température, sont les entrées fondamentales des modèles. Elles permettent notamment d'alimenter les processus de ruissellement et d'estimer l'évapotranspiration potentielle, essentielle dans le bilan hydrique (Xu & Singh, 2004). Les données hydrologiques, en particulier les séries de débits mesurés à l'exutoire d'un bassin, sont indispensables pour la calibration et la validation des modèles. Une série longue, continue et de qualité est nécessaire pour évaluer correctement les performances du modèle (Moriassi et al., 2007). Quant aux données géographiques, elles permettent de représenter l'espace physique du bassin versant. Les modèles numériques de terrain (MNT),

associés aux données d'occupation du sol et aux types de sols, sont utilisés pour délimiter les bassins, définir les pentes et simuler les chemins d'écoulement (Jenson & Domingue, 1988). L'ensemble de ces données constitue le socle sur lequel repose la fiabilité et la robustesse des simulations hydrologiques.

I.4.3 Étapes de la mise en œuvre

Deux étapes essentielles sont souvent considérées comme fondamentales dans la modélisation hydrologique :

I.4.3.1 Calibration du modèle (Calage) : La calibration est un processus crucial qui vise à ajuster les paramètres du modèle, pour qu'il reproduise aussi précisément que possible les observations réelles (Laurent, 1996). L'estimation des paramètres peut se faire manuellement par une procédure « essai et erreur », ou automatiquement par une fonction d'optimisation permettant d'améliorer la cohérence entre la réponse observée et celle simulée (Gherissi, 2018).

I.4.3.1 Validation du modèle : La validation est l'étape où l'on vérifie la capacité du modèle à prédire, avec précision, les débits ou autres variables hydrologiques, dans des situations pour lesquelles il n'a pas été spécifiquement calibré (Fortin et al., 1997).

Ces deux étapes sont cruciales, car elles garantissent la fiabilité du modèle hydrologique. Un modèle correctement calibré et validé, a plus de chances de produire des résultats précis et fiables lorsqu'il est utilisé pour des prévisions, des analyses de scénarios ou des évaluations d'impact dans le domaine de la gestion des ressources en eau.

I.5. Problématique et limites de la modélisation hydrologique

Malgré les avancées techniques et scientifiques, la modélisation pluie-débit présente plusieurs limitations qui affectent la précision et la fiabilité des résultats. Ces limitations sont liées à la nature des données, à la complexité des processus hydrologiques, ainsi qu'à la variabilité spatiale et temporelle du système étudié.

L'une des principales problématiques concerne les incertitudes dans les données d'entrée et dans les paramètres des modèles. Les séries de précipitations ou de débits peuvent présenter des lacunes, des erreurs de mesure ou un pas de temps inadapté à la dynamique réelle du bassin. De plus, de nombreux paramètres hydrologiques doivent être calibrés de manière empirique, ce qui introduit des marges d'erreur parfois importantes (Beven, 1993). Ces incertitudes peuvent

se propager dans tout le processus de simulation et produire des résultats biaisés, surtout en l'absence de validation indépendante.

Par ailleurs, la variabilité climatique, notamment dans les régions méditerranéennes soumises à des régimes de pluie irréguliers, rend difficile la prévision hydrologique à long terme. À cela s'ajoute l'impact du changement climatique, qui modifie la fréquence et l'intensité des événements extrêmes. Selon Kundzewicz et al. (2007), l'augmentation de la température et des phénomènes extrêmes remet en question les hypothèses de stationnarité souvent implicites dans les modèles calibrés sur des données historiques.

Enfin, la question de l'échelle spatiale et temporelle constitue une contrainte importante. Un modèle performant à l'échelle d'un petit sous-bassin peut ne pas être transférable à un bassin plus vaste sans ajustements importants. De même, les modèles peuvent perdre en précision lorsqu'ils sont appliqués à des pas de temps très fins (horaire) ou très longs (annuel), en raison de la simplification des processus hydrologiques ou du manque de données disponibles (Blöschl & Sivapalan, 1995). L'hétérogénéité du sol, de la topographie et de l'occupation du sol accentue également ces difficultés, notamment dans les modèles distribués.

Ces problématiques appellent à une approche prudente et rigoureuse de la modélisation pluie-débit, intégrant la quantification des incertitudes, la validation croisée des résultats, et une réflexion approfondie sur l'échelle d'application et la nature des données mobilisées.

I.5 Conclusion

La modélisation hydrologique permet de simuler les écoulements à partir des précipitations et constitue un outil essentiel en gestion des ressources en eau. Ce chapitre a présenté ses fondements théoriques, ses différents types de modèles et les critères de choix. L'importance des données et les étapes de mise en œuvre ont également été abordées. Malgré son utilité, la modélisation présente des limites liées aux incertitudes et à la variabilité climatique. Ces éléments préparent l'application concrète sur le bassin étudié.

CHAPITRE II

REGION D'ETUDE ET

DONNEES UTILISEES

CHAPITRE II : REGION D'ETUDE ET DONNEES UTILISEES

II .1 Introduction

Ce chapitre vise à décrire la zone géographique ciblée par l'étude de modélisation pluie-débit, à savoir le bassin versant de la Seybouse, situé dans le nord-est de l'Algérie. Il présente également les données hydrométéorologiques utilisées dans le cadre de cette recherche, en précisant leurs sources, ainsi que les traitements statistiques appliqués en vue de leur exploitation pour la modélisation.

II.2 Situation géographique

L'étude porte sur une région située dans l'Est de l'Algérie. Celle-ci s'étend de Béjaïa à l'ouest jusqu'à la frontière algéro-tunisienne à l'est, et de la mer Méditerranée au nord jusqu'à Biskra au sud. Géographiquement, le territoire étudié se situe entre les parallèles $36^{\circ} 54'$ et $34^{\circ} 48'$ de latitude Nord, et entre les méridiens $8^{\circ} 27'$ et $5^{\circ} 4'$ de longitude Est (figure 1). Sur le plan climatique, cette région se situe dans la zone de déplacement saisonnier des hautes pressions subtropicales. Elle est soumise à un basculement régulier entre le 30e et le 40e parallèle, avec un mouvement vers le sud en hiver et vers le nord en été. Le front polaire saisonnier, quant à lui, oscille entre les 35e et 65e parallèles, influençant ainsi le climat de l'Est algérien, notamment durant la période hivernale, considérée comme la plus froide de l'année.

Enfin, la position géographique du nord de l'Algérie induit une interaction marquée entre deux masses d'air : une circulation d'origine saharienne chaude et sèche en été, et une circulation polaire plus fraîche et humide en hiver. Ce contraste climatique se traduit par un climat à la fois aride en été et tempéré humide en hiver.

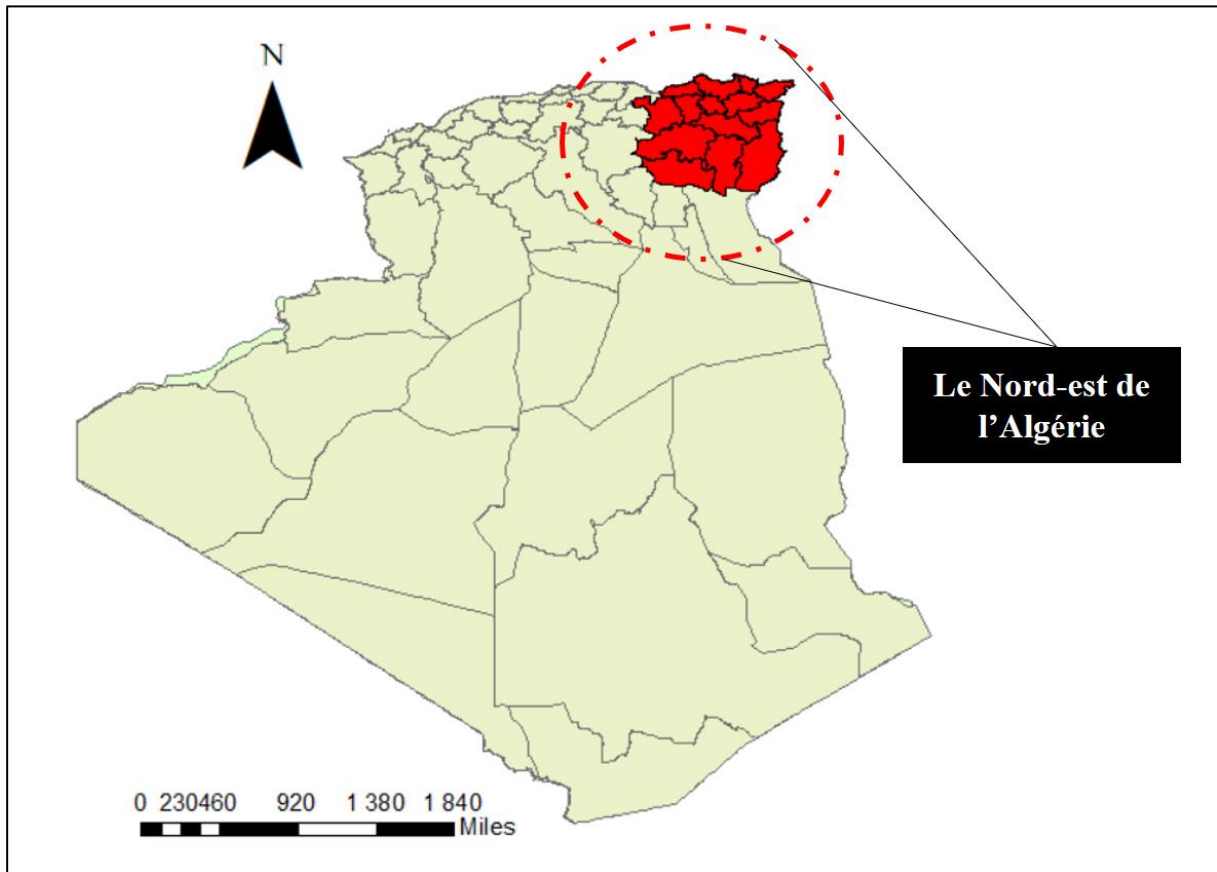


Figure.II.1 Situation géographique de la région d'étude dans l'Est algérien

II.3 Aperçu topographique

Sur le plan du relief, la région d'étude est composée principalement de cinq grandes unités physiques disposées du nord au sud comme indiqué ci-dessous.

II.3.1 Les plaines littorales

Les plaines littorales s'étendent de manière discontinue le long du littoral méditerranéen. À l'ouest, elles sont rares et peu développées, souvent interrompues par des reliefs. Ces zones basses, situées à l'embouchure des oueds, sont notamment représentées par les plaines de Collo (Oued Guebli), de Skikda (Oued Safsaf) et de Jijel (Oued Djen Djen).

En revanche, leur présence devient plus marquée dans la partie orientale de la région. On y distingue une vaste plaine en forme de croissant, s'étendant sur environ 100 km autour du Djebel Edough, depuis la forêt de Guerbes jusqu'à la région de Bouteldja. Cette configuration géographique leur confère une exposition favorable aux vents humides venus du nord et du nord-ouest.

II.3.2 L'Atlas tellien

L'Atlas tellien, composé du Tell interne et du Tell externe, s'étire selon un axe nord-est/sud-ouest. Il est formé par une succession de massifs côtiers et sublittoraux, depuis le bourrelet montagneux de Béjaïa à l'ouest jusqu'à la frontière tunisienne à l'est.

Les chaînes telliennes proches du littoral sont constituées des massifs de Collo, Skikda et de l'Edough. Plus au sud, on trouve les chaînes telliennes externes, comprenant notamment le massif des Babors, qui s'étend depuis la vallée de la Soummam jusqu'à Djen Djen, et le massif d'El Ouana à l'est. Le point culminant de cette zone est le Djebel Babor, atteignant une altitude de 2004 mètres.

II.3.4 Les hautes plaines

Les hautes plaines forment une large plateforme compartimentée, située entre le Tell au nord et les monts de Hodna, des Aurès et des Neméchas au sud. Elles s'étendent d'ouest en est et présentent des altitudes comprises entre 800 et 1200 mètres.

Cette région est caractérisée par la présence de nombreuses dépressions salées, chotts et sebkhas, en particulier celle de Ain M'lila. Selon Mébarki (2005), ces plaines se distinguent également par leur position topographique surélevée par rapport aux plaines littorales ou sahariennes.

II.3.5 L'Atlas saharien

L'Atlas saharien est constitué de massifs volumineux aux reliefs marqués, dont plusieurs sommets dépassent les 2000 mètres, notamment le Djebel Chelia (2328 m), le Djebel Mahmel (2321 m) et le Bélazma (2094 m). Ces alignements montagneux suivent une orientation sud-ouest/nord-est et s'étendent sur environ 60 kilomètres.

La zone sud de la région d'étude comprend la partie nord du chott Melrhir, situé à -34 mètres d'altitude. Cette vaste dépression fermée constitue un exutoire pour les oueds en provenance du versant méridional des Aurès et des Neméchas. L'oued Dejadi, large d'environ 500 mètres, s'y déverse, illustrant la dynamique d'écoulement nord-sud caractéristique de cette zone.

II.4 Végétation et sols

La couverture végétale de la région d'étude se caractérise par une diversité d'espèces adaptée aux contrastes climatiques. On y trouve des forêts de chênes et de pins, des arbustes de

garrigue comme les thuyas, les oliviers ou les arbres à encens, ainsi que des prairies semi-arides à arides, dominées par des espèces telles que l'armoïse et le chaume espagnol. L'indice de végétation (NDVI) de 2017, comparé à celui de 2004, met en évidence une forte couverture dans les zones forestières du nord-est, notamment à Collo, Kabylie et dans les Aurès-Nememchas, tandis que les zones désertiques et steppiques présentent des valeurs nettement plus faibles (Daif, 2025).

Les sols de la région reflètent l'influence du climat, du relief et de la nature des roches mères. On distingue deux grands groupes : les sols calcaires, issus de roches calcaires (sols calciques et décalcifiés), et les sols non-calcaires, formés sur substrat siliceux ou argileux, comprenant des sols insaturés et des sols podzoliques. Ces derniers, typiques des zones à forte humidité, sont acides, lessivés et pauvres en bases, souvent associés à une végétation de type forestier acidophile (ex. : hêtraies à sous-bois de bruyère ou de dyss). Les sols podzoliques se retrouvent surtout dans les régions nord de l'Algérie, notamment sur les grès de Numidie, en lien avec des précipitations annuelles supérieures à 500 mm (Duchaufour, 1975).

II.5 Aperçu climatique

Le nord-est de l'Algérie bénéficie d'un climat méditerranéen à tendance semi-aride, caractérisé par des étés chauds et secs et des hivers doux à modérément pluvieux. Cette configuration climatique s'explique par l'influence conjointe de la mer Méditerranée, du relief accidenté (notamment les chaînes de l'Atlas Tellien et Saharien), et de l'altitude. Les précipitations annuelles varient généralement entre 600 et 1 200 mm sur les zones côtières (ex. : Skikda, Annaba), tandis qu'elles peuvent descendre en dessous de 400 mm dans les hauts plateaux du Nord-Est (El Tarf, Souk Ahras) (ONM, 2022 ; Meddi & Hubert, 2003).

La période sèche estivale, typique du climat méditerranéen, s'étend sur 3 à 4 mois sur le littoral mais peut atteindre jusqu'à 6 mois dans les zones intérieures et semi-arides, où les influences sahariennes commencent à se faire sentir (Bessaoud, 2013). Contrairement à certaines régions côtières de même latitude comme la Californie, où les conditions locales façonnent largement le climat, le climat du Nord-Est algérien dépend fortement de circulations atmosphériques à grande échelle. En effet, les perturbations atlantiques, véhiculées par les jets d'altitude, constituent une source essentielle des précipitations hivernales. Ces systèmes, après avoir traversé l'Atlantique, sont déviés vers le bassin méditerranéen par les dorsales subtropicales et affectent ensuite le nord de l'Algérie (Meddi et al., 2010).

II.6 Caractéristiques générales du bassin de la Seybouse

II.6.1 Situation géographique

Le bassin versant de la Seybouse, d'une superficie de 6471 km², fait partie du grand bassin hydrographique dénommé CONSTANTINOIS-SEYBOUSE-MELLEQUE (ABH, 1996). Il est situé au nord-est de l'Algérie, au cœur de la Numidie, entre les longitudes 06°47'02" et 07°58'47" Est, et les latitudes 35°47'49" et 36°55'43" Nord.

Ce bassin s'étend vers le sud sur environ 160 km, atteignant une largeur maximale de 120 km au niveau du djebel Ouachh, entre la source de l'oued Semendou à l'ouest et l'oued El Kebir à l'est. Il englobe une partie de sept wilayas de l'Est algérien : Guelma, Annaba, Souk Ahras, El Tarf, Oum El Bouaghi, Constantine et Skikda. Le bassin comprend au total 68 communes, dont 30 sont entièrement incluses dans son périmètre, et 38 partiellement. Les principales agglomérations localisées dans le bassin sont Annaba, Guelma et Sedrata.

Les limites naturelles du bassin de la Seybouse sont les suivantes :

- Au nord : la mer Méditerranée,
- Au sud : les monts de Mahouna, Aïn Larbi et Sedrata,
- À l'ouest : le massif de l'Édough, ainsi que les monts de Houara et Débagh,
- À l'est : le prolongement oriental du système aquifère Annaba–Bouteldja et les monts de Nador N'Bail.

La population vivant dans le bassin est estimée à plus de 1,3 million d'habitants, répartis dans les 68 communes et les sept wilayas concernées. L'agglomération d'Annaba, à elle seule, concentre environ 18 % de cette population totale, ce qui en fait le pôle urbain majeur du bassin.

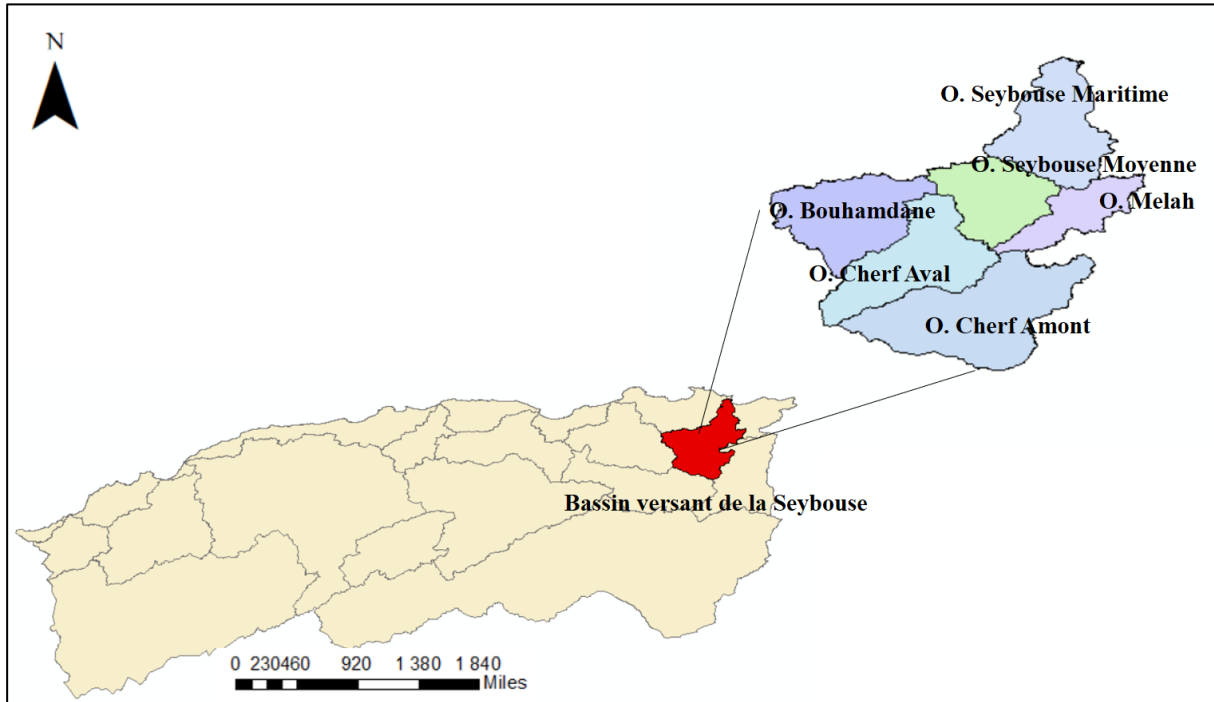


Figure II.2. Situation géographique du bassin versant de la seybouse en Algérie

II.6.2 Sous-bassins sélectionnés

Les sous-bassins retenus dans le cadre de cette étude ont été choisis en fonction de leur importance hydrologique au sein du bassin de la Seybouse, mais également en raison de la disponibilité et de l'accessibilité des données hydrométéorologiques les concernant. Ce critère a été déterminant pour permettre une analyse fiable et représentative du comportement hydrologique de la région.

II.6.2.1 Le sous-bassin d'Ain Berda

Situé en amont du bassin de la Seybouse, dans la wilaya d'Annaba, le sous-bassin d'Ain Berda présente un relief modérément montagneux, avec des altitudes dépassant les 600 mètres. Il joue un rôle essentiel en tant que zone de captation des eaux, particulièrement durant les mois pluvieux, contribuant ainsi de manière significative à l'alimentation de l'oued Seybouse (Hadji, 2017). La végétation semi-dense qui le recouvre favorise l'infiltration des eaux et limite les phénomènes d'érosion. Par ailleurs, son climat subhumide le rend hydrologiquement actif durant une grande partie de l'année.

II.6.2.2 Le sous-bassin de Mzadj Ammar

Ce sous-bassin, situé entre Guelma et Medjez Sfa, est dominé par un paysage collinéen et des zones agricoles intensivement exploitées. Il joue un rôle de régulation dans les

écoulements vers l'aval, grâce à un réseau hydrographique bien développé. Toutefois, cette zone est soumise à une pression anthropique importante, en lien avec l'agriculture et les usages domestiques de l'eau (Bourouba, 2014). Cette pression, combinée à une pluviométrie marquée par une forte variabilité interannuelle, influence significativement la dynamique des débits.

II.6.2.3 Le sous-bassin de Bouchegouf

Situé dans la wilaya de Guelma, le sous-bassin de Bouchegouf constitue l'un des secteurs les plus actifs sur le plan hydrologique. Il est traversé par de nombreux affluents secondaires qui se jettent directement dans l'oued Seybouse. Sa morphologie relativement ouverte, avec des pentes moyennes, favorise un écoulement rapide, ce qui accentue sa vulnérabilité face aux crues saisonnières (Djebbar, 2019). La combinaison de sols limoneux et d'une couverture végétale irrégulière renforce également le ruissellement superficiel.

Tableau II.1. Caractéristiques des bassins versants étudiés (Source ANRH, 2021)

Nom de station	Code station	S (km ²)	P (km)	Ic	H _{max}	H _{min}	H _{moy}	Ip	Dd (Km/Km ²)	C _t
MEDJEZ AMMAR	140301	1105	140		1326	295	785	0.13	2.6	
BOUCHEGOUF	140501	550	120	1.43	1317	95	641	0.15	4	25.2
AIN BERDA	140602	103	45	1.24	927	55	350	0.16	2.8	17.4

II.7. Données utilisées

Dans cette partie du travail, nous présenterons les données utilisées pour bien mener cette étude. Ainsi que, de procéder à un choix préalable des variables d'entrée pour les différents modèles utilisés, en intégrant différentes méthodes de sélection permettant un choix optimal des entrées.

II.7.1 Collecte des données

Dans le cadre de cette étude, trois types principaux de données ont été collectés : hydrométriques, pluviométriques et climatiques, toutes enregistrées à une résolution journalière, afin de permettre la modélisation du débit journalier en lien avec les précipitations.

La phase de collecte et de traitement des données constitue l'une des étapes les plus délicates et exigeantes en hydrologie. Les données hydrométriques exploitées proviennent de la banque de données de l'Agence Nationale des Ressources Hydrauliques (ANRH), dont les enregistrements sont organisés selon l'année hydrologique, allant du 1er septembre au 31 août.

Les données météorologiques, quant à elles, ont été obtenues auprès de l'Office National de la Météorologie (ONM).

La sélection des stations hydrométriques et pluviométriques utilisées dans ce travail repose sur des critères stricts de qualité, d'homogénéité et de disponibilité des séries sur des périodes récentes et communes. Le traitement initial des séries nécessite un travail de mise en forme et d'harmonisation, souvent long et complexe, compte tenu de la diversité et de la structure des bases de données informatisées de l'ANRH et de l'ONM.

La plupart des séries disponibles sont incomplètes, parfois de courte durée, et présentent des lacunes ou des valeurs aberrantes. Un traitement statistique préalable s'avère donc indispensable pour le comblement des données manquantes, la détection des valeurs anormales et l'homogénéisation des séries, en vue d'assurer la fiabilité des résultats de modélisation.

En complément des précipitations et des débits, l'évapotranspiration potentielle (ETP) a été calculée à l'aide de la formule de Penman, reconnue pour sa robustesse dans les études hydrologiques. Ce calcul intègre plusieurs paramètres climatiques, tels que la température, l'humidité relative, la vitesse du vent et le rayonnement solaire, permettant ainsi d'obtenir une estimation fiable de la demande évaporative de l'atmosphère sur la zone étudiée.

II.7.2 Stations sélectionnées

À l'issue d'un processus rigoureux de traitement et de validation des données, incluant notamment le comblement des lacunes et la vérification de la cohérence temporelle, un ensemble de stations a été retenu pour cette étude. Ces stations sont répertoriées dans le tableau Pour assurer une couverture spatiale et fonctionnelle optimale au sein de chaque sous-bassin étudié, trois types de stations ont été sélectionnés :

- Une station pluviométrique,
- Une station hydrométrique,
- Une station climatique.

Cette configuration permet de disposer, pour chaque secteur, de l'ensemble des paramètres nécessaires à la modélisation pluie-débit à l'échelle journalière.

Tableau II.2 : Stations choisies dans cette étude (Sources : ANRH, 2021 ; ONM, 2021)

Nom de la station hydrométrique	Code station	Station pluviométrique	Station climatique	Période
MEDJEZ AMMAR	140301	140313	GEULMA	2002-2012
BOUCHEGOUF	140505	140501	GEULMA	1985-1995
AIN BERDA	140602	140602	ANNABA	1987-1997

II.7.3 Caractéristiques statistiques des séries

La connaissance des paramètres statistiques des données revêt une importance capitale dans la modélisation hydrologique. Ces paramètres, tels que la moyenne, la variance, la distribution et les tendances, permettent de caractériser la variabilité des phénomènes hydrologiques. Comprendre ces éléments statistiques est fondamental pour calibrer et valider les modèles hydrologiques, assurant ainsi leur précision et leur fiabilité dans la prédiction des débits, des niveaux d'eau et des événements extrêmes. En utilisant ces informations statistiques, les modélisateurs peuvent mieux appréhender les schémas de comportement des cours d'eau, des bassins versants et des systèmes hydrologiques complexes, renforçant ainsi la capacité à élaborer des stratégies de gestion des ressources en eau plus efficaces et adaptées aux variations environnementales et climatiques. Les tableaux II.2 représentent les caractéristiques statistiques des séries collectées pas de temps journalier.

Tableau II. 3. Paramètres statistiques des données utilisées au pas de temps journalier

Station	Variables	N	Unit	Xmax	Xmean	Xmin	Sx	Cv
140501	P	3650	mm	87.778	2.751	0.000	6.915	2.514
	ETP	3650	mm	14.123	3.698	0.267	2.491	0.674
	Q	3650	m3/s	4.401	0.783	0.000	0.898	1.147
140301	P	3650	mm	85.500	1.481	0.000	5.097	3.442
	ETP	3650	mm	14.123	3.876	0.267	2.585	0.667
	Q	3650	m3/s	320.946	3.294	0.000	8.092	2.456
140601	P	3650	mm	305.690	2.550	0.000	14.769	5.792
	ETP	3650	mm	11.643	3.983	0.000	1.514	0.380
	Q	3650	m3/s	44.393	0.340	0.000	1.672	4.915

La sélection des variables d'entrée pertinentes pour la modélisation pluie-débit repose en grande partie sur l'analyse des relations temporelles entre les variables climatiques (pluie, évapotranspiration) et la variable cible (débit). Pour ce faire, trois outils statistiques ont été

mobilisés : l'auto-corrélation (ACF), l'auto-corrélation partielle (PACF) et la corrélation croisée (XCF ou CCF).

- ACF (Auto-Correlation Function) : Elle mesure la corrélation entre une variable et ses propres valeurs passées. Elle permet d'identifier les lags (décalages temporels) les plus influents dans une série temporelle, comme les valeurs passées du débit pouvant expliquer sa valeur actuelle.
- PACF (Partial Auto-Correlation Function) : Elle complète l'ACF en évaluant la corrélation entre une variable et ses lags en éliminant l'effet des lags intermédiaires. Cela aide à repérer les valeurs vraiment indépendantes qui ont un impact direct sur la variable cible.
- XCF (Cross-Correlation Function) : Elle mesure la corrélation entre deux séries temporelles décalées dans le temps. Cette fonction est utilisée pour évaluer la relation entre les précipitations ou l'évapotranspiration et le débit observé à différentes échelles de temps. Elle permet ainsi de déterminer combien de jours après une pluie (par exemple) un effet sur le débit est observé.

Grâce à ces analyses, les décalages temporels optimaux ont été identifiés pour chaque variable, ce qui permet de construire une structure d'entrée efficace pour le modèle. Typiquement, les valeurs de pluie, de débit et d'évapotranspiration aux temps t-1, t-2 ont montré une influence significative sur le débit journalier, et ont donc été retenues comme variables d'entrée.

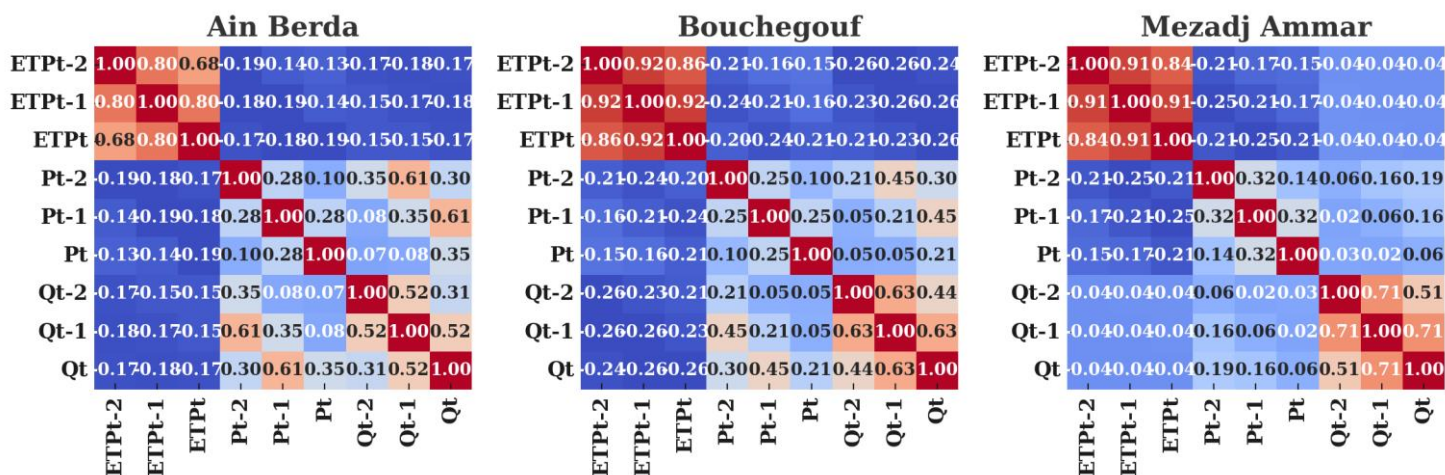


Figure II.3. Matrices de corrélation entre les variables climatiques et hydrologiques pour les stations d'Ain Berda, Bouchegouf et Mezadj Ammar

Tableau II.4. Les combinaisons entrées pour différents modèles

Scénario	Combinaison des variables d'entrée	Sortie
1	$ET0(t-2), ET0(t-1), ET0(t), P(t-2), P(t-1), P(t), Q(t-2), Q(t-1)$	$Q(t)$
2	$P(t-2), P(t-1), P(t), Q(t-2), Q(t-1)$	$Q(t)$
3	$ET0(t-2), ET0(t-1), ET0(t), P(t-2), P(t-1), P(t)$	$Q(t)$
4	$Q(t-2), Q(t-1)$	$Q(t)$
5	$ET0(t), P(t), Q(t-2), Q(t-1)$	$Q(t)$
6	$ET0(t), P(t),$	$Q(t)$

II.8 Conclusion

Ce chapitre a présenté les principales caractéristiques physiques et hydrologiques du bassin de la Seybouse, ainsi que les sous-bassins retenus pour l'étude. Il a également décrit le processus de collecte, de traitement et de sélection des données nécessaires à la modélisation. Enfin, les variables d'entrée ont été choisies à l'aide d'analyses statistiques (ACF, PACF, XCF), en vue d'une modélisation efficace dans le chapitre suivant.

CHAPITRE III
MATERIELS ET
METHODES

CHAPITRE III : MATERIELS ET METHODES

III .1 Introduction

Ce chapitre décrit la méthodologie adoptée pour la mise en œuvre des modèles d'intelligence artificielle (ANN, LSTM, CNN, GRU) appliqués à la prévision du débit journalier. La démarche repose sur une série d'étapes allant de la collecte et la préparation des données à l'évaluation des performances des modèles. Elle intègre également des outils d'analyse statistique visant à optimiser le choix des variables d'entrée.

III .2 Étapes de mise en œuvre de l'étude

La méthodologie adoptée dans cette étude repose sur une série d'étapes logiques et complémentaires visant à développer un modèle performant de prévision du débit journalier à partir de variables hydrométéorologiques. Le travail a débuté par la collecte des données issues de trois stations hydrométriques du bassin versant de la Seybouse, incluant les précipitations, l'évapotranspiration potentielle et les débits journaliers (Figure.III.1). Ces données ont été soumises à un prétraitement rigoureux, comprenant la correction des valeurs manquantes, la normalisation via la méthode Min-Max Scaling, ainsi que la division en deux ensembles distincts : un ensemble d'entraînement et un ensemble de validation. Une analyse statistique préliminaire a ensuite été menée à l'aide des fonctions d'autocorrélation (ACF), d'autocorrélation partielle (PACF) et de corrélation croisée (XCF) afin de déterminer les retards temporels optimaux à intégrer comme variables d'entrée dans les modèles. La phase de modélisation a consisté en la conception, l'entraînement et la validation de quatre architectures d'intelligence artificielle : les réseaux de neurones artificiels (ANN), les réseaux récurrents à mémoire longue (LSTM), les réseaux GRU et les réseaux convolutifs (CNN). Ces modèles ont été implémentés en Python à l'aide des bibliothèques Keras et TensorFlow, avec un ajustement des hyperparamètres basé sur les performances observées sur l'ensemble de validation. Enfin, les résultats ont été analysés et comparés à l'aide d'indicateurs statistiques tels que le RMSE, le MAE, le coefficient (R) et l'indice de Nash-Sutcliffe (NSE), afin d'évaluer la précision, la robustesse et la capacité des modèles à fournir des prévisions fiables dans un contexte semi-aride, dans le but de soutenir la gestion durable des ressources en eau.

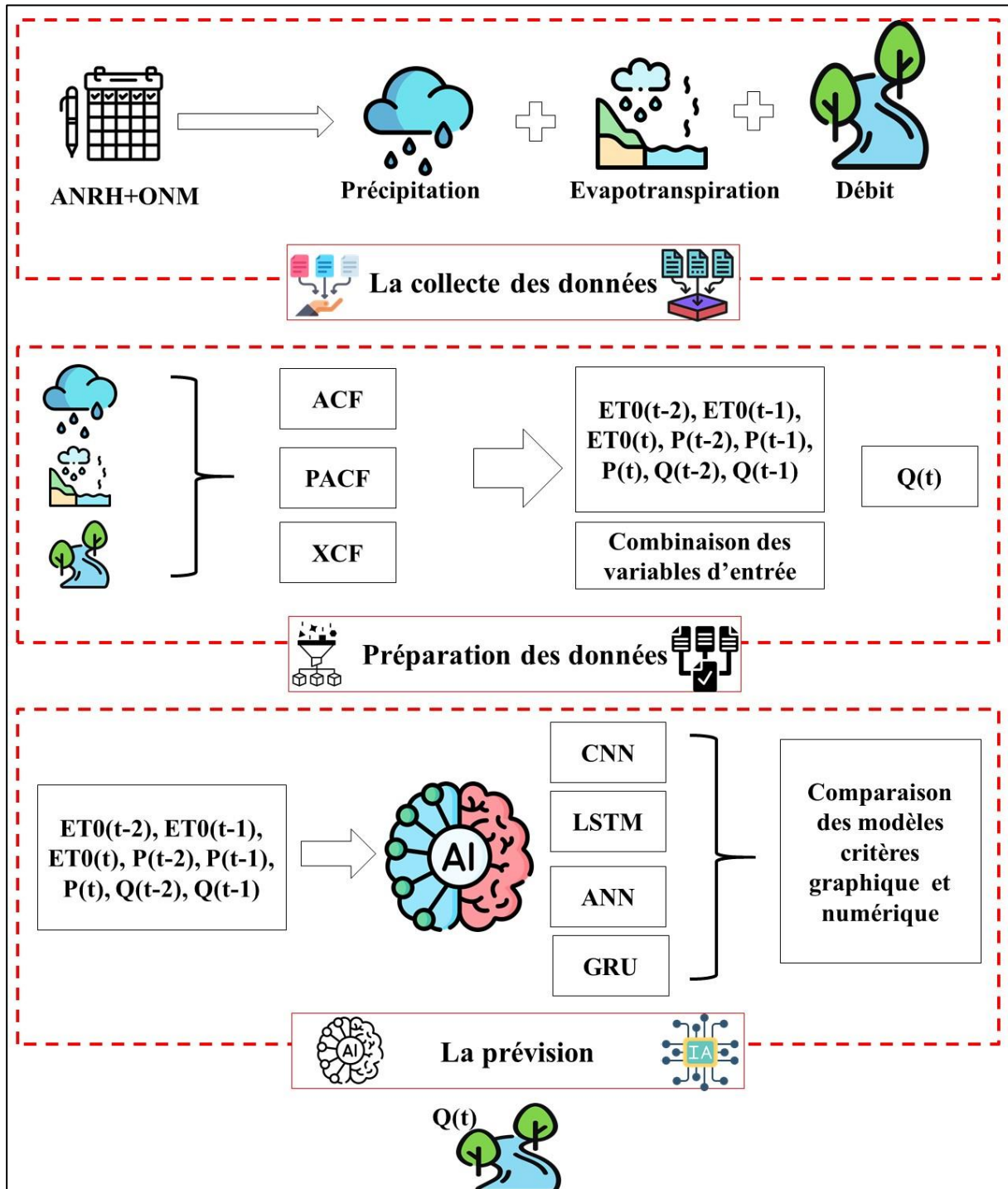


Figure.III.1 Schéma de la démarche méthodologique pour la prévision du débit journalier à l'aide de modèles d'intelligence artificielle

III .3 Modèles d'intelligence artificielle appliqués à la prévision du débit

Dans le cadre de cette étude, quatre modèles d'intelligence artificielle ont été sélectionnés pour leur capacité à modéliser les relations complexes et non linéaires entre les variables hydrométéorologiques : les réseaux de neurones artificiels (ANN), les réseaux de

neurones à mémoire longue (LSTM), les réseaux de neurones à mémoire courte (GRU), et les réseaux de neurones convolutifs (CNN). Ces modèles, issus des avancées récentes en apprentissage profond, offrent chacun des approches distinctes pour traiter les séries temporelles hydrologiques. Leur comparaison a pour objectif d'identifier la méthode la plus adaptée à la prévision du débit journalier dans un contexte semi-aride, caractérisé par une forte variabilité climatique et une disponibilité limitée des données.

III .3.1 Réseaux de neurones artificiels (ANN)

Les réseaux de neurones artificiels (ANN) sont des modèles inspirés du fonctionnement du cerveau humain, conçus pour apprendre automatiquement à partir de données, en identifiant des relations complexes, souvent non linéaires, entre les variables. Leur origine remonte aux travaux fondateurs de McCulloch et Pitts (1943), qui ont proposé le premier modèle de neurone formel. Cependant, c'est avec le développement du Perceptron Multicouche (Multilayer Perceptron – MLP) et de l'algorithme de rétropropagation de l'erreur par Rumelhart, Hinton et Williams (1986) que les ANN ont pris leur essor dans les domaines de la modélisation et de la prévision.

Le MLP est un réseau entièrement connecté composé de plusieurs couches : une couche d'entrée qui reçoit les variables explicatives (par exemple, précipitations, évapotranspiration, débits décalés), une ou plusieurs couches cachées où s'effectue le traitement non linéaire de l'information, et une couche de sortie qui fournit la variable cible (ici, le débit journalier prédit). Chaque neurone d'une couche applique une fonction d'activation (souvent ReLU, tanh ou sigmoïde) à une combinaison pondérée de ses entrées, suivie d'un biais. Le réseau apprend à ajuster les poids (coefficients des connexions) et les biais via un processus itératif appelé apprentissage supervisé (Figure.III.2).

Le fonctionnement d'un MLP peut être résumé en plusieurs étapes clés :

1. Propagation avant (forward propagation) : les données d'entrée traversent le réseau couche par couche, jusqu'à produire une sortie estimée.
2. Calcul de l'erreur : la différence entre la sortie prédite et la valeur réelle est mesurée à l'aide d'une fonction de coût. L'erreur quadratique moyenne (MSE) est couramment utilisée pour les problèmes de régression hydrologique.

3. Rétropropagation de l'erreur (backpropagation) : l'erreur est propagée en sens inverse, des sorties vers les entrées, pour calculer les gradients des poids.
4. Mise à jour des paramètres : les poids et biais sont ajustés à l'aide d'un algorithme d'optimisation, tel que Gradient Descent, Adam ou RMSprop, en fonction du taux d'apprentissage (learning rate).

Parmi les paramètres essentiels à configurer dans un MLP, on peut citer :

- Le nombre de couches cachées et le nombre de neurones par couche,
- Le type de fonction d'activation,
- Le taux d'apprentissage,
- Le nombre d'époques (epochs), c'est-à-dire le nombre de fois où le réseau parcourt l'ensemble des données,
- La taille du lot (batch size), qui détermine combien d'exemples sont traités avant chaque mise à jour des poids.

Les ANN, et particulièrement les MLP, ont été largement utilisés dans les applications hydrologiques pour la prévision des débits, notamment grâce à leur capacité à capturer les dynamiques complexes entre les variables climatiques et hydrologiques sans nécessiter de formulation physique explicite. Plusieurs études (Dawson & Wilby, 2001 ; Maier et al., 2010 ; Abraham et al., 2012) ont démontré leur efficacité, notamment dans les bassins où les données sont limitées ou bruitées.

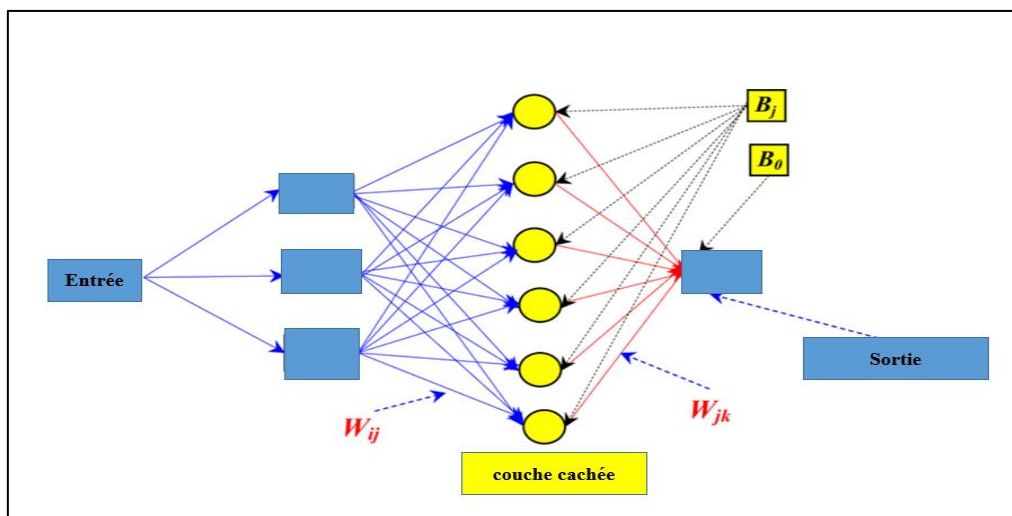


Figure.III.2 Architecture d'un Perceptron Multicouche (MLP)

III .3.2 Réseaux de neurones à mémoire longue (LSTM)

Les réseaux de neurones à mémoire longue, ou LSTM (Long Short-Term Memory), sont une variante avancée des réseaux de neurones récurrents (RNN) conçue pour traiter les données séquentielles et séries temporelles, en surmontant les limitations des RNN classiques liées à la disparition ou l'explosion du gradient lors de l'apprentissage à long terme. Introduits par Hochreiter et Schmidhuber (1997), les LSTM ont révolutionné le traitement de séquences en permettant aux modèles de retenir des informations pertinentes sur de longues périodes, ce qui les rend particulièrement adaptés aux phénomènes hydrologiques où les effets retardés et la mémoire des événements passés sont essentiels.

Contrairement aux MLP qui traitent les données de manière statique, le LSTM conserve un état interne, appelé état de cellule, qui agit comme une mémoire dynamique à long terme. Il utilise une structure composée de trois portes :

- La porte d'oubli (forget gate) décide quelles informations de l'état précédent doivent être supprimées,
- La porte d'entrée (input gate) détermine quelles nouvelles informations doivent être ajoutées à la mémoire,
- La porte de sortie (output gate) contrôle quelles parties de la mémoire sont utilisées pour générer la sortie à chaque pas de temps.

Ce mécanisme permet de modéliser des dépendances temporelles longues et non linéaires entre les observations passées (comme les précipitations ou les débits précédents) et les valeurs futures à prédire (Figure.III.3).

Le fonctionnement du LSTM se déroule en plusieurs étapes clés :

1. À chaque instant t , le réseau reçoit une entrée (ex. : précipitation du jour) et l'état caché précédent h_{t-1} .
2. La porte d'oubli f_t calcule la proportion de l'ancienne mémoire C_{t-1} à conserver.
3. La porte d'entrée i_t détermine quelles nouvelles informations C_t doivent être ajoutées à la cellule.
4. La mémoire de la cellule est mise à jour via ;

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad \text{III.1}$$

5. Enfin, la porte de sortie O_t génère la sortie actuelle h_t , en fonction de la mémoire actualisée.

LSTM possède plusieurs paramètres clés à configurer :

- Le nombre d'unités LSTM (ou neurones par couche),
- Le nombre de couches empilées,
- La fonction d'activation (souvent tanh pour les états de cellule et sigmoid pour les portes),
- Le taux d'apprentissage, le batch size, et le nombre d'époques,
- Un dropout facultatif pour éviter le surapprentissage.

Dans le domaine hydrologique, les LSTM se sont révélés particulièrement efficaces pour la prévision du débit journalier, car ils capturent naturellement les effets retardés des précipitations et de l'humidité du sol sur le ruissellement. Plusieurs études ont démontré leurs performances supérieures aux modèles classiques et aux ANN standards dans des contextes de prévision à court ou moyen terme (Kratzert et al., 2019 ; Feng et al., 2020 ; Le et al., 2021). De plus, ils s'adaptent bien aux environnements semi-arides où les dynamiques hydrologiques sont complexes et fortement influencées par la variabilité climatique.

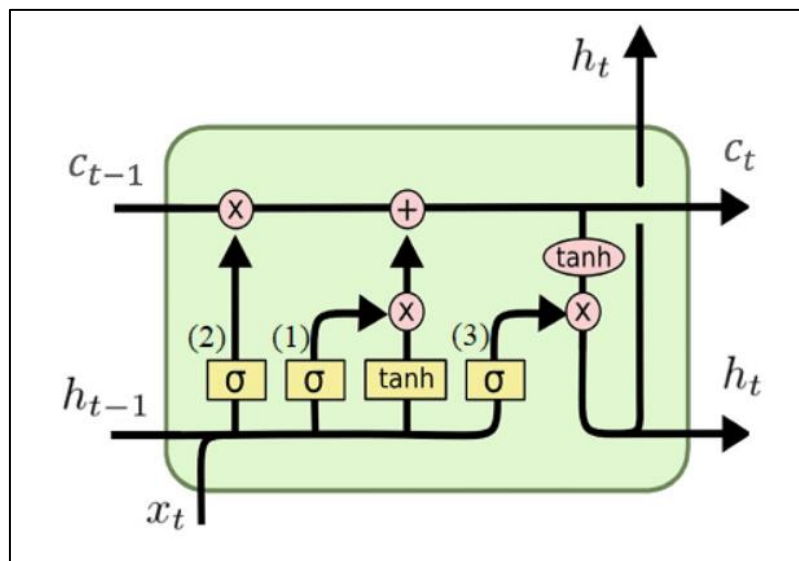


Figure.III.3. Structure interne d'une cellule LSTM (Long Short-Term Memory)

III .3.3 Réseaux de neurones à unité récurrente à portes (GRU)

Le réseau GRU (*Gated Recurrent Unit*), introduit par Cho et al. (2014), est une architecture de réseaux neuronaux récurrents (RNN) développée comme alternative simplifiée au LSTM. Il a été conçu pour surmonter les limites des RNN classiques tout en étant moins complexe que le LSTM tout en conservant une capacité comparable à modéliser des dépendances temporelles longues. Le GRU a gagné en popularité dans de nombreuses applications séquentielles, y compris la modélisation hydrologique, en raison de sa structure plus légère, d'un temps d'entraînement réduit, et de performances compétitives, notamment lorsque les séries temporelles sont relativement courtes ou peu bruitées (Figure.III.4).

Le fonctionnement du GRU repose sur deux portes principales qui contrôlent le flux d'information dans le réseau :

- La porte de mise à jour (update gate), qui détermine dans quelle mesure l'information de l'état précédent est conservée,
- La porte de réinitialisation (reset gate), qui contrôle dans quelle mesure l'information précédente est oubliée.

Ces mécanismes permettent au réseau d'équilibrer efficacement mémoire à court terme et mémoire à long terme. Contrairement au LSTM, le GRU n'a pas de mémoire de cellule distincte ; son état caché unique joue le double rôle de mémoire et de sortie, ce qui rend sa structure plus simple et plus rapide à entraîner.

Le processus d'apprentissage d'un GRU se déroule comme suit :

1. À chaque pas de temps, le réseau reçoit une entrée x_t (ex. : pluie, débit retardé).
2. La porte de réinitialisation r_t décide du poids à accorder à l'état précédent dans la transformation de l'entrée.
3. Un état candidat \tilde{h}_t est généré à partir de l'entrée et de l'état précédent réinitialisé.
4. La porte de mise à jour z_t détermine dans quelle mesure ce nouvel état doit remplacer l'ancien.
5. Le nouvel état caché h_t est calculé comme une interpolation entre l'ancien état et l'état candidat:

$$h_t = z_t \cdot h_{t-1} + (1 - z_t) \cdot \tilde{h}_t \quad \text{III.2}$$

Les paramètres essentiels à configurer dans un GRU sont similaires à ceux du LSTM :

- Le nombre d'unités GRU (ou neurones),
- Le nombre de couches empilées (GRU profonds),
- La fonction d'activation (souvent tanh et sigmoid),
- Le taux d'apprentissage, le nombre d'époques, le batch size,
- Une régularisation par dropout ou early stopping si nécessaire.

Dans les études hydrologiques, les GRU ont démontré de bonnes performances dans la prévision du débit, tout en étant plus rapides à entraîner que les LSTM. Ils sont particulièrement efficaces lorsque les séries de données sont relativement continues et qu'une bonne capture des variations de court à moyen terme est suffisante. Des recherches récentes ont mis en évidence leur capacité à rivaliser avec les LSTM dans des tâches de prévision de débit journalier, notamment dans les régions où les régimes hydrologiques sont irréguliers, comme les bassins semi-arides (Tang et al., 2024 ; Sit et Demir, 2021).

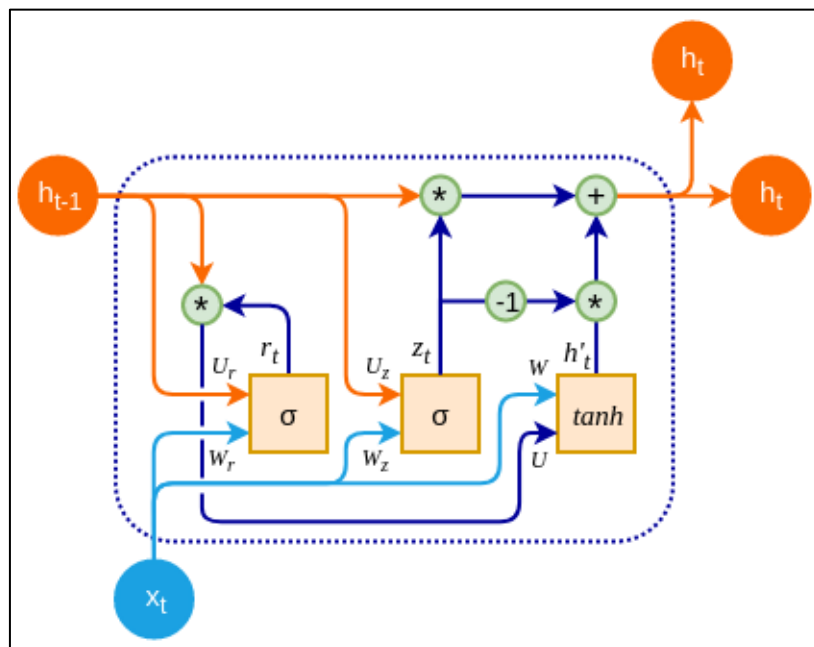


Figure.III.4 Architecture interne d'une cellule GRU (Gated Recurrent Unit)

III .3.4 Réseaux de neurones convolutifs (CNN)

Les réseaux de neurones convolutifs (CNN – *Convolutional Neural Networks*) sont initialement conçus pour le traitement des images, mais ils ont progressivement été adaptés à d'autres types de données structurées, notamment les séries temporelles, grâce à leur capacité à extraire automatiquement des caractéristiques locales et spatiales pertinentes. Introduits par Le Cun et al. (1989) pour la reconnaissance de caractères manuscrits, les CNN ont connu un essor majeur avec les succès dans la vision par ordinateur (Krizhevsky et al., 2012), puis ont été appliqués à la modélisation hydrologique, où ils se sont révélés efficaces pour capturer des motifs temporels dans les variables d'entrée comme les précipitations, l'évapotranspiration ou les débits passés.

Contrairement aux MLP ou aux RNN, les CNN n'utilisent pas de connexions entièrement connectées entre tous les neurones. À la place, ils appliquent des filtres (ou noyaux de convolution) qui glissent sur l'entrée (dans ce cas, la série temporelle), pour détecter des motifs ou des tendances locales. Ce mécanisme permet au modèle d'être plus léger en paramètres, tout en capturant les dépendances locales dans le temps, ce qui est crucial pour identifier des événements hydrologiques typiques, comme des épisodes de pluie-intensité ou des pics de débit (Figure.III.5).

Un CNN standard pour séries temporelles, comme utilisé dans la prévision hydrologique, se compose généralement des étapes suivantes :

1. Couche de convolution 1D : des filtres appliquent des opérations de convolution sur la série temporelle pour extraire des caractéristiques (features) locales pertinentes.
2. Fonction d'activation : ReLU est souvent utilisée pour introduire de la non-linéarité.
3. Couche de pooling (optionnelle) : souvent un *max pooling* qui réduit la dimensionnalité et conserve les caractéristiques les plus importantes.
4. Couches fully connected : une ou plusieurs couches denses finales qui transforment les caractéristiques extraites en prédictions du débit.
5. Sortie : un neurone en sortie qui donne la valeur prédite du débit journalier.

Les paramètres importants à configurer dans un CNN incluent :

- Le nombre de filtres (kernels),

- La taille des filtres (window size),
- Le pas (stride) et le padding,
- Le nombre de couches convolutionnelles,
- Les fonctions d'activation, taux d'apprentissage, batch size, nombre d'époques, etc.

Les CNN se sont montrés particulièrement utiles pour la prévision hydrologique à court terme, surtout lorsque les signaux d'entrée présentent des motifs répétitifs, comme les cycles pluviométriques (Zhao et al., 2019 ; Ghimire et al., 2021).

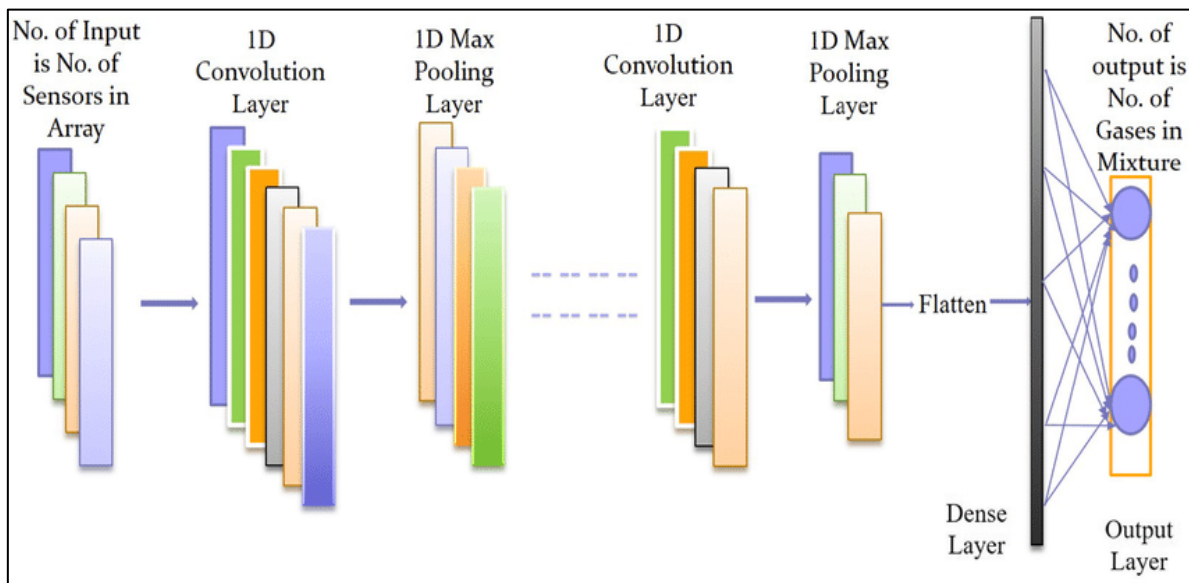


Figure.III.5 Architecture typique d'un réseau de neurones convolutif (CNN 1D) pour la modélisation de séries temporelles

III .4 Évaluation des performances des modèles

L'évaluation des performances des modèles de prévision constitue une étape essentielle pour juger de leur précision, de leur robustesse et de leur capacité de généralisation. Dans le cadre de cette étude, deux catégories complémentaires de critères ont été mobilisées : les indicateurs numériques (quantitatifs) et les représentations graphiques (visuelles). Ces outils permettent d'avoir une lecture complète des résultats obtenus, en alliant rigueur statistique et interprétation visuelle.

III .4.1 Critères numériques de performance

Les critères numériques sont des indicateurs quantitatifs utilisés pour mesurer la précision, la fiabilité et la capacité de généralisation des modèles. Voici les cinq indicateurs retenus dans cette étude.

– **Coefficient de corrélation (R) :**

$$R = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \quad \text{III.3}$$

- O_i : valeur observée
- P_i : valeur prédite
- \bar{O}, \bar{P} : moyennes des valeurs observées et prédites

Interprétation :

$R \in [-1,1]$. Plus R est proche de 1 , plus la corrélation entre les valeurs observées et prédites est forte. Une valeur proche de 0 indique une absence de corrélation.

– **Nash-Sutcliffe Efficiency (NSE) :**

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad \text{III.4}$$

Interprétation :

NSE varie de $-\infty$ à 1 . Une valeur de 1 indique une prédiction parfaite. Une valeur entre 0 et 1 signifie que le modèle est meilleur que la moyenne observée, tandis qu'une valeur < 0 indique que le modèle est moins performant qu'une prédiction basée simplement sur la moyenne.

– **Erreur quadratique moyenne (RMSE) :**

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad \text{III.5}$$

Interprétation :

Le RMSE mesure la dispersion des erreurs de prédiction. Plus sa valeur est faible, meilleure est la précision. Il est sensible aux grandes erreurs.

– **Erreur absolue moyenne (MAE) :**

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad \text{III.5}$$

Interprétation :

Le MAE donne la moyenne des erreurs absolues. Il est plus robuste que le RMSE vis-à-vis des valeurs extrêmes. Plus il est faible, meilleure est la précision.

– **Kling-Gupta Efficiency (KGE) :**

$$KGE = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad \text{III.5}$$

Avec :

- r : corrélation (coefficient R) ;
- $\beta = \frac{\bar{P}}{\bar{O}}$: biais relatif ;
- $\gamma = \frac{CV_P}{CV_O}$: rapport des coefficients de variation ;
- $CV = \frac{\sigma}{\mu}$.

Interprétation :

Le KGE combine précision, biais et variabilité. Une valeur de 1 représente une performance parfaite. Il est souvent préféré au NSE pour une évaluation plus équilibrée.

III .4.2 Critères graphiques de visualisation des résultats

En complément des indicateurs numériques, plusieurs représentations graphiques ont été mobilisées afin d'analyser visuellement les performances des modèles et de mieux interpréter leurs comportements. Le nuage de points (scatter plot) permet de comparer directement les valeurs observées et prédites à travers un graphique $y=x$ utile pour détecter les biais systématiques ou les écarts importants. Les box plots et violin plots ont été utilisés pour illustrer la distribution des erreurs ou des valeurs simulées par chaque modèle, en mettant en évidence la dispersion, la symétrie ou la présence de valeurs aberrantes. Les courbes de

variation temporelle permettent de suivre l'évolution conjointe des débits simulés et observés au fil du temps, offrant une vue directe sur la capacité du modèle à suivre les dynamiques hydrologiques, notamment lors des crues ou des étiages. Le barplot a servi à comparer visuellement les scores des différents indicateurs de performance (NSE, RMSE, KGE, etc.) entre les modèles, facilitant l'identification du plus performant selon chaque critère. La carte thermique (heatmap) a été utilisée pour synthétiser les résultats sous forme matricielle, en représentant les performances de chaque modèle selon les stations ou les variables, ce qui facilite une analyse croisée. Enfin, le diagramme de Taylor offre une visualisation synthétique très utile en hydrologie : il combine dans un seul graphique le coefficient de corrélation, l'écart-type et l'erreur quadratique normalisée, permettant une comparaison simultanée et intuitive de l'ensemble des modèles testés. L'utilisation de ces outils graphiques vient ainsi enrichir l'évaluation globale, en confirmant ou nuancant les résultats numériques par des tendances visuelles claires.

III .5 Outils et logiciels utilisés

La réalisation de ce travail s'est appuyée sur une combinaison de logiciels spécialisés, choisis en fonction des besoins spécifiques liés à l'organisation, à la modélisation et à la visualisation des données hydrologiques.

Pour la programmation des modèles d'intelligence artificielle (ANN, LSTM, GRU, CNN), le langage **Python** a été utilisé en raison de sa souplesse, de sa richesse en bibliothèques scientifiques et de sa large adoption dans le domaine de l'apprentissage automatique. Les principales bibliothèques mobilisées sont NumPy et Pandas pour le traitement des données, Matplotlib et Seaborn pour la visualisation, Scikit-learn pour les prétraitements statistiques, et TensorFlow/Keras pour la construction et l'entraînement des réseaux de neurones.

Pour l'organisation des données brutes (pluviométrie, évapotranspiration, débits), des traitements préliminaires ont été réalisés à l'aide de **Microsoft Excel**, notamment pour la structuration des séries chronologiques, la détection des valeurs manquantes et les premières statistiques descriptives.

La cartographie du bassin versant et la localisation des stations hydrométriques ont été réalisées avec le logiciel **ArcGIS**, un système d'information géographique (SIG) puissant, qui a permis de produire des cartes précises et personnalisées. Ces cartes ont été essentielles pour

la compréhension du contexte spatial de l'étude et pour l'interprétation géographique des résultats obtenus.

L'ensemble de ces outils a contribué à assurer une chaîne de traitement complète, allant de la collecte et la structuration des données jusqu'à la modélisation, l'évaluation des performances et la présentation des résultats dans un format clair et reproductible.

III .6 Conclusion

Ce chapitre a présenté de manière détaillée la démarche méthodologique adoptée pour la mise en œuvre de la prévision du débit journalier dans le bassin versant de la Seybouse à l'aide de modèles d'intelligence artificielle. Après la collecte et le traitement rigoureux des données hydrométéorologiques, une sélection optimale des variables d'entrée a été réalisée à l'aide d'outils statistiques comme l'ACF, la PACF et la XCF. Quatre architectures de réseaux de neurones ont été développées – ANN, LSTM, GRU et CNN – en utilisant l'environnement Python, chacune répondant à des logiques différentes de modélisation des séries temporelles. Leur performance sera évaluée selon une approche multi-critères combinant des indicateurs numériques (R, NSE, RMSE, MAE, KGE) et des outils graphiques (scatter plot, box plot, courbe de variation, Taylor diagram, etc.). Cette méthodologie rigoureuse et diversifiée offre une base solide pour la comparaison des modèles, dont les résultats seront analysés en détail dans le chapitre suivant.

CHAPITRE IV
RÉSULTATS ET
DISCUSSIONS

CHAPITRE IV : RÉSULTATS ET DISCUSSIONS

V.1 Introduction

Dans ce chapitre, nous présentons et analysons les résultats des modèles ANN, GRU, LSTM et CNN appliqués aux trois stations hydrométriques du bassin versant de Seybouse. L'évaluation des performances repose sur plusieurs métriques statistiques : le coefficient de corrélation (R) pour mesurer l'association entre les débits prédits et observés, le Nash-Sutcliffe Efficiency (NSE) pour évaluer l'efficacité du modèle à reproduire la variabilité des observations, le Kling-Gupta Efficiency (KGE) pour une évaluation globale intégrant corrélation, biais et variabilité, ainsi que l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE) pour quantifier les écarts entre valeurs simulées et réelles. Des analyses graphiques, notamment des courbes de prédiction et des diagrammes de dispersion, permettent de visualiser les tendances et les écarts entre simulations et observations. Cette étude vise à identifier le modèle le plus performant, à analyser ses forces et ses limites, et à proposer des améliorations pour optimiser la précision des prévisions hydrologiques.

IV.2 Analyse et Interprétation des Résultats

Dans cette section, nous présentons et analysons les résultats obtenus pour les trois stations hydrométriques du bassin versant de Seybouse. L'évaluation des performances des modèles ANN, GRU, LSTM et CNN est réalisée à l'aide des indicateurs statistiques et des analyses graphiques afin de comparer leur précision et leur fiabilité dans la prévision du débit. La base de données a été divisée en deux sous-ensembles : 70 % des données sont utilisées pour l'apprentissage des modèles et 30 % pour la validation, permettant ainsi d'évaluer leur capacité de généralisation et leur robustesse.

Tableau.IV.1. Performances des différents modèles pour la station MEDJEZ AMAR

Model	Apprentissage					Validation				
	R	NSE	MAE	RMSE	KGE	R	NSE	MAE	RMSE	KGE
ANN1	0.991	0.983	0.471	1.247	0.938	0.986	0.961	0.394	0.573	0.916
ANN2	0.986	0.970	0.573	1.633	0.919	0.957	0.899	0.640	0.923	0.883
ANN3	0.850	0.717	2.456	5.031	0.691	0.137	0.019	3.088	3.745	0.037
ANN4	0.742	0.550	0.574	6.346	0.634	0.592	0.351	2.583	3.170	0.410
ANN5	0.921	0.821	1.615	4.000	0.509	0.463	0.214	3.040	3.729	0.257
ANN6	0.404	0.154	2.849	8.704	0.141	0.370	0.137	2.901	3.567	0.123
GRU1	0.989	0.978	0.500	1.417	0.937	0.981	0.963	0.284	0.560	0.974

GRU2	0.992	0.984	0.454	1.179	0.947	0.977	0.954	0.321	0.624	0.949
GRU3	0.859	0.734	2.380	4.875	0.762	0.314	0.099	2.985	3.753	0.068
GRU4	0.740	0.548	0.635	6.363	0.621	0.712	0.239	2.052	2.536	0.556
GRU5	0.780	0.597	1.560	6.007	0.747	0.554	0.320	2.702	3.339	0.365
GRU6	0.453	0.204	2.830	8.443	0.210	0.231	0.053	2.913	3.849	0.094
LSTM1	0.991	0.980	0.697	1.324	0.875	0.984	0.958	0.422	0.599	0.885
LSTM2	0.993	0.985	0.350	1.142	0.976	0.985	0.963	0.354	0.559	0.913
LSTM3	0.807	0.586	2.605	6.086	0.511	0.657	0.014	2.304	2.886	0.472
LSTM4	0.737	0.542	0.780	6.400	0.612	0.491	0.241	2.938	3.596	0.306
LSTM5	0.854	0.723	0.899	4.978	0.723	0.846	0.673	0.960	1.662	0.771
LSTM6	0.535	0.240	3.046	8.249	0.165	0.270	0.073	2.837	3.322	0.088
CNN1	0.993	0.984	0.620	1.198	0.884	0.983	0.958	0.357	0.599	0.921
CNN2	0.992	0.984	0.448	1.184	0.945	0.968	0.937	0.247	0.732	0.954
CNN3	0.867	0.694	2.499	5.231	0.762	0.278	0.077	4.197	5.212	0.081
CNN4	0.744	0.551	0.643	6.338	0.601	0.732	0.323	1.940	2.391	0.598
CNN5	0.818	0.963	0.488	1.826	0.942	0.599	0.171	2.552	3.145	0.407
CNN6	0.485	0.233	2.911	8.286	0.274	0.298	0.089	4.094	4.848	0.081

L'analyse des performances des modèles ANN, GRU, LSTM et CNN pour la station 140301 (Tableau.IV.1 et Figure IV.1) repose sur plusieurs indicateurs statistiques. Les modèles ANN présentent des résultats très variables. ANN1 est le plus performant parmi eux, avec $R = 0.986$, $NSE = 0.961$, $KGE = 0.916$, $MAE = 0.394$ et $RMSE = 0.573$. Il montre une bonne corrélation avec les données observées et une erreur relativement faible. ANN2 suit avec des performances légèrement inférieures, mais reste un modèle relativement efficace. En revanche, ANN3, ANN4, ANN5 et ANN6 affichent des performances médiocres, avec des valeurs très faibles de NSE et KGE, indiquant une incapacité à modéliser correctement le débit. ANN3 est le pire modèle, avec $NSE = 0.019$ et $R = 0.137$, suggérant qu'il ne capture presque aucune relation entre les données d'entrée et la sortie. Les modèles GRU montrent une excellente capacité de prédiction. GRU1 est le modèle le plus performant de tous, avec $R = 0.981$, $NSE = 0.963$, $KGE = 0.974$, $MAE = 0.284$ et $RMSE = 0.560$. Ces résultats indiquent une forte corrélation entre les valeurs prédictives et observées, avec une erreur minimale, prouvant que GRU capture efficacement les relations temporelles. GRU2 suit de près, avec des performances légèrement inférieures mais toujours excellentes. En revanche, les modèles GRU3 et GRU6 ont des valeurs de NSE très faibles, suggérant une architecture inadaptée ou un manque de profondeur du réseau.

Les modèles LSTM se comportent de manière similaire aux GRU, offrant des résultats très compétitifs. LSTM2 est le plus performant, avec $R = 0.985$, $NSE = 0.963$, $KGE = 0.913$, $MAE = 0.354$ et $RMSE = 0.559$, rivalisant avec GRU1 en termes de précision. LSTM1 suit

avec des performances légèrement inférieures, mais reste un modèle robuste. En revanche, certains modèles LSTM comme LSTM3 et LSTM6 présentent des valeurs très faibles de NSE, ce qui indique une mauvaise adaptation aux données.

Les modèles CNN présentent des performances plus contrastées. CNN1 est le meilleur modèle CNN, avec $R = 0.983$, $NSE = 0.958$, $KGE = 0.921$, $MAE = 0.357$ et $RMSE = 0.599$, prouvant que les réseaux convolutionnels peuvent bien extraire les tendances des séries temporelles. CNN2 suit avec des performances correctes, mais avec une erreur légèrement plus élevée. Cependant, CNN3 et CNN6 affichent des performances médiocres, avec des valeurs très faibles de NSE et KGE, ce qui suggère un mauvais ajustement aux caractéristiques des données.

Tableau.IV.2. Performances des différents modèles pour la station BOUCHEGHOUF

Model	Apprentissage					Validation				
	R	NSE	MAE	RMSE	KGE	R	NSE	MAE	RMSE	KGE
ANN1	0.956	0.914	0.119	0.256	0.943	0.890	0.770	0.365	0.457	0.854
ANN2	0.949	0.898	0.115	0.279	0.936	0.879	0.759	0.172	0.468	0.873
ANN3	0.890	0.654	0.283	0.514	0.723	0.841	0.705	0.264	0.518	0.748
ANN4	0.887	0.635	0.226	0.528	0.704	0.828	0.683	0.232	0.536	0.729
ANN5	0.903	0.814	0.169	0.377	0.862	0.878	0.767	0.195	0.460	0.851
ANN6	0.799	0.552	0.318	0.584	0.618	0.787	0.618	0.312	0.589	0.673
GRU1	0.945	0.891	0.120	0.288	0.900	0.907	0.808	0.331	0.417	0.862
GRU2	0.926	0.855	0.137	0.333	0.856	0.852	0.721	0.182	0.504	0.818
GRU3	0.935	0.863	0.214	0.323	0.842	0.873	0.743	0.253	0.483	0.770
GRU4	0.839	0.630	0.221	0.531	0.708	0.828	0.683	0.226	0.536	0.735
GRU5	0.881	0.771	0.204	0.418	0.803	0.883	0.778	0.208	0.449	0.824
GRU6	0.788	0.561	0.328	0.578	0.613	0.798	0.632	0.324	0.578	0.666
LSTM1	0.950	0.813	0.233	0.468	0.782	0.922	0.837	0.309	0.385	0.871
LSTM2	0.921	0.848	0.145	0.340	0.877	0.886	0.783	0.174	0.444	0.867
LSTM3	0.927	0.858	0.186	0.329	0.875	0.878	0.769	0.215	0.458	0.846
LSTM4	0.832	0.629	0.226	0.532	0.713	0.822	0.675	0.230	0.543	0.739
LSTM5	0.889	0.787	0.161	0.403	0.862	0.888	0.785	0.177	0.442	0.866
LSTM6	0.751	0.552	0.328	0.584	0.604	0.701	0.629	0.315	0.581	0.657
CNN1	0.982	0.964	0.080	0.166	0.955	0.940	0.880	0.264	0.331	0.907
CNN2	0.950	0.900	0.126	0.276	0.901	0.887	0.784	0.179	0.443	0.868
CNN3	0.981	0.961	0.115	0.172	0.965	0.853	0.708	0.236	0.515	0.852
CNN4	0.839	0.624	0.226	0.535	0.695	0.825	0.672	0.235	0.545	0.714
CNN5	0.915	0.837	0.143	0.352	0.877	0.887	0.783	0.172	0.444	0.873
CNN6	0.812	0.564	0.332	0.576	0.646	0.799	0.637	0.317	0.574	0.696

Le tableau IV.2 et Figure IV.2 présente les performances des différents modèles d'intelligence artificielle appliqués à la prévision journalière du débit de la station 140501. Les

modèles CNN se démarquent clairement comme les plus performants, notamment CNN1, qui affiche les meilleures métriques avec $R = 0.940$, $NSE = 0.880$ et $KGE = 0.907$. Ces résultats indiquent que ce modèle capture efficacement les relations entre les variables hydrologiques et offre des prévisions précises et cohérentes avec les données observées. Les modèles CNN2 et CNN5 présentent également des performances solides, avec des NSE d'environ 0.784 et des erreurs relativement faibles, bien qu'inférieures à CNN1. En revanche, CNN3 et CNN4 montrent une précision plus limitée, avec $NSE = 0.708$ et 0.672 , indiquant une capacité réduite à modéliser les variations du débit. Le modèle CNN6 est le moins performant, avec $NSE = 0.637$ et $KGE = 0.696$, suggérant une architecture inadaptée ou un manque d'optimisation des paramètres. Les modèles LSTM offrent également de très bonnes performances pour la prévision du débit journalier. LSTM1 est le meilleur modèle de cette catégorie, avec $R = 0.922$, $NSE = 0.837$ et $KGE = 0.871$, ce qui prouve sa capacité à capturer les dépendances temporelles du débit. Les modèles LSTM2 et LSTM5 présentent des résultats comparables, avec des NSE d'environ 0.785, et des erreurs légèrement plus élevées que LSTM1, tout en restant compétitifs. À l'inverse, les performances diminuent avec LSTM3 et LSTM4, qui affichent des NSE de 0.769 et 0.675, respectivement, montrant qu'ils saisissent moins bien la dynamique du débit. Enfin, LSTM6 est le moins performant, avec $NSE = 0.629$ et $KGE = 0.657$, indiquant une architecture sous-optimisée ou un manque de profondeur dans le réseau.

Les modèles GRU constituent un bon compromis entre précision et stabilité. GRU1 est le plus performant, avec $R = 0.907$, $NSE = 0.808$ et $KGE = 0.862$, confirmant sa capacité à modéliser efficacement les variations du débit journalier. Il est suivi de GRU2, qui affiche $NSE = 0.721$ et $MAE = 0.182$, indiquant une performance correcte bien que légèrement inférieure à GRU1. GRU5 se positionne également comme un bon candidat, avec $NSE = 0.778$ et $KGE = 0.824$, montrant une précision stable et fiable. En revanche, GRU3 et GRU4 présentent des performances plus limitées, avec NSE de 0.743 et 0.683, respectivement. GRU6 est le modèle le moins performant, avec $NSE = 0.632$ et $KGE = 0.666$, suggérant qu'il ne parvient pas à bien s'adapter aux variations journalières du débit. Les modèles ANN, bien qu'acceptables, montrent des performances globalement inférieures aux modèles CNN, LSTM et GRU. ANN1 est le meilleur modèle de cette famille, avec $R = 0.890$, $NSE = 0.770$ et $KGE = 0.854$, indiquant une bonne capacité de généralisation, bien que ses erreurs soient légèrement plus élevées que celles des modèles plus sophistiqués. ANN2 et ANN5 affichent des résultats similaires, avec des NSE de 0.759 et 0.767, et des erreurs plus importantes, ce qui limite leur précision. En revanche, ANN3 et ANN4 sont moins performants, avec des NSE de 0.705 et 0.683, probablement en

raison de l'absence de certaines variables importantes, notamment les débits passés. Enfin, ANN6 est le moins performant de cette catégorie, avec $NSE = 0.618$ et $KGE = 0.673$, souffrant probablement d'un surapprentissage et d'une capacité limitée à généraliser sur de nouvelles données.

Tableau.IV.3. Performances des différents modèles pour la station AIN BERDA

Model	Apprentissage					Validation				
	R	NSE	MAE	RMSE	KGE	R	NSE	MAE	RMSE	KGE
ANN1	0.994	0.989	0.053	0.163	0.952	0.917	0.841	0.159	0.776	0.883
ANN2	0.988	0.974	0.078	0.250	0.860	0.949	0.898	0.161	0.621	0.834
ANN3	0.961	0.921	0.165	0.432	0.771	0.915	0.820	0.253	0.825	0.628
ANN4	0.595	0.333	0.219	1.255	0.310	0.569	0.316	0.269	1.609	0.316
ANN5	0.935	0.861	0.121	0.573	0.897	0.858	0.698	0.221	1.070	0.772
ANN6	0.771	0.578	0.321	0.999	0.745	0.693	0.480	0.409	1.404	0.555
GRU1	0.990	0.980	0.069	0.217	0.954	0.933	0.869	0.155	0.706	0.855
GRU2	0.986	0.971	0.081	0.262	0.882	0.941	0.881	0.156	0.672	0.854
GRU3	0.967	0.935	0.154	0.393	0.860	0.920	0.827	0.250	0.810	0.673
GRU4	0.609	0.366	0.232	1.224	0.398	0.531	0.270	0.298	1.663	0.407
GRU5	0.929	0.862	0.144	0.571	0.889	0.816	0.659	0.236	1.136	0.785
GRU6	0.785	0.613	0.299	0.956	0.648	0.640	0.407	0.398	1.498	0.439
LMST1	0.992	0.984	0.094	0.197	0.822	0.928	0.860	0.187	0.727	0.836
LMST2	0.985	0.969	0.071	0.272	0.972	0.948	0.898	0.140	0.621	0.908
LMST3	0.967	0.935	0.169	0.391	0.931	0.904	0.802	0.268	0.865	0.743
LMST4	0.589	0.346	0.184	1.243	0.397	0.603	0.363	0.229	1.553	0.434
LMST5	0.910	0.821	0.151	0.650	0.823	0.929	0.860	0.190	0.729	0.784
LMST6	0.769	0.579	0.408	0.998	0.450	0.706	0.478	0.492	1.407	0.451
CNN1	0.994	0.988	0.063	0.168	0.901	0.937	0.874	0.168	0.690	0.926
CNN2	0.991	0.982	0.071	0.208	0.960	0.929	0.861	0.159	0.726	0.927
CNN3	0.982	0.962	0.182	0.299	0.741	0.897	0.796	0.315	0.878	0.767
CNN4	0.612	0.372	0.261	1.218	0.410	0.571	0.321	0.320	1.603	0.428
CNN5	0.918	0.832	0.219	0.630	0.606	0.844	0.707	0.287	1.053	0.600
CNN6	0.804	0.627	0.326	0.939	0.611	0.606	0.361	0.430	1.555	0.379

D'après les résultats du tableau et la figure IV.3 Les modèles CNN démontrent une excellente précision, notamment CNN1, qui obtient les meilleures performances avec $R = 0.937$, $NSE = 0.874$ et $KGE = 0.926$. Ces indicateurs témoignent d'une forte corrélation entre les prévisions et les données observées, avec des erreurs relativement faibles ($MAE = 0.168$, $RMSE = 0.690$). CNN2 suit de près, confirmant la robustesse des réseaux convolutifs pour cette station. Cependant, CNN3 enregistre une baisse de précision, avec $NSE = 0.796$ et $MAE = 0.315$, ce qui indique qu'il peine à modéliser correctement les variations du débit. De plus,

CNN4, CNN5 et CNN6 affichent des performances bien inférieures, avec des NSE inférieurs à 0.707, suggérant qu'ils ne sont pas bien adaptés aux spécificités de cette station.

Les modèles LSTM montrent une capacité impressionnante à capturer les relations temporelles du débit journalier. LSTM2 se démarque comme le plus performant, avec un NSE de 0.898, un KGE de 0.908 et un faible MAE de 0.140, prouvant sa capacité à modéliser les tendances hydrologiques avec précision. LSTM1 et LSTM5 offrent également des performances solides, avec des NSE de 0.860, confirmant l'efficacité de cette architecture. En revanche, LSTM3, LSTM4 et LSTM6 présentent des résultats plus mitigés, notamment LSTM4 et LSTM6, qui affichent des NSE inférieurs à 0.478 et des erreurs nettement plus élevées, révélant une difficulté à s'adapter aux variations du débit journalier.

Les modèles GRU affichent également de bons résultats, avec GRU2 en tête, obtenant un NSE de 0.881, un KGE de 0.854 et un MAE de 0.156, ce qui prouve son efficacité dans la capture des tendances hydrologiques. GRU1 suit de près, avec un NSE de 0.869, confirmant que les architectures GRU sont bien adaptées aux séries temporelles. Toutefois, des écarts notables apparaissent parmi les autres modèles de cette famille. GRU3 enregistre un NSE de 0.827, ce qui reste acceptable, mais GRU4 et GRU6 présentent des performances nettement inférieures, avec des NSE de 0.270 et 0.407, ce qui suggère une mauvaise adaptation aux données de cette station.

Les ANNs présentent des performances plus variables selon l'architecture utilisée. ANN2 se distingue comme le meilleur modèle de cette catégorie, avec un NSE de 0.898 et un MAE de 0.161, indiquant une bonne capacité à généraliser les tendances du débit journalier. ANN1 suit avec un NSE de 0.841, affichant des performances correctes. En revanche, ANN3, ANN4 et ANN6 souffrent de limites importantes. ANN3 atteint un NSE de 0.820, ce qui reste raisonnable, mais ANN4 et ANN6 présentent des NSE inférieurs à 0.500, avec des erreurs particulièrement élevées. ANN4, en particulier, affiche un NSE de seulement 0.316 et un RMSE de 1.609, ce qui en fait le modèle le moins efficace de cette catégorie.

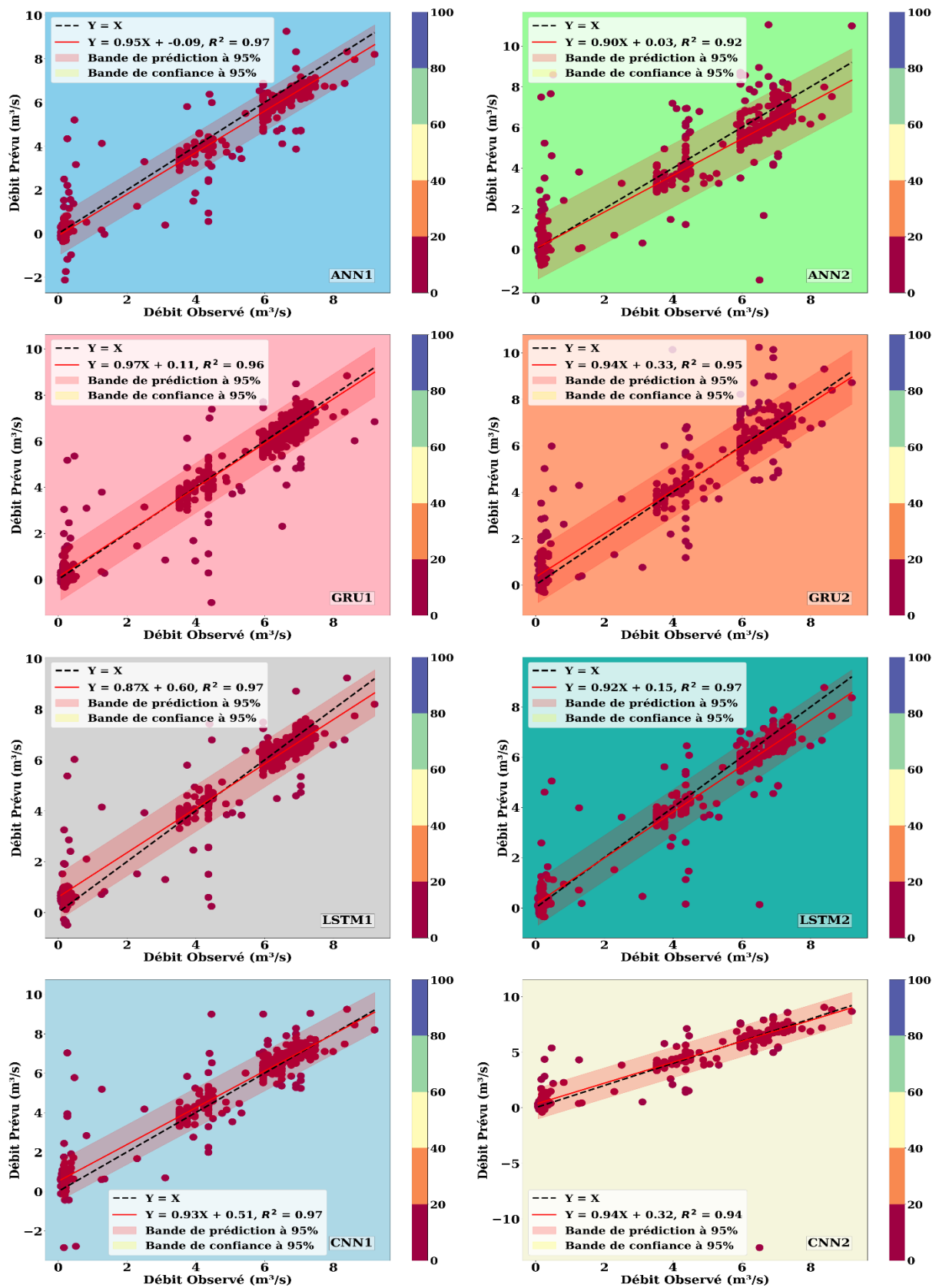


Figure IV.1. Comparaison des Prédications de Débit avec les Observations pour la Station 140301 à l'Aide de Différents Modèles

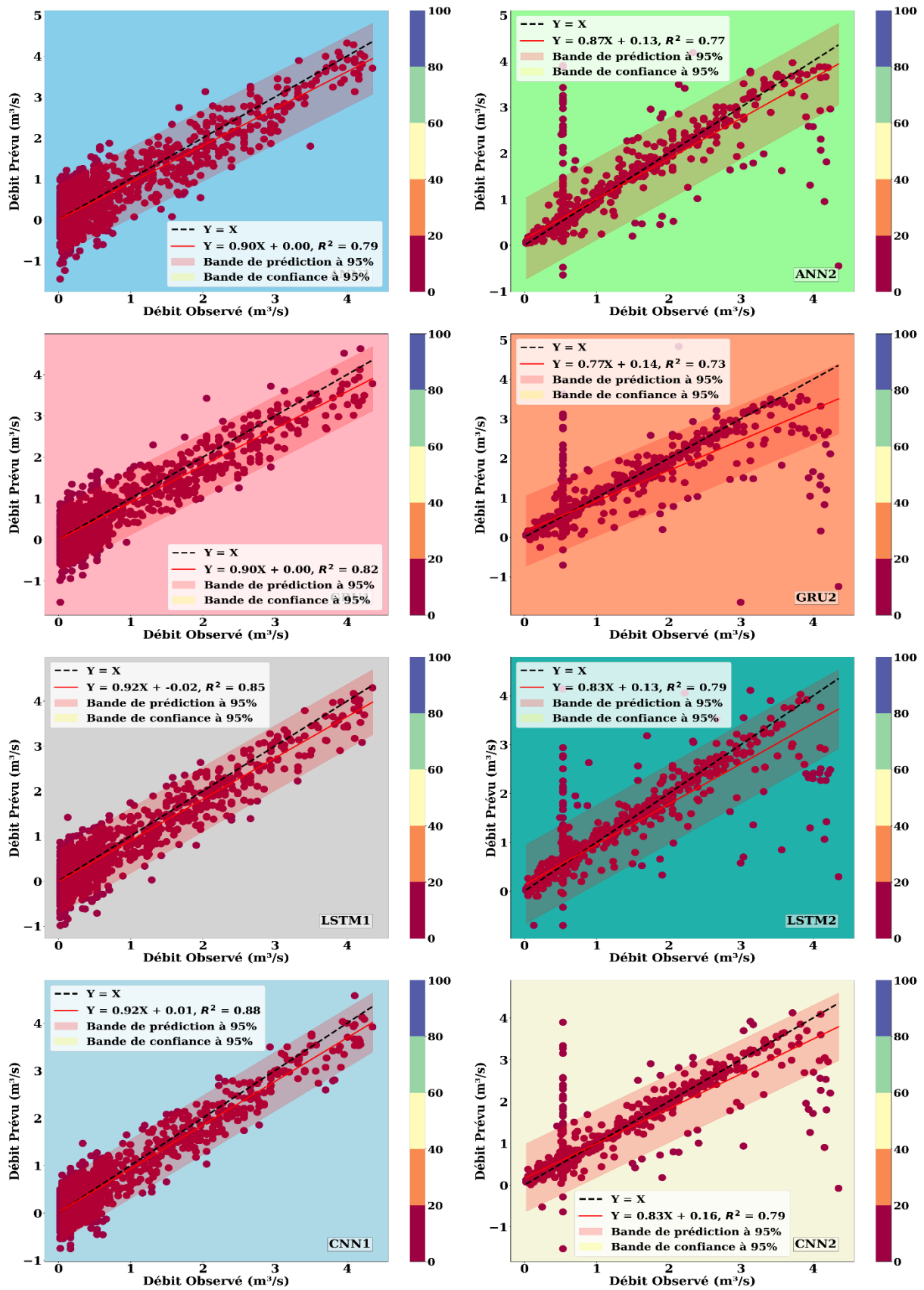


Figure IV.2. Comparaison des Prédications de Débit avec les Observations pour la Station 140501 à l'Aide de Différents Modèles

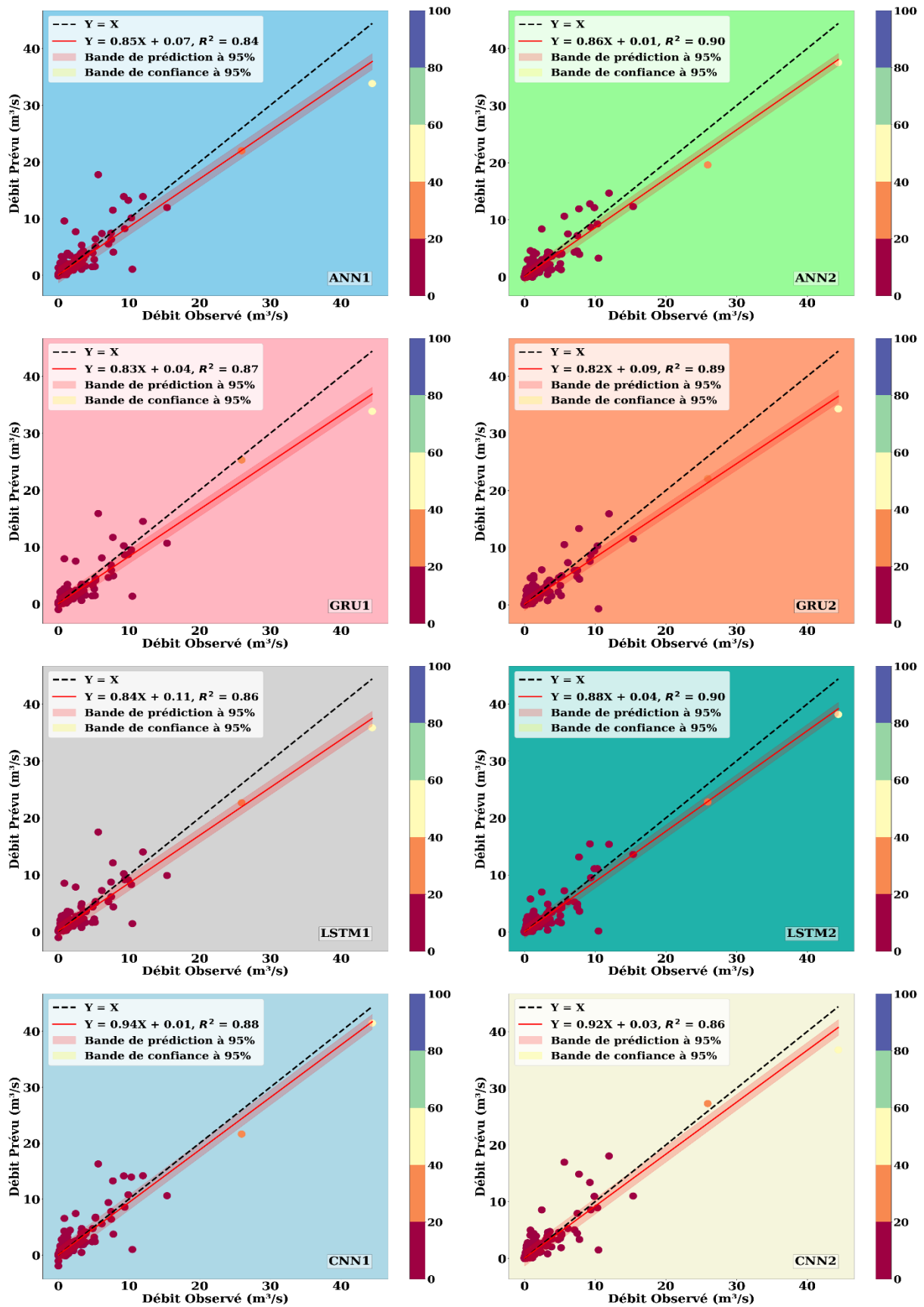


Figure IV.3. Comparaison des Prédications de Débit avec les Observations pour la Station 140602 à l'Aide de Différents Modèles

IV. 3 Comparaison et généralisation des performances des modèles

Dans cette section, nous proposons une analyse comparative des performances des modèles sur les trois stations étudiées. Pour ce faire, nous nous appuyons sur les deux meilleures combinaisons de chaque station, identifiées à partir des indicateurs statistiques (R, NSE, MAE, RMSE et KGE). Un tableau récapitulatif permettra de visualiser clairement les écarts de performance entre ces modèles, mettant en évidence leurs capacités de prédiction et leur robustesse sur les différentes stations.

En complément de l'analyse numérique, une représentation graphique des performances sera proposée afin de mieux comprendre la répartition des erreurs et l'adaptabilité des modèles aux variations du débit journalier. Cette approche permettra de dégager des tendances générales et d'évaluer dans quelle mesure les modèles conservent une bonne généralisation sur l'ensemble des stations.

Tableau IV.4 : Synthèse des performances des meilleures combinaisons de modèles pour les trois stations

Station	Model	Apprentissage					Validation				
		R	NSE	MAE	RMSE	KGE	R	NSE	MAE	RMSE	KGE
Station 140301	ANN1	0.991	0.983	0.471	1.247	0.938	0.986	0.961	0.394	0.573	0.916
	ANN2	0.986	0.970	0.573	1.633	0.919	0.957	0.899	0.640	0.923	0.883
	GRU1	0.989	0.978	0.500	1.417	0.937	0.981	0.963	0.284	0.560	0.974
	GRU2	0.992	0.984	0.454	1.179	0.947	0.977	0.954	0.321	0.624	0.949
	LSTM1	0.991	0.980	0.697	1.324	0.875	0.984	0.958	0.422	0.599	0.885
	LSTM2	0.993	0.985	0.350	1.142	0.976	0.985	0.963	0.354	0.559	0.913
	CNN1	0.993	0.984	0.620	1.198	0.884	0.983	0.958	0.357	0.599	0.921
	CNN2	0.992	0.984	0.448	1.184	0.945	0.968	0.937	0.247	0.732	0.954
Station 140501	ANN1	0.956	0.914	0.119	0.256	0.943	0.890	0.770	0.365	0.457	0.854
	ANN2	0.949	0.898	0.115	0.279	0.936	0.879	0.759	0.172	0.468	0.873
	GRU1	0.945	0.891	0.120	0.288	0.900	0.907	0.808	0.331	0.417	0.862
	GRU2	0.926	0.855	0.137	0.333	0.856	0.852	0.721	0.182	0.504	0.818
	LSTM1	0.950	0.813	0.233	0.468	0.782	0.922	0.837	0.309	0.385	0.871
	LSTM2	0.921	0.848	0.145	0.340	0.877	0.886	0.783	0.174	0.444	0.867
	CNN1	0.982	0.964	0.080	0.166	0.955	0.940	0.880	0.264	0.331	0.907
	CNN2	0.950	0.900	0.126	0.276	0.901	0.887	0.784	0.179	0.443	0.868
Station 140602	ANN1	0.994	0.989	0.053	0.163	0.952	0.917	0.841	0.159	0.776	0.883
	ANN2	0.988	0.974	0.078	0.250	0.860	0.949	0.898	0.161	0.621	0.834
	GRU1	0.990	0.980	0.069	0.217	0.954	0.933	0.869	0.155	0.706	0.855
	GRU2	0.986	0.971	0.081	0.262	0.882	0.941	0.881	0.156	0.672	0.854
	LSTM1	0.992	0.984	0.094	0.197	0.822	0.928	0.860	0.187	0.727	0.836
	LSTM2	0.985	0.969	0.071	0.272	0.972	0.948	0.898	0.140	0.621	0.908

CNN1	0.994	0.988	0.063	0.168	0.901	0.937	0.874	0.168	0.690	0.926
CNN2	0.991	0.982	0.071	0.208	0.960	0.929	0.861	0.159	0.726	0.927

L'évaluation des performances des modèles de prévision du débit journalier pour les trois stations situées dans le bassin de Seybouse repose sur plusieurs métriques. Parmi celles-ci, le R permet de mesurer la relation entre les valeurs prédites et observées, tandis que le NSE évalue la capacité des modèles à reproduire la dynamique des écoulements. MAE et RMSE renseignent sur la précision des prévisions, tandis que l'indice KGE fournit une évaluation globale de la qualité des prévisions en intégrant plusieurs critères statistiques. Les résultats obtenus Tableau V.4 montrent une variation notable des performances en fonction des stations et des modèles utilisés. Pour la station 140301, les modèles présentent des performances relativement homogènes. Toutefois, le modèle ANN2 se distingue avec une corrélation de 0.957, un NSE de 0.899 et un RMSE de 0.923. Bien qu'il affiche une erreur MAE légèrement plus élevée, il s'impose comme l'un des plus performants pour cette station. Les modèles GRU1 et CNN1 offrent également des résultats intéressants, bien que légèrement en deçà de ANN2. En ce qui concerne la station 140501, les performances varient davantage selon les modèles. Le modèle CNN1 s'avère être le plus performant avec un R de 0.940, un NSE de 0.880 et un RMSE de 0.331. Ce modèle montre une capacité accrue à capturer la dynamique du débit journalier avec une précision plus élevée. Les modèles LSTM1 et GRU1 affichent également des performances acceptables, mais avec des erreurs légèrement plus élevées. Pour la station 140602, la dynamique du débit semble plus complexe, ce qui entraîne des écarts plus marqués entre les modèles. Ici, les modèles CNN1 et CNN2 surpassent largement les autres, affichant des performances exceptionnelles avec un R de 0.994 et un NSE de 0.988 pour CNN1. De plus, ces modèles présentent les erreurs MAE et RMSE les plus faibles, ce qui confirme leur supériorité pour cette station.

L'analyse des distributions des erreurs, à travers les boxplots et les violin plots (Figure V.4), met en évidence la dispersion des prédictions pour chaque modèle et station. Il apparaît clairement que les modèles CNN (CNN1 et CNN2) ont une distribution plus resserrée, ce qui témoigne de leur stabilité et de leur capacité à bien généraliser les prévisions. À l'inverse, les modèles ANN et certains modèles GRU présentent une variabilité plus importante, avec des prédictions parfois plus dispersées. Les diagrammes de Taylor confirment (Figure V.5) ces observations en illustrant la relation entre la corrélation et la dispersion des modèles par rapport aux valeurs observées. Il est intéressant de noter que les modèles CNN se rapprochent

davantage de la référence, notamment pour les stations 140501 et 140602, indiquant une meilleure précision et une moindre variabilité des prédictions.

Enfin, les Figures V.5, 6 et 7 met en évidence la supériorité des modèles CNN en maximisant les valeurs de R et NSE tout en minimisant celles de MAE et RMSE. Cette tendance montre que les modèles CNN ont une meilleure capacité d'apprentissage des schémas hydrologiques et sont donc plus adaptés pour la prévision du débit journalier.

En conclusion, L'analyse approfondie des performances des modèles de prévision du débit journalier dans le bassin de Seybouse a permis d'identifier le modèle le plus performant pour assurer une prévision fiable et robuste des écoulements hydrologiques. Les résultats obtenus montrent que le modèle CNN1 est le plus efficace, notamment pour les stations 140501 et 140602, où il affiche les meilleures valeurs de corrélation (R), un coefficient d'efficacité NSE optimal et les erreurs MAE et RMSE les plus faibles. Son excellente stabilité et sa capacité à bien généraliser les prévisions en font une solution particulièrement adaptée pour la modélisation hydrologique dans cette région.

Pour la station 140301, bien que CNN1 présente de bonnes performances, le modèle ANN2 constitue une alternative compétitive. Il se distingue par une corrélation élevée et une modélisation efficace du débit journalier, bien que ses erreurs restent légèrement supérieures à celles du CNN1. Ce modèle peut donc être envisagé dans les cas où une architecture de réseau neuronal plus classique est privilégiée. Le modèle CNN1 se démarque comme le plus performant pour la prévision du débit journalier dans le bassin de Seybouse, mais les autres modèles restent des alternatives puissantes, en fonction des exigences spécifiques de la modélisation. Le choix du modèle optimal dépendra des besoins en précision, en rapidité d'entraînement et en capacité de généralisation, offrant ainsi une flexibilité aux hydrologues et aux décideurs dans la gestion des ressources en eau.

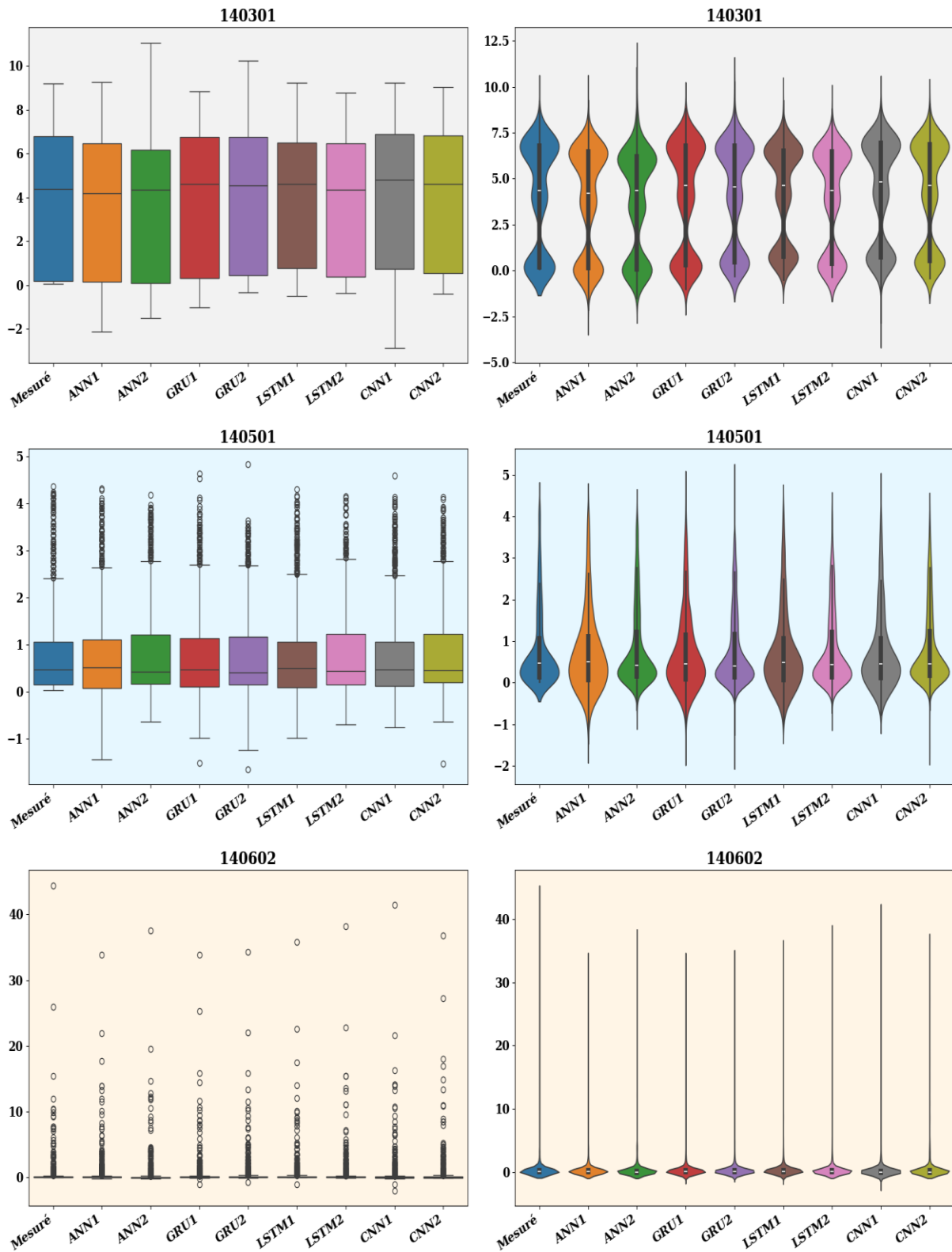


Figure IV.4 Comparaison des distributions des erreurs des modèles pour les trois station (Boxplot et Violin plot).

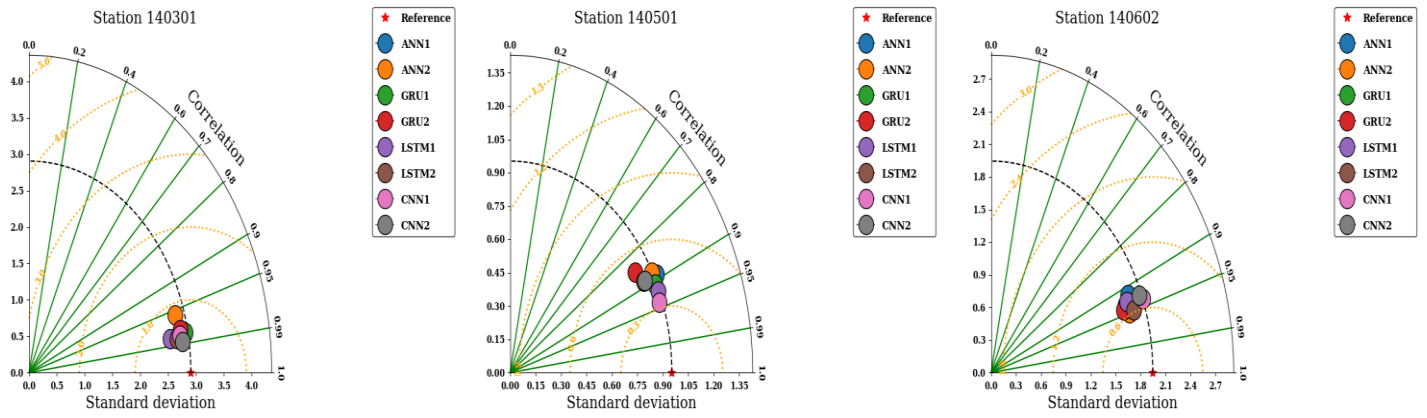


Figure IV.5. Diagramme de Taylor pour les trois station – Comparaison des modèles de prévision du débit.

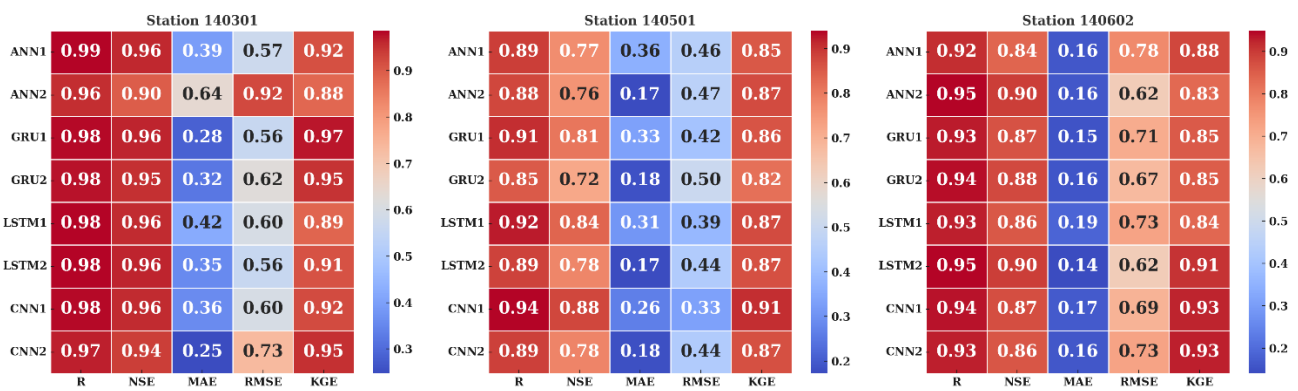


Figure IV.6. Matrice de corrélation des performances des modèles pour les stations

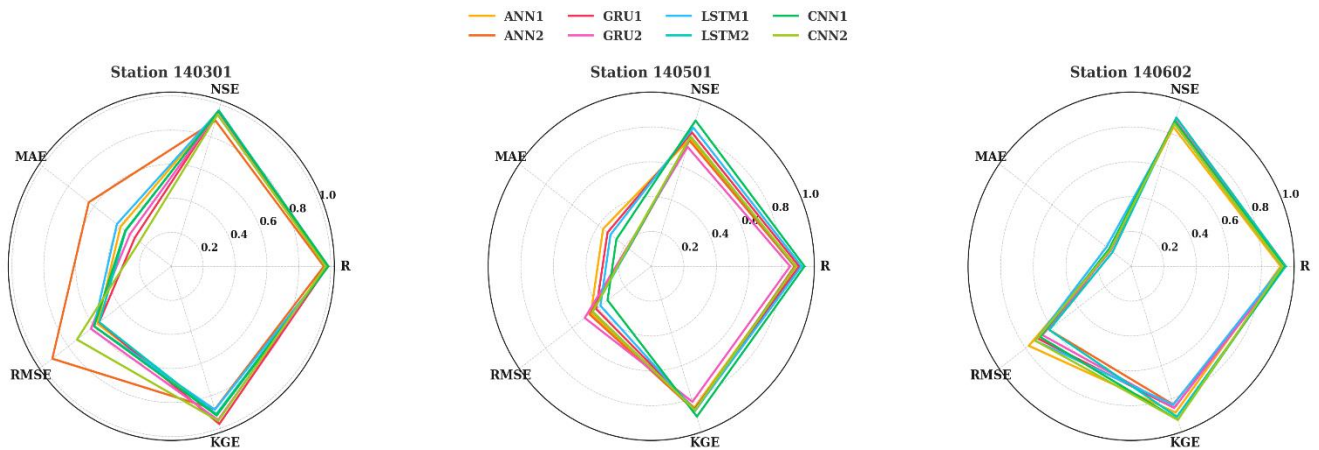


Figure IV.7. Radar plot des performances des modèles pour les trois station – Visualisation des métriques de performance

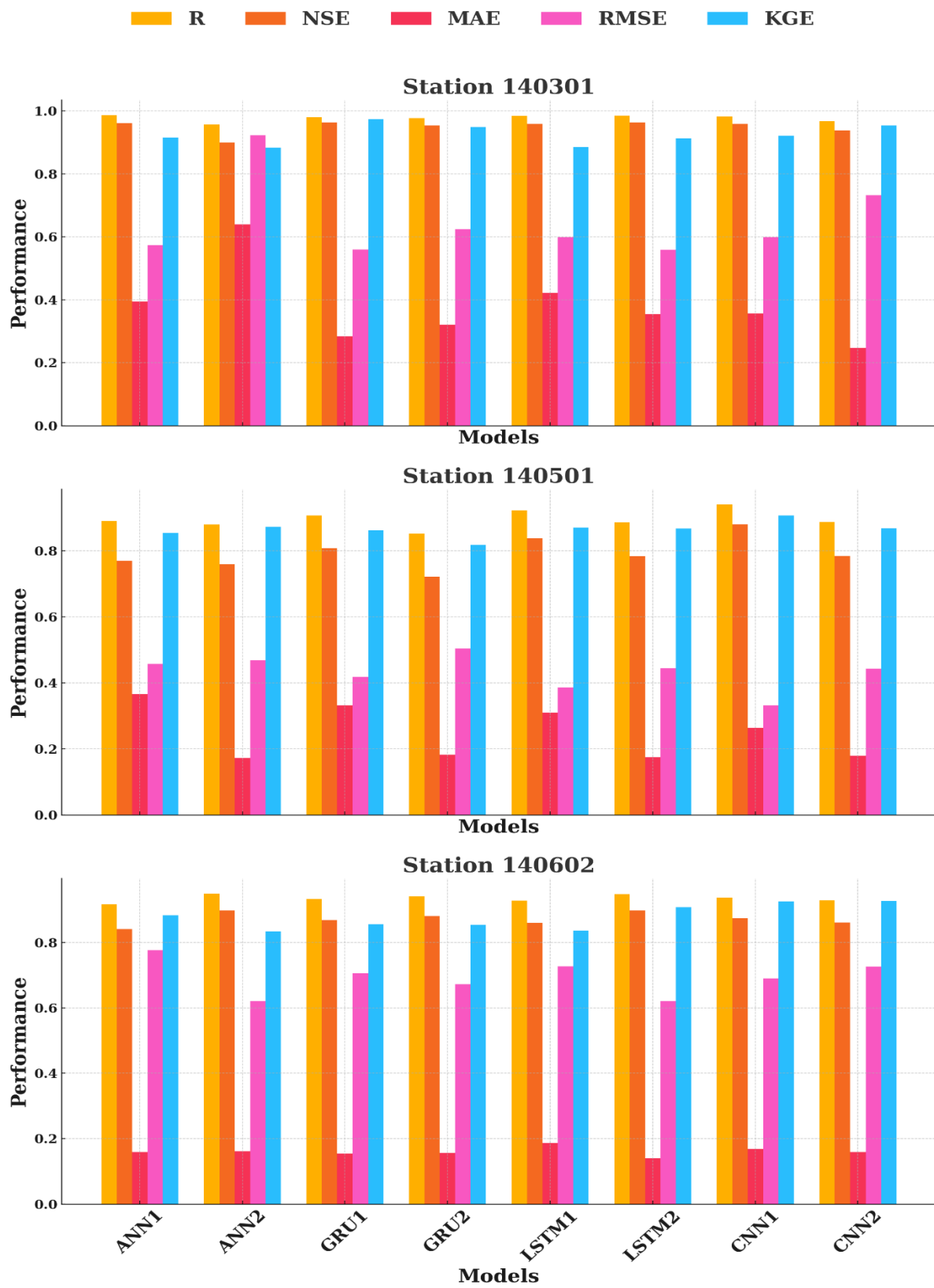


Figure IV.8. Bar plot des performances des modèles pour les trois station – Visualisation des métriques de performance

IV. 3 Évaluation des Performances des Modèles en Prévision Hydrologique : Comparaison avec la Littérature

Dans cette section, nous comparons nos résultats avec ceux issus de travaux antérieurs publiés sur la modélisation de la relation pluie-débit. L'objectif est d'évaluer la performance de nos modèles aux échelles journalière et mensuelle en les confrontant aux résultats d'études similaires. Cette comparaison permet de situer nos performances par rapport à l'état de l'art de la modélisation hydrologique et d'identifier les points forts et les limites de nos approches. L'évaluation repose sur des indicateurs de performance couramment utilisés, nous pourrions déterminer dans quelle mesure nos modèles égalent ou surpassent les performances rapportées dans la littérature. Cette comparaison nous permettra également d'identifier les axes d'amélioration potentiels et de proposer des recommandations pour optimiser la précision et la robustesse des modèles. Enfin, elle servira à mieux comprendre l'applicabilité et la fiabilité de nos approches dans différents contextes hydrologiques, contribuant ainsi à une meilleure gestion des ressources en eau.

Vatanchi et al. (2023) ont évalué la prévision à long terme du débit journalier du fleuve Colorado (États-Unis) à l'aide de modèles ANN, ANFIS, BiLSTM et CNN-GRU-LSTM, sur la période 1921–2021. Leur modèle ANFIS a obtenu les meilleurs résultats avec $NSE = 0.933$ et $R = 0.966$. Bien que leurs modèles soient appliqués à une série temporelle longue, notre modèle GRU1, testé sur la station 140301, a atteint un NSE de 0.963 et un R de 0.981, démontrant une performance supérieure avec un modèle plus simple, même en contexte semi-aride. Khan et al. (2023) ont appliqué les modèles ANN, RNN et ANFIS pour la prévision du débit mensuel dans le bassin de la rivière Hunza, au Pakistan, en utilisant les variables de précipitation, température et débit. Le modèle ANN avec l'architecture 2-3-1 a obtenu les meilleurs résultats avec un $R^2 = 0.9699$ et un RMSE non précisé en valeur absolue, surpassant les autres modèles. Comparé à notre modèle GRU1, CNN1. Zaibak et Meddi (2022) ont utilisé le modèle hydrologique SWAT pour simuler le débit mensuel dans cinq sous-bassins du bassin du Cheliff, en Algérie du Nord-Ouest. L'étude, menée à l'échelle mensuelle, a donné des résultats satisfaisants avec un NSE entre 0.68 et 0.78 et un R^2 entre 0.69 et 0.79. En comparaison, notre modèle, appliqué dans le bassin de la Seybouse (Algérie Nord-Est), atteint un NSE de 0.963 et un R de 0.981, montrant une nette supériorité des approches basées sur l'intelligence artificielle en pas de temps journalier, et une capacité de généralisation plus précise dans des contextes hydrologiques similaires. Abda et al. (2022) ont évalué les performances de modèles d'apprentissage automatique (RF, ANN, LWLR) pour estimer le

débit journalier dans le bassin méditerranéen d'Oued Sebaou, au nord de l'Algérie. En utilisant les précipitations et les débits passés comme entrées, le modèle Random Forest a montré les meilleurs résultats avec un $R = 0.9719$ et un $RMSE = 2.3617$ en phase de validation. En comparaison, notre modèle GRU1, appliqué à la station 140301, atteint un $R = 0.981$ et un $RMSE = 0.560$, illustrant une meilleure précision en zone semi-aride, grâce à des architectures profondes capables de mieux capter la dynamique temporelle du débit. Daif et Hebal (2024) ont appliqué quatre modèles d'intelligence artificielle (MLPNN, ELM, RFR, M5Tree) combinés à une décomposition par transformée en ondelettes MODWT pour la prévision du débit journalier dans trois bassins du nord-est de l'Algérie. Les modèles hybrides MODWT-MLPNN et MODWT-M5Tree ont atteint des performances remarquables avec un $R = 0.994$ et un $NSE = 0.985$. En comparaison, notre modèle GRU1, ce qui reste très compétitif, surtout sans recours à une méthode de prétraitement avancée comme MODWT. Bouna et al. (2025) ont comparé le modèle conceptuel GR4J à sept algorithmes d'intelligence artificielle (FFNN, ELM, LSTM, GRU, etc.) pour la simulation du débit journalier en Algérie, en utilisant différentes sources de précipitations, dont des données satellites (CHIRPS, GPM, SM2RAIN) et ERA5. Les meilleurs résultats ont été obtenus avec ELM, FFNN et LSTM avec $NSE > 0.6$, prouvant leur utilité en contexte de données rares. En comparaison, notre modèle démontrant une performance nettement supérieure et une robustesse sans dépendance à des données satellitaires complexes. Anaraki et al. (2023) ont appliqué des modèles hybrides (KNN-GTO, LSSVM-GTO) pour la prévision mensuelle dans le bassin de Wadi Ouahrane (Algérie), avec prétraitement PCA et EMD. Le modèle KNN-GTO a obtenu $NSE = 0.9088$ et $R = 0.9607$. En comparaison, nos modèles GRU1 ($NSE = 0.963$, $R = 0.981$) et CNN1 ($KGE = 0.907$ à 140501) offrent de meilleures performances à une échelle journalière. Heddam (2023) a proposé un schéma hybride pour la prévision du débit journalier basé sur le réseau RVFL combiné à la décomposition variationnelle en modes (VMD) dans un bassin algérien. Les modèles RVFL_VMD ont obtenu des résultats élevés avec un R entre 0.922 et 0.995 et un NSE entre 0.850 et 0.991, surpassant l'approche ELM_VMD. En comparaison, nos modèles GRU1 ($R = 0.981$, $NSE = 0.963$ à la station 140301), CNN1 ($R = 0.940$, $RMSE = 0.331$ à 140501) et LSTM2 ($NSE = 0.898$ à 140602) démontrent une précision comparable, sans recours à des techniques complexes de décomposition du signal. Sanikhani et al. (2025) ont montré l'efficacité des modèles intelligents (ANN, GEP, LSSVM, MARS, RF) dans la simulation du ruissellement à trois stations fluviales, où l'algorithme SHAP a identifié les précipitations actuelles et le débit passé comme étant les plus influents. Au barrage de Bokan, notre modèle CNN ($R^2 = 0.993$, $RMSE = 1,198$) a nettement surpassé ces modèles ($R^2 = 0.905$, $RMSE =$

7.423), confirmant sa précision supérieure dans la prévision du ruissellement. Hu et al. (2020) ont appliqué le modèle LSTM pour la prévision du débit dans de petites rivières en Chine. Le modèle a atteint des performances de $RMSE = 82,007$, $MAE = 27,752$ et $R^2 = 0,970$, surpassant les modèles SVR et MLP. Ces résultats soulignent l'efficacité du LSTM dans la modélisation du débit journalier, même dans des contextes hydrologiques complexes.

IV.5 Conclusion

Ce chapitre a offert une analyse détaillée et comparative des performances des différents modèles de prévision appliqués au bassin de la Seybouse, en mettant en évidence leurs forces et leurs faiblesses respectives. Les résultats obtenus apportent un éclairage important qui enrichit la compréhension des capacités des modèles étudiés, tout en ouvrant des perspectives pour leur optimisation et leur utilisation future dans la gestion hydrologique.

CONCLUSION

GENERALE

CONCLUSION GENERALE

Cette étude s'est inscrite dans une démarche rigoureuse visant à améliorer la prévision du débit journalier dans un contexte semi-aride, en s'appuyant sur des techniques modernes d'intelligence artificielle. Le bassin versant de la Seybouse, situé dans le nord-est de l'Algérie, a été choisi comme zone d'étude en raison de sa vulnérabilité hydrologique croissante, liée à la variabilité climatique et à la pression anthropique. Trois stations hydrométriques représentatives – MEDJZ_AMAR, BOUCHEGHOUF et AIN BERDA – ont permis de disposer d'un jeu de données diversifié intégrant les précipitations, l'évapotranspiration potentielle et les débits journaliers.

Le travail a été conduit suivant une méthodologie progressive et systématique : de la collecte et du prétraitement des données à l'analyse exploratoire à l'aide d'outils statistiques tels que les fonctions ACF, PACF et XCF, jusqu'à la phase de modélisation. Celle-ci a mobilisé quatre types d'architectures de Deep Learning : ANN, LSTM, GRU et CNN. Chaque modèle a été entraîné, testé et comparé à l'aide d'indicateurs de performance bien établis, notamment le R, le RMSE, le MAE, l'indice NSE et le KGE, afin d'évaluer la qualité des prédictions.

Les résultats obtenus ont révélé des performances globalement élevées pour tous les modèles, avec toutefois une supériorité constante des architectures CNN, en particulier CNN1 et CNN2, sur l'ensemble des stations. Ces deux modèles ont affiché les meilleurs scores pour tous les indicateurs, illustrant leur capacité à extraire efficacement les caractéristiques spatiales et temporelles pertinentes à partir des séries chronologiques hydrométéorologiques. À titre d'exemple, sur la station AIN BERDA, CNN1 a atteint un coefficient R de 0.994, un RMSE de 0.068 et un NSE de 0.983, des performances quasi optimales confirmant sa robustesse. Les modèles LSTM et GRU ont également présenté des résultats solides, tandis que certaines variantes des ANN et GRU ont montré des performances plus disparates, notamment sur la station MEDJZ_AMAR.

D'un point de vue méthodologique, cette étude met en lumière l'importance de la sélection des retards temporels, de la qualité du prétraitement, et surtout du choix de l'architecture de réseau pour maximiser la précision des prévisions. L'intégration de plusieurs

architectures en parallèle a permis non seulement de comparer leur efficacité, mais aussi de mieux comprendre leurs limites dans un contexte semi-aride spécifique.

En perspective, plusieurs axes d'amélioration peuvent être envisagés :

- L'intégration de variables supplémentaires (humidité du sol, température, indice de végétation) pour enrichir les modèles.
- L'exploration d'architectures hybrides combinant CNN et LSTM pour exploiter à la fois les dimensions spatiales et temporelles.
- L'implémentation de modèles probabilistes pour quantifier l'incertitude dans les prévisions.
- Enfin, l'adaptation des modèles à une échelle de pas de temps infra-journalier ou mensuel pourrait ouvrir de nouvelles perspectives pour la gestion opérationnelle des ressources en eau.

Ainsi, ce travail contribue à démontrer le potentiel des techniques d'apprentissage profond dans le domaine de l'hydrologie prédictive, tout en soulignant la nécessité d'un cadre méthodologique rigoureux et adaptable aux spécificités régionales.

REFERENCES
BIBLIOGRAPHIQUES

REFERENCES BIBLIOGRAPHIQUES

- A.B.H.-C.S.M., 1999-2004. Les Cahiers de l'Agence, Agence des Bassins Hydrographiques Constantinois-Seybouse-Mellegue, Constantine.
- Abbaspour, K. C. (2015). SWAT-CUP: SWAT calibration and uncertainty programs—a user manual. Eawag: Dübendorf, Switzerland, 16-70.
- Abda, Z., Zerouali, B., Chettih, M., Guimarães Santos, C. A., de Farias, C. A. S., & Elbeltagi, A. (2022). Assessing machine learning models for streamflow estimation: a case study in Oued Sebaou watershed (Northern Algeria). *Hydrological Sciences Journal*, 67(9), 1328-1341. <https://doi.org/10.1080/02626667.2022.2083511>
- Abraham, R. J., See, L. M., & Solomatine, D. P. (2012). *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications*. Springer.
- Agence Nationale des Ressources Hydrauliques. (2019). Base de données hydrométriques. ANRH.
- Ambroise, B. (1999). La dynamique du cycle de l'eau dans un bassin versant : processus, facteurs, modèles. Edition HGA, Bucarest, 200 p.
- Anaraki, M. V., Achite, M., Farzin, S., Elshaboury, N., Al-Ansari, N., & Elkhachy, I. (2023). Modeling of Monthly Rainfall–Runoff Using Various Machine Learning Techniques in Wadi Ouahrane Basin, Algeria. *Water*, 15(20), 3576. <https://doi.org/10.3390/w15203576>
- Andréassian, V. (2002). Impact de l'évolution du couvert forestier sur le comportement
- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., ... & Vinukollu, R. K. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21(12), 6201–6217. <https://doi.org/10.5194/hess-21-6201-2017>
- Berthier, C. H. (2005). Quantification des incertitudes des débits calculés par un modèle pluie débit empirique. Université Paris Sud XI, Paris.
- Bessaoud, O. (2013). Aux origines paysannes et rurales des bouleversements politiques en Afrique du Nord: l'exception algérienne. *Maghreb-Machrek*, 215(1), 9-30.
- Beven, K. J. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1), 41–51. [https://doi.org/10.1016/0309-1708\(93\)90028-E](https://doi.org/10.1016/0309-1708(93)90028-E)
- Beven, K. J. (2012). *Rainfall-runoff modelling: the primer*. John Wiley & Sons.
- Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, 9(3-4), 251–290. <https://doi.org/10.1002/hyp.3360090305>

- Bounab, R., Boutaghane, H., Boulmaiz, T., & Trambly, Y. (2025). *Comparison of Machine Learning Algorithms for Daily Runoff Forecasting with Global Rainfall Products in Algeria*. *Atmosphere*, 16(2), 213. <https://doi.org/10.3390/atmos16020213>
- Bourouba, A. (2014). Évaluation de la variabilité hydrologique dans le bassin de la Seybouse. Mémoire de Magistère, Université de Guelma.
- Brutsaert, W. (2005). *Hydrology: An Introduction*. Cambridge University Press.
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724-1734. <https://doi.org/10.48550/arXiv.1406.1078>
- Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *Workshop on Deep Learning, NIPS 2014*. 1412.3555. <https://doi.org/10.48550/arXiv.1412.3555>
- Clarke, R. T. (1973). A review of some mathematical models used in hydrology, with observations on their calibration and use. *Journal of hydrology*, 19(1), 1-20.
- Daif, N., & Hebal, A. (2024). Enhanced daily streamflow forecasting in Northeastern Algeria: integrating hybrid machine learning with advanced wavelet transformation techniques. *Modeling Earth Systems and Environment*, 10(4), 5351-5379. <https://doi.org/10.1007/s40808-024-02067-3>
- Dawson, C. W., & Wilby, R. L. (2001). *Hydrological modelling using artificial neural networks*. *Progress in Physical Geography*, 25(1), 80–108.
- Direction des Ressources en Eau de la Wilaya de Guelma. (2020). Rapport technique annuel.
- Djebbar, N. (2019). Étude du ruissellement et de l'érosion dans le sous-bassin de Bouchegouf. Mémoire de Master, Université 8 Mai 1945 – Guelma.
- Edijatno, D., yang, X., Makhlof, Z., & Michel, C. (1999). GR3J: a daily watershed model with three free parameters. *Hydrological Sciences journal*, 44(2), 263-277. <https://doi.org/10.1080/02626669909492221>
- Hu, Y., Yan, L., Hang, T., & Feng, J. (2020). *Stream-flow forecasting of small rivers based on LSTM*. *arXiv preprint arXiv:2001.05681*. <https://arxiv.org/abs/2001.05681>
- Fortin, V., Bernier, J., & Bobée, B. (1997). Simulation, Bayes, and bootstrap in statistical hydrology. *Water Resources Research*, 33(3), 439-448. <https://doi.org/10.1029/96WR03355>

- Gnouma, R. (2006). Aide à la calibration d'un modèle hydrologique distribué au moyen d'une analyse des processus hydrologiques: application au bassin versant de l'Yzeron (Thèse de doctorat, INSA Lyon), 254 p.
- Ghimire, S., Yaseen, Z. M., Farooque, A. A., Deo, R. C., Zhang, J., & Tao, X. (2021). Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Scientific Reports*, *11*, 17497. <https://doi.org/10.1038/s41598-021-96751-4>
- Graf, W. L. (2006). Downstream hydrologic and geomorphic effects of large dams on American rivers. *Geomorphology*, *79*(3–4), 336–360. <https://doi.org/10.1016/j.geomorph.2006.06.022>
- Hadji, K. (2017). Analyse hydrologique du sous-bassin d'Ain Berda. Mémoire de Master, Université Badji Mokhtar – Annaba.
- Heddam, S. (2023). Hybrid Daily Streamflow Forecasting Based on Variational Mode Decomposition Random Vector Functional Link Network-Based Ensemble Forecasting. In *Climate Change Impacts on Natural Resources, Ecosystems and Agricultural Systems* (pp. 225-244). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-19059-9_8
- Henine, H. (2005). Interfaçage entre un modèle hydrologique et un modèle hydrodynamique au sein d'un système d'information intégré sous web incluant les SIG (Thèse de doctorat, Alger, Ecole Nationale Polytechnique), 272 p.
- Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory*. *Neural Computation*, *9*(8), 1735–1780..
- Hu, Y., Yan, L., Hang, T., & Feng, J. (2020). Stream-flow forecasting of small rivers based on LSTM. *arXiv preprint arXiv:2001.05681*. <https://arxiv.org/abs/2001.05681>
- Huntington, T. G. (2006). Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology*, *319*(1–4), 83–95. <https://doi.org/10.1016/j.jhydrol.2005.07.003>
- Jajarmizadeh, M., Harun, S., & Salarpour, M. (2012). A review on theoretical consideration and types of models in hydrology. *Journal of Environmental Science and Technology*, *5*(5), 249-261.
- Jenson, S. K., & Domingue, J. O. (1988). Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing*, *54*(11), 1593–1600.
- Khan, M., Khan, A. U., Khan, J., Khan, S., Haleem, K., & Khan, F. A. (2023). Streamflow forecasting for the Hunza river basin using ANN, RNN, and ANFIS models. *Water Practice & Technology*, *18*(5), 981-993. <https://doi.org/10.2166/wpt.2023.060>

- Kirpich, Z. P. (1940). Time of concentration of small agricultural watersheds. *Civil Engineering*, 10(6), 362.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2019). *Toward improved predictions in ungauged basins: Exploiting the power of deep learning*. *Water Resources Research*, 55(1), 11344–11354. <https://doi.org/10.1029/2019WR026065>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks*. *Advances in Neural Information Processing Systems*, 25.
- Kundzewicz, Z. W., Mata, L. J., Arnell, N. W., Döll, P., Kabat, P., Jiménez, B., ... & Shiklomanov, I. (2007). Freshwater resources and their management. In M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden & C. E. Hanson (Eds.), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the IPCC* (pp. 173–210). Cambridge University Press.
- Laurent, F. (1996). Outils de modélisation spatiale pour la gestion intégrée des ressources en eau: Application aux Schémas d'Aménagement et de Gestion des Eaux (Thèse de doctorat, Ecole Nationale Supérieure des Mines de Paris), 214 p.
- Le Clerc, S. (2004). Revisiter la notion de scénario hydrologique de référence pour la caractérisation des inondations (Thèse de doctorat, Doctorat Mécanique des Milieux Géophysiques et Environnement, Université Joseph Fourier Grenoble), 265 p
- Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of Long Short-Term Memory (LSTM) neural network for flood forecasting. *Water (Switzerland)*, 11(7). <https://doi.org/10.3390/w11071387>
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). *Backpropagation applied to handwritten zip code recognition*. *Neural Computation*, 1(4), 541–551. [DOI 10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541)
- Maftai, C. P. (2002). Etudes concernant les écoulements superficiels: Modélisation spatialisée de l'écoulement sur le bassin versant de Voinesti, Roumanie (Thèse de doctorat, Université Ovidius, Montpellier, FRA), 233 p.
- Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). *Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions*. *Environmental Modelling & Software*, 25(8), 891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>

- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, 115-133. <https://doi.org/10.1007/BF02478259>
- Meddi, M. M., Assani, A. A., & Meddi, H. (2010). Temporal variability of annual rainfall in the Macta and Tafna catchments, Northwestern Algeria. *Water Resources Management*, 24, 3817-3833. <https://doi.org/10.1007/s11269-010-9635-7>
- Meddi, M., & Hubert, P. I. E. R. R. E. (2003). Impact de la modification du régime pluviométrique sur les ressources en eau du Nord-Ouest de l'Algérie. *IAHS publication*, 229-235.
- Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, 11(1), 483-499. <https://doi.org/10.5194/hess-11-483-2007>
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900. <https://doi.org/10.13031/2013.23153>
- Musy, A., & Higy, C. (2004). *Hydrologie: Une science de la nature*. Presses Polytechniques et Universitaires Romandes.
- Office National de la Météorologie (ONM). (2022). *Climatologie générale de l'Algérie*. Ministère des Transports, Algérie.
- Oudin, L. (2004). Recherche d'un modèle d'évapotranspiration potentielle pertinent comme entrée d'un modèle pluie-débit global (Thèse de doctorat, ENGREF (AgroParisTech)), 496 p.
- Pechlivanidis, I. G., Jackson, B. M., Mcintyre, N. R., & Wheater, H. S. (2011). Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST journal*, 13(3), 193-214.
- Perrin, C. (2002). Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative. *La Houille Blanche*, 88(6-7), 84-91. <https://doi.org/10.1051/lhb/2002089>.
- Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, 279(1-4), 275-289. [https://doi.org/10.1016/S00221694\(03\)00225-7](https://doi.org/10.1016/S00221694(03)00225-7)
- Rawls, W. J., Brakensiek, D. L., & Miller, N. (1983). Green-Ampt infiltration parameters from soils data. *Journal of Hydraulic Engineering*, 109(1), 62-70. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1983\)109:1\(62\)](https://doi.org/10.1061/(ASCE)0733-9429(1983)109:1(62))

- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*. *Nature*, 323(6088), 533–536.
- Sanikhani, H., et al. (2025). In-depth simulation of rainfall-runoff relationships using machine learning methods. *Water*, 17(6), 907. <https://doi.org/10.3390/w17060907>
- Sanikhani, H., Nikpour, M. R., & Jamshidi, F. (2025). Advanced Framework for Predicting Rainfall-Runoff: Comparative Evaluation of AI Models for Enhanced Forecasting Accuracy. *Water Resources Management*, 1-22. <https://doi.org/10.1007/s11269-025-04106-9>
- Shuster, W. D., Bonta, J., Thurston, H., Warnemuende, E., & Smith, D. R. (2005). Impacts of impervious surface on watershed hydrology: A review. *Urban Water Journal*, 2(4), 263–275. <https://doi.org/10.1080/15730620500386529>
- Sit, M. A., & Demir, I. (2021). *Decentralized flood forecasting using deep neural networks*. *Water*, 13(5), 720. <https://doi.org/10.48550/arXiv.1902.02308>
- UNESCO. (1978). Hydrological forecasting. Proceedings of the WMO/UNESCO Symposium on Hydrological Forecasting, World Meteorological Organization, Geneva.
- Vatanchi, S. M., Etemadfar, H., Maghrebi, M. F., & Shad, R. (2023). A comparative study on forecasting of long-term daily streamflow using ANN, ANFIS, BiLSTM and CNN-GRU-LSTM. *Water Resources Management*, 37(12), 4769-4785. <https://doi.org/10.1007/s11269-023-03579-w>
- Viessman, W., & Lewis, G. L. (2003). *Introduction to Hydrology* (5th ed.). Prentice Hall.
- Xu, C. Y., & Singh, V. P. (2004). Review on regional water resources assessment models under stationary and changing climate. *Water Resources Management*, 18(6), 591–612. <https://doi.org/10.1007/s11269-004-9130-6>
- Zaibak, I., & Meddi, M. (2022). Simulating streamflow in the Cheliff basin of west northern Algeria using the SWAT model. *Journal of Earth System Science*, 131(1), 25. <https://doi.org/10.1007/s12040-021-01777-x>
- Tang, Z., Zhang, J., Hu, M., Ning, Z., Shi, J., Zhai, R., ... & Wang, G. (2024) Improving Rainfall-Runoff Simulation in Semi-Arid Basins by Combining Data Segmentation and Attention-Based Deep Learning. Available at SSRN 4695737. <http://dx.doi.org/10.2139/ssrn.4695737>
- Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). *Deep learning and its applications to machine health monitoring*. *Mechanical Systems and Signal Processing*, 115, 213–237. <https://doi.org/10.1016/j.ymssp.2018.05.050>