engineering proceedings

MDPI

Proceeding Paper

Operational Effects on Water Quality Evolution in Water Distribution Systems †

Laura González [1], Yesid Coy [1], Dominic L. Boccelli [2] and Juan Saldarriaga [3,*]

[1] Water Distribution and Sewerage Systems Research Center, Universidad de los Andes, Bogotá 111711, Colombia; lk.gonzalez@uniandes.edu.co (L.G.); b.coy@uniandes.edu.co (Y.C.)
[2] Department of Civil and Architectural Engineering and Mechanics, University of Arizona, Tucson, AZ 85721, USA; dboccelli@arizona.edu
[3] Civil and Environmental Engineering Department, Universidad de los Andes, Bogotá 111711, Colombia
* Correspondence: jsaldarr@uniandes.edu.co; Tel.: +57-1-339-49-49 (ext. 2805)
† Presented at the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry (WDSA/CCWI 2024), Ferrara, Italy, 1–4 July 2024.

Abstract: Water distribution systems (WDSs) are subject to operational changes due to variability in demands, the availability of flow rates, system maintenance and unexpected events. This study aims to assess the behavior of water quality in a distribution network associated with operational changes that the system typically undergoes. For this work, two black-box models were compared to predict chlorine concentration at different nodes in a small network and a large network. The model was trained with synthetic data from simulations through the EPANET-Python Toolkit. The results show that the black-box model can be implemented to predict water quality in real time.

Keywords: water quality; WDS; chlorine decay; water age; black-box models

check for updates

Citation: González, L.; Coy, Y.; Boccelli, D.L.; Saldarriaga, J. Operational Effects on Water Quality Evolution in Water Distribution Systems. *Eng. Proc.* **2024**, *69*, 30. https://doi.org/10.3390/engproc2024069030

Academic Editors: Stefano Alvisi, Marco Franchini, Valentina Marsili and Filippo Mazzoni

Published: 2 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

An accurate representation of the underlying water quality dynamics within a distribution system network model can assist with water quality management [1]. The purpose of this research is to evaluate the effect of operational changes on the evolution of chlorine. Although the preferred methods for modeling include physically based equations, an extensive exploration of data-based or black-box approaches has been performed as well [2]. As an alternative to physically based models, two black-box models can be implemented in order to predict residual chlorine levels. This document is divided into three sections. The methodology presents the structure of the algorithms used in this work. The results show the prediction graphs of selected nodes and a consolidation of the model's performance metrics. The conclusions present the most relevant findings and future work.

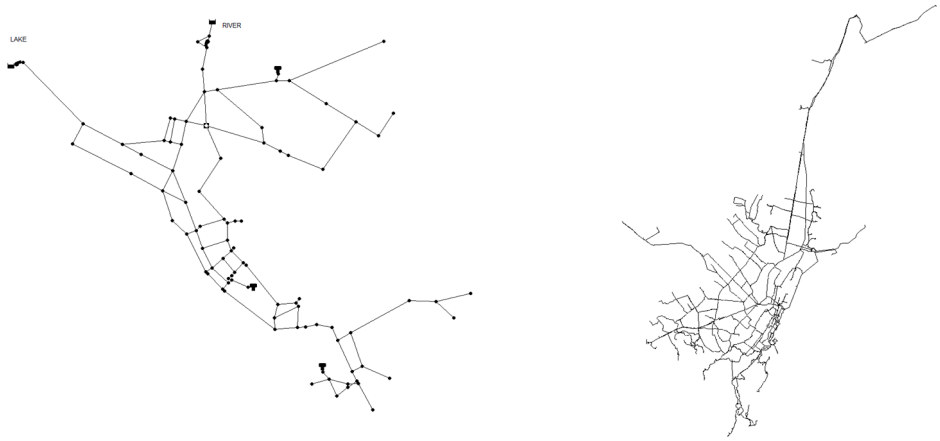
2. Methodology

This section presents the methodology applied for (a) the hydraulic and water quality models of the case studies, (b) the algorithms used for the prediction of chlorine concentrations and (c) model inputs and hyperparameter optimization as follows.

2.1. Hydraulic and Water Quality Simulation

The WDSs used as case studies in this work are EPANET Net3 and Bogotá trunk network (as shown in Figure 1). For Net 3, 5 operational and water quality scenarios were simulated considering different initial chlorine concentrations in the reservoirs, as well as different demand patterns, to simulate 5 months of historical data for developing two black-box models. In the second case study, due to the hydraulic complexity and because the network represents a real system, the model requires robust calibration. Thus, for the

second case study, only 11 days of data were generated for the preliminary evaluation. The EPANET-Python Toolkit [3] and EPANET-MSX were used for the simulations.



(**a**) (**b**)

Figure 1. (**a**) Net3; (**b**) Bogotá trunk network.

2.2. LSTM Algorithm

Long Short-Term Memory is a form of RNN with a more complex cell architecture for more accurately maintaining the memory of important correlations. This algorithm is good at extracting patterns in input feature space, where the input data span over long sequences [4]. The algorithm uses two separate paths, long-term memory data, i.e., the cell state (C_t), and short-term memory data, i.e., the hidden state (h_t) [5]. Each input vector x_t is processed sequentially by means of a sigmoid activation function (σ). The forget gate determines which information from the previous cell is discarded and which is preserved by means of a forget gate vector (f_t). The input gate determines which new information is stored in the cell state (C_t) by means of an input gate (i_t) and candidate cell state ($\widetilde{C}_t$) vectors. The cell state (C_t) is updated by combining the previous cell state with the candidate cell state ($\widetilde{C}_t$) weighted by the forget gate (f_t) and the input gate (i_t). The output gate (o_t) is produced to determine which information from the current cell state (C_t) should be output as the hidden state (h_t). Then, an output for the current time-step is produced.

2.3. CNN Algorithm

The Convolutional Neural Network (CNN) model consists of three layers: the input, hidden and output layers. The CNN is composed of a sequence of convolutional layers, where each neuron in a layer is connected to a small local region of neurons in the input data. This is executed by sliding a weight matrix, called a filter, over the input and the convolution computed at each point, which is referred to as the feature map, between the input and the filter [6]. The output feature map from the first layer is then given by convolving each filter w_h^1 for $h = 1, \ldots, M_1$ with the input as follows [7]:

$$a^1(i,h) = (w_h^1 * x)(i) = \sum_{j=-\infty}^{\infty} w_h^1(j)x(i-j)$$

where $w_h^1 \in \mathbb{R}^{1\text{x}k\text{x}1}$, and $a^1 \in \mathbb{R}^{1xN-k+1\text{x}M_1}$.

2.4. Inputs and Hyperparameters

- Inputs: Nodes were selected as sensors to take their hydraulic characteristics (pressure and demand) and residual chlorine. For Net 3, we used the same ones used by [8]; in the case of the Bogotá trunk network, we took some nodes chosen by [2]. Abrupt changes in both hydraulic and water quality patterns influence the accu-

racy and performance of the proposed models; then, seasonality was included as an additional variable.

- Hyperparameters: Bayesian optimization was used to create different models with a given parameter assignment, specifying a range of variation depending on the available computational cost. This optimization aims to minimize the final validation loss. As hyperparameters, the LSTM algorithm used N Units, Epochs and Batch Size; the CNN only used Epochs and Batch Size.

3. Results

Figure 2 shows the predicted performance of the test data set for Net 3 at the 1, 3, 6 and 12 h forecast using the CNN at node 143. As the prediction time increases, the algorithm tends to reduce performance; therefore, the accuracy of the prediction is lower. R^2 decreases from a value of 0.98 to 0.90. Table 1 presents the prediction of the other nodes.

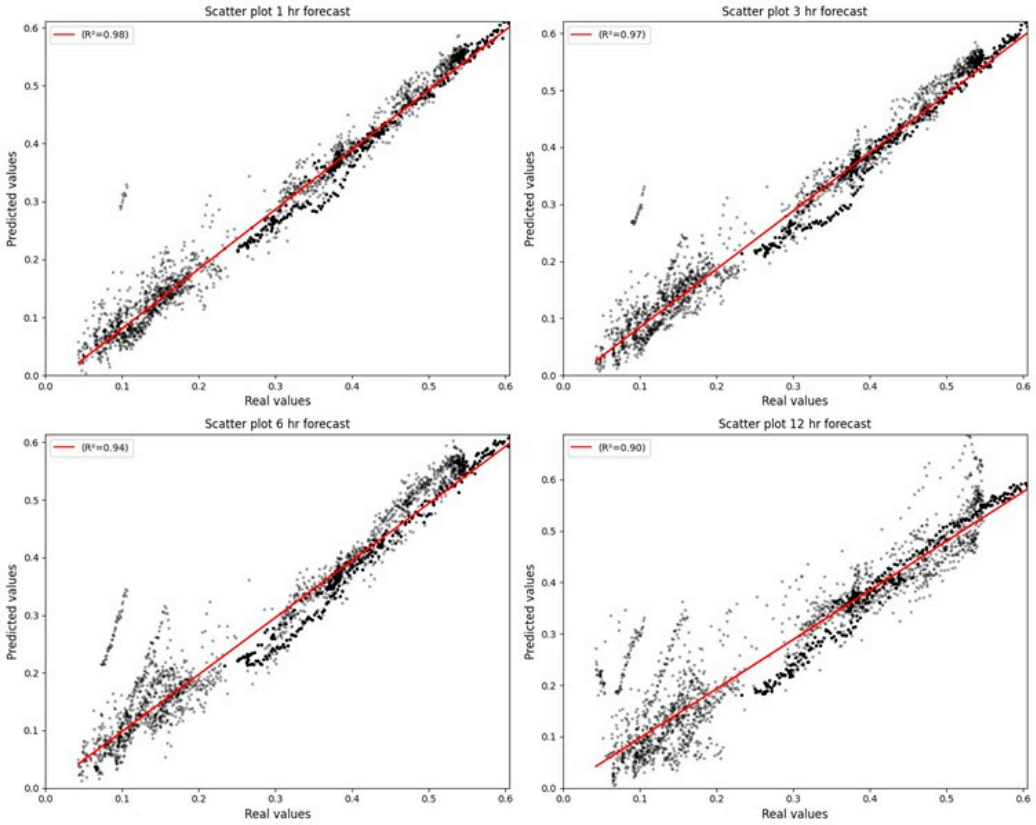


Figure 2. Scatter plot for 1, 3, 6 and 12 h forecast with CNN for node 143—Net3.

Table 1. Data set test performance at different prediction times for Net3.

Node	h-Forecast	LSTM R^2	CNN R^2	Node	h-Forecast	LSTM R^2	CNN R^2
121	1	0.98	0.97	159	1	0.98	0.98
	3	0.96	0.96		3	0.97	0.97
	6	0.94	0.94		6	0.96	0.95
	12	0.81	0.81		12	0.9	0.86
143	1	0.97	0.98	197	1	0.94	0.96
	3	0.96	0.97		3	0.94	0.94
	6	0.93	0.94		6	0.93	0.92
	12	0.89	0.90		12	0.77	0.75

Regarding the Bogotá trunk network, as mentioned, only 11 days of data were available for training. However, despite the decrease in time, the model still has very high performance, due to the constant patterns in this scenario, as indicated in Figure 3.

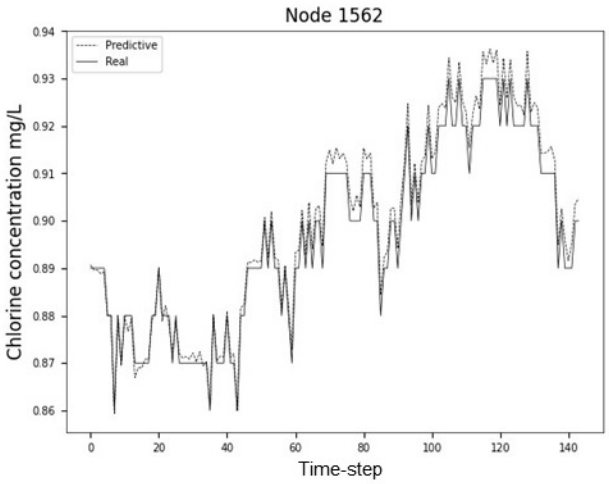


Figure 3. Chlorine concentration prediction at node 1562 of Bogotá trunk network with CNN.

4. Conclusions

This paper provides the preliminary results of research conducted on the prediction of the chlorine residual in WDSs. Two algorithms were evaluated, LSTM and the CNN, demonstrating high predictive capability, although evaluated with cyclic patterns. The Net3 network showed that the models were optimal-fitting; however, in this research, the location of the sensors was not evaluated; therefore, an analysis to identify the most predictive nodes would be recommended. The Bogotá trunk network obtained an R^2 of 0.99, as the patterns had a more stable behavior. Finally, the CNN requires shorter computational times; however, it predicts the first hours better, but by hour 12, LSTM showed higher accuracy. Future work can focus on generating synthetic data with changes closer to reality or with the implementation of real data.

Author Contributions: Conceptualization, L.G., D.L.B. and J.S.; methodology, L.G. and Y.C.; software, L.G. and Y.C.; validation, L.G., Y.C., D.L.B. and J.S.; formal analysis, L.G., Y.C., D.L.B. and J.S.; investigation, L.G., Y.C., D.L.B. and J.S.; data curation, L.G. and Y.C.; writing—original draft preparation, L.G. and Y.C.; writing—review and editing, D.L.B. and J.S.; visualization, L.G. and Y.C.; supervision, D.L.B. and J.S.; project administration, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy issues from water utilities.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Qin, T.; Boccelli, D.L. Estimating Distribution System Water Demands Using Markov Chain Monte Carlo. *J. Water Resour. Plan. Manag.* **2019**, *145*, 04019023. [CrossRef]
2. Enriquez, L.; González, L.; Saldarriaga, J.G. Application of Black-Box Models Based on Artificial Intelligence for the Prediction of Chlorine and TTHMs in the Trunk Network of Bogotá, Colombia. *J. Hydroinformatics* **2023**, *25*, 1396–1412. [CrossRef]
3. Kyriakou, M.S.; Demetriades, M.; Vrachimis, S.G.; Eliades, D.G.; Polycarpou, M.M. EPyT: An EPANET-Python Toolkit for Smart Water Network Simulations. *J. Open Source Softw.* **2023**, *8*, 5947. [CrossRef]
4. Muharemi, F.; Logofătu, D.; Leon, F. Machine Learning Approaches for Anomaly Detection of Water Quality on a Real-World Data Set. *J. Inf. Telecommun.* **2019**, *3*, 294–307. [CrossRef]
5. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

6. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-Term Water Quality Variable Prediction Using a Hybrid CNN–LSTM Deep Learning Model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [CrossRef]
7. Borovykh, A.; Bohte, S.; Oosterlee, C.W. Conditional Time Series Forecasting with Convolutional Neural Networks. *arXiv* **2017**, arXiv:1703.04691.
8. Yang, X.; Boccelli, D.L. Dynamic Water-Quality Simulation for Contaminant Intrusion Events in Distribution Systems. *J. Water Resour. Plan. Manag.* **2016**, *142*, 04016038. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.