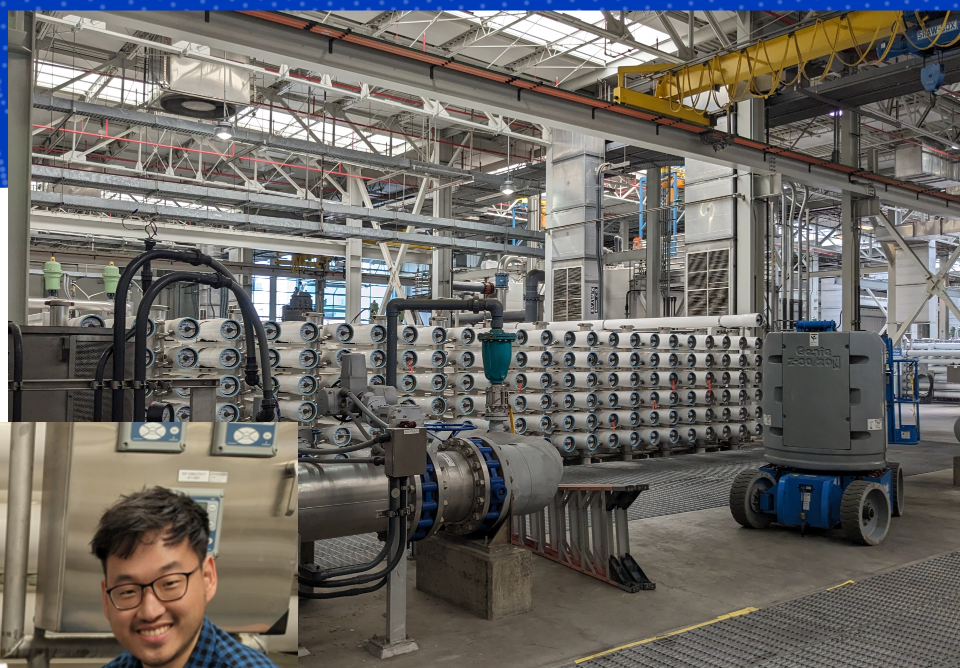
Reverse Osmosis Process Monitoring for Reuse

Intelligent Water System Challenge

July 2024



ORANGE COUNTY WATER DISTRICT

carollo®

Table of Contents

The Team

Six (6) members from Orange County Water District (OCWD) and Carollo Engineers (herein Carollo) respectively constitute this team for the Intelligent Water Systems (IWS) Challenge.

Andrew Huang, MS (ahuang@ocwd.com) is *a laboratory analyst/data scientist* in the R&D department at OCWD. He specializes in developing dashboards to monitor treatment processes such as ultrafiltration (UF) and reverse osmosis (RO) using performance and water quality (WQ) data. He is an expert in R and Python programming languages. Using R, Andrew developed an online RO dashboard to monitor sensors in real-time. Andrew will continue to develop and expand the dashboard's capabilities for the challenge solution. Andrew is the **Team Lead** (Figure 1)**.**

Jana Safarik, MBA (jsafarik@ocwd.com) is *a principal scientist* in OCWD's R&D department. She manages studies and projects related to UF and RO fouling mechanisms. She was the principal investigator and lead researcher on projects investigating novel sensors and surrogates for improved RO monitoring.

Han Gu, PhD (hgu@ocwd.com) is *a senior scientist/process specialist* in OCWD's R&D department. He is a lead on projects ranging from modernizing OCWD's in-house membrane pilots, trialing commercially and in-house available RO online "dashboards," and participates in artificial intelligence research in three funded projects.

Andy Salveson, PE (asalveson@carollo.com) is *a chief technologist* for Carollo and a nationally renowned expert in wastewater and reuse who specializes in disinfection and advanced treatment. Andy has overseen numerous research projects and is Principal Investigator of the National Alliance for Water Innovation (NAWI) project associated with this IWS solution: *Project 5.17 Data-driven Fault Detection and Process Control for Potable Reuse with Reverse Osmosis.*

Kyle Thompson, PhD, PE (kthompson@carollo.com) is *a senior reuse technologist* for Carollo and an expert in machine learning (ML) and reuse, and authored many high-impact peer-reviewed papers, including three on ML or statistical process monitoring. Kyle conducted the desktop analysis in an earlier phase of NAWI Project 5.17 that laid the foundation for this solution.

Yoko Koyama, MS (ykoyama@carollo.com) is *a ML technologist* for Carollo and an expert in modeling and data analytics for water and experienced science communicator who has spoken at many national and regional conferences on ML in the water sector. Yoko has developed RShiny dashboards for interactive machine learning tools for membrane bioreactors and granular activated carbon.

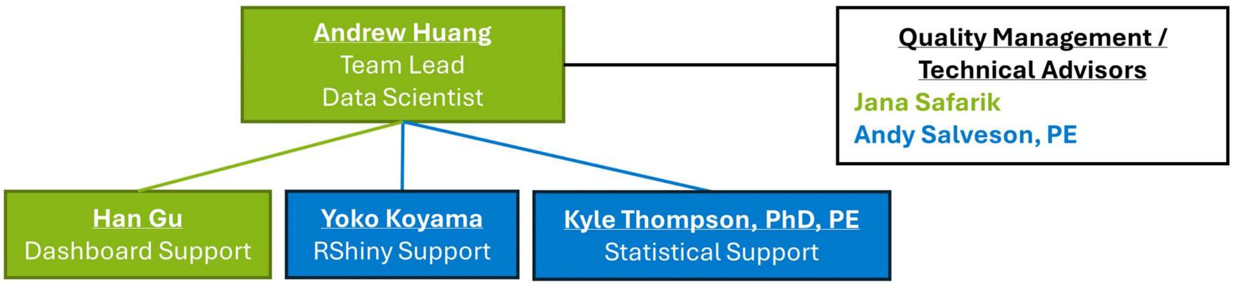


Figure 1. Challenge Team Org Chart

Green indicates Orange County Water District (OCWD) and blue indicates Carollo

Problem Statement

Advanced purification of municipal wastewater effluent is a critical, affordable water supply for drought-stricken communities. Reverse osmosis (RO) plays an important role in virus and protozoa removal for potable reuse, with stringent regulations requiring high log reduction values (LRV). Due to the lack of real-time virus and protozoa detection methods, surrogates like total organic carbon (TOC) or electrical conductivity (EC) are used to monitor RO integrity. These can be measured with online, real-time sensors in RO feed and permeate. However, even a slight error in the permeate sensor results in an erroneous low outlier in the calculated LRV, potentially triggering a false alarm for RO integrity failure. Too many false alarms would result in (1) wasted operator time on unnecessary troubleshooting, (2) a pause in water purification, (3) diversion of purified water back to wastewater headworks, leading to increased cost, energy and water waste, or (4) complacency about future alarms.

Giving treatment processes the pathogen LRV credit they merit is becoming increasingly important for water reuse. For example, California's regulations for direct potable reuse (DPR), adopted by the State Water Resources Control Board (SWRCB) in 2023[1], established stringent requirements to ensure public health safety. Specifically, DPR facilities must achieve a minimum LRV of 20 for viruses, 14 for *Giardia*, and 15 for *Cryptosporidium*. This high level of pathogen control is necessary to meet California's rigorous standards, which are more stringent than those in other states like Colorado and Texas. Due to precedence, California's rules may have national implications for states that have not yet developed their own reuse regulations. To comply with the California DRP regulations:

"For the reverse osmosis treatment process, a [reuse agency] shall propose … ongoing performance monitoring using at least one surrogate … that is capable of being monitored continuously and recorded and have associated alarms that indicate when the integrity of the reverse osmosis membrane has been compromised. … During full-scale operation of the reverse osmosis treatment process, a [reuse agency] shall continuously monitor and record the surrogate … that indicate[s] when the integrity of the process has been compromised and record when the critical limits … are exceeded, according to an approved operations plan..."

Common alarm-setting approaches assume a normal data distribution, but this can lead to excessive false alarms for even moderately skewed data.[2] A better, data-driven approach is needed to promptly detect true changes in RO performance without excessive alerts and enable DPR agencies to demonstrate high LRVs even with occasional data outliers. This approach should balance the need for prompt response with avoiding unnecessary downtime due to false alarms.

The Solution

RO Dashboard System Architecture

The Intelligent Water System is a real-time dashboard that monitors removal of RO surrogates, such as TOC and Peak C fluorescence, to ensure they stay above target values, using advanced statistical criterion. It receives continuous real-time data from a full-scale 5-mgd RO unit, with sensors sending information to the cloud via a Raspberry Pi, where LRVs are calculated and visualized in real-time (Figure 2).

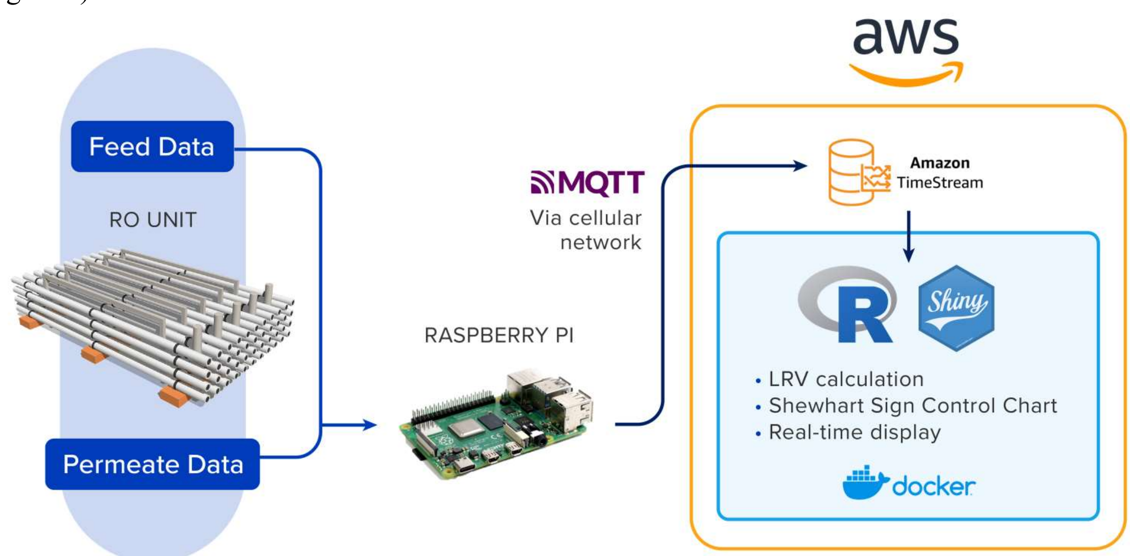


Figure 2 OCWD's Intelligent Water System

OCWD's Intelligent Water System transmits feed and permeate data via MQTT[3] over a cellular network to Amazon Web Services (AWS) for cloud storage and processing.[4] The data is visualized on an RShiny[5] dashboard, which includes fault detection algorithms coded in R.[6]

The dashboard displays raw data such as TOC and Peak C fluorescence, and provides real-time advanced analytics, all of which are detailed further in the subsequent sections.

Data Collection and Management

For data collection and transmission, OCWD implemented a robust system leveraging Message Queuing Telemetry Transport (MQTT)[3] protocol and a Raspberry Pi device. This setup was chosen to efficiently transmit data from the RO unit to OCWD's cloud database, addressing the challenge of poor Wi-Fi coverage at the site. MQTT, designed for low-bandwidth, high-latency networks, enables efficient transmission of frequent, small data packets. The data collection and MQTT transfer are programmed in Python,[7] utilizing its extensive libraries for seamless integration with hardware interfaces and network protocols.

For data management, AWS Timestream was selected as the cloud database solution. This fully managed time-series database is optimized for large-scale, time-stamped data, offering efficient data retrieval, retention policy management, high ingest rates, and SQL-like querying capabilities. These features simplify time-series data management and enable quick data retrieval for analytics and visualization. Historical data which date back to 2024-01-03 are stored and accessible via the dashboard user interface detailed in a later section.

Peak C fluorescence and TOC data are continuously collected from the RO feed and permeate of a GWRS 5-mgd RO unit. The Raspberry Pi transmits these datasets in real-time to OCWD's AWS Timestream database via MQTT protocol, recording at 2-minute intervals.

Data Quality Check

Raw TOC and fluorescence data undergo quality check measures prior to data analytics. For TOC, data points below the detection limit were removed; for peak C fluorescence, any data points equal to zero are removed.

Data Analytics

Introduction to Shewhart Sign Control Chart

The dashboard's advanced analytics component features a Shewhart sign control chart[2] to assess if surrogate LRVs have dropped below the target level over a past rolling time period. The Shewhart sign control chart[2] does not assume normality, unlike conventional control charts. It treats each datapoint as a "coin flip" above or below a target median; it triggers an alarm if too many recent datapoints below the target indicate there's a certain probability median has probably dropped the target. Analysis confirms that both conventional and novel RO monitoring surrogates tend to be non-normal in full-scale data, supporting this technique's suitability (Figure 3).

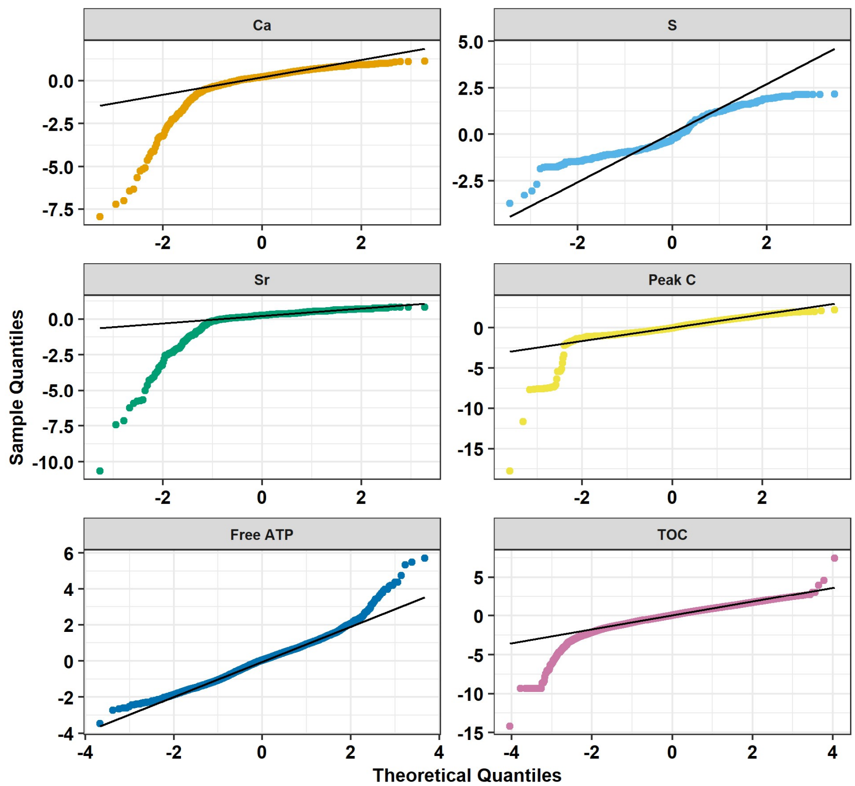


Figure 3 Normal Probability Plots of Free Adenosine Triphosphate (ATP), Ca, S, and Sr LRVs at OCWD

Sample quantiles are the actual z scores of the datapoints, and theoretical quantiles are the expected z-scores for datapoints.

Shewhart Sign Control Chart Applied to OCWD Historical Data

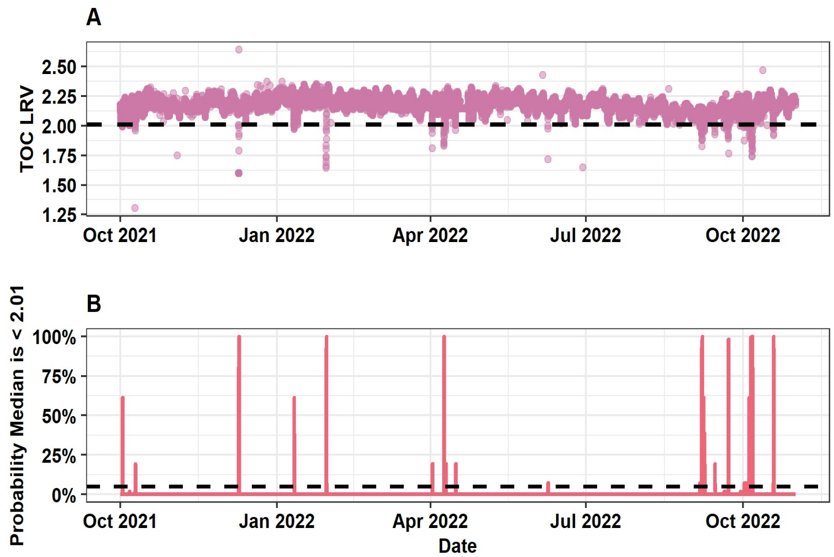


Figure 4 Shewhart Sign Control Chart Results on TOC Data from OCWD

(A) TOC LRV. The dashed line indicates a 2.01 LRV target. (B) Probability median TOC LRV has dropped below 2.01 based on the past 12 points. The dashed line represents a p-value > 0.05 threshold to trigger an alarm based on the Shewhart sign control chart.

In prior work, conventional and novel surrogate sensor data were collected at OCWD and shared with Carollo. The Shewhart sign control chart, applied to 12-datapoint windows, could detect LRV losses in as few as three measurements, faster than current protocols. For most surrogates, it resulted in at least 50 percent fewer alarm events compared to single-datapoint alarms, potentially saving operator time. Fluorescence and TOC are effective LRV surrogates for early RO operation warning. Peak C fluorescence is economically advantageous for verifying high LRVs. While previous research showed less benefit in reducing false compared to other surrogates, its non-normal distribution makes the Shewhart sign control chart appropriate. Simulation on TOC LRV data from OCWD achieved a 50 percent reduction in false alarms compared to single-point alarms (Figure 4).

Shewhart Sign Control Chart Applied to Real-Time Data

While WQ parameters are collected at 2-minute intervals in the Timestream database, the Sign Test is applied to a rolling 3-hour window, using data retrieved at 15-minute intervals. This approach, yielding 12 sample points per window, offers a more robust analysis than using the 2-minute interval data, as it better captures real water quality changes while minimizing false alarms from short-term fluctuations.

User Interface (UI) and Data Visualization

OCWD's R Shiny-based[5] dashboard offers a UI to access the comprehensive functionality for both Peak C fluorescence and TOC data analysis. It features automated real-time data visualization over the past 24-hour window, historical data plotting capabilities, and data table download options for both parameters. Additionally, the dashboard incorporates Shewhart sign control chart results for each metric, providing advanced statistical analysis.

For the real-time data display, feed and permeate WQ parameters are displayed at 2-minute intervals. For accessing historical data, users can look up WQ parameters that date back 2024-01-03 and download data tables in csv format.

For the advanced statistical component, the dashboard features the following functions:

1. Calculate and record whether a Shewhart sign control chart alarm has been triggered.
2. Display the hypothesized median, probability of exceedance, and an indication of the alarm on live charts.
3. Record of times when the alarm has been triggered, if any.

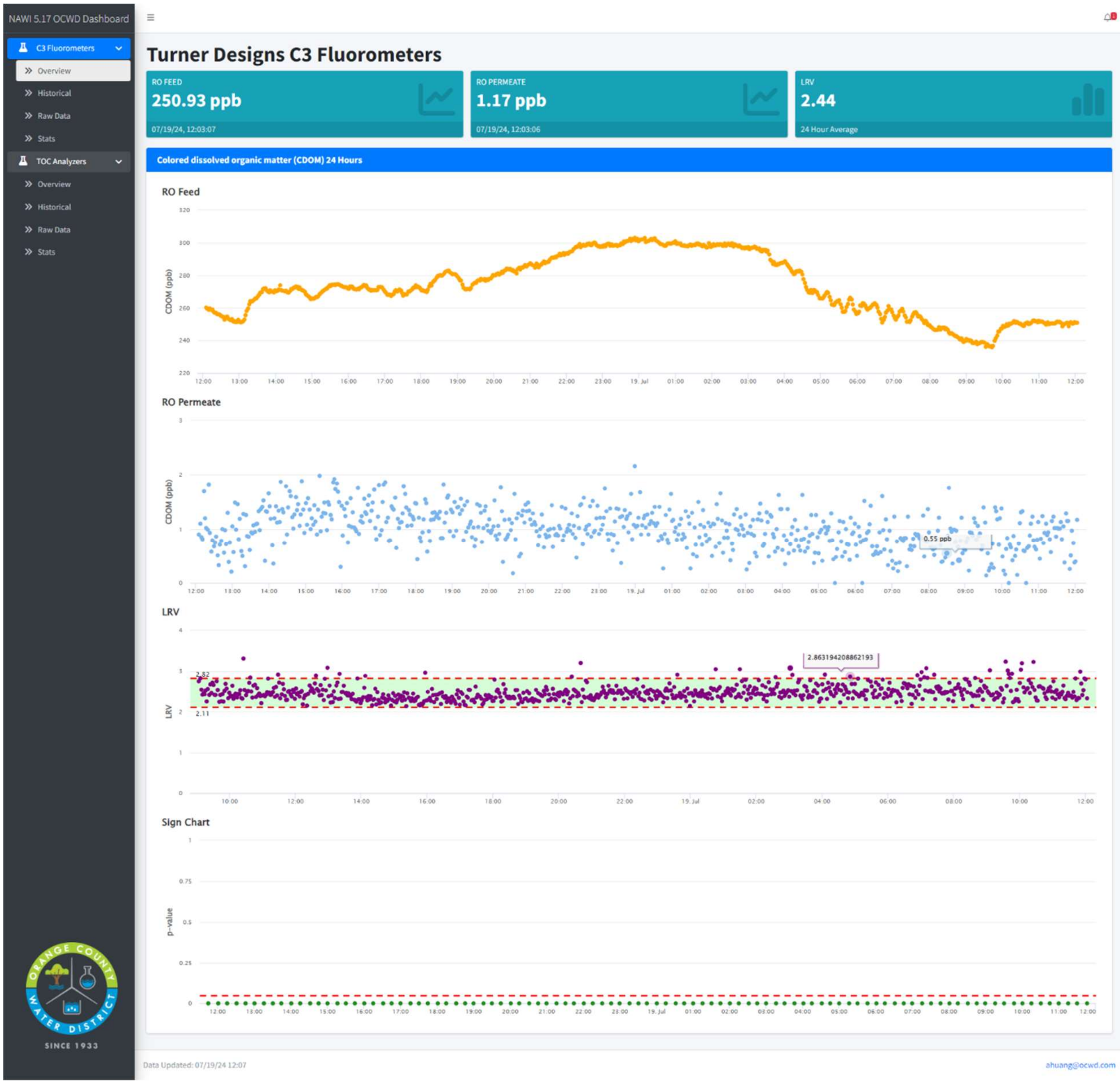


Figure 5 A Screenshot of the Application's User Interface

The main overpage page displays current RO feed and RO permeate concentrations as well as LRV for the selected parameter. Time-series charts plot the last 24 hours of concentration, LRV, and p-value of the Shewhart Sign Control test.

Solution Implications

This integrated approach of data collection, retrieval, visualization, and analysis allows users to monitor, analyze, and extract valuable insights from both Peak C fluorescence and TOC data. It uses a single, user-friendly interface, enhancing the ability to track and assess RO system performance at a glance. The system's design, utilizing widely available components and open-source technologies, makes it readily replicable for other utilities. The use of Raspberry Pi devices, Python programming, MQTT protocol, and AWS cloud services provides a scalable and cost-effective solution that can be adapted to various operational scales and budgets. Furthermore, the modular nature of the system allows utilities to customize and expand the setup to meet their specific monitoring needs, making it an accessible and flexible option for water/wastewater/reuse facilities looking to enhance their RO system monitoring capabilities.

Project Evolution and Next Steps

The cloud-based dashboard was developed in a sequential manner as shown in Figure 5 below.

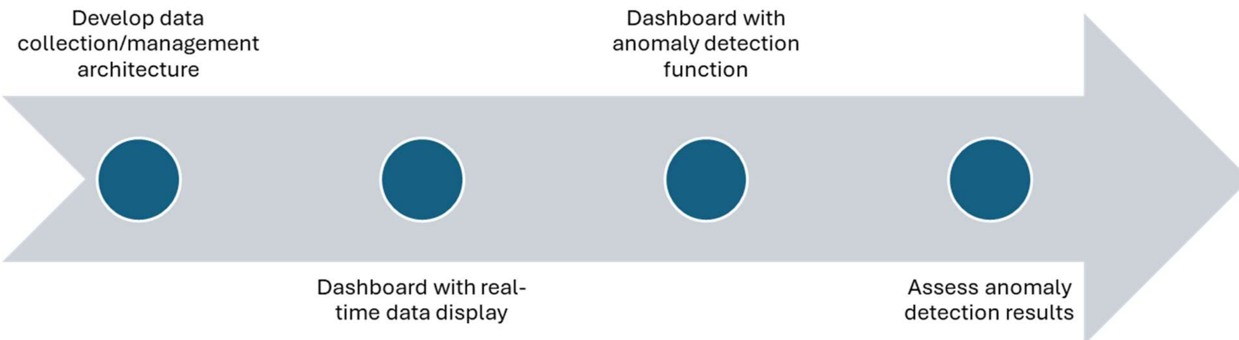


Figure 6 Project Evolution Timeline

In the upcoming period, preferably spanning at least two months, we will assess whether the number of alarms triggered by Shewhart sign control chart alarms is lower than the number of alarms triggered by individual data points assuming a normal distribution.

To disseminate this innovative approach and encourage its adoption across the water industry, OCWD and Carollo plan to present this work at several key conferences. Presentations are scheduled for the upcoming American Water Works Association (AWWA) Water Quality Technology Conference (WQTC) and pending with the AWWA/WEF Utility Management Conference. These platforms will provide opportunities to showcase the system's effectiveness, discuss implementation strategies, and engage with other utilities and researchers interested in advanced RO monitoring solutions. Through these presentations, we aim to share practical insights, lessons learned, and best practices, fostering collaboration and knowledge exchange within the water industry.

Reference

(1) State Water Resources Control Board. *Regulating Direct Potable Reuse in California | California State Water Resources Control Board*. https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/direct_potable_reuse.html (accessed 2024-05-09).

(2) Amin, R. W.; Reynolds Jr, M. R.; Saad, B. Nonparametric Quality Control Charts Based on the Sign Statistic. *Communications in Statistics-Theory and Methods* **1995**, *24* (6), 1597–1623.

(3) Amazon Web Service. *What is MQTT? - MQTT Protocol Explained - AWS*. Amazon Web Services, Inc. https://aws.amazon.com/what-is/mqtt/ (accessed 2024-05-17).

(4) *Cloud Computing Services - Amazon Web Services (AWS)*. https://aws.amazon.com/?nc2=h_lg (accessed 2024-05-17).

(5) Chang, W.; Cheng, J.; Allaire, J. J.; Sievert, C.; Schloerke, B.; Xie, Y.; Allen, J.; McPherson, J.; Dipert, A.; Borges, B. Shiny: Web Application Framework for R, 2024. https://shiny.posit.co/.

(6) R Core team. *R: The R Project for Statistical Computing*. https://www.r-project.org/ (accessed 2021-06-09).

(7) Van Rossum, G.; others. Python Programming Language. In *USENIX annual technical conference*; Santa Clara, CA, 2007; Vol. 41, pp 1–36.