Ann. Limnol. - Int. J. Lim. 52 (2016) 137–150
DOI: 10.1051/limn/2016003

Available online at:
www.limnology-journal.org

A multiple process univariate model for the prediction of chlorophyll-a concentration in river systems

Sanghyun Kim*

Department of Environmental Engineering, Pusan National University, Busandaehak-ro 63 beon-gil, Geumjung-gu, Busan 46241, Korea

Received 23 September 2015; Accepted 4 January 2016

Abstract – The concentration of chlorophyll-a (Chl-a) in river systems is dependent on various hydrometric and biochemical factors, including an intricate array of corresponding growth and extinction mechanisms. This complex and interactive assortment of factors makes prediction of algal blooms difficult. This paper introduces an innovative time-series model structure that predicts Chl-a concentration in inland waters. To improve the prediction accuracy of existing models, we assume that the predicting variable is determined by multiple and independent drivers. An enhanced stochastic model, namely a multiple process univariate model (MPUM), is developed to address the impacts of the distinct mechanisms associated with each regulator (*e.g.*, hydrometeorological factors and anthropogenic activities). Observations of the algae concentration at 16 weirs along four major river systems in South Korea are used to model Chl-a concentration. Comparisons between traditional models and the proposed method demonstrate the strengths of the MPUM, both in making predictions and in the parsimony of the model structure. The robustness of the developed model was further validated by modeling algae concentration before and after river-flow regulation procedures.

Key words: Water quality / algae prediction / chlorophyll-a concentration / multiple process model

Introduction

Since "the Four Rivers Restoration Project of Korea" was completed in 2011, major river systems in South Korea have often suffered from algal bloom (Park, 2012). Chlorophyll-a (Chl-a) concentrations are used to estimate algal biomass in coastal and water systems (Padisák, 2004). Algal blooms lead to the deterioration of fish habitats by increasing turbidity and, also cause the decline of specific aquatic species, such as invertebrates. High Chl-a concentrations often result in unpleasant tastes and odor problems in drinking water, as well as, clogged filters in water treatment plants. Some species of cyanobacteria produce toxins (cynatoxins) that cause significant ecological and human health concerns (WHO, 2003). Prediction of algal bloom occurrences is instrumental in the implementation of appropriate management practices, such as low flow augmentation for river systems.

Variations in algae concentrations have been simulated using process-based mathematical models (Thomann and Mueller, 1987; Whitehead *et al.*, 1997; U. S. Environmental Protection Agency, 2002, 2015; Wu and Xu, 2011). However, the processes associated with algal blooms are extremely complicated, and the prediction of algae concentrations that are made using ecological and water quality models are often limited. This is mainly due to the uncertainty associated with the kinetic coefficients and the structural complexity of two- and three-dimensional (3D) models which require a substantial amount of input data and demanding calibration and validation procedures (U. S. Environmental Protection Agency, 2015). Alternatively, predictions of algae concentrations have been explored using several statistical methods or artificial neural networks (Lee *et al.*, 2003; Marsili-Libelli, 2004; Hamilton *et al.*, 2009; Cha *et al.*, 2014; Coad *et al.*, 2014; Chen *et al.*, 2015; Muttil and Lee, 2005; Oh *et al.*, 2007; Malek *et al.*, 2011). Depending on data availability, meta-heuristic approaches do not always produce consistent results, which may indicate a violation of the stationary process assumption in time series modeling. In general, statistical models yield distinct results for different locations, even within identical watersheds. The phenomenon is related to the large number of model parameters associated with black box modeling.

*Corresponding author: kimsangh@pusan.ac.kr

Transfer function approaches were used to predict biochemical oxygen demand (BOD) (Novotny and Olem, 1994) in a combined sewer system and to model $\delta^{18}O$ in hillslope runoff responses (Weiler *et al.*, 2003; Iorgulescu *et al.*, 2007) using hydrological components (*e.g.*, rainfall and groundwater) as the input. When used to predict Chl-a concentrations, an Autoregressive Moving Integrated Moving Average (ARIMA) model showed the potential and feasibility of using a simple univariate model (Chen *et al.*, 2015).

In this study, an alternative approach to the existing methods is proposed for the prediction of Chl-a concentrations in river systems. Based on an autoregressive moving average process (Box and Jenkins, 1976; Salas *et al.*, 1988), we developed an innovative and feasible modeling structure, the multiple process univariate model (MPUM), that considers the multiple and mutually-independent processes that play a role in the temporal variation of algae concentration. In order to provide a deterministic basis for the stochastic structure associated with the time-series model for temporal changes in Chl-a concentrations, we present a mathematical derivation for the process-based variations in algae concentration. Seasonal variations in Chl-a concentration have been noted in various inland water systems (Mallin *et al.*, 1991; Sin *et al.*, 2000; French and Petticrew, 2007). In order to evaluate the seasonal influences on the variations in algae concentration within an existing modeling platform, we used the seasonal autoregressive integrated moving average (SARIMA) model. The proposed method and existing approaches were used to model Chl-a concentrations over a 2-year period at 16 locations along four river systems in South Korea. The validity of the developed model is further investigated through: (1) an extended modeling process that used data collected over an 8-year-time period; and (2) a comparison of our model results with results from more traditional approach. The strengths of MPUM are assessed in the context of model performance and parsimony of the model structure, as well as the feasibility and effectiveness of parameter evaluation.

Materials and methods

Study area and data

We selected 16 points (weirs) in four major river systems (Han River, Geum River, Nakdong River and Yeoungsan River) in South Korea to model Chl-a concentration (see Fig. 1). The weirs were created through "the Four Rivers Restoration Project of Korea", which was designed to relieve water-related problems. This project was funded through investments from multiple government ministries between 2010 and 2011. The Korean government estimated that approximately one billion tons of water would be needed to prevent water scarcity in 2016 (Woo, 2009; Jun and Kim, 2011). In addition to managing water, this project also focused

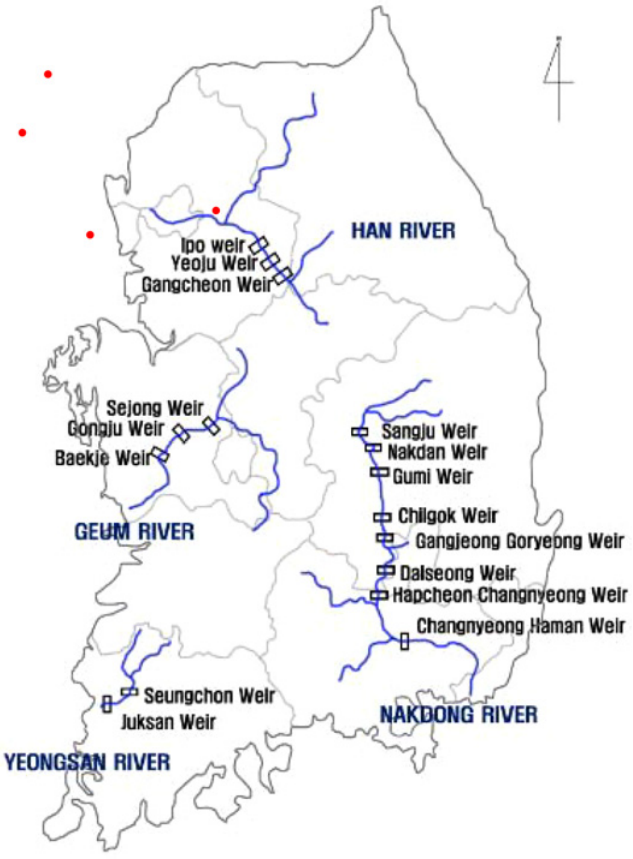

Fig. 1. Chl-a modeling locations along four major rivers in South Korea.

on improving water quality by managing pollutants that could result in eutrophication of the waterways (*e.g.*, chemical oxygen demand (COD) and total phosphorus (TP)). However, sections of the river just below some of the weirs (see Fig. 1) were converted into standing water, which was responsible for the development of algal blooms (Park, 2012). The Ministry of Environment collected data on several indicators of water quality at a weekly interval (*e.g.*, COD, BOD, total suspended solid, TP, total nitrogen and Chl-a concentrations) between 2011 and 2014. Nutrients at all points were high and concentrations of TP and total nitrogen did not control the algal blooming process (Kim *et al.*, 2007). Actually, nitrogen and phosphorus balance per hectare of agricultural land in Korea were three to four times the averages from OECD countries (OECD, 2013). The weekly concentrations of Chl-a (as determined by the Ministry of Environment) between 2012 and 2013 are used in this study.

Autoregressive moving average and seasonal autoregressive integrated modeling average models

One advantage of the univariate autoregressive moving average model (ARMA) is that it is possible to fit a model with a relatively small number of parameters ($p + q$). The ARMA(p,q) model, where p and q are orders of

autoregressive and moving average operators, respectively, can be expressed as (Box and Jenkins, 1976; Salas *et al.*, 1988),

$$Z_t = \phi_1 Z_{t-1} + \cdots + \Phi_p Z_{t-p} + e_t - \theta_1 e_{t-1} - \cdots \theta_q e_{t-q} \quad (1)$$

where Z_t is the modeling variable at time step t and e_t is an independent random variable.

Using the backward operator B, the ARMA(p, q) model can be expressed as

$$\Phi(B) Z_t = \Theta(B) e_t \quad (2)$$

where

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \text{ and}$$
$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$
$$(e.g.,\ B^p Z_t = Z_{t-p}, B^q e = e_{t-q}).$$

A deterministic justification of ARMA models for time series modeling of Chl-a concentrations is presented in the Appendix.

The variations in Chl-a concentration tends to exhibit a substantial seasonal effect. The incorporation of seasonal differencing or appropriate differencing into the model structure is accomplished by the following general multiplicative model, called the seasonal autoregressive integrated moving average (SARIMA) model (Salas *et al.*, 1988):

$$\Phi_p(B)\Phi_p(1-B)^{\mathrm{d}}(1-B^{\mathrm{s}})^{\mathrm{D}} Z_t = \Theta_q(B)\Theta_Q(B^{\mathrm{s}}) e_t \quad (3)$$

where superscript s indicates the order of the backward operator for seasonal differencing. The model in equation (9) can be denoted as the SARIMA $(pdq) \times (PDQ)_{\mathrm{s}}$ model.

Multiple process univariate model

The variations in Chl-a concentration in river systems are the result of multiple processes, which are controlled by hydro-meteorological, bio-chemical and anthropogenic drivers. The temporal aspects of these various factors cause distinctive patterns in Chl-a variations. This means that the stochastic process for predicting Chl-a concentration is not required to be restricted to traditional single process-based approaches, such as ARMA or SARIMA. Therefore, we can relax this underlying assumption of time series modeling, in order to address the complicated processes affecting Chl-a concentration. In fact, one common feature in natural systems is the sinusoidal variation in an annual period. For example, variations in rainfall, runoff and temperature have a strong annual periodicity and their corresponding human responses, such as the operation of dams for drought or flood control are also partially related to extreme hydro-meteorological drivers (*e.g.*, storms and typhoons). Furthermore, seasonal production pulses of nitrogen loading (Mallin *et al.*, 1991), seasonality in pH and TP concentration (French and Petticrew, 2007) are also associated with the periodicity of the nature system.

The variation of Chl-a concentration can be expressed as a linear superposition of factors based on multiple processes as follows:

$$Z_t = Z_t^1 + Z_{t-1}^1 + \cdots + Z_t^n + \cdots + Z_{t-m}^n \quad (4)$$

where the subscripts t and $t-1$,..., $t-m$ represent the corresponding time steps, the numbers in superscript indicate the corresponding stochastic processes, and the subscript n denotes the number of processes.

The stochastic structures introduced in equation (4) are based on two hypotheses. Firstly, the concentration of Chl-a can be expressed as a summation of multiple stochastic processes. This assumption seems acceptable when we consider that the occurrence of algal blooms is frequently associated with low flow conditions in river systems and, the availability of sufficient nutrients, and that these factors are completely independent of each other. Secondly, no causal link exists between each stochastic process, and all processes are associated with mutually exclusive drivers such as rainfall, and fertilizer and pesticide applications.

The MPUM based on equation (4) can be expressed as follows:

$$Z_t = \varphi_{1,1} Z_t^1 + \varphi_{2,1} Z_{t-1}^1 + \cdots + \varphi_{1,2} Z_t^2 + \varphi_{2,2} Z_{t-1}^2 + \cdots + \theta_{1,1} e_t^1 + \theta_{2.1} e_{t-1}^1 + \cdots + \theta_{1,2} e_t^2 + \theta_{2,2} e_{t-1}^2 + \cdots \quad (5)$$

where $\varphi_{1,1}$and $\varphi_{1,2}$ are autoregressive coefficients for the first and second stochastic processes at the current time step, respectively and $\varphi_{2,1}$ and $\varphi_{2,2}$ are autoregressive coefficients for the first and second stochastic processes at the time step $t-1$, respectively. Furthermore, $\theta_{1,1}$ and $\theta_{1,2}$ are moving average coefficients for the first and second stochastic processes at the current time step, respectively, and $\theta_{2,1}$ and $\theta_{2,2}$ are moving average coefficients for the first and second stochastic processes at the time step $t-1$, respectively. We assume that the random processes, $e_t^1, e_{t-1}^1, \ldots$ and $e_t^2, e_{t-1}^2, \ldots$ are independent of each other.

Procedure and model developments

Development procedure of univariate model

The modeling of all components in equation (5) also follows four distinct procedures (Box and Jenkins, 1976; Salas *et al.*, 1988) (*i.e.*, statistical analysis and pretreatment, identification of model structure, estimation of parameters and diagnostic checking). If the delineated model structure is found to be inappropriate in the checking procedure, the modeling is restarted from the second step (model identification) to determine a better model structure. Figure 2 presents a flowchart for the time-series model development process.

A preliminary evaluation of the Chl-a concentration statistics indicates that a proper transformation is necessary to improve normality and stationary. A Box–Cox transformation, $y = x^{\lambda} - 1/\lambda$, (Box and Cox, 1964), is

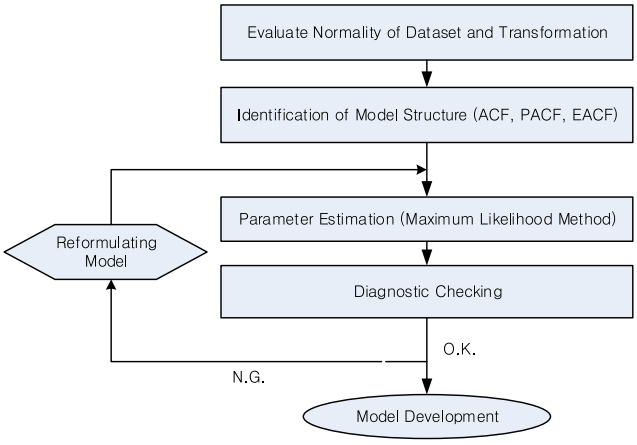

Fig. 2. Flowchart for the time-series model.

introduced to improve the normality of the data. The standardization of the transformed time series resulted in an appropriate data distribution for time series modeling. Table 1 presents Chl-a concentration statistics for four of the study plots. The skewness and kurtosis of the original data were improved through the transformation and the mean and standard deviation were transformed into 0 and 1 for the four points. The statistics of the other 12 points were similar to those shown in Table 1.

An inspection of the time series plots for each Chl-a concentration indicates that as a whole, the data exhibits no notable trends, but the annual seasonality is significant. The autocorrelation function (ACF) and partial ACF (PACF) function of the data present temporal correlation structures as shown in Figure 3. The ACF of the corrected Chl-a concentration for the Ipo and Sejong weirs show that autocorrelations were significant through lags three to five, after which the functions tailed off (see Fig. 3(a)). The PACF revealed significant responses for only lag one (Fig. 3(b)). Candidate model structures for the prediction of Chl-a concentration can also be selected through the evaluation of the corner table (Liu and Hanssens, 1982). Depending on how the matrix elements are distributed in the estimated corner table, candidate models can be developed through further analysis. The corner tables appearing in Figures 3(c) and (d) corresponding to ARMA(1,0) and ARMA(1,1) are candidate models for Chl-a concentrations at the Ipo and Sejong weirs, respectively. Potential model structures for the other 14 points in the four rivers were also determined based on evaluated ACFs, PACFs and corner tables.

In order to estimate the model parameters, maximum likelihood estimates (Box and Jenkins, 1976) and conditional likelihood estimates were used to determine the sum of the square surface for a range of parameters and the location of the minimum of the sum of square in space. Model estimates were calculated simultaneously for the autoregressive and moving average parameters and the standard deviation of the residuals. Model estimates were calculated using the time series modeling procedure of the Scientific Computing Associate (SCA) package (Liu, 2006).

To assure that the assumptions of the modeling process are correct, model adequacy can be checked through a series of diagnostic tests, "Over-fitting" is one such method used to determine if a model structure is optimal. In order to check if additional parameters can improve model performance, progressively more complicated models are fit to the Chl-a concentration time series. Results were obtained for multiple Chl-a concentration time series models that were tested in the over-fitting process. Based on the initial estimations of the parameters for several candidate models, both the statistical index and physical inference are needed to determine whether a model with estimated parameters appropriately represents the Chl-a concentration variation of a certain point. The Student's t statistics indicate the significance of the parameter, provided that doing so did not influence the estimates of other parameters. If the t statistics were within an absolute value of 2, then the null hypothesis of the Student's t test was not rejected at the 5% significance level. The continuity of the parameter structure in the time series model is another important criterion for explaining the physical aspects of Chl-a concentration. If one of the parameters for an intermediate time step of the autoregressive or moving average term is negligible, or the t statistics of the parameter is very small, then the corresponding model structure can be said

Table. 1. Statistics associated with the original, transformed, and standardized Chl-a concentrations at four points along the four major rivers.

Points		Data	Mean	S.D.	Skewness	Kurtosis
Han River	Ipo	Original	9.61	12.31	2.07	3.17
		Transformed	1.34	0.75	0.05	− 0.86
		Standardized	0.00	1.00	0.05	− 0.86
Geum River	Sejong	Original	22.75	22.98	1.54	1.83
		Transformed	2.28	0.78	− 0.10	− 0.91
		Standardized	0.00	1.00	− 0.10	− 0.91
Yeongsan River	Seungchon	Original	47.03	35.15	0.66	− 0.11
		Transformed	10.58	5.50	− 0.03	− 1.02
		Standardized	0.00	1.00	− 0.03	− 1.02
Nakdong River	Changnyeong Haman	Original	34.68	30.40	1.59	1.91
		Transformed	3.83	1.17	0.10	− 0.29
		Standardized	0.00	1.00	0.10	− 0.29

to be unsuitable for explaining the variations in Chl-a concentration. For example, at time step $t-2$, Φ_{t-2}, the Chl-a concentration could not exert a larger impact on the current Chl-a concentration, Z_t, than that from the time step $t-1$, Φ_{t-1}.

If multiple candidates were obtained through this procedure, a parsimony test was applied in order to select the final model. Akaike's information criterion (AIC) was used to find a balance between the variance of the residuals and the number of autoregressive and moving average parameters (Akaike, 1974), which is defined as

$$\mathrm{AIC}(p,q) = N \cdot \ln(\sigma_\varepsilon^2) + 2 \cdot (p+q) \quad (6)$$

where N is the sample size, σ_ε^2 is the maximum likelihood estimate of the variance of the residuals, and p and q are the orders of the autoregressive and moving average parameters, respectively.

Table 2 presents the parameters, residual standard deviation estimates and t statistics of the selected models for four points in the four major rivers. All of the models (including the other 12 points) satisfied the condition of model stability (Box and Jenkins, 1976). Diagnostic checks on the model residuals were performed to test whiteness and normality. The ACF of the residual series provided correlations between individual residuals. Figure 4 presents the ACF and PACF of the residuals for points at

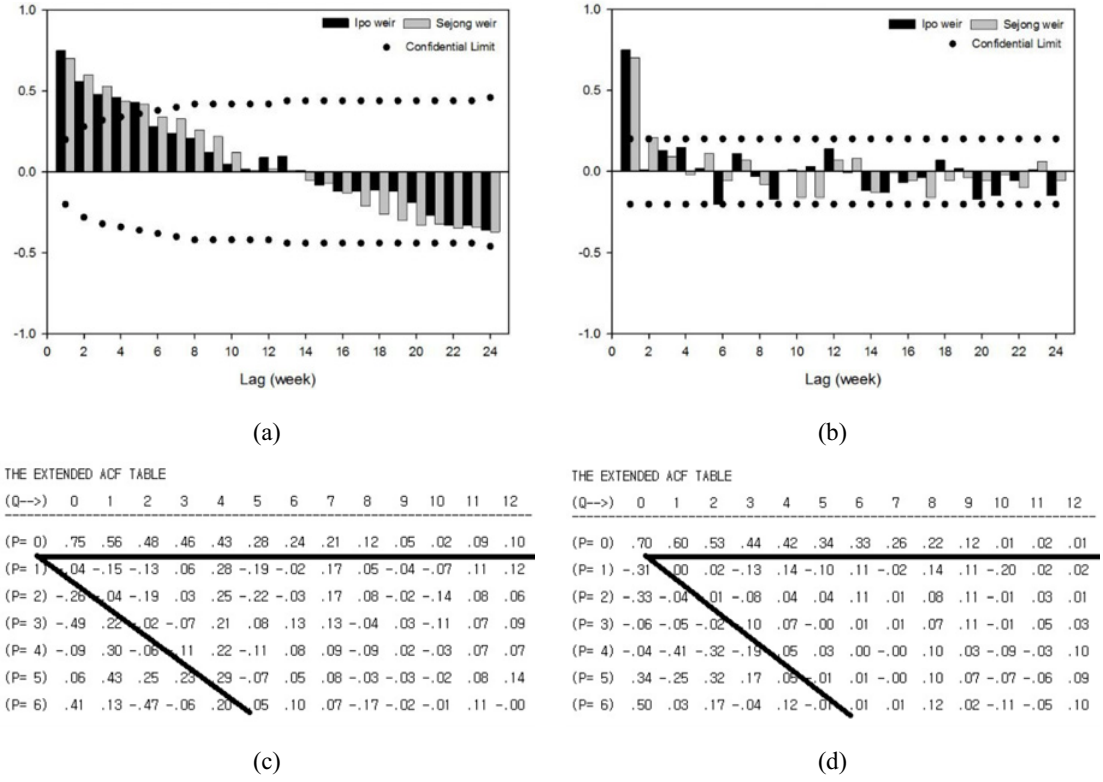

Fig. 3. ACF, PACF and EACF for Chl-a concentrations at the Ipo and Sejong weirs.

Table 2. Estimated model parameters for the prediction of Chl-a concentration for the four points in Table 1 using the ARMA model.

Points		p, q order	Statistics	φ_1	θ_1
Han River	Ipo	(1, 0)	Value	0.7863	–
			SE	0.0781	
			t-value	10.07	
Geum River	Sejong	(1, 1)	Value	0.8758	0.3578
			SE	0.0624	0.1202
			t-value	14.04	2.98
Yeongsan River	Seungchon	(1, 0)	Value	0.6949	–
			SE	0.1285	
			t-value	5.41	
Nakdong River	Changnyeong Haman	(1, 1)	Value	0.8127	0.3828
			SE	0.0864	0.1379
			t-value	9.40	2.78

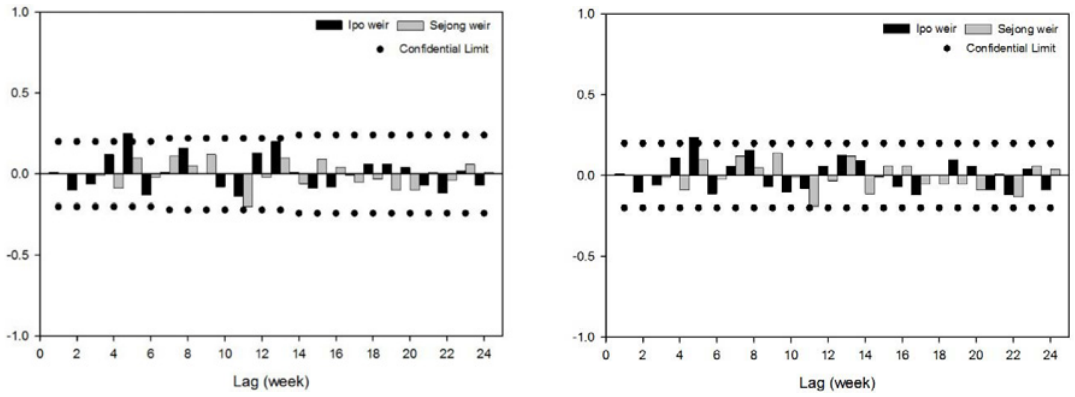

Fig. 4. The ACF (a) and PACF (b) of the ARMA model residuals at the Ipo and Sejong weirs.

the Ipo and Sejeong weirs, which were not significant to confidence intervals. The plots of the ACF and PACF of the residuals did not show any significant stochastic structure over the confidence intervals. Therefore, the residual diagnostics checks indicated that the selected models appropriately represent the stochastic structures of a time series.

SARIMA Modeling

The seasonality in Chl-a concentration can be incorporated with the SARIMA model using equation (3). In order to create the most appropriate model, multiple SARIMA candidate models were considered. The weekly data on Chl-a concentration indicates that the annual periodicity should be addressed using a 52 differencing structure. Analytical derivations using various assumptions about the models yield the following SARIMA models:

Model $(1, 0, 1) \times (1, 0, 1)_{52}$ as SARIMA1

$$(1 - \phi_1 B - \phi_{52} B^{52} - \phi_{53} B^{53}) Z_t = (1 - \theta_1 B - \theta_{52} B^{52} - \theta_{53} B^{53}) e_t \quad (7)$$

Model $(1, 0, 2) \times (1, 0, 1)_{52}$ as SARIMA2

$$(1 - \phi_1 B - \phi_{52} B^{52} - \phi_{53} B^{53}) Z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_{52} B^{52} - \theta_{53} B^{53} - \theta_{54} B^{54}) e_t \quad (8)$$

Model $(2, 0, 1) \ (1, 0, 1)_{52}$ as SARIMA3

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_{52} B^{52} - \phi_{53} B^{53} - \phi_{54} B^{54}) Z_t = (1 - \theta_1 B - \theta_{52} B^{52} - \theta_{53} B^{53}) e_t \quad (9)$$

Model $(1, 1, 2) \ (1, 0, 1)_{52}$ as SARIMA4

$$(1 - 2\phi_1 B - \phi_2 B^2 - \phi_{52} B^{52} - 2\phi_{53} B^{53} - \phi_{54} B^{54}) Z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_{52} B^{52} - \theta_{53} B^{53} - \theta_{54} B^{54}) e_t \quad (10)$$

Model $(2, 1, 1) \times (1, 0, 1)_{52}$ as SARIMA5

$$(1 - 2\phi B - \phi_3 B^3 - \phi_{52} B^{52} - 2\phi_{53} B^{53} - \phi_{55} B^{55}) Z_t = (1 - \theta_1 B - \theta_{52} B^{52} - \theta_{53} B^{53}) e_t \quad (11)$$

Model $(1, 0, 0) \times (1, 0, 1)_{52}$ as SARIMA6

$$(1 - \phi_1 B - \phi_{52} B^{52} - \phi_{53} B^{53}) Z_t = (1 - \theta_{52} B^{52}) e_t \quad (12)$$

Model $(2, 0, 0) \times (1, 0, 1)_{52}$ as SARIMA7

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_{52} B^{52} - \phi_{53} B^{53} - \phi_{54} B^{54}) Z_t = (1 - \theta_{52} B^{52}) e_t \quad (13)$$

Model $(1,1,1) \times (1, 0, 1)_{52}$ as SARIMA8

$$(1 - 2{*}\phi_1 {*} B - \phi_2 {*} B^2 - \phi_{52} {*} B^{52} - 2{*}\phi_{53} {*} B^{53} - \phi_{54} {*} B^{54}) Z_t = (1 - \theta_1 {*} B - \theta_{52} {*} B^{52} - \theta_{53} {*} B^{53}) e_t \quad (14)$$

The analytical expression for SARIMA8 is recognized as an identical model to SARIMA3, despite the fact that the parameters Φ_1 and Φ_{53} differ by a multiple of two.

Identical parameter estimation and diagnostic checking procedure were applied to the SARIMA modeling of Chl-a concentrations. Table 3 illustrates the model identification processes for the seven distinct models appearing in equations (7)–(13). As presented in Table 3, SARIMA1, SARIMA2, SARIMA5 and SARIMA6 were converted into ARMA(1,0) as the insignificant parameters (having small *t*-statistics) were eliminated. A similar model identification process was applied to all other modeling points.

MPUM Modeling procedure

In order to incorporate all relevant processes into modeling of the weekly concentration of Chl-a, we presumed that two distinct processes affect the growth and fate of algae. The first (process one) is hydrometeological processes (*e.g.*, rainfall, stream flow and temperature), which display a strong annual periodicity. The second (process two) is random anthropogenic effects (*e.g.*, the application of fertilizers or pesticides). Assuming that seasonality can be approximated using a sinusoidal function, the time series for Chl-a concentration can be decomposed into two components. Firstly, the periodic component (process one), which is approximated through the optimization of constant, amplitude, phase, and frequency

Table 3. Estimated model parameters for the prediction of Chl-a concentrations using SARIMA models for the Ipo weir; numbers in red represent insignificant parameters while numbers in blue are the parameters for the final models.

Points		$(p, d, q) \times (P,D,Q)_{52}$	Statistics	φ_1	φ_2	φ_3	φ_{52}	φ_{53}	φ_{54}	φ_{55}	θ_1	θ_2	θ_{52}	θ_{53}	θ_{54}
Han River	Ipo	(1, 0, 1) ×	Value	0.7802	–	–	0.1061	− 0.0890	–	–	0.0483	–	0.0241	0.0079	–
		$(1, 0, 1)_{52}$	SE	0.1283			0.0961	0.0966			0.1912		0.1071	0.1372	
			t-value	6.08			1.10	− 0.92			0.25		0.22	0.06	
			Value	0.7736	–	–	0.0445	–	–	–	–	–	–	–	–
			SE	0.1228			0.0712								
			t-value	6.30			0.62								
			Value	0.7863	–	–	–	–	–	–	–	–	–	–	–
			SE	0.0781											
			t-value	10.07											
		(1, 0, 2) ×	Value	0.7739	–	–	0.1092	− 0.1019	–	–	0.0568	− 0.0535	0.0142	0.0154	0.0116
		$(1, 0, 1)_{52}$	SE	0.1720			0.0944	0.0986			0.2130	0.1738	0.1314	0.1287	0.1331
			t-value	4.50			1.16	− 1.03			0.27	− 0.31	0.11	0.12	0.09
			Value	0.7779	–	–	0.1049	− 0.0867	–	–	–	–	–	–	–
			SE	0.1258			0.0962	0.0964							
			t-value	6.18			1.09	− 0.90							
			Value	0.7736	–	–	0.0445	–	–	–	–	–	–	–	–
			SE	0.1228			0.0712								
			t-value	6.30			0.62								
			Value	0.7863	–	–	–	–	–	–	–	–	–	–	–
			SE	0.0781											
			t-value	10.07											
		(2, 0, 1) ×	Value	0.7297	0.0353	–	0.1087	− 0.1806	0.1175	–	− 0.0028	–	0.0275	0.0114	–
		$(1, 0, 1)_{52}$	SE	0.6850	0.5230		0.0972	0.1313	0.10581.11		0.6863		0.1160	0.1410	
			t-value	1.07	0.07		1.12	− 1.37			− 0.01		0.24	0.08	
			Value	0.7798	–	–	0.1071	− 0.1845	0.1200	–	–	–	–	–	–
			SE	0.1266			0.0957	0.1191	0.0912						
			t-value	6.16			1.12	− 1.55	1.31						
		(1, 1, 2) ×	Value	0.5917	− 0.4149	–	0.1029	− 0.1126	0.1936	–	0.5223	− 0.3090	0.0039	− 0.0073	− 0.0020
		$(1, 0, 1)_{52}$	SE	0.1681	0.2797		0.0911	0.0624	0.0858		0.3378	0.1732	0.1413	0.1586	0.1414
			t-value	3.52	− 1.48		1.13	− 1.81	2.26		1.55	− 1.78	0.03	− 0.05	− 0.01
			Value	0.6086	− 0.4434	–	0.1007	− 0.1117	0.1943	–	0.5575	− 0.3167	–	–	–
			SE	0.1616	0.2684		0.0907	0.0627	0.0853		0.3239	0.1716			
			t-value	3.77	− 1.65		1.11	− 1.78	2.28		1.72	− 1.85			
		(2, 1, 1) ×	Value	0.4611	–	− 0.1294	0.0737	− 0.0548	–	0.0938	0.2156	–	0.0182	0.0000	–
		$(1, 0, 1)_{52}$	SE	0.0900		0.1399	0.0984	0.0516		0.0718	0.2394		0.1271	0.1428	
			t-value	5.13		− 0.92	0.75	− 1.06		1.31	0.90		0.14	0.00	
			Value	0.4096	–	–	–	− 0.0330	–	0.0889	–	–	–	–	–
			SE	0.0592				0.0393		0.0729					
			t-value	6.92				− 0.84		1.22					
			Value	0.3951	–	–	–	–	–	0.0567	–	–	–	–	–
			SE	0.0588						0.0655					
			t-value	6.72						0.87					
			Value	0.3932	–	–	–	–	–	–	–	–	–	–	–
			SE	0.0391											
			t-value	10.07											

Table 3. (*Contd.*)

Points	$(p, d, q) \times (P,D,Q)_{52}$	Statistics	φ_1	φ_2	φ_3	φ_{52}	φ_{53}	φ_{54}	φ_{55}	θ_1	θ_2	θ_{52}	θ_{53}	θ_{54}
	$(1, 0, 0) \times$	Value	0.7577	–	–	0.1114	− 0.0888	–	–	–	–	0.0364	–	–
	$(1, 0, 1)_{52}$	SE	0.0996			0.0962	0.0927					0.0990		
		t-value	7.61			1.16	− 0.96					0.37		
		Value	0.7736	–	–	0.0445	–	–	–	–	–	–	–	–
		SE	0.1228			0.0712								
		t-value	6.30			0.62								
		Value	0.7863	–	–	–	–	–	–	–	–	–	–	–
		SE	0.0781											
		t-value	10.07											
	$(2, 0, 0) \times$	Value	0.7222	0.0478	–	0.1082	− 0.1787	0.1124	–	–	–	0.0408	–	–
	$(1, 0, 1)_{52}$	SE	0.1413	0.1453		0.0956	0.1141	0.0929				0.1000		
		t-value	5.11	0.33		1.13	− 1.57	1.21				0.41		
		Value	0.7798	–	–	0.1071	− 0.1845	0.1200	–	–	–	–	–	–
		SE	0.1266			0.0957	0.1191	0.0912						
		t-value	6.16			1.12	− 1.55	1.31						
	$(1, 1, 1) \times$	Value	0.3649	0.0353	–	0.1087	− 0.0903	0.1175	–	− 0.0028	–	0.0275	0.0114	–
	$(1, 0, 1)_{52}$	SE	0.3425	0.5230		0.0972	0.0657	0.1058		0.6863		0.1160	0.1410	
		t-value	1.07	0.07		1.12	− 1.37	1.11		− 0.01		0.24	0.08	
		Value	0.3899	–	–	0.1071	− 0.0923	0.1200	–	–	–	–	–	–
		SE	0.0633			0.0957	0.0596	0.0912						
		t-value	6.16			1.12	− 1.55	1.31						

parameters of the sinusoidal function using an objective function that minimizes the root mean square error (RMSE) between approximations and the original series. Secondly, the residual time series component (process two), obtained from the differences between the original time series and the periodic component.

Figures 5 and 6 show the ACF, partial ACF and extended ACF (EACF) of Chl-a concentrations for the Ipo and Sejong weirs for processes one and two, respectively.

The ACFs of process one for the Ipo and Sejong weirs, which are extended to infinite, exhibit damped waves. The PACFs range between 1 and 2 (see Figs. 5(a) and (b)), which indicates that the AR(2) process is appropriate for process one. In fact, the EACFs presented in Figures 5(c) and (d) also indicate that the AR(2) model has an appropriate structure for explaining the stochastic features for process one. As shown in Figures 6(a) and (b), the ACFs and PACFs of process two for the Ipo and Sejong weirs are extend to infinity and exhibit irregular damped waves, indicating that the ARMA process constitutes a potential stochastic model (Salas *et al.*, 1988). The EACFs for process two are presented in Figures 6(c) and (d), and indicate that the ARMA(1,1) model is a suitable candidate structure for process two. Other modeling procedures, such as the parameter estimation and diagnostic checking procedure, are similar to those for the ARMA or SARIMA models.

Table 4 presents the model parameters for MPUM for four points in the four rivers. The model structures are similar to each other, except that the Seungchon point in the Yeongsan River yields model AR(1) for process two that is slightly different from all other models. Standard errors and *t* statistics of all parameters indicate that the proposed model structures are robust (see Table 4). All of the other 12 points from the four rivers result in similar model structures to those presented in Table 4.

Results and discussions

Comparison of model performances

The simulation performances of the three distinct model structures (*i.e.*, ARMA, SARIMA and MPUM) can be compared using the coefficient of determination (R^2), RMSE and AIC. Even though the ARMA model and SARIMA models are different, many SARIMA models are reduced to something similar to ARMA through the iterative model identification processes presented in Table 3. All of the SARIMA models, namely SAM1–SAM7 (see equations (7)–(13)) were used to achieve a comprehensive performance evaluation of the SARIMA models.

Table 5 presents R^2 for ARMA, SARIMA and MPUM for all 16 points used. The maximum and minimum differences between MPUM and ARMA are 0.62 and 0.16, respectively, and the mean and standard deviation for the corresponding differences are 0.39 and 0.15,

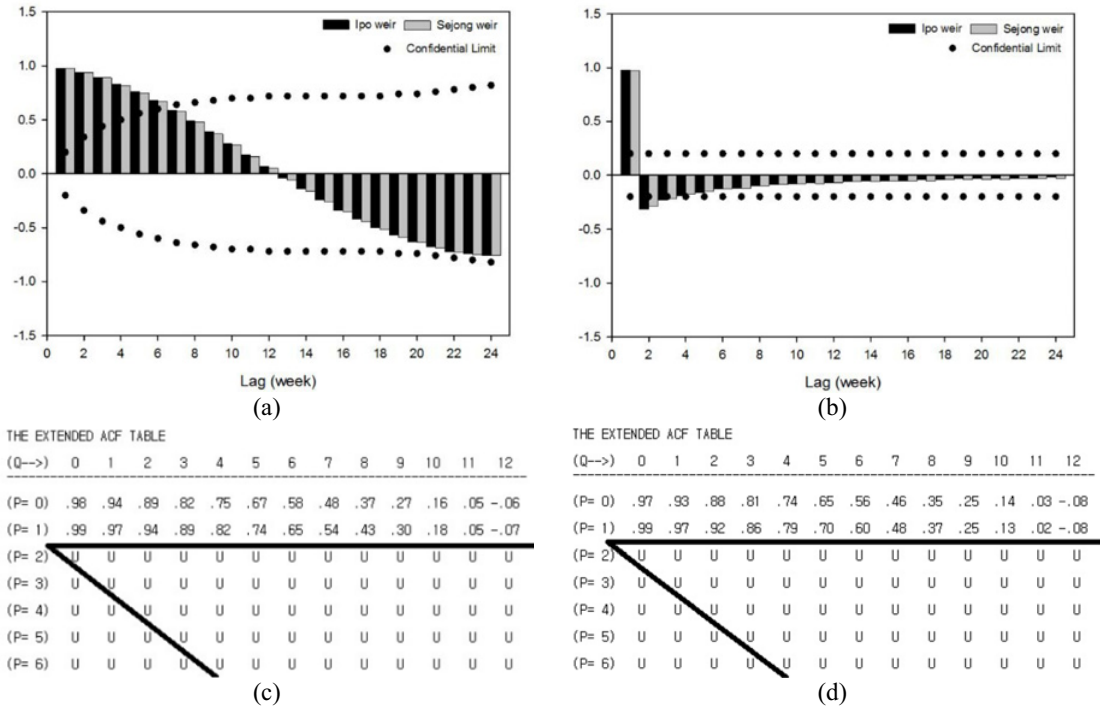

Fig. 5. ACF (a), PACF (b) and EACF (c) and (d) process one for Ipo and Sejong weirs.

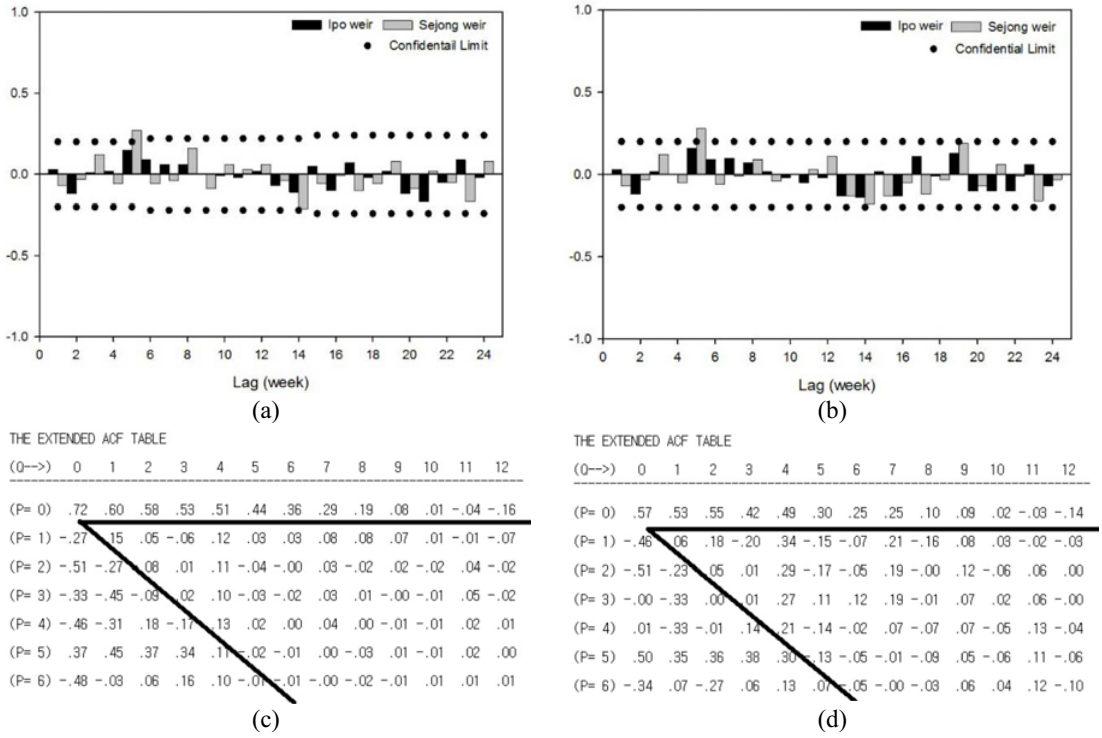

Fig. 6. ACF (a), PACF (b) and EACF (c) and (d) process two for Ipo and Sejong weirs.

respectively. The differences between MPUM and the best SARIMA (*i.e.*, SAM1, SAM2, …, SAM7) are also significant, yielding 0.42 and 0.08 as the maximum and minimum, respectively. The corresponding mean and standard deviation are 0.24 and 0.11, respectively. The R^2 values presented in Table 5 indicate that the modeling performance of MPUM is better than those of ARMA and SARIMA at all of the observed points.

Table 6 presents the RMSEs for ARMA, SARIMA and MPUM for all points. The maximum and minimum differences between ARMA and MPUM and between the smallest SARIMA and MPUM are 13.7 and 0.1

Table 4. Estimated model parameters for the prediction of Chl-a concentrations using MPUM.

Points		p, q order	Statistics	φ_1	φ_2	θ_1
Han River	Ipo	(2, 0)	Value	1.9854	− 1.0000	–
			SE	0.7905E-07	0.7717E-07	
			t-value	0.25E + 08	− 0.1E + 08	
		(1, 1)	Value	0.8950	–	0.4278
			SE	0.0516		0.1112
			t-value	17.34		3.85
Geum River	Sejong	(2, 0)	Value	1.9854	− 1.0000	–
			SE	0.9064E-08	0.9046E-08	
			t-value	0.22E + 09	− 0.1E + 09	
		(1, 1)	Value	0.9109	–	0.5509
			SE	0.0494		0.1003
			t-value	18.42		5.49
Yeongsan River	Seungchon	(2, 0)	Value	1.9854	− 1.0000	–
			SE	0.1185E-07	0.1184E-07	
			t-value	0.17E + 09	− 0.8E + 08	
		(1, 0)	Value	0.4583	–	–
			SE	0.2423		
			t-value	1.89		
Nakdong River	Changnyeong Haman	(2, 0)	Value	1.9854	− 1.0000	–
			SE	0.2248E-07	0.2230E-07	
			t-value	0.88E + 08	− 0.4E + 08	
		(1, 1)	Value	0.8260	–	0.5958
			SE	0.1121		0.1497
			t-value	7.37		3.98

Table 5. Coefficients of determination (R^2) for ARMA, SARIMA and MPUM for all points.

	Model								
Point	ARMA	SAM1	SAM2	SAM3	SAM4	SAM5	SAM6	SAM7	MPUM
Ipo	0.32	0.32	0.32	0.33	0.35	0.32	0.31	0.33	0.74
Yeoju	0.45	0.45	0.45	0.45	0.45	0.46	0.45	0.45	0.70
Gangcheon	0.34	0.34	0.35	0.44	0.61	0.61	0.34	0.43	0.75
Sejong	0.12	0.24	0.31	0.37	0.51	0.41	0.15	0.34	0.74
Gonju	0.38	0.45	0.47	0.47	0.48	0.48	0.46	0.46	0.66
Baekje	0.50	0.53	0.54	0.54	0.54	0.56	0.54	0.54	0.76
Sangju	0.18	0.30	0.30	0.34	0.38	0.32	0.23	0.34	0.51
Nakdan	0.15	0.21	0.22	0.25	0.27	0.37	0.18	0.20	0.69
Gumi	0.17	0.20	0.21	0.22	0.23	0.21	0.20	0.21	0.61
Chilgok	0.27	0.34	0.40	0.39	0.40	0.37	0.32	0.36	0.71
Gangjeong	0.22	0.29	0.35	0.31	0.37	0.32	0.28	0.30	0.46
Dalseong	0.52	0.56	0.56	0.58	0.60	0.57	0.56	0.58	0.68
Hapcheon	0.49	0.53	0.55	0.55	0.55	0.56	0.53	0.55	0.72
Changyeong	0.14	0.21	0.22	0.22	0.23	0.25	0.21	0.21	0.47
Seungchon	0.09	0.18	0.28	0.28	0.29	0.20	0.19	0.19	0.71
Juksan	0.04	0.24	0.24	0.20	0.23	0.20	0.21	0.19	0.63

and 10.4 and 0.1, respectively. The mean and standard deviation for the differences between ARMA and MPUM and between the smallest SARIMA and MPUM are 5.0 and 3.3 and 3.2 and 2.5, respectively. Even though the SARIMA model structures address seasonality in the time series, the model structure of MPUM provides a significant improvement in reducing the RMSE for all points.

In terms of AIC, the model parsimony (Akaike, 1974) describes modeling residuals and the number of parameters. Table 7 presents the AICs for ARMA, SARIMA and MPUM for all points. The AICs for MPUM are smaller than those for ARMA, except the AIC for the Yeoju point (see Table 7). In facts, the RMSEs for the Yeoju point are substantially lower than those for other points (see Table 6), indicating that the variation of Chl-a concentration is much smaller there than at the other points. The mean and standard deviation of the differences between the AICs of ARMA and those of MPUM are 30.8 and 15.5, respectively. The differences

Table 6. RMSE for ARMA, SARIMA and MPUM for all points.

	Model								
Point	ARMA	SAM1	SAM2	SAM3	SAM4	SAM5	SAM6	SAM7	MPUM
Ipo	4.5	4.5	4.5	4.4	4.4	4.5	4.6	4.4	3.2
Yeoju	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.2
Gangcheon	2.4	2.4	2.4	2.2	1.9	1.9	2.4	2.2	1.5
Sejong	17.1	15.2	14.5	13.8	12.2	13.4	16.2	14.1	8.9
Gonju	23.5	21.8	21.3	21.4	21.1	21.2	21.7	21.6	16.8
Baekje	19.1	18.7	18.5	18.7	18.6	18.5	18.5	18.7	13.4
Sangju	14.8	13.7	13.7	13.2	12.9	13.3	14.6	13.2	11.3
Nakdan	19.7	18.9	18.8	18.5	18.2	17.0	19.5	19.3	12.4
Gumi	19.1	18.3	18.2	18.1	18.1	18.2	18.3	18.3	14.4
Chilgok	17.3	16.2	15.5	15.8	15.5	16.0	16.4	16.0	11.5
Gangjeong	16.2	15.3	14.5	15.1	14.4	15.0	15.4	15.2	13.7
Dalseong	16.0	15.0	14.9	14.6	14.3	14.7	15.0	14.6	12.8
Hapcheon	16.0	14.7	14.4	14.4	14.4	14.2	14.8	14.5	11.3
Changyeong	21.7	20.3	20.3	20.1	20.0	19.7	20.4	20.4	16.7
Seungchon	30.5	28.5	27.2	27.6	27.5	28.2	28.3	28.2	16.8
Juksan	19.0	16.9	16.6	17.1	16.6	17.1	17.3	17.2	11.9

Table 7. AIC for ARMA, SARIMA and MPUM for all points.

	Model								
Point	ARMA	SAM1	SAM2	SAM3	SAM4	SAM5	SAM6	SAM7	MPUM
Ipova	151.6	159.2	163.3	161.6	164.7	162.9	156.6	157.7	123.0
Yeoju	85.4	93.0	97.0	97.1	100.5	95.5	89.6	93.0	86.4
Gangcheon	91.7	98.7	102.0	94.8	81.6	80.3	95.2	91.0	48.3
Sejong	282.1	278.4	278.0	273.1	264.9	270.4	281.0	272.4	221.9
Gonju	313.5	313.8	513.8	316.1	318.9	315.1	309.5	313.1	284.6
Baekje	293.0	298.9	302.2	303.0	306.3	302.0	293.9	298.8	262.6
Sangju	268.1	268.8	272.4	268.9	270.4	270.0	270.8	265.2	245.8
Nakdan	296.3	300.2	303.6	302.0	304.3	293.6	299.0	302.0	254.6
Gumi	292.9	296.9	300.6	300.0	303.6	300.2	293.0	296.7	269.2
Chilgok	283.4	285.1	284.6	286.3	288.9	287.5	282.5	283.6	247.7
Gangjeong	277.1	279.3	278.3	282.1	281.7	281.4	276.3	278.7	264.5
Dalseong	275.7	277.2	280.5	278.6	280.5	279.3	273.1	274.6	257.8
Hapcheon	275.8	75.6	277.7	277.3	281.4	275.8	271.9	273.9	245.6
Changyeong	305.8	307.3	310.9	310.3	313.6	308.6	303.6	307.4	284.0
Seungchon	339.1	340.1	339.8	341.3	344.7	343.2	335.7	339.1	284.3
Juksan	292.4	289.3	291.4	294.0	295.6	294.1	287.2	291.0	250.4

between the minimum SARIMA out of SAM1, SAM2, … , SAM7 and MPUM results in a mean and standard deviation of 28.7 and 12.1, respectively, indicating that the model parsimony of MPUM is substantially better than the other model.

Extended modeling of Chl-a for pre- and post-river flow regulation

Variations in Chl-a was modelled using field measurements gathered between 2006 and 2013. The performance of MPUM is compared with that of traditional models using this extensive field dataset. As illustrated in Tables 5–7, the performances of SARIMA are similar to those of ARMA. In fact, a substantial portion of the SARIMA models was reduced into simpler ARMA models, as can be seen in Table 3. Figure 7 presents modeling results obtained with the ARMA and MPUM models at the Dalseong point in the Nackdong River. The R^2 and the RMSE of ARMA are 0.40 and 21.86, respectively, while the corresponding statistics for MPUM are 0.65 and 16.89, indicating that the model behavior of MPUM is better than that of the existing models even in the extensive data taken both before and after the river regulation project in 2011. The incorporation of seasonal effects into the MPUM structure using independent modeling processes results in a robust capability for modeling the weekly variation of Chl-a concentration, whether the flow control in the river is stronger (after 2012), or more dominated by the natural hydro-meteorologic driver such as rainfall (before 2010).

Structural strengths of MPUM compared with other models

All models have distinct structures and their corresponding model parameters, and the estimation of parameters is an essential modeling procedure. Physical based

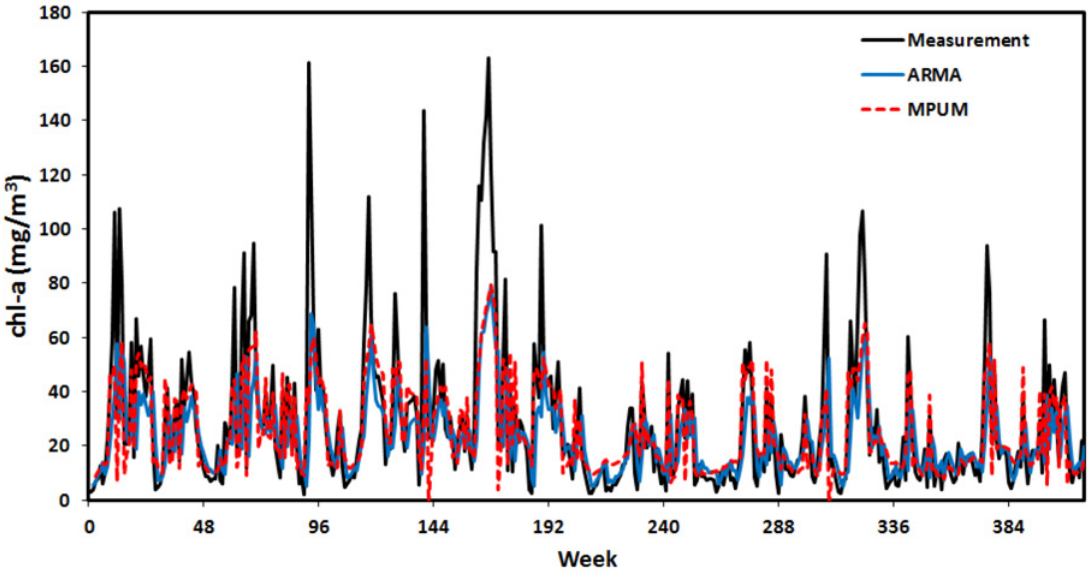

Fig. 7. Extended modeling of Ch-a between 2006 and 2013 at a Dalseong point in Nackdong River, South Korea.

numerical models provide a corresponding explanation for each process that is related to variation in algae concentration, which can be useful in water quality management. Mathematical models also aide in understanding the response of water quality to external drivers (*e.g.*, rainfall and nutrients). However, the number of parameters associated with mathematical water quality models (*e.g.*, HSPF and EFDC) is generally orders of between two and four, depending on the systems and dimensions (2D or 3D) of the problem (U. S. Environmental Protection Agency, 2002, 2015), whereas the number of parameters used by MPUM is less than or equal to four (see Table 4). Considering the entrapment of local optima in the parameter evaluation process, the simple structure of MPUM results in a substantial strength in the optimization of parameters in terms of both accuracy and efficiency. In addition, unlike other models, MPUM does not require additional environmental input data (such as phosphorus or nitrogen), which can often involve uncertainty in terms of data acquisition. As can be seen in equation (A6), the structure of MPUM considers the impact of other environmental variables on the past history of Chl-a concentration, even though the impact of contemporary input variables (which is minor) cannot be considered.

Therefore, MPUM provides several advantages over approaches when predicting of Chl-a concentration. The performance of MPUM in terms of predictability and model parsimony is better than other univariate models. Feasibility of application is improved because specific conditions of the physical world such as construction of weir and regulation of flow, do not need to be addressed. Furthermore, development of a commercially available online sensor for Chl-a also provides an additional strength to univariate model for early warning of algal blooms (Chen *et al.*, 2015).

Conclusion

Existing mathematical and Black–Box models for Chl-a concentrations in inland water systems exhibit various strengths and weaknesses in terms of process representation and simulation accuracy. They also share the common issues of over-parameterization and the occurrence of multiple local optima during in the model calibration procedure. Our proposed model based on the autoregressive moving average operator, addresses algal modeling issues such as predictability, model parsimony and efficiency, through the implementation of multiple independent processes. Through analytical derivations we provided a deterministic basis for univariate time series modeling even with a simplified process assumption. The application of our method and other existing approaches to 16 points in four major rivers demonstrated the potential of the MPUM in terms of the predictability and parsimony of a Chl-a prediction model. Extended modeling results for a period extending both before and after "the Four Rivers Restoration Project of Korea" indicated the robustness of our model even with the presence of substantial flow regulation. The efficient incorporation of seasonality into the model structure as an independent process is responsible for the merits of the proposed method. The relatively simple input requirements make this a feasible tool for environmental managers to use for making algal bloom predictions, with minimum concerns for uncertainty in other environmental variables. The application of our method to other water systems, such as lakes, estuaries, and coastal regions or to the evaluation of impacts of climate change with the introduction of enhanced higher order operators in the model structure, would be topics for future research.

Acknowledgements. This research was partially funded by a research grant from the Ministry of Environment (South Korea) through the National Institute of Environmental Research.

References

Akaike H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, *19*, 716–723.

Box G. and Cox D.R., 1964. An analysis of transformations. *J. R. Stat. Soc., B*, *26*, 211–252.

Box G. and Jenkins G., 1976. Time Series Analysis: Forecasting and Control. revised edn., Prentice-Hall, Englewood Cliffs, N.J.

Cha Y., Park S.S., Kim K., Byeon M. and Stow C.A., 2014. Probabilistic prediction of cyanobacteria abundance in a Korean reservoir using a Bayesian Poisson model. *Water Resour. Res.*, *50*, 2518–2532.

Chen Q., Guan T., Yun L., Li R. and Recknagel F., 2015. Online forecasting chlorophyll a concentrations by an autoregressive integrated moving average model: feasibilities and potentials. *Harmful Algae*, *43*, 58–65.

Coad P., Cathers B., Ball J.E. and Kadiuczka R., 2014. Proactive management of estuarine algal blooms using an automated monitoring buoy coupled with an artificial neural network. *Environ. Model. Softw.*, *61*, 393–409.

French T.D. and Petticrew E.L., 2007. Chlorophyll a seasonality in four shallow eutrophic lakes (northern British Columbia, Canada) and the critical roles of internal phosphorus loading and temperature. *Hydrobiologia*, *575*, 285–299.

Hamilton G., McVinish R. and Mengersen K., 2009. Bayesian model averaging for harmful algal bloom prediction. *Ecol. Appl.*, *19*, 1805–1814.

Iorgulescu I., Beven K.J. and Musy A., 2007. Flow, mixing, and displacement in using a data-based hydrochemical model to predict conservative tracer data, *Water Resour. Res.*, *43*, W03401.

Jun K.S. and Kim J.S., 2011. The four major rivers restoration project: impacts on river flows. *KSCE J. Civil Eng.*, *15*, 217–224.

Kim D., Cao H., Jeong K., Recknagel F. and Joo G., 2007. Predictive function and rules for population dynamics of *Microcystis aeruginosa* in the regulated Nakdong River (South Korea), discovered by evolutionary algorithms. *Ecol. Model.*, *203*, 147–156.

Lee J.H.W., Huang Y., Dickman M. and Jayawadema A.W., 2003. Neural network modeling of costal algal blooms. *Ecol. Model.*, *159*, 179–201.

Liu L.M., 2006. Time Series Analysis and Forecasting. 2nd edn., Scientific Computing Associates Corp., Villa Park, Illinois.

Liu L.M. and Hanssens D.M., 1982. Identification of multiple input transfer function models. *Commun. Stat.- Theory Methods*, *11*, 297–314.

Malek S., Ahmad S.M.S., Singh S.K.K. and Salleh A., 2011. Assessment of predictive models for chlorophyll-a concentration of a tropical lake. *BMC Bioinfor.*, *12*(Suppl 13), S12.

Mallin M.A., Paerl H.W. and Rudek J., 1991. Seasonal phytoplankton composition, productivity and biomass in the Neuse River estuary, North Carolina. *Estuar. Coast. Shelf Sci.*, *32*, 609–623.

Marsili-Libelli S., 2004. Fuzzy prediction of the algal blooms in the Orbetello lagoon. *Environ. Model. Softw.*, *19*, 799–808.

Muttil N. and Lee J.H.W., 2005. Genetic programming for analysis and real time prediction of coastal algal blooms. *Ecol. Model.*, *189*, 363–376.

Novotny V. and Olem H., 1994. Water Quality Prevention, Identification and Management of Diffusive Pollution. Van Nostrand Reinhold, New York, 549–551.

Oh H.M., Ahn C., Lee J.W., Chon T.S., Choi K.H. and Park Y.S., 2007. Community patterning and identification of predominant factors in algal bloom in Daechung Reservoir (Korea) using artificial neural networks. *Ecol. Model.*, *203*, 109–118.

OECD 2013. OECD Compendium of agri-environmental indicators, OECD Publishing. doi: 10.1787/9789264181151-en.

Padisák J., 2004. Phytoplankton. *In*: O'Sillivan P.E. and Reynolds C.S. (eds.), The Lakes Handbook Vol. 1 Limnology and Limnetic Ecology. Balckwell Publishing, Oxford, UK.

Park S.B., 2012. Algal blooms hit South Korea rivers. *Nature*.

Salas J.D., Delleur J.W., Yevjeuch V. and Lane W.L., 1988. Applied Modeling of Hydrologic Time Series. Water Resource Publication, Chelsea, Michigan.

Sin Y., Wetzel R.L. and Anderson I.C., 2000. Seasonal variations of size-fractionated phytoplankton along the salinity gradient in the York River estuary, Virginia (USA). *J. Plank. Res.*, *22*, 1945–1960.

Thomann R.V. and Mueller J.A., 1987. Principles of Surface Water Quality Modeling and Control. Harper and Row, New York.

U. S. Environmental Protection Agency 2002. User Manual for Environmental Fluid Dynamics Code Hydro Version (EFDC-HYDRO)., Tetra Tech, Inc., Fairfax, VA.

U. S. Environmental Protection Agency 2015. Application of BASIN/HSPF to Data-scarece Watersheds. EPA/600/R-15/007, Washington, DC.

Weiler M., McGlynn B.L., McGuire K.J. and McDonnell J., 2003. How does rainfall become runoff? A combined tracer and runoff transfer function approach. *Water Resour. Res.*, *49*, 1315, doi: 10.1029/2003WR002331.

Whitehead P.G., Howard A. and Arulmani C., 1997. Modeling algae growth and transport in rivers: a comparison of time series analysis, dynamic mass balance and neural networks techniques. *Hydrobiologia*, *349*, 39–46.

WHO, 2003. Algae and cyanobacteria in fresh water. *In*: Bartram, J. and Rees G. (eds.), Guidelines for Safe Recreational Water Environments. Vol. I, Coastal and Fresh Water. World Health Organization, Geneva, 136–158.

Woo H., 2009. Korea to launch a major project on river rehabilitation. *J. Hydr. Res.*, *47*, 74–75.

Wu G. and Xu Z., 2011. Prediction of algal blooming using EFDC model: case study in the Daoxiang Lake. *Ecol. Model.*, *222*, 1245–1252.

Appendix

Assuming that Chl-a concentration in river systems is determined by both bio-chemical and hydrometeorological drivers, the time series of dimensionless Chl-a (through the centralization process (Salas *et al.*, 1988), z_t, can be expressed as follows:

$$z_t = f(S_{t-1}, \ldots, S_{t-m}) + g(X_t, \ldots, X_{t-n}) \quad \text{(A1)}$$

where $f(S_{t-1}, \ldots, S_{t-m})$ is a function of the time series of dimensionless bio-chemical components, $S_{t-1}, \ldots, S_{t-m}$, and $g(X_t, \ldots, X_{t-n})$ is a function of the time series of dimensionless hydrometeological components, $X_t, \ldots X_{t-n}$. The functions in equation (A1) can be approximated by polynomials as $a_1 S_{t-1} + \ldots + a_m S_{t-m}$ and $b_1 X_t + \ldots + b_n X_{t-n}$.

If the Chl-a concentration at the current time step is assumed to be determined by a hydrometeorological variable at the current time step and a bio-chemical variable at previous time step, then

$$z_t = aS_{t-1} + bX_t \quad \text{(A2)}$$

where a and b are coefficients $(a, b < 1)$ for the corresponding components.

By considering mass conservation and the impact of the hydrologic driver, the bio-chemical component S can be expressed as,

$$S_t = S_{t-1} + C'X_t - aS_{t-1} \quad \text{(A3)}$$

where the coefficient C' is given by cS_t/X_t and c is the hydrologic component to bio-chemical component conversion factor. In fact, transfer functions from the hydrological component into the water quality concentration have previously been modeled using several approaches (Novotny and Olem 1994; Iorgulescu *et al.*, 2007; Weiler *et al.*, 2003).

By expressing the previous time step using equation (A2) we can express S as

$$S_{t-2} = \frac{1}{a} z_{t-1} - \frac{b}{a} x_{t-1} \quad \text{(A4)}$$

The previous time step can be expressed using equation (A3) as,

$$S_{t-1} = (1-a)S_{t-2} + C'X_{t-1} \quad \text{(A5)}$$

By substituting equation (A5) into equation (A2), and using this with equation (A4), we obtain

$$z_t = (1-a)z_{t-1} + bX_t + (a(C'+b) - b)X_{t-1} \quad \text{(A6)}$$

Equation (A6) has the form of an ARMA(1,1) model, if the hydrologic variable X is an independent series. Depending on the assumptions on the functions in equation (A2), more complicated ARMA(p,q) models can be formulated.