

Évaluation des performances de l'intelligence artificielle et de l'apprentissage automatique pour la prévision des crues : étude de cas du bassin versant de l'III

Artificial intelligence and machine learning performance assessment for flood forecasting: a case study of the III watershed

■ N. REIMINGER^{1,2*}, X. JURADO¹, L. SAUNIER¹, L. MAURER^{2,3}, E. REIMINGER⁴, L. WEBER¹, T.H.L. NGUYEN^{1,2}, C. WEMMERT²

¹ AIR&D – Strasbourg

² Laboratoire ICube (UMR 7357) – Illkirch-Graffenstaden

³ École nationale du génie de l'eau et de l'environnement de Strasbourg (Engees) – Strasbourg

⁴ Direction régionale de l'environnement, de l'aménagement et du logement (Dreal) Grand Est – Strasbourg

Mots-clés :

Prévision
Crue
Inondation
Hauteur d'eau
Apprentissage automatique
Intelligence artificielle

RÉSUMÉ

Les inondations, un aléa récurrent pouvant se montrer dévastateur, peuvent causer de lourds dommages humains et économiques, et sont susceptibles de s'intensifier avec le changement climatique. Leurs impacts soulignent la nécessité d'anticiper les crues en développant constamment des outils efficaces et à la pointe pour la gestion du risque inondation. En France, la prévision des crues repose sur des outils de modélisation spécifiques basés sur la physique, la donnée ou une combinaison des deux. L'intelligence artificielle (IA), en plein essor, offre de nouvelles possibilités et perspectives pour les ingénieurs et chercheurs du domaine, mais n'est cependant pas encore utilisée à l'échelle nationale pour la prévision des crues. Bien que les algorithmes complexes d'apprentissage automatique (ML, *Machine Learning*) soient prometteurs pour la prévision des crues, leur mise en œuvre et leur utilisation courante restent difficiles pour les non-spécialistes. Cet article explore l'utilisation de six algorithmes d'apprentissage automatique, incluant des modèles simples d'ensemble, pour prévoir les hauteurs d'eau à 24 heures au niveau de deux stations de prévision des crues en France situées dans le Grand Est. Les résultats montrent que les niveaux d'eau, et donc les crues, peuvent être anticipés 24 heures à l'avance en utilisant uniquement des données de hauteur d'eau et par l'intermédiaire de modèles d'apprentissage automatique simples appartenant aux modèles ensemblistes (*Random Forest* et *Gradient Boosting*). Ces derniers se révèlent performants, tout en restant accessibles aux non-spécialistes, tandis que les modèles plus complexes ne permettent pas d'obtenir des résultats supérieurs en l'absence de données météorologiques supplémentaires telles que les pluies. Ce travail ouvre la voie vers l'utilisation de l'apprentissage automatique et de l'intelligence artificielle pour la prévision des crues en France.

Keywords:

Forecasting
Flood
Water level
Machine learning
Artificial intelligence

ABSTRACT

Floods, a recurring hazard with devastating potential, can cause significant human and economic damage, and are likely to intensify with climate change. Their impacts underscore the need to anticipate floods by continually developing effective and cutting-edge tools for flood risk management. In France, flood forecasting relies on specific modeling tools based on physics, data, or a combination of both. Artificial intelligence (AI), currently expanding, offers new possibilities and perspectives for engineers and researchers in the field, but is not yet used nationally for flood forecasting. While complex machine learning algorithms show promise for flood prediction, their implementation and widespread use remain challenging for non-specialists. This article explores the use of six machine learning algorithms, including simple ensemble models, to forecast water levels 24 hours ahead at two flood forecasting stations in the Grand Est region of France. The results demonstrate that water levels, and consequently floods, can be anticipated using only water level data through simple machine learning models belonging to ensemble models (*Random Forest* and *Gradient Boosting*). These models perform well while remaining accessible to non-specialists, whereas more complex models do not yield superior results in the absence of additional meteorological data such as rainfall. This work paves the way for the use of machine learning and artificial intelligence in flood forecasting in France.

Introduction

Les inondations ont des effets néfastes importants sur la vie humaine, l'agriculture et l'économie, et constituent l'une des catastrophes naturelles les plus dévastatrices [MOSAVI *et al.*, 2018]. Des mesures existent

pour prévenir le risque inondation, notamment la surveillance des précipitations [LIU *et al.*, 2022], la mise en place d'ouvrages de rétention des crues [HADID *et al.*, 2019] ou encore l'infiltration à la parcelle [YANG *et al.*, 2023]. Cependant, le changement climatique, la croissance démographique, l'urbanisation et l'imperméabilisation des sols continuent d'accroître le risque

* Auteur correspondant – Courriel : nreiminger@air-d.fr

de tels événements [ZHENG et HUANG, 2023]. Il est donc d'autant plus nécessaire d'améliorer les stratégies de gestion du risque inondation et, parmi elles, la prévision des crues [DA SILVA et al., 2022].

La prévision des hauteurs d'eau est l'une des différentes possibilités disponibles pour anticiper les crues. En effet, en raison de son importance dans la gestion des inondations, cette prévision s'avère être un élément clé pour la manœuvre efficace et pertinente des ouvrages, ainsi que l'optimisation de la gestion de crise [SAPITANG et al., 2020]. L'objectif principal de la prévision des hauteurs d'eau est de prévoir avec précision l'arrivée des crues afin de minimiser les dommages causés par les inondations dans les zones concernées, ce qui fait de la fiabilité des prévisions une préoccupation majeure dans ce domaine [PHAM et al., 2020] : plus les résultats sont précis, plus ils sont bénéfiques, car des prévisions précises contribuent à une bonne gestion du risque ainsi qu'à des stratégies d'évacuation efficaces [XIE et al., 2017]. La grande variété de méthodes dédiées à la prévision des hauteurs d'eau peut être classée en deux catégories principales : les modèles basés sur la physique et les modèles basés sur la donnée [ZAKARIA et al., 2021]. Les modèles basés sur la physique reposent sur la compréhension des processus hydrologiques et leurs prévisions sont basées sur la reconstruction de ces processus. Néanmoins, ces modèles ont l'inconvénient d'être complexes et nécessitent une quantité importante d'informations au niveau de la zone concernée. Les modèles basés sur la donnée, en revanche, s'appuient sur les données historiques des variables liées aux inondations, telles que le niveau des rivières, le débit des cours d'eau ainsi que les précipitations, pour créer des modèles mathématiques pouvant être utilisés pour prévoir les hauteurs d'eau et, par conséquent, les crues. En France, notamment dans la région Grand Est, deux modèles sont actuellement utilisés par les autorités locales chargées de la prévision des inondations : le modèle GRP et le modèle LARSIM. Le modèle LARSIM (*Large Area Runoff Simulation Model*) est un modèle hydrologique qui décrit les processus d'écoulement continu dans les bassins versants et les réseaux hydrauliques, et qui modélise tous les aspects majeurs du cycle de l'eau terrestre [LEG, 2019; ISHIKAWA et al., 2021]. Il appartient donc à la catégorie des modèles basés sur la physique. À l'inverse, le modèle GRP (Génie Rural journalier pour la Prévision des crues) est utilisé pour la prévision à court terme et nécessite une base de données interne, alimentée par les réseaux de mesure disponibles [VIATGÉ et al., 2019; TIBERI-WADIER et al., 2021]. Ce dernier appartient quant à lui à la catégorie de modèles basés sur la donnée.

Dans la catégorie des modèles basés sur la donnée, les modèles d'apprentissage automatique (ML, *Machine Learning*) et d'intelligence artificielle (IA) ont montré un grand intérêt ces dernières années pour la prévision des crues dans le monde [MOSAVI et al., 2018; IBRAHIM et al., 2022]. Pour la prévision des hauteurs d'eau, en particulier, les chercheurs ont principalement utilisé des réseaux neuronaux artificiels (**ANN**, *Artificial Neural Network*) par le biais de perceptrons multicouches (**MLP**, *Multi-Layer Perceptron*) [AMNATSAN et al., 2018; ZEYNOUDDIN et al., 2020], ainsi que des systèmes de type **ANFIS** (*Adaptive Neuro Fuzzy Inference System*) [KALOOB et al., 2017; ÜNEŞ et al., 2019] ou **SVM** (*Support Vector Machine*) [KHAI et al., 2019; KIM et KIM, 2021], tandis que d'autres algorithmes populaires tels que les techniques d'apprentissage ensembliste utilisant le *Random Forest* (**RF**) ou le *Gradient Boosting* (**GB**), plus faciles à mettre en œuvre, ne sont presque jamais utilisés [WEE et al., 2021]. En outre, une des critiques majeures que l'on peut faire à ces études est la quantification des performances des modèles uniquement basée sur des critères statistiques classiquement utilisés en IA ou en ML, tels que le coefficient de détermination (R^2) [SHAMIM et al., 2016; PIRI et KAHKHA, 2017; CHEIFETZ et al., 2021], l'erreur quadratique moyenne (*RMSE*, *Root Mean Squared Error*) [ZHANG et al., 2018; ZHU et al., 2020] ou encore l'erreur absolue moyenne (*MAE*, *Mean Absolute Error*) [DAS et al., 2016; ZAJI et al., 2018]. L'utilisation de critères hydrologiques spécifiques, tels que le coefficient de Nash-Sutcliffe (*NSE*), n'apparaît que de façon marginale [IBRAHIM et al., 2022] tandis que d'autres critères, tels que le critère de persistance (*PI*, *Persistence Index*) ou le *CSI* (*Critical Success Index*), ne sont jamais mentionnés. Ce dernier point est très préoccupant, car certains modèles peuvent donner un excellent score R^2 mais un très mauvais *NSE* [KIM et KIM, 2021], ce qui peut donc fausser l'analyse de leur efficacité et précision d'un point de vue hydrologique.

L'objectif de la présente étude est d'évaluer les performances d'algorithmes d'apprentissage automatique de complexité variable pour la prévision des hauteurs d'eau à 24 heures. Une attention particulière est accordée aux algorithmes simples qui sont rarement, voire jamais utilisés à cette fin, tels que les techniques d'apprentissage ensembliste, incluant notamment le *Random Forest* (**RF**) et le *Gradient Boosting* (**GB**). Leurs performances sont comparées à celles d'algorithmes complexes couramment utilisés, tels que le **MLP** et le réseau de neurones récurrent (**RNN**, *Recurrent Neural Network*) à l'aide de critères statistiques classiques tels que *MAE* et *RMSE*, mais aussi des critères de perfor-

mance hydrologique spécifiques, notamment le *NSE*, le *PI* et le *CSI*. Ce travail présente un défi particulier en raison de l'absence volontaire d'informations sur les précipitations : seules des données sur les hauteurs d'eau sont utilisées. En outre, l'entraînement des divers algorithmes a été réalisé en considérant l'historique complet des hauteurs d'eau plutôt que les seuls événements de crue, ce qui permet une représentation plus large de la dynamique hydrologique par le modèle. Le contexte de l'étude, les détails quant aux algorithmes et paramètres utilisés pour prédire les hauteurs d'eau ainsi que le détail des critères pris en compte pour évaluer la performance des algorithmes sont exposés en section 1. Les résultats de l'étude et leur discussion sont quant à eux donnés en section 2.

1. Matériels et méthodes

1.1. Sites d'étude et caractéristiques hydrologiques

1.1.1. Site principal

L'étude a été menée sur une station de prévision des crues située dans la ville de Kogenheim (6 813 774,873/1 036 738,924 Lambert-93), dans la région Grand Est, présentée en figure 1A. Cette station est située en aval du bassin versant de l'III et a été installée en septembre 2006 pour suivre les niveaux d'eau de l'III, l'un des affluents du Rhin. La station de Kogenheim est une station dite de « propagation ». Les hauteurs d'eau à cet endroit sont donc principalement influencées par les écoulements provenant de l'amont plutôt que par les précipitations locales. L'III au niveau de Kogenheim est caractérisée par un débit moyen de 35,4 m³/s et des niveaux d'eau allant de 0,3 m à 2,5 m avec un niveau moyen d'environ 0,9 m entre 2007 et 2022. La répartition

des niveaux d'eau au cours de ces mêmes années est donnée en figure 1B.

Trois stations hydrométriques sont situées en amont de la station de Kogenheim, à savoir la station de Colmar (première station en amont de Kogenheim située sur l'III), la station d'Ostheim (située sur la Fecht, un affluent de l'III) et la station de Sélestat (située sur le Giessen, un autre affluent de l'III). Ces stations sont caractérisées par des débits moyens de 19,3 m³/s, 6,59 m³/s et 2,64 m³/s respectivement entre 2007 et 2022. Leur emplacement est également indiqué en figure 1A.

Enfin, chaque station de prévision des crues en France dispose de ses propres seuils d'alerte pour avertir les autorités d'un risque imminent d'inondation. Néanmoins, ces seuils ne sont pas accessibles au public. Dans le cadre de cette étude, on considère que le seuil d'alerte à la station de Kogenheim est $h_{alert} \geq 2$ m, ce qui correspond à 2 % du total des hauteurs d'eau enregistrées entre 2007 et 2022.

1.1.2. Site secondaire

Pour renforcer les conclusions de l'étude, des analyses complémentaires ont été réalisées sur une seconde station de prévision des crues située sur la Zorn à Waltenheim, à 70 km au nord de Kogenheim. Cette dernière, de type « propagation » et « pluie-débit », est davantage influencée par les précipitations que la station de Kogenheim.

La rivière Zorn à cet endroit est caractérisée par un débit moyen de 5,4 m³/s, et des niveaux d'eau allant de 0,3 m à 3,1 m avec un niveau moyen d'environ 0,6 m entre 2007 et 2020. Plus de détails sur la localisation de ces stations hydrométriques ainsi que sur la distribution des hauteurs d'eau à Waltenheim entre 2007 et 2020 sont donnés en annexe 1.

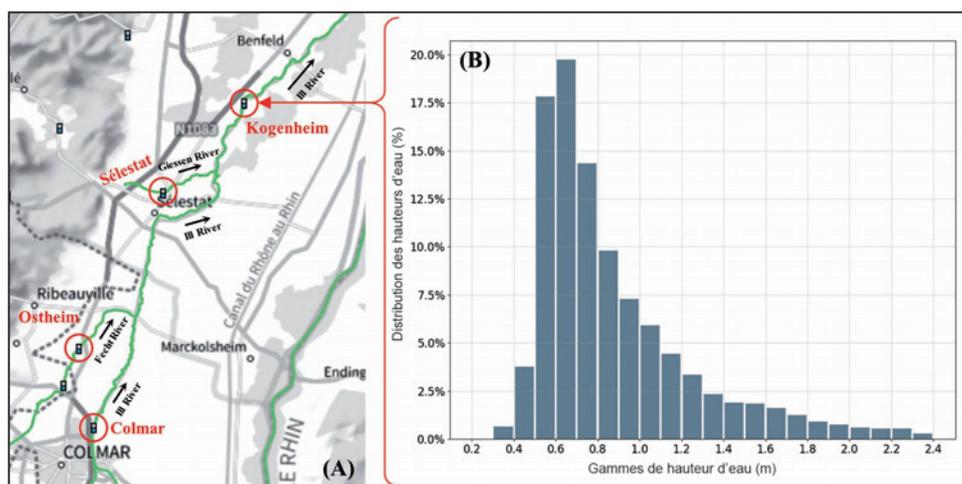


Figure 1. Site principal d'étude. (A) Localisation des stations hydrométriques (www.vigicrues.gouv.fr) et (B) distribution des hauteurs d'eau à Kogenheim entre 2007 et 2022

1.2. Architectures d'apprentissage automatique

1.2.1. Caractérisation des données et prétraitement

Les données utilisées dans le cadre de ce travail correspondent à des hauteurs d'eau horaires observées au niveau de stations hydrométriques bas-rhinoises. Ces données ont été obtenues par l'intermédiaire de l'application HydroPortail⁵ qui permet l'accès en ligne aux données hydrométrique. En particulier, les stations et périodes suivantes ont été considérées dans ce travail :

- site principal : Kogenheim, Ostheim, Sélestat et Colmar entre le 1^{er} janvier 2007 et le 31 décembre 2022;
- site secondaire : Waltenheim, Oberhof et Saverne entre le 1^{er} janvier 2007 et le 31 décembre 2019.

Aucun prétraitement de consolidation ou de nettoyage de la base de données ne fut nécessaire. En effet, les données utilisées correspondent aux données dites « vérifiées » disponibles sur HydroPortail. Aucune valeur manquante ou aberrante n'était à déplorer.

On notera enfin que certains algorithmes de *machine learning* ont nécessité un prétraitement de normalisation ou de standardisation de la base de données de hauteurs d'eau afin de faciliter leur entraînement et améliorer leurs prédictions. Les algorithmes en question sont précisés dans le *tableau I*.

⁵ <https://hydro.eaufrance.fr>

1.2.2. Algorithmes

Six différents algorithmes d'apprentissage supervisé ont été utilisés dans cette étude :

- la régression multilinéaire (**MLR**, *Multi-Linear Regression*) est une méthode statistique utilisée pour modéliser la relation entre plusieurs variables dépendantes ou indépendantes. L'objectif de cette approche est de trouver les coefficients de chaque variable indépendante qui permettent de prédire au mieux la valeur de la variable dépendante ;
- le *Random Forest* (**RF**) est une méthode d'apprentissage d'ensemble qui consiste à construire une multitude de prédicteurs faibles indépendants (arbres de décision) et à combiner leurs prédictions pour obtenir une prédiction moyenne ;
- le *Gradient Boosting* (**GB**) est une méthode d'apprentissage d'ensemble similaire au **RF**. Leur principale différence réside dans le fait que, pour le **GB**, chaque nouveau prédicteur cherche à corriger les erreurs de son prédécesseur en se concentrant davantage sur les cas qu'il a mal interprétés. Les prédicteurs ne sont donc pas indépendants puisqu'ils sont formés sur la base des résultats des prédicteurs précédents ;
- le **perceptron** multicouche (**MLP**) est un type de réseau de neurones artificiel (**ANN**) composé de plusieurs couches cachées de neurones interconnectés. Les données reçues dans la couche d'entrée sont traitées dans les couches cachées où chaque neurone

Algorithmes		Features scaling	Hyperparamètres	Fonction de perte
Nom	Librairie			
MLR	<i>LinearRegression</i>	STD	–	–
RF	<i>RandomForestRegressor</i>	–	n_estimators = 200 depth = 10	mse
BG	<i>GradientBoostingRegressor</i>	–	n_estimators = 200 depth = 10	mse
MLP	<i>SequentialDense</i>	–	6 couches <i>Dense</i> (150, 100, 50, 25, 10, 1 neurones) <i>Relu</i> (fonction d'activation), <i>Adam optimizer</i> batch_size = 100, epochs=10, learning_rate = 10 ⁻⁵	mae
RNN	<i>SequentialSimpleRNNDense</i>	NOR	1 couche <i>SimpleRNN</i> (100 neurones) and 3 <i>Dense</i> layers (25, 10, 1 neurons) <i>relu</i> activation function, <i>Adam optimizer</i> batch_size = 250, epochs=5, learning_rate = 10 ⁻⁴	mse
CNN	<i>SequentialConv1DMaxpooling1DDense</i>	NOR	1 couche <i>Conv1D</i> (64 filtres, kernel_size = 2) 1 couche <i>Maxpooling1D</i> (pool_size = 2) 2 couches <i>Dense</i> (50 and 1 neurones) <i>relu</i> (fonction d'activation), <i>Adam optimizer</i> batch_size = 250, epochs=5	mse

STD (resp. NOR) pour la standardisation (resp. normalisation) utilisant *StandardScaler* (resp. *MinMaxScaler*) de *scikit-learn* / mse : erreur quadratique moyenne / mae : erreur absolue moyenne. *Features scaling* : mise à l'échelle des caractéristiques.

Tableau I. Algorithmes et hyperparamètres retenus

applique une fonction d'activation non linéaire avant de la transmettre à la couche suivante pour finalement produire la sortie finale dans la couche de sortie ;

- le réseau de neurones récurrent (**RNN**) est un type d'**ANN** conçu pour traiter des données séquentielles, telles que des séries temporelles. Il diffère des autres réseaux neuronaux tels que le **MLP** par sa capacité à conserver les informations relatives aux entrées antérieures, ce qui lui permet de prendre en compte le contexte et les dépendances d'une entrée donnée par rapport aux entrées précédentes ;
- le réseau de neurones convolutif (**CNN**) est un type d'**ANN** principalement utilisé pour les tâches de traitement et de reconnaissance d'images, mais qui peut néanmoins également être utilisé pour tout besoin de traitement de séquences. Il est composé de plusieurs couches, notamment des couches convolutives, des couches de mise en commun et des couches entièrement connectées : les couches convolutives sont conçues pour apprendre les caractéristiques de l'entrée à l'aide d'un ensemble de filtres, qui sont ensuite convolués et passent par des couches de mise en commun pour réduire leur dimension et augmenter leur invariance avant d'être classées par les couches dites entièrement connectées.

Ces différents algorithmes ont été utilisés sous Python (v 3.7) par l'intermédiaire des bibliothèques *scikit-learn* (v 1.0.2), *Keras* (v 2.9.0) et *Tensorflow* (v 2.9.1). Les bibliothèques spécifiques, ainsi que les hyperparamètres et fonctions de perte utilisés qui ont conduit aux meilleurs résultats pour chaque algorithme sont répertoriés dans le *tableau I*. On notera que lorsqu'un hyperparamètre donné n'est pas mentionné, c'est le paramètre par défaut de la bibliothèque qui est utilisé.

1.2.3. Entraînement des modèles

Le jeu de données principal utilisé dans cette étude correspond à quatre chroniques de hauteurs d'eau relevées entre le 1^{er} janvier 2007 et le 31 décembre 2022, une pour chaque station considérée : Kogenheim, Colmar, Sélestat et Ostheim. Chaque chronique est composée de 140 256 données horaires de hauteur d'eau pour un total de 561 024 données. Ces données ont été obtenues via HydroPortail, une plateforme française rassemblant en libre accès les données hydrologiques au niveau national.

Lors de l'entraînement d'algorithmes d'apprentissage automatique, il est nécessaire de définir un jeu de données dit « d'entraînement » sur lequel les algorithmes vont s'ajuster automatiquement, un jeu de données dit de « validation » utilisé pour suivre l'évolu-

tion de l'entraînement, et un jeu de données de « test », utilisé en dernier lieu, pour évaluer les performances de l'algorithme sur des données qui n'ont pas servi à son entraînement. Il convient de noter que le choix a été fait d'entraîner les algorithmes non seulement sur les événements de crue, mais également sur toutes les autres données disponibles, y compris les périodes d'étiage, de moyennes eaux, etc. Les jeux d'entraînement, de validation et de test ont été répartis de la façon suivante :

- le jeu d'entraînement correspond aux 12 premières années de données (entre 2007 et 2018, soit 75 % des données totales). En considérant que le seuil d'alerte de crue à la station de Kogenheim est de $h_{alert} \geq 2$ m (voir *section 1.1*), le jeu d'entraînement contient donc 24 événements de crue ;
- le jeu de validation correspond aux 10 % des dernières données du jeu d'entraînement et comprend ainsi trois événements de crue. Ces données ne servent pas à entraîner le modèle, mais à suivre leur convergence ;
- le jeu de test correspond aux quatre dernières années de données (entre 2019 et 2022, soit 25 % des données totales) et comprend huit événements de crue.

On notera que les jeux d'entraînement et de test ont été divisés afin que les années les plus récentes soient utilisées pour tester les performances des modèles, dans le but d'évaluer leur capacité à réaliser des prédictions fiables à l'avenir. Enfin, la même répartition (25 % de données de test et 75 % de données d'entraînement, dont 10 % de données de validation) a été appliquée aux données correspondant au site secondaire de Waltenheim.

Deux paramétrages différents ont ensuite été considérés dans cette étude pour entraîner les algorithmes :

- paramétrage P1 : utilisation de l'historique des niveaux d'eau des stations en amont uniquement ;
- paramétrage P2 : utilisation de l'historique des niveaux d'eau des stations en amont ainsi que de celui de la station de prévision des crues ciblée.

Plus de précisions à propos de ces différents cas étudiés sont données dans le *tableau II*.

Enfin, un total de 12 combinaisons entre les algorithmes d'apprentissage automatique (**MLR**, **RF**, **GB**, **MLP**, **RNN** et **CNN**) et les paramétrages retenus (P1 et P2) a donc été considéré dans cette étude. Ces combinaisons seront ci-après désignées par « modèles ». Pour une meilleure compréhension, et à titre d'exemple, le modèle nommé « **MLR1** » correspond donc à l'algorithme **MLR** avec le paramétrage P1 tandis que le modèle nommé « **GB2** » correspond à l'algorithme **GB** avec le paramétrage P2, etc.

Paramétrage	Cas			
	Kogenheim		Waltenheim	
	Paramétrage 1	Paramétrage 2	Paramétrage 1	Paramétrage 2
Nombre de stations d'entrée	3	4	2	3
Stations d'entrée	Colmar, Sélestat et Ostheim	Kogenheim, Colmar, Sélestat et Ostheim	Oberhof et Saverne	Waltenheim, Oberhof et Saverne
Station cible	Kogenheim (t+24h)		Waltenheim (t+24h)	

Tableau II. Résumé des paramétrages étudiés

1.3. Critères de performance

Plusieurs critères ont été utilisés pour évaluer et comparer les performances des modèles. Cela inclut notamment des paramètres statistiques comparant les erreurs résiduelles entre les hauteurs d'eau observées et prévues (MAE, rMAE, RMSE et rRMSE), mais également des critères de performances plus consensuels dans le domaine de la prévision des crues (NSE, PI et CSI). Les équations de ces différents critères sont données en annexe 2 tandis que leurs valeurs cibles sont résumées dans le tableau III.

Pour les besoins de ce travail, un nouveau critère de performance appelé « taux de décalage d'alerte » (TDA) a été créé. Cette métrique est un indicateur des performances de détection du seuil d'alerte et donne des informations sur la capacité d'un modèle à prévoir en temps voulu un événement donné (par exemple, une hauteur d'eau considérée comme un seuil de crue pour déclencher une alerte), en tenant compte du temps de prévision souhaité par le modèle. Deux formulations de ce nouveau critère sont données en équations 1 et 2, incluant respectivement une version non bornée (TDA), mais aussi une version bornée (TDAb).

$$TDA = 1 - \frac{|\hat{t}_{seuil} - t_{seuil}|}{pp} \quad [\text{équation 1}]$$

$$TDAb = \begin{cases} 0, & \text{si } TDA < 0 \\ TDA, & \text{si } TDA \geq 0 \end{cases} \quad [\text{équation 2}]$$

avec TDAb et TDA les taux de décalage d'alerte borné et non borné, t_{seuil} le moment observé où le seuil d'alerte est atteint, \hat{t}_{seuil} le moment prédit où le seuil d'alerte est atteint, et pp la période de prévision.

Le TDA varie entre $-\infty$ et 1 et fournit des informations relatives au décalage de prévision du modèle. Une prévision en temps voulu donnerait une valeur de 1, tandis que 0 indique que la prévision par le modèle de la hauteur d'eau seuil est retardée par la valeur de la période de prévision du modèle (négativement ou positivement). Les valeurs négatives correspondent quant à elles à des prévisions retardées d'une valeur supérieure à la période de prévision du modèle. Le TDAb est la formulation bornée du TDA, variant entre 0 et 1. Une prévision en temps voulu donnerait une valeur de 1, tandis que 0 indique que la prévision du seuil par le modèle est retardée d'une valeur égale ou supérieure à la période de prévision du modèle (négativement ou positivement). À titre d'exemple, pour un événement de crue donné, si le dépassement du seuil d'alerte a été détecté avec 5 heures d'avance ou de retard en considérant un modèle de prévision à 24 heures (la prévision aurait donc lieu soit à t+19h soit à t+29h selon si le modèle est respectivement en retard ou en avance), on aurait $TDA = TDAb = 0,79$.

2. Résultats et discussion

2.1. Prévision de la dynamique globale d'évolution des hauteurs d'eau à Kogenheim

Les résultats des modèles ont d'abord été comparés sur l'ensemble du jeu de test (non utilisé pour leur entraînement) afin d'évaluer leurs performances dans la

Critère	Type	Valeurs		
		Minimale	Satisfaisante	Idéales
MAE	Absolu	$+\infty$	-	0
rMAE	Relatif	$+\infty$	-	0
RMSE	Absolu	$+\infty$	-	0
rRMSE	Relatif	$+\infty$	-	0
NSE	Relatif	$-\infty$	≥ 0	1
NSE ₂	Relatif	$-\infty$	≥ 0	1
NSE ₃	Relatif	$-\infty$	≥ 0	1
PI	Relatif	$-\infty$	≥ 0	1
CSI	Relatif	0	-	1
ALT	Relatif	$-\infty$	0	1
bALT	Relatif	0	-	1

Tableau III. Résumé des critères de performance utilisés et de leurs valeurs minimales, satisfaisantes et idéales

modélisation de la dynamique globale d'évolution des hauteurs d'eau à Kogenheim.

La figure 2 montre les niveaux d'eau prévus à Kogenheim 24 heures à l'avance par les différents modèles en comparaison des niveaux d'eau réellement observés. Sont distingués sur cette figure les résultats issus des deux paramétrages retenus (gauche : paramétrage n° 1/ droite : paramétrage n° 2) pour (A) une période de moyennes eaux et (B) une période de hautes eaux. Comme le montre cette figure, les modèles entraînés en utilisant le paramétrage n° 2 montrent globalement de meilleures performances que les modèles entraînés uniquement avec les stations en amont (paramétrage n° 1), aussi bien pour les moyennes eaux que les hautes eaux. L'historique des hauteurs d'eau à la station de

prévision apparaît donc comme étant une information pertinente pour améliorer la prévision des niveaux d'eau à 24 heures dans le cas de l'apprentissage automatique.

En comparant les résultats obtenus par les modèles un par un, il apparaît que les prévisions les moins bonnes sont obtenues par les algorithmes **MLR**, **RNN** et **CNN** pour les deux paramétrages considérés. Cependant, les prévisions faites par les modèles **MLR2**, **RNN2** et **CNN2** sont respectivement meilleures que celles faites par les modèles **MLR1**, **RNN1** et **CNN1**. La figure 2 montre enfin que les meilleures prévisions sont obtenues par les algorithmes **MLP**, **RF** et **GB**, et particulièrement par les modèles **MLP2**, **RF2** et **GB2**. Enfin, il est à noter que les modèles d'apprentissage d'ensemble utilisés dans ce travail (**RF2** et **GB2**) sont plus efficaces pour prévoir

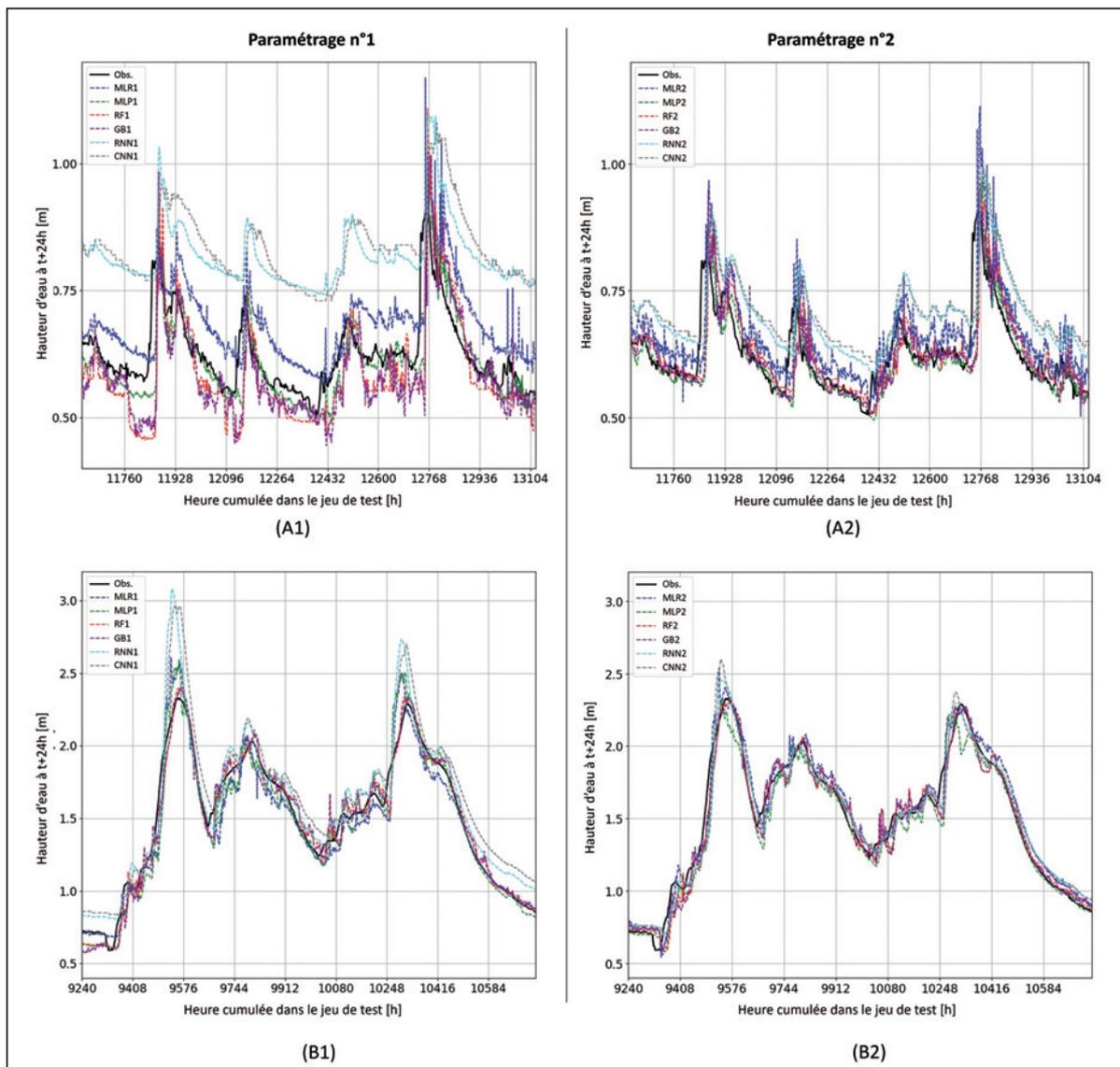


Figure 2. Comparaison entre les niveaux d'eau observés (ligne continue) et prévus (lignes en pointillé) à 24 heures à Kogenheim en fonction des heures cumulées dans l'ensemble de tests pour une période de moyennes eaux (A) et de hautes eaux (B) en distinguant les deux paramétrages retenus

les niveaux d'eau à 24 heures pendant les périodes de crue que les modèles **MLP2** et inversement.

Après cette analyse qualitative, les critères *MAE* et *RMSE* ainsi que leur formulation relative (*rMAE* et *rRMSE*) ont été calculés pour chaque donnée horaire dans l'ensemble de tests afin d'obtenir une comparaison quantitative des modèles. Les résultats moyens sont présentés en *figure 3*. Ces résultats renforcent les observations précédentes. En effet, avec un *rMAE* moyen de 23 % et 26 % respectivement, et un *rRMSE* moyen de 27 % et 29 % respectivement, les modèles **RNN1** et **CNN1** prévoient moins bien les hauteurs d'eau à 24 heures pour la dynamique hydrologique globale. Le meilleur est le modèle **MLP2** avec un *rMAE* moyen et un *rRMSE* moyen de 4,0 % et 6,4 % respectivement, légèrement devant les méthodes d'apprentissage par ensemble (modèles **GB2** et **RF2**) qui donnent un *rMAE* moyen d'environ 4,5 % et un *rRMSE* moyen d'environ 7 %. De plus, avec un *rMAE* moyen et un *rRMSE* moyen de 6,5 % et 8,0 % respectivement, le modèle **MLR2** performe mieux que ce qui a été qualitativement estimé précédemment à l'aide de la *figure 2*.

Enfin, lors de la comparaison des résultats obtenus par les modèles en fonction de leur paramétrage, les modèles utilisant le paramétrage n° 2 ont systématiquement généré de meilleurs *MAE*, *RMSE*, *rMAE* et *rRMSE* pour un même algorithme. Les valeurs obtenues avec le paramétrage n° 2 peuvent être jusqu'à deux fois meilleures qu'avec le paramétrage n° 1 en fonction de l'algorithme utilisé (par exemple, **GB2** est presque deux fois meilleur que **GB1** selon le critère *rMAE* : 4,5 % contre 8,7 % respectivement).

2.2. Prédiction des crues à Kogenheim

Après avoir comparé les résultats des modèles sur l'ensemble du jeu de test, l'intérêt s'est ensuite porté

plus particulièrement sur les résultats obtenus durant les événements de crue.

La *figure 4* permet de comparer les hauteurs d'eau réelles avec les hauteurs d'eau prévues à Kogenheim 24 heures à l'avance par les différents modèles entraînés en utilisant le paramétrage n° 2 et pour les huit événements de crue inclus dans les données du jeu de test (à noter que les *figures 4C* et *4E* incluent deux événements chacun). Comme le montre cette figure, tous les modèles semblent capables de reproduire les fluctuations des niveaux d'eau lors des crues. Cependant, en examinant de plus près les résultats, certaines divergences apparaissent entre les modèles. En effet, les modèles **MLR2**, **RNN2** et **CNN2** ont tendance à surestimer les pics de crue ainsi que les hauteurs d'eau lors de la décrue. Le modèle **MLP2**, quant à lui, tend à sous-estimer la plupart des pics de crue tandis que la décrue semble correctement prédite. Enfin, les modèles **RF2** et **GB2** semblent donner les meilleures prévisions, à la fois pour les pics de crue et les décrues.

Un point important à noter est qu'aucun des modèles ne parvient à reproduire correctement la montée des niveaux d'eau, quel que soit l'événement considéré. Tous les modèles sont en retard, un retard compensé ensuite plus ou moins rapidement selon le modèle. Ce manquement pourrait être la conséquence de l'absence d'informations sur les précipitations, qui est en réalité la principale source d'augmentation rapide des hauteurs d'eau, rendant ainsi plus difficile pour les modèles de prévoir la montée des niveaux d'eau au bon moment.

Pour appuyer les observations qualitatives précédentes, plusieurs critères de performance (*rMAE*, *rRMSE*, *NSE₁*, *NSE₂*, *NSE₃*, *PI*, *CSI*, *TDAb*, faux positifs et faux négatifs) ont été calculés pour chaque événement de crue dans

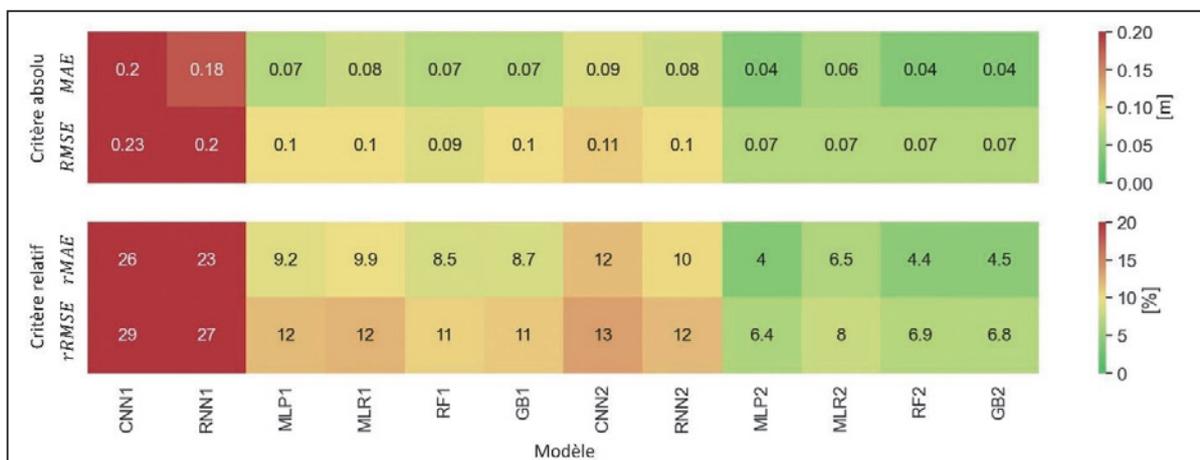


Figure 3. Carte thermique des *MAE*, *RMSE*, *rMAE* et *rRMSE*, et moyens obtenus sur l'ensemble du jeu de test de Kogenheim en fonction du modèle

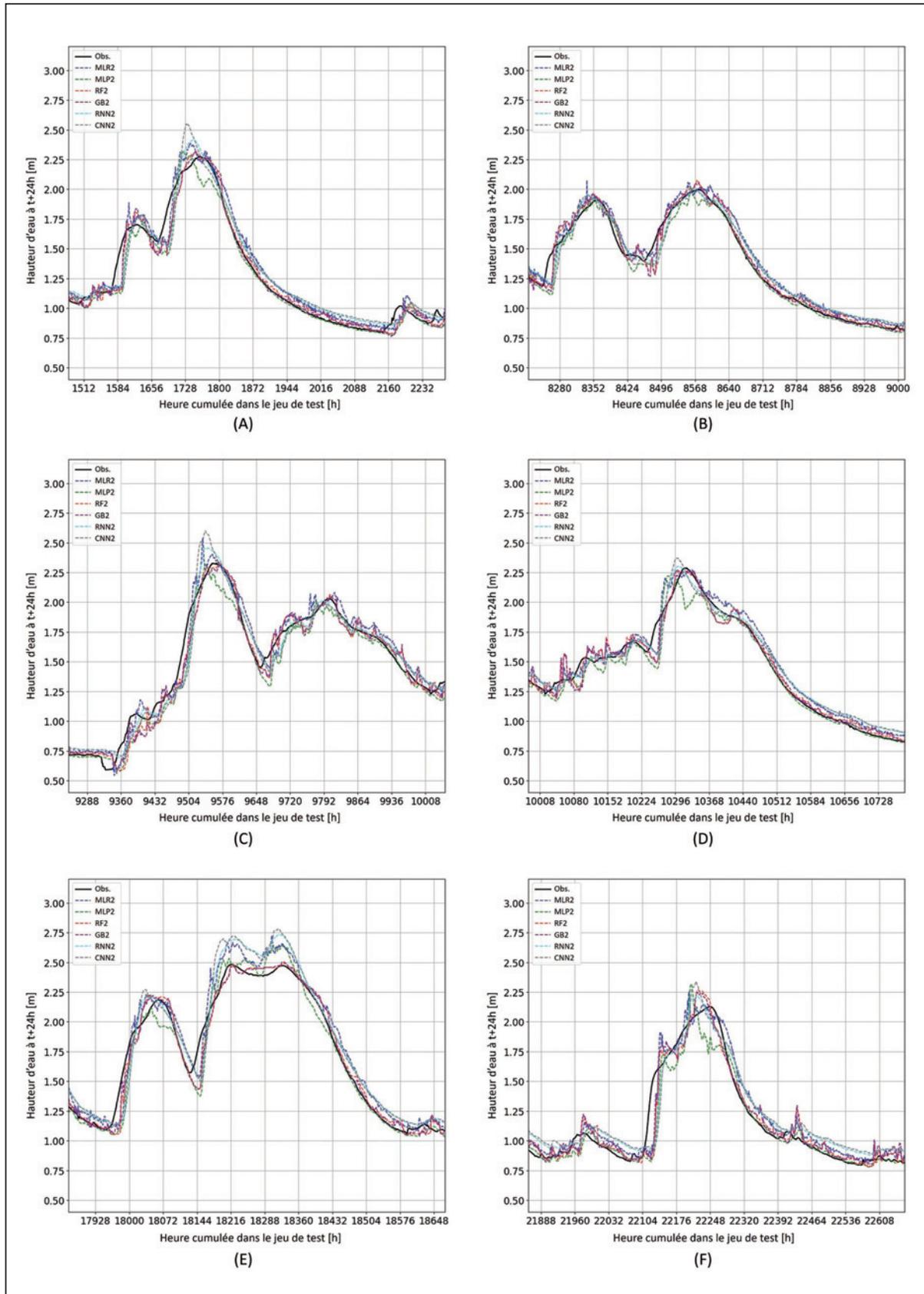


Figure 4. Comparaison entre les niveaux d'eau observés (ligne continue) et prévus (lignes en pointillé) à 24 heures à Kogenheim en fonction des heures cumulées pour les huit événements de crue présents dans le jeu de test (deux événements en C et E) et pour les modèles utilisant le paramétrage n° 2

l'ensemble de tests. Les résultats moyens sont présentés dans la *figure 5*, tandis que les résultats détaillés sont donnés en *annexe 3* (à l'exception du critère *CSI* qui ne peut pas être calculé pour des événements individuels). En premier lieu, et exception faite du modèle **CNN2** qui a commis un faux négatif sur huit événements (12,5 %), aucun des autres modèles n'a commis de faux négatifs. Cela est un bon point pour les modèles de prévision de crues où les faux négatifs doivent absolument être évités.

On observe également en *figure 5* que les modèles entraînés en utilisant le paramétrage n° 2 sont, encore une fois, systématiquement meilleurs que ceux entraînés en utilisant le paramétrage n° 1, ce qui soutient les résultats obtenus précédemment en *section 2.1*. Ce dernier point est très clair lorsqu'on compare les modèles **MLR**, **MLP**, **RNN** et **CNN** entre eux selon divers critères de performance (NSE_1 , NSE_2 , NSE_3 , PI et faux positifs), mais beaucoup moins évident lorsqu'il s'agit de comparer les

modèles **RF** et **GB**. Il est important de noter que, pour ces deux derniers modèles, seuls les faux positifs et deux critères de performance hydrologique (le *Critical Success Index*, généralement inutilisé dans la littérature scientifique, et le Taux de Décalage des Alertes, construit pour cette étude) montrent des différences suffisamment significatives pour suggérer que **RF2** et **GB2** sont respectivement meilleurs que **RF1** et **GB1**. Cela illustre particulièrement l'importance de considérer plusieurs critères de performances pour qualifier la performance des modèles hydrologiques basés sur l'apprentissage automatique, car ces modèles n'auraient pas pu être différenciés uniquement sur la base de paramètres statistiques usuels tels que $rMAE$ et $rRMSE$. Enfin, avec des coefficients d'efficacité de Nash-Sutcliffe (NSE_2) de 0,77 et 0,80 pour **RF2** et **GB2** respectivement, ces deux modèles peuvent être qualifiés de bons et très bons respectivement [MORIASI et al., 2015].

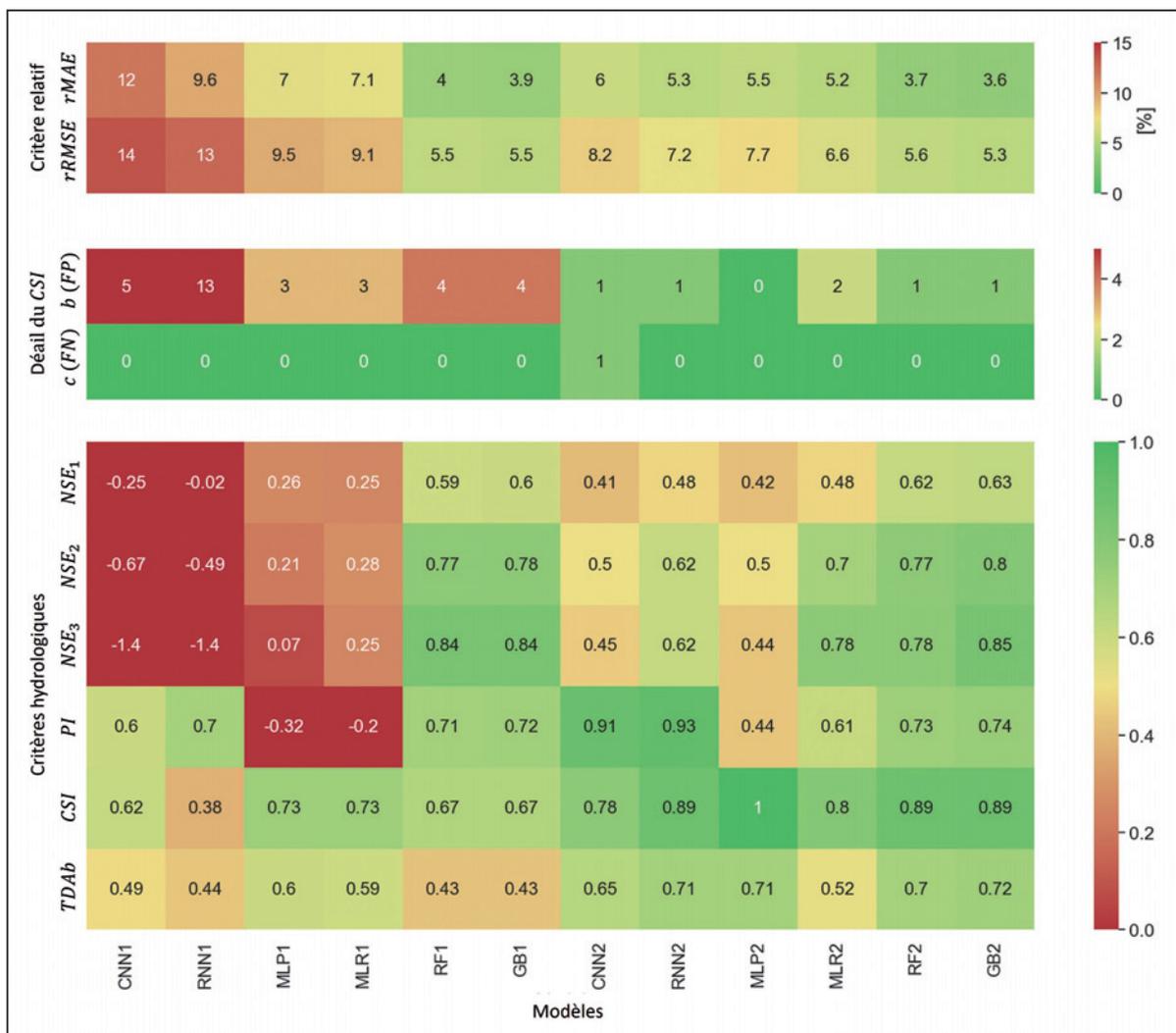


Figure 5. Cartes thermiques des $rMAE$, $rRMSE$, NSE_1 , NSE_2 , NSE_3 , PI , CSI , $TDA b$ obtenus, moyennés sur les huit événements de crue de l'ensemble de tests de Kogenheim, et nombre de faux négatifs (FN) et positifs (FP) obtenus selon le modèle

L'utilisation de nombreux critères de performance soulève également un problème important pour déterminer le ou les meilleurs modèles. En effet, il est difficile de comparer directement les résultats présentés dans la figure 5 pour classer les modèles en fonction de leur score global sur tous les critères, car ces critères n'évoluent pas de la même manière ni vers le même optimum, et il est difficile de savoir dans quelle mesure deux valeurs sont statistiquement différentes. Ainsi, la méthodologie développée par FAWAZ *et al.* [2019] a été utilisée pour résoudre ce problème et obtenir un classement robuste des modèles. Cette méthodologie permet de comparer différents modèles en les classant selon leur performance sur un critère donné appliqué à des sous-ensembles de données distincts correspondant ici à chaque événement de crue (un score de 1 étant le meilleur résultat). Ce classement peut ensuite être utilisé pour établir des diagrammes de différence critique, comme présenté en figure 6 pour les critères $rMAE$, $rRMSE$, NSE_1 , NSE_2 , NSE_3 , PI et $TDAb$ en considérant le paramétrage n° 2. Un classement global est également donné dans cette figure en considérant tous ces critères simultanément.

Une première observation que l'on peut faire est que, pour quatre des sept critères évalués en figure 6 (NSE_1 , NSE_2 , $rMAE$ et $rRMSE$), deux grands clusters apparaissent avec, d'une part, les modèles **MLR2**, **MLP2**, **RNN2** et **CNN2** et, d'autre part, les modèles d'ensemble **RF2** et **GB2**, ces derniers obtenant les meilleurs résultats. En considérant le critère NSE_3 , les modèles **RF2** et **GB2**

restent les meilleurs, mais partagent cette position avec **MLR2** et, un peu plus loin, avec **RNN2**. Pour le critère PI , les meilleurs résultats sont obtenus par **RNN2**, suivi de **CNN2**, puis par **RF2** et **GB2** simultanément, suivi de **MLR2** et enfin de **MLP2**. Lors de la comparaison des modèles sur la base du $TDAb$, aucune tendance spécifique ne peut être dégagée puisque les résultats ne sont pas distinguables selon la méthodologie du diagramme de différence critique. Ainsi, conformément à ces observations et en constatant également les résultats obtenus en considérant ces sept métriques simultanément, **RF2** et **GB2** sont les meilleurs modèles pour la prévision des crues à Kogenheim sur la seule base de l'utilisation des hauteurs d'eau comme données d'entrée.

2.3. Vérification des résultats précédents pour la station de Waltenheim

Tous les modèles évalués dans cette étude correspondent à des modèles d'apprentissage automatique. Leurs résultats sont donc fortement influencés par le jeu de données, tant en termes de quantité (historique des données) que de qualité (valeurs manquantes ou fausses, précision, incertitudes, etc.), ainsi que par le choix des entrées données (nombre de stations en amont, paramètres d'apprentissage, etc.). Ainsi, pour éviter les défauts de reproductibilité de ces résultats et, par conséquent, afin de fiabiliser les conclusions, les différents modèles étudiés ont également été testés sur une autre station de prévision des crues (Waltenheim) en considérant le paramétrage n° 2 ayant permis

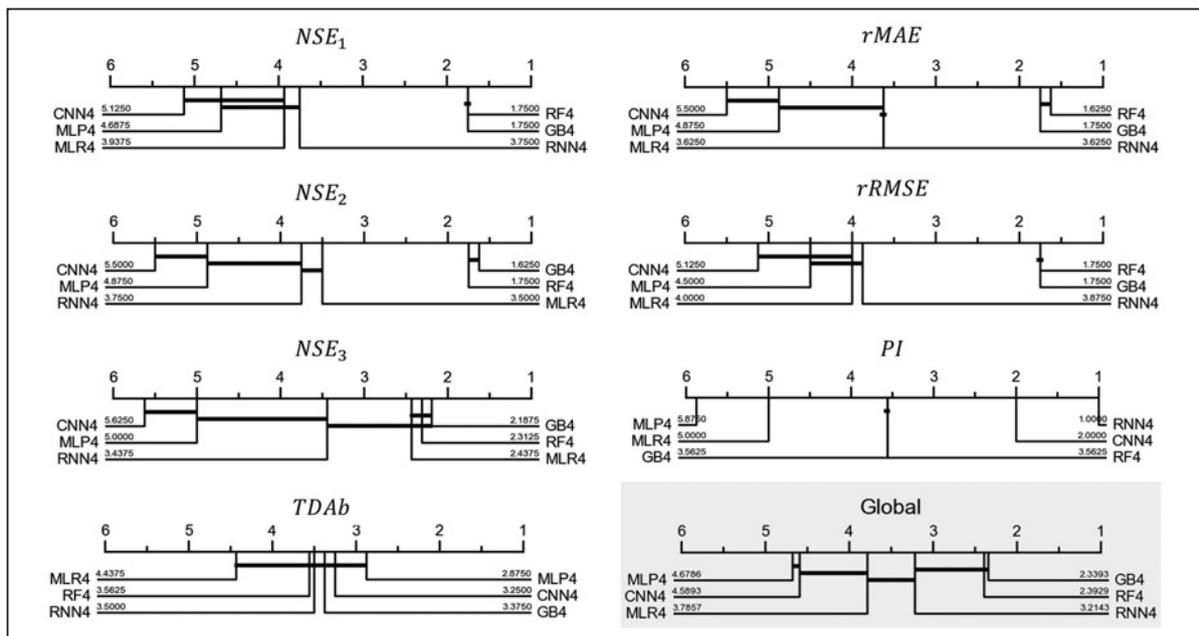


Figure 6. Diagrammes de différence critique classant les performances des modèles issus du paramétrage n° 2 sur la station de prévision de Kogenheim selon divers critères de performance statistiques ($rMAE$ et $rRMSE$) et hydrologiques (NSE_1 , NSE_2 , NSE_3 , PI et $TDAb$) ainsi que l'ensemble des critères simultanément (global)

d'obtenir les meilleurs résultats de prévision des hauteurs d'eau à Kogenheim. Les diagrammes de différence critique obtenus pour les critères $rMAE$, $rRMSE$, NSE_1 , NSE_2 , NSE_3 , PI et $TDAb$ ainsi qu'en considérant tous ces critères simultanément sont donnés dans la figure 7. D'après les résultats obtenus, les modèles RF2 et GB2 restent les meilleurs en considérant les critères NSE_1 , NSE_2 , $rMAE$, et $rRMSE$. Ils partagent cette place avec MLR2 qui était précédemment le troisième meilleur modèle sur ces critères. Pour les autres critères, de légères différences apparaissent, mais les méthodes d'ensemble (RF2 et GB2) restent principalement les meilleures, à l'exception de RF2 en considérant le critère PI , où RNN2 et CNN2 le devançant. D'un point de vue général, et en considérant tous les critères simultanément, GB2 est le meilleur modèle pour la prévision des hauteurs d'eau à Waltenheim.

2.4. Limitations

Les résultats de cette étude démontrent que l'utilisation de modèles d'ensemble simples, tels que les algorithmes RF ou GB, plus faciles d'accès pour un public non spécialiste de l'apprentissage automatique, donnent des résultats satisfaisants tant pour la prévision des crues que pour la prévision globale de l'évolution des hauteurs d'eau. Ces algorithmes surpassent d'autres algorithmes plus complexes comme le RNN, surtout lorsqu'ils sont entraînés en tenant compte des niveaux

d'eau historiques observés au niveau de la station cible de prévision en tant qu'entrée supplémentaire, en plus des niveaux d'eau mesurés en amont. Dans ce cas, de faibles erreurs sont obtenues quant à l'évolution globale des hauteurs d'eau, avec environ 5 % d'erreur en moyenne et environ 7 % d'erreur quadratique moyenne. Une très bonne performance en matière de prévision des crues est également obtenue, conduisant à des coefficients de Nash-Sutcliffe d'environ 0,8, ce qui est du même ordre de grandeur, voire mieux, que ceux obtenus par des modèles hydrologiques complexes tels que SWAT [KATEB et al., 2019] ou LARSIM [ISHIKAWA et al., 2021] nécessitant beaucoup plus de données d'entrée, tant en qualité qu'en quantité. Enfin, des valeurs de PI de 0,73 et de CSI d'environ 0,9, sans aucun faux négatif sur huit événements de crue et seulement un faux positif, témoignent également de la bonne performance de ces algorithmes pour la prévision des crues.

Ces performances ont été obtenues en entraînant les algorithmes non spécifiquement sur les événements de crue, mais sur l'ensemble de l'historique des niveaux d'eau et en ne considérant aucune donnée additionnelle telles que la pluie, l'humidité des sols, etc. Ce dernier point est probablement à l'origine d'un retard dans le début de la montée des eaux systématiquement observé avant les pics de crue sur les deux ensembles de données. Cela ne pose pas de problème spécifique dans

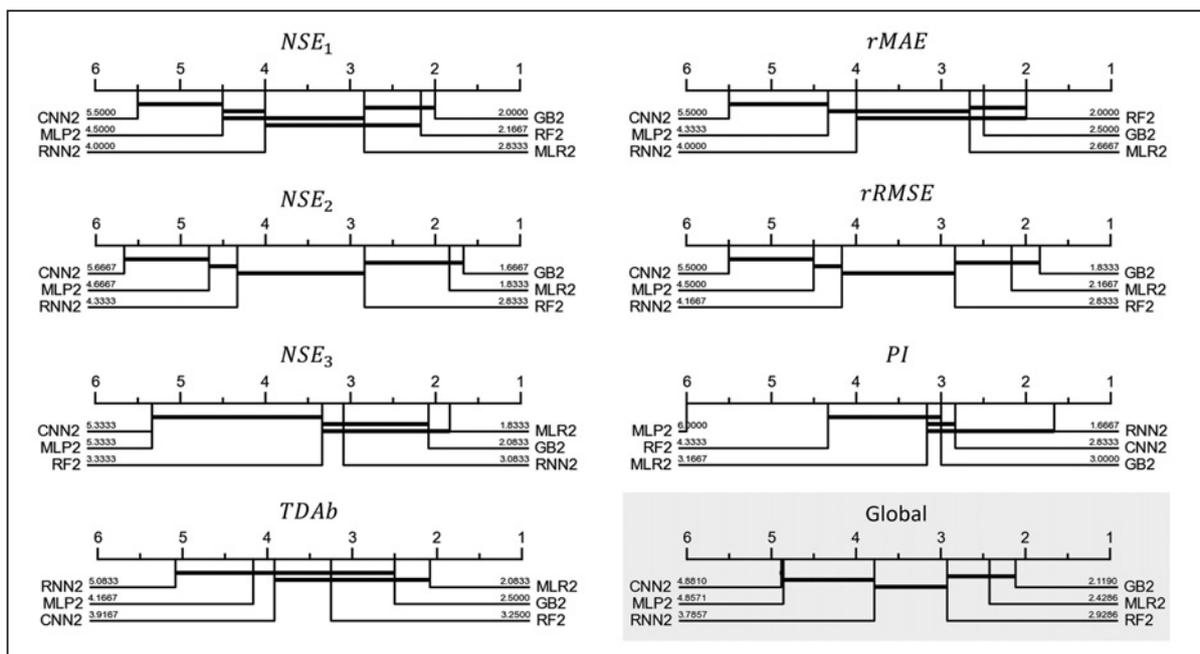


Figure 7. Diagrammes de différence critique classant les performances des modèles issus du paramétrage n° 2 sur la station de prévision de Waltenheim selon divers critères de performance statistique ($rMAE$ et $rRMSE$) et hydrologiques (NSE_1 , NSE_2 , NSE_3 , PI et $TDAb$) ainsi que l'ensemble des critères simultanément (global)

ce cas, car les modèles compensent ensuite ce retard, et tous les événements de crue inclus dans l'ensemble de tests sont modélisés avec précision. Cependant, cela n'est probablement pas systématique, par exemple pour des stations de prévision des crues très différentes en matière d'hydrologie. Une certaine vigilance est donc à conserver, et l'utilisation et l'optimisation de modèles pour la prévision des crues basés sur l'apprentissage automatique doivent être réalisées au cas par cas.

Conclusion

Six algorithmes d'apprentissage automatique, à savoir **MLR**, **MLP**, **RF**, **GB**, **RNN** et **CNN**, ont été entraînés selon deux paramétrages donnés, résultants en 12 modèles pour prédire les hauteurs d'eau 24 heures à l'avance. Les résultats obtenus par ces algorithmes au niveau d'une station de prévision des crues (Kogenheim) ont été présentés et discutés, puis confirmés sur une station additionnelle (Waltenheim), renforçant ainsi les conclusions. Les résultats de l'étude ont démontré que les crues et, plus généralement, les hauteurs d'eau, peuvent être pré-

vues avec précision 24 heures à l'avance en se basant uniquement sur les données des niveaux d'eau, sans besoin de données de précipitations. L'utilisation de modèles d'apprentissage automatique simples, plus faciles d'accès aux non-spécialistes, a prouvé son intérêt en procurant des résultats très satisfaisants. Par exemple, l'application de méthodes d'apprentissage par ensemble, telles que le *Gradient Boosting* et le *Random Forest*, a donné des coefficients d'efficacité de Nash-Sutcliffe d'environ 0,8, des indices de persistance de 0,73 et des indices de succès critique d'environ 0,9. À l'inverse, l'utilisation de modèles d'apprentissage automatique plus complexes, tels que le Perceptron Multi-Couche (MLP) ou les Réseaux de Neurones Récurrents (RNN), n'a pas amélioré les résultats tandis qu'ils nécessitent une expertise accrue pour leur mise en œuvre. Le manque de précision de ces algorithmes complexes pourrait provenir du manque de données d'entrée supplémentaires, telles que les niveaux de précipitation ou les taux de perméabilité et de saturation des sols. Ce travail ouvre la voie vers l'utilisation de l'apprentissage automatique et de l'intelligence artificielle pour la prévision des crues en France.

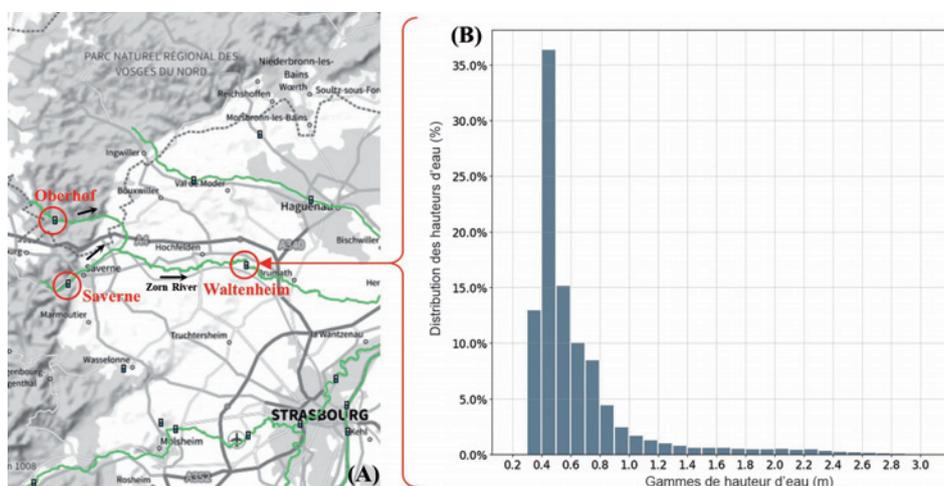
Bibliographie

- AMNATSAN S., YOSHIKAWA S., KANAE S. (2018) : « Improved forecasting of extreme monthly reservoir inflow using an analogue-based forecasting method: A case study of the Sirikit Dam in Thailand ». *Water* ; 10(11) : 1614.
- CHEIFETZ N., GUÉRY L., DELABRE K., HEIM V. (2021) : « Anticipation de crues sur la Seine, la Marne et l'Oise pour protéger la production d'eau potable : Anticipating rises in the level of Seine, Marne and Oise to protect waterproduction plants ». *Techniques Sciences Méthodes* ; 11 : 45-52.
- DA SILVA L.B.L., ALENCAR M.H., DE ALMEIDA A.T. (2022) : « A novel spatiotemporal multi-attribute method for assessing flood risks in urban spaces under climate change and demographic scenarios ». *Sustainable Cities and Society* ; 76(3-4) : 103501.
- DAS M., GHOSH S. K., CHOWDARY V. M., SAIKRISHNAVENI A., SHARMA R. K. (2016) : « A probabilistic nonlinear model for forecasting daily water level in reservoir ». *Water Resources Management* ; 30(9) : 3107-22.
- FAWAZ H.I., FORESTIER G., WEBER J., IDOUMGHAR L., MULLER P.-A. (2019) : « Deep learning for time series classification: a review ». *Data Mining and Knowledge Discovery* ; 33(4) : 917-63.
- HADID B., DUVELLA E., CHIRON P., ARCHIMÉDE B. (2019) : « A flood mitigation control strategy based on the estimation of hydrographs and volume dispatching ». *IFAC-PapersOnLine* ; 52(23) : 17-22.
- IBRAHIM K.S.M.H., HUANG Y.F., AHMED A.N., KOO C.H., EL-SHAFIE A. (2022) : « A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting ». *Alexandria Engineering Journal* ; 61(1) : 279-303.
- ISHIKAWA M., HAAG I., KRUMM J., TELTSCHER K., LORKE A. (2021) : « The effect of stream shading on the inflow characteristics in a downstream reservoir ». *River Research and Applications* ; 37(7) : 943-54.
- KALOOP M.R., EL-DIASTY M., HU J.W. (2017) : « Real-time prediction of water level change using adaptive neuro-fuzzy inference system ». *Geomatics, Natural Hazards and Risk* ; 8(2) : 1320-32.
- KATEB Z., BOUCHEKIA H., BENMANSOUR A., BELARBI F. (2019) : « Hydrological modeling using the SWAT model based on two types of data from the watershed of Beni Haroun dam, Algeria ». *Journal of Water and Land Development* ; 43(1) : 76-89.
- KHAI W.J., ALRAIH M., AHMED A.N., FAI C.M., EL-SHAFIE A., EL-SHAFIE A. (2019) : « Daily forecasting of dam water levels using machine learning ». *International Journal of Civil Engineering and Technology* ; 10(6) : 314-23.
- KIM C., KIM C.-S. (2021) : « Analysis of AI-based techniques for forecasting water level according to rainfall ». *Tropical Cyclone Research and Review* ; 10(4) : 223-8.
- LEG L.-E. (2019) : « Das Wasserhaushaltsmodell LARSIM - Modellgrundlagen und Anwendungsbeispiele ». Disponible en ligne : <http://www.larsim.info/dokumentation/LARSIM-Dokumentation.pdf>
- LIU J., CHO H.-S., OSMAN S., JEONG H.-G., LEE K. (2022) : « Review of the status of urban flood monitoring and forecasting in TC region ». *Tropical Cyclone Research and Review* ; 11(2) : 103-19.
- MORIASI D. N., GITAU M. W., PAI N., DAGGUPATI P. (2015) : « Hydrologic and water quality models: Performance measures and evaluation criteria ». *Transactions of the ASABE* ; 58(6) : 1763-85.

- MOSAVI A., OZTURK P., CHAU K. (2018) : « Flood prediction using machine learning models: Literature review ». *Water* ; 10(11) : 1536.
- PHAM Q.B., AFAN H.A., MOHAMMADI B., AHMED A.N., LINH N.T.T., VO N. D., et al. (2020) : « Hybrid model to improve the river streamflow forecasting utilizing multi-layer perceptron-based intelligent water drop optimization algorithm ». *Soft Computing* ; 24(23) : 18039-56.
- PIRI J., KAHKHA M.R.R. (2017) : « Prediction of water level fluctuations of Chahnimeh reservoirs in zabol using ANN, ANFIS and cuckoo optimization algorithm ». *Iranian Journal of Health, Safety and Environment* ; 4(2) : 706-15.
- SAPITANG M., M. RIDWAN W., FAIZAL KUSHIAR K., NAJAH AHMED A., EL-SHAFIE A. (2020) : « Machine learning application in reservoir water level forecasting for sustainable hydro-power generation strategy ». *Sustainability* ; 12(15) : 6121.
- SHAMIM M.A., HASSAN M., AHMAD S., ZEESHAN M. (2016) : « A comparison of artificial neural networks (ANN) and local linear regression (LLR) techniques for predicting monthly reservoir levels ». *KSCCE Journal of Civil Engineering* ; 20(2) : 971-7.
- TIBERI-WADIER A.-L., GOUTAL N., RICCI S., SERGENT P., VIATGÉ J., MONTEIL C., BOUTTIER F. (2021) : « Comparative global sensitivity analysis on the hydrologic models GRP and MORDOR-TS ». *Hydroscience Journal* ; 107(1) : 1-11.
- ÜNEŞ F., DEMIRCI M., TAŞAR B., KAYA Y.Z., VARÇIN H. (2019) : « Estimating Dam Reservoir Level Fluctuations Using Data-Driven Techniques ». *Polish Journal of Environmental Studies* ; 28(5) : 3451-62.
- VIATGÉ J., PINNA T., FICCHÍ A., PERRIN C., DORCHIES D., GARANDEAU L., TILMANT F. (2019) : « Vers une plus grande flexibilité temporelle du modèle opérationnel de prévision des crues GRP ». *La Houille Blanche* ; 105(2) : 72-80.
- WEE W. J., ZAINI N.B., AHMED A.N., EL-SHAFIE A. (2021) : « A review of models for water level forecasting based on machine learning ». *Earth Science Informatics* ; 14(4) : 1707-28.
- XIE K., OZBAY K., ZHU Y., YANG H. (2017) : « Evacuation zone modeling under climate change: A data-driven method ». *Journal of Infrastructure Systems* ; 23(4) : 04017013.
- YANG W., ZHANG J., KREBS P. (2023) : « Investigating flood exposure induced socioeconomic risk and mitigation strategy under climate change and urbanization at a city scale ». *Journal of Cleaner Production* ; 387: 135929.
- ZAJI A. H., BONAKDARI H., GHARABAGHI B. (2018) : « Reservoir water level forecasting using group method of data handling ». *Acta Geophysica* ; 66(2): 717-730.
- ZAKARIA M.N.A., ABDUL MALEK M., ZOLKEPLI M., NAJAH AHMED A. (2021) : « Application of artificial intelligence algorithms for hourly river level forecast: A case study of Muda River, Malaysia ». *Alexandria Engineering Journal* ; 60(4) : 4015-28.
- ZEYNODDIN M., BONAKDARI H., EBTEHAJ I., AZARI A., GHARABAGHI B. (2020) : « A generalized linear stochastic model for lake level prediction ». *The Science of the Total Environment* ; 723: 138015.
- ZHANG X., LIU P., ZHAO Y., DENG C., LI Z., XIONG M. (2018) : « Error correction-based forecasting of reservoir water levels: Improving accuracy over multiple lead times ». *Environmental Modelling & Software* ; 104 : 27-39.
- ZHENG J., HUANG G. (2023) : « Towards flood risk reduction: Commonalities and differences between urban flood resilience and risk based on a case study in the Pearl River Delta ». *International Journal of Disaster Risk Reduction* ; 86 : 103568.
- ZHU S., HRNJICA B., PTAK M., CHOIŃSKI A., SIVAKUMAR B. (2020) : « Forecasting of water level in multiple temperate lakes using machine learning models ». *Journal of Hydrology* ; 585 : 124819.

Annexe 1

Site secondaire d'étude



A) Localisation des stations hydrométriques (www.vigicrues.gouv.fr) ; B) distribution des hauteurs d'eau à Waltenheim entre 2007 et 2022

Annexe 2

Critères d'évaluation et de comparaison des performances des modèles

- **Erreur moyenne absolue (MAE, Mean Absolute Error) et sa version relative :**

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad [\text{équation 2.1}]$$

$$rMAE = \frac{1}{n} \sum_{i=1}^n rAE_1 = \frac{1}{n} \sum_{i=1}^n \frac{|O_i - P_i|}{O_i} \quad [\text{équation 2.2}]$$

avec MAE l'erreur moyenne absolue, rAE_j l'erreur relative pour la $j^{\text{ème}}$ prédiction, $rMAE$ l'erreur moyenne absolue relative à la moyenne des observations, n le nombre de paires d'observations et de prédiction dans le jeu de données, O_i l' $i^{\text{ème}}$ observation et P_i l' $i^{\text{ème}}$ prédiction.

- **Erreur quadratique moyenne (RMSE, Root Mean Squared Error) et sa version relative :**

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad [\text{équation 2.3}]$$

$$rRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(O_i - P_i)^2}{O_i^2}} \quad [\text{équation 2.4}]$$

avec RMSE l'erreur quadratique moyenne, $rRMSE$ l'erreur quadratique moyenne relative à la moyenne des observations, n le nombre de paires d'observations et de prédiction dans le jeu de données, O_i l' $i^{\text{ème}}$ observation et P_i l' $i^{\text{ème}}$ prédiction.

- **Coefficient de Nash-Sutcliffe (NSE, Nash-Sutcliffe Efficiency coefficient) et sa généralisation :**

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2} \quad [\text{équation 2.5}]$$

$$NSE_j = 1 - \frac{\sum_{i=1}^n |O_i - P_i|^j}{\sum_{i=1}^n |O_i - \bar{O}_i|^j} \quad [\text{équation 2.6}]$$

avec NSE le coefficient de Nash-Sutcliffe, NSE_j le coefficient de Nash-Sutcliffe généralisé à $j^{\text{ème}}$ degré, n le nombre de paires d'observations et de prédiction dans le jeu de données, O_i l' $i^{\text{ème}}$ observation et P_i l' $i^{\text{ème}}$ prédiction.

- **Indice de persistance (PI, Persistence Index) :**

$$PI = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - O_{i-1})^2} \quad [\text{équation 2.7}]$$

avec PI l'indice de persistance, n le nombre de paires d'observations et de prédiction dans le jeu de données, O_i l' $i^{\text{ème}}$ observation, O_{i-1} la $(i-1)^{\text{ème}}$ observation et P_i la $i^{\text{ème}}$ prédiction.

- **Indice critique de succès (CSI, Critical Success Index) :**

$$CSI = \frac{a}{a + b + c} \quad [\text{équation 2.8}]$$

avec CSI l'indice critique de succès, a le nombre de vrais positifs, b le nombre de faux positifs et c le nombre de faux négatifs.



DOMAINES
D'INTERVENTION MULTITECHNIQUE

**Conception - Etudes - Installations
Mise en service**

Hydromécaniques - Électromécaniques

Électricité HT / BT - Automatismes

Serrurerie - Manutention industrielle

ACTEMIUM Paris Hydro Process

conçoit et réalise des solutions techniques
dans le domaine de l'eau

01 30 09 63 00 - CHATOU

2 avenue de l'Europe 78400 CHATOU

01 58 42 51 61 – CRETEIL

2 chemin des Marais 94046 Créteil



Expertises et savoir-faire en :

- Stations de pompage
- Equipements électromécaniques associés
- Electricité, automatismes, supervision
- En eaux usées, potables et industrielles

En tant qu'ensemblier, **ACTEMIUM** intervient de la conception jusqu'à la mise en service.

Annexe 3.1

Résultats détaillés des critères de performance obtenus pour chacun des huit événements de crue de l'ensemble de tests (E1 à E8) en considérant le paramétrage n° 1 (Moy. : résultat moyen sur l'ensemble des événements de crue)

Modèle	Événement	Critère								
		NSE ₁	NSE ₂	NSE ₃	MAE	rMAE [%]	RMSE	rRMSE [%]	PI	TDA _b
MLR1	E1	0,54	0,68	0,74	0,11	5,54	0,15	7,45	0,53	0,96
	E2	0,62	0,80	0,87	0,06	3,37	0,08	4,91	0,67	0,17
	E3	0,47	0,66	0,76	0,14	7,48	0,18	9,37	0,58	0,83
	E4	0,04	0,19	0,27	0,11	6,29	0,13	7,49	-0,71	0,17
	E5	0,12	0,21	0,23	0,13	6,88	0,17	8,41	-0,22	0,42
	E6	0,27	0,35	0,38	0,12	6,94	0,17	10,19	0,62	0,63
	E7	-0,31	-0,74	-0,99	0,31	13,28	0,37	15,71	-3,61	0,96
	E8	0,24	0,12	-0,29	0,13	6,71	0,18	9,58	0,52	0,58
	Moy.	0,25	0,28	0,25	0,14	7,06	0,18	9,14	-0,20	0,59
MLP1	E1	0,48	0,58	0,63	0,12	6,09	0,17	8,23	0,39	0,92
	E2	0,44	0,52	0,58	0,09	4,98	0,12	7,41	0,22	0,25
	E3	0,53	0,7	0,79	0,13	6,83	0,17	9,11	0,63	1,00
	E4	0,28	0,49	0,61	0,08	4,69	0,10	5,93	-0,08	0,29
	E5	0,30	0,23	0,12	0,11	5,39	0,16	8,22	-0,19	0,38
	E6	0,37	0,48	0,57	0,11	5,98	0,15	9,00	0,70	0,33
	E7	-0,41	-1,10	-1,66	0,34	14,29	0,41	17,24	-4,57	0,83
	E8	0,12	-0,20	-1,10	0,15	7,76	0,21	11,1	0,34	0,83
	Moy.	0,26	0,21	0,07	0,14	7,00	0,19	9,53	-0,32	0,60
RF1	E1	0,60	0,79	0,86	0,09	4,71	0,12	6,09	0,69	0,25
	E2	0,58	0,78	0,87	0,06	3,64	0,08	4,87	0,64	0,00
	E3	0,70	0,88	0,95	0,08	4,48	0,10	5,78	0,85	0,58
	E4	0,57	0,81	0,91	0,05	2,97	0,06	3,88	0,60	0,75
	E5	0,67	0,83	0,88	0,05	2,70	0,08	4,18	0,73	0,79
	E6	0,37	0,59	0,73	0,11	5,84	0,13	7,67	0,76	0,58
	E7	0,74	0,86	0,90	0,06	3,07	0,11	5,52	0,63	0,33
	E8	0,51	0,65	0,62	0,08	4,44	0,11	6,43	0,81	0,13
	Moy.	0,59	0,77	0,84	0,07	3,98	0,10	5,55	0,71	0,43
GB1	E1	0,60	0,78	0,84	0,09	4,70	0,12	6,33	0,67	0,29
	E2	0,59	0,79	0,86	0,06	3,60	0,08	4,85	0,65	0,00
	E3	0,73	0,89	0,95	0,07	4,06	0,10	5,56	0,87	0,63
	E4	0,60	0,82	0,91	0,05	2,74	0,06	3,78	0,62	0,58
	E5	0,68	0,85	0,89	0,05	2,56	0,07	3,96	0,76	0,92
	E6	0,40	0,62	0,76	0,10	5,55	0,13	7,40	0,78	0,67
	E7	0,72	0,87	0,92	0,07	3,25	0,10	5,24	0,66	0,33
	E8	0,46	0,60	0,60	0,09	4,88	0,12	6,76	0,78	0,04
	Moy.	0,60	0,78	0,84	0,07	3,92	0,10	5,49	0,72	0,43
RNN1	E1	0,13	-0,37	-1,47	0,20	9,64	0,30	14,1	0,59	1,00
	E2	0,41	0,68	0,83	0,09	5,10	0,10	5,70	0,93	0,00
	E3	0,15	-0,34	-1,41	0,23	11,09	0,35	16,36	0,80	0,79
	E4	0,31	0,58	0,75	0,08	4,57	0,09	5,40	0,90	0,21
	E5	0,02	-0,44	-1,48	0,15	7,51	0,22	10,87	0,67	0,46
	E6	-0,34	-1,15	-2,28	0,23	11,71	0,31	15,60	0,81	0,04
	E7	-0,43	-1,35	-2,47	0,34	14,62	0,43	18,43	0,07	1,00
	E8	-0,45	-1,53	-3,89	0,24	12,76	0,30	15,53	0,83	0,00
	Moy.	-0,02	-0,49	-1,43	0,20	9,63	0,26	12,75	0,70	0,44
CNN1	E1	-0,08	-0,3	-0,74	0,25	12,35	0,30	13,95	0,61	0,79
	E2	0,17	0,44	0,66	0,13	7,35	0,13	7,84	0,87	0,00
	E3	-0,16	-0,47	-1,04	0,32	15,53	0,37	17,38	0,78	0,96
	E4	0,19	0,44	0,63	0,09	5,42	0,11	6,30	0,87	0,13
	E5	-0,10	-0,26	-0,56	0,17	8,44	0,21	10,14	0,71	0,63
	E6	-0,62	-1,06	-1,31	0,28	14,19	0,30	15,15	0,82	0,33
	E7	-1,17	-3,32	-6,47	0,52	22,32	0,59	24,74	-0,71	0,88
	E8	-0,22	-0,81	-1,99	0,20	10,81	0,25	13,32	0,88	0,25
	Moy.	-0,25	-0,67	-1,40	0,25	12,05	0,28	13,60	0,60	0,49

LA MEILLEURE FAÇON DE DÉCOUPER VOS TUBES EN TOUTE SÉCURITÉ !

- Coupe de la fonte ductile, plastique, PEHD ...
- Aucun effet rebond
- Meilleur contrôle de la coupe
- Sécurité maximale
- Excavation minimale
- Réduction de l'effort de l'opérateur



4 modèles disponibles :

695XL-PG

680ES-PG

890F-PG

701A-PG



Votre sécurité, notre priorité

Rue Marie Curie - Z.I Mitry Compans - 77290 Mitry Mory - Tel : 01 60 21 64 00
atdv@atdv.fr - <https://atdv.fr>

Plus d'info ici



Spécialistes du traitement de l'eau

Vendeurs **Installateurs** **Mainteneurs**

- **Secteurs :** particulier, collectif, tertiaire, industriel, CHR (café, hôtels, restaurants)
- **Prestations :** analyses d'eau, matériel de traitement, consommables et produits de traitement

Maintenance / Installation / Dépannage
Désembouage / Conditionnement /
Désinfection Réseaux fermés (EC/EG)
Réseaux sanitaires Habilitation
disconnecteurs

Régions : Centre, IdF, Vendée

Contact : Tél. : 01 69 83 83 27 - @ : contact@i-w-s.fr

PRIMUS LINE® REHAB GAINE DE RÉHABILITATION FLEXIBLE

La seule gaine pour réhabiliter
les réseaux d'eau avec Avis Technique.



PRIMUS LINE

SAFE.RELIABLE.SUSTAINABLE.

www.primusline.com

Annexe 3.2

Résultats détaillés des critères de performance obtenus pour chacun des huit événements de crue de l'ensemble de tests (E1 à E8) en considérant le paramétrage n° 2 (Moy. : résultat moyen sur l'ensemble des événements de crue)

Modèle	Événement	Critère								
		NSE ₁	NSE ₂	NSE ₃	MAE	rMAE [%]	RMSE	rRMSE [%]	PI	TDAb
MLR2	E1	0,57	0,78	0,87	0,10	5,34	0,12	6,53	0,68	0,92
	E2	0,57	0,81	0,92	0,07	3,94	0,08	4,72	0,70	0,13
	E3	0,60	0,81	0,90	0,11	6,08	0,13	7,57	0,76	0,83
	E4	0,47	0,75	0,89	0,06	3,62	0,07	4,29	0,47	0,00
	E5	0,45	0,71	0,84	0,08	4,47	0,10	5,40	0,55	0,42
	E6	0,34	0,55	0,69	0,11	6,36	0,14	8,44	0,73	0,25
	E7	0,44	0,73	0,89	0,13	6,21	0,15	6,90	0,28	1,00
	E8	0,42	0,42	0,21	0,10	5,52	0,14	8,65	0,68	0,58
	Moy.	0,48	0,70	0,78	0,10	5,19	0,12	6,56	0,61	0,52
MLP2	E1	0,51	0,66	0,71	0,11	5,71	0,15	7,74	0,50	0,83
	E2	0,61	0,80	0,89	0,06	3,36	0,08	4,74	0,67	0,46
	E3	0,57	0,75	0,84	0,12	6,20	0,15	8,32	0,69	1,00
	E4	0,47	0,66	0,73	0,06	3,52	0,09	4,99	0,28	0,33
	E5	0,36	0,39	0,37	0,10	4,90	0,15	7,30	0,05	0,50
	E6	0,28	0,34	0,37	0,12	6,76	0,17	10,16	0,61	0,83
	E7	0,62	0,80	0,87	0,09	4,16	0,13	6,24	0,46	0,88
	E8	-0,07	-0,41	-1,22	0,18	9,57	0,22	12,28	0,23	0,88
	Moy.	0,42	0,50	0,44	0,11	5,52	0,14	7,72	0,44	0,71
RF2	E1	0,67	0,82	0,87	0,08	4,13	0,11	5,87	0,74	0,46
	E2	0,68	0,85	0,90	0,05	2,91	0,07	4,32	0,75	0,33
	E3	0,75	0,90	0,95	0,07	3,83	0,10	5,61	0,87	0,58
	E4	0,58	0,81	0,91	0,05	2,84	0,06	3,77	0,61	0,83
	E5	0,65	0,79	0,83	0,05	2,85	0,09	4,64	0,67	1,00
	E6	0,49	0,68	0,76	0,09	4,89	0,12	7,39	0,81	0,79
	E7	0,77	0,91	0,94	0,06	2,67	0,09	4,50	0,75	0,63
	E8	0,37	0,40	0,07	0,10	5,69	0,15	8,50	0,67	0,96
	Moy.	0,62	0,77	0,78	0,07	3,73	0,10	5,58	0,73	0,70
GB2	E1	0,68	0,83	0,88	0,07	3,95	0,11	5,80	0,74	0,50
	E2	0,70	0,86	0,90	0,05	2,71	0,07	4,19	0,77	0,67
	E3	0,72	0,87	0,93	0,08	4,19	0,11	6,14	0,84	0,58
	E4	0,53	0,78	0,89	0,05	3,18	0,07	4,05	0,54	0,75
	E5	0,64	0,78	0,81	0,06	2,90	0,09	4,74	0,66	0,96
	E6	0,55	0,77	0,88	0,08	4,26	0,10	6,04	0,87	0,79
	E7	0,76	0,90	0,94	0,06	2,74	0,09	4,65	0,73	0,67
	E8	0,43	0,58	0,57	0,09	5,17	0,12	6,96	0,77	0,88
	Moy.	0,63	0,80	0,85	0,07	3,64	0,10	5,32	0,74	0,72
RNN2	E1	0,55	0,71	0,78	0,10	5,43	0,14	7,18	0,91	0,63
	E2	0,64	0,83	0,91	0,06	3,37	0,07	4,60	0,96	0,50
	E3	0,55	0,73	0,81	0,12	6,72	0,16	8,79	0,96	0,92
	E4	0,61	0,83	0,92	0,05	2,65	0,06	3,64	0,96	0,71
	E5	0,59	0,77	0,84	0,06	3,27	0,09	4,73	0,95	0,79
	E6	0,24	0,29	0,32	0,13	7,38	0,18	10,54	0,94	0,54
	E7	0,32	0,58	0,78	0,16	7,23	0,18	8,19	0,84	0,75
	E8	0,31	0,18	-0,41	0,12	6,47	0,17	10,18	0,94	0,88
	Moy.	0,48	0,62	0,62	0,10	5,32	0,13	7,23	0,93	0,71
CNN2	E1	0,47	0,56	0,57	0,12	6,33	0,17	8,60	0,87	0,71
	E2	0,60	0,76	0,83	0,06	3,75	0,09	5,47	0,94	0,63
	E3	0,44	0,57	0,62	0,15	8,05	0,20	10,55	0,94	0,92
	E4	0,55	0,77	0,87	0,05	3,04	0,07	4,12	0,95	0,00
	E5	0,530	0,67	0,74	0,07	3,69	0,11	5,51	0,92	0,79
	E6	0,2	0,25	0,29	0,14	7,72	0,18	10,7	0,93	0,42
	E7	0,28	0,48	0,67	0,17	7,64	0,20	9,08	0,80	0,79
	E8	0,19	-0,09	-0,96	0,13	7,49	0,19	11,56	0,93	0,96
	Moy.	0,41	0,50	0,45	0,11	5,96	0,15	8,20	0,91	0,65