Using Generalized Time Series Transfer Function Model for Automated Water Quality Monitoring

Cheolyong Park, Ji Eun Moon, Gyu Moon Song, Tae Yoon Kim*

Department of Statistics, Keimyung University, Daegu 704-701, Korea

(Received March 25, 2013; Revised April 20, 2013; Accepted May 5, 2013)

ABSTRACT

In this article we propose a generalized time series transfer function model and apply it to building a water quality monitoring algorithm (WQMA) and waste dumping simulation algorithm (WDSA). Our empirical experiments indicate that WQMA is quite effective against various dumping situations simulated by WDSA. In particular, it is interesting to report that WQMA might even identify furtive waste dumping such as dumping with rain.

Key words : Automated water quality monitoring, Generalized transfer function model, Waste dumping simulation algorithm

Generalized Transfer Function Model

The transfer function model relates a variable to the present and the past of other variables. The influence of one variable on another can be spread over several time periods. Instantaneous and lagged effects of an input variable X_t on an output variable Z_t can be represented by a model in the form

$$Z_t = v(B)X_t \tag{1}$$

where

$$v(B) = \frac{\omega(B)B^b}{\delta(B)}. \tag{2}$$

The operators $v(B)=v_0+v_1B+v_2B^2+\cdots$, $\omega(B)=\omega_0-\omega_1B-\cdots-\omega_sB^s$ and $\delta(B)=1-\delta_1B-\cdots-\delta_rB^r$ are polynomials in backshift operator B, and b is a parameter representing the delay between the variables. Furthermore, it is assumed that the roots of $\delta(B)=0$ are on or outside the unit circle. The relation between coefficients v_k and parameters $\mathbf{w}=(\omega_0,\cdots,\omega_s)'$, $\mathbf{d}=(\delta_1,\cdots,\delta_r)'$ and b can be obtained by equating the coefficients of B^k in

$$\delta(B)v(B) = \omega(B)B^b. \tag{3}$$

Model (1) is extended to the transfer function-noise model of the form

$$Z_t = v(B)X_t + N_t \tag{4}$$

where N_t is assumed to be uncorrelated with X_t. Refer to [1]. In this article we introduce a generalized time series transfer function model

$$Z_t = \sum_{i=1}^{q} v_i(B:\mathbf{w}_i,\mathbf{d}_i(X_{it}))X_{it} + u_t, \tag{5}$$

which includes q independent input variables and allows the dependence of $\mathbf{d}_i$ on X_{it}. Here, u_t can be represented by an ARIMA model. Note that this model assumes that for a given i, the past X_{it}'s not only influence future Z_t's, but also determine the strength of influence for the past X_{it}s over a long period of time via dependence of $\mathbf{d}_i$ on X_{it}. Note that when δ_i is close to 1, it implies a strong or long range dependence of X_{it}.

Time series are frequently affected by exogenous events usually referred to as interventions. Interventions can affect the response in several ways. They can change the level of a series either abruptly or after some delay, change the trend, or lead to other, more complicated, effects. The generalized transfer function model in (5) can be utilized to determine whether there is evidence that such a change in the series has actually occurred

* Correspondence should be addressed to Dr. Tae Yoon Kim, Department of Statistics, College of Natural Sciences, Keimyung University, Daegu 704-701, Korea. Tel: +82-53-580-5533, Fax: +82-53-580-5164, E-mail: tykim@kmu.ac.kr

and, if so, its nature and magnitude. In fact, we introduce, for the effect of interventions, the transfer function-noise model of the form

$$Z_t = \sum_{i=1}^{q} v_i(B{:}\mathbf{w}_i, \mathbf{d}_i(I_{it}))I_{it} + u_t, \tag{6}$$

where I_{it} is an indicator sequence reflecting the absence and presence of an intervention (or $I_{it}(T)=1$ if $t=T$ and zero, otherwise). In (6), prior to the intervention, u_t and Z_t are the same and can be represented by an ARIMA model. In (6), we have q types of interventions and a given ith type intervention I_{it}, which has spread influence according to $\mathbf{d}_i(I_{it})$ and $\mathbf{w}_i$. For possible choices of $v(B{:}\mathbf{w},\mathbf{d})$ in (6), one may consider $v(B{:}\mathbf{w},\mathbf{d}) = \omega_{01}/(1-B)$ for a step change at time T, $v(B{:}\mathbf{w},\mathbf{d}) = \omega_{02}/(1-\delta B)$ for an initial increase followed by a gradual decrease without the lasting effect and $v(B{:}\mathbf{w},\mathbf{d}) = [\omega_{01}/(1-\delta B)] + [\omega_{02}/(1-B)]$ for an initial increase followed by a gradual decrease with lasting effect ω_{02}.

Time series are influenced by repetitious interventions, e.g., waste dumping against water quality time series. If the timing of such interventions is known, intervention models can be used to account for their effects. However, in practice, the timing is frequently unknown. Because the effects of interventions can bias the parameter estimates, forecasts, and seasonal adjustments, it is important to develop procedures that can help detect and remove such effects. This is known as the problem of outliers or spurious observations [2]. discusses two characterizations of outliers in the context of time series models. The aberrant observation model is as follows:

$$u_t^* = u_t + \omega I_t(T) \text{ and, } \phi(B)u_t = \theta(B)a_t \tag{7}$$

the aberrant innovation model is as follows:

$$u_t^* = u_t + \phi^{-1}(B)\theta(B)\omega I_t(T) = \phi^{-1}(B)\theta(B)[a_t + \omega I_t(T)] \tag{8}$$

Here, u_t^* denotes the observed time series, u_t is the underlying ARIMA process without the impact of outliers and a_t refers to white noise. In the aberrant observation model, only the level of the T^{th} observation is affected. In the aberrant innovation model, the outlier affects the shock at time T, which in turn influences u_T, u_{T+1}. Refer to [1] for its detailed explanation.

Water Quality Monitoring vs Waste Dumping

The traditional approach to water quality monitoring involves manually sampling water at remote sites and transporting it to a laboratory for chemical analysis. This approach, while relatively non-technical and easily repeatable, does not allow for continuous data collection and monitoring. Recently, new technology and instrumentation have developed water quality monitoring stations, which are able to monitor water quality continuously regardless of weather and accessibility (see, e.g., [3-6]). This advanced water quality monitoring station provides water quality data for documenting spatial and temporal changes in water quality, identify and respond to pollution or other water quality episodes, and compare water quality to various water standards (e.g., drinking, agriculture, industry or fishery standards). In this section, we are mainly concerned about using the generalized time series transfer function model for monitoring water quality, particularly identifying illegal dumping waste by individual or firm through continuous automated station monitoring. Note that dumping still occurs illegally almost everywhere, although dumping waste is well known to have serious economic and health impacts by killing aquatic life and damaging the habitats and ecosystems (see [7-9]).

In order to operate the station for identifying or detecting dumping waste, the water quality monitoring algorithm (WQMA) was built, which focuses and finds an appropriate model for water quality time series data "under the normal condition." The underlying idea of WQMA is that if the normal movement of water quality over time is figured out successfully, then one would be able to identify waste dumping effectively by detecting the moment when the temporal water quality path deviates from its normal path significantly. This approach is quintessential for detecting waste dumping because it is almost impossible to model waste dumping into a temporal process based on past waste dumping episodes. More specifically, in most cases, we do not have sufficient information for building a wasting dumping process because most waste dumping frequently occurs without being caught and recorded. Note that waste dumping tends to be performed during nighttime, weekends or when it rains in order to avoid surveillance by government authorities. For modeling the normal temporal path of water quality, one should obtain water quality data under normal conditions, which are not only free from waste dumping but which also reflect the change of the water quality due to natural occasions. For example, the normal condition should relate to the seasonal effect (i.e., water quality changes periodically with season) or rainfall effect (i.e., water quality improves with rainfall). However, it is hard to gather such data since in reality, water quality data tend to be contaminated by unidentified sources, including waste dumping. Such difficulty is resolved by employing the generalized time series intervention model of Section 1, which effectively extracts or isolates waste dumping from the contaminated water quality time series data. Keep in mind that the time series intervention model is useful in isolating the unusual external events that influence or intervene in the normal path of the time series of concern.

In order to investigate the performance of WQMA, it is applied and then tested against in both empirical and simulated situations. Our test results show that WQMA reports quite accurate and sensitive performances in identifying waste dumping, particularly furtive ones such as rain dumping. At this juncture, it is worth mentioning that we newly built a waste dumping simulation algorithm (WDSA) that simulates various waste dumping episodes. Taking into account the difficulty for testing WQMA with the real data whose waste dumping episodes are completely known, the construction of WDSA is an essential pillar of the current work. It will be noticed later that we have employed model (6) and the outlier removal procedure based on model (7) for constructing WQMA, and that model (5) has been used for WDSA. Finally, it should be noted that until now, the time series model for water quality data has been employed mainly for prediction; however, we extend the model for monitoring and simulation purposes (see [10-14]).

WQMA and WDSA

1. WQMA construction

In order to develop WQMA using the generalized transfer function model, let Z be the selected variable for water quality monitoring by the station and $\{Z_t\colon t=1,2,\cdots\}$ be a sequence of their observations. Assume that $\{Z_t\colon t=1,2,\cdots\}$ follows a time series intervention model, i.e.,

$$\begin{aligned} Z_t &= f_0(t)+f_1(I_{11}(t),\cdots,I_{1p_1}(t))+f_2(I_{21}(t),\cdots,I_{2p_2}(t))+\eta_t \\ &= f_0(t)+f_1(t)+f_2(t)+\eta_t \end{aligned} \tag{9}$$

where $f_0(t)$ denotes a time trend function, $f_1(t)$ and $f_2(t)$ denote the time series intervention model for the rainfall effect and dumping waste effect, respectively, where

$$I_{1j}(t)=\begin{cases} 1 & t=R_j \\ 0 & \text{otherwise} \end{cases} \quad \text{for} \quad j=1,\cdots,p_1$$

and

$$I_{2j}(t)=\begin{cases} 1 & t=D_j \\ 0 & \text{otherwise} \end{cases} \quad \text{for} \quad j=1,\cdots,p_2.$$

Indeed rainfall and dumping wastes occur at $t=R_1,\cdots,R_{p_1}$ and $t=D_1,\cdots,D_{p_2}$ respectively and $\{\eta_t\colon t=1,2,\cdots\}$ is a stationary ARMA (u,v) (auto-regressive moving average) process (i.e., $\phi(B)\eta_t=\theta(B)a_t$).

As mentioned in Section 2, the main objective of WQMA is to figure out $\eta_t+f_0(t)+f_1(t)$, which describes the normal evolution of Z_t over time, say $Z_t^{(N)}$. In order to specify f_0 and f_1, we employ

$$f_0(t)=\alpha_0 \text{ and } f_1(t)=f_1(I_{11}(t),\cdots,I_{1p_{11}}(t))=\alpha_1\sum_{j=1}^{p_1} I_{1j}(t). \tag{10}$$

For f_2, we employ the following time intervention model:

$$\begin{aligned} f_2(t) &= f_2(I_{21}(t),\cdots,I_{2p_2}(t)) \\ &= \frac{\omega_1(B)B^{b_1}}{\delta_1(B)} I_{21}(t)+\cdots+\frac{\omega_{p_2}(B)B^{b_{p_2}}}{\delta_{p_2}(B)} I_{2p_2}(t), \end{aligned} \tag{11}$$

where $\omega_i(B)=\omega_{i0}-\omega_{i1}B-\cdots-\omega_{is_i}B^{s_i}$ and $\delta_i(B)=\delta_{i0}-\delta_{i1}B-\cdots-\delta_{ir_i}B^{r_i}$ for $i=1,\cdots,p_2$, B is a backshift operator, and b_i denotes the delayed time for the intervention effect. Also, f_2 is used mainly because Z_t is subject to the influence of waste dumping over a certain duration of time after its occurrence. Now, model (9) is in fact model (6), where

$$u_t=f_0(t)+\eta_t, \quad v_1(B\colon \mathbf{w}_1,\mathbf{d}_1(I_{1t}))I_{1t}=\sum_{j=1}^{p_1}\alpha_1 I_{1j}(t)=f_1(t)$$

$$v_2(B\colon \mathbf{w}_2,\mathbf{d}_2(I_{2t}))I_{2t}=\sum_{j=1}^{p_2}\frac{\omega_j(B)B^{b_j}}{\delta_j(B)} I_{2j}(t)=f_2(t).$$

Assuming that the training data set $\{Z_t\colon t=1,\cdots,n_0\}$ and $\{R_1,\cdots,R_{p_1}\}$, WQMA is trained by the data of size n_0 currently available and is applied to the test data of size n_1 streaming into WQMA immediately after its training. Thus, the main feature of the WQMA is that it is designed for *automatic implementation* on a relatively short period of time so that it might be easily updated for real-time monitoring of water quality. Let $\{Z_t\colon t=1,2,\cdots,n_0\}$ and $\{Z_t\colon t=n_0+1,\cdots,n_0+n_1\}$ denote training and testing data set, respectively. Then, WQMA can be described as follows:

(Step 1) With the training data and precipitation points $R_1,\cdots,R_{p_1}$ available, fit $\hat{Z}_t^{(N)}=\hat{\eta}_t+\hat{f}_0(t)+\hat{f}_1(t)$ and then find $\hat{D}_1,\cdots,\hat{D}_{P_2}$ as outliers.

(Step 2) Using $\hat{D}_1,\cdots,\hat{D}_{P_2}$ obtained from Step 1 together with the training data and precipitation points, build the final model $\hat{Z}_t=\hat{\eta}_t+\hat{f}_0(t)+\hat{f}_1(t)+\hat{f}_2(t)$.

(Step 3) Using the results from Step 2, calculate the residuals $e_t=Z_t-\hat{f}_2(t)-\hat{Z}_t^{(N)}$ for $t=1,\cdots,n_0$, where $\hat{Z}_t^{(N)}=\hat{\eta}_t+\hat{f}_0(t)+\hat{f}_1(t)$. With these residuals $\mathbf{e}=(e_1,\cdots,e_{n_0})$, estimate its marginal distribution $N(\hat{\mu}_e,\hat{\sigma}_e^2)$ by assuming the normality of e.

(Step 4) For $t=n_0+1,\cdots,n_0+n_1$, calculate the corresponding residuals $e_t^{(1)}=Z_t-\hat{Z}_t^{(N)}$. If $e_t^{(1)}=Z_t-\hat{Z}_t^{(N)}$ exceeds the 95$^{\text{th}}$-percentile of $N(\hat{\mu}_e,\hat{\sigma}_e^2)$, a warning is issued.

Remarks. There are several technical aspects to be discussed regarding the automated WQMA. First, models (7) and (6) (generalized time series intervention models) provide the basic tools to (Step 1) and (Step 2). Recall that because it is usually hard or impossible to obtain the exact information about $D_1,\cdots,D_{p_2}$ (p_2 waste dumping time points) from the training data in practice, they are to be estimated as $\hat{D}_1,\cdots,\hat{D}_{P_2}$ in order to build $\hat{f}_2$. In order to estimate them, we predefine $D_1,\cdots,D_{p_2}$ at which the Z_ts are outliers from the training data and find them by hiring an outlier detection process (model (7)). Second, real-time monitoring usually refers to fine scale monitoring on the time domain and hence, a relatively short period (or microscopic) of monitoring is preferred. Because the constant function would be sufficient as the basis function over a short period of time, we use the constant function for f_0 and f_1. Third, residual e_t of (Step 3) estimates η_t in model (9), whereas residual $e_T^{(1)}$ of (Step 4) estimates the error when there is no waste dumping. More specifically, the difference between e_T and $e_T^{(1)}$ is estimating the effect purely due to dumping waste, if it exists, and hence the difference is clearly a reasonable test statistics that tests the null hypothesis H_0: there is no dumping waste, or the underlying difference is zero. Using R, Song et al. [15] completed an algorithm that implements WQMA automatically. Their algorithm uses $n_0=200$ training data and $n_1=40$ testing data and is conducted by the aid of automatic ARMA (u,v) model selection and outlier detection R algorithm.

2. WDSA construction

For testing the efficacy of auto-WQMA, we need the test data set to have exact information regarding waste dumping points $D_1,\cdots,D_{p_2}$. However, it will be almost impossible to obtain such data from the real situation because illegal dumping moments are always unknown unless they get caught. This strongly suggests the need to develop an algorithm which virtually simulates dumping under various situations. Regarding the precipitation simulation, on the other hand, one needs to consider not only the precipitation itself, but also its amount. In particular, we are interested in the situation where dumping is made with rain. This is needed because waste dumping tends to conspicuously increase under heavy rain, such as hurricanes and typhoons. Note that heavy rain will dilute and wash out the polluted surface water with its lasting impact, whereas light rain will not be able to dilute the polluted surface water almost at all.

Assuming model (9) based on model (5), we consider X_t, the precipitation amount at t, as the predictor for f_1, and model f_1 as

$$f_1^{(s)}(t)=v_1(B{:}\mathbf{w}_1,\mathbf{d}_1(X_tI_{1t}))X_tI_{1t}=\sum_{j=1}^{p_1}\frac{g_1(X_t)}{g_2(X_t,B)}I_{1j}(t), \tag{12}$$

where $I_{1j}(t)=1$ if it rains at t and $=0$ otherwise, $g_1(X_t)=\beta_1X_t$ for positive constants β_0 and β_1, and $g_2(X_t,B)=1-\rho(X_t,\beta_2)B$, where $\rho(X_t,\beta_2)=\dfrac{X_t}{X_t+\beta_2}$ for some $\beta_2>0$. Here, $g_1(X_t)$ assumes that X_t proportionately determines the degree of water quality improvement and $g_2(X_t,B)$ assumes that X_t influences the lasting impact of the precipitation decaying exponentially over time. Note that $\rho(X_t,\beta_2)=\dfrac{X_t}{X_t+\beta_2}$ as a function of $X_t(\geq 0)$ approaches 1 from below as X_t increases to infinity. In other words the lasting impact of precipitation over time elongates as precipitation amount X_t increases. As a result, we employ

$$f_1^{(s)}(t)=v_1(B{:}\mathbf{w}_1,\mathbf{d}_1(X_tI_{1t}))X_tI_{1t}=\sum_{j=1}^{p_1}\frac{\beta_1X_t}{1-\dfrac{X_t}{X_t+\beta_2}B}I_{1t}(t). \tag{13}$$

Here, (13) is reasonable because it stipulates that precipitation X_t has a linear relation on Z_t (water quality measurement) and its exponentially decaying impact on Z_t, over time comes strong with large X_t.

For modeling f_2, recalling that the intervention term could measure the lasting impact of the external events, we assume

$$f_2^{(s)}(t)=\sum_{j=1}^{p_2}\frac{\omega_j(B)B^{b_j}}{\delta_j(B)}I_{2j}(t)=\sum_{j=1}^{p_2}\frac{\beta_3}{1-\beta_4B}I_{2j}(t), \tag{14}$$

where $\beta_4=\beta_3/c$ and $0<\beta_3<c$ for some fixed positive constant c. Here, β_3 might be related to the degree of waste dumping and β_4 (<1) is proportional to β_3. A more sophisticated model, such as $f_1^{(S)}$ above, is possible by quantifying the degree of waste dumping quantitatively. Now our waste dumping simulation algorithm produces Z_t s from

$$Z_t=f_0^{(s)}(t)+f_1^{(s)}(t)+f_2^{(s)}(t)+\eta_t, \tag{15}$$

where $f_0^{(S)}(t)=\beta_0$ and η_t is an ARMA process. Note that mo-

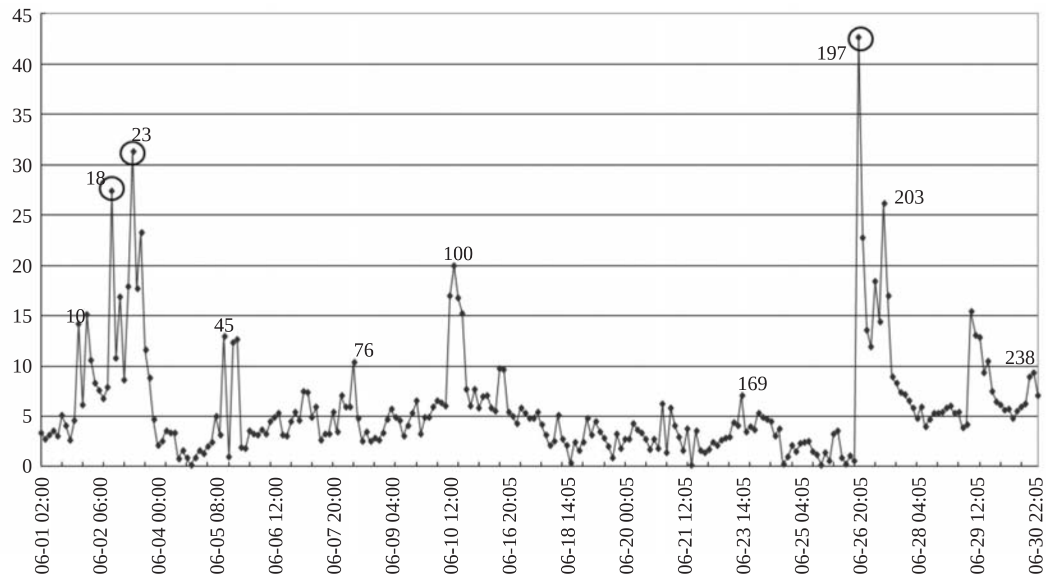

Fig. 1. Plot of COD during 6.1-6.30, 2005.

del (15) is a generalized transfer function model with exogenous X_t variable.

Empirical Experiment

A good example of Z_t, for our model and algorithm, is the chemical oxygen demand (COD). When COD is the variable of major interest or Z_t, one may consider a sudden increase of COD (or outlier of Z_t) as a form of possible waste dumping because a high COD value tends to relate to a serious effluent from the pollution source. In fact, most governments impose strict regulations regarding the maximum COD for water in possible effluent tracts [16]. Based on this, our empirical experiment considers the following episode. For monitoring the water quality of "River A" close to Seoul in Korea, an automated water quality measurement station was installed by the government environment agency, which monitors continuously the related parameters, such as nitrate and COD. One of the main aims of the station is to monitor water quality of the river against possible dumping waste. It is known that a primary source of dumping waste is the industrial firms near the river. From her past experience, the concerned government agency comprehends that rainfall dumping tends to be made quite often. In order to handle this factor in detecting waste dumping into River A, we first apply manual WQMA to the COD values of River A and study how WQMA works before the automated WQMA (a-WQMA) is tested against the data simulated by WDSA. This proves that WQMA is quite efficient in identifying waste dumping into the river.

COD measurements $Z_1,\cdots,Z_{240}$ were made and recorded every two hours during June 1-June 30, 2005. For constructing a properly estimated time series intervention model, the first 200 data $Z_1,\cdots,Z_{200}$ are used for fitting and the remaining 40 data $Z_{201},\cdots,Z_{240}$ are reserved for testing. The time plot for $Z_1,\cdots,Z_{200}$ is given in Fig. 1. For building f_1, it is confirmed that a significant amount of rain fell at t=43, 44, 45, 46, and 49. Also, WQMA manually detects four outliers at t=18, 23, 45, 197 for building f_2 by eye examination. Here, t=45 is detected because it shows an unusually high value during the rain period. Using the 200 training data together with this information, the time series intervention model is fitted to yield

$$\hat{Z}_t = 4.70+0.61Z_{t-1}+0.20Z_{t-2}+17.50I_{18}(t)+14.39I_{23}(t)$$
$$+9.69I_{45}(t)+\frac{40.63}{1-0.51B}I_{197}(t)-4.66I_R(t), \tag{16}$$

where $I_k(t)=1$ if $t=k$ or 0 otherwise, and $I_R(t)=1$ if it rains at t. The indicator variable I's are used for handling the outliers (t=197, 18, 23, and 45, see Fig. 1) and the rain fall effect. Note that $I_R(t)$ and $I_k(t)$ correspond to $I_{1j}(t)$ and $I_{2j}(t)$, respectively. Thus, by using (16), we have

$$\hat{\eta}_t=0.61Z_{t-1}+0.20Z_{t-2},\quad \hat{f}_0(t)=4.70,\quad \hat{f}_1(t)=-4.66I_R(t), \text{ and}$$

$$\hat{f}_2(I_t)=17.50I_{18}(t)+14.39I_{23}(t)+9.69I_{45}(t)+\frac{40.63}{1-0.51B}I_{197}(t). \tag{17}$$

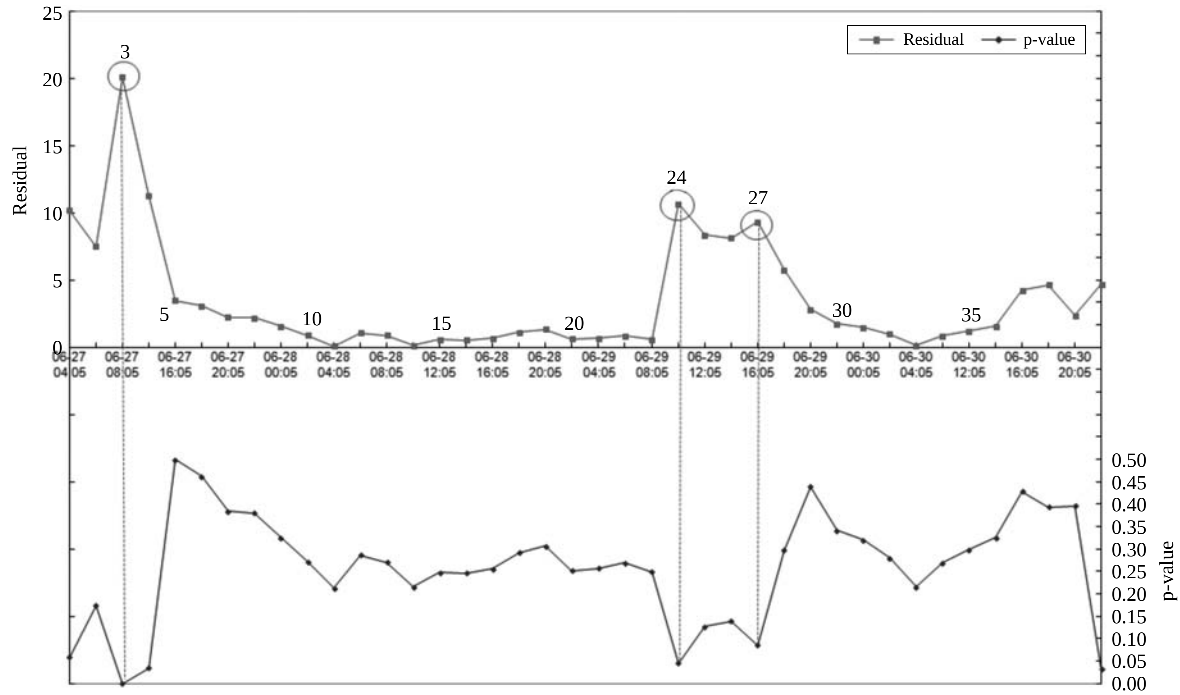

Fig. 2. Plot of residuals and their p-values.

From (17), one may observe the followings: (i) Outlier at $t=197$ has a lasting impact cut by half every two hours (i.e., $\frac{1}{1-0.51B}=1+0.51B+(0.51B)^2+\cdots$), whereas outliers at $t=18$, 23, 45 appear to have an isolated impact (recall it rained on $t=45$). (ii) Rain tends to decrease the COD value significantly (i.e., $-4.66I_R(t)$). (Step 1) and (Step 2) have been implemented now.

For (Step 3), we have

$$\hat{Z}_t^{(N)}=\hat{\eta}_t+\hat{f}_1(I_t)=4.70+0.61Z_{t-1}+0.20Z_{t-2}-4.66I_R(t). \quad (18)$$

Using (18), we can calculate the residuals for $t=1,\cdots,200$ by using the following:

$$e_t=Z_t-\hat{Z}_t^{(N)} \quad (19)$$

and then, a corresponding distribution was estimated by $N(\hat{\mu}_e,\hat{\sigma}_e^2)$. With the estimated $\hat{F}_e$, one may calculate the p-value $p(t)$ for 40 test data, which completes (Step 3) and (Step 4) of WQMA. In fact, the plot for e_t and the corresponding p-value $p(t)$ for the test data is given in Fig. 2. From Fig. 2, one may easily observe that WQMA issues three warnings at $t=201$, 203 and 224 due to their corresponding p-values, which are less than 0.05. While dumping at $t=203$ and 224 could be suggested strongly in terms of COD values, $t=201$ appears to be related to the lasting impact of the previous major dumping made at $t=197$.

In order to evaluate the performance of a-WQMA (and hence WQMA) under simulated situations, we simulate various waste dumping situations via WDSA. Indeed we employ

$$Z_t^{(s)}=f_0^{(s)}(t)+f_1^{(s)}(t)+f_2^{(s)}(t)+\eta_t^{(s)} \quad (20)$$

where $\eta_t^{(s)}=0.61Z_{t-1}^{(s)}+0.20Z_{t-2}^{(s)}+\varepsilon_t^{(s)}$, $\varepsilon_t^{(s)}\sim N(0,1)$, $f_0^{(s)}(t)=-1.7$

$$f_1^{(s)}(t)=v_1(B{:}\mathbf{w}_1,\mathbf{d}_1(X_tI_{1t}))X_tI_{1t}=\sum_{j=1}^{p_1}\frac{0.062X_t}{1-\frac{X_t}{X_t+25}B}I_{1j}(t)$$

and $$f_2^{(s)}(t)=\sum_{j=1}^{p_2}\frac{\beta_3}{1-\beta_3B/10}I_{2j}(t)$$

for selections of $0<X\le 50$ and $0<\beta_0\le 10$. For evaluation purposes, we generate the 300 data $T=\{Z_1^{(s)},\cdots,Z_{300}^{(s)}\}$, which consists of training data $T_1=\{Z_1^{(s)},\cdots,Z_{200}^{(s)}\}$ and testing data $T_2=\{Z_{201}^{(s)},\cdots,Z_{300}^{(s)}\}$, i.e., $T=T_1\cup T_2$. Here, T is designed to contain 10 dumping waste and 10 precipitation time points (i.e., $p_1=p_2=10$ or 10 I_1s and 10 I_2s). Furthermore, T_2 is designed to contain at least 4 cases or time points at which raining and dumping occur simultaneously. Such cases are selected randomly over $t=1,\cdots,300$. In addition, note that the selection of X value (the precipitation amount) for I_1 or the selection of β_3 value

Table 1. Comparison of average detection rates among three methods

Simulated situation I	Algorithm	Detection rate DR_I
Dumping with rain	a-WQMA	63.6%
	Fixed boundary=8.4	35.0%
	Upper 5 percentile	35.7%
Dumping without rain	a-WQMA	71.6%
	Fixed boundary=8.4	64.8%
	Upper 5 percentile	66.2%
No dumping	a-WQMA	1.4%
	Fixed boundary=8.4	3.6%
	Upper 5 percentile	3.9%

for I_2 is made randomly from the given ranges in each case.

As competitors with a-WQMA, two simple algorithms for waste dumping detection are considered, i.e., upper 5th percentile method and fixed boundary method. The upper 5th percentile method issues a warning when $Z_t^{(s)}$ for t=201,···,300 (COD value in test data T_2) exceeds the upper 5th percentile of $T_1=\{Z_1^{(s)},\cdots,Z_{200}^{(s)}\}$ (COD values in training data) while the fixed boundary method issues a warning when $Z_t^{(s)}$ for t=201, ···,300 exceeds the fixed bound 8.4. For comparison of the three methods, we consider 3 simulated situations, i.e., dumping with rain, dumping without rain and no dumping. Recall that each dumping or rain has a different impact on COD values afterwards due to the dependence on its own X and β_3. Also, recall that our simulation produces 10 t's of dumping and 10 t's of rain with at least 4 t's of dumping rains inside T_2. For a quantitative performance evaluation of a-WQMA, 110 realizations of (20) are performed in order to obtain the detection rates by each method for the three possible situations. Here, the detection rate is formally given by

$$DR_I = (\text{the number of detections made for the situation } I \text{ on } 201 \le t \le 300) \,/\, (\text{the number of } t\text{'s belonging to the situation } I \text{ on } 201 \le t \le 300).$$

For instance, note that the denominator of DR_I is between 4 and 10 if the situation I is dumping with rain.

In Table 1, the experiment results are summarized. It is worth mentioning that across all situations a-WQMA excels other competitors uniformly. For dumping with rain situation it dominates the others outstandingly. For dumping without rain, a similar notice can be made through a-WQMA dominance, which is not as strong as the dumping with rain situation. Note that for the above two situations, a higher detection rate means better performance. For no dumping situation, a-WQMA excels the other competitors. Note here that for the no dumping situation, the lower detection rate means better performance.

References

1. Abraham B, Ledolter J. Statistical methods for forecasting. New York: John Wiley & Sons, Inc; 2005.
2. Martin RD. Robust estimation of autoregressive models. In: Brillinger DR, Tiao GC, editors. Directions in time series. Hayward (CA): Institute of Mathematical Statistics, 1980. p. 228-254.
3. Wagner R, Mattrow H, Ritz G, Smith B. Guidelines and standard procedures for continuous water-quality monitors: site selection. Field operation, calibration, record computation and reporting. Water resources investigation report 00-4252, Reston: U.S. Geological Survey; 2000.
4. http://www.evrotas.gr.
5. Kim G, Lee K, Lee J, Yi M. Case study for determination of a water level monitoring frequency for nationwide groundwater monitoring networks in Korea. J Hydrol 2007;342:223-237.
6. Vaseashta A, Vaclavikova M, Vaseashta S, Gallios G, Roy P, Pummakarnchana O. Nanostructures in environmental pollution detection, monitoring, and remediation. Sci Technol Adv Mat 2007;8:47-59.
7. Mariolakos I, Fountoulis I, Spyridonos E, Dritasa C, Kapourani E, Andreadakis E. Holistic methodology for water resources management in semi-acid regions. Case study in Mani (S Peloponnesus, Greece). Proceedings of the 10th Biennial Symposium on Artificial Recharge of Ground Water; 1-40. Arizona; 2001.
8. Luo Q, Cai ZW, Wong MH. Polybrominated diphenylethers in fish and sediment from river polluted by electronic waste. Sci Total Environ 2007;383:115-127.
9. Kobr M, Linhar I. Geophysical survey as a basis for regeneration of waste dump Halde 10, Zwickau, Saxony. J Appl Geophys 1994;31:107-116.
10. Novotny EV, Stefan HG. Stream flow in Minnesota: indicator of climate change. J Hydrol 2007;334:319-333.
11. Mariolakos I, Fountoulis I, Andreadakis E, Kapourani E. Real-time monitoring on Evrotas River (Laconia, Greece): dissolved oxygen as a critical parameter for environmental status classification and warning. Desalination 2007;213:72-80.
12. Huck PM, Farquhar GJ. Water quality models using the Box-Jenkins method, J Env Eng Div ASCE 1974;100:733-751.
13. Jayawardena A.W, Lai F. Time series analysis of water quality data in pearl river, China. J Env Eng Div ASCE 1989;115:590-607.
14. Chapra S. Surface water quality modeling. New York: Mcgraw-

Hil; 1997.
15. Song GM, Moon JE, Park CY. Realization of an outlier detection algorithm using R. J Korean Data Inf Sci Soc 2011;22: 449-458. In Korean.
16. Sawyer CN, McCarty PL, Parkin GF. Chemistry for environmental engineering and science. 5th ed. New York: McGraw-Hill; 2003.