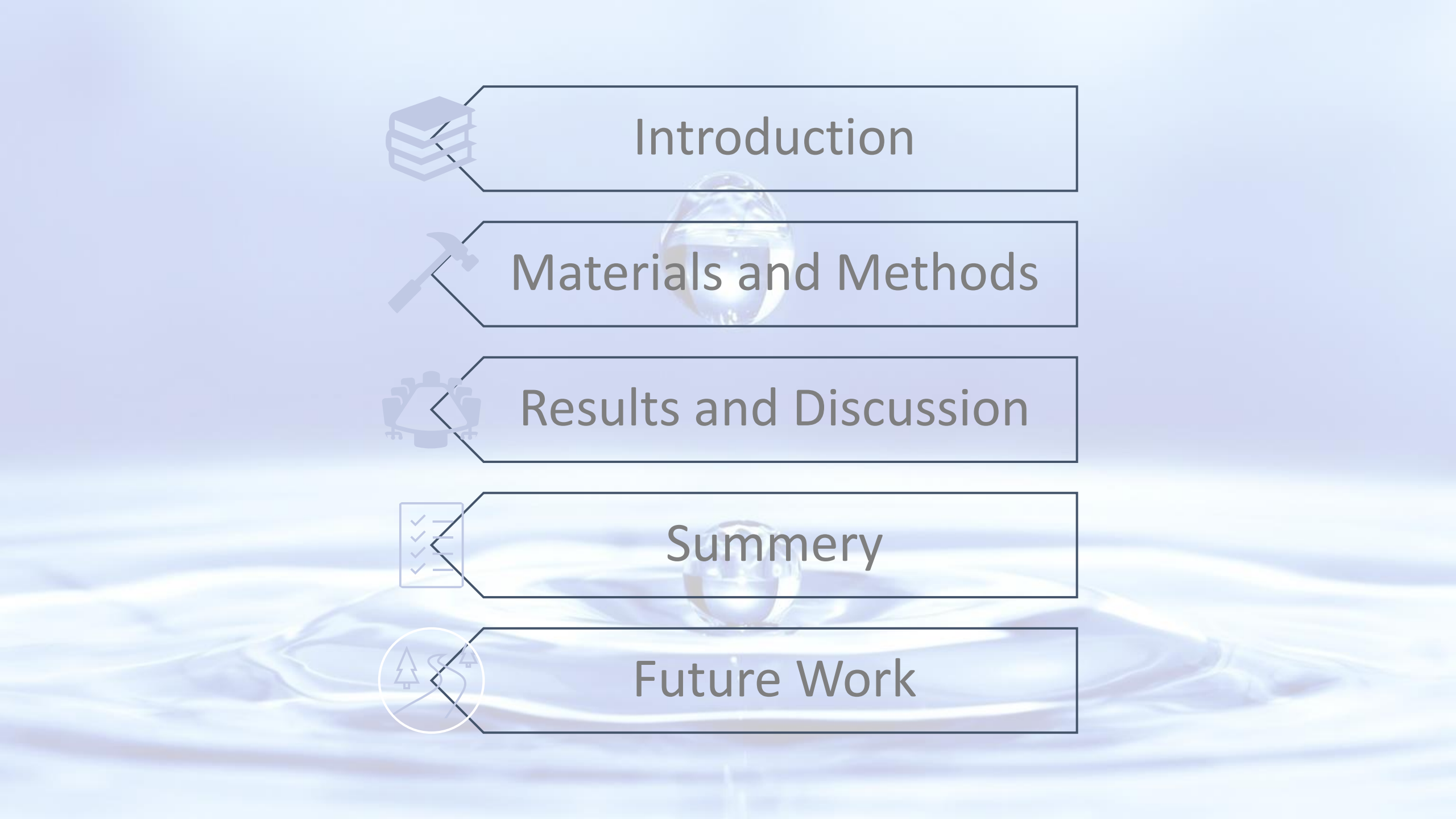
OPTIMIZED CHLORINE BULK DECAY MODELS AND A MACHINE-LEARNING-GUIDED WATER QUALITY RESPONSIVE KINETIC MODEL FOR RESIDUAL CHLORINE PREDICTION

Iman Jafari, Rongmo Luo, Fang Yee Lim, Ng Szu Hui and Hu Jiangyong* (*: corresponding author: ceehujy@nus.edu.sg)

National University of Singapore (NUS)

Department of Civil and Environmental Engineering

March 2022


Introduction
Materials and Methods
Results and Discussion
Summery
Future Work

Introduction

Water Disinfection

- ***Chlorine is one** of the most commonly used **disinfectants** for water **disinfection**.*

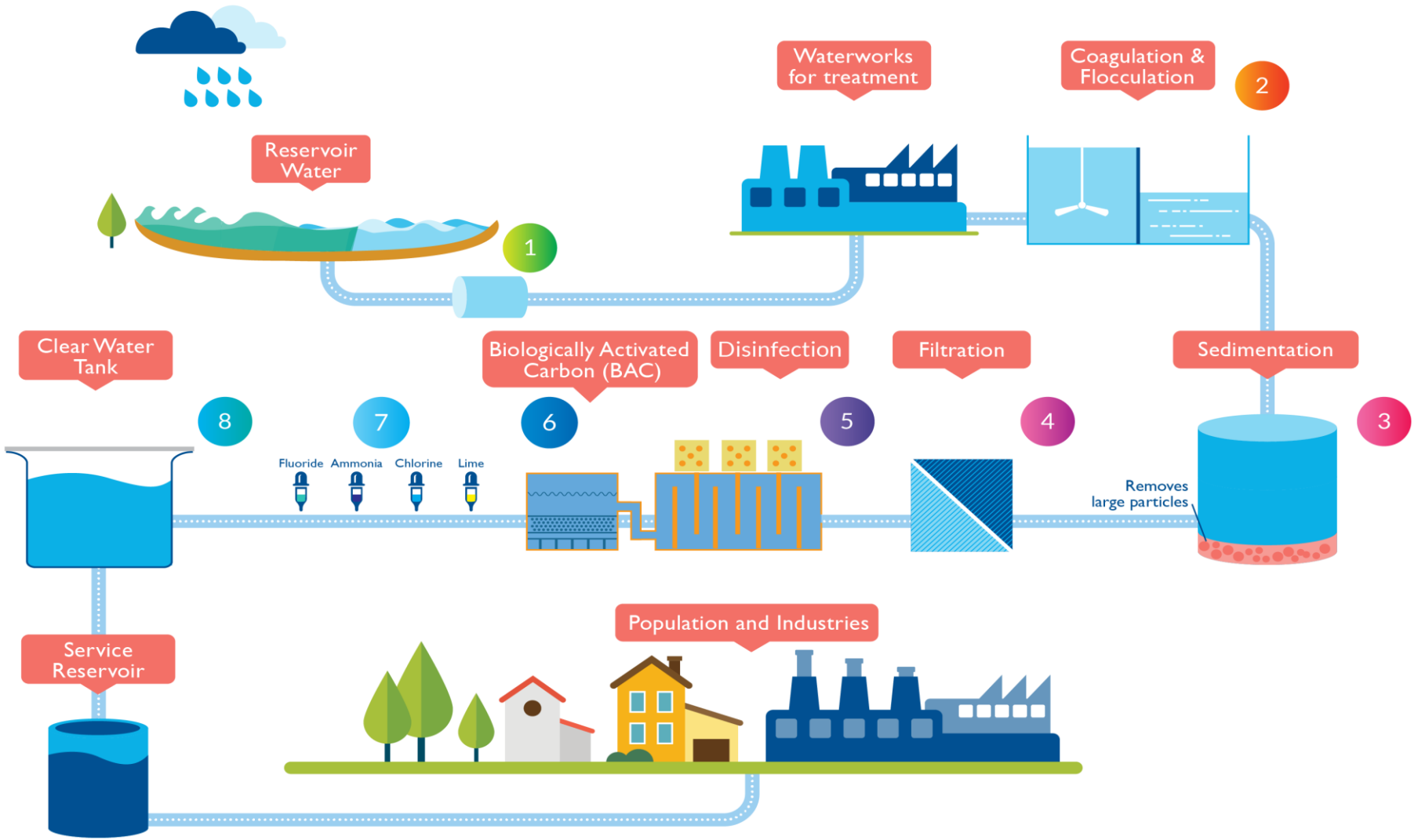
Chlorine Decay

- ***Bulk decay:*** Due to the reaction of chlorine with the substances available in the bulk water
- ***Wall Decay:*** Related to the reaction of the chlorine with the pipeline walls and the biofilms and corrosion products which are attached to them

Adding more chlorine to compensate its decay

- ***If the initial chlorine dosing concentration is too low:*** There may not be a residual left at the end of the distribution system to protect water against reoccurrence of any potential contamination
- ***If the dosing rate is too high:*** 1- Customer complaints, 2- Corrosion of the pipe network, 3- Formation of carcinogenic by-products

Introduction



Introduction

Question

❖ ***How to determine the right amount of chlorine that needs to be added to water before it leaves the treatment plant?***

Perform trials frequently

- Very ***time and resource consuming*** and thus is cost prohibitive.

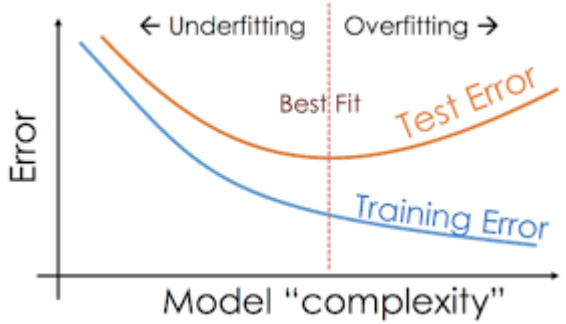
Mathematically model chlorine decay within distribution system

- The most cost-effective method to achieve the right chlorination regime using ***chlorine decay models***

Introduction

❖ The trade-off between model complexity and performance:

- More parameters can improve accuracy, but increase complexity and overfitting risk
- Overfitting leads to poor performance on new, unseen data
- A higher number of parameters may not necessarily lead to better performance on new data
- Additional parameters may fit to noise or random fluctuations, rather than underlying trends
- More complex models may not be practical or reliable for predicting new data
- Carefully consider the trade-off when choosing a model for a particular application



First Order Model (FOM)

Parallel First Order Model (PFOM)

Second Order Model (SOM)

Parallel Second Order Model (PSOM)

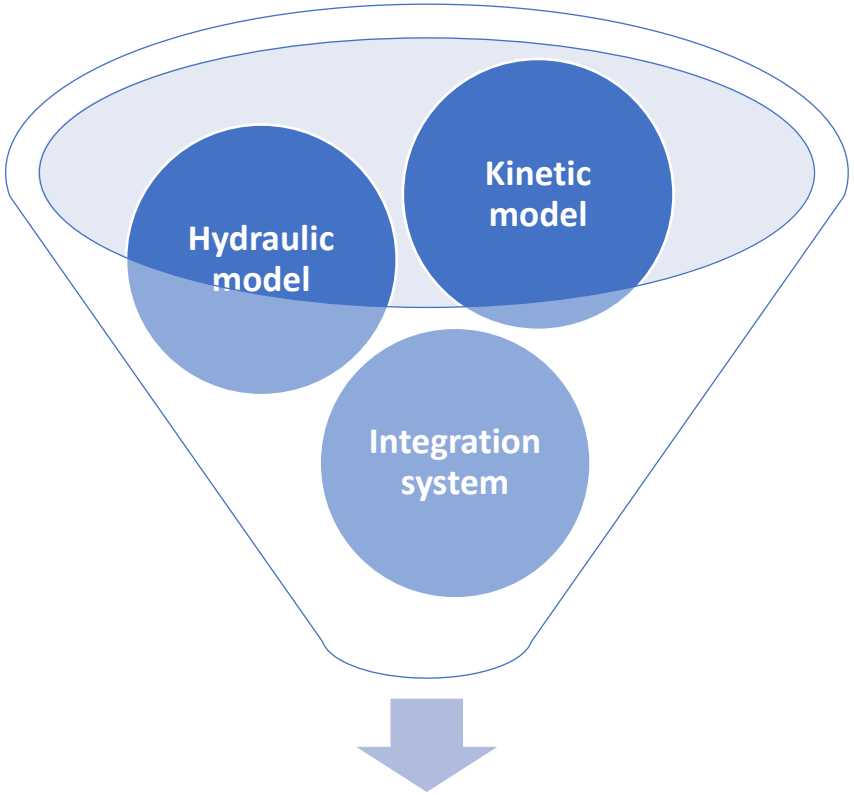
From simple to complex

Making a Balance

Complexity ⚖ ***Performance***

Introducing new version of chlorine bulk decay kinetic models by adding one additional parameter as "chlorine demand"

Introduction



Real-time water quality model

Rates of disinfectant decay

- Measuring in laboratory environment

Network hydraulics

- Only in time period of interest

They would not be robust to expected changes in source water

Prediction Of Disinfectant Residuals

Hydraulic Variability

-Production flows at all treatment sources
-Hydraulic heads at all treatment source points of entry
-Distribution storage tank or reservoir levels
-Transmission and distribution control valve status or setting
-Individual booster or high service pump status (on/off)
-Individual booster or high service VFD pump speed feedback
- etc.

Water Quality Variability

-Type of source water
-Initial chlorine concentration (ICC)
-Temperature
-pH
-NOM
-Age of infrastructure
- etc.

Introduction

Research Outlines

Modifying currently existing decay models from first order to parallel second order model

- Introducing a new parameter in the equations as “Chlorine Demand”
- Were assessed against their accuracy in explaining chlorine bulk decay behaviour in an attempt to correct the flaws of existing models
- Regarded as a noteworthy achievement that will add to the novelty of the current study

Predicting bulk decay coefficients based on water quality parameters through an analytical process instead of running bulk decay experiments in a laboratory environment

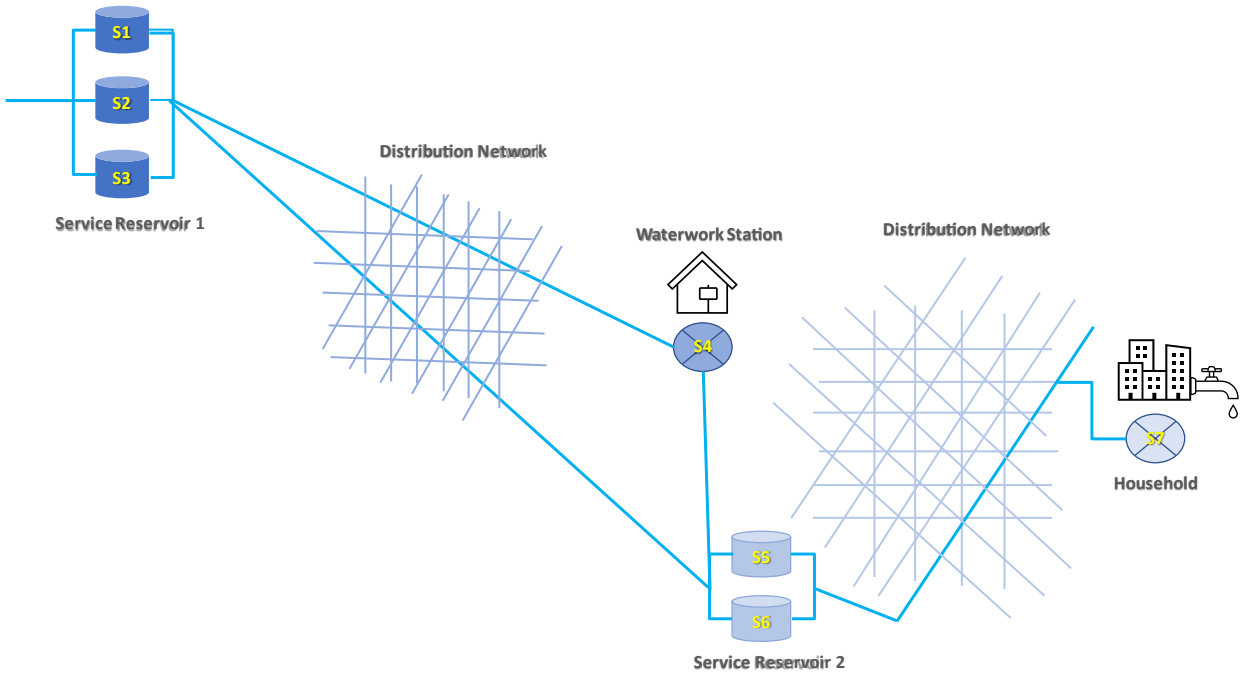
- To robust the predicting chlorine decay model to cover expected changes in source water, demands, or system operation over the ensuing weeks, months, or seasons
- The system assigns new values for the rates for each water sample with different water quality properties using a trained machine learning model

Implementation plan in real water distribution networks

- In this proposed methodology the system automatically updates the values for decay rate coefficients based on the changes in water quality parameters in an online manner

Materials and Methods

1. Sample Preparation:



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

Materials and Methods

2. Water Quality Parameter Measurements

Parameter	Unit	Method/Instrument
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)

❖ ***In this study the time intervals that were selected to measure the above-mentioned water quality parameters are 1h, 2h, 4h, 8h, 12h, 24h, 48h, 72h, 96h, and 168h after collection time***

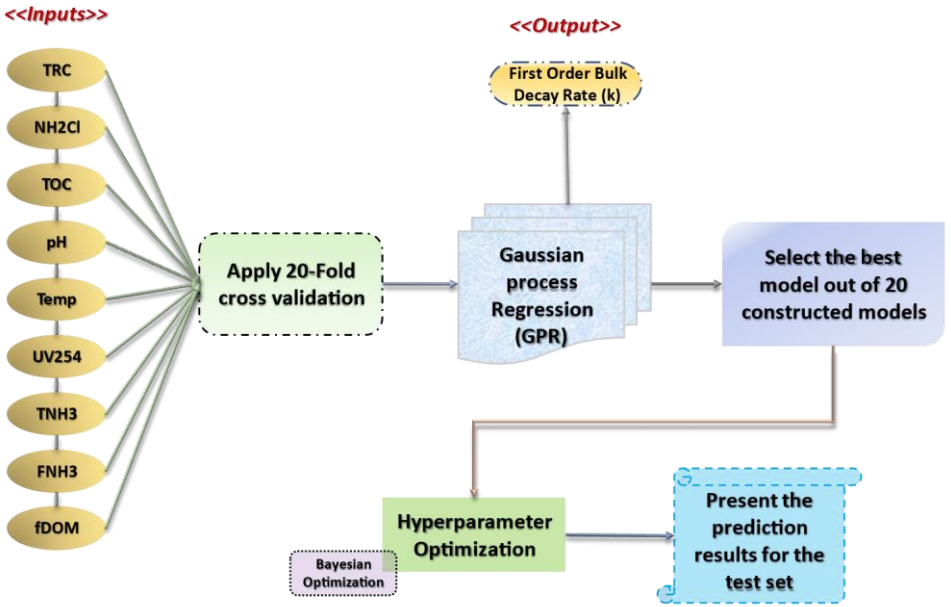
Materials and Methods

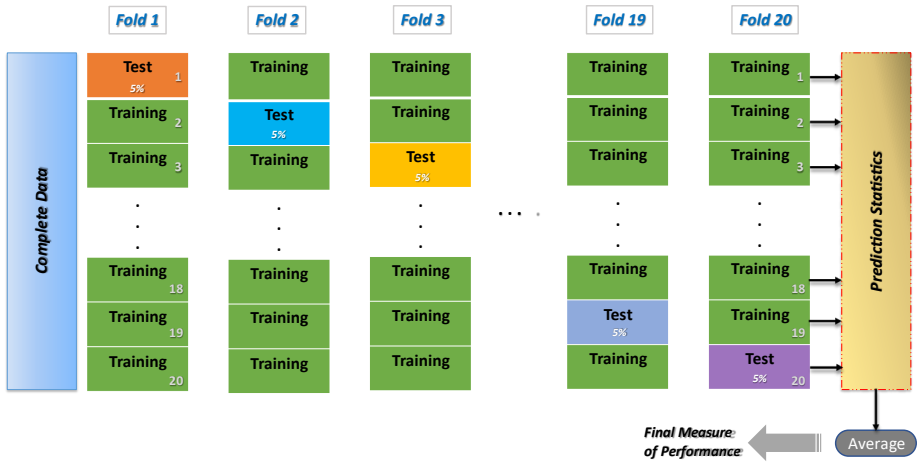
3. Modified Kinetic Bulk Decay Models

Model	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 $(1-R_2)$

Materials and Methods

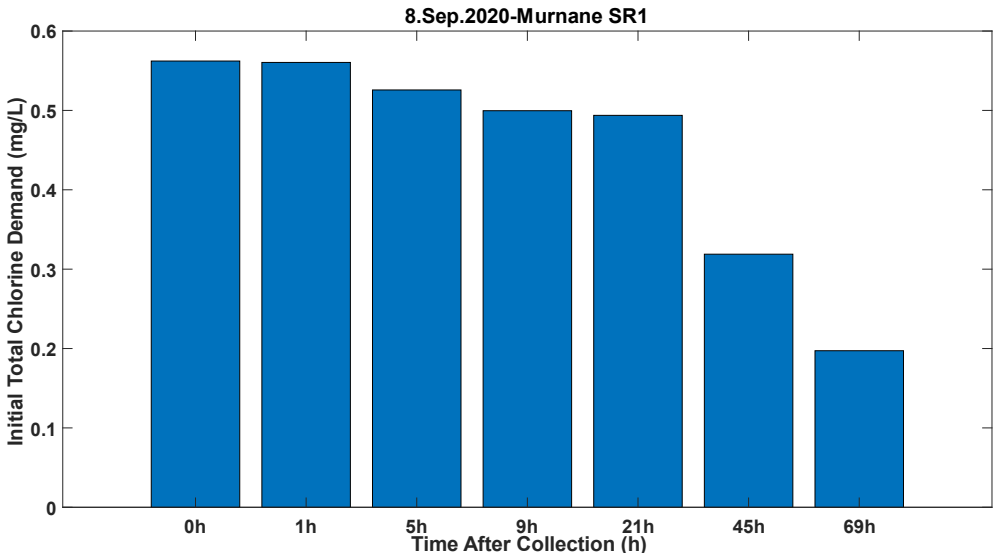
4. Responsive Kinetic Model





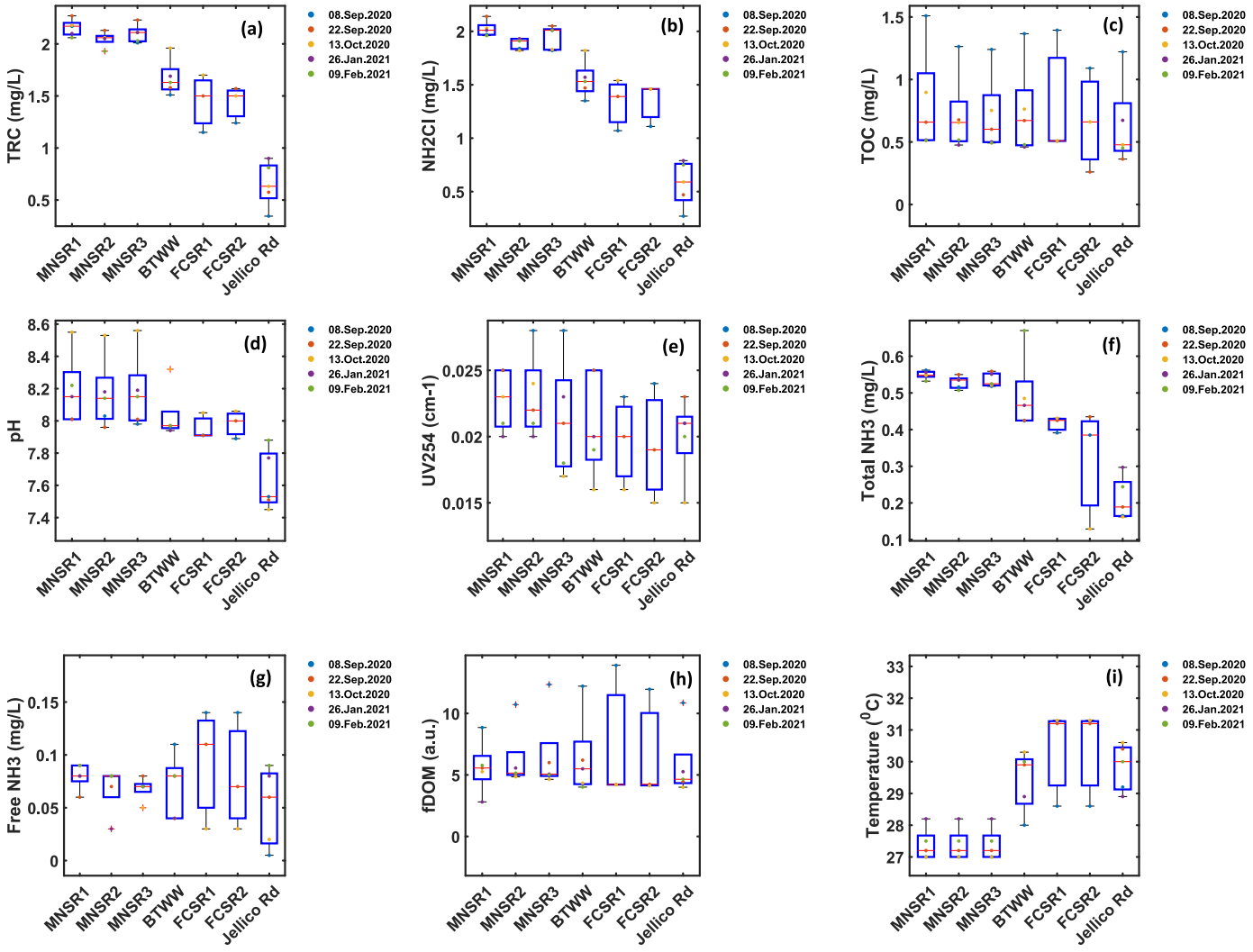
This technique could increase the number of datasets from 31 to 217

Time (h)	Total Cl_2	NH_2Cl	TOC	pH	uv254	Total NH_3	Free NH_3	fDOM	Temp	k_b	k_d	Cl_d
0	2.17	1.97	1.509	8.01	0.025	0.562	0.08	8.851	27	0.001658	0.010020	0.5622
1	2.16	2.05	1.491	8.01	0.026	0.533	0.09	11.162	27	0.001625	0.009649	0.5605
5	2.15	1.99	1.272	7.94	0.026	0.551	0.06	11.57	27	0.001638	0.010832	0.5258
9	2.14	2	1.047	7.79	0.021	0.465	0.005	9.442	27	0.001649	0.011887	0.4996
21	2.04	1.89	1.352	7.65	0.025	0.537	0.11	10.843	27	0.001310	0.007489	0.4938
45	1.98	1.81	1.308	7.57	0.011	0.515	0.11	9.214	27	0.001302	0.014377	0.3189
69	1.9	1.73	1.021	7.41	0.016	0.515	0.11	9.061	27	0.001103	0.025391	0.1972
93	1.81	1.69	0.944	7.38	0.016	0.505	0.005	8.915	27			
165	1.72	1.44	0.854	7.29	0.015	0.512	0.005	8.114	27			



Results and Discussion

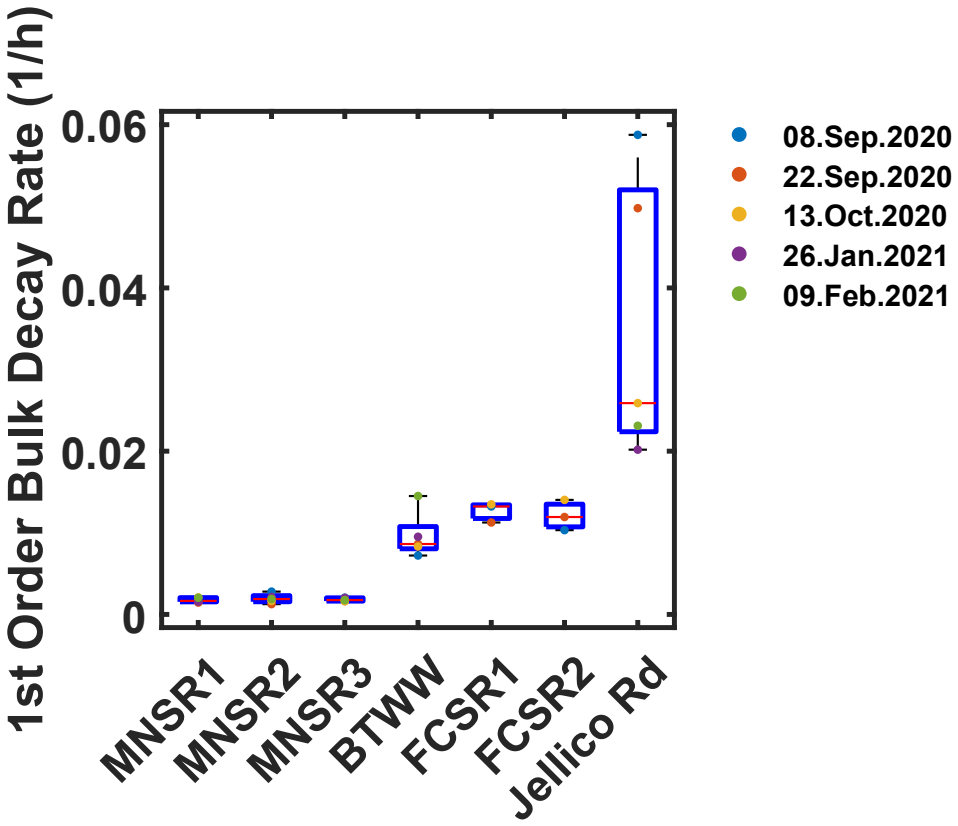
1. Data Visualization of Key Water Quality Parameters



- ✓ The level of total chlorine concentration in this part of distribution network is in an acceptable range recommended by many organizations such as American Water Works Association (AWWA).
- ✓ TOC levels lower than 2 mg/L is showing the full consistency with regulations that can reduce the formation of DBP to comply with the requirements for a safe tap water delivery
- ✓ The total residual chlorine is entirely contributed by monochloramines, and there is almost no contribution from free chlorine.
- ✓ This indicates that the entire distribution network has optimal chlorine dosage sufficient to oxidise all organic compounds whilst reacting with all free ammonia present, leaving the latter remaining to be as close to 0
- ✓ The total free ammonia concentration depletes further down the distribution network, which reinforces the fact that whatever free chlorine present (residual free chlorine) was used to react with ammonia to form monochloramines

Results and Discussion

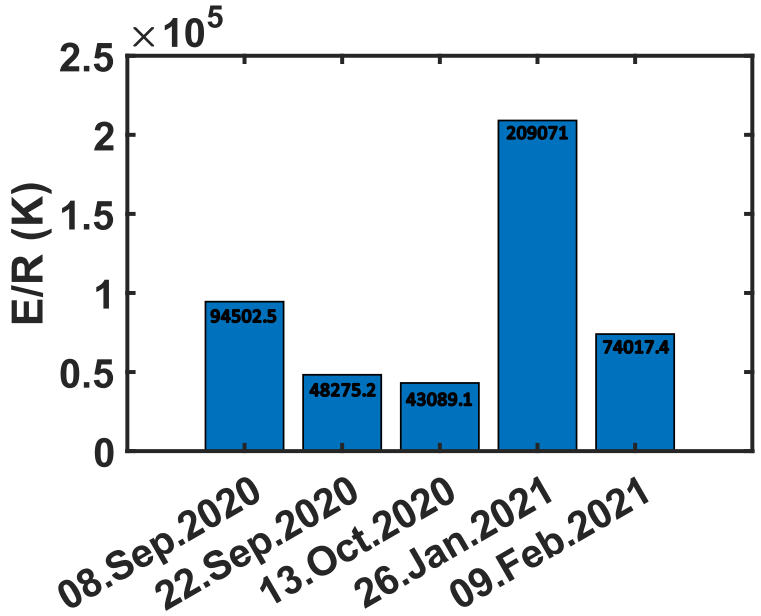
2. Bulk Decay Experiments



❖ The bulk decay rate was generally increasing by moving from MNSR towards Jellico Rd

The major reason of this increment in bulk decay rate coefficients can be explained by the Arrhenius theory (Equation 3-1) due to the increase in temperature by moving from MNSR towards Jellico Rd and also the reverse effect of initial chlorine on bulk decay rate as TRC is decreasing from MNSR to Jellico Rd.

$$K_T = K_{base} * e^{\frac{-\frac{E}{R}*(base-T)}{(273+base)*(273+T)}} \quad \textit{(Equation 3-1)}$$



Assumption: Water samples collected in the same day having the same activation coefficients

E/R can be estimated for each day by considering the lowest temperature of the collected samples as the base temperature and estimating the bulk decay rate of this sample with the lowest temperature and also another sample with different temperature.

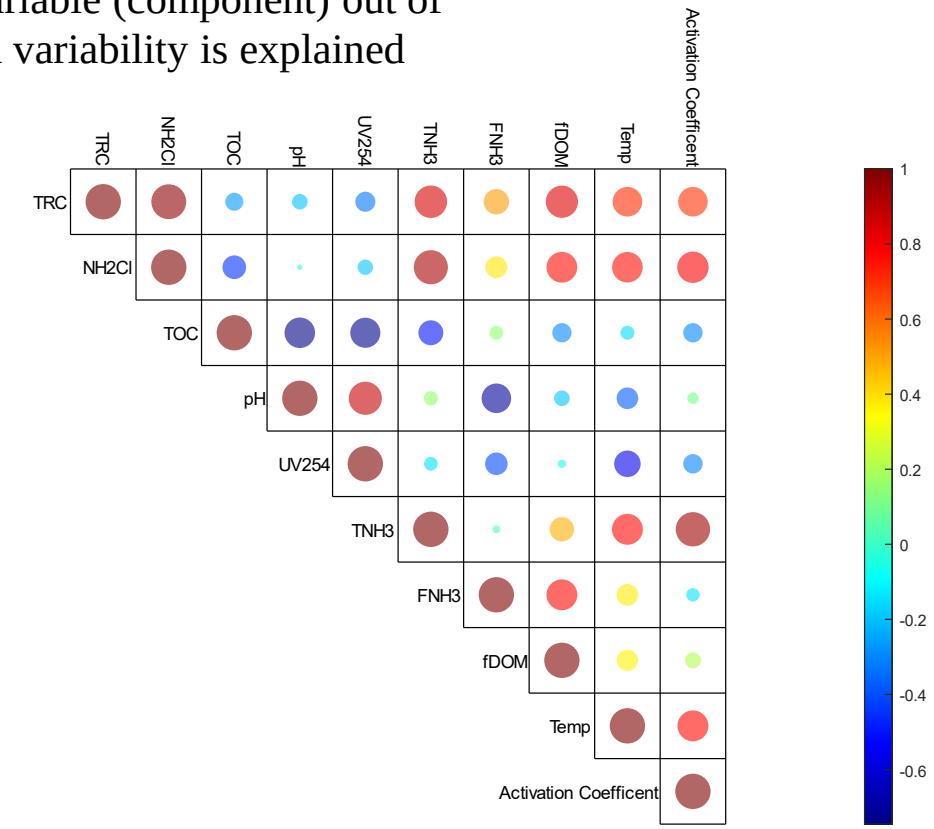
*Ababu T. Tiruneh, Tesfamariam Y. Debessai, Gabriel C. Bwembya, Stanley J. Nkambule, "A Mathematical Model for Variable Chlorine Decay Rates in Water Distribution Systems", *Modelling and Simulation in Engineering*, vol. 2019, Article ID 5863905, 11 pages, 2019. https://doi.org/10.1155/2019/5863905

Results and Discussion

2. Bulk Decay Experiments

- To explain the variations in activation coefficients (E/R) estimated for 5 different days, the correlation between E/R with different water quality parameters was calculated
- A parallel Component Analysis (PCA) was carried out to obtain a single variable (component) out of seven variables for each water quality parameter by which most of the data variability is explained

	TRC	NH_2Cl	TOC	pH	UV254	TNH_3	FNH_3	fDOM	Temp
	Percentage of data variability explained by each component								
Comp1	83.86	90.36	89.33	73.42	83.52	79.48	68.25	68.00	97.89
Comp2	13.44	6.89	7.02	23.78	10.45	16.36	16.33	21.92	2.11
Comp3	2.44	2.36	3.14	1.58	5.60	3.81	10.49	6.14	0.00
Comp4	0.17	0.38	0.49	1.06	0.42	0.31	4.89	3.91	0.00
Comp5	0.08	0.00	0.01	0.16	0.00	0.03	0.05	0.03	0.00



- ✓ As can be seen, there is a high correlation between activation coefficients with TRC, NH2Cl, TNH3, and temperature with correlation coefficients equal to 0.6995, 0.7883, 0.9450 and 0.7637, respectively.

Results and Discussion

2. Bulk Decay Experiments

- ❖ The exponential term in the Arrhenius equation implies that the rate constant of a reaction increases exponentially when E/R decreases.
- ✓ Accordingly, it can be concluded that bulk decay coefficients are inversely proportional to TRC, NH_2Cl, TNH_3 as those water qualities have a significant positive linear correlation with E/R.
  - However, the significant positive correlation between temperature and E/R is not fundamentally valid as the activation energy is not affected by temperature.
- ❖ To confirm these results, the Pearson correlation between bulk decay rate coefficients with water quality parameters were also reported:

i. ***The results are consistent with the previous results***

ii. ***pH also showed a significant negative correlation with bulk decay rate coefficient:*** can be due to the higher concentration of mono-chloramine in the pH levels >7 which is contributing to lower reaction rates of TRC in bulk water.

iii. ***Temperature also had a significant positive linear correlation with bulk decay rate as expected.***

Results and Discussion

3. Modified bulk Decay Kinetic Models

Collection Date	Site	First Order Model (FOM)						
		Original			Modified			
		k	MSE	R	TCD_0	k_d	MSE	R
08.Sep.2020	MNSR1	0.0017	0.0014	0.9774	0.5622	0.0100	0.0002	0.9963
	MNSR2	0.0028	0.0040	0.9760	0.7480	0.0124	0.0008	0.9934
	MNSR3	0.0021	0.0022	0.9767	0.6379	0.0097	0.0010	0.9846
	BTWW	0.0072	0.0037	0.9880	1.0662	0.0131	0.0014	0.9939
	FCSR1	0.0132	0.0012	0.9957	1.0180	0.0168	0.0006	0.9977
	FCSR2	0.0103	0.0015	0.9951	1.2570	0.0101	0.0014	0.9952
	Jellico Rd	0.0588	0.0008	0.9899	0.3055	0.0825	0.0002	0.9927

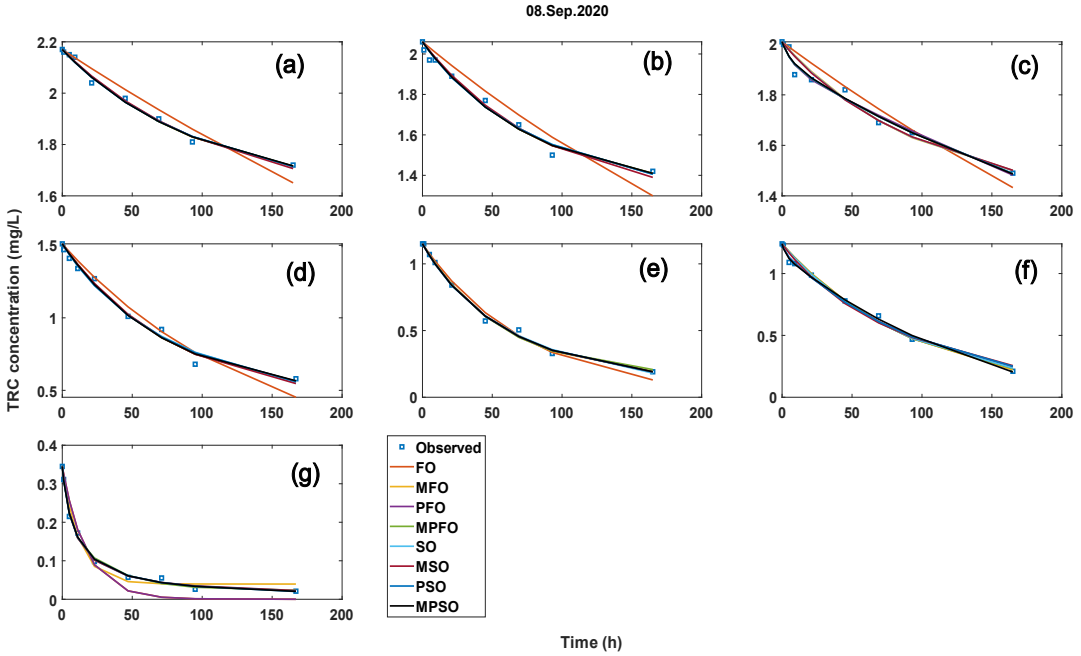
Collection Date	Site	Parallel First Order Model (PFOM)										
		Original					Modified					
		x	k_1	k_2	MSE	R	x	k_{d1}	k_{d2}	TCD_0	MSE	R
08.Sep.2020	MNSR1	0.2637	0.0097	0.0000	0.0002	0.9962	0.4073	0.0100	0.0000	1.3815	0.0002	0.9963
	MNSR2	0.6369	0.0000	0.0124	0.0008	0.9934	0.5347	0.0000	0.0126	1.5997	0.0008	0.9934
	MNSR3	0.9517	0.0015	0.1219	0.0006	0.9898	0.0719	0.1576	0.0038	1.0424	0.0006	0.9902
	BTWW	0.7061	0.0131	0.0000	0.0014	0.9939	0.2486	0.0135	0.0129	1.0704	0.0014	0.9940
	FCSR1	0.4257	0.0063	0.0226	0.0005	0.9979	0.5000	0.0176	0.0176	0.9973	0.0006	0.9977
	FCSR2	0.0478	0.5253	0.0095	0.0007	0.9970	0.9880	0.0106	0.0090	1.2208	0.0015	0.9949
	Jellico Rd	0.5000	0.0588	0.0588	0.0008	0.9899	0.4659	0.2104	0.0315	0.3225	0.0001	0.9977

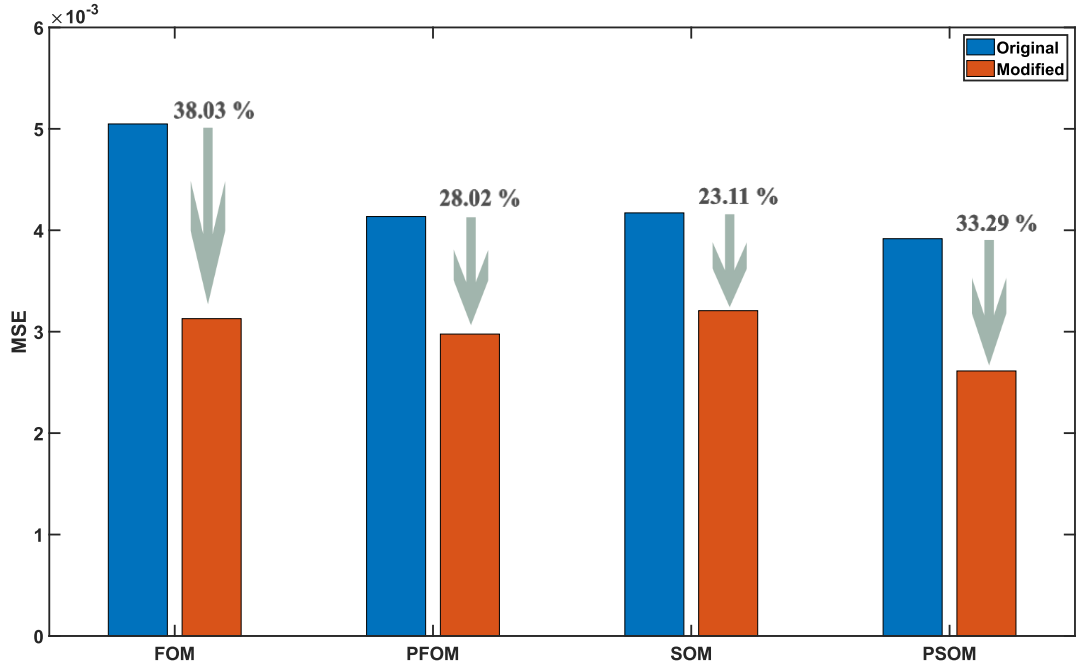
Collection Date	Site	Parallel Second Order Model (PSOM)														
		Original							Modified							
		Z	R_1	R_2	K_1	K_2	MSE	R	TCD_0	Z	R_1	R_2	K_{d1}	K_{d2}	MSE	R
08.Sep.2020	MNSR1	0.3646	0.2181	0.3192	0.0105	0.0092	0.0002	0.9962	4.9147	0.5692	0.1177	0.1154	0.0101	0.0101	0.0002	0.9963
	MNSR2	0.7929	0.4006	0.4006	0.0121	0.0121	0.0008	0.9928	8.9858	0.5673	0.0863	0.0805	0.0128	0.0125	0.0008	0.9932
	MNSR3	0.5788	2.8707	0.0873	0.0011	0.1626	0.0006	0.9901	1.6058	0.7777	2.2945	0.1994	0.0013	0.1691	0.0006	0.9899
	BTWW	0.2261	0.0000	2.3667	0.0027	0.0055	0.0015	0.9933	18.220	0.5141	0.0586	0.0586	0.0133	0.0135	0.0014	0.9938
	FCSR1	0.3632	1.2680	6.6334	0.0073	0.0029	0.0005	0.9979	4.8686	0.3265	0.0975	0.2779	0.0248	0.0159	0.0005	0.9979
	FCSR2	0.0448	14.420	14.734	0.0388	0.0007	0.0008	0.9967	1.6084	0.0514	12.275	6.3691	0.0344	0.0010	0.0004	0.9982
	Jellico Rd	0.3421	1.3619	1.0100	0.1865	0.0580	0.0001	0.9981	0.3345	0.2725	1.3920	1.1259	0.2104	0.0608	0.0001	0.9981

Collection Date	Site	Second Order Model (SOM)								
		Original				Modified				
		C_{A0}	k	MSE	R	TCD_0	C_{A0}	k_d	MSE	R
08.Sep.2020	MNSR1	0.6116	0.0044	0.0002	0.9962	0.6041	20.178	0.0004	0.0002	0.9960
	MNSR2	0.8260	0.0058	0.0008	0.9928	0.8004	76.527	0.0001	0.0008	0.9938
	MNSR3	0.7041	0.0046	0.0009	0.9850	0.9293	0.9371	0.0078	0.0009	0.9852
	BTWW	1.5358	0.0067	0.0015	0.9930	1.1069	76.992	0.0002	0.0014	0.9940
	FCSR1	2.0702	0.0080	0.0006	0.9978	1.0586	4.1442	0.0042	0.0005	0.9979
	FCSR2	9.8056	0.0011	0.0014	0.9948	1.6149	1.6137	0.0059	0.0015	0.9939
	Jellico Rd	0.3374	0.3168	0.0001	0.9980	0.3407	0.3393	0.3224	0.0001	0.9979

Results and Discussion

3. Modified bulk Decay Kinetic Models





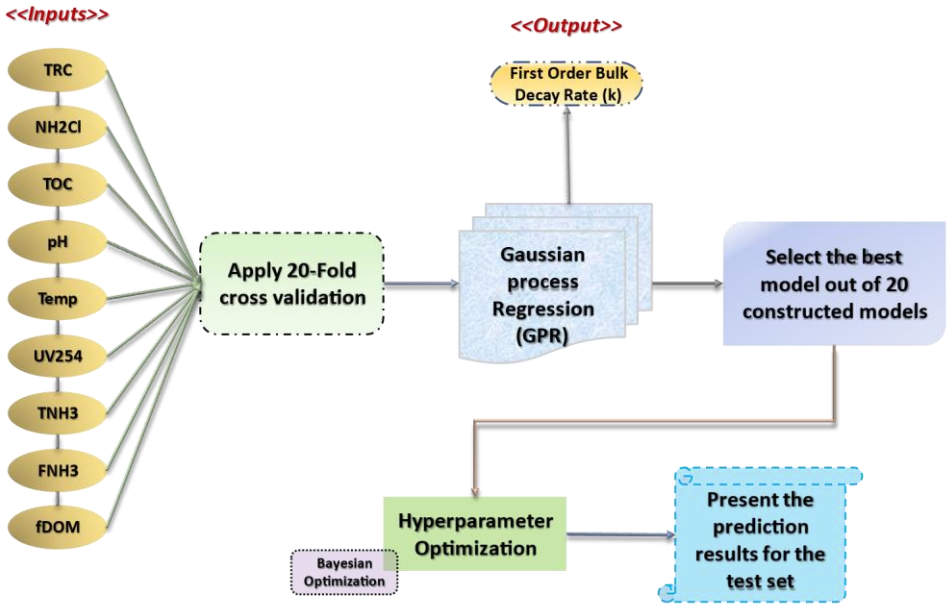
- As can be seen, generally, all of the models are performing almost similar except FOM which is showing lower accuracy in fitting the chloring decay data

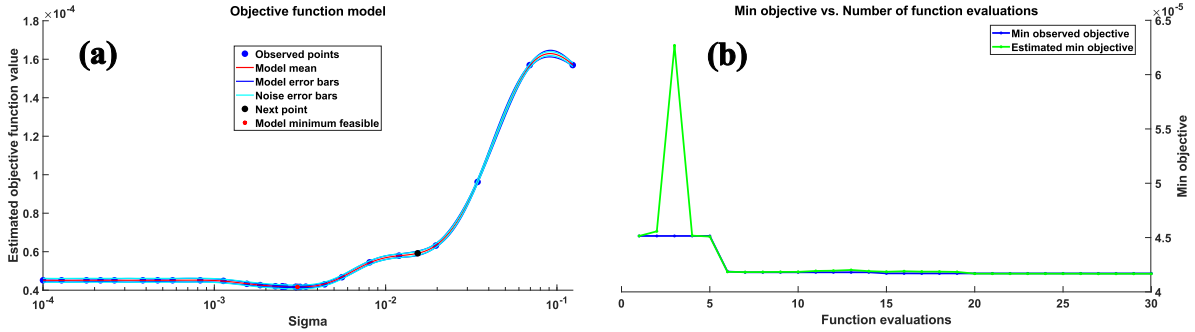
- Although FOM is very simple to use and therefore it has always been popular, it has not provided a good data fitting for various chlorine decay data and modelling applications.

✓ *However, as can be seen, adding only one additional parameter to the original FOM equation as initial total chlorine demand (TCD_0) can considerably increase its accuracy in MFOM so that the performance of this widely used and simple model will be similar to other existing models with higher complexity.*

Results and Discussion

4. Responsive Bulk Decay Kinetic Model



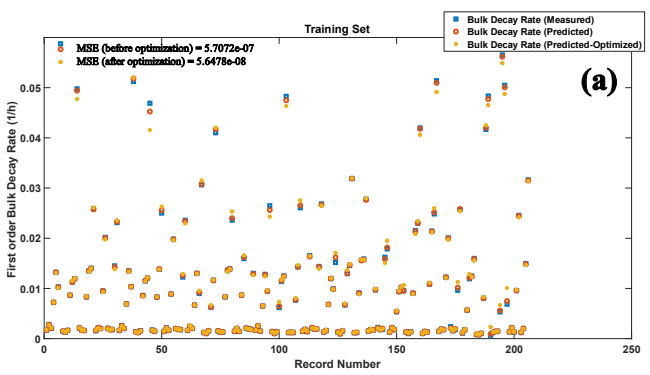


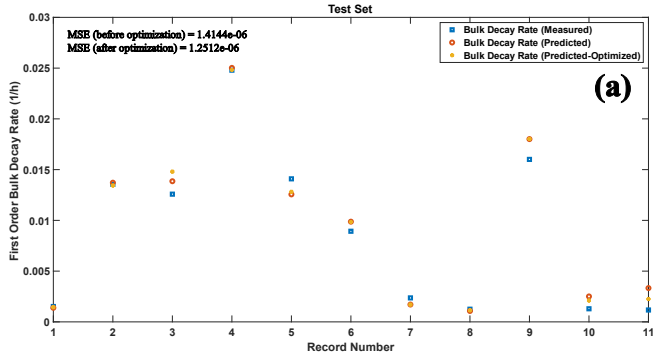
Fold Number	Number of data in training set	Number of data in test set	Correlation coefficient (Training)	MSE (Training)	Correlation coefficient (Test)	MSE (Test)
1	207	10	0.99982	5.91E-08	0.75361	2.80E-05
2	206	11	0.99983	5.65E-08	0.98028	5.45E-06
3	206	11	0.99982	6.06E-08	0.98242	1.50E-05
4	206	11	0.99982	5.89E-08	0.96544	9.91E-06
5	206	11	0.99978	7.02E-08	0.85615	6.70E-05
6	206	11	0.99983	5.32E-08	0.96840	2.02E-05
7	206	11	0.99983	5.86E-08	0.59104	2.79E-05
8	206	11	0.99981	6.01E-08	0.87296	8.93E-05
9	206	11	0.99983	5.65E-08	0.97362	3.90E-06
10	206	11	0.99983	5.84E-08	0.90646	1.34E-05
11	206	11	0.99985	5.15E-08	0.99063	3.58E-06
12	206	11	0.99986	4.93E-08	0.98703	4.08E-06
13	206	11	0.99983	5.64E-08	0.95560	2.78E-06
14	206	11	0.99980	6.41E-08	0.96299	2.47E-05
15	206	11	0.99980	6.31E-08	0.83063	7.68E-05
16	206	11	0.99977	7.04E-08	0.74933	1.36E-04
17	206	11	0.99983	5.37E-08	0.97595	1.55E-05
18	206	11	0.99983	5.65E-08	0.99002	1.41E-06
19	207	10	0.99981	5.85E-08	0.94507	3.13E-05
20	207	10	0.97390	7.45E-06	0.66907	1.71E-04
Average	206	11	0.998524	4.28E-07	0.895335	3.74E-05

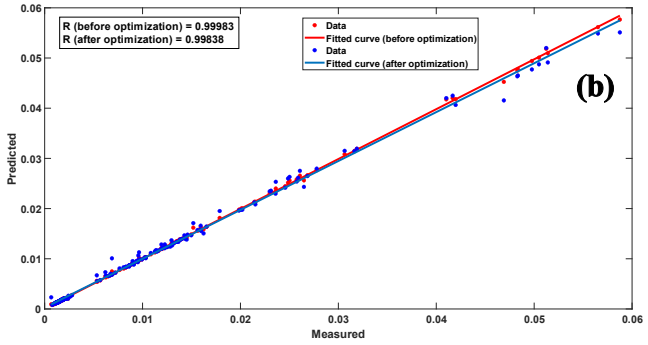
Results and Discussion

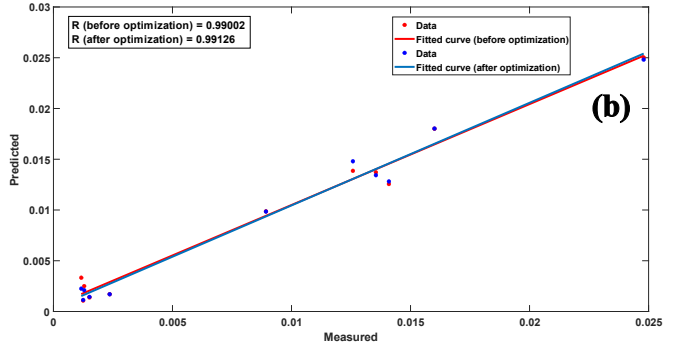
3. Modified bulk Decay Kinetic Models

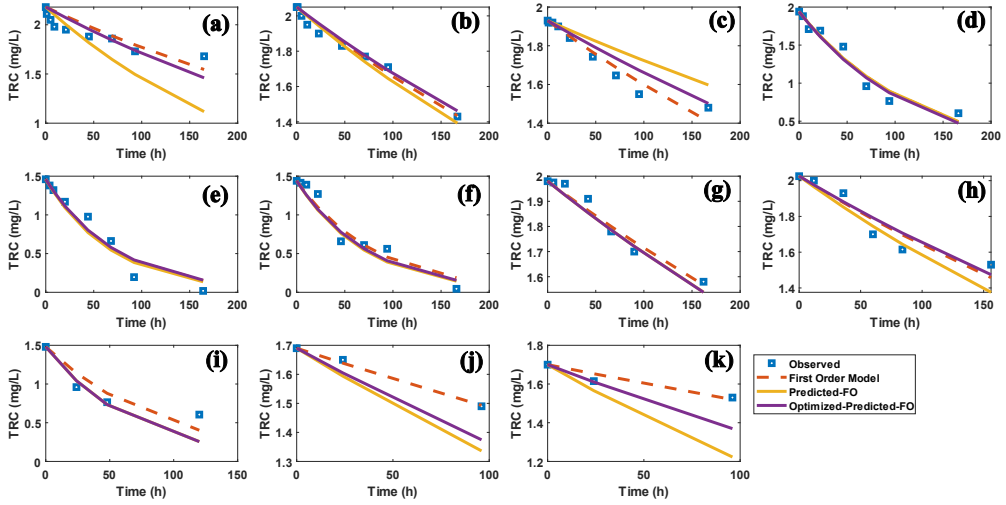
Test Set	FOM		Predicted-FOM		Optimized-Predicted-FOM	
	R	MAE	R	MAE	R	MAE
a	0.9783	0.0316	0.9587	0.0342	0.9588	0.0346
b	0.9910	0.0380	0.9754	0.0568	0.9761	0.0427
c	0.9729	0.0350	0.9597	0.1352	0.9560	0.0684
d	0.9550	0.0269	0.9885	0.0280	0.9885	0.0278
e	0.9989	0.0147	0.9707	0.0268	0.9702	0.0203
f	0.9591	0.0354	0.9873	0.0588	0.9874	0.0521
g	0.9831	0.0192	0.9518	0.0330	0.9518	0.0235
h	0.9945	0.0238	0.9800	0.0509	0.9797	0.0310
i	0.9196	0.0262	0.9906	0.0304	0.9905	0.0296
j	0.9837	0.0021	0.9970	0.1320	0.9973	0.1106
k	0.9639	0.0063	0.9694	0.1334	0.9667	0.0849
Average	0.9727	0.0236	0.9754	0.0654	0.9748	0.0478





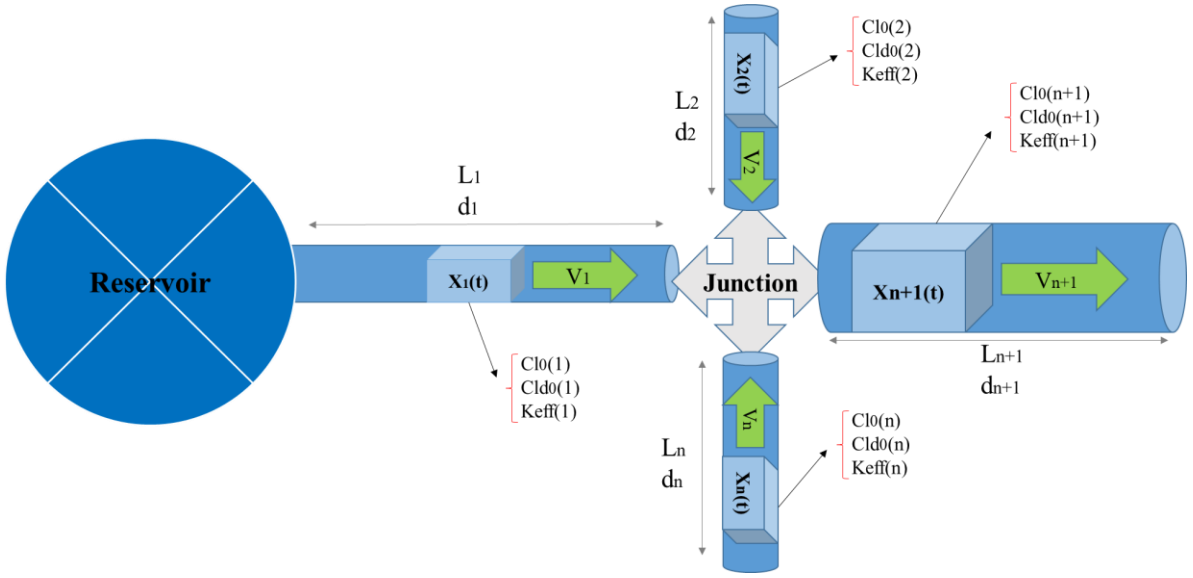






Results and Discussion

4. Implementation in Real Water Distribution Networks



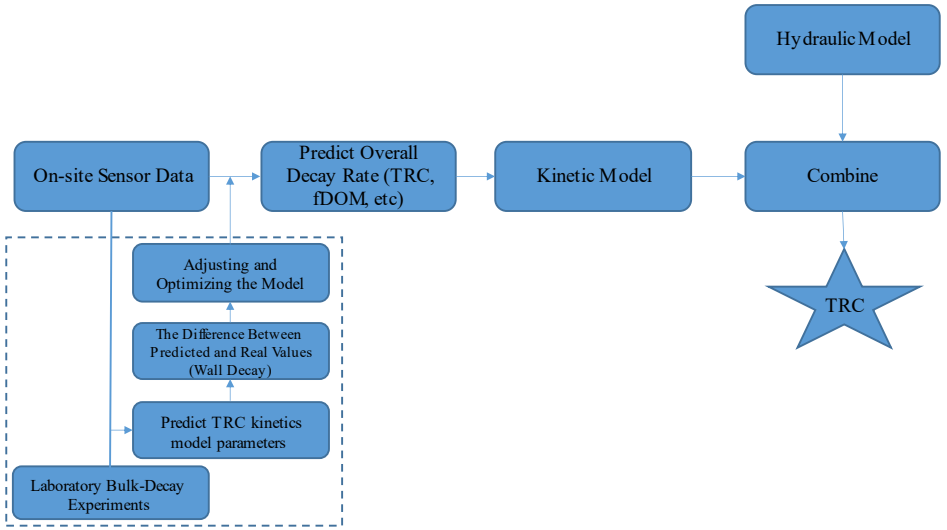
$$Kb1(n+1) = \frac{\sum_{i=1}^{n} Qi*Kb1i}{\sum_{i=1}^{n} Qi} \qquad Cl_0(n+1) = \sum_{i=1}^{n} Cli\left(\frac{Li}{Vi}\right)$$

$$Kb2(n+1) = \frac{\sum_{i=1}^{n} Qi*Kb2i}{\sum_{i=1}^{n} Qi} \qquad NOM_{10}(n+1) = \sum_{i=1}^{n} NOM10i\left(\frac{Li}{Vi}\right)$$

$$K1(n+1) = \frac{\sum_{i=1}^{n} Qi*K1i}{\sum_{i=1}^{n} Qi} \qquad NOM_{20}(n+1) = \sum_{i=1}^{n} NOM20i\left(\frac{Li}{Vi}\right)$$

$$K2(n+1) = \frac{\sum_{i=1}^{n} Qi*K2i}{\sum_{i=1}^{n} Qi}$$

- $\frac{Li}{Vi}$ is a time duration that water segment (i) travels it's corresponding pipeline length until the junction.



❖ *In this proposed methodology the system automatically updates the values for the parameters in kinetic model based on the changes in water quality*

Summery

All water quality parameters in the studied portion of distribution network were in an optimal range to maintain safe and high-quality water and preserving drinking water quality from the point-of-entry to the point-of-use

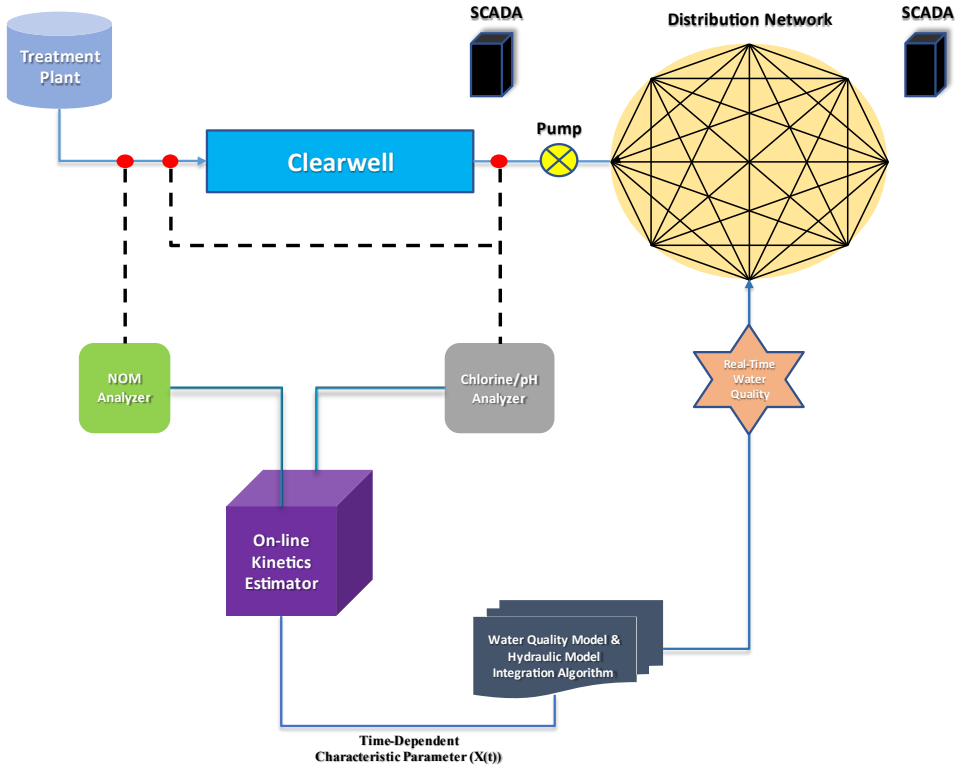
Bulk decay coefficients were inversely proportional to TRC, NH_2Cl, TNH_3, pH and directly proportional to Temperature

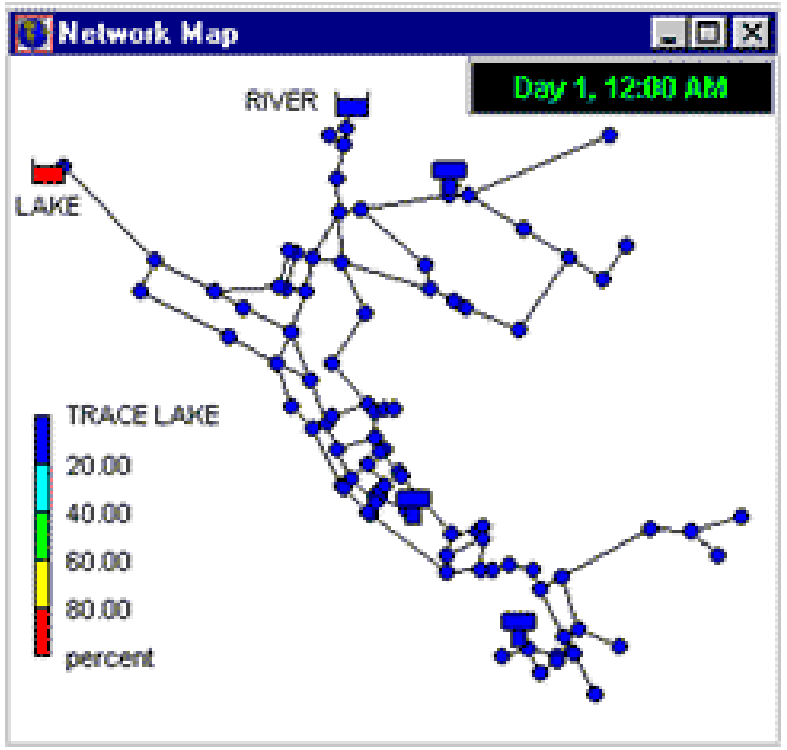
By applying the proposed modification in this study on all four nominated models, the MSE values were decreased by 38.03%, 28.02%, 23.11%, and 33.29% for FOM, PFOM, SOM and PSOM, respectively.

A new methodology is 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.

Future Work

Robust Prediction of Disinfectant Degradation in Drinking Water Distribution Systems Integrating Water Quality Sensing and Digital Twin Technologies






No. :
Date :
Thank You
For Your
Attention