

Apprentissage profond appliqué à la prévision saisonnière des volumes d'apport dans les réservoirs hydroélectriques

par

Dylane MARTEL

MÉMOIRE PAR ARTICLE 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.

MONTREAL, LE 4 DÉCEMBRE 2025

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Dylane Martel, 2025



Cette licence [Creative Commons](https://creativecommons.org/licenses/by-nc-nd/4.0/) 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 Annie Levasseur, présidente du jury
Département de génie de la construction à l'École de technologie supérieure

M. Reda Snaiki, membre du jury
Département de génie de la construction à l'École de technologie supérieure

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

LE 21 NOVEMBRE 2025

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

J'aimerais tout d'abord remercier mon directeur de maîtrise, Richard Arsenault, pour sa passion contagieuse pour le domaine, son soutien tout au long du projet et sa disponibilité exemplaire. Je t'ai d'abord eu comme professeur en CTN336 – Mécanique des fluides. J'ai adoré ton énergie et ta façon d'enseigner. Au début de mon BAC, je ne savais pas trop dans quel domaine me spécialiser, mais depuis ton cours, j'ai su que je voulais m'orienter vers l'hydrologie. Je me suis par la suite tournée vers toi pour mon dernier stage et j'étais très heureuse que tu m'acceptes comme stagiaire. Merci pour tes nombreux conseils et ta confiance tout au long de mon parcours.

Je tiens également à remercier mon co-directeur, Jean-Luc Martel, qui a lui aussi su susciter mon intérêt pour le domaine grâce à son enseignement du cours CTN446 – Hydraulique et hydrologie. Un des rares profs que j'ai vus avec autant d'énergie ! Merci pour ta passion contagieuse, ton soutien et ton aide dans la rédaction de mon mémoire.

J'aimerais également remercier Rio Tinto et le CRSNG, qui ont financé ce projet, ainsi que l'ÉTS, le CRSNG et Hydro-Québec pour les bourses d'études qui m'ont offert la possibilité de me consacrer entièrement à mon mémoire.

Sur le plan personnel, j'aimerais remercier mes amis et ma famille. On me demandait souvent : « Et puis, t'en es où ? Tu penses finir quand ? » et moi de dire : « Ça avance ! Bientôt ! » Désolée pour les réponses un peu floues. La réalité, c'est que ça m'a pris plus de temps que je ne le pensais et aussi que je n'avais aucune idée réellement de quand j'allais terminer. Malgré tout, vous avez su me motiver et m'encourager à me surpasser. Merci aussi au soutien financier de mes parents aka la Fondation Martel haha. C'est aussi grâce à vous que j'ai pu payer mes études et m'y consacrer pleinement.

Enfin, je réserve mes plus profonds remerciements à mon fiancé, Francis. Tu es avant tout la personne qui partage mon quotidien et supporte mes hauts et mes bas. Tu as su me soutenir

tout au long de mon parcours scolaire, et ce, depuis la fin de mon secondaire. Merci de croire en moi depuis toutes ces années, de me pousser à me dépasser, de me faire rire même quand mon moral est bas. Je ne te le dis peut-être pas assez, mais merci pour ta patience, tes conseils et ton écoute. Je suis vraiment chanceuse de t'avoir à mes côtés et j'ai hâte de continuer à bâtir un avenir avec toi. *Je t'aime.*

Apprentissage profond appliqué à la prévision saisonnière des volumes d'apport dans les réservoirs hydroélectriques

Dylane MARTEL

RÉSUMÉ

Les réservoirs hydroélectriques nécessitent des prévisions d'apports afin de permettre aux gestionnaires des ressources en eau d'optimiser les taux de vidange et d'améliorer l'efficacité des infrastructures. Habituellement, les exploitants utilisent des modèles hydrologiques conceptuels ou basés sur la physique pour prévoir les débits à des échéances allant d'un mois à un an, selon l'horizon de planification. Ces prévisions sont ensuite agrégées pour estimer les volumes d'apports cumulés, qui sont essentiels pour la gestion saisonnière des réservoirs. Des prévisions fiables de volumes à moyenne échéance (quelques semaines) soutiennent non seulement l'optimisation du remplissage des réservoirs et de la production hydroélectrique, mais aussi la préparation aux crues saisonnières et aux périodes d'étiage.

Pour répondre plus directement à ces besoins, cette étude propose une méthode novatrice qui prédit directement les volumes d'apport cumulés sur 45 jours à l'aide d'un modèle d'apprentissage profond entraîné sur des prévisions météorologiques ensemblistes et des données observées de volume, sans passer par la simulation des débits journaliers, et en ciblant le volume total sur l'ensemble de la période en une seule opération. De plus, le modèle exploite de vastes jeux de données durant sa phase d'apprentissage, en utilisant des données issues de 200 bassins versants canadiens. Il est ensuite appliqué à huit bassins canadiens pour estimer les volumes d'apport prévisionnels. Enfin, pour le système hydroélectrique du Lac-Saint-Jean, au Québec, les performances des réseaux LSTM, un type de réseau de neurones récurrent conçu pour capter les dépendances temporelles, sont comparées à celles d'un modèle hydrologique conceptuel opérationnel.

Les résultats montrent que les prévisions provenant des modèles LSTM sont généralement précises mais manquent de fiabilité. Les prévisions ensemblistes ont tendance à être trop confiantes et sous-dispersées. L'entraînement régional améliore les performances dans certains cas, mais aucun schéma universel ne se dégage à travers toutes les saisons ou tous les bassins. Comparativement au modèle conceptuel opérationnel, les modèles LSTM le surpassent durant la saison automnale, tant en termes d'exactitude que de fiabilité, soulignant ainsi leur potentiel complémentaire. En complément, l'entraînement direct d'un modèle LSTM entraîné localement avec les données de prévisions météorologiques d'ensembles a permis d'évaluer les limites face à cette approche. Dans l'ensemble, cette étude suggère que les modèles basés sur les LSTM représentent une alternative prometteuse pour la planification stratégique des réservoirs, en particulier lorsqu'ils sont combinés à des modèles hydrologiques physiques, afin d'améliorer la représentation de l'incertitude.

Mots-clés : prévision hydrologique ensembliste, apprentissage profond, volume d'apport, LSTM, apprentissage à grande échelle

Deep learning applied to seasonal inflow volume forecasts in hydroelectric reservoirs

Dylane MARTEL

ABSTRACT

Hydropower reservoirs require inflow forecasts to allow water resources managers to optimize drawdown rates and improve infrastructure efficiency. Usually, operators use physically-based or conceptual hydrological models to forecast streamflow for a lead-time ranging from one month to a year, depending on the planning horizon. These forecasts are then aggregated to evaluate cumulative inflow volumes, which are essential for seasonal operations. Reliable medium-range volume forecasts over a few weeks support not only the optimization of reservoir filling and hydropower generation, but also preparedness for seasonal floods and low-flow periods.

To address these needs more directly, this study presents a novel method that directly forecasts 45-day cumulative inflow volumes using a deep learning model trained on ensemble meteorological forecasts and observed volume data, avoiding the need to simulate daily streamflow and targeting the total volume over the full 45-day period in a single operation. Furthermore, the model leverages large-scale datasets during its training, by using data from 200 basins in Canada. The model is then applied to eight Canadian basins to estimate inflow volume forecasts. For the Lac-Saint-Jean hydroelectric system in Quebec, the performance of LSTM networks, a type of recurrent neural network designed to capture temporal dependencies, is also compared with an operational conceptual hydrological model.

The results show that forecasts produced by LSTM models are generally accurate but lack reliability. Ensemble forecasts tend to be overconfident and underdispersed. Regional training improves performance in some cases, but no universal pattern emerges across all seasons or basins. Compared to the operational conceptual model, LSTM models outperform it during the autumn season, both in terms of accuracy and reliability, highlighting their complementary potential. Additionally, the direct training of a locally trained LSTM model using ensemble meteorological forecasts helped assess the limitations of this approach. Overall, this study suggests that LSTM-based models offer a promising alternative for strategic reservoir planning, particularly when combined with physically based hydrological models to improve uncertainty representation.

Keywords: Ensemble hydrological forecasting, deep learning, inflow volume, LSTM, Large-sample training

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 REVUE DE LITTÉRATURE	5
1.1 Importance de l'hydrologie.....	5
1.2 Modèles hydrologiques traditionnels.....	6
1.2.1 Modèles hydrologiques conceptuels	7
1.2.2 Modèles hydrologiques physiques.....	7
1.3 Modèles basés sur l'intelligence artificielle.....	8
1.4 Surapprentissage	10
1.5 Modélisation hydrologique régionale appliquée à la simulation des débits	11
1.6 Forçages météorologiques	14
1.6.1 Données de réanalyse.....	14
1.6.2 Prévisions météorologiques	16
1.7 Prévision hydrologique	17
1.7.1 Prévisions déterministes et ensemblistes	19
1.7.2 Incertitudes en prévision hydrologique.....	20
1.7.3 Prévision avec un modèle régional	21
1.8 Prévisions des volumes d'eau	22
1.9 Objectifs de recherche.....	25
CHAPITRE 2 DEEP LEARNING APPLIED TO SEASONAL INFLOW VOLUME FORECASTS IN HYDROELECTRIC RESERVOIRS	27
2.1 Abstract	27
2.2 Introduction.....	28
2.3 Study area and data	32
2.3.1 Main study basin	32
2.3.2 Regional training and complementary analysis basins	34
2.3.3 Reanalysis data.....	36
2.3.4 Ensemble forecast data	37
2.3.5 Basin descriptor data.....	37
2.3.6 Streamflow and inflow volumes	38
2.4 Methodology	39
2.4.1 Forecasting models and approaches.....	39
2.4.2 Data preparation.....	40
2.4.3 LSTM model and structures.....	43
2.4.4 Training.....	44
2.4.5 The CEQUEAU hydrological model	45
2.4.6 Performance metrics	45
2.5 Results.....	48

2.5.1	Model training and verification	48
2.5.2	Overall comparison of model performance in forecasting on the Lac-Saint-Jean basin	49
2.5.3	Accuracy of annual and seasonal forecasting models across the test basins.....	50
2.5.4	Reliability of the LSTM-based inflow volume forecasting models.....	52
2.6	Discussion	56
2.6.1	The performance of LSTM in testing	56
2.6.2	Influence of physiographic and climatic factors	57
2.6.3	Seasonal performance differences for the Lac-Saint-Jean basin	60
2.6.4	Seasonal variability in forecast performance across the other basins	62
2.6.5	Forecast reliability issues	63
2.6.6	Limitations	66
2.7	Conclusion and recommendations	67
2.8	Acknowledgments.....	68
	CHAPITRE 3 DISCUSSION	71
3.1	Adaptation de la méthode à divers horizons de prévisions.....	71
3.2	Défis liés à la programmation et l'entraînement des modèles LSTM	72
3.3	Exploration d'un scénario base sur les prévisions ensemblistes.....	75
3.3.1	Limites de l'apprentissage à partir des prévisions ensemblistes.....	81
	CONCLUSION.....	85
	RECOMMANDATIONS	87
ANNEXE I	DÉBIT QUOTIDIEN MOYEN AVEC ENVELOPPE MIN-MAX POUR LES HUIT BASSINS TESTS	89
ANNEXE II	COMPOSANTES DU KGE POUR LES SCÉNARIOS LOC ET REG	91
ANNEXE III	CRPS NORMALISÉ DES PRÉVISIONS D'ENSEMBLE DES SCÉNARIOS LOC ET REG	93
ANNEXE IV	CRPS DES PRÉVISIONS D'ENSEMBLE DES VOLUMES POUR LE BASSIN DU LAC-SAINT-JEAN POUR LES MODÈLES INDIVIDUELS ET LES COMBINAISONS MULTI-MODÈLES	95
	LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES	97

LISTE DES TABLEAUX

	Page
Tableau 2.1	Characteristics details of the study basins (1979–2023)..... 36
Tableau 2.2	Basin descriptors used in this study 38
Tableau 2.3	KGE scores for the training of the LOC and REG scenarios 48
Tableau 3.1	KGE scores for the training of the LOC, REG and L_FC scenarios 76

LISTE DES FIGURES

	Page
Figure 2.1	Location of the Lac-Saint-Jean basin in Canada, as well as the hydropower generating station at the reservoir outlet..... 33
Figure 2.2	Daily average and minimum to maximum range hydroclimatology for the Lac-Saint-Jean basin (1979 to 2023): (a) Streamflow; (b) Air temperature; (c) Total precipitation. The dominant spring peak in discharge reflects seasonal snowmelt, with a moderate autumnal rise also evident, largely attributed to increased rainfall and reduced evapotranspiration 34
Figure 2.3	Location of the 200 basins distributed across the provinces of Canada. The green basin outline represents the LSJ basin and the seven red outlines show the locations of the basins used to test the method generalization. All eight outlined basins are part of the 200 basins used for model training 35
Figure 2.4	Example of the implementation of a training sample for a 45-day cumulative volume forecast model, trained with observed data (a), as well as an example of how such volume forecast models are applied over an independent period (b). Panel (a) shows a single sample, but the process is repeated for all available data used to train the models, as well as for the validation and testing periods. Similarly, panel (b) illustrates an example of a single 45-day cumulative volume forecast, but the process is repeated for all available forecast dates 41
Figure 2.5	Structures of the LOC LSTM model (a) and REG model (b)..... 44
Figure 2.6	Ensemble forecasts of 45-day inflow volumes for the LSJ basin: (a) Min–max envelopes, (b) Daily mean forecasts for three scenarios: LOC, REG and CEQ. The forecast period covers March 10, 2016 to May 29, 2023 50
Figure 2.7	CRPS of the volume ensemble forecast by season for the LOC and REG models over the entire period of the forecast. Each panel represents a basin: a) Lac-Saint-Jean [1], b) Grande River [2], c) Cascapedia River [3], d) Bear Brook [4], e) Hayes River [5], f) Swift Current Creek [6] , g) Crowsnest River [7], and h) Tuya River [8]. The CEQUEAU model is also shown in panel a) for Lac-Saint-Jean. Seasons are defined as winter (DJFM), spring

	(AM), summer (JJA), fall (SON), and annual (Y). The forecast period covers March 10, 2016 to May 29, 2023	51
Figure 2.8	Forecast reliability for the Lac-Saint-Jean basin under three modeling scenarios. The figure presents Talagrand diagrams illustrating the forecast rank distributions (in percentiles) of 45-day cumulative water volumes for each scenario. The scenarios include: LOC (local LSTM model), REG (regional LSTM model trained on 200 basins), and CEQUEAU (hydrological reference model). Forecast reliability is assessed over the entire available prediction period (2016–2023)	54
Figure 2.9	Annual forecast reliability for three LSTM-based modeling scenarios: LOC (local model) and REG (regional model trained from 200 basins). The evaluation period spans the full range of available forecasts (2016–2023). Talagrand diagrams represent the forecast rank distribution (in percentiles) of 45-day cumulative water volumes for each scenario, across seven Canadian basins, numbered 2 to 8: Grande River [basin 2, panels a-b], Cascapedia River [3, c-d], Bear Brook [4, e-f], Hayes River [5, g-h], Swift Current Creek [6, i-j], Crowsnest River [7, k-l], and Tuya River [8, m-n]	55
Figure 2.10	Talagrand diagrams of 45 day cumulative inflow forecasts (2016–2023) for the Lac Saint Jean basin for five scenarios: LOC (local LSTM), REG (regional LSTM trained on 200 basins), CEQUEAU (reference hydrological model), and two combined ensembles (LOC+CEQUEAU and REG+CEQUEAU)	65
Figure 3.1	CRPS des prévisions ensemblistes de volume par saison pour les scénarios : LOC, REG et L_FC. Le modèle CEQUEAU est également présenté dans le panneau a) pour le Lac-Saint-Jean. Chaque panneau représente un bassin versant : a) Lac-Saint-Jean [1], b) Grande River [2], c) Cascapedia River [3], d) Bear Brook [4], e) Hayes River [5], f) Swift Current Creek [6], g) Crowsnest River [7], and h) Tuya River [8]. Les saisons sont définies comme suit : hiver (DJFM), printemps (AM), été (JJA), automne (SON) et annuel (Y). La période de prévision correspondant aux derniers 20 % de la période de test du scénario L_FC, entre 2016 et 2023	78
Figure 3.2	Fiabilité des prévisions pour le bassin du Lac-Saint-Jean selon quatre scénarios de modélisation : LOC, REG, L_FC et CEQUEAU. La figure présente des diagrammes de Talagrand illustrant la distribution des rangs de prévision (en percentiles) des volumes d'eau cumulés sur 45 jours pour chaque scénario. La fiabilité des prévisions est évaluée sur les dates communes des	

scénarios, du 20 décembre 2021 au 29 mai 2023 inclusivement. Les sous-figures dont le texte ABDU est rouge correspondent aux cas où la distribution des rangs ne s'écarte pas significativement d'une loi uniforme (selon un test de Kolmogorov–Smirnov avec un seuil de 0,05), suggérant une fiabilité satisfaisante. À l'inverse, les sous-figures dont le texte ABDU est noir indiquent une divergence statistiquement significative par rapport à l'uniformité, suggérant un manque de fiabilité..... 79

Figure 3.3 Fiabilité annuelle des prévisions pour trois scénarios de modélisation basés sur des LSTM : LOC, REG et L_FC. Les diagrammes de Talagrand représentent la distribution des rangs de prévision (en percentiles) des volumes d'eau cumulés sur 45 jours pour chaque scénario, sur sept bassins canadiens numérotés de 2 à 8 : Grande Rivière [bassin 2, panneaux a–b], Rivière Cascapédia [3, c–d], Bear Brook [4, e–f], Rivière Hayes [5, g–h], Swift Current Creek [6, i–j], Rivière Crowsnest [7, k–l] et Rivière Tuya [8, m–n]. La période de prévision correspondant aux derniers 20 % de la période de test du scénario L_FC, entre 2016 et 2023..... 80

Figure 3.4 Prévisions ensemblistes des volumes d'apport sur 45 jours pour le bassin du Lac-Saint-Jean : (a) Enveloppes min–max, (b) Moyennes journalières des prévisions pour les quatre scénarios : LOC, REG, L_FC et CEQUEAU. La période de prévision s'étend du 20 décembre 2021 au 29 mai 2023 81

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ABDU	Average Bin Distance to Uniformity
AI	Intelligence artificielle (Artificial Intelligence)
AM	Saison de la fonte de neige - Avril à Mai
ANN	Réseaux de neurones artificiels (Artificial Neural Networks)
CEQUEAU	Modèle hydrologique développé par l’Institut national de la recherche scientifique (INRS) de l’Université du Québec (CEntre QUébécois des sciences de l’EAU)
CRPS	Continuous rank probability score
DJFM	Saison hivernale – Décembre à Mars
DL	Apprentissage profond (Deep Learning)
ECCC	Environnement et Changement climatique Canada
ECMWF	European centre for medium-range weather forecasts
GR4J	Modèle hydrologique Génie Rural à 4 paramètres Journalier
HBV	Modèle hydrologique Hydrologiska Byråns Vattenbalansavdelning
HUC	Code d'unité hydrologique (Hydrologic Unit Code)
JJA	Saison d’été - Juin à Août
KGE	Kling-Gupta Efficiency
KS	Kolmogorov-Smirnov
LSJ	Lac-Saint-Jean
LSTM	Long Short-Term Memory
MOHYSE	MOdèle HYdrologique Simplifié à l'Étrême
RNCan	Ressources Naturelles Canada
RNN	Réseau de neurones récurrents (Recurrent Neural Networks)

SEAS5	Seasonal Forecasting System version 5
SON	Saison d'automne – Septembre à Novembre
SRM	Snowmelt Runoff Model
TOPMODEL	TOPOgraphy based hydrological MODEL
WASA	Water Availability in Semi-Arid Environments

LISTE DES SYMBOLES ET UNITÉS DE MESURE

UNITÉS GÉOMÉTRIQUES

Longueur

m	mètre
mm	millimètre

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

UNITÉS MÉCANIQUES

Vitesse

m/s	mètre par seconde
-----	-------------------

INTRODUCTION

Grâce à ses nombreux plans d'eau, le Canada compte près de 600 centrales hydroélectriques qui ont permis de produire 57,8% de l'électricité nationale en 2023 (Statistique Canada, 2024; RNCAN, 2024). L'hydrologie occupe ainsi une place centrale dans la gestion des ressources et l'approvisionnement énergétique du pays. De plus, face aux changements climatiques, comprendre et prévoir le comportement de l'eau sur le territoire devient indispensable pour gérer les risques, optimiser les ressources et se préparer à des événements météorologiques extrêmes de plus en plus fréquents (Coumou & Rahmstorf, 2012).

La prévision hydrologique débute par l'utilisation d'un modèle capable d'estimer le débit à l'exutoire d'un bassin versant. Selon l'approche choisie, ce modèle peut représenter plus ou moins en détail le comportement de l'eau sur le territoire, en incluant ou non certains processus internes comme l'infiltration, le ruissellement ou l'évaporation. Avant de pouvoir produire des prévisions, le modèle doit être alimenté par des données, généralement observées, afin d'être calibré. La calibration consiste à ajuster les paramètres pour que les résultats du modèle se rapprochent le plus possible des données historiques afin de représenter le comportement hydrologique du bassin versant en dépit des lacunes dans la représentation directe des processus. Une fois calibré, le modèle peut être utilisé de différentes façons pour générer des prévisions. Par exemple, il est possible d'y intégrer des prévisions météorologiques en entrée, ce qui permet d'estimer les débits futurs. Une autre méthode consiste à utiliser des données historiques, en partant du principe que ce qui s'est produit dans le passé a des chances de se reproduire dans le futur (Troin et al., 2021).

Plusieurs études ont exploré différentes façons d'améliorer les outils de modélisation hydrologique et la performance des prévisions de débit. L'arrivée, au cours des dernières décennies, de l'intelligence artificielle a permis le développement de modèles capables de modéliser et conserver les états d'un système dans le temps, permettant de modéliser des processus se déroulant sur de longues périodes. Ces modèles, en particulier les réseaux de neurones récurrents de type LSTM (Long Short-Term Memory), ont démontré des

performances supérieures rapport aux approches traditionnelles (Arsenault, Martel, Brunet, Brissette, & Mai, 2023a; Kratzert et al., 2018). Appliqués à des horizons variés, ils affichent d'excellentes performances à court terme (quelques jours ; Armstrong et al., 2025 ; Sabzipour et al., 2023) et, malgré le caractère instable et chaotique de l'atmosphère qui complique la prévision sur des échéances plus longues, plusieurs études soulignent leur potentiel à moyen et long terme (Cheng et al., 2020 ; Herbert et al., 2021). Un élément clé pour assurer la performance de ces modèles est la quantité de données disponibles. Plusieurs stratégies permettent d'enrichir les jeux de données, notamment par l'utilisation de sources variées ou par l'exploitation de données issues de nombreux bassins versants. La modélisation régionale, qui consiste à entraîner un modèle à partir de plusieurs bassins simultanément, a démontré son efficacité dans de nombreuses études en améliorant la capacité de généralisation et la robustesse des simulations et des prévisions (He, Xu, Wu, Kang, & Huang, 2025; Kratzert et al., 2018).

L'amélioration de la fiabilité et de la précision des prévisions est particulièrement cruciale pour les centrales hydroélectriques, qui dépendent de ces informations pour optimiser la gestion des réservoirs. Celles-ci s'appuient sur des prévisions de volumes à court, moyen et long terme pour assurer la gestion quotidienne des réservoirs ainsi que la planification stratégique des apports et des déversements. Or, ces volumes sont généralement obtenus en agrégeant des prévisions de débit journalier. Jusqu'à présent, les modèles d'intelligence artificielle appliqués à l'hydrologie ont surtout été calibrés en se basant sur les débits, comme il est nécessaire de faire en modélisation hydrologique traditionnelle. Une question pertinente se pose alors : pourquoi ne pas entraîner directement ces modèles sur les volumes, c'est-à-dire sur l'indicateur opérationnel lui-même ? Cela permettrait de réduire la complexité et la durée de calcul, de potentiellement mieux représenter la plage d'incertitude des prévisions, et surtout, d'aligner les prévisions avec les besoins concrets des gestionnaires hydriques.

C'est dans cette optique, qui vise à améliorer les prévisions sur un horizon reconnu comme particulièrement incertain, à utiliser des techniques d'apprentissage régional et à cibler directement les volumes d'apport, que s'inscrit cette étude. L'objectif principal est de prévoir

les volumes de crue cumulés sur un horizon de 45 jours, à partir d'un ensemble de données provenant de 200 bassins versants répartis à travers le Canada. L'analyse évalue la valeur ajoutée d'un modèle LSTM entraîné de façon régionale, c'est-à-dire sur plusieurs bassins simultanément, comparé à une approche locale, où un modèle est calibré séparément pour chaque bassin. La méthodologie est testée sur huit bassins versants issus de l'échantillon régional, sélectionnés pour analyser la capacité et la robustesse de la méthodologie. L'un de ces bassins fait également l'objet d'une comparaison avec un modèle opérationnel actuellement utilisé.

Ce mémoire est constitué de quatre chapitres. Le premier présente une revue de littérature qui détaille les avancées récentes ainsi que les pistes explorées pour améliorer les prévisions hydrologiques en général, et plus spécifiquement celles portant sur la prévision des volumes d'eau. Le deuxième chapitre présente l'article scientifique soumis au journal *Journal of Hydrology*, décrivant la zone d'étude, la méthodologie, les résultats et leur analyse. Le troisième chapitre propose une réflexion plus large et la présentation de résultats supplémentaires. Enfin, le quatrième chapitre présente une conclusion générale ainsi que des recommandations pour les travaux futurs.

CHAPITRE 1

REVUE DE LITTÉRATURE

Cette section présente une revue de littérature sur la modélisation et la prévision hydrologique, en mettant particulièrement l'accent sur les avancées des modèles LSTM fondés sur l'intelligence artificielle, la prévision des volumes d'eau, ainsi que sur les approches et techniques afférentes.

1.1 Importance de l'hydrologie

La science de l'hydrologie joue un rôle essentiel dans la compréhension et la gestion de l'eau à l'échelle des bassins versants. Elle permet d'expliquer comment la pluie et la neige deviennent du ruissellement, s'infiltrent dans le sol ou s'accumulent temporairement à la surface, en fonction des conditions météorologiques et des échanges entre le sol, la surface et l'atmosphère.

Comprendre ces processus est crucial pour plusieurs domaines, comme l'approvisionnement en eau potable, la prévention des inondations, ou encore la production d'hydroélectricité. Par exemple, dans le cas des barrages hydroélectriques, il est important de pouvoir prévoir les volumes d'eau à venir pour déterminer les meilleures actions à entreprendre quant à la quantité d'eau à relâcher des réservoirs, tout en permettant de s'y prendre suffisamment d'avance pour éviter des problèmes de surcharge ou trop vider le réservoir, pour ainsi optimiser la production d'hydroélectricité. Lorsqu'il est question de risques comme les sécheresses ou les crues, l'hydrologie permet aussi d'anticiper et de mieux se préparer à ces événements extrêmes, ce qui aide à protéger les populations et les infrastructures (Kundzewicz et al., 2003). Dans ce contexte, la modélisation hydrologique constitue un outil central, que ce soit pour simuler le fonctionnement des bassins versants à partir de données observées, ou pour produire des prévisions correspondant à différents scénarios futurs.

Cette revue de littérature s'ouvre sur une présentation des différents types de modèles hydrologiques, puis aborde les types de forçages utilisés dans la modélisation. Elle se poursuit avec une synthèse des approches de prévision hydrologique, pour se concentrer enfin sur les travaux portant spécifiquement sur la prévision des volumes d'apport en eau.

1.2 Modèles hydrologiques traditionnels

Il existe une grande variété de modèles hydrologiques, chacun conçu pour répondre à des besoins spécifiques et être performant dans certains contextes : par exemple, certains modèles appliqués aux bassins dominés par la neige intègrent finement les processus nivaux, notamment l'accumulation et la fonte graduelle de la neige, comme le font des modèles tels que HBV (Bergström & Forsman, 1973) ou SRM (Martinec, 1975). En revanche, en climat tropical, où les précipitations intenses et l'évapotranspiration élevée dominent, certains modèles hydrologiques simulent de manière particulièrement détaillée l'infiltration rapide, l'interception végétale et les échanges sol-atmosphère, tel que SWAT (Arnold et al., 1998). Une revue menée par Singh et Woolhiser (2002) identifie près de 70 modèles hydrologiques largement utilisés à l'échelle planétaire, bien qu'il en existe de nombreux autres. Pour structurer cette diversité, les modèles hydrologiques peuvent être classés selon leur degré de spatialisation (globaux, semi-distribués, distribués) et leur approche (conceptuelle vs fondée sur la physique). Les modèles globaux traitent le bassin comme un tout, sans représentation spatiale explicite, tandis que les modèles semi-distribués le découpent en sous-unités hydrologiques (souvent des sous-bassins) afin de saisir une partie de l'hétérogénéité. Les modèles pleinement distribués, pour leur part, maillent finement le territoire et résolvent localement les processus. Plus la résolution et le pas de temps sont élevés, plus les besoins en données (sol, topographie, couverture, etc.) et plus le temps de calcul et la puissance augmentent. Le choix d'un modèle dépend donc à la fois des objectifs de modélisation, des données disponibles et des contraintes de temps et de matériel: il doit être assez sophistiqué pour répondre aux besoins, sans jamais être plus complexe que nécessaire (Beven & Young, 2013).

1.2.1 Modèles hydrologiques conceptuels

Les modèles conceptuels cherchent à représenter le fonctionnement d'un bassin versant de manière simplifiée, en utilisant quelques réservoirs et flux d'eau reliés entre eux par des formules généralement simples, inspirées d'une représentation simplifiée des processus hydrologiques (Liu, Wang, Xu, & Dang, 2017). Ils sont le plus souvent globaux (p. ex. GR4J, MOHYSE) ou semi-distribués comme HBV. Bien que la majorité présente une résolution spatiale grossière, certains modèles adoptent un maillage plus fin : TOPMODEL (Beven & Kirkby, 1979) représente la variabilité du stockage et du ruissellement à l'aide d'un indice topographique, tandis que WASA (Water Availability in Semi-Arid Environments) utilise une discrétisation hiérarchique du paysage en cinq niveaux spatiaux, du bassin versant jusqu'au profil de sol, permettant de représenter la variabilité des caractéristiques du terrain, du sol et de la végétation (Güntner & Bronstert, 2004). Dans leur forme globale, ces modèles nécessitent relativement peu de données d'entrée et de paramètres, s'exécutent rapidement et fournissent une estimation efficace du débit. À titre d'exemple, GR4J (Perrin, Michel & Andréassian, 2003) ne comporte que six paramètres, soit les quatre paramètres hydrologiques de base du modèle GR4J, auxquels s'ajoutent deux paramètres liés à la neige lorsque le modèle de neige CEMANEIGE (Valéry et al. 2010) y est ajouté. Cette structure relativement simple facilite l'ajustement du modèle, tandis que MOHYSE (Fortin & Turcotte, 2007) en compte dix, offrant une flexibilité accrue pour représenter divers processus tout en restant accessible en termes de calcul et de données.

1.2.2 Modèles hydrologiques physiques

Les modèles hydrologiques basés sur la physique cherchent à reproduire explicitement les échanges et processus hydrologiques sur le territoire, en étant généralement distribués. Ces modèles sont non seulement utiles pour simuler et prévoir les débits, mais aussi pour analyser d'autres variables internes du bassin. Par exemple, les interactions surface-subsurface, le comportement de la neige ou l'impact de l'usage du sol peuvent être investigués à l'aide de ces modèles. Ils sont particulièrement pertinents lorsque l'objectif dépasse la simple simulation de débit. Par exemple, ils servent souvent à analyser l'impact des changements climatiques,

car ils peuvent montrer de manière plus fidèle comment le système réagit lorsque les conditions évoluent, incluant l'évolution de l'occupation du sol (Talbot et al., 2025).

1.3 Modèles basés sur l'intelligence artificielle

Les progrès récents en intelligence artificielle, et en particulier dans l'apprentissage profond (DL), ont ouvert de nouvelles perspectives pour la modélisation hydrologique. Depuis les années 1990, des réseaux de neurones artificiels (ANN) ont été explorés pour reproduire les relations non linéaires entre précipitations et débits, offrant une alternative aux approches hydrologiques traditionnelles (Daniel, 1991; Hsu, Gupta, & Sorooshian, 1995 ; Raman & Sunilkumar, 1995; Minns & Hall, 1996 ; Dawson & Wilby, 1998 ; Tokar & Markus, 2000 ; ASCE Task Committee, 2000). Contrairement aux modèles à base physique, ces approches apprennent directement à partir des données, sans nécessiter la formalisation explicite des processus physiques. En contrepartie, ces modèles restent difficiles à interpréter : les réseaux de neurones se comportent souvent comme une « boîte noire », rendant difficile d'expliquer précisément pourquoi ils accordent plus d'importance à certaines variables d'entrée qu'à d'autres ou comment elles affectent les sorties du modèle.

Une limitation importante des premiers réseaux était leur absence de mémoire interne, ce qui rendait difficile la capture des dépendances temporelles pourtant cruciales en hydrologie (ASCE Task Committee, 2000 ; Maier & Dandy, 2000). Pour y répondre, les réseaux de neurones récurrents (RNN) ont été introduits (Govindaraju & Rao, 2000; Caterini & Chang, 2018). Leur architecture maintient un état caché mis à jour au fil du temps, ce qui leur permet de mémoriser les conditions passées et de les utiliser pour prédire l'évolution future. Cependant, bien que les RNN soient particulièrement adaptés à l'analyse de séries temporelles, leur capacité de mémoire demeure limitée : en pratique, ils peinent à conserver l'information au-delà d'une dizaine de pas de temps (Bengio et al., 1994), ce qui restreint leur efficacité pour des horizons de prévision plus longs. Malgré cette contrainte, ils se sont avérés utiles pour la prévision hydrologique à court terme (Kang et al., 1993; Coulibaly & Anctil, 1999; Kumar et al., 2005; Yang et al., 2019). Par ailleurs, Pan & Wang (2005) ont montré l'intérêt des RNN pour simuler le processus pluie-débit.

Une avancée majeure a ensuite été l'introduction des réseaux à mémoire longue à court terme (LSTM, Long Short-Term Memory) par Hochreiter et Schmidhuber (1997). Conçus pour capturer à la fois les dépendances à court et à long terme dans des séries temporelles, les LSTM sont particulièrement bien adaptés aux systèmes hydrologiques, où des effets tels que l'accumulation et la fonte de la neige ou la recharge des eaux souterraines se manifestent avec un certain retard sur le débit (Kratzert et al., 2018; Kratzert et al., 2019). Ozdemir et Yildirim (2023) ont d'ailleurs montré que les LSTM obtiennent de très bons résultats pour la prévision du niveau d'eau sur un horizon de 60 jours, dépassant les performances des prévisions provenant du modèle ANN.

Depuis, de nombreuses variantes ont été proposées. Les réseaux bidirectionnels (Bi-LSTM) lisent la séquence dans les deux sens afin de mieux capter le contexte (Graves & Schmidhuber, 2005; Alfwzan et al., 2023; Zhang et al., 2023). Les LSTM empilés superposent plusieurs couches pour apprendre des relations plus complexes (Mirzaei et al., 2021; Waqas et al., 2024). Les Gated Recurrent Units (GRU) simplifient quant à eux l'architecture des LSTM et s'entraînent plus rapidement, tout en offrant des performances comparables. Par exemple, dans le cadre de prévisions à 120 jours des niveaux d'eau en Turquie, les GRU ont surpassé un ANN, un LSTM standard, un LSTM empilé et un Bi-LSTM (Ozdemir et Yildirim, 2023).

D'autres extensions combinent des architectures complémentaires. Les ConvLSTM, par exemple, utilisent des « convolutions », une opération qui fait passer un petit filtre sur les données afin de repérer automatiquement des motifs récurrents. Grâce à cela, ils prennent en compte à la fois l'évolution dans le temps et la répartition dans l'espace, ce qui les rend particulièrement adaptés aux données organisées en grille, comme les cartes de précipitations (Shi et al., 2015). Les LSTM avec attention mettent l'accent sur les instants ou variables les plus informatifs (Alizadeh et al., 2021), tandis que les schémas encodeur-décodeur condensent la séquence d'entrée pour ensuite générer une séquence de sortie complète. Cette architecture est particulièrement adaptée aux prévisions multi-étapes, qui visent à anticiper simultanément

l'évolution du débit jour par jour sur un horizon étendu plutôt que de prédire un seul jour à la fois (Cui et al., 2022; Kao et al., 2020).

Plusieurs études ont confirmé la capacité des LSTM à surpasser les modèles hydrologiques traditionnels, notamment dans des contextes marqués par une forte variabilité hydrométéorologique, tels que des expériences de simulation (Kratzert et al., 2018), des prévisions dans des bassins non jaugés (Arsenault et al., 2023a), des analyses de fréquence des crues (Martel et al., 2024) et des évaluations des impacts des changements climatiques (Martel et al., 2025). Ces modèles nécessitent toutefois d'importants volumes de données, incluant de nombreux bassins, de longues séries chronologiques d'observation et de multiples variables hydrométéorologiques pertinentes (par exemple, précipitations, température, équivalent en eau de neige), pour atteindre leur plein potentiel prédictif. Leur performance est directement liée à leur capacité à apprendre à partir de vastes échantillons de données (Kratzert et al., 2019 ; Feng, Fang et Shen, 2020). Cette forte capacité d'apprentissage a toutefois un coût : l'entraînement des LSTM peut être exigeant en calcul et en mémoire. Ils traitent des séquences et ajustent un grand nombre de paramètres, ce qui implique beaucoup d'opérations matricielles et la conservation d'états intermédiaires.

1.4 Surapprentissage

Un problème fréquent lors de la calibration des modèles hydrologiques est le surapprentissage (*overfitting*). Celui-ci survient lorsque le modèle reproduit trop fidèlement les données de la phase d'entraînement, au point d'apprendre leurs particularités ou leur bruit, ce qui entraîne une dégradation des performances lors de la validation ou du test. Ce phénomène peut être causé par une complexité excessive du modèle par rapport à la quantité de données disponibles, par une période d'entraînement trop courte ou trop homogène, ou encore par l'absence de régularisation (Tripathy et Mishra, 2024). Dans le cas des modèles d'apprentissage profond, plusieurs techniques existent pour limiter ce problème. Le *dropout*, par exemple, consiste à ignorer aléatoirement une partie des informations utilisées par le modèle pendant l'apprentissage, ce qui l'oblige à apprendre des relations plus générales (Srivastava, 2014). La

pénalisation L2, aussi appelée *weight decay*, ajoute une contrainte pendant l'entraînement pour éviter que les paramètres du modèle ne deviennent trop grands. Cela permet d'obtenir un modèle plus simple, qui généralise mieux. La version découplée du *weight decay*, quant à elle, applique cette pénalité directement sur les poids, sans l'intégrer dans le calcul du gradient de la fonction de perte, ce qui peut être plus efficace avec certains optimiseurs (Bjorck, Weinberger, & Gomes, 2021). Finalement, pour bien entraîner un modèle d'intelligence artificielle et éviter qu'il mémorise simplement les données, il est important de séparer les données en trois parties : l'entraînement, la validation et le test. Le modèle apprend d'abord avec les données d'entraînement. Ensuite, à la fin de chaque étape d'apprentissage, sa performance est évaluée avec les données de validation. Cela permet de vérifier s'il s'améliore vraiment ou s'il commence à faire du surapprentissage et apprendre les détails trop spécifiques (le « bruit ») des données d'entraînement. L'entraînement du modèle se termine lorsque les performances en validation cessent de s'améliorer. Ce principe est appelé l'arrêt anticipé (Prechelt, 2012). Enfin, l'ensemble de données pour la période test est utilisé à la toute fin pour évaluer la performance finale du modèle sur une période indépendante. Ces approches permettent d'améliorer la capacité de généralisation et de limiter les pertes de performance lorsqu'un modèle est appliqué à de nouveaux bassins ou à des périodes indépendantes.

1.5 Modélisation hydrologique régionale appliquée à la simulation des débits

La modélisation hydrologique régionale repose sur l'idée d'exploiter les données de plusieurs bassins versants pour entraîner et calibrer un modèle. Cette stratégie est particulièrement bénéfique avec les approches d'intelligence artificielle, car elle enrichit l'apprentissage par une plus grande diversité d'exemples, ce qui améliore la généralisation et permet de compenser le manque de données locales, qui s'avère être un défi majeur en apprentissage automatique où des jeux d'entraînement trop petits ou peu représentatifs conduisent souvent à du surapprentissage ou à une mauvaise généralisation. Un autre avantage de la modélisation régionale est sa capacité à réaliser des prévisions de débit dans des bassins non jaugés, en transférant l'information apprise à l'échelle régionale vers un bassin cible dépourvu de mesures de débit (Razavi & Coulibaly, 2013).

Selon le type de modèle hydrologique utilisé, les méthodes de régionalisation et les contraintes associées varient considérablement. Dans le cas des modèles conceptuels, la régionalisation repose sur l'utilisation de bassins jaugés présentant des caractéristiques physiques ou climatiques similaires à celles du bassin non jaugé, ou situés à proximité. Chaque bassin jaugé est alors calibré individuellement, et les paramètres obtenus sont transférés vers le bassin non jaugé, soit directement (transfert d'un bassin donneur), soit en les combinant (par exemple via une moyenne pondérée) selon des critères de similarité (Arsenault, Poissant, & Brissette, 2015; Arsenault and Brissette, 2016). Cette approche est relativement simple à mettre en œuvre, mais sa fiabilité dépend fortement de la correspondance entre bassins.

De leur côté, les modèles issus de l'intelligence artificielle sont plus flexibles. Il est possible de créer un modèle unique capable d'apprendre une structure commune à partir de nombreux bassins jaugés, tout en tenant compte des différences entre les bassins grâce à des variables descriptives telles que la superficie, la pente ou l'occupation des sols. Une fois entraîné, ce modèle peut être appliqué à un bassin non jaugé, à condition de disposer des données d'entrée nécessaires, comme les séries météorologiques et les caractéristiques du bassin. Lorsque le bassin cible est partiellement ou complètement jaugé, il devient possible d'inclure une partie des observations locales dans l'entraînement. Il est également envisageable, dans ce cas, de procéder à un réentraînement local (fine-tuning) afin d'adapter le modèle plus spécifiquement au bassin, que ce soit pour étendre une simulation au-delà de la période observée ou pour effectuer des prévisions hydrologiques.

Plusieurs études récentes confirment l'efficacité de ces approches régionales fondées sur l'apprentissage automatique, notamment lorsqu'elles sont comparées à des modèles hydrologiques traditionnels calibrés localement ou régionalement. Mai et al. (2022) comparent douze modèles hydrologiques calibrés localement et régionalement à un LSTM entraîné globalement sur le bassin des Grands Lacs ; ce dernier obtient des KGE médians supérieurs en validation temporelle (0,82 vs 0,79), spatiale (0,76 vs 0,61) et spatio-temporelle (0,75 vs 0,60). Le KGE (Kling-Gupta Efficiency) est un indicateur combinant biais, variance et corrélation

pour évaluer la qualité d'une simulation hydrologique, où une valeur de 1 correspond à une simulation parfaite (Gupta et al., 2009). De leur côté, Arsenault et al. (2023a) montrent que des LSTM entraînés régionalement surpassent, dans 93 à 97 % des cas, trois méthodes de régionalisation fondées sur GR4JCN, HMETs et HSAMI dans 148 bassins non jaugés en Amérique du Nord. Enfin, Koch et Schneider (2022) observent que des LSTM entraînés sur 301 bassins danois dépassent un modèle distribué à base physique calibré localement (DK) dans 80 % des cas, la médiane du KGE passant alors de 0,7 à 0,8.

Hashemi et al. (2022) prolongent l'approche combinée en regroupant des bassins hydrologiquement homogènes en France pour entraîner les LSTM, puis en ajustant localement les hyperparamètres (longueur de séquence, nombre d'unités cachées, etc.). Ils observent une amélioration significative des performances, en particulier pour les régimes à réponse lente tels que les régimes nivaux et uniformes.

En revanche, Chen et al. (2025) observent des bénéfices plus limités : le réentraînement local de modèles régionaux améliore les résultats dans 55,1 % des 472 bassins CAMELS-US étudiés, mais dégrade les performances dans les autres cas.

La littérature ne propose pas de seuil fixe de bassins « idéal », mais les gains apparaissent surtout lors du passage d'un très petit nombre à une centaine de bassins. Dans leur analyse sur 531 bassins CAMELS, Kratzert et al. (2024) montrent que la performance (NSE) augmente rapidement jusqu'à environ 100 bassins, puis continue de croître de façon moins prononcée, suggérant qu'une centaine de bassins représentatifs fournit déjà une base robuste. Ils constatent également que regrouper les bassins versants selon leurs caractéristiques hydrologiques ou physiques similaires (par exemple, en formant des groupes à partir d'attributs comme la taille, le climat ou le type de sol) permet d'obtenir des résultats plus précis et stables. En revanche, des découpages plus rigides et administratifs, comme ceux utilisés dans le système HUC aux États-Unis, donnent généralement de moins bonnes performances. Cela dit, les bénéfices liés à l'utilisation de bassins similaires dépendent fortement de la ressemblance entre les bassins utilisés pour entraîner le modèle et ceux sur lesquels il sera appliqué. Si les bassins cibles

ressemblent beaucoup à ceux de l'ensemble d'entraînement, le modèle donne généralement de bons résultats. Par contre, ces bons résultats peuvent chuter si les bassins sont très différents. Pour que le modèle soit plus stable et efficace sur une grande variété de territoires, il est donc préférable d'utiliser un ensemble d'entraînement large et diversifié, contenant des bassins de types variés. Même si cela peut parfois réduire légèrement la précision pour certains cas particuliers, cela rend le modèle plus robuste, c'est-à-dire capable de bien performer même lorsque les conditions changent. En pratique, viser une centaine de bassins constitue un bon compromis entre coût et rendement, même si intégrer davantage de bassins peut encore affiner légèrement les prévisions.

1.6 Forçages météorologiques

Les forçages météorologiques désignent les variables atmosphériques ou environnementales qui varient dans le temps et qui servent d'entrées au modèle hydrologique (par exemple, les précipitations, le rayonnement solaire, le vent ou la température). Ces données peuvent être issues de différentes sources. Elles proviennent soit d'observations in situ, mesurées directement par des instruments au sol (stations météorologiques), soit d'observations à distance, obtenues par télédétection (satellites, radars), soit de produits de modélisation numérique de la Terre, incluant les jeux de réanalyse (données pseudo-historiques) ainsi que les prévisions météorologiques. Seules les données de réanalyse et les prévisions météorologiques seront abordées en détail dans cette section, puisqu'il s'agit des sources utilisées dans le cadre de l'article présenté.

1.6.1 Données de réanalyse

Les modèles hydrologiques, qu'ils soient conceptuels, basés sur la physique ou encore fondés sur l'apprentissage automatique, nécessitent de longues séries temporelles complètes de variables météorologiques afin d'être correctement calibrés et validés. Disposer d'observations continues sur plusieurs décennies demeure toutefois un défi, car les stations météorologiques peuvent présenter des lacunes spatiales ou temporelles, notamment dans les régions isolées ou peu instrumentées. Pour pallier ces limitations, les données de réanalyse se sont avérées être

un bon substitut aux observations directes lorsqu'elles sont absentes ou incomplètes (Chen & Brissette, 2017; Essou, Brissette, & Lucas-Picher, 2017; Tarek, Brissette, & Arsenault, 2020) et elles sont aujourd'hui largement utilisées dans les applications environnementales et hydrologiques (p. ex. Sabzipour et al., 2023, Do et al., 2020; Gualtieri, 2022; Yoshimura et al., 2008). Elles résultent de la combinaison de mesures in situ ou satellitaires avec un modèle de prévision météorologique, les observations étant assimilées au modèle pour reconstituer l'état passé de l'atmosphère et produire un ensemble cohérent de variables climatiques atmosphériques, terrestres et océaniques sur une couverture spatiale et temporelle continue.

Les premières données de réanalyse atmosphérique ont été produites en 1979 dans le cadre du First GARP Global Experiment (FGGE) (Bengtsson, 1982), marquant le point de départ d'une série d'améliorations successives en résolution et en précision. Cette première génération de réanalyses a ouvert la voie à des produits plus avancés, tels qu'ERA-15 (Gibson et al., 1999), ERA-40 (Uppala et al., 2005) et ERA-Interim (Dee et al., 2011), qui ont progressivement intégré un volume croissant d'observations satellitaires, des modèles atmosphériques plus performants et des systèmes d'assimilation plus sophistiqués. Cette évolution a conduit à ERA5, publiée en 2019 par le Copernicus Climate Change Service (C3S) de l'ECMWF (Hersbach et al., 2020), actuellement la réanalyse globale la plus complète et la plus précise, et largement utilisée comme jeu de données de référence en modélisation hydrologique (Tarek et al., 2020; Cantoni et al., 2022; Armstrong et al., 2025). Distribuée sur une grille mondiale de 31 km de résolution, elle couvre la période 1940 à aujourd'hui. Plus récemment, ERA5-Land, publiée en 2020 (Muñoz-Sabater et al., 2021), a été développée comme version optimisée pour la composante terrestre. Offrant une résolution horizontale affinée à 9 km et ajustant notamment la température et l'humidité près de la surface en fonction du relief, elle améliore la représentation des variables de surface. ERA5-Land capte mieux l'hétérogénéité des précipitations et de l'humidité des sols, grâce à cette finesse spatiale et à un recalcul des flux de surface, ce qui la rend particulièrement adaptée aux études hydrologiques à l'échelle de bassins de petite et moyenne taille.

1.6.2 Prévisions météorologiques

Les prévisions météorologiques jouent un rôle central pour la société. Elles soutiennent l'agriculture, aident à anticiper les événements extrêmes comme les inondations ou les sécheresses, facilitent la planification du transport et guident la gestion des ressources énergétiques et hydriques (World Meteorological Organization & World Bank, 2015). L'idée de prévoir la météo à l'aide d'équations mathématiques remonte au début des années 1920, notamment avec les travaux de Lewis Fry Richardson (1922), mais il a fallu attendre la fin de la Seconde Guerre mondiale et l'arrivée des premiers ordinateurs pour que cette vision devienne réalité. La toute première prévision numérique de 24 heures est réalisée en 1950 (Charney, Fjørtoft & von Neumann, 1950), marquant le début de la prévision numérique du temps.

Depuis, l'évolution des modèles, des observations satellitaires et de la puissance de calcul a profondément transformé les systèmes de prévision. C'est dans ce contexte qu'est créé, en 1975, le Centre européen pour les prévisions météorologiques à moyen terme (ECMWF), organisation intergouvernementale qui joue depuis un rôle important dans le développement de la prévision numérique. Sa mission est de fournir à ses pays membres des prévisions fiables pour soutenir la sécurité des populations et la gestion des ressources face aux phénomènes extrêmes. Le centre joue un rôle à la fois opérationnel et scientifique, en développant et maintenant l'un des modèles de prévision les plus performants au monde (Haiden et al., 2024). Ses systèmes couvrent plusieurs horizons temporels : les prévisions à moyen terme (jusqu'à quinze jours), produites quotidiennement avec une résolution ensembliste d'environ 9 km, les prévisions étendues (jusqu'à 46 jours), émises deux fois par semaine avec une résolution d'environ 36km, et les prévisions saisonnières (jusqu'à sept mois), actualisées chaque mois avec la même résolution (36km). Chaque prévision est générée à l'aide d'un modèle numérique de la Terre, c'est-à-dire une représentation mathématique des processus physiques et dynamiques de l'atmosphère, de l'océan, des surfaces continentales et de la cryosphère, découpée en un maillage tridimensionnel de cellules. La prévision est initialisée à partir d'un état actuel de l'atmosphère, obtenu grâce à l'assimilation d'observations récentes issues de

satellites, de stations in situ et d'autres sources, puis simule son évolution future en résolvant les équations du modèle à intervalles réguliers. Cet état est ensuite réajusté au fil de nouvelles observations, afin de corriger les écarts et relancer la prévision (Haiden et al., 2023; Johnson et al., 2019).

Au Canada, l'organisme responsable de la surveillance et de la prévision météorologique est Environnement et Changement climatique Canada (ECCC), via son Centre météorologique canadien (CMC). Le système de prévision ensembliste global (GEPS) fournit des prévisions jusqu'à 16 jours (avec une extension à 39 jours le lundi et le jeudi) à une résolution horizontale d'environ 25 km (Deng et al., 2024b). ECCC dispose aussi de prévisions ensemblistes saisonnières avec le CanSIPS (Canadian Seasonal to Interannual Prediction System), produites mensuellement et couvrant un horizon de 12 mois avec une résolution d'environ 100 km (Merryfield et al., 2013).

1.7 Prévision hydrologique

Les prévisions hydrologiques font généralement référence à l'anticipation du débit naturel futur des cours d'eau. Toutefois, d'autres variables peuvent également faire l'objet de prévisions, comme le niveau d'eau ou le volume d'eau entrant dans un réservoir. Ces prévisions sont essentielles pour gérer les ressources en eau et limiter les risques. Elles servent, par exemple, à anticiper les crues et prévenir et préparer les citoyens en cas d'inondations, à mieux réguler les réservoirs et la production hydroélectrique, à sécuriser l'approvisionnement en eau potable et l'irrigation, ou encore à préserver les débits nécessaires au bon fonctionnement des écosystèmes aquatiques.

L'horizon temporel des prévisions hydrologiques n'est pas strictement défini et varie en fonction des objectifs visés. De manière générale, les prévisions de court terme, de quelques heures à quelques jours, sont surtout utiles pour les petits bassins à réponse rapide pour réagir face aux crues soudaines (Saavedra Valeriano et al., 2010) ou pour prendre des décisions de gestion hydroélectrique en fonction de la disponibilité et de l'efficacité des groupes turbine-

alternateur et de la demande d'énergie. Celles de moyen terme (plusieurs jours à quelques semaines) et à long terme (plusieurs mois) soutiennent la planification stratégique de l'exploitation des grands réservoirs, par exemple pour anticiper des périodes de sécheresse ou organiser la gestion de l'eau à grande échelle (Balthazar et al., 2024; Turner, Xu, & Voisin, 2020). Par exemple, Deng et al. (2024a) ont comparé un modèle LSTM sur un seul bassin à un modèle hydrologique déterministe basé sur les processus physiques (`wflow_sbm`) pour la station de Lobith sur le Rhin. Le modèle LSTM s'est montré plus performant pour la détection des sécheresses à courte échéance, en particulier entre 1 et 11 jours, tandis que les deux modèles affichaient des performances similaires entre 11 et 21 jours. Pour les prévisions à plus long terme, soit au-delà de 31 jours et jusqu'à un horizon de 46 jours, `wflow_sbm` donnait de meilleurs résultats, ce qui souligne l'importance d'adapter l'outil de prévision à l'horizon temporel visé.

Les prévisions hydrologiques peuvent être regroupées en trois grandes catégories (Troin et al., 2021). La première regroupe les prévisions basées sur des méthodes statistiques (SBSP, *statistical-based streamflow prediction*), telles que les modèles d'apprentissage automatique ou les modèles autorégressifs. Ces méthodes reposent sur l'analyse des débits passés et, selon le cas, de variables météorologiques ou climatiques afin d'estimer les débits futurs. La deuxième catégorie correspond à l'approche ESP (*Ensemble Streamflow Prediction*), introduite par Day (1985), qui utilise l'état hydrologique initial du bassin (neige, humidité du sol, etc.) combiné à des séquences météorologiques historiques. Ces séquences sont interprétées comme des scénarios plausibles du climat futur, ce qui permet de générer un ensemble de trajectoires de débits. Enfin, la troisième catégorie regroupe les prévisions hydrologiques forcées par des prévisions météorologiques numériques, connues sous le nom de EPS (*Ensemble Prediction Systems*). Cette approche s'appuie sur les sorties de modèles météorologiques numériques, qui fournissent des prévisions du temps à venir, utilisées comme forçages dans les modèles hydrologiques (Cloke et Pappenberger, 2009).

1.7.1 Prévisions déterministes et ensemblistes

Il existe deux types de prévisions : les prévisions déterministes, qui reposent généralement sur un unique scénario météorologique et produisent une seule estimation des débits, et les prévisions probabilistes, qui génèrent un ensemble de scénarios équiprobables. Une première approche concernant les prévisions probabilistes consiste à recourir à des prévisions météorologiques d'ensemble, composées de plusieurs scénarios (appelés membres), obtenus en perturbant légèrement les conditions initiales d'un modèle atmosphérique. Chaque membre donne lieu à une trajectoire possible de l'évolution des variables météorologiques (telles que la température ou les précipitations), lesquelles, une fois introduites dans un modèle hydrologique, produisent un éventail de débits futurs. Une deuxième approche consiste à perturber les états initiaux du modèle hydrologique lui-même (par exemple, le stockage de neige ou l'humidité des sols), en utilisant un forçage météorologique déterministe ou probabiliste, afin de représenter l'incertitude liée aux conditions hydrologiques de départ. Enfin, une troisième approche consiste à combiner plusieurs séries de débits déterministes issues de différents modèles hydrologiques, ou en agrégeant plusieurs prévisions probabilistes produites par des modèles distincts. Ces approches multi-modèles permettent de mieux représenter l'incertitude structurelle liée au choix des modèles hydrologiques et d'obtenir des prévisions de meilleure qualité, comme l'ont montré Dion, Martel & Arsenault (2021). Toutefois, cela engendre un temps de calcul plus coûteux (Jiang et al., 2014).

Selon Krzysztofowicz (2001), une seule possibilité de prévision donne l'illusion de la certitude, qui peut mener à des erreurs coûteuses, voire parfois irréparables, si la réalité diverge du scénario anticipé. La prévision probabiliste, quant à elle, permet de mieux représenter l'incertitude inhérente aux prévisions et d'anticiper divers comportements hydrologiques plausibles sur la période ciblée, offrant ainsi une vision plus complète des risques. Elle favorise une gestion plus flexible des réservoirs, en contribuant à optimiser la production d'énergie hydroélectrique tout en augmentant les bénéfices qui y sont associés. En outre, la possibilité de prendre en compte les événements extrêmes améliore la capacité des gestionnaires à s'y préparer efficacement, réduisant ainsi les risques liés aux crues ou aux pénuries d'eau.

1.7.2 Incertitudes en prévision hydrologique

Des incertitudes sont présentes tout au long de la chaîne de modélisation et de prévision hydrologique, à commencer par les données d'entrée fournies aux modèles hydrologiques. Les données historiques utilisées pour calibrer les modèles peuvent comporter des lacunes ou des erreurs, notamment en raison des limites instrumentales. Certains capteurs présentent intrinsèquement une incertitude plus élevée, par exemple ceux qui mesurent l'équivalent en eau de la neige à l'aide de pluviomètres chauffants (Grossi et al., 2017). De plus, au fil du temps, les bris, les remplacements d'appareils ou les changements de site peuvent modifier la cohérence des séries, nécessitant un traitement statistique préalable. Les données diffusées publiquement ont généralement déjà subi une série d'étapes de contrôle de qualité, de correction et d'homogénéisation pour limiter ces sources d'erreur. Les données pseudo-historiques issues de réanalyses, comme ERA5, ne sont pas exemptes d'incertitudes : elles proviennent d'un modèle numérique de l'atmosphère qui intègre les observations disponibles, mais qui reste soumis à des biais de modélisation. ERA5 fait d'ailleurs l'objet d'un post-traitement avant diffusion, ce qui réduit certaines erreurs mais n'élimine pas l'incertitude. Il peut être pertinent de post-traiter localement les données d'entrée provenant de réanalyses pour un bassin versant, car des biais régionaux peuvent subsister et l'échelle spatiale ou temporelle des données disponibles ne correspond pas toujours à celle requise par le modèle (Probst & Mauser, 2022).

Les prévisions météorologiques introduisent également de l'incertitude, qui augmente avec l'horizon temporel : elles sont généralement fiables sur quelques jours, mais leur performance diminue rapidement à mesure que l'échéance s'allonge, en raison du caractère instable et non linéaire de la dynamique atmosphérique. Les ensembles multi-membres, obtenus en perturbant les conditions initiales pour générer plusieurs trajectoires de prévision, constituent un outil efficace pour quantifier l'incertitude. Néanmoins, ces ensembles demeurent sujets à des biais systématiques, qui peuvent être corrigés par des méthodes statistiques telles que le quantile mapping ou les régressions (Aguilar Andrade et al., 2024; Anghileri et al., 2019; Fang et al., 2015), ou, plus récemment, par des techniques d'apprentissage automatique (Zhang & Ye,

2021; Zarei, Najarchi & Mastouri, 2021). À ces sources s'ajoutent les incertitudes propres aux modèles hydrologiques, liées à leur structure, aux simplifications conceptuelles, à la représentation spatiale des processus, aux paramètres calibrés, et aux conditions initiales. Dans le cas des modèles d'intelligence artificielle, il n'y a pas de paramètres calibrés au sens traditionnel, puisque le modèle apprend lui-même les relations entre les différentes variables à partir des données. Cependant, certains réglages du modèle, appelés hyperparamètres, comme la taille du réseau, la vitesse à laquelle il apprend ou la manière dont il transforme les données à chaque étape, influencent directement la façon dont il interprète et assimile l'information. Ces choix, effectués par l'utilisateur, introduisent une part d'incertitude, car différentes combinaisons peuvent conduire à des résultats très différents. L'exploration des différentes combinaisons possibles est essentielle afin de trouver la meilleure combinaison possible. Il a d'ailleurs été démontré qu'une recherche aléatoire dans l'espace des hyperparamètres est souvent plus efficace que de tester systématiquement toutes les combinaisons possibles à intervalles réguliers, car elle permet d'explorer plus rapidement des configurations variées et potentiellement performantes (Bergstra & Bengio, 2012).

Enfin, des approches complémentaires comme l'assimilation continue de nouvelles observations (Liu et al., 2012; Magnusson et al., 2020; Samuel et al., 2019), la modélisation multi-modèles, qui consiste à utiliser plusieurs modèles hydrologiques distincts (Dion et al., 2021; Solanki et al., 2025), ou un post-traitement adapté contribuent à limiter l'impact de ces incertitudes et à améliorer la fiabilité des prévisions (Li et al., 2017; Matthews et al., 2022; Solanki et al., 2025; Xu, Anctil & Boucher, 2022).

1.7.3 Prévision avec un modèle régional

Les études en lien avec la modélisation régionale reposent essentiellement sur des simulations de débit, et leur application à la prévision des débits reste encore peu documentée. He et al. (2025) ont évalué quatre architectures d'apprentissage profond sur 596 bassins issus de l'ensemble de données CAMELS. Pour des horizons de prévision allant jusqu'à sept jours, ils montrent que l'approche multi-bassins, enrichie par un réentraînement local, surpasse

l'entraînement isolé sur un seul bassin. En revanche, l'entraînement local traditionnel affiche de meilleures performances que la modélisation régionale non ajustée localement. Une autre étude, menée par Quang et al. (2025), a comparé cinq architectures d'intelligence artificielle selon trois configurations : un modèle local, un modèle régional, et un modèle régional suivi d'un réentraînement local. L'expérience, réalisée à partir de données issues de 11 stations météorologiques dans le bassin transfrontalier du fleuve Rouge (Chine, Laos, Vietnam), portait aussi sur des prévisions journalières à 7 jours d'échéance. Les meilleurs résultats ont été obtenus pour des prévisions à un ou deux jours, avec une dégradation marquée au-delà de trois jours. L'approche régionale-locale combinée à un modèle CNN-LSTM s'est révélée être la plus performante, offrant la meilleure concordance avec les observations sur l'ensemble des stations.

1.8 Prévion des volumes d'eau

Dans le contexte de la gestion des réservoirs hydroélectriques, la maîtrise des volumes d'apport est essentielle pour assurer un fonctionnement optimal du système. D'une part, il est impératif de ne pas dépasser la capacité maximale du réservoir, afin d'éviter tout risque de surverse ou d'inondation en aval et sur les berges. D'autre part, un volume élevé est généralement recherché, car une hauteur d'eau plus importante permet une meilleure production hydroélectrique pour un même débit turbiné. Il est également nécessaire de conserver un volume minimal, notamment dans des réservoirs naturels régulés comme le Lac Saint-Jean, afin de maintenir un niveau d'eau adéquat en amont pour répondre aux usages riverains, tout en assurant un débit suffisant en aval, conformément aux exigences environnementales et réglementaires (Yin, Yang et Petts, 2011).

Quelle que soit la complexité des approches modernes utilisées pour optimiser les opérations hydroélectriques (simulations avancées, modèles d'optimisation ou contrôles prédictifs), la capacité à anticiper et ajuster dynamiquement les volumes stockés et libérés en fonction des conditions futures demeure au cœur des décisions, comme le soulignent Becker, Kim et Pummer (2023).

Actuellement, les exploitants estiment les volumes d'apport en générant des prévisions de débit, puis en agrégeant ces débits sur la période souhaitée (journée, semaine, saison) (Herbert et al., 2021; Swift-LaPointe, White, & Radić, 2025; Viterbo et al., 2020). Dans les petits bassins sans réservoir ou avec un faible stockage, les prévisions à court terme sont particulièrement utiles pour la gestion quotidienne, en lien avec la demande énergétique : combien de turbines doivent être activées, à quel moment, et à quel régime de fonctionnement ? Elles permettent également de réagir rapidement aux événements extrêmes, en facilitant l'allocation dynamique de l'eau vers les turbines ou les déversoirs en cas de crue imminente, lorsque les apports excèdent la capacité de stockage. À l'inverse, les grands bassins équipés de réservoirs importants ou influencés par des processus à réponse lente, comme l'accumulation et la fonte nivale, bénéficient davantage de prévisions à moyen et long terme. C'est notamment le cas dans les climats nordiques, comme le Canada, où la neige s'accumule en hiver puis fond rapidement au printemps, générant un pic de volume important (Magnusson et al., 2020). Dans ces contextes, anticiper de manière fiable et précise les volumes saisonniers permet non seulement d'optimiser la gestion du remplissage et la production hydroélectrique, mais aussi de mieux se préparer aux crues saisonnières et aux épisodes de faibles apports. Ces prévisions renforcent la capacité des gestionnaires à planifier les ajustements nécessaires et à prendre des décisions fiables et adaptées face à l'incertitude hydrologique. Par exemple, si les prévisions annoncent une importante fonte nivale dans les semaines à venir, il peut être pertinent de commencer à abaisser progressivement le niveau du réservoir afin de libérer de la capacité de stockage. Cette démarche permet d'anticiper l'arrivée de volumes d'eau élevés et de limiter les risques de débordement ou d'inondation en aval. À l'inverse, si les prévisions indiquent un faible apport en eau pour les mois suivants, notamment en contexte de sécheresse, les décisions prises à court terme viseront plutôt à conserver un maximum d'eau en réserve, par exemple en réduisant temporairement les débits turbinés. Dans les deux cas, l'objectif est d'utiliser chaque goutte d'eau au moment le plus opportun, en tenant compte à la fois des contraintes opérationnelles et des conditions futures. Cependant, la question du moment précis de l'arrivée de l'eau peut amener à penser qu'il serait préférable de disposer de prévisions détaillées jour par jour, afin de savoir exactement quelle quantité d'eau est attendue et à quelle date. En réalité, dans le cas de réservoirs de grande capacité, comme ceux exploités par Rio-Tinto, la différence

entre une arrivée au jour 35 ou au jour 38 a peu d'importance. La gestion des volumes se fait de manière progressive, en ajustant quotidiennement les décisions selon les conditions observées, tout en maintenant les objectifs de remplissage à plus long terme. Ce qui importe surtout est de connaître le volume total attendu sur un horizon pertinent, que ce soit 30, 60 ou 90 jours, selon le contexte saisonnier et les objectifs de gestion. Cette information permet de guider les grandes décisions, même si la date précise d'arrivée de l'eau varie légèrement. Ainsi, les prévisions à moyen ou long terme orientent directement la gestion opérationnelle à court terme, en fournissant un cadre stratégique pour ajuster les volumes libérés ou stockés en fonction des conditions anticipées.

De nombreux travaux se sont penchés sur la prévision des débits, en explorant une grande diversité d'approches et d'horizons temporels, dans le but d'améliorer la qualité des outils de prévision hydrologique (Armstrong et al., 2025; Bennett, Wang, Li, Robertson, & Schepen, 2016; Deng et al., 2024; Dion et al., 2021; Sabzipour et al., 2023). Or, ces prévisions sont souvent utilisées pour estimer les volumes d'apport, qui constituent une information clé dans la prise de décision en contexte hydroélectrique. Cela soulève une question importante : pourquoi ne pas chercher à prévoir directement ces volumes, plutôt que de les obtenir par traitement ultérieur des prévisions de débit? Une telle approche permettrait de cibler directement la variable utilisée en gestion opérationnelle, tout en réduisant potentiellement les incertitudes liées à l'agrégation des débits ainsi que la complexité et le coût du processus de prévision.

Dans le cadre d'un projet connexe, Soucy (2025) a récemment appliqué un modèle LSTM à la prévision directe des volumes d'apport sur un horizon de 14 jours. Son approche s'appuie sur l'intégration de rétro-prévisions météorologiques afin d'augmenter le nombre d'échantillons disponibles pour l'entraînement. Les résultats montrent que les modèles LSTM offrent des prévisions précises mais sous-dispersées, avec une fiabilité limitée, et que l'intégration de rétro-prévisions améliore la précision au prix d'une fiabilité réduite, notamment au printemps et en hiver, où les conditions hydrologiques sont les plus complexes.

À ce jour, aucune autre étude ne semble avoir exploré la prévision directe des volumes pour des horizons plus éloignés, malgré le fait que les volumes représentent un indicateur clé pour la gestion opérationnelle des ressources en eau. Cela est probablement dû au fait que les modèles hydrologiques traditionnels ne permettent pas de modéliser directement des volumes, alors que les modèles d'apprentissage profond récents ne sont pas obligés de simuler les débits journaliers pour estimer des volumes directement.

1.9 Objectifs de recherche

L'objectif principal de cette étude est d'améliorer la prévision des volumes de crue cumulés sur 45 jours à partir d'un large éventail de données issus de 200 bassins versants au Canada. Plus précisément, il s'agit d'évaluer la valeur ajoutée d'un modèle LSTM entraîné de manière régionale, sur plusieurs bassins, par rapport à un modèle local, sachant que les LSTM sont particulièrement performants lorsqu'ils disposent de grandes quantités de données. L'étude vise également à tester la possibilité d'entraîner directement un modèle à partir de prévisions météorologiques d'ensemble. La méthodologie est appliquée à huit bassins versants situés dans différentes provinces du Canada, permettant ainsi d'examiner la capacité des modèles à se généraliser à des contextes présentant des caractéristiques physiographiques et hydrologiques variées. Afin de mieux refléter l'hétérogénéité des conditions hydrologiques et climatiques selon les saisons, les résultats sont analysés de manière saisonnière à l'aide de plusieurs critères de performance. Par ailleurs, pour mettre en perspective les résultats, les prévisions obtenues pour le bassin du Lac-Saint-Jean sont comparées à celles générées par le modèle conceptuel actuellement utilisé par Rio Tinto, un acteur majeur de l'industrie mondiale de l'aluminium, dans la gestion opérationnelle de ses centrales hydroélectriques. Enfin, une analyse multi-modèle est également explorée afin d'examiner la complémentarité potentielle des différentes approches et leur valeur ajoutée en contexte opérationnel.

CHAPITRE 2

DEEP LEARNING APPLIED TO SEASONAL INFLOW VOLUME FORECASTS IN HYDROELECTRIC RESERVOIRS

Dylane Martel¹, Richard Arsenault¹, Frédéric Talbot^{1,2}, Jean-Luc Martel¹

¹ Hydrology, Climate and Climate Change Laboratory, École de technologie supérieure,
Université du Québec, Montréal, H3C 1K3, Canada

² Rio Tinto Aluminum, Saguenay, G7S 4L2, Canada

Article soumis pour publication à la revue « Journal of Hydrology » le 4 octobre 2025

2.1 Abstract

Hydropower reservoirs require inflow forecasts to allow water resources managers to optimize drawdown rates and improve infrastructure efficiency. Usually, operators use physically-based or conceptual hydrological models to forecast streamflow for a lead-time ranging from one month to a year, depending on the planning horizon. These forecasts are then aggregated to evaluate cumulative inflow volumes, which are essential for seasonal operations. Reliable medium-range volume forecasts over a few weeks support not only the optimization of reservoir filling and hydropower generation, but also preparedness for seasonal floods and low-flow periods.

To address these needs more directly, this study presents a novel method that directly forecasts 45-day cumulative inflow volumes using a deep learning model trained on ensemble meteorological forecasts and observed volume data, avoiding the need to simulate daily streamflow and targeting the total volume over the full 45-day period in a single operation. Furthermore, the model leverages large-scale datasets during its training, by using data from 200 basins in Canada. The model is then applied to eight Canadian basins to estimate inflow volume forecasts. An operational conceptual hydrological model is also compared to these LSTM models for the Lac-St-Jean hydropower system in Québec, Canada.

Results show that LSTM models can achieve good forecasting skill for 45-day inflow volumes, with forecasting performance influenced by basin characteristics, seasonal conditions, and the training strategy used. While accuracy is generally high, reliability remains a challenge, as ensemble forecasts tend to be overconfident and underdispersed. Regional training improves performance in some cases, but no universal pattern emerges across all seasons or basins. Compared to the operational conceptual model, LSTM models outperform it during the autumn season, both in terms of accuracy and reliability, highlighting their complementary potential. Overall, this study suggests that LSTM-based models offer a promising alternative for strategic reservoir planning, particularly when combined with physically-based models, to improve uncertainty representation.

2.2 Introduction

Streamflow forecasting is essential for the effective management of reservoirs and dams, supporting both operational decisions and long-term planning. In practice, operators typically use conceptual or physically based hydrological models to forecast streamflow at a given time step, then assess the total expected inflows over the forecast period (Herbert, Asghar, & Oroza, 2021; Viterbo et al., 2020). These forecasts help manage risks related to water supply variability and optimize hydroelectric power production. For example, anticipating high inflows allows operators to pre-emptively lower reservoir levels, maximizing available storage, preventing uncontrolled overflows, and harnessing future inflow energy. Conversely, in anticipation of drought, more water can be retained to secure long-term energy supply. Forecasts are also essential for flood and drought risk management, protecting infrastructure and ecosystems from extreme events (Kundzewicz et al., 2003), and for improving hydropower planning and risk mitigation (Graham et al., 2022).

The distinction between short-, medium-, and long-term forecasts is not always clear-cut, but they are often described as follows: short term refers to lead times of one to several days, medium-term forecasts extend up to several weeks, and long-term forecasts cover periods of several months. Short-range forecasts support daily operations by enabling quick adjustments

to reservoir levels and outflows, while medium- and long-term forecasts play a strategic role in seasonal planning, especially during winter or in regions with high hydrological variability (Turner, Xu, & Voisin, 2020). The relevance of each forecast range depends heavily on the size, location, and operational context of the basin. For small basins with little or no storage capacity and fast hydrological response times, short to medium range forecasts may be sufficient. In contrast, larger basins with significant reservoirs, or those influenced by slow-response processes such as snow accumulation and melt, benefit more from reliable long-term forecasts. This also varies geographically. In snow-free regions such as parts of Africa, long-term forecasting often focuses on drought anticipation and seasonal rainfall variability. In northern climates like Canada, the main challenge lies in representing snow processes and their delayed impact on streamflow. This study focuses on the role of long-term forecasting in Canadian basins, where extended winters and substantial snow accumulation strongly influence inflow dynamics and reservoir operations.

Despite their proven utility, current hydrological models still face limitations, particularly when it comes to extrapolating long-term meteorological trends (Arnal et al., 2018; Peñuela, Hutton, & Pianosi, 2020; Duethmann, Blöschl, & Parajka, 2020). Persistent biases, both in meteorological forecasts (e.g., over- or underestimation of temperature and precipitation) and within the modeling chain (e.g., related to model structure or parameterization), make these forecasts less reliable for optimal water resource management. This is especially critical in regions like Canada, where snow accumulation and melt conditions complicate reservoir operations. It is therefore crucial to improve the quality of long-term forecasts, especially during the winter period, to maximize the efficiency of hydroelectric facilities year-round in basins characterized by long winter and a significant accumulation. Recent advances in long-term forecasting have focused on techniques such as data assimilation, statistical post-processing, pre-processing of meteorological ensembles, and multi-model forecasting strategies (Troin et al., 2021). While these approaches have shown value, they remain constrained by the structural limitations of traditional hydrological models. Changing the model itself could offer another avenue for improvement, but this was difficult to realize in

practice, as conventional models provided limited gains and usually traded their strong and weak points depending on the models.

The emergence of artificial intelligence, and particularly deep learning, has opened new perspectives for hydrological modeling. Among these, the development of Long Short-Term Memory (LSTM) networks by Hochreiter and Schmidhuber (1997) marked a major milestone in sequence modeling, as they were designed to capture both short and long term dependencies in time series. This makes them particularly well suited to hydrological systems, where processes such as snow accumulation and melt, or groundwater recharge, can have delayed impacts on streamflow. Although LSTMs were introduced in the late 1990s, their transformative impact on hydrology only materialized two decades later, notably around 2018, when high-resolution meteorological data, sufficient computational resources, and large-scale basin databases such as CAMELS, which includes 671 basins in the United States, became widely accessible. Several studies have confirmed the ability of LSTMs to outperform traditional hydrological models, especially in contexts marked by strong hydrometeorological variability such as simulation experiments (Kratzert et al., 2018), predictions in ungauged basins (Arsenault et al., 2023a), flood frequency analyses (Martel et al., 2024), and climate change impact assessments (Martel et al., 2025). These models, however, require large volumes of data, including numerous basins, long observational time series, and multiple relevant hydro-meteorological variables (e.g., precipitation, temperature, snow water equivalent), to reach their full predictive potential. Their performance is directly linked to their capacity to learn from vast swaths of data samples (Kratzert et al., 2019; Feng et al., 2020).

Given that deep learning models require substantial data to perform well, regional modeling offers a promising solution for poorly gauged basins while allowing models to learn hydrological dynamics representative of a broader region (Mic, Galéa, & Javelle, 2002; Li et al., 2022). Moreover, even in well-gauged basins with long historical records, regional models can still yield better performance than locally trained models, due to the added value of learning from diverse hydrological conditions across multiple basins. Kratzert et al. (2018) demonstrated that an LSTM model pre-trained on 241 basins from the CAMELS dataset in the

United States, and then fine-tuned on a target basin, outperformed an LSTM trained solely on the target basin. Their approach achieved $NSE > 0.8$ in 27.4% of basins, compared to only 17.4% for the benchmark SAC-SMA hydrological model coupled with Snow-17. Koch and Schneider (2022) applied LSTM models to 301 Danish basins and found that they outperformed the physically based distributed DK hydrological model in 80% of cases, increasing the median KGE from 0.7 to 0.8. Mai et al. (2022) compared twelve hydrological models calibrated at local and regional scales to a globally trained LSTM model over the Great Lakes basin. The LSTM achieved higher median KGE scores across all validation types: temporal (0.82 vs. 0.79), spatial (0.76 vs. 0.61), and spatiotemporal (0.75 vs. 0.60). Finally, Arsenault et al. (2023a) evaluated the performance of LSTM neural networks for continuous streamflow prediction in 148 ungauged basins in northeastern North America. Compared with three hydrological regionalization methods based on the GR4J-CN, HMETs, and HSAMI models, LSTM outperformed traditional approaches in 93% to 97% of basins. Most studies on regional modeling in hydrology have focused on streamflow simulation, with relatively few exploring its application to forecasting. He et al. (2025) evaluated four deep learning architectures across 596 CAMELS basins and found that a regional model with local fine-tuning outperformed both single-basin training and purely regional models for lead times up to seven days. Similarly, Quang et al. (2025), using data from 11 meteorological stations in the Red River basin, showed that a CNN-LSTM model combining regional training with local retraining achieved the best performance, particularly for 1- or 2-day forecasts, with accuracy declining noticeably beyond three days.

Recent studies have shown promising results for using deep learning to forecast streamflow or inflow volume forecasts. A study conducted by Sabzipour et al. (2023) on the same Lac-Saint-Jean watershed as the present study found that LSTM-based streamflow forecasts were more reliable up to 7 days and more accurate up to 9 days than those from a semi-distributed model with data assimilation (also included in this study). At longer lead times, Herbert et al. (2021) trained four LSTM models on historical SWE (snow water equivalent) and inflow data from a mountainous basin in northeastern Utah (USA), aiming to predict inflow volumes up to four months ahead. Their results showed that LSTM models could rival the ESP (Ensemble

Streamflow Prediction) approach based on the HBV hydrological model structure and outperform traditional statistical models such as SARIMA, VAR, and TBATS for long-term forecasts. On average, the LSTM encoder–decoder model improved accuracy by about 50% over ESP. Nevertheless, research in this field remains limited, particularly in operational northern contexts.

LSTM clearly showed improvement over traditional hydrological modelling in many contexts, notably forecasting. However, a potential advantage that remains underexplored is bypassing daily streamflow prediction and its post-processing to forecast inflow volumes directly at the operational lead time. This study investigates that idea. The work focuses on the Lac-Saint-Jean basin and snowmelt-dominated catchments, where reliable long-lead inflow forecasts are essential for reservoir management. It proposes an LSTM-based methodology that learns directly from cumulative inflow volumes over a fixed lead time, aligning the target variable with operational needs. The study also evaluates the feasibility of a regional LSTM trained across multiple basins with diverse hydrological characteristics to assess transferability and regional applicability.

2.3 Study area and data

2.3.1 Main study basin

The study area primarily focuses on the Lac-Saint-Jean (LSJ) basin, located in the province of Quebec, Canada (Figure 2.1, green dot in Figure 2.3). This basin, with a surface area of approximately 45,000 km², is operated by Rio Tinto, an aluminum producer that generates its own electricity to meet the energy demands of its industrial activities (Arsenault & Côté, 2019). The basin is characterized by a snowmelt-dominated hydrological regime (Figure 2.2). Runoff drains into a 1,053 km² reservoir with a storage capacity of 4,550 hm³, which supplies the Isle-Maligne hydropower generating station, with an installed capacity of 448 MW. In terms of land use, the basin is predominantly covered by forests (72.5%), followed by water bodies (9.7%), shrublands (9.3%), grasslands (3.3%), and wetlands (1.9%). Agricultural lands account for about 2.6% of the area, while urban areas represent only 0.6%.

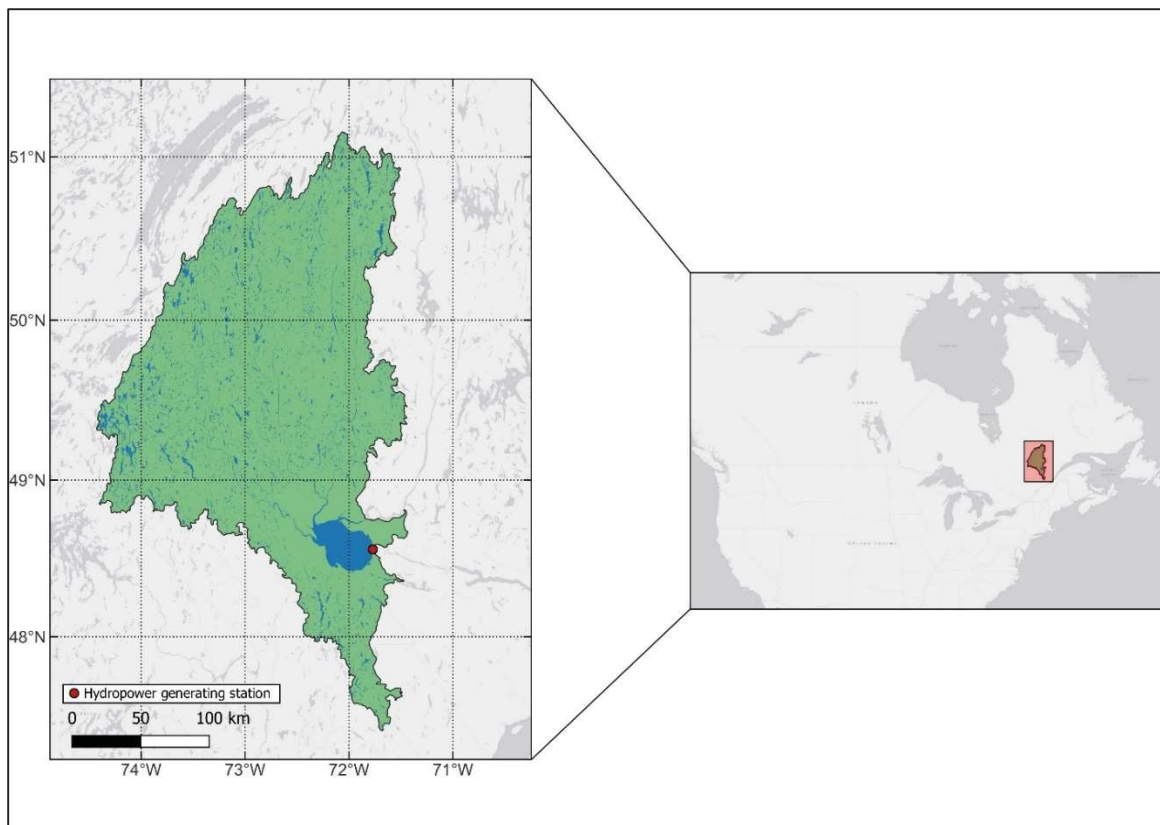


Figure 2.1 Location of the Lac-Saint-Jean basin in Canada, as well as the hydropower generating station at the reservoir outlet

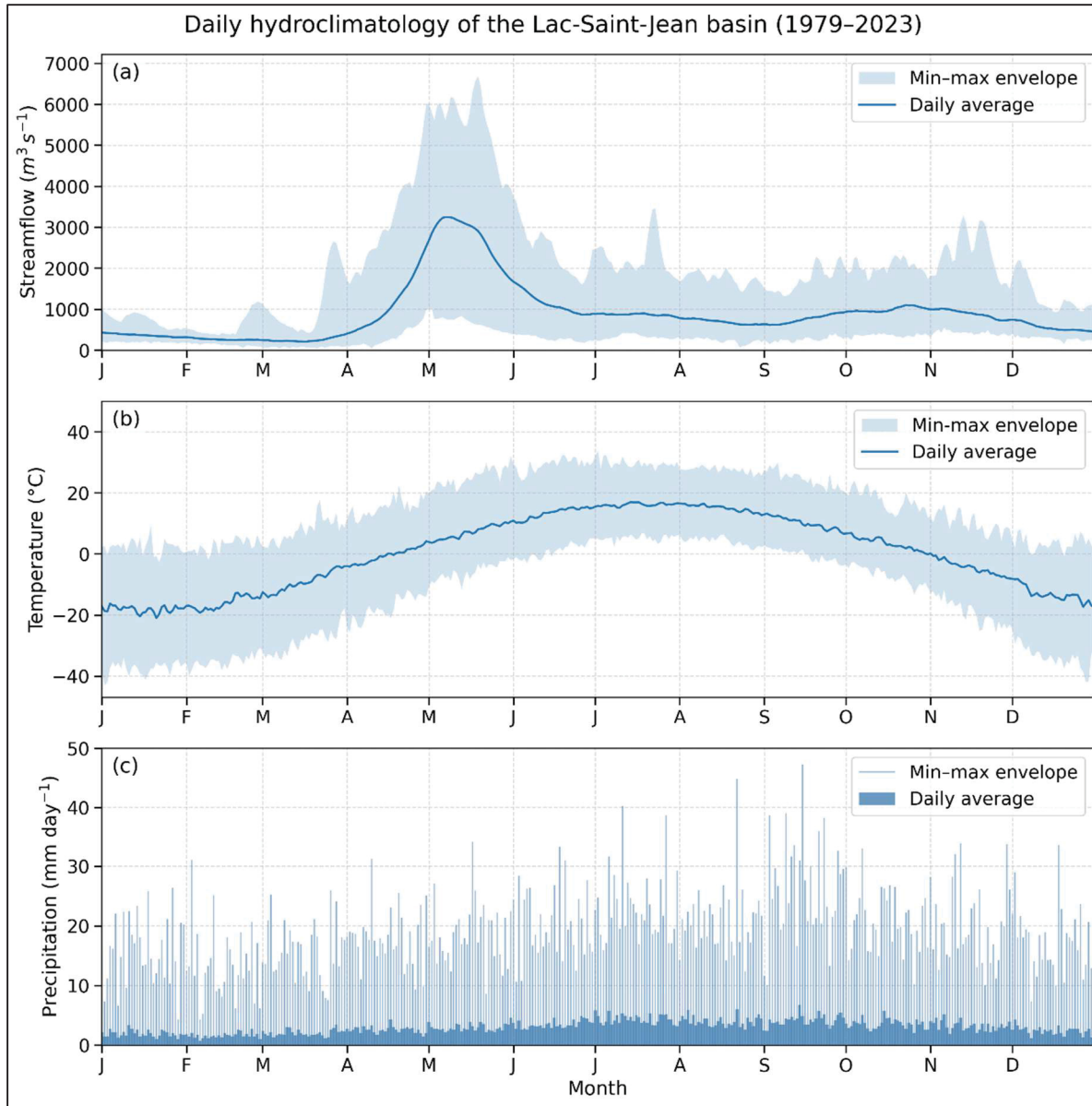


Figure 2.2 Daily average and minimum to maximum range hydroclimatology for the Lac-Saint-Jean basin (1979 to 2023): (a) Streamflow; (b) Air temperature; (c) Total precipitation. The dominant spring peak in discharge reflects seasonal snowmelt, with a moderate autumnal rise also evident, largely attributed to increased rainfall and reduced evapotranspiration

2.3.2 Regional training and complementary analysis basins

Figure 2.3 displays the set of 200 basins in Canada that were used to train the regional LSTM model, which includes the LSJ basin (outlined in green) and the seven test basins (outlined in

red). These basins were selected from the HYSETS database, which contains hydrometeorological data for over 14,000 basins in North America (Arsenault et al. 2020). The selection of these 200 basins was based on the availability of discharge data, favoring those with the most available records, and aimed to achieve a spatial distribution as equitable as possible across the different Canadian provinces, so that the regional model could learn from the full diversity of hydroclimatic conditions. The methodology is then tested on seven basins, selected based on the availability of meteorological forecast data, the presence of continuous discharge data over the entire forecast horizon, and their geographic spread across several provinces. These test basins were chosen to represent a wide range of hydrological regimes across Canada and to assess the robustness of the LSTM models on basins that differ from LSJ and exhibit varied dynamic and static (i.e. geophysical) characteristics. All other basins, used exclusively for regional model training, are represented by blue dots. Only basins from Canada are used due to the limited availability of weather forecasting data.

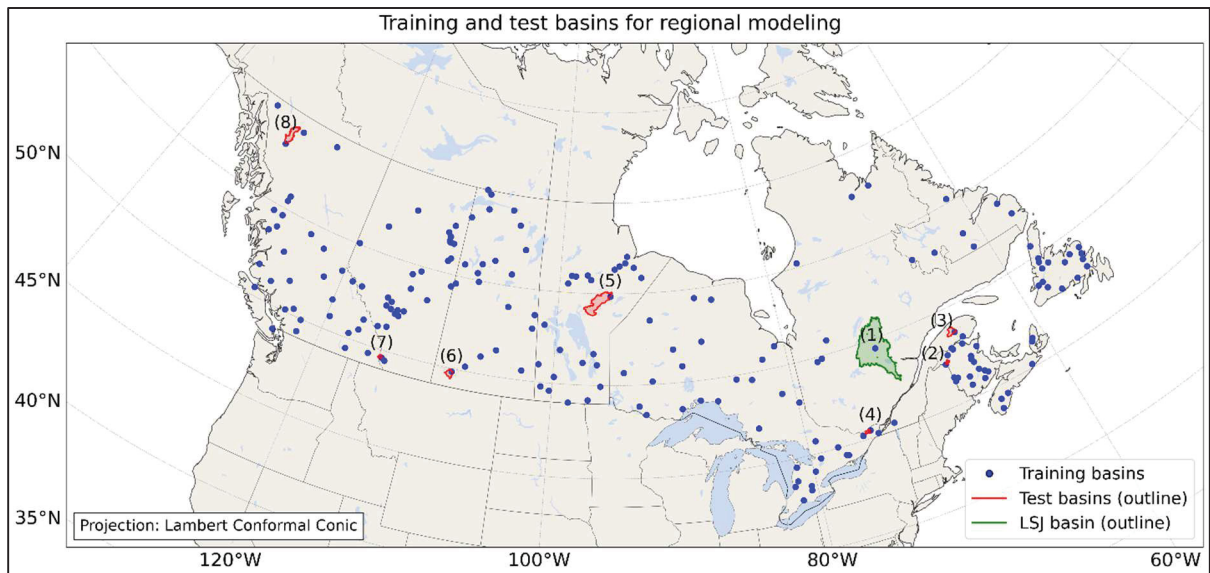


Figure 2.3 Location of the 200 basins distributed across the provinces of Canada. The green basin outline represents the LSJ basin and the seven red outlines show the locations of the basins used to test the method generalization. All eight outlined basins are part of the 200 basins used for model training

The seven Canadian test basins exhibit predominantly nival flow regimes, with varying levels of rainfall influence. Basins 4 and 6 display classic nival patterns, characterized by a single dominant spring freshet. However, at basin 4, flows drop more slowly after the spring freshet, showing that summer rain has a bigger influence on the flow regime. Basins 5, 7 and 8 also follow nival regimes, although their peak flows occur later in the summer, likely due to the delayed snowmelt associated with continental subarctic climates. Basins 2 and 3 show similar nival dominance but with additional autumn flow peaks, suggesting a more substantial contribution from rainfall during the fall season. These seasonal dynamics are illustrated in the hydrographs of all eight test basins (see Figure-A I-1-1 in Appendix I).

Tableau 2.1 Characteristics details of the study basins (1979–2023)

N ^o	Basin name	Area [km ²]	Elevation [m]	Slope [deg]	Annual mean total precipitation [mm]	Annual mean temperature [deg C]	Annual mean discharge [m ³ s ⁻¹]	Available discharge years
1	Lac-Saint-Jean	44,877.19	404.7	3.41	1,107.1	0.36	893.54	44
2	Grande River	328.44	266.7	3.16	1,217.8	3.61	7.44	42
3	Cascapedia River	1,652.36	479.2	8.26	1,234.5	0.83	43.45	43.2
4	Bear Brook	448.17	75.3	0.61	1,080.89	6.26	6.16	43.2
5	Hayes River	9,190.87	221.2	0.66	608.69	-0.94	45.01	41.1
6	Swift Current Creek	1,433.43	975.9	1.47	463.8	4.17	1.07	44
7	Crowsnest River	402.65	1,703.1	14.94	649	1.22	4.54	44
8	Tuya River	3,554.87	1,193.4	4.98	732.25	-2.61	36.9	42.9

2.3.3 Reanalysis data

The meteorological data for each basin was taken from the HYSETS database (Arsenault et al., 2020). It consists of ERA5 reanalysis data (Hersbach et al., 2020) averaged at the basin scale and on a daily timestep from 1979 to 2023 inclusively, which was used to train the LSTM models in this study. It has been demonstrated by Tarek, Brissette, and Arsenault (2020) that these data can be used as a replacement for weather-station data for hydrological modelling

purposes in Canada. The variables retained are total precipitation (mm), minimum and maximum air temperature ($^{\circ}\text{C}$), dewpoint temperature ($^{\circ}\text{C}$), snow water equivalent (mm), surface pressure (Pa) and wind velocity in both easterly and northerly directions ($\text{m}\cdot\text{s}^{-1}$).

2.3.4 Ensemble forecast data

In order to evaluate the predictive performance of the LSTM model, 50-member ensemble forecasts were extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) integrated forecasting system (IFS) for the target basins over the period covering March 2016 to May 2023, as this was the only period for which the forecasts were available. The extracted forecasts are divided into two sets: the medium-range forecasts covering a 15-day horizon with a 6-hour time step and issued daily, and the extended-range forecasts extending from the 16th to the 46th day also with a 6-hour time step and issued twice a week (Mondays and Thursdays). The former has a resolution of about $9\text{ km} \times 9\text{ km}$, while the latter has a resolution of about $36\text{ km} \times 36\text{ km}$. For both sets, the forecasts were extracted and spatially averaged over the corresponding target basin, reducing each ensemble to a single virtual station representing the entire basin. The two sets were then concatenated along the time-step dimension, using only the common issue dates, to form the final prediction horizon of up to 46 days. The initial time step, set at 06:00 UTC, was then converted to a daily time step in Central Standard Time (CST), the closest time zone to all Canadian basins given the available data. This conversion reduced the forecast horizon to 45 complete days in local time. The variables used are identical to those in the ERA5 database used for model training.

2.3.5 Basin descriptor data

The static predictors listed in Tableau 2.2 were extracted from the HYSETS database and originate from three complementary sources. First, the geographic and topographic attributes were derived from the EarthEnv-DEM90 digital elevation model (Robinson, Regetz, & Guralnick, 2014) and subsequently processed on the PAVICS-Hydro platform (Arsenault et al., 2023b). Second, drainage area was obtained directly from the National Water Data Archive (HYDAT) maintained by the Water Survey of Canada (WSC) of Environment and Climate

Change Canada (ECCC). Finally, land-use and land-cover information was extracted by the North American Land Change Monitoring System (NALCMS; Latifovic et al., 2012).

Tableau 2.2 Basin descriptors used in this study

Descriptor	Minimum	Median	Maximum	Units
Drainage Area	46.4	1 882.6	132 585.0	km ²
Elevation	75.30	431.25	2 137.30	m
Slope	0.19	2.02	27.59	deg
Aspect	2.0	138.3	357.5	deg
Gravelius	1.33	2.16	4.31	-
Land use forest	0.00	61.54	96.56	%
Land use grass	0.08	3.54	38.15	%
Land use wetland	0.00	1.27	50.75	%
Land use water	0.06	4.73	27.54	%
Land use urban	0.00	0.50	10.35	%
Land use shrubs	0.00	6.33	47.41	%
Land use crops	0.00	0.00	90.54	%
Land use snow-ice	0.00	0.00	36.58	%

2.3.6 Streamflow and inflow volumes

The daily inflow discharge to the LSJ reservoir, provided by Rio Tinto, spans the period from January 1979 to December 2023. Streamflow data for the other 199 basins were obtained from the HYSETS database, which consolidates daily discharge measurements from the ECCC hydrometric network. For the selected basins, the lengths of the available records range from 16 to 44 years. In all cases, new inflow volume time-series were constructed by summing flows over 45-day rolling windows to serve as targets for training and testing forecasting models, as

the variable of interest is flow volume and not daily flow timeseries. The decision to train the model directly on cumulative volumes rather than daily flows is motivated by several potential benefits. First, it is hypothesized that this choice helps reduce the propagation of daily errors that would otherwise accumulate when summing daily forecasts over longer horizons. Second, it ensures that the target variable is directly aligned with operational needs for flood or reservoir volume management. It is further hypothesized that cumulative volumes are inherently easier to predict than daily flows; consequently, LSTM models may perform better when trained directly on volumes. By contrast, traditional hydrological models cannot be trained in this way and must simulate streamflow that must then be aggregated at the desired horizon.

To minimize data loss when calculating the cumulative volume, a strategy was implemented to handle missing values. If the 45-day window contains five or fewer missing streamflow values, these values are replaced by the mean of the non-missing values in the window, thus reducing information loss. Otherwise, this data point is discarded. The five-day threshold was chosen arbitrarily: a lower tolerance would have resulted in excessive data loss, while a higher threshold could have introduced too much bias into the volumes.

2.4 Methodology

2.4.1 Forecasting models and approaches

In this study, three modeling strategies were used to forecast 45-day inflow volumes. The first approach is a local LSTM (LOC), which is based on a simple LSTM model trained using the ERA5 data on the target basin. LOC is included as a baseline; however, training an LSTM on a single basin is generally ill-advised, as discussed by Kratzert et al. (2024). Once the model is trained, the ensemble weather forecasts are supplied one at a time as inputs, which generates an ensemble forecast of inflow volumes. The second approach uses a more complex LSTM model trained from the ERA5 data of all 200 basins, making it a regionally trained model (REG). The model is then used to predict inflow volumes on a target basin using the ensemble weather forecasts. The third modelling approach is only used on the main Lac-Saint-Jean (LSJ) basin. It consists of using a conceptual semi-distributed hydrological model (CEQUEAU) used operationally by Rio Tinto as a reference point to allow assessing the performance of the LSTM-based models. The ensemble forecasts of LSJ are used as inputs to the CEQUEAU model.

Since the CEQUEAU model is already calibrated by Rio Tinto for this basin, no additional tuning step is carried out; only a hydrological forecast is produced using the meteorological forecast data. This scenario thus enables a direct comparison between the deep-learning-based approaches and a proven hydrological model used in an operational context.

2.4.2 Data preparation

The LSTM model requires that the data be presented in a specific format. For the dynamic data, the model expects data windows in the form of multidimensional arrays, including: the number of training samples, the length of the time sequence for each sample, and the number of features (or variables) measured at each time step in the sequence. In this study, these dimensions correspond to: the number of forecast days (e.g., 16,000 days), the length of the time sequence (365 days), and the dynamic meteorological variables (8 variables). Using a 365-day input window is a safe option to ensure that the entire winter season is included in the input data, regardless of how far north the basin is located. Each sequence consists of 320 days of observed meteorological data (from ERA5) prior to the target forecast date, followed by 45 days of observed meteorological data starting from the forecast date (Figure 2.4). These sequences are used to train the model sequentially, where the target variable is the cumulative volume over the same 45 forecast days, obtained by aggregating the daily streamflows. Meanwhile, the static data are organized as two-dimensional arrays, representing the number of forecast days and the types of static variables.

In addition, further processing was applied to the 45-day cumulative volume values. The training volumes were divided by the area of their respective basin for the REG approach, which involves using data from 200 basins. This normalization yields a volume per unit area, making the forecasts more comparable between basins of different sizes. Although each basin's area is already provided as input, it guides the model explicitly: without this normalization, it would be very difficult for the network to deduce on its own that flow volumes should be scaled by drainage area in order to make physical sense.

a) Training example Target date : 1990-11-17		1990-01-01	1990-01-02	1990-01-03	...	1990-11-14	1990-11-15	1990-11-16	1990-11-17	1990-11-18	1990-11-19	...	1990-12-31
Training dataset	LOC scenario												
	Input: ERA5 sequences for the LSJ basin	320d (ERA5 - LSJ)							45d (ERA5 - LSJ)				
	REG scenario												
	Input: ERA5 sequences from 200 basins	320d (ERA5 - 200 basins)							45d (ERA5 - 200 basins)				
	All scenarios												
	Target (Output): Observed inflow volume								45d LSJ				

b) Forecasting example Target date : 2023-11-17		2023-01-01	2023-01-02	2023-01-03	...	2023-11-14	2023-11-15	2023-11-16	2023-11-17	2023-11-18	2023-11-19	...	2023-12-31
Forecasting dataset	All scenarios												
	Input: 50-member forecasts – LSJ basin	320d (ERA5 - LSJ)							45d (forecasts - LSJ)				
	Target (Output): Observed inflow volume								45d LSJ				

Note : LSJ stands for Lac-Saint-Jean.

	Observed meteorological data from ERA5 (8 meteorological variables)
	Forecast meteorological data from the 50-member ensemble forecasts (8 meteorological variables)
	Target 45-day cumulative volume data (from daily data, Rio Tinto)

Figure 2.4 Example of the implementation of a training sample for a 45-day cumulative volume forecast model, trained with observed data (a), as well as an example of how such volume forecast models are applied over an independent period (b). Panel (a) shows a single sample, but the process is repeated for all available data used to train the models, as well as for the validation and testing periods. Similarly, panel (b) illustrates an example of a single 45-day cumulative volume forecast, but the process is repeated for all available forecast dates

Finally, the dynamic, static, and inflow volume data were normalized to ensure better convergence of the LSTM model and improve prediction accuracy. In this study, the MinMaxScaler was configured with the training period data only, to rescale each variable to a common range between 0 and 1, thus preventing the model from giving excessive importance to certain variables simply due to their larger numerical values, which ensures fair treatment of each variable. The scaler was then applied on all data, including for the training, validation and testing subsets. This eliminates the possibility of data leaking and ensures robust results.

Regarding the data split, the dataset was divided into three distinct sets: training, validation, and testing. The training period allows the model to learn the relationships between the input data and the predictions by adjusting the weights and internal parameters. The validation period enables the model to stop training when the validation scores no longer improve, serving as a stopping criterion to avoid overfitting. Finally, the testing period serves to assess the robustness of the model, by testing on data that is completely independent from the training and validation steps. The splits are varied depending on the implemented scenario, but all splits are done in continuous chronological blocks to avoid having testing data on a day following a training or validation data point. Some differences arise depending on the scenario:

- 1) In Scenario 1 (LOC), only ERA5 data from the period 1979 to 2015 are used to train the model, as the 2015–2023 period is reserved for simulating ensemble forecasts. Thus, 60% of the data from 1979 to 2015 are allocated to training, 20% to validation, and the remaining 20% to model evaluation during the testing phase.
- 2) In Scenario 2 (REG), only ERA5 data for the target basin for the period 1979–2015 are used for training, since the subsequent period is reserved for running the model in ensemble forecast mode, as in Scenario 1 (LOC). However, all data from the 199 additional basins are used for training/validation. Thus, 60% of the ERA5 data for the target basin and 80% of the data from the 199 complementary basins are used for training. The validation data consist of the remaining 20% of the ERA5 data for the target basin and 20% of the remaining data for each additional basin. Finally, the testing period consists of the final 20% of the ERA5 data for the target basin.

The main target basin used in this study is LSJ, but the same methodology is independently applied to the seven other basins mentioned in Section 2.2. In each case, the selected test basin is treated as the target, with its data split into training, validation, and testing subsets following the same protocol. Although these basins are part of the 200 selected basins, only a portion of their historical data is used for model training in Scenario 2, to assess how well the methodology generalizes to varied hydrological contexts. When a basin other than LSJ is targeted, the data from LSJ and the six remaining test basins are fully used for training and

validation, just like the other basins in the dataset. While these test basins are not strictly out-of-sample, this setup offers a controlled framework to evaluate the robustness of the approach across diverse conditions. By rotating the target basin and ensuring geographic and hydroclimatic diversity, the experiments provide valuable insight into the method's generalization capacity, though broader applications to entirely unseen basins remain for future work.

2.4.3 LSTM model and structures

As mentioned previously, two LSTM models are used in this study: a local model and a regional model. The regional model adopts an architecture inspired by Arsenault et al. (2023a) and was refined through an iterative trial-and-error approach, whereas the local model uses a simplified configuration. These two variants each correspond to a specific application context and are presented in Figure 2.5. The LOC model is detailed in Figure 2.5a and the REG model is shown in Figure 2.5b. The LOC model is not suitable for regional modeling because it requires not only dynamic data but also the static characteristics of each basin, which are essential for understanding the relationships between the physical characteristics of the basins and the properties of their hydrological regimes. Hyperparameters such as the number of layers and LSTM units, as well as the activation functions, were optimized using a trial-and-error approach to ensure the best possible convergence when training the models on the target LSJ basin. For the REG model, a dropout rate of 0.3 was applied to the LSTM layers and 0.1 to the static layer. A LeakyReLU activation function is used before the final Dense layer for the LOC model (a) and before the penultimate Dense layer for the REG model (b). The LSTM model structures and the selected hyperparameters were then applied to the seven other complementary basins to test the application of the LOC and REG models trained on the LSJ basin to other basins.

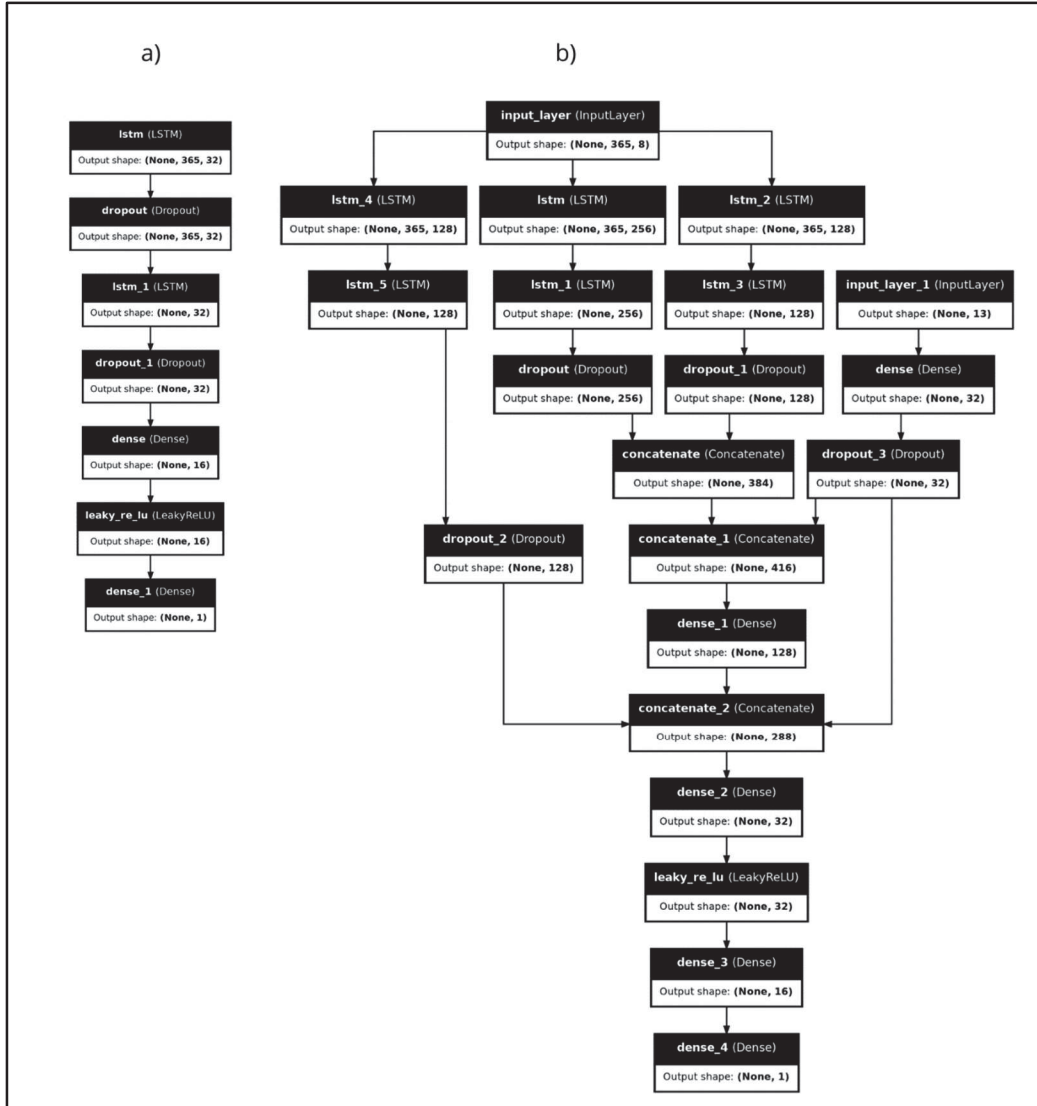


Figure 2.5 Structures of the LOC LSTM model (a) and REG model (b)

2.4.4 Training

All three LSTMs were trained with a batch size of 512 sequences, which corresponds to the maximum allowed by the available GPU memory, using the Adam optimiser and the mean-squared error (MSE) loss. MSE quantifies the average squared difference between the predicted values and the corresponding observations.

It is defined as:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.1)$$

where N is the total number of observations, $\hat{y}_i$ is the predicted value, and y_i is the corresponding observed value.

Initial learning rates were scenario-specific: 1×10^{-4} for the REG model and 1×10^{-3} for the LOC model. The learning rate was adjusted dynamically during training using the “ReduceLROnPlateau” algorithm with a reduction factor of 0.1 for REG, while LOC followed a fixed step-decay schedule. Early stopping was applied with a patience of 15 epochs, based solely on the validation MSE.

2.4.5 The CEQUEAU hydrological model

The semi-distributed CEQUEAU hydrological model (Morin & Paquet, 2007) used in this study follows the configuration of Sabzipour et al. (2023) and serves as the reference against which LSTM approaches are evaluated on the LSJ catchment (45,000 km²). It divides the catchment into 28 hydrological units and incorporates continuous data assimilation of streamflow. CEQUEAU uses daily total precipitation and mean air temperature as inputs. It generates daily streamflow forecasts that were summed over the 45-day lead-time to obtain simulated inflow volume forecasts.

2.4.6 Performance metrics

The modified Kling-Gupta Efficiency score (KGE) is used in this study to assess the testing performance of the LSTM models. Introduced by Gupta et al. (2009) and modified by Kling, Fuchs, and Paulin (2012), it is defined as:

$$1 - KGE = \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (2.2)$$

where β is the bias ratio, defined as:

$$\beta = \frac{\mu_{sim}}{\mu_{obs}} \quad (2.3)$$

and γ is the variability ratio, defined as:

$$\gamma = \frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}} \quad (2.4)$$

KGE values range from $-\infty$ to 1. A score greater than -0.41 indicates that the model improves on the long-term mean-flow benchmark (Knoben, Freer, & Woods, 2019) and a value of 1 indicates perfect agreement between simulated and observed data.

The performance of the forecasting models is evaluated using two complementary metrics, each capturing a different aspect of forecast quality. The Continuous Ranked Probability Score (CRPS; Hersbach, 2000) is a widely used metric for assessing the skill of probabilistic forecasts. It simultaneously accounts for accuracy, defined as the closeness of the forecast distribution to the observed value, and sharpness, which refers to the concentration (or spread) of the forecast distribution, independently of the observations. A lower CRPS value indicates a forecast that is both close to the observed outcome and confidently narrow in its spread. Formally, the CRPS measures the distance between the cumulative distribution function (CDF) of the forecast and the CDF of the observed value. It is defined as:

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{+\infty} \left[\left(F(V_{sim_i}) - F(V_{obs}) \right) \right]^2 dV \quad (2.5)$$

where N is the number of ensemble members, $F(V_{sim,i})$ is the forecast CDF, and $F(V_{obs})$ is the step function corresponding to the observed value. The CRPS is expressed in the same units as the forecast variable, which is cubic meters (m^3) for inflow volumes. Its values range from 0 to $+\infty$, with 0 indicating a perfect forecast. Compared to deterministic metrics such as

KGE, the CRPS provides a more informative assessment in a probabilistic context, as it simultaneously penalizes both bias and variance.

In addition to CRPS, Talagrand diagrams (Talagrand & Vautard, 1997) are used to evaluate forecast reliability, which reflects the statistical consistency between the predicted ensemble distribution and the observed outcomes. In this study, observation percentiles within the forecast ensemble are binned in 10-percentile intervals, leading to 10-bin histograms. A perfectly flat histogram denotes ideal reliability, indicating that the observation is statistically indistinguishable from any ensemble member. Conversely, a U-shaped curve signals under-dispersion, because the observation falls outside the ensemble range too frequently. Finally, an L-shaped histogram reveals a systematic bias, with the observation consistently higher or lower than the forecasts. To further quantify the uniformity of the rank histogram, the Absolute Bin Distance from Uniformity (ABDU) score is computed (Arsenault, Latraverse, & Duchesne, 2016):

$$ABDU = \frac{1}{N} \sum_{k=1}^N \left| S_k - \frac{M}{N} \right| \quad (2.6)$$

where S_k is the number of occurrences in bin k , M is the total number of samples, and N is the number of histogram bins. Lower ABDU values indicate greater consistency between the ensemble and the observations, i.e., better forecast reliability.

2.5 Results

2.5.1 Model training and verification

A summary of the training performance of the two LSTM models (LOC and REG) is presented in Tableau 2.3. Among the two scenarios, the REG model stands out for its robustness and stability, particularly between the training and validation phases. It shows KGE values close to 0.97, indicating very good reproduction of the structure of the hydrological signal, including temporal correlation, bias (mean ratio), and relative variability (coefficient of variation). However, these high KGE values, while impressive, should not be interpreted in the same way as KGE scores for daily streamflow. It is also important to note that training and validation performance are only used as diagnostic variables given the strong likelihood of overfitting, thus making their analysis in absolute terms much less useful. The performance on the testing period, on the other hand, is a direct and independent evaluation of the model performance.

Tableau 2.3 KGE scores for the training of the LOC and REG scenarios

N ⁰	Basin	LOC			REG		
		Training	Validation	Testing	Training	Validation	Testing
1	Lac-Saint-Jean	0.976	0.917	0.945	0.974	0.968	0.926
2	Grande River	0.914	0.915	0.880	0.969	0.967	0.953
3	Cascapedia River	0.943	0.921	0.914	0.968	0.974	0.937
4	Bear Brook	0.875	0.806	0.840	0.972	0.972	0.871
5	Hayes River	0.702	0.738	0.667	0.966	0.967	0.600
6	Swift Current Creek	0.801	0.655	0.519	0.980	0.971	0.611
7	Crowsnest River	0.902	0.905	0.912	0.970	0.969	0.933
8	Tuya River	0.861	0.816	0.706	0.968	0.972	0.791

In the test period, the performance of the REG scenario varies considerably from one basin to another, with KGE values ranging from 0.600 to 0.953. The best results are obtained for basins

2, 3, 7 and 8. The LOC scenario stands out mainly for basin 1 (LSJ), where it outperforms the other configuration, with a testing KGE of 0.945 compared to 0.926 for REG. In general, LOC offers good performance for five basins (1 (LSJ), 2, 3, 4, and 7), with KGE values ranging from 0.840 to 0.945 during the test period. The other three basins show more mixed results, with test KGE values between 0.519 and 0.706. Apart from being situated in the Prairies and western Canada, the three basins with lower performance do not share a clear common physiographic context or hydrological regime, as two are prairie basins while one is mountainous. Basin 7 is also located in western Canada and in a mountainous setting, yet it constitutes an exception, as it still demonstrates good performance despite the challenges faced in such environments. For detailed values of the correlation coefficient, bias, and variability, refer to Tableau-A I-1 to A I-3 in the Appendix II.

2.5.2 Overall comparison of model performance in forecasting on the Lac-Saint-Jean basin

This section compares the performance of the LOC and REG scenarios against that of the reference hydrological model CEQUEAU (CEQ) over the full forecast period available. Overall, the two LSTM models achieve similar mean performance in terms of KGE (Figure 2.6). Across all ensemble members, the mean KGE computed by taking the ensemble mean reaches 0.908 for the regional model (REG) and 0.904 for the local model (LOC), both exceeding the CEQ score of 0.849 for the same period. However, LOC systematically underestimates winter inflows. REG tracks the observations more closely but still underestimates winter volumes. Consequently, during winter, the min–max range of both LOC and REG does not always encompass the observations, as in the winter of 2017. By contrast, CEQ produces a wider forecast band that consistently covers the observations.

For LOC, the daily forecast spread at the spring peak is generally narrower than for the other two models. This tighter band fails to capture the observation and lies below it in certain years, such as 2017 and 2019, indicating that the model struggles to anticipate some spring high flows.

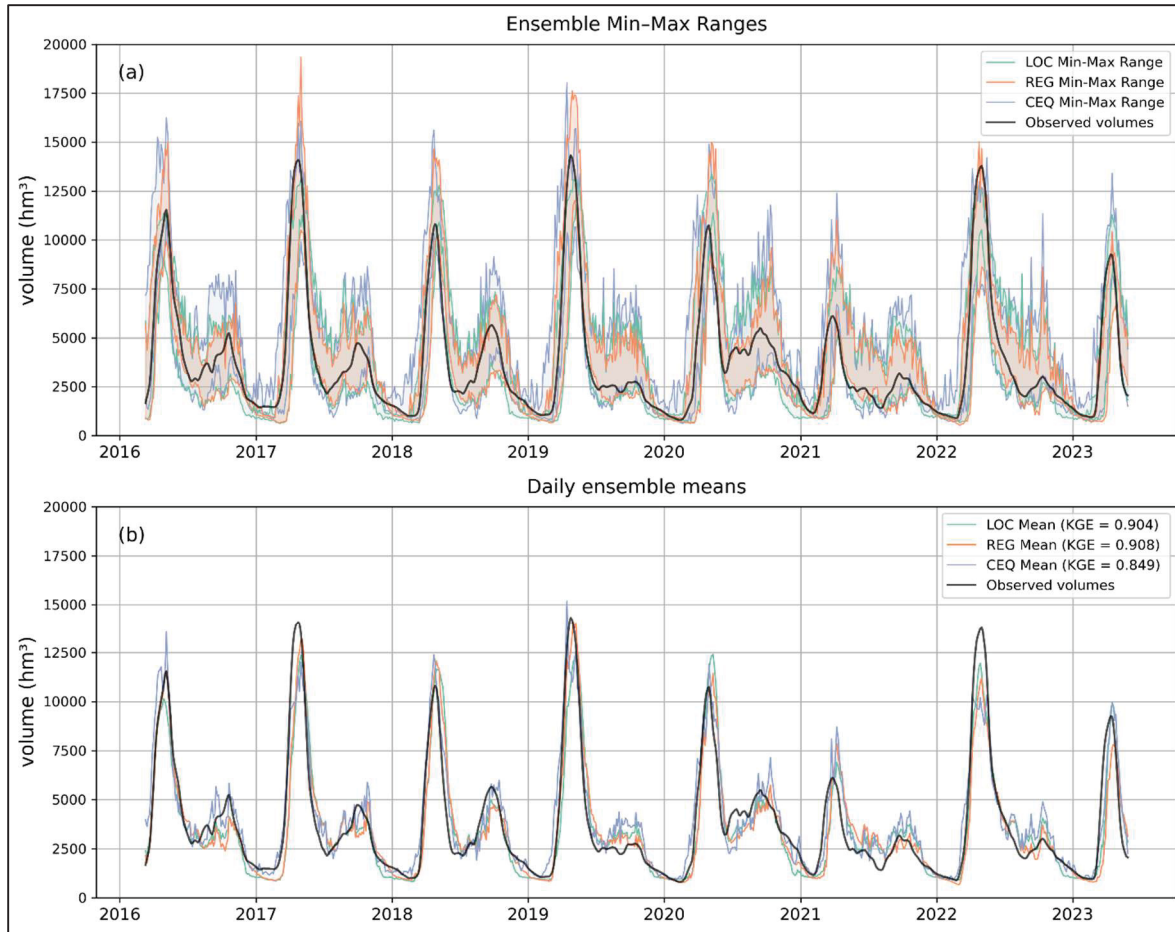


Figure 2.6 Ensemble forecasts of 45-day inflow volumes for the LSJ basin: (a) Min-max envelopes, (b) Daily mean forecasts for three scenarios: LOC, REG and CEQ. The forecast period covers March 10, 2016 to May 29, 2023

2.5.3 Accuracy of annual and seasonal forecasting models across the test basins

The aim of this section is to compare how the LSTM models simulate 45-day cumulative inflow volumes over the forecast period for the LSJ basin and seven additional basins, at both seasonal and annual scales. As described in Section 3.6, forecast accuracy is assessed using the CRPS metric. The CRPS values obtained for the 45-day cumulative-volume ensemble forecasts for the 2016–2023 period are shown in Figure 2.7. Note that the CRPS values are expressed in cubic hectometers (hm³).

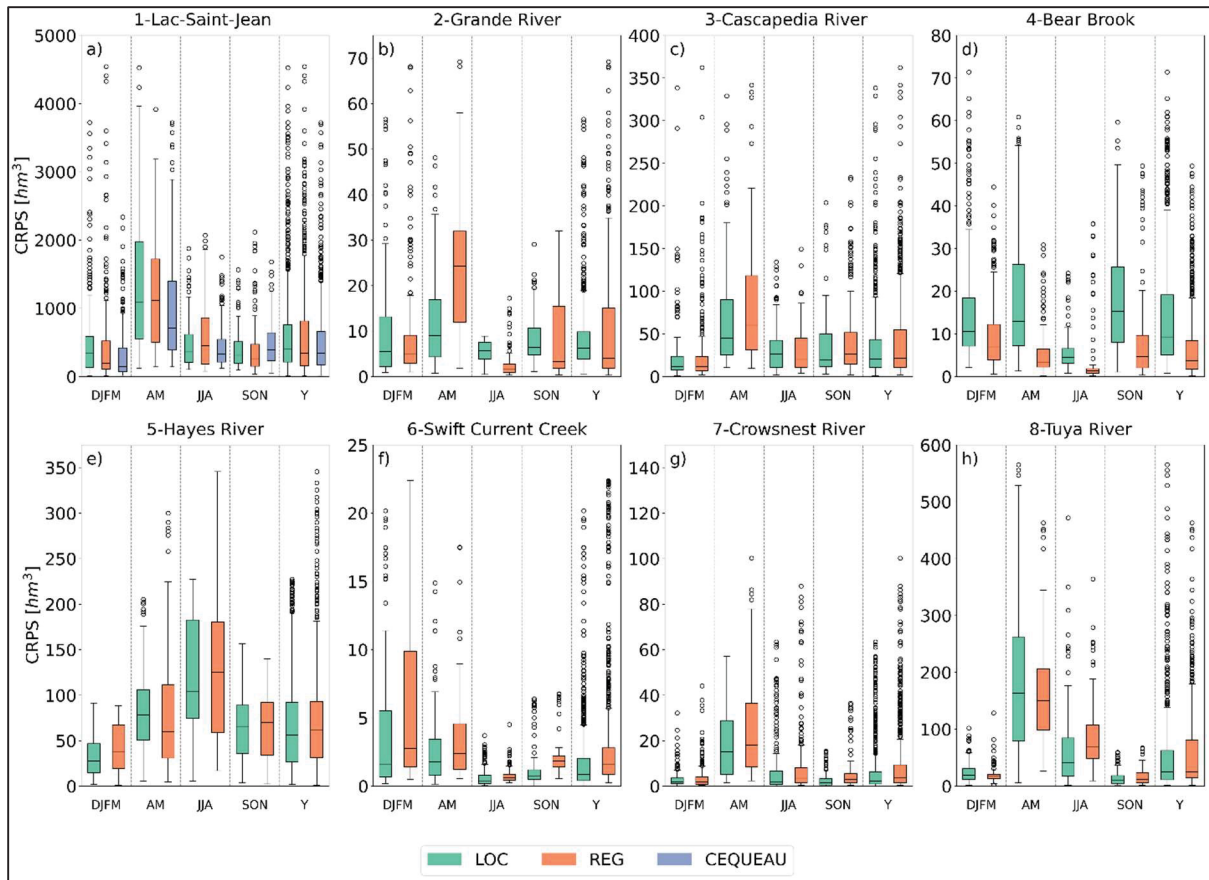


Figure 2.7 CRPS of the volume ensemble forecast by season for the LOC and REG models over the entire period of the forecast. Each panel represents a basin: a) Lac-Saint-Jean [1], b) Grande River [2], c) Cascapedia River [3], d) Bear Brook [4], e) Hayes River [5], f) Swift Current Creek [6], g) Crowsnest River [7], and h) Tuyá River [8]. The CEQUEAU model is also shown in panel a) for Lac-Saint-Jean. Seasons are defined as winter (DJFM), spring (AM), summer (JJA), fall (SON), and annual (Y). The forecast period covers March 10, 2016 to May 29, 2023

The CRPS values vary markedly from one season to another: they are often higher in spring, except for basin 5 (Figure 2.7e) and basin 6 (Figure 2.7f). These larger values chiefly reflect the greater magnitude of flows during the peak-flood season and do not necessarily indicate a loss of model skill. To compare performance on a common basis, the seasonal CRPS values can also be expressed as a percentage of the mean volume for the corresponding season, whereas the annual CRPS values are expressed as a percentage of the mean annual volume, as shown in Figure-A I-1 in the Appendix III.

For basin 1 (LSJ), the reference model CEQUEAU is more accurate than the LOC and REG models (Figure 2.7a). Its interquartile CRPS lies in the lowest value band for most seasons (winter, spring, and summer) whereas LOC and REG occupy higher bands. In autumn, however, REG records the lowest interquartile CRPS. When looking at the annual CRPS distributions, CEQUEAU remains the most accurate overall, followed by REG and then LOC.

When the LSTM model performances for basin 1 are compared with those for the seven additional basins, no clear seasonal pattern emerges: the ranking of the models by performance varies widely from one basin to another. For instance, at basins 4 and 7 (Figures 2.7d and 2.7g), REG is consistently more accurate than LOC, whereas the opposite holds for basin 3, although the gap is less pronounced. Other basins, such as basin 8 (Figure 2.7h), present a mixed picture: REG leads in winter (narrower interquartile range) and spring (lower median error), while in summer and autumn LOC remains the best performer.

2.5.4 Reliability of the LSTM-based inflow volume forecasting models

This section first presents an analysis of the reliability of the three LSTM models, evaluated both globally (on an annual scale) and by season for the LSJ basin, in comparison with the operational reference hydrological model CEQUEAU (Figure 2.8). As previously described in Section 3.6, reliability is assessed using Talagrand diagrams and the associated ABDU values, which quantify the uniformity of the rank histogram. Secondly, the results obtained for LSJ are compared with those of the seven other complementary basins.

Regarding the LSJ basin, the semi-distributed CEQUEAU model shows relatively flat Talagrand diagrams, indicating better overall reliability compared to the two LSTM model scenarios, especially on an annual scale as well as during winter, summer, and autumn seasons. However, in autumn, the LOC (ABDU = 6) and REG (ABDU = 9) models, although they display moderate underdispersion (U-shaped form), show better performance than CEQUEAU (ABDU = 10), whose histogram has a more L-shaped form, indicating a pronounced underestimation of forecasts.

The two LSTM scenarios show notable underdispersion for several seasons: in winter, the histograms take on an inverted L shape, indicating frequent underestimation, while in spring and summer, the U shape indicates that the spread is too narrow. These biases are reflected in the annual diagrams, which generally display a U shape, illustrating a lack of reliability in the forecast envelope.

Figure 2.9 shows the annual Talagrand diagrams for the two LSTM model scenarios applied to the seven complementary basins where the study's methodology was also tested. A pattern similar to that of the LSJ basin emerges: each of the LSTM scenarios shows, for all basins, signs of underdispersion (U shape) or systematic biases, whether positive (inverted L shape) or negative (L shape). These distortions in the rank distributions indicate that the forecast ensembles struggle to reliably represent the observed variability, thereby compromising the probabilistic quality of the predictions.

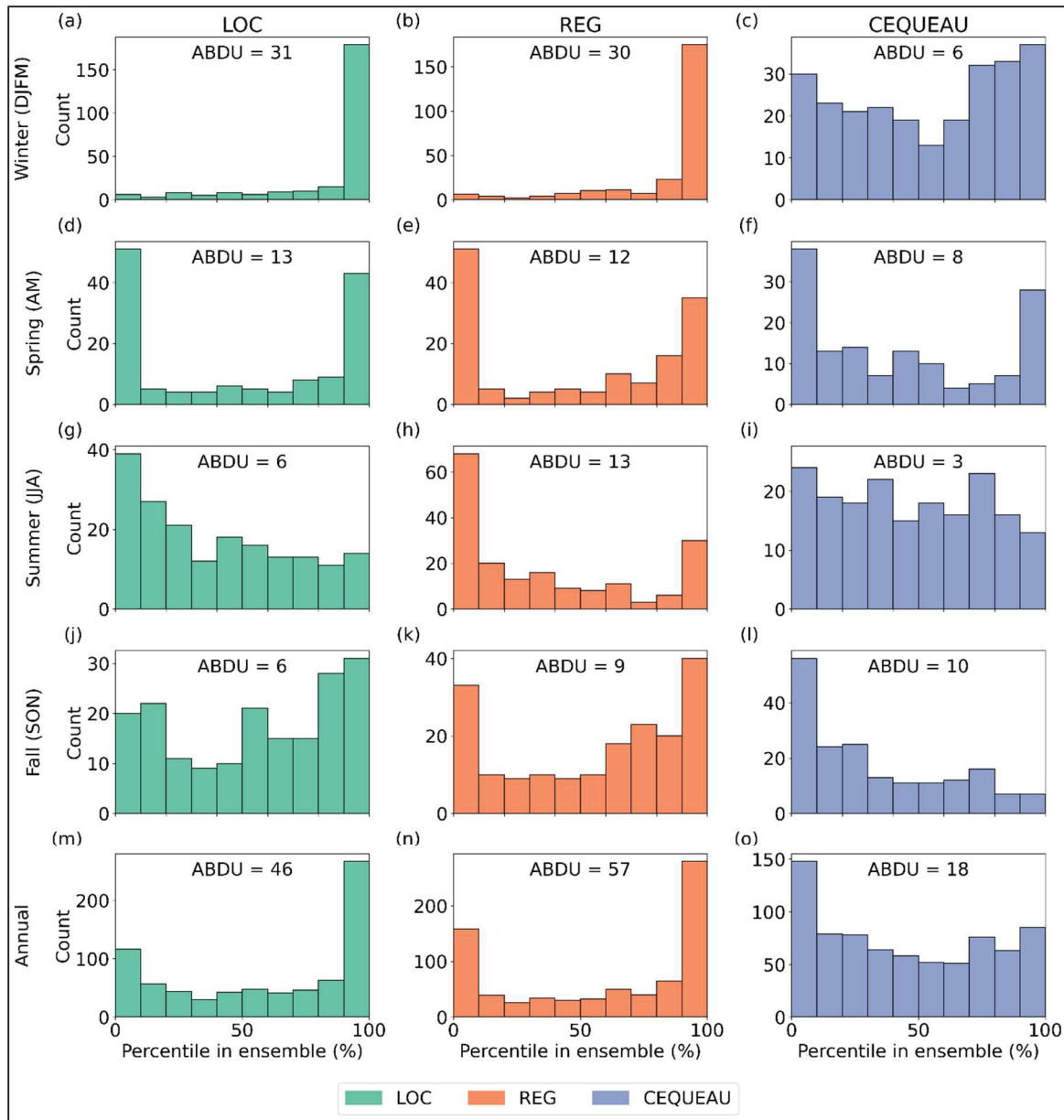


Figure 2.8 Forecast reliability for the Lac-Saint-Jean basin under three modeling scenarios.

The figure presents Talagrand diagrams illustrating the forecast rank distributions (in percentiles) of 45-day cumulative water volumes for each scenario. The scenarios include:

LOC (local LSTM model), REG (regional LSTM model trained on 200 basins), and CEQUEAU (hydrological reference model). Forecast reliability is assessed over the entire available prediction period (2016–2023)

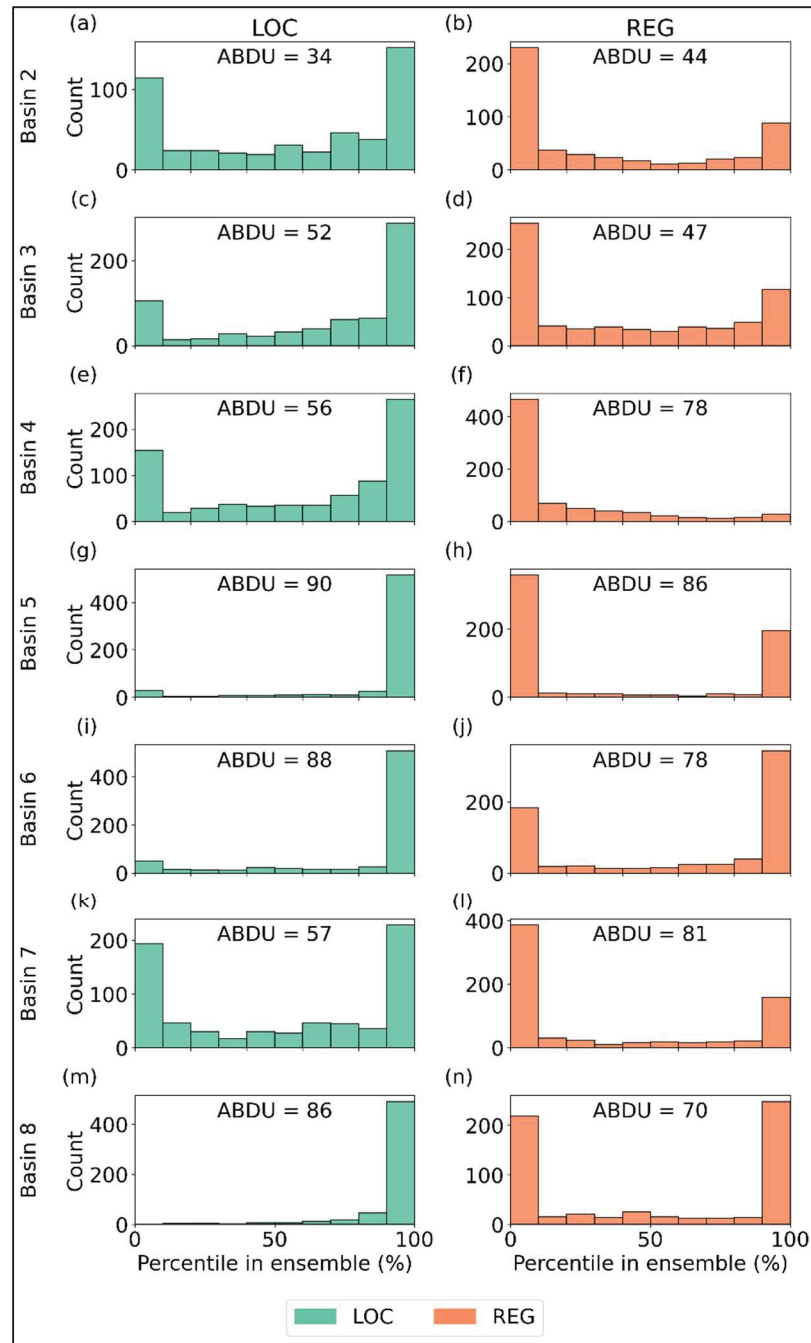


Figure 2.9 Annual forecast reliability for three LSTM-based modeling scenarios: LOC (local model) and REG (regional model trained from 200 basins). The evaluation period spans the full range of available forecasts (2016–2023). Talagrand diagrams represent the forecast rank distribution (in percentiles) of 45-day cumulative water volumes for each scenario, across seven Canadian basins, numbered 2 to 8: Grande River [basin 2, panels a-b], Cascapedia River [3, c-d], Bear Brook [4, e-f], Hayes River [5, g-h], Swift Current Creek [6, i-j], Crowsnest River [7, k-l], and Tuya River [8, m-n]

2.6 Discussion

2.6.1 The performance of LSTM in testing

Clear differences in calibration performance are observed between the local (LOC) and regional (REG) models. This reflects the combined influence of basin diversity, training data type, and hyperparameter selection focused on basin 1 (LSJ).

In the LOC scenario, the basin with the best performance is basin 1 (LSJ) which is not surprising given that the model structure was chosen to maximize its performance, which could partly explain this result. Conversely, basins 5 (Manitoba), 6 (Saskatchewan), and 8 (British Columbia) show more limited performance (KGE in testing phase between 0.519 and 0.706), and their weaknesses can be traced to distinct component behaviors. Basin 5 in LOC suffers from low correlation in testing ($r = 0.675$), whereas SCC displays limited correlation ($r = 0.682$), a poor bias representation ($\beta = 0.643$), but good variability ($\gamma = 1.056$) in testing, pointing to challenges in representing the hydrological signal. Basin 8, by contrast, displays acceptable correlation ($r = 0.879$) and good bias representation ($\beta = 0.985$), but with poor variability ($\gamma = 0.732$), indicating over-smoothing that undermines amplitude reproduction despite reasonable mean and temporal alignment.

The REG scenario yields similar overall performance to LOC but slightly surpasses it for six of the eight basins in test-period KGE (Tableau 2.3), implying that pooling a larger and more diverse set of basins generally improves calibration robustness. Nonetheless, generalization in testing is heterogeneous depending on the metric: basin 5 exhibits a slightly worse correlation than with the LOC model ($r = 0.634$ vs $r = 0.675$, respectively), but all others remain the same or see improvements, such as SCC whose correlation jumps from $r = 0.682$ for the LOC model to 0.813 for the REG model. However, in terms of bias, the REG model seems to either remain inline with the LOC model or reduces the performance slightly. It is notable that the bias ratios for the REG model are all smaller than 1, whereas the LOC model does see some ratios above 1. This seems to show that the REG model, in general, slightly underestimates the mean flows

overall compared to the LOC model. Variability under REG again shows similar results to LOC, with, basin 5 and 8 tending to underestimate variability while SCC overestimates it ($\gamma = 1.167$), highlighting inconsistencies in how variance is preserved across basins. Interestingly, the best-performing basin in REG is basin 3 ($KGE = 0.953$), slightly outperforming LSJ (0.926), which suggests positive transfer when a basin's local hydrological behavior aligns well with patterns represented in the regional training pool. One possible explanation for the persistently lower performance of basin 5, 6, and 8 across both scenarios relates to their geographic and hydrological distinctiveness, which may limit both local and regional models' ability to generalize. These results should be interpreted with caution, since model performance in simulation mode does not necessarily reflect forecast quality (Davidson-Chaput et al., 2025).

2.6.2 Influence of physiographic and climatic factors

The characteristics of a basin and its hydrological behaviour vary considerably from one region of Canada to another, which partly explains the differences in calibration performance observed among the eight testing basins. As noted earlier, the LOC scenario shows stable and consistent performance in several basins in Quebec (basins 1 (LSJ) and 3), Ontario (basin 4), New Brunswick (basin 5), and southern Alberta (basin 7). These basins generally share moderate topographic complexity, which favours a more stable relationship between input data and simulated flows. Basin 7 is a notable exception: despite being grouped with these basins, its location in the Canadian Rockies introduces additional complexity. With a relatively high elevation (1703 m), steep average slopes (14.9°), and a nival flow regime (Tableau 2.1), the basin is strongly influenced by snowpack melt, subject to marked altitudinal precipitation gradients, and highly sensitive to interannual climate variability. The basin's consistently good calibration performance is thus somewhat surprising, given the well documented challenges of calibrating hydrological models in high altitude, snow dominated basins (Brandt et al., 2022; Evin et al., 2024).

In contrast, basin 6 in Saskatchewan is dominated by cropland (70.86 %) and grassland (22.81 %), and its extremely flat topography (mean slope 1.47° ; Tableau 2.1) greatly extends

the time of concentration and produces diffuse flow, spreading out runoff events and dampening peak flows. Across the North American Prairies, this near zero relief is punctuated by numerous depressions (“potholes”) whose wet or dry states and interconnectivity vary, leading to high seasonal and interannual variability in water inputs from prolonged droughts to extreme flood events. Such nonlinear hydrological responses are especially challenging to model over nearly flat basins. Ahmed et al. (2020) refer to the Canadian Prairies as a “graveyard of hydrological models” due to the lack of relief and the complexity of seasonal processes. Moreover, Mekonnen et al. (2015) demonstrate that a hybrid approach, combining a process-based model with a data driven method, significantly improves flood simulations on these plateaus compared to using either approach alone.

Basin 5, located in northern Manitoba within the Canadian Shield, clearly illustrates another type of modelling challenge. Although it lies outside the geographical boundaries of the Prairie Pothole Region, it shares some comparable hydrological mechanisms, notably an extremely flat topography (slope of 0.66° , see Tableau 2.1) that promotes the accumulation of water in wetlands and peatlands typical of the boreal forest. Land cover further reinforces these conditions, with nearly half of the basin covered by forest (48.54 %), and roughly equal shares of wetlands (22.14 %) and surface water (19.87 %) (see Tableau 2.1). Together, these extensive wetlands and water bodies act as temporary storage zones that slow and modulate runoff, resulting in a slow and weakly responsive hydrological regime. In addition, the region’s subarctic climate, with its long cold winters, leads to substantial spring snowmelt, whose amount and timing depend mainly on soil frost depth, ice-rich layers, and the development of the active layer (i.e., the seasonally thawed surface soil above permafrost) (Metcalf & Buttle, 2001). The combination of these factors makes it particularly uncertain to predict the precise moment when the stored water will be released: a phase error in the timing of snowmelt can lead to a significant shift in the cumulative runoff volume and flow timing. Interestingly, aside from basin 1 (LSJ), basin 5 is the only other basin that does not appear to benefit from the regional model training: its test-period KGE is indeed lower for the REG scenario (0.600) than for LOC (0.667). This may suggest that the regional sample includes too few basins with

similarly flat, wetland-dominated boreal hydrological behaviour, which limits the model's ability to generalize well to the basin 5 context.

Finally, basin 8 in British Columbia illustrates another type of calibration challenge, as it is located within the Boreal Cordillera ecozone. It has a lower elevation (1193 m), gentler mean slope (4.98°), and a substantially higher annual mean discharge ($36.9 \text{ m}^3/\text{s}$) than basin 7 (see Tableau 2.1). However, its drainage area is nearly nine times larger than that of basin 7, and its Gravelius index (2.31 vs. 1.48 for basin 7) indicates a much less compact basin shape. This greater size and morphological complexity amplify hydrological variability and likely make model calibration more demanding. In addition, its Boreal Cordillera context, combining mountainous terrain with a nival flow regime, produces a strongly seasonal hydrograph dominated by a pronounced summer snowmelt peak. These characteristics heighten the sensitivity of streamflow to interannual climate variability, making calibration particularly challenging (Brandt et al., 2022).

These regional particularities can complicate the training of a generic model and underline the need for finer calibration, adapted to the specific characteristics of each basin. A test was thus carried out to retrain the regional LSTM model using local forecast data from basin 1 (LSJ) (fine-tuning), but the results proved unsatisfactory. The input data format remained consistent with that used during the initial training, except that each sequence now consisted of 320 days of historical meteorological data followed by 45 days of ensemble-based meteorological forecasts. This resulted in 50 parallel input sequences (one per ensemble member), each composed of a 365-day window leading to the same 45-day cumulative volume target, thereby simulating realistic forecasting conditions. Despite this, difficulties with convergence were observed, as well as overfitting during training and a simulated value range that was too narrow. One hypothesis is that the complexity of the regional model was too high in relation to the limited size of the new dataset. Additionally, providing 50 distinct input sequences that all lead to the same target value may have artificially reduced the spread of the predicted outputs, thereby limiting the model's ability to express uncertainty. These observations also suggest that part of the limitations observed could be directly related to the availability and

nature of the forecast data, which only cover the 2016-2023 period and as such do not represent a full sampling of the climate of each region. It should also be noted that some of these basins are not necessarily representative of those typically associated with hydropower production. As such, the primary objective of this study may not directly apply to all, and future research should investigate the performance in a variety of hydroclimatological conditions.

2.6.3 Seasonal performance differences for the Lac-Saint-Jean basin

CEQUEAU is only used as a reference for the LSJ basin, which makes seasonal comparisons relevant exclusively for this basin. Therefore, this section focuses on how the LSTM models perform seasonally compared to CEQUEAU for the LSJ basin. Seasonal performance was evaluated using normalized CRPS and Talagrand diagrams to capture both forecast accuracy and reliability for the 45-day inflow volumes. Overall, the results show that CEQUEAU generally provides better seasonal reliability for LSJ, especially during winter and spring when snowmelt processes dominate, but also during summer when inflows are more driven by rainfall events. It is also well known that LSTM-based models are generally overconfident, which leads to lower reliability (Liu et al., 2023; Wang, Si, & Chu, 2022).

However, the comparison also highlights that the LSTM scenarios can occasionally outperform CEQUEAU under specific seasonal conditions. For example, in autumn, they both display flatter Talagrand diagrams and lower CRPS values than CEQUEAU, suggesting better reliability and less bias during this period. This indicates that deep learning models, when properly trained with high-quality historical data, can capture certain hydrological patterns that a process-based model may not fully resolve, especially for transitional seasons with complex dynamics.

More specifically, Talagrand diagrams reveal distinct seasonal bias patterns for LOC and REG: in autumn and winter, both scenarios show a slight negative bias (underprediction). In spring, forecasts are mostly underdispersed, while in summer, they tend to display a positive bias

(overprediction). This illustrates how forecast reliability varies by season and supports the idea that targeted post-processing could help correct systematic errors.

LOC outperforms REG in terms of reliability during summer and autumn, as indicated by lower ABDU values. For example, REG shows essentially similar performance in winter (ABDU 30 vs. 31) and spring (ABDU 12 vs. 13), but is less reliable in summer (ABDU 13 vs. 6) and autumn (ABDU 9 vs. 6). In terms of accuracy, REG is slightly more precise overall, except in summer, where LOC achieves lower normalized CRPS values. These results illustrate that a regional approach can help capture seasonal dynamics more effectively, while local training (LOC) can still be advantageous under certain seasonal conditions despite the broader regional training of REG.

Nonetheless, these improvements remain limited and highly dependent on the season. On an annual basis, CEQUEAU still outperforms the LSTM scenarios in terms of overall reliability. Therefore, while the LSTM models show promise in terms of their precision and accuracy, they are not yet reliable enough to fully replace an operational model like CEQUEAU for long-term inflow forecasting for hydropower reservoir management, at least for the LSJ basin. For the LSTM models, seasons with better precision (lower CRPS) also tend to exhibit better reliability, as reflected by flatter Talagrand rank histograms, particularly in autumn and summer. The reasons for CEQUEAU's higher reliability remain unclear; further investigation is warranted, especially since direct volume forecasting is still an emerging field.

These results suggest that deep-learning models, which have been shown to address key limitations of conventional hydrological modeling such as scaling, equifinality, and parameter regionalization (Shen 2018), could be beneficial when used alongside traditional models, for example in a multi-model or hybrid framework (e.g., Armstrong et al., 2025; Cho & Kim, 2022; Konapala et al., 2020; Ougahi & Rowan, 2025; Solanki, Vegad, Kushwaha, & Mishra, 2025). Such an approach would combine the physical realism and proven operational performance of CEQUEAU with the data-driven flexibility of LSTM networks, potentially

offering better forecasts for decision-making during certain seasons. Such an approach is discussed in more detail in section 2.6.5.

2.6.4 Seasonal variability in forecast performance across the other basins

While the results for the LSJ basin highlight both the strengths and limitations of the LSTM scenarios compared to a conceptual model, the strong variability already observed between basins suggests that seasonal performance patterns may differ substantially from one basin to another. Therefore, this section focuses on examining these seasonal differences across the seven other test basins, comparing the LSTM scenarios in the broader study area where CEQUEAU is not available.

Over the full forecast period, the regional model often performs better during winter and spring for several basins. This may be explained by the greater diversity of snow- and rain-snow-dominated conditions within the regional dataset, which likely helps the model generalize more effectively in these seasons. In summer, however, the local model sometimes achieves lower CRPS values, especially in basins where floods are primarily driven by convective precipitation and sudden runoff events. In autumn, the differences are less clear-cut, with some basins favoring LOC and others REG, without a consistent pattern.

In summary, these findings make clear that no single scenario is optimal under all conditions. The results reveal strong intra-basin variability (between seasons within the same basin) as well as inter-basin variability (from one basin to another). This emphasizes that forecast performance depends heavily on local hydrological regimes and specific basin characteristics, highlighting the value of adaptive or multi-model approaches for improving regional operational forecasts.

This observed variability also underscores some fundamental limitations of LSTM models. These models minimize statistical errors through automatic learning, without incorporating explicit hydrological constraints. Unlike conceptual or semi-distributed models, which are

based on physical or conceptual equations, LSTMs offer no guarantee of physical consistency or transferability between basins. This structure makes them more sensitive to extrapolation, particularly when applied to conditions that were underrepresented during training. If trained on a sufficiently large and diverse set of basins, and with adequate architectural complexity, an LSTM should theoretically be able to adapt to a wide range of hydrological behaviors. However, the regional model used here may lack either the internal complexity or the diversity of training basins required to generalize well to highly contrasting contexts such as those included in this study. The variability observed from one basin or season to another therefore highlights these structural limitations. This emphasizes the importance of expanding the training database by including more representative examples of each basin type, in order to improve the model's generalization ability.

2.6.5 Forecast reliability issues

Forecasts generated by the LSTM models consistently exhibit a lack of reliability across all eight basins analyzed. On an annual basis, most of these basins (Figures 2.8 and 2.9) demonstrate significant underdispersion, illustrated by U-shaped distributions in the Talagrand diagrams. Additionally, the presence of seasonal biases in the seasonal diagrams (as seen for the LSJ basin in Figure 2.8) highlights that some post-processing could help improve the forecast skill to some degree.

These findings are consistent with those reported by Armstrong et al. (2025) in a recent study conducted on the LSJ basin. Their study focused on multi-model ensemble streamflow forecasts and included an evaluation of the individual performance of an LSTM model compared to the same CEQUEAU model with data assimilation, using ensemble meteorological forecasts for the period 2015–2019 with forecast horizons ranging from 1 to 9 days. Armstrong et al. (2025) similarly identified reliability issues in LSTM-based forecasts, although their study focused on daily streamflow forecasts, whereas the present analysis specifically addresses 45-day cumulative volumes. They also demonstrated that a multi-model combination significantly improved forecast reliability. To explore a simple remedy for these

reliability shortfalls, a multi model ensemble was tested by joining CEQUEAU forecasts with those from the LOC and REG models. This combined ensemble produced markedly flatter Talagrand diagrams (Figure 2.10), indicative of better dispersion, and reduced ABDU values in both spring and autumn. In terms of CRPS performance, the multi-model approaches demonstrated enhanced accuracy during winter for both combinations (LOC + CEQUEAU and REG + CEQUEAU). In spring, CEQUEAU and REG + CEQUEAU produced comparable results, while in summer, the LOC + CEQUEAU ensemble was slightly more accurate than CEQUEAU alone. In autumn, the most accurate results were obtained with the REG + CEQUEAU combination (see Figure-A IV-I-1 in Appendix IV). On an annual basis, the multi model approach also yielded a net improvement in both overall reliability and precision compared to any single model. Additionally, a Kolmogorov–Smirnov (KS) test was applied by season to the five scenarios (three single model and two multi-model ensembles) followed a uniform distribution. The KS statistic measures the maximum distance between the empirical and theoretical cumulative distribution functions, and larger p-values indicate greater consistency with the hypothesized uniform distribution. Only CEQUEAU in summer did not reject uniformity (p-value = 0.097). Since $p > 0.05$, the null hypothesis of uniformity is not rejected, indicating statistical compatibility with a uniform distribution; this does not prove uniformity, but it merely suggests that any deviations from uniformity are not statistically significant given the sample size. Nevertheless, the multi-model ensembles generally exhibited larger p-values than the standalone LSTM models, consistent with flatter Talagrand diagrams and lower ABDU scores, indicating better-calibrated spread and improved reliability. These encouraging results suggest that even a basic LSTM + CEQUEAU fusion can substantially mitigate LSTM overconfidence and warrants further investigation in future forecasting studies.

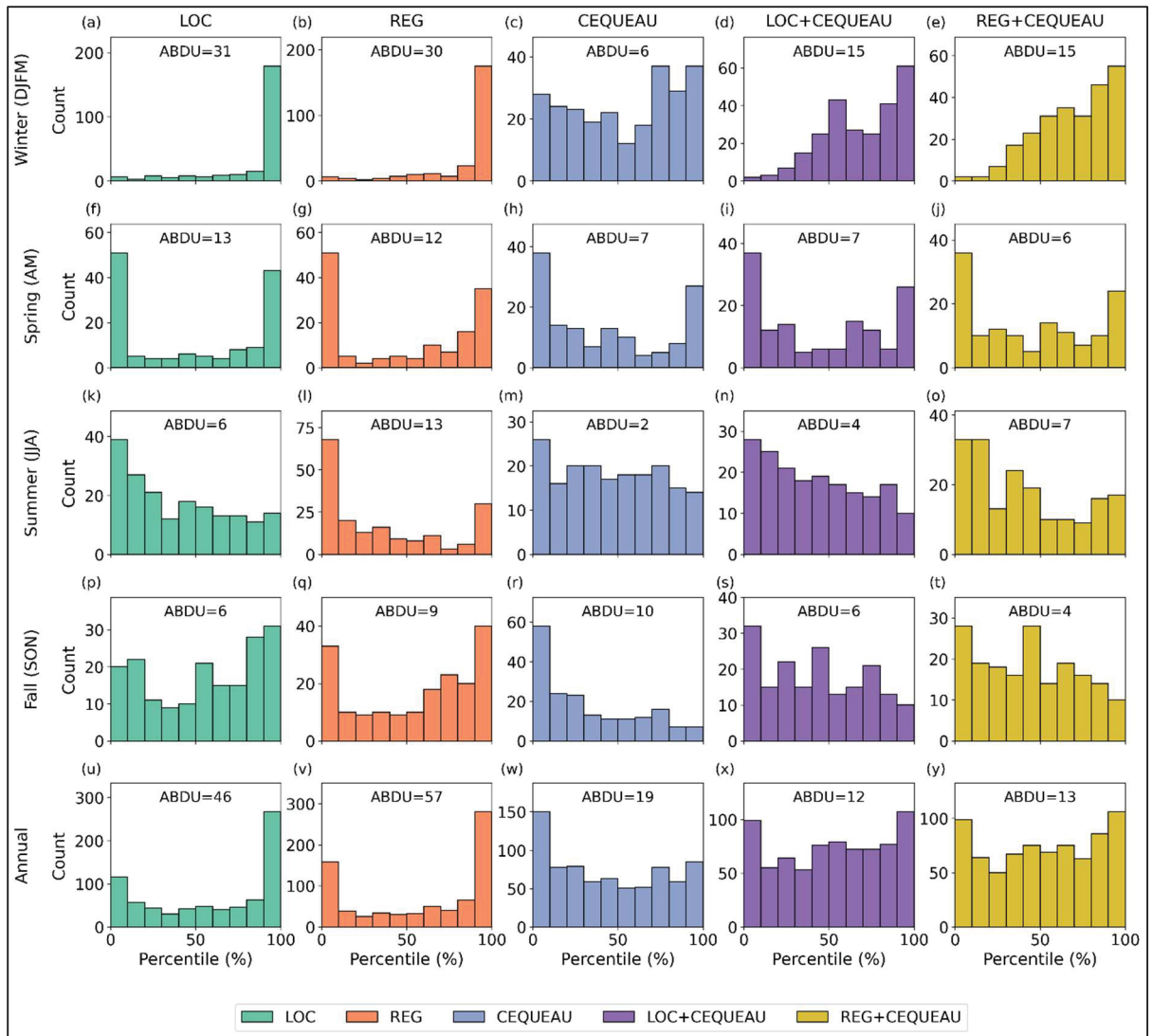


Figure 2.10 Talagrand diagrams of 45 day cumulative inflow forecasts (2016–2023) for the Lac Saint Jean basin for five scenarios: LOC (local LSTM), REG (regional LSTM trained on 200 basins), CEQUEAU (reference hydrological model), and two combined ensembles (LOC+CEQUEAU and REG+CEQUEAU)

An additional point of interest is the difference between the two studies regarding CEQUEAU's results. While Armstrong et al. (2025) reported U-shaped Talagrand diagrams for CEQUEAU, the present study reveals diagrams characterized by better-distributed bands, particularly during winter and summer (flatter distributions). Additionally, the systematic bias observed here in autumn is L-shaped (positive bias), contrasting with the symmetric underdispersion (U-shaped) reported by Armstrong et al. (2025). These differences are most

likely attributable to the distinct temporal ranges, forecast horizons and target variables (daily flow rates vs total inflow volumes) considered in the two studies. Moreover, in the present study, CEQUEAU shows better overall reliability for most seasons for the LSJ basin, except in cases such as the LOC scenario, which displays a more evenly distributed Talagrand diagram. This suggests that while CEQUEAU generally performs well for this basin, certain LSTM scenarios can occasionally outperform it in terms of reliability under specific seasonal conditions. For example, both REG and LOC outperform CEQUEAU in autumn. Further testing of this hypothesis on the seven other test basins, where CEQUEAU cannot be applied, would require comparison with other traditional hydrological models.

Although none of the LSTM-based scenarios demonstrate fully satisfactory reliability, their performance still varies depending on the training strategy used and basin. For example, when comparing the annual ABDU values across the eight basins individually (Figures 2.8 and 2.9), LOC and REG each perform better an equal number of times (4 out of 8). Overall, no clear pattern emerges indicating that one scenario systematically outperforms the others; instead, performance depends strongly on both season and basin. These observations confirm that the structure of the training dataset (LOC vs. REG) has a measurable impact on forecast spread and bias, but also highlight the importance of exploring how different configurations could be combined, adjusted, or complemented. For example, multi-model or hybrid approaches could be implemented to better address these reliability limitations.

2.6.6 Limitations

This study has a few limitations that could be addressed in future studies to improve the robustness of the obtained results. First, the available forecast period is relatively short, covering approximately eight years. This limited time frame reduces the representativeness of rare or extreme events in the forecast data, which may affect the model's performance when evaluated under unusual climatic conditions. Second, it would have been relevant to assess all 200 basins in order to determine whether the methodology performs better in certain regions and to identify any emerging trends. However, this broader evaluation was constrained by the

lack of available weather forecast data and by insufficient or missing streamflow observations during the forecast period. Adding more basins would also help the model train on a more diversified set of basins which could help with robustness and generalizability. In addition, the model hyperparameters (i.e. model structure) were optimized solely for the LSJ basin. As a result, the performance of the scenarios for the seven other basins tested may not be fully optimal. This specific calibration may have limited the immediate transferability of the models, even though one of the main objectives of this study was precisely to evaluate the possibility of directly applying a generic methodology to different contexts. Finally, because the comparison relies on a single CEQUEAU configuration, it is difficult to ascertain whether the finding that CEQUEAU still outperforms LSTM for forecasting would hold across alternative calibrations, periods, and versions of the model, and it should therefore be interpreted with caution.

2.7 Conclusion and recommendations

This study aimed to assess the ability of LSTM models to predict cumulative 45-day flood volumes for the LSJ basin in Canada and to evaluate the potential added value of training with an expanded regional dataset of 200 basins across the country. To evaluate the robustness of the methodology, seven target basins were tested. Finally, to provide a benchmark against operational practice, the results of the different forecasting scenarios were compared with those produced by the physically based, semi-distributed CEQUEAU model, which is already calibrated and used by Rio Tinto for their hydropower operations. Overall, the results show that while the ensemble forecasts are accurate on average, they lack reliability and tend to be overconfident. A second objective was to assess whether regional training improves model performance by leveraging a more diverse set of basins. The use of regional models can improve forecast quality depending on the season and the basin. There does not appear to be a clear pattern where one scenario consistently performs better in a specific season, but the regional model did improve upon the LOC model in general, with a few notable exceptions. The method thus shows promise and could be further improved by further increasing the size of the training dataset (i.e. adding more training basins). Another goal was to verify whether

scenarios calibrated for the LSJ basin could be transferred to other basins. The results indicate that while some transferability is possible, performance varies greatly between basins and depends strongly on local hydrological conditions. Finally, the LSTM scenarios were compared to the CEQUEAU model for LSJ currently used in operational hydropower for Rio Tinto. While forecasts produced with CEQUEAU generally remain more performant overall, LSTM models showed better performance in fall in both reliability and accuracy, highlighting their potential as complementary tools in a multi-model framework rather than standalone replacements.

This study shows that LSTM models demonstrate real potential for seasonal volume forecasting, especially in the context of strategic planning. However, their effectiveness strongly depends on the type of input data, the training strategy, the target basin, the season, and above all, the operational objectives. Hybrid approaches combining physical models and deep learning, as suggested by Armstrong et al. (2025), could be a promising path to reconcile accuracy, reliability, and the representation of uncertainty. In summary, while LSTM models show strong potential for seasonal inflow volume forecasting, enhancing their reliability and representation of uncertainty is crucial to support robust operational hydrological management.

2.8 Acknowledgments

This work was partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number ALLRP-580910-22. The authors would like to thank Rio Tinto for providing observed streamflow data and access to the calibrated CEQUEAU hydrological model. We also acknowledge the European Centre for Medium-Range Weather Forecasts (ECMWF) for access to the medium-range and extended-range ensemble forecasts, as well as the Copernicus Climate Change Service (C3S) for providing ERA5 reanalysis data. The use of ChatGPT-4 for minor spelling and stylistic revisions during manuscript preparation is gratefully acknowledged. The base map in Figure 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

Cette étude a démontré que les modèles LSTM présentent un potentiel réel pour la prévision des volumes d'eau sur un horizon de 45 jours. Bien que la valeur ajoutée d'un modèle régional par rapport à un modèle local varie considérablement selon les bassins versants et les saisons, cette approche demeure pertinente dans une perspective opérationnelle. Cependant, la méthodologie présente encore certaines limites sur le plan de la fiabilité. Un bref test de l'application d'une approche multi-modèle combinant un réseau LSTM et un modèle conceptuel, réalisé sur le bassin du Lac-Saint-Jean, a permis d'améliorer significativement la fiabilité des prévisions saisonnières, en particulier au printemps et en automne. Ces résultats, bien que préliminaires, soutiennent la pertinence de poursuivre les recherches dans cette direction.

Les sections qui suivent abordent successivement : l'adaptation de la méthode à différents horizons de prévision, ainsi que les défis rencontrés lors de la programmation des modèles LSTM. Enfin, des résultats exploratoires complémentaires, non inclus dans l'article principal, présentent un scénario basé sur l'entraînement d'un modèle LSTM à partir des prévisions météorologiques d'ensemble, plutôt que de données historiques, afin d'examiner les limites et le potentiel associées à cette approche.

3.1 Adaptation de la méthode à divers horizons de prévisions

La méthode présentée dans l'article peut être adaptée pour produire des prévisions de volumes à différentes échéances, que ce soit pour des horizons plus courts ou plus longs que 45 jours. Par exemple, le rapport de Soucy (2025) a utilisé un réseau LSTM pour prévoir les volumes cumulés à 14 jours sur le bassin du Lac-Saint-Jean. Pour des échéances plus courtes, comme une prévision à 5 jours, le fonctionnement reste semblable. La fenêtre temporelle transmise au modèle LSTM peut toujours couvrir 365 jours afin de rester cohérente avec l'approche initiale, mais il faut ajuster la quantité de données météorologiques disponibles avant le début de la

période à prévoir. Si l'objectif est de prédire un volume sur 5 jours, la séquence d'entrée comprendra alors 360 jours de données météorologiques observées, suivis des 5 jours suivants à partir de la date d'émission de la prévision. Lors de l'entraînement, ces 5 jours sont connus et utilisés comme observations. En phase de prévision, ils sont remplacés par les données issues des prévisions météorologiques. Pour des échéances dépassant 45 jours, la méthode doit être adaptée, car les prévisions météorologiques disponibles ne couvrent pas ces périodes de manière fiable et ne sont pas émises régulièrement. Une solution consiste à prolonger les séquences à l'aide de données historiques. Par exemple, pour une prévision sur 90 jours débutant le 1er janvier 2024, les 45 premiers jours (du 1er janvier au 14 février) peuvent être pris directement à partir des prévisions météorologiques du modèle ECMWF. Les 45 jours suivants (du 15 février au 30 mars), qui dépassent l'horizon de ces prévisions, peuvent être comblés à l'aide de données météorologiques historiques. Pour cela, les données observées entre le 15 février et le 30 mars sont extraites pour chacune des années antérieures disponibles, par exemple 30 ans, ce qui permet de construire 30 prolongements possibles de la séquence initiale. Chaque année historique représente un scénario météorologique plausible pour cette deuxième moitié de l'horizon de prévision. En combinant ces 30 années avec les 50 membres ensemblistes d'ECMWF, un total de 1350 trajectoires météorologiques différentes est obtenu. Cette approche permet d'allonger l'horizon de prévision tout en intégrant l'incertitude liée aux conditions futures, grâce à une combinaison de prévisions réelles et de données historiques. Le fait d'utiliser à la fois des données de prévisions et des données historiques pour prolonger l'horizon de prévision a d'ailleurs été exploré par Swift-LaPointe et al. (2025). Leur approche visant à estimer un volume cumulé sur neuf mois est toutefois différente, puisqu'elle repose sur l'utilisation de modèles LSTM entraînés sur des valeurs mensuelles de débit. Le modèle est généré neuf fois, une pour chaque mois de la période cible, puis les résultats sont additionnés afin d'obtenir le volume total.

3.2 Défis liés à la programmation et l'entraînement des modèles LSTM

L'intelligence artificielle appliquée à l'hydrologie représente encore un domaine émergent. L'exploration initiale des outils de programmation et des techniques de d'apprentissage

automatisé s'accompagne naturellement de plusieurs défis. Cette section présente plusieurs constats issus des expérimentations menées avec les modèles LSTM, en mettant l'accent sur les aspects techniques liés au traitement des données et à l'entraînement des réseaux.

Parmi les éléments importants à prendre en compte figure notamment la question des ressources informatiques nécessaires à la préparation et à l'entraînement des modèles. L'analyse simultanée d'un grand nombre de bassins versants exige un ordinateur performant, doté d'une carte graphique configurée pour l'apprentissage profond (GPU) et d'une grande capacité de mémoire vive (RAM). Parmi les étapes les plus énergivores, la préparation des données se distingue lorsque le nombre de bassins versants est nombreux, en raison du volume de données météorologiques à intégrer, structurer et normaliser sur de longues périodes. L'étape de normalisation consiste à mettre à l'échelle les variables d'entrée en les ajustant sur une échelle commune. Cette opération évite que le modèle accorde automatiquement plus d'importance à une variable uniquement parce que ses valeurs numériques sont plus grandes que celles des autres (par exemple, la pression exprimée en pascals comparée à la température en degrés Celsius). En plus de mieux équilibrer l'influence des variables, la normalisation contribue à accélérer l'apprentissage, en simplifiant le traitement des données et en favorisant une convergence plus rapide du modèle. Dans cette étude, le traitement d'un grand nombre de bassins versants a rendu la préparation des données particulièrement exigeante. Par exemple, le jeu de données utilisé comprenait 200 bassins, 8 variables météorologiques et environ 30 ans de données journalières en moyenne. En créant des fenêtres d'apprentissage de 365 jours pour entraîner le modèle, cela a généré environ 6,4 milliards de valeurs individuelles à traiter simultanément, sans même compter les valeurs cibles ni les variables statiques. Ce chiffre donne un bon aperçu de la quantité importante de données à manipuler avant même d'entamer la phase d'entraînement du modèle. La normalisation ne peut pas être appliquée avant la création de ces séquences de 365 jours, puisque celle-ci doit être effectuée exclusivement à partir des séquences vues en entraînement. Il est donc nécessaire, dans un premier temps, de transformer les données brutes en séquences, puis de les diviser en trois ensembles distincts (entraînement, validation et test) avant d'appliquer la normalisation. Cette contrainte explique en grande partie la forte consommation de mémoire associée à cette étape du processus. Enfin,

une fois les données préparées, l'entraînement du modèle constitue la phase la plus coûteuse en ressources. L'optimisation du code s'avère alors cruciale, notamment pour éviter que des variables intermédiaires ne soient conservées inutilement en mémoire, ce qui peut entraîner des erreurs liées à sa surcharge.

D'autres défis concernent les différences entre l'entraînement des modèles régionaux et celui des modèles locaux. Ces derniers, puisqu'ils sont calibrés à partir d'un volume de données bien plus restreint, présentent une plus grande variabilité d'un entraînement à l'autre, même lorsque les hyperparamètres sont identiques. Ce genre de variation est courant avec les modèles LSTM, qui utilisent des méthodes d'optimisation stochastique et dont le fonctionnement peut changer légèrement selon la façon dont l'apprentissage est initialisé. Cela rend parfois difficile de savoir si un changement de paramètre a réellement amélioré le modèle, ou si la différence observée est simplement due au hasard. Pour mieux prendre en compte ce côté aléatoire, chaque combinaison de paramètres a été testée dix fois, et la meilleure des dix simulations a été conservée pour l'analyse. À l'inverse, le modèle régional a donné des résultats beaucoup plus stables : les performances restaient presque identiques d'un essai à l'autre, ce qui rend l'évaluation de l'impact des changements liés aux hyperparamètres plus fiable. Toutefois, cette stabilité vient avec un coût : alors qu'un modèle local s'entraîne en environ 3 minutes, un modèle régional a nécessité jusqu'à 3 heures de calcul. Ce temps pourrait toutefois être réduit avec un ordinateur plus performant, par exemple en utilisant un GPU plus puissant, ce qui permettrait d'accélérer considérablement l'entraînement. Là où le modèle LSTM devient particulièrement intéressant, c'est qu'une fois entraîné, il peut produire des prévisions de volumes très rapidement, même à partir de données complexes comme les prévisions météorologiques ensemblistes. À l'inverse, un modèle hydrologique traditionnel comme CEQUEAU doit être exécuté séparément pour chaque membre météorologique, et cela pour chaque jour de prévision, puisqu'il fonctionne de manière séquentielle dans le temps. Ce fonctionnement rend le processus particulièrement long lorsque des prévisions doivent être émises sur plusieurs années : le temps total de calcul peut alors atteindre jusqu'à 5 heures pour 7 ans de données de prévision. De plus, à la fin de chaque simulation, il est nécessaire d'effectuer une conversion des débits journaliers en volumes cumulés, ce qui ajoute une étape

de traitement supplémentaire. En comparaison, le modèle LSTM, une fois calibré, est capable de générer l'ensemble des prévisions en quelques minutes seulement.

Un autre défi a été la construction du modèle régional, qui a nécessité une approche progressive, en testant différentes configurations intermédiaires. Avant d'aboutir à un modèle régional fonctionnel entraîné sur 200 bassins versants, plusieurs configurations intermédiaires ont été testées, incluant des regroupements de 5, 10, 30 ou 50 bassins. Ce processus a soulevé plusieurs défis, notamment parce que toutes les combinaisons de bassins ne donnaient pas nécessairement de bons résultats. L'ajout progressif de bassins permettait en général d'améliorer les performances globales, mais cette amélioration n'était pas systématique. Plus le nombre de bassins augmentait, plus il devenait nécessaire de réévaluer les hyperparamètres du modèle, car une configuration qui fonctionne bien pour un petit groupe de bassins ne s'applique pas toujours à un ensemble plus large. De plus, au moment de lancer un entraînement LSTM, des problèmes techniques pouvaient survenir, comme un arrêt complet du processus ou un comportement anormal, où le modèle produisait exactement les mêmes résultats pendant plusieurs itérations, sans progrès visible. Dans ces cas, il suffisait parfois de relancer l'entraînement ou simplement de laisser tourner plus longtemps pour que le modèle « débloque » et reprenne une trajectoire d'apprentissage normale.

3.3 Exploration d'un scenario base sur les prévisions ensemblistes

Afin d'explorer davantage les possibilités d'amélioration des prévisions, une autre expérience a été conçue : l'entraînement du modèle LSTM local, mais cette fois directement entraîné à partir des données de prévisions météorologiques ensemblistes (L_FC). Cette approche pourrait permettre au réseau non seulement d'apprendre les biais systématiques présents dans les prévisions météo afin de mieux les corriger lors de la prévision des volumes, mais aussi de mieux représenter l'incertitude grâce à l'apprentissage sur des scénarios ensemblistes. À première vue, disposer de seulement huit années de données semble insuffisant pour entraîner efficacement un modèle LSTM. Toutefois, l'utilisation de données ensemblistes, composées de 50 membres et émises 2 fois par semaine, augmente fortement la taille de l'échantillon.

Un total de 60 % des données de prévision est utilisé pour l'entraînement, 20 % pour la validation, et les 20 % restants pour le test. Il est à noter que la vérification des prévisions sur l'ensemble de test repose ainsi sur une période plus courte que pour les scénarios LOC et REG. Cette limitation est prise en compte en comparant tous les modèles sur le sous-ensemble commun des données de prévision utilisées pour la période test du scénario L_FC.

Le tableau 3.1 présente les performances d'entraînement des trois scénarios. Étant donné que les résultats de calibration du scénario L_FC sont de nature ensembliste, contrairement aux scénarios LOC et REG, qui sont déterministes, les valeurs de KGE sont calculées sur la moyenne des 50 membres.

Tableau 3.1 KGE scores for the training of the LOC, REG and L_FC scenarios

N° bassin	Training			Validation			Testing		
	LOC	REG	L_FC	LOC	REG	L_FC	LOC	REG	L_FC
1	0.976	0.974	0.899	0.917	0.968	0.690	0.945	0.926	0.832
2	0.914	0.969	0.922	0.915	0.967	0.930	0.880	0.953	0.597
3	0.943	0.968	0.940	0.921	0.974	0.751	0.914	0.937	0.880
4	0.875	0.972	0.772	0.806	0.972	0.828	0.840	0.871	0.877
5	0.702	0.966	0.662	0.738	0.967	0.625	0.667	0.600	0.740
6	0.801	0.980	0.815	0.655	0.971	0.738	0.519	0.611	0.800
7	0.902	0.970	0.830	0.905	0.969	0.849	0.912	0.933	0.740
8	0.861	0.968	0.945	0.816	0.972	0.675	0.706	0.791	0.489

Le scénario L_FC, entraîné directement sur les prévisions météorologiques ensemblistes, a produit des résultats mitigés, avec des valeurs de KGE en test variant de 0,489 à 0,880, et a montré une instabilité marquée entre les phases d'entraînement, de validation et de test. Par exemple, pour le bassin 1 (LSJ), le KGE atteint 0,899 en entraînement, chute à 0,690 en validation, puis remonte à 0,832 en test, ce qui suggère un surapprentissage potentiel dû à une variabilité limitée des données d'entrée. De manière générale, cette instabilité contraste avec

la plus grande robustesse des scénarios LOC et REG, qui bénéficient probablement d'une meilleure cohérence temporelle et d'une plus grande diversité grâce aux données historiques observées. Cependant, L_FC peut parfois offrir des résultats plus stables pour certains bassins. C'est le cas du bassin 6 : le KGE reste relativement constant avec 0,815 en entraînement, 0,738 en validation, et 0,800 en test. Il est toutefois important de garder à l'esprit que ces résultats s'appuient sur une période d'analyse beaucoup plus courte, les phases de validation et de test représentant chacune seulement 20 % des données disponibles sur huit ans. Cela limite les comparaisons directes avec les autres scénarios, dont les périodes de données s'étendent de 41 à 44 ans par bassin.

Sur la période commune à l'ensemble des scénarios, L_FC présente des performances très variables et obtient rarement les médianes de CRPS les plus faibles, toutes saisons confondues (Figure 3.1). En revanche, à l'échelle globale par bassin, L_FC apparaît comme le plus précis pour les bassins 3 et 7. Pour le bassin 1 (LSJ), les diagrammes de Talagrand saisonniers (Figure 3.2) montrent une sous-dispersion en automne et en été. En hiver, les prévisions présentent un biais négatif plus faible, comparable à celui de CEQUEAU. Au printemps, tous les scénarios affichent un niveau de fiabilité similaire (ABDU compris entre 3 et 4). À l'échelle annuelle, le scénario présente des prévisions sous-dispersées tant pour le bassin 1 (LSJ) que pour les sept autres bassins (Figure 3.3). Cependant, pour les bassins 6, 7 et 8, le scénario L_FC est celui dont les valeurs d'ABDU sont les plus faibles (signe d'une meilleure fiabilité). Il ne faut toutefois pas comparer directement les valeurs d'ABDU d'un bassin à l'autre : le nombre de jours disponibles diffère selon les bassins, ce qui influence cet indicateur. La comparaison est pertinente uniquement entre les scénarios d'un même bassin.

Le test KS a été mené afin de déterminer si, statistiquement, les prévisions suivent une loi uniforme. Les diagrammes de Talagrand dont le texte ABDU noir sont ceux pour lesquels la valeur-p est inférieure au seuil de 0,05, ce qui indique un écart statistiquement significatif par rapport à une loi uniforme et une fiabilité jugée discutable. La valeur-p représente ici la probabilité d'obtenir une différence aussi marquée entre la distribution observée des rangs de prévision et celle attendue d'un modèle parfaitement fiable. À l'inverse, les diagrammes de

Talagrand dont le texte ABDU est rouge correspondent aux cas où l'uniformité n'a pas été rejetée, indiquant une fiabilité jugée satisfaisante selon un seuil de confiance de 95 %.

Selon les résultats du test KS, le scénario L_FC est le moins fiable, avec seulement les prévisions au moins d'hiver jugées fiables. CEQUEAU est le modèle le plus fiable, avec un niveau de fiabilité acceptable pour les mois d'hiver, de printemps et d'été. Il ne faut toutefois pas comparer ces résultats avec ceux de l'article, car ils ne portent pas sur les mêmes périodes de données.

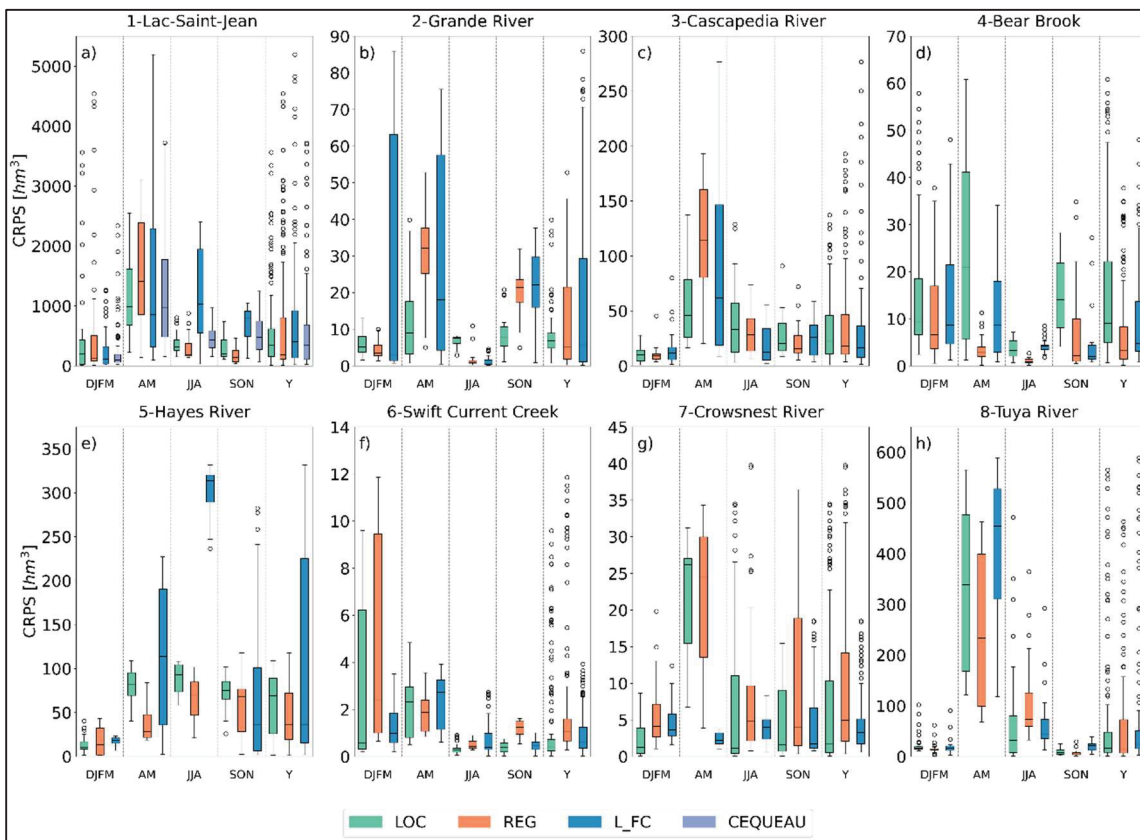


Figure 3.1 CRPS des prévisions ensemblistes de volume par saison pour les scénarios : LOC, REG et L_FC. Le modèle CEQUEAU est également présenté dans le panneau a) pour le Lac-Saint-Jean. Chaque panneau représente un bassin versant : a) Lac-Saint-Jean [1], b) Grande River [2], c) Cascapedia River [3], d) Bear Brook [4], e) Hayes River [5], f) Swift Current Creek [6], g) Crowsnest River [7], and h) Tuya River [8]. Les saisons sont définies comme suit : hiver (DJFM), printemps (AM), été (JJA), automne (SON) et annuel (Y). La période de prévision correspondant aux derniers 20 % de la période de test du scénario L_FC, entre 2016 et 2023

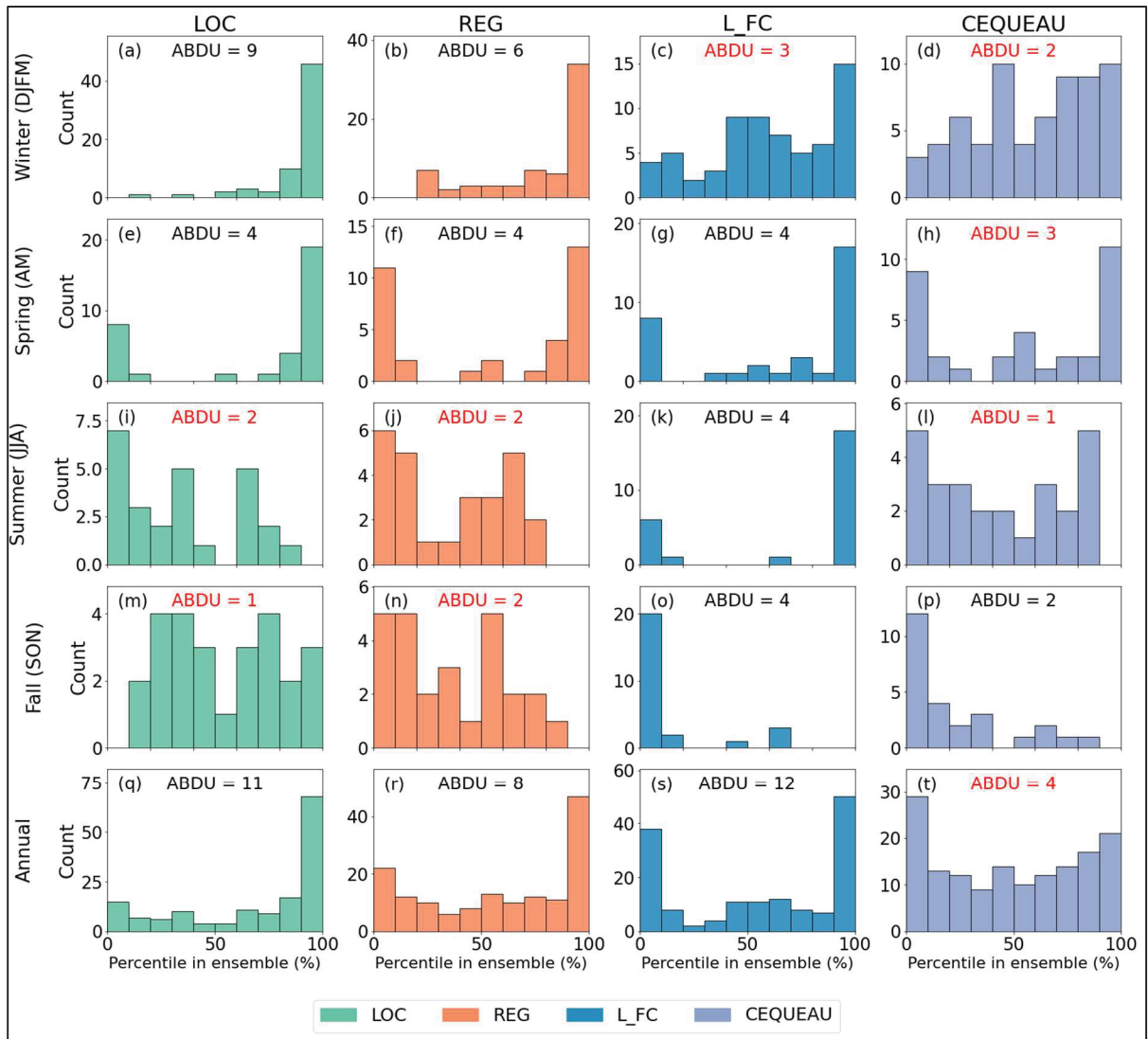


Figure 3.2 Fiabilité des prévisions pour le bassin du Lac-Saint-Jean selon quatre scénarios de modélisation : LOC, REG, L_FC et CEQUEAU. La figure présente des diagrammes de Talagrand illustrant la distribution des rangs de prévision (en percentiles) des volumes d'eau cumulés sur 45 jours pour chaque scénario. La fiabilité des prévisions est évaluée sur les dates communes des scénarios, du 20 décembre 2021 au 29 mai 2023 inclusivement. Les sous-figures dont le texte ABDU est rouge correspondent aux cas où la distribution des rangs ne s'écarte pas significativement d'une loi uniforme (selon un test de Kolmogorov-Smirnov avec un seuil de 0,05), suggérant une fiabilité satisfaisante. À l'inverse, les sous-figures dont le texte ABDU est noir indiquent une divergence statistiquement significative par rapport à l'uniformité, suggérant un manque de fiabilité

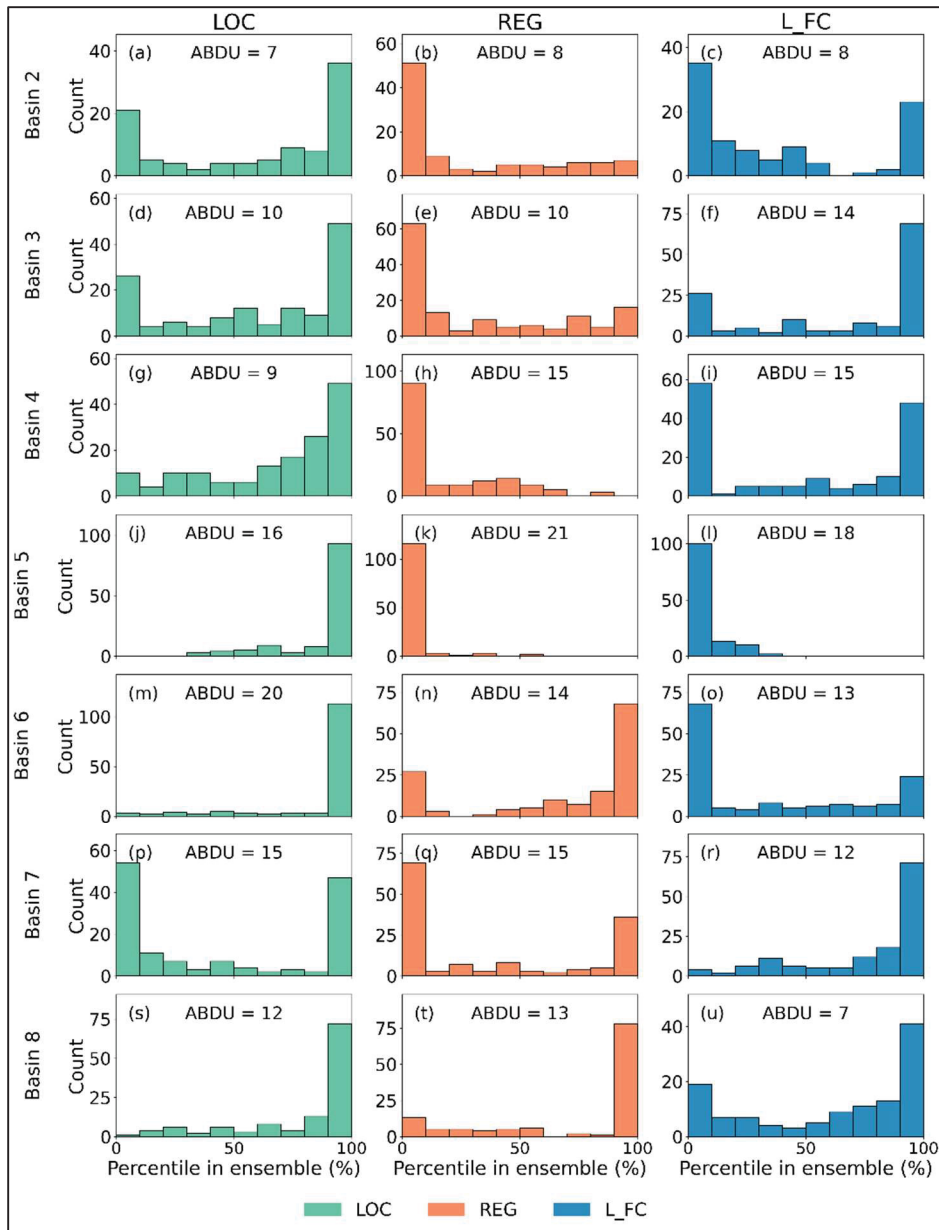


Figure 3.3 Fiabilité annuelle des prévisions pour trois scénarios de modélisation basés sur des LSTM : LOC, REG et L_FC. Les diagrammes de Talagrand représentent la distribution des rangs de prévision (en percentiles) des volumes d'eau cumulés sur 45 jours pour chaque scénario, sur sept bassins canadiens numérotés de 2 à 8 : Grande Rivière [bassin 2, panneaux a–b], Rivière Cascapédia [3, c–d], Bear Brook [4, e–f], Rivière Hayes [5, g–h], Swift Current Creek [6, i–j], Rivière Crowsnest [7, k–l] et Rivière Tuya [8, m–n]. La période de prévision correspondant aux derniers 20 % de la période de test du scénario L_FC, entre 2016 et 2023

3.3.1 Limites de l'apprentissage à partir des prévisions ensemblistes

La figure 3.4 illustre clairement les limites du scénario L_FC de l'entraînement direct à partir de prévisions météorologiques ensemblistes.

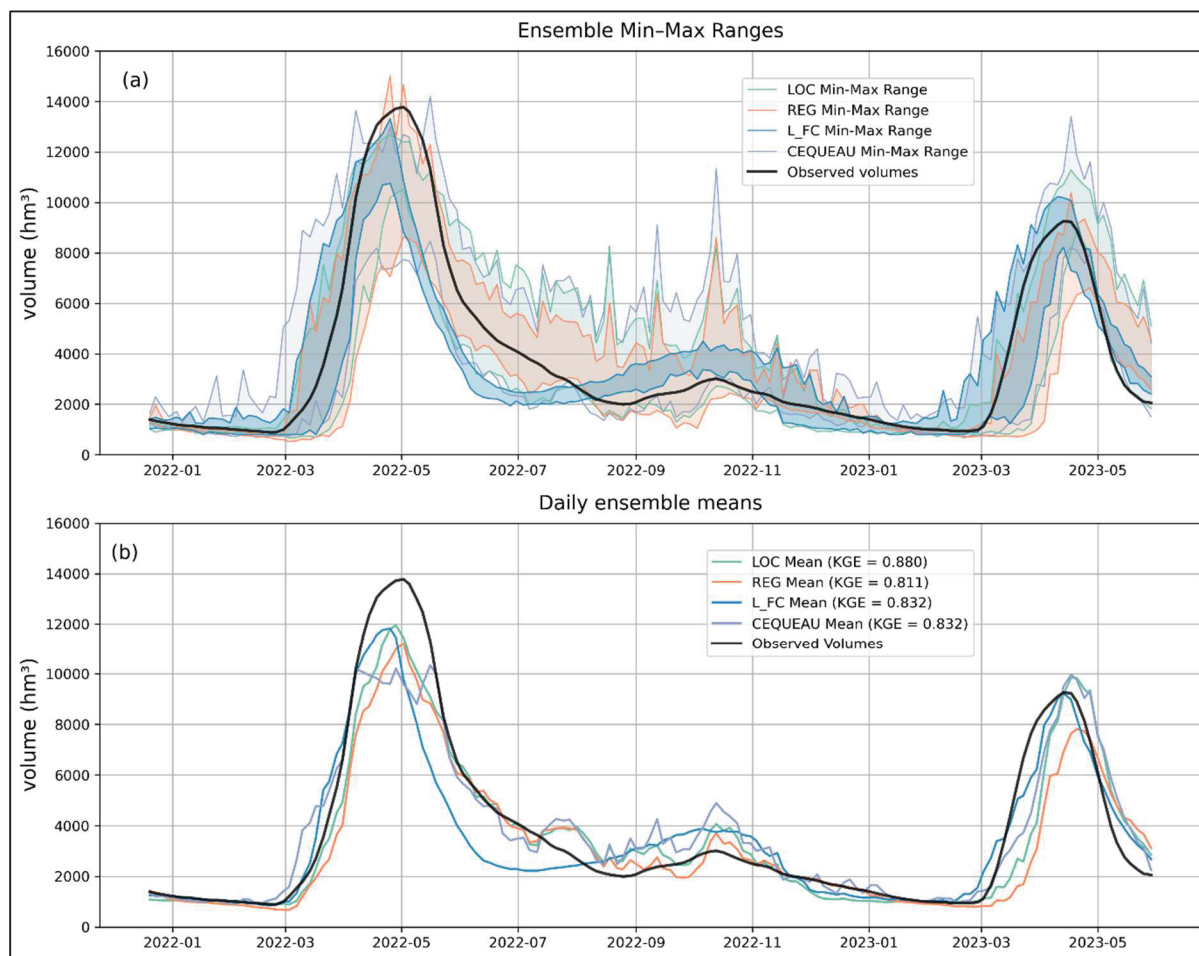


Figure 3.4 Prévisions ensemblistes des volumes d'apport sur 45 jours pour le bassin du Lac-Saint-Jean : (a) Enveloppes min-max, (b) Moyennes journalières des prévisions pour les quatre scénarios : LOC, REG, L_FC et CEQUEAU. La période de prévision s'étend du 20 décembre 2021 au 29 mai 2023

Une hypothèse possible pour expliquer la sous-dispersion marquée des prévisions est que les données d'entrée provenant des prévisions météorologiques réduisent considérablement la diversité de l'ensemble d'entraînement. En effet, même si chaque jour comporte 50 scénarios météorologiques différents, ils sont tous associés à une seule valeur cible de débit observé.

Cette situation peut pousser le modèle à chercher une solution qui fonctionne en moyenne pour l'ensemble des membres, plutôt que de capter la variabilité propre à chacun. En d'autres termes, au lieu d'apprendre des différences fines entre les scénarios météo, le modèle pourrait préférer donner toujours à peu près la même réponse, car c'est ce qui minimise l'erreur moyenne pendant l'apprentissage. Ce comportement peut conduire à des prédictions sous-dispersées, indépendamment des différences réelles entre les membres d'entrée. Une telle dispersion artificiellement réduite ne garantit ni la précision, ni la fiabilité des prévisions : l'ensemble peut ne pas inclure la valeur observée. Ainsi, les prévisions générées pour le bassin du LSJ apparaissent anormalement concentrées, ce qui reflète un manque d'incertitude réaliste malgré la diversité initiale des conditions météorologiques simulées. Un important décalage entre les prévisions et les observations est observé au printemps et à l'été 2022. Ce phénomène pourrait notamment expliquer la sous-dispersion observée dans les diagrammes de Talagrand (Figures 3.2 et 3.3), ainsi qu'une représentation inadéquate de la plage des volumes de crue possibles (comme le montre la Figure 3.4). Une solution envisageable serait d'entraîner le modèle LSTM sur la moyenne des 50 membres, puis de lui fournir l'ensemble des membres en entrée afin de générer la plage complète de prévisions. Toutefois, compte tenu du nombre limité d'années disponibles, cette approche reste difficilement applicable avec un modèle entraîné sur un seul bassin: un test mené sur le bassin du Lac-Saint-Jean avec le modèle local LSTM a montré que le modèle ne convergeait pas de manière satisfaisante dû au manque de données. Une telle stratégie pourrait néanmoins devenir envisageable à l'avenir si le modèle régional existant est utilisé pour augmenter la diversité et le volume des données d'entraînement. Ainsi, le scénario REG entraîné sur les données météorologiques historiques pourrait être repris et recalibré localement avec la moyenne des données de prévisions d'ensembles du bassin cible.

L'avantage théorique de l'apprentissage direct à partir des données de prévision est qu'il permet au modèle d'intégrer les biais systématiques présents dans les prévisions météorologiques, plutôt que d'être uniquement calibré sur des valeurs observées, puis appliqué à des prévisions qui peuvent différer de manière significative. Même si le modèle est effectivement entraîné sur la moyenne des membres ensemblistes, ce qui revient à ignorer la

variabilité entre les scénarios individuels, le modèle reste exposé aux biais systématiques des prévisions globales, ce qui lui permettrait d'apprendre à les corriger implicitement.

CONCLUSION

En conclusion ce mémoire répond à la question posée en introduction: pourquoi ne pas entraîner directement les modèles sur les volumes d'eau, en montrant que les LSTM peuvent effectivement prévoir directement les volumes cumulés, alignant ainsi les prévisions sur l'indicateur opérationnel? Cependant, cette approche n'est pas encore parfaite, des problèmes de fiabilité subsistent, notamment des biais, une variabilité saisonnière et des performances qui varient selon les bassins. Afin d'évaluer la capacité des LSTM à prévoir directement les volumes, trois scénarios de prévision utilisant des modèles LSTM ont été explorés : un modèle local entraîné avec les données météorologiques historiques d'un seul bassin versant, un modèle régional entraîné sur les données historiques de 200 bassins versants répartis à travers le Canada, et un modèle local entraîné directement à partir des données de prévisions météorologiques ensemblistes d'un bassin versant. Des tests complémentaires ont également permis de dégager des pistes quant à l'amélioration de la qualité des prévisions.

L'évaluation de ces scénarios a révélé plusieurs points clés concernant les performances des modèles LSTM. Premièrement, les modèles LSTM sont capables de prévoir les volumes d'eau cumulés sur 45 jours de façon précise, mais ont tendance à être surconfiants, ce qui entraîne une baisse de fiabilité. Deuxièmement, l'ajout de bassins régionaux donne des résultats mitigés: dans certains cas, les performances sont meilleures comparativement à un modèle entraîné uniquement sur un seul bassin, mais cela varie d'un bassin à l'autre et d'une saison à l'autre. Troisièmement, la combinaison des prévisions provenant d'un LSTM avec un modèle conceptuel est prometteuse, car elle améliore à la fois la précision et la fiabilité, notamment en automne et au printemps pour le bassin du Lac-Saint-Jean. Quatrièmement, l'entraînement direct sur les prévisions ensemblistes s'est montré moins stable que les approches basées sur les données observées, en raison d'un jeu de données plus court et moins varié ; toutefois, certains bassins ont affiché de bonnes performances selon les saisons, ce qui suggère un potentiel dans des contextes ciblés. Cinquièmement, l'entraînement direct des prévisions météo réduit considérablement la plage d'incertitude des prévisions de volume, ce qui génère également de forts biais.

Enfin, ce mémoire a permis d'ouvrir la porte aux études portant sur la prévision directe de volumes d'eau. Bien que la méthodologie testée soit pour l'instant moins fiable que l'approche basée sur l'agrégation des prévisions journalières d'un modèle conceptuel, il s'agit d'un domaine encore très récent et peu documenté. Les prochaines études devraient se concentrer sur d'autres horizons de prévision et d'autres bassins, explorer des architectures d'apprentissage profond plus récentes (par exemple des modèles de type Transformer comme le TFT) ainsi qu'intégrer une étape de post-traitement.

RECOMMANDATIONS

Les prochaines études devraient se concentrer sur les faiblesses identifiées par cette nouvelle approche de prévision des volumes. Entraîner un modèle d'intelligence artificielle directement sur les volumes d'eau reste une méthode encore peu explorée, mais qui présente un bon potentiel d'amélioration. Plusieurs pistes peuvent être envisagées :

- 1) Effectuer une sélection plus ciblée des variables d'entrée, en testant le retrait de certaines variables pour évaluer si leur absence nuit aux performances du modèle, ou au contraire permet de les améliorer.
- 2) Tester l'approche sur d'autres bassins versants, dans une optique multimodale, pour vérifier si les gains observés dans cette étude se généralisent à d'autres contextes hydroclimatiques.
- 3) Ajouter une étape de correction des biais dans les données de prévision météorologique, en amont de l'entraînement, afin d'améliorer la qualité des entrées et de réduire les erreurs systématiques.
- 4) Intégrer un post-traitement des prévisions de volume, afin de corriger les biais résiduels du modèle et ainsi renforcer la précision et la fiabilité des résultats finaux.
- 5) Tester d'autres horizons de prévision de volumes d'eau.
- 6) Changer le modèle LSTM pour un modèle plus récent comme le Temporal Fusion Transformer (TFT). Ce type d'architecture d'apprentissage profond a montré son efficacité sur des données séquentielles et constitue une piste prometteuse pour de futurs travaux.

Une autre piste, plus méthodologique, consisterait à entraîner un modèle régional directement sur la moyenne des prévisions ensemblistes, plutôt que sur l'ensemble complet. Cette approche permettrait de contourner la contrainte liée à la faible disponibilité des données de prévision sur certains bassins, tout en simplifiant la phase d'apprentissage. Une fois le modèle entraîné, les scénarios météorologiques de l'ensemble prévisionnel pourraient ensuite être utilisés en entrée pour générer des prévisions ensemblistes, même si le modèle n'a été exposé qu'à la moyenne durant l'entraînement. Cette stratégie permettrait ainsi de restituer une distribution des prévisions. L'intérêt théorique d'une telle approche réside dans le fait que l'entraînement

directement sur les données de prévision permettrait au modèle d'apprendre à corriger les biais systématiques présents dans ces prévisions météorologiques, ce qui contribuerait à éviter les décalages potentiels pouvant survenir lorsqu'un modèle est calibré uniquement à partir d'observations, mais utilisé avec des données biaisées en opérationnel. Cela dit, il faudrait aussi considérer le fait que cette approche réduirait de manière significative le nombre de données disponibles pour l'entraînement, utilisant un seul membre moyen plutôt que l'ensemble complet de 50 membres.

ANNEXE I

DÉBIT QUOTIDIEN MOYEN AVEC ENVELOPPE MIN-MAX POUR LES HUIT BASSINS TESTS

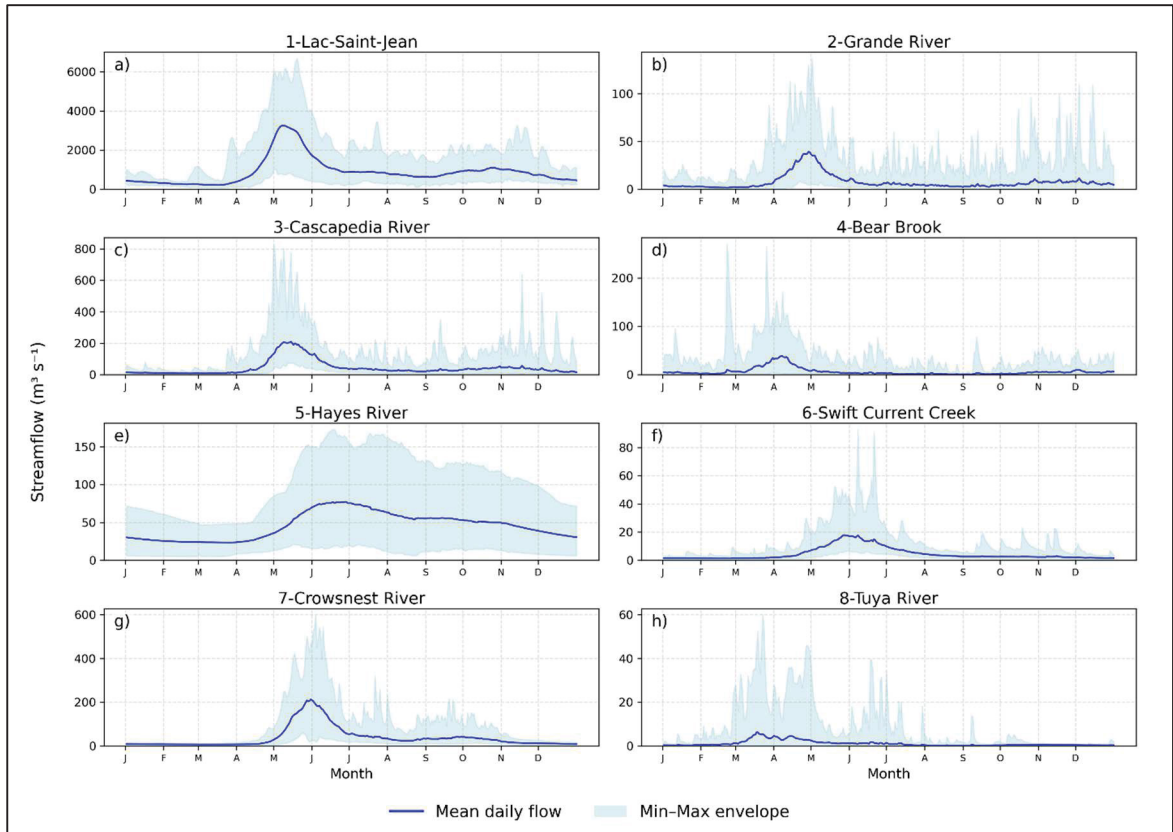


Figure-A I-1 Débit quotidien moyen avec enveloppe min-max pour les huit bassins tests (1979–2023). Chaque panneau représente un bassin versant différent : a) Lac-Saint-Jean [1], b) Rivière Grande [2], c) Rivière Cascapédia [3], d) Bear Brook [4], e) Rivière Hayes, f) Ruisseau Swift Current [6], g) Rivière Crowsnest [7], et h) Rivière Tuya [8]

ANNEXE II

COMPOSANTES DU KGE POUR LES SCÉNARIOS LOC ET REG

Tableau-A II-1 Coefficient de corrélation sans dimension r pour la calibration des deux scénarios : LOC et REG

N ⁰	Basin	LOC			REG		
		Training	Validation	Testing	Training	Validation	Testing
1	Lac-Saint-Jean	0.976	0.942	0.967	0.984	0.972	0.961
2	Grande River	0.957	0.916	0.911	0.984	0.971	0.971
3	Cascapedia River	0.970	0.943	0.951	0.981	0.974	0.978
4	Bear Brook	0.928	0.828	0.888	0.982	0.973	0.941
5	Hayes River	0.805	0.811	0.675	0.982	0.972	0.634
6	Swift Current Creek	0.870	0.663	0.682	0.984	0.973	0.813
7	Crowsnest River	0.945	0.918	0.958	0.983	0.972	0.978
8	Tuya River	0.915	0.835	0.879	0.981	0.973	0.945

Tableau-A II-2 Ratio de biais sans dimension β pour la calibration des deux scénarios : LOC et REG

N ⁰	Basin	LOC			REG		
		Training	Validation	Testing	Training	Validation	Testing
1	Lac-Saint-Jean	1.002	0.940	0.967	0.985	1.010	0.948
2	Grande River	0.971	1.002	0.925	0.983	1.000	0.991
3	Cascapedia River	0.954	0.993	1.008	0.976	0.996	0.956
4	Bear Brook	0.909	0.969	0.889	0.984	1.005	0.888
5	Hayes River	1.092	0.857	0.958	0.993	1.009	0.876
6	Swift Current Creek	1.060	0.962	0.643	0.995	1.010	0.702
7	Crowsnest River	0.956	0.977	1.070	0.981	1.000	0.972
8	Tuya River	1.104	0.950	0.985	0.976	0.997	0.829

Tableau-A II-3 Ratio de variabilité sans dimension γ pour la calibration des deux scénarios :
LOC et REG

N°	Basin	LOC			REG		
		Training	Validation	Testing	Training	Validation	Testing
1	Lac-Saint-Jean	0.996	0.996	1.029	0.985	0.988	1.036
2	Grande River	0.931	0.987	1.031	0.979	0.984	1.036
3	Cascapedia River	0.983	0.946	0.930	0.991	0.999	0.961
4	Bear Brook	0.953	1.084	1.027	0.986	0.997	0.974
5	Hayes River	0.794	0.888	0.942	0.972	0.985	0.898
6	Swift Current Creek	0.861	1.061	1.056	0.989	0.999	1.167
7	Crowsnest River	0.932	0.958	1.034	0.984	0.987	1.057
8	Tuya River	1.035	0.935	0.732	0.990	0.995	0.894

ANNEXE III

CRPS NORMALISÉ DES PRÉVISIONS D'ENSEMBLE DES SCÉNARIOS LOC ET REG

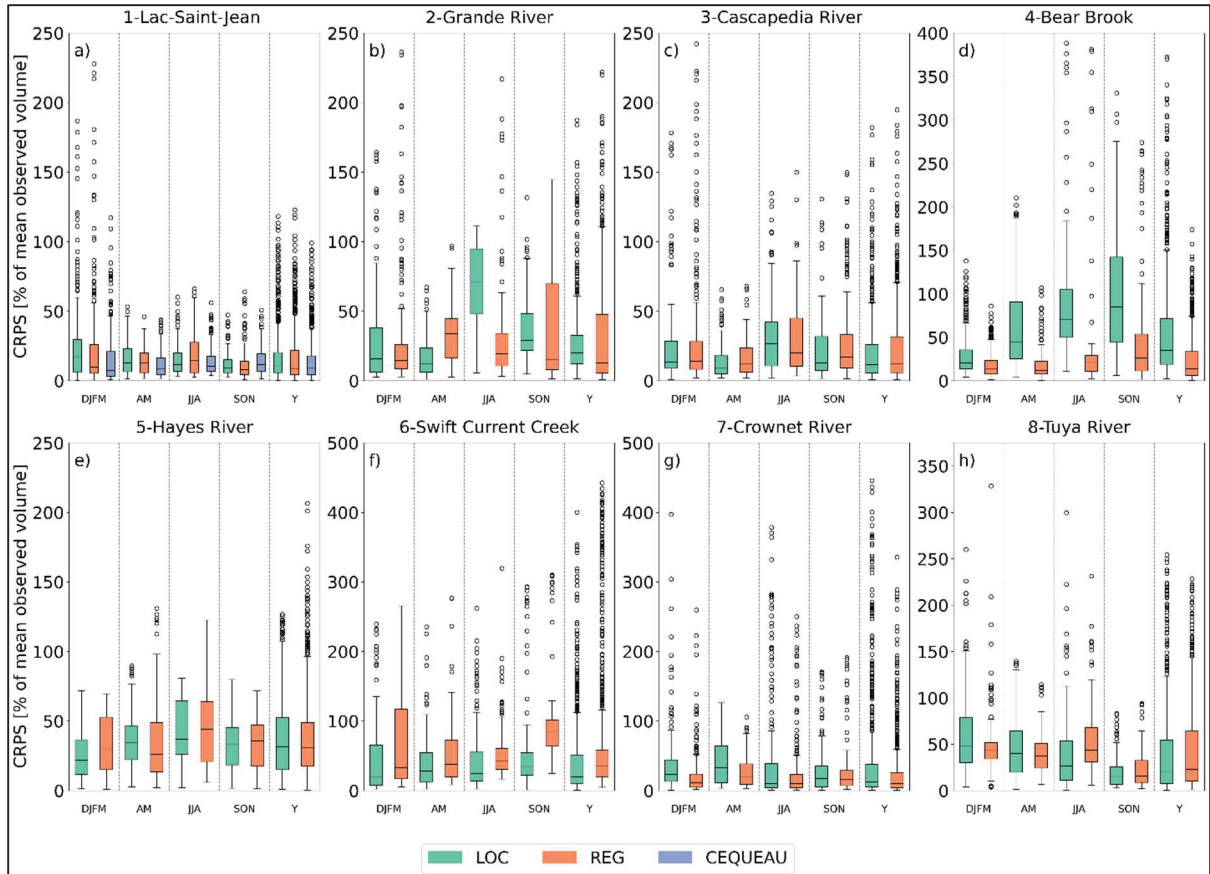


Figure-A III-1 CRPS normalisé des prévisions d'ensemble des volumes par saison pour deux scénarios principaux : LOC (modèle local) et REG (modèle régional). Pour le bassin du Lac-Saint-Jean uniquement, le modèle de référence CEQUEAU pour ce bassin est également inclus. Chaque panneau représente un bassin versant différent : a) Lac-Saint-Jean [1], b) Rivière Grande [2], c) Rivière Cascapedia [3], d) Bear Brook [4], e) Rivière Hayes [5], f) Swift Current Creek [6], g) Rivière Crowsnest [7], et h) Rivière Tuya [8]. Les valeurs saisonnières normalisées ont été calculées en divisant le CRPS de chaque saison par le volume saisonnier moyen du bassin correspondant. Les valeurs annuelles normalisées ont été calculées de la même manière, en utilisant le volume annuel moyen. La période de prévision couvre du 10 mars 2016 au 29 mai 2023

ANNEXE IV

CRPS DES PRÉVISIONS D'ENSEMBLE DES VOLUMES POUR LE BASSIN DU LAC-SAINT-JEAN POUR LES MODÈLES INDIVIDUELS ET LES COMBINAISONS MULTI-MODÈLES

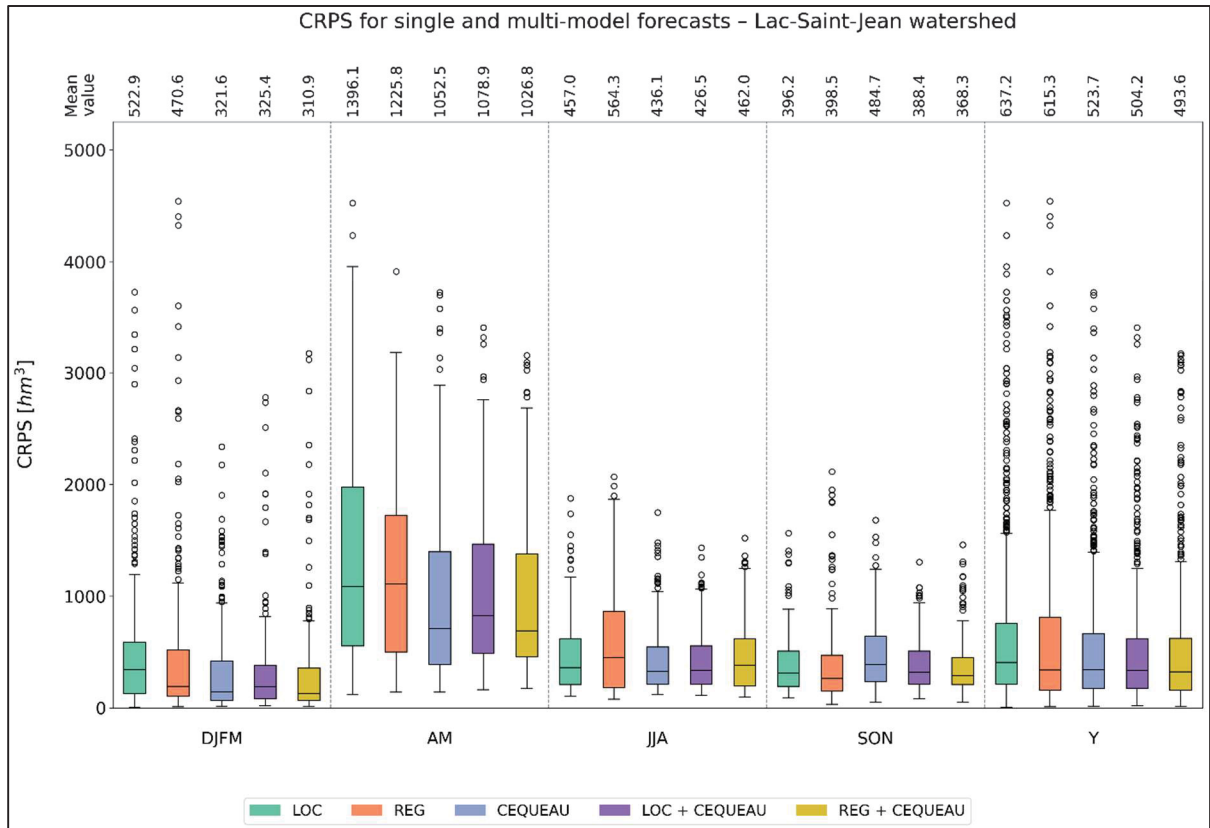


Figure-A IV-1 CRPS des prévisions d'ensemble des volumes pour le bassin du Lac-Saint-Jean, par saison, pour les modèles individuels (LOC, REG et CEQUEAU) et les combinaisons multi-modèles (LOC+CEQUEAU et REG+CEQUEAU). Les saisons sont définies comme suit : hiver (DJFM), printemps (AM), été (JJA), automne (SON) et annuel (Y). La période de prévision couvre du 10 mars 2016 au 29 mai 2023

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Aguilar Andrade, F. S., Arsenault, R., Poulin, A., Troin, M., & Armstrong, W. (2024). Application of weather post-processing methods for operational ensemble hydrological forecasting on multiple catchments in Canada. *Journal of Hydrology*, 642, 131861. <https://doi.org/10.1016/j.jhydrol.2024.131861>
- Ahmed, M. I., Elshorbagy, A., & Pietroniro, A. (2020). Toward simple modeling practices in the complex Canadian Prairie Watersheds. *Journal of Hydrologic Engineering*, 25(6), 04020024. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001922](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001922)
- Alfwzan, W. F., Selim, M. M., Althobaiti, S., & Hussin, A. M. (2023). Application of Bi-LSTM method for groundwater quality assessment through water quality indices. *Journal of Water Process Engineering*, 53, 103889. <https://doi.org/10.1016/j.jwpe.2023.103889>
- Alizadeh, B., Ghaderi Bafti, A., Kamangir, H., Zhang, Y., Wright, D. B., & Franz, K. J. (2021). A novel attention-based LSTM cell post-processor coupled with Bayesian optimization for streamflow prediction. *Journal of Hydrology*, 601, 126526. <https://doi.org/10.1016/j.jhydrol.2021.126526>
- Anghileri, D., Monhart, S., Zhou, C., Bogner, K., Castelletti, A., Burlando, P., & Zappa, M. (2019). The value of subseasonal hydrometeorological forecasts to hydropower operations: How much does preprocessing matter? *Water Resources Research*, 55(12), 10159–10178. <https://doi.org/10.1029/2019WR025280>
- Armstrong, W., Arsenault, R., Martel, J. L., Troin, M., Dion, P., Sabzipour, B., Mai, J. (2025). Improving multi-model ensemble streamflow forecasts by combining lumped, distributed and deep learning hydrological models. *Hydrological Sciences Journal*, 70(7), 1091–1108. <https://doi.org/10.1080/02626667.2025.2471430>
- Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., & Pappenberger, F. (2018). Skilful seasonal forecasts of streamflow over Europe? *Hydrology and Earth System Sciences*, 22(4), 2057–2072. <https://doi.org/10.5194/hess-22-2057-2018>
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment: Part I. Model development. *Journal of the American Water Resources Association*, 34(1), 73–89. <https://doi.org/10.1111/j.1752-1688.1998.tb05961.x>

- Arsenault, R., Poissant, D., & Brissette, F. (2015). Parameter dimensionality reduction of a conceptual model for streamflow prediction in Canadian, snowmelt dominated ungauged basins. *Advances in Water Resources*, 85, 27–44. <https://doi.org/10.1016/j.advwatres.2015.08.014>
- Arsenault, R., & Brissette, F. (2016). Analysis of continuous streamflow regionalization methods within a virtual setting. *Hydrological Sciences Journal*, 61(15), 2680–2693. <https://doi.org/10.1080/02626667.2016.1154557>
- 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. <https://doi.org/10.1007/s11269-016-1425-4>
- 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, 2735–2750. <https://doi.org/10.5194/hess-23-2735-2019>
- Arsenault, R., Brissette, F., Martel, J.-L., Troin, M., Lévesque, G., Davidson-Chaput, J., ... Poulin, A. (2020). A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. *Scientific Data*, 7, 243. <https://doi.org/10.1038/s41597-020-00583-2>
- Arsenault, R., Martel, J.-L., Brunet, F., Brissette, F., & Mai, J. (2023a). Continuous streamflow prediction in ungauged basins: Long short-term memory neural networks clearly outperform traditional hydrological models. *Hydrology and Earth System Sciences*, 27, 139–157. <https://doi.org/10.5194/hess-27-139-2023>
- Arsenault, R., Huard, D., Martel, J.-L., Troin, M., Mai, J., Brissette, F., Jauvin, C., Vu, L., Craig, J. R., Smith, T. J., & Logan, T. (2023b). The PAVICS-Hydro platform: A virtual laboratory for hydroclimatic modelling and forecasting over North America. *Environmental Modelling & Software*, 168, 105808. <https://doi.org/10.1016/j.envsoft.2023.105808>
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000). Artificial Neural Networks in Hydrology. I: Preliminary Concepts. *Journal of Hydrologic Engineering*, 5(2), 115–123. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:2\(115\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115))
- Balthazar, L. D., Miranda, F., Cândido, V. B. R., Capriles, P., Moraes, M., Ribeiro, C. B. M., Fayer, G., & Goliatt, L. (2024). Long-term natural streamflow forecasting under drought scenarios using data-intelligence modeling. *Water Cycle*, 5, 266–277. <https://doi.org/10.1016/j.watcyc.2024.07.001>

- Becker, B., Kim, J., & Pummer, E. (2023). Reservoir operations under uncertainty with moving-horizon approach and ensemble forecast optimization. *Journal of Applied Water Engineering and Research*, 12(3), 265–276. <https://doi.org/10.1080/23249676.2023.2276948>
- Bennett, J. C., Wang, Q. J., Li, M., Robertson, D. E., & Schepen, A. (2016). Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model. *Water Resources Research*, 52(10), 8238–8259. <https://doi.org/10.1002/2016WR019193>
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166. <https://doi.org/10.1109/72.279181>
- Bengtsson, L., Kanamitsu, M., Kållberg, P. W., & Uppala, S. M. (1982). FGGE 4-dimensional data assimilation at ECMWF. *Bulletin of the American Meteorological Society*, 63(1), 29–43.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281–305.
- Bergström, S., & Forsman, A. (1973). Development of a conceptual deterministic rainfall runoff model. *Nordic Hydrology*, 4(3), 147–170. Repéré à <https://iwaponline.com/hr/article/4/3/147/1357/DEVELOPMENT-OF-A-CONCEPTUAL-DETERMINISTIC-RAINFALL>
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, 24(1), 43–69. <https://doi.org/10.1080/02626667909491834>
- Beven, K. J., & Young, P. C. (2013). A guide to good practice in modeling semantics for authors and referees. *Water Resources Research*, 49(8), 5092–5098. <https://doi.org/10.1002/wrcr.20393>
- Bjorck, J., Weinberger, K. Q., & Gomes, C. P. (2021). Understanding decoupled and early weight decay. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), 6777–6785. <https://doi.org/10.1609/aaai.v35i8.16837>
- Brandt, W. T., Haleakala, K., Hatchett, B. J., & Pan, M. (2022). A review of the hydrologic response mechanisms during mountain rain-on-snow. *Frontiers in Earth Science*, 10, 791760. <https://doi.org/10.3389/feart.2022.791760>

- Cantoni, E., Trambly, Y., Grimaldi, S., Salamon, P., Dakhlaoui, H., Dezetter, A., & Thiemig, V. (2022). Hydrological performance of the ERA5 reanalysis for flood modeling in Tunisia with the LISFLOOD and GR4J models. *Journal of Hydrology: Regional Studies*, 42, 101169. <https://doi.org/10.1016/j.ejrh.2022.101169>
- Caterini, A. L., & Chang, D. E. (2018). Recurrent neural networks. In *Deep neural networks in a mathematical framework* (pp. 59-79). Cham: Springer International Publishing.
- Charney, J. & Fjortoft, R. & Neumann, J. (1950). Numerical integration of the barotropic vorticity equation. *Tellus A: Dynamic Meteorology and Oceanography*, 2(4), Repéré à https://www.eas.ualberta.ca/jdwilson/EAS372_13/Charney_etal50.pdf
- Chen, J., & Brissette, F. P. (2017). Hydrological modelling using proxies for gauged precipitation and temperature. *Hydrological Processes*, 31(22), 3881–3897. <https://doi.org/10.1002/hyp.11304>
- Chen, X., Zhang, Y., Ye, A., Li, J., Hsu, K., & Sorooshian, S. (2025). Fine-tuning long short-term memory models for seamless transition in hydrological modelling: From pre-training to post-application. *Environmental Modelling & Software*, 186, 106350. <https://doi.org/10.1016/j.envsoft.2025.106350>
- 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. <https://doi.org/10.1016/j.jhydrol.2020.125376>
- Cho, K., & Kim, Y. (2022). Improving streamflow prediction in the WRF-Hydro model with LSTM networks. *Journal of Hydrology*, 605, 127297. <https://doi.org/10.1016/j.jhydrol.2021.127297>
- Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, 375(3–4), 613–626. <https://doi.org/10.1016/j.jhydrol.2009.06.005>
- Coulibaly, P., & Anctil, F. (1999). Real-time short-term natural water inflows forecasting using recurrent neural networks. In *Proceedings of the 1999 International Joint Conference on Neural Networks (IJCNN'99)* (pp. 3802–3805). IEEE. <https://doi.org/10.1109/IJCNN.1999.830759>
- Coumou, D., & Rahmstorf, S. (2012). A decade of weather extremes. *Nature Climate Change*, 2(7), 491–496. <https://doi.org/10.1038/nclimate1452>
- Cui, Z., Zhou, Y., Guo, S., Wang, J., & Xu, C.-Y. (2022). Effective improvement of multi-step-ahead flood forecasting accuracy through encoder–decoder with an exogenous input structure. *Journal of Hydrology*, 609, 127764. <https://doi.org/10.1016/j.jhydrol.2022.127764>

- Daniel, T. M. (1991). Neural networks: Applications in hydrology and water resources engineering. In *Proceedings of the International Hydrology and Water Resources Symposium*. Institution of Engineers, Australia.
- 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*, 658, 133186. <https://doi.org/10.1016/j.jhydrol.2025.133186>
- Dawson, C. W., & Wilby, R. (1998). An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal*, 43(1), 47–66. <https://doi.org/10.1080/02626669809492102>
- Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management*, 111(2), 157–170. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1985\)111:2\(157\)](https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157))
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>
- Deng, J., Couasnon, A., Dahm, R., Hrachowitz, M., van Heeringen, K.-J., Korving, H., Weerts, A., & Taormina, R. (2024a). Operational low-flow forecasting using LSTMs. *Frontiers in Water*, 5, 1332678. <https://doi.org/10.3389/frwa.2023.1332678>
- Deng, X. et al. (2024b). Global Ensemble Prediction System (GEPS): Update from version 7.1.0 to version 8.0.0 (Canadian Meteorological Centre Technical Note). Government of Canada. https://collaboration.cmc.ec.gc.ca/cmc/cmoe/product_guide/docs/tech_notes/technote_geps-800_e.pdf
- Dion, P., Martel, J.-L., & Arsenault, R. (2021). Hydrological ensemble forecasting using a multi-model framework. *Journal of Hydrology*, 600, 126537. <https://doi.org/10.1016/j.jhydrol.2021.126537>
- Do, H. X., Westra, S., Leonard, M., & Gudmundsson, L. (2020). Global-scale prediction of flood timing using atmospheric reanalysis. *Water Resources Research*, 56(1), e2019WR024945. <https://doi.org/10.1029/2019WR024945>
- Duethmann, D., Blöschl, G., & Parajka, J. (2020). Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? *Hydrology and Earth System Sciences*, 24(7), 3493–3511. <https://doi.org/10.5194/hess-24-3493-2020>

- Essou, G. R. C., Brissette, F., & Lucas-Picher, P. (2017). The use of reanalyses and gridded observations as weather input data for a hydrological model: Comparison of performances of simulated river flows based on the density of weather stations. *Journal of Hydrometeorology*, 18(2), 497–513. <https://doi.org/10.1175/JHM-D-16-0088.1>
- Evin, G., Le Lay, M., Fouchier, C., Penot, D., Colleoni, F., Mas, A., Garambois, P.-A., & Laurantin, O. (2024). Evaluation of hydrological models on small mountainous catchments: Impact of the meteorological forcings. *Hydrology and Earth System Sciences*, 28, 261–281. <https://doi.org/10.5194/hess-28-261-2024>
- Fang, G. H., Yang, J., Chen, Y. N., & Zammit, C. (2015). Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. *Hydrology and Earth System Sciences*, 19(6), 2547–2559. <https://doi.org/10.5194/hess-19-2547-2015>
- Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*, 56(9), e2019WR026793. <https://doi.org/10.1029/2019WR026793>
- Fortin, V., & Turcotte, R. (2007). *Le modèle hydrologique MOHYSE, bases théoriques et manuel de l'utilisateur*. Notes de cours pour SCA7420, Département des sciences de la Terre et de l'atmosphère, Université du Québec à Montréal, Montréal, 17p.
- Gibson, J.K., Kållberg, P., Uppala, S.M., Hernandez, A., Nomura, A. and Serrano, E. (1999). ECMWF re-analysis project report 1, ERA-15 description (version 2). Technical Report, ECMWF, Reading, UK.
- Govindaraju, R. S., & Rao, A. R. (Eds.). (2000). *Artificial neural networks in hydrology*. Kluwer Academic Publishers.
- Graham, R. M., Browell, J., Bertram, D., & White, C. J. (2022). The application of sub-seasonal to seasonal (S2S) predictions for hydropower forecasting. *Meteorological Applications*, 29(2), e2047. <https://doi.org/10.1002/met.2047>
- Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5–6), 602–610. <https://doi.org/10.1016/j.neunet.2005.06.042>
- Grossi, G., Lendvai, A., Peretti, G., & Ranzi, R. (2017). Snow precipitation measured by gauges: Systematic error estimation and data series correction in the Central Italian Alps. *Water*, 9(7), 461. <https://doi.org/10.3390/w9070461>

- Gualtieri, G. (2022). Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review. *Renewable and Sustainable Energy Reviews*, 167, 112741. <https://doi.org/10.1016/j.rser.2022.112741>
- Güntner, A., & Bronstert, A. (2004). Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas. *Journal of Hydrology*, 297(1–4), 136–161. <https://doi.org/10.1016/j.jhydrol.2004.04.008>
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Haiden, T., Janousek, M., Vitart, F., Ben Bouallègue, Z., & Prates, F. (2023). Evaluation of ECMWF forecasts, including the 2023 upgrade (Technical Memorandum No. 911). European Centre for Medium-Range Weather Forecasts. <https://doi.org/10.21957/d47ba5263c>
- Haiden, T., Janousek, M., Vitart, F., Tanguy, M., Prates, F., & Chevalier, M. (2024). Evaluation of ECMWF forecasts (ECMWF Technical Memoranda No. 918). European Centre for Medium-Range Weather Forecasts. <https://doi.org/10.21957/52f2f31351>
- He, M., Xu, X., Wu, S., Kang, C., & Huang, B. (2025). Multi-step ahead forecasting of daily streamflow based on the transform based deep learning model under different scenarios. *Scientific Reports*, 15, 5451. <https://doi.org/10.1038/s41598-025-89837-w>
- 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. <https://doi.org/10.1016/j.jhydrol.2021.126676>
- Hersbach, Hans. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, 15(5), 559–570. [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)
- Hersbach, H. et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hsu, K., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall–runoff process. *Water Resources Research*, 31(10), 2517–2530. <https://doi.org/10.1029/95WR01955>

- Jiang, S., Ren, L., Yang, X., Ma, M., & Liu, Y. (2014). Multi-model ensemble hydrologic prediction and uncertainties analysis. In *Evolving Water Resources Systems: Understanding, Predicting and Managing Water–Society Interactions*, Proceedings of ICWRS2014, Italy, June 2014. IAHS Publication
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., ... Monge-Sanz, B. M. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3), 1087–1117. <https://doi.org/10.5194/gmd-12-1087-2019>
- Kang, K. W., Kim, J. H., Park, C. Y., & Ham, K. J. (1993). Evaluation of hydrological forecasting system based on neural network model. In *Proceedings of the 25th Congress of the International Association for Hydraulic Research* (pp. 257–264). International Association for Hydraulic Research.
- Kao, I-Feng & Zhou, Yanlai & Chang, Li-Chiu & Chang, Fi-John. (2020). Exploring a Long Short-Term Memory based Encoder-Decoder Framework for Multi-Step-Ahead Flood Forecasting. *Journal of Hydrology*, 583, 124631. [10.1016/j.jhydrol.2020.124631](https://doi.org/10.1016/j.jhydrol.2020.124631).
- Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424-425, 264-277. <https://doi.org/10.1016/j.jhydrol.2012.01.011>
- Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10), 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>
- Koch, J., & Schneider, R. (2022). Long short-term memory networks enhance rainfall-runoff modelling at the national scale of Denmark. *GEUS Bulletin*, 49. <https://doi.org/10.34194/geusb.v49.8292>
- Konapala, G., Kao, S.-C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, 15(10), 104022. <https://doi.org/10.1088/1748-9326/aba927>
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). 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. <https://doi.org/10.5194/hess-23-5089-2019>

- Kratzert, F., Gauch, M., Klotz, D., & Nearing, G. (2024). HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin. *Hydrology and Earth System Sciences*, 28(17), 4187–4201. <https://doi.org/10.5194/hess-28-4187-2024>
- Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, 249(1–4), 2–9. [https://doi.org/10.1016/S0022-1694\(01\)00420-6](https://doi.org/10.1016/S0022-1694(01)00420-6)
- Kundzewicz, Z. W., Budhakooncharoen, S., Bronstert, A., Hoff, H., Lettenmaier, D., Menzel, L., & Schulze, R. (2003). Coping with variability and change: Floods and droughts. *Natural Resources Forum*, 27(4), 263–274. <https://doi.org/10.1111/1477-8947.00029>
- Kumar, D. N., Raju, K. S., & Sathish, T. (2004). River flow forecasting using recurrent neural networks. *Water Resources Management*, 18, 143–161. <https://doi.org/10.1023/B:WARM.0000024727.94701.12>
- Lamontagne, J. R., Barber, C. A., & Vogel, R. M. (2020). Improved estimators of model performance efficiency for skewed hydrologic data. *Water Resources Research*, 56(9), e2020WR027101. <https://doi.org/10.1029/2020WR027101>
- Latifovic, R., Homer, C., Ressler, R., Pouliot, D. A., Hossian, S., Colditz, R., Olthof, I., Giri, C., & Victoria, A. (2012). North American Land Change Monitoring System. In C. Giri (Ed.), *Remote sensing of land use and land cover: Principles and applications* (1st ed., pp. 303–324). CRC Press. <https://doi.org/10.1201/b11964-24>
- Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., & Di, Z. (2017). A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *WIREs Water*, 4(6), e1246. <https://doi.org/10.1002/wat2.1246>
- Li, X., Khandelwal, A., Jia, X., Cutler, K., Ghosh, R., Renganathan, A., ... Kumar, V. (2022). Regionalization in a global hydrologic deep learning model: From physical descriptors to random vectors. *Water Resources Research*, 58(8), e2021WR031794. <https://doi.org/10.1029/2021WR031794>
- Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, ... Restrepo, P. (2012). Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, 16(10), 3863–3887. <https://doi.org/10.5194/hess-16-3863-2012>
- Liu, S., Lu, D., Painter, S. L., Griffiths, N. A., & Pierce, E. M. (2023). Uncertainty quantification of machine learning models to improve streamflow prediction under changing climate and environmental conditions. *Frontiers in Water*, 5, 1150126. <https://doi.org/10.3389/frwa.2023.1150126>

- Liu, Z., Wang, Y., Xu, Z., & Duan, Q. (2017). Conceptual hydrological models. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. Cloke, & J. Schaake (Eds.), *Handbook of hydrometeorological ensemble forecasting*. Springer. https://doi.org/10.1007/978-3-642-40457-3_22-1
- Magnusson, J., Nævdal, G., Matt, F., Burkhart, J. F., & Winstral, A. (2020). Improving hydropower inflow forecasts by assimilating snow data. *Hydrology Research*, 51(2), 226–237. <https://doi.org/10.2166/nh.2020.025>
- Mai, J., Shen, H., Tolson, B. A., Gaborit, É., Arsenault, R., Craig, J. R., ... Waddell, J. W. (2022). The Great Lakes Runoff Intercomparison Project Phase 4: The Great Lakes (GRIP-GL). *Hydrology and Earth System Sciences*, 26(13), 3537–3572. <https://doi.org/10.5194/hess-26-3537-2022>
- Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water-resources variables: A review of modelling issues and applications. *Environmental Modelling & Software*, 15(1), 101–124. [https://doi.org/10.1016/S1364-8152\(99\)00007-9](https://doi.org/10.1016/S1364-8152(99)00007-9)
- Martel, J.-L., Arsenault, R., Turcotte, R., Castañeda-Gonzalez, M., Brissette, F., Armstrong, ... Caron, L.-P. (2024). Exploring the ability of LSTM-based hydrological models to simulate streamflow time series for flood frequency analysis. *EGUsphere*. <https://doi.org/10.5194/egusphere-2024-2134>
- Martel, J.-L., Brissette, F., Arsenault, R., Turcotte, R., Castañeda-Gonzalez, M., Armstrong, ... Caron, L.-P. (2025). Assessing the adequacy of traditional hydrological models for climate change impact studies: A case for long short-term memory (LSTM) neural networks. *Hydrology and Earth System Sciences*, 29(13), 2811–2836. <https://doi.org/10.5194/hess-29-2811-2025>
- Martinec, J. (1975). Snowmelt-runoff model for stream flow forecasts. *Hydrology Research*, 6(3), 145–154. <https://doi.org/10.2166/nh.1975.0010>
- Matthews, G., Barnard, C., Cloke, H., Dance, S. L., Jurlina, T., Mazzetti, C., & Prudhomme, C. (2022). Evaluating the impact of post-processing medium-range ensemble streamflow forecasts from the European Flood Awareness System. *Hydrology and Earth System Sciences*, 26(11), 2939–2968. <https://doi.org/10.5194/hess-26-2939-2022>
- Mekonnen, B. A., Nazemi, A., Mazurek, K. A., Elshorbagy, A., & Putz, G. (2015). Hybrid modelling approach to prairie hydrology: fusing data-driven and process-based hydrological models. *Hydrological Sciences Journal*, 60(9), 1473–1489. <https://doi.org/10.1080/02626667.2014.935778>

- Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., ... Hu, Z.-Z. (2013). The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization. *Monthly Weather Review*, 141(8), 2910–2945. <https://doi.org/10.1175/MWR-D-12-00216.1>
- Metcalfe, R. A., & Buttle, J. M. (2001). Soil water dynamics in a forested boreal catchment, central Canada. *Hydrological Processes*, 15(1), 65–77. <https://doi.org/10.1002/hyp.1120>
- Mic, R., Galéa, G., & Javelle, P. (2002). Modélisation régionale des débits de crue du bassin hydrographique du Cris : approche régionale classique et par modèles de référence. *Revue des sciences de l'eau / Journal of Water Science*, 15(3), 677–700. <https://doi.org/10.7202/705475ar>
- Minns, A. W., & Hall, M. J. (1996). Artificial neural networks as rainfall–runoff models. *Hydrological Sciences Journal*, 41(3), 399–417. <https://doi.org/10.1080/02626669609491511>
- Mirzaei, M., Yu, H., Dehghani, A., Galavi, H., Shokri, V., Mohsenzadeh Karimi, S., & Sookhak, M. (2021). A Novel Stacked Long Short-Term Memory Approach of Deep Learning for Streamflow Simulation. *Sustainability*, 13(23), 13384. <https://doi.org/10.3390/su132313384>
- Moore, R. D., Trubilowicz, J. W., & Buttle, J. M. (2011). Prediction of streamflow regime and annual runoff for ungauged basins using a distributed monthly water balance model. *Journal of the American Water Resources Association*, 47(5), 958–971. <https://doi.org/10.1111/j.1752-1688.2011.00595.x>
- Morin, G., & Paquet, P. (2007). Modèle hydrologique CEQUEAU (Rapport de recherche No. R000926). INRS, Centre Eau, Terre et Environnement.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., ... Thépaut, J.-N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9), 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Nagesh Kumar, D., Srinivasa Raju, K. & Sathish, T. River Flow Forecasting using Recurrent Neural Networks. *Water Resources Management*, 18, 143–161 (2004). <https://doi.org/10.1023/B:WARM.0000024727.94701.12>
- Ougahi, J. H., & Rowan, J. S. (2025). Enhanced streamflow forecasting using hybrid modelling integrating glacio-hydrological outputs, deep learning and wavelet transformation. *Scientific Reports*, 15, 2762. <https://doi.org/10.1038/s41598-025-87187-1>

- Ozdemir, S., & Ozkan Yildirim, S. (2023). Prediction of Water Level in Lakes by RNN-Based Deep Learning Algorithms to Preserve Sustainability in Changing Climate and Relationship to Microcystin. *Sustainability*, 15(22), 16008. <https://doi.org/10.3390/su152216008>
- 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. <https://doi.org/10.5194/hess-24-6059-2020>
- Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1), 275–289. doi: [https://doi.org/10.1016/S0022-1694\(03\)00225-7](https://doi.org/10.1016/S0022-1694(03)00225-7).
- Prechelt, L. (2012). Early stopping — but when? In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (2nd ed., vol. 7700, pp. 53–67). https://doi.org/10.1007/978-3-642-35289-8_5
- Probst, E., & Mauser, W. (2022). Evaluation of ERA5 and WFDE5 forcing data for hydrological modelling and the impact of bias correction with regional climatologies: A case study in the Danube River Basin. *Journal of Hydrology: Regional Studies*, 40, 101023. <https://doi.org/10.1016/j.ejrh.2022.101023>
- Quang, N. H., Van An, N., & Viet, T. Q. (2025). From local to regional: Deep learning models for daily water discharge forecasting in a data-scarce basin and engineered river. *Water Resources Management*, 39, 6539–6580. <https://doi.org/10.1007/s11269-025-04261-z>
- Raman, H., & Sunilkumar, N. (1995). Multivariate modelling of water resources time series using artificial neural networks. *Hydrological Sciences Journal*, 40(2), 145–163. <https://doi.org/10.1080/02626669509491401>
- Razavi, T., & Coulibaly, P. (2013). Streamflow prediction in ungauged basins: Review of regionalization methods. *Journal of Hydrologic Engineering*, 18(8), 958–975. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000690](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690)
- Ressources naturelles Canada. (2024). L'Énergie hydroélectrique. Repéré à <https://natural-resources.canada.ca/energy-sources/renewable-energy/hydroelectric-energy>
- Richardson, L. F. (1922). *Weather prediction by numerical process*. Cambridge University Press.
- Robinson, N., Regetz, J., & Guralnick, R. P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90 m digital elevation model from fused ASTER and SRTM data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 57–67. <https://doi.org/10.1016/j.isprsjprs.2013.11.002>

- 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, W10544. <https://doi.org/10.1029/2010WR009502>
- 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(Part A), 130380. <https://doi.org/10.1016/j.jhydrol.2023.130380>
- Samuel, J., Rousseau, A. N., Abbasnezhadi, K., & Savary, S. (2019). Development and evaluation of a hydrologic data-assimilation scheme for short-range flow and inflow forecasts in a data-sparse high-latitude region using a distributed model and ensemble Kalman filtering. *Advances in Water Resources*, 130, 198–220. <https://doi.org/10.1016/j.advwatres.2019.06.004>
- Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, 54(10), 8558–8593. <https://doi.org/10.1029/2018WR022643>
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems* (Vol. 28, pp. 802–810). MIT Press.
- Singh, V. P., & Woolhiser, D. A. (2002). Mathematical modeling of watershed hydrology. *Journal of Hydrologic Engineering*, 7(4), 270–292. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2002\)7:4\(270\)](https://doi.org/10.1061/(ASCE)1084-0699(2002)7:4(270))
- Solanki, H., Vegad, U., Kushwaha, A., & Mishra, V. (2025). Improving streamflow prediction using multiple hydrological models and machine learning methods. *Water Resources Research*, 61(1), e2024WR038192. <https://doi.org/10.1029/2024WR038192>
- Soucy, L. (2025). *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* (Mémoire de maîtrise, École de technologie supérieure, Université du Québec, Montréal).
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958. <http://jmlr.org/papers/v15/srivastava14a.html>
- Statistique Canada. (2024). Approvisionnement et écoulement de l'électricité, 2023 (données définitives). [Communiqué]. Le Quotidien. Repéré à <https://www150.statcan.gc.ca/n1/daily-quotidien/241022/dq241022e-fra.htm>

- Swift-LaPointe, T., White, R. H., & Radić, V. (2025). A hybrid statistical-dynamical forecast of seasonal streamflow for a catchment in the Upper Columbia River basin in Canada. *Frontiers in Water*, 7, Article 1595898. <https://doi.org/10.3389/frwa.2025.1595898>
- Talagrand, O., & Vautard, R. (1997). Evaluation of probabilistic prediction systems. Communication présentée au Workshop on Predictability, Paris. Repéré à : <https://www.ecmwf.int/en/elibrary/76596-evaluation-probabilistic-prediction-systems>
- Talbot, F., Sylvain, J.-D., Drolet, G., Poulin, A., Martel, J.-L., & Arsenault, R. (2025). Assessing hydroclimatic impacts of climate change in snowy catchments using a physically based hydrological model. *Journal of Hydrology: Regional Studies*, 59, 102453. <https://doi.org/10.1016/j.ejrh.2025.102453>
- Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, 24(5), 2527–2544. <https://doi.org/10.5194/hess-24-2527-2020>
- Tokar, A. S., & Markus, M. (2000). Precipitation–runoff modeling using artificial neural networks and conceptual models. *Journal of Hydrologic Engineering*, 5(2), 156–161. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:2\(156\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(156))
- Tripathy, K. P., & Mishra, A. K. (2024). Deep learning in hydrology and water resources disciplines: Concepts, methods, applications, and research directions. *Journal of Hydrology*, 628, 130458. <https://doi.org/10.1016/j.jhydrol.2023.130458>
- Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J.-L. (2021). Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years. *Water Resources Research*, 57, e2020WR028392. <https://doi.org/10.1029/2020WR028392>
- Turner, S. W. D., Xu, W., & Voisin, N. (2020). Inferred inflow forecast horizons guiding reservoir release decisions across the United States. *Hydrology and Earth System Sciences*, 24, 1275–1291. <https://doi.org/10.5194/hess-24-1275-2020>
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Da Costa Bechtold, V., Fiorino, ... Woollen, J. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612), 2961–3012. <https://doi.org/10.1256/qj.04.176>
- Valéry, A., 2010. *Modélisation précipitations débit sous influence nivale: Elaboration d'un module neige et évaluation sur 380 bassins versants* (Doctoral dissertation, AgroParisTech).

- Viterbo, F., Read, L., Nowak, K., Wood, A. W., Gochis, D., Cifelli, R., & Hughes, M. (2020). General Assessment of the Operational Utility of National Water Model Reservoir Inflows for the Bureau of Reclamation Facilities. *Water*, 12(10), 2897. <https://doi.org/10.3390/w12102897>
- Wang, Z., Si, Y., & Chu, H. (2022). Daily streamflow prediction and uncertainty using a Long Short Term Memory (LSTM) network coupled with bootstrap. *Water Resources Management*, 36, 4491–4507. <https://doi.org/10.1007/s11269-022-03264-4>
- Waqas, M., & Humphries, U. W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*, 13, 102946. <https://doi.org/10.1016/j.mex.2024.102946>
- World Meteorological Organization & World Bank. (2015). *Valuing weather and climate: Economic assessment of meteorological and hydrological services* (WMO-No. 1153). <https://documents1.worldbank.org/curated/en/711881495514241685/pdf/Valuing-weather-and-climate-economic-assessment-of-meteorological-and-hydrological-services.pdf>
- Xu, J., Anctil, F., & Boucher, M.-A. (2022). Exploring hydrologic post-processing of ensemble streamflow forecasts based on affine kernel dressing and non-dominated sorting genetic algorithm II. *Hydrology and Earth System Sciences*, 26(4), 1001–1017. <https://doi.org/10.5194/hess-26-1001-2022>
- Yang, T., Akbari Asanjan, A., Welles, E., Gao, X., Sorooshian, S., & Liu, X. (2017). Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water Resources Research*, 53(4), 2786–2812. <https://doi.org/10.1002/2017WR020482>
- Yang, S., Yang, D., Chen, J., & Zhao, B. (2019). Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. *Journal of Hydrology*, 579, 124229. <https://doi.org/10.1016/j.jhydrol.2019.124229>
- Yin, X.-A., Yang, Z.-F., & Petts, G. E. (2011). Reservoir operating rules to sustain environmental flows in regulated rivers. *Water Resources Research*, 47(8), W08509. <https://doi.org/10.1029/2010WR009991>
- Yoshimura, K., Kanamitsu, M., Noone, D., & Oki, T. (2008). Historical isotope simulation using Reanalysis atmospheric data. *Journal of Geophysical Research: Atmospheres*, 113(D19). <https://doi.org/10.1029/2008JD010074>
- Zarei, M., Najarchi, M., & Mastouri, R. (2021). Bias correction of global ensemble precipitation forecasts by Random Forest method. *Earth Science Informatics*, 14(3), 677–689. <https://doi.org/10.1007/s12145-021-00577-7>

- Zhang, X., Wang, X., Li, H., Jian, J., & Liu, H. (2023). Monthly runoff prediction based on a coupled VMD-SSA-Bi-LSTM model. *Scientific Reports*, 13, Article 13149. <https://doi.org/10.1038/s41598-023-39606-4>
- Zhang, Y., & Ye, A. (2021). Machine learning for precipitation forecasts postprocessing: Multimodel comparison and experimental investigation. *Journal of Hydrometeorology*, 22(11), 3065–3085. <https://doi.org/10.1175/JHM-D-21-0096.1>

