Innovative data collection and management strategies for improved water treatment efficiency

Tom Meuzelaar[1], Shannon D. Zahuranec[2], James P. Jonas[3]

[1]*Life Cycle Geo, LLC, tom@lifecyclegeo.com, ORCID 0000-0003-3119-8390*

[2]*Life Cycle Geo, LLC, shannon@lifecyclegeo.com, ORCID 0009-0006-6804-8854*

[3]*Life Cycle Geo, LLC, jim@lifecyclegeo.com, ORCID 0009-0009-9934-0673*

Abstract

The performance of active and passive water treatment systems can be negatively influenced by seasonal and diurnal water quality and quantity fluctuations of the feedwater. Treatment performance can be improved by proactively modifying the water management strategy in response to these fluctuations. Employing modern, full, or partially automated configurations as part of the water management strategy can efficiently optimize these types of water treatment systems.

These automated system configurations build upon existing components with emerging technologies resulting in the following innovative data workflow: (a) automated, frequent collection of data at various treatment system monitoring points using sensor-based technologies, (b) automated alarms that can signal remote system upset conditions (compliance exceedances, pump malfunction, clogging, fouling etc.), (c) telemetry-based data upload and ingestion by a cloud-based data management system, (d) automated data cleaning and preparation pipelines, and (e) the use of conventional statistical and computational techniques and, when necessary, more advanced algorithms such as machine learning to analyze incoming data streams.

This workflow promotes intelligent, real-time guidance on water treatment and management decisions, such as treatment methods, dosage frequency, water diversion, and more, can be provided in near real-time through visualization, reporting, dashboards, and PLC controls. Results of the implementation of various components of this workflow demonstrate benefits such as improved treatment efficiency, more reliable operation, compliance with standards at discharge points, and overall reduction in labor, reagent costs, and energy demand. Additionally, transferring all system data, including sensor data, images, operator logs, and legacy PDFs, to a cloud-hosted data warehouse opens important opportunities for enhancing value extracted from collected data.

An example application is provided for a remote semi-passive treatment system designed to treat waste rock drainage prone to upset conditions due predominantly to large fluctuations in water volume and difficulty staffing an experienced operator. Finally, the authors discuss how this workflow could be used to optimize a full-scale active treatment system with multiple sensor locations and numerous real-time data streams using a digital twin/machine learning approach.

Keywords: Data management, data collection, mine drainage, passive treatment, groundwater, surface water

Introduction

The advent of modern sensor technology, cloud-based data management, and advanced data analytics methods (e.g. machine learning) provide operators with new opportunities to optimize the full spectrum of passive to active water treatment systems. In this paper, the authors summarize the components of a fully automated water treatment system data collection, management, analysis, and reporting/visualization workflow and discuss the benefits of such a workflow.

This innovative workflow addresses the following challenges that many water treatment operators currently face. Automated sensor-based data collection substantially reduces or eliminates the need for expensive, manual sampling and lengthy laboratory analysis, and provides a source of high frequency data. Treatment operators often struggle to maintain and train staff to collect consistent and reliable system performance data. Data is often not collected at a sufficient frequency to adequately monitor system performance towards optimization and commercial laboratory analysis can take a month or longer, especially when labs are backed up. Sensor-based data can be processed on edge devices (for real-time decision making), within the Supervisory Control and Data Acquisition (SCADA) system or uploaded via telemetry to cloud. Data management on the cloud includes development of a universal schema, data validation, feature engineering, redundancy and disaster recovery, and integration with field-based monitoring systems. And finally, real-time or batch analytics on the cloud facilitates intelligent decision-making by way of visualization, reporting, dashboarding, recommendation engines, and automation of Programmable Logic Controller (PLC) devices for situations where typical PLC feedback loops are insufficient.

Field Data Collection and Management Methods

A typical water treatment control system is comprised of passive components (i.e. sensors), active components (e.g. valves and pumps), Programmable Logic Controller (PLC), and Supervisory Control and Data Acquisition (SCADA) system.

Sensors provide critical input to the treatment system and can come in many different forms with equally as many functions. The array of sensors in a treatment system are often distributed at the influent, across the treatment train at key locations, and at the effluent. Measurements made by the sensors are transmitted back to the PLC (via a wired connection, radio, cell, or even satellite) where logical commands are applied, based on the sensor data, to adjust controls such as valves and pumps. Many sensors currently on the market utilize microcontrollers to manage power, data requests, and data transmissions in a digital format rather than transmitting raw analog (e.g. 4–20 mA) measurements. The computing power of these microcontrollers has grown substantially over the past decade and many microcontrollers on the market today have the capability of efficiently storing and running machine learning algorithms that could be beneficial for locally identifying imminent failures of the sensor or equipment.

Peng, et al. (2021) describe a principal component analysis (PCA) model for predicting fault detections in submersible pumps and Yang, et al. (2022) describe a denoising autoencoder and support vector machine (DAE-SVM) approach for fault detection in submersible pumps. Both methods may be applicable for deployment on edge devices. The benefits of employing PCA for fault detection include improved efficiency, simplicity, and speed. Limitations include the assumption of data linearity and an emphasis on variance, which might not point to the most relevant features for fault detection. Application of DAE-SVM for fault detection has the advantages of noise reduction, high accuracy combined with robust feature representations, and the ability to capture non-linear relationships in the dataset. For environments where computational resources are not severely limited and where accuracy in detecting complex fault patterns is paramount, DAE-SVM would likely be the preferable choice. However, for simpler scenarios or when computational efficiency is a priority, PCA could provide a viable and effective solution.

PLCs act as a relay bank wherein PLCs apply a few different basic algorithms (e.g. on/off, feedback–proportional, integral, derivative) to control the active components based on the passive component inputs. PLCs are suitably designed to provide real-time control based on immediate feedback, trigger alarms when equipment or sensors are not operating within predefined parameters, locally controlled via Human Machine Interfaces (HMIs), and remotely controlled via the SCADA system. PLCs communicate with peripheral devices (sensors, valves, pumps) via one of many different communication protocols. Since all of the logical algorithms to run the treatment system are inherently embedded in the PLC, the PLC can autonomously control the treatment system without internet connection.

The SCADA system is both hardware and software designed to provide a higher level of supervision over the treatment process, store data collected through the PLC, monitor and control at an enterprise level (i.e. over more than one treatment system), provide remote access to the PLC, and provide data analytics and visualization. Many SCADA systems deployed today have the ability to store data locally and in the cloud. Pushing treatment system data into the cloud via the SCADA can accommodate higher order data analytics and data visualization.

Overall, in typical water treatment systems, PLCs act as front-end devices interfacing with the passive and active components of the treatment system while the SCADA system provides a centralized location for visualizing the treatment process and treatment process control. This type of control system is very effective in systems that have little to no lag time in the "dose-response" reactions or are overseen by highly skilled operators that have spent many seasons overseeing the particular treatment system. When systems require substantial foresight by experienced operators to anticipate potential issues or "ramp up" the treatment system's capacity (in the case of some semi-passive treatment systems), transmission of data to a centralized server (or cloud) can provide a means for effective communication between the operators and the decision-makers through the use of visualizations (i.e. a dashboard) as well as making the data available to more sophisticated predictive algorithms that ultimately can assist both operators and decision-makers with making prompt and accurate decisions.

Cloud Management

Field sensor-based data uploaded by telemetry can be stored in popular cloud-based data storage platforms such as Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP) etc. These platforms provide ample advantages to data management workflows such as scalability, accessibility, security collaboration and cost-effectiveness. Plus a cloud platform can be used to host highly different data types. For example, site photos can be seamlessly incorporated with water chemistry data, alongside information extracted from legacy PDFs, utilizing cloud OCR services like AWS Textract.

The process of managing sensor data in the cloud includes cloud database setup: the initial step in the process is selection of the cloud platform suitable to the needs of the project followed by creating a cloud database. Transmitting data to the cloud relies on remote sensors that can be configured to transmit data to the cloud as explained above; cloud platforms provide encryption tools and services to ensure the security of the data. An automated workflow can be created using Python scripts to streamline the process of ingesting data into cloud storage. This workflow incorporates a data dictionary to standardize datasets according to project requirements. Quality assurance and quality control (QA/QC) measures, including verifying data units and organizing information for convenient retrieval and analysis, are integrated into the workflow. Cloud platforms provide various tools for real-time processing of the data. For example, AWD Lambda enables code execution bypassing the need to provision or manage servers. With AWS Lambda, code can be executed in response to events or triggers without the need to worry about the underlying infrastructure. Analytics techniques range from simple statistical, to numerical and graphical methods, to complex unsupervised and supervised machine learning algorithms. Reports, graphs, dashboards, and recommendation engines are implemented to help operators make more informed decisions. The cloud-based database can be integrated to track the performance and health of the remote sensors and the overall data management infrastructure. Backup and disaster recovery strategies can be implemented to safeguard against data loss. Cloud platforms often provide tools for automated backup and recovery. Workflows can be developed to conduct thorough testing to ensure that data is being collected and handled accurately.

Case Study

It is common to witness the failure of semi-passive treatment systems after only a few years of operation. Poor or inadequate design is often thought to be the culprit, but recent advances in automation have allowed for restored operation of such systems and suggest difficulty with manual system operation may be the most critical contributor to underperformance.

One such example exists with a hypothetical site in a rural area (reflective of two separate, confidential sites we've worked on and combined here for ease of discussion), where a remote bio-treatment system designed to treat elevated concentrations of sulfate and nitrate in seepage emanating from the toe of a waste rock pile partially failed after only two years of operation. Failure of the system was related to unmanaged seasonal fluctuations in water volume and due to inconsistent manual operation. The semi-passive system was selected, in part, to accommodate ease of operation given the remote site setting; however, this expectation of easy operation accompanied by a strong seasonal hydrologic pattern led to annual losses in treatment efficiency followed by complete failure after four years.

System maintenance and automation implemented nearly a decade after initial system start-up led to improved treatment performance by providing a linkage between inflow volume and conductivity (through sulfate-conductivity correlations established over the period of operation), whereby sulfate loading could be automatically calculated and reagent dosing could be adjusted in real time without the need for manual adjustments by a dedicated operator.

Substantial additional benefits can be realized through application of predictive machine learning tools that allow anticipation of variables such as the increased flow and load as observed in the case study example. One such univariate tool was described by Do et al. (2022), where they utilized a seasonal autoregressive integrated moving average (SARIMA) forecasting model to predict wastewater inflow at a wastewater treatment plant. Azad et al. (2022) identified that a SARIMA model fused with an artificial neural network model (ANN) outperformed the SARIMA model alone for predicting water levels within a reservoir. Other variables that may incur delays in the system that exceed the PLC feedback loop and therefore benefit from predictive automation could include microbial growth and oxidation reactions (i.e. kinetic variables).

Conclusions

The implementation of modern sensor technology in water treatment systems, when coupled with edge computing, data upload to the cloud via telemetry, and cloud-based data management, analysis, and decision-making can provide operators with new opportunities to optimize the full spectrum of passive to active water treatment systems. This workflow, when fully automated can be deemed a Digital Twin, which can be defined as a digital replica (the cloud-based data management workflow for remote sensor data) of a physical asset (the water treatment system) for purposes of remote, automated management and optimization of the asset. While a full-scale digital twin should be considered aspirational for certain water treatment systems as its development is highly dependent on available sensor technology, and the history, volume and types of data that have been collected, it represents the future of water treatment systems management. Once a history of data has been collected, digital twins can be used to forward predict operating conditions, which is useful for anticipating changes to influent water chemistry and volume, potential compliance violations, and systems failure. When such issues are predicted in advance, they can be managed proactively.

References

Azad, A.S., R. Sokkalingam, H. Daud, S.K. Adhikary, H. Khurshid, S.M.A. Mazlan, M.B.A. Rabbani (2022) Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study. Sustainability 2022, 14, 1843. DOI: 10.3390/su14031843

Do, P., C.W.K. Chow, R. Rameezdeen, N. Gorjian (2022) Wastewater inflow time series forecasting at low temporal resolution using SARIMA model: a case study in South Australia. Environmental Science and Pollution Research, Vol 29, pg 70987-70999. DOI: 10.1007/s11356-022-20777-y

Peng, L., G. Han, X. Sui, A.L. Pagou, L. Zhu, J. Shu (2021) Predictive Approach to Perform Fault Detection in Electrical Submersible Pump Systems. ACS Omega, Vol 6, pg 8104–8111. DOI: 10.1021/acsomega.0c05808

Yang, P., J. Chen, H. Zhang, S. Li (2022) A Fault Identification Method for Electric Submersible Pumps Based on DAE-SVM. Hindawi, Shock and Vibration, Volume 2022. Article ID 5868630, 16 pages. DOI: 10.1155/2022/5868630