INFORMED AI IN THE WATER SECTOR – THE ROADMAP FROM POSSIBLE TO PLAUSIBLE

Silvia Vlad, Jacobs

ABSTRACT

As water suppliers throughout New Zealand tackle the twin challenges of climate change and limited investment budgets, artificial intelligence (AI) and machine learning applications can help us achieve both sustainability and efficiency goals, giving utilities the insights needed to streamline and optimise operations and capital investments. While the conceptual promise of AI is often lauded, many utilities find that real-world considerations such as historical data quality and quantity can interpose significant limitations on the value of AI applications.

This paper will discuss key factors in identifying viable machine learning applications, highlighting data requirements and methods for supplementing real-world historical data, as well as outlining opportunities for tiered AI implementations that grow with an organisation as it gains digital maturity. Starting from available data and planning out future data collection, machine learning applications can increase in sophistication and value as additional datasets become available.

By further augmenting our existing datasets with synthetic data, "informed" AI allows machine learning models to overcome the challenges of narrow historical datasets, which frequently capture only a small range of operating conditions. Leveraging both scenario analysis and operational forecasting capabilities, AI applications are giving utilities an ever-expanding set of insights into chemical and energy optimisation, greenhouse gas, and cost-reduction opportunities. Key case studies from around the world will showcase how global best practices in leveraging machine learning can be used to help New Zealand utilities achieve sustainability and cost-savings goals.

KEYWORDS

Artificial intelligence, optimisation, machine learning, data mining, scenario analysis

PRESENTER PROFILE

Silvia Vlad, MASc, P.Eng.

Silvia is a Jacobs process engineer with experience delivering water treatment design, optimisation and planning projects for utilities spanning Canada, the United States, Asia and New Zealand. She works collaboratively with Jacobs' global Digital team, to advance the use of digital tools in delivering water and wastewater infrastructure projects.

INTRODUCTION – SMART UTILITIES

The “smart utility” landscape is rapidly evolving along with our understanding of the array of tools, techniques, benefits and pitfalls of digitalisation across our business. Referred to interchangeably as a “smart water grid,” “digital water utility,” “intelligent water utility,” “intelligent water system” and “data-driven water utility,” the label of “smart utility” carries with it an implication of overlayed data collection, information creation and insight extraction to inform decision making (Water Research Foundation Project *[WRF]* 4714, 2020). Critically, the informed decision making at a smart utility targets and realises better business outcomes in alignment with the utility’s goals, which typically including high service reliability, cost efficiency, energy efficiency and carbon footprint minimisation.

The New Zealand water industry is squarely situated in the midst of numerous drivers to reduce greenhouse gas emissions, curtail energy consumption and deliver high-caliber three-waters services at an affordable cost, and as such is well positioned to see the potential benefits of digitalisation. The global Covid 19 pandemic and recent droughts, storms and extreme events have also fueled a resilience focus for the water industry, with an emphasis on business and operational continuity, cyber protection and data accountability and accessibility.

In many utilities, the adoption of digital tools and “smart” practices is significantly higher during planning and design phases of a (physical) asset lifecycle, and conversely, lowest during the operational and maintenance phase (WRF Project 4836, 2017). Furthermore, data flow across the lifecycle of our physical assets is often suboptimal with respect to feeding forward and feeding back information from different stages of the asset lifecycle.

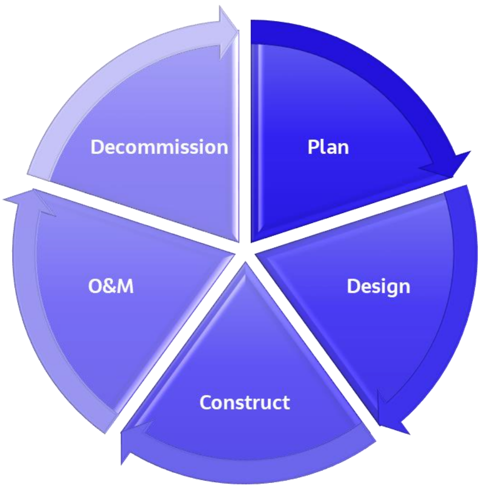

Figure 1: Digital adoption is typically lower in the later phases of the asset lifecycle

For example, questions from the design stage of a new physical asset could include:

- What kinds of pumps are currently in operation across our system?
- Which pump type has the lowest maintenance requirements, in our historic experience?
- How much down time is associated with pump maintenance?
- What are the cost implications of selecting exclusively high-quality, low-maintenance pumps in our future designs?

A utility looking to answer these questions may not have easy access to query the appropriate data. While the necessary information may be captured and available somewhere within the organisation, that data loses significant value by not being accessible to the right people at the right time. There is increasing acknowledgement of the need to have a digital thread that spans the asset lifecycle, with digital artifacts created at each stage that can be leveraged both up- and downstream.

A survey of global water utilities found that while immense volumes of data are being collected, only an estimated 10% of data collected is actually used (WRF Project 4836, 2017), and in general the digital maturity of water utilities Two pivotal issues hamper a utility's ability to leverage data and take insight-based action: data quality and a (perceived) lack of talent in the utility to plan, deploy, integrate and manage the digital practices required.

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Amidst the backdrop of the "smart utility" landscape, artificial intelligence describes a set of tools which can help us interpret both stockpiled and live data, and can form part of the overall smart utility approach. A subset of artificial intelligence, machine learning uses computational power to "identify patterns, make decisions and improve themselves through experience and data" (Columbia University, 2022). Figure 2 provides an overview of three key learning methods: supervised, unsupervised and reinforcement, and their typical applications in the water industry.

Supervised learning is sometimes referred to as "learning by example." It uses data that have paired inputs with "correct" or "desired" outputs . These data sets are referred to as "labeled data." During training, patterns are created that correlate the inputs with the desired outputs within the machine learning model. Selecting appropriate "features" (parameters or system attributes) to include in a supervised learning training dataset can be a critical step, which sometimes requires an additional step for preliminary data investigation.

Unsupervised learning refers to the identification of patterns in data sets containing data points that are neither classified nor labeled. Unsupervised learning is uses machine learning algorithms to explore relationships between data, and can be useful in discovering how different parameters are related in complex systems. This type of learning can be used to help select the most meaningful "features" to use in follow-on learning.

Reinforcement learning uses a trial and error approach to establish a relationship model. The model is trained by rewarding correct or desireable outcomes and penalising less desireable outcomes. It icould be hought of as similar to training a pet using treats and punishment.

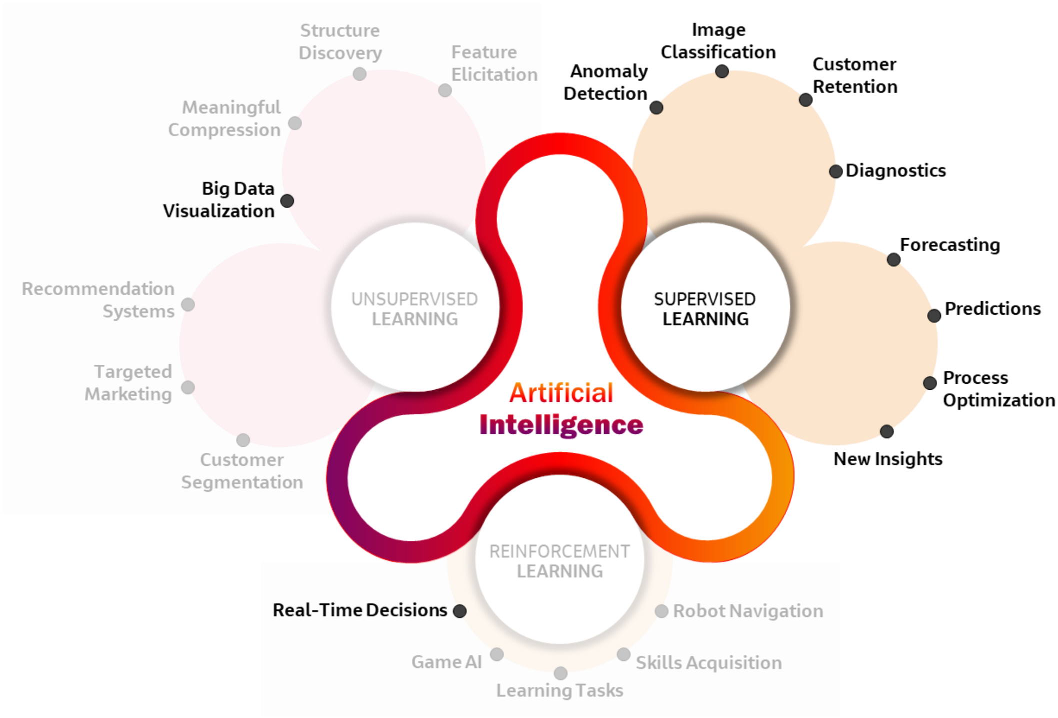

Figure 2: Artificial intelligence learning methods and typical applications in the water industry

MACHINE LEARNING ADOPTION

The value of data deminishes with time, and data is most valuable with context. Machine learning is typically used because it can enable us to *quickly* analyse large volumes of data, and provide timely input for descisions, based on historical data and outcomes rather than rules, equations or anecdotes. It is typically most useful in modelling complex systems, where a mechanistic or theoretical model may struggle to capture all the variability in outcomes. As highlighted in Figure 2, many water industry applications of machine learning use supervised learning approaches, which critically rely on appropraite feature selection, and on labelled datasets.

Utiltiy goals for implementing machine learning applications typically fall into two major goals:

- Optimising operations (be it for better treatment outcomes, lower climate impact, lower chemical and energy costs, or a combination of optimisation targets)
- Predicting system failure (to support operational resiliency and capital/maintenance planning)

Several example applications for machine learning are outlined in Table 1, with the location of key case studies shown in Figure 3.

Table 1: Machine learning applications and example implementations

Machine Learning Goal	Application	Example Implementation(s)
Treatment optimisation	Identify chemical and energy savings, improve treatment performance	Treatment Optimisation – Melbourne Winneke Water Treatment Plant (WTP), Woodland Davis WTP, Singapore Public Utilities Board CCK WTP
Image classification	Condition assessment image coding	Dragonfly - Jacobs-Hitachi partnership, using machine learning to complete sewer condition assessment scoring of CCTV inspection footage
Equipment performance optimisation	Identify equipment performance optimisation opportunities	Pipe Failure Prediction – Yarra Valley Water AquaDNA: UK Pump and Equipment Health monitoring
New insights	Discover hidden value in online or grab sample data	Forest Fire water quality impacts – Eugene Water and Electric
Forecasting	Predict raw and treated water quality based on historical and on-line information	Influent/Raw Water Quality Changes and cyanobacterial taste and odour compound prediction – Clayton County, Georgia Demand/Production Forecasting – Silverdale, Washington
Anomaly detection	Instrument Malfunction, Maintenance Monitoring, Contaminant Warning	Contaminant Warning Systems – New York, Philadelphia, Dallas, San Francisco, Cincinnati, Glendale

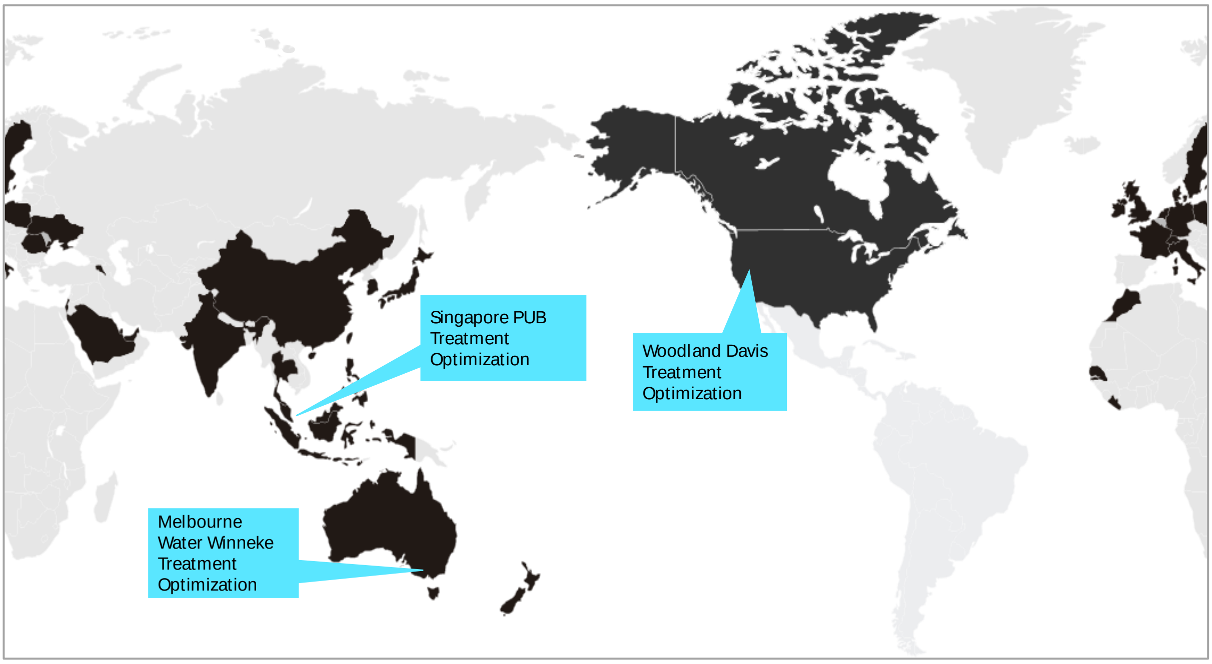

Figure 3: locations of treatment optimisation machine learning case studies

In drinking water treatment, coagulation control is a prime candidate for machine learning based on several factors:

- There is no simple coagulation "equation"
- Long detention times make it difficult to react to changes in coagulation *outcomes*
- Coagulation decisions are frequently made based on experience, intuition, and trial and error
- Existing practices (e.g. regular jar testing) are manual and labour intensive
- Optimisation can improve water quality, enhance chemical and energy efficiency, and lower costs for chemicals, backwash pumping and residuals management

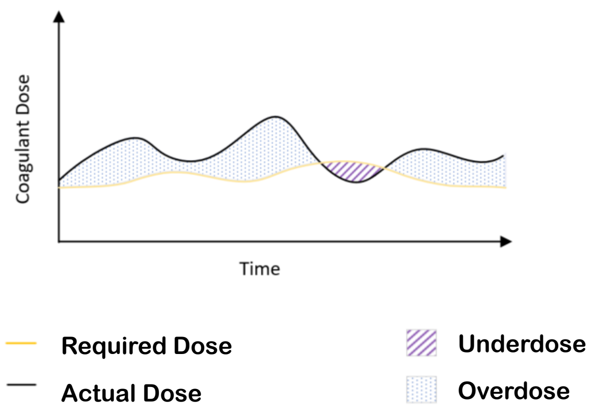

Figure 4: coagulation control optimisation opportunity

Chronic over- or under-dosing is difficult to combat while maintaining plant operations within acceptable treated water quality parameters, creating an optimisation opportunity if treated water outcomes can be accurately predicted in real time (see Figure 4).

To this end, machine learning approaches are being applied for coagulant dosing optimisation at the Winneke WTP in Melbourne, Australia, the CCK WTP in Singapore and the Woodland Davis WTP in the United States, as shown in Figure 3.

WOODLAND DAVIS WTP CASE STUDY

The objective of machine learning application at the Woodland Davis WTP were two-fold:

1. Develop a machine learning model to optimise chemical usage
2. Create a dashboard display recommended changes in real time

Figure 5: Woodland Davis WTP

The treatment process is shown in Figure 6, with the model input monitoring data highlighted. The optimisation target for the machine learning model was based on a combination of three factors:

1. Minimise the chemical cost
2. Minimise settled water turbidity
3. Minimise the frequency of changes to the coagulant dosing setpoint

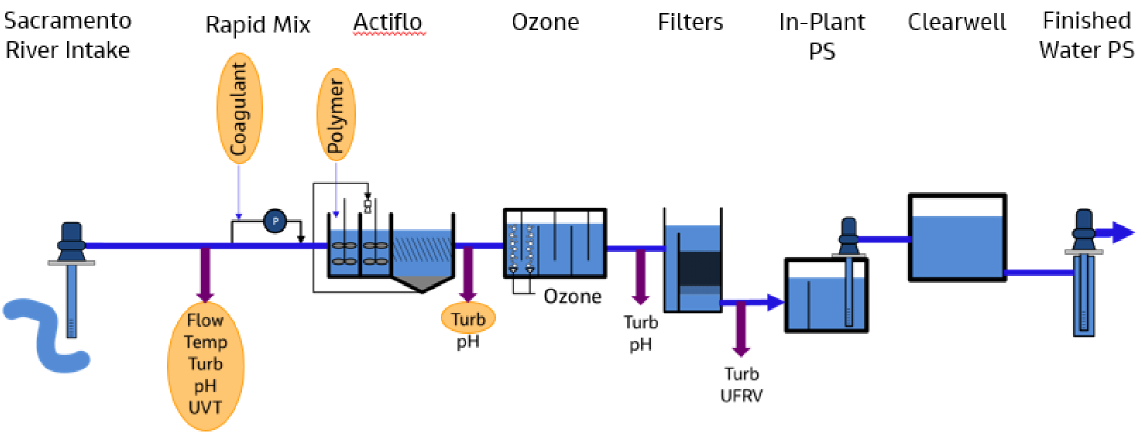

Figure 6: Woodland Davis WTP treatment train and water quality monitoring

The first step in model development was to clean and review the available data, identifying any erroneous, blank or invalid data, then selecting the parameters relevant to the optimisation target. While a significant volume of data was available through the plant SCADA, not all of it was material to the prediction of settled water turbidity. The Woodland Davis model was developed based on the available historical plant data from 2017 through 2019 taken at a 15-minute increments. 80% of the data was used to train the model, while the remaining 20% was used for model validation, as shown in Figure 7.

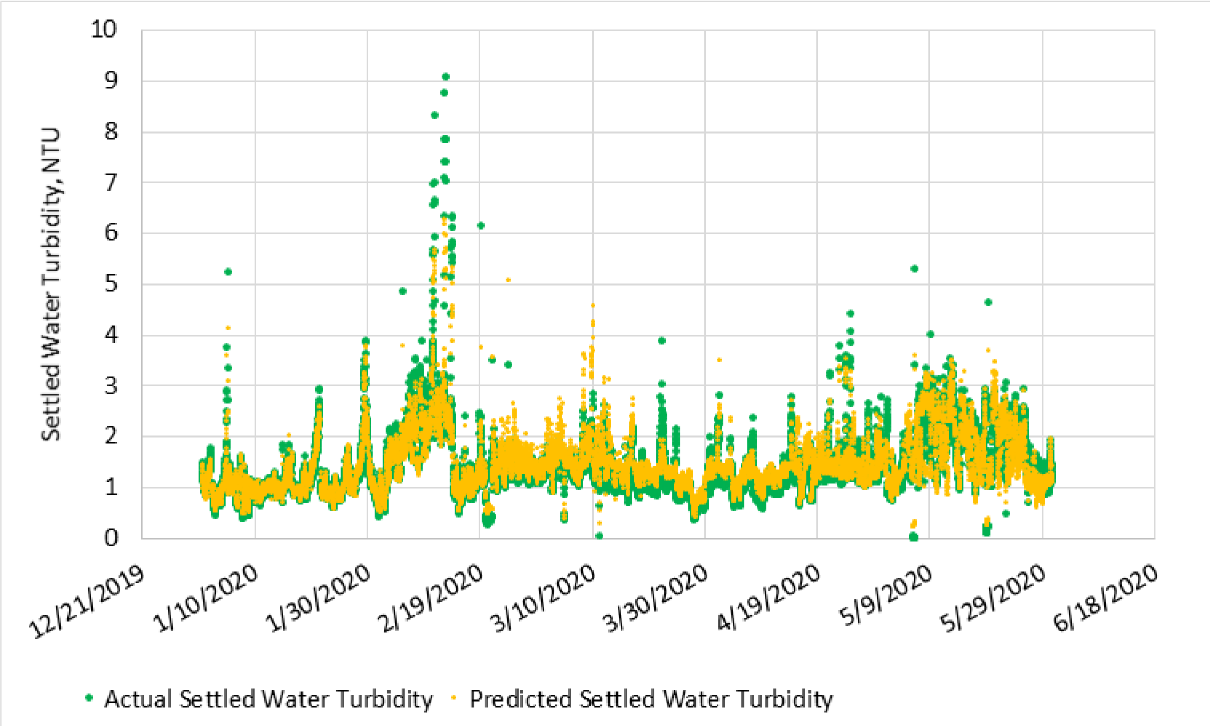

Figure 7: Trained model projections of settled water turbidity

Based on the trained model, a recommended coagulant dose was projected by the model to meet settled water turbidity target (as shown in Figure 8) with the recommendation and financial

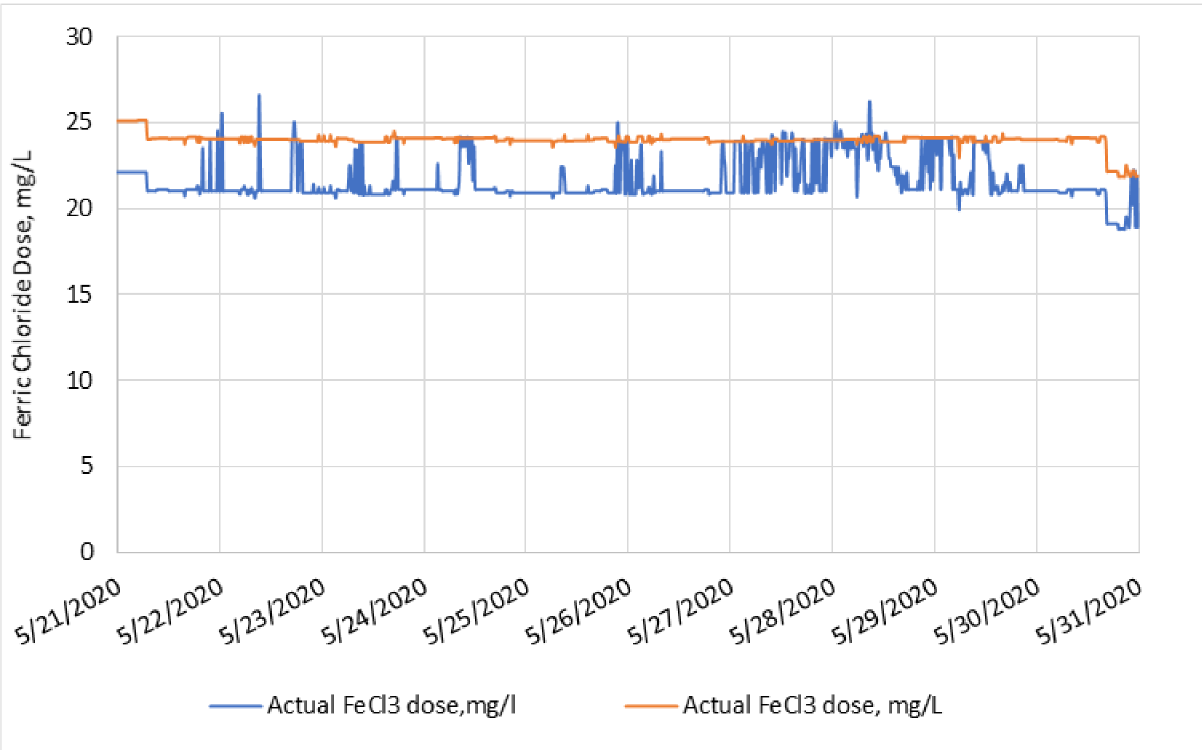

Figure 8: Trained model projections of settled water turbidity

implications displayed on a dashboard for operators' consideration in making dosing decisions (Figure 9). A 9% reduction in overdosing was projected on average, indicating general overdosing historically, and cumulative annual cost savings of $54,000-$72,000 were projected, depending on the settled water turbidity target.

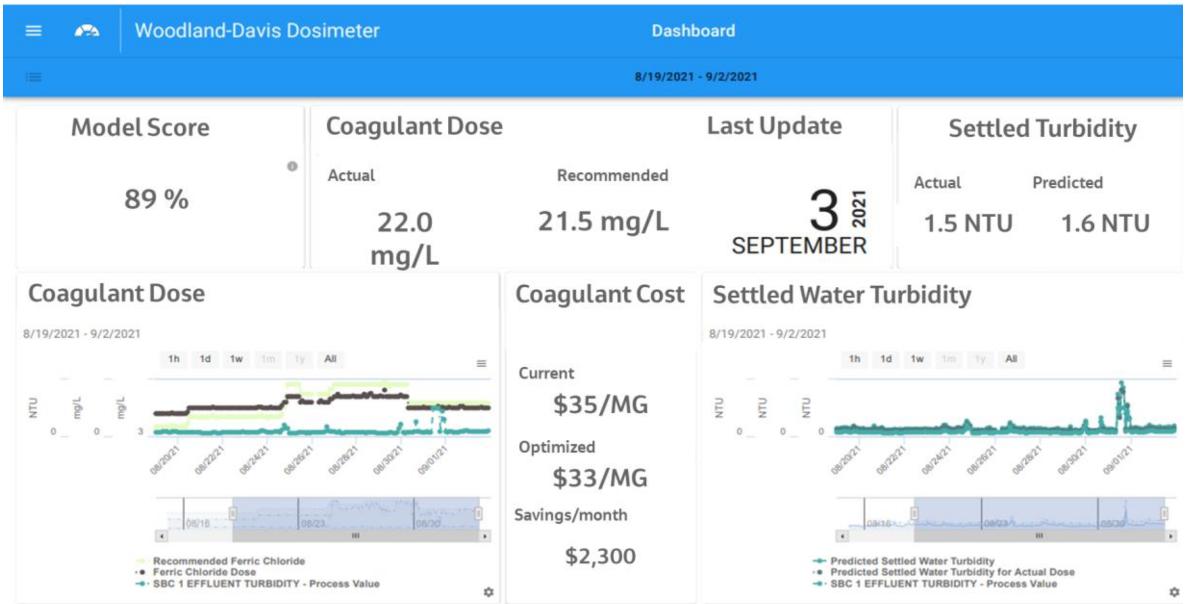

Figure 9: model output dashboard

HYBRID MACHINE LEARNING

As noted above, candidate applications for machine learning are often focused on areas where mechanistic understanding doesn't capture the true system complexity and variability. While machine learning algorithms can help us work with these complex systems on the basis of empirical outcomes, the black-box nature of these algorithms can make it challenging to interpret their outputs in the context of our theoretical understanding. In some cases, more advanced machine learning applications can also build on our mechanistic understanding, leveraging the benefits of both approaches through a hybrid model, as outlined in Table 2.

Table 2: Machine learning, mechanistic and hybrid modelling tradeoffs

	Pure Machine Learning	**Mechanistic Models**	**Hybrid**
Benefits	- Predictive abilities for systems with unclear relationships - Fast	- Well-understood by industry - Transparent reasons for actions	- Combines the strengths of both approaches
Detractions	- Does not handle uncommon situations well - "Black Box" does not provide understanding for model users	- Inherent uncertainty of water/wastewater processes invalidates discrete answers	- Limited full-scale experience

		- Requires frequent maintenance to keep relevance /accuracy	

As an example application, a hybrid approach could allow a model to manage influent unknowns and variability of biological and chemical reactions to different stressors in a wastewater treatment context. This is the topic of a current Water Research Foundation Project (#5121), investigating Innovative Predictive Control Strategies for Nutrient Removal. This work is considering multiple applications of a proposed hybrid controller at a range of wastewater facilities, blending machine learning and process mechanistic modelling for:

- Enhanced biological phosphorus removal (EBPR) stability & prediction of metal salt addition requirements for Total Phosphorus (TP) compliance.
- Enhanced nitrogen removal (ENR) performance
- Energy reduction

The project is targeting a demonstration of short-term optimisation functions and longer-term (>10 days) predictive capabilities of potential issues related to phosphorus and nitrogen management, and is part of industry-leading efforts to advance use of hybrid modelling in a smart utility environment.

INFORMED AI

While hybrid modelling addresses on potential issue of machine learning – namely, the black-box nature of the model itself – it also presents an opportunity to enhance the training datasets used in developing the machine learning models through the production of “synthetic data” from a mechanistic model.

Generally, data captured in water and wastewater operations lacks “persistency of excitation” – i.e. it is fairly stable withing typical operating windows and does *not* explore the system limitations. By nature of the reliability of our systems, and the need to provide continuous service to the public, it is generally undesirable and often impractical to obtain real data from a live system operating under “extreme” conditions. However, persistency of excitation is necessary to identify the underlying dynamics of complex systems – if inputs are consistently static, the machine learning algorithms cannot learn much form the complex relationships between different factors.

A classical example of an informed artificial intelligence approach is in the training of autonomous vehicles. While a self-driving car may have driven millions of kilometers on real roads to train a driving algorithm, they are typically trained on billions of kilometers of “simulated” roadways as well to teach the algorithm to manage atypical but often vitally important circumstances.

In a water utility context, mechanistic models – for instance, hydraulic and/or process digital twins – give us the means to produce model-training data corresponding to different scenarios, which *could* happen in the real world but may not have been captured in the historical data. The use of this synthetic data produces an “informed” AI model.

By further augmenting our existing datasets with synthetic data, “informed” AI allows machine learning models to overcome the challenges of narrow historical datasets, which frequently capture only a small range of operating conditions.

DIGITAL ASSET PLANNING

Machine learning, hybrid models and informed artificial intelligence present significant opportunities to make use of water utility data. However, it should be noted that significant investment is often needed to realise the full value proposition of digital assets, including machine learning. Like a physical asset, these digital entities have a life cycle of their own, with planning, design, construction, operational and maintenance needs to be considered as part of their implementation. And like our physical assets, they do not stand alone must be integrated with existing assets, and designed to accommodate future expansions and changing needs.

WRF project 4714 (2020) developed a digital maturity model to benchmark utility and industry status, and highlighted 7 key characteristics of a “Smart Utility:”

- Strategy & Vision
- Data Management
- Integration & Interoperability
- Analytics & Information Use
- Risk & Resiliency
- Workforce
- Asset Management

Given the rapid evolution of digital platforms, and our changing understanding of data needs, there can be an understandable impulse to “wait and see” rather than investing in digital assets that may not be compatible with our future systems. However, by developing a strategy and vision for long-term organisational digitsation, utilities can look to establish their smart water systems incrementally, realising short term benefits while building towards greater combined benefits once multiple tools are in place.

In the artificial intelligence context, a pure machine learning application could be developed, providing immediate benefits in optimisation. A stand-alone mechanistic digital twin model could similarly provide short term benefits for scenario testing and operator training. If each is developed incorporating interoperability considerations, these discrete tools can later be linked to provide hybrid modelling and synthetic data generation, providing even greater benefits as a combined system. This staged approach allows utilities to adopt, test and prove out each tool without the need to develop a full combined system before deriving any operational benefits.

CONCLUSIONS

Leveraging both scenario analysis and operational forecasting capabilities, artificial intelligence applications are giving utilities an ever-expanding set of opportunities for chemical and energy optimisation, greenhouse gas, and cost-reduction opportunities.

By pairing pure machine learning with a hybrid mechanistic modelling approach, we can peer inside the black box to glean further insights into the relationships and parameters influencing treatment outcomes. We can also further leverage mechanistic models to augmenting our existing datasets with synthetic data. In this way, an informed artificial intelligence approach allows machine learning models to overcome the challenges of narrow historical datasets, which typically capture only a small range of operating conditions.

By designing our data systems and digital assets with long term integration and interoperability in mind, we can make staged investments to build new tools, such as machine learning applications for short-term insights, then obtain incremental value when pairing them with mechanistic digital twins down the road.

ACKNOWLEDGEMENTS

Raja Kadiyala, Jacobs Digital Market Director

Kim Ervin, Jacobs Global Technology Leader for Applied Digital Tools in Drinking Water and Reuse

REFERENCES

Water Research Foundation Project 4836, 2017. *Leveraging Other Industries - Big Data Management Phase I*.

Water Research Foundation Project 4714, 2020. *Intelligent Water Systems: Digital Maturity Model*

Water Research Foundation Project 4978, 2021. *Application of Big Data for Energy Management in Water Utilities*

Columbia University, Fu Foundation School of Engineering and Applied Science. *Artificial Intelligence (AI) vs. Machine Learning* <https://ai.engineering.columbia.edu/ai-vs-machine-learning/#:~:text=Put%20in%20context%2C%20artificial%20intelligence,and%20improve%20themselves%20through%20experience>, accessed 16 August, 2022