Citation: Bertone, E., Stewart, R.A., Zhang, H., and O'Halloran, K. (2016) Hybrid water treatment cost prediction model for raw water intake optimization. *Environmental Modelling & Software*, Volume 75, January 2016, Pages 230-242, <http://dx.doi.org/10.1016/j.envsoft.2015.10.020>

Hybrid water treatment cost prediction model for raw water intake optimization

Edoardo Bertone¹, Rodney A. Stewart^{1*}, Hong Zhang¹ and Kelvin O'Halloran²

¹Griffith School of Engineering, Griffith University, Gold Coast Campus, QLD 4222, Australia.

²Seqwater, Ipswich, QLD 4305, Australia

*Corresponding author:

Email: r.stewart@griffith.edu.au

Phone: +61 7 555 28778

Highlights:

- Prediction model built to estimate treatment costs for a dual source drinking WTP
- Model optimises source selection proportions based on treatment and pumping costs
- Model aids WTP plant operators to alter the source selection strategy in extreme events
- Model utilisation for source selection optimization delivers life cycle monetary savings

Abstract

In order to reduce the total cost of a dual source drinking water treatment plant operation, a comprehensive hybrid prediction model was built to estimate the necessary chemicals dosage and pumping energy costs for alternative source selection scenarios. Correlations between the water quality parameters and the required treatment chemicals were estimated using historical data and the expected pH variations associated with each chemical addition, which was based on the Caldwell-Lawrence diagram. The pumping energy costs were also estimated using a data-driven approach that was based on historical plant data. The research has practical implications for water treatment operators seeking to minimize plant operational costs through selecting raw water intake volumes for their treatment plant based on multiple source options and offtake tower gate levels. Future research seeks to better link current and future water treatment dosage cost predictions directly to water quality measurements taken from vertical profiling systems.

Keywords

Water treatment plant; water treatment optimization; decision support system; data-driven modelling.

1. Introduction

1.1. Water treatment process

The delivery of safe drinking water (i.e. without harmful chemicals or waterborne pathogens) is an essential task for any bulk water supplier. The treatment of raw water from lakes, rivers, or wells is, therefore, required in order to meet the drinkability criteria defined by different national regulators.

In a conventional water treatment plant (WTP), the process of treating water usually consists of the following five steps: (1) raw water is adjusted for alkalinity and pH with the addition of hydrated lime and carbon dioxide; (2) particulate matter congregates together with the addition of aluminum sulphate and other coagulants such as polymers and then the water flows over a cascade that mixes chemicals and raw water with the coagulants; (3) water is slowly mixed in the clarifiers where larger particles settle down to the bottom and are periodically removed (sedimentation); (4) water is directed from the clarifiers to the filters (e.g. anthracite and sand filter) in order to entrap any smaller particles that survived the clarification process; and (5) sodium hydroxide is added to adjust the final pH/alkalinity, sodium hypochlorite for disinfection and fluoride for fluoridation (Sarai, 2006). Sometimes, as an alternative to sedimentation, dissolved air flotation can be effectively used in those WTP receiving waters from lakes that overturns once/twice each year leading to algae blooms, or taste and odors problems (Kawamura, 2000).

Estimating the monetary cost associated with water treatment is fundamental to a practical planning approach for potable water supply (Abdullahi, 2013). The operation of a WTP, nevertheless, is significantly different from most other industrial operations, as raw water quality constantly changes according to the season, wet weather events, or anthropogenic activities in the catchment. The water treatment cost is clearly related to the amount of the chemical dosage needed to adequately improve the sourced water quality, however the creation of accurate cost estimate predictions is challenging due to the variance in water quality parameters in the raw water. (Abdullahi, 2013). Therefore, although it is clear that an accurate algorithm is a prerequisite to predicting the chemical dosage for optimum treatment, it must be based on water quality data that often does not exist for most WTP's (Mirsepassi, 2004). Typically, chemical dosages are estimated with jar tests, which are expensive and

time-consuming (Maier et al., 2004). Moreover, jar tests are not ideal for handling sudden changes in water quality that can occur, which require a prompt adjustment of chemical dosing. However, recent advancements in the field of environmental monitoring technologies such as vertical profiling systems (Rouen et al., 2005), or data storage and analytics, greatly enhanced the potential for the creation of decision support systems for WTP operators seeking to make urgent water treatment decisions.

1.2. Review of reported optimization models

In recent years, a number of optimization models have been developed for different stages of water and wastewater treatment processes. Mathematical data-driven models, in particular, appear to be more prevalent in the literature than process-based models. Their main virtues include their simplicity, relative precision and cost-effectiveness. Additionally, data-driven simulations tend to be faster than process-based (e.g. hydraulic) models, which often require outputs from multiple component models to be combined in order to provide useful predictions, making them less feasible for treatment optimization (Regneri et al., 2010). On the other hand, mathematical models can be effective in situations where a solution is needed urgently (e.g. because of an outbreak of a waterborne contagious disease), since the sample analysis procedure can be time consuming (Al-Ali et al., 2011). In general, data mining techniques have been found to be promising for modeling industrial applications (Kusiak and Wei, 2011). In relation to the water industry, Savic et al. (1999) provided comprehensive review of data mining and analytics techniques applied to urban and surface water engineering problems. More than a decade has passed since their study, and a significant amount of new research has been completed that has improved these existing techniques and also added a range of novel water treatment management applications.

A good part of previous research has focused on Wastewater Treatment Plant (WWTP) optimization and costs reduction rather than on the potable water treatment processes improvement. For WWTP applications, there are a few examples of process-based models. For example, McCorquodale et al. (2005) used physical modeling to investigate the optimization of the hydraulic conditions of a high-purity oxygen-activated sludge. Also, Seggelke et al. (2005) optimized the dynamic control of a treatment plant inflow by using an online simulator running parallel to the real WWTP operation; this provided model information to a prognosis tool which determines the best inflow options.

Interestingly, there are many examples of formulated data-driven models that have surfaced in recent years. For instance, Baruch et al. (2005) built three adaptive neural network control structures to regulate a biological wastewater treatment process. Another neural network-based control system was developed by Lee et al. (2005) to efficiently operate small plants with significant variance in the influent loadings. They used an internet-based remote monitoring system that input oxidation-reduction potential (ORP) as the main sensor for the control. Interestingly, Gillot et al. (1999) created a novel economic index that considered both WWTP fixed and variable costs in order to compare and decide on potential lowest cost treatment scenarios for a WWTP. The integration of variable costs was deemed crucial for more clearly identifying treatment options that could deliver cost savings. Both Bozkurt et al., (2015) and Hakanen et al. (2013) developed similar optimization models that could be used for multi-objective WWTP design and operation. Al-Ali et al. (2011) provides another example of a data-driven statistical model that could correlate biogeochemical and chemical oxygen demand with total suspended solids and other anions of the wastewater samplings from a drug factory. Beyond WWTP, Alcoea et al. (2009) developed an optimization tool for desalination plants that attempted to achieve the multiple objectives of reduced environmental impacts and minimized the operational costs (Alcoea et al., 2009).

For potable water systems, Chen et al. (2014) recently built an interesting hybrid model that optimized the source selection proportions of a water resource system that included both surface and groundwater. The groundwater flow was simulated with a physical model and incorporated into an artificial neural network to run the optimization. One of the first WTP optimization models was attempted by Baruch et al. (2005) for an Iranian plant; however, the optimization tool covered turbidity and total organic carbon removal only, with the chemicals optimization being conducted by jar tests. Similarly, a Decision Support System (DSS) was built by Slavik et al. (2010) to assess different reservoir raw water management strategies. Even though the model is very comprehensive, and also considers such aspects as flood risks, the water quality (and thus treatment cost) is assessed by means of organic load and turbidity parameters only. The model does not consider the extra treatment costs associated with the day-to-day variations of other relevant water quality parameters (e.g. pH, alkalinity, or peaks in manganese). Interestingly, Rietveld et al. (2010) developed a number of optimization models for a drinking WTP aimed to improve the operation of treatment sub-processes and, thus, reduce the costs.

In the Rothberg, Tamburini and Winsor (RTW) model, more details are provided to support water operators (Rothberg et al., 1993; RTW, 1996). This model is a spreadsheet-based tool designed to help users assess the effects of chemical additions on the stability of water, and to predict changes in the water quality parameters, such as pH, alkalinity, or calcium carbonate precipitation potential. The RTW model is often used by water engineers to develop corrosion control strategies, optimize coagulation, determine pH impacts on the precipitation of metals, and evaluate the chemical dosage options and their economics.

A quite relevant body of work, in terms of chemical dosages and treatment cost prediction modeling was undertaken by Abdullahi and his colleagues. Firstly, he determined the amount of alum (Abdullahi and Odigure, 2006), then the amount of chlorine (Abdullahi and Abdulkarim, 2010) and, finally, the amount of lime (Abdullahi et al., 2012) required for water treatment, using mathematical models and the existing interrelationships between the parameters. It must be acknowledged that an alum prediction model, based on historical jar tests, had already been built by van Leeuwen et al. (2001), while Artificial Neural Networks modeling was used by Maier et al. (2004) to predict optimal alum doses and, thus, potentially avoid the jar tests.

Abdullahi (2013) was able to put all the models together and, by using a new, integrated data-mining approach, he was able to estimate the WTP operational costs. The new model covers energy, administration, maintenance, and chemical costs. However, it should be noted that the only modeled chemicals are lime, alum, and chlorine using the findings of his previous studies in 2006, 2010 and 2012.

In the Mudgeeraba WTP, location of this study, more chemicals are used (e.g. carbon dioxide and sodium hydroxide) which also cause pH variations during the process, and must be taken into account in order to properly adjust the final pH of the treated water. Also, one of the two water sources that can be drawn from is derived from a reservoir that has a lower elevation than the WTP, while the second reservoir provides gravity fed raw water to the WTP. Therefore, for this study, the cost of all the estimated chemicals as well as the pumping costs must be calculated for each water reservoir source selection scenario.

Due to the amount of historical water treatment data available to the researchers and the literature presenting the successful development of several data-driven models for similar applications, the authors embarked on this study to apply data-mining and analytics techniques coupled with some existing models for pH variation prediction in order to develop a comprehensive WTP cost prediction tool. The proposed hybrid methodology has elements of novelty since it takes advantage of data-driven approaches whenever historic data reveals good prediction potential for certain variables, and at the same time by mathematically estimating pH/alkalinity changes at different stages of the treatment, it is possible to reliably predict those chemicals (e.g. lime, carbon dioxide, sodium hydroxide) used mainly to reach a targeted pH and alkalinity in the finished water. Such hybrid approach, to the authors' knowledge, has not been applied before; likewise, a comprehensive prediction tool of all the main chemicals and pumping costs based on real-time lake water quality has not yet been developed. The tool should allow WTP operators to estimate the optimum blend from two different raw water sources in order to minimize the total WTP operational costs. The model development process has some inherent risks since data-driven models are, by definition, not mechanistic. Consequently, in some circumstances, the models might not describe the plant processes appropriately. Nonetheless, if a sufficient level of prediction accuracy can be reached, they can provide a valuable support tool, along with additional information for optimal plant control (Dürrenmatt and Gujer, 2012).

2. Methods

2.1. Research domain and data collection

The Mudgeeraba WTP (Figure 1) is the second largest drinking water treatment facility in the Gold Coast region, Queensland, Australia, after the Molendinar WTP; in its current state, it can treat a maximum of 110 ML/day of raw water (Rogers et al., 2008). The Mudgeeraba WTP has two conventional clarifiers and sixteen mono media sand filters.

Raw water is redirected to the Mudgeeraba WTP from two reservoirs (Hinze Dam and Little Nerang Dam), which are located about 3 and 8 kilometres west and south-west of Mudgeeraba, respectively (Figure 2). Hinze dam, the main reservoir servicing Gold Coast City (maximum storage capacity: 310,700 ML), has two intake structures: the lower intake tower (which redirects water to the Molendinar WTP), and the upper intake tower (which is connected to the Mudgeeraba WTP). The Hinze dam Upper Intake tower (HUI) is located 4.5

kilometres upstream of the main embankment of Hinze Dam, on the Little Nerang Creek arm of the impoundment; it consists of a 58 metre high dry well reinforced concrete tower, housing nine 900 mm diameter inlet pipes at about 5 metre intervals. Three electric pumps transfer up to 87.4 ML/day of water to the header tank, where the water then flows by gravity to the Mudgeeraba WTP, at a rate of up to 74 ML/day (DERM, 2010). The Little Nerang Dam (LND), on the other hand, is a much smaller reservoir (maximum storage capacity: 8,400 ML), and is usually kept as a backup for emergency situations (such as power outages, and the inability to draw water from the HUI due to water quality issues). The LND's only intake tower is made of reinforced concrete; it houses five screened 450 mm diameter inlet pipes at 5-7 metre intervals. An 850 mm diameter gravity raw water main that is 7.845 km long transports the raw water to the Mudgeeraba WTP (DERM, 2010). In the case of this study, the geographical location of the two water sources (i.e. one requiring pumping and the other gravity), maximum capacity (i.e. one small and prone to rapid depletion without inflow and one large), and water quality differences at various times, means that a robust source selection tool can potentially derive source selection options that can yield significant operational cost savings when compared to traditional historical practices.

Figure 1: Aerial photo of Mudgeeraba WTP (Rogers et al., 2008)

Figure 2: Locations of LND, HUI and Mudgeeraba WTP (see green dot)

Seqwater, the main bulk water supplier in the South-East Queensland region, and custodian of the Mudgeeraba WTP, provided historical water quality datasets for the current study. The data from the two reservoirs included manual weekly samplings from 2008 to 2014. The parameters measured included: water temperature (T_w), pH, turbidity (Tb), color, manganese (Mn), and dissolved oxygen (DO). Additionally, the data from the recently installed Vertical Profiling System (VPS) was provided. A VPS consists of a buoy with a probe underneath which is automatically winched up and down and collects, with a 3-hour frequency and 1-metre interval, vertical profiles of key-parameters such as water temperature, dissolved oxygen, conductivity, turbidity, pH, and redox potential. The profile data can be accessed in real-time from the Seqwater server, enabling a real-time detection of sudden drastic changes in some parameters. Moreover, daily data for the Mudgeeraba WTP, for the period 2010-2014, were also available. From a water quality point of view, the measured parameters were: pH, alkalinity, T_w , Tb , color, and Mn . From a chemical usage perspective, the following dosages were recorded on a daily basis: aluminum sulphate, slaked lime, carbon dioxide,

sodium hydroxide, powdered activated carbon (PAC), sodium hypochlorite, potassium permanganate, fluoride, and the coagulant polydiallyldimethylammonium chloride (shortened polydadmac). The monthly energy usage costs for the Mudgeeraba WTP and associated pump HUI station were also provided. Other available data included: the daily flow from each reservoir, the dam draw-off levels, and the chemicals costs.

2.2. Data analysis

In Figure 3, the total monthly raw water flows from the HUI and the LND reservoirs are represented in the bar chart. As displayed in the figure, the total flow increases with increasing summer demand. Despite LND being the main source between October 2010 and August 2012, it was drawn upon much less after that date despite the higher (energy) cost related to drawing the water from the HUI, due to water quality differences and the desire for a backup water supply source in case of HUI pumping system failure.

Figure 3: Monthly raw water flows from LND and HUI to Mudgeeraba WTP, 2010-2014.

The relationship between pumping electricity costs of Mudgeeraba WTP and the volume of raw water (ML) pumped from the HUI was numerically established through regression analysis (Figure 4). Pumping costs from HUI equated to approximately AUD\$ 30/ML (\$1 AUD = \$0.71 USD; October 2015), which resulted in a daily pumping cost of AUD\$ 1500 for an average daily volume of 50 ML. It was also calculated that the pumping costs represent the vast majority of the total electricity costs, with the other combined costs to run the plant

typically being about 5% of the pumping costs. Thus, it can be considered of paramount importance to be able to separately assess and predict the costs associated with the treatment process, as well as those associated with the pumping costs for HUI, especially in view of a better estimation, on a daily basis, of the relative cost-benefit of avoiding the use of the LND source.

Figure 4: Relationship between pumping electricity costs for Mudgeeraba WTP (PC) and HUI daily volume drawn (VOL).

The first step in the analysis of the available water quality and the chemical usage data was the development of Self-Organizing Maps (SOMs) for the most relevant available input data regarding water quality and water treatment; they were used to concurrently reveal any significant correlations between the multiple variables. SOMs (Kohonen, 1990) are a class of unsupervised learning Artificial Neural Networks (ANN) that perform their own classification of the presented data and the dimensionality reduction of the feature space, to yield topologically ordered maps. These ordered, color contour maps enable the researcher to notice similar patterns between them, meaning that high/low levels of different variables typically matched over time. That is, the visual inspection of the component planes of the SOM provides a rapid and intuitive means of examining the covariance between the variables

and how variables vary against each other, and exploring the hypotheses for increased understanding (Mounce et al., 2014), especially in complex problems. The SOM was run using Matlab R2012a software on an Intel®Core™ i5-2400 CPU @ 3.10 GHz processor, with simulations taking only a few seconds per run. Figure 5 illustrates the SOM for the main chemicals and water quality parameters at Mudgeeraba WTP for the period 2010-14. This map contains color coded hexagons that summarize all of the component planes that represent individual variables. In the component planes for individual variables, the coloring corresponds to actual normalized numerical values for the input variables that are referenced in the scale bars adjacent to each plot. Blue shades show low values and red corresponds to high values. The x axis represents a range of input variables (water quality and chemicals parameters) for SOM analysis, while in the y axis the corresponding variables (i.e. ANN output variables) are represented.

Figure 5: SOM for main chemicals and water quality parameters, Mudgeeraba WTP 2010-14.

As illustrated in Figure 5, some variables have similar color patterns, thus indicating relevant correlations. In particular, the alum (i.e. aluminum sulphate) has a relatively similar pattern of the Tb and true water color. This is an expected finding, as it illustrates how the amount of the coagulant alum is proportional to the amount of particles and organic matter to be coagulated and removed. The Tb , as well as the Mn , also had similar patterns to the sodium hypochlorite. This relationship is also expected since sodium hypochlorite helps in the oxidation of, for instance, the soluble form of Mn (e.g. Mn^{2+}), into their insoluble ones (e.g. MnO_2), which consequently precipitates and, then, can be more easily removed. Potassium

permanganate (i.e. KMnO_4), which is dosed in the case of extremely high Mn concentration in support of pre-filter chlorination was applied on a few occasions only, when very high Mn spikes were recorded. On the other hand, polydadmec, which, like alum, is supposed to follow a path proportional to Tb and/or color, seems to be related to alkalinity. After consultation with process engineers, with experience in water treatment processes, we concluded that a possible explanation is given by the fact that: (1) alkalinity is related to pH (thus higher alkalinity implies higher pH); (2) higher pH implies a poorer floc structure; and (3) a poorer floc structure implies an increased dosage of polydadmec, which helps in bridging the flocs and improves the filtration process. Importantly, despite being intrinsically correlated to pH and alkalinity, no clear relationship was found between these variables and lime, carbon dioxide, or sodium hydroxide. This outcome occurred because their addition occurred during some intermediate step of the process; thus, the initial pH or alkalinity is only partially relevant information. However, previous work (e.g. Rothberg et al., 1993) was able to estimate pH and alkalinity changes during the treatment process by use of the Caldwell-Lawrence diagrams for pH and CaCO_3 equilibrium; these nonlinear variations are due to the addition of a certain amount of a particular chemical to the water, with pre-defined chemical characteristics. Therefore, it is actually possible, in order to predict lime, carbon dioxide and sodium hydroxide, to estimate the dosage of these chemicals required for adjusting pH and alkalinity at intermediate steps of the treatment process. Hence, the final model will make use of data-driven approaches to estimate the dosage of certain chemicals, based on the quality of the raw water, and the existing available mathematical/chemical models, in order to estimate the dosages of the remaining chemicals.

2.3. Model development

The core modelling objective of this current study was to predict the total cost of treating one ML of water, based on the expected water quality of water drawn from all of the different available gates of the two reservoirs. Thus, despite the multiple variables and objectives, as it typically happens in environmental modeling (Hauduc et al., 2015), one single target variable was identified. By comparing the costs as well as supply reliability at the different gates and reservoir sources (both in terms of water treatment and electricity), the operators will be able to better determine the optimal location source for raw water to be drawn for the Mudgeeraba WTP.

In order to achieve such objectives, different models for each predicted variable must be developed. As suggested in the previous section, different approaches were followed according to the data analysis and the available information. Following the RTW model relationships, and based on the Caldwell-Lawrence diagrams (Caldwell and Lawrence, 1953), an algorithm was developed to look for the closest-matching cell in a large database containing a number of pH variations, caused by the addition of certain amounts of the chemicals of interest. The pH variation is largely nonlinear and predominately depends on: water temperature, total dissolved solids, alkalinity, initial pH, and chemical type and dosage. This algorithm is a key-component of the full hybrid cost prediction model, as it accounts for intermediate pH variations during the water treatment, and allows the estimation of chemicals such as carbon dioxide or sodium hydroxide, which are dosed for pH neutralization. Also, for potassium permanganate, not enough data were available to build a data-driven model with *Mn*, as it is very rarely used. However, Seqwater internal operational guidelines recommend dosing an amount of permanganate three times higher than the measured concentration of soluble *Mn*. Therefore, the model relies on this information, leaving a user-defined threshold for triggering its use. The other chemicals (i.e. alum, polydadmec, and sodium hypochlorite) could be predicted accurately enough with data-driven models, based on historical data. The SOM illustrated in Figure 5 helped in reducing the number of possible relevant inputs for each chemical and setting up the regression analysis. The alum dosage could be predicted with a polynomial regression model ($R^2 = 0.54$), based on both true color and turbidity. The model is nonlinear and must consider both color and turbidity since high turbidity is often coupled with high color; the coagulation process is more efficient than when just treating high turbidity alone. Polydadmec usage was predicted based on the alkalinity values with an exponential regression model ($R^2 = 0.67$). Finally, sodium hypochlorite was predicted by using water temperature and soluble *Mn* data. The relationship with temperature is expected, as warm temperatures may lead to a quicker loss of disinfectant residuals in the water supply distribution systems (Fisher et al., 2011; Kohpaei et al., 2011; Fisher et al., 2012). The three aforementioned data-driven models for alum, polydadmec and sodium hypochlorite dosage predictions are based on historical data and thus are site-specific; therefore, they would require a recalibration (or possibly even a new identification of relevant inputs with SOM) if the model needs to be replicated for another location. Instead, predictions of lime, carbon dioxide and sodium hydroxide are based on pH variations, and thus can be applied to any location.

The flow chart in Figure 6, as well as Table 1, illustrates the step-by-step model analysis procedure, which encapsulates all of the aforementioned models, performed by the developed software tool. Firstly, the lime is added in order to adjust the alkalinity to a target value.

$$D_{lime} = \frac{(Alk_{TG} - Alk_{Raw})}{\Delta Alk_{Lime}} \quad (1)$$

where:

D_{lime} = the amount of lime to be dosed [mg/L]

Alk_{TG} = the targeted alkalinity as per $CaCO_3$ per mg/L at the end of the treatment

Alk_{Raw} = the alkalinity measured in the raw water as per $CaCO_3$ per mg/L

ΔAlk_{Lime} = the variation of alkalinity in mg/L as per $CaCO_3$ with 1 mg/L of dosage of lime.

This is constant and it is 1.35.

This, however, will cause a pH variation, raising the pH to $pH_{interim1}$ calculated through the Caldwell-Lawrence diagrams approach used in the RTW model. The pH is then corrected back to 7 by calculating the optimal dosage of carbon dioxide (CO_2) through the same theoretical principle. Before that, the potassium permanganate ($KMnO_4$) is dosed, based on the Seqwater guidelines equation (i.e. 3 times the concentration of the soluble Mn), which provided soluble Mn that exceeds the threshold chosen by the user (i.e. the WTP operators):

$$D_{KMnO_4} = \begin{cases} 0 & Mn_{sol} < Mn_{sol,TR} \\ 3 \cdot Mn_{sol} & Mn_{sol} \geq Mn_{sol,TR} \end{cases} \quad (2)$$

where:

D_{KMnO_4} = Potassium permanganate to be dosed [mg/L]

Mn_{sol} = soluble manganese measured in the raw water [mg/L]

$Mn_{sol,TR}$ = soluble manganese threshold concentration (user-defined) triggering the use of permanganate [mg/L].

Following the CO_2 calculation, alum, polyadamac and $NaClO$ are added, based on the water quality, as per the data-driven model findings, illustrated in Equations (3), (4) and (5).

$$D_{Alum} = 2.242 \log Tb + 0.2538 \cdot C + 22.08 \quad (3)$$

where:

D_{Alum} = amount of alum to be dosed [mg/L]

Tb = turbidity measured in the raw water [NTU]

C = true water colour measured in the raw water [HU]

$$D_{Pdma} = \begin{cases} 0 & Alk_{Raw} < Alk_{TR} \\ 0.0332 \cdot Alk_{Raw} - 0.035 & Alk_{Raw} \geq Alk_{TR} \end{cases} \quad (4)$$

where:

D_{Pdma} = polydadmac to be dosed [mg/L]

Alk_{TR} = the alkalinity as per $CaCO_3$ per mg/L which triggers the use of polydadmac (for this case study, $Alk_{TR} = 15$)

Alk_{Raw} = the alkalinity measured in the raw water as per $CaCO_3$ per mg/L

$$D_{NaClO} = \begin{cases} 0.2334 \cdot T_w - 0.7518 & Mn_{sol} < 0.04 \\ 0.2334 \cdot T_w + 6.4982 \cdot Mn_{sol} - 0.0334 & Mn_{sol} \geq 0.04 \end{cases} \quad (5)$$

where:

D_{NaClO} = sodium hypochlorite to be dosed [mg/L]

T_w = water temperature measured in the raw water [$^{\circ}C$]

Mn_{sol} = soluble manganese measured in the raw water [mg/L]

These again imply variations in pH and alkalinity, calculated through the same chemical approach. Finally, the pH is adjusted to the correct level by adding NaOH (i.e. sodium hydroxide), according to the same nonlinear pH variation estimation. The alkalinity is then checked to be within an acceptable range. If the final alkalinity falls outside the acceptable range, the estimated dosage of NaOH is progressively changed (i.e. by adding/subtracting 5%, then 10% etc.) until both pH and alkalinity are within acceptable boundaries. Eventually, having the cost per unit of each chemical, the total water treatment (WT) cost can be predicted by using the estimated dosages of chemicals calculated with the aforementioned procedure. Likewise, the variable electricity cost can be estimated through a data-driven approach. Applying the linear regression trend illustrated in Figure 4, the cost per ML of water drawn from HUI can be determined. Water drawn from LND is gravity fed and has no pumping requirement.

Table 1 – Cost prediction model step-by-step analysis processing procedure

Step	Parameter predicted	Predictors	Model approach
1	Lime	Alkalinity	Chemical
2	KMnO ₄	Mn_{sol}	Mathematical
3	CO ₂	pH _{interim1} - pH _{target}	Chemical
4	Alum	Tb , color	Data-driven
5	Polydadmac	Alkalinity	Data-driven
6	NaClO	T_w , Mn_{sol}	Data-driven
7	NaOH	pH _{target} -pH _{interim2} Alkanity _{target} - Alkalinity _{interim1}	Chemical
8	Total WT cost	Chemicals steps 1 to 6, costs	Mathematical
9	Electricity cost	HUI volume drawn	Data-driven
10	Total cost	Total WT cost, Electricity cost	Summation

Figure 6: Hybrid costs prediction model flow chart

3. Model Results

3.1. Chemical dosages and cost predictions

Using Figure 7, the actual model performance can be evaluated for each chemical prediction performed. The predicted dosages are multiplied by the relative costs, thus obtaining the actual and predicted total daily costs of each chemical. As the potassium permanganate chart was used only once during the historical period analyzed, it is not presented here. By using the chemical approach, which aims to dose lime and CO_2 in order to adjust the pH and alkalinity levels, it is apparent how the models tend to underestimate the actual dosage performed by the WTP operators, for the historical period analyzed. Interestingly, the

underestimation of CO₂ seems to be a logical consequence of the underestimation of the lime level; the more lime is dosed, the more the pH is raised, and the more CO₂ the WTP operators use to readjust the pH to an optimal, neutral value. Hence, the only real underestimation is for lime dosage quantification. Despite the reason being unclear, by simply adding a coefficient to take into account the underestimation, it would be possible to obtain a nearly perfect model. Further, despite the underestimations, the model predictions closely reflect the trends and peaks of the real dosage curves.

Next, the data-driven model for alum prediction yielded a very high performance, with the model outputs matching, or being very close to, the real dosages for most of the simulation period. Thus, color and turbidity, the only two inputs to the model, explain much of the variance. Similarly, the data-driven model for polydadmec prediction also yielded a satisfactory performance; while there is imprecision in quantifying the actual concentration, it was successfully able to predict when this coagulant was actually used. Moreover, as mentioned before, this data-driven model makes use of an unexpected input for a coagulant (i.e. alkalinity); nevertheless, an explanation for such a correlation was previously provided.

The least performing model modules are those predicting chemicals added in the latter stages of the treatment process (e.g. sodium hydroxide), as there is higher uncertainty related to the previous chemical predictions. For example, sodium hypochlorite was not predicted accurately for the 2010-11 period. However, after this the correlation with the water temperature became much stronger, resulting in much more accurate predictions for the remaining years. Also, the sodium hydroxide prediction was the poorest. However, this may have been a consequence of the underestimation of lime and CO₂, leading to the different interim pHs to be adjusted. Further investigations need to be undertaken. Interestingly, unlike the lime and CO₂, sodium hydroxide was typically overestimated, which, interestingly, balances the previous models' underestimations.

Overall, the costs for coagulants (i.e. alum and polydadmec) represent the highest fractions of the chemical costs for treatment, being approximately AUD\$ 800/1000 per day. The chemicals used for the pH and alkalinity adjustment (i.e. lime, CO₂, and sodium hydroxide) also represent a high cost (about AUD\$ 700/850 per day), while chlorination (about AUD\$ 200/400 per day) and fluoridation (about AUD\$ 100/day, not shown in this paper) play a secondary role in determining the total water treatment cost.

Figure 7: Actual and predicted total daily cost of the chemicals (14-days moving average), Mudgeeraba WTP 2010-14

3.2. Total costs assessment and optimization considerations

After each sub-model performance was assessed, the total predicted cost associated with the estimated dosages of the chemicals was compared to the actual total costs incurred during the 2010-14 period (Figure 8). The analysis showed that the model was accurate, with only a few minor errors. As a consequence, more assessments and evaluations over possible cost reductions could be undertaken. First of all, when the HUI is selected as main source of raw water, the pumping costs play a major role in determining the total treatment-associated costs

(see Figure 9). Those increase by over 50% when HUI is the only source of raw water. It is important to investigate the cost-benefit of such an intake choice. As a consequence, the model was used to predict the total costs, with a different raw water selection (i.e. LND 100%), using one of the worst-case scenarios of water quality recorded in the historical data (i.e. the LND water quality was much worse than in the HUI). The findings can be summarised as follows, and are presented in Table 2.

The day selected for this assessment was taken from the historical records; the raw water coming from the LND presented a very high turbidity (140 NTU), while the blended raw water (about 49% HUI and 51% LND) recorded about 70 NTU only, as the water quality in HUI was better. By assuming that 100% of the water was withdrawn from the LND, the model was used to back-predict the chemicals necessary to guarantee the delivery of safe treated water. Focusing on the turbidity-alum only (as the other water quality parameters played only a minor role in changing the final treatment costs), the relationship between them was found to be not linear. Therefore, despite dealing with a turbidity level twice as high as what the WTP actually received, the required level of alum to guarantee proper coagulation was estimated to only rise from 38 to 50mg/L. This would have implied extra chemical costs of AUD\$ 270; on the other hand, the model predicted savings in pumping costs of about AUD\$ 650, since all the water would have flowed by gravity from the LND, rather than pumping 18.6 ML from HUI.

Therefore, when no other considerations are taken into account (e.g. LND water supply availability, potential taste and odour complaints, etc.), the applied model proved that savings of about AUD\$ 400 per day could have been achieved. Moreover, those savings could be much higher in a normal scenario where the water quality from the LND was on average similar to that of HUI. Further, the chemicals costs, with a supposed 100% LND use, would be similar to the current mixed intake choice.

In support of these conclusions, Figure 10 analyses the variations in historical (i.e. 2010-14) water treatment costs by assuming either increased or decreased usage of the LND source. As a reference, LND usage was 47% on average, over the 2010-14 period. The estimated savings on the y axis are based on the predicted percentages multiplied by the average daily water treatment costs (i.e. for chemicals dosage) for the period 2010-14, which was AUD\$ 2,122, or the average daily pumping costs, which was AUD\$ 868. The chart was built using average

results to simplify this assessment; in reality each daily estimated cost would vary around the calculated averages based on water quality, water demand, and initial HUI usage. However, a best-case scenario (BS) and worst-case scenario (WS) for water treatment costs (i.e. assuming that the water quality in LND is consistently better/worse than in HUI) were considered according to the model outputs based on the historical data. In this way, the actual savings could be estimated to be between the two boundary conditions.

Figure 8: Actual and predicted total costs of chemicals for water treatment for Mudgeeraba WTP, 2010-14

Figure 9: Predicted treatment costs (i.e. chemicals costs), pumping costs and total costs for Mudgeeraba WTP, 2010-14.

Table 2 – Optimised intake selection considering historical worst-case water quality scenario

Parameter	Actual data	Optimised Scenario
Source volume [ML]	HUI=18.6; LND=19.6	HUI=0; LND=38.2
Turbidity [NTU]	70	140
Predicted Alum [mg/L]	38	50
Predicted Alum cost [AUD\$]	880	1150
Predicted pumping costs [AUD\$]	650	0
Total variable costs [AUD\$]	1530	1150

The total costs of treatment (i.e. pumping + chemicals) would decrease even in a sustained adverse scenario of water quality in LND. To explain, in the case of a 100% increase in usage of LND (i.e. historical LND usage of 47% doubles to 94%) and poor LND water quality, the extra chemicals costs would amount to AUD\$ 500/day. It should be noted that this scenario is unlikely since there were many days historically in which the water quality in LND was better than in HUI; so in this best case the higher LND draw could reduce treatment costs by up to AUD\$ 350/day. Treatment costs were summed with pumping cost reductions, which was directly proportional to the average amount of water drawn from HUI and not dependent on water quality. Pumping cost savings of AUD\$ 800/day could be achieved with a doubling of LND usage. In summary, by considering a twofold increase in raw water drawn from

LND, it would be possible to achieve daily total monetary savings of between AUD\$ 350 (WS) and AUD\$ 1100 (BS).

However, it is unlikely that the smaller LND would have sufficient capacity to be the primary source all year round. Nevertheless, even a 50% increase would still lead to substantial monetary savings (i.e. between AUD\$ 175 and AUD\$ 550 per day on average). Additionally, LND often spills during the wet season due to its small size and high inflow, which also replenishes it rather quickly. Thus an assessment of the possibility of an increased usage of LND, especially during wet periods, (i.e. typically from December to late March) is strongly recommended.

Figure 10: Spider diagram with predicted variations in cost of treatment, pumping and total, considering a best-case (BS) scenario and a worst-case (WS) scenario of water quality and different percentages of LND usage variation compared to the historical dataset.

3.3. Reservoir to raw water models

In order to develop an effective water intake optimization support tool for WTP operators, it must be related to the reservoirs' water quality, rather than to the raw water quality recorded at the WTP. By doing so, the operators can select the best reservoir and, importantly, the best depth to take the water from, based on the water quality and the subsequent predicted

chemicals/energy costs. These water quality parameters could change remarkably from the moment the water is drawn from the reservoir to the moment it reaches the WTP. Hence, data analysis was performed and data-driven models were built for this purpose. Data was collected from historical lake samplings and Vertical Profiling Systems (VPS) installed at HUI and LND. The intake depth was calculated for each day of the historical dataset, based on the height of the gates and dam level, while the water quality, at the intake depth, was estimated by interpolating the available data.

The developed models achieved very good correlations to predict raw water quality parameters such as water colour, turbidity, and water temperature for both reservoirs ($R^2 > 0.7$ for all the models) and acceptable correlation with pH for HUI ($R^2 = 0.4$). As the final purpose was to apply the model to the data collected remotely by the VPSs, a model to link alkalinity (not measured by the VPS) to pH was successfully developed (Figure 11, $R^2 = 0.64$). The manganese levels at both dams, which can be indirectly predicted through the VPS data, using previously developed data-driven manganese prediction models (Bertone et al., 2015), showed poor correlations with the levels recorded at the WTP. However, little overlapping data was available and the correlations are expected to improve, over time, when more data for the analysis is collected. On the other hand, the pH from the LND was found to noticeably, and unpredictably, change before the water reaches the WTP (Figure 12). Despite these observable fluctuations and peaks in the pH at the reservoir (possibly due to algae), the pH at the WTP, after travelling for almost 8 km of pipeline, appears to typically settle in the 6.5 to 7 range. Further investigation is required to clarify the reasons of this pH change behaviour.

Figure 11: Alkalinity and pH time series for the actual blended raw water for Mudgeeraba WTP, 2010-14

Figure 12: Time series of pH recorded at LND (depth of the intake) and at the WTP, 2010-14

4. CONCLUSIONS

A hybrid prediction model, consisting of chemical, mathematical, and data-driven sub-models, was developed for a WTP in order to estimate: (1) the variations in water quality from the two potential raw water source reservoirs for the WTP; (2) the dosages of the chemicals required to treat the raw water for a range of gate and source scenarios; and finally (3) the associated variable costs of source selection options, including water quality treatment and the additional variable costs for pump electricity. The aggregated water treatment cost prediction model was demonstrated to be accurate when compared against the historical

dataset. The final developed model was used to run various scenarios with different historical intake and reservoir source selection choices. Scenario analysis showed that there is significant potential for improved operational decision making with respect to the traditional source selection and intake selection practices. Specifically, scenario analysis showed that there is significant life cycle cost savings in pumping costs if LND was utilized more extensively. Even considering scenarios of poor water quality from LND, requiring considerably more water treatment, the total overall variable operating cost is still lower than selecting the raw water from HUI due to the pumping requirements for this source. The approach and analytical techniques developed and applied herein has applicability to other WTP having multiple source selection options.

Future research will be undertaken to assess the feasibility of greater utilization of the LND reservoir source. LND, being a relatively small reservoir, often spills after wet weather events, resulting in a number of periods where water that could have been used for Mudgeeraba WTP at no pump energy cost is wasted. Therefore, to determine the extent to which LND can be reliability drawn, the variable cost model developed herein will need to be coupled with a LND rainfall-runoff model, thereby indicating the reliable yield from LND over a particular period having certain weather conditions. Moreover, further research will be completed to improve the level of understanding of the pH variations from the LND to the WTP, and to build a better model for dosing requirements for excessive levels of soluble manganese.

The ultimate goal of this research is the development of a decision support tool for the optimized operational cost of Mudgeeraba WTP using predominately the near-real time measurements taken from the lake diagnostic sensors (i.e. VPS) located near the offtake towers of both HUI and LND. This decision support tool will assist the WTP operators with chemicals dosages estimation as well as the optimal reservoir and intake tower depth selection. Such a tool has a range of benefits, including: (1) enhanced value added to the existing deployed VPS instrumentation; (2) reduced WTP operational costs; (3) enhanced ability for operators to better prepare for poor raw water quality entering the plant; (4) better utilization of the low energy LND water source; and (5) evidence available for decisions pertaining to the mixing of the two source waters for optimized dual source water treatment.

ACKNOWLEDGEMENTS

Our industry partner to this research (Seqwater) is acknowledged for the technical and financial support required to complete this collaborative project. In particular, the authors would like to thank Seqwater operators located at the Mudgeeraba water treatment plant for providing the valuable input data for the model. Geoff Hamilton is also acknowledged for his specialist knowledge on plant operations and chemical treatment requirements.

References

Abdullahi, M.E., Odigire, J.O. 2006. Development of mathematical model for determining the quantity of alum required for water treatment. *Journal of Engineering Technology and Industrial Application*, **2**(2).

Abdullahi M.E., Abdulkarim, B.I. 2010. Development of mathematical model for determining the quantity of chlorine required for water treatment. *Journal of Applied Sciences Research*, **6**, 1002-1007.

Abdullahi M.E., Aloko, D.F., Baba, G.A., Mohammed, J. 2012. Predictive model for lime dosage in water treatment plant. *International Journal of Scientific and Research Publications*, **2** (12), 1-5.

Abdullahi, M.E. 2013. Development of a model for the estimation of water treatment cost: a data mining approach. *International Journal of Scientific & Engineering Research*, **4**(4), 610-615.

Al-Ali, M.I., Abdullah, A.I., Mohammad, A.M., Al-Ali, K.I. 2011. Estimating and modelling the interrelationships between physicochemical pollutants of Samara drug factory wastewater. *European Journal of Scientific Research*, **61**, 2, 230-241

Alcolea, A., Renard, P., Mariethoz, G., Bertone, F. 2009. Reducing the impact of a desalination plant using stochastic modeling and optimization techniques. *Journal of Hydrology*, **365**(3), 275-288.

Baruch, I. S., Georgieva, P., Barrera-Cortes, J., De Azevedo, S. F. 2005. Adaptive recurrent neural network control of biological wastewater treatment. *International Journal of Intelligent Systems*, **20**(2), 173-193.

Bertone E., Stewart R.A., Zhang H., Bartkow, M., Hacker, C. Decision support system for manganese forecasting and proactive treatment in sub-tropical water reservoirs. *Environmental Modelling and Software* [revised May 2015].

Borazjan, M., Bina, B., Amin, M., Taheri, E., Hassanzadeh, A., Modarresi, M. 2014. Optimization of treatment process of Bushigan treatment plant in respect of turbidity and total organic carbon reduction. *International Journal of Environmental Health Engineering*, 3:9.

Bozkurt, H., Quaglia, A., Gernaey, K.V., Singh, G. 2015. A mathematical programming framework for early stage design of wastewater treatment plants. *Environmental Modelling and Software*, **64**, 164–176.

Caldwell, D.H, Lawrence, W.B. 1953. Water Softening and Conditioning Problems: Solution by Chemical Equilibrium Methods. *Industrial and Engineering Chemistry*, **45**(3), 535-548.

Chen, C.W., Wei, C.C., Liu, H.J., Hsu, N.S. 2014. Application of neural networks and optimization model in conjunctive use of surface water and groundwater. *Water Resources Management*, **28**(10), 2813-2832.

Department of Environment and Resource Management - Water Planning, South East and Water Services South, South East Region. 2010. Gold Coast Resource Operations Plan.

Dürrenmatt, D.J., Gujer, W. 2012. Data-driven modeling approaches to support wastewater treatment plant operation. *Environmental Modelling and Software* **30**, 47–56.

Fisher, I., Kastl, G., Sathasivan, A., Jegatheesan, V. 2011. Suitability of chlorine bulk decay models for planning and management of water distribution systems. *Critical Reviews in Environmental Science and Technology*, **41**(20), 1843-1882.

Fuchs-Hanusch, D., Friedl, F., Scheucher, R., Kogseder, B. Muschalla, D. 2013. Effect of seasonal climatic variance on water main failure frequencies in moderate climate regions. *Water Science and Technology: Water Supply*, **13**(2), 435-446.

Gillot, S., De Clercq, B., Defour, D., Simoens, F., Gernaey, K., Vanrolleghem, P.A. 1999. Optimisation of wastewater treatment plant design and operation using simulation and cost analysis. Proceedings 72nd WEF Conference. New Orleans, USA, October 9–13 1999.

Gunaratnam, M., Alva-Argaez, A., Kokossis, A., Kim, J.K., Smith, H. 2005. Automated design of total water systems. *Industrial and Engineering Chemistry Research*, **44**(3), 588.

Hakanena, J., Sahlstedt, K., Miettinen, K. 2013. Wastewater treatment plant design and operation under multiple conflicting objective functions. *Environmental Modelling and Software* **46**, 240–249.

Hauduc, H., Neumann, M.B., Muschalla, D., Gamerith, V., Gillot, S., Vanrolleghem, P.A. 2015. Efficiency criteria for environmental model quality assessment: A review and its application to wastewater treatment. *Environmental Modelling and Software* **68**, 196–204.

Kawamura, S. 2000. Integrated design and operation of water treatment facilities – 2nd edition. Wiley, New Jersey.

Kohonen, T. 1990. The Self-Organizing Map. Proceedings of the IEEE, 78(9), 1464-1480.

Kohpaei, A. J., Sathasivan, A. and Aboutaleb, H. 2011. Evaluation of second order and parallel second order approaches to model temperature variation in chlorine decay modelling. *Desalination and Water Treatment*, **32**(1-3), 100-106.

Kusiak, A., Wei, X. 2011. Prediction of methane production in wastewater treatment facility: a data-mining approach. *Annals of operation research*, **216**(1), 71-81

Lee, M. W., Joung, J. Y., Lee, D. S., Park, J. M., Woo, S. H. 2005. Application of a moving-window-adaptive neural network to the modeling of a full-scale anaerobic filter process. *Industrial and Engineering Chemistry Research*, **44** (11), 3973-3982.

Maier, H.R., Morgan, N., Chow, C.W.K. 2004. Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environmental Modelling and Software* **19**(5), 485-494.

McCorquodale, A., Gerges, H., Georgiou, I., Yu, W., Ullensvang, T. 2005. Field tests and physical modeling of the Sacramento WWTP. Proceedings of the 78nd Annual Water Environment Federation Technical Exposition and Conference [CD-ROM]; Washington, D.C., Oct 29–Nov 2; Water Environment Federation: Alexandria, Virginia.

Mirsepasi, A. 2004. Application of Intelligent System for Water Treatment Plant Operation. *Iranian Journal of Environmental Health Science and Engineering*, **1** (2), 51-57.

Mounce, S.R., Sharpe, R., Speight, V., Holden, B., Boxall, J.B. 2014. Knowledge discovery from large disparate corporate databases using self-organising maps to help ensure supply of high quality potable water. 11th International Conference on Hydroinformatics HIC 2014, New York City, USA, 17-21 August 2014.

Regneri, M., Klepiszewski, K., Ostrowski, M., Vanrolleghem, P.A. 2010. Fuzzy decision making for multi-criteria optimization in integrated wastewater system management. Proceedings of the IWA 6th International Conference on Sewer Processes and Networks, Surfers Paradise (Australia), 7-10 November 2010.

Rietveld, L.C., van der Helm, A.W.C., van Schagen, K.M., van der Aa, L.T.J. 2010. Good modelling practice in drinking water treatment, applied to Weesperkarspel plant of Waternet. *Environmental Modelling and Software* **25**(5), 661-669.

Rogers, P., Lockyer, J., Stevenson, S. 2008. Mudgeeraba WPP filtration evolution. 33rd Annual Qld Water Industry Operations Workshop Indoor Sports Centre, Carrara – Gold Coast City, Australia, 3-5 June, 2008.

Rothberg, M.R., Scuras, S.E., Leonardo, D., Hale, J. 1993. AQUALINK: Computerized Corrosion Control. *Journal - American Water Works Association* **85** (7), 26.

Rouen, M., George, D.G., Kelly J.L., Lee M.J. 2005. High-resolution automatic water quality monitoring systems applied to catchment and reservoir monitoring. *Freshwater Forum* **23**, 20–37

RTW Professional Engineers and Consultants. 1996. The Rothberg, Tamburini and Winsor model for corrosion control and process chemistry (3.0). American Water Works Association, Denver, Co, 1996.

Sarai, D.S. 2006. Water treatment made simple for operators. Wiley, Hoboken, New Jersey

Savic, D.A., Davidson, J.W., Davis, R.B. 1999. Data mining and knowledge discovery for the water industry. In Savic, D.A. and Walters, G.A. Eds., *Water Industry Systems*, Vol. 2. Research Studies Press Ltd, UK, pp. 155-163, 1999.

Seggelke K., Rosenwinkel K.H., Vanrolleghem P.A., Krebs P. 2005. Integrated operation of sewer system and WWTP by simulation-based control of the WWTP inflow. *Water Science and Technology*, **52**(5), 195-203.

Slavics, J.J. 1973. Standard and specifications. In: “Food safety through the Hazard Analysis and Critical Control Point system”, pp. 275-289. Pillsbury Co., Minneapolis, Minnesota.

Slavik, I., Uhl, W., Völker, J., Lohr, H., Funke, M., Rolinski, S., Paul, L. 2010. IntegTa: a procedure for integrative management of dammed raw water reservoirs for drinking water production and their lower reaches. *Water Science and Technology: Water Supply*, **10**(5), 783-792.

van Leeuwen, J. A., Fabris, R., Sledz, L. and van Leeuwen, J. K. (2001) Modelling enhanced alum treatment of southern Australian raw waters for drinking purposes. International Congress on Modelling and Simulation. MODSIM Canberra, 10-13 December 2001.