Evaluation of Fluorescence Sensor to Monitor Wastewater, Assess Treatment Performance and Develop Warning Alarm Using Machine Learning

Project Team

Adelaide University: Christopher Chow, Jing Gao, Ben van den Akker, Ke Xing, Rameez Rameezdeen, Linda Shi

City of Alexandrina: Gary Lyons, Chloe Downes

Intelligent Sensing Technologies: Chris Moon

Local Government Association – Community Wastewater Innovation Fund. Project ID: 2025 CWMS-011

Acknowledgements

The authors gratefully acknowledge the financial support provided by the LGA Community Wastewater Innovation Fund 2025 (Local Government Association of South Australia). The authors also thank City of Alexandrina and Intelligent Sensing Technologies for their collaboration and technical support throughout the project.

CONTENTS

Contents **3**

Figures **4**

Tables **5**

1 Introduction **6**

2 Fluorescence Sensor applications overview **8**

3 Materials and Methods **10**

3.1 Goolwa CWMS 10

3.2 Sensor, Installation and Calibration 10

3.3 Influent Wastewater Sampling and Data Collection 12

3.4 Custom-Build Data Display Portal Development 13

3.5 Evaluation Procedures 13

4 Results and Discussion **15**

4.1 Custom-Build Data Portal 15

4.2 Influent Wastewater Flow 16

4.3 Influent Wastewater Quality (Laboratory Results) 17

4.4 Comparison between Laboratory and Fluorescence Sensor Measurements 19

4.5 Performance of the Conventional pH, Turbidity and Conductivity Sensors 20

4.6 Influent Wastewater Quality (Sensor Results) 21

4.7 Interrelationships among Wastewater Quality Parameters 25

4.8 Pattern Analysis Using Advanced Computing Techniques 27

4.9 Warning Alarm Development Using Machine Learning 31

5 Lesson Learned **34**

5.1 Sensor Setup, Calibration and Maintenance 34

5.2 Data Processing Software Comparison 34

6 Summary and Conclusions **35**

8 References **36**

Appendix I - Proteus Sensor Information **39**

Appendix II – The Custom-Build Data Display Portal **43**

Appendix III - Pattern Analysis Using Advanced Computing Techniques **44**

Appendix IV - H_2S Sensor Results **50**

FIGURES

Figure 1: Goolwa wastewater treatment plant. 10
Figure 2: Balance Tank of Goolwa wastewater treatment plant. 11
Figure 3: Fluorescence sensor installation and powered by solar with battery backup. 12
Figure 4: The custom-build data portal display with all the measured parameters. 16
Figure 5: Influent flow trends and hourly flow distribution of Goolwa WWTP from 25 September to 27 October 2025. 17
Figure 6: Wastewater quality trends of influent NH_3-N, pH and PO_4-P at Goolwa WWTP on 25 September 2025. 18
Figure 7: Wastewater quality trends of influent BOD5, COD, SS, TDS and alkalinity at Goolwa WWTP on 25 September 2025. 18
Figure 8: Linear regression between fluorescence sensor and laboratory measurements for (a) BOD_5 (mg/L), (b) COD (mg/L) and (c) PO_4-P (mg/L) for the influent wastewater of Goolwa WWTP. 19
Figure 9: Comparisons between sensor and laboratory (lab) measurements for (a)sensor and lab pH, (b) sensor turbidity and lab suspended solids, and (c) sensor conductivity and lab total dissolved solids for the influent wastewater of Goolwa WWTP. 20
Figure 10: Comparison between sensor and laboratory trends of influent wastewater quality parameters, BOD_5 (mg/L), COD (mg/L), and PO_4-P (mg/L), a) 1-hour average sensor trends and b) 15-minute sensor trends. 22
Figure 11: Trends of influent wastewater quality parameters, a) BOD (mg/L), COD (mg/L), and PO_4-P (mg/L), and b) turbidity (HU), conductivity (S/m), c) pH and temperature (^oC), and d) ADO (mg/L), measured by online probes at the Goolwa WWTP from 25 September to 27 October 2025. 23
Figure 12: Observed linear relationships between (a) laboratory COD and SS, and (b) sensor temperature and pH. 25
Figure 13: Trends of influent wastewater flow and dissolved oxygen concentration measured by online probes at the Goolwa WWTP from 25 September to 27 October 2025. 26
Figure 14: Trends of influent wastewater flow, COD and BOD_5 measured by online probes at the Goolwa WWTP from 25 September to 27 October 2025 with the School Holidays and non-School Holidays labelled. 27
DTW < 10 = high similarity 28
Figure 15: Daily pattern comparison in DTW (Group by individual days): a. Flow; b. COD reading. 28
Figure 16: Daily flow pattern comparison in DTW (Grouped by weekdays and weekends): a. 15 mins interval; b. 30 mins interval. 29
Figure 17: Daily COD pattern comparison in DTW (Grouped by weekdays and weekends): a. 15 mins interval; b. 30 mins interval. 30
Figure 18: Daily COD pattern comparison in DTW (Grouped by individual days, 30 mins interval): a. Flow; b. COD. 31
Figure 19: Flow AR vs LTSM 32
Figure 20: COD AR vs LTSM 33

TABLES

Table 1: Flow data timestamp issue 15
Table 2: H_2S data timestamp issue. 15
Table 3: Mean Absolute Error (MAE) and bias of sensor and laboratory measurements of BOD_5, COD, PO_4-P and pH. 21
Table 4: Flow prediction model comparison – Lower value is better 32
Table 5: COD prediction model comparison – Lower value is better 33

1 INTRODUCTION

Wastewater services are essential service provided by water entities to their customers. Community Wastewater Management Systems (CWMSs) are typically used to service regional areas and rural areas and usually consist of small-scale wastewater treatment plants (WWTPs). A CWMS comprises a wastewater collection and sewerage system that transports wastewater to a WWTP for treatment, storage, and subsequent disposal or reuse. CWMSs include different types of wastewater schemes, such as the Septic Tank Effluent Drainage Scheme (STEDS), a full sewer, or a combination of the two. The majority of the existing CMWSs are STEDS, which utilise on-site septic tanks connected to provide primary treatment before wastewater is conveyed via a community sewer network to a centralised WWTP. Discharge from the septic tanks typically flows by gravity through the collection system and wastewater pump stations by gravity or is pumped through sewer mains to the treatment plant. In South Australia, CWMS are primarily managed by local councils, which often have limited resources and service relatively small population (LGA 2023). As a result, influent wastewater quality of CWMSs is typically not routinely monitored, given the absence of specific regulatory requirements and resources limitations. However, influent quality in CWMSs can be highly variable and unpredictable. The presence of unwanted substances or shock loads can negatively affect the performance of the treatment process, such as upsetting the biological treatment process. Monitoring the influent wastewater quality is therefore important to support operational management and improve the reliability of treatment performance.

Conventional wastewater quality monitoring relies on grab sampling and standard laboratory analysis. As the majority of CWMSs are in the regional and rural areas, significant time is often required to travel to sites, collect samples, transport samples to laboratories and received analytical results. Due to limited number of grab samples typical collected, the results may not adequately represent the true variability of the wastewater quality over time. Short-duration or transient wastewater quality events may be missed, resulting in treatment plants being managed largely in a reactive manner (Shi *et al*, 2024). In contrast, online wastewater quality monitoring provides real-time measurements and continuous data, which allows earlier detection of wastewater quality changes and more proactive operational responses when a wastewater event occurs, such as a sudden increase in trade waste load (Zhang *et al*., 2023). Over the last decade, the use of online instruments to monitor CWMSs has become increasingly common, particularly for simple parameters such as pH and temperature. A wide range of online instruments is available for wastewater monitoring, ranging from relatively simple sensors such as turbidity meters to sophisticated sensors such as fluorescence sensors.

Fluorescence spectroscopy is optical device used for the real-time detection of organic pollutants and microbes. Fluorescence sensors are primarily used to monitor dissolved organic matter (DOM), protein-like organic compounds, indicators of microbial activity, and overall wastewater contamination and treatment performance. Water utilities use fluorescence sensors and probes because they are fast, highly sensitive to changes in organic matter, reagent-free, and well suited for early-warning applications. Fluorescence sensors are routinely used in certain applications, particularly in recycled water schemes and advanced treatment processes. They are not yet universally adopted for all wastewater or treated water quality monitoring applications; however, their operational use is steadily increasing as confidence in the technology grows. Fluorescence sensors have been trialled by many

Australian water utilities, mainly through collaborative research involving utilities and universities. Multiple full-scale utility trials have been conducted across New South Wales, Queensland, Victoria, and South Australia since 2008. These trials were led or supported by organisations such as Sydney Water and Melbourne Water. Applications include recycled water plants, dual reticulation schemes, advanced water recycling facilities, and drinking water treatment plants.

Advanced data analytics techniques, including machine learning, have been increasingly applied to online wastewater monitoring in recent decades. In the context of fluorescence-based monitoring of wastewater, most machine-learning applications rely on fluorescence spectral data (Carstea *et al*., 2016; Khan *et al*., 2022; Marino *et al*., 2025). For example, Yang et al. (2024) developed multi-target random forest regression models using fluorescence spectra to predict the removal of trace organic contaminants from municipal secondary effluent during an advanced oxidation process, achieving high predictive accuracy with R^2 values ranging from 0.83 to 0.95. Similarly, Nurhayati et al. (2023) applied an artificial neural network to estimate DOC concentrations in mixed agricultural and domestic wastewater influent using fluorescence spectra. In their study, fluorescence Peak T (tryptophan-like fluorescence maximum at excitation 275 nm and emission 340 nm) was identified as a strong predictor, yielding an R^2 of 0.89. However, very few studies have reported on applying machine learning to fluorescence sensor monitoring of small wastewater treatment systems.

This project aimed to address the current challenge of lacking online / real-time analytical tool by developing rapid assessment capability that reveals the status of the systems in “real-time” with decision support. This project assessed the feasibility of using a fluorescence sensor to continuous monitor the influent wastewater quality of a CWMS in regional SA. Goolwa wastewater plant was selected for the fluorescence sensor real-time monitoring trial. In addition, the fluorescence sensor (Proteus) is a multi-sensor system which contains other sensors were included in the trial. Those recorded sensor signals were used to develop an early warning system of the influent wastewater quality changes to support the treatment operations using machine learning**.**

2 FLUORESCENCE SENSOR APPLICATIONS OVERVIEW

Fluorescence sensor operates on the principles of fluorescence spectroscopy by measuring fluorescent light emitted by molecular in water. The intensity of the emitted light can be used to identify specific contaminants. Many organic and biological compounds present in wastewater naturally fluoresce when exposed to light of a specific wavelength. Fluorescence sensors are mainly used to monitor DOM, protein-like organic compounds, microbial activity indicators and wastewater contamination and treatment performance. They are highly sensitive tools for detecting and characterising DOM. These sensors can be deployed in situ, allowing continuous monitoring without the need for sample collection and laboratory analysis. However, in-situ applications face several challenges, including biofouling, changes in turbidity, signal interference, and the need for regular calibration to maintain accuracy. Site-specific algorithms are often required to correct for multiple parameters such as turbidity, temperature, and inner-filter effects. While factors like turbidity and quenching can affect fluorescence readings, modern fluorometers incorporate advanced correction algorithms that help compensate for these interferences, improving data reliability even under difficult operating conditions. Numerous studies have demonstrated that fluorescence spectroscopy can be used as an online surrogate monitoring and early-warning tool for real-time wastewater quality assessment (Khan *et al*., 2022). In one study, a fluorescence sensor was applied to track the performance of a decentralised treatment plant treating domestic wastewater from a community in South Africa. Changes in fluorescence-based metrics, particularly the humification index, showed strong correspondence with chemical oxygen demand (COD) removal efficiencies under normal operation and during disturbance events associated with scum removal (Mladenov *et al*., 2024). These results indicated that fluorescence-based indices can serve as robust indicators for identifying performance variations and operational issues in decentralised wastewater treatment systems.

Several studies have also demonstrated the potential of fluorescence spectroscopy to monitor the removal of micropollutants during advanced treatment processes. Fluorescence spectroscopy has been proposed for monitoring contaminants of emerging concern (CECs) based on correlations between CEC concentrations and characteristics of dissolved organic matter (Fazzino *et al*., 2025). Fluorescence sensors, using indices measured at specific excitation/emission wavelength pairs, can be integrated into commercially available or custom-designed field-deployable systems, enabling near real-time monitoring of CECs in wastewater. Membrane bioreactor studies and review papers suggest that such systems are potentially suitable for real-time contaminant monitoring (Sgroi *et al*., 2017; Xin *et al*., 2020; Augsburger *et al*., 2021).

Fluorescence excitation/emission matrices were used as a tool to monitor the removal of organic micropollutants from wastewater effluents by adsorption onto activated carbon (Guillossou *et al*., 2021). Correlations were found between the removal of spectroscopic indexes and most organic micropollutants. This review is demonstrated that the application of fluorescence spectroscopy to the monitoring of adsorption processes at the industrial scale can accurately track the evolution of organic micropollutants. Fluorescence excitation–emission matrices (EEMs) have also been successfully used to monitor the removal of organic micropollutants from wastewater effluents through adsorption onto activated carbon (Guillossou *et al*., 2021). Strong correlations were observed between the reduction of selected

spectroscopic indices and the removal of most organic micropollutants. Guillossou *et al*. (2021) believe that fluorescence spectroscopy can be effectively applied at industrial scale to track the evolution and removal of organic micropollutants during adsorption processes.

Real-time control of tertiary wastewater treatment was demonstrated through nine days of continuous, 24-hour monitoring of tertiary effluents from two municipal WWTPs in Sicily, Italy. High correlations between fluorescence sensor signals and COD concentrations were observed, suggesting that fluorescence sensors are promising tools for rapid wastewater quality assessment, particularly for COD monitoring (Marino *et al*., 2025). Furthermore, a multi-parameter sonde incorporating two fluorescence peaks, turbidity, temperature, and electrical conductivity was used to provide biochemical oxygen demand (BOD) surrogate measurements at three WWTPs in the United Kingdom over a ten-month period (Khamis *et a*l., 2021). The sensors were installed after primary settling and in final effluent streams. Laboratory analyses were used for calibration, and site- and process-specific sensor calibrations yielded the most accurate results. This study demonstrated the potential of online optical sensors to function as robust BOD surrogates in WWTP monitoring.

3 MATERIALS AND METHODS

3.1 Goolwa CWMS

Goolwa Wastewater Treatment Plant (WTTP) was selected for this study. The Goolwa CWMS is a combination of STEDS and full sewered connections. Wastewater is conveyed via a network of gravity drains and pumping stations as well as vacuum mains to Goolwa WWTP. The influent consists primarily of domestic wastewater, with a small proportion of trade wastewater. The Goolwa WWTP services the townships of Goolwa and Hindmarsh Island and is located on the northern outskirts of Goolwa along Kessell Road. The treatment system is a lagoon-based process, as shown in Figure 1, and includes sequencing batch reactors (SBRs) for treatment, lagoon is for effluent storage and disinfection via 50-day detention time only. The lagoon system consists of well-maintained ponds surrounded by grassy embankments and native vegetation. All domestic wastewater collected by the CWMS is treated and recycled. The recycled wastewater is used for agriculture in Goolwa, contributing to reduced demand for mains water and River Murray water supplies. The designed treatment capacity of the CWMS is 1.5 ML/day, while the current average influent flow is approximately 0.9 ML/day.



Figure 1: Goolwa wastewater treatment plant.

3.2 Sensor, Installation and Calibration

A Proteus multiparameter water quality sensor – Proteus 35, Proteus Instruments, was installed in the balance tank (Figure 2) of the Goolwa WWTP, co-located with an ISCO autosampler (24 individual 1 litre sampling bottles). The instrument is powered by a solar energy system with battery backup, as shown in Figure 3. The Proteus multiparameter probe utilises advanced fluorescence spectroscopy to provide real-time, high-sensitivity measurements of organic pollution. The probe targets naturally occurring compounds in water that emit fluorescence when excited by specific ultraviolet wavelengths. The instrument incorporates manufacturer-provided algorithms designed to compensate for environmental factors that may interfere with fluorescence signals, such as temperature variation and turbidity. To enable long-term field deployment, the sensor is equipped with an integrated automatic cleaning mechanism to reduce the impacts of biofouling. The Proteus platform supports a wide range of water quality measurements and can be configured to simultaneously monitor multiple parameters, including BOD, COD, DOC, TOC, Coliforms (Total, E. coli, Faecal), TAN, Ortho-phosphate, Temperature, Dissolved Oxygen, Pressure, pH, ORP/REDOX, EC/Salinity/TDS, Turbidity/TSS, Ammonium, Nitrate, Chloride, Sodium, Calcium, Bromide, Chlorophyll a (red excitation), Chlorophyll a (blue excitation), Rhodamine, Blue

Green Algae (freshwater), Blue Green Algae (marine), CDOM, Optical Brighteners, Tryptophan, Fluorescein Dye, Refined Oil, Crude Oil. The probe (Proteus 35) can be configured with up to 11 sensors on a single instrument (Proteus 40 up to 13 sensors), allowing the real-time output of more than 30 water quality parameters. For this trial, eight parameters, BOD, COD, Ortho-phosphate, Temperature, Dissolved Oxygen, pH, EC/Salinity/TDS, Turbidity/TSS, were selected. Prior to installation, the sensor was cleaned using ultrapure water and calibrated in accordance with the manufacturer's specifications. A Consibio gas phase H_2S sensor was also installed to monitor H_2S (the sensor was available at that time and added in as an additional monitoring tool).



Figure 2: Balance Tank of Goolwa wastewater treatment plant.

Further details of Proteus Sensor properties (including linear range and detection limits) are outlined in the Appendix 1. The sensor was attached to metal railings using a stainless-steel chain and all sensing elements were submerged to a minimum depth of 0.3 m at a location that was well mixed. An autosampler was used to collect discrete water samples from the balance tank. The inlet tube was situated in the flow at the same depth as the sensors and a coarse strainer was fitted to ensure no blockage of the tubing occurred during sample collection.





Figure 3: Fluorescence sensor installation and powered by solar with battery backup.

Initial sensor calibrations were conducted on site, pH, Conductivity, Turbidity calibrated using reference standards available off-the-shelf. DO by shaking a bottle vigorously to oxygenate water, 100% only. BOD / COD / Ortho-P via provided sample results from samples collected in Day 1 of the trial and analysed by commercial laboratory (described below) using Proteus proprietary algorithms (calibration procedure in Appendix 1).

3.3 Influent Wastewater Sampling and Data Collection

Measurements were taken every 15 minutes using the Proteus sensor and the data was relayed to a remote server through a custom telemetry setup. The web-based portal was hosted by the Intelligent Sensing Technologies (local distributor), it provided real-time data display functions, single parameter against time and overlapping of several parameters with zoom and zoom out option (screen capture in Appendix 1). Users can use this portal for real-time monitoring and basic data processing. The ISCO autosampler was synchronised with the Proteus (internal clocks cross-referenced on each deployment) and programmed to start sampling at 7 am on the sampling day (25th September 2025), and to continue at one-hourly intervals until 6 pm, giving a total of 12 samples (350 mL was collected for each sample time using a 1 L sample bottle) for the initial sensor calibration. Ten water quality parameters, NH_3-N, NO_3-N, NO_2-N, pH, PO_4-P, BOD_5, COD, SS, TDS and alkalinity, were sent to laboratory for analysis. In addition, the autosampler continued to collect additional samples for BOD_5 analysis until 5:00am the following day (a total of 24 BOD_5 samples were collected). Additional weekly grab samples were obtained manually for 4 weeks. Samples were kept on ice and sent to a laboratory where BOD_5 was determined according to standard methods. Full chain of custody was ensured with samples temperature controlled during transport and BOD_5 analysis initiated within 24 h. A four-week (25/9/25 to 27/10/25) monitoring trial was conducted at the Goolwa WWTP using the Proteus Multiparameter Water Quality Probe, to monitor BOD, COD, Conductivity (TDS), Turbidity (TSS), pH, Dissolved Oxygen, Orthophosphate (PO_4-P), temperature to cover both school holiday period vs normal period. TDS and TSS were post-calibrated using the analytical results of the hourly samples on 25/9/25.

Flow data from a flow meter were extracted from a historical database provided by Alexandrina Council. The flow measurements were timestamped and recorded at irregular intervals, ranging from approximately 2 to 10 measurements per minute in the extracted

dataset. To align with the sensor reading interval, the data were averaged over 15-minute periods.

H_2S data were collected from the H_2S sensor at one minute interval. H_2S measurement was not initially planned for this project, however, a new sensor was available from Intelligent sensing technologies and installed into the monitoring system. H_2S sensor result is presented in Appendix 4.

3.4 Custom-Build Data Display Portal Development

A custom-build web-based data display portal was developed using Apache Superset, an open-source platform for modern data exploration and interactive visualization. The portal was connected to the database containing the ten monitored parameters—BOD, COD, conductivity (TDS), turbidity (TSS), pH, dissolved oxygen, orthophosphate (PO_4-P), temperature, flow, and H_2S. It provided four time-series visualizations (parameter vs. time) for Mean, Aggregate, Peak, Median, and Hourly Flow Distribution displays, each with user-selectable to display which parameters. Users could also adjust four display options: Time Gain (default: every 15 minutes), Datetime Range (default: full trial period, with custom start and end dates allowed), Weekday/Weekend (default: both, with the option to view only weekdays or weekends), and Day of Week (default: all days, with the option to select a specific day). These options were specific designed for this trial based on previous experience. A scaling factor was applied to each parameter to bring the data into a comparable range for visualization; this adjustment is discussed later in the report.

3.5 Evaluation Procedures

All analytical data including flow measurements, sensor results, and laboratory results were organised and stored in Microsoft Excel spreadsheets according to sample collection times (for laboratory analyses) or sensor measurement times. Excel was initially used for basic data processing, including temporal averaging of measurements, data visualisation, and statistical analyses such as linear regression and error analysis. Subsequently, Excel was replaced by a custom-build data display portal and Python-based workflows. Python is an open-source programming language widely used for web development, data science, artificial intelligence, and automation.

Recorded sensor parameters including flow from the flow sensor; COD, BOD_5, and PO_4-P from the fluorescence sensor; and pH, turbidity (as suspended solids, SS), and conductivity (as total dissolved solids, TDS) from other sensors were systematically analysed. Statistical analyses were conducted to compare sensor measurements with laboratory grab sample results. Interrelationships among key parameters were also examined, including COD versus SS, pH versus temperature, dissolved oxygen versus flow, and the relationships between COD, BOD, and flow during school holiday and non-school holiday periods.

Further analysis of the flow and COD data was conducted using Dynamic Time Warping (DTW). DTW was selected to identify temporal patterns, assess correlations between parameters, and detect abnormal behaviour in the data. Python was used as the programming platform for this analysis. Other parameters were processed and presented in Appendix 3.

A time-series prediction model based on flow and COD data was developed for the warning alarm system. Long Short-Term Memory (LSTM) network, a special form of Artificial Neural Network (ANN), was implemented using Python to construct and evaluate the model. The LTSM prediction model results on flow data and COD were compared against the conventional Autoregressive (AR) modelling method.

4 RESULTS AND DISCUSSION

4.1 Custom-Build Data Portal

Upon data collection, this project requires a robust integration of data visualisation and advanced analytics, specifically employing neural networks for a machine-learning-based forecasting model. Crucially, the chosen software platform must remain user-friendly to ensure participating organisations can undertake data exploration with minimal training. After evaluating various options—ranging from generic business intelligence tools like Power BI to scientific libraries such as scikit-learn—the project team has selected Apache Superset as the demonstration platform, supported by the machine learning capabilities of scikit-learn.

While generic tools like Power BI excel at addressing "known" questions through polished, drag-and-drop dashboards, Apache Superset is architected for the "unknown unknowns" inherent in scientific discovery. By treating SQL as a first-class citizen rather than a hidden backend detail, it provides a transparent IDE that ensures research logic is auditable and reproducible—a feature critical to this project's integrity. For instance, datasets such as flow sensor and H_2S readings (as shown below in Table 1 and Table 2) often exhibit different sampling frequencies and non-equidistant intervals. To address this, normalisation is required during pre-processing; Superset's built-in SQL functions allow these normalisations (e.g., 15, 30, or 60-minute intervals) to be performed dynamically.

Table 1: Flow data timestamp issue

Flow Sensor reading	Inlet Flow (L/s)
24/09/2025 6:24:54	0
24/09/2025 6:25:24	2
24/09/2025 6:25:26	3
24/09/2025 6:25:42	2
24/09/2025 6:25:46	1

Table 2: H_2S data timestamp issue.

H_2S Sensor reading	
24/09/2025 12:10	187.104
24/09/2025 12:10	131.659
24/09/2025 12:11	83.5235
24/09/2025 12:11	51.13575
24/09/2025 12:12	18.748

Furthermore, Superset's open-source extensibility ensures it is not restricted to standard charts. Researchers can develop bespoke visualisation plugins for specialised geospatial or statistical plots that proprietary tools often struggle to render. By shifting the focus from static reporting to dynamic, code-driven exploration, Superset functions less like a mere presentation layer and more like a high-powered microscope for data, serving as an essential bridge between raw database architecture and high-integrity scientific analysis.

It is important to note that while Apache Superset excels as a sophisticated exploration layer, it does not natively train deep learning models. However, its built-in analytics suite provides immediate tools for time-series resampling, rolling averages, and period-over-period comparisons. To complement this, scikit-learn has been selected as the advanced analytics engine. Scikit-learn remains a cornerstone of the machine learning ecosystem, offering a refined framework for classical predictive modelling. It also serves as a robust entry point for neural network development through its Multi-layer Perceptron (MLP) implementations, allowing for the rapid prototyping of supervised tasks without the architectural overhead of more complex frameworks.

Ultimately, combining scikit-learn with Apache Superset creates a powerful end-to-end research pipeline that bridges the gap between sophisticated data science and interactive discovery. While scikit-learn provides the mathematical "engine room" for model training, Apache Superset acts as the transparent "command centre" where these complex outputs are visualised and interrogated. This synergy allows researchers to move beyond static reports; instead, they can use Superset to monitor model performance in real-time, perform ad-hoc "what-if" analysis on prediction residuals, and share insights with stakeholders through an auditable, SQL-backed interface.

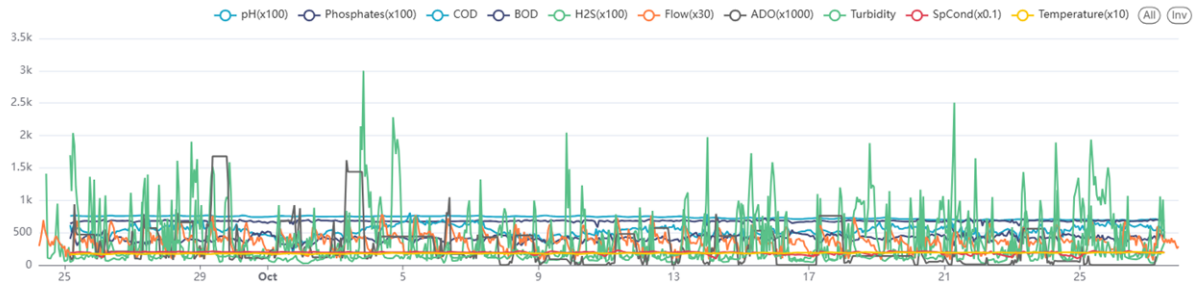


Figure 4: The custom-build data portal display with all the measured parameters.

An example of the custom-build data portal is shown in Figure 4. The portal features a sliding display that allows users to view parameter concentrations at specific points in time. Applying scaling factors to normalise variables to comparable levels proved to be an effective approach for enhancing data visualisation. In addition, the portal supports multiple simultaneous instances, enabling users with two or more computer screens to display and compare different datasets or time periods side-by-side. This functionality is particularly useful for comparative analysis across different operational conditions or timeframes.

4.2 Influent Wastewater Flow

Figure 5 illustrates the temporal variation and distribution of influent flow at the Goolwa WWTP over the monitoring period from 25 September to 27 October 2025. Based on hourly averaged data, Influent flow rate ranges from around 3 L/s to 26 L/s and it reveals distinct diurnal patterns, with lower flows typically occurring overnight and higher flows during the morning and daytime periods, reflecting potential wastewater generation patterns and pump station operations within the CWMS. In general, according to the hourly flow distribution, the influent flow rate increases from around 6 am, reaches a peak at approximately 9 am, then declines until about 11 am. The flow remains relatively steady until around 7 pm, after which it decreases again and stabilises overnight until 6 am next day. The distribution of hourly flows

indicates that most inflow volumes cluster around the average operating range, with occasional higher-flow events that may be associated with short-term increases in wastewater input, such as trade waste discharges or operational variability. Overall, the flow data provide important hydraulic context for interpreting influent wastewater quality trends, as variations in flow can influence pollutant concentrations and affect loading to the treatment system.

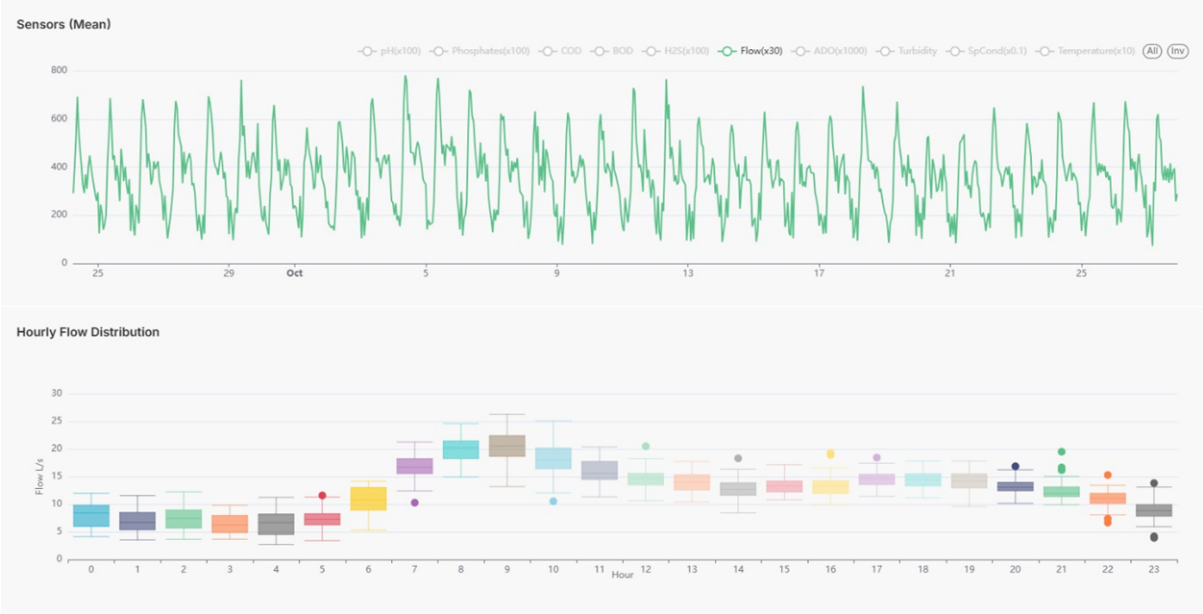


Figure 5: Influent flow trends and hourly flow distribution of Goolwa WWTP from 25 September to 27 October 2025.

The observed influent flow pattern is consistent with typical diurnal flow behaviour reported for municipal wastewater treatment plants, characterised by a morning peak associated with domestic water use, followed by lower and more stable flows overnight (Sydney Water, 2021; Do et al., 2022). Similar diurnal patterns are commonly observed in Australian domestic and mixed domestic–trade waste catchments, including small and regional systems (Simmonds & Bristow, 2019). Understanding these hydraulic trends is critical, as flow variations directly influence pollutant loading and concentration through dilution. From the initial visual inspection between weekdays and weekends, three are not observable differences (additional data analysis later in advanced computing section).

4.3 Influent Wastewater Quality (Laboratory Results)

Laboratory results for influent wastewater on 25 September 2025 are shown in Figures 6 and 6, including NH_3-N, pH, PO_4-P, BOD_5, COD, SS, TDS and alkalinity. Figure 6 shows that influent pH remained relatively stable through the day (7.4 to 7.7), whereas NH_3-N varied more strongly (65-71 mg/L), with distinct peaks at approximately 09:50 and 14:30. These NH_3-N concentrations are indicative of relatively strong nitrogen strength compared with typical influent domestic wastewater with high trade waste ranges (40-80 mg/L) reported in Australia (Simmonds and Bristow, 2019; Sydney Water, 2021). NO_3-N concentrations are below reporting limit (<0.01 mg/L), while NO_2-N mostly below reporting limit (<0.01 mg/L) and only two samples with 0.02 mg/L and 0.03 mg/L. PO_4-P ranged from 6.3 to 7.0 mg/L and showed some co-variation with NH_3-N, but the magnitude and timing of changes were less consistent. PO_4-P concentrations of 6.3-7.0 mg/L in the influent municipal wastewater are considered

normal to moderately high and fall well within typical municipal wastewater ranges. High BOD_5 generally means higher treatment demand and stronger sewage. High SS can cause operational issues (blockages, settling load, sludge production). TDS can affect some sensors and biological performance at extremes.

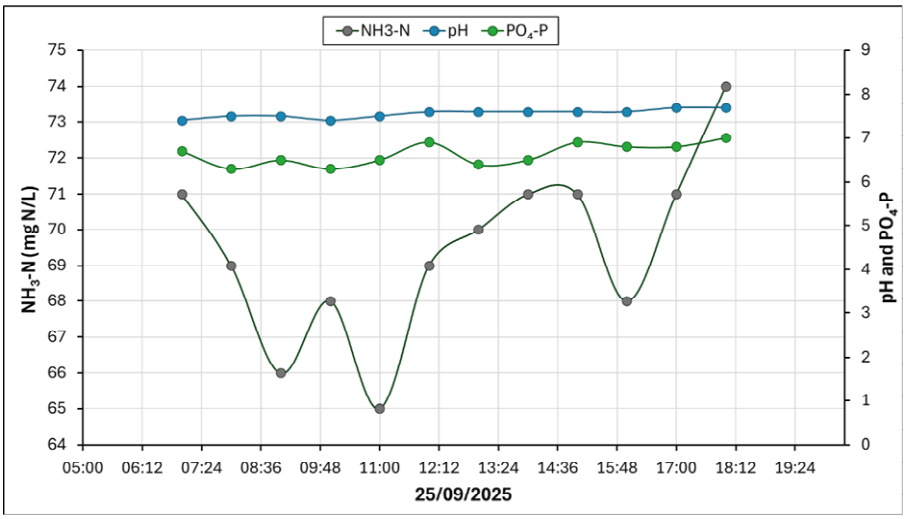


Figure 6: Wastewater quality trends of influent NH_3-N, pH and PO_4-P at Goolwa WWTP on 25 September 2025.

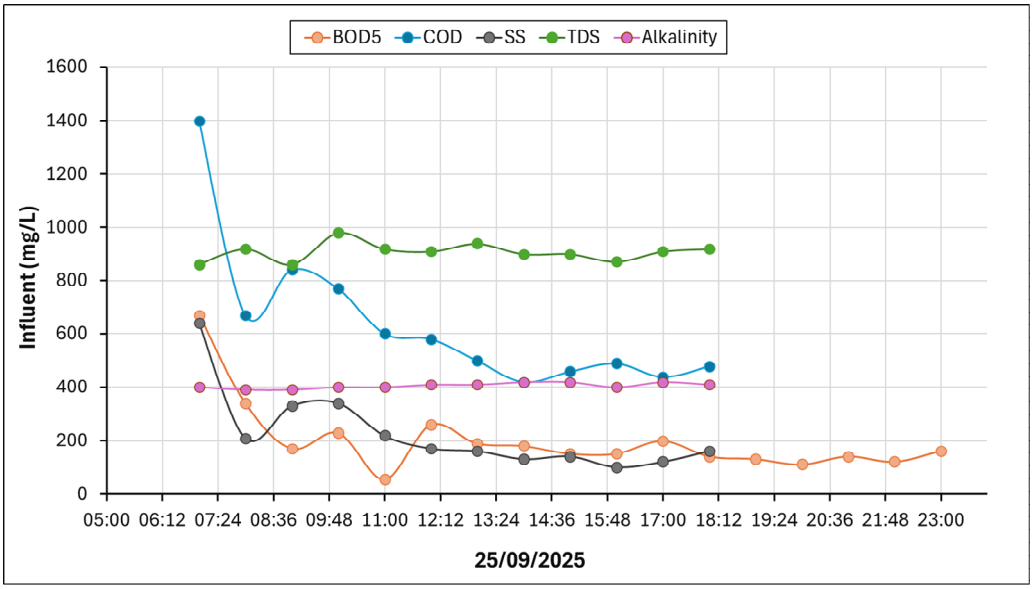


Figure 7: Wastewater quality trends of influent BOD5, COD, SS, TDS and alkalinity at Goolwa WWTP on 25 September 2025.

Figure 7 shows that BOD_5 and COD concentrations fluctuated throughout the day and did not always vary in parallel, reflecting temporal changes in influent composition. SS, TDS, and alkalinity trends provide additional insight into these within-day variations: changes in suspended and dissolved solids, as well as buffering capacity, can coincide with periods of increased organic loading and may influence treatment process performance. COD values ranging from 440 to 1400 mg/L align with typical influent characteristics observed in domestic systems with low to high proportions of trade waste. Elevated COD values may indicate potential trade waste discharges. BOD_5 concentrations ranged from 55 to 670 mg/L, with most

measurements below 400 mg/L, suggesting that the influent primarily consists of domestic wastewater with light trade waste contribution.

The elevated concentrations of BOD_5, COD, NH_3-N, and SS observed at 7 am indicate a high likelihood of trade waste discharge during this period. Additional monitoring may be needed if there are concerns about the impact on wastewater treatment performance.

4.4 Comparison between Laboratory and Fluorescence Sensor Measurements

A fluorescence sensor was used to monitor influent wastewater quality of Goolwa WWTP for about a month from 25 September to 27 October 2025. BOD_5, PO_4-P and COD concentrations derived from the fluorescence sensor measurements of influent wastewater were assessed against the standard laboratory measurements. BOD_5, COD and PO_4-P from the sensor were calibrated by the instrument supplier using proprietary Proteus algorithms and the corresponding laboratory data of these three parameters was collected on 25 September 2025. The number of paired sensor-laboratory data used to assess the performance of the fluorescence was 12 for both COD and PO_4-P but 18 for BOD_5, and laboratory results were based on discrete grab samples. Prior to the assessment, hourly averages of the online influent wastewater sensor measurements were calculated to align with the sampling times of the corresponding laboratory measurements. Linear regression, Mean Absolute Error (MAE), and Bias were selected for the assessments as these three methods have been used by many studies to evaluate water quality-based sensor calibration and accuracy (Harvey, 2016; Samuelsson *et al*., 2023; Zimmerman *et al*., 2017). MAE is used to assess average sensor errors, widely used in calibration studies. Bias is used to measure whether a sensor consistently overestimates or underestimates compared to standard laboratory measurements.

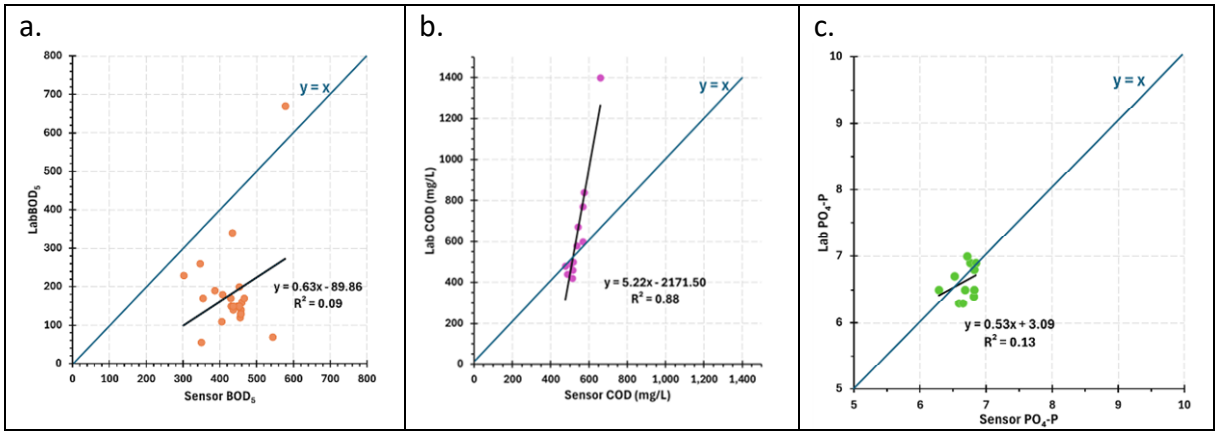


Figure 8: Linear regression between fluorescence sensor and laboratory measurements for (a) BOD_5 (mg/L), (b) COD (mg/L) and (c) PO_4-P (mg/L) for the influent wastewater of Goolwa WWTP.

According to the assessment results in Figure 8 and Table 3, for BOD_5, linear regression between sensor and laboratory data indicated weak agreement ($y = 0.63x - 89.86$, $R^2 = 0.09$), suggesting that the sensor calibration did not reliably capture variability in laboratory BOD_5. The slope (<1) further indicates under-response at higher concentrations. The fluorescence sensor showed a high error, with an MAE of 237.29 mg/L, and a substantial negative bias (-227.01 mg/L), demonstrating systematic underestimation of BOD_5 concentrations. For COD, linear regression showed a strong relationship between sensor and laboratory data (y = 5.22x

- 2171.50, R^2 = 0.88), indicating that the sensor captured temporal variability reasonably well. However, the slope and intercept suggest a scaling discrepancy that would benefit from additional calibration. Error was lower than for BOD_5, with an MAE of 136.96 mg/L and a relative MAE of 21.48%. The positive bias (99.30 mg/L) indicates a tendency to overestimate the COD concentrations. For PO_4-P, linear regression indicated large deference between sensor and laboratory measurements (y = 0.53x + 3.09, R^2 = 0.13), suggesting that sensor-derived values did not reliably track laboratory variability. The slope (< 1) and positive intercept indicate both scaling and offset effects. The sensor showed relatively low absolute error (MAE = 0.21 mg/L) and a small positive bias (0.06 mg/L), suggesting slight overestimation.

4.5 Performance of the Conventional pH, Turbidity and Conductivity Sensors

In Figure 9, although pH, turbidity and conductivity sensors are integrated within the same multiparameter platform as the fluorescence sensor, these sensors are conventional and operate independently of fluorescence-based measurements. The pH sensor uses a glass electrode to measure hydrogen ion activity, providing a direct indication of wastewater acidity or alkalinity. The conductivity sensor employs electrodes to measure the ability of water to conduct an electrical current, which is commonly used as a surrogate indicator of TDS. Turbidity is measured using an optical light-scattering method, which quantifies the cloudiness of water caused by suspended particles.

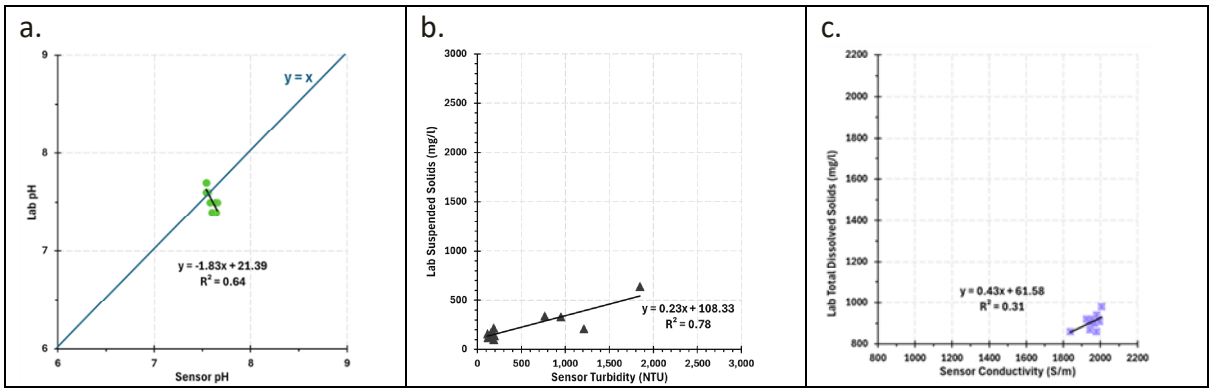


Figure 9: Comparisons between sensor and laboratory (lab) measurements for (a)sensor and lab pH, (b) sensor turbidity and lab suspended solids, and (c) sensor conductivity and lab total dissolved solids for the influent wastewater of Goolwa WWTP.

The number of paired sensor-laboratory data used to assess the performance of the fluorescence was 12 for all three pairs. A linear regression between sensor and laboratory data indicated moderate agreement between sensor pH and lab pH (y = 1.83x+21.39, R^2 = 0.64), suggesting that the sensor calibration did not reliably capture variability in laboratory pH. The MAE of 0.46 pH units indicates a moderate level of disagreement between sensor and laboratory pH measurements. The negative bias (−0.26 pH units) shows that the sensor systematically underestimates pH relative to laboratory values. While the magnitude of error is relatively small compared to many other wastewater quality parameters, regular field validation and recalibration are required to ensure confidence in pH measurements.

There is a reasonably strong agreement between the sensor turbidity and lab suspended solids, y = 0.23x +108.33, R^2= 0.78, indicating that with a simple site-specific conversion,

turbidity can be used as a reliable surrogate for SS in influent wastewater monitoring. This relationship suggests that continuous turbidity measurements provide a practical means of tracking suspended solids variability and capturing short-duration particulate loading events that may be missed by conventional grab sampling. A very weak relationship between sensor conductivity and lab TDS ($y = 0.43x + 61.58$, $R^2 = 0.31$) indicates further site-specific calibration is needed for the conductivity sensor. These results emphasise the need for site-specific calibration prior to operational use of turbidity as an SS surrogate, conductivity as an TDS surrogate in influent wastewater.

Table 3: Mean Absolute Error (MAE) and bias of sensor and laboratory measurements of BOD_5, COD, PO_4-P and pH.

	MAE	**Bias**
BOD_5	237.29 mg/L	227.01 mg/L
COD	136.96 mg/L	-99.30 mg/L
PO_4-P	0.21 mg/L	0.06 mg/L
pH	0.46	-0.26

These discrepancies likely reflect limitations in the proprietary calibration. However, the sample sizes used were relatively small, which may limit the robustness of the calibrations. This raised the question of analytical accuracy in comparison to laboratory measurement versus the benefit of continuous real-time monitoring of water quality provided to the operators. In addition, the calibration procedure recommended by the manufacturer was not conducted in full due to sample volume limitation of the autosampler, it was recommended to have duplicate samples sent to two different laboratories for the same analysis (Appendix 1 for details). Overall, the fluorescence sensor performed undesirably in terms of measurement accuracy when compared with laboratory results. This is consistent with findings from previous studies, which have shown that achieving accurate wastewater measurements using fluorescence sensors can be challenging due to the highly variable nature of the wastewater matrix. Wastewater often contains multiple fluorophores and interferents that vary independently of the target parameter, e.g., BOD_5 (Carstea et al., 2016). In addition, high and fluctuating turbidity and suspended solids can attenuate excitation and emission light through inner-filter and scattering effects. Environmental factors such as temperature, pH, and dissolved constituents can also reduce fluorescence intensity and alter sensor response (Khamis et al., 2015; Goffin et al., 2020). During long-term in situ deployments, biofouling, scaling, and sensor drift may further degrade optical signals, even with automatic cleaning systems in place. This increases the need for regular validation and site-specific recalibration (Booth et al., 2023).

4.6 Influent Wastewater Quality (Sensor Results)

The previous section discussed the statistical analysis by comparison laboratory and sensor results (Day 1). Figure 10 shows a visual comparison by superimposing laboratory results with the sensor results. Figure 10a shows the 1-hour averaged sensor signals and Figure 10b shows the original 15-minute sensor signals of COD, BOD_5 and PO_4-P.

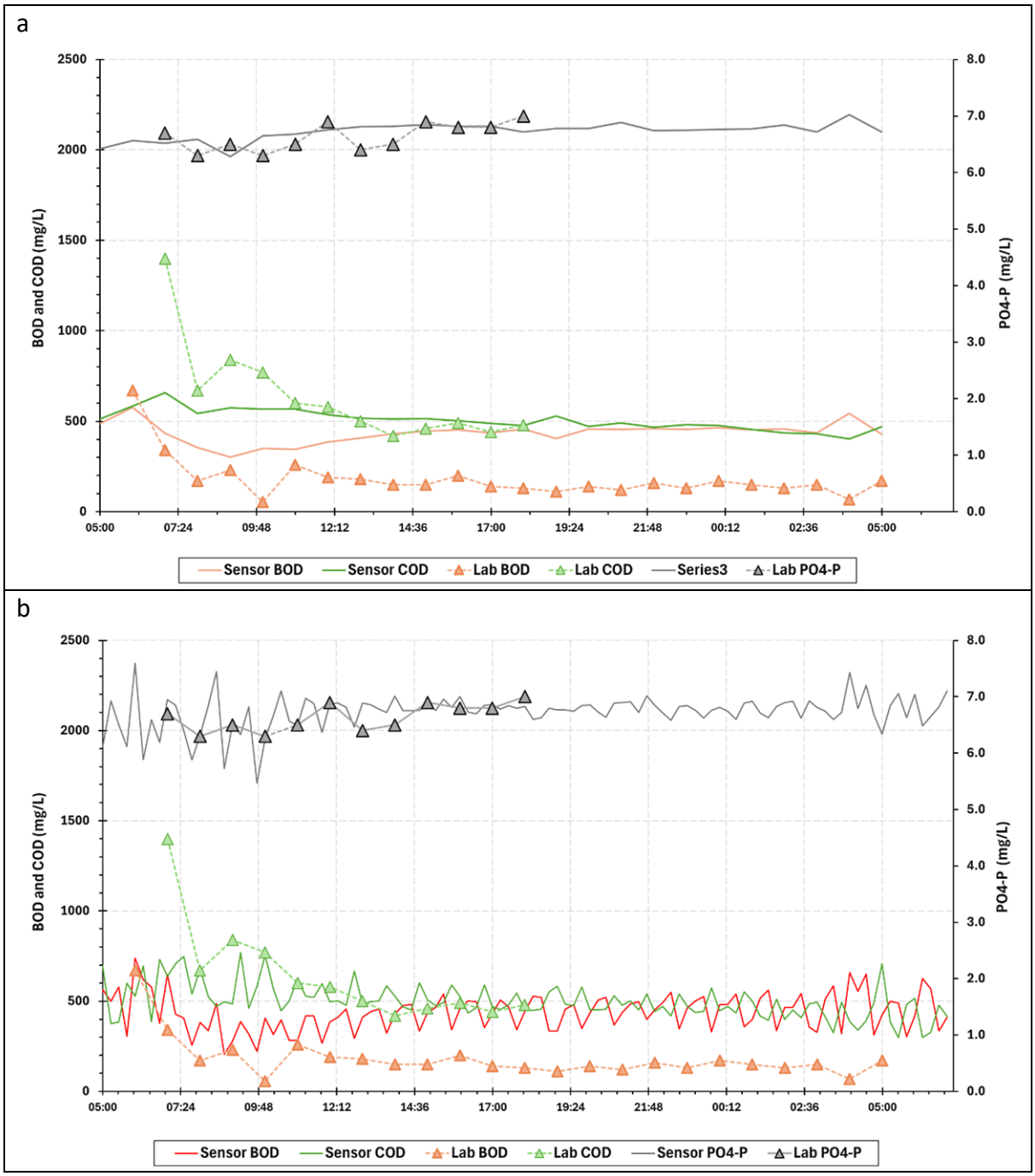


Figure 10: Comparison between sensor and laboratory trends of influent wastewater quality parameters, BOD_5 (mg/L), COD (mg/L), and PO_4-P (mg/L), a) 1-hour average sensor trends and b) 15-minute sensor trends.

As discussed in the previous section, PO_4-P exhibited the strongest correlation, followed by COD, while BOD_5 showed the weakest correlation for the 1-hour averaged sensor signal. When the original 15-minute sensor signal was used, greater temporal variability was observed; however, both PO_4-P and COD demonstrated improved visual agreement with laboratory results. BOD_5 laboratory values were inferred from trends observed in the sensor signal rather than direct correlation. For other sensor-measured parameters—turbidity, specific conductivity, and pH—the sensors measured turbidity and specific conductivity as proxies for suspended solids (SS) and total dissolved solids (TDS), respectively. Statistical comparisons of these parameters, along with pH measurements, were presented in the preceding section.

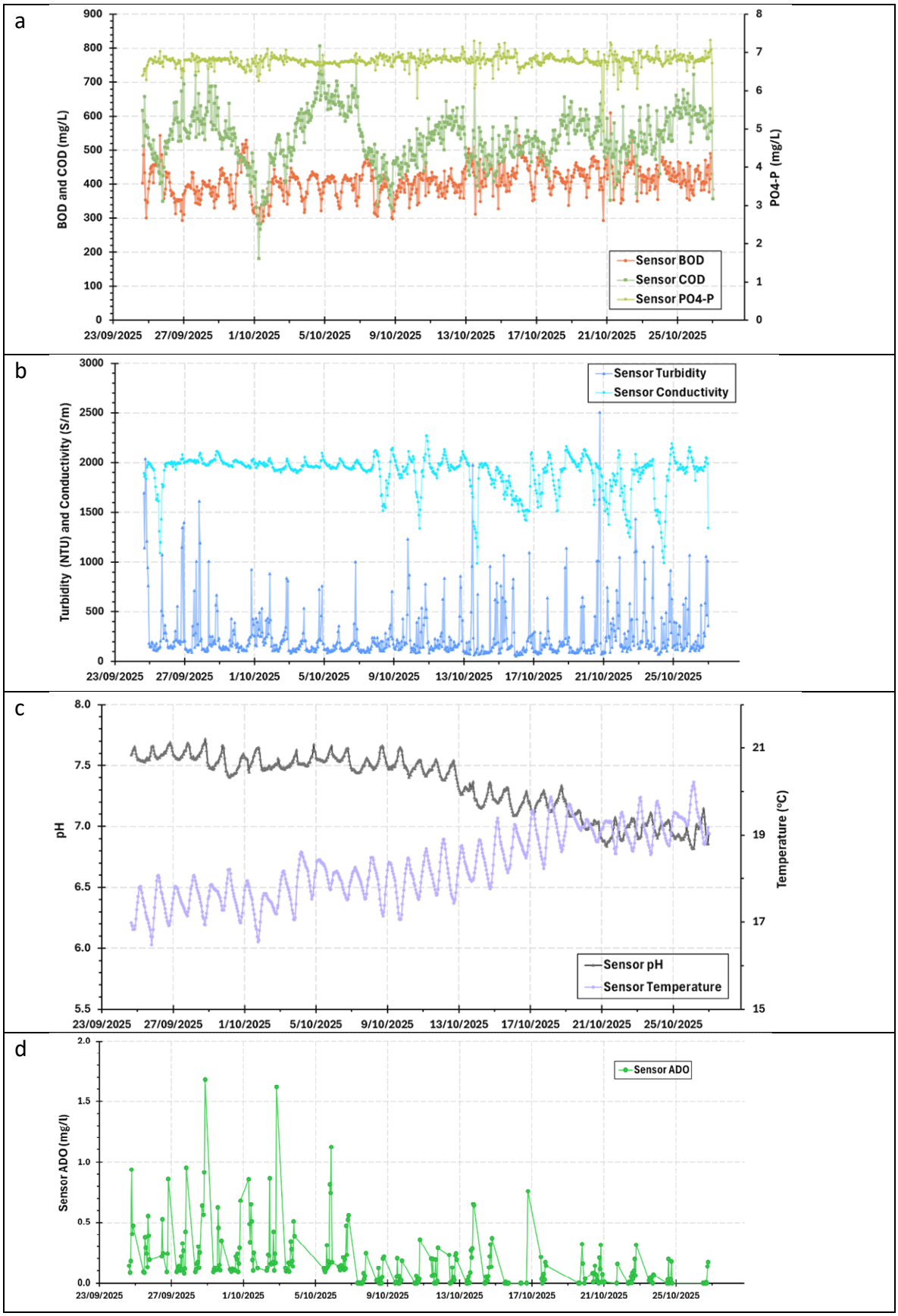


Figure 11: Trends of influent wastewater quality parameters, a) BOD (mg/L), COD (mg/L), and PO_4-P (mg/L), and b) turbidity (HU), conductivity (S/m), c) pH and temperature (°C), and d) ADO (mg/L), measured by online probes at the Goolwa WWTP from 25 September to 27 October 2025.

Figure 11 presents the trends of influent wastewater quality parameters measured continuously (converted into hourly averages) by the online sensor between 25 September and 27 October 2025 at the Goolwa WWTP. Figure 11a shows sensor-derived BOD, COD, and PO_4-P concentrations. Sensor-derived BOD and COD display the largest variability among all the parameters over the monitoring period, reflecting strong fluctuations in influent organic strength typical of small and regional wastewater catchments (Simmonds & Bristow, 2019; Sydney Water, 2021). Several periods of elevated COD are observed during the monitoring period. In contrast, BOD shows a narrower overall measurement range and a more stable temporal profile. While BOD also fluctuates over time, the pattern is characterised by frequent small-scale ups and downs rather than pronounced peaks. These reflect the dynamic nature of influent wastewater quality in the CWMS catchment. BOD measured by the sensor is within the range of 320-480 mg/L in general. COD measured by the sensor is within the range of 300-720 mg/L in general. Periods of elevated organic indicators suggest intermittent increases in organic loading, potentially associated with diurnal wastewater generation patterns or episodic trade waste inputs. In contrast, PO_4-P was almost stable (mostly within 6. 5-7 mg/L over the same period.

Figure 11b presents turbidity and conductivity trends measured by the online probes. Turbidity shows clear fluctuations and occasional peaks, indicating variability in suspended solids content and the occurrence of short-duration influent events that would be difficult to capture using conventional grab sampling (Carstea et al., 2016; Khan et al., 2022). Conductivity remains comparatively stable over time, mostly within the range of 1500 and 2100 S/m, suggesting that the dissolved ionic strength of the influent wastewater is relatively consistent despite variations in organic and particulate loads.

Figure 11c shows pH measured by the sensor, within 7-8, was decreasing overtime throughout the study period. Temperature followed a clear seasonal and diurnal pattern over the monitoring period, trended slightly increased overtime from 16 to 20 ^{o}C, reflecting ambient conditions and influent characteristics, and acts as an important co-varying parameter influencing both wastewater chemistry and sensor response (Khamis *et al*., 2015; Goffin *et al*., 2020).

Figure 11d illustrates the temporal trend in sensor ADO (mg/L), characterised by intermittent spikes and generally low dissolved oxygen levels over the monitoring period. ADO concentrations range from 0.01 to 1.68 mg/L. Dissolved oxygen levels below 1.0 mg/L are generally typical of domestic wastewater influent due to high biochemical oxygen demand. Occasional ADO values exceeding 1.0 mg/L may suggest industrial discharge (Yankovych et al., 2023). However, the ADO dataset exhibits substantially lower temporal coverage than other monitored parameters, which may reflect sensor or operational limitations rather than true process variability.

Overall, it highlights the ability of online sensors to capture high-resolution temporal variability in influent wastewater quality at a CWMS. While the accuracy of some sensor-derived parameters requires additional calibration and validation against laboratory measurements, the observed trends demonstrate the value of continuous monitoring for understanding influent trends, supporting operational awareness, and providing a foundation

for developing early-warning and decision-support tools for small and decentralised wastewater treatment systems.

4.7 Interrelationships among Wastewater Quality Parameters

Interrelationships among wastewater quality parameters were observed. Figure 12 illustrates key relationships observed for laboratory and sensor-measured parameters. Figure 12(a) shows a strong linear relationship between laboratory-measured chemical oxygen demand (COD) and suspended solids (SS), indicating that increases in particulate matter are associated with higher organic load in the influent wastewater, reflecting the substantial contribution of particulate organic matter to total COD in the influent wastewater. The relationship implies that increases in suspended solids reflect higher particulate organic content, which contributes to elevated COD, and that SS may therefore be used as an indirect indicator of organic strength under certain conditions. This finding was similar to other researcher's work (Ahnert *et al*., 2021, Shi *et al*., 2024). A conversion factor of 0.54 could be used to calculate the SS concentration based on the concentration of COD for monitoring the TSS content in the influent wastewater.

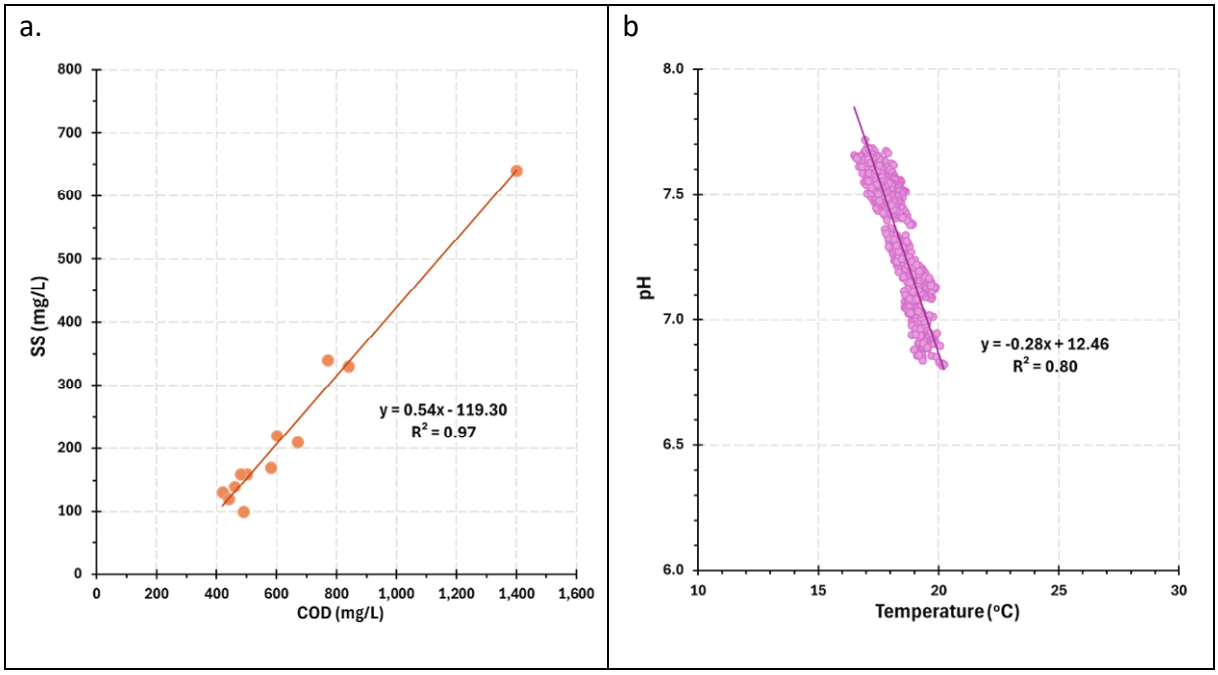


Figure 12: Observed linear relationships between (a) laboratory COD and SS, and (b) sensor temperature and pH.

Figure 12(b) presents the relationship between temperature and pH measured by online probes, showing a strong correlation in which pH varies systematically with temperature. This relationship reflects the influence of temperature on both water chemistry and measurement processes in wastewater systems (Khamis *et al.*, 2015; Goffin *et al.*, 2020). Changes in temperature can alter acid–base equilibria and reaction kinetics, affecting the dissociation of weak acids and bases and, consequently, the measured pH of the wastewater (Goffin *et al.*, 2020). In addition, biological processes such as microbial metabolism and ammonification are temperature dependent and may indirectly influence pH through changes in biochemical reaction rates (Booth *et al.*, 2023).

Temperature can also affect sensor response and accuracy, particularly for electrochemical pH probes. Although modern pH sensors apply automatic temperature compensation, residual temperature effects may remain due to limitations in compensation algorithms, sensor drift, or differences between calibration and field conditions. In wastewater applications, these effects may be exacerbated by high ionic strength, variable conductivity, and the presence of interfering substances.

The observed temperature–pH relationship therefore highlights the importance of considering temperature as a co-varying factor when interpreting online pH data. Failure to account for temperature effects may lead to misinterpretation of short-term pH fluctuations as changes in wastewater composition rather than temperature-driven variability. This finding underscores the need for appropriate temperature compensation (Khamis *et al.*, 2015; Booth *et al.*, 2023), regular calibration, and contextual interpretation of pH data in continuous online monitoring systems, particularly in small and decentralised wastewater treatment plants.

When inspecting the dissolved oxygen concentration readings recorded by the sensor, the sensor recorded quite a lot of blank readings, after consulted the manufacturer, it was likely to be signal overloaded, this could mean during the trialling period, there were external factor affected the dissolved oxygen sensor. Figure 13 shows the flow and the dissolved oxygen sensor readings, apparently when flat lines indicated there were no readings recorded by the dissolved oxygen sensor, those flat lines coincided with the during the high flow periods. Potentially during high flow periods, the sensor could be affected, further investigating later in the report.

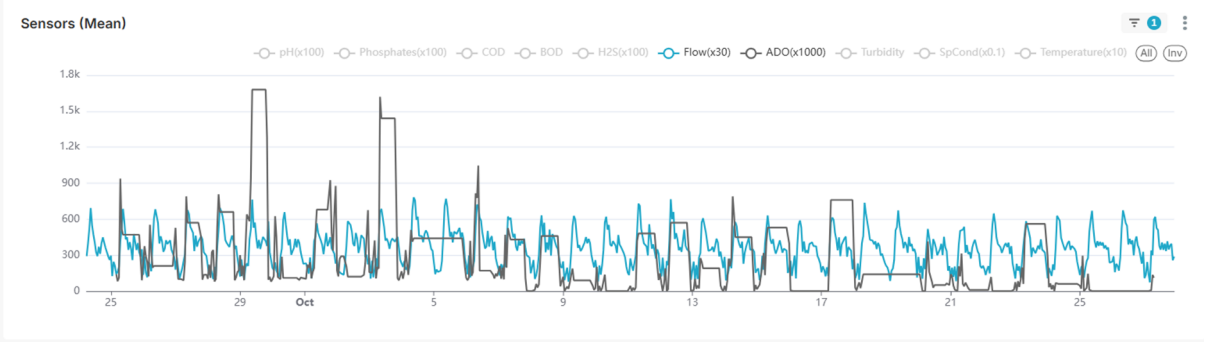


Figure 13: Trends of influent wastewater flow and dissolved oxygen concentration measured by online probes at the Goolwa WWTP from 25 September to 27 October 2025.

Figure 14 shows the COD, BOD_5, flow with the School Holidays and Non-School Holidays period marked, there are two distinctive high COD periods in the School Holidays period compared to the Non-School Holidays period.

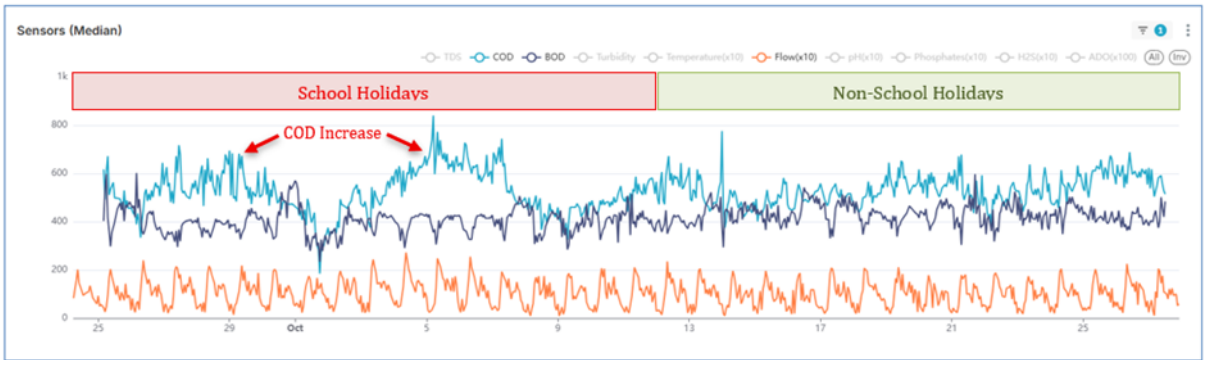
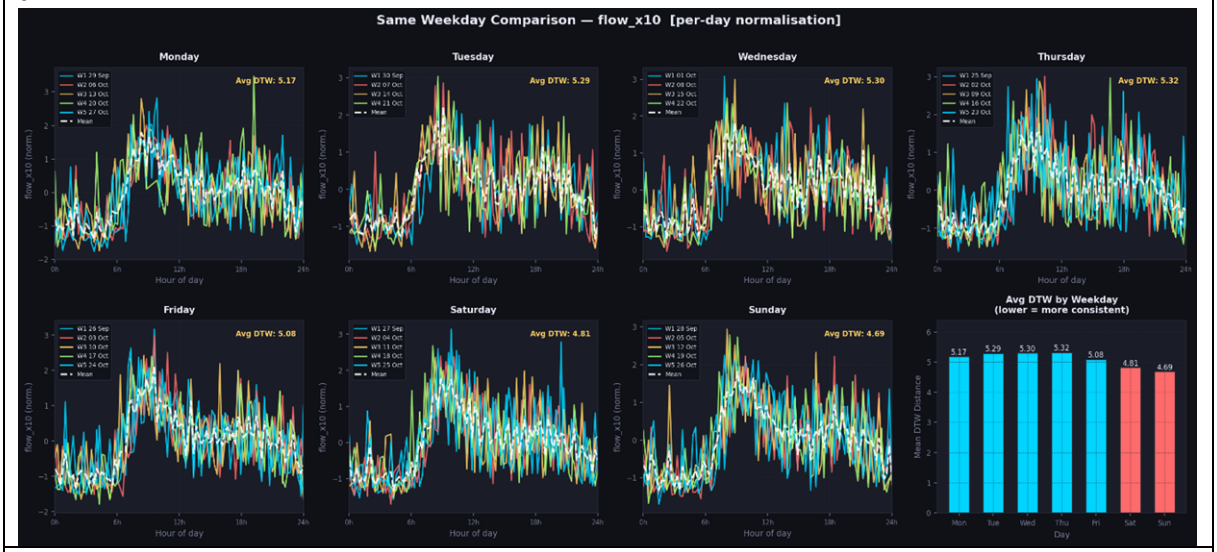


Figure 14: Trends of influent wastewater flow, COD and BOD_5 measured by online probes at the Goolwa WWTP from 25 September to 27 October 2025 with the School Holidays and non-School Holidays labelled.

4.8 Pattern Analysis Using Advanced Computing Techniques

The section uses Dynamic Time Warping (DTW) to analyse sensor signals in a daily basis to understand whether these is any regular daily patterns, particularly, the difference between weekdays and weekends using the flow and COD data. DTW is a well-known algorithm for measuring the similarity between two temporal sequences which may vary in speed (Müller, 2007). Unlike standard Euclidean distance, which compares points at fixed time intervals, DTW finds an optimal alignment between two sequences by "warping" the time axis.

a



b



DTW < 10 = high similarity

Figure 15: Daily pattern comparison in DTW (Group by individual days): a. Flow; b. COD reading.

This makes it exceptionally suitable for comparing time-series data against a reference signal, particularly when the timing of events is inconsistent. DTW is highly effective when the pattern matters more than the duration (e.g. monitoring a machine cycle where a specific mechanical signature occurs, but the cycle speed varies based on load.)

Two comparisons were performed based on each day of the flow data.

1. 15 mins average flow reading, constituting a day pattern
2. 30 mins average flow reading, constituting a day pattern

The analysis results are presented in DTW Heatmaps and hieratical clusters (Figure 15). In DTW Distance Heatmaps, a square matrix where every row and column is one day in the dataset (e.g. "06 Jan Mon", "07 Jan Tue", ...). Each cell's colour indicates DTW distance between those two days, where

- Cyan (low end of colormap) = very similar daily profiles
- Black (mid) = moderate difference
- Coral (high end) = very different profiles

The diagonal is always zero (a day compared to itself). The matrix is reordered by the clustering algorithm (similar days are grouped together).

Furthermore, the hierarchical clustering of DTW distances is visualised in dendrograms to showing strongest repeating patterns across days (Figure 16). In the tree-like diagrams, a tight cluster of many days merging at a short distance indicates a stable recurring pattern, while a day that only joins the tree very late suggests an anomaly.

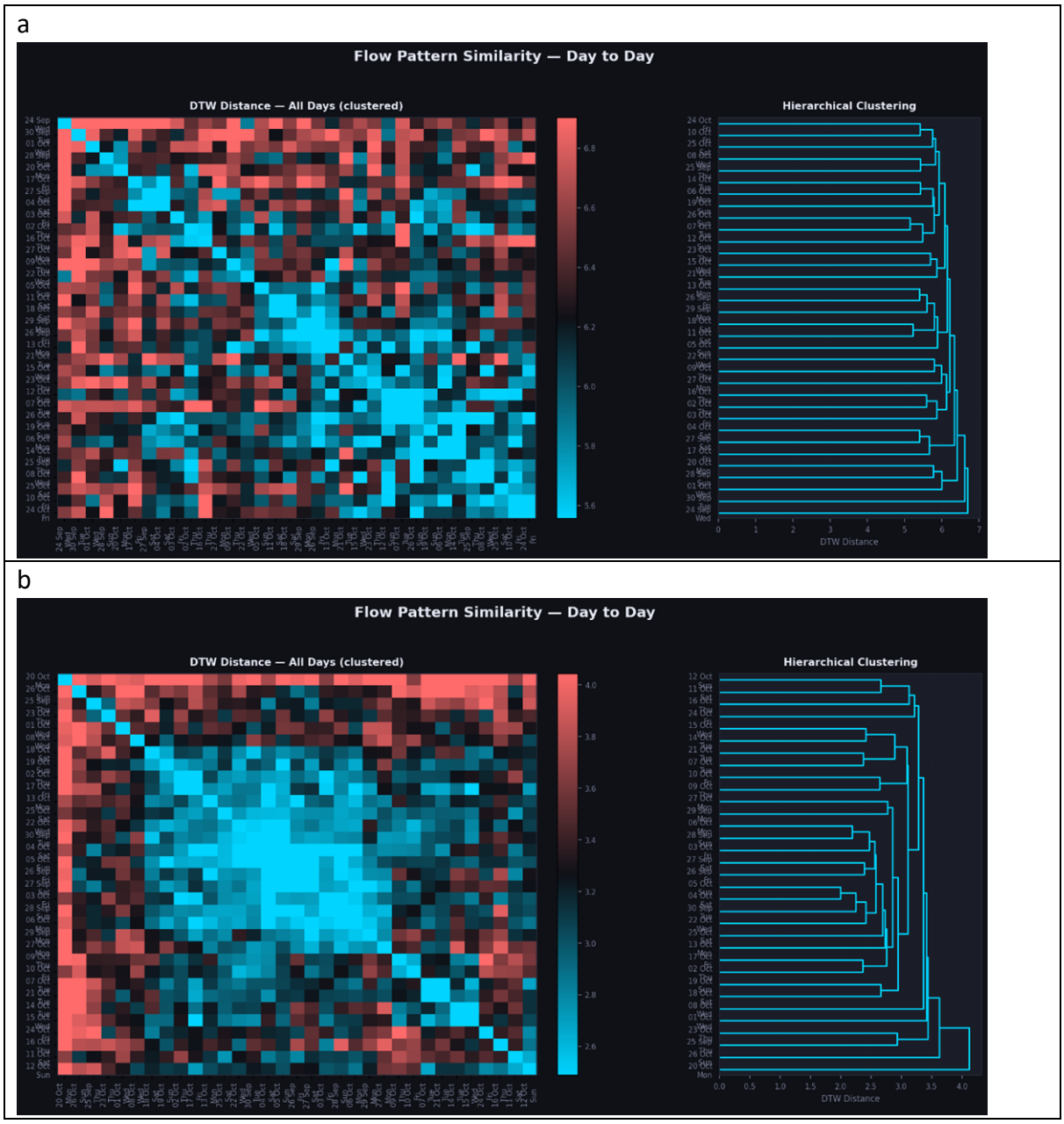


Figure 16: Daily flow pattern comparison in DTW (Grouped by weekdays and weekends): a. 15 mins interval; b. 30 mins interval.

Both the 15 minute and the 30 minutes interval day patterns in Figure 16 indicate strong repeating patterns for the weekdays and weekends as observed from the cyan dots in the heatmap. The DTW clustering analysis further verified this observation. For example, in the 15 minutes DTW clustering analysis, the top diagram shows a strong recurring pattern for Oct 10 (Friday) and Oct 24 (Friday) but is significantly distinct to Oct 25 (Sat). Similarly, in the 30 minutes DTW clustering analysis, the top diagram shows a strong recurring pattern for Oct 11 (Sat) and Oct 12 (Sun) but is significantly distinct to Thursday Oct 16 and Friday Oct 24.

Following from the above, 15 minutes interval was used for the rest of the report. DTW distance was also calculated for the COD readings (Figure 17). The results are shown below.

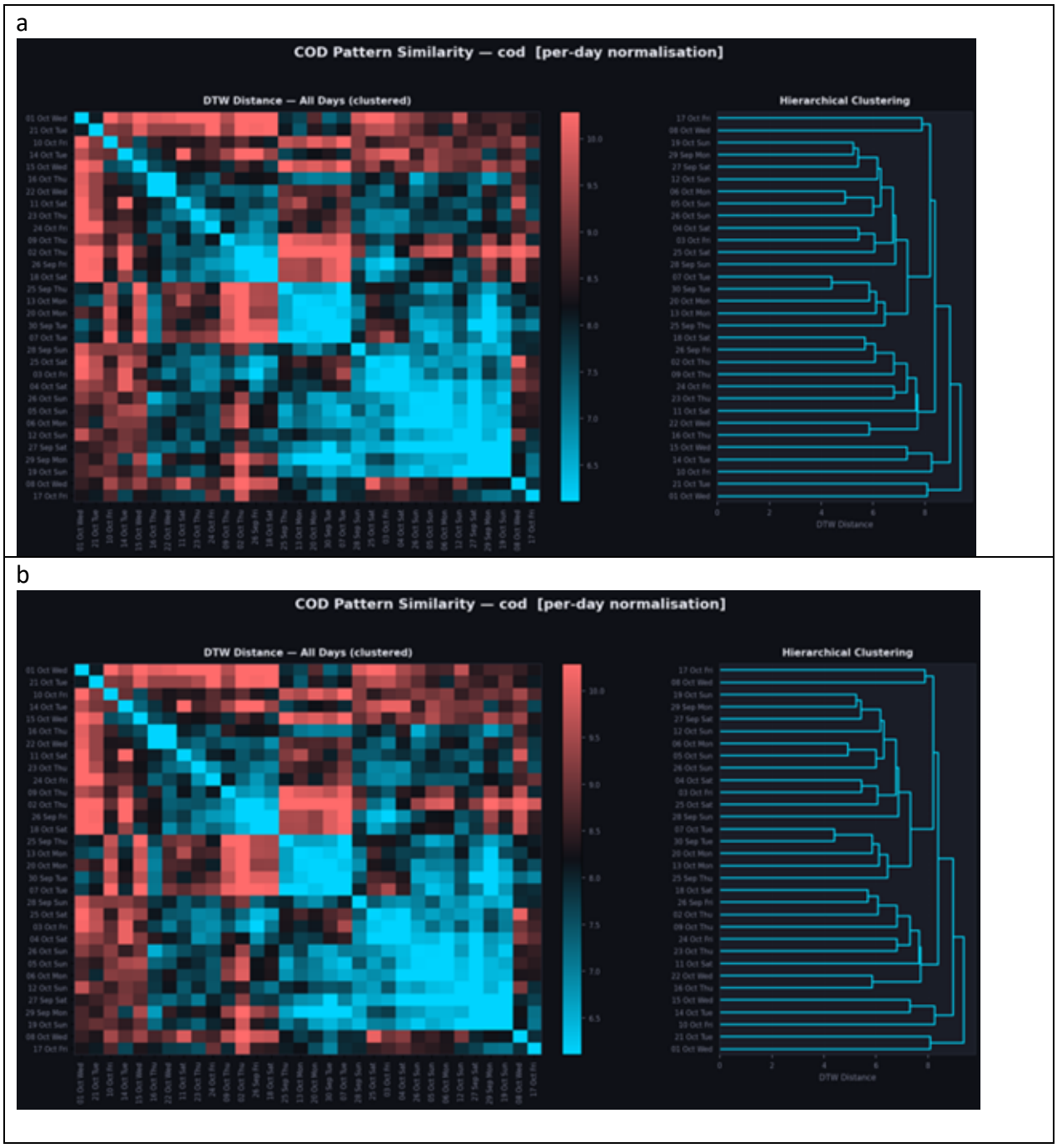


Figure 17: Daily COD pattern comparison in DTW (Grouped by weekdays and weekends): a. 15 mins interval; b. 30 mins interval.

Examination of the COD DTW results did not reveal clear weekday–weekend patterns, or any distinct recurring patterns, in contrast to the flow DTW results. Flow data primarily reflect wastewater discharge timing and community usage patterns, whereas COD represents the composition of wastewater within the sewer network. As such, COD variability may reflect inputs from different sources, such as domestic and industrial discharges. This finding suggests that the catchment includes multiple contributing sources, resulting in less consistent and non-uniform COD patterns.

By using a line chart as shown below (Figure 18), it suggests that the flow daily patterns (weekdays vs weekends) are more obvious than COD daily patterns (weekdays vs weekends).

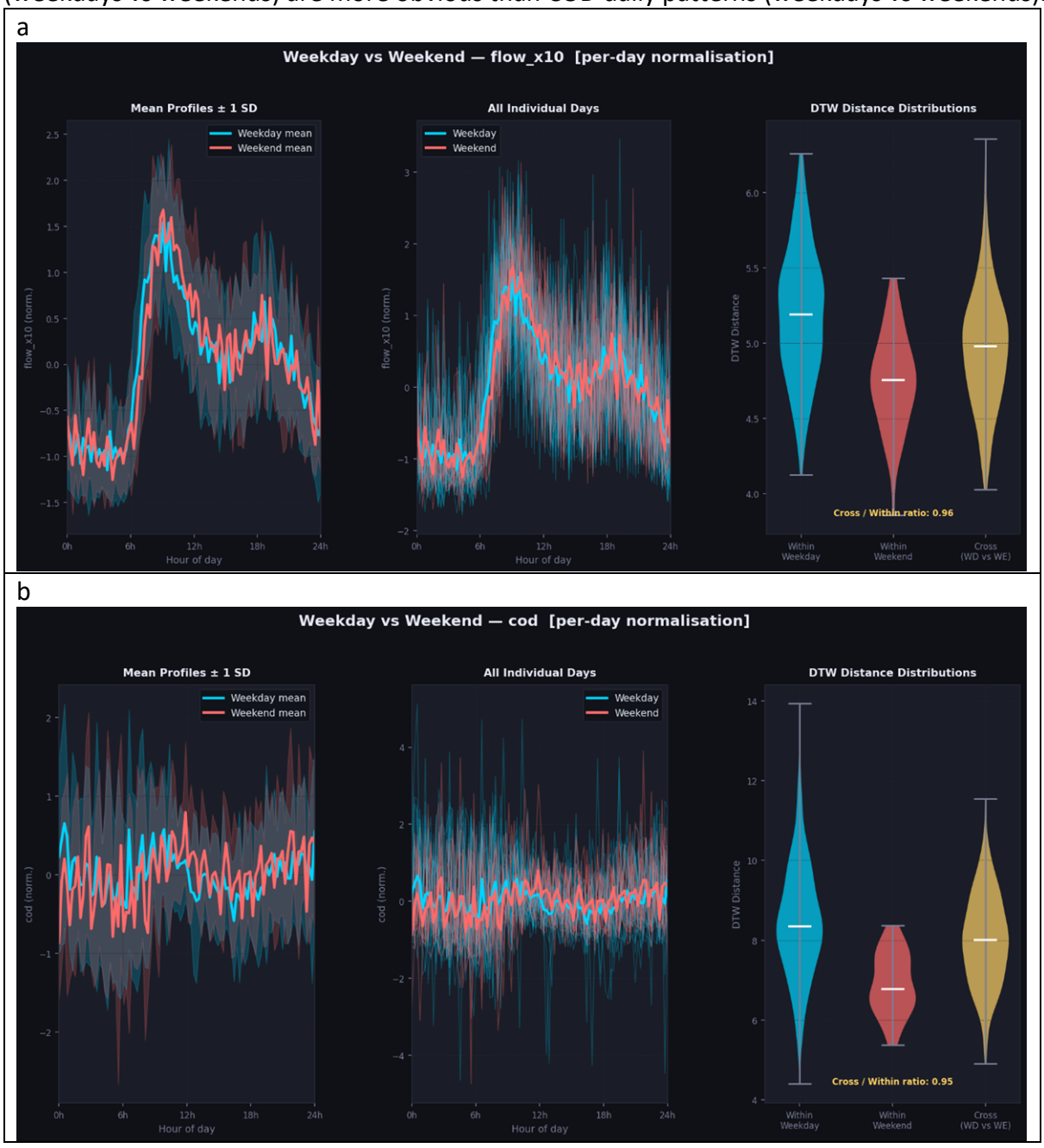


Figure 18: Daily COD pattern comparison in DTW (Grouped by individual days, 30 mins interval): a. Flow; b. COD.

The DTW results of all the other parameters are in Appendix 3.

4.9 Warning Alarm Development Using Machine Learning

When building a time-series prediction model using artificial neural networks (ANNs), the Long Short-Term Memory (LSTM) network, a specialised Recurrent Neural Networks (RNNs), is an effective choice because it addresses the core machine learning challenge of temporal dependencies. Time-series data isn't just a collection of independent variables; it's a sequence where the order is everything. Traditional deep learning architectures often struggle with the

vanishing gradient problem, where the signal from earlier time steps peters out during backpropagation through time (BPTT). LSTMs solve this by using a gated cell architecture, consisting of forget, input, and output gates, which allows the model to learn which features are worth keeping in the long-term memory and which are just redundant noise. In this project, this means the network can effectively capture complex non-linear relationships and seasonal trends, ensuring the gradient flow remains stable enough to learn from patterns that happened weeks or even months ago.

The LTSM results on flow data and COD prediction model are presented together with the baseline Autoregressive (AR) models (linear regression) as shown below (Figure 19 and Table 4).

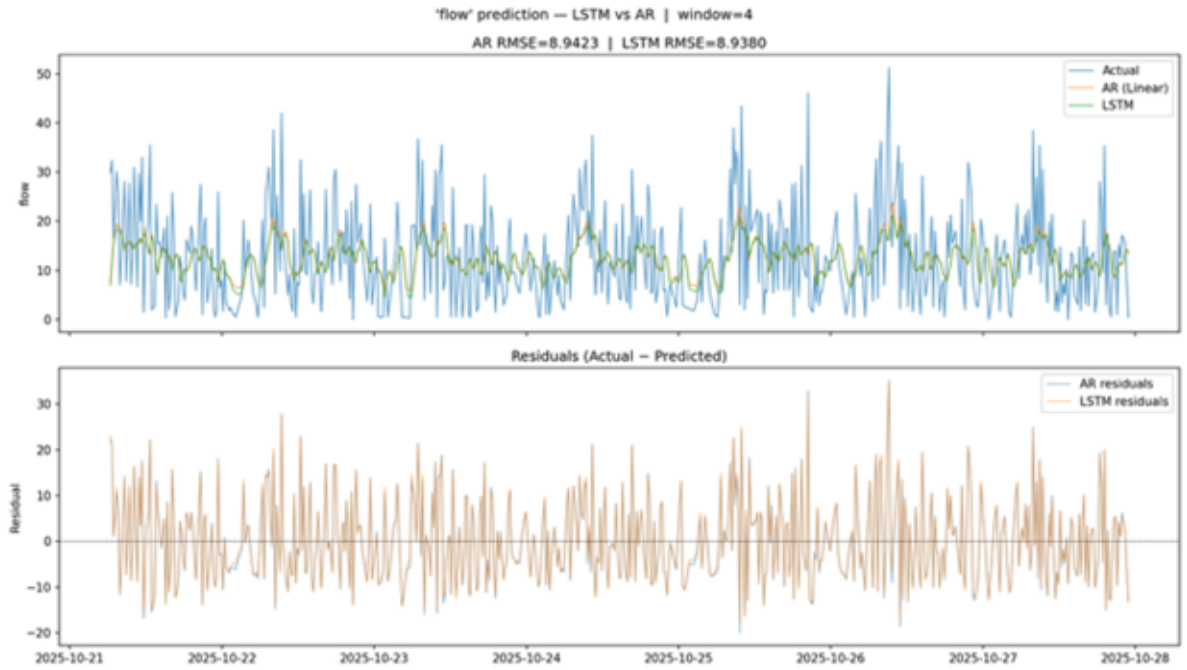


Figure 19: Flow AR vs LTSM

Table 4: Flow prediction model comparison – Lower value is better

AR - Linear Regression	**LTSM**
RMSE : 8.9423	RMSE : 8.9380
MAE: 7.2375	MAE : 7.2153
R^2: 0.0937	R^2: 0.0945

Winner by Root Mean Square Error (RMSE): LSTM (0.0% better)

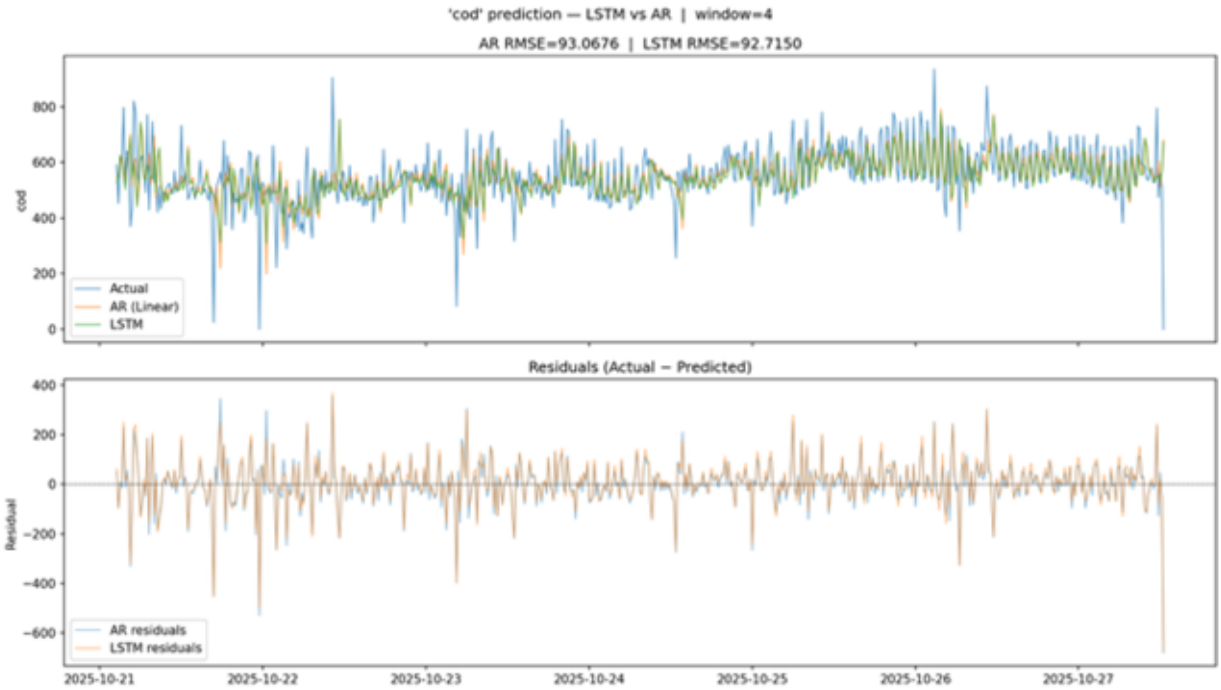


Figure 20: COD AR vs LTSM

Table 5: COD prediction model comparison – Lower value is better

AR - Linear Regression	**LTSM**
RMSE : 93.0676	RMSE : 92.7150
MAE: 61.6650	MAE : 61.8441
R^2: 0.1674	R^2: 0.1737

Winner by RMSE: LSTM (0.4% better)

The results (Figure 20 and Table 5) suggest that LTSM does not present any advantages in flow prediction. However, LTSM shows a slight advantage in COD prediction. When a LSTM network outperforms Linear Regression (LR), it is usually because the underlying data contains complexities that a linear, "memoryless" model cannot capture. Comparing to flow data readings, COD readings may affect by various latent variables (factors that influence the system but aren't explicitly measured in the dataset). For example, in this project, bubbles may affect the COD readings. While an LSTM cannot "magically" see data that isn't there, it is exceptionally good at capturing the proxies and patterns those missing variables leave behind in the temporal sequences.

5 LESSON LEARNED

5.1 Sensor Setup, Calibration and Maintenance

Sensor and Laboratory results comparison, we generally assume laboratory results are the true results and as the reference point. However, if there are some sampling issues to affect laboratory results, this will be a difficult situation to confirm sensor results, in this case, the recommended calibration used two separate laboratories to ensure no discrepancy between the two laboratories, unfortunately due to autosampler issue, not able to collect sufficient volume of sample to conduct the laboratory analyses and divide the samples to two groups to use two commercial laboratories. This may affect the comparison between sensor and laboratory results. There is another contributing factor that the sensor was position slightly apart from autosampler sampling intake. The project team done the best we can to ensure the samples collected are the same as the presented to the sensor. This issue can be managed if the equipment is permanently installed.

5.2 Data Processing Software Comparison

Excel is the most used universal data processing package with graphical display function. However, to handle the amount of data generated from an online instruction can often exceed the Excel storage capacity, the graphical function can struggle to correctly display the graph even before the cells filled up.

In this case that different data sets corrected data in different time intervals, the time axis is required to be lined with those recorded parameters. Excel has provided functionality to complete the task, however, quite a lot of manual processing is needed.

The instrument portal provided by the Intelligent Sensing Technologies is a real-time display portal for the Proteus sensors (Appendix 2). It provides all the necessary functionalities to view and display sensor signals. The zoom in and zoom out functions allow users to closer inspect a particular region of interest. For this work which requires incorporating with different databases, including flow. The custom-build portal is serving the purpose for this project with all the functionalities for data processing and comparison.

As mentioned earlier that Excel can perform data averaging to line up sensor measurement time, in addition, Excel was used to investigate different averaging times, such as averaged using 15 minutes versus 1 hour, impact on data interpretation. Excel has provided the result, however, when we used the custom-build portal, this comparison can be completed in much shorter time.

6 SUMMARY AND CONCLUSIONS

This case study represents the first trial of a fluorescence sensor for a CWMS in South Australia. The project enabled the team to gain in-house experience in deploying sensors with real-time monitoring capabilities to support wastewater treatment operations, as well as to appreciate the volume and complexity of data generated by continuous monitoring systems.

Under the first objective, evaluation of the suitability and usefulness of fluorescence sensors for real-time monitoring showed that the accuracy of some sensor-derived parameters was lower than expected when compared with laboratory measurements. Nevertheless, the observed trends clearly demonstrate the value of continuous monitoring in understanding influent behaviour and enhancing operational awareness.

The second objective—applying machine-learning algorithms to develop artificial intelligence capability for operational support—demonstrated that several existing algorithms available in the public domain can be readily adapted to this application. This provides a solid foundation for the development of early-warning and decision-support tools for small and decentralised wastewater treatment systems. The application of DTW for analysing sensor signals demonstrated its effectiveness in identifying similarities through pattern analysis and in visualising comparisons between different parameters. In addition, the use of time-series prediction models based on advanced computational technique, LSTM networks—showed promising results, indicating strong potential for further development.

Finally, the custom-build data portal proved effective for data visualisation and analysis, offering significantly greater flexibility and functionality than Excel for managing and interpreting real-time monitoring data.

8 REFERENCES

Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) and Australian and New Zealand Environment and Conservation Council (ANZECC), 1997. *Australian guidelines for sewerage systems: effluent management*. National Water Quality Management Strategy (NWQMS) Paper 11. Canberra: Commonwealth of Australia. Available at: https://www.waterquality.gov.au/sites/default/files/documents/effluent-management.pdf (Accessed: 22 March 2026).

Ahnert, M.; Schalk, T.; Brückner, H.; Effenberger, J.; Kuehn, V.; Krebs, P. Organic matter parameters in WWTP–a critical review and recommendations for application in activated sludge modelling. Water Science and Technology. 2021, 84, 2093–2112.

Augsburger, N., Zaouri, N., Cheng, H. and Hong, P.Y., 2021. The use of UV/H2O2 to facilitate removal of emerging contaminants in anaerobic membrane bioreactor effluents. Environmental research, 198, p.110479.

Booth, A. M., Fleck, J. A., Pellerin, B. A., Hansen, A. M., Etheridge, A. B., Foster, G. M., Graham, J. L., Bergamaschi, B. A., Carpenter, K. D., Downing, B. D., Rounds, S. A. and Saraceno, J. F., 2023. Field techniques for fluorescence measurements targeting dissolved organic matter, hydrocarbons, and wastewater in environmental waters: principles and guidelines for instrument selection, operation and maintenance, quality assurance, and data reporting. U.S. Geological Survey Techniques and Methods, Book 1, Chap. D11. https://doi.org/10.3133/tm1d11.

Booth, A.M., Fleck, J.A., Pellerin, B.A., Hansen, A.M., Etheridge, A.B., Foster, G.M., Graham, J.L., Bergamaschi, B.A., Carpenter, K.D., Downing, B.D., Rounds, S.A. and Saraceno, J.F., 2023. Field techniques for fluorescence measurements targeting dissolved organic matter, hydrocarbons, and wastewater in environmental waters: principles and guidelines for instrument selection, operation and maintenance, quality assurance, and data reporting. *U.S. Geological Survey Techniques and Methods*, Book 1, Chap. D11. https://doi.org/10.3133/tm1d11.

Carstea, E.M., Bridgeman, J., Baker, A. and Reynolds, D.M., 2016. Fluorescence spectroscopy for wastewater monitoring: a review. *Water Research*, 95, pp.205–219. https://doi.org/10.1016/j.watres.2016.03.021.

Do, P., Chow, C.W.K., Rameezdeen, R., Gorjian, N., 2022. Wastewater inflow time series forecasting at low temporal resolution using SARIMA model: a case study in South Australia, Environmental Science and Pollution Research, 29, p.70984–70999. https://doi.org/10.1007/s11356-022-20777-y.

Fazzino, F., Roccaro, P., Di Bella, A., Gagliano, E., Vagliasindi, F.G. and Korshin, G.V., 2025. Use of fluorescence for real-time monitoring of contaminants of emerging concern in (waste) water: Perspectives for sensors implementation and process control. Journal of Environmental Chemical Engineering, 13(2), p.115916.

Goffin, A., Vasquez-Vergara, L.A., Guérin-Rechdaoui, S., Rocher, V. and Varrault, G., 2020. Temperature, turbidity, and the inner filter effect correction methodology for analyzing fluorescent dissolved organic matter in urban sewage. *Environmental Science and Pollution Research*, 27, pp.35712–35723. https://doi.org/10.1007/s11356-020-09889-5.

Guillossou, R., Le Roux, J., Goffin, A., Mailler, R., Varrault, G., Vulliet, E., Morlay, C., Nauleau, F., Guérin, S., Rocher, V. and Gasperi, J., 2021. Fluorescence excitation/emission matrices as a tool to monitor the removal of organic micropollutants from wastewater effluents by adsorption onto activated carbon. Water Research, 190, p.116749.

Harvey, D, 2016. Analytical Chemistry 2.1 – Linear Regression and Calibration Curves. Chemistry LibreTexts.

Khamis, K., Bradley, C., Gunter, H.J., Basevi, G., Stevens, R. and Hannah, D.M., 2021. Calibration of an in-situ fluorescence-based sensor platform for reliable BOD5 measurement in wastewater. Water Science and Technology, 83(12), pp.3075-3091.

Khamis, K., Sorensen, J.P.R., Bradley, C., Hannah, D.M., Lapworth, D.J. and Stevens, R., 2015. In situ tryptophan-like fluorometers: assessing turbidity and temperature effects for freshwater applications. *Environmental Science: Processes & Impacts*, 17(4), pp.740–752. https://doi.org/10.1039/C5EM00030K.

Khan, M.F.S., Akbar, M., Wu, J. and Xu, Z., 2022. A review on fluorescence spectroscopic analysis of water and wastewater. Methods and applications in fluorescence, 10(1), p.012001.

LGA 2023. State Infrastructure Strategy – Discussion Paper. Local Government Association (LGA) of South Australia.

Marino, L., Todesco, R., Gagliano, E., Santoro, D. and Roccaro, P., 2025. Real-time wastewater quality monitoring by fluorescence sensors: Validation for COD and CEC monitoring and implication for carbon footprint reduction. Science of The Total Environment, 963, p.178464.

Mladenov, N., Sanfilippo, S., Panduro, L., Pascua, C., Arteaga, A. and Pietruschka, B., 2024. Tracking performance and disturbance in decentralized wastewater treatment systems with fluorescence spectroscopy. Environmental Science: Water Research & Technology, 10(6), pp.1506-1516.

Müller, Meinard. (2007). Dynamic time warping. Information Retrieval for Music and Motion. 2. 69-84. 10.1007/978-3-540-74048-3_4.

Nurhayati, M., You, Y., Park, J., Lee, B.J., Kang, H.G. and Lee, S., 2023. Artificial neural network implementation for dissolved organic carbon quantification using fluorescence intensity as a predictor in wastewater treatment plants. Chemosphere, 335, p.139032.

Samuelsson, O., Lindblom, E. U., Björk, A., & Carlsson, B., 2023. To calibrate or not to calibrate, that is the question. Water Research, 229, 119338.

Sgroi, M., Roccaro, P., Korshin, G.V., Greco, V., Sciuto, S., Anumol, T., Snyder, S.A. and Vagliasindi, F.G., 2017. Use of fluorescence EEM to monitor the removal of emerging contaminants in full scale wastewater treatment plants. Journal of hazardous materials, 323, pp.367-376.

Shi, Z., Chow, C.W., Gao, J., Xing, K., Liu, J. and Li, J., 2024. Using Surrogate Parameters to Enhance Monitoring of Community Wastewater Management System Performance for Sustainable Operations. Sensors, 24(6), p.1857.

Simmonds & Bristow Pty Ltd, 2019. *Characterisation of raw sewage and design implications for household sewage treatment systems: for the review of AS1546.3:2017*. Technical report C1911-007 prepared for the National Onsite Providers Association (NOPA), December. Available at: http://www.nopa.org.au/wp-content/uploads/2020/12/Characterisation-of-Raw-Sewage.pdf (Accessed: 22 March 2026).

Sydney Water, 2021. *Wastewater treatment planning guidelines* (D0001891), Version 1.0. Sydney: Sydney Water. Available at: https://www.sydneywater.com.au/content/dam/sydneywater/documents/provider-information/standards-specifications/wastewater-treatment-planning-guidelines.pdf (Accessed: 22 March 2026).

Xin, C., Cheng, Z., You, Z., Bai, H. and Wang, J., 2020. Using EEM fluorescence to characterize the membrane integrity of membrane bioreactor (MBR). Journal of Membrane Science, 610, p.118356.

Yang, Y., Shan, C. and Pan, B., 2024. Machine learning modeling of fluorescence spectral data for prediction of trace organic contaminant removal during UV/H2O2 treatment of wastewater. Water Research, 255, p.121484.

Yankovych, H., Vaclavikova, M. and Melnyk, I., 2023. A review on adsorbable organic halogens treatment technologies: approaches and application. Sustainability, 15(12), p.9601.

Zhang, Q.; Xie, Y.; Tian, Y.; Dzakpasu, M.; Wang, X.C. Acceptable risk assessment and management of environmental pollution emergency events base on distance model. Sci. Total Environ. 2023, 866, 161298

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., & Subramanian, R. (2017). A machine learning calibration model to improve low-cost sensor performance. Atmospheric Measurement Techniques (AMT) Preprint, 1–32.

APPENDIX I - PROTEUS SENSOR INFORMATION

Sensor Specifications (Data Sheet)

	Units	Range	Resolution	Accuracy	Comments
BOD	mg/L	0-2000 mg/L	0.01 mg/L	±5 % of reading*	Local site calibration can improve accuracy.
COD	mg/L	0-4000 mg/L	0.01 mg/L	±5 % of reading*	Local site calibration can improve accuracy.
TDS (Conductivity)	g/L	0 to 65 g/L	0.1 g/L	±5 % of reading	Calculated from specific conductance.
TSS (Turbidity)	mg/L	0 to 500 mg/L	4 digits with maximum of two decimals	±2 % of reading	Calculated using the correlation between turbidity and a sediment standard or sample. Local calibration can be applied.
pH	pH unit	0 to 14	0.01	±0.1 within 10°C of calibration, 0.2°C otherwise	Refillable reference electrode; corrected for temperature; typical sensor life > 4 years.
Dissolved Oxygen	mg/L	0-20 mg/L 20-30 mg/L 30-50 mg/L	0.10 mg/L 0.15 mg/L 0.1 mg/L	±0.1 ±0.15 ±5 of reading	Compensated for temperature and salinity; EPA approved “lifetime” luminescence method; typical sensor cap life > 4 years.
Phosphate (PO_4^+)	mg/L	0-100 mg/L	0.01 mg/L	±0.1 mg/L 0 -2 mg/L ±0.5 mg/L 2 -100 mg/L	Environmental/river applications typically 0 - 1mg/L Wastewater applications typically 0 - 50 mg/L.
Temperature	Water Temperature	-5 to 50°C	0.01	±0.1	Never needs calibration.

Proteus Sensor calibration: SOP for samples for providing a site calibration

It is vital that samples are taken under carefully controlled conditions, as they are the key to an accurate and robust site calibration. Please follow the guidelines below:

1. It is essential that the samples you take capture the full range of readings you would realistically expect to occur in the location you are planning to monitor.
2. To achieve this, we would recommend a **minimum** of 15 paired samples which span the entire range of expected/normal reasons.

    For example, if you are monitoring coliforms where the day-to-day average is 1-3 cfu/100 ml, but during storms this increases to 100+ cfu/100 ml, you should aim start sampling before a forecasted storm and continue to sample through the rising limb to the point at which coliforms peak. You can, of course, sample beyond this but as a minimum, the entire range must be covered.

3. Paired samples require two separate samples taken simultaneously; this can be achieved by either taking two separate samples together or by splitting a single, larger sample. The samples are then to be sent off for analysis either at the same or separate laboratories. • Ensure the laboratories are accredited in the particular test being carried out (e.g. BOD5).
4. The timings of the samples **must** be recorded and **must** coincide with the time the Proteus is also taking readings – **deviation from this will produce an imprecise calibration.**

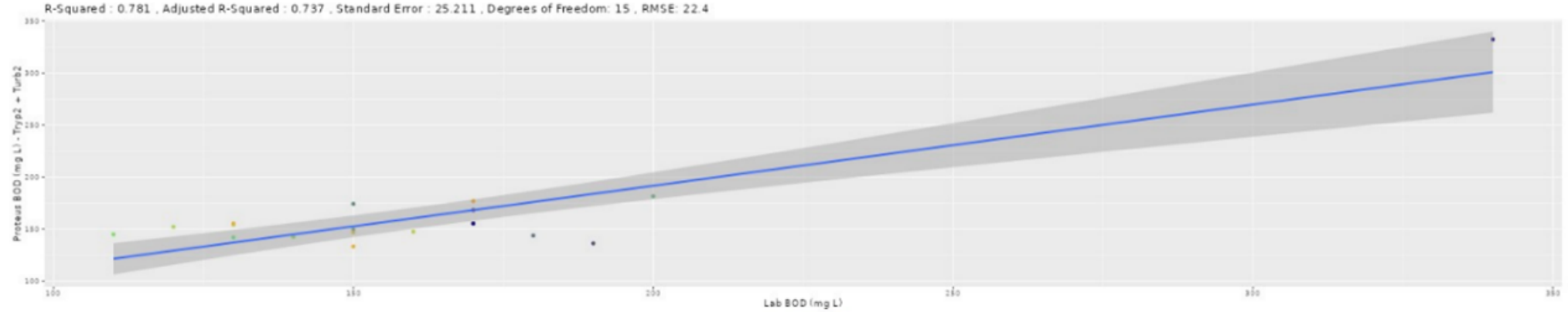
For example: Date	**Time**	**Lab 1 BOD**	**Lab 2 BOD**	**Proteus BOD**
1/1/2020	12:00	25.3	28.2	26.00
1/1/2020	13:00	23.6	28.3	25.32

5. In addition, the samples must be taken at exactly the same point and depth as the Proteus instrument – we recommend physically attaching the sampler end to the Proteus to aid this.
6. We recommend taking 1l samples as this protects against loss of samples during the laboratory testing.

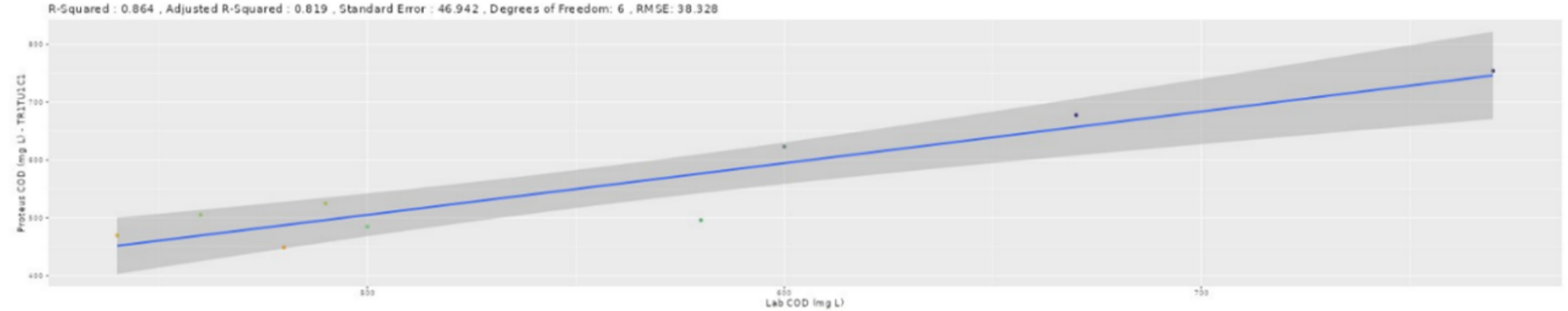
Additional Good Practice:

1. Always use the bottles supplied by the laboratory to avoid contamination of samples; also ensure the hose and/or other equipment used to collect the sample is thoroughly cleaned prior to sampling using acid washes or similar.
2. Seasonal trends are easily capturable by recalibrating throughout the year as part of maintenance routines.
3. When providing a calibration for BOD, the sample procedure should follow the protocol outlined by ISO 5815-2:2003. Samples should be kept both refrigerated and in the dark from the moment of collection to the beginning of laboratory testing. Use cool boxes with ice blocks to ensure this requirement is met. If samples for BOD cannot be analysed within 4 hours, they need to be refrigerated and analysed within 24 hours.

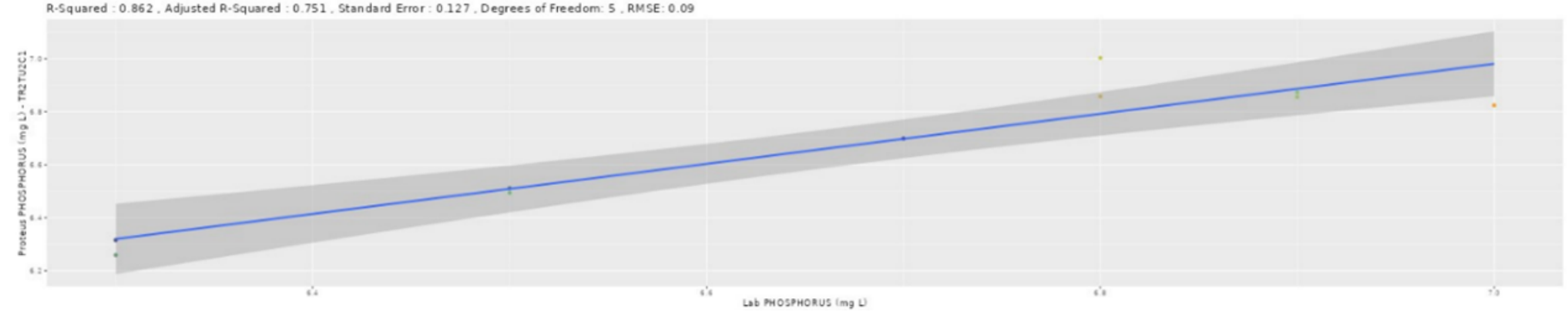
BOD Calibration



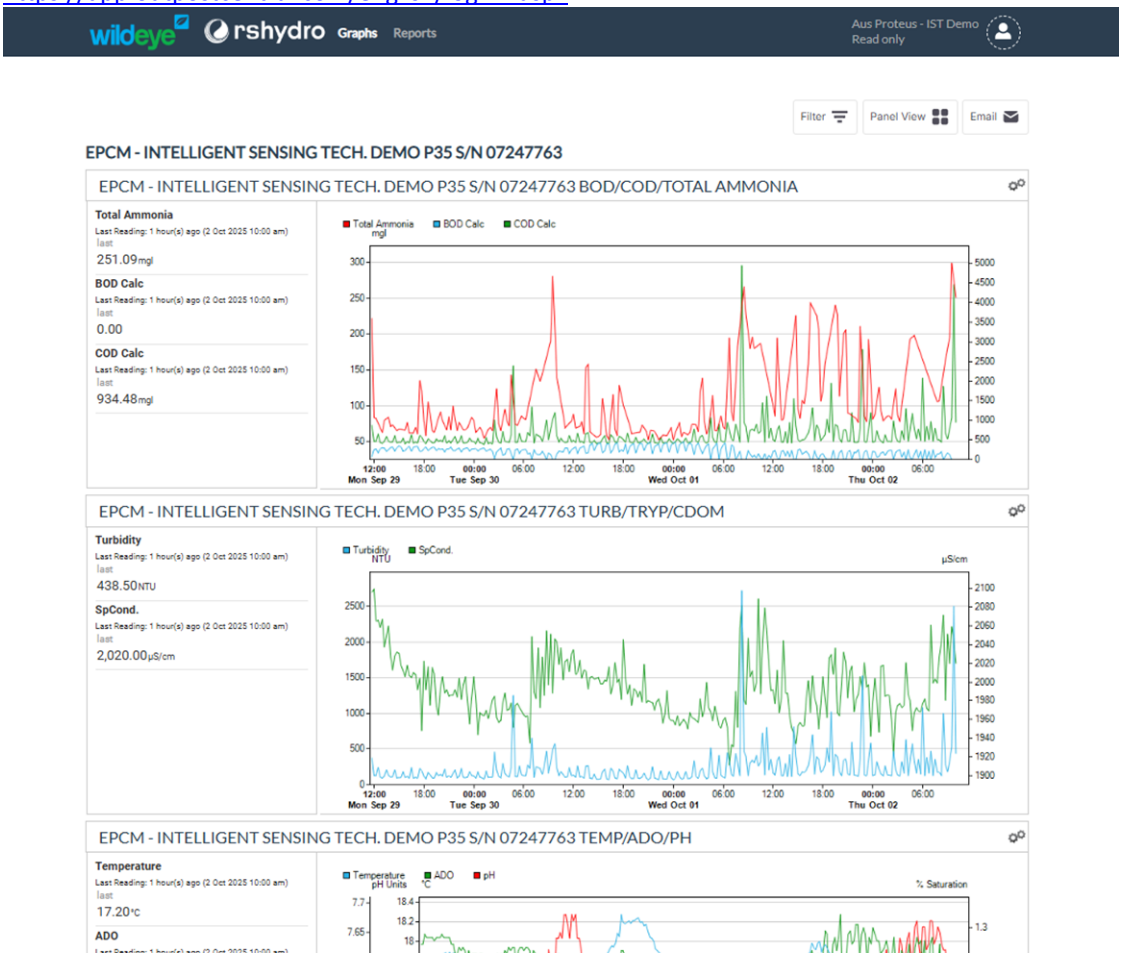
COD Calibration



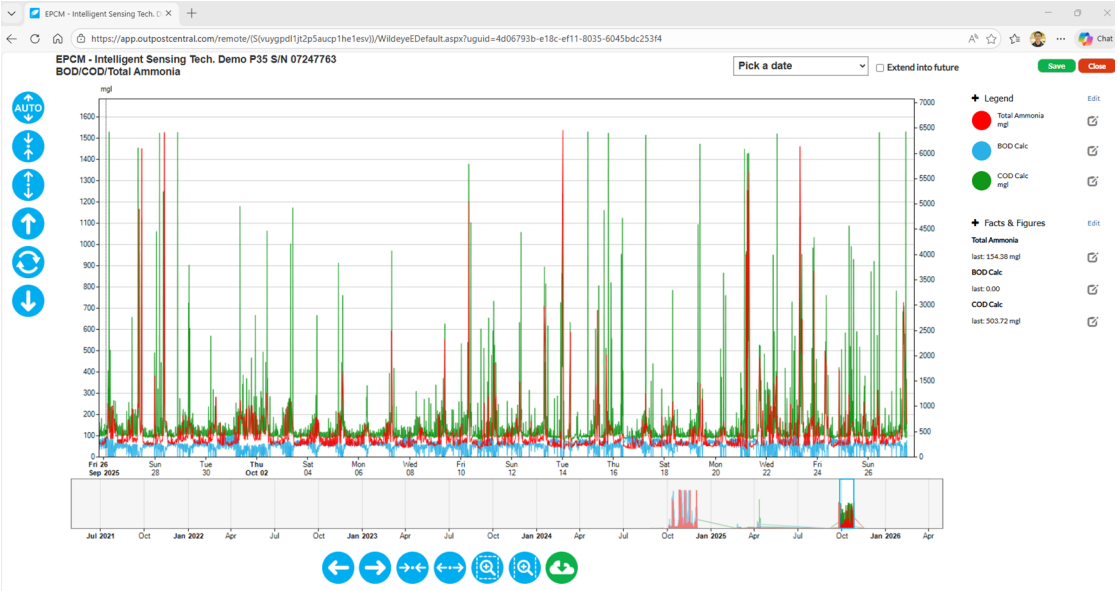
Phosphate Calibration



Sensor web-portal provided by Intelligent Sensing Technologies during the trial. https://app.outpostcentral.com/english/login2.aspx



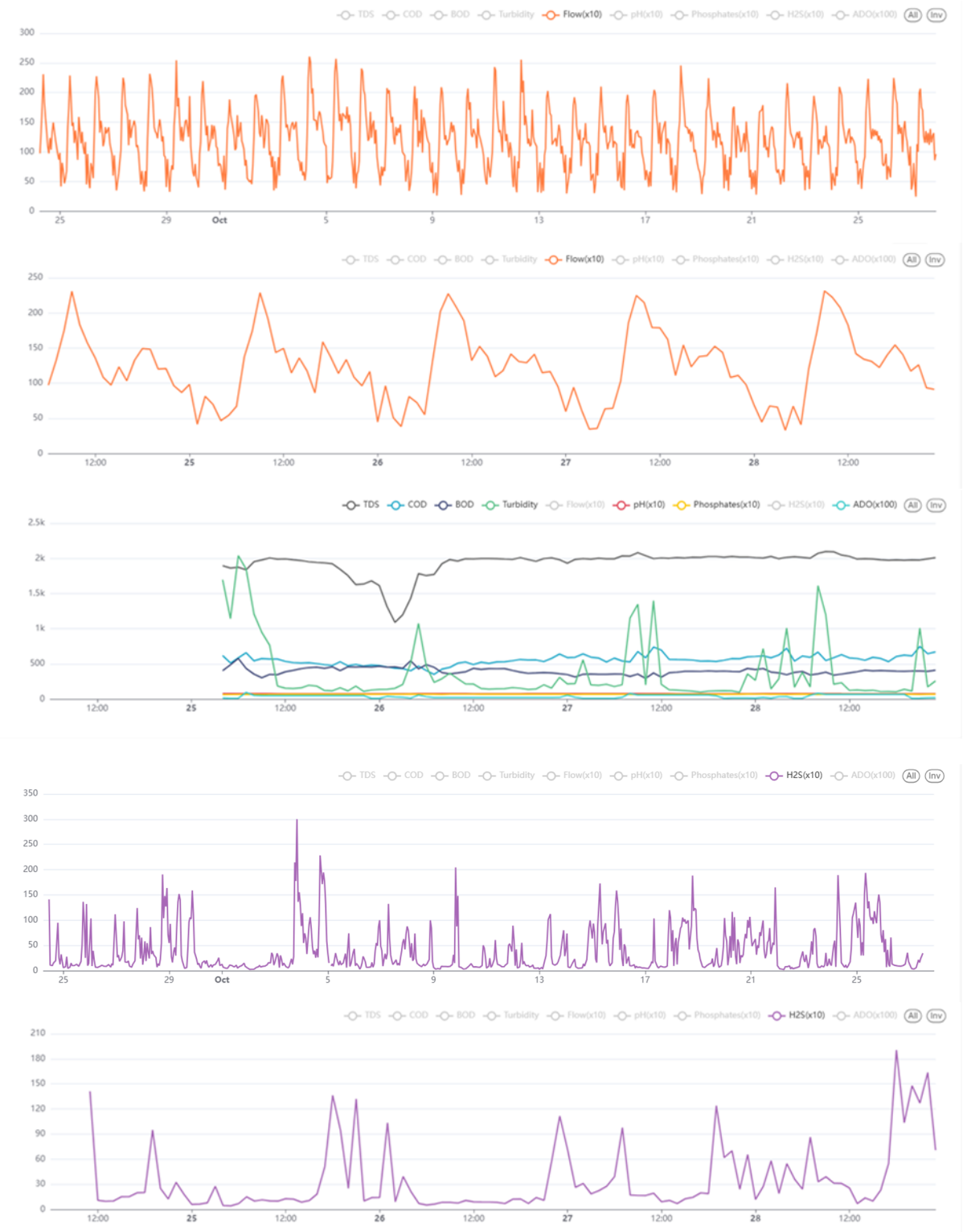
Data analysis after trial Zoom out



Note: The ammonia concentration shown here is simply a default display from the portal. The ammonia probe has not been calibrated and is not included in this trial.

APPENDIX II – THE CUSTOM-BUILD DATA DISPLAY PORTAL

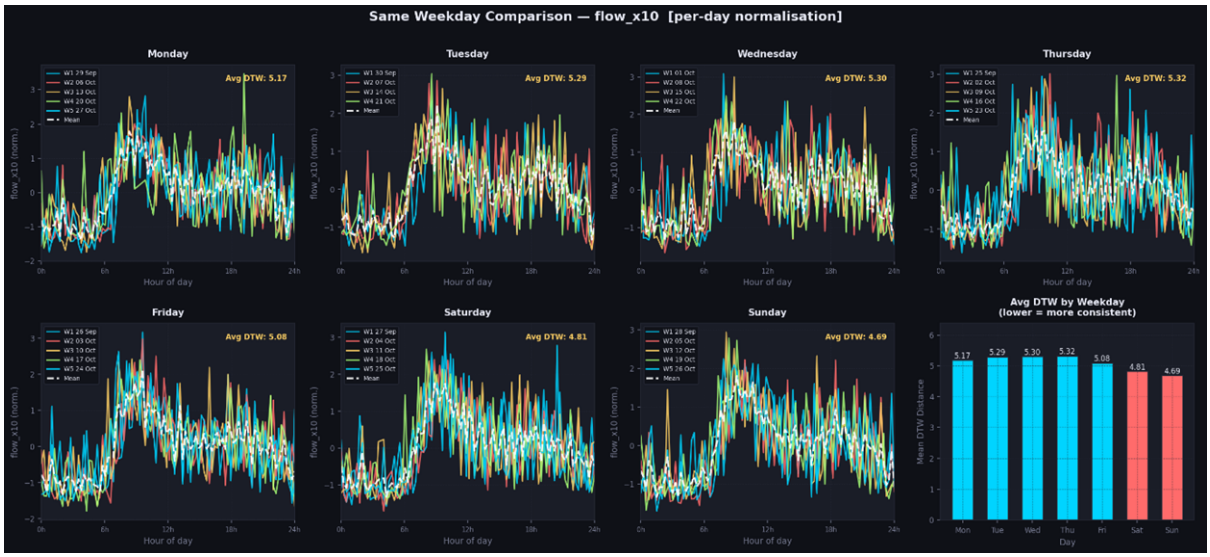
Portal: url: https://data-visualisation.cwms.dns-cloud.net/



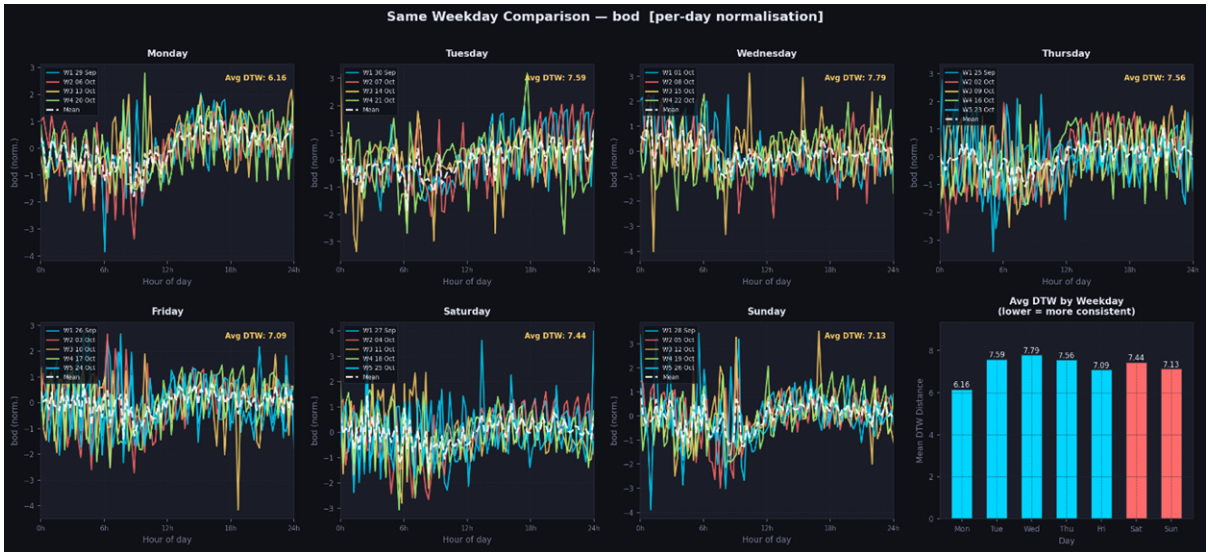
APPENDIX III - PATTERN ANALYSIS USING ADVANCED COMPUTING TECHNIQUES

Dynamic Time Warping (DTW), Flow, BOD, COD, PO_4-P, pH, Conductivity (TDS), Turbidity (SS), temperature and H_2S

A **DTW distance < 10** indicates a high degree of similarity between two sequences
Lower = more similar



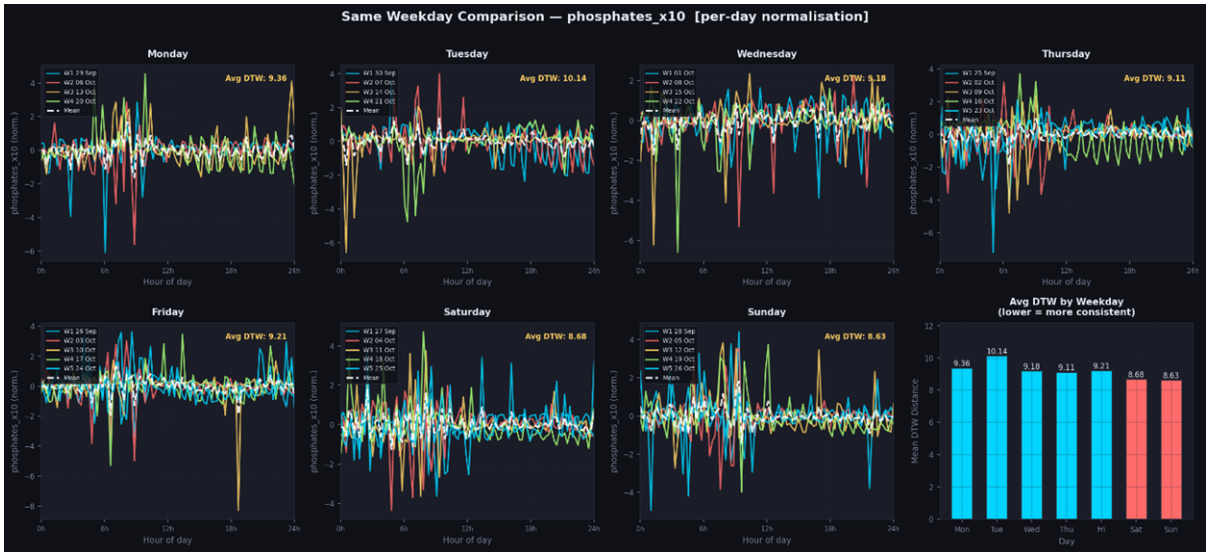
Daily flow pattern comparison in DTW (Grouped by individual days)



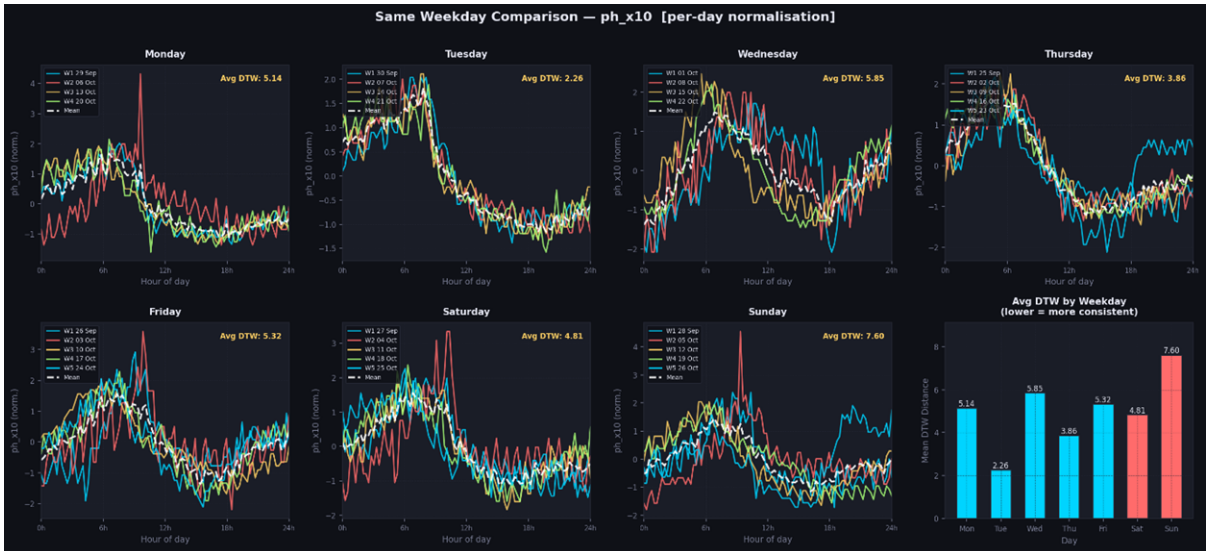
Daily BOD pattern comparison in DTW (Grouped by individual days)



Daily COD pattern comparison in DTW (Grouped by individual days)



Daily PO_4-P pattern comparison in DTW (Grouped by individual days)



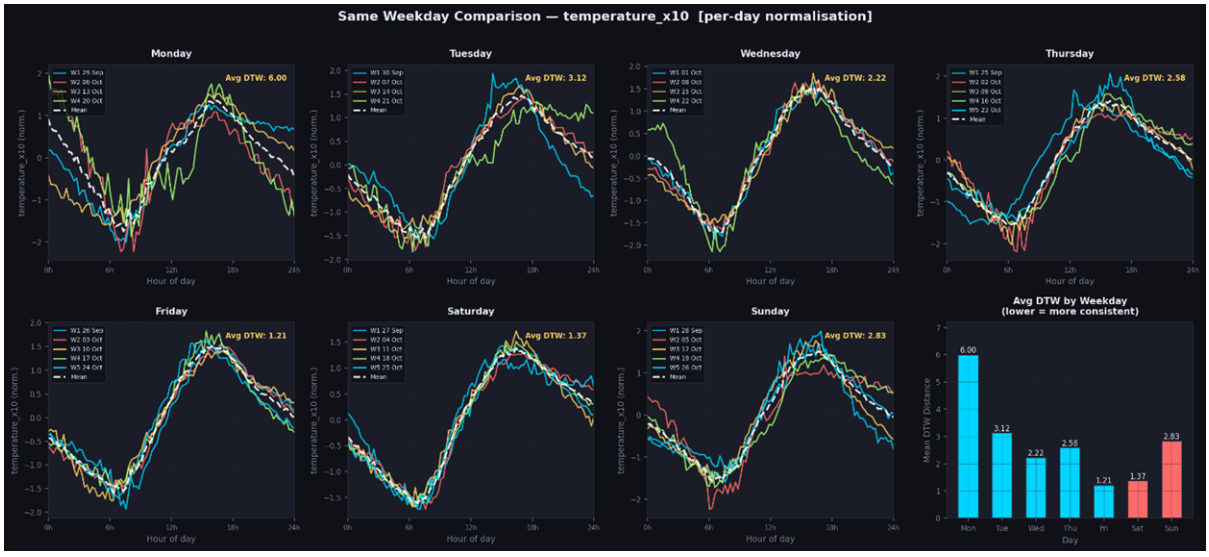
Daily PH pattern comparison in DTW (Grouped by individual days)



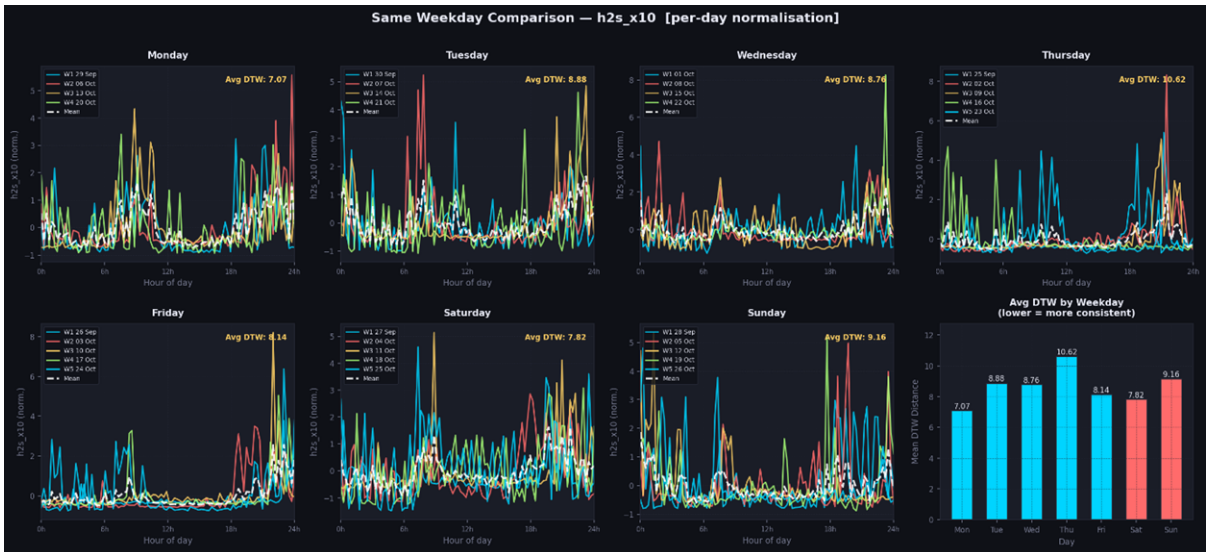
Daily TDS pattern comparison in DTW (Grouped by individual days)



Daily Turbidity pattern comparison in DTW (Grouped by individual days)

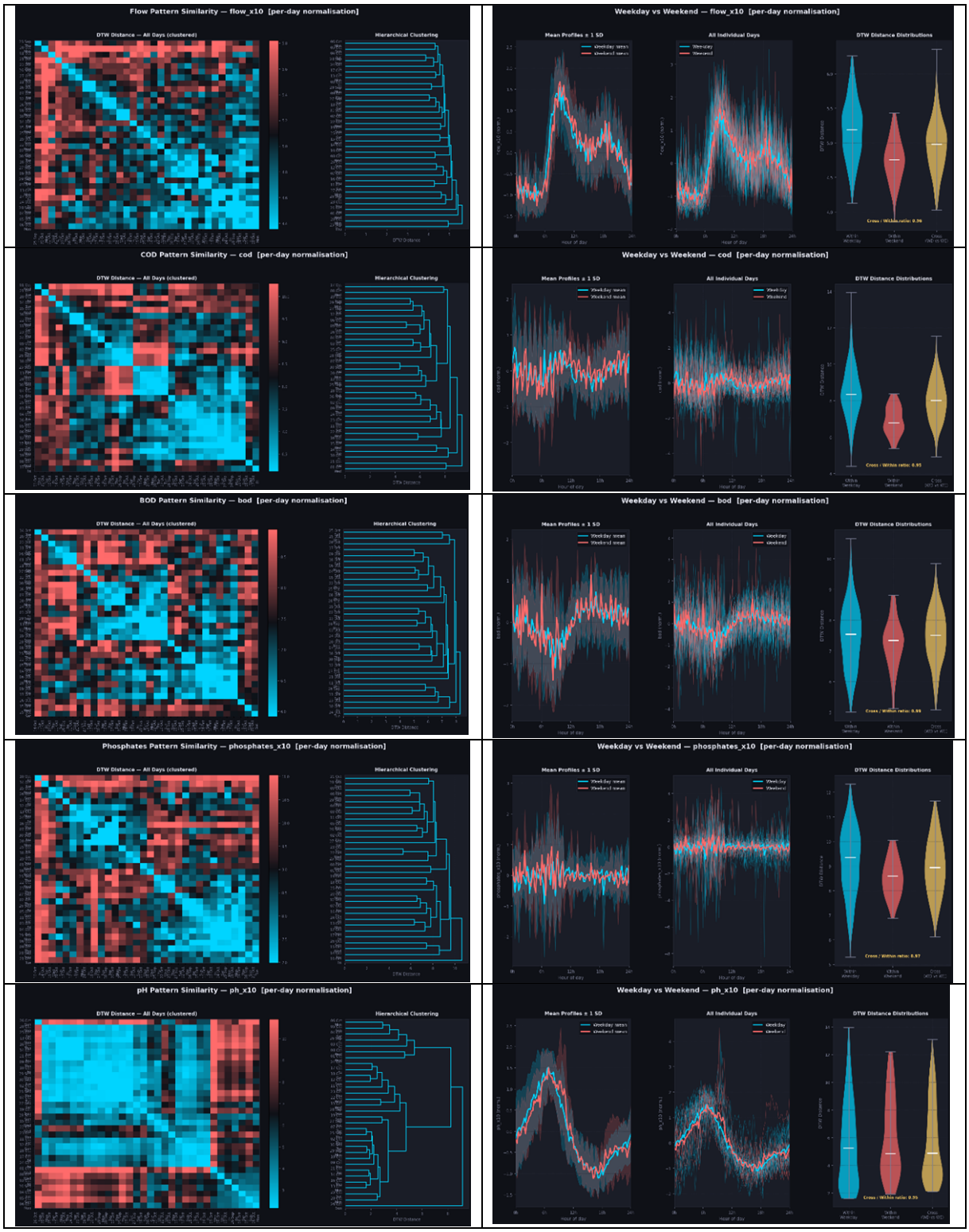


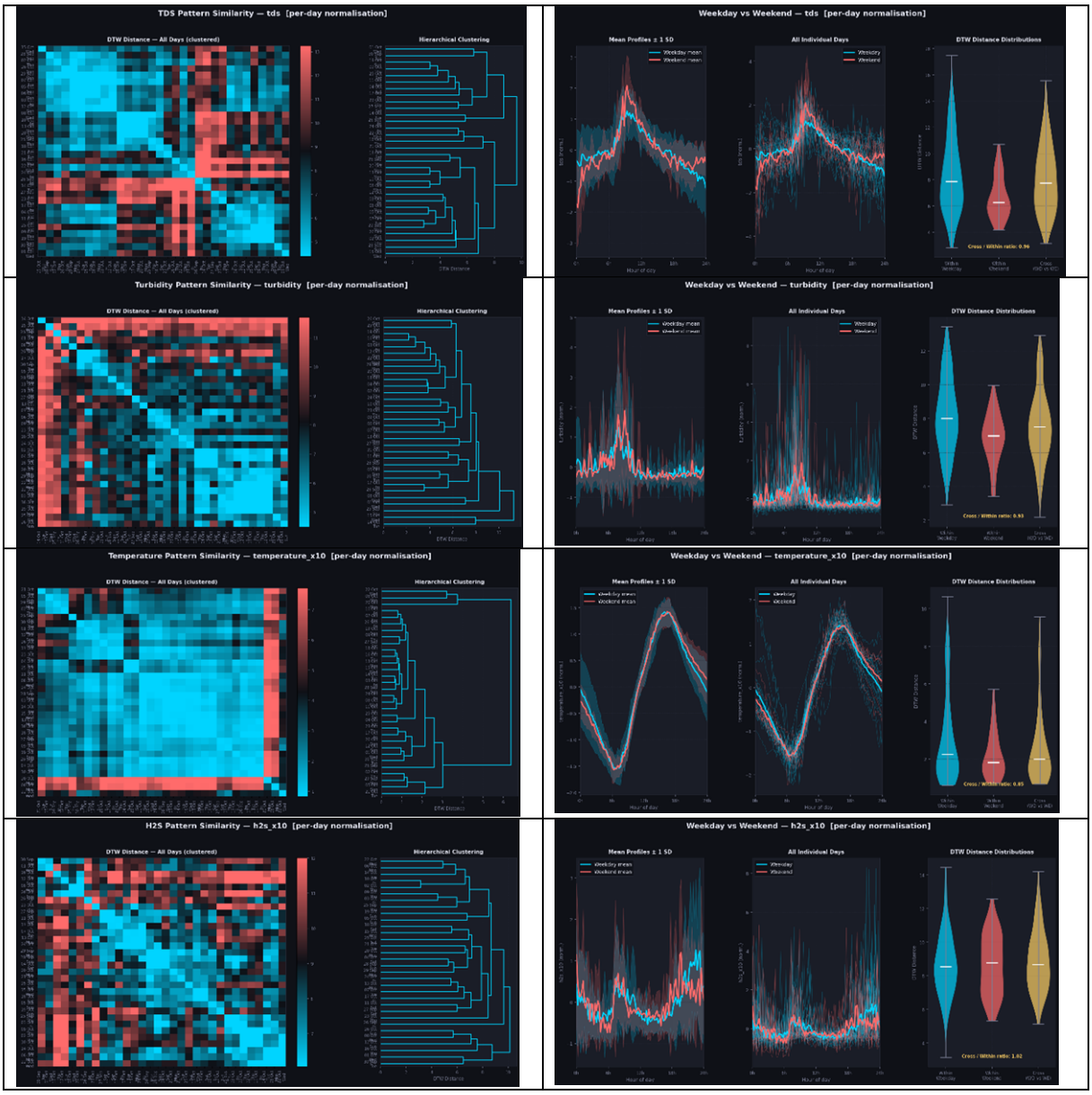
Daily Temperature pattern comparison in DTW (Grouped by individual days)



Daily H_2S pattern comparison in DTW (Grouped by individual days)

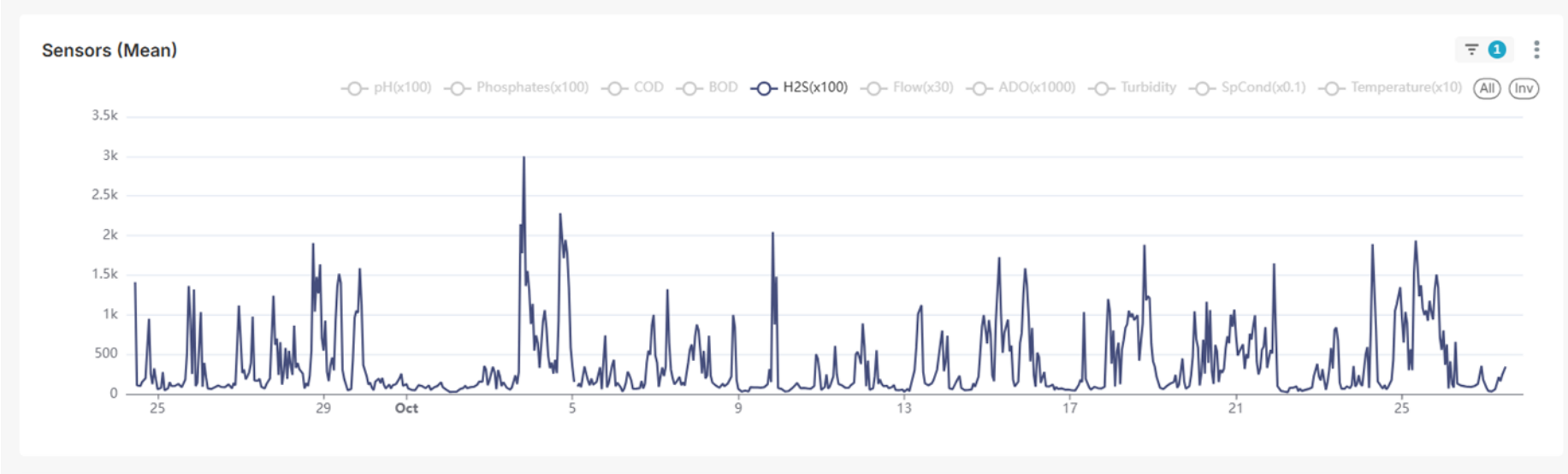
Heatmap and clustering analytics are also used to identify similar patterns across weekdays and weekends as shown below including Flow, COD, BOD, PO_4-P, pH, Conductivity (TDS), Turbidity (SS), temperature and H_2S.

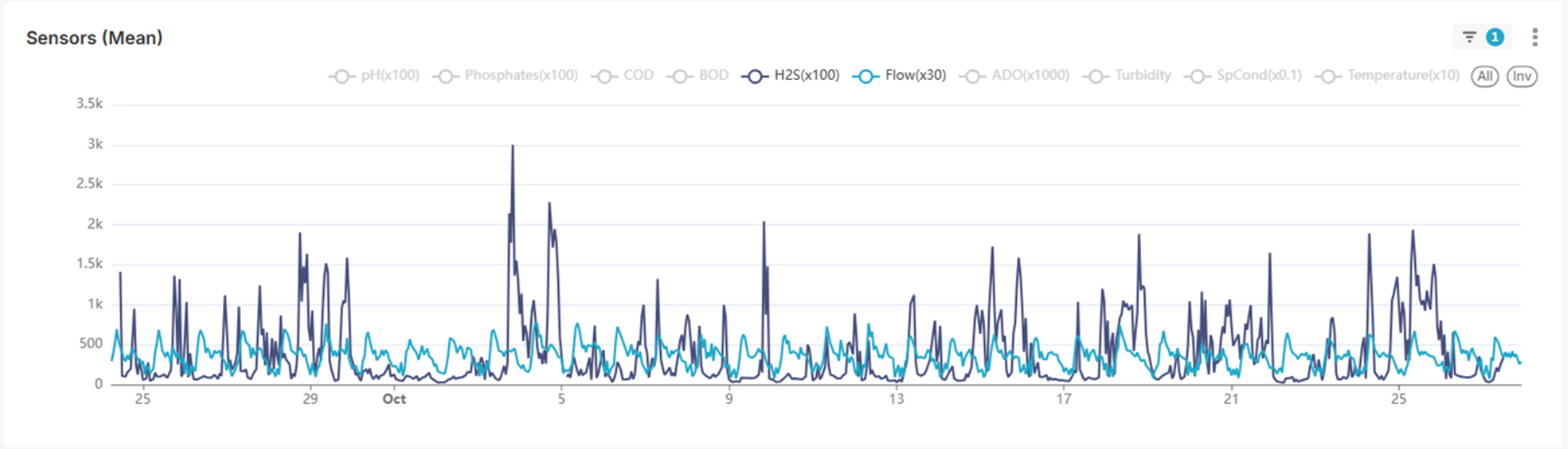


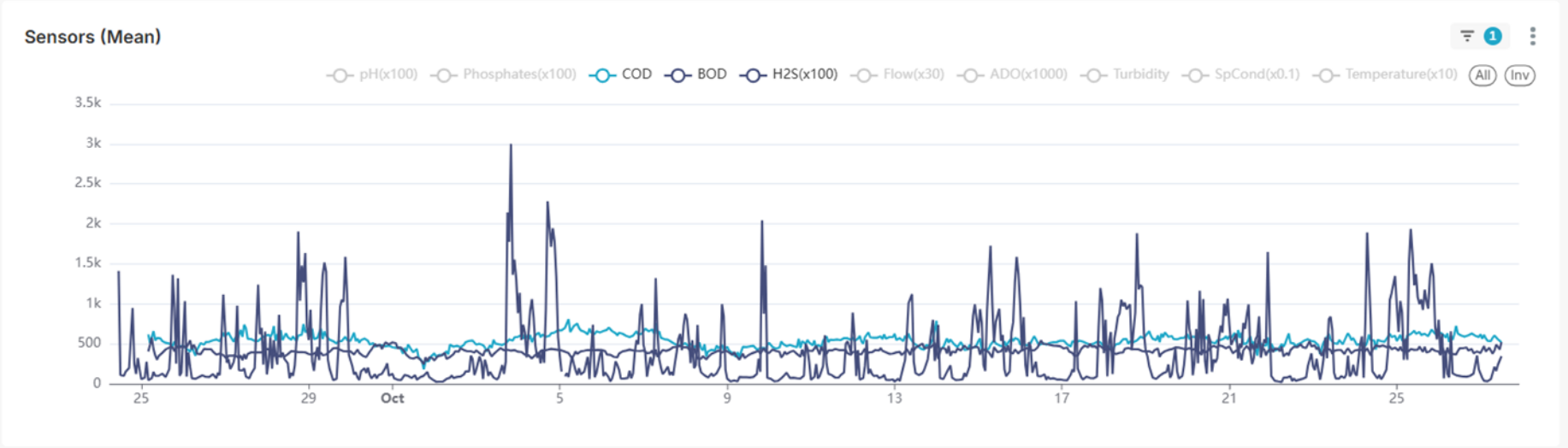


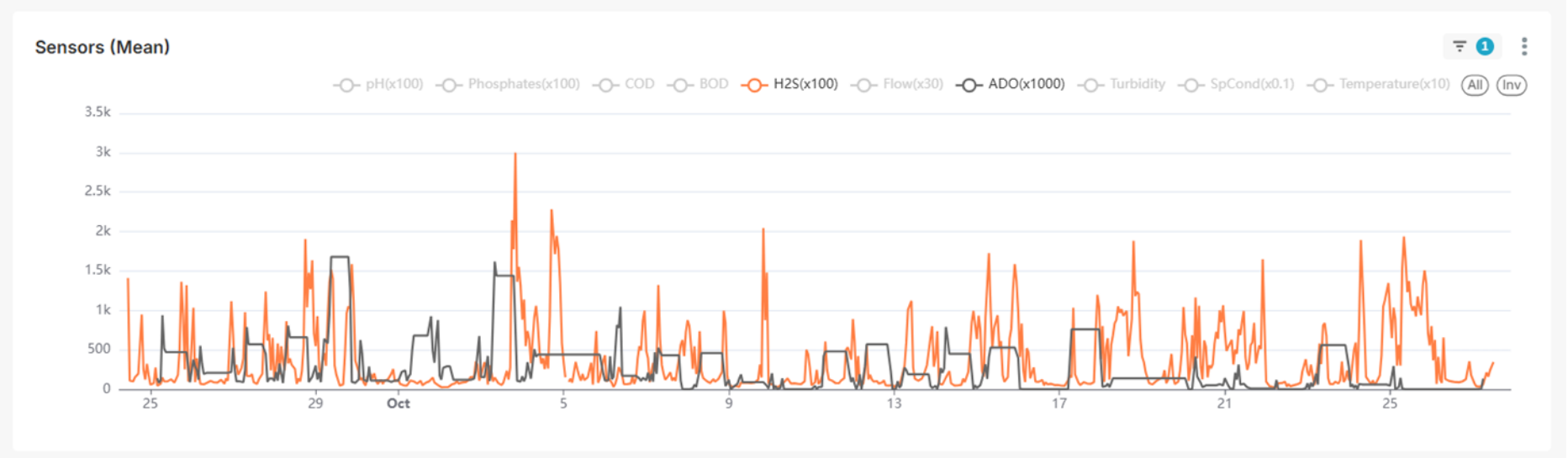
Daily pattern comparison in DTW (Grouped by weekdays and weekends)

APPENDIX IV - H_2S SENSOR RESULTS




Sensors (Mean)
pH(x100)
Phosphates(x100)
COD
BOD
H2S(x100)
Flow(x30)
ADO(x1000)
Turbidity
SpCond(x0.1)
Temperature(x10)
All
Inv
3.5k
3k
2.5k
2k
1.5k
1k
500
0
25
29
Oct
5
9
13
17
21
25


Sensors (Mean)
pH(x100)
Phosphates(x100)
COD
BOD
H2S(x100)
Flow(x30)
ADO(x1000)
Turbidity
SpCond(x0.1)
Temperature(x10)
All
Inv
3.5k
3k
2.5k
2k
1.5k
1k
500
0
25
29
Oct
5
9
13
17
21
25


Sensors (Mean)
pH(x100)
Phosphates(x100)
COD
BOD
H2S(x100)
Flow(x30)
ADO(x1000)
Turbidity
SpCond(x0.1)
Temperature(x10)
All
Inv
3.5k
3k
2.5k
2k
1.5k
1k
500
0
25
29
Oct
5
9
13
17
21
25