Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2019.23375

143 (2019) 69–77
March

The principal component analysis for detecting leaks in water pipe networks utilizing flow and leak record data

Suwan Park[a,*], Jaehong Ha[a], Kimin Kim[b]

[a]*Department of Civil and Environmental Engineering, Pusan National University, Busan 609-735, South Korea, Tel. +82 51 510 2734; Fax: +82 51 513 9596; email: swanpark@pusan.ac.kr (S. Park)*
[b]*Seoyong Engineering, Yongin 16950, South Korea*

Received 31 August 2018; Accepted 15 October 2018

ABSTRACT

In this paper, the potential of the principal component analysis (PCA) technique for the application of detecting leaks in water pipe networks was evaluated. For this purpose, the PCA was conducted to evaluate the relevance of the calculated statistical outliers of a PCA model utilizing the recorded inflows of district metered areas (DMAs) and the records of leak repair completion of a case study water pipe network. The PCA technique was enhanced by applying the computational algorithm developed in this study which was designed to use flow data in a time window from the original 24-h flow data so that the effective outlier detection rate was maximized. Sensitivity analyses of the parameters of the PCA model and the developed algorithm on the results of the study were conducted. Consideration on how to apply the parameters in the practical applications was also presented. The developed algorithm may be applied in determining whether further leak detection field work for DMAs needs to be performed.

Keywords: Principal component analysis (PCA); DMA; Water pipe network; Leak detection; Computational algorithm; Flow data

1. Introduction

Recently, smart water management technology has been actively developed worldwide in order to reduce unaccounted-for water quantity due to water leakage, pipeline rupture, and weighing error in the water supply networks. As one of the areas of smart water management technology, studies on the development of techniques for estimating the leakage in water pipe networks such as International Water Association method based on the top-down method of Lambert and Hirner [1], the bottom-up method of nighttime minimum flow analysis by Covas et al. [2] had been developed. In addition, water leak detection techniques in water supply networks were developed based on the optimization and mathematical modeling techniques of Kapelan et al. [3] and Stathis and Loganathan [4], acoustic logging method of Muggleton et al. [5] and Pilcher et al. [6], and the correlation method of Muggleton and Brennan [7].

The use of computational algorithms to predict or detect water leakage in water supply networks has been attempted relatively recently. Some examples are the applications of an artificial neural network [8], fuzzy inference, and hydraulic simulation method of Xia and Guo-jin [9], the support vector [10,11], and the Kalman filtering technique [12]. However, in case of method of Kalman filtering, it is impossible to perform analysis if some portion of data is missing. In case of the method using artificial neural network, it is necessary to newly train the artificial neural network by using up-to-date data when the state of water supply pipe network, for example, the status of valve and layout of pipe network, has been changed. In the case of support vector machines, hundreds or thousands of artificial leaks must be generated in the real water supply network in order to obtain the data sets needed to train the artificial intelligence. Due to the discussed limitations, application of the above methods for the real water supply networks is severely limited. Contrary to the above

* Corresponding author.

methods, the principal component analysis (PCA) technique, which is one of the statistical multivariate data analysis techniques, does not need to train an artificial neural network and can still be applied to data sets that have missing values.

Palau et al. [13] used the PCA technique to analyze the flow data of the water supply network. Palau et al. [13] performed PCA using the flow data observed from the laboratory-scale water supply network. The T^2 Hotelling and distance to model (DMOD) statistics was obtained and compared with the threshold value of each statistic to detect leakage. Park et al. [14] applied the PCA to the nighttime flow data observed for 4 months in four small district metered areas (DMAs) in real water pipe network to determine the optimal PCA period. Park et al. [15] developed a computational algorithm to determine the minimum water use time in a day for a representative water use type within a water supply DMA in an attempt to more efficiently perform leak detection at nighttime.

In this study, the PCA technique was applied to the flow data observed in a water supply network, and the relationship between the detected statistical outliers and the maintenance records related to leakage of the water supply pipelines were evaluated. Based on the results of the analysis, the applicability of the PCA to the early detection of leakage of water pipe network was analyzed. A computational algorithm that can improve the accuracy of leak detection by using the PCA model was also developed and applied to the flow data in the DMAs of an actual water distribution pipe network in Korea.

The method developed in this paper may be used to determine whether to perform preemptive pipe maintenance activity for a DMA. The analysis developed in this paper may be performed every day for accumulated flow and leak record data for a DMA to check if the flow data during the past day represent an outlier of a PCA model. The developed method can be used to discern whether a detected outlier is considered to be relevant to a leak incident, which can also be used to make a decision for preemptive pipe maintenance activity for a DMA. The sensitivities of the parameters for the PCA and the algorithm developed in this study were also performed.

2. Material and methods

2.1. The PCA and DMOD statistics

PCA is one of the analytical techniques for analyzing multivariate data. It is a technique for converting the data contained in multidimensional data to low-dimensional data with minimal loss of information. The principal concept of PCA is to represent the whole information through fewer variables than the original data. Principal components are statistically independent from each other, and there is no loss of information when all the main components of the data are used. The first principal component best describes the variability of the data, and the explanatory powers of the principal components diminish gradually.

The PCA consists of a principal component score matrix T of $n \times f$ dimensions and a loading matrix of $n \times m$ dimensions, where n is the number of observations and m the number of variables. Eq. (1) shows that an original data matrix may be factorized into loading matrix P of $m \times f$ dimensions and a residual matrix E of $n \times m$ dimensions.

$$X = T \times P^T + E \tag{1}$$

where f is the number of principal components, and $f < m$. The column of the loading matrix P represents an eigenvector for the eigenvalue of the variance–covariance matrix of X. The eigenvectors, that is, the principal components are arranged in the order of the corresponding eigenvalues, and the principal components can be selected only partially according to the purpose of analysis. The most optimal partitioning of the raw data is to minimize the residual matrix by partitioning. The T^2 Hotelling statistic or DMOD statistics may be used to determine outliers of a PCA model.

According to Palau et al. [13], the T^2 Hotelling statistic is more suitable for detecting abnormal demand in a pipe network, and the DMOD statistic is more suitable for detecting leakage in a pipe network. The equation for obtaining a DMOD statistic is given by Eqs. (2) and (3), respectively.

$$S_i = \sqrt{\frac{\sum_{k=1}^{K} e_{ik}^2}{(K-A)}} \tag{2}$$

$$S_0 = \sqrt{\frac{\sum_{i=1}^{N}\sum_{k=1}^{K} e_{ik}^2}{(N-A-A_0)(K-A)}} \tag{3}$$

where S_i represents the absolute DMOD and S_0 the normalized distance of the model, K the number of primitive variables, A the number of principal components, N the number of observations, and A_0 1 when it is normalized and 0 otherwise. E_{ik} residual of i-th observation of variable K. Eq. (4) represents the criterion for determining whether an estimated DMOD statistic is an outlier.

$$(S_i / S_0)^2 > F_{\text{critical}} \tag{4}$$

where $(S_i/S_0)^2$ has an F-Snedecor probability distribution with $(N–A–1)$ $(K–A)$ degrees of freedom and $(K–A)$ as parameters. For example, if the calculated DMOD statistic $(S_i/S_0)^2$ is larger than the p-value of the F-Snedecor probability distribution, a DMOD statistic of the flow data is determined to be an outlier. In this study, as shown in Palau et al. [13], it was considered that a DMOD outlier may be regarded as an indication of a leak incident in a DMA.

2.2. The flow data

This study used inflow data of 11 DMAs in a water pipe network in Korea. The flow measurement data are the 24-h flow data recorded every hour for 426 d from September 1, 2011, to October 30, 2012. The leakage records for the study area present the dates of completion of leakage repair with a total of 120 dates of completion of leakage repair. Table 1 shows some portion of the DMA inflow data of the water pipe network.

3. Results and discussions

3.1. Analysis on the 24-h flow data

For each DMA of the study network, the dates that have reports of completion of leak repairs and the dates on which the flow data were determined as DMOD outliers of the PCA model were compared. As the first attempt for the analyses, the PCA model was constructed using the hourly flow data measured for a whole day of the DMAs. The outliers of the PCA model used in the analyses were the outliers of the DMOD statistics.

It is considered normal that the time of completion of leakage repair is different from the time of leakage occurrence. Because the dates that have reports of completion of leak repairs of the study network may not be the dates on which actual leaks occurred, the dates within each week before and after a reported date of leakage repair completion were compared with the dates that have the outliers of the PCA model. Moreover, there may be an error in recording the correct date of completion of leakage repair.

Inclusion of the outliers occurred within 7 d prior to the recorded date of leakage repair in the analysis was due to an assumption that a report on the occurrence of a leak and/or completion of leakage repair may be recorded after up to 7 d of an actual outbreak of leakage depending on the leak size and the delay in recognition of leak occurrence. On the other hand, inclusion of the outliers occurred within 7 d after the recorded date of leakage repair in the analysis was due to the consideration of the possibility that up to 7 d of an error might occur in recording the correct date of completion of leakage repair.

Eq. (5) shows the formula used to calculate the effective outlier detection rate (EODR) inside the outlier detection periods (ODPs).

Table 1
Sample inflow data in the case study area water pipe network (m^3/h)

Date	Hour	DMA A	DMA B	DMA C
2012-06-19	23:00	63	19	101
2012-06-19	24:00	36	11	23
2012-06-20	1:00	23	12	88
2012-06-20	2:00	23	12	13
2012-06-20	3:00	24	8	12
2012-06-20	4:00	22	9	12
2012-06-20	5:00	31	11	19
2012-06-20	6:00	44	21	35
2012-06-20	7:00	80	94	112
2012-06-20	8:00	68	148	169
2012-06-20	9:00	71	94	165

$$\text{EODR}\left(\%\right) = \frac{\text{No. of Outliers in ODP}}{\text{Total No. of Outliers}} \times 100 \tag{5}$$

Each ODP is 15 d which has the date on which completion of leakage repair was recorded in the center. Table 2 shows the calculated EODRs for each DMA in the study area.

As shown in Table 2, the EODRs calculated from the PCA modeling using hourly flow data resulted in rather low values implying that the use of PCA in early detection of leaks will not be reliable.

In this study, a computational algorithm that utilizes various parts of the hourly flow data was developed to enhance the EODRs. Furthermore, the developed algorithm was examined to verify the applicability of the algorithms in detecting leaks of real-world water distribution pipe networks. For convenience, the term a “leak report” is also used in this paper to mean a “report on completion of leakage repair”.

3.2. The developed computational algorithm

In this study, a method to improve the efficiency of detecting the outliers was developed which uses a portion of flow data among the hourly flow data rather than the whole 24-h flow data. The portion of flow data used in the algorithm was defined as the hourly flow data measured in a consecutive time window during a day which can be represented with the corresponding “center time” and “time range” of the time window. For example, if the “center time” and “time range” of a time window is 4 and 7, respectively, the data used for the analysis are the flows in the time window of Lambert and Hirner [1], Covas et al. [2], Kapelan et al. [3], Muggleton et al. [5], Muggleton and Brennan [7], Mounce et al. [8], and Borges and Ramirez [10].

Fig. 1 shows the computational algorithms developed in this study to improve the efficiency of detecting the outliers in the ODPs. Based on the developed algorithm, a computer program was developed using the Matlab programming software. Using the computer program, the EODRs were calculated for the hourly flow data inside all of the possible time windows in a day which were generated for all possible combinations of center times and time ranges. The time range was increased from 3 to 23 h in the increment of 2 h. The center time of each time window was varied from 1:00 to 24:00.

An EODR was calculated using the flow data in each time window. The best time window (BTW) which has the highest EODR (maximum EODR, M-EODR) was obtained for each DMA using the developed algorithm. As with the analysis using the 24-h flow data, the outliers of the PCA model were checked if they reside in the ODPs of a DMA. Furthermore, the changes in the calculated M-EODRs were analyzed according to the amount of flow data used.

Table 2
EODRs using 24-h flow data for the case study water pipe network

Name of DMA	A	B	C	D	E	F	G	H	I	J	K
Number of leak records	17	12	9	5	9	6	8	2	17	10	25
EODR using 24-h flow data (%)	54	55	51	32	10	9	38	15	19	39	92
Number of principal component	9	12	12	8	7	12	14	14	8	7	14

Table 3 shows the calculated M-EODRs using the 426 d of flow data and the algorithm shown in Fig. 1.

The center time and range in Table 3 correspond to the time window that resulted in the M-EODR shown in Table 3. Comparing Tables 2 and 3, the EODRs in Table 3 resulted in higher EODRs than the ones in Table 2 except one DMA. Therefore, it was found that there is a specific time zone (or BTW) for each DMA in which the residuals (DMOD statistics) of a PCA model and the leak records have a stronger relationship than the case of using 24-h flow data.

3.3. Analysis on the EODRs for various amounts of flow data

Additional analyses were performed to check if the computed M-EODRs and BTW show stable results as more data are accumulated. It was considered that if the computed M-EODRs and the BTW are drastically different depending on the amount of data used, the method may not be reliable to use in determining whether to perform physical leak detection activity based on the occurrence of the outliers of the PCA model. Table 4 shows the computed M-EODRs for the monthly accumulated flow data starting from September, 2011, to October, 2012.

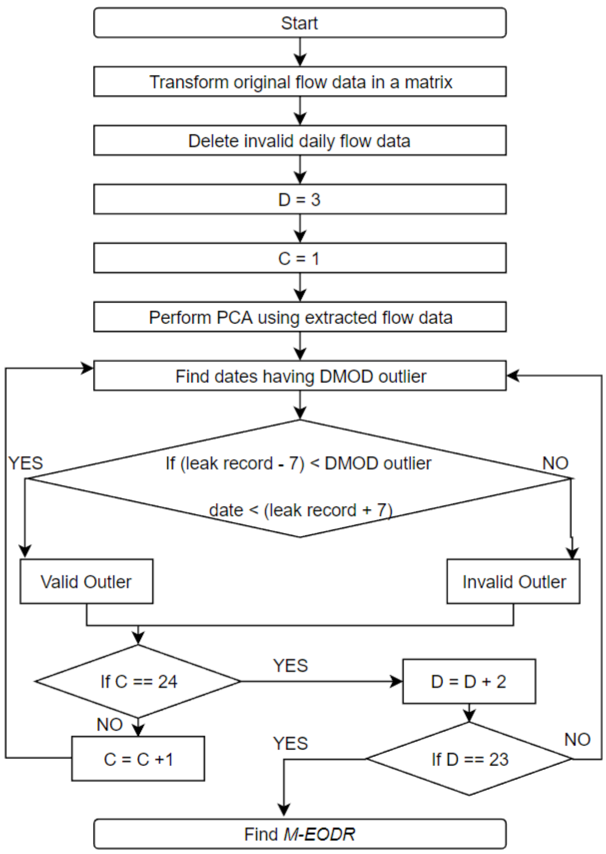

Fig. 1. Computational algorithms for assessing the capability of the PCA for detecting leaks in the DMAs.

In Table 4, the value 90% for DMA I at Month 14 represents the M-EODR for DMA I using the flow data from September, 2011, to October, 2012.

Analysis of Table 4 reveals that the DMAs show stable values for the center times and time ranges after the 9th month except DMAs F and H. The 9th month indicates May. Therefore, this phenomenon is considered to be due to the change of the season which results in less frequent number of leak reports.

The changes in the EODRs with time were estimated as the difference of the EODRs in consecutive months. Fig. 2 shows the change of the EODRs according to the amount of accumulated data over time. As Fig. 2 shows, the EODRs have less variability after the 9th month. Therefore, the variability of the EODR of each DMA over time is also considered to be affected by the accumulated number of leak reports.

However, by analyzing Table 3 general relationship between the EODRs and the number of leak reports for a fixed time was not found other than DMA H and K. In other words, as the number of leak reports increased among the DMAs for a fixed time, the EODRs of the different DMAs did not show any noticeable increase or decrease for a given amount of data used.

DMA H and K showed an exceptional trend from other DMAs. It is conjectured that the EODR of DMA H is considerably low due to the low number of leak reports and the EODR of DMA K is very high because of the high number of leak reports. It is considered that if the number of leak reports is too small, the EODR tends to be estimated as very low because the most of the outliers tend to reside outside of the ODPs resulting in non-effective outliers.

In addition, Table 4 shows that the EODRs of the DMAs tend to have relatively higher values in the early period of flow data observations. The reason of this trend is considered to be due to the characteristics of the developed algorithm and the specific phenomenon of the study area where the outliers and other leak reports coexist inside an ODP in the early period. In this case, the outliers are double counted as effective outliers leading toward higher EODRs than 100%, which is the case DMA K belongs to. Because the occurrences of the outliers and the leak reports were concentrated during the early period of flow data, the effects of the detected outliers diminish as more data are accumulated and the calculated EODRs resulted in a decreasing trend.

Table 3
M-EODRs using hourly flow data for the case study DMAs

DMA	A	B	C	D	E	F	G	H	I	J	K
Number of leak records	17	12	9	5	9	6	8	2	17	10	25
M-EODR (%)	83	81	104	60	104	24	58	9	90	40	163
Center time (h)	15	12	21	7	6	21	21	18	9	16	21
Time range (h)	5	3	3	3	5	3	3	3	3	5	3

Table 4
M-EODRs, center time, and time duration for the flow data

DMA	Category	Cumulative number of months													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	M-EODR (%)	100	250	167	150	200	211	143	124	124	105	94	85	79	83
	Center time (h)	4	13	13	12	15	15	15	14	14	10	10	12	15	14
	Time range (h)	3	3	3	5	3	3	7	5	5	3	3	5	5	7
	Cumulative number of leaks	2	7	7	9	12	13	13	14	15	15	16	16	16	17
B	M-EODR (%)	100	100	100	109	108	100	83	80	81	81	81	81	81	81
	Center time (h)	11	11	3	16	16	16	12	14	11	14	14	14	14	15
	Time range (h)	3	5	3	9	9	9	7	5	19	5	5	5	5	5
	Cumulative number of leaks	1	2	3	5	8	8	8	8	9	9	10	10	10	12
C	M-EODR (%)	200	150	167	140	133	120	119	122	112	108	107	107	104	104
	Center time (h)	12	17	17	4	10	17	10	22	22	22	22	22	22	22
	Time range (h)	3	19	19	3	5	19	23	5	5	5	5	5	5	5
	Cumulative number of leaks	2	3	6	7	8	8	8	8	8	8	9	9	9	9
D	M-EODR (%)	N/A	300	263	233	210	147	129	92	86	71	63	63	63	60
	Center time (h)	N/A	12	10	10	10	8	8	7	7	7	7	7	7	7
	Time range (h)	N/A	3	3	3	3	3	3	3	3	3	3	3	3	3
	Cumulative number of leaks	0	3	3	4	4	4	4	4	4	4	4	4	4	5
E	M-EODR (%)	400	200	144	167	188	208	200	207	200	173	173	157	109	104
	Center time (h)	6	8	5	6	23	2	1	3	3	2	2	2	6	6
	Time range (h)	3	3	9	5	9	13	13	11	11	13	13	13	5	5
	Cumulative number of leaks	5	6	6	7	7	9	9	9	9	9	9	9	9	9
F	M-EODR (%)	N/A	N/A	N/A	0	0	0	0	50	31	33	31	31	29	24
	Center time (h)	N/A	N/A	N/A	1	1	1	1	22	21	21	21	21	21	21
	Time range (h)	N/A	N/A	N/A	3	3	3	3	5	3	3	3	3	3	3
	Cumulative number of leaks	0	0	0	1	1	1	1	4	5	5	6	6	6	6
G	M-EODR (%)	100	100	100	100	91	92	85	83	70	63	65	59	58	58
	Center time (h)	9	5	13	20	4	17	22	15	11	17	17	15	13	13
	Time range (h)	3	3	5	3	17	17	5	13	5	17	17	13	7	7
	Cumulative number of leaks	1	3	4	5	5	6	7	7	7	7	7	7	7	8
H	M-EODR (%)	N/A	N/A	N/A	33	25	25	20	17	10	10	17	17	10	9
	Center time (h)	N/A	N/A	N/A	1	1	1	1	1	4	1	12	12	2	2
	Time range (h)	N/A	N/A	N/A	5	5	5	5	5	5	7	3	3	9	9
	Cumulative number of leaks	0	0	0	1	1	1	1	1	1	2	2	2	2	2
I	M-EODR (%)	200	200	200	200	200	150	200	200	150	144	136	117	94	90
	Center time (h)	2	6	2	6	1	14	1	17	17	17	17	15	9	9
	Time range (h)	3	3	3	3	5	7	5	7	7	7	7	11	3	3
	Cumulative number of leaks	3	5	5	8	9	9	11	13	15	15	16	17	17	17
J	M-EODR (%)	200	125	167	167	150	150	133	120	85	65	71	47	38	40
	Center time (h)	12	12	3	5	19	2	23	3	11	12	13	15	12	13
	Time range (h)	3	5	3	9	3	13	9	11	5	3	7	9	3	3
	Cumulative number of leaks	2	5	7	8	9	9	9	10	10	10	10	10	10	10
K	M-EODR (%)	200	300	220	257	233	213	214	200	185	200	200	200	200	163
	Center time (h)	3	1	7	8	23	10	7	10	19	15	15	15	15	13
	Time range (h)	3	3	15	15	21	7	15	7	7	7	7	7	7	9
	Cumulative number of leaks	4	6	9	13	14	14	17	17	19	21	22	22	22	25

In summary, it was concluded that the changes in the number of leak reports over time, or the change of season, had influence on the changes in the BTWs and the M-EODRs for a DMA. Meanwhile, there were no noticeable differences in the BTWs and M-EODRs among the DMAs with regard to the number of leak reports.

According to Table 4, DMAs A, B, C, E, and I had relatively higher values of M-EODRs for the 14 months of the flow data, and the BTWs did not show much fluctuation in recent months. Therefore, if a criterion of 70% of M-EODR was used, it is considered that preemptive leak detection may be performed for DMAs A, B, C, E, and I in the case study network if the flow data of the precious day turns out to be an DMOD outlier of the PCA method developed in this paper.

3.4. Sensitivity analysis of the parameters of the model and algorithm

The sensitivities of the explanatory power of the eigenvector and the *p*-value of the F-distribution that was used to determine DMOD outliers were analyzed for the BTWs and M-EODRs. Additional analysis on the sensitivity of the ODPs was analyzed for the BTWs and M-EODRs.

Figs. 4–6 show the changes of the BTWs according to different values for the explanatory power of the eigenvalues using the *p*-value of 0.1 and the whole flow data. Fig. 3 shows the changes of the M-EODRs according to different values for the explanatory power of the eigenvalues using the *p*-value of 0.1 and the whole flow data. As can be analyzed from Figs. 3–6, the sensitivities of the explanatory power of the eigenvector analyzed for the BTWs and M-EODRs were not high.

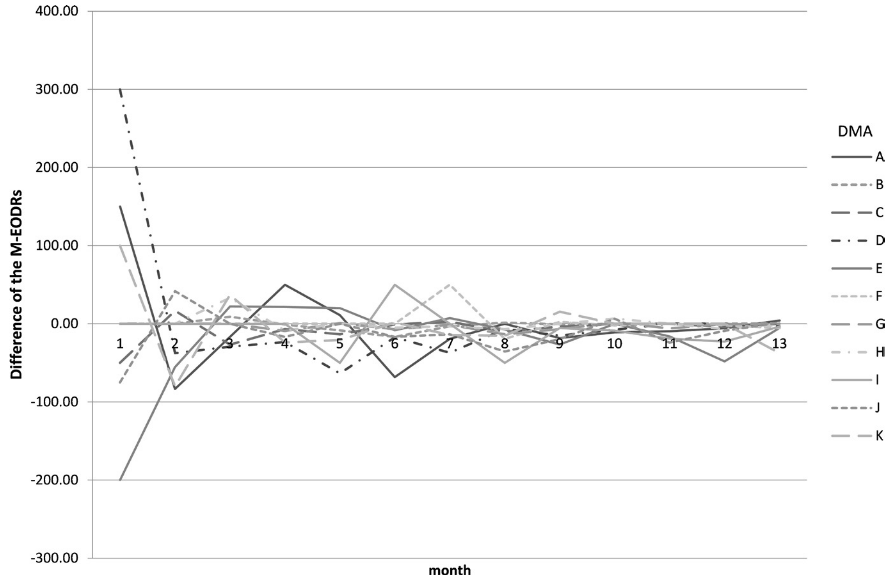

Fig. 2. Change of the M-EODRs according to the amount of accumulated data over time.

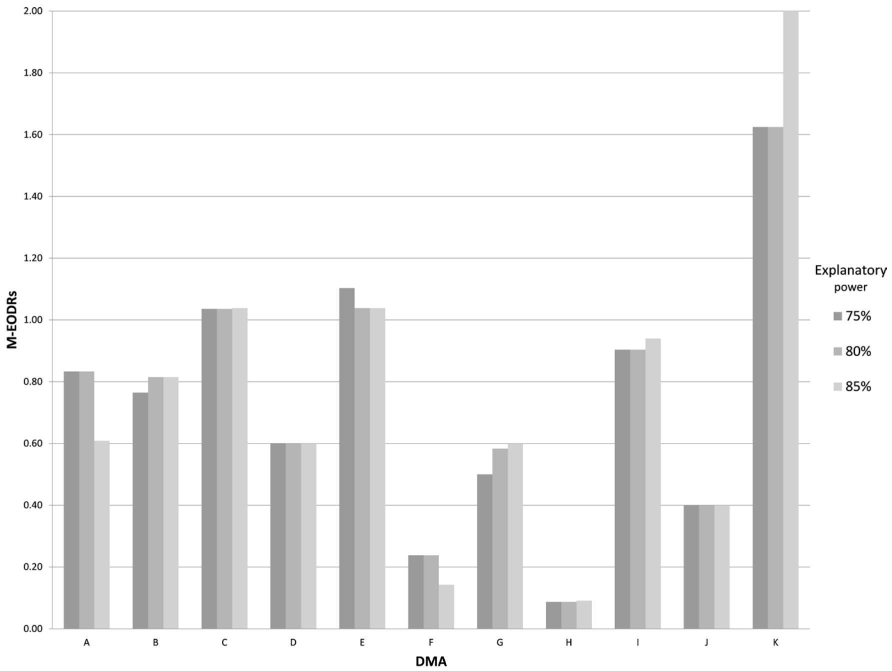

Fig. 3. M-EODRs according to the explanatory power of the eigenvectors.

Table 5 shows the changes in the BTWs and M-EODRs according to different p-values of F-distribution using the whole flow data. As can be analyzed from Table 5, the sensitivity of the p-value for the BTWs and M-EODRs were relatively high.

Table 6 shows the changes in the BTWs and M-EODRs according to different ODPs using the whole flow data. As can be analyzed from Table 6, the sensitivity of the ODPs for the BTWs and M-EODRs were relatively high.

As Table 5 shows, the M-EODRs calculated using the p-value of 0.05 resulted in higher M-EODRs with less number of M-EODRs that are greater than 100% than the case of the p-value of 0.03. It is better to have the values of the M-EODRs as less than or equal to 100% because the M-EODRs that are greater than 100% may not represent the realistic EODR due to the double-counted outliers.

If the manager of a water pipe network will rely on the estimated M-EODRs in performing preemptive pipe leak detection, he/she needs to be able to distinguish DMAs of higher priority for preemptive action due to a limited budget. Therefore, it is recommended that the manager will choose the p-value of F-distribution that produces various values of M-EODRs among the DMAs and less number of M-EODRs that are greater than 100%. The ODP is considered to be dependent on the reliability of leak records in a water pipe network. For instance, a network that has leak reports recorded on time will have a relatively short ODP. Therefore, it is recommended that the manager will choose an appropriate ODP based on the condition of the leak records keeping of the pipe network.

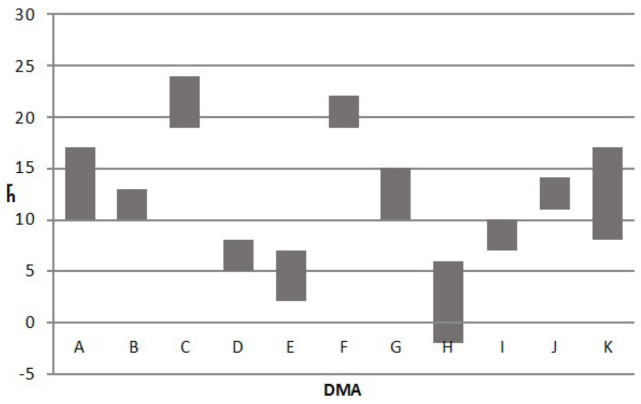

Fig. 4. BTW for the explanatory power of the eigenvectors of 75%.

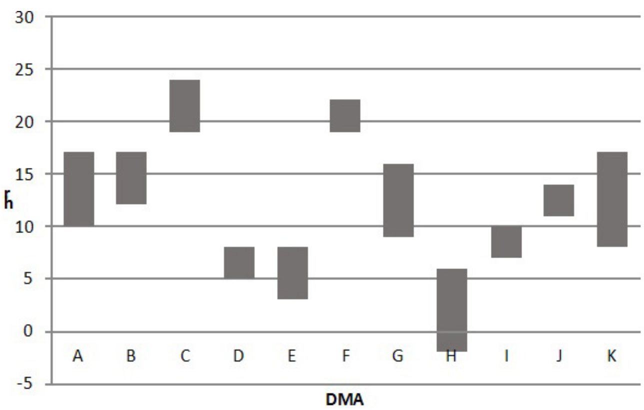

Fig. 5. BTW for the explanatory power of the eigenvectors of 80%.

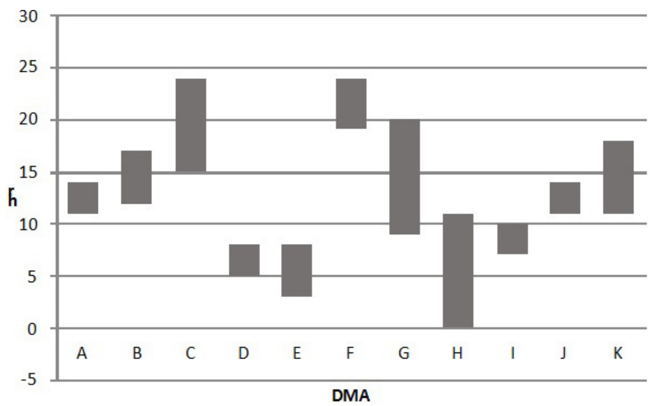

Fig. 6. BTW for the explanatory power of the eigenvectors of 85%.

4. Conclusions

In this paper, we have developed a computational algorithm to detect leak in a water pipe network using flow data of the DMAs in a case study area. The algorithm developed in this study was based on the PCA, one of the multivariate data analysis techniques, and was designed to improve the leak detection efficiency of a PCA which uses

Table 5
Changes in the BTWs and M-EODRs according to the p-value

p-value	Category	A	B	C	D	E	F	G	H	I	J	K
0.1	M-EODR (%)	83	81	104	60	104	24	58	9	90	40	163
	Center time (h)	14	15	22	7	6	21	13	2	9	13	13
	Time range (h)	7	5	5	3	5	3	7	9	3	3	9
0.05	M-EODR (%)	100	89	118	100	105	29	100	13	200	100	217
	Center time (h)	13	16	23	11	5	21	11	6	19	1	14
	Time range (h)	3	9	5	5	7	3	5	9	3	3	15
0.03	M-EODR (%)	65	109	150	200	111	100	200	13	129	133	250
	Center time (h)	16	16	20	8	5	21	19	4	14	9	23
	Time range (h)	7	9	5	9	7	3	5	5	13	3	21

Table 6
Changes in the BTWs and M-EODRs according to different ODPs

ODP (d)	Category	A	B	C	D	E	F	G	H	I	J	K
15	M-EODR (%)	83	81	104	60	104	24	58	9	90	40	163
	Center time (h)	14	15	22	7	6	21	13	2	9	13	13
	Time range (h)	7	5	5	3	5	3	7	9	3	3	9
11	M-EODR (%)	60	59	82	56	80	13	58	8	63	29	140
	Center time (h)	14	15	22	12	20	16	13	4	9	1	12
	Time range (h)	7	5	5	5	3	3	7	5	3	3	7
7	M-EODR (%)	33	43	57	56	67	13	30	6	38	29	100
	Center time (h)	14	14	23	12	6	16	8	7	9	1	12
	Time range (h)	7	5	5	5	3	3	3	7	3	3	7

24-h flow data. It is an algorithm that compares the dates which have outliers of the DMOD statistics from the PCA and the dates of completion of leak repair using only the flow data of a specific time window in a day.

The technique developed in this study may be used by a manager of a water pipe network to confirm whether the flow data of the previous day at present analysis time is calculated as the DMOD outliers by performing the developed PCA algorithm. Occurrence of the outliers and the calculated values of the M-EODRs may assist the manager of a water pipe network in deciding whether to conduct a leakage test for a suspected leak zone in a DMA.

Sensitivity analyses on the results of this study showed that the explanatory power of the eigenvectors did not have a high sensitivity on the BTWs and M-EODRs. Meanwhile, the sensitivity of the *p*-value and ODP for the BTWs and M-EODRs were relatively high. Therefore, it was concluded that the generally accepted value of 80% may be used for the explanatory power of the eigenvectors for the analyses. In the meantime, it is considered that the *p*-value of F-distribution needs to be chosen so that various values of the M-EODRs among the DMAs are produced. It is also considered that the manager will choose an appropriate ODP based on the condition of the leak records keeping of the pipe network.

Considering the analyses on the changes in the BTWs and M-EODRs according to the amount of data accumulated over time, applications of the developed method are considered to be dependent on the accuracy of the calculated M-EODR and the variability of the BTW for each DMA. In other words, a manager of a water pipe network may be able to perform a preemptive leak detection for a DMA with confidence if the calculated M-EODR is good enough, for example, more than 70%, and the estimated BTW does not change in a recent period, say for the last 6 months. However, the exact criteria regarding the M-EODR and BTW to use for preemptive maintenance activities for a DMA must be decided based on the rational knowledge and experience of the managers on the financial condition and maintenance practices of a water pipe network.

The method developed in this paper examined the changes in the BTW which resulted in the highest EODR (M-EODR) for a DMA. It was considered that if the BTWs do not change much in a recent period, say for 6 months, the calculated M-EODR may be used to determine whether preemptive leak detection is performed for a DMA. Therefore, BTW was used in the analyses under the assumption that there will be a specific time zone of each DMA in which the residuals (DMOD statistics) of a PCA model and the leak records have a stronger relationship than the case of using 24-h flow data. The leak records used in this study have only the dates of completion of leak repair.

Therefore, the specific reasons why some specific time zones had stronger relationship between the DMOD outliers and the leak records were hard to be analyzed due to limited information on the leak records used. It is conjectured that the estimated BTWs of a DMA may represent the time zone in which leakage can be distinguished more easily than the normal flows in a DMA. It is considered that further analyses need to be performed using more detailed information regarding the actual leak occurrence times.

Acknowledgment

This subject is supported by Korea Ministry of Environment as "Global Top Project (code 2016002130005)".

References

[1] A. Lambert, W. Hirner, Losses from Water Supply Systems: Standard Terminology and Performance Measures, IWA the Blue Pages, 2000, pp. 1–13.

[2] D. Covas, H. Ramos, N. Lopes, A.B. Almeida, Water Pipe System Diagnosis by Transient Pressure Signals, Water Distribution Systems Analysis Symposium, 2006, pp. 1–19.

[3] Z. Kapelan, D.A. Savic, G.A. Walters, Incorporation of prior information on parameters in inverse transient analysis for leak detection and roughness calibration, Urban Water J., 1 (2004) 129–143.

[4] J.A. Stathis, G.V. Loganathan, Analysis of Pressure-dependent Leakage in Water Distribution Systems, Analysis of Pressure-dependent Leakage in Water Distribution Systems, 29th Annual Water Resources Planning and Management Conference: Preparing for the 21st Century, 1999, pp. 1–10.

[5] J.M. Muggleton, M.J. Brennan, R.J. Pinnington, Y. Gao, A novel sensor for measuring the acoustic pressure in buried plastic water pipes, J. Sound Vib., 295 (2006) 1085–1098.

[6] R. Pilcher, S. Hamilton, H. Chapman, B. Ristovski, S. Strapely, Leak Location and Repair Guidance Notes, International Water Association, Water Loss Task Forces: Specialist Group Efficient Operation and Management, 2007.

[7] J.M. Muggleton, M.J. Brennan, Axisymmetric wave propagation in buried, fluid-filled pipes: effects of wall discontinuities, J. Sound Vib., 281 (2005) 849–867.

[8] S.R. Mounce, A.J. Day, A.S. Wood, A. Khan, A neural network approach to burst detection, Water Sci. Technol., 45 (2002) 237–246.

[9] L. Xia, L. Guo-jin, Leak Detection of Municipal Water Supply Network Based on the Cluster-Analysis and Fuzzy Pattern Recognition, International Conference on E-Product, E-Service, and E-Entertainment (ICEEE), 2010, pp. 1–5.
[10] L.A. Borges, M.A. Ramirez, Acoustic Water Leak Detection System, Water Distribution Systems Analysis Symposium, 2010, pp. 1–3.
[11] M. Tajima, A. Mita, Automatic Leakage Detection for Water Supply Systems Using Principal Component Analysis, Vol. 33, Materials Forum – Parkville – CD ROM Ed., 2009, pp. 10.
[12] G. Ye, R.A. Fenner, Kalman filtering of hydraulic measurements for burst detection in water distribution systems, J. Pipeline Syst. Eng. Pract., 2 (2010) 14–22.
[13] C.V. Palau, F. Arregui, A. Ferrer, Using multivariate principal component analysis of injected water flows to detect anomalous behaviors in a water supply system: a case study, Water Supply, IWA, 4 (2003) 169–181.
[14] S. Park, D. Jeon, S. Jung, J. Kim, D. Lee, Identifying an appropriate analysis duration for the principal component analysis of water pipe flow data, J. Korean Soc. Water Wastewater, 27 (2013) 351–361.
[15] S. Park, S.Y. Jung, V. Sahleh, Determining the time of least water use for the major water usage types in district metered areas, J. Korean Soc. Water Wastewater, 29 (2015) 415–425.