Influent generator: Towards realistic modelling of wastewater flowrate and water quality using machine-learning methods

Thèse

Feiyi Li

Doctorat en génie des eaux
Philosophiae doctor (Ph.D.)

Québec, Canada

Influent generator: Towards realistic modelling of wastewater flowrate and water quality using machine-learning methods

Thèse

Feiyi Li

Sous la direction de :

Peter A. Vanrolleghem, directeur de recherche

Résumé

Depuis que l'assainissement des eaux usées est reconnu comme un des objectifs de développement durable des Nations Unies, le traitement et la gestion des eaux usées sont devenus plus importants que jamais. La modélisation et la digitalisation des stations de récupération des ressources de l'eau (StaRRE) jouent un rôle important depuis des décennies, cependant, le manque de données disponibles sur les affluents entrave le développement de la modélisation de StaRRE.

Cette thèse vise à faire progresser la modélisation des systèmes d'assainissement en général, et en particulier en ce qui concerne la génération dynamique des affluents. Dans cette étude, différents générateurs d'affluent (GA), qui peuvent fournir un profil d'affluent dynamique, ont été proposés, optimisés et discutés. Les GA développés ne se concentrent pas seulement sur le débit, les solides en suspension et la matière organique, mais également sur les substances nutritives telles que l'azote et le phosphore. En outre, cette étude vise à adapter les GA à différentes applications en fonction des différentes exigences de modélisation. Afin d'évaluer les performances des GA d'un point de vue général, une série de critères d'évaluation de la qualité du modèle est décrite.

Premièrement, pour comprendre la dynamique des affluents, une procédure de caractérisation des affluents a été développée et testée pour une étude de cas à l'échelle pilote. Ensuite, pour générer différentes séries temporelles d'affluent, un premier GA a été développé. La méthodologie de modélisation est basée sur l'apprentissage automatique en raison de ses calculs rapides, de sa précision et de sa capacité à traiter les mégadonnées. De plus, diverses versions de ce GA ont été appliquées pour différents cas d'études et ont été optimisées en fonction des disponibilités des données (la fréquence et l'horizon temporel), des objectifs et des exigences de précision.

Les résultats démontrent que : i) le modèle GA proposé peut être utilisé pour générer d'affluents dynamiques réalistes pour différents objectifs, et les séries temporelles résultantes incluent à la fois le débit et la concentration de polluants avec une bonne précision et distribution statistique; ii) les GA sont flexibles, ce qui permet de les améliorer selon différents objectifs d'optimisation; iii) les GA ont été développés en considérant l'équilibre entre les efforts de modélisation, la collecte de données requise et les performances du modèle.

Basé sur les perspectives de modélisation des StaRRE, l'analyse des procédés et la modélisation prévisionnelle, les modèles de GA dynamiques peuvent fournir aux concepteurs et aux modélisateurs un profil d'affluent complet et réaliste, ce qui permet de surmonter les obstacles liés au manque de données d'affluent. Par conséquent, cette étude a démontré l'utilité des GA et a fait avancer la modélisation des StaRRE en focalisant sur l'application de méthodologies d'exploration de données et d'apprentissage automatique. Les GA peuvent

donc être utilisés comme outil puissant pour la modélisation des StaRRE, avec des applications pour l'amélioration de la configuration de traitement, la conception de procédés, ainsi que la gestion et la prise de décision stratégique. Les GA peuvent ainsi contribuer au développement de jumeaux numériques pour les StaRRE, soit des système intelligent et automatisé de décision et de contrôle.

Mots-clés : apprentissage automatique, pilotage par la donnée, exploration des données, eaux numériques, la conception et le contrôle des StaRRE

Abstract

Since wastewater sanitation is acknowledged as one of the sustainable development goals of the United Nations, wastewater treatment and management have been more important then ever. Water Resource Recovery Facility (WRRF) modelling and digitalization have been playing an important role since decades, however, the lack of available influent data still hampers WRRF model development.

This dissertation aims at advancing the field of wastewater systems modelling in general, and in particular with respect to the dynamic influent generation. In this study, different WRRF influent generators (IG), that can provide a dynamic influent flow and pollutant concentration profile, have been proposed, optimized and discussed. The developed IGs are not only focusing on flowrate, suspended solids, and organic matter, but also on nutrients such as nitrogen and phosphorus. The study further aimed at adapting the IGs to different case studies, so that future users feel comfortable to apply different IG versions according to different modelling requirements. In order to evaluate the IG performance from a general perspective, a series of criteria for evaluating the model quality were evaluated.

Firstly, to understand the influent dynamics, a procedure of influent characterization has been developed and experimented at pilot scale. Then, to generate different realizations of the influent time series, the first IG was developed and a data-driven modelling approach chosen, because of its fast calculations, its precision and its capacity of handling big data. Furthermore, different realizations of IGs were applied to different case studies and were optimized for different data availabilities (frequency and time horizon), objectives, and modelling precision requirements.

The overall results indicate that: i) the proposed IG model can be used to generate realistic dynamic influent time series for different case studies, including both flowrate and pollutant concentrations with good precision and statistical distribution; ii) the proposed IG is flexible and can be improved for different optimization objectives; iii) the IG model has been developed by considering the balance between modelling efforts, data collection requirements and model performance.

Based on future perspectives of WRRF process modelling, process analysis, and forecasting, the dynamic IG model can provide designers and modellers with a complete and realistic influent profile and this overcomes the often-occurring barrier of shortage of influent data for modelling. Therefore, this study demonstrated the IGs' usefulness for advanced WRRF modelling focusing on the application of data mining and machine learning methodologies. It is expected to be widely used as a powerful tool for WRRF modelling, improving treatment configurations and process designs, management and strategic decision-making, such as when transforming a conventional WRRF to a digital twin that can be used as an intelligent and automated system.

Key words: machine learning, data-driven models, data mining, digital water, WRRF design and control