Sujet de thèse CIFRE CNR (2025 – 2028)

Intégration de l'expertise humaine dans les prévisions d'ensemble hydrométéorologiques

Entreprise : CNR
Lieu principal : 2 rue André Bonin, 69004 Lyon
Laboratoire : INRAE HYCAR, 1 rue Pierre-Gilles de Gennes CS 10030 92761 Antony CEDEX
Ecole doctorale : Sorbonne Université, ED 398 (Ecole Doctorale Géosciences, Ressources Naturelles et Environnement)
Salaire : 31 180 bruts annuels, soit 2398,47 euros bruts / mois sur 13 mois + 2 primes variables + TR
Candidature (CV, lettre de motivation) : l.caillouet@cnr.tm.fr (+ copie a maria-helena.ramos@inrae.fr)

Contexte

CNR (Compagnie Nationale du Rhône) est le premier producteur français d'électricité d'origine 100% renouvelable (eau, vent, soleil) et le concessionnaire à vocation multiple du Rhône de la frontière suisse à la Méditerranée : production d'hydroélectricité, développement du transport fluvial, agriculture durable et environnement pour un développement durable des territoires.

Du fait de sa nature, la production de CNR dépend fortement des conditions météorologiques, ce qui a conduit au développement d'un ensemble d'outils de prévision hydrométéorologique. Ces outils sont utilisés au quotidien par les prévisionnistes CNR pour établir des prévisions expertisées dans le but d'optimiser la gestion des actifs et la programmation de la production hydro-électrique.

Au vu des enjeux opérationnels, la nécessité de quantifier les différentes incertitudes a rendu indispensable l'usage des prévisions probabilistes. A CNR, cette approche a été adoptée en intégrant un ensemble de méthodologies et d'outils pour améliorer les processus de prise de décision. Chaque jour, un ensemble de scénarios météorologiques et hydrologiques sont crées afin de prévoir au mieux l'évolution des précipitations et des débits dans les jours à venir, en de multiples points sur le territoire Rhodanien [Caillouet 2022]. Cette approche ensembliste est également adoptée par d'autres services tels que le SCV (ex SCHAPI) [Demargne 2021], qui prévoit d'ici 2030 d'exploiter son système de prévision des crues "Vigicrues" en version probabiliste [Godet 2023]. Une revue mondiale de systèmes ensemblistes hydrologiques est visible dans [Troin 2021], ainsi qu'une discussion plus ancienne dans [Cloke 2009].

Bien que les chaines élaborées fournissent des prévisions pertinentes, celles-ci peuvent être affinées grâce à l'expertise humaine [Célié et al., 2019]. Ainsi, la chaîne de prévision déterministe de CNR permet aux prévisionnistes d'ajouter leur expérience à chaque étape de la chaîne de prévision. Cependant, le cadre probabiliste rend cette expertise plus difficile, en raison de la multitude de données à prendre en compte, qui recouvrent plusieurs échéances et zones d'études. Présenter des données utilisables pour une analyse rapide par les prévisionnistes représente alors un vrai défi de mise en forme et de création d'outils.

A CNR, l'une des pistes étudiées pour résumer les ensembles hydrologiques et météorologiques est celle du clustering, déjà utilisée par certains centres de prévisions, tels que l'ECMWF [Straus et al., 2007]. Le but est alors de regrouper des scénarios prévus par groupes d'évolution type, permettant de passer de plusieurs dizaines de traces en maximum 6 groupes d'évolution. La question de la modification de ces ensembles (bruts ou résumés par clusters) et des conséquences sur leurs propriétés statistiques reste une question ouverte dont le sujet a commencé à être traité lors d'un stage.

Objectifs de la thèse

L'objectif est de progresser dans la prévision des précipitations et des débits en combinant différentes sources d'informations sur leur évolution : prévision numérique, expertise humaine et observations, afin d'alimenter un système de prévision hydrologique qui modélise les incertitudes par des ensembles. Concrètement, on étudiera comment corriger de manière optimale les prévisions d'ensemble météorologiques et hydrologiques court terme (J+4). Cela inclura l'ajout de l'expertise humaine qui permettra notamment de conserver la cohérence entre ces prévisions et celles issues des autres outils (prévisions moyen terme, prévisions court terme déterministes), tout en visant de conserver les propriétés statistiques de l'ensemble intactes (comme la fiabilité). Ce problème lie trois questions :

- Sous quelle forme produire les ensembles en vue de leur modification ultérieure ? Cette partie reprendra des études sur le clustering et affinera les méthodes et moyens mis en place.
- Quelle est la meilleure manière d'intervenir sur une prévision d'ensemble ? Différentes techniques seront étudiées et évaluées sur de grands échantillons, comme la sélection de clusters, leur pondération ou des modifications des volumes prévus.
- Comment optimiser les performances des prévisions d'ensemble ainsi modifiées ? On étudiera comment ces interventions impactent les qualités objectives des prévisions hydrométéorologiques, notamment leur fiabilité et leur valeur décisionnelle.

Sujet détaillé

L'***objectif principal*** de cette thèse est de concevoir une maquette d'un système temps réel dans lequel des prévisionnistes humains pourraient intervenir sur une prévision d'ensemble afin d'exprimer leur expertise. L'objectif opérationnel de cette modification peut dépasser l'amélioration directe des débits prévus, puisqu'il s'agit d'obtenir des modifications pertinentes sur les précipitations également, tout en respectant la cohérence inhérente entre les pluies et les débits. On peut voir cela comme une problématique d'interaction humain/machine : comment outiller un humain pour modifier une donnée aussi complexe qu'une prévision d'ensemble ? Les études sur le sujet [Fundel 2019, Demuth 2020] insistent sur la nécessité de co-construire les systèmes de prévision d'ensemble avec leurs usager(e)s, pour une bonne prise en main opérationnelle. Une étude préliminaire a déjà été menée à CNR par Mesure [2023], notamment sur l'intérêt d'un clustering pour faciliter la manipulation d'un ensemble, idée qui sera reprise et améliorée dans cette thèse.

Une première difficulté est la caractérisation d'interventions humaines, typiquement des corrections spécifiées par des expert(e)s prévisionnistes. En l'absence d'un mode opératoire standardisé, il faut **identifier des corrections-types**, par exemple le choix d'un (ou plusieurs) scénario privilégié, ou un souhait d'amplifier ou atténuer des pluies ou débits prévus par les modèles. Il faut ensuite décider **comment traduire numériquement ces corrections**, en construisant un nouvel ensemble, tout en conservant la cohérence physique entre les variables manipulées. Il s'agit d'un problème ouvert, les procédures documentées d'expertise d'ensembles météorologiques se limitant à la production de décision d'alerte ou de produits assez simples [Ramos 2010, Fundel 2019]. Cela soulève déjà des questions complexes sur la chaîne de valeur allant des modèles numériques à l'interprétation humaine de leurs prévisions [Golding 2022]. Les publications scientifiques sur la correction humaine d'ensembles sont très rares [Celie 2019]. Cette thèse s'intéressera pour la première fois à la production d'un ensemble corrigé complet, notamment en termes de débits.

Une seconde difficulté est l'optimisation de la qualité de l'ensemble ainsi produit. Cela dépend de la manière dont l'ensemble d'entrée a été élaboré, notamment la qualité des modèles calculant les membres et les perturbations utilisées pour y introduire de la dispersion, qui ne seront pas remises

en cause dans la thèse. Cet ensemble peut avoir été calibré (c'est à dire corrigé statistiquement par apprentissage sur des données passées, par exemple avec des réseaux de neurones ou autres) afin d'optimiser certains aspects de sa fiabilité statistique pertinents pour les applications branchées en aval. On souhaite comprendre comment les corrections humaines modifient ces propriétés de l'ensemble, et, si elles les dégradent, d'identifier une **technique de correction d'ensemble qui soit un compromis entre la correction humaine et ces propriétés statistiques**.

Cela peut être vu comme un problème d'assimilation de données dans un ensemble, sur lequel des études ont été déjà menées, par exemple l'algorithme de "filtre à particule" [Van Leeuwen 2019] : l'expertise humaine est ici l'équivalent d'une observation, et l'on doit fabriquer la meilleure prévision d'ensemble possible en la combinant avec une ébauche d'ensemble a priori. Un travail complémentaire gagnera donc à être mené sur l'**utilisation d'observations pour corriger une prévision d'ensemble de précipitations :** c'est une approche prometteuse pour la prévision d'ensemble à courte échéance des précipitations intenses, notamment par identification de précurseurs dans les corrélations ensemblistes pour construire des corrections qui aient des chances d'être des améliorations des prévisions [Bouttier 2024].

Profil souhaité

La thèse se déroulera à CNR à la suite d'un cursus Master M2 ou équivalent. Elle nécessitera les compétences suivantes :

- Capacité à réaliser un projet scientifique supervisé dans une équipe d'ingénieurs d'études, à présenter les résultats à l'écrit et à l'oral.
- Connaissances en modélisation numérique météorologique ou hydrologique, en statistiques, machine learning ou data science. Une expérience sur la prévision d'ensemble ou probabiliste sera un avantage.
- Maîtrise du langage python ou R en autonomie pour traitement scientifique de données et leur visualisation (outils actuels en RShiny).
- Bonnes compétences en anglais : un objectif de la thèse est la rédaction d'un article scientifique en anglais, ainsi que des présentations en conférences.

Des déplacements de l'ordre d'une semaine tous les deux mois seront prévus dans le laboratoire d'accueil (INRAE HYCAR). **Une fois le candidat sélectionné, le dossier doit être envoyé à l'ANRT pour validation, ce qui prend entre 2 et 3 mois.**

Références

- Bellier, J., Zin, I., Bontron, G. 2017: Sample stratification in verification of ensemble forecasts of continuous scalar variables: Potential benefits and pitfalls. Monthly Weather Review, 145(9), 3529-3544. http://dx.doi.org/10.1175/MWR-D-16-0487.1.
- Bontron, G. (2004). Prévision quantitative des précipitations : adaptation probabiliste par recherche d'analogues ; utilisation des réanalyses NCEP/NCAR et application aux précipitations du sud-est de la France. Thèse de Doctorat. Grenoble INP. https://theses.hal.science/tel-01090969
- Bouttier, F., M. Mandement, 2024: Vers une anticipation à 1-6h des risques de pluies intenses quasi-stationnaires. Accepté pour publication dans La Houille Blanche le 26 février 2024. https://doi.org/10.1080/27678490.2024.2325698 et https://hal.science/hal-04482783

- Caillouet L., S. Celie, O. Vannier, G. Bontron, S. Legrand, 2022 : Operational hydrometeorological forecasting on the Rhône River in France: moving toward a seamless probabilistic approach. LHB, 108:1, 2061312. https://doi.org/10.1080/27678490.2022.2061312
- Celie, S., Bontron, G., Ouf, D., & Pont, E. (2019). Apport de l'expertise dans la prévision hydro-météorologique opérationnelle. La Houille Blanche, 105(2), 55–62. https://doi.org/10.1051/lhb/2019015
- Cloke, H.L., Pappenberger, F., 2009: Ensemble flood forecasting: A review. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2009.06.005
- Demargne, J., Fouchier, C., Organde, D., Piotte, O., Belleudy, A., 2021: Advances and challenges of the French operational flash flood warning system, Vigicrues Flash, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-16063, https://doi.org/10.5194/egusphere-egu21-16063
- Demuth, J., R. Morss, I. Jankov, T. Alcott, C. Alexander, D. Nietfeld, T. Jensen, D. Novak, S. Benjamin, 2020: recommendations for developing useful and usable convection-allowing model ensemble information for NWS forecasters. Wea. For. 35:1381-1406. https://doi.org/10.1175/WAF-D-19-0108.1
- Fundel VJ, Fleischhut N, Herzog SM, Goeber M, Hagedorn R., 2019: Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. Quart. J. Roy. Met. Soc. 145:210-231. https://doi.org/10.1002/qj.3482
- Godet J., O. Payrastre, P. Javelle, and F. Bouttier, 2023: Assessing the ability of a new seamless short-range ensemble rainfall product to anticipate flash floods in the French Mediterranean area. Nat. Haz. Earth. Sys. Sci. 23, 3355–3377, https://doi.org/10.5194/egusphere-2023-907
- Golding B. (Ed.), 2022: towards the perfect weather warning. Bridging discplinary gaps through partnership and communication. WMO Book, United Nations Office for Disaster Risk Reduction, Springer Open Access. 277pp.
- Marchal, H., F. Bouttier, O. Nuissier, 2025 : Is considering runs (in)consistency so useless for weather forecasting ? EGUsphere preprint nhess-2024-208, submitted on 7 Nov 2024. https://nhess.copernicus.org/preprints/nhess-2024-208/
- Mesure, T. 2023: Intégration du clustering dans une chaîne de prévision hydrométéorologique, optimisation du nombre de clusters et impact de l'expertise. Rapport de stage ingénieur IENM3, encadrants L. Caillouet, O. Vannier, S. Celie, G. Bontron, CNR.
- Ramos, M.-H., Mathevet, T., Thielen, J. and Pappenberger, F., 2010: Communicating uncertainty in hydro-meteorological forecasts: mission impossible? Met. Apps, 17: 223-235. https://doi.org/10.1002/met.202
- Straus, D. M., S. Corti, and F. Molteni, 2007. Circulation regimes: Chaotic variability versus sst-forced predictability. Journal of Climate 20: 2251-2272.
- Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J.-L. (2021). Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years. Water Resources Research, 57, e2020WR028392. https://doi.org/10.1029/2020WR028392
- Van Leeuwen PJ, Künsch HR, Nerger L, Potthast R, Reich S. Particle filters for high-dimensional geoscience applications: A review. Q J R Meteorol Soc. 2019; 145: 2335–2365. https://doi.org/10.1002/qj.3551