ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
Journal of Korean Society of Water and Wastewater Vol. 27, No. 3, pp. 351-361, June, 2013
http://dx.doi.org/10.11001/jksww.2013.27.3.351

Identifying an Appropriate Analysis Duration for the Principal Component Analysis of Water Pipe Flow Data

상수도 관망 유량관측 자료의 주성분 분석을 위한 분석기간의 설정

Park, Suwan* · Jeon, Daehoon · Soyeon Jung · Kim, Joohwan · Lee, Doojin

박수완* · 전대훈 · 정소연 · 김주환 · 이두진

부산대학교 공과대학 사회환경시스템공학부

Abstract : In this study the Principal Component Analysis (PCA) was applied to flow data in a water distribution pipe system to analyze the relevance between the flow observation dates, which have the outliers of observed night flows, and the maintenance records. The data was obtained from four small size water distribution blocks to which 13 maintenance records such as pipe leak and water meter leak belong. The flow data during four months were used for the analysis. The analysis was carried out to identify an appropriate analysis period for a PCA model for a water distribution block. To facilitate the analyses a computational algorithm was developed. MATLAB was utilized to realize the algorithm as a computer program. As a result, an appropriate PCA period for each of the case study small size water distribution blocks was identified.

Key words : analysis period, computational algorithm, night flows, Principal Component Analysis, water distribution pipe system
주제어 : 분석기간, 계산 알고리즘, 야간 유량, 주성분 분석, 상수관망

1. 서 론

1908년 우리나라 최초의 상수도 시설이 도입된 이후 약 100년의 역사를 지닌 한국의 상수도는 많은 발전을 이뤄 왔다. 2010년 12월말 현재 우리나라에서는 전국 162개 지방상수도사업자(특 · 광역시 7, 특별자치도 1, 시 73, 군81) 및 1개 광역 상수도사업자로부터 전체인구의 97.7 %인 약 50,264천명이 상수도를 공급받고 있다(상수도 통계, 2010). 특히 시 · 도 상수도 보급현황을 보면, 총 5곳의 수도보급률이 100 %를 나타낼 만큼 급속하게 발전되고 있는 현실이다. 우리나라 상수도 관망의 총연장은 165,800 km에 달하고 있으며 그중 6.4 %가 송수관, 50.8 %가 배수관, 40.8 %가 급수관으로 알려져 있고, 이를 통한 급수 총량은 5,910백만 m^3에 이른다. 이 가운데 누수량 등을 제외한 실제 유효수량은 5,267백만 m^3 이며, 수도요금이 부과된 양(유수량)은 4,920백만 m^3으로 나타났다(상수도 통계, 2010). 우리나라의 상수도 보급률은 2010년 97.7 %로 해가 갈수록 꾸준히 증가 하고 있지만 동기간 유수율은 83.2 %, 누수율은 10.8 %로 여전히 공급의 효율성에 문제가 있다는 것을 알 수 있다.

상수 관망은 노후가 될수록 관망의 파손 및 누수의 비율이 증가 하고, 도수능력 또한 감소한다. 그 결과 상수관망의 파손과 누수와 같은 관망의 이상 징후로 인하여 매년 막대한 손실이 발생하

• Received 16 December 2012, revised 10 June 2013, accepted 12 June 2013.
* Corresponding author: Tel : 051-510-2734 Fax : 051-513-9596 E-mail : swanpark@pusan.ac.kr

고, 유수율과 누수율의 관리를 어렵게 하며, 수자원 확보가 시급한 현실에서 많은 양의 수자원 손실과 정수가 된 물의 누수라는 이중적 손실을 가져 온다. 이러한 직접적 피해와 더불어 누수된 물이 토사로 침투되면서 지반 또한 약하게 만들어 지반 침하 사고를 초래하기도 한다. 상수도 관련 민원 현황을 살펴봐도 2007년 이후 부터 현재 까지 누수에 관한 민원이 꾸준히 증가 하고 있는 것은 상수 관망의 노후화로 인한 누수에 따른 현상일 것으로 짐작할 수 있다. 이러한 문제점을 해결하기 위해서는 상수도 시설의 발전과 적절한 유지관리를 통하여 누수율을 감소시키고 유수 율을 향상하여야 한다. 누수율 감소와 유수율 향상을 위해서는 상수 관망의 누수와 같은 관망의 이상 징후를 탐지하여 누수를 예방하는 것이 하나의 방편이 될 수 있다. 본 논문에서는 누수 혹은 비정상적 수요발생과 같은 관망의 이상 징후를 예측하는데 적용될 수 있는 통계분석 기법 중 주성분 분석기법의 모형구축 방안에 대하여 논하였다.

본 연구에서는 주성분 분석기법으로 계산되는 상수관망 관측유량의 주성분 분석 이상치 발생 날짜와 유지관리 기록을 비교분석하였으며, 신뢰도 높은 주성분 분석 모형을 구축하는데 필요한 주성분 분석기간 선정 방법론에 대하여 논하였다. 주성분 분석 기법을 이용하여 T^2Hotelling 통계치와 DMOD 통계치의 이상치 판정 방법을 사용하되, 실제 관망의 유량관측 자료를 이용하여 누수와 같은 유지관리 기록이 존재하는 날짜와 관측유량의 주성분 분석이상치 발생 일자를 비교함으로써 신뢰도 높은 주성분 분석 모형을 구축하는데 가장 적합한 주성분 분석기간을 결정하는 기법을 개발하였다.

2. 상수관망의 누수관리 기법

관망의 누수관리와 관련하여 연구되어온 내용은 누수량 산정 방법, 누수 위치 탐지법, 그리고 누수 제어 기법으로 대별된다. 각 연구 방법의 최근 연구 사례를 들면, 누수량 산정 방법은 Top-down 방식을 사용한 Lambert and Hirner(2000)의 IWA 방법이 있으며, Bottom-up 방식을 이용한 Covas, et al.(2006)의 야간 최소유량 분석법이 있다. 누수 위치 탐지 법으로는 최적화 기법과 수리학적 모형(Kapelan, et. al, 2004; Stathis and Loganathan, 1999) 또는 천이류 흐름해석 기법을 이용하는 누수 발생 판정방법이 있고, 음향 로깅 (Muggleton, et al. 2006), 단계적 시험법(Pilcher, et al., 2007) 또는 지반 운동 센서와 지반 투과 레이더(O'Brien, et al., 2003)를 사용하는 누수 고립(localisation) 방법이 있으며, 또한 누수-소음 상관관계법(Muggleton and Brennan, 2005), 가스 투입법 (Farley and Trow, 2003) 또는 피그 장착 음향 탐지법(Mergelas and Henrich, 2005)을 이용하는 정확한 누수위치 탐지(leakage pin-pointing) 방법이 있다.

누수 제어 기법은 크게 수동적 누수 제어와 능동적 누수 제어로 나눌 수 있으며, 능동적으로 누수를 제어하기 위하여 압력관리와 같은 기법이 많이 개발되어 왔다. 최근 들어서는 효율적인 누수관리를 위하여 수리학적 모델링 소프트웨어와 GIS 및 SCADA를 통합하여 하나의 패키지(Tabesh and Delavar, 2003)로 만들려는 시도가 있었으며, 인공지능의 한 분파인 인공 신경망을 이용하여 수요량을 예측(Bougadis, et al., 2005)한다거나, 관로 파손사건을 탐지하기 위하여 방대한 양의 관망 운영 자료와 과거 기록을 탐색한 사례(Mounce, et al., 2009)도 누수 제어 기법 개발의 하나로 볼 수 있다.

상수관망 누수감지 알고리즘으로는 유량 및 압력 시계열 자료에 대해 인공신경망을 적용하여 DMA 내의 누수 판정 및 하루 후의 누수 여부를 예측한 Mounce, et al.(2002)의 연구사례가 있다. Mounce, et al.(2002)은 누락된 자료를 보강하기 위해 통계적 기법(ARIMA based filter)을 사용하였으며, 자료를 정규화(linear re-scaling)한 뒤 자료를 재구성(tapped delay

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
Journal of Korean Society of Water and Wastewater Vol. 27, No. 3, pp. 351-361, June, 2013

line format)하여 Fuzzy 추론 기법에 따라 누수를 판정하였다. Mounce, et al.(2002)은 인공신경망을 훈련시키기 위해 몇 개월 정도의 자료가 필요하며, 모니터링 되는 지역의 특성(밸브 작동 등)이 바뀔 경우에는 새로운 자료가 필요하다고 분석하였다. Xia and Guojin(2010)은 군집분석과 퍼지 추론 기법을 결합한 방법론을 제시하였는데, 이들은 수리 시뮬레이션을 통하여 누수 사건에 대해 비슷한 압력 변화를 보이는 관로들을 그룹으로 구분하고 데이터베이스로 구축하였다. 이를 이용하여 Xia and Guojin(2010)은 새로운 누수 사건이 발생하였을 경우 이 누수 사건에 따라 모니터 되는 압력의 변화가 어떤 '압력 변화 관로그룹'에 속하는 지를 판정하였다.

또 다른 상수관망 누수감지 알고리즘은 SVM (Support Vector Machine)으로 인공신경망과 비교하여 고차원의 자료를 효율적으로 사용 가능하고, 적은 자료로 트레이닝이 가능하다. SVM은 초음파 센서 자료를 이용하여 상수관망뿐 아니라 송유관 및 가스관의 관두께 감소 및 파열 등의 이상 징후를 감시하는데도 쓰이고 있으나, 상수관망의 누수탐지 분야는 아직 실험실 규모의 소구경 관로에 대해 연구되고 있는 실정이다. 상수관망에 적용된 예로는 Borges and Ramirez(2010)가 실제 관망에서 수집된 전자청음 자료를 분석한 예가 있으나, 그 결과는 그리 만족스럽지 못한 것으로 보고되었다. 한편, Tajima and Mita(2009)는 SVM을 이용한 연구에서 누수 판정 적중률 90 %의 결과를 보고하였다.

칼만 필터(Kalman Filter)를 누수감지에 활용한 사례를 보면 Ye and Fenner(2011)는 다수의 DMA에서 측정한 유량과 압력 자료에 칼만 필터링을 적용하여, 필터의 출력과 측정값의 차이로부터 관망의 이상 징후를 판정하였다. DMA(블록) 인입 지점의 압력 및 유량 모니터링 자료를 사용하여, 15분 간격으로 모니터하고, mobile network 내의 GPRS(general packet radio service)를 통하여 30분 간격으로 전송하였다. 전송된 자료를 토대로 정상적인 '시그널'과 비정상적인 '노이즈'를 '필터링' 하고, 필터의 오차(관측한 유량 및 압력 값과 예측한 유량 및 압력 값과의 차이)를 분석하여 누수 여부를 판정하였는데, 즉 하루 중 같은 시간대의 과거 유량 및 압력 자료의 변화를 분석하여 비정상적인 유량 또는 압력 발생 여부를 판정(유량 자료의 경우 누수량도 산정 가능)하였다.

상수 관망 이상 징후 판정 방법 중 주성분 분석을 이용한 Palau, et al.(2003)는 상수 관망에서 관측한 유량 자료를 이용하여 T^2Hotelling 통계치와 DMOD 통계치를 구하고 이를 T^2Hotelling 임계값 및 DMOD 임계값과 비교하여 통계치가 임계값을 초과하면 이상치로 판정하였다. 그러나 Palau, et al.(2003)는 신뢰도 높은 주성분 분석 모형을 구축하기 위한 분석기간을 설정하기 위한 연구는 수행하지 않았다.

3. 주성분 분석 기본 이론

3.1 주성분 분석의 개념

주성분 분석(Principal Component Analysis, PCA)이란 다변량 자료 분석 기법 중의 하나로 해석하려고 하는 다차원의 자료(data)에 포함된 정보의 손실을 최소한으로 하여 다차원의 변수를 보다 낮은 차원의 자료로 축약하는 기법이다. 따라서 가급적 적은(2 ~ 3개) 새로운 변수(주성분)로 전체의 정보를 표현하려고 하는 것이 주성분 분석의 주된 개념이다. 주성분들은 서로 통계적으로 독립적이며, 도출된 모든 주성분을 사용할 경우 원자료가 가지는 정보의 손실이 없다. 제 1 주성분은 자료의 변동을 가장 많이 설명하고, 2번째 및 3번째 주성분 등으로 구해지는 그 외의 주성분들은 제 1 주성분으로 설명되지 않은 나머지 자료의 변동을 설명하며 그 설명력은 점차 줄어든다.

주성분 분석은 $n \times m$ 차원(여기서, n은 관찰의 수, m은 변수의 수를 의미)을 가지는 행렬에 대한 분산-공분산(variance-covariance) 행

렬 X를 주성분 점수(scores) 행렬 T($n \times f$ 차원)와 부하(loading) 행렬이라고 불리는 $m \times f$ 차원의 행렬 P 및 잔차(residual)행렬 E($n \times m$ 차원)로 분할(factorization)하는 방법으로 식(1)과 같다.

$$X = T \times P^T + E \quad (1)$$

여기서, f는 주성분의 개수이며, $f < m$의 크기를 갖는다. 부하 행렬 P의 열(column)은 X의 분산-공분산 행렬의 고유값(eigenvalue)에 대한 고유벡터(eigenvector)를 나타낸다. 보통 고유벡터는 해당 고유값의 크기순으로 정렬되며, 분석 목적에 적합한 주성분의 개수만큼 선택된다. 가장 최적으로 분할하는 조건은 주어진 요소에 대하여 잔차 행렬을 최소화하는 것이다. 이 기준을 만족시키기 위하여 P 행렬의 열(column)은 X의 분산-공분산 행렬의 고유값 중에서 큰 순서대로 f만큼 택하여 이에 해당하는 고유 벡터로 구성한다. 고유값은 주성분(Principal Component, PC)으로부터 데이터를 다시 복원할 때 해당하는 데이터에 대한 가중치의 역할을 하게 된다. 구해진 고유값 중 큰 고유값에 대한 고유벡터를 원자료에 곱해주면 원자료를 고유벡터 축 상의 값으로 변환(투영)할 수 있다.

3.2 주성분 분석의 이상치(outlier)

유량자료의 주성분 분석 이상치를 판정하는 목적은 이상치를 가지는 날짜의 유량자료를 주성분 분석에서 제외함으로써 더욱 신뢰도 높은 주성분 분석 모형을 구축하기 위함이다. 이러한 신뢰도 높은 주성분 분석 모형을 사용함으로써 탐지된 비정상적 수요 및 누수와 같은 관망의 이상징후에 대한 신뢰도를 더욱 높일 수 있다. 유량자료의 이상치를 판정하기 위하여 두 가지 방법을 쓸 수 있는데, 하나는 자료의 분산을 이용하여 비교적 큰 이상치(strong outlier)를 판정하는 T^2Hotelling 통계치를 이용하는 방법과 주성분 분석 모델과 관측 값의 오차를 이용하여 비교적 크지 않은 보통의 이상치(moderate outlier)를 판정하는 DMOD 통계치를 이용하는 방법이 있다.

3.2.1 T^2Hotelling 통계치를 이용한 방법

T^2Hotelling 통계치은 매일 관측되는 유량 관측치에 의해 계산된 Score와 주성분 중심축 사이의 차이를 나타낸다. T^2Hotelling 통계치는 식(2)와 같다.

$$T^2Hotelling_i = \sum_{a=1}^{A} \frac{t_{iA^2}}{S_{tA^2}} = \quad (2)$$

$$\begin{bmatrix} t_{i,1} \\ t_{i,2} \\ \vdots \\ t_{i,A} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{S_{t1^2}} & 0 & 0 \\ 0 & \frac{1}{S_{t2^2}} & 0 \\ 0 & 0 & \frac{1}{S_{t3^2}} \end{bmatrix} \cdot [t_{i,1}, t_{i,2} \cdots t_{i,A}] = t_i^T \cdot S_t^{-1} \cdot t_i$$

여기서 A는 주성분의 개수, t_i는 i번째 유량 관측치를 주성분 축에 투영한 주성분 점수를 나타낸다. S_t는 Score Matrix T의 공분산이며, S_{tA}는 Score Matrix T의 A번째 열의 공분산을 나타낸다.

T2 Hotelling 통계치는 m과 $n-m-1$의 자유도를 가지는 F-Snedecor 확률 분포를 가지며, 일반적으로 사용되는 신뢰도 판정 기준은 95 %이다. n은 관측의 개수, m은 변수의 개수이다. 다음의 식(3)으로 이상치를 판정한다.

$$T_i^2 > \frac{m(n-2)}{n-m-1} \cdot F_{(m,n-m-1)} \quad (3)$$

따라서 $m(n-2) \times F_{\text{critical}}$ $(p=0.05)/(n-m-1)$과 비교하여 T_i^2 가 크다면 i번째 관측일의 유량자료에 대한 T^2Hotelling 통계치는 이상치로 판정된다.

3.2.2 DMOD 통계치를 이용한 방법

DMOD 통계치를 이용한 방법은 주성분 분석 모델의 예측값과 관측 값의 오차를 이용하여 비

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
Journal of Korean Society of Water and Wastewater Vol. 27, No. 3, pp. 351-361, June, 2013

교적 크지 않은 이상치(moderate outlier)를 판정하는 방법으로, 주성분 그림에 이상치로 표시하기에 강하지 않은 보통의 이상치를 판정하는 방법이다. 보통의 이상치는 각 관측의 잔차에 의해 확인된다. DMOD의 기하학적 의미를 개념적으로 나타내면 Fig. 1과 같은데, DMOD는 3차원 자료에 대한 주성분 분석으로 부터 2개의 고유벡터를 사용할 경우 3차원의 원자료의 값과 도출된 고유벡터에 따른 PC1 축과 PC2 축에 의한 2차원 평면까지의 거리를 나타낸다.

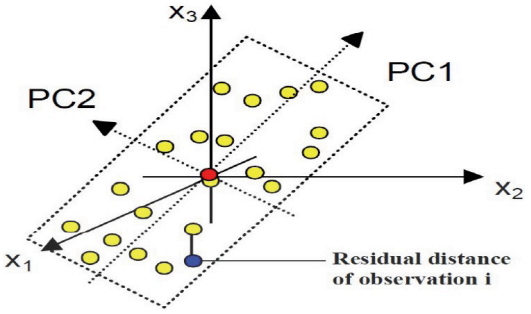

Fig. 1. Geometric Analysis of Model Residual Distance

DMOD를 구하는 식은 S_i/S_o이다. S_i는 관측 값의 모델에 대한 절대잔차를 나타내며, S_o는 모델의 평준화된 잔차를 나타낸다. $(S_i/S_o)^2$은 $(K-A)$와 $(N-A-1)(K-A)$의 자유도를 가진 F-Snedecor 확률 분포를 가진다. S_i와 S_o는 식 (4)와 (5)로 구할 수 있다.

$$S_i = \sqrt{\frac{\sum_{k=1}^{K} e_{ik}^2}{(K-A)}} \tag{4}$$

$$S_o = \sqrt{\frac{\sum_{i=1}^{N}\sum_{k=1}^{K} e_{ik}^2}{(N-A-A_0)\times(K-A)}} \tag{5}$$

여기서 e_{ik}는 k번째 변수에 대한 i번째 관측 값의 잔차, K는 원자료 변수의 개수, N은 관측 자료의 개수, A는 주성분의 개수, A_0는 자료를 표준화했으면 1을, 그렇지 않다면 0을 사용한다. 산정한 DMOD 통계치의 값을 이용하여 다음의 식(6)으로 이상치를 판정한다.

$$(S_i/S_o)^2 > F_{(m,n-m-1)} \tag{6}$$

4. PCA 이론의 적용 및 알고리즘 개발

4.1 사용된 유량관측 자료 및 유지관리 기록

본 연구에서 사용된 유량자료는 A시 상수 관망의 소블록에 대한 4개월가량의 유량 관측자료이다. 본 연구에 사용된 관망의 유지관리 기록이 존재하는 관망 내 소블록은 A-1, A-2, A-3 및 A-4의 총 4개 소블록이다. 이러한 소블록에서 관측되어 수집된 유량자료는 2011년 9월 1일부터 12월 31일까지의 자료이며, Table 1은 수집된 유량 자료를 A-1블록의 예를 들어 보인 것이다. 유지관리 기록이란 본 연구에 이용된 연구 대상의 관망에서 누수 사고 및 복구와 같은 누수와 관련된 기록으로서 관망에서 누수 사고 등이 발생하여 이러한 내용이 기록된 날짜들을 의미한다.

누수 혹은 비정상적 수요량 발생 등과 같은 관망의 이상징후는 주간시간대 보다 야간시간대

Table 1. Observed flow data for A-1 Block

(unit of flow : m³/h)

Time (hr) / Date	1	2	3	4	5	6	…	19	20	21	22	23	24
2011-09-01	26	21	22	19	26	49		74	70	65	52	37	34
2011-09-02	29	23	21	19	35	41		79	76	62	59	45	30
2011-09-03	29	24	21	27	32	41		94	76	63	56	47	33
							⋮						
2011-12-30	42	39	39	35	50	46		61	70	64	49	47	45
2011-12-31	37	34	34	30	35	38		-	-	-	-	-	-

에 그 현상이 더욱 잘 탐지될 수 있다. 따라서 본 연구에서는 수집된 유량자료 중 0시 부터 6시까지의 야간 시간대의 유량자료를 분석에 사용하였다.

4.2 야간유량 관측 자료를 이용한 주성분 분석 과정

관망의 야간시간대에 관측된 유량자료를 이용하여 주성분 분석을 실시하는 과정은 다음과 같다. 관망에서 준비된 유량자료 중 야간시간대(0시에서 6시 사이)에 측정된 유량자료만을 추출하여 식 (7)과 같은 행렬의 형태로 준비한다.

$$Z=\begin{bmatrix} Q_{1,0} \cdots Q_{1,6} \\ \vdots \quad \ddots \quad \vdots \\ Q_{N,0} \cdots Q_{N,6} \end{bmatrix} \tag{7}$$

여기서 Q_{ij}는 i 번째 관측일의 j 시간에서의 관측 유량을, N은 관측치의 개수를 의미한다.

다음으로 식(7)의 Z-matrix를 식(8)과 같이 표준화시킨다.

$$X=\begin{bmatrix} \frac{(Q_{1,0}-\overline{Q_0})}{\sigma_0} \cdots \frac{(Q_{1,6}-\overline{Q_6})}{\sigma_6} \\ \vdots \quad \ddots \quad \vdots \\ \frac{(Q_{N,0}-\overline{Q_0})}{\sigma_0} \cdots \frac{(Q_{N,6}-\overline{Q_6})}{\sigma_6} \end{bmatrix} \tag{8}$$

$$\overline{Q_k}=\frac{\sum_{n=1}^{k} Q_{n,k}}{N} \qquad \sigma_k=\frac{\sum_{n=1}^{k}(Q_{n,k}-\overline{Q_k})^2}{N-1} \tag{9}$$

여기서 $\overline{Q_k}$는 k시간에서의 평균 유량, σ_k는 k시간에서의 유량에 대한 표준편차이다.

행렬 X의 주성분은 Xp=λp 를 만족시키는 고유벡터 p와 스칼라 값을 가지는 고유값 λ를 구하고 다음 식(10)을 이용하여 계산된 행렬 X에 대한 주성분 점수(score) 행렬 T의 각 열로서 계산된다.

$$T = X \times p \tag{10}$$

여기서 X는 크기가 $N\times7$인 야간시간대 유량 관측 자료의 분산-공분산 행렬을 나타내며, p는 크기가 7×1인 X에 대한 고유벡터이다. 고유값 λ는 원자료의 분산-공분산 행렬 X와 단위행렬 I를 이용한 행렬 X의 특성방정식인 |X-λI| = 0인 조건을 이용하여 구한다. 고유벡터는 각 고유값에 대한 $Xp_i=\lambda p_i$ 혹은 (X-λI)p = 0을 만족하는 0이 아닌 벡터 p_i가 λ_i에 대한 고유벡터이다.

원자료로부터 도출된 모든 주성분을 사용할 경우 원자료가 가지는 정보의 손실은 없으나, 원자료가 가지는 차원을 축약하는 것이 주성분 분석의 목적이므로 원자료의 특성을 충분히 설명할 수 있는 주성분의 개수를 적절히 결정하여야 한다. 주성분의 개수는 주성분 점수 행렬 T의 열의 개수를 결정하며, 본 연구에서는 주성분의 원자료 특성 설명력이 90 %를 초과하는 주성분의 개수를 주성분 분석 모형의 구축에 사용하였다. 주성분 분석 모형의 구축을 통하여 계산된 T^2Hotelling 통계치의 이상치와 DMOD 통계치의 이상치를 가지는 유량관측일을 판정할 수 있으며, 보다 신뢰도 높은 주성분 분석 모형을 구축하기 위해서는 이러한 이상치를 가지는 유량관측일을 제거한 후 다시 주성분 분석 모형을 구축하여야 한다. 여기서 T^2Hotelling 통계치와 DMOD 통계치는 매 관측일에 대해 계산되며, 따라서 매 관측일의 T^2Hotelling 통계치와 DMOD 통계치가 각 통계치의 이상치를 초과하는지의 여부는 식 (3)과 (6)에 의해 판단된다.

한편 이상치를 가지는 유량관측일은 주성분 분석 기간에 따라 달라지므로, 신뢰도 높은 주성분 분석 모형을 구축하기 위해서는 이상치를 가지는 유량관측일과 실제 유지관리기록이 존재하는 날짜와의 연관성이 높게 나오는 적절한 주성분 분석기간을 결정하여야 한다. 이는 일차적으로 주성분 분석 모형을 구축한 다음 T^2Hotelling 통계치와 DMOD 통계치의 이상치를 가지는 날짜의 유량자료를 제거한 후 다시 주성분 분석 모형을 구축하기 위함이다. 이러한 신뢰도 높은 주성분 분석 모형을 이용하여 미래에 발생할 수 있는 관망의 누수 혹은 비정상적 수요량의 발생과

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
Journal of Korean Society of Water and Wastewater Vol. 27, No. 3, pp. 351-361, June, 2013

같은 관망의 이상징후를 탐지할 수 있다.

본 연구에서는 적절한 주성분 분석기간을 결정하기 위한 기준으로 탐지된 이상치가 존재하는 관측일과 누수보수와 같은 유지관리 기록 간에 높은 관련성을 가지는 분석기간을 파악하는 것으로 하였다. 따라서 본 연구에서는 주성분 분석 기법의 T^2Hotelling 통계치와 DMOD 통계치의 이상치 판정 방법을 이용하여 누수보수와 같은 유지관리 기록이 존재하는 날짜와 관측유량의 주성분 분석 이상치 발생 일자를 비교하므로써 신뢰도 높은 주성분 분석 모형을 구축하는데 가장 적합한 주성분 분석기간을 결정하기 위한 컴퓨터 계산 알고리즘을 개발하였다.

4.3 주성분 분석기간 결정 알고리즘

본 연구에서 분석에서 제거되어야 할 유량자료를 판정하기 위한 이상치의 기준은 주성분 분석 이상치 발생 날짜와 유지관리 기록일의 일치여부로 하였다. 그러나 실제 누수가 발생하여 민원이 발생하거나 사업자가 인지하기까지는 3일에서 7일까지의 시간이 걸리는 것으로 가정하여 탐지된 이상치가 존재하는 관측일과 유지관리 기록 간의 관련성을 분석할 수 있는 컴퓨터 계산 알고리즘을 개발하였다. Fig. 2는 이와 같은 분석을 수행하기 위하여 본 연구에서 개발된 계산 알고리즘을 나타낸다.

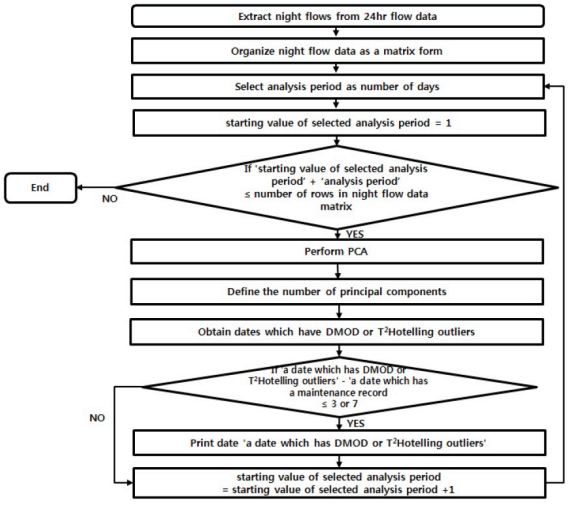

Fig. 2. Algorithm for the Principal Component Analysis of Night Flow Data

주성분 분석과 이상치가 존재하는 관측일에 대한 판정은 그 결과가 자료의 크기에 따라 바뀔 수 있다. 따라서 본 연구에서 개발된 알고리즘에서는 Fig. 2와 같이 '주성분 분석 기간 설정'에서 분석하고자 하는 관측기간을 설정한 다음, 설정된 기간에 해당하는 유량자료에 대한 주성분 분석을 실시하였다. 사용된 주성분 분석 기간은 20일 부터 10일 간격으로 90일까지이다. 설정된 각 분석 기간 동안 매일의 DMOD 통계치와 T^2Hotelling 통계치를 계산한 후 분석 기간 동안의 DMOD 통계치와 T^2Hotelling 통계치에 대한 임계값을 초과하는 날짜를 탐지하고, 이러한 날짜가 유지관리 기록이 존재하는 날짜보다 3일 또는 7일 이전 내에 발생한 날짜를 파악하였다.

설정된 각 분석기간은 원자료가 존재하는 총 기간인 2011년 9월 1일부터 2011년 12월 31일까지의 4개월에 비하여 작으므로, 각 분석기간에 대해 Fig. 2에서 '설정된 기간의 시작값'을 2011년 9월 1일부터 '1'만큼 증가시켜 유량관측 자료가 시작하는 날짜를 하루만큼 옮긴 후 분석 기간의 크기에 따라 재설정된 기간에 해당하는 유량자료에 대한 주성분 분석을 실시하였다. 각 분석 기간에 대한 주성분 분석 실시 횟수는 각 분석 기간에 따라 다른데, 예를 들어 분석 기간을 30일로 하였을 경우 첫 번째 분석 기간의 시작일을 2011년 9월 1일로 하여 분석을 하고, 그 이후로는 분석 시작일을 하루씩 옮겨 분석 기간의 마지막 날이 2011년 12월 31일이 될 때까지 분석을 실시함으로써 총 93번의 주성분 분석을 실시하였다. 즉 분석기간을 30일로 하였을 경우 첫 번째 주성분 분석기간은 2011년 9월 1일부터 2011년 9월 30일까지이며, 두 번째 분석기간은 2011년 9월 2일부터 2011년 9월 31일까지이다. 본 연구에서는 이상과 같은 알고리즘을 범용 과학기술계산용 소프트웨어인 MATLAB으로 구현하였다.

탐지된 이상치가 존재하는 관측일과 유지관리 기록 간의 관련성을 분석하는 목적은 주성분 분석에 따라 계산된 이상치 발생 날짜와 유지관리 기록을 비교하여 가장 적절한 주성분 분석기간을 결정하기 위함이다. 이를 위하여 본 연구에서 개발된 Fig. 2의 알고리즘을 이용하여 탐지되는 주성분 분석 이상치 발생 날짜가 분석기간이 바뀜에 따라 반복해서 발생하는 현상을 분석하였다. 탐지되는 주성분 분석 이상치 발생 날짜 중 비교적 적은 횟수로 탐지되는 이상치 발생 날짜는 비교적 많은 횟수로 탐지되는 이상치 발생 날짜에 비하여, 그 날 관측된 유량이 분석에 사용된 전체 유량자료에 대한 전반적인 주성분 분석 모형의 이상치 발생에 기여하는바가 크지 않을 것으로 판단하였다. 왜냐하면 이상치는 전체 관측유량자료와의 상대적 관계에 의해서 판정되기 때문이다. 따라서 반복해서 발생하는 이상치 발생 날짜의 개수가 5개 또는 10개 미만인 날짜는 탐지된 이상치가 존재하는 관측일과 유지관리 기록 간의 관련성 분석에서 제외하였다.

탐지된 이상치가 존재하는 관측일과 유지관리 기록 간의 관련성 판정을 위한 분석 방법을 정리하면 실제 누수가 발생하여 민원이 발생하거나 사업자가 인지하기까지의 시간 및 각 분석기간에 따른 총 분석횟수 동안 반복해서 탐지되는 이상치 발생 날짜의 개수로 요약된다. 이러한 방법을 사용하여 두 가지 경우에 대한 분석을 실시하였는데, 그 중 하나는 이상치 발생 날짜와 유지관리 기록의 차이가 3일 이내인 날짜를 선별하되, 분석기간에 따른 분석들에서 반복해서 발생하는 이상치 발생 날짜의 개수가 5개 미만인 날짜를 분석에서 제외한 경우(분석유형 A)이며, 다른 하나는 이상치 발생 날짜와 유지관리 기록의 차이가 7일 이내인 날짜를 선별하되, 분석기간에 따른 분석들에서 반복해서 발생하는 이상치 발생 날짜의 개수가 10개 미만인 날짜를 분석에서 제외한 경우(분석유형 B)이다. 이러한 분석 결과는 Table 2 및 Table 3과 같다.

Table 2 및 Table 3에서 보인, 괄호를 사용하지 않은 숫자는 DMOD 및 T^2Hotelling 통계치를 초과하는 이상치가 발생하는 날짜가 '유지관리 기록'이 존재하는 날짜의 3일 또는 7일 이내인 날짜를 분석 기간의 최초날짜(2011년 9월 1일)로부터의 경과 일수로 나타낸 것이다. 한편 Table 2 및 Table 3에서 괄호 안의 숫자는 각 분석기간에 대해 계산된 이상치 발생 날짜, 즉 괄

Table 2. Analysis type A

Analysis Period	A-1		A-2		A-3		A-4	
	DMOD(%)	T^2(%)	DMOD(%)	T^2(%)	DMOD(%)	T^2(%)	DMOD(%)	T^2(%)
20	50 (100)	- (0)	- (0)	- (0)	- (0)	- (0)	- (0)	- (0)
30	50 (25)	- (0)	57,116 (29)	55,64,113,114 (31)	61 (9)	35,36,64 (27)	- (0)	- (0)
40	50 (9)	- (0)	55,116 (15)	55,64,114 (20)	61 (6)	35,36,61,64,76 (36)	56 (10)	- (0)
50	- (0)	50 (20)	57 (8)	55,64,114 (23)	61 (8)	35,36,61,64,76 (33)	56 (8)	- (0)
60	- (0)	50 (13)	57 (8)	64,114 (17)	61,105 (12)	35,36,61,64,76 (36)	56 (8)	- (0)
70	- (0)	- (0)	- (0)	64,114 (17)	61 (7)	35,36,61,64,76 (36)	56 (8)	- (0)
80	- (0)	- (0)	- (0)	64,114 (17)	61 (9)	35,36,61,64,76 (38)	- (0)	- (0)
90	- (0)	- (0)	- (0)	55,64,114 (23)	61 (10)	35,36,64,76 (29)	- (0)	- (0)

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
Journal of Korean Society of Water and Wastewater Vol. 27, No. 3, pp. 351-361, June, 2013

Table 3. Analysis type B

Analysis Period	A-1		A-2		A-3		A-4	
	DMOD(%)	T^2(%)	DMOD(%)	T^2(%)	DMOD(%)	T^2(%)	DMOD(%)	T^2(%)
20	- (0)	- (0)	- (0)	- (0)	- (0)	- (0)	- (0)	- (0)
30	50 (50)	- (0)	57 (100)	52,53,59,64 (44)	57,61,71,101 (80)	31,36,101 (50)	- (0)	- (0)
40	50 (33)	46 (20)	44 (50)	52,53,59,64 (40)	57,58,60,61,72,100,101 (88)	31,36,61,64,76,101 (60)	56 (13)	- (0)
50	46 (25)	46,50 (50)	44 (17)	52,53,59,64 (40)	57,60,61,72,101 (45)	31,35,36,57,60,61,64,76,101 (69)	56 (13)	- (0)
60	46 (25)	- (0)	44,53 (25)	52,53,59,64 (40)	57,60,61 (38)	35,36,57,60,61,64,76,101 (73)	56 (11)	- (0)
70	46 (20)	- (0)	44,53 (29)	52,53,59,64 (40)	57,60,61 (38)	35,36,60,64,76 (50)	- (0)	- (0)
80	46 (20)	48 (10)	44,53 (33)	52,53,59,64 (40)	57,61 (25)	35,36,60,67,76 (45)	- (0)	- (0)
90	46 (20)	46 (13)	44 (14)	52,53,59,64 (40)	57 (17)	35,36,57,60,64,76 (50)	- (0)	- (0)

호를 사용하지 않은 숫자, 중 유지관리 기록이 존재하는 날짜의 3일 또는 7일 이내에 속하는 이상치 발생 날짜의 백분율을 나타낸 것이다. 예를 들어 A-4블록에 대해 분석기간 40일에 대한 33번째 분석결과에 의하면 DMOD 결과값에서 28번째, 즉 2011년 9월 1일로부터 54일째와 58일째인 10월 24일과 10월 28일의 유량값들이 이상치로 판정되었고, 누수보수기록은 2011년 9월 1일로부터 56일째인 10월 26일로 이상치가 발생한 날짜인 10월 24일 이후 3일 이내이므로 10월 24일의 이상치 판정이 누수를 감지한 것으로 판단하였다. 한편 분석기간 40일을 사용한 여러 번의 분석과정에서 발생한 DMOD 이상치 중 5번 이상 반복해서 발생한 이상치 발생 날짜는 9월 1일로부터 21, 27, 45, 56, 60, 74, 87, 86, 95 및 102번째 날짜로 총 10건의 이상치 탐지날짜가 계산되었다. 이중 56일째 이상치 만이 누수보수기록이 존재하는 날짜의 3일이내 이므로 계산된 이상치 발생 날짜의 누수탐지 백분율은 10%로 계산된다. 이러한 과정으로 계산된 결과를 Table 2에 정리하였다.

소블록별 적절한 주성분 분석기간을 결정하기 위한 기준은 각 분석기간에 대해 계산된 이상치 발생 날짜의 백분율(Table 2 및 Table 3에서 괄호 안의 숫자)이 가장 높은 분석기간을 찾는 것으로 하였다. T^2Hotelling은 주로 큰 이상치를 탐지하는데 쓰이고 일반적으로 DMOD 이상치는 T^2Hotelling 이상치에 포함된다(Palau, et al., 2003). 따라서 T^2Hotelling 통계치를 기준으로 이상치 발생 날짜의 백분율을 분석하고, T^2Hotelling 통계치의 값이 존재하지 않은 경우에는 DMOD 이상치 발생 날짜의 백분율이 가장 높은 분석기간을 찾는 것으로 하였다

A-1 소블록의 경우 모든 분석유형에서 분석기간 50일의 T^2Hotelling 이상치 발생 날짜 백분율이 가장 높고, A-2 소블록의 경우 모든 분석유형에서 분석기간 30일의 T^2Hotelling 이상치 발생 날짜 백분율이 가장 높은 것으로 나타났다. 따라서 A-1 소블록과 A-2 소블록에 대한 적절한 주성분 분석기간은 각각 50일 및 30일인 것으로 분석된다. A-4 소블록의 경우 모든 분석유형에서 T^2Hotelling 이상치 발생 날짜가 존재하지 않으므로 DMOD 이상치 발생 날짜의 백분율을 분석한 결과, 분석유형 A에서는 40일, 분석유형 B에서는 40일과 50일이 적절한 주성분 분석기간으로 파악된다. 따라서 A-4 소블록

은 40일 또는 50일이 적절한 주성분 분석기간인 것으로 분석된다. 한편 A-3 소블록의 경우는 분석유형 A에서는 80일이, 분석유형 B에서는 60일의 이상치 발생 날짜 백분율이 가장 높다. 그러나 분석유형 A에서 분석기간 60일과 80일의 이상치 발생 날짜 백분율이 큰 차이를 보이지 않으므로 A-3 소블록의 적절한 주성분 분석기간은 60일을 사용하여야 할 것으로 분석된다.

5. 요약 및 결론

본 연구에서는 탐지된 이상치가 존재하는 관측일과 유지관리 기록 간의 관련성을 분석하기 위하여 다변량 자료 통계분석 기법 중 주성분 분석법을 적용하였다. 연구대상 관망의 4개 소블록에서 4개월 동안 관측된 야간유량 자료에 대한 주성분 분석기법의 적용과 탐지된 이상치가 존재하는 관측일과 유지관리 기록 간의 관련성을 분석하기 위한 컴퓨터 계산 알고리즘을 개발하였으며, 이를 MATLAB 프로그램으로 구현하였다. 본 연구를 통한 결과를 요약하면 다음과 같다.

주성분 분석을 이용하여 연구대상인 A시의 소블록 별로 주성분 분석 기간을 20일, 30일, 40일, 50일, 60일, 70일, 80일 및 90일 단위로 하여 T^2Hotelling 통계치와 DMOD 통계치의 이상치를 가지는 날짜를 구하였다. 탐지된 이상치가 존재하는 관측일과 유지관리 기록 간의 관련성을 분석하기 위해 각 분석기간에 대해 계산된 이상치 발생 날짜 중 유지관리 기록이 존재하는 날짜의 3일 또는 7일 이내에 속하는 기준을 만족하고, 분석기간별 분석결과에서 반복해서 발생되는 이상치를 가지는 날짜의 개수가 5개 또는 10개 미만인 날짜를 제외하여 이상치가 존재하는 관측일과 유지관리 기록 간의 관련성을 분석하였다. T^2Hotelling 통계치를 기준으로 이상치 발생 날짜 중 유지관리 기록이 존재하는 날짜의 3일 또는 7일 이내에 속하는 날짜의 백분율이 가장 높은 분석기간을 찾고, T^2Hotelling 통계치의 값이 존재하지 않은 경우에는 DMOD 이상치 발생 날짜의 백분율이 가장 높은 분석기간을 확인하였다. 그 결과 분석대상 소블록에 대해 적절한 주성분 분석기간을 파악할 수 있었다.

관망의 이상징후를 판정하기 위한 최적의 주성분 분석기간은 관망의 특성에 따라 달리 도출되므로 주성분 분석기간 결정법을 이용한 주성분 분석 모형도 관망의 특성에 따라 달리 모형화된다. 본 연구에서 사용된 관망은 4개의 소블록이며, 유량관측 자료는 4개월에 유지관리 기록은 13개로 불과해 비교적 소규모의 자료가 사용되었다. 따라서 향후 좀 더 장기간에 걸친 유량관측 자료와 다수의 유지관리 기록을 여러 관망으로부터 획득하여, 본 연구를 통하여 개발된 주성분 분석기간 결정법을 이용한 주성분 분석 모형을 누수와 같은 관망의 이상징후 탐지에 적용하기 위한 추가적인 연구가 필요한 것으로 사료된다.

사 사

이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었음.

참고문헌

Ministry of Environment (2010) Water Suppy Statistics

Borges, L.A. and Ramirez, M.A. (2010) Acoustic Water Leak Detection System, In: Proceedings of the 7th International Telecommunications Symposium, pp. 1-3.

Bougadis, J. Adamowski, K. and Diduch, R. (2005) Shortterm municipal water demand forecasting, *Hydrological Processing*, **19** (1), 137-148.

Covas, D. Ramos, H. Lopes, N. and Almeida, A.B. (2006) Water pipe system diagnosis by transient pressure signals, In: Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium. Cincinnati, USA.

Farley, M. and Trow, S. (2003) Losses in water distribution networks, London: IWA Publishing.

Kapelan, Z. Savic, D.A. and Walters, G.A. (2004) Incorporation of prior information on parameters in inverse transient analysis for leak detection and roughness calibration, *Urban Water Journal*, **1**(2), pp.129－143.

Lambert, A. and Hirner, W.H. (2000) Losses from Water Supply Systems: Standard Terminology and Performance Measures, IWSA Blue Pages.

Mergelas, B. and Henrich, G. (2005) Leak locating method for precommissioned transmission pipelines: North American case studies, In: Leakage 2005 Conference Proceedings. Halifax, Canada.

Mounce, S.R. Boxall, J. and Machell, J. (2009) Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows, *Journal of Water Resources Planning and Management*, **136**(3), May/June 2010, pp. 309－318.

Mounce, S.R. Day, A.J. Wood, A.S. and Khan, A. (2002) A neural network approach to burst detection, *Water science and technology*, **45**(4－5), pp. 237－246.

Muggleton, J.M. and Brennan, M.J. (2005) Axisymmetric wave propagation in buried, fluid－filled pipes: effects of wall discontinuities, *Journal of Sound and Vibration*, **281**(3－5), pp.849－867.

Muggleton, J.M. Brennan, M.J. Pinnington, R.J. and Gao, Y. (2006) A novel sensor for measuring the acoustic pressure in buried plastic water pipes, *Journal of Sound and Vibration*, **295**(3－5), pp.1085－1098.

O'Brien, E. Murray, T. and McDonald, A. (2003) Detecting leaks from water pipes at a test facility using ground penetrating radar, In: Proceedings of PEDS 2003 (Pumps, Electromechanical Devices and Systems Applied to Urban Water Management), Valencia, Spain.

Palau, C.V. Arregui, F. and Ferrer, A. (2004) 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**(3), pp. 169－181.

Pilcher, R. Hamilton, S. Chapman, H. Ristovski, B. and Strapely, S. (2007) Leak location and repair guidance notes, In: International Water Association, Water Loss Task Forces: Specialist Group Efficient Operation and Management, Bucharest, Romania.

Stathis, J.A. and Loganathan, G.V. (1999) Analysis of pressure－dependent leakage in water distribution systems, Analysis of pressure－dependent leakage in water distribution systems, In: Preparing for the 21st Century, 29th Annual Water Resources Planning and Management Conference, Tempe, AZ, USA.

Tabesh, M. and Delavar, M.R. (2003) Application of integrated GIS and hydraulic models for unaccounted for water studies in water distribution systems, In: Proceedings of the International Conference on Advances in Water Supply Management. London: Balkema.

Tajima, M. and Mita, A. (2009) Automatic Leakage Detection for Water Supply Systems Using Principal Component Analysis, MATERIALS FORUM －PARKVILLE－ CD ROM EDITION－, **33**, pp. 10.

Xia, L. and Guo－jin, L. (2010) 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).

Ye, G and Fenner, Ra. (2011) Kalman fifilter－ing of hydraulic measurements for burst detection in water distribution systems, *Journal of Pipeline Systems Engineering and Practice (ASCE)*, **2**. pp. 14－22. ISSN 1949－1190.