8th IWA Water Resource Recovery Modelling Seminar
Stellenbosch, South Africa, 15-18 Jan 2023
Affiliation: International Water Association (IWA)

Modelling of Urban Wastewater Treatment Processes covering the entire Data-Pipeline: from Raw Data to a Digital Twin to Optimize Performance and Increase Efficiency.

M. Serrao*, J. Sparks, K. Mesta, J.D. Therrien, N. Covre, F. Li, N. Nicolaï, P. A. Vanrolleghem
*model*EAU* - Département de génie civil et génie des eaux - Pavillon Adrien-Pouliot - 1065 avenue de la Médecine, Québec, QC, Canada, G1V 0A7 (marcello.serrao@enpc.fr)

Keywords: data treatment; hybrid model; machine learning; process control; N-removal

Introduction

The treatment of municipal wastewater is an important contributor to the protection of public and ecological health. The development and application of “smart” process control tools helps Water Resource Recovery Facilities (WRRFs) increases optimization and efficiency, facilitates resource recovery, and promotes the shift towards energy neutrality as part of a circular economy. Smart systems allow visualizing the effects of recommended control strategies and provide indications of the expected economic and ecological benefits (e.g., lowering CO_2 and N_2O emissions). This builds the confidence required for decision making and bridges the gap between superior WRRF effluent quality and desired effluent quality achieved cost-effectively.

Therrien et al. (2020) point out that these developments share a strong need for process data, which has historically been difficult to collect. With sensors more ubiquitous than ever, WRRFs now produce an unprecedented amount of data. For sensors and novel data-driven tools to help close the digitalization gap and have a significant impact on the operation they must create value for water professionals. Simply put, the information they yield must be relevant enough to be worth looking at; be clear enough to be understood; be convincing enough to be believed and be reliable enough to be acted upon (Therrien et al., 2020).

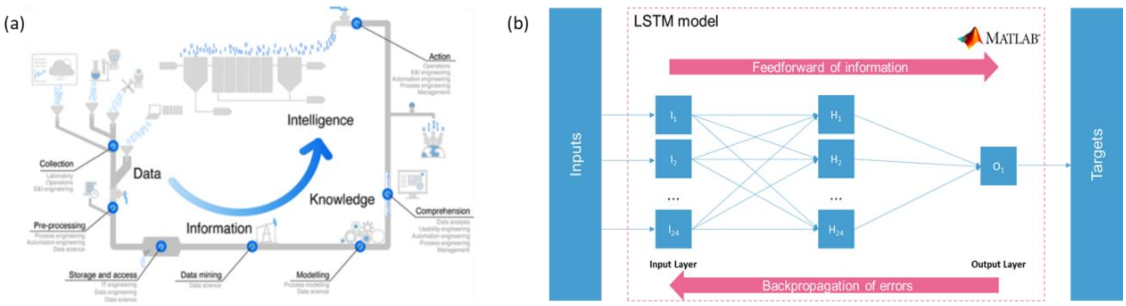

Figure 1 (a): From data to intelligence. For each step in the pipeline, the most essential professions are listed (Therrien et al., 2020). **(b)**: Schematic structure of the LSTM model architecture for automatic tuning of data validation methods developed by Nathalia Soares Covre.

As indicated in Figure 1a, the data processing steps required to create understanding and intelligence from raw measurements can collectively be thought of as a pipeline.

When knowledge-based tools are transparent and robust enough to be trusted and are supported by statistical modelling approaches, intelligent decisions can be made with confidence (Therrien et al., 2020).

The research that will be reported upon aims at deepening the understanding of WRRF processes and to develop smart tools spanning the entire data pipeline to turn data into intelligent WRRF actions. This paper aims to summarize current collective work at model*EAU*.

Methodological Approach

model*EAU*'s work spans multiple fields of enquiry, all of which are mobilized to facilitate WRRF digitalization:

- Data pipeline framework: Niels Nicolaï and Jean-David Therrien reviewed the WRRF digitalization literature, pinpointed sources of difficulties and proposed paths towards improving the value of collected data for all stakeholders.
- Data storage and access: Jean-David Therrien and Niels Nicolaï contributed to the group's ongoing development of a water quality database for raw and validated data with an emphasis on structured metadata (Plana et al. 2019). Tools for data visualization and extraction were developed to facilitate data reuse.
- Automatic data calibration: Nathalia Soares Covre is applying machine-learning fundamentals to improve data quality assessment methods to tune themselves. A LSTM Artificial Neural Network is being developed, shown in figure 1b, which can detect and replace data from outliers or other small sensor failures, and treating signals in real-time during process monitoring.
- Automated influent generator: the aim of Feiyi Li's research project is to create an artificial neural network that generates realistic datasets describing the wastewater flow rate and composition at a WRRF inlet.
- Automated respirometry-based control: Karen Mesta's project is based on quantifying the amount of rapidly oxidizable material (including nitrifiable nitrogen) in raw wastewater using a model that can interpret the data of an automated respirometer (figure 2a). The main objective is to optimize plant operations through automated control that interprets the respirometer's system response to assess the wastewater pollutant load and to identify the system dynamics.
- Model-based process operation and control: Jeffrey Sparks is researching predictive and model-based ammonia and aeration controls to minimize PID loop errors, energy usage and the negative impacts from industrial slug loads using statistical, mechanistic and hybrid models. Neural networks, tree-based and biological models are elements of this research implemented at full-scale at Nansemond WRRF (Virginia).
- Adaptive hybrid model: Marcello Serrao develops an adaptive hybrid digital twin integrating the assets of plant-wide process models based on our understanding of the physical, biological and chemical processes, and the strengths of data-based algorithms to autonomously process data and recognize patterns (Figure 2b). This work is implemented at full-scale on one of Europe's largest WRRF, the Seine Aval (France) nutrient removing biofilter plant servicing almost 6 million inhabitants.

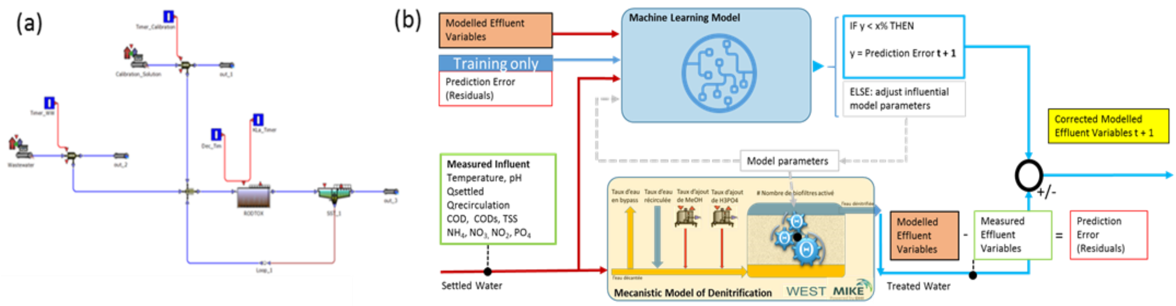

Figure 2: **(a)** Digital representation of the pil*EAU*te's automated respirometer developed in WEST (DHI). **(b)** Hybrid model consisting of a data-driven model (blue box) which serves as a corrector to the knowledge-based model (orange box) for the Seine Avail biofilter plant.

<u>Data sources</u>

The data acquired by the researchers at model*EAU* come from various sources and cover a variety of treatment processes.

Université Laval is equipped with a state-of-the-art laboratory and pilot-scale wastewater treatment system (pil*EAU*te 12 m^3 pilot for nutrient removal with two parallel trains for comparative studies) that treats water from a student dormitory and adjacent parking lots. The pil*EAU*te is highly instrumented (worth emphasizing are an on-line RODTOX respirometer and wet chemistry analysers for all inorganic nitrogen species as well as total and ortho-phosphate) The lab is well-equipped to study particle properties, in particular particle size (Lasentec FBRM and Malvern particle sizers) and particle settling velocity (ViCAs and Elutriation set-up), and biodegradability and toxicity of wastewater samples using respirometry.

model*EAU*'s research is supported by operational data from municipal facilities in France and the US. The Seine-Aval treatment plant managed by SIAAP (Paris region; 1,7 million m^3/day dry weather; 6 million PE) uses inclined lamella separators in chemically enhanced primary settlers for fast settling of suspended solids, followed by a 3-stage submerged biofiltration for carbon, nitrogen and phosphate removal. In turn, the HRSD Nansemond Treatment Plant (NTP) is located in eastern Virginia (US). The NTP design flow is 114,000 m^3/day (380,000 PE), but current flows average only 64,000 m^3/day with a current influent BOD concentration also slightly lower than the basis of design. The plant performs Biological Nutrient Removal (BNR) utilizing a 5-stage Bardenpho process.

<u>Modelling toolkit</u>

For the research work on mechanistic process models, many model*EAU* students and SIAAP use WEST (DHI). The work at the HRSD facility is being supported by Dynamita and their modelling software known as Sumo. Work on automatic data validation at the pil*EAU*te plant is being developed in MATLAB and Python for the training of machine learning models.

Results and Discussion

For the automatic data quality assessment, Nathalia Covre applies a Long-Short Term Memory Neural Network designed in MATLAB with 3 layers (24 - 24 - 1). Datasets selected for training include measured variables that have previously been treated to remove empty values and to aggregate to an hourly frequency. Early results show a good outlier removal performance and straightforward tuning of the method.

In Karen Mesta's research project, an automated respirometer (RODTOX, Kelma) is used, in conjunction with a mechanistic model in WEST, to estimate kinetic parameters for the pil*EAU*te plant's activated sludge and to fractionate the biochemical oxygen demand. The respirometer's signal is automatically forwarded to a custom water quality database, which is then queried to perform biodegradation model identification.

The hybrid model developed by Marcello Serrao uses a mechanistic model (WEST) to simulate a biofilm reactor for the nitrification and denitrification processes in a series of submerged biofilters. The data-driven component (MATLAB) compensates for unmodelled dynamics and uncertainty due to the inherent process complexity.

Jeffrey Sparks uses a similar hybrid scheme that uses a mechanistic model to output DO setpoints and keep the ammonia concentrations at the end of the aeration tank at a setpoint. The data-driven component improves the mechanistic model's DO setpoint predictions for minimizing the controller error.

Conclusion

Across the wastewater sector, researchers are looking closer at the various data processing and management components of smart systems, such as automated data quality control, storage and accessibility, model calibration and interpretation. Hybrid models allow for integrating existing knowledge of wastewater treatment processes with AI-based techniques for data (pre-)processing and insight generation. As the robustness of techniques used at each step of the pipeline grow, it will become possible to integrate them into larger systems such as digital twins which will enable the production of actionable information for WRRF process control and optimization in real-time. Though such complex systems are still far from completion, model*EAU* is continuing its work to better understand, design and combine their components.

Acknowledgements

Sincere thanks are extended to SIAAP (France) and HRSD (US) for their financial support. Peter A. Vanrolleghem holds the Canada Research Chair on Water Quality Monitoring.

References

Therrien, J. D., Nicolaï, N., & Vanrolleghem, P. A. (2020). *A critical review of the data pipeline: how wastewater system operation flows from data to intelligence*. Water Science and Technology, 82(12), 2613-2634.

Torfs, E., Nicolaï, N., Daneshgar, S., Copp, J. B., Haimi, H., Ikumi, D., B. Johnson; B.B. Plosz; S. Snowling; L.R. Townley; B. Valverde-Pérez; P.A. Vanrolleghem; L. Vezzaro & Nopens, I. (2022). *The transition of WRRF models to digital twin applications.* Water Science and Technology.

Plana, Q., Alferes, J., Fuks, K., Kraft, T., Maruéjouls, T., Torfs, E., & Vanrolleghem, P. A. (2019). *Towards a water quality database for raw and validated data with emphasis on structured metadata.* Water Quality Research Journal, 54(1), 1-9.