Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2021.26685

212 (2021) 8–16
February

Application of LPCF model based on ARIMA model to prediction of water quality change in water supply system

No-Suk Park[a], Seong-Su Kim[b], Inseok Seo[b], Sukmin Yoon[a,*]

[a]*Department of Civil Engineering and Engineering Research Institute, Gyeongsang National University, 501, Jinju-Daero, Jinju 52828, Republic of Korea, Tel. +82 55 772 1798; Fax: +82 772 1799; emails: gnuysm@gmail.com (S. Yoon), nspark@gnu.ac.kr (N.-S. Park)*
[b]*Korea Water Resources Corporation, K-water Institute, 125, Yuseong-daero 1689 beon-gil, Yuseong-gu, Daejeon 34045, Republic of Korea, emails: kssman@kwater.or.kr (S.-S. Kim), seois@ kwater.or.kr (I. Seo)*

Received 18 May 2020; Accepted 2 October 2020

ABSTRACT

Currently, various water quality parameters (WQPs) are monitored for real-time contamination warning (CW) in the water supply system (WSS) of South Korea. If the measured values of WQP exceed the threshold value, CWs are issued. However, the U.S. Environmental Protection Agency (EPA) reported the following limitations of the CW system based on these thresholds. First, irregular and sudden hydraulic changes in WSS caused by pump or valve malfunction may cause measurement error of the WQP sensors, which may cause nuisance and unnecessary false-positive alarms. Second, in the case of long-term outflow of micropollutants, WQPs change is 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 CW, which is based on the autoregressive (AR) model. The main purpose of this study is to develop a CW technique to be applied to WSSs in South Korea. For the development of a real-time CW technique, the LPCF model was applied with reference to previous research of the U.S. EPA. However, the time series of the WQP observed in WSSs often does not satisfy stationarity even though they are important fundamental assumptions of the AR model. Therefore, in this study, we developed an LPCF model by applying the autoregressive integrated moving average model considering nonstationary WQPs.

Keywords: Contamination warning; Linear prediction–correction filter model; Autoregressive model; Autoregressive integrated moving average model

1. Introduction

The water supply system (WSS) is a public facility that continuously supplies tap water of good quality whereby each component, such a reservoir or water treatment plant (WTP), is connected through a pipe. Therefore, if external contaminants are introduced through joints or cracks where the pipe integrity is relatively weak or if internal contamination occurs owing to corrosion in the pipe, the stability of the entire WSS and water quality management will be negatively affected [1,2]. In addition, changes in various hydraulic conditions owing to temporal changes in the use of tap water, the use of fire-fighting water, and the sudden operation of fittings (e.g., valves, and pumps) in the pipe network make real-time contamination warning (CW) in WSSs more difficult.

Currently, most of the WSSs in South Korea use the set-point method, in which CWs are issued when changes in water quality parameters (WQPs) observed in real-time, such as pH and turbidity, exceed the threshold value [3,4].

* Corresponding author.

However, the U.S. Environmental Protection Agency (EPA) reported that a false alarm could be raised when using the set-point method. First, irregular and sudden hydraulic changes in WSSs caused by pump or valve malfunction may cause a 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, the WQPs changes are slightly within the thresholds values, which causes a serious monitoring error of false-negative that cannot be detected even in actual contamination. As a result, these false alarms are a major cause of lowered operational efficiency and reliability of CW in WSSs [5,6].

In this regard, the U.S. EPA has developed a CW system that can detect all possible contamination events, including intentional water pollution accidents by terrorist acts, at an early stage in the entire WSS. In order to build a reliable CW system, the U.S. EPA referred to the successful development of the event detection system (EDS) based on on-line WQPs monitoring as an essential element and conducted an EDS challenge in 2007–2014 for six isolated WSSs in Cincinnati, Ohio [7–13].

Here, the basic algorithm of the EDS is briefly described. The EDS predicts the WQPs states in real-time based on a statistical estimation model and defines the difference between the predicted value and the actual observation value as the residual. On the basis of the probabilistic analysis of the residuals, an analytical result is derived that minimizes the false alarms pointed out as a disadvantage of the set-point method. The time series increments (TSI) model, the linear prediction–correction filter (LPCF) model, and the multivariate nearest neighbor (MVNN) approach are applied to the EDS; the LPCF model and the MVNN approach are known to be more effective than the TSI model. The MVNN approach returns only similarities between the past and current values of WQP; it does not provide predictive values. The LPCF model provides estimates for the near future based on the statistical time series model. That is, it predicts real-time WQPs changes in the WSS by repeating the generation and updating process of the autoregressive (AR) model. The AR model assumes stationarity for the observed time series, which means that the mean and covariance of the observed data are not affected by changes in time. In reality, however, most of the WQP observed in a WSS often does not satisfy the assumptions of stationarity such as changes in mean and covariance over time in addition to trend and seasonality. In general, if the observed time series indicate nonstationarity, transformations are required such as variance stabilizing transformations or differencing transformations. In particular, a random trend such as a seasonal factor can be transformed into a stationary time series through continuous differencing. In such a case, the autoregressive integrated moving average (ARIMA) model, which is an extension of the AR model, is applied as a tentative estimation model for time series [14,15].

As demonstrated in the 2016 water contamination accident in Flint, Michigan, the USA owing to lead leakage, suspension of the WSS as a result of water pollution causes enormous social and economic damages and has a negative impact on the reliability and consumption rate of tap water. In this study, we analyze the precedent research of the EDS, which is being studied and developed by the EPA in order to develop real-time CW and rapid response capability in the WSS of South Korea. In order to develop the WQP estimation model, which is the essential technology of EDS, the LPCF model is applied to the G_WTP in South Korea, and its performance is verified. However, unlike previous studies of the EPA based on the AR model, the ARIMA model as a fundamental estimation model is applied here considering possible nonstationarity in the observed WQP.

2. Theoretical background

The LPCF model is a method that predicts the current value of WQP in real-time through a linear combination of the weighted sum of past values. It is a process in which the AR model is continuously generated and updated at the same time as the WQP observation period. The AR model of order p, which is denoted as AR (p), is:

$$\hat{z}_{t+1} - \mu = \varnothing_1\left(z_t - \mu\right) + \varnothing_2\left(z_{t-1} - \mu\right) + \ldots + \varnothing_p\left(z_{t-p+1} - \mu\right) + a_{t+1} \quad (1)$$

where z_t with mean μ is the value of WQP at the current time, $\hat{z}_{t+1}$ is the estimate of the value of WQP at the next time step, $\varnothing_p$ is the coefficient for the AR model, p is the order of the estimation filter polynomial (number of previous measurements), and a_{t+1} is the estimation error or residual.

The residual calculated by this estimate is:

$$a_{t+1} = z_t - \hat{z}_{t+1} \quad (2)$$

where a_{t+1} is referred to as a zero-mean Gaussian white noise process.

By introducing the backshift operator $B^j z_t = z_{t-j}$ and $Z_t = z_t - \mu$, Eq. (1) can be rewritten as Eq. (3) in the following compact form:

$$\left(1 - \varnothing_1 B - \varnothing_2 B^2 \ldots - \varnothing_p B^p\right) Z_{t+1} = a_{t+1} \quad (3)$$

For the real-time LPCF model, $\varnothing_p$ is updated at every time step and uses only the most recent p observations.

The time series of the WQP observed in WSSs often does not satisfy the stationarity even though it is an important fundamental assumption of the AR model. For example, the nonstationary time series of the WQP could have non-constant mean μ time-varying second moments such as nonconstant variance σ^2, or it has both of these properties. Although many time series are nonstationary, a homogeneous nonstationary series can be reduced to a stationary series by taking the appropriate difference of the general series [16]. Thus, the times series Z_t is nonstationary, but the dth differenced series, $w_t = (1 - B)^d Z_t$ for integers $d \geq 1$, is stationary.

Obviously, a stationary process resulting from differences in a homogeneous nonstationary series is not necessarily white noise. Generally, the difference in time series w_t follows the general stationary autoregressive moving average (ARMA) process, which is:

$$\varnothing_p\left(B\right)\left(1 - B\right)^d Z_t = \theta_o + \theta_q\left(B\right) a_t \quad (4)$$

where $\varnothing_p$ is the stationary AR operator, $\theta_q(B)$ is the invertible MA operator, and θ_q and θ_o are parameters of the MA model, respectively.

The resulting homogeneous nonstationarity model in Eq. (4) is referred to as the ARIMA model of order (p,d,q) and is denoted as the ARIMA (p,d,q) model.

3. Study area and procedure

The G_WTP in South Korea, with a capacity of 207,000 m^3/d, was selected for this study. The target water quality of G_WTP is pH 5.8–8.5, 0.1–4.0 mg/L free residual chlorine (F–Cl), and turbidity less than 0.5 NTU; each WQP has measured automatically in 1 min intervals. In this study, we used F–Cl and pH data from January 1 to December 31, 2013, observed at G_WTP for development and verification of the LPCF model.

Fig. 1 illustrates the monthly average and standard deviation of the WQP dataset observed at the G_WTP. The standard error of each observation value increased slightly during the summer season of June to September. The annual average value of the F–Cl was 0.59 mg/L (σ = 0.078 mg/L), and the annual average pH was 7.19 (σ = 0.56), which satisfies the target water quality of G_WTP. The LPCF model developed by the U.S. EPA is a process in which the AR model is continuously generated and updated during the observation period of the WQPs. Therefore, in order to develop the LPCF model, the size of the analysis window, and the model identification process of the appropriate AR (p) model needs to be first determined. Here, the analysis window can be defined as the sample space for estimating the parameters of the AR (p) model that moves continuously according to the observation period, as shown in Fig. 2.

The optimal AR (p) model is identified when determining the proper pth order. However, the ARIMA model, which is an extension of the AR model, is applied when nonstationarity such as trends or seasonal components appear in the time series of WQP. Therefore, identification of the optimal ARIMA (p,d,q) model can be summarized by determining the proper order of (p,d,q). In this study, to develop and test the LPCF model based on the ARIMA model, a simple random sampling of 20,160 sets of WQP data for 2 weeks was performed; the detailed characteristics of all data are summarized in Table 1.

A time series from April 1 to April 7, 2013 (10,080 sets), was used as training data to develop the LPCF model, and the optimal size of the analysis window and optimal ARIMA (p,d,q) model for the building the LPCF model were then derived. A time series from May 6 to May 12, 2013 (10,080 sets), was applied as test data to verify the performance of the LPCF model.

4. Results

4.1. Development of LPCF model

Fig. 3 presents the time series plot and the sample autocorrelation function (SACF) for Data No. C-1. As shown in Fig. 3a, Data No. C-1 did not indicate a deterministic trend, although it showed strong autocorrelation between adjacent

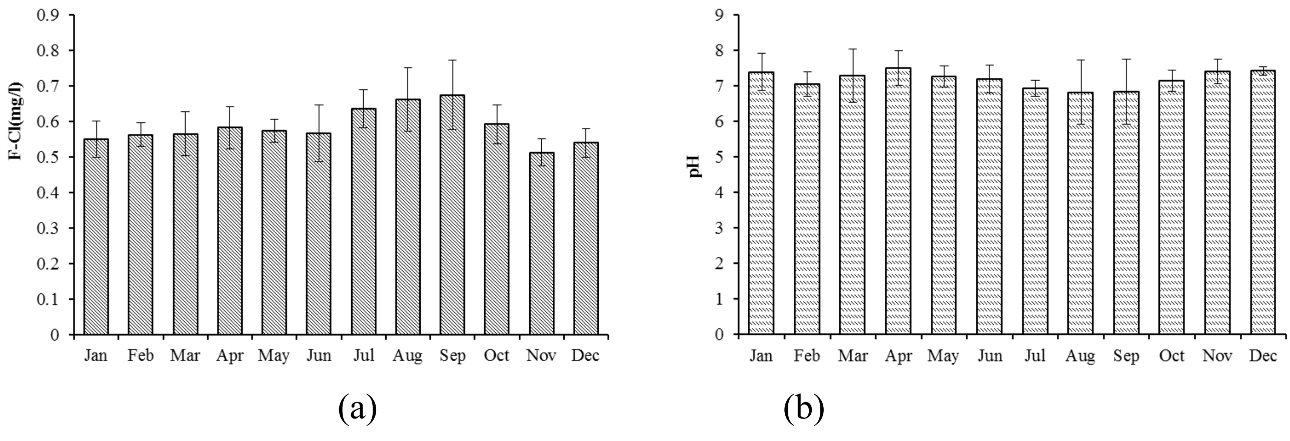

Fig. 1. Monthly changes in the WQP dataset: (a) F–Cl and (b) pH. The bar plot indicates the average, and the error bar indicates the standard error.

Table 1
Description of sampling data from G_WTP

Variable	Data No.	Data period	Mean	Std.
F–Cl	No. C-1	2013/04/01–2013/04/07 (Training)	0.596	0.056
	No. C-2	2013/05/06–2013/05/12 (Test)	0.575	0.03
pH	No. P-1	2013/04/01–2013/04/07 (Training)	7.727	0.478
	No. P-2	2013/05/06–2013/05/12 (Test)	7.199	0.151

Observation Time ; t		
Analysis Window	Observation	Estimate
$z_{t-p} \sim z_t$	z_t	$\hat{z}_{t+1}$

Observation time ; t+1		
Analysis Window	Observation	Estimate
$z_{t-p+1} \sim z_{t+1}$	z_{t+1}	$\hat{z}_{t+2}$

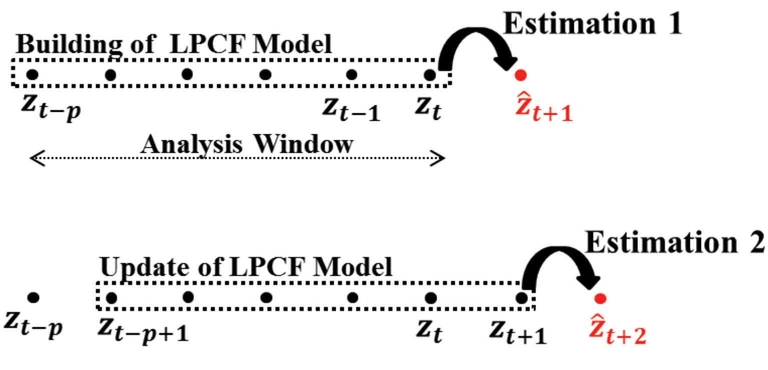

Fig. 2. Concept of the analysis window.

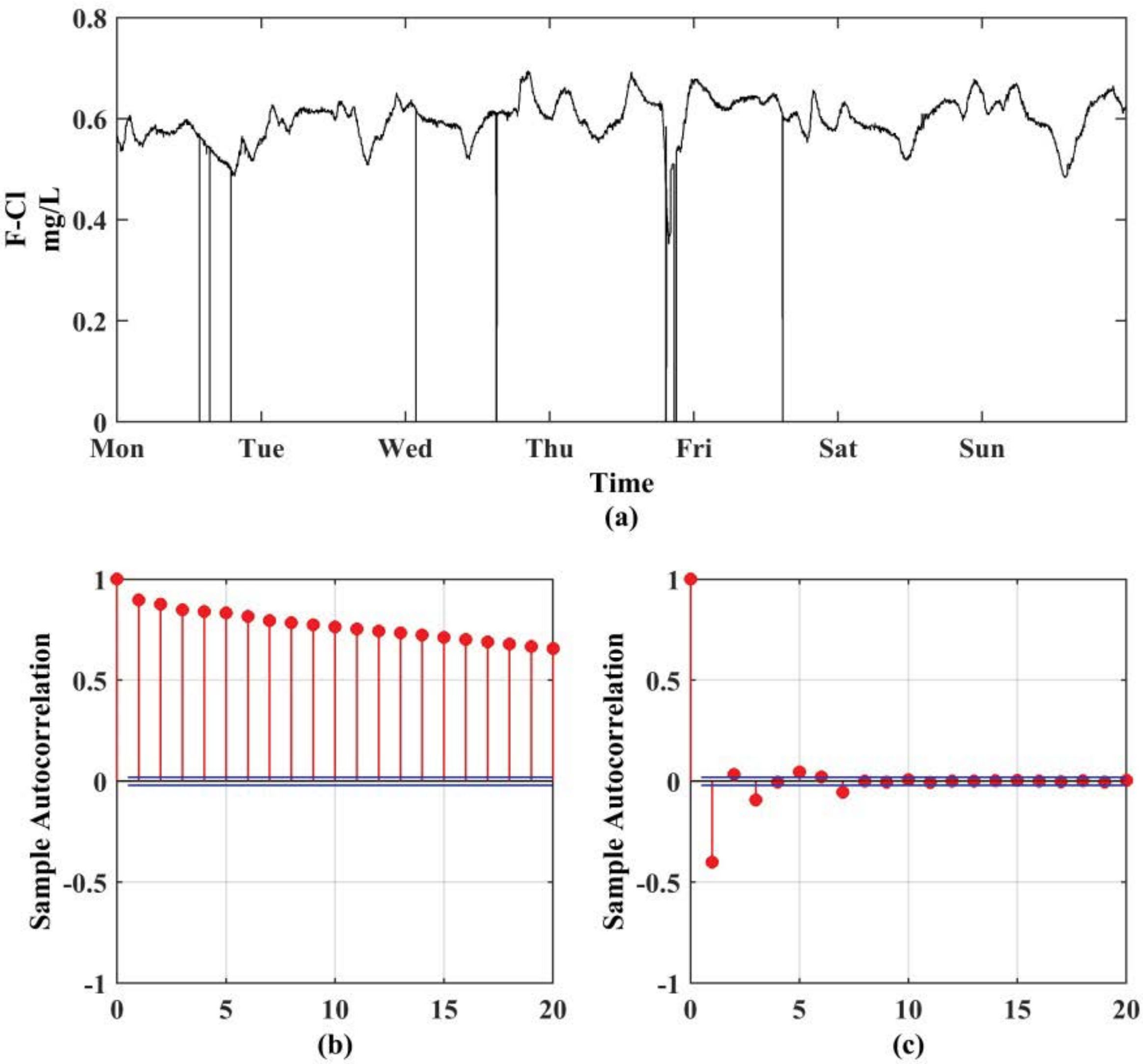

Fig. 3. Time series plot and SACF of the Data No. C-1: (a) the time series plot, (b) SACF of raw time series, and (c) SACF of first-order differencing time series.

data. The SACF of Data No. C-1, shown in Fig. 3b, gradually decreased, whereas the SACF of the first-order differencing data (Fig. 3c) showed a relatively large spike at lag 1. Therefore, Data No. C-1 indicates that the times series is nonstationary and has a probabilistic trend.

Fig. 4 presents a plot of time series and the SACF for Data No. P-1. As shown in Fig. 4a, Data No. P-1 indicates that the times series was clearly nonstationary with an increasing trend. The SACF of Data No. P-1 (Fig. 4b) gradually decreased, whereas that of the first-order differencing data (Fig. 4c) showed a relatively large spike at lag 1, 2, and 3.

To quantitatively analyze the nonstationarity in Data No. C-1 and Data No. P-1, the order of the ARIMA (p,d,q) model was set at $p \leq 2$, $d \leq 1$, and $q \leq 2$ according to the principle of parsimony in model building recommended by Box and Jenkins [15]. Schwartz's Bayesian Criterion (SBC) was calculated to compare the raw time series group (d = 0) and the differenced time series data groups (d = 1). Here, SBC is an information criterion used to assess the quality of the model fitting and is defined as:

$$\mathrm{SBC}(M) = n\ln\hat{\sigma}_\epsilon^2 + M\ln n \quad (5)$$

where M is the number of parameters in the model, $\hat{\sigma}_\epsilon^2$ is the maximum likelihood estimation of σ_ϵ^2, and n is the effective number of observations.

Table 2 summarizes the SBC calculation results for Data No. C-1 and Data No. P-1. In the case of Data No. C-1, the average of SBC for the models with the raw time series group was –6.145, whereas the average of SBC for the models with the differenced time series groups was –7.571 (SBC_{max} = –7.590, ARIMA (2,1,1)), which is about 23% lower. In the case of Data No. P-1, the average of the SBC for the models with the raw time series group was –1.447, whereas that for the models with the differenced time series groups was –2.357 (SBC_{max} = –2.393, ARIMA (2,1,2)),

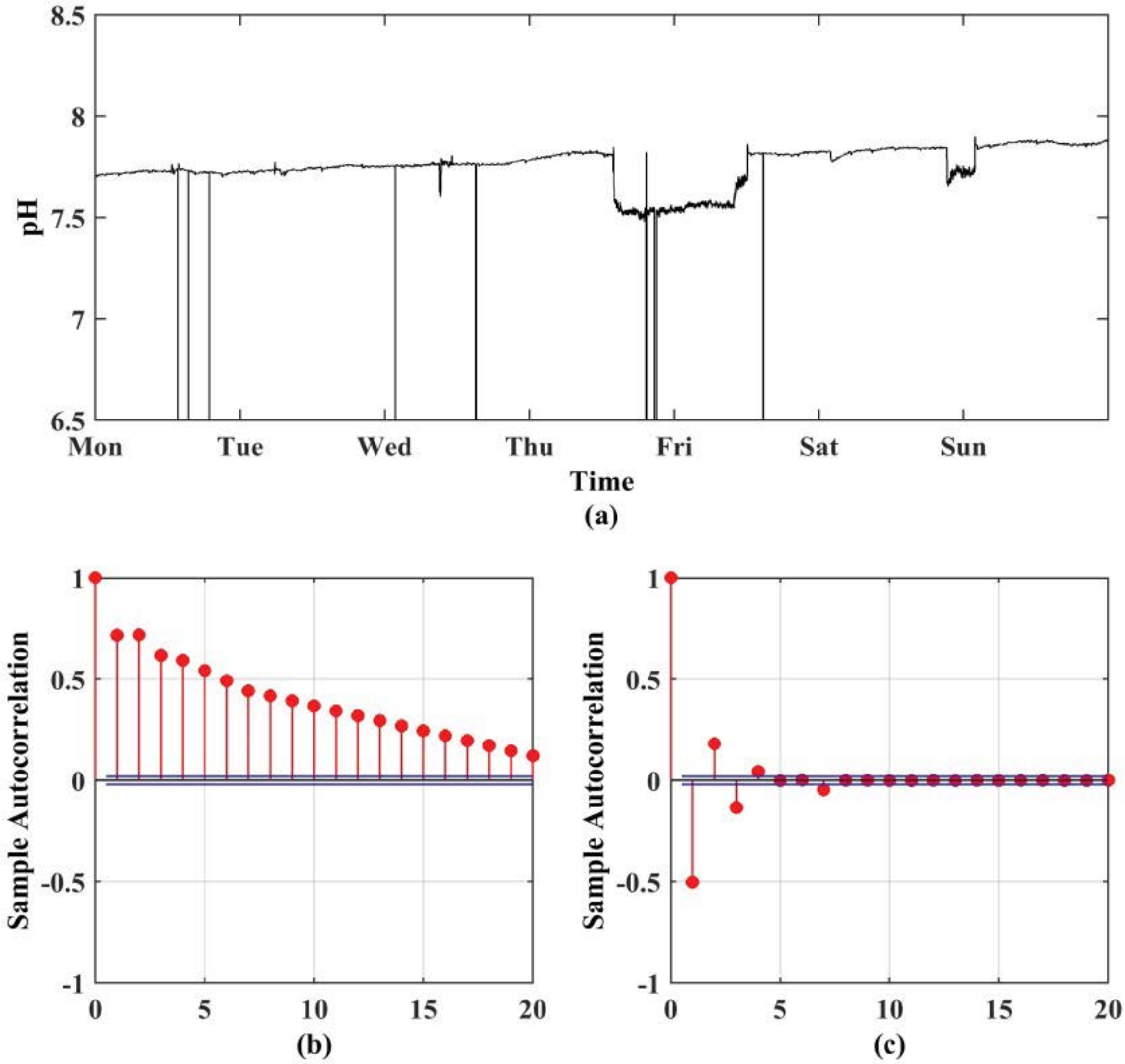

Fig. 4. Time series plot and SACF of the Data No. P-1: (a) time series plot, (b) SACF of raw time series, and (c) SACF of first-order differencing time series.

Table 2
SBC calculation results of Data No. C-1 and Data No. P-1

Data No.	ARMA (p,q)	SBC	
		d = 0	d = 1
No. C-1	1,0	–7.303	–7.523
	1,1	–7.519	–7.586
	0,1	–2.329	–7.582
	2,0	–7.464	–7.545
	2,1	–7.521	–7.590
	2,2	–7.518	–7.590
	0,2	–3.363	–7.584
No. P-1	1,0	–2.001	–2.335
	1,1	–2.281	–2.351
	0,1	–2.812	–2.341
	2,0	–2.275	–2.344
	2,1	–2.290	–2.392
	2,2	–2.297	–2.393
	0,2	–1.799	–2.346

which is about 39% lower. Therefore, in this study, the LPCF model based on the ARIMA model with differencing transformation was developed considering the nonstationarity in the WQPs. The detailed results are given below.

In order to develop the LPCF model, it is necessary to identify the appropriate ARIMA (p,d,a) model and the size of the analysis window. As previously mentioned, the analysis window is a sampling space for estimating the parameters of the LPCF model and should be shifted to the same WQP observation period. In a previous study of the U.S. EPA for pipe networks, the optimal size of the analysis window was 2,880 min [7,8]. In this study, the size of the analysis window was set to 2,880 min, and the tentative LPCF model based on the ARIMA model for each WQP was identified. In addition, the size of the analysis window was reanalyzed because the WTP was selected for the study area, unlike the previous study of the U.S. EPA for pipe networks. To identify the optimal model, seven ARIMA models ($p \le 2$, $d \le 1$, and $q \le 2$) were built, and the mean squared prediction error (MSE), mean absolute prediction error (MAE), and correlation coefficient (CC) were selected for statistical evaluation of each model. The MSE, MAE, and CC are expressed as:

$$\text{MSE} = \frac{\sum_{i=1}^{n}\left(Z_i - \hat{Z}_i\right)^2}{n} \tag{6}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}\left|Z_i - \hat{Z}_i\right|}{n} \tag{7}$$

$$\text{CC} = \frac{\text{Cov}\left(Z_i, \hat{Z}_i\right)}{\sqrt{\text{Var}\left(Z_i\right)\text{Var}\left(\hat{Z}_i\right)}} \tag{8}$$

where Z_i is the value of observed WQPs at time i, $\hat{Z}_i$ is the value of estimated WQPs at time i, and n is the number of data.

Table 3 summarizes the MSE, MAE, and CC calculation results for Data No. C-1 and Data No. P-1. In the case of Data No. C-1, the CC of all models was more than 0.90 and showed a strong positive correlation with the observations. The MSE showed the same results in the seven models, and the difference of MAEs in each model was small but showed the minimum value in the ARIMA (1,1,0) model. In the case of Data No. P-1, the ARIMA (2,1,0), ARIMA (2,1,1), and ARIMA (2,1,2) models were excluded from this analysis because of the parameters of the ARIMA model were not converged in some points. The CC of the ARIMA (1,1,0), ARIMA (1,1,1), and ARIMA (0,1,1) models satisfying the parameter estimation showed a good fit with the observations. The MSE showed the same result in the three models, and the MAE showed the minimum value in the ARIMA (1,1,0) model. Considering the results of the test criterion, the difference in prediction performance between models was small when the order of the ARIMA model was $p \le 2$, $d \le 1$, and $q \le 2$. However, the ARIMA (1,1,0) model is expected to be the best time series model for building the LPCF model considering the parsimony and the efficiency of parameter estimation in the identification of the ARIMA model.

To estimate a reasonable analysis window size for WTP, the analysis window was set from 180 to 2,880 min, and the ARIMA (1,1,0) model was applied as a fundamental prediction model of the LPCF model.

Fig. 5a illustrates the CC and MAE results according to the analysis window for Data No. C-1. In all of the analysis windows, the CC showed a positive correlation of 0.8 or more. The MAE decreased gradually with an increase in the analysis window; a relatively large decrease was found at 1,080 min of the analysis window. Fig. 5b shows the CC and MAE results according to the analysis window for Data No. P-1. In the entire analysis window, the CC showed strong fitness of 0.99 or more. The MAE increased gradually with an increase in the analysis window, which is different from that of Data No. C-1, and a relatively large increase appeared after 720 min of the analysis window. Therefore, considering the results of this study, it is expected that the analysis window of a shorter interval than that for the pipe network can be applied when developing an LPCF model for WTP.

4.2. Validation of LPCF model

The ARIMA (1, 1, 0) model was selected as an optimal prediction model for F–Cl and pH through analysis of the training dataset. To test the LPCF model based on this ARIMA model, Data No. C-2 and Data No. P-2, shown in Table 1, were adopted as test data. In the development of the LPCF model, the analysis window was applied for 1,440 min equally; the predicted value of the LPCF model was substituted when the observation data were missing.

Fig. 6 illustrates the time series plot of Data No. C-2, the simulation results of the LPCF model, and the residual. Except for partial missing points, Data No. C-2 oscillated up and down around the mean value without rapid

Table 3
MSE, MAE, and CC calculation results of Data No. C-1 and Data No. P-1

Data No.	ARIMA (*p*,*d*,*q*)	MSE	MAE	CC
No. C-1	**1,1,0**	**0.0021**	**0.0030**	**0.908**
	1,1,1	0.0021	0.0039	0.907
	0,1,1	0.0021	0.0038	0.907
	2,1,0	0.0021	0.0036	0.908
	2,1,1	0.0021	0.0038	0.907
	0,1,2	0.0021	0.0039	0.907
No. P-1	**1,1,0**	**0.0001**	**0.0042**	**0.996**
	1,1,1	0.0001	0.0054	0.995
	0,1,1	0.0001	0.0054	0.995
	2,1,0	–	–	–
	2,1,1	–	–	–
	2,1,2	–	–	–

fluctuation (Fig. 6a). As shown in Fig. 6b, the estimates of the LPCF model were distributed in a similar manner as that of the observations; the distribution of the residuals is typical of white noise.

Fig. 7 illustrates the time series plot of Data No. P-2, the simulation results of the LPCF model, and the residual. As shown in Fig. 7a, the time series of Data No. P-2 had a partial missing value and a base change in the middle part of the observation. Moreover, the time series was more static than that of Data No. C-2. The LPCF model for Data

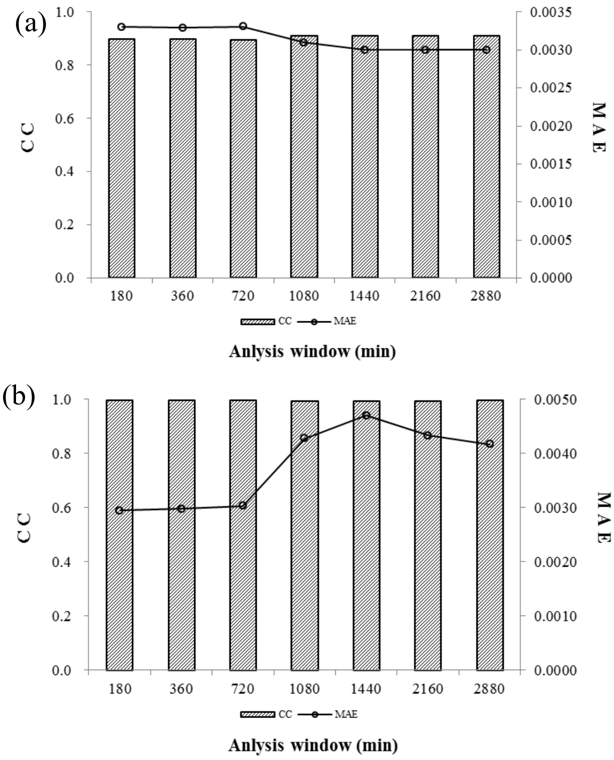

Fig. 5. Results of CC and MAE according to the analysis window: (a) Data No. C-1 and (b) Data No. P-1.

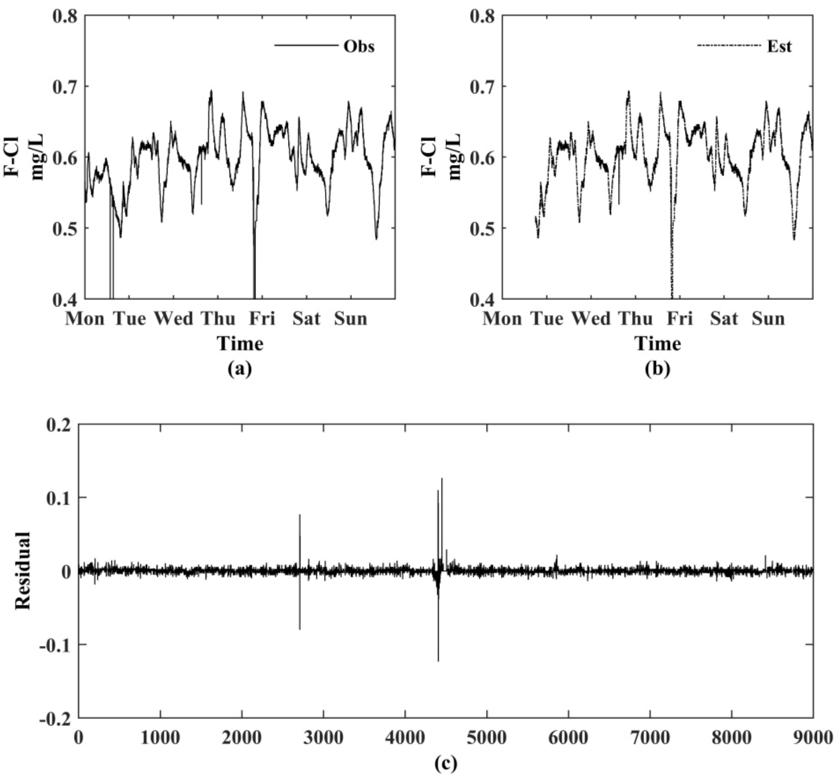

Fig. 6. Simulation results of Data No. C-2 using the LPCF model (ARIMA (1,1,0)) and a window analysis size of 1,440 min.

Table 4
MSE, MAE, and CC calculation results of Data No. C-2 and Data No. P-2

Data No.	Data period	MSE	MAE	CC
No. C-2	2013/05/06–2013/05/12	0.0001	0.0014	0.950
No. P-2	2013/05/06–2013/05/12	0.0000	0.0015	0.998

No. P-2 simulated the overall variation of pH including the base fluctuations shown in Fig. 7b. The distribution of the residuals also showed typical white noise type except for the missing and base fluctuation points. The CC, MSE, and MAE were calculated to quantitatively analyze the performance of both LPCF models.

Table 4 summarizes the estimation results of the test criterion for Data No. C-2 and Data No. P-2. The correlation between observation and estimates was higher than 0.95 for both models. The MSEs of Data No. C-2 and Data No. P-2 were 0.0001 and 0.0000, respectively, and the MAEs were 0.0014 and 0.015, respectively. The CC, MSE, and MAE indicate that the LPCF model based on the ARIMA (1,1,0) model developed in this study can accurately simulate the changes in time series of WQPs observed at the WTP.

5. Conclusion

The purpose of this study was to develop real-time CW technology in the WSS of South Korea. For this purpose, an LPCF model, a real-time CW model developed by U.S. EPA, was applied to G_WTP in South Korea. However, in this study, the ARIMA model was applied as a fundamental prediction model of the LPCF model considering the nonstationarity of the time series of WQPs. This application is different from that of the previous study by the U.S. EPA. The results are summarized in the following points.

- Analysis of the time series plot and SACF of the collected WQP dataset in the G_WTP of South Korea, including F–Cl and pH, revealed that both times series were nonstationary and showed a probabilistic trend.
- Considering the nonstationary of the time series of WQPs, an LPCF model based on the ARIMA model was developed. As a result, the prediction performance of the models including the first-order differencing transformation indicated more accurate estimation results. Moreover, the ARIMA (1,1,0) model was identified as the optimal time series model for both the time series of the WQPs. Therefore, it is considered that the ARIMA model including first-order differencing transformation is a more suitable fundamental prediction model for building the LPCF model for real-time CW of WSS.
- In order to estimate a reasonable size for the analysis window for the LPCF model, the analysis window was set from 180 to 2,880 min, and the ARIMA (1,1,0) model was applied as the fundamental prediction model of the LPCF model. In a previous study by the U.S. EPA on pipe networks, the size of the analysis window was suggested as more than 2,440 min. However, in this study

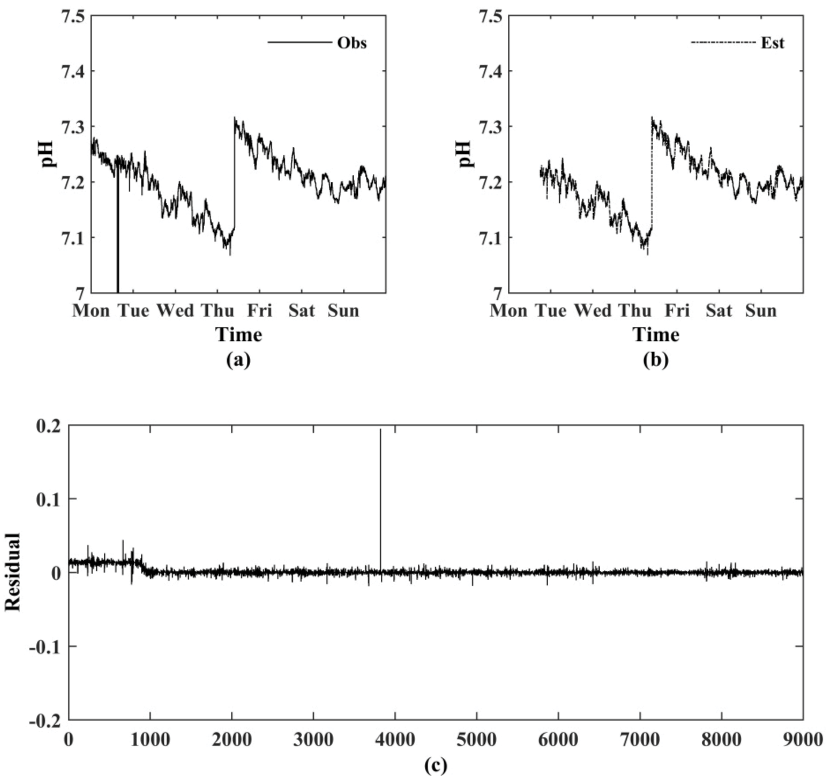

Fig. 7. Simulation results of Data No. P-2 using the LPCF model (ARIMA (1,1,0)) and a window analysis size of 1,440 min.

on WTPs, an analysis window shorter than 2,440 min was applicable.

Further, compared with the U.S. EPA, which conducts leading research on the development of CW and surveillance systems, South Korea's research environment and performance are relatively inadequate. Therefore, it is expected that the results of this study can be used as a useful basic tool for developing CW and surveillance systems in the WSSs that are newly established or undergoing expansion.

Acknowledgment

This research was supported by a grant (16AWMP-B113766-01) from the Research & Development (R&D) Program on Water Management funded by the Ministry of Environment of the Korean Government.

References

[1] W.E. Hart, R. Murray, Review of sensor placement strategies for contamination warning systems in drinking water distribution systems, J. Water Resour. Plann. Manage., 136 (2010) 611–619.
[2] B.H. Tangena, P.J.C.M. Janssen, G. Tiesjema, E.J. van den Brandhof, M. Klein Koerkamp, J.W. Verhoef, A. Filippi, W. van Delft, A Novel Approach Early Warning of Drinking Water Contamination Events, U. Borchers, J. Gray, K.C. Thompson, Eds., Water Contamination Emergencies: Monitoring, Understanding and Acting, Royal Society of Chemistry, Cambridge, 2011, pp. 13–31.
[3] N.-S. Park, Y.-J. Lee, S.-H. Chae, S.M. Yoon, A study on the statistical predictability of drinking water qualities for contamination warning system, J. Korean Soc. Water Wastewater, 29 (2015) 469–479.
[4] S.M. Yoon, S.-S. Kim, S.-H. Chae, N.-S. Park, Introducing new outlier detection method using robust statistical distance in water quality data, Desal. Water Treat., 149 (2019) 157–163.
[5] U.S. EPA, Online Water Quality Monitoring Primer For Water Quality Surveillance and Response Systems (EPA 817-B-15–002A), United States Environmental Protection Agency, 2015a.
[6] U.S. EPA, Summary of Implementation Approaches and Lessons Learned From the Water Security Initiative Contamination Warning System Pilots (EPA 817-R-15–002), United States Environmental Protection Agency, 2015b.
[7] WRF, Interpreting Real-Time Online Monitoring Data for Water Quality Event Detection (Project 4182), Water Research Foundation, 2014.
[8] U.S. EPA, Water Security Initiative: System Evaluation of the Cincinnati Contamination Warning System Pilot (EPA 817-R-14–001A), United States Environmental Protection Agency, 2014a.
[9] U.S. EPA, Water Security Initiative: Evaluation of the Water Quality Monitoring Component of the Cincinnati Contamination Warning System Pilot (EPA 817-R-14–001B), United States Environmental Protection Agency, 2014b.
[10] U.S. EPA, Event Detection, Deployment, Integration and Evaluation System (EPA 817-B-14–004), United States Environmental Protection Agency, 2014c.
[11] U.S. EPA, Water Quality Event Detection System Challenge: Methodology and Findings (EPA 817-R-13–002), United States Environmental Protection Agency, 2013.
[12] U.S. EPA, Water Quality Event Detection System for Drinking Water Contamination Warning System: Development, Testing, and Application of CANARY (EPA 600-R-010–036), United States Environmental Protection Agency, 2013.
[13] U.S. EPA, Cincinnati Pilot Post-Implementation System Status: Covering the Pilot Period: December 2005 Through December 2007 (EPA 817-R-08–004), United States Environmental Protection Agency, 2008.
[14] W.W.S. Wei, Time Series Analysis: Univariate and Multivariate Methods, 2nd ed., Pearson Addison Wesley, California, 2006.
[15] P.J. Brockwell, R.A. Davis, Introducing Time Series and Forecasting, 2nd ed., Springer, New York, 2001.
[16] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed., John Wiley and Sons, New Jersey, 2015.