KYAMBOGO UNIVERSITY

DIRECTORATE OF RESEARCH AND GRADUATE TRAINING

COMPARISON OF PHYSICAL AND STATISTICAL MODELS IN PREDICTING SPACE-TIME DECAY OF RESIDUAL CHLORINE IN WATER DISTRIBUTION SYSTEM

BY

JULIUS CAESAR KWIO-TAMALE
B.Eng. CBE (KyU)

18/U/GMEW/22153/PD

A DISSERTATION SUBMITTED TO KYAMBOGO UNIVERSITY DIRECTORATE OF RESEARCH AND GRADUATE TRAINING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN WATER AND SANITATION ENGINEERING DEGREE OF KYAMBOGO UNIVERSITY

October 2022

DECLARATION

I, *Julius Caesar Kwio-Tamale*, hereby declare that this submission is my own work and that to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree of the university or other institute of higher learning except where due acknowledgement has been made in the text and reference list.

Julius Caesar Kwio-Tamale

Signed: .... Date: 8-11-2022

APPROVAL

The undersigned approve that they have read and hereby recommend for submission to Directorate of Research and Graduate Training of Kyambogo University a dissertation entitled "Comparison of Physical and Statistical Models in Predicting Space-Time Decay of Residual Chlorine in Gravity Water Distribution Systems:" in partial fulfillment of the requirements for the award of Master of Science in Water and Sanitation Engineering Degree of Kyambogo University.

Dr. Charles Onyutha .... Supervisor

Date: 08.11.2022

Dr. Anne Nakagiri .... Supervisor

Date: ....-.

ABSTRACT

Chlorine is the most widely used disinfectant in water distribution due to its efficacy, ease of application, low cost and extended disinfection durability. The World Health Organization recommends concentrations of residual chlorine in drinking water to be within 0.2 – 5 mg/l. Concentrations lower than 0.2 mg/l expose water consumers to secondary water-borne diseases. Chlorine concentrations more than 5 mg/l expose consumers to carcinogenic disinfection by-products. Studies on comparative analysis of performance of physical and statistical models in predicting chlorine decay in drinking water distribution system are lacking. The specific objectives of this study were: (1) characterization of residual chlorine decay parameters in water distribution, (2) assessment of space-time decay of chlorine in water distribution, (3) comparison of performance of models in predicting chlorine decay in water distribution and (4) identification of appropriate model(s) for predicting chlorine decay in water distribution system. Performance of EPANET physical model was compared with statistical models of multiple linear regression (MLR), principal component regression (PCR), lasso regression (LR), ridge regression (RR), decision tree (DT), random forest (RF) and artificial neural network (ANN). ANN performed best with R^2 of 94% followed by MLR (63%), PCR (61%), RF (55%) and DT (41%). Initial chlorine and electrical conductivity were the two most significant parameters in water distribution that together contributed to about 90% of chlorine decay. Based on generalizability, dimensionality control and interpretability as desired factors for a good model, linear regression with R-squared of 63% and 0.045 mg/l error estimate performed best in predicting residual chlorine. Water zoning is recommended with existing water reservoirs as secondary chlorination points to maintain residual chlorine concentrations within 0.2 – 5 mg/l. In return, high and low dosages that cause carcinogenic disinfection by-products and predispose public health to secondary pathogenic infectious water-borne diseases respectively will both be avoided throughout water distribution network.

ACKNOWLEDGEMENT

I am grateful to the Government of Uganda through Kyambogo University in financing this research. I thank the Head Office of the National Water and Sewerage Corporation (NWSC) for granting permission to undertake this research on Lirima gravity flow scheme which they own and operate. Equally, I thank the NWSC Tororo office for its field support through assigning its key staff of water quality technician and plumbers who assisted in identifying water sampling points, water sampling and water quality testing both online and in laboratory. I am also grateful to my supervisor Dr. Charles Onyutha for his guidance in planning and guiding the research process from its start to successful completion. I thank Eng. Denis Alunyu of Manafwa District for his support in organizing some of the key logistics like hand-held GPS toolkit, field transport and also participating in pre-field planning meetings in Tororo.

DEDICATION

To my family, late parents and late uncles who contributed morally and financially to my education that has made me what I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

Abbreviations

ANN	Artificial neural network
BPT	Break pressure tank
CRC	Constant rate coefficient
DBP	Disinfection by-product
DDA	Demand driven analysis
DWD	Directorate of Water Development (of Uganda)
EC	Electrical conductivity
GPS	Geographical position(ing) system
GRNN	Generalized regression neural network
GV	Gate valve
GWF(S)	Gravity water flow (scheme)
HDPE	High density polyethylene
ICC	Initial chlorine concentration
KMO	Kaiser-Meyer-Olkin (statistics or coefficient)
LASSO or lasso	Least Absolute Shrinkage and Selection Operator
MAE	Maximum Absolute Error
MLP	Multi-Layer Perceptron
MLR	Multi-Linear (or Multi-variate) Regression
MWE	Ministry of Water and Environment (of Uganda)
NTU	Nephelometric turbidity units
NWSC	National Water and Sewerage Corporation
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
pH	Hydrogen ion potential
PVC	Polyvinyl chloride
RMSE	Root Mean Square Error

SR	Single Reactant
Temp	Temperature
UTM	Universal Time Mercator
WDN	Water Distribution Network
WTN	Water Transmission Network
WTP	Water Treatment Plant

Acronyms

FIFO	First in First Out
ReLU	Rectified Linear Unit(s)

Units and Symbols

K_b	Bulk chlorine decay rate
Km	Kilometer
mg/l	Milligram per liter
m	Meter
mm	Millimeter
°C	Degrees Celsius
%	Percent(age)
µS	Micro siemens

CHAPTER ONE

INTRODUCTION

1.1 Background

Chlorine is the most widely used disinfectant in drinking water because of its low cost, extended disinfectant durability up to point of consumption, ease of use and efficacy in killing infectious pathogens (Hyunjun and Sanghyun, 2017; Mahendrarajah, 2014). Under-dosage of chlorine exposes consumers to microbial regrowth and consumer infection (Mahendrarajah (2014). Over-dosage corrodes pipeline infrastructure and also discourages consumption because of taste and pungent odour (Tiruneh et al., 2019a). Over-dosage produces carcinogenic Disinfection By-Products (DBPs) that potentially cause cancer, reproductive disorders, liver and kidney damage, birth defects, miscarriage and other human health complications (Vuta and Dumitran, 2019). This, therefore means that control of chlorine dosage within safe regulatory limits is very important for public health.

The World Health Organization (WHO) periodically revises and recommends drinking water standards. The WHO (2014) recommended residual chlorine in drinking water as 0.2 – 5 mg/l. Consequently, there is need to balance these two diametrically opposing but important requirements in drinking water distribution practice. Water utilities use various models to balance these requirements. In Australia, performance of general regression neural networks was compared with multi-linear regression models only without consideration of other chlorine decay models (Bowden et al., 2019). However, there has not been any detailed comparison of chlorine decay models to guide water utilities on which one(s) to use (Gibbs et al., 2019). This study was undertaken to fill in this knowledge and research gap. The idea is to identify model(s) that perform appropriately in predicting amounts of initial chlorine to achieve the recommended range 0.2 – 5 mg/l of free chlorine in drinking water.

1.2 Structure of research dissertation report

This research report contains five chapters chapters including chapter one (introduction), chapter two (literature review), chapter three (research methodology), chapter four (results and discussion) and chapter five (summary, conclusion and recommendations).

Chapter one Section 1.1 introduces the background for the research problem that is defined in Section 1.3. Section 1.4 and Section 1.5 contain the main objective and specific objectives of the research respectively. Chapter one ends with Section 1.8 on conceptual framework of the study.

Chapter Two reviews literature on process modelling and statistical modeling about residual chlorine decay in water distribution systems. Chapter Three explains the methodology in terms of research strategy, sampling design, data collection, research instruments, data analysis and ethical considerations undertaken in conducting the research. Chapter Four presents results specific to each research objective in chapter one's Section 1.5. These results are also discussed accordingly. Chapter five summarizes research methodology, findings before concluding with recommendations and necessary future actions.

1.3 Problem statement

Water managers tend to be confronted with two opposing but equally important requirements regarding the need to attain the residual chlorine concentration range of 0.2 - 5 mg/l in drinking water as recommended by the WHO (2014). The first requirement is avoidance of low chlorine concentrations that exposes water consumers to pathogenic infections (Mahendrarajah, 2014). The second requirement is avoidance of high chlorine concentrations that forms carcinogenic disinfection by-products (Vuta and Dumitran, 2019). High chlorine concentrations also corrode water supply infrastructure (Tiruneh et al., 2019a).

Physical (or process-based) and statistical models exist for predicting residual chlorine decay in water distribution. However, there seems to be an insufficient understanding on the underlying knowledge that governs process-based models (Vuta and Dumitran, 2019). Equally, some statistical models are too simple to capture variation in observations. On the other hand, some statistical models are too complex and overfit observations in modelling realities. Besides, existing literature shows that performance of process models and statistical models in predicting final residual chlorine in water distribution have not been compared systematically (Gibbs et al., 2019). Therefore, there is need to evaluate several models and determine which achieves the most acceptable results. Manual adjusting of residual chlorine levels causes upstream over-dosage and downstream under-dosage. Upstream over-dosage increases water treatment costs, consumer water bills, infrastructure corrosion, unpleasant taste and odour that discourages water acceptance. Downstream under-dosage exposes public health to multiple infections and water-borne diseases.

Therefore, it is important to identify an appropriate chlorine decay model that achieves acceptable range of chlorine concentrations. Appropriate chlorine decay model can promote safe drinking water in line with the United Nations Sustainable Development Target 6 (SDG6) on the need for universal access to safe drinking water to enhance human health and productivity.

1.4 Main objective of research

The main aim of the research was to compare space-time performance of chlorine decay models in water distribution system.

1.5 Specific objectives of research

The specific objectives of the research were:

(i) to characterize residual chlorine decay parameters in water distribution,

(ii) to assess space-time decay of chlorine in water distribution systems,

(iii) to compare performance of various models in water distribution systems, and

(iv) to identify the appropriate model(s) for predicting residual chlorine decay in water distribution systems.

1.6 Research questions

The research questions were:

(i) Which water quality and water system parameters influence space-time decay of chlorine?

(ii) How does chlorine decay in space and time?

(iii) How does the various models that predict residual chlorine concentration in distribution systems compare in performance?

(iv) Which model(s) is/are appropriate for managing residual chlorine decay?

1.7 Scope of research

The time, geographical and content scopes of the study were as follows:

1.7.1 Time frame

The study was done during the dry spell of the year which was February to early March 2021. This was deliberately intended to avoid disruptions like landslides that are endemic in this area during rain seasons. Therefore, data from rain season was out of scope.

1.7.2 Geographical coverage

The study was done on the 90 Km length of Lirima Gravity Flow Scheme in Mount Elgon region in eastern Uganda as shown in Figure 1.1. The Geographical Position System (GPS) coordinates of the water treatment plant of Lirima gravity flow scheme was 36N 0657122 Northing, 0098196 Easting, 1,812 meters above mean sea level. This gravity scheme is owned and operated by National Water and Sewerage Corporation. Data was collected from three out of the four water zones representing 75%. The three zones were Musiye zone, Manyeke zone and Vermiculite zone. Data was not collected from Butiru zone. This coverage was considered good enough for generalization to the scheme itself and other gravity water schemes.

1.7.3 Content

The study was done mainly on plastic materials in gravity water distribution system. For simplicity, single-reactant one-phase constant-rate-coefficient first-order model was assumed in process modelling of chlorine decay in water distribution systems. This assumption was motivated by the common use of single-reactant one-phase constant decay-rate modelling in water quality studies because of (1) lower order of reaction of less than one (i.e. $n < 1$) predict better than higher order models (Hyunjun and Sanghyun, 2017), (2) marginal performance benefit of higher orders over simple first-order models (Goyal and Patel, 2014), (3) low 0%-15% with average 12% error in reaction rates between first-order and second-order models (Hyunjun and Sanghyun, 2017; Tiruneh et al., 2019b) and (4) first order reactions performed well with regression R-squared of 0.89 – 0.95 (Tiruneh et al., 2019b). The fifth justification for this assumption was that previous studies showed that the very low ratio of fast to slow reactants in the order (10:10,000) (1%) in two-phase or multi-phase reactant models was insignificant (Jamwal and Kumar, 2016; Tiruneh et al., 2019b). The advantage of parallel model was also insignificant for first-order decay model.

1.8 Study area

The study was carried out on Lirima gravity water scheme located on the slopes of Mount Elgon in eastern Uganda. Lirima gravity water scheme is owned and operated by National Water and Sewerage Corporation. The GPS of Lirima's water treatment plant is 36 N, 0657123 Northing, 0098196 Easting and 1,812 meters above mean sea level. Detail maps on transmission and distribution networks of Lirima gravity flow scheme are in Figure 3.1 and Figure 3.2 respectively in the methodology chapter.

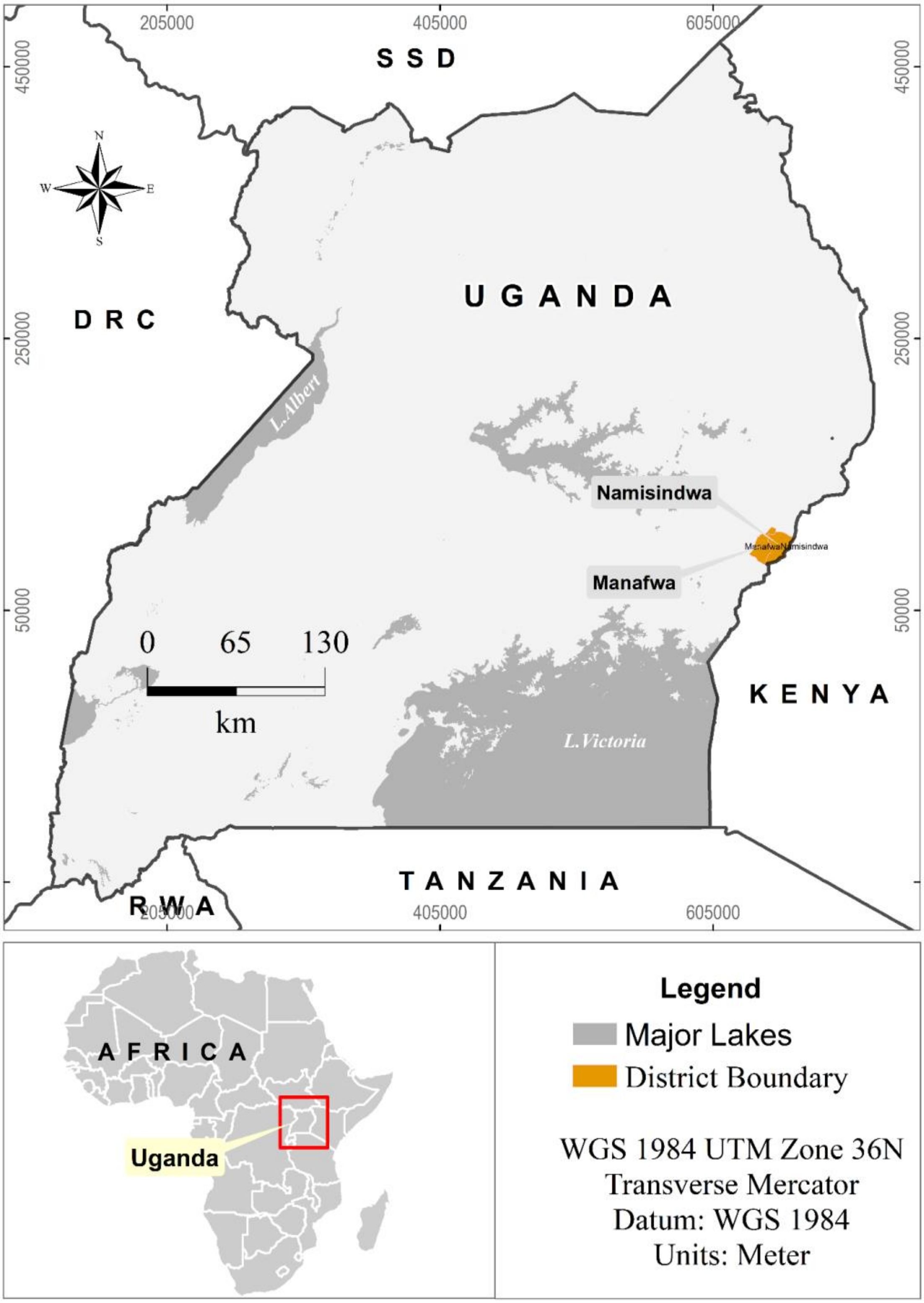


Figure 1. 1 : Research project area location map

1.9 Conceptual framework

The physical and statistical models as described in Section 1.1 were used to model residual chlorine at any point in the water distribution system. The independent variables comprise water quality parameters and pipe system network parameters. The dependent variable is residual chlorine downstream from upstream chlorination point. The conceptual model for the study is as shown in Figure 1.2.

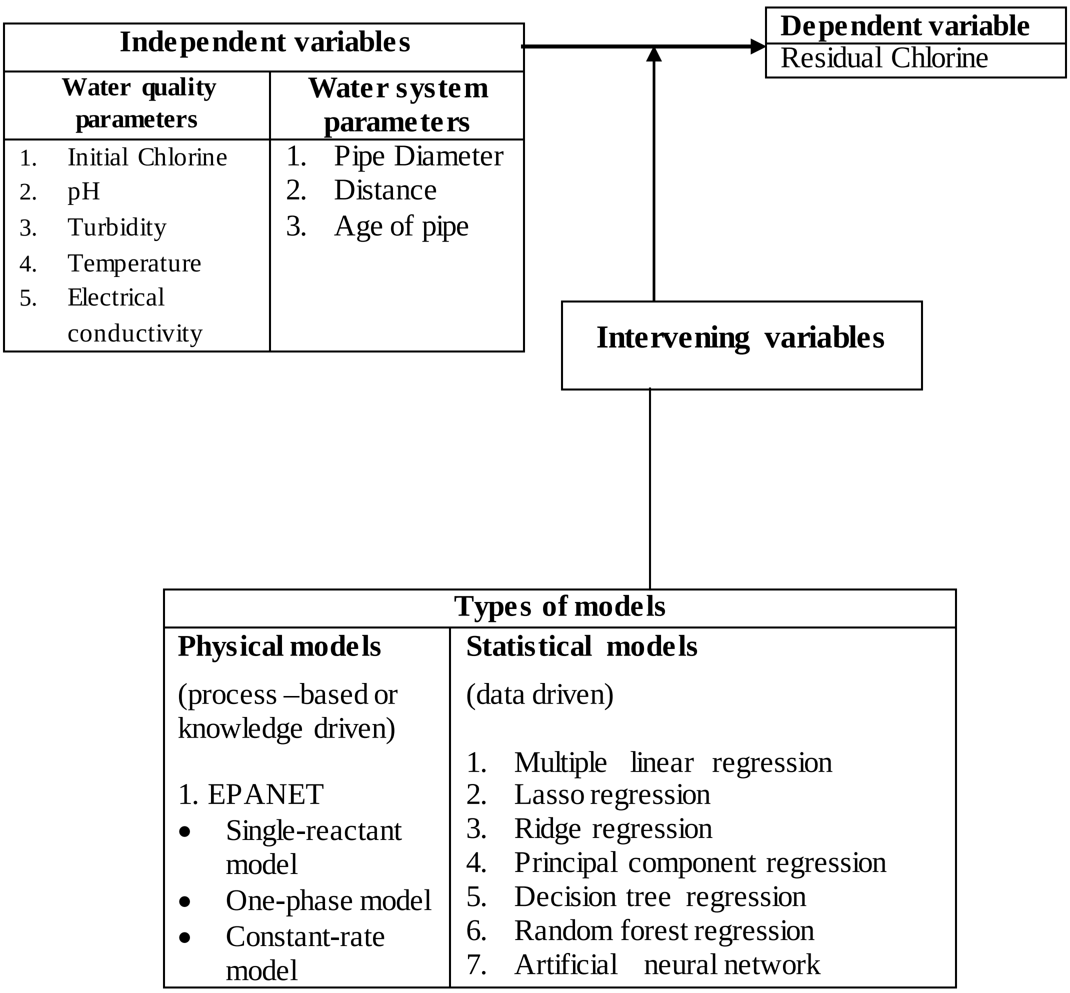


Figure 1. 2: Conceptual framework

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter reviews past works on water quality modelling that are broadly divided into process (knowledge-driven) and statistical (data-driven) models (Gibbs et al., 2019). Process models are developed based on the science of chlorine decay (Bowden et al., 2019; Vuta and Dumitran, 2019). In contrast, statistical models are developed from observed data of water quality parameters that influence residual chlorine decay in water distribution (Chirwa and Madzivhandil, 2017; Mahendrarajah, 2014; Bowden et al., 2019). Within each of these two broad water quality models, there are various sub-models.

2.2 Characterization of water distribution parameters of chlorine decay

Distance from the tank in the distribution system deteriorates water quality due to water age while high residence time increases formation of disinfection by-products (DBPs) (Shamsaei, Jaafar and Basri, 2013). Pipe length and residence time were also identified as influential factors of residual chlorine decay in water distribution (Chirwa and Madzivhandil, 2017). These observations suggest that distance (pipe length) is of crucial importance. Minimizing water pipe lengths in water distribution network design would limit loss of chlorine at distant water consumption points.

2.2.1 Initial chlorine concentration of chlorine decay

Initial chlorine concentration (ICC) affects the bulk decay rate of chlorine (K_b) such that low K_b guarantees high water quality (Karadirek et al., 2015). In the same line, it is known that K_b depends on water quality in water distribution network (Nono et al., 2019). This implies that ICC is an important parameter to be considered in predicting residual chlorine decay modeling in water distribution networks. Initial chlorine concentrations of 1.00-1.5 mg/l (Blokker et al., 2014) and 1.8 mg/l (Wu and Dorea, 2020) can be sufficient. In Africa, ICC of 1.14 mg/l (South Africa) and 0.6 – 1.0 mg/l (Botswana) were reported by Madhivindhila and Chirwa (2017) and Nono et al. (2019) respectively. The low ICC at water treatment plants is a good strategy for avoiding formation of carcinogenic disinfection by-products (DBPs). However, low ICCs may result in residual chlorine below the lower limit of 0.2 mg/l as recommended by WHO (2014) if consumer water draw-off points are far away from water chlorination. A typical case of very low ICC of 0.85 mg/l was reported by Monteiro et al. (2017). Therefore, there is need to balance low ICC without compromising on residual chlorine at far ends of water distribution networks. A sustainable solution for this is use of intermediate water tanks / reservoirs as secondary chlorination points.

2.2.2 Turbidity and pH of treated water

Turbidity below 5 NTU and pH under 8 are the two key water quality parameters to control free chlorine residual to achieve disinfection efficacy in drinking water (Branz et al., 2017). However, turbidity is believed to be an improper proxy for determining chlorine dosage (Wu and Dorea, 2020). This, therefore calls for a new approach of chlorine dosage that excludes turbidity as an input parameter.

2.2.3 Temperature of treated water

Reactive natural organic matter that influence slow decay phase after 30 minutes are temperature-dependent (Wu and Dorea, 2020). Investigators like Zhang et al. (2016)

and Vargas et al. (2021) included temperature as key parameter in influencing water quality in water distribution systems. These are important observations although these researchers did not show how temperature influences chlorine decay. However, the influence of temperature on water age was related to empirical linear equation represented as $T = 0.\ 0839.Age + 16.3$ with R^2 of 0.58 (Monteiro et al., 2017). This finding was significant in view of the moderate correlation between temperature and water age. The direct role temperature plays in chlorine reaction in terms of c-t (concentration-time) value in water treatment was emphasized by Garcia-Avila et al. (2020), Government of Sudan (2017) and WHO (2014). Hassan et al. (2019) showed strong relationship between temperature and chlorine bulk reaction rate for different reaction orders as follows: $R^2 = 0.83$ for $n = 0$, $R^2 = 0.95$ for $n=2$ and $R^2 = 0.71$ for $n =2$ where n = order of residual chlorine decay kinetics.

2.2.4 Hydraulic transients in water distribution networks

High flow velocities cause turbulence which is associated with low chlorine concentrations (Jamwal and Kumar, 2016; Kim et al., 2014; Vuta and Dumitran, 2019). The case of velocity increase causing free chlorine decay because of increased mixing and transport of chemicals between bulk and wall in water pipes was advanced by Stoinov and Aisopou (2014). These findings are important in underscoring maximum flow velocity limits that regulatory agencies impose in design of water distribution networks. The guidelines for water infrastructure design manual of Directorate of Water Development (Uganda) (2013) is maximum velocity of 3.5 m/s. Water age of over 20 hrs have residual chlorine of less than 0.04 mg/l (Gibson et al., 2019) which is below the lower limit set by WHO (2014). Water age that is a function of velocity emphasizes that the longer water stays in pipes, the more are the chances for it to interact with pipe biofilm that consume and deplete chlorine. This implies that the design of water distribution network should ensure that treated water released from water treatment plant should reach consumer points within a day.

Low pressures increase chlorine decay in water distribution network (Alsaydalani, 2019). Similarly, Ataoui and Ermini (2017) found that pressure was the most important parameter (among pressure, velocity and chlorine concentration) in minimizing risk of low chlorine concentration at water demand nodes. This result emphasizes the need to operate water distribution system above regulatory minimum pressure. This means that the pressure at peak-hour must be above 1 bar (10 m) as per guidelines for water infrastructure design manual of Directorate of Water Development (Uganda) (2013).

However, Hyunjun and Sanghyun (2017) claimed that transient events inhibit consumption of chlorine. This disagrees with Stoinov and Aisopou (2014), Jamwal and Kumar (2016), Kim et al. (2014) and Vuta and Dumitran (2019). A 2-D (two-dimensional) model to investigate the effect of transient events (changes in pressure and velocity) on chlorine decay was used by Hyunjun and Sanghyun (2017). This 2-D that is not real-life on 3-D model could have contributed to the fundamental difference in results. To compare model performance, evaluation by model performance metrics is necessary on 3-D model.

2.2.5 Distance and residence time in water distribution networks

Water quality deteriorates with distance due to water age and residence time that increases formation of disinfection by-products (DBPs) (Shamsaei and Basri, 2013; Tiruneh et al., 2019a). Pipe length and residence time were also identified to influence residual chlorine decay in water distribution (Chirwa and Madzivhandil, 2017). Similarly, Kulkami et al. (2018) found that long hydraulic retention times increase nitrification and decay of monochloramine in gravity-fed water distribution systems. These observations are significant as they suggest that distance (pipe length) is of crucial importance. This, therefore means that shortening water pipe lengths in distribution network design limits loss of chlorine at distant water consumption points.

2.2.6 Conclusion on chlorine decay parameters in water distribution network

In conducting on-line water quality monitoring in water distribution systems for quality surveillance, Environmental Protection Agency (2018) advised that threshold analysis for parameter values should not surpass pre-defined thresholds within the normal range of values. The limitation of water quality parameters within pre-defined threshold values is a good practice to guarantee health and safety of water consumers. Low but adequate chlorine concentrations, electrical conductivities and turbidities ensure safe drinking water. This was demonstrated by Bowden et al. (2019) on velocity, temperature and turbidity as the only water parameters. Similarly, Cuesta et al. (2014) used water quality parameters of velocity, temperature and initial chlorine concentration as the only parameters. This trend illustrates the importance of water quality parameters more than water distribution physical parameters.

2.3 Assessment of space-time decay of chlorine in gravity flow systems

The two broad categories of residual chlorine decay models are process-based models and statistical-based models (Gibbs et al., 2019).

2.3.1 Process-based modelling of chlorine decay

Process models that rely on chemical reaction kinetics are deterministic models (Soyupak et al., 2011). This suggests that process models do not account for probabilistic variations in water quality modelling in water distribution networks. This is a setback for process models because random fluctuations in water quality parameters are unavoidable in practice. Velocity and pressure were important hydraulic parameters that influence chlorine decay in water distribution (Hyunjun and Sanghyun, 2017). Therefore, probabilistic modeling of chlorine residual is important because of hydraulic transients that occur in water distribution. Process-based models that assume exponential decay kinetics were not yet fully understood (Vuta and Dumitran, 2019). Process-based models are difficult because of the imprecise apriori knowledge of chemistry and mathematics that govern chlorine behaviour in process models (Bowden et al., 2019). These findings illustrate that process modelling of residual chlorine in water distribution cannot be fully relied on.

Uniform bulk reaction constant is one of the ways used in modelling chlorine decay in water distribution (Georgescu and Georgescu, 2012). This approach is potentially inaccurate because chlorine-reactive substances vary along water distribution mains. Therefore, assumption or use of constant chlorine decay is inaccurate and should be discouraged. The most common exponential decay model for residual chlorine in water distribution network is first-order decay kinetics (Gitu and Egbe, 2016; Nono et al., 2019; Vuta and Dumitran, 2019). This assertion contradicts Hyunjun and Sanghyun (2017) who advocated for as many as 14 higher order decay kinetics. Residual chlorine decay kinetics occurs mainly within the bulk of water flow (Castro and Neves, 2010; Vuta and Dumitran, 2019). This implies that bulk decay dominates at the expense of wall decay. However, Castro and Neves (2010) demonstrated that wall decay can contribute more than bulk decay in overall chlorine decay in water distribution networks. Therefore, underestimation of wall reaction can introduce large errors especially in metallic pipes. The basic form of first order exponential decay kinetics by Mahendrarajah (2014) and Bowden et al. (2019) is as in Equation 2.1.

$$C_t = C_o . exp\,(-K_b t) \qquad [2.1]$$

where:

C_0 = initial Concentration of chlorine, (mg/l)

C_t = concentration of chlorine at time t, (mg/l)

K_b = bulk reaction co-efficient of chlorine, (hr^{-1} or day^{-1})

t = time (Hr. or Day)

Wall reaction co-efficient K_w that contributes to chlorine decay follow zero order decay (Vuta and Dumitran, 2019). The lack of sufficient knowledge of exponential decay solution for residual chlorine has been reported by Bowden et al (2019), Castro and Neves (2010) and Vuta and Dumitran (2019). In view of the claim of inadequate knowledge on determination of wall reaction co-efficient, Vuta and Dumitran (2019) recommended apriori calibration for wall reaction co-efficient. The relationship between total decay constant, bulk decay and wall decay was given by Jamwal and Kumar (2016) in Equation 2.2.

$$K_{total} = K_b + K_w \qquad [2.2]$$

where:

K_{total} = overall decay constant (Day^{-1}),

K_b = bulk decay constant (Day^{-1}),

K_w = wall decay constant (Day^{-1}),

2.3.1.1 Variants of process-based models of chlorine decay

Several variants of process models for residual chlorine decay in water have been proposed. These are (1) single reactant (SR) model versus (2) double reactant (2R) model, (3) variable reaction rate constant model (VRC) versus (4) constant reaction rate model, (5) slow versus (6) fast models and (7) *n*-order chlorine reaction kinetics.

The double-reactant (2R) model can be decomposed into fast and slow reactants (Fischer et al., 2011; Monteiro et al., 2017). This model deviates from the traditional single-reactant (SR) model that has been used in most cases. Fischer et al. (2011) argued that 2R model was simpler and generally more suitable than the SR model that uses single set of invariant parameters. The variable reaction constant (VRC) was preferred in favour of constant reaction models by Fischer et al. (2011) and Tiruneh et al. (2019b). Reported 0% – 15% error between variable reaction rate and constant reaction rate models exist (Tiruneh et al., 2019b). This margin of error is substantial because it cumulatively contributes to high cost of chlorination. The variable rate reaction model is more realistic than the constant reaction rate constant model. This is because water quality and temperature that influence chlorine reaction vary in distribution networks.

Arithmetic and harmonic averages and concentration-weighed aggregate reaction rate in reaction kinetics were used by Tiruneh et al. (2019a). The use of various means is good as it provides alternative approaches in chlorine reaction kinetics. However, there was no conclusive position on which of these is better. No universal chlorine decay model was suitable for all system conditions (Hyunjun and Sanghyun, 2017).

Considering many variants of process-based models for chlorine decay in water distribution, this observation by Hyunjun and Sanghyun (2017) seems to be true.

Existence of several competing process-based chlorine decay models demonstrates lack of unanimity on current knowledge on science of chlorine decay in water distribution network. Therefore, preference of any of these process-based models to another cannot be made with good certainty. This situation tends to promote statistical-based models until when there is unanimity among the scientific community on process-based chlorine decay modeling.

2.3.2. Water quality softwares for modelling piped water flow

Best practice requires development of EDSSs (Environmental Decision Support Systems) for managing the environment (Walling and Vaneeckhaute, 2020). Therefore, water quality models can be considered to be EDSSs. Several water quality models like EPANET, OpenFLows WaterGEMs, WaterCAD, CivilDesigner exist. The requirements for a good water quality model include (1) interoperability, interface and graphical editing, (2) hydraulics operations, (3) model building and data connection, (4) model management, (5) result presentation, (6) optimization using genetic algorithms and (7) energy and capital cost management (US Environmental Protection Agency, 2022).

EPANET is widely used globally, is free and primarily designed to model movement and fate of water constituents in water distribution system (US Environmental Protection Agency, 2022). Other water quality models like WaterGEMs and WaterCAD also posses the required specifications for water quality modelling mentioned above. However, they were not designed primarily to model and monitor movement and fate of water constituents in water distribution system movement and fate of water constituents in water distribution system. In this regard, EPANET has advantage over other water quality models because EPANET has in-built provision for monitoring the movement and fate of chlorine in water distribution system.

2.3.3 Statistical modelling of chlorine decay

Statistical-based models use verifiable independent data of parameters that influence chlorine decay to determine terminal residual chlorine as dependent variable at water consumer points (Bowden et al., 2019; Chirwa and Madzivhandil, 2017; Mahendrarajah, 2014). This suggests that statistical-based models seem to appeal to practical use more than process-based models. Examples of statistical models include linear regression models (Jones, 2014), principal component regression model (Jolliffe and Cadima, 2016), lasso and regression models (Melkumova and Shatskikh, 2017), decision trees (Louppe, 2014) and random forests (Louppe, 2014).

2.3.3.1 Linear regression models

Linear regression models are of the general form as shown in Equation 2.3 (Jones, 2014).

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots + \beta_n x_n \qquad [2.3]$$

where:

Y = response (outcome or dependent) variable,

β_0 = value of Y when all independent variables have zero value,

β_i = marginal change in Y due to unit change in x_i,

x_i = i^{th} independent (predictor) variable,

n = number of independent (predictor) variable.

Regression models are useful for descriptions and predictions (Jones, 2014). Predictor variables can be a combination of continuous and categorical variables (Jones, 2014). However, the outcome of regression models must be quantifiable (Jones, 2014). These conditions are applicable to statistical modeling of residual chlorine where water quality parameters as reviewed in Section 2.2 like initial chlorine concentration, turbidity, electrical conductivity, pipe diameter and pipe distance are all continuous variables. Model outcome of residual chlorine concentration is quantitative and measured in mg/l.

The rule of thumb for a good regression model is that the number of predictors, n, should not exceed 10 to avoid multicollinearity between predictors (Jones, 2014). When sample size is small, it is difficult for a model to detect real relationships(s) even if such relationship(s) exist (Jones, 2014). Prediction accuracy of Ordinary Least Squares (OLS) linear regression increases when the number of predictors, k, is much less than the sample size, n, (Melkumova and Shatskikh, 2017). This is because of low variance when $k << n$ (Melkumova and Shatskikh, 2017). One way to determine enough sample size is to ensure that the number of observations per predictor variable is at least 25 (Jones, 2014). This approach is important in determining the number of sample points in water reticulation network and number of repetitions of sampling on different days and at different times in order to generate adequate sample data for analysis. Statistical power is the probability of detecting underlying relationships in datasets when there is actual relationship and the minimum statistical power should be 0.8 (80%) (Jones, 2014).

Construction validity in regression modeling is a form of internal validity for which relationships derived from data analysis exist (Jones, 2014). Construction validity in regression modeling can be ascertained by diagnostics of linearity (for systematic part of model), homoscedasticity and independence (for random part of model) (Jones, 2014). Internal validity in regression modeling examines whether assumed relationships are causal (Jones, 2014). Internal validity is promoted by avoiding omitted variable bias and endogeneity (Jones, 2014). Omitted variable bias is avoidable by including as many as possible of all relevant predictors determined from domain knowledge of research problem (Jones, 2014). This is complied with by including all the water quality and system variables identified in Section 2.2. Endogeneity is the case of direct influence of response on a predictor (Jones, 2014). Among the water quality and system variables identified in Section 2.2, endogeneity does not arise.

2.3.3.2 Principal component analysis regression models

Principal component analysis (PCA) is a technique for reducing data dimensionality of large datasets in order to enhance data interpretation without losing much statistical information and variability (Jolliffe and Cadima, 2016). PCA works by creating new uncorrelated variables known as principal components (PCs) to maximize variance. (Jolliffe and Cadima, 2016). The PCs are defined from the pre-existing variables and are linear functions (linear combinations) of the original variables (Jolliffe and Cadima, 2016).

PCA is developed by SVD (Singular Value Decomposition) of centred data covariance matrix or centred data correlation matrix (Jolliffe and Cadima, 2016). The basic form of PCA models is as shown in Equation 2.4 (Jolliffe and Cadima, 2016).

$$\mathbf{Sa} - \lambda\mathbf{a} = \mathbf{0} \Longleftrightarrow \mathbf{Sa} = \lambda\mathbf{a} \qquad [2.4]$$

where:

$\mathbf{S}$= covariance matrix,

$\mathbf{a}$ = corresponding eigenvector,

λ = corresponding eigenvalue of $\mathbf{a}$.

To avoid the problem of different scales and units of original variables, they must be standardized before their relationships are investigated in covariance matrix (Jolliffe and Cadima, 2016).

PCA is a statistical method related to linear regression (principal component regression) (Jolliffe and Cadima, 2016). The main use of PCA is descriptive instead of being inferential (Jolliffe and Cadima, 2016). Eigenvalues of eigenvectors of the centred data covariance matrix or centred data correlation matrix represent the variance of interest (Jolliffe and Cadima, 2016). When the number of PCs increase, it becomes difficult to interpret it because of the growing number of non-trivial coefficients (Jolliffe and Cadima, 2016). Rotation in PCA is a trade-off technique between interpretation and variance in PCA (Jolliffe and Cadima, 2016).

2.3.3.3 LASSO regression models

LASSO is an acronym of “least absolute selection and shrinkage operator” (Tibshirani, 2013). Lasso is a regularized linear regression model (Tibshirani, 2013). Regularization is the shrinkage (reduction) of the estimated beta coefficients of predictors in OLS regression model to improve predictive accuracy (Melkumova and

Shatskikh, 2017). Regularization is done by penalizing (i.e. reducing) the beta coefficients of predictors of OLS regression model. (Melkumova and Shatskikh, 2017; Tibshirani, 2013). The LASSO regularized beta coefficient is as shown in Equation 2.5 by Tibshirani (2013).

$$\beta^{lasso} = \text{argmin } ||y\text{-}x\beta||^2 + \lambda \sum_{j=1}{}^{p} ||\beta_j||^2 \qquad \ldots\ldots\ldots\ldots [2.5]$$

where:

β_j= predictor coefficient estimate of the j^{th} predictor,

x = predictor,

λ = penalty (regularization) term which is a non-zero scalar quantity.

p = number of predictors

LASSO that is also known as L1 because it uses power (exponent) one of sum of beta coefficients in regularization (Tibshirani, 2013) shrinks some coefficients to zero and retains some (Tibshirani, 2013). In this regard, it performs the additional role of variable selection in a model (Melkumova and Shatskikh, 2017; Tibshirani, 2013). This suggests that LASSO is good to use in large samples. Cross-validation is the technique used to find the optimal (i.e. tuning) value of a parameter by running analysis on test (i.e. unseen) data to yield optimal performance metrics results (Melkumova and Shatskikh, 2017).

2.3.3.4 Ridge regression models

Unlike LASSO, ridge regression shrinks coefficients to small values but never reduces any to zero (Melkumova and Shatskikh, 2017). This means that both significant and insignificant predictors remain in a model. It also means that ridge regression model is good to use in sparse samples as it allows weak predictors to augment contribution of the few strong predictors to ensure overall better prediction accuracy. The ridge regularized beta coefficient is as shown in Equation 2.6 (Tibshirani, 2013).

$$\beta^{ridge} = \text{argmin } ||y\text{-}x\beta||^2 + \lambda \sum_{j=1}^{p} ||\beta_j||^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2.6]$$

where:

β_j= predictor coefficient estimate of the j^{th} predictor,

x = predictor,

λ = penalty (regularization) term which is a non-zero scalar quantity,

p = number of predictors

Ridge regression that is also known as L2 because it uses power (exponent) two of sum of beta coefficients in regularization (Tibshirani, 2013). The cross-validation technique of determining the optimal penalty term is basically the same as for lasso.

2.3.3.5 Decision tree regression models

A DT (decision tree) is a directed graph in which information for decision-making starts from a root (a parent node) and moves unidirectionally to other parent node(s) until it ends at terminal node(s) (leaf or leaves or child nodes) where final decision is made (Louppe, 2014). At every parent node, a DT splits incoming information into two sub-spaces based on purity (homogeneity of data). Regression output at terminal and child nodes is then computed from the probability of nodal purity and nodal split value (Louppe, 2014). Figure 2.1 shows a schematic representation of a decision tree.

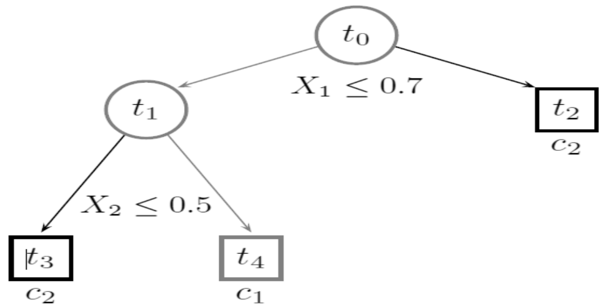


Figure 2. 1: Schematic representation of decision tree

Source: Louppe (2014, p.28)

The shorter a DT is (i.e. the shallower or smaller depth it has), the more interpretive it is even by less technical people (Louppe, 2014). Shallow DT depths suggest avoidance of overfitting a DT model. This property is helpful for quick decision making by less skilled personnel in water quality management.

DTs are non-parametric as they do not need assumptions in modelling complex relations between predictor and response variables (Louppe, 2014). DTs are good for feature selection because they are robust to irrelevant and noisy variables (Louppe, 2014). This property is useful in assisting water managers decide on which chlorine decay parameters to focus on for control and management. DTs are also almost immune to outliers (Louppe, 2014).

2.3.3.6 Random forest regression models

RFs (Random Forests) are collections of randomized decision trees (Louppe, 2014). Collection of randomized decision trees (DTs) reduces prediction generalization error in RF ensemble model by decreasing variance error in the bias-variance decomposition (Louppe, 2014). This means that RFs are better models than DTs. However, DTs are very significant in context of RFs because they (DTs) are the building blocks for RFs. RFs combines the individual predictions of DTs to produce better prediction output than for the individual DT output (Louppe, 2014). The regression prediction of RF is computed by averaging the individual predictions of the aggregated DTs as shown in Equation 2.7 (Louppe, 2014).

$$\psi_{\mathcal{L},\theta_1,\ldots,\theta_M}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} \varphi_{\mathcal{L},\theta_m}(\mathbf{x}) \qquad \text{.......................... [2.7]}$$

where:

m = start number of decision tree

M = number of decision trees,

φ= decision tree parameters

ψ = average of randomized decision tree performance

θ_i = individual performance of the i^{th} decision tree

2.3.3.7 Artificial neural networks

Artificial neural networks (ANNs) are structures composed of dense interconnected nodes (also known as neurons) that perform massive computations on data (Basheer and Hajmeer, 2000; Chau, 2006). ANN are biologically inspired computer programmes designed to mimic functioning of human brain (Agatonovic-Kustrin and Beresford, 2000). The concept of ANNs was thus developed from human brain perceptrons (Basheer and Hajmeer, 2000). The artificial neurons which are data processing elements are connected with weights to constitute the ANN structure organized in layers (Agatonovic-Kustrin and Beresford, 2000). Each neuron has a weighted input, transfer function and output (Agatonovic-Kustrin and Beresford, 2000). Figure 2.2 shows a simplified basic structure of (a) a single perceptron and (b) multi-layer perceptron MLP ANN.

(a) (b)

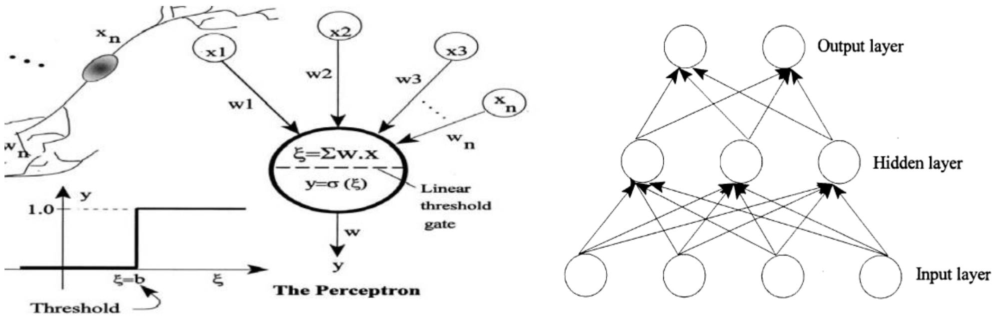


Figure 2. 2: Schematic of (a) single perceptron and (b) MLP (multi-layer perceptron) ANN (artificial neural network)

Source: Basheer and Hajmeer (2000, pp. 5, 7).

The behaviour of an ANN is influenced by the transfer function of its neurons, architecture and learning rule (Agatonovic-Kustrin and Beresford, 2000). A transfer function works on activation signal of neuron to produce output signal that is passed to the next neuron in the next layer (Agatonovic-Kustrin and Beresford, 2000). Weights signify the importance of each input in contributing to activation of a neuron (Agatonovic-Kustrin and Beresford, 2000; Basheer and Hajmeer, 2000). Desirable features of ANN are (1) non-linearity, (2) high-parallelism, (3) fault and failure tolerance, (4) learning ability and (5) generalization capability (Basheer and Hajmeer, 2000). Non-linearity that is introduced by transfer functions (Agatonovic-Kustrin and Beresford, 2000) allows better fit to data (Basheer and Hajmeer, 2000). Noise tolerance allows better predictions in the presence of uncertain data and measurement errors (Basheer and Hajmeer, 2000). High parallelism means fast data processing, learning means adaptation, updating and modification of internal ANN structure to changing environment while generalization allows model application to unlearned data (Basheer and Hajmeer, 2000).

The weighted sum of inputs constitutes the activation of a neuron (Agatonovic-Kustrin and Beresford, 2000). Activation of a neuron happens when bias *b* which is the threshold excitation minimum is exceeded (Basheer and Hajmeer, 2000). After activation (excitation or firing) of/by a neuron, the combined weighted sum of inputs is passed through a transfer function to another neuron in the next layer (Agatonovic-Kustrin and Beresford, 2000; Basheer and Hajmeer, 2000). The mathematical formulation for activation or excitation is as shown in Equation 2.8 (Basheer and Hajmeer, 2000).

$$y = \begin{cases} 1, & \text{if} \sum_{i=1}^{n} w_i x_i \geq b, \\ 0, & \text{if} \sum_{i=1}^{n} w_i x_i < b, \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2.8]$$

where:

y = output,

x_i = i^{th} input

w_i = weight of i^{th} input

b = bias of neuron

n = Number of inputs

Activation (excitation) occurs when y = 1 and activation fails (inhibition) when y = 0 (Basheer and Hajmeer, 2000). ANN is a heavily parametrized system (Agatonovic-Kustrin and Beresford, 2000). This assertion is correct because ANN has many parameters that include input weights, neuron bias, activation functions, number of neurons per layer, number of layers, learning rate, number of iterations, etc. The permutation between neurons per layer and number of layers alone yields large number of neurons in ANN. Since the minimum number of layers in MLP is three, there are potentially many weights. The assertion of Agatonovic-Kustrin and Beresford (2000) that large training dataset is required in ANN is thus correct.

Backpropagation allows the fine-tuning of the weights and bias parameters of ANN (Agatonovic-Kustrin and Beresford, 2000). This is done by using the learning rate where prediction errors in each forward pass are compared with the pre-known target (output) values (Agatonovic-Kustrin and Beresford, 2000; Basheer and Hajmeer, 2000).

2.4 Comparison of performance of models that predict residual chlorine

Performance of EPANET model for residual chlorine at water distribution end-points in Algeria were evaluated by RMSE (root mean square error) and correlation coefficients by Bensoltane et al. (2018). Similarly, Bowden et al. (2019) in Australia used RMSE in deciding that GRNN chlorine models were better than MLR chlorine models. Similar study in Czech Republic showed that RMSE of ANN models were better (lower) than those of EPANET (Cuesta et al., 2014). However, in Ecuador, Nash-Sutcliffe Efficiency (NSE) index and correlation coefficients were used (Garcia-Avila et al., 2021). In Australia, RMSE, MAE (mean absolute error) and Max (maximum absolute error) methods were used to evaluate ANN models at 95% confidence interval (Gibbs et al., 2019). Non-linear neuro-fuzzy algorithms were found to be 1.63 times better in performance than multi-linear regression models in South Korea (Lee et al., 2016). This finding suggests that non-linear algorithms handle non-linear behaviours better.

These studies demonstrate that each model performance metric has some merits and demerits. Therefore, it is a good practice to use multiple metrics to evaluate models.

2.5 Identification of model(s) for chlorine decay in gravity water system

In Section 2.3, various performance metrics were used to evaluate the goodness of fit of chlorine decay models. The importance of these performance metrics is best appreciated when more than one is used for effective comparison and decision making. Therefore, all of these would be used in comparing and deciding which model to choose.

2.6 Conclusion on literature review

Conclusions on physical and statistical modelling were as follows:

2.6.1 Conclusion on physical (process-based or knowledge-driven) modelling

First-order chlorine decay kinetics is the dominant model used in process-models for residual chlorine decay in water distribution networks/systems. First-order decay kinetics is also mainly used in single-reactant models. The reason for first-order single-reactant models is their simplicity and insignificant deviation in performance when compared to higher-order, double-reactant, variable reaction and parallel first-order models according to majority of researchers. There subjectivity in preferring first-order chlorine decay kinetics to higher-order chlorine decay kinetics due to simplicity may be erroneous. This is because many researchers who use first-order chlorine decay kinetics admit current lack of adequate knowledge on the underlying processes in process models for residual chlorine decay. This situation demonstrates the challenge of physical models in residual chlorine decay modelling.

2.6.2 Conclusion on statistical (data-driven) modelling

From literature review, high prediction accuracy due to low variance in linear regression models is achievable when the number of predictors is much less compared to sample size. Quantitatively, the *p/n* ratio should be small (< 1/25) or *n/p* ratio should be large (> 25). This also helps in capturing the underlying relationships, patterns and trends in data. The number of predictors also should not exceed 10 ($p <$ 10) to avoid multicollinearity. This rationale is helpful as one of the ways to determine sample size in a study that has regression modelling. Therefore, this method was used to triangulate sample size in chapter three.

Principal component analysis (PCA), lasso regression, decision tree and random forest models do variable (feature) selection in regression modelling. Standardized beta coefficients and associated p-values also can suggest importance of predictors in regression modelling. Ridge regression is limited to modelling small sample data. It does this by retaining all variables such that weak predictors augment the few strong predictors so that the overall prediction is better than if any of the weak predictors were eliminated. The list of statistical models reviewed in this work is not exhaustive. What has been considered here are deemed adequate for modelling residual chlorine decay in water distribution network. The statistical models considered include ordinary least multiple linear regression models, regularized (i.e. LASSO and ridge) regression models, artificial neural networks and decision tree and random forests. These models have been used variously and singularly for residual chlorine decay modelling in water distribution system. However, their respective performances have not been compared at the same time on the same water distribution system for decision making.

The challenge in statistical models is on the choice of performance metric whose effectiveness depends on (1) type of data, (2) data size, (3) data distribution and (4) algorithms that these statistical models use. Another challenge with statistical models is their very nature of being data-driven models. This means that there should be large datasets to ensure that there is variability in the different variables for underlying relationships and patterns to be revealed. The challenge is big in tree-based models but bigger in artificial neural network modeling where large training data is required. However, the disagreement among the scientific community on performance of statistical models is not as high as that on process (physical) modeling. The next chapter of methodology covers the collection of these chlorine decay parameters and their analyses.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This research used observational and comparison case research designs to predict space-time decay of residual chlorine in gravity water flow scheme. Quantitative observed field data was used to predict residual chlorine concentrations at various points and times in water distribution network.

3.2 Location of research area

This research was conducted on Lirima Gravity Flow Scheme located in Manafwa and Namisinde districts in the Mount Elgon region in Eastern Uganda. This gravity scheme is owned and operated by National Water and Sewerage Corporation (NWSC) which is a government parastatal. Figure 3.1 shows the water transmission main from treatment plant.

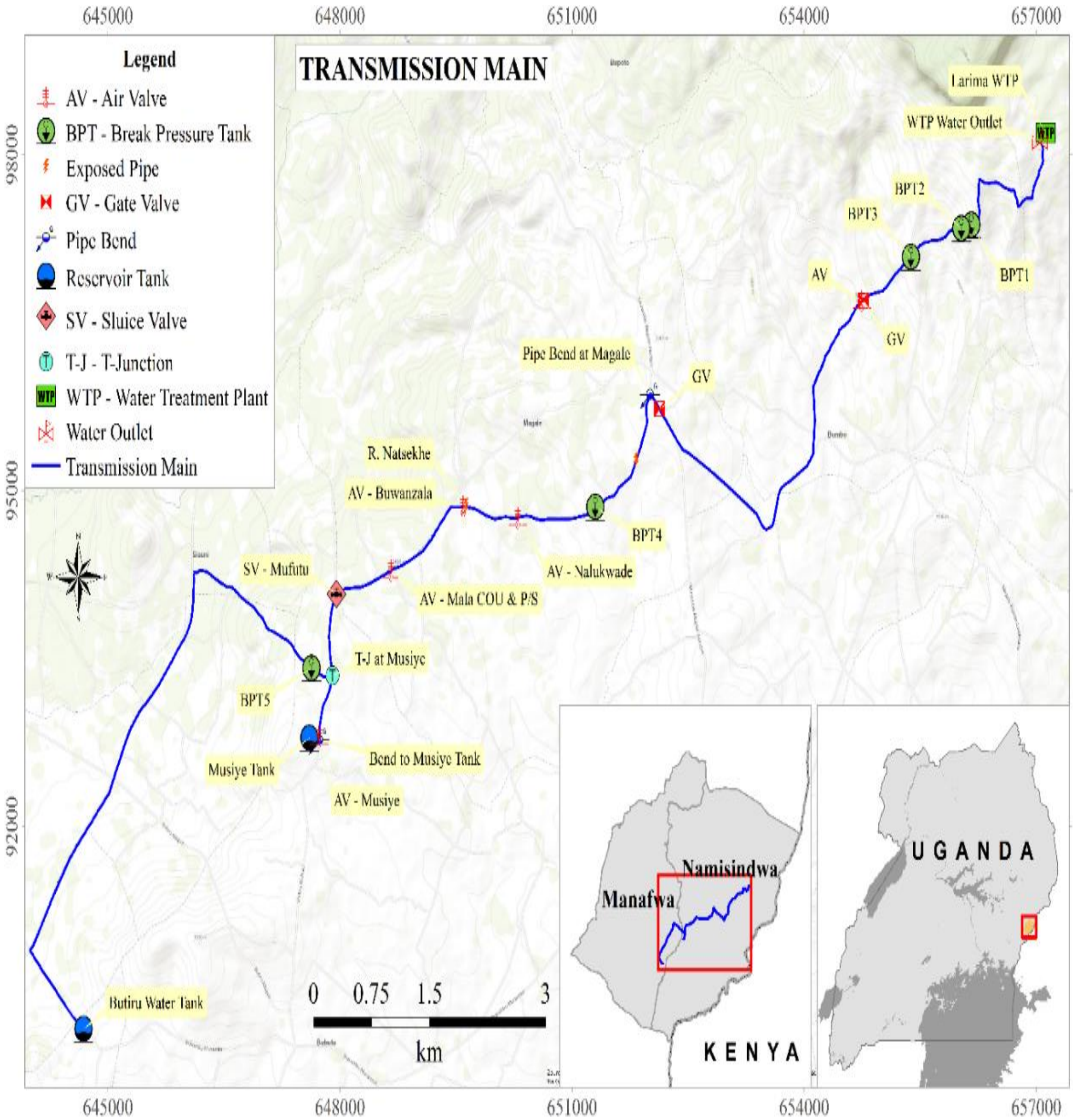


Figure 3. 1: Lirima gravity transmission water main from water treatment plant

3.3 Types of data

The parameters identified in chapter 2 (literature review) that influence residual chlorine decay in water distribution networks were broadly divided into water quality and system (physical) parameters. Water quality data collected included residual chlorine concentrations, pH, turbidity and electrical conductivity. Water distribution system parameters collected were distance (length) of sample points from head/start of water distribution/transmission lines, pipe size (diameter), flow velocity and pressure.

Data was collected from 8th to 20th February 2021. The dry weather during this time enabled access to water distribution network in this difficult and rugged terrain.

3.4 Data collection

Data collection was managed as follows:

3.4.1 Data collection strategy

At least two runs (morning and afternoon) were conducted each day on a particular distribution main. On each run, data was collected at sampling points at approximate intervals of 1 Km. This spacing interval was based on the advice of NWSC that closer intervals may not reveal significant variations in residual chlorine concentrations. Data collection was replicated on different days to account for variations in study data. Replication of data on different days was also a strategy to increase on sample size of study data. A total of 128 datasets were collected. Appendix Tables A.2, A.3, A.4 and A.5 contain the GPS co-ordinates of water transmission and distribution mains, sampling yard taps and water tanks (break pressure tanks and intermediate reservoirs) respectively.

3.4.2 Data collection procedure

Water was sampled at clear water tank outlets, wash outs and nearest functional yard taps that were on direct supply lines from water distribution and transmission mains. The yard water taps from which water was sampled were those that were very close to distribution mains within off-sets of less than 5 m as shown in Figure 3.2. It was assumed that water quality parameters at yard taps close to distribution networks would not have varied significantly from the water in the nearby distribution lines hence would be practically representative of water quality parameter values.

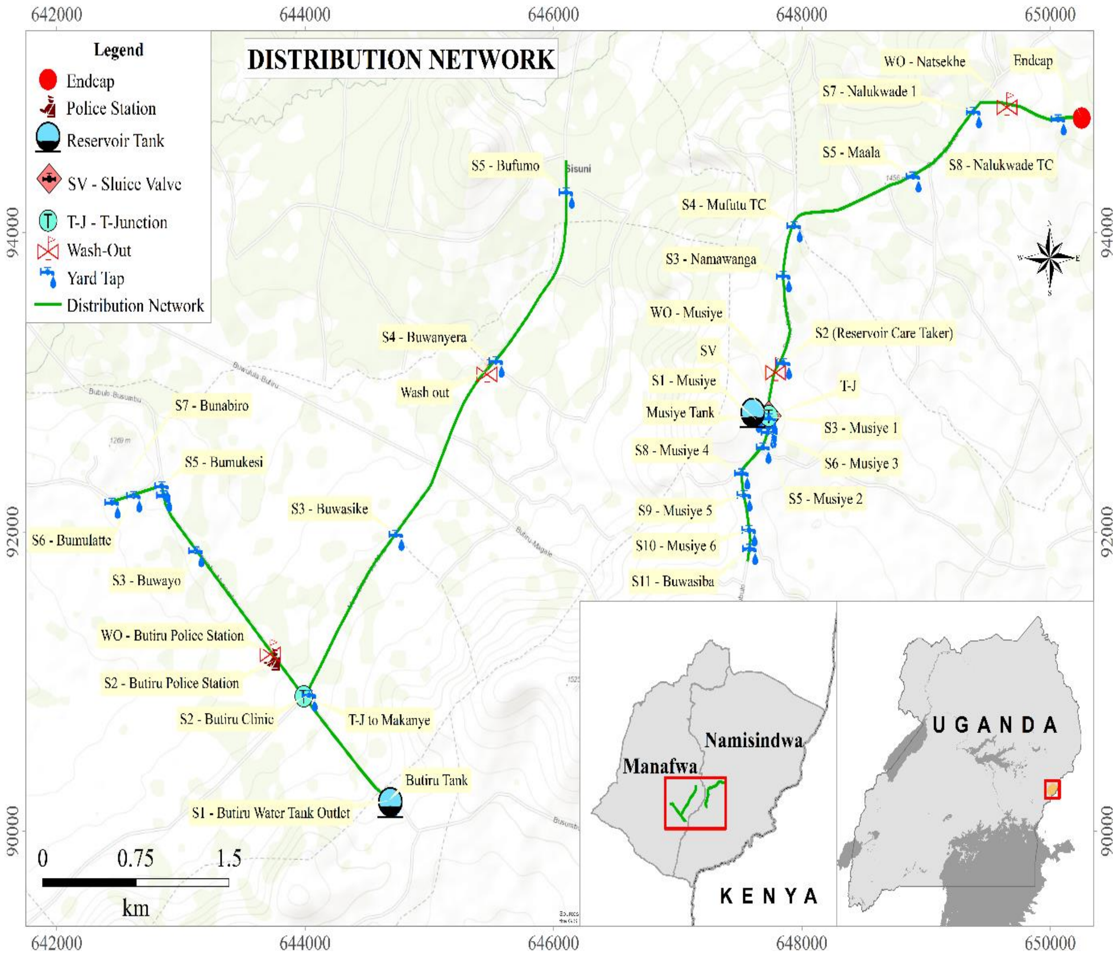


Figure 3. 2: Yard taps from adjacent distribution water mains of Lirima gravity scheme

3.4.3 Data collection personnel

The research team comprised four people as follows: (1) research student, (2) NWSC water quality analyst and (3) two NWSC plumbing technicians. The research student coordinated field data collection and ensured that correct and standard data collection procedures were observed in order to obtain quality data. The researcher also picked GPS coordinates of pipeline mains and water taps. The water quality analyst sampled and analysed water quality parameters at each sample point. The plumbing technicians traced and identified pipeline and yard tap locations for picking GPS coordinates and water sampling. The plumbing technicians also would fix pressure gauge on water taps (faucets) for pressure testing and remove it for reuse at subsequent sample points.

3.5 Data collection instruments

Field data was collected using the following portable instruments as summarized in Table 3.1 whose photographs are shown in Plate 3.1.

Table 3. 1: Instruments used for data collection

Item	Instrument	Purpose
1	Lovibond MD 600 digital meter	1. Tests residual chlorine in the range of 0-6 mg/l 2. Tests turbidity (FAU)
2	pH and Conductivity 901 digital meter	1. Tests temperature (°C) 2. Tests pH 3. Tests electrical conductivity (µS)
3	Pressure gauge EN837-1	1. Tests pressure (Bars)
4	Garmin GPSMAP 64s	1. Picking GPS coordinates of water mains and yard taps
5	Standard 1 Liter plastic bottles	1. Water sampling
6	Standard ice box	1. Preservation of sampled water
7	Calibrated 5 Liter plastic jerrycan	1. Water flow rate measurement
8	Stop-watch	1. Timing filling of calibrated 5 Liter plastic jerrycan
9	Area topographic map	1. Tracing and identification of water mains

The details about the photographs are as summarized in Table 3.2.

Table 3. 2: Details of water sampling and testing photographs

Item	Plate	Detail
1	a	Signboard at Lirima Gravity Water Treatment Plant
2	b	Typical cylindrical break-pressure tanks and water reservoirs
3	c	Atop Butiru water tank overlooking Mbale-Lwakhakha Road
4	d	Water tank roof access to tank interior
5	e	Testing on-line water sample at Musiye wash-out
6	f	Lovibond MD 600 digital meter test kit
		Picking GPS coordinates of water distribution line
7	g	At Butiru Clinic on Butiru-Manyeke Water Line
8	h	On exposed transmission line to BPT 3
9	i	At Gate valve at Magale Town Council
10	j	At T-junction to Bumbo Town
		Pressure-testing and water flow rate (velocity) testing
11	k	Pressure-testing at Bunyangabo cell
12	l	Pressure-testing at Bunyangabo cell
		Water sampling at Natshekhe wash-out on Musiye-Nalukwade line
13	m	Accessing and cleaning of Natshekhe wash-out
14	n	Draining-off accumulated dirt and dirty water
15	o	Clean water
16	p	Sampling clean water
17	q	Clean water

Plate 3. 1: (a) – (q): Lirima gravity water sampling and testing

3.6 Data analysis

EPANET 2.0, ArcGIS, Python, MATLAB and IBM SPSS Version 25 softwares were used to analyze data. Analysis of same aspect using different softwares was meant to triangulate analysis results. Triangulation helps in decision making on whether there was consistency on analysis results obtained from different analytical tools. Correlation analysis was used to investigate relationship of water and system (physical) parameters mentioned in Section 3.3 with residual chlorine decay. Mass continuity equation was used to determine the flow rate and velocity in distribution mains by using time to fill a calibrated five (5) liter jerrycan from yard taps of known cross-sectional areas. Flow rate (flow rate) was used as a proxy variable for water age because actual water age required tracers to use. Laboratory and field test results for chlorine decay constants were compared in deciding which one to use in chlorine decay reaction.

3.6.1 Data analysis methods for specific objective number one

The influence of water quality and water system parameters was investigated using triangulated methods of (1) decision tree analysis importance score, (2) random forest ensemble importance score, (3) principal component analysis equamax rotated matrix loadings and communalities and (4) p-values and standardized beta coefficients of independent variables in backward elimination in ordinary least squares regression models.

3.6.2 Data analysis methods for specific objective number two

Mathematical models and an algorithm for analyzing space-time decay of residual chlorine in water would be as follows:

3.6.2.1 Mathematical models for space-time decay of residual chlorine in water

Space-time decay of residual chlorine in gravity water flow system was determined using six models of (1) linear, (2) logarithmic, (3) inverse, (4) quadratic, (5) cubic and (6) Weibull models for the statistical modelling of space-time decay. Physical (process) modelling of space-time decay of residual chlorine that is based on chlorine decay coefficients (constants) was investigated using (1) laboratory determined chlorine decay coefficients (constants) and (2) single-reactant chlorine decay coefficients (constants) based on the in-situ residual chlorine decay in water pipelines.

3.6.2.2 Algorithm for space-time chlorine decay in gravity water distribution system

This section presents an algorithm for secondary chlorination in gravity water distribution network.

The eight (8) steps of the algorithm are as follows:

1. Given a gravity water distribution pipeline that begins from start-point A (upstream) and distributes water to end-point B (downstream).
2. The pipeline of total length L comprises pipe segments L_i such that $\sum L_i = L$.
3. At start-point A, water is supplied from a reservoir tank that with water at depth h.
4. From hydraulics and assuming negligible exit head loss, the velocity v in the first pipe segment L_1 is:

   $v = \sqrt{2gh}$. ……………………………………………………………………..…[3.1]
5. From mass continuity equation, velocity in two connected pipe segments Li and L_{i+1} with functional arguments (Length: L, velocity: V, Area: A, diameter, d) is related as follows:

   $A_1 V_1 = A_2 V_2$, $A_1 = \pi d_1^2/4$, $A_2 = \pi d_2^2/4$ and $V_2 = V_1 d_1^2/ d_2^2$
6. The time of travel t_i in a pipe segment L_i from start-node (upstream) to end-node (downstream) with water travel velocity V_i is $t_i = (L_i / V_i)$.

7. At intermediate (i.e. shared) nodes, the residual chlorine concentration $[Cl]_{shared\text{-}node}$ is such that:

$$[Cl]_{\text{end-node of pipe } Li} = [Cl]_{\text{start-node of pipe } Li+1}$$

8. Using first-order (*n=1*) exponential chlorine decay equation of $C_t = C_0 \exp(L_i K_i/V_i)$ can be applied to any pipe segment L_i from where K_i is the single-value (i.e. $K_{total} = K_b + K_w$) chlorine decay constant. The single-value K_i of $n=1$ is that calibrated by demand flow and cross-validated by chlorine residuals at nodes in EPANET hydraulic analysis.

These eight (8) steps of the algorithm that repeat for each pipe segment can be performed in a "for loop" in a computer programming language. The attributes of pipe segments can be defined in global variables as sequence of objects in a list [] or dictionary {}.

To operationalize this secondary chlorination algorithm in a computer programme, a function for determining residual chlorine at downstream nodes can be defined. Arguments / parameters of (Length: *L*, velocity: *V*, Area: *A*, diameter: *d*, Single-value decay constant: *K*) would be passed into the function. A "for loop" with "if-else" control structure of (1) if initial chlorine dosage was less than 3 mg/l, continue else break (i.e. stop) and return would be included in the function block statement. The limit of 3 mg/l (Nouri et al. 2017) minimizes formation of carcinogenic disinfection by-products. (2) A nested loop within the first loop if velocity of flow was less than 3.5 m/s (Directorate of Water and Development (Uganda) 2013), continue else break (i.e. stop) and return.

3.6.3 Data analysis methods for specific objective number three

Comparison of performance of physical and statistical models in predicting chlorine decay in water systems was evaluated using (1) goodness of fit metrics of (a)

Pearson's coefficient r, (b) R-squared, (c) adjusted R-squared and (d) standard error of model estimate. Performance accuracy metrics used were (1) RMSE and (2) maximum absolute error. All these evaluations were estimated at minimum of 95% confidence interval and maximum of 5% level of significance.

3.6.4 Data analysis methods for specific objective number four

The choice of appropriate model for use in optimizing residual chlorine concentrations in gravity water system was based on considerations / factors of (1) generalizability of model determined from the performance and comparison of scores of train and test datasets, (2) dimensionality control in terms of the number of control predictors in the model for determining final residual chlorine and (3) interpretability of model in terms of the scale of the resulting predictors.

3.7 Ethical considerations

Three deliberate approaches of permission, collaboration with utility area field officers and proxy variable for water age were used to promote ethical considerations in this research.

3.7.1 Permission for research

Kyambogo University sought for permission from NWSC headquarters in Kampala (*as in Appendix Letter A.1)* for study on Lirima gravity flow scheme. Permission for the study was granted by the Research and Development unit of NWSC (*see Appendix Letter A.2,*).

3.7.2 Collaboration with utility area field officers

The NWSC local area office in Tororo attached its key staff like the senior water quality technician and plumbers in identifying locations of water pipes and associated appurtenance like wash-outs. These key staff also participated in carrying out both laboratory and on-line water quality tests. This approach ensured easy access to sampling points and management of emerging concerns of water consumers.

3.7.3 Proxy variable for water age

Water age was identified as one of the confounding variables as identified in Section 1.8 and Figure 1.2 on conceptual framework. Calculated water velocity was used as a proxy variable for water age. Tracers which produce more accurate results were not used due to cost and public health implications.

3.8 Summary of methodology

Figure 3.3 summarizes the linkage between methodology and the preceding chapters.

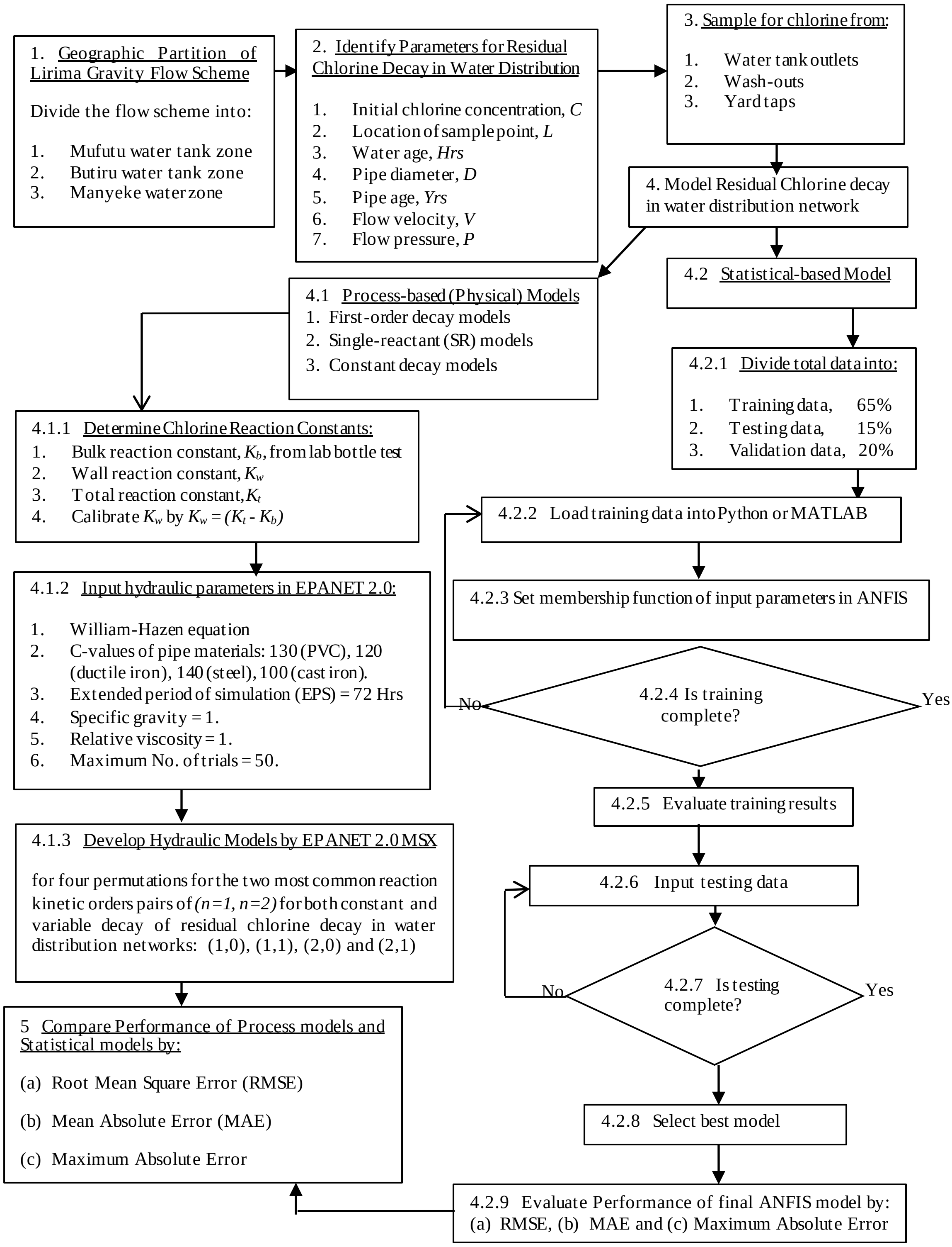


Figure 3. 3: Methodology for comparing performance of chlorine decay models

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the findings of this research and is structured as follows: (1) Univariate analysis of chlorine decay parameters, (2) bivariate analysis of relationships between chlorine decay parameters with chlorine, (3) process model calibration of residual chlorine decay in Water Distribution Networks (WDN) by EPANET, (4) performance analysis of EPANET-based process model in predicting residual chlorine concentration in WDN, (5) data-driven model calibration of chlorine concentration in WDN by multi-variate regression analysis and multi-perceptron sequential deep neural network analysis. (6) Comparison of process model and data-driven model(s) in predicting residual chlorine concentration in WDN and (7) scenario analysis performed within WHO (2011) regulated chlorine concentrations of 0.2-5.0 mg/l are then presented. This is to enable decision on (i) chlorine dosage at treatment plant and (ii) need of secondary chlorination and if so at what network location(s). The results under each of these sections are discussed after their presentation. The chapter concludes with summary of key results and proposed future action.

4.2 Univariate analysis results for parameters of chlorine decay

The summary statistics from Python analysis for four key water quality parameters (temperature, pH, turbidity and electrical conductivity), two system parameters (length, pipe diameter) and two hydraulic parameters (pressure and velocity) are as presented in Table 4.1.

Table 4. 1: Descriptive statistics for residual chlorine decay parameters in water

	count	mean	std	min	25%	50%	75%	max
Water quality parameter	128	64.50	37.09	1.00	32.75	64.50	96.25	128.00
Residual chlorine (mg/l)	128	0.14	0.07	0.00	0.09	0.14	0.19	0.37
Distance (Km)	128	2.50	2.2	0.01	0.67	1.71	4.60	7.50
travel time (min)	128	46.13	42.63	5.00	15.00	30.00	65.00	190.00
Diametre (mm)	128	108.28	51.92	50.00	80.00	100.00	100.00	250.00
Turbidity (NTU)	128	0.96	0.77	0.00	0.75	1.07	1.07	5.00
Electrical Conductivity (µS/cm)	128	70.01	2.53	65.40	68.38	70.01	70.03	78.50
pH	128	7.53	0.17	6.71	7.48	7.53	7.60	7.83
Temperature (°C)	128	23.98	1.06	20.10	23.59	23.98	24.31	27.05
Pressure (Bar)	128	2.00	1.08	0.00	1.73	2.00	2.00	6.0
Velocity (m/s)	128	0.04	0.02	0.001	0.02	0.04	0.05	0.10

Table 4.1 shows that the mean residual chlorine of 0.14 mg/l is below the lower limit of 0.2 – 0.5 mg/l specified by WHO (2014). The pH that ranged from 6.71 – 7.83 were within the acceptable range of 6.5 – 8.5 as specified by US EAS 12 (UNBS, 2014).

4.3 Compliance of parameters of chlorine decay for modeling

Table 4.2 lists the test statistics for linearity, homoscedasticity (constant variance) and independence as necessary conditions for normally distributed data to satisfy in modelling.

Table 4. 2: Data quality tests for residual chlorine decay parameters in water

		Compliance test statistics		
		1. Linearity	**3. Homoscedasticity**	**4. Independence**
Item	**Independent variables**	**(Pearson's *r*)**	**(Limits of *Z*)**	**(VIF)**
1	Initial chlorine	0.69		1.35
2	Distance	- 0.11	[-2, 2]	3.38
3	travel time	- 0.08	[-1, 3]	2.75
4	diametre	- 0.09	[-2, 3]	1.95
5	turbidity	- 0.02		1.16
6	Electrical conductivity	- 0.20	[-2, 2]	1.15
7	pH	0.15		1.10
8	temperature	- 0.21	[-3, 2]	1.41
9	pressure	0.03	[-2, 3]	1.64
10	velocity	- 0.17	[-2, 2]	1.20

VIF is Variable Inflationary Factor that measures multicollinearity between independent variables in a regression model, Z = Standardized Z scores

Table 4.2 shows that all independent variables are linearly correlated with final chlorine. The independent variables were normally distributed. The variance of the independent variables were also fairly uniform as they all lie within the restricted maximum residual limits of [-3, +3]. The independent variables are unavoidably autocorrelated because of their large number of 10. However, the autocorrelation between the independent variables are tolerable as they all are below variable inflationary factor of five. These independent variables therefore qualified for modelling residual chlorine decay in drinking water reticulation.

4.3 Characterization of water parameters in relation to chlorine decay

This section presents and discusses the results for specific objective no. 1 which was "to characterize gravity water distribution parameters in relation to chlorine decay". The associated research question for this objective was "Which water quality and water system parameters influence space-time decay of chlorine?" The relationship of each parameter of chlorine decay in water distribution system is summarized by (1) scatter plots in Figure 4.1 and (2) correlation matrix in Table 4.3.

4.3.1 Variation of chlorine decay with water quality and water system parameters

The scatter plots of each parameter of chlorine decay with residual chlorine concentration in water distribution are as in Figure 4.1.

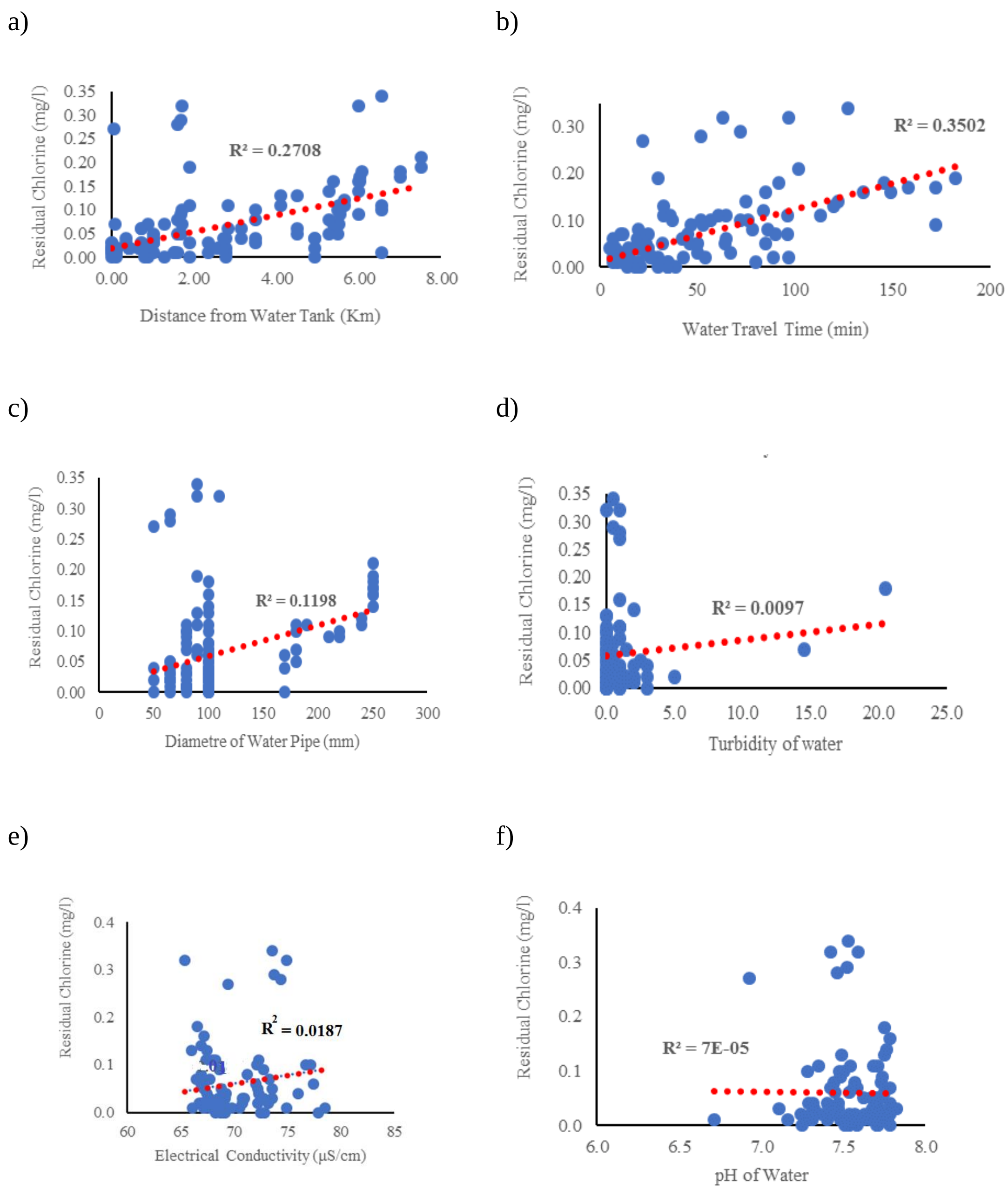


Figure 4. 1: Scatter plots of residual chlorine decay parameters in water distribution

For significance in prediction, an independent variable should have Pearson's correlation coefficient $r > 0.3$. R^2 and R-squared are used interchangeably to mean the same thing. However, in order to increase model performance by avoiding too few predictors, independent variables with R^2 values close to 0.3 can be included such as diametre of water pipe. From Figure 4.1 above, (1) distance from water tank and (2) water travel time qualify to be predictors.

4.3.2 Correlation analysis of parameters of chlorine decay

The correlation matrix of chlorine decay parameters with residual chlorine in water analyzed by Python is as presented in Table 4.3 as follows:

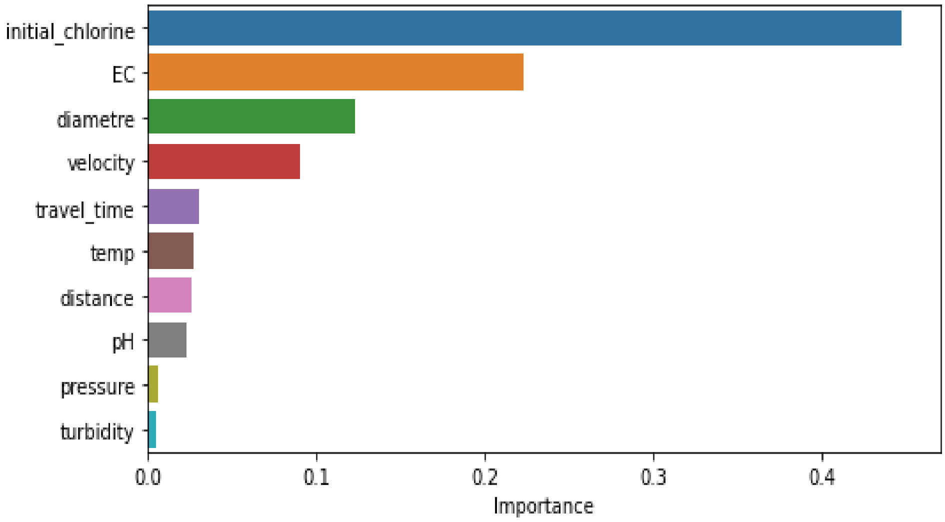
Table 4. 3: Correlation matrix of water quality and water system parameters with chlorine

	RC	**IC**	**dist**	**tt**	**dia**	**tur**	**EC**	**pH**	**temp**	**pres**	**vel**
RC (mg/l)	1.00										
IC (mg/l)	0.69	1.00									
Dist (Km)	- 0.11	0.30	1.00								
tt (min)	- 0.08	0.31	0.71	1.00							
Dia (mm)	- 0.09	0.20	0.63	0.52	1.00						
tur (NTU)	- 0.02	- 0.07	- 0.03	0.05	0.08	1.00					
EC (μSiem^{-1})	- 0.20	- 0.05	- 0.05	-0.06	- 0.10	- 0.29	1.00				
pH	0.15	0.11	0.14	0.17	0.12	0.03	-0.11	1.00			
temp (°C)	- 0.21	- 0.21	- 0.05	- 0.002	0.05	- 0.10	0.21	-0.06	1.00		
pres (Bar)	0.03	0.19	0.31	0.17	0.02	- 0.11	-0.15	0.17	0.23	1	
vel (m/s)	-0.17	- 0.07	- 0.03	-0.004	-0.06	-0.01	-0.15	0.20	-0.001	0.32	1

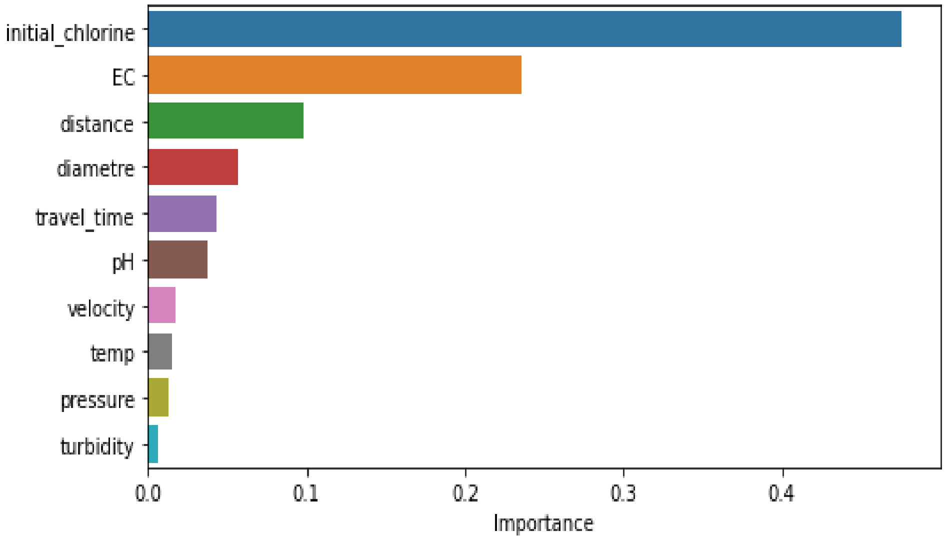
<u>**Legend**</u>:
RC= residual chlorine, IC= initial chlorine, dist = distance, tur = turbidity, EC= electrical conductivity, temp= temperature, pre= pressure, vel = velocity, tt = travel time, dia = diameter

4.3.3 Tree-based analysis of importance of residual chlorine decay parameters

The importance of residual chlorine decay parameters in decision tree and random forest models are as shown in Figure 4.2 (a) and (b).



(a) Feature importance score from decision tree model



(b) Feature importance score from random forest model

Figure 4. 2: Tree-based importance of parameters of residual chlorine decay

Figure 4.2 demonstrates that water quality parameters contribute more to residual chlorine decay than system parameters. Water quality parameters are not as easily controllable as system parameters that are mostly fixed after construction of water infrastructure.

4.3.4 Statistical analysis of importance of residual chlorine decay parameters

The importance of residual chlorine decay parameters was analyzed by ordinary least squares regression and principal component analysis as shown in the following respective sections.

4.3.4.1 Backward elimination of importance of chlorine decay parameters

The importance of residual chlorine decay parameters analyzed by statistical methods in ordinary least squares regression model are as shown in Table 4.4 and Table 4.5. Model 1 in Table 4.4 (a) and (b) from IBM SPSS V25 analysis presents model summary and coefficients of linear regression with all the 10 independent variables.

Table 4. 4: Linear regression model for all independent variables

(a) Model summary

Model	R	R Squared	Adjusted R Squared	Std. Error of the Estimate
1	0.836[a]	0.699	0.672	0.0422

a. Predictors: (Constant), velocity, turbidity, diametre, initial chlorine, pH, EC = electrical conductivity, temp, pressure, travel time, distance

(b) Model coefficients

	Unstandardized coefficients		Standardized coefficients			Collinearity statistics	
Model 1	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	- 0.059	0.229		- 0.257	0.798		
1. initial chlorine	0.563	0.041	0.816	13.740	0.000	0.742	1.347
2. distance	- 0.005	0.0003	- 0.162	-1.718	0.089	0.296	3.381
3. travel time	- 0.001	0.000	- 0.283	-3.331	0.001	0.363	2.752
4. diameter	-3.890E-5	0.000	- 0.028	- 0.386	0.700	0.513	1.948
5. turbidity	- 0.001	0.003	- 0. 001	- 0.191	0.849	0.864	1.158
6. EC	- 0.006	0.002	- 0. 019	-3.469	0.001	0.871	1.148
7. pH	0.071	0.024	0.160	2.980	0.004	0.908	1.101
8. temperature	0.001	0.004	0.010	0.160	0.873	0.711	1.406
9. pressure	-5.041E-5	0.005	- 0.001	- 0.011	0.991	0.611	1.636
10. velocity	- 0.670	0.0242	- 0.155	-2.766	0.007	0.835	1.198

a. Dependent Variable: final chlorine   b. EC = electrical conductivity

From Table 4.4, diameter, turbidity, temperature and pressure were all not statistically significant with very high p-values greater than 0.05. Although travel time and velocity were statistically significant, they were highly correlated with distance as shown by correlation matrix in Table 4.3. Distance correlated more strongly with final residual chlorine than each of travel time and velocity correlated with final residual chlorine. This made distance a better predictor for final residual chlorine than travel time and velocity. This also explains the high multicollinearity of distance with variable inflationary factor of 3.38 and marginal statistical insignificance with p-value of 0.089. Model 2 in Table 4.5 excludes diametre, turbidity, temperature and pressure which were all not statistically significant. However, distance was included in addition to initial chlorine and electrical conductivity which were statistically significant. Table 4.5 (a) and (b) presents model summary and coefficients respectively of linear regression model 2 for three independent and statistically significant variables.

Table 4. 5: Linear regression model for statistically significant variables

(a) Model summary

Model	R	R Squared	Adjusted R Squared	Std. Error of the Estimate
2	0.793[a]	0.628	0.619	0.0453

a. Predictors: (1) Constant, (2) initial chlorine, (3) distance, (4) EC

Durbin-Watson = 1.262

(b) Model coefficients

Model 2	Predictors	Unstandardized coefficients B	Unstandardized coefficients Std. Error	Standardized coefficients Beta	t	Sig.	Collinearity statistics Tolerance	Collinearity statistics VIF
	Constant	0.415	0.112		3.692	0.000		
	1. initial chlorine	0.548	0.040	0.795	13.761	0.000	0.906	1.104
	2. distance	- 0.012	0.002	- 0.365	- 6.250	0.000	0.907	1.105
	3. EC	- 0.005	0.002	- 0.175	- 3.182	0.002	0.995	1.005

a. Dependent Variable: final chlorine b. EC = electrical conductivity

The multicollinearity for all independent variables were low ranging from 1.005 – 1.105 and their p-values were also all below 0.05 for statistical significance at 5% level of significance.

4.3.4.2 Principal component analysis of importance of chlorine decay parameters

Based on Kaiser criterion of eigenvalues of greater than one for important principal components, Figure 4.3 shows that three principal components had eigenvalues above one out of the ten independent variables needed to explain most of the variation in residual chlorine decay in drinking water reticulation system.

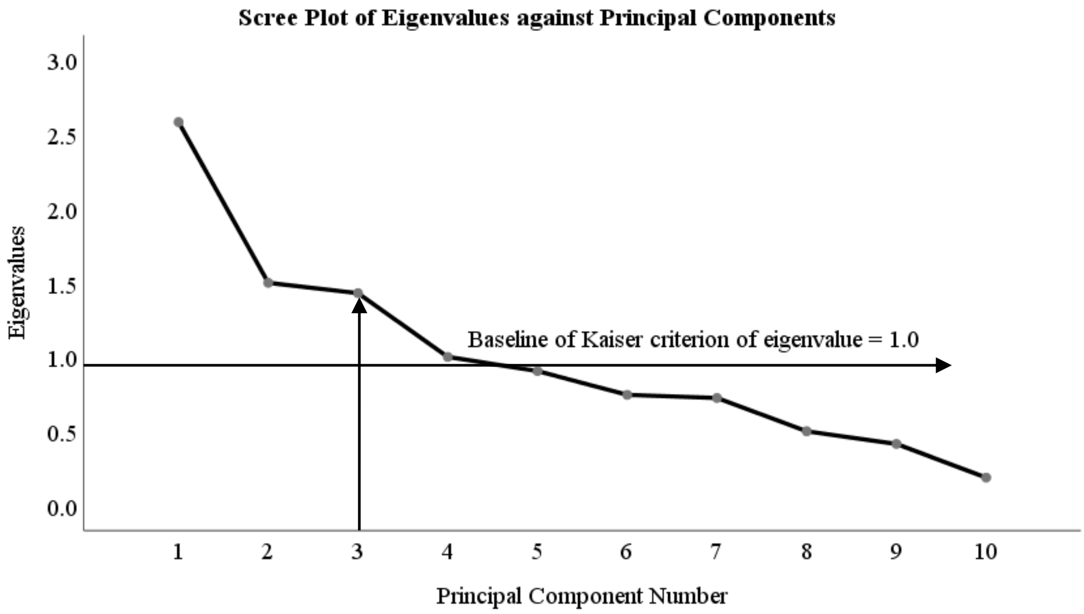


Figure 4. 3: Scree plot of eigenvalues against principal components

The associated principal components are as summarized in Table 4.6.

Table 4. 6: Principal component analysis of residual chlorine decay parameters

Item	No. of Principal Components (PCs)	KMO Measure of sampling adequacy	Bartlett's Test of sphericity	PC1 (%)	PC2 (%)	PC3 (%)	PC4 (%)	Explained total variance (%)
1	2	0.510	0.005	44.06	32.77	NA	NA	76.83
2	3	0.510	0.005	44.06	32.77	23.17	NA	100.00
3	4	0.547	0.000	23.15	25.26	14.45	12.28	65.14

Legend: 1. PC = Principal component, 2. KMO = Kaiser-Meyer-Olkin

Table 4.6 provides three cases for water quality and water physical parameters to explain residual chlorine decay in drinking water distribution systems. Case one is a two principal component solution that explains 77% of residual chlorine decay. Cases two and three comprise three and four principal component solutions that explains 100% and 65% respectively of residual chlorine decay. All these cases satisfy the two requirements of KMO being above 0.5 and Bartlett's test of sphericity being less than 0.05 for statistical significance. However, a four principal component solution explains a much lower 65% total variance compared to the cases of two and three principal components that explain 77% and 100% variance respectively. This could have been because the eigenvalue of the four principal component solution was marginally above one. Therefore, the four principal component solution was not considered. Therefore, there is need to evaluate several models and determine which achieves the most acceptable results. The equations of the principal components for the best case of three principal component solution are as follows:

$$PC1 = 0.999\ Electrical\ conductivity \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [4.1]$$

$$PC2 = 0.988\ Distance \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [4.2]$$

$$PC3 = 0.988\ Initial\ Chlorine \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [4.3]$$

The other related statistics of extraction communalities, structure and pattern matrices for the cases of principal component solutions are found in *Appendix Tables A6 to A8.*

4.3.5 Summary of importance of residual chlorine decay parameters

Table 4.7 summarizes the importance of the different variables that influence residual chlorine decay in drinking water reticulation system. Importance metrics for variables were calculated using different methods of (1) decision tree, (2) random forest, (3) principal component analysis equamax rotated component loading and (4) backward elimination in ordinary least square linear regression.

Table 4. 7: Importance of variables in residual chlorine decay in water system

Item	Independent variable	Decision tree (Score)	Random forest (Score)	PCA (Loading)	OLS Backward elimination p-value / Standardized Beta coefficient
1	Initial chlorine	0.437	0.470	0.825	0.000 / 0.816
2	distance	0.019	0.093	0.817	0.089 / - 0.162
3	travel time	0.030	0.041	0.734	0.001 / - 0.283
4	diametre	0.119	0.052	0.705	0.700 / - 0.028
5	turbidity	0.004	0.007	0.639	0.849 / - 0. 001
6	electrical conductivity	0.217	0.226	0.439	0.001 / - 0. 019
7	pH	0.015	0.030	0.312	0.004 / 0.160
8	temperature	0.022	0.013	0.748	0.873 / 0.010
9	pressure	0.006	0.056	0.693	0.991 / - 0.001
10	velocity	0.089	0.015	0.602	0.007 / - 0.155

Legend: PCA = Principal Component Analysis, OLS = Ordinary Least Square

Table 4.7 demonstrates consistency in ranking of water quality and water system parameters as measured by the different ranking systems of decision tree, random forest, PCA and OLS linear regression.

4.3.6 Bivariate analysis of parameters of chlorine decay

The relationship of each parameter of chlorine decay with chlorine decay in water distribution system is discussed in the following sections:

4.3.6.1 Hydrogen ion concentration (pH)

Hydrogen ion concentration (pH) correlated weakly and insignificantly with residual chlorine decay as shown in Table 4.3. This finding is consistent with those of Cuesta et al. (2014) and Powell et al. (2004) who observed that pH doesn't influence free chlorine decay in water distribution. This finding suggests that the role of pH in chlorinating water in water distribution system is limited to influencing dissociation (ionization) of free chlorine into hypochlorite ions ($OHCl^{-1}$) and chlorine ions (Cl^{-1}). After free molecular chlorine (Cl) dissociates, the subsequent decay of hypochlorite ions ($OHCl^{-1}$) and chlorine ions (Cl^{-1}) is influenced by other parameters.

4.3.6.2 Temperature of water

Temperature is a water quality parameter that correlated weakly and insignificantly with residual chlorine decay as shown in Table 4.3 This result contradicts findings of Al-Jasser (2006) and Karadirek et al. (2015) that suggested that temperature as a measure of thermal energy catalyzes chemical kinetic reactions. The distribution of temperature in this study was as narrow as 20.10 °C – 27.05 °C with mean of 23.89 °C as shown in Table 4.1. The limitation of temperature to influence chlorine decay could have been due to its small variability. The small variability in temperature could have limited its impact on residual chlorine decay in this study. From Section 2.3.2.1, variability of an independent variable is important in revealing the underlying relationship between it and a target (dependent) response variable.

4.3.6.3 Turbidity

Turbidity is a water quality parameter correlated weakly and not significantly as shown in Table 4.3. Variability of turbidity was low hence one possible reason for not impacting on free chlorine decay in water. However, the composition of the constituents of slow and fast reactants in turbid water was not known. Wu and Dorea (2020) reported that turbidity was an improper proxy for determining chlorine dosage in water. Underground water, as was the case for source of water supply in this study, was less contaminated with organic compounds compared to surface water that normally is more polluted with organics because of open exposure. Underground water tends to have more inorganic pollutants due to dissolution of subsurface minerals than surface water. Therefore, the use of other measures for organic constituents in water such as DOC (dissolved organic compound) and UV-254 nm are better measures for organic reactants in water.

4.3.6.4 Distance and water age

Both distance and water travel time (a good proxy for water age) correlated weakly with residual chlorine decay as shown in Table 4.3. The strong multi-collinearity between distance and water travel time was in order because water travel time is a function of (depends directly on) distance travelled. Distance as a physical water distribution network metric is easier to measure than water age that is water quality parameter. It is therefore advisable to prefer distance to water age in dimensionality (feature) reduction in prediction of residual chlorine.

4.3.6.5 Pressure and velocity

Both pressure and velocity correlated weakly and not significantly with residual chlorine decay as shown in Table 4.3. Pressure and velocity were the two main hydraulic transient variables in water distribution system. The strong multi-collinearity between pressure and velocity was consistent with the law of energy conservation in hydrodynamics. As potential energy (pressure energy) is lost, kinetic energy (velocity energy) is gained. According to Monteiro et al. (2017), EPANET does not simulate water quality parameters well at low water velocities that associate with minimum diffusion. The unexpected average performance of EPANET was also attributed to the few sampling points per distribution line in some water zones.

4.3.6.6 Electrical conductivity

Electrical conductivity is a water quality parameter that correlated moderately with residual chlorine decay as shown in Table 4.3. The distribution of electrical conductivity in this study was as narrow as 65.35 μS/cm – 78.50 μS/cm with mean of 70.01 μS/cm as shown in Table 4.1. These values of electrical conductivity were low compared to those observed by Monteiro et al. (2017) who had high electrical conductivity of 213 μS/cm for comparable pH of 7.4 and temperature of 16.3°C. Similarly, the observed electrical conductivities were also low compared to 110.47 μS/cm in the study of gravity water distribution systems in Ecuador by Garcia et al. (2021) for comparable pH of 7.24. This study was similar to those of Monteiro et al. (2017) and Garcia et al. (2021) because all of them were conducted on gravity water flow schemes in mountainous regions. The small variability in electrical conductivity could have limited its impact on residual chlorine decay in this study. However, compared to travel time, diameter and turbidity, electrical conductivity correlated ten times better. Electrical conductivity suggests dissolved salts hence electrolytic activity. The correlation of electrical conductivity with residual chlorine decay agrees with the findings of Nono et al. (2019) of insufficient or small quantities of inorganics like iron (Fe^{2+}(aq)), manganese (Mn^{2+}(aq)) below 0.3 mg/l in treated water having minimal effect in residual chlorine decay. This suggests that if electrical conductivity values had been high, it would influence residual chlorine decay significantly.

4.4 Assessment of space-time decay of chlorine in water distribution systems

This section presents and discusses the results for specific objective no. 2 which was "to assess space-time decay of chlorine in water distribution systems". The space-time decay of chlorine was investigated in two ways. The first method was statistical modelling and the second was process modelling using EPANET.

4.4.1 Statistical modelling of space-time decay of chlorine in gravity water

Correlation matrix Table 4.3 on p.50 showed that both distance and water age (travel time) correlated weakly with final chlorine. Distance correlated weakly with Pearson's r = - 0.111 with p-value of 0.216. Travel time correlated even more weakly with Pearson's r = - 0.079 with p-value of 0.379. However, there was strong multi-collinearity of Pearson's r = 0.712 with p-value of 0.000 even at the 0.01 significance level between distance and travel time. This suggests that distance would be a stronger and statistically significant predictor of final residual chlorine in the absence of water age (travel time). Backward elimination of travel time in ordinary least square regression revealed that in Table 4.5, distance had standardized beta = - 0.365 with p-value of near 0.000. Curve fitting of linear, logarithmic, inverse, quadratic, cubic and Weibull models is as shown in Figure 4.4 and Figure 4.5.

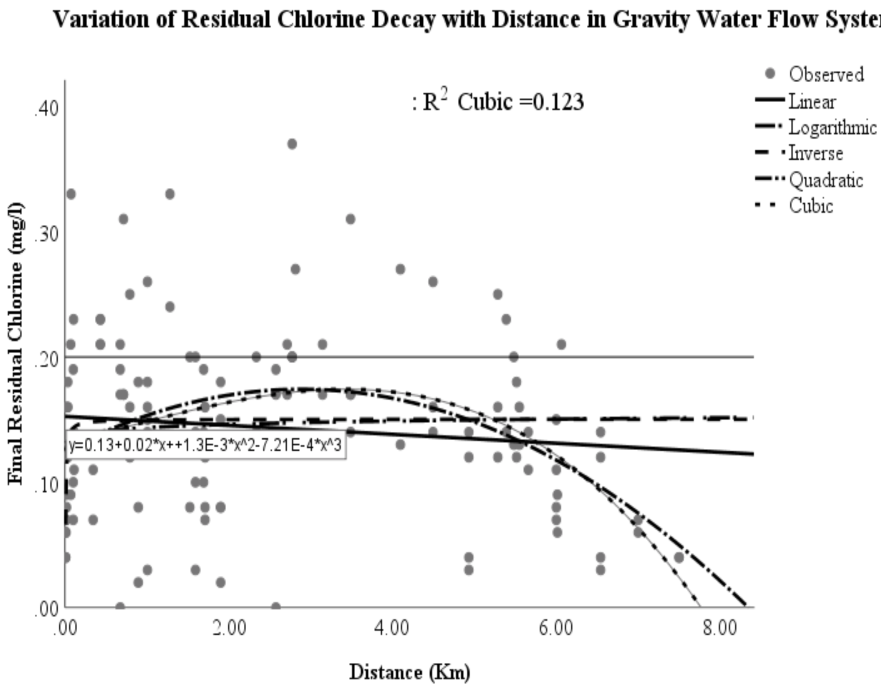


Figure 4. 4: Statistical models for space decay of chlorine in gravity water systems

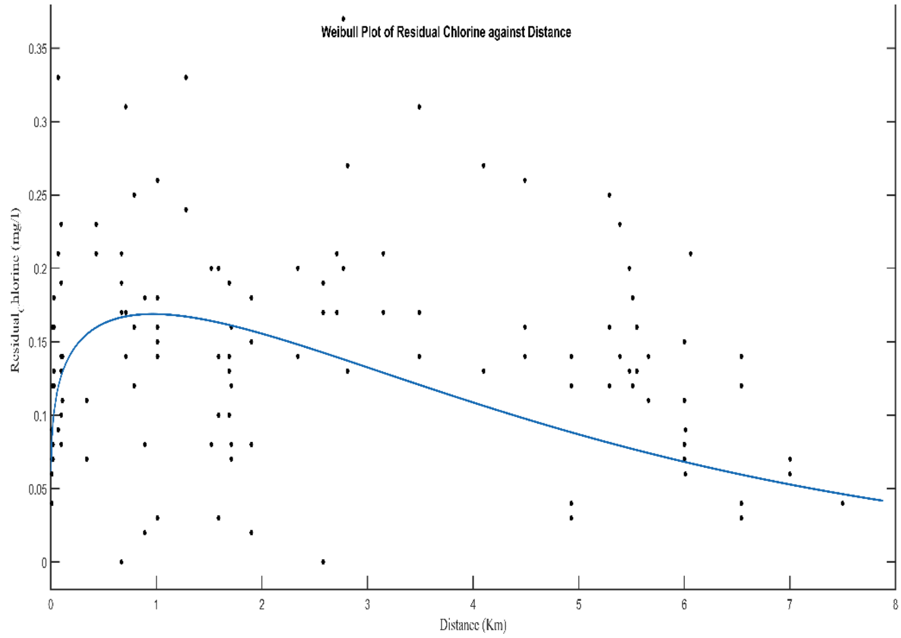


Figure 4. 5: Weibull model for space decay of residual chlorine in gravity water

The summary statistics for evaluating performance of these six models is in Table 4.8.

Table 4. 8: Curve fitting statistics of models for chlorine decay with distance

Item	Model	Model summary				ANOVA statistics	
		R	R^2	Adjusted R^2	Standard error	F-score	p-value
1	Linear	0.110	0.012	0.004	0.073	1.553	0.216
2	Logarithmic	0.098	0.010	0.002	0.073	1.223	0.271
3	Inverse	0.255	0.065	0.058	0.071	0.777	0.004
4	Quadratic	0.340	0.116	0.102	0.069	0.174	0.000
5	Cubic	0.350	0.123	0.102	0.069	5.785	0.001
6	Weibull		- 0.087	- 0.096	0.077		

The associated equations for the models in Table 4.8 are as below.

1. Linear model

$$Final\_Chlorine = 0.153 - 0.004Distance \qquad [4.4]$$

2. Logarithmic model

$$Final\_Chlorine = 0.144 - 0.004\ \ln(Distance) \qquad [4.5]$$

3. Inverse model

$$Final\_Chlorine = 0.150 - 0.004x/Distance \qquad [4.6]$$

4. Quadratic model

$$Final\_Chlorine = 0.125 + 0.034Distance - 0.004Distance\hat{}2 \qquad [4.7]$$

5. Cubic model

$$Final\_Chlorine = 0.131 + 0.016Distance + 0.001Distance^2 - 0.001Distance\hat{}3 \qquad [4.8]$$

4.4.2 Statistical models of space-time decay of chlorine in water

All models were consistent with expectation in showing that residual chlorine decays with distance. This was shown by the negative coefficient of distance in linear, inverse, quadratic and cubic models. For logarithmic model, the low coefficient of logarithmic distance reduces residual chlorine as well. The short rising arm of Weibull model could have been due to chlorine addition from incoming water from upstream. The short buildup of chlorine suppresses its decay causing a short spike in chlorine levels. The first part of the descending limb of Weibull model showed that chlorine decay was fast at start and decreases with time. This clearly shows that chlorine decay represented by chlorine decay constant is not constant. This finding seems to agree with the study of Tiruneh et al. (2019a) that showed that constant bulk chlorine constant underestimates residual chlorine. This was consistent with the single reactant process model for chlorine decay kinetics as postulated by researchers like Bowden et al. (2019) and Jamwal and Kumar (2016). Eventually, the later parts of the descending limb of Weibull model levels to a more sustained gentle slope. This could be due to the slow velocities at downstream points that does not support diffusion and dispersion of chlorine for its consumption as observed by Monteiro et al. (2017). Overall, the R-squared for all the models were low. This indicates that there was high unexplained variance. The high unexplained variance is consistent with the current debate of limited knowledge about residual chlorine decay in drinking water distribution systems as advanced by Soyupak et al. (2011), Vuta and Dumitran (2019) and Zhang et al. (2016). This suggests that a lot of probabilistic variations still exist in water quality in water conveyance. Typical stochastic parameters that should be modelled in understanding chlorine decay in water distribution could be hydraulic transients (pressure and velocity) as advocated by Hyunjun and Sanghyun (2017).

Linear and logarithmic models had very low R^2 and high p-values to justify their choice. Although inverse, quadratic and cubic models had low R^2, they were statistically significant even at the 0.01 significance level. Quadratic and cubic models practically had the same adjusted R^2 = 0.102 and same standard error of 0.069 mg/l. However, the cubic model had much better F-score of 5.785 than quadratic model whose F-score was 0.174. In view of cubic model having the highest F-score and second lowest p-value, it was therefore preferred to other models.

4.4.3 Process modelling of space-time decay of chlorine in gravity water

The laboratory determined chlorine bulk decay, K_b, wall decay, K_w, and total decay, K_T, constants for three gravity water distribution zones of Lirima water gravity scheme. Chlorine decay model was developed by calibrating measured water flows that dissolves and transports chlorine within gravity flow distribution system. The calibrated models for the three water distribution zones were then validated by observed residual chlorine concentrations at water demand junctions/nodes.

4.4.4 Laboratory-determined chlorine decay constants

Table 4.9 presents chlorine decay constants for three water distribution zones in Lirima gravity water flow scheme.

Table 4. 9: Chlorine decay constants in Lirima gravity water distribution zones

(a) Musiye-Nalukwade water distribution zone

Item	Pipe link	K_b (/Day)	K_T (/Day)	K_w (/Day)
1	Musiye to Namawanga	2.29	2.34	0.05
2	Namawanga to Mufutu	12.48	15.60	3.12
3	Natsekhe to Nalukwade	1.76	17.80	16.04

(a) Butiru-Manyeke water distribution zone

Item	Pipe link	K_b (/Day)	K_T (/Day)	K_w (/Day)
1	Butiru Clinic to Buwasike	13.37	25.65	12.28
2	Buwanyera to Bufumo	20.26	58.39	38.13

(c) Butiru – Vermiculite water distribution zone

Item	Pipe link	K_b (/Day)	K_T (/Day)	K_w (/Day)
1	Bunangabo Cell to Bumulatte	7.17	7.34	0.16
2	Bumulatte to Bunabiro	19.21	43.78	24.57

Table 4.9 shows that there is high variation in the chlorine decay constants throughout the distribution network. This could be due to the differences in water quality in terms of the different temperatures, turbidity and electrical conductivities at different locations as indicated in Section 4.2. Wall decay, K_w, constants of 0.05 day^{-1} (Musiye Tank to Namawanga) and 0.16 day^{-1} (Bunangabo to Bumulatte) were comparable to the range of 1.2 day^{-1} (PVC) - 2.16 day^{-1} (medium HDPE) determined by Hallam et al. (2002) for plastic pipes. However, the high wall demand K_w, constants beyond the 1.2 day^{-1} (PVC) to 2.16 day^{-1} range contradicts strongly the studies of Al-Jasser (2006), Bensoltane et al. (2018) and Powell et al. (2004) that considers wall decay constants of plastic pipes to be insignificantly small in residual chlorine modeling. Some of the laboratory determined bulk chlorine decay constants, K_b, such as 1.76 day^{-1} (Natsekhe to Nalukwade), 2.29 day^{-1} (Musiye to Namawanga) and 7.17 day^{-1} (Bunangabo to Bumulatte) were low and realistic. However, they vary from the average 0.55 day^{-1} that Rossman et al. (1994) argued to be the case for treated water. The high values of chlorine bulk decay, K_b, decay constants determined by laboratory analysis could have been due to different laboratory environmental conditions from those inside water distribution system. The differences in temperature, turbidity (as measure of organics) and electrical conductivity (as measure of inorganics) as shown in Section 4.2 affect chlorine bulk decay constants consistent with the assertion of Karadirek et al. (2015). Therefore, several bulk decay constants occur in water distribution network for the varying water quality parameters of temperature, electrical conductivity, organic and inorganic compounds that also vary. The use of single chlorine decay constant to represent the entire water distribution network should be discouraged as it potentially gives wrong answers in process modelling of residual chlorine decay. Variable reaction constants proposed by Tiruneh et al. (2019b) would thus be more realistic though not simple for practical residual chlorine modelling.

4.4.5 Chlorine decay constants based on residence times in water pipes

Total first-order (*n=1*) chlorine decay constants derived from the residence time (water age) in distribution pipes are as shown in Table 4.10 for three gravity water zones

Table 4. 10: Calibration of total chlorine decay constant in water distribution

(a) Musiye-Nalukwade water distribution zone

Item	Pipe link (from junction *i* to next junction (*i+1*))	Velocity (m/s)	Length (m)	Free chlorine residual concentration (mg/l): At start of pipe link	Free chlorine residual concentration (mg/l): At end of pipe link	Total chlorine decay constant. K_T (/Day)
1	Musiye Tank to Namawanga	0.0327	2,722.38	0.21	0.20	- 0.000001
2	Namawanga to Mufutu trading center	0.0327	714.04	0.20	0.17	- 0.000007
3	Mufutu Trading center to Maala	0.0610	1,015.46	0.17	0.16	- 0.000004
4	Maala to Natsekhe wash-out	0.0243	783.82	0.16	0.16	0.000000
5	Natsekhe wash-out to Nalukwade 1	0.0390	490.37	0.16	0.14	- 0.000011

(a) Butiru-Manyeke water distribution zone

Item	Pipe link (from junction *i* to next junction (*i+1*))	Velocity (m/s)	Length (m)	Free chlorine residual concentration (mg/l): At start of pipe link	Free chlorine residual concentration (mg/l): At end of pipe link	Total chlorine decay constant. K_T (/Day)
1	Butiru to Buwasike	0.0749	1,281.44	0.40	0.33	-0.000011
2	Buwasike to Buwanyera	0.0860	2,814.22	0.33	0.27	-0.000006
3	Buwanyera to Bufumo	0.1148	1,503.33	0.27	0.08	-0.000093

(c) Butiru-Vermiculite water distribution zone

Item	Pipe link (from junction *i* to next junction (*i+1*))	Velocity (m/s)	Length (m)	Free chlorine residual concentration (mg/l): At start of pipe link	Free chlorine residual concentration (mg/l): At end of pipe link	Total chlorine decay constant.K_T (/Day))
1	Butiru to Bunangabo	0.0347	1,592.86	0.36	0.08	-0.000033
2	Bunangabo to Bumulatte	0.0647	1,000.00	0.08	0.07	-0.000009
3	Bumulatte to Bunabiro	0.2970	1,000.00	0.07	0.03	-0.000252

The total single-value first-order chlorine decay constants, K_T, determined from residence time using the methods of Hallam et al. (2002), and Madzivhandila and Chirwa (2017) in water distribution pipes were small. These results were consistent with the studies of Al-Jasser (2006), Bensoltane et al. (2018), Monteiro et al. (2017) and Powell et al. (2004) that reported very low hence insignificant chlorine wall demands for non-metallic pipes like plastic pipes in particular. The variation in single decay coefficient in the different pipe links accounts for the variations in other factors from point to point and this is consistent with earlier observation by Tiruneh et al. (2019a). The low values of the calibrated wall demand coefficients are also consistent with low flow velocities as was the case in this study. With the very low flow velocities in this study as shown in Table 4.10 (a), (b) and (c), the flow regime was largely laminar in compliance with Rossman et al. (1994), Karikari and Ampofo (2013), Jamwal and Kumar (2016), and Vuta and Dumitran (2019). Under such laminar flow, chlorine residual decay is mass-transport limited (Hallam et al., 2002) in which chlorine bulk decay dominates. This means that wall decay, K_w, constants are almost zero which is consistent with the values determined from residence time in this study in Table 4.10 (a), (b) and (c). Use of the low values of wall decay, K_w, constants determined from residence time of water in distribution system would therefore be appropriate for modelling residual chlorine decay in plastic pipes. Both water transmission and distribution pipes in Lirima gravity water flow scheme were plastic.

4.4.6 Calibration and validation of space-time chlorine decay in EPANET

This section presents results for models calibrated by measured flows and cross-validated by chlorine concentrations. The purpose of this was to enable a calibrated water flow model to carry dissolved chlorine from one part to another in water distribution network. The calibrated water flow models were then cross-validated by observed residual chlorine concentrations at the same water demand junctions. Model cross-validation was done by using measured residual chlorine concentrations on different days and run times from those used in water demand flow calibration. Table 4.11 shows EPANET hydraulic model calibration and validation for the three water distribution zones of Lirima gravity flow scheme.

Table 4. 11: EPANET hydraulic model calibration and validation for gravity water

(a) Musiye-Nalukwade water distribution zone

Calibration data	Demand model calibration Statistics		Test data	Chlorine model cross validation statistics	
	Mean error (m3/d)	R-Squared		Mean error (mg/l)	R-Squared
Day1_Run3 (Evening)	1.749	0.997	Day8_Run1 (Morning)	0.168	0.722
Tuesday 9th Feb 2021			Wed 24th Feb 2021		

(b) Butiru-Manyeke water distribution zone

Calibration data	Demand model calibration statistics		Test data	Chlorine model Cross Validation Statistics	
	Mean error (m3/d)	R-Squared		Mean error (mg/l)	R-Squared
Day5_Run1 (Afternoon)	0.000	1.000	Day2_Run1 (Afternoon)	0.002	0.997
Tuesday 23rd Feb 2021			Wed 10th Feb 2021		

(c) Butiru-Vermiculite water distribution zone

Calibration data	Demand model Calibration statistics		Test Data	Chlorine model Cross validation statistics	
	Mean error (m3/d)	R-Squared		Mean error (mg/l)	R-Squared
Day3_Run2 (Afternoon)	0.653	0.999	Day7_Run1 (Morning)	0.773	0.937
Thursday 11th Feb 2021			Tue 23rd Feb 2021		

Table 4.11 shows varied performance of EPANET model in space-time decay of residual chlorine with R-squared ranging from 72% with model error estimate of

0.168 mg/l to 98% with model error estimate of 0.002 mg/l. The result shows that the performance of EPANET varies with water distribution zones within the same gravity flow scheme. This may be due to the variation in the distance from zonal water reservoir to the furthest water consumption time. The deposits of reactants within the water distribution pipes may also vary due to differences in network infrastructure maintenance schedules These and other unknown factors can influence space-time decay of residual chlorine even within the same water distribution zone.

4.4.7 Algorithm result for chlorine residuals in gravity flow distribution

Using the algorithm in section 3.6.2.2, Table 4.12 gives chlorine residuals for secondary chlorination based on this algorithm.

Table 4. 12: Algorithm results for secondary chlorination in gravity water flow

(a) Musiye-Nalukwade water zone 5.34 Km long

Item	Initial chlorine (mg/l)	Depth of water in tank (m)	Pipe DN (mm)	Velocity in pipe (m/s)	Residual chlorine at pipe end (mg/l)	Limit of distance (Km)
1	0.32	0.50	100	3.13	0.3148	5.34
		0.75	100	3.84	0.3158	5.34
2	0.36	0.50	100	3.13	0.3542	5.34
		0.75	100	3.84	0.3552	5.34
3	0.44	0.50	100	3.13	0.4329	5.34
		0.75	100	3.84	0.4342	5.34

(a) Butiru-Vermiculite water zone 4.9 Km long

Item	Initial chlorine (mg/l)	Depth of water in tank (m)	Pipe DN (mm)	Velocity in pipe (m/s)	Residual chlorine at pipe end (mg/l)	Limit of distance (Km)
1	0.37	0.50	75	3.13	0.3668	2.90
	0.37	0.50	63	4.44	0.3629	2.90
2	0.37	0.25	75	2.21	0.3600	2.90
3	0.41	0.50	75	3.13	0.3629	2.90
			63	4.44	0.3620	3.00
		0.25	75	2.21	0.3989	3.00
4	0.49	0.50	75	3.13	0.4806	2.90
		0.25	75	2.12	0.4767	2.90

(a) Butiru-Manyeke water zone 7.56 Km long

Item	Initial chlorine (mg/l)	Depth of water in tank (m)	Pipe DN (mm)	Velocity in pipe (m/s)	Residual chlorine at pipe end (mg/l)	Limit of distance (Km)
1	0.31	0.50	75	3.13	0.3067	5.34
		0.75	75	3.84	0.3158	1.96
2	0.35	0.50	75	3.13	0.3542	5.34
		0.75	75	3.84	0.3552	5.34
3	0.43	0.50	100	3.13	0.4329	5.34
		0.75	100	3.84	0.4342	5.34

4.4.8 Discussion of algorithm results for secondary chlorination in water zones

The aim of the algorithm for secondary chlorination was twofold as follows: (1) to ensure that there is adequate and safe residual chlorine in all water distribution zones that meets the range of 0.2 – 5.0 mg/l as per standards of WHO (2014). The second objective of the algorithm for secondary chlorination was to ensure that water flow velocities are within the range of 0.3 – 3.5 m/s as specified in Design Guidelines for Water Supply Infrastructure manual by the Directorate of Water and Development (Uganda) (2013). Simulation of this algorithm shows that in order to operate water reservoir tank to ensure flow velocity equal to the regulatory maximum limit of 3.5 m/s requires water depth of 0.625 m. Both intermediate reservoir and break-pressure tanks in Lirima have maximum (water over-flow) design height of 3.5 m. This means that the tanks will be operated at only 17.8% capacity. During data collection, the observed depth of water in the tanks of the gravity scheme was between 0.30 m – 0.50 m which falls within the 17% capacity operation for below maximum velocity flows. Future designs of especially intermediate reservoirs that can act as secondary chlorination points should ensure that the plan area of tanks should be increased for maximum design height of 0.625 m for a given design volume of such tanks.

The decay of residual chlorine from an upstream point to a downstream point based on flow velocity was low. This could have been because of the low velocities that ranged from 0.001 – 0.10 m/s averaging 0.04 m/s as shown in Table 4.1 of descriptive statistics for water quality and water system parameters. Low flow velocities are largely laminar and therefore minimize turbulence (Jamwal and Kumar, 2016; Kim et al., 2014; Vuta and Dumitran, 2019) that is necessary for mass-transport (Stoinov and Aisopou, 2014). The dominant process reaction model most likely was phase two with slow-reactants. This assumption was based on more than 30 minutes lapse of time wherein fast reactants are consumed at treatment plant during initial chlorine dosage. The reaction therefore conforms to Wu and Dorea (2020) for slow decay. Diffusion and dispersion of dissolved chlorine was low due to the low velocities consistent with what Monteiro et al. (2017) asserted. Smaller pipe sizes than 75 mm transmit water at excessive velocities higher than the limiting velocity of 3.5 m/s as specified by Directorate of Water and Development (Uganda) (2013). This implies that this algorithm is applicable to pipe sizes of 75 mm and larger.

4.5 Comparison of model performance in predicting residual chlorine

This section presents and discusses the results for specific objective no. 3 which was "to compare performance of various models". Physical (process) model used was EPANET applied in modelling residual chlorine decay in three water distribution zones. Statistical (data-driven) models developed for residual chlorine decay analysis include multiple linear regression model, lasso regression model, ridge regression model, decision tree regression model, random forest and artificial neural network. The statistically significant water quality parameters with low multi-collinearity identified in Section 4.3.4 were used in these statistical models.

4.5.1 EPANET model

Table 4.13 presents the performance of EPANET in modelling residual chlorine decay in three separate zones and for all the combined zones.

Table 4. 13: Performance of EPANET in modeling chlorine decay in water distribution

Item	Water Zone	Sample Points	Performance Metrics	
			RMSE (mg/l)	MAE (mg/l)
1	Musiye	91	0.4250	0.63
2	Vermiculite	45	0.8718	1.02
3	Manyeke	10	0.0000	0.00
4	Combined	149	0.5829	1.02

Legend: RMSE = Root Mean Square Error, MAE = Maximum Absolute Error

Table 4.13 shows the analysis results for performance of EPANET in predicting residual chlorine decay in water distribution network. The size of datasets used clearly influences the analysis results. Small datasets tend to overfit while larger datasets tend to model chlorine decay better.

4.5.2 Multi-linear regression model

The equation for final residual chlorine in water distribution network based on statistically independent and statistically significant water quality and water distribution system parameters was:

$$Final\ chlorine\ = 0.415\ +\ 0.548\ Initial\ chlorine\ -\ 0.012\ distance\ -\ 0.005\ EC. \ldots\ldots.[4.9]$$

where final chlorine and initial chlorine are measured in mg/l, distance is measured in Km and EC (electrical conductivity) is measured in µS.

4.5.2.1 Discussion of multi-linear regression model

From Equation 4.9, calculation of initial chlorine required to ensure final residual chlorine of 0.2 mg/l can be inferred and obtained from Equation 4.10 as follows:

$$Initial\ chlorine = \ ((0.2 + 0.012D + 0.005EC - 0.415))/0.548 \ldots\ldots\ldots\ldots.[4.10]$$

where:

D = Distance of water distribution pipe from chlorine dosing point (upstream) to water consumption draw-off point (downstream)

This equation explains 63% of residual chlorine decay in gravity water distribution with error of 0.045 mg/l. In accounting for over 60% of residual chlorine in gravity water distribution, the final estimated residual chlorine using this model is expected to be within 0.135 mg/l (three standard deviations) from the true residual chlorine at 95% confidence level. This error is reasonably small hence tolerable. This model structure is simple as it uses only three parameters of (1) initial chlorine dosage at water treatment plant, (2) distance from water treatment plant to any water consumption point downstream and (3) electrical conductivity that are easy and to determine and are conveniently controllable. The standardized beta coefficients of the OLS regression model show the importance of initial chlorine (0.795), electrical conductivity (0.365) and distance (0.175) as predictors of final residual chlorine in that decreasing order.

4.5.3 Principal component analysis regression model

The equations for the three principal components developed from the rotated component matrix for three principal component solution based on electrical conductivity, distance (length) and initial chlorine (chlorine dose) as the key predictors were as follows:

$$PC1 = 0.999\ Electrical\ conductivity \quad [4.11]$$

$$PC2 = 0.988\ Distance \quad [4.12]$$

$$PC3 = 0.988\ Initial\ Chlorine \quad [4.13]$$

The linear regression model with the three principal components as predictors developed in Equation 4.11 to Equation 4.14 is presented in Table 4.13.

Table 4. 14: Details of principal component analysis-based regression model

(a) Model summary

Model	R	R squared	Adjusted R squared	Std. error of estimate
1	0.788[a]	0.620	0.611	0.04564

a. Predictors: (Constant), chlorine dose, length, electrical conductivity

(b) Coefficients[a]

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.	Collinearity statistics	
	B	Std. error	Beta			Tolerance	VIF
(Constant)	0.144	0.004		35.597	0.000		
electrical_ conductivity	- 0.014	0.004	- 0.189	-3.423	0.001	1.000	1.000
length	- 0.017	0.004	- 0.230	- 4.155	0.000	1.000	1.000
chlorine_dose	0.053	0.004	0.729	13.171	0.000	1.000	1.000

a. Dependent Variable: residual chlorine

Legend: B = Unstandardized bete coefficient , t = test statistic = (B/std.error) , VIF = Variable Inflationary Factor

The resulting linear regression model from the three principal components in Table 4.14 as predictors for final residual chlorine in water reticulation is as shown in Equation 4.14.

$$Resdual\ chlorine = 0.144 - 0.014\ \ EC - 0.017\ length + 0.053\ chlorine_{dose} \ldots \ldots [4.14]$$

4.5.3.1 Comparison of OLS linear regression and PCA regression models

The performance of linear and PCA regression models in predicting residual chlorine decay in gravity water reticulation systems is compared in Table 4.15. The performance comparison is based on initial chlorine, distance and electrical conductivity that were found to be the three parameters contributing to over 90% of residual chlorine decay.

Table 4. 15: Performance of linear and principal component regression models

Item	Model characteristic statistics	Regression models: Linear regression model	Regression models: Principal component regression model	Deviation from each other
1.0	**Model summary**			
1.1	Pearson's *r*	0.793	0.788	0.005
1.2	R-squared	0.623	0.620	0.003
1.3	Adjusted R-squared	0.619	0.611	0.008
1.4	Standard error of model estimate	0.0453	0.0456	0.0003
2.0	**ANOVA statistics**			
2.1	F-score	69.25	67.48	1.77
2.2	Model significance (p-value)	0.000	0.000	0.000
3.0	**Collinearity statistics (VIF) for key predictors**			
3.1	Electrical conductivity	1.005	1.000	0.005
3.2	Distance (length)	1.103	1.000	0.103
3.3	Initial chlorine (Chlorine dose)	1.004	1.000	0.004

4.5.4 Lasso regression model

Lasso regression model performed as shown in Table 4.16.

Table 4. 16: Performance metrics of lasso regression model for residual chlorine

Item	Dataset	Performance metrics		
		RMSE	**MAE**	**R^2**
1	Training	0.07	0.05	0.00
2	Test	0.08	0.06	-0.03

Legend: RMSE = Root Mean Square Error, MAE = Maximum Absolute Error

Table 4.16 shows that lasso regression marginally over-fitted data. This is because the train data statistics of RMSE and MAE were both slightly better than the corresponding statistics for test data. Since the differences are marginal, lasso regression can be compared with other statistical models.

4.5.5 Ridge regression model

Ridge regression model performed as shown in Table 4.17.

Table 4. 17: Performance metrics of ridge regression model for residual chlorine

Item	Dataset	Performance metrics		
		RMSE	MAE	R^2
1	Training	0.05	0.04	0.42
2	Test	0.06	0.05	0.39

Legend: RMSE = Root Mean Square Error, MAE = Maximum Absolute Error

Table 4.17 shows that ridge regression marginally over-fitted data. This is because the train data statistics of RMSE, MAE and R-squared were all slightly better than the corresponding statistics for test data. Since the differences are marginal, ridge regression can be compared with other statistical models.

4.5.6 Lasso and ridge regression models

Lasso regression model generalizes well as its training and test performance metrics compare closely. However, it has a low R^2. This can be attributed due to elimination of variables that correlate weakly with final residual chlorine.

Ridge regression model also generalizes well as its training and test performance metrics compare closely. Its test R^2 is low because of inclusion of variables that correlate weakly with final residual chlorine.

4.5.7 Decision tree model

The model in Table 4.18 presents the training and test performance and schematic decision tree regressor respectively of decision tree as analyzed by Python.

Table 4. 18: Performance metrics of decision tree model for residual chlorine

Item	Dataset	Performance metrics		
		RMSE	**MAE**	**R^2**
1	Training	0.03	0.01	0.41
2	Test	0.03	0.01	0.41

Legend: RMSE = Root Mean Square Error, MAE = Maximum Absolute Error

4.5.8 Random forest model

The model in Table 4.19 presents the training and test performance of random forest as analyzed by Python.

Table 4. 19: Performance metrics of random forest model for residual chlorine

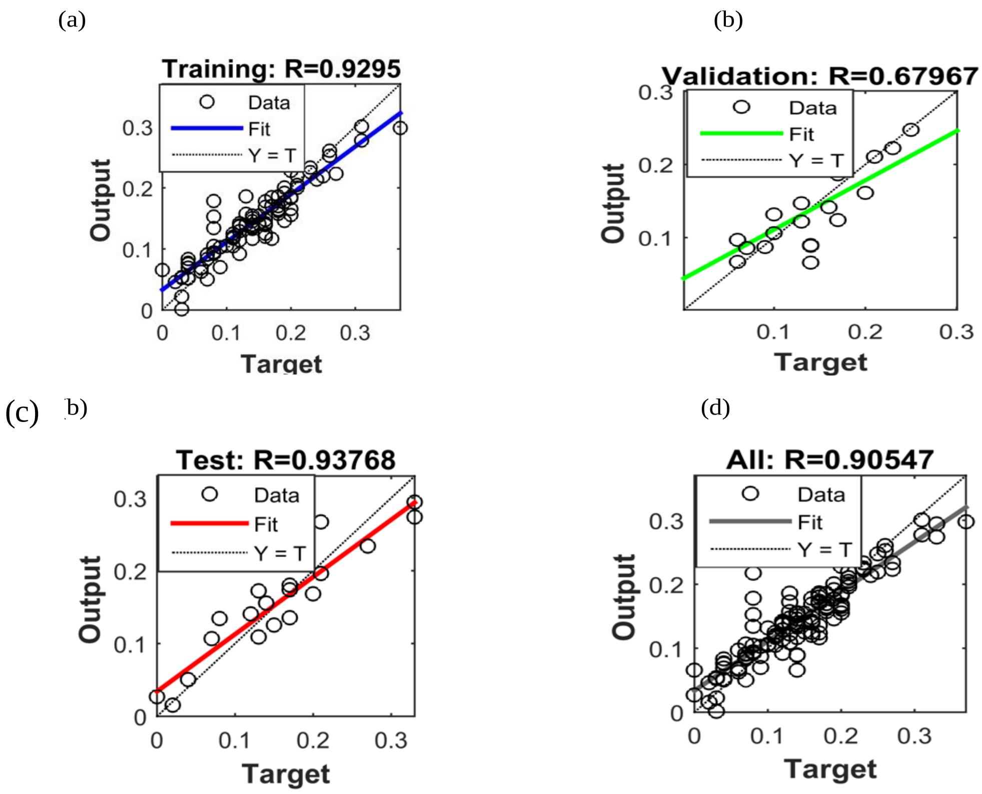
Item	Dataset	Performance metrics		
		RMSE	**MAE**	**R^2**
1	Training	0.02	0.02	
2	Test	0.05	0.04	0.55

Legend: RMSE = Root Mean Square Error, MAE = Maximum Absolute Error

Decision tree that normally tends to overfit in most cases compared to other regression models had comparable performance to random forest as shown in Table 4.19. This in part could have been due to pruning the decision tree to a maximum depth of four to minimize its inherent tendency to overfit. Since the random forest is an ensemble of decision trees as base models to minimize individual decision tree overfitting, it follows that the random forest model performed expectedly at least as well as decision tree model. This conforms to domain knowledge on decision trees and random forests.

4.5.9 Artificial neural network model

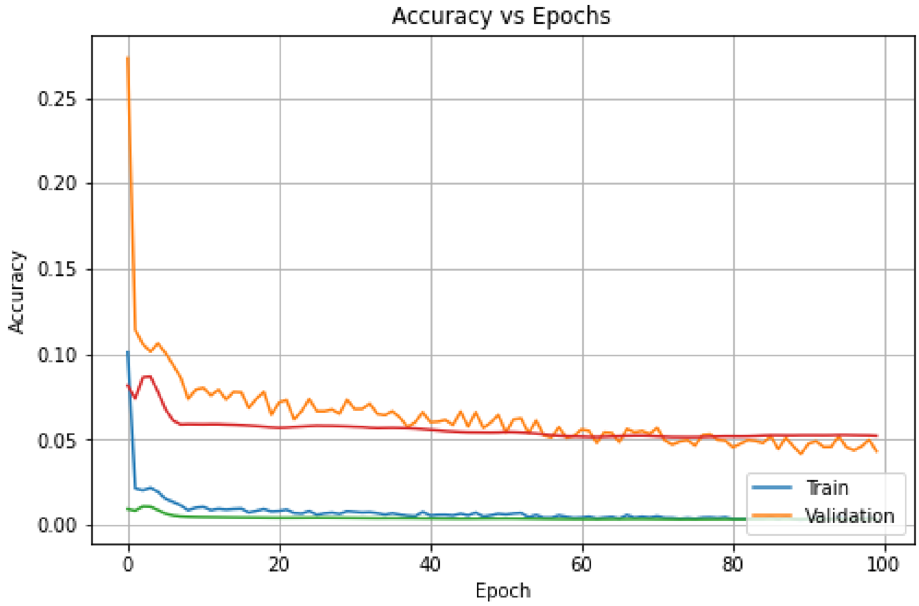
Performance of artificial neural network for residual chlorine in gravity water distribution is summarized in Figures 4.6. The outputs and targets in Figure 4.6 are the predicted and actual residual chlorine concentrations respectively. The R-score values were 0.93, 0.68 and 0.94 for training, validation and test datasets respectively.



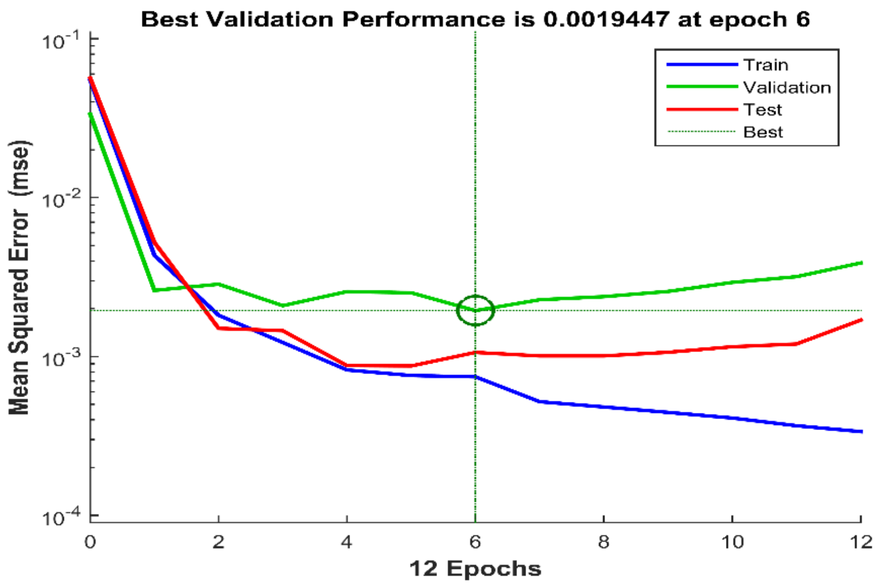
Legend: Y = Output T = Target All= Whole (unsplit) dataset

Figure 4. 6: Training and validation performance of artificial neural network model for residual chlorine

The performance accuracy plots for the generated artificial neural network for residual chlorine were as shown in Figure 4.7.



(a) Python-generated



(b) MATLAB-generated

Figure 4. 7: Training and validation accuracies in artificial neural network

Figure 4.7 (a) shows better accuracy for validation data than that for training data. Similarly, Figure 4.7 (a) shows better accuracy in terms of lower RMSE at lower epochs. Both Figure 4.7 (a) and (b) emphasize better results at lower epochs. This means that there should be care not to over-train in artificial neural network modelling.

Figure 4.8 shows the architecture / structure of the artificial neural network for residual chlorine decay in gravity water distribution.

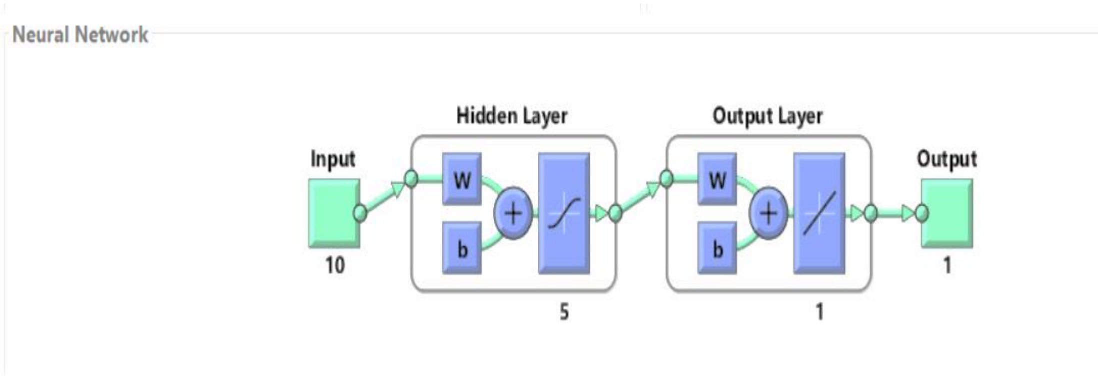


Figure 4. 8: Structure of artificial neural network model for residual chlorine

Ten predictor variables (inputs) were fed in the artificial neural network (ANN) that had five hidden layers. The single output was the final residual chlorine as target variable.

4.6 Performance of chlorine decay models in predicting chlorine decay in water

Table 4.20 summarizes the performance of process and statistical models based on performance metrics from test (validation) data. The statistics for multi-linear regression model are those contained in Table 4.5 (Linear regression model for statistically significant parameters of chlorine decay in water distribution network).

Table 4. 20: Performance of chlorine decay models in water distribution network

Item	Model	Goodness of fit statistics		Performance accuracy (mg/l)		
		Adjusted R^2	Std/ error (mg/l)	RMSE (mg/l)	MAE (mg/l)	Rank
1	EPANET	0.235		0.43	0.63	8
2	OLS Linear regression	0.619	0.0453			2
3	Lasso regression			0.06		7
4	Ridge regression			0.05		6
5	PCA regression	0.611	0.0456			3
6	Decision tree	0.409		0.05	0.04	5
7	Random forest	0.545		0.05	0.04	4
8	Artificial neural network	0.938		0.04	0.05	1

Legend: RMSE = Root Mean Square Error, MAE = Maximum Absolute Error

The eight models in Table 4.20 were ranked using performance accuracy statistics as mentioned in methodology section 3.6.2 (data analysis method for specific objective No. 3). Both multi-linear regression and random forest models were consistent in emphasizing the importance of water quality (initial chlorine and electrical conductivity) and physical (distance) parameters in residual chlorine decay. The standardized beta coefficients of multi-linear regression in Section 4.5.2 and importance ranks of random forest in Section 4.5.9 are summarized in Table 4.21 for ease of appreciation.

Table 4. 21: Influence of parameters in residual chlorine decay in gravity water

Chlorine decay parameter category	**Chlorine decay parameter**	**Linear regression: standardized beta coefficients**	**Random forest: feature importance rank**	**Order of decreasing importance**
Water Quality	Initial chlorine	0.795 (59.55%)	0.4696 (62.79%)	1
	Electrical conductivity	0.365 (27.34%)	0.2261 (30.23%)	2
Water System	Distance	0.175 (13.11%)	0.0522 (6.98%)	3

Table 4.21 further shows that water quality parameters contribute between 87% - 93% of residual chlorine decay requirements compared to 7% - 13% of water system contribution. This calls for more control in setting the right chlorine dosage during water treatment. Once water is treated, electrical conductivity in water distribution network (pipes and tanks) accounts for residual chlorine during water transport. In design of new water distribution systems and improvement of existing ones, the distance of water distribution network from water treatment plant to water delivery points should be well optimized to minimize residual chlorine loss.

4.7 Identification of model(s) for predicting residual chlorine decay in water

This section presents and discusses the results for specific objective no. 4 which was "to identify the appropriate model(s) for predicting residual chlorine decay". The factors that were considered in selection of model include (1) generalizability identified by minimum overfitting on test (validation) data, (2) dimensionality control in terms of balance between model density versus model sparsity and (3) model interpretability. Each of these factors are required in a good model. Based on performance metrics in Table 4.20, the order of model performance from best to worst is artificial neural network, multi-linear regression model, principal component regression model, random forest, decision tree, ridge regression model, lasso regression model and EPANET.

4.7.1 Model factors and model ranks

Table 4.20 shows that artificial neural network (ANN) performed best because of having the highest R-score of 0.938, least RMSE of 0.04 mg/l and second lowest MAE of 0.05 mg/l (just above the lowest MAE of 0.04 mg/l). Besides, Figure.4.6 also shows that the test R-score of ANN of 0.9377 as in Figure.4.6(c) was more than its training R-score of 0.9295 as in Figure.4.6(a). This suggest that the ANN model did not overfit. However, ANNs have the weakness of difficulty in interpretation. The contribution of each parameter in predicting a target variable like final chlorine is important. This is not the case with ANN because it is a "blackbox".

Regression models like multi-linear ordinary least square (OLS), LASSO (L1 regularization model) and ridge (L2 regularization model) regression models are easy to interpret because their regressor (predictor parameter) coefficients shows how much the target variable (in this case final residual chlorine) changes with unit change in each predictor variable. However, model parsimony requires restriction in the number of predictors to avoid the model from underfitting (when too few predictors are used) as opposed to model from becoming complex (when too many predictors are used). In this regard OLS, and ridge regression models are better than lasso regression model in terms of avoiding too few predictors. The performance of the tree-based models (decision tree and random forest) is also encumbered by interpretation weakness for lack of knowledge about how much each predictor influences chlorine decay.

Therefore, regression models tend to appeal more for choice because of ease of interpretation. The influence of each parameter as shown by the beta coefficients in regression models is important in controlling and managing each predictor with respect to the desired final residual chlorine at water consumption points. This leaves us with OLS, lasso, ridge and PCA regression models.

Lasso regression model eliminates insignificant models. This is good as long as the remaining number of predictors are not too few to underfit final model. On the other hand, ridge regression model allows the insignificant model to remain with their insignificant contributions. This may abuse the consideration for model parsimony that is about avoiding overfitting. PCA regression model also suffers from interpretation because its predictors are linear combinations of individual predictors. Table 4.15 (comparison of performance of OLS and PCA regression models) showed that OLS and PCA regression models performed almost the same. The R-squared values were 61.9% (OLS), 61.1% (PCA); standard error of estimates were 0.0453 mg/l (OLS) and0.0456 mg/l (PCA) and ANOVA F-scores were 69.25 (OLS) and

67.48 (PCA). In view of requirement for interpretability in modelling, OLS emerges a better model than PCA. Therefore, in view of the above considerations and arguments, the OLS regression model is a better choice. However, the other models can be used as a background check on the performance of the OLS regression model itself.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

The main purpose of this research was to compare space-time performance of chlorine decay models in order to identify user-friendly decay model for gravity water distribution systems. This goal had four specific objectives of (1) characterization of water distribution parameters of chlorine decay, (2) assessment of space-time decay of chlorine in gravity flow distribution systems, (3) comparison of performance of various models that predict residual chlorine concentration and (4) identification of appropriate model(s) for residual chlorine decay in gravity water systems.

The associated research questions for research specific objectives were (1) which water quality and water system parameters influence space-time decay of chlorine in gravity water systems? (2) how does chlorine in gravity distribution systems decay in space and time? (3) how does the various models that predict residual chlorine concentration in gravity distribution systems compare in performance? and (4) which model(s) is/are appropriate for managing residual chlorine decay in gravity water system?

The research was done on Lirima Gravity Flow Scheme found in Mount Elgon in eastern Uganda. This water scheme is owned and operated by National Water and Sewerage Corporation.

5.2 Summary of findings

Chlorine dosage was being done centrally at the treatment plant. The chlorine dosage ranged from 0.70 to 1.0 mg/l with mean dosage of 0.76 mg/l as shown in appendix Table A.1. This dosage was consistent with the WHO (2014) recommended range of 0.2 -5.0 mg/l. However, residual chlorine at most consumer points that averaged 0.14 mg/l was below the lower limit of 0.2 mg/l. The turbidity of the water was low with mean of 0.96 NTU. The insignificance of turbidity to explain chlorine decay was consistent with the findings of Wu and Dorea (2020) who discouraged turbidity for determining chlorine dosage in water. The electrical conductivity of treated water at Lirima treatment plant was high ranging from 95.4 µS -141.0 µS with mean of 122.4 µS as shown in Table A.1. This compares sharply with the low electrical conductivity in water transmission and distribution lines ranging from 65.4 µS – 78.50 µS with mean of 70.01 µS as shown in Table 4.1. Initial chlorine concentration, electrical conductivity and distance correlated significantly with chlorine decay. However, pipe diametre, turbidity, pH, temperature and pressure correlated weakly and insignificantly with residual chlorine.

Based on generalizability, dimensionality control and interpretability factors, linear regression with R-squared of 63% and model error of 0.045 mg/l was the best model for residual chlorine decay. The performance of models in Table 4.20 in predicting decay of residual chlorine in gravity water systems ranked from best to worst is artificial neural network, multi-linear regression model, principal component regression model, random forest, decision tree, ridge regression model, lasso regression model and EPANET.

5.3 Conclusions

The following conclusions based on results for research specific objectives in Section 1.5 and research questions in Section 1.6 are made.

5.3.1 Conclusion on characterization of parameters of chlorine decay in water

Initial chlorine dose and electrical conductivity together on average influence 90% of chlorine decay in gravity drinking water distribution system. Distance from water treatment to water consumption point on average influence 10%. The other parameters like pH, turbidity and temperature had insignificant effect on residual chlorine decay in water distribution network. These results support the conclusion that water quality parameters influence residual chlorine decay much more than physical parameters.

5.3.2 Conclusion on space-time decay of chlorine in water distribution systems

Distance had p-value of less than 0.01 even at the 0.01 significance level resulting in regression model with adjusted R-squared of 62% with standard estimate error of 0.045 mg/l. These results support the conclusion that distance is a much better space-time chlorine decay parameter.

5.3.3 Conclusion on comparison of performance of models for chlorine decay

The better performance of statistical models supports the conclusion that they should be preferred to use of EPANET model.

5.3.4 Conclusion on identification of models for predicting residual chlorine decay

The better performance of multi linear regression model with known effect of each water quality and physical parameter that is interpretable supports the conclusion that they should be used in predicting of residual chlorine decay in water distribution system.

5.4 Limitations and future action

The following limitations should be noted:

5.4.1 Poor performance of EPANET

EPANET model did not predict residual chlorine decay well. This is attributed to the low velocities that averaged 0.04 m/s that could not allow diffusion of chlorine in water. National Water and Sewerage Corporation advised that the water sampling points should not be close if there was to be observable differences in residual chlorine concentrations between any two adjacent points. This advice was based on the low chlorine dosage at the treatment plant that ranged from 0.70 mg/l-1.0 mg/l that averaged 0.76 mg/l as shown in Table A.1. Consequently, the interval of water sampling points was large ranging from 0.8 Km-1.2 Km. This strategy ensured that there was noticeable decay in residual chlorine from one upstream point to the next downstream point to introduce variability needed in modelling. The unavoidable effect of this strategy was the few sampling points per distribution line in some water zones ranging from four to six. Each hydraulic analysis for a particular run on a given gravity water distribution zone depends on the number of sampling points. This requirement of hydraulic connectivity is one of the key basis for EPANET water quality modelling. Although several runs were performed on different days and at different times within a day on given gravity water distribution zone, the challenge of reduced variability in residual chlorine due to few water sampling points on a given water distribution line remained.

5.4.2 Use of one process model against several statistical models

EPANET was the only process-based model that was used in comparison with seven statistical models. The decision to use EPANET as the only process model was influenced by its popularity and dominant in water industry and practice and availability as well compared to other water modelling tools like WaterCAD, Mike, Pipe, Civil Designer. etc. It is not proven that any of these other process models could have performed worse than EPANET against any of the statistical models used in this study. Therefore, there is need for a separate and similar study to find out how these models would perform compared to statistical models

5.4.3 Non-variable decay constant in EPANET

The assertion by Soyupak et al. (2011) that chlorine decay assumes a deterministic reaction kinetic model may not be true at all times. This is because the distribution of parameters that influenced chlorine decay like chlorine decay coefficient, electrical conductivity etc. were non-uniform in this study. This could have affected the performance of EPANET that was modelled on the basis of single-reactant and constant decay constants as was assumed in Section 1.7.3. Although effort was made by using different decay constants in individual pipe segments, this may not have been adequate to reflect the true state of events even in individual pipe segments because some pipe segments were long.

5.4.4 Narrow variability in some water quality parameters

Temperature and turbidity were envisaged as key water quality parameters that influence residual chlorine decay in water distribution networks. However univariate analysis of parameters of residual chlorine decay in water reticulation networks found temperature and turbidity to have small variabilities as shown in Table 4.1 (descriptive statistics of chlorine decay parameters). The small variabilities could have affected the role of these parameters in statistical modelling for chlorine decay.

5.4.5 Cross-sectional study

This was cross-sectional study as it was conducted during the dry season. The results are therefore applicable to the months of February and March only. Therefore, there is need to conduct similar study during wet months of the year to introduce adequate variation in especially temperature given that temperature is a very important influence in chlorine decay in water distribution systems. This will allow development of a generalized model that applies throughout a year (Garcia-Avila et al., 2020). The pipelines in the study area were of plastic (HDPE) material. Similarly, application of results is therefore limited to plastic pipelines.

5.5 Recommendations

The following recommendations follow from this research:

1. National Water and Sewerage Corporations in particular and water utilities in general should remove dissolved salts as much as possible at water treatment plants to minimize uptake of residual chlorine immediately at start of water supply.
2. National Water and Sewerage Corporations in particular and water utilities in general should clean (flush) pipeline routinely and regularly to avoid build-up of sediments of salt precipitates that can re-dissolve to deplete residual chlorine.
3. National Water and Sewerage Corporations in particular and water utilities in general should adopt water zoning approach for managing residual chlorine in

gravity water distribution. Intermediate reservoirs that traditionally are used to balance pressure and regulate water demand should as well be used as points of secondary chlorination (online chlorine boosting) points. This will reduce distance and boost final residual chlorine to above 0.2 mg/l. Table 5.1 contains the recommended initial chlorine at water zone intermediate tanks. The initial chlorine concentrations are based on minimum, mean and maximum electrical conductivities observed in gravity distribution network in Table 4.1. Equation 4.10 was used to determine these proposed initial chlorine concentrations.

Table 5. 1: Recommended initial chlorine for each Lirima gravity water zones.

S. N	Water zone	Farthest distance (Km)	Initial chlorine concentration (mg/l)		
			Electrical conductivity (μSiem/cm)		
			Minimum 65.40	Mean 70.01	Maximum 78.50
1	Musiye-Nalukwade	5.34	0.32	0.36	0.44
2	Butiru-Manyeke	4.90	0.31	0.35	0.43
3	Butiru-Vermiculite	7.56	0.37	0.41	0.49

4. Regulatory bodies should revise design manuals to allow future designs of intermediate water tanks that should be used as secondary chlorination points to (1) be limited to maximum water depth of 0.63 m to minimize high and excessive flow velocities and (2) plan area be increased to accommodate design volume.
5. The practice of using laboratory determined bulk chlorine decay constant to design decay in water distribution networks should be avoided because the environmental conditions in laboratories and within water distribution system differ. Instead, chlorine decay constants determined from water age should be used.

6. Not until future works show otherwise, data-driven models should be preferred over process models in water quality management to achieve predictability of water quality parameters in water distribution networks.

REFERENCES

Agatonovic-Kustrin, S., and Beresford, R. (2000) 'Basic concepts of artificial neural network (ANN) modelling and its applications in pharmaceutical research', *Journal of Pharmaceutical and Biomedical Analysis*, 22(2000), 717-727. [Online] https://www.elsevier.com/locate/jpba

Al-Jasser, A. O. (2006) 'Chlorine decay in drinking-water transmission and distribution systems: pipe service age effect,' *Water Research*, 4(2007), 387-396. [Online] doi: https://doi.org/10.1016/j.watres.2006.08.032

Alsaydalani, M.O.A. (2019) 'Simulation of pressure head and chlorine decay in a water distribution network: a case study', *The Open Civil Engineering Journal*, 13, 58-68. [Online] doi: https://doi.org/10.2174/1874149501913010058

Angulo, F., Urueta, E., Valverde, G. and Paternina, O. (2017) 'Cartagena's water distribution system', *18th International Conference on Water Distribution System Analysis,* WDSAI2016", 186(2017), 28-35. [Online] doi: https://doi.org/10.1016/j.proeng.2017.03.204

Araya, A. and Sanchez, L.D. (2018) 'Residual chlorine behaviour in a distribution network of a small water supply system', *Journal for Water, Sanitation, Hygiene and Development*, 8(2), 349-358. [Online] doi: https://doi.org/10.2166/washdev.2018.162

Ataoui, R. and Ermini, R. (2017) 'Risk assessment of water distribution service', *18th International Conference on Water Distribution System Analysis, WDSAI2016*, 186(2017), 514-521. [Online]doi: https://doi.org/10.1016/j.proeng.2017.03.264

Azad, A., Karami, H., Farzin, S., Mousavi, S. F. and Kisi, O. (2019) ‘Modelling river quality parameters using modified adaptive neuro fuzzy inference system’, *Water Science and Engineering*, 12(1), 45-54. [Online] doi: https://doi.org/10.1016/j.wse.2018.11.001

Baig, S.A., Lou, Z., Baig, M.A., Qasim, M., Shams, D.F, Mahmood, Q. and Xu, X. (2017) ‘Assessment of tap water quality and corrosion scales from the selected distribution systems in northern Pakistan’, *Environmental Monitoring Assessment*, 189(194). [Online] doi: https://doi.org/10.1007/s10661-017-5907-5

Basheer, I.A., and Hajmeer, M. (2000) ‘Artificial neural networks: fundamentals, computing, design and applications’, *Journal of Microbiological Methods*, 43(2000), 3-31. [Online] https://www.elsevier.com/ locate /jmicmeth

Bensoltane, M. A., Zeghadnia, L., Djemili, L., Gheid, A. and Djebbar, Y. (2018) ‘Enhancement of the free residual chlorine concentration at the ends of the water supply network: case study of Souk Ahras City-Algeria’, *Journal of Water and Land Development*, 38 (VII–IX), 3-9. [Online] doi: https://doi.org/ 10.2478/jwld-2018-0036

Bereskie, T., Manuel, J. R., and Sadiq, R., (2017) ‘Drinking water management and governance in Canada: an innovative plan-do-check-act (PDCA) framework for a safe drinking water supply’, *Environmental Management*, 60, 243-262. [Online] doi: https://doi.org/10.1007/s00267-017-0873-9

Blokker, M., Vreeburg, J. and Speight, V., (2014) 'Residual chlorine in the extremities of the drinking water distribution: the influence of stochastic water demands', *12th International Conference on Computing and Control for the Water Industry, CCW12013*, Procedia Engineering, 70 (2014), 172-180. [Online] doi: https://doi.org/10.1016/j.proeng.2014.02.020

Bowden, G.J., Nixon, J.B., Dandy, G.C., Maier, H.R. and Holmes, M, (2019) 'Forecasting chlorine residuals in a water distribution system using generalized neural network'. [Online] doi: htpps://doi.org/ 10.1.1.519.8859.pdf

Branz, A., Levine, M., Lehmann, L., Bastable, A., Ali, S. I., Kadir, K., Yates, T., Bloom, D. and Lantagne, D. (2017) 'Chlorination of drinking water in emergencies: a review of knowledge to develop recommendations for implementation and research needed', *Waterlines*, 36(1). [Online] doi: https://dx.doi.org/10.336/1756-333488.2017.002

Castro, P. and Neves, M, (2010) 'Chlorine decay in water distribution systems: case study of Lousada network', *Environmental 2010: Situation and Perspectives for the European Union*, 6-10 May 2003, Porto, Portugal

Chau, K.W., (2006) 'A review on integration of artificial intelligence into water quality modelling', *Marine Pollution Bulletin*, 52 (2006). 726-733. [Online] doi: https://10.1016/j.marpolbul.2006.04.003

Chirwa, E. and Madzivhandil, V, (2017) 'Modelling chlorine decay in water distribution systems using Aquasim', *Chemical Engineering Transactions (CET)*, 57. [Online] https:/www.aidiic.it/cet

Clark, R. M. (2011) 'Chlorine fate and transport in drinking water distribution systems: results from experimental and modelling studies', *Front. Earth Sci. 2011*, 5(4), 334-340. [Online] doi: https://doi.org/ 10.1007/s11707-011-0194-x

Cuesta, C.G.A., Tuhovcak, L. and Taus, M. (2014) 'Using artificial neural networks to assess water quality in water distribution networks', *12th International Conference on Computing and Control for the Water Industry*, CCWI2013, 70, 399-408. [Online] doi: https://doi.org/ 10.1016/j.proeng.2014.02.045

Dore, M. H. (2015) *Water policy in Canada: problems and possible solutions.* Switzerland: Springer International Publishing AG. [Online] doi: https://doi.org/ 10.1007/978-3-319-15883-9_211

Environmental Protection Agency, (2018) *Online water quality monitoring systems in distribution systems: for water quality surveillance and response systems*, MC 140, Cincinnati, USA: Office of Water.

Fattoruso, G., De Chiara, D., De Vito, S., La Ferrara, V., Di Francia, G., Leopardi, A., Cocozza, E. Viscusi, M. and Fontana, M. (2013) 'Simulation of chlorine decay in drinking water distribution systems: Case study of Santa Sofia network', (Southern Italy). *Proceedings of 17th National 467 Conference*, 5-7 February 2013, Brescia, Italy, 467-470. [Online] doi: https://doi.org/ 10.1007/978-3-319-00684-0_90

Fisher, I., Kastl, G., Clark, and Sathasivan, A. (2011) 'Evaluation of suitable chlorine bulk-decay models for water distribution systems', *Water Research*, 45(2011), 4896-4908. [Online] doi: https://doi.org/ 10.1016/j.watres.2011.06.032

Garcia-Avila, F., Anazco, A. A., Ordonez-J, J., Guanuchi-Quezada, C., Flores del Pino, L. and Ramos-Fernandez, L. (2021) 'Modeling of residual chlorine in a drinking water network in times of pandemic of the SARS-COV-2 (COVID-19)', *Sustainable Environmental Research*, (2021) 31 (12). [Online] doi: https://doi.org/ 10.1186/s42834-021-00084-w

Garcia-Avila, F., Sanchez-Alvarracin, C., Cadme-Galabav, M., Conchado-Martinez, J., Garcia_Mera, G. and Zhindon-Arevalo, C. (2020) 'Relationship between chlorine decay and temperature in the drinking water', *MethodsX7*, (2020), 101002. [Online] doi: https://doi.org/ 10.1016/j.mex.2020.101002

Geetha, S. and Gouthami, S, (2017) 'Internet of things enabled real time water quality monitoring system', *Smart Water*, 2(1), 1-19. [Online] doi: https://doi.org/ 10.1186//s40713-0005-y

Geogerscu, A. M. and Geogerscu, S.C. (2012) 'Chlorine concentration decay in the water distribution system of a town with 50,000 inhabitants', *U.P.B. Scientific Bulletin, Series D: Mechanical Engineering*, 74(1). [Online] https://www.researchgate.net/publication/235800357

Gibbs, M. et al. (2019) 'Use of artificial neural networks for modelling chlorine residuals in water distribution networks. *United Water International Pty Ltd*, Adelaide-Australia. [Online] https://www.mssanz.org.au

Gitu, I. and Egbe, J. (2016) 'Residual chlorine in water distribution network', *International Journal of Scientific Research and Engineering Studies*, 3 (3). [Online] https://www.ijsres.com

Gorzalski, A. S., Harrington, G.W. and Coronell, O. (2019) ‘Assessing flow segregation and mixing by modelling residual disinfectant conversion’, *AWWA Water Science*, 1-18. [Online] doi: https://doi.org/10.1002/aws2.1154

Goyal, R. P. and Patel, H, M. (2014) ‘Analysis of residual chlorine in simple drinking water distribution system with intermittent water supply’, *Applied Water Science*, 5, 311-319. [Online] doi: https://doi.org/10.1007/s13201-014-0193-7

Government of Sudan, (2017) *Protocols for the chlorination of drinking water for small to medium sized supplies,* Khartoum: Federal Ministry of Health and Ministry of Water Resources, Irrigation and Electricity. [Online] https://www.humanitarianresponse.info

Hallam, N. B, West, J. R., Forster, C. F., Powell, J.C. and Spencer, I. (2002) ‘The decay of chlorine associated with the pipe wall in water distribution systems’, *Water Research*, 36(2002), 3479-3488. [Online] https:// www.elsevier.com/locate/watres

Hassan, G., AL-Moameri, H. H. and Ismail, A. H. (2019) ‘Chlorine concentration-time (C-T) plot for Echhornia crassipes in water’, 2nd *International Conference on Sustainable Engineering Techniques, (ICSET 2019)*’, IOP Conference Series: Materials Science and Engineering, 518 (2019)062017. [Online] doi: https://doi.org/10.1088//1757-899X/518/6/062017

Hyunjun, K., and Sanghyun, KP. (2017) ‘Evaluation of chlorine decay models under transient conditions in a water distribution system’, *Current Science*, 3(8), 522-537. [Online] doi: https://doi.org/10.2166/hydro.2017.082

Jamwal, P. and Kumar, M.S.M. (2016) 'Effect of flow velocity on chlorine decay in a water distribution network: a pilot study', *Journal of Hydroinformatics*, 19(4), 1349-1354. [Online] doi: https://doi.org/10.18520/cs/v111/i8/1349-1354

Jollife, I.T. and Cadima, J. (2016) 'Principal component analysis: a review and recent developments', *Phil. Trans. R. Soc.*, A374, 1-16. [Online] doi: https://doi.org/10.1098/rsta.2015.0202

Jones, K, (2014) 'An introduction to statistical modelling', *Chemical Engineering Transactions (CET*), 241-255. [Online] https:/www. researchgate.net/publication/256801601

Karadirek, I., Kara, S., Muhammetoglu, A., Muhammetoglu, H. and Soyupak, S. (2015) 'Management of chlorine dosing rates in urban water distribution networks using online continuous monitoring and modelling', *Urban Water Journal*, 19(4), 1-15. [Online] doi: htpps://doi/10.1080/1573062X.2014.992916

Karikari, A. Y. and Ampofo, J, A. (2013) 'Chlorine treatment effectiveness and physico-chemical and bacteriological characteristics of treated water supplies in distribution network of Acrra-Tema metropolis- Ghana', *Applied Water Science*, 3, 535-543. [Online] doi: https://doi.org/10.1007/s13201-013-0101-6

Kim, S., Kim, H., and Koo, J. (2014) 'Prediction of chlorine in various hydraulic conditions from a pilot-scale water distribution system', *12th International Conference on Computing and Control for the Water Industry*, 70 (2014), 934-942. [Online] doi: https://doi.org/10.1016/j.proeng.2014.02.104

Kim, J. D., Yoo, D.G., Lee, S.M., Lee, H.M. and Choi, Y.H. (2018) ‘Optimizing re-chlorination injection points for water supply networks using harmony search algorithm’, *Water 2018*, 10(547). [Online] doi: https://doi.org/10.3390/w10040547

Kim, S. (2019) ‘Valve manoeuvre prediction in simple and complicated pipe systems.’, *Water Resources Management*, 33, 4671-4685. [Online] doi: https://doi.org/10.1007/s11269-019-02262-3

Kulkami, V., Awad, J., Medlock, A., Monis, P., Lau, M., Drigo, B. and Leeuwen, J.V. (2018) ‘Field based pilot-scale drinking water distribution system: simulation of long hydraulic retention times and microbiological mediated monochloramine decay’, *Journal List MethodsX*, 5, 684-696. [Online] doi: https://doi.org/10.1016/j.mex.2018.06.015

Kumar, A., Kumar, K, Bharanidharan, B., Neha, M., Eshita, D., Singh, M., Thakur, V., Sharma, S. and Malhotra, N. (2015) ‘Design of water distribution system using EPANET’, *International Journal of Advanced Research*, 3(9), 789-812. [Online] https://www.journalijar.com

Lee, H., Shin, G., S., Hong, S., Choi, J. and Chun, M. (2016) ‘Post-chlorination process control based on flow prediction by time series neural network in water treatment plant.’, *International Journal of Fuzzy Logic and Intelligent Systems*, 16(3), 197-207. [Online] doi: https://dx.doi.org// 10.5391/IJFIS.2016.16.3.197

Loucks, D.P. and van Beek, E. (2017) *Water resource systems management and planning: an introduction to methods, models and applications*. 2nd edn. Cham, Switzerland: Springer International Publishing AG. [Online] doi: https://doi.org/10.1007/978-3-319-44234-1

Louppe, G. (2014) *Understanding random forests-from theory to practice,* PhD Thesis, Department of Electrical Engineering & Computer Science, Faculty of Applied Sciences, University of Liège. [Online] https//arxiv.org

Madzivhandila, V. A. and Chirwa, E.M.N. (2017) 'Modelling chlorine decay in drinking water distribution systems using aquasim', *Chemical Engineering Transactions,* 57. [Online] https://www.aidic.it/cet

Mahendrarajah, R. (2014) 'Chlorine demand analysis in distribution systems using hydraulic models and laboratory tests', *77th Annual WIOA Victorian Water Industry Operations Conference and Exhibitions*, Bendigo Exhibition Centre, 2-4 September 2014.

Melkumova, L.E. and Shatskikh, S.Y. (2017) 'Comparing rigde and lasso estimators for data analysis', *3rd International Conference on Information Technology and Nanotechnology*, Samara, Russia, April 2017. 25-27.

Mao, Q., Feng, J., Wang, W., Wang, Q., Hu, Z. and Yuan, S. (2016) 'Chlorination of parabens: reaction kinetics and transformation product identification', *Environmental Science Pollution Research*, 23, 23081-23091. [Online] doi: https://doi.org/10.1007/s11356-016-7499-y

Mentes, A., Galiatsatou, P., Spyrou, D., Samara, A. and Stournara, P. (2020) 'Hydraulic simulation and analysis of an urban centre's aqueducts using scenario analysis for network operations: the case of Thessaloniki City in Greece', *Water* ***2020*** *MDPI*, 12, 1627. [Online] doi: https://doi.org/10.3390/w12061627

Moghaddam, A. Afsharnia, M. and Minaee, R.P. (2020) 'Preparing the optimal emergency response protocols by MOPSO for a real-world water distribution network', *Environmental Science and Pollution Research*, 27, 30625-30637. [Online] doi: https://doi.org/10.1007/s11356-020-08892-0

Monteiro, L., Figueiredo, D., Dias, S., Freitas, R., Covas, D. Menaia, J. and Coelho, P. (2014) 'Modelling of chlorine decay in drinking water supply systems using EPANET MSX', *12th International Conference on Computing and Control for the Water Industry, CCWI2013*, 70, 1192-1200. [Online] https://www.sciencedirect.com

Monteiro, L., Figueiredo, D., Covas, D. and Menaia, J. (2017) 'Integrating water temperature in chlorine decay modelling: a case study', *Urban Water Journal*, 19(4), 1-5. [Online] doi: https://dx.doi.org/10.1080/1573062X.2017.1363249

Mounaouer, B. and Abdennaceur, H. (2016) 'Modelling and kinetic characterization of wastewater disinfection using chlorine and UV irradiation', *Environmental Science Pollution Research*, 23, 19861-19875, [Online] doi: https://doi.org/10.1007/s11356-016-7173-4

Musz, A., Kowalska, B. and Widomski, K. M. (2009) 'Some issues concerning the problems of water quality modelling in distribution systems', *Ecological Chemistry and Engineering Systems*, 16(S2), 176-184.

Nejjari, V., Puig, V., Perez, Z., Quevedo, J., Cuguero, M.A., Sanz, G. and Mirats, J. M. (2014) 'Chlorine decay model calibration and comparison: application to real water network', *12th International Conference on Computing and Control for the Water Industry, CCW12013*, Procedia Engineering, 70 (2014), 1221-1230. [Online] doi: https://doi.org/10.1016/j.proeng.2014.02.135

Powell, J. C., Hallam, N. B., West, J. R., Forster, C. F. and Simms, J. (2000) 'Factors which control bulk chlorine decay rates', *Water Research*, 34(1); 117-126. [Online] https:// www.elsevier.com/locate/watres

Niu, Z. Hu, X., Zhang, Y. and Sun, Y. (2016) 'Effect of chlorine dosage in prechlorination on trihalomethanes and haloacetic acids during water treatment process', *Environmental Science Pollution Research*, 24, 5068-5077. [Online] doi: https://doi.org/10.1007/s11356-016-8265-x

Nono, D., Odirile, P.T, Basupi, I. and Parida, P.T. (2019) 'Assessment of probable causes of chlorine decay in water distribution systems of Gaborone city-Botswana', *Water SA,* 45(2), 190-198. [Online] doi: https://doi.org/10.4314/wsa.v45i2.05

Nouri, I. (2017) 'Optimal design and management of chlorination in drinking water networks: a multi-objective approach using genetic algorithms and the Pareto optimality concept', *Applied Water Science*, 7, 3527-3538. [Online] doi: https://doi.org/10.1007/s13201-017-0620-7

Oladipupo, B., Adnan, M.A, Hamam, Y., Page, P.R., Adedeji, K.B. and Piller, O. (2019) 'Solving management problems in water distribution networks: a survey of approaches and mathematical models,' *Water **2019** MDPI*, 11, 562. [Online] doi: https://doi.org/10.3390/w11030562

Ozdemir, O.N. and Ucak, A. (2002) 'Simulation of chlorine decay in drinking water distribution systems,' *Journal of Environmental Engineering*, 128, 31-39. [Online] doi: https://doi.org/10.1061/(ASCE) 0733-9372 (2002) 128: 1 (31)

Rajasingham, A., Harvey, B., Taye, Y., Kamwaga, S., Martinsen, A., Sirad, M., Aden, M., Gallagher, K., and Thomas, H. (2020) 'Improved chlorination and rapid water quality assessment in response to an outbreak of acute watery diarrhea in Somali region, Ethiopia,' *Journal of Water, Sanitation, Hygiene and Development*, 10(3), 596-602. [Online] doi: https://doi.org/10.2166/washdev.2020.146

Ramos, H. M., Loureiro, D., Lopes, A., Fernandes, C., Covas, D., Reis, L.F. and Cunha, M.C. (2010) 'Evaluation of chlorine decay in drinking water systems for different flow conditions: from theory to practice,' *Water Resources Management*, 24, 815-835. [Online] doi: https://doi.org/10.1007/s11269-009-9472-8

Ricca, H., Aravinthan, V. and Mahinthakumar, G. (2019) 'Modelling chloramine decay in full-scale drinking water supply systems,' *Water Environment Federation*, 441–454. [Online] doi: https://doi.org/10.1002/wer.1046

Rossman, L.A., Clark, R.M. and Walter, M. G. (1994) 'Modelling chlorine residuals in drinking-water distribution systems,' *Journal of Environmental Engineering*, 120 (4), 803-820.

Shamsaei, H., Jaafar, O. and Basri, N.E.A. (2013) 'Effects of residence time on water quality in large water distribution systems, *Engineering Scientific Research*, 5, 449-457. [Online] doi: https://dx.doi.org/10.4236/eng.2013.54054

Sharif, A. T, Farahat, A., Haider, H., Al-Zahrrani, M.A., Rodriguez, M.J. and Sadiq, R. (2017) 'Risk-based framework for optimizing residual chlorine in large water distribution systems,' *Environmental Monitoring Assessment*, 189(307). [Online] doi: https://doi.org/10.1007/s10661-017-5989-0

Soyupak, S., Kilic, H., Karadirek, I. E., and Muhammetoglu, H. (2011) 'On the usage of artificial neural networks in chlorine control applications for water distribution networks with high quality water,' *Journal of Water Supply, Research and Technology-AQUA*, 60(1), 51-61. [Online] doi: https://doi.org/10.2166/aqua.2011.086

Stoianov, I. and Aisopou, A. (2014) 'Chlorine decay under steady and unsteady state hydraulic conditions,' *12th International Conference on Computing and Control for the Water Industry, CCW12013*, Procedia Engineering, 70 (2014), 1592-1601. [Online] doi: https://doi.org/10.1016/j.proeng.2014.02.176

Tibrishirani, R. (2013) 'Modern regression 2: The Lasso,' *Data Mining*, 36-462/36-662, March 21 2013

Tiruneh, A. T, Debessai, T.Y., Bwembya, G.C., Nkambule, S.J. and Zwane, L. (2019a) 'Variable chlorine decay rate modelling of the Matsapha town water network using EPANET program,' *Journal of Water Resource and Protection*, 11, 37-52. [Online] doi: https://doi.org/10.4236/jwarp.2019.111003

Tiruneh, A. T, Debessai, T.Y., Bwembya, G.C. and Nkambule, S.J. (2019b) 'A mathematical model for variable chlorine decay rates in water distribution systems,' *Hindawi Modelling and Simulation in Engineering*, 2019. [Online] doi: https://doi.org/10.1155/2019/5863905

Tiwari, S, Babbar, R., and Kaur, G. (2018) 'Performance evaluation of two ANFIS models for predicting water quality index of River Satluj (India),' *Hindawi Advances in Civil in Engineering*, 2018. [Online] doi: https://doi.org/10.1155/2018/8971079

Torretta, V., Tolkou, A.K., Katsoyoyiannis, I.A., Katsoyoyiannis, A., Trulli, E., Magaril, E. and Rada, E.C, (2019) 'Consumption of free chlorine in an aqueduct with low protection: case study of the new aqueduct Simbrivio-Castelli (NASC), Italy,' *Water 2018 MDPI*, 10(127). [Online] doi: https://doi.org/10.3390/w10020127 www.mdpi.com/journal/water

US Environmental Protection Agency, (2020) 'Application for modelling drinking water distribution systems'. [Online] https:/www.epa.gov

Vargas, T. F, Baia, C.C., Machado, T.L., Dorea, C.C. and Bastos, W.R. (2021) 'Decay of free residual chlorine in wells water of northern Brazil,' *Water 2021*, 13(7), 992. [Online] doi: htpp:// doi:10.3390/w13070992

Vuta, L. and Dumitran, G.E. (2019) ‘Some aspects regarding chlorine decay in water distribution networks,’ *Aerul si Compaonente ale Mediului*, 253-259. [Online] htpps://www.researchgate.net>2284…

Water Research Australia Limited, (2015) *Good practice guide to the operation of drinking water supply systems for the management of microbial risk*, final report project 1074. Adelaide-Australia: World Health Organization Regional Office for South-East Asia. [Online] https:// www.waterra.com.au.

Wijesinghe, L., Ilangangedara, D. and Gunarathne, L.H.P. (2019) ‘Sustainable rural water supply schemes and community-based organizations,’ *Indian Journal of Public Administration*, 65(3), 702-717. [Online] doi: https://doi.org/10.1177/0019556119840924

World Health Organization, (2014) *Water safety in distribution systems*. [Online] https://www.who.int/water_sanitation_health/publications/Water_Safety_in_Distribution_System/en

World Health Organization, (2017) *Principles and practices of drinking-water chlorination: a guide to strengthening chlorination practices in small-to medium sized water supplies*. [Online] https://creativecommons.org/licenses/by-nc-sa/3.0/igo).

Wu, H. and Dorea, C.C. (2020) ‘Towards a predictive model for initial chlorine dose in humanitarian emergencies’ *Water 2020*, 12, 1506. [Online] doi: https://doi.org/10.3390/w12051506

Yanga, Y. J., Goodrich, J. A., Clark, R. M. and Sylvana, Y. L. (2006) 'Modelling and testing of reactive contaminant transport in drinking water pipes: chlorine response and implications for online contaminant detection,' *Water Research*, 42(2008); 1397-1412. [Online] doi: https://doi.org/10.1016/j.watres.2007.10.009

Yi, J.C., Lee, J., Jung, H., Park, P.K. and Noh, S.H. (2018) 'Reduction of bacterial regrowth in treated water by minimizing water stagnation in the filtrate line of a gravity-driven membrane system,' *Environmental Engineering Research*, 24(1), 17-23. [Online] doi: https://doi.org/10.4491/eer.2018.048

Zhang, C., Li, C., Zheng, X., Zhao, J., He, G. and Zhang, T. (2016) 'Effect of pipe materials on chlorine decay, trihalomethanes formation and bacterial communities in pilot-scale water distribution systems,' *International Journal of Environmental Science Technology*, 14, 84-94. [Online] doi: https://doi.org/10.1007/s13762-016-1104-2

Zhou, L., Zhang, Y. and Zheng, G. (2013) 'Monochloramine decay for two chlorine to ammonia ratios in bulk water', *Water Environment Federation*, 85(11), 2194–2200. [Online] doi: https://doi.org/10.2175/106143013X13736496909509

Zlatanovic, L., Moerman, A., van der Hoek, J.P., Vreeburg, J. and Blokker, M. (2017) 'Development and validation of drinking water temperature model in domestic drinking water supply systems,' *Urban Water Journal*, 14(10), 1031–1037. [Online] doi: https://doi:10.1080/1573062X.2017.132550

APPENDICES

Appendix I: Letter A.1: Request by Kyambogo University to NWSC for permission to study Lirima gravity scheme

KYAMBOGO UNIVERSITY

Department of Civil and Building Engineering

P. O. BOX 1, KYAMBOGO – P. O. BOX 7181 KAMPALA, UGANDA

Website: www.kyu.ac.ug, Email: civil@kyu.ac.ug

TEL: +256-41-4287340, FAX: +256-41-4289056/4222643

November 11, 2020

Manager – Research and Development,

International Resource Center - National Water and Sewerage Corporation,

Old Port Bell Road, Kampala, Uganda.

Dear Eng./Dr./Sir,

RE: REQUEST TO CONDUCT RESEARCH ON MODELLING CHLORINE DECAY IN WATER DISTRIBUTION USING LIRIMA GRAVITY WATER FLOW SCHEME AS A CASE STUDY

Mr. Julius Caesar Kwio-Tamale is a student of Kyambogo University undertaking Master of Science in Water and Sanitation Engineering at the Department of Civil and Building Engineering. He is conducting a research study on "Comparison of Performance of Models in Predicting Space-Time Decay of Chlorine in Gravity Water Flow Systems: Case Study of Lirima Gravity Water Flow Scheme". The researcher is being supervised and co-supervised by Dr. Charles Onyutha and Eng. Dr. Anne Nakagiri, respectively.

According to the World Health Organization report of 2014, the minimum and maximum residual chlorine concentrations in drinking water are recommended to be 0.2 mg/l and 5 mg/l, respectively. Ensuring the recommended range of chlorine concentration at all time in drinking water at the water draw off points is normally a difficult task. Either there is over-dosage or under-dosage of the chlorine in drinking water. Chlorine over-dosage leads to production of disinfection by-products that may cause cancer, reproductive disorders, liver and kidney damage, birth defects, and miscarriage. Over-dosage of chlorine also makes the water corrosive to pipes and objectionable in terms of taste and color. On the other hand, under-dosage of chlorine promotes re-growth of microbials thereby leading to infections of consumers. Attempts to keep chlorine concentrations in drinking water within the recommended range of 0.2-5 mg/l tends to be possible through modelling and predictions.

The main problem with modelling chlorine lies in the complexity of capturing the dynamics of chlorine decay amidst several practical factors such as, temperature, pH, and turbidity. This research attempts to partially solve this problem by modelling the decay of chlorine in both space and time. To do so, this research requires large amount of data (such as chlorine dosage, free chlorine, water temperature, and turbidity) to be measured or tested based on samples from a reasonable number of locations from a water distribution network. This research has a number of specific objectives (see next page) and will yield information that can be used to improve knowledge regarding chlorine dosage amidst practical field conditions to ensure the water quality meets the recommended World Health Organization concentration limits 0.2-0.5 mg/l of chlorine.

The purpose of this letter is to request your office to allow Julius Caesar Kwio-Tamale (with the help of the support from your team in the National Water and Sewerage Corporation branch where the study area is located) to (i) identify suitable locations for sampling, (ii) take samples, and (iii) test the samples in your laboratory, (iv) avail the student any other necessary information or services that will be required for the research. In the same line, this request is also to allow you ensure the student conducts laboratory tests at subsidized prices since there will be many samples.

In line with ethical considerations and courtesy to your permission, Kyambogo University will share and discuss results of the study with National Water and Sewerage Corporation before dissemination.

I shall be grateful for any assistance rendered to Mr. Julius Caesar Kwio-Tamale to allow him conduct his research study timely.

Yours Sincerely,

Dr. Lawrence Muhwezi,

Head of Department of Civil and Building Engineering

Cc. Dean, School of Graduate Studies, Kyambogo University

Dr. Charles Onyutha - Department of Civil and Building Engineering, Kyambogo University

Eng. Dr. Anne Nakagiri - Department of Civil and Building Engineering, Kyambogo University

The specific objectives of this study are;

a) to characterize gravity water distribution parameters in relation to chlorine decay,
b) to assess space-time decay of chlorine in gravity flow distribution systems,
c) to compare performance of various models that predict residual chlorine concentration, and
d) to identify the appropriate model(s) for residual chlorine decay in gravity water flow system.

Appendix II: Letter A.2: Permission letter of NWSC for research on Lirima gravity flow scheme

NATIONAL WATER & SEWERAGE CORPORATION

HEAD OFFICE

TELEGRAMS WATERS KAMPALA
Telephone : 256 - 313 - 315000/100
Twitter : @nwscug
Facebook : www.facebook.com/waterug
Email : info@nwsc.co.ug

P.O. Box 7053
Plot 3 Nakasero
KAMPALA

24th November 2020

Our Ref: BSS/R&D/20-11

Mr. Julius Caesar Kwio Tamale
Student, Master of Science in Water and Sanitation Engineering
Kyambogo University.

Re: REQUEST TO CONDUCT RESEARCH ON MODELLING CHLORINE DECAY IN WATER DISTRIBUTION USING LIRIMA GRAVITY WATER FLOW SCHEME AS A CASE STUDY

Reference is made to your supervisors' request to carry out research at National Water and Sewerage Corporation's water treatment and distribution systems in Tororo Area; under the topic: ***"Comparison of Performance of Models in Predicting Space-Time Decay of Chlorine in Gravity Water Flow Systems: Case Study of Lirima Gravity Water Flow Scheme".*** It is indicated in the request that you need to have access to the Lirima water supply system and carry out sampling and laboratory analysis for free residual chlorine, temperature and turbidity; besides obtaining any other related information with the help of the technical staff in the Area.

Permission is hereby granted, valid until 28th February 2021.

By copy of this letter, the Area Manager, Tororo is requested to provide the student with access to the water supply systems and laboratory, and the information necessary for successful running of this research study. This is based on condition that the researcher provides all the necessary field and laboratory reagents and consumables. It is expected that the findings from this research will be made available to NWSC and key recommendations beneficial to the corporation will be adopted to improve our operations. In this regard, I recommend that upon completion of the data collection (and analysis), a presentation of the report of findings is made, details of which will be organised by the Senior Manager, WQ Management, based at the Central Laboratory, Bugolobi-Kampala. You are therefore urged to get in touch with her at the appropriate timing.

I wish you all the best as you carry out this research study.

Christopher Kanyesigye
Manager, Research and Development
c.c. =Senior Manager, WQ Management, NWSC HQ; =Area Manager, NWSC Tororo

"NW&SC - Water is Good Water for Public Health"

Appendx III: Questionnaire for Data on NWSC Lirima gravity flow scheme
Date: ……………………..… **Respondent:** ……………………………………………….

1. Background

This research aims to model residual chlorine decay in water from treatment plant to water consumption points. To do this, key water quality and water infrastructure parameters are to be collected. Other operational variables will also be collected.

2. Chlorine dosage at Lirima treatment plant

What is the initial chlorine dosage at the treatment plant after chlorination?

………………………………………………………………………………. (mg/l)

How frequent is chlorine dosage at the treatment plant? *Please state your response below*: …………………………………………………………………………………

3. Secondary chlorination:

Is secondary chlorination done in the transmission and distribution network? *Please tick the applicable box*. Yes No.

4. Network water quality monitoring

(a) At what points is water quality monitoring done in the transmission and distribution network? *Please state your response below*.

…………………………………………………………………………………………

(b) How frequently is water quality monitoring done in the transmission and distribution network? *Please state your response below*.

…………………………………………………………………………………………

5. Nature of water supply

(a) What is the nature of water supply? *Please tick the applicable box*.

Continously Intermittently

(b) If water supply is intermittent, for how many hours is it supplied on a given day? *Please state your response below*.

…………………………………………………………………………………………

END OF QUESTIONNAIRE

THANK YOU FOR YOUR RESPONSE

Appendix IV: Table A.1: Water quality parameter test values at Lirima gravity water treatment plant

Treatment Plant

Research Collection Data

SN	DAY & DATE	TIME OF SAMPLE	RESIDUAL	ELECTRICAL CONDUCTIVITY	PH	TURGIDITY (NTU)	TEMPERATURE	COMMENTS (IF ANY)
	Tuesday 9th/02/2021	9:00am	1.0	120.9	7.70	0.99	22.1	
		10:00am	0.90	95.4	7.66	0.81	23.0	
		11:00am	0.85	95.7	7.55	0.80	23.5	
		12:00am	0.83	132.4	7.52	0.78	23.7	
		2:00am	0.80	141.0	7.50	0.64	20.1	
		3:00am	0.75	131.0	7.43	0.42	21.7	
		4:00am	0.75	130.5	7.42	0.40	22.5	
		5:00am	0.70	132.0	7.42	0.40	22.0	
	Wednesday 10TH /02/2021	9:00AM	0.95	139.6	7.80	0.87	18.70	
		10:00AM	0.93	139.0	7.71	0.81	20.8	
		11:00AM	0.93	141.0	7.66	0.78	22.5	
		12:00AM	0.90	151.0	7.60	0.70	23.5	
		2:00AM	0.90	130.5	7.60	0.66	22.1	
		3:00AM	0.85	132.0	7.60	0.64	23.0	
		4:00AM	0.80	141.0	7.55	0.50	23.5	
		5:00AM	0.75	141.5	7.50	0.40	23.7	
	Thursday 11TH /02/2021	9:00AM	1.0	121.0	7.75	0.93	20.8	
		10:00AM	0.95	130.0	7.70	0.91	21.5	
		11:00AM	0.93	131.2	7.70	0.90	22.5	
		12:00AM	0.85	132.0	7.68	0.82	23.0	
		2:00PM	0.80	141.0	7.65	0.80	20.1	
		3:00PM	0.82	141.5	7.48	0.72	21.7	
		4:00PM	0.75	151.0	7.40	0.68	22.5	
		5:00PM	0.73	151.2	7.41	0.68	23.0	

NB. (1) Type of equipment used might also be important to be captured

PH, EC & Temperature used a entrench elite PCTS achieve

Turbidity – I used a lovibond turbid meter

Residual chlorine – a PCTS comparator.

Appendix V: Table A.2: GPS co-ordinates of water transmission lines of Lirima Gravity Scheme

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Lirima WTP	36 N	0657122	0098196	1812	HDPE	250
2	BPT1	36 N	0656161	0097372	1757	HDPE	250
3	BPT2	36 N	0656025	0097335	1722	HDPE	250
4	BPT3		065538	0097085	1661	HDPE	250
5	GV//T-Junction to Bumbo TC	36 N	0654761	0096706	1618	HDPE	250
6	AV	36 N	0654747	0096701	1614	HDPE	250
7	GV at Magale TC	36 N	0652130	0095728	1571	HDPE	250
8	Bend at Magale	36 N	0652002	0095850	1559.5	HDPE	250
9	Exposed Pipe	36 N	0651842	0095303	1532	HDPE	250
10	T-Junction at Magale	36 N	0651605	0095025	1525	HDPE	250
10	BPT4	36 N	0651300	0094852	1521	HDPE	250
11	AV at Nalukwade	36 N	0650304	0094755	1449.5	HDPE	250
12	Exposed Pipe @ R. Natsekhe	36 N	0649612	0094865	1425	HDPE	250
13	AV at Buwanzaala	36 N	0649612	0094862	1434	HDPE	250
14	AV at Maala	36 N	0648664	0094287	1453	HDPE	250
15	SV at Mufutu	36 N	0647909	0094073	1435	HDPE	250
17	Musiye T-Junction	36 N	06477907	0094069	1438	HDPE	250/ 100
18	BPT5	36 N	0647589	0098151	1417	HDPE	250
18	Musiye Wash-out	36 N	0647787	0093089	1443	HDPE	100
19	AV at Musiye	36 N	0647726	0092798	1465	HDPE	100
20	90 turn to Musiye Tank	36 N	0637725	0092786	1463	HDPE	100
21	Musiye Reservoir	36 N	0647596	0092792	1468	HDPE	100
18	BPT5	36 N	0647589	0098151	1417	HDPE	250
22	Bufumo	36 N	0649032	0092351	1240	HDPE	80
23	Buwanyera	36 N	0647532	0092251	1241	HDPE	80
24	Busike	36 N	0644730	0091989	1302	HDPE	80
25	Butiru T-Junction		0643987	0090896	1339	HDPE	80
26	Butiru Reservoir	36 N	0644672	0090204	1354	HDPE	250

Appendix VI: Table A.3: GPS co-ordinates of water distribution lines of Lirima Gravity Scheme

(a) Mufutu-Nalukwade trading centre water distribution zone

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Mufutu ClearWater Tank Outlet	36 N	0657047	0098134	1442.69	HDPE	100
2	Wash out	36 N	0647782	0093081	1448	HDPE	100
3	Mufutu village	36 N	0647939	0094061	1440	HDPE	100
4	Maala village	36 N	0648893	0094381	1447	HDPE	100
5	Nalukwade village	36 N	0649591	0094872	1432	HDPE	100
6	Nalukwade trading centre	36 N	0650255	0094766	1457	HDPE	100

(b) Butiru-Makenye water distribution zone

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Butiru Water Tank Outlet	36 N	0644672	0090204	1354	HDPE	80
2	T-Junction to Makenye	36 N	0643987	0090896	1339	HDPE	80
3	At Clinic	36 N	0644031	0090915	1260	HDPE	80
4	Buwasike village:	36 N	0644730	0091989	1302	HDPE	80
5	Buwanyera village:	36 N	00647532	0092251	1241	HDPE	80
6	Wash out	36 N	0646051	0093782	1276	HDPE	80
7	Bumafumo village	36 N	00649032	0092351	1240	HDPE	80

(c) Butiru-Vermiculite water distribution zone

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Butiru Police Station	36 N	0643763	0091147	1262	HDPE	80
2	Bunangabo Cell	36 N	0642610	0092246	1261	HDPE	80
3	Buwayo Cell	36 N	0643119	0091879	1262	HDPE	65
4	Bamukesi village	36 N	0642609	0092229	1257	HDPE	65
5	Bamulatte village	36 N	0642609	0092229	1224	HDPE	50

(d) Musiye-Bubuto water distribution zone

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Mufutu ClearWater Tank Outlet	36 N	0657047	0098134	1442.69	HDPE	100
2	S1	36 N	0647830	0093135	1451	HDPE	100
3	SV & T-J	36 N	0647734	0092776	1471	HDPE	100
4	S3	36 N	0647734	0092757	1470	HDPE	65
5	S4	36 N	0647734	0092740	1470	HDPE	65
6	S5	36 N	0647729	0092709	1471	HDPE	65
7	S6	36 N	0647725	0092673	1470	HDPE	65
8	S7	36 N	0647685	0092571	1469	HDPE	50
9	S8	36 N	0647541	0092260	1459	HDPE	50
10	S9	36 N	0647532	0092251	1451	HDPE	50
11	S10	36 N	0647532	0092251	1474	HDPE	50

Table A.4: GPS co-ordinates of water yard taps on Lirima Gravity Water Scheme

(a) Mufutu-Nalukwade trading centre water distribution zone

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Namawanga	36 N				Cast Iron	15
2	Mufutu	36 N	0647939	0094061		Cast Iron	15
3	Maala	36 N				Cast Iron	15
4	Natsekhe wash-out	36 N				Cast Iron	15
5	Nalujwade	36 N	0649597	0094876	1466	Cast Iron	15

(b) Butiru-Makenye water distribution zone

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Butiru Clinic	36 N	0644031	0090915	1270	Cast Iron	15
2	Buwasike Village:	36 N	0647532	0092251	1252	Cast Iron	15
3	Buwanyera Village:	36 N	0647532	0092251	1242	Cast Iron	15
4	Bufumo Village:	36 N				Cast Iron	15

(c) Butiru-Vermiculite water distribution zone

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Butiru Police Station	36 N	0643705	0091118	1291	Cast Iron	15
2	Buwayo Cell	36 N	0643198	0091951	1259	Cast Iron	15
3	Bunangabo Cell	36 N	0642610	0092246	1232	Cast Iron	15
4	Bamukesi Village	36 N	0642609	0092229	1265	Cast Iron	15
5	Bamulatte Village	36 N	0642609	0092229	1226	Cast Iron	15
6	Bunabiro Village	36 N	0642609	0092229	1214	Cast Iron	15

(d) Musiye-Bubuto water distribution zone

Item	Description	Latitude	Northing (Y)	Easting (UTM) (X)	Elevation (Z)	Pipe material	DN
1	Musiye caretaker's home	36 N	0647830	0093135	1452	Cast Iron	15
2	S3	36 N	0647734	0092757	1471	Cast Iron	15
3	S4 residence of LC3 Chairperson Elect-Buwambo SC	36 N	0647734	0092740	1470	Cast Iron	15
4	S5	36 N	0647742	0092705	1470	Cast Iron	15
5	S6	36 N	0647734	0092670	1470	Cast Iron	15
6	S7 (Musiye P/Sch)	36 N	0647658	0092594	1473	Cast Iron	15
7	S8	36 N	0647532	0092251	1466	Cast Iron	15
8	S9	36 N	0647532	0092251	1451	Cast Iron	15
9	S10	36 N	0647532	0092251	1466	Cast Iron	15
10	S11	36 N	0647532	0092251	1475	Cast Iron	15

Appendix VII: Table A.5: GPS co-ordinates of water tanks on Lirima Gravity Water Scheme

Item	Tank type	Location	Size		GPS Co-ordinates		
			Diameter (m)	Height (m)	Northing (Y)	Easting (UTM) (X)	Elevation (Z)
1	WTP	Lirima	3.00	3.00			
2	BPT 1	Lirima	3.00	3.00	0656161	0097372	1757
3	BPT 2	Lirima	3.00	3.00	0656025	0097335	1722
4	BPT 3	Bumbo	3.00	3.00	0644041	0099917	1661
5	BPT 4	Magale	6.30	3.00	0651300	0094752	1521
6	BPT 5	Musiye	6.30	3.00	0647589	0093751	1417
7	BPT 6	Butiru	10.50		0644672	0090204	1354

NB: WTP denotes Lirima Water Treatment Plant

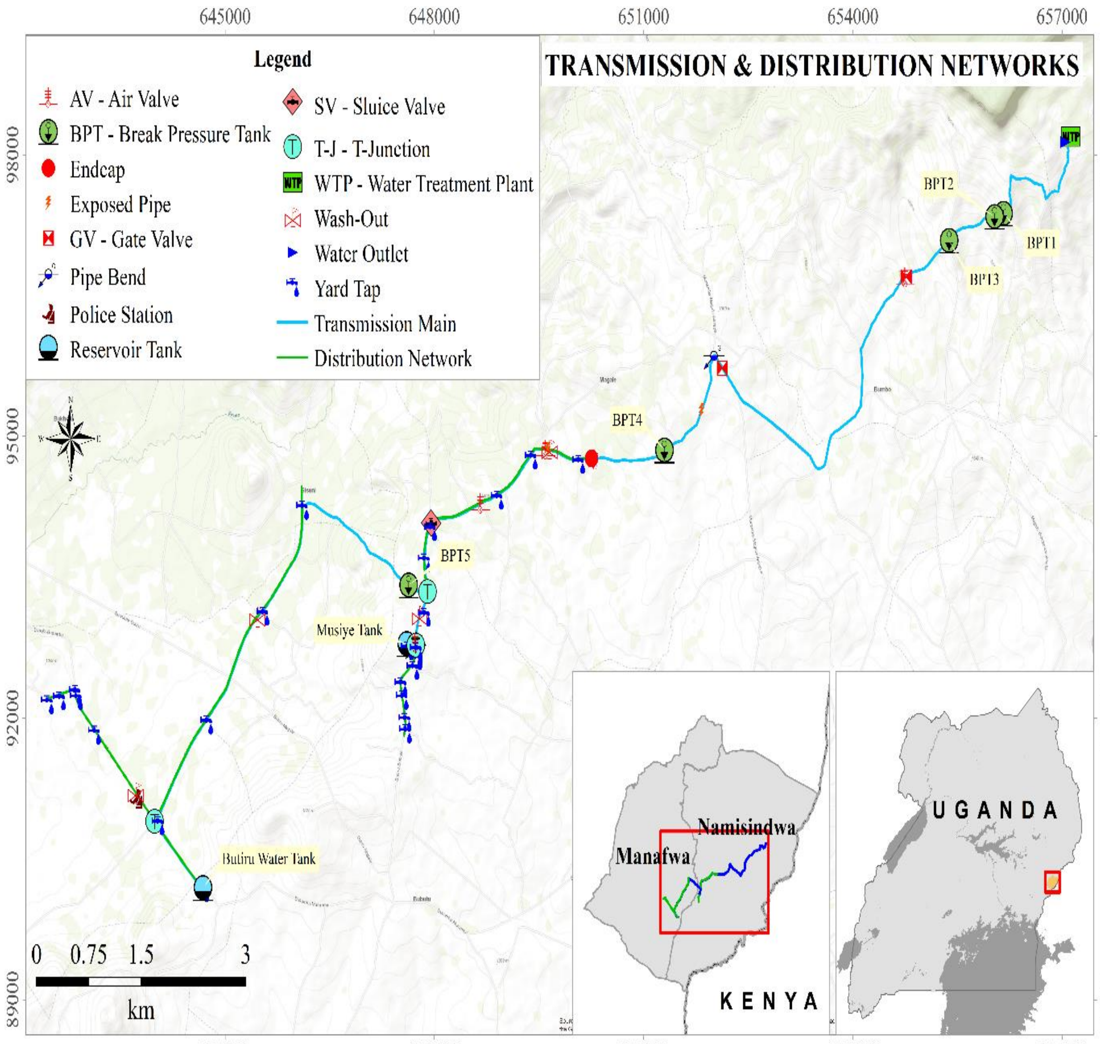
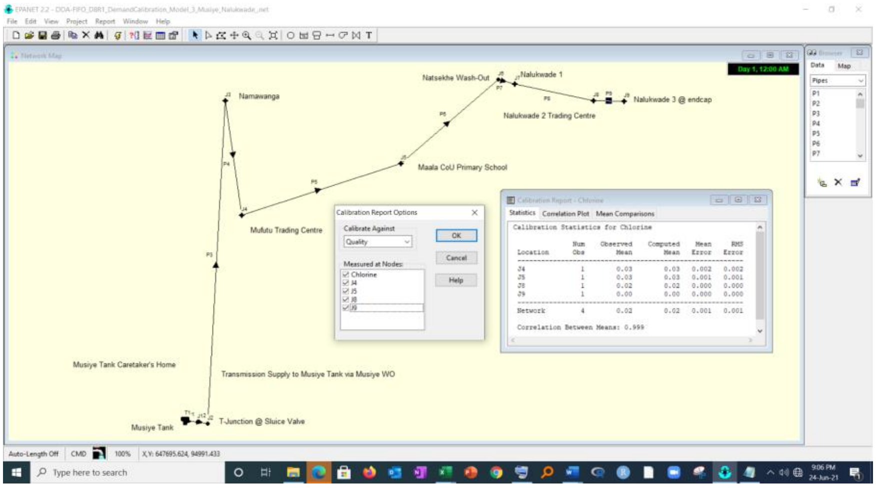
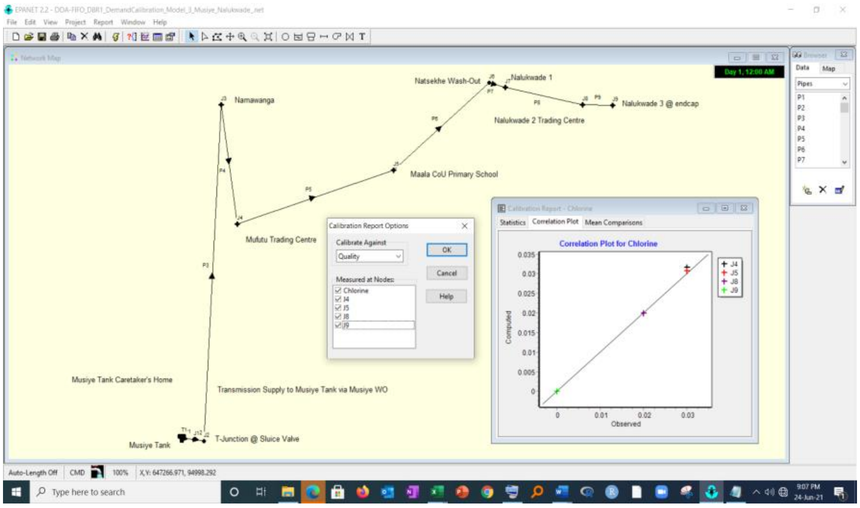


Figure A.1: Combined water transmission and distribution mains of Lirima gravity scheme

(a) Chlorine calibration statistics: (1) Mean error = 0.001mg/l (2) R^2 = 0.999



(b) Correlation plot for observed (**blue**) and modeled (**red**) chlorine concentrations



(c) Comparison of observed (**blue**) and modeled (**red**) chlorine concentrations

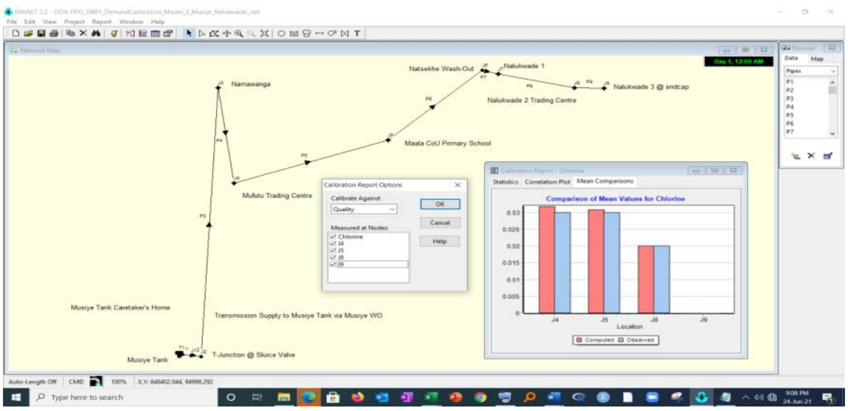
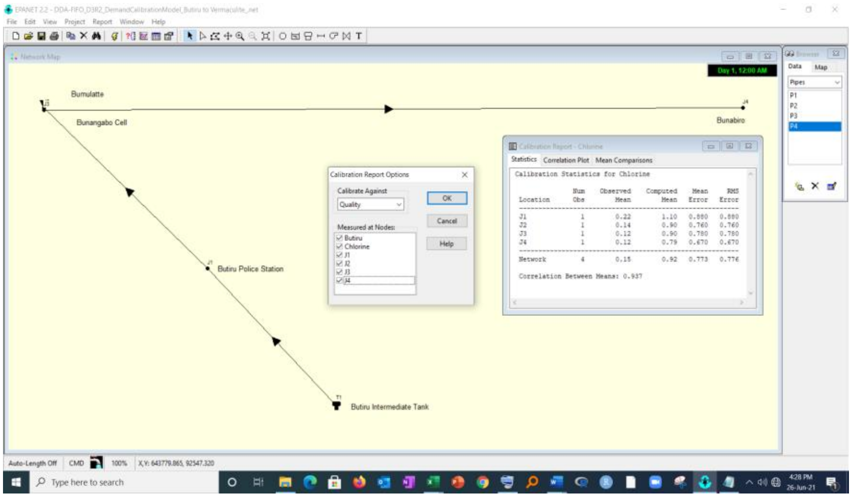
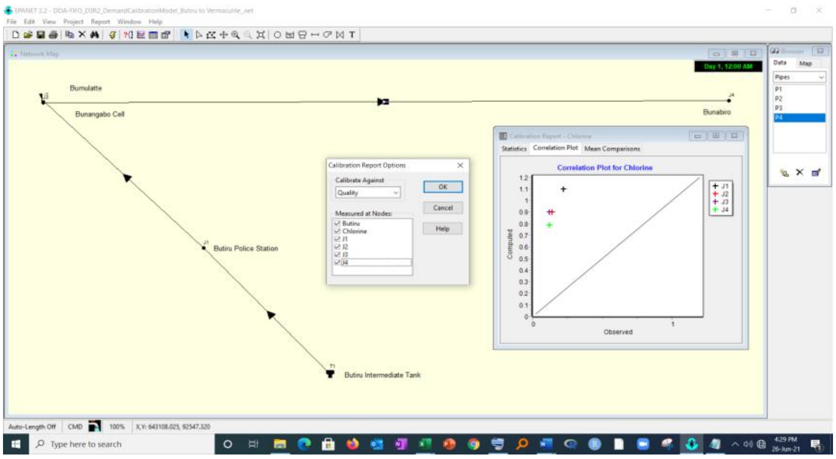


Figure A.2: Performance of EPANET model for Mufutu-Nalukwade water zone: Day8_Run1 demand model cross-validated with Day1_Run3 chlorine data

(a) Chlorine calibration statistics: (1) Mean error = 0.773 mg/l (2) R^2 = 0.937



(b) Correlation plot for observed (**blue**) and modeled (**red**) chlorine concentrations



(c) Comparison of observed (**blue**) and Modeled (**red**) chlorine concentrations

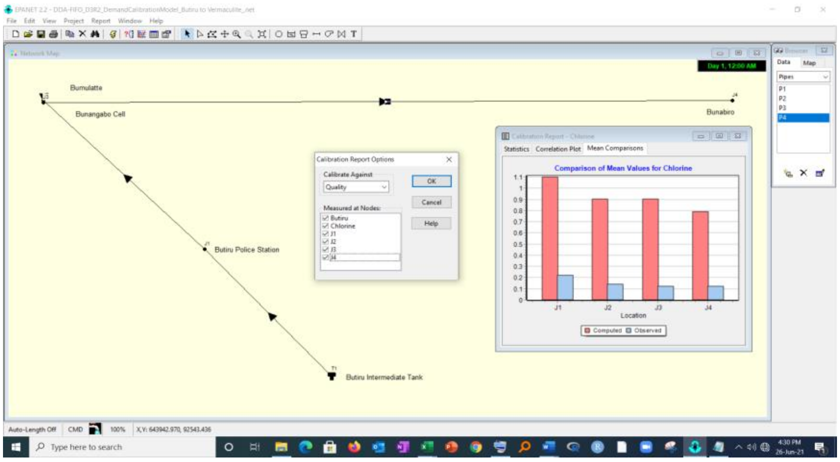
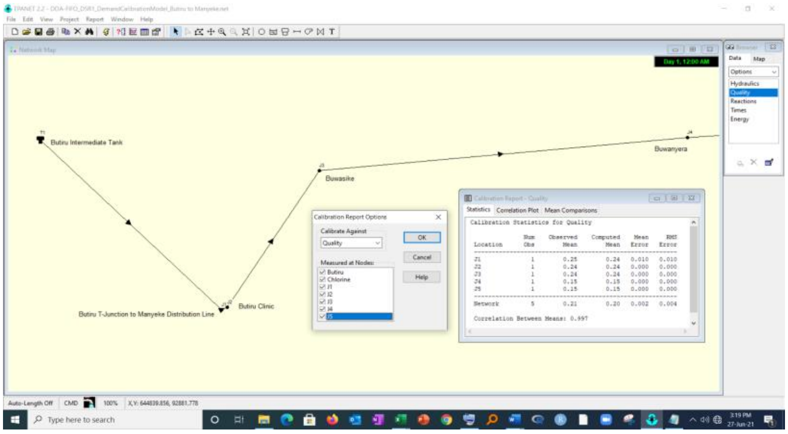
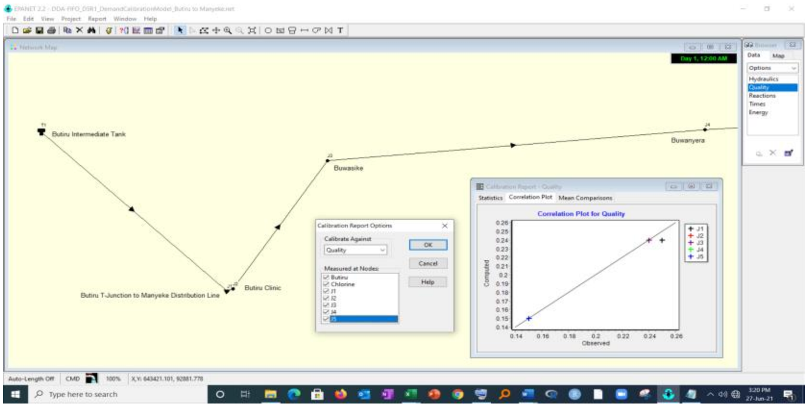


Figure A.3: Performance of EPANET model for Butiru-Vermiculite water zone: Day3_Run2 demand model cross-validated with Day7_Run1 chlorine data

(a) Chlorine calibration statistics: (1) Mean error = 0.002 mg/l (2) R^2 = 0.997



(b) Correlation plot for observed (**blue**) and modeled (**red**) chlorine concentrations



(c) Comparison of observed (**blue**) and modeled (**red**) chlorine concentrations

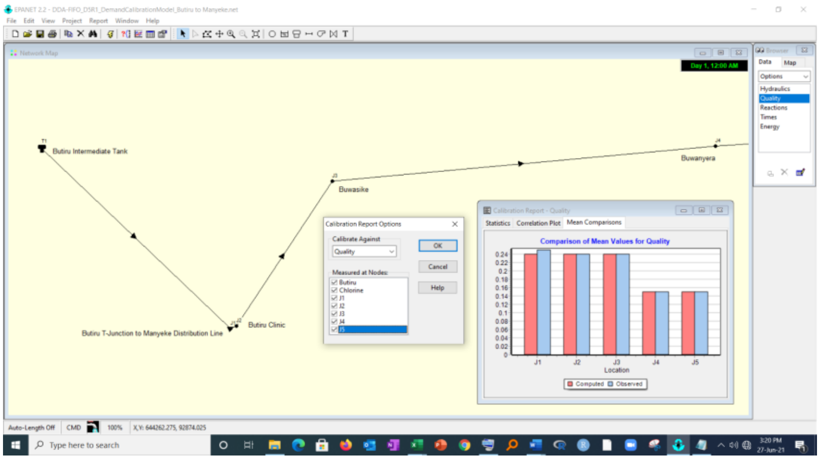


Figure A.4: Performance of EPANET model for Butiru-Manyeke water zone: Day5_Run1 demand model cross-validated with Day2_Run1 chlorine data

Appendix VIII: Table A.6: Two principal component solution analysis statistics

KMO and Bartlett's Test

Kaiser-Meyer-Olkin measure of sampling adequacy.		0.510
Bartlett's test of sphericity	Approx. Chi-Square	12.741
	df	3
	Sig.	0.005

Communalities

	Initial	Extraction
initial_chlorine	1.000	0.652
distance	1.000	0.653
EC	1.000	1.000

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component	
	1	2
distance	0.808	
initial_chlorine	0.807	
EC		0.999

Extraction Method: Principal Component Analysis.
Rotation Method: Equamax with Kaiser Normalization.[a]
a. Rotation converged in 3 iterations.

Total Variance Explained

	Initial Eigenvalues			Extraction sums of squared Loadings			Rotation sums of squared Loadings		
Component	Total	% of Variance	Cum %	Total	% of Variance	Cum %	Total	% of Variance	Cum%
1	1.322	44.061	44.061	1.322	44.061	44.061	1.305	43.488	43.488
2	0.983	32.771	76.832	0.983	32.771	76.832	1.000	33.344	76.832
3	0.695	23.168	100.000						

Extraction Method: Principal Component Analysis.

Appendix IX: Table A.7: Three principal component solution analysis statistics

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.510
Bartlett's Test of Sphericity	Approx. Chi-Square	12.741
	df	3
	Sig.	0.005

Communalities

	Initial	Extraction
initial_chlorine	1.000	1.000
distance	1.000	1.000
EC	1.000	1.000

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component		
	1	2	3
EC	0.999		
distance		0.988	
initial_chlorine			0.988

Extraction Method: Principal Component Analysis.
Rotation Method: Equamax with Kaiser Normalization.[a]
a. Rotation converged in 3 iterations.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cum %	Total	% of Variance	Cum %	Total	% of Variance	Cum %
1	1.322	44.061	44.061	1.322	44.061	44.061	1.000	33.336	33.336
2	0.983	32.771	76.832	0.983	32.771	76.832	1.000	33.332	66.668
3	0.695	23.168	100.000	0.695	23.168	100.000	1.000	33.332	100.000

Extraction Method: Principal Component Analysis.

Appendix X: Table A8: Four principal component solution analysis statistics

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.547
Bartlett's Test of Sphericity	Approx. Chi-Square	271.410
	df	45
	Sig.	0.000

Communalities

	Initial	Extraction
initial_chlorine	1.000	0.825
distance	1.000	0.817
travel_time	1.000	0.734
diametre	1.000	0.705
turbidity	1.000	0.639
EC	1.000	0.439
pH	1.000	0.312
temp	1.000	0.748
pressure	1.000	0.693
velocity	1.000	0.602

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
initial_chlorine	0.244	0.122	0.132	0.856
distance	0.871	0.149	- 0.028	0.190
travel_time	0.834	0.095	- 0.065	0.160
diametre	0.834	- 0.069	- 0.061	- 0.020
turbidity	0.106	- 0.052	- 0.786	- 0.090
EC	- 0.029	- 0.231	0.614	- 0.091
pH	0.127	0.457	- 0.276	0.101
temp	0.148	0.233	0.534	- 0.622
pressure	0.180	0.782	0.207	0.080
velocity	- 0.152	0.738	- 0.146	- 0.110

Extraction Method: Principal Component Analysis.
Rotation Method: Equamax with Kaiser Normalization.[a]
a. Rotation converged in 6 iterations.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cum %	Total	% of Variance	Cum%	Total	% of Variance	Cum %
1	2.582	25.819	25.819	2.582	25.819	25.819	2.315	23.153	23.153
2	1.500	15.001	40.820	1.500	15.001	40.820	1.526	15.263	38.416
3	1.431	14.307	55.126	1.431	14.307	55.126	1.445	14.452	52.869
4	1.002	10.017	65.144	1.002	10.017	65.144	1.228	12.275	65.144
5	0.907	9.067	74.211						
6	0.746	7.464	81.675						
7	0.726	7.255	88.930						
8	0.501	5.013	93.943						
9	0.416	4.162	98.104						
10	0.190	1.896	100.000						

Extraction Method: Principal Component Analysis.