

Article

Estimation of Non-Revenue Water Ratio for Sustainable Management Using Artificial Neural Network and Z-Score in Incheon, Republic of Korea

Dongwoo Jang and Gyewoon Choi *

Department of Civil & Environmental Engineering, Incheon National University, Incheon 22012, Korea; nightray@inu.ac.kr

* Correspondence: gyewoon@inu.ac.kr; Tel.: +82-32-835-4760; Fax: +82-32-851-5730

Received: 21 September 2017; Accepted: 16 October 2017; Published: 25 October 2017

Abstract: The non-revenue water (NRW) ratio in a water distribution system is the ratio of the loss due to unbilled authorized consumption, apparent losses and real losses to the overall system input volume (SIV). The method of estimating the NRW ratio by measurement might not work in an area with no district metered areas (DMAs) or with unclear administrative district. Through multiple regression analyses is a statistical analysis method for calculating the NRW ratio using the main parameters of the water distribution system, although its disadvantage is lower accuracy than that of the measured NRW ratio. In this study, an artificial neural network (ANN) was used to estimate the NRW ratio. The results of the study proved that the accuracy of NRW ratio calculated by the ANN model was higher than by multiple regression analysis. The developed ANN model was shown to have an accuracy that varies depending on the number of neurons in the hidden layer. Therefore, when using the ANN model, the optimal number of neurons must be determined. In addition, the accuracy of the outlier removal condition was higher than that of the original data used condition.

Keywords: water distribution system; non-revenue water; artificial neural network; multiple regression analysis

1. Introduction

Non-revenue water (NRW) includes water lost from physical incidents such as pipe leaks caused by bursts in a water distribution system and water-related commercial losses stemming from illegal connections, unmetered public use and meter error [1]. NRW ratio is 5–50% for major countries. Singapore, Denmark and Netherlands have the lowest NRW ratio (5–6%), while Chile (34%) and Mexico (51%) have the highest NRW ratio [2]. According to data from Korea waterworks 2015 [3], the NRW ratio of major cities in Korea is the lowest in Seoul at 4.9% and the highest in Gwangju at 56.8%. Incheon has an NRW ratio of 11.2%, lower than the national average of 16.3% percent.

Incheon takes its tap water from Paldang Dam via a single pipeline, thus making it vulnerable to pipe breakage due to accident or disaster [4,5]. This makes consumers likely to suffer damage due to suspension of water supply. To prevent this, the management of hydraulic pressure in the pipe network and regular evaluation of pipe deterioration are recommended measures. A decrease in the NRW ratio correlates to the reduction of leak quantity by optimal operation management in a district metered areas (DMA).

Analysis of the effects of pipe damage on the overall water distribution system helps determine what to improve first in the water pipeline [5]. A systematic plan for replacement and remediation is in effect for the maintenance of the city waterworks [6–8]. Though improvement projects for old waterworks are being implemented, it is difficult to reduce the system's economic losses and improve

its function via the evaluation of old pipes and accident prevention, which depend on empirical judgment [9,10].

Therefore, research and analysis of the factors affecting leaks when deciding the priority of water distribution system maintenance are needed, as well as identifying the physical and operational factors affecting leaks with parameters such as hydraulic pressure, deteriorated pipe ratio and water supply quantity. To decrease the NRW ratio, studies such as those on pipe network analysis, reliability enhancement, diagnosis of pipe network technology and evaluation of pipe deterioration for optimal water distribution were conducted in previous research.

Determining the level of leaks and bursts in the overall volume of NRW, a performance indicator was found for comparing leak management in water supply system: The Infrastructure Leakage Index (ILI) [11–13].

In addition, studies have been carried out on the parameters of a water distribution system. A regression equation for predicting the NRW ratio was developed using statistical analysis by acquiring main parameter and statistical data on the analysis of water distribution system [14]. And water supply and the operating and maintenance cost of a water distribution system was suggested [15]. The system for performance indicators revised for small water supply utilities. Principal component analysis (PCA) was used to reduce the dimensionality of the original data [16,17].

These statistical techniques and performance indicators were helpful in forecasting NRW, and a number of parameters of water distribution systems were proposed and analyzed. This suggested numerous approaches to improve the accuracy of NRW ratio prediction, as well as a scientific approach toward the sustainable management of water distribution systems.

A well-established DMA in water distribution systems can be analyzed through physical and operational parameters [18]. To estimate the NRW ratio, including the amount of water leaks, the main parameters of water distribution systems appropriate for regional characteristics are selected, and the NRW calculation model, which was developed by statistical analysis, plays an important role in the planning and operating of DMA.

An artificial neural network (ANN) is a model used for predicting dependent variables through statistical learning algorithms when sufficient data on independent variables are available to describe dependent variables. Due to the lack of sufficient learning data, however, the ANN model has not been widely used in the estimation of the NRW ratio.

Major ANN studies applied to water distribution systems in recent years are as follows. A procedure to devise a general operating policy toward reservoir operation from a dynamic programming using neural network (DPN) was suggested [19]. Relatively new technique of using ANNs researched for forecasting short-term water demand [20]. ANNs in water quality modeling, as well as for the process and control of treating drinking water used in water distribution systems [21]. Research on the application of ANNs for analysis of data from sensors measuring hydraulic parameters are presented [22]. Additionally, the efficiency of computational intelligence techniques was compared in water demand forecasting [23].

Recent research about ANN used it as a means of estimating the temporal variation of analytic factors such as real-time water quality, operation of reservoir and short-term demand forecasting. The application of an ANN to water distribution systems for estimating NRW and parameter analysis, however, proved insufficient.

In this study, a model for NRW ratio calculation for Incheon was developed by considering an ANN and parameters of major water distribution systems. The statistical method was used to compare the results of the ANN and real measured values according to the removal of outliers through the use of Z-score standardization.

The results of the NRW ratio by multiple regression analysis and an ANN were compared through accuracy assessment analysis. To estimate the NRW ratio, parameters including deteriorated pipe ratio, water supply quantity per demand junction and demand energy ratio were selected in the previous research [24]. Demand energy was calculated using simulated nodal hydraulic pressure

and demand using EPANET 2.0 (Environmental Protection Agency, Cincinnati, OH, USA, 2000), a hydraulic numerical analysis model for water distribution systems.

2. Theoretical Background

2.1. Analysis of Water Supply Energy in Water Distribution Systems

The EPANET 2.0 model developed by the U.S. Environmental Protection Agency was used for the hydraulic modeling of DMA. This model used the gradient algorithm for pipe network analysis, and the extended period simulation was applied to analyze the hydraulic flow in the pipe network under the time-series condition [25].

Under the EPANET 2.0 model, a hybrid node-loop approach is used to calculate continuity and energy equations for in-pipe flow analysis. Continuous equations, the main theory of water network analysis, and energy equations for the analysis of energy losses were used in the model. The energy required in the pipe network can be divided into water supply and leakage energy that represented the water velocity and pressure head in the inner pipe. If velocity and pressure head are high in a pipe network, this raises the leakage quantity, so the required level of water supply energy in each demand junction that can be supplied will maintain minimum hydraulic pressure. In addition, the energy arising from difference between the total hydraulic head and minimum hydraulic head for stable water supply at junction is regarded as excessive energy that affects leakage.

The water supply energy is calculated by Equation (1) as the energy arising from the minimum residual head required for water supply. The estimation method for water supply energy is calculated by multiplying water demand and the hydraulic head via analysis of the EPANET pipe network in each junction.

$$E_w = (\rho_w g \times Q \times H) \times \Delta t \quad (1)$$

where E_w is water supply energy ($\text{kg} \cdot \text{m}^2 \cdot \text{s}^{-2} \times 10^3$), ρ_w is density of water ($1.0 \times 10^3 \text{ kg/m}^3$), g gravitational acceleration (9.8 m/s^2), Q is water demand (m^3/s) and H is hydraulic head (m), Δt is unit time.

The minimum residual head varies depending on the building level of the direct water supply according to related regulations of the water service provider. The Incheon Water Supply Ordinance allows direct water supply up to the fifth floor from ground level on condition of no pumping system, and the minimum level of residual hydraulic head is set at 25 m. This hydraulic head regarded as standard of available water supply energy in Incheon. The excessive energy is the difference between the total supply and available supply energy, which affects leaks in the pipe network and available supply energy is a condition of hydraulic pressure at 25 m. The demand energy ratio is the percentage of total supply energy divided by available supply energy considering the energy loss in the pipe network. Excessive energy can be defined as energy excluding the available supply of energy from the total supply of energy. When excessive energy is high, the demand energy ratio increases proportionally, which causes a higher volume of leakage.

2.2. Statistical Analysis

Statistical analysis was performed to find correlations of main parameters in water distribution systems. The method was to clarify and verify the functional relationship between parameters and analyze the correlation between the parameters of water distribution systems and the relationship between the selected dependent variable and independent variables.

Correlation Analysis

Correlation analysis studies the linear relationship between two variables in probability theory and statistics. Both variables can be correlated with each other from an independent relationship, and the strength of their relationship is called a Pearson correlation coefficient as defined as an

Equation (2) [26]. The correlation analysis was used to compare the accuracy between the ANN simulation and the actual measured values.

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

where r_{xy} is the correlation coefficient and \bar{x} , \bar{y} the mean values of x and y .

The correlation coefficient is obtained between minus 1.0 and 1.0 and has the following characteristics. Multiple regression analysis is an analytical technique that estimates causality between variables by statistical methods, as well as a method to analyze the regression model with a dependent variable and two or more independent variables. The multiple linear regression model with independent variables is expressed as (3).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3)$$

where x is the independent variable, y is the dependent variable, β is the regression coefficient and β_0 is the regression intercept.

A method for estimating the coefficients of multiple regression equations is a simultaneous input method for analyzing all independent variables and a method for removing specified variables at once, making a model consisting of constant terms only. In addition, the backward method eliminates all variables one by one according to the removal criterion after selection, and the stepwise method determines selection and exclusion of variables in each step [27].

2.3. Artificial Neural Network

An ANN is a massively parallel distributed processor with a natural propensity for storing experiential knowledge and making it available for use. It resembles the human brain in two respects: knowledge is acquired by the network through a learning process and inter-neuron connection strengths, known as synaptic weights, are used to store the knowledge [28].

The ANN procedure used is a feed-forward network type with input, hidden and output layers, as shown in Figure 1. Neurons in the input layer simply act as a buffer. Neurons in various layers are interconnected through weights. Neurons in the hidden and output layers are called the activation function, and the activation function used here is a sigmoidal activation function. The input for each neuron j in the hidden layer is the sum of the weighted input signal x_i . ($\sum w_{ji}x_i = net_j$, in which w_{ji} is the interconnecting weight between neuron j in the hidden layer and neuron i in the input layer.) The output y_j from the neuron given by the neuron output in the output layer is computed similarly.

$$y_i = f\left(\sum w_{ji}x_i\right) = \frac{1}{1 + e^{-net_j}} \quad (4)$$

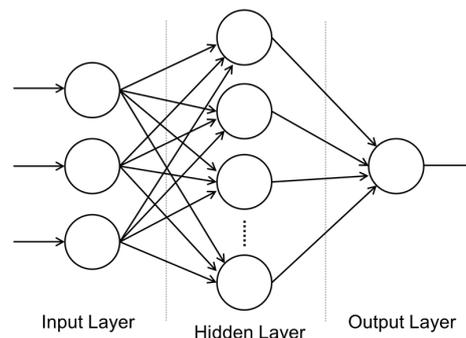


Figure 1. Schematic diagram of multilayer feed-forward neural network [28].

3. Status and Data Collection of Waterworks in the Target Area

The target area for this study was the Korean city of Incheon. The data collected included the status of the area, waterworks facilities and operational status, and the water supply indicators of the Incheon Waterworks Basic Plan of 2015. In addition, various hydraulic design data of the water distribution system and hydraulic simulation results were collected.

3.1. Status of Waterworks in Target Area

The water population of Incheon is 2,851,491 and the water supply rate is 98.3%. The daily water supply per person is 343 L, and the water supply area is divided into nine districts. The city has 24 reservoirs and 68 pumping stations. The total length of the network is 3634 km. DMAs were built in Incheon that divide all water supply districts into separate ones instead of directly supplying water from the water purification plant to tap. The DMA system of Incheon consists of six large DMAs within the boundary of the water purification plant, 32 DMAs in the reservoir boundary and 367 detailed small DMAs from reservoir boundary [29]. Table 1 shows the classification of Incheon's DMA system.

Table 1. Classification of Incheon's district metered area (DMA) system.

Large DMA	Middle DMA	Small DMA
Gongchon (1)	Incheon Int'l Airport, Geomam, Yeonhui, Ganghwa, Seongnam, Gajwa (16)	(96)
Namdong (1)	Manwol, Mt. Subong Mt. Jangsu, Songhyeun (4)	(87)
Bupyeong (1)	Chunma, Mt. Wonjeok, Mt. Heemang stream (3)	(79)
Susan (1)	Ohbong, Mt. Yeonsu, Songdo, Hagik, Dohwa, Munhak, Seochang (7)	(101)
Yeongheung (1)	Yeongheung (1)	(2)
Noon (1)	Noon (1)	(2)
Total : 6	Total: 32	Total: 367

Source: Waterworks HQ, Incheon Metropolitan City (2015).

The observed NRW ratio in 135 DMAs is shown in Figure 2.

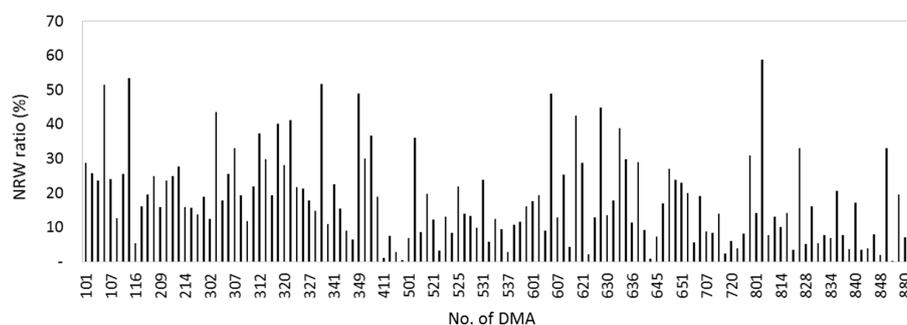


Figure 2. Observed non-revenue water (NRW) ratio in Incheon's DMAs.

3.2. Hydraulic Analysis of Water Distribution Systems

Analysis of Incheon's pipe network was done using the diagnosis data of water pipe technology established in 2015. Incheon Metropolitan City Waterworks, based on the GIS, built the pipe network by acquiring data such as pipe diameter and length, valve, flowmeter and ground level.

The hydraulic simulation of the network was performed for each DMA and the demand energy ratio (total supply energy/available supply energy) for each junction of a small DMA was calculated from the results of the analysis. Data such as pipe length, average pipe diameter, number of demand junctions and water supply amount for each DMA were used to construct the EPANET model.

The condition of EPANET simulation is that of the designated maximum water supply in 2015, and the demand energy ratio is obtained by calculating the pressure of the nodal point based on the demand amount at each node of a DMA. Based on the modeling simulation, the demand energy ratio of each DMA is shown in Figure 3.

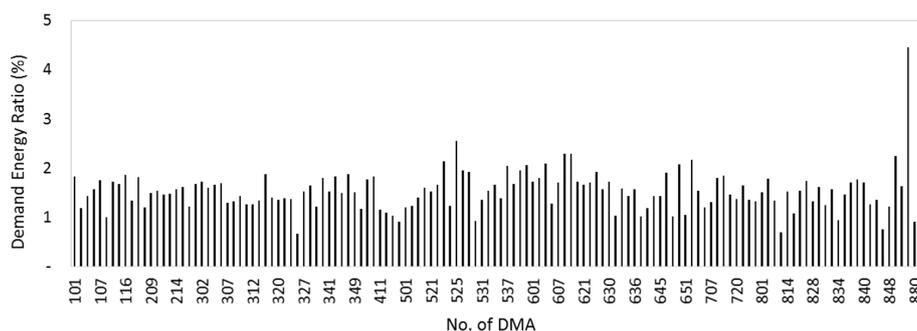


Figure 3. Simulated demand energy ratio in the DMA of Incheon.

4. Statistical Analysis of Main Parameters in Water Distribution Systems

4.1. Selection and Characteristics of Main Parameters

Analysis of the technical diagnosis results of Incheon's water pipe network established in 2015 showed that water pipe deterioration in the DMA system greatly influences NRW [29]. The deteriorated pipe ratio, pipe length, mean pipe diameter, number of demand junctions, water supply quantity, number of leaks and demand energy ratio of DMAs were selected as parameters that could affect the NRW ratio.

To derive the parameters with high correlation with the NRW ratio, three parameters were selected: the deteriorated pipe ratio, demand energy ratio and water supply quantity per junction through multiple regression analysis. From the previous research, the main parameters selected according to the statistically significant order of multiple regression analysis [24]; this is described in detail in Section 4.3.

The demand energy ratio is calculated by dividing the actual supply energy by the minimum required energy in the water supply network. The deteriorated pipe ratio is a parameter determined by pipe installation by year and pipe material. The number of leaks tends to increase as the degree of aging rises, and the water supply quantity per demand junction increases in apartments and densely populated districts.

4.2. Correlation Analysis of Each Parameter

To analyze the correlations between the parameters of water distribution systems, the physical and operational data of selected parameters in each DMA were used based on a diagnosis of Incheon's water network technology done in 2015. Data on 135 DMAs in Incheon were collected.

Table 2 shows the correlation analysis results for each parameter. The deteriorated pipe ratio and the number of leaks had a high correlation with the NRW ratio [24]. A positive correlation tendency was seen with the NRW ratio in the number of demand junctions and demand energy ratio, but the Pearson correlation coefficient of under 0.5 shows a low relationship with the measured NRW ratio. And the same coefficient between the water supply quantity and pipe length was 0.71, showing the highest correlation among the 10 used parameters.

As a result of the correlation analysis, the Pearson correlation coefficient was less than 0.5, except for the deteriorated pipe ratio, and the correlation between the NRW ratio and used parameters were found to be not high. The negative correlation coefficient was represented by figures such as the mean pipe diameter, mean pipe length per demand junction, water supply quantity per demand junction and water supply quantity.

Table 3 is results of basic statistical analysis of used parameters of Incheon, 135 DMAs were selected and data collection was done.

Table 2. Correlation analysis of main parameters.

Parameters	NRW Ratio	Pipe Length	Mean Pipe Diameter	No. of Demand Junctions	Pipe Length/Demand Junction	Water Supply Quantity/Demand Junction	Water Supply Quantity	Deteriorated Pipe Ratio	Demand Energy Ratio	No. of Leaks
NRW ratio	1.00	-	-	-	-	-	-	-	-	-
Pipe length	0.00	1.00	-	-	-	-	-	-	-	-
Mean pipe diameter	-0.11	0.09	1.00	-	-	-	-	-	-	-
No. of demand junctions	0.17	0.40	0.06	1.00	-	-	-	-	-	-
Pipe length/demand junction	-0.18	0.06	0.19	-0.29	1.00	-	-	-	-	-
Water supply quantity/demand junction	-0.13	-0.10	0.31	-0.25	0.36	1.00	-	-	-	-
Water supply quantity	-0.03	0.71	0.32	0.33	-0.09	0.12	1.00	-	-	-
Deteriorated pipe ratio	0.50	-0.06	0.05	0.25	-0.18	0.06	0.01	1.00	-	-
Demand energy ratio	0.16	-0.19	0.09	-0.05	-0.13	0.03	-0.05	0.05	1.00	-
No. of Leaks	0.36	-0.13	0.00	0.37	-0.26	-0.11	-0.05	0.50	0.14	1.00

Table 3. Data of parameters related to NRW ratio in water distribution systems.

Parameters	Minimum Value	Maximum Value	Mean Value	Standard Deviation
NRW ratio	0.00	64.29	19.24	13.29
Pipe length	0.70	52.50	8.80	6.37
Mean pipe diameter	89.57	288.51	160.70	36.18
No. of demand junctions	3.00	3824.00	806.56	679.54
Pipe length per demand junction	0.00	0.73	0.03	0.08
Water supply quantity per demand junction	0.49	454.67	11.19	43.17
Water supply quantity	28.00	12,340.00	2570.89	2055.70
No. of leaks	0.00	39.00	12.04	7.86
Deteriorated pipe ratio	0.00	46.10	11.39	10.08
Demand energy ratio	0.67	4.45	1.54	0.45

4.3. Selection of Main Parameters for Estimation of NRW Ratio

To analyze the correlation between the NRW ratio and the main parameters of water distribution systems, 135 DMAs were used excluding those unfinished, non-operating or abnormally operating among 367 DMAs of Incheon underwent multiple regression analysis. For this analysis, the number of demand junctions, pipe length, mean pipe diameter, water supply quantity per demand junction, number of leaks, deteriorated pipe ratio, demand energy ratio, pipe length per demand junction and water supply quantity were selected as independent variables in the multiple regression model, and the NRW ratio was selected as the dependent variable.

As a result of the multiple regression analysis using the stepwise selection method, the deteriorated pipe ratio (%), water supply quantity per demand junction ($\text{m}^3/\text{day}/\text{junction}$) and demand energy ratio (%) were selected under the condition that satisfied statistical significance (T-statistics and probability value are statistically satisfied). A multiple regression equation with three independent variables was thus derived for estimation of the NRW ratio. Table 4 shows the statistical results of all parameters used to estimate the NRW ratio using multiple regression analysis.

Table 4. Results of Multiple regression analysis using all parameters.

Parameters	Coefficient	Standard Error	T-Statistic	<i>p</i> -Value
NRW ratio	8.063	6.254	1.289	0.199
Pipe length	0.294	0.346	0.849	0.397
Mean pipe diameter	−0.041	0.032	−1.285	0.201
No. of demand junctions	−0.00062	0.0017	−0.349	0.727
Pipe length per demand junction	−2.005	14.866	−0.135	0.892
Water supply quantity per demand junction	−0.046	0.0411	−1.123	0.263
Water supply quantity	−0.00023	0.00083	−0.274	0.784
No. of leaks	0.211	0.164	1.284	0.201
Deteriorated pipe ratio	0.604	0.115	5.217	7.32×10^{-7}
Demand water ratio	4.253	2.204	1.929	0.055

In statistical hypothesis testing, the probability value (*p*-value) is the probability for a given statistical model that, when the null hypothesis is true, the statistical summary (such as the sample mean difference between two compared groups) is the same as or of higher than the measured results. If the *p*-value is higher than 0.05 and the T-statistic is lower than 1.196, this means it is not statistically significant [30].

Table 5 shows the results of multiple regression analysis with the NRW ratio as a dependent variable. This is considered reliable because the T-statistic of independent supply variables is more than ± 1.96 and the *p*-value is less than 0.05 [24].

Table 5. Results of multiple regression analysis using three parameters.

	Coefficient	Standard Error	T-Statistic	<i>p</i> -Value
NRW ratio (%)	4.684	3.397	1.379	0.170
Water supply quantity/demand junction	−0.069	0.032	−2.183	0.031
Deteriorated pipe ratio (%)	0.663	0.091	7.245	2.58×10^{-11}
Demand energy ratio	4.310	2.036	2.117	0.036

From the multiple regression analysis of Table 5, the regression equation of the NRW ratio can be defined as Equation (5). As the parameter affecting the NRW ratio, the deteriorated pipe ratio was 0.663, the demand energy ratio was 4.310, and the amount of water supply per demand junction 0.069. The value of each parameter is calculated according to Equation (5). In addition to these three parameters, the NRW ratio is fixed at 4.684 percent as the constant, and the ratios of deteriorated pipe

and demand energy are increasing parameters. The water supply quantity per demand junction is a decreasing parameter in the estimation of the NRW ratio.

$$y = 4.684 + 0.663x_1 + 4.310x_2 - 0.069x_3 \quad (5)$$

where, y is the NRW ratio (%), x_1 is the deteriorated pipe ratio (%), x_2 is the demand energy ratio (%), and x_3 is the amount of water supply per demand junction ($\text{m}^3/\text{day}/\text{junction}$)

As the demand energy ratio of DMAs in Incheon is calculated between 1 and 2 except for those on high elevation ground, it shows that the NRW ratio can be raised within 10% according to the energy ratio. In an area with high water supply such as apartment and dense population areas, the NRW ratio will decrease.

5. Estimation of NRW Ratio Using ANN

5.1. Model Construction of ANN

To estimate the NRW ratio using an artificial neural network (ANN), the results of multiple regression analysis were used to determine independent variables with the three parameters of the ratios of deteriorated pipe and demand energy and the water supply quantity per demand junction. The objective function was used to calculate the NRW ratio (%) via ANN. Figure 4 represents the constructed ANN model used in this study.

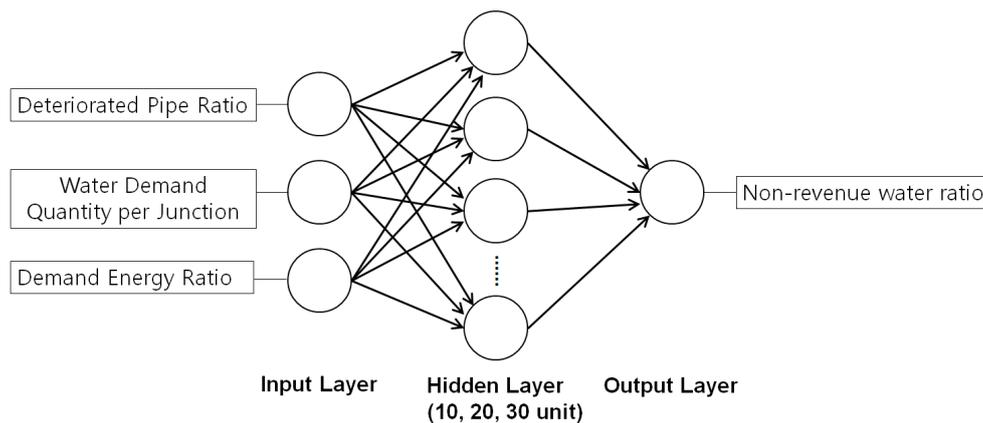


Figure 4. ANN model for analyzing the NRW ratio.

If many parameters are used, the problem of over-fitting could occur in ANN simulation, so the modeling case is made with a minimum number of parameters. An ANN simulation was performed by using 10, 20 and 30 neurons in the hidden layer.

5.2. Estimation of NRW Ratio via ANN

The ANN model was built using a single layer of an ANN structure and a back propagation algorithm. In the learning method of back propagation, an input signal to an input layer is transferred to hidden and output layers through the transfer function between layers. By comparing the transmitted signal with the desired one, the error between the target and learning values is determined in the final output layer. The error is again transmitted in the reverse direction and then the weight of each layer is updated.

This study implemented an ANN using the MATLAB program. A neural network toolbox was used in MATLAB and the Levenberg-Marquardt method of back propagation was used for training. This network training function updated weight and bias values according to the Levenberg-Marquardt optimization.

Figure 5 is the NRW ratio derived from ANN. The grey solid line shows the result of NRW by measurement, and the estimated NRW ratio of each DMA is shown when the number of neurons in the hidden layer is set to 10, 20 and 30, respectively. The measured NRW ratio was 0.5–58.9 percent, while the NRW ratio by ANN was estimated to be within 0.5–49.1 percent. The mean error rate was 18.4 percent for the measured NRW ratio and 19.3, 18.0 and 20.4 percent for the 10, 20 and 30 hidden layers, respectively. And the multiple regression equation showed the closest value of 18.5 percent.

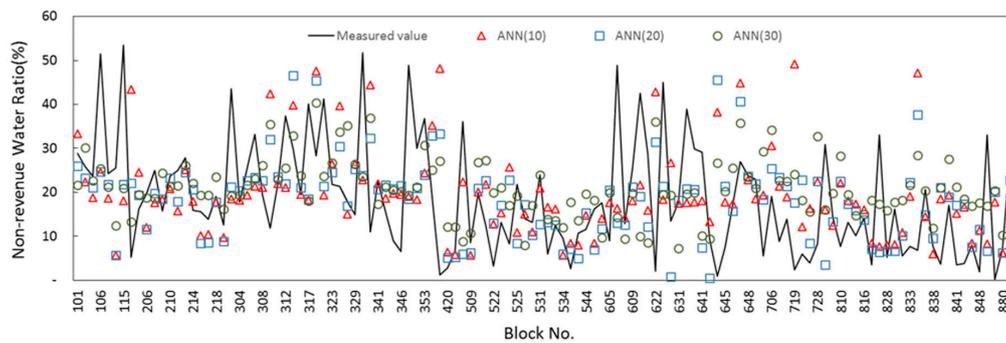


Figure 5. NRW ratio by artificial neural network (ANN) model simulation in each DMA.

5.3. Estimation of NRW Ratio Using ANN with Outlier Removal Case

The Z-score method can be used to distinguish the difference and distribution of the data used when conducting the result analysis. The Z-score is a dimensionless quantity obtained by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation [31]. This conversion process is called standardizing or normalizing. The mean and standard deviation are used to determine how far the data deviate from the average when the standard deviation is taken as a unit, and the method of Z-score is shown in Equation (6).

$$z = \frac{x - \mu}{\sigma} \quad (6)$$

where μ is mean of the population and σ is the standard deviation.

The outlier can be estimated through the Z-score method. The mean of the standardized Z-scores calculated is 0, and the standard deviation is 1. As a result, values above ± 3 are considered far away from the mean. In this study, the analysis was performed after excluding the DMA data for the parameter with the absolute value of the standardized Z-score of 3 or more among the main parameters of water distribution systems.

Finally, 122 sets of DMA data satisfying the Z-score among 135 sets of data were selected and used in the ANN analysis. Figure 6 show the results of the NRW ratio derived from the ANN estimated after excluding the abnormal value by the Z-score.

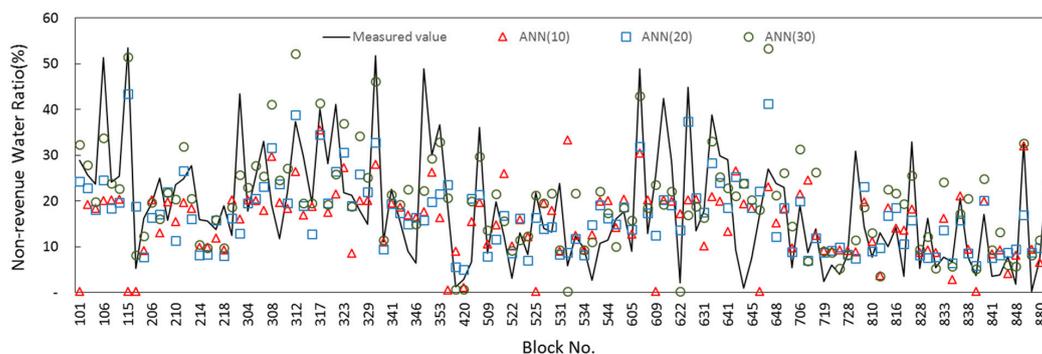


Figure 6. NRW ratio by ANN model simulation in each DMA with outlier removal condition.

ANN (10) shows a tendency toward underestimation than the measured NRW ratio, and part of the results largely deviate from measured values. Under the condition in which the outlier was removed, the higher the number of neurons, the higher the accuracy with the measured value.

5.4. Analysis of Estimation Results of NRW Ratio via ANN

To evaluate the accuracy of the multiple regression equations as proposed in the multiple regression analysis and the results of the ANN model developed in this study, an error ratio analysis was performed to evaluate the difference between the measured and model values. Accuracy analysis can be estimated by comparing the measured value with the value generated by the ANN model.

For this purpose, the mean absolute error (MAE), mean square error (MSE), PBIAS (percent of BIAS) which evaluates the bias of the estimation result, and the G-value which is represent the goodness of fit were used as prediction methods. The calculation method of each equation is shown in Equations from (7) to (10) [29], and the comparison between the measured and model estimation values can be more accurately evaluated through regression analysis.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n [|z(x_i) - \hat{z}(x_i)|] \quad (7)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n [z(x_i) - \hat{z}(x_i)]^2 \quad (8)$$

$$\text{PBIAS} = \frac{1}{n} \sum_{i=1}^n [z(x_i) - \hat{z}(x_i)] \quad (9)$$

$$G = \left(1 - \frac{\sum_{i=1}^n [z(x_i) - \hat{z}(x_i)]^2}{\sum_{i=1}^n [z(x_i) - \bar{z}]^2}\right) \times 100 \quad (10)$$

where $\hat{z}(x_i)$ is the estimated value at i and \bar{z} is the mean value of data.

If MAE and MSE are smaller, the estimated value is more accurate. If PBIAS is close to 0, the estimation result represents less bias. A G value of 100 is a perfect estimation. If the G value is negative, it is less reliable than using the average of data values as a predictor. MSE, MSE, PBIAS and G-value were used to verify the accuracy of the NRW ratio (%) estimated by the ANN. Table 6 shows the results of the NRW ratio's accuracy assessment by the ANN and the multiple regression equation.

Table 6. Accuracy evaluation by ANN and multiple regression analysis for estimating NRW ratio. Mean absolute error (MAE); mean square error (MSE); PBIAS (percent of BIAS).

	Multiple Regression Analysis	Original Data			Outlier Removed Data		
		ANN (10)	ANN (20)	ANN (30)	ANN (10)	ANN (20)	ANN (30)
PBIAS	−0.15	−0.97	0.40	−1.98	3.13	1.73	−1.28
MAE	8.33	8.14	7.84	8.76	9.26	7.55	7.03
MSE	115.84	114.14	104.17	125.20	155.02	97.18	86.95
G (Goodness of fit)	28.6	29.7	35.8	22.8	0.8	37.8	44.4

The ANN (20) with 20 neurons using original data satisfies the MAE, MSE and G-value and the resulting range of values closest to the measured NRW ratio (%). In the case of PBIAS, which shows the data's bias, the multiple regression equation shows the lowest value, indicating less biased results than others from the ANN. The ANN (30) with 30 neurons showed the highest accuracy among all assessment regulations of the PBIAS, MAE, MSE and G-value when the outlier was removed by the Z-score method. As a result of analyzing the data with the outlier removed by the Z-score, the accuracy of ANN (20) and ANN (30) was found higher than that of the original data used as a condition.

ANN (30) showed the highest accuracy among all results, and ANN (20) of the original data represented the least biased NRW ratio. Figure 7 shows the results of a scatter plot analysis of original data without using the Z-score method. The R^2 of the ANN model with 20 hidden layers was 0.3663 and the correlation coefficient was higher than the ANN model with 10 or 30 hidden layers and multiple regression analysis. These are the same results in Table 4 and the ANN model with 20 hidden neurons seems highly accurate.

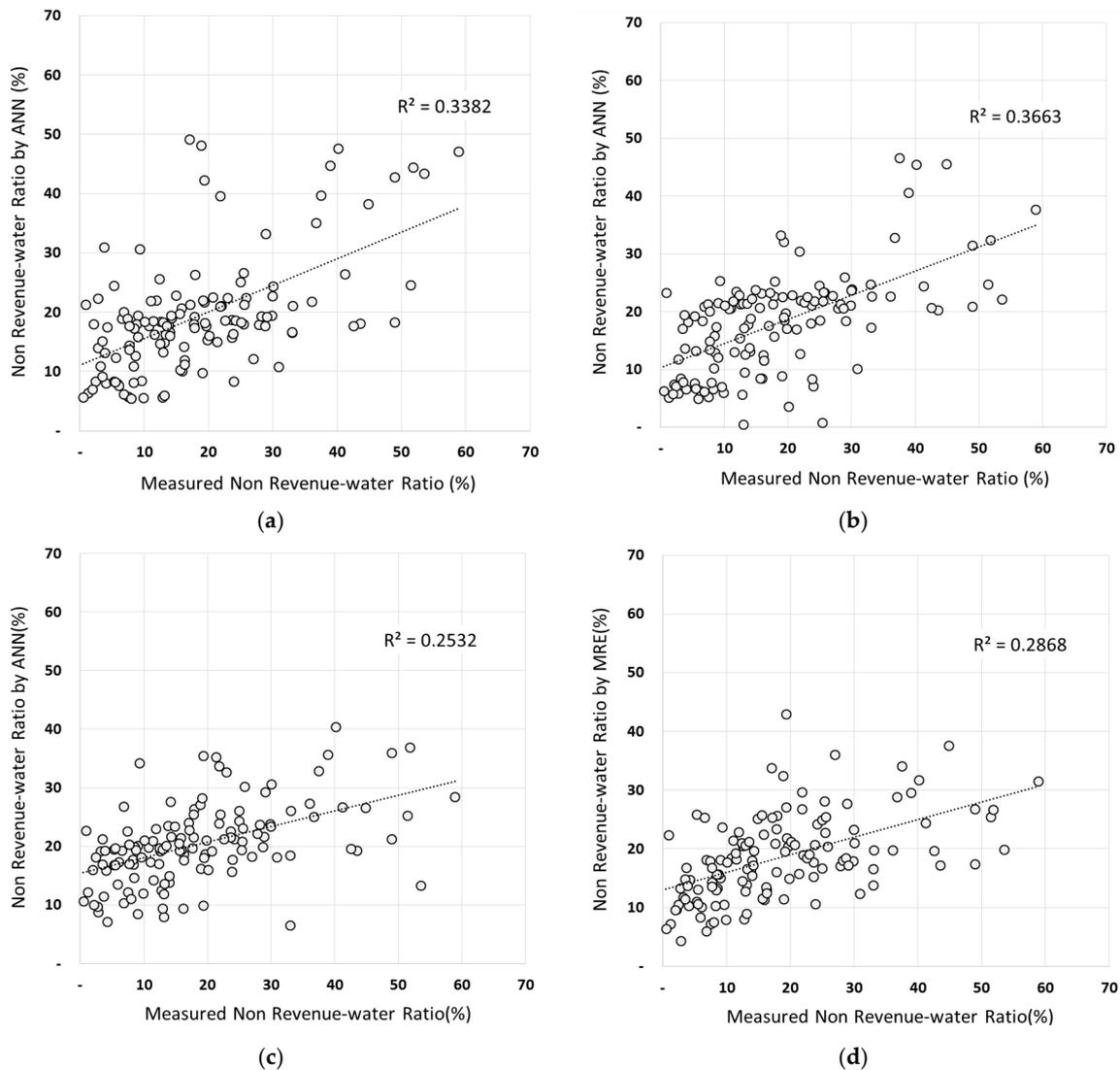


Figure 7. Scatter analysis results of NRW ratio: (a) ANN using 10 neurons in hidden layer; (b) ANN using 20 neurons in hidden layer; (c) ANN using 30 neurons in hidden layer; (d) Equation using multiple regression analysis.

Figure 8 shows the results after excluding the abnormal values using the Z-score method. The accuracy of the ANN model was found to be the most accurate under the condition of 30 hidden layers. An R^2 of 0.476 denotes high similarity than other neuron cases. In the case of the ANN model, six cases were used to estimate the NRW ratio, and the accuracy was high or low depending on the number of hidden layers compared with the multiple regression equation comparing the previous research [24].

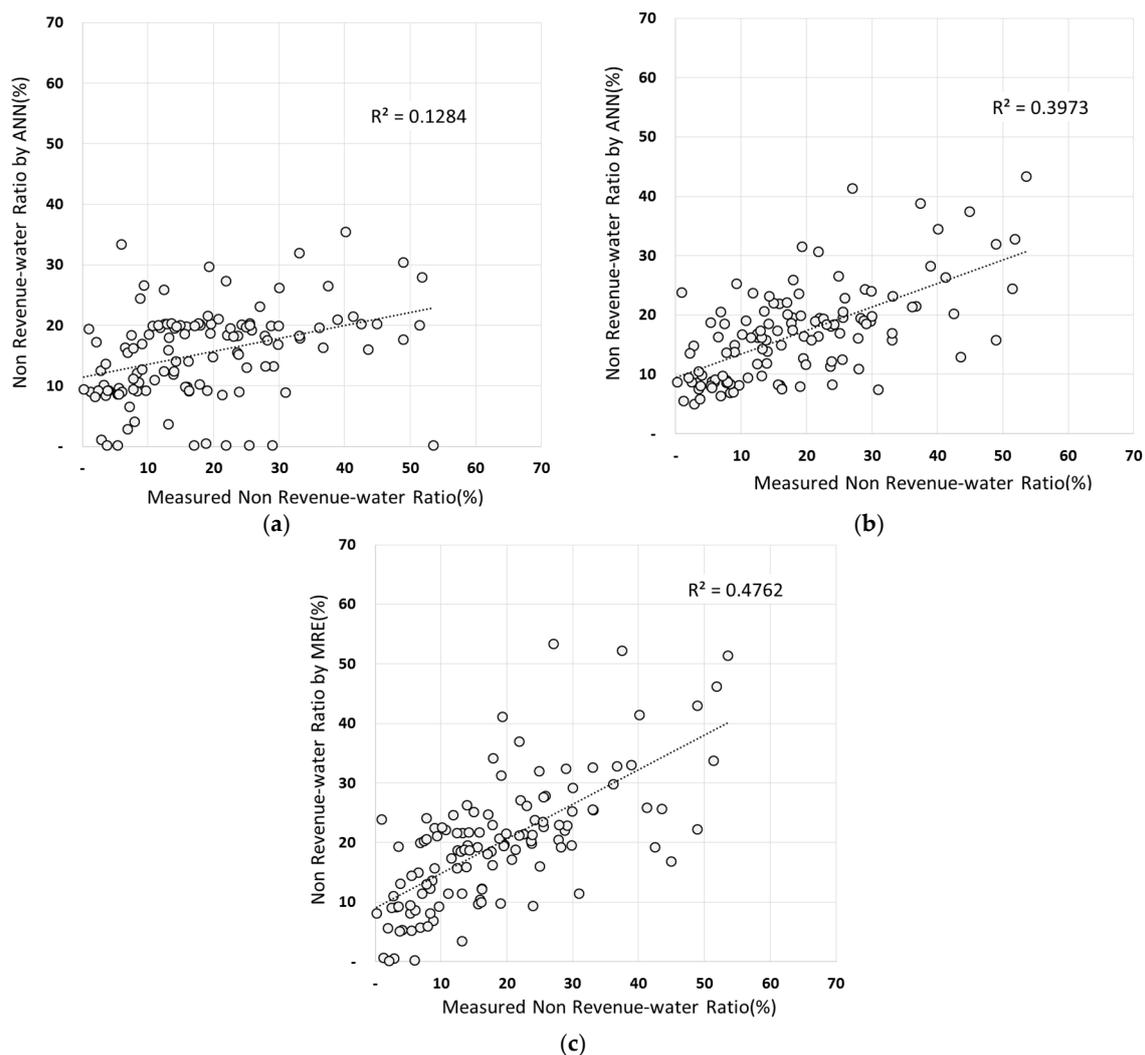


Figure 8. Scatter analysis results of NRW ratio with outlier remove condition: (a) ANN using 10 neurons in hidden layer; (b) ANN using 20 neurons in hidden layer; (c) ANN using 30 neurons in hidden layer.

6. Conclusions

The present study developed a model for estimating the NRW ratio using an ANN based on specific parameters affecting leaks in the water distribution systems of Incheon. Accuracy assessment and scatter plot analysis were used to select the optimal ANN model cases. The following conclusions were therefore drawn.

First, the estimation model for the NRW ratio was developed by an ANN in the water distribution systems of Incheon. In comparison with the multiple regression equation, the ANN-estimated NRW ratio was more accurate when the appropriate number of hidden layers was applied. Improvement of about 40 percent occurred compared with the NRW ratio derived from a multiple regression equation. This proves that the selected parameters such as water supply quantity per demand junction, deteriorated pipe ratio and demand energy ratio are valid for estimating the NRW.

Second, analysis of the outlier of independent variables is crucial when applying the ANN model. If the NRW ratio was applied to the ANN model by eliminating the outlier data through the Z-score method, the results of the NRW ratio would have been similar to the measured value than in cases in which the outlier data were not removed. The accuracy of NRW prediction can be improved through the accuracy and outlier verification of the collected data of each DMA.

Third, the optimal number of hidden layers is needed when estimating the NRW ratio via ANN. When developing the ANN model, this study set hidden layers with 10, 20 and 30 neurons. If the number of hidden layers is set up with more detailed numbers, however, more accurate results from an ANN can be expected.

The estimation model for the NRW ratio developed through this study can be applicable to the water distribution systems of Incheon. The development model is expected to help set the direction of improvement of the analysis of water distribution systems and the optimal operation of water supply and waterworks facilities for the construction of DMAs in Incheon. The model can also help enhance the revenue water ratio and diagnostic operation of water distribution systems.

Acknowledgments: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Education) (NRF-2014S1B2A1A01034684). The contents in this paper include the outcomes from HydroAsia 2016.

Author Contributions: Jang led the work performance and wrote the manuscript; Choi coordinated the research, contributed to writing the article.

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

References

1. Choi, G.W.; Jang, Y.G.; Lee, S.W. Effect of Estimation Method of Demand Water on the Analysis of Water Distribution System. In Proceedings of the Korea Water Resources Association Conference, Daejeon, Korea, 18–19 May 2006; pp. 1425–1430. (In Korean) Available online: [http://www.kwra.or.kr/wonmun/KWRA_4_2006_05_1425\(C\).pdf](http://www.kwra.or.kr/wonmun/KWRA_4_2006_05_1425(C).pdf) (accessed on 23 October 2017).
2. Wikipedia, Non-Revenue Water. Available online: https://en.wikipedia.org/wiki/Non-revenue_water#cite_note-11 (accessed on 15 September 2017).
3. Waterworks Headquarters Incheon Metropolitan City. *Waterworks Status*; Incheon Metropolitan City: Incheon, Korea, 2015; Available online: http://kosis.kr/statisticsList/statisticsList_02List.jsp?vwcd=MT_ATITLE01&parmTabId=M_02_01_01#SubCont (accessed on 23 October 2017).
4. Byeon, S.J. Development of Water Management System for Stable Water Supply in Isolated Region. Ph.D. Thesis, Incheon National University, Incheon, Korea, 2015.
5. Kanakoudis, V.K. Vulnerability based Management of Water Resources Systems. *J. Hydroinform.* **2004**, *6*, 133–156.
6. Park, S.W.; Kim, T.Y.; Lim, K.Y.; Jun, H.D. Fuzzy Techniques to Establish Improvement Priorities of Water Pipes. *J. Korea Water Resour. Assoc.* **2011**, *44*, 903–913. (In Korean) [[CrossRef](#)]
7. Park, Y.S. A Study on Long Term Replacement and Maintenance Plan for Multi-Region Water Pipelines Considering Economics. Master's Thesis, Seoul National University, Seoul, Korea, 2014. (In Korean)
8. Kanakoudis, V.K.; Tolikas, D.K. The role of leaks and breaks in water networks—Technical and economical solutions. *J. Water Supply* **2001**, *50*, 301–311.
9. Park, I.C.; Kwon, K.W.; Cho, W.C.; Cho, K.H. Study on the Decision Priority of Rehabilitation for Water Distribution Network Based on Prediction of Pipe Deterioration. In Proceedings of the Korea Water Resources Association Conference, Daejeon, Korea, 18–19 May 2006; pp. 1391–1394. (In Korean) Available online: [http://www.kwra.or.kr/wonmun/KWRA_4_2006_05_1391\(C\).pdf](http://www.kwra.or.kr/wonmun/KWRA_4_2006_05_1391(C).pdf) (accessed on 23 October 2017).
10. Kanakoudis, V.K.; Tolikas, D.K. Assessing the Performance Level of a Water System. *Water Air Soil Pollut.* **2004**, *4*, 307–318. [[CrossRef](#)]
11. Winarni, W. Infrastructure Leakage Index (ILI) as Water Losses Indicator. *Civil Engineering Dimension.* **2009**, *11*, 126–134.
12. Mukundi, M.J. Determinants of High Non-Revenue Water: A Case of Water Utilities in Murang' A County, Kenya. Master's Thesis, Kenyatta University, Kahawa, Kenya, 2014.
13. Shilehwa, C.M. Factors Influencing Water Supply's Non-Revenue Water: A Case of Webuye Water Supply Scheme. Ph.D. Thesis, University of Nairobi, Nairobi, Kenya, 2013.
14. Chung, S.H.; Lee, H.K.; Koo, J.Y.; Yu, M.J. Characterization of the ratio of revenue water in the 79 cities by Principal Component Analysis and Clustering Analysis. In Proceedings of the Korean Society of Environmental Engineers and Korea Water and Wastewater Works Association

- Joint Conference, Daejeon, Korea, 3–4 November 2004; pp. 133–142. (In Korean) Available online: http://kiss.kstudy.com/search/download.asp?ftproot=http://210.101.116.13/kiss3/inFTP_Journal.asp&inst_key=5012&a_img=07702915.pdf&isDownload=1&publ_key=29369 (accessed on 23 October 2017).
15. Kim, J.H.; Yoo, K.T.; Jun, H.D.; Jang, J.S. An Investigation of the Relationship between Revenue Water Ratio and the Operating and Maintenance Cost of Water Supply Network. *J. Korean Soc. Water Environ.* **2012**, *28*, 202–212. (In Korean)
 16. Shinde, V.R.; Hirayama, N.; Mugita, A.; Itoh, S. Revising the Existing Performance Indicator System for Small Water Supply Utilities in Japan. *Urban Water J.* **2013**, *10*, 377–393. [[CrossRef](#)]
 17. Vishwakarma, A. A Frontier Approach to Integrate Quality Parameters in Benchmarking Analysis of Water Supply Utilities. *Int. J. Adv. Res. Sci. Eng.* **2015**, *4*, 1187–1192.
 18. Gonelas, K.; Chondronasios, A.; Kanakoudis, V.; Patelis, M.; Korkana, P. Forming DMAs in a Water Distribution Network Considering the Operating Pressure and the Chlorine Residual Concentration as the Design Parameters. *J. Hydroinform.* **2017**, *19*. [[CrossRef](#)]
 19. Raman, H.; Chandramouli, V. Deriving a General Operating Policy for Reservoirs using Neural Network. *J. Water Resour. Plan. Manag.* **1996**, *122*, 342–347. [[CrossRef](#)]
 20. Jain, A.; Varshney, A.K.; Joshi, U.C. Short-Term Water Demand Forecast Modelling at IIT Kanpur Using Artificial Neural Networks. *Water Resour. Manag.* **2001**, *15*, 299–321. [[CrossRef](#)]
 21. Baxter, C.W.; Zhang, Q.; Stanley, S.J.; Shariff, R.; Tupas, T.; Stark, H.L. Drinking Water Quality and Treatment: The Use of Artificial Neural Networks. *Can. J. Civ. Eng.* **2001**, *28*, 26–35. [[CrossRef](#)]
 22. 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](#)]
 23. Msiza, I.S.; Nelwamondo, F.V.; Marwala, T. Water Demand Prediction Using Artificial Neural Networks and Support Vector Regression. *J. Comput.* **2008**, *3*, 1–8. [[CrossRef](#)]
 24. Jo, H.G.; Choi, G.W.; Jang, D.W. Development of the Non-revenue Water Analysis Equation through the Statistical Analysis of Main Parameter in Waterworks System in Incheon City. *Crisisonomy* **2016**, *12*, 63–75. (In Korean)
 25. Jang, Y.K. Demand Patterns Optimization of Water Storage Facilities for Saving Energy Costs in Water Distribution Systems. Ph.D. Thesis, Incheon National University, Incheon, Korea, 2013. (In Korean)
 26. Kang, M.W.; Kim, K.K.; Kim, B.Y.; Kim, Y.J.; Kim, Y.W.; Yeo, I.K.; Lee, U.Y.; Hwang, S.Y. *Introduction to Statistics*; Freedom Academy: Chicago, IL, USA, 2013. (In Korean)
 27. Gwak, J.M. *Research and Statistical Analysis*; Informa: Seoul, Korea, 2013. (In Korean)
 28. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Macmillan Publishing Company: New York, NY, USA, 1994.
 29. Waterworks Headquarters Incheon Metropolitan City. *Technical Diagnostics Report for Re-establish Basic Plan of Waterworks Maintenance in Incheon Water Distribution Network*; Incheon Metropolitan City: Incheon, Korea, 2015; (In Korean). Available online: http://waterworksh.incheon.kr/cop/bbs/selectBoardArticle.do?menuNo=1020600&bbsId=BBSMSTR_00000000113&nttId=2&bbsTyCode=BBST01&bbsAttrbCode=BBSA03&authFlag=Y&pageIndex=4&searchCnd=0&searchWrd=&searchCtgr=&searchEtcFld1= (accessed on 24 October 2017).
 30. Wasserstein, R.L.; Lazar, N.A. The ASA’s Statement on *p*-Values: Context, Process, and Purpose. *Am. Stat.* **2016**, *70*, 129–133. [[CrossRef](#)]
 31. Kreyszig, E. *Advanced Engineering Mathematics*, 4th ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1979.

