Développement de modèles d'apprentissage automatisé pour la prévision hydrologique d'ensemble au lac Saint-Jean

par

Marc-Alexandre MORIN

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE
LA MAITRISE EN GÉNIE CONCENTRATION PERSONNALISÉE
M. Sc. A.

MONTRÉAL, LE 17 AVRIL 2025

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Richard Arsenault, directeur de mémoire
Département de génie de la construction à l'École de technologie supérieure

M. François Brissette, président du jury
Département de génie de la construction à l'École de technologie supérieure

M. Jean-Luc Martel, membre du jury
Département de génie de la construction à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 15 AVRIL 2025

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire.

En premier lieu, je tiens à exprimer ma profonde gratitude à mon directeur de recherche, Richard Arsenault, pour son encadrement exceptionnel et ses précieux conseils. Son expertise, sa disponibilité et son enthousiasme ont fait de cette expérience de recherche un véritable plaisir, transformant chaque étape, aussi intense soit-elle, en un apprentissage enrichissant et stimulant.

Enfin, je tiens à adresser mes remerciements les plus sincères à ma conjointe et à ma famille pour leur soutien inconditionnel, leur patience et leur encouragement tout au long de cette aventure.

À tous, je vous suis infiniment reconnaissant.

Développement de modèles d'apprentissage automatisé pour la prévision hydrologique d'ensemble au lac Saint-Jean

Marc-Alexandre MORIN

RÉSUMÉ

L'objectif de cette étude est de développer et d'évaluer des modèles d'intelligence artificielle avancés afin d'améliorer la prévision des volumes de crue dans un contexte de gestion des réservoirs hydroélectriques. En se concentrant sur le bassin versant du Lac-Saint-Jean, cette recherche vise à optimiser la précision des prévisions hydrologiques sur un horizon de 15 jours, en intégrant des approches ensemblistes. Pour ce faire, elle exploite la réanalyse atmosphérique ERA5 en plus des données de prévisions météorologiques archivées du centre ECMWF, permettant ainsi de mieux anticiper les variations hydrologiques associées aux événements météorologiques d'importance pour l'hydrologie locale, notamment la fonte des neiges et les précipitations extrêmes.

La méthodologie adoptée repose sur la comparaison de deux architectures principales, soit un modèle encodeur-décodeur standard à base de LSTM et une version similaire mais qui intègre un mécanisme d'attention. Ces modèles ont été conçus pour capter les dynamiques hydrologiques complexes en s'appuyant sur un apprentissage supervisé. L'évaluation repose sur des métriques telles que le KGE et le CRPS, ainsi que sur les diagrammes de Talagrand, qui permettent d'analyser la précision et la fiabilité des prévisions d'ensemble ainsi générées. En complément, l'étude explore l'impact du transfert d'apprentissage. Cette approche consiste à pré-entraîner le modèle sur un long historique de données avant de l'adapter à des données de prévision, ce qui favorise une meilleure généralisation.

L'association du mécanisme d'attention et du transfert d'apprentissage permet une réduction des erreurs, avec un KGE maintenu à 0,91 au 15^{e} jour. L'effet du transfert d'apprentissage est particulièrement marqué lors des crues printanières, où la dynamique des débits est plus complexe. L'ajout du mécanisme d'attention apporte un gain supplémentaire, mais son impact varie selon les saisons. Il est bénéfique en hiver et en automne, tandis qu'il peut accentuer la variabilité des prévisions au printemps et en été. Si les modèles offrent une bonne précision en moyenne, les diagrammes de Talagrand révèlent un défaut de calibration, caractérisé par une sous-dispersion des prévisions, en particulier à mesure que l'horizon temporel s'allonge. Cela indique que les modèles sont trop confiants, car ils sous-estiment l'incertitude associée à leurs prévisions.

En somme, cette étude met en avant les bénéfices du transfert d'apprentissage et de l'attention pour la prévision hydrologique, tout en soulignant la nécessité d'améliorer la fiabilité et la calibration des prévisions d'ensemble, ouvrant ainsi la voie à de futures recherches.

Mots-clés : Prévision hydrologique, LSTM encodeur-décodeur, Mécanisme d'attention, Intelligence artificielle, Transfert d'apprentissage

Development of machine learning models for ensemble hydrological forecasting on the lac Saint-Jean basin

Marc-Alexandre MORIN

ABSTRACT

The objective of this study is to develop and evaluate advanced artificial intelligence models to improve flood volume forecasting in the context of hydroelectric reservoir management. Focusing on the Lac-Saint-Jean watershed, this research aims to optimize the accuracy of 15 day hydrological forecasts by integrating ensemble-based approaches. To achieve this, it leverages historical and meteorological data from sources such as ERA5 and ECMWF, enabling better anticipation of hydrological variations associated with climatic events, particularly snowmelt and extreme precipitation.

The adopted methodology is based on the comparison of two main architectures: a standard encoder-decoder model and a version incorporating an attention mechanism. These models are designed to capture complex hydrological dynamics through supervised learning. The evaluation relies on metrics such as the KGE and CRPS, as well as Talagrand diagrams to assess the calibration of probabilistic forecasts. Additionally, this study explores the impact of transfer learning, which involves pre-training the model on a long historical dataset before fine-tuning it with more recent data, thus enhancing generalization.

The combination of the attention mechanism and transfer learning leads to a reduction in errors, with a KGE maintained at 0.91 on day 15. The effect of transfer learning is particularly pronounced during spring floods, when streamflow dynamics are more complex. The addition of the attention mechanism provides an additional gain, although its impact varies by season. It proves beneficial in winter and autumn, while it can increase forecast variability in spring and summer. Although the models perform well on average, Talagrand diagrams reveal a calibration issue, characterized by an under-dispersion of forecasts, especially as the lead time increases. This indicates that the models are overly confident, as they underestimate the uncertainty associated with their predictions.

Overall, this study highlights the benefits of transfer learning and attention for hydrological forecasting, while also emphasizing the need to enhance the reliability and calibration of ensemble predictions, paving the way for future research.

Keywords: Hydrological forecasting, LSTM encoder-decoder, Attention mechanism, Artificial intelligence, Transfer learning

TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

Page

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ANN	Réseaux de neurones artificiels
CRPS	Score de probabilité ordonnée continue
ECMWF	Centre Européen pour les Prévisions Météorologiques à Moyen Terme
ED	Encodeur-Décodeur
ED-MM	Encodeur-Décodeur Multi-membres
EPS	Ensemble Prediction Systems
ERA5	Réanalyse de données climatiques par le ECMWF
GRN	Mécanisme de porte
IA	Intelligence Artificielle
KGE	Efficacité Kling-Gupta
LSTM	Mémoire à court et long terme
MHA	Attention multi-tête
MSE	Erreur de la moyenne au carré
RNN	Réseau de Neurones Récurrent
Seq2Seq	Séquence à Séquence
Seq2Vec	Séquence à Vecteur
TA	Transfert d'apprentissage
TF	Teacher Forcing
TFT	Transformeur de fusion temporelle

LISTE DES SYMBOLES ET UNITÉS DE MESURE

J/m^2	Joules par mètre carré
K	Kelvin
m	Mètre
mm	Millimètre
m/s	Mètres par seconde
m3/s	Mètres cube par seconde
Pa	Pascal

INTRODUCTION

La production d'aluminium est un procédé énergivore qui nécessite un approvisionnement en électricité considérable. Rio Tinto, qui exploite une raffinerie d'aluminium dans la région du Saguenay–Lac-Saint-Jean au Québec, dépend directement d'une centrale hydroélectrique pour alimenter son processus industriel. Comme toute centrale hydroélectrique efficace, la gestion de l'eau permet d'assurer une production d'énergie fiable. Cependant, la gestion des réservoirs hydroélectriques dans ce secteur est particulièrement complexe en raison de la variabilité naturelle du climat et du caractère chaotique des processus atmosphériques. En effet, le bassin versant du Lac-Saint-Jean est soumis à d'importantes fluctuations des débits d'eau, principalement causées par la fonte des neiges au printemps, rendant la prévision des volumes d'apports difficile.

Cette gestion est nécessaire pour deux raisons principales. D'une part, l'efficacité de la production hydroélectrique repose sur l'optimisation du niveau des réservoirs, puisque la puissance générée par une turbine est proportionnelle à la hauteur de chute de l'eau. D'autre part, la capacité d'emmagasinement étant limitée et les capacités de déversement restreintes obligent l'anticipation de ces apports afin d'éviter les débordements et de préserver l'intégrité des infrastructures du barrage. Ainsi, un équilibre doit être maintenu entre le débit turbiné, l'apport en eau et la hauteur de retenue afin d'assurer à la fois la performance énergétique et la sécurité des infrastructures.

Pour relever ces défis, les hydrologues développent des modèles visant à représenter la dynamique d'emmagasinement de l'eau dans les réservoirs et à anticiper les variations de débit. Ces approches intègrent des données météorologiques et hydrologiques pour estimer les apports en eau et ajuster les décisions de déversement et de turbinage en tenant compte des incertitudes associées aux prévisions. Parallèlement aux modèles classiques, de nouvelles méthodes basées sur l'intelligence artificielle (IA) émergent et leur intégration en hydrologie se fait de manière tout à fait naturelle. En effet, l'hydrologie consiste à mettre en relation des données d'intrant, telles que des variables météorologiques, avec une donnée d'extrant, le

débit, afin d'établir un modèle représentant un volume de contrôle, en l'occurrence, un bassin versant, à l'aide d'un ensemble de règles. C'est exactement le principe de l'intelligence artificielle qui vise à déterminer des relations mathématiques reliant des données en entrée à une sortie. Cette convergence permet ainsi de développer des modèles performants pour optimiser la gestion des réservoirs hydroélectriques.

Ce document est structuré en cinq chapitres qui suivent une progression allant de la théorie aux résultats et à leur interprétation. Il commence par une revue de littérature qui présente les concepts liés à la prévision hydrologique et aux approches d'IA. Ensuite, la zone d'étude et les données utilisées sont décrites afin de poser le cadre de l'analyse. La méthodologie dans le troisième chapitre expose les modèles développés et les techniques employées pour améliorer la prévision des débits. Les résultats obtenus sont ensuite analysés et comparés à l'aide de différentes métriques afin d'évaluer leur performance. Enfin, une discussion met en perspective les contributions du travail en abordant l'impact des mécanismes d'apprentissage utilisés et leurs limites.

CHAPITRE 1

REVUE DE LITTÉRATURE

1.1 Prévision hydrologique

La prévision hydrologique est essentielle pour la gestion hydrique, que ce soit pour prévenir les inondations ou optimiser la production hydroélectrique. Pour ce faire, la prévision dynamique des débits repose sur deux étapes selon Troin, Arsenault, Wood, Brissette, & Martel (2021). Tout d'abord, les modèles hydrologiques sont calibrés puis assimilés en fonction des données hydrométéorologiques afin d'obtenir des conditions initiales qui représentent l'état actuel du bassin versant. Une fois ces états initiaux établis, les prévisions météorologiques des prochains jours sont intégrées au modèle, qui simule ensuite l'évolution des débits futurs. Différentes approches peuvent être mises en œuvre pour élaborer un modèle hydrologique. Cette section a pour objectif de présenter les deux principales catégories de prévision hydrologique, à savoir les approches de prévision hydrologique déterministe et ensembliste.

1.1.1 Prévision déterministe

Une prévision hydrologique déterministe est une approche qui, à partir de conditions initiales spécifiques et de données météorologiques, produit une unique prévision des débits futurs. Fan, Schwanenberg, Alvarado, Assis dos Reis, Collischonn et Naumman (2016) ont montré que ces modèles étaient efficaces pour la gestion à court terme des réservoirs hydroélectriques. Leur étude sur le réservoir de Três Marias au Brésil indique que sur un horizon de 4 jours, les prévisions déterministes permettent une optimisation fiable des débits.

Cependant, ils ont observé que la fiabilité des prévisions déterministes diminuait rapidement au-delà de 4 jours, en raison de l'augmentation de l'incertitude météorologique. Ce constat rejoint les conclusions de Boucher, Tremblay, Delorme, Perreault, & Anctil, (2012), qui ont comparé les performances des prévisions déterministes et ensemblistes pour la gestion hydroélectrique. Bien que les prévisions déterministes, notamment celles à haute résolution,

offrent généralement de meilleures performances que les modèles probabilistes en optimisant le débit turbiné et la production énergétique, leur précision se détériore avec l'augmentation de l'horizon de prévision. En effet, à partir de 48 heures, les prévisions probabilistes deviennent plus performantes, offrant une meilleure capacité d'anticipation des apports hydriques et de leur incertitude, permettant ainsi une gestion plus efficace des réservoirs.

Bien que cette approche soit utile dans le cadre de la prise de décision pour la gestion des ressources en eau, elle présente une limitation majeure. Cette limitation s'explique par le fait que les prévisions déterministes reposent sur un seul scénario, ce qui empêche de représenter l'incertitude associée au processus de prévision (Troin et al., 2021).

1.1.2 Prévision d'ensemble

La prévision d'ensemble a été développée pour mieux évaluer l'incertitude en générant plusieurs scénarios possibles de débit futur en variant les conditions initiales ou les paramètres d'un modèle hydrologique. Contrairement à la prévision déterministe, elle permet d'estimer la probabilité d'atteindre certains seuils critiques, offrant ainsi une vue globale des risques. Les *Ensemble Prediction Systems* (EPS), qui sont aujourd'hui considérés comme l'état de l'art dans ce domaine (Cloke & Pappenberger, 2009), fournissent des prévisions plus adaptées à la prise de décision opérationnelle. Plutôt que de s'appuyer sur une seule prévision déterministe, cette approche repose sur l'exécution de plusieurs simulations avec des conditions initiales légèrement différentes afin d'explorer une gamme de scénarios possibles.

Boucher et al. (2012) démontrent que les prévisions d'ensemble surpassent les prévisions déterministes dans la gestion des barrages hydroélectriques en offrant une meilleure prise en compte des incertitudes, une optimisation de la production énergétique et une réduction des risques d'inondation et de pertes d'eau. Ces résultats justifient l'adoption des EPS dans la gestion des ressources en eau en les positionnant comme une solution incontournable pour les opérateurs hydroélectriques. De même, Wetterhall et al. (2013) confirment que ces systèmes permettent une meilleure anticipation des apports en eau, optimisent la production énergétique

et réduisent les pertes par déversement. Ainsi, en intégrant les incertitudes et en optimisant la prise de décision, les EPS s'imposent comme approche pour la gestion hydroélectrique moderne.

1.2 Évaluation de la qualité des prévisions

Cette section présente les principales métriques d'évaluation utilisées dans le domaine de l'hydrologie, en mettant en avant leurs particularités. Chaque métrique offre des avantages spécifiques pour évaluer la performance des modèles de prévision. Les sous-sections suivantes détaillent les différentes mesures, soit l'efficacité de Kling-Gupta (KGE), l'erreur moyenne quadratique (MSE), le score de probabilité ordonnée continue (CRPS) et les diagrammes de Talagrand.

1.2.1 KGE

Le KGE est une métrique standard largement utilisée en hydrologie. Contrairement à d'autres métriques plus unidimensionnelles, comme le NSE ou la RMSE, le KGE offre une évaluation plus globale de la qualité des prévisions déterministes. En effet, il prend en compte la corrélation, la variabilité relative et le biais. Concrètement, le KGE mesure la performance d'un modèle en calculant la distance euclidienne entre ces trois composantes.

$$KGE = 1 - \sqrt[2]{(r-1)^2 + (\gamma-1)^2 + (\beta-1)^2} \tag{1.1}$$

La première variable r, est le coefficient de corrélation de Pearson. Ensuite, la variable γ est un ratio de la variabilité $((\sigma_{q_{sim}}/\mu_{q_{sim}})/(\sigma_{q_{obs}}/\mu_{q_{obs}}))$ qui mesure la correspondance entre la dispersion des débits simulés q_{sim} et observés q_{obs}. Finalement, la variable β est le ratio de biais $(\mu_{q_{sim}}/\mu_{q_{obs}})$ indiquant si le modèle a tendance à surestimer ou sous-estimer les valeurs prédites (Kling, Fuchs, & Paulin, 2012). Dans le cas des prévisions d'ensemble, une moyenne est calculée sur tous les membres pour chaque journée de prévision avant d'y calculer la métrique.

1.2.2 MSE

Le MSE est une métrique très utilisée dans le domaine de l'hydrologie (Gupta, Kling, Yilmaz, & Martinez, 2009). Elle est définie par l'équation suivante:

$$MSE = \frac{\sum_{i=1}^{N}\left(q_{sim_i} - q_{obs_i}\right)^2}{N} \tag{1.2}$$

Où N est le nombre de données, q_{sim_i} représente les valeurs de débits simulées et q_{obs_i} les valeurs observées. Le MSE mesure la moyenne des carrés des écarts entre les valeurs prévues et observées, donnant ainsi plus de poids aux erreurs importantes. Tout comme pour le KGE, avant d'appliquer cette équation, une moyenne des prévisions d'ensemble est effectuée afin d'obtenir une valeur unique de MSE par jour de prévision. Comme pour le KGE, la moyenne des membres de l'ensemble est utilisée pour calculer le MSE de manière déterministe, en fournissant une valeur unique de débit simulé par jour à comparer aux observations.

1.2.3 Score de probabilité ordonnée continue (CRPS)

Le CRPS est une métrique adaptée aux prévisions probabilistes. Il mesure la distance moyenne entre la fonction de densité de probabilité observée et celle des prévisions d'ensemble (Hersbach, 2000). Mathématiquement, il est exprimé par l'équation suivante:

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\left[F\left(q_{sim_i}\right) - F(q_{obs})\right]^2 dq \tag{1.3}$$

Le CRPS varie de 0 à +∞, une valeur de zéro indiquant une prévision parfaite. Par ailleurs, ses unités sont les mêmes que celles de la variable utilisée, soit le m³/s pour les prévisions de débit. Il est généralement calculé indépendamment pour chaque journée de prévisions d'ensemble et pour chaque délai de prévision (*lead-time*), puis moyenné sur l'ensemble des prévisions pour chaque *lead-time*. Ainsi, cette métrique permet d'évaluer la précision et la fiabilité des

prévisions probabilistes, en analysant la dispersion des ensembles et en pénalisant les écarts importants par rapport aux observations.

1.2.4 Diagrammes de Talagrand

Le diagramme de Talagrand (O. Talagrand, 1997) est une métrique classique dans le domaine de la prévision hydrologique pour évaluer la performance des prévisions d'ensemble. Il permet d'analyser la distribution des prévisions par rapport aux observations et d'identifier les biais ainsi que les éventuelles sous- ou sur-dispersions du modèle. L'objectif est d'obtenir une distribution uniforme, indiquant un bon calibrage des prévisions. Une sur-dispersion se manifeste par une distribution en forme de cloche (gaussienne), traduisant une trop grande variabilité des prévisions. Enfin, une distribution en U (parabolique positive) révèle une sous-dispersion, où les prévisions manquent de diversité et sont trop concentrées autour de certaines valeurs. Ainsi, le diagramme permet d'évaluer la calibration d'une prévision d'ensemble, c'est-à-dire la capacité de la dispersion des prévisions à bien représenter la probabilité que le débit se réalise.

1.3 Modèles d'intelligence artificielle

Cette section explore différentes approches de prévision des débits en hydrologie. Les données météorologiques, sous forme de séries temporelles, peuvent être traitées efficacement à l'aide de modèles d'apprentissage automatique. L'état de l'art recommande l'utilisation d'architectures de réseaux de neurones récurrents (RNN) avec des cellules mémoire à court et long terme (LSTM), pouvant être enrichis par des modules supplémentaires tels que le mécanisme d'attention ou encore des modèles plus complexes tels que les transformeurs ou les transformeurs de fusion temporelle (TFT). Ces architectures sont présentées dans les sections suivantes.

1.3.1 Réseau de neurones

Les approches traditionnelles de prévision hydrologique reposent sur trois grandes catégories de modèles, soit empiriques, conceptuels et basés sur la physique (Devia, Ganasri, & Dwarakish, 2015). Parmi ceux-ci, les modèles empiriques, fondés sur l'analyse statistique des observations passées, ont évolué avec l'essor des méthodes d'IA. En effet, les techniques d'apprentissage automatique, capturant des relations complexes entre variables, permettent de modéliser un comportement hydrologique sans recourir explicitement aux lois physiques.

Les réseaux de neurones artificiels (ANN) comptent parmi les premières applications de l'apprentissage automatique en hydrologie (Maier & Dandy, 2000). Ces modèles utilisent plusieurs couches de neurones interconnectés avec des poids qui apprennent les relations entre les variables d'entrée, telles que les précipitations, la température et l'humidité du sol, et les variables de sortie, comme le débit ou le niveau d'eau. En hydrologie, le problème à résoudre est principalement un problème de régression, où l'objectif est de prédire une valeur continue, le débit d'une rivière, en fonction des conditions météorologiques antécédentes et à venir.

L'apprentissage d'un réseau de neurones artificiel suit généralement quatre étapes. Tout d'abord, les poids d'apprentissage sont initialisés aléatoirement, attribuant des valeurs aléatoires aux connexions entre les neurones. Ensuite, les données d'entrée sont propagées à travers le réseau pour générer une première estimation des débits, correspondant à la sortie de la première itération. Cette prédiction est ensuite comparée aux valeurs observées à l'aide d'une fonction de coût, telle que le MSE, qui mesure l'écart entre les valeurs prédites et les valeurs réelles. Enfin, un algorithme d'optimisation, tel que via des méthodes de descente du gradient, est utilisé pour mettre à jour les poids du réseau. Ce processus permet de réduire progressivement l'erreur et d'améliorer la précision des prédictions au fil des itérations, en ajustant les paramètres du modèle afin d'optimiser sa capacité d'apprentissage.

Cependant, les ANN ne tiennent pas compte de la structure temporelle des données, ce qui limite leur efficacité pour la prévision hydrologique à long terme. En effet, ces modèles

considèrent chaque instant indépendamment, sans exploiter la dépendance temporelle entre les observations passées et futures. Cette limitation a conduit au développement de modèles plus avancés, tels que les réseaux de neurones récurrents (RNN).

1.3.2 Réseaux de neurones récurrents

Les RNN (Rumelhart, Hinton, & Williams, 1986) représentent une architecture neuronale conçue pour traiter les données séquentielles. Contrairement aux réseaux de neurones classiques, les RNN conservent une mémoire des états précédents grâce à une connexion récurrente tel qu'illustré à la Figure 1.1.

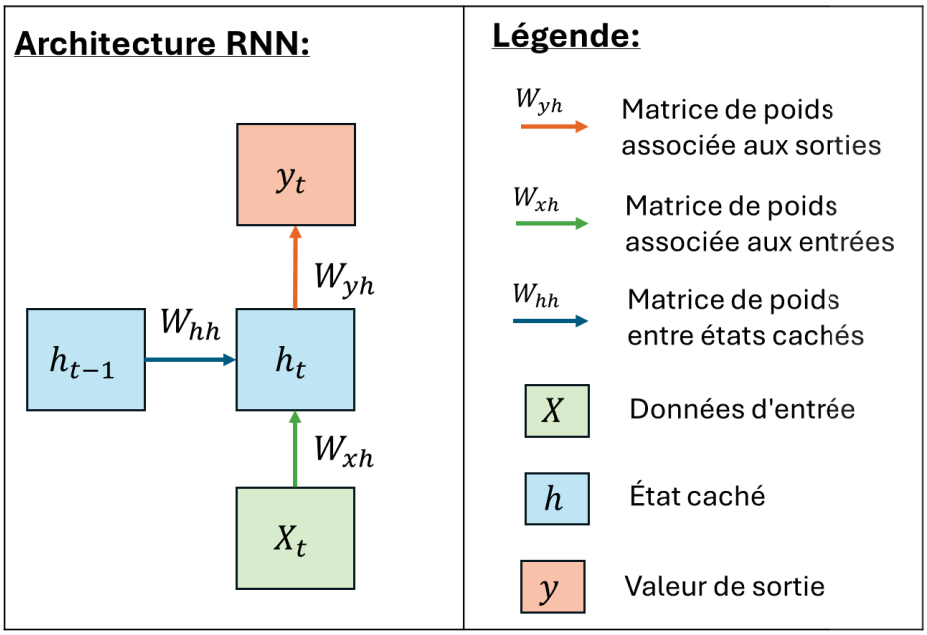

Figure 1.1 Architecture d'un Réseau de Neurones Récurrent où X_t (vert) représente l'entrée, h_t (bleu) est l'état caché et y_t (orange) est la sortie. Les flèches indiquent les matrices de poids reliant les éléments à travers le réseau

L'équation de mise à jour de l'état caché (h_t) dans un RNN décrit comment l'information est propagée d'un instant à l'autre. Elle est définie comme suit :

$$h_t = \tanh(W_{xh}X_t + W_{hh}\, h_{t-1} + b_h) \tag{1.4}$$

Cette équation modélise la dynamique interne du neurone récurrent en combinant l'entrée actuelle X_t et l'état caché précédent h_{t-1}. L'entrée X_t représente la donnée d'entrée du moment présent, comme une mesure de précipitation. Avant d'être combinés par une sommation, ils sont multipliés par une matrice de poids W_{xh} et W_{hh} dont le rôle est d'apprendre les relations entre les vecteurs de données d'entrée et l'état du réseau. D'ailleurs, un biais b_h, est ajouté pour permettre au modèle d'apprendre des décalages spécifiques pour l'hyperplan généré par l'équation.

Une fois cette somme pondérée calculée, une fonction d'activation prenant la forme d'une tangente hyperbolique (*tanh*) est appliquée afin d'y ajouter de la non-linéarité. De plus, cette fonction maintient les valeurs de sortie dans l'intervalle [-1,1], ce qui évite des problèmes liés à l'explosion ou à la disparition du gradient lors de l'apprentissage. L'explosion du gradient amplifie excessivement les poids, rendant la mise à jour instable et provoquant la divergence des poids du modèle, tandis que la disparition du gradient affaiblit progressivement les gradients, réduisant l'influence des états passés sur l'apprentissage.

L'état caché h_t est ensuite utilisé pour générer la sortie du modèle selon l'équation 1.5.

$$y_t = tanh\left(W_{yh}h_t + b_y\right) \tag{1.5}$$

Dans cette équation, W_{yh} est la matrice de poids qui relie l'état caché h_t à la sortie y_t, tandis que b_y est un biais appliqué à la sortie. Finalement, la fonction *tanh* est appliquée pour obtenir la sortie y_t.

Dans le cadre de la modélisation des séries temporelles, trois architectures principales illustrées à la Figure 1.2 sont couramment utilisées pour capturer les relations entre les entrées et les sorties.

La première architecture correspond à un modèle séquence en entrée et séquence en sortie (Seq2Seq), où chaque valeur de la séquence d'entrée est associée à une valeur de la séquence

de sortie. En hydrologie, cela signifie que pour chaque donnée météorologique en entrée, une prévision du débit est générée en sortie. Toutefois, cette approche ne tire pas pleinement partie de l'information la plus pertinente pour la prévision hydrologique, soit le débit observé (Wagener et al., 2001). En effet, comme le souligne la littérature, la variable la plus utile pour estimer le débit d'une journée est le débit de la journée précédente. Ainsi, une architecture mieux adaptée consiste à exploiter directement les valeurs observées récentes du débit pour améliorer la précision des prévisions hydrologiques.

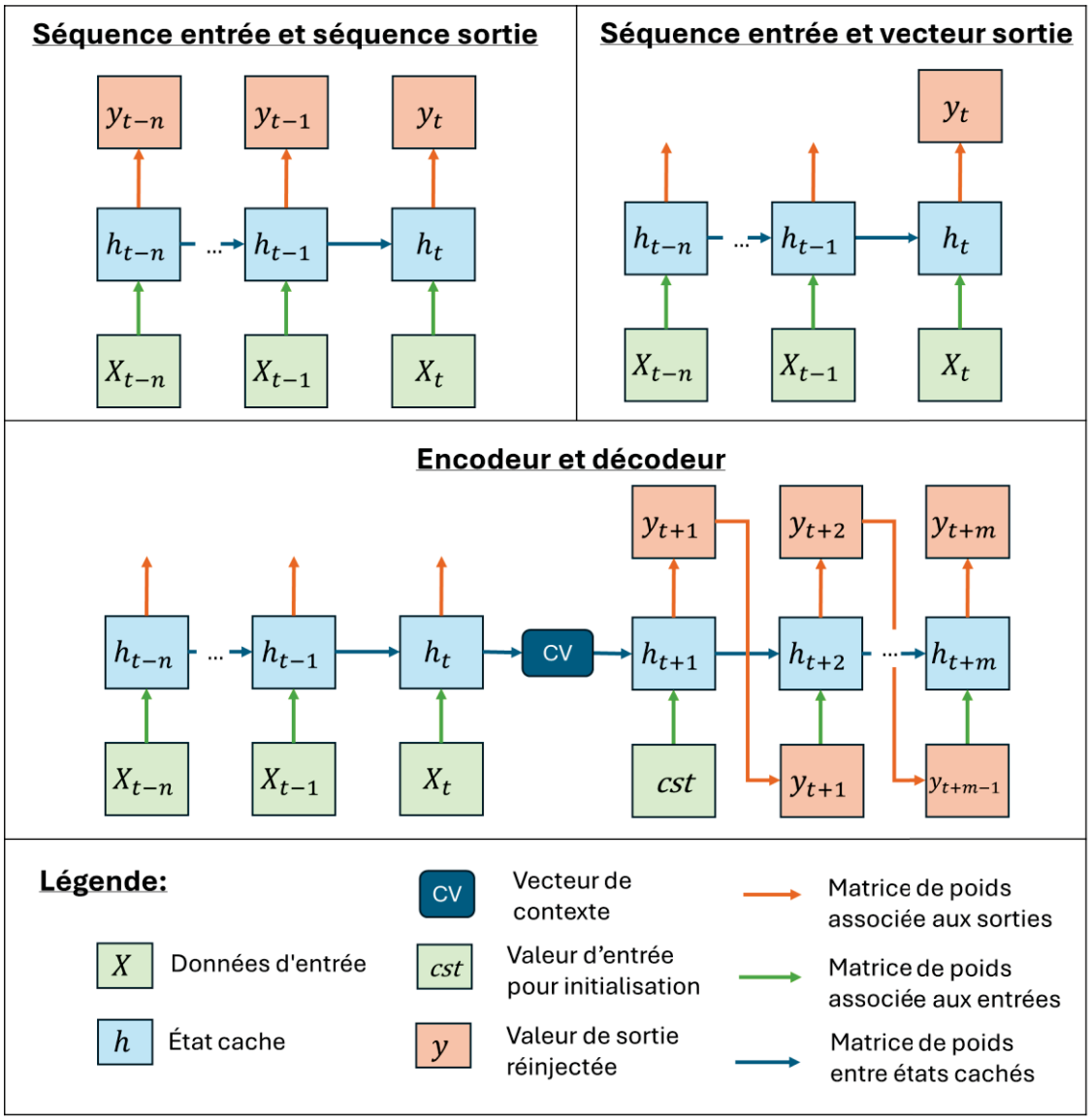

Figure 1.2 Configurations différentes de réseaux de neurones récurrents (RNN) utilisées pour la prévision des séries temporelles : (1) Séquence entrée et séquence sortie (Seq2Seq),

(2) Séquence entrée et vecteur sortie (Seq2Vec), et (3) Architecture Encodeur-Décodeur (ED) avec vecteur de contexte. Quant aux variables, X_t (vert) représente l'entrée, h_t (bleu) est l'état caché et y_t (orange) est la sortie. Les flèches indiquent les matrices de poids et la boite CV est le vecteur de contexte

La seconde architecture repose sur une séquence en entrée avec un vecteur en sortie (Seq2Vec). Cette configuration est particulièrement adaptée aux tâches nécessitant une seule valeur en sortie, comme la classification (Li, Kiaghadi, & Dawson, 2021). Bien qu'elle soit similaire à l'architecture Seq2Seq, la principale différence réside dans le fait que la séquence de sortie complète n'est pas conservée. En effet, seul le dernier vecteur y_t est retenu.

En hydrologie, cette approche permet d'intégrer directement les débits observés dans les entrées du modèle et de produire une seule prévision du débit à la fin de la séquence. Bien que l'objectif soit également de produire une sortie séquentielle, Sabzipour et al., (2023) ont démontré que l'utilisation de cette architecture parallélisée sur 9 jours permettait d'atteindre des résultats très satisfaisants, avec un KGE de 0.88 pour la valeur moyenne des prévisions d'ensemble de la dernière journée. Il convient de préciser que la cellule utilisée était un LSTM, dont le fonctionnement sera abordé dans la section suivante.

La troisième architecture repose sur une architecture encodeur-décodeur (ED), qui est largement utilisée dans les prévisions de séries temporelles lorsque la longueur de la séquence d'entrée ne correspond pas nécessairement à celle de la séquence de sortie. Contrairement aux approches classiques où chaque entrée est directement associée à une sortie (comme dans Seq2Seq), cette architecture introduit une phase d'encodage, suivie d'une phase de décodage. L'encodeur reçoit en entrée une séquence de données passées, telles que les précipitations et les débits observés des jours précédents, et les transforme en une représentation compacte, appelée vecteur de contexte, qui synthétise l'information. Ce vecteur est ensuite transmis au décodeur, qui l'utilise pour générer progressivement les prévisions pour les jours à venir. Dans la phase de décodage, les valeurs d'entrée sont généralement issues des prévisions précédentes du modèle lui-même, ce qui permet de produire une séquence complète de manière itérative.

En somme, cette architecture est très flexible, mais elle présente certains inconvénients, dont le goulot d'étranglement. Cela signifie que toute l'information de la séquence d'entrée doit être compressée dans un unique vecteur avant d'être transmise au décodeur, ce qui peut entraîner des pertes d'information et une dégradation des performances pour les longues séquences.

1.3.3 LSTM

Les RNN classiques présentent une limitation majeure, soit la disparition ou l'explosion du gradient, surtout lorsqu'ils traitent de longues séquences. Pour pallier ce problème, l'architecture de base des RNN a été conservée, mais tel qu'illustré à la Figure 1.3, la cellule a été modifiée pour créer les LSTM (Hochreiter & Schmidhuber, 1997). Ainsi, cette nouvelle structure permet de gérer efficacement les dépendances à long et court terme.

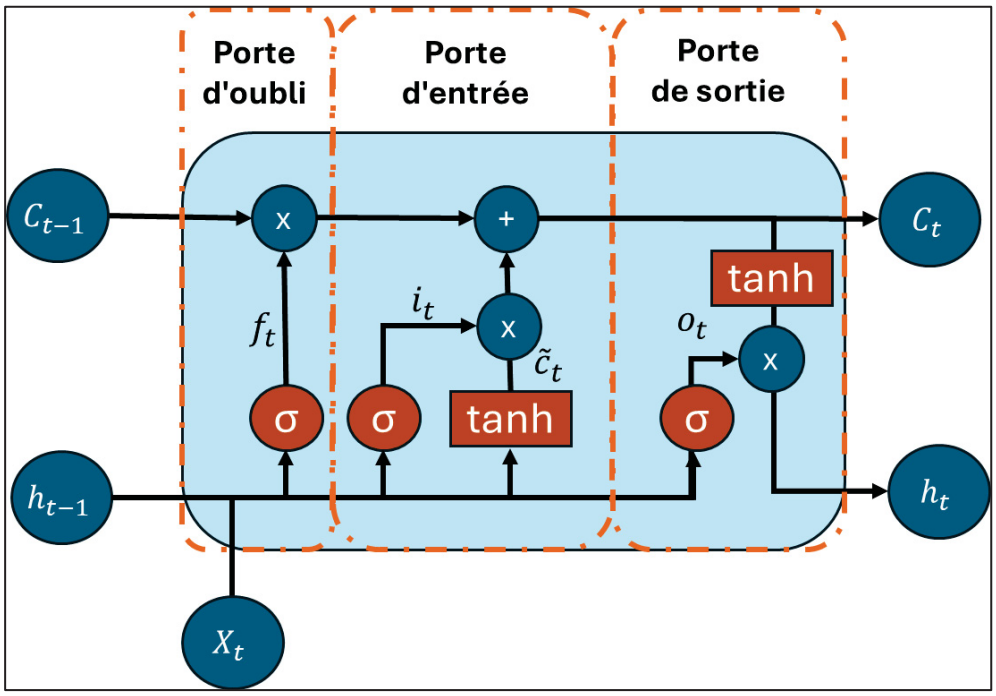

Figure 1.3 Architecture de la cellule LSTM, où X_t est l'entrée, h_t est l'état caché court terme et c_t représente l'état de la mémoire à long terme. La cellule est composée de trois portes : (1) la porte d'oubli, qui contient une activation sigmoïde (σ), (2) la porte d'entrée, qui combine une activation sigmoïde (σ) et une transformation tanh (orange), (3) la porte de sortie qui applique une activation sigmoïde (σ) suivie d'une fonction tanh (orange), pour produire h_t. Les flèches indiquent les interactions entre ces composants, où les multiplications (x) et additions (+) définissent les opérations appliquées aux données

La Figure 1.3 illustre la structure d'une cellule LSTM, composée de trois portes, chacune munie d'une fonction sigmoïde agissant comme un régulateur. Ces portes contrôlent le flux d'information en décidant quelles données doivent être conservées pour l'état suivant. Dans cette architecture, les entrées principales incluent X_t, représentant l'entrée actuelle, h_{t-1}, correspondant à la mémoire à court terme et c_t pour la mémoire à long terme. Les éléments en orange représentent des fonctions d'activation, où σ correspond à la fonction sigmoïde qui contraint la sortie à prendre une valeur entre 0 et 1, indiquant ainsi la proportion d'information conservée.

La porte d'oubli est chargée de déterminer la proportion des informations de la mémoire à long terme à conserver. Son fonctionnement est régi par l'équation 1.6.

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{1.6}$$

Les variables W_f et U_f sont des matrices de poids qui appliquent une transformation linéaire à l'entrée actuelle et à l'état caché précédent, influençant l'activation de cette porte. Elles permettent ainsi au modèle de capturer les relations entre les informations et d'apprendre à ajuster dynamiquement l'importance des données passées et présentes. Ensuite, le biais b_f ajuste l'hyperplan pour affiner cette transformation. Enfin, la matrice résultante est passée à travers une fonction d'activation sigmoïde, introduisant de la non-linéarité.

La porte d'entrée évalue le potentiel des nouvelles informations à intégrer dans la mémoire à long terme. Cette étape est déterminée par l'équation 1.7.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{1.7}$$

Les variables W_f et U_f contrôlent l'interprétation des relations des vecteurs de x_t et h_{t-1} respectivement. Parallèlement, un candidat de mise à jour de la mémoire est généré par l'équation suivante :

$$\tilde{C}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (1.8)$$

Cette transformation applique une non-linéarité aux nouvelles informations avant leur intégration dans la mémoire cellulaire. Ainsi, $\tilde{c}_t$ représente une proposition de mise à jour des informations issues des entrées et des états passés. Ensuite, la mémoire à long terme c_t est mise à jour en combinant les informations retenues par la porte d'oubli et celles ajoutées par la porte d'entrée selon l'équation 1.9.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (1.9)$$

Enfin, la porte de sortie détermine la proportion d'informations issues de la mémoire à long terme c_t qui sera transmise à l'état caché suivant h_t. La matrice o_t joue un rôle de régulation similaire à f_t, mais avec ses propres paramètres d'apprentissage. Cette activation contrôle l'intensité de l'information issue de la mémoire cellulaire et transmise à l'étape suivante.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (1.10)$$

L'état caché final h_t à l'équation 1.11 est obtenu en appliquant une transformation non linéaire à la mémoire cellulaire c_t et en la pondérant par o_t. Ainsi, la sortie finale h_t est déterminée par la fraction de la mémoire à long terme c_t qui doit être transmise à l'état caché et propagée dans le réseau.

$$h_t = o_t \odot \tanh(c_t) \quad (1.11)$$

Dans le domaine de l'hydrologie, plusieurs études ont montré que les LSTM surpassent les RNN classiques en termes de performance (Sahoo, Jha, Singh, & Kumar, 2019 ; Waqas & Humphries, 2024). D'ailleurs, Sabzipour et al. (2023) comparent les performances d'un LSTM avec celles d'un modèle hydrologique basé sur la physique pour la prévision du débit du même bassin. L'étude montre que le modèle LSTM réduit les écarts entre les prévisions et les observations par rapport aux modèles physiques comme CEQUEAU (Charbonneau, Fortin, &

Morin, 1977). Cette réduction de la disparité entre les scénarios prévus et les observations permet d'améliorer la précision des prévisions en rapprochant les débits simulés des débits observés. Ainsi, le biais observé lors des premiers jours de prévision aura un impact moindre sur les résultats à plus long terme.

Enfin, l'utilisation des LSTM en hydrologie se révèle particulièrement efficace. Par exemple, Arsenault, Martel, Brunet, Brissette, & Mai, (2023) ont démontré leur utilité pour la prévision des débits dans les bassins non jaugés. Leur étude sur 148 bassins du nord-est de l'Amérique du Nord montre que les LSTM surpassent les modèles hydrologiques classiques dans plus de 93 % des cas et offrent des prévisions plus précises sur des sites non jaugés dans 78 % des situations. De plus, un autre avantage souligné par les auteurs est que, contrairement aux modèles traditionnels nécessitant une paramétrisation spécifique, les LSTM exploitent directement l'ensemble des données régionales.

1.3.4 Mécanisme d'attention

Une problématique associée à l'architecture ED est le phénomène du goulot d'étranglement. Cette problématique se traduit par une incapacité à encoder toute l'information dans le vecteur de contexte. Le mécanisme d'attention (Bahdanau, Cho, & Bengio, 2016), qui constitue une amélioration de l'architecture ED, pallie ce problème en accumulant la mémoire et en attribuant une importance relative à chaque état caché produit par l'encodeur. Le mécanisme d'attention permet au modèle de se concentrer sur les parties pertinentes de la séquence de l'encodeur lors des prévisions. Pour ce faire, il attribue un poids à chaque élément de la séquence en fonction de son importance pour la prévision à chaque pas de temps. Cela réduit le problème du goulot d'étranglement en accumulant et pondérant les informations qui sont jugées pertinentes par le modèle. Malgré un gain en précision, ce mécanisme ajoute un niveau de complexité en ressources computationnelles.

Girihagama et al. (2022) comparent deux modèles dans leur ouvrage pour la prévision des débits dans 10 bassins versants du Canada, situés dans le bassin de la rivière des Outaouais. Le

premier modèle est un LSTM avec un encodeur et décodeur standard. Le second modèle possède aussi un encodeur et un décodeur, mais il est équipé d'un mécanisme d'attention. Les résultats de cette étude sont assez catégoriques. En effet, le modèle qui intègre le mécanisme d'attention est nettement plus performant. Selon le premier modèle, les résultats des évaluations RMSE et KGE sont dans l'ordre de 40.39 m^3/s et 0.827, respectivement, tandis que le second, modèle était dans l'ordre de 8.2 m^3/s et 0.957.

Ainsi, il est possible de conclure que l'intégration du mécanisme d'attention dans une architecture encodeur-décodeur permet une amélioration de la précision des prévisions. Cette approche ne se limite pas uniquement à une meilleure exactitude des prédictions, mais contribue également à maintenir des performances optimales sur des horizons de prévision plus longs par rapport à un LSTM standard (Dai, Zhang, Nedjah, Xu, & Ye, 2023 ; Girihagama et al., 2022 ; Yan, Chen, Hang, & Hu, 2021).

En effet, l'étude de Dai et al., (2023) propose un modèle LSTM-seq2seq équipé d'un mécanisme d'attention pour la prévision des niveaux d'eau du fleuve Chu. Ce modèle a été comparé à plusieurs approches, dont d'autres modèles LSTM et des variantes en combinant des multi-modèles. Les conclusions montrent que l'ajout du mécanisme d'attention améliore la précision des prévisions, en particulier face aux données bruitées, tout en accélérant la convergence du modèle.

Enfin, Yan et al., (2021) appliquent leur modèle LSTM avec le mécanisme d'attention à des données provenant du bassin versant de Tunxi, afin d'évaluer sa capacité à améliorer la prévision des débits. Comparé aux modèles traditionnels tels que BPNN, SVM et LSTM, leur approche se distingue par une NSE de 0,961, surpassant nettement les autres méthodes testées. Selon cette étude, l'ajout du mécanisme d'attention permet d'améliorer la stabilité des prévisions ce qui est important lorsque les précipitations ont de forte variabilité.

1.3.5 Transformeur

Le modèle transformeur (*Transformer*) a été introduit par le papier révolutionnaire « *Attention is All You Need* » (Vaswani et al., 2017). Ce modèle, illustré à la Figure 1.4, repose sur une architecture ED qui permet d'analyser des séquences sans utiliser de récurrence. Il repose sur deux stratégies pour y parvenir.

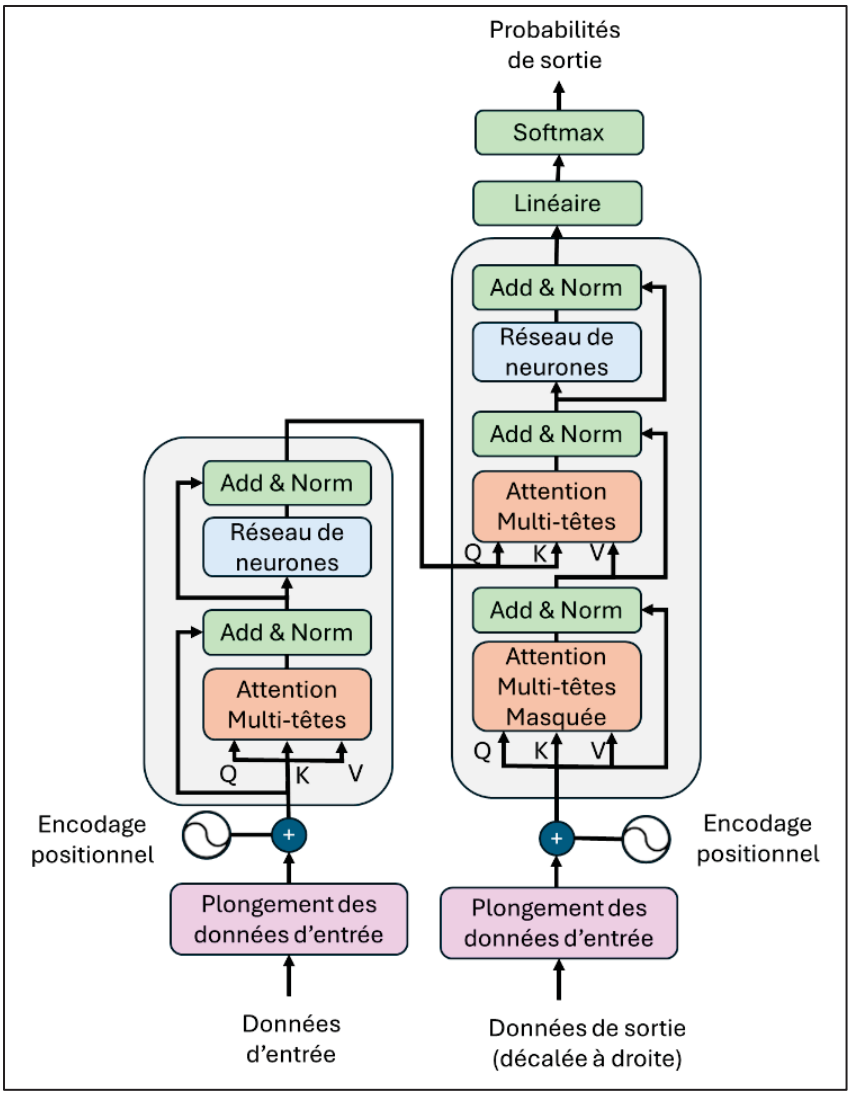

Figure 1.4 Architecture du transformeur comprenant un encodeur et un décodeur, chacun constitué de couches d'attention multi-têtes, de normalisation et de réseaux de neurones. Le décodeur intègre une couche d'attention multi-têtes masquée pour le traitement séquentiel des entrées
Adapté de Vaswani et al., (2017)

La première stratégie est l'encodage positionnel, qui permet au modèle de tenir compte de l'ordre des éléments dans la séquence malgré l'absence de récurrence ou de convolution. Les formules des encodages positionnels sont les suivantes :

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10\,000^{\frac{2i}{d_{model}}}}\right) \tag{1.12}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10\,000^{\frac{2i}{d_{model}}}}\right) \tag{1.13}$$

Ici, pos représente la position d'un élément dans la séquence d'entrée, i correspond à l'indice de la dimension et d_{model} est la dimension d'espace d'intégration, c'est-à-dire la taille du vecteur utilisé pour représenter chaque élément d'entrée dans l'encodage positionnel.

Chaque dimension de l'encodage positionnel est associée à une fonction sinusoïdale, ce qui permet au modèle de capturer les relations séquentielles entre les éléments d'entrée sans dépendre d'une structure récurrente.

La deuxième stratégie implémentée s'agit de l'attention multi-têtes (MHA). Le MHA est une extension du mécanisme d'attention simple, telle que décrite dans la section ci-dessus. Ce mécanisme permet au modèle de se concentrer simultanément sur différentes parties de la séquence en parallèle, augmentant ainsi sa capacité à capturer des dépendances complexes entre les données d'entrée.

Le mécanisme d'attention standard calcule une « attention » unique, où chaque requête **Q** est associée à un ensemble de clés **K** et de valeurs **V**, selon l'équation 1.14.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1.14}$$

Le produit QK^T permet d'obtenir un indice de similarité entre chaque requête et toutes les clés de la séquence. Ensuite, l'application de la fonction softmax permet de transformer ces indices en probabilités pour données une pondération attribuées aux différentes valeurs de V.

Dans le contexte de la prévision hydrologique, l'efficacité des modèles basés sur les transformeurs a été étudiée par Demiray & Demir (2024) pour anticiper l'évolution du débit des cours d'eau sur un horizon de 120 heures à partir des observations des 72 heures précédentes. Ces modèles ont été comparés à des approches plus conventionnelles, notamment les architectures Seq2Seq et LSTM. Les résultats mettent en évidence une performance supérieure des transformeurs, démontrant leur capacité à mieux modéliser les dynamiques hydrologiques.

Cette conclusion est également soutenue par l'étude menée par Amanambu, Mossa, & Chen, (2022) qui ont qui a évalué la capacité des modèles transformeur et LSTM à prédire les niveaux d'eau sur plusieurs échéances dans le cadre de la prévision de la sécheresse hydrologique sur la rivière Apalachicola en Floride. Les résultats de leur analyse montrent que le transformeur a systématiquement surpassé le LSTM, démontrant une meilleure précision sur toutes les périodes de prévision. Plus précisément, le transformeur a atteint un coefficient de détermination moyen R^2 supérieur à 0,92, tandis que le LSTM affichait une performance inférieure, avec un R^2 moyen inférieur à 0,88.

1.3.6 Transformeur de fusion temporelle (TFT)

Ensuite, Lim, Arik, Loeff, & Pfister (2020) proposent le modèle TFT, illustré à la Figure 1.5, conçu pour effectuer des prévisions multi-horizons en se basant sur l'architecture des transformeurs. Pour atteindre cet objectif, ils intègrent cinq stratégies : le mécanisme de porte, le réseau de sélection de variables, l'encodeur de covariables statiques, le traitement temporel, et les intervalles de prévision. Chacun de ces éléments est présenté ci-dessous.

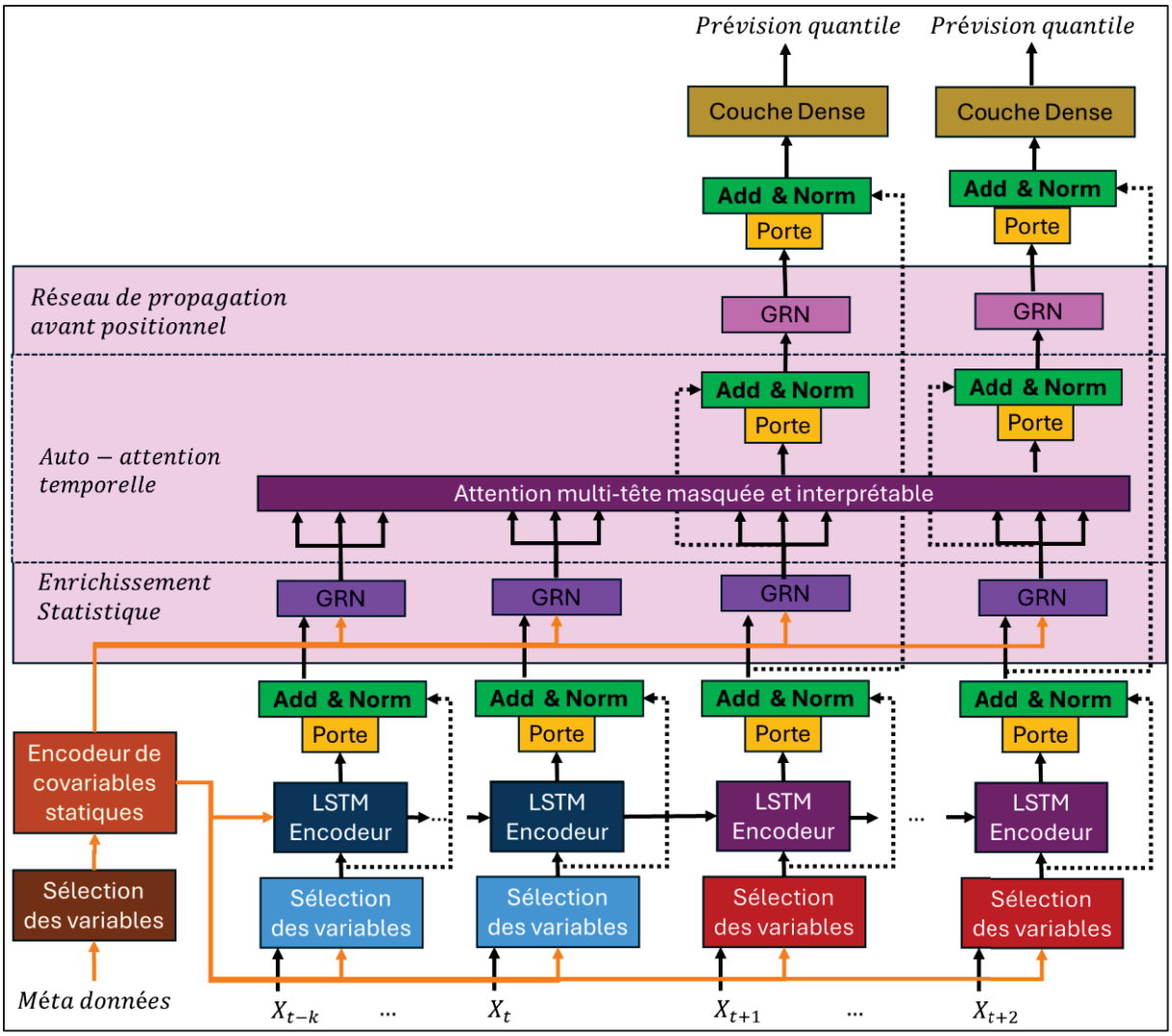

Figure 1.5 Architecture du Transformeur de Fusion Temporelle (TFT) intégrant des variables statiques et dynamiques, une sélection de variables adaptative et une attention multi-tête masquée et interprétable
(Adapté de Lim et al., 2020)

Mécanisme de porte (GRN) :

Les Mécanismes de porte (Gated Recurrent Network; GRN) dans le modèle TFT, illustrés à la Figure 1.6, permettent de contourner les composants inutilisés de l'architecture, adaptant ainsi la profondeur et la complexité du réseau aux besoins de chaque jeu de données.

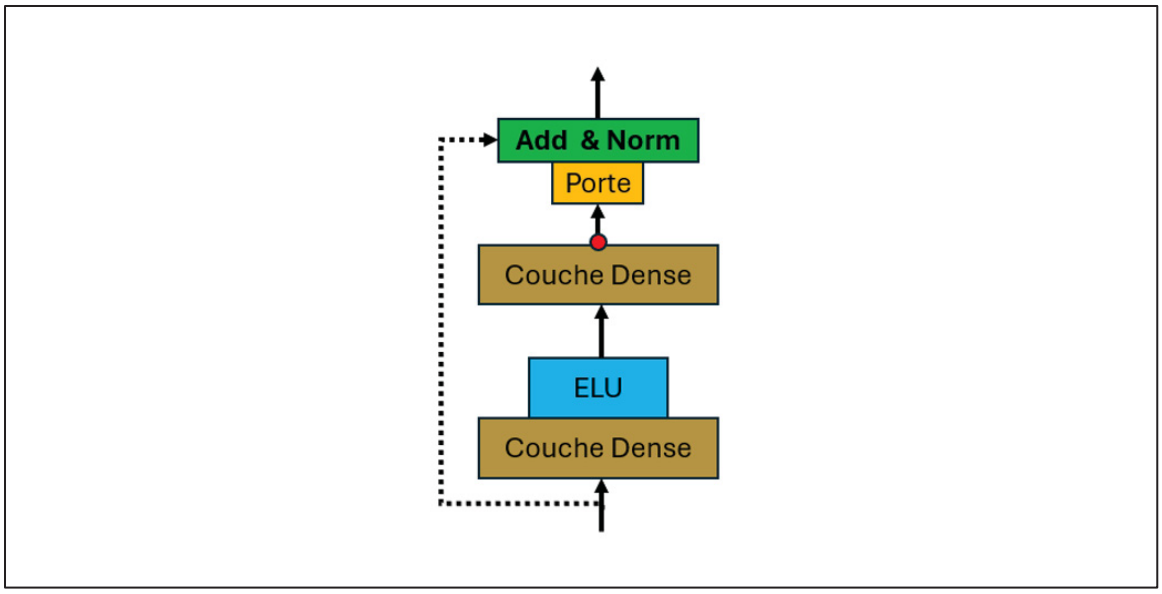

Figure 1.6 Mécanisme de porte normalisée (*Gated Residual Network*, GRN) utilisé dans le Transformeur de Fusion Temporelle (TFT), combinant des couches denses, une activation ELU et une connexion résiduelle avec normalisation
Adapté de Lim et al., (2020)

En utilisant le GRN, le modèle peut moduler le niveau de traitement non linéaire appliqué aux données en fonction de leur pertinence. D'ailleurs, la fonction qui introduit cette non-linéarité est la fonction ELU, qui retourne la valeur d'entrée si elle est positive et applique une transformation exponentielle pour les valeurs négatives. Si un traitement complexe n'est pas nécessaire, le mécanisme GRN peut complètement ignorer certaines couches, rendant le modèle plus simple et mieux adapté aux petits ensembles de données ou aux données bruitées (Lim et al., 2020).

Réseaux de sélection de variables:

Le modèle TFT intègre des réseaux de sélection de variables qui identifient dynamiquement les variables les plus pertinentes à chaque pas de temps. Cette sélection permet d'éliminer les entrées non pertinentes ou bruitées, améliorant ainsi la performance globale du modèle. La sélection des variables s'effectue à la fois pour les covariables statiques et temporelles, ce qui permet de concentrer la capacité d'apprentissage sur les caractéristiques les plus prédictives (Lim et al., 2020).

Encodeurs de covariables statistiques :

Le modèle TFT se distingue par sa capacité à intégrer des covariables statiques, qui peuvent avoir une influence sur la dynamique temporelle. Les encodeurs spécifiques aux covariables statiques génèrent des vecteurs de contexte qui sont utilisés à différents points de l'architecture, notamment pour enrichir le traitement des variables temporelles et pour sélectionner les variables pertinentes.

Traitement temporel :

Le modèle TFT utilise un MHA modifié pour apprendre les relations temporelles à court et à long terme. Ce mécanisme d'attention interprétable permet de capturer les dépendances à long terme entre les différentes étapes temporelles des données. En plus, le modèle emploie une couche de séquence à séquence pour traiter les données localement, ce qui permet de mieux capturer les patrons locaux comme les anomalies ou les points de changement tout en préservant les relations globales.

Intervalles de prévision via les prévisions quantiles :

Le modèle génère des intervalles de prévision en calculant des prévisions quantiles, ce qui permet d'estimer une plage de valeurs probables pour chaque horizon de prévision. Au lieu de fournir une seule valeur prédite, le modèle génère plusieurs percentiles, permettant ainsi d'estimer la distribution des valeurs possibles et d'améliorer la gestion de l'incertitude dans les prévisions probabilistes.

Quant à l'application dans le domaine de l'hydrologie, Rasiya Koya & Roy (2024) présentent une étude qui évalue ce modèles comparativement au transformeur et au LSTM. Dans cette étude, les modèles ont été évalués sur 2 610 bassins à travers le monde, et les résultats ont démontré que le TFT surpassait à la fois le LSTM et les transformeurs, tant en précision des prévisions qu'en interprétabilité. Les performances, mesurées par le KGE, étaient respectivement de 0.679 pour le LSTM, 0.661 pour le transformeur et 0.704 pour le TFT.

Selon Rasiya Koya & Roy (2024) il est possible de conclure trois éléments. Premièrement, le LSTM est performant pour capturer les relations à court terme grâce à sa structure récurrente. Deuxièmement, le modèle transformeur est meilleur pour capturer les dépendances à long terme grâce à ses mécanismes d'attention, qui lui permettent de traiter l'ensemble de la séquence simultanément. Troisièmement, le TFT se distingue par sa capacité à combiner les points forts du LSTM et du transformeur, en offrant des prévisions plus précises tout en capturant à la fois les patrons locaux et les tendances à long terme.

De récentes études confirment la supériorité du TFT en prévision hydrologique, grâce à sa capacité à intégrer divers types de données, capter les dépendances temporelles et fournir des prévisions précises et interprétables (Francisco & Matos, 2024 ; Mbuvha et al., 2023).

En effet, l'étude de Francisco & Matos, (2024) explore l'utilisation des TFT pour la prévision des débits au Portugal et le compare au modèle hydrologique conceptuel HBV (Ouachani, Bargaoui, & Ouarda, 2007). Les résultats montrent que le TFT offre des prévisions plus précises, surpassant le HBV dans plusieurs bassins, notamment en représentant mieux les pics de débit et en améliorant la généralisation.

D'ailleurs, Mbuvha et al., (2023) proposent un nouveau flux de travail pour la prévision des débits en présence de données manquantes, testé sur 10 stations hydrologiques au Bénin. Leur approche combine des techniques d'imputation des données avec l'utilisation de modèles TFT et LSTM pour la prévision des débits journaliers. Les résultats montrent que le TFT a surpassé le LSTM pour toutes les stations.

1.4 Objectifs de la présente étude

Cette étude vise à concevoir et développer un modèle de prévision des débits basé sur l'intelligence artificielle pour le bassin versant du Lac-Saint-Jean, en intégrant les dynamiques hydrologiques complexes telles que la fonte des neiges, les précipitations et les variations saisonnières. En s'appuyant sur l'étude de Sabzipour (2023), qui a établi une base de référence

pour la prévision hydrologique du bassin versant du Lac-Saint-Jean, cette étude vise à exploiter les mêmes ressources et données pour concevoir une architecture plus performante. L'objectif est d'améliorer la précision des modèles existants en développant et comparant deux approches. Soit un modèle encodeur-décodeur LSTM et un modèle encodeur-décodeur LSTM avec mécanisme d'attention.

Bien que les architectures transformeur et TFT semblent plus prometteuses, l'approche la plus simple est souvent privilégiée dans le développement des modèles d'IA. D'ailleurs, l'intégration des prévisions météorologiques ensemblistes (50 membres) combinée aux données historiques reste encore peu explorée dans la littérature. Les objectifs secondaires de cette étude incluent d'évaluer l'apport de ces informations et d'optimiser la précision des prévisions hydrologiques sur un horizon allant jusqu'à 15 jours.

CHAPITRE 2

ZONE D'ÉTUDE ET DONNÉES

Ce chapitre a pour objectif de présenter la zone d'étude en décrivant les caractéristiques du bassin versant du Lac Saint-Jean. Il est suivi d'une section dédiée aux données utilisées, comprenant les débits observés, les données météorologiques historiques et les prévisions météorologiques.

2.1 Zone d'étude

Cette étude est menée sur le bassin versant du Lac-Saint-Jean (LSJ), Figure 2.1, situé au Québec, Canada.

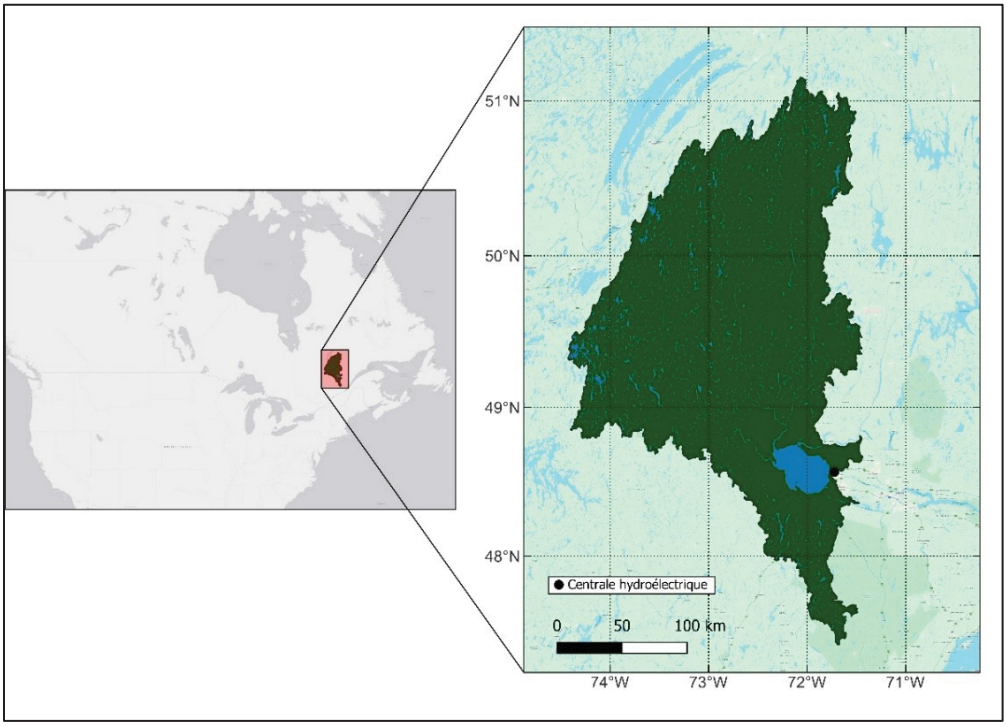

Figure 2.1 Bassin versant du Lac-Saint-Jean, QC, Canada, avec la centrale hydroélectrique (carré noir) de Rio Tinto

L'un des principaux défis est de capturer la dynamique de la fonte des neiges lors des crues printanières. Ce bassin, couvrant une superficie de plus de 45 000 km², alimente le réservoir du LSJ, d'une capacité de 4 550 hm³ (Arsenault & Côté, 2019). Ce réservoir approvisionne la centrale hydroélectrique d'Isle-Maligne, qui produit l'électricité destinée à l'aluminerie de Rio Tinto.

2.2 Base de données

Cette étude utilise trois types de données pour l'entraînement du modèle, soit les débits, les prévisions météorologiques et les données météorologiques historiques. Cette sous-section vise à fournir une description détaillée de chacune de ces catégories de données et des traitements appliqués. D'ailleurs, il est important de préciser que les paramètres statiques, tels que la topographie et la superficie du bassin, ne sont pas utilisés dans cette étude, car elle porte sur un seul bassin versant dont les caractéristiques physiques sont constantes.

Conformément aux objectifs définis, un modèle encodeur-décodeur sera mis en place. L'encodeur aura pour rôle de traiter les données météorologiques passées, tandis que le décodeur exploitera les prévisions météorologiques pour générer les sorties. Bien que ces données soient similaires à bien des égards, elles présentent des caractéristiques différentes qui seront présentées dans cette section. Étant donné que l'encodeur et le décodeur disposent chacun de leurs propres poids, ils peuvent établir indépendamment des relations entre les caractéristiques intégrées à leur structure respective. Par conséquent, la position et la quantité de caractéristiques dans le vecteur d'entrée n'ont, en théorie, aucun impact sur le modèle. Toutefois, pour des raisons qui deviendront plus évidentes dans le prochain chapitre quant au sujet de l'approche par phase, cette étude privilégie l'utilisation des variables communes entre les prévisions météorologiques et les observations.

2.2.1 Débit

Les données de débit sont fournies par l'entreprise Rio Tinto. Elles sont obtenues en calculant un bilan de masse au réservoir du Lac Saint-Jean, soustrayant les débits turbinés et déversés

des accumulations de volume d'emmagasinement du réservoir. Les relevés sont systématiquement enregistrés de façon journalière, représentant le débit moyen à 12h00. La série temporelle de ces mesures débutent le 1er janvier 1950, offrant ainsi un historique continu.

2.2.2 Données météorologiques observées

Le jeu de données utilisé pour les observations météorologiques est une réanalyse atmosphérique produite par le Centre Européen pour les Prévisions Météorologiques à Moyen Terme (*European Center for Medium-Range Weather Forecasts*; ECMWF). Ce jeu de données, ERA5, compile des données climatiques et météorologiques mondiales disponibles depuis 1940, en utilisant des techniques avancées d'assimilation qui combinent observations et modèles, offrant ainsi un historique détaillé de l'atmosphère. Avec sa résolution spatiale et temporelle élevée, ERA5 fournit des informations les plus précises possibles sur les conditions météorologiques historiques (Hersbach et al., 2020). Par ailleurs, l'étude de Tarek, Brissette, & Arsenault (2020) a confirmé l'efficacité d'ERA5 en tant que jeu de données de référence pour remplacer les observations météorologiques pour la modélisation hydrologique.

Dans le cadre de cette recherche, l'ensemble des données d'ERA5 s'étend du 1er janvier 1978 jusqu'au 31 décembre 2023. Ces données sont structurées en trois dimensions: le temps, la longitude et la latitude, qui sont appliquées à chacun des attributs météorologiques. D'ailleurs, afin de permettre au modèle de capturer les variations saisonnières, deux variables explicites, le mois et le jour, ont été ajoutées.

Tel que discuté plus tôt, les variables météorologiques sélectionnées incluent seulement celles communes aux prévisions. De plus, les données historiques de débit ont été intégrées directement dans le vecteur d'entrée, en excluant ceux des prévisions, afin de permettre au modèle de mieux relier les conditions météorologiques passées aux variations du débit. D'ailleurs, pour simplifier la tâche de l'encodeur, les données initialement réparties en fonction

des dimensions spatiales (latitude et longitude) ont été moyennées spatialement. Finalement, le vecteur d'entrée utilisé pour l'encodeur est présenté au Tableau 2.1.

Tableau 2.1 Variables météorologiques provenant d'ERA5 sélectionnées pour l'encodeur

Variable	Description	Unité
mois	Mois de l'année	-
Jour	Jour du mois	-
v10	Composante V du vent à 10 mètres	m/s
u10	Composante U du vent à 10 mètres	m/s
tp	Précipitations totales	m
tmax	Température maximale	K
ssr	Rayonnement solaire de surface net	J/m^2
sp	Pression de surface	Pa
sd	Équivalent en eau de la neige	m
d2m	Température du point de rosée à 2 mètres	K
Débit	Apports naturels d'eau calculés par bilan de masse	m^3/s

2.2.3 Prévision météorologique

Les données de prévision météorologique fournies par le ECMWF couvrent la période allant du 9 mars 2016 jusqu'au 31 mai 2023. Ces données sont structurées en quatre dimensions soient (1) le temps, (2) les 50 scénarios prévisionnels, (3) l'horizon de chaque prévision sur 15 jours (lead-time) au rythme de 4 prévisions par jour, ainsi que (4) les coordonnées géographiques (longitude et latitude sous forme de vecteur plutôt qu'en grille). Cette structure multidimensionnelle est appliquée à chaque variable des prévisions météorologiques. Les données météorologiques issues des prévisions ECMWF sont fournies à une fréquence de quatre fois par jour. Afin de simplifier leur traitement, une moyenne des valeurs a été calculée pour obtenir des agrégations quotidiennes sur une période de 15 jours. De plus, une moyenne a été appliquée sur les dimensions de longitude et de latitude afin de réduire la complexité spatiale.

Par ailleurs, dans le contexte des simulations, la valeur observée du débit de rivière qu'on retrouve au Tableau 2.1 est remplacée par la valeur du débit simulé dans la section dédiée au décodeur. Finalement, la variable de la couverture nuageuse totale a été retirée puisqu'elle n'était pas dans les données d'ERA5. Ainsi, le Tableau 2.2 présente une synthèse de variables sélectionnées.

Tableau 2.2 Variables météorologiques provenant d'ECMWF sélectionnées pour le décodeur

Variable	Description	Unité
Mois	Mois de l'année	-
Jour	Jour du mois	-
v10	Composante V du vent à 10 mètres	m/s
u10	Composante U du vent à 10 mètres	m/s
tp	Precipitations totales journalières	m
mx2t6	Température maximale sur 6 heures, converti en température maximale journalière	K
ssrd	Rayonnement solaire atteignant la surface décumulé	J/m^2
sp	Pression atmosphérique à la surface	Pa
sd	Équivalent en eau de la neige	m
d2m	Température du point de rosée à 2 mètres	K
Débit	Débit simulé du cours d'eau	m^3/s

CHAPITRE 3

MÉTHODOLOGIE

L'un des principaux avantages des modèles présentés dans la section suivante réside dans sa capacité à intégrer simultanément l'ensemble des valeurs disponibles. Toutefois, cette approche présente une contrainte majeure, elle limite la taille des données exploitables au plus petit échantillon commun. Dans le cadre de cette étude, les prévisions de l'ECMWF, couvrant une période de sept ans, restreignent l'utilisation des décennies de données disponibles d'ERA5 et des débits. Ce chapitre détaille les méthodes adoptées, en expliquant la division des données, les modèles utilisés, ainsi que l'approche par phases mise en place pour contourner cette limitation temporelle.

3.1 Division des données

Pour assurer une bonne généralisation du modèle et limiter le surapprentissage, les données sont réparties en trois ensembles : entraînement (50 %), validation (25 %) et test (25 %). L'ensemble de validation est utilisé pour ajuster l'entraînement sans influencer directement le modèle, tandis que l'ensemble de test permet d'évaluer sa performance finale sur des données totalement indépendantes. Cette répartition, bien que peu conventionnelle, permet d'avoir quatre crues printanières dans l'entraînement et deux périodes de crue dans la validation et les tests. Une fois les données collectées, il est nécessaire d'appliquer des méthodes de normalisation pour stabiliser l'entraînement. Dans cette étude, la normalisation basée sur l'écart-type a été privilégiée en utilisant la fonction *StandardScaler* fournie par la bibliothèque scikit-learn. De plus, afin d'éviter toute fuite de données, la moyenne et l'écart-type utilisés proviennent exclusivement des données d'entraînement avant d'être appliqués aux ensembles de validation et de test. Sans l'étape de normalisation, le modèle risque d'avoir des difficultés à attribuer une pondération équilibrée aux données d'entrée, en raison des disparités entre elles.

3.2 Modèles d'apprentissage profond

Cette section présente la structure du modèle utilisé dans cette étude. Elle commence par expliquer le modèle de base, qui repose sur une architecture encodeur-décodeur. Ensuite, elle aborde l'intégration du mécanisme d'attention, qui améliore la précision des prévisions en pondérant l'importance des étapes temporelles. Enfin, le choix des hyperparamètres ainsi que la fonction-objectif utilisée pour l'entrainement du modèle sont décrits.

3.2.1 Modèle encodeur-décodeur

Le modèle de base adopté pour cette étude repose sur l'architecture encodeur-décodeur (ED) telle qu'illustrée à la Figure 3.1. Dans la section de l'encodeur, les cellules bleues représentent des cellules LSTM qui reçoivent en entrée les données météorologiques observées (W_{obs}) et les débits enregistrés (Q_{obs}). Grâce à leur capacité à conserver des informations sur une séquence temporelle, les LSTM de l'encodeur peuvent extraire les caractéristiques pertinentes des données observées, offrant ainsi un contexte historique pour les prévisions.

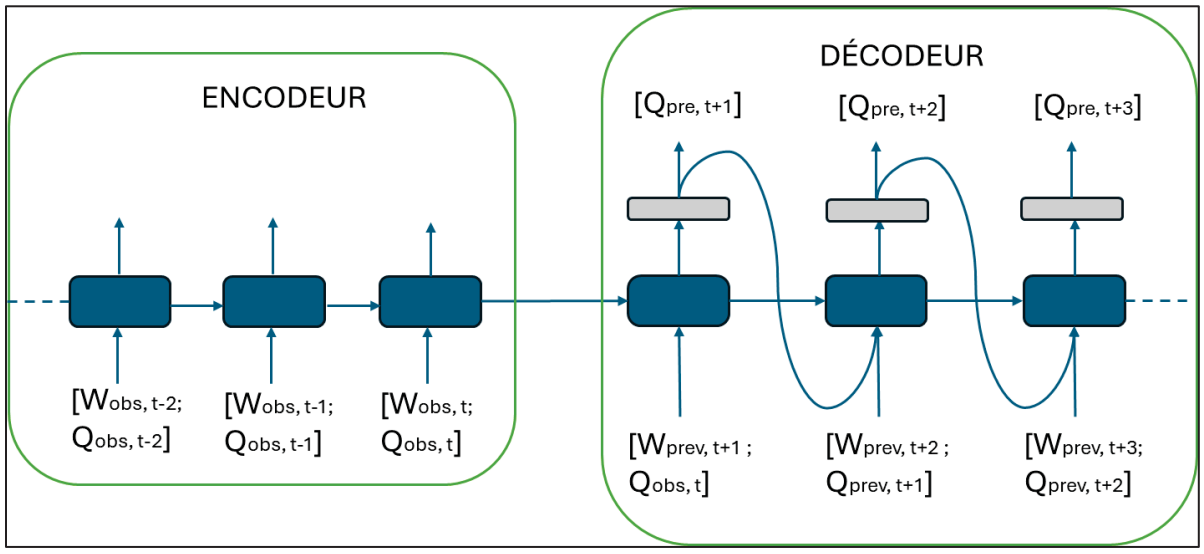

Figure 3.1 Architecture encodeur-décodeur basée sur des LSTM pour la prévision hydrologique, intégrant les données météorologiques observées et les prévisions météorologiques

Le décodeur, situé à droite, est également constitué de cellules LSTM qui utilisent les informations synthétisées par l'encodeur pour générer les prévisions. Les cellules grises dans le décodeur représentent des couches denses, qui ajustent la dimension des sorties des LSTM afin de produire les valeurs de prévision de débit (Q_{pre}). D'ailleurs, les poids attribués aux cellules LSTM du décodeur diffèrent de ceux de l'encodeur, ce qui rend possible l'intégration de vecteurs d'entrée de dimensions variables.

L'un des atouts majeurs de cette architecture réside dans la flexibilité du décodeur. Contrairement à une simple utilisation des informations de l'encodeur, le décodeur prend en entrée la dernière valeur prédite (Q_{pre}) et la concatène avec les prévisions météorologiques (W_{pre}). Cette approche permet d'incorporer les débits simulés dans le processus de prévisions des séquences suivantes, en tenant compte des conditions météorologiques futures. Par conséquent, cette méthode a le potentiel d'améliorer la précision globale du modèle, mais aussi de propager des erreurs.

Cette approche repose sur une technique inspirée du *teacher forcing* (TF), une méthode bien connue pour optimiser les paramètres des modèles séquentiels. Le TF consiste à remplacer de façon aléatoire la prévision du temps t-1 par la valeur réelle lors de l'entraînement (Brownlee, 2017). Dans le modèle proposé ici, au lieu d'intégrer systématiquement la valeur réelle, le modèle incorpore les prévisions météorologiques ainsi que le débit passé. L'hypothèse sous-jacente est qu'en intégrant ces deux types de données, le modèle sera encouragé à réaliser de meilleures prévisions.

3.2.2 Modèle avec mécanisme d'attention

Le modèle enrichi avec le mécanisme d'attention, illustré à la Figure 3.2, s'appuie sur l'architecture ED tout en ajoutant une complexité supplémentaire.

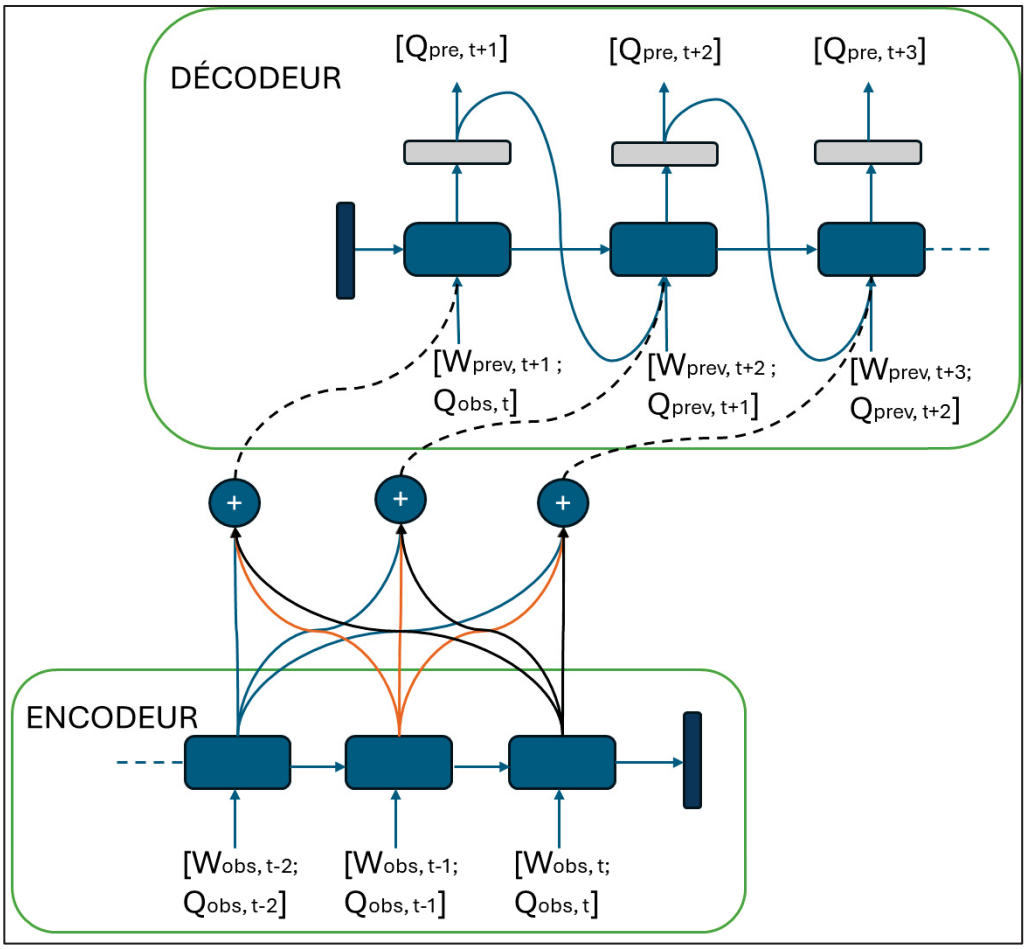

Figure 3.2 Architecture encodeur-décodeur enrichie par un mécanisme d'attention permettant une pondération adaptative des étapes temporelles pour améliorer la précision des prévisions hydrologiques

Dans ce modèle, l'attention est appliquée pour pondérer l'importance des différentes étapes temporelles des données extraites par l'encodeur. Cette approche permet de résoudre le problème du goulot d'étranglement de l'espace latent, où toute l'information doit habituellement passer à travers un seul vecteur de contexte fixe. En concaténant les états cachés de l'encodeur avec l'état actuel du décodeur, le mécanisme d'attention calcule une « énergie » (symbolisé par le cercle avec un « + ») qui détermine l'importance de chaque étape temporelle. Grâce à cette énergie, le décodeur n'est plus limité à un contexte global, il peut accéder directement à l'ensemble des informations de l'encodeur et ajuster son focus en temps réel pour chaque prévision. L'intégration de l'attention permettrait donc d'améliorer la précision des prévisions de débit. Cependant, elle introduit également une certaine complexité.

3.2.3 Modèle encodeur-décodeur multi-membres

Le modèle encodeur-décodeur mutli-membres (ED-MM) est basé sur le modèle encodeur-décodeur de base. Il fonctionne en plaçant en parallèle autant de modules LSTM indépendants que de membres de l'ensemble de prévisions, tous étant alimentés par le même encodeur. Chaque ensemble, composé d'un LSTM et de ses couches denses associées, prédit les débits sur une période de longueur désirée.

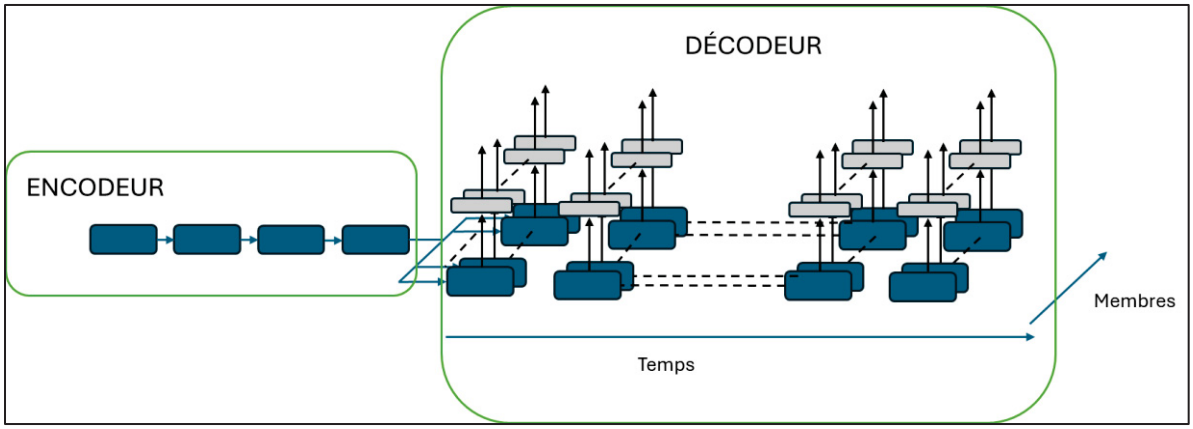

Figure 3.3 Architecture du modèle encodeur-décodeur multi-membres (ED-MM) intégrant plusieurs décodeurs indépendants pour capturer l'incertitude des prévisions hydrologiques

Cette étude vise non seulement à évaluer l'approche multi-membres, mais aussi à la comparer avec l'ajout d'un mécanisme d'attention. L'objectif est de déterminer si l'augmentation de la complexité du modèle est nécessaire pour mieux capturer les relations complexes présentes dans les données.

3.2.4 Choix des hyperparamètres

L'optimisation des hyperparamètres a été réalisée afin d'identifier la configuration offrant les meilleures performances. Lors de l'entraînement, une taille de lot de 128 a été utilisée, signifiant que 128 échantillons de données étaient traités simultanément avant chaque mise à jour des poids du modèle. Cela permet de maintenir un bon compromis entre l'efficacité computationnelle et la précision.

Le taux d'apprentissage est une valeur scalaire qui sert de facteur de mise à l'échelle pour le gradient lors de la mise à jour des poids du modèle. Plus précisément, il est multiplié au gradient calculé à chaque itération de la descente de gradient afin de contrôler l'amplitude des ajustements appliqués aux paramètres du réseau. Un taux d'apprentissage trop élevé peut provoquer une divergence du modèle, tandis qu'un taux trop faible ralentit la convergence et coincé le modèle dans un minimum local. Pour cette étude, un taux d'apprentissage initial de 0,001 est utilisé, avec une réduction progressive de 25% toutes les 5 itérations, permettant d'affiner les mises à jour des poids.

La longueur de la séquence de l'encodeur, fixée à 90 jours, a été choisie pour capturer un historique contextuel suffisant tout en minimisant les pertes de données au début des ensembles d'entraînement, de validation et de test. Bien qu'une séquence plus longue, de 180 ou 365 jours, aurait été préférable pour capturer l'entièreté du cycle saisonnier, la contrainte stricte de division des données a limité cette possibilité. En effet, il est impossible d'utiliser les données d'entraînement pour fournir le contexte historique de la validation et ainsi de suite pour la validation et les tests. Cela entraîne donc une perte inévitable de données équivalente à la longueur de la séquence de l'encodeur, réduisant ainsi le nombre de jours disponibles de prévision pour toutes les catégories de données. Afin de préserver les crues printanières, une séquence plus courte et une division contrôlée des données se révèlent plus adaptées.

Par la suite, l'architecture repose sur trois couches LSTM, introduisant une complexité supplémentaire pour modéliser des dynamiques temporelles plus fines. Le mécanisme d'attention additive a été privilégiée et la taille des couches cachées était fixée à 256. Enfin, une désactivation aléatoire de 0,4 a été appliqué sur l'encodeur et le décodeur pour limiter le surapprentissage. Cette technique de régularisation désactive aléatoirement 40 % des neurones à chaque itération, obligeant ainsi le modèle à entraîner l'ensemble des neurones plutôt que de se focaliser sur quelques-uns.

3.2.5 Fonction-objectif

La fonction-objectif, utilisée comme critère de performance lors de l'entrainement du modèle, repose sur le MSE, qui mesure l'écart quadratique moyen entre les prédictions et les valeurs cibles. Dans le domaine de l'hydrologie, le MSE est particulièrement adapté, car il incite le modèle à apprendre à partir des valeurs aberrantes, grâce à l'effet du gradient. Cela peut potentiellement aider le modèle à mieux prédire les débits extrêmes, ce qui est essentiel pour les prévisions de débit en période de crue. Cependant, cela peut aussi avoir l'effet de négliger les périodes de plus faible hydraulicité, où les erreurs absolues sont plus faibles qu'en période de crue.

Pour mieux prendre en compte d'autres aspects tels que les valeurs extrêmes et l'incertitude des prévisions, des pondérations spécifiques ont été appliquées.

Tout d'abord, une pondération w_{ij} basée sur la valeur cible y_{ij} est introduite où i représente la journée de prévision et j l'ensemble. Certaines valeurs de débit sont considérées comme moins représentés que d'autres dans l'apprentissage. Ainsi, les valeurs inférieures à 400 m^3/s et 300 m^3/s se voient attribuer des pondérations de 3 et 5 respectivement, tandis que toutes les valeurs supérieures à 1000 m^3/s sont multipliées par un facteur de 3. Cette stratégie permet de renforcer l'importance de certaines plages de valeurs des prévisions.

$$w_{ij} = \begin{cases} 5, & si\ y_{ij} < 300 \\ 3, & si\ y_{ij} < 400 \\ 3, & si\ y_{ij} > 1000 \\ 1, & sinon \end{cases} \tag{3.1}$$

Une autre correction importante concerne les valeurs cibles y_{ij} qui se situent en dehors de la plage de prédictions de l'ensemble. Si la valeur cible est inférieure au minimum ou supérieure au maximum des prédictions, une pénalité de 5.0 est appliquée. En complément, une seconde pénalisation d'une valeur de 3.0 est introduite pour les valeurs cibles qui sortent de l'intervalle

quartiles Q1 et Q3 des prédictions de l'ensemble. Cette approche permet de favoriser une meilleure prise en compte des valeurs centrales.

$$v_{ij} = \begin{cases} 5, & si\ y_j \notin \left[\min\left(\widehat{y_{ij}}\right), \max\left(\widehat{y_{ij}}\right)\right] \\ 3, & si\ y_j \notin \left[\mathrm{Q1}\left(\widehat{y_{ij}}\right), \mathrm{Q3}\left(\widehat{y_{ij}}\right)\right] \\ 1, & sinon \end{cases} \tag{3.2}$$

L'ensemble de ces pénalités *L* est combiné sous la forme d'une pondération globale qui ajuste la perte MSE de base où *N* représente le nombre de jour de prévision, soit de 15 et *M* le nombre d'ensembles soit 50.

$$L = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} 0.1 * w_{ij} + 0.5 * v_{ij} + 0.4 * \left(y_{ij} - \widehat{y_{ij}}\right)^2 \tag{3.3}$$

La pondération finale résulte d'un équilibre entre l'importance des valeurs cibles, l'incertitude des prévisions et l'écart entre les valeurs cibles et les plages statistiques des prévisions.

3.3 Stratégie d'entraînement par phases pour l'intégration des données météorologiques

Un minimum de 15 années d'historique est nécessaire pour établir des prévisions hydrologiques fiables (Kratzert, Klotz, Brenner, Schulz, & Herrnegger, 2018). Cependant, la période couverte par les prévisions météorologiques fournies par l'ECMWF limite la quantité de données disponibles pour l'entraînement à sept années. Afin de pallier cette contrainte, cette étude propose une méthode alternative permettant de générer des pseudo-données de prévisions météorologiques à partir des données historiques d'ERA5. Cette transformation vise à structurer les données ERA5 de manière cohérente avec celles des prévisions ECMWF, en alignant leurs formats et leurs caractéristiques. Cette approche permet ainsi d'élargir l'ensemble d'entraînement.

Le Tableau 3.1 présente un résumé des stratégies d'entraînement réalisées. Les dates en bleu correspondent aux données issues d'ECMWF, tandis que celles en vert représentent les données provenant d'ERA5.

Tableau 3.1 Stratégies et phases d'entraînement, de validation et de test des modèles avec données ERA5 et ECMWF

	Entrainement		Validation		Test	
	Encodeur	Décodeur	Encodeur	Décodeur	Encodeur	Décodeur
ED-MM	2016-2021	2016-2021	2021-2022	2021-2022	2022-2023	2022-2023
ED-MM-TA Phase 1	1978-2016	1978-2016	2016-2021	2016-2021	2021-2023	2021-2023
ED-MM-TA Phase 2	2016-2021	2016-2021	2021-2022	2021-2022	2022-2023	2022-2023

3.3.1 Stratégie sans transfert d'apprentissage:

Lors de la stratégie sans transfert d'apprentissage, le modèle (ED-MM) est entraîné avec les données ECMWF et ERA5. Bien que ERA5 couvre la période 1978 à 2023, la fenêtre utilisable est limitée à 2016-2023 en raison de la disponibilité des données d'ECMWF. Ainsi, l'objectif est d'évaluer la performance du modèle avec les données disponibles sur cette période commune.

3.3.2 Stratégie avec transfert d'apprentissage :

Phase 1:

La phase 1 repose sur l'utilisation exclusive des données historiques ERA5 lors de l'entrainement couvrant la période 1978 à 2016. Pour ce faire, des pseudo-prévisions ont été créées à partir des données d'ERA5. Cette approche vise à tirer parti d'un jeu de données plus étendu. Les ensembles de validation et de test sont ensuite sélectionnés sur les périodes 2016-2021 et 2022-2023, respectivement, en utilisant les valeurs d'ECMWF pour réaliser des prévisions. L'intention de cette phase est d'avoir un modèle avec des poids pré-entrainé. Par conséquent, il ne sera pas évalué comme un modèle indépendant.

D’ailleurs, pour simuler le ssrd d’ECMWF, la variable ssr d’ERA5 est employée. De plus, afin de mieux imiter la structure et les caractéristiques des données de prévision météorologique, une transformation des données de pseudo-prévision a été effectuée. Une fonction d'ajout de bruit suivant une distribution normale a été appliquée, où le niveau de bruit est proportionnel à la distance temporelle par rapport à la journée considérée. Par exemple, un bruit de 1% est ajouté pour la première journée, 2% pour la deuxième, et ainsi de suite. Cette approche permet de simuler des erreurs de mesure ou des variations aléatoires dans les données, rendant le modèle plus résilient face aux incertitudes. Cette méthode de transformation permet que les pseudo-prévisions reproduisent non seulement les caractéristiques structurelles des données ECMWF, mais intègrent également une dimension d’incertitude.

Phase 2:

Enfin, la phase 2 adopte une approche par transfert d’apprentissage (TA), où, plutôt que d'initialiser les poids de manière aléatoire, le modèle pré-entraîné en phase 1 est réutilisé. Cette stratégie permet de tirer parti des connaissances acquises lors de la première phase, optimisant ainsi l’apprentissage et la performance du modèle. Donc, dans un premier temps, le modèle est pré-entraîné sur les données historiques ERA5 (1978-2016) afin de bénéficier de l’ensemble des tendances météorologiques passées. Ensuite, le modèle avec ses poids pré-entrainés est réutilisé et est affiné sur la période 2016-2021 en intégrant les prévisions du ECMWF, ce qui permet d'améliorer sa précision sur des données récentes tout en apprenant à gérer l'incertitude des prévisions météorologiques.

Ainsi, l’objectif est de comparer quatre modèles pour mesurer l'impact du mécanisme d’attention et du TA sur la précision des prévisions hydrologiques. Deux modèles utilisent l’attention (avec ou sans TA), tandis que les deux autres fonctionnent sans attention. D’ailleurs, tous ces modèles reposent sur l’architecture encodeur-décodeur multi-membres. Pour des raisons de simplification, cette architecture sera désignée sous l’appellation ED dans la suite du document.

Tableau 3.2 Synthèse du nom des modèles selon l'architecture et la phase

	Sans attention	Avec Attention
Sans TA	ED	ED-ATT
Avec TA	ED-TA	ED-ATT-TA

Cette approche permet d'évaluer séparément l'influence de l'attention et du TA sur la performance des prévisions d'ensemble et la gestion de l'incertitude. Les quatre configurations des modèles entraînés sont ainsi définies par le Tableau 3.2.

CHAPITRE 4

RÉSULTATS

Le chapitre suivant présente les résultats obtenus après l'entraînement des quatre modèles présentés au Tableau 3.2. Chacun de ces modèles a été entraîné avec les mêmes hyperparamètres et sur la même durée. Ainsi, pour l'ensemble du chapitre, la période de test s'étend de 2022 à 2023 sur une durée de 555 jours. D'ailleurs, lorsqu'une analyse est détaillée par saison, les données sont divisées selon la classification astronomique des saisons, c'est-à-dire en fonction des équinoxes et des solstices.

4.1 Analyse des prévisions et dispersion des ensembles

Cette section présente une analyse des prévisions hydrologiques obtenues lors des tests. La Figure 4.1 illustre des échantillons de prévisions sélectionnés pour représenter chaque saison. Cette comparaison a pour objectif de mettre en évidence les performances des quatre modèles selon le CRPS moyenné sur les 15 jours de prévisions pour une date commune, permettant ainsi de quantifier la précision des prévisions probabilistes.

L'analyse est réalisée pour cinq périodes distinctes, soit des dates spécifiques pour l'hiver, le printemps 1 et le printemps 2, qui représentent respectivement le début de la crue printanière et son maximum, ainsi que l'été et l'automne. Chaque colonne de la Figure 4.1 correspond à un modèle spécifique, tandis que chaque ligne représente une saison. Les courbes noires indiquent la moyenne des prévisions d'ensemble, les courbes vertes pointillées représentent les observations, et les lignes grises illustrent les membres de l'ensemble simulé. Cette visualisation permet d'évaluer la capacité de chaque modèle à capter les variations saisonnières et à quantifier l'incertitude associée aux prévisions.

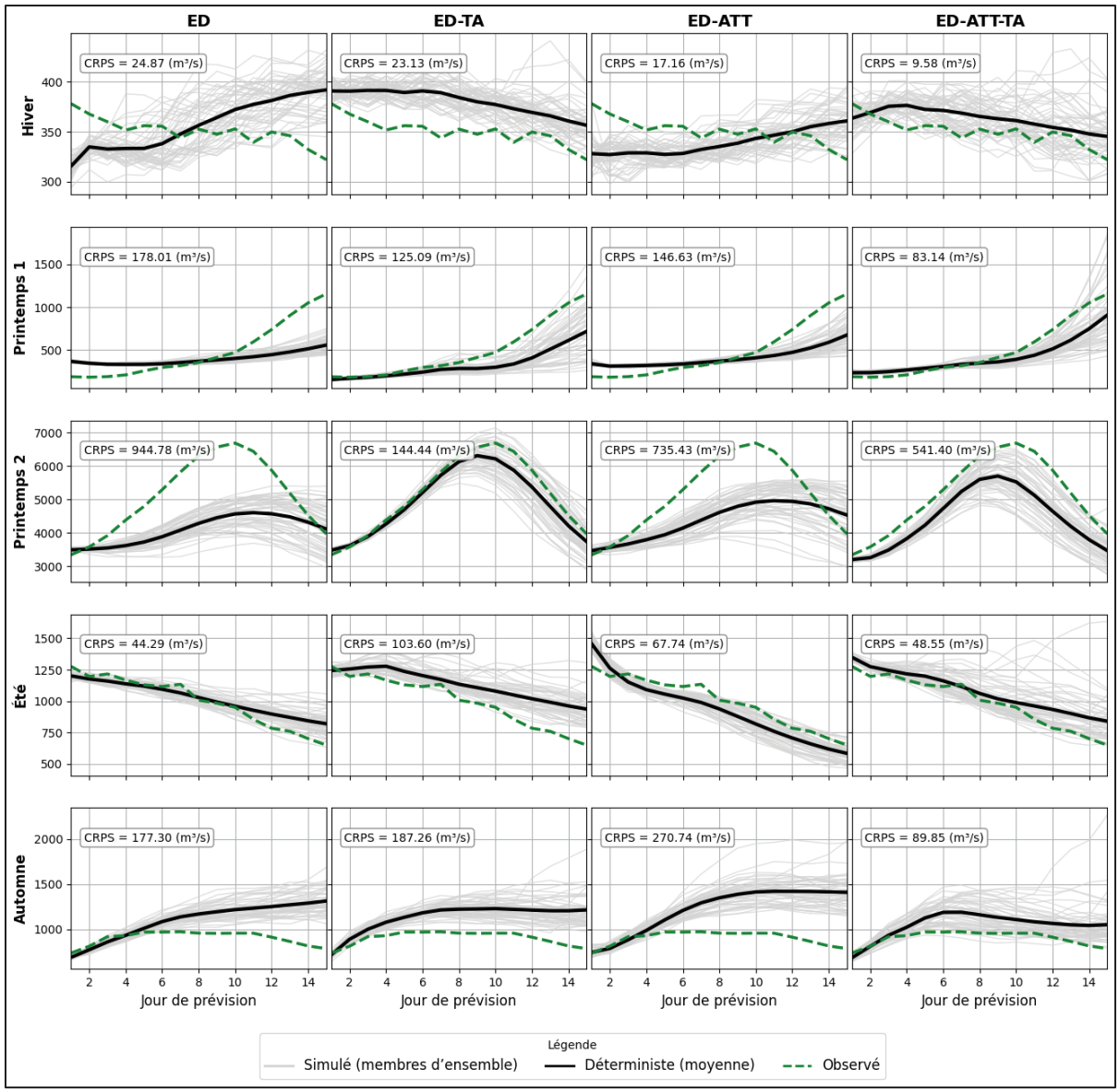

Figure 4.1 Prévisions sur un horizon de 15 jours selon les saisons et les modèles. Les prévisions d'ensemble des quatre modèles sont représentées par des courbes grises, la moyenne déterministe par une courbe noire, et les observations réelles par une courbe verte pointillée. Chaque ligne correspond à une saison, tandis que chaque colonne représente un modèle spécifique. Le CRPS, affiché pour chaque scénario, sert de métrique pour évaluer la précision des prévisions d'ensemble

En hiver, soit le 29 janvier 2023, les valeurs de CRPS indiquent que le modèle ED-ATT-TA, avec un score de 9.58 m³/s, offre les prévisions les plus précises, suivi du modèle intégrant uniquement l'attention, qui atteint 17.16 m³/s. À l'inverse, le modèle ED, affichant un CRPS

de 24.87 m³/s, présente les performances les plus faibles, suggérant que l'absence d'attention et de transfert d'apprentissage nuit à la qualité des prévisions. De plus, il est possible de constater que les modèles utilisant la stratégie de TA captent mieux la tendance décroissante du débit. Une fois le mécanisme d'attention ajouté, les débits observés sont beaucoup mieux représentés dans la distribution.

Quant au printemps 1 (prévision du 16 avril 2022), qui correspond au début de la crue printanière, une période particulièrement critique pour les gestionnaires hydriques, le modèle ED-ATT-TA se distingue par la meilleure performance, réduisant le CRPS à 83.14 m³/s, comparé à 178.01 m³/s pour le modèle ED. L'ajout du TA seul pour le modèle ED-TA améliore également la précision avec un CRPS de 125.09 m³/s. Sans le TA, les prévisions apparaissent beaucoup plus grossières, tandis que son intégration affine les courbes et permet de mieux suivre les subtilités des débits observés. De plus, l'ajout du mécanisme d'attention contribue à maintenir la qualité des prévisions sur un horizon plus long.

Lors du pic de la crue printanière, représenté par le printemps 2 (prévision du 24 mai 2022), les performances des modèles présentent des différences considérables. Contrairement aux tendances observées lors du début de la crue, le modèle ED-TA, bien que dépourvu de mécanisme d'attention, affiche un CRPS plus faible (144.44 m³/s) que le modèle ED-ATT (735.43 m³/s), suggérant que l'ajout de l'attention dans ce cas particulier ne contribue pas nécessairement à améliorer la prévision. De plus, bien que le modèle ED-ATT-TA sous-estime les débits observés et présente un CRPS plus élevé (541.40 m³/s), il est important de noter que cette période correspond à une année record, marquée par l'une des plus grandes crues printanières enregistrées. Cela met en évidence un défi pour les modèles intégrant l'attention, qui peuvent être plus sensibles aux événements extrêmes et nécessiter des ajustements supplémentaires pour capter adéquatement ces dynamiques hydrologiques exceptionnelles. Toutefois, il paraît évident que l'ajout du TA est nécessaire pour mieux capturer la dynamique des crues printanières.

Quant à l'été (prévision du 14 aout 2022), les modèles présentent des performances plus homogènes par rapport aux autres saisons, avec des valeurs de CRPS relativement faibles. Contrairement aux autres périodes, le modèle sans attention ni TA offre ici la meilleure précision avec un CRPS de 44.29 m^3/s, tandis que le modèle ED-ATT-TA affiche un CRPS légèrement plus élevé de 48.55 m^3/s. Ceci souligne que, dans ce contexte particulier, l'apport du mécanisme d'attention et du TA n'apporte pas d'amélioration et introduit une complexité supplémentaire sans bénéfice. De plus, l'ajout de la stratégie TA tend à augmenter la complexité de la distribution des prévisions, rendant les sorties plus dispersées. En revanche, les modèles sans transfert d'apprentissage produisent des prévisions plus homogènes, ce qui peut constituer un avantage pour cette période. Cette stabilité dans les prévisions simplifiées peut être préférable dans un contexte où les variations des débits sont plus stables.

Selon l'échantillon de l'automne (prévision du 28 octobre 2022), le modèle ED-ATT-TA se distingue avec un CRPS de 89.85 m^3/s, confirmant l'intérêt de la combinaison du TA et de l'attention pour améliorer la précision des prévisions. Il capture mieux la dynamique des débits par rapport aux autres configurations. Le modèle ED, avec un CRPS de 177.30 m^3/s, bien qu'inférieur en performance, surpasse néanmoins certaines variantes intégrant partiellement ces mécanismes, illustrant ainsi l'impact limité de l'ajout isolé du TA ou de l'attention.

4.2 Analyse quantitative du CRPS

Cette section analyse l'évolution du CRPS, d'abord sur l'ensemble de la période de test, puis par saison. La Figure 4.2 illustre son évolution sur 15 jours de prévision, tandis que la Figure 4.3 détaille les performances saisonnières pour les différents modèles.

Selon la Figure 4.2, lorsqu'aucun mécanisme d'attention n'est intégré, un écart est observé entre les modèles avec et sans TA, en faveur du TA, qui réduit l'erreur initiale. Cependant, cet écart diminue progressivement au fil des jours pour devenir presque négligeable à l'horizon de 15 jours. Cette tendance confirme les observations précédentes. En effet, bien que le TA améliore la dispersion des prévisions, il entraîne une légère perte de précision à long terme.

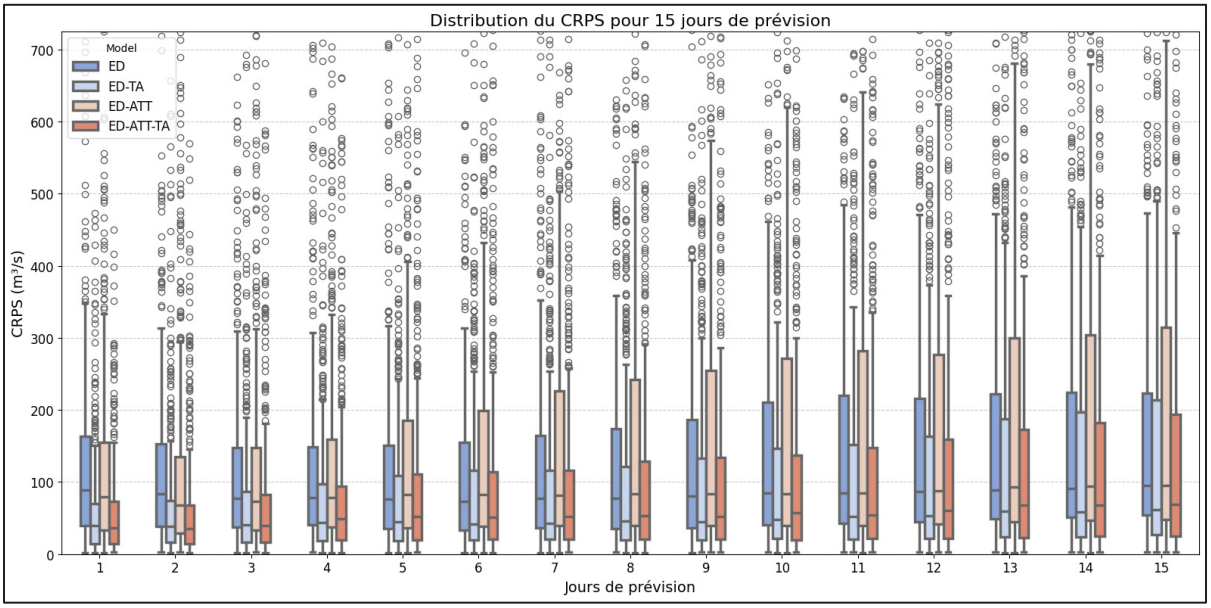

Figure 4.2 Distribution du CRPS en fonction des 15 jours de prévision pour les quatre modèles à base de LSTM

L'ajout de l'attention modifie cette dynamique. Si, au début de la prévision, un écart similaire est observé entre les modèles avec et sans TA, l'erreur du modèle avec l'attention augmente progressivement, maintenant un écart relativement stable avec le modèle ED-ATT-TA. De plus, il est intéressant de noter que le modèle sans TA et sans attention surpasse à long terme celui intégrant l'attention mais sans TA. En somme, l'ajout du TA est particulièrement bénéfique à court terme, améliorant la précision des premières journées de prévision. Cette tendance continue surtout avec la combinaison du mécanisme d'attention.

Par la suite, la Figure 4.3 illustre la distribution du CRPS en fonction de l'horizon de prévision, cette fois segmenté par saison. Cette approche permet d'analyser plus en détail l'impact du TA et du mécanisme d'attention.

En hiver, les valeurs de CRPS sont relativement faibles et stables sur l'ensemble des jours de prévision. L'ajout du TA et de l'attention permet une amélioration considérable, avec une réduction systématique du CRPS. Les modèles intégrant les deux stratégies affichent une meilleure performance par rapport aux autres qui présentent une dispersion plus importante

des erreurs. L'impact du TA est plus marqué que celui de l'attention, d'ailleurs, l'attention seule n'apporte aucune amélioration sans l'apport du TA.

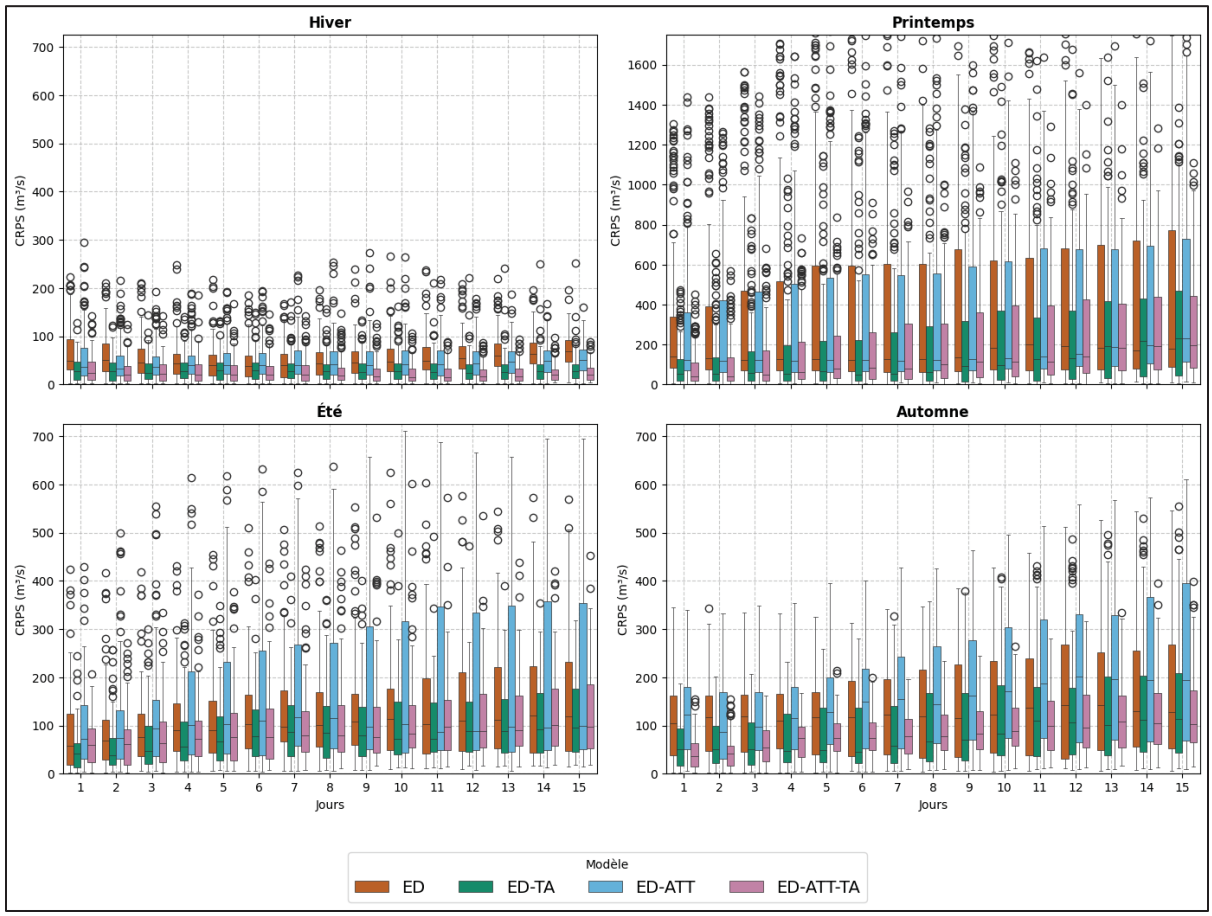

Figure 4.3 Distribution du CRPS en fonction des 15 jours de prévision selon les saisons pour les quatre modèles. Il est à noter que l'échelle verticale est commune à toutes les saisons sauf le printemps en raison de la grande amplitude des CRPS pour cette saison

Le printemps se distingue par une forte augmentation du CRPS au fil des jours de prévision, particulièrement en raison de l'augmentation des débits. L'écart entre les modèles est plus marqué, mettant en évidence la complexité de cette période pour la prévision hydrologique. Le modèle ED sans aucune stratégie affiche les erreurs les plus élevées, tandis que ED-ATT-TA réduit le CRPS. L'attention seule semble moins efficace sans l'apport du TA, suggérant que la captation des tendances historiques est essentielle pour prévoir la montée des eaux avec précision.

L’été, les erreurs de prévision sont globalement plus faibles qu’au printemps, tout en présentant une plus grande dispersion que les deux autres saisons. Contrairement aux autres saisons, lorsque le TA est absent, le modèle sans attention performe relativement bien à long terme. Le TA introduit une plus grande variabilité dans les prévisions, ce qui peut être un avantage pour mieux représenter l’incertitude, mais un inconvénient en termes de précision. L’attention seule semble ajouter de la complexité sans améliorer les performances, alors que le modèle ED-TA maintient une bien meilleure précision.

L’automne présente une tendance intermédiaire entre l’hiver et l’été, avec des erreurs modérées et une augmentation progressive du CRPS. La combinaison des deux stratégies offre la meilleure performance, réduisant les erreurs à long terme. L’attention seule ne semble pas améliorer les prévisions, confirmant que son utilité dépend d’un apprentissage avec deux phases.

4.3 Analyse selon la métrique KGE

L'analyse du KGE permet d'évaluer la performance des modèles en tenant compte simultanément de la corrélation, du biais et de la variabilité. La Figure 4.4 ci-dessus illustre l'évolution du KGE sur 15 jours de prévision pour les différentes configurations de modèles. Pour plus de détails, l’Annexe I présente le Tableau-A I-1 et le Tableau-A I-2 d’évaluation du KGE pour les phases d'entraînement, de validation et de test des quatre modèles.

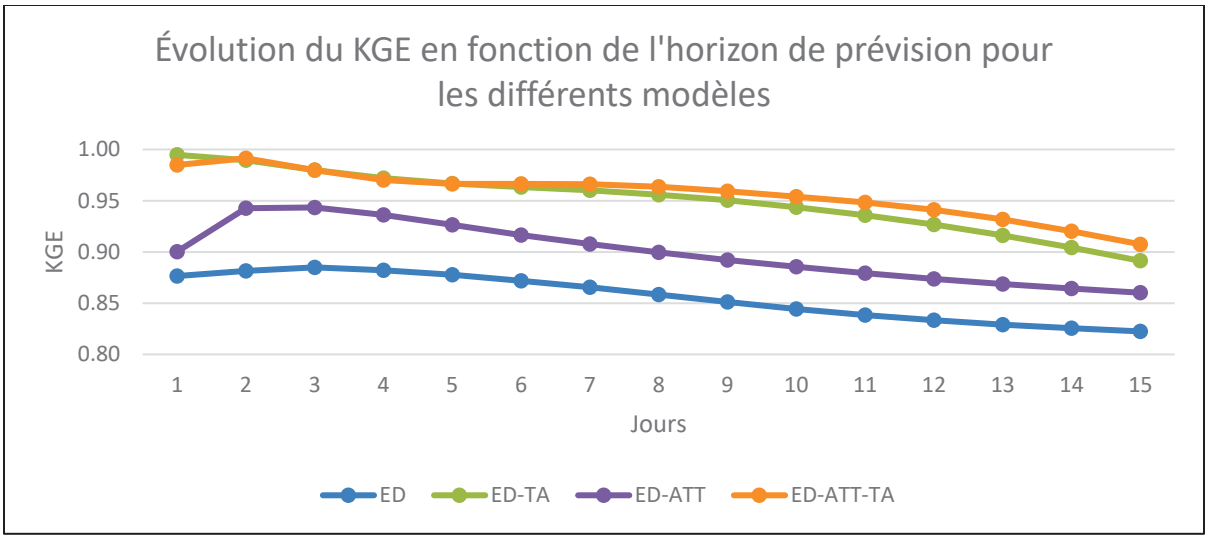

Figure 4.4 Évolution du KGE sur 15 jours de prévision pour les quatre modèles

Les résultats montrent une diminution progressive du KGE avec l'horizon de prévision, ce qui est attendu étant donné l'accumulation des erreurs au fil du temps. Toutefois, les modèles intégrant le TA et l'attention affichent clairement de meilleures performances. Le modèle ED-ATT-TA maintient les valeurs de KGE les plus élevées sur l'ensemble de la période de prévision.

En revanche, le modèle ED présente les performances les plus faibles, avec une dégradation plus marquée du KGE au fil du temps. L'ajout du TA améliore les résultats, réduisant la décroissance du KGE sur les 15 jours de prévision. Cependant, lorsqu'il est combiné avec le mécanisme d'attention, l'effet est encore plus marqué, surtout sur les premiers jours de prévision.

Il est également intéressant de noter que le modèle ED-ATT affiche un KGE inférieur à celui du ED-TA, ce qui suggère que l'attention seule ne suffit pas à améliorer la qualité des prévisions sans le bénéfice du transfert d'apprentissage.

4.4 Analyse selon les diagrammes Talagrand

L'analyse des diagrammes de Talagrand permet d'évaluer la fiabilité des prévisions probabilistes en examinant la répartition des observations.

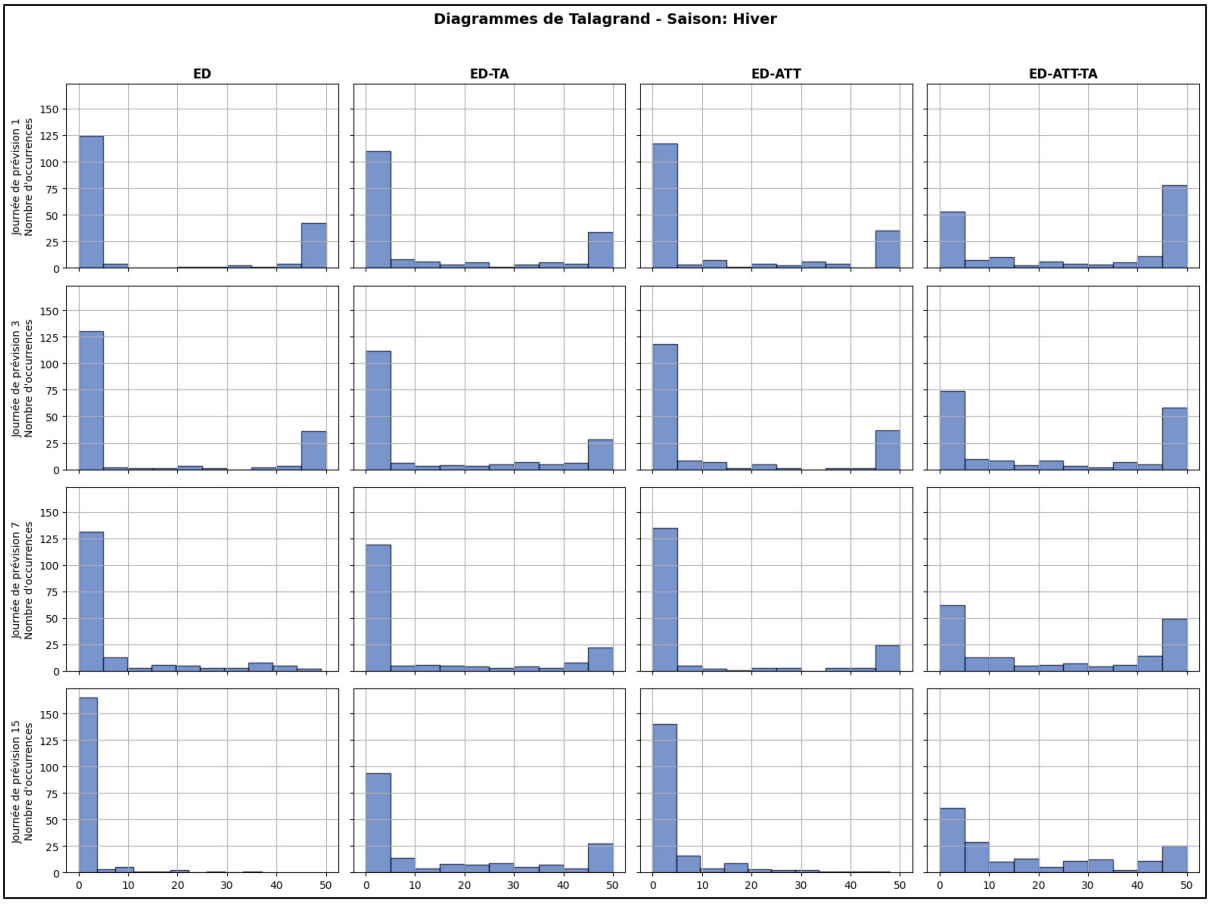

Figure 4.5 Diagrammes de Talagrand pour l'hiver. Les colonnes représentent les quatre modèles évalués (ED, ED-TA, ED-ATT, ED-ATT-TA), tandis que les lignes correspondent aux différentes échéances de prévision (jours 1, 3, 7 et 15)

Pour l'hiver, les diagrammes présentés à la Figure 4.5 montrent une forte concentration des observations aux bornes inférieures et quelquefois à la borne supérieure pour l'ensemble des modèles, indiquant un biais systématique qui tend à surestimer les débits. L'ajout du TA améliore légèrement la distribution, mais une asymétrie persiste. Toutefois, on observe qu'avec le TA, plus l'horizon de prévision avance, plus la répartition des observations tend vers une distribution plus uniforme. L'introduction du mécanisme d'attention semble avoir peu d'effet

en hiver, mais combiné avec le TA, l'attention permet de réduire les biais et améliorer la dispersion.

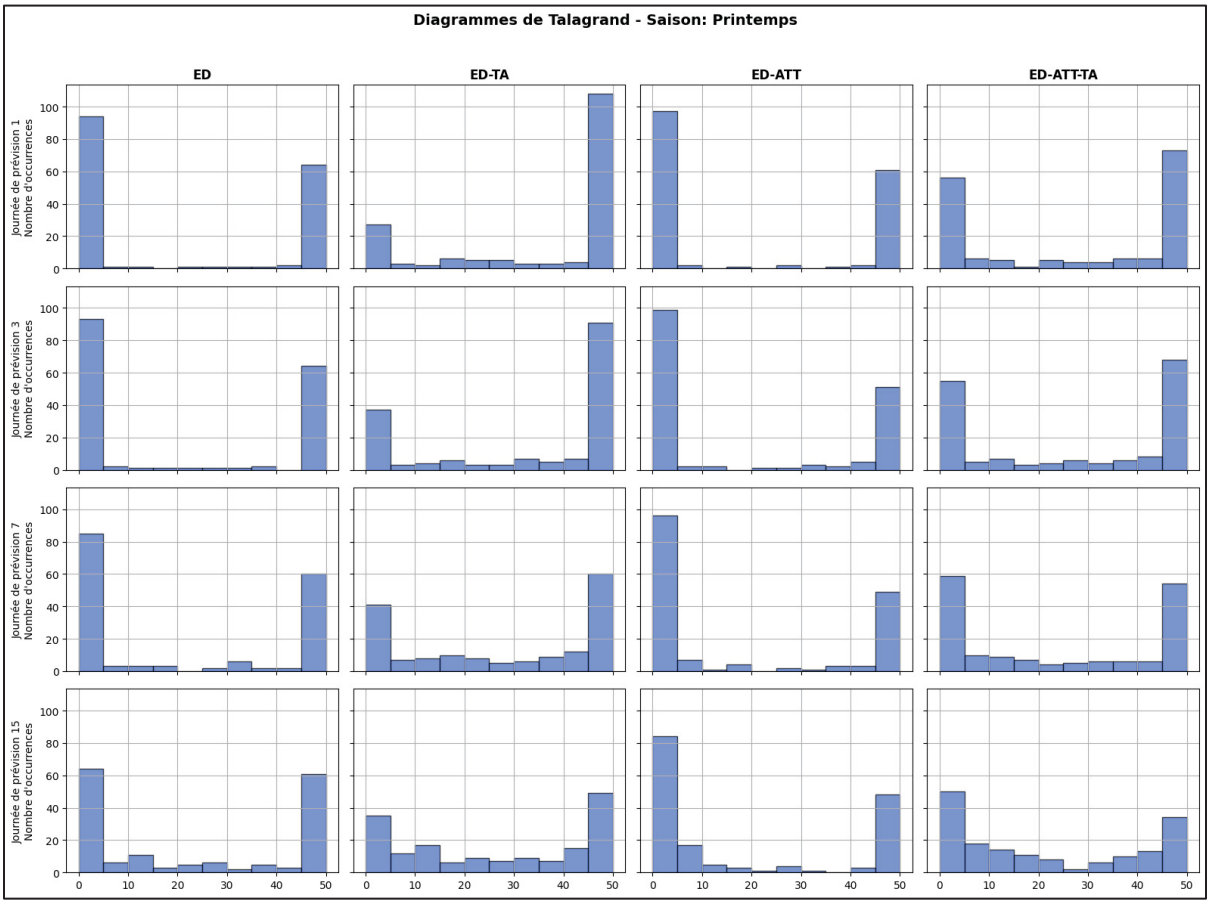

Figure 4.6 Diagrammes de Talagrand pour le printemps. Les colonnes représentent les quatre modèles évalués (ED, ED-TA, ED-ATT, ED-ATT-TA), tandis que les lignes correspondent aux différentes échéances de prévision (jours 1, 3, 7 et 15)

Par la suite, les diagrammes de Talagrand présentés à la Figure 4.6 pour le printemps montrent une forte concentration aux bornes, indiquant une sous-dispersion des prévisions. Le TA améliore progressivement la répartition des prévisions, réduisant l'asymétrie au fil des jours. L'attention seule n'apporte qu'un gain limité, mais combinée au TA, elle permet une meilleure dispersion des prévisions.

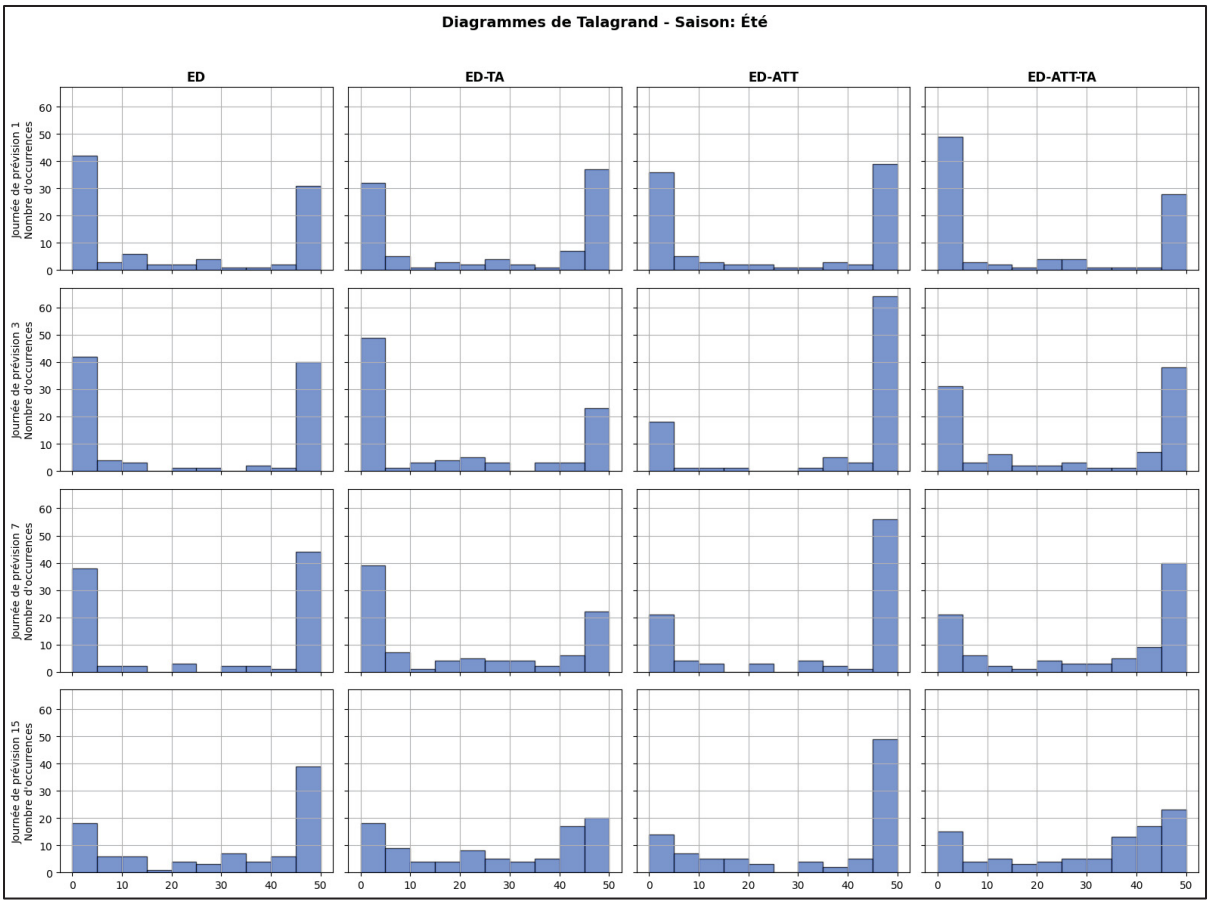

Figure 4.7 Diagrammes de Talagrand pour l'été. Les colonnes représentent les quatre modèles évalués (ED, ED-TA, ED-ATT, ED-ATT-TA), tandis que les lignes correspondent aux différentes échéances de prévision (jours 1, 3, 7 et 15)

Les résultats pour l'été, présentés à la Figure 4.7, suivent les mêmes tendances que celles observées en hiver et au printemps. Les modèles sans TA présentent une forte concentration aux bornes, indiquant de la sous-dispersion systématique. L'ajout du TA améliore progressivement la répartition des prévisions, rendant la distribution plus uniforme avec l'augmentation du nombre de jours. Le modèle ED-ATT semble légèrement sous-estimer les valeurs, mais son comportement reste similaire aux autres saisons.

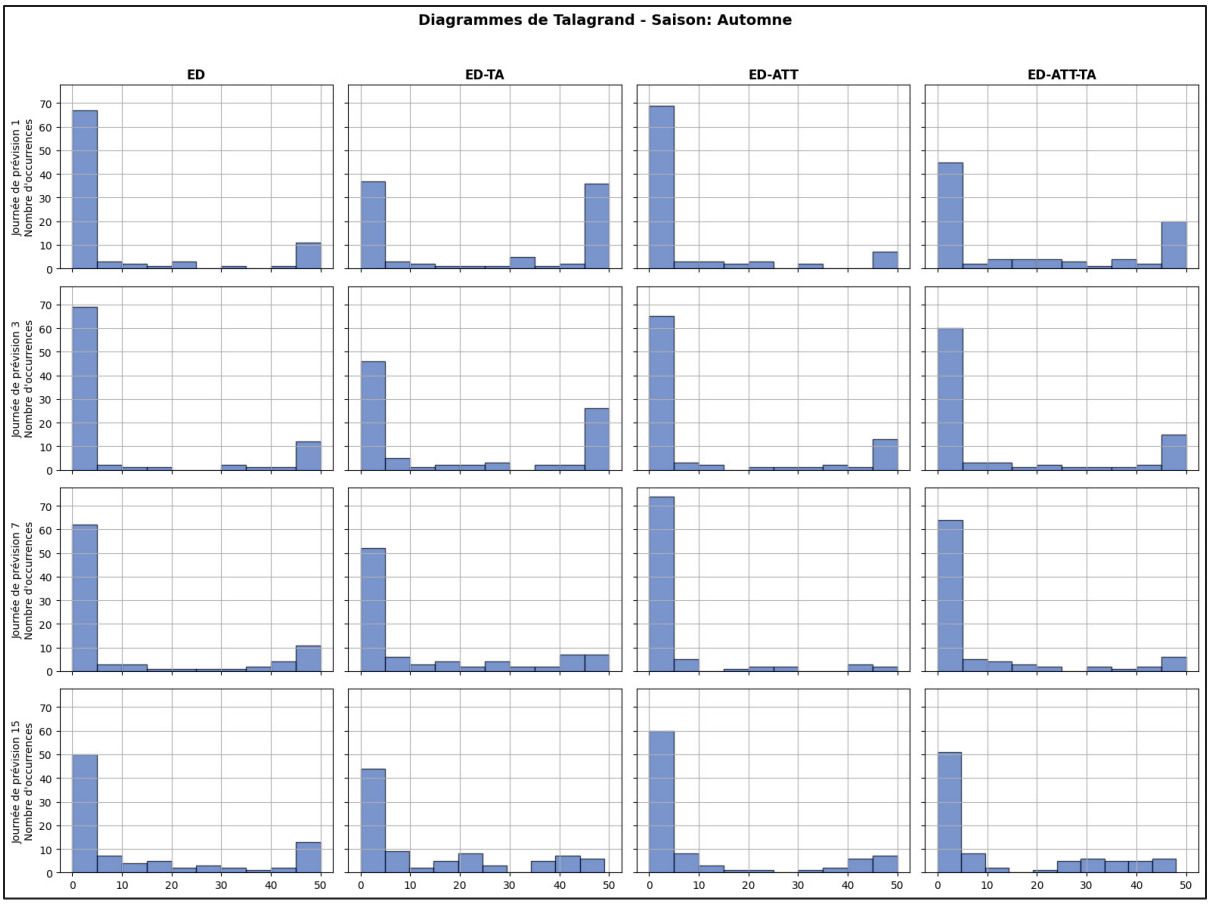

Figure 4.8 Diagrammes de Talagrand pour l'automne. Les colonnes représentent les quatre modèles évalués (ED, ED-TA, ED-ATT, ED-ATT-TA), tandis que les lignes correspondent aux différentes échéances de prévision (jours 1, 3, 7 et 15)

En automne, on observe à la Figure 4.8 une accumulation disproportionnée des prévisions dans la première classe de distribution, ce qui indique une tendance des modèles à surestimer les débits. Cette caractéristique est particulièrement marquée pour le modèle ED-ATT, suggérant que l'ajout du mécanisme d'attention seul ne corrige pas ce biais. En revanche, l'ajout du TA permet d'atténuer cette tendance, bien que la répartition des prévisions reste asymétrique.

CHAPITRE 5

DISCUSSION

Lorsqu'on évoque la notion de meilleur modèle, cela ne signifie pas nécessairement qu'un modèle plus complexe, comme le TFT, surpasse systématiquement un LSTM plus simple. En effet, bien que le TFT soit une architecture plus avancée, cela ne garantit pas automatiquement de meilleures performances. Ce qui détermine réellement la qualité d'un modèle est sa capacité à capturer la complexité inhérente aux données et à généraliser efficacement. Autrement dit, un modèle plus complexe ne sera bénéfique que si la nature des données justifie cette complexité et que le volume de données disponible est suffisant pour un apprentissage. Ainsi, à la lumière de ce chapitre, il sera possible de déterminer si l'intégration de l'approche multi-membres, du transfert d'apprentissage et du mécanisme d'attention apporte une amélioration des performances et si cela justifie une exploration plus approfondie de modèles encore plus complexes dans le cadre de l'approche ED.

Pour ce faire, cette discussion est structurée en cinq sections. La première section porte sur l'analyse des modèles sans TA, et sera suivie d'une analyse des modèles avec TA. Ensuite, une analyse inter-métrique est réalisée afin de comparer les performances des modèles selon les différentes métriques d'évaluation. La section suivante est consacrée à la comparaison des résultats obtenus avec ceux de l'état de l'art, permettant de situer les performances des modèles développés dans le contexte des travaux existants. Enfin, la dernière section discute des limitations de l'approche proposée.

5.1 Analyse des modèles sans le transfert d'apprentissage

Les résultats ont démontré que les modèles sans transfert d'apprentissage parviennent à capturer les tendances générales des séries hydrologiques. Cependant, ils rencontrent des difficultés importantes face aux fluctuations complexes des débits. Cela ne signifie pas nécessairement qu'ils sont inefficaces, mais plutôt qu'ils pourraient bénéficier d'une réduction de leur complexité, par exemple en diminuant la taille du vecteur d'état caché.

D’ailleurs, cette hypothèse est appuyée par l’analyse du CRPS, où l’impact du mécanisme d’attention reste incertain lorsque le modèle ne bénéficie pas du TA. Si l’on compare les performances entre le modèle de base et celui avec l’attention sans le TA pour toutes les saisons, l’ajout de l’attention n’apporte pas systématiquement d’amélioration. En hiver, par exemple, les résultats montrent des écarts instables. En effet, certains jours, l’attention semble aider à mieux suivre la dynamique des débits, tandis que d’autres jours, elle détériore la performance de manière aléatoire. Ce manque de cohérence suggère que, sans TA, le mécanisme d’attention est moins efficace et parfois même nuisible.

Des tendances similaires sont observées au printemps, où l’impact de l’attention varie d’un jour à l’autre sans effet net en moyenne. En revanche, en été et en automne, les conclusions sont claires. L’ajout de l’attention augmente systématiquement les erreurs, ce qui pourrait indiquer une mauvaise exploitation des dépendances temporelles dans ces contextes de plus faible variabilité hydrologique.

Toutefois, l'analyse basée sur la métrique KGE offre un angle d'interprétation différent. Comme illustré à la Figure 4.4, le modèle intégrant l’attention affiche systématiquement des valeurs de KGE plus élevées, avec un écart moyen de plus de 0.05 point par rapport au modèle sans attention. À première vue, cela peut sembler contradictoire avec les observations issues du CRPS, mais cela peut s’expliquer par la nature des indicateurs utilisés.

En effet, le KGE repose sur une évaluation déterministe qui mesure la performance globale du modèle en termes de corrélation, de biais et de variabilité, sans tenir compte de la dispersion des prévisions d’ensemble. Ainsi, même si le mécanisme d’attention n’améliore pas systématiquement les prévisions quotidiennes, il peut favoriser une représentation plus fidèle des tendances générales du débit sur l’ensemble de la période analysée. À l’inverse, le CRPS, les échantillons de prévision et le diagramme de Talagrand sont des métriques probabilistes, plus sensibles à la calibration ponctuelle, c’est-à-dire, la dispersion des ensembles jour par jour. Ils révèlent ainsi des déséquilibres dans la distribution des prévisions.

En somme, l'architecture proposée semble donner de bons résultats, malgré le fait qu'elle n'ait bénéficié que de moins de quatre années de données pour l'entraînement. Au vu des performances obtenues, il serait pertinent de poursuivre l'analyse en explorant deux pistes d'amélioration. Tout d'abord, il faudrait réduire la complexité du modèle afin d'éviter un sur-apprentissage. Ensuite, il faudrait intégrer des termes de régularisation plus adaptés, optimisant ainsi la généralisation des prévisions.

5.2 Analyse des modèles avec le transfert d'apprentissage

Lorsqu'on observe les modèles intégrant la stratégie de TA, il apparaît immédiatement que les résultats sont meilleurs. Cette sous-section vise à mettre en évidence la valeur ajoutée de l'initialisation des poids avant l'entraînement.

D'un point de vue déterministe, l'analyse de la métrique KGE (Figure 4.4) confirme la supériorité des modèles utilisant le TA. En particulier, la comparaison entre le modèle TA-ATT et TA seul révèle que leurs performances sont quasiment identiques entre les journées 1 et 6. Cependant, au-delà de cette période, la version intégrant le mécanisme d'attention maintient une courbe plus stable.

Ainsi, ces résultats suggèrent que le modèle proposé est suffisamment complexe pour capturer les relations sous-jacentes entre les données, et que l'ajout du mécanisme d'attention affine encore davantage les prévisions, bien que l'amélioration demeure modérée. D'ailleurs, cette conclusion se reflète clairement dans la Figure 4.1, où l'on constate que l'ajout du TA permet non seulement de mieux représenter les tendances générales, mais aussi de mieux capter des subtilités hydrologiques. Cette amélioration est particulièrement notable lors de la période du printemps 2, qui correspond au pic de la crue printanière. Contrairement aux modèles sans TA, ceux qui en bénéficient parviennent à mieux suivre l'évolution des débits extrêmes, réduisant ainsi les écarts entre les prévisions et les observations.

Quant au CRPS, il devient de plus en plus évident que l'ajout du TA améliore la qualité des prévisions, même en hiver, une période généralement stable. En effet, comme l'illustre la Figure 4.1, les deux modèles affichant les CRPS les plus faibles sont ceux intégrant le transfert d'apprentissage, confirmant son impact positif.

On pourrait penser qu'en hiver, une période où les variations de débit sont relativement limitées, un plus petit ensemble de données suffirait à bien modéliser les tendances. Toutefois, l'augmentation du volume de données grâce au TA permet au modèle de capter des subtilités plus complexes et d'optimiser la calibration des prévisions. Ce résultat s'inscrit dans la continuité des conclusions précédentes, à savoir que l'intégration du TA améliore la capacité du modèle à généraliser les dynamiques hydrologiques, en particulier dans des contextes où des tendances plus fines doivent être représentées.

L'analyse des résultats révèle une tendance cohérente entre les saisons, confirmant l'apport du TA. Comme observé en hiver, les prévisions du printemps et de l'été suivent une dynamique similaire, où le CRPS des modèles avec TA est systématiquement plus faible que celui des modèles sans TA. Cela suggère que l'intégration d'un pré-entraînement améliore considérablement la représentation des tendances hydrologiques saisonnières. En revanche, l'ajout du mécanisme d'attention ne semble pas offrir une plus-value, la différence entre les modèles avec et sans attention étant pratiquement négligeable pour ces saisons.

En automne, toutefois, une amélioration notable du CRPS avec ATT commence à émerger, indiquant que l'attention pourrait jouer un rôle plus pertinent dans des contextes hydrologiques plus complexes. Contrairement aux autres saisons, où le TA domine largement en termes de gains de précision, cette période met en évidence un impact positif du mécanisme d'attention, notamment lorsque combiné au TA.

Cependant, l'évaluation du bénéfice de l'attention uniquement à travers le CRPS peut être plus délicate. En effet, l'attention tend à augmenter la variance des prévisions d'ensemble, ce qui

peut dégrader les scores de CRPS si cette dispersion n'est pas accompagnée d'une amélioration de la capacité du modèle à capter la valeur observée dans l'ensemble prévisionnel.

Ce phénomène d'augmentation de variabilité s'explique par la manière dont les entrées sont intégrées par le décodeur multi-membres dans le cadre du modèle avec attention. Initialement, tous les membres reçoivent la même représentation encodée. Toutefois, dès l'étape de décodage, la structure des données intrants est modifiée. En effet, au lieu d'avoir une entrée basée sur 11 valeurs, chacune étant une série temporelle, le modèle avec attention ajoute 256 unités au vecteur d'entrée, correspondant à une représentation condensée de l'état caché moyen de l'encodeur. De plus, chaque membre de l'ensemble choisit indépendamment une fenêtre de focalisation sur les 90 jours analysés par l'encodeur, introduisant une plus grande diversité dans les prédictions. Ce mécanisme génère une pluralité de débits simulés, augmentant ainsi la variance globale des prévisions.

Cette diversification des scénarios simulés peut être observée visuellement dans la Figure 4.1, qui illustre une plus grande dispersion des ensembles en présence du mécanisme d'attention. Toutefois, les diagrammes de Talagrand indiquent qu'il n'y a pas de gain clair en termes de capacité à inclure systématiquement la valeur cible dans l'enveloppe des prévisions. Autrement dit, l'attention élargit l'éventail des possibles sans recentrer les prévisions autour des observations. Ainsi, bien que l'attention soit potentiellement bénéfique pour enrichir la quantification de l'incertitude, cette diversité n'est pas toujours alignée sur la valeur observée. En conséquence, elle entraîne une hausse du CRPS.

5.3 Analyse inter métrique

D'un point de vue déterministe, l'analyse du KGE, qui représente la performance moyenne des prévisions, montre des résultats très satisfaisants. En particulier, le modèle ED-ATT-TA atteint un KGE d'environ 0.91 après 15 jours, indiquant que la moyenne des prévisions est bien alignée avec la réalité. Cela suggère que, globalement, les modèles capturent correctement l'évolution des débits.

Cependant, une analyse ensembliste plus approfondie, en utilisant le CRPS et les diagrammes de Talagrand, met en évidence une mauvaise calibration de la dispersion des ensembles. Bien que les résultats du CRPS soient très bons, en particulier le modèle ED-ATT-TA, qui affiche les meilleures performances, certaines limites apparaissent lorsque l'on examine la répartition des prévisions.

En effet, si le CRPS est relativement faible, indiquant que les prévisions sont précises, il ne pénalise pas suffisamment les erreurs situées aux extrêmes de la distribution (min, max, Q1 et Q3). Cette observation est confirmée par les diagrammes de Talagrand, qui révèlent une forte concentration des prévisions aux bornes inférieures et supérieures, ce qui suggère que les modèles ont tendance à sous-estimer l'incertitude réelle. Autrement dit, les valeurs cibles sont trop fréquemment situées aux extrêmes de la distribution, plutôt qu'au centre.

Cette observation suggère que le modèle présente un excès de confiance dans ses prévisions d'ensemble, en particulier aux horizons courts, où la variance des prédictions est plus faible. Cela s'explique d'une part par le fait que l'information transmise au décodeur est identique pour tous les membres de l'ensemble, et d'autre part par la faible incertitude des prévisions météorologiques aux premiers jours, ce qui réduit la dispersion des simulations. Ainsi, plus l'horizon de prévision est court, plus le modèle tend à générer des prévisions homogènes, limitant la diversité des scénarios explorés.

Pour atténuer ce biais, étant donné que l'architecture ED et les prévisions météorologiques restent constantes, une solution envisageable consisterait à modifier la fonction de coût en intégrant une pénalisation basée sur la variance des prévisions des membres. Cette approche permettrait de contraindre le modèle à maintenir une dispersion suffisante dans les prévisions d'ensemble, en particulier lors des premiers jours, afin de mieux refléter l'incertitude réelle.

Cela suggère que, même si le CRPS et le KGE valident la précision du modèle, une optimisation de la répartition des ensembles est nécessaire pour éviter une concentration

disproportionnée des prévisions dans les bornes extrêmes et mieux représenter l'incertitude hydrologique.

5.4 Comparaison à l'état de l'art

L'objectif de cette étude était d'optimiser l'utilisation des données ECMWF et ERA5 et de comparer les performances du modèle avec l'état de l'art. Toutefois, la comparaison entre différentes études reste complexe, car elle nécessiterait des tests sur les mêmes bassins et sur une période identique. Ainsi, pour évaluer les bénéfices du modèle proposé, il a été confronté aux résultats de Sabzipour et al. (2023), qui ont obtenu d'excellentes performances sur le même bassin versant avec les mêmes sources de données, bien que la période de test soit différente.

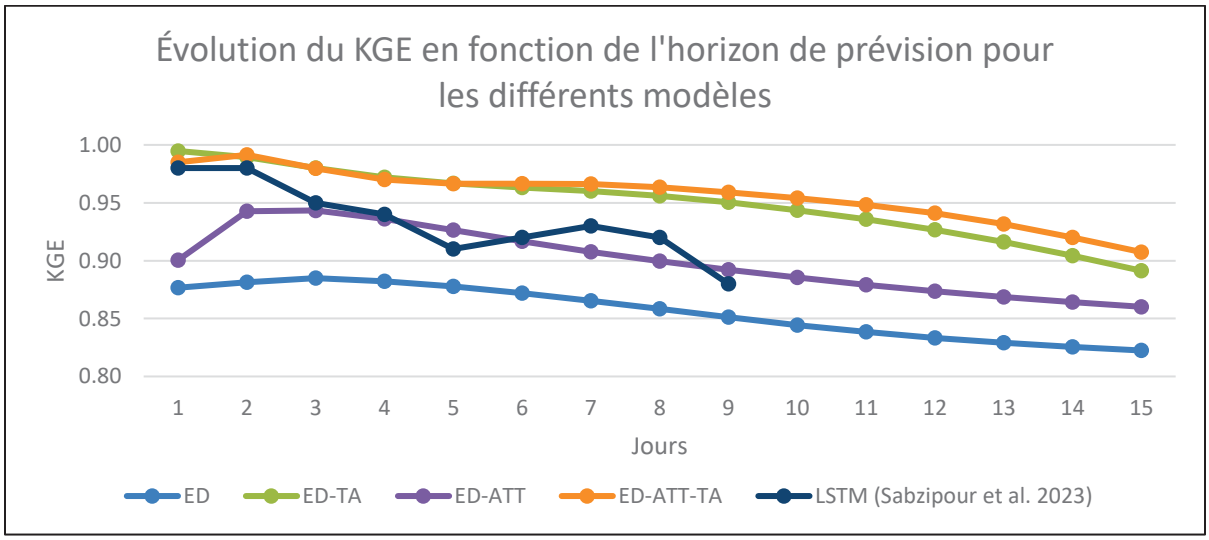

Figure 5.1 Évolution du KGE des cinq modèles, incluant la comparaison avec Sabzipour (2023), en fonction de l'horizon de prévision sur les données de test

L'analyse de la Figure 5.1 met en évidence la supériorité des modèles intégrant le TA par rapport au modèle LSTM de Sabzipour et al. (2023). Les modèles avec TA maintiennent des valeurs de KGE plus élevées sur l'ensemble des 9 premiers jours de prévision, traduisant une meilleure stabilité et une moindre dégradation des performances dans le temps. En particulier, le modèle intégrant à la fois le mécanisme d'attention et le TA affiche un KGE de 0,96 au jour

9, suivi du modèle utilisant uniquement le TA avec 0,95, alors que le modèle de Sabzipour atteint seulement 0,88, indiquant une performance nettement inférieure.

D'ailleurs, il est possible d'établir un parallèle avec l'étude de Kratzert, Gauch, Klotz, & Nearing, (2024) qui suggère d'entraîner un modèle sur plusieurs bassins versants afin d'améliorer ses performances. Ce principe repose sur l'idée qu'en intégrant un plus grand volume de données issues de divers bassins, le modèle peut apprendre des relations hydrométéorologiques plus généralisables, malgré les différences spécifiques à chaque bassin. Ainsi, cette approche permet d'accéder à une quantité substantielle de données supplémentaires, offrant une meilleure représentation des relations entre les variables météorologiques et les débits, sans toutefois capturer pleinement les incertitudes inhérentes aux prévisions météorologiques.

Un modèle dont les performances se rapprochent davantage de celles de l'étude de Sabzipour et al. (2023) est le modèle sans TA mais avec ATT. En effet, pour les deux premiers jours, le modèle de LSTM_Sabzipour semble légèrement supérieur, mais à partir du jour 3, les performances des deux modèles deviennent comparables. Enfin, le modèle ED sans TA ni ATT affiche les résultats les plus faibles, avec un KGE nettement inférieur sur toute la période de prévision. Cet écart souligne l'importance du transfert d'apprentissage et de l'attention dans l'amélioration des prévisions, notamment pour maintenir la précision sur un horizon de prévision plus long.

5.5 Contributions de l'étude

Bien que l'étude de Sabzipour et al. (2023) ait démontré de bonnes performances avec un LSTM classique, la présente étude propose d'introduire plusieurs niveaux de complexité afin de mieux représenter les dynamiques hydrologiques et l'incertitude associée aux prévisions météorologiques.

Tout d'abord, l'adoption d'une architecture encodeur-décodeur permet de dissocier la représentation des données historiques, considérées comme fiables, de celle des prévisions météorologiques, intrinsèquement incertaines. L'encodeur est ainsi dédié à l'apprentissage des conditions passées, tandis que le décodeur apprend à modéliser l'incertitude future.

Ensuite, pour tirer pleinement parti des prévisions d'ensemble fournies par ECMWF, un décodeur multi-membre a été implémenté. Ce dernier introduit une dimension supplémentaire dans les données en générant plusieurs scénarios de prévision à partir d'un même contexte initial. Il devient ainsi possible de contrôler explicitement la variance des membres via la fonction-objectif, offrant un levier direct sur la calibration des ensembles.

Afin d'enrichir le jeu de données d'entraînement, une stratégie de transfert d'apprentissage a également été intégrée. Les données d'observation ont été utilisées comme pseudo-prévisions couvrant une période historique plus longue, et un bruit aléatoire, croissant avec l'éloignement temporel, a été ajouté pour simuler l'incertitude associée aux véritables prévisions météorologiques. Cette méthode permet d'accroître la diversité des données tout en maintenant une structure d'incertitude.

Par ailleurs, le modèle a été équipé d'un mécanisme d'attention permettant d'extraire des informations temporelles plus pertinentes avant chaque prévision. Concrètement, l'attention évalue la similarité entre l'état caché courant et les représentations passées, ce qui permet au modèle d'identifier des journées antérieures présentant des conditions hydrométéorologiques analogues. Ce processus contribue à des prévisions plus contextualisées et éclairées. Il serait d'ailleurs pertinent, dans une perspective future, d'analyser ces journées sélectionnées par l'attention afin de mieux comprendre les critères sur lesquels le modèle s'appuie pour faire ses prévisions.

Enfin, un aspect peu abordé dans la littérature concerne le rôle de la fonction-objectif dans le contrôle de l'apprentissage des membres de l'ensemble. En adaptant la fonction de perte pour

intégrer des contraintes sur la variance, la calibration, ou des sections critiques tel que les crues, cette étude démontre qu'il est possible de réguler les prévisions de manière explicite.

5.6 Limitations

Cette section met en évidence cinq principales limitations de cette étude, qui influencent tant la qualité des prévisions que la portée des résultats obtenus. Ces limitations concernent la disponibilité des données météorologiques, les choix architecturaux du modèle, la durée d'entraînement restreinte, la généralisation des résultats à d'autres bassins versants et l'absence de comparaison avec des modèles hydrologiques classiques.

L'une des principales limitations concerne la disponibilité des données météorologiques d'observation issues d'ERA5. En effet, ces données sont fournies avec un délai de trois jours, ce qui constitue un obstacle pour une mise en production en temps réel. Afin de pallier cette contrainte, il serait nécessaire d'intégrer une autre source de données météorologiques pour couvrir ces données manquantes en contexte opérationnel.

L'architecture adoptée repose sur 50 décodeurs parallèles, chacun associé à un membre d'ensemble. Cette approche permet d'optimiser l'entraînement en exploitant la diversité des ensembles de prévision. Toutefois, si l'on envisageait une approche séquentielle, où un seul décodeur traiterait les 50 ensembles un à un, un problème de sur-apprentissage surviendrait. Dans cette configuration, le modèle mémorise les informations de l'encodeur et reproduit les prévisions, compromettant ainsi sa capacité de généralisation aux ensembles de validation et de test.

Il est aussi à noter que la période de disponibilité des données de prévisions du ECMWF était courte, limitant de surcroit le potentiel d'entrainement des modèles sur ces données. Il serait intéressant de reprendre cette étude avec des données d'autres centres qui archivent leurs données depuis plus longtemps, malgré une moins bonne qualité prévisionnelle. Cependant, il faudrait s'assurer que les modèles prévisionnels demeurent stables dans le temps pour ne pas

introduire des biais au fur et à mesure que les modèles de prévision s'améliorent avec la recherche continue réalisée dans ce domaine.

De plus, cette étude a été réalisée sur un seul bassin versant, ce qui est une limitation en soi car il est difficile de statuer sur la généralisation des résultats à d'autres sites. Il est possible que le bassin versant du Lac Saint-Jean soit particulier et que d'autres bassins versants ne soient pas aussi facilement modélisables. Idéalement, il faudrait reprendre l'étude sur une multitude de bassins versants et analyser les performances en fonction des conditions hydrométéorologiques et géophysiques des sites.

Enfin, cette étude ne propose pas une comparaison directe entre le modèle développé et des modèles hydrologiques physiques ou conceptuels. Une telle comparaison pourrait fournir un cadre plus clair sur la valeur ajoutée de l'approche basée sur l'apprentissage profond par rapport aux méthodes traditionnelles. L'intégration d'un modèle physique de référence permettrait d'évaluer si les modèles empiriques peuvent capturer efficacement les processus hydrologiques sous-jacents ou s'ils nécessitent des ajustements pour mieux représenter la dynamique des bassins versants.

CONCLUSION ET RECOMMANDATIONS

Cette étude visait à développer un modèle de prévision hydrologique basé sur l'intelligence artificielle pour le bassin versant du Lac-Saint-Jean, en intégrant des approches avancées telles que le transfert d'apprentissage et le mécanisme d'attention. À travers une méthodologie combinant l'utilisation de données issues d'ERA5 et ECMWF, plusieurs modèles encodeur-décodeur multi-membres ont été évalués et comparés afin de déterminer leur capacité à améliorer la précision des prévisions hydrologiques sur un horizon de 15 jours.

Les résultats ont mis en évidence l'impact positif du transfert d'apprentissage, qui permet d'exploiter un plus large historique de données et d'améliorer la généralisation du modèle. Il a été démontré que les modèles intégrant cette approche surpassent systématiquement ceux entraînés uniquement sur des données plus récentes, réduisant ainsi la perte de précision sur les prévisions à long terme. L'analyse des métriques, notamment le KGE et le CRPS, a également révélé que l'ajout du mécanisme d'attention pouvait, dans certaines conditions, améliorer la qualité des prévisions en renforçant la prise en compte des relations temporelles complexes entre les variables hydrométéorologiques.

Cette étude s'inscrit dans la continuité des travaux antérieurs en modélisation hydrologique par réseaux de neurones récurrents, notamment celui de Sabzipour et al. (2023). En s'appuyant sur ces fondations, elle introduit plusieurs couches de complexité supplémentaires, telles que l'architecture encodeur-décodeur, le décodeur multi-membre, le transfert d'apprentissage, et le mécanisme d'attention. Ces éléments ont permis de mieux représenter l'incertitude et de capter les relations temporelles complexes entre les variables hydrométéorologiques. Les résultats obtenus confirment la valeur ajoutée de ces innovations, justifiant ainsi la poursuite et l'approfondissement de ces approches avancées dans les futures études de prévision hydrologique.

Malgré ces avancées, certaines limites ont été identifiées. D'une part, la disponibilité restreinte des données de prévision ECMWF a limité la période d'entraînement du modèle, réduisant

ainsi son potentiel d'apprentissage à long terme. D'autre part, l'étude s'est concentrée sur un seul bassin versant, ce qui pose des questions quant à la généralisabilité des résultats à d'autres contextes hydrologiques. Une validation sur plusieurs bassins serait donc nécessaire pour confirmer l'efficacité du modèle dans des conditions variées. Dans la même veine, le modèle est actuellement entraîné et testé sur un unique bassin hydrologique. Une généralisation à plusieurs bassins permettrait d'améliorer les résultats en augmentant les jeux de données (Kratzert et al., 2024).

Actuellement, chaque ensemble est attribué à un décodeur unique, ce qui limite le nombre total de prévisions générées à 50 par phase de test. Une alternative intéressante serait d'implémenter une rotation des ensembles sur l'ensemble des décodeurs, permettant ainsi de produire jusqu'à 2 500 prévisions. Bien que la pertinence de cette approche ne soit pas garantie, elle offrirait une opportunité d'analyser la dispersion des prévisions et d'ajuster les méthodes de calibration en conséquence.

D'ailleurs, le mécanisme d'attention, quant à lui, a démontré une plus-value dans l'amélioration des prévisions, surtout lorsqu'il est combiné au transfert d'apprentissage. Une piste d'amélioration consisterait à tester des architectures plus avancées, telles que les transformeurs ou les TFT.

Finalement, ces travaux constituent une avancée dans l'application des modèles d'apprentissage profond pour la prévision hydrologique en contexte opérationnel. Ils ouvrent la voie à de nombreuses pistes d'exploration, que ce soit en optimisant l'architecture des modèles, en élargissant les sources de données ou en intégrant des approches hybrides. Les résultats obtenus dans ce mémoire démontrent ainsi le potentiel des approches de type encodeur-décodeur pour améliorer la gestion des ressources hydriques, contribuant ainsi à une prise de décision plus éclairée pour la production hydroélectrique au Lac Saint-Jean et ailleurs dans le monde.

ANNEXE I

ÉVALUATION SELON LE KGE

Tableau-A I-1 Évaluation des performances des modèles ED et ED-TA selon les ensembles d'entraînement, de validation et de test selon la métrique KGE

	ED			ED-TA		
Jour	Entrainment	Validation	Test	Entrainment	Validation	Test
1	0.98	0.87	0.88	0.98	0.96	0.99
2	0.99	0.86	0.88	0.98	0.97	0.99
3	0.99	0.84	0.88	0.98	0.96	0.98
4	0.99	0.81	0.88	0.98	0.95	0.97
5	0.98	0.79	0.88	0.98	0.94	0.97
6	0.98	0.76	0.87	0.97	0.94	0.96
7	0.98	0.73	0.87	0.97	0.93	0.96
8	0.98	0.71	0.86	0.96	0.92	0.96
9	0.98	0.69	0.85	0.96	0.91	0.95
10	0.98	0.68	0.84	0.95	0.90	0.94
11	0.98	0.68	0.84	0.95	0.89	0.94
12	0.98	0.67	0.83	0.94	0.88	0.93
13	0.97	0.67	0.83	0.93	0.87	0.92
14	0.97	0.66	0.83	0.93	0.87	0.90
15	0.97	0.66	0.82	0.93	0.86	0.89

Tableau-A I-2 Évaluation des performances des modèles ED-ATT et ED-ATT-TA selon les ensembles d'entraînement, de validation et de test selon la métrique KGE

	ED-ATT			ED-ATT-TA		
Jour	Entrainment	Validation	Test	Entrainment	Validation	Test
1	0.99	0.92	0.90	0.97	0.99	0.98
2	0.97	0.88	0.94	0.97	0.99	0.99
3	0.97	0.86	0.94	0.98	0.98	0.98
4	0.98	0.84	0.94	0.98	0.96	0.97
5	0.99	0.82	0.93	0.98	0.95	0.97
6	0.99	0.79	0.92	0.98	0.95	0.97
7	0.99	0.77	0.91	0.98	0.94	0.97
8	0.99	0.75	0.90	0.98	0.93	0.96
9	0.99	0.73	0.89	0.98	0.93	0.96
10	0.99	0.71	0.89	0.97	0.93	0.95
11	0.99	0.70	0.88	0.96	0.92	0.95
12	0.98	0.69	0.87	0.95	0.91	0.94
13	0.98	0.68	0.87	0.95	0.90	0.93
14	0.98	0.67	0.86	0.94	0.89	0.92
15	0.98	0.66	0.86	0.94	0.88	0.91

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

Amanambu, A. C., Mossa, J., & Chen, Y.-H. (2022). Hydrological Drought Forecasting Using a Deep Transformer Model. *Water*, *14*(22), 3611. https://doi.org/10.3390/w14223611

Arsenault, R., & Côté, P. (2019). Analysis of the effects of biases in ensemble streamflow prediction (ESP) forecasts on electricity production in hydropower reservoir management. *Hydrology and Earth System Sciences*, *23*(6), 2735-2750. https://doi.org/10.5194/hess-23-2735-2019

Arsenault, R., Martel, J.-L., Brunet, F., Brissette, F., & Mai, J. (2023). Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models. *Hydrology and Earth System Sciences*, *27*(1), 139-157. https://doi.org/10.5194/hess-27-139-2023

Bahdanau, D., Cho, K., & Bengio, Y. (2016, 19 mai). Neural Machine Translation by Jointly Learning to Align and Translate. arXiv. https://doi.org/10.48550/arXiv.1409.0473

Boucher, M.-A., Tremblay, D., Delorme, L., Perreault, L., & Anctil, F. (2012). Hydro-economic assessment of hydrological forecasting systems. *Journal of Hydrology*, *416-417*, 133-144. https://doi.org/10.1016/j.jhydrol.2011.11.042

Brownlee, J. (2017, 5 décembre). What is Teacher Forcing for Recurrent Neural Networks? *MachineLearningMastery.com*. Repéré à https://machinelearningmastery.com/teacher-forcing-for-recurrent-neural-networks/

Chang, A. Y.-Y., Bogner, K., Ramos, M.-H., Harrigan, S., Domeisen, D. I. V., & Zappa, M. (2024). *Using Temporal Fusion Transformer (TFT) to enhance sub-seasonal drought predictions in the European Alps* (Rapport No. EGU24-12981). Communication présentée au EGU24, Copernicus Meetings. https://doi.org/10.5194/egusphere-egu24-12981

Charbonneau, R., Fortin, J.-P., & Morin, G. (1977). The CEQUEAU model: description and examples of its use in problems related to water resource management / Le modèle CEQUEAU: description et exemples d'utilisation dans le cadre de problèmes reliés à l'aménagement. *Hydrological Sciences Bulletin*, *22*(1), 193-202. https://doi.org/10.1080/02626667709491704

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, *375*(3), 613-626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Dai, Z., Zhang, M., Nedjah, N., Xu, D., & Ye, F. (2023). A Hydrological Data Prediction Model Based on LSTM with Attention Mechanism. *Water*, *15*(4), 670. https://doi.org/10.3390/w15040670

Demiray, B. Z., & Demir, I. (2024, 11 juin). Towards Generalized Hydrological Forecasting using Transformer Models for 120-Hour Streamflow Prediction. arXiv. https://doi.org/10.48550/arXiv.2406.07484

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia*, *4*, 1001-1007. https://doi.org/10.1016/j.aqpro.2015.02.126

Fan, F. M., Schwanenberg, D., Alvarado, R., Assis dos Reis, A., Collischonn, W., & Naumman, S. (2016). Performance of Deterministic and Probabilistic Hydrological Forecasts for the Short-Term Optimization of a Tropical Hydropower Reservoir. *Water Resources Management*, *30*(10), 3609-3625. https://doi.org/10.1007/s11269-016-1377-8

Francisco, R., & Matos, J. P. (2024). Deep Learning Prediction of Streamflow in Portugal. *Hydrology*, *11*(12), 217. https://doi.org/10.3390/hydrology11120217

Girihagama, L., Naveed Khaliq, M., Lamontagne, P., Perdikaris, J., Roy, R., Sushama, L., & Elshorbagy, A. (2022). Streamflow modelling and forecasting for Canadian watersheds using LSTM networks with attention mechanism. *Neural Computing and Applications*, *34*(22), 19995-20015. https://doi.org/10.1007/s00521-022-07523-8

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80-91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. Repéré à https://journals.ametsoc.org/view/journals/wefo/15/5/1520-0434_2000_015_0559_dotcrp_2_0_co_2.xml

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. https://doi.org/10.1002/qj.3803

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424-425*, 264-277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Kratzert, F., Gauch, M., Klotz, D., & Nearing, G. (2024). HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin. *Hydrology and Earth System Sciences*, *28*(17), 4187-4201. https://doi.org/10.5194/hess-28-4187-2024

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005-6022. https://doi.org/10.5194/hess-22-6005-2018

Li, Z., Kiaghadi, A., & Dawson, C. (2021). Exploring the best sequence LSTM modeling architecture for flood prediction. *Neural Computing and Applications*, *33*, 1-10. https://doi.org/10.1007/s00521-020-05334-3

Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2020, 27 septembre). Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. arXiv. https://doi.org/10.48550/arXiv.1912.09363

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, *15*(1), 101-124. https://doi.org/10.1016/S1364-8152(99)00007-9

Mbuvha, R., Adounkpe, J. Y. P., Houngnibo, M. C. M., Mongwe, W. T., Nikraftar, Z., Marwala, T., & Newlands, N. K. (2023). A novel workflow for streamflow prediction in the presence of missing gauge observations. *Environmental Data Science*, *2*, e23. https://doi.org/10.1017/eds.2023.11

O. Talagrand, R. V. (1997). Evaluation of probabilistic prediction systems. *ECMWF*. [text]. Repéré à https://www.ecmwf.int/en/elibrary/76596-evaluation-probabilistic-prediction-systems

Ouachani, R., Bargaoui, Z., & Ouarda, T. (2007). Integration d'un filtre de Kalman dans le modele hydrologique HBV pour la prevision des debits / Integration of a Kalman filter in the HBV hydrological model for runoff forecasting. *Hydrological Sciences Journal-journal Des Sciences Hydrologiques - HYDROLOG SCI J*, *52*, 318-337. https://doi.org/10.1623/hysj.52.2.318

Rasiya Koya, S., & Roy, T. (2024). Temporal Fusion Transformers for streamflow Prediction: Value of combining attention with recurrence. *Journal of Hydrology*, *637*, 131301. https://doi.org/10.1016/j.jhydrol.2024.131301

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533-536. https://doi.org/10.1038/323533a0

Sabzipour, B., Arsenault, R., Troin, M., Martel, J.-L., Brissette, F., Brunet, F., & Mai, J. (2023). Comparing a long short-term memory (LSTM) neural network with a physically-based

hydrological model for streamflow forecasting over a Canadian catchment. *Journal of Hydrology*, *627*, 130380. https://doi.org/10.1016/j.jhydrol.2023.130380

Sahoo, B. B., Jha, R., Singh, A., & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica*, *67*(5), 1471-1481. https://doi.org/10.1007/s11600-019-00330-1

Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, *24*(5), 2527-2544. https://doi.org/10.5194/hess-24-2527-2020

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J.-L. (2021). Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years. *Water Resources Research*, *57*(7), e2020WR028392. https://doi.org/10.1029/2020WR028392

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is All you Need. Dans *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. Repéré à https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., & Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, *5*(1), 13-26. https://doi.org/10.5194/hess-5-13-2001

Waqas, M., & Humphries, U. W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*, *13*, 102946. https://doi.org/10.1016/j.mex.2024.102946

Wetterhall, F., Pappenberger, F., Cloke, H. L., Thielen-del Pozo, J., Balabanova, S., Daňhelka, J., … Holubecka, M. (2013, 20 février). Forecasters priorities for improving probabilistic flood forecasts. Hydrometeorology/Modelling approaches. https://doi.org/10.5194/hessd-10-2215-2013

Yan, L., Chen, C., Hang, T., & Hu, Y. (2021). A stream prediction model based on attention-LSTM. *Earth Science Informatics*, *14*(2), 723-733. https://doi.org/10.1007/s12145-021-00571-z