



UNIVERSITE D'ABOMEY CALAVI (UAC)

INSTITUT NATIONAL DE L'EAU(INE)

DEPARTEMENT D'HYDROLOGIE ET DE GESTION DES RESSOURCES EN EAU

Mémoire de fin de formation pour l'obtention du diplôme de

Licence Professionnelle en Hydrologie

OPTION : HYDROLOGIE QUANTITATIVE

Vers le développement d'un modèle de prévision des débits à partir des réanalyses ERA5 et LAND à l'aide des méthodes de l'intelligence artificielle dans le bassin versant du Mono.

Rédigé et soutenu par :

AGADJI Koffi Fabrice

Sous la direction de :

Maître de Mémoire :

Dr NTCHA MPO Yèkambèssoun

Maître de Stage :

M. René BODJRENOU

Mémoire soutenu le 07 Décembre 2024 avec mention excellente en face du Jury composé de :

Président du Jury : Professeur Mathieu Maurice AHOUANSSOU

Examineur : Docteur BONGOU Ernesto

Rapporteur : Docteur N'TCHA M'PO Yèkambèssoun

ANNEE ACADEMIQUE 2023-2024

CERTIFICATION

Nous soussignés :

- **Président du Jury** : Professeur Mathieu Maurice AHOUANSON. Maître de Conférence des Universités de CAMES, spécialiste en Aménagement et gestion des ressources naturelles

Membres du Jury

- **Docteur N'TCHA M'PO Yèkambèssoun** : Enseignant chercheur à l'Institut National de l'Eau de l'Université d'Abomey Calavi, Hydrologue et géostatisticien
- **Docteur BONGOU Ernesto** : Enseignant chercheur à l'Institut National de l'Eau de l'Université d'Abomey Calavi, Hydrologue

Jury chargé de juger le mémoire de fin de formation pour l'obtention du diplôme de Licence professionnelle en hydrologie quantitative et gestion des ressources en eau de Monsieur AGADJI Koffi Fabrice certifie que les différents apports, contributions et corrections énoncés au cours de sa soutenance dont le thème est « Vers le développement d'un modèle de prévision des débits à partir des réanalyses ERA5 et LAND à l'aide des méthodes de l'intelligence artificielle dans le bassin versant du Mono. », tenue le 07 Décembre 2024 de 14h à 15h, ont été intégrés au mémoire pour l'amélioration de la qualité scientifique et esthétique de ce dernier.

En foi de quoi, le présent certificat lui est délivré pour servir et valoir ce que de droit.

Président du Jury

Date: / /

Signature :

Examineur

Date: / /

Signature :

Rapporteur

Date: / /

Signature :

DEDICACE

Je dédie ce document à la cour céleste, tout particulièrement à la mémoire de ma mère regrettée,
HOUNNOUVI Agathe, et à mon très cher père AGADJI Houngue

REMERCIEMENT

Nous tenons à remercier DIEU et toutes les personnes sans qui la réalisation de ce travail n'aurait pas été possible.

Notre profonde reconnaissance va notamment à l'endroit de :

- Dr N'TCHA M'PO Yèkambèssoun pour son encadrement précieux, ses conseils avisés et son soutien constant qui ont largement contribué à la réussite de ce travail.,
- M. René BODJRENOU pour son accompagnement et sa disponibilité sans faille tout au long de ce travail,
- Tous les enseignants du Département d'Hydrologie et Gestion des Ressources en Eau de l'Institut National de l'Eau, recevez ma profonde gratitude ; vos efforts ne sont pas vains,
- Tous les Honorables membres du Jury qui accepteront d'apprécier ce modeste travail, merci pour le temps que vous sacrifierez au détriment de vos occupations pour l'analyse et l'amélioration de ce document,
- Tous nos camarades de promotion pour l'ambiance et l'entraide qui ont régné tout au long de notre formation, bonne suite à tous,
- Messieurs Comlan HOUNOUI et Rogatien HOUNKPATIN pour leurs conseils et encouragements,
- Tous les membres de ma famille et autres, pour leur soutien et leur amour inconditionnel.

RESUME

Dans un contexte où le changement climatique exacerbe le risque d'inondations, il devient essentiel d'adopter des stratégies novatrices pour anticiper et gérer ces événements. Ce mémoire porte sur le développement d'un modèle de prévision des débits et l'évaluation des performances des simulations des réanalyses ERA5 et LAND dans le bassin versant du Mono, à l'exutoire d'Athiémé. Pour ce faire, différentes méthodes de correction de biais, telles que le Quantile Mapping et celles d'intelligence artificielle, ont été explorées et comparées. Les résultats montrent que, bien que les réanalyses initiales soient relativement peu fiables, l'application de ces méthodes de correction permet d'améliorer la précision de ces réanalyses. Le modèle stacking a surpassé les méthodes classiques telles que Quantile Mapping et le Random Forest avec un R^2 de 0.84, offrant ainsi des prédictions plus fiables. L'analyse des réanalyses corrigés a également permis de confirmer la stabilité et la cohérence des données, grâce à des tests statistiques. Par ailleurs, l'ajustement des lois de distribution des réanalyses a permis une meilleure estimation des débits de crue, facilitant ainsi l'identification des risques associés aux différentes périodes de retours. Enfin, les réseaux neuronaux récurrents **LSTM** (Long Short-Term Memory), sont développés pour permettre de mieux capturer les dépendances temporelles complexes des réanalyses corrigés par rapport aux observations afin d'améliorer les prévisions à court et à moyen terme des réanalyses. Ce qui se justifie avec une performance R^2 allant de 0.89 à 0.78.

Mots clés : Prévision hydrologique, Modèle LSTM, Réanalyse, Bassin du Mono, Correction de biais, Intelligence artificielle

ABSTRACT

In a context where climate change exacerbates the risk of flooding, it has become essential to adopt innovative strategies to anticipate and manage these events. This thesis focuses on the development of a flow forecasting model and the evaluation of the performance of ERA5 and LAND reanalysis simulations in the Mono River Basin, at the Athiémé outlet. To achieve this, various bias correction methods, such as Quantile Mapping and artificial intelligence techniques, have been explored and compared. The results show that, although the initial reanalysis data are relatively unreliable, the application of these correction methods significantly improves their accuracy. The stacking model outperformed traditional methods like Quantile Mapping and Random Forest, achieving an R^2 of 0.84, thus providing more reliable predictions. The analysis of the corrected reanalysis data also confirmed the stability and consistency of the data, supported by statistical tests. Furthermore, adjusting the distribution laws of the reanalysis data led to better flood discharge estimates, facilitating the identification of risks associated with different return periods. Finally, Long Short-Term Memory (LSTM) recurrent neural networks were developed to better capture the complex temporal dependencies of the corrected reanalysis data in comparison to observations, thereby improving short- and medium-term forecasting performance. This is evidenced by R^2 values ranging from 0.89 to 0.78.

Keywords : Hydrological forecasting, LSTM model, Reanalysis, Mono River Basin, Bias correction, Artificial intelligence

LISTE DES SIGLES ET ACRONYMES

AIC	: Akaike Information Criterion (Critère d'Information d'Akaike)
AGCM	: Atmospheric General Circulation Model (Modèle de circulation Générale atmosphérique)
AR6	: Sixième rapport d'évaluation du GIEC
BIC	: Bayesian Information Criterion (Critère d'Information Bayésien)
CART	: Classification and Regression Trees (Arbres de classification et de Régression)
CEPMET	: Centre Européen pour les Prévisions Météorologiques à Moyen Terme
CRU	: Climate Research Unit
ERA5	: Cinquième génération des réanalyses atmosphériques du CEPMMET
ETR	: Évapotranspiration réelle
GIEC	: Groupe d'experts intergouvernemental sur l'évolution du climat
KGE	: Kling-Gupta Efficiency
LAND	: Version haute résolution des réanalyses ERA5 pour les surfaces Terrestres
L-moments	: Méthode des moments pondérés par la probabilité
MERRA2	: Modern-Era Retrospective analysis for Research and Applications, Version 2
ML	: Machine Learning (Apprentissage automatique)
NARR	: North American Regional Reanalysis (Réanalyse Régionale Nord-Américaine)
NSE	: Nash-Sutcliffe Efficiency (Efficacité de Nash-Sutcliffe)
OMM	: Organisation météorologique mondiale
Q95	: Débit au 95e centile (débit dépassé 95% du temps)
QM	: Quantile Mapping (Méthode de correction par ajustement de quantiles)
R²	: Coefficient de détermination
RBF	: Radial Basis Function (Fonction à base radiale)
RMSE	: Root Mean Square Error (Erreur quadratique moyenne)
SVR	: Support Vector Regression (Régression à vecteurs de support)

WFDE5 : Watch Forcing Data methodology applied to ERA5 reanalysis

LISTE DES TABLEAUX

Tableau 1 : Les différentes fonctions d'activation	26
Tableau 2: Qualité des modèles en fonction des valeurs du NSE	28
Tableau 3: Efficacité des corrections sur ERA5 et LAND à pas de temps journalier.	30
Tableau 4: Les performances des réanalyses empilés après correction	32
Tableau 5 : Statistique descriptive des données d'observation et réanalyses.....	32
Tableau 6 : Résultat du test d'hypothèse.....	33
Tableau 7 : Critères AIC et BIC des débits historiques	34
Tableau 8 : Quantiles des séries de débits de crues à différentes périodes de retour.....	35
Tableau 9 : Résultats des critères déterministes calculés sur les temps de prévisions allant de 1 à 7 jours.	37
Tableau 10 : Résultats des critères probabilistes calculés sur les temps de prévisions allant de 1 à 7 jours suivant les terciles.....	38

LISTE DES FIGURES

<i>Figure 1 : Composante du cycle de l'eau</i>	<i>3</i>
<i>Figure 2 : la carte de la localisation.....</i>	<i>9</i>
<i>Figure 3 : la carte topographique du bassin du mono à Athieme.....</i>	<i>10</i>
<i>Figure 4 : La différence entre la régression linéaire et polynomiale</i>	<i>18</i>
<i>Figure 5 : L'arbre de décisions.....</i>	<i>22</i>
<i>Figure 6: Fonctionnement interne d'un réseau de neurones récurrent</i>	<i>24</i>
<i>Figure 7: Débits observés et débits des réanalyses bruts et corrigés en 2010</i>	<i>31</i>
<i>Figure 8: Nuage de point des débits d'observation et des réanalyses journaliers avant et après correction.</i>	<i>31</i>
<i>Figure 9 : Ajustement graphiques des observations et les réanalyses par la loi de log Pearson III.....</i>	<i>34</i>
<i>Figure 10 : Variabilité des Q95 historiques des réanalyses en fonction des années.....</i>	<i>35</i>
<i>Figure 11 : L'évolution des MSE en fonction des époques à l'entraînement et à la validation.</i>	<i>36</i>
<i>Figure 12 : Courbes ROC pour des Échéances de 1 à 6 Jours</i>	<i>39</i>
<i>Figure 13 : Développement d'un modèle de prévision à chaque échéance.</i>	<i>41</i>

CHAPITRE 1 : INTRODUCTION GENERALE

1.1-Contexte et justification de l'étude

La caractérisation du cycle de l'eau est essentielle pour garantir la sécurité des populations, prévenir les inondations et assurer l'approvisionnement en eau, notamment pour le développement technologique dans des domaines tels que la production d'hydroélectricité. Cette caractérisation repose sur la modélisation des termes du bilan de l'eau, passés et présents, et de leurs projections futures (Levesque, 2022). Les inondations récentes ont provoqué des conséquences socio-économiques et écologiques significatives, avec des pertes humaines et des dégâts matériels. Dans son sixième rapport de synthèse (AR6), le GIEC alerte sur les risques accumulés liés au changement climatique, en particulier pour le continent africain, particulièrement vulnérable à ces impacts. En Afrique de l'Ouest, et plus particulièrement dans les départements du Mono et du Couffo, cette vulnérabilité s'est accumulée, les populations étant de plus en plus exposées aux zones à risque (Amoussou, 2015). La variabilité climatique impacte directement le développement social, économique et environnemental, alors que les stratégies gouvernementales pour y faire face restent souvent insuffisantes (Ahonnon, 2016). Les inondations de juin 2024 en témoignent, car elle a perturbé les activités locales de façon significative. Ainsi, les études hydrologiques sont fondamentales pour établir des stratégies de gestion durable, préservant les ressources en eau et anticipant les risques (Bodjrenou et Comandan, 2023a). Cette gestion nécessite une connaissance approfondie des processus régissant les variations hydriques, obtenue grâce à des modèles hydrologiques basés sur des données précises (OBADA, 2016). Cependant, la qualité des données disponibles reste limitée en raison de la répartition irrégulière des stations de mesure et d'insuffisance de moyens financier (Bodjrenou et al., 2024a). De plus, des facteurs externes, tels que les feux de brousse peuvent endommager les stations de mesures, créant donc des lacunes dans les séries de données (Amoussou et al., 2022). Pour pallier ces difficultés, les réanalyses climatiques, comme ERA5 et LAND, fournissent des données de grande échelle, bien qu'elles aient nécessité souvent des ajustements pour mieux s'adapter aux réalités locales. En parallèle, les avancées en intelligence artificielle (IA) offrent des perspectives inédites pour la modélisation des débits d'eau en capturant les complexes dynamiques temporelles des systèmes aquatiques (Hounnondaho, 2023). Ce travail s'appuie sur deux axes : la correction des biais dans les réanalyses climatiques et l'utilisation de l'IA pour la prévision des débits des réanalyses ERA5 et LAND, afin d'évaluer la capacité des techniques d'IA à fournir des prévisions fiables, adapté

aux besoins locaux. Ce sujet de fin de formation en licence professionnelle d'hydrologie s'intitule ainsi : « **Vers le développement d'un modèle de prévision des débits à partir des réanalyses ERA5 et LAND à l'aide des méthodes de l'intelligence artificielle dans le bassin versant du Mono.** »

1-2 Revue de littérature

Les simulations hydrologiques sont essentielles pour comprendre et gérer les ressources en eau dans les bassins versants. L'avènement de modèles de simulation basés sur des données de réanalyses telles qu'ERA5 et LAND offre de nouvelles opportunités pour évaluer et prédire les variations hydrologiques. Cette revue de littérature examine la pertinence et les défis de l'utilisation d'ERA5 et LAND dans les bassins versants en général et celui du Mono en particulier.

1-2-1 Clarification conceptuelle

L'hydrologie : c'est la science qui étudie le cycle de l'eau et ses interactions entre l'atmosphère, la surface terrestre et le sous-sol.

Un **modèle** est une représentation simplifiée d'un système, utilisée pour comprendre, analyser, ou prédire son comportement. Il peut prendre différentes formes, y compris physiques, conceptuelles, ou mathématiques, et s'appliquer dans de nombreux domaines, tels que la science, l'ingénierie, l'économie et la météorologie. Dans le contexte de l'hydrologie superficielle, la définition et la prévision des écoulements en rivière sont réalisées à l'aide des modèles hydrologiques (Arsenault et al, 2014).

Les modèles hydrologiques sont des représentations mathématiques qui simulent le comportement des systèmes d'eau dans un bassin versant, intégrant des processus tels que le ruissellement, l'infiltration et l'évapotranspiration. Ils sont utilisés pour évaluer les ressources en eau, prévoir les inondations et analyser les impacts environnementaux, en se basant sur des équations physiques et des données empiriques. La figure 1 présente le cycle de l'eau.

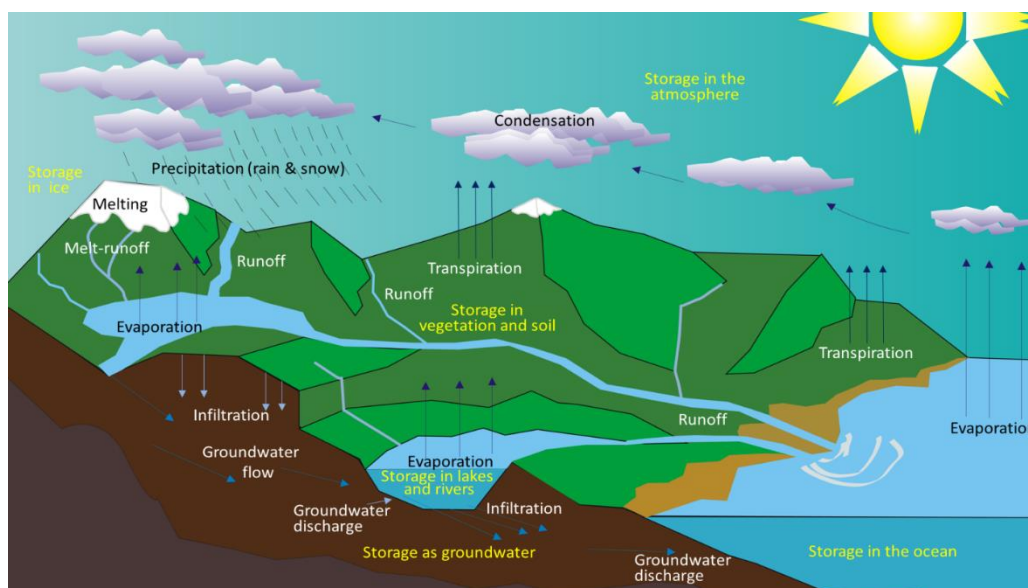


Figure 1 : Composante du cycle de l'eau

Simulation hydrologique : Selon **CEREMA** (Centre d'Études et d'Expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement) la simulation hydrologique représente le résultat final de la modélisation hydrologique (Seiller et al., 2012).

Le CEPMMT : défini comme Centre européen pour les prévisions météorologiques à moyen terme, c'est une organisation intergouvernementale indépendante soutenue par 35 États. Elle fournit des prévisions météorologiques numériques mondiales quatre fois par jour, ainsi que des analyses de la qualité de l'air, de la composition atmosphérique, et des prévisions hydrologiques, entre autres (<https://www.ecmwf.int/en/about/who-we-are>).

Données de réanalyse ERA5 et LAND

La réanalyse est une méthode scientifique permettant de développer un enregistrement complet des changements météorologiques et climatiques au fil du temps. Elle combine des observations et des modèles numériques pour générer une estimation synthétisée de l'état du système terrestre, couvrant généralement plusieurs décennies (<https://reanalyses.org/>). La réanalyse climatique ERA5 comporte deux variantes en son sein. Nous avons la première variante qui a une résolution de $0,25^\circ \times 0,25^\circ$ et la seconde variante plus précise à $0,1^\circ \times 0,1^\circ$ (Land). ERA5 offre des données horaires basées sur l'assimilation d'un large éventail d'observations in situ et télédéteectées dans un modèle de circulation générale atmosphérique (AGCM) (Bodjrénou et al., 2023b ; Reichle et al., 2017).

Analyse fréquentielle

L'analyse fréquentielle est une méthode statistique qui examine des événements passés pour définir les probabilités d'apparition future (Koumassi et al., 2014). Elle est un outil fondamental pour analyser l'occurrence d'événements extrêmes. Cette méthode consiste à rechercher la loi de probabilité la mieux adaptée aux valeurs observées d'une variable aléatoire, telle que la pluie maximale journalière, en comparant différentes lois et méthodes d'estimation (Habibi et al., 2013).

Inondation : c'est la submersion par des eaux douces ou salées d'une zone émergée, qui peut se produire lentement ou rapidement, et qui peut être causée par la crue d'un cours d'eau, des pluies intenses, ou la rupture d'un ouvrage (WWW.futura-sciences.com/planete/définitions/meteorologie-inondation-6305/).

Crue : Une augmentation du niveau d'eau dans les fleuves, lacs, et rivières. Elle est souvent associée à un débit particulièrement élevé ou de faible fréquence (Ababou et Al-Bitar, 2006).

Période de retour : L'intervalle de temps nécessaire pour que l'intensité d'un événement atteigne ou dépasse un certain seuil. Elle est l'inverse de la fréquence et correspond à l'inverse de la probabilité de dépassement de cet événement (F. Métivier, 2010).

Méthodes de correction de biais

L'article de Maraun (2016) présente une revue des techniques de post-traitement, c'est-à-dire la correction du biais. Plusieurs méthodes ont été développées pour corriger les biais des modèles climatiques. Certaines appliquent une correction aux observations, tandis que d'autres modifient directement les simulations. La méthode delta, par exemple, est utilisée pour ajuster les données d'observation en fonction du signal de changement climatique du modèle. En revanche, les méthodes directes, telles que la mise à l'échelle proposée par Widmann et Bretherton (2000), utilisent directement les séries temporelles simulées. Cependant, ces méthodes peuvent ne pas tenir compte des biais affectant la variance et l'aplatissement des données. Une méthode plus flexible est l'ajustement par quantiles (Quantile Mapping), qui ajuste la fonction de répartition de la précipitation simulée pour mieux correspondre aux observations (Panofsky et Brier, 1968).

L'Intelligence artificielle

L'intelligence artificielle (IA) est un domaine de l'informatique qui vise à créer des systèmes capables de simuler des comportements intelligents, semblables à ceux de l'humain, tels que la reconnaissance de formes, la prise de décision et l'apprentissage. L'IA s'appuie sur des

algorithmes et des modèles pour traiter des données massives et tirer des conclusions ou faire des prédictions, souvent de manière autonome. Elle est utilisée dans des domaines variés comme la santé, les finances, la logistique et bien d'autres. Les avancées en IA, notamment en apprentissage automatique et en réseaux de neurones, permettent aujourd'hui des innovations impressionnantes, tout en soulevant des questions éthiques et de sécurité.

1-2-2 Synthèse bibliographique

Historiquement, l'utilisation des données de réanalyses pour la modélisation hydrologique était peu répandue, souvent limitée à des études à petite échelle en raison de l'incertitude et des biais associés. Par exemple, Choi et al. (2009) ont étudié la réanalyse NARR sur trois bassins versants du Manitoba, trouvant que les données de sortie étaient peu fiables sans correction des biais. Cependant, des études plus récentes montrent un meilleur potentiel des données de réanalyses. Essou et al. (2016, 2017) indiquent que, bien que les observations soient généralement plus performantes, les débits simulés par ERA5 peuvent être satisfaisants, en particulier dans les bassins versants avec moins d'une station par 1000 km². Une étude récente de Bodjrenou (2023) a évalué les produits de réanalyse ERA5, WFDE5 et MERRA2 au Bénin, révélant que ERA5 et WFDE5 reproduisent mieux le régime pluviométrique bimodal, tandis que MERRA2 montre des biais significatifs. Ces travaux soulignent que, malgré les recommandations d'utiliser des données d'observation, ERA5 reste une alternative valable dans les régions où les données de qualité sont rares (Levesque, 2022).

2. But de la recherche

Le but de ce mémoire est d'évaluer la capacité prédictive des techniques d'intelligence artificielle dans le domaine hydrologique, en vue d'améliorer la précision des prévisions hydrologiques.

3. Objectifs du mémoire

3.1 Objectif général

L'objectif général de ce mémoire est d'évaluer la performance des simulations hydrologiques des réanalyses ERA5 et LAND dans le bassin versant du Mono à l'exutoire d'Athiémé, en corrigeant ces données et en développant un modèle de prévision des débits à l'aide de techniques d'intelligence artificielle.

3.2 Objectifs spécifiques

Pour atteindre l'objectif général de l'étude, plusieurs objectifs spécifiques sont définis :

- Évaluer la performance des simulations hydrologiques d'ERA5 et de LAND dans le bassin du Mono à l'exutoire d'Athiémé.
- Corriger les données des simulations hydrologiques d'ERA5 et de LAND dans le bassin du Mono à l'exutoire d'Athiémé
- Faire une analyse fréquentielle des Q95 à partir des réanalyses corrigées.
- Développer un modèle prédictif des débits journaliers.

4. Questions de recherche

La présente étude répond aux trois questions de recherche présentées ci-dessous.

- Quelle est la qualité des débits des réanalyses ERA5 et LAND dans le bassin versant du Mono à l'exutoire d'Athiémé ?
- Quelle méthode de correction est-elle adaptée pour réduire les biais des réanalyses ERA5 et LAND dans le bassin versant du Mono à l'exutoire d'Athiémé ?
- Comment les méthodes de correction de biais influencent-elles l'analyse fréquentielle des débits Q95 à partir des données de réanalyse corrigées dans le bassin du Mono à Athiémé ?
- Quel modèle fournit les prédictions les plus précises pour les débits journaliers du bassin du Mono à l'exutoire d'Athiémé ?

5. Les hypothèses de recherche

- Les simulations hydrologiques d'ERA5 et de LAND présentent des écarts significatifs par rapport aux observations.
- Les méthodes de correction des biais, amélioreront la précision des réanalyses ERA5 et LAND pour le bassin du Mono à l'exutoire d'Athiémé.
- Les réanalyses corrigées permettront une meilleure estimation des quantiles Q95.
- Un modèle prédictif basé sur des techniques d'intelligence artificielle peut fournir des prévisions de débits journaliers plus précises.

6. Structuration du mémoire

Le présent mémoire est structuré en trois (03) chapitres (mis à part la conclusion générale et les perspectives).

- Le premier chapitre est celui de l'introduction générale, décrivant le contexte et la justification de l'étude ainsi que la revue de littérature.

- Le deuxième chapitre traite des données, du matériel et des méthodes. Il est composé de six sections principales, incluant la présentation de la zone d'étude, la description des données utilisées, les logiciels employés et, surtout, la méthodologie adoptée.
- Le troisième chapitre est consacré aux résultats et aux discussions, avec quatre (04) sous-sections, dont trois pour présenter les résultats obtenus et la dernière pour les discussions.

CHAPITRE 2 : DONNEES, MATERIELS, METHODES ET SECTEUR D'ETUDE

Ce chapitre présente les bases de données utilisées, ainsi que les caractéristiques physiographiques et géographiques du secteur d'étude, les matériels et les méthodes employés.

2.1 Présentation du milieu d'étude

Dans cette section, nous présenterons la zone d'étude, sa localisation géographique, son relief, son sol, sa végétation, son climat, son hydrographie, ainsi que les aspects socio-économiques et démographiques.

2.1.1 Localisation géographique

Le bassin-versant du Mono situé dans la région du golfe de Guinée est partagé entre le Bénin et le Togo. Les parties hautes et moyennes du bassin sont en grande partie sur le territoire togolais (figure 2). Le complexe s'étend entre 06°16' et 09°20'N et 0°42' et 2°25'E et il couvre une superficie de 27 870 km². Le bassin porte le fleuve Mono sur 560 km avant d'aboutir à l'océan Atlantique par le biais du système lagunaire. D'orientation nord-sud, le bassin s'est formé essentiellement lors du chevauchement de formations des unités externes des Dahoméyides (couverture sédimentaire, roches métamorphiques et le socle granitique) sur celles de l'unité de la plaine bénino-togolaise. (Amoussou, 2010). Le bassin du Mono abrite un barrage hydroélectrique, le Nangbéto, situé au Togo.

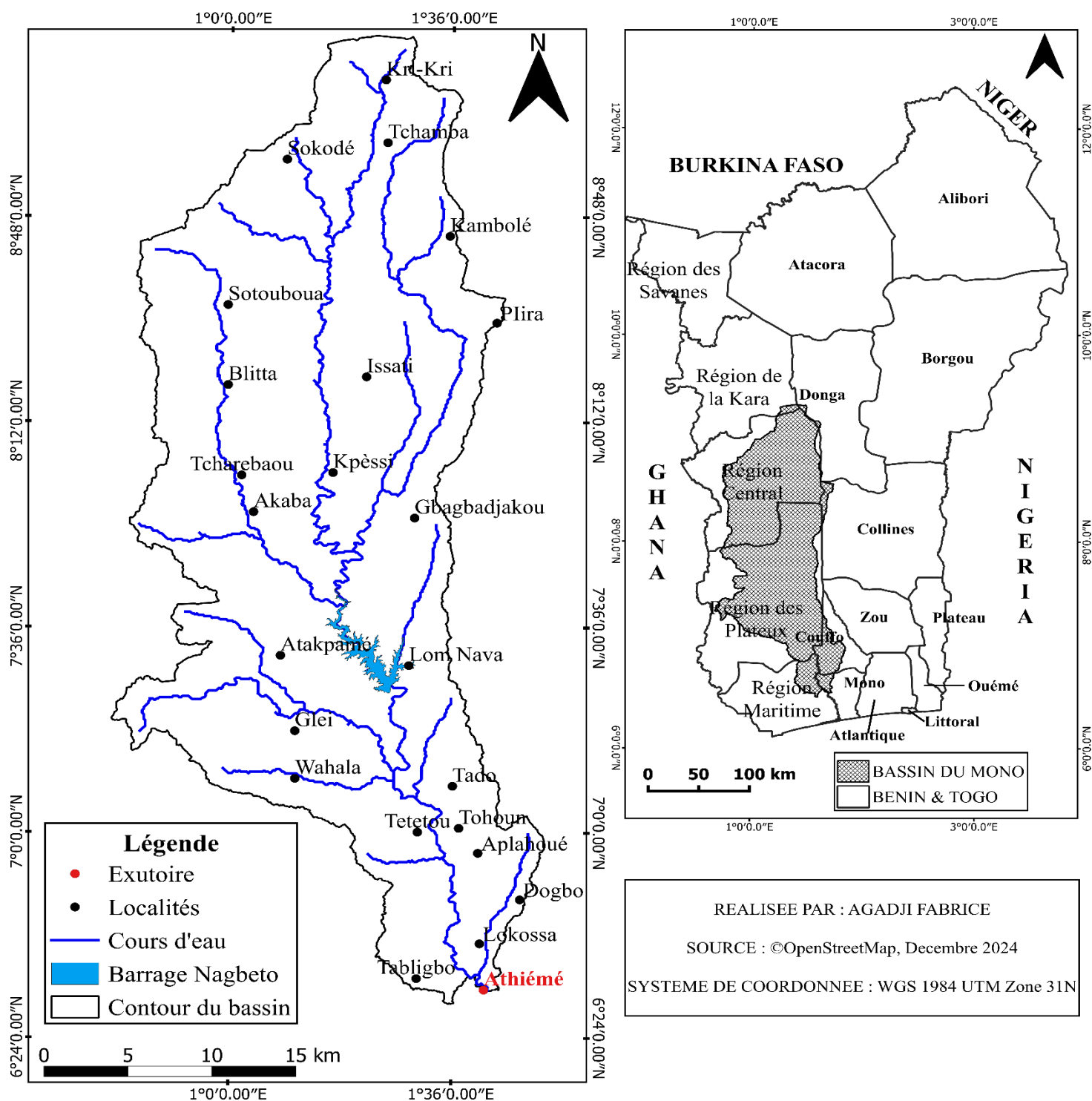


Figure 2 : la carte de la localisation

2.1.2 Relief

Le bassin-versant comporte un bassin sédimentaire côtier au sud, modelé en plaine littorale et plateaux et des plus hauts reliefs au nord comprenant les monts Atacora et leurs prolongements méridionaux, les monts du Togo. Ces reliefs sont disséqués en plusieurs massifs, dont les crêtes orientées SSW-NNE, culminant entre 600 et 941 m à Atilakoutsé, constituent un véritable

château d'eau. Celui-ci alimente plusieurs cours d'eau à écoulement sub-méridien à méridien (Klassou, 1996). Cependant, la plus grande partie du bassin-versant, est à basse altitude (30 à 250 m). Le bassin inférieur du complexe a une topographie adoucie du fait de son recouvrement partiel par les alluvions sablo-argileuses. Il présente toutefois les différents paliers qui marquent sans doute des niveaux de reprise d'érosion.

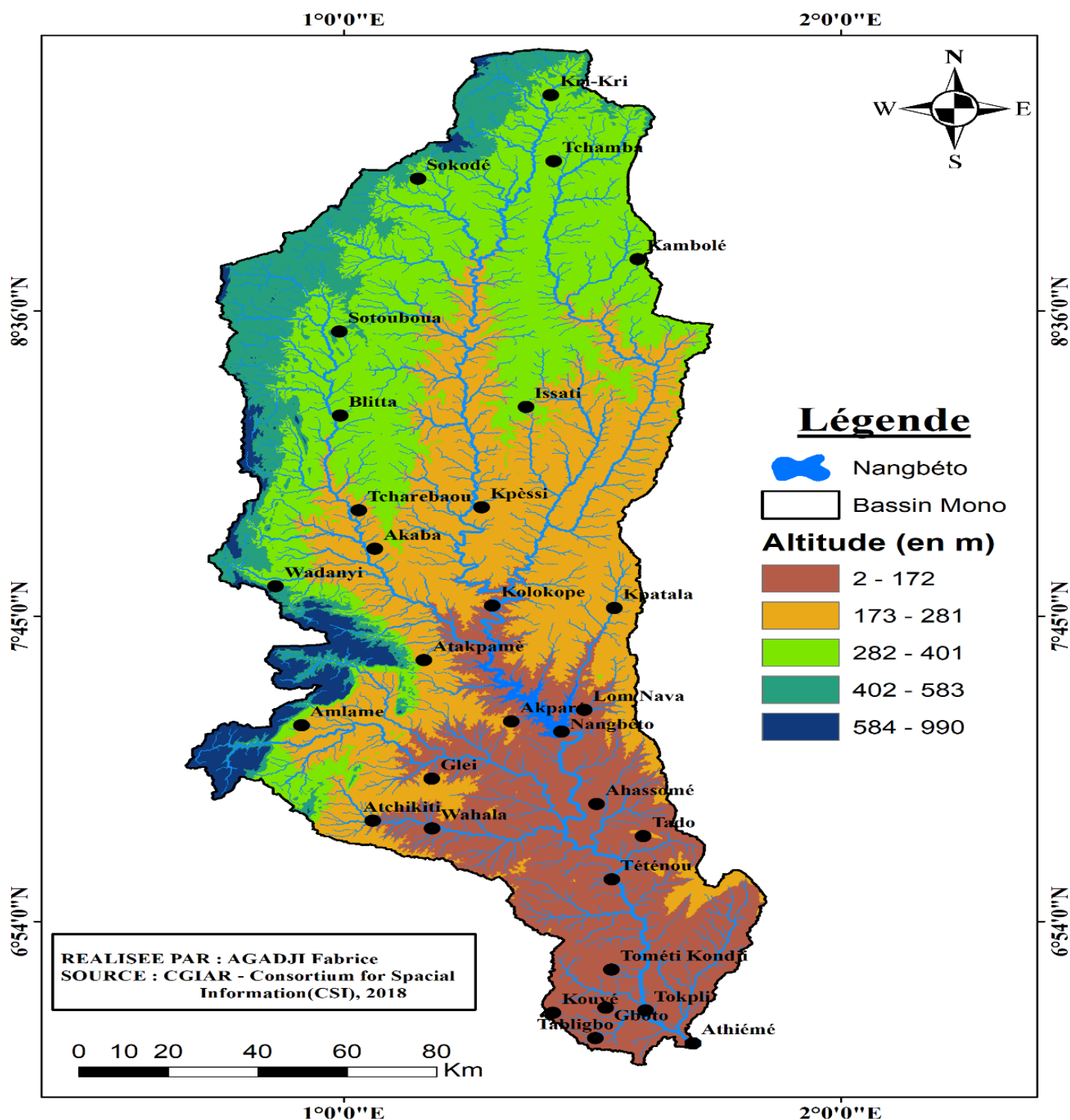


Figure 3 : la carte topographique du bassin du mono à Athiémé

2.1.3 Végétation et sol

Le bassin du Mono est couvert de la savane herbeuse, les formations marécageuses et quelques mangroves. La végétation est dominée par le palmier à huile et les graminées, mais on y trouve également quelques reliques de forêts galeries. Les sols, de type alluvial et colluvial, sont plus ou moins hydromorphes, riches en matière organique et très fertiles, mais inondés de façon saisonnière par les crues du fleuve Mono. Les sédiments sont dans l'ensemble fins, avec des sables et graviers observés dans les chenaux, et des limons et vases dans les dépressions et plaines d'inondation. On rencontre également des sols de type sableux, peu fertiles et aptes aux cocotiers (Gbèyétin, 2014).

2.1.4 Climat et réseaux hydrographiques

Ce bassin, soumis à la mousson ouest-africaine, s'étend sur deux domaines climatiques : au Sud, un climat subéquatorial à régime bimodal (deux saisons sèches et deux saisons pluvieuses alternées), et au centre et au Nord, un climat tropical à régime unimodal (une saison sèche et une saison pluvieuse). Cette diversité au sein d'un espace de petite taille, associée à l'évolution climatique marquée des dernières années, notamment la construction de la retenue d'eau de Nangbéto, est susceptible d'avoir des effets complexes sur les régimes hydrologiques. Le bassin du Mono-Couffo bénéficie d'une pluviométrie moyenne annuelle de 1100 mm, et la température moyenne annuelle est de l'ordre de 27°C (Amoussou et al., 2010).

2.1.5 Démographie et activités socio-économiques

Le bassin du Mono couvre, dans la partie Bénin, les départements du Couffo et du Mono. Selon les données du recensement de la population béninoise de 2013 (RPGH4), élaboré par l'INSAE, la croissance démographique est significative, avec un doublement de la population du Mono (281 245 à 497 243 habitants) et un triplement de celle du Couffo (395 132 à 745 328 habitants) entre 1992 et 2013.

Les facteurs humains responsables de la survenance des inondations dans la basse vallée du Mono incluent la croissance démographique, l'évolution horizontale des villes, le manque d'un système urbain de drainage des eaux de pluie, les comblements des marécages et l'ouverture de nouvelles pistes ou dessertes rurales après recharge des espaces humides. La basse vallée est essentiellement peuplée des ethnies comme Ouatchi, Kotafon, Aïzo, Adja, Mina, Xla et minoritairement de Yoruba et Nago. Les activités principales sont l'agriculture, la pêche, l'élevage, la transformation des produits agricoles, l'artisanat et le commerce.

2.2 Données utilisées

Les données d'observation de débit utilisées proviennent de la base de données de la Direction Générale de l'Eau (DGEau). En ce qui concerne les données de réanalyses, nous avons utilisé les données hydrologiques fournies par le CEPMMT, qui est le Centre européen pour les prévisions météorologiques à moyen terme.

❖ **Prétraitement des données :**

Au cours du processus d'apprentissage automatique, il est essentiel de prétraiter les données en utilisant des techniques qui améliorent leur distribution pour les algorithmes d'apprentissage. Le prétraitement comprendra les étapes suivantes :

Nettoyage des données : Suppression des bruits et incohérences pour assurer l'intégrité des données d'entrée. Ceci se fait par une exploration des données pour détecter ces incohérences.

Gestion des valeurs aberrantes : Les valeurs aberrantes sont des données qui s'écartent significativement de la tendance générale des autres données et peuvent être dues à des erreurs de mesure ou à des variations réelles. Les méthodes statistiques pour détecter les valeurs aberrantes incluent l'utilisation de diagrammes en boîte (boxplots), qui graphiquement présentent la distribution des données par quartiles et permettent d'identifier les points qui se situent en dehors des moustaches du boxplots comme des valeurs potentiellement aberrantes.

Normalisation : Elle consiste à ramener toutes les variables de données à la même échelle tout en préservant la forme de la distribution de chaque variable et les rendant facilement comparables.

2.3 Matériels utilisés

- Word 2013 pour la rédaction du document
- Excel 2013 pour les premières visualisations des données
- ArcGIS et Q-GIS pour la réalisation des cartes
- Hyfran pour l'analyse fréquentielle
- RStudio et Python (PyCharm et Jupyter Notebook) pour le traitement et l'analyse des données
- Zotero pour la gestion des références

2-4 Méthodes et critères d'évaluation de la performance des simulations

2-4-1 Critères d'évaluation de la performance des simulations

2-4-1-1 Critères déterministes

2-4-1-1-1 RMSE (Root Mean Squared Error)

Le RMSE est un critère d'évaluation largement utilisé pour mesurer la précision d'un modèle de prévision, de régression ou de classification. Il est couramment utilisé dans divers domaines tels que l'apprentissage automatique, la statistique, la météorologie et l'économétrie. Le RMSE est utilisé pour quantifier l'écart moyen entre les valeurs prédites par un modèle et les valeurs réelles observées. Il est calculé en prenant la racine carrée de la moyenne des carrés des écarts entre les valeurs prédites (ou estimées) et les valeurs réelles.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{obs} - Q_{sim})^2}{n}}$$

Avec Q_{obs} le débit observé, Q_{sim} le débit simulé et n le nombre de donnée

- **RMSE faible** : Indique une meilleure précision du modèle, car cela signifie que l'écart moyen entre les valeurs prédites et les valeurs réelles est plus petit.
- **RMSE élevé** : Indique une moins bonne précision du modèle, car les écarts entre les valeurs prédites et réelles sont plus grands.

2-4-1-1-2 Nash-Sutcliffe Efficiency (NSE)

L'un des critères largement utilisés en hydrologie est le coefficient d'efficacité de Nash-Sutcliffe (Nash et Sutcliffe, 1970). C'est un indicateur de performance construit à partir de la normalisation du MSE, dont les valeurs sont comprises dans l'intervalle $]-\infty ; 1]$. Il permet d'estimer la capacité d'un modèle à reproduire un comportement observé. Il est défini comme suit :

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{obsi} - Q_{simi})^2}{\sum_{i=1}^n (Q_{obsi} - \overline{Q_{obs}})^2}$$

Avec Q_{obsi} le débit observé, Q_{sim} le débit simulé, n le nombre de donnée et $\overline{Q_{obs}}$ la moyenne des débits observés

Un NSE de 1 indique que les données simulées correspondent exactement aux données observées. Un NSE de zéro indique que les données simulées sont aussi précises que la moyenne des données observées, tandis qu'un NSE inférieur à zéro signifie que la valeur moyenne des données observées est un meilleur estimateur que les données simulées.

2-4-1-1-3 Le critère de Kling-Gupta (KGE)

Le critère de Kling-Gupta (KGE), formulé par Kling et Gupta en 2009, est une mesure de performance utilisée pour évaluer l'efficacité des modèles hydrologiques. Développé par Gupta, Kling, Yilmaz et Martinez, le KGE combine plusieurs aspects statistiques pour fournir une évaluation globale de la performance du modèle. Le KGE s'exprime comme suit :

$$KGE = 1 - DE \text{ Avec } DE = \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$$

r : le coefficient de corrélation de Pearson entre la simulation et l'observation
 α : σ simulé / σ observé : le coefficient qui caractérise la variabilité de la simulation par rapport à l'observation
 β : μ simulé / μ observé : le rapport entre la moyenne de la simulation et la moyenne de l'observation. Interprétation KGE =1 : La simulation correspond parfaitement aux observations. KGE <1 : La simulation dévie des observations, avec des composantes spécifiques (r , α , β) indiquant où se situe la déviation. KGE proche de 0 : Les simulations ne sont pas meilleures qu'une prédiction utilisant la moyenne des observations. KGE < 0 : Les simulations sont moins fiables qu'une prédiction utilisant la moyenne des observations.

L'objectif du KGE est d'atteindre des valeurs optimales des composantes r , α et β , ce qui minimise la distance euclidienne (DE) vers zéro, maximisant ainsi le KGE.

2-4-1-1-4 Coefficient de détermination (R^2)

Le coefficient de détermination (carré du coefficient de corrélation de Pearson) est habituellement utilisé pour évaluer la performance des modèles hydrogéologiques et hydrologiques (LEGATES, 1999). Il est obtenu en calculant la régression linéaire entre les valeurs (débits) calculées et les valeurs (débits) observées ou mesurées. Il représente la proportion de la variance d'une variable dépendante qui s'explique par une ou plusieurs variables indépendantes dans le modèle de régression. Il est donné par la relation

$$R^2 = 1 - \frac{\sum_{i=1}^n (Q_i - Q_i')^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2}$$

R^2 Le coefficient de détermination, Q_i est la valeur mesurée du débit, Q_i' est le débit calculé par le modèle, $\bar{Q}$ est le débit moyen mesuré et n est le nombre d'observations de l'ensemble de calage.

2-4-1-1-5 Pourcentage de biais (pbiais)

Le "PBIAS" (Pourcentage de Biais) est un critère d'évaluation utilisé pour mesurer le biais systématique d'un modèle de prévision ou de régression par rapport aux observations réelles. Il

est souvent utilisé dans le domaine de l'hydrologie et de la modélisation hydrologique, mais peut également être appliqué à d'autres domaines où l'évaluation des biais est essentielle.

Le pbias mesure la tendance moyenne des valeurs simulées à être plus grandes ou plus petites que celles observées. Le PBIAS exprime le biais moyen en pourcentage par rapport aux valeurs réelles. Un PBIAS de 0% indique que le modèle présente un biais nul, c'est-à-dire que les valeurs prédites sont en moyenne égales aux valeurs réelles. Un PBIAS positif indique un biais de surestimation, tandis qu'un PBIAS négatif indique un biais de sous-estimation.

$$PBIAS = \frac{\sum (Obs - Prev)}{\sum Prev} \times 100$$

2-4-1-1-6 Corrélation de Pearson

Elle mesure la force et la direction de la relation linéaire entre deux variables quantitatives. Elle est représentée par le coefficient (**cor**), qui varie entre -1 et 1 : un coefficient de 1 indique une corrélation positive parfaite, -1 une corrélation négative parfaite, et 0 l'absence de corrélation. Le coefficient est calculé par la formule suivante :

$$Cor = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

Où X_i et Y_i représentent les deux variables, individuelles, $\bar{Y}$ et $\bar{X}$ sont les moyennes des variables respectives.

1-4-1-2 Critères probabilistes

- **Le tableau de contingence**

Le **tableau de contingence** est un outil statistique essentiel pour évaluer les performances des prévisions, notamment dans le cadre d'événements météorologiques extrêmes comme les pluies intenses. Il confronte les prévisions d'un modèle aux observations réelles et permet de calculer divers scores probabilistes. Les quatre catégories qui composent le tableau sont :

1. **Vrais positifs (VP)** : Le modèle prévoit correctement un débit, qui se produit effectivement.
2. **Faux positifs (FP)** : Le modèle prédit un débit qui ne se matérialise pas.
3. **Vrais négatifs (VN)** : L'absence d'un débit est correctement prévue.
4. **Faux négatifs (FN)** : Le modèle n'anticipe pas un débit qui survient.

1-4-1-2-1 Rapport des vraies Alertes ou la précision ou le taux des prédictions positives

La **précision** mesure la capacité du modèle à faire des prédictions positives correctes. Elle indique, parmi toutes les prédictions positives faites par le modèle, combien étaient effectivement correctes.

$$Pr = \frac{VP}{(FP+VP)}$$

1-4-1-2-2 Taux de succès (HR) ou le recall

Le **recall** mesure la capacité du modèle à détecter tous les cas réellement positifs. Il est important, parmi toutes les prédictions positives réelles, combien ont été correctement identifiées. Un recall élevé signifie que le modèle ne manque presque aucun cas positif, mais il peut parfois prédire des faux positifs en contrepartie.

$$HR = \frac{VP}{VP+FN}$$

1-4-1-2-3. Score F1

Le **F1-score** est une mesure qui combine la précision et le rappel en une seule métrique harmonisée. Il est particulièrement utile lorsque les données sont déséquilibrées, car il équilibre l'importance entre le nombre de vrais positifs correctement prédits (rappel) et l'exactitude des prédictions positives (précision). Un F1-score élevé indique que le modèle est à la fois précis et capable de détecter la majorité des cas positifs.

$$F1 = \frac{2 * HR * Pr}{HR + Pr}$$

1-4-1-3 Évaluation de la performance par la Courbe ROC

La courbe ROC (Receiver Operating Characteristic) illustre graphiquement la relation entre la sensibilité (taux de vrais positifs) et la spécificité (1 - taux de faux positifs) d'un test ou d'un modèle prédictif. Elle est utilisée pour évaluer la performance des prévisions corrigées, en mettant en évidence les compromis entre les taux de réussite et de fausses alarmes pour différents seuils de décision (MetOffice, n.d.).

Un bon système de prévision se caractérise par une courbe ROC inclinée vers le coin supérieur gauche du graphique, indiquant un taux de réussite élevé et un faible taux de fausses alarmes. En revanche, un système de prévision sans compétence produira une courbe suivant la diagonale, où les taux de réussite et de fausses alarmes sont égaux.

Le score ROC, correspondant à l'aire sous la courbe (AUC, pour "Area Under the Curve"), fournit une mesure quantitative de la performance. Une prévision parfaite obtient une AUC de

1,0, où toutes les occurrences sont correctement prédites, tandis qu'une prévision sans compétence atteint une AUC proche de 0,5. Ainsi, le score ROC permet de comparer différentes approches et d'évaluer leur capacité à prédire les événements de manière fiable (Mason & Graham, 1999).

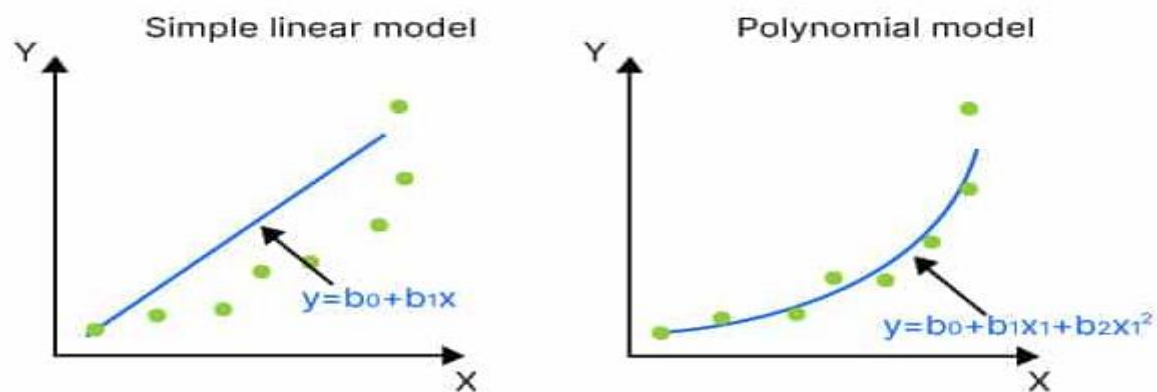
2-5-2 Les méthodes de corrections

2-5-2-1 La régression linéaire

Dans le modèle de régression linéaire simple, « x » désigne la ou les variables indépendantes, la covariable ou les prédicteurs, et « y » désigne la variable dépendante ou la réponse. Dans la formule $Y = \beta_0 + \beta_1 X$, la régression linéaire simple calcule les coefficients β_0 et β_1 d'un modèle linéaire qui prédit la valeur d'un « y » par rapport à une variable « x ». https://www.researchgate.net/publication/275018715_Statistical_Hypothesis_Testing.

2-5-2-2 Régression polynomiale

La régression polynomiale est souvent considérée comme une régression linéaire multiple spéciale. La régression polynomiale est une méthode statistique permettant de déterminer la relation entre une variable indépendante (x) et une variable dépendante (y) et de modéliser leur relation comme le polynôme de nième degré. La relation de la variable indépendante et dépendante sur un graphe s'avère être une relation curviligne à l'aide d'une équation polynomiale. La régression polynomiale est utilisée lorsqu'il n'y a pas de corrélation linéaire entre les variables. Par conséquent, cela explique pourquoi il ressemble plus à une fonction non linéaire. L'équation de régression polynomiale de nième degré peut s'écrire comme suit : $Y = b_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n$. Le problème avec la régression linéaire était qu'elle utilise la droite du meilleur ajustement. En d'autres termes, lorsque nous avons un ensemble de données et que nous le traçons sur un graphique, il doit y avoir une droite où se trouvent les nuages de points. Mais que se passe-t-il si nous avons un ensemble de données qui ne nous donne pas de droite mais une courbe ? C'est à ce moment-là que la régression polynomiale entre en jeu. Ainsi, lorsque la régression linéaire ne parvient pas à déterminer une relation linéaire entre les variables, la régression polynomiale le fait pour nous. <https://www.voxco.com/fr/blog/regression-polynomiale-tout-ce-que-vous-devez-savoir/>



2-5-2-3 Méthode Quantile mapping

La méthode de correction par quantiles consiste à faire correspondre la fonction de répartition de la variable aléatoire à celle de la variable aléatoire de référence. Dans une variante largement utilisée, un quantile de la distribution d'une variable explicative biaisée pour une période de calibration est remplacé par le même quantile de la distribution de référence pour la même période de calibration. Cette méthode se présente sous différentes formes :

2-5-2-3-1 Méthode Quantile mapping PTF une transformation exponentielle asymptotique.

La méthode de *Quantile Mapping PTF* (Parametric Transformation Function) avec une transformation exponentielle asymptotique est utilisée pour approcher asymptotiquement une distribution cible. La transformation est définie comme :

$$P_{ob} = (a + b \cdot P_m) \cdot (1 - \exp(-P_m / \tau))$$

Dans cette équation : Le terme linéaire $4a + b \cdot P_m$ ajuste l'échelle et le décalage des quantiles. Le terme exponentiel $(1 - \exp(-P_m / \tau))$ introduit une asymptote, garantissant que la transformation se stabilise progressivement pour des valeurs plus élevées de P_m , Où P_{ob} représente les quantiles observés, P_m les quantiles modélisés, et a , b , et τ sont des paramètres à ajuster. Cette transformation ajuste les données modélisées pour refléter la distribution observée par mise à l'échelle et translation, le terme exponentiel contrôlant le comportement asymptotique. Elle est particulièrement utile pour capturer des valeurs extrêmes tout en lisant la tendance centrale des données. <https://rdrr.io/cran/qmap/man/fitqmapptf.html>.

2-5-2-3-2. Méthode Quantile mapping PTF avec une transformation personnalisée

La **transformation « myff »** est une fonction de puissance personnalisable qui permet d'ajuster la distribution à l'aide d'une loi de puissance simple et qui est définie par l'utilisateur :

$$P_{ob} = a \cdot P_m^b$$

Avec **P_{ob}** est le quantile observé, **P_m** est le quantile modélisé, a et b sont des paramètres ajustés aux données.

Les paramètres sont généralement ajustés en minimisant une fonction de coût telle que la somme résiduelle des carrés (RSS) ou l'erreur absolue moyenne (MAE). La transformation « myff » est très flexible, permettant d'ajuster la forme et l'échelle de la distribution globale, ce qui la rend utile pour divers types d'ensembles de données où la relation entre les données modélisées et observées n'est pas linéaire. <https://rdr.io/cran/qmap/man/fitqmapptf.html>.

2-5-2-3-3. Méthode Quantile mapping DIST avec une distribution de Bernoulli- Gamma

La **transformation Bernoulli-Gamma** est utilisée pour gérer des ensembles de données contenant un mélange de valeurs nulles et non nulles. Elle modélise les zéros à l'aide d'une distribution de Bernoulli et les valeurs non nulles à l'aide d'une distribution Gamma. La fonction de densité de probabilité (PDF) pour les valeurs non nulles est :

$$g(x) = \begin{cases} \pi \cdot \gamma(x) & \text{si } x > 0 \\ 1 - \pi & \text{si } x = 0 \end{cases}$$

Où : π est la probabilité d'un événement, $\gamma(x)$ est la fonction de distribution Gamma. La distribution Gamma est définie par sa forme α et l'échelle β paramètres :

$$\gamma(x) = \frac{x^{\alpha-1} \cdot \exp(-x/\beta)}{\beta^{\alpha} \Gamma(\alpha)}$$

La composante de Bernoulli attribue une probabilité $1-\pi$ pour les zéros. Pour les valeurs supérieures à zéro, la distribution suit une forme Gamma. Cette méthode est particulièrement efficace pour les données de précipitations, où il y a de nombreux jours sans précipitations (valeurs nulles) et quelques jours avec des précipitations importantes (valeurs non nulles). La distribution Bernoulli-Gamma peut bien capturer ce double comportement. Cette distribution Bernoulli-Gamma avec la fonction `fitqmapdist()`, similaire à la méthode Bernoulli-Weibull. Cette méthode dans R, s'utilise avec la fonction `fitqmapdist()` dans le package `qmap`, qui permet d'effectuer un ajustement basé sur des distributions paramétriques, y compris Bernoulli-Weibull. Les fonctions comme `dbernweibull()`, `pbernweibull()`, et `qbernweibull()` permettent de

calculer respectivement la densité, la fonction de répartition, et les quantiles de la distribution Bernoulli-Weibull. <https://rdr.io/cran/qmap/man/berngamma.html>

2-5-2-3-4-Méthode Robuste Quantile mapping avec transformation "linéaire"

La méthode **RQUANT** améliore quantile mapping de base en utilisant une régression linéaire locale robuste pour estimer la relation quantile-quantile (QQ). Au lieu de simplement interpoler entre les quantiles, RQUANT ajuste une régression linéaire locale des moindres carrés à chaque quantile, en utilisant des points de données proches pour améliorer l'ajustement.

1. Les données modélisées et observées sont d'abord divisées en quantiles (par exemple, tous les 0,01 percentiles).
2. Pour chaque quantile, la méthode identifie les points de données les plus proches et ajuste une ligne de régression linéaire locale à ces points.
3. La régression linéaire locale est utilisée pour estimer la valeur du quantile observé correspondant à chaque quantile modélisé.

Cette méthode peut gérer les irrégularités et résiste aux valeurs aberrantes. Elle utilise également l'amorçage (c'est-à-dire l'échantillonnage avec remplacement) pour générer plusieurs estimations de la transformation, qui sont ensuite moyennées pour produire des résultats robustes. `fitQmapRQUANT` estime les valeurs de la relation quantile-quantile des séries temporelles observées et modélisées pour des quantiles régulièrement espacés en utilisant la régression linéaire locale des moindres carrés. `doQmapRQUANT` effectue un mappage quantile en interpolant les quantiles empiriques. <https://rdr.io/cran/qmap/man/fitqmaprquant.html>

2-5-2-3-5-Quantile mapping avec une transformation spline

La méthode **SSPLIN** utilise des splines de lissage pour créer une transformation spline lisse entre les quantiles observés et modélisés. Une spline de lissage est une courbe flexible qui s'adapte au tracé quantile-quantile et minimise l'erreur résiduelle tout en maintenant la fluidité

$$\mathbf{P}_{ob} = \mathbf{S}(\mathbf{P}_m)$$
 Où : $\mathbf{P}_{ob}$ est le quantile transformé (observé), $\mathbf{P}_m$ est le quantile modélisé et $\mathbf{S}(\mathbf{P}_m)$ est la spline de lissage ajusté au tracé quantile-quantile. La courbe spline est construite en ajustant une fonction aux quantiles qui équilibre la précision de l'ajustement (minimisation des erreurs) et la régularité (réduction des courbures prononcées). Le degré de régularité peut être contrôlé par un paramètre (par exemple, le facteur de lissage), qui ajuste le compromis entre la flexibilité de l'ajustement et sa régularité. Cette méthode est particulièrement utile lorsque la relation quantile-quantile est non linéaire et

complexe, car elle permet de capturer des changements subtils dans la distribution sans surajustement au bruit dans les données. Elle est souvent appliquée dans des scénarios où des transitions continues et fluides entre les quantiles sont essentielles, comme les données hydrologiques avec des variations saisonnières graduelles.
<https://rdr.io/cran/qmap/man/fitqmaprquant.html>

2-5-2-4 Random forest

Le Random forest est un algorithme composé de différents arbres de décision, chacun avec les mêmes nœuds, mais utilisant des données différentes qui mènent à des feuilles différentes. Il fusionne les décisions de plusieurs arbres de décision afin de trouver une réponse, qui représente la moyenne de tous ces arbres de décision. Théorie mathématique derrière les forêts aléatoires : **Construction d'arbres** : les forêts aléatoires créent plusieurs arbres de décision. Chaque arbre est construit en sélectionnant un sous-ensemble aléatoire de données (bootstrapping) et de caractéristiques (méthode du sous-espace aléatoire). L'objectif est de réduire le biais entre les arbres tout en conservant leur pouvoir prédictif. Chaque arbre est associé à une série de divisions ou « nœuds », qui partitionnent les données en fonction d'un critère **la réduction de la variance**. Critère pour la **Régression** : Pour la régression, le critère utilisé est la réduction de la variance. Soient $C_1, C_2, \dots, C_n$ le nombre de sous échantillon, et $\text{Var} (C)$ représente la variance du jeu de données. Elle est donc calculée par la formule suivante

$$\Delta \text{Var} = \left\{ \text{Var} (C) - \left(\frac{|C_1|}{|C|} \text{Var} (C_1) + \dots + \frac{|C_n|}{|C|} \text{Var} (C_n) \right) \right\}$$

Soit $T_1, T_2, \dots, T_m$ arbres de décision générés dans une forêt. Pour une nouvelle observation x , la prédiction p de la forêt pour un problème de régression est la moyenne des prédictions des arbres : <https://tikz.net/random-forest/>

$$p = \frac{1}{Tm} * \sum_{b=0}^m Tm(x)$$

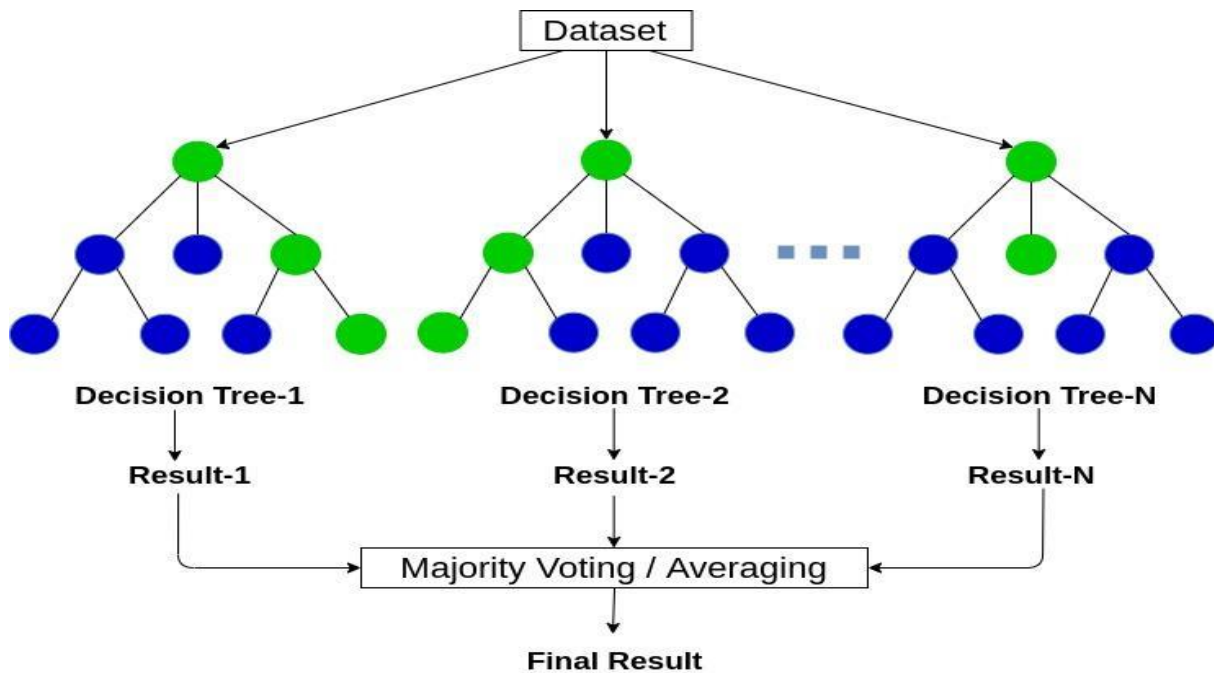


Figure 5 : L'arbre de décisions

2-5-2-5 Support Vector Regression (SVR)

Le Support Vector Regression (SVR) est une version de la Support Vector Machine (SVM) adaptée aux tâches de régression. Contrairement à la classification, où SVM cherche à trouver un hyperplan optimal pour séparer les classes, SVR vise à ajuster un modèle qui minimise l'erreur de prédiction tout en restant autour de la fonction cible. Les principes fondamentaux du SVR sont présentés dans les lignes suivantes.

Marge d'Insensibilité ϵ : L'objectif principal de SVR est de trouver une fonction qui a une déviation d'au plus ϵ des valeurs cibles, tout en maximisant la marge (similaire à SVM pour la classification). Cela signifie que tant que les prédictions du modèle sont à l'intérieur de la bande ϵ , l'erreur n'est pas pénalisée.

La fonction à minimiser peut-être représentée ainsi :

$$\frac{1}{2} * ||\mathbf{w}||^2 \text{ Sous contrainte } f(x) = \begin{cases} y_i - (\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}), & x < \epsilon \\ (\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}) - y_i, & x \geq \epsilon \end{cases}$$

Où $\mathbf{w}$ est le vecteur de poids, $\mathbf{x}_i$ est l'entrée, y_i est la sortie et $\mathbf{b}$ est le biais.

SVR minimise une fonction de *cout* $\frac{1}{2} * ||\mathbf{w}||^2$ tout en maintenant les erreurs en dessous de ϵ . Les erreurs qui dépassent ϵ sont pénalisées avec une marge définie par le paramètre C (qui contrôle la régularisation). Les erreurs en dehors de cette marge sont appelées slack variables

et leur somme contribue à la fonction de coût totale. Le paramètre **C** détermine le compromis entre la minimisation de l'erreur et la maximisation de la marge. Un **C** élevé réduit la marge et donne une plus grande importance à l'erreur, tandis qu'un **C** plus faible favorise une marge plus large, au risque de sous-ajuster les données.

Noyaux (Kernel Trick) : SVR utilise des noyaux pour mapper les données dans un espace de plus haute dimension où il devient plus facile de trouver un hyperplan linéaire pour la régression. **Noyau Linéaire** : Si les données sont linéaires, un noyau linéaire simple est utilisé. **Noyau RBF (Radial Basis Function)** : Pour capturer des relations non linéaires complexes. Le noyau RBF est défini par $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$ où γ détermine la portée d'influence d'un seul point de données. **Polynomial** : Capture des interactions non linéaires de manière polynomiale.

2-5-2-6 Stacking

Le modèle stacking une approche avancée, qui combine plusieurs algorithmes d'apprentissage automatique pour améliorer la précision des prédictions. D'abord, plusieurs modèles de base, chacun possédant des caractéristiques distinctes et complémentaires, sont entraînés pour estimer les débits à partir des données issues des réanalyses climatiques ERA5 et LAND. Ces modèles incluent des algorithmes tels que Random Forest , Extra Trees , SVR , K-Nearest Neighbours (KNN) , AdaBoost , Gradient Boosting , et un réseau de neurones multi-couches (MLP). Les données d'entrée sont soigneusement normalisées et transformées par une transformation polynomiale pour mieux capturer les non-linéarités présentes dans les relations entre les réanalyses climatiques et les débits observés. Après l'entraînement, chaque modèle de base produit ses propres prédictions des débits. Ces prédictions sont ensuite empilées pour constituer un nouvel ensemble de données de prédictions, capturant ainsi les diverses perspectives fournies par chaque modèle de base. Un **modèle de régression linéaire** intermédiaire est ensuite appliqué sur ces prédictions empilées. Cette étape intermédiaire affine davantage les prédictions, en combinant de manière optimale les résultats des modèles de base et en particulier leurs biais individuels. Pour la phase finale, un modèle de **Gradient Boosting** est entraîné sur les prédictions affinées de la régression linéaire. Ce modèle de Gradient Boosting, qui combine les prédictions intermédiaires, ajuste encore les estimations pour produire une prédiction finale des débits manquants. L'approche de stacking utilisée dans ce modèle vise à tirer partie de la capacité de chaque modèle de base à capturer différents aspects des données, ce qui conduit à une estimation plus robuste et plus précise des débits.

2-5-2-7 Les réseaux de neurones artificiels

Les réseaux de neurones artificiels sont inspirés du fonctionnement du cerveau humain, où des neurones interconnectés traitent et transmettent l'information. Dans ces réseaux, chaque neurone reçoit des signaux, les traite à l'aide de poids et de biais, et utilise une fonction d'activation pour générer une sortie. Les réseaux de neurones sont très efficaces pour reconnaître des motifs complexes dans les données, d'où leur popularité dans les domaines de la reconnaissance d'images, du traitement du langage naturel, et bien d'autres. Cependant, les réseaux de neurones classiques ont des limitations pour traiter des données séquentielles, car ils ne retiennent pas d'information d'un instant à l'autre. C'est pour cela que les réseaux de neurones récurrents (RNN) et les modèles de mémoire à long court terme (LSTM) ont été créés. Les modèles LSTM ont été introduits pour pallier les problèmes de mémoire courte des RNN standards, intègrent des "cellules" capables de mémoriser des informations sur de longues périodes grâce à des mécanismes de portes (d'entrée, d'oubli et de sortie).

Ces portes permettent au réseau de décider quelles informations conserver ou oublier, faisant des LSTM un outil puissant pour les séries temporelles et les séquences de données longues.

<https://www.deeplearningbook.org/>

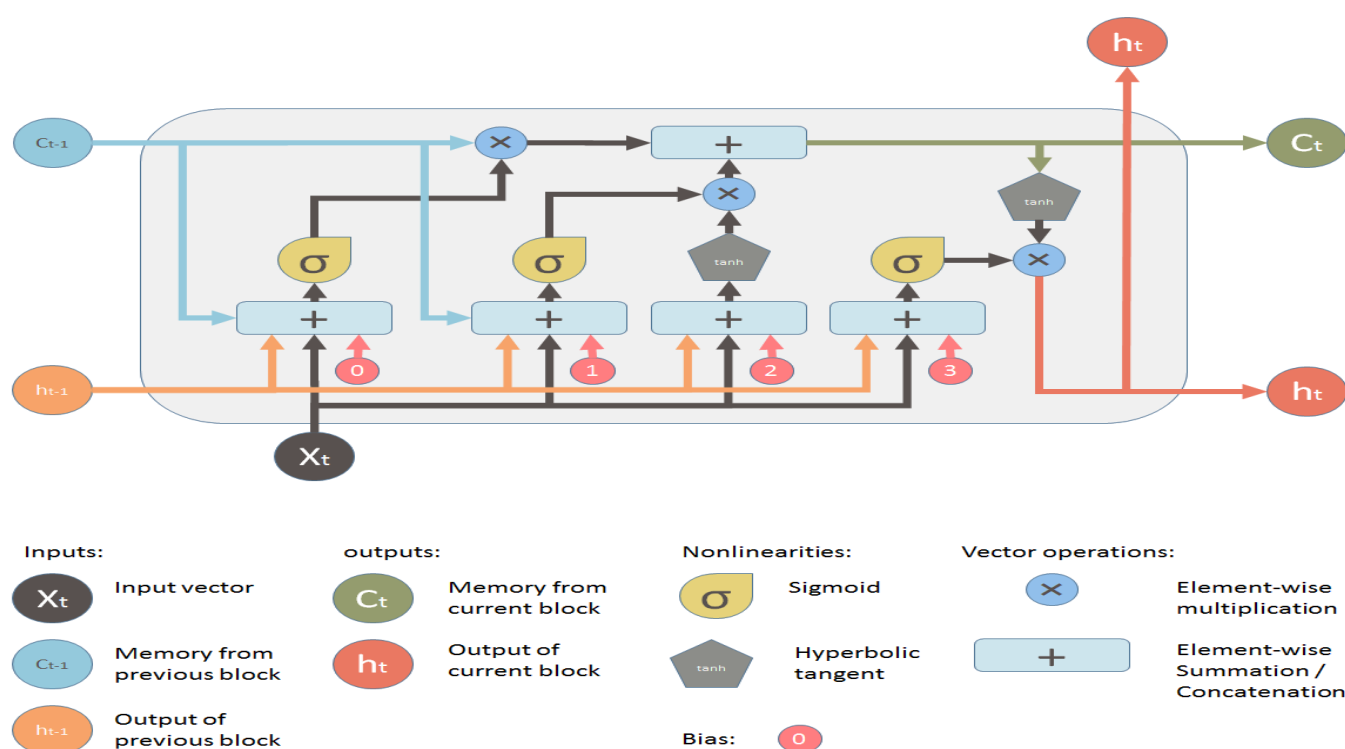


Figure 6: Fonctionnement interne d'un réseau de neurones récurrent

2-5-3-1 Les éléments principaux de l'architecture LSTM

1. **Les entrées** : LSTM reçoit à chaque étape une nouvelle entrée vectorielle X_t la mémoire précédente C_{t-1} , et la sortie précédente h_{t-1} .
2. **Les portes (sigmoïde et tanh)** : Trois portes (entrée, oubli, sortie) régulent le flux d'information. Elles permettent de contrôler ce qui doit être retenu ou oublié, grâce aux fonctions sigmoïde (qui déterminent l'importance des informations) et tanh (qui introduisent des non-linéarités pour moduler les nouvelles informations).
3. **Cellule mémoire (C_t)** : Elle stocke l'information pertinente à long terme. Les portes permettent de mettre à jour cette mémoire.
4. **La sortie (h_t)** : Elle représente l'état caché de l'unité LSTM, qui est influencé par les nouvelles données et l'état de mémoire.

2-5-3-2 Les fonctions d'activation

Il existe plusieurs fonctions d'activation utilisées dans les réseaux de neurones, dont les plus courantes incluent :

1. Sigmoid :

Description : La fonction sigmoïde prend une valeur d'entrée et la transforme en une valeur entre 0 et 1, ce qui est utile pour les problèmes de classification binaire. Elle peut souffrir du problème de saturation, où les gradients deviennent très petits, ralentissant l'apprentissage (problème de vanishing gradient).

2. Tanh :

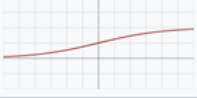
Description : La fonction tanh est similaire à la sigmoïde, mais elle transforme les valeurs d'entrée en un intervalle de -1 à 1. Cela permet une meilleure représentation des valeurs négatives. Bien qu'elle résolve en partie le problème de saturation, elle peut également être affectée par le vanishing gradient.

3. ReLU (Rectified Linear Unit) :

Description : La fonction ReLU transforme toutes les valeurs négatives en zéro tout en laissant les valeurs positives inchangées. Cela aide à introduire de la non-linéarité dans le modèle. Elle permet une convergence plus rapide lors de l'entraînement et réduit le problème de vanishing gradient, rendant ReLU très efficace pour les réseaux profonds. Sa simplicité computationnelle en fait un choix privilégié pour de nombreux types de modèles, y compris les LSTM.

<https://www.deeplearningbook.org/>.

Tableau 1 : Les différentes fonctions d'activation

Name ⇄	Plot	Function, $f(x)$ ⇄	Derivative of f , $f'(x)$ ⇄	Range ⇄
Identity		x	1	$(-\infty, \infty)$
Binary step		$\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$	$\begin{cases} 0 & \text{if } x \neq 0 \\ \text{undefined} & \text{if } x = 0 \end{cases}$	$\{0, 1\}$
Logistic, sigmoid, or soft step		$\sigma(x) = \frac{1}{1 + e^{-x}}$ [1]	$f(x)(1 - f(x))$	$(0, 1)$
tanh		$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$1 - f(x)^2$	$(-1, 1)$
Rectified linear unit (ReLU) [11]		$\begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$ $= \max\{0, x\} = x \mathbf{1}_{x>0}$	$\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \\ \text{undefined} & \text{if } x = 0 \end{cases}$	$[0, \infty)$

<https://inside-machinelearning.com/fonction-dactivation-comment-ca-marche-une-explication-simple/>

Les époques représentent le nombre de fois où l'ensemble de données d'entraînement est parcouru pendant l'entraînement d'un modèle. Un nombre d'époques élevé permet au modèle d'apprendre des motifs complexes, mais peut aussi entraîner un surapprentissage, où le modèle devient trop spécifique aux données d'entraînement. Il est donc crucial de surveiller la performance sur un ensemble de validation pour déterminer le nombre optimal d'époques. Enfin, un nombre excessif d'époques peut augmenter considérablement le temps d'entraînement.

2-5-3-3 Hyperparamètres utilisés dans ce modèle :

1. **Nombre d'époques** : le modèle est entraîné pendant 60 époques. Cela détermine combien de fois le modèle passe par l'ensemble de données d'entraînement.
2. **Taux de Dropout** : un taux de dropout de 20 % pour réduire le risque de surapprentissage. Le dropout est un mécanisme pour régulariser le modèle en désactivant aléatoirement certains neurones durant l'entraînement.
3. **Taille du batch** : la taille du batch à 32, ce qui influence la manière dont les données sont présentées au modèle lors de l'entraînement.
4. **Nombre d'unités LSTM** : 50 unités sont choisies pour chaque couche LSTM, ce qui affecte la capacité d'apprentissage du modèle.

2-5-3-4 Étapes d'implémentation

Préparation des données :

Création des séquences : Puisque nous travaillons avec des séries temporelles, nous avons défini une fonction pour transformer les données en séquences de longueur spécifiée et capturer la dynamique temporelle. Cette fonction crée des séquences en déplaçant une fenêtre de la longueur spécifiée à travers les données. Les données ont été divisées en sous-séries en utilisant une sous-série de 20 jours, appelée `time_step`. Chaque sous-série devient une observation pour le modèle, avec la cible étant les valeurs des sept (07) jours suivants.

Partitionnement des données : Les données sont scindées en différents sous-ensembles. Une partition en ensembles d'entraînement de 1980 à 2009, de validation de 2010 à 2016, et de test de 2017 à 2020 garantira que le modèle est bien formé, optimisé, et évalué sur des données distinctes.

1-Ensemble d'entraînement

Description : L'ensemble d'entraînement est constitué des données utilisées pour apprendre les paramètres du modèle. C'est sur cet ensemble que le modèle est ajusté, lui permettant de "comprendre" les relations entre les variables d'entrée et les valeurs cibles. Il permet au modèle de minimiser l'erreur entre les prédictions et les vraies valeurs à l'aide de l'optimisation des paramètres.

2. Ensemble de validation

Description : Cet ensemble est constitué d'un sous-ensemble de données distinct de l'entraînement, mais utilisé pour valider et ajuster le modèle pendant le processus d'entraînement. Aide à évaluer la performance du modèle en dehors des données d'entraînement, ce qui permet de détecter et d'atténuer les problèmes de surapprentissage ou de sous-apprentissage. Cet ensemble sert également pour les ajustements d'hyperparamètres, comme le taux d'apprentissage ou la profondeur des arbres, sans que le modèle ne "voie" l'ensemble de test final.

3. Ensemble de test

Description : L'ensemble de test est un ensemble de données totalement indépendant utilisé pour la dernière évaluation des performances du modèle. Fournit une estimation fiable de la performance du modèle sur des données qu'il n'a jamais vues. L'ensemble de test est la référence pour évaluer la capacité de généralisation réelle du modèle.

2-5-3-5 Structure du modèle développé :

1. Couche LSTM :

La première couche LSTM avec `return_sequences=True` permet de transmettre la séquence complète d'états cachés à la couche suivante. Cela est crucial pour la gestion des dépendances temporelles dans des données qui ont plusieurs pas de temps.

La deuxième couche LSTM (sans `return_sequences`) produit uniquement l'état caché final, ce qui est approprié pour des tâches de prédiction où seule la dernière information est nécessaire.

2. Dropout : le Dropout est une technique régulière utilisée pour éviter le surapprentissage (overfitting). En désactivant aléatoirement des neurones pendant l'entraînement, cela force le réseau à apprendre des représentations plus robustes et généralisables.

3. Fonction d'activation : Par défaut, les unités LSTM utilisent une fonction d'activation tanh pour leur état caché et une fonction sigmoïde pour leurs portes (entrée, oubli, sortie). Cela aide à contrôler le flux d'informations dans le modèle.

4. Optimiseur et fonction de perte

Optimiseur Adam : l'optimiseur Adam est un choix populaire pour l'entraînement des réseaux de neurones, notamment pour sa capacité à adapter les taux d'apprentissage pendant l'entraînement, ce qui peut améliorer la convergence.

Fonction de perte MSE : La fonction de perte MSE (mean squared error) est particulièrement adaptée pour les problèmes de régression, car elle pénalise les grandes erreurs de manière quadratique, ce qui aide à guider le modèle vers des prédictions plus précises.

5. **Entraînement et Validation** : Le modèle a été entraîné pendant 60 époques avec l'optimiseur Adam, en utilisant l'erreur quadratique moyenne (`mean_squared_error`) comme fonction de perte pour minimiser l'écart entre les valeurs prédites et observées. Pour évaluer la performance du modèle, les données ont été séparées en ensembles d'entraînement, de validation et de test, afin de vérifier la généralisation du modèle.

Tableau 2: Qualité des modèles en fonction des valeurs du NSE

NSE	Appréciation du modèle
$\text{NSE} \geq 90\%$	Excellente
$80\% \leq \text{NSE} < 90\%$	Très bonne
$60\% \leq \text{NSE} < 80\%$	Bonne
$\text{NSE} < 60\%$	Mauvaise

2-5-4 Les lois d'ajustement

Dans cette section, nous abordons l'ajustement de quatre lois de distribution aux crues extrêmes afin de déterminer celle qui représente le mieux les séries de données. Avant d'effectuer cet ajustement, il est essentiel de vérifier que les données respectent certaines hypothèses de base, qui seront testées statistiquement. Les tests incluent le test d'indépendance de Wald-Wolfowitz, qui évalue si les observations sont indépendantes, le test de stationnarité de Mann-Kendall pour détecter les tendances dans des séries chronologiques, et le test d'homogénéité de Wilcoxon pour comparer plusieurs échantillons. Les lois choisies pour l'ajustement sont **la loi log-normale, la loi de Gumbel, la loi Pearson type III, la loi log-Pearson type III et GEV**. Pour choisir la meilleure loi, nous utiliserons des critères tels que l'AIC (Critère d'Information d'Akaike) et le BIC (Critère d'Information Bayésien), qui ont été utilisés par plusieurs chercheurs dans des études antérieures. Enfin, pour estimer les paramètres des lois retenues, nous appliquerons la méthode du maximum de vraisemblance et la méthode des L-moments, cette dernière étant particulièrement efficace pour la distribution des valeurs extrêmes généralisées.

CHAPITRE 3 : RESULTATS ET DISCUSSIONS

Ce chapitre expose d'abord les résultats obtenus en appliquant diverses méthodes de correction de biais sur les données du bassin du Mono à Athiémé. Ensuite, les performances de ces méthodes sont éprouvées à l'aide de plusieurs critères déterministes, notamment le KGE, le RMSE, le R^2 , le PBIAS et la corrélation. Nous identifions également les probabilités d'occurrence des crues Q95 issues de la meilleure méthode, avant de développer le modèle de prévision. Enfin, une discussion approfondie sera menée sur les différents résultats obtenus.

3-1 Evaluation de la qualité des méthodes de correction par les critères d'évaluation

Tableau 3: Efficacité des corrections sur ERA5 et LAND à pas de temps journalier.

	KGE		RMSE		Cor		R2		PBIAS		NSE	
	ERA5	LAN D	ERA 5	LAN D	ERA5	LAN D	ERA 5	LAN D	ERA 5	LAN D	ERA 5	LAN D
Données brutes	0.04	0.13	0.46	0.44	0.1	0.18	0.01	0.03	-24	16.56	-1.26	-1.4
Régression linéaire	-7.4	-3.38	0.3	0.29	0.1	0.19	0.01	0.03	0	0	0.01	0.03
Régression polynomiale degré 3	-0.24	-0.09	0.3	0.3	0.12	0.22	0.01	0.05	0	0	0.01	0.05
QM avec « Expo-asymptique »	0.13	0.22	0.4	0.37	0.13	0.22	0.02	0.05	0.4	10.42	-0.7	-0.56
QM avec « muff »	0.12	0.22	0.39	0.37	0.13	0.22	0.02	0.05	-0.9	8.41	-0.7	-0.53
QM avec « bergamma »	0.12	0.2	0.4	0.4	0.13	0.2	0.02	0.04	-2.51	6.58	-1	-0.86
QM avec « ssplin »	-3.6	-1.88	0.5	0.29	0.11	0.27	0.01	0.07	0	13.36	0.03	0.07
RQUANT linéaire	0.12	0.21	0.4	0.37	0.12	0.21	0.02	0.04	0.26	0	-0.7	-0.64
SVR	-0.26	-0.13	0.31	0.3	0.14	0.24	-0.05	0.003	25.91	22.8	-0.05	0.031
RandomForest	0.31	0.52	0.25	0.2	0.56	0.74	0.3	0.53	0.09	0.16	0.31	0.53
Stacking	0.42	0.67	0.24	0.19	0.59	0.78	0.35	0.6	0.0	0.0	0.35	0.6

Ce tableau compare les performances de différentes méthodes de correction de biais sur les données climatiques ERA5 et LAND à l'aide de plusieurs métriques : KGE, RMSE, Corrélation (Cor), R^2 , PBIAS, et NSE. Les données brutes montrent des scores faibles, particulièrement avec des valeurs de KGE, R^2 , et NSE proches de zéro, indiquant une performance initiale médiocre. Parmi les méthodes testées, les approches de quantile mapping (QM) avec des variantes comme « Expo-asymptique » et « ssplin » améliorent modestement les résultats, mais le stacking atteint les meilleures performances, particulièrement pour LAND (KGE = 0.67, R^2 = 0.6, Corrélation = 0.78, et NSE = 0.6), surpassant les autres modèles. La Random Forest et

la régression polynomiale apportent également une par rapport aux données brutes, mais restent en deçà du stacking en termes de précision et de réduction du biais. En conclusion, l'approche de stacking se révèle la plus efficace pour la correction de biais, améliorant de manière significative la corrélation et la précision sur les deux ensembles de données ERA5 et LAND.

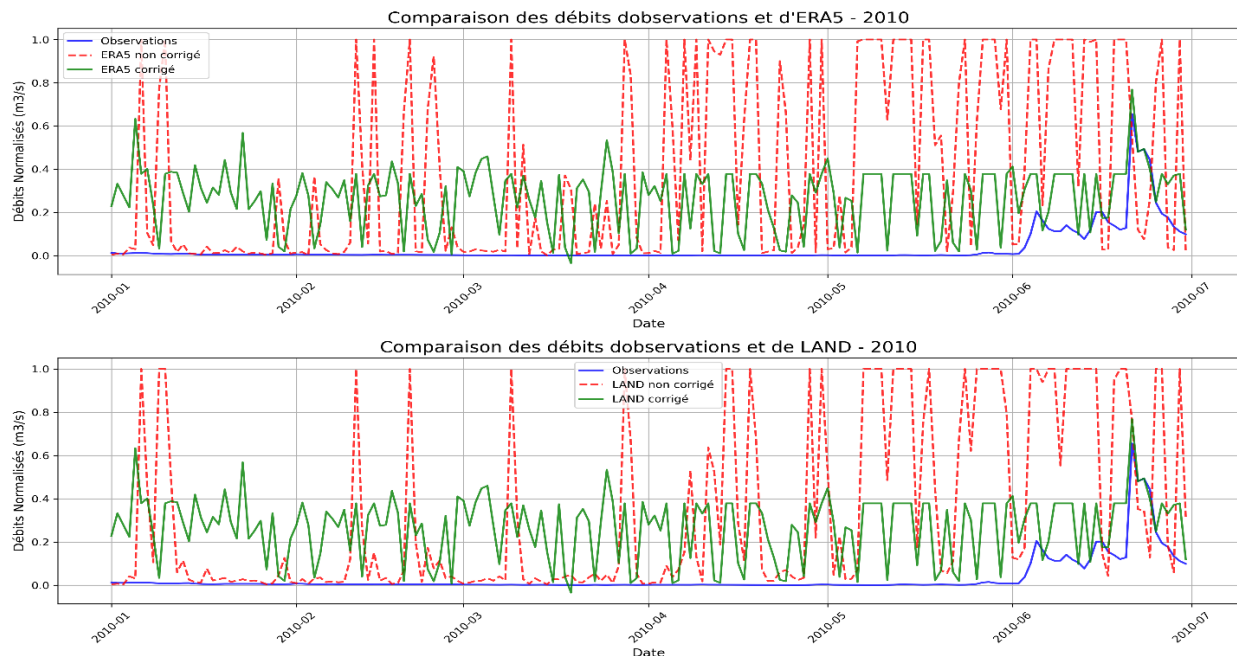


Figure 7: Débits observés et débits des réanalyses bruts et corrigés en 2010

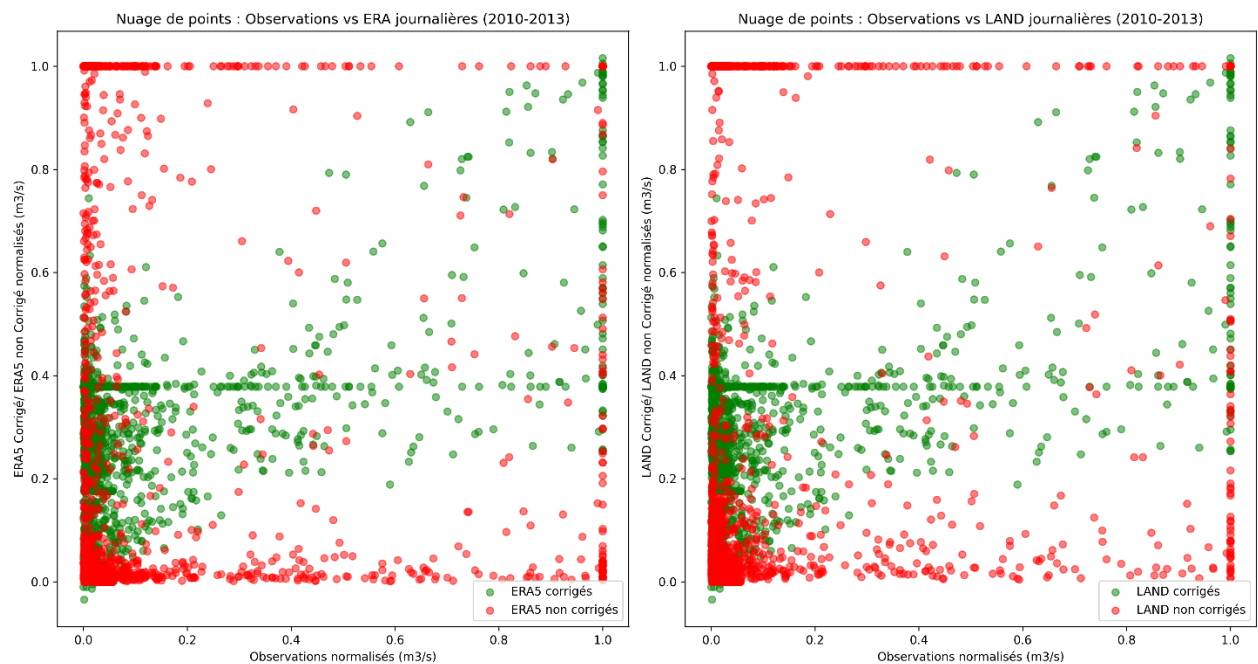


Figure 8: Nuage de point des débits d'observation et des réanalyses journaliers avant et après correction.

3-2 Analyse fréquentielle des débits de crues (Q95).

Après correction des réanalyses ERA5 et LAND, les performances obtenues restaient insuffisantes. Pour améliorer les résultats, une combinaison des deux réanalyses a été réalisée et intégrée dans un modèle de stacking. Cette méthode a significativement amélioré les performances, offrant une base plus solide pour conduire l'analyse fréquentielle. Cette approche, s'est avérée essentielle pour tirer des conclusions pertinentes dans l'étude des fréquences des crues.

Tableau 4: Les performances des réanalyses empilés après correction

Indicateurs Types de données	KGE	RMSE	Cor	R ²	Pbiais	NSE
Réanalyses empilés(corrigé)	0.88	0.12	0.91	0.84	0.0	0.84

Tableau 5 : Statistique descriptive des données d'observation et réanalyses

Indicateurs Types de données	Min	Med	Moy	Max	Ecart type	Cv
Observations	76.6	395	405	787	182	0.45
Réanalyses	76.0	398	406	795	184	0.45

3-2-1. Tests d'hypothèses, de stationnarité et d'homogénéité

Il s'agit de procéder à la vérification des tests statistiques pour voir l'indépendance, l'homogénéité ou la stationnarité de chaque donnée. Ainsi, les hypothèses d'indépendance, d'homogénéité et de stationnarité ont été effectuées. L'hypothèse d'indépendance a été validée à l'aide du test de Wald-Wolfowitz. Les tests d'homogénéité et de stationnarité utilisés sont respectivement ceux de Wilcoxon et de Kendal.

Tableau 6 : Résultat du test d'hypothèse

Données	U	P-value	Nature des données
Test d'indépendance			
Observation	1.38	0.167	Indépendantes
Réanalyses	1.4	0.163	Indépendantes
Test de stationnarité			
Observation	1.67	0.094	Stationnaires
Réanalyses	1.7	0.08	Stationnaire
Test de d'homogénéité			
Observation	1.92	0.055	Homogène
Réanalyses	1.9	0.05	Homogène

|U| représente la valeur de la statistique. Il ressort de l'analyse des résultats de ce tableau, que la série chronologique (1980-2020) est indépendante, pour les données d'observation et les reanalyses car il n'y a pas de liens entre les observations successives au seuil de défaillance de 5%. |K| : valeur de la statistique ; P : p-value L'application du test de Kendal révèle qu'il n'y a pas de tendance dans les observations, et les réanalyses au seuil de défaillance de 5% et |W| : valeur de la statistique ; P : p-value. Dans le test de Wilcoxon particulièrement, nous avons subdivisé les observations de chaque type de données en deux sous échantillons égales afin de vérifier l'homogénéité. L'hypothèse H0 considérée ici est : Les moyennes des deux sous-échantillons sont égales. L'analyse des résultats de ce tableau révèle que la moyenne des deux sous-échantillons est égale au seuil de défaillance de 5% pour l'observation et les réanalyses. Par conséquent, les données sont homogènes. On remarque ici que toutes les données sont : indépendantes, stationnaires et homogènes.

3-2-2 Ajustement des lois aux Q95

3-2-2-1 Ajustement sur critères des Q95.

Pour les données historiques, le choix de la meilleure loi, s'est basé sur deux critères. Il s'agit du critère d'information de Bayésien (BIC) et du critère d'information d'Akaike (AIC). La meilleure loi à retenir est celle qui a simultanément les deux critères faibles.

Tableau 7 : Critères AIC et BIC des débits historiques

Données	Lois	Paramètres	AIC	BIC
OBSERVATION	Log Pearson de type III	3	545.17	550.31
	Gumbel	2	547.2	550.7
	GEV	3	546.29	552.03
	Log normale	3	545.9	551.07
	Pearson de type III	3	547.3	552.48
Réanalyses	Log Pearson de type III	3	545.91	551.05
	Gumbel	2	547.78	551.21
	GEV	3	547.56	552.7
	Log normale	3	548.08	553.2
	Pearson de type III	3	447.95	553.09

3-2-2-2 Ajustement graphique des Q95

L'analyse des graphes révèle que les Q95 des observations suivent très bien la loi de Log Pearson de type III, les Q95 des réanalyses et des observations suivent la loi de log Pearson de type III aussi. Aussi, tous les points sont dans l'intervalle de confiance de 95% pour cette loi.

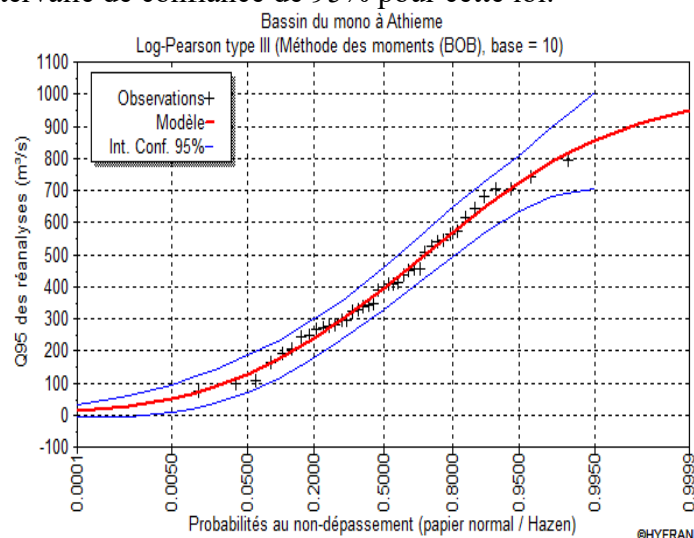
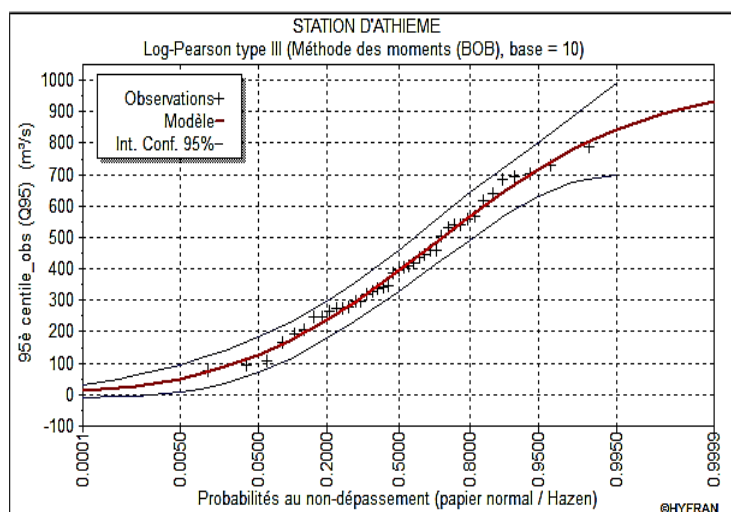


Figure 9 : Ajustement graphiques des observations et les réanalyses par la loi de log Pearson III

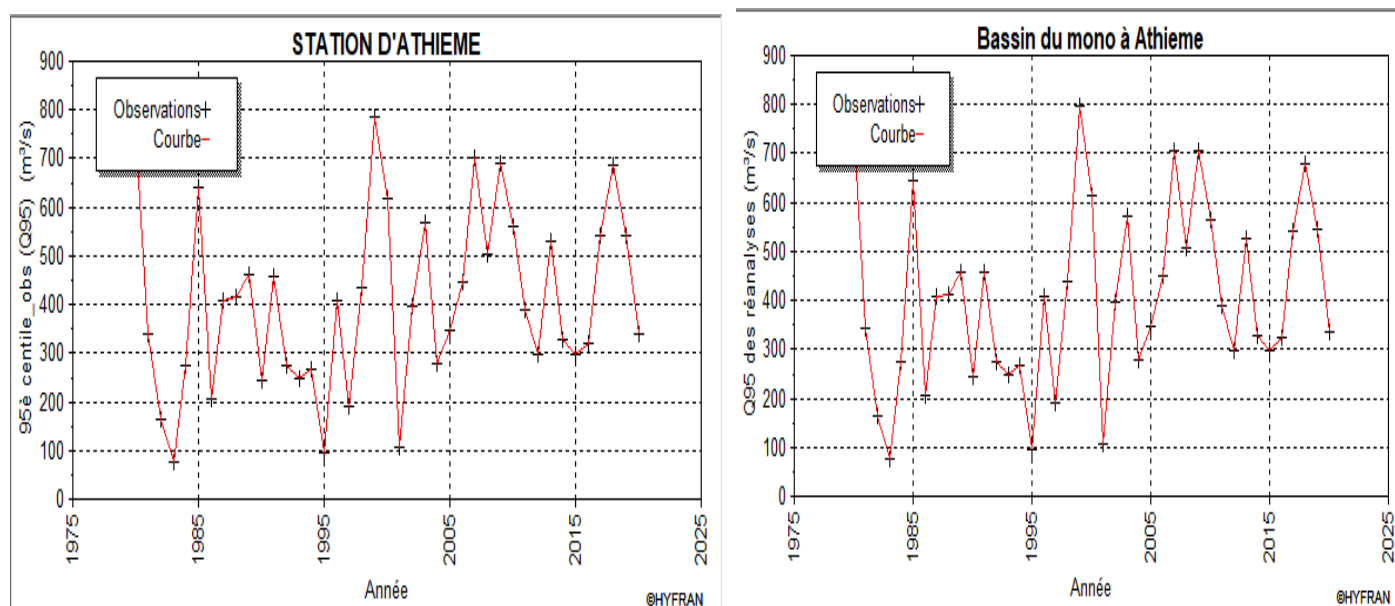


Figure 10 : Variabilité des Q95 historiques des réanalyses en fonction des années

L'analyse de la figure nous permet de révéler les variations au niveau des observations et des réanalyses. Ainsi, on note de cette analyse une forte variation 95è centile (Q95) au-dessus de la moyenne. On constate alors une évolution en dents de scies des Q95 dans la période d'étude et que, soit les réanalyses et les observations dans ce bassin ont une tendance hydrométrique à la hausse parce que plus de la moitié des années observées sont au-dessus de la moyenne hydrométrique. Ce qui se justifie par le test de Mann Kendall avec le taux qui est égale 0.18 et p value égale à 0.09

3-2-2-3 Estimation des quantiles des séries de débits de crues étudiées.

A partir de l'équation issue de l'ajustement de chaque loi, les quantiles ont été calculés pour différentes période de retour. Pour le calcul des quantiles, les périodes de retour retenues sont celles de 2 ans, 5 ans, 10 ans, 20 ans, 50 ans et 100 ans.

Tableau 8 : Quantiles des séries de débits de crues à différentes périodes de retour.

Période	2 ans	5 ans	10 ans	20 ans	50 ans	100 ans
Observation	395 [329-461]	568 [493-643]	653 [576-731]	717 [632-802]	780 [674-885]	816 [690-941]
Réanalyses	395 [329-461]	570 [494-646]	657 [578-736]	723 [635-810]	787 [679-896]	825 [695-956]

De son analyse, on remarque que l'évolution des quantiles est proportionnelle à celles des périodes de retour. Après correction, les réanalyses parviennent à reproduire la variabilité temporelle observée, ce qui montre l'efficacité de l'ajustement pour capturer les fluctuations hydrologiques dans le bassin du Mono. En outre, pour cette l'analyse fréquentielle, les Q95 des réanalyses corrigées ajuste correctement aux observations. Ces résultats indiquent la capacité des modèles d'IA pour la correction des réanalyses et à estimer précisément les événements extrêmes.

3-3 Développement d'un modèle LSTM pour la prévision des débits journaliers.

Pour prédire les valeurs de la série des débits, nous avons implémenté un modèle de réseau de neurones à mémoire court longue, ou LSTM (Long Short-Term Memory), un type de réseau récurrent (RNN) bien adapté aux données séquentielles.

3-3-1 Convergence du MSE en fonction des époques :

Pour évaluer la performance de notre modèle durant l'entraînement, nous avons analysé la convergence du MSE (Mean Squared Error) au fil des époques. Cette étape permet d'observer la manière dont le modèle apprend progressivement et réduit son erreur moyenne, ce qui est important pour garantir une bonne capacité de généralisation.

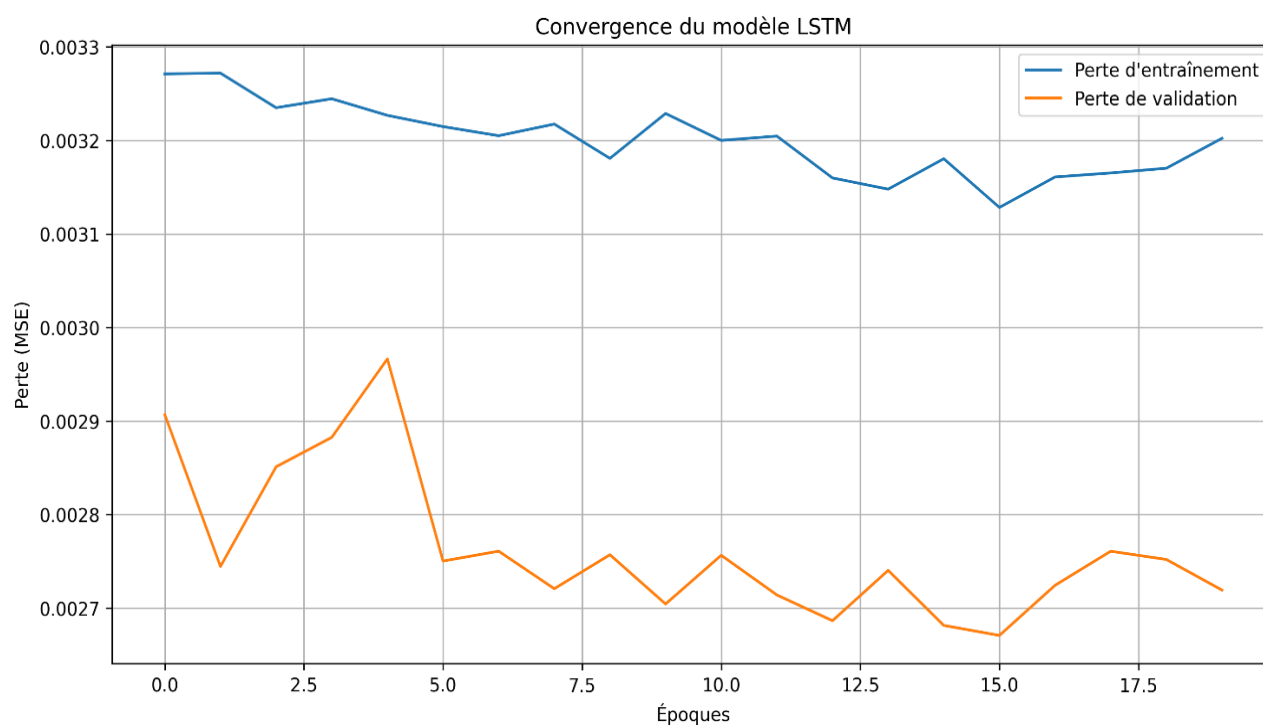


Figure 11 : L'évolution des MSE en fonction des époques à l'entrainement et à la validation.

3-3-2 Performance des prévisions pour Chaque Jour

Les performances globales du modèle ont été évaluées avec les métriques NSE, RMSE, KGE et R². Ces indicateurs fournissent un aperçu complet de la précision des prédictions, ainsi que des points forts et des limites du modèle dans la prédiction des données de test.

Tableau 9 : Résultats des critères déterministes calculés sur les temps de prévisions allant de 1 à 7 jours.

Critère de performance Echéances	NSE	RMSE	KGE	R2
1	0.94	32.10	0.95	0.94
2	0.93	34.76	0.95	0.93
3	0.92	36.82	0.95	0.92
4	0.91	40.24	0.94	0.91
5	0.90	42.20	0.94	0.90
6	0.89	45.33	0.93	0.89
7	0.87	47.24	0.92	0.8

Le modèle montre une **bonne précision** pour des prévisions allant jusqu'à sept jours, avec des scores NSE supérieurs à 0,80. Toutefois, la précision commence par **diminuer** légèrement avec l'éloignement temporel des échéances. Globalement, le modèle est **particulièrement bien adapté** aux prévisions.

Tableau 10 : Résultats des critères probabilistes calculés sur les temps de prévisions allant de 1 à 7 jours suivant les terciles.

Station	Echéance	Critères	Valeur par catégorie		
			Au-dessous du 1 ^{er} Tercile	Entre le 1 ^{er} et le 2 ^{ème} Tercile	Entre le 2 ^{ème} et le maximum
Athiémé	1	HR	0.85	0.76	0.95
		Pr	0.82	0.78	0.95
		F1	0.83	0.77	0.95
		AUC	0.89	0.85	0.97
	2	HR	0.85	0.76	0.95
		Pr	0.82	0.78	0.95
		F1	0.83	0.77	0.95
		AUC	0.88	0.83	0.96
	3	HR	0.81	0.72	0.93
		Pr	0.79	0.73	0.93
		F1	0.80	0.73	0.93
		AUC	0.86	0.81	0.95
	4	HR	0.81	0.72	0.93
		Pr	0.79	0.73	0.93
		F1	0.80	0.73	0.93
		AUC	0.85	0.80	0.94
	5	HR	0.77	0.73	0.91
		Pr	0.79	0.70	0.92
		F1	0.78	0.72	0.92
		AUC	0.84	0.79	0.94
	6	HR	0.78	0.70	0.92
		Pr	0.78	0.70	0.91
		F1	0.78	0.70	0.92
		AUC	0.84	0.78	0.94
	7	HR	0.74	0.71	0.90
		Pr	0.79	0.67	0.91
		F1	0.76	0.69	0.90
		AUC	0.82	0.77	0.93

Ces résultats transmettent la capacité du modèle à effectuer de bonnes prévisions pour chaque classe, en fonction des catégories définies par les terciles. Les métriques HR, Pr et F1 montrent que le modèle parvient à prédire les occurrences dans chaque intervalle (au-dessous du 1^{er} tercile, entre le 1^{er} et le 2^e tercile, et entre le 2^e tercile et le maximum.), bien que la précision diminue légèrement avec l'éloignement temporel des échéances. Le critère AUC,

mesurant la qualité globale du modèle, reste élevé tout au long des échéances, avec des scores variants entre 0,77 et 0,96. Cela indique que le modèle maintient un bon niveau de performance dans la différenciation et la prévision des débits des classes, même à long terme.

3-3-3 Evaluer de la performance du modèle en termes de sensibilité et de spécificité.

La courbe ROC (Receiver Operating Characteristic) permet d'évaluer la performance du modèle en visualisant la relation entre le taux de faux positifs (1 - spécificité) et le taux de vrais positifs (sensibilité) pour chaque classe.

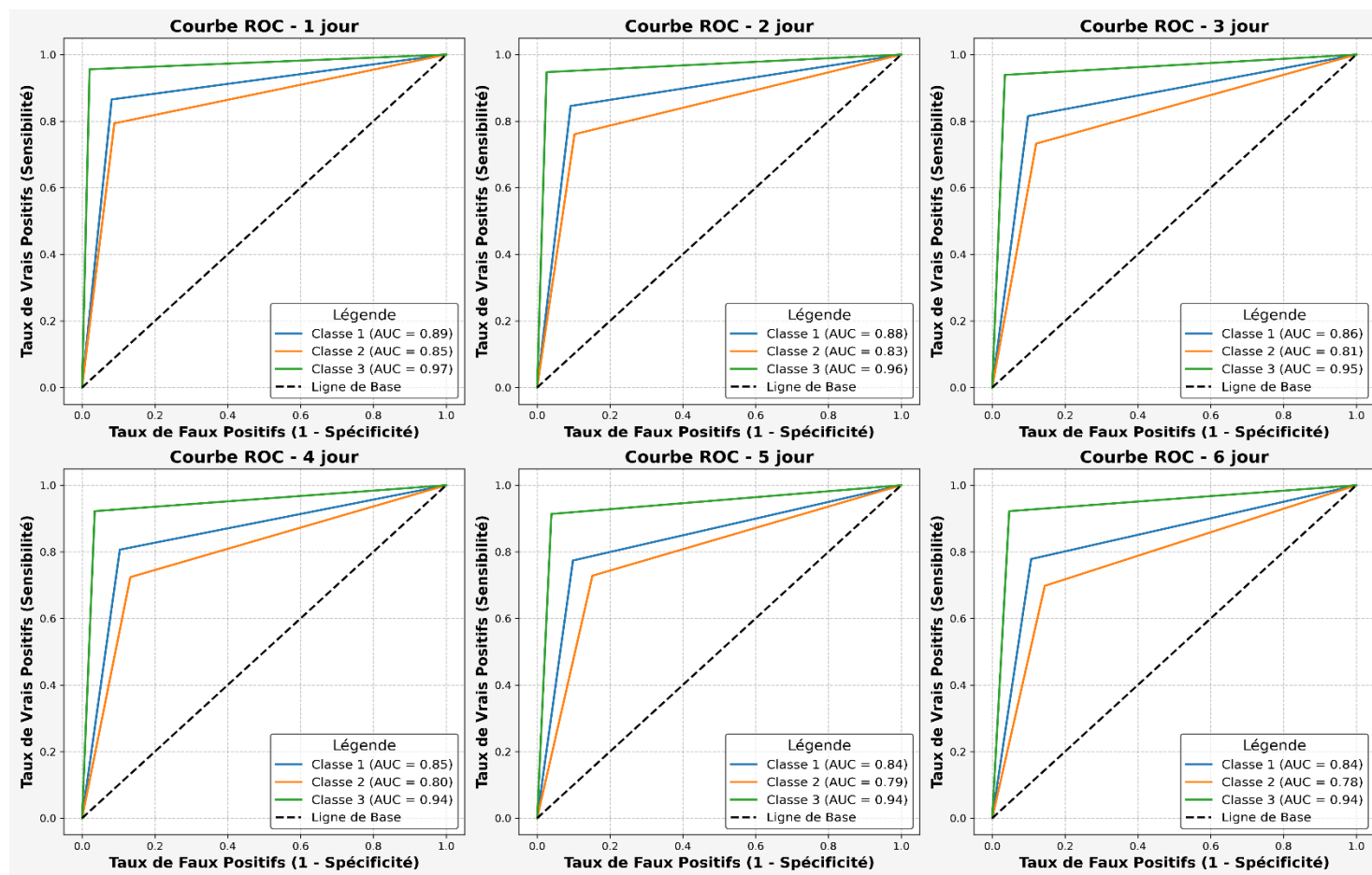


Figure 12 : Courbes ROC pour des Échéances de 1 à 6 Jours

Les courbes ROC présentées montrent que le modèle de prévision est performant pour toutes les classes (Tercile 1, Tercile 2 et Tercile 3) et sur toutes les périodes, de 1 à 6 jours. Chaque courbe est inclinée vers le coin supérieur gauche, indiquant un taux de réussite élevé (sensibilité) et un faible taux de fausses alertes (1 - spécificité). Les valeurs d'AUC (Area Under the Curve) confirment cette performance, dépassant 0.8 pour toutes les classes, ce qui témoigne d'une capacité globale du modèle à faire bonne prévision. Par exemple, pour la classe 3, l'AUC atteint jusqu'à 0.96, indiquant un modèle presque optimal avec un très faible

taux de fausses alertes. La classe 1, avec une AUC oscillant entre 0.90 et 0.84, maintient également une excellente performance sur l'ensemble des horizons temporels.

En conclusion, les courbes ROC montrent que le modèle est efficace pour toutes les classes, avec une gestion satisfaisante des fausses alertes et des taux de réussite constants sur des périodes allant jusqu'à 6 jours. Cela reflète la fiabilité du système pour la prévision multi-classes.

3-3-4 Graphique des résultats du Modèle pour Chaque Jour de Prédiction

Ce graphique illustre la performance du modèle sur plusieurs jours de prévision, réalisé à l'aide des métriques NSE, RMSE, KGE, et R^2 . Il met en évidence la précision du modèle pour des prévisions à court terme, avec une qualité qui tend à diminuer à mesure que l'horizon de prévision s'étend.

Vers le développement d'un modèle de prévision des débits à partir des réanalyses ERA5 et LAND à l'aide des méthodes de l'intelligence artificielle dans le bassin versant du Mono.

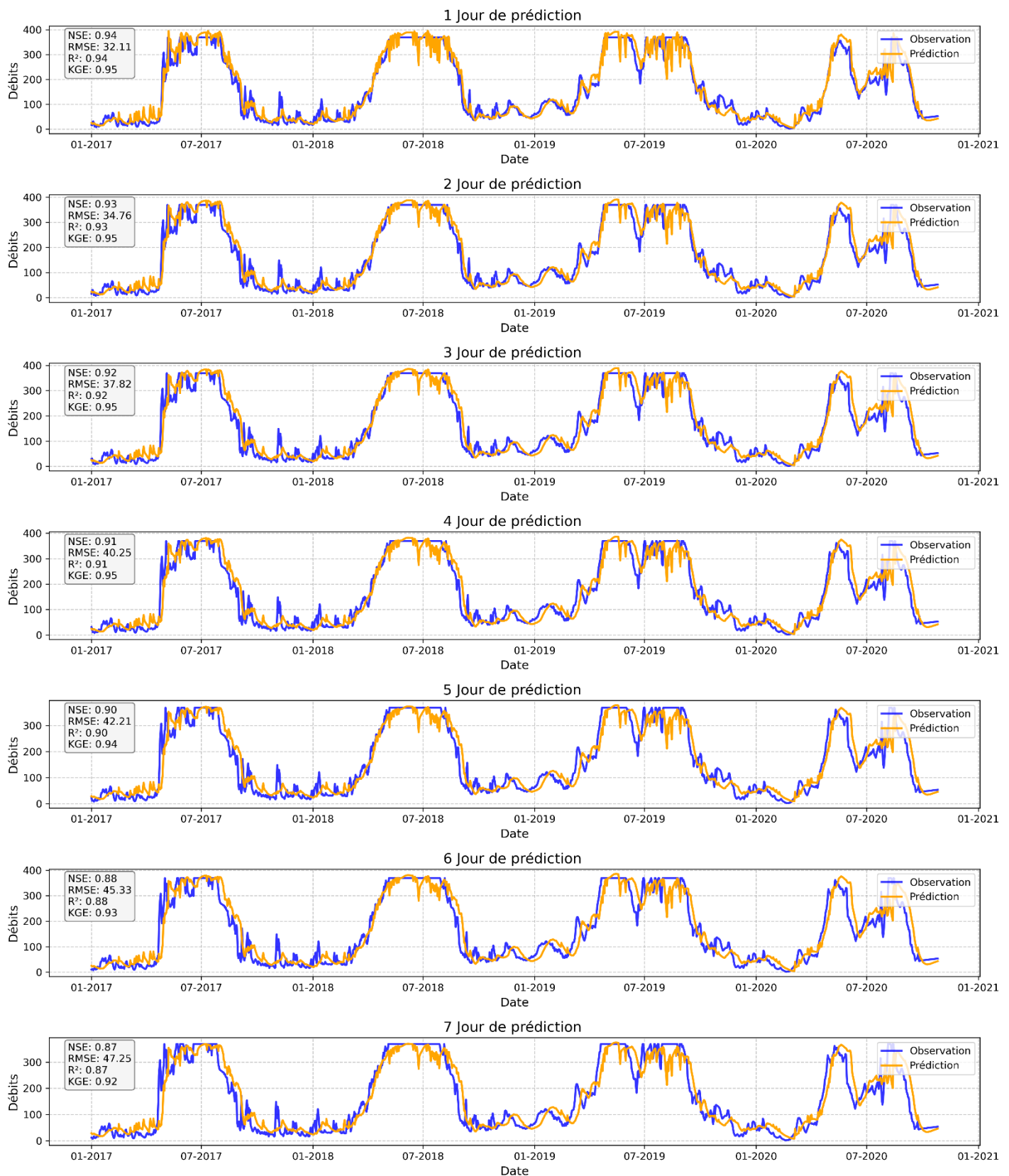


Figure 13 : Développement d'un modèle de prévision à chaque échéance.

3-4 Discussion

La présente étude a pour but d'évaluer la capacité prédictive des techniques d'intelligence artificielle, en particulier les réseaux de neurones, l'algorithme Random Forest (RF) et le Stacking, dans le domaine hydrologique. Dans un premier temps, l'évaluation des réanalyses ERA5 et LAND a montré des performances limitées en comparaison aux observations, confirmant les biais dans les données des réanalyses, comme observé dans d'autres travaux (Zogo, 2023). Ces résultats sont cohérents avec les conclusions de plusieurs chercheurs comme (Bodjrénou et al., 2024b) soulignant la nécessité de correction de biais pour aligner les réanalyses avec les conditions locales. Pour corriger ces biais, l'algorithme Random Forest et Stacking ont été utilisés comme méthode d'apprentissage automatique capable de capturer les relations complexes entre les variables climatiques et les débits observés. Ce choix est renforcé par des études antérieures ; Jenkins et al. (2018) ont comparé le modèle Random Forest avec des méthodes de régression classiques et ont démontré que Random Forest permet de mieux prédire les pertes en eau en se basant sur des covariables météorologiques, confirmant ainsi son efficacité dans la réduction des erreurs entre les prévisions et les observations. D'après (Dehghani et al., 2023), les réseaux de neurones de type LSTM montrent une efficacité particulière dans des contextes où la répartition spatiale des données revêt une importance primordiale. Ce facteur explique en partie les bonnes performances observées pour notre modèle LSTM dans des applications analogues. Par ailleurs, les travaux de recherche réalisés par (Hounnondaho, 2023) ont appliqué des modèles LSTM et les modèles ConvLSTM pour la prévision des hauteurs d'eaux dans le cadre de la modélisation hydrologique du bassin versants de Ouémé à l'exutoire de Béterou. Leur modèle principal a atteint un coefficient de détermination (R^2) de 0.96 pour le modèle ConvLSTM et 0.95 pour le modèle LSTM ce qui est proche des performances obtenues avec notre modèle LSTM. Dans une étude menée par Johnson, Jones et Thompson (2019), les auteurs ont procédé à une comparaison approfondie de diverses techniques d'apprentissage automatique en vue de la prévision des débits d'eau. Le modèle LSTM qu'ils ont conçu a atteint un coefficient de détermination remarquable de 0.96, un résultat qui se rapproche de celui obtenu par notre propre modèle LSTM, utilisé pour prédire les débits d'eau. La recherche de (Zogo, 2023) indique que les produits de CEPMMT sont plus fiables pour les prévisions à court terme, en particulier pour les pluies extrêmes. Cela rejoint les résultats obtenus dans cette étude, où le modèle LSTM a montré une bonne précision pour les prévisions à court terme. Cette performance met en lumière l'efficacité des réseaux de neurones dans le domaine hydrologique.

CONCLUSION ET PERSPECTIVES

Cette étude a permis de démontrer l'importance et l'efficacité des corrections de biais dans les réanalyses climatiques, en particulier ERA5 et LAND, pour la prévision des débits dans le bassin du Mono, à l'exutoire d'Athiémé. Sans correction, les réanalyses brutes présentent des écarts significatifs par rapport aux observations, limitant leur capacité à reproduire fidèlement les caractéristiques des débits, notamment lors d'événements extrêmes. En appliquant des méthodes de correction, en particulier le modèle de forêt aléatoire (Random Forest) et Stacking, nous avons observé une amélioration significative de la correspondance entre les prévisions des réanalyses et les valeurs mesurées. Ces résultats confirment que les techniques de correction de biais sont essentielles pour une modélisation hydrologique dans les contextes où les observations sont rares ou incomplètes. Par ailleurs, cette étude met en avant le potentiel des techniques d'intelligence artificielle, telles que les réseaux de neurones LSTM, les modèles Random Forest et Stacking pour la modélisation des processus hydrologiques complexes. Leur capacité à capter les dépendances temporelles et à s'adapter aux variations locales en fait des outils puissants pour des prévisions robustes dans des environnements hydrologiques divers. Ces résultats ouvrent la voie à l'utilisation de méthodes d'intelligence artificielle dans la gestion des ressources en eau et la prévention des risques liés aux événements extrêmes, contribuant ainsi à une meilleure résilience des systèmes hydrologiques face aux changements climatiques. Les méthodes d'intelligence artificielle utilisées, comme les modèles LSTM et le Stacking, ont démontré leur efficacité en améliorant respectivement la précision des prévisions hydrologiques et la correction de biais. Dans le cadre de notre étude sur l'amélioration des prévisions hydrologiques, deux axes de développement se démarquent : l'intégration des images satellitaires à haute résolution pour affiner la modélisation des crues et optimiser la précision des modèles, ainsi que le déploiement opérationnel des systèmes prédictifs en partenariat avec les agences locales. Cette collaboration permettrait d'incorporer les modèles dans des systèmes d'alerte précoce pour renforcer les dispositifs de résilience face aux crues dans les zones à risque.

REFERENCES BIBLIOGRAPHIQUES

- Ababou, R., & Al-Bitar, A. INP/ENSEEIH : Hydrologie statistique 2005-06.
- Adolphe, AHONNON. (2016). « Impacts socio-économiques et environnementaux des inondations sur le développement de la Commune d'Athiémé (Bénin) ». Repéré à https://www.researchgate.net/publication/370042765_Impacts_socio-economiques_et_environmentaux_des_inondations_sur_le_developpement_de_la_Commune_d'Athieme_Benin
- Amoussou, E., Camberlin, P., & Mahé, G. (2012). Impact de la variabilité climatique et du barrage Nangbéto sur l'hydrologie du système Mono-Couffo (Afrique de l'Ouest). *Hydrological Sciences Journal*, 57(4), 805–817. <https://doi.org/10.1080/02626667.2011.643799>.
- Arsenault, R., Poulin, A., Côté, P., & Brissette, F. (2014). « Comparaison d'algorithmes d'optimisation stochastique dans l'étalonnage de modèles hydrologiques ». *Journal of Hydrologic Engineering*, 19(7), 1374-1384. doi: 10.1061/(ASCE)HE.1943-5584.0000938.
- Bodjrenou, R., & Comandan, F. (2023a). « Influence de la dynamique d'occupation du sol sur les ressources en eau au Bénin (Afrique de l'Ouest) ». *lhb: hydrosience journal*. doi: <https://doi.org/10.1080/27678490.2023.2233482>
- Bodjrenou, R., Cohard, J.-M., Hector, B., Lawin, EA, Chagnaud, G., Danso, DK, & Ahamide, B. (2023b). « Evaluation des estimations de réanalyse des précipitations, du rayonnement et de la température au Bénin (Afrique de l'Ouest) ». *Journal of Applied Meteorology and Climatology*, 1005-1022. Repéré à <https://hess.copernicus.org/preprints/hess-2023-311/hess-2023-311.pdf>
- Bodjrènou, R., Sintondji, LO, & Comandan, F. (2023b). « Modélisation hydrologique à base de modèles physiques dans le bassin de l'Ouémé : enjeux et perspectives pour l'optimisation des simulations ». *Journal of Hydrology: Regional Studies*, 48. doi: 10.1016/j.ejrh.2023.101448.
- Bodjrenou, René, Luc Ollivier Sintondji, Yekambessun M'Po N'Tcha, Diane Germain, Francis Esse Azonwade, Aymar Yaovi Bossa, Silvère Fernand Sohindji, Gilbert Hounnou, Edid Amouzouvi, et Arthur Freud Segnon Kpognin. (2024). « Évaluation des estimations hydrologiques issues des réanalyses ERA5 au Bénin (Afrique de l'Ouest) ». *Discussions sur l'hydrologie et les sciences du système terrestre*, 1–23. est-ce que je: 10.5194/hess-2023-311.

- Choi, W., Kim, SJ, Rasmussen, PF, & Moore, AR (2009). « Utilisation de la réanalyse régionale nord-américaine pour la modélisation hydrologique au Manitoba ». *Revue canadienne des ressources en eau* , 34(1), 17-36. doi: 10.4296/cwrj3401017.
- Cover, TM, et Thomas, JA (2006). *Éléments de théorie de l'information* (2e éd.). Wiley-Interscience.
- Dallaire, G. (2019). « Étude de l'incertitude liée à la modélisation de l'évapotranspiration potentielle dans les études d'impact des changements climatiques sur les ressources en eau au Canada, aux États-Unis et au Mexique ». École de technologie supérieure, Montréal.
- Dehghani, A., et al. (2023). « Évaluation comparative de LSTM, CNN et ConvLSTM pour la prévision horaire du débit à court terme à l'aide d'approches d'apprentissage profond ». *Ecological Informatics* , 75, 102119. doi: 10.1016/j.ecoinf.2023.102119.
- Dembélé, M., Schaepli, B., Giesen, Van De. et Mariéthoz, G. (2020). « Adéquation de 17 jeux de données maillées sur les précipitations et les températures pour la modélisation hydrologique à grande échelle en Afrique de l'Ouest ». *Hydrologie et sciences du système terrestre* , 5379-5406. Repéré à <https://hess.copernicus.org/preprints/hess-2023-311/hess-2023-311.pdf>
- Ernest AMOUSSOU, Anon. 2015. « Application d'un modèle conceptuel à l'analyse de la dynamique hydrométéorologique des crues dans un bassin-versant en milieu tropical humide : cas du fleuve Mono ».
- Essou, GRC, Brissette, F., & Lucas-Picher, P. (2017). « Utilisation de réanalyses et d'observations maillées comme données météorologiques d'entrée pour un modèle hydrologique : comparaison des performances des débits fluviaux simulés en fonction de la densité des stations météorologiques ». *Journal of Hydrometeorology* , 18(2), 497-513. doi: 10.1175/jhm-d-16-0088.1.
- Essou, GRC, Sabarly, F., Lucas-Picher, P., Brissette, F., & Poulin, A. (2016). « Les précipitations et la température issues de réanalyses météorologiques peuvent-elles être utilisées pour la modélisation hydrologique ? ». *Journal of Hydrometeorology* , 17(7), 1929-1950. doi: 10.1175/jhm-d-15-0138.1.
- Filikibirou, Tassou Zakari, Ibidon Firmin Akpo, Kassimou Issaka, Francis Oninkitan Agani, Jacob Afouda Yabi. (2022). « Typologie des systèmes de production de maïs (*Zea mays* L.) dans un contexte de changement climatique au Nord Bénin ». *Revue Africaine d'Environnement et d'Agriculture* .

- Habibi, B., Meddi, M. et Boucefiane, A. (2013). « Analyse fréquentielle des pluies journalières maximales Cas du Bassin Chott-Chergui ». *Nature & Technologie* , (08), 41-48.
- Hounnondaho Freddy Zinsè. « Utilisation de l'Apprentissage Automatique et des Données Satellitaires pour l'Amélioration de la Prédiction des Crues : Cas du Bassin de l'Ouémé à Bétérou ».
- Jenkins, JM, Kowalski, M., & Alvarenga, EFS (2018). « Modélisation prédictive des pertes en eau à l'aide de forêts aléatoires sur les covariables météorologiques ». *Water Supply* , 18(6), 2180-2187.
- Johnson, C., Jones, R., & Thompson, S. (2019). « Une comparaison des techniques d'apprentissage automatique pour la prévision des débits ». *Journal of Hydrology* , 568, 1058-1072.
- Klassou, SD (1996). Evolution climato-hydrologique récente et conséquences sur l'environnement : l'exemple du bassin versant du fleuve Mono (Togo-Bénin) . *Thèse de Doctorat, Université de Bordeaux III*, 472 p.
- Koumassi. (2019). Caractérisation spatiale du risque d'inondation dans le bassin versant de la Volta au Bénin . Thessalonique - Grèce.
- Lévesque, Guillaume. sd « école de technologie supérieure université du québec ».
- Maraun, D. (2016). « Simulations de changement climatique avec correction des biais - Une revue critique ». *Current Climate Change Reports* , 2(4), 211-220. doi: <https://doi.org/10.1007/s40641-016-0050-x>
- Mason, S., & Graham, N. E. (1999). *Conditional probabilities, relative operating characteristics, and relative operating levels. Weather and Forecasting.*
- Métivier, F. (2010). *Hydrologie continentale : Notes de cours, problèmes et solutions* .
- Murphy, KP (2012). *Apprentissage automatique : une perspective probabiliste* . MIT Press.
- Obada, Ézéchiél. (2016). « Évaluation des Performances de Neuf (09) Modèles Hydrologiques Pluie-débit Globaux sur le Bassin de la Mékrou à L'exutoire de Kompongou (Bénin) », 140, 411-424. Repéré à [[https ://www .researchgate .net /publication /318468022 Evaluation des Performances de Neuf 09 Modeles Hydrologiques Pluie - debit Globaux sur le Bassin de la Mekrou a L'exutoire de Kompongou Benin](https://www.researchgate.net/publication/318468022_Evaluation_des_Performances_de_Neuf_09_Modeles_Hydrologiques_Pluie_-_debit_Globaux_sur_le_Bassin_de_la_Mekrou_a_L'exutoire_de_Kompongou_Benin) .
- Pierre, Robyns, et al. (2019). *Hydrologie : principes, analyse et applications*
- ZOGO Lucrèce Xavière (2023) « Evaluation des produits du ECMWF en termes de performances prévisionnelles des pluies extrêmes sur le bassin de l'Ouémé »

TABLE DES MATIERES

DEDICACE.....	ii
REMERCIEMENT	iii
RESUME.....	iv
ABSTRACT	v
LISTE DES SIGLES ET ACRONYMES	vi
LISTE DES TABLEAUX.....	viii
LISTE DES FIGURES.....	ix
CHAPITRE 1 :	1
INTRODUCTION GENERALE.....	1
1.1-Contexte et justification de l'étude	1
1-2 Revue de littérature	2
1-2-1 Clarification conceptuelle.....	2
1-2-2 Synthèse bibliographique	5
2. But de la recherche	5
3. Objectifs du mémoire	5
3.1 Objectif général	5
3.2 Objectifs spécifiques	6
4. Questions de recherche.....	6
5. Les hypothèses de recherche	6
6. Structuration du mémoire.....	6
CHAPITRE 2 :	8
DONNEES, MATERIELS, METHODES ET SECTEUR D'ETUDE	8
2.1 Présentation du milieu d'étude	8
2.1.1 Localisation géographique	8
2.1.2 Relief	9
2.1.3 Végétation et sol.....	11
2.1.4 Climat et réseaux hydrographiques	11
2.1.5 Démographie et activités socio-économiques	11
2.2 Données utilisées.....	11
2.3 Matériels utilisés	12
2-4 Méthodes et critères d'évaluation de la performance des simulations.....	12

2-4-1 Critères d'évaluation de la performance des simulations.....	12
2-4-1-1 Critères déterministes.....	12
2-4-1-2 Critères probabilistes.....	15
2-5-2 Les méthodes de corrections.....	17
2-5-2-1 La régression linéaire	17
2-5-2-2 Régression polynomiale	17
2-5-2-3 Méthode Quantile mapping.....	18
2-5-2-4 Random forest	21
2-5-2-5 Support Vector Regression (SVR)	22
2-5-2-6 Stacking.....	23
2-5-2-7 Les réseaux de neurones artificiels.....	24
2-5-3-1 Les éléments principaux de l'architecture LSTM.....	25
2-5-3-2 Les fonctions d'activation.....	25
2-5-3-3 Hyperparamètres utilisés dans ce modèle :	26
2-5-3-4 Étapes d'implémentation.....	27
2-5-3-5 Structure du modèle développé :.....	27
2-5-4 Les lois d'ajustement	29
CHAPITRE 3 :	30
RESULTATS ET DISCUSSIONS.....	30
3-1 Evaluation de la qualité des méthodes de correction par les critères d'évaluation.....	30
3-2 Analyse fréquentielle des débits de crues (Q95).....	32
3-2-1. Tests d'hypothèses, de stationnarité et d'homogénéité	32
3-2-2 Ajustement des lois aux Q95	33
3-2-2-1 Ajustement sur critères des Q95.....	33
3-2-2-2 Ajustement graphique des Q95	34
3-2-2-3 Estimation des quantiles des séries de débits de crues étudiées.....	35
3-3 Développement d'un modèle LSTM pour la prévision des débits journaliers.	36
3-3-1 Convergence du MSE en fonction des époques :	36
3-3-2 Performance des prévisions pour Chaque Jour.....	37
3-3-3 Evaluer de la performance du modèle en termes de sensibilité et de spécificité.....	39
3-3-4 Graphique des résultats du Modèle pour Chaque Jour de Prédiction.....	40
3-4 Discussion	42
CONCLUSION ET PERSPECTIVES	43
.....	43