Inflow prognoses to a large wastewater treatment plant (WWTP) in Copenhagen forecasted by machine learning (ML)

Three project videos

Sten Lindberg*, Laura Frølich*, Dennis Wanninger*, Nina Donna Sto. Domingo*, Arne Møller*, Barbara Greenhill**

*DHI Group – Hørsholm, Denmark
sl@dhigroup.com, lafr@dhigroup.com, dewa@dhigroup.com, nsd@dhigroup.com, arm@dhigroup.com

**BIOFOS, Copenhagen, Denmark
bg@biofos.dk

Introduction

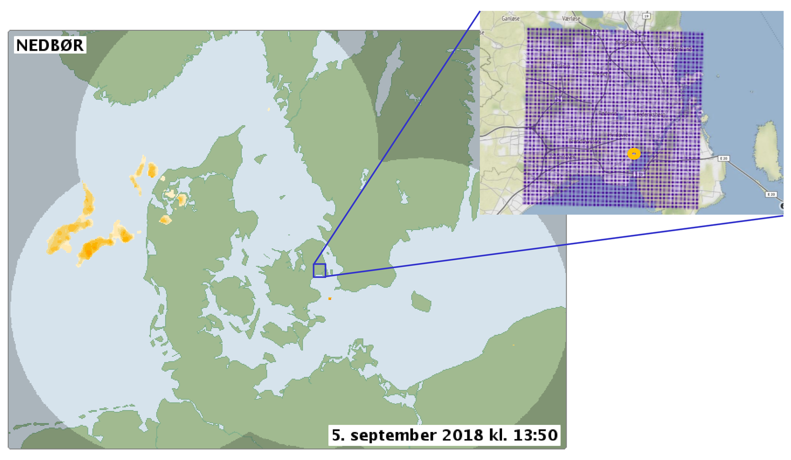

Figure 1: Radar data from Danish Meteorological Institute (DMI). Purple marks the radar reflectivity pixels used. The orange circle marks the location of the wastewater treatment plant (WWTP)

To comply with the EU Water Framework Directive (EU 2000), the total annual emissions of total nitrogen (N) from Copenhagen plants have to be reduced by 240 tons. One important measure is the reduction of bypass water (mechanically treated wastewater).

The complexity in managing the entire system has called for integrated digital solutions, including data collection and sharing.

To enhance the wastewater treatment processes and to meet the overall goals of reducing emissions, BIOFOS joined the Digital Water City project, under the EU Horizon2020 program. Development and adaptation of three digital solutions has been executed during the last two years, and the technology is now being tested and validated.

The three digital solutions are “Improved machine learning (ML) sewer inflow forecast”, “Decision Support System for real-time control of WWTP operations and in-sewer retention”, and “Web-based prototype platform for decision support at city scale”

Methodology

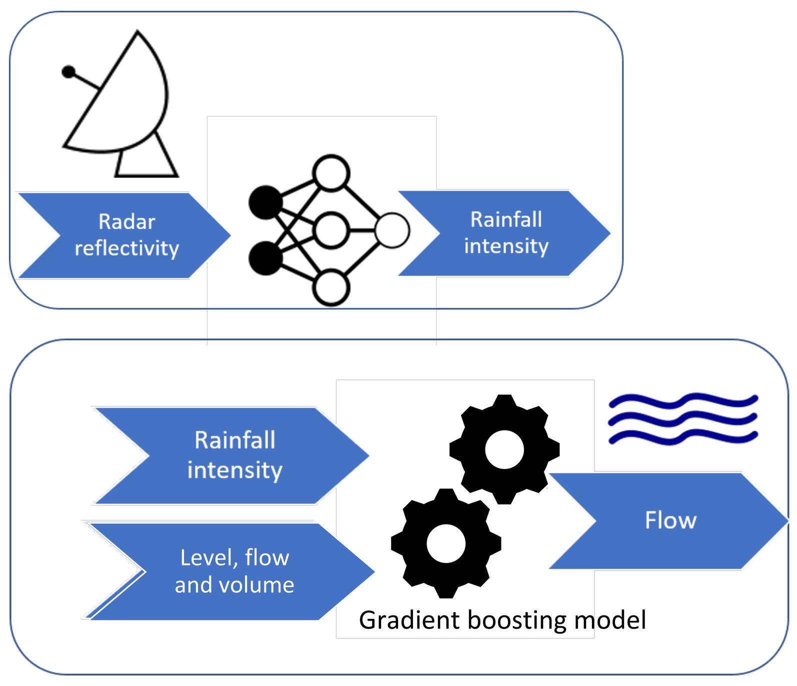

Figure 2: Radar to Rain to Flow

The three different digital solutions

- Inflow forecast based on Machine learning
- Decision Support System for Real-time Control
- Web-based visualization platform (Future City Flow)

Results & Discussion

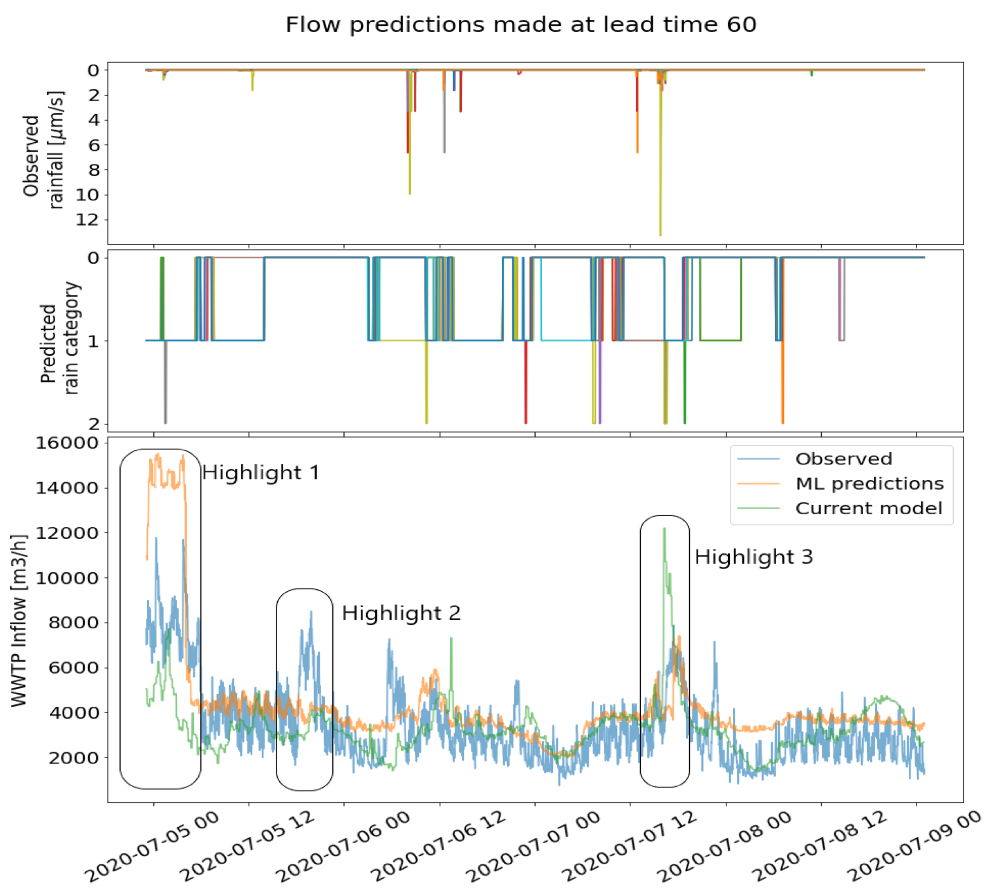

Figure 3: Comparison of forecasted and observed rain and the impact on the forecasted flows

Figure 3 shows an example of ML results. The top part shows observed rain data while rain categories derived from weather radar data are shown in the middle plot. Colors in these two plots correspond to rain gauge locations. The bottom plot shows observed flow (blue) and predictions from the grey-box model currently in use (green) and from the ML model (orange). In the beginning of the time series (highlight 1), the ML model overpredicted while the current model underpredicted. Rain categories 1 and 2 were predicted for several locations (middle plot), while little to no rain was observed (top plot), explaining the overprediction from the ML model. In highlight 2, both models underpredicted. Since no rain was observed here, the higher observed flow must be due to other effects, e.g., storage basin emptying. In highlight 3, the ML model predicted the magnitude of flow more accurately than the current model, but the current model pinpointed the time of increased flow better. These observations indicate future directions, such as improving rain category predictions to obtain more accurate ML model input.

Figure 4 shows a rain event (rain in upper parts of plots), as well the measurements and predictions of inflow from the current production model and a machine learning model that has been operationalized using Azure MLOps and trained on data from the end of 2020 to March 6th, 2021. The operationalized model uses the scikit-learn [1] HistGradientBoostingRegressor to speed up training. Inputs to this model consist of rain gauge data and upstream level, flow, and volume measurements.

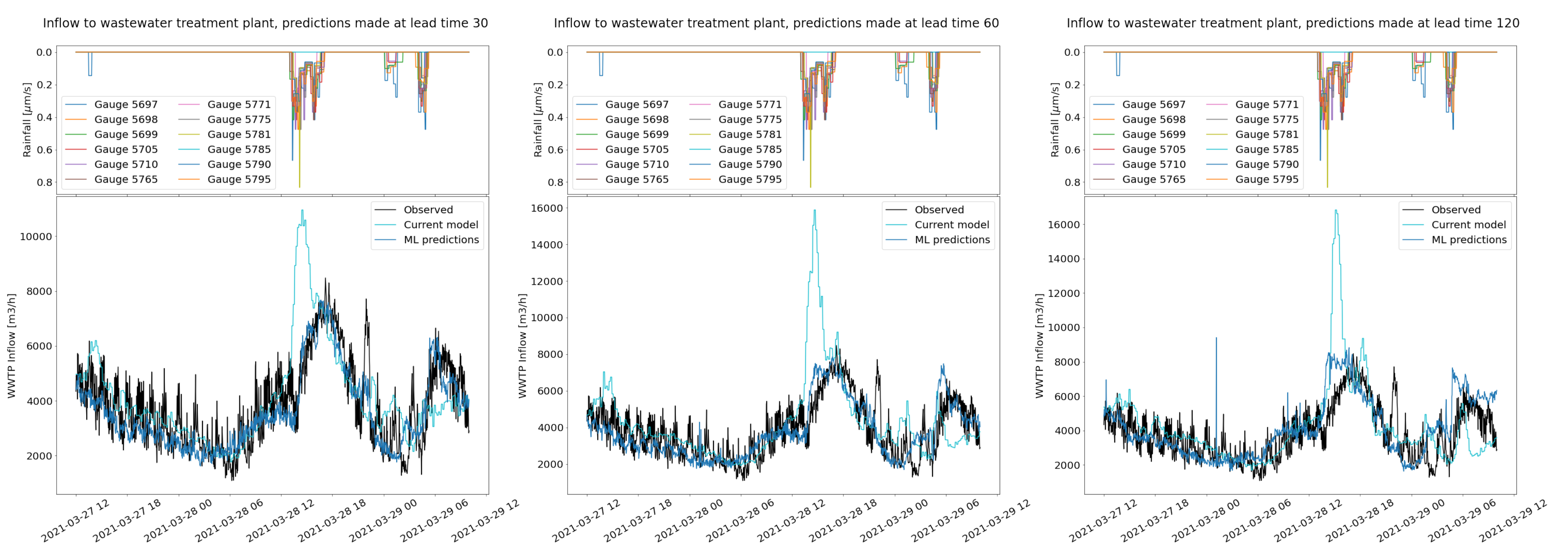

Figure 4: Measures and predictions for a rain event. Rainfall is shown in the top and predictions and measurements (black) are in the bottom for 30 minutes (left), 60 minutes (middle), and 120 minutes lead times (right)

The forecasted inflow from the machine learning is visualized together with forecasts produced with hydrodynamic models, running different controls scenarios. The wastewater treatment plant operators, use the forecasts to optimize the control strategies – increasing the capacity of the plant and reducing emissions.

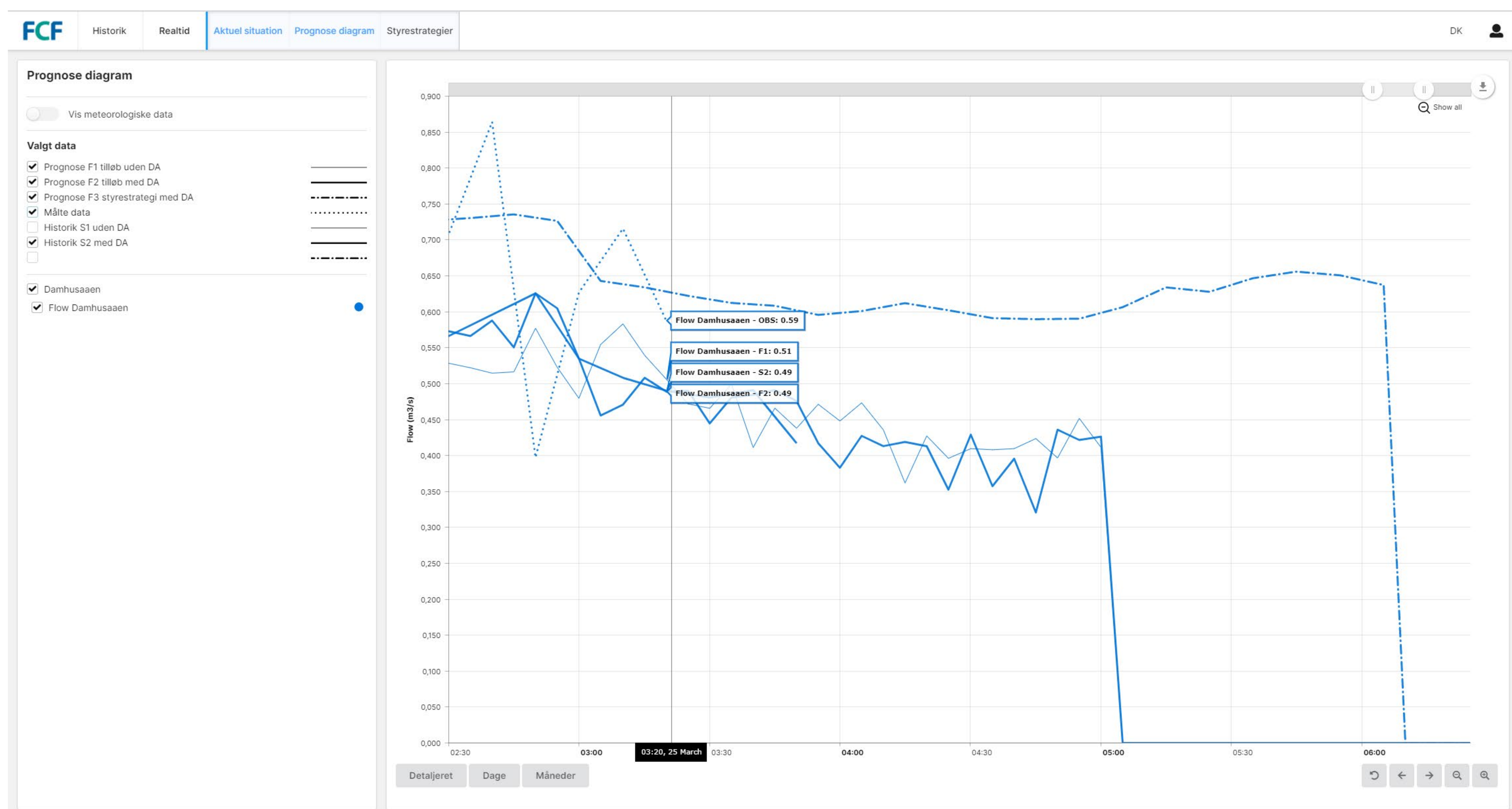

Figure 5: 2 hours forecast with hydrodynamic model and 3 hours with ML engine

Conclusions

The digital solutions described above, have been installed and deployed in Q4 2021 and Q1 2022. The very first results confirm the assumptions that use of ML based forecasts improves the accuracy, while there still are some challenges with the reliability of the radar forecast predictions. The value of gathering, displaying, and sharing both sensor data and predictions to the stakeholders have been documented.

Reference [1] Pedregosa, F. et al., 2011 Scikit-learn: Machine Learning in Python JMLR 12, pp. 2825-2830.

Digital Water .City