water

MDPI

Article

A Digital Twin of a Water Distribution System by Using Graph Convolutional Networks for Pump Speed-Based State Estimation

Carlos A. Bonilla [1,2], Ariele Zanfei [3], Bruno Brentan [2,4], Idel Montalvo [2,5] and Joaquín Izquierdo [2,*]

1 Department of Civil, Environmental and Chemical Engineering, Faculty of Engineering and Architecture, University of Pamplona, Pamplona 543050, Colombia; carlos.bonilla@unipamplona.edu.co
2 Fluing-Institute for Multidisciplinary Mathematics, Universitat Politècnica de València, 46022 Valencia, Spain; brentan@ehr.ufmg.br (B.B.); imontalvo@ingeniousware.net (I.M.)
3 Faculty of Science and Technology, Free University of Bozen-Bolzano, Piazza Università 5, 39100 Bolzano, Italy; ariele.zanfei@natec.unibz.it
4 Hydraulic Engineering and Water Resources Department, School of Engineering, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil
5 Business Development, IngeniousWare GmbH, Jollystrasse 11, 76137 Karlsruhe, Germany
* Correspondence: jizquier@upv.es

Abstract: Water distribution system monitoring is currently carried out using advanced real-time control technologies to achieve a higher operational efficiency. Data analysis techniques can be implemented for condition estimation, which are crucial tools for managing, developing, and operating water networks using the monitored flow rate and pressure data at some network pipes and nodes. This work proposes a state estimation methodology that enables one to infer the hydraulic state of the operating speed of pumping systems from these pressure and flow measurements. The presented approach suggests using graph convolutional neural network theory linked to hydraulic models for generating a digital twin of the water system. It is validated on two benchmark hydraulic networks: the Patios-Villa del Rosario, Colombia, and the C-Town networks. The results show that the proposed model effectively predicts the state estimation in the two hydraulic networks used. The results of the evaluation metrics indicate low values of mean squared error and mean absolute error and high values of the coefficient of determination, reflecting high predictive ability and that the prediction results adequately represent the real data.

Keywords: graph convolutional neural networks; machine learning; state estimation; water distribution system; hydraulic modeling; digital twin

Citation: Bonilla, C.A.; Zanfei, A.; Brentan, B.; Montalvo, I.; Izquierdo, J. A Digital Twin of a Water Distribution System by Using Graph Convolutional Networks for Pump Speed-Based State Estimation. *Water* **2022**, *14*, 514. https://doi.org/10.3390/w14040514

Academic Editors: Gabriele Freni and Mariacrocetta Sambito

Received: 30 December 2021
Accepted: 7 February 2022
Published: 9 February 2022

1. Introduction

Water distribution systems (WDS) are currently made up of different hydraulic structures and complex elements interconnected to supply the requirements regarding the water demand of a community [1,2]. To better supply the users, pump stations should be controlled by on/off operations, and in many cases, using variable speed drivers (VSD). As pointed by [3], these devices can improve the efficiency of the system when well designed and controlled. The advances in information technology and computational devices lead to civil systems' digital transformation [4]. Due to the complexity of these systems, it is necessary to implement monitoring and control techniques that allow for optimal management of the networks [1–5]. These techniques/methodologies are typically based on supervisory control and data acquisition systems (SCADA), which provide real-time measurements taken in the field, the WDS, and are transmitted to a central control system [5,6].

The comprehensive monitoring of a WDS is a complex task. Nevertheless, it is one of the necessary actions for good practices in managing the system [7]. The water system management could be fully understood since control and monitoring devices could be installed. Nevertheless, abnormal situations, such as failures in the monitoring system, remote control, cyber-attacks etc., lead to a loss of knowledge of the real system state. For this reason, it is crucial to analyze the monitored data from an optimal set of sensors placed at critical nodes and pipes of the system within a calibrated hydraulic model to efficiently anticipate, detect and manage abnormal operating conditions [8,9]. These techniques are known as the state estimation (SE) of a WDS [10,11].

Some approaches to SE of WDSs have been developed in recent years, mainly focused on hydraulic parameter estimations (e.g., pressure and flow). Díaz et al. [12] implemented an algebraic method evaluating the impacts of measurements on inferring the hydraulic state of WDSs through an observability analysis. The authors presented a new methodology for sensor placement and uncertainty evaluation of various hydraulic variables. Combining the methods from Díaz et al. [10] and the algebraic analysis for clustering presented by [13], a set of research has been presented for improving various needs of SE, such as sensitivity analysis [14], calibration of a WDS [15], SE in changing WDS topologies [16,17], and leak detection using SE [18].

Fusco et al. [19] analyzed complementary SE approaches in WDSs with control devices and monitoring data, iteratively solving the unknown variable estimation problem using weighted least squares minimization and gradient methods. By performing residual analysis, it was possible to identify the changes in a WDS due to the opening or closing of control elements, such as pressure-reducing valves.

In addition to statistical or stochastic methodologies, new research approaches have been developed in the WDS field. In recent years, Hydroinformatics has integrated water sciences, data sciences, artificial intelligence, and social sciences [20]. The models proposed from data mining and machine learning have become popular in the last decade [21]. Some of the developed applications using machine learning are applied to optimal pressure management and district metered area design [22], leak detection in WDSs [23–27], water demand estimation [28–30], and detection of cyberattacks, physical attacks, and contamination in WDSs [31,32].

Recently, research based on graph convolutional neural networks (GCNs) [33–35] has been applied to different areas such as the classification of texts, images, and videos, social network analysis, biological sciences, material sciences, and engineering [36,37]. Similarly, GCNs have been implemented in the research of vehicular traffic prediction based on urban road networks [38] and traffic speed prediction by analyzing the influence of external factors such as weather conditions and the distribution of monitored points [39,40]. GCNs are successfully applied in several fields; however, dynamic systems can be better modeled if time-dependencies can be captured by the modeling method. For instance, temporal GCNs (T-GCNs) are proposed in the literature for passenger prediction [41], electric vehicle charging [42], or flood forecasting [43]. Many T-GCNs add recurrent layers based on long-short-term demand in the architecture of classical GCNs.

Moreover, in recent years, research in water networks has considered the possibility of implementing digital twins (DTs), a recent technique for modeling civil infrastructure systems [44]. In WDS, the implementation of DTs allows for the creation of hydraulic models that enable the development of the simulation of dynamic processes to improve the design of new infrastructures, reduce risk scenarios and optimize the management of the network and its elements [45]. In this case, new SE techniques, based on T-GCNs, contribute to the initial implementation of DTs in WDSs.

Although SE applications on WDSs have recently been explored, most works intend to estimate hydraulic parameters (e.g., pressure and flow). In general, these parameters can be obtained by well-calibrated hydraulic models. Nevertheless, these models should be fed with information on control devices, such as relative pump speeds. This paper proposes a SE algorithm using T-GCNs that enables one to infer the real operating pump speeds in the network based on the available pressure and flow rate measurements from the sensors installed in the WDS. The relative speeds are then injected into a hydraulic model, allowing for the estimation of pressure and flow for the entire water network. Monitoring data from two benchmark water networks; the Patios-Villa del Rosario network in Colombia and the

C-Town network are used to develop and evaluate the algorithm. The proposed algorithm results are robust and provide reliable performance that allows one to infer pump speeds in the analyzed networks. The results of the estimated pump speeds allow the hydraulic models to calculate the hydraulic state of the entire network.

The major advance in the study is the possibility it offers for producing a model that estimates the relative operating speed of the pumps, a parameter that is usually input data for classical hydraulic models. These results have two functions; the first one, which we exploit in this article, is the ability to build a DT; the second one is the possibility of detecting anomalies in a network.

2. Materials and Methods

In this section, we present the methodologies developed for this study. T-GCNs are presented as algorithms for estimating relative pump speeds. Used error evaluation metrics are then stated. Finally, the pressure and flow rate calculation from the estimated relative pump speed rates is described. The general diagram shown in Figure 1 illustrates the entire procedure developed in this article.

Figure 1. General diagram of the developed procedure.

2.1. Temporal-Graph Convolutional Neural Networks

A graph is a mathematical representation of a network. A generic graph $G = (V, E, A)$ is defined by a set V of vertices or nodes connected through the edges or lines in set E. The graph is represented by the adjacency matrix A, which embodies the graph structure and the related connections.

This study proposes to represent a WDS through a graph structure, where the vertices represent the network status on some defined points that are assumed to be connected based on their correlation. For the data set, the correlation matrix is obtained. Based on a threshold defined by the user, if the correlation coefficient between two variables is higher than the threshold, a link between those variables is created.

Given the proposed structure, this work aims to create a model that extracts the necessary and fundamental features from the graph to estimate the pump statuses from the available information in monitored nodes. This can be summarized as:

$$S = f(P, Q), \tag{1}$$

where S represents the pump speeds, P the pressures at the nodes, and Q the flow rates in the pipes. In other words, the graph is built with some nodes that represent the pressures, some nodes that represent the flow rates, and some nodes, the pump speeds. The task of the model is to identify the relationship of Equation (1) among these variables in order to estimate the unknown pump speeds of each of the pumps operating in a network.

Graph convolutional neural networks are computational models designed to process graph data and are useful in many applications [46]. The main idea of a GCN is to apply to graphs the same extraction features of conventional convolutional neural networks. However, the proposed problem has an essential dependency on time. Pressures, flow rates, and pump speeds are variables that change continuously in time. For this reason, it is proposed to model Equation (1) with a temporal graph convolutional neural network.

The T-GCN is designed to extract temporal and spatial information from structured graph data. The T-GCN is intended to be a sequence of graph convolutional layers with recurrent layers to perform this task. Specifically, the T-GCN designed for this study comprises three stages of layers. Firstly are the GCN layers, which receive the graph input.

It is worth noting that, in order to process the graph, the GCN layers receive, as input, the adjacency matrix of the graph (i.e., the correlation between measurements) and the nodes' features. Secondly, the output of the GCNs layers is processed by the recurrent layers. Finally, the output of the recurrent layers is further processed by a dense layer that provides the final output. This study implements a similar architecture of T-GCN as the one proposed by [39], where further details about the T-GCN model can be found. Furthermore, similar T-GCN models have also been successfully designed [47]. The Keras version 2.6.0, TensorFlow version 2.6.0, and Spektral version 1.0.8 [48,49] libraries of Python have been used to build the model practically.

2.2. Evaluation Parameters

The performance of the proposed T-GCN model for SE is evaluated using 3 statistical metrics: the root mean square error (RMSE) (Equation (2)), the mean absolute error (MAE) (Equation (3)), and the determination coefficient (r^2) (Equation (4)),

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{3}$$

$$r^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \hat{Y})^2}, \tag{4}$$

where the y_i are the measured pump speeds, $\hat{y}_i$ the predicted pump speeds data of sample i, for a dataset with N samples, and $\hat{Y}$ is the mean of the data set of $\hat{y}_i$.

Low RMSE and MAE values reflect a low prediction error and a good performance of the proposed model. In contrast, high r-squared (r^2) values (close to 1) represent better predictive ability, and the prediction results have a high similarity to the real data [38–40].

2.3. Pressure and Flow Calculation from the Estimated Relative Speed

After estimating the relative speeds for pumps, it is possible to obtain the pressure and flow rate of the entire network through a hydraulic model. The estimated relative pump speed is used as input to a hydraulic model built using Epanet 2.2. The simulations are carried out using the WNTR library with the Epanet hydraulic solver [50]. In general, for hydraulic models, the operational conditions of control devices are required as input for simulations. Nevertheless, these controls are unknown in many cases and require a deep calibration process. Using the estimated relative speeds, it is possible to calculate all of the pressures and flow rates of water network models based only on measured pressure and flow rate. The process of the proposed methodology is illustrated in Figure 2.

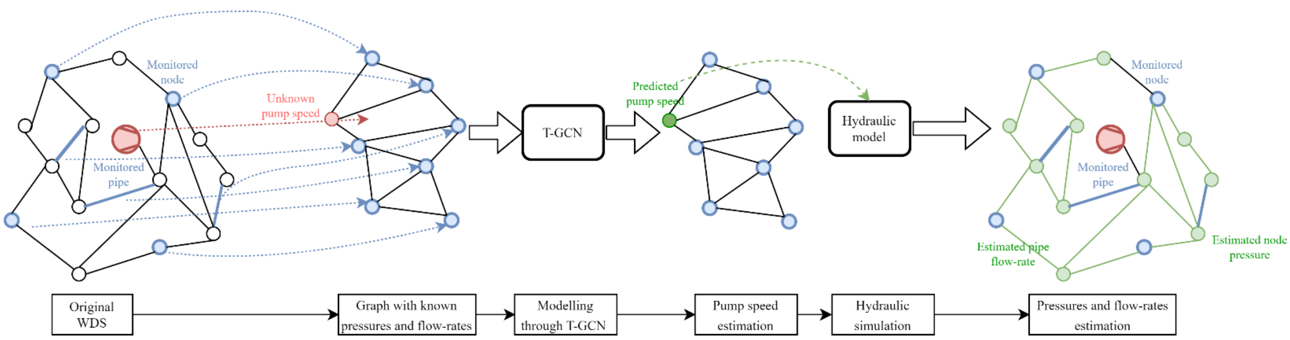


Figure 2. Methodology for pressure and flow rate estimation based on estimated relative pump speeds.

Figure 2 highlights the entire procedure proposed in this study. The first part shows the novel use of the T-GCN to estimate pump speeds in a specific location of a WDS (already presented in Section 2.1). Specifically, the process starting from the original WDS where the measurements are taken is highlighted. With the flow rates and pressure measurements from the network, the following step is the graph creation, which is the basis for the T-GCN model. Once the graph is built based on the correlation between its nodes, the graph is processed by the T-GCN, estimating the pump speed. Hereafter, the estimated pump speeds are used together with the already measured pressures and flow rates to calculate the flow rates in the non-monitored pipes and the pressures in the non-monitored nodes. The result is a complete estimation of the network status, which originates a digital twin of the base network hydraulic model. This resulting model is called the digital twin due to the capability of reproducing the operations found in the real network, even if no information about the control of pumps is provided.

Finally, pressures and flow rates on monitored elements are compared with their estimates provided by the hydraulic model to evaluate the methodology's performance. This evaluation is conducted based on the metrics presented in Section 2.2.

3. Case Studies

As previously mentioned, the two hydraulic networks used in this research are the Patios-Villa del Rosario network in Colombia and the C-Town network used in the well-known case study "The Battle of the Water Networks II". Usually, in a real water supply system, related to the variability and stochasticity of demand and the presence of anomalies in the network, such as leaks, make the estate estimation a challenging task. In this work, for both case studies, demand variability and stochasticity have been considered, as explained later. The two reference networks are described in the following sections.

3.1. Network 1: Patios Network-Villa del Rosario

The first case study is a water distribution system from northeastern Colombia (Norte de Santander), a network that supplies the municipalities of Villa del Rosario and Los Patios. The characteristics of the pumps are a flow rate of 9.39 L/s, a head loss of 57.06 m, and a characteristic curve formed by three operating points as follows: flow (0.0; 63.09; 100.94) L/s and head (57.06; 55.53; 35.72) m, respectively. The network comprises 67 pipes, 62 junction nodes, five reservoirs, and two pumps. The main characteristics of Network 1 include a total pipe length of 43.54 km, pipe material with a roughness coefficient of 0.0015 mm, and pipe diameters ranging from 75 to 762 mm. This network was created using the Darcy Weisbach loss equation.

3.2. Network 2: C-Town Network

The second case study is the C-Town water network. The network consists of 429 pipes, 388 junction nodes, seven tanks, 1 reservoir, 11 pumps, and 5 valves. The characteristics of each of the pumps in this network can be consulted in the inp. file, available online. Network 2 has a total pipe length of 56.73 km, pipe material with the Hazen-Williams roughness coefficient ranging from 60 to 140, and the pipe diameter ranging from 51 to 610 mm. The network is divided into five district metered areas (DMAs).

Table 1 summarizes the elements of each network described in Sections 3.1 and 3.2.

3.3. Data Set Generation for T-GCN Application

Since these networks do not have an associated data set of real data, a methodology of synthetic data generation is used for our T-GCN application. To generate the set of monitoring data used in this research, the Water Network Tool for Resilience (WNTR) [49] is used to model both networks.

Table 1. Main characteristics of the elements of the case study networks.

Parameter	Network 1	Network 2
Total length	43.54 km	56.73 km
Roughness Coefficient	0.0015 mm (Darcy-Weisbach)	60–140 (Hazen-Williams)
Pipe diameter	75–762 mm	51–610 mm
Number of pipes	67	429
Number of nodes	62	388
Number of reservoirs	5	1
Number of pumps	2	11
Number of tanks	0	7
Number of valves	0	5

WNTR is an open-source Python library based on EPANET that integrates hydraulic simulation, water quality, and several metric options for the comprehensive resilience assessment of a water network. It allows us to generate and modify the structure of water networks, simulates different analysis scenarios and response strategies, simulates and analyzes network pressure-dependent demands, analyzes water quality, calculates resilience metrics, and visualizes the results [47].

Hydraulic modeling is performed using an already-existing hourly demand pattern for each network. Considering the stochastic behavior of water demand, the data set generation randomly changes the base demand every time step between 90% and 110% of the original base demand. A random vector was created with the function "random.uniform"; this random change represents the existing variations between the hourly demands of each day of the week and the variations in the population regarding water consumption over time. The simulations for dataset generation are conducted for one year, with a hydraulic time step of one hour for both water networks. In the sequel, we refer to this data as real data.

From the generated hourly variation pattern, the relationship between the hourly pattern and the rotational speed of each pump in each network was applied. The maximum value of each pattern was taken as the maximum pump rotation speed, which is represented as 100%. The minimum value of each pattern was taken as the minimum pump speed, represented by 70% of the pump operating speed. This rule made it possible to generate a real speed pattern for each pump and hydraulically model the networks to obtain a data set of network pressures and flow rates.

For Network 1, Patios-Villa del Rosario network, five nodes are selected to simulate pressure monitoring, and four pipes are set for simulation of flow meters. For Network 2, C-Town water network, 12 nodes are selected as pressure sensors, and 10 pipes are selected as flow meters. The selected pipes and monitored nodes were randomly distributed in the network; no optimal sensor location criteria were used. Many drinking water utilities in Latin America do not yet employ methodologies for optimal sensor placement. Figure 3 shows the location of the monitoring sensors in each WDS highlighted as red dots and lines.

Each dataset is compiled with measured pressures and flows recorded at selected nodes and pipes, respectively. Each network has thus data for one year from 1 January to 31 December, recorded hourly. A data split was performed from this data set, taking 50% of the data for the training of the T-GCN model and the other 50% for the validation procedure.

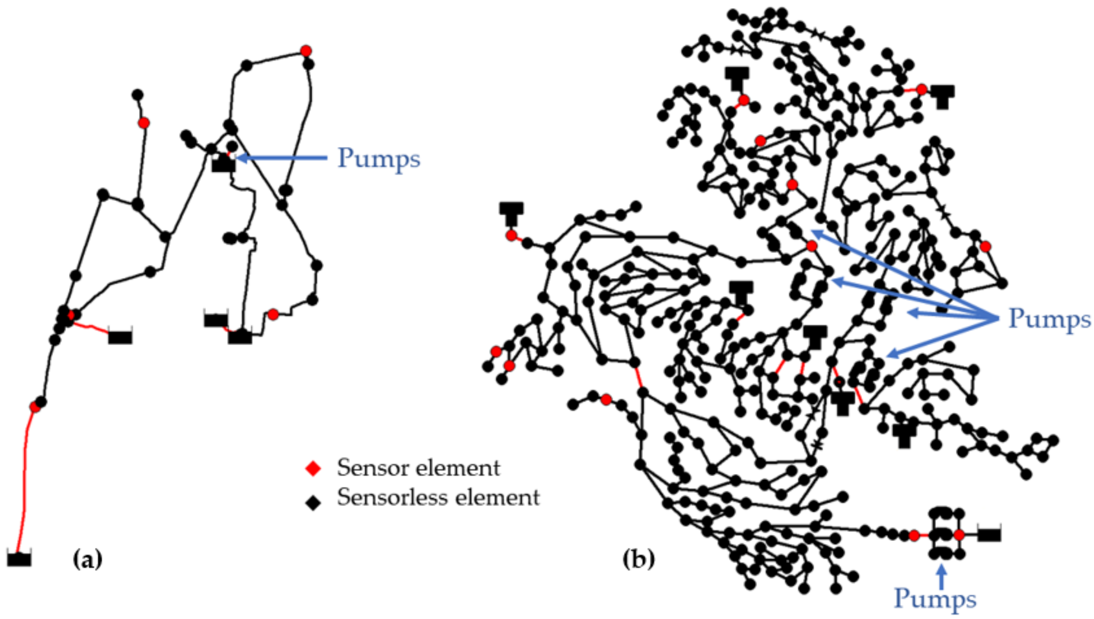


Figure 3. Topology and spatial distribution of sensors in (**a**) Patios-Villa del Rosario (Network 1), (**b**) C-Town (Network 2).

4. Results

4.1. T-GCN Evaluation for Pump Speed Estimation

The T-GCN model proposed is applied to Network 1 and Network 2. The results for each case study are shown and discussed in this section.

Considering the pressure and flow rate monitored data of Network 1, a first data processing is required to get the graph that will be the input structure to T-GCN. This graph is built based on data correlation, as commented in Section 2.1. For Network 1, the dataset turns out to be fully connected. Furthermore, the pressure head and the flow rate positively correlate with the relative pump speed. This is expected since, by reducing the relative speed of pumps, the operational pump curve is shifted to a new curve with a lower hydraulic head for a specific flow rate. The algorithm is tested for the rest of the data. Figure 4 shows the comparison of the estimated relative pump speed and the real values of the relative speed.

Based on Figure 4, it is possible to observe that T-GCN achieves a reliable performance on estimating relative speed. The estimation accuracy during the evening and night (low relative speed values) is better than that during the day. This is related to the operational rules for pump speed based on delivered flow in the network. Delivered flow is strongly correlated to the water demand, leading to more uncertainties during the day's peaks. For better exploring the results of the T-GCN, estimated and real data are plotted in scatter graphs, shown in Figure 5. The errors for low relative speeds are lower than errors for higher speeds. For relative speeds lower than 0.85, the average error is lower than 0.05. Nevertheless, for relative speeds higher than 0.85, the error can achieve values around 0.1.

For quantitively characterizing the performance of the T-GCN on estimating the relative speed of pumps, the indicators presented in Section 2.2 are calculated. The *RMSE* is equal to 0.015 and the *MAE* is equal to 0.011; the results of these two metrics reflect a high predictive capacity. The r^2 obtained for Network 1 is 0.972, a high value, showing, in general, the high accuracy of the model.

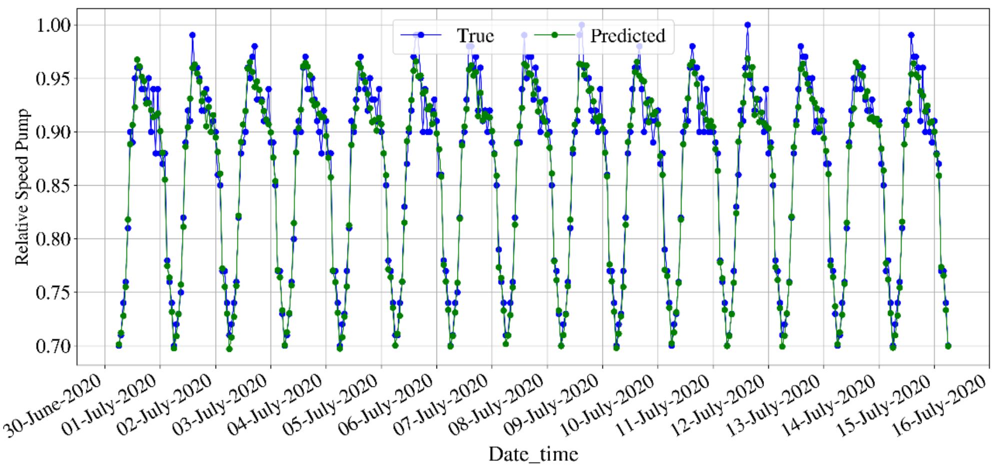


Figure 4. Comparison between estimated and real relative pump speed for Patios-Villa del Rosario WDS.

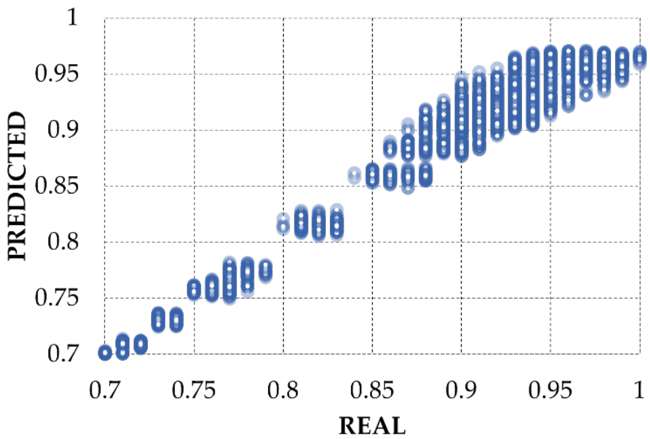


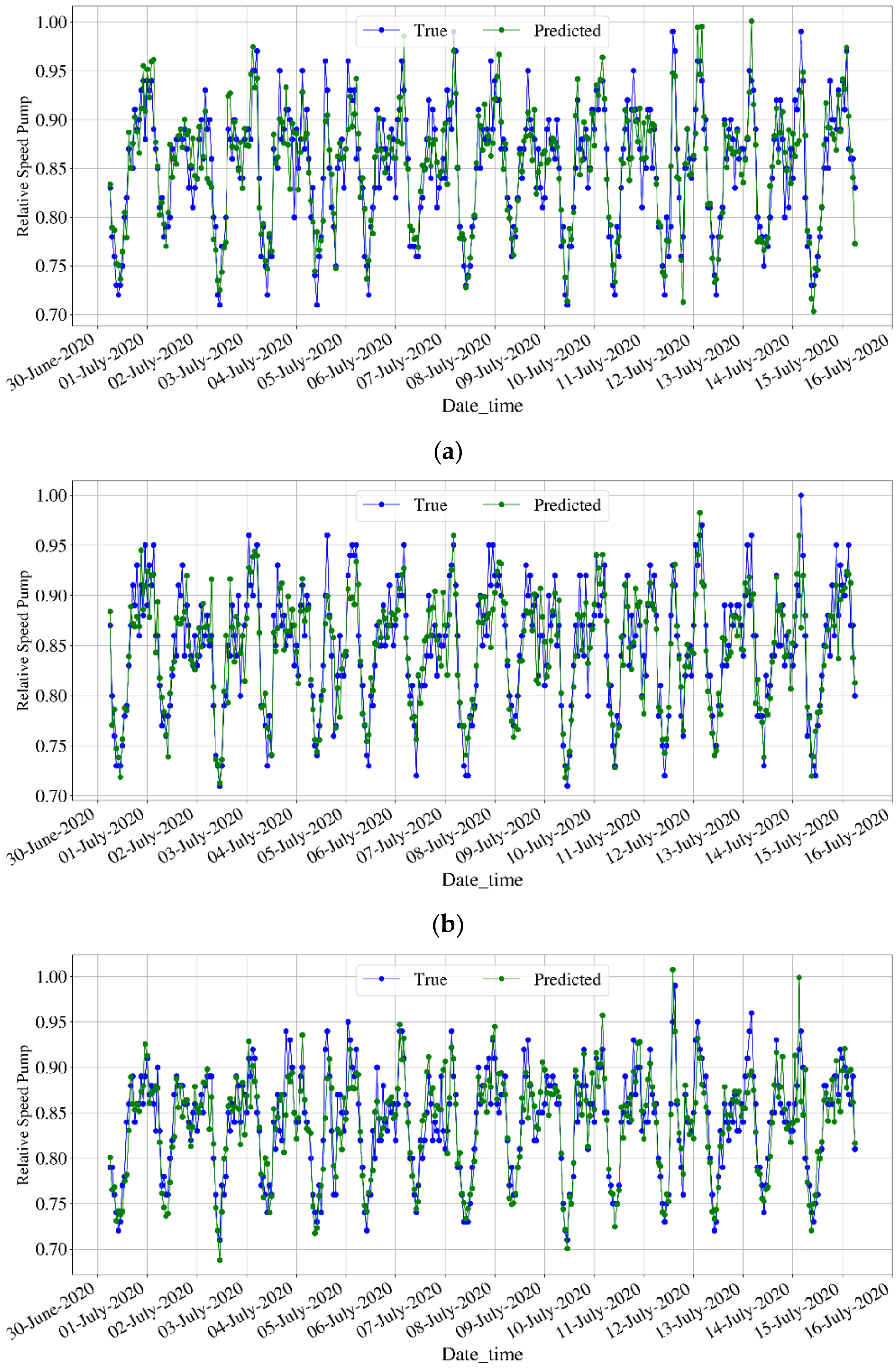
Figure 5. Scatter plot of estimated and real values of operational speed rotation for Network 1.

To implement the proposed model in Network 2, a T-GCN model is trained on each pump to estimate the operating speed of the element. As mentioned in 3.2., the network has 5 hydraulic zones, and the pump speeds are estimated for each zone. Due to the sub-division of the network, five models are developed, one for each zone. The evaluated models are named as follows: Model 1-PU2, Model 2-PU4, Model 3-PU6, Model 4-PU8, and Model 5-PU10.

Figure 6 shows the graphical comparison of estimated values and the real values of the operating speed of each driving equipment of Network 2. A high predictive capability of the GCN model is observed. A similar trend can be observed between the real relative speed pattern and the estimated speed by the T-GCN models. This suggests the model's validity in prediction and the ability to estimate pump speeds.

Scatter plots of the results for Network 2 are presented in Figure 7 and again highlight an error trend with the speed values. In fact, for low-speed values, the errors are around 5% of the expected values, while for higher speeds, the errors can reach 15%.

In network 2, each model used is independently analyzed, and the $RMSE$ results range between 0.025 and 0.027, and the MAE values between 0.020 and 0.021. The r^2 values obtained range from 0.799 to 0.815, indicating that the prediction results adequately represent the real data. The values obtained, which allows one to assess the T-GCN's performance in the different analyses performed, are shown in Table 2.

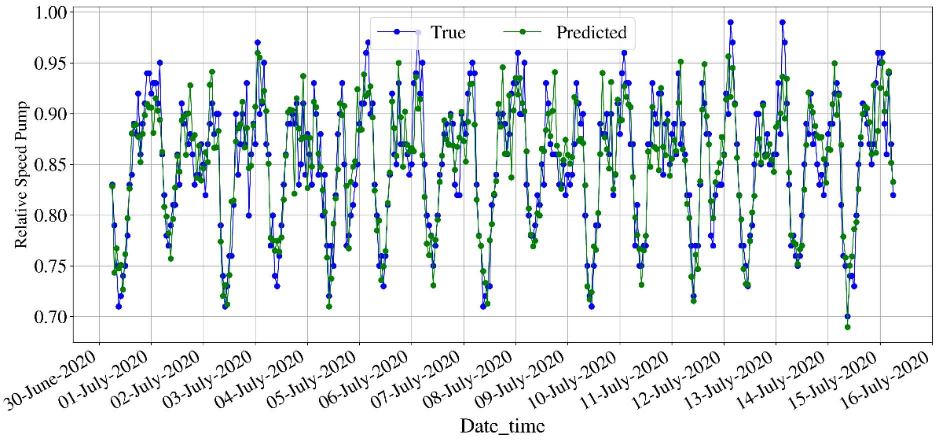


(**a**)

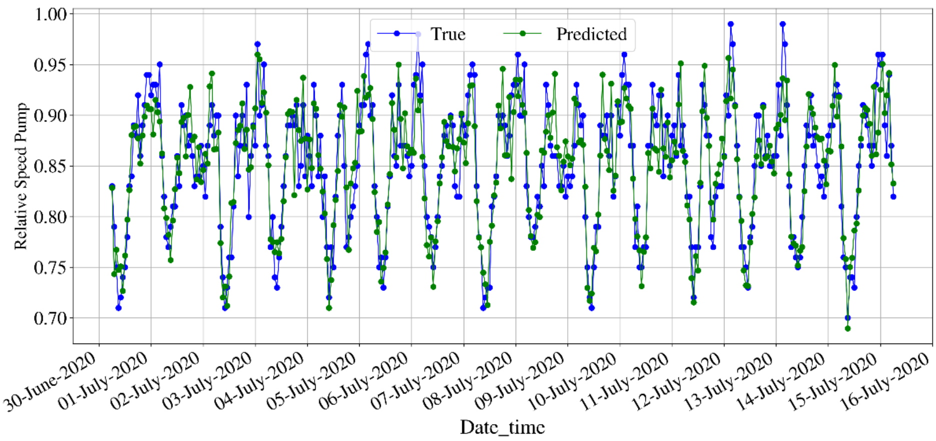
(**b**)

(**c**)

Figure 6. *Cont.*



(**d**)



(**e**)

Figure 6. Comparison curve of model predicted values and real values of the relative pump speeds for Network 2. (**a**) PU2, (**b**) PU4, (**c**) PU6, (**d**) PU8, (**e**) PU10.

4.2. Estimation of Pressure and Flowrate Using Estimated Pump Speeds

The relative speed pattern generated by the T-GCN is used in the hydraulic model to estimate pressure and flow for the entire water network. New hydraulic modeling of the two networks is performed using the WNTR tool, and the pressures at the nodes and the flows in the pipes are calculated. These results are compared with the real values calculated based on the real relative pump speeds. The same demand pattern is used for this analysis for the real and estimated relative pump speeds. The monitored nodes and pipes indicated in Figure 3 are used for comparison results.

Table 3 shows the error metrics ($RMSE$, MAE, and r^2) calculated for a set of nodes of Network 1 and Network 2, and pipes for Network 2. Flows in Network 1 do not change due to new pump speeds because of the topology of the water network and the hydraulic model. In effect, since Network 1 has no tanks and the hydraulic model is built as demand-driven, the flows in the pipes do not change due to the relative pump speed differences. For this reason, only pressures at some nodes of Network 1 are used in the evaluation process.

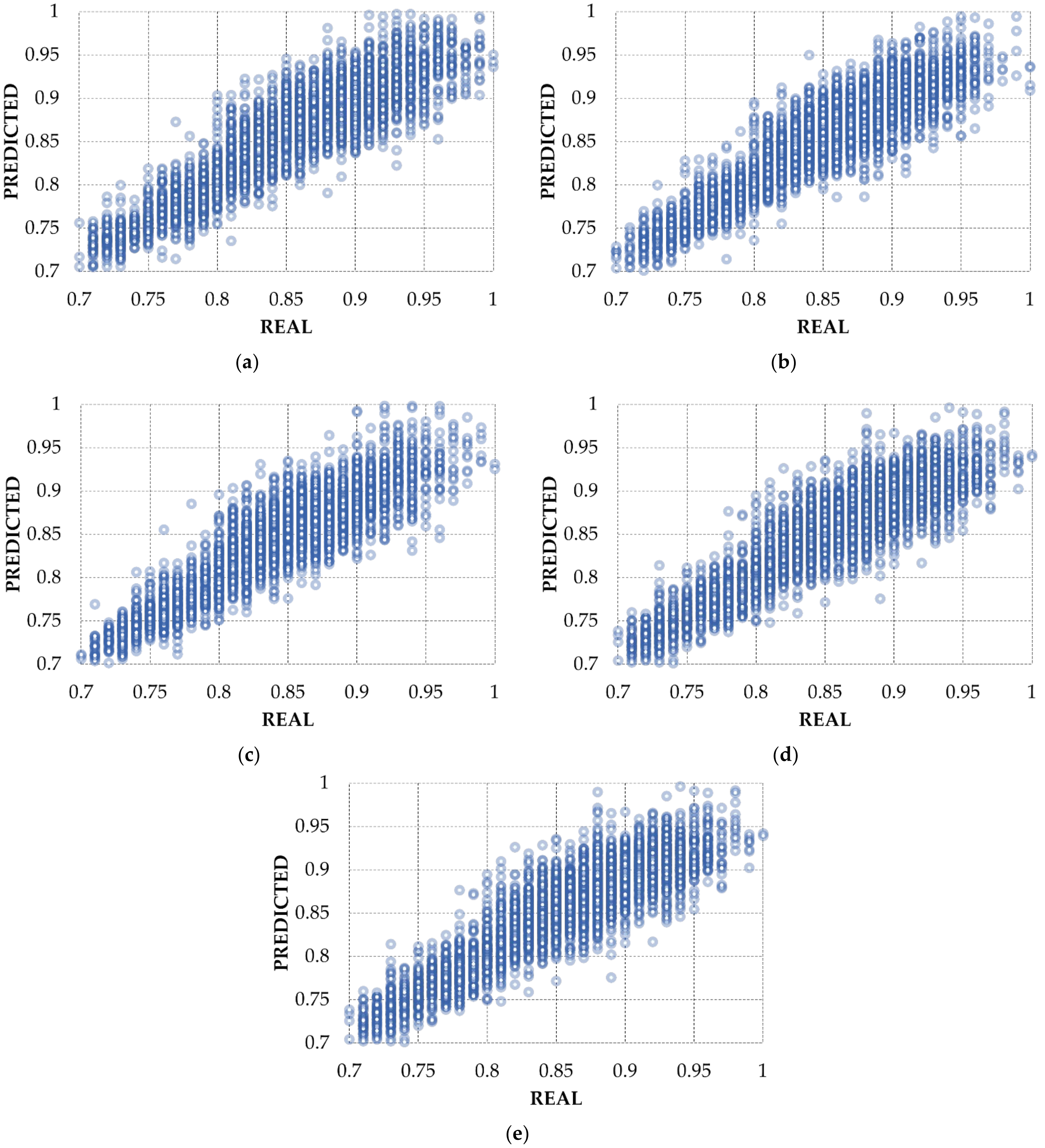


Figure 7. Scatter plots of predicted and real values of relative pump speeds for Network2. (**a**) PU2 (**b**) PU4 (**c**) PU6 (**d**) PU8 (**e**) PU10.

Table 2. Performance evaluation parameters of the T-GCN model.

Parameter	PU2	PU4	PU6	PU8	PU10
RMSE	0.028	0.026	0.026	0.027	0.027
MAE	0.021	0.020	0.020	0.021	0.021
r^2	0.801	0.815	0.799	0.802	0.802

Table 3. Evaluation parameters of the T-GCN model in pressure and flow rate.

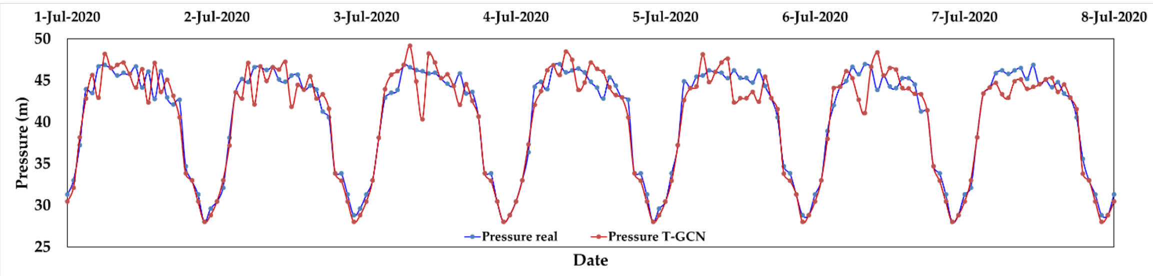
	Network 1		Network 2		
Parameter	Node 14	Node J269	Node J256	Pipe p397	Pipe J379
RMSE	1.8	2.9	1.7	1.8	9.7
MAE	1.1	2.0	1.3	1.2	2.6
r^2	0.923	0.448	0.592	0.949	0.632

Regarding Network 2, *RMSE* for nodal pressure varies from 1.7 m to 2.9 m, while, for pipes flow, ranges from 1.8 L/s to 9.7 L/s. Some locations' estimated flow rates and pressures show more significant differences from the original model values. These differences can mainly be attributed to nodes and pipes close to pump stations. These latter elements are more affected by the relative pump speed variation. A slight error in the speed prediction can cause an essential difference in the digital twin model's resulting pressures and flow rates.

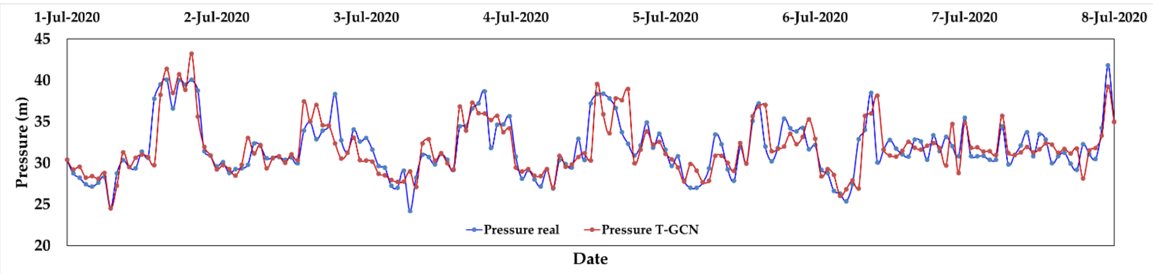
Figure 8 shows the comparison between the estimated pressure using a relative pump speed generated by the T-GCN and the real relative pump speed. It is observed that the predicted pressure variation at each node shows a similar behavior trend to the real pressure, which suggests that the T-GCN model is very effective in predicting SE in WDSs. For node J256, the relative pump speeds' errors impact pressure estimation more, although the general trend remarkably matches the real values.

An analysis of the various graphs between the real flow rates of the C-Town network and the predicted flow rates is also performed. Figure 9 compares the flow calculated by the hydraulic model using real and estimated relative pump speeds. Figure 9a shows that the variation between predicted and real flow rates in the p397 pipe is minimal. In contrast, Figure 9b shows a more significant difference in the p379 pipeline when using the T-GCN model, but with results still close to the real ones.

In summary, the T-GCN model yielded flow and pressure prediction values that demonstrate efficiency and robustness in SE. Consequently, a new methodology for SE in WDS could be implemented using this proposal.

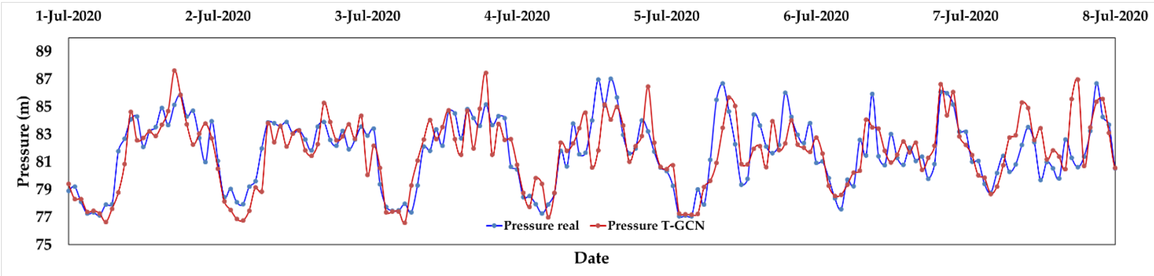


(**a**) Pressure at node 14 (Network 1) calculated using real and estimated relative pump speed



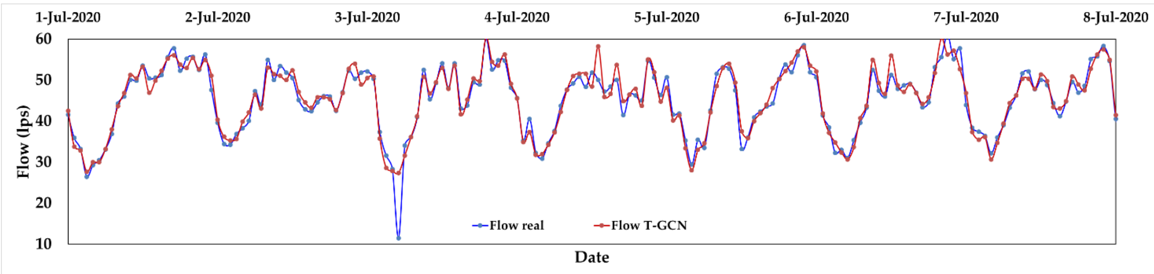
(**b**) Pressure at node J269 (Network 2) calculated using real and estimated relative pump speed

Figure 8. *Cont.*

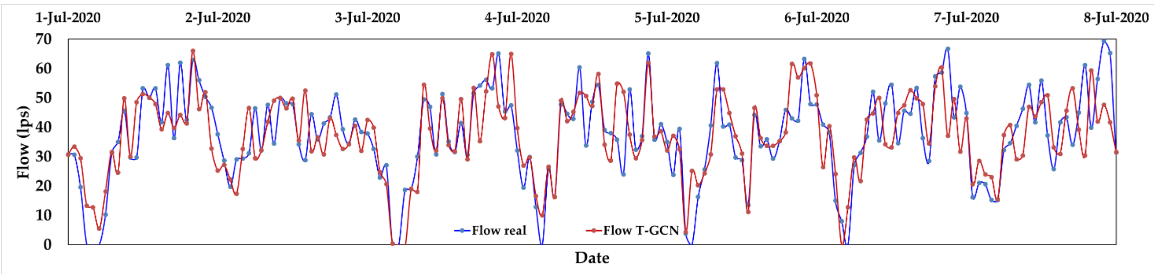


(**c**) Pressure at node J256 (Network 2) calculated using real and estimated relative pump speed

Figure 8. Comparison between estimated pressure based on real and estimated relative pump speed.



(**a**) Flow in pipe p397 (Network 2) calculated using real and estimated relative pump speed



(**b**) Flow in pipe p379 (Network 2) calculated using real and estimated relative pump speed

Figure 9. Comparison between estimated flow based on real and estimated relative pump speed.

5. Conclusions

This paper presents an SE methodology that allows one to infer the hydraulic state of a WDS based on available monitoring measurements and uses graph convolutional network theory. The proposed T-GCN model is adopted to predict the pump speeds of two WDSs from the monitoring data of pressure at some nodes and flow rate in certain pipelines. Afterwards, it is proposed that one should jointly use the estimated pump speed with the available pressure and flow rates to estimate the pressures and flow rates in the non-monitored elements. This latter operation allows one to build a digital twin model of the WDS.

The proposed procedure is tested using monitoring data from the Patios-Villa del Rosario hydraulic network, Colombia, and the C-Town network. The validation of the model in the networks yielded solid results with high performance in prediction capacity. The representation of the data obtained using scatter plots shows a good correlation and a perfect fit of the estimated data versus the prediction model data in each of the analyses performed in the two case studies.

The results of the evaluation metrics produce $RMSE$ values between 0.015 and 0.028, MAE values between 0.011 and 0.021, and r^2 values between 0.799 and 0.972. Considering both networks, the best results are obtained for Network 1. Network 1 is simpler than Network 2, where several control devices and tanks drive the system. The values derived from the evaluation metrics can be very favorably interpreted and reflect a high predictive and prognostic accuracy. In turn, they demonstrate the validity of the T-GCN model for SE in WDSs.

The main limitation of this study for application in a real network is the need to have monitored data of the relative operating speed of the pumps in a WDS, and it is important to clarify that these data are not always available since not all companies that operate distribution networks measure this parameter.

The proposed approach and the obtained results open a new path for research ideas in the management, development, and operation of water networks through the applicability of T-GCNs to SE. The model developed has the practical utility of contributing to the creation and application of digital twins in WDS, which becomes a valuable tool for detecting anomalies in WDSs.

Author Contributions: Conceptualization and methodology, B.B. and I.M.; software and data analysis, C.A.B. and A.Z.; writing and preparation of the original draft, C.A.B., A.Z., B.B. and J.I.; visualization, I.M.; supervision, J.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Herrera, M.; Ayala-Cabrera, D.; Izquierdo, J.; Montalvo, I. Smart Data Analysis for Smart Water Networks. In Proceedings of the Congress on Numerical Methods in Engineering—CMN 2017, Valencia, Spain, 3–5 July 2017; pp. 1665–1677.
2. Herrera, M.; García-Díaz, J.; Izquierdo, J.; Pérez-García, R. Municipal Water Demand Forecasting: Tools for Intervention Time Series. *Stoch. Anal. Appl.* **2011**, *29*, 998–1007. [CrossRef]
3. Mala-Jetmarova, H.; Sultanova, N.; Savic, D. Lost in optimisation of water distribution systems? A literature review of system operation. *Environ. Model. Softw.* **2017**, *93*, 209–254. [CrossRef]
4. Makropoulos, C.; Savić, D. Urban Hydroinformatics: Past, Present and Future. *Water* **2019**, *11*, 1959. [CrossRef]
5. Gutiérrez-Pérez, J. Monitorización, detección y estimación de estados de fallo en la calidad del agua de redes de distribución urbanas. Ph.D. Thesis, Universitat Politècnica de València, Valencia, Spain, 2021.
6. Taormina, R.; Galelli, S.; Tippenhauer, N.; Salomons, E.; Ostfeld, A. Characterizing Cyber-Physical Attacks on Water Distribution Systems. *J. Water Resour. Plan. Manag.* **2017**, *143*, 04017009. [CrossRef]
7. Tshehla, K.; Hamam, Y.; Abu-Mahfouz, A. State estimation in water distribution network: A review. In Proceedings of the 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 July 2017; pp. 1247–1252.
8. Righetti, M.; Bort, C.; Bottazzi, M.; Menapace, A.; Zanfe, A. Optimal selection and monitoring of nodes aimed at supporting leakages identification in WDS. *Water* **2019**, *11*, 629. [CrossRef]
9. Zanfei, A.; Menapace, A.; Santopietro, S.; Righetti, M. Calibration Procedure for Water Distribution Systems: Comparison among Hydraulic Models. *Water* **2020**, *12*, 1421. [CrossRef]
10. Díaz, S.; González, J.; Mínguez, R. Observability Analysis in Water Transport Networks: Algebraic Approach. *J. Water Resour. Plan. Manag.* **2016**, *142*, 04015071. [CrossRef]
11. Letting, L.; Hamam, Y.; Abu-Mahfouz, A. Estimation of water demand in water distribution systems using particle swarm optimization. *Water* **2017**, *9*, 593. [CrossRef]
12. Díaz-García, S. Comprehensive Approach for On-Line Monitoring Water Distribution Systems via State Estimation Related Techniques. Ph.D. Thesis, Universidad de Castilla-La Mancha, Ciudad Real, Spain, 2017.
13. Díaz, S.; Mínguez, R.; González, J. Aproximación estocástica al análisis de observabilidad en redes de abastecimiento de agua. *Ing. Del Agua* **2016**, *20*, 139. [CrossRef]
14. Díaz, S.; Mínguez, R.; González, J.; Savic, D. Explicit Expressions for State Estimation Sensitivity Analysis in Water Systems. *J. Water Resour. Plan. Manag.* **2018**, *144*, 06018001. [CrossRef]
15. Díaz, S.; Mínguez, R.; González, J. Calibration via Multi-period State Estimation in Water Distribution Systems. *Water Resour. Manag.* **2017**, *31*, 4801–4819. [CrossRef]
16. Díaz, S.; Mínguez, R.; González, J. Topological State Estimation in Water Distribution Systems: Mixed-Integer Quadratic Programming Approach. *J. Water Resour. Plan. Manag.* **2018**, *144*, 04018026. [CrossRef]

17. Díaz, S.; Mínguez, R.; González, J. Topological Observability Analysis in Water Distribution Systems. *J. Water Resour. Plan. Manag.* **2017**, *143*, 06017001. [CrossRef]
18. Díaz, S.; Mínguez, R.; González, J. Probabilistic leak detectability assessment via state estimation in water transport networks. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2111–2128. [CrossRef]
19. Fusco, F.; Arandia, E. State Estimation for Water Distribution Networks in the Presence of Control Devices with Switching Behavior. *Procedia Eng.* **2017**, *186*, 592–600. [CrossRef]
20. Holz, K.; Cunge, J.; Lehfeldt, R.; Savic, D. Hydroinformatics Vision 2011. In *Advances in Hydroinformatics*; Springer: Singapore, 2014; pp. 545–560.
21. Rozos, E. Machine learning, urbanwater resources management and operating policy. *Resources* **2019**, *8*, 173. [CrossRef]
22. Novarini, B.; Brentan, B.M.; Meirelles, G.; Luvizotto-Junior, E. Optimal pressure management in water distribution networks through district metered area creation based on machine learning. *Rev. Bras. Recur. Hidricos.* **2019**, *24*, 1–11. [CrossRef]
23. Mounce, S.R.; Machell, J. Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water J.* **2006**, *3*, 21–31. [CrossRef]
24. Capelo, D.; Brentan, M.; Monteiro, B.; Covas, L. Near–Real Time Burst Location and Sizing in Water Distribution Systems Using Artificial Neural Networks. *Water* **2021**, *13*, 1841. [CrossRef]
25. Manzi, D.; Brentan, B.; Meirelles, G.; Izquierdo, J.; Luvizotto, E. Pattern recognition and clustering of transient pressure signals for burst location. *Water* **2019**, *11*, 2279. [CrossRef]
26. Bohorquez, J.; Alexander, B.; Simpson, A.R.; Lambert, M.F. Leak Detection and Topology Identification in Pipelines Using Fluid Transients and Artificial Neural Networks. *J. Water Resour. Plan. Manag.* **2020**, *146*, 04020040. [CrossRef]
27. Bohorquez, J.; Simpson, A.R.; Lambert, M.F.; Alexander, B. Merging Fluid Transient Waves and Artificial Neural Networks for Burst Detection and Identification in Pipelines. *J. Water Resour. Plan. Manag.* **2021**, *147*, 04020097. [CrossRef]
28. Brentan, B.M.; Luvizotto, E.; Herrera, M.; Izquierdo, J.; Pérez-García, R. Hybrid regression model for near real-time urban water demand forecasting. *J. Comput. Appl. Math.* **2017**, *309*, 532–541. [CrossRef]
29. Bennett, C.; Stewart, R.A.; Beal, C.D. ANN-based residential water end-use demand forecasting model. *Expert Syst. Appl.* **2013**, *40*, 1014–1023. [CrossRef]
30. Msiza, I.S.; Nelwamondo, F.V.; Marwala, T. Artificial neural networks and support vector machines for water demand time series forecasting. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montreal, QC, Canada, 7–10 October 2007; pp. 638–643.
31. Tsiami, L.; Makropoulos, C. Cyber—Physical attack detection in water distribution systems with temporal graph convolutional neural networks. *Water* **2021**, *13*, 1247. [CrossRef]
32. Zhou, Y.; Jiang, J.; Qian, K.; Ding, Y.; Yang, S.H.; He, L. Graph convolutional networks based contamination source identification across water distribution networks. *Process Saf. Environ. Prot.* **2021**, *155*, 317–324. [CrossRef]
33. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In Proceedings of the 5th International Conference on Learning Representations, Conference Track Proceedings, Toulon, France, 24–26 April 2017; pp. 1–14.
34. Zhou, J. Graph neural networks: A review of methods and applications. *AI Open* **2020**, *1*, 57–81. [CrossRef]
35. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE Trans. Neural Netw.* **2009**, *20*, 61–80. [CrossRef]
36. Singh, R.; Bathla, S.; Meel, P. State-of-the-Art Applications of Graph Convolutional Neural Networks. In Proceedings of the 6th International Conference on Recent Trends in Computing, Delhi, India, 3–4 July 2020; pp. 107–115.
37. Michaël-Defferrard, P.V.; Xavier, B. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. *Adv. Neural Inf. Process. Syst.* **2016**, *59*, 395–398.
38. Bai, J. A3t-gcn: Attention temporal graph convolutional network for traffic forecasting. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 485. [CrossRef]
39. Zhao, L. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3848–3858. [CrossRef]
40. Zhu, J.; Wang, Q.; Tao, C.; Deng, H.; Zhao, L.; Li, H. AST-GCN: Attribute-augmented spatiotemporal graph convolutional network for traffic forecasting. *IEEE Access* **2021**, *9*, 35973–35983. [CrossRef]
41. Bai, L.; Yao, L.; Wang, X.; Li, C.; Zhang, X. Deep spatial–temporal sequence modeling for multi-step passenger demand prediction. *Future Gener. Comput. Syst.* **2021**, *121*, 25–34. [CrossRef]
42. Hüttel, F.; Peled, I.; Rodrigues, F.; Pereira, F. *Deep Spatio-Temporal Forecasting of Electrical Vehicle Charging Demand*; Cornell University: New York, NY, USA, 2021; pp. 1–6.
43. Ding, Y.; Zhu, Y.; Feng, J.; Zhang, P.; Cheng, Z. Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing* **2020**, *403*, 348–359. [CrossRef]
44. Curl, J.; Nading, T.; Hegger, K.; Barhoumi, A.; Smoczynski, M. Digital twins: The next generation of water treatment technology. *J.-Am. Water Work. Assoc.* **2019**, *111*, 44–50. [CrossRef]
45. Callcut, M.; Cerceau, J.; Varga, L.; McMillan, L. Digital Twins in Civil Infrastructure Systems. *Sustainability* **2021**, *13*, 11549. [CrossRef]
46. Bronstein, M.M.; Bruna, J.; LeCun, Y.; Szlam, A.; Vandergheynst, P. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Process. Mag.* **2017**, *34*, 18–42. [CrossRef]

47. Ma, Z.; Mei, G.; Prezioso, E.; Zhang, Z.; Xu, N. A deep learning approach using graph convolutional networks for slope deformation prediction based on time-series displacement data. *Neural Comput. Appl.* **2021**, *33*, 14441–14457. [CrossRef]
48. Grattarola, D.; Alippi, C. Graph Neural Networks in TensorFlow and Keras with Spektral [Application Notes]. *IEEE Comput. Intell. Mag.* **2021**, *16*, 99–106. [CrossRef]
49. Chollet, F. *Deep learning with Python*, 2nd ed.; Manning Pubkications: New York, NY, USA, 2021; p. 504.
50. Klise, K.; Murray, R.; Haxton, T. An overview of the Water Network Tool for Resilience (WNTR). In Proceedings of the 1st International WDSA/CCWI Joint Conference, Kingston, ON, Canada, 23–25 July 2018.