

## Comparaison des modèles LSTM et Transformée en Ondelettes-LSTM pour la prédiction de la température dans une partie du Bassin du Congo

### [ Comparison of LSTM and Wavelet Transform-LSTM Models for Temperature Prediction in a part of Congo Basin ]

Djes-Fresy Bilenga Moukodouma<sup>1,2</sup>, Christophe Denis<sup>2,3</sup>, Donald Romarick Rotimbo Mbourou<sup>1</sup>, and Christiane Atteke Nkoulembene<sup>4</sup>

<sup>1</sup>Laboratoire de Recherche Multidisciplinaire en Environnement (LARME), Université des Sciences et Techniques de Masuku (USTM), bâtiment du département de Chimie, Mbaya, route nationale 1, BP 901 Franceville, Gabon

<sup>2</sup>Unité de Modélisation Mathématique et Informatique des Systèmes Complexes (UMMISCO), Sorbonne Université, Institut de Recherche pour le Développement, 32 avenue Henri Varagnat, 93140 Bondy Cedex, France

<sup>3</sup>Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST), Panthéon Sorbonne, 13 rue du Four, 75006 Paris, France

<sup>4</sup>Département de Biologie, Université des Sciences et Techniques de Masuku, Mbaya, route nationale 1, BP 901 Franceville, Gabon

---

Copyright © 2024 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT:** Currently, the Congo Basin represents the most important center in terms of biodiversity concentration, especially with the increasing deforestation observed in the Amazon. The available climate models are mostly at larger scales, and few of them focus on specific areas of the Congo Basin, such as the locality of Makokou in Gabon. A new approach is therefore needed to predict temperature changes in this particular region. Although some work focus on temperature prediction, most do not use deep learning algorithms. This contribution aims to compare the predictions of a Long Short-Term Memory (LSTM) model with those from the combination of Wavelet Transform and LSTM (WT-LSTM). The developed LSTM model includes two LSTM layers, two Dropout layers (with a rate of 50 %) and a Dense layer to output the predicted value. The WT-LSTM model shows superior results compared to the LSTM model, with a root mean square error of 0.45 °C, a mean absolute error of 0.35 °C, and a Spearman correlation coefficient of 0.97 °C. These results highlight the importance of using advanced approaches to improve climate forecasts in areas crucial for biodiversity conservation. The increased accuracy of predictions could help better anticipate and mitigate the impacts of local climate change, thereby contributing to the sustainable management of this ecologically sensitive region.

**KEYWORDS:** climate change, deep learning, Long Short-Term Memory model (LSTM), Wavelet Transform, temperature prediction, biodiversity, Congo Basin.

**RESUME:** Actuellement, le Bassin du Congo représente le centre le plus important en termes de concentration en biodiversité, surtout avec la déforestation croissante observée en Amazonie. Les modèles climatiques disponibles sont majoritairement à des échelles plus grandes et peu d'entre eux se concentrent sur des zones spécifiques du Bassin du Congo, comme la localité de Makokou au Gabon. Une nouvelle approche est donc nécessaire pour prédire les changements de température dans cette région particulière. Bien qu'il existe quelques travaux portant sur la prédiction des températures, la majorité n'utilisent pas les algorithmes d'apprentissage profond. Cette contribution vise à comparer les prédictions d'un modèle de mémoire à court et long terme (LSTM) avec celles issues de la combinaison Transformée en Ondelettes et LSTM (TO-LSTM). Le modèle LSTM développé comprend deux couches LSTM, deux couches Dropout (à un taux de 50%) et une couche Dense pour afficher la valeur prédite. Le modèle TO-LSTM présente des résultats supérieurs à ceux du modèle LSTM, avec une racine carrée d'erreur quadratique moyenne de 0,45 °C, une erreur absolue moyenne de 0,35 °C et un coefficient de corrélation de Spearman de 0,97 °C. Ces résultats soulignent l'importance d'utiliser des approches avancées pour améliorer les prévisions climatiques dans des zones cruciales pour la conservation de la biodiversité. La précision accrue des prévisions pourrait aider

à mieux anticiper et atténuer les impacts des changements climatiques locaux, contribuant ainsi à la gestion durable de cette région écologiquement sensible.

**MOTS-CLEFS:** changements climatiques, apprentissage profond, modèle de mémoire à court et long terme LSTM, Transformée en Ondelettes, prédiction des températures, biodiversité, Bassin du Congo.

## 1 INTRODUCTION

Défi majeur du 21<sup>ème</sup> siècle, les changements climatiques constituent un véritable danger (Belle et al. 2016) pour la flore, la faune et le quotidien des populations du Bassin du Congo (BC) (Dellink et al. 2017; Adamo et al. 2018; Grooten and Almond 2018). En pratique, les variations climatiques inhibent la reproduction de certaines espèces floristiques (Chakanyuka 2019; Bush et al. 2020). Les animaux se déplacent de plus en plus à la recherche de points d'eau et de ressources alimentaires (Moukodouma et al. 2023). La survie des populations est menacée par des inondations fréquentes, des éboulements de terrain, et parfois même des séismes (Adamo et al. 2018; Chirwa and Adeyemi 2020; Moukodouma et al. 2023). Aux impacts précédents, s'ajoute la sécheresse des sols de plus en plus présente dans la région (Chirwa and Adeyemi 2020). N'ayant plus de terres cultivables, de nombreux habitants du Bassin du Congo meurent progressivement de faim (Chirwa and Adeyemi 2020; Toto 2023).

Les répercussions des changements climatiques au niveau de chaque espèce (végétale, animale et humaine) sont davantage visibles. Les variations de températures sont un critère clé des changements climatiques. L'intérêt de prédire l'évolution des températures provient du fait que celles-ci jouent un rôle fondamental pour tous les êtres vivants. Par exemple, certains arbres de la forêt ont besoin d'une température minimale pour déclencher leur floraison (Bush 2018; Ren et al. 2021). Le corps humain a aussi en réalité besoin d'une certaine température pour maintenir le bon fonctionnement de l'organisme (Vecellio et al. 2022; Vanos et al. 2023). Ces différentes observations montrent l'utilité de s'intéresser aux variations des températures de nos jours (Lee et al. 2023).

Dans ce sens, plusieurs modèles basés sur des approches statistiques et l'apprentissage machine sont de plus en plus utilisés pour prédire l'évolution des températures (Piccolroaz et al. 2016; Cifuentes et al. 2020). Les modèles statistiques sont couramment employés pour extraire des motifs à partir des données et pour prédire des nouvelles observations (Piccolroaz et al. 2016). Dans une étude menée à trois points distincts du fleuve Ouémé au Bénin, les prévisions de la pluviométrie, de la température moyenne et de l'évapotranspiration ont été rendues possibles grâce à un modèle ARIMA (AutoRegressive Integrated Moving Average) (Nounangnonhou and Fifatin 2016). Au cours d'une autre étude, les données d'une période future ont été prédites grâce aux données météorologiques antérieures de la province d'Aceh Besar sur l'île de Sumatra, à partir du modèle ETS (Error, Trend, Seasonal) de prévision univariée de séries temporelles (Jofipasi et al. 2018).

En plus des modèles purement statistiques, l'amélioration des performances prédictives de l'apprentissage machine a favorisé l'émergence de nombreux autres modèles (Guillaume 2019; Zameer et al. 2023). A cet effet, dans une étude, l'algorithme supports vecteurs machine (SVM) s'est avéré être meilleur dans la prédiction de la température globale terre-océan par rapport aux réseaux neuronaux artificiels utilisés (Abubakar et al. 2016). L'algorithme SVM (Support Vector Machine) a été développé par l'auteur Vapnik (Vapnik 2013) et est utilisé aussi bien pour les problèmes de classifications (la variable cible est une valeur qualitative), que de régressions (la variable cible est une variable quantitative).

Dans un autre contexte, les réseaux neuronaux artificiels (Artificial Neural Network ou ANN) ont été exploités pour la prédiction de la température horaire de l'air (Li et al. 2020; Haque et al. 2021; Gong et al. 2022). Et il ressort de ces études qu'à l'échelle régionale, les modèles d'apprentissage profond apportent des prévisions bien plus précises que celles de l'apprentissage machine (Abubakar et al. 2016).

Toutes ces méthodes présentent des résultats tout de même acceptables mais elles possèdent un inconvénient majeur. Les modèles statistiques par exemple calculent la probabilité qu'un phénomène météorologique se produise (Moazenzadeh et al. 2022). Cependant, le mécanisme et les facteurs affectant les changements de température de l'air sont très complexes et non linéaires, ce qui accroît la difficulté à saisir les changements dynamiques de températures lors de la prédiction de longues séries temporelles (Hou et al. 2022). Les réseaux neuronaux artificiels sont un outil puissant qui n'arrive pas toujours à garder en mémoire la dynamique lorsque la série temporelle est très importante (Bharadiya 2023).

Au regard de cet inconvénient majeur qui caractérise chacune des méthodes décrites précédemment, il apparaît utile de migrer vers l'approche des réseaux de neurones récurrents avec la mémoire à court et long terme (LSTM). Cette approche semble avoir fait déjà des preuves remarquables dans le traitement de longues séries temporelles (Yadav et al. 2020; Karevan and Suykens 2020; Hu et al. 2020).

Le problème lié à l'étude des variations de températures touche de nombreux pays tels que ceux du Bassin du Congo (BC), dont le Gabon fait partie. L'approche LSTM pourrait être une véritable option, utile pour prédire l'évolution des températures dans cette région. En effet, la LSTM a le mérite de mieux s'adapter aux données séquentielles (Xia et al. 2020; Kang et al. 2020). Dans une étude menée à

la station météorologique de Yinchuan (Chine), un modèle basé sur l'algorithme bidirectionnel LSTM a permis de prédire avec une erreur absolue moyenne de 1,02 °C, la température horaire de l'air (Hou et al. 2022). D'autres travaux ont mis en évidence un modèle LSTM pour prédire à la fois les niveaux de température et d'humidité dans les bâtiments (Wang et al. 2021). Une récente étude menée en Australie a révélé une p-value moyenne égale à 0,9994 pour le test de khi-2, prouvant ainsi que le modèle LSTM conçu dans ce cas a prédit avec une très grande précision la température future en Australie pour une année (Qiu 2023). En Amazonie, un réseau LSTM a été proposé pour prévoir les températures minimales, moyennes et maximales jusqu'en 2030 dans 20 grandes villes traversées par la forêt (Dominguez et al. 2023). Dans une autre étude, toujours en Amazonie, une architecture de réseau neuronal a été développée pour apprendre à détecter la déforestation de bout en bout à partir de séries temporelles d'images de radar (Karaman et al. 2023). Pour ce qui est de notre région cible, des études de climat basées sur les méthodes de l'Intelligence Artificielle (IA) sont quasi-inexistantes. Une revue fait état de certains chiffres concernant les publications avec l'IA en Afrique. A ce jour, sur 2468 articles publiés en IA en Afrique, le Cameroun a un pourcentage de représentativité de 0,85% puis on note 0,45% pour la République Démocratique du Congo (RDC), et 0,16% pour le Gabon (Ezugwu et al. 2023). Néanmoins au Maroc, une étude basée sur l'algorithme LSTM a permis de prédire le rayonnement solaire (Boutahir et al. 2022). Au Kenya, pays dont les caractéristiques sont très proches de celles de notre région d'étude, un modèle de réseau neuronal récurrent LSTM a permis avec une bonne précision de prévoir les totaux mensuels des précipitations (Beunk 2021).

Contrairement à l'Amazonie où règne un nombre important d'études sur les variations des températures, la question de la modélisation des températures demeure très peu étudiée dans la région du Bassin du Congo.

Ce travail compare deux modèles pour prédire les températures à Makokou, au Gabon. Le premier modèle utilise l'architecture LSTM seule, tandis que le second combine la Transformée en Ondelettes avec l'architecture LSTM. L'objectif est de montrer comment l'intégration de la tendance obtenue par Transformée en Ondelettes améliore la précision des prévisions de température par rapport au modèle LSTM seul.

La présente contribution s'avère nécessaire dans une région tant oubliée, mais qui demeure et pourtant le plus grand puits de carbone actuellement au monde (Atyi et al. 2022). Les gouvernants et acteurs des Conférences Des Parties (Conference Of Parties ou COP) et le Groupe Intergouvernemental des Experts sur l'étude du Climat (GIEC) souhaitent maintenir l'augmentation de la température globale mondiale à un seuil de 1,5°C (MAIDOU 2020; Beaudoin and Chaloux 2023; Torre-Schaub 2023). Cette étude pourrait ainsi aider les gouvernants du Bassin du Congo (BC) à mieux planifier leurs politiques environnementales en vue de répondre à cet objectif.

Notre contribution s'organise de la manière suivante: la section II Matériels et Méthodes présente la zone d'étude et la méthodologie suivie. La section III Résultats met en évidence l'ensemble des résultats majeurs obtenus de l'étude. La section IV Discussion quant à elle permet d'étayer les différents résultats à la lumière des références bibliographiques. La section V est la conclusion qui nous permet de faire un bilan de l'étude puis d'énoncer les travaux futurs qui peuvent être menés pour enrichir davantage l'actuel travail.

## 2 MATERIELS ET METHODES

### 2.1 ZONE D'ÉTUDE

L'étude se déroule dans la localité de Makokou, chef-lieu de la province de l'Ogooué Ivindo au nord-est du Gabon (Figure 1). Le Gabon occupe la plus grande partie de forêt du Bassin du Congo (Balada 2021). Nous retrouvons majoritairement cette forêt dans l'Ogooué Ivindo avec notamment la réserve de biosphère Ipassa-Makokou et le parc national de l'Ivindo (Roland et al. 2016). Ces deux réserves à elles seules regroupent une diversité d'espèces floristiques et fauniques (Roland et al. 2016; Laguardia et al. 2021). Parmi celles-ci, nous dénombrons environ 18 espèces de mammifères, 424 espèces d'oiseaux, 65 espèces de reptiles et 47 espèces d'amphibiens (Roland et al. 2016). A ces différentes espèces s'ajoute l'éléphant, l'ingénieur écologue de la forêt qui est présent en grand nombre dans la province de l'Ogooué Ivindo (Berzaghi et al. 2019; Laguardia et al. 2021; Kermabon 2022). A l'instar du pays, la localité de Makokou possède un climat de type équatorial caractérisé par l'alternance des saisons sèches et pluvieuses (Roland et al. 2016). La température moyenne de Makokou oscille autour de 24°C et sa pluviométrie moyenne annuelle varie entre 1600 mm et 1800 mm (Roland et al. 2016).

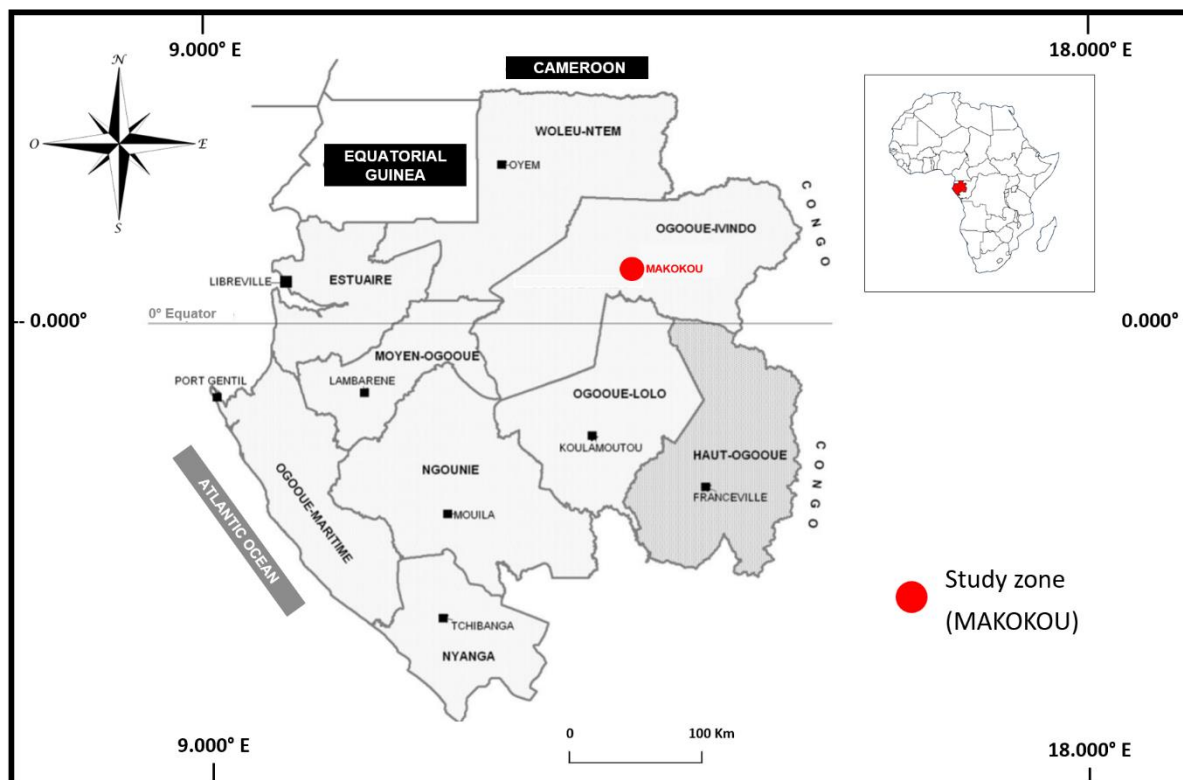


Fig. 1. Localisation de la zone d'étude

Ainsi, la localité de Makokou apparaît comme une zone très riche en concentration de biodiversité. Cette importante biodiversité nécessite donc qu'on s'intéresse à cette zone, au regard des rôles multiples qu'elle pourrait avoir tant pour la survie de l'homme que pour la faune et la flore. Celle-ci pourrait également être un véritable moteur en ce qui concerne la lutte contre les changements climatiques.

## 2.2 DESCRIPTION DE LA BASE DE DONNEES

Les données climatiques de notre étude sont celles de la localité de Makokou sur la période 2000 - 2020. Elles ont été collectées par la National Data Climatic Center (NCDC). La NCDC est une grande base mondiale des données météorologiques (Gad and Hosahalli 2022). En ce qui concerne notre zone d'étude, seule la période 2000 - 2020 était répertoriée avec tout de même un nombre important d'observations journalières (On dénombreait environ 4695 observations).

Comme l'illustre les tableaux 1a et 1b de nombreux paramètres climatiques sont mesurés dans cette station. La base de données contient la date du jour (YEARMODA), la température moyenne (TEMP) en degrés Fahrenheit, le point de rosée moyen en degrés Fahrenheit (DEWP), la pression moyenne au niveau de la mer en millibars (SLP), la pression moyenne de la station pour la journée en millibars (STP), la visibilité en mètres (distance jusqu'à laquelle les objets peuvent être clairement vus), la température maximale en degrés Fahrenheit (MAX), la température minimale en degrés Fahrenheit (MIN), les précipitations totales en pouces (PRCP) et l'ensemble des conditions météorologiques (FRSHTT). Particulièrement dans le sigle FRSHTT, nous retrouvons Fog (Brouillard), Rain (Pluie), Snow (Neige), Hail (Grêle), Thunder (Tonnerre) et Tornado (Tornado).

Tableau 1. Caractéristiques des 5 premières lignes pour les 5 premières variables [source: NCDC]

YEARMODA	TEMP	DEWP	SLP	STP	VISIB
01/01/2000	70.5	69.2	1006.9	950.2	6.2
02/01/2000	74.1	68.7	1006.9	949.9	4.7
03/01/2000	74.6	70.6	1006.2	949.2	6.5
04/01/2000	74.8	70.7	1005.8	949.4	7.4
05/01/2000	74.2	70.8	1005.2	948.8	5

Au cours de notre étude, la base de données disponible a été subdivisée en deux parties, à savoir une partie servant pour l'entraînement des futurs modèles (ce qui correspondait à 80 % des données), et les 20 % de données restantes ont servi à l'entraînement des modèles de réseaux de neurones mis au point dans cette étude.

## 2.3 MÉTHODES

### 2.3.1 TRANSFORMÉE EN ONDELETTES DISCRÈTE (TOD)

La Transformée en Ondelettes Discrète (TOD) est une version discrétisée de la Transformée en Ondelettes Continue (TOC), où le signal est analysé à des échelles et des positions spécifiques (Alessio 2016; Guo et al. 2018). Contrairement à la TOC, la TOD ne produit pas de redondance, ce qui la rend plus efficace pour le stockage et le traitement des données (Ponni alias Sathya and Ramakrishnan 2020; Chen et al. 2021).

Dans notre étude, nous avons opté pour les ondelettes Symlets d'ordre 4 (Sym4). En effet, contrairement aux ondelettes de Daubechies qui présentent un support non symétrique, les Symlets sont d'abord presque symétriques. Cette symétrie rend ces ondelettes plus adaptées à certaines applications telles que le traitement d'images et le débruitage où les effets de bords sont moins tolérables (Shahbazzabar et al. 2018; Isabona and Kehinde 2019; Arfaoui et al. 2021). Ensuite, les Symlets, tout comme les Daubechies, sont des ondelettes orthogonales (Guo et al. 2022; Daud and Sudirman 2022). Cette propriété permet une décomposition et une reconstruction sans perte d'information, ce qui est essentiel pour de nombreuses applications de traitement de signal (Kumar and Satyanarayana 2022). Enfin, les Symlets sont efficaces pour lisser un signal, c'est-à-dire pour éliminer les petites irrégularités tout en conservant les grandes tendances (Gossler et al. 2023).

Avant de passer les données de températures au modèle de réseau de neurones LSTM, nous avons appliqué une Transformée en Ondelettes Discrète sur la série temporelle de température moyenne de la localité de Makokou. Ensuite, nous avons extrait la dernière composante dans la décomposition en ondelettes. Cette dernière composante est la tendance, généralement observée dans les basses fréquences. La tendance extraite du signal a été sauvegardée dans la base de données pour être utile pour le développement des modèles.

### 2.3.2 DÉVELOPPEMENT DES MODÈLES DE RÉSEAUX DE NEURONES

#### 2.3.2.1 LONG SHORT-TERM MEMORY MODEL

La mémoire à long et court terme est un type de réseau de neurones récurrent, conçue pour surmonter certaines limitations des réseaux de neurones récurrents traditionnels. Les limitations que posent ces traditionnels réseaux de neurones récurrents sont liées à l'apprentissage de dépendances à long terme dans les séquences de données. La LSTM a en effet le mérite de mieux s'adapter aux données séquentielles (Xia et al. 2020; Bagastio et al. 2023).

Un réseau LSTM est constitué d'une série de cellules (ou blocs) LSTM. Chacune de ces cellules, est dotée d'une structure interne complexe qui permet de retenir, oublier ou modifier des informations en fonction de la séquence de données en entrée (Qiu 2023).

La figure 2 suivante présente l'architecture générale d'un réseau LSTM.

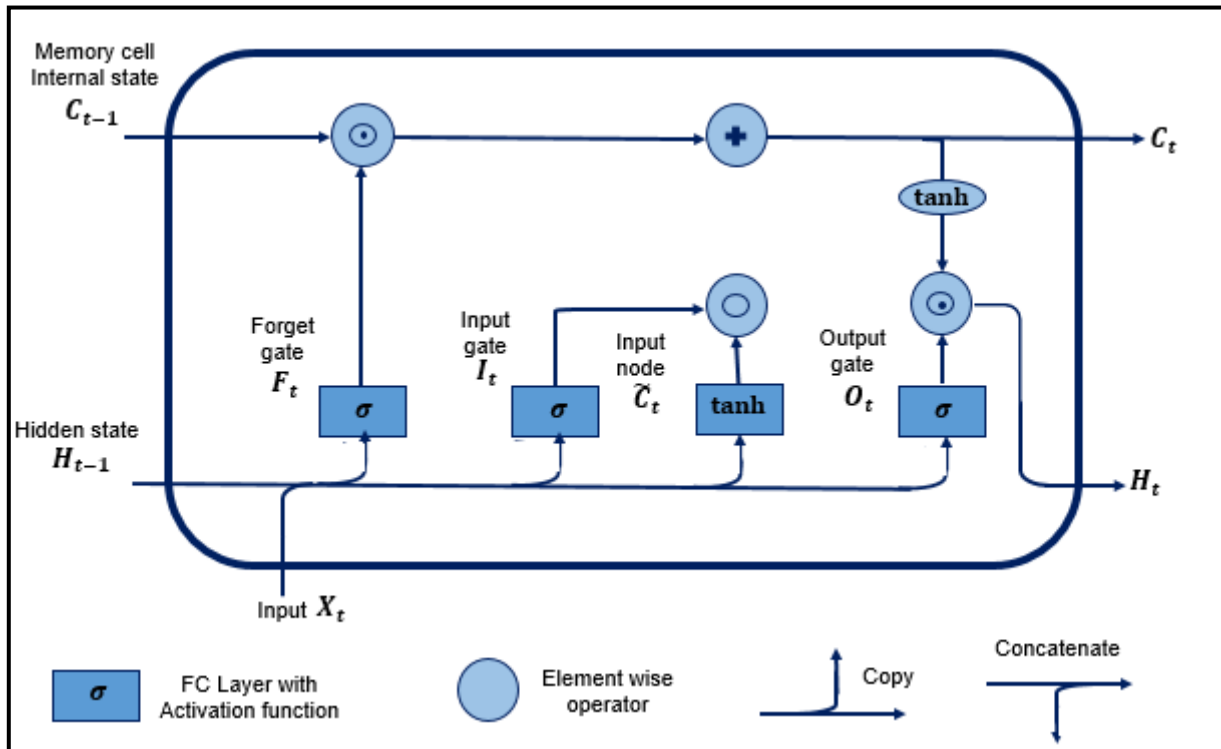


Fig. 2. Architecture of LSTM Network [source: internet]

Dans le contexte de notre étude, nous avons développé un modèle LSTM à cinq couches. Ledit modèle possédait deux couches LSTM, deux couches Dropout et une couche Dense. En pratique, les deux couches LSTM (ayant chacune 50 neurones) se comportent comme des blocs de mémoire qui peuvent mémoriser des informations à long terme. Pour éviter le surapprentissage du modèle, nous avons ajouté deux couches Dropout à un taux de 50 %, c'est-à-dire que 50 % de neurones ont été désactivés aléatoirement à chaque itération dans le modèle LSTM développé. Cette technique de régularisation avait pour but d'éviter que le réseau devienne trop dépendant de neurones spécifiques au cours de son entraînement.

Une fois le modèle construit, celui-ci a été compilé grâce à l'optimiseur Adam et la fonction de perte « Mean\_Squared\_Error ». La compilation est nécessaire car celle-ci configure le modèle pour l'apprentissage tout en précisant clairement comment il doit être entraîné.

Juste après la compilation du modèle, nous avons entraîné ce dernier à partir de la base de données d'entraînement (80 % des données).

Nous avons par la suite utilisé notre modèle pour réaliser des prédictions sur les données de test. Ensuite, nous avons inversé la normalisation afin de ramener les valeurs prédites à leur échelle d'origine. Enfin, nous avons comparé visuellement ces prédictions aux valeurs réelles, afin d'évaluer la correspondance entre elles.

La dernière étape de notre modélisation concerne l'interprétabilité. A ce stade, il s'agit d'évaluer la contribution de chaque variable explicative au modèle. Pour cela, nous avons utilisé les valeurs de Shapley, qui offrent une méthode cohérente et équitable pour quantifier l'impact de chaque variable explicative sur les prédictions du modèle (Denis and Varenne 2022).

### 2.3.2.2 CRITERES D'EVALUATION DES PERFORMANCES DES MODELES

#### 2.3.2.2.1 MÉTRIQUES TRADITIONNELLES

Les modèles ont principalement été évalués à partir de trois métriques traditionnelles, notamment l'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne (MAE), et l'erreur absolue moyenne en pourcentage (MAPE) (Boutahir et al. 2022; Ezugwu et al. 2023).

La racine carrée de l'erreur quadratique moyenne est définie par l'équation (1) ci-après:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - y_i)^2} \tag{1}$$

Dans cette formule,  $N$  représente le nombre d'observations,  $Y_i$  la valeur réelle et  $y_i$  la valeur prédite.

L'erreur absolue moyenne (MAE) définie à l'équation (2) est la moyenne de la valeur absolue de l'écart entre la prédiction  $y_i$  et la valeur réelle  $Y_i$  sur un échantillon de  $N$  observations

$$MAE = \frac{1}{N} \sum_{i=1}^N (|Y_i - y_i|) \quad (2)$$

Partant de l'équation (2), nous pouvons affirmer que plus le MAE est petit, meilleur est le modèle au regard des données.

L'erreur absolue moyenne en pourcentage (MAPE) est un indicateur statistique utilisé pour mesurer la précision d'un modèle de prévision. Il représente l'erreur moyenne entre les valeurs réelles et les valeurs prédites, exprimée en pourcentage. Plus précisément, le MAPE calcule la moyenne des erreurs absolues en pourcentage des valeurs réelles, ce qui permet de comprendre à quel point les prédictions d'un modèle s'éloignent ou se rapprochent des valeurs observées. La formule du MAPE est la suivante:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - y_i}{Y_i} \right| \times 100 \quad (3)$$

Où  $N$  représente le nombre d'observations,  $Y_i$  la valeur réelle et  $y_i$  la valeur prédite.

Si l'erreur absolue moyenne en pourcentage (MAPE) est faible, cela indique que le modèle est précis (les prédictions sont proches des valeurs réelles). Par contre si la MAPE est élevée, cela suggère qu'il y'a un écart considérable entre les prédictions du modèle et les valeurs réelles.

#### 2.3.2.2.2 TEST DE RANGS SIGNÉS DE WILCOXON

Le Test de rangs signés de Wilcoxon est utilisé pour comparer les échantillons appariés comme les valeurs réelles et les prédictions d'un modèle.

L'hypothèse nulle de ce test est que les médianes des différences entre les paires sont égales à zéro. L'hypothèse alternative stipule que les médianes des différences entre les paires ne sont pas égales à zéro.

Le Test de rangs signés de Wilcoxon effectue d'abord le calcul des différences pour chaque paire de valeurs appariées. Ensuite, il attribue des rangs absolus aux différences en les classant de la plus petite à la plus grande. Par la suite, les rangs des différences sont réattribués en fonction du signe original. Juste après, le Test effectue la somme des rangs pour les différences positives et pour les différences négatives. Enfin, on utilise la somme des rangs signés pour calculer la statistique du test, qui est comparée à une distribution de référence pour déterminer la significativité.

En ce qui concerne l'interprétation, si la statistique du test est grande ou petite et que la p-value est au-dessus du seuil de significativité (0.05), on ne rejette pas l'hypothèse nulle. Par contre, si les conditions précédentes ne sont pas vérifiées, l'hypothèse nulle est rejetée.

#### 2.3.2.2.3 COEFFICIENT DE CORRELATION DE RANGS DE SPEARMAN

Le coefficient de corrélation de rangs de Spearman est une mesure statistique qui évalue la force et la direction d'une relation monotone entre deux variables (Astivia and Zumbo 2017). Il repose sur trois principes clés à savoir la relation monotone, les rangs et sa formule (Al-Hameed and Khawla 2022).

Une relation est monotone si, à mesure que les valeurs d'une variable augmentent, celles de l'autre variable augmentent ou diminuent de manière cohérente, sans nécessairement suivre une ligne droite.

Plutôt que de comparer directement les valeurs brutes des variables, Spearman s'appuie sur les rangs. Les valeurs des variables sont d'abord classées, puis les différences entre ces rangs sont utilisées pour calculer la corrélation. Etant donné qu'il se base sur les rangs et non sur les valeurs brutes, il est moins sensible aux valeurs aberrantes.

Le coefficient de corrélation de rang de Spearman ( $\rho$ ) est calculé à partir de la différence des rangs des deux variables, selon la formule donnée à l'équation (4) (Eltehiwy and Abdul-Motaal 2023)

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (4)$$

Avec  $d_i$  la différence entre les rangs de chaque paire de variables et  $n$  est le nombre de paires de données.

Si  $\rho = 1$ , la relation est parfaitement monotone et croissante, et si  $\rho = -1$  la relation est parfaitement monotone et décroissante. Dans le cas où  $\rho$  est nulle, il n'y a aucune relation monotone entre les variables.

#### 2.3.2.2.4 TEST DE MANN-WHITNEY U

Afin de renforcer davantage les résultats des modèles, nous avons utilisé le test de Mann-Whitney U pour comparer les résidus médians des différents modèles afin de voir s'il existe une différence significative entre leurs distributions (Vengatesan et al. 2018; Vierra et al. 2023).

L'hypothèse nulle ( $H_0$ ) de ce test est qu'il n'y a pas de différence significative entre les deux échantillons. Autrement dit les échantillons ont des distributions similaires. Par contre l'hypothèse alternative ( $H_1$ ) stipule qu'il y'a une différence significative entre les distributions c'est-à-dire que l'un des modèles a tendance à produire des valeurs plus grandes ou plus petites que l'autre.

Le Test de Mann-Whitney U débute par le rang des données. En effet les données des deux échantillons sont d'abord combinées et ordonnées du plus petit au plus grand. Ensuite, des rangs sont assignés aux valeurs. Si deux ou plusieurs valeurs sont identiques, elles reçoivent la moyenne des rangs qu'elles occuperaient si elles étaient distinctes.

Après la détermination des rangs des données, s'ensuit le calcul des statistiques U. Pour chaque échantillon, le Test de Mann-Whitney U calcule un score U en fonction de la somme des rangs. Les formules de base sont données aux équations (5) et (6) (Wall Emerson 2023)

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \tag{5}$$

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2} \tag{6}$$

Avec  $R_1$  et  $R_2$  les sommes des rangs des échantillons 1 et 2 respectivement, et  $n_1, n_2$  les tailles des échantillons 1 et 2.

La troisième étape du Test est le choix de la statistique U. En pratique le plus petit des deux U est utilisé pour interpréter le test.

Enfin, la p-value associée est calculée à partir de la statistique U pour déterminer si la différence observée entre les échantillons est significative. Si la p-value est inférieure à un seuil (souvent fixé à 0.05), on rejette l'hypothèse nulle en faveur de l'hypothèse alternative.

### 3 RESULTATS

#### 3.1 EXPLORATION DES DONNÉES

L'exploration de la base de données révèle que les températures moyennes journalières dans la localité de Makokou oscillent entre 20 °C et 32 °C au cours de la période d'étude (2000-2020) (Figure 3).

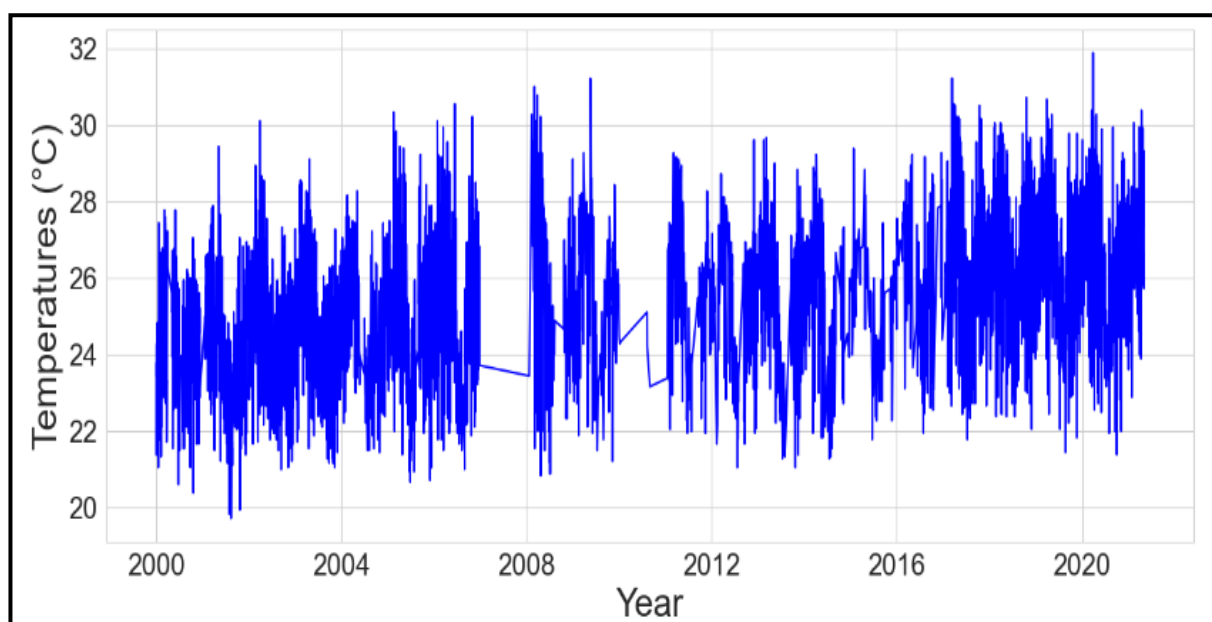


Fig. 3. Evolution of Temperature Time Serie



Comme l'illustre la figure 3, la tendance extraite après application de la Transformée en Ondelettes présente une évolution similaire aux données brutes.

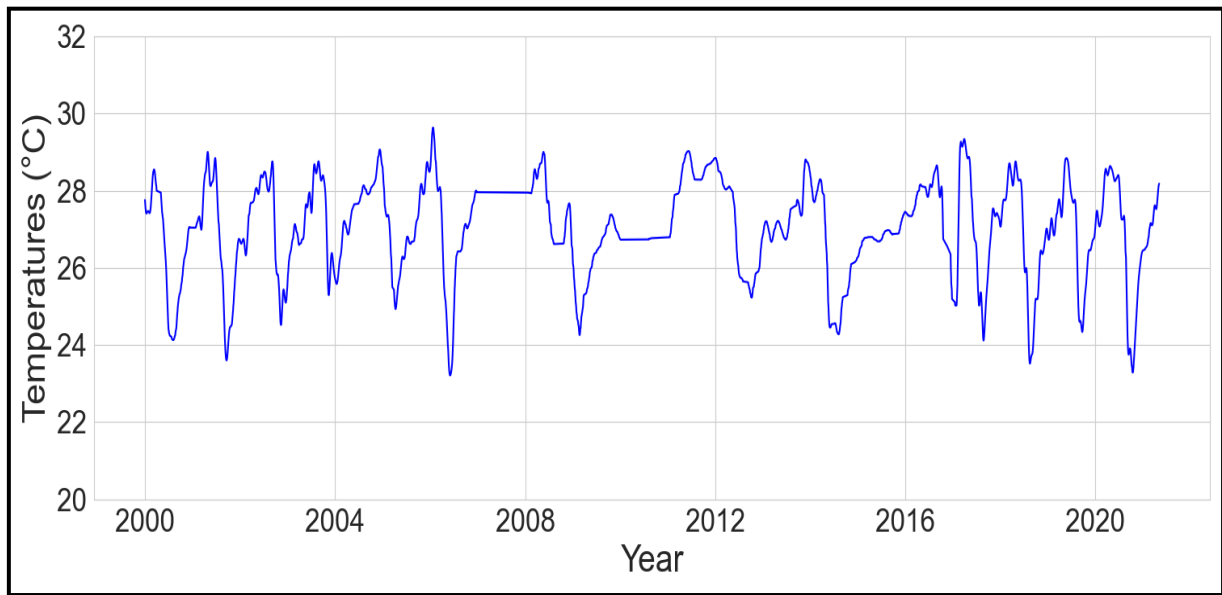


Fig. 4. *Tendance des Températures*

La ressemblance avec les données brutes semble indiquer que la tendance est pertinente pour la modélisation des températures dans la localité de Makokou.

### 3.2 MODELES DE RESEAUX DE NEURONES ET RESULTATS

#### 3.2.1 LONG SHORT-TERM MEMORY MODEL (LSTM) AND PERFORMANCE EVALUATION.

Une fois le modèle LSTM conçu, nous l'avons entraîné (figure 5) puis testé avec les données de test (figure 6). La figure 5 ci-après montre comment évolue les pertes d'entraînement et de validation au cours de la phase d'apprentissage du modèle.

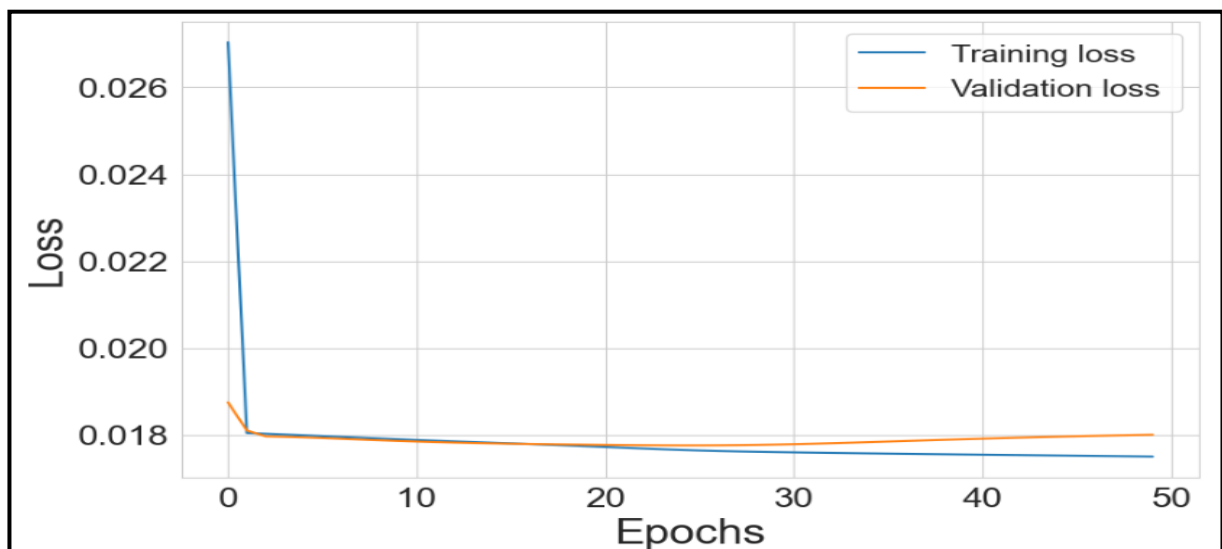


Fig. 5. *Courbes de pertes d'entraînement et de validation du modèle LSTM*

Les prédictions ont ensuite été faites après l'entraînement du modèle, La figure 6 met en évidence les résultats obtenus.

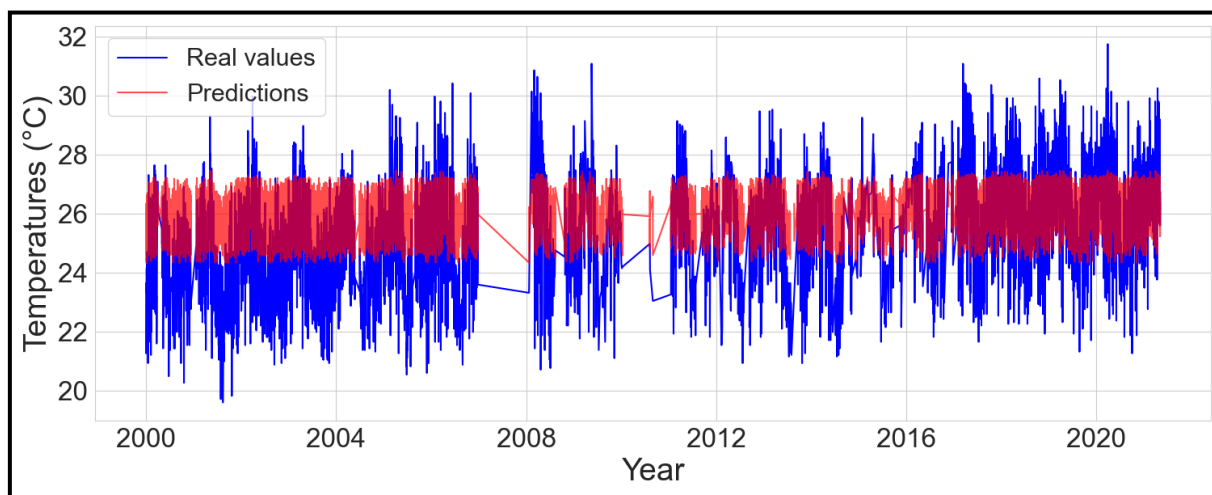


Fig. 6. Valeurs réelles et Prédictions du modèle LSTM

L'observation de la figure 6 révèle que le modèle LSTM simple c'est-à-dire celui n'ayant subi aucune transformation au préalable ne prédit pas de manière satisfaisante les températures.

Les métriques traditionnelles d'évaluation ont donné respectivement une valeur de 1.62 °C pour l'erreur quadratique moyenne, 1.30 °C pour l'erreur absolue moyenne et un pourcentage d'erreur absolu moyen de 4.89 %.

Afin d'évaluer davantage le modèle LSTM simple, nous avons effectué les tests non paramétriques de rangs signés de Wilcoxon et le coefficient de corrélation de Spearman.

Le Test de rangs signés de Wilcoxon révèle une différence significative entre les valeurs réelles et les valeurs prédites (Tableau 2).

Tableau 2. Résultats du Test de rang signés de Wilcoxon

YEARMODA	MAX	MIN	PRCP	FRSHTT
01/01/2000	75.2	68.4	0.39E	10
02/01/2000	84.9	65.5	0.03G	100000
03/01/2000	84.6	66.9	0.00G	100000
04/01/2000	84.6	67.1	0.00D	100000
05/01/2000	80.6	68.4	0.02E	100010

Le coefficient de corrélation de Spearman (Tableau 3) suggère une corrélation significative modérée entre les valeurs originales et les prédictions du modèle

Tableau 3. Résultats du Test de corrélation de Spearman

	Wilcoxon sign-rank	
	Statistic of Test Z	P-value
LSTM	137748	$1.12 \times 10^{-21}$

Les hyperparamètres du réseau de neurones LSTM développé sont consignés dans le tableau 4. Les hyperparamètres sont des paramètres que nous avons choisi intuitivement avant l'entraînement du modèle, dans le but de minimiser les métriques d'évaluation.

Tableau 4. LSTM Model hyperparameters

	Spearman's rank correlation	
	Correlation coefficient $\rho$	P-value
LSTM	0.40	$8.25 \times 10^{-34}$

### 3.2.2 TRANSFORMEE EN ONDELETTES (TO) ASSOCIEE À LA MEMOIRE À LONG ET COURT TERME (LONG SHORT-TERM MEMORY MODEL (LSTM))

En vue d'améliorer les performances de notre modèle LSTM, conçu précédemment, nous avons remplacé les données brutes (températures moyennes) par la tendance extraite à partir de la Transformée en Ondelettes. Ensuite, le modèle a été entraîné avec un nombre d'époques (epochs) de 50. La figure 7 décrit l'évolution des pertes d'entraînement et de validation.

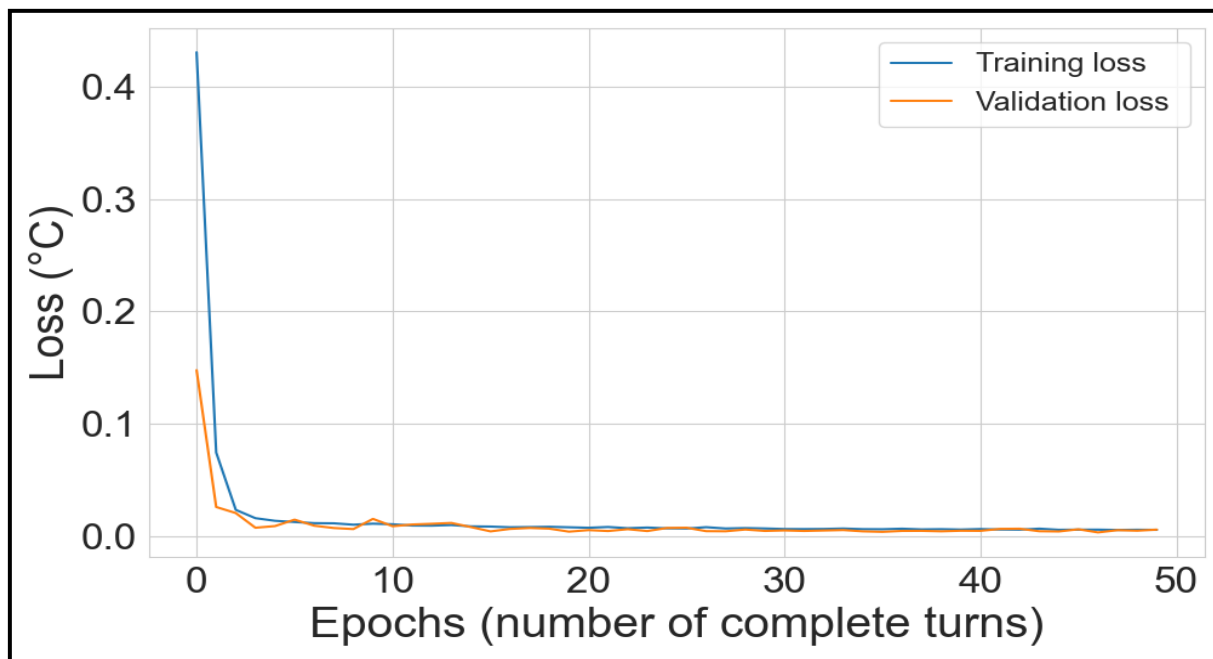


Fig. 7. Evolution des courbes de perte et de validation pendant l'entraînement

L'on semble apercevoir une superposition des courbes de perte et de validation au cours de l'entraînement (figure 7).

La figure 8 met en évidence les prédictions obtenues à partir de ce nouveau modèle LSTM. Nous constatons que cette fois ci, il y'a une correspondance presque parfaite entre les données originales et les prédictions.

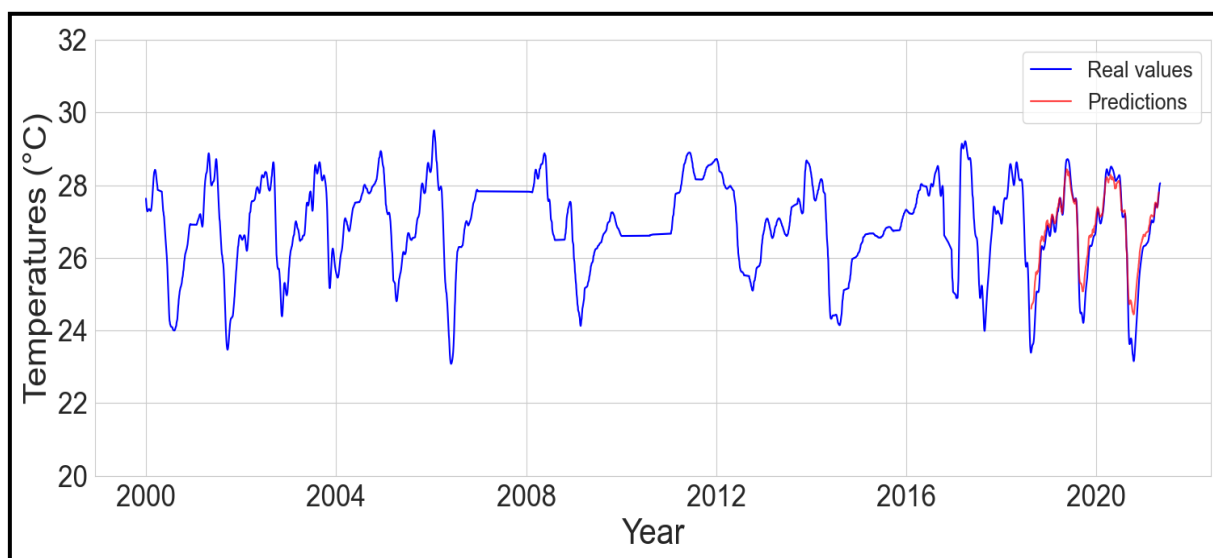


Fig. 8. Valeurs réelles et Prédictions du modèle TO-LSTM

L'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne (MAE) et le pourcentage d'erreur absolue moyen (MAPE) sont respectivement de 0.45 °C, 0.35 °C et 1.35 %.

Le Test de rangs signés de Wilcoxon indique qu'il n'y a pas de différence significative entre les valeurs réelles et les valeurs prédites (Tableau 5), ce qui semble confirmer que les valeurs prédites se rapprochent des valeurs originales.

**Tableau 5. Résultats du Test de rang signés de Wilcoxon**

Parameters	Values
Number of Units LSTM	50
Epochs	50
Batch_size	32
Optimizer	Adam
Learning rate	0.001
Dropout rate	0.5

Le coefficient de corrélation de Spearman (Tableau 6) suggère une corrélation forte et significative entre les valeurs initiales et les prédictions du modèle.

**Tableau 6. Résultats du Test de corrélation de Spearman**

	Wilcoxon sign-rank	
	Statistic of Test Z	P-value
TO-LSTM	126748	0.08

Le Test de Mann-Whitney U révèle une différence significative entre les résidus des modèles LSTM simple et TO-LSTM (Tableau 7)

**Tableau 7. Résultats du Test de Mann-Whitney U**

	Spearman's rank correlation	
	Correlation coefficient $\rho$	P-value
TO-LSTM	0.97	0

Enfin, nous avons évalué les contributions de chaque variable explicative à la sortie du modèle TO-LSTM grâce aux indices de Shapley (figure 9). Il ressort de ce calcul que le point de rosée moyen (DEWP), la pression moyenne au niveau de la mer (SLP) et la pression moyenne de la station pour la journée (STP) influencent fortement la sortie du modèle.

Ceci semble révéler que ces variables jouent un rôle crucial dans les prédictions. Juste après ces trois premières variables, s'ensuivent respectivement d'autres variables à savoir la visibilité (VISIB), les températures maximale (MAX) et minimale (MIN) avec des contributions sensiblement égales (0.018). Enfin on retrouve les précipitations totales (PRCP) et la variable FRSHTT résumant les conditions météorologiques.

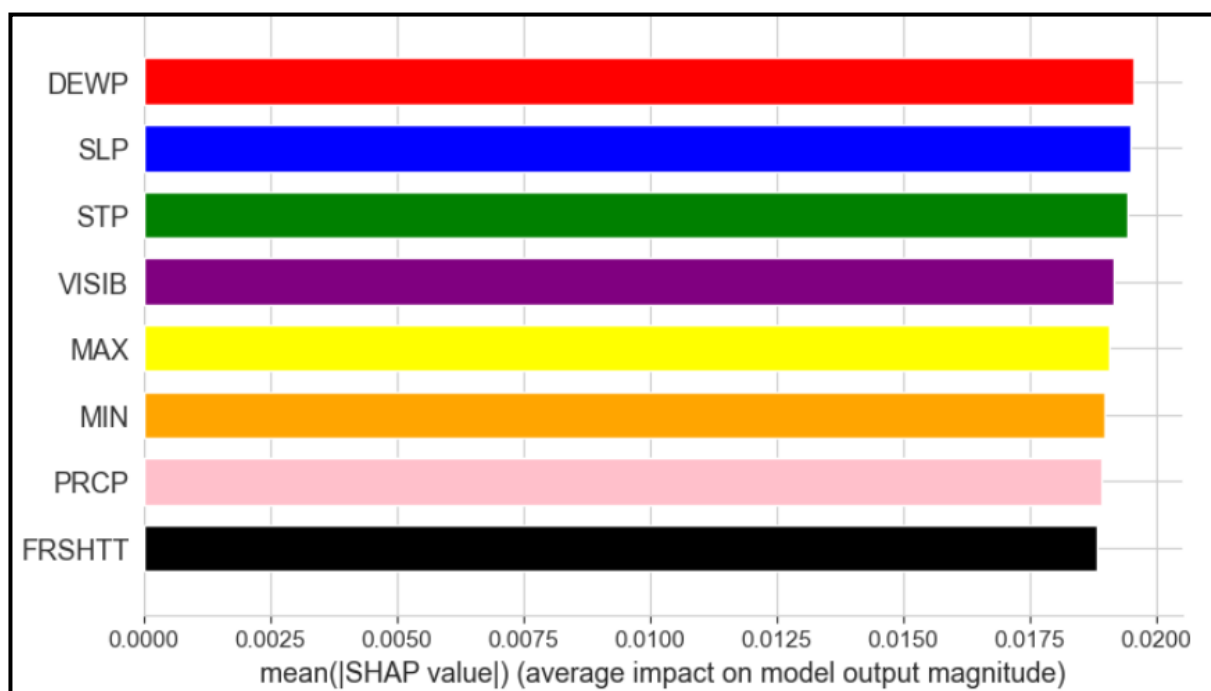


Fig. 9. Contributions des variables explicatives au modèle TO-LSTM

#### 4 DISCUSSION

Dès la 10<sup>ème</sup> époque lors de la phase d'entraînement de notre modèle LSTM simple, une convergence du modèle semble apparaître. Cependant, pour le modèle TO-LSTM, la convergence est observée plutôt (autour de la troisième époque) et est perceptible tout au long de l'entraînement avec des pertes d'entraînement et de validation qui oscillent quasiment autour de zéro. Cette convergence très rapide du modèle LSTM-TO pourrait s'expliquer par la présence des données d'entraînement assez représentatives des données et d'un choix efficace des hyperparamètres des modèles. En effet les métriques d'évaluation que sont les RMSE, MAE et MAPE valent respectivement 0.45 °C, 0.35 °C et 1.35 %. De telles valeurs pour nos indicateurs de performances suggèrent que l'écart entre les prédictions et les valeurs réelles est minime. Ces résultats semblent se rapprocher de ceux obtenus dans une étude (Park et al. 2019). Dans cette étude, un modèle LSTM a également été conçu pour prédire la température sur des courtes périodes (6, 12 et 24 heures) puis étendu à 7 et 14 jours. Ce modèle LSTM comportait 4 couches LSTM et 384 neurones (ou unités). Les RMSE obtenues dans cette étude étaient de 0.79 °C pour la prédiction de 24 heures, 2.84 °C pour la prédiction de 7 jours et 3.06 °C pour la prédiction de 14 jours. De plus, nos résultats semblent se rapprocher de ceux obtenus dans une autre étude (Inik et al. 2022). Celle-ci a analysé une base de données de températures au sol de 2013 à 2021 en Turquie et un modèle LSTM a servi à prédire la température moyenne. Effectivement, au cours de cette analyse, les auteurs ont mis au point un modèle hybride LSTM-GRU capable de prédire la température moyenne du sol dans la localité de Bingöl (Turquie). Le modèle développé avait pour hyperparamètres, une vitesse d'apprentissage de 0.001, deux couches (LSTM et GRU) avec chacune 200 unités et une couche Dense pour la sortie finale du modèle. L'erreur quadratique moyenne (RMSE) obtenue était de 1.25 °C. Toutes les valeurs des métriques obtenues dans ces différentes études sont proches des nôtres. Bien que n'ayant pas travaillé sur la même base de données dans chacun des cas, cette comparaison avec ces différentes études nous est utile tout de même dans la mesure où il s'agit du même problème à résoudre mais aussi et surtout de la même stratégie LSTM utilisée. Nous montrons dans notre travail, qu'avec un faible nombre de couches (2 couches LSTM) et neurones (50), notre modèle LSTM simple obtient certes des prédictions moins satisfaisantes, mais après l'extraction de la tendance via une Transformée en Ondelettes, le modèle TO-LSTM ayant toujours le même nombre de couches LSTM et d'unités, prédit maintenant mieux les températures, avec des écarts minimes entre les valeurs réelles et les valeurs prédites.

Le test de rangs signés de Wilcoxon révèle une différence significative entre les valeurs réelles et prédites pour le modèle LSTM simple, avec notamment une p-value de  $1.12 \times 10^{-21}$ . Par contre dans le cas du modèle TO-LSTM, le test de rangs signés de Wilcoxon suggère qu'il n'y a pas de différence significative entre les valeurs réelles et prédites. Ces résultats montrent que le modèle LSTM simple bien qu'efficace parfois dans certaines situations, n'arrive pas ici à capturer certaines fluctuations spécifiques des séries de températures. En revanche, la p-value (0.08) supérieure au seuil de significativité (0.05) dans le second cas souligne que la combinaison Transformée en Ondelettes et LSTM est adaptée pour mieux capturer les relations complexes sous-jacentes dans les données de températures. L'intégration de la Transformée en Ondelettes dans la structure du LSTM améliore grandement les capacités de ce dernier à capturer les variations multi-échelles et les fluctuations des séries temporelles. Plusieurs études confirment l'efficacité de l'intégration de la

Transformée en Ondelettes dans les modèles prédictifs. Dans ce sens, une récente étude explore l'utilisation conjointe de la Transformée en Ondelettes et du LSTM pour la prévision des données météorologiques et montre que cette combinaison permet une amélioration significative de la précision des prédictions, en particulier pour les séries temporelles non stationnaires. L'étude a en effet montré que la prédiction de la température de la colonne de distillation était améliorée de 10 % avec le réseau LSTM, lorsque les données étaient transformées via la Transformée en Ondelettes au préalable (Kwon et al. 2022). Un autre article montre que les modèles de réseaux de neurones basés sur la Transformée en Ondelettes obtiennent de meilleures performances en matière de prédiction des séries temporelles climatiques avec une meilleure capture des variations à différentes échelles. En effet cette étude a utilisé l'association Transformée en Ondelettes et réseau LSTM pour prédire la température moyenne journalière dans la ville de Mugla (Turquie). L'étude précise qu'avec cette combinaison le modèle s'est avéré meilleur que le LSTM standard, avec notamment une erreur quadratique moyenne de  $0.56\text{ }^{\circ}\text{C}$  (Ghasemlounia et al. 2024).

Le coefficient de corrélation de Spearman est de 0.40 pour le modèle LSTM. Bien que cette valeur soit statistiquement significative avec une P-value de  $8.25 \times 10^{-34}$  inférieure à 0.05, elle indique une corrélation modérée entre les valeurs prédites et les valeurs réelles. Cette corrélation modérée suggère que le modèle LSTM simple ne capture pas suffisamment les relations complexes existantes dans les données. Cependant, pour le modèle TO-LSTM, le coefficient de corrélation de Spearman atteint une valeur impressionnante de 0.97 avec P-value nulle, indiquant une corrélation presque parfaite entre les valeurs prédites et réelles. Cela confirme que le modèle TO-LSTM est efficace pour capturer et reproduire les tendances des données réelles, offrant ainsi des prédictions précises et fiables.

Dans une étude complémentaire (Vikas Goyal et al. 2022), les auteurs ont démontré que le coefficient de corrélation de Spearman est un critère de performance pertinent pour évaluer des modèles de prévision de températures, tels que ceux utilisant les réseaux de neurones LSTM. Ils ont observé que le coefficient de Spearman s'est avéré statistiquement significatif dans tous les modèles évalués, renforçant son utilité dans la validation des performances des modèles.

De plus, le Test de Mann-Whitney U révèle une différence significative entre les résidus des modèles LSTM simple et TO-LSTM. Cette différence indique que les deux modèles ne sont pas équivalents en termes de capacité à prédire la température moyenne dans la localité de Makokou. Les résultats du Test de Mann-Whitney U, combinés au coefficient de corrélation de Spearman, suggèrent que le modèle TO-LSTM pourrait offrir une performance supérieure par rapport au modèle LSTM simple. Ces tests corroborent ainsi l'idée que le modèle TO-LSTM est non seulement plus performant en termes de précision des prédictions mais aussi plus robuste pour capturer les complexités des données. Dans une autre étude des modèles ont été développés pour prédire la température dans des centrales éoliennes (Mishra et al. 2020). La comparaison desdits modèles a montré que les modèles basés sur les Transformées en Ondelettes ou de Fourier présentent des meilleures performances, montrant ainsi la plus-value qu'apportent ces Transformées aux prédictions des modèles d'apprentissage profond.

Les valeurs de Shapley révèlent que le point de rosée moyen (DEWP), la pression moyenne au niveau de la mer (SLP) et la pression moyenne de la station pour la journée (STP) influencent fortement la sortie du modèle. En effet le point de rosée moyen est un indicateur important pour prédire la formation de brouillard, de pluie ou de neige, ce qui influence directement les prédictions des températures via des modèles météorologiques. Une étude a analysé la relation entre la température de l'air et le point de rosée moyen de rosée aux Etats-Unis (Russell 2024). L'étude démontre que le point de rosée moyen joue un rôle clé dans la prédiction des températures et précipitations, car il est étroitement lié à l'humidité de l'air. En effet, plus le point de rosée moyen est élevé, plus l'air contient de l'humidité, ce qui affecte directement les modèles météorologiques utilisés pour prédire les températures. Les auteurs concluent que la prise en compte du point de rosée moyen dans les modèles de prédictions permet d'améliorer l'exactitude des prévisions, notamment pour les événements climatiques extrêmes tels que les vagues de chaleur ou les périodes de fortes pluies.

Ensuite, les pressions moyennes au niveau de la mer et de la station influencent de manière significative les phénomènes météorologiques à grande échelle tels que les températures. En effet, la pression atmosphérique influence la température en affectant la densité de l'air et sa capacité à retenir la chaleur. Par exemple dans une zone de haute pression, l'air descend ce qui pourrait entraîner une augmentation des températures en raison de la compression adiabatique. A l'inverse, dans une zone de basse pression l'air monte ce qui peut entraîner une diminution des températures (Ning et al. 2018). Dans ce sens, une étude a montré que les modèles prédictifs d'apprentissage profond (comme les LSTM), exploitent généralement la pression au niveau de la mer pour prévoir des événements comme les tempêtes, ce qui confirme bien l'apport considérable de la pression au niveau de la mer dans la prédiction des variables climatiques (Rus et al. 2023).

Bien qu'ayant la plus faible contribution de Shapley, la variable FRSHTT résumant les conditions météorologiques Fog (Brouillard), Rain (Pluie), Snow (Neige), Hail (Grêle), Thunder (Tonnerre) et Tornado (Tornade), influence aussi la sortie du modèle. Ceci à juste titre confirme tout de même le rôle prépondérant des conditions météorologiques sur la prédiction des températures. Cela est confirmé par de récents travaux (Cifuentes et al. 2020; Azari et al. 2022) qui présentent un nombre important d'études dans lesquelles les différentes conditions météorologiques comprenant bien évidemment les précipitations totales et la température maximale influencent grandement les prédictions de température globale de la terre, régionale ou même locale.

## 5 CONCLUSION

L'objectif majeur de notre étude était de prédire les températures dans la localité de Makokou en exploitant une base de données de paramètres climatiques couvrant la période 2000 à 2020. Dans cette optique, deux modèles d'apprentissage profond ont été proposés: le premier basé sur une architecture LSTM classique, et le second intégrant une Transformée en Ondelettes (TO-LSTM) pour capturer les tendances et variations sous-jacentes dans les séries temporelles.

La validation des modèles s'est appuyée sur des métriques traditionnelles telles que la racine carrée de l'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne (MAE) et le pourcentage d'erreur absolu moyen (MAPE). En parallèle, des tests non paramétriques, à savoir le Test de rangs signés de Wilcoxon, le coefficient de corrélation de Spearman, et le test de Mann-Whitney U, ont été employés pour renforcer les critères de validation des modèles.

Les résultats obtenus ont montré une RMSE de 0.45°C pour le modèle TO-LSTM, contre 1.62°C pour le modèle LSTM simple ou classique. Ces résultats indiquent que le modèle basé sur la Transformée en Ondelettes (TO-LSTM) capture efficacement les relations complexes dans les données météorologiques. Le test de rangs signés de Wilcoxon a révélé qu'il n'existait pas de différences significatives entre les prédictions du modèle TO-LSTM et les valeurs réelles. Cette conclusion a été renforcée par le coefficient de corrélation de Spearman, qui a atteint une valeur de 0.97 pour le modèle TO-LSTM. Cette valeur, bien supérieure à celle obtenue pour le modèle simple (0.40), souligne une forte corrélation significative entre les prédictions du modèle TO-LSTM et les valeurs réelles.

De plus, le test de Mann-Whitney U a mis en évidence une différence significative entre les résidus des deux modèles, démontrant que les deux architectures ne capturent pas de manière identique les relations sous-jacentes dans les données. Cela confirme que le modèle TO-LSTM, qui intègre la Transformée en Ondelettes est plus apte à modéliser les tendances climatiques complexes dans les séries temporelles de température.

En conclusion, cette étude démontre que l'ajout de la Transformée en Ondelettes dans une architecture LSTM permet d'améliorer considérablement la performance prédictive des réseaux de neurones récurrents, en particulier dans le contexte de la prévision des températures à long terme. Il serait intéressant, dans des études futures, d'explorer l'intégration de ce modèle avec d'autres techniques avancées d'apprentissage profond et d'appliquer cette méthode à d'autres paramètres climatiques tels que les précipitations, afin de généraliser ces résultats et d'améliorer la compréhension des dynamiques météorologiques dans diverses régions du Bassin du Congo.

## DISPONIBILITÉ DU CODE

Les codes Python pour tous les scripts utilisés dans cette étude sont disponibles sur demande auprès du principal correspondant.

## REMERCIEMENTS

Nous tenons à remercier l'Institut de Recherche pour le Développement (IRD) pour son soutien multiforme tout au long de ces années de thèse.

## CONFLITS D'INTÉRÊTS

Les auteurs déclarent qu'ils n'ont aucun conflit d'intérêts.

## REFERENCES

- [1] Abubakar A, Chiroma H, Zeki A, Uddin M (2016) Utilising key climate element variability for the prediction of future climate change using a support vector machine model. *International Journal of Global Warming* 9: 129–151. <https://doi.org/10.1504/IJGW.2016.074952>.
- [2] Adamo N, Al-Ansari N, Sissakian VK, et al (2018) Climate Change : Consequences on Iraq's Environment. *Journal of Earth Sciences and Geotechnical Engineering* 8: 43–58.
- [3] Alessio SM (2016) Discrete Wavelet Transform (DWT). In: Alessio SM (ed) *Digital Signal Processing and Spectral Analysis for Scientists: Concepts and Applications*. Springer International Publishing, Cham, pp 645–714.
- [4] Al-Hameed AA, Khawla (2022) Spearman's correlation coefficient in statistical analysis. *International Journal of Nonlinear Analysis and Applications* 13: 3249–3255. <https://doi.org/10.22075/ijnaa.2022.6079>.
- [5] Arfaoui S, Mabrouk AB, Cattani C (2021) *Wavelet Analysis: Basic Concepts and Applications*. Chapman and Hall/CRC, New York.
- [6] Astivia OLO, Zumbo BD (2017) Population models and simulation methods: The case of the Spearman rank correlation. *British Journal of Mathematical and Statistical Psychology* 70: 347–367. <https://doi.org/10.1111/bmsp.12085>.

- [7] Atyi R, Hiol H, Lescuyer G, et al (2022) Les Forêts du bassin du Congo : Etat des Forêts 2021. In: Centre de Recherche Forestière Internationale (CIFOR).
- [8] Azari B, Hassan K, Pierce J, Ebrahimi S (2022) Evaluation of Machine Learning Methods Application in Temperature Prediction. *CRPASE* 8: 1–12. <https://doi.org/10.52547/crpase.8.1.2747>.
- [9] Bagastio K, Oetama RS, Ramadhan A (2023) Development of stock price prediction system using Flask framework and LSTM algorithm. *Journal of Infrastructure, Policy and Development* 7: <https://doi.org/10.24294/jipd.v7i3.2631>.
- [10] Balada A (2021) Le Bassin du Congo, deuxième puits de carbone du monde entre préservation et exploitation. *Le Monde.fr*.
- [11] Beaudoin S, Chaloux A (2023) Négociations climatiques : COP 27 à la COP 28, Observatoire Multilatéralisme & Organisations internationales. <https://observatoire-multilateralisme.fr/publications/negociations-climatiques-cop-27-a-la-cop-28/>. Accessed 10 Jan 2024.
- [12] Belle EMS, Burgess ND, Misrachi M (2016) Impacts du changement climatique sur la biodiversité et les aires protégées en Afrique de l'Ouest, Résumé des résultats du projet PARCC, Aires protégées résilientes au changement climatique en Afrique de l'Ouest. Rapport UNEP-WCMC 52p.
- [13] Berzaghi F, Longo M, Ciaï P, et al (2019) Carbon stocks in central African forests enhanced by elephant disturbance. *Nat Geosci* 12: 725–729. <https://doi.org/10.1038/s41561-019-0395-6>.
- [14] Beunk J (2021) Seasonal Forecasting of Rainfall in Equatorial East Africa using an Artificial Neural Network. Master Thesis.
- [15] Bharadiya J (2023) Exploring the Use of Recurrent Neural Networks for Time Series Forecasting. *International Journal of Innovative Research in Science Engineering and Technology* 8: 2023. <https://doi.org/10.5281/zenodo.8002429>.
- [16] Boutahir MK, Farhaoui Y, Azrour M (2022) Machine Learning and Deep Learning Applications for Solar Radiation Predictions Review: Morocco as a Case of Study. In: Yaseen SG (ed) *Digital Economy, Business Analytics, and Big Data Analytics Applications*. Springer International Publishing, Cham, pp 55–67.
- [17] Bush ER (2018) Tropical Phenology In A Time Of Change. <https://doi.org/Phd Thesis, Stirling University>.
- [18] Bush ER, Jeffery K, Bunnefeld N, et al (2020) Rare ground data confirm significant warming and drying in western equatorial Africa. *PeerJ* 8: e8732. <https://doi.org/10.7717/peerj.8732>.
- [19] Chakanyuka TL (2019) The Conservation of African Elephants under the CITES International Ivory Trade Ban. *Kathmandu Sch L Rev (KSLR)* Vol.7, Issue 2: p.71-83.
- [20] Chen G, Li K, Liu Y (2021) Applicability of Continuous, Stationary, and Discrete Wavelet Transforms in Engineering Signal Processing. *Journal of Performance of Constructed Facilities* 35: 04021060. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0001641](https://doi.org/10.1061/(ASCE)CF.1943-5509.0001641).
- [21] Chirwa PW, Adeyemi O (2020) Deforestation in Africa: Implications on Food and Nutritional Security. In: Leal Filho W, Azul AM, Brandli L, et al. (eds) *Zero Hunger*. Springer International Publishing, Cham, pp 197–211.
- [22] Cifuentes J, Marulanda G, Bello A, Reneses J (2020) Air Temperature Forecasting Using Machine Learning Techniques: A Review. *Energies* 13: 4215. <https://doi.org/10.3390/en13164215>.
- [23] Daud SNSS, Sudirman R (2022) Wavelet Based Filters for Artifact Elimination in Electroencephalography Signal: A Review. *Ann Biomed Eng* 50: 1271–1291. <https://doi.org/10.1007/s10439-022-03053-5>.
- [24] Dellink R, Hwang H, Lanzi E, Chateau J (2017) International trade consequences of climate change. OCDE, Paris.
- [25] Denis C, Varenne F (2022) Interprétabilité et explicabilité de phénomènes prédits par de l'apprentissage machine. *Revue Ouverte d'Intelligence Artificielle* 3: 287–310. <https://doi.org/10.5802/roia.32>.
- [26] Dominguez D, Barriuso Pastor J, Pantoja-Díaz O, González-Rodríguez M (2023) Forecasting Worldwide Temperature from Amazon Rainforest Deforestation Using a Long-Short Term Memory Model. *Sustainability* 15: 15152. <https://doi.org/10.3390/su152015152>.
- [27] Eltehiwy M, Abdul-Motaal AB (2023) A new Method for Computing and Testing The significance of the Spearman Rank Correlation. *Computational Journal of Mathematical and Statistical Sciences*. <https://doi.org/10.21608/cjmss.2023.229746.1015>.
- [28] Ezugwu AE, Oyelade ON, Ikotun AM, et al (2023) Machine Learning Research Trends in Africa: A 30 Years Overview with Bibliometric Analysis Review. *Arch Computat Methods Eng* 30: 4177–4207. <https://doi.org/10.1007/s11831-023-09930-z>.
- [29] Gad I, Hosahalli D (2022) A comparative study of prediction and classification models on NCDC weather data. *International Journal of Computers and Applications* 44: 414–425. <https://doi.org/10.1080/1206212X.2020.1766769>.
- [30] Ghasemlounia R, Gharehbaghi A, Ahmadi F, Albaji M (2024) Developing a novel hybrid model based on deep neural networks and discrete wavelet transform algorithm for prediction of daily air temperature. *Air Qual Atmos Health*. <https://doi.org/10.1007/s11869-024-01595-2>.
- [31] Gong B, Langguth M, Ji Y, et al (2022) Temperature forecasting by deep learning methods. *Geoscientific Model Development* 15: 8931–8956. <https://doi.org/10.5194/gmd-15-8931-2022>.
- [32] Gossler FE, Duarte MAQ, Villarreal F (2023) Design of Nearly-Orthogonal Symmetric Wavelet Filter Banks Based on the Wavelet Orthogonalization Process. *Circuits Syst Signal Process* 42: 234–254. <https://doi.org/10.1007/s00034-022-02111-6>.
- [33] Grooten M, Almond REA (2018) Living Planet Report 2018: Summary. In: CABI Digital Library. WWF.
- [34] Guillaume S-C (2019) Apprendre Le Machine Learning en UNE semaine.
- [35] Guo T, Zhang T, Lim E, et al (2022) A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities. *IEEE Access* 10: 58869–58903. <https://doi.org/10.1109/ACCESS.2022.3179517>.



- [36] Guo Y, Zhao R, Zeng Y, et al (2018) Identifying scale-specific controls of soil organic matter distribution in mountain areas using anisotropy analysis and discrete wavelet transform. *CATENA* 160: 1–9. <https://doi.org/10.1016/j.catena.2017.08.016>
- [37] Haque E, Tabassum S, Hossain E (2021) A Comparative Analysis of Deep Neural Networks for Hourly Temperature Forecasting. *IEEE Access* 9: 160646–160660. <https://doi.org/10.1109/ACCESS.2021.3131533>.
- [38] Hou J, Wang Y, Zhou J, Tian Q (2022) Prediction of hourly air temperature based on CNN–LSTM. *Geomatics, Natural Hazards and Risk* 13: 1962–1986. <https://doi.org/10.1080/19475705.2022.2102942>.
- [39] Hu J, Wang X, Zhang Y, et al (2020) Time Series Prediction Method Based on Variant LSTM Recurrent Neural Network. *Neural Process Lett* 52: 1485–1500. <https://doi.org/10.1007/s11063-020-10319-3>.
- [40] Inik O, Inik Ö, Öztaş T, Yuksel A (2022) Soil Temperature Prediction with Long Short Term Memory (LSTM). *Türk Tarım ve Doğa Bilimleri Dergisi* 9: 779–785. <https://doi.org/10.30910/turkjans.1101753>.
- [41] Isabona J, Kehinde R (2019) MULTI-RESOLUTION BASED DISCRETE WAVELET TRANSFORM FOR ENHANCED SIGNAL COVERAGE PROCESSING AND PREDICTION ANALYSIS. *FUDMA JOURNAL OF SCIENCES* 3: 6–15.
- [42] Jofipasi CA, Miftahuddin, Hizir (2018) Selection for the best ETS (error, trend, seasonal) model to forecast weather in the Aceh Besar District. *IOP Conf Ser: Mater Sci Eng* 352: 012055. <https://doi.org/10.1088/1757-899X/352/1/012055>.
- [43] Kang H, Yang S, Huang J, Oh J (2020) Time Series Prediction of Wastewater Flow Rate by Bidirectional LSTM Deep Learning. *Int J Control Autom Syst* 18: 3023–3030. <https://doi.org/10.1007/s12555-019-0984-6>.
- [44] Karaman K, Sainte Fare Garnot V, Wegner JD (2023) DEFORESTATION DETECTION IN THE AMAZON WITH SENTINEL-1 SAR IMAGE TIME SERIES. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences X-1-W1-2023*: 835–842. <https://doi.org/10.5194/isprs-annals-X-1-W1-2023-835-2023>.
- [45] Karevan Z, Suykens JAK (2020) Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks* 125: 1–9. <https://doi.org/10.1016/j.neunet.2019.12.030>
- [46] Kermabon M (2022) Vers une atténuation des conflits hommes-éléphants : Cas de l'aire protégée de Moukalaba-doudou au Gabon. Mémoire de recherche, Université Le Mans.
- [47] Kumar BBS, Satyanarayana PS (2022) A mixture of Noise Image Denoising using Sevenlets Wavelet Techniques. *Trends in Sciences* 19: 4186–4186. <https://doi.org/10.48048/tis.2022.4186>.
- [48] Kwon H, Choi Y, Park H, et al (2022) Distillation Column Temperature Prediction Based on Machine-Learning Model Using Wavelet Transform. In: Yamashita Y, Kano M (eds) *Computer Aided Chemical Engineering*. Elsevier, pp 1651–1656.
- [49] Laguardia A, Gobush KS, Bourgeois S, et al (2021) Assessing the feasibility of density estimation methodologies for African forest elephant at large spatial scales. *Global Ecology and Conservation* 27: e01550. <https://doi.org/10.1016/j.gecco.2021.e01550>.
- [50] Lee H, Calvin K, Dasgupta D, et al (2023) IPCC, 2023: Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. <https://doi.org/10.59327/IPCC/AR6-9789291691647.001>. Accessed 19 Aug 2024.
- [51] Li C, Zhang Y, Ren X (2020) Modeling Hourly Soil Temperature Using Deep BiLSTM Neural Network. *Algorithms* 13: 173. <https://doi.org/10.3390/a13070173>.
- [52] MAIDOU HM (2020) Gouvernance forestière et REDD+: le cas du projet de renforcement des capacités institutionnelles pour les forêts du bassin du Congo. *La gouvernance forestière en Afrique centrale: Entre pratiques et politiques* 261.
- [53] Mishra S, Bordin C, Taharaguchi K, Palu I (2020) Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature. *Energy Reports* 6: 273–286. <https://doi.org/10.1016/j.egy.2019.11.009>
- [54] Moazenzadeh R, Mohammadi B, Duan Z, Delghandi M (2022) Improving generalisation capability of artificial intelligence-based solar radiation estimator models using a bio-inspired optimisation algorithm and multi-model approach. *Environ Sci Pollut Res* 29: 27719–27737. <https://doi.org/10.1007/s11356-021-17852-1>.
- [55] Moukodouma D-FB, Mbourou DRR, Nkoulembene CA, Denis C (2023) A temperatures variation favor human-elephant conflict in Gabon's Lékédi National Park. *IJAERS* 10: 007–026. <https://doi.org/10.22161/ijaers.108.2>.
- [56] Ning G, Wang S, Yim SHL, et al (2018) Impact of low-pressure systems on winter heavy air pollution in the northwest Sichuan Basin, China. *Atmospheric Chemistry and Physics* 18: 13601–13615. <https://doi.org/10.5194/acp-18-13601-2018>.
- [57] Nounangnonhou TC, Fifatin F-XN (2016) Modélisation et simulation des tendances climatiques à l'horizon 2040 sur le bassin du fleuve Ouémé en République du Bénin.
- [58] Park I, Kim HS, Lee J, et al (2019) Temperature Prediction Using the Missing Data Refinement Model Based on a Long Short-Term Memory Neural Network. *Atmosphere* 10: 718. <https://doi.org/10.3390/atmos10110718>.
- [59] Piccolroaz S, Calamita E, Majone B, et al (2016) Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches. *Hydrological Processes* 30: 3901–3917. <https://doi.org/10.1002/hyp.10913>.
- [60] Ponni alias Sathya S, Ramakrishnan S (2020) Non-redundant frame identification and keyframe selection in DWT-PCA domain for authentication of video. *IET Image Processing* 14: 366–375. <https://doi.org/10.1049/iet-ipr.2019.0341>.
- [61] Qiu C (2023) A Method Using LSTM Networks to Impute Missing Temperatures in Temperature Datasets and to Predict Future Temperatures. *Highlights in Science, Engineering and Technology* 46: 116–124. <https://doi.org/10.54097/hset.v46i.7691>.

- [62] Ren P, Liu Z, Zhou X, et al (2021) Strong controls of daily minimum temperature on the autumn photosynthetic phenology of subtropical vegetation in China. *Forest Ecosystems* 8: 31. <https://doi.org/10.1186/s40663-021-00309-9>.
- [63] Roland ZKC, Ornella MN, Donald MI, et al (2016) Repartition Des Glossines Dans La Province De L'ogoooue Ivindo Ancien Foyer De Trypanosomose Humaine Africaine. *ESJ* 12: 281. <https://doi.org/10.19044/esj.2016.v12n12p281>.
- [64] Rus M, Fettich A, Kristan M, Ličer M (2023) HIDRA2: deep-learning ensemble sea level and storm tide forecasting in the presence of seiches – the case of the northern Adriatic. *Geoscientific Model Development* 16: 271–288. <https://doi.org/10.5194/gmd-16-271-2023>.
- [65] Russell K (2024) Seasonal Variation Of Dew-Point Temperatures In The United States. <https://temperatures.com/weather-and-climate/seasonal-variation-of-dew-point-temperatures-in-the-united-states/>. Accessed 16 Sep 2024.
- [66] Shahbazzabar D, Alirezaee S, Ahmadi M, Heydari R (2018) A MC-CDMA system based on orthogonal filter banks of wavelet transforms and partial combining. *AEU - International Journal of Electronics and Communications* 94: 128–138. <https://doi.org/10.1016/j.aeue.2018.05.026>
- [67] Torre-Schaub M (2023) Agir sans attendre pour le climat: la clé d'un avenir viable Commentaire de la synthèse du 6e rapport du GIEC approuvé et publié le 19 mars 2023. *Énergie - Environnement - Infrastructures : actualité, pratiques et enjeux* 3.
- [68] Toto E (2023) Agroecology alliance calls for more food at less cost to nature in Congo Basin. In: *Mongabay Environmental News*. <https://news.mongabay.com/2023/08/agroecology-alliance-calls-for-more-food-at-less-cost-to-nature-in-congo-basin/>. Accessed 17 Sep 2024.
- [69] Vanos J, Guzman-Echavarría G, Baldwin JW, et al (2023) A physiological approach for assessing human survivability and liveability to heat in a changing climate. *Nat Commun* 14: 7653. <https://doi.org/10.1038/s41467-023-43121-5>.
- [70] Vapnik V (2013) *The Nature of Statistical Learning Theory*. Springer Science & Business Media.
- [71] Vecellio DJ, Wolf ST, Cottle RM, Kenney WL (2022) Evaluating the 35°C wet-bulb temperature adaptability threshold for young, healthy subjects (PSU HEAT Project). *Journal of Applied Physiology* 132: 340–345. <https://doi.org/10.1152/jappphysiol.00738.2021>.
- [72] Vengatesan K, Mahajan SB, Sanjeevikumar P, et al (2018) Performance Analysis of Gene Expression Data Using Mann–Whitney U Test. In: Konkani A, Bera R, Paul S (eds) *Advances in Systems, Control and Automation: ETAEERE-2016*. Springer, Singapore, pp 701–709.
- [73] Vierra A, Razzaq A, Andreadis A (2023) Chapter 27 - Continuous variable analyses: t-test, Mann–Whitney U, Wilcoxon sign rank. In: Eltorai AEM, Bakal JA, Newell PC, Osband AJ (eds) *Translational Surgery*. Academic Press, pp 165–170.
- [74] Vikas Goyal, Ayay Yadav, Rahul Mukherjee (2022) Performance Evaluation of Machine Learning and Deep Learning Models for Temperature Prediction in Poultry Farming. <https://ieeexplore.ieee.org/abstract/document/9791771>. Accessed 16 Sep 2024.
- [75] Wall Emerson R (2023) Mann-Whitney U test and t-test. *Journal of Visual Impairment & Blindness* 117: 99–100. <https://doi.org/10.1177/0145482X221150592>.
- [76] Wang X, Wang X, Wang L, et al (2021) A Distributed Fusion LSTM Model to Forecast Temperature and Relative Humidity in Smart Buildings. In: *2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA)*. pp 1–6.
- [77] Xia K, Huang J, Wang H (2020) LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access* 8: 56855–56866. <https://doi.org/10.1109/ACCESS.2020.2982225>.
- [78] Yadav A, Jha CK, Sharan A (2020) Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science* 167: 2091–2100. <https://doi.org/10.1016/j.procs.2020.03.257>
- [79] Zameer A, Jaffar F, Shahid F, et al (2023) Short-term solar energy forecasting: Integrated computational intelligence of LSTMs and GRU. *PLOS ONE* 18: e0285410. <https://doi.org/10.1371/journal.pone.0285410>.