상수도 시스템 내에서의 실시간 수질감시를 위한 LPCF 모형의 적용 및 검증

Application of the LPCF Model for Real-Time Contamination Warning in the Water Supply System

박창경 · 윤석민 · 윤상진 · 김성수* · 박노석[†]

Changgyeong Park · Sukmin Yoon · Sangjin Yun · Seong-Su Kim* · No-Suk Park[†]

경상대학교 건축도시토목공학부 및 공학연구원 · *한국수자원공사

Division of Architectural, Urban and Civil Engineering and

Engineering Research Institute, Gyeongsang National University · *K-water

(Received September 17, 2018; Revised October 15, 2018; Accepted October 19, 2018)

Abstract : Currently, various water quality parameters are monitored for real-time contamination warning in the water supply system of South Korea. If the measured values of water quality parameters exceed the threshold, contamination warnings are issued. However, the U.S. Environmental Protection Agency (EPA) reported the following limitations of the contamination warning system based on these thresholds. First, irregular and sudden hydraulic changes in water supply system caused by pump or valve malfunction may cause measurement error of the water quality sensors, which may cause nuisance and unnecessary false-positive alarms. Second, in the case of long-term outflow of micropollutants, water quality changes are slightly within the thresholds, which causes a serious monitoring error of false-negatives that cannot be detected even in actual contamination. Therefore, the U.S. EPA applied a linear prediction - correction filter (LPCF) model for real-time contamination warning, which is based on the autoregressive (AR) model. The main purpose of this study is to develop a contamination warning technique to be applied to water supply system of South Korea. For the development of a real-time contamination warning technique, the LPCF model was applied and Testing with water quality data collected from G_water supply system of South Korea.

Key Words : Water Supply System, Real-Time Contamination Warning, Autoregressive Model, Linear Prediction-Correction Filter Model

요약 : 현재 국내의 상수도 시스템 내에서는 수질감시등의 목적으로 다양한 수질인자들이 실시간으로 관측되고 있으며, 실시간 관측되는 수질인자들의 변화가 특정 한계값(Threshold)을 초과하는 경우 오염경보를 발령하는 ‘Set Point’방법을 적용하고 있다. 하지만 미국 EPA에서는 ‘Set Point’방법을 활용한 수질오염감시는 다음과 같은 심각한 허위경보(False alarm)를 발생시킬 수 있다고 보고하였다. 첫째, 한계값에 의존한 ‘Set Point’방법은 상수도 시스템 내의 불규칙적인 수리조건에 의한 급격한 수질변화 및 계측기의 측정오류로 인해 발생하는 빈번한 이상값(Outlier)들의 영향으로 인해 과도하면서도 성가신 위양성(False-Positive)의 거짓경보를 발생시킨다. 둘째, ‘Set Point’방법은 미소오염물질의 장기간 유출과 같이 설정된 한계값 이내에서 발생하는 미세한 수질오염에 대해서는 실제적으로 대응할 수 없는 위음성(False-Negative)의 심각한 감시오류를 발생시키게 되며, 결국 이러한 허위경보들은 수질오염감시의 운영 효율과 신뢰도를 저하시키는 주요원인이 된다고 지적하였다. 이에 본 연구에서는 국내 상수도 시스템에 적용 가능한 실시간 수질오염감시 기술을 개발하고자 미국 EPA의 LPCF 모형을 적용하였다. 그리고 LPCF 모형의 개발을 위해 국내 G_정수장에서 관측된 pH, 탁도, 잔류염소, ORP 및 전기전도도 자료를 수집하였으며, 모의 수질오염사고를 이용하여 LPCF 모형의 적용성을 검증하였다.

주제어 : 상수도시스템, 수질오염감시, AR(p)모형, LPCF모형

1. 서론

현대의 상수도 시스템은 양질의 수돗물을 연속적으로 공급하기 위한 공공시설로서 각 구성요소(정·배수지, 펌프장 등)들은 상호 관로(Pipe)를 통해 연결된다. 따라서 관의 연결성이 상대적으로 취약한 접합부 또는 균열부 등을 통해 외부 오염원이 유입되거나, 관내 부식으로 인한 내부 오염이 발생하는 경우 전체 상수도 시스템의 안정성 및 수질관리에 부정적인 영향을 미치게 된다. 또한 수돗물 사용의 시간적인 변화, 소방용수의 사용, 관내 부속품(밸브, 펌프 등)의 급격한 작동과 같이 상수도 시스템 내에서 발생하는 다양한 수리조건의 변화는 수질오염감시를 더욱 어렵게 한다

고 할 수 있다.^{1~3)}

실제 지난 2015년 미국 미시간주 플린트 시에서는 취수원을 염화물의 농도가 상대적으로 높은 플린트 강으로 변경한 이후 노후화된 급수관의 부식이 촉진되어 수돗물에 납이 침출되는 수질오염사고가 발생되었다. 당시 수돗물에 대한 플린트 시 거주민들의 민원이 지속적으로 제기되었으나 플린트 시 당국의 소극적인 수질오염감시와 수질검사를 위한 샘플링 오류 등으로 인해 1년 이상 수돗물 공급을 중단시키지 않았다. 그 결과 해당지역 영·유아 8천여 명에게 납중독이 발생하는 등 심각한 후유증을 발생시켰다.

현재 국내 외 대부분의 상수도 시스템에서는 수질오염감시를 위해 실시간 관측되는 수질인자들(Ex; pH, Chlorine,

[†] Corresponding author E-mail: nspark@gnu.ac.kr Tel: 055-772-1798 Fax: 055-772-1799

etc)의 변화가 특정 한계값(Threshold)을 초과하는 경우 오염 정보를 발령하는 'Set Point' 방법을 적용하고 있다. 하지만 미국 EPA에서는 'Set Point' 방법을 활용한 수질오염감시는 다음과 같은 심각한 허위경보(False alarm)를 발생시킬 수 있다고 보고하였다.

첫째, 한계값에 의존한 'Set Point' 방법은 상수도 시스템 내의 불규칙적인 수리조건에 의한 급격한 수질변화 및 계측기의 측정오류로 인해 발생하는 빈번한 이상값(Outlier)들의 영향으로 인해 과도하면서도 성가신 위양성(False-Positive)의 거짓경보를 발생시킨다. 둘째, 'Set Point' 방법은 미소 오염물질의 장기간 유출과 같이 설정된 한계값 이내에서 발생하는 미세한 수질오염에 대해서는 실제적으로 대응할 수 없는 위음성(False-Negative)의 심각한 감시오류를 발생시키게 되며, 결국 이러한 허위경보들은 수질오염감시의 운영 효율과 신뢰도를 저하시키는 주요원인이 된다고 지적하였다^{4,5)}

이와 관련해 미국 EPA에서는 상수도 시스템 내에서 발생 가능한 모든 수질오염사고들을 조기에 감지하고, 수질오염 발생 시 신속한 정보와 처리를 수행할 수 있는 총체적인 대응으로서 'Contamination Warning System (CWS)'의 개념을 수립하였다. 그리고 미국 EPA에서는 신뢰도 높은 CWS의 구축을 위해 온라인 수질 모니터링 기반의 'Event Detection System (EDS)'의 성공적인 개발을 필수요소로서 언급하였으며, 2007년부터 2014년까지 미국 신시내티 내 고립 배·급수 관망 6곳을 대상으로 'EDS Challenge'라는 기술 개발 프로젝트를 진행한바 있다.

여기서, 미국 EPA에서 수립한 EDS의 기본 알고리즘에 관해 간략히 살펴보면 상수도 시스템 내에서 관측되는 수질인자들의 변화를 실시간으로 예측할 수 있는 수질상태추정 모형(Water qualities estimation model)을 개발하고, 모형으로부터 구해진 예측값과 실제 관측값과의 차이를 잔차(Residual)로서 정의하게 된다. 그리고 예측잔차들에 대한 확률적론 분석을 기반으로 'Set Point' 기반의 수질오염감시기술의 단점으로 지적된 허위경보들을 최소화한 분석 결과를 도출하게 된다. 그리고 EDS에 적용된 수질상태추정 모형으로서는 'Time Series Increments (TSI) Model', 'Linear Prediction-Correction Filter (LPCF) Model', 'Multivariate Nearest Neighbor (MVNN) Approach'이 적용되며, LPCF 모형과 MVNN Approach가 TSI 모형에 비해 보다 효과적인 것으로 알려져 있다. 그리고 MVNN Approach는 과거 수질상태와 현재 수질상태와의 유사도(Similarity)만 반환할 뿐 예측값은 제공하지 못하는 반면 LPCF 모형은 통계적 시계열 모형을 기반으로 가까운 미래시점에 대한 예측값을 제공한다. 즉, LPCF 모형은 통계적 시계열 모형인 Autoregressive(AR) 모형의 생성과 갱신과정을 반복함으로써 상수도 시스템내의 실시간의 수질상태 변화를 예측하게 된다.⁶⁻¹¹⁾

앞서 언급한 미국 플린트시의 수질오염사고 사례에서와 같이 수도물 오염사고에 따른 용수공급의 중단은 막대한 사회·경제적 피해를 발생시킬 뿐만 아니라 수도물에 대한 신

뢰도와 음용률에도 부정적인 영향을 미치게 된다. 이에 본 연구에서는 국내 상수도 시스템에 적용 가능한 실시간 수질오염감시 기술을 개발하고자 미국 EPA의 LPCF 모형을 적용하였다. 그리고 LPCF 모형의 개발을 위해 국내 G_정수장에서 관측된 pH, 탁도, 잔류염소, ORP 및 전기전도도 자료를 수집하였으며, 모의수질오염사고를 이용하여 LPCF 모형의 적용성을 검증하였다.

2. 이론적 배경

LPCF 모형은 과거 수질 관측값들의 선형결합(Linear combination)을 통해 현재의 수질상태를 실시간으로 예측하는 방법론으로서 Autoregressive, AR(p)모형을 기반으로 한다. 임의의 시간 t 에서의 평균이 μ 인 관측값 Z_t 에 대한 정상 AR(p)과정은 식 (1)과 같이 정의된다.^{12,13)}

$$Z_t - \mu = \Phi_1(Z_{t-1} - \mu) + \Phi_2(Z_{t-2} - \mu) + \dots + \Phi_p(Z_{t-p} - \mu) + \epsilon_t \\ = \sum_{j=1}^p \Phi_j(Z_{t-j} - \mu) + \epsilon_t \quad (1)$$

여기서, ϵ_t 는 확률오차(Random error)로서 관측시점에 관계없이 독립항등분포(Independently and identically distributed, I.I.D.)를 따르는 백색잡음과정(White noise process)이다. 그리고 Φ 는 AR(p) 모형의 추정계수로서 Yule-Walker 방정식을 이용해 각 시차 p 에 대응되는 자기상관계수(Autocorrelation coefficient) ρ_k 와 관계를 식 (2)와 같이 정의할 수 있다.

$$\rho_k = \Phi_1\rho_{k-1} + \dots + \Phi_p\rho_{k-p}, k \geq 1 \quad (2)$$

식 (2)에 정의된 Yule-Walker 방정식을 행렬의 형태로 나타내면 식 (3)과 같으며,

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \rho_1 & \rho_1 & 1 & \dots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{bmatrix} \quad (3)$$

식 (3)을 식 (4)와 같이 정리하면

$$\rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{bmatrix}, P = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \rho_1 & \rho_1 & 1 & \dots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix}, \Phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{bmatrix} \quad (4)$$

추정계수 Φ_i 는 식 (5)와 같이 나타낼 수 있다.

$$\Phi = P^{-1}\rho_p \quad (5)$$

Fig. 1. Concept of LPCF model and analysis window.

따라서 식 (5)에서 ρ_p 의 추정량인 표본자기상관계수(Sample autocorrelation coefficient) $\hat{\rho}_p$ 를 대입하면 ϕ_i 들의 Yule-Walker 추정량을 구할 수 있다.

LPCF 모형은 Fig. 1에서 나타난 것과 같이 수질 관측주기와 동일하게 AR(p)모형을 연속적으로 생성·갱신하며 수질변화를 예측하게 된다. 따라서 LPCF 모형을 개발하기 위해서는 첫째, 분석윈도우(Analysis Window)의 적정크기 둘째, 최적 AR(p)모형의 식별과정이 필요하다. 여기서 분석윈도우란 AR(p) 모형의 적정차수 및 모수를 추정하기 위한 시계열 구간 또는 표본공간으로 정의할 수 있으며, Fig. 1에서 나타난 것과 같이 관측주기에 따라 연속적으로 이동되어야 한다. 그리고 분석윈도우의 크기는 수질인자들의 특성과 예측의 정확도 등에 따라 가변적이거나 일반적으로 크기가 증가하는 경우 모수추정은 강건(Robust)해지는 특성이 있다. 반면 분석윈도우가 지나치게 크게 되면 모수추정과 모형의 구현에 있어서 과도한 연산시간이 요구될 수 있으므로 주의해야 한다. 그리고 LPCF 모형에 의해 감시단계가 진행되는 동안 관측값이 '0'으로 기록되는 결측값 및 설정된 예측잔차 한계를 초과하는 이상값이 발생 되는 경우 AR(p) 모형의 모수추정에 부정적인 영향을 미치게 된다. 따라서 LPCF 모형은 결측값과 이상값의 출현 빈도가 설정된 유의 확률을 초과하지 않는 경우 예측값으로 대체되며 이러한 의미에서 Linear Prediction-Correction Filter Model이라 명명된다.

3. 대상지점 및 연구방법

3.1. 대상지점

본 연구에서는 LPCF 모형을 개발하고 성능을 검증하기 위해 2017년 7월 1일~2017년 8월 31일 기간 관측된 수질 데이터를 국내 G_정수장으로부터 수집하였다. 수집된 수질항목은 pH, 탁도, 잔류염소, ORP 및 전기전도도이며 각 항목들의 1분주기로 관측되고 있다.

국내 G_정수장으로부터 수집된 수질 데이터 중 2017년 7월 27일~2017년 8월 14일 기간의 데이터들은 LPCF모형의 개발을 위한 학습 데이터(Training Data), 2017년 8월 14

Table 1. Statistical characteristic of sampling data from G_WTP

Water quality	Training data		Test data	
	Avg.	Std.	Avg.	Std.
pH	6.96	0.13	7.01	0.08
Turbidity (NTU)	0.05	0.00	0.05	0.00
Residual Chlorine (mg/L)	0.52	0.03	0.53	0.02
ORP (mV)	836.20	8.93	819.45	3.48
EC (μS/cm)	151.83	10.11	128.49	15.01

일~2017년 8월 31일 기간의 데이터들은 LPCF 모형의 성능을 검증하기 위한 테스트 데이터(Test Data)로서 적용되었으며, 각 데이터들의 통계적 특성은 Table 1에 정리하였다.

3.2. 연구방법

LPCF 모형은 짧은 수질관측주기(본 연구에서는 1분) 동안 수질 데이터의 수집, AR(p) 모형 매개변수 추정/예측, 잔차계산이라는 복잡한 연산과정을 수행해야 한다. 따라서 주어진 관측주기를 만족하면서 모수추정의 강건성을 확보할 수 있는 최적 분석윈도우의 크기를 추정할 필요가 있어서 본 연구에서는 학습 데이터를 활용하여 분석윈도우의 크기를 60분~1440분으로 조정한 후 적정 분석윈도우의 크기를 분석하였다.

그리고 본 연구에서는 LPCF 모형의 성능 평가를 위한 모의 수질사고를 개발하고자 미국 EPA의 선행연구 결과를 참고하였다. 미국 EPA에서는 원시 수질 데이터 내에서의 모의 수질사고를 발생시키기 위해 식 (6)을 제안하였다.

$$Z_E(t) = Z_0(t) + E_{Ind}(t) \cdot e \cdot E_{max} \cdot \sigma_Z \quad (6)$$

여기서, $Z_E(t)$ 는 수질사고, $Z_0(t)$ 는 원시 수질 데이터, $E_{Ind}(t)$ 는 수질사고 Indicator로서 정상 상태인 경우 '0', 수질사고인 경우에는 '1'이 적용된다. 그리고 e 는 추세변수로서 상승기에는 '1', 하강기에는 '-1'이 적용된다. E_{max} 는 σ_Z 의 강도변수이며 σ_Z 는 원시 수질 데이터들의 표준편차이다.

본 연구에서는 식 (6)을 이용해 모의 수질사고의 형태를 Table 2에 나타난 것과 같이 서서히 수질 관측값이 증가 또

Table 2. Description of artificial event

Event type	Description
I : Trend event
II : Peak event
III : Pulse event
IV : Wave event

는 감소하는 추세형(Type I, Trend), 수질 관측값이 상승 후 하강 또는 하강 후 상승하는 첨두형(Type II, Peak), 수질 관측값이 급격히 상승 또는 하강 후 유지되는 맥동형(Type III, Pulse) 그리고 수질 관측값이 상승과 하강 또는 하강과 상승을 반복하는 파동형(Type IV, Wave)로 구성하였다.

그리고 모의 수질사고의 강도 E_{max} 는 0.5~3.0의 범위에서 적용하였으며 E_{max} 의 발생시점은 Table 2에서 나타난 것과 같이 각 Type별 모의사고에서 빨간색 점으로 표시하였다. 예를 들어 Type I의 모의 수질사고에서 E_{max} 를 0.5로 설정하는 경우 E_{max} 를 관측시간의 진행과 함께 단계적으로 상승시켜 최종시점에서는 $E_{max} = 0.5$ 가 되도록 하였다. 아울러 본 연구에서 적용된 LPCF 모형 및 모의사고들은 공학용 소프트웨어인 Matlab (R2012a)을 활용하여 개발되었다.

4. LPCF모형의 개발 및 검증

4.1. LPCF모형의 개발

본 연구에서는 LPCF 모형을 개발하기 위해 국내 G_정수장으로부터 수집된 2017년 7월 27일 ~ 2017년 8월 14일 기간의 데이터(1분 간격, 10,800set)를 활용하였다.

Fig. 2는 수집된 학습 데이터들의 시계열분포를 도시한 것으로서 5개의 수질항목들의 시계열변화에서는 결측값이나 이상값은 분포하지 않는 등 안정적인 시계열 분포를 나타내는 것을 알 수 있다.

LPCF 모형을 개발하기 위해서는 적정 AR(p) 모형의 식별이 요구된다. 일반적으로 AR(p) 모형의 식별은 차수 'p'의 선택이며, 이 단계에서 가장 유의해야 할 사항은 'p'를 크게 잡으면 추정해야 할 모수의 개수 또한 'p'개로 증가하므로 비효율적이라는 사실이다. 실제로 추정할 모수의 개수가 증가하면 최종 예측모형이 복잡해질 뿐만 아니라 추정의 효율성도 떨어지게 되므로 가능한 간단한 모형을 선택하는 것이 통례이며 이처럼 간단한 모형을 선택하는 것을 간결의 원칙(Principal Of Parsimony)이라 한다. 한편, LPCF 모형의 개발과 관련해 국내 C_정수장을 대상으로한 Park¹⁾의 선행연구에서는 pH, 탁도 및 잔류염소를 대상으로 LPCF 모형의 적정 AR(p)모형을 식별한바 있으며, AR(p)모형의

Fig. 2. Time series distribution of training data.

차수에 따른 LPCF 모형의 예측오차의 변동은 미소하다고 밝힌바 있다. 따라서 본 연구에서는 수집된 수질자료의 시계열분포 및 선행연구결과를 참고하여 AR(1) 모형을 LPCF 모형의 기본 시계열 모형으로 채택하였다.

LPCF 모형의 개발을 위해서는 AR(p) 모형의 식별과 더불어 적정 분석윈도우의 크기를 설정해야 한다. 따라서 주어진 관측주기를 만족하면서 모수추정의 강건성을 확보할 수 있는 최적 분석윈도우의 크기를 추정하기 위해 본 연구에서는 분석윈도우의 크기를 60분~1440분으로 조정하여 적정 크기를 분석하였다.

Fig. 3(a)은 학습데이터 중 pH를 대상으로 분석윈도우의 크기를 180분으로 설정한 후 LPCF 모형을 구동하여 수질변화를 예측한 결과를 나타낸 것이며, Fig. 3(b)는 예측잔차의 분포를 나타낸 것이다. 본 연구에서는 적정 분석 윈도우의 크기를 결정하기 위해 분석윈도우 변화에 따른 예측잔차들의 MAE(Mean Absolute Prediction Error)를 산정하였다.

Table 3은 분석윈도우 변화에 따른 학습 데이터들의 수질항목별 MAE 산정 결과를 나타낸 것으로 pH, 탁도, 잔류염소의 경우 MAE는 분석윈도우의 크기 변화와 무관하게 일정한 값이 반환되었다. 그리고 ORP와 전기전도도의 경우 MAE는 분석윈도우의 크기가 증가함에 따라 감소하는 경향

을 나타내었으나 변화범위는 미소하게 나타났다.

미국 EPA의 선행연구에서는 분석윈도우의 크기가 증가하게 되면 모수추정이 강건하게 되어 예측오차의 크기와 변동범위가 감소하는 것으로 보고되었으나 이는 수질변화가 급격히 발생하는 관망을 대상으로 한 분석 결과이다. 본 연구는 미국 EPA와는 상이하게 수질항목들의 변동범위가 낮은 정수장을 대상으로 하였기 때문에 분석윈도우의 크기가 모수추정의 강건성에 미치는 영향은 미소한 것으로 판단된다. 따라서 1분이라는 상대적으로 짧은 허용 연산시간을 고려하는 경우 분석윈도우의 크기를 최소 60분정도로 설정하는 것이 가장 효율적일 수 있으나, 수도물의 소비 패턴에 따른 일사용량변화를 반영하기 위해서는 최소 360분(6시간) 이상의 분석윈도우가 필요할 것으로 판단된다.

4.2. LPCF모형의 검증

본 연구에서는 LPCF 모형의 검증을 위해 국내 G_정수장으로부터 수집된 2017년 8월 14일 ~ 2017년 8월 31일 기간의 데이터(1분 간격, 10,800set)를 활용하였다. 그리고 본 연구에서 개발한 LPCF모형의 적정성을 검증하기 위해 AR(1), 분석윈도우 360분을 적용한 후 예측잔차를 분석하였고 그 결과를 학습데이터의 결과와 비교하였다.

Table 4는 학습 데이터와 테스트 데이터에 대한 MAE산

Table 3. MAE calculation results of training data

Water quality	Analysis window (min)					
	60	180	360	720	1080	1440
pH	0.0023	0.0023	0.0022	0.0022	0.0022	0.0022
Turbidity (NTU)	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Residual chlorine (mg/L)	0.0008	0.0008	0.0007	0.0007	0.0007	0.0006
ORP (mV)	0.3210	0.3140	0.3020	0.3030	0.3010	0.2980
EC (μ S/cm)	0.0320	0.0320	0.0310	0.0310	0.0310	0.0310

Table 4. Comparison of MAE of training data and test data

Water quality	MAE	
	Training data	Test data
pH	0.0022	0.0024
Turbidity (NTU)	0.0001	0.0001
Residual chlorine (mg/L)	0.0007	0.0008
ORP (mV)	0.3020	0.3321
EC (μ S/cm)	0.0310	0.0330

Fig. 3. Forecasting result of pH (Training Data, AR (1), Analysis Window = 180 min).

Fig. 4. Description of artificial event in test data (Test Data, pH).

정 결과를 비교한 것으로, ORP의 MAE가 학습 데이터에 비해 다소 높기는 하나 전반적으로 테스트 데이터들의 예측 오차들은 학습 데이터들의 예측오차와 유사한 범위에서 분포하는 것으로 나타났다.

LPCF 모형의 수질감시 성능을 검증하기 위해 Fig. 4에서 나타낸 것과 같이 테스트 데이터 내에 4가지 Type의 모의 수질사고들을 임의 분포한 후, 수질사고의 최대강도인 $E_{\max}$ 의 범위를 0.5~3.0 (6 Case)로 조정하였다.

LPCF 모형을 이용해 수질감시를 수행하기 위해서는 사고 식별을 위한 확률공간 및 예측잔차의 허용한계를 설정해야 한다. 본 연구에서는 사고식별을 위한 확률공간을 '의사결정 윈도우(Discrimination Window)'로 정의하였으며, 미국 EPA 선행연구를 참고하여 확률모형으로 이항확률분포(Binomial Distribution)를 적용하였다. 그리고 예측잔차의 허용한계는 유의수준을 5~10% 수준에서 고려해 식 (7)과 같이 설정하였다.

$$\begin{cases} th_{\bar{\epsilon}} = \bar{\epsilon} + 1.65\sigma_{\bar{\epsilon}}, & \alpha = 0.1 \\ th_{\bar{\epsilon}} = \bar{\epsilon} + 1.96\sigma_{\bar{\epsilon}}, & \alpha = 0.05 \end{cases} \quad (7)$$

여기서, $th_{\bar{\epsilon}}$ 는 예측잔차의 허용한계, $\bar{\epsilon}$ 는 예측잔차 그리고 $\sigma_{\bar{\epsilon}}$ 는 예측잔차의 표준편차이다.

따라서 설정된 의사결정윈도우 내에서 예측잔차가 허용한계이내에 분포하는 경우 '0', 한계를 초과하는 경우 '1'이 반환하며, 발생빈도가 설정된 이항분포의 유의수준을 초과하게 되면 수질사고로 식별하게 된다. 본 연구에서는 의사결정윈도우의 크기를 5분(Trial 5회)으로 설정하였으며, 사고 식별을 위한 이항확률분포의 유의수준은 5%를 적용하였다.

Fig. 5는 5개 수질항목들의 예측잔차의 허용한계 및 $E_{\max}$ 의 변화에 따른 평균 사고식별 비율을 나타낸 것이다. 예측잔차의 허용한계 변화에 따른 평균 사고식별 비율을 살펴보면 $\alpha = 0.10$ 에서는 81%, $\alpha = 0.05$ 에서는 79%로 나타났다. 그리

Fig. 5. Event detection results according to residual threshold.

Fig. 6. Event detection results according to event type.

고 $E_{\max}$ 의 변화에 따른 식별 성능을 살펴보면 사고식별 비율은 55%~90%로서 사고식별 비율이 55%인 $E_{\max} = 0.5$ 제외하면 모두 80%이상의 높은 사고식별 성능을 나타내었다.

Fig. 6는 5개 수질항목들의 모의사고 형태에 따른 평균 사고식별 비율을 나타낸 것이다. $\alpha = 0.10$ 에서는 Type I 80%, Type II 87%, Type III 57% 그리고 Type IV 100% 그리고 $\alpha = 0.05$ 에서는 Type I 80%, Type II 80%, Type III 57% 그리고 Type IV 100%로서 맥동형 모의사고에 대한 식별 성능은 상대적으로 낮은 반면 파동형 모의사고에 대한 식별 성능은 매우 우수한 것으로 나타났다.

5. 결론

본 연구의 주요목적은 국내 상수도 시스템 내에서 일반적으로 관측되고 있는 다항목의 수질데이터를 활용하여 실시간 수질오염감시 기술을 개발하는데 있다. 실시간 수질오염감시 기술과 관련해 선도적 연구를 수행하고 있는 미국 EPA에의 선행연구를 참고하여 본 연구에서는 AR(p)모형 기반의 LPCF모형을 개발하였고, 국내 G_정수장을 대상으로 검증을 수행하였다. 그 결과들을 요약하면 다음과 같다.

1) 실시간 수질감시를 위해 LPCF모형을 적용하는 경우 적정 AR(p)모형 및 분석윈도의 크기를 설정해야하며, 미국 EPA는 분석윈도우의 크기가 증가하게 되면 보수추정이 강건하여 예측오차가 감소한다고 밝힌바 있다. 본 연구에서는 국내 선행연구를 참고하여 AR(1)모형을 기반으로 LPCF모형을 개발하고 분석윈도우 변화에 따른 예측오차를 분석한 결과 예측오차들의 변동 범위는 미소한 것으로 나타났다. 이는 미국 EPA의 경우 수질변동이 상대적으로 높은 관망을 대상으로 한 반면 본 연구는 상대적으로 수질변동이 낮은 정수장을 대상으로 한 것이 주요원인으로 판단된다.

2) LPCF 모형의 수질감시 성능을 검증하기 위해 4가지 형태의 모의 수질사고를 개발하였고, 사고식별을 위한 확률

분포모형으로서 이항확률분포를 적용하였다. 모의 수질사고의 최대강도($E_{\max}$)를 6 Case로 변화하여 예측잔차의 허용한계에 따른 평균 사고식별 비율을 분석한 결과 허용한계 $\alpha = 0.10$ 에서는 81%, $\alpha = 0.05$ 에서는 79%로 나타났다. 그리고 모의 수질사고의 최대 강도 변화에 따른 성능을 분석한 결과 사고식별 비율의 범위는 55%~90%로서, 사고식별 비율이 55%인 $E_{\max} = 0.5$ 제외하면 모두 80%이상의 높은 사고식별 성능을 나타내었다.

3) 모의사고 형태에 따른 평균 사고식별 비율을 분석한 결과 예측잔차의 허용한계 $\alpha = 0.10$ 에서는 Type I 80%, Type II 87%, Type III 57% 그리고 Type IV 100% 그리고 $\alpha = 0.05$ 에서는 Type I 80%, Type II 80%, Type III 57% 그리고 Type IV 100%로서 맥동형 모의사고에 대한 식별 성능은 상대적으로 낮은 반면 파동형 모의사고에 대한 식별 성능은 매우 우수한 것으로 나타났다.

맥동형 모의사고에 대한 식별성능이 상대적으로 낮게 나타나는 원인으로는 수질상태추정 모형으로 적용된 LPCF 모형이 통계 시계열 모형인 AR(p)모형에 기초한다는 점을 고려할 수 있다. 즉, AR(p)모형은 과거 시계열자료의 패턴이 미래에도 지속적으로 유지된다는 가정 하에서 현재까지 수집된 자료를 분석하여 미래에 대한 예측을 하는 방법론이다. 따라서 LPCF 모형은 설정된 과거 시차의 시계열 패턴을 연속적으로 반환하기 때문에 동적인 시계열 변화가 발생되지 않는 맥동형 모의사고의 경우 초기의 맥동충격이 설정된 예측잔차의 허용한계 내에서 분포하는 경우 일정한 예측값만이 반환됨으로서 수질사고에 대한 식별 성능 또한 감소하는 것으로 판단된다.

Acknowledgement

본 연구는 환경부 물관리연구사업의 연구비 지원(과제번호: 16AWMP-B113766-01)에 의해 수행되었습니다.

KSEE

References

1. Park, N. S., Lee, Y., Chae, S. and Yoon, S., "A Study on the Statistical Predictability of Drinking Water Qualities for Contamination Warning System," *J. Korean Soc. Water and Wastewater*, **29**(4), 469~479(2015).
2. Yoon, S., Park, N. S., Ahn J., Song Y., Shin, C. S. and Kim, S. H., "Feasibility study for the introduction of contamination warning system for establishing smart water grid," *J. Korean Soc. Water and Wastewater*, **29**(2), 271~281(2015).
3. Park, N. S., Kim, S., Lee, Y. J. and Yoon, S., "A Study on the Turbidity Estimation Model Using Data Mining Techniques in the Water Supply System," *J. Korean Soc. Environ. Eng.*, **38**(2), 87~95(2016).
4. U.S. EPA, "Water Quality Event Detection Systems for Drinking Water Contamination Warning Systems," National Homeland Security Research Center(2010).
5. U.S. EPA, "Water Security Initiative: Guidance for building laboratory capabilities to response to drinking water contamination," Office of water(2013).
6. U.S. EPAa, "Water Security Initiative: System Evaluation of the Cincinnati Contamination Warning System Pilot," U.S EPA Water Security Division(2014).
7. U.S. EPAb, "Water Security Initiative: Evaluation of the Water Quality Monitoring Component of the Cincinnati Contamination Warning System Pilot," U.S EPA Water Security Division(2014).
8. U.S. EPAc, "Water Security Initiative: Evaluation of the Enhanced Security Monitoring Component of the Cincinnati Contamination Warning System Pilot," U.S EPA Water Security Division(2014).
9. U.S. EPAd, "Water Security Initiative: Evaluation of the Customer Complaint Surveillance Component of the Cincinnati Contamination Warning System Pilot," U.S EPA Water Security Division(2014).
10. U.S. EPAe, "Water Security Initiative: Evaluation of the Public Health Surveillance Component of the Cincinnati Contamination Warning System Pilot," U.S EPA Water Security Division(2014).
11. U.S. EPAf, "Water Security Initiative: Evaluation of the Consequence Management Component of the Cincinnati Contamination Warning System Pilot," U.S EPA Water Security Division(2014).
12. Wei, W. W. S., "Time Series Analysis : Univariate and Multivariate Methods," Addison-Wesley Pub.(1990).
13. Brockwell, P. J. and Davis, R. A., "Introduction to Time Series and Forecasting," 2nd Edition, Taylor & Francis(2002).