Demo Case Update (2/2)

KWR – KWR WATER B.V.

Waternet – Wastewater Treatment Plant Amsterdam West

EURECAT – Technology Centre in Catalonia

26 november 2021

Waternet has dedicated one of the seven treatment lanes of the wastewater treatment plant (WWTP) Amsterdam West to investigate AI based process control strategies for energy optimization and N2O reduction using FIWARE system components for data exchange.

Researchers from KWR, Eurecat and Waternet have been working on influent data collection, training of the reinforcement learning (RL) agent for real-time control of the research lane, and development and deployment of near real-time soft sensors within the FIWARE system to estimate air flow, respectively. Furthermore, KWR has been working on training of an AI auto-encoder model for anomaly detection and correction of real-time ammonium and nitrate measurements.

This article is the second of a series of two articles.

Reinforcement learning

Reinforcement Learning (RL) agent's models are compared with real data in order to analyze their quality and weaknesses for hyperparameter tunning, algorithm selection and to improve the learning phase. This leads us to study the predictive capacity and error propagation in continuous episodes of the environment model. The results are used to improve the environment model performance studying the most significant variables for prediction and the future error propagation on the predicted variables. Therefore, it has been developed an easy and automated system to integrate changes on the environment model, such as new versions, to the RL process, with the aim of facilitating the integration of improvements. It is essential to speed up and encourage advancements in an automated way.

Focusing on RL training steps, the RL agent must visit as much as different scenarios as possible to learn from the experience and better generalize to non-visited states. The number of steps per episode considered during the RL training phase have been determined based on the error propagation of the environment model, ensuring quality of the predicted results. Moreover, random initializations of the environment model to different moments in time gives more overview of the problem and provides dynamic resets to random new situations. Finally, in order to have more information about the performance of the control RL models in benchmark analysis, optimization algorithms based on heuristics have been developed.

Evaluations on the RL trained models, who have learned coherent strategies, show that the current challenges are related to the reward signals used to guide the agent's learning. The reward function definition is a critical point in RL. We are currently working on redefining this guidance to ensure the correct representation of all the implications and benefits of taking any decision and / or sequence of decisions in the system.

AI-based Data Validation of Key Process Parameters

A key aspect that must be considered prior to the training, testing and implementation of an AI based digital twin and data driven intelligent control is ensuring the quality of data. Data quality can be impacted by sensor faults, sensor calibration issues, fouling or obstruction of the sensors, and connectivity problems during data transfer between sensors and the data management system, here referred to as the Process Information Management System (PIMS). As a result, there is a need for data quality checks and corrections. The use of (sophisticated) on-line real-time sensors yields high volumes of data, which makes manual detection and correction labor intensive, susceptible to human error and impractical in case of automated control. Therefore advanced, automated data handling and validation is necessary to mitigate erroneous monitoring and control of critical water operations.

For this purpose, an automated and online data validation framework has been developed that can be used as a data screening and correction layer to validate critical sensor data, prior to the sensor signals being used for further analytics or visualization. Additionally, the framework will be integrated completely within the FIWARE ecosystem that has been implemented for the demo case, as depicted in Figure 2 below. In order to demonstrate the data validation pipeline and the implementation of the smart application, ammonium (NH4) and nitrate (NO3) sensor data from the aerobic zone of the bioreactor unit in the research lane has been used. Initially, sensor data are passed through an anomaly detector that flags obvious outliers through user-defined and statistical measures, i.e. threshold-based detection and extreme shifts or jumps. Furthermore, contextual based anomalies, i.e. anomalous events occurring over time, are also detected in the form of a flatline detection, where the sensor or PIMS system provides the same value for an unrealistically long period of time. Subsequent to the flagging of anomalies, data reconciliation is conducted to correct anomalous values. Data reconciliation is conducted using recurrent neural networks based autoencoder models, consisting of Long Short Term Memory (LSTM) and dense layers. For each process parameter (NH4 and NO3 signals), two autoencoder models have been trained to capture different dynamics of the signal using resampled data and different sequence sizes as input. Finally, the outputs or predictions of the two models per parameter are resampled and aggregated using exponential weighted smoothing of the short-time constant autoencoder as well as weighted smoothing of the long-time constant autoencoder for medium to long forecasting horizons (1 hour or more). This is particularly crucial for long contextual anomalous events, where subsequent autoencoder forecasts must be made while taking the most recent predictions by then autoencoder as input. The trained autoencoder models exhibited a very high accuracy for the NH4 signal, as can be seen from the figure 3. The (near) real-time validated data as achieved through the AI-based data validation will be relayed through a proxy within the FIWARE setup and will be pushed to a dashboard. Finally, validated data can also be used as input to the AI-based data-driven smart applications, such as the digital twin model by Waternet and the smart control agent of the WWTP using reinforcement learning as developed by Eurecat.

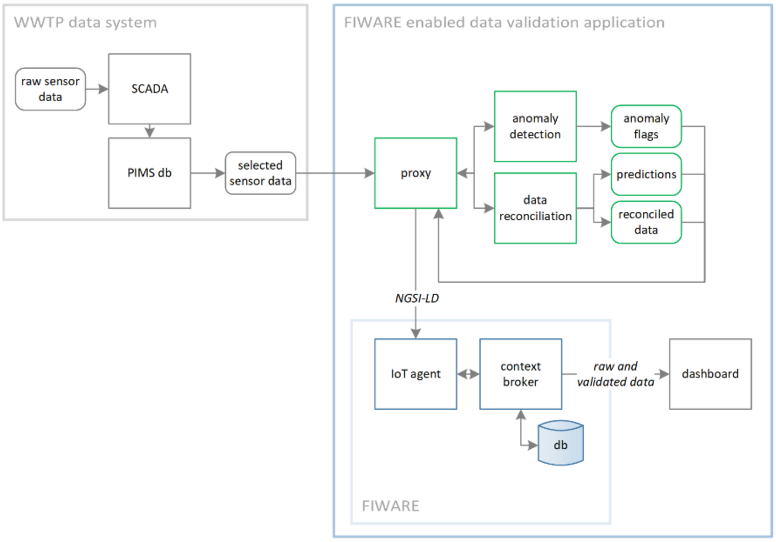
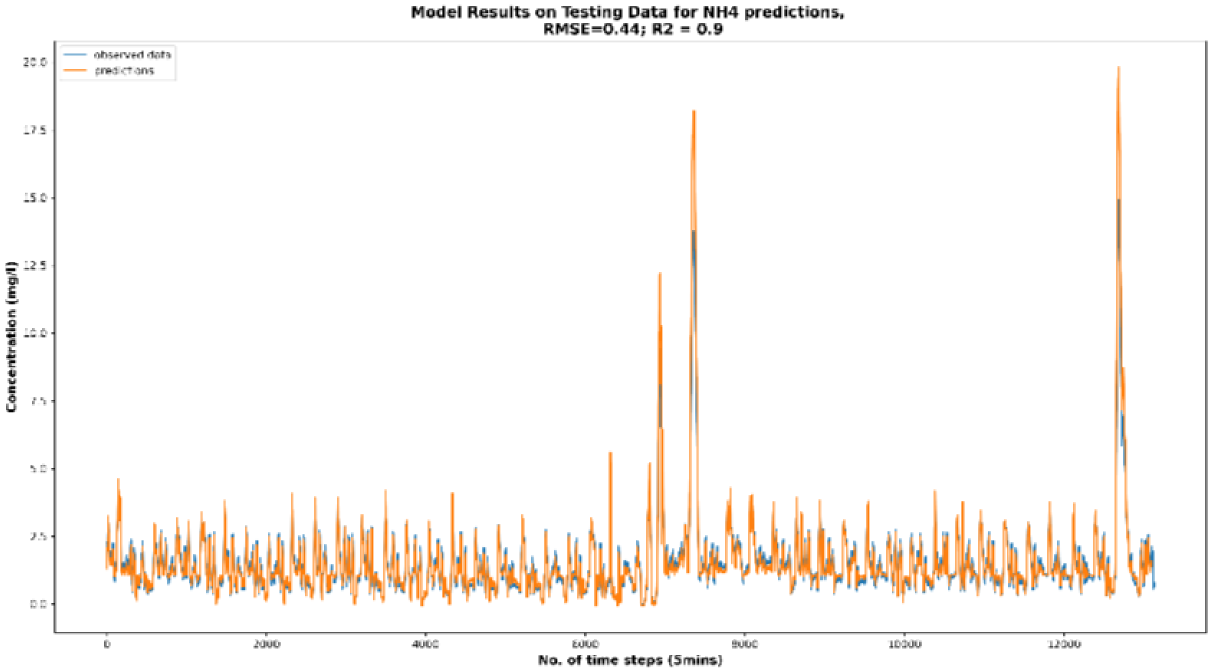


Figure 1 Simplified schematic overview of the AI-based data validation integrated with the legacy system of Waternet and the Fiware architecture



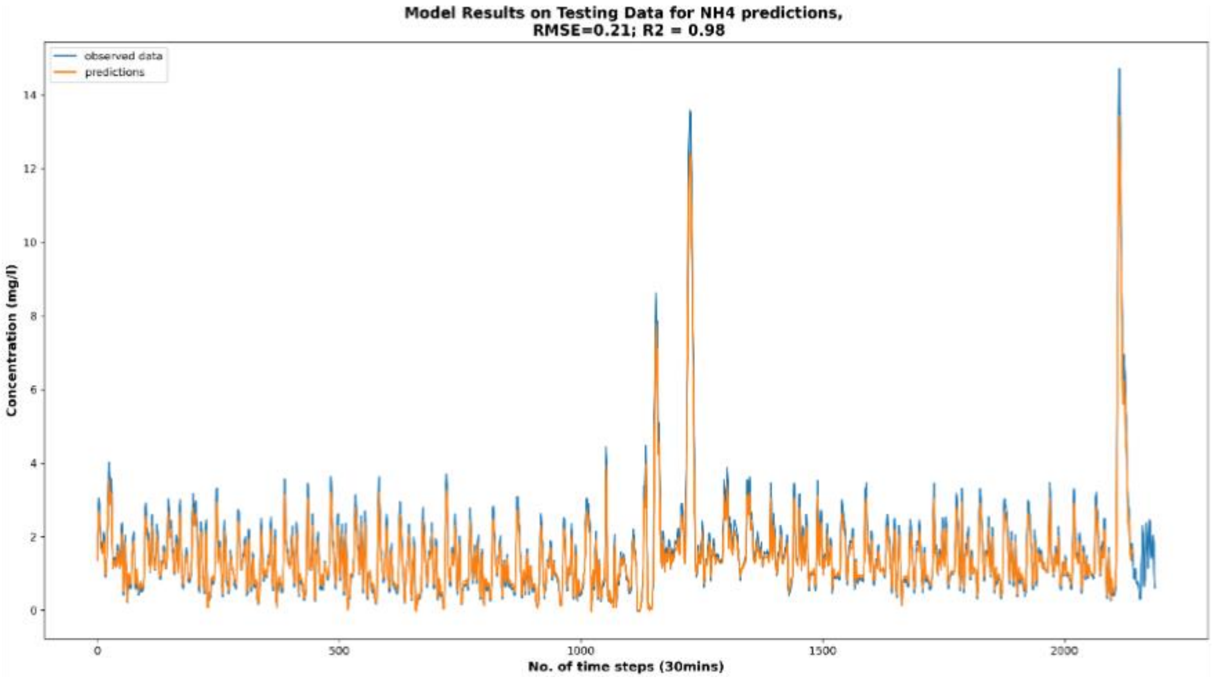


Figure 2 Raw data of NH4 sensor (blue line) and predictions of the autoencoder for data resampled to 5 minutes (upper graph) and 30 minutes (lower graph)

Author: KWR, WNT & EUT

Disclaimer

This document reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Intellectual Property Rights

© 2021, Fiware4Water consortium

All rights reserved.

This document contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

This document is the property of the Fiware4Water consortium members. No copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights. In addition to such written permission, the source must be clearly referenced.

Project Consortium

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No. 821036.