International Journal of Industrial Engineering & Production Research September 2020 Vol. 31, No. 3: 423-433
DOI: 10.22068/ijiepr.31.3.423

RESEARCH PAPER

A Novel Data-driven and Feature-based Forecasting Framework for Wastewater Optimization of Network Pressure Management System

Pegah Rahimian[*1] & Sahand Behnam[2]

Received 10 August 2020; Revised 15 August 2020; Accepted 24 August 2020; Published online 30 September 2020

ABSTRACT

In this paper, a novel data-driven approach to improving the performance of wastewater management and pumping system is proposed, in which necessary data are obtained by data mining methods as the input parameters of optimization problem to be solved in nonlinear programming environment. In this regard, first, CART classifier decision tree is used to classify the operation mode, or the number of active pumps, based on the historical data of Austin-Texas infrastructure. Then, SOM is utilized to classify the customers and select the most important features that might have effect on the consumption pattern. Further, the extracted features is fed to Levenberg-Marquardt (LM) neural network that predicts the required outflow rate of the period for each operation mode classified by CART. The results showed that the prediction F-measures were measured 90%, 88%, and 84% for each operation mode 1, 2, and 3, respectively. Finally, the nonlinear optimization problem is developed based on the data and features extracted from the previous steps solved by artificial immune algorithm. The results of the optimization model were compared with the observed data, showing that the proposed model could save up to 2%-8% of the outflow rate and wastewater, regarded as a significant improvement in the performance of pumping system.

KEYWORDS: *Network Pressure Management; Data mining; Neural network; Nonlinear programming; Artificial Immune network.*

1. Introduction

Nowadays, water scarcity is gaining increasing importance for countries, especially the developing ones, due to global warming and droughts (IWMI 2009). Under such conditions, crucial management problems such as wastewater and resource management may arise. In recent decades, water resources have been exhaustively consumed for agricultural, irrigational, industrial, urban, and residential purposes. This multiplicity of demands for water complicates the wastewater management more than ever. Therefore, the need to propose a comprehensive approach to tackle the problem of multiple sources and coming up with a robust optimal solution is felt. Furthermore, recognizing the consumers' consumption patterns become more complex than ever since their different behaviors vary from region to region and period to period. A majority of consumption databases comprise numerous features that would confuse managers with making decisions about what features greatly affect consumption that must be taken under control. In this respect, a data-driven and feature-based framework is beneficial to select the key features of consumption and control them in order to optimize the wastewater.

Several researches have been conducted to examine the multi-objective environment of wastewater management and pumping systems. Lopez et al. proposed Evolutionary Algorithms (AEs) to introduce the multi-objective approach to minimizing the pumping cost and maximizing the stop time [1], and Kernan et al. (2017) studied the Genetic Algorithms (GA) used for optimizing and scheduling the pumping configuration along with EPANET hydraulic solver [2]. In addition, Torregrossa et al. (2019) compared Genetic Algorithm (GA) with Particle Swarm Optimization (PSO) to activate dynamic pump and minimize the costs [3].

Furthermore, optimization process in this area requires convex Mixed-Integer Nonlinear Programming (MINLP). Zhuan et al. (2013)

* **Corresponding author:** ***Pegah Rahimian***
pegah.rahimian@tmit.bme.hu

1. *PhD Student, Budapest University of Technology & Economics, Telecommunication & Media Informatics, HSNLab.*
2. *CEO & Founder, Teleminer GmbH.*

studied the Evolutionary Algorithms and different variants of dynamic programming [4]. Guo et al. investigated simulation models using Particle Swarm Optimization (PSO) to optimize water demand supply in different locations by means of effective reservoir operations [5]. Karamouz et al. suggested an implicit stochastic method for optimal operation function using historical time series to statistically obtain the optimal decision rules [6].

Wang et al. utilized artificial neural network to tackle the problems caused by the nonlinear and complex nature of the relationships among the independent features of water demand [7]. Mehta and Jain employed Fuzzy technology to extract the reservoir operation rules and compared the performances of several Fuzzy approaches [8].

However, there are some common characteristics and limitation in all of the abovementioned studies:

1. *They were designed only for fixed speeds in pumping systems; therefore, there would not be any real-time controlling and optimization.*
2. *They did not consider consumers' features and characteristics that would significantly affect water demand and supply operation.*
3. *They assumed the operational mode of pumping system (number of pumps) to be fixed or the pump activation were not obtained from the historical data.*

This paper aims to introduce a comprehensive approach to compensate the limitations and constraints of previous studies in several ways. First, the proposed approach determines the operational modes of pumping system infrastructure and historical data of demands. Therefore, it ensures more accuracy in demand satisfaction. Second, this model is capable of extracting the most important features that have significant effects on consumers' demands. Among these features is location which affects the consumers' habits and consumption. Finally, it is a developed approach to optimization in real time since it considers small windows of time and the optimization is dynamically done over a specific time period at various pump speeds.

The rest of the present paper is structured as follows: Section 2 describes the dataset and pumping system infrastructure of Austin-Texas. Section 3 discusses the data mining process for three different methods, namely CART classification diagram for operation mode (number of active pumps), Self-Organized Map (SOM) for consumers' clustering and feature extraction, and L-M neural network for outflow rate prediction. Section 4 presents the optimization model which is solved by means of artificial immune network (aiNet) algorithm.

2. Methodology

2.1. Dataset

The dataset used in this study consists of several tables including Austin-Texas consumers' data, pumping system data, and Energy consumption data. Each table includes water consumption data according to the characteristics of consumers and pumping system. The whole dataset includes the data obtained over the years from 2013 to 2019; however, in this paper, the analysis is limited to the 4th quarter of the year 2017.
(https://doi.org/10.26000/007.000004)

2.2. Pumping system and wastewater management process

Wastewater management process can be divided into different phases. Yoo et.al (2001) [9] categorized the important phases as pre-treatment, primary treatment, secondary treatment (biological), tertiary treatment, disinfection, sludge treatment, and odor control.

This paper primarily emphasized on the intermediate wastewater and pumping management which can be put between primary and secondary phases. The primary phase is the process of solid material separation using lamellar and the secondary phase is the process of biological filtering.

A majority of previous studies on wastewater optimization introduced an optimization approach applied to several pump settings. However, in this paper, attempts were made to use the determined numbers of pumps supported by Pressure Reducing Valve (PRV). Hence the speed and frequency of the pumps are regarded as decision variables, and the requirements including the output of the data mining phase is fed to the PRV setting. Figure.1 shows the pump and PRV setting provided for Austin-Texas water consumers.

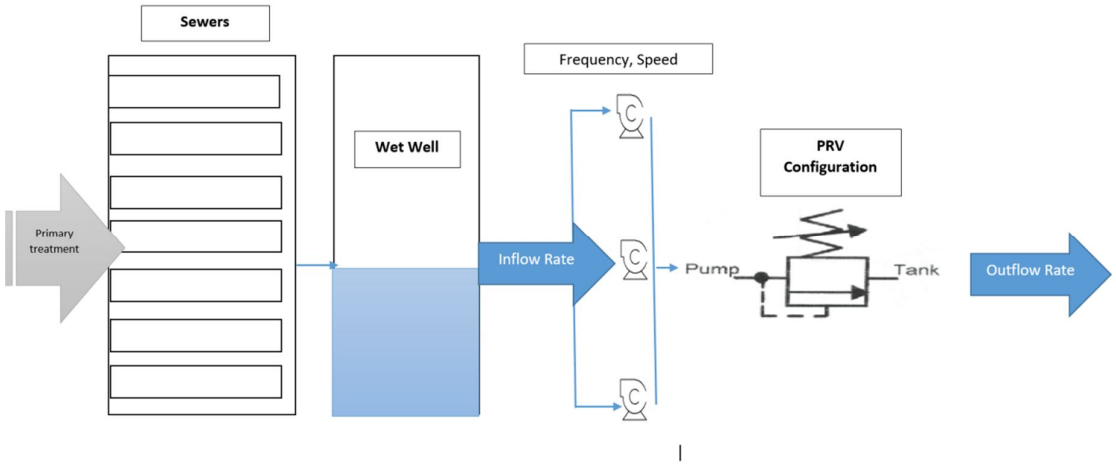

Fig. 1. Pumping system & wastewater infrastructure

2.3. PRV setting

Generally, PRV is installed to reduce the water pressure or outflow rate, while there are not excessive demands. However, this existing equipment is not intelligent enough to control the demands (consumers' requirements) and infrastructure facilities at the same time. Therefore, significant water leakage and pipe breaks are observed due to high outflow rate in off-peak and low flow rates in peak demand.

The proposed approach is capable to be set on PRVs and easily measures the outflow rate, optimizes it, and reduces the pressure according to the estimated outflow rate. Therefore, it works totally automatic and intelligent, significantly manages wastewater, decreases pipe breaks and water leakage, and comprehensively satisfies the consumers' demands at any time.

2.4. Pump scheduling:

In order to distinguish among the operative pumps, idle pumps, and backup pumps, some instructions should be followed in pump settings. To this end, the recursive procedure is used to classify the observations set in CART. Pumping system data table with several characteristics in a certain time period was also used that comprised the number of active, idle, and backup pumps, wet well level, frequency/speed of each pump, and amount of Energy consumption.

The other factor that significantly affects the CART diagram and node separation is influent or input flow rate. However, these observations are not available in our dataset since not all sewers are equipped with sensors to track the influent flow rate; even if they are equipped with sensor, the distances among the sensors and wet are not equal. Therefore, several studies suggested different formulas to estimate the influent flow rate. In this paper, it was assumed that the influent flow rate is directly affected by wet well level and wastewater flow rate after pumping. Both of these factors are available in our dataset and are used in estimation [10].

$$\text{Inflow}(t) = \text{Outflow}(t).\Delta T + \frac{(\Delta L * AL)}{\Delta T} \quad (1)$$

where t is the current time, ΔT is the sampling period, ΔL is the difference between the wet well levels at two sampling points, and A_L is the area of wet well level. This estimation was done to determine the decision parameter of CART diagram in Subsection 3.1.

Therefore, pumping time must pass from the minimum level of wet well to the maximum level of wet well and can be defined as:

$$\Delta T = \frac{(\Delta L * AL)}{Outflow(t) - Inflow(t)} \quad (2)$$

In the rest of this paper, this time window is used as the sampling period and directly utilized in optimization problem in Section 4.

3. Data Analysis

3.1. CART classification for operation mode:

According to these observations, CART diagram classification is constructed using Gini Coefficient as the index of impurity. Figure 2 shows the classification of the rules.

Index of impurity can be defined through the segmentation rule in decision tree nodes. Therefore, Gini Coefficient proposed by

Speybroeck et.al (1998) was employed and shown as follows [11]:

$$\text{Gini}(n) = 1-\sum I^2(j|n) \quad (3)$$

where 'n' is the node, Gini(n) is the Gini Coefficient of n, and 'I 'is the proportion of class 'j' in node 'n'.

If the Gini coefficient is available, the recursive procedure of segmentation of the root node begins. To this end, the difference in impurity of the parent node and sub-nodes is calculated as [12]:

$$\Delta i(s,n) = \text{Gini}(n) - I_L[i(n_L)] - I_R[i(n_R)] \quad (4)$$

where I_L and I_R are proportions of sample in left and right sides of the node and $i(n_{L/R})$ is the impurity of the left and right sub-nodes. The nodes with the maximum value of $\Delta i(s,n)$ are used for segmentation in each step, and the recursive process is run until CART diagram is constructed as shown in Figure2.

Figure 2 classifies the operation mode or number of active pumps in each time period based on the provided dataset. The system first measures the wet well level; if it is higher than 6m, it will activate the first pump with frequency "F" and then again, measure it to be approximately 50Hz. If the frequency is less than 50Hz, Pump "P1" will suffice in this time period, and if it is more than 50Hz, the second pump "P2" will be activated and its speed measured. If the speed is higher than the rated speed, the third pump "P3" will be activated; otherwise, "P1+P2" will suffice. On the contrary, if the wet well level is less than 6m, the inflow rate will be estimated using Equation 1. If the inflow rate is higher than 230k (m3/day), Pump "P2" will be activated; otherwise, Pump "P1" will suffice. Of note, these rules in this tree are extracted from the historical data of the pump allocation based on the wet well level, pump frequency, speed, and inflow rate.

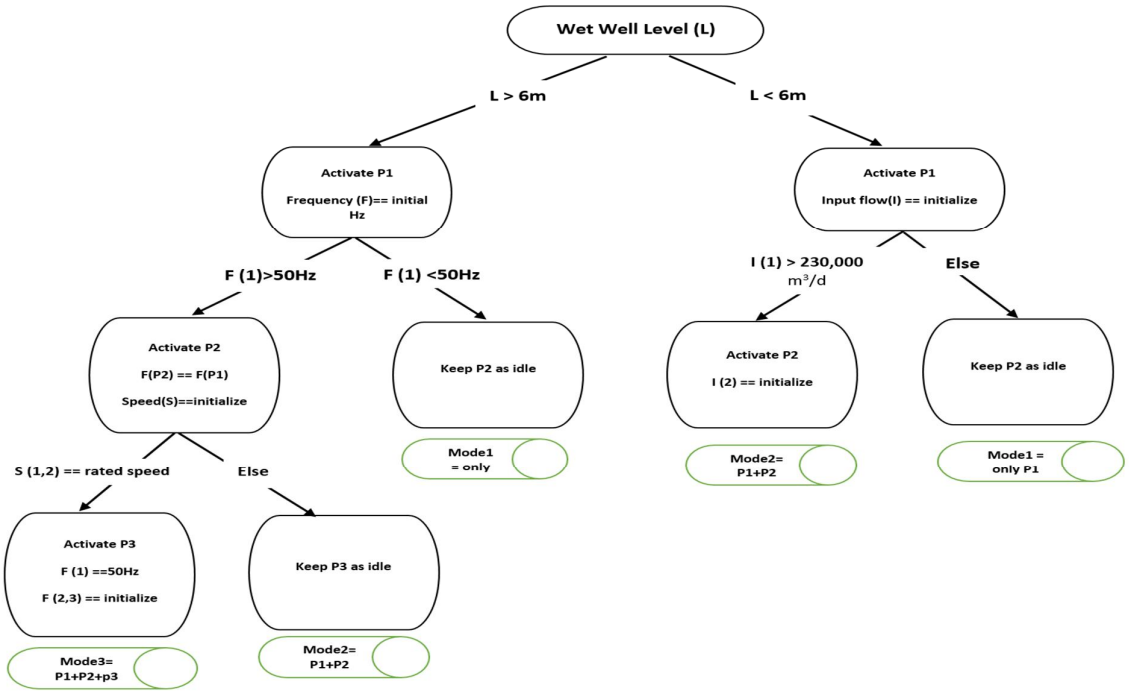

Fig. 2. CART classification diagram for recognizing operation mode (M1: {P1}, M2: {P1+P2}, M3: {P1+P2+P3})

3.2. Feature selection and consumer's classification:

Water consumption datasets usually consist of numerous features that describe the consumers' patterns. Hence the data system is noisy that lessens the accuracy of the prediction models. This is why selecting top independent features and dropping highly correlated ones are significant tasks which should be done prior to consumption prediction.

The customers' dataset itself classifies customers into 4 different categories with their specific features and consumption data:

- Irrigation-Residential
- Irrigation-multi-family
- Residential

- Multi-family

The observed consumption of these different classes for the whole year 2017 is shown in Fig. 2.

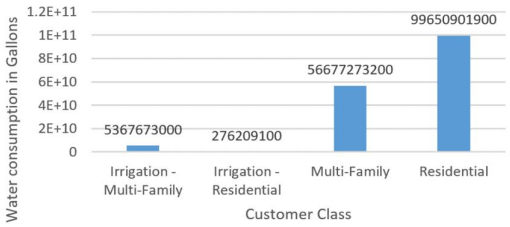

Fig. 3. Visualization of Austin consumer's classes in dataset

However, this classification does not seem rational since no relationship is found between the residential and multi-families or Irrigation and non-irrigation classes. In such situations, the possibility of automatic classification of water consuming is investigated in order to achieve the important and effective features affecting water consumption.

Kohonen Self-Organizing Maps (SOMs) was used in the analysis of automatic classification. This method helps reduce the number of unnecessary dimensions and indicates a transparent graphic of two-dimensional images. This is the unsupervised technique used to cluster customers on the basis of their effective features in water consumption.

In other words, SOM functions as in the following: if a large number of consumers share similar characteristics and tend to be in the same map cluster, it can be concluded that the mentioned cluster would have significant effect on water consumption. Therefore, it is regarded as one of the significant factors on water consumption.

Chrysi et al. (2015) [13] suggested the effective features of the input vector of SOM as shown in Table 1. Q shows the time period of a different quarter of the year, and Q3 is considered the basis time period due to the increasing amount of water consumption in summer.

Tab. 1. Input vector of SOM

Input Feature in SOM	Name
Avg of quarterly consumption	Q_{Avg}
Avg consumption over maximum consumption ratio	Q_{Avg} / Q_{max}
Ratio of first quarter Avg. over the third quarter	Q_{1Avg} / Q_{3Avg}
Ratio of second quarter Avg. over the third quarter	Q_{2Avg} / Q_{3Avg}
Ratio of fourth quarter Avg. over the third quarter	Q_{4Avg} / Q_{3Avg}

To obtain this analysis, SOM was implemented using a special clustering tool of neural network in Matlab and Rapidminer with the same result. The training parameters are:

Tab. 2. Features of the self-organized map

Size of the Map	Radius for Neighbouring	Learning Rate	Epochs
4 x 4	1	0.9	200

After the successful NN algorithm training, SOM map is created and the water consumption data is clustered into 4 different categories according to Figs. 3,4, presenting the output of Rapidminer and Matlab, respectively.

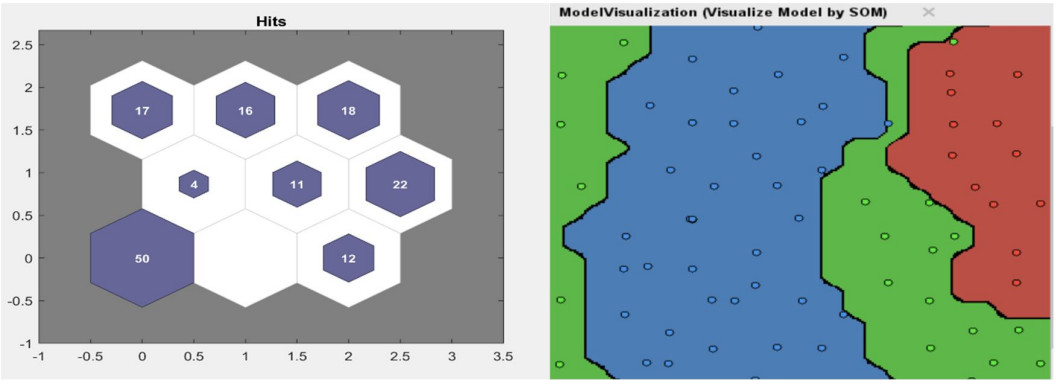

Fig. 3-4. Consumers' clustering results and hitting numbers for each cluster

Figures 3 and 4 both show a 4x4 map of training data which clusters the water consumers in 4 different categories. In other words, these maps suggest that water consumption depends on 4 different significant features. However, there are consumers with other features that were not included in SOM due to the long radius of neighborhood, showing the feature's trivial effect. In this case, those insignificant features are excluded and the dimension of the map are reduced to 4x4.

The next step is to estimate the potentiality of consumers' different features that are clustered in one of the 4 different categories based on similarity of characteristics. To this end, attempts were made to calculate the percentage of customers with one specific features, i.e., the number of individuals per household in each cluster. If a significant relationship is observed among different clusters, it can be concluded that the feature belongs to one of the 4 main categories shown in SOM.

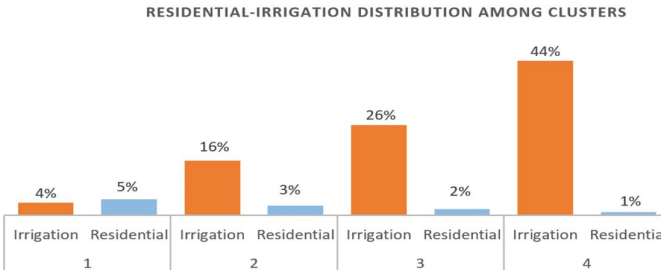

Fig. 5. Residential – Irrigation distribution in each cluster

According to Fig. 5, approximately 44% of consumers are clustered in the fourth category. Upon a shift from the first to the fourth cluster, the percentages of consumers in the irrigation and residential category increase and decrease, respectively. This is an evidence that residential-irrigation is one of the categories provided by SOM. An analysis of the number of people per household was done and positive results were obtained as well. In order to increase the certainty of this experiment, regression analysis was done to show the correlation of the water consumption and number of people per household. Therefore, this feature is considered as another significant categories of water consumer's properties. Fig. 6 shows the regression analysis and correlation of two factors.

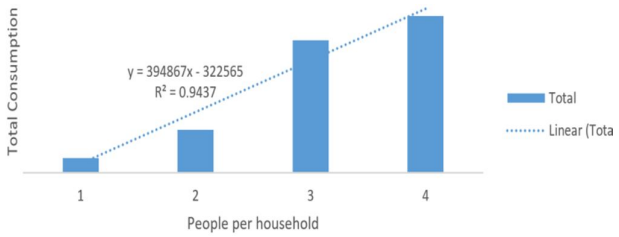

Fig. 6. Regression analysis of the independent variables

The same analysis was done considering other potential characteristics of consumers. However, it is possible to skip the details of analysis for rest of characteristics and draw a reliable conclusion with respect to most significant clusters as the output of SOM, incluing:

- Number of people per household
- Income of family
- Residential-Irrigation
- Number of Pumps for each unit

3.3. Data-driven evaluation of pumping system performance:

In this section, the performance of the pumping system is evaluated by developing a prediction model. The results obtained from CART classification regarding the mode of the system (number of active pumps) and SOM regarding the feature and requirement selection were used in order to get the desired outputs from the proposed model.

First, a model was developed to estimate the outflow rate per household or business. Zhang et al. [14] compared the performance of several datamining algorithms such as Random Forest (RF), Support Vector Machine (SVM), and K nearest neighbor to resolve the problem of complex and none-linear nature of wastewater pumping systems and concluded that Neural Network (NN) outperformed others with high accuracy of estimation. Therefore, NN was utilized to estimate the outflow rate in this study. The variables and structure of network are given in the following. Table 3 lists all the parameters used for modeling the outflow rate. Of note, the total outflow rate of the district can be easily calculated via summation of estimated outflow rates per household or business.

Tab. 3. Description of the parameters

Parameters' Names	Parameter description and possible values
NP	Number of people per household or business
FI	Total family or business income
RI	Consumer type (Binary): Residential=0, Irrigation=1
M	Operation mode = number of active pumps = 1,2,3
$S_{t,m}$	Speed of pump 'm' at time 't'
L_t	Wet well level at time 't'
O_t	Outflow rate at time 't'

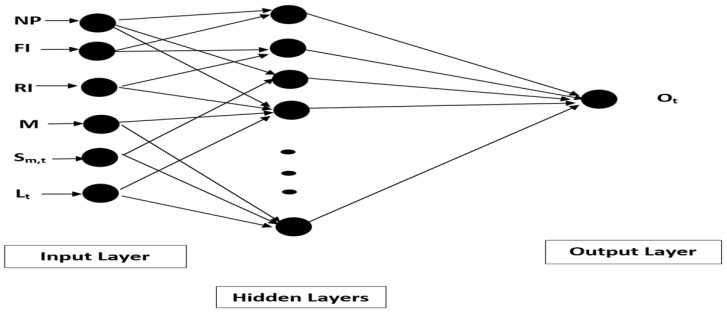

Fig. 7. LM neural network structures and input/output nodes

LM algorithm, i.e., the advanced form of Gaussian-Newton algorithm, was chosen to train this network. It has the capabilities of both Gaussian-Newton and gradient methods. The structure of the proposed LM network is described in the following:

1. There are 6 input nodes which are the outputs of the previous section analysis. NP, FI, and RI were selected from SOM and M was regarded as the operation mode or the number of active pumps from CART classification.
2. The empirical formula was used for calculating the number of hidden layers:

$$N = \sqrt{Ni + No} + a \quad (5)$$

where 'Ni' is the input nodes (6), 'No' the output nodes (1), and 'a' the constant number. After several trials of changing 'a' and error calculation, the optimal number of hidden layers was calculated as 6.

Now, the outflow rate should be modeled according to the operation modes:

$O_{m1,t} = NN_{m1}(O_{m1,t-\Delta T}, S_{m1,t}, S_{m1,t-\Delta T}, L_t, L_{t-\Delta T}, NP, FI, RI)$ (6)

$O_{m2,t} = NN_{m2}(O_{m2,t-\Delta T}, S_{m1,t}, S_{m1,t-\Delta T}, S_{m2,t}, S_{m2,t-\Delta T}, L_t, L_{t-\Delta T}, NP, FI, RI)$ (7)

$O_{m3,t} = NN_{m3}(O_{m3,t-\Delta T}, S_{m1,t}, S_{m1,t-\Delta T}, S_{m2,t}, S_{m2,t-\Delta T}, S_{m3,t}, S_{m3,t-\Delta T}, L_t, L_{t-\Delta T}, NP, FI, RI)$ (8)

where $O_{mi,t}$ is the estimated outflow rate in the operation mode 'i:1,2,3' at time 't' and NN_{mi} is the LM neural network prediction model for outflow wastewater. The rest of the parameters are described in Table 3.

In the following section, the accuracy of the constructed model is calculated by evaluating different metrics such as Mean Squared Error (MSE), precision, recall, and F1 through the confusion matrix.

3. Results

The model in this paper is fitted to the R package and the results of the confusion matrix are directly derived from the R code after fitting. Therefore, the code is run 3 times for 3 different operation modes and each time, the metrics are obtained. Table 4 shows the results of the matrix for each operation mode. Of note, F1 score is calculated using the following expression, demonstrating the performance of our mode:

$$F1 = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}} \quad (9)$$

Tab. 4. Result of the training and test analysis of the model

Operation Mode	MSE	Precision	Recall	F1
M1: {P1}	0.01	92%	89%	90%
M2: {P1+P2}	0.01	87%	88%	88%
M3: {P1+P2+P3}	0.07	85%	82%	84%

Moreover, the robustness of our approach is emphasized by plotting the ROC curve, as shown in Fig. 8.

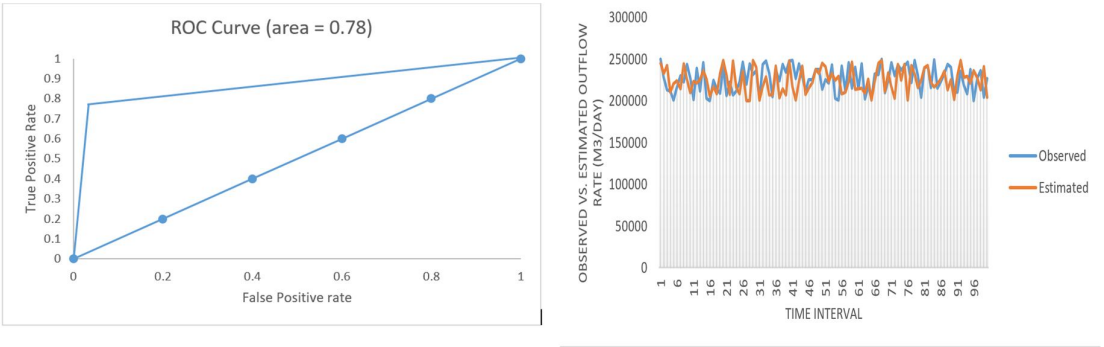

Fig. 8-9. ROC Curve of the fitted model & visualization of observed estimated points

So far, based on the inputs such as pumps speed, wet well level, and characteristics of customers, customers' requirements were predicted by estimating the outflow rate. However, some important factors of PRV are still lacking. The following questions should be answered. What is the optimal speed of the pumps at each time stamp? What is the maximum outflow rate to be fed on PRVs to prevent pipe leakage during low-demand periods while satisfying all the constraints from Texas wastewater infrastructure? The following section discusses the optimization model of the proposed approach.

4. Optimization Process for PRV Setting

In this study, an optimization model was designed to maximize the outflow rate while satisfying the requirements of consumers and Austin infrastructure. Thus, the objective function is expressed as follows:

$$O_{opt,T} = \int_T O(t)\, dt \quad (10)$$

where O_{opt} is the optimal outflow rate to be set in PRV setting, and T is expressed as time window of sampling, estimated using Equation (2) in Subsection 2.3.

There are 3 operation modes (i:1,2,3), and the equation should be discretized so that it can be solved. Thus, the objective function can be expressed as follows:

$$O_{opt} = \sum_{i=1}^{3} \sum_{t=1}^{T} O(mi,t)\, \Delta t \quad (11)$$

The decision variables used in this study are pump speed and wet well level in each time window. In Austin infrastructure, the rotational speed can be set as the maximum grid frequency which is 50Hz corresponding to 3000 rpm, and the minimum is assumed to be 1500 rpm to prevent the system from overheating. Moreover, the wet well level changes at the time window Δt according to the following Equation:

$$L_{t+\Delta t} = L_t + \frac{(O(t)-I(t)) * \Delta t}{AL} \quad (12)$$

where O_t is the estimated outflow rate at time t in Section 4, I_t is the estimated inflow rate at time t according to Equation 1, and A_L (m^2) is the constant area of the wet well level. Moreover, the wet well level should always be kept between the allowed minimum and maximum which varies from time to time; however, at this specific time window, it is assumed constant.

Moreover, PRV requires a momentum parameter which will be directly affected by the rotational speed changes over time. Thus, the PRV momentum is expressed as:

$$\tau(\text{PRV}) = \frac{S(t-\Delta t)}{S(t)} * \Delta t \quad (13)$$

In order to decrease the number of constraints and complexity of the problem, the wet well level constraint is integrated in the objective function with Lagrangian Coefficient (λ). Therefore, the model for outflow optimization is formulated as:

$$\text{Min } O_{opt} = \sum_{i=1}^{3} \sum_{t=1}^{T} O\ (mi, t)\ \Delta t \quad (14)$$

St.

$O_{m1,t} = NN_{m1}\ (O_{m1,t-\Delta T}, S_{m1,t}, S_{m1,t-\Delta T}, L_t, L_{t-\Delta T}, NP, FI, RI) + \lambda(\max\{0, L_{t+\Delta t} - L_{max}\})$

$O_{m2,t} = NN_{m2}\ (O_{m2,t-\Delta T}, S_{m1,t}, S_{m1,t-\Delta T}, S_{m2,t}, S_{m2,t-\Delta T}, L_t, L_{t-\Delta T}, NP, FI, RI) + \lambda(\max\{0, L_{t+\Delta t} - L_{max}\})$

$O_{m3,t} = NN_{m3}(O_{m3,t-\Delta T}, S_{m1,t}, S_{m1,t-\Delta T}, S_{m2,t}, S_{m2,t-\Delta T}, S_{m3,t}, S_{m3,t-\Delta T}, L_t, L_{t-\Delta T}, NP, FI, RI) + \lambda(\max\{0, L_{t+\Delta t} - L_{max}\})$

$L_{t+\Delta t} = L_t + \frac{(O(t)-I(t)) * \Delta t}{AL}$

$\tau(\text{PRV}) = \frac{S(t-\Delta t)}{S(t)} * \Delta t$

$1500 < S_t < 3000$

$L_{min} < L_t < L_{max}$

4.1. Problem solving with aiNet

In order to solve the optimization model proposed in the previous section, several options with respect to different algorithms are available. The model is complex and nonlinear which makes the process more challenging. However, several studies have proposed different algorithms for solving such a problem. Most of such frequent studies are in the field of evolutionary algorithms or artificial immune systems.

A method of Mixed-Integer Non-Linear Problem (MINLP) was proposed by Oreste et al. (2019) [10], which would solve the problem of having both linear and nonlinear constraints in the optimization model. In their studies, given that time window is quite large, a Mixed-Integer Non-Linear Programming is required for a NP-hard problem.

Genetic algorithm has been suggested in several studies for optimization of nonlinear problems in water distribution systems, and it has been concluded that the GA would yield a low-cost solution [2,18].

Nanas et al. (2007) compared the performance of evolutionary algorithms and artificial immune systems methods and suggested the superiority of artificial immune network for this kind of problem [19]. Therefore, the optimization problem solving was applied using artificial immune network (aiNet) since it was more robust and required lower computational cost than other methods.

4.2. Data points and optimization steps for aiNet

In this section, 3 different scenarios as the three observed operation modes in Section 2 are elaborated. Then, a time window from each operation mode test dataset should be set the duration of each of which is one hour. Each time window consists of 11 data points, and the optimization result in each step is used as the input of next step.

In the first step, the population with size N is randomly generated, and the following steps are iterated until the stop criterion is met. The full procedure of this algorithm was described by Xian et al. (2007) [21], and the researcher customized their process in this paper to solve the optimization problem:

1. Create a fixed size clone for each parent;
2. Calculate the fitness of each parent in the clone;
3. Compare the antibody of the parents with that of the highest fitness in each clone;
4. Replace the parent with the highest antibody value in the clone;

5. Search for the local optimal value, until the difference between 2 average fitness value of a clone is less than 0.02 (stop criterion).

Table 5 shows the summary of the optimization results of each scenario or operation mode, along with the outflow rates in the optimization process for the selected time windows and data points.

Tab. 5. Results of optimization

Operation Mode	Calculated outflow rate M^3/day	Observed outflow rate in test dataset	Result of saving in optimization
M1	227,000	221,498	0.02
M2	359,000	329,659	0.08
M3	573,000	570,390	0

5. Conclusion

The present study proposes a data-driven and feature-based framework and aims to optimize the wastewater management and pumping system performance by minimizing the outflow rate to prevent the system from pipe leakage and to satisfy consumers' demands while it is adjusted to PRVs. Numerous studies have been conducted to improve the performance. However, they did not take into consideration the confusing nature of wastewater datasets due to the increasing number of features and sources provided for consumers. Therefore, this study considered some pre-processing over the dataset obtained from the result of our data analysis to increase the accuracy of the proposed optimization model and decrease its complexities. In the present study, the infrastructure data of Austin-Texas pumping system in 2017 was utilized to construct the CART classification diagram, helping the decision makers to specify the optimum number of pumps in each period according to the wet well level, inflow rate, frequency, and speed of the pumps; This process is defined as operation mode. Then, SOM is used to select the most significant features of the consumer's data with respect to the consumption pattern. Therefore, only the output features of the map are used to reach the next step and rest of the features in our dataset are neglected. For the next step, the extracted features of SOM are used as the input of Neural network to predict the outflow rate of the pumping system. This estimation shows the required amount of consumption in each time period. The F-measures of this estimation are 90%, 88%, and 84% for operation mode 1, 2, and 3 respectively, which shows the robustness of our prediction method. However, in order to make the pumping system intelligent, the optimal amount of outflow rate should be set on the PRVs, and this number is regarded as the solution of the nonlinear optimization problem. The decision variables include wet well level and speed of the pumps, and the problem is solved with artificial immune network algorithm. The optimization process leads to 2% to 8% more saving in wastewater and outflow rate in comparison with the observed rate; and it indicates that the performance of pumping system is directly affected by the operation mode, feature selection of consumers, and outflow rate estimation model.

References

[1] M. López-Ibáñez, T. D. Prasad, B. Paechter, Optimal pump scheduling: Representation and multiple objectives, in: Proceedings of the eighth International Conference on Computing and Control for the Water Industry, Vol. 1, pp. 117-122.

[2] AD Savic, GA Waiters·, RM Atkinson· & MR Smitht, Genetic Algorithm Optimization of Large Water Distribution System Expansion; Measurement + Control, Vol. 32, (1999).

[3] D. Torregrossa, F. Capitanescu, Optimization models to save energy and enlarge the operational life of water pumping systems, Applied Energy Vol. 213 (2019), pp. 89-98.

[4] X. Zhuan, X. Xia, Optimal operations cheduling of apumping station with multi plepumps, Applied Energy Vol. 104 (2013), pp. 250-257.

[5] Ai, X.; Gao, Z. Combined optimal method of drawing reservoir optimal operation figure. In Advances in Computer Science, Intelligent System and Environment; Springer: Berlin/Heidelberg, Germany, (2011), pp. 103-107.

[6] Karamouz, M.; Houck, M.H. Annual and monthly reservoir operating rules generated by deterministic optimization. Water Resour. Res. Vol. 18, (1982), pp. 1337-1344.

[7] Wang, Y.-M.; Chang, J.-X.; Huang, Q. Simulation with RBF neural network model for reservoir operation rules. Water Resour. Res. Vol. 24, (2010), pp. 2597-2610.

[8] Mehta, R.; Jain, S.K. Optimal operation of a multi-purpose reservoir using neuro-fuzzy technique. Water Resour. Res. Vol. 23, (2009), pp. 509-529.

[9] C. K. Yoo, D. S. Kim, J.-H. Cho, S. W. Choi, I.-B. Lee, Process system engineering in wastewater treatment process, Korean Journal of Chemical Engineering Vol. 18 (2001), pp. 408-421.

[10] Oreste Fecarotta , Riccardo Martino and Maria Cristina Morani; Wastewater Pump Control under Mechanical Wear; https://www.mdpi.com/journal/water, (2019).

[11] Speybroeck, N. Classification and Regression Trees; Chapman & Hall/CRC: Boca Raton, FL, USA, (1998), pp. 1174-1176.

[12] Yi Ji, Xiaohui Lei, Siyu Cai, and Xu Wang; Application of a Classifier Based on Data Mining Techniques in Water Supply Operation; Water Vol. 8, (2016), p. 599. Doi:10.3390/w8120599 www.mdpi.com/journal/water

[13] Chrysi Laspidoua, Elpiniki Papageorgiou, Konstantinos Kokkinos , Sambit Sahud , Arpit Guptae , Leandros Tassiulas; Exploring patterns in water consumption by clustering; 13th Computer Control for Water Industry Conference, CCWI (2015).

[14] Zhang Z, Kusiak A. Models for optimization of energy consumption of pumps in a wastewater processing plant. ASCE J Energy Eng Vol. 137, No. 4, (2011), pp. 159-68.

[15] Bessler,F.T.;Savic,D.A.;Walters,G.A.Water reservoir control with datamining. J.WaterResour. Plan. Manag. Vol. 129, (2003), pp. 26-34.

[16] L.N. de Castro, F.J. Von Zuben, and H. Knidel (Eds.): Multimodal Dynamic Optimization: From Evolutionary Algorithms to Artificial Immune Systems; ICARIS 2007, LNCS 4628, (2007), pp. 13-24.

[17] Burer, S.; Letchford, A.N. Non-convex mixed-integer nonlinear programming: A survey. *Surv. Oper. Res. Manag. Sci.* Vol. 17, (2012), pp. 97-106.

[18] Indrani Gupta, A Gupta, P Khanna, Genetic algorithm for optimization of water distribution systems; Environmental Modelling & Software, (1999).

[19] Nanas N, De Roeck A. Multimodal dynamics optimization: from evolutionary algorithms to artificial immune system. In: Proceedings of the 6th international conference on artificial immune systems; (2007), pp. 13-24.

[20] R. Kernan, X. Liu, S. McLoone, B. Fox, Demand side management of an urban water supply using wholesale electricity price, Applied Energy Vol. 189, (2017), pp. 395-402.

[21] Xian Shen, X. Z. Gao, and Rongfang Bie, Artificial Immune Networks: Models and Applications; International Journal of Computational Intelligence Systems, Vol. 1, No. 2, (2008), pp. 168-176 Published by Atlantis Press.

Follow This Article at The Following Site:

Rahimian P, Behnam S. A novel data driven and feature based forecasting framework for wastewater optimization of network pressure management system. IJIEPR. 2020; 31 (3) :423-433
URL: http://ijiepr.iust.ac.ir/article-1-1098-en.html