



Contrat post-doctoral sur la

Simulation jointe de la qualité des eaux et du débit de petits cours d'eau sous changement climatique

Durée : 24 mois

Début : 1^{er} février 2024 ou à convenir

Lieu : Avignon, France

Employeur : INRAE - BioSP

Contact : edith.gabriel@inrae.fr et lionel.benoit@inrae.fr

Contexte

L'Observatoire Pérenne de l'Environnement (OPE, <https://ope.andra.fr>) est un dispositif mis en place il y a une dizaine d'années par l'Agence nationale pour la gestion des déchets radioactifs (ANDRA) pour établir un suivi à long terme de l'environnement autour du Centre industriel de stockage géologique (Cigéo) consacré au stockage profond de déchets radioactifs.

L'ANDRA souhaite désormais tirer profit de la grande quantité de données collectées par l'OPE pour construire une prospective sur l'état des cours d'eau autour du Cigéo sous l'influence jointe du changement climatique et de la mise en exploitation du site. Pour réaliser cette étude, l'ANDRA s'associe à la chaire GeoLearning (2022-2027, INRAE - Mines Paris, <https://chaire-geolearning.org/>) qui développe des travaux de recherche en Géostatistique, événements extrêmes et Machine Learning pour la transition climatique.

Objectif du post-doc

L'objectif est de proposer une modélisation statistique spatio-temporelle et multivariée des débits et de la qualité des eaux dans la zone OPE. Le modèle envisagé permettra d'étudier l'impact de différents scénarii de changement climatique et/ou anthropique sur l'environnement de Cigéo, et d'émuler certains paramètres hydro-météorologiques utilisés en entrée de modèles d'impact (hydrogéologie, écologie des cours d'eau, etc.).

Programme de recherche détaillé

Données et zone d'étude

L'OPE est situé autour du village de Bure à la frontière de la Meuse et de la Haute-Marne dans une zone à faible densité de population. Les paysages sont caractérisés par une alternance de forêts, de grandes cultures et de prairies.

D'un point de vue hydrologique la zone comprend trois bassins versants interconnectés par un épikarst qui sont suivis en une quinzaine de sites dans le cadre de l'OPE. Le but de cet observatoire est d'évaluer la qualité des eaux (de surfaces et souterraines) à l'aide de prélèvements ponctuels réguliers renseignant un grand nombre de paramètres (environ 300) qui caractérisent en particulier le débit, et les teneurs de l'eau en matières organique et minérales ainsi qu'en micro- et macro-polluants. Six stations automatiques complètent ce dispositif en mesurant en temps continu un nombre réduit de paramètres (conductivité, température, oxygène dissous, HAP, pH, carbone organique dissous, nitrates et hauteur d'eau) afin de mieux appréhender la variabilité temporelle de la qualité des eaux superficielles. Enfin, des échantillonneurs intégratifs permettent d'estimer la concentration à long terme de certains polluants, et d'évaluer leur impact sur les organismes sensibles (p. ex. batraciens).

Pour cette étude, nous nous concentrerons sur un sous-ensemble des paramètres mesurés par l'OPE pour caractériser la qualité de l'eau. Nous retiendrons les paramètres physico-chimiques généraux de caractérisation de l'état des cours d'eau définis dans la Directive Cadre sur l'Eau (DCE), à savoir le bilan de l'oxygène (oxygène dissous et carbone organique dissous), la température de l'eau, la concentration en nutriments (phosphore, nitrates, ammonium), pH, et conductivité.

Pour compléter ces mesures effectuées sur les cours d'eau, un réseau de forages instrumentés dispersés dans le paysage comprenant actuellement une cinquantaine d'ouvrages (capteurs de niveau, plus dans certains cas de conductivité et température) fournit des informations sur l'état des nappes phréatiques connectées aux cours d'eau d'intérêt. De plus, de nombreuses données environnementales auxiliaires sont disponibles pour décrire l'environnement de surface (p. ex. occupation du sol, parcellaires et pratiques agricoles) et atmosphérique (p. ex. météorologie, qualité de l'air) afin de contextualiser le suivi de la qualité de l'eau et les interactions nappes-rivières.

Questions de recherche

La personne retenue devra traiter les questions suivantes :

1. Identifier les proxys / créer les indicateurs permettant de prédire l'évolution de la qualité de l'eau et des débits,
2. Construire, en chaque station, puis pour l'ensemble des stations, un modèle d'apprentissage automatique liant les proxys environnementaux aux variables hydrologiques cibles,
3. Réaliser une étude de l'impact des changements climatique et anthropique sur les débits des cours d'eau et la qualité de leurs eaux,
4. Interpoler, en temps continu et sur l'ensemble du réseau des cours d'eau, les prédictions issues des modèles des étapes 2 et 3.

Méthodologie

L'étape 1 d'analyse de données consiste à identifier, synthétiser et formater les données pertinentes en interaction avec les spécialistes de l'OPE, puis à créer des indicateurs de l'état de l'environnement et du climat pouvant influencer les débits et la qualité de l'eau (Messenger et al., 2021). Elle cherche également à identifier la saisonnalité, la variabilité inter-annuelle et les possibles évolutions à long terme de ces indicateurs ainsi que des paramètres de la qualité de l'eau. Enfin, l'étape 1 vise à sélectionner les indicateurs qui influencent le plus les variables cibles.

L'étape 2 consiste à construire des modèles d'apprentissage automatique (p.ex XGBoost) reliant les variables hydrologiques cibles (états d'écoulement, débits) ainsi que les descripteurs de la qualité de l'eau aux proxys environnementaux et climatiques identifiés dans l'étape 1 (Beaufort *et al.*, 2019). Une modélisation par site sera tout d'abord effectuée, puis un modèle multi-stations sera envisagé dans un second temps pour prendre en compte les dépendances inter-sites à l'échelle régionale (Sigrist *et al.*, 2022). La construction de ces modèles nécessite de lever des verrous méthodologiques liés à la gestion des fréquences d'échantillonnage de la qualité de l'eau (suivi ponctuel versus continu), à la complétion des données manquantes (p.ex engendrées par des problèmes sur les capteurs) ou encore aux incertitudes de mesure.

L'étape 3 permettra d'analyser l'impact de différents scénarii de changements environnementaux sur la qualité de l'eau à travers un exercice de simulations numériques. Les scénarii de changements climatiques pourront être simulés à l'aide d'un générateur stochastique de conditions météorologiques en cours de développement à BioSP, et les réponses des débits et de la qualité des eaux des cours d'eau seront simulées à l'aide des modèles d'apprentissage automatique élaborés à l'étape 2.

Enfin, l'étape 4 d'interpolation des sorties des modèles de l'étape 2 et 3 s'appuiera sur des méthodes de géostatistique pour estimer les paramètres de débit et de qualité de l'eau en tout point du réseau

hydrographique (Isaak *et al.*, 2014). Cette tâche est complexe car il s'agit d'interpoler sur un réseau et en temps continu des variables qui sont corrélées (Porcu *et al.*, 2022) et qui présentent un certain nombre de valeurs nulles (p. ex en présence d'assec).

Environnement de travail

Ce post-doc est financé pour 2 ans par la chaire GeoLearning, et sera encadré par Edith Gabriel et Lionel Benoit (INRAE-BioSP) avec un suivi de Sylvain Gigneux, Timothée Robineau et Maxime Savatier (ANDRA-OPE). La personne retenue rejoindra l'unité BioSP à Avignon (<https://biosp.mathnum.inrae.fr/>) pour conduire ses travaux de recherche. Le salaire brut mensuel sera compris entre 2640 et 3340 euros. Des opportunités d'enseignement pourront lui être proposées en collaboration avec Avignon Université selon sa motivation et son projet professionnel.

Compétences recherchées et candidature

Les candidat(e)s doivent être titulaires d'un doctorat en statistique, en apprentissage automatique, en science des données, ou dans un domaine connexe. Elle/il doit montrer un fort intérêt pour les applications hydro-climatiques et les problématiques liées aux ressources en eau.

Les candidat(e)s doivent avoir de solides compétences analytiques, ainsi que de bonnes compétences en programmation scientifique (de préférence en R ou Python) et en traitement de données. De bonnes compétences interpersonnelles, la volonté d'interagir avec d'autres chercheurs et étudiants et la capacité de présenter les résultats de ses travaux lors de conférences internationales sont attendues. La maîtrise de l'anglais et de solides compétences en communication orale et écrite sont requises.

Les dossiers de candidature sont à envoyer par mail à edith.gabriel@inrae.fr et lionel.benoit@inrae.fr, et doivent contenir : une lettre de motivation, un CV, les informations de contact de trois références, et la version électronique d'une publication (de préférence un article dans une revue, à défaut lien vers le manuscrit de thèse).

L'appel à candidature est ouvert jusqu'à ce que le poste soit pourvu.

Références

- Beaufort, A., Carreau, J., and Sauquet, E. (2019). A classification approach to reconstruct local daily drying dynamics at headwater streams. *Hydrological Processes*, 33(13), 1896-1912.
- Isaak, D. J., Peterson, E. E., Ver Hoef, J. M., Wenger, S. J., Falke, J. A., Torgersen, C. E., and Monestiez, P. (2014). Applications of spatial statistical network models to stream data. *Wiley Interdisciplinary Reviews: Water*, 1(3), 277-294.
- Messenger, M. L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., and Datry, T. (2021). Global prevalence of non-perennial rivers and streams. *Nature*, 594(7863), 391-397.
- Porcu, E., White, P. A., and Genton, M. G. (2022). Nonseparable space-time stationary covariance functions on networks cross time. *arXiv preprint arXiv:2208.03359*.
- Sigrist, F. (2022). Latent Gaussian model boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2), 1894-1905.



GEOLEARNING
CHAIRE /// Data Science for the Environment