REDUCING OPEX, SENSOR DRIFT AND EMISSIONS THROUGH DIGITAL INNOVATION

Anna Whitmore (Mott MacDonald), Kenny Williamson (Watercare Services Ltd.), Glenys Rule (Watercare Services Ltd.), Owen Flanagan (Mott MacDonald), Evan Harwin (Mott MacDonald)

ABSTRACT

Poor quality plant operating data is a common issue for water utilities and limits operators' ability to understand and improve plant performance. Emerging technology such as soft sensors and digital twins can improve confidence by supplementing and enhancing operating plant data.

As well as enabling operators to better understand plant performance, digital twins and soft sensors can also assist with the development of control strategies to reduce operational expenditure and greenhouse gas (GHG) emissions. Watercare forecast a greater than $50M spend on wastewater asset operating costs for FY2022. The adoption of digital technology therefore has the potential to realise significant savings.

The Rosedale Wastewater Treatment Plant (WWTP) consists of primary sedimentation, biological treatment through the Modified Ludzack-Ettinger (MLE) process, clarification, UV disinfection, and anaerobic digestion for solids stabilisation.

A digital twin of Rosedale WWTP was developed in 2020 to provide real-time insights into plant performance and scenario testing, to support commissioning of the fourth MLE reactor. The solution combines biological modelling, machine learning, predictive rainfall and scenario analysis. This paper describes the development methodology and functionality, including the data, technology and analytics that were used.

Watercare's 2022 Rosedale Innovation programme comprises two workstreams which build on the existing digital twin solution; soft sensors for instrument failure detection and opex and greenhouse gas baselining and prediction.

The soft sensors workstream is focused on the development of soft sensors to proactively identify instrumentation drift and failure events. The soft sensors use a combination of machine learning and biological models to predict instrumentation drift and alert operators to irregular operation. This paper describes the development methodology including data exploration and modelling techniques.

The objective of the opex and GHG workstream is to develop real-time operational insights to enable the baselining, monitoring, and control of GHG and opex. This paper outlines the different data sources and calculations and describes how they were codified and surfaced through the digital twin. This workstream also includes analysing process emissions. This involved installing nitrous oxide (N_2O) monitors on MLE4 and calibrating the digital twin to act as a soft sensor for N_2O emissions.

The challenge of unreliable data and instrumentation failure is not unique to Rosedale WWTP. The application of emerging digital technology can enhance operational plant data through the creation of new data sources. This data provides new insights into plant operations and enables operators to consider operational strategies to reduce opex and emissions. This paper addresses how the Rosedale technology can be scaled to other treatment works to improve data quality, optimise opex and GHG emissions and improve instrumentation reliability.

KEYWORDS

SOFT SENSORS, OPEX, GREENHOUSE GAS, ASSET MANAGEMENT DECISION MAKING, DIGITAL TWIN, DATA SCIENCE, MACHINE LEARNING

PRESENTER PROFILE

Anna is a senior consultant in Mott MacDonald’s Digital Ventures team. She leads the development and delivery of impactful smart infrastructure solutions on Mott MacDonald’s digital twin platform Moata, including flagship digital twin projects.

INTRODUCTION

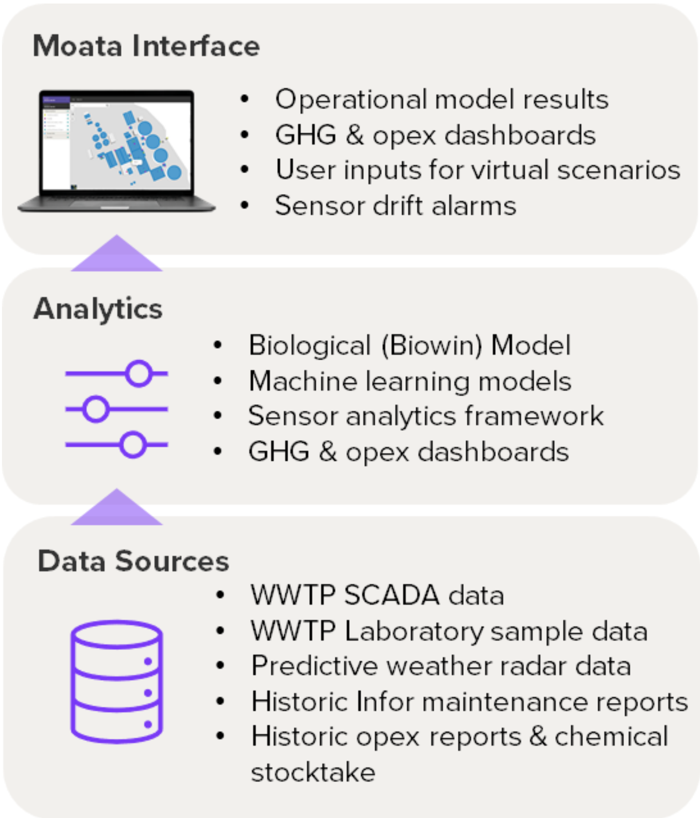
PROJECT BACKGROUND

The Watercare Rosedale Wastewater Treatment Plant (WWTP) located in Auckland, serves a population equivalent (PE) of about 240,000. The process consists of primary sedimentation, biological treatment through the Modified Ludzack-Ettinger (MLE) process, clarification, UV disinfection, and anaerobic digestion for solids stabilisation.

THE ROSEDALE WWTP DIGITAL TWIN

In 2020 a digital twin was developed for the Rosedale WWTP to support commissioning of the fourth MLE reactor. The digital twin combines biological modelling and machine learning with real-time data and dashboards to generate new insights into plant performance. Figure 1 provides a conceptual overview of the digital twin components.

Figure 1: Rosedale Digital Twin Components



The digital solutions developed through the Rosedale innovation programme leveraged the Rosedale digital twin and are underpinned by the Moata platform. Two workstreams were delivered as part of the Rosedale Programme:

WORKSTREAM 1 – OPERATIONAL COST AND GREENHOUSE GAS

The objective of workstream one was to identify, control and reduce WWTP operational expenditure ($\$/m^3$) and Greenhouse Gas (GHG) emissions. This was achieved through developing operational dashboards which reported opex and GHG emissions of the WWTP.

WORKSTREAM 2 – SOFT SENSORS FOR DRIFT DETECTION

The objective of workstream two was to identify sensor drift through the development of soft sensors. Using a combination of biological and machine learning models, soft sensors were developed to predict plant characteristics and alert the operators to potential instrument drift.

The methodology and results of each workstream are described in the following section.

WORKSTREAM 1 – OPEX & GHG WORKSTREAM

PROBLEM STATEMENT

Watercare have a target to reduce operational greenhouse gas emissions by 50% by the year 2030. Current estimates indicate approximately 70% of Watercare's operational GHG emissions is generated from WWTP process emissions.

GHG reporting is done on an annual basis with many of the values based on theoretical factors. For example, emissions from reactors are calculated on a static population and IPCC guideline emissions factors.

It is crucial to increase our understanding of GHG and opex sources to enable improved control and reduction. By developing real-time dashboards Watercare will be able to identify hotspots, trial and monitor control strategies, and view trends over time.

METHODOLOGY

The GHG and Opex dashboards were developed through the following steps:

1. Each process area was reviewed to identify any GHG or opex sources. These were broadly split into the following categories: energy, maintenance, labour, solids disposal, process emissions, and chemicals. Further detail is provided in the following section.
2. Each source was reviewed to identify what data was available to calculate the CO2 equivalent or cost. The report inputs 40 unique data streams - captured from plant SCADA, laboratory sampling results, an operational biological model, enterprise maintenance systems (Infor), monthly reporting for solids disposal and historic opex reports.
3. Calculations were developed for each source. These calculations were derived from the NZ Carbon Accounting Guidelines (Andrews, J 2021), the Watercare opex reports or process calculations.
4. Data pipelines were developed to automatically ingest data into the Moata Digital Twin. This included direct integration of SCADA and sampling data into Moata.
5. The calculations were codified in Moata to automatically calculate the daily GHG (CO2 equivalent) and opex (NZD) values.
6. Visual dashboards were developed to surface the results through a simple, easy-to-use interface.

OPERATIONAL COST AND GHG SOURCES

The opex sources at Rosedale WWTP included power consumption, chemical consumption, labour & maintenance, and solids disposal. The greenhouse gas (GHG) sources at Rosedale WWTP included process emissions, power, chemical use, and biosolids disposal & effluent discharge.

Table 1: Operational Cost and GHG Sources

Source	Consumption Calculation	Opex factor	GHG factor
Chemical – Gravity Belt Thickeners (GBT) polymer	GBT 1 polymer consumption is calculated from the average historic consumption (kg/tDS) over 2021/2022 and sludge throughput from SCADA. GBT 3 polymer consumption calculated from polymer flow rate & % concentration from SCADA.	Chemical costs taken from 2021/22 Rosedale Opex. Thickening Polymer powder ($/kg)	Polymer = 0.813kg CO_2e/kg Average value for a range of polymer, does not include transport emissions. From Moata Carbon Portal
Chemical – Centrifuge Dewatering polymer	Centrifuge 1,2,3 & 4 polymer consumption calculated from Polymer dosing rates and feed flowrates	Dewatering Polymer emulsion ($/kg)	Refer above for dewatering polymer GHG
Chemical - Utilities sodium hypochlorite (NaClO)	NaClO consumption calculated from storage tank volume and level sensor from SCADA	NaClO ($/L)	NaClO = 0.372 CO_2e/L (assumes 15.7% active chlorine) From Moata Carbon Portal
Electricity	Total electricity imported from grid is calculated from SCADA. Total electricity consumption is portioned across process areas based on equipment consumption. Equipment consumption was based on power rating and operating hours. Equipment from the load schedule <5kWh were excluded, as well as UV as it is on a separate meter. Equipment >5KW accounts for 94% of total load.	Rosedale electricity contract rates for Feb 22 to Jan 23 provided from WSL 02.05.22. Rates vary with hour and day.	Emission factor 0.101 kg CO_2e/kWh for purchased grid-average electricity (Ministry for the Environment, 2022)

Diesel engine	Diesel consumption is a manual input from Rosedale Chemical Stocktake monthly reporting.	Diesel ($/L)	Diesel emission factor 2.66 kg CO_2e/L (Ministry for the Environment, 2022)
Biosolids disposal	Biosolids production calculated from digested sludge storage tank % solids and Centrifuge feed flowrate from SCADA. Assumed a 95% solids capture rate in centrifuge.	Biosolids disposal & transport ($/wet tonne)	Assume 0.027 kg CH_4/kg TS disposed at landfill (Andrews, J 2021). Solids transport emissions not included in GHG baseline
Grit and screening disposal	Grit and screening load manually calculated from Watercare monthly reporting of biosolids removal. Reporting based on contractor invoices.	Grit and screening disposal & transport ($/wet tonne)	N/A
Process Emissions from MLE reactors	MLE, CH_4 & N_2O emissions from online biological model. Model results to be validated during N_2O monitoring campaign.	N/A	CO_2e Conversion factors: 28 kg CO_2e/kg CH_4 298 kg CO_2e/ kg N_2O
Digester emissions	Total digester methane production calculated from SCADA - Digester 1,2,3 & 4 Gas production and CH4%.	N/A	Fugitive emissions factor 0.0141 – assume a high-quality biogas digester low quality gas storage in a cool climate (Andrews, J 2021).
Emissions from flare	Volume of biogas flared calculated from SCADA - Volume gas flared and digester CH_4 %	N/A	Assume 98% destruction of CH_4 from flaring (Andrews, J 2021).
Discharge to ocean	Effluent BOD and TN load calculated from SCADA.	N/A	N_2O and CH_4 emissions based on emission factors from effluent discharge to water bodies (Andrews, J 2021).
Labour & sampling	Labour based on annual salaries of the following full time equivalent roles: 5x Process controllers, 6x	Labour and sampling split across process area - defined	N/A

	process operators, 3x plant operators and a % of operations management. Solids and liquid sampling costs for 2021/22 from WSL reporting.	by WSL operations manager. Assume even split of cost across the year.	
Maintenance	Maintenance data based on historic Infor maintenance records. Records define equipment cost, work cost and type e.g., corrective or preventative.	Process area derived from maintenance records. Assume even split across the year.	N/A

THE ROSEDALE BIOWIN MODEL

BioWin is a wastewater treatment process simulator that ties together biological, chemical, and physical process models. BioWin is created by EnviroSim Limited and used world-wide to design, upgrade, and optimise wastewater treatment plants.

A core component of the Rosedale Digital Twin is a live Biowin Model. The BioWin model was operationalised with an application programming interface (API) developed between Moata and BioWin to push/pull plant data and model results. The existing BioWin model was developed pre-2020 for the design upgrade and has known issues around the mass balance. Where plant data was not available the Biowin model outputs have been used for the GHG dashboards. A well-calibrated model is crucial for the accuracy of these results and at the time of publishing this paper there are ongoing efforts to improve calibration, including an intensive sampling survey being carried out in August 2022.

DISCUSSION

Extracts of the operational dashboards are shown in Figure 2, Figure 3 and Figure 4 below.

Figure 2: Daily GHG emissions per process area

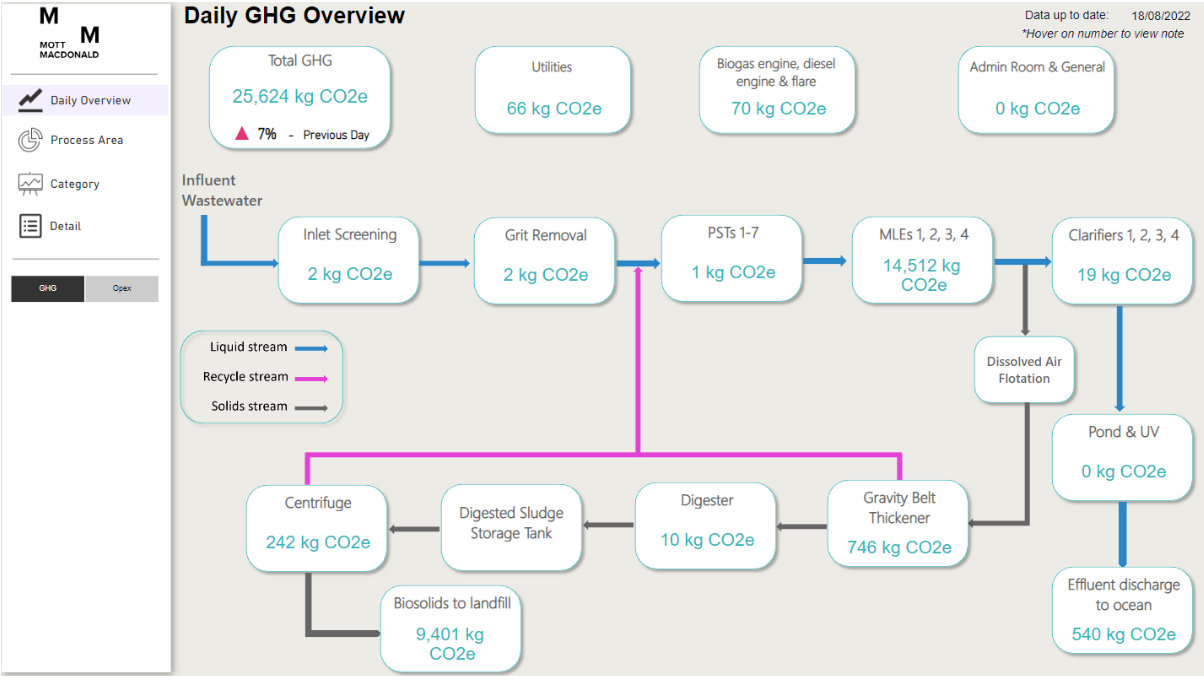


Figure 3: GHG by process area and weekly trends

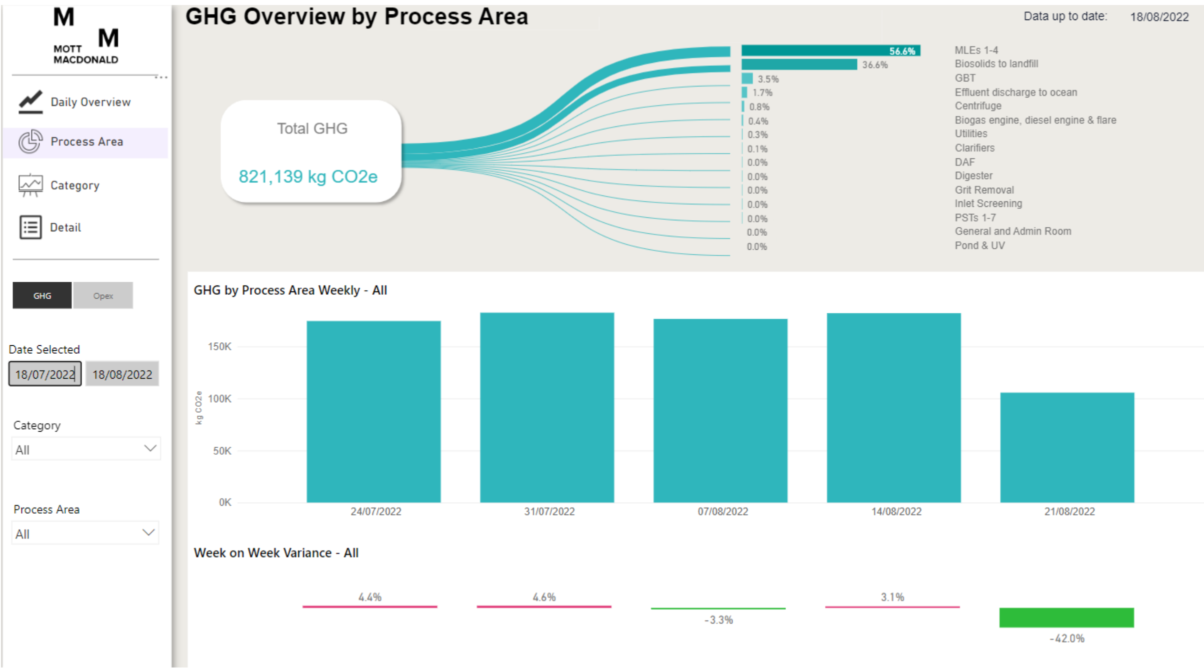
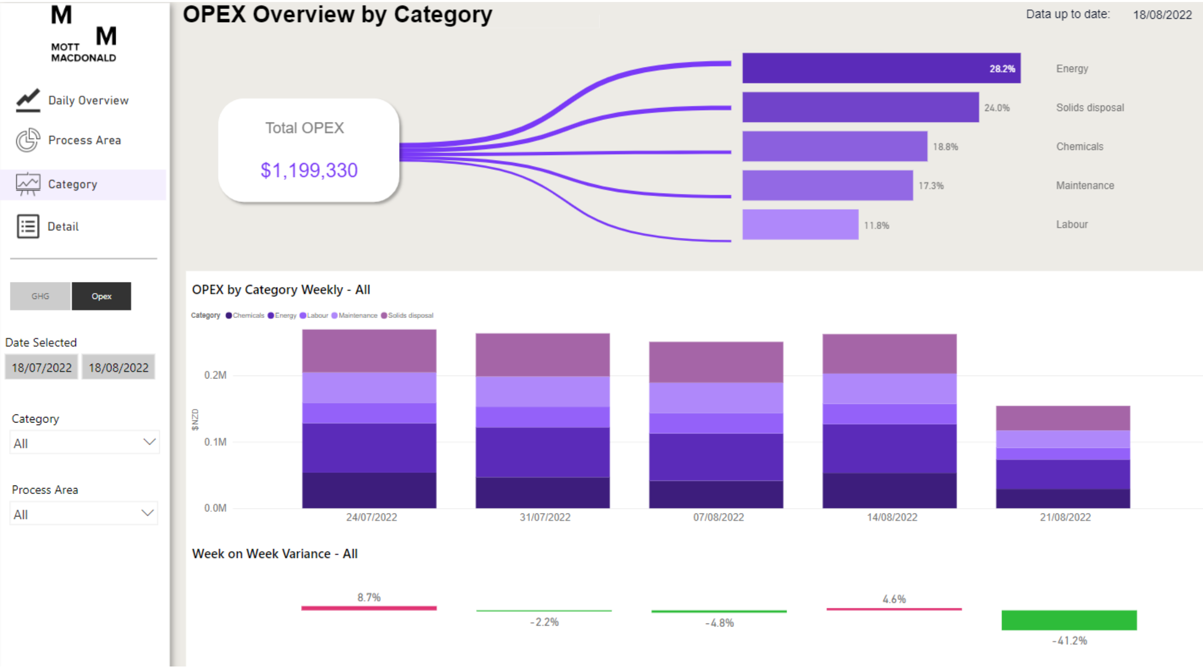


Figure 4: Opex values by category and weekly trends



The core objective of the dashboards was to identify the hot spots across the WWTP. Process emissions has been highlighted as the major contributor to GHG emissions, contributing 94% of the total GHG. Based on the operational Biowin model 57% is produced in the MLE reactors with 37% occurring at the landfill where the solids are taken. In 2019 the Intergovernmental Panel on Climate Change (IPCC, 2019) published guidelines including recommendations on GHG inventory methods. This highlighted the need to include facility-level data in inventories. To verify the modelled results and broad emissions factors, a N_2O monitoring campaign is underway at the Rosedale WWTP.

One of the challenges faced when developing the dashboards was the input data quality. Multiple SCADA data input streams had issues with completeness and accuracy. To generate the reports data cleaning was applied, for example excluding negative values and forward filling data where none was available. Further investigation into the input data quality would improve the robustness of the dashboards.

CONCLUSION

GREENHOUSE GAS EMISSIONS

The purpose of the operational dashboards is to identify hotspots of GHG and Opex production, and to monitor the impacts of control strategies. Initial results indicate that process emissions from the MLE contribute 57% of the overall WWTP GHG emissions, and ~90% of the onsite GHG emissions (if landfill emissions form solids disposal are excluded).

These results will be validated with actual GHG N_2O monitoring data once this is available. These results reinforce the need to better understand process emissions and develop and trial control strategies. The GHG dashboards indicate ~100 tonnes CO2e is produced from MLE process emissions per week. This equates to $442,000 per year in carbon credits (at current spot rates of ~$85/TCO2e).

OPERATIONAL COST

The opex dashboards indicate solids disposal as a major operational cost – contributing approximately 25% of total opex. Solids disposal costs include transport and disposal for biosolids, screening and grit. The volume of Biosolids produced is calculated from centrifuge SCADA data (% solids and feed flowrate) and multiplied by an annual rate for transport and disposal. The production of biosolids has been highlighted as an area that could be further optimized for cost and emissions. A factor of 0.027 kg CH_4/kg TS has been used for biosolids disposed at landfill – resulting in Biosolids contributing 37% of total emissions.

It is important to consider the interrelationships between effluent quality, GHG emissions and Opex. By creating dashboards which include both GHG and Opex we can identify relationships and monitor trends of both values when trialling control strategies. This will be pertinent for the MLEs where aeration has a significant impact on both MLE Opex (10% of the total plant Opex) and N_2O production (57% of GHG emissions).

AMPLIFICATION ACTIONS

The dashboards developed through this study have enabled the Rosedale team to identify target areas, trial control strategies and monitor results. This was made possible by the combination of connected plant data, online process calculations, biological modelling, and clear visual representation. The key focus is now on obtaining a comprehensive set of data for N_2O, verifying the model, and trialing reduction strategies.

WORKSTREAM 2 – SOFT SENSORS FOR DRIFT DETECTION

PROBLEM STATEMENT

Accurate, reliable data on plant parameters is critical for improving and optimizing plant performance. This includes the development of reliable, effective digital tools as described in Workstream 1.

Unreliable plant instrumentation is a universal experienced at treatment plants. This can be attributed to multiple reasons - one of which being sensor drift, where instruments increase or decrease over time. This can be difficult for operators to identify visually due to the gradual gradient and similarities to actual plant upsets.

The objective of this workstream is to develop soft 'virtual' sensors of plant parameters to identify sensor drift events. Operators are alerted to sensor drift event as they occur and can then calibrate, clean or replace the sensor, improving the quality of data captured.

METHODOLOGY

The following steps were followed to develop the soft sensors:

1. Data Analysis
    a. Historic sensor data was reviewed alongside maintenance records to identify historic drift events
    b. Engage with instrument technicians and plant operators to understand general sensor quality.
    c. Data cleansing to remove anomalies and spikes from standard activities e.g., cleaning.
2. Model development
    a. Determine appropriate model type e.g., Machine learning model, biological model, process logic. Model selection is discussed further below.
    b. Identify model inputs through engagement with process engineers and plant operators
    c. Develop / train based on historic data
    d. Deploy soft sensor connected to SCADA data.
3. Drift Detection
    a. Calculate rolling-average period for physical sensor and soft sensor. Period to be determined based on nature of parameter.
    b. Monitor deviation between physical sensor and soft sensor rolling averages
    c. Determine threshold for sensor drift. The threshold value was initially identified experimentally by comparing the results of a large number of parameters on historical data in order to find values that resulted in the best detection performance.
    d. Apply wet weather exceptions to prevent alarming due to the impact of increased flows.
4. Performance monitoring
    a. Configure email and/or text alerts to operators when a threshold is exceeded
    b. Review sensor and capture feedback – e.g. correctly identified drift, false positive, false negative.
    c. Adjust configuration if required

DISCUSSION

The soft sensors were developed using a range of modelling techniques including mechanistic biological models and data driven machine learning models. An overview on how each soft sensor was developed and comparison of modeling techniques is discussed below. **Error! Reference source not found.**2 provides an overview of the models used for each soft sensor and contributing factors.

Table 2: Soft sensor modelling approach overview

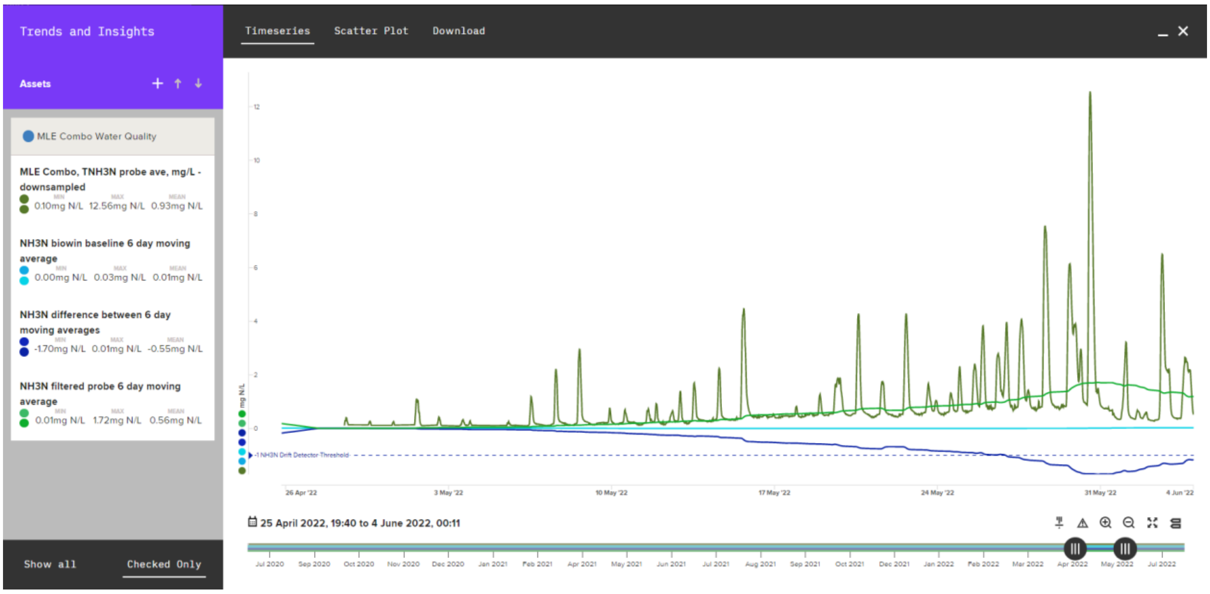
Parameter	**Ammonia**	**Suspended solids, Nitrite, Nitrate**	**Dissolved Oxygen**
Model Type	BioWin model	Machine Learning model	Statistical model
Reason	The BioWin model was already calibrated for ammonia & historic ammonia probe data was poor quality so machine learning wasn't a valid option.	The BioWin model was not calibrated for the parameters. Probe and/or sampling data was available to train the models.	DO operates at a defined set point. The relationship between the valve position and DO can indicate if the probe is fouling. A statistical model was used to monitor this relationship.
Location	Combined MLE outlet	Combined MLE outlet	MLEs 1,2,3 & 4
Alarm Configuration	The difference between probe & soft sensor 2-day moving averages was monitored. A filter was applied to exclude alarms during high rainfall periods.	The difference between the probe & soft sensor 6-day moving averages was monitored. Filter applied to exclude probe anomalies.	Threshold limit based on relationship between valve position % and DO probe value.

BIOLOGICAL MODEL – AMMONIA SOFT SENSOR

The biological (BioWin) model is a mechanistic model which determines plant values based on the chemical and biological reactions occurring in the treatment plant. Significant effort is required to calibrate a biological model – including comprehensive plant-wide sampling data and expert process knowledge. Once calibrated a biological model can be used to virtually trial control strategies and interventions.

The Ammonia soft sensor is generated from a biological model. The biological model inputs influent data to calculate the expected ammonia concentration in the reactors.

Figure 5: Ammonia soft sensor generated from biological model



During rainfall events the model inputs influent flow, rainfall and a fixed ammonia diurnal load profile to determine the approximate influent concentration. The digital twin has been configured to assume increased flow attributed to increased rainfall is dilute and does not contain additional load. This does not account for first flush phenomenon where a short ammonia peak occurs with the initial flow increase. This causes a limitation in the ammonia soft sensor, where the concentration is not representative during the initial rainfall. Increasing the quantity of influent load data captured would improve the accuracy of the soft sensor during wet weather events.

MACHINE LEARNING MODEL – NITRITE, NITRATE AND TSS SOFT SENSOR

The machine learning (ML) models used a Temporal Convolutional Neural Network (TCN) model. This is a highly effective algorithm for time series data and can automatically identify complex relationships in the input data and output data that increase predictive performance (Bai, 2018).

ML soft sensors were developed for Suspended Solids, Nitrite and Nitrate. By combining process engineering and data science knowledge the team were able to identify key patterns and features for each quality parameter. For example, the process engineer would identify typical parameters that impact TSS (WAS, flowrate, and load) and the data scientist would train the model based on the features that best predicted the data.

For some traces the probe data was clearly offset from the laboratory sampled data in the training period. In order to make this data usable, the training data was adjusted by subtracting the difference between the moving averages of the probe and sampled data. This adjusted training data was then used to train the model. The difference between the probes is not used in the prediction process.

By training the model on previous probe readings, fit to the trend of the sample data (taking an aggregated window on the lab-sampled data to eliminate noise) it was possible to get the best of both worlds: the short-term patterns modelled from the probe (modelling the diurnal behaviour); and the trend of the lab data. This trend is considered more trustworthy as the lab process is less likely to suffer from gradual drift than the online sensor.

As a result, the output looks somewhat like the real probe readings, but with the trend sitting closer to the laboratory samples, which were taken to be the true value.

Figure 6: TSS soft sensor generated from ML model

STATISTICAL MODEL – DISSOLVED OXYGEN SOFT SENSOR

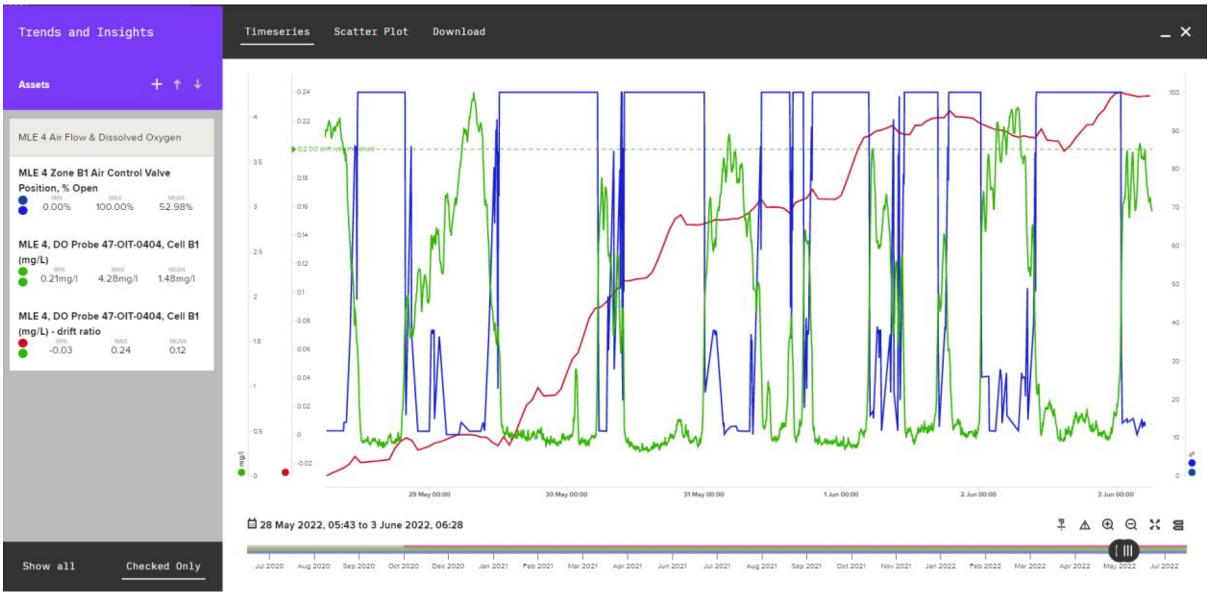
The performance of the MLEs bioreactors is heavily influenced by the levels of dissolved oxygen (DO) in the tanks. The Rosedale MLEs are controlled to a DO setpoint, meaning that the plant will adjust airflow into the MLE cells based on DO probe values to maintain a target level.

A statistical model was used to determine Dissolved Oxygen probe sensor drift, through monitoring the relationship between DO and valve position to identify drift.

DO drift events identified by plant operators are short term phenomena, defined as: DO probe drifts, the air flow and air valve position adjust to maintain the DO set-point, DO probe value remains constant even though probe is incorrect, the DO and air flow are much higher than required.

To identify these events the relationship between the air control valve position and DO probe rolling averages were monitored. Where the % of valve position opening doesn't cause the DO value to increase as it had previously, an alarm is raised.

Figure 7: Drift detection at MLE 4 Zone B1



Drift detection alerts have been deployed at 16 DO probe across MLE 1,2,3 and 4.

CONCLUSION

The use of soft sensors for drift detection has been demonstrated through the Rosedale innovation programme.

At the time of publishing this paper there are 20 operational soft sensors alerts at Rosedale WWTP. The performance and accuracy of the sensors are continuously monitored, with the Rosedale operational team responding to alarms and feeding back to the Moata data science team.

The continuous recording of drift alerts and outcomes e.g. true positive, false positive, or false negative, will support the verification of the soft sensors. A potential benefit of successful soft sensors is reduced reliance on physical sampling and probes. Long-term performance monitoring will build confidence in soft sensor reliability and open opportunity for further application.

When scaling to other plants or parameters, a decision between a mechanistic approach using a biological model or a data driven machine learning approach will be required.

As discussed above, a biological model requires significant upfront investment for calibration and relies on quality accurate influent flow and load data. While a machine learning model is faster and cheaper to deploy, it is not able to support trialling asset management scenarios.

Rosedale results comparing the two techniques indicate ML models can generate more accurate soft sensors, although the model is reliant on representative training data. As the model learns relationships from historic data, an operational change in the plant will not be immediately reflected in the model.

The decision between techniques therefore is dependent on whether there is an existing calibrated biological model, the availability of online flow and load influent data, and the availability of historic training data sets.

Soft sensors have the potential to significantly improve plant data quality, either through improved sensor maintenance or reducing reliance or unreliable sensors. Plant control systems and operators rely on probe data to control processes. It is therefore critical for any plant optimisation that this data is accurate and reliable. It is also fundamental for digital solutions, such as the GHG and opex dashboards developed in workstream 1.

For example, DO is a key parameter in the production of N_2O – too high or too low, depending on the metabolic conditions, and the risk of production increases. It is also a major contributor to opex through energy consumption, approximately $25,000 per week. DO probes are typically unreliable and intensive to maintain - the DO sensor drift alerts will improve the quality of data.

CONCLUSIONS

The two workstreams in Watercare's 2022 Rosedale innovation programme involved the creation of real time operational dashboards for GHG and Opex for the entire Rosedale WWTP, and the implementation of a soft sensor and drift detection programme for a number of key sensors in the MLEs.

The operational dashboards revealed hotspots for GHG and Opex in the Rosedale WWTP. The MLEs were identified as a hotspot contributing approximately 50% of the overall WWTP GHG emissions. The other major operational GHG contribution identified was the biosolids disposal which contribute approximately 40% of the overall GHG emissions for the plant. These two areas together represent the vast majority of the process emissions so it is clear that any mitigation strategies should be focused there. The opex dashboards also indicate that biosolids disposal is a major operational cost, representing 25% of total opex.

A variety of techniques were trialled as part of the soft sensor workstream: biological models, machine learning models, and statistical models. As outlined, different situations require different techniques, with a biological model being the best solution when a well calibrated model already exists, and a machine learning model being a faster and cheaper solution than calibrating a biological model when quality training data is available.

It has been shown that soft sensors have great potential for improving plant data quality by either improving sensor maintenance routines or reducing reliance on unreliable sensors. High quality data is crucial both for operational purposes such as controlling the WWTP, and for strategic analysis such as the operational dashboards created in the GHG and Opex workstream. Soft sensors therefore will play an important role in the future of waste water treatment.

ACKNOWLEDGEMENTS

The authors would like to thank the other members of the team for their contribution to this project, including Apra Boyle-Gotla (Watercare), David Hume (Mott MacDonald), Nick Dempsey (Mott MacDonald), and Helen Wang (Mott MacDonald).

REFERENCES

Andrews, J (2021) Carbon accounting guidelines for wastewater treatment: CH_4 and N_2O, Water New Zealand.

IPCC (2019). 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories Volume 1, Chapter 1.

Ministry for the Environment (2022). Measuring emissions: A guide for organisations: 2022 detailed guide. Wellington: Ministry for the Environment.

Bai, S (2018). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. Retrieved from the arxiv database.