

Optimisation de la prévision hydrologique d'ensemble par
l'intégration de l'apprentissage profond dans une méthode
multi-modèle hybride

par

William ARMSTRONG

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

MONTRÉAL, LE 12 FÉVRIER 2024

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



William Armstrong, 2024



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

Mme Annie Poulin, présidente du jury
Département de génie de la construction à l'École de technologie supérieure

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

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

LE 9 FÉVRIER 2024

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je veux premièrement remercier mon directeur de maîtrise, Richard Arsenault, le professeur qui a changé mon parcours scolaire grâce à son discours de 10 min, durant le cours CTN336, sur le processus pour réaliser des études supérieures. Sans ce bref moment, je n'aurais pas accompli un stage en recherche, été auxiliaire de cours, gagné la bourse du CRSNG (merci au Conseil de recherches en sciences naturelles et en génie du Canada), rédigé un mémoire et je ne serais pas rendu là, maintenant, futur ingénieur en expertise hydrologique chez Hydro-Québec. Avoir un professeur qui répond à tes questions passé 23h00 est un atout exceptionnel qui ajoute une dynamique particulière dans le parcours d'une maîtrise. Je te le dis encore, car je te l'ai dit souvent, merci Richard pour ton temps, ta patience et ton écoute tout au long de mon parcours.

Je veux ensuite remercier l'ensemble des co-auteurs de mon article, sans vous il ne serait pas aussi développé et aussi bien écrit. Vous m'avez aidé énormément, à l'aide de vos commentaires, à débloquer ma rédaction et à m'amener sur de nouvelles pistes d'idées.

Finalement, je voulais remercier mes amis et ma famille qui m'ont supporté au cours de mes études. Je sais, je n'étais pas « obligé » de faire une maîtrise pour travailler comme ingénieur, mais c'est mon choix et aujourd'hui ça me permet d'accéder à un poste chez Hydro-Québec. Une chance que vous étiez là pour me répéter « Et puis, tu finis quand ta maîtrise? » ou encore « Ça avance bien? Tu es rendu où là? » encore et encore. Évidemment, la réponse à ces questions était toujours « Ça va bien, je devrais finir bientôt. » quand la vraie réponse était « Je n'ai aucune idée! Lâchez-moi avec ça! » quand je sais très bien que sans ses petits commentaires je n'aurais probablement pas avancé aussi vite.

Optimisation de la prévision hydrologique d'ensemble par l'intégration de l'apprentissage profond dans une méthode multi-modèle hybride

William ARMSTRONG

RÉSUMÉ

L'application récente de l'apprentissage profond dans le domaine de l'hydrologie a démontré un potentiel prometteur. Cette recherche se concentre sur le développement d'une approche hybride multi-modèle, mettant en œuvre huit modèles hydrologiques globaux et conceptuels, un modèle semi-distribué, et un modèle d'apprentissage profond (Deep Learning, DL), pour la prévision des apports à un réservoir hydroélectrique. L'introduction du modèle à mémoire long terme et court terme (Long Short-Term Memory, LSTM) au sein de ce cadre multi-modèle est particulièrement mise de l'avant pour son potentiel à améliorer la précision des prévisions de débit à court terme sur le bassin versant du Lac-Saint-Jean. Les résultats montrent que l'association des modèles hydrologiques avec le LSTM améliore notablement la performance de prévision, en particulier pour les prévisions à court terme jusqu'à 9 jours, surpassant d'autres combinaisons multi-modèles en diminuant, en moyenne, jusqu'à 45% les valeurs du CRPS. L'étude souligne également que l'apprentissage profond, lorsqu'il est combiné avec d'autres approches de modélisation, permet d'atteindre les performances maximales notamment au niveau de la distribution des prévisions où les valeurs ABDU (Average Bin Distance to Uniformity, valeur mesurant la fiabilité des prévisions d'ensemble) ont diminué de 64%. La méthode hybride multi-modèle se révèle efficace pour capturer la variabilité hydrologique à l'échelle du bassin versant.

En complément, une section du mémoire porte sur l'opérationnalisation de ces outils et présente l'application de cette méthode dans un contexte pratique, notamment pour améliorer les systèmes de prévision hydrologique chez Rio Tinto. Utilisant le modèle hydrologique CEQUEAU7, déjà en place chez Rio Tinto, l'étude propose une méthode d'intégration simple, consistant à ajouter et fusionner des modèles hydrologiques complémentaires au système existant. Cette approche augmente la robustesse et la précision des prévisions. L'inclusion du modèle LSTM, en particulier, a montré une amélioration significative de la précision des prévisions produisant en moyenne des valeurs de CRPS 50% plus faibles, confirmant l'efficacité de l'apprentissage profond pour mieux saisir les dynamiques complexes des systèmes hydrologiques. L'intégration des modèles GR4J et LSTM offrent les meilleures performances, illustrant l'efficacité d'une approche multi-modèle qui tire parti des forces spécifiques de différents modèles hydrologiques pour des prévisions plus précises et robustes.

Mots-clés : LSTM, prévisions hydrologiques d'ensemble, multi-modèle, Deep Learning, modélisation hydrologique, assimilation de données

Optimization of ensemble hydrological forecasting by integrating deep learning into a hybrid multi-model method

William ARMSTRONG

ABSTRACT

The recent application of deep learning in the field of hydrology has shown promising potential. This research focuses on the development of a hybrid multi-model approach, implementing eight lumped and conceptual hydrological models, a semi-distributed model, and a deep learning (DL) model, for forecasting inflows to a hydroelectric reservoir. The introduction of the Long Short-Term Memory (LSTM) model within this multi-model framework is particularly highlighted for its potential to improve the accuracy of short-term flow forecasts in the Lac-Saint-Jean watershed. The results show that the combination of hydrological models with LSTM significantly improves forecast performance, especially for short-term forecasts up to 9 days, outperforming other multi-model combinations by reducing, on average, up to 45% the CRPS values. The study also emphasizes that deep learning, when combined with other modeling approaches, can achieve maximum performance, particularly in terms of forecast distribution where ABDO scores decreased by 64%. The hybrid multi-model method proves effective in capturing hydrological variability at the watershed scale.

In addition, a section of the thesis focuses on the operationalization of these tools and presents the application of this method in a practical context, particularly for improving hydrological forecasting systems at Rio Tinto. Using the CEQUEAU7 hydrological model, already in place at Rio Tinto, the study proposes a simple integration method, consisting of adding and merging complementary hydrological models to the existing system. This approach increases the robustness and accuracy of the forecasts. The inclusion of the LSTM model, in particular, has shown a significant improvement in forecast accuracy, producing on average CRPS values 50% lower, confirming the effectiveness of deep learning in better capturing the complex dynamics of hydrological systems. The integration of the GR4J model with LSTM offers the best performance, illustrating the effectiveness of a multi-model approach that leverages the specific strengths of different hydrological models for more accurate and robust forecasts.

Keywords: LSTM, ensemble hydrological forecasts, multi-model, Deep Learning, hydrological modeling, data assimilation

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 REVUE DE LITTÉRATURE.....	5
1.1 L'hydrologie et modélisation hydrologique	5
1.1.1 Modèles hydrologiques conceptuels.....	6
1.1.2 Modèles hydrologiques à base physique	6
1.1.3 Modèles hydrologiques basés sur l'intelligence artificielle.....	6
1.1.4 Discrétisation spatiale des modèles hydrologiques	8
1.2 Prévisions météorologiques	8
1.3 Prévisions hydrologiques.....	9
1.3.1 Prévisions déterministes	10
1.3.2 Prévisions probabilistes et d'ensemble	11
1.4 Incertitude en prévision hydrologique	12
1.5 Assimilation de données	14
1.6 Prévision multi-modèle.....	16
1.7 Évaluation de la qualité des prévisions hydrologiques.....	17
1.7.1 L'erreur moyenne absolue (MAE).....	18
1.7.2 Continuous Ranked Probability Score (CRPS)	18
1.7.3 Diagrammes de Talagrand.....	19
1.8 Les objectifs de recherche	20
CHAPITRE 2 A HYBRID MULTI-MODEL APPROACH FOR HYDROLOGICAL ENSEMBLE FORECASTING: THE ADDED VALUE OF DEEP LEARNING MODELS.....	21
2.1 Abstract.....	21
2.2 Introduction.....	22
2.3 Study area	25
2.4 Datasets.....	26
2.5 Methodology.....	27
2.5.1 Multiple-models and Data Assimilation	27
2.5.2 Model calibration.....	29
2.5.3 Model forecast performance and evaluation criteria	30
2.6 Results.....	33
2.6.1 Model calibration.....	33
2.6.2 Model comparison in forecasting	35
2.6.3 Two model combination comparison in forecasting	36
2.6.4 Three-model combination comparison in forecasting	38
2.6.5 Four-model combination comparison in forecasting	40
2.6.6 Effect of incremental multi-model ensemble size	42
2.6.7 Validation of the multi-model approach.....	44
2.6.8 Reliability of the multi-model approach.....	46

2.7	Discussion.....	50
2.8	Conclusion and recommendations.....	54
2.9	Acknowledgments	55
CHAPITRE 3 DISCUSSION.....		57
3.1	Analyse des performances de la méthode multi-modèle hybride.....	57
3.2	Analyse des performances selon les différentes saisons.....	59
3.3	Opérationnalisation.....	61
3.4	Limitations et améliorations	64
3.4.1	Modèles hydrologiques.....	64
3.4.2	Période des données météorologiques.....	64
3.4.3	Localisation de l'étude.....	65
3.4.4	Assimilation de données	65
3.4.5	Post-traitement des résultats	66
CONCLUSION.....		67
RECOMMANDATIONS		69
ANNEXE I RÉSULTATS DES VALEURS MAE SUPPLÉMENTAIRES		71
ANNEXE II DIAGRAMME DE TALAGRAND		77
LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES.....		81

LISTE DES TABLEAUX

	Page
Tableau 2.1	Number of parameters for each hydrological model used in this study (Adapted from Dion et al., 2021).....29
Tableau 2.2	KGE scores for the nine hydrological models for the calibration on the LSJ watershed (Adapted from Dion et al., 2021)34
Tableau 2.3	KGE scores for the LSTM model calibrated on the LSJ catchment (Adapted from Sabzipour et al., 2023b) and simulated with lead-times from 1 to 9 days34

LISTE DES FIGURES

		Page
Figure 1.1	Exemple d’hydrogramme représentant une prévision hydrologique déterministe.....	11
Figure 1.2	Exemple d’hydrogramme représentant une prévision hydrologique probabiliste.....	12
Figure 1.3	Exemple d’hydrogrammes observé, simulés et moyennés pour les approches multi-modèles. Adapté d’Arsenault (2015).....	17
Figure 1.4	Exemples de dispersion possible d’un ensemble de 100 membres sur des diagrammes de Talagrand. Adapté de Dion (2020).....	20
Figure 2.1	The Lac-Saint-Jean (LSJ) catchment in the province of Quebec, Canada (subset panel). The LSJ catchment is composed of ten sub-catchments contributing inflows to the reservoir, which is the target flow of this study. The “other tributaries” (light purple coloured area around the main reservoir in southern part of LSJ catchment) include all smaller streams that contribute flows directly to the reservoir without passing through major river systems. Drainage directions (red arrows) are also provided to show flow routing in the catchment. Adapted from Sabzipour et al. (2023b).....	26
Figure 2.2	CRPS of the streamflow ensemble forecasts for the LSJ catchment by season for all models over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot	36
Figure 2.3	Results of two-model combination: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the LSTM alone (red) and when combined with other models over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot.....	38
Figure 2.4	Results of three-model combinations: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the LSTM alone (red), the LSTM-GR4J (blue), and the LSTM-GR4J combined with other models over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot.....	40

Figure 2.5 Results of four-model combinations: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the Best 3 (i.e. LSTM-GR4J-CEQUEAU7_DA) model combination (red), the Best 3 and CEQUEAU model (blue), the Best 3 and HBV model (green), the Best 3 and HYMOD model (purple), the Best 3 and IHACRES model (orange), the Best 3 and MOHYSE model (yellow), the Best 3 and SIMHYD model (brown), and the Best 3 and TOPMODEL model (pink) over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot42

Figure 2.6 Comparison of best one-, two-, and three-model ensembles: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the LSTM (red), the combination of LSTM-GR4J (blue) and LSTM-GR4J-CEQUEAU7_DA (green) over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot.....43

Figure 2.7 Comparison of best one-, two-, and three-model ensembles: MAE of streamflow ensemble forecasts for the LSJ catchment by season for the LSTM (red), the combination of LSTM-GR4J (blue) and LSTM-GR4J-CEQUEAU7_DA (green) over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average MAE values for each lead-time are indicated above each boxplot.....44

Figure 2.8 Hybrid vs purely conceptual hydrologic model ensembles: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the combination of LSTM-GR4J-CEQUEAU7_DA (red), GR4J-HYMOD-TOPMODEL (blue) and GR4J-TOPMODEL-CEQUEAU7_DA (green) over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot45

Figure 2.9 Reliability of forecasts using individual models: Talagrand diagrams representing the rank (in percentiles) of the 5-day streamflow forecasts by the 10 hydrological models on the LSJ catchment. A reliable forecast is expected to lead to flat distributions (similar count in each bin). A p-value less than or equal to 0.05 indicates that the uniformity hypothesis is rejected by the K-S test (red highlighted values).....47

Figure 2.10 Reliability of forecasts using individual models and two- and three-model combinations: Talagrand diagrams representing the rank (in percentiles) of the 5-day streamflow forecasts for the individual CEQUEAU7_DA, LSTM, and GR4J models, as well as for the combination of the LSTM-GR4J and LSTM-GR4J-CEQUEAU7_DA models in the LSJ catchment. A reliable forecast is expected to lead to flat distributions (similar count in each bin). A p-value less than or equal to 0.05 indicates that the uniformity hypothesis is rejected by the K-S test (red highlighted values).....49

Figure 2.11 Comparison between observed and simulated streamflow over a 9-day forecasting period from the GR4J (a-b), CEQUEAU7_DA (c-d), LSTM (e-f), LSTM-GR4J (g-h) and LSTM-GR4J-CEQUEAU7_DA (i-j) models. Left hydrographs are during winter (February 2, 2018, to February 10, 2018) and right hydrographs during summer (August 4, 2018, to August 12, 2018).51

Figure 3.1 CRPS des prévisions hydrologiques d'ensemble pour le bassin versant du LSJ, séparées par saison, pour CEQUEAU7_OL (rouge), CEQUEAU7_DA (bleu), CEQUEAU7_DA & LSTM (vert) et CEQUEAU7_DA, LSTM & GR4J (violet) sur une période de prévision de 2015 à 2019 : hiver (DJFM; a), printemps (AM; b), été (JJA; c) et automne (SON; d).....63

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ABDU	Average Bin Distance to Uniformity
AI	Artificial Intelligence
AM	Saison de la fonte de neige - Avril à Mai
CEQUEAU	Modèle hydrologique développé par l'Institut national de la recherche scientifique (INRS) de l'Université du Québec
CMAES	Covariance Matrix Adaptation Evolution Strategy
CRPS	Continuous rank probability score
DA	Assimilation de données
DJFM	Saison hivernale – Décembre à Mars
DL	Deep Learning
ECCC	Environnement et Changement climatique Canada
ECMWF	European centre for medium-range weather forecasts
EnKF	Filtres d'ensemble de Kalman (Ensemble Kalman filter)
GR4J	Modèle hydrologique Génie rural à 4 paramètres – Journalier, développé par le Centre national du machinisme agricole du génie rural, des eaux et des forêts (CEMAGREF) en France, maintenant devenu l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE)
HBV	Modèle hydrologique « Hydrologiska Byråns Vattenbalansavdelning model » développé par le Swedish meteorological and Hydrological Institute (SMHI)
HOOPLA	Hydrological Prediction Laboratory
HYMOD	Modèle hydrologique « HYdrological MODel » développé par l'Institute of Hydrology à Wallingford en Angleterre.
IA	Intelligence Artificielle

IHACRES	Modèle hydrologique développé conjointement par le Centre for Ecology & Hydrology en Angleterre et le Centre for Resource and Environmental Studies à l'Université national d'Australie.
JJA	Saison d'été - Juin à Août
KGE	Kling-Gupta Efficiency
LSJ	Lac-Saint-Jean
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MOHYSE	Modèle hydrologique simplifié à l'extrême développé par l'Institut national de la recherche scientifique (INRS) de l'Université du Québec NWPE Prévision météorologique d'ensemble (Numerical weather prediction ensemble)
OL	Open Loop (sans utilisation d'assimilation de données)
RNN	Réseau de neurones récurrents (Recurrent Neural Networks)
SIMHYD	Modèle hydrologique simplifié à partir du modèle HYDROLOG et développé par le Cooperative Research Centre for Catchment Hydrology en Australie
SON	Saison d'automne – Septembre à Novembre
TOPMODEL	Modèle hydrologique topographique développé par l'Institute of Hydrology à Wallingford en Angleterre.

INTRODUCTION

Au Québec, l'hydroélectricité est la principale source de production d'électricité, fournissant environ 95% de l'énergie électrique produite dans la province (HEC Montréal, 2023). Cette dépendance à l'hydroélectricité met en lumière l'importance de l'hydrologie. En ingénierie, un des objectifs premiers de l'hydrologie est de prévoir les débits à l'exutoire d'un bassin versant en fonction des événements météorologiques tels que la pluie ou la fonte des neiges. Ces prévisions sont un facteur clé dans la planification et l'exploitation des centrales hydroélectriques. De nos jours, ce sont des modèles hydrologiques qui simulent les processus hydrologiques pour fournir une estimation des débits. Avec la hausse des températures, les changements climatiques affectent notre planète de manière importante et des phénomènes météorologiques extrêmes comme les inondations et les sécheresses deviennent plus fréquents et plus intenses (Arnell & Gosling, 2013). Il est donc essentiel de prendre en compte les changements climatiques dans tous nos projets pour protéger notre environnement et assurer un avenir plus sûr. En hydrologie il est particulièrement important de tenir compte des changements climatiques, car ils ont un impact important sur la disponibilité de l'eau, avec une augmentation de la variabilité et de l'incertitude des précipitations et des débits d'eau, ce qui conduit à plus d'inondations ou de périodes de sécheresse (Fluixá-Sanmartín & Escuder-Bueno, 2018). Ces considérations mettent en lumière l'importance d'une précision accrue dans les prévisions hydrologiques. Bien que ce projet se concentre sur les prévisions hydrologiques à court terme, de 1 à 9 jours, basées sur des états hydrologiques connus, une meilleure prévision permettrait de maximiser l'efficacité des centrales hydroélectriques, un enjeu de plus en plus crucial dans un contexte de valorisation accrue de nos ressources en réponse aux changements climatiques. Ainsi, en améliorant les modèles hydrologiques, il est possible de contribuer à une gestion plus efficace et responsable de cette ressource essentielle, en protégeant à la fois nos communautés et l'environnement.

La modélisation hydrologique peut servir à faire de la simulation hydrologique ou de la prévision hydrologique. La simulation hydrologique consiste à recréer le débit à un endroit dans la rivière (généralement l'exutoire) à partir des données hydrométéorologiques

historiques. Plus généralement, il s'agit de reproduire le cycle de l'eau le plus fidèlement possible. La prévision hydrologique est très similaire, mais plutôt que de recréer un évènement, elle génère un débit futur en fonction des prévisions météorologiques. Il y a deux techniques principales pour effectuer de la prévision, soit de la prévision déterministe ou de la prévision probabiliste.

Concernant les techniques de prévision, on distingue principalement deux approches : la prévision déterministe et la prévision probabiliste. La prévision déterministe repose sur une évaluation spécifique de conditions initiales et de paramètres pour produire une seule estimation du débit futur. Cette méthode offre une réponse directe et claire, mais elle peut manquer de flexibilité face à l'incertitude inhérente aux systèmes météorologiques et hydrologiques. En contraste, la prévision probabiliste prend en compte une série de conditions initiales, produisant ainsi un ensemble de prévisions hydrologiques potentielles. Cette approche se distingue par sa capacité à quantifier l'incertitude, permettant une meilleure évaluation des risques et des probabilités associés aux évènements hydrologiques futurs.

Il existe plusieurs techniques pour améliorer les prévisions hydrologiques qui permettent de minimiser les biais et optimiser leur distribution. L'assimilation de données et le post-traitement sont deux techniques efficaces, mais plutôt que de se fier à un modèle unique, l'approche multi-modèle combine les prévisions de plusieurs modèles pour produire une prévision consolidée. Cette méthodologie a prouvé son potentiel dans divers domaines, y compris en hydrologie, comme l'ont démontré Arsenault et al. (2015) sur 429 bassins versants, et particulièrement dans le domaine de la prévision hydrologique (Dion et al., 2021).

L'introduction de l'Intelligence Artificielle (IA) dans le domaine de l'hydrologie a été très bénéfique. Les techniques d'IA et d'apprentissage profond (Deep Learning, DL), telles que les réseaux neuronaux artificiels, sont de plus en plus utilisées en alternative aux modèles hydrologiques dans les études de prévision (Hu et al., 2019). Le DL utilise des réseaux neuronaux à plusieurs couches ce qui lui permet de modéliser des relations non linéaires complexes et de reconnaître des motifs dans de grands ensembles de données (LeCun et al.,

2015). Les réseaux LSTM (Long Short-Term Memory) sont une forme avancée de réseaux neuronaux récurrents (RNN) introduits par Hochreiter et Schmidhuber (1997) et surpassent les RNN traditionnels (Kratzert et al., 2018). Une fois les LSTM adaptés au traitement des données en séries temporelles, ils présentent un fort potentiel pour améliorer la prévision hydrologique (Zhang et al., 2018). Ils sont considérés comme une alternative fiable aux modèles hydrologiques pour le développement de systèmes de prévision de débit de rivière (Girihagama et al., 2022; Sabzipour et al., 2023b).

Compte tenu des éléments mentionnés ci-dessus liés à la prévision hydrologique d'ensemble, ce mémoire vise à évaluer si l'introduction du modèle LSTM dans un cadre multi-modèle améliore la précision des prévisions hydrologiques à court terme, le tout évalué sur le bassin versant du Lac-Saint-Jean (LSJ). Comme il n'existe pas d'application de prévision multi-modèle hybride incluant trois types de modèles (modèles hydrologiques globaux, distribués et à base de DL), la fiabilité de l'approche multi-modèle hybride est étudiée pour identifier ses avantages pour la prévision hydrologique dans cette région nordique. En particulier, cette étude cherche à déterminer si, en exploitant la performance maximale de chaque modèle de prévision, l'approche multi-modèle hybride proposée améliore la précision des prévisions hydrologiques.

Ce mémoire est structuré en quatre chapitres suivant cette introduction. Le premier chapitre consiste en une revue de littérature, offrant une analyse détaillée des écrits existants sur les méthodes de simulation et de prévision hydrologique, ainsi que sur les approches d'optimisation de ces prévisions. Dans le deuxième chapitre se trouve un article scientifique soumis et destiné à être publié dans le journal "*Journal of Hydrology*". Cet article englobe l'ensemble des résultats du projet de recherche, incluant les informations sur les données utilisées, la zone d'étude, la méthodologie employée, les résultats et leur interprétation. Le troisième chapitre se consacre à une discussion approfondie. Il vise à étendre les conclusions de l'article par l'ajout de résultats supplémentaires, tout en explorant les possibilités d'application pratique de la méthode étudiée. Finalement, la conclusion et les recommandations sont présentées à la 4^e et dernière section de ce mémoire.

CHAPITRE 1

REVUE DE LITTÉRATURE

Cette section présente la revue de littérature portant sur les avancées actuelles en lien avec les prévisions hydrologiques d'ensemble. Elle comprendra une analyse des dernières avancées en matière de prévisions hydrologiques, des différents types de modèles hydrologiques (conceptuel, à base physique et à base de DL), de l'utilisation de la technique multi-modèle et de l'assimilation de données.

1.1 L'hydrologie et modélisation hydrologique

L'hydrologie est principalement la science du cycle de l'eau (Davie., 2019), qui s'étend sur plusieurs facettes comme les précipitations et l'évapotranspiration, le déplacement de l'eau sur et sous la terre, la fonte des neiges qui sont primordiales dans certaines régions et pour des domaines d'application particuliers tels le génie civil. L'hydrologie en génie civil représente une composante fondamentale pour la conception et la gestion des infrastructures liées à l'eau. Le rôle de l'hydrologie s'étend à de multiples facettes de l'ingénierie, de la gestion des cours d'eau et des réservoirs pour la production hydroélectrique, à la protection de l'environnement. Afin de modéliser le cycle hydrologique, des modèles hydrologiques permettant de simuler les processus hydrologiques sont utilisés. Il est important de souligner les différents aspects de la modélisation hydrologique, soit la simulation et la prévision hydrologique. La simulation hydrologique se concentre sur la reconstitution du débit d'un cours d'eau, généralement à son exutoire, en utilisant des données hydrométéorologiques historiques. Cette méthode cherche à reproduire avec la plus grande exactitude le cycle hydrologique. Ensuite, la prévision hydrologique, utilisant une démarche similaire, se distingue par son objectif de prédiction. Plutôt que de reproduire des événements historiques, elle utilise les prévisions météorologiques pour calculer le débit futur des cours d'eau.

1.1.1 Modèles hydrologiques conceptuels

Les modèles conceptuels sont des représentations simplifiées des processus hydrologiques, où les paramètres peuvent ne pas avoir de signification physique directe, mais sont ajustés pour reproduire les observations. Par exemple, le modèle GR4J ne possède que quatre paramètres, mais il est très efficace puisqu'il est rapide à calibrer et reproduit bien les débits. Au lieu de modéliser l'ensemble des processus hydrologiques et considérer le plus de caractéristiques physiques possible du bassin versant, ils se composent uniquement d'équations mathématiques simplifiant la relation entre la météo et les débits (Sood et al., 2015). Les avantages du modèle conceptuel résident dans sa simplicité, sa réduction des besoins en données et temps de calcul, et sa capacité à générer des prévisions à grande échelle.

1.1.2 Modèles hydrologiques à base physique

Les modèles à base physique tentent de représenter les processus hydrologiques réels à partir de principes physiques. Bien que plus précis, ils nécessitent souvent plus de données, sont plus complexes à mettre en œuvre et impliquent une plus grande puissance de calcul par rapport aux modèles conceptuels (Faticchi et al., 2016). Les modèles hydrologiques à base physique fonctionnent à une échelle spatiale fine, en utilisant généralement des grilles ou des sous-bassins versants pour diviser le bassin versant principal en plusieurs unités. Chaque unité est ensuite modélisée individuellement, ce qui permet une représentation complète des processus hydrologiques en fonction des variations spatiales des caractéristiques du terrain, du sol et de la végétation (Clark et al., 2015).

1.1.3 Modèles hydrologiques basés sur l'intelligence artificielle

Grâce à l'abondance de données et à la puissance informatique actuelle, il est désormais possible d'utiliser des modèles de prévision hydrologique basés sur l'intelligence artificielle (IA), qui font partie des approches dites « data-driven ». Ces modèles exploitent la capacité de l'IA à apprendre de grandes quantités de données pour prédire avec précision des événements hydrologiques. Les modèles hydrologiques ont commencé à intégrer des techniques

d'apprentissage profond (Deep Learning, DL). Le DL, une sous-catégorie de l'IA, utilise des réseaux de neurones à plusieurs couches pour apprendre des représentations de données complexes. Ces réseaux peuvent identifier des motifs et des relations dans de grandes quantités de données que les méthodes traditionnelles ne peuvent pas facilement détecter. En hydrologie, cela a permis des prévisions plus précises et robustes en tenant compte de multiples variables et interactions complexes dans le cycle de l'eau (Sit et al., 2020). Ils sont utilisés pour leur précision supérieure à reproduire les débits en rivière (Tanty et Desmukh, 2015).

Les modèles à mémoire à long terme et à court terme (Long Short-Term Memory, LSTM), introduit par Hochreiter et Schmidhuber (1997), sont un type avancé de réseau de neurones récurrents (Recurrent Neural Networks, RNN) utilisé dans le DL, maintenant possible d'être mis en application avec l'avancement de la technologie (ordinateurs plus performants) et l'abondance de données historiques. Ils sont particulièrement adaptés pour traiter des séquences de données, comme les séries temporelles, qui sont courantes en hydrologie. Contrairement aux réseaux neuronaux traditionnels, les LSTM ont la capacité de retenir des informations sur de longues périodes, ce qui est essentiel pour comprendre les tendances et les cycles dans les données hydrologiques. Zhang et al. (2018) ont démontré que le modèle LSTM performe bien même dans un contexte de gestion de réservoir, en particulier pour sa capacité à simuler des conditions de faible débit et les périodes de pointe. L'utilisation des modèles LSTM en hydrologie représente une avancée significative, car ils peuvent modéliser avec précision les dynamiques temporelles complexes des systèmes hydrologiques. Même pour les bassins versants non jaugés, le modèle LSTM surpasse d'autres modèles, comme le démontrent Kratzert et al. (2019). Leur étude révèle que, sur 531 bassins versants, le modèle LSTM est plus efficace que le modèle à base physique SAC-SMA, même quand ce dernier est utilisé sur des bassins jaugés. Plus récemment, Sabzipour et al. (2023b) ont démontré que les prévisions hydrologiques à court terme réalisées par un modèle LSTM sont plus fiables et précises que celles d'un modèle semi-distribué couplé à de l'assimilation de données. Ils soulignent que le modèle LSTM génère des valeurs de l'erreur absolue moyenne (Mean Absolute Error, MAE) 42% plus faible par rapport au modèle hydrologique assimilé.

1.1.4 Discrétisation spatiale des modèles hydrologiques

Dans cette recherche, deux techniques de discrétisation spatiale des modèles hydrologiques sont utilisées, il y a des modèles globaux et un modèle distribué. Un modèle hydrologique global est un modèle qui représente les processus hydrologiques de façon uniforme sur l'ensemble d'un bassin versant comme une moyenne. Un modèle global se retrouve sous différentes formes, notamment sous forme conceptuelle ou à base d'IA, comme le modèle GR4J et un modèle LSTM respectivement. Un modèle hydrologique distribué peut être autant un modèle conceptuel ou à base physique. Le principe d'un modèle distribué, ou semi-distribué, est de diviser un bassin versant sous forme de grille ou en sous-bassins versants pour représenter les processus hydrologiques à plus petite échelle, afin de prendre en considération la variabilité hydroclimatique et les caractéristiques physiques du bassin versant. Généralement, un modèle à base physique est automatiquement distribué, mais un modèle distribué peut être conceptuel ou même mixte.

1.2 Prévisions météorologiques

La météorologie joue un rôle primordial dans notre quotidien. Que ce soit pour l'agriculture, la navigation, le transport aérien, ou simplement pour planifier nos activités, les prévisions météorologiques influencent bon nombre de nos décisions. De nos jours, les techniques de prévision météorologique ont beaucoup évolué. Historiquement, les prévisions étaient basées sur des observations empiriques et avec le temps, des instruments comme le baromètre, l'anémomètre, le pluviomètre et le thermomètre ont été développés, permettant des mesures plus systématiques. Toutefois, avant même l'arrivée de ces instruments, l'idée que la météo pouvait être prédite à l'aide de calculs mathématiques était déjà dans l'air (Abbe, 1901). Ce n'est pas avant 1954 que la première prévision météorologique a été réalisée (Bolin, 1955) et encore, il a fallu attendre 20 ans d'évolution pour avoir un ordinateur assez puissant pour fournir une prévision complète. C'est grâce à cette évolution qu'il est maintenant possible d'effectuer des prévisions météorologiques à différentes échelles, allant de prévisions horaires (court terme), sur 48 heures, jusqu'à une quinzaine de jours pour les prévisions journalières (moyen terme). Malgré d'énormes progrès, la prévision de phénomènes extrêmes reste

complexe. Cependant, des avancées dans les modèles numériques telles que l'apprentissage automatique et l'intelligence artificielle commencent à être intégrées pour améliorer la précision des modèles. La prévision météorologique, autrefois basée sur des observations simples, est devenue une science complexe qui allie observations de pointe, modélisation avancée et techniques de calcul. L'European Centre for Medium-Range Weather Forecasts (ECMWF), ou Centre Européen pour les Prévisions Météorologiques à Moyen Terme en français, est une organisation de renommée internationale dédiée à l'amélioration des prévisions météorologiques à moyen terme à l'échelle mondiale. En plus de l'ECMWF, Environnement et Changement climatique Canada (ECCC) joue également un rôle important en fournissant des prévisions météorologiques de haute qualité et des informations cruciales sur le climat. Leur rôle principal est de collecter, analyser et diffuser des données météorologiques et climatiques afin de fournir des prévisions météorologiques précises et de contribuer à la compréhension des phénomènes météorologiques et climatiques. Les progrès réalisés ces dernières décennies offrent une meilleure compréhension et prédiction du temps, au bénéfice de tous (Bauer, Thorpe & Brunet, 2015). Toutefois, un grand nombre d'incertitudes reste présent (Leutbecher et al., 2017).

1.3 Prévisions hydrologiques

La prévision hydrologique est une discipline clé de l'hydrologie, qui s'intéresse à la prédiction du débit d'eau, généralement dans une rivière, à un endroit et un moment précis. Ces prévisions jouent un rôle important dans la gestion des ressources en eau, la prévention des inondations et la production d'énergie hydroélectrique (Schaake et al., 2007). La base pour réaliser une prévision hydrologique est le résultat de la prévision météorologique, à savoir les précipitations et la température. Ensuite, pour des modèles hydrologiques plus complexes (par exemple, les modèles à base plus physiques et distribués), d'autres intrants peuvent être ajoutés (par exemple, la vitesse du vent, l'humidité relative de l'air, etc.). Cependant, pour les modèles plus simples (par exemple, les modèles conceptuels globaux), les intrants nécessaires sont uniquement les précipitations et la température moyennes à l'échelle du bassin versant. Historiquement, les prévisions hydrologiques, comme des prévisions météorologiques,

s'appuyaient principalement sur des observations empiriques (Eagleson, 1994). Cela signifie qu'elles étaient fondées sur des données collectées au fil du temps, telles que les niveaux d'eau des rivières et les précipitations. En parallèle, les méthodes statistiques étaient employées pour analyser ces données. Ces méthodes incluaient diverses techniques de calcul pour estimer la probabilité d'événements futurs, basées sur les tendances et les modèles observés dans les données historiques. Avec les progrès dans le domaine des technologies informatiques et l'arrivée des méthodes numériques, une évolution significative s'est produite dans le domaine des prévisions hydrologiques. Les modèles hydrologiques ont évolué pour intégrer des approches plus sophistiquées. Les modèles conceptuels, par exemple, utilisent des algorithmes pour simuler les processus hydrologiques de manière simplifiée (Sood & Smakhtin, 2015). Les modèles à base physique, quant à eux, tentent de reproduire plus fidèlement les processus physiques réels, tels que l'infiltration d'eau dans le sol, le type de végétation ou la géométrie du bassin versant (Clark et al., 2015). Par ailleurs, l'intégration de l'intelligence artificielle a permis de développer des modèles capables d'apprendre de grandes quantités de données et d'améliorer la précision des prédictions. Ces avancées technologiques ont conduit à une popularité accrue de ces méthodes plus avancées, offrant une compréhension plus profonde et des prévisions plus précises des phénomènes hydrologiques (Troin et al., 2021). La prévision hydrologique est une discipline en constante évolution, qui bénéficie des progrès technologiques et méthodologiques.

1.3.1 Prévisions déterministes

Les prévisions déterministes génèrent une seule valeur de débit puisqu'elles n'utilisent qu'une série de conditions initiales déterminées en fonction de la météo la plus propice (voir Figure 1.1). L'exactitude des prévisions déterministes dépend fortement de la qualité des données en entrée (Chao et al., 2015), de la précision des modèles utilisés, et de la compréhension des processus hydrologiques. Les progrès technologiques ont amélioré la collecte et le traitement des données, renforçant ainsi la fiabilité des prévisions. Cependant, malgré leur précision, les prévisions déterministes ont leurs limites, notamment face à l'incertitude inhérente aux conditions météorologiques (Her et al., 2019). Elles sont de moins en moins utilisées, car les

prévisions déterministes ne représentent pas l'ensemble des débits possibles et négligent donc d'informer quant à l'incertitude de la valeur prédite (Demargne et al., 2014). Pour pallier ces incertitudes, les hydrologues combinent souvent les approches déterministes avec des méthodes probabilistes, offrant ainsi une vision plus complète et robuste des prévisions hydrologiques.

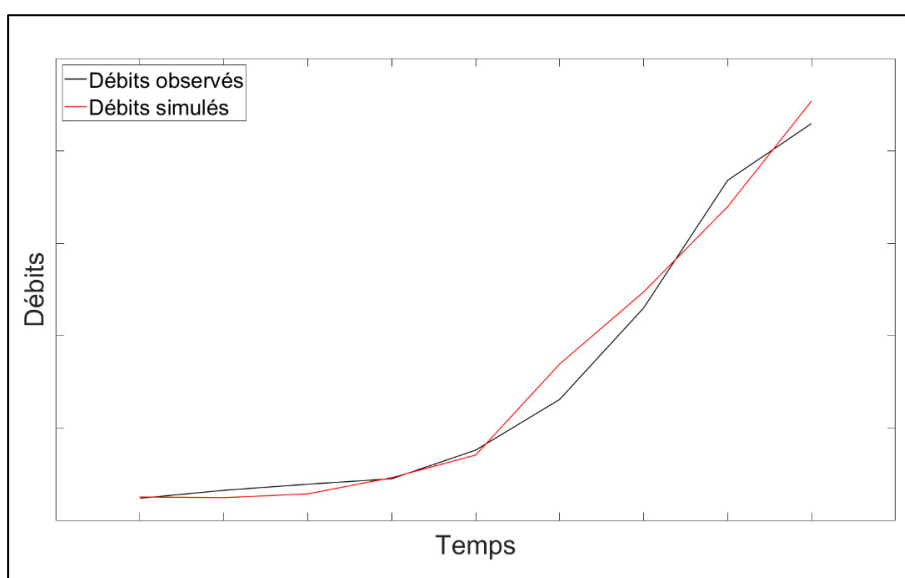


Figure 1.1 Exemple d'hydrogramme représentant une prévision hydrologique déterministe

1.3.2 Prévisions probabilistes et d'ensemble

Reconnaissant les incertitudes inhérentes à la prévision hydrologique, les hydrologues ont adopté des approches probabilistes. La prévision probabiliste est généralement considérée comme supérieure parce qu'elle fournit une vue plus complète des scénarios futurs, permettant de contextualiser les résultats obtenus lors d'une prévision et permettant de prendre des décisions de gestion des ressources hydriques en contexte de gestion des risques éclairée (Krzysztofowicz, 2001). Les prévisions par ensembles, par exemple, génèrent plusieurs scénarios à partir de conditions initiales légèrement différentes, offrant ainsi une gamme de résultats possibles et une estimation de leur probabilité (voir Figure 1.2). Comme les modèles météo utilisent aussi des méthodes probabilistes, il est possible d'utiliser l'ensemble des

extrants d'un modèle météo comme intrants pour un modèle hydrologique et, du même coup, d'obtenir une prévision d'ensemble. En général, la moyenne des membres d'une prévision d'ensemble est plus performante qu'une prévision déterministe (McCollor & Stull, 2008; Wetterhall et al., 2013). De plus, il est possible de quantifier la probabilité des évènements extrêmes avec l'approche probabiliste (Wetterhall et al., 2014).

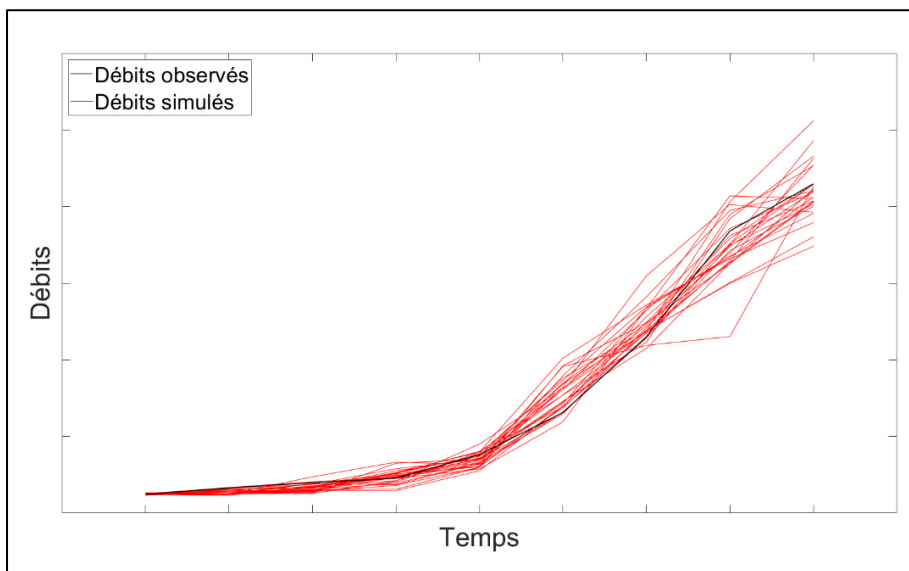


Figure 1.2 Exemple d'hydrogramme représentant une prévision hydrologique probabiliste

1.4 Incertitude en prévision hydrologique

L'incertitude est inhérente aux prévisions hydrologiques, mais il est possible de les quantifier (Bruen et al., 2010; Rossa et al., 2011; Zappa et al., 2010). Elle peut provenir des erreurs de mesure, de l'imprécision des modèles ou de la variabilité naturelle des processus hydrologiques. Les incertitudes primaires sont les intrants des modèles hydrologiques, soient les données météorologiques (précipitation et température), ainsi que les débits observés utilisés lors du calage (paramétrisation) des modèles hydrologiques. Les prévisions météorologiques sont aussi empreintes d'incertitudes liées à l'estimation des états de l'atmosphère et de la nature chaotique du système météorologique. L'incertitude des prévisions météorologiques est ensuite propagée à travers le modèle hydrologique lors de la prévision

hydrologique, augmentant l'incertitude totale des processus. Comme les modèles hydrologiques sont calibrés avec des données météorologiques historiques afin de reproduire le plus fidèlement à la réalité que possible le débit à l'exutoire d'un bassin versant et qu'ils sont dépendants des événements précédant la prévision hydrologique, il est de plus en plus important que la prévision météorologique soit représentative de la météo historique. Les approches comme l'assimilation de données ou les prévisions d'ensemble tentent de quantifier et de réduire ces incertitudes.

Une seconde source d'incertitude serait les paramètres utilisés pour les modèles hydrologiques. L'incertitude peut provenir de la sensibilité des modèles aux paramètres. Certains modèles sont extrêmement sensibles à de légères variations dans les valeurs des paramètres, ce qui peut entraîner de grandes différences dans les résultats de prévision (Ratto et al., 2007). L'incertitude paramétrique est aussi liée à l'interaction entre les paramètres. Dans de nombreux cas, les paramètres ne fonctionnent pas de manière isolée, mais interagissent les uns avec les autres. Cela peut compliquer le processus de calibration, car l'ajustement d'un paramètre peut influencer la performance d'un autre, ce qui rend difficile de trouver la combinaison optimale de paramètres. Dépendant de la méthode de calibration utilisée, selon le type de modèle, les paramètres ne seront jamais exacts et souvent différents d'une calibration à une autre dépendant des critères de calage et de l'hydrologue. Par exemple, le choix des critères de performance, comme la fonction-objectif Kling-Gupta Efficiency (KGE; Gupta et al., 2009; Kling et al., 2012) ou le Nash-Sutcliffe (Nash & Sutcliffe, 1970), pour évaluer la qualité du calage peut influencer les valeurs des paramètres obtenues comme il est expliqué dans les travaux de Kavetski et al. (2003).

Une autre source d'incertitude est liée à la structure même du modèle hydrologique. Chaque modèle est basé sur des hypothèses et des simplifications de la réalité du cycle de l'eau. Par exemple, certains modèles peuvent ne pas prendre en compte les interactions entre les eaux de surface et les eaux souterraines, ce qui peut mener à des prédictions inexactes dans certaines conditions. Ceci amène un autre point : ce ne sont pas tous les modèles hydrologiques qui sont performants sur tous les bassins versants (Kavetski et al., 2006). Par exemple, certains modèles

peuvent être plus performants dans des régions nordiques que dans des régions chaudes et arides.

Enfin, les changements climatiques et les altérations humaines sur le bassin versant (comme l'urbanisation, la déforestation, la construction de barrages, etc.) ajoutent une couche supplémentaire d'incertitude (Hattermann, Huang, & Koch, 2015). Les modèles hydrologiques sont souvent calibrés sur des données historiques, et les changements dans les conditions climatiques ou dans l'utilisation des terres peuvent rendre ces modèles moins précis pour prédire les conditions futures. Des recherches, comme celle effectuée par Bastola, Murphy & Sweeney (2011), se penche sur le rôle des incertitudes des modèles hydrologiques face aux changements climatiques et sur des techniques pour mitiger cette problématique.

1.5 Assimilation de données

Comme mentionné dans les sections précédentes, les prévisions hydrologiques sont accompagnées d'incertitudes provenant des conditions initiales, de paramètres de modèle inexacts et d'une déficience structurelle du modèle hydrologique (Kavetski et al., 2006). Une approche visant à réduire l'incertitude prédictive et, par conséquent, à améliorer les prévisions hydrologiques consiste à utiliser des techniques d'assimilation de données (data assimilation en anglais, DA) (Mazrooei & Sankarasubramanian, 2019). L'assimilation de données est utilisée en hydrologie avant la simulation pour faire varier les états initiaux des modèles hydrologiques. L'assimilation de données améliore les performances des prévisions de débit en améliorant les états initiaux du modèle pour générer la prévision (Piazzi, Thirel & Perrin, 2021). Cette manipulation est effectuée, car lors d'une simulation hydrologique, les états initiaux du modèle ne correspondent pas nécessairement au débit observé dû au fait que la simulation récente n'est pas parfaite et cumule des écarts entre les débits observés et simulés avec le temps. Alors l'assimilation de données vient corriger cet écart en modifiant les états initiaux dans le but de le rapprocher le plus possible de l'observation ou de la meilleure valeur de débit possible. De plus, l'assimilation de données permet de créer une variabilité dans les conditions initiales en produisant un plus grand échantillon de résultats.

Dans le cas de cette étude, seule l'assimilation de données automatique a été utilisée. Parmi les méthodes d'assimilation de données automatique, l'assimilation de données séquentielle a suscité de l'intérêt pour les applications à court terme en raison de la mise à jour des états du modèle à mesure que de nouvelles observations deviennent disponibles (Leisenring & Moradkhani, 2011). De nombreuses études ont exploré le potentiel des méthodes d'assimilation de données séquentielles pour améliorer la précision des prévisions hydrologiques en utilisant le filtre de Kalman (Kalman filter, KF ; Kalman, 1960), une variante telle que le filtre d'ensemble de Kalman (ensemble Kalman filter, EnKF ; Evensen, 1994) et le filtre particulaire (particle filter, PF; Arulampalam et al., 2002). Piazzzi et al. (2021) ont évalué la performance du EnKF et du PF sur un modèle hydrologique conceptuel global sur 232 bassins versants en France et leurs résultats montrent que les prévisions basées sur l'EnKF surpassent celles basées sur le PF. Dans la plupart des applications, l'EnKF parvient à améliorer la précision des prévisions hydrologiques comme dans les travaux de Sabzipour et al. (2023a), qui ont mis en œuvre l'EnKF dans un modèle hydrologique semi-distribué, démontrent que l'EnKF contribue largement à augmenter la précision des prévisions en caractérisant l'incertitude des hyperparamètres individuellement.

En résumé, la méthode d'assimilation de données EnKF commence par générer un ensemble initial de scénarios basés sur des états et des paramètres du modèle hydrologique, puis prédit le comportement hydrologique pour une période donnée. Les données de débit observationnelles réelles sont ensuite comparées aux débits simulés par le modèle alimenté par la météo observée, et une relation est établie entre les états du modèle et les débits simulés pour chacun des scénarios. L'EnKF génère ensuite une nouvelle série d'états qui représente la distribution la plus probable des états réels du bassin versant. Ce processus itératif se poursuit à chaque pas de temps avant l'étape de prévision, permettant à l'ensemble de générer des débits à partir d'états réalistes et donc moins incertains. Ceci permet d'améliorer la précision des prévisions hydrologiques au fil du temps et de mieux quantifier l'incertitude. Selon les travaux de Thiboult, Anctil & Boucher (2016), l'EnKF est plus efficace lorsqu'il est combiné avec des prévisions hydrologiques d'ensemble et une méthode multi-modèle.

1.6 Prévision multi-modèle

Face à la complexité des systèmes hydrologiques et à la multitude des modèles disponibles, l'approche multi-modèle hydrologique est apparue comme une méthode prometteuse pour améliorer la qualité et la robustesse des prévisions (Velazquez et al., 2011). L'approche multi-modèle consiste à utiliser simultanément plusieurs modèles hydrologiques pour une même prévision. Dans le contexte de cette étude, les prévisions individuelles sont ensuite combinées pour obtenir une prévision d'ensemble et ainsi créer un grand ensemble multi-modèle pour générer plus de variabilité dans les prévisions. Évidemment, l'utilisation de plusieurs modèles hydrologiques demande plus de ressources informatiques (Jiang et al., 2014).

D'autres techniques multi-modèles sont également disponibles dans la littérature comme la méthode de Granger et Ramanathan (Granger et Ramanathan 1984) qui consiste à effectuer une sorte de moyenne pondérée des modèles, ou encore la méthode utilisée dans les travaux de See et Openshaw (2000) qui consiste à sélectionner le meilleur modèle pour chaque jour de prévision. La variabilité des processus hydrologiques, les échelles spatiales et temporelles considérées, ainsi que la disponibilité des données font qu'un seul modèle ne peut pas toujours capturer l'ensemble des dynamiques d'un bassin versant. La prévision hydrologique multi-modèle offre une approche pour améliorer la qualité des prévisions en tirant parti de la force de chaque modèle tout en compensant leurs faiblesses individuelles (Arsenault et al., 2015; Arsenault & Brissette, 2016). Comme les prévisions d'ensemble, la moyenne des membres d'un ensemble multi-modèle est généralement plus performante que celle d'un seul modèle puisque certains modèles peuvent surestimer un événement tandis que d'autres peuvent sous-estimer le même événement (voir Figure 1.3). Le type de modèle hydrologique utilisé (conceptuel, à base physique ou à base de DL) est également un facteur à considérer lors de l'utilisation d'une technique multi-modèle. Comme mentionné à la section 1.3.5, un modèle à base physique est censé être plus précis qu'un modèle global conceptuel, mais celui-ci demande beaucoup plus de paramètres et est plus complexe à calibrer. Un avantage de la méthode multi-modèle est qu'il serait possible d'utiliser plusieurs modèles globaux conceptuels, plus simples et plus rapides à calibrer qu'un seul modèle à base physique, et

obtenir de meilleures prévisions (plus précises et fiables) que celles produites par le modèle plus physique. Il est également possible de combiner les différents types de modèle afin d'améliorer les prévisions comme dans les travaux de Thébault et al. (2023). Sur ce même ordre d'idées, une méthode multi-modèle combinant des modèles conceptuels globaux, distribués et à base de DL n'a jamais été appliquée.

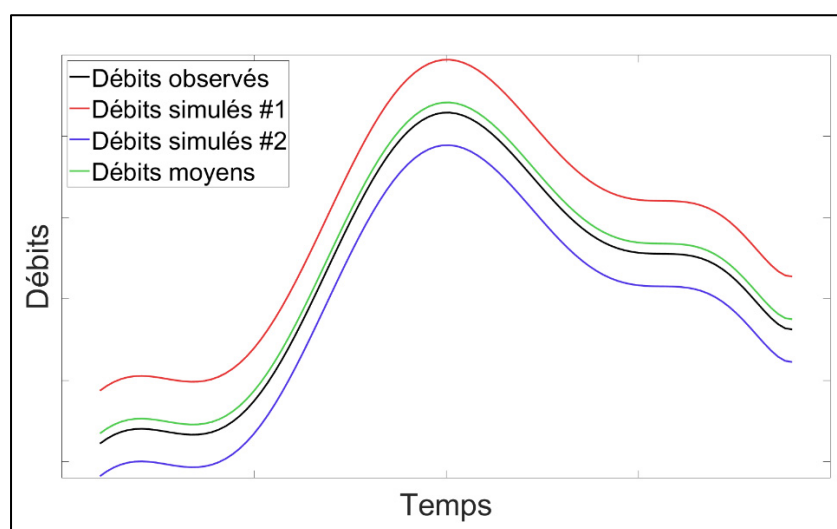


Figure 1.3 Exemple d'hydrogrammes observés, simulés et moyennés pour les approches multi-modèles. Adapté d'Arsenault (2015)

1.7 Évaluation de la qualité des prévisions hydrologiques

Il est parfois avantageux d'évaluer les performances des modèles de prévision hydrologique sur une base saisonnière, car l'efficacité de ces modèles peut varier en fonction de la saison. Par exemple, les modèles peuvent rencontrer plus de difficultés pour prévoir les crues du printemps en raison de la fonte des neiges et des changements rapides des conditions météorologiques. En effectuant une évaluation des performances des modèles par saison, on obtient plus de renseignements sur leurs forces et faiblesses sur une période spécifique, ce qui permet une évaluation plus précise. Selon les travaux de Peñuela, Hutton & Pianosi, (2020), il existe également des avantages opérationnels, notamment en gestion de réservoir, à utiliser des prévisions basées sur une approche saisonnière. Ces avantages peuvent contribuer à une

gestion plus efficace et proactive des ressources en eau. Il existe plusieurs méthodes pour évaluer la qualité des prévisions hydrologiques. Elles se concentrent principalement sur le calcul des écarts entre les débits simulés et les débits observés. Cette évaluation peut se faire de différentes manières, en utilisant une variété de mesures, afin de déterminer avec précision l'exactitude des prévisions hydrologiques. Les sections suivantes se concentreront sur la présentation des techniques utilisées dans cette étude, soit l'erreur moyenne absolue (Mean Absolute Error, MAE), le Continuous Ranked Probability Score (CRPS) et les diagrammes de Talagrand.

1.7.1 L'erreur moyenne absolue (MAE)

Le MAE est une mesure de performance courante pour les modèles de prévision hydrologique déterministes qui mesure la différence moyenne entre les prévisions et les observations. Même si le MAE n'est pas censé être utilisé avec des prévisions probabilistes, il est possible de l'utiliser si la moyenne des prévisions est utilisée. Le MAE est facile à calculer et à interpréter et il est souvent utilisé pour évaluer la précision des modèles hydrologiques et pour comparer différents modèles entre eux (Thiboult et al., 2016; Sun et al., 2020; Sabzipour et al., 2023b). Pour évaluer la performance d'un modèle hydrologique avec le MAE, la différence, en valeur absolue, entre la prévision du modèle et l'observation correspondante est calculée, ce qui représente l'erreur absolue pour un point. La moyenne des différences absolues pour l'ensemble des données est ensuite d'être calculée, ce qui fournit une mesure de l'erreur absolue moyenne entre les prévisions et les observations (MAE). Un MAE faible indique une meilleure performance du modèle. Plus de détails sur le MAE sont présentés à la section 2.5.3.

1.7.2 Continuous Ranked Probability Score (CRPS)

Le CRPS est une mesure de performance pour les modèles de prévision probabiliste. Il mesure l'écart entre la distribution de probabilité des prévisions et des observations. Le CRPS fournit une mesure de la performance du modèle en termes de biais et de dispersion des prévisions par rapport aux observations. Le CRPS est un outil couramment utilisé pour évaluer la performance des prévisions hydrologiques probabilistes (Pagano et al., 2013; Boucher et al.,

2015; Dion et al., 2021; Sabzipour et al., 2023b). Un CRPS faible indique une meilleure performance, soit des prévisions plus précises et fiables comme le démontrent les travaux d'Hersbach (2000). Plus de détails sur le CRPS sont présentés à la section 2.5.3.

1.7.3 Diagrammes de Talagrand

Les diagrammes de Talagrand (Talagrand & Vautard, 1997), aussi connus sous le nom d'histogramme de rang (Hamill & Collucci, 1997) ou d'ensemble de probabilité par contenant (Anderson, 1996), sont un outil statistique en météorologie et hydrologie particulièrement approprié pour évaluer la qualité des prévisions d'ensemble (Hamill, 2001). Les diagrammes de Talagrand ont été initialement développés dans le domaine de la météorologie numérique, ensuite ils ont trouvé une application en hydrologie, plus particulièrement pour l'analyse des prévisions hydrologiques d'ensemble. Ils mesurent la distribution des rangs des observations réelles parmi les prévisions d'ensemble, classant chaque observation selon sa position parmi les prévisions triées. Cette méthode permet d'évaluer la fiabilité des prévisions, en montrant si les prévisions d'ensemble couvrent adéquatement la gamme des résultats possibles. Une surdispersion indique que la distribution des prévisions d'ensemble est trop large, ce qui entraîne la présence de membres aberrants qui sont soit beaucoup trop bas, soit beaucoup trop élevés par rapport aux observations (Figure 1.4c). Les biais systématiques négatifs ou positifs indiquent que les prévisions d'ensemble ne correspondent pas aux observations, soit des prévisions toujours soit plus élevées (biais positif, Figure 1.4d) soit plus basses (biais négatif, Figure 1.4f) que les membres des prévisions d'ensemble. Une distribution uniforme des rangs indique une bonne fiabilité.

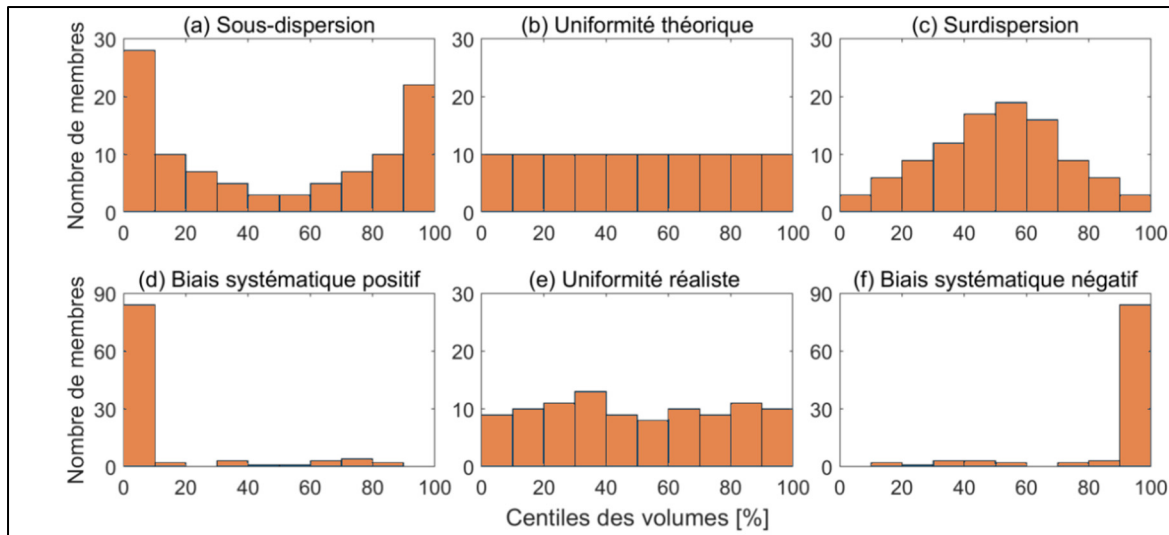


Figure 1.4 Exemples de dispersions possible d'un ensemble de 100 membres sur des diagrammes de Talagrand. Adapté de Dion (2020)

1.8 Les objectifs de recherche

Cette étude se concentre sur une approche hybride multi-modèle qui combine huit modèles hydrologiques globaux, un modèle semi-distribué et un modèle d'apprentissage profond, spécifiquement le modèle à mémoire long terme et court terme (LSTM). L'objectif principal est d'améliorer la précision des prévisions de débit à court terme dans le bassin versant du Lac-Saint-Jean. Cette étude cherche à déterminer si l'apprentissage profond, lorsqu'il est combiné avec d'autres méthodes de modélisation utilisant de l'assimilation de données, permet d'atteindre une performance supérieure. De plus, l'étude aborde l'opérationnalisation de cette méthode en contexte pratique, en l'appliquant pour améliorer les systèmes de prévision hydrologique chez Rio Tinto. En utilisant le modèle hydrologique CEQUEAU7 déjà en place, l'étude propose une approche d'intégration simple en ajoutant et en fusionnant des modèles hydrologiques complémentaires au système existant. Cette démarche renforce la fiabilité et la précision des prévisions. Cette recherche tiendra également compte de l'impact des différentes saisons, car en région nordique le régime hydrologique est très variable, et utilisera plusieurs critères de performance pour évaluer les résultats obtenus.

CHAPITRE 2

A HYBRID MULTI-MODEL APPROACH FOR HYDROLOGICAL ENSEMBLE FORECASTING: THE ADDED VALUE OF DEEP LEARNING MODELS

William Armstrong ¹, Richard Arsenault ¹, Jean-Luc Martel ¹, Magali Troin ^{1,2}, Patrice Dion ³, Behmard Sabzipour ¹, François Brissette ¹, Juliane Mai ⁴

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

² HydroClimat| TVT, Maison du Numérique et de l'Innovation, place Georges Pompidou, 83 000, Toulon, France

³ CIMA+, 740 R. Notre Dame O Bureau 900, Montreal, Quebec, Canada H3C 3X6

⁴ Earth and Environmental Science, University of Waterloo, 200 University Ave W, Waterloo, ON, N2L 3G1

This paper has been submitted for publication
in “*Journal of Hydrology*”,
January 2024

2.1 Abstract

Deep learning has recently shown promise for hydrological applications. The accuracy of a hybrid multi-model neural network approach for streamflow forecasting is investigated with the use of nine conceptual hydrological models, i.e. eight lumped and one semi-distributed, and one deep learning (DL) model. This study aims to evaluate if the introduction of the Long Short-Term Memory (LSTM) DL model within the multi-model framework improves the accuracy of short-term streamflow forecasts over a Canadian catchment. Results show that combining the hydrological models with the LSTM leads to an improved forecast skill of streamflow, especially for the early lead-times up to 9 days. We found a better performance of the LSTM when combined with other hydrological models, suggesting that DL achieves optimal performance when paired with different modelling approaches. The hybrid multi-model approach captures hydrological variability, both spatially and temporally at the catchment scale, indicating the ability of the approach to be used in regions sensitive to climate variability. Furthermore, insights derived from Talagrand diagrams underscore the importance of conducting a model selection among the multi-model approach when streamflow forecasts are intended for specific hydrological conditions and seasons. Even though the LSTM is

computationally time expensive during training, such DL model appear to be a promise predictive tool for hydrological forecasting when integrated within a hybrid multi-model framework.

2.2 Introduction

Hydrological forecasting plays an important role in water management and planning, supporting socio-economic development and water-related risk decisions (Boucher et al., 2011; Anghileri et al., 2016; Cassagnole et al., 2021). With the emergence of new technologies and modeling methods, various hydrological forecasting approaches have been proposed in the last few decades to predict catchment streamflow with increasing accuracy (Troin et al., 2021; Ng et al., 2023).

Short-term streamflow forecasts are essential for flood early prediction and emergency management, where hourly and daily forecasts can reduce flood risks and damage. Conversely, medium and long-term streamflow forecasts are useful in water resource planning and reservoir management, addressing time horizons from several days to several months (Davie., 2019). In these two fields, deterministic and probabilistic forecasts of streamflow can be delivered. A deterministic forecast model provides a single forecast based on a set of input conditions, without any information on the likelihood of that outcome occurring (Her et al., 2019). In contrast, a probabilistic forecast model acknowledges the range of possible forecasts and assigns a probability of happening to each outcome (Zappa et al., 2008). It is widely recognized that the comprehensive view of potential outcomes increases the application value of probabilistic forecasts (Krzysztofowicz, 2001). By providing information regarding the confidence range of streamflow forecasts, decision-makers can assess risks more effectively and implement measures that account for multiple potential scenarios.

Lumped hydrological models offer a macroscopic perspective on the hydrological behavior of a catchment (Jajarmizadeh et al., 2012). They rely on empirical or conceptual equations and describe the catchment as a single homogeneous spatial entity. The water balance is averaged over the entire catchment, without considering the spatial variability of hydrological processes

and input data (Sood & Smakhtin, 2015). The advantages of the lumped models lie in their simplicity with a reduced data requirement; they are particularly suited to the generation of large-scale streamflow forecasts. However, they may occasionally lack accuracy, especially when dealing with specific hydrological events or regions characterized by substantial hydrological heterogeneity (Carpenter et al., 2006). In contrast, distributed hydrological models are operated at a finer spatial scale, where the catchment is discretized into grids or homogeneous sub-catchments. They allow a comprehensive representation of hydrological processes based on spatial variations in terrain, soil, and vegetation characteristics (Clark et al., 2015). Although they are more demanding in terms of input data and computation time compared to the lumped models, they provide a better representation of local hydrological phenomena (Fatichi et al., 2016).

However, hydrological forecasts are attached with uncertainty stemming from initial conditions, inaccurate model parameters, and structural deficiency of the hydrological model (Kavetski et al., 2006). One approach to reduce the predictive uncertainty and consequently improve the hydrological forecasts is the use of Data Assimilation (DA) techniques (Mazrooei & Sankarasubramanian, 2019). By integrating various sources of observational data (e.g., river discharge, soil moisture, and snow water equivalent) with model forecasts, DA increases the skills of streamflow forecasts by improving initial model parameters and states used to issue the forecast (Piazzini et al., 2021). Various DA techniques are used to assimilate, for example, streamflow data that are available until the day the forecast is issued into hydrological models for forecasting purposes (Liu et al., 2012). Sequential DA methods have gained interest for short-term applications because of the sequentially updating of the model states and/or parameters as new observations become available (Leisenring & Moradkhani, 2011). Many studies have explored the potential of sequential DA methods to improve the accuracy of streamflow forecasts using Kalman filter (KF; Welch & Bishop, 1995) and variants such as ensemble Kalman filter (EnKF; Evensen, 1994) and extended Kalman filter (EKF; Shamir et al., 2010) (Mazrooei & Sankarasubramanian, 2019; Sun et al., 2020; Piazzini et al., 2021; Sabzipour et al., 2023a). In most applications EnKF succeeds in enhancing the streamflow forecast skills; for example, Sabzipour et al. (2023a) that implemented the EnKF in a semi-

distributed hydrological model, demonstrates that the EnKF largely contributes to increasing prediction accuracy by characterizing the uncertainty of individual hyperparameters (i.e., temperature) in forecasting outcomes for snow-dominated catchments in Canada.

Artificial intelligence (AI) and deep learning (DL) techniques, such as artificial neural networks, are being increasingly used as alternative to hydrological models in forecasting studies (Hu et al., 2019). DL, a subset of machine learning, employs neural networks with many layers (hence "deep") to analyze various forms of data. When it comes to forecasting, DL has shown significant promise due to its ability to model complex nonlinear relationships and to recognize patterns in large datasets (LeCun et al., 2015). LSTM (Long Short-Term Memory) networks are an advanced type of Recurrent Neural Networks (RNN) introduced by Hochreiter and Schmidhuber (1997), which surpass traditional RNNs (Kratzert et al., 2018). LSTMs are well-suited for processing time series data and hold great potential for improving hydrological forecasting (Zhang et al., 2018). They are seen as a reliable alternative to hydrological models for developing streamflow prediction systems (Girihagama et al., 2022; Sabzipour et al., 2023b).

Rather than relying on a single model, the multi-model approach combines forecasts from multiple models to produce a consolidated forecast, often resulting in greater accuracy. This methodology has demonstrated its potential in various fields, including hydrology, as demonstrated by Arsenault et al., (2015) on 429 watersheds, and particularly in the field of hydrological forecasting (Dion et al., 2021) where eight different hydrological models were used on five sub-basins of the Lac-Saint-Jean (LSJ) catchment in Canada. Considering the above-mentioned challenges in hydrological ensemble forecasting, this study aims at evaluating if the introduction of the LSTM within a multi-model framework improves the accuracy of short-term streamflow forecasts over the LSJ catchment. In particular, this study seeks to address if, by exploiting the maximum performance of each forecasting model, the proposed hybrid multi-model approach enhances the accuracy of streamflow forecasts. As there are no applications of hybrid multi-model forecasting that include three types of model (lumped and distributed hydrological models using DA, combined with a DL-base model), the

reliability of the hybrid multi-model approach is investigated to identify its benefits for streamflow forecasting in a Nordic region.

Section 2.5 presents the experimental design of the study. Section 2.6 depicts the key results of the hybrid multi-model approach followed by a discussion in Section 2.7. Concluding remarks are provided in Section 2.8.

2.3 Study area

This study focuses on the Lac-Saint-Jean (LSJ) catchment, located in the province of Quebec, Canada. The LSJ catchment is a vast region covering an area of approximately 45,000 km², used by the company Rio Tinto for aluminum production through hydropower generation. It comprises ten sub-catchments that ultimately feed into a central reservoir (See Figure 2.1).

The study region experiences a continental climate characterized by cold and snowy winters, which significantly influence the hydrological regime and the occurrence of spring snowmelt floods (Arsenault et al., 2013). The LSJ reservoir, spanning an area of 1,000 km², plays an important role in the regional hydropower generation (Arsenault and Côté, 2019). The production of short-term streamflow forecasts is of great importance to optimize water management, and to ensure constant monitoring and regular maintenance of the hydroelectric power plant for safety and efficiency operation in the LSJ catchment.

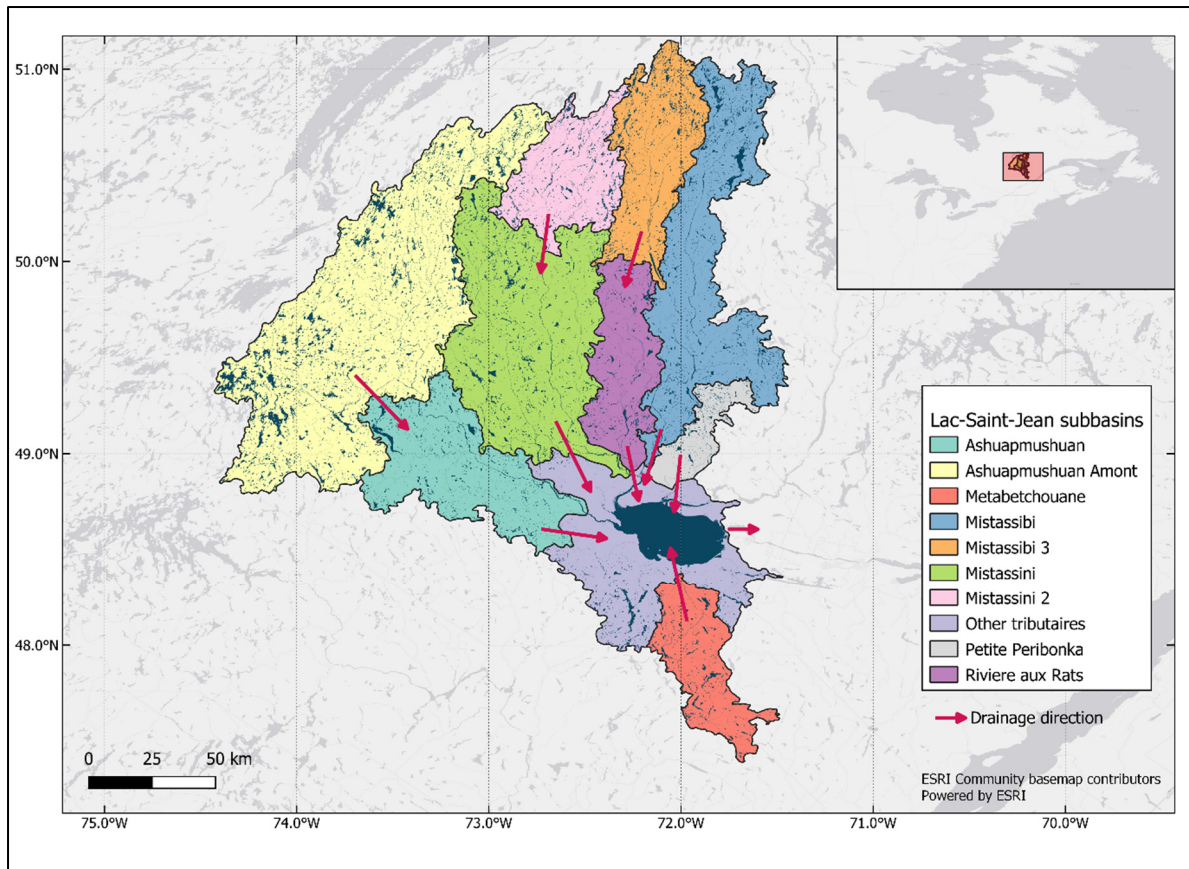


Figure 2.1 The Lac-Saint-Jean (LSJ) catchment in the province of Quebec, Canada (subset panel). The LSJ catchment is composed of ten sub-catchments contributing inflows to the reservoir, which is the target flow of this study. The “other tributaries” (light purple coloured area around the main reservoir in southern part of LSJ catchment) include all smaller streams that contribute flows directly to the reservoir without passing through major river systems. Drainage directions (red arrows) are also provided to show flow routing in the catchment. Adapted from Sabzipour et al. (2023b)

2.4 Datasets

The meteorological datasets are provided by Rio Tinto and consist of spatially interpolated grid data from 16 weather stations within the study area. Meteorological datasets include daily values of minimum and maximum temperatures, and total precipitation over the 1954-2019 period.

The hydrometric data, covering the same historical period as the meteorological data (1954 to 2019), are also supplied by Rio Tinto. However, it is worth noting that the inflow data are derived from calculations involving the reservoir level, turbines outflows, and flood spillway releases. Given the uncertainty around these data, three days rolling averages were computed to ensure that a consistent water volume was maintained at the catchment outlet.

In addition, weather forecasts came from the European Center for Medium-Range Weather Forecasts (ECMWF). This consists of 50-member perturbed forecasts of daily precipitation, minimum and maximum temperatures for 1 to 9 days of lead time for the period ranging between January 2015 and December 2019.

2.5 Methodology

This section provides an overview of the methodology used in this study. It encompasses several key elements, including a description of the hydrological models and the data assimilation scheme (2.5.1) as well as the implementation/calibration (2.5.2) and the evaluation methods (2.5.3) of the multi-model approach.

2.5.1 Multiple-models and Data Assimilation

The multi-model approach consists in using several hydrological models hoping to take advantage of the strength of each one while canceling out the situations where other models might have challenges. For this study, a range of hydrological models was chosen that differ regarding model structure, parametrization, processes, and spatial discretization. The 10 hydrological models chosen for this study are listed in Table 2.1.

The conceptual lumped models were selected from the HOOPLA library (Thiboult et al., 2019), from which eight models were chosen based on the study of Dion et al. (2021). The selected lumped hydrological models differ in terms of representation of hydrological processes (e.g., number of reservoirs and flow routing), but they all use the same evapotranspiration calculation formula (the Oudin PET formulation; Oudin et al., 2005) and

snowmelt method (CEMANEIGE; Valéry et al., 2014). All these models were used with the EnKF DA to ensure proper initial states for the forecasting step. Their inclusion in this study facilitates the exploration and comparison of modeling approaches towards improving streamflow forecasts.

CEQUEAU7, the operational model of Rio Tinto, was chosen to illustrate the semi-distributed category of hydrological models. CEQUEAU7 incorporates three conceptual reservoirs to represent interactions among surface water, soil moisture, and groundwater. Evapotranspiration is estimated with the Oudin method (Oudin et al., 2005) and snow processes are simulated with the CEMANEIGE model (Valéry et al., 2014). Here, CEQUEAU7 was used to generate streamflow forecasts over a predefined period, with a comparative analysis of forecasts obtained with a DA scheme based on the EnKF (CEQUEAU7_DA) and without any DA scheme (the Open Loop mode; CEQUEAU7_OL). The calibration of the CEQUEAU7 model was performed by Rio Tinto over the 1954-2014 period. Parameters for CEQUEAU7 are shared among all discretized units. For the Lac-Saint-Jean catchment, there are 28 hydrological units that share the same parameters which are more or less active depending on the land use and other catchment characteristics.

The LSTM model was selected to represent the DL-based category. A LSTM consists of a memory cell with a neuron with a self-recurrent connection and three nonlinear gates to control the flow of information into and out of the cell. The forget gate within the memory cell allows the LSTM to retain information over extended periods, enabling them to learn long-term dependencies that prove challenging for other RNNs to capture. Consequently, LSTMs are adept at overcoming issues like short-term memory limitations which make them excellent candidates for hydrological modelling and forecasting (Xu et al., 2020).

Tableau 2.1 Number of parameters for each hydrological model used in this study (Adapted from Dion et al., 2021)

Hydrological models	Spatial discretization	Number of parameters	Reference
CEQUEAU	Lumped	9	Girard et al. (1972)
GR4J	Lumped	4	Perrin et al. (2003)
HBV	Lumped	9	Bergström and Forsman (1973)
HYMOD	Lumped	6	Wagener et al. (2001)
IHACRES	Lumped	7	Jakeman et al. (1990)
MOHYSE	Lumped	7	Fortin and Turcotte (2007)
SIMHYD	Lumped	8	Chiew et al. (2002)
TOPMODEL	Lumped	7	Beven et al. (1984)
CEQUEAU7	Semi-Distributed	23	St-Hilaire et al. (2015)
LSTM	DL-based	-	Hochreiter and Schmidhuber (1997)

It should be noted that the number of streamflow simulations is different between models. Due to the high computational requirements of CEQUEAU7, the model provides streamflow forecasts (nine days forecast) every three days (day 1, 4, 7...), and to make a fair comparison, only the same three days interval was kept by the other models. For the eight lumped models, each streamflow forecast generation involves using 50-member weather forecast and 25 initial condition members with the EnKF DA approach, so 1250 simulations are generated for each of the nine forecast days. For CEQUEAU7, only the average of the forecasts generated by the 25 DA members was retained. Finally, only the 50-member weather forecasts are used for the LSTM.

2.5.2 Model calibration

The set of hydrological models is calibrated over a 66-year period (1954-2019) and then applied to generate streamflow forecasts over a lead time of a 9 days. The model performance is evaluated using the modified Kling-Gupta Efficiency (KGE) objective function (Gupta et al., 2009; Kling et al., 2012) defined by:

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

where r is the linear correlation coefficient, α is the ratio of coefficients of variation, and β is the bias ratio. All these parameters are dimensionless and are calculated based on the comparison between observed and simulated flows. The KGE value ranges from $-\infty$ to 1, with a perfect fit between observed and simulated flows resulting in a KGE value of 1.

The calibration of lumped hydrological models is performed using the Shuffled Complex Evolution – University of Arizona (SCE-UA) algorithm (Duan et al., 1992), which was shown to be robust and efficient for model calibration (Arsenault et al., 2014). Only one calibration period is used, which includes all available data to maximize the information content, as recommended in Arsenault et al. (2018). The model performance is then assessed during the forecasting period. The semi-distributed CEQUEAU7 model was calibrated manually by Rio Tinto hydrologists using a functional data analysis implementation (Larabi et al. 2018). The deep-learning based LSTM model is trained through a training-validation-testing approach based on the Adam optimizer and specific hyperparameter tuning (Sabzipour et al., 2023b).

2.5.3 Model forecast performance and evaluation criteria

It is often advantageous to assess the performance of hydrological forecasting models on a seasonal basis, as the efficacy of the models can vary depending on the season. For instance, models may encounter greater challenges when forecasting spring floods because of snowmelt and the rapid changing weather conditions. By conducting a performance evaluation of the models by season, more insights into their strengths and weaknesses are get over a specific period which allows a more accurate adjustment. For instance, if a model demonstrates a relatively weak performance during the spring flood season, this might indicate the need to adjust parameters or modeling techniques to capture the dynamics of snowpack processes more precisely. The evaluation of model performance is done over the 2015-2019 period where the 50-member ECMWF forecasts for temperature and precipitation are available.

To ensure a consistent basis for data analysis, a data processing step is implemented to categorize data into four groups where, each group is corresponding to a specific season. The seasons are selected to match with the hydro-climatic patterns of the LSJ catchment: winter (December to March), spring (April and May), summer (June to August), and fall (September to November).

The performance of the forecasting models is evaluated for three characteristics of the forecasts. Firstly, forecast accuracy which assesses the closeness of the forecasts to the actual observed outcomes, ensuring that forecasts are consistently aligned with observational values. Second, forecast sharpness which depicts the concentration of the forecasted distributions. A sharp forecast indicates a high degree of confidence expressed by narrow predictive intervals. However, the value of sharpness depends on the forecast accuracy, thus guaranteeing the high confidence of the model. Lastly, forecast reliability, which ensures that the dispersion of the forecast ensemble represents the real probabilities of occurrence of the streamflow events.

The evaluation of those characteristics of the streamflow forecasts is conducted with three metrics: the Continuous Ranked Probability Score (CRPS; Hersbach, 2000), which reflects accuracy and sharpness, the Mean Absolute Error (MAE; Mather and Johnson, 2016), targeting accuracy, and Talagrand diagrams for forecast reliability. The CRPS quantifies the average distance between the observed ($F(q_{obs})$) and the simulated ($F(q_{sim})$) probability density function generated by the forecast ensemble as:

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

Where N represents the size of the forecast ensemble. CRPS values range from 0 to $+\infty$, with 0 indicating a perfect forecast. The CRPS has the same unit as the streamflow variable (q), i.e., m^3/s .

The MAE quantifies the difference between observed and forecasted outcomes within a deterministic multi-model framework. The MAE is equivalent to the CRPS when one member is considered, making thus the CRPS an extension of the MAE to ensembles. The MAE calculation involves identifying the mean absolute errors expressed as:

$$\text{MAE}(t) = \frac{1}{N} \sum_{i=1}^N |q_{sim_i}(t) - q_{obs}(t)| \quad (2.3)$$

where q_{obs} is the observed streamflow, $q_{sim,i}$ is the simulated streamflow for the i -th member of the ensemble of size N at time t . A MAE(t) value of zero is considered as optimal.

CRPS and MAE are computed for each forecast day and each lead time, and their distributions are presented for each of the four seasons (winter, spring, summer, and fall) to provide a comprehensive assessment of the model's performance throughout the year. To conduct a comprehensive comparison between forecasts and observations, Talagrand diagrams are generated for each individual model and model combinations. The Talagrand diagram offers an overall view of forecast quality in terms of forecast reliability. They are constructed by computing the percentile (or rank) of the observation within the forecasted ensemble for each forecast date, and then verifying the histogram of the distribution of these percentiles. The shapes of the Talagrand diagram gives insights into the reliability of ensemble forecasts between an under-dispersion (u-shape) and an over-dispersion (n-shape).

In addition, the Kolmogorov-Smirnov (K-S) test is employed to reject (or accept) the hypothesis of uniformity ($\alpha = 5\%$). The K-S test is a non-parametric method used to compare the distributions of two samples or a sample with a reference distribution. It provides a way to test if data follow a particular distribution or if two datasets come from the same underlying distribution.

The Average Bin Distance to Uniformity (ABDU) is used to quantify the deviation of the rank histogram from perfect uniformity (Arsenault et al. 2016). ABDU provides a concise way to evaluate the reliability of ensemble forecasts. The ABDU is expressed as:

$$ABDU = \frac{1}{N} \sum_{k=1}^N \left| S_k - \frac{M}{N} \right| \quad (2.4)$$

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. A perfect uniformity would have an ABDU of zero. However, ABDU is not suitable to compare different seasons due to its sensitivity to the total sample size (M) in the Talagrand diagram and the varying lengths of the seasons (winter ~ 120 days, spring ~ 60 days, summer and fall ~ 90 days). In contrast, ABDU provides a numerical measure of the distribution of two ensembles, determining which ensemble is closer to uniformity for a specific season.

2.6 Results

2.6.1 Model calibration

This section presents a summary of calibration results of the hydrological models as provided in Dion et al. (2021) and LSTM model as provided in Sabzipour et al. (2023b). It is important to note that all models use the same forecasting period of 2015-2019 inclusively, but calibration and validation differ between models. The 8 lumped models were calibrated over the entire 1954-2019 period, and CEQUEAU7 was calibrated over the 1954-2014 period, using all available data at the time of calibration as per recommendations of Arsenault et al. (2018). The LSTM was calibrated over the 1979-2014 period, but the method used is different from other models. By using a random selection, 70% of the years were used for training, 15% for the validation and the remaining 15% for independent testing. The nine hydrological models perform well during the calibration with KGE values up to 0.94 (Table 2.2) using the historic meteorologic forcings and deterministic model runs. On the other hand, individual LSTMs are built for each lead time and then trained and validated for that lead-time. By cumulating these LSTM responses one day at a time, complete hydrographs can be generated for the training, validation and testing periods. The training period is used to train the model weights, and the validation period is used to evaluate the model weights during training and to stop training

when validation scores stop improving (calibration stopping criteria). The final model setup is then tested using an independent portion of data never used for training/validation. The LSTM performance over the training period is very good with KGE values above 0.90 for the 9-day lead time (Tables 2.3). However, for lead-times day nine, the LSTM performance over the testing period is slightly lower, even though the KGE values remain very good ($KGE \geq 0.88$).

Tableau 2.2 KGE scores for the nine hydrological models for the calibration on the LSJ watershed (Adapted from Dion et al., 2021)

Hydrological models	Calibration KGE
CEQUEAU	0.93
GR4J	0.94
HBV	0.94
HYMOD	0.94
IHACRES	0.92
MOHYSE	0.94
SIMHYD	0.92
TOPMODEL	0.93
CEQUEAU7	0.94

Tableau 2.3 KGE scores for the LSTM model calibrated on the LSJ catchment (Adapted from Sabzipour et al., 2023b) and simulated with lead-times from 1 to 9 days

Lead times (days)	Nb. of LSTM units	Dropout rate	Nb. of training epochs	Batch size	Training KGE	Validation KGE	Testing KGE
1	128	0.2	100	128	0.97	0.99	0.98
2	128	0.2	200	128	0.98	0.99	0.98
3	128	0.2	200	128	0.95	0.95	0.95
4	128	0.2	250	128	0.95	0.91	0.94
5	128	0.2	250	128	0.92	0.94	0.91
6	128	0.2	300	128	0.93	0.91	0.92
7	64	0.1	1000	64	0.94	0.94	0.93
8	64	0.1	1000	64	0.95	0.93	0.92
9	64	0.1	1000	64	0.90	0.92	0.88

2.6.2 Model comparison in forecasting

The focus of this section is to compare the performance of the ten models at simulating streamflow over the LSJ catchment at the seasonal scales during the forecasting period; the semi-distributed CEQUEAU7 model is evaluated twice—once in its open-loop setup and once with Data Assimilation.

The skill of streamflow ensemble forecasts for one to nine days of lead time is presented in Figure 2.2 through the CRPS values obtained by model and by season. The average CRPS values for each lead-time are indicated above each boxplot. The results show that a great variation among the models in winter (Figure 2.2a). LSTM and CEQUEAU7_DA give the lowest CRPS values, with consistent performance throughout the 9-day winter forecasts.

The LSTM outperforms the other models for the lead time up to three days in spring and fall, and for the lead time up to four days in summer, followed by GR4J and TOPMODEL (Figures 2.2b to 2.2d). After the 4-day forecast, a decreasing performance of the LSTM is seen with wider spread in the CRPS as lead-times increases for the three seasons. Overall, GR4J provides the best performance in spring across the entire lead time, compared to other models. GR4J CRPS values averaging 477 for day one to 506 for day 9, compared to LSTM who start with a very good average of 64 for day one, but end with an average value of 753 for day nine, which is almost 12 times increase. In summer, after the 4-day forecast, the LSTM shows comparable performance to other models such as CEQUEAU, GR4J, MOHYSE, SIMHYD, and TOPMODEL. Even though some discrepancies between models are seen in fall for the first six days of the forecast, a similar trend is observed for the last three days with close CRPS spread. The results in Figure 2.3 indicate the added value of the LSTM model to the overall quality of ensemble forecasts in the four seasons for the 9-day lead-times. This best model to forecast streamflow, with a lower CRPS value for more than half (19/36) of the cases, is considered for further analysis in the next section.

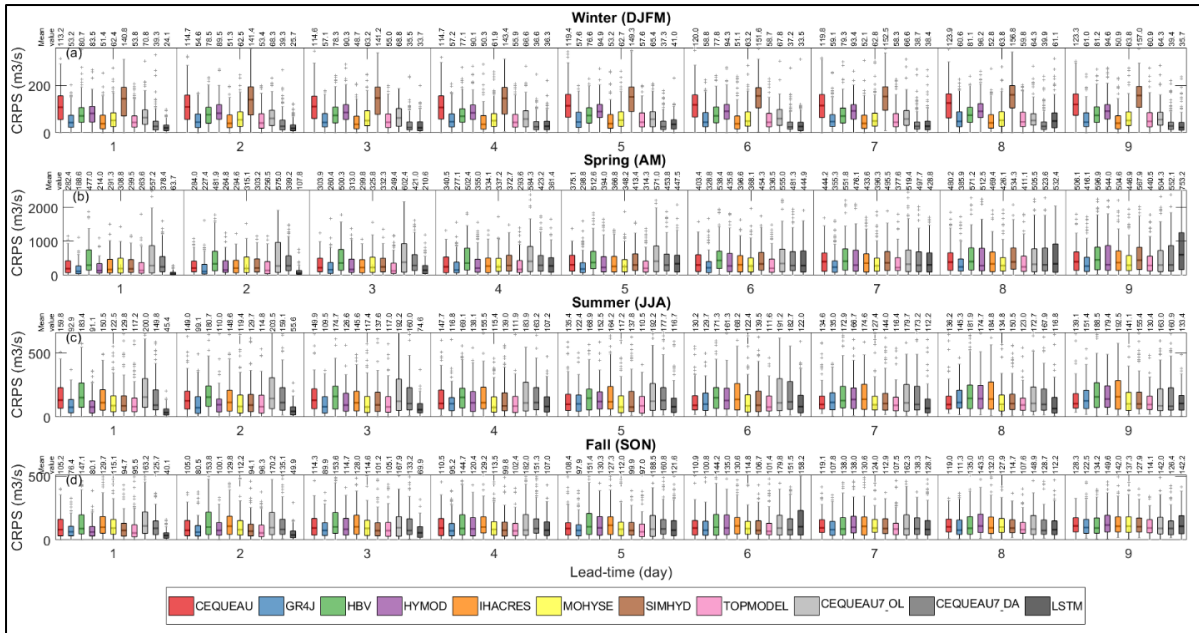


Figure 2.2 CRPS of the streamflow ensemble forecasts for the LSJ catchment by season for all models over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot

2.6.3 Two model combination comparison in forecasting

To further examine the influence of model combination on the streamflow forecast skill, all model combinations are evaluated for each season using the LSTM as the optimal model as seen in the previous section. The ensemble forecasts generated with the various multi-model combinations are compared to the ensemble forecasts generated with the individual LSTM. The results reveal a pronounced variation in the CRPS values in winter compared to other seasons (Figure 2.3a). There are three combinations that stand out a little more than the others during winter: LSTM-GR4J (green), the LSTM-TOPMODEL (grey), and the LSTM-CEQUEAU7_DA (dark grey) with average CRPS value for day one ranging from 23.2, 23.0 and 22.7 respectively and 35.5, 35.4 and 28.8 for day nine. The LSTM-CEQUEAU7_DA model combination (in dark grey) produced the lowest CRPS for each of the 9 forecast days. The LSTM-TOPMODEL combination (in grey) also performed well in winter, achieving the second-lowest average CRPS on five out of nine days.

In spring, the overall skill of the ensemble forecasts is improved by model combination after the 2-day lead time (Figure 2.3b). For days 1 and 2, the standalone LSTM model (in red) yielded the best results with average CRPS values of 64 and 107, respectively. Close behind, the LSTM-GR4J combination (in green) produced the second lowest CRPS values for days 1-3. Then, for days 4-5, LSTM-CEQUEAU7_DA (in dark gray) had the lowest CRPS, followed by LSTM-GR4J (in green) in second place. For days 6-9, LSTM-GR4J (in green) led with the best results, with average CRPS values ranging from 295 to 500. The LSTM-GR4J (green) combination provides ensemble forecasts with better accuracy (lower spread and score of the CRPS) than the other multi-model combinations and the individual LSTM for days 3-9 of the lead time.

In summer, the model combinations have comparable performance to that of the individual LSTM results for the first three days of lead time (Figure 2.3c). However, using a combination of models provides an increase summer forecast skill from day 4. For day 1, the standalone LSTM model (in red) again produced the best result with an average CRPS of 45. However, for the rest of the hydrological forecasts, the LSTM-GR4J (in green), LSTM-TOPMODEL (in gray), and LSTM-CEQUEAU7_DA (in dark gray) combinations shared the best outcomes. They achieved the lowest CRPS values for day 2, days 3, 5, and 6, and days 4 and 9, respectively.

Similarly to summer, considering a combination of models in fall does lead to large gains in performance, as the ensemble forecasts generated by the LSTM were already of good quality, but similar performance to the other individual models seen in Figure 2.3d. However, choosing to combine models in fall from the day one of the lead time does have a clear impact on the accuracy of the ensemble forecasts which are improved against the LSTM. Nevertheless, the same combinations as in summer performed best also in fall. The LSTM-GR4J (in green), LSTM-TOPMODEL (in gray), and LSTM-CEQUEAU7_DA (in dark gray) combinations achieved their lowest CRPS values on days 1-3, days 4-7, and days 8-9, respectively.

According to these results, the model combination to select to get the best performance of the ensemble forecasts with a low CRPS for the four seasons is the LSTM-GR4J combination. This combination has results very close to the LSTM-TOPMODEL and LSTM-CEQUEAU7_DA combinations, sometimes equivalent or even a slightly higher CRPS for a few forecast days; however, the LSTM-GR4J combination provides the best results for the spring season, which is the most important season for the Nordic study region. This LSTM-GR4J model combination in the multi-model approach is considered for further analysis in the next section.

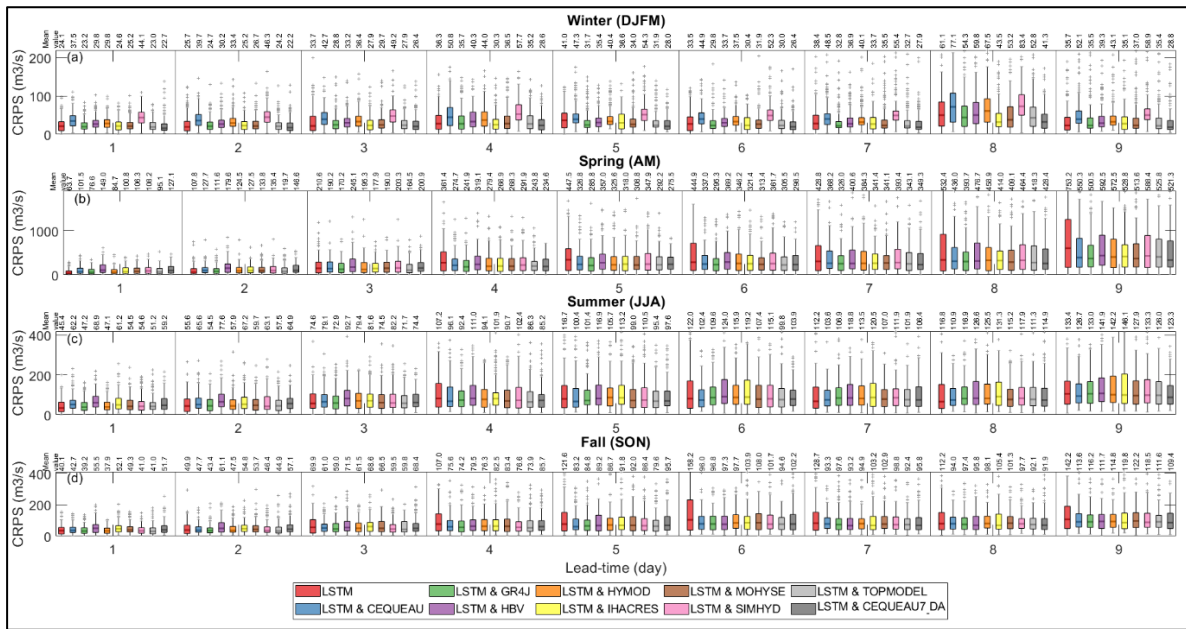


Figure 2.3 Results of two-model combination: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the LSTM alone (red) and when combined with other models over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot

2.6.4 Three-model combination comparison in forecasting

Using the previous best model combination (LSTM-GR4J), we can then identify a second level of model combination (LSTM-GR4J-X) to gain as much forecast skill as possible. The

objective is to assess if the multi-model approach defined as the combination of 3 dissimilar models, i.e. DL (LSTM) – lumped (GR4J) – distributed (CEQUEAU7_DA), can surpass the performance of the individual models or combinations of only two types of models.

Figure 2.4 presents the skill of streamflow ensemble forecasts generated by the multiple model combinations. We show that the ensemble forecasts of streamflow are clearly improved with the LSTM-GR4J-CEQUEAU7_DA combination with lower CRPS values across the four seasons compared to other combinations. In winter, this combination manages to produce the lowest CRPS value for nine days in row, it's not surprising given the good results obtained previously by CEQUEAU7_DA. LSTM-GR4J-CEQUEAU7_DA CRPS averaging 22.8 for day one to 32.6 for day 9 when the second best average between 24.3 and 33.5 for LSTM-GR4J-IHACRES (Figure 2.4a).

An exception is observed for days 1 and 2 in spring, where the individual LSTM model (in red) produces the lowest average CRPS value with 64 and 108 respectively, compared with combination LSTM-GR4J-CEQUEAU7_DA which produced an average CRPS value of 99 and 121 (Figure 2.4b). The same behavior is observed for days 1 in summer, where the LSTM (in red) produces the lowest average CRPS value with 45, compared with 47 for LSTM-GR4J and 50 for LSTM-GR4J-CEQUEAU7_DA (Figure 2.4c). This can be attributed to the setup of the LSTM model, as it focuses on prior and upcoming weather conditions without accounting for the initial states of the catchment. This prevents the LSTM from gathering errors for consecutive forecast days. However, this does not prevent from potential deviations over longer forecasting periods.

In fall it is slightly more difficult to determine which combination produces a lower CRPS, as the results are very similar. LSTM-GR4J achieved a lower CRPS for the initial two days of forecasting, whereas LSTM-GR4J-CEQUEAU7_DA demonstrated a lower CRPS during the final three days of the forecast. For the intervening period, the performance of all models was remarkably similar (except for LSTM alone which is worse), with a negligible difference of less

than 5% in average CRPS values. From here onwards the combination LSTM-GR4J-CEQUEAU7_DA will be denoted as Best 3.

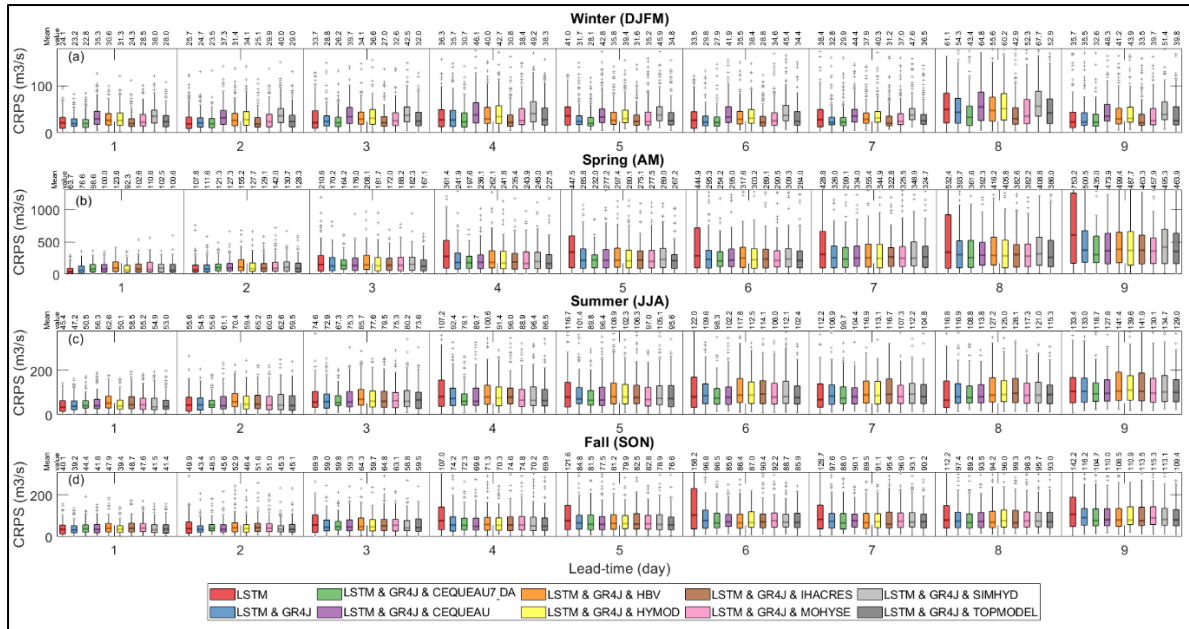


Figure 2.4 Results of three-model combinations: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the LSTM alone (red), the LSTM-GR4J (blue), and the LSTM-GR4J combined with other models over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot

2.6.5 Four-model combination comparison in forecasting

From the evaluation of the LSTM-GR4J-CEQUEAU7_DA combination, other model combinations can be further explored in view of optimizing streamflow ensemble forecasts. In the following, a lumped model is added to the best 3-model combination to build a combination of four models. This is presented in Figure 2.5 with the CRPS values for each season over the 9-day lead time.

In winter, for day 1-2 and 4-5 there is no advantage to add a fourth model, but the Best 3 and IHACRES model combination (in orange) consistently exhibits a lower CRPS value compared

to the other combinations for day 3 and 6-9. On day eight this combination produces a CRPS 12% lower than the best 3-model, but it managed to produce a CRPS only 1% lower, or less, for the remaining four days (Figure 2.5a). In the spring, there is no advantage in adding a fourth hydrological model for days 1 and 2 of the forecast period. For days four to seven, incorporating a fourth model, in this case TOPMODEL (in pink), offers a marginal benefit by reducing the CRPS by less than 1%. However, for days three and eight to nine, the addition of the fourth model (IHACRES (in orange), MOHYSE (in yellow) or TOPMODEL (in pink)) results in a more significant CRPS reduction, ranging between 3 and 5% (Figure 2.5b). For the summer and fall seasons, there is no significant advantage in adding a fourth hydrological model. Nonetheless, there is a slight improvement in CRPS, with a reduction of 5% for days 5-7 in the fall (Figure 2.5d), and a reduction of less than 1% in 72% of the cases. However, these improvements are not as significant as those observed previously in sections 3.3 and 3.4. We show that the benefit of incorporating an additional model into the top 3-model combination is not substantial enough to justify its continued use, even though the results do not deteriorate.

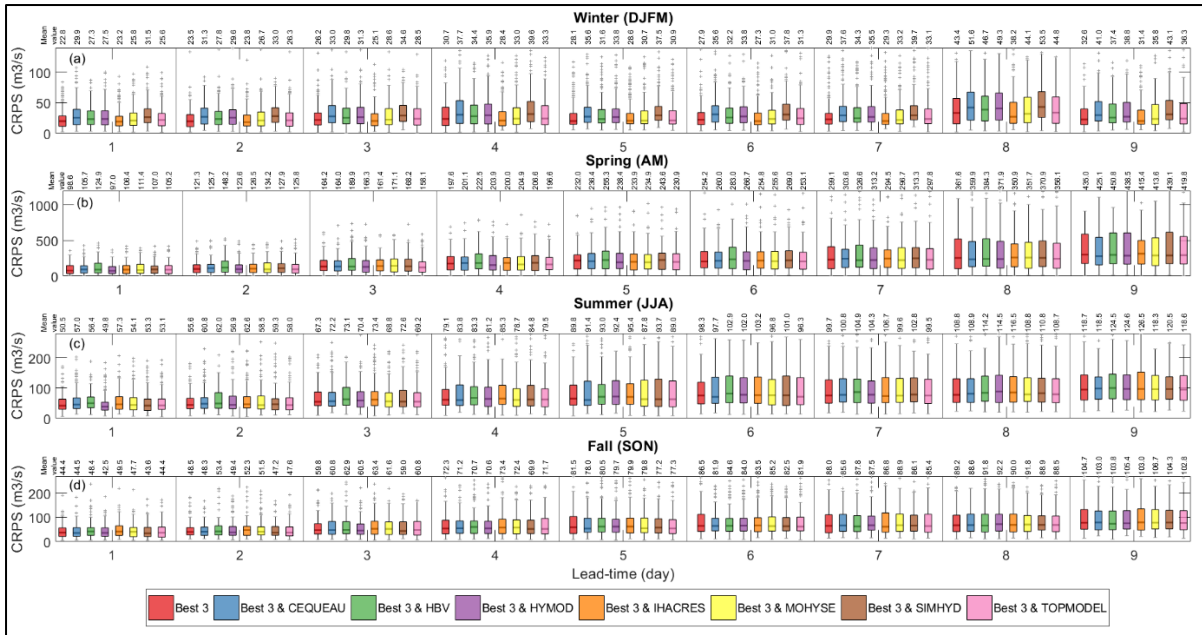


Figure 2.5 Results of four-model combinations: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the Best 3 (i.e. LSTM-GR4J-CEQUEAU7_DA) model combination (red), the Best 3 and CEQUEAU model (blue), the Best 3 and HBV model (green), the Best 3 and HYMOD model (purple), the Best 3 and IHACRES model (orange), the Best 3 and MOHYSE model (yellow), the Best 3 and SIMHYD model (brown), and the Best 3 and TOPMODEL model (pink) over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot

2.6.6 Effect of incremental multi-model ensemble size

Figure 2.6 and 2.7 summarizes the CRPS and MAE values obtained with the best performing one-, two-, and three-model combinations. These results indicate that the most effective combination involves pairing the LSTM with the GR4J and CEQUEAU7_DA models. However, the introduction of an additional (fourth) hydrological model into this combination does not significantly improve the skill of the streamflow ensemble forecasts as demonstrated in the previous section. In this study, the similarity in outcomes from CRPS and MAE (Figures 2.6 vs 2.7) suggests that our streamflow ensemble forecasts are generally close to the observed values, with a narrow distribution. This indicates that the forecasts are neither systematically overestimating nor underestimating the observed data. Such a pattern in the forecasts leads to

a situation where both the probabilistic measure (CRPS) and the average error measure (MAE) reflect similar levels of forecast reliability and accuracy.

In winter, the CRPS goes down by an average of 18%, with reductions ranging from 5% to as much as 32% on different days. In spring, the average drop in CRPS is even bigger at 22%, with some days seeing a huge decrease of up to 48% (day five Figure 2.6b). During the summer, the average reduction is 11%, with certain days reaching up to 26%. In fall, there's an average CRPS decrease of 22%, with day-to-day changes between 3% and a high of 45%. Overall, the multi-model approach consistently yields lower CRPS and MAE values than those obtained by the individual LSTM model (in red) for the four seasons, with the exception for days 1 and 2 of the lead time (Figures 2.6b and 2.7b). This shows the importance of a suitable selection of models to be combined in view of achieving skillfully short-term streamflow forecasts.

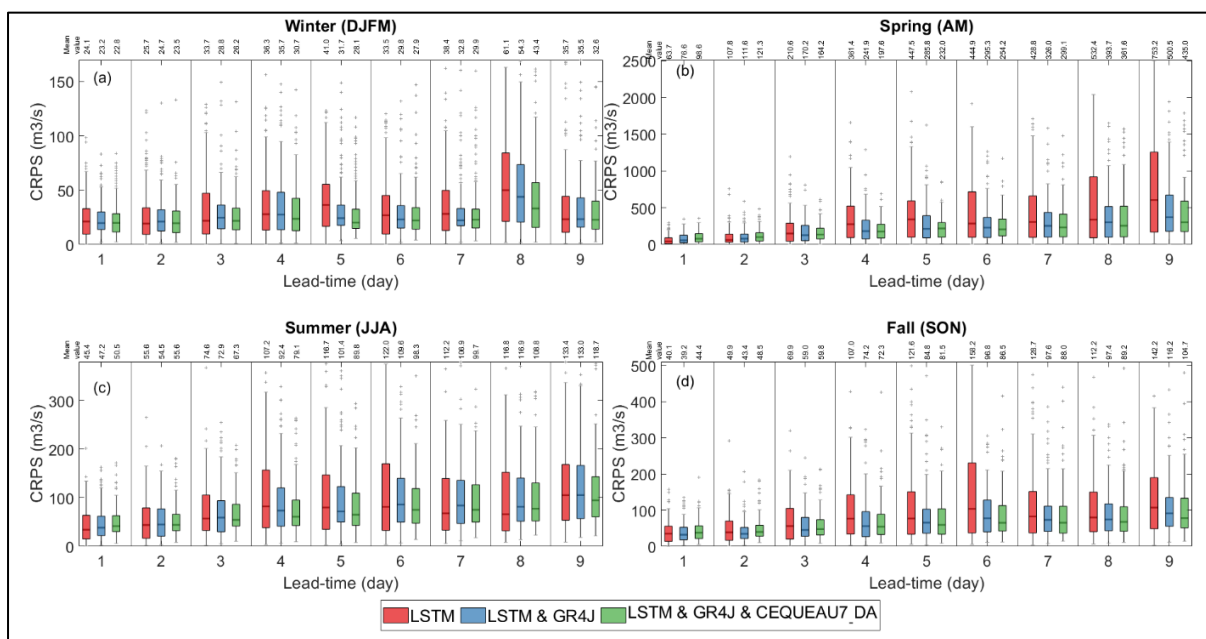


Figure 2.6 Comparison of best one-, two-, and three-model ensembles: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the LSTM (red), the combination of LSTM-GR4J (blue) and LSTM-GR4J-CEQUEAU7_DA (green) over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot

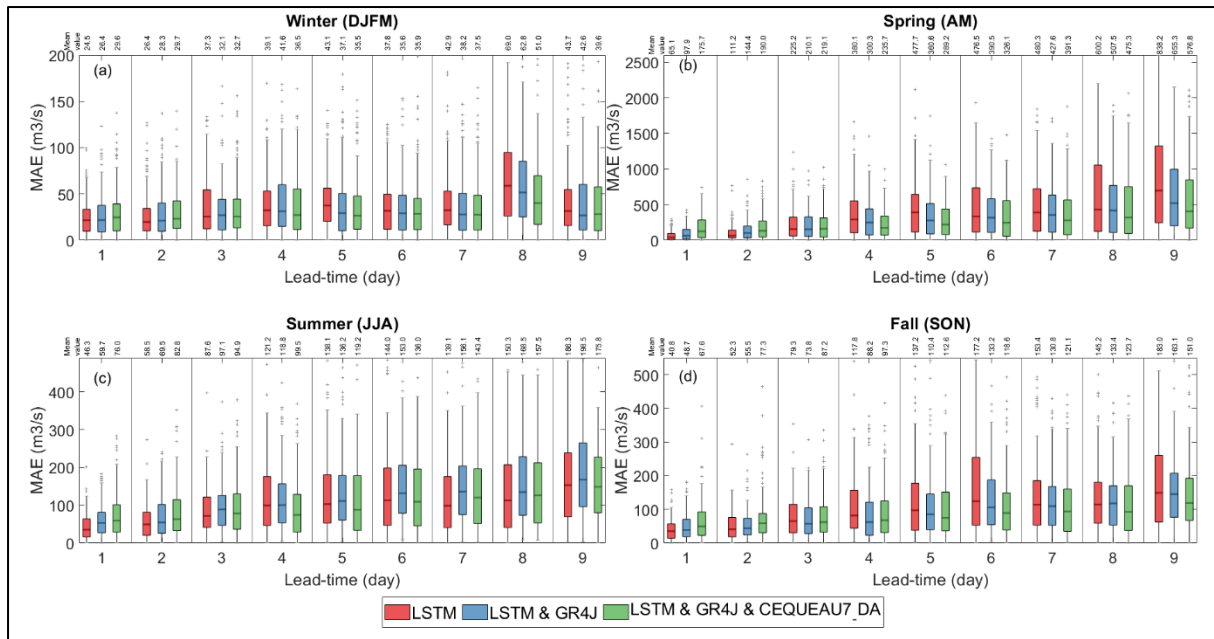


Figure 2.7 Comparison of best one-, two-, and three-model ensembles: MAE of streamflow ensemble forecasts for the LSJ catchment by season for the LSTM (red), the combination of LSTM-GR4J (blue) and LSTM-GR4J-CEQUEAU7_DA (green) over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average MAE values for each lead-time are indicated above each boxplot

2.6.7 Validation of the multi-model approach

To validate the performance of the hybrid multi-model combination (DL, lumped, and distributed models), an additional test is made to compare Best 3 (LSTM-GR4J-CEQUEAU7_DA) with two combinations of three models excluding the LSTM model to assess its impact. The first combination consists of three conceptual lumped models (GR4J, HYMOD, and TOPMODEL) and the second combination consists of two lumped and one semi-distributed models (GR4J, TOPMODEL, and CEQUEAU7_DA). Figure 2.8 shows the added value of the hybrid multi-model combination to the overall quality of streamflow ensemble forecasts particularly in winter for lead time up to nine days, which surpasses the other two multi-model combinations.

In the winter season, the LSTM-GR4J-CEQUEAU7_DA (in red) impressively achieves an average CRPS 45% lower than GR4J-HYMOD-TOPMODEL (in blue) and outperforms the

GR4J-TOPMODEL-CEQUEAU7_DA model (in green) by a 25% difference (Figure 2.8a). In spring, the LSTM-GR4J-CEQUEAU7_DA combination (in red) achieves an average CRPS 14% lower than GR4J-HYMOD-TOPMODEL (in blue). It maintains a narrow lead with just 1% lower CRPS compared to GR4J-TOPMODEL-CEQUEAU7_DA (in green), which shows stronger results on days 8 and 9 (Figure 2.8b). During the summer, the LSTM-GR4J-CEQUEAU7_DA model (in red) again leads the way by producing an average CRPS 20% lower than GR4J-HYMOD-TOPMODEL (in blue), and surpasses the GR4J-TOPMODEL-CEQUEAU7_DA (in green) by 10% (Figure 2.8c). In fall, it continues with LSTM-GR4J-CEQUEAU7_DA (in red) achieving an average CRPS 13% lower than GR4J-HYMOD-TOPMODEL (in blue) and 9% lower than GR4J-TOPMODEL-CEQUEAU7_DA (in green) (Figure 2.8d). This consistent pattern across the seasons highlights the robustness and superior performance of the hybrid multi-model combination (DL, lumped, and distributed models).

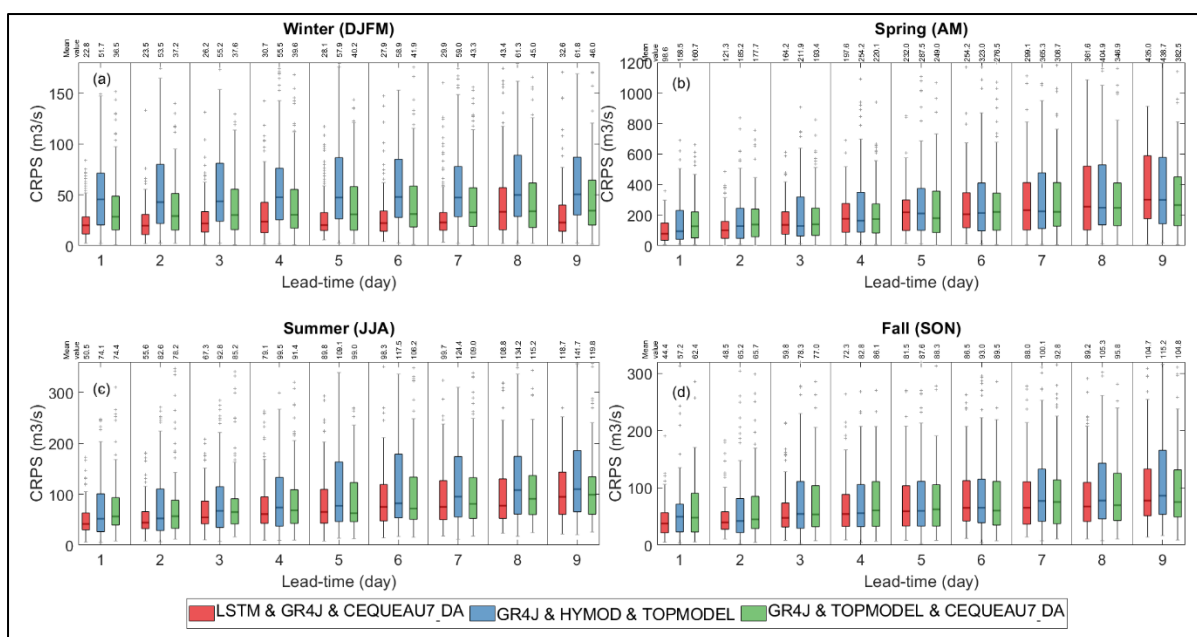


Figure 2.8 Hybrid vs purely conceptual hydrologic model ensembles: CRPS of streamflow ensemble forecasts for the LSJ catchment by season for the combination of LSTM-GR4J-CEQUEAU7_DA (red), GR4J-HYMOD-TOPMODEL (blue) and GR4J-TOPMODEL-CEQUEAU7_DA (green) over the 2015-2019 forecasting period: winter (DJFM; a), spring (AM; b), summer (JJA; c), and fall (SON; d). Average CRPS values for each lead-time are indicated above each boxplot

2.6.8 Reliability of the multi-model approach

Figure 2.9 present the Talagrand diagrams for the 5-day streamflow ensemble forecasts by the 10 individual hydrological models on the LSJ catchment. In an ideal ensemble every percentile in the ensemble (x-axis) should have the same likelihood (y-axis). A reliable forecast would hence manifest in a flat distribution (same count in every bin).

In winter, achieving a uniform dispersion appears difficult for the models, suggesting that the complex snow-related hydrological processes struggle the models to give accurate streamflow forecasts. Only CEQUEAU7_DA does not reject the uniformity hypothesis in winter, while the other models tend to reject the uniformity hypothesis. For the other three seasons, CEQUEAU, GR4J, HYMOD, MOHYSE, SIMHYD, and TOPMODEL provide relatively flat Talagrand diagrams compared to other models. These accurate representations of distribution are not reflected by the LSTM that exhibits a high degree of under-dispersion across all four seasons. This is somewhat unexpected, considering the recognized ability of the DL models to learn and understand complex systems.

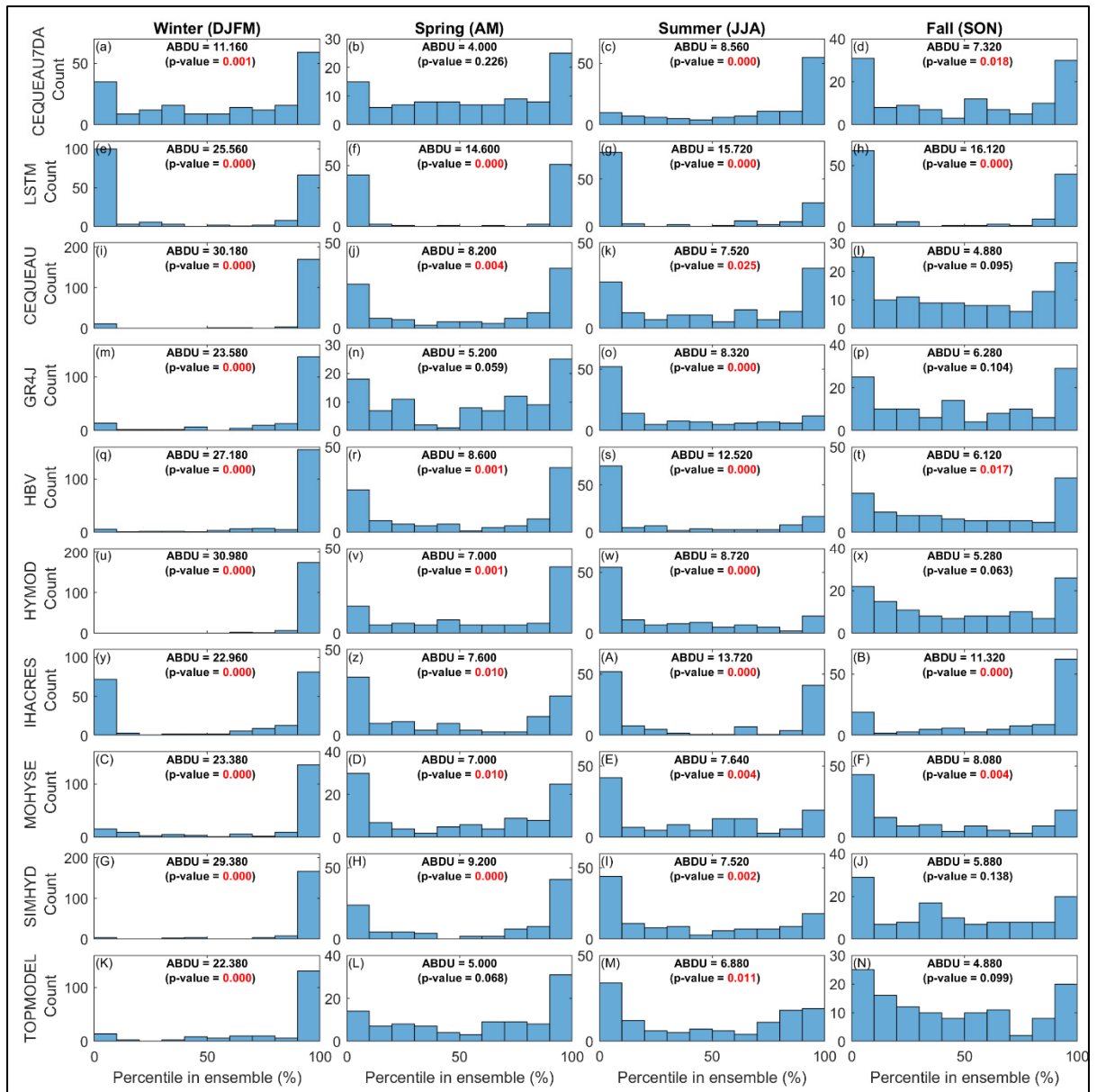


Figure 2.9 Reliability of forecasts using individual models: Talagrand diagrams representing the rank (in percentiles) of the 5-day streamflow forecasts by the 10 hydrological models on the LSJ catchment. A reliable forecast is expected to lead to flat distributions (similar count in each bin). A p-value less than or equal to 0.05 indicates that the uniformity hypothesis is rejected by the K-S test (red highlighted values)

When introducing new types of models in the model combination (two- and three-model combinations), the dispersion of the streamflow ensemble forecasts is significantly better represented than for the individual models. The variability is one of the key aspects of the multi-model approach that leads to an overall improved reliability of streamflow forecasts. This implies a shift from U-shaped Talagrand diagrams towards more uniform diagrams when combining models and using multi-model ensembles. This transition can be seen in Figure 2.10 where the Talagrand diagrams go from an under-dispersion for the individual LSTM, GR4J, and CEQUEAU7_DA models to a more uniform dispersion when combined in two-model (LSTM-GR4J) and three-model (LSTM-GR4J-CEQUEAU7_DA) combinations.

The ABDU index, which measures the deviation from uniformity, of the LSTM-GR4J-CEQUEAU7_DA combination is lower than that of the individual models, except for the GR4J in summer. Indeed, the LSTM exhibits an under-dispersion with U-shaped Talagrand diagrams, while the GR4J shows a more uniform distribution. When combined, the models create a distribution with too few elements between the extreme (0th-10th and 90th-100th percentile) and central (40th-60th percentile) quartiles of the distribution (see Figure 2.10 m-p). The inclusion of the CEQUEAU7_DA model in this ensemble results in a more uniform distribution, indicating the potential of these models to be used combined (see Figure 2.10 q-t).

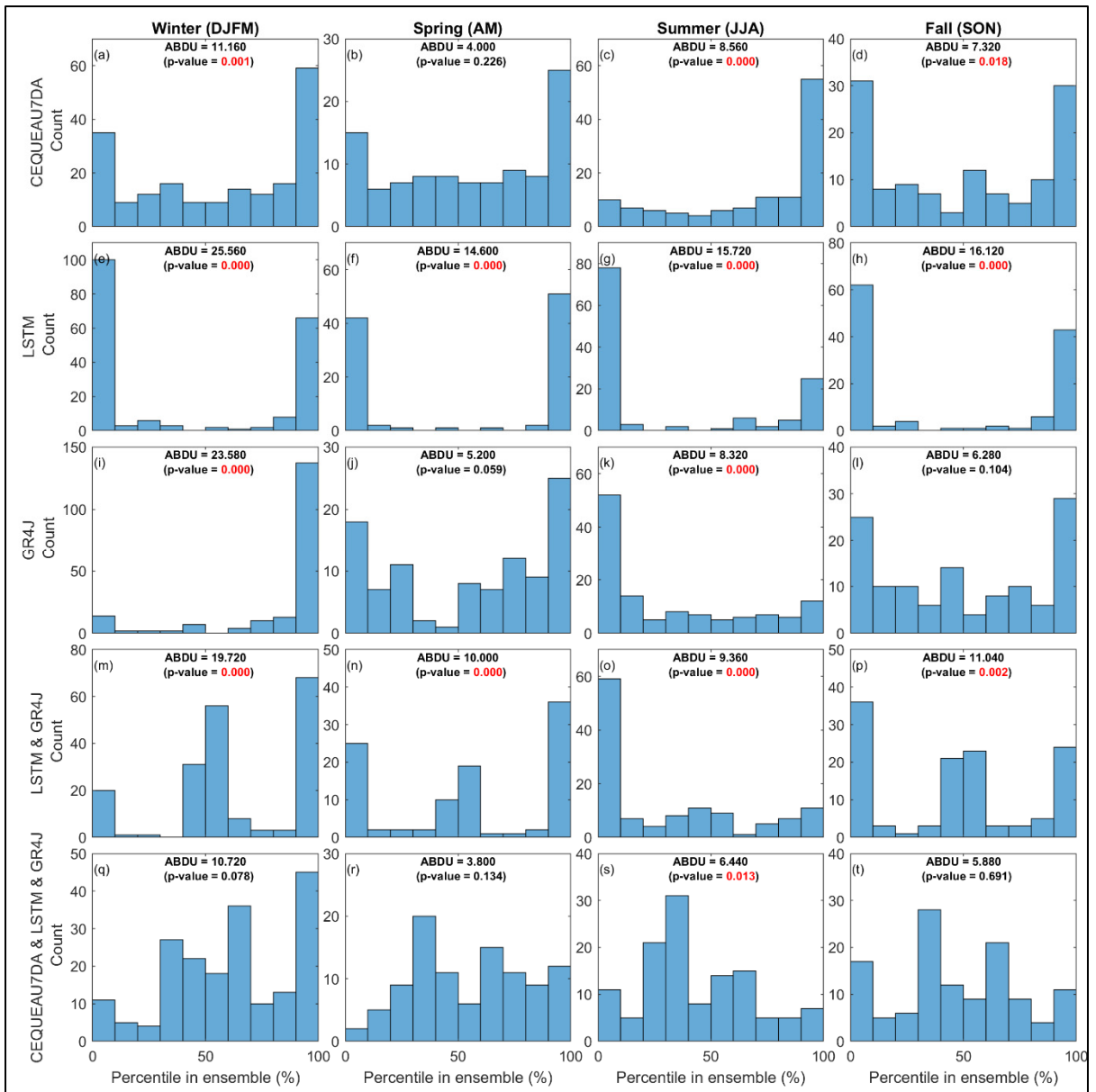


Figure 2.10 Reliability of forecasts using individual models and two- and three-model combinations: Talagrand diagrams representing the rank (in percentiles) of the 5-day streamflow forecasts for the individual CEQUEAU7_DA, LSTM, and GR4J models, as well as for the combination of the LSTM-GR4J and LSTM-GR4J-CEQUEAU7_DA models in the LSJ catchment. A reliable forecast is expected to lead to flat distributions (similar count in each bin). A p-value less than or equal to 0.05 indicates that the uniformity hypothesis is rejected by the K-S test (red highlighted values)

2.7 Discussion

The multi-model approach aims at combining different streamflow simulations generated from multiple models to create a richer streamflow ensemble forecast. Through the combination of multiple hydrological models, the strengths of each model are considered to provide more reliable and accurate forecasts of streamflow (Thiboult & Anctil, 2015). This approach improves streamflow forecasts through the minimization of predictive uncertainties stemming from the general model structure and parameter estimation. Additionally, the integration of diverse model types offers a more comprehensive representation of the inherent hydrological processes while accounting for the spatial and temporal variations of hydrological conditions. Consequently, the multi-model approach enables capturing a larger range of scenarios, enhancing the overall ability of models to address the complexity and variability within the hydrological systems (Velasquez et al. 2011). This can be seen in Figure 2.11, where the combination of models shows a wider coverage for both seasons. It can also be seen that the larger number of simulations from GR4J provides a wider spread from the start but combining it with LSTM and CEQUEAU7 forecasts still improves the forecast skill, despite their smaller representation in the ensemble.

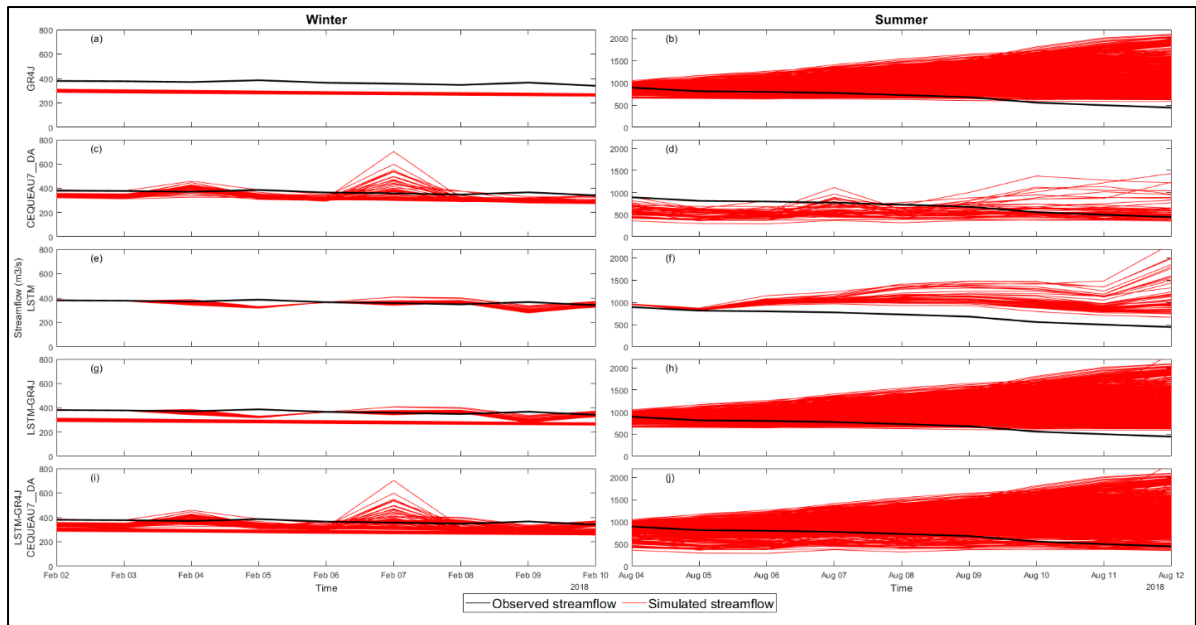


Figure 2.11 Comparison between observed and simulated streamflow over a 9-day forecasting period from the GR4J (a-b), CEQUEAU7_DA (c-d), LSTM (e-f), LSTM-GR4J (g-h) and LSTM-GR4J-CEQUEAU7_DA (i-j) models. Left hydrographs are during winter (February 2, 2018, to February 10, 2018) and right hydrographs during summer (August 4, 2018, to August 12, 2018).

The multi-model approach examined in this study attempts to identify the optimal combination and number of models to improve the accuracy of streamflow ensemble forecasts. To do so, a combination of three model types, i.e. lumped and distributed hydrological models as well as deep-learning (DL)-based models, is explored to yield enhanced streamflow ensemble forecasts. We found an improvement of the streamflow ensemble forecasts accuracy when generated with the hybrid multi-model combination (section 2.6.6). However, additional work is required to identify the clear benefits of this approach compared to alternative techniques, such as post-processing (e.g., Boucher et al., 2015). This could enhance the current results even further and will be addressed in a subsequent study.

The results show that the added value of the DL-based LSTM model when combined with other conceptual hydrologic models to obtain an overall improved quality of streamflow ensemble forecasts (section 2.6.7). This illustrates the potential of the DL models, when applied judiciously, to yield substantial benefits in hydrological forecasting. However, it also

raises a question: Are the purely DL-based models better suited to operate autonomously, or is their potential fully reached when they are combined with other model types?

The optimal number of models in the multi-model approach agrees with the recent study of Thébault et al. (2023) who shows a maximal gain in streamflow simulations with a combination of two or three models. We found that the addition of a fourth model did not lead to a significant further improvement compared to the three-model combination (see Figure 2.5). This observation suggests the existence of an optimal threshold beyond which the inclusion of additional models may not yield significant benefits. This optimal threshold might be variable according to a multitude of factors, such as the model type (i.e., lumped and distributed models or data-driven approach) or the catchment location and other properties. Further exploration on the optimal threshold according to the type of models need to be conducted for an effective combination of multiple models.

Moreover, the multi-model approach has the potential to give a more comprehensive representation of hydrological processes by combining models each with a distinct representation of hydrological processes. This ability to capture diverse scenarios and to adapt to a large range of conditions is an advantage, particularly in regions subject to rapid climate changes. The analysis of the Talagrand diagrams shows the overall robustness of the streamflow ensemble forecasts as generated by the combination of LSTM-GR4J-CEQUEAU7_DA models. Such a model combination significantly improves the accuracy of the streamflow ensemble forecasts in winter, indicating the importance of selecting and fine-tuning models to specific hydrological conditions. The good performance of the CEQUEAU7_DA model, which maintains consistency with the uniformity hypothesis while other models fail, is noticeable. This is mainly because the CEQUEAU7 model provides a better representation of the hydrological processes by considering a wide array of variables, an ability rooted in its distributed structure. This model is also more suited for longer-term forecasts in spring, where the streamflow estimates are more challenging due to the complex underlying physical processes, whereas the LSTM model might not be trained sufficiently (as seen in Figure 2.6 and 2.7).

The results obtained with the LSTM model serve as a reminder that, despite its complex algorithmic nature, computational power, or complexity, this does not necessarily translate into streamflow forecasts with a better accuracy/reliability. However, when used in combination with other models, the LSTM *strengths become a more effective approach for streamflow forecasting*. However, this study has limitations as the multi-model approach encompasses one distributed model and one DL-based model. One avenue could be to test the benefit of the inclusion of additional distributed and/or additional DL-based models in view of exploring the potential gain in performance. Expanding the experimental design to include multiple catchments with varied hydrological responses and extending the forecasting period for a longer duration would greatly enrich the study's results. This more comprehensive strategy would allow for a deeper and more diverse analysis, facilitating the generalization of results across different hydrological conditions and timescales. Such an approach would provide a more complete understanding of this methodology effectiveness and adaptability to various settings.

While recognizing the improved accuracy brought by the multi-model combinations, it is essential to emphasize that the forecast quality also relies on a deep understanding of hydrological systems and the ability at judiciously integrating data and models. Multi-model combinations present promising avenues for improving the accuracy and reliability of streamflow ensemble forecasts. However, further studies will be essential to seamlessly incorporate them into operational hydrological forecasting systems.

2.8 Conclusion and recommendations

This study provides insight into the potential effectiveness of a hybrid multi-model approach with the use of three types of hydrological models to enhance the accuracy of short-term streamflow ensemble forecasts at the seasonal scale. By combining lumped and distributed hydrological models with a deep-learning (DL)-based model, the advantages of each model were considered while minimizing their individual shortcomings. The main findings of this study are as follows:

- The LSTM DL-based model achieves good performance at forecasting streamflow at the seasonal scale.
- The LSTM provides more skillful ensemble forecasts at the seasonal scale when combined with other model types instead of using only the LSTM.
- Although several models need to be selected in multi-model combinations, it is shown that the largest improvement is achieved when combining three models with different model structure: 1) lumped and conceptual, 2) more physical and distributed, and 3) DL-based.

The findings of this study highlight the advantages, limitations, and specific evaluation of the multi-model approach in streamflow forecasting for all seasons. Overall, this study shows that the combination of a lumped and distributed models with a DL-based LSTM model is a promising multi-model approach for forecasting short-term streamflow in snow-dominated regions. While this study provides a solid basis, further research in this field is required to make the hybrid multi-model approach operational.

Overall, this study demonstrates the potential of model combinations in the field of hydrological forecasting. A judicious selection of models has significant potential to improve streamflow forecast accuracy. Future research could be dedicated to exploring the underlying mechanisms that contribute to the superiority of certain model combinations or alternative evaluation metrics for a more comprehensive validation process.

2.9 Acknowledgments

This study was partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Collaborative Research and Development grant CRDPJ-522126-17. The authors would also like to thank Rio Tinto for sharing their hydrometeorological data on the Lac-Saint-Jean catchment and the European Center for Medium-Range Weather Forecasts (ECMWF) for providing access to the historical forecast data from their MARS computing and archiving facilities. In this study, the ERA5 reanalysis dataset produced by Hersbach et al., (2018) was used. It has been downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store:

<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single->

[levels?tab=overview](https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview). The base map in Figure 2.1 was created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license.

CHAPITRE 3

DISCUSSION

Cette étude a démontré que le modèle LSTM est très performant seul et encore plus lorsqu'il est intégré dans une combinaison hybride avec d'autres modèles hydrologiques. L'efficacité du modèle LSTM peut être attribuée à sa capacité d'apprendre les relations complexes et non-linéaires entre la météo et les débits, tout en minimisant l'accumulation d'erreurs sur des jours de prévision consécutifs. Cependant, cette méthode présente des limites, surtout dans le cadre de périodes de prévision plus longues. En effet, la qualité des prévisions tend à diminuer dans ces cas, principalement à cause de la réduction de la précision des prévisions météorologiques sur lesquelles elle repose. L'utilisation combinée de modèles hydrologiques globaux, distribués et à base de DL a démontré des bénéfices significatifs. Les résultats présentés à section précédente seront discutés ainsi que le potentiel opérationnel de cette méthode sur les installations de Rio Tinto.

3.1 Analyse des performances de la méthode multi-modèle hybride

L'application de la méthode multi-modèle en hydrologie est simple, sauf lorsqu'il s'agit de choisir les modèles appropriés pour la combinaison. Cette décision ne dépend pas seulement de la structure des modèles, mais dépend aussi largement de l'expertise et de l'interprétation de l'hydrologue. Le choix des modèles à intégrer dans l'ensemble n'est pas une tâche standardisée, il nécessite une compréhension approfondie des particularités de chaque modèle, ainsi qu'une capacité à évaluer comment leurs forces et faiblesses individuelles peuvent se compléter dans le contexte spécifique de l'étude. Cette étape est très importante, car c'est l'hydrologue qui doit interpréter les résultats finaux. Il doit être capable de comprendre et d'utiliser les différentes informations fournies par chaque modèle pour arriver à des conclusions claires et utiles. Dans le cadre de cette étude, afin de simplifier le plus possible le choix des modèles, ils ont été sélectionnés selon les performances de chacun selon le CRPS et le MAE.

L'approche multi-modèle adoptée vise à exploiter les forces de différents types de modèle hydrologique pour générer des prévisions hydrologiques d'ensemble plus fiables et précises. Cette méthodologie repose sur la diversité des modèles pour diminuer le biais des prévisions. La méthode multi-modèle utilisée a déjà fait ses preuves dans le domaine de l'hydrologie et des prévisions d'ensemble (Thibault & Anctil, 2015; Arsenault et al., 2015; Arsenault & Brissette., 2016; Dion et al., 2021), mais avec l'intérêt des modèles à base de DL, une combinaison hybride multi-modèle (modèles globaux, semi-distribués et à base de DL) n'avait pas été testée. La combinaison des 3 types de modèle a permis de réduire la sous-dispersion et les biais des prévisions hydrologiques d'ensemble sur l'ensemble des saisons. Il a été constaté que l'ajout d'un modèle LSTM à d'autres modèles améliore la qualité globale des prévisions d'ensemble en diminuant significativement le biais, comme l'indiquent les résultats du CRPS. Cela démontre le potentiel considérable des modèles d'apprentissage profond qui, utilisés judicieusement, peuvent apporter des avantages significatifs à la prévision hydrologique. Toutefois, le modèle LSTM seul montre une sous-dispersion de ses prévisions d'ensemble, car elles sont plus regroupées que les autres modèles, mais légèrement au-dessus ou en dessous du débit observé. C'est pour cette raison que le modèle LSTM, une fois combiné à un autre modèle, conserve son faible biais, soit la précision de ses prévisions d'ensemble, et réduit sa sous-dispersion. L'efficacité de la combinaison des modèles LSTM-GR4J-CEQUEAU7_DA est particulièrement notable. Les diagrammes de Talagrand (Figure 2.10) illustrent la diminution de la sous-dispersion des prévisions d'ensemble, surtout en hiver.

L'analyse de la Figure 2.8 se penche sur la performance des combinaisons hybrides multi-modèles intégrant un modèle à base de DL et les modèles hydrologiques globaux et semi-distribués. La comparaison avec les autres combinaisons, soit celle des trois modèles globaux (GR4J, HYMOD et TOPMODEL) et celle des modèles globaux et semi-distribués (GR4J, TOPMODEL et CEQUEAU7_DA), révèle que le modèle hybride surpasse ces combinaisons jusqu'à un certain délai de prévision et confirme que l'intégration de modèles basés sur l'apprentissage profond peut effectivement améliorer la précision des prévisions. Cette approche hybride se distingue particulièrement en hiver pour des prévisions allant jusqu'à neuf jours, surpassant les autres combinaisons multi-modèles avec un CRPS moyen 45% plus faible

par rapport à GR4J-HYMOD-TOPMODEL. Durant les saisons printanières, le modèle hybride garde une précision supérieure, une valeur CRPS plus faible, jusqu'à sept jours. Au-delà de cette période, la performance du modèle hybride tend à rejoindre celle des deux autres combinaisons. Durant l'été et l'automne, la combinaison hybride maintient une valeur de CRPS plus faible sur les neuf jours. Des recherches futures pourraient se concentrer sur l'identification des facteurs contribuant au rendement décroissant des performances des modèles sur les longs délais de prévision et chercher des moyens de maintenir ou même d'améliorer la capacité prédictive sur des périodes prolongées. Il serait intéressant de poursuivre les travaux pour discerner clairement les avantages de cette approche par rapport à d'autres techniques, comme le post-traitement mentionné par Boucher et al. (2015) et Dion et al. (2021), qui pourrait compléter et améliorer nos résultats actuels.

3.2 Analyse des performances selon les différentes saisons

L'analyse des performances de modèles hydrologiques selon les différentes saisons est essentielle pour une compréhension approfondie des processus hydrologiques, particulièrement dans une région nordique. Chaque saison, avec ses caractéristiques climatiques et environnementales spécifiques, influence différemment le cycle de l'eau. En hiver, par exemple, les températures basses modifient le stockage de l'eau en favorisant l'accumulation de neige, ce qui diminue le ruissellement et l'infiltration. Le printemps, marqué par la fonte des neiges, présente des défis en termes de prévision des crues avec une augmentation du ruissellement, mais une infiltration toujours faible en raison du sol encore gelé. L'été, souvent caractérisé par un climat plus sec ou des événements pluvieux intenses, influence considérablement le comportement des bassins versants. Enfin, l'automne, avec des précipitations plus fréquentes et moins d'évapotranspiration, peut également influencer les modèles hydrologiques et leur capacité à simuler les débits. Cette approche est utile pour la gestion des ressources en eau et des événements extrêmes tels que les inondations ou les sécheresses. En bref, une analyse saisonnière des modèles hydrologiques contribue non seulement à la précision scientifique, mais aussi à une gestion plus efficace et durable de l'eau dans des contextes environnementaux changeants.

L'efficacité des modèles hydrologiques varie selon les saisons, en raison principalement de la variabilité des débits. Si l'évaluation de leurs performances était sur une base annuelle plutôt que saisonnière, les conclusions pourraient différer, car chaque saison influence différemment les valeurs du CRPS, du MAE et du ABDU. Par exemple, le modèle LSTM affiche d'excellentes performances dans la plupart des scénarios, à l'exception du printemps, où pour le neuvième jour de prévision, son CRPS moyen est environ 30% plus élevé que celui des autres modèles (Figure 2.2b). Cette différence influencerait le résultat des autres saisons dans une évaluation annuelle, car les valeurs de CRPS au printemps sont plus élevées. Ainsi, une analyse saisonnière est cruciale pour une compréhension précise et détaillée des performances du modèle LSTM.

Il est possible d'analyser la performance de la méthode multi-modèle hybride selon les saisons à l'aide du CRPS et du MAE (Figure 2.4, 2.5, 2.6, 2.7, 2.8, 3.1, A I-3 et A I-5). Initialement, sur les 4 saisons, une diminution du CRPS et du MAE est observée lorsqu'une comparaison entre un modèle seul à une combinaison multi-modèle est effectuée. Cela dit, il y a qu'un seul modèle qui n'a pas bénéficié d'une amélioration de performance de la méthode multi-modèle à certains moments. L'analyse a révélé que le modèle LSTM, lorsqu'il est utilisé de manière indépendante, produit des prévisions plus précises et fiables que lorsqu'il est intégré dans une méthode multi-modèle. Cette tendance se manifeste en hiver, le modèle LSTM est plus performant pour les jours 1 et 2, comme le démontre le MAE (voir Figure 2.7a). Au printemps, il affiche une meilleure précision pour les jours 1 et 2, tant selon le MAE que le CRPS (Figures 2.6b et 2.7b). Durant l'été, il est plus précis pour les jours 1, 2, 3 et 7 d'après le MAE (Figure 2.7c), ainsi que pour les jours 1 et 2 selon le CRPS (Figure 2.6c). En automne, le modèle montre de meilleurs résultats pour les jours 1 et 2, selon le MAE (Figure 2.7d).

D'après les diagrammes de Talagrand, il est possible d'évaluer si la méthode multi-modèle améliore les prévisions hydrologiques en fonction des saisons. Les Figures A II-1 et A II-2 illustrent les distributions des ensembles de prévisions générés par les modèles individuels. Pour les deux premiers jours de prévision (Figure A II-1) ainsi que pour les neuf jours (Figure A II-2), les résultats en hiver montrent que les prévisions sont généralement soit sous-

dispersées (Figures A II-1 et A II-2, diagrammes a, e, et y) soit systématiquement biaisées négativement (Figures A II-1 et A II-2, diagrammes i, m, q, u, C, G et K). Dans les autres saisons, les prévisions sont principalement sous-dispersées, avec le test de Kolmogorov-Smirnov (p -value = 5%) rejetant l'hypothèse d'uniformité dans 33 des 40 cas pour la Figure A II-1 et dans 35 des 40 cas pour la Figure A II-2. Les Figures A II-3 et A II-4 démontrent l'impact de la méthode multi-modèle sur les prévisions. Il est observé que les valeurs ABDU de la méthode multi-modèle sont majoritairement inférieures à celles obtenues par les modèles individuels. De plus, la distribution des prévisions de la méthode multi-modèle semble visuellement plus uniforme comparée à celle des modèles individuels. Cette observation est soutenue par le test de Kolmogorov-Smirnov, qui ne rejette pas l'hypothèse d'uniformité en hiver, au printemps et en automne.

3.3 Opérationnalisation

Cette section du mémoire présente le potentiel de la méthode utilisée dans cette étude pour renforcer les systèmes de prévision utilisés par Rio Tinto. Cette recherche, tout en restant modeste dans ses affirmations, suggère des améliorations pratiques et réalisables pour les opérations de prévision hydrologique de l'entreprise.

L'étude a utilisé le modèle hydrologique semi-distribué CEQUEAU7, qui est également le modèle actuellement en usage chez Rio Tinto. Cette correspondance offre une opportunité d'intégrer directement la méthode utilisée. Les résultats obtenus dans l'étude démontrent le potentiel de l'application des méthodes développées avec les systèmes existants chez Rio Tinto. La méthode proposée se définit par sa simplicité d'intégration. Elle consiste essentiellement à ajouter des modèles hydrologiques complémentaires au système déjà en place et à fusionner leurs prévisions. Cette approche multi-modèle augmente la robustesse et la précision des prévisions. La Figure 3.1 présente les résultats de la méthode utilisée avec CEQUEAU7 comme modèle de base. L'introduction de l'assimilation de données (CEQUEAU7_DA en bleu) améliore la précision par rapport au modèle de base (CEQUEAU7_OL en rouge), comme le montrent les valeurs du CRPS inférieures dans la plupart des cas, particulièrement en hiver

où celui-ci est en moyenne 43% plus faible (Figure 3.1a). L'intégration du modèle à apprentissage profond LSTM et du modèle GR4J améliore la qualité des prévisions hydrologiques à travers toutes les saisons et les délais prévisionnels.

L'ajout du modèle LSTM à CEQUEAU7_DA (en vert) montre l'amélioration la plus significative avec une diminution de 50% du CRPS moyen allant jusqu'à 70% en été en comparaison avec CEQUEAU7_OL (Figure 3.1c), ce qui suggère que l'apprentissage profond peut mieux capturer les dynamiques complexes des systèmes hydrologiques. Il est à noter que Rio Tinto utilise une approche d'assimilation de données particulière et le point de départ est donc meilleur que celui identifié par CEQUEAU7_OL et plus similaire à CEQUEAU7_DA. L'intégration du modèle GR4J avec LSTM produit les valeurs du CRPS les plus bas, ce qui indique que cette combinaison est la plus efficace dans cet ensemble de données. L'approche multi-modèle s'avère particulièrement efficace, car elle tire parti de la variabilité structurelle des différents modèles hydrologiques. Cela permet d'obtenir un ensemble de prévisions qui couvre plus fréquemment le débit observé, réduisant ainsi le biais dans les prévisions et le problème de sous-dispersion. En combinant un modèle global, un modèle semi-distribué et un modèle LSTM, comme dans cette étude, on bénéficie d'un spectre plus large de compétences. Cette combinaison exploite les forces de chaque modèle pour un ensemble de prévision hydrologique nettement plus précis et fiable.

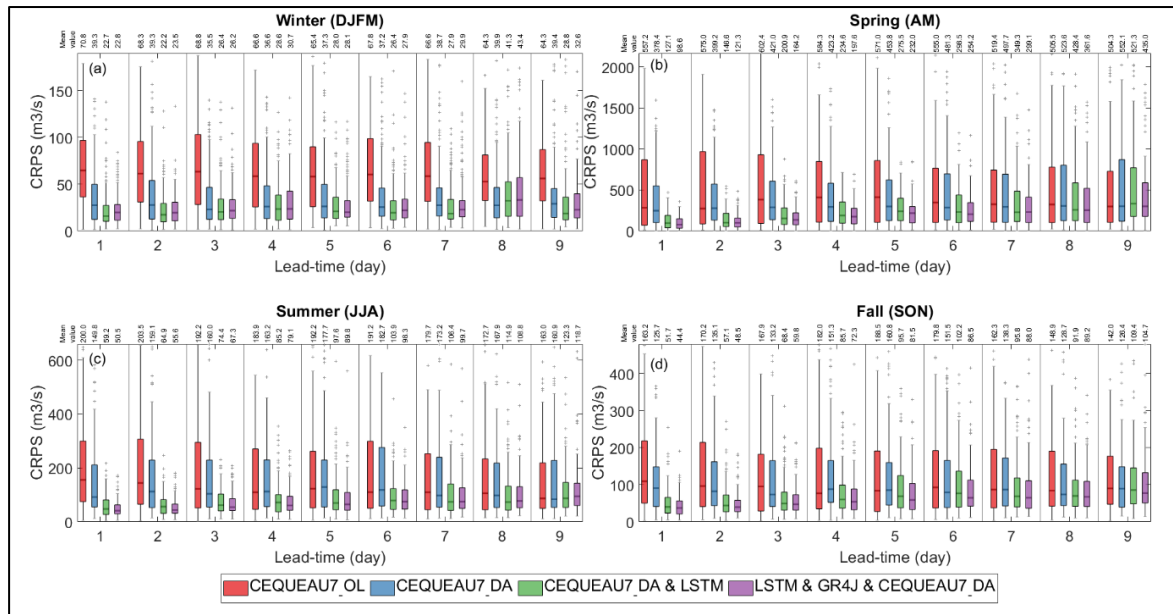


Figure 3.1 CRPS des prévisions hydrologiques d'ensemble pour le bassin versant du LSJ, séparées par saison, pour CEQUEAU7_OL (rouge), CEQUEAU7_DA (bleu), CEQUEAU7_DA & LSTM (vert) et CEQUEAU7_DA, LSTM & GR4J (violet) sur une période de prévision de 2015 à 2019 : hiver (DJFM; a), printemps (AM; b), été (JJA; c) et automne (SON; d)

De plus, l'analyse comparative des diagrammes de Talagrand entre CEQUEAU7_DA et la combinaison CEQUEAU7_DA, LSTM et GR4J illustre des différences (Figure 2.10). En utilisant la méthode multi-modèle, soit CEQUEAU7_DA, LSTM et GR4J, on observe une distribution plus uniforme des valeurs de prévision, ce qui indique une amélioration de la qualité des prévisions. Une diminution du biais de la distribution est également confirmée par des valeurs ABDU plus faibles sur l'ensemble des saisons. Cette tendance se confirme tant dans les premiers jours de prévision (Figure-A II-3) que pour le dernier jour de prévision (Figure-A II-4).

Une considération importante est la capacité des installations informatiques existantes de Rio Tinto à gérer et exécuter ces modèles dans des délais acceptables. Heureusement, les modèles LSTM et GR4J, qui sont des composantes clés de notre approche, sont réputés pour leur efficacité et leur rapidité d'exécution. Ces caractéristiques les rendent plus simples et plus rapides à exécuter comparativement à CEQUEAU7. Il est donc raisonnable de s'attendre à ce

que leur intégration ne pose pas de problèmes significatifs en termes de capacité de traitement ou de temps de réponse.

3.4 Limitations et améliorations

Cette section aborde les diverses limitations rencontrées dans l'étude actuelle et propose des voies d'amélioration possibles. Elle couvre l'impact des modèles hydrologiques, de la période des données météorologiques, de la localisation de l'étude, ainsi que le nombre de membres utilisé en lien avec l'assimilation de données.

3.4.1 Modèles hydrologiques

L'étude s'est concentrée sur un nombre relativement limité de modèles hydrologiques (huit modèles globaux, un modèle semi-distribué et un modèle à base de DL). Cette limitation pourrait affecter la reproductibilité et la robustesse des conclusions. Ce ne sont pas les modèles globaux qui auraient le plus grand effet sur les résultats, car il y en a déjà huit et il a été remarqué qu'un maximum de gain d'une méthode multi-modèle est atteint après une combinaison de deux ou trois modèles comme le travail de Thébault et al. (2023) l'indique également. Toutefois, avoir un nombre plus élevé de modèles semi-distribués ou distribués et à base de DL pourrait modifier les conclusions de cette étude. De plus, utiliser différents modèles distribués serait très pertinent afin de comparer les différentes représentations numériques des processus hydrologiques comme ceux du modèle WaSiM-ETH (Water balance Simulation Model ETH; Schulla & Jasper, 2007) qui intègre une représentation complexe des mouvements souterrains des eaux.

3.4.2 Période des données météorologiques

La période utilisée pour effectuer cette étude est légèrement courte et pourrait être modifiée pour inclure une période plus récente. La période de simulation et de vérification (2015-2019) peut ne pas couvrir suffisamment de variabilité climatique et hydrologique pour tester pleinement l'efficacité des modèles dans différentes conditions. Les données météorologiques

pourraient être combinées à d'autres sources que ECMWF comme Environnement Canada et Changement Climatique et le récent modèle de prévision GraphCast. De plus, comparer les différentes sources de données météorologiques permettrait potentiellement de diminuer le biais puisqu'il est essentiel d'identifier la source des incertitudes comme mentionné dans les récents travaux de Shu et al. (2023) afin de les quantifier et des réduire.

3.4.3 Localisation de l'étude

Pour surmonter les limitations liées à l'utilisation d'un seul bassin versant, il serait bénéfique d'étendre les recherches à plusieurs bassins versants, couvrant différentes caractéristiques géographiques et climatiques. Cela permettrait de tester la robustesse et la flexibilité des modèles dans diverses conditions environnementales comme dans les travaux d'Arsenault et al. (2015), où 429 bassins versants ont été utilisés. Valider la performance de cette technique à travers d'autres bassins versants nordiques de superficie variable, ainsi que sur des bassins versants où le climat est plus aride ou subtropical humide serait également intéressant.

3.4.4 Assimilation de données

Comme mentionné dans la section 2.5.1 de l'article, les prévisions ont été effectuées aux trois jours et le nombre de simulations est différent entre les modèles à cause des exigences de calcul élevées du modèle CEQUEAU7. Pour les huit modèles globaux, 1250 simulations (50 membres de météo \times 25 membres de conditions initiales) sont générées pour chacun des neuf jours de prévision, en comparaison à 50 simulations pour les modèles CEQUEAU7 et LSTM. Une étude supplémentaire pourrait être effectuée en utilisant la moyenne des 25 états initiaux d'assimilation de données avec les 50 membres de météo pour les huit modèles globaux afin d'obtenir le même nombre de prévisions pour chaque modèle utilisé dans l'étude. Cela permettrait de vérifier si les membres supplémentaires de prévision des modèles globaux ont affecté les conclusions de l'étude.

3.4.5 Post-traitement des résultats

Le post-traitement, tel que suggéré dans les travaux de Boucher et al. (2015) et Dion et al. (2021), représente une voie prometteuse pour améliorer les prévisions hydrologiques. Cette approche peut inclure des techniques statistiques pour ajuster les résultats des modèles, en tenant compte des erreurs systématiques et en améliorant la précision des prévisions. Les futures recherches pourraient se concentrer sur l'intégration de techniques de post-traitement avec les modèles à base de DL pour affiner davantage les prévisions particulièrement pour les jours sept à neuf des prévisions.

CONCLUSION

En conclusion, ce mémoire offre un aperçu de l'efficacité potentielle d'une approche hybride multi-modèle en utilisant trois types de modèles hydrologiques pour améliorer la précision des prévisions hydrologiques d'ensemble à court terme à l'échelle saisonnière. En combinant des modèles hydrologiques conceptuels globaux et semi-distribué avec un modèle basé sur l'apprentissage profond, les avantages de chaque modèle ont été pris en compte. Les principales conclusions de cette étude sont les suivantes :

- 1) Le modèle LSTM obtient de meilleures performances en prévision du débit sur l'ensemble des saisons.
- 2) Le LSTM permet d'obtenir des prévisions d'ensemble plus précises lorsqu'il est combiné avec d'autres types de modèles plutôt que lorsqu'il est utilisé seul.
- 3) L'approche multi-modèle utilisée a démontré que la plus grande amélioration est obtenue lorsque trois modèles de structures différentes sont combinés (global et conceptuel, semi-distribué et basé sur le DL).

Cependant, ces conclusions ouvrent également une interrogation : les modèles purement basés sur l'apprentissage profond sont-ils plus adaptés à fonctionner de manière autonome, ou leur potentiel est-il pleinement exploité lorsqu'ils sont combinés avec d'autres types de modèles ? Dans ce contexte, il est intéressant de noter que le modèle LSTM semble présenter des écarts plus marqués par rapport au débit observé durant le printemps, notamment aux jours huit et neuf, comparativement aux autres modèles étudiés. Cette divergence suggère une possible limitation dans la capacité du modèle LSTM à prédire de manière précise des débits plus élevés sur des horizons temporels plus longs. De plus, le modèle LSTM se révèle sous-dispersé, comme le montre clairement la Figure 2.10. Cette observation renforce l'idée que, malgré leurs avantages, les modèles basés sur le DL peuvent bénéficier de manière significative de l'intégration avec d'autres types de modèles hydrologiques, permettant ainsi d'atteindre une compréhension et une précision prévisionnelle plus équilibrées et fiables.

De plus, cette étude met l'accent sur l'application de cette méthode de prévision hydrologique pour améliorer la fiabilité et la précision des systèmes de prévision chez Rio Tinto. En utilisant le modèle hydrologique semi-distribué CEQUEAU7 comme point de départ, cette recherche a démontré la pertinence et la simplicité de l'intégration de modèles hydrologiques complémentaires. Les résultats montrent que l'introduction du modèle à apprentissage profond LSTM et le modèle GR4J améliorent significativement la précision des prévisions hydrologiques à différentes échelles temporelles pour toutes les saisons. Notamment, l'ajout du modèle LSTM à CEQUEAU7_DA a entraîné une amélioration particulièrement marquée, soulignant l'efficacité de l'apprentissage profond dans la capture des dynamiques complexes des systèmes hydrologiques. Enfin, l'intégration du modèle GR4J avec le modèle LSTM a produit les meilleures performances, confirmant la valeur d'une approche multi-modèle qui prend avantage des forces spécifiques de chaque modèle. Dans l'ensemble, cette étude illustre la pertinence d'une telle approche pour améliorer la prévision hydrologique, ouvrant ainsi la voie à des applications pratiques dans des contextes opérationnels comme celui de Rio Tinto.

Les résultats de cette étude mettent en lumière les avantages, les limites et l'évaluation spécifique de l'approche multi-modèle dans la prévision du débit pour toutes les saisons. Dans l'ensemble, cette étude montre que la combinaison de modèles conceptuels et semi-distribué avec un modèle LSTM basé sur l'apprentissage profond constitue une approche multi-modèle prometteuse pour la prévision du débit à court terme dans les régions où le régime hydrologique est principalement dominé par la neige.

Il serait également intéressant de vérifier si les modèles de prévisions hydrologiques LSTM sont plus performants en combinaison avec un modèle de prévision météorologique à base d'IA comme GraphCast présenté dans les travaux de Lam et al. (2023). GraphCast, avec sa capacité supérieure à prévoir la météo, pourrait révolutionner les modèles de prévision hydrologique. En combinant GraphCast avec des modèles LSTM, on pourrait mieux prédire des événements comme les inondations et sécheresses. Cette combinaison offre une meilleure gestion de l'eau et une réponse plus rapide aux conditions météo extrêmes.

RECOMMANDATIONS

Sur la base des limitations identifiées dans l'étude, voici des recommandations clés pour améliorer les futures recherches portant sur la précision et la fiabilité des modèles de prévision:

- 1) Intégrer davantage de modèles hydrologiques semi-distribués et distribués, ainsi que des modèles à base d'apprentissage profond, pour enrichir l'analyse et renforcer la robustesse des conclusions. Utiliser une variété de modèles pour comparer et valider les résultats, afin de garantir une meilleure généralisation et reproductibilité des conclusions.
- 2) Inclure une période de données plus longue et plus récente pour couvrir une plus grande variabilité climatique et hydrologique. Combiner les données météorologiques issues de différentes sources, telles que GraphCast et Environnement et Changement climatique Canada, pour enrichir l'analyse et la rendre plus complète.
- 3) Étendre les recherches à plusieurs bassins versants présentant différentes caractéristiques géographiques et climatiques, pour tester la flexibilité et la robustesse des modèles dans divers environnements.
- 4) Incorporer des techniques de post-traitement pour ajuster les résultats des modèles, en tenant compte des erreurs systématiques et en améliorant la précision des prévisions.
- 5) Réaliser une étude sur l'impact du nombre de membres de prévision utilisé par modèle.

Ces recommandations visent à surmonter les limitations actuelles et à améliorer l'efficacité des modèles hydrologiques, en les rendant plus robustes, fiables et adaptés aux diverses conditions environnementales et climatiques.

ANNEXE I

RÉSULTATS DES VALEURS MAE SUPPLÉMENTAIRES

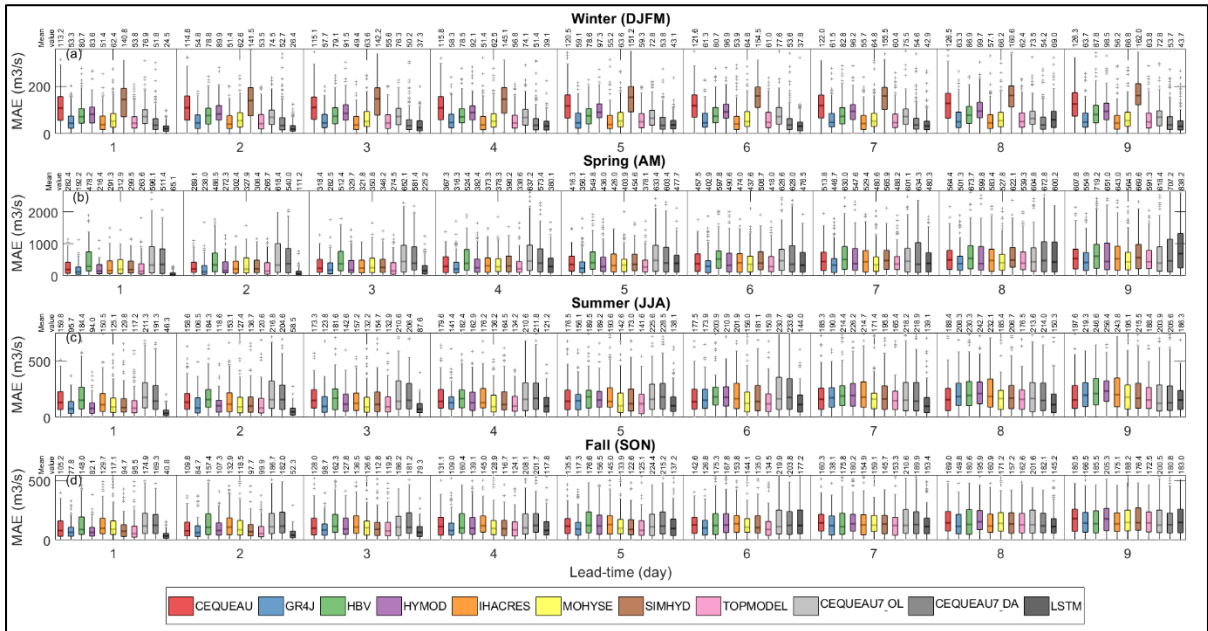


Figure-A I-1 MAE des prévisions hydrologiques d'ensemble pour le bassin versant du LSJ par saison pour tous les modèles sur la période de prévision 2015-2019 : hiver (DJFM ; a), printemps (AM ; b), été (JJA ; c) et automne (SON ; d)

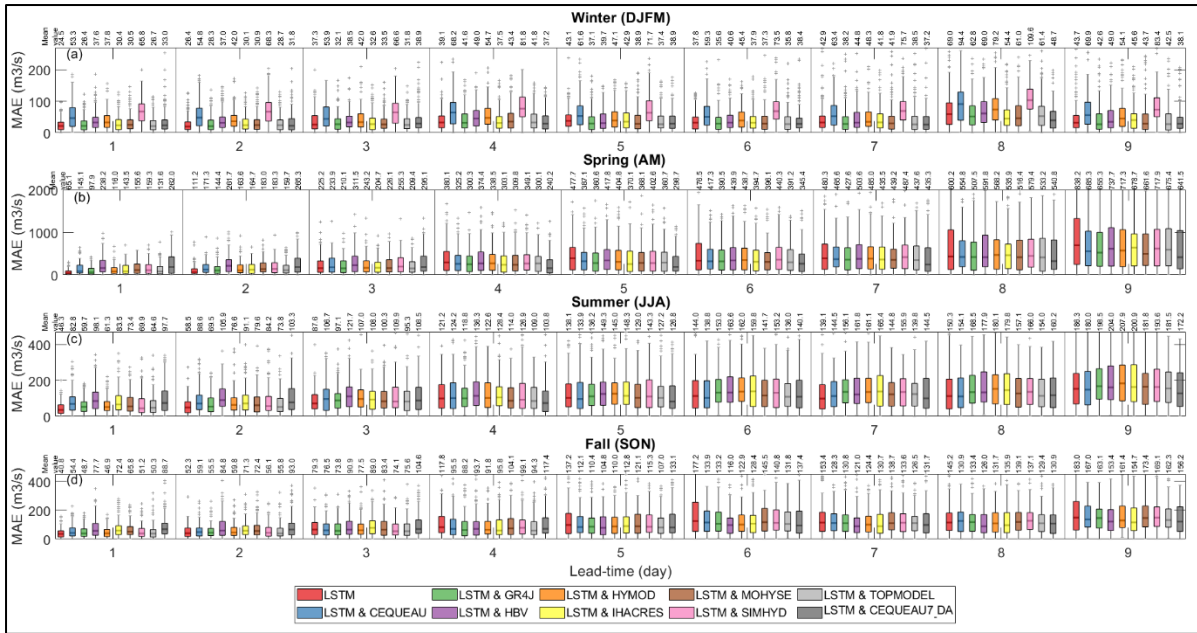


Figure-A I-2 MAE des prévisions hydrologiques d'ensemble pour le bassin versant LSJ par saison pour le LSTM seul (rouge) et lorsqu'il est combiné avec d'autres modèles sur la période de prévision 2015-2019 : hiver (DJFM ; a), printemps (AM ; b), été (JJA ; c) et automne (SON ; d)

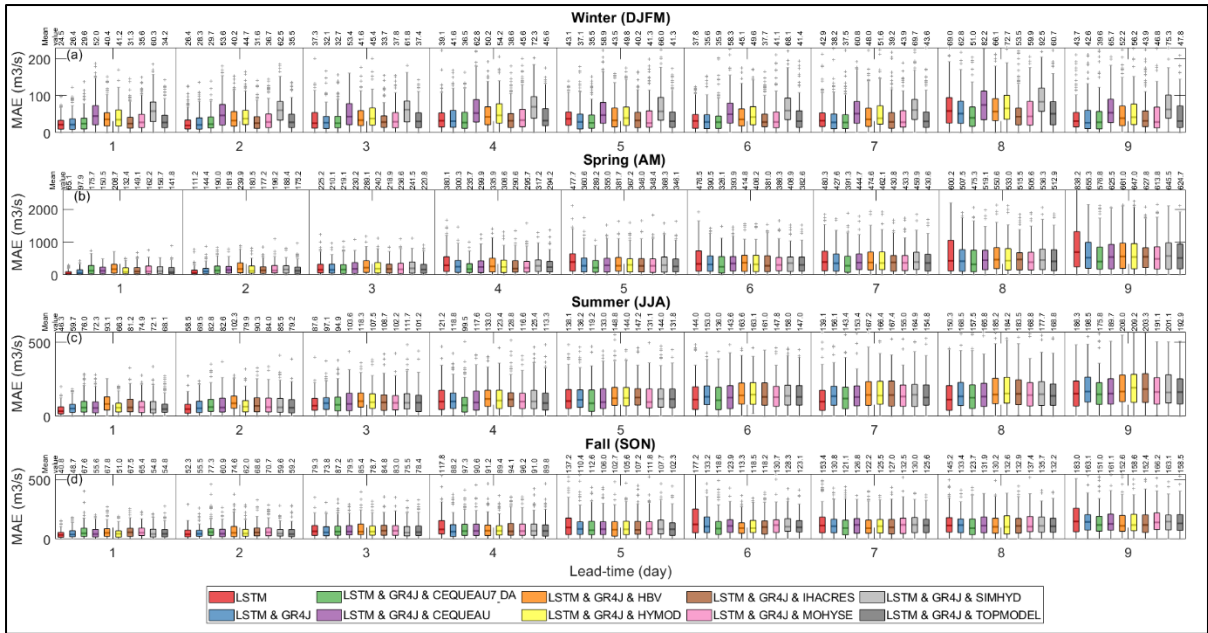


Figure-A I-3 MAE des prévisions hydrologiques d'ensemble pour le bassin versant LSJ par saison pour le LSTM seul (rouge), le LSTM-GR4J (bleu), et le LSTM-GR4J combiné avec d'autres modèles sur la période de prévision 2015-2019 : hiver (DJFM ; a), printemps (AM ; b), été (JJA ; c), et automne (SON ; d)

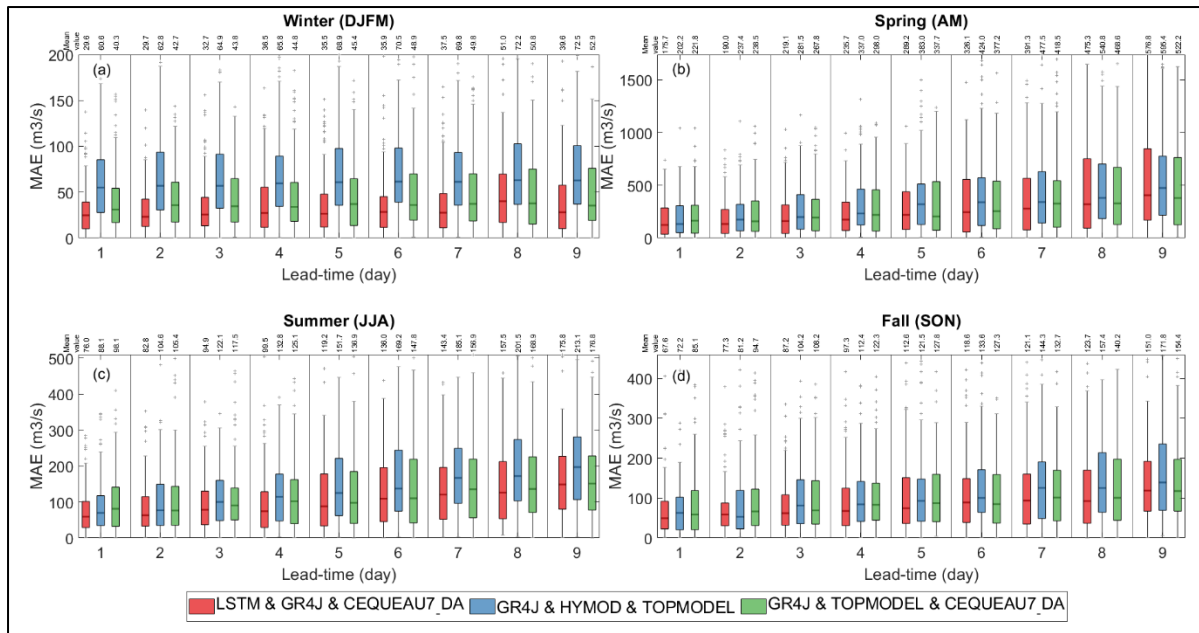


Figure-A I-4 MAE des prévisions hydrologiques d'ensemble pour le bassin versant LSJ par saison pour la combinaison LSTM-GR4J-CEQUEAU7_DA (rouge), GR4J-HYMOD-TOPMODEL (bleu) et GR4J-TOPMODEL-CEQUEAU7_DA (vert) sur la période de prévision 2015-2019 : hiver (DJFM ; a), printemps (AM ; b), été (JJA ; c) et automne (SON ; d)

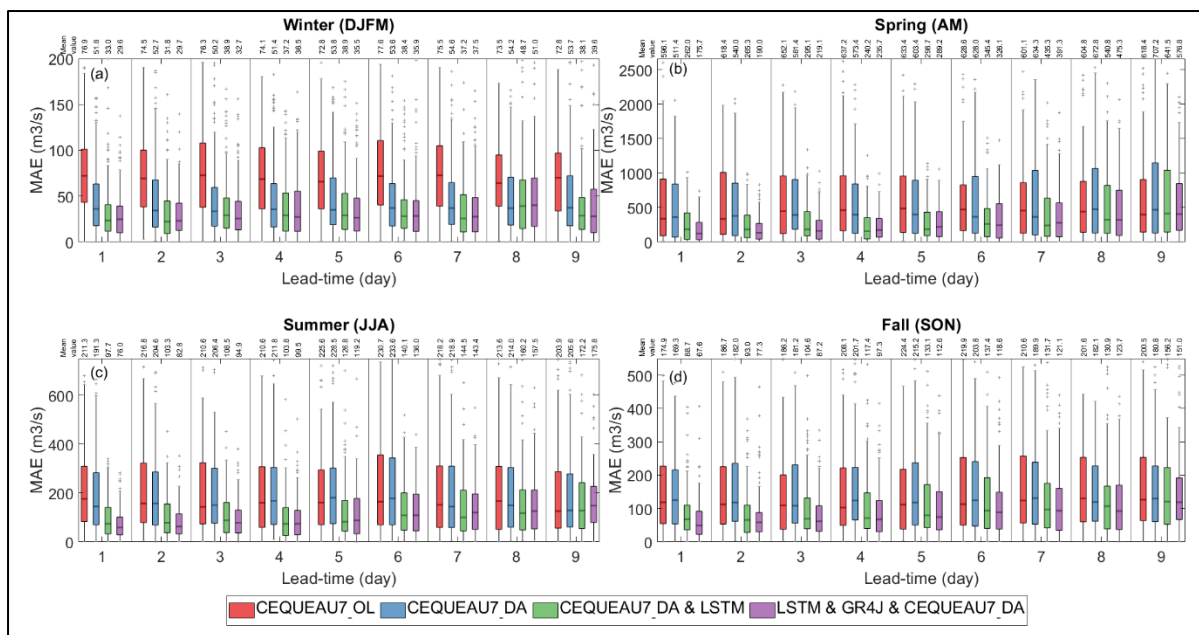


Figure-A I-5 MAE des prévisions hydrologiques d'ensemble pour le bassin versant du LSJ, séparées par saison, pour CEQUEAU7_OL (rouge), CEQUEAU7_DA (bleu), CEQUEAU7_DA & LSTM (vert) et CEQUEAU7_DA, LSTM & GR4J (violet) sur une période de prévision de 2015 à 2019 : hiver (DJFM; a), printemps (AM; b), été (JJA; c) et automne (SON; d)

ANNEXE II

DIAGRAMME DE TALAGRAND

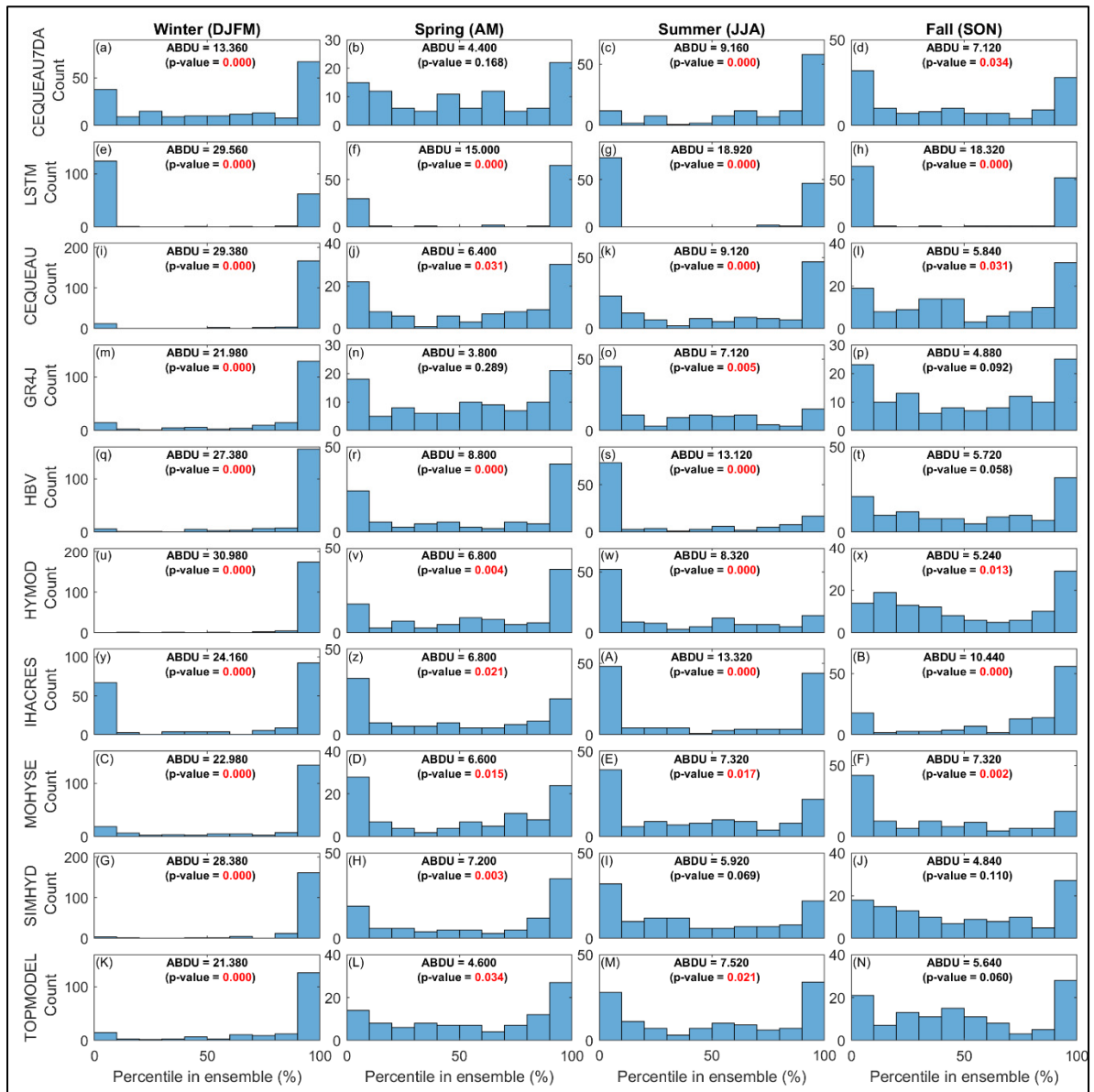


Figure-A II-1 Reliability of forecasts using individual models: Talagrand diagrams representing the rank (in percentiles) of the 2-day streamflow forecasts by the 10 hydrological models on the LSJ catchment. A reliable forecast is expected to lead to flat distributions (similar count in each bin). A p-value less than or equal to 0.05 indicates that the uniformity hypothesis is rejected by the K-S test (red highlighted values)

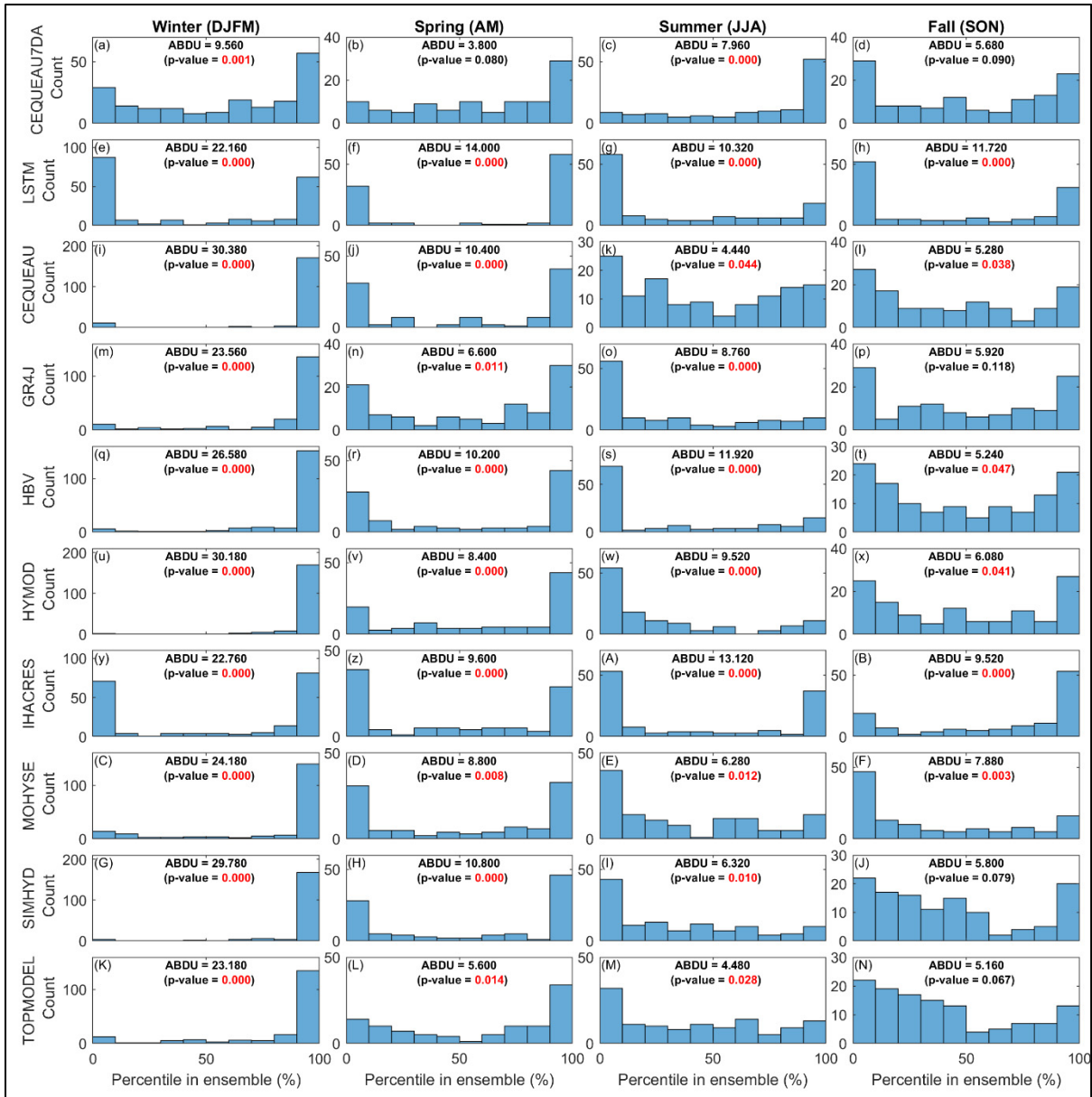


Figure-A II-2 Reliability of forecasts using individual models: Talagrand diagrams representing the rank (in percentiles) of the 9-day streamflow forecasts by the 10 hydrological models on the LSJ catchment. A reliable forecast is expected to lead to flat distributions (similar count in each bin). A p-value less than or equal to 0.05 indicates that the uniformity hypothesis is rejected by the K-S test (red highlighted values)

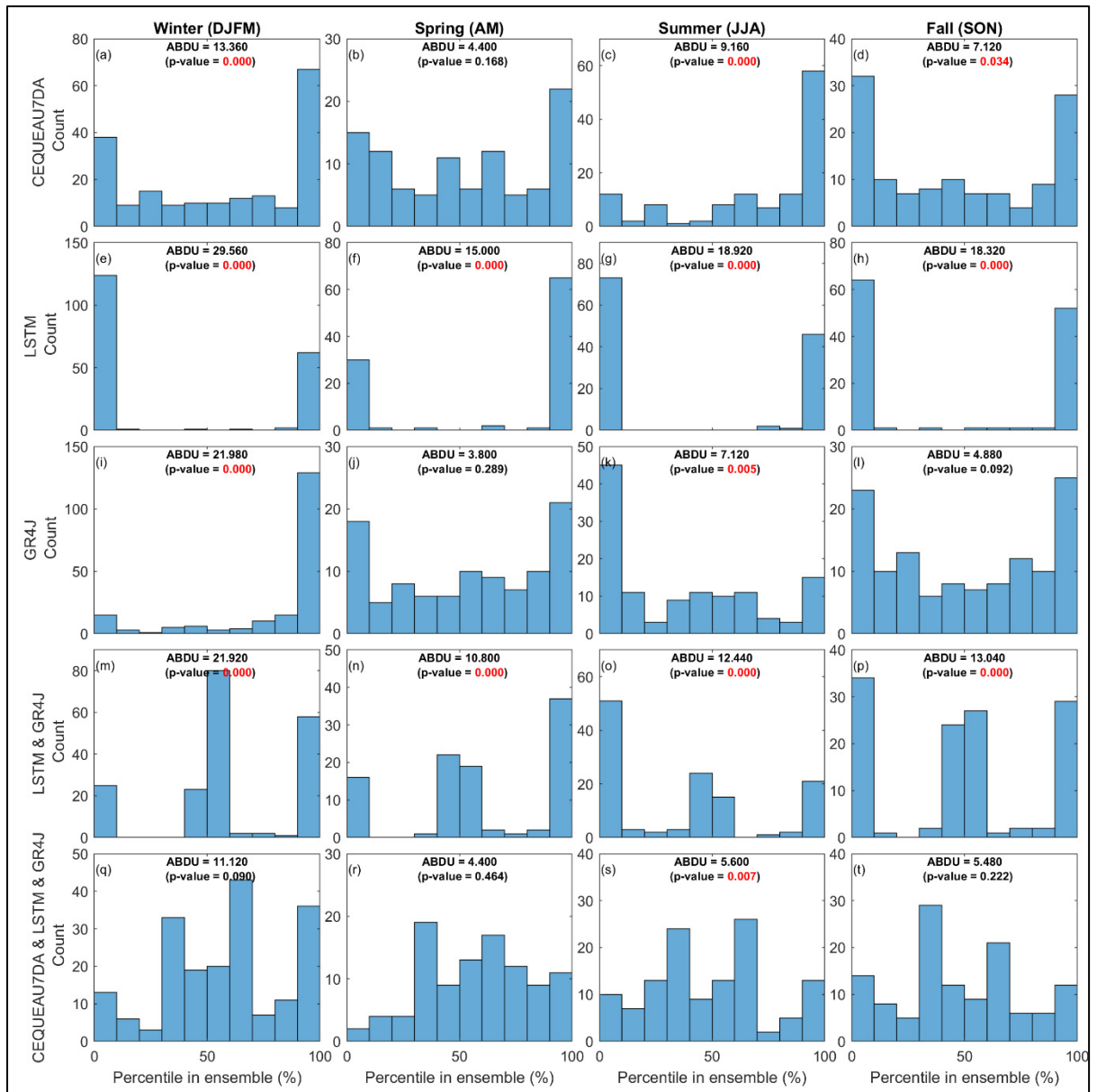


Figure-A II-3 Reliability of forecasts using individual models and two- and three-model combinations: Talagrand diagrams representing the rank (in percentiles) of the 2-day streamflow forecasts for the individual CEQUEAU7_DA, LSTM, and GR4J models, as well as for the combination of the LSTM-GR4J and LSTM-GR4J-CEQUEAU7_DA models in the LSJ catchment. A reliable forecast is expected to lead to flat distributions (similar count in each bin). A p-value less than or equal to 0.05 indicates that the uniformity hypothesis is rejected by the K-S test (red highlighted values)

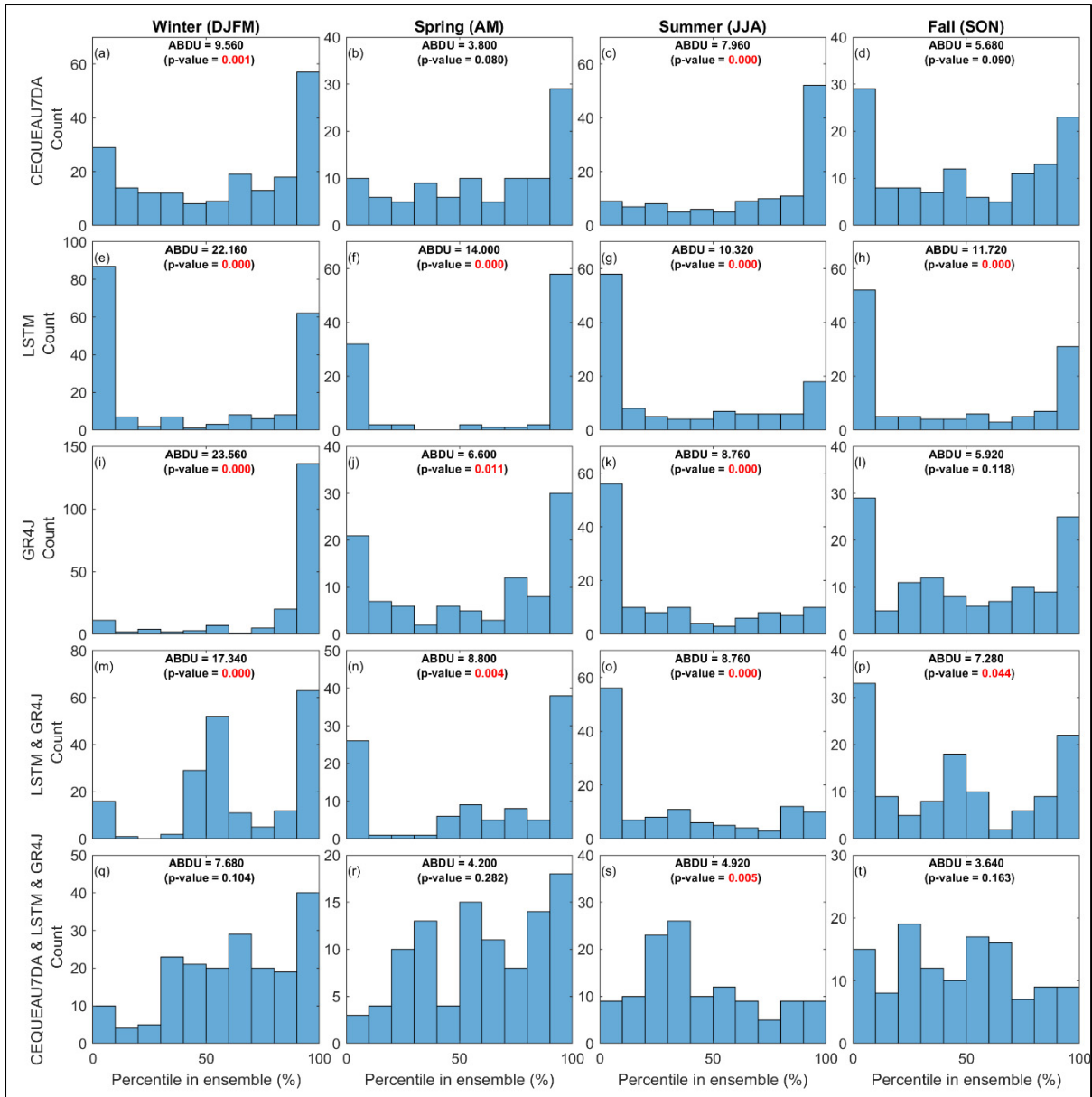


Figure-A II-4 Reliability of forecasts using individual models and two- and three-model combinations: Talagrand diagrams representing the rank (in percentiles) of the 9-day streamflow forecasts for the individual CEQUEAU7_DA, LSTM, and GR4J models, as well as for the combination of the LSTM-GR4J and LSTM-GR4J-CEQUEAU7_DA models in the LSJ catchment. A reliable forecast is expected to lead to flat distributions (similar count in each bin). A p-value less than or equal to 0.05 indicates that the uniformity hypothesis is rejected by the K-S test (red highlighted values)

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

Abbe, C. (1901). The physical basis of long-range weather forecasts. *Monthly Weather Review*, 29(12), 551-561.

Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of climate*, 9(7), 1518-1530.

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier, D. P. (2016). Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resources Research*, 52(6), 4209-4225.

Arnell, N. W., & Gosling, S. N. (2013). The impacts of climate change on river flow regimes at the global scale. *Journal of Hydrology*, 486, 351-364.

Arsenault, R. (2015). Équifinalité, incertitude et procédures multi-modèle en prévision hydrologique aux sites non-jaugés. (Thèse de doctorat, École de technologie supérieure, Montréal, QC). Repéré à <https://espace.etsmtl.ca/id/eprint/1488/>

Arsenault, R., & Brissette, F. (2016). Multi-model averaging for continuous streamflow prediction in ungauged basins. *Hydrological Sciences Journal*, 61(13), 2443-2454.

Arsenault, R., Brissette, F. and Martel, J.L., 2018. The hazards of split-sample validation in hydrological model calibration. *Journal of hydrology*, 566, pp.346-362.

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

Arsenault, R., Gatién, P., Renaud, B., Brissette, F., & Martel, J. L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology*, 529, 754-767.

Arsenault, R., Latraverse, M., & Duchesne, T. (2016). An efficient method to correct under-dispersion in ensemble streamflow prediction of inflow volumes for reservoir optimization. *Water resources management*, 30, 4363-4380.

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

Arsenault, R., Poulin, A., Côté, P. and Brissette, F., 2014. Comparison of stochastic optimization algorithms in hydrological model calibration. *Journal of Hydrologic Engineering*, 19(7), pp.1374-1384.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing*, 50(2), 174-188.

Bastola, S., Murphy, C., & Sweeney, J. (2011). The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. *Advances in water resources*, 34(5), 562-576.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567), 47-55.

Bolin, B. (1955). Numerical forecasting with the barotropic model 1. *Tellus*, 7(1), 27-49.

- Boucher, M. A. (2011). *Le défi de l'incertitude: de la production de prévisions hydrologiques d'ensemble à leur utilisation opérationnelle* (Doctoral dissertation, Université Laval).
- Boucher, M. A., Tremblay, D., Delorme, L., Perreault, L., & Anctil, F. (2012). Hydro-economic assessment of hydrological forecasting systems. *Journal of Hydrology*, *416*, 133-144.
- Boucher, M. A., Perreault, L., Anctil, F., & Favre, A. C. (2015). Exploratory analysis of statistical post-processing methods for hydrological ensemble forecasts. *Hydrological Processes*, *29*(6), 1141-1155.
- Cassagnole, M., Ramos, M. H., Zalachori, I., Thirel, G., Garçon, R., Gailhard, J., & Ouillon, T. (2021). Impact of the quality of hydrological forecasts on the management and revenue of hydroelectric reservoirs—a conceptual approach. *Hydrology and Earth System Sciences*, *25*(2), 1033-1052.
- Carpenter, T. M., & Georgakakos, K. P. (2006). Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of hydrology*, *329*(1-2), 174-185.
- Chao, L., Hui, Z., & Xiaofeng, Z. (2015). Data quality assessment in hydrological information systems. *Journal of Hydroinformatics*, *17*(4), 640-661.
- Chevuturi, A., Tanguy, M., Facer-Childs, K., Martínez-de la Torre, A., Sarkar, S., Thober, S., ... & Blyth, E. (2023). Improving global hydrological simulations through bias-correction and multi-model blending. *Journal of Hydrology*, *621*, 129607.
- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... & Rasmussen, R. M. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*(4), 2498-2514.

Davie, T. (2019). *Fundamentals of hydrology*. Routledge.

Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., ... & Zhu, Y. (2014). The science of NOAA's operational hydrologic ensemble forecast service. *Bulletin of the American Meteorological Society*, 95(1), 79-98.

Dion, P. (2020). Amélioration des prévisions hydrologiques d'ensemble dans un contexte de modélisation multi-modèle. masters. École de technologie supérieure. Repéré à <https://espace.etsmtl.ca/id/eprint/2684/>

Dion, P., Martel, J. L., & Arsenault, R. (2021). Hydrological ensemble forecasting using a multi-model framework. *Journal of Hydrology*, 600, 126537.

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research*, 28(4), 1015-1031.

Eagleson, P. S. (1994). The evolution of modern hydrology (from watershed to continent in 30 years). *Advances in water resources*, 17(1-2), 3-18.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5), 10143-10162.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., ... & Tarboton, D. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, 537, 45-60.

Feizi, H., Apaydin, H., Sattari, M. T., Colak, M. S., & Sibtain, M. (2022). Improving reservoir inflow prediction via rolling window and deep learning-based multi-model approach: case

study from Ermenek Dam, Turkey. *Stochastic Environmental Research and Risk Assessment*, 36(10), 3149-3169.

Fluixá-Sanmartín, J., Altarejos-García, L., Morales-Torres, A., & Escuder-Bueno, I. (2018). Climate change impacts on dam safety. *Natural Hazards and Earth System Sciences*, 18(9), 2471-2488.

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

Granger, C. W., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of forecasting*, 3(2), 197-204.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, 377(1-2), 80-91.

Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129(3), 550-560.

Hamill, T. M., & Colucci, S. J. (1997). Verification of Eta-RSM short-range ensemble forecasts. *Monthly Weather Review*, 125(6), 1312-1327.

Hattermann, F. F., Huang, S., & Koch, H. (2015). Climate change impacts on hydrology and water resources.

Hemri, S., Lisniak, D., & Klein, B. (2015). Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research*, 51(9), 7436-7451.

Her, Y., Yoo, S. H., Cho, J., Hwang, S., Jeong, J., & Seong, C. (2019). Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. *Scientific reports*, 9(1), 4974.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5), 559-570.

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

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water*, 10(11), 1543.

Jaiswal, R. K., Ali, S., & Bharti, B. (2020). Comparative evaluation of conceptual and physical rainfall-runoff models. *Applied water science*, 10, 1-14.

Jajarmizadeh, M., Harun, S., & Salarpour, M. (2012). A review on theoretical consideration and types of models in hydrology. *Journal of Environmental Science and Technology*, 5(5), 249-261.

Jeong, D. I., & Kim, Y. O. (2009). Combining single-value streamflow forecasts—A review and guidelines for selecting techniques. *Journal of Hydrology*, 377(3-4), 284-299.

Jiang, S., Ren, L., Yang, X., Ma, M., & Liu, Y. (2014). Multi-model ensemble hydrologic prediction and uncertainties analysis. *Proceedings of the International Association of Hydrological Sciences*, 364, 249-254.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.

Kavetski, D., Franks, S. W., & Kuczera, G. (2003). Confronting input uncertainty in environmental modelling. *Calibration of watershed models*, 6, 49-68.

Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water resources research*, 42(3).

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of hydrology*, 249(1-4), 2-9.

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of hydrology*, 424, 264-277.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12), 11344-11354.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ... & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science*, eadi2336.

Larabi, S., St-Hilaire, A., Chebana, F. and Latraverse, M., 2018. Using functional data analysis to calibrate and evaluate hydrological model performance. *Journal of Hydrologic Engineering*, 23(7), p.04018026.

Latif, S. D., & Ahmed, A. N. (2023). A review of deep learning and machine learning techniques for hydrological inflow forecasting. *Environment, Development and Sustainability*, 1-28.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.

Leisenring, M., & Moradkhani, H. (2011). Snow water equivalent prediction using Bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, 25, 253-270.

Leutbecher, M., Lock, S. J., Ollinaho, P., Lang, S. T., Balsamo, G., Bechtold, P., ... & Weisheimer, A. (2017). Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143(707), 2315-2339.

Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., & Di, Z. (2017). A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdisciplinary Reviews: Water*, 4(6), e1246.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., ... & Restrepo, P. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and earth system sciences*, 16(10), 3863-3887.

Liu, Y., Zhang, J. Y., Elmahdi, A., Yang, Q. L., Guan, X. X., Liu, C. S., ... & Wang, G. Q. (2021). Transferability of a lumped hydrologic model, the Xin'anjiang model based on similarity in climate and geography. *Water Supply*, 21(5), 2191-2201.

Mather, A. L., & Johnson, R. L. (2016). Forecasting turbidity during streamflow events for two Mid-Atlantic US streams. *Water Resources Management*, 30, 4899-4912.

Ma, X., Hu, H., & Ren, Y. (2023). A hybrid deep learning model based on feature capture of water level influencing factors and prediction error correction for water level prediction of cascade hydropower stations under multiple time scales. *Journal of Hydrology*, 617, 129044.

- Mazrooei, A., & Sankarasubramanian, A. (2019). Improving monthly streamflow forecasts through assimilation of observed streamflow for rainfall-dominated basins across the CONUS. *Journal of Hydrology*, 575, 704-715.
- McCollor, D., & Stull, R. (2008). Hydrometeorological short-range ensemble forecasts in complex terrain. Part II: Economic evaluation. *Weather and Forecasting*, 23(4), 557-574.
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, 10(3), 282-290.
- Ng, K.W., Huang, Y.F., Koo, C.H., Chong, K.L., El-Shafie, A. and Ahmed, A.N., 2023. A review of hybrid deep learning applications for streamflow forecasting. *Journal of Hydrology*, p.130141.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, 303(1-4), 290-306.
- Pagano, T. C., Shrestha, D. L., Wang, Q. J., Robertson, D., & Hapuarachchi, P. (2013). Ensemble dressing for hydrological applications. *Hydrological Processes*, 27(1), 106-116.
- Parker, D., & Fordham, M. (1996). An evaluation of flood forecasting, warning and response systems in the European Union. *Water Resources Management*, 10, 279-302.
- 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.

Piazzzi, G., Thirel, G., Perrin, C., & Delaigue, O. (2021). Sequential data assimilation for streamflow forecasting: Assessing the sensitivity to uncertainties and updated variables of a conceptual hydrological model at basin scale. *Water Resources Research*, 57, e2020WR028390. <https://doi.org/10.1029/2020WR028390>

Ratto, M., Young, P. C., Romanowicz, R., Pappenberger, F., Saltelli, A., & Pagano, A. (2007). Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology. *Hydrology and Earth System Sciences*, 11(4), 1249-1266.

Sabzipour, B., Arsenault, R., Troin, M., Martel, J. L., & Brissette, F. (2023a). Sensitivity analysis of the hyperparameters of an Ensemble Kalman Filter application on a semi-distributed hydrological model for streamflow forecasting. *Journal of Hydrology*, 130251.

Sabzipour, B., Arsenault, R., Troin, M., Martel, J.L., Brissette, F., Brunet, F. and Mai, J., (2023b). 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*, p.130380.

See, L., & Openshaw, S. (2000). A hybrid multi-model approach to river level forecasting. *Hydrological Sciences Journal*, 45(4), 523-536.

Schaake, J. C., Hamill, T. M., Buizza, R., & Clark, M. (2007). HEPEX: the hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society*, 88(10), 1541-1548.

Schulla, J., & Jasper, K. (2007). Model description wasim-eth. Institute for Atmospheric and Climate Science, Swiss Federal Institute of Technology, Zürich.

Shu, Z., Zhang, J., Wang, L., Jin, J., Cui, N., Wang, G., ... & Liu, C. (2023). Evaluation of the impact of multi-source uncertainties on meteorological and hydrological ensemble forecasting. *Engineering*, 24, 212-228.

Shamir, E., Lee, B.-J., Bae, D.-H., & Georgakakos, K. P. (2010). Flood forecasting in regulated basins using the ensemble extended kalman filter with the storage function method. *Journal of Hydrologic Engineering*, 15, 1030–1044. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000282](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000282)

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, 82(12), 2635-2670.

Sood, A., & Smakhtin, V. (2015). Global hydrological models: a review. *Hydrological Sciences Journal*, 60(4), 549-565.

St-Hilaire, A., Boucher, M.A., Chebana, F., Ouellet-Proulx, S., Zhou, Q.X., Larabi, S., Dugdale, S. and Latraverse, M., 2015, April. Breathing a new life to an older model: The CEQUEAU tool for flow and water temperature simulations and forecasting. In Proceedings of the 22nd Canadian Hydrotechnical Conference.

Sun, Y., Bao, W., Valk, K., Brauer, C. C., Sumihar, J., & Weerts, A. H. (2020). Improving forecast skill of lowland hydrological models using ensemble Kalman filter and unscented Kalman filter. *Water Resources Research*, 56(8), e2020WR027468. <https://doi.org/10.1029/2020wr027468>

Talagrand, O. (1999). Evaluation of probabilistic prediction systems. In Workshop Proceedings" Workshop on Predictability", 20-22 October 1997, ECMWF, Reading, UK.

Tanty, R., & Desmukh, T. S. (2015). Application of artificial neural network in hydrology— A review. *Int. J. Eng. Technol. Res*, 4(6), 184-188.

Thébault, C., Perrin, C., Andréassian, V., Thirel, G., Legrand, S., & Delaigue, O. (2023). Multi-model approach in a variable spatial framework for streamflow simulation. *EGUsphere*, 2023, 1-34.

Thibault, A., Anctil, F., & Boucher, M. A. (2016). Accounting for three sources of uncertainty in ensemble hydrological forecasting. *Hydrology and Earth System Sciences*, 20(5), 1809-1825.

Thibault, A., Seiller, G., Poncelet, C., & Anctil, F. (2019, 2019/05/15/). HOOPLA: un outil multifonction pour la modélisation hydrologique présentée à Colloque AQT/RHQ – La télédétection de l'eau, dans tous leurs états, Sherbrooke, Canada. Repéré à <https://aqtrhq2019.sciencesconf.org/>

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.

Valéry, A., Andréassian, V., & Perrin, C. (2014). ‘As simple as possible but not simpler’: What is useful in a temperature-based snow-accounting routine? Part 2–Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. *Journal of hydrology*, 517, 1176-1187.

Velazquez, J.A., Anctil, F., Ramos, M.H. and Perrin, C., 2011. Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Advances in Geosciences*, 29, pp.33-42.

Wan, Y., Chen, J., Xu, C.-Y., Xie, P., Qi, W., Li, D., Zhang, S., 2021. Performance dependence of multi-model combination methods on hydrological model calibration strategy and ensemble size. *J. Hydrol.* 603, 127065. <https://doi.org/10.1016/j.jhydrol.2021.127065>

Welch, G., & Bishop, G. (1995). An introduction to the Kalman filter.

Wetterhall, F., Pappenberger, F., Alfieri, L., Cloke, H., & Thielen, J. (2014, May 01, 2014). Forecaster priorities for improving probabilistic flood forecasts présentée à EGU General Assembly Conference Abstracts. Repéré à <https://ui.adsabs.harvard.edu/abs/2014EGUGA..16.8159W>

Wetterhall, F., Pappenberger, F., Alfieri, L., Cloke, H. L., Thielen-del Pozo, J., Balabanova, S., . . . Holubecka, M. (2013). HESS Opinions "Forecaster priorities for improving probabilistic flood forecasts". *Hydrol. Earth Syst. Sci.*, 17(11), 4389-4399. doi: 10.5194/hess-17-4389-2013. Repéré à <https://www.hydrol-earth-syst-sci.net/17/4389/2013/>

Wetterhall, F., & Smith, P. (2019). Hydrological Challenges in Meteorological Postprocessing. Dans Q. Duan, F. Pappenberger, A. Wood, H. L. Cloke & J. C. Schaake (Éds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 239-253). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-39925-1_15. Repéré à https://doi.org/10.1007/978-3-642-39925-1_15

Zappa, M., Rotach, M. W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., ... & Wunram, C. (2008). MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. *Atmospheric Science Letters*, 9(2), 80-87.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., ... & Zhuang, J. (2018). Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*, 565, 720-736.

Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of hydrology*, 561, 918-929.

Zhang, Z., Tang, H., Qin, H., Luo, B., Zhou, C., & Zhou, H. (2023). Multi-step ahead probabilistic forecasting of multiple hydrological variables for multiple stations. *Journal of Hydrology*, 617, 129094.