environmental sciences proceedings

MDPI

Proceeding Paper

Optimized Chlorine Bulk Decay Models and a Machine-Learning-Guided Water Quality Responsive Kinetic Model for Residual Chlorine Prediction †

Iman Jafari, Rongmo Luo, Fang Yee Lim, Ng Szu Hui and Hu Jiangyong *

* Correspondence: ceehujy@nus.edu.sg

† Presented at the 7th International Electronic Conference on Water Sciences, 15–30 March 2023; Available online: https://ecws-7.sciforum.net.

Abstract: Effective treatment and disinfection of source waters and the safe delivery of potable water to customers require a comprehensive understanding of water quality changes from source to tap. Prediction of water quality within distribution systems, including disinfectant residual loss and by-product generation, has been a subject of applied research since the early 1990's. In this study the currently existing process based bulk decay models were modified by replacing initial chlorine concentration parameter with chlorine demand in their equations and the results showed that this modification could improve the performance of the models by 38.0%, 28.0%, 23.1% and 33.3% in average for First Order Model (FOM), Parallel First Order Model (PFOM), Second Order Model (SOM) and Parallel Second Order Model (PSOM), respectively. Furthermore, an online predictive method based on a machine learning algorithm was introduced and implemented in this study to predict first order chlorine bulk decay rate by feeding water quality parameters as the inputs. In addition, a novel methodology was introduced and suggested in this study based on the obtained results to be applied in real water distribution system for an optimized online prediction of residual chlorine.

Keywords: chlorine decay; distribution system; bulk water decay; water quality; modelling; machine learning

Citation: Jafari, I.; Luo, R.; Lim, F.Y.; Hui, N.S.; Jiangyong, H. Optimized Chlorine Bulk Decay Models and a Machine-Learning-Guided Water Quality Responsive Kinetic Model for Residual Chlorine Prediction. *Environ. Sci. Proc.* **2023**, *5*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: 15 March 2023

1. Introduction

Chlorine is a widely used chemical disinfectant in water treatment plants, and its residual concentration in treated water is an important indicator of the effectiveness of the treatment process. Accurate prediction of the residual chlorine concentration is essential for ensuring the safety and quality of treated water, as well as for optimizing the operation of water treatment plants [1,2].

Chlorine bulk decay models have been widely studied in the literature as a means of predicting the residual chlorine concentration in treated water. These models are based on the premise that the concentration of chlorine in water decreases over time due to various physical, chemical, and biological processes. Many studies have focused on developing empirical models that describe the relationship between chlorine decay and various parameters such as water temperature, pH, and the presence of other chemical species. However, these models are often based on limited data and may not be accurate for all water systems [3].

The development of accurate and reliable chlorine bulk decay models is a challenging task due to the complexity of the processes involved and the large number of parameters that can affect the rate of chlorine decay. Many studies on chlorine bulk decay modeling are based on limited data sets, which may not be representative of all water systems. This can lead to models that are not accurate or reliable for all water treatment plants [4].

The rate of chlorine decay is influenced by a large number of parameters, which can have complex relationships with each other. This makes it difficult to develop accurate models that capture all of these relationships. The relationships between the rate of chlorine decay and various parameters are often nonlinear, which makes it challenging to model them accurately using traditional methods. Water quality can vary significantly depending on the source and treatment processes, which can affect the rate of chlorine decay. The presence of microorganisms and other biological processes can significantly affect the rate of chlorine decay. These processes are difficult to model accurately and can lead to significant errors in the prediction of residual chlorine concentration [5].

Overall, the development of accurate and reliable chlorine bulk decay models is a challenging task due to the complexity of the processes involved and the large number of parameters that can affect the rate of chlorine decay. Despite these challenges, the development of improved models for predicting residual chlorine concentration is essential for ensuring the safety and quality of treated water and optimizing the operation of water treatment plants [6].

In this study new chlorine decay models were proposed by modifying currently existing decay models from first order to parallel second order model by introducing a new parameter called "Total Chlorine Demand" and were assessed against their accuracy in explaining chlorine bulk decay behavior in an attempt to correct the flaws of existing models. Furthermore, in order to make the predicting chlorine decay model cover expected changes in source water, demands, or system operation over the ensuing weeks, months, or seasons, the required parameters in kinetics model should be estimated online rather than offline measurements in laboratory environment which restricts the models to offline use and planning-level analysis. In this regard, a new methodology was used in this study to predict bulk decay coefficients based on water quality parameters through an analytical process instead of running bulk decay experiments in a laboratory environment. Accordingly, to cover the effects of water quality variabilities on chlorine bulk decay kinetics, a predictive tool based on machine learning algorithms was used to train a model to predict chlorine bulk decay rate coefficients based on water quality parameters as the input and kinetic parameters as the output in such a way that the system assigns new values for the rates for each water sample with different water quality properties. In other words, in this proposed methodology the system automatically updated the values for decay rate coefficients based on the changes in water quality parameters in an online manner.

2. Methods

2.1. Sample Preparation

Water samples were collected from a portion of Singapore water distribution system from different locations within the site, anonymously described by "S" in Figure 1, in order to provide an appropriate representation of water quality throughout the network.

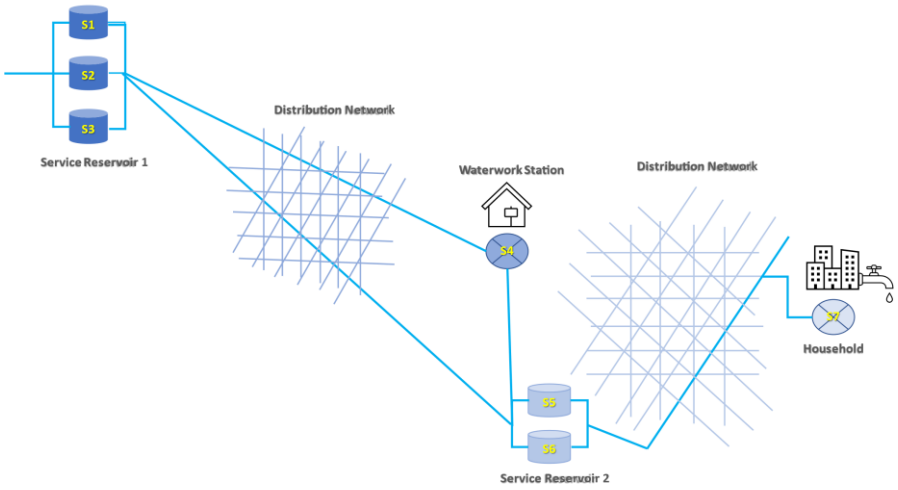

Figure 1. Sampling locations in Singapore water distribution network.

From September 2020 to February 2021, five water sample collection attempts were done from seven different locations throughout water distribution system as explained in Table 1 and totally 31 samples were collected during this period.

Table 1. Description of water sample collection dates and locations.

Date	Time			
	S1 S2 S3	**S4**	**S5 S6**	**S7**
8/9/2020	9:00 AM (27 °C)	11:00 AM (28 °C)	2:00 PM (28.6 °C)	4 PM (29.2 °C)
22/9/2020	9:00 AM (27.2 °C)	10 AM (29.9 °C)	2:00 PM (31.2 °C)	3 PM (30.4 °C)
13/10/2020	9:00 AM (27.0 °C)	10:00 AM (30.3 °C)	2:00 PM (31.3 °C)	3 PM (30.6 °C)
26/1/2021	9:40 AM (27.5 °C)	10:20 AM (30.0 °C)	Under maintenance	2:00 PM (30.0 °C)
9/2/2021	9:30 AM (28.2 °C)	10:40 AM (28.9 °C)	Under maintenance	2:20 PM (28.9 °C)

2.2. Water Quality Parameter Measurements

To achieve a well-defined chlorine decay profile for a particular water sample, selection of appropriate times for measurement of water quality parameters is crucial [7]. In this study the time intervals selected to measure the water quality parameters were 1 h, 2 h, 4 h, 8 h, 12 h, 24 h, 48 h, 72 h, 96 h, and 168 h after collection time and the details on the methods and techniques used to measure these parameters are provided in Table 2.

Table 2. Measured water quality parameters and their corresponding measurement methods and instruments.

Parameters	Unit	Method/Instrument
Total Residual Chlorine (TRC)	mg/L Cl_2	Method 867 (DPD method)
NH_2Cl	mg/L Cl_2	Method 10171 (Indophenol method)
TOC	mg/L	TOC analyzer (Shimadzu TOC-L)
pH / Temperature	No unit/°C	pH meter (Horiba Scientific pH1100)

UV254	cm^{-1}	HACH Spectrophotometer (DR6000)
Total NH_3	mg/L NH_3–N	HACH Spectrophotometer (DR6000)
Free NH_3	mg/L NH_3–N	HACH Spectrophotometer (DR6000)
fDOM	a.u. Ex 365 Em 480	Fluorescence Spectrophotometer (Agilent Technologies Cary Eclipse)

2.3. Modified Kinetic Bulk Decay Models

Given to the fact that each water sample based on its quality indices needs a specific amount of chlorine concentration in order to react with all of the organic and inorganic reactants existing in this water sample, in this case initial chlorine concentration (Cl_0) is replaced by a new variable called initial Total Chlorine Demand (TCD_0), which is a portion of initial chlorine concentration [8]. As a result, previous kinetic equations were modified accordingly and are reported in Table 3.

Table 3. Original versus modified equations of existing chlorine decay models.

Models	Original Equation	Modified Equation	Description on New Parameters
First order Model (FOM)	$Cl(t) = Cl_0 \times e^{-k \times t}$	$Cl(t) = TCD_0 \times e^{-k_d \times t} + (Cl_0 - TCD_0)$	TCD_0: Initial Total Chlorine demand K_d: First Order decay rate associated with initial total chlorine demand
Parallel First order Model (PFOM)	$Cl(t) = f(Cl_0, t) = Cl_0 \times x \times e^{-k_1 \times t} + Cl_0 \times (1 - x) \times e^{-k_2 \times t}$	$Cl(t) = TCD_0 \times x \times e^{k_{1d} \times t} + TCD_0 \times (1 - x) e^{k_{2d} \times t} + (Cl_0 - TCD_0)$	K_{1d}: First order fast reaction rate constant of the chlorine decay associated with $TCD_0 \times x$ K_{2d}: First order slow reaction rate constant of the chlorine decay associated with $TCD_0 \times (1 - x)$
Second Order Model (SOM)	$C_{Cl}(t) = \frac{C_{Cl_0} - C_{A_0}}{1 - \frac{C_{A_0}}{C_{Cl_0}} \times e^{-(C_{Cl_0} - C_{A_0}) \times k \times t}}$	$Cl(t) = \frac{TCD_0 - C_{A_0}}{1 - \frac{C_{A_0}}{TCD_0} \times e^{-(TCD_0 - C_{A_0}) \times k_d \times t}} + (Cl_0 - TCD_0)$	K_d: Second order decay rate associated with initial total chlorine demand (TCD_0) and initial notional reactant (C_{A_0})
Parallel Second Order Model (PSOM)	$Cl(t) = \frac{Cl_0 Z (1 - R_1)}{1 - R_1 \times e^{-(1 - R_1) \times k_1 \times t}} + \frac{Cl_0 (1 - Z)(1 - R_2)}{1 - R_2 \times e^{-(1 - R_2) \times k_2 \times t}}$	$Cl(t) = \frac{TCD_0 Z (1 - R_1)}{1 - R_1 \times e^{-(1 - R_1) \times k_{d1} \times t}} + \frac{TCD_0 (1 - Z)(1 - R_2)}{1 - R_2 \times e^{-(1 - R_2) \times k_{d2} \times t}} + (Cl_0 - TCD_0)$	k_{d1}: Second order fast reaction rate constant of the chlorine decay associated with $TCD_0 \times Z$ and $(1 - R_1)$ k_{d2}: Second order slow reaction rate constant of the chlorine decay associated with $TCD_0 \times (1 - Z)$ and $(1 - R_2)$

2.4. Responsive Kinetic Model

A machine learning based predictive tool was introduced and developed in this study based on water quality parameters as the input and kinetic parameters as the output. A schematic of this predictive model developed in this study using Gaussian Process Regression (GPR) algorithm is provided in Figure 2.

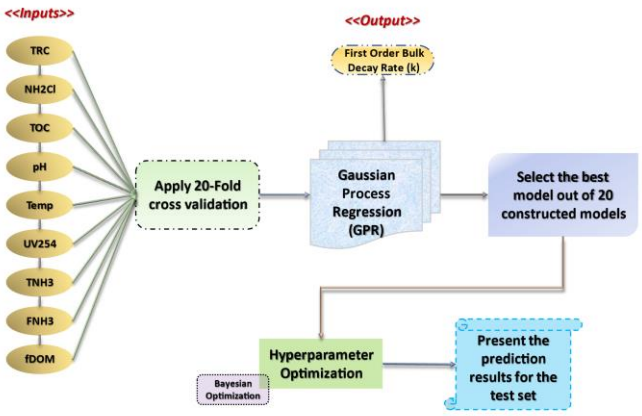

Figure 2. Machine Learning (ML) model set-up for chlorine bulk decay predictions.

3. Results and Discussion

Figure 3 presents the MSE values graphically for all nominated models and their modified versions. As can be seen from Figure 3, PSOM and PFOM were ranked first and second, respectively amongst all nominated models. One of the reasons that PFOM and PSOM generated more accurate results compared to FOM and SOM is that the parallel first and second order models (PFOM and PSOM) contain more parameters that need to be estimated and therefore have more flexibility to fit the data. Although PSOM was ranked better than PFOM with regards to accuracy of the predicted chlorine residuals, it should be noted that PSOM contains five parameters compared to only three parameters of the PFOM and thus it can be said that PSOM produces more accurate results than PFOM due to having more model parameters. In addition, by applying the proposed modification in this study on all four nominated models, the MSE values were decreased by 38.0%, 28.0%, 23.1%, and 33.3% for FOM, PFOM, SOM and PSOM, respectively.

Moreover, as noted from the results, adding only one additional parameter (TCD_0) to the general first order model increased the accuracy of the model to the extent that its performance was even better than all previously existing models with higher complexity such as PFOM, SOM and PSOM. Therefore, it is concluded that MFOM is the most accurate process-based chlorine decay model for modelling chlorine decay in bulk water with the least complexity in converging process. It is generally true that models with more parameters can fit a given set of data more accurately, but this comes at the cost of increased complexity and a greater risk of overfitting. Overfitting occurs when a model fits the training data too well, but does not generalize well to new, unseen data. This can lead to poor performance on out-of-sample data. In the case of kinetic chlorine bulk decay models, a model with a higher number of parameters may be more accurate at fitting the data used to train the model, but this may not necessarily lead to better performance on new, unseen data. This is because the additional parameters may be fitting to noise or random fluctuations in the training data, rather than capturing the underlying trends and relationships. Therefore, while more complex models may be more accurate on the training data, they may not be more practical or reliable for predicting the behaviour of the system on new, unseen data. In general, it is important to carefully consider the trade-off between model complexity and performance when choosing a model for a particular application.

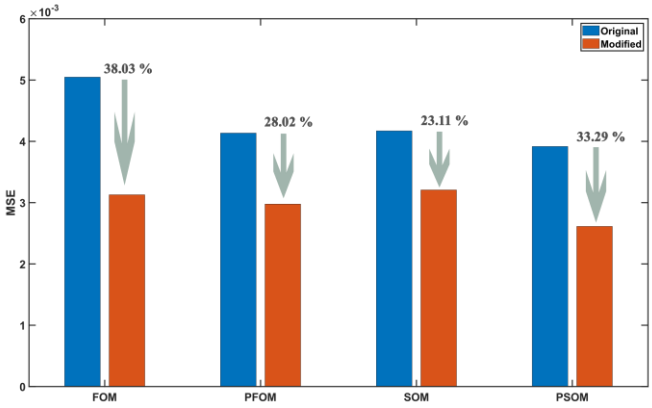

Figure 3. The MSE values for all nominated models and their modified versions.

The obtained results from the ML model were used in fitting FOM equation on its corresponding test set that were not involved in training the model to examine the performance of this proposed responsive bulk decay model in predicting TRC decay. The plots of raw TRC decay data and corresponding fits of original FOM as well as the responsive FOM for 11 sets of data that were separated before training the ML model are shown in Figure 4. As can be seen, for all datasets the predicted-FOM after the optimization showed closer fits to the original FOM comparing to the one before the optimization process. Furthermore, the results showed that the prediction of the first order rate coefficients by the ML model after optimization was acceptable in all of the cases except in a few circumstances where the distance between the fitted line using the original FOM and also the one using optimized-predicted-FOM was higher than expected probably due to the limited amount of data for training the model, the presence of other affecting factors on first order decay rate variations which were not considered in this study, as well as the high sensitivity of first order bulk decay kinetic model to minor errors in predicted rate coefficients.

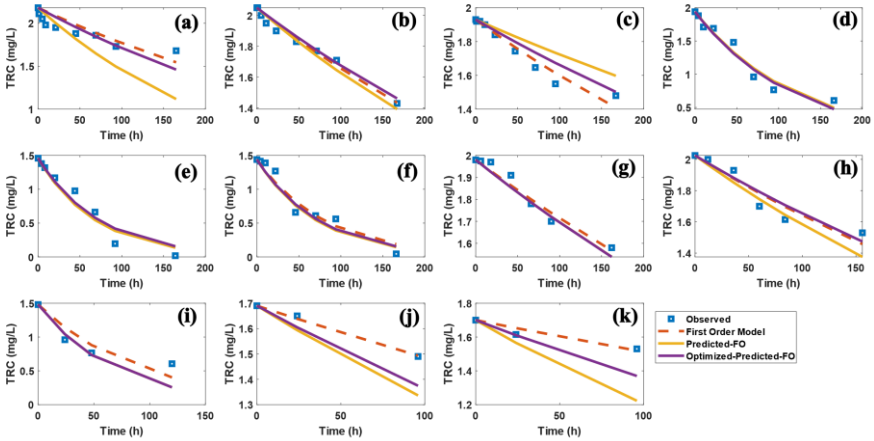

Figure 4. The plots of raw TRC decay data and corresponding fits of original FOM as well as the responsive FOM for 11 sets of data that were separated before training the ML model.

Based on the results obtained in previous sections, this part of the study provides an implementation plan in real water distribution network that can potentially fill the gap in disinfection decay monitoring by developing a predictive tool that integrates various real-time water quality sensors, supervisory control and data acquisition (SCADA) systems, and hydraulic and chemical kinetics models. Having the overall chlorine decay kinetic parameter is not enough to estimate the residual chlorine concentration in different locations of the distribution system over time. Other requisites like a suitable and calibrated hydraulic model are needed to predict the chlorine concentration versus time in different

locations. In this regard, it was proposed here to consider each water flow as a segment inside of which the chlorine was reacting with organic and inorganic matters based on the chemical kinetic equations. This water segment was also supposed to flow throughout the distribution system with an average velocity obtained from the calibrated hydraulic model. The kinetic parameters in above mentioned equations can be predicted by ML model in such a way that once a new water flow with different properties flows in the distribution system, new values are assigned for the kinetics parameters as the input variables into the ML model would change. Hence, by the combination of a calibrated hydraulic model and a suitable kinetic model, as well as an accurate predictive model (ML model), the concentration of chlorine can be estimated in different locations of the water distribution system versus time. In Figure 5, a summary of this proposed methodology is presented.

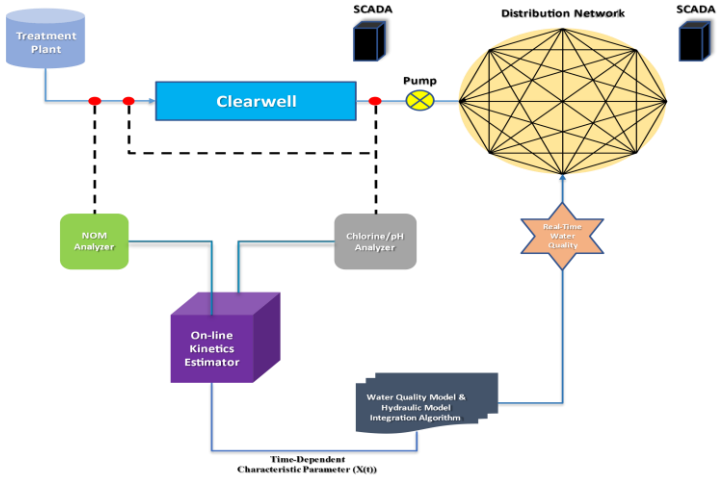

Figure 5. A summary of proposed methodology for online TRC prediction in water distribution system.

4. Conclusions

In this study the currently existing process based bulk decay models were modified by replacing initial chlorine concentration parameter with Total Chlorine Demand in their equations and the results showed that this modification could improve the performance of the models by 38.0%, 28.0%, 23.1% and 33.3% in average for FOM, PFOM, SOM and PSOM, respectively. In addition, it was proven that the chlorine decay prediction in water distribution system can be modified and robust to be used as an online tool for predicting residual chlorine in different locations of distribution system over time rather than to be restricted by off-line use and planning-level analysis. In this regard, an online predictive method based on a machine learning algorithm was introduced and implemented in this study to predict first order chlorine bulk decay rate by feeding water quality parameters as the inputs. Hence, a GPR model was trained and used to predict the kinetic parameter in FOM, and the results showed that although the accuracy of predictions for the test set was high for most of the cases, the high sensitivity of the FOM to its kinetic parameter (first order decay rate coefficient) resulted in high MSE values in some of the TRC predictions. However, the high correlation coefficients between the predicted and actual TRC values represented the fact that the model could properly identify the substantial process behind TRC prediction based on water quality parameters. In addition, a novel methodology was introduced and suggested in this study based on the obtained results to be applied in real water distribution system for an optimized online prediction of residual chlorine. By incorporating the variability of source natural organic matter, along with operational actions and water demands, the proposed approach seeks to address a research

challenge to develop high fidelity and robust water quality predictions—well suited to providing operational decision support for optimized distribution system management.

Acknowledgments: This research is supported by the National Research Foundation, Singapore, and PUB, Singapore's National Water Agency under its Urban Solutions & Sustainability Competitive Research Program (Water) PUB-1804-0084. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and PUB, Singapore's National Water Agency.

References

1. George, K.; Fisher, I.; Sathasivan, A. Scan of water treatment processes to achieve desirable chlorine stability in water supply systems. *Process Saf. Environ. Prot.* **2017**, *112*, 265–273.
2. Zhang, C.; Li, C.; Zheng, X.; Zhao, J.; He, G.; Zhang, T. Effect of pipe materials on chlorine decay, trihalomethanes formation, and bacterial communities in pilot-scale water distribution systems. *Int. J. Environ. Sci. Technol.* **2017**, *14*, 85–94.
3. Ian, F.; Kastl, G.; Sathasivan, A.; Jegatheesan, V. Suitability of chlorine bulk decay models for planning and management of water distribution systems. *Crit. Rev. Environ. Sci. Technol.* **2011**, *41*, 1843–1882.
4. Ian, F.; Kastl, G.; Sathasivan, A. Evaluation of suitable chlorine bulk-decay models for water distribution systems. *Water Res.* **2011**, *45*, 4896–4908.
5. Powell, J.C.; Hallam, N.B.; West, J.R.; Forster, C.F.; Simms, J. Factors which control bulk chlorine decay rates. Water Research **2000**, *34*, 117–126.
6. Peirovi, M.R.; Afsharnia, M.; Moghaddam, A.; Ebrahimi, A.A.; Askarishahi, M.; Mokhtari, M. Calibration of water quality model for distribution networks using genetic algorithm, particle swarm optimization, and hybrid methods. *MethodsX* **2019**, *6*, 540–548.
7. Jeanne, H.J.; McBean, E.A. Using Bayesian statistics to estimate the coefficients of a two-component second-order chlorine bulk decay model for a water distribution system. *Water Res.* **2007**, *41*, 287–294.
8. Powell, J.C.; West, J.R.; Hallam, N.B.; Forster, C.F.; Simms, J. Performance of various kinetic models for chlorine decay. *J. Water Resour. Plan. Manag.* **2000**, *126*, 13–20.