UNIZIK JOURNAL OF AGRICULTURAL AND BIOSYSTEMS ENGINEERING

UJABE

Journal homepage: https://journals.unizik.edu.ng/ujabe/

June (2026), 1(3): 66-89

ISSN: 3141-2912

WATER QUALITY MODELING USING ARTIFICIAL NEURAL NETWORK AT THE DEAD-END SECTIONS OF DRINKING WATER DISTRIBUTION SYSTEM

***J.I. Ubah[1], Tochukwu Chibueze Ogwueleka[2], Ogunkuade O.G.[2], J.C. Chukwuneke[3]**

[1]*Department of Agricultural and Bioresources Engineering, Nnamdi Azikiwe University Awka, Nigeria*

[2]*Department of Civil Engineering, University of Abuja, FCT, Nigeria*

[3]*Department of Civil Engineering, Chukwuemeka Odimegwu Ojukwu University, Anambra State*

**Corresponding Author's E-mail:* ji.ubah@unizik.edu.ng

Received: 12th April 2026; **Accepted:** 20th May 2026; **Available Online:** 30th June 2026

ABSTRACT

Despite efforts in treating water in water distribution systems, water borne disease outbreaks persist as a result of undergoing factors within the treatment tank, one of which is the water condition at the dead-end section. Existing models have addressed these problems in the mains and distribution, but that of dead end is still lagging. Therefore, this research seeks to address issue of water deterioration at the dead-end section of water distribution system of Abuja, Federal Capital territory of Nigeria. The Artificial Neural Network (ANN) model was deployed, for its ability to accurately model data of both short and long duration so as to establish dependability of the governing water quality parameters. The ANN models developed for training, testing, validation, and overall analysis indicated no significant variation between the predicted outputs and the original data values, demonstrating strong predictive capability and reliability. Among the different RANDIV distribution ratios examined, the 70:15:15 ratio produced better results for the Bayesian Regularization (BR), Levenberg-Marquardt (LM), and Scaled Conjugate Gradient (SCG) algorithms compared to the 80:10:10 and 60:20:20 ratios, with the BR algorithm showing superior overall performance. The ANN model equation, Output = 0.68(Target) + 0.02, showed close agreement between predicted and observed residual chlorine values, with only about 10 out of 259 data points exhibiting deviations up to 0.45%, while the majority differed by approximately 0.01%. The model achieved high reliability with R^2 values of 0.81 for training and between 0.62 and 0.75 for testing, validation, and overall datasets. Furthermore, the model-generated initial residual chlorine value of 0.02 mg/l complied with the USEPA recommended standard of less than 0.1 mg/l. The BR algorithm also demonstrated the best computational efficiency with a performance value of 0.0092 achieved within only 10 epochs, while LM and SCG required over 1000 epochs. The Nash-Sutcliffe Coefficient of Efficiency of 0.9236 (92.36%) further confirmed the accuracy and effectiveness of the developed model.EPANET simulations of the water distribution network revealed no significant chlorine loss around the dead-end sections, thereby maintaining acceptable residual chlorine concentrations and minimizing water quality deterioration. Simulations conducted over critical periods of 24 hours, 144 hours, and 240 hours demonstrated that the distribution system consistently maintained recommended chlorine residual levels under varying flow velocity, demand, and head conditions. Residual chlorine concentrations at selected dead-end nodes remained within safe operational limits, with a maximum value of 0.44 mg/l, which is below the recommended limit of 0.5 mg/l. The findings indicate that the distribution system is capable of preventing the growth and regrowth of disease-causing microorganisms while ensuring safe water quality throughout the network.

Keywords: Artificial Neural Network, Dead-end, Water Quality, Water Treatment, Wastewater

1.0 INTRODUCTION

Water supply systems fit into the group of serious infrastructure systems that play a major function in the socio-economic growth of our modern societies and the safeguard of public health. Their chief duty is to make available access to clean and safe water for all our municipal, industrial, and fire-fighting use. Corrosion of pipe materials may also bring about an increase in the regularity of outbreaks and discoloration of the drinking water (Cunningham and Ghahramani, 2014), or the discharge of toxic heavy metals (CERT,2015). The large scale of water distribution systems impacts to this degeneration mainly as a result of the extended transport times between treatment and consumption points, which is generally known as the “water age”, since more time is available for interactions to occur (Eliades and Polycarpou, 2012). Operating under restricted budgets, water utilities are confronted with a considerable task to sustain water quality in their old distribution systems and to stay in agreement with all regulatory guidelines. Bad water quality administrations are the major determinants allowing regulatory breaches, chemical pollution events and disease outbreaks to take place. The increased concern in employing smart network technologies over the past decade has been supplemented with a reliable growth in the advancement of related industrials devices and solutions, such as developed metering technologies (Izquierdo et al, 2007), sensor networks, data analytics tools and automation systems. A well designed water supply system is required to function optimally such that consumers could have access to drinking water with good quality (Bwire et al, 2015). In order to make water available for human consumption it must be properly treated. However, these problems do not stop at the point of treatment but, further arise as the water travels through the distribution system to the consumer (Baker et al, 2019). The sources of contamination threat along the distribution networks are associated with the dead ends, point of pipe leakages, settlement of disinfectants, corrosion of pipes and several other factors which usually lead to recontamination of the treated water before getting to the consumers (Wang et al,2014). In spite of huge amount of money budgeted for water treatment yearly, studies have shown that there are still many disease outbreaks; causes of which have been traced to poor handling of disinfectants at the point of treatment, transportation (leakages and crossing points of potable water and sewer lines), storage and handling by the consumers ((Baker et al, 2019; Wang et al, 2014).

In all, there is a need to ensure healthy living through water supply system that is free from contaminants and pathogenic organisms capable of causing diseases right from the source to consumers (Pedersen, 1990; Zlatanovic et al, 2017; Liu et al, 2017; Grayman and Clark, 1990). Many models based on long-term data collection with the usual long equations have been used to address this issue in the mains and distribution networks but that of dead-ends still lingers. This work centres on the option of using Artificial Neural Networks which can utilize both short- and long-term data with simple equations to model the dead-end due to its ability to quickly minimize error while analysing water quality data using the Multilayer Perceptron Algorithm (MLPA) with forward distribution and backward transmission feed forward techniques until tolerable error value is obtained. Many previously developed models have addressed water quality in the mains and distribution systems while that of dead ends are yet to receive sufficient attention. Abokifa, Yang, Lo and Biswas, (2016) developed a multi-component mathematical model to investigate how pipe biofilms and multi-species chemical dynamics accelerate the formation of hazardous disinfection byproducts (DBPs), specifically trihalomethanes (THMs), in distribution systems. Aparicio and Li, (2016) built upon Buchberger's work to launch the Washington University Dead End Simulator (WUDESIM), a model coupled with stochastic demand generation to evaluate spatial transport parameters and chemical dispersion in dead-end pipe sections. This study addresses these gaps by utilizing, the artificial neural networks (ANN) method, for its ability to accurately model data of both long and short duration with simple equation using the four years (2013-2016) water quality data obtained from the WDS and reservoir tanks of various Districts in Abuja, Nigeria.

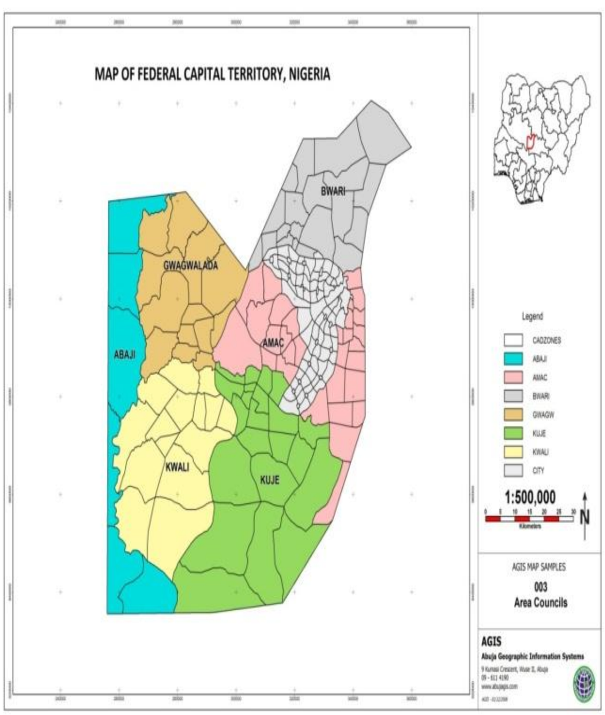
1.1 Water Quality Modeling at Dead–end Sections of Water Distribution Systems

According to Horta,2007; Housh and Ohar, (2017a),it was investigated that at dead end sections of supply system, the growth of bacteria is elevated that the quantity of residual chlorine was between 0.4 and 0.6 mg/l which demonstrates that while growth of HPC has increased, chlorine use has been lessened,(which makes water harmful).The likely cause being that, the velocity at those points are already low demonstrating that Chlorine didn't influence bacterial HPC at dead ends as earlier studies demonstrates that there was no association between residual Chlorine and HPC (Liu,2017). In 2015 alone, nearly one-quarter of the U.S. population was supplied by water systems that had as minimum one or more reported infringements of the Safe Drinking Water Act (SDWA) (Ahmed et al, 2016). Also, sporadic low flow velocities and regular stagnation are related with the dead-end mains of the distribution system **(**Barbeau et al, 2005); investigated that, these sections of the distribution system are well identified as problematic positions for undue long residence times, leading to quick water quality fall, disinfectant residuals loss and great potential for bacterial re-emergence **(**Galvin, 2011). Housh and Ohar, 2017a reported that the depreciation in water quality at dead end section is as a result of various determinants. Buchberger and Lee, (1999) and Tzatchkov et al, (2002) in their investigations saw that modeling of water quality in dead-ends is yet to benefit much popularity among investigators even when it is acknowledged that the water distribution system (WDS) comprises of over 25% of the total infrastructure of an area servicing sufficient share of the residential consumers.

2 Materials and Methods

2.1 Study Area

Abuja; the Federal Capital Territory is located in the central part of Nigeria (Plate 3.1) at Longitude $7^{0}29^{'}28.69''$ and Latitude $9^{0}4' 20.15''$ with an area of about 7135 km^{2}and a population of about 5,000,000 presently but expected to rise to 10,000,000 with time. The water supply is from the Lower Usman Dam Water Treatment Plant (LUDWTP) with a treatment capacity of 720 million litres per day. The Abuja Water Distribution System (WDS) is divided into five zones; viz: The City, Asokoro/Karu, Kubwa, Gwagwalada and Kuje. The city is made of four developmental Phases out of which Phases I and II have been substantially developed while the remaining are developing in some axis at a very faster rate and at slower rates. The water distribution system is designed in such a way that all supplies are by gravity.



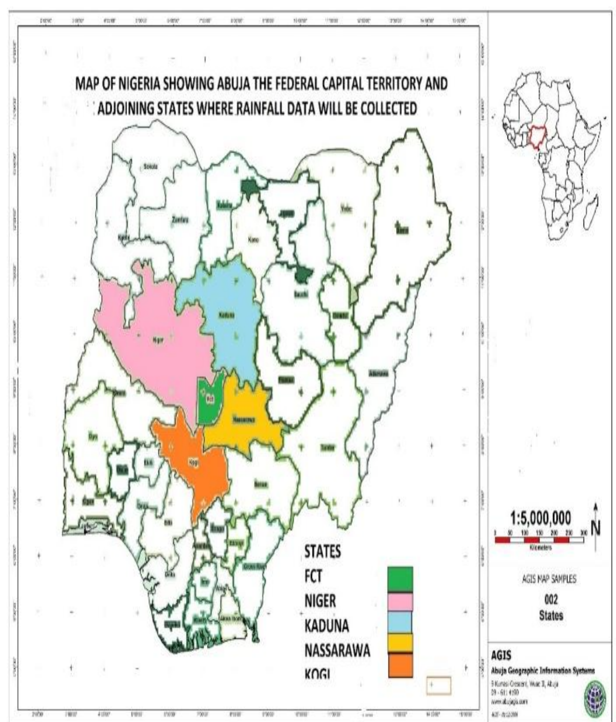


Fig. 1: Map of Abuja (FCT) showing the Area Councils   Figure 2: Map of Nigeria showing Abuja
Source: Google Search- Map of Nigeria States, May 2018

2.2 Data Collection

Four-year data consisting of water quality physicochemical parameters like: odour, taste, pH, temperature, salinity, conductivity, total dissolved solids, turbidity, dissolved oxygen, residual chlorine, chlorine ion, total alkalinity and total hardness with the bacteriological results, covering year 2013 to 2016, was collected from the FCT Water Board. These data were collected from the following locations within the Phase 1 development of the Federal Capital City: Areas 1, 2, 3 and 8 of Garki 1 District, Garki 2 District, Zone 1, 2, 3, 4, 5 and 6 of Wuse 1 District, Wuse 2 District, Asokoro District, Maitama District, Karu Site Resident, Apo-Gudu District, Tanks 3,4and 6 and Karu Tank. The skeletonized FCT water distribution reservoir.

2.3 Model Structure

Each data group for the rainy season, dry season and combined was trained using Levenberg-Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG) with different Hidden Layers, Ratios and Epochs, the best results from each of the combinations were recorded. The SCG is a supervised learning algorithm for feed forward neural network (FFNN), with its step size scaling mechanism which minimized duration of iteration, increased efficiency and performance of LM algorithm but cannot reduce error oscillation. The data for the Rainy Season, Dry Season and Combined were tagged as ANN I, ANN II and ANN III respectively. Also, different ANN models were built and verified for optimal number of nodes in the hidden layer to avoid over-fitting by applying different numbers of Hidden Layers, Delays and RANDIV data distribution ratios for Training, Test and Validation from which three different ANNs were obtained using LM, BR and SCG Algorithms. Model for the Artificial Neural Networks was obtained.

2.3.1 Input/ Output Variables for Physicochemical (Quality) Data-ANN

The physicochemical parameters obtained from the Water Board was trimmed down to pH, Temperature, Conductivity, Total Dissolve Solids, Turbidity, chlorine ion, Total Alkalinity and Total Hardness as input parameters based on the strength of their relationships with the residual chlorine as the output variable. Salinity and dissolved oxygen were not considered as part of the parameters that significantly affect the residual chlorine as right from inception, the values recorded for the salinity is almost the same throughout, they are 0.3 and 0.4 all through.

The Architecture of the system is presented in Figure 3

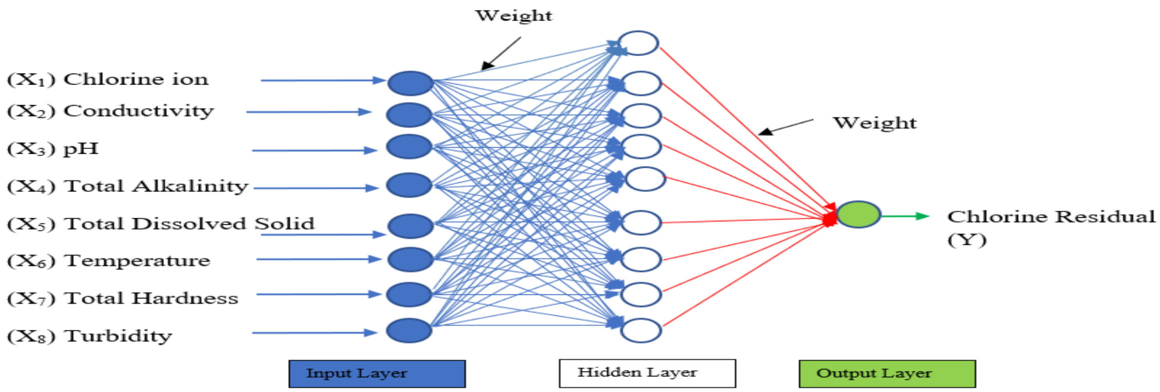


Figure 3: Multilayer Perceptron Neural Network for the Chosen Water Quality Parameters

2.3.2 Model Descriptions

Artificial Neural Networks (ANN)

ANN is a process consisting of interrelated components that learn to identify outlines of available data and which then utilise what it has learnt during training to detect similar patterns and use for prediction in future. Moghadassi et al, 2009, inferred that prior to training, random small weights ranging between -1 and +1 are introduced within the network and the required output for any of the input patterns is produced during training. A process known as *forward pass* is introduced by applying the input pattern to calculate an output different from the expected value as a result of the randomly weights applied. An error which is the difference between the Target and Actual Output is then estimated. A *back-pass process* using the calculated error with an activation process will now produce an output of each neuron that is closer to the target. The process is repeated until the error is minimized. The commonly used ANN is the Back Propagation (BP) type **(**Maier and Dandy, 2000), which is also known as Feedback Networks.

2.3.3 Structure of ANN

The typical structure of a 3-layer ANN is represented as shown in figure 4

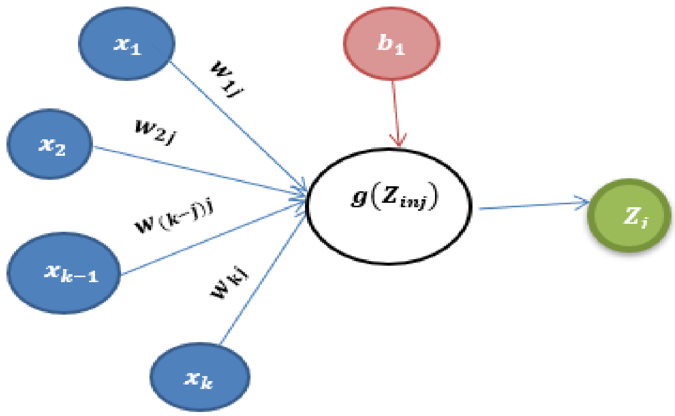


Figure 4: Typical 3-Layer Artificial Neural Networks.

2.3.4 Data Processing for Quality-ANN

The total number of data (Combined Data) for the four years is 385 out of which 259 and 126 constitute the Rainy and Dry Season data respectively. For each of the seasons and combined, the data was divided through a method of RANDIV into ratios 70:15:15, 60:20:20 and 80:10:10 for training, test and validation respectively. The essence of this process is to obtain the model that has the best performance and efficiency. Having achieved this, the ANN models were used to simulate the rainy season data.

2.3.5 EPANET

EPANET is a model developed by EPA that performs an extended period simulation of hydraulic and water quality behaviour within pressurized pipe networks. Water Distribution Systems (WDS) is made of numerous numbers of water conveying units such as ground/overhead tanks, pipelines sizes (lengths and diameters), materials and fittings that forms a network **(**Blokker et al, 2008). The sizes and material types and fittings change the hydraulic behaviour of flow in the network which determines the required hydraulic model of a flow. The hydraulic models which simulate the flow quantity, flow direction, and pressure in the system, is expressed by the Advection-Dispersion-Reaction (ADR) equation below:

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} = E\frac{\partial^2 C}{\partial x^2} + f(c) \text{ .... } 1$$

where; ***C;*** represents the cross-sectional average concentration of the water quality parameter in mg/l), ***t;*** is the time (s), ***u;*** is the mean flow velocity in m/s, ***x;*** is the direction of the flow, ***E;***

represents the mixing coefficient of the axial dispersion in one-dimensional flow in m^2/s, and ***f* (*C*);** is the reaction function = $-KC$ with K the reaction constant.

Input/Output Variables for Hydraulic Data- EPANET

The hydraulics data required for modelling in the EPANET are in the following two parts:

1. **Links:** These are the pipes with the following parameters:

Length, L (m), Diameter, D (m), Pipe Roughness Coefficient, C (dimensionless)

2. **Nodes:** They are junction at which the various pipes meet and are of the following parameters:

(i) Elevation, H (m), (ii) Demand, Q (l/s)

The distribution system is an existing one and the point of interest is that of quality and not hydraulic, hence there was no need of redesigning the pipe sizes since these parameters are already known, therefore the values were obtained and inputted. There are three different pipe types in the distribution system and this consist of Cast Iron, uPVC and PE. In view of this, three different roughness coefficients corresponding to the material types were used. The pressure heads at each node were obtained as the difference between the elevation at the inlet and outlet of the pipe. The demand at each junction was calculated as a product of per capital demand of 130 l/c/d and the number of dwellers in those houses that get supply from the node with a multiplying factor to cater for gardening, wastages and other usages based on the planning criteria in the FCT Development Control.

Table 2: Sizes and Roughness Coefficients of Different Pipes Used in the Water Supply Network

Pipe material	Pipe size (mm)	Roughness coefficient (Hazen Williams, C)
PE	80	150
uPVC	100, 150, 200, 250, 300	140
Cast Iron CI	300, 400, 600, 800, 900, 1500	130

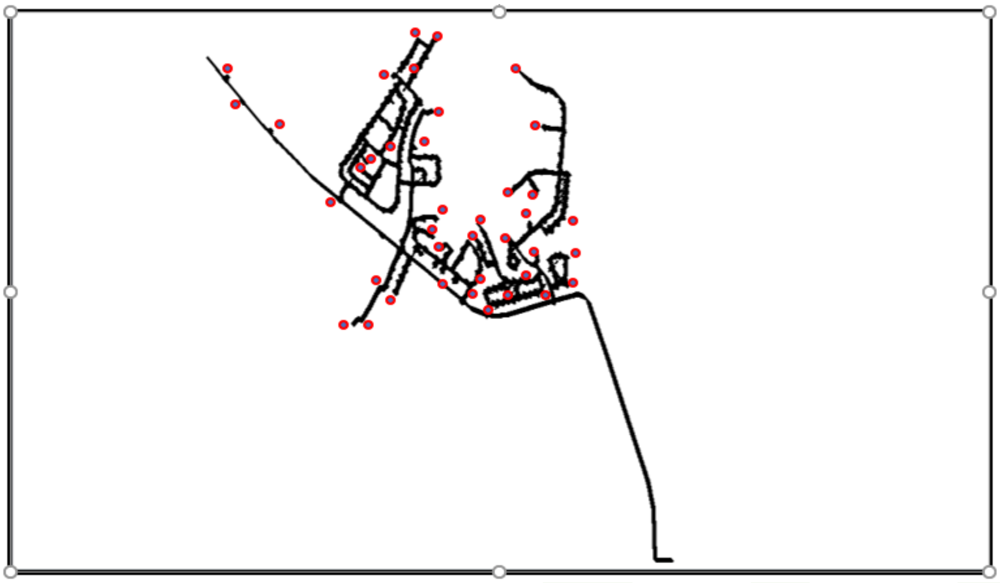


Figure 5: Skeletonized Gudu Water Distribution Pipe Network Showing Dead-Ends –EPANET

3.0 Results and Discussion

3.1 Results

For the three proposed models, ANN I and ANN II represent models for Rainy and Dry Seasons Data while ANN III is for the Combined data. When each of the data set was trained with Bayesian Regularization (BR), Levenberg Marquardt (LM) and Scaled Conjugate Gradient (SCG), with Hidden Layer =9, Delays=1:2, n=259, Data Distribution Ratio=70:15:15 and the results obtained are as presented in Figures 5 to 11 and Table 3.1 in the following sections.

Result for ANN I (Rainy Season)

(a) Train Bayesian Regularization (BR)

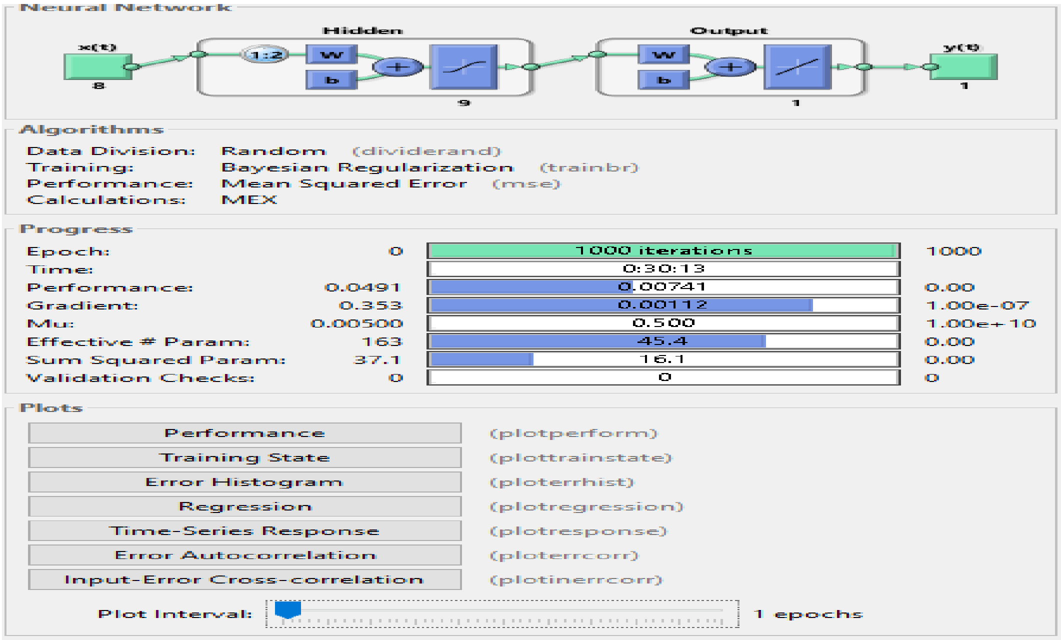


Figure 5: Neural Network of Bayesian Regularization for ANN I (Train BR)

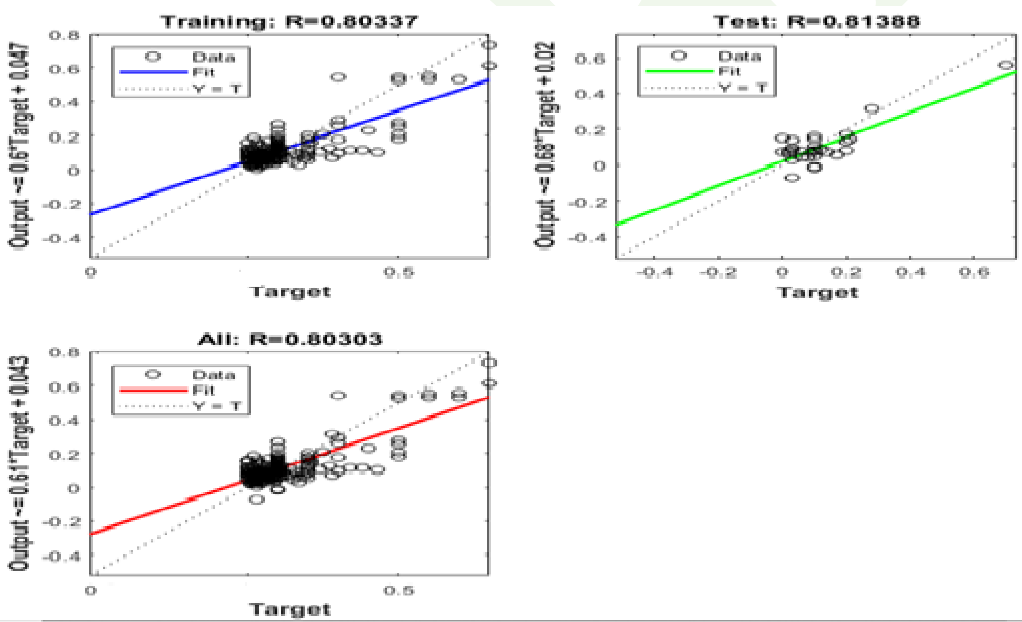


Figure 6: Regression Analysis for ANN I (Train BR)

The regression for the ANN I as indicated in Figure 6 is as high as 0.80 for both Training and while that of the Testing is 0.81 having a Performance of 0.00741for 1000 iterations within a period of 30 minutes 13 seconds, which shows that the constituent parameters are strongly good for the modeling.

(b) Train Levenberg Marquardt (LM)

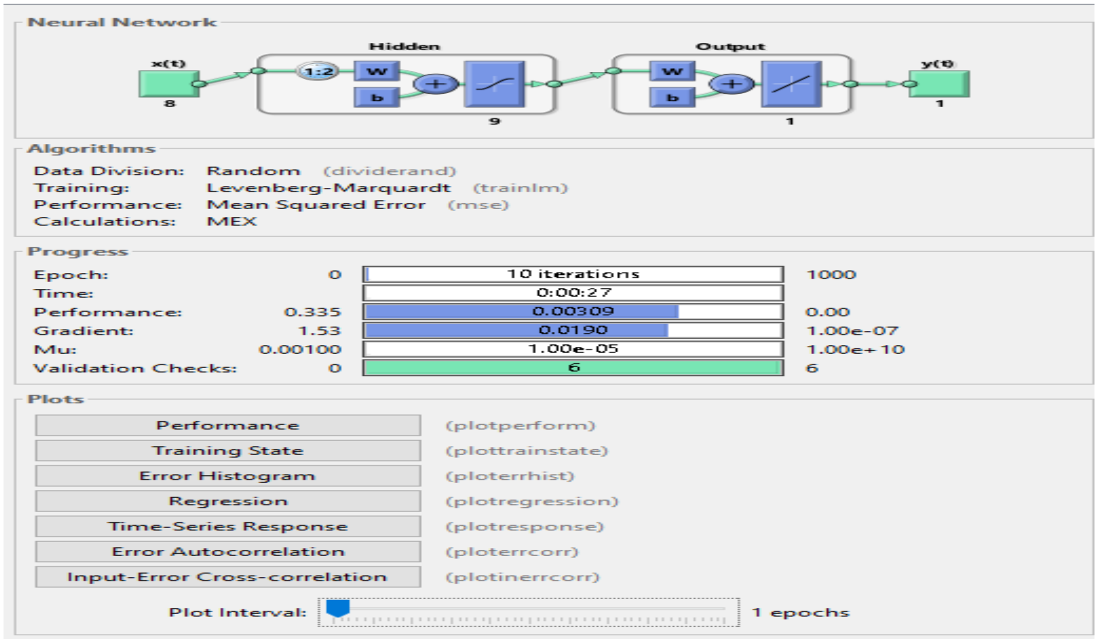


Figure 7: Neural Network of Levenberg Marquardt for ANN I (Train LM)

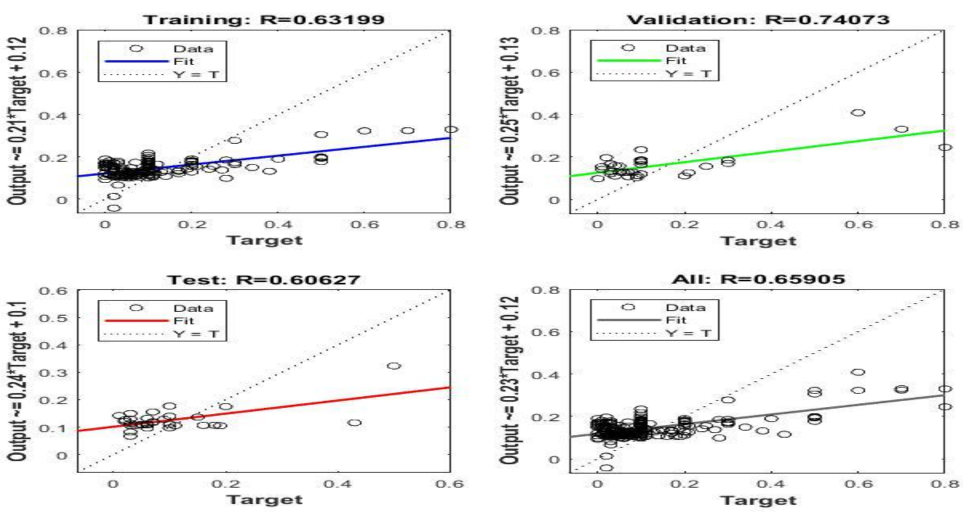


Figure 8: Regression Analysis for ANN I (Train LM)

For the Levenberg Marquardt, the regression as in Figure 3.4 is moderately high at values ranging from 0.60 to 0.74 for Training, Test Validation and All yielding a Performance of 0.00309with 10 Iteration and Performance =0.00309 within a period of 27 seconds. This is also an indication that the constituent parameters are suit for the modelling.

(c) Train Scaled Conjugate Gradient(SCG)

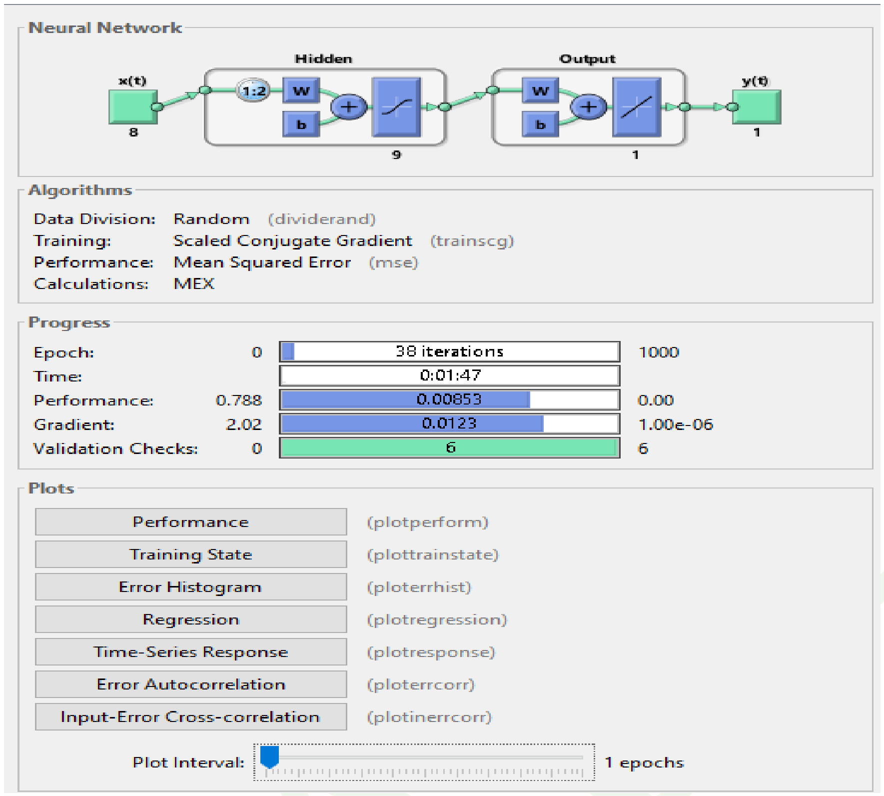


Figure 9: Neural Network of Scaled Conjugate Gradient for ANN I (Train SCG)

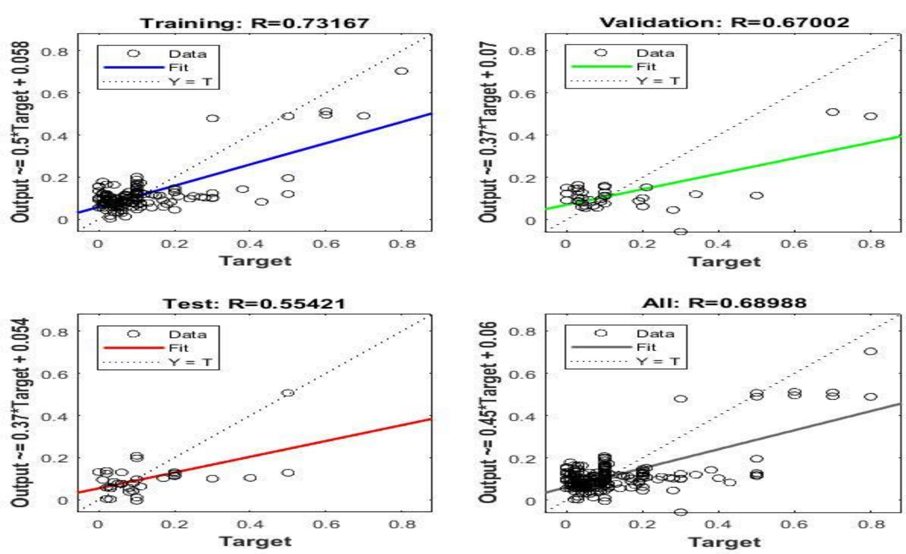


Figure 10: Regression Analysis for ANN I (Train SCG)

The regression value of 0.55 for test under SCG for the data is lower than those of training, validation and All with 0.73, 0.67 and 69 respectively for 1000Iterations and a Performance 0f 0.00853 achieved within 1 minute 47secoonds. These values are equally good for choosing the parameters.

Table 3. Summary of Results for ANN I, ANN II & ANN III Models (Rainy, Dry, Combined Season Data respectively)

			Output			R^2	
ALGORITHM		Bayesian Regularization (BR)	Levenberg-Marquardt (LM)	Scaled Conjugate Gradient (SCG)	BR	LM	SCG
ANN I MODEL	Training	0.6*Target + 0.05	0.21*Target + 0.12	0.51*Target + 0.06	0.8	0.63	0.73
	Test	**0.68 Target + 0.02**	0.24*Target + 0.10	0.37*Target + 0.05	**0.81**	0.61	0.55
	Validation	BR does not validate	0.25*Target + 0.13	0.30*Target + 0.07		0.74	0.67
	All	0.61*Target + 0.04	0.23*Target + 0.12	0.45*Target + 0.06	0.8	0.66	0.69
ANN II MODEL	Training	0.89*Target + 0.02	0.17*Target+ 0.01	0.35*Target + 0.11	0.98	0.25	0.62
	Test	0.20*Target + 0.07	0.042*Target + 0.05	0.005*Target+ 0.15	0.17	-0.04	0.03
	Validation	BR does not validate	0.36*Target + 0.11	0.35*Target + 0.09		-0.24	0.72
	All	0.78*Target + 0.01	0.10*Target + 0.02	0.19*Target + 0.13	0.77	0.13	0.44
ANN III MODEL	Training	0.25*Target + 0.10	0.13*Target + 0.10	0.17*Target + 0.11	0.53	0.37	0.41
	Test	0.24*Target + 0.11	0.20*Target + 0.09	0.19*Target + 0.11	0.41	0.56	0.27
	Validation	BR does not validate	0.39*Target + 0.08	0.29*Target + 0.11		0.78	0.47
	All	0.25*Target + 0.10	0.16*Target + 0.10	0.18*Target + 0.11	0.51	0.46	0.39

From Table 3 it can be seen that the R^2 values of ANN Model I for Training and Test using BR, LM and SCG Algorithms are higher than those of Models II and III. This is an indication that the rainy season data is more significant when compared with the other two seasonal data.

3.2 Results for EPANET:

The EPANET model gives the estimate of residual elements found at dead end-sections and analysis of the result obtained is presented in the Figures 11 to 18 while comparison of Observed, EPANET and ANN values shown in Figure 4. Figure 11 shows the initial water quality before simulation to give better understanding of the varying residual elements before reaching dead end-sections while Figure 12 shows the simulated network with velocity as simulation progressed.

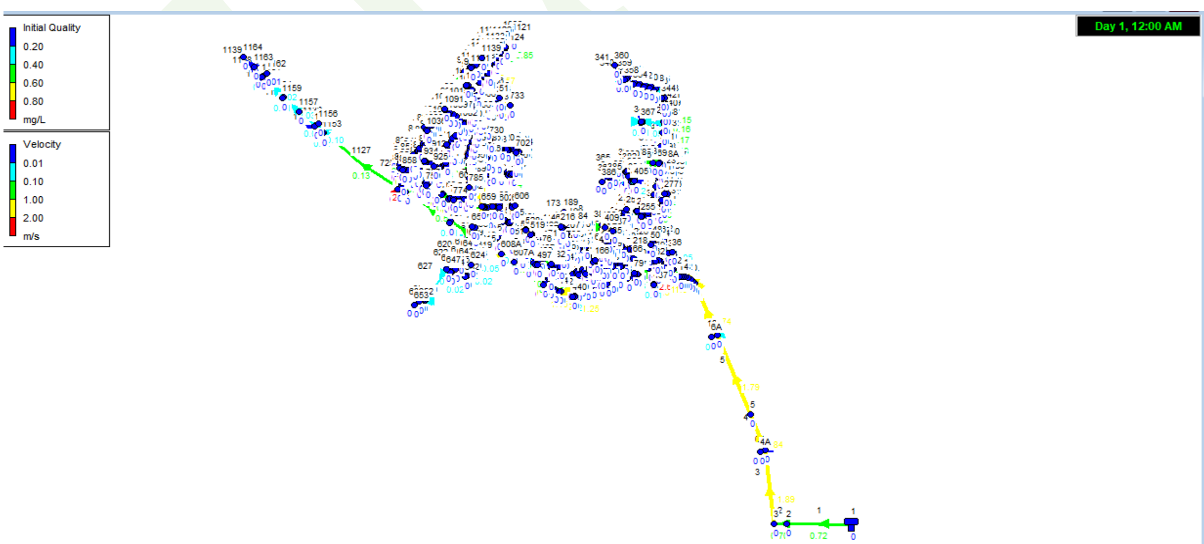


Figure 11: Process of simulation of initial water quality with velocity

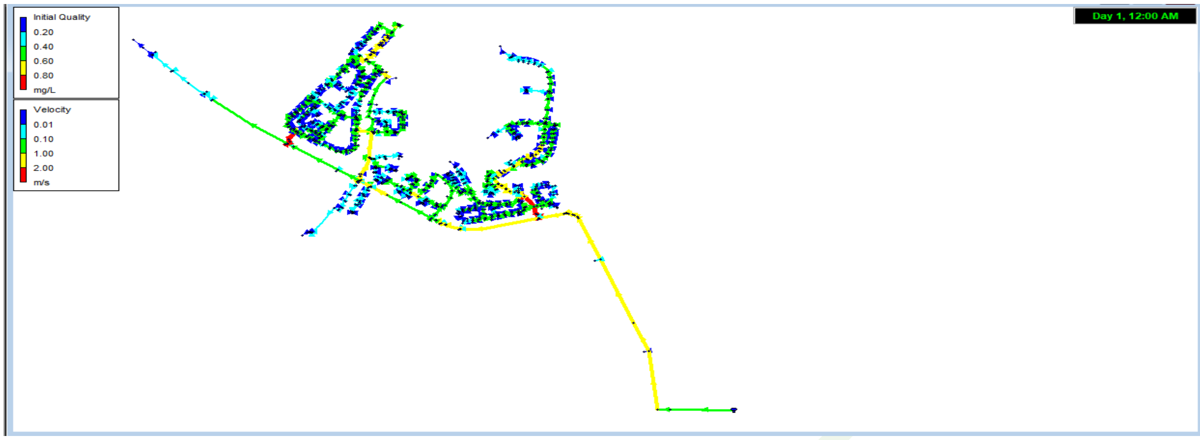


Figure 12: Result of Simulated water quality with velocity

The chlorine demands within the distribution pipe lines and the dead ends for the varying hours as shown in Figure 13 falls within the specified chlorine demands of 0.2-0.5mg/l while that of the mains exceeds this value. The blue region is the area where the dead-ends are found and within 12 to 24hrs of dosing they have stabilized and maintaining the specified standard of 0.25 to 0.5 mg/l while in the same period the mains maintained higher values beyond the standard values as expected. This indicate that there is no significant lack or excessive deposition of disinfectant at the dead-ends. The values in the mains keep varying between 1.00,0.75,0.50 and 0.25 which are above the specified values.

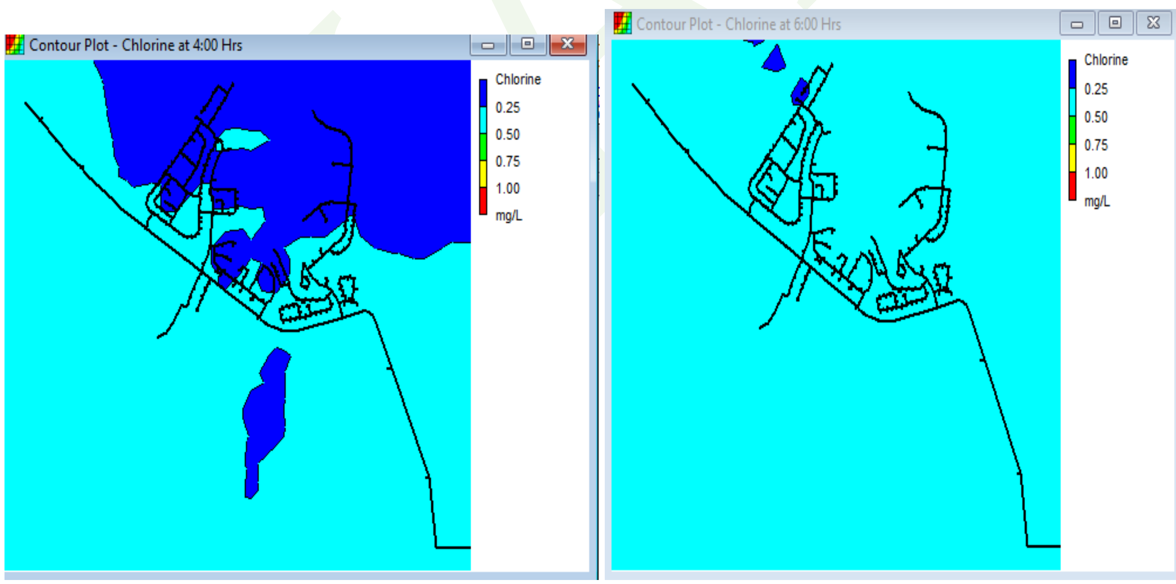


Figure 13: Contour plot for chlorine

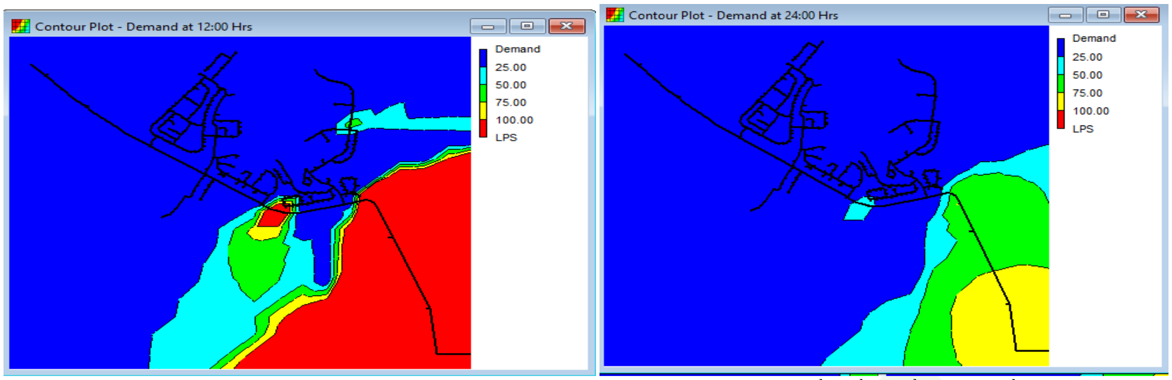


Figure 13: Contour Plot of Water Quality Demand at 4th, 6th, 12th, and 24th Hours

Towards the final distribution of the disinfectant for the water treatment, it was observed that the depletion continued at a constant rate. Figure 14 shows the distribution of the disinfectant and based on this graph, it is better to replenishment at the 64% point as suggested by the model at an hourly rate instead of waiting till the point when over 92%ofthe substituents would have been depleted.

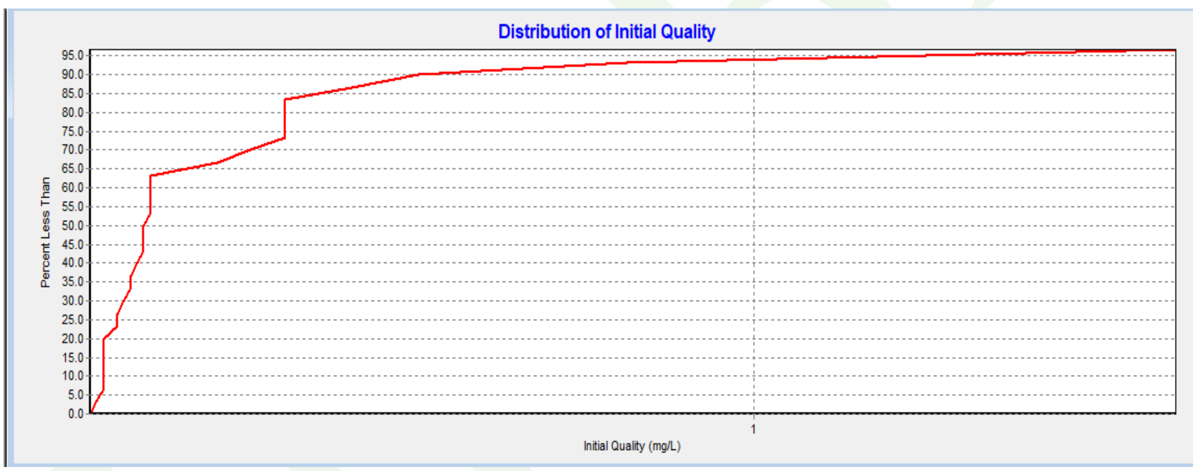
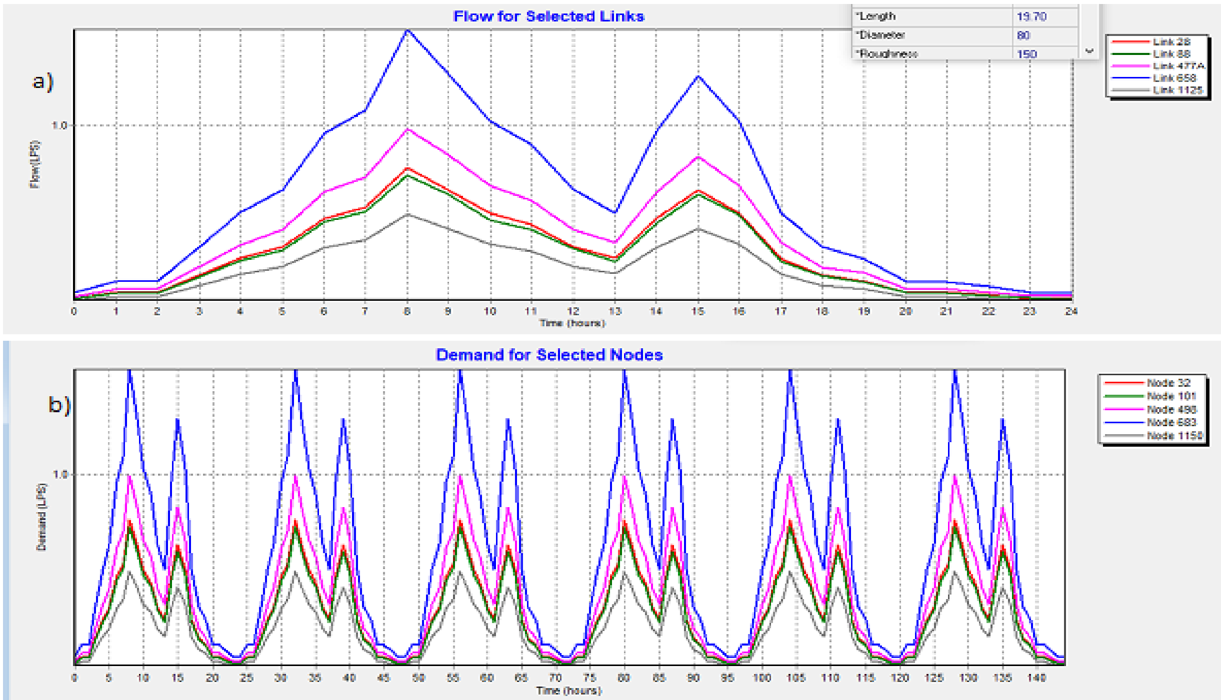


Figure 14: Distribution of Quality

Figure 15a shows hourly Demand at Selected Dead-End Nodes 28, 88, 477A, 658 and 1125 while 15b are for 144-hour demand at odes 32, 101, 498, 683 and 1150. The figures show the peak period trends for morning-rush and evening-recession. For 24 hours and 144 hours, certain nodes were selected as seen in figure 15a.



Figures 15 a-b:24-Hourly Demand for Nodes 28, 88, 477A, 658 and 1125 and144-Hour Demand at Selected Dead-End Nodes 32, 101, 498, 683 and 1150

Figures 16- 18, illustrate the rate of chlorine consumption on a 24-hour demand, 144-hour demand and finally 240 hour demands for selected nodes. It was observed that the substituents continue to decrease in water treatment, over certain distances and time. The nodes close to the dosage point utilizes the disinfectant almost immediately while other nodes farther away could not utilize it until four or more hours after. However, there are no indications of inadequate or excess chlorine in any of the dead-end sections.

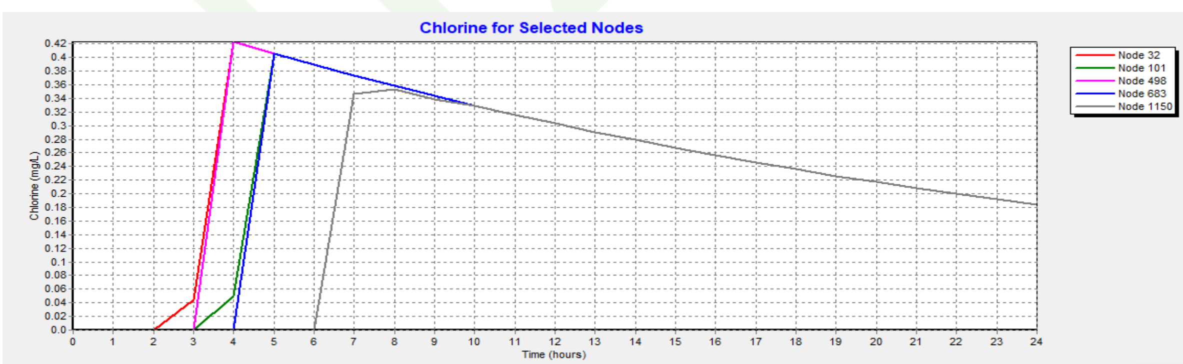


Figure 16: 24-Hour Chlorine Consumption at Selected Dead-End Nodes 32, 101, 498, 683 and 1150

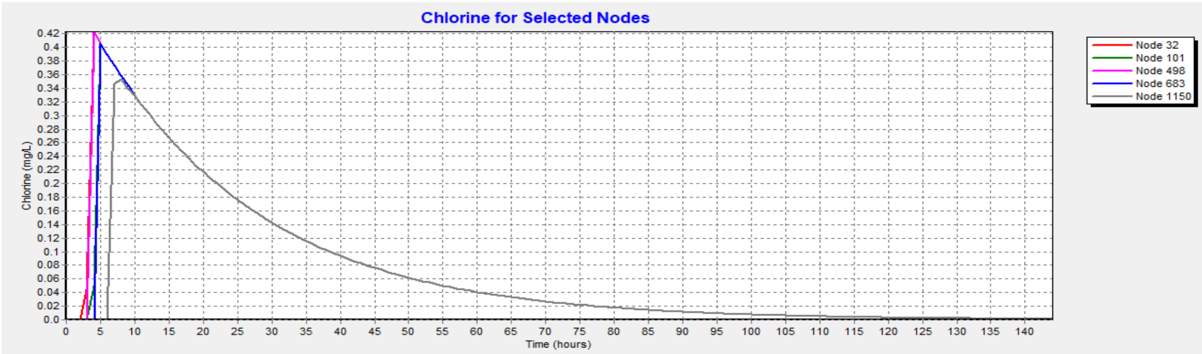


Figure 17: 144-Hour Chlorine Consumption at Selected Dead-End Nodes 32, 101, 498, 683 and 1150

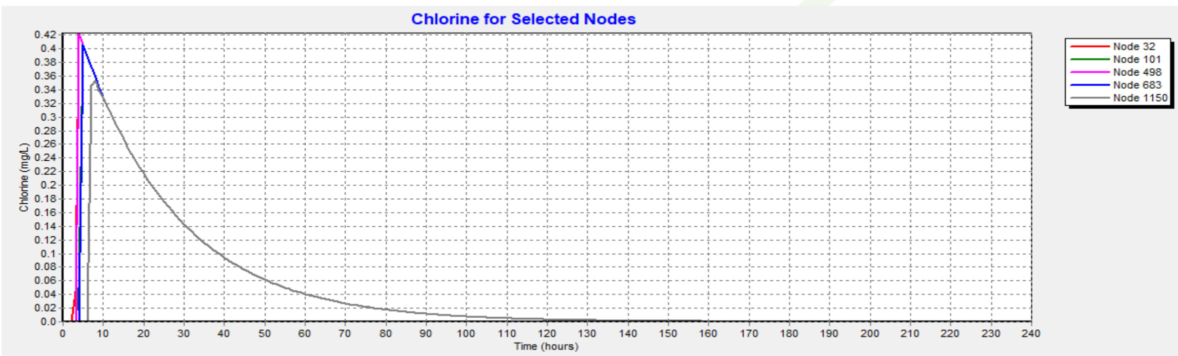


Figure 18: 240-Hour Chlorine Consumption at Selected Dead-End Nodes 32, 101, 498, 683 and 1150

Table 3. 1 performance evaluation of EPANET and ANN for the dead-end nodes Abuja WDS.

Node	EPANET Value (**X_{ob}**)	ANN Simulated Value (**X_{sim}**)						
		Output = 0.68*Target + 0.02						
N=	29	Slope =	0.68	Intercept=	0.02			
Dead-end Nodes	**X_{ob}**	**X_{sim}**	**$\lvert X_{ob}-X_{sim}\rvert$**	**$X_{ob}*X_{sim}$**	**X_{ob}^2**	**X_{sim}^2**	**$(X_{ob}-X_{sim})^2$**	**$(X_{ob}-X_{ob,av})^2$**
32	0.44	0.3192	0.1208	0.1404	0.1936	0.1019	0.0146	0.0013
101	0.42	0.3056	0.1144	0.1284	0.1764	0.0934	0.0131	0.1764
216	0.41	0.2988	0.1112	0.1225	0.1681	0.0893	0.0124	0.1681
189	0.41	0.2988	0.1112	0.1225	0.1681	0.0893	0.0124	0.1681
226	0.44	0.3192	0.1208	0.1404	0.1936	0.1019	0.0146	0.1936
250	0.42	0.3056	0.1144	0.1284	0.1764	0.0934	0.0131	0.1764
297	0.39	0.2852	0.1048	0.1112	0.1521	0.0813	0.0110	0.1521
367	0.38	0.2784	0.1016	0.1058	0.1444	0.0775	0.0103	0.1444
386	0.38	0.2784	0.1016	0.1058	0.1444	0.0775	0.0103	0.1444
406	0.36	0.2648	0.0952	0.0953	0.1296	0.0701	0.0091	0.1296
408	0.44	0.3192	0.1208	0.1404	0.1936	0.1019	0.0146	0.1936
462	0.44	0.3192	0.1208	0.1404	0.1936	0.1019	0.0146	0.1936
498	0.42	0.3056	0.1144	0.1284	0.1764	0.0934	0.0131	0.1764
509	0.41	0.2988	0.1112	0.1225	0.1681	0.0893	0.0124	0.1681
538	0.39	0.2852	0.1048	0.1112	0.1521	0.0813	0.0110	0.1521

567	0.41	0.2988	0.1112	0.1225	0.1681	0.0893	0.0124	0.1681
572	0.41	0.2988	0.1112	0.1225	0.1681	0.0893	0.0124	0.1681
595	0.42	0.3056	0.1144	0.1284	0.1764	0.0934	0.0131	0.1764
606	0.42	0.3056	0.1144	0.1284	0.1764	0.0934	0.0131	0.1764
632	0.42	0.3056	0.1144	0.1284	0.1764	0.0934	0.0131	0.1764
653	0.41	0.2988	0.1112	0.1225	0.1681	0.0893	0.0124	0.1681
654	0.42	0.3056	0.1144	0.1284	0.1764	0.0934	0.0131	0.1764
683	0.41	0.2988	0.1112	0.1225	0.1681	0.0893	0.0124	0.1681
733	0.4	0.292	0.108	0.1168	0.1600	0.0853	0.0117	0.1600
912	0.37	0.2716	0.0984	0.1005	0.1369	0.0738	0.0097	0.1369
934	0.34	0.2512	0.0888	0.0854	0.1156	0.0631	0.0079	0.1156
996	0.4	0.292	0.108	0.1168	0.1600	0.0853	0.0117	0.1600
1122	0.38	0.2784	0.1016	0.1058	0.1444	0.0775	0.0103	0.1444
1150	0.37	0.2716	0.0984	0.1005	0.1369	0.0738	0.0097	0.1369
Sum	**11.73**	**8.5564**	**3.1736**	**3.4730**	**4.7623**	**2.5327**	**0.3491**	**4.5700**
Mean	**0.4045**	**0.2950**	**0.1094**					
MAD			**0.1094**					
RMSE			**1**					
R^2			**1**					
CE			**0.9236**	**92.36**	**%Efficiency (Applying Nash-Sutcliffe Efficiency)**			

The result of Observed, EPANET and simulated ANN values indicates that they respond similarly while that of simulated ANN is closer to the Observed values more than those of EPANET. Hence, simulating water quality data at dead ends of water distribution network with ANN model is as good as other models.

3.3 Result for Multiple Regression Analysis using Wolfram Mathematica

The pre-processed data for each of the Rainy, Dry and Combined season was analyzed using 10,000 iterations at a confidence Level of 0.99 to obtain the Best Fitted Models. The model produced by ANN I was better than those of the other two.

3.3.1 Output Model for Rainy Season Data

The following output equations of models for Rainy, Dry and Combined season data show the effect of each parameter on the final residual chlorine in the distribution system and the dead-ends. The constant value in the equations indicates that whatever changes being contributed by any of the variables X_1, X_2, X_3, X_4, ... X_8 does not affect the initial required value but the overall value based on each variable's contribution. This can be seen when all the independent variables are set to zero. Though, initial value 0.745931 for the rainy season model is high but better than the other two periods Dry and Combined seasons with 1.131333 and 0.835307 when compared with specified standards. Also, its R^2 of 0.496765 though low but higher than those of the other two seasons data sets and they all exhibit dependency.

$$Y = 0.745931 - 0.004940\ X_1 - 0.000624\ X_2 - 0.0729413\ X_3 + 0.000652\ X_4 + 0.001183\ X_5 + 0.000271\ X_6 - 0.000151\ X_7 - 0.003164\ X_8$$

$\mathbf{R^2 = 0.496765}$

Where; Y and X_1 to X_8, are as defined in Section 2.7.2 and R^2 = the correlation coefficient

3.3.2 Output Model for Dry Season Data

The equation below shows the output model for the Dry Season Data with the same independent variables X_1, X_2, X_3, …… X_8. The initial value 1.131333 is high compared to specified standard though its R^2 of 0.417241 is low but there is still dependency.

$$Y = 1.131333 - 0.005683\ X_1 - 0.004620\ X_2 - 0.098916\ X_3 - 0.00191704\ X_4 + 0.006382\ X_5 + 0.0013473\ X_6 + 0.000234262\ X_7 - 0.00957355\ X_8$$

R^2 = **0.417241**

3.3.3 Output Model for Combined Data

The ANN III output model equation for the combined season also with the same independent variables show dependency. Fig

$$Y = 0.835307 - 0.004747\ X_1 - 0.001367\ X_2 - 0.086326\ X_3 + 0.000886\ X_4 + 0.002184\ X_5 + 0.000956\ X_6 - 0.000145\ X_7 - 0.002856\ X_8 \qquad R^2 = \mathbf{0.436597}$$

Table 3.3: Summary for Multiple Regression Results

Model	Require Starting Value of Disinfectant (mg/l)	R^2	Remarks
Rainy Season	0.74593	0.49676	Moderate value required. Highest R^2
Dry Season	1.13133	0.41724	Higher value required. Lower R^2
Combined Season	0.83531	0.43659	Closer to ANN I value. Higher R^2

4.0 DISCUSSION

4.1 ANN

The results obtained from the Artificial Neural Network (ANN) models for training, testing, validation, and overall simulation demonstrated strong agreement between the predicted outputs and the original residual chlorine data. This indicates that the ANN model possesses high predictive capability and can effectively simulate chlorine residual concentrations in water distribution systems. The insignificant differences between the observed and simulated values further confirm the robustness and reliability of the developed model. Similar findings were reported by Nourani et al. (2019), who noted that ANN models are highly efficient in capturing nonlinear relationships in water quality and hydraulic systems because of their adaptive learning and self-organizing abilities. Likewise, Rajaee et al. (2020) observed that ANN techniques provide accurate prediction results for complex environmental datasets where conventional statistical models often fail. The study showed that the RANDIV distribution ratio of 70:15:15 for training, testing, and validation produced better results for Bayesian Regularization (BR), Levenberg–Marquardt (LM), and Scaled Conjugate Gradient (SCG) algorithms than the 80:10:10 and 60:20:20 ratios. This suggests that a balanced data partitioning structure improves model generalization and prediction efficiency. According to Kim and Pachepsky (2016), appropriate data splitting is essential for preventing overfitting and ensuring stable predictive performance in ANN applications related to environmental and water quality modelling. The improved performance observed in the 70:15:15 ratio indicates that sufficient data were available for both model training and validation, thereby improving the model's ability to predict unseen data accurately.

The ANN model equation, Output = 0.68(Target) + 0.02, revealed a strong linear relationship between the predicted and observed residual chlorine concentrations. Out of the 259 rainy-season data points analyzed, only about 10 points showed deviations up to 0.45%, while the remaining values differed by approximately 0.01%, indicating excellent approximation and prediction accuracy. Similar studies by Mosavi et al. (2018) reported that ANN models are capable of achieving minimal prediction errors in water resource simulations due to their strong pattern-recognition and nonlinear mapping characteristics. The very small deviations observed in this study therefore indicate that the ANN model can effectively reproduce actual field conditions and accurately estimate residual chlorine concentrations within the distribution network. The coefficient of determination (R^2) values obtained in this study also demonstrate satisfactory model reliability. Although the training phase recorded the highest R^2 value of 0.81, the testing, validation, and overall values ranging from 0.62 to 0.75 are still considered acceptable for environmental and water quality modeling. Adnan et al. (2021) explained that R^2 values above 0.60 indicate strong correlation and acceptable predictive reliability in hydrological and environmental simulations. Therefore, the ANN model developed in this study can be considered suitable for predicting chlorine residual behaviour in water distribution systems under varying environmental conditions.

The initial residual chlorine concentration of 0.02 mg/l generated by the ANN model falls within the permissible limit recommended by the United States Environmental Protection Agency (2023) for potable water systems. This implies that the predicted chlorine dosage is safe and sufficient to maintain microbial water quality without introducing excessive chlorination effects. Similar observations were made by Deborde and von Gunten (2018), who explained that maintaining residual chlorine within acceptable regulatory standards is critical for ensuring microbiological safety while minimizing the formation of harmful disinfection by-products. Among the ANN algorithms examined, the Bayesian Regularization (BR) algorithm produced the best performance compared to the Levenberg–Marquardt (LM) and Scaled Conjugate Gradient (SCG) algorithms. The BR algorithm achieved a performance value of 0.0092 within only 10 epochs, whereas LM and SCG required over 1000 epochs to obtain comparable results. This demonstrates that BR has superior convergence speed, better generalization capability, and improved computational efficiency. Similar conclusions were drawn by Aljarah et al. (2020), who found that Bayesian Regularization effectively minimizes overfitting and improves ANN stability when handling noisy or limited environmental datasets. The faster convergence achieved by BR in this study further confirms its suitability for residual chlorine prediction applications. The Nash–Sutcliffe Efficiency (NSE) value of 0.9236, representing 92.36% efficiency, further validates the accuracy and reliability of the developed model. NSE values close to unity indicate strong agreement between observed and simulated data. According to Tiyasha et al. (2022), NSE values greater than 0.90 indicate excellent model performance and predictive strength in hydrological and environmental studies. Therefore, the high NSE value obtained in this study confirms that the ANN model is highly reliable for predicting residual chlorine concentrations and can serve as an effective decision-support tool for water quality monitoring and management in distribution systems.

4.2 EPANET

Observations through simulation of the network shown in Figures 11 to 13 and Table 3 indicated that there was no significant loss of chlorine around the dead-end in this network, leaving the residual chlorine within an acceptable value, hence changes in water quality at the dead ends are equally not significant. Figures 11-18 explained the utilization and performances of the disinfectant at selected dead-ends 32, 101, 498, 683 and 1150under flow velocity, demand and head variations. The decay processes for 24-hour (1-day), 144-hour (6-day) and 240-hour (10-day) indicated that the distribution system was able to maintain the required values during these critical periods. Also, nodes other than dead-ends that are extremely closer (Node 2) to and farther (Node 701) from the Tank were simulated to know their responses in the usage of chlorine. The results indicated that closer ones start utilizing the disinfectant right from the zeroth-hour while others lagged till the fourth-hour. The Maximum value of residual chlorine generated at the dead-end is 0.44mg/l which is within the 0.5mg/l recommended (Galvin, 2011)**.** This is an indication that the distribution system may not give room for the growth and regrowth of disease-causal agents**.**

5.0 CONCLUSION

The good results obtained from the 2013 to 2016 water quality data obtained from FCT Water Board, further prove that Artificial Neural Network model can effectively be utilized for modeling water quality data of shorter duration and attain high percentage of accuracy as could with other models that utilize long-term data in the dead-end water distribution network. This was seen when ANN models through various Algorithms were compared with the result of EPANET and Multiple Regression checking the squared correlation coefficient (R^2), the accuracy, reliability and performance. Its ability to model without knowing the history or relationships of parameters required to be part of the model’s equation, even while dealing with complex issues makes it superior over some of the existing models. The values obtained for the performance and efficiency, compete very well with existing models and equations with extra advantage of achieving them at a faster rate. Hence, the ANN model could safely be used for monitoring the performance level of disinfectants in the water distribution system in the FCT.

REFERENCES

Abokifa, A. A., Yang, Y. J., Lo, C. S., & Biswas, P. (2016). Investigating the role of biofilms in trihalomethane formation in water distribution systems with a multi-component model. Water Research, 104, 208–219.

Ahmed A., Yang Y. J, Lo C. S., Biswas P., 2016. “Water Quality Modelling In The Dead End Sections Of Drinking Water Distribution Networks,” *Water Research,* vol. 89, pp. 107-117.

Aparicio, J., and Li, Z. (2016). Water quality modeling in the dead end sections of drinking water distribution networks. Water Research, 89, 212–225.

Baker H. B., Aldridge C., Omer A., 2016 "Water: Availability and Use," Researchgate Publication/324226678, November 2016.

Barbeau B., Gauthier V., Julienne K. and Carriere A., 2005. "Dead-End Flusshing of a Distribution System: Short and Long Term Effecte on Water Quality," *Water Supply: Research an Technology-AQUA*, vol. 54, no. 6, pp. 371-383.

Buchberger S. G. and Lee Y., 1999 "Evidence Supporting the Poisson Pulse Hypothesis for Residential Water Demand," in *Proceeding CCWI International Conference Computer Control Water Ind.*

Bwire C., Onchiri R., Mburu N., 2015. *International Journal of Civil Engineering and Technology (IJCIET)*, vol. 6, no. 4, pp. 28-38.

CERT D.-I., 2015. *Industrial Control Systems- Cyber Emergency Response Team, Year in Review*, U.S. Department of Homeland Security.

Cunningham J. P. and Ghahramani Z., 2014. "Linear Dimensionality Reduction: Survey, Insights, and Generalizations," *Journal of Machine Learning Research*, vol. 16, pp. 2859-2900.

Danhassan M. A., Olasehinde P. I., Amadi A. N. and Yisa J., 2012 "Spatial and Temporal Distribution of Nitrate Pollution in Ground Water of Abuja, Nigeria".

Galvin, R., 2011."Eliminate Dead-End Water," pp. 20-21.

Grayman W. M., Clark R. M., 1990 "An Overview of Water Quality Modelling in a Distribution System," in *American Water Works Annual Conference Proceedings*, Ohio.

Horta R., 2007 "The City of Boca Raton: A Case Study In Water Utility Cybersecurity," *Journal -American Water* Works *Association*, vol. 99, no. 3, pp. 48-55.

Housh M. and Hoar Z., 2017 "Integrating physically based simulators with Event Detection 346 Systems: Multi-site detection," *Water Research*, vol. 110, pp. 180-191.

Housh M. and Ohar Z., 2018 "Model-based Approach for Cyber Physical Attack Detection in Water Distribution Systems," Water Research, vol. 139, pp. 132-143.

Izquierdo J., Lopez P. A., Martinez F. J. and Perez R., 2007 "Fault Detection In Water Supply Systems Using Hybrid (Theory And Data-Driven) modelling," *Mathematical and Computer Modelling*, vol. 46, no. 3-4, pp. 341-350.

Liu G., Zhang Y., Knibbe W. J., Feng C., Liu W., Medema G. and Van Der Meer W., 2017. "Potential Impact of Changing Supply Water Quality on Drinking Water Distribution," *Journal of Water Research*, pp. 135-148.

Pedersen K., 1990. "Biofilm development on stainless steel and pvc surface in drinking water," *Water Resources*, pp. 239-243.

Tzatchkov, A. A. Aldama and F. I. Arreguin, 2002 "Advection-Dispersion-Reaction Modelling in Water Distribution Networks," *Journal of Water Resources Planning Management*, pp. 334-342.

Wang H., Masters S., Edwards M. A., Falkinham III J. O., A. Pruden, 2014. "Effect Of Disinfectant, Water Age and Pipe Materials on Bacterial and Eukaryotic Community Structure in Drinking Water Biofilms," Environmental Science Technology, vol. 14, no. 18, pp. 1426-1435.

Zlatanovic L. V., Van Der Hoek J. P. and Vreeburg J. H., 2017. "An Experimental Study on the Influence of Water Stagnation and Temperature Change on Water Quality in a Full-Scale Domestic Drinking Water System," *Water Research,* Vols. 123,, pp. 761-772

To cite this article:
J.I. Ubah, Tochukwu Chibueze Ogwueleka, Ogunkuade O.G., and J.C. Chukwuneke. 2026. Water Quality Modeling Using Artificial Neural Network at The Dead-End Sections of Drinking Water Distribution System. 1(3): 66-89. https://journals.unizik.edu.ng/ujabe/