

Article

# An Efficient Burst Detection and Isolation Monitoring System for Water Distribution Networks Using Multivariate Statistical Techniques

KiJeon Nam <sup>1,†</sup> , Pouya Ifaei <sup>1,†</sup>, Sungku Heo <sup>1</sup>, Gahee Rhee <sup>1</sup>, Seungchul Lee <sup>2,\*</sup> and ChangKyoo Yoo <sup>1,\*</sup>

<sup>1</sup> Department of Environmental Science and Engineering, College of Engineering, Kyung Hee University, Seocheon-dong 1, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Korea; spirit1058@gmail.com (K.N.); pooya\_if@hotmail.com (P.I.); energizer@khu.ac.kr (S.H.); gahee.re@gmail.com (G.R.)

<sup>2</sup> Technology Development Center, Samsung engineering, Woncheon-dong, Yeongtong-gu, Suwon-si, Gyeonggi-do 446-701, Korea

\* Correspondence: sangchu.lee@samsung.com (S.L.); ckyoo@khu.ac.kr (C.Y.); Tel.: +82-31-260-6021 (S.L.); +82-31-201-3824 (C.Y.)

† The first and second authors contributed equally to this paper.

Received: 29 April 2019; Accepted: 19 May 2019; Published: 24 May 2019



**Abstract:** Detection and isolation of burst locations in water distribution networks (WDN) are challenging problems in urban management because burst events cause considerable economic, social, and environmental losses. In the present study, a novel monitoring and sensor placement approach is proposed for rapid and robust burst detection. Accordingly, a hybrid principal component analysis (PCA) and standardized exponential weighted moving average (EWMA) system is proposed for WDN monitoring and management. In addition, the optimal sensor configuration is obtained using PCA, *k*-means clustering, and a sensitivity analysis considering the diurnal patterns and the noises of pressure and flowrate data in the WDN. The proposed system is applied to a branched WDN, and the results are compared to those obtained with conventional monitoring systems. The results show that the proposed system detected the burst occurrence regardless of noise size with a detection rate of 93%. Compared to conventional systems, the isolation ratio improved by 10%, indicating that the bursts were isolated more accurately. In addition, the corresponding sensor configuration was 40% less expensive than the conventional systems.

**Keywords:** burst isolation; burst monitoring; optimal sensor location; pipe burst; water distribution network

## 1. Introduction

Public access to high-quality freshwater has accounted for successful development in the last century. With rapid urbanization, water distribution networks (WDNs) spread over cities to supply freshwater to residential, industrial, and commercial districts. Over time, design approaches for water distribution networks improved, and more efficient facilities were employed in complicated WDNs [1]. The quality of the end-pipe product was also controlled more strictly by environmental departments. However, in the present century, monitoring of complicated water networks has become a serious concern as WDNs age. Thus, advanced monitoring techniques need to be developed to meet the contemporary challenges in WDN maintenance.

One of the most common problems in WDN maintenance is a burst event in pipelines [2]. Pipeline bursts, which can be caused by aging, corrosion, and deterioration of facilities, lead to significant economic, social, and environmental costs. The burst causes negative effects on the conveying pipes

and other instruments of WDN, and this is directly or indirectly associated with additional economic costs due to water loss, diagnosis and repair of the WDN, and interruption in the water supply [3,4]. In the aspect of the social costs, a pipe rupture in a WDN results in potable water loss and disruptions of customer service [5]. Moreover, corrosion in the surrounding soil, leakage to underground water resources, and risks to public health are environmental impacts of water leakage due to pipe burst [6]. Thus, prevention of water loss caused by bursts in WDNs has become a critical challenge in urban management in the recent two decades [7].

Rapid detection of bursts is a promising remedy to the problem because it reduces the above-mentioned costs [2]. A burst can occur for several reasons, including poor pipe conditions, inappropriate operation of the system, extreme weather conditions, and extraordinary environmental pressure. Hence, detecting the exact time and location of the burst is a sophisticated problem involving complicated factors. Conventional burst detection and isolation techniques are often lengthy procedures. In the conventional methods, a candidate area of analysis is uncovered by digging after visual inspection. This conventional technique entails extra costs due to the associated ground digging, traffic jams, and water losses [4,8]. Therefore, research has been recently devoted to developing faster and more accurate burst detection approaches to decrease these costs [8].

Statistical computation techniques based on observation of transient changes in recorded data have been extensively implemented to detect bursts. The statistical techniques extract information from a system, monitor data variations, and continuously reflect dynamic conditions of the system [9]. Misiunas et al. suggested pressure-based cumulative sum (CUSUM) control charts to detect burst occurrence. They showed that outliers of the CUSUM indicated that pressure drop was induced by burst occurrence [10]. Palau et al. used a principal component analysis (PCA) to extract key information from flow data sets measured by an supervisory control and data acquisition (SCADA) system. Control charts driven by the PCA identified anomalous behavior in the extracted flow data [11]. Loureiro et al. implemented a region-based outlier statistical process control method to detect abnormal parameters in flowrates [12]. It should be noted that monitoring burst occurrence using only pressure or flowrate in the pipelines does not lead to rapid detection [13]. In this way, a multi-objective optimization sectorization method was proposed that considered the hydraulics, water quality, and economic factors together [14]. Romano et al. obtained significantly accurate results on leakage detection using statistical process control [15]. While their univariate statistical technique was easy to apply, a single variable was not sufficient to interpret the complex problem [16]. However, a multivariate monitoring system can be used to overcome the limitations of the univariate burst monitoring systems and to improve accuracy on burst detection [17].

When a burst is detected with a monitoring system, the burst location should be identified to hinder water leakage prior to repair of the WDN. However, WDN-related detection methods mainly concentrate on finding optimal sensor placements to develop early warning systems (EWSs) and improve data quality [18,19]. A Fisher Discriminant Analysis was applied to identify the leakage location by sensor measurements. The results were satisfactory in a case study. However, the burst monitoring system could not provide comprehensive guidelines to detect burst locations [17,20]. Thus, burst monitoring systems should be developed to not only detect accurate burst occurrence time, but also to isolate the exact burst location. Sarrate et al. proposed sensor locations for leak detection using a structural model and a clustering technique [21]. Costanzo et al. suggested a model calibration tool to monitor and isolate sensor's location using multivariate data [22]. While both temporal and local detection were considered in their approach, hydraulic fluctuations were restricted. Pressure and flowrate time series have diurnal patterns as a result of diurnal water consumption patterns in the WDN. Moreover, the daily patterns of pressure and flowrate obtained from nodes are similar because of the geometric location of the nodes and demands in a WDN. Due to these complex dynamic patterns in WDNs, a burst isolation system with optimal sensor location is necessary.

The aim of the present study was to detect burst occurrence, determine optimal sensor configuration, and isolate burst location in a comprehensive monitoring and maintenance system.

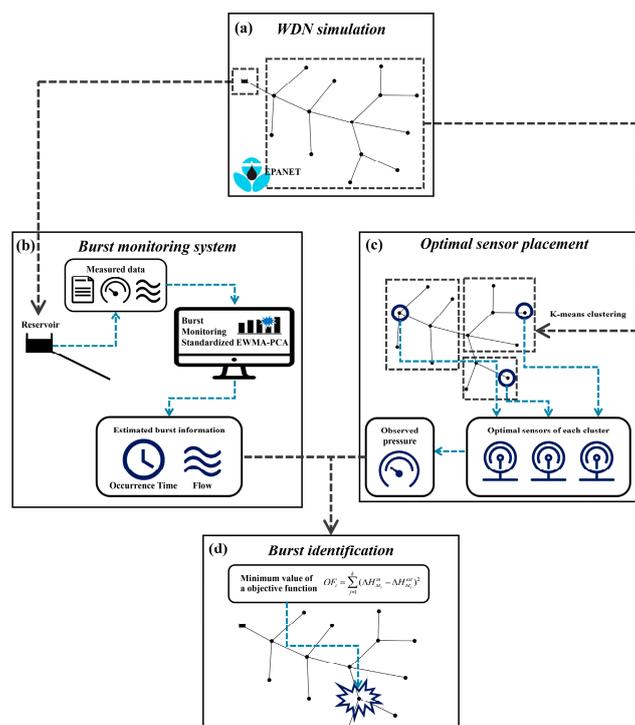
Accordingly, multivariate statistical and analytical techniques were applied to detect burst occurrence using flowrate and pressure data simultaneously. The results of the multivariate monitoring system were compared to those obtained by conventional univariate monitoring systems. The novel multivariate monitoring system consisted of standardized exponential weighted moving average (EWMA) and PCA to overcome the weaknesses of the conventional monitoring systems. The optimal sensor configuration was detected using the PCA and a  $k$ -means clustering algorithm. A sensitivity analysis was conducted to consider hydraulic and diurnal characteristics of the WDN. Finally, the proposed monitoring system and optimal sensor locating were interlinked to isolate the burst location in the WDN. The comprehensive system was compared with conventional systems using burst scenarios for the simulated WDN.

The present paper consists of four major parts. First, we present a branched WDN simulation that verifies monitoring systems and optimizes sensor configurations. Second, a novel standardized EWMA-PCA monitoring system is detailed for burst detection in the WDN. Two conventional systems using CUSUM charts and standardized EWMA were employed to compare the performance of the novel methodology in 10 burst scenarios. Third, optimal sensor configuration using PCA,  $k$ -means approach, and sensitivity analysis is discussed. The results are then compared with those of other methods. Fourth, a burst identification approach that minimizes an objective function in a mathematical model is presented.

## 2. Materials and Methods

### 2.1. Research Framework

One of the goals of this study was to obtain optimal sensors' location by detecting and isolating a burst event in a WDN for continuous monitoring of the burst and maintaining a robust WDN. For this, a WDN needed to be simulated to verify the novel monitoring approach. Accordingly, the research framework is graphically shown in Figure 1. The proposed method can be divided into three parts: WDN simulation, detection of burst occurrence and determination on optimal sensor location, and identification of the node in which the burst occurs.



**Figure 1.** Research framework for an efficient burst detection and isolation monitoring system.

The first step was WDN modelling to simulate burst occurrence in pipes, as shown in Figure 1a. For this, EPANET software, which is widely used for WDN simulation, was employed. The EPANET is a powerful and effective tool to simulate natural and engineering water systems such as a WDN. Accurate simulation framework of the EPANET facilitates its application for simple networks on flat terrains. Therefore, it has been globally implemented in the field of design and operational analysis of WDN [23,24]. The EPANET simulated the modeled WDN for an operational period of two days (48 h) in one-minute time intervals to reflect diurnal variations in the WDN. Ten scenarios were generated during the simulation to validate robust performance of the proposed system. Pressure and flowrate were continuously monitored in the proposed system at the entrance of the modeled WDN. When a burst event occurs, a hybrid standardized EWMA-PCA system at the entrance of the WDN detected the event based on monitored data. Then, the system started estimating the burst information (occurrence time of the burst and discharge flowrate in the pipe), as shown in Figure 1b. According to the monitoring system responses, the optimal pressure sensor location was determined with respect to the greatest sensitivity of responses to variations of pressure. A *k*-means clustering algorithm was applied in this step and is graphically shown in Figure 1c. The optimal pressure sensors were selected in each cluster with common characteristics of the WDN on considered nodes. The third step, shown in Figure 1d, was identification of the burst location based on the information obtained in the second step. The pressure sensors found to obtain the burst location comparing measured pressure by the sensors and the estimated pressure by the monitoring system. All nodes were burst location candidates in the WDN. These nodes were individually calculated to locate the burst using an objective function. The node with the smallest difference between measured and estimated pressure was selected as the burst location. Since a burst causes pressure decrease in the WDN, the variation in pressure was detected by the pressure sensors, and the estimated pressure drop from the selected node by the monitoring system coincided with the measured pressure drop on an actual busted node. The performance of the proposed standardized EWMA-PCA-based monitoring system equipped with *k*-means clustering for burst detection and burst isolation was compared with other monitoring systems. Two systems were applied for comparative performance evaluation with the proposed system. In the first system, a cumulative sum (CUSUM) chart was applied for monitoring, and a sensitivity analysis was used for burst detection. In the second system, standardized EWMA and *k*-means clustering were applied for system monitoring and burst detection, respectively.

## 2.2. Burst Monitoring System

The proposed hybrid standardized EWMA-PCA system was anticipated to enhance the burst detection performance because the standardized EWMA detects slight variations in the process, and the PCA considers the correlations between pressure and flowrate simultaneously to detect burst occurrence. First, measured data at the entrance of the WDN were filtered by mean trajectory removal to eliminate strong diurnal characteristics of flowrate and pressure. Second, the filtered data were converted to weight average values  $z_i$  using the standardized EWMA. Third, burst occurrence was detected using a Hotelling  $T^2$  chart of the PCA model according to the weighted average values  $z_i$ .

### 2.2.1. Standardized Exponential Weighted Moving Average (Standardized EWMA)

An EWMA is an advanced statistical tool to monitor small shifts in processes [25]. It uses both the information in the last sample observation and any information in the previous sample observations based on a weighting factor. The weighting factor is iteratively multiplied by sample observations in terms of time to represent reduced significance related to the present sample observation. EWMA is mathematically represented as follows [26]:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1}, \quad (1)$$

$$z_0 = \mu_0 \text{ or } z_0 = \bar{x}, \quad (2)$$

where  $z_i$  is the weighted average of the measured data at time  $i$ ,  $x_i$  is the current data,  $z_{i-1}$  is the previous weighted average, and  $\lambda$  is the weight of the EWMA that varies between 0 and 1. 0.4 was selected as the weight value of the EWMA since this value was recommended for the EWMA using the  $3\sigma$  [27]. At time step 0, it is assumed that a population mean ( $\mu_0$ ) or a sample mean ( $\bar{x}$ ) is equal to the first weighted average value ( $z_0$ ). For the standardized EWMA, the sample ( $x_i$ ) is normalized before calculating the weighted average  $z_i$  using Equation (3).

$$\text{normalized } x_i = \frac{x_i - \mu_0}{\sigma}, \quad (3)$$

where  $\sigma$  is the standard variation of the samples. The standardized EWMA is used to generate input data from measured pressure and flowrate quantities, and the calculated weighted average value is used as the input of the burst monitoring system.

### 2.2.2. Principal Component Analysis

Principal component analysis (PCA) is a statistical linear transformation that is widely used in data analysis, model compression, and multivariate process monitoring [28]. The dimensionality of the data is reduced in the PCA to form a meaningful data structure [29]. In PCA, the data matrix,  $X$ , is decomposed into a number of principal components (PCs) that are new axes to explain variance-covariance in the data. The data are divided into two parts; the first part is explained by the PCA and represents the sum of outer production of vector  $t_i$  by vector  $p_i$ . The second part is the residual part ( $E$ ) [9].

$$X = TP^T + E = \sum_{i=1}^m t_i p_i^T + E, \quad (4)$$

where  $m$  is the number of PCs determined by cumulative percent variance,  $t_i$  is the orthogonal score vector that contains information on the relationship between observations, and  $p_i^T$  is the orthonormal loading vector that includes information on the relationship of variables. To monitor a multivariate process in accordance with time, a  $T^2$  chart is generally used because it converts multivariate data into statistical values ( $T^2$  statistic).  $T^2$  is given in Equation (5).

$$T^2 = x^T P \Lambda^{-1} P^T x = x^T D x, \quad (5)$$

where  $\Lambda$  is the diagonal matrix of the eigenvalues. The lower confidence limit for  $T^2$  is equal to zero, and the upper confidence limit is obtained based on  $\chi^2$  with degree of freedom  $d$  and probability  $\alpha$ , as presented in Equation (6).

$$UCL = \chi_{\alpha}^2(d) \quad (6)$$

Here, a  $T^2$  chart is used to monitor pressure and flowrate measured at the entrances to detect burst occurrence in the WDN. The  $T^2$  chart reduces detection time based on EWMA as an input variable to the PCA, and it can determine the relationship between induced variations of pressure and flowrate caused by the burst.

### 2.3. Optimal Pressure Sensor Placement

Optimal sensor location should address similarities of pressure variations at adjacent nodes. In addition, the sensor location should be sensitive enough to represent neighboring nodes. Therefore, in this study, a sensitivity analysis-based optimal pressure sensor placement was performed using clusters. A  $k$ -means clustering algorithm was employed to classify the WDN. Here, the  $k$ -means algorithm clusters the WDN into several classes that contain nodes with similar pressure and flowrate. The nodes in the same cluster represent candidates for optimal sensor location. Subsequently, the node that has the highest sensitivity value in the cluster was selected as the optimal sensor location.  $k$ -means clustering formed groupings within the WDN using seven variables obtained at each node: three

parameters of the second-order regression model, mean and standard deviation of flowrate, and mean and standard deviation of pressure. Three parameters of the second-order regression model were used to characterize patterns of pressure and flowrate, which have a second order relationship represented by Bernoulli's Equation [19]. The second-order regression model and parameters  $p_1$ ,  $p_2$ , and  $p_3$  are expressed in Equation (7).

$$pressure = p_1 \cdot flow\ rate^2 + p_2 \cdot flow\ rate + p_3 \quad (7)$$

Since the seven data sets had different scale and variation, the  $k$ -means clustering algorithm was not efficient. The mean values of flowrate and pressure had a large scale and could be the major influences on similarity among the nodes. That is why a pretreatment method was required to normalize the data and reduce dimensionality of the input data. In this study, PCA was used to modify the input data to the  $k$ -means algorithm.

### 2.3.1. K-Means Clustering Approach

Clustering analysis is a method to identify natural characteristics of data based on similarities among them. It is extensively used in various research fields, including statistics, engineering, and electronics [30,31]. There are two major types of clustering algorithms: Hierarchical clustering and partitioned clustering. Hierarchical clustering obtains nested clusters by merging adjacent data points into larger clusters or by separating each cluster into smaller clusters continuously [32]. On the other hand, partitioned clustering can obtain all clusters simultaneously. Among the portioned clustering methods, the  $k$ -means clustering algorithm is a mature and effective clustering approach [30,31]. The  $k$ -means algorithm is a well-known approach that clusters the observations or variables according to the highest similarities and lowest dissimilarities among them [33,34].  $k$ -means clustering starts by selecting a random number of centroids for desired clusters. Then, all data are allocated to the closest clusters based on Euclidian distance from centroids. These two steps are iterated to minimize the objective function given in Equation (8).

$$J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2, \quad (8)$$

where  $x_i$  ( $i = 1, \dots, n$ ) is an observed dataset that has a  $d$ -dimensional vector,  $\mu_k$  is the mean of cluster  $c_k$ ,  $c_k$  ( $k = 1, \dots, K$ ) is an index of each cluster, and  $K$  is the number of clusters. In this study,  $k$ -means clustering analysis was used to classify the data, and the number of groups was equal to the number of desired sensors based on the similarity of hydraulics at the nodes. Thus, the clustering approach was referred to as a supervised method in which the final number of clusters was initially determined. Based on the results of clustering, the most sensitive node in each cluster was selected as the optimal sensor location that represented the dynamics of pressure and flowrate in the cluster.

### 2.3.2. Sensitivity Analysis

The sensitivity of the nodes with respect to any possible burst location in the WDN was derived using the estimated average value of the burst flowrate, as given in Equation (9).

$$S_{i,j} \equiv \left| \frac{\delta H_j}{\delta Q_{D,i}} \right| = \left| \frac{H_j(Q_{D,i}) - H_j(Q_{D,i}^*)}{Q_{D,i} - Q_{D,i}^*} \right| \quad \forall i \in [1, N], \quad \forall j \in [1, N], \quad (9)$$

where  $H_j(Q_{D,i})$  is the simulated head at node  $j$  for the demand discharge  $Q_{D,i}$  at node  $i$ ,  $H_j(Q_{D,i}^*)$  is the simulated head at node  $j$  for  $Q_{D,i}^*$  is the sum of  $Q_{D,i}$  and estimated burst flow rate, and  $N$  is the number of nodes in the WDN. The variable  $Q_{D,i}$  was assumed to be the average flowrate according to the daily

demand pattern. Then, the cumulative sensitivity of each node was calculated as an indicator of the system's sensitivity, as shown in Equation (10).

$$S_i = \sum_{j=1}^N (S_{i,j})^2 \quad \forall i \in [1, N], \quad \forall j \in [1, N] , \quad (10)$$

where  $S_i$  is the cumulative sensitivity of node  $i$ . The optimal pressure sensor location in each cluster was determined as the node that had the greatest amount of cumulative sensitivity.

#### 2.4. Burst Identification

As the burst was detected by the proposed monitoring system, the burst flowrate and occurrence time were estimated. At the same time, the pressure was measured by the selected optimal pressure sensors. Based on the estimated flowrate and the related burst information obtained from the monitoring system, the pressure ( $H_{M_i}^{est}$ ) was estimated assuming that the burst could occur at all candidate nodes. Thus, the node minimizing the objective function given in Equation (11) was identified as the burst location [18].

$$OF_i = \sum_{j=1}^k (\Delta H_{M_i}^m - \Delta H_{M_i}^{est})^2, \quad (11)$$

where  $i$  is the node number,  $\Delta H_{M_i}^m$  is the measured pressure change recorded by the pressure sensor,  $\Delta H_{M_i}^{est}$  is the estimated pressure change obtained by the monitoring system, and  $k$  is the number of pressure sensors.

### 3. Results and discussion

#### 3.1. Simulation of a Water Distribution Network

A branched water supply network in a single district metering area (DMA) was simulated to evaluate the performance of the proposed system. A layout of the WDN is graphically shown in Figure 2. The WDN consists of 14 nodes, 14 pipes, and one reservoir. The pipe diameters and lengths varied between 150 mm and 400 mm and between 300 and 1200 m, respectively. The cast-iron pipes were assumed to have a Hazen-Williams roughness coefficient of 100. The node elevation varied from 105 m to 125 m, and the demand discharge fluctuated between 86 m<sup>3</sup>/day (0.9954 l/s) and 173 m<sup>3</sup>/day (2.0023 l/s). The total water demand of the network was supplied from a fixed head reservoir at 250 m altitude.

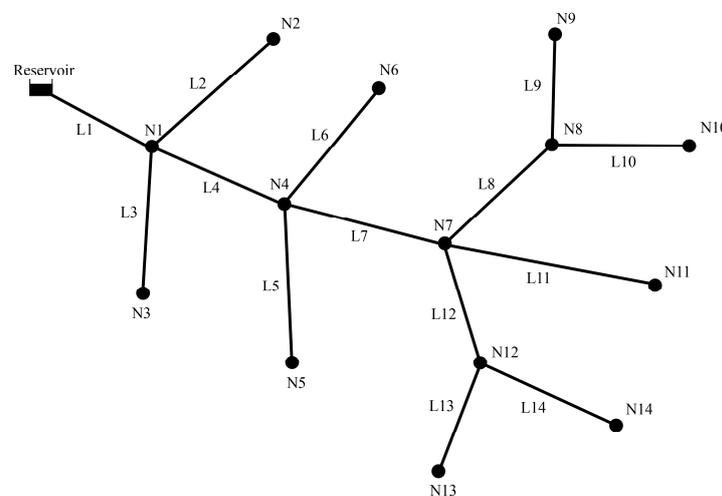
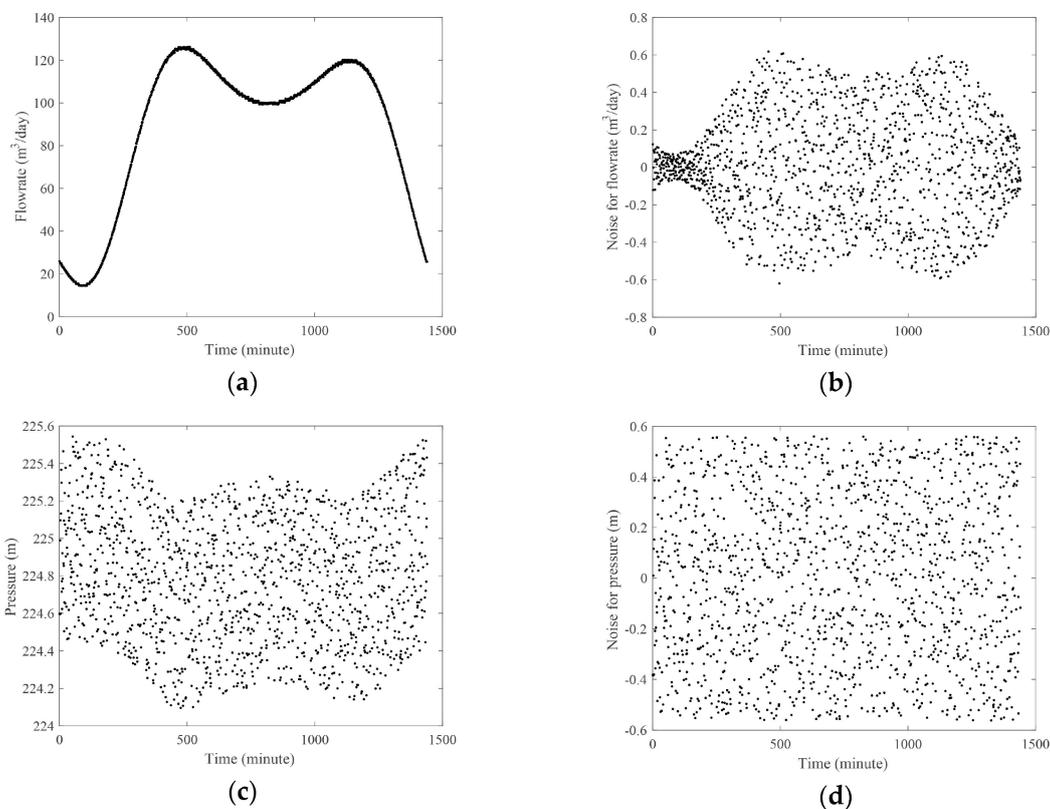


Figure 2. A schematic representation of the water distribution network.

Specific properties of the WDN are summarized in Table 1. The daily pattern of demand load at the nodes of the WDN was a function of living pattern. The periodic daily flowrate and pressure at node 1 are shown in Figure 3a,c, respectively. According to Figure 3a, the diurnal flowrate reaches two peak points: in the morning between 06:00 and 09:00 and in the evening between 18:00 and 22:00. The peak hours result in pressure drop in the WDN, which can be seen in Figure 3c. Since the measurement equipment had noise, signal, and noise ratios were assumed to be 0.5% and 0.25% of the flowrate and pressure data, respectively. The diurnal flowrate diagram and pressure diagram (including noise) are shown in Figure 3b,d, respectively.

**Table 1.** Properties of nodes and links in the simulated WDN.

| Node Number | Elevation (m) | Demand (m <sup>3</sup> /day) | Link | Length (m) | Diameter (mm) |
|-------------|---------------|------------------------------|------|------------|---------------|
| N1          | 125           | 86                           | L1   | 1200       | 400           |
| N2          | 120           | 130                          | L2   | 1400       | 150           |
| N3          | 121           | 86                           | L3   | 500        | 150           |
| N4          | 120           | 86                           | L4   | 700        | 350           |
| N5          | 110           | 173                          | L5   | 400        | 150           |
| N6          | 116           | 86                           | L6   | 400        | 125           |
| N7          | 117           | 86                           | L7   | 600        | 350           |
| N8          | 115           | 173                          | L8   | 300        | 250           |
| N9          | 110           | 173                          | L9   | 400        | 200           |
| N10         | 111           | 130                          | L10  | 500        | 200           |
| N11         | 114           | 173                          | L11  | 400        | 200           |
| N12         | 110           | 173                          | L12  | 400        | 250           |
| N13         | 105           | 173                          | L13  | 350        | 200           |
| N14         | 110           | 86                           | L14  | 500        | 150           |



**Figure 3.** Periodic daily patterns at node 1: (a) flowrate, (b) noisy flowrate, (c) pressure, and (d) noisy pressure.

### 3.2. Monitoring of Burst Occurrence in the WDN

To evaluate the proposed system, 10 burst scenarios were generated with varying size and occurrence time. The flowrate, location, time, and duration of the bursts are summarized in Table 2. All scenarios were simulated using EPANET software in a 48 h period with respect to the diurnal demand pattern. The burst size varied from 10 m<sup>3</sup>/day to 100 m<sup>3</sup>/day, and the burst location, occurrence time, and duration were randomly selected. The burst events occurred two times in the first and the second 24 h during the simulation at the same time interval. As an instance, the burst appeared at 384 min (in the first 24 h) and 1824 min (in the second 24 h) in scenario 6. Table 3 shows the three monitoring and detection systems applied in this study, as described in Section 2. System 1 employed CUSUM and a sensitivity analysis, as detailed in [10]. The sensitivity analysis was used to measure a perturbation in demand changes at each node, and the most sensitive node was selected as the location of the burst sensor. In system 1, the burst was detected by the CUSUM charts using time series data sets. The CUSUM chart detected demand fluctuations in the water distribution network. In system 2, *k*-means clustering and a sensitivity analysis were employed to determine the sensor location, and standardized EWMA charts were used to detect burst occurrence. The standardized EWMA diagnosed the burst using the upper and lower control limits of the algorithm. In system 3, the PCA was used, and the results were compared to system 2. The PCA was applied to pre-treat the data used for sensor location and to monitor the burst according to the data converted by the standardized EWMA. System 3 was compared to systems 1 and 2 in terms of efficiency and robustness. Performance evaluation involved detection and isolation of burst occurrence, and the robustness was evaluated by allocating the signal and noise ratios.

**Table 2.** Burst occurrence scenarios.

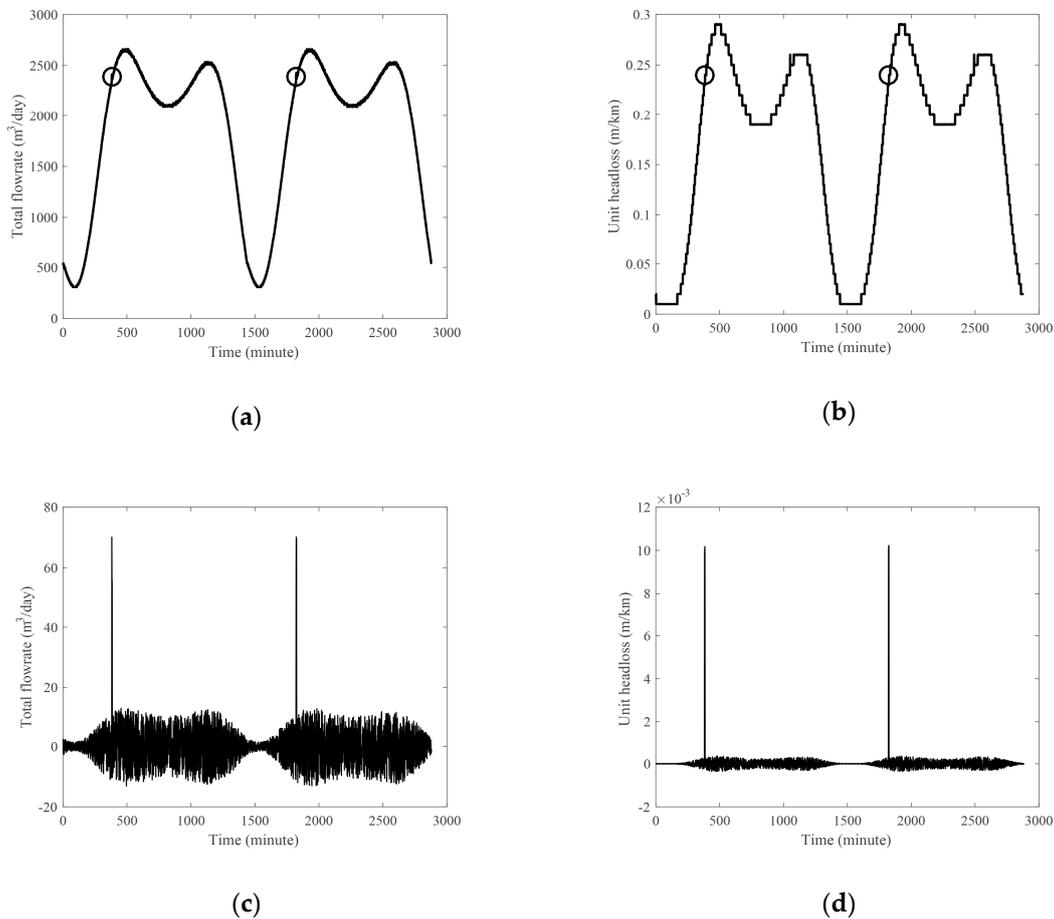
| Scenario | Burst Flowrate (m <sup>3</sup> /day) | Burst Location | Occurrence Time (min) | Duration (min) |
|----------|--------------------------------------|----------------|-----------------------|----------------|
| S1       | 10                                   | N14            | 902                   | 2              |
| S2       | 20                                   | N6             | 685                   | 5              |
| S3       | 30                                   | N7             | 551                   | 7              |
| S4       | 40                                   | N10            | 62                    | 8              |
| S5       | 50                                   | N2             | 589                   | 7              |
| S6       | 60                                   | N4             | 384                   | 4              |
| S7       | 70                                   | N7             | 348                   | 3              |
| S8       | 70                                   | N14            | 540                   | 7              |
| S9       | 80                                   | N1             | 744                   | 4              |
| S10      | 100                                  | N7             | 980                   | 5              |

**Table 3.** Three burst monitoring and sensor placement systems.

| System   | Burst Monitoring Method   | Sensor Placement Approach                                 | Reference  |
|----------|---------------------------|---|------------|
| System 1 | CUSUM chart               | Sensitivity analysis                                      | [9]        |
| System 2 | Standardized EWMA chart   | <i>k</i> -means clustering and sensitivity analysis       | This study |
| System 3 | Standardized EWMA and PCA | PCA, <i>k</i> -means clustering, and sensitivity analysis | This study |

The systems were evaluated using the scenarios, and the results for scenario 6 are shown in Figure 4. In scenario 6, the burst flowrate, time, and duration were assumed to be 60 m<sup>3</sup>/day, 384 min, and 4 min at N4, respectively. Figure 4a,b show the variations in flowrate and unit head loss at the entrances of the WDN. The unit head loss was used instead of pressure at the entrance because the entrance was directly linked to the reservoir, and the pressure was not changed. Two small peaks, highlighted by circles in Figure 4a,b, show the variations of flowrate and unit loss. However, it is hard to identify these two peaks (which are induced by the burst at N4) in the measured data set due to the considerable variations in the daily pattern and the noise of the flowrate and unit head loss. This shows that a method to remove strong diurnal patterns in the WDN is necessary to identify the

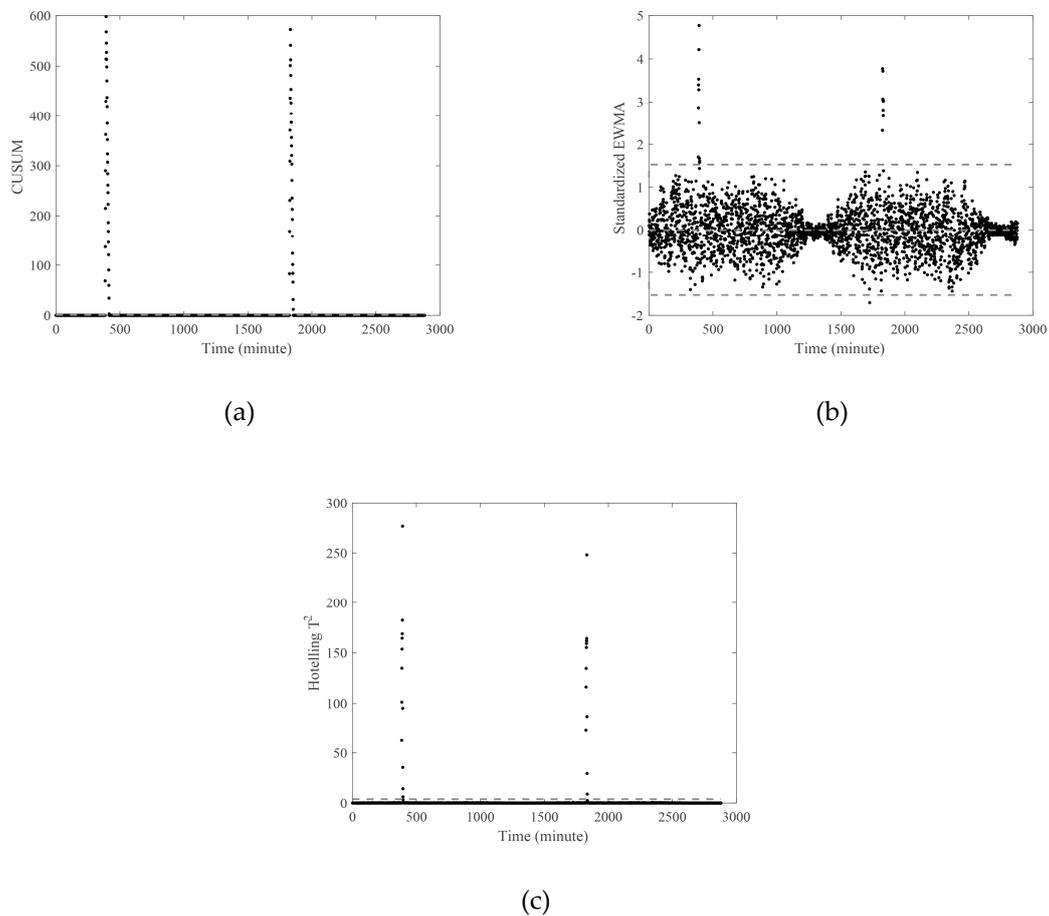
burst in the measured data set. Therefore, a mean trajectory removal technique was implemented prior to applying the monitoring system to remove the periodic pattern from the measurements. Figure 4c,d show the data set on which mean trajectory filtering was implemented. This data was input to the monitoring system.



**Figure 4.** Variation of flowrate and pressure at the entrance over a 48 h period: (a) Flowrate, (b) unit head loss, (c) flowrate without periodic pattern, and (d) head loss without periodic pattern.

Three monitoring systems, the CUSUM, the standardized EWMA, and the standardized EWMA-PCA, were applied to the cleansed dataset according to scenario 6. The obtained results with the monitoring limit lines are shown in Figure 5a–c. The flowrate and unit head loss were used in the standardized EWMA-PCA simultaneously. However, only the flowrate was used in the CUSUM and the standardized EWMA because these monitoring systems were developed based on univariate statistical techniques. The monitored burst times were similar to the real burst times of 384 and 1824 min. The burst detection times obtained from applying all monitoring systems to the 10 scenarios are summarized in Table 4. According to the table, shorter detection time was obtained with the CUSUM chart system compared to the standardized EWMA and the standardized EWMA-PCA systems. However, the burst in S1(10 m<sup>3</sup>/day) and S2(20 m<sup>3</sup>/day) was not detected by the CUSUM chart system. Moreover, the system 2 could not detect the burst in S2. This indicates that the systems 1 and 2 have limitations in detecting small burst flow rate despite of fast detection time. In addition, the detection time of the CUSUM and standardized EWMA systems increased as the burst size decreased from 100 m<sup>3</sup>/day in S10 to 10 m<sup>3</sup>/day in S1. This means that the univariate statistical monitoring systems were not efficient in detecting small bursts, and their monitoring performance was affected by noise, which hid small bursts. On the other hand, the standardized EWMA-PCA could detect bursts in all generated scenarios of which burst flowrates were from 10 m<sup>3</sup>/day to 100 m<sup>3</sup>/day, and the detection time was

relatively fast, even for small bursts. This is because the detection performance of the proposed system improved by considering unit head loss alongside flowrate.



**Figure 5.** Burst detection at the entrance of the WDN according to S6 using (a) CUSUM, (b) standardized EWMA, and (c) standardized EWMA-PCA.

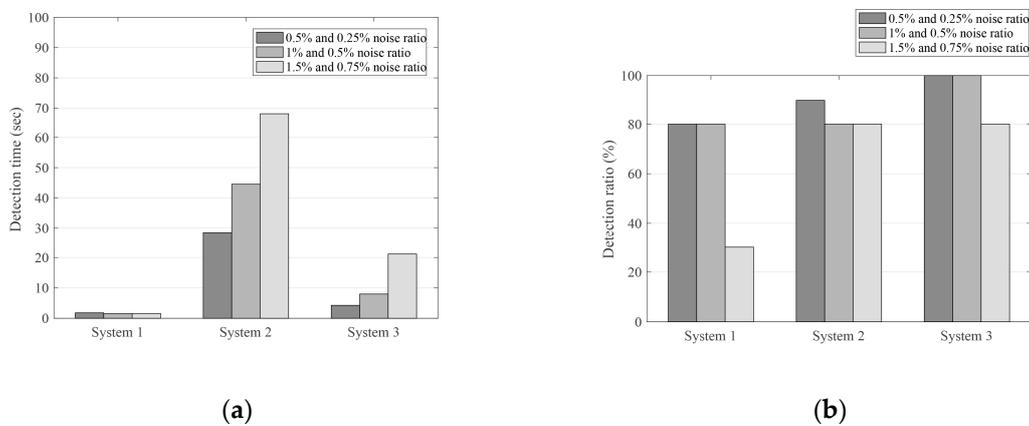
**Table 4.** Detection times of the burst monitoring systems in each scenario.

| Scenario | CUSUM<br>(system 1) | Standardized EWMA<br>(system 2) | Standardized EWMA-PCA<br>(system 3) |
|----------|---------------------|---------------------------------|-------------------------------------|
| S1       | -                   | -                               | 0.88 s                              |
| S2       | -                   | 47.72 s                         | 1.63 s                              |
| S3       | 4.11 s              | 30.27 s                         | 6.04 s                              |
| S4       | 3.32 s              | 35.52 s                         | 12.98 s                             |
| S5       | 1.83 s              | 20.87 s                         | 4.62 s                              |
| S6       | 1.15 s              | 27.51 s                         | 2.59 s                              |
| S7       | 1.01 s              | 25.99 s                         | 1.56 s                              |
| S8       | 0.95 s              | 25.36 s                         | 5.33 s                              |
| S9       | 1.09 s              | 17.41 s                         | 2.84 s                              |
| S10      | 0.83 s              | 25.84 s                         | 4.10 s                              |
| Average  | 1.79 s              | 28.50 s                         | 4.26 s                              |

### 3.3. Robustness of the Three Monitoring Systems

To evaluate robustness of the three monitoring systems against noise, detection performance was compared based on detection time and detection ratio. The comparative results are shown in Figure 6a,b. The detection ratio is the ratio of correct detection time to the whole period, and it decreased as noise increased in the CUSUM, although the average detection time was less than that of

the two other monitoring systems. The detection ratio of the standardized EWMA was not significantly affected by the increase in noise; however, more time was required to detect the burst occurrence as the noise increased. This indicates that the standardized EWMA was more robust than the CUSUM chart.



**Figure 6.** A comparison of the three monitoring systems according to noise in flowrate and pressure considering (a) detection time and (b) detection ratio.

Control limits should be established in CUSUM and standardized EWMA to differentiate between normal and abnormal process conditions. This limit was obtained according to the average quantity of burst size in the CUSUM. However, the limit was obtained with respect to the variance quantity of normalized data instead of using the measured data in standardized EWMA. Therefore, the control limit of standardized EWMA was more flexible than in CUSUM with respect to dynamic noise. Consequently, standardized EWMA would be more suitable than CUSUM if the measured data showed a relatively large noise ratio or dynamic noise size regardless of the detection time. Moreover, standardized EWMA-PCA detected the burst using the measured noise more rapidly and stably than CUSUM and standardized EWMA. Using the weighted average values as input data in PCA, an abrupt change was clearly identified in the measured data according to noise. Furthermore, PCA enhances the peak size due to burst occurrence by considering the flowrate and unit head loss simultaneously. Thus, the standardized EWMA-PCA monitoring system was very effective in identifying abnormal conditions in the measured data with a high noise ratio.

### 3.4. Optimal Sensor Placement

Sensitivity analysis was applied to determine the optimal pressure sensor placement to isolate burst location using Equations (10) and (11). It was assumed that the average burst size was  $50 \text{ m}^3/\text{day}$ , and the burst occurred on an individual node at a time interval. The sensitivity analysis was performed under steady state conditions. The results of cumulative sensitivity of all candidate nodes were obtained and are summarized in Table 5. The cumulative sensitivity varied in a range between  $5.16 \times 10^{-7}$  and  $1.10 \times 10^{-5}$ . The most and the least sensitive sections in the WDN were, respectively, N2 and N1 according to variation in pressure drop. Having conducted the sensitivity analysis in all nodes, the clustering algorithm was implemented to classify the nodes according to hydraulic similarities. Thus, the optimal sensor location was obtained in a node that had the highest sensitivity among all nodes in the same cluster.

Table 6 summarizes input data employed in the  $k$ -means clustering algorithm. The PCA was preliminarily conducted on raw data prior to  $k$ -means implementation to reduce dimensionality of the raw data and to harmonize the data from difference scales. Input data to the PCA were obtained using three parameters of the regression model (Equation (8)) and mean and standard deviation of the flowrate and pressure. Two PCs were obtained by the PCA and explained 92.4% of the variance in the raw data. The PCs were then used in the  $k$ -means clustering algorithm. The average flowrate versus the average

pressure, as well as the score plots of the PCA are graphically shown in Figure 7. The flowrate-pressure plot shows a different distribution of nodes compared to the PCA score plot. As an instance, the location of N5 was considerably different in the score plot than in the flowrate-pressure plot. The N5 was very close to N9 and N12 in the flowrate-pressure plot, and it was hard to distinguish it from the adjacent nodes. However, N5, N9, and N12 were separately located in the score plots. Thus, the PCA efficiently extracted key information from seven data sets and reflected co-relationships between flowrate and pressure in the distribution of all nodes.

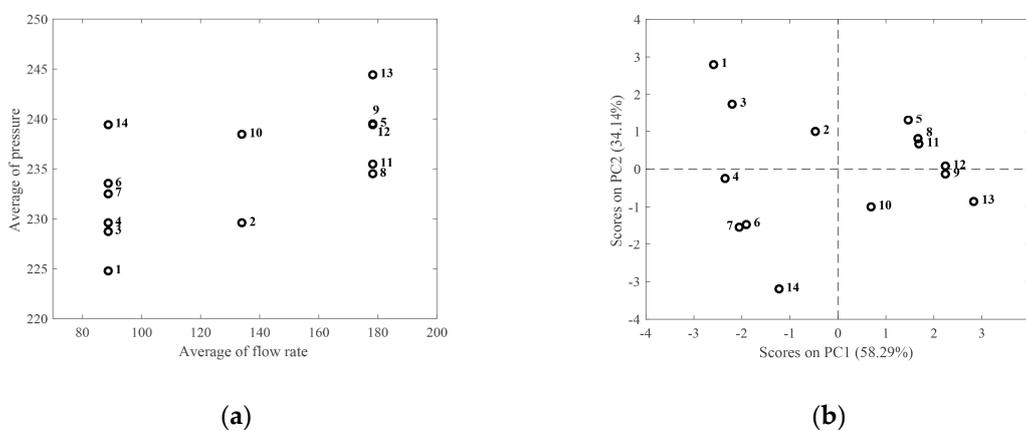
**Table 5.** Cumulative sensitivity of all candidate nodes.

| Node                             | N1                    | N2                    | N3                    | N4                    | N5                    | N6                    | N7                    |
|----------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Cumulative sensitivity ( $S_i$ ) | $5.16 \times 10^{-7}$ | $1.10 \times 10^{-5}$ | $1.53 \times 10^{-6}$ | $1.88 \times 10^{-6}$ | $2.35 \times 10^{-6}$ | $3.78 \times 10^{-6}$ | $1.88 \times 10^{-6}$ |
| Node                             | N8                    | N9                    | N10                   | N11                   | N12                   | N13                   | N14                   |
| Cumulative sensitivity ( $S_i$ ) | $2.12 \times 10^{-6}$ | $2.76 \times 10^{-6}$ | $4.14 \times 10^{-6}$ | $3.76 \times 10^{-6}$ | $2.48 \times 10^{-6}$ | $2.76 \times 10^{-6}$ | $4.09 \times 10^{-6}$ |

**Table 6.** *k*-means input data according to regression model.

| Node | Parameters of Regression Model |                       |       | Flow Rate |                    | Pressure |                    |
|------|--------------------------------|-----------------------|-------|-----------|--------------------|----------|--------------------|
|      | $p_1$                          | $p_2$                 | $p_3$ | Mean      | Standard Deviation | Mean     | Standard Deviation |
| N1   | $-1.94 \cdot 10^{-5}$          | $-3.46 \cdot 10^{-4}$ | 225   | 88.6      | 35.34              | 224.79   | 0.11               |
| N2   | $-1.70 \cdot 10^{-5}$          | $-5.73 \cdot 10^{-4}$ | 230   | 134.0     | 53.43              | 229.58   | 0.23               |
| N3   | $-2.26 \cdot 10^{-5}$          | $-4.51 \cdot 10^{-4}$ | 229   | 88.6      | 35.34              | 228.75   | 0.13               |
| N4   | $-3.36 \cdot 10^{-5}$          | $-8.12 \cdot 10^{-4}$ | 230   | 88.6      | 35.34              | 229.63   | 0.20               |
| N5   | $-1.09 \cdot 10^{-5}$          | $-4.52 \cdot 10^{-4}$ | 225   | 178.3     | 71.10              | 239.52   | 0.26               |
| N6   | $-4.05 \cdot 10^{-5}$          | $-8.83 \cdot 10^{-4}$ | 234   | 88.6      | 35.34              | 233.56   | 0.24               |
| N7   | $-4.21 \cdot 10^{-5}$          | $-9.16 \cdot 10^{-4}$ | 233   | 88.6      | 35.34              | 232.54   | 0.25               |
| N8   | $-1.16 \cdot 10^{-5}$          | $-4.45 \cdot 10^{-4}$ | 235   | 178.2     | 71.10              | 234.50   | 0.27               |
| N9   | $-1.20 \cdot 10^{-5}$          | $-5.10 \cdot 10^{-4}$ | 240   | 178.3     | 71.10              | 239.47   | 0.28               |
| N10  | $-2.09 \cdot 10^{-5}$          | $-6.85 \cdot 10^{-4}$ | 239   | 134.0     | 53.43              | 238.48   | 0.28               |
| N11  | $-1.10 \cdot 10^{-5}$          | $-4.80 \cdot 10^{-4}$ | 236   | 178.3     | 71.10              | 235.51   | 0.26               |
| N12  | $-1.16 \cdot 10^{-5}$          | $-4.67 \cdot 10^{-4}$ | 240   | 178.3     | 71.10              | 239.49   | 0.27               |
| N13  | $-1.20 \cdot 10^{-5}$          | $-5.26 \cdot 10^{-4}$ | 245   | 178.3     | 71.10              | 244.47   | 0.28               |
| N14  | $-5.06 \cdot 10^{-5}$          | $-9.67 \cdot 10^{-4}$ | 240   | 88.6      | 35.34              | 239.46   | 0.29               |

Notes:  $p_1$ ,  $p_2$ , and  $p_3$  indicate parameters in the second-order regression model expressed by Equation (7).



**Figure 7.** Distribution of the nodes in (a) flowrate and pressure plot and (b) PCA score plot.

The nodes were clustered according to the desired number of sensors using the score matrix obtained by the PCA. The differences in the distribution of nodes were not reflected if a small number

of clusters (such as two) were employed in the  $k$ -means approach. However, four clusters included characteristics of the nodes in detail. The optimal pressure sensor locations are given in Table 7 based on the number of sensors. The optimal pressure sensor configuration was obtained when isolating the burst location. The optimal measurement locations were obtained among all probable configurations for each number of sensors.

Table 7. Optimal sensor placement according to number of sensors.

| Number of Sensors | Number of Probable Configurations | Selected Measurement Points                   |
|-------------------|-----------------------------------|---|
| 1                 | 14                                | 2   |
| 2                 | 91                                | 2, 10   |
| 3                 | 364                               | 2, 10, 14                                     |
| 4                 | 1001                              | 2, 6, 10, 14                                  |
| 5                 | 2002                              | 2, 3, 6, 10, 14                               |
| 6                 | 3003                              | 2, 3, 6, 9, 11, 14                            |
| 7                 | 3432                              | 2, 3, 4, 9, 10, 11, 14                        |
| 8                 | 3003                              | 2, 3, 4, 6, 9, 10, 11, 14                     |
| 9                 | 2002                              | 2, 3, 4, 9, 10, 11, 12, 13, 14                |
| 10                | 1001                              | 2, 3, 4, 5, 6, 9, 10, 11, 13, 14              |
| 11                | 364                               | 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14          |
| 12                | 91                                | 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14       |
| 13                | 14                                | 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14    |
| 14                | 1                                 | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 |

### 3.5. Burst Isolation

The burst was detected using the optimal sensor location as described in the previous sections. The performance of the three burst monitoring systems in isolating the burst event is graphically compared in Figure 8. As shown in the figure, the sensitivity analysis-assisted CUSUM chart-monitoring system employed nine sensors (more than half the number of nodes) and had an isolation ratio of 10%. Furthermore, 8.57% of burst events in 10 scenarios were successfully detected and isolated using this monitoring system. Despite the CUSUM chart rapidly detecting the burst occurrence, it failed in burst monitoring in many cases. The sensitivity analysis-based method for selecting the optimal sensor location could not describe the dynamics of pressure and flowrate in the nodes. Therefore, system 1 was not appropriate to monitor the burst in the presence of dynamic noises. System 2, in which the standardized EWMA chart was employed to monitor the burst, and the  $k$ -means clustering algorithm and sensitivity analysis were used to locate the pressure sensors, improved the isolation ratio (14.29%) significantly. Thus, the standardized EWMA was accurate and robust against dynamic noise, though it needed longer computation time to detect the burst. In addition, the  $k$ -means clustering approach was acceptable to determine the optimal sensor location considering the dynamics of pressure in the nodes.

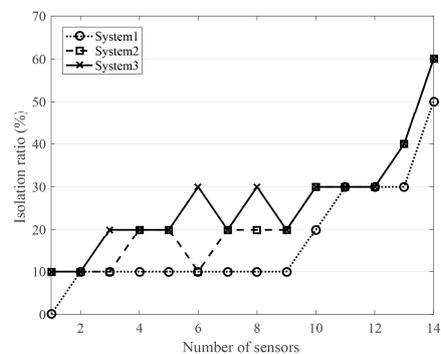


Figure 8. Isolation performance of three systems according to number of pressure sensors.

System 3 (in which the standardized EWMA and the PCA were employed to monitor the burst, and PCA, *k*-means clustering, and sensitivity analysis were used to select the sensor locations) had an isolation ratio of 18.57%. The detection time decreased in system 2 by applying a multivariate statistical monitoring method (PCA), and the isolation ratio was enhanced by adding a periodic pattern of the pressure and flowrate in the nodes through the PCA. As shown in Figure 8, system 3 improved both monitoring and isolation of the burst in comparison with systems 1 and 2. When applying six pressure sensors at N2, N3, N6, N9, N11, and N14, system 3 improved the isolation ratio by 20% compared to systems 1 and 2. It should be noted that the performance of six sensors in system 3 corresponded to employing 10 sensors in system 2 and 11 sensors in system 1. Thus, the third monitoring system was the most economic. This system is anticipated to have better performance than other two systems considering more complicated WDNs. However, the novel system was validated considering a simple WDN for the restrictions in applying real world data in this study. The research can be extended taking various burst scenarios in different locations of a complex network into account.

#### 4. Conclusions

A novel burst monitoring, isolation, and sensor placement system was proposed for water distribution networks based on multivariate statistical and analytical techniques. In the proposed method, a hybrid standardized EWMA-PCA approach and PCA, *k*-means, and sensitivity analyses were employed for burst monitoring and sensor placement, respectively. The proposed system was validated in a simulated branched WDN in 10 scenarios, and its performance was compared to those of conventional monitoring systems. The proposed system efficiently improved burst monitoring regardless of noise pattern and isolated the burst location. The average monitoring performance considering the noise was 93%, and the isolation ratio improved by 10% compared to the conventional systems because both pressure and flowrate were considered in the multivariate monitoring system, and the hydraulic characteristics of the nodes were included in the approach. Furthermore, the system could monitor small burst events that were not detected by the conventional monitoring systems. The sensitivity analysis results showed that the proposed system was robust to diverse burst events. In addition, the proposed system used optimal pressure sensors that were more economic than the flow sensors. The installation ratio improved by 20%, while the optimal number of sensors decreased by 40%. Alongside the economic benefits, the proposed burst detection and isolation system is anticipated to detect and isolate burst events in WDNs much faster than conventional systems.

Complex real WDNs are required to validate the proposed burst detection and isolation monitoring system in industrial settings. Furthermore, more scenarios including burst occurrence along the pipes need to be mentioned to compare the performance of the system in detecting the burst by the nodes and far from them. For this, big datasets should be employed using open source data generated in complicated networks. A deep learning model can be applied to the proposed system to predict, interpret and analyze the big-data in the WDNs. The application of deep learning algorithms to the proposed system can enhance the performance of burst monitoring and isolation in the complex real WDNs. It is anticipated that the monitoring system can be improved by these changes in the future in terms of the system's efficiency and economic-social-environmental costs.

**Author Contributions:** Conceptualization, K.N., S.L., and C.Y.; methodology, K.N. and P.I.; formal analysis, K.N. and S.L.; data curation, S.H. and G.H.; writing—original draft preparation, K.N.; writing—review and editing, P.I. and C.Y.

**Acknowledgments:** This work was supported by a National Research Foundation (NRF) grant funded by the Korean government (MSIT) (No. NRF-2017R1E1A1A03070713), the Korea Ministry of Environment (MOE) as Graduate School specialized in Climate Change, and a grant from the Railway Technology Research Project of the Ministry of Land Infrastructure and Transport (19QPPW-B152307-01).

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

## References

1. Jang, D.; Park, H.; Choi, G. Estimation of leakage ratio using principal component analysis and artificial neural network in water distribution systems. *Sustainability* **2018**, *10*, 750. [[CrossRef](#)]
2. Choi, D.Y.; Kim, S.-W.; Choi, M.-A.; Geem, Z.W. Adaptive Kalman filter based on adjustable sampling interval in burst detection for water distribution system. *Water* **2016**, *8*, 142. [[CrossRef](#)]
3. Mazumder, R.K.; Salman, A.M.; Li, Y.; Yu, X. *Performance Evaluation of Water Distribution Systems and Asset Management*; American Society of Civil Engineers: Reston, VA, USA, 2018.
4. Choi, G.B.; Kim, J.W.; Suh, J.C.; Jang, K.H.; Lee, J.M. A prioritization method for replacement of water mains using rank aggregation. *Korean J. Chem. Eng.* **2017**, *34*, 2584–2590. [[CrossRef](#)]
5. Jung, S.; Vanli, O.A.; Kwon, S.D. Wind energy potential assessment considering the uncertainties due to limited data. *Appl. Energy* **2013**, *102*, 1492–1503. [[CrossRef](#)]
6. Farmani, R.; Kakoudakis, K.; Behzadian, K.; Butler, D. Pipe Failure Prediction in Water Distribution Systems Considering Static and Dynamic Factors. *Procedia Eng.* **2017**, *186*, 117–126. [[CrossRef](#)]
7. Rojek, I.; Studzinski, J. Detection and Localization of Water Leaks in Water Nets Supported by an ICT System with Artificial Intelligence Methods as a Way Forward for Smart Cities. *Sustainability* **2019**, *11*, 518. [[CrossRef](#)]
8. Lee, S.J.; Lee, G.; Suh, J.C.; Lee, J.M. Online burst detection and location of water distribution systems and its practical applications. *J. Water Resour. Plan. Manag.* **2015**, *142*, 04015033. [[CrossRef](#)]
9. Wang, B.; Yan, X.; Jin, Y. Fault detection based on polygon area statistics of transformation matrix identified from combined moving window data. *Korean J. Chem. Eng.* **2017**, *34*, 275–286. [[CrossRef](#)]
10. Misiunas, D.; Vítkovský, J.; Olsson, G.; Lambert, M.; Simpson, A. Failure monitoring in water distribution networks. *Water Sci. Technol.* **2006**, *53*, 503–511. [[CrossRef](#)] [[PubMed](#)]
11. Palau, C.; Arregui, F.; Carlos, M. Burst detection in water networks using principal component analysis. *J. Water Resour. Plan. Manag.* **2011**, *138*, 47–54. [[CrossRef](#)]
12. Loureiro, D.; Amado, C.; Martins, A.; Vitorino, D.; Mamade, A.; Coelho, S.T. Water distribution systems flow monitoring and anomalous event detection: A practical approach. *Urban Water J.* **2016**, *13*, 242–252. [[CrossRef](#)]
13. Hutton, C.; Kapelan, Z. Real-time Burst Detection in Water Distribution Systems Using a Bayesian Demand Forecasting Methodology. *Procedia Eng.* **2015**, *119*, 13–18. [[CrossRef](#)]
14. Zhang, K.; Yan, H.; Zeng, H.; Xin, K.; Tao, T. A practical multi-objective optimization sectorization method for water distribution network. *Sci. Total Environ.* **2019**, *656*, 1401–1412. [[CrossRef](#)] [[PubMed](#)]
15. Romano, M.; Woodward, K.; Kapelan, Z. Statistical Process Control Based System for Approximate Location of Pipe Bursts and Leaks in Water Distribution Systems. *Procedia Eng.* **2017**, *186*, 236–243. [[CrossRef](#)]
16. Saccenti, E.; Hoefsloot, H.C.; Smilde, A.K.; Westerhuis, J.A.; Hendriks, M.M. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **2014**, *10*, 361–374. [[CrossRef](#)]
17. Wu, Y.; Liu, S. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water J.* **2017**, *14*, 972–983. [[CrossRef](#)]
18. Diao, K.; Rauch, W. Controllability analysis as a pre-selection method for sensor placement in water distribution systems. *Water Res.* **2013**, *47*, 6097–6108. [[CrossRef](#)] [[PubMed](#)]
19. Farley, B.; Mounce, S.; Boxall, J. Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. *Urban Water J.* **2010**, *7*, 345–356. [[CrossRef](#)]
20. Romero-Tapia, G.; Fuente, M.J.; Puig, V. Leak Localization in Water Distribution Networks using Fisher Discriminant Analysis. *IFAC-PapersOnLine* **2018**, *51*, 929–934. [[CrossRef](#)]
21. Sarrate, R.; Blesa, J.; Nejjari, F.; Quevedo, J. Sensor placement for leak detection and location in water distribution networks. *Water Sci. Technol. Water Supply* **2014**, *14*, 795–803. [[CrossRef](#)]
22. Costanzo, F.; Morosini, A.F.; Veltri, P.; Savić, D. Model calibration as a tool for leakage identification in WDS: A real case study. *Procedia Eng.* **2014**, *89*, 672–678. [[CrossRef](#)]
23. Sayyed, M.A.; Gupta, R.; Tanyimboh, T. Modelling pressure deficient water distribution networks in EPANET. *Procedia Eng.* **2014**, *89*, 626–631. [[CrossRef](#)]
24. Constantin, A.; Nițescu, C.S. Water Distribution Network Design Based on Numerical Simulation in EPANET. In Proceedings of the International Scientific Conference People, Buildings and Environment, Brno, Czech Republic, 17–19 October 2018.

25. Haq, A.; Khoo, M.B.C. An adaptive multivariate EWMA chart. *Comput. Ind. Eng.* **2019**, *127*, 549–557. [[CrossRef](#)]
26. Nam, K.; Kim, M.; Lee, S.; Hwangbo, S.; Yoo, C. Interpretation and diagnosis of fouling progress in membrane bioreactor plants using a periodic pattern recognition method. *Korean J. Chem. Eng.* **2017**, *34*, 2966–2977. [[CrossRef](#)]
27. Montgomery, D.C. *Introduction to Statistical Quality Control*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
28. Ifaei, P.; Karbassi, A.; Lee, S.; Yoo, C. A renewable energies-assisted sustainable development plan for Iran using techno-econo-socio-environmental multivariate analysis and big data. *Energy Convers. Manag.* **2017**, *153*, 257–277. [[CrossRef](#)]
29. Yao, J.; Pan, Y.; Yang, S.; Chen, Y.; Li, Y. Detecting Fraudulent Financial Statements for the Sustainable Development of the Socio-Economy in China: A Multi-Analytic Approach. *Sustainability* **2019**, *11*, 1579. [[CrossRef](#)]
30. Ay, M.; Kisi, O. Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques. *J. Hydrol.* **2014**, *511*, 279–289. [[CrossRef](#)]
31. Hatamlou, A.; Abdullah, S.; Nezamabadi-Pour, H. A combined approach for clustering based on K-means and gravitational search algorithms. *Swarm Evol. Comput.* **2012**, *6*, 47–52. [[CrossRef](#)]
32. Ifaei, P.; Karbassi, A.; Jacome, G.; Yoo, C. A systematic approach of bottom-up assessment methodology for an optimal design of hybrid solar/wind energy resources—Case study at middle east region. *Energy Convers. Manag.* **2017**, *145*, 138–157. [[CrossRef](#)]
33. Ifaei, P.; Farid, A.; Yoo, C. An optimal renewable energy management strategy with and without hydropower using a factor weighted multi-criteria decision making analysis and nation-wide big data—Case study in Iran. *Energy* **2018**, *158*, 357–372. [[CrossRef](#)]
34. Liu, L.; Peng, Z.; Wu, H.; Jiao, H.; Yu, Y.; Zhao, J. Fast identification of urban sprawl based on K-means clustering with population density and local spatial entropy. *Sustainability* **2018**, *10*, 2683. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).