
n

RÉPUBLIQUE DU CAMEROUN

Paix - Travail - Patrie

UNIVERSITÉ DE YAOUNDÉ I

FACULTÉ DES SCIENCES

DÉPARTEMENT D'INFORMATIQUE



REPUBLIC OF CAMEROON

Peace - Work - Fatherland

UNIVERSITY OF YAOUNDE I

FACULTY OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

PRÉVISION DES PRÉCIPITATIONS DANS LA ZONE DE MAGA, EXTRÊME-NORD CAMEROUN, PAR UN MODÈLE HYBRIDE LSTM-GRU.

En vue de l'obtention du diplôme de Master Recherche en Informatique

Option :

Science des données

Présenté par :

PERIVOU AGOULA

Matricule :

15T2562



Sous la direction de : Pr NGONGO Séraphin Isidore, UY1

Année académique 2022 / 2023



Université de Yaoundé 1
Département d'Informatique

**Prévision des précipitations dans la zone de MAGA,
Extrême-Nord Cameroun, par un modèle hybride
LSTM-GRU.**

Présenté par :
PERIVOU AGOULA

Matricule :
15T2562

Superviseur : Pr NGONGO Séraphin Isidore, UY1

Année académique 2022 / 2023

DÉDICACE

Je dédie ce mémoire à ma grande soeur ATILWA AGOULA.

REMERCIEMENTS

Avant toute chose, je remercie notre créateur Dieu le père pour la grâce qu'il m'a accordé d'être en santé afin de pouvoir effectuer ce travail et son soutien indéfectible durant toute cette période. Je témoigne ensuite ma gratitude et ma reconnaissance à toutes ces personnes qui ont contribué à la réalisation de ce travail :

À mon directeur, Pr NGONGO SERAPHIN ISIDORE, pour m'avoir donné l'opportunité d'explorer un thème très intéressant, d'avoir toujours été disponible face aux préoccupations et difficultés rencontrées. Ce dernier m'a toujours orienté dans la bonne direction, apporté des conseils, incité à cultiver le sens du travail bien fait. Je leur adresse ma reconnaissance infinie et mes sincères remerciements pour m'avoir encadré, avoir été disponible, pour les multiples encouragements et conseils qui m'ont accompagné tout au long de ce travail.

Aux distingués membres de mon jury de soutenance, pour l'honneur qu'ils m'ont fait en acceptant d'évaluer ce travail, pour leurs précieux commentaires et suggestions qui ont permis d'améliorer la qualité de ce mémoire, et pour le temps qu'ils ont généreusement consacré à sa lecture et évaluation.

À l'Université de Yaoundé 1 et au département d'informatique en particulier, pour la qualité de la formation que j'ai reçue.

À l'équipe TOGETTECH pour l'environnement de travail confortable mis à notre disposition.

À mes très chers parents pour leur immense amour, leur patience, leur soutien indéfectible durant toute cette période, leurs encouragements et leurs conseils. À toutes mes soeurs, mes oncles et tantes pour leurs encouragements et conseils.

À monsieur MOJO Fabrice pour ses conseils et ses remarques durant lors de mes différents exposés.

Enfin, j'exprime toute ma gratitude à mes amis pour tous leurs encouragements et leur soutien, nos différents échanges. Je remercie GHISLAIN POUM et NAGUEU LIONEL qui m'ont aidé et orienté dès le début et à chaque fois que j'avais des difficultés, Abdelkarim qui m'a aidé dans les aspects mathématiques de mon travail, THSUISSEU SIPEUWOU AUDREY qui m'a toujours encouragé et soutenu pour la finalisation de ce travail, et tous ceux que je n'ai pas cité, merci beaucoup les amis.

RÉSUMÉ

La prévision des précipitations est essentielle pour la gestion agricole, la planification des ressources en eau et l'atténuation des risques liés aux catastrophes naturelles, particulièrement dans les régions arides comme Maga, dans l'Extrême-Nord du Cameroun. C'est un problème complexe en raison des variations du changement climatique dues à de nombreux facteurs. Des études antérieures ont utilisé des approches numériques de temps, des approches statistiques et des approches de machine learning pour prédire les précipitations à partir d'autres variables météorologiques. En s'appuyant sur les travaux existants, nous avons proposé dans ce mémoire d'utiliser un modèle hybride combinant les réseaux de neurones LSTM et GRU, en se basant sur un jeu de données téléchargé depuis le site de la NASA^{1 2} afin d'améliorer la précision des prévisions de précipitations dans cette région. L'approche repose sur l'exploitation des capacités des réseaux de neurones LSTM et GRU à modéliser les séries temporelles complexes et à capturer les dépendances à long terme à partir de données historiques telles que la température, l'humidité et la vitesse du vent collectées sur plusieurs années. Après plusieurs expérimentations et ajustements des hyperparamètres, la configuration optimisée prédit avec succès les quantités mensuelles de précipitations, malgré les changements climatiques. Elle affiche un MAE de 11,974, un R^2 de 0,905 et un RMSE de 15,982 sur les ensembles de données de validation et de test. Dans les études futures, l'intégration de données supplémentaires et l'exploration de nouvelles architectures de modèles pourraient encore améliorer la précision des prévisions, renforçant ainsi la résilience des communautés face aux défis climatiques croissants.

Mots-clés : Précipitation, Réseaux de neurones à long et court terme (LSTM), Unités récurrentes à portes (GRU), Prévision des précipitations.

1. NASA : National Aeronautics and Space Administration.

2. Jeu de données disponible à <https://power.larc.nasa.gov/data-access-viewer/>.

ABSTRACT

Precipitation forecasting is essential for agricultural management, water resource planning, and mitigating risks related to natural disasters, particularly in arid regions like Maga in the Extreme North of Cameroon. It is a complex issue due to climate change variations caused by numerous factors. Previous studies have employed numerical time approaches, statistical methods, and machine learning techniques to predict precipitation from other meteorological variables. Building on existing work, we proposed in this thesis to use a hybrid model combining LSTM and GRU neural networks, based on a dataset downloaded from the NASA website^{3 4} to improve the accuracy of precipitation forecasts in this region. The approach relies on leveraging the capabilities of LSTM and GRU neural networks to model complex time series and capture long-term dependencies from historical data such as temperature, humidity, and wind speed collected over several years. After several experiments and adjustments of hyperparameters, the optimized configuration successfully predicts monthly precipitation amounts, despite climate changes. It shows a MAE of 11.974, an R^2 of 0.905, and an RMSE of 15.982 on the validation and test datasets. In future studies, integrating additional data and exploring new model architectures could further enhance forecasting accuracy, thereby strengthening community resilience to growing climate challenges.

Keywords : Rainfall, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Rainfall prediction

3. NASA : National Aeronautics and Space Administration.

4. Dataset available at <https://power.larc.nasa.gov/data-access-viewer/>.

TABLE DES MATIÈRES

Introduction Générale	1
0.1 Contexte	1
0.2 Problématique	1
0.3 Méthodologie de la recherche	2
0.4 Organisation du mémoire	2
1 Généralités et revue de la littérature	3
1.1 Cadre naturel d'étude : zone de Maga	3
1.2 Notion de précipitation	4
1.2.1 Définition	4
1.2.2 Mécanismes de formation des précipitations	5
1.2.3 Types de précipitations et mesures	5
1.3 Notions des séries temporelles	5
1.4 Les approches de modélisation des séries temporelles	6
1.4.1 Les approches numériques de prévision du temps (PNT)	6
1.4.2 Les méthodes statistiques	8
1.4.3 Les réseaux de neurones	10
1.4.4 Les méthodes de deep learning	11
1.5 Tableaux récapitulatifs	12
1.6 Conclusion	13
2 DESCRIPTION DU MODÈLE LSTM-GRU POUR LA PRÉVISION DES PRÉCIPITATIONS	14
2.1 Définition du problème de prévision des précipitations	14
2.2 Réseaux LSTM (Long Short-Term Memory)	15
2.3 Réseaux GRU (Gated Recurrent Unit)	16
2.4 Architecture du modèle hybride LSTM-GRU	17
2.5 Algorithme de prévision séquentielle multi-horizon avec LSTM-GRU	20
2.5.1 Explication de l'algorithme	22
2.6 Analyse de la complexité de l'algorithme	23
2.6.1 Complexité temporelle	23
2.6.2 Complexité spatiale	24
2.7 Conclusion	24

3	RESULTATS D'EXPERIMENTATION	25
3.1	Protocole expérimental	25
3.1.1	Environnement et outils de travail	25
3.1.2	Description du jeu de données	26
3.1.3	Relation entre les variables météorologiques et les précipitations	26
3.1.4	Préparation et Prétraitement des Données	28
3.1.5	Les Métriques d'évaluation des performances	29
3.1.6	Les configurations du Modèle LSTM-GRU	30
3.2	Présentation des résultats	30
3.2.1	Résultats des métriques d'évaluation	30
3.2.2	Graphiques des fonctions de pertes et de validation	30
3.2.3	Prévisions sur les données de Test	31
3.2.4	Graphiques des données historiques et celles prédites	32
3.3	Interprétation des résultats	33
3.3.1	Tableau des métriques	33
3.3.2	Graphiques des fonctions de perte et de validation	33
3.3.3	Prévisions sur les données de test	34
3.4	Discussion	34
3.4.1	Prévision des quantités des précipitations 2024	34
3.4.2	Comparaison de nos résultats avec les résultats de l'article de base	35
3.5	Conclusion	35
	Conclusion et perspectives	35
	Bibliographie	38

ABBREVIATIONS

- **AIC** : Akaike Information Criterion
- **ANN** : Artificial Neural Network
- **AR** : AutoRegressive
- **ARIMA** : AutoRegressive Integrated Moving Average
- **ARMA** : AutoRegressive Moving Average
- **BIC** : Bayesian Information Criterion
- **BLSTM** : Bidirectional Long Short-Term Memory
- **CNN** : Convolutional Neural Network
- **DWT** : Discrete Wavelet Transform
- **GCM** : General Circulation Model
- **GRU** : Gated Recurrent Unit
- **LAM** : Limited-Area Model
- **LSTM** : Long Short-Term Memory
- **MA** : Moving Average
- **MLP** : Multi-Layer Perceptron
- **MSE** : Mean Squared Error
- **NASA** : National Aeronautics and Space Administration
- **NSE** : Nash-Sutcliffe Efficiency
- **PNT** : Prévision Numérique de Temps
- **PRECTOTCORR** : Corrected Total Precipitation
- **RH2M** : Relative Humidity at 2 Meters
- **RMSE** : Root Mean Square Error
- **RNN** : Recurrent Neural Network
- **SARIMAX** : Seasonal AutoRegressive Integrated Moving Average with eXogenous factors
- **T2M_MAX** : Maximum Temperature at 2 Meters
- **T2M_MIN** : Minimum Temperature at 2 Meters
- **WS10M_MAX** : Maximum Wind Speed at 10 Meters
- **WS10M_MIN** : Minimum Wind Speed at 10 Meters

TABLE DES FIGURES

1.1	cartographie de Maga. (Source :[6])	4
1.2	précipitation. (Source :[9])	5
1.3	Réseau de neurones. (Source : [21])	10
2.1	Architecture d’une cellule LSTM. Source : [32]	16
2.2	Structure d’un bloc GRU	17
2.3	Architecture du modèle hybride LSTM-GRU.	19
3.1	Précipitations vs Humidité relative	27
3.2	Précipitations vs Température maximale	27
3.3	Précipitations vs Température minimale	27
3.4	Précipitations vs Vitesse maximale du vent	27
3.5	Précipitations vs Vitesse minimale du vent	27
3.6	Évolution des précipitations	27
3.7	Corrélation entre les précipitations et les autres variables	28
3.8	Prétraitement des données	29
3.9	Graphique des historiques des fonctions de perte et validation	31
3.10	Résultats des prévisions sur les données Test 2023	32
3.11	Précipitations de 1982 à 2023	33
3.12	Précipitations 2024	35

LISTE DES TABLEAUX

1.1	Chronologie des principales contributions à la prévision des précipitations.	12
3.1	Informations du dataset	26
3.2	Configurations des modèles LSTM-GRU et leurs paramètres	30
3.3	Résultats des différentes configurations des modèles	31
3.4	Tableau comparatif des valeurs réelles, prédites et des écarts pour chaque mois de l'année 2023.	32

INTRODUCTION GÉNÉRALE

0.1 Contexte

Les précipitations jouent un rôle crucial dans les écosystèmes et la vie quotidienne des populations dans le monde entier et en particulier dans la zone de Maga. La prévision précise des précipitations s'avère d'une importance capitale pour la planification des activités humaines et la gestion des ressources naturelles, en particulier dans les milieux où l'agriculture dépend fortement des conditions météorologiques. L'Extrême-Nord du Cameroun, et plus particulièrement la zone de Maga, est une zone qui connaît des changements climatiques considérables ce qui est d'ailleurs à l'origine des événements météorologiques extrêmes, tels que des précipitations abondantes provoquant des inondations [1] ou des sécheresses prolongées [2]. En date du 28 Septembre 2019 [3] par exemple, les localités situés le long du fleuve Logone, à l'Est du département du Logone (Arrondissement de Zina) et au sud du Lac de Maga (Arrondissement de Kaï Kaï- Département du Mayo Danay) ont subi des inondations en raison des fortes précipitations qui ont provoquées la montée des eaux du fleuve Logone, occasionnant au passage la rupture de plusieurs digues communautaires. Par conséquent, la prévision précise des précipitations dans cette zone est très capital pour les activités agro-pastorales, la gestion des ressources en eau et la prévention des risques liés aux inondations ou à la sécheresse.

0.2 Problématique

Depuis plusieurs années, de nombreuses techniques de prévision des précipitations ont été développées, partant des modèles statistiques anciennes aux méthodes basées sur l'apprentissage automatique. Parmi ces méthodes, les réseaux de neurones récurrents (RNN) ont montré des performances assez intéressants dans la modélisation des séries temporelles, y compris la prévision des précipitations [4]. En particulier, les architectures LSTM (Long Short-Term Memory) et GRU (Gated Recurrent Unit) ont été utilisées pour capturer les dépendances temporelles complexes et améliorer les prévisions des séries temporelles [5]. Dans ce mémoire, nous nous concentrons spécifiquement sur l'application de l'architecture LSTM-GRU pour la prévision des précipitations dans la zone de Maga de l'Extrême-Nord du Cameroun. L'objectif est d'évaluer les performances de cette approche combinant LSTM et GRU dans un contexte tropical et d'analyser sa capacité à capturer les éléments de précipitations spécifiques à la zone.

Ainsi, la question principale de ce travail est de savoir si l'architecture LSTM-GRU peut donner

des prévisions précises des précipitations dans la zone de Maga, tout en prenant en compte les caractéristiques du climat de cette zone.

0.3 Méthodologie de la recherche

La méthodologie adoptée dans cette étude repose sur l'utilisation de données historiques de précipitations et de variables météorologiques connexes, notamment la température, l'humidité relative et la vitesse du vent. Ces données, issues de la base NASA POWER, couvrent la période allant de 1982 à 2023 et concernent spécifiquement la zone de Maga, dans la région de l'Extrême-Nord du Cameroun.

Dans un premier temps, un prétraitement des données est effectué, incluant la gestion des valeurs manquantes par interpolation et la normalisation des variables via la méthode Min-Max. Ensuite, l'ensemble des données est divisé en deux sous-ensembles : un ensemble d'apprentissage (de 1/01/1982 à 31/12/2022) et un ensemble de test (de 1/01/2023 à 31/12/2023).

La modélisation repose sur une architecture hybride combinant des couches LSTM (Long Short-Term Memory) et GRU (Gated Recurrent Unit), choisie pour sa capacité à capturer les dépendances temporelles complexes des séries climatiques. L'entraînement du modèle est réalisé en utilisant l'optimiseur Adam, et une recherche empirique des hyperparamètres est menée pour améliorer les performances.

La performance du modèle est évaluée à l'aide de métriques classiques : l'erreur absolue moyenne (MAE), la racine de l'erreur quadratique moyenne (RMSE) et le coefficient de détermination (R^2). Enfin, les résultats prédits sont comparés aux observations réelles afin d'évaluer la capacité du modèle à reproduire fidèlement les précipitations futures dans la zone étudiée.

0.4 Organisation du mémoire

Dans l'introduction, nous présentons le contexte général de l'étude, en précisant la problématique, les objectifs poursuivis et la méthodologie adoptée.

Le **chapitre 1** est consacré aux généralités et à la revue de littérature.

Le **chapitre 2** décrit en détail le modèle de prévision des précipitations appliqué à la région de Maga.

Le **chapitre 3** présente les résultats des expérimentations réalisées à partir des données climatiques historiques.

Enfin, dans la conclusion, nous synthétisons les principaux résultats obtenus et proposons des pistes pour des recherches futures.

CHAPITRE 1

GÉNÉRALITÉS ET REVUE DE LA LITTÉRATURE

Ce chapitre examine les éléments clés pour comprendre et modéliser les précipitations. Nous débutons par une présentation du cadre naturel d'étude, suivie d'une explication des mécanismes et types de précipitations, ainsi que des méthodes de mesure associées. Nous introduisons ensuite les concepts fondamentaux des séries temporelles et explorons diverses approches de modélisation, y compris les techniques numériques, ainsi que les méthodes statistiques. Ensuite, nous abordons les réseaux de neurones et les méthodes de deep learning, en soulignant leur rôle dans la prévision des précipitations. Enfin, nous donnons un tableau comparatif des solutions antérieures proposées pour la prévisions des séries temporelles.

1.1 Cadre naturel d'étude : zone de Maga

Maga, située dans la région de l'Extrême-Nord du Cameroun, est le cadre naturel d'étude de notre travail. La zone s'étend sur environ 2000 km², entre les latitudes 10°32'57" et 11°58'00" Nord, et les longitudes 14°34' et 15°10'13" Est. Le climat de Maga est sahélo-soudanien, caractérisé par une longue saison sèche d'octobre à avril et une courte saison des pluies de juin à septembre. Les précipitations annuelles varient entre 530 et 630 mm [1], avec des pics en août et septembre. La température moyenne est de 28°C, atteignant un maximum de 32°C en mars, avril et mai. Le relief de la région est dominé par une plaine inondable faisant partie d'un vaste bassin géomorphologique. La végétation se compose principalement d'une savane arbustive, de steppes herbacées et de zones clairsemées. Les sols de Maga comprennent des alluvions récents et anciens, des vertisols, des sols ferrugineux tropicaux, des sols hydromorphes et des sols halomorphes [1].

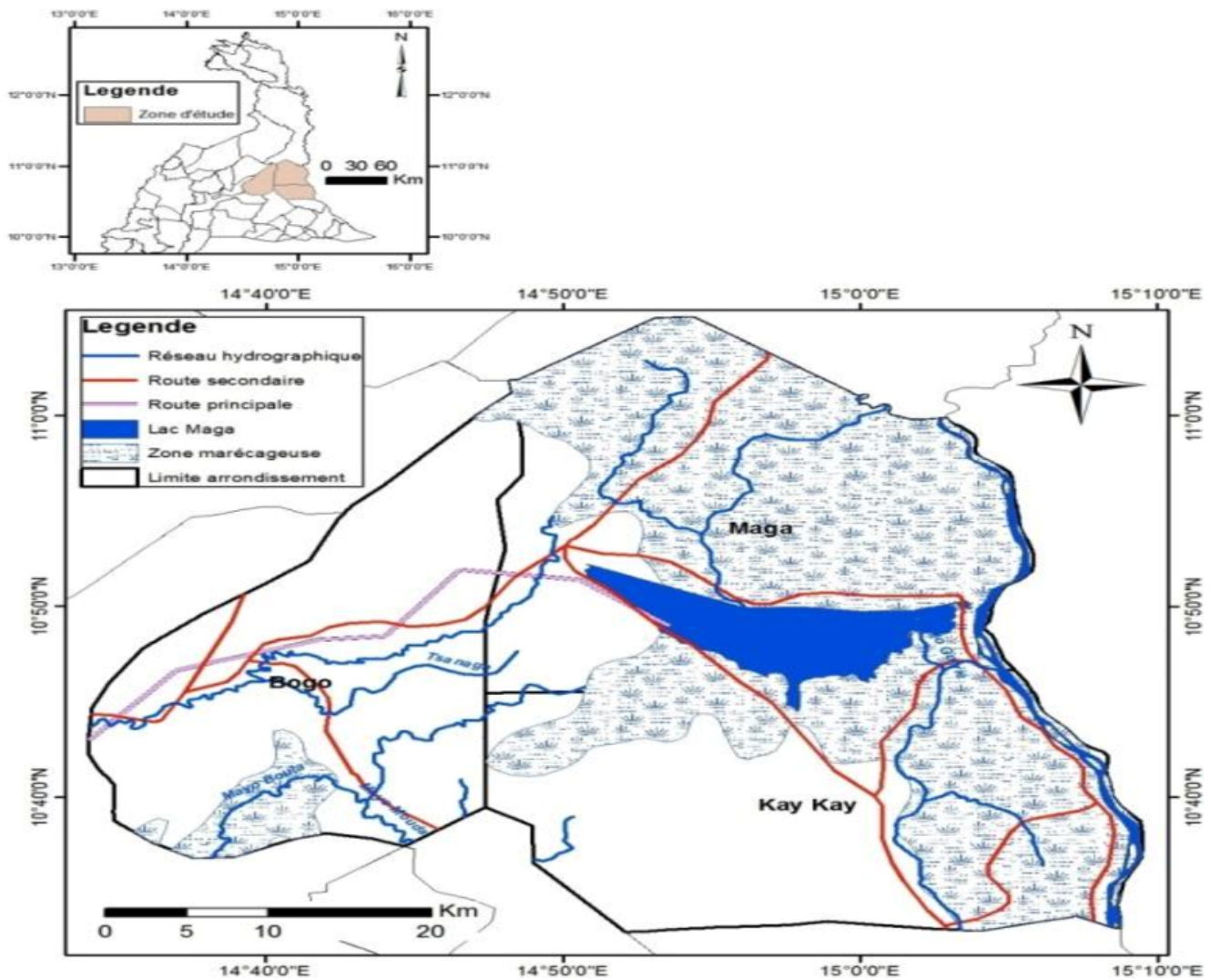


FIGURE 1.1 – cartographie de Maga. (Source :[6])

Sur le plan humain, Maga abrite une diversité de peuplements tels que les Massa, les Toupouri, les Musgum, les Mousey, les Peulhs, les Kanuri et les Kéra [1]. Les principales activités économiques de la région sont l'agriculture, l'élevage et la pêche. Les populations locales occupent notamment le lit majeur du Logone et le lac de Maga, qui est réputé pour sa richesse en poissons, en quête de zones fertiles propices à leurs activités. Cette interaction entre les ressources naturelles, le relief, le climat et les activités humaines crée un contexte spécifique dans lequel se déroule l'étude de Maga [1, 7].

1.2 Notion de précipitation

1.2.1 Définition

Les précipitations désignent les particules d'eau qui tombent de l'atmosphère vers le sol, sous forme liquide (pluie, bruine) ou solide (neige, grêle) [8]. Elles résultent d'un changement de température ou de pression et sont essentielles dans les systèmes hydrologiques continentaux comme les bassins versants. Les précipitations jouent un rôle clé dans le cycle de l'eau en apportant de l'humidité aux surfaces terrestres.

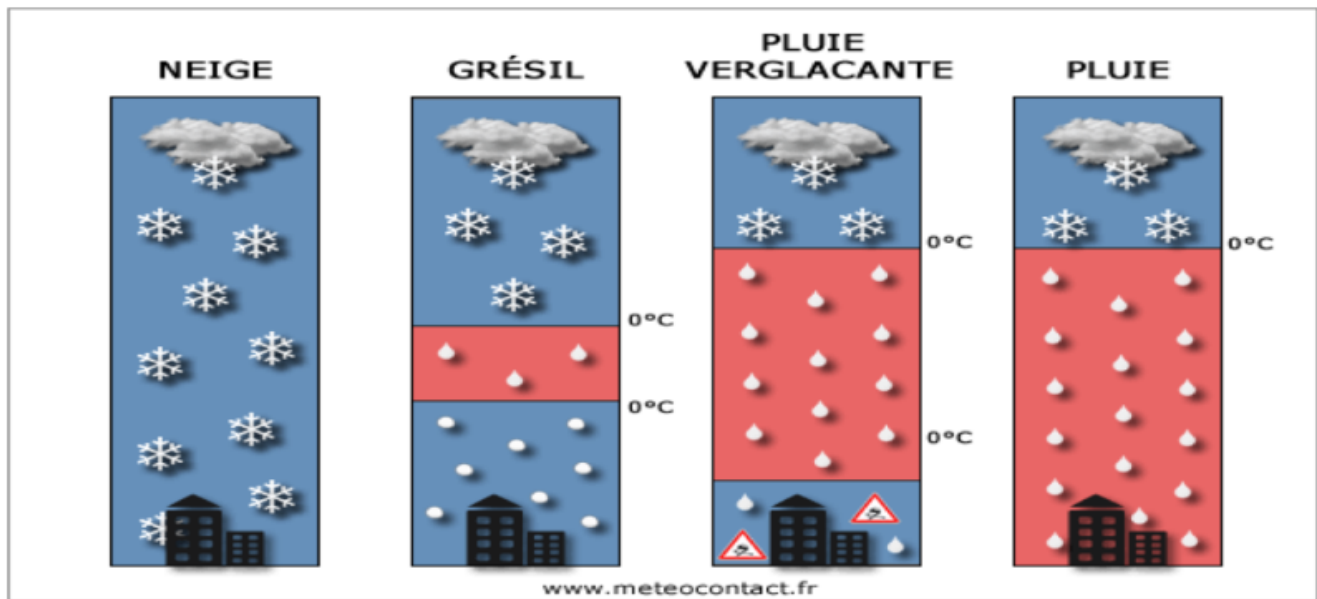


FIGURE 1.2 – précipitation. (Source :[9])

1.2.2 Mécanismes de formation des précipitations

Les précipitations se forment principalement par condensation de la vapeur d'eau atmosphérique. Ce processus peut être déclenché par plusieurs mécanismes thermodynamiques, dont le refroidissement isobare, la détente adiabatique, l'apport de vapeur d'eau, et le mélange par turbulence. La présence de vapeur d'eau, un noyau de condensation, et des turbulences sont nécessaires pour la formation des gouttelettes ou des cristaux de glace, qui, lorsqu'elles se rejoignent et deviennent suffisamment lourdes, tombent sous forme de précipitations.

1.2.3 Types de précipitations et mesures

Il existe plusieurs types de précipitations : convectives, orographiques, et frontales. Les précipitations convectives résultent d'une montée rapide d'air, les orographiques se produisent lorsque l'air rencontre une barrière topographique, et les précipitations frontales se forment aux interfaces entre masses d'air. La mesure des précipitations, qui peut être liquide ou solide, est effectuée à l'aide de pluviomètres pour la quantité totale et de pluviographes pour la répartition temporelle. Ces mesures sont exprimées en hauteur de précipitation (mm) et en intensité (mm/h).

1.3 Notions des séries temporelles

Définition 1 (Série temporelle) — *Stricto sensu, une **série temporelle** est une série d'observations numériques (mesures) indexées par le temps [10]. Ces observations seront représentées par :*

$$y_1, y_2, \dots, y_n$$

— *1, 2, ..., n représentent les marques temporelles, y_i est la valeur de la mesure réalisée au temps i .*

Les séries temporelles sont largement utilisées dans divers domaines tels que l'économie, la finance, la météorologie, et bien d'autres, pour analyser les tendances, les cycles et les modèles de comportement qui évoluent au fil du temps. Un exemple de série temporelle est celui des précipitations, qui enregistrent les variations des quantités de pluie ou de neige au fil des saisons et des années.

Une série temporelle (ou chronologique) Y_t est généralement constituée des composants suivants :

- **Tendance** T_t : elle représente l'évolution à long terme de la série à une échelle interannuelle. Par exemple, cela peut correspondre à la croissance économique ou à l'évolution climatologique à long terme. La tendance peut être modélisée de différentes façons :
 - Tendance linéaire : $T_t = a + bt$
 - Tendance quadratique : $T_t = a + bt + ct^2$
 - Tendance logarithmique : $T_t = \log(t)$
- **Saisonnalité** S_t : elle correspond à un phénomène périodique de période identifiée. Par exemple, en météorologie, cela peut représenter l'évolution de la température au fil des saisons.
- **Erreur** ε_t : c'est la partie aléatoire de la série, idéalement stationnaire.

Parfois, une autre composante, le **cycle** C_t , peut être ajoutée. Le cycle correspond à un phénomène répétitif régulier, de période inconnue ou changeante, qui est prévisible. La décomposition de la série peut être :

- additive : $Y_t = T_t + S_t + \varepsilon_t$;
- multiplicative : $Y_t = T_t \cdot S_t \cdot \varepsilon_t$;
- multiplicative et additive : $Y_t = (T_t + S_t) \cdot \varepsilon_t$ ou $Y_t = (T_t \cdot S_t) + \varepsilon_t$.

1.4 Les approches de modélisation des séries temporelles

1.4.1 Les approches numériques de prévision du temps (PNT)

Les méthodes de prévision numérique du temps (PNT) sont des techniques utilisées pour estimer et prédire les quantités et la répartition des précipitations dans une zone géographique sur une période donnée [11]. Ces méthodes s'appuient sur des modèles numériques qui simulent l'évolution de l'atmosphère en utilisant des équations mathématiques complexes.

On distingue généralement trois types de modélisation en fonction de la résolution spatiale adoptée.

Modèles de circulation générale (GCM)

Les modèles de circulation générale (GCM) constituent des outils essentiels pour simuler la dynamique atmosphérique à grande échelle et prévoir le climat sur le long terme. L'étude de Forget et al. (1999) illustre cette approche en développant un GCM spécifiquement adapté à l'atmosphère martienne, prenant en compte ses particularités telles qu'une composition majoritairement en CO, une pression atmosphérique très faible et une forte concentration en poussières [12].

Le modèle proposé par Forget et al. intègre les équations fondamentales de la dynamique atmosphérique en combinant un schéma spectral pour les grandes échelles et une discrétisation en grille pour la modélisation des processus locaux. Il inclut également des paramétrisations physiques

avancées, notamment pour les transferts radiatifs, la condensation du CO, la formation des nuages et les ondes de gravité générées par la topographie martienne.

La méthodologie repose sur la confrontation des résultats de simulation avec les observations issues des missions Viking et Mars Global Surveyor. L'architecture modulaire du modèle permet de l'adapter à diverses conditions, allant de la surface martienne jusqu'à plus de 100 km d'altitude.

Parmi les avantages notables de ce GCM figurent sa capacité à reproduire fidèlement la circulation atmosphérique globale de Mars ainsi que l'intégration de processus physiques clés. Toutefois, le modèle présente certaines limites, notamment une résolution spatiale encore insuffisante pour capturer les phénomènes locaux tels que les tempêtes de poussière, une simplification des paramétrisations turbulentes, et un coût computationnel relativement élevé.

Modèles de prévisions globaux

Les modèles de prévisions globaux, avec une résolution de 20 à 50 km [13], utilisent des équations simplifiées nécessitant la paramétrisation de phénomènes comme la convection. Dans sa thèse, Benmahdjoub (2011) applique un réseau de neurones à retard temporel (TDNN) pour prévoir les précipitations, exploitant la capacité du modèle à gérer les variations temporelles complexes des séries de précipitations.

Cette méthode apprend à partir de données historiques pour ajuster ses paramètres et prédire les précipitations futures, montrant une bonne corrélation avec les observations. Ses avantages incluent la modélisation efficace de la non-linéarité des précipitations, tandis que ses limites concernent la sensibilité aux choix d'architecture et aux données d'apprentissage.

Modèles à aire limitée (LAM)

Les modèles à aire limitée (LAM) sont utilisés pour des prévisions météorologiques à court terme à l'échelle régionale ou locale, avec une résolution fine permettant de mieux représenter la topographie et la convection atmosphérique, notamment dans les zones montagneuses. Météo-France déploie des modèles opérationnels tels qu'ALADIN, un modèle bi-spectral développé en Europe, et AROME, un modèle non hydrostatique à très haute résolution (environ 1 à 2,5 km), capables de simuler des phénomènes convectifs locaux comme les orages [14].

Ces modèles fonctionnent sur un domaine géographique limité avec des conditions aux limites fournies par des modèles globaux. Cette approche permet de mieux capturer les processus locaux et les interactions relief-atmosphère, essentiels pour la précision des prévisions régionales. L'assimilation continue des données améliore la qualité des simulations.

Cependant, les LAM restent dépendants des conditions aux limites imposées par les modèles globaux, ce qui peut introduire des erreurs. Malgré leur haute résolution, ils ne saisissent pas toutes les complexités atmosphériques, notamment certains processus turbulents et convectifs, et leur coût computationnel élevé limite la durée et la fréquence des prévisions.

Les modèles de prévision numérique du temps (PNT) sont limités par la sensibilité aux conditions initiales, la résolution spatiale et temporelle insuffisante pour capturer les phénomènes à petite échelle, ainsi que par la nécessité de paramétrer des processus physiques complexes comme la convection et la turbulence, ce qui introduit des incertitudes. De plus, leur dépendance aux conditions aux limites et le coût informatique élevé restreignent la fréquence et la précision des prévisions, limitant ainsi leur capacité à représenter pleinement la complexité atmosphérique.

1.4.2 Les méthodes statistiques

Les approches statistiques utilisent des archives de données pour établir des relations entre variables observées à des instants différents et prédire des valeurs futures en fonction de ces relations [13]. Elles se concentrent sur la modélisation des relations statistiques plutôt que sur les processus physiques sous-jacents.

Modèles autoregressifs et moyennes mobiles

En 1927, Yule et Slutsky [15] ont introduit des modèles autoregressifs et des moyennes mobiles, respectivement. Yule a proposé des modèles autoregressifs (AR) de la forme

$$Y_t = \alpha Y_{t-1} + \beta Y_{t-2}. \quad (1.1)$$

Le modèle autoregressif général de Yule est

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t, \quad (1.2)$$

où (ϵ_t) est un bruit blanc.

Slutsky a introduit les moyennes mobiles (MA) sous la forme

$$Y_t = \mu_0 \epsilon_t + \mu_1 \epsilon_{t-1} + \mu_q \epsilon_{t-q}, \quad (1.3)$$

où (ϵ_t) est un bruit blanc.

Les processus AR et MA peuvent être hybrides en processus ARMA.

Les modèles ARMA, ARIMA et SARIMAX

Développés par Box et Jenkins en 1976 [16], les modèles ARMA, ARIMA et SARIMAX sont utilisés pour modéliser les séries chronologiques.

Modèle ARMA

Le modèle ARMA (AutoRegressive Moving Average) combine l'autorégression (AR) et la moyenne mobile (MA) [15]. Il est utilisé pour modéliser des séries chronologiques stationnaires, où la stationnarité est une condition nécessaire pour garantir la constance des propriétés statistiques dans le temps. Le modèle ARMA est défini par les équations :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \quad (1.4)$$

où (ϵ_t) est un bruit blanc.

Modèle ARIMA

Le modèle ARIMA (AutoRegressive Integrated Moving Average) étend le modèle ARMA pour traiter les séries chronologiques non stationnaires en ajoutant une composante d'intégration (I) [15]. Cela permet de modéliser les tendances en différenciant les données pour les rendre stationnaires :

$$\Pi(L)X_t = \Phi(L)(1-L)^d X_t = \Theta(L)\epsilon_t, \quad (1.5)$$

où d représente le degré de différenciation.

Modèle SARIMAX

Le modèle SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors) est une extension du modèle ARIMA qui prend en compte les facteurs exogènes et les effets saisonniers. Il est particulièrement utile pour les séries chronologiques présentant des comportements saisonniers ou influencées par des variables externes. Le modèle SARIMAX est défini par les équations :

$$\Pi(L)X_t = \Phi(L)(1 - L)^d(1 - L^s)^D X_t = \Theta(L)\epsilon_t + \beta Z_t, \quad (1.6)$$

où s représente la saisonnalité, D le degré de différenciation saisonnière, et Z_t les facteurs exogènes.

Ces modèles offrent une gamme complète d'outils pour modéliser des séries chronologiques dans divers contextes, en prenant en compte la stationnarité, les tendances non stationnaires, et les influences saisonnières et exogènes.

Bien que ces méthodes soient simples et largement utilisées, elles présentent des limitations pour les séries non linéaires et les données non stationnaires [17].

Modèle SARIMA-LSTM de Armathé et al. (2024)

Armathé et al. (2024) [18] ont proposé un modèle hybride SARIMA-LSTM pour la prévision climatique dans la zone soudano-sahélienne du Cameroun. Ce modèle combine la capacité de SARIMA à capturer les tendances saisonnières avec celle de LSTM à modéliser les dépendances non linéaires à long terme.

Le modèle est entraîné sur des données historiques de température et de précipitations, issues de sources in-situ et satellitaires. Cette approche permet une meilleure représentation des dynamiques climatiques régionales.

Parmi les avantages, on note une précision importante des prévisions et une meilleure prise en compte des variations saisonnières locales. Toutefois, le modèle présente certaines limites : dépendance à la qualité des données, difficulté de généralisation à d'autres zones, l'absence de détection des pics de précipitations.

Prévision des pluies mensuelles en agriculture

Balah et Belkacem (2018) [19] ont développé un modèle ARIMA pour la prediction des précipitations en agriculture. Ce modèle exploite les composantes d'auto-régression et de moyenne mobile afin de capturer les tendances et les variations saisonnières des précipitations, offrant ainsi une modélisation statistique adaptée aux séries temporelles climatiques.

La méthode consiste à analyser les données historiques de pluies mensuelles, à identifier les paramètres ARIMA optimaux, puis à utiliser ce modèle pour générer des prévisions futures. L'approche permet de prendre en compte les cycles saisonniers tout en restant relativement simple à mettre en œuvre, ce qui la rend accessible pour des applications agricoles.

Les avantages du modèle résident dans sa capacité à fournir des prévisions précises pour les tendances générales des pluies mensuelles, utiles pour la planification agricole. Toutefois, ses limites incluent une sensibilité aux variations saisonnières irrégulières et une difficulté à modéliser les événements extrêmes, comme les fortes précipitations ou les sécheresses soudaines, ce qui peut restreindre son efficacité dans des contextes climatiques très variables.

1.4.3 Les réseaux de neurones

Les réseaux de neurones, inspirés du fonctionnement du cerveau humain, sont des systèmes informatiques conçus pour reconnaître des motifs et résoudre des problèmes complexes. Leur architecture comprend une couche d'entrée qui reçoit les données, des couches cachées où des transformations et des calculs sont effectués, et une couche de sortie qui produit les résultats finaux [20].

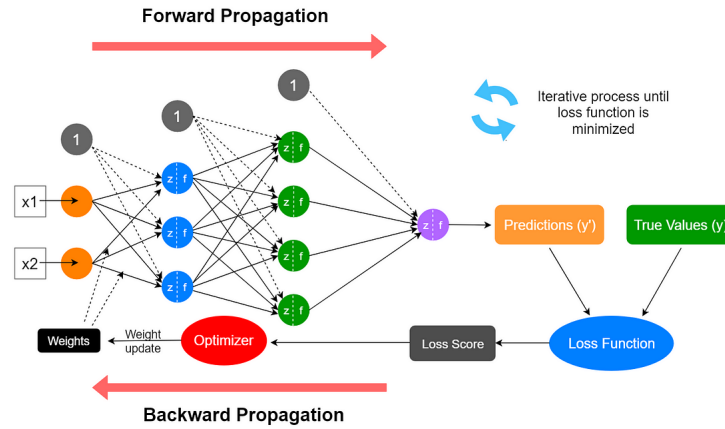


FIGURE 1.3 – Réseau de neurones. (Source : [21])

Le fonctionnement des réseaux de neurones 1.3 repose sur l'apprentissage supervisé¹ et la rétropropagation. L'apprentissage supervisé forme le réseau à partir de paires d'entrée-sortie étiquetées pour découvrir des relations sous-jacentes, tandis que la rétropropagation ajuste les poids des connexions neuronales pour réduire les erreurs de prévision [20].

Les réseaux de neurones sont largement utilisés pour diverses applications telles que la prévision de valeurs futures, la classification de données et la reconnaissance de motifs complexes.

Prévision des précipitations à la station hydrologique de Ca Mau

Tran Anh et al. (2019) [22] ont développé un modèle hybride combinant la transformée en ondelettes discrètes (DWT) et des réseaux de neurones artificiels (ANN) pour améliorer la prévision mensuelle des précipitations à la station hydrologique de Ca Mau, au Vietnam. La DWT décompose les séries temporelles en composantes fréquentielles, permettant d'isoler les signaux importants et de réduire le bruit.

La méthodologie consiste à utiliser les composantes issues de la DWT comme entrées pour un réseau de neurones à propagation avant, qui modélise les relations non linéaires complexes des données climatiques. Les résultats montrent que ce modèle DWT-ANN offre de meilleures performances que les ANN seuls ou d'autres méthodes de prétraitement.

Cette approche améliore la précision des prévisions, utile pour la gestion de l'eau et l'agriculture. Cependant, elle dépend de la qualité des données, du calibrage des paramètres, et sa complexité peut compliquer son utilisation opérationnelle. De plus, ce modèle ne conserve pas explicitement les données à long terme, limitant sa capacité à modéliser les tendances climatiques sur de longues périodes.

1. méthode d'apprentissage machine où un modèle est formé en utilisant un ensemble de données étiquetées.

Prévision des précipitations dans le nord de l'Inde

Mishra et al. (2018) [23] ont développé des modèles de réseaux de neurones artificiels (ANN) pour prévoir les précipitations un et deux mois à l'avance dans le nord de l'Inde. Ils ont utilisé un réseau de neurones à propagation avant, entraîné par rétropropagation avec la fonction d'optimisation Levenberg-Marquardt, en exploitant des données mensuelles de précipitations couvrant la période 1871-2012.

La méthodologie repose sur l'utilisation des précipitations des mois précédents comme seules variables d'entrée pour prédire la pluie future. Le modèle a été calibré par essais et erreurs pour déterminer le nombre optimal de neurones cachés, et évalué par l'erreur quadratique moyenne et la corrélation entre valeurs prédites et observées. Les résultats montrent que le modèle à un mois d'avance est plus performant que celui à deux mois, dont la précision diminue sensiblement avec l'allongement de l'horizon de prévision.

Ce modèle ANN permet de modéliser des relations non linéaires complexes et fournit des prévisions utiles pour la gestion de l'eau et l'agriculture. Cependant, ses limites résident dans la baisse de performance pour les prévisions à plus long terme et dans le fait qu'il ne prend en compte que les précipitations passées, sans intégrer d'autres variables météorologiques susceptibles d'améliorer la qualité et la robustesse des prévisions.

1.4.4 Les méthodes de deep learning

Le deep learning se distingue des réseaux de neurones traditionnels par sa profondeur, caractérisée par un grand nombre de couches cachées permettant de modéliser des relations complexes et de capturer des caractéristiques à plusieurs niveaux d'abstraction [24]. Parmi les architectures spécifiques du deep learning, les LSTM (Long Short-Term Memory) et les GRU (Gated Recurrent Unit) sont des types de réseaux de neurones récurrents (RNN) conçus pour saisir des dépendances à long terme dans les séries temporelles [25].

Prévision des précipitations mensuelles dans la région de Simtokha, Bhoutan

Chhetri et al. (2020) [26] ont comparé plusieurs modèles d'apprentissage profond pour la prévision des précipitations mensuelles à Simtokha, Bhoutan. Leur étude s'est concentrée sur un modèle hybride BLSTM-GRU, combinant les avantages des réseaux LSTM bidirectionnels et des unités GRU pour mieux capturer les dépendances temporelles dans les séries de précipitations.

La méthodologie s'appuie sur des données climatiques mensuelles sur 23 ans, intégrant plusieurs variables météorologiques comme entrées du modèle. Les performances du BLSTM-GRU ont été évaluées et comparées à celles d'autres architectures, montrant une réduction de l'erreur quadratique moyenne (MSE) de 41,1 % par rapport au LSTM classique.

Ce modèle se distingue par sa précision et sa capacité à modéliser les relations complexes entre variables climatiques. Cependant, il nécessite un grand volume de données de qualité, une puissance de calcul importante, et sa généralisation à d'autres régions reste à valider.

prévision des précipitations dans la zone de Palembang

Surta et al. (2023) [5] ont proposé des modèles qui permettent de prédire les précipitations quotidiennes à Palembang pour le mois à venir en utilisant deux réseaux de neurones récurrents avancés : LSTM (Long Short-Term Memory) et GRU (Gated Recurrent Unit). Ces modèles sont

conçus pour exploiter les dépendances temporelles dans les séries chronologiques et modéliser les relations complexes entre les variables météorologiques.

La méthode repose sur dix ans de données météorologiques quotidiennes, incluant la pluie, la température, l'humidité, l'évaporation, l'ensoleillement et la vitesse du vent. Après un prétraitement (normalisation, découpage en ensembles d'entraînement et de test), les modèles sont entraînés avec optimisation manuelle des hyperparamètres. Les performances sont évaluées via le RMSE et le coefficient de détermination R^2 .

Le modèle LSTM s'est montré plus performant que le GRU (RMSE de 7,45 et R^2 de 0,70 contre RMSE de 9,33 et R^2 de 0,54 pour le GRU), démontrant une bonne capacité à modéliser les relations non linéaires et à exploiter des données multivariées pour des prévisions plus précises.

Cependant, l'étude n'évalue pas spécifiquement la capacité des modèles à capter les pics de précipitations ou événements extrêmes. De plus, la performance dépend fortement de la qualité des données, l'optimisation des hyperparamètres reste manuelle, et la généralisation à d'autres régions ou contextes n'est pas démontrée.

Prévision des précipitations dans certaines villes marocaines

Sutskever et al. (2014) [27] ont proposé un modèle LSTM à plusieurs couches pour traiter des séquences temporelles complexes et apprendre les dépendances à long terme, ce qui est essentiel pour la prévision de paramètres météorologiques.

Le modèle repose sur l'empilement de plusieurs couches LSTM, chaque couche traitant la sortie de la précédente, afin d'extraire progressivement des motifs temporels de plus en plus abstraits dans les séries de données météorologiques.

Les LSTM profonds offrent des prévisions précises grâce à leur capacité à capturer les motifs temporels complexes, tout en étant robustes sur de longues séquences. Cependant, ils nécessitent beaucoup de données et de calcul, leur interprétation est complexe, et leur capacité à prédire les événements extrêmes reste limitée.

1.5 Tableaux récapitulatifs

La prévision météorologique, et en particulier la prévision des précipitations, a évolué au fil des années grâce à divers modèles et algorithmes. Le tableau 1.1 présente une chronologie de quelques contributions dans ce domaine, en mettant en évidence les principes, algorithmes, avantages et limites de chaque méthode.

TABLE 1.1 – Chronologie des principales contributions à la prévision des précipitations.

Auteurs	Principe	Méthode	Avantages	Limites
Forget et al. (1999)	Simulation de la dynamique atmosphérique globale	Modèle de circulation générale (GCM) adapté à Mars	Reproduction fidèle de la circulation martienne, modulaire et physiquement complet	Résolution spatiale limitée, coût computationnel élevé, pas spécifique aux petites zones à fort aléa
Benmahdjoub (2011)	Réseau neuronal appliqué à la météo	TDNN sur données de précipitations	Bonne modélisation des non-linéarités	Sensible à l'architecture et aux données

Auteurs	Principe	Méthode	Avantages	Limites
Armathé et al. (2024)	Hybridation SARIMA-LSTM pour séries climatiques	SARIMA pour saisonnalité + LSTM pour non-linéarités	Bonne précision, adapte les variations saisonnières	Sensible aux données, pas adapté à toutes les zones, ne capte pas les pics
Balah et Belkacem (2018)	Modèle statistique pour l'agriculture	ARIMA sur données mensuelles	Simplicité, utile pour tendances globales	Faible prise en compte des extrêmes, peu robuste aux irrégularités saisonnières
Tran Anh et al. (2019)	Hybridation ondelettes-neurones pour capturer les signaux utiles	DWT + ANN	Bonne précision prévisionnelle, extraction multi-échelle	Dépendance aux données et paramètres, complexité opérationnelle
Mishra et al. (2018)	ANN pour précipitations mensuelles	ANN avec fonction Levenberg-Marquardt	Prédiction utile à court terme, non-linéarités captées	Performance diminue sur prévision long terme, variables d'entrée limitées
Chhetri et al. (2020)	Apprentissage profond hybride pour données climatiques	BLSTM-GRU	Très précis, modélisation complexe des séries	Besoin de beaucoup de données et de calcul, difficulté de généralisation
Surta et al. (2023)	Prévision quotidienne à l'aide de LSTM et GRU	Comparaison LSTM vs GRU sur données météo multivariées	Bonne précision LSTM, relations non-linéaires exploitées	Ne capte pas les extrêmes, dépendance à l'optimisation manuelle
Sutskever et al. (2014)	LSTM profond pour séquences temporelles	Empilement de couches LSTM	Modélisation fine des motifs temporels, robustesse sur longues séquences	Besoin massif de données, difficulté d'interprétation, peu d'informations sur les extrêmes

1.6 Conclusion

En somme, ce chapitre a fourni une vue d'ensemble des fondamentaux nécessaires à la modélisation des précipitations dans la zone de Maga. Nous avons exploré les aspects essentiels liés aux précipitations, y compris leur formation, leurs types et les méthodes de mesure. Nous avons également abordé les séries temporelles, en mettant en exergue les approches de modélisation anciennes et modernes, des réseaux de neurones et du deep learning. Cette compréhension intégrée des concepts et méthodes nous prépare à appliquer ces outils de manière efficace pour améliorer les prévisions des précipitations et mieux répondre aux défis spécifiques de la zone de Maga. Nous avons terminé le chapitre en donnant un tableau comparatif des solutions antérieures proposées pour la prévisions des séries temporelles.

CHAPITRE 2

DESCRIPTION DU MODÈLE LSTM-GRU POUR LA PRÉVISION DES PRÉCIPITATIONS

Dans ce chapitre, nous détaillons le problème de prévision des précipitations à partir de données météorologiques multivariées. Nous commençons par formaliser ce problème, puis présentons les architectures récurrentes LSTM et GRU, en exposant leurs principes fondamentaux et avantages respectifs. Ensuite, nous introduisons notre modèle hybride combinant ces deux architectures, à pierre angulaire de notre approche. Enfin, nous décrivons l'algorithme de prévision basé sur ce modèle ainsi que son analyse de complexité théorique.

2.1 Définition du problème de prévision des précipitations

La prévision des précipitations repose sur l'exploitation de variables météorologiques telles que la température, l'humidité relative, la pression atmosphérique, la vitesse et la direction du vent, la nébulosité, le rayonnement solaire, le point de rosée, et l'indice de chaleur.

On considère une série temporelle multivariée de données météorologiques :

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]$$

où chaque vecteur $\mathbf{x}_t \in \mathbb{R}^d$ regroupe les observations météorologiques à l'instant t .

La série correspondante des précipitations est :

$$\mathbf{Y} = [y_1, y_2, \dots, y_T]$$

avec $y_t \in \mathbb{R}$ la valeur des précipitations à l'instant t .

L'objectif est d'estimer la valeur future des précipitations à l'horizon $t + \tau$ à partir d'une séquence historique, c'est-à-dire de construire un modèle prédictif non linéaire f_θ , paramétré par θ ,

tel que :

$$y_{t+\tau} = f_{\theta}([\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-l+1}]) + \varepsilon_{t+\tau} \quad (2.1)$$

où $\varepsilon_{t+\tau}$ est un terme d'erreur.

La tâche consiste à apprendre les paramètres θ minimisant l'erreur de prédiction sur un ensemble d'apprentissage, en exploitant la richesse temporelle et multivariée des données. Les réseaux de neurones récurrents et leurs variantes avancées, tels que LSTM et GRU, sont particulièrement adaptés à la modélisation des dépendances temporelles complexes [28?].

2.2 Réseaux LSTM (Long Short-Term Memory)

Les réseaux LSTM sont une variante avancée des RNN conçue pour pallier la difficulté des RNN classiques à apprendre des dépendances à long terme dans les séquences [?].

Le LSTM introduit une cellule mémoire C_t régulée par trois portes principales :

- La **porte d'oubli (forget gate)** f_t permet au réseau de sélectionner les informations historiques à oublier [29]. La portion d'information à oublier est déterminée par la formule (2.2).
- La **porte d'entrée** (également appelée **porte de mémoire**) i_t est utilisée pour renforcer la mémoire de certaines informations historiques [30, 31]. Les équations (2.3) et (2.4) représentent l'information d'entrée.
- Enfin, la **porte de sortie** o_t permet de réviser complètement les informations contenues dans la mémoire à long terme et de générer les signaux de sortie. Elle est définie par les formules (2.6) et (2.7).

Les équations principales sont :

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2.2)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2.3)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (2.4)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (2.5)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2.6)$$

$$h_t = o_t \odot \tanh(C_t) \quad (2.7)$$

où σ est la fonction sigmoïde, $\tanh$ la tangente hyperbolique, et $\odot$ le produit de Hadamard (élément par élément).

Cette architecture permet au LSTM de conserver efficacement des informations sur de longues séquences, ce qui est crucial pour modéliser des phénomènes météorologiques complexes. L'architecture d'une cellule LSTM est représentée à la Figure 2.1.

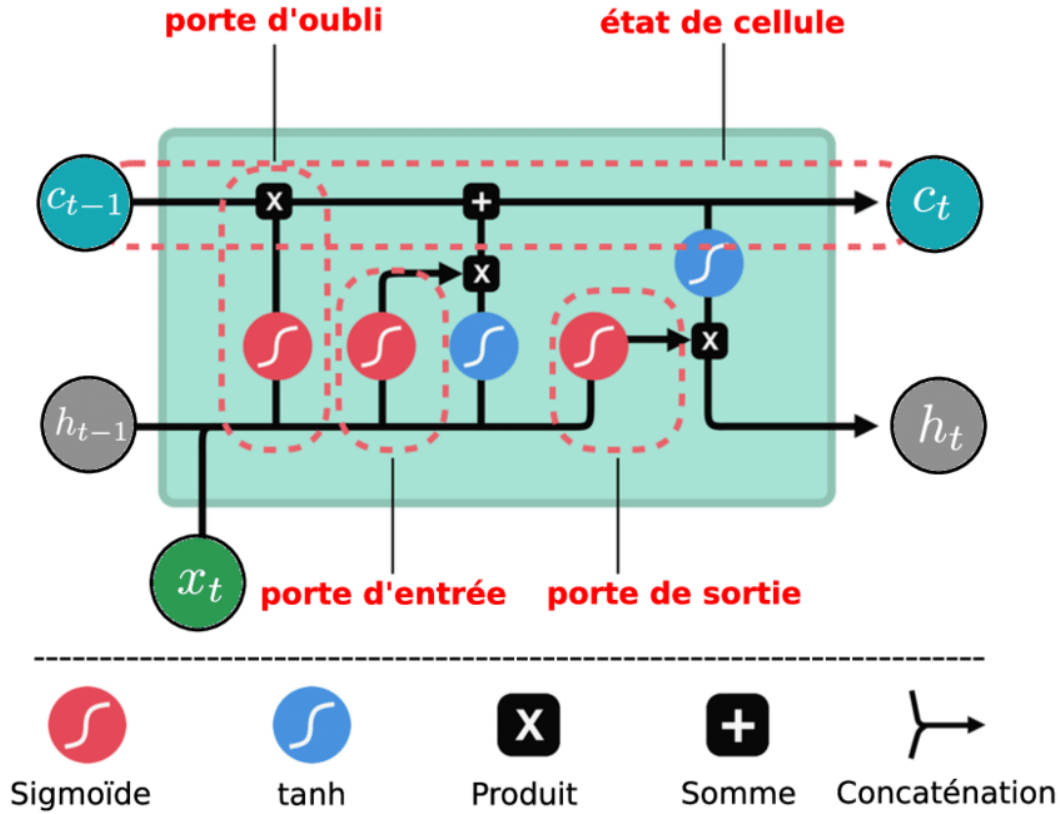


FIGURE 2.1 – Architecture d'une cellule LSTM. Source : [32]

2.3 Réseaux GRU (Gated Recurrent Unit)

Introduite initialement par Cho et al. [33], la GRU est une variante simplifiée du LSTM, conçue pour conserver les performances tout en réduisant la complexité du modèle. Contrairement au LSTM, la GRU ne possède pas d'état de cellule séparé, mais repose uniquement sur l'état caché pour le transfert d'information entre les unités récurrentes.

Elle comporte deux portes principales :

- **La porte de réinitialisation** r_t , qui contrôle la quantité d'informations passées à oublier.
- **La porte de mise à jour** z_t , qui détermine la proportion d'informations passées à conserver dans l'état caché actuel. Elle combine les fonctions des portes d'entrée et d'oubli du LSTM.

Ce mécanisme permet à la GRU de capturer efficacement des dépendances à long terme tout en conservant une architecture plus légère. Ces deux portes régulent le flux d'information et assurent la stabilité du gradient, ce qui aide à résoudre le problème de disparition du gradient [31, 34].

La Figure 2.2 présente la structure d'un bloc GRU avec ses différentes composantes.

Le fonctionnement d'une GRU peut être formellement exprimé à l'aide des équations suivantes :

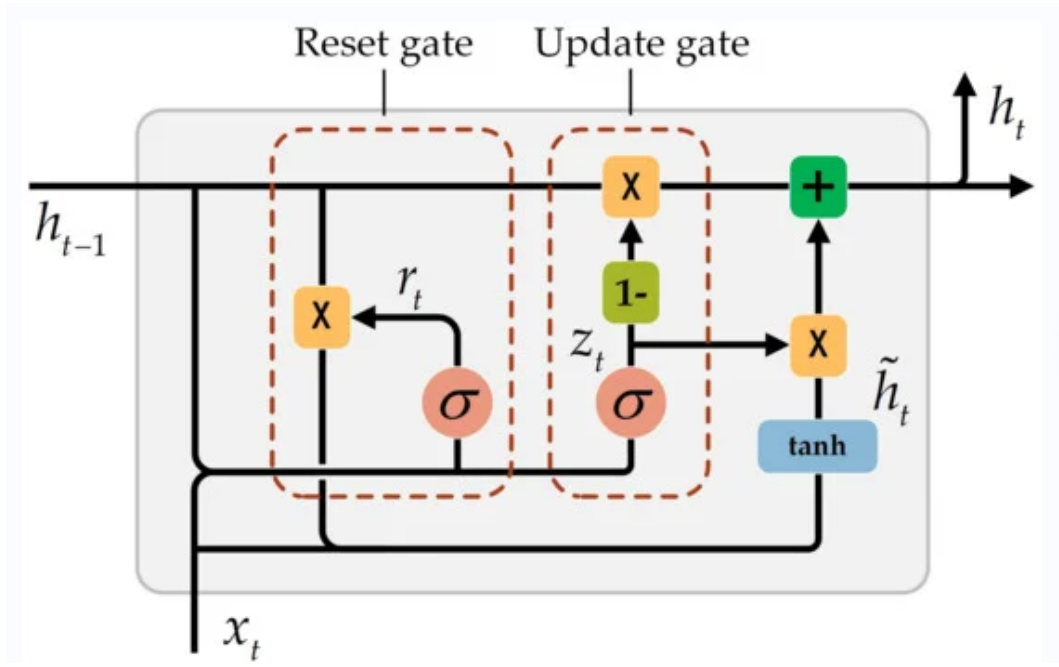


FIGURE 2.2 – Structure d'un bloc GRU (Source : [35])

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (2.8)$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2.9)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \quad (2.10)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2.11)$$

où x_t est l'entrée à l'instant t , h_{t-1} est l'état caché précédent, r_t est la porte de réinitialisation, z_t est la porte de mise à jour, $\tilde{h}_t$ est l'état candidat, et h_t est l'état caché mis à jour.

Les matrices W_* , U_* et les vecteurs b_* sont respectivement les poids et biais associés à chaque composant.

2.4 Architecture du modèle hybride LSTM-GRU

Le modèle hybride LSTM-GRU, illustré à la Figure 2.3, combine les avantages des couches LSTM, capables de capturer les dépendances à long terme, et des couches GRU, plus légères et efficaces pour affiner les représentations temporelles. Ce choix d'architecture permet de bénéficier à la fois de la mémoire longue durée offerte par les LSTM et de l'efficacité computationnelle des GRU, aboutissant à une solution adaptée à la prédiction de séries temporelles climatiques telles que les précipitations.

Le modèle fonctionne comme suit :

- En entrée, il reçoit une séquence temporelle de vecteurs d'observations $[\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-l+1}]$,

où l est la longueur de la séquence historique.

- Il est composé de $L \in \mathbb{N}^*$ couches LSTM suivies de $G \in \mathbb{N}^*$ couches GRU. Après chaque couche, les processus de Dropout et de Batch Normalization sont intégrés afin d'assurer la régularisation et la stabilité de l'entraînement.
- La sortie est produite par une couche dense avec une activation ReLU, afin de garantir la positivité des valeurs de précipitation prédites.

Phase d'apprentissage : Le modèle est entraîné pour prédire la précipitation à l'instant t à partir de la séquence historique d'entrées $[\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-l+1}]$. Formellement :

$$[h_1^{\text{LSTM}}, \dots, h_l^{\text{LSTM}}] = \text{LSTM}([\mathbf{x}_t, \dots, \mathbf{x}_{t-l+1}]) \quad (2.12)$$

$$[h_1^{\text{GRU}}, \dots, h_l^{\text{GRU}}] = \text{GRU}([h_1^{\text{LSTM}}, \dots, h_l^{\text{LSTM}}]) \quad (2.13)$$

$$\hat{y}_t = \text{ReLU}(W_{\text{dense}} h_l^{\text{GRU}} + b_{\text{dense}}) \quad (2.14)$$

où h_l^{GRU} est l'état caché final de la dernière couche GRU à la fin de la séquence d'entrée.

Phase de prédiction multi-horizon : Pour prédire les précipitations à un horizon futur $t + \tau$, le modèle entraîné est appliqué sur la séquence d'entrées correspondante décalée dans le temps :

$$[\mathbf{x}_{t+\tau}, \mathbf{x}_{t+\tau-1}, \dots, \mathbf{x}_{t+\tau-l+1}]$$

où les $\mathbf{x}$. dans notre cas sont des valeurs des observations prédites à partir du modèle hybride. La prédiction s'écrit alors :

$$\hat{y}_{t+\tau} = \text{ReLU}(W_{\text{dense}} h_l^{\text{GRU}} + b_{\text{dense}}),$$

avec h_l^{GRU} obtenu en traitant la séquence d'entrée décalée à travers les couches LSTM puis GRU.

Couche LSTM

Chaque couche LSTM traite la séquence d'entrée et produit une séquence d'états cachés h_t^{LSTM} , calculés à chaque pas de temps à l'aide des équations suivantes :

$$h_t^{\text{LSTM}} = o_t \odot \tanh(C_t)$$

Les portes d'oubli, d'entrée et de sortie suivent les équations classiques des LSTM (2.2 à 2.7).

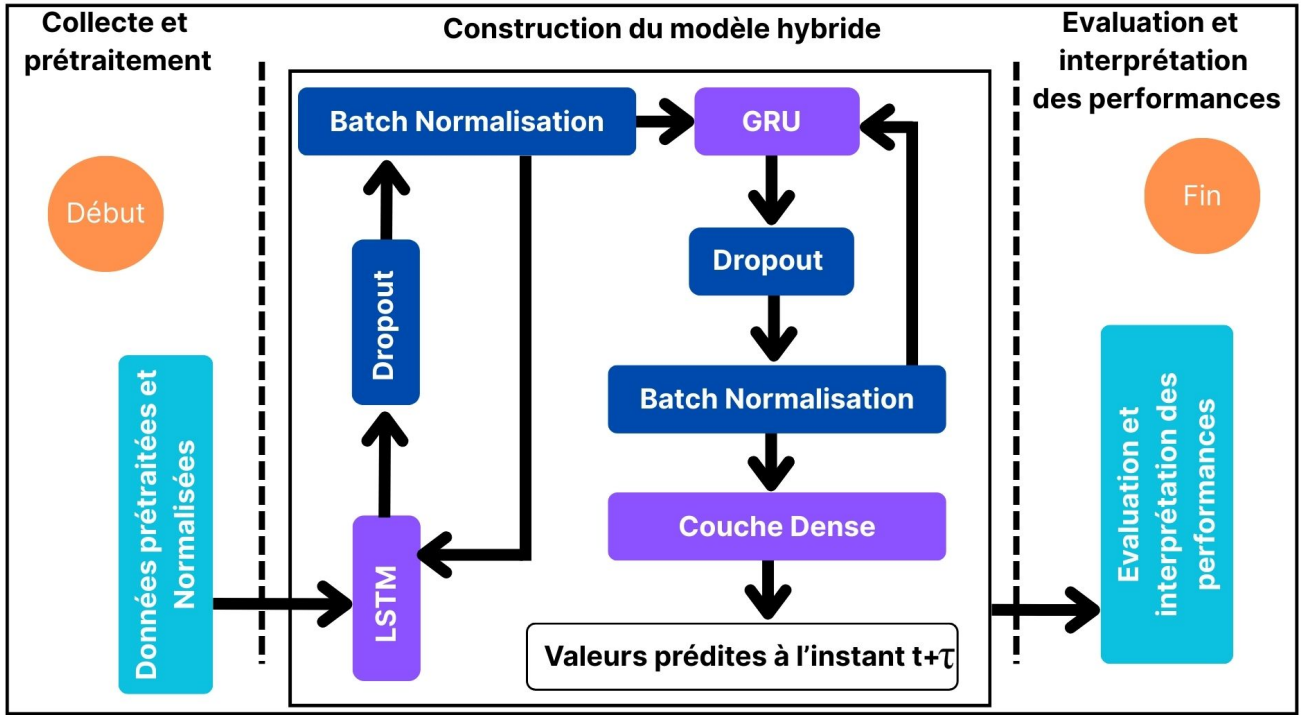


FIGURE 2.3 – Architecture du modèle hybride LSTM-GRU.

Couche GRU

Les couches GRU raffinent la représentation extraite par les LSTM. Elles utilisent deux portes principales : la porte de réinitialisation r_t , qui contrôle la quantité d'informations passées à oublier, et la porte de mise à jour z_t , qui régule la quantité d'informations à conserver.

Ces portes sont calculées à partir de la concaténation des états cachés précédents h_{t-1}^{GRU} et des sorties LSTM h_t^{LSTM} :

$$r_t = \sigma(W_r[h_{t-1}^{\text{GRU}}; h_t^{\text{LSTM}}] + b_r) \quad (2.15)$$

$$z_t = \sigma(W_z[h_{t-1}^{\text{GRU}}; h_t^{\text{LSTM}}] + b_z) \quad (2.16)$$

$$\tilde{h}_t = \tanh(W_h[r_t \odot h_{t-1}^{\text{GRU}}; h_t^{\text{LSTM}}] + b_h) \quad (2.17)$$

$$h_t^{\text{GRU}} = (1 - z_t) \odot h_{t-1}^{\text{GRU}} + z_t \odot \tilde{h}_t \quad (2.18)$$

où $[\cdot; \cdot]$ désigne la concaténation vectorielle.

L'état initial h_0^{GRU} est initialisé à zéro.

Paramètres du modèle

- $\mathbf{x}_t \in \mathbb{R}^d$: vecteur d'entrée à l'instant t
- $h_t^{\text{LSTM}}, h_t^{\text{GRU}} \in \mathbb{R}^h$: états cachés des couches LSTM et GRU

- $W_f, W_i, W_C, W_o \in \mathbb{R}^{h \times (d+h)}$: matrices de poids des portes LSTM
- $W_r, W_z, W_h \in \mathbb{R}^{h \times 2h}$: matrices de poids des portes GRU
- $W_{\text{dense}} \in \mathbb{R}^{1 \times h}, b_{\text{dense}} \in \mathbb{R}$: paramètres de la couche de sortie
- σ : fonction sigmoïde, $\tanh$: tangente hyperbolique, $\odot$: produit de Hadamard (élément par élément)

2.5 Algorithme de prévision séquentielle multi-horizon avec LSTM-GRU

Notre démarche repose sur l'utilisation séquentielle et itérative d'un même modèle hybride LSTM-GRU pour modéliser la dynamique des variables météorologiques et prédire les précipitations futures. Elle s'organise en deux grandes phases : l'apprentissage supervisé des données historiques, suivi d'un processus de prédiction multi-horizon.

1. Apprentissage supervisé des précipitations

Le modèle LSTM-GRU est entraîné pour apprendre la relation entre les séquences historiques des variables météorologiques $\mathbf{X}$ et la variable cible $\mathbf{Y}$, à savoir les précipitations. Formellement, on apprend à prédire y_t à partir de la séquence d'entrée $[\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-l+1}]$, où l est la fenêtre temporelle d'entrée.

L'objectif est d'optimiser les paramètres θ du modèle pour minimiser l'erreur de prédiction sur l'ensemble d'apprentissage, mesurée par la fonction de perte **Mean Squared Error (MSE)**. La rétropropagation du gradient est utilisée pour ajuster les poids du modèle en fonction de cette erreur.

2. Prédiction séquentielle multi-horizon

La prédiction se déroule en deux étapes imbriquées :

1. **Prédiction des variables météorologiques futures** : Pour chaque variable météorologique x^j (avec $j = 1, \dots, d$), on prédit ses valeurs futures à partir de ses observations passées selon :

$$\hat{x}_{t+1}^j, \dots, \hat{x}_{t+\tau}^j = f_{\theta}^j([x_t^j, x_{t-1}^j, \dots, x_{t-l+1}^j])$$

Chaque variable est prédite individuellement à l'aide du modèle LSTM-GRU adapté à ses spécificités temporelles.

2. **Prédiction des précipitations futures** : À chaque horizon $k = 1, \dots, \tau$, on constitue le vecteur d'entrée :

$$\hat{\mathbf{x}}_{t+k} = [\hat{x}_{t+k}^1, \dots, \hat{x}_{t+k}^d]$$

en concaténant les valeurs prédites des différentes variables climatiques. Le modèle entraîné est ensuite utilisé pour prédire la précipitation correspondante :

$$\hat{y}_{t+k} = f_{\theta}(\hat{\mathbf{x}}_{t+k})$$

Cette stratégie permet d'expliciter la dynamique des variables climatiques avant de prédire la variable cible, renforçant ainsi la précision et la cohérence des prévisions multi-horizon.

Algorithm 1: Prédiction séquentielle multi-horizon avec modèle hybride LSTM-GRU

Input: Données historiques des variables météorologiques $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_T]$,

Données historiques des précipitations $\mathbf{Y} = [y_1, \dots, y_T]$,

Horizon de prédiction τ ,

Longueur de la fenêtre temporelle l

Output: Prédictions des précipitations futures $\hat{y}_{T+1}, \dots, \hat{y}_{T+\tau}$

```

1 Phase 1 : Apprentissage supervisé
2 for chaque instant  $t = l, \dots, T$  do
3   Extraire la séquence d'entrée :  $\mathbf{S}_t = [\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-l+1}]$ 
4   Associer la cible :  $y_t$ 
5   Calculer la prédiction :  $\hat{y}_t = f_\theta(\mathbf{S}_t)$ 
6   Calculer la perte MSE :  $\mathcal{L}_t = \frac{1}{N} \sum (y_t - \hat{y}_t)^2$ 
7   Mettre à jour les paramètres  $\theta$  par rétropropagation du gradient pour minimiser  $\mathcal{L}_t$ 
8 Répéter jusqu'à convergence sur tout l'ensemble d'apprentissage

9 Phase 2 : Prédiction multi-horizon
10 for chaque variable météorologique  $j = 1, \dots, d$  do
11   Initialiser la séquence historique  $S^j = [x_{T-l+1}^j, \dots, x_T^j]$ 
12   for  $k = 1$  to  $\tau$  do
13     Prédire la valeur future :
           
$$\hat{x}_{T+k}^j = f_\theta^j(S^j)$$

           Mettre à jour la séquence  $S^j \leftarrow [S^j[2:], \hat{x}_{T+k}^j]$  (fenêtre glissante)
14 for  $k = 1$  to  $\tau$  do
15   Construire le vecteur d'entrée des variables climatiques prédites :
           
$$\hat{\mathbf{x}}_{T+k} = [\hat{x}_{T+k}^1, \dots, \hat{x}_{T+k}^d]$$

           Prédire la précipitation correspondante :
           
$$\hat{y}_{T+k} = f_\theta(\hat{\mathbf{x}}_{T+k})$$

16 return  $\hat{y}_{T+1}, \dots, \hat{y}_{T+\tau}$ 

```

2.5.1 Explication de l'algorithme

Cet algorithme s'organise en deux phases complémentaires :

- **Apprentissage supervisé :** Le modèle hybride LSTM-GRU est entraîné à partir des données historiques pour apprendre la relation entre les séquences des variables météorologiques et les précipitations. La fonction de perte utilisée est la Mean Squared Error (MSE), qui mesure la

moyenne des carrés des écarts entre les valeurs réelles et prédites. La rétropropagation du gradient, notamment la rétropropagation dans le temps (BPTT), est appliquée pour corriger les erreurs en ajustant les paramètres du modèle.

- **Prédiction multi-horizon** : La prédiction s’effectue en deux étapes imbriquées. D’abord, chaque variable météorologique est prédite individuellement sur l’horizon futur τ en utilisant ses propres observations passées. Ensuite, à chaque instant futur, les variables climatiques prédites sont assemblées pour former le vecteur d’entrée du modèle, qui prédit la précipitation correspondante. Cette approche explicite la dynamique propre à chaque variable climatique avant la prédiction de la variable cible, améliorant la précision et la cohérence des prévisions.

2.6 Analyse de la complexité de l’algorithme

Cette section présente l’analyse de la complexité temporelle et spatiale de l’algorithme de prévision séquentielle multi-horizon basé sur le modèle hybride LSTM-GRU.

2.6.1 Complexité temporelle

La complexité temporelle correspond au temps de calcul nécessaire pour effectuer une passe avant (inférence) et une passe arrière (rétropropagation) dans le réseau, pour une séquence de données.

- **Complexité d’une couche LSTM** : Chaque couche LSTM traite une séquence de longueur l , avec des vecteurs d’entrée de dimension d et un état caché de dimension h . À chaque pas de temps, les opérations principales sont des multiplications matricielles entre les poids $W \in \mathbb{R}^{4h \times (d+h)}$ (correspondant aux 3 portes plus la mise à jour du candidat d’état) et les vecteurs concaténés $[x_t, h_{t-1}]$, ainsi que des calculs non linéaires (sigmoïde, tanh). La complexité par pas est donc en $O(h(d+h))$, et pour la séquence complète $O(l \times h(d+h))$.
- **Complexité d’une couche GRU** : Une couche GRU effectue des opérations matricielles avec des poids $W \in \mathbb{R}^{3h \times (d+h)}$ (correspondant aux 2 portes plus le candidat d’état), avec une complexité par pas similaire, $O(h(d+h))$. Pour la séquence complète, la complexité est $O(l \times h(d+h))$, légèrement inférieure à celle des LSTM.
- **Complexité totale du modèle hybride** : Le modèle comprend L couches LSTM et G couches GRU, donc la complexité temporelle totale par séquence est :

$$O(l \times h(d+h) \times (4L + 3G))$$

où les coefficients 4 et 3 correspondent au nombre de matrices de poids (liées aux portes et candidats) dans LSTM et GRU respectivement.

- **Phase d’apprentissage** : L’apprentissage implique une passe avant et une rétropropagation (BPTT) pour chaque séquence, doublant approximativement le coût de calcul par itération.

- **Phase de prédiction multi-horizon** : La prédiction des variables météorologiques futures pour chaque variable j sur l’horizon τ est itérative, avec une complexité linéaire en τ . La prédiction des précipitations à chaque horizon est également linéaire en τ .

2.6.2 Complexité spatiale

La complexité spatiale correspond à la mémoire nécessaire pour stocker les paramètres du modèle et les états intermédiaires durant l’entraînement et l’inférence.

- **Paramètres du modèle** :
 - Pour chaque couche LSTM, les matrices de poids ont une taille approximative de $4h \times (d + h)$, soit $4h(d + h)$ paramètres.
 - Pour chaque couche GRU, la taille est environ $3h \times (d + h)$, soit $3h(d + h)$ paramètres.
 - La couche dense finale a h paramètres plus un biais.

Le total est donc :

$$4Lh(d + h) + 3Gh(d + h) + h + 1$$

- **États intermédiaires** : Durant l’entraînement, il faut stocker les états cachés h_t , les cellules C_t (pour LSTM), et les sorties intermédiaires pour la rétropropagation. La mémoire nécessaire est proportionnelle à $l \times h \times (L + G)$.
- **Données d’entrée** : Stockage des séquences d’entrée de taille $l \times d$.

Le modèle hybride LSTM-GRU présente une complexité temporelle et spatiale linéaire en la longueur de la séquence l et en l’horizon de prédiction τ , mais polynomiale en la taille des états cachés h et en la dimension d’entrée d .

2.7 Conclusion

Dans ce chapitre, nous avons d’abord décrit formellement le problème de la prévision des précipitations. Nous avons ensuite présenté les architectures LSTM et GRU individuellement. Après cela, nous avons introduit l’architecture hybride combinant ces deux modèles, LSTM et GRU. Enfin, nous avons détaillé l’algorithme d’apprentissage du modèle LSTM-GRU et analysé sa complexité théorique tant spatiale que temporelle. Nous présenterons l’étude expérimentale dans le chapitre suivant.

CHAPITRE 3

RESULTATS D'EXPERIMENTATION

Ce chapitre a pour but de présenter les expérimentations et les différents résultats obtenus, suivis de leur interprétation.

3.1 Protocole expérimental

3.1.1 Environnement et outils de travail

Les expériences de cette étude ont été menées sur une machine équipée d'un processeur Dell Intel Core i7-8650U cadencé à 1.90GHz (jusqu'à 2.11 GHz) avec 16GB de RAM. L'implémentation de l'algorithme a été réalisée en utilisant Google Colab et Jupyter Notebook version 7.0.6 en langage Python.

Pour la réalisation de ce projet de prévision des précipitations, plusieurs bibliothèques ont été utilisées. NumPy et Pandas facilitent la manipulation et l'analyse des données. TensorFlow, à travers Keras, est utilisé pour construire des modèles de réseaux de neurones LSTM et GRU, intégrant des couches de régularisation et de normalisation pour améliorer la performance. L'optimiseur Adam est choisi pour la descente de gradient. Les callbacks ReduceLROnPlateau et EarlyStopping ajustent dynamiquement le taux d'apprentissage et arrêtent l'entraînement en cas de stagnation des performances. Matplotlib permet de visualiser les données et résultats, tandis que scikit-learn est utilisé pour évaluer les prévisions avec des métriques de performances. Enfin, MinMaxScaler normalise les données pour optimiser les performances du modèle.

3.1.2 Description du jeu de données

Le jeu de données "NASA/POWER CERES/MERRA2 Native Resolution Daily Data"¹ fournit des mesures météorologiques quotidiennes de janvier 1982 à décembre 2023, pour une latitude de 10.8449°N et une longitude de 14.9498°E. Il comprend 15 340 observations avec six variables clés :

- **T2M_MIN** : Température minimale à 2 mètres (°C).
- **T2M_MAX** : Température maximale à 2 mètres (°C).
- **PRECTOTCORR** : Précipitations corrigées (mm/jour).
- **RH2M** : Humidité relative à 2 mètres (%).
- **WS10M_MAX** : Vitesse maximale du vent à 10 mètres (m/s).
- **WS10M_MIN** : Vitesse minimale du vent à 10 mètres (m/s).

Le tableau 3.1 présente un aperçu détaillé des données utilisées, fournissant des informations sur les conditions météorologiques et les précipitations dans la zone d'étude. Ces données sont essentielles pour comprendre les conditions climatiques et développer notre modèle prédictif des précipitations.

TABLE 3.1 – Informations du dataset

	max width=
Dates	01/01/1982 à 31/12/2023
Emplacement	Latitude 10.8449°, Longitude 14.9498°
Élévation	311.18 mètres (moyenne pour la région 0.5 x 0.625 degré lat/lon)
Valeur pour les données manquantes	-999
Variables	T2M_MIN T2M_MAX PRECTOTCORR RH2M WS10M_MAX WS10M_MIN
Nombre d'observations	15 340

Ces données sont utilisées dans notre travail pour mieux comprendre les conditions météorologiques et proposer un modèle prédictif des précipitations.

3.1.3 Relation entre les variables météorologiques et les précipitations

Cette section explore les relations entre les précipitations et plusieurs variables météorologiques, illustrées par divers graphiques. Les graphiques 3.5, 3.1, 3.4, 3.3 et 3.2 montrent comment les

1. Source : <https://power.larc.nasa.gov/data-access-viewer/>

précipitations évoluent en fonction de l'humidité relative, des températures minimales et maximales, ainsi que des vitesses maximales et minimales du vent.

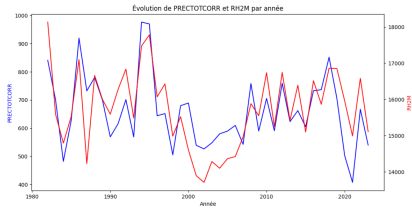


FIGURE 3.1 – Précipitations vs Humidité relative

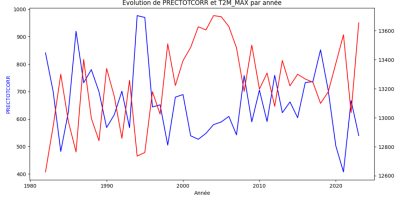


FIGURE 3.2 – Précipitations vs Température maximale

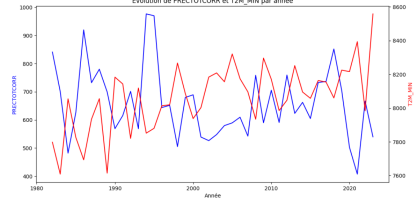


FIGURE 3.3 – Précipitations vs Température minimale

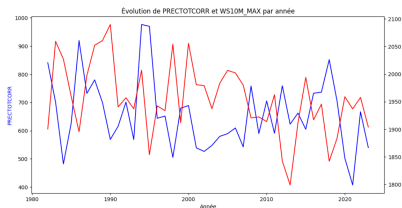


FIGURE 3.4 – Précipitations vs Vitesse maximale du vent

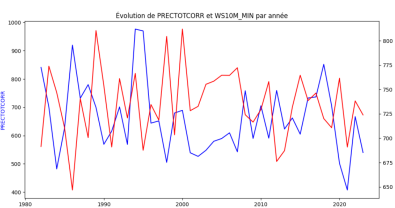


FIGURE 3.5 – Précipitations vs Vitesse minimale du vent

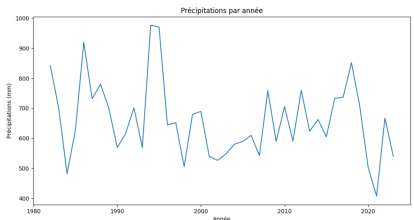


FIGURE 3.6 – Évolution des précipitations

Les analyses graphiques révèlent que les précipitations sont fortement corrélées positivement avec l'humidité relative, suggérant qu'une humidité accrue entraîne une augmentation des précipitations. En revanche, les précipitations présentent des corrélations négatives modérées avec les températures maximales et minimales ainsi qu'avec les vitesses maximales et minimales du vent, indiquant que des températures plus élevées et des vitesses de vent plus importantes sont associées à des niveaux de précipitations plus faibles.

La figure 3.7 illustre ces relations, mettant en évidence la forte corrélation positive entre les précipitations ("PRECTOTCORR") et l'humidité relative ("RH2M"), ainsi que les corrélations négatives avec les températures maximales ("T2M_MAX") et les vitesses du vent ("WS10M_MAX" et "WS10M_MIN").

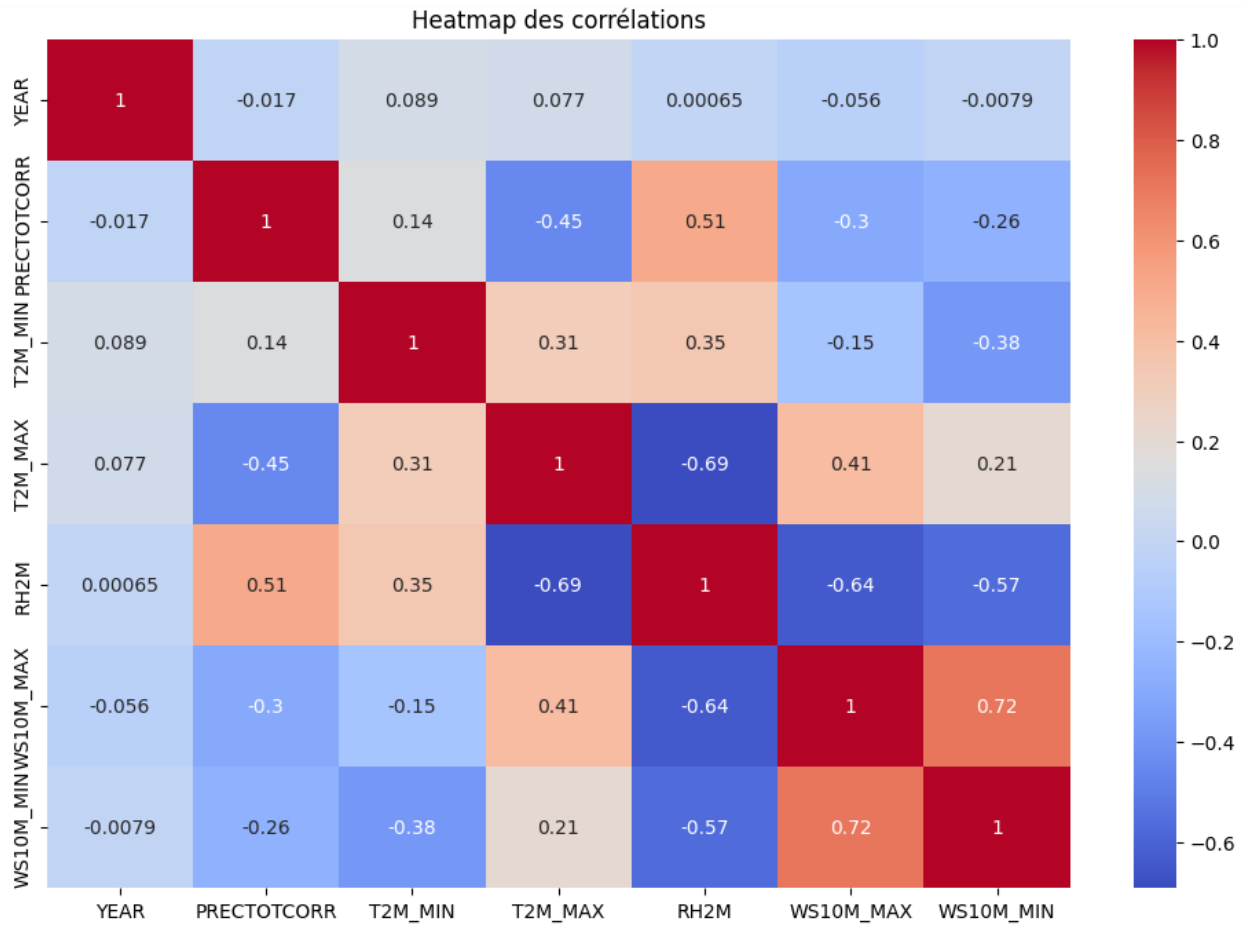


FIGURE 3.7 – Corrélation entre les précipitations et les autres variables

Ces relations soulignent l'importance de ces variables dans la prévision des précipitations.

3.1.4 Préparation et Prétraitement des Données

Pour préparer les données pour l'entraînement du modèle, nous avons d'abord monté le disque Google Drive dans Colab et chargé le fichier CSV dans un DataFrame Pandas. Les colonnes de températures (T2M_MIN, T2M_MAX) et de vitesses du vent (WS10M_MAX, WS10M_MIN) ont été converties en valeurs numériques. Les valeurs manquantes ont été remplacées par les moyennes ou la médiane, tandis que les valeurs manquantes dans PRECTOTCORR ont été remplacées par zéro. Les valeurs aberrantes ont également été supprimées.

Ensuite, nous avons effectué une normalisation Min-Max des colonnes sélectionnées (T2M_MIN, T2M_MAX, PRECTOTCORR, RH2M, WS10M_MAX, WS10M_MIN) pour les mettre dans une plage de 0 à 1. La formule utilisée est :

$$X_{\text{normalisé}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (3.1)$$

où X est la valeur d'origine, $X_{\min}$ et $X_{\max}$ sont les valeurs minimales et maximales respectivement, et $X_{\text{normalisé}}$ est la valeur normalisée.

Nous avons ensuite agrégé les données mensuellement en calculant les moyennes pour les variables pertinentes et ajouté une colonne **MONTH**. Les données ont été divisées en ensembles d'entraînement (jusqu'à décembre 2022) et de test (à partir de janvier 2023). Enfin, les données ont été organisées en séquences temporelles de 12 mois pour l'entraînement et la validation du modèle. La figure 3.8 illustre ces étapes de prétraitement.



FIGURE 3.8 – Prétraitement des données

3.1.5 Les Métriques d'évaluation des performances

Pour évaluer notre modèle de prévision des précipitations, nous utilisons quatre indicateurs : le R^2 , le RMSE et le MAE.

Le RMSE (*Root Mean Squared Error*) évalue la dispersion des erreurs de prévision, offrant une mesure de la magnitude des erreurs :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.2)$$

Le coefficient de détermination R^2 mesure la proportion de la variance des valeurs observées expliquée par le modèle. Il varie entre 0 et 1, avec 1 indiquant une parfaite concordance entre les valeurs observées et prédites :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.3)$$

Le MAE (*Mean Absolute Error*) calcule la moyenne des erreurs absolues, étant moins sensible aux valeurs aberrantes que le MSE :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.4)$$

Ici, y_i est la valeur observée, $\hat{y}_i$ est la valeur prédite, $\bar{y}$ est la moyenne des valeurs observées, et n est le nombre total d'échantillons. Ces métriques permettent d'évaluer la performance du modèle et de comparer différentes approches.

3.1.6 Les configurations du Modèle LSTM-GRU

Nous avons testé plusieurs configurations de modèles LSTM-GRU, détaillées dans le tableau 3.2. Chaque configuration utilise le MSE comme fonction de perte. Les colonnes du tableau présentent les différentes combinaisons d'unités, d'époques, de tailles de lot, et d'autres paramètres. La validation du modèle se fait en utilisant les données de validation pour surveiller la performance sur des données non vues pendant l'entraînement.

TABLE 3.2 – Configurations des modèles LSTM-GRU et leurs paramètres

Modèle	Nombre d'unités	Nb époques	Batch size	Learning rate
1	LSTM(512, 256), GRU(128, 128)	400	16	0.001
2	LSTM(512, 256), GRU(128, 128)	400	12	0.0001
3	LSTM(256), GRU(128, 128)	400	10	0.0001
4	LSTM(512, 256), GRU(128, 128)	400	12	0.01
5	LSTM(512, 256), GRU(128, 128)	400	10	0.001
6	LSTM(512, 256), GRU(128, 128)	200	12	0.001
7	LSTM(512, 256, 256), GRU(128, 128)	300	12	0.001
8	LSTM(512, 256), GRU(128, 128)	300	6	0.001
9	LSTM(128), GRU(64)	300	6	0.001
10	LSTM(512), GRU(128, 128)	300	6	0.001

3.2 Présentation des résultats

Cette section présente les résultats obtenus des différentes configurations du modèle LSTM-GRU lors de l'entraînement et du test. Les performances des modèles sont essentielles pour évaluer leur capacité à prédire avec précision les précipitations dans la zone de Maga.

3.2.1 Résultats des métriques d'évaluation

Le tableau 3.3 récapitule les performances des différents modèles LSTM-GRU configurés pour la prévision des précipitations à partir de variables températures, humidité relative et les directions de vent. Chaque ligne représente une configuration spécifique du modèle, incluant le nombre d'unités pour chaque couche LSTM et GRU, le nombre d'époques d'entraînement, le batch size, le learning rate, le dropout, ainsi que les mesures de MAE, RMSE et R^2 pour évaluer la précision des prévisions.

3.2.2 Graphiques des fonctions de pertes et de validation

Le graphique 3.9 illustre l'évolution des pertes (loss) d'entraînement et de validation du modèle au fil des époques (epochs). La courbe bleue représente la perte d'entraînement, tandis que la courbe jaune indique la perte de validation. Un espace gris entre les deux courbes met en évidence l'écart (Gap) entre les pertes d'entraînement et de validation. Les valeurs de perte varient

TABLE 3.3 – Résultats des différentes configurations des modèles

Modèle	MAE	R ²	RMSE
1	17.572	0.8287	21.478
2	18.328	0.78952	23.8113
3	16.712	0.8392	20.809
4	20.659	0.76738	25.0326
5	11.974	0.90517	15.982
6	16.607	0.848	20.177
7	18.110	0.8178	22.153
8	12.497	0.8965	16.9964
9	14.099	0.87745	18.1689
10	12.479	0.888	17.344

de 0 à environ 8000, montrant une tendance générale à la baisse au fur et à mesure que le nombre d'époques augmente.

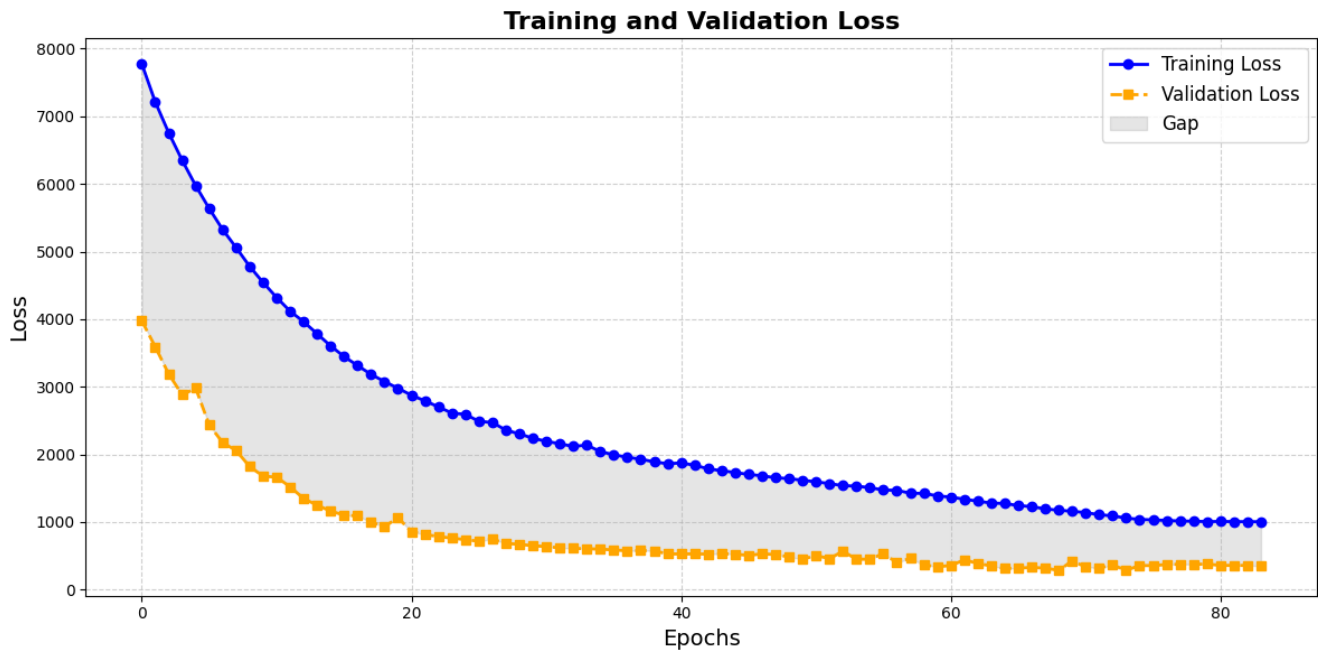


FIGURE 3.9 – Graphique des historiques des fonctions de perte et validation

3.2.3 Prévisions sur les données de Test

Le graphique 3.10 illustre les précipitations réelles et prédites sur l'année 2023, avec la courbe bleue représentant les valeurs réelles et la courbe orange les prévisions du modèle. Les précipitations sont mesurées en millimètres pour chaque mois. Le modèle suit globalement la tendance annuelle des précipitations, avec des pics observés pendant la saison des pluies et des niveaux presque nuls pendant les mois secs.

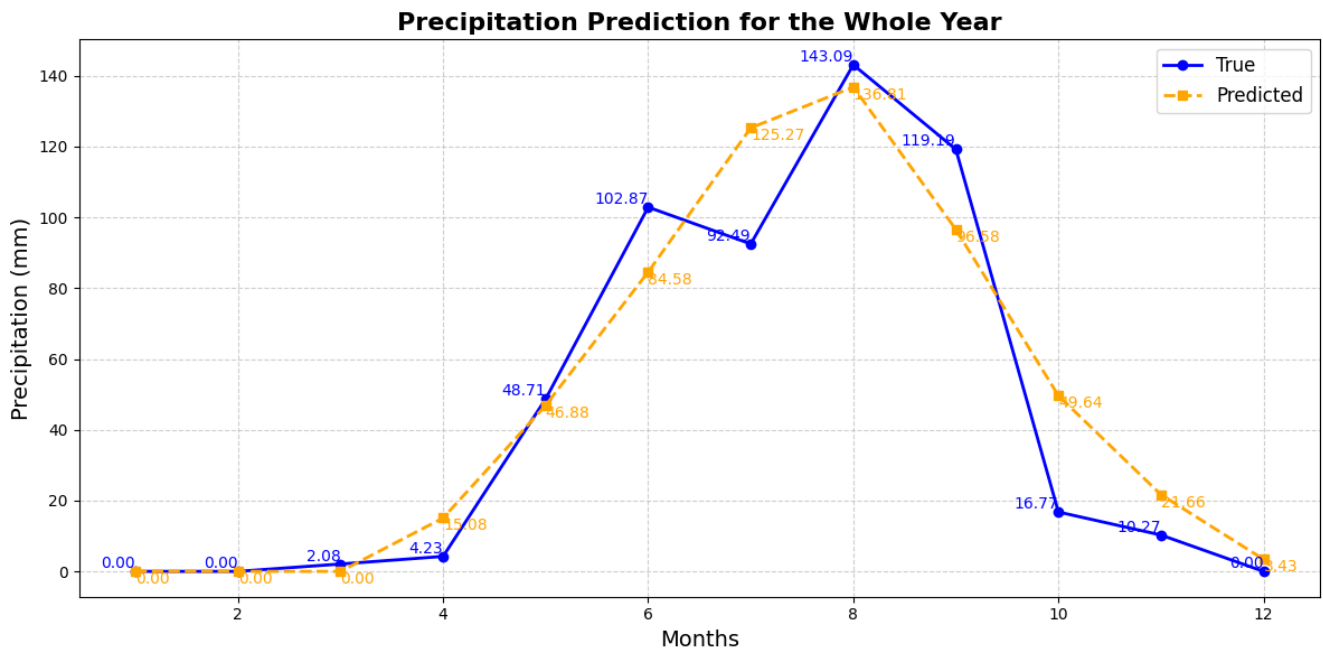


FIGURE 3.10 – Résultats des prévisions sur les données Test 2023

Le tableau 3.4 présente une comparaison entre les précipitations observées (valeurs réelles) et celles prédites par le modèle pour chaque mois de l’année. La dernière ligne montre la différence entre les valeurs réelles et prédites, indiquant les écarts pour chaque mois.

TABLE 3.4 – Tableau comparatif des valeurs réelles, prédites et des écarts pour chaque mois de l’année 2023.

Mois	Jan	Fév	Mar	Avr	Mai	Juin	Juil	Août	Sept	Oct	Nov	Déc
Réel	0.00	0.00	2.08	4.23	48.71	102.87	92.49	143.09	119.19	16.77	10.27	0.00
Prédit	0.00	0.00	0.00	15.08	46.88	84.58	125.27	136.81	96.58	49.64	21.66	3.43
Écart	0.00	0.00	2.08	10.85	1.83	18.29	32.78	6.28	22.61	32.87	11.39	3.43

3.2.4 Graphiques des données historiques et celles prédites

Le graphique 3.11 présente l’évolution des précipitations de 1982 à 2023. Il illustre à la fois les valeurs observées et prédites de l’année 2023. Nous observons que la courbe des précipitations prédites pour l’année 2023, obtenue à l’aide du meilleur modèle issu de notre expérimentation, suit une tendance similaire à celle des valeurs réelles de la même année. Cela suggère que notre modèle est capable de capturer efficacement les variations saisonnières et inter-annuelles des précipitations.

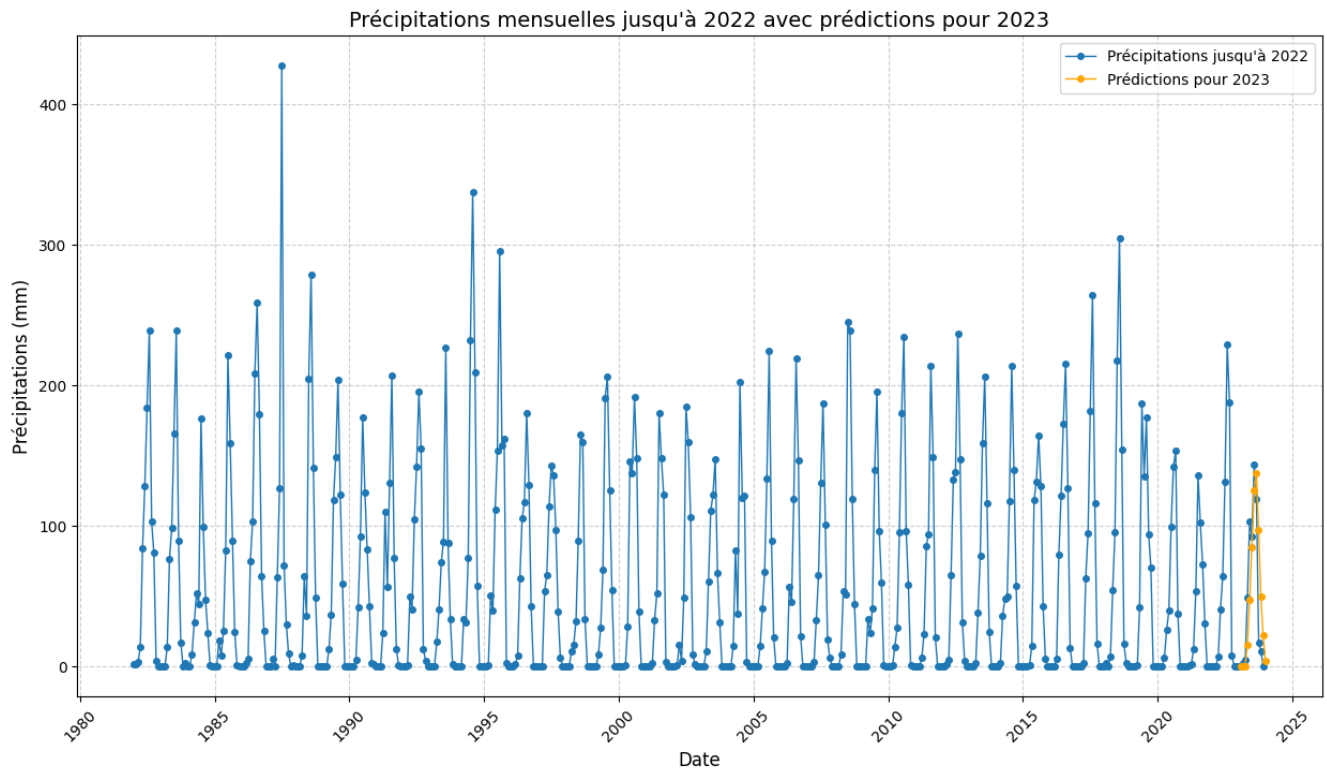


FIGURE 3.11 – Précipitations de 1982 à 2023

3.3 Interprétation des résultats

3.3.1 Tableau des métriques

L'analyse des résultats indique que la configuration la plus performante est celle du modèle numéro **5**, avec une architecture **LSTM (512, 256)** et **GRU(128, 128)**. Ce modèle a montré un **MAE** de **11.974** et un **R²** de **0.905** et un **RMSE** de **15.982**, montrant une capacité du modèle à prédire avec une précision acceptable les précipitations dans la zone de Maga. Les variations du learning rate et du batch size semblent influencer significativement les performances, soulignant l'importance de l'optimisation des paramètres pour améliorer les prévisions.

3.3.2 Graphiques des fonctions de perte et de validation

Les courbes du graphique 3.9 démontrent une diminution significative des pertes, ce qui indique que le modèle apprend efficacement. Au départ, l'écart entre les pertes d'entraînement et de validation est large, mais il se réduit au fil du temps, suggérant une bonne généralisation du modèle. Vers la fin de l'entraînement, la courbe de perte de validation se stabilise, indiquant que le modèle a atteint une performance optimale. Cela montre qu'il est peu probable que des améliorations significatives soient obtenues en poursuivant l'entraînement, ce qui est un signe positif pour la capacité du modèle à prédire sur des données nouvelles.

3.3.3 Prévisions sur les données de test

Les résultats des précipitations prédites pour l'année 2023, obtenus grâce au modèle hybride LSTM-GRU, montrent une comparaison entre les valeurs réelles et prédites pour chaque mois. Les valeurs réelles varient considérablement, avec des pics notables en mai et juin, tandis que les valeurs prédites suivent une tendance similaire. La figure 3.10 illustre cette comparaison, montrant que le modèle capture efficacement les pics durant les mois pluvieux et les baisses pendant les mois secs. Il reflète ainsi les variations saisonnières, permettant d'identifier les périodes de haute et basse précipitation et facilitant la planification saisonnière.

Le tableau 3.4 présente les écarts entre les valeurs réelles et prédites des précipitations pour chaque mois de l'année. On observe que ces écarts sont généralement faibles, ce qui montre une performance assez acceptable du modèle dans la plupart des mois.

En somme, le modèle est généralement en phase avec les observations, permettant d'anticiper les fluctuations saisonnières et d'optimiser la planification en fonction des variations climatiques attendues.

3.4 Discussion

Dans cette étude, les résultats obtenus avec le modèle démontrent une promesse significative pour la prévision des précipitations dans des conditions climatiques complexes, telles que celles de la zone sahélienne. Les performances supérieures du modèle suggèrent que l'utilisation d'une combinaison LSTM-GRU bien configurée, adaptée aux particularités temporelles des données météorologiques, peut améliorer de manière substantielle la précision des prévisions.

De plus, les configurations des hyperparamètres ont été identifiées comme un facteur crucial influençant les performances des modèles. En particulier, les modèles dotés de taux d'apprentissage raisonnables, d'une taille de lot moyenne et d'une distribution équilibrée de neurones LSTM et GRU montrent une convergence plus stable et des performances de validation améliorées.

Les résultats du tableau 1.1 montrent que les mois de juin, juillet et septembre présentent des écarts significatifs entre les valeurs réelles et prédites. Cela suggère que le modèle peine à capturer les variations saisonnières, ainsi que les événements de précipitations extrêmes, dans un contexte de changements climatiques au fil des ans.

Pour optimiser les prévisions dans un contexte climatique en constante évolution, il est essentiel d'apporter des améliorations. Des ajustements supplémentaires et l'exploration de techniques avancées pourraient renforcer la robustesse du modèle face à ces défis.

3.4.1 Prévision des quantités des précipitations 2024

Le graphique 3.12 présente les quantités des précipitations prédites par notre modèle pour l'année 2024. On observe que les précipitations augmentent progressivement à partir de janvier, atteignent

un pic en juillet, puis diminuent de manière significative jusqu'en décembre.

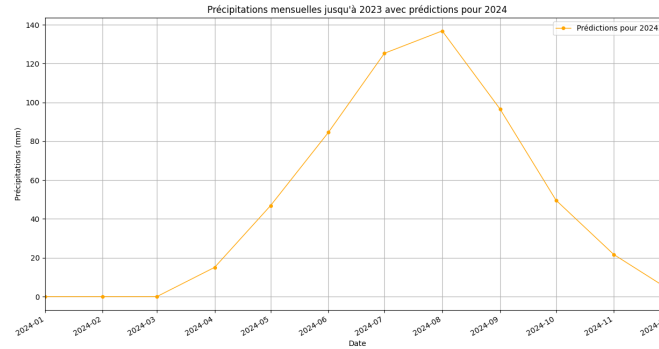


FIGURE 3.12 – Précipitations 2024

La forme du graphique 3.12 montre que notre modèle parvient à prédire de façon approximative les précipitations de l'année 2024. Les mois de juillet et août semblent être les plus pluvieux, ce qui correspond clairement à la saison de pluie. En revanche, les mois de janvier, février, et décembre présentent les précipitations les plus faibles, montrant une saison sèche pendant ces périodes.

3.4.2 Comparaison de nos résultats avec les résultats de l'article de base

Comparé à l'étude de Surta et al. (2023) [5] sur la prévision des précipitations à Palembang, notre modèle LSTM-GRU présente plusieurs avantages. Tandis que l'étude précédente montre que le LSTM seul offre une meilleure précision de prévision par rapport au GRU, notre modèle hybride obtient des métriques de performance globalement meilleures avec un R^2 de 0.905, contre 0.70 pour le LSTM de Surta et al. Cela démontre la capacité de notre modèle à mieux reproduire les données réelles.

Cependant, notre modèle présente également certaines limites. Il reste sensible aux valeurs extrêmes, ce qui affecte la précision des prévisions, notamment pendant les périodes où ces valeurs sont observées et lors d'événements météorologiques particuliers. Cette sensibilité a conduit à des erreurs de prévision significatives dans ces situations.

3.5 Conclusion

En résumé, ce chapitre a présenté le protocole expérimental pour l'évaluation des modèles LSTM-GRU. Nous avons décrit l'environnement de travail, le jeu de données, et les relations entre variables météorologiques et précipitations. Le prétraitement des données, incluant la gestion des valeurs manquantes et la normalisation, a été détaillé. Les configurations des modèles ont été spécifiées, et les résultats obtenus ont été analysés à l'aide des métriques d'évaluation, des graphiques de perte, et des prévisions sur les données de test. La comparaison des données historiques et prédites a permis d'évaluer la performances du modèle.

CONCLUSION ET PERSPECTIVES

Ce mémoire a exploré l'application des réseaux neuronaux récurrents (RNN), notamment les architectures LSTM et GRU, pour la prévision des précipitations dans la région de Maga. Face à la variabilité saisonnière et aux conditions météorologiques changeantes, ces modèles avancés se révèlent essentiels pour obtenir des prévisions fiables. Nous avons examiné diverses approches, passant des modèles numériques de prévision du temps (PNT) aux techniques de machine learning et deep learning, en intégrant les méthodes statistiques. Bien que ces approches aient montré des performances intéressantes, elles sont souvent spécifiques à des zones climatiques différentes de la zone soudano-sahélienne.

Nous avons d'abord présenté la formulation mathématique du problème, puis exploré brièvement les architectures LSTM et GRU pour souligner leur importance dans notre approche. Nous avons ensuite décrit notre modèle hybride combinant ces deux architectures de réseaux neuronaux récurrents, en détaillant l'algorithme de prévision des précipitations ainsi que l'analyse de la complexité de l'algorithme.

Après avoir présenté notre jeu de données de la zone de Maga téléchargé sur le site de la NASA et les relations entre les variables météorologiques et les précipitations, nous avons décrit les différentes étapes de prétraitement nécessaires à la préparation des données pour les expérimentations. Suite à plusieurs expérimentations et variations des hyperparamètres, la configuration 5, avec une architecture **LSTM (512, 256) et GRU (128, 128)**, a démontré de bonnes performances avec un **MAE** de **11.974**, un **R²** de **0.905** et un **RMSE** de **15.982**. Ces résultats attestent de la capacité du modèle à capturer efficacement les variations saisonnières des précipitations.

Les prévisions obtenues avec le modèle optimal ont montré une bonne concordance avec les valeurs réelles, illustrant ainsi l'efficacité des approches utilisées. Pour l'avenir, il serait pertinent d'intégrer de nouvelles variables météorologiques, d'optimiser les paramètres et d'explorer des techniques de machine learning complémentaires afin d'améliorer la qualité des résultats, qui sont fortement influencés par les changements climatiques au fil des années. De plus, le développement

de modèles pour des prévisions journalières pourrait affiner encore davantage les prévisions et mieux répondre aux besoins locaux en matière de gestion des ressources hydriques.

BIBLIOGRAPHIE

- [1] O Leumbe, Dieudonné Bitom, Lionnelle Mamdem, Denis Tiki, and Achille Ibrahim. Cartographie des zones à risques d'inondation en zone soudano-sahélienne : cas de maga et ses environs dans la région de l'extrême-nord cameroun. *Afrique Science : Revue Internationale des Sciences et Technologie*, 11(3) :45–61, 2015.
- [2] G Liénou, F Delclaux, and P Noupa. La plaine du lac tchad dans l'extrême-nord cameroun : de la sécheresse et des processus d'adaptation des populations à la restauration des inondations, 2011.
- [3] Frédéric Saha, Mesmin Tchindjang, Jean-Guy Dzana, and Djasrabé Nguemadjita. Risques naturels dans la région de l'extrême-nord du cameroun et dynamique des extrêmes hydrologiques du système chari-logone. *Physio-Géo. Géographie physique et environnement*, (Volume 15) :69–88, 2020.
- [4] S Aswin, P Geetha, and R Vinayakumar. Deep learning models for the prediction of rainfall. In *2018 International Conference on Communication and Signal Processing (ICCSP)*, pages 0657–0661. IEEE, 2018.
- [5] Wijaya Surta, Kurniawan Tri Basuki, Negara Edi Surya, and Kunang Yesi Novaria. Rainfall prediction in palembang city using the gru and lstm methods. *Journal of Data Science*, 2023(04) :1–13, 2023.
- [6] Auteur(s) non spécifié(s). Mapping of flood risk areas in sudano-sahelian zone : Maga and surrounding area case in the far north region of cameroon. *African Journal of Science, Technology, Innovation and Development*, année non spécifiée. Consulté le 29 mai 2025.
- [7] D Tiki. Etude morphologique, physico-chimique et cartographique des sols de maga (extrême-nord cameroun). *Mém. Master en Géosciences et Environnement. Univ. Ngaoundéré*, page 71, 2014.

- [8] Aniss Zebiri. *Estimation de la dynamique à partir de l'assimilation d'images radars multi-échelles : application à la prévision des taux de précipitation à courte échéance, à partir des images radars*. PhD thesis, Sorbonne université, 2020.
- [9] Météo contact - précipitations. <https://www.meteocontact.fr>, 2024. Consulté le 10 Octobre 2024.
- [10] Daniel Eilstein, Alain Le Tertre, Abdelkrim Zeghnoun, Sylvie Cassadou, Laurent Filleul, Laurence Pascal, Hélène Prouvost, Christophe Declercq, Philippe Saviuc, Agnès Lefranc, et al. Séries temporelles et modèles de régression. *Rapport technique, Institut de veille sanitaire. Département santé environnement*, 2003.
- [11] Zhaoxia Pu and Eugenia Kalnay. Numerical weather prediction basics : Models, numerical methods, and data assimilation. *Handbook of hydrometeorological ensemble forecasting*, pages 67–97, 2019.
- [12] François Forget, Frédéric Hourdin, Richard Fournier, Christophe Hourdin, Olivier Talagrand, Matthew Collins, Stephen R Lewis, Peter L Read, and Jean-Paul Huot. Improved general circulation models of the martian atmosphere from the surface to above 80 km. *Journal of Geophysical Research : Planets*, 104(E10) :24155–24175, 1999.
- [13] Khalida Benmahdjoub. *Prévision des précipitations par réseaux de neurones*. PhD thesis, Université Mouloud Mammeri, 2011.
- [14] Claude Fischer, François Bouyssel, Pierre Brousseau, Ryad El Khatib, Patricia Pottier, Yann Seity, Éric Wattrelot, and Alain Joly. Les modèles opérationnels de prévision numérique à aire limitée de météo-france. *La Météorologie*, 2018(100) :18–28, 2018.
- [15] Arthur Charpentier. *Modèles de prévision séries temporelles*. Paris, France, 2012.
- [16] Chaouchkhoulane Meriem. *Méthode de box et jenkins*.
- [17] MD Mahsin, Yesmin Akhter, and Monira Begum. Modeling rainfall in dhaka division of bangladesh using time series analysis. *Journal of mathematical modelling and application*, 1(5) :67–73, 2012.
- [18] Joseph Armathé, Isidore Sé, Patrick Forghab Mbomba, Romain Armand Soleil Batha, Paul Ghislain Poum Bimbar, et al. Cameroon climate predictions using the sarima-lstm machine learning model : Adjustment of a climate model for the sudano-sahelian zone of cameroon. *Open Journal of Statistics*, 14(3) :394–411, 2024.
- [19] Belkacem Balah. Frontières d'utilisation de la prévision des pluies mensuelles avec le modèle arima en agriculture. In *1er Séminaire International Agriculture 4.0 : Le Génie Rural au service de l'Environnement*, 2018.

- [20] Claude Touzet. *les réseaux de neurones artificiels, introduction au connexionnisme*. Ec2, 1992.
- [21] Deniz Mogulkoc. The neural network club : Deep learning. <https://denizmogulkoc.medium.com/the-neural-network-club-deep-learning-cabe7013b691>, 2025. Consulté le 29 mai 2025.
- [22] Duong Tran Anh, Thanh Duc Dang, and Song Pham Van. Improved rainfall prediction using combined pre-processing methods and feed-forward neural networks. *J*, 2(1) :65–83, 2019.
- [23] Neelam Mishra, Hemant Kumar Soni, Sanjiv Sharma, and AK Upadhyay. Development and analysis of artificial neural network models for rainfall prediction by using time-series data. *International Journal of Intelligent Systems and Applications*, 10(1) :16, 2018.
- [24] Juncheng Zhu, Zhile Yang, Monjur Mourshed, Yuanjun Guo, Yimin Zhou, Yan Chang, Yanjie Wei, and Shengzhong Feng. Electric vehicle charging load forecasting : A comparative study of deep learning approaches. *Energies*, 12(14) :2692, 2019.
- [25] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv :1412.3555*, 2014.
- [26] Manoj Chhetri, Sudhanshu Kumar, Partha Pratim Roy, and Byung-Gyu Kim. Deep blstm-gru model for monthly rainfall prediction : A case study of simtokha, bhutan. *Remote sensing*, 12(19) :3174, 2020.
- [27] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.
- [28] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016.
- [29] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber, et al. Gradient flow in recurrent nets : the difficulty of learning long-term dependencies, 2001.
- [30] Weicong Kong, Zhao Yang Dong, Youwei Jia, David J Hill, Yan Xu, and Yuan Zhang. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE transactions on smart grid*, 10(1) :841–851, 2017.
- [31] Xiaolei Liu, Zi Lin, and Ziming Feng. Short-term offshore wind speed forecast by seasonal arima-a comparison against gru and lstm. *Energy*, 227 :120492, 2021.
- [32] The structure of the lstm network model. https://www.researchgate.net/figure/The-structure-of-the-LSTM-network-model_fig3_374408883, 2024. Consulté le 10 Octobre 2024.

- [33] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation : Encoder-decoder approaches. *arXiv preprint arXiv :1409.1259*, 2014.
- [34] Rui Fu, Zuo Zhang, and Li Li. Using lstm and gru neural network methods for traffic flow prediction. In *2016 31st Youth academic annual conference of Chinese association of automation (YAC)*, pages 324–328. IEEE, 2016.
- [35] Architectural elements in a tcn : a dilated causal convolution with dilation factors. https://www.researchgate.net/figure/Architectural-elements-in-a-TCN-a-A-dilated-causal-convolution-with-dilation\protect\penalty\z@-factors-d_fig3_373294748, 2024. Consulté le 10 Octobre 2024.