

Towards intelligent Process Control of Municipal Wastewater Treatment: Development of a Hybrid Model that aims to improve Simulation Performance and Process Optimization.

École des Ponts / ParisTech

École doctorale N° 531, Sciences, Ingénierie et Environnement SIE

Spécialité : Sciences et Techniques de l'Environnement

PhD dissertation defended on 4th of July, 2023 by Marcello SERRAO

This thesis was prepared at the
Laboratoire Eau, Environnement et Systèmes Urbains LEESU (ENPC)
in co-direction with modelEAU (ULaval)
within the framework of the Mocopée phase II research program
in collaboration with SIAAP and W-SMART.



Bruno Tassin



UNIVERSITÉ
LAVAL

Peter Vanrolleghem



Vincent Jauzein - Eloïse DeTredern
Sam Azimi - Vincent Rocher



Ilan Juran

Abstract

Wastewater treatment models are recognized as important decision support tools for process design, scenario analysis, impact assessment, optimization of treatment operation, and for educational purposes. Several biofiltration models have been developed so far based on modified Activated Sludge Models. They allow simulating the removal of organic matter and nutrients, as well as greenhouse gas emissions. Generally, these models have been calibrated with short-term experimental data in pilot set-ups and have reached reasonable performance. The greater complexity of mechanistic models of biofiltration makes their parameterization, calibration and validation much more laborious, requiring more dedicated measurement data. This can become a bottleneck to model application in an operational context.

Alternatively, hybridization of mechanistic models with data-driven models allows taking advantage of the strong points of each model type: calculations of water quality variables are based on the fundamental understanding of the physical-chemical-biological processes, which are supported by a fast computing data-driven model to correct for missed dynamics. The data-driven model learns unknown relationships from the data and helps improving simulation quality, especially for interpolation purposes. In recent literature, hybrid models describing activated sludge processes have improved model performance and supported process monitoring and control. As far as known to the authors, in this study, the approach is applied for the first time to biofilm systems. Hybrid models have a higher benefit/cost ratio for solving complex problems, which is a key factor for process systems engineering.

In this study, a hybrid biofiltration model was developed that improves model prediction power compared to the pure mechanistic one. A previously developed biofiltration model was recalibrated with recent data from the biofiltration lane of the 6 million PE, 1.5 million m³/day wastewater treatment plant that uses submerged upflow biofilters for denitrification with controlled methanol dosing. The mechanistic model is a 1-dimensional dynamic biofilm reactor model describing a ‘fixed-culture’ in an upward flow biological filter. The model simulates in total 22 biological and physical-chemical conversions of carbon, nitrogen and phosphorus by describing the evolution of all components with mass balances. The recalibrated model effluent nitrate concentrations are however overestimated by a relative mean error of up to 25% (calibration for 1 month).

Data-driven models were selected among linear regression models, regression decision tree models; support vector machines and artificial neural networks with up to three hidden layers. The hybrid model architecture adopted is a cooperative parallel hybrid structure with a 3-layer feedforward neural network with 12, 25 and 1 neurons for the input, middle and output layer and a ReLU activation function. It was trained to calculate the error between the simulations of the mechanistic model and the observations to allow correcting the simulated effluent quality variables. The results obtained show that the output of the hybrid model is much closer to the observations for the effluent nitrate concentrations with a mean error of 2% (training) and 8% (testing). Low Janus coefficients in the range of 1 to 2 confirm the validity of the hybrid model.

The hybrid model was subsequently applied in a process control system for methanol dosing in a predenitrification biofilter to reach a certain set point of the effluent nitrate concentration. The

proposed hybrid model predictive controller first searches for an optimal dosing rate by running multiple scenarios through the mechanistic biofilter model and the data-driven model described earlier for a two-hour prediction horizon. The HMPC then selects the scenario with the closest agreement with the nitrate set point. The proposed action corresponding to the first 15 minutes time step is then applied to a second biofilter model that represents the real physical system. This allows studying the effect of the proposed optimal dosing rate. Feedback to the HMPC optimizer is provided for so that the optimizer can correct of any divergence between the model included in the HMPC and the model playing the role of reality. The results show that the HMPC control strategy leads to an improvement in effluent quality compared to the current feedforward controller implemented at the plant, albeit with an increased operational cost of methanol dosing. Further research is needed to study the effect of a multi-objective optimization strategy that searches for optimal dosing rates, while at the same time considering the economic costs and associated greenhouse gas emissions.

Résumé

En contexte industriel, les modèles mécanistes de traitement des eaux usées décrivant les procédés de boues activées sont reconnus comme des outils importants d'aide à la décision. Ils intègrent un grand nombre de paramètres et nécessitent une phase complexe de calage préalable. Dans le cas du traitement par biofiltration, un procédé compact et largement appliqué en région urbanisée parisienne, la nécessité de modéliser les processus bio-physico-chimiques au sein du biofilm complexifie ces modèles mécanistes. Ces modèles pour la biofiltration nécessitent une puissance de calcul relativement élevée et ne sont pas encore bien établis.

Des modèles de *Machine Learning (ML)* basés sur les données décrivent quant à eux le système uniquement sur les informations extraites des données. Ils présentent une force d'interpolation et sont beaucoup plus rapides en calcul, ce qui les rend très intéressants pour les applications en temps réel.

L'hybridation combine un modèle mécaniste qui intègre des connaissances pertinentes sur les processus, avec un modèle basé sur les données qui augmente la précision des estimations en incluant des informations sur des sous-processus mal décrits.

Dans cette étude, un modèle mécaniste existant est recalibré avec des données de mesure haute fréquence provenant de l'étage de dénitrification amont de la filière de biofiltration d'une usine à grande échelle. Un modèle hybride, qui intègre un modèle de *Machine Learning* d'estimation des résidus, est décrit. Il permet une réduction significative des erreurs d'estimation.

1 Introduction

Le programme de recherche Mocopée phase I (2014 – 2017) a permis de développer des modèles mathématiques de prédiction des variables de qualité d'eau pour les procédés de traitement appliqués aux stations d'épuration de la région parisienne (SIAAP, 2018). Cet effort a abouti à la conception de modèles mécanistes de décantation primaire physico-chimique (Bernier et al., 2016), de traitement secondaire par biofiltration (Zhu, 2020) et par bioréacteur à membrane (Nadeem, 2021). L'étape suivante a consisté à développer un modèle filière qui intègre les trois phases de biofiltration : la dénitrification amont (DAM), la nitrification (NIT) et la dénitrification aval (DAV).

Au contraire des modèles mécanistes ASM (Activated Sludge Models, Henze et al., 2000) qui décrivent les processus de croissance de la biomasse en suspension (présents dans les procédés de boues activées) et qui ont une base solide au sein de la communauté scientifique, les modèles mécanistes de biofiltration, qui est un procédé à base de culture fixée de microorganismes, sont encore en plein développement. Les équations biologiques prises en compte par l'ASM ne peuvent pas être directement appliquées dans un modèle de biofiltration en raison de l'importance des processus de transport par diffusion, des cinétiques biologiques qui se déroulent dans le biofilm et du rétrolavage. De plus, il n'existe pas encore de protocoles largement acceptés pour développer des modèles de biofilm (Rittman et al., 2018). Les travaux de thèse de Jialu Zhu (2020) ont permis de développer un premier modèle filière, qui intègre les trois phases de biofiltration, à partir de données collectées au mois de mai 2018. Les

résultats montrent que le modèle est généralement capable de prédire les valeurs moyennes quotidiennes des variables en sortie.

Cependant, la complexité des modèles mécanistes rend leur paramétrage et leur calage laborieux, consommateurs en temps et en données ; cela peut devenir une contrainte limitant leur application (Vanrolleghem et al., 2005). Ils nécessitent une puissance de calcul élevée, ce qui les rend moins adaptés pour les applications de contrôle en temps réel avec un court horizon temporel d'action (Schneider et al., 2022).

A l'opposé, les modèles de ML basés sur les données utilisent des méthodes d'apprentissage automatique pour trouver des relations inconnues entre les données, et sont très rapides en temps de calcul. Ces modèles décrivent le système uniquement sur la base d'informations extraites des données et ont de fortes capacités d'interpolation (Newhart et al., 2019). A cause des caractéristiques non-stationnaires et de la variabilité temporelle potentiellement importante des liens existants entre les données, un réentraînement fréquent de ce type de modèle reste nécessaire (Torfs et al., 2022).

La modélisation hybride, combinant l'utilisation de modèles mécanistes, basés sur la connaissance du système, avec des modèles de ML, basés sur les données, est capable de cumuler les avantages de chaque approche (Von Stosch et al., 2014). Le principal avantage est un rapport bénéfice/coût plus élevé pour résoudre des problèmes complexes, ce qui est un facteur clé pour l'ingénierie des systèmes de procédés (Schneider et al., 2021). Ce travail s'intègre dans une démarche de développement de jumeaux numériques dans le domaine de l'assainissement.

L'objectif du travail présenté est donc quadruple ; premièrement, mettre à jour le modèle de biofiltration développé précédemment (Zhu, 2020) en calant ses paramètres grâce à des données acquises sur site industriel en période 2019 – 2020 et préalablement qualifiées et identifiées grâce à une analyse de sensibilité; deuxièmement, développer un modèle de *ML* permettant l'identification des erreurs du modèle mécaniste; et troisièmement, intégrer ces deux composantes dans un modèle hybride afin d'augmenter la précision des estimations ; et quatrièmement, l'application d'un modèle hybride comme étude de cas de contrôle de processus dans un cadre opérationnel avec des données à haute résolution.

Le modèle hybride développé (Figure 1) est une structure hybride parallèle coopérative (Schneider et al., 2022) où le modèle de ML prédit la différence entre les valeurs des simulations du modèle mécaniste et les observations (l'erreur résiduelle, Δ). Ceci nécessite que le modèle mécaniste fasse d'abord des estimations des valeurs de variables en sortie du procédé (ligne bleue dans la figure), qui sont ensuite analysées par le modèle de *ML* (boîte bleue), qui reçoit également les informations sur la qualité d'eau à l'entrée du procédé (ligne noire). Enfin, les résultats du modèle mécaniste et du modèle de *ML* sont combinés pour en déduire une valeur plus précise. Dans le cas de simulations prédictives, en l'absence de mesure en sortie du procédé, le modèle hybride nous permet de prédire les valeurs de l'eau traitée.