

Utilisation d'un modèle de mémoire long-court terme entraîné
sur des rétro-prévisions pour la prévision des apports en eau
dans le réservoir hydroélectrique du Lac Saint-Jean

par

Léo SOUCY

MÉMOIRE PAR ARTICLES PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE
SUPÉRIEURE COMME EXIGENCE PARTIELLE À L'OBTENTION DE
LA MAÎTRISE AVEC MÉMOIRE EN GÉNIE DE LA CONSTRUCTION
M. Sc. A.

MONTRÉAL, LE 15 AVRIL 2025

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Léo Soucy, 2025



Cette licence [Creative Commons](#) signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

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. Jean-Luc Martel, codirecteur de mémoire
Département de génie de la construction à l'École de technologie supérieure

Mme. Janie Masse-Dufresne, présidente du jury
Département de génie de la construction à l'École de technologie supérieure

Mme. Annie Poulin, 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 10 AVRIL 2025

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens avant tout à remercier mon directeur de mémoire Richard Arsenault, Rich pour les intimes, qui a été bien plus qu'un simple directeur. Tu as également joué les rôles de mentor, de confident, d'ami et, dans mes nombreux moments d'anxiété, de psychologue. Richard, tu es une personne exceptionnelle, et je me considère incroyablement chanceux de t'avoir eu comme guide tout au long de ma maîtrise. Il serait difficile de lister tout ce qui te rend unique, mais une mention spéciale s'impose pour ta disponibilité inégalée, littéralement 24/7. Pour dire, je n'étais même plus surpris de recevoir des messages de ta part en pleine nuit de samedi à dimanche ! Plus sérieusement, tu as su me diriger d'une main de maître, m'offrant des conseils judicieux qui m'ont permis de progresser à un rythme soutenu, sans jamais m'imposer de pression. Malgré ton statut émérite, tu ne démontres aucune prétention et restes très terre à terre, des qualités qui te rendent si agréable à côtoyer. Ton ouverture d'esprit, ta capacité à m'écouter et à comprendre mes besoins m'ont aidé à maintenir le cap et à rester motivé, même dans les moments difficiles où, je l'avoue, je n'étais probablement pas toujours facile à approcher. J'aurais adoré que tu m'enseignes au moins un cours, mais travailler avec toi (souvent à distance !) durant cette étape charnière de mon parcours m'a permis de comprendre pourquoi tu es autant apprécié par tes étudiants et tes collègues. Merci de m'avoir offert cette opportunité de recherche et de m'avoir appuyé sans relâche, j'ai bien de la misère à m'imaginer ce à quoi aurait pu ressembler ma maîtrise sans toi. Avec tes multiples talents et ton ardeur à l'ouvrage, je n'ai aucun doute que ta carrière de professeur et de chercheur continuera à prospérer. J'espère sincèrement que nous resterons en contact.

Merci également à mon codirecteur de mémoire Jean-Luc Martel, aka JL, qui a su me conseiller et prendre le relai de façon formidable lors des rares moments d'indisponibilité de Richard. Tu es un gars passionnant, à l'énergie contagieuse et très facile d'approche, malgré que ton débit de paroles puisse être intimidant! J'ai énormément apprécié t'avoir comme professeur en hydraulique urbaine, et tes enseignements sur l'apprentissage machine ont posé des bases solides pour mon projet de maîtrise. Tu as tout pour connaître une carrière remarquable et tu

es sans conteste sur la bonne voie. Merci pour ton soutien et ta passion, JL. Je te souhaite sincèrement le meilleur pour la suite.

Je tiens également à exprimer ma gratitude envers Rio Tinto qui a financé ce projet de recherche ainsi qu'envers les différentes institutions qui m'ont accordé une bourse d'études aux cycles supérieurs : l'ÉTS, le CRSNG, le FRQNT et Hydro-Québec. Grâce à leur soutien financier, j'ai pu me consacrer pleinement à ma maîtrise, en travaillant librement, sans avoir à me préoccuper des contraintes financières. Ces bourses m'ont permis de me concentrer sur mes recherches et de tirer le meilleur de cette expérience académique.

À un niveau plus personnel, merci à mes collègues à la maîtrise et au doctorat du laboratoire HC3, qui sont devenus au fil du temps des amis précieux. Je suis extrêmement fier des liens d'amitié solides que nous avons tissés et reconnaissant de pouvoir compter sur des personnes en qui j'ai autant confiance. J'espère sincèrement que notre petit groupe restera uni, peu importe où nos chemins respectifs nous mèneront. Je souhaite que nous continuions à nous retrouver et à partager ces moments remplis de plaisir !

Finalement, je remercie du plus profond du cœur ma famille et plus particulièrement mes parents pour leur support tout au long non seulement de ma maîtrise, mais également de l'ensemble de mon parcours scolaire. Avec ma tendance à l'anxiété de performance (léger héritage génétique il faut dire!), il m'est souvent arrivé de me sentir dépassé et de « capoter » un peu beaucoup alors qu'il n'y avait pas lieu de m'inquiéter autant. Dans ces moments de tension, vous avez toujours su m'écouter, garder votre calme et me faire réaliser qu'en fait je m'en faisais pour... pas grand-chose. Je sais que vous éprouvez un amour inconditionnel envers moi et sachez que c'est réciproque, même si mon attitude peut parfois le mettre en doute. Merci de m'avoir offert un environnement si sain et propice à la réussite, un chez-nous dans lequel je me sens chez-moi, un cadre empreint de bienfaisance, de bonheur et d'amour. Cet équilibre m'a permis d'être dans les meilleures dispositions pour avancer et m'épanouir. Je n'aurais pu demander mieux et vous avez une immense part dans l'homme confiant en ses moyens et épanoui que je suis aujourd'hui.

Utilisation d'un modèle de mémoire long-court terme entraîné sur des rétro-prévisions pour la prévision des apports en eau dans le réservoir hydroélectrique du Lac Saint-Jean

Léo SOUCY

RÉSUMÉ

L'essor de l'intelligence artificielle (IA) transforme de nombreux secteurs clés de notre société, y compris l'hydrologie. Parmi les avancées marquantes et à fort potentiel se trouve l'apprentissage profond, une branche de l'IA. Ce travail exploite un modèle d'apprentissage profond (Deep Learning, DL) pour prévoir les apports en eau cumulés sur une période de 14 jours vers un réservoir hydroélectrique alimenté par le bassin versant du Lac-Saint-Jean. Le modèle utilisé, basé sur les réseaux à mémoire long terme et court terme (Long Short-Term Memory, LSTM), s'est récemment distingué par ses performances encourageantes dans diverses applications de prévision hydrologique. Par ailleurs, l'intégration de rétro-prévisions météorologiques dans l'entraînement de certains modèles est évaluée afin de déterminer leur impact sur la performance des modèles et la qualité des prévisions volumétriques produites.

Les résultats révèlent que les modèles LSTM génèrent des prévisions de qualité tout en mettant en évidence des avantages et des limites spécifiques. Globalement, les prévisions d'ensemble des modèles se distinguent par leur précision, mais elles manquent de fiabilité et présentent une sous-dispersion persistante, quelle que soit la saison. Les modèles rencontrent notamment des difficultés à reproduire avec exactitude le moment et l'amplitude de la crue printanière, échouant fréquemment à prévoir correctement les débits extrêmes. Par ailleurs, la recherche montre que l'intégration de rétro-prévisions lors de l'entraînement améliore la précision et la netteté des prévisions, mais au détriment de leur robustesse.

En complément, une évaluation saisonnière a permis d'examiner le comportement des modèles en fonction des conditions hydroclimatiques propres à chaque saison. Les performances des modèles s'avèrent plus limitées en hiver et au printemps, saisons où les conditions hydrologiques complexes, telles que la présence de neige et les variations rapides des débits, posent des défis supplémentaires. Dans ces périodes, les modèles entraînés avec des rétro-prévisions sont supérieurs en termes de précision. En revanche, les prévisions estivales et automnales sont de meilleure qualité, mais l'utilisation des rétro-prévisions durant l'entraînement tend à diminuer la précision des ensembles. Sur le plan de la fiabilité, une sous-dispersion importante est constatée pour tous les modèles. Cependant, le modèle sans rétro-prévisions se révèle plus robuste dans l'ensemble, particulièrement en été et en automne. La pertinence des rétro-prévisions en contexte opérationnel dépend donc des objectifs établis, comme la gestion des conditions hydriques extrêmes.

Mots-clés : Prévision hydrologique d'ensemble, modèle d'apprentissage-machine, rétro-prévisions, volumes d'eau, Deep Learning, LSTM

Application of a long short-term memory model trained on meteorological reforecasts for forecasting inflow volumes to the Lac-Saint-Jean hydroelectric reservoir

Léo SOUCY

ABSTRACT

The rise of artificial intelligence (AI) is transforming numerous key sectors of our society, including hydrology. Among the most impactful and promising advances is deep learning, a branch of AI. This study leverages a deep learning (DL) model to forecast cumulative water inflows over a 14-day period into a hydroelectric reservoir supplied by the Lac Saint-Jean watershed. The model employed is based on Long Short-Term Memory (LSTM) networks, which have recently gained attention for their strong performance in various hydrological forecasting applications. Additionally, meteorological reforecasts are integrated into the training of some models to assess their influence on model performance and the quality of volumetric forecasts.

The results show that LSTM models generate high-quality forecasts while highlighting specific benefits and limitations. Overall, the ensemble forecasts produced by the models are accurate but lack reliability, exhibiting persistent under-dispersion across all seasons. The models notably struggle to accurately capture the timing and magnitude of the spring freshet, frequently failing to predict extreme flows. Furthermore, the research indicates that incorporating reforecasts during training enhances forecast precision and sharpness but at the expense of robustness.

In addition, a seasonal evaluation was conducted to analyze model behavior under the unique hydroclimatic conditions of each season. Model performance is more limited in winter and spring, where complex hydrological conditions, such as snow presence and rapid flow variations, pose additional challenges. During these seasons, models trained with reforecasts outperform the baseline model in terms of accuracy. Conversely, summer and autumn forecasts are of higher quality, but the use of reforecasts during training tends to reduce ensemble precision. Regarding reliability, significant under-dispersion is observed across all models. However, the baseline model without reforecasts demonstrates greater overall robustness, particularly during summer and autumn. The operational relevance of reforecasts thus depends on the specific forecasting objectives, such as managing extreme hydrological conditions.

Keywords: Ensemble hydrological forecasting, machine learning model, reforecasts, inflow volumes, deep learning, LSTM

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 REVUE DE LITTÉRATURE	5
1.1 Hydrologie	5
1.2 Prévision météorologique	6
1.2.1 Prévisions déterministes et ensemblistes	8
1.2.2 Rétro-prévisions	9
1.3 Modélisation hydrologique	11
1.3.1 Modèles classiques	11
1.3.2 Modèles basés sur l'IA	13
1.4 Prévision hydrologique	17
1.4.1 Prévisions déterministes et ensemblistes	17
1.4.2 Sources d'incertitude en prévision hydrologique	19
1.4.3 Prévision de volumes d'eau	21
1.5 Traitement statistique	22
1.6 Prévision des apports volumétriques en eau	23
1.7 Objectifs de la recherche	25
CHAPITRE 2 ASSESSING THE VALUE OF METEOROLOGICAL REFORECAST DATA TO PREDICT INFLOW VOLUMES OVER A CANADIAN SNOW-DOMINATED CATCHMENT USING A DEEP LEARNING MODEL	27
2.1 Abstract	27
2.2 Introduction	28
2.3 Methods and data	32
2.3.1 Study area	32
2.3.2 Datasets	34
2.3.3 Hydrological modelling using LSTM models	37
2.3.3.1 ERA5-based (E5)	39
2.3.3.2 Reforecast-based (RF)	40
2.3.3.3 Pretrained reforecast model (E5-RF)	41
2.3.4 LSTM Model development	41
2.3.4.1 Data scaling	42
2.3.4.2 Model structure and hyperparameterization	42
2.3.4.3 Model training	43
2.3.5 Application to inflow forecasting	44
2.4 Results	46
2.4.1 LSTM model training, validation and testing results	46
2.4.2 Results of using reforecasts on the testing period	47
2.4.3 Performance of the LSTM models in forecasting	50
2.4.4 Comparison of model performance on the seasonal scale	56

2.4.5	Forecast reliability	58
2.5	Discussion	60
2.5.1	The performance of LSTM models in testing	60
2.5.2	The performance of LSTM models in forecasting	61
2.5.2.1	Issue with timing delay in peak inflow volumes	61
2.5.2.2	Biases in the simulation of low-flow periods	62
2.5.2.3	Challenges in forecasting the spring freshets	64
2.5.2.4	Forecast reliability issues	65
2.5.3	On the impact of using reforecast data in the training of LSTM models	66
2.5.3.1	Impact of reforecast training on LSTM model performance	66
2.5.3.2	Seasonal performance differences	67
2.5.3.3	Forecast reliability and ensemble sharpness	67
2.5.4	Added value of the LSTM models	68
2.5.5	Review of the study's objectives	68
2.5.6	Limitations	69
2.6	Conclusion and recommendations	70
2.7	Acknowledgments	71
CHAPITRE 3 DISCUSSION		73
3.1	Analyse de la performance globale des modèles	73
3.2	Analyse de la performance des modèles sur une base saisonnière	78
3.3	Utilité d'intégrer des rétro-prévisions durant l'entraînement des modèles	80
3.4	Limitations rencontrées et améliorations envisagées	83
3.4.1	Ajout de variables en tant qu'entrants au modèle	83
3.4.2	Étude portant sur un seul bassin versant	84
3.4.3	Fonction objectif	85
3.4.4	Hyperparamètres	85
3.4.5	Comparaison avec un modèle opérationnel	86
CONCLUSION		87
RECOMMANDATIONS		89
ANNEXE I	FONCTIONS DE DENSITÉ CUMULATIVE DES VOLUMES OBSERVÉS DURANT LA CALIBRATION DES MODÈLES	91
ANNEXE II	STRUCTURE OPTIMALE DES MODÈLES	93
ANNEXE III	PRÉVISIONS DE VOLUMES GÉNÉRÉES PAR LES MODÈLES DURANT LA PÉRIODE D'ENTRAÎNEMENT	95
LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES		97

LISTE DES TABLEAUX

	Page
Tableau 2.1	35
Tableau 2.2	43
Tableau 2.3	47
Tableau 2.4	57

LISTE DES FIGURES

	Page	
Figure 2.1	Study area and location of the downstream hydropower generating station. The large lake (Lac-Saint-Jean) is the 1,000 km ² reservoir managed by Rio Tinto for hydropower generation.....	33
Figure 2.2	Example implementation of a training sample for the E5 and RF forecasting models (a) and an example implementation of a forecasting sample for all models (b). Panel (a) presents a single sample. However, this process is repeated across the entire dataset available for the training period and similarly applied to the validation and testing periods. Similarly, panel (b) shows an example for a single forecast issue date, but the same process is repeated for all forecast issue dates.....	39
Figure 2.3	Envelopes of the 14-day inflow volumes ensemble forecasts for the LSJ catchment for all models over the reforecast testing period from 2007 to 2009: E5 (a), RF (b), and E5-RF (c). Lighter colors depict the ensemble's full range and darker colors represent the part of the ensemble located within the 20 th and 80 th percentiles	49
Figure 2.4	Average of the 50-member ensemble forecast for the 14-day inflow volume.....	50
Figure 2.5	Envelopes of the 14-day inflow volume curve ensemble forecasts for the E5 model: entire forecasting period (a), strongest (b) and smallest (c) spring freshet. Lighter colors depict the ensemble's full range, while darker colors represent the forecasts located within the 20 th and 80 th percentiles	52
Figure 2.6	Envelopes of the 14-day inflow volume curve ensemble forecasts for the RF model: entire forecasting period (a), strongest (b) and smallest (c) spring freshet. Lighter colors depict the ensemble's full range, while darker colors represent the forecasts located within the 20 th and 80 th percentiles	53
Figure 2.7	Envelopes of the 14-day inflow volume curve ensemble forecasts for the E5-RF model: entire forecasting period (a), strongest (b) and smallest (c) spring freshet. Lighter colors depict the ensemble's full range, while darker colors represent the forecasts located within the 20 th and 80 th percentiles	54

Figure 2.8	Boxplots of 14-day volume ensemble forecasts CRPS values for the E5 (green), RF (orange) and E5-RF (purple) models over winter (a), spring (b), summer (c), and fall (d). The number of points in each boxplot represents the number of issued forecasts for that season. Note that the y-axis range of spring (b) is larger from the other all panels	57
Figure 2.9	Reliability of the forecasts generated by the E5 (green), RF (orange) and E5-RF (purple) models as demonstrated with Talagrand diagrams showing the rank (in percentiles) for the 14-day observed inflow volumes within the 50-member forecast ensembles	59

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ABDU	Average Bin Distance to Uniformity
AI	Artificial Intelligence
AM	Saison de la fonte de neige - Avril à Mai
ANN	Réseaux de neurones artificiels (Artificial Neural Networks)
CRPS	Continuous rank probability score
DJFM	Saison hivernale – Décembre à Mars
DL	Deep Learning
ECMWF	European centre for medium-range weather forecasts
GDPS	Global Deterministic Prediction System
GFS	Global Forecast System
IA	Intelligence artificielle
IPCC	Groupe d'experts intergouvernemental sur l'évolution du climat (Intergovernmental Panel on Climate Change)
JJA	Saison d'été - Juin à Août
KGE	Kling-Gupta Efficiency
LSJ	Lac-Saint-Jean
LSTM	Long Short-Term Memory
RNN	Réseau de neurones récurrents (Recurrent Neural Networks)
SON	Saison d'automne – Septembre à Novembre

LISTE DES SYMBOLES ET UNITÉS DE MESURE

UNITÉS GÉOMÉTRIQUES

Longueur

m	mètre
mm	millimètre
km	kilomètre

Aire

km ²	kilomètre carré (= 1 000 000 m ²)
-----------------	---

Volume

m ³	mètre cube
hm ³	hectomètre cube

UNITÉS DE TEMPS

s	seconde
h	heure
d	jour

Contrainte, pression

Pa	pascal
----	--------

UNITÉS CALORIFIQUES

K	kelvin
°C	degré Celsius

INTRODUCTION

À une époque où les changements climatiques occupent une place centrale dans les débats, l'électricité se distingue comme une source d'énergie dont la demande croît à un rythme effréné, alors que les sociétés modernes s'efforcent de s'éloigner des combustibles fossiles, plus polluants. Ces derniers sont les principaux responsables du changement climatique, représentant plus de 75 % des émissions mondiales de gaz à effet de serre et près de 90 % des émissions de dioxyde de carbone (United Nations, 2024). La transition énergétique repose en grande partie sur l'adoption de sources d'énergie propres, telles que l'électricité. Pour être véritablement durable, cette électricité doit provenir de sources renouvelables, parmi lesquelles figurent l'énergie solaire, éolienne, la géothermie et l'hydroélectricité. Environ 94% de l'électricité produite au Québec provient de l'hydroélectricité (Régie de l'énergie du Canada, 2024), ce qui en fait de loin la principale source d'énergie de la province. La majorité de cette production est assurée par Hydro-Québec grâce à ses nombreuses centrales hydroélectriques. L'importance cruciale d'une gestion adéquate/optimale des ressources en eau apparaît alors évidente, sachant que l'approvisionnement énergétique de la province repose en grande partie sur l'hydroélectricité.

L'hydrologie est un domaine vaste aux applications concrètes et essentielles pour le bon fonctionnement de nos sociétés. Parmi ses nombreux champs d'action, la prévision des débits issus d'un bassin versant en fonction de l'évolution des conditions météorologiques est particulièrement importante. Les prévisions hydrologiques bénéficient à des secteurs variés tels que la production d'hydroélectricité, la prévision et la mitigation des inondations, ainsi que la gestion des ressources en eau (Dion et al., 2021) et sont cruciales pour optimiser l'exploitation des infrastructures hydroélectriques, préserver leur intégrité et garantir la pérennité des écosystèmes environnants. Cependant, générer des prévisions de débits précises demeure complexe en raison d'une multitude de facteurs, dont l'incertitude climatique et des prévisions météorologiques qui en découlent. En réalité, ces prévisions hydrologiques sont des estimations réalisées au mieux des connaissances disponibles, produites par des modèles hydrologiques qui simulent les processus hydrologiques. Il est désormais connu que les

changements climatiques provoqueront une hausse des extrêmes de températures et de précipitations, occasionnant des phénomènes météorologiques extrêmes (sécheresses, inondations, tempêtes, etc.) plus fréquents et intenses (IPCC et al., 2019). Ces perturbations influencent directement le régime hydrologique des bassins versants et modifient les débits à leur exutoire. Dans les bassins nordiques, où la neige joue un rôle clé, les projections indiquent une augmentation des débits hivernaux, une diminution des débits estivaux et une réduction de l'intensité des crues printanières (Chen et al., 2011), bien que ce dernier point fasse encore l'objet de divergences dans la littérature. Ainsi, les projets hydrologiques doivent désormais tenir compte des changements climatiques afin de considérer leurs impacts sur l'hydrologie des bassins versants à l'étude. Adapter nos pratiques à cette nouvelle réalité est essentiel pour limiter les problèmes liés à la disponibilité de l'eau et garantir un usage optimal de cette ressource vitale.

La génération des prévisions hydrologiques de qualité est centrale à une prise de décision éclairée qui tient compte des événements hydrométéorologiques à venir. Ce projet développe et applique des méthodes de prévision hydrologique volumétrique à moyen-terme (14 jours), dont l'amélioration permettrait d'optimiser l'efficacité, la rentabilité et la pérennité des ouvrages hydroélectriques. Plus concrètement, des modèles hydrologiques permettant de prédire des volumes d'apports de manière plus précise et fiable ont le potentiel d'accroître notre capacité à gérer et utiliser plus efficacement et durablement la ressource en eau.

La modélisation hydrologique peut être appliquée en contexte de simulation ou de prévision. La simulation hydrologique consiste à reproduire les débits d'une période passée en s'appuyant sur des données hydrométéorologiques historiques, alors que la prévision hydrologique vise à estimer un débit futur à partir de prévisions météorologiques. Deux approches bien établies sont généralement employées aux fins de la prévision: la prévision déterministe et son équivalent probabiliste.

La prévision déterministe génère une seule valeur de débit en se basant sur la meilleure estimation possible des conditions météorologiques et hydrologiques actuelles et futures. Cette

méthode présente une limitation majeure : elle ne parvient pas à représenter adéquatement les incertitudes de nature météorologique et hydrologique inhérentes à la modélisation hydrologique. Les prévisions déterministes sont encore utilisées de nos jours, même si leurs utilisateurs sont conscients que la divergence d'un scénario unique par rapport à la réalité est susceptible d'entraîner des conséquences majeures.

Pour sa part, la prévision probabiliste produit un ensemble de scénarios équiprobables de débits en intégrant des variations dans les éléments clés du processus de modélisation hydrologique. L'obtention de plusieurs prévisions permet de représenter une gamme de résultats possibles et de mieux quantifier l'incertitude inhérente aux différentes étapes de modélisation. Elle offre ainsi un outil précieux pour améliorer la gestion et la prise de décision, particulièrement en contexte de risque (Krzysztofowicz, 2001).

Les modèles hydrologiques traditionnels sont encore largement utilisés pour émettre des prévisions hydrologiques opérationnelles. Or, l'essor de l'informatique et l'accès croissant à des volumes de données considérables ont favorisé l'émergence de modèles basés sur l'intelligence artificielle (IA), notamment les modèles d'apprentissage profond (Deep Learning ou DL), une sous-catégorie de l'IA prometteuse pour le domaine de l'hydrologie. Les modèles de DL peuvent identifier et exploiter les tendances complexes et les informations variées inhérentes aux systèmes hydrologiques, souvent caractérisés par leur forte non-linéarité et leur complexité. Plus particulièrement, les réseaux de longue mémoire à court terme (Long Short-Term Memory ou LSTM), introduits par Hochreiter et Schmidhuber (1997), suscitent un intérêt croissant dans plusieurs domaines de la recherche en hydrologie, notamment pour la prévision hydrologique, où des études récentes démontrent qu'ils surpassent les modèles hydrologiques traditionnels (Kratzert et al., 2018, 2019a; Hu et al., 2018; Hunt et al., 2022; Sabzipour et al., 2023; Arsenault et al., 2023).

La prévision hydrologique ne se limite pas aux débits; elle peut également concerter les volumes d'eau cumulés sur une certaine période temporelle. La prévision volumétrique, similairement à celle des débits, s'effectue à différents horizons temporels selon les

circonstances et objectifs visés. Le court terme sert davantage aux réservoirs de faible volume utile et à l'optimisation des opérations de turbinage, alors que le long terme est utile aux larges réservoirs pour lesquels la planification de la production d'hydroélectricité s'effectue à plus long terme. L'échelle saisonnière, établie en fonction de périodes d'une durée spécifique, notamment les saisons hydrologiques particulières à un bassin versant, est également d'intérêt. Par exemple, pour les réservoirs nordiques fortement influencés par la crue printanière, des prévisions volumétriques saisonnières précises jouent un rôle central dans la gestion et la planification à long terme des opérations aux réservoirs (Arsenault et al., 2016).

Le contexte présenté ci-dessus établit les bases de ce mémoire, dont l'objectif principal est d'évaluer si l'utilisation de modèles LSTM permet d'améliorer la qualité des prévisions des volumes d'eau cumulés sur une période de 14 jours, avec comme cas d'étude le bassin versant du Lac-Saint-Jean (LSJ). Ce mémoire comporte quatre chapitres, qui suivent l'introduction. Le premier chapitre propose une revue de littérature approfondie, offrant une analyse détaillée des travaux réalisés jusqu'à ce jour sur la modélisation hydrologique et d'autres éléments centraux de ce mémoire, dont les modèles LSTM et la prévision des volumes d'eau. Le deuxième chapitre présente un article scientifique soumis au journal *Journal of Hydrology*. Cet article montre l'évolution du projet de recherche, incluant les données utilisées, la description de la zone d'étude, la méthodologie mise en œuvre, ainsi que les résultats et leur analyse. Le troisième chapitre est dédié à une discussion approfondie, qui se base sur les résultats principaux de l'article pour y ajouter des détails complémentaires et soulever certaines limitations du projet de recherche. Enfin, le quatrième chapitre conclut le mémoire en résumant les principales contributions et en formulant des recommandations pour les recherches futures.

CHAPITRE 1

REVUE DE LITTÉRATURE

Cette section présente une revue de littérature centrée sur la modélisation hydrologique, en portant une attention particulière aux avancées en matière d'utilisation des modèles LSTM, de prévision des volumes d'eau, ainsi que des approches et techniques existantes dans le domaine de la prévision hydrologique.

1.1 Hydrologie

L'hydrologie est la science qui étudie le mouvement de l'eau sur Terre. Elle traite d'où on retrouve l'eau, de sa circulation et de sa distribution, de ses propriétés chimiques et physiques, ainsi que de sa réaction avec l'environnement, y compris sa relation avec les êtres vivants (Devia et al., 2015). Le cycle de l'eau, ou cycle hydrologique, englobe les processus continus permettant à l'eau de circuler sur la planète tels que les précipitations et l'évaporation. Au sein du secteur d'activités du génie civil, l'hydrologie occupe une place prépondérante dans la conception et l'exploitation des ouvrages, en compagnie d'autres domaines d'ingénierie tels l'hydraulique et la géotechnique. Elle englobe un large éventail de responsabilités, de la saine gestion de la ressource en eau à la protection de l'environnement, en passant par la sécurité des ouvrages. Malgré les avancées réalisées et les connaissances développées au fil du temps, la compréhension des mécanismes prenant part au cycle hydrologique continue d'évoluer. Toutefois, la complexité et la nature chaotique du cycle nécessitent que des approximations soient faites lorsqu'on tente de le reproduire. Pour modéliser le cycle hydrologique et ses processus le plus fidèlement possible, des modèles hydrologiques sont utilisés. Un grand nombre de modèles ont été développés, chacun détenant ses propres caractéristiques, entre autres au niveau de sa structure, ses paramètres et ses applications potentielles. Les modèles permettent d'effectuer ce qui est appelé de la modélisation hydrologique, qui est subdivisée en deux catégories : la simulation et la prévision hydrologique. Ces deux approches se distinguent par l'intervalle temporel séparant le débit modélisé du moment de la modélisation. La simulation a pour but de reconstituer le débit en un point précis d'un cours d'eau ou d'un

réservoir. Autrement dit, en se concentrant sur une période historique pour laquelle les données sont connues, le modèle tente de réduire au maximum l'écart entre les valeurs simulées et observées, signifiant une reproduction fidèle du cycle hydrologique en place. Pour sa part, la prévision hydrologique calque presque la simulation, à l'exception qu'elle vise une période future. Elle cherche en effet à prédire des débits futurs, pour laquelle aucune référence actuelle n'existe. Cette distinction est essentielle puisque les deux types de modélisation ne font pas usage des mêmes données pour parvenir à leurs fins. La simulation se base sur les données historiques alors que le processus prévisionnel calcule les valeurs futures à partir de prévisions météorologiques.

Les sections à venir présentent le parcours et les avancées qu'ont connu la prévision hydrologique au fil du temps, un aspect essentiel au domaine de l'hydrologie, sur lequel beaucoup de recherche est effectuée depuis plusieurs décennies. Tout d'abord, les étapes venant en amont de la prévision hydrologique, soit la prévision météorologique puis la modélisation hydrologique, sont détaillées. Une section est ensuite consacrée à la prévision hydrologique elle-même avant d'aborder une méthode reconnue pour améliorer ces prévisions, le traitement statistique. Puis, la revue de littérature se termine par l'indication des objectifs de recherche et s'ouvre vers le cœur de ce mémoire, contenu dans un article soumis dans un journal scientifique.

1.2 Prévision météorologique

La météorologie joue un rôle clé dans de multiples sphères de la société moderne comme le transport de passagers et de marchandises, l'agriculture, le tourisme, l'énergie et la sécurité publique de même que la simple planification d'activités quotidiennes. La prévision météorologique applique la science et la technologie afin de prédire l'état de l'atmosphère pour un emplacement désigné (Abhishek et al., 2012) et contribue à la prise de décision éclairée et une mitigation des risques et dangers climatiques. Un long chemin a été parcouru pour en arriver aux techniques de prévision modernes, qui reposent grandement sur l'intelligence artificielle et la prévision numérique. En effet, les premières prévisions étaient beaucoup plus rudimentaires, en l'occurrence basées sur des données recueillies par observation. Puis,

l'avènement d'instruments de mesure simples tels que le thermomètre, le baromètre et la girouette a permis un suivi et des mesures plus fiables de variables météorologiques fondamentales comme la température, la pression atmosphérique et le vent. Malgré les limites en matière de précision et de couverture de ces instruments, les avancées en météorologie qu'ils ont permises ont jeté des bases solides pour le développement des techniques actuelles. La prévision météorologique moderne s'appuie sur des modèles informatiques, de l'équipement sophistiqué (satellites et radars) et la connaissance des tendances en fonction de l'historique météorologique (Saxena et al., 2013). Ces progrès ont contribué à étendre l'horizon temporel des prévisions météorologiques, qui sont émises à différentes échelles, comprenant le court (jusqu'à cinq jours), le moyen (entre six et quinze jours) et le long terme (plus de quinze jours) (Ehrendorfer, 1997). Il est à noter que ces durées sont assez subjectives et varient selon les objectifs visés et le contexte opérationnel. Au niveau du pas de temps des prévisions, le court terme est souvent représenté par des prévisions effectuées au pas de temps horaire alors que le pas de temps journalier sied mieux le long terme. Il est reconnu que la précision des prévisions diminue avec le temps en raison de la nature chaotique de l'atmosphère, la puissance de calcul massive nécessaire à la résolution des équations, les erreurs potentielles dans les conditions initiales et une compréhension incomplète des processus atmosphériques (Abhishek et al., 2012). À court terme, la capacité de prévision repose davantage sur les détails ponctuels d'événements météorologiques tandis qu'à long terme, elle est liée aux dynamiques se déroulant à grande échelle (Bauer et al., 2015). Le développement de l'intelligence artificielle a donné un nouveau souffle aux modèles météorologiques, qui s'appuient notamment sur cette avancée numérique pour améliorer leurs performances (Saima et al., 2011). De nombreuses organisations produisent des prévisions météorologiques qu'elles s'efforcent de rendre les plus fiables possible, dans le but de mieux comprendre la dynamique des systèmes météorologiques partout sur Terre. Une des plus reconnues, l'European Centre for Medium-Range Weather Forecasts (ECMWF), produit des prévisions météorologiques sur de nombreux horizons temporels à l'échelle mondiale à l'aide de modèles numériques. Au Canada, l'organisme Environnement et Changement climatique Canada (ECCC) fournit des prévisions météorologiques à l'échelle du pays en plus d'offrir d'autres services reliés au climat. L'accès

et le partage des prévisions météorologiques entre organisations contribue à une compréhension plus étendue des phénomènes météorologiques.

1.2.1 Prévisions déterministes et ensemblistes

Deux approches prévisionnelles existent : les prévisions déterministes et celles probabilistes. Une prévision déterministe résulte en un estimé ponctuel de la variable prédictive (Krzysztofowicz, 2001). Autrement dit, une telle prévision génère un seul membre de prévision, soit une valeur pour chaque variable modélisée. Cette approche qui vise la meilleure estimation possible se démarque par sa simplicité et sa faible demande en ressources informatiques, ce qui explique son utilisation répandue dans le passé et encore de nos jours. Malgré que la capacité de prévision soit en constante progression compte tenu des technologies modernes et d'un historique croissant de données, la qualité des prévisions météorologiques demeure minée par plusieurs sources d'incertitudes telles la fiabilité et l'exactitude des données ainsi que la variabilité naturelle du climat. Ces éléments font en sorte qu'il est sensé de ne pas vouloir se fier au seul extrant généré par une prévision déterministe, car en réalité, il existe un ensemble des possibilités futures. L'utilité de telles prévisions individuelles dans un contexte opérationnel est restreinte (J. C. Thompson, 1952), car la probabilité que le scénario généré s'éloigne du scénario réel est bien présente, ce qui pourrait engendrer des conséquences.

La présence de nombreux éléments susceptibles d'intégrer des erreurs dans les prévisions a alimenté la transition des grands centres prévisionnels vers le probabilisme. En effet, à la recherche de prévisions plus précises, mais surtout plus fiables, la communauté météorologique s'est orientée vers l'approche probabiliste. Une prévision probabiliste rend compte sur la distribution envisageable de la variable prédictive (Krzysztofowicz, 2001). Les prévisions météorologiques probabilistes génèrent plusieurs valeurs équiprobales pour chacune des variables modélisées. Une prévision probabiliste regroupe plusieurs membres de prévision, formant ce qu'on appelle communément un ensemble, au sein duquel un membre de prévision correspond à une série de valeurs. Cette approche vise à quantifier l'incertitude inhérente au processus de prévision, qui d'un point de vue météorologique provient surtout de l'estimation des états initiaux et des imprécisions des modèles numériques (Leutbecher et al.,

2008). En contexte d'optimisation à court terme d'opération d'un réservoir hydroélectrique, Fan et al. (2016) ont démontré la supériorité des prévisions météorologiques probabilistes en comparaison à leurs contreparties déterministes. Les résultats de Thiboult et al. (2016) montrent également que l'utilisation de prévisions météorologiques probabilistes améliore la fiabilité des prévisions hydrologiques à long terme. Pour sa part, Krzysztofowicz (2001) vante les mérites de la prévision probabiliste, notamment la quantification de l'incertitude, la prise de décision éclairée en situation de risque et la gestion améliorée en comparaison à la prévision déterministe.

Un contrecoup des prévisions probabilistes est le besoin supérieur en ressources informatiques. En effet, plus il y a de membres de prévisions à générer, plus la capacité informatique requise est élevée. Pour retrouver des besoins informatiques raisonnables, les résolutions spatiales et temporelles des prévisions probabilistes sont souvent révisées à la hausse (plus grossières). La résolution spatiale fait référence à la dimension de la taille des cases d'un modèle alors que la résolution temporelle est le pas de temps auquel le modèle génère des données. Par exemple, le système de prévision AIFS du ECMWF, qui produit des prévisions sur une étendue de 15 jours, a un modèle déterministe de résolution temporelle variant de 1 à 3 heures et de résolution spatiale 0.1° et un modèle probabiliste de résolution temporelle 6 heures et de résolution spatiale 0.2° .

1.2.2 Rétro-prévisions

Afin d'analyser l'erreur historique des modèles prévisionnels et de statuer sur leurs performances antérieures, plusieurs méthodes de correction existent, notamment basées sur les rétro-prévisions (*reforecasts* ou *hindcasts* en anglais). Ce produit de prévision météorologique ensembliste consiste en des prévisions météorologiques générées sur des périodes passées, et ce à partir d'un modèle numérique fixe (Verkade et al., 2013). Les rétro-prévisions sont produites à partir de modèles de prévision météorologiques modernes jumelés à des techniques d'assimilation de données et sont très similaires aux réanalyses, un type de produit météorologique reconnu pour ses capacités d'estimation des états atmosphériques historiques.

Alors que les réanalyses font usage d'un modèle de prévision à court terme pour établir strictement l'état actuel de l'atmosphère, les rétro-prévisions emploient un modèle de prévision à plus long terme. La différence réside dans l'aspect temporel, car au niveau spatial, les deux produits génèrent des données sur une grille de calcul fixe, bien qu'à des résolutions différentes. Dans le cas des réanalyses, le modèle de prévision établit le portrait le plus fidèle possible de l'atmosphère au moment identifié comme étant le temps initial de la prévision météorologique à produire seulement. Pour les rétro-prévisions, le modèle de prévision génère des prévisions d'une durée prolongée et ne se limite donc pas au strict temps initial. Ces prévisions historiques peuvent servir à l'ajustement statistique des modèles de prévision en comparant leurs prévisions aux observations (ce qui s'est réellement produit durant la même période lors des années passées) et à l'analyse de leur habileté à reproduire des phénomènes spécifiques (Abaza et al., 2017). Les forces et les faiblesses des modèles prévisionnels peuvent ainsi être détectées et permettre d'améliorer les modèles en tenant compte des erreurs passées. Les rétro-prévisions sont émises périodiquement et ce pour plusieurs années antérieures, offrant une grande quantité de données fiables et de bonne qualité, étant générées par des modèles de prévision récents. L'évolution rapide des systèmes de prévision météorologiques ont pour effet de raccourcir les historiques de prévision des systèmes (Boucher et al., 2011) ainsi que leur durée de vie opérationnelle, ce qui peut affecter l'utilité des rétro-prévisions. Plusieurs organisations effectuant de la prévision météorologique émettent des rétro-prévisions disponibles publiquement, notamment l'ECMWF et ECCC. L'ECMWF génère des rétro-prévisions à l'échelle planétaire, dont les conditions initiales sont établies par ses modèles d'assimilation opérationnels, créant ainsi une large base de données accessible qui reconstituent les conditions climatiques de nombreuses années passées. Plusieurs études ont analysé la pertinence d'utiliser des rétro-prévisions à des fins de correction statistique de prévisions. Hamill et al. (2006) ont vanté la capacité des rétro-prévisions à améliorer drastiquement les prévisions météorologiques en corrigeant statistiquement les modèles de prévision numérique actuels. Hagedorn (2008) a démontré que les prévisions météorologiques d'un système prévisionnel pouvaient fortement bénéficier de la calibration à l'aide du produit de rétro-prévisions correspondant. D'autres études ont analysé le potentiel des rétro-prévisions en contexte hydrologique. Verkade et al. (2013) sont parvenus à améliorer légèrement la

robustesse des prévisions hydrologiques au détriment de leur résolution en faisant du post-traitement à l'aide de données de rétro-prévisions du ECMWF. Jorg-Hess et al. (2015) ont démontré que la correction statistique de rétro-prévisions, particulièrement les données de précipitations, améliore la modélisation hydrologique durant des périodes de faible hydraulicité.

1.3 Modélisation hydrologique

Un modèle hydrologique peut être défini comme un ensemble d'équations menant à l'estimation du débit en fonction de divers paramètres décrivant les caractéristiques du bassin versant (Devia et al., 2015). Dans la chaîne prévisionnelle, le modèle hydrologique lie la météorologie à l'hydrologie. Ces modèles prennent en entrée les données météorologiques issues des prévisions météorologiques et les traduisent en prévisions hydrologiques. De nombreux modèles hydrologiques ont été développés depuis les débuts de la recherche en hydrologie, chacun se distinguant par sa complexité, ses utilités opérationnelles et ses forces et faiblesses. Les modèles hydrologiques se déclinent en deux types, soit ceux dits « classiques » et ceux basés sur les données. Les modèles classiques sont classifiés comme empiriques, conceptuels ou physiques en fonction de leur paramétrisation ainsi que leur représentation des processus hydrologiques (Avila et al., 2023) et varient en complexité. Les modèles conceptuels et physiques sont des modèles dits *process-driven*.

1.3.1 Modèles classiques

Les modèles conceptuels ou « boîte grise » font partie des plus simples modèles basés sur les processus hydrologiques. Ils représentent ces processus et les différents flux au sein du bassin versant par le biais de conceptualisations et de relations mathématiques simplifiées (Avila et al., 2023). Les équations construites sur la base de relations observées ou supposées entre différentes variables hydrologiques permettent d'optimiser les paramètres internes du modèle pour reproduire le plus fidèlement possible les débits observés. Ces modèles comportent généralement peu de paramètres, ce qui accélère leur calibration, par contre une quantité importante de données hydrométéorologiques est nécessaire à leur calibration (Devia et al.,

2015). Un large éventail de modèles conceptuels existe, chacun possédant sa propre structure interne, équations dictant la façon dont circule l'eau et l'énergie et processus hydrologiques représentés (Knoben et al., 2019). Le choix du modèle varie selon le bassin versant à l'étude puisque la performance des modèles varie grandement selon les conditions hydrologiques rencontrées. En ce qui concerne leur mise en place, les modèles conceptuels nécessitent moins de données et une infrastructure de calcul plus simple que les modèles basés sur la physique, une des raisons pour laquelle ils sont souvent utilisés à des fins opérationnelles (Clark et al., 2008). Malgré leur simplicité, ils peuvent démontrer des performances similaires à d'autres types de modèles.

Les modèles conceptuels globaux sont plus simples en comparaison aux modèles physiques, qui comportent généralement davantage d'équations et d'intrants requis (épaisseur de neige, évapotranspiration, etc.) et qui ont souvent une discréétisation spatiale plus fine. Les modèles physiques ou boite blanche représentent les processus hydrologiques de manière plus détaillée en se basant sur les lois physiques. Ils utilisent les principes de conservation de l'énergie et de la masse dans les flux d'eau et tiennent compte de l'hétérogénéité spatiale du bassin versant (Avila et al., 2023) en le subdivisant en plusieurs unités distinctes selon une grille. La façon avec laquelle cette division est effectuée dicte si le modèle physique est semi-distribué ou distribué. Cette discréétisation spatiale entraîne une paramétrisation et modélisation par unité, permettant une représentation fidèle aux caractéristiques physiques propres à chaque unité. Ces modèles contiennent un grand nombre de paramètres décrivant les caractéristiques physiques du bassin versant. Ils peuvent demander en intrants de nombreuses variables telles que l'humidité du sol ainsi que son type et sa topographie (Devia et al., 2015). La combinaison de ces éléments rend les modèles physiques assez complexes et exigeants en termes d'expertise et de ressources informatiques requises, ce qui complique leur mise en œuvre opérationnelle. En revanche, ils sont applicables à une grande variété de situations et permettent de mieux comprendre et interpréter les principes physiques régulant le bassin versant.

1.3.2 Modèles basés sur l'IA

Une alternative aux modèles hydrologiques décrits précédemment sont les modèles basés sur les statistiques ou boîte noire, connus sous le terme *data-driven*. L'approche statistique ne considère pas explicitement les mécanismes et phénomènes hydrologiques pour supporter ses prévisions. Elle tire sa capacité de prévision en extrayant des informations et tendances au sein des données disponibles par le biais d'analyses statistiques (McCuen, 2016). Avec l'essor de l'informatique et la disponibilité accrue de la donnée, l'utilisation d'une approche particulière parmi les modèles statistiques s'est popularisée : les modèles basés sur l'intelligence artificielle (IA), qui puisent leur pouvoir prédictif uniquement à partir de données.

Une sous-catégorie de l'IA en particulier, l'apprentissage profond (Deep learning ou DL), est une approche prometteuse pour le domaine de l'hydrologie. Elle fait usage de réseaux de neurones artificiels (Artificial neural networks ou ANN) multicouches. Les ANNs sont composés d'une couche d'entrée, de couches intermédiaires cachées et d'une couche de sortie et sont capables d'apprendre efficacement des relations autant linéaires que non-linéaires entre les données d'entrée et de sortie (Sit et al., 2020). La présence de plusieurs couches engendre une meilleure représentation des différentes sources de données (Saba et al., 2019), résultant en une transmission des informations plus complète au sein du modèle. Le DL rend possible l'apprentissage des tendances et des informations diverses contenues au sein de systèmes complexes et non-linéaires, comme les bassins versants rencontrés en hydrologie, et ne nécessitent pas d'instructions explicites vis-à-vis les phénomènes hydrologiques, réduisant ainsi le travail de paramétrisation requis en amont de leur utilisation. En effet, il n'est pas nécessaire de renseigner les modèles avec une structure représentant les processus physiques guidant le comportement hydrologique du bassin versant. Toutefois, les ANN traditionnels ne se rappellent pas l'ordre des entrées du modèle (Kratzert et al., 2018), ce qui constitue un inconvénient compte tenu des séquences temporelles généralement utilisées en prévision hydrologique.

Plusieurs types d'ANN existent, dont les réseaux de neurones récurrents (Recurrent neural networks ou RNN), qui possèdent une couche récurrente permettant un partage des informations apprises au sein d'une même couche du réseau. Ils sont spécifiquement conçus pour comprendre les dynamiques temporelles en considérant les entrées du modèle dans leur ordre séquentiel (Rumelhart et al., 1986). Cette capacité est essentielle pour la prévision hydrologique compte tenu de la nécessité de traiter des séries temporelles, où l'ordre des données a de l'importance pour comprendre puis bien représenter les processus. Toutefois, les RNN ne sont pas en mesure d'apprendre les dépendances à long terme représentant par exemple les effets de stockage, qui peuvent jouer un rôle majeur dans les processus hydrologiques (Kratzert et al., 2018). Les RNN sont donc peu adaptés aux longues séries temporelles habituellement utilisées en prévision hydrologique, ce qui réduit leur pertinence.

Un type spécial de RNN, les modèles d'apprentissage-machine avec mémoire long-court terme (Long Short-Term Memory ou LSTM), introduits par Hochreiter et Schmidhuber (1997), viennent pallier ce problème. Ils possèdent en effet l'habileté d'apprendre les dépendances à long terme entre les entrées et sorties du réseau, ce qui est crucial en hydrologie, notamment pour mieux représenter les effets de stockage comme les processus neigeux (Kratzert et al., 2018). L'architecture interne des LSTM diffère de celle des RNN traditionnels. Une cellule LSTM possède une mémoire supplémentaire pouvant contenir à long terme de l'information inchangée et trois portes qui contrôlent l'information passant par la cellule (Kratzert et al., 2018). Cette capacité de mémorisation rend les LSTM mieux adaptés aux prévisions à plus long terme dans lesquelles plusieurs pas de temps doivent être analysés. Cependant, la rétention étendue d'information requiert davantage de capacité informatique.

Plusieurs études démontrent que les LSTM semblent être une alternative viable aux modèles hydrologiques classiques, autant en situation de simulation que de prévision de débits. En appliquant pour la simulation journalière de débits un modèle LSTM de base et peu optimisé, Kratzert et al. (2018) ont obtenu des résultats similaires et même légèrement supérieurs à ceux du modèle hydrologique de référence SAC-SMA + Snow-17. Par après, Kratzert et al. (2019a) ont démontré l'aptitude de régionalisation pour la simulation hydrologique en entraînant un

modèle LSTM sur 531 bassins versants américains, surpassant significativement les résultats obtenus avec différents modèles hydrologiques. Cheng et al. (2020) ont confronté les modèles ANN aux LSTM dans un contexte de prévision long-terme de débits journaliers et mensuels pour deux bassins versants thaïlandais, révélant que des prévisions journalières précises sont obtenues jusqu'à 20 jours et que les LSTM performent mieux pour les prévisions journalières long-terme, le manque de données affectant toutefois ses prévisions mensuelles. Une autre étude comparant ces deux types de modèles d'IA pour la simulation hydrologique dans un bassin versant chinois a démontré qu'ils surclassent les modèles hydrologiques classiques, les LSTM étant plus performants et stables en raison de la présence d'une porte d'oubli dans leur architecture (Hu et al., 2018). Hunt et al. (2022) indiquent que les prévisions d'un modèle LSTM alimenté par des variables météorologiques et hydrologiques sont supérieures à celles brutes, mais aussi corrigées d'un modèle physique opérationnel. Cette supériorité est également constatée envers d'autres types de modèles comme l'arbre décisionnel XGBoost (Gauch et al., 2021). Récemment, Sabzipour et al. (2023) ont prouvé la supériorité des prévisions hydrologiques à court terme réalisées par un modèle LSTM en comparaison à celles d'un modèle utilisé opérationnellement, semi-distribué et jumelé à l'assimilation de données. Dans une configuration de régionalisation appliquée à des bassins non-jaugés, Arsenault et al. (2023) ont démontré qu'un modèle LSTM entraîné sur les données provenant de tous les bassins versants surpassé les modèles hydrologiques sur la plupart des bassins, et dépasse même les modèles hydrologiques calibrés sur les données cibles de bassins faussement non jaugés.

Des variations du modèle LSTM standard ont été explorées. Girihagama et al. (2022) ont montré qu'un modèle LSTM contenant un mécanisme d'attention offre des performances surpassant largement celles d'un modèle LSTM de base pour la prévision hydrologique à court terme. Le modèle LSTM équipé d'un algorithme encodeur-décodeur (LSTM E-D) testé par Herbert et al. (2021) en contexte de prévisions d'apports d'eau à long terme rivalise la précision des prévisions d'ensemble basées sur la climatologie.

D'autres applications hydrologiques ont également été étudiées, indiquant le potentiel multidisciplinaire et la pertinence des modèle d'IA dans des domaines autres que la stricte prévision hydrologique. Le modèle d'apprentissage profond proposé par Kao et al. (2021) parvient à projeter les précipitations régionales sur les caractéristiques d'inondation de la région et d'ainsi générer des prévisions précises de profondeurs d'inondation régionales. Zhang et al. (2018) ont prouvé que les LSTM sont un choix viable pour la gestion de réservoir, pouvant simuler précisément l'opération courante, notamment lors de conditions de faible hydraulité et de périodes de pointe.

Les LSTM ne sont toutefois pas exempts de failles et leur mise en place opérationnelle peut être fragile. Ils ont des limitations, nécessitant par exemple une quantité de données largement supérieure par rapport à des modèles hydrologiques simples pour atteindre une calibration adéquate (Bai et al., 2021). Aussi, leur nature de boite noire rend leur comportement plus imprévisible que les modèles classiques en empêchant l'accès à leurs états internes ainsi qu'aux relations qu'ils établissent entre les entrées et sorties du modèle. Malgré ces contraintes, ils parviennent à maintenir de bonnes performances en simulation, même lorsqu'ils n'apprennent pas correctement les relations physiques explicites entre les variables d'entrée et la variable cible (Yokoo et al., 2022).

Au final, l'offre de modèles hydrologiques est très variée, une diversité dans les approches de modélisation qui reflète les divers besoins des utilisateurs, l'incertitude quant à la meilleure façon de représenter la circulation et la distribution d'eau et d'énergie ainsi que les compromis en termes d'exigence en données, de degré de réalisme et de ressources informatiques nécessaires (Beven, 2012). L'approche retenue repose généralement sur la disponibilité des données, l'hétérogénéité du bassin versant et la pertinence dans le cadre de l'application voulue (Avila et al., 2023; Darbandsari et Coulibaly, 2020).

1.4 Prévision hydrologique

La prévision hydrologique vise à prédire des variables hydrologiques telles que le débit, le volume, le niveau ou la qualité de l'eau en un endroit donné et un moment précis. Elle est d'une importance capitale dans de nombreux secteurs d'activités comme la production d'hydroélectricité, la prédition et mitigation des inondations et la gestion de la ressource en eau (Dion et al., 2021). Une meilleure connaissance des débits d'eau à venir engendre une prise de décisions plus éclairée qui mitige les risques, soutient l'économie et fait bonne figure d'un point de vue social. Ce savoir accru peut profiter de manière significative à l'ensemble des acteurs sociaux, qu'il s'agisse des instances politiques, des gestionnaires d'infrastructures ou de la population générale. Les avancées technologiques des dernières décennies, conjuguées à l'émergence des prévisions probabilistes, ont favorisé une diversification marquée des approches de prévision hydrologique (Troin et al., 2021). Cet éventail de méthodes optimisées et mieux adaptées aux contraintes modernes permet de gérer plus efficacement la ressource en eau, mais de nombreux défis demeurent.

Les prévisions hydrologiques sont générées par des modèles hydrologiques, eux-mêmes alimentés par des prévisions météorologiques. Le développement de la technologie informatique, la numérisation des modèles et l'avènement de l'intelligence artificielle ont propulsé la prévision hydrologique vers de nouveaux sommets. En effet, ces avancées permettent d'améliorer continuellement les modèles existants et de développer de nouveaux modèles puissants, augmentant du même coup la qualité des prévisions. Entre autres, des modèles basés sur l'IA sont en pleine croissance et profitent d'immenses quantités de données pour apprendre et reproduire fidèlement les processus hydrologiques.

1.4.1 Prévisions déterministes et ensemblistes

Tout comme les prévisions météorologiques, les prévisions hydrologiques se trouvent sous la forme déterministe ou ensembliste.

Une prévision hydrologique déterministe produit une seule valeur de débit. L'extrant d'une telle prévision est unique, car la prévision météorologique générée puis passée dans le modèle hydrologique ainsi que les conditions initiales de ce dernier sont uniques. Cette approche se démarque par la clarté et le caractère direct de la réponse fournie, facilitant ainsi son interprétation et son intégration dans les processus décisionnels. Toutefois, malgré le raffinement des méthodes de prévision et des équipements de collecte et de traitement des données hydrométéorologiques, la fiabilité des données est loin d'être parfaite et d'autres sources d'incertitude existent, nuisant à la confiance vouée à l'approche déterministe. De plus, elle ne transmet pas d'information sur l'incertitude alors que chaque étape du processus prévisionnel introduit de l'incertitude pouvant dégrader la qualité et l'application opérationnelle des prévisions (Bogner & Kalas, 2008; Demirel et al., 2013; Pappenberger et al., 2005). En effet, il est impossible de décrire parfaitement toutes les sources d'incertitude associées à la prévision hydrologique, ce qui écarte la possibilité d'obtenir une estimation parfaite de l'incertitude totale de la prévision (Boucher et al., 2011). Plus concrètement, une prévision hydrologique déterministe erronée est susceptible d'engendrer d'importantes pertes auprès de plusieurs sphères, dont celles économiques et sociales (Krzysztofowicz, 2001). D'autres éléments tels que la variabilité climatique et l'imperfection des modèles hydrologiques utilisés ont appuyé la transition vers la prévision dite d'ensemble. En dépit de ces contre-indications, de nombreux prévisionnistes utilisent toujours des prévisions déterministes pour leurs besoins opérationnels (Boucher et al., 2012).

Une prévision hydrologique probabiliste, ou d'ensemble, génère plusieurs valeurs de débits en fonction de différents facteurs engendrant des variations dans les prévisions émises. En générant un ensemble, l'approche probabiliste illustre le spectre des résultats possibles, quantifiant du même coup l'incertitude associée aux prévisions (Troin et al., 2021). Les prévisions d'ensemble peuvent être obtenues en faisant varier les états initiaux du modèle hydrologique, en alimentant le modèle hydrologique de prévisions météorologiques probabilistes, ou bien en utilisant plus d'un modèle hydrologique, une approche dite multi-modèle. Pour une variété d'applications hydrologiques, les prévisions probabilistes ou basées sur des ensembles ont le potentiel de générer des bénéfices et d'optimiser les décisions en

comparaison à ce qui serait obtenu via des prévisions déterministes (Ramos et al, 2013). Boucher et al. (2011) ont démontré que les prévisions hydrologiques d'ensemble sont supérieures à leur contrepartie déterministe basées sur un modèle météorologique plus avancé, surtout lorsqu'un post-traitement minimal est appliqué. Donnant suite à cette étude et dans un contexte opérationnel de production hydroélectrique, Boucher et al. (2012) ont montré que l'utilisation de l'approche ensembliste réduit les déversements et améliore globalement la production d'électricité. Roulin (2007), s'appuyant sur des prévisions de précipitations ensemblistes du ECMWF, a aussi montré que les prévisions probabilistes permettaient de prendre des décisions à plus grande valeur économique pour se protéger d'inondations. Malgré ces avantages indéniables, de nombreux défis persistent et limitent l'utilisation opérationnelle de prévisions d'ensemble tels que la communication et la compréhension auprès des utilisateurs (Ramos et al., 2010), la complexité de l'information transmise et la forte demande en ressources informatiques (Troin et al., 2021). Toutefois, la prise en compte de multiples scénarios envisageables et l'incertitude prévisionnelle omniprésente, deux éléments amplifiés par les changements climatiques, justifient l'utilisation de prévisions d'ensemble.

La qualité des prévisions hydrologiques, qu'elles soient déterministes ou probabilistes, peut être évaluée de diverses façons permettant de statuer sur leur fiabilité et leur précision. Par contre, certaines mesures sont strictement applicables aux prévisions déterministes alors que d'autres sont adaptées à l'approche probabiliste. Les méthodes d'évaluation se basent la plupart du temps sur la comparaison entre les débits simulés ou prédits et les débits observés, notamment sous la forme d'analyse des écarts. Ces approches se déclinent en plusieurs catégories, notamment des mesures statistiques quantitatives, mais aussi des représentations graphiques.

1.4.2 Sources d'incertitude en prévision hydrologique

La chaîne prévisionnelle comporte plusieurs éléments qui, combinés, intègrent un haut degré d'incertitude dans les prévisions hydrologiques. L'incertitude prévisionnelle sur laquelle il est possible d'agir vient de trois sources : les conditions initiales, les prévisions météorologiques

et les modèles hydrologiques (Troin et al., 2021). Plusieurs outils, tels que la perturbation des conditions initiales, la prévision météorologique d'ensemble et la combinaison de plusieurs modèles ont été développés afin de s'y attaquer, démontrant leur capacité à capturer une certaine part de l'incertitude totale et améliorer globalement les prévisions (Thiboult et al., 2016). Ces méthodes peuvent aussi être combinées pour former un ensemble plus large qui a le potentiel de mieux évaluer l'incertitude des prévisions en améliorant la distribution et l'uniformité des prévisions (Dion et al. 2021).

L'assimilation de données, qui intègre des observations récentes afin de mieux estimer l'état initial du modèle, est une approche préconisée pour minimiser le problème des conditions initiales. La méthode du filtre de Kalman (KF) et ses dérivées, notamment le filtre de Kalman d'ensemble (EnKF) sont largement utilisées et ont le potentiel d'améliorer les prévisions hydrologiques, comme les résultats de Sabzipour et al. (2023) le démontrent.

La nature chaotique de l'atmosphère rend les prévisions météorologiques hautement incertaines, surtout à long terme. Ainsi, l'utilisation de prévisions météorologiques d'ensemble, qui génère une fourchette de possibilités, permet de mieux représenter l'incertitude climatologique.

La qualité des données de calibration et l'exactitude du modèle caractérisent l'incertitude liée aux modèles hydrologiques. D'une part, l'imprécision des données météorologiques et hydrologiques utilisées lors de la calibration du modèle peut se transmettre à ce dernier. Les instruments de mesure comportant une large marge d'erreur, il est essentiel de valider les données historiques récoltées instrumentalement. Les paramètres trouvés en calibration tout comme les états initiaux du modèle peuvent être imprécis et s'ajoutent à la simplification faite du bassin versant par le modèle. Tout compte fait, la combinaison de plusieurs modèles hydrologiques permet d'obtenir un ensemble flexible et mieux dispersé, à condition que ces modèles soient différents intrinsèquement (Dion et al., 2021).

1.4.3 Prévision de volumes d'eau

La prévision des volumes vise des objectifs similaires à celle des débits. Elle cherche à estimer les apports totaux au cours d'une certaine période, une information utile pour faire de la gestion opérationnelle sur un horizon temporel étendu. En même temps, compte tenu de la forte demande d'électricité engendrée par la société moderne, la planification de la production d'hydroélectricité à court terme est importante, requérant des prévisions court terme de volumes précises pour en tirer profit (Stokelj et al., 2002). La connaissance de la quantité d'eau issue d'un système hydrologique est également pertinente pour faire la conception et le dimensionnement de réservoirs, tout comme pour l'établissement des pratiques optimales pour les opérer (Papamichail et Georgiou, 2001). L'objectif de la prévision volumétrique étant d'estimer un volume d'eau total, l'importance de variables dépendantes du temps telles que la distribution et l'amplitude des débits est moindre. La méthode usuelle consiste à faire le calcul des débits sur la période désirée, qui sont ensuite agrégés. Cette façon de faire permet d'utiliser les modèles hydrologiques traditionnels, qui ont été construits spécifiquement pour faire de la prévision de débits. Il est également possible d'agréger préalablement à la modélisation les débits observés en volumes, comme le fait Bai et al. (2015, 2016). Ces volumes deviennent alors les nouvelles valeurs cibles que le modèle cherche à reproduire. Les modèles statistiques sont compatibles avec cette approche, qui n'est toutefois pas adaptée aux modèles hydrologiques traditionnels voués à modéliser des débits. De nombreuses études de prévision volumétrique s'attardent aux périodes critiques d'un point de vue opérationnel, comme la crue printanière dans des régions fortement influencées par la neige (Arsenault et al., 2016; Herbert et al., 2021). Malgré que la littérature au sujet de la prévision volumétrique soit davantage exhaustive que pour la prévision de débits, de nombreux horizons temporels ont été explorés, notamment le journalier (Bai et al., 2016; Sattari et al., 2012), le mensuel (Hashino et al., 2007; Papamichail et Georgiou, 2001) et le saisonnier (Herbert et al., 2021). L'utilisation de prévisions saisonnières permet d'optimiser l'atteinte des objectifs opérationnels, qu'il s'agisse de gestion des niveaux d'eau, de production énergétique ou de préservation environnementale (Ashbolt et Perera, 2018; Peñuela et al., 2020).

1.5 Traitement statistique

Les modèles de prévision possèdent des biais provenant des données en entrée, de leurs paramètres ou des hypothèses simplificatrices utilisées dans le modèle (Hashino et al., 2007). Ces biais sont propagés dans la chaîne prévisionnelle et dégradent la précision et la fiabilité des prévisions. Afin d'améliorer leur qualité et mieux estimer l'incertitude, il est recommandé d'appliquer des méthodes de traitement statistique (Schaake et al., 2010) qui visent à retirer les biais et améliorer la dispersion dans les ensembles de prévisions (Zalachori et al., 2012). Les stratégies de correction peuvent être appliquées avant la modélisation hydrologique (pré-traitement ou *pre-processing*) ou après la modélisation hydrologique (post-traitement ou *post-processing*), deux approches qui peuvent aussi être combinées. Les techniques de pré-traitement, appliquées aux prévisions météorologiques, traitent seulement la partie des sources d'incertitude de nature météorologique. Ainsi, elles retirent uniquement une certaine part des biais et erreurs retrouvés dans les prévisions hydrologiques. Le post-traitement est alors nécessaire pour obtenir des prévisions hydrologiques d'ensemble non-biaisées et dont la distribution est adéquate (Madadgar et al., 2014). Pour cette raison, le post-traitement, appliqué à la variable hydrologique prédictive, est souvent la stratégie de correction favorisée. Les résultats d'Hashino et al. (2007), qui ont effectué du post-traitement de prévisions de volumes d'eau saisonniers, montrent que cette correction statistique améliore significativement la qualité des prévisions en éliminant les biais inconditionnels. Le mécanisme de post-traitement de Zhao et al. (2011), entraîné sur des données de la période de calibration, réduit efficacement le biais moyen et rapproche les distributions de probabilité des simulations post-traitées de celles des débits observés. L'étude de Kang et al. (2010) sur les prévisions hydrologiques d'ensemble indique que le post-traitement réduit l'incertitude de manière plus efficace que le pré-traitement, mais que dans la majorité des scénarios étudiés, c'est la combinaison des deux approches qui s'avère la plus avantageuse. La littérature portant sur les bénéfices apportés par la correction de biais, une méthode de post-traitement assez répandue, est assez faste en matière de prévisions hydrologiques saisonnières (Peñuela et al., 2020).

1.6 Prévision des apports volumétriques en eau

La prévision hydrologique s'étend sur plusieurs horizons temporels, de court terme (quelques heures) à long terme (quelques mois) en passant par le moyen terme qui vient combler l'écart entre ces deux périodes, durées qui dépendent toutefois des buts visés par les prévisions hydrologiques. La prévision saisonnière est une déclinaison de la prévision à long terme dont la durée, généralement de trois ou quatre mois, s'établit en fonction de la durée des saisons hydrologiques.

Les prévisions à court terme sont utiles pour les réservoirs de moindre capacité gérés pour des opérations quotidiennes telles que la protection face aux inondations (Saavedra Valeriano et al., 2010) et autres événements extrêmes. Elles contribuent également à optimiser l'allocation des ressources en eau à plus petite échelle, par exemple au niveau des turbines de centrales hydroélectriques. La qualité des prévisions à court terme est principalement affectée par les conditions initiales des modèles (Troin et al., 2021), source d'incertitude pouvant être tempérée à l'aide de méthodes d'assimilation de données, qui approchent les états initiaux des modèles de la réalité.

Les prévisions à long terme sont davantage pertinentes aux grands réservoirs dont les objectifs opérationnels s'inscrivent dans la durée, comme la génération d'hydroélectricité et l'approvisionnement en eau (Maurer and Lettenmaier, 2004; Sankarasubramanian et al., 2009). Elles permettent de mieux planifier la production d'électricité ainsi que la maintenance des infrastructures et d'optimiser les niveaux des réservoirs sur des périodes prolongées. Cette meilleure planification génère des bénéfices économiques, notamment par la réduction des déversements et la flexibilité accordée aux planificateurs pour synchroniser les périodes de production avec celles de forte demande énergétique. La qualité des prévisions à long terme est limitée par les lacunes climatiques dans la structure et la paramétrisation des modèles ainsi que par le caractère intrinsèquement chaotique du climat. En effet, l'influence des conditions initiales sur les prévisions se dissipe à long terme, faisant place à la variabilité climatique (Troin et al., 2021).

Les prévisions saisonnières supportent la gestion des réservoirs en estimant l'apport en eau envisagé en fonction de périodes d'une durée spécifique, notamment les saisons hydrologiques particulières à un bassin versant. Ces prévisions permettent d'élaborer des plans d'exploitation à court terme pour atteindre les objectifs visés comme maximiser la sécurité de l'approvisionnement en eau, minimiser la consommation d'énergie, réduire les coûts opérationnels ou fournir les débits environnementaux (Ashbolt et al., 2018). Dans les climats nordiques, où la crue printanière constitue la principale source d'apport en eau, les prévisions saisonnières peuvent aider à établir une stratégie de gestion annuelle (Bazile et al., 2017). Elles permettent aussi de mieux répondre à des problèmes concrets de gestion de l'eau, comme la détermination des volumes à libérer de réservoirs autant hydroélectriques que responsables de l'approvisionnement en eau ou aux objectifs intersectoriels (Turner et al., 2017). Appliquées à la production hydroélectrique, ces prévisions permettent d'optimiser les décisions et opérations quotidiennes en fonction de la stratégie de gestion à long terme alignée sur les tendances anticipées. Par exemple, si une importante crue printanière est attendue et que le niveau actuel du réservoir est élevé, les opérations seront planifiées en vue d'éviter d'éventuels déversements. Le potentiel des prévisions saisonnières pour l'optimisation de la planification opérationnelle de réservoirs a été reconnu dans plusieurs études. Les résultats d'Ashbolt et al. (2018) indiquent que les règles d'opération élaborées selon les prévisions saisonnières performent mieux que celles optimisées à partir de données historiques pour maximiser le niveau du réservoir et minimiser les coûts opérationnels ainsi que les déversements. Sankarasubramanian et al. (2009) ont démontré que la simulation de l'allocation intra saisonnière d'eau à l'aide de prévisions hydrologiques issues de prévisions mensuelles de précipitations surpassé celle basée sur l'historique en réduisant les déversements, en augmentant la production hydroélectrique et en respectant davantage les niveaux cibles des réservoirs. Peñuela et al. (2020) ont perçu des avantages opérationnels en intégrant des prévisions saisonnières dans un système d'optimisation en temps réel pour améliorer la gestion de réservoirs, notamment en termes d'efficacité des opérations. Les prévisions saisonnières n'ont pas que des avantages, alors que Pechlivanidis et al. (2020) ont montré que la capacité prévisionnelle diminue avec le temps, surtout en hiver, et que la qualité des prévisions de

volumes saisonniers varie notamment selon la saison et le bassin. Ainsi, malgré le long historique de recherche et l'utilisation opérationnelle bien répandue des prévisions hydrologiques saisonnières, de nombreuses opportunités subsistent pour améliorer leur performance et leur valeur (Mendoza et al., 2017).

1.7 Objectifs de la recherche

Le principal objectif de cette étude est d'améliorer la prévision des volumes d'eau reçus sur une période de 14 jours vers un réservoir hydroélectrique alimenté par le bassin versant du Lac-Saint-Jean. Pour cela, l'étude s'appuie sur un modèle d'apprentissage profond aux capacités prometteuses pour la prévision hydrologique, le modèle de longue mémoire à court terme (LSTM), ainsi que sur les rétro-prévisions météorologiques. Ce produit météorologique offre une source de données non seulement fiables, mais aussi volumineuses, ce qui offre aux modèles une quantité plus importante de données pour leur entraînement que des données de prévision régulières. L'étude vise à vérifier si les modèles LSTM peuvent être utilisés pour prévoir directement les volumes, sans passer au préalable par des prévisions de débits. Elle évalue aussi l'apport des données de rétro-prévisions météorologiques dans l'entraînement des modèles, afin de déterminer si elles améliorent la performance des modèles et la qualité des prévisions volumétriques. En entraînant chaque modèle différemment, l'influence spécifique des rétro-prévisions est isolée, renforçant ainsi la fiabilité des conclusions. Compte tenu des importantes variations saisonnières de la région d'étude, cette recherche prend également en compte l'impact des différentes saisons et procède à une évaluation complète par le biais de plusieurs critères de performance qui évaluent les caractéristiques essentielles des résultats.

CHAPITRE 2

ASSESSING THE VALUE OF METEOROLOGICAL REFORECAST DATA TO PREDICT INFLOW VOLUMES OVER A CANADIAN SNOW-DOMINATED CATCHMENT USING A DEEP LEARNING MODEL

Léo Soucy¹, Richard Arsenault¹, Jean-Luc Martel¹

¹ Hydrology, Climate and Climate Change Laboratory, École de technologie supérieure, Université du Québec, 1100 Notre-Dame West, Montréal, Quebec, Canada, H3C 1K3

Article soumis pour publication à la revue « Journal of Hydrology » le 25 février 2025

2.1 Abstract

Accurate inflow volume forecasts are critical for effective hydropower reservoir management. This study presents a novel framework that leverages deep learning models and atmospheric reforecasts to predict 14-day cumulative inflow volumes in a Canadian catchment. Specifically, it evaluates whether Long Short-Term Memory (LSTM) models can directly forecast inflow volumes without intermediate daily streamflow predictions and examines the impact of training with meteorological reforecast data on model performance and forecast quality.

Three LSTM models were trained using various combinations of meteorological data inputs from the European Centre for Medium-Range Weather Forecasts (ECMWF), including ERA5 reanalysis data and probabilistic reforecast datasets. The first model was trained exclusively on ERA5 data, the second used reforecast data in its training, and the third was a hybrid of the two first models, integrating the training process of both models. The models were then applied to ECMWF ensemble meteorological forecasts and assessed using quantitative reliability and accuracy metrics.

Results indicate that the three LSTM models successfully predict 14-day cumulative inflow volumes, demonstrating strong performance metrics. However, no single model consistently

outperforms the others. The model trained solely on ERA5 data exhibits greater variability, resulting in lower accuracy but higher reliability. Performance also varies seasonally, with reforecasts-trained models improving accuracy in winter and spring, but underperforming in summer and fall.

These findings suggest that incorporating meteorological reforecast data in LSTM model training can enhance forecast accuracy in specific seasons but may compromise forecast robustness. The choice to integrate reforecasts should therefore depend on the desired trade-off between accuracy and reliability in an operational forecasting context.

2.2 Introduction

Hydrological forecasting is essential across multiple sectors, including hydropower generation, flood prediction and mitigation, and water resource management (Dion et al., 2021). A better understanding of future water flows enables more informed decision-making, reduces risks to infrastructures, supports economic stability, and provides social benefits.

While hydrological forecasts often focus on streamflow, inflow volume is another valuable variable that can be predicted. Inflow forecasting estimates the total water volume received over a specific period, a critical measure for effective water allocation across diverse sectors such as drinking water supply, agriculture, and hydropower (Apaydin et al., 2020). However, most studies approach inflow forecasting indirectly by predicting streamflow rates and subsequently aggregating them, rather than directly forecasting inflow volumes (Arsenault et al., 2016; Herbert et al., 2021).

Accurate reservoir inflow forecasts are crucial as they significantly aid in mitigating flood and drought impacts, optimizing irrigation scheduling, and enhancing the efficiency of hydropower operations (Bai et al., 2015, Ionno et al. 2024). Reliable inflow volume forecasts are essential for successful operation and management of reservoirs (Bai et al., 2016). From an operational standpoint, they maximize hydropower generation benefits and minimize risks associated with

spillway operations (Ahmed et al., 2015). Short-term forecasts are especially useful for smaller reservoirs designed for short-term operations, such as flood protection, while long-term forecasts are more beneficial for large reservoirs with long-term objectives, including hydropower generation and water supply (Anghileri et al., 2016).

Deep learning (DL) models, which derive their predictive power from large datasets, are increasingly being used as alternatives to traditional hydrological models for hydrological forecasting tasks (Kratzert et al., 2018; 2019a; 2019b; Damavandi et al., 2019). These models employ complex and interconnected artificial neural networks (ANNs), which can efficiently learn both linear and non-linear relationships between input features and target variables (Sit et al., 2020).

ANNs come in various architectures, including recurrent neural networks (RNNs), first developed by Rumelhart et al. (1986), which are specifically designed to capture temporal dependencies by processing sequential inputs (Sit et al., 2020; Kratzert et al., 2018). A more advanced type of RNN, the Long Short-Term Memory (LSTM) network, introduced by Hochreiter and Schmidhuber (1997), can learn long-term dependencies between network inputs and outputs. This capability is particularly relevant in hydrology, as it improves the representation of storage effects such as snow accumulation and melt processes (Kratzert et al., 2018). Unlike RNNs, which often struggle with short-term memory limitations, LSTMs can effectively capture long-term dependencies, making them particularly well-suited for hydrological modeling and forecasting applications (Xu et al., 2020).

Numerous studies have established LSTMs as a viable alternative to traditional hydrological models across various forecasting applications. Hunt et al. (2022) reported that LSTM forecasts driven by meteorological and hydrological inputs outperformed not only raw forecasts, but also bias-corrected forecasts produced by an operational physical model. Recently, Sabzipour et al. (2023) demonstrated the short-term hydrological forecasts generated by an LSTM model outperformed those from an operational semi-distributed model with data assimilation. Cheng et al. (2020) evaluated LSTM models for long-term daily and monthly

streamflow forecasting in two basins in Thailand, finding that accurate daily forecasts could extend up to 20 days. However, they also noted that data scarcity affected the accuracy of monthly forecasts.

In a regionalization context for ungauged basins, Arsenault et al. (2023) showed that an LSTM model trained on a comprehensive dataset outperformed traditional hydrological models across most basins and even exceeded the performance of models calibrated on target data for pseudo-ungauged basins. Kratzert et al. (2018; 2019a, 2019b) demonstrated that regional LSTM models effectively transfer global hydrological knowledge to individual catchments, consistently outperforming a range of benchmark hydrological models.

To remain relevant for future studies, hydrological models must also adapt to changing climatic conditions. Bai et al. (2021) compared an LSTM network with two conceptual hydrological models for runoff prediction under evolving climate scenarios, revealing that LSTMs are more robust when sufficient calibration data is available. Similarly, Martel et al. (2024a) suggested that regional LSTMs are better suited than traditional hydrological models for climate change impact assessments, positioning them as promising alternatives.

In a related study on flood frequency analysis, Martel et al. (2024b) showed that LSTM-based models simulate peak streamflow with equal or greater accuracy compared to a distributed hydrological model across various catchments in Quebec, Canada. Zhang et al. (2022) further highlighted the potential of LSTMs for flood peak prediction in ungauged or data-sparse catchments, where they outperformed a conceptual hydrological model, particularly for catchments that are underrepresented in the dataset.

However, a key limitation of LSTM models is their high data requirement for effective learning. When training data is limited, their performance can decline significantly, occasionally falling below that of models they typically outperform (Ayzel & Heistermann, 2021; Bai et al., 2021; Gauch et al., 2021). When observed data is insufficient to properly train LSTM-based models, or to further improve LSTM models performance, alternative datasets

can be leveraged to increase the amount of available training data. Various methodologies are employed to increase the amount of training data needed to properly learn hydrological processes. For instance, Arsenault et al. (2023), Kratzert et al. (2019a), and Kratzert et al. (2019b) adopted a regionalization approach, where data is collected from multiple catchments, while others use reanalysis datasets to compensate for short or incomplete meteorological records (Sabzipour et al., 2023; Hunt et al., 2022). Similar challenges arise in the statistical post-processing of meteorological forecasts, where short observational records can hinder the development of robust statistical correction models. In such cases, meteorological reforecasts can be used to analyze local biases and errors (anomalies) in the weather forecasts (Abaza et al., 2017).

Reforecasts are retrospective meteorological predictions generated for past periods with a fixed numerical model (Verkade et al., 2013). These datasets provide a large volume of historical forecast data, as they typically generate predictions for the same calendar day as the operational forecast, but across multiple years. While reforecasts are commonly used for post-processing, they could also be directly integrated into LSTM training, enabling the model to learn and adjust for biases in the meteorological model forecasts. Despite their potential, few studies have explored the use of reforecasts in hydrological forecasting, with the large majority of them relying on traditional hydrological models. Among the limited research in this area, Verkade et al. (2013) and Jorg-Hess et al. (2015) reported modest improvements in hydrological model simulation and forecast quality following the statistical correction of forecasts using reforecasts.

This study aims to evaluate the potential of LSTM models in simulating and predicting 14-day cumulative inflow volumes over the Lac-Saint-Jean (LSJ) catchment in Quebec, Canada. Given the region's substantial annual snowfall, it provides an ideal environment for testing LSTM forecasts across a wide range of hydrometeorological conditions and assessing the reliability of these models in an operational context.

Nordic regions such as LSJ are strongly influenced by snowpack dynamics, making the long-term hydrological processes more complex and challenging to capture compared to regions with more uniform weather. Specifically, this study seeks to determine whether LSTM models can directly forecast inflow volumes without relying on intermediate daily streamflow predictions and whether training with meteorological reforecasts can enhance model performance and forecast quality.

2.3 Methods and data

The following sections outline the methodology employed in this study. Section 2.3.1 introduces the study region and its hydroclimatic characteristics, while Section 2.3.2 details the hydrometeorological datasets utilized. Section 2.3.3 explores hydrological forecasting using LSTM models, providing an in-depth description of each model developed. Section 2.3.4 explains the model construction and parameterization processes. Finally, Section 2.3.5 discusses the application and evaluation of the models in forecasting scenarios.

2.3.1 Study area

This study focuses on the LSJ catchment in Quebec, Canada, a vast region spanning 45,000 km². The catchment is used by Rio Tinto, a major international mining and metals company, to generate hydropower for aluminum production. The average daily streamflow is 873 m³/s, with annual peak flows averaging 4,180 m³/s but occasionally reaching up to 9,000 m³/s. The LSJ reservoir spans 1,000 km² and plays a crucial role in ensuring water availability for downstream hydropower stations (Arsenault & Côté, 2019). The catchment's location is presented in Figure 2.1.

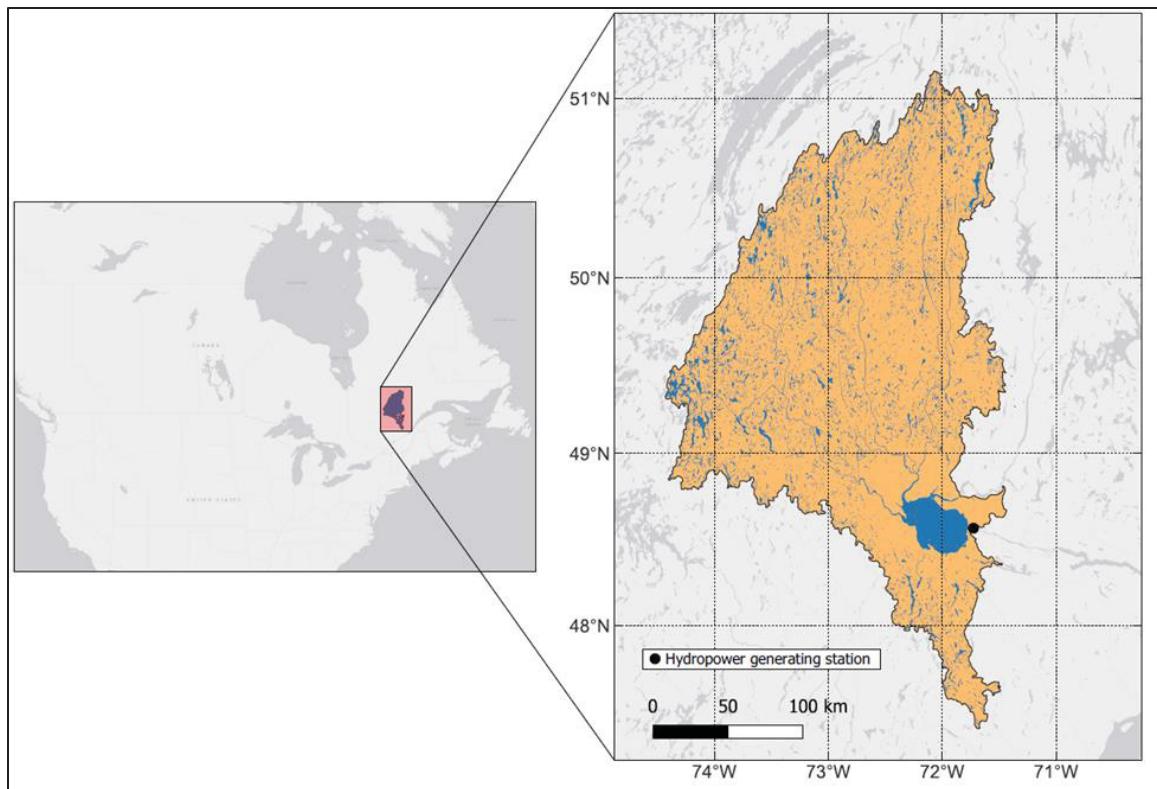


Figure 2.1 Study area and location of the downstream hydropower generating station.

The large lake (Lac-Saint-Jean) is the 1,000 km² reservoir managed by Rio Tinto for hydropower generation

The region experiences a continental climate characterized by cold winters and substantial snowfall from November to April. These significant snow accumulations strongly impact the hydrological regime, particularly during the spring freshet (Arsenault & Côté, 2019).

Accurate inflow volume forecasts are essential for optimizing water resource management and improving hydropower operations planning. Such forecasts are particularly important for ensuring safe and efficient hydroelectric plant operations while mitigating risks associated with extreme hydrological events, such as the spring freshet. These events can threaten the integrity of hydropower plants, as well as downstream infrastructures and communities.

2.3.2 Datasets

In this study, multiple datasets were used, including hydrometric data and observed, forecast, and reforecast meteorological data. This section provides an overview of all datasets, and the pre-processing methodology adopted.

Two meteorological datasets were used to train the LSTM models. The first dataset consists of pseudo-observed meteorological data, also known as reanalysis data, obtained from the European Center for Medium-Range Weather Forecasts (ECMWF) fifth generation reanalysis (ERA5; Hersbach et al., 2020). This dataset was used both as historical observations for training and in forecasting mode. ERA5 provides global coverage at a spatial resolution of $0.25^\circ \times 0.25^\circ$, with data available at an hourly time step. For this study, the hourly data was aggregated to a daily scale.

To ensure relevance to the study area, the raw ERA5 data was filtered to keep only the points located within the catchment and then spatially averaged across the entire catchment, resulting in a single vector for each meteorological variable over the study area. The ERA5 data spans the period from 1978 to 2023, resulting in a total 16,801 days of data available for the study. The meteorological variables used in this study are presented in Table 2.1.

Tableau 2.1 Meteorological variables used in this study

Meteorological variable	ECMWF short name	Units
Dew point temperature 2 meters above ground	d2m	K
Minimum temperature 2 meters above ground	mn2t6	K
Maximum temperature 2 meters above ground	mx2t6	K
Snow depth	sd	m of water equivalent
Surface pressure	sp	Pa
Total precipitation	tp	m
Wind U-component 10 meters above ground	u10	$m s^{-1}$
Wind V-component 10 meters above ground	v10	$m s^{-1}$

The second meteorological dataset used to train the LSTM models consists of retrospectively forecasted meteorological data, also known as reforecast data. This dataset was collected from the ECMWF Integrated Forecasting System (IFS) medium range forecasts. The 10-member ensemble weather reforecast data is issued every Monday and Thursday, with a lead-time of 15 days. Each reforecast includes hindcasts for the previous 20 years on the same calendar day as the reforecast issue date. Thus, for each reforecast issue date, a total of 200 hindcasts are generated (10 ensemble members for each of the previous 20 years).

The weather variables are available at a spatial resolution of approximately 18 km with a 6-hour time step. To align with the study's temporal resolution, the four 6-hour timesteps were aggregated to a daily scale. However, due to a time zone offset, the final day of each 15-day forecast was incomplete for the study location, resulting in the use of a 14-day ensemble

weather reforecast instead. Similarly to the ERA5 dataset, the ensemble reforecasts were spatially aggregated over the entire catchment.

The reforecasts span from March 2016 to April 2023, amounting a total of 745 reforecast issue dates. Given the 20-year lookback window, reforecasts are available from March 1996 (20 years prior to the first forecast date) to April 2022 (one year prior to the last forecast date). Combining the 745 reforecast issue dates with the 20-year lookback period and the 10-member ensembles, the reforecast dataset amounts to a total of 149,000 14-day meteorological forecasts. The meteorological variables included in this dataset are the same as those in the ERA5 dataset, as listed in Table 2.1.

The next category of meteorological data consists of actual operational forecast data, which was required to evaluate the performance of the trained models on independent data and in an operational context. This data comprises the 50-member operational ensemble weather forecast data from the ECMWF's IFS. These medium-range forecasts are issued daily with a 15-day lead-time. The weather variables are available at a spatial resolution of 18 km with a 6-hour time step. The same spatial and temporal pre-processing steps applied to the reforecast data were also applied to the operational forecast data. The forecasts span the period between March 9, 2016, to May 31, 2023, resulting in a total of 2,639 ensemble forecasts. Given that each forecast consists of 50 ensemble members, the dataset contains a cumulative total of 131,950 14-day forecasts. The meteorological variables included in this dataset are the same as those from the ERA5 and reforecast datasets, as listed in Table 2.1.

The observed hydrometric data, consisting of daily reservoir inflows, were provided by Rio Tinto, the operator of the hydropower generating station and owner of water rights for the LSJ system. The dataset spans from January 1950 to March 2024. Given the diverse and widespread sources inflow across the territory, inflow values were derived using mass balance calculations that account for reservoir level, turbine outflows, and flood spillway releases. While measurements are taken at key reservoir entry points, mass conservation equations offer more reliable estimates of inflow values. However, mass-balance derived inflows can be noisy,

therefore a five-day rolling average was computed and used to produce a more consistent and stable water volume at the catchment outlet, following the methodology of Davidson-Chaput et al. (2025). The daily inflows to the reservoir were then aggregated to compute 14-day cumulative inflow volumes, which serve as the target variable in the forecasting procedure in this study.

2.3.3 Hydrological modelling using LSTM models

Given their recent yet widespread and promising application in hydrological modelling and forecasting, this study explores the use of LSTM models for inflow volumes forecasting. However, due to their complex nature, lacking a physical basis and involving numerous trainable parameters (i.e., weights and biases), the training procedure must be carefully designed. Unlike traditional hydrological models, which typically divide data into two sets (calibration and validation), LSTM models require three distinct datasets: training, validation, and testing.

The training and validation phases occur simultaneously: the training dataset is used to optimize the model weights, while the validation dataset evaluates the model's performance and determines when to stop training based on predefined stopping criteria. The testing phase follows training and assesses the model's robustness on an independent dataset before operational forecasting. In this study, 70 % of the data was allocated for training, 15 % for validation, and the remaining 15 % for testing. Each phase contains independent data, i.e. not shared with any of the two other phases to increase model robustness when exposed to unseen data.

The LSTM models developed in this study were trained and tested on sequences of 365 days of meteorological data. Each sequence is composed of 351 days of observations (prior to the forecast date) and 14 days of forecasted meteorological data (starting from the forecast date). Therefore, the input data to the LSTM models consist of 365 consecutive days of

meteorological data, while the 14-day cumulative inflow volumes corresponding to the forecasted meteorological data period serve as the target variable.

In this study, three LSTM models were trained, each using a different combination of datasets and processes to assess the impact of using reforecasts during training. The models are as follows:

- 1) E5 model: Trained exclusively with ERA5 data,
- 2) RF model: Trained with ERA5 data for the 351-day lookback period and reforecasts for the 14-day forecast period, and
- 3) E5-RF model: A hybrid approach where the model is pre-trained using the E5 model data and then fine-tuned by retraining with the RF model data.

These three modeling approaches are detailed in the following subsections and are presented in Figure 2.2.

a) Training example Target date : 2005-12-18		2005-01-01	2005-01-02	...	2005-12-16	2005-12-17	2005-12-18	2005-12-19	2005-12-20	2005-12-21	2005-12-22	2005-12-23	2005-12-24	2005-12-25	2005-12-26	2005-12-27	2005-12-28	2005-12-29	2005-12-30	2005-12-31
Training dataset		E5 model																		
		Input : ERA5 series	351-day (ERA5)		14-day (ERA5)															
		RF model																		
		Input : Reforecast series	351-day (ERA5)		14-day (Reforecasts)															
		All models																		
		Target (Output) : Observed volume											14-day							

b) Forecasting example Forecast issue date : 2017-12-18		2017-01-01	2017-01-02	...	2017-12-16	2017-12-17	2017-12-18	2017-12-19	2017-12-20	2017-12-21	2017-12-22	2017-12-23	2017-12-24	2017-12-25	2017-12-26	2017-12-27	2017-12-28	2017-12-29	2017-12-30	2017-12-31
Forecasting dataset		All models																		
		Input : Forecast series	351-day (ERA5)		14-day (Forecasts)															
		Target (Output) : Observed volume											14-day							

 Observed meteorological data from ERA5 (8 meteorological variables)

 Forecasted meteorological data from the 10-member ensemble reforecasts (8 meteorological variables)

 Forecasted meteorological data from the 50-member ensemble forecasts (8 meteorological variables)

 Target 14-day cumulative volume data (aggregation of the daily reservoir inflows provided by Rio Tinto)

Figure 2.2 Example implementation of a training sample for the E5 and RF forecasting models (a) and an example implementation of a forecasting sample for all models (b). Panel (a) presents a single sample. However, this process is repeated across the entire dataset available for the training period and similarly applied to the validation and testing periods. Similarly, panel (b) shows an example for a single forecast issue date, but the same process is repeated for all forecast issue dates

2.3.3.1 ERA5-based (E5)

The E5 model is based solely on ERA5 input data. For each forecast date, the input sequence consists of 365 consecutive days from the ERA5 dataset, representing the 351 days preceding the forecast issue date and the 14 days starting from the forecast issue date. While the last 14 days are not forecast data, they function as such in the training context, as they simulate forecast conditions during model training. However, in the operational evaluation phase, these 14 days will be replaced with actual forecast data. The purpose of using this model is to isolate the contribution of the ERA5 dataset and examine its predictive performance without

incorporating reforecast data. As such, the E5 model serves as the reference model against which the reforecast-based models (RF and E5-RF) will be compared.

To maximize data availability, the entire ERA5 period from 1978 to 2023 was used. The dataset was divided as follows:

- Training: December 18, 1978, to December 17, 2009 (31 years),
- Validation: December 18, 2009, to December 17, 2016 (8 years), and
- Testing: December 18, 2016, to December 18, 2023 (8 years).

These date ranges were selected to maintain appropriate training-to-testing ratios, while also ensuring that inflow volume distributions across different calibration phases remained similar. Ensuring consistent distribution is crucial, as exposure to extreme inflow events (both high and low) during training enables the model to make more accurate predictions for similar events in the validation and testing phases. The distributions of observed inflows for each period are provided in the Supplementary materials (Figure-A I-1).

2.3.3.2 Reforecast-based (RF)

The RF model combines ERA5 data with 14-day reforecasts, using 351 days of ERA5 data preceding the forecast issue date and 14 days of reforecast data corresponding to the 14-day lead-time starting from the forecast issue date. In the testing phase, operational forecasts will replace reforecast data for the 14-day lead-time. By incorporating reforecast data during training, this model may learn patterns that better characterize forecast data compared to historical observations alone. By exclusively using reforecast data for the 14-day lead-time during training, it is possible to isolate its contribution and evaluate the true predictive potential resulting of its usage.

The reforecast dataset was divided into three independent sets for model training, validation and testing. The complete period of reforecast availability, from March 1996 to April 2022,

was used, considering the 20-year lookback window and the forecast period spanning from March 2016 to April 2023. Given the nature of reforecast, a procedure was developed to ensure that no data was shared between the three sets. The testing phase was first defined from 2007 to 2009, while the remaining data was split into training 80% and validation 20%, resulting in:

- Training: March 1996 to November 2015 (excluding 2007 to 2009),
- Validation: November 2015 to April 2022, and
- Testing: 2007 to 2009.

The distributions of observed inflows for each period are provided in the Supplementary materials (Figure-A I-2).

2.3.3.3 Pretrained reforecast model (E5-RF)

The E5-RF model is a hybrid approach that combines both the E5 and RF models. This was done by first training the E5 model and then performing a secondary training phase using the same method as the RF model to fine-tune the model weights, following a transfer learning approach (Kratzert et al., 2018; Yosinski et al., 2014). This approach allows the model to initially learn general patterns and relationships between input and output data from the E5 dataset, which spans a longer historical period and encompasses a wider range of hydrometeorological conditions. The model is then refined using the RF dataset, which is more representative of forecast data. The main purpose of this hybrid model is to determine whether combining both datasets and incorporating an additional fine-tuning step leads to improved forecasting performance. The same dates used for training the E5 and RF models were used for the training and the fine-tuning, respectively.

2.3.4 LSTM Model development

The three LSTM models were parameterized and structured slightly differently to optimize the balance between the data availability and model performance. Each model was trained using

its respective combinations of input data and model architecture to assess the value of integrating reforecast data in the training process. This procedure not only increased the likelihood of obtaining better predictions but also ensured a fair comparison between models. To identify the most effective configuration for each model, multiple architectures and parameterizations were tested. The model configuration that yielded the best results was selected for further evaluation and is presented in Table 2.2.

2.3.4.1 Data scaling

Data scaling is an important pre-processing step used in DL model training to ensure efficient training and convergence. For this study, a “Min-Max” scaling technique was used, where each feature in the dataset is rescaled to a specific range between 0 and 1. The scaler was parameterized using only the training period data to prevent data contamination in the validation and testing datasets. Within each model, two scalers were parametrized:

- 1) Input data or feature variables (meteorological series of 365 days), and
- 2) Output/target data (14-day cumulative inflow volumes).

2.3.4.2 Model structure and hyperparameterization

A two-layer LSTM model was found to be most optimal, with its structure illustrated in the Supplementary materials (Figure-A II-1). The model consists of a single branch of sequential layers stacked in the following order: an input layer, two LSTM layers, a dropout layer, a dense layer, a LeakyReLU activation layer, and two additional dense layers. The final dense layer has a single neuron, allowing the model to output a single inflow volume forecast when fed a sequence of 365 days of meteorological data.

Different hyperparameters and model structures were explored and tested to optimize model performance relative to forecast quality. For each model, the learning rate, maximum number of epochs, batch size, number of LSTM units, dropout rate and number of neurons in the dense layers were adjusted through trial and error to improve performance, particularly during

validation and testing. The key hyperparameters of each LSTM model are summarized in Table 2.2.

Tableau 2.2 LSTM model hyperparameters

Hyperparameter	E5	RF	E5-RF
Learning rate	1×10^{-3}	5×10^{-5}	5×10^{-5}
Number of epochs	50	50	50
Batch size	128	128	256
LSTM units (first layer)	64	64	64
LSTM units (second layer)	64	64	64
Dropout rate	0.1 (10%)	0.1 (10%)	0.1 (10%)
Dense neurons (first layer)	32	64	32
Dense neurons (second layer)	8	16	8

2.3.4.3 Model training

Each model was trained by generating an inflow volume estimation for each training data sample and comparing it to the observed inflow volume for the corresponding period. This process was repeated for all training samples, allowing the resulting time series of simulated and observed inflow volumes to be compared.

Training was performed using the Keras library from Tensorflow in Python, with the Adam optimizer and the modified Kling-Gupta Efficiency (KGE) objective function (Gupta et al., 2009; Kling et al., 2012) as the loss function (Equation (2.1)).

$$1 - KGE = \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (2.1)$$

where r is the linear correlation coefficient, α is the ratio of the coefficients of variation, representing the relative variability between the observed and simulated data, and β is the bias ratio, which assesses the systematic deviation between the observed and simulated volumes.

The KGE value ranges from $-\infty$ to 1, with 1 indicating perfect agreement between the observed and simulated volumes.

2.3.5 Application to inflow forecasting

Given the variability of hydrometeorological conditions experienced throughout the year, model performance and optimal parametrization can vary from season to season (Muleta, 2012). For example, forecasting spring floods is particularly challenging due to snowmelt processes and change in weather patterns, resulting in poorer performance. Therefore, to account for seasonal variations, assessing hydrological forecasting models on a seasonal basis can be beneficial. Forecasts were separated into four groups corresponding to the LSJ catchment's hydrological seasons: winter (December to March, DJFM), spring (April and May, AM), summer (June to August, JJA), and fall (September to November, SON). Spring typically displays poorer performance metric values, as its complex snow-related processes and highly variable hydrological conditions make it the most difficult season to simulate.

Forecast quality is evaluated by analyzing specific aspects of the forecasts, as summarized in Troin et al. (2021). In addition to the deterministic KGE, other metrics were used to evaluate the ensemble properties of the volumetric forecasts based on three key characteristics: accuracy, sharpness and reliability. Accuracy represents the mean difference between individual forecasts and observations. Sharpness quantifies the variability of forecasts, typically illustrated by the average width of prediction intervals. Reliability assesses how well forecasts align with observations or how often observations fall within the ensemble predictions. Additionally, two other performance attributes are used throughout this paper to evaluate forecast quality: skill and bias. Skill measures the accuracy of a forecast relative to a reference forecast, making it useful for comparative analysis. Bias, on the other hand, captures the difference between the mean forecast and the mean observation, helping to identify whether a forecast systematically overestimates or underestimates actual values.

Two verification metrics were chosen to evaluate forecast quality: the Continuous Ranked Probability Score (CRPS; Hersbach, 2000; Equation (2.2), which captures both accuracy and sharpness, as well as Talagrand diagrams (Talagrand et al., 1999), which assess forecast reliability.

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{+\infty} [F(V_{sim_i} - F(V_{obs}))]^2 dV \quad (2.2)$$

where N represents the size of the forecast ensemble, and the integral represents the squared difference between the cumulative distribution functions of the forecasted and observed values, where the observed value is a scalar. A lower CRPS value indicates a smaller discrepancy, signifying higher accuracy and improved sharpness of the forecast distribution. CRPS values range from 0 to $+\infty$, with 0 indicating a perfect forecast. Since CRPS has the same unit as the predicted variable, it is expressed in hm^3 ($10^6 m^3$) for this study. The CRPS is computed for each forecast date, and its seasonal distribution is obtained by aggregating all values within each season (DJFM, AM, JJA, SON) to observe performance fluctuations throughout the year. Talagrand diagrams, also known as rank histograms (Hamill & Colucci, 1997), offer a visual representation of how observed values are distributed within forecast ensembles. They are created by determining the observation's percentile (or rank) within the forecast ensemble for each forecast date and then examining the histogram of the resulting percentile distribution. Ideally, observations should be uniformly distributed across all forecast ranks, producing a flat histogram where all columns have the same height. Deviations from uniformity indicate biases or issues with forecast spread: U-shaped histograms suggest under-dispersion (forecasts are too narrow), while \cap -shaped histograms indicate over-dispersion (forecasts are too broad). Although Talagrand diagrams effectively visualize forecast reliability, they do not provide a quantitative measure. The Average Bin Distance to Uniformity (ABDU; Arsenault et al. 2016) quantifies the deviation of the rank histogram from perfect uniformity, offering a concise manner to measure the reliability of probabilistic forecasts. It is computed with Equation (2.3):

$$ABDU = \frac{1}{N} \sum_{k=1}^N \left| S_k - \frac{M}{N} \right| \quad (2.3)$$

where S_k represents the number of occurrences in bin k , M is the total number of samples, and N is the number of histogram bins. A perfectly uniform Talagrand diagram would yield an ABDU of zero. However, ABDU is not ideal for comparing a single model's performance, as it is sensitive to the total sample size (M), which varies due to season length (approximately 120 days in winter, 60 in spring, and 90 in both summer and fall). Instead, ABDU is best suited for comparing two models within the same season as it offers a numerical assessment of how closely each ensemble distribution approaches uniformity.

2.4 Results

The following sections outline the main results of the study. Section 3.1 shows how the LSTM models performed during the calibration phases, while Section 3.2 particularly focuses on the reforecast testing period. Section 3.3 presents global model performance in the forecasting application. Section 3.4 compares the accuracy of the forecasts generated by the models on a seasonal basis. Finally, Section 3.5 shows the reliability evaluation of the forecasts.

2.4.1 LSTM model training, validation and testing results

The KGE scores obtained in training, validation and testing for each model are presented in Table 2.3. Additionally, for completeness, the E5 model's performance is included when evaluated using reforecast data in simulation mode over the three periods to facilitate comparison. This test assesses whether the E5-trained model can maintain good simulation performance when replacing the ERA5 data with the reforecast data in the last 14 days of the 365-day samples.

Tableau 2.3 KGE values in training, validation and testing for each of the three LSTM model variants of this study

Model	Training period	Validation period	Testing period
	KGE	KGE	KGE
E5 (training)	0.974	0.920	0.962
RF	0.954	0.961	0.924
E5-RF	0.957	0.961	0.939
E5 (reforecast simulation)	0.949	0.934	0.910

These results indicate that the RF and E5-RF models outperform the E5 model when simulating with reforecast data, even though all models maintain strong KGE values. This outcome was expected, as the E5 model was not trained with reforecast data.

As expected, KGE scores are lower in testing for all models, since the test data is entirely new to them. However, for the E5 in reforecast simulation mode, one might have expected similar KGE scores in all three phases, given that reforecast data was never seen during training. This suggests that the reforecast testing phase may contain specific patterns difficult to predict, leading to poorer model performance. For the RF and E5-RF models, the similarity in KGE scores between the training and validation periods simply indicates effective training and absence of overfitting.

2.4.2 Results of using reforecasts on the testing period

Figure 2.3 shows the ensemble envelopes for each model when forced with meteorological reforecast data on the 2007-2009 testing period. The results indicate that all models perform similarly, even though the E5 model exhibits significantly higher variability in inflow volume compared to RF and E5-RF models. This variability is particularly high during the second half of the year, following the spring freshet. However, the variability is not as high at the occurrence of the annual volume peak, even in absolute terms. This is surprising, considering this period is usually the most challenging to model due to its complex underlying physical

processes. Nonetheless, the KGE values are similar and range from 0.910 (E5) to 0.939 (E5-RF), indicating overall very good performance. The KGE score for each model is presented in each panel's title and is computed with the averages over the full reforecast ensemble members for each reforecast day of the period. Note that the E5 model refers to the E5 simulations with the reforecasts.

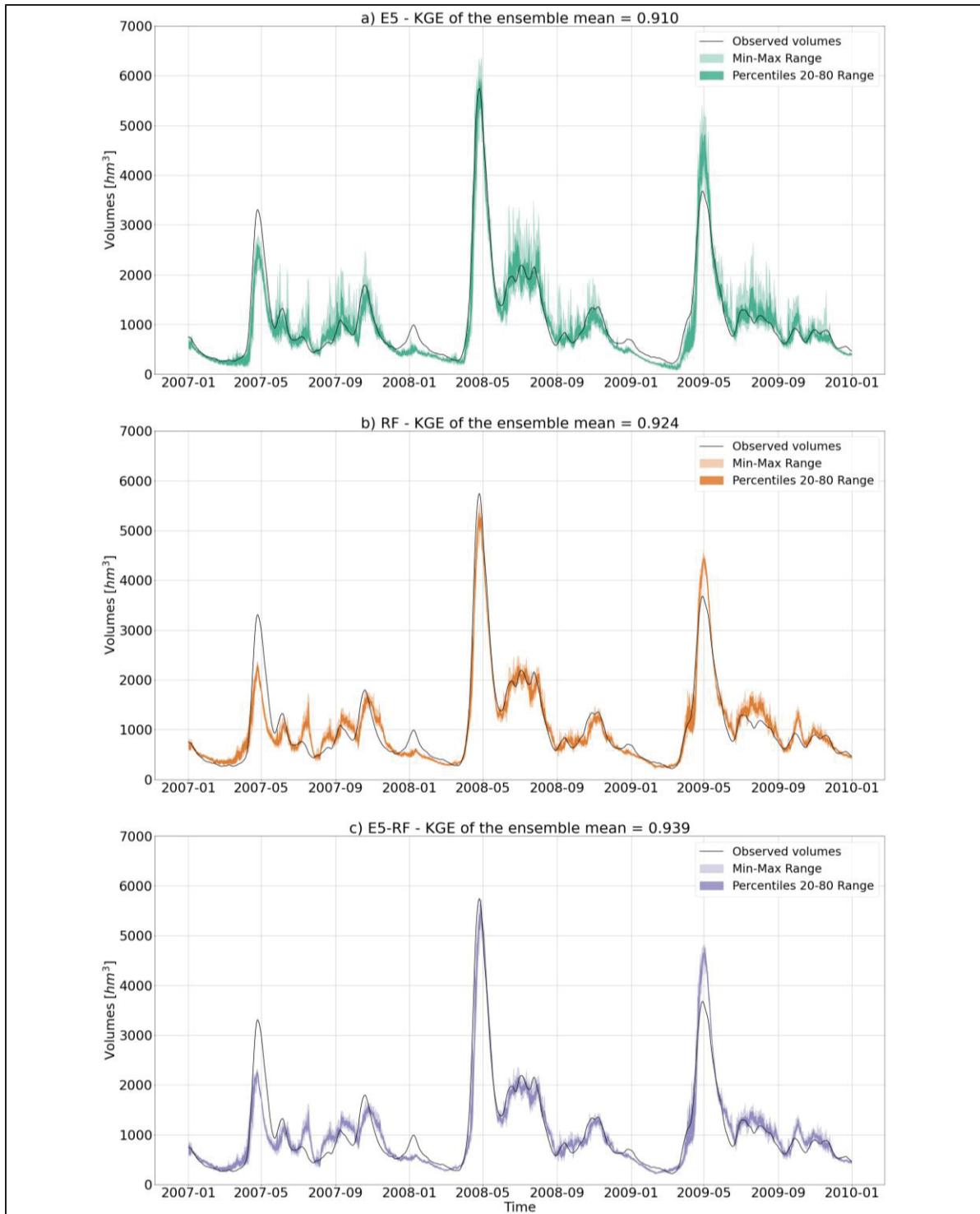


Figure 2.3 Envelopes of the 14-day inflow volumes ensemble forecasts for the LSJ catchment for all models over the reforecast testing period from 2007 to 2009: E5 (a), RF (b), and E5-RF (c). Lighter colors depict the ensemble's full range and darker colors represent the part of the ensemble located within the 20th and 80th percentiles

However, all models struggle with the spring freshet forecasts, notably underestimating 2007 and overestimating 2009. While E5 slightly reduces the underestimation of the 2007 peak, the difference from the other models is not significant. Additionally, the small peaks of January 2008 and 2009 are underestimated, with the models failing to capture any significant volume increase or only a small portion of it. In general, 2007 proves to be most challenging year to forecast, as the models struggle to reproduce the observed patterns. In contrast, the highest volume peak of the period, observed in May 2008, is well forecasted by all models.

2.4.3 Performance of the LSTM models in forecasting

In Figure 2.4, the average of the 50-member forecasts is computed to generate a pseudo-deterministic forecast for each model, allowing the computing of the KGE metric on the operational forecasting setup. Results (KGE value) suggest that the E5 (0.891) model slightly underperforms compared to the RF (0.946) and E5-RF (0.939) models.

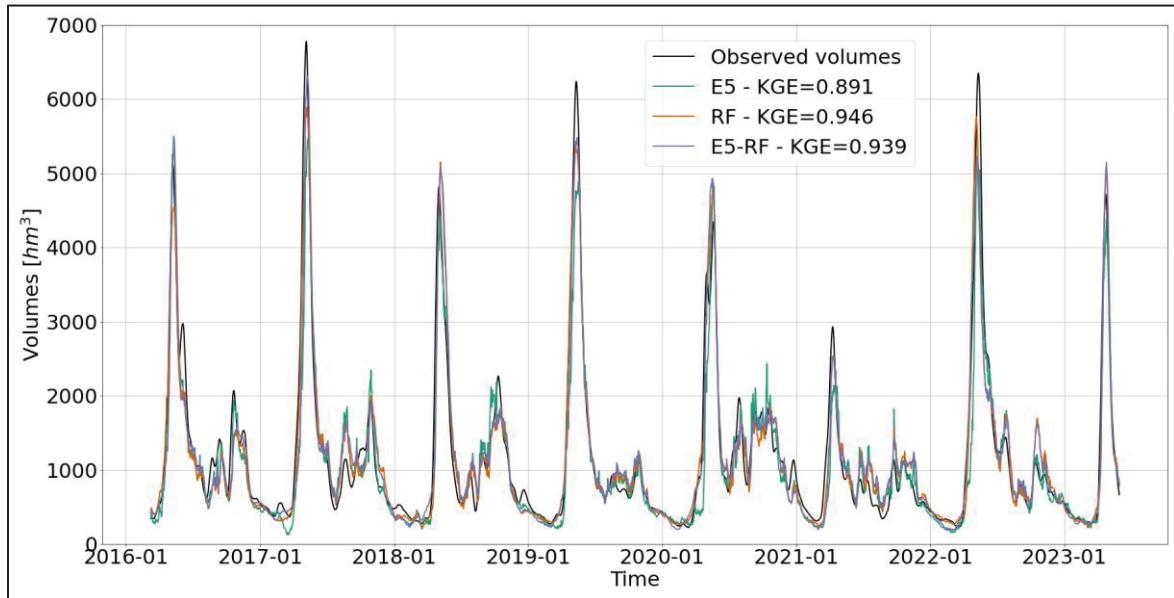


Figure 2.4 Average of the 50-member ensemble forecast for the 14-day inflow volume.

The E5 model (green curve) generally produces the worst predictions, particularly struggling with the magnitude of the spring freshet, which corresponds to the annual peak volume. The RF and E5-RF models demonstrate strong performance over the entire forecasting period, with both models predicting the annual peak volume with similar accuracy. None of the three models exhibit any distinctive bias outside of slightly underestimating the winter low flows most of the years.

The skill of ensemble forecasts is analyzed in Figures Figure 2.5 through Figure 2.7. These figures present the inflow volume curves of the ensemble forecasts generated by the E5 (Figure 2.5), RF (Figure 2.6) and E5-RF (Figure 2.7) models. The top panels cover the entire forecasting period, while the two smaller plots focus on the years with the strongest (2017) and smallest (2021) spring freshet. The average KGE and CRPS scores above each plot quantify the overall performance and skill of the forecasts. The KGE was computed using the average of the 50 members for each forecast day, while the CRPS was averaged over all forecast days. Overall, the three models achieved strong KGE and CRPS scores over the full forecast period.

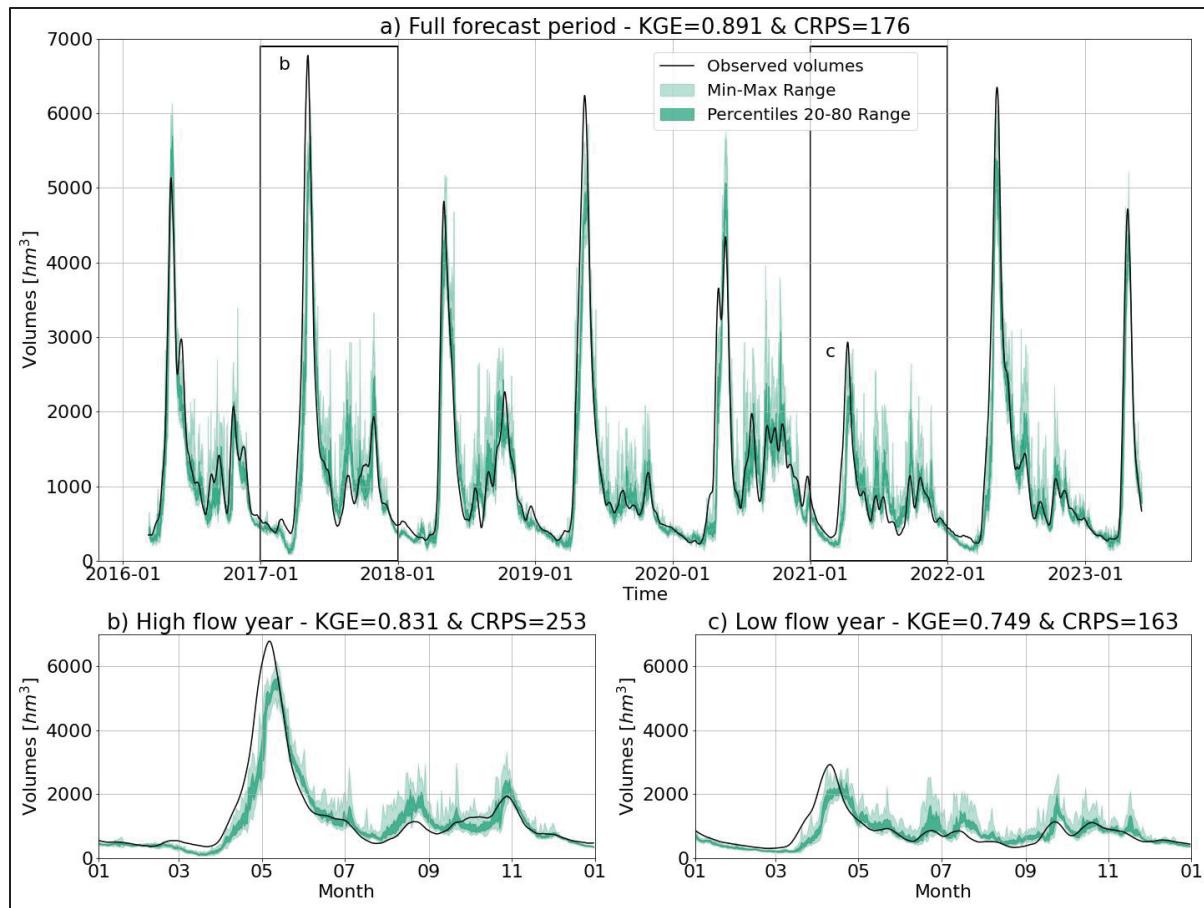


Figure 2.5 Envelopes of the 14-day inflow volume curve ensemble forecasts for the E5 model: entire forecasting period (a), strongest (b) and smallest (c) spring freshet. Lighter colors depict the ensemble's full range, while darker colors represent the forecasts located within the 20th and 80th percentiles

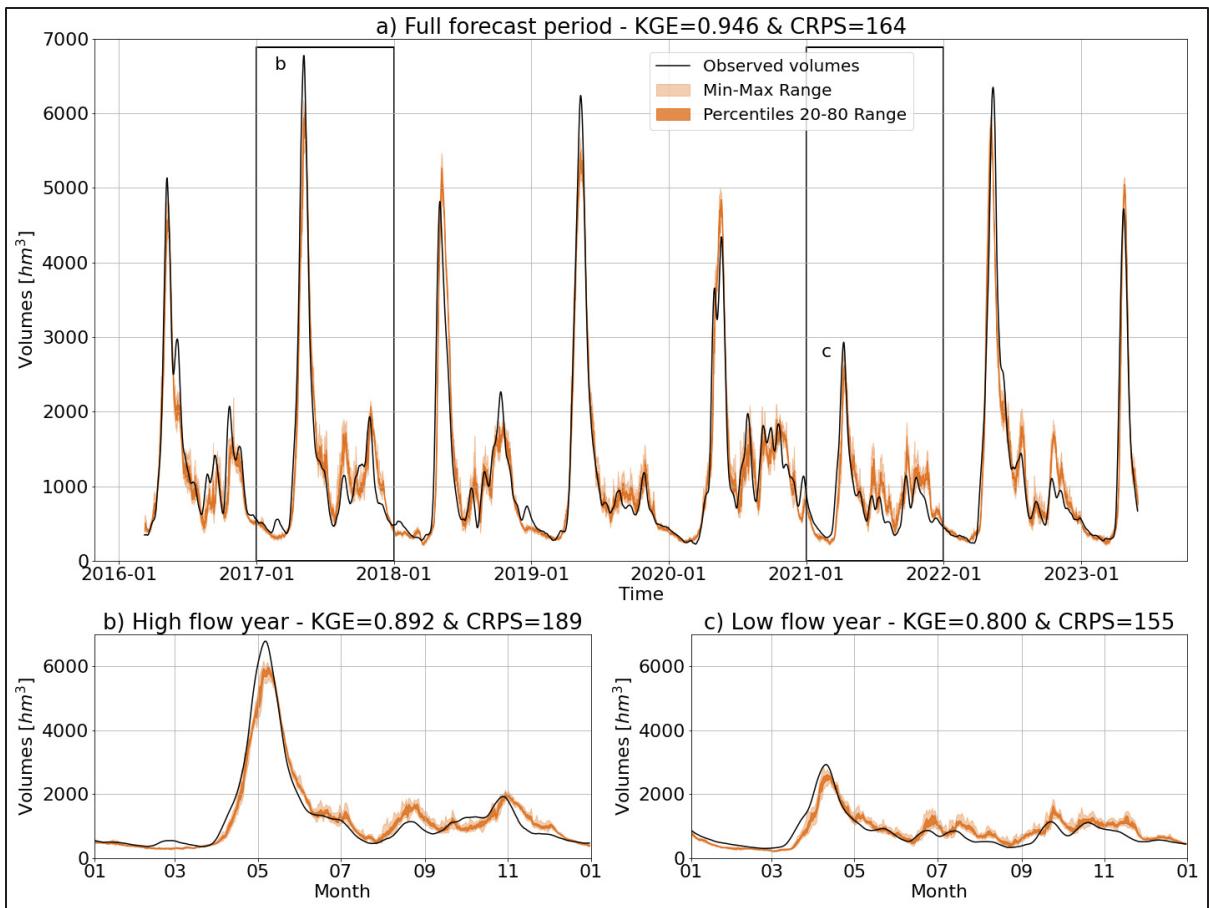


Figure 2.6 Envelopes of the 14-day inflow volume curve ensemble forecasts for the RF model: entire forecasting period (a), strongest (b) and smallest (c) spring freshet. Lighter colors depict the ensemble's full range, while darker colors represent the forecasts located within the 20th and 80th percentiles

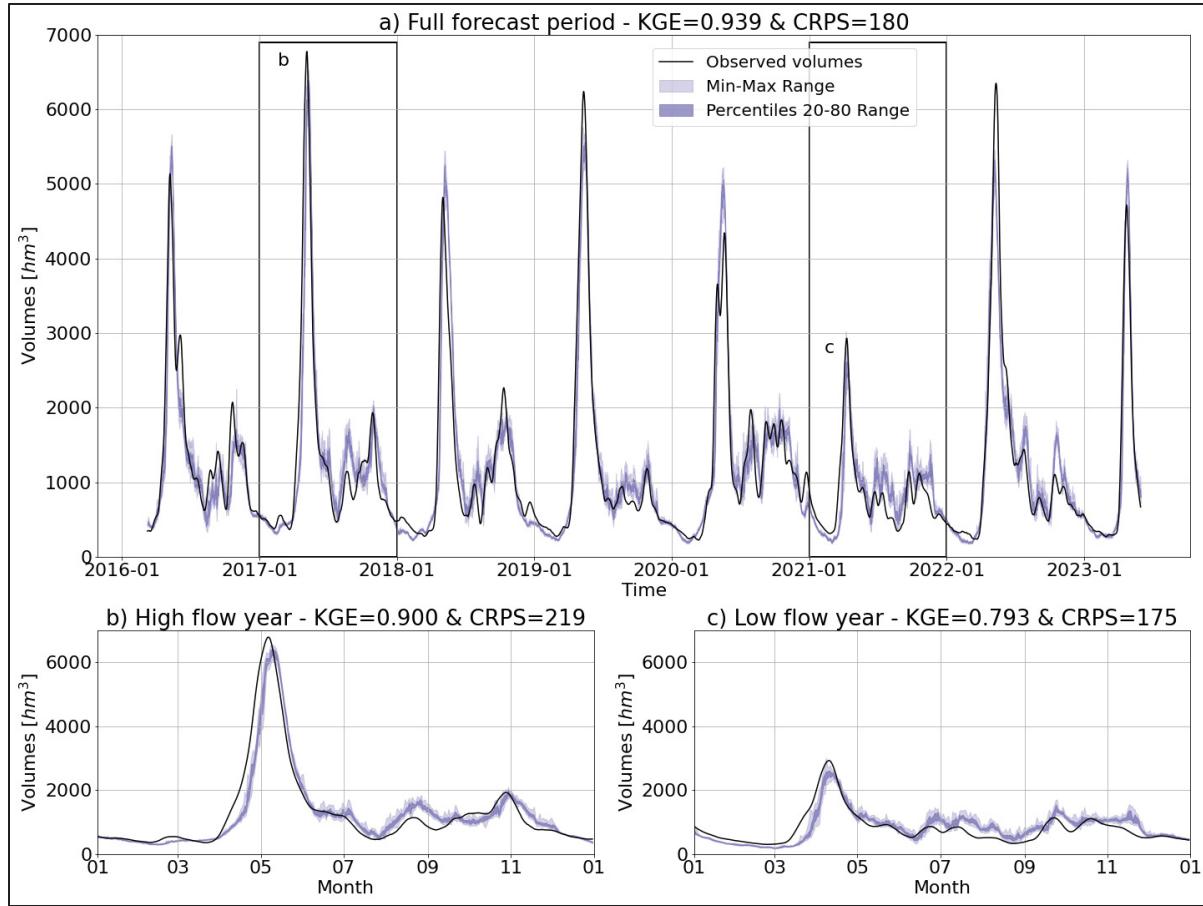


Figure 2.7 Envelopes of the 14-day inflow volume curve ensemble forecasts for the E5-RF model: entire forecasting period (a), strongest (b) and smallest (c) spring freshet. Lighter colors depict the ensemble's full range, while darker colors represent the forecasts located within the 20th and 80th percentiles

The E5 model forecasts demonstrate a high degree of variability, especially the spring freshet through the end of the year, similar to when it was forced with reforecast data. In contrast, the RF and E5-RF models, both trained with reforecast data, generate ensembles with significantly lower variability, as depicted by the thinner envelopes throughout the forecasting period and the reduced width between the light and dark envelopes. Both RF and E5-RF models generate similar forecasted inflow volume curves, with under- and overestimations occurring at similar periods. E5-RF has a better KGE, but a slightly worse CRPS score compared to E5, indicating possibly better reliability and better uncertainty representation by the latter. The RF model obtains the best KGE and CRPS scores, with values of 0.946 and 164, respectively.

While the models generally reproduce well the observed inflow volume curve and demonstrate no evident bias, some discrepancies remain. Most notably, apart from 2022, the models underestimate the observed volume from the beginning of the year until the start of the freshet receding and then over-estimate from the falling limb and onwards, as highlighted in panels b) and c) of Figures Figure 2.5 through Figure 2.7. Such behavior aligns with the average inflow volume curves shown in Figure 2.4. A more detailed analysis of this model behavior is presented in the discussion.

In terms of peak volumes, the E5 model performs poorly in predicting the spring freshet. The ensemble's 20th and 80th percentiles rarely encompass the observed volume, and in some cases, the complete ensemble fails to capture the target volume. In contrast, the RF and E5-RF models demonstrate greater accuracy and consistency, though their ensembles sometimes completely under- or overestimate the observed volume due to their lower spread. Panel a) of figures Figure 2.5, Figure 2.6 and Figure 2.7 clearly illustrate the models' failure to accurately predict the magnitude and timing of the spring freshet, as the prediction envelope rarely encompasses the observed peak volume. Panel b) reveal a tendency to underestimate the larger peak flows observed in 2017, 2019, and 2022, while smaller peak flows are generally overestimated, as shown in panel a). This pattern suggests that the models struggle to differentiate the unique hydrologic conditions of each year's spring freshet, leading to the generation of similar peak volumes regardless of the actual hydrometeorological scenario. An exception to this trend is the 2021 spring freshet, which had significantly lower volumes compared to other years but was still underestimated, as shown in panel c). This anomaly resulted from lower snowfall combined with frequent temperature increases, leading to reduced snow accumulation and consequently lower snowmelt inflows. Nevertheless, panel c) demonstrates that the models effectively captured these atypical snowpack dynamics, even slightly overestimating the reduction in freshet volume.

A time delay throughout the forecast period can be seen, shown more evidently in panels b) and c). This slight lag between the observations and the forecast envelope, particularly during

the spring freshet, results in the forecasted peak being reached a few days after the observed peak. The lag can reach one week for the E5 model, while it ranges from 1 to 3 days for the RF and E5-RF models.

All models perform better in 2017 than in 2021, as indicated by the higher KGE scores. However, the CRPS scores are worse in 2017 compared to 2021 due to the substantially greater inflows observed in 2021. These larger volumes amplify the absolute differences between observed and predicted values, even though, in relative terms, the forecasts are actually closer to the observation.

2.4.4 Comparison of model performance on the seasonal scale

Figure 2.8 presents the boxplots of CRPS scores for each forecast day and Table 2.4 indicates the median CRPS, at the season scale. A non-parametric Wilcoxon test was also conducted to determine whether the forecast distributions are significantly different or not.

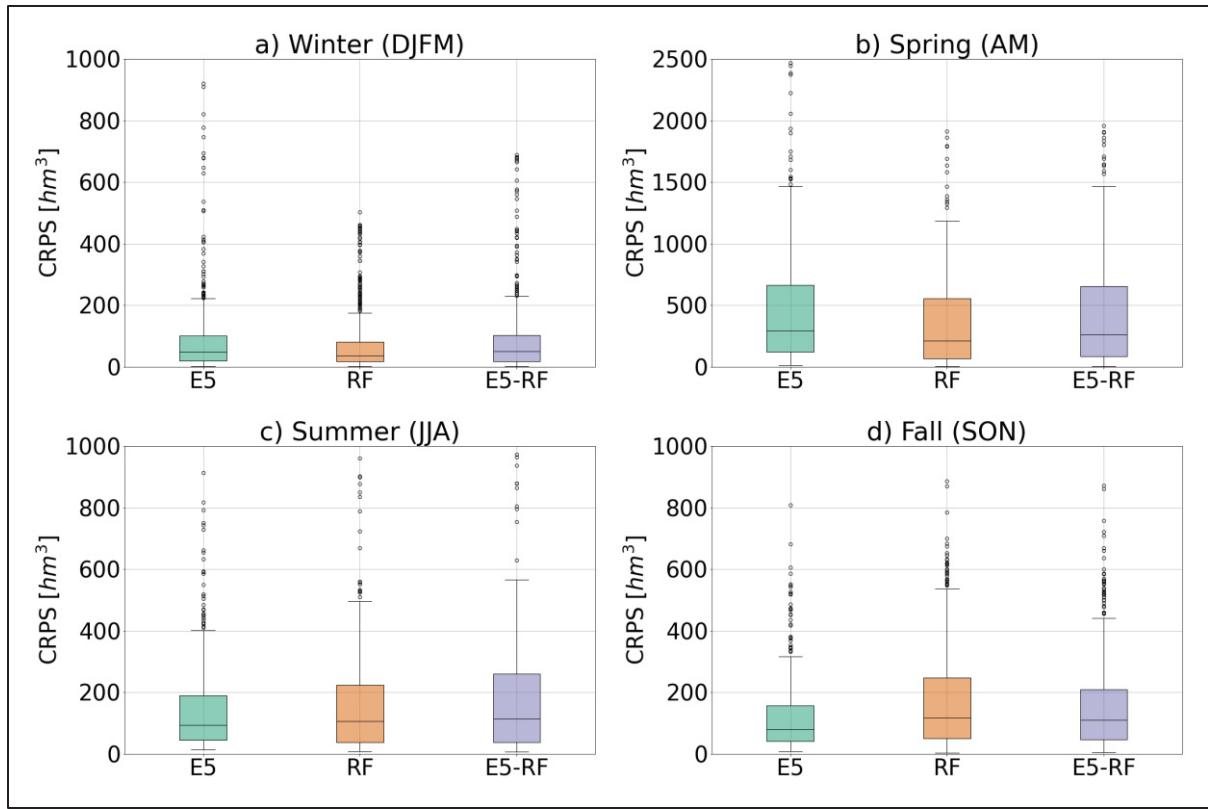


Figure 2.8 Boxplots of 14-day volume ensemble forecasts CRPS values for the E5 (green), RF (orange) and E5-RF (purple) models over winter (a), spring (b), summer (c), and fall (d).

The number of points in each boxplot represents the number of issued forecasts for that season. Note that the y-axis range of spring (b) is larger from the other all panels

Tableau 2.4 Median CRPS results of the forecasts generated by the three models. Bold CRPS values indicate a significantly different forecast according to the Wilcoxon rank test with a significance level of 5%

Model	Winter	Spring	Summer	Fall	Annual
E5	49	293	94	79	85
RF	36	214	107	117	82
E5-RF	50	263	114	111	92

At the seasonal scale, no model consistently outperforms the others. In summer, although the boxplots in Figure 2.8 appear to be different, the Wilcoxon test results on the median (Table

2.4) indicate that the forecast distributions are statistically similar. The E5 model performs best in summer and fall and remains competitive in winter, but it performs significantly worse in spring, as indicated by its much higher median CRPS. Also, this model has a higher spread, notably for winter and spring seasons. The RF model outperforms the two other models in winter and spring, but its skill declines in fall, where it performs slightly worse than the E5-RF model. The E5-RF generally demonstrates the worst performance across seasons, never leading in any of the four. However, in spring, it performs similarly to RF even though its median CPRS is higher, with both models outperforming E5.

Overall, training with reforecast data improves inflow forecasts in spring, a crucial season for hydropower reservoir management. However, the opposite is observed in fall, where the E5 model (which does not use reforecast data in training) significantly outperforms the other two models. Accurately simulating inflows for this season is also essential in Nordic regions, where reservoir levels must be carefully managed to ensure sufficient water availability for meeting high winter electricity demands.

2.4.5 Forecast reliability

The reliability of ensemble forecasts at the seasonal and annual scales is analyzed in Figure 2.9 using Talagrand diagrams. Each diagram contains a number of points corresponding to the number of issued forecasts for that period, equal to 870 (winter), 488 (spring), 644 (summer), 637 (fall) and 2,639 (annual).

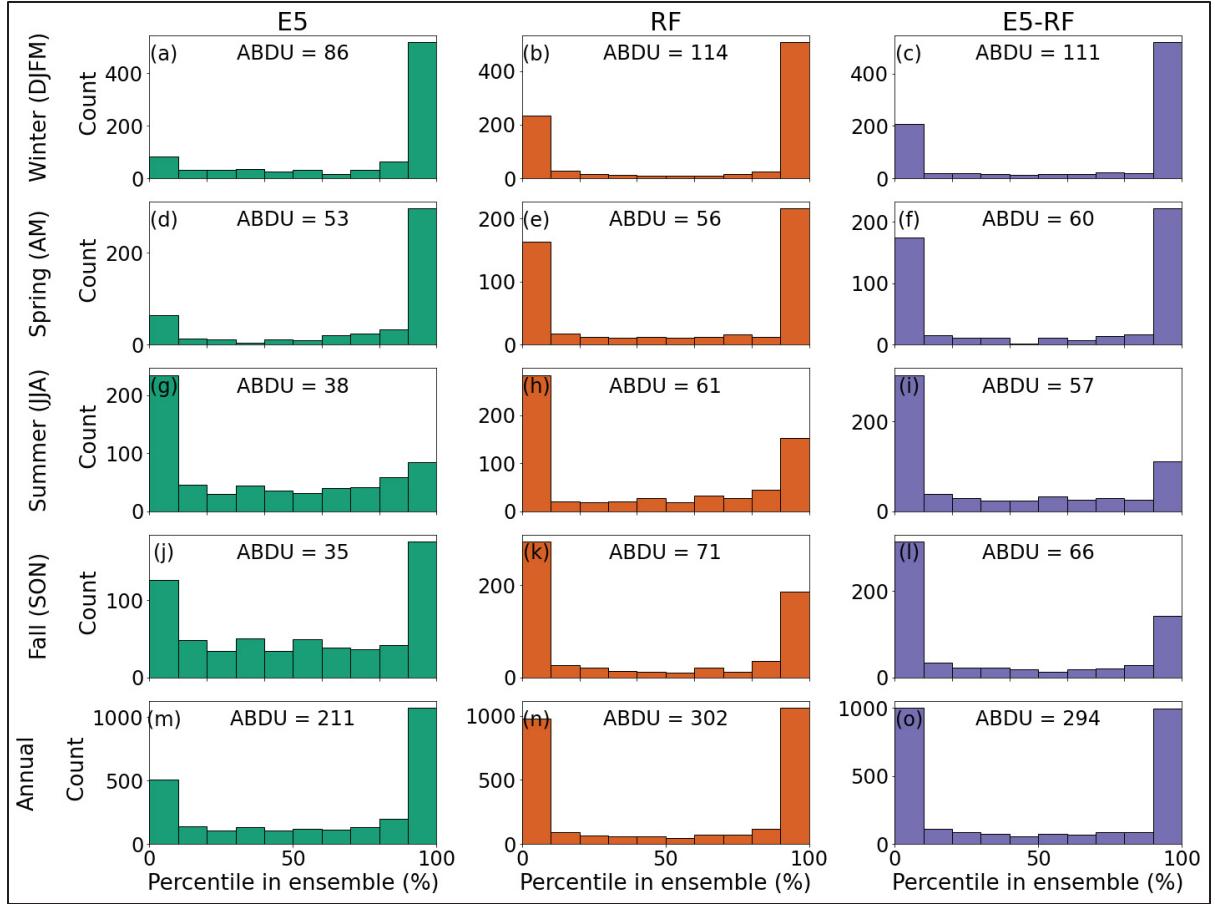


Figure 2.9 Reliability of the forecasts generated by the E5 (green), RF (orange) and E5-RF (purple) models as demonstrated with Talagrand diagrams showing the rank (in percentiles) for the 14-day observed inflow volumes within the 50-member forecast ensembles

All models exhibit clear under-dispersion across all four seasons and the full forecast period, as indicated by the frequent U and L-shaped diagrams, far from being uniform distributions. This suggests that the forecasts are generally biased and/or under-dispersed, leading to unreliable predictions. The inverted L-shaped histograms (e.g., E5 in winter and spring) indicate that forecasts tend to underestimate the observed inflow volume. Despite this behavior, the E5 model is more reliable than the other models during summer and fall, although it shows an overestimation bias in summer.

Overall, the E5 model generates more uniform forecast distributions throughout the year, obtaining the lowest ABDU values in every season. The largest ABDU difference between E5

and the other models is seen in fall, but winter and summer also show noticeable improvements in uniformity for the E5 model's forecasts.

2.5 Discussion

The following sections provide a detailed analysis of the results. Section 4.1 evaluates model performance during the reforecast training phases, while Section 4.2 examines their forecasting performance. Section 4.3 investigates the value of reforecasts, and Section 4.4 discusses the added benefits of LSTM models. Section 4.5 revisits the study's objectives, and Section 4.6 outlines the associated limitations.

2.5.1 The performance of LSTM models in testing

Figure 2.3 indicates that all models achieve good performance throughout the reforecasts training phases, with KGE scores ranging from 0.910 to 0.939 in testing. These results confirm the training process successfully enabled the models to learn and simulate hydrological processes in the LSJ catchment. The testing period of the reforecasts was chosen to assess model performance after their training (in their final configuration) since it closely resembles an operational forecasting framework.

A slight decrease in performance from training and validation to testing is expected. However, despite the fact that the E5 model was not trained on reforecast data, it still exhibits a similar performance drop in testing as the RF and E5-RF models. This suggests that the E5 model learned patterns in a comparable manner. One possible reason for this distinct drop in skill even for the E5 model is that the reforecast testing period contained particularly difficult patterns to predict, leading to poorer performance. Also, Figure 2.3 illustrates that the E5 model underestimates winter low flows more than any other model. This could be due to lower inflow volumes encountered in training, causing this model to reproduce those lower inflow volumes magnitudes in the reforecast testing period.

In general, the E5 model appears slightly less well-calibrated than the two other models, as indicated by the KGE scores. However, these differences are minimal in the forecasting context, where uncertainties in meteorological data and hydrological processes modelling play a significant role. As demonstrated by Davidson-Chaput et al. (2025), simulation skill does not necessarily translate directly to forecasting skill.

2.5.2 The performance of LSTM models in forecasting

This section discusses the results obtained from the forecasting application, focusing on key areas for improvement. Specifically, it examines issues related to the magnitude and timing inaccuracies affecting the simulation of the spring freshet, biases observed during low flow periods, and the reliability of the ensemble forecasts.

2.5.2.1 Issue with timing delay in peak inflow volumes

As previously shown in panel a) of figures Figure 2.5, Figure 2.6 and Figure 2.7, all models produce forecasts of good quality with ensemble mean KGE values above 0.891 and CRPS values smaller than 180 hm³. This difference may seem high, but relative to the average 14-day inflow volume over the entire period of 1177 hm³, it represents a reasonable maximum error, equivalent to 15.3% of the average. However, many forecast attributes remain sub-optimal.

Panels b) and c) of figures Figure 2.5, Figure 2.6 and Figure 2.7 highlight that forecast timing is a recurring issue, which was also noticed in the testing period (Figure 2.3). This behavior is more evident for the E5 and E5-RF models. This timing delay contrasts somewhat with the findings of Sabzipour et al. (2023), who observed proper timing in their hydrographs for short-term streamflow simulations. However, they noted that timing issues became more pronounced as lead time increased, which could explain the temporal shifts observed in the 14-day inflow volume forecasts. It is worth noting that their study focused on streamflow rather than inflow volumes, as examined here. Such delays result in an underestimation of the rising limb and peak flow, and an overestimation of the falling limb, and a persistent offset from observed

inflow volumes, which affects model performance. The RF model exhibits smaller timing error, indicating that training with reforecast data may help capture biases between ERA5 and forecast data. Since the same forecasting system generates both forecasts and reforecasts, RF may have internally learned the meteorological model's biases and properly adjusted for them when using forecast data.

The poor simulation of the spring freshet could be attributed to the complex snowpack dynamics occurring at that time. However, the consistent timing delay throughout the year challenges this hypothesis. One potential solution to reduce the delay is to provide the model with observed streamflow or inflow volumes preceding the forecast date. This could, similarly to traditional hydrologic models using data assimilation, offer crucial insights into the current catchment hydrogeologic state, helping the model to generate more realistic inflow forecasts. Sabzipour et al. (2023) used the daily streamflow from the year preceding the forecast date but encountered challenges in accurately capturing timing for longer lead-time streamflow simulations. This suggests the possibility that including this information may not significantly improve inflow volume forecasts at the timescales used in this study. An alternative approach could involve providing the model with the historical mean streamflow for the same day of the year as the forecast date, such as in Hunt et al. (2022). However, this information may be less representative of the catchment's current conditions.

2.5.2.2 Biases in the simulation of low-flow periods

The models struggle to accurately predict winter low-flows, particularly before the rising limb of the inflow volume curve. Panel c) (and panel b in a reduced fashion) in Figures Figure 2.5, Figure 2.6 and Figure 2.7 show that the models underestimate inflow volumes during winter low-flow periods, which are characterized by colder temperatures, reduced snowmelt capacity, and lower soil infiltration. The meteorological forecast model may predict temperatures that are too cold during that period, generating less inflow from snowmelt. Although winter snowmelt is less frequent, it can still occur during winter when temperatures exceed 0 °C. This underestimation bias suggests that the models struggle to accurately capture snowpack

dynamics during the coldest months. This pattern is also observed during the reforecast testing period (Figure 2.3), where the models underestimate winter low flows in two out of three years. If the same issue occurred during training, it would indicate that the models failed to learn the dynamics of winter low flows. However, this problem is less pronounced during the training phase, with some years even showing overestimation. The inflow volumes generated by the models during training are provided in the Supplementary Materials (Figure-A III-1). Minor weather variations between the training and forecasting periods could partly explain the bias. Although subtle, these differences mean the weather conditions encountered during training and forecasting are not statistically identical. This discrepancy may result in bias if the models do not adapt accordingly. During winter, mean temperatures were about 1°C higher, and total precipitation was approximately 0.3 mm/day greater during the forecasting period, which represents a significant rise considering the average total precipitation was around 2 mm/day for this season. The models may struggle to account for the increased inflow volumes resulting from enhanced snowmelt and precipitation. Additionally, warmer temperatures could have triggered earlier snowmelt, leading to reduced snowpack and consequently less inflow available before the freshet. Still, the issue may also stem from insufficient snowmelt generation under cold temperatures or an overall underestimation of baseflow, as in years with milder winter temperatures, a smaller snowpack may result in more water infiltrating the soil in late fall, therefore increasing baseflow. This early underestimation propagates into the spring freshet, where the models struggle to capture peak inflow volumes, similar to findings in Muluye & Coulibaly (2007).

Summer flows, which are also of low magnitude although higher than winter low flows, are predominantly over-estimated. This outcome is unexpected, as models typically perform better during periods with higher-probability total inflows (i.e., medium to low flows), as noted by Muluye and Coulibaly (2007) and Herbert et al. (2021). Given the models' extensive exposure to such flow magnitudes during training, they should theoretically be more robust and better adapted to these conditions. One potential reason for this overestimation bias could be due to the forecasting model still being too cold in summer, which would underestimate the actual evaporation and evapotranspiration, and thus result in higher inflows. Similar to winter, the

weather conditions during the training and forecasting periods in summer are not statistically identical, which may have contributed to the observed bias. Although the differences are once again quite minor, with mean temperatures about 0.75°C higher and total precipitation 0.4 mm/day greater (with an average total precipitation of around 4 mm/day for the season) during the forecasting period, these changes can still affect model performance if the models fail to consider these biases. The models may overemphasize the impact of total precipitation on inflow volumes, leading to overestimations, even though temperature rises should partially offset this effect by promoting greater evaporation and evapotranspiration. However, if the models do not adequately account for these compensatory processes, the resulting forecasts may be biased accordingly. Baseflow contributions, which are influenced by the input data provided, may also be misrepresented by the models. Given they do not have any physical representation of hydrological processes, it would seem that the LSTM models have some difficulty with these processes that are affected over longer periods.

2.5.2.3 Challenges in forecasting the spring freshets

Despite studies showing that LSTMs perform well under more extreme conditions (e.g., Lees et al., 2021; Damavandi et al., 2019), the spring freshet remains challenging to forecast, especially for 14-day inflow volumes. Even minor lag at the beginning of the forecast window can amplify error throughout the forecast period.

In general, lowest and highest spring freshet flows are misrepresented, smaller freshets are overestimated, while larger ones are underestimated, except in 2021, which had the weakest spring freshet of the forecast period. Underestimating high inflows is a frequent issue in hydrological forecasting using data-driven models (e.g., Sabzipour et al., 2023; Herbert et al., 2021; Sattari et al., 2012). Since peak flows occur less frequently, the models may struggle to generalize their prediction, as training data is dominated by low-to-medium inflows.

E5 mainly underestimates peak flows (Figure 2.4), but it also generated the highest peak flows out of all models when forced with reforecast data (Figure 2.3). E5's poor prediction of the

2017, 2019 and 2022 spring freshets, the strongest events, could be explained by fewer high-magnitude inflow volumes in its training dataset, as shown in Supplementary materials (Section S1). The reforecasts dataset structure (10-member ensemble for the previous 20 years per reforecast date) gives RF and E5-RF models greater exposure to high inflow volumes during training, which may explain their better performance in more extreme conditions compared to E5.

The models generate forecasts with similar accuracy, but seasonal variations are noticeable. The RF model performs best in winter and spring, while E5 outperforms in summer and fall. E5-RF is more balanced, neither excelling nor underperforming except for summer, where it shows poorer skill. In spring, both models trained with reforecast data (RF and E5-RF) perform better, particularly RF. However, all models exhibit lower accuracy, as indicated by the higher CRPS values in Figure 2.8b. This behavior is likely influenced by the spring freshet, which presents two challenges: complexity and magnitude. Spring freshet is difficult to simulate due to snow-related processes and highly variable hydrometeorological conditions. For the LSJ catchment, the spring freshet represents the highest flows and inflow volumes of the year, leading to higher absolute CRPS values, similar to winter, but with the opposite effect (e.g. Figure 2.8a).

2.5.2.4 Forecast reliability issues

Forecast reliability remains an issue for all models, especially for RF and E5-RF. LSTM forecasts are under-dispersed in all seasons as shown by the U-shaped Talagrand diagrams in Figure 2.9, similar to Armstrong et al. (2025). Winter and spring are particularly problematic, given the complex physical processes involving snow, but summer and fall also deviate from a uniform distribution. E5 exhibits the largest underestimation bias (reversed L-shape diagram) in winter and spring, suggesting poor accuracy; its ensemble rarely encompasses observations, even with its wider spread. Figures Figure 2.5, Figure 2.6 et Figure 2.7 suggest that forecast variability is proportional to inflow volume. This means that winter forecasts show smaller variability, indicating the winter bias could be less impactful than the spring bias, which also

exhibit greater variability. However, summer and fall forecast are more reliable, especially for the E5 model. This increase in robustness could be explained by higher forecast variability, and steadier hydrometeorological conditions, which facilitates modeling. It is notable that LSTM models tend to be overconfident compared to conceptual or physically-based traditional hydrological models (e.g., Armstrong et al. 2025; Sabzipour et al. 2023). Consequently, post-processing is necessary to correct high underdispersion levels (i.e., the U-shaped Talagrand from Figure 2.9).

2.5.3 On the impact of using reforecast data in the training of LSTM models

To our knowledge, no other study has explored the effect of using reforecast data to train deep learning models in a forecasting context, limiting further comparisons with our results. Given that additional calibration data generally improves model performance (e.g., Ayzel & Heistermann, 2021), we expected the RF and E5-RF models, which were trained with reforecast data, to outperform the E5 model due to their longer calibration period. However, our findings indicate that using reforecast data during training impacts model performance in complex ways, improving forecast skill in some circumstances while reducing it in others.

2.5.3.1 Impact of reforecast training on LSTM model performance

One key difference is that E5 produces highly variable forecasts, while RF and E5-RF generate sharper ensembles predictions. This behavior may be linked to training exposure. E5 was trained exclusively on observations (i.e., ERA5) data, making it less robust to forecast uncertainty and potentially unable to adapt to the wider range of conditions found in forecast data. Training with reforecasts may help models develop a more comprehensive understanding of hydrological processes, enabling them to better capture diverse scenarios in forecasting mode.

2.5.3.2 Seasonal performance differences

Both RF and E5-RF models perform better in winter and spring, especially RF (Table 2.4). This is important for water managers, as accurately forecasting the spring freshet is essential for hydropower management and flooding protection. Training with reforecasts improves model adaptation to winter low flows and spring high flows, as the longer datasets provides exposure to a broader range of meteorological scenarios, including extreme events (Hammil et al., 2013). E5 is more skillful than RF and E5-RF in summer and fall, indicating that training solely on ERA5 is sufficient for modeling steadier hydrometeorological conditions. In contrast, reforecast training may introduce unnecessary complexity, leading models to learn unsuitable and overcomplicated patterns that hinder performance under less extreme conditions.

2.5.3.3 Forecast reliability and ensemble sharpness

Figure 2.9 shows that training with reforecasts diminishes forecast reliability. E5 is the most reliable model across all seasons, likely due to its higher ensemble variability, which allows it to better capture observed inflow volumes even when forecasts are biased. RF and E5-RF models are consistently under-dispersed, meaning that when a bias is present, most ensemble members reproduce the bias, failing to encompass the observation. In spring, errors in timing and magnitude of the rising and falling limbs of the inflow volume curve lead to inaccurate ensemble forecasts. The leftmost bin in the Talagrand diagram reflects falling limb overestimation, while the rightmost bin represents rising limb underestimation. Since the spring freshet covers much of the season, these biases are clearly seen in panels d, e and f of Figure 2.9.

The decision to incorporate reforecasts into training should depend on forecasting priorities. For winter low flows and spring freshet peak flows, reforecasts training adds significant value. For summer and fall predictions, reforecasts training tends to degrade forecast skill. Training with reforecasts enhances ensemble sharpness, but this can reduce forecast usefulness when reliability is a priority, and the forecast is regularly biased.

2.5.4 Added value of the LSTM models

As previously noted, LSTM models outperform traditional hydrological models in various streamflow forecasting applications (e.g., Arsenault et al., 2023; Sabzipour et al., 2023). However, their operational value extends beyond superior forecast quality. LSTM models provide a more streamlined framework compared to conventional hydrological models, which involve multiple complex steps such as input pre-processing, data assimilation, modeling hydrological processes, output post-processing, and streamflow aggregation into inflow volumes. In contrast, LSTMs can eliminate these steps by directly processing meteorological data inputs and producing inflow volumes as outputs. This faster and simplified approach enhances their practicality, facilitating operational implementation and reducing computational demands.

However, LSTMs are not without limitations. One of their main drawbacks is their lack of adaptability for performing diverse tasks. In this study, the LSTM models were specifically designed to forecast inflow volumes, meaning that if streamflow forecasts were required, an entirely new model would need to be trained. Consequently, it is advisable to develop multiple models in advance and assess their forecast quality before deploying them operationally, as an LSTM model optimized for one task may not perform well for another.

2.5.5 Review of the study's objectives

The main objective of this study was to evaluate the performance of LSTM models in forecasting cumulative 14-day inflow volumes without relying on intermediate daily streamflow predictions. This type of forecast is highly valuable for reservoir managers, as it helps them adjust operating rules to optimize hydropower generation.

All three models tested demonstrated good performance, with average KGE scores ranging from 0.891 to 0.946. Even though performance metrics (KGE, CRPS, and ABDU) varied across models, they also exhibited common trends, as previously discussed. However, a clear issue identified in this study was forecast reliability, as the ensembles were under-dispersed

across all seasons. This leads to the second question of this study: Does training with meteorological reforecast data enhance model performance and forecast quality?

The results indicate that training with meteorological reforecasts influences several characteristics of ensemble forecasts. On one hand, the integration of reforecasts in training led to improved accuracy in winter and spring, particularly in forecasting peak inflow volumes during major spring freshets, which is critical for reservoir safety and profitability. On the other hand, forecast reliability is not improved, the E5 model (trained only on ERA5) being more robust in all seasons. Models trained with reforecasts are less skilled in summer and fall, exhibiting lower accuracy and reliability. Ultimately, there is no clear consensus on the benefits of adding reforecast data in training; it depends on the forecasting objectives.

2.5.6 Limitations

This study has several limitations that provide potential avenues for future research:

- 1) Limited generalizability of the findings: Having conducted the study on a single Nordic catchment, the results may be site-specific and not transferable to other regions with differing climatic, hydrological, or land use characteristics. Indeed, this study focuses exclusively on the LSJ catchment, partly due to the complexity of obtaining reforecast data for other regions. However, data quantity is not a limitation, as reforecast datasets offer a comprehensive and extensive database. To further enhance model calibration, an alternative approach could involve dividing the LSJ catchment into smaller subcatchments during the training phase, while maintaining the same forecasting application method for the entire catchment. For instance, splitting it into 10 sub-basins would effectively increase the training data tenfold. Given that longer calibration length typically improves model performance, exploring this regionalization strategy at the catchment scale could be beneficial.
- 2) Hyperparameter optimization: LSTM model hyperparameters were determined through a trial-and-error approach, which may not have resulted in the optimal configuration.

However, reasonable ranges for the hyperparameters were tested under a variety of configurations, suggesting that the results should be robust despite this limitation.

- 3) Inclusion of hydrological predictors: LSTM models were trained using eight meteorological variables, which was the maximum available across all datasets. However, to better align with operational forecasting, other hydrological predictors such as recently observed streamflow could be incorporated as model inputs. This could help the model minimize forecast error by providing a more realistic representation of hydrological state at the time of prediction.
- 4) Alternative objective functions: The KGE metric used in this study does not account for the target variable's magnitude, meaning that it does not prioritize peak flow timing, which is crucial for flood risk mitigation and hydropower management.
- 5) Benchmarking against operational forecasts: This study did not compare model performance against a 14-day operational weather and climatology forecasts. This comparison would be valuable to assess the model performance in the absence of meteorological forecast data.

2.6 Conclusion and recommendations

This study evaluates the usefulness of reforecasts data in LSTM model training for forecasting 14-day cumulative inflow volumes over the Lac-Saint-Jean (LSJ) catchment in Canada. Three LSTM models were developed and trained using different combinations of input data before being applied to real 14-day meteorological forecast data from ECMWF. An objective was to determine whether incorporating a reforecast dataset in training improves model performance and forecast quality.

The key findings of this study are as follows:

- 1) Good model performance in training: All models achieved good performance during training, with a minimum KGE of 0.910. The models successfully forecasted the strongest inflow volume of the testing period.

- 2) Accurate but unreliable ensemble forecasts: The forecasts were under-dispersed across all seasons, indicating low ensemble reliability.
- 3) Challenges in peak flow prediction: The models struggled with the timing and magnitude of the spring freshet. Peak flow events were generally not well captured, limiting forecast accuracy.
- 4) Trade-offs in using reforecast data: Models trained with reforecast data were more accurate and produced sharper ensembles. However, these models were less robust, meaning their usefulness depends on the forecasting objectives.

This study shows that LSTM models have strong potential in inflow volume forecasting, producing high-quality forecasts. However, training with reforecast data leads to sharper but less robust ensemble. While LSTM models present a promising alternative to current operational models, further research is needed to fully exploit their potential, and to ensure applicability in an operational forecasting context.

To improve LSTM-based inflow forecasting, future research should investigate combining existing strategies with the framework developed in this study, such as: multi-model combination to leverage complementary model strengths, which has shown to improve ensemble forecast accuracy and reliability in all seasons (e.g., Armstrong et al., 2025) or multi-basin data aggregation to improve robustness and transferability (e.g., Kratzert et al., 2019a, 2019b).

While LSTM models hold great promise for inflow volume forecasting, continued advancements in forecast reliability, extreme event prediction, and operational integration are necessary to fully leverage their capabilities in real-world hydrological applications.

2.7 Acknowledgments

This study was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Alliance grant ALLRP 580910-22. The authors would also like to thank

Rio Tinto for sharing their hydrometeorological data on the Lac-Saint-Jean catchment and the European Center for Medium-Range Weather Forecasts (ECMWF) for providing access to the historical forecast data from their MARS computing and archiving facilities. In this study, the ERA5 reanalysis dataset produced by Hersbach et al., (2018) was used. It has been downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store:

<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>.

The base map in Figure 2.1 was created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.

CHAPITRE 3

DISCUSSION

L'utilisation des modèles LSTM en prévision hydrologique s'est largement répandue ces dernières années, eux qui ont connu un gain significatif en popularité. Ces modèles ont démontré leur fort potentiel dans diverses applications, surpassant souvent les modèles hydrologiques traditionnels encore utilisés dans des contextes opérationnels. Cependant, à notre connaissance, aucune étude n'a exploré la prévision volumétrique directe en s'appuyant sur des données de rétro-prévisions météorologiques, une lacune que cette étude vise à combler.

Les résultats obtenus confirment le fort potentiel des modèles LSTM pour la prévision de volumes cumulés sur une période de 14 jours, produisant des prévisions de bonne qualité. La précision des prévisions représente leur principal atout, tandis que leur fiabilité constitue leur principale faiblesse. L'intégration de données de rétro-prévisions météorologiques lors de l'entraînement des modèles a modifié les caractéristiques des prévisions émises : bien qu'elle ait amélioré leur précision et leur netteté, elle a également réduit leur fiabilité. Ce constat mitigé souligne que l'utilisation des rétro-prévisions doit être envisagée avec précaution, leur pertinence dépendant étroitement des objectifs spécifiques de la prévision hydrologique.

Cette section discutera en détail des résultats présentés et explorera des pistes d'amélioration susceptibles de compléter cette étude et d'enrichir les approches futures.

3.1 Analyse de la performance globale des modèles

La performance des modèles peut être évaluée à travers trois caractéristiques clés des prévisions d'ensemble qu'ils produisent: la précision, la netteté et la fiabilité. Ces attributs ainsi que plusieurs autres permettant de caractériser la qualité des prévisions sont définis dans Troin et al. (2021). La précision identifie dans quelle mesure les volumes prévus correspondent aux volumes observés. La netteté réfère à la concentration de l'ensemble des prévisions, traduisant

le degré de confiance intrinsèque du modèle. Une netteté élevée se manifeste par une distribution étroite des valeurs au sein de l'ensemble, signifiant que les prévisions sont regroupées dans un intervalle restreint. Cependant, il est important de souligner que des prévisions précises et nettes ne suffisent pas à garantir leur qualité. Si ces prévisions sont biaisées par rapport à la valeur cible, elles perdent leur utilité, car elles témoignent d'une confiance mal placée du modèle en des résultats incorrects. Enfin, la fiabilité évalue si la distribution de l'ensemble des prévisions reflète fidèlement les probabilités réelles des événements. Une répartition uniforme des observations au sein de l'ensemble indique une fiabilité élevée, car cela signifie que les prévisions englobent correctement les observations sans démontrer de biais marqué. Cette dimension est essentielle pour la viabilité opérationnelle des modèles, car les décideurs doivent pouvoir s'appuyer sur des prévisions fiables pour prendre des décisions éclairées, en particulier face à des événements extrêmes.

Dans cette étude, trois métriques de performance ont été utilisées pour évaluer les prévisions, soient le Kling-Gupta Efficiency (KGE), qui permet d'analyser la performance globale des modèles tout en étant utilisé pour déterminer la structure et les hyperparamètres optimaux lors de leur calibration, le Continuous Ranked Probability Score (CRPS), qui évalue à la fois la précision et la netteté des prévisions ainsi que les diagrammes de Talagrand, qui mesurent la fiabilité des prévisions d'ensemble. En combinant ces métriques, l'étude offre une évaluation complète des performances des modèles dans des contextes variés.

Les trois modèles LSTM développés produisent des prévisions de bonne qualité, atteignant un KGE minimal de 0,891 et un CRPS maximal de 180 hm^3 , une performance raisonnable compte tenu du volume observé moyen de 1 177 hm^3 sur la période de prévisions (écart relatif maximal d'environ 15% du volume observé moyen). Globalement, la courbe des volumes observés est bien reproduite, comme l'illustre le panneau a) des figures Figure 2.5, Figure 2.6 et Figure 2.7. Cependant, les panneaux b) et c) de ces figures mettent en évidence certaines lacunes qui limitent le rendement des modèles. Ces panneaux révèlent qu'un décalage temporel entre les courbes apparaît lors de la crue printanière, autour du mois d'avril. Ce décalage, qui peut atteindre une semaine pour le modèle E5 et varie de 1 à 3 jours pour les modèles RF et E5-RF,

entraîne une sous-estimation de la montée de la crue et une surestimation de sa descente. Bien que le décalage soit particulièrement marqué en période de crue, il persiste jusqu'à la fin de l'année.

Par ailleurs, les modèles rencontrent des difficultés à reproduire les volumes des périodes de plus faible hydraulicité associées à l'hiver et l'été. En hiver, où l'accumulation de neige réduit le ruissellement et l'infiltration, les volumes, déjà faibles, sont sous-estimés, suggérant que les modèles météorologiques prévisionnels sont probablement trop froids et ne génèrent pas assez de fonte de neige. De légères variations dans les conditions météorologiques des périodes hivernales d'entraînement et de prévision, subtiles mais qui font en sorte qu'elles sont statistiquement différentes, pourraient en partie expliquer le biais si les modèles ne s'adaptent pas correctement aux conditions changeantes. En hiver, la température moyenne est environ 1°C plus élevée et les précipitations totales moyennes environ 0,3 mm/jour supérieurs durant la période de prévision, ce qui représente une augmentation notable compte tenu de la précipitation totale moyenne d'environ 2 mm/jour. Les modèles pourraient ainsi éprouver de la difficulté à discerner l'augmentation des volumes résultant d'une fonte des neiges et de précipitations accrues.

À l'inverse, les volumes d'été, aussi de faible amplitude malgré que supérieurs à ceux d'hiver, sont majoritairement surestimés. Ce résultat est inattendu, car les modèles performent généralement mieux lors des périodes caractérisées par des débits usuels (c'est-à-dire de moyens à faibles), comme l'ont souligné Muluye et Coulibaly (2007) ainsi que Herbert et al. (2021). Étant largement exposés à des volumes d'une amplitude similaire durant l'entraînement, les modèles devraient en théorie être plus robustes et mieux adaptés aux conditions les générant. Une explication possible de ce biais de surestimation pourrait être que les modèles de prévision demeurent trop froids en été, sous-estimant ainsi l'évaporation et l'évapotranspiration réelles, ce qui pousse à la hausse les volumes simulés. Comme en hiver, les conditions météorologiques diffèrent statistiquement entre les périodes estivales d'entraînement et de prévision, ce qui pourrait avoir contribué au biais observé. Bien que ces différences soient une fois de plus relativement faibles — avec une température moyenne

supérieure d'environ $0,75^{\circ}\text{C}$ et des précipitations totales moyennes supérieures de 0,4 mm/jour durant la période de prévision (avec des précipitations totales moyennes d'environ 4 mm/jour) — elles peuvent néanmoins influencer les performances des modèles s'ils ne prennent pas correctement en compte ces écarts. Dans le cas présent, les modèles pourraient attribuer un poids excessif aux précipitations totales, conduisant à une surestimation du volume d'eau, alors que la hausse des températures devrait en partie compenser cet effet en favorisant une évaporation et une évapotranspiration accrues. Le biais de surestimation observé pourrait signifier que les modèles ne parviennent pas à tenir compte de ces processus compensatoires.

La sous-estimation des volumes hivernaux se maintient pour certaines années au niveau de la crue printanière, particulièrement pour le modèle E5. L'incapacité des modèles à simuler précisément ces pointes de crue peut s'expliquer par la variabilité des conditions hydrologiques et hydrauliques, ainsi que par les dynamiques liées au manteau neigeux qui caractérisent cette période. La sous-estimation des périodes de forte hydraulité est un problème courant rencontré par divers modèles d'IA (Muluye & Coulibaly, 2007; Sattari et al., 2012) dont les LSTM (Sabzipour et al., 2023).

Cette difficulté à prédire avec précision les volumes extrêmes, autant faibles qu'élevés, peut être attribuée à leur fréquence réduite, limitant l'exposition des modèles à ces événements lors de leur entraînement. Une meilleure gestion de cette hétérogénéité dans les données pourrait constituer une piste d'amélioration pour les modèles LSTM dans ce contexte. À la base, l'idéal est de concentrer les valeurs extrêmes de la variable cible dans la phase d'entraînement, de manière à ce que les modèles soient exposés à une échelle de valeurs étendue. La technique du suréchantillonnage, qui consiste à injecter artificiellement des valeurs peu représentées dans les données d'entraînement, serait également à considérer pour réduire l'impact de cette hétérogénéité.

La présence du décalage temporel et de la sous-estimation des volumes extrêmes est soulignée pour les années 2017 et 2021, mais ces phénomènes sont également observés pour la majorité des années incluses dans la période de prévisions (figures Figure 2.5a, Figure 2.6a et Figure

2.7a). Cependant, le constat est moins homogène en ce qui concerne les pointes de crue printanière, alors que les pointes plus faibles tendent à être surestimées, tandis que les pointes plus élevées de 2017, 2019 et 2022 sont généralement sous-estimées. Ces tendances suggèrent que les modèles ont du mal à différencier les conditions hydrologiques propres à chaque crue printanière, générant des volumes de pointe similaires, et ce indépendamment du scénario hydrométéorologique. Une exception notable est la crue printanière de 2021, dont le volume était nettement inférieur à celui des autres années mais qui a néanmoins été sous-estimée, comme le montre le panneau c) des figures Figure 2.5, Figure 2.6 et Figure 2.7. Cette année anormale a été marquée par des précipitations plus faibles combinées à des hausses fréquentes de température, réduisant ainsi l'accumulation de neige et, par conséquent, les apports venant de la fonte. Néanmoins, le panneau c) montre que les modèles ont su capter ces dynamiques inhabituelles du manteau neigeux, allant même jusqu'à surestimer légèrement la réduction du volume de la crue.

Pour tous les modèles, la valeur du KGE est plus élevée pour l'année de forte hydraulicité (Figures Figure 2.5b, Figure 2.6b et Figure 2.7b), tandis que l'année de faible hydraulicité (Figures Figure 2.5c, Figure 2.6c et Figure 2.7c) affiche un meilleur CRPS, ce qui pourrait sembler contradictoire. Toutefois, ces résultats s'expliquent par l'influence de la magnitude des volumes observés durant ces deux années, alors que des volumes plus élevés entraînent des valeurs absolues de CRPS plus importantes. Il est donc logique que les valeurs de CRPS soient gonflées pour l'année 2017, qui présente des volumes largement supérieurs à ceux de 2021. Par exemple, pour le modèle E5-RF, les CRPS de 219 hm^3 et 175 hm^3 correspondent respectivement à des erreurs relatives de 17% et 22% si l'on considère le volume moyen de 1298 hm^3 en 2017 et 805 hm^3 en 2021. Cela illustre que, bien que les valeurs absolues de CRPS soient plus élevées pour l'année 2017, ces erreurs sont proportionnelles à la magnitude des volumes observés. Par conséquent, les métriques d'évaluation suggèrent que les modèles tendent à mieux simuler les volumes plus élevés, pour lesquels ils obtiennent de meilleurs KGE et CRPS relativement à la magnitude des volumes observés.

3.2 Analyse de la performance des modèles sur une base saisonnière

Il est essentiel d'évaluer la performance des modèles en fonction des différentes saisons afin de mieux cerner leur comportement, leurs points forts et leurs faiblesses tout au long de l'année. Cette analyse saisonnière revêt une importance particulière dans le contexte des bassins versants nordiques, comme celui étudié ici, qui subissent des variations hydrométéorologiques et hydrauliques significatives d'une saison à l'autre. En hiver, les précipitations tombent principalement sous forme de neige, et les débits demeurent faibles en raison de l'accumulation de neige et du sol gelé, limitant le ruissellement et l'infiltration. Au printemps, la fonte des neiges, combinée à des précipitations pluvieuses, provoque une crue printanière marquée, caractérisée par des débits élevés et une grande variabilité hydrologique. L'été se distingue par une stabilisation des débits, dans un contexte de températures élevées favorisant une évapotranspiration importante. Enfin, l'automne est marqué par une augmentation des précipitations pluvieuses et des températures plus modérées, ce qui réduit l'évapotranspiration. Cette variabilité saisonnière impose des défis spécifiques aux modèles, soulignant la nécessité d'une analyse approfondie pour évaluer leur capacité à représenter fidèlement les processus hydrologiques propres à chaque saison.

Les modèles produisent des prévisions avec une précision comparable dans l'ensemble, tout en affichant des variations notables selon les saisons. Le modèle RF se démarque par sa performance supérieure en hiver et au printemps, tandis que le modèle E5 surpasse ses homologues en été et à l'automne. Le modèle E5-RF, quant à lui, offre une performance globalement équilibrée, sans exceller ni sous-performer de manière significative, à l'exception de l'été, où il montre une nette infériorité. Pour l'hiver, qui est généralement plus facile à modéliser car l'état d'une journée est fortement dépendant de la précédente, et le printemps, qui est davantage complexe à simuler en raison des dynamiques neigeuses, les modèles entraînés avec des données de rétro-prévisions, et particulièrement le modèle RF, produisent des résultats supérieurs (Figure 2.8a et Figure 2.8b). Les difficultés éprouvées en hiver démontrent une faiblesse des modèles LSTM, qui ne possèdent pas de loi physique exprimant le poids du débit de récession transmis à la prochaine journée, malgré une exposition à des

situations similaires en entraînement. Malgré ces performances, les prévisions printanières restent généralement moins précises pour toutes les approches, comme en témoignent les valeurs plus élevées du CRPS. Ce résultat peut être attribué à la difficulté de simuler la crue printanière, caractérisée par des processus hydrologiques complexes liés à l'enneigement et à des conditions hydrométéorologiques très variables. Un autre facteur à considérer est la forte hydraulité associée au printemps dans le bassin du Lac-Saint-Jean. Cette saison est marquée par les débits les plus élevés et, par conséquent, par les volumes les plus importants, ce qui tend à gonfler les valeurs absolues de CRPS en comparaison aux autres saisons. L'impact de la proportionnalité des volumes se manifeste également en hiver, mais dans le sens inverse, où les débits plus faibles réduisent naturellement les valeurs absolues de CRPS. Ces résultats soulignent l'importance d'adopter une évaluation saisonnière pour interpréter les performances des modèles avec nuance et précision.

La fiabilité des prévisions est problématique pour tous les modèles, en particulier pour RF et E5-RF, et varie également selon les saisons. Les prévisions estivales et automnales sont relativement plus fiables que celles réalisées pour l'hiver et le printemps, ces dernières étant reconnues pour leur complexité en raison de processus physiques liés à la neige. Malgré cette amélioration saisonnière, les distributions des ensembles restent éloignées de la distribution uniforme caractéristique d'un ensemble parfaitement fiable. Toutes les saisons montrent des prévisions sous-dispersées, comme l'ont également observé Armstrong et al. (2025). Cette sous-dispersion est mise en évidence par les diagrammes en forme de U présentés à la Figure 2.9.

Le modèle E5 présente un biais de sous-estimation particulièrement marqué (forme en L) durant l'hiver et le printemps, indiquant une faible précision : l'ensemble de prévisions englobe rarement l'observation, malgré une dispersion relativement large. Ce constat est cohérent avec les observations précédentes, selon lesquelles les prévisions hivernales et printanières sont moins précises. Les figures Figure 2.5, Figure 2.6 et Figure 2.7 illustrent également que la variabilité des ensembles semble proportionnelle aux volumes observés, avec une variabilité faible en hiver et élevée au printemps. Par conséquent, bien que le biais hivernal soit notable,

il pourrait être moindre comparé à celui du printemps, car la moindre variabilité hivernale contribue moins à ce qu'un ensemble biaisé couvre l'observation. En été et en automne, les prévisions sont nettement plus fiables, particulièrement pour le modèle E5. Cette amélioration pourrait s'expliquer par une dispersion plus importante des ensembles, combinée à des conditions hydrométéorologiques plus stables, ce qui facilite la modélisation durant ces saisons.

Pour remédier à l'enjeu de sous-dispersion, l'application de techniques éprouvées devrait être envisagée, comme le post-traitement des prévisions (Dion et al., 2021) ou la combinaison de plusieurs modèles (Arsenault et al., 2015; Armstrong et al., 2025). Ces approches pourraient contribuer à améliorer la fiabilité des prévisions et à rapprocher les distributions des ensembles d'une configuration optimale.

3.3 Utilité d'intégrer des rétro-prévisions durant l'entraînement des modèles

Des rétro-prévisions ont été utilisées lors de la calibration de certains modèles, une approche à notre connaissance inexplorée dans la sphère de la prévision de volumes d'eau, afin d'évaluer l'apport de telles données sur la performance des modèles et en cerner la pertinence. En étendant artificiellement la période de calibration grâce à des données supplémentaires, cette méthode offrait le potentiel d'améliorer les performances des modèles, comme suggéré par Ayzel et al. (2021). Un autre avantage de ce produit météorologique est qu'il emploie le même modèle prévisionnel que les prévisions opérationnelles, permettant un ajustement intrinsèque aux biais du modèle de prévision durant l'entraînement. Les résultats obtenus montrent que l'utilisation des rétro-prévisions influence les prévisions générées, apportant à la fois des avantages et des inconvénients.

Les courbes de volumes présentées (Figures Figure 2.5, Figure 2.6 et Figure 2.7) révèlent que les prévisions émises par le modèle entraîné sans rétro-prévisions, soit le modèle E5, affichent une variabilité nettement plus élevée en comparaison à celles des modèles intégrant les rétro-prévisions en entraînement, soit RF et E5-RF. Cette variabilité accrue pourrait s'expliquer par

les particularités de l'entraînement du modèle E5. En effet, le modèle E5 s'appuie exclusivement sur des observations issues de la réanalyse ERA5, ce qui peut limiter sa robustesse face à l'incertitude inhérente au processus prévisionnel. L'absence de données additionnelles représentant des scénarios variés et incertains pourrait restreindre la capacité du modèle à s'adapter efficacement à des conditions atypiques ou extrêmes. L'entraînement avec des rétro-prévisions, générées par le même modèle météorologique que les prévisions utilisées en opération, permet au modèle d'être exposé à une gamme élargie de scénarios prévisionnels. Cela peut renforcer sa capacité à interpréter et à anticiper divers contextes hydrométéorologiques, améliorant ainsi sa stabilité et réduisant la variabilité de ses prévisions.

Les modèles entraînés avec des données de rétro-prévisions, en particulier RF, démontrent une précision accrue en hiver et au printemps, comme en témoigne le Tableau 2.4. Cette supériorité saisonnière suggère que l'intégration des rétro-prévisions dans l'entraînement enrichit les modèles d'une compréhension plus étendue des conditions hydrométéorologiques variées. Cela leur permet de mieux s'adapter à la faible hydraulité hivernale et à la forte hydraulité printanière. Les rétro-prévisions, de par leur nature répétitive, exposent les modèles à une diversité accrue de scénarios météorologiques, incluant des événements rares (Hamill et al., 2013) susceptibles de produire des débits extrêmes. La capacité à reproduire adéquatement les volumes printaniers est particulièrement précieuse dans le contexte du bassin versant à l'étude, où une prévision précise de la fonte printanière est cruciale pour optimiser la gestion hydroélectrique et prévenir les inondations. Cependant, le Tableau 2.4 montre également qu'E5 surpassé les modèles entraînés avec des rétro-prévisions durant l'été et, de manière encore plus marquée, en automne. L'entraînement plus simple d'E5, basé uniquement sur des pseudos-observations, semble ainsi suffisant pour capturer les dynamiques hydrologiques moins complexes de ces périodes. En résumé, l'entraînement avec des rétro-prévisions offre un avantage considérable dans la gestion des saisons aux conditions extrêmes, mais peut engendrer une légère perte de performance dans les périodes plus stables où les processus sont moins complexes.

La Figure 2.9 met en évidence une réduction de la fiabilité des prévisions pour les modèles entraînés avec des rétro-prévisions, alors qu'E5 se distingue par une fiabilité supérieure en toutes saisons. Cette supériorité est particulièrement notable en été et en automne, où les prévisions d'E5 affichent une distribution plus proche de l'uniformité, contrastant avec les fortes surestimations observées chez RF et E5-RF. La meilleure fiabilité d'E5 peut être expliquée par la grande variabilité de son ensemble, qui, bien qu'associée à un biais, parvient à englober en partie les volumes observés. En revanche, les prévisions plus nettes de RF et E5-RF amplifient leur sous-dispersion. La netteté accentue l'effet des biais : la majorité des membres des ensembles reproduisent le biais au lieu de refléter la variabilité des conditions observées, ce qui compromet leur fiabilité.

Au printemps, les trois modèles présentent des résultats décevants en raison des difficultés à simuler la crue printanière. Les biais de prévision observés dans la Figure 2.9 (panneaux d, e et f) s'expliquent en partie par le délai identifié dans la simulation de la crue printanière. Ce décalage engendre une sous-estimation des volumes lors de la montée, reflétée par la partie droite des diagrammes et une surestimation lors de la descente, reflétée par la partie gauche des diagrammes. La crue printanière occupant une grande portion de la saison, ces biais influencent fortement les indices de fiabilité pour cette période.

Le bilan sur la pertinence d'intégrer les rétro-prévisions à l'entraînement des modèles est donc mitigé. Pour la prévision des périodes d'hydraulique extrême, comme les faibles débits hivernaux et les pics de crue printanière, les rétro-prévisions apportent une valeur ajoutée significative. Cependant, elles tendent à dégrader la qualité des prévisions estivales et automnales en tout point. Finalement, les modèles entraînés avec des rétro-prévisions génèrent des ensembles dont la variabilité est fortement réduite. Cette caractéristique peut être avantageuse dans certaines circonstances, mais peut également compromettre leur utilité lorsque l'objectif principal est la fiabilité et que les prévisions présentent régulièrement des biais. En somme, l'intégration des rétro-prévisions doit être envisagée en fonction des priorités spécifiques de l'application. Si l'objectif principal est de mieux gérer les périodes d'hydraulique extrême, cette approche est justifiée et même recommandée. Toutefois, pour des

besoins axés sur la fiabilité générale ou la performance dans des conditions hydrologiques stables, leur utilisation pourrait s'avérer contre-productive.

3.4 Limitations rencontrées et améliorations envisagées

Cette section explore les différentes limitations de l'étude et propose des axes d'amélioration. Elle examine notamment l'intégration de nouvelles variables comme intrants du modèle, la focalisation sur un unique bassin versant, le choix de la fonction objectif et des hyperparamètres du modèle, ainsi que l'éventuelle comparaison avec un modèle opérationnel.

3.4.1 Ajout de variables en tant qu'intrants au modèle

D'abord, les modèles LSTM ont été entraînés à partir de huit variables météorologiques, soit le maximum disponible en fonction de ce qui était partagé par les jeux de données météorologiques utilisés en entraînement et en prévision. Ce nombre pourrait être élargi en intégrant d'autres variables disponibles dans les ensembles météorologiques telles que la séparation des précipitations en précipitations liquides (pluie) et solides (neige). Ces variables additionnelles permettraient aux modèles de capturer de nouvelles tendances et relations entre les conditions météorologiques et les volumes d'eau.

En parallèle, l'intégration de prédicteurs hydrologiques, comme les débits ou volumes récemment observés, pourrait représenter une avancée notable, aidant le modèle à reconnaître l'état actuel du bassin versant, ce qui pourrait se traduire par une réduction des erreurs de prévision. Sabzipour et al. (2023) ont démontré l'efficacité de cette approche, utilisant les débits des jours précédents la prévision comme intrants pour des prévisions de débits sur neuf jours. Toutefois, l'impact de tels prédicteurs hydrologiques serait probablement moindre sur des prévisions de volumes à moyen terme que sur des débits à court terme, car l'influence de l'état initial du bassin s'estompe avec la durée de prévision. De plus, le gain en précision pourrait se dissiper au cours du processus d'agrégation des débits en volume cumulatif, qui constitue une sortie de modèle plus brute qu'un débit ponctuel.

3.4.2 Étude portant sur un seul bassin versant

Aussi, il est important de noter que cette étude porte sur un seul bassin versant et que ce qui en est retiré est donc applicable à ce seul espace, qui possède ses propres caractéristiques hydrologiques, météorologiques et géologiques. Afin de confirmer l'adaptabilité de la méthodologie et d'étendre les conclusions de l'étude, d'autres bassins devraient être explorés dans de futures recherches pour vérifier si des performances similaires sont obtenues. Ces recherches pourraient de nouveau s'attaquer à des bassins nordiques ou bien envisager des bassins soumis à des conditions hydroclimatiques différentes afin de mettre à l'épreuve les modèles dans divers scénarios.

Il est essentiel de souligner que cette étude s'est concentrée sur un seul bassin versant, limitant ainsi l'applicabilité des conclusions aux spécificités hydrologiques, météorologiques et géologiques de ce dernier. Afin de confirmer la généralisation de la méthodologie employée et d'élargir les conclusions, il serait pertinent de mener des recherches sur d'autres bassins versants. Ces études pourraient cibler des bassins nordiques similaires ou explorer des bassins soumis à des conditions hydroclimatiques distinctes, mettant ainsi à l'épreuve les modèles dans divers scénarios. Une telle démarche contribuerait à évaluer la robustesse et l'adaptabilité des modèles dans des contextes variés.

L'utilisation des rétro-prévisions en prévision hydrologique demeure encore peu explorée. L'accès limité à ces données et leur documentation restreinte en font un outil novateur mais encore largement inexploité. Cette étude visait à faire une première évaluation de leur potentiel en prévision volumétrique. Or, dans le but de maximiser la performance des modèles LSTM, il aurait été avantageux d'étendre l'entraînement à des données issues de plusieurs bassins versants. Des travaux récents (Kratzert et al., 2024) ont démontré que les LSTM offrent des performances supérieures lorsqu'ils sont entraînés sur un ensemble varié de données couvrant un grand nombre de bassins.

3.4.3 Fonction objectif

Les résultats de cette étude dépendent fortement de la fonction objectif utilisée pour calibrer et configurer les modèles LSTM. En effet, cette fonction, en orientant l'entraînement des modèles, influence directement leurs performances et les conclusions qui en découlent. Dans cette étude, le Kling-Gupta Efficiency (KGE) a été choisi comme fonction objectif, car il évalue globalement la performance des prévisions, permettant de statuer sur le potentiel général de l'approche adoptée. Cependant, il est évident que l'utilisation d'une autre fonction objectif aurait pu conduire à des résultats différents. Par exemple, une fonction objectif qui accorde un poids plus élevé aux pointes de volumes mettrait l'accent sur la réduction des erreurs lors des événements critiques, mais pourrait potentiellement diminuer la performance globale sur les périodes moins dynamiques. Pour des études futures souhaitant cibler des aspects spécifiques de la prévision hydrologique, il serait pertinent d'explorer des fonctions objectif adaptées à ces priorités.

3.4.4 Hyperparamètres

Le choix des hyperparamètres dans les modèles LSTM joue un rôle critique dans leurs performances. Dans cette étude, une méthode itérative d'essais-erreurs a été utilisée pour ajuster les hyperparamètres, permettant d'identifier des configurations performantes, mais pas nécessairement optimales. Cette approche, bien que méthodique, repose sur des ajustements manuels limités par le temps et les ressources, ce qui peut restreindre l'exploration de l'espace des hyperparamètres. L'intégration d'algorithmes d'optimisation automatisés pourrait potentiellement offrir une meilleure exploration et identification des hyperparamètres optimaux. Or, compte tenu des nombreux tests effectués pour établir des plages raisonnables d'hyperparamètres, qui ont ensuite été évalués sous diverses configurations, cette amélioration potentielle doit être mise en perspective. Malgré tout, il demeure pertinent d'analyser la valeur ajoutée des logiciels d'optimisation à la méthode d'essais-erreurs.

3.4.5 Comparaison avec un modèle opérationnel

Cette étude s'est concentrée exclusivement sur les modèles de type LSTM. Ce choix méthodologique visait à maintenir une cohérence dans l'évaluation des modèles, tout en utilisant un modèle de référence (non exposé aux rétro-prévisions) pour des comparaisons internes. Cependant, l'absence de comparaison avec un modèle opérationnel limite la portée des conclusions dans un contexte pratique. Pour évaluer l'intérêt opérationnel de la méthodologie développée, il serait pertinent de procéder à une application en cas opérationnel puis de comparer les performances des modèles LSTM avec celles d'un modèle déjà utilisé dans un cadre opérationnel.

CONCLUSION

En conclusion, rappelons que ce travail avait pour objectif de réaliser des prévisions de volumes d'eau cumulés sur 14 jours pour le bassin versant du Lac-Saint-Jean, au Canada, en utilisant des modèles LSTM. Elle visait également à évaluer l'apport des rétro-prévisions météorologiques dans l'entraînement des modèles, dans le but d'améliorer leurs performances et, par conséquent, la qualité des prévisions. Trois modèles LSTM ont été développés, chacun utilisant différentes combinaisons de données d'entrée, avant d'être testés sur des prévisions météorologiques réelles.

Les résultats obtenus démontrent le potentiel prometteur des modèles LSTM pour la prévision volumétrique à moyen terme. Ces modèles produisent des prévisions de bonne qualité tout en mettant en lumière des bénéfices et des limites spécifiques. Les principales conclusions de l'étude menée sont les suivantes :

- 1) Pendant la période de calibration, les modèles affichent des performances élevées, avec un KGE minimal de 0,910, et reproduisent efficacement les volumes de grande magnitude.
- 2) Bien que précises, les prévisions d'ensemble des modèles manquent de fiabilité et présentent une sous-dispersion constante, indépendamment de la saison.
- 3) Tous les modèles éprouvent des difficultés à reproduire le moment et l'ampleur de la crue printanière, échouant souvent à prévoir correctement les débits extrêmes.
- 4) L'intégration des rétro-prévisions durant l'entraînement améliore la précision et la netteté des ensembles, mais diminue leur robustesse. Ainsi, leur utilité en contexte opérationnel dépend des objectifs spécifiques de prévision, comme la gestion des crues ou des conditions hydriques extrêmes.

Ainsi, les résultats de cette étude indiquent que les modèles LSTM constituent une alternative prometteuse aux approches opérationnelles actuelles pour la prévision volumétrique, bien qu'une marge d'amélioration subsiste. Des travaux complémentaires sont nécessaires pour

comprendre comment tirer pleinement parti de leurs capacités et garantir leur applicabilité et transférabilité dans un contexte opérationnel de prévision.

Une évaluation des modèles sur une base saisonnière a permis de détecter des variations de performance en fonction des saisons, malgré un bon rendement annuel global. Cette analyse met en évidence que les modèles ne sont pas tous aussi performants au cours d'une même saison, et offre un aperçu précieux sur l'impact des caractéristiques spécifiques à chaque saison. Cela revêt une importance particulière dans le contexte d'un bassin versant nordique, où les conditions hydroclimatiques peuvent varier considérablement tout au long de l'année.

En termes de précision, les modèles rencontrent plus de difficultés en hiver et au printemps, avec des écarts importants proportionnellement aux volumes observés. Cette baisse de performance pourrait être attribuée aux conditions hydrologiques particulièrement variables de ces saisons, ainsi qu'à la présence de neige, des facteurs connus pour complexifier la modélisation hydrologique. Les modèles basés sur les rétro-prévisions surpassent celui de référence pour ces deux saisons. L'inverse est constaté en été et en automne, pour lesquelles la précision est meilleure et le modèle de référence domine. Concernant la fiabilité, les modèles éprouvent toujours des difficultés en hiver et au printemps, sous-estimant notamment les débits hivernaux et rencontrant des difficultés à reproduire l'amplitude et le moment de la crue printanière. Une sous-dispersion marquée est observée pour tous les modèles, comme l'illustre la Figure 2.9. Cependant, le modèle de référence E5 se distingue par sa robustesse, en particulier en été et en automne.

Les efforts futurs devraient se concentrer sur les faiblesses identifiées de l'approche actuelle, notamment la fiabilité des prévisions à travers toutes les saisons et la précision des prévisions des débits extrêmes. Pour améliorer ces aspects spécifiques ainsi que les performances globales des prévisions LSTM, les recherches pourraient intégrer des stratégies éprouvées telles que la combinaison des sorties de modèles hydrologiques variés (comme proposé par Armstrong et al., 2025) ou l'agrégation de données provenant de plusieurs bassins pour l'entraînement des modèles LSTM (comme décrit par Kratzert et al., 2019a, 2019b).

RECOMMANDATIONS

La portée du travail réalisé est contrainte par plusieurs limitations dont les plus importantes ont été soulignées précédemment. En se basant sur ces limitations, des recommandations spécifiques sont proposées pour orienter de futures recherches visant à améliorer la qualité des prévisions générées par des modèles hydrologiques. Ces recommandations visent également au progrès de la prévision hydrologique, un domaine où des opportunités d'avancement demeureront toujours présentes.

- 1) Combiner la méthodologie avec des méthodes éprouvées, telles que l'application d'un post-traitement (Zhao et al., 2011; Madadgar et al., 2014), l'adoption d'une approche multi-modèle (Zhang et al., 2020) ou même la combinaison de ces deux techniques (Xu et al., 2019), pourrait améliorer la performance des modèles.
- 2) Étendre le jeu de données afin de stimuler l'entraînement des modèles. En exposant les modèles à une période historique plus longue, ces derniers seraient confrontés à une plus grande variété de conditions hydroclimatiques, renforçant leur capacité d'adaptation. De plus, l'utilisation de données plus récentes permettrait d'améliorer la qualité de l'information tout en assurant une meilleure compatibilité avec les produits météorologiques actuels.
- 3) Utiliser des prévisions météorologiques de divers modèles, par exemple celui du ECMWF, le GFS ou le GDPS. Une telle diversité permettrait d'analyser l'impact des jeux de données sur la performance des modèles et d'identifier ceux générant les meilleurs résultats.
- 4) Considérer plusieurs bassins versants aux caractéristiques variées. En élargissant l'échantillon à des bassins divergents sur les plans hydrologiques, climatiques, topographiques, géologiques et géographiques, il serait possible de tester l'adaptabilité des modèles dans des contextes diversifiés.
- 5) Comparer la performance des modèles avec celle des modèles couramment utilisés dans des applications opérationnelles de prévision hydrologique pour évaluer leur potentiel en contexte opérationnel.

- 6) Augmenter la durée de prévision vers la prévision saisonnière. En allongeant la période de prévision pour viser une approche saisonnière, il devient possible d'optimiser l'atteinte des objectifs opérationnels, qu'il s'agisse de la gestion des niveaux d'eau, de la production énergétique ou de la préservation environnementale (Turner et al., 2017 ; Ashbolt et al., 2018 ; Peñuela et al., 2020). Les jours de prévision non couverts par les rétro-prévisions pourraient être remplis par la climatologie.

Ces éléments visent à combler les lacunes identifiées dans ce travail et à orienter les recherches vers des avenues prometteuses pour améliorer la performance globale des modèles hydrologiques et, à terme, assurer une gestion saine et durable des ressources en eau.

ANNEXE I

FONCTIONS DE DENSITÉ CUMULATIVE DES VOLUMES OBSERVÉS DURANT LA CALIBRATION DES MODÈLES

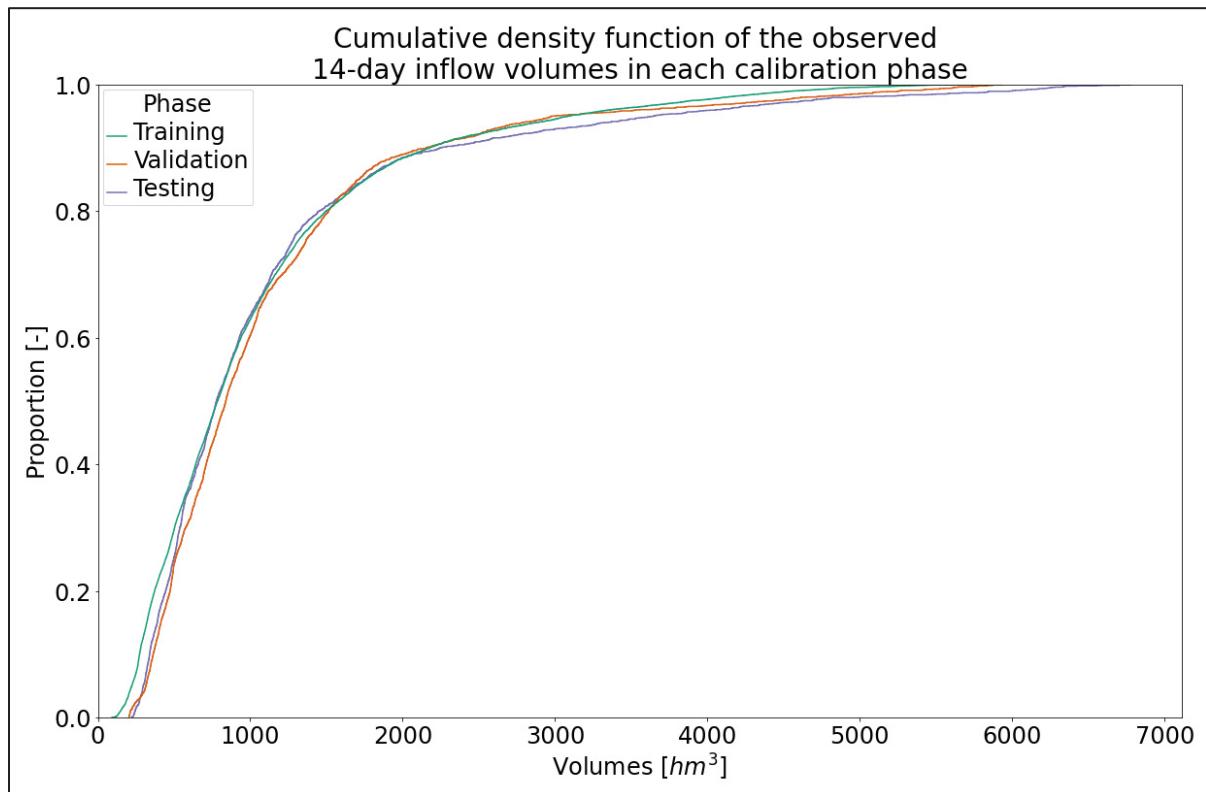


Figure-A I-1 Distribution des volumes observés cumulatifs sur 14 jours pour chaque phase de calibration du modèle E5. La phase d'entraînement s'étend du 18/12/1978 au 17/12/2009, celle de validation du 18/12/2009 au 17/12/2016 et celle de test du 18/12/2016 au 18/12/2023

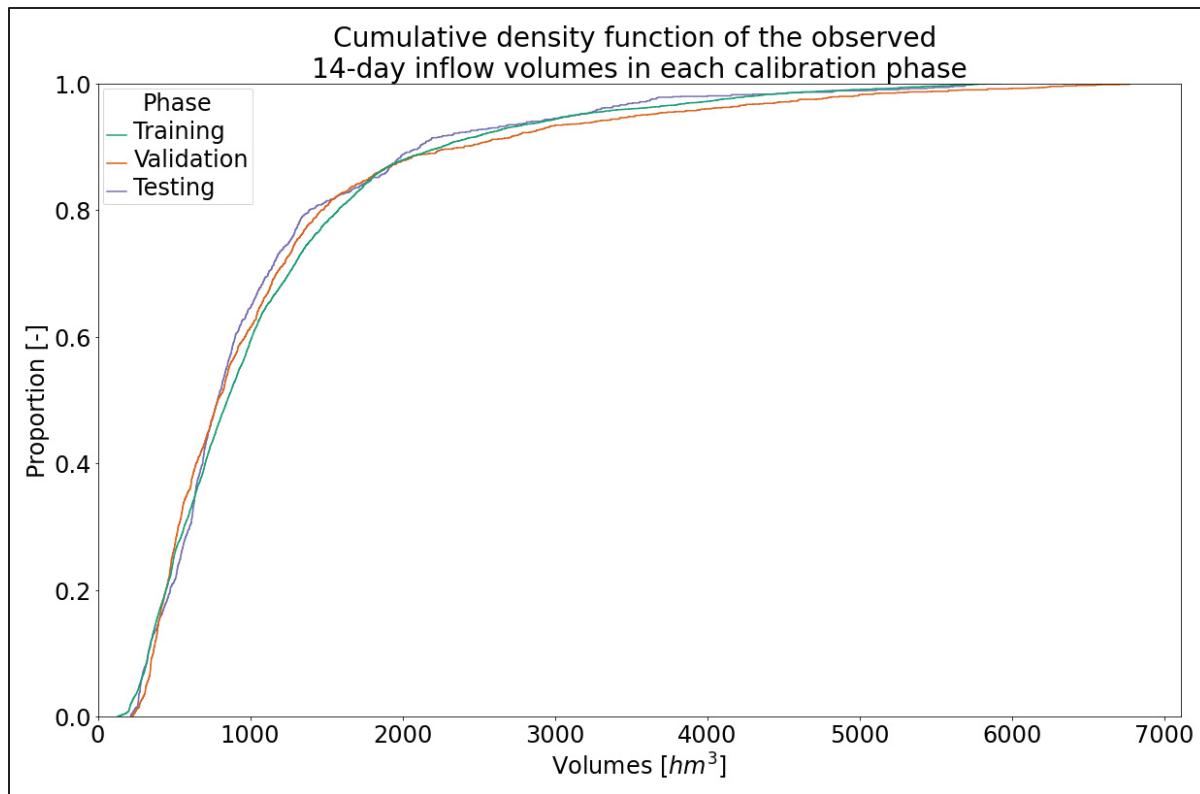


Figure-A I-2 Distribution des volumes observés cumulatifs sur 14 jours pour chaque phase de calibration du modèle RF. La phase d'entraînement s'étend de mars 1996 à novembre 2015 (excluant 2007 à 2009), celle de validation de novembre 2015 à avril 2022 et celle de test de 2007 à 2009

ANNEXE II

STRUCTURE OPTIMALE DES MODÈLES

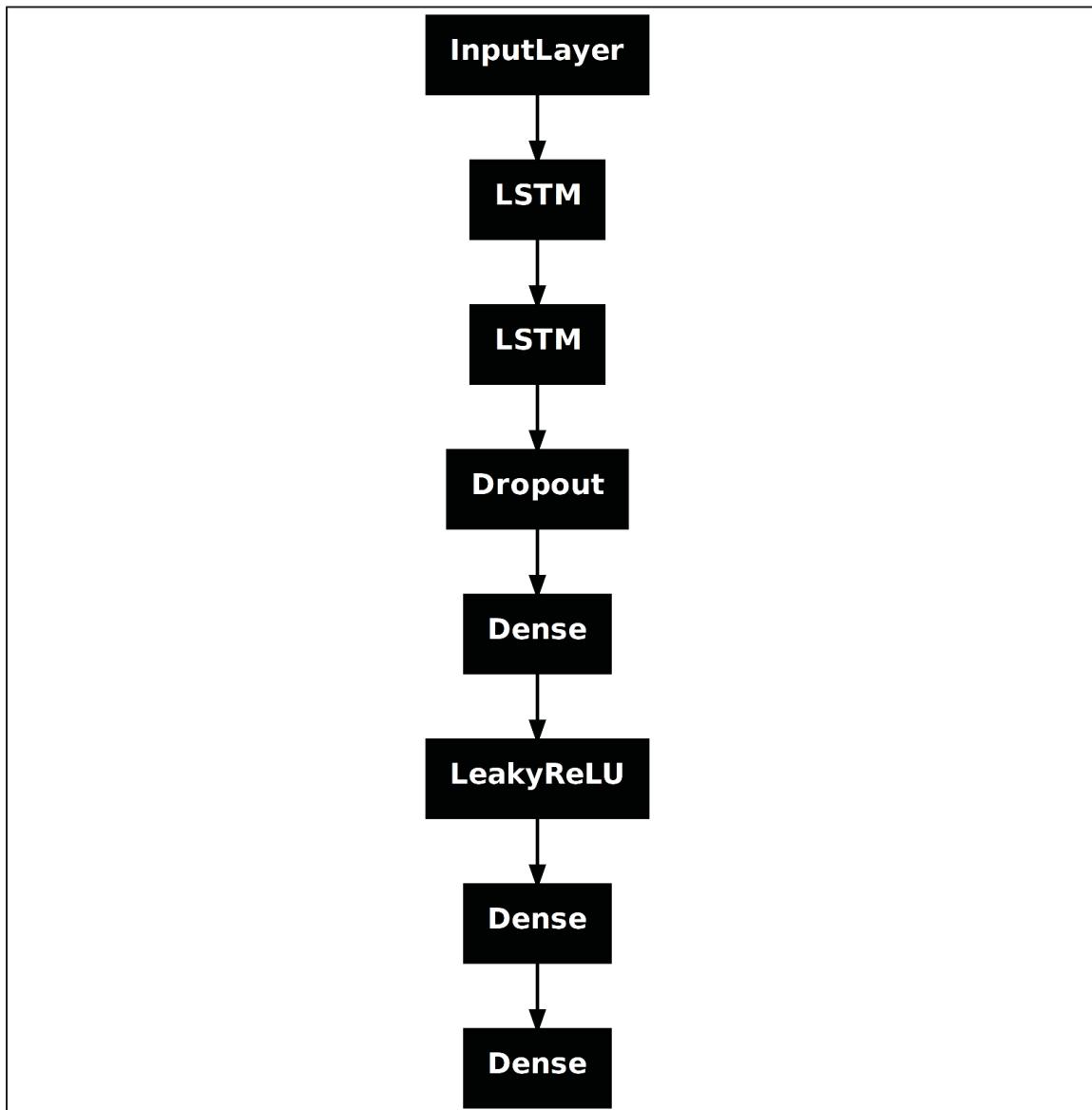


Figure-A II-1 Schéma de la structure utilisée pour les modèles de cette étude. La couche d'entrée est de dimension (365,8) correspondant aux 365 jours consécutifs de huit variables météorologiques alors que la couche de sortie est de dimension (1,) correspondant à l'unique simulation d'un volume. Tous les modèles possèdent cette structure, mais certains hyperparamètres (nombre d'unités LSTM ou de neurones Dense) diffèrent. Les valeurs d'hyperparamètres sont présentées dans le Tableau 2.2 pour la version finale de chaque modèle

ANNEXE III

PRÉVISIONS DE VOLUMES GÉNÉRÉES PAR LES MODÈLES DURANT LA PÉRIODE D'ENTRAINEMENT

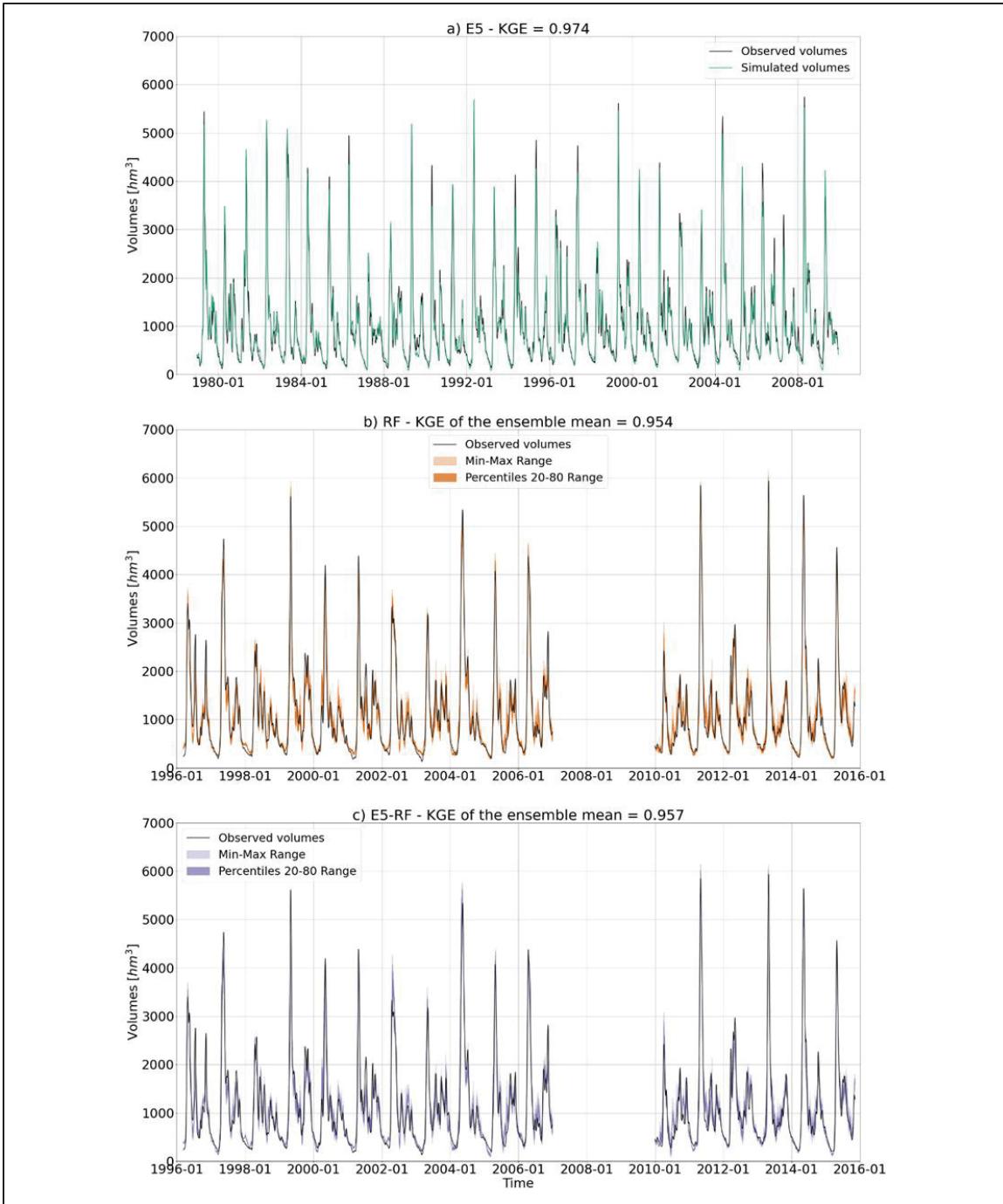


Figure-A III-1 Prévisions des volumes cumulatifs sur 14 jours pour le bassin versant du LSJ pour tous les modèles sur la période d'entraînement : E5 (a), RF (b) et E5-RF (c). La phase d'entraînement s'étend de décembre 1978 à décembre 2009 pour E5 et de mars 1996 à novembre 2015 pour RF (à l'exception de la période 2007-2009, illustrée par l'espace blanc). E5-RF a été entraîné sur les deux ensembles de données d'entraînement, mais seules les prévisions de la dernière phase d'entraînement sont présentées ici pour des raisons de représentativité (plus rapprochées de l'application en prévision)

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Abaza, M., Anctil, F., Fortin, V., & Perreault, L. (2017). Hydrological Evaluation of the Canadian Meteorological Ensemble Reforecast Product. *Atmosphere-Ocean*, 55(3), 195–211.
- Abhishek, K., Singh, M. P., Ghosh, S., & Anand, A. (2012). Weather Forecasting Model using Artificial Neural Network. *Procedia Technology*, 4, 311-318.
- Ahmed, S., Coulibaly, P., & Tsanis, I. (2015). Improved Spring Peak-Flow Forecasting Using Ensemble Meteorological Predictions. *Journal of Hydrologic Engineering*, 20(2), 04014044.
- Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier, D. P. (2016). Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resources Research*, 52(6), 4209–4225.
- Apaydin, H., Feizi, H., Sattari, M. T., Colak, M. S., Shamshirband, S., & Chau, K.-W. (2020). Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting. *Water*, 12(5), 1500.
- Armstrong, W., Arsenault, R., Martel, JL., Troin, M., Dion, P., Sabzipour, B., Brissette, F., & Mai, J. (2025) A Hybrid Multi-Model Approach for Hydrological Ensemble Forecasting: The Added Value of Deep Learning Models. *Hydrological Sciences Journal*. Sous presse.
- Arsenault, R., Gatien, P., Renaud, B., Brissette, F., & Martel, J.-L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology*, 529, 754-767.
- Arsenault, R., Latraverse, M., & Duchesne, T. (2016). An Efficient Method to Correct Under-Dispersion in Ensemble Streamflow Prediction of Inflow Volumes for Reservoir Optimization. *Water Resources Management*, 30(12), 4363–4380.
- 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.
- 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.

- Ashbolt, S. C., & Perera, B. J. C. (2018). Multiobjective Optimization of Seasonal Operating Rules for Water Grids Using Streamflow Forecast Information. *Journal of Water Resources Planning and Management*, 144(4), 05018003.
- Ávila, L., Silveira, R., Campos, A., Rogiski, N., Freitas, C., Aver, C., & Fan, F. (2023). Seasonal Streamflow Forecast in the Tocantins River Basin, Brazil : An Evaluation of ECMWF-SEAS5 with Multiple Conceptual Hydrological Models. *Water*, 15(9), 1695.
- Ayzel, G., & Heistermann, M. (2021). The effect of calibration data length on the performance of a conceptual hydrological model versus LSTM and GRU: A case study for six basins from the CAMELS dataset. *Computers & Geosciences*, 149, 104708.
- Bai, Y., Wang, P., Xie, J., Li, J., & Li, C. (2015). Additive Model for Monthly Reservoir Inflow Forecast. *Journal of Hydrologic Engineering*, 20(7), 04014079.
- Bai, Y., Chen, Z., Xie, J., & Li, C. (2016). Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of Hydrology*, 532, 193–206.
- Bai, P., Liu, X., & Xie, J. (2021). Simulating runoff under changing climatic conditions: A comparison of the long short-term memory network with two conceptual hydrologic models. *Journal of Hydrology*, 592, 125779.
- Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567), 47-55.
- Bazile, R., Boucher, M.-A., Perreault, L., & Leconte, R. (2017). Verification of ECMWF System 4 for seasonal hydrological forecasting in a northern climate. *Hydrology and Earth System Sciences*, 21(11), 5747-5762.
- Beven, K. (2012). Rainfall-Runoff Modelling : The Primer. Dans *Rainfall-Runoff Modelling : The Primer : Second Edition* (Vol. 15).
- Bogner, K., & Kalas, M. (2008). Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. *Atmospheric Science Letters*, 9(2), 95-102.
- Boucher, M.-A., Anctil, F., Perreault, L., & Tremblay, D. (2011). A comparison between ensemble and deterministic hydrological forecasts in an operational context. *Advances in Geosciences*, 29, 85-94.
- 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.
- Chen, J., Brissette, F. P., & Leconte, R. (2011). Uncertainty of downscaling method in

- quantifying the impact of climate change on hydrology. *Journal of Hydrology*, 401(3), 190-202.
- Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., & Pain, C. C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590, 125376.
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., ... Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, 44(12).
- Damavandi, H. G., Shah, R., Stampoulis, D., Wei, Y., Boscovic, D., & Sabo, J. (2019). Accurate prediction of streamflow using long short-term memory network: A case study in the Brazos river basin in Texas. *International Journal of Environmental Science and Development*, 10(10), 294-300.
- Darbansari, P., & Coulibaly, P. (2020). Inter-comparison of lumped hydrological models in data-scarce watersheds using different precipitation forcing data sets: Case study of Northern Ontario, Canada. *Journal of Hydrology: Regional Studies*, 31, 100730.
- Davidson-Chaput, J., Arsenault, R., Martel, J.L., & Troin, M. (2025). Value of various elements of the hydrological forecasting chain: is there a successful pathway for improving the overall performance? *Journal of Hydrology*, En révision.
- Demirel, M. C., Booij, M. J., & Hoekstra, A. Y. (2013). Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resources Research*, 49(7), 4035-4053.
- Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia*, 4, 1001-1007.
- Dion, P., Martel, J.-L., & Arsenault, R. (2021). Hydrological ensemble forecasting using a multi-model framework. *Journal of Hydrology*, 600, 126537.
- Ehrendorfer, M. (1997). Predicting the uncertainty of numerical weather forecasts: A review. *METEOROLOGISCHE ZEITSCHRIFT-BERLIN-*, 6, 147-183.
- Fan, F. M., Schwanenberg, D., Alvarado, R., Assis dos Reis, A., Collischonn, W., & Naumann, 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.
- Gauch, M., Mai, J., & Lin, J. (2021). The proper care and feeding of CAMELS: How limited

training data affects streamflow prediction. *Environmental Modelling & Software*, 135, 104926.

Giriagama, 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.

Hagedorn, R. (2008). *Using the ECMWF reforecast dataset to calibrate EPS forecasts*.

Hamill, T. M., & Colucci, S. J. (1997). Verification of Eta–RSM Short-Range Ensemble Forecasts. *Monthly Weather Review*, 125(6), 1312–1327.

Hamill, T. M., Whitaker, J. S., & Mullen, S. L. (2006). Reforecasts : An important dataset for improving weather predictions. *Bulletin of the American Meteorological Society*, 87(1), 33-46. Scopus.

Hamill, T. M., Bates, G. T., Whitaker, J. S., Murray, D. R., Fiorino, M., Galarneau, T. J., ... Lapenta, W. (2013). NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. *Bulletin of the American Meteorological Society*, 94(10), 1553–1565.

Hashino, T., Bradley, A. A., & Schwartz, S. S. (2007). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences*, 11(2), 939-950.

Herbert, Z. C., Asghar, Z., & Oroza, C. A. (2021). Long-term Reservoir Inflow Forecasts: Enhanced Water Supply and Inflow Volume Accuracy Using Deep Learning. *Journal of Hydrology*, 601, 126676.

Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, 15(5), 559–570.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtol, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., ... Zuo, H. (2018, Dec). Operational global reanalysis: progress, future directions and synergies with NWP. ERA report series No. 27.

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.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

- Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. *Water*, 10(11), 1543.
- Hunt, K. M. R., Matthews, G. R., Pappenberger, F., & Prudhomme, C. (2022, February 8). *Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States*.
- Ionno, A., Arsenault, R., Troin, M., Martel, J.-L., & Brissette, F. (2024). Impacts of climate change on flood volumes over North American catchments. *Journal of Hydrology*, 630, 130688.
- IPCC, P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.- O. Pörtner, . . . Malley, J. (2019). *Summary for Policymakers: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. The Intergovernmental Panel on Climate Change. Repéré à <https://www.ipcc.ch/srccl/chapter/summary-for-policymakers/>
- Jörg-Hess, S., Kempf, S. B., Fundel, F., & Zappa, M. (2015). The benefit of climatological and calibrated reforecast data for simulating hydrological droughts in Switzerland. *Meteorological Applications*, 22(3), 444–458.
- Kang, T.-H., Kim, Y.-O., & Hong, I.-P. (2010). Comparison of pre- and post-processors for ensemble streamflow prediction. *Atmospheric Science Letters*, 11(2), 153-159.
- Kao, I.-F., Liou, J.-Y., Lee, M.-H., & Chang, F.-J. (2021). Fusing stacked autoencoder and long short-term memory for regional multistep-ahead flood inundation forecasts. *Journal of Hydrology*, 598, 126371.
- 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, 264-277.
- Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., & Woods, R. A. (2019). Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1.2 : An open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. *Geoscientific Model Development*, 12(6), 2463-2480.
- 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.
- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S.

- (2019a). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12), 11344-11354.
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019b). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, 23(12), 5089-5110.
- 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.
- Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of hydrology*, 249(1-4), 2-9.
- Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., & Dadson, S. J. (2021). Benchmarking data-driven rainfall-runoff models in Great Britain : A comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. *Hydrology and Earth System Sciences*, 25(10), 5517-5534.
- Leutbecher, M., & Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, 227(7), 3515-3539.
- Madadgar, S., Moradkhani, H., & Garen, D. (2014). Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes*, 28(1), 104-122.
- Martel, J.-L., Brissette, F., Arsenault, R., Turcotte, R., Castañeda-Gonzalez, M., Armstrong, W., Mailhot, E., Pelletier-Dumont, J., Rondeau-Genesse, G., & Caron, L.-P. (2024a). Assessing the adequacy of traditional hydrological models for climate change impact studies: A case for long-short-term memory (LSTM) neural networks. *EGUsphere*, 2024, 1-44.
- Martel, J.-L., Arsenault, R., Turcotte, R., Castañeda-Gonzalez, M., Brissette, F., Armstrong, W., Mailhot, E., Pelletier-Dumont, J., Lachance-Cloutier, S., Rondeau-Genesse, G., & Caron, L.-P. (2024b). Exploring the ability of LSTM-based hydrological models to simulate streamflow time series for flood frequency analysis. *EGUsphere*, 2024, 1-32.
- Maurer, E. P., & Lettenmaier, D. P. (2004). Potential Effects of Long-Lead Hydrologic Predictability on Missouri River Main-Stem Reservoirs. *Journal of Climate*, 17(1), 174-186.
- McCuen, R. H. (2016). *Modeling Hydrologic Change: Statistical Methods*. CRC Press, Boca Raton, FL.
- Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., ... Arnold, J.

- R. (2017). An intercomparison of approaches for improving operational seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, 21(7), 3915-3935.
- Muleta, M. K. (2012). Improving Model Performance Using Season-Based Evaluation. *Journal of Hydrologic Engineering*, 17(1), 191–200.
- Muluye, G. Y., & Coulibaly, P. (2007). Seasonal reservoir inflow forecasting with low-frequency climatic indices: a comparison of data-driven methods. *Hydrological Sciences Journal*, 52(3), 508–522.
- Papamichail, D. M., & Georgiou, P. E. (2001). SEASONAL ARIMA INFLOW MODELS FOR RESERVOIR SIZING. *JAWRA Journal of the American Water Resources Association*, 37(4), 877-885.
- Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J., & de Roo, A. P. J. (2005). Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). *Hydrology and Earth System Sciences*, 9(4), 381-393.
- Pechlivanidis, I. G., Crochemore, L., Rosberg, J., & Bosshard, T. (2020). What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts? *Water Resources Research*, 56(6), e2019WR026987.
- Peñuela, A., Hutton, C., & Pianosi, F. (2020). Assessing the value of seasonal hydrological forecasts for improving water resource management : Insights from a pilot application in the UK. *Hydrology and Earth System Sciences*, 24(12), 6059-6073.
- Ramos, M.-H., Mathevet, T., Thielen, J., & Pappenberger, F. (2010). Communicating uncertainty in hydro-meteorological forecasts : Mission impossible? *Meteorological Applications*, 17(2), 223-235.
- Ramos, M. H., van Andel, S. J., & Pappenberger, F. (2013). Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences*, 17(6), 2219-2232.
- Régie de l'énergie du Canada (2024). Profils énergétiques des provinces et territoires – Québec. Repéré à <https://www.cer-rec.gc.ca/fr/donnees-analyse/marches-energetiques/profils-energetiques-provinces-territoires/profils-energetiques-provinces-territoires-quebec.html>
- Roulin, E. (2007). Skill and relative economic value of medium-range hydrological ensemble predictions. *Hydrology and Earth System Sciences*, 11(2), 725-737.

- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. (1986). Sequential thought processes in PDP models. *Parallel distributed processing: explorations in the microstructures of cognition*, 2, 3-57.
- Saavedra Valeriano, O. C., Koike, T., Yang, K., Graf, T., Li, X., Wang, L., & Han, X. (2010). Decision support for dam release during floods using a distributed biosphere hydrological model driven by quantitative precipitation forecasts. *Water Resources Research*, 46(10).
- Saba, L., Biswas, M., Kuppili, V., Godia, E. C., Suri, H. S., Edla, D. R., ... Suri, J. S. (2019). The present and future of deep learning in radiology. *European Journal of Radiology*, 114, 14-24.
- 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.
- Saima, H., Jaafar, J., Belhaouari, S., & Jillani, T. A. (2011). Intelligent methods for weather forecasting: A review. *2011 National Postgraduate Conference*, 1-6.
- Sankarasubramanian, A., Lall, U., Devineni, N., & Espinueva, S. (2009). The Role of Monthly Updated Climate Forecasts in Improving Intraseasonal Water Allocation. *Journal of Applied Meteorology and Climatology*, 48(7), 1464-1482.
- Sattari, M. T., Yurekli, K., & Pal, M. (2012). Performance evaluation of artificial neural network approaches in forecasting reservoir inflow. *Applied Mathematical Modelling*, 36(6), 2649–2657.
- Saxena, A., Verma, N., & Tripathi, K. C. (2013). A review study of weather forecasting using artificial neural network approach. *Int. J. Eng. Res. Technol*, 2(11), 2029-2036.
- Schaake, J., Paillex, J., Thielen, J., Arritt, R., Hamill, T., Luo, L., Martin, E., McCollor, D. and Pappenberger, F. (2010), Summary of recommendations of the first workshop on Postprocessing and Downscaling Atmospheric Forecasts for Hydrologic Applications held at Météo-France, Toulouse, France, 15–18 June 2009. *Atmosph. Sci. Lett.*, 11: 59-63.
- Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, 82(12), 2635-2670.
- Stokelj, T., Paravan, D., & Golob, R. (2002). Enhanced Artificial Neural Network Inflow Forecasting Algorithm for Run-of-River Hydropower Plants. *Journal of Water Resources Planning and Management*, 128(6), 415-423.

- Talagrand, O., Vautard, R., & Strauss, B. (1999). Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability*, 4.2.
- Thiboult, A., Anctil, F., & Boucher, M.-A. (2016). Accounting for three sources of uncertainty in ensemble hydrological forecasting. *Hydrology and Earth System Sciences*, 20(5), 1809-1825.
- Thompson, J. C. (1952). On the Operational Deficiencies in Categorical Weather Forecasts. *Bulletin of the American Meteorological Society*, 33(6), 223-226.
- Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J. (2021). Generating Ensemble Streamflow Forecasts : A Review of Methods and Approaches Over the Past 40 Years. *Water Resources Research*, 57(7), e2020WR028392.
- Turner, S. W. D., Bennett, J. C., Robertson, D. E., & Galelli, S. (2017). Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. *Hydrology and Earth System Sciences*, 21(9), 4841-4859.
- United Nations (2024). Renewable energy – powering a safer future. Repéré à <https://www.un.org/en/climatechange/raising-ambition/renewable-energy>
- Verkade, J. S., Brown, J. D., Reggiani, P., & Weerts, A. H. (2013). Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology*, 501, 73–91.
- Xu, J., Anctil, F., & Boucher, M.-A. (2019). Hydrological post-processing of streamflow forecasts issued from multimodel ensemble prediction systems. *Journal of Hydrology*, 578, 124002.
- Xu, W., Jiang, Y., Zhang, X., Li, Y., Zhang, R., & Fu, G. (2020). Using long short-term memory networks for river flow prediction. *Hydrology Research*, 51(6), 1358–1376.
- Yokoo, K., Ishida, K., Ercan, A., Tu, T., Nagasato, T., Kiyama, M., & Amagasaki, M. (2022). Capabilities of deep learning models on learning physical relationships : Case of rainfall-runoff modeling with LSTM. *Science of The Total Environment*, 802, 149876.
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. *Advances in neural information processing systems*, 27.
- Zalachori, I., Ramos, M.-H., Garçon, R., Mathevret, T., & Gailhard, J. (2012). Statistical processing of forecasts for hydrological ensemble prediction : A comparative study of different bias correction strategies. *Advances in Science and Research*, 8(1), 135-141.
- Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., ... Zhuang, J. (2018).

Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*, 565, 720-736.

Zhang, J., Chen, J., Li, X., Chen, H., Xie, P., & Li, W. (2020). Combining Postprocessed Ensemble Weather Forecasts and Multiple Hydrological Models for Ensemble Streamflow Predictions. *Journal of Hydrologic Engineering*, 25(1), 04019060.

Zhang, Y., Ragettli, S., Molnar, P., Fink, O., & Peleg, N. (2022). Generalization of an Encoder-Decoder LSTM model for flood prediction in ungauged catchments. *Journal of Hydrology*, 614, 128577.

Zhao, L., Duan, Q., Schaake, J., Ye, A., & Xia, J. (2011). A hydrologic post-processor for ensemble streamflow predictions. *Advances in Geosciences*, 29, 51-5.