Prewhitened causality analysis for the chlorophyll-a concentration in the Yeongsan River system
Eunhyung Lee, Sanghyun Kim, Eunhye Na and Kyunghyun Kim

ABSTRACT
Blooming of algae has been a primary issue of concern for heavily polluted aquatic ecosystems. The chlorophyll-a (Chl-a) concentration depends on various hydrological, biochemical and anthropogenic components, which makes prediction of algal blooms complicated. A river regulation project in Yeongsan River, South Korea, involving the construction of a weir, had substantially altered the flow regime. A prewhitened time series analysis is a useful method for delineation of a causal relationship between two environmental variables. This study explores the impact of river regulation on algal blooming using both the prewhitened cross-correlation method and principal factor analysis. Both individual and comprehensive causality structures were configured for the variation in Chl-a concentration. A prewhitened cross-correlation analysis indicates that the water quality response patterns of the river system were changed to those of a reservoir after the river regulation project. A principal factor analysis of correlations indicates that the weir construction had a stronger impact on algal concentration than both the hydro-meteorological factor and difference in sampling location. Variation in stochastic structures from nutrients and water quality factors to algal bloom was substantially reduced by the construction of a weir, which can be explained by the relatively uniform flow pattern throughout the river regulation practice.

Key words | algal prediction, chlorophyll-a concentration, factor analysis, prewhitening analysis, water quality

INTRODUCTION
It is well known that algal blooms not only result in the decline of specific species, such as invertebrates, but also lead to deterioration of fish habitats due to increasing turbidity. Significant algal blooms often result in unpleasant taste and odor problems in drinking water as well as issues with filter clogging in water treatment plants. Furthermore, some species of cyanobacteria produce toxins (cyanotoxins) that cause significant ecological and human health concerns (WHO 2003).

Algal bloom problems in major river systems in South Korea have become critical social issues since the Four Rivers Restoration Project was completed in 2011 (Park 2012). The effect of the large weirs constructed along the rivers during the project on the frequency and intensity of algal bloom is still controversial. Therefore, consideration of the existence of the weirs is particularly important when investigating the causality of algal blooms. Also, prediction of algal blooms is important because it is advantageous in the implementation of appropriate management practices, such as low flow augmentation (the release of water stored in a reservoir) for river systems.

The variation of algal blooms has been modeled using process-based approaches (Thomann & Mueller 1987; Whitehead et al. 1997; USEPA 2002, 2015; Wu & Xu 2011). However, processes associated with behavior of algae are extremely complicated, and the predictions that can be achieved through mathematical modeling are often limited. This is due to the uncertainty in the kinetic coefficients and...
the structural complexity of two- or three-dimensional models, which usually require a substantial amount of input data and demanding calibration and validation procedures (USEPA 2015).

Alternatively, the prediction of behavior of algae has been explored using several statistical methods or techniques based on artificial neural network (ANN) (Lee et al. 2003; Marsili-Libelli 2004; Mutil & Lee 2005; Oh et al. 2007; Hamilton et al. 2009; Malek et al. 2011; Cha et al. 2014; Coad et al. 2014). Applications of ANN or statistical methods have shown potential in the prediction of algal variation, with feasible input data and high accuracy in simulations. However, ANN does suffer from the drawbacks of slow convergence and the presence of multiple local optima, which are associated with the feed forward back propagation algorithm of ANN. Furthermore, ANN or statistical approaches do not provide any process-based insight in their modeling results. Depending on data availability, meta-heuristic approaches frequently result in different modeling results, which may indicate violation of the stationary process assumption in time series modeling. In general, ANN and statistical models also yield distinct modeling results for different locations, even within an identical watershed, which is related to the large number of model parameters (weights) appearing in black box modeling. Redundancies in inputs and interactions in the model structure often result in over-parameterization in the statistical model. Therefore, optimum identification of the relevant environmental factors is important to determine the appropriate model structure for prediction of algal biomass.

When we find a relationship between any environmental variable and chlorophyll-α (Chl-α) concentration, using common index to represent the amount of algal biomass, the conventional cross-correlation function often shows limitations due to the autocorrelation structure of each time series (Kim et al. 2002). This issue can be largely addressed through the introduction of a prewhitening method (Kim 2002). Case studies for the Elbe River in Germany and Murray River in Australia also indicated that the delineation of actual correlation between water quality variables can be achieved through filtering a common stochastic structure between input and output time series (Maier & Dandy 1997; Lehmann & Rode 2001).

The relationships among many variables can be elucidated by the introduction of factor analysis with minimum loss of information. The biplot of factor analysis efficiently describes the correlation among different variables and was found to be useful in evaluating water quality at monitoring points (Mohamed et al. 2015) and in identifying contaminant sources associated with heavy metals (Basamba et al. 2013).

To understand the possible variation in causality due to environmental factors that influence algal blooms in terms of Chl-α concentration before and after the river regulation project, various time series of meteorological, hydrological and water quality parameters and Chl-α concentration were collected upstream and downstream of the weir, which was constructed in 2012 in the Yeongsan River, South Korea. This study addresses two issues based on the analysis of measurements of hydro-meteorological, nutrient, and water quality parameters during both pre and post periods of a river regulation project: (i) how a prewhitened relationship (Granger causality) between various environmental parameters and Chl-α concentration can be obtained as a stochastic structure associated with common drivers by eliminating seasonality from each time series; (ii) how multiple relationships between hydro-meteorological, nutrient, and water quality factors and algal bloom can be expressed through factor analysis. The first and second questions address linear and comprehensive causality structures, respectively, for variation of Chl-α concentration with and without the impact of river regulation. Finally, we analyzed how the construction of the weir changed the algal concentration response pattern in the context of various environmental factors.

**MATERIALS AND METHODS**

**Study area and data acquisition**

Considering data availability during pre and post periods of the river regulation project, two locations in the Yeongsan River were selected for this study (see Figure 1). The Seungchon Weir was constructed between 2010 and 2011 under ‘the Four Rivers Restoration Project of Korea’, which aimed to address water-related issues such as preventing
floods and storing more water for droughts (Woo 2009). Therefore, we define the period from 2007 to 2009 as the pre period of the project, 2010 to 2011 as the construction period, and 2012 to 2014 as the post period. For those periods, water quality samples were collected and analyzed at weekly intervals at the two locations, Kwangsan and Najoo, for upstream and downstream points of the Seungchon Weir (see Figure 1). The Ministry of Environment had been responsible for obtaining data for 15 water quality variables, which include biochemical oxygen demand (BOD), chemical oxygen demand (COD), dissolved oxygen (DO), pH, suspended solids (SS), electric conductivity (EC), water temperature (WT), Chl-α concentration, total nitrogen (TN), ammonium nitrogen (NH₃N), nitrate nitrogen (NO₃N), dissolved total nitrogen (DTN), total phosphorus (TP), dissolved total phosphorus (DTP), and phosphate phosphorus (PO₄P). Chl-α was extracted in acetone and concentrations measured using spectrometry methods.

Meteorological data were obtained from datasets generated by the Korea Meteorological Administration. The dataset from the Kwangjoo Regional Meteorological Office (126° 53′ 00″, 35° 10′ 00″), which is located close to the Seungchon Weir, was selected as the representative meteorological data. In order to match the temporal resolution to the weekly water quality data for time series analysis, meteorological data were converted to weekly interval datasets. The meteorological data used in this study were atmosphere temperature (AT), rainfall (RAIN), wind speed (WS), humidity (HU), cloud amount (WIND), solar radiation (SoR), and duration of sunshine (SuD). Temporal resolution of other hydrological data such as rainfall and flowrate were also converted to a weekly interval.

**Procedure for the time series analysis**

If statistics of data show high skewness, a significant trend, or periodicity, then proper pretreatment is required before
the primary analysis. A suitable transformation of the time series can improve the normality of the series, to eliminate trends and seasonality. The Box–Cox transformation (Box & Cox 1964) has been the most widely used method to remove non-normality from a time series, which can be defined as:

\[ x(t) = (z(t) + C)^\gamma, \quad \gamma \neq 0 \]  
(1)

\[ x(t) = \ln(z(t) + c), \quad \gamma = 0 \]  
(2)

where \( z(t) \) is the time series, and \( C \) and \( \gamma \) are constants.

The cross-correlation function, the degree of the linear relationship between two series, can be computed as:

\[ r_{xy}(t) = \frac{C_{xy}(t)}{(C_x(0) \cdot C_y(0))^{0.5}} \]  
(3)

where \( C_{xy}(t) \) is the cross-covariance between two series, i.e., \( x \) and \( y \), and \( C_x(0) \) and \( C_y(0) \) are the variance estimates for the \( x \) and \( y \) series, respectively.

Depending on the stochastic structure of the two series, i.e., whether they are mutually cross-correlated or auto-correlated, and both are white noise processes, the variance of the cross-correlation function is distinctively expressed. In this approach, the confidence intervals of the cross-correlation function are estimated based on the assumption that the time series had no pre-stochastic process.

To remove the stochastic signal from the pretreated series, an appropriate univariate model can be introduced. The autocorrelation and partial autocorrelation structures of the reference series identify a suitable model structure (Box & Jenkins 1976). The time series model for the prewhitening process can be expressed as:

\[ \phi_x(B) \cdot x(t) = \theta_x(B) \cdot u(t) \]  
(4)

where \( B \) is the backshift operator defined as \( B^k x(t) = x(t-k) \), \( k \) is a positive lag, \( \phi_x(B) = 1 - \phi_1B - \phi_2B^2 - \ldots - \phi_mB^m \), and \( \theta_x(B) = 1 - \theta_1B - \theta_2B^2 - \ldots - \theta_mB^m \).

Once the structure of the model has been determined, the identical model is used to remove the stochastic structure of the other series as:

\[ \phi_y(B) \cdot y(t) = \theta_y(B) \cdot v(t) \]  
(5)

Therefore, the prewhitened series, \( u(t) \) and \( v(t) \) from Equations (4) and (5), can be obtained as residual estimates from Equations (4) and (5), respectively. The independence of the residuals can be checked by computing the autocorrelation and autocorrelation functions (ACFs) of the residuals and comparing the confidence intervals.

The prewhitened causality between the two series could be checked using the cross-correlation function of model residuals as follows:

\[ r_{uv}(k) = \frac{C_{uv}(k)}{(C_u(0) \cdot C_v(0))^{0.5}} \]  
(6)

where \( C_{uv}(t) \) is the cross covariance estimate, which can be calculated as:

\[ C_{uv}(k) = \frac{1}{N} \sum_{t=1}^{N-k} u(t+k) \cdot v(t), \quad k \geq 0 \]  
(7)

\[ C_{uv}(k) = \frac{1}{N} \sum_{t=1-k}^{N} u(t+k) \cdot v(t), \quad k < 0 \]  
(8)

and \( C_u(0) \) and \( C_v(0) \) are the variance estimates for the prewhitened series.

Cross-correlation functions (CCFs) between hydrochemical components data and Chl-a concentration data (e.g., Equation (3)) often resulted in unacceptable relationships. The CCFs between original time series suffer from their autocorrelation structure and periodicity. They make cross-correlation analysis difficult by obscuring original causality relationships between environmental factors and algae. In order to delete and weaken the obstructive factors, we implemented the prewhitening procedure. Prewhitening is necessary to delineate the causality of algal blooms accurately (Maier & Dandy 1997; Lehmann & Rode 2001).

PRINCIPAL FACTOR ANALYSIS

Principal factor analysis (PFA) is a multivariate analysis method to describe clustered relations among several variables. It is also referred to as a dimensional reduction method, which uses simplified factors to represent substantial amounts of information. Using CCFs obtained through...
prewhitened causality analysis, datasets were categorized into multiple time series depending on the location and sampling time. We characterized stochastic structures of the time series with the spatial distribution of factor analysis through biplots of factors such as meteorological, water quality, and nutrient components.

If X is a probability vector which has p variables, the average vector, $\mu$, for the corresponding X vector can be defined. Since it is impossible to visually illustrate p dimensions, the dimension and sample size of the vector need to be downsized. An application of PFA reduces the sample size of a vector to s, which is lower than p. Equation (9) represents PFA for reduction of vector size as:

$$X - \mu = LF + \varepsilon$$

where $L$ is the matrix of factor loadings having a dimension as $p \times s$, which consists of factor loadings ($\lambda_i$). $F$ is a common factor vector, $(f_1, f_2, \ldots, f_s)^T$, with zero as an average and $I_{ss}$ as variance, $\varepsilon$ is a specific factor, $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_s)^T$, with zero as an average and $(\phi_1, \phi_2, \ldots, \phi_s)^T$ as variance.

The total variance in PFA can be separated into the common factor variance and the specific variance. The common factor variance is connected with other variables, but the specific variance is for an independent variable (Wrigley 1957).

The covariance matrix ($\Sigma$) can be defined as the sum of the factor loading matrix and a specific covariance matrix ($\varphi$) through a factor decomposition as:

$$\Sigma_{pp} = \text{Cov}(X - \mu) = LL^T + \varphi$$

The variance of X can be expressed as the sum of factor loadings and a specific variance as:

$$\text{Var}(x_i) = \lambda_{i1}^2 + \cdots + \lambda_{is}^2 + \varphi_{ii} = h_i^2 + \varphi_{ii}$$

where $h_i^2$ is a common variance and $\varphi_{ii}$ is a specific variance which is independent of other variables.

Singular value decomposition (SVD) was used to estimate L and $\varphi$ using a reduced sample covariance matrix. A standardized matrix ($Y$) can be calculated as:

$$Y = UD_1V^T = \sum_{k=1}^{p} \lambda_k u_k v_k^T$$

where $v_k$ is the eigenvector obtained through the spectral decomposition of the sample covariance matrix.

The relationship between sample covariance matrix ($S$) and standardized matrix ($Y$) can be expressed as:

$$S = (n - 1)^{-1}Y^TY$$

where $n$ is the total number of environmental variables. The covariance matrix can be estimated through spectral decomposition as:

$$S = (n - 1)^{-1}VD_1V^T = (n - 1)^{-1}\sum_{k=1}^{p} \lambda_k^2 v_k v_k^T$$

where the diagonal matrix $D_1$ is $\text{diag}(\lambda_1, \ldots, \lambda_p)$.

If n is known, Y, $\lambda_k$, and $v_k$ can be obtained from Equation (12). The matrix of factor loadings can be calculated as:

$$L_{(s)} = (n - 1)^{-1/2}V_{(s)}D_{(s)} = (n - 1)^{-1/2}(v_1 \lambda_1, \ldots, v_p \lambda_p)$$

The total variance is used to estimate the specific variance matrix ($\varphi$) using the equation $S - L_{(s)}L_{(s)}^T$. Therefore, all variables can be expressed in terms of loading factors of factor 1 and factor 2, which can be obtained through a biplot of the corresponding coordinates. The relationship among different variables can be expressed through their spatial distributions (i.e., position and angle) in the biplot of factor analysis (Choi & Byun 1996).

In order to quantify the comprehensive relationships, the angles of biplots between vectors can be used to evaluate similarity between two vectors. The vector in a biplot can be expressed through its magnitude and direction, representing variance and relative location, respectively. Similarity between two vectors in a biplot can be mathematically defined in terms of the angle between them, which can be calculated as:

$$\theta = \cos^{-1} \left( \frac{v_1 \cdot v_2}{||v_1|| ||v_2||} \right)$$

where $\theta$ is the angle between two vectors, and $||v_i||$ is a norm of vector $v_i$. The smaller the angle between two vectors, the higher the relationship between two factors.
RESULTS AND DISCUSSION

Statistics of chlorophyll-a concentration

Concentrations of Chl-a in the pre and post periods of the river regulation project were substantially different. The variation of algal blooms also appears to have changed as a result of the river regulation practice, because variances for the period between 2012 and 2014 were significantly greater than those for the period between 2007 and 2009 (see Figure 2). Figure 2 presents a boxplot of Chl-a concentrations for 12 datasets, which were obtained by combination of the sampling locations and the years of the pre and post periods (i.e., KS07 represents the data at Kwangsan in 2007). The black dots show the outliers of the data. Mean concentrations of Chl-a after construction of the Seungchon Weir (KS12 ∼ 14 and NJ12 ∼ 14) were higher than those for the earlier period (KS07 ∼ 09 and NJ07 ∼ 09) and intensity of algal blooming seems higher at the downstream point (NJ) than upstream (KS).

Prewhitening time series analysis

In order to investigate the underlying stochastic structures of datasets, the univariate time series modeling procedure (i.e., Equation (4)) can be applied. However, statistics of the datasets such as skewness coefficient indicated that proper pretreatment using the Box-Cox transformation was necessary (Box & Cox 1964). A heuristic approach was used to find the most appropriate parameters, c and γ in Equations (1) and (2), for our datasets, and in most cases, the transformation generally improved the normality of the data. We investigated stochastic structures of the datasets through evaluations of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the datasets. Figure 3(a) and 3(b), respectively, present the ACF and PACF for each confidence interval of the rainfall time series data in 2008. The pattern of decreasing sinusoidal wave for ACF and infinite extent with damped waves after lag 1 for PACF indicates that ARMA(1,1) is a potential model for the time series data. The parameter estimation for the model was conducted through sequential applications of maximum likelihood estimates, which was used to find the sum of the squares surface for a range of parameters and locate the minimum with the corresponding parameters. As a diagnostic checking procedure, the residuals of the modeling for the same time series data were evaluated to detect any stochastic feature. Figure 3(c) and 3(d) showed the ACF and PACF of the residuals with 95% confidence intervals, which reveal no notable stochastic structure which contains periodicity and autocorrelation. An identical prewhitening procedure was applied to delineate the residual time series for all datasets, which addressed the effects of location, year, and distinct environmental factors.

The results of prewhitening are demonstrated in Figure 4 for data from the upstream and downstream points.

Figure 2 | Boxplot of Chl-a concentration for the upstream point (a) and downstream point (b) before and after weir construction, respectively.
Figure 4(a) and 4(b), respectively, show significant CCFs at −8 to −3 weeks for rainfall and Chl-a concentration and sinusoidal CCF response pattern for TP and Chl-a concentration. Figure 4(c) and 4(d) show the CCFs between prewhitened rainfall and Chl-a concentration in 2008, and between prewhitened TP and Chl-a concentration in 2012, respectively. CCFs obtained through the prewhitening procedure are different from those for the original time series (Figure 4(a) and 4(b)). The periodicity in CCF for the original time series is eliminated through the prewhitened causality analysis. Rainfall showed a significant negative relationship in lag 0 with negligible correlations for all other lags (Figure 4(c)) and TP did not show any causality structure related to Chl-a concentration (Figure 4(d)).

The significant negative CCF at 0-lag is explained by the increasing flowrate due to rainfall, which can remove the algae in the river. Considering the change of flowrate after rainfall in the river, the reduction of Chl-a concentration seems physically meaningful. Also, the significant positive CCF at 1-lag can be explained by the nutrient inflow from agricultural areas, forest, and livestock farms after rainfall, which is the main cause of algal growth. No significant relationship between TP and Chl-a can be explained by the generally high TP concentration in all datasets. That is, the TP concentration is high enough that it is not limiting algal growth. Both nitrogen and phosphorus balance per hectare of the agricultural land in Korea were three to four times the corresponding average values for OECD countries (OECD 2015). Concentrations of TP did not control the algal blooming process for major rivers in South Korea (Kim et al. 2007).

Using prewhitening and cross-correlation methods, the prewhitened causality analysis between environmental
factors and Chl-α concentration can be accurately evaluated. In particular, the comparison of correlation analysis at 0-lag between the conventional and prewhitened methods provides improved analysis using prewhitening for algal responses. Figure 5(a)–5(d), respectively, show cross-correlations for points upstream and downstream of the weir before and after weir construction. As shown in Figure 5, correlation relationships of Chl-α with SoR, HU, BOD, COD, and pH were downscaled in the prewhitened causality analysis, indicating less impact of these environmental factors. In addition, notable differences in correlations were found in rainfall and phosphorus levels. Significant negative correlations in rainfall can only be found in the prewhitened correlations, indicating the importance of the hydrometric driver in algal blooms. Comparisons of the prewhitened correlations for conditions pre- and post-weir construction indicate that forms of phosphorus, such as DTP and PO₄P, played an important role in determining algal blooms only after river regulation had begun.

**Correlation analysis of Chl-α concentration in the Yeongsan River**

The results of prewhitened correlation at lag 0 between various environmental factors and Chl-α concentrations are summarized in Table 1. Meaningful correlations at 0 lag are ±0.28, which is the criterion for the 95% confidence level. Positive and negative correlations indicate the increasing tendency of Chl-α concentrations to increase and...
decrease with the corresponding environmental factor, respectively.

As presented in Table 1, the CCF value characteristics can be separated into two groups showing different trends, i.e. those with dissimilar CCF values and those with similar CCF values between pre and post weir construction. For the first group, differences of CCF value between pre and post weir construction can be expressed by the changing of the water system characteristics from a current flow system to a reservoir system. There are three meaningful examples: DO, WT and phosphorus components. Correlations of Chl-α concentrations with DO are significantly different between the pre and post periods of weir construction. All the CCF values before weir construction are not statistically significant, except for CCF of DO at KS08. This shows that the cross-correlation of DO gets stronger after weir construction. Also, this high correlation with oxygen indicates that the characteristic of the flow has been converted from an open to a closed system in the following two aspects. First, phytoplankton production in the epilimnion may be attributed to the increase of flow residence time caused by the weir (Heiskary & Markus 2000). The flow velocities between 2007 and 2009 were faster than those between 2012 and 2014. Stagnation of water by the river regulation enhances
the causality between Chl-a and DO concentrations (Park 2012). Second, the significant relationship between DO and Chl-a can be explained by increasing the effects of photosynthesis and respiration following algal proliferation. The impact of photosynthesis is stronger than that of respiration because all data were collected in the middle of the day. As phytoplankton are buoyant, they remain in the epilimnion, where conditions are favorable for algal photosynthesis; therefore, oxygen concentration increases in reaction to phytoplankton activity (Thomann & Mueller 1971). The dynamics of phytoplankton can be explained by the positive relationship between Chl-a concentration and DO (Hardy 1973).

Correlations between WT and Chl-a concentrations also showed more frequent and higher causality during the period of post weir construction than pre construction (see Table 1). As flow residence time increases after the weir construction, the surface water temperature is more subject to change. WT not only affects the size of the phytoplankton through controlling enzymatic reactions but also plays an important role as a limiting factor in controlling proliferation and microbial growth (Kormas et al. 2010). WT between 15 °C and 25 °C is known to be appropriate for a rapid growth rate of algae (Davis et al. 2009), but at WT below 13 °C and above 30 °C, the growth rates tend to decrease (Reynolds 1984; Chu & Rienzo 2013). As WT increases and remains high, the algal growth increases according to changing temperature.

For phosphorus components (TP, DTP, PO₄-P), the trend differences including negative correlation after weir...
construction and positive correlation before weir construction come from changing phosphorus dynamics, which affect algal growth. The total phosphorus is composed of two different components: total dissolved phosphorus (DTP) and a total particulate component. DTP includes dissolved reactive phosphorus (ortho-phosphorus). The reactive phosphorus is a potential source for phytoplankton growth. Strong negative correlations between PO₄P and Chl-a concentration between 2012 and 2014 indicate that construction of the weir makes the water environment of the Yeongsan River similar to that of a reservoir. In the epilimnion of the lake, the ortho-phosphorus is the main form of P transported through entrainment, which can be explained by the fact that the ortho-phosphorus ions reach the surface layer via pore water (Gibbs 1977). Algae readily consume the ortho-phosphorus ions in the surface layer (Thomann & Mueller 1987). In contrast, Chl-a is more influenced by TP in riverine conditions (Table 1), partially because the phytoplankton biomass does not have enough time to grow using the ortho-phosphorus supply present in the current system (Mainstone & Parr 2002). However, the phytoplankton increases with TP in the study area, which comes in the form of point and non-point sources.

For the similar CCF values trend, environmental factors like BOD, COD, rainfall, SoR, and SuD are not affected by pre and post weir construction. Relationships between oxygen demands (e.g., BOD and COD) and Chl-a concentrations can be explained through algal decomposition and respiration and their role in the production of oxygen demand (Heiskary & Markus 2001). As the number of algae increases, the contribution to BOD and COD gets higher. Frequent rainfall drastically changes water characteristics, such as WT, DO, and illumination intensity, because of complete or partial dilution. During successive rainfall events, conditions for continuous algal growth can be severely limited. Heavy rainfall may wash away most of the algal growth. The negative correlation between rainfall and Chl-a concentration looks reasonable, and slightly decreased correlations for post river regulation indicate the reduced mitigation impact of rainfall on algal blooms. The other environmental factors studied, i.e., pH, EC, SS, TN, NH₃N, TDN, AT, WS, HU, and WIND, did not show any noticeable correlations with the Chl-a concentration.

Prewhitened correlations for negative lags can be even more important in the configuration of the causality between environmental factors and Chl-a concentration than those with no lag. Considering the time scale of algal growth and decay, we investigated meaningful stochastic structures of algal causality for lags shorter than three weeks.

Table 2 presents causality structures of negative lags for the upstream point. More environmental factors were found to be meaningful for variation of Chl-a concentration during the post river regulation than those prior to river regulation. Significant correlations appeared in hydro-meteorological components such as WS, SoR, and SuD for the post river regulation period. Positive correlations in WS, SoR, and SuD indicate that the meteorological impact on the algal growth is increased due to the weir construction. The reduction of flow velocity originating from the backwater impact of the weir can be strongly associated with the hydro-meteorological factor, which results in algal blooming.

Significant correlations of negative lags for the downstream point are summarized in Table 3. The downstream
point also showed that more environmental factors were related to Chl-a concentration during post river regulation than those prior to weir construction. A notable stochastic structure of causality was found in factors related to nitrogen (e.g., TN and DTN). Significant correlations in lags 1 and 3 for 2012 and 2013, respectively, indicate that the impact of nitrogen on algal blooming increased since the construction of the weir.

### Comprehensive relationships to Chl-a concentration

Using the results of prewhitened correlation analysis, we investigated comprehensive structural distributions of environmental variables for generation of Chl-a concentration using about 12 distinct datasets, which were classified depending on the sampling period and location. The dataset is made up of CCF values at 0-lag, and this dataset represents the correlation with environmental factors within the same sampling period and location. Results of prewhitened causality analysis can be further categorized depending on the characteristics of environmental factors such as hydro-meteorological, water quality, and nutrient factors. The hydro-meteorological factors include AT, Rain, WS, HU, Wind, SoR, and SuD. Water quality factors include pH, DO, BOD, COD, SS, WT, and EC. The nutrient factors are related to nitrogen and phosphorus: TN, NO3N, NH3N, TDN, TP, PO4P, and TDP.

Figure 6 presents biplots of factor analysis using conventional CCFs. The dashed and solid lines, respectively, represent vectors of before and after construction of Seungchon Weir. All eigenvalues were greater than 80%, which indicates that biplots explain relationships among all vectors well. Each vector in Figure 6 represents the corresponding characteristic depending on period and location. Distributions of solid lines and dashed lines in Figure 6 were hardly distinguishable except those for meteorological factors (Figure 6(b)), which show small differences except vectors in 2007. This means that the impact of weir construction on comprehensive relationships cannot be delineated by the existing method.

Figure 7 presents biplots of the factor analysis using the prewhitened CCFs. Figure 7(a) shows the biplot of factor analysis for all environmental factors; the summation of eigenvalues addresses 86.81% of total variation. The results for total environmental factors show distinct distributions between solid and dashed lines in the biplot (Figure 7(a)). The differences between pre and post periods of construction of the weir indicates that the temporal impact is more important than the spatial location impact, such as upstream (KS) or downstream (NJ).

The biplot of factor analysis for the hydro-meteorological factor is presented in Figure 7(b). The summation of eigenvalue incorporates 91.51% of total variation. Spatial distributions of most datasets were similar except for 2007. This means that the impact of the hydro-meteorological factor is not significant for the causality of algal bloom during either pre or post periods of the river regulation project.

Figure 7(c) presents biplots of factor analysis for nutrient factors. Unlike other biplots, both dimension 1 and dimension 2 have significant eigenvalues: 57.74% and 27.21%, respectively. The biplot patterns for the period between 2007 and 2009 look widely distributed, but those for periods after the weir construction are concentrated on the left middle part (Figure 7(c)). This means that the weir construction introduces a notable similarity in the causality structure.

### Table 3 | Significant CCFs of negative lag at the downstream point (Najoo) in the Yeongsan River

| Downstream | pH   | DO   | COD  | WT   | EC   | TN   | DTN  | WS   |
|------------|------|------|------|------|------|------|------|------|
| NJ07       | 0.41 (−2) | 0.31 (−2) | 0.31 (−3) | 0.36 (−2) |
| NJ08       |      |      | 0.32 (−1) |      |      |      |      |      |
| NJ09       |      | 0.38 (−1) | 0.30 (−3) |      |      |      |      |      |
| NJ12       | 0.30 (−3) |      |      | 0.31 (−1) | 0.31 (−1) | −0.32 (−3) |
| NJ13       | 0.30 (−1) |      | 0.33 (−3) | 0.30 (−3) |      |      |      |
| NJ14       | −0.32 (−1) | 0.30 (−2) | 0.30 (−1) |      |      |      |      |

Numbers in parentheses are lag.
between the nutrient factors and algal blooms. The diversity of algal variation in the river system can be switched to a less variable nutrient process to algal bloom in a system similar to a reservoir. Furthermore, the differences between the upstream point (KS) and the downstream point (NJ) are reduced after the construction of the weir.

Figure 7(d) presents the biplot of factor analysis for the water quality factors. Dimension 1 is important because the eigenvalue of dimension 2 is not significant. Distribution for the period of pre weir construction showed a wide range of angles, but the datasets from the post weir construction period showed less variation.
in both distance and angle. Although the difference between pre and post periods of weir construction is smaller than that for the nutrient factor, the water quality factor also shows that the difference between upstream and downstream points becomes less after the construction of the weir. Stronger stagnation flow pattern and longer residence time due to the weir construction are responsible for the similarity of causality in nutrient and water quality factors regardless of the location of the sampling point.

The comparison between Figures 6 and 7 indicates that the factor analysis using prewhitened CCFs can identify the impact of weir construction for configuration of comprehensive causality. The prewhitened CCFs by filtering

![Figure 7](http://iwaponline.com/wqrj/article-pdf/54/2/161/682865/wqrjc0540161.pdf)

**Figure 7** Biplots of factor analysis based on the prewhitened CCFs considering all factors (a), meteorological factors (b), nutrient factors (c), and water quality factors (d). Eigenvalues are displayed in the upper right hand of each part.
autocorrelation and periodicity produce a better explanation than the conventional approach.

So, we calculate the angle by using Equation (16) based on vector coordinates. Figure 8 presents angles between vectors from biplots using conventional CCFs representing temporal and spatial differences in various environmental factors (e.g., meteorological, nutrient, and water quality factors). Figure 8(a) indicates that the spatial difference between upstream and downstream of the weir is minor and the relationship due to nutrient is negligible. The location differences between Kwangsan and Najoo are not significant. Figure 8(b) shows that the temporal difference between upstream and downstream of the weir is significant for the meteorology factor. Except for the meteorology factor, the angle differences between pre and post weir construction are less than 20° degrees, which is small considering the maximum possible 180° degrees in a biplot.

Figure 9 presents the angle differences for spatial and temporal perspectives of total, hydrometeorological, water quality, and nutrient factors. The angle differences between Kwangsan and Najoo (Figure 9(a)) are similar to those for Figure 8(a). However, the angle differences from the weir construction (pre and post) were substantial (Figure 9(b)). The angles before the weir construction range from 32.9° to 81.3°, while those after the weir construction range from 9.4° to 15.2°. The angles after weir construction were more than two times smaller than the angles before weir construction for all environmental factors.
characteristics. Large angles before the weir construction indicate distinct characteristics for all environmental factors, and small angles after weir construction can be explained by the similarity in causality between different environmental factors.

CONCLUSIONS

In this study, a systematic procedure is proposed to configure a prewhitened causality structure of Chl-α concentration considering multiple correlated stochastic structures of water quality data. Observations of hydrological and water quality variables and algae concentration at upstream and downstream points of the constructed weir at the Yeongsan River in South Korea are used in this prewhitened cross-correlation study. Performance comparisons between a conventional cross-correlation approach and the prewhitening method demonstrate advantages in using the latter method, both in determining important influencing factors for model prediction as well as in the configuration of controls for water quality management. Water temperature, oxygen related parameters, and phosphorus showed strong concurrent relationships to Chl-α concentration during the post river regulation period. The construction of the weir was responsible for significant causalities between hydro-meteorological components or nitrogen in negative lags and algae concentration. By taking spatial and temporal water system characteristics as factors in biplots, the introduction of factor analysis into the datasets of prewhitened cross-correlations can express the multi-relationship among the changes of water system characteristics clearly, while the traditional method fails to identify the impact of the weir in causality analysis. According to the axes and coordinates in the factor analysis biplots, the impact of hydro-meteorological factors on algal blooming was smaller than that of the river regulation project. The weir construction introduces a substantial similarity in the causality structures between nutrient or water quality factors and algal bloom for both the upstream and downstream points. Evaluation of angle differences between biplot vectors of prewhitened CCFs provides quantitative insight between different environmental factors and Chl-α concentration.

ACKNOWLEDGEMENTS

This study was financially supported by the research project (2015-01-02-033) from Institute of Environmental Research, Ministry of Environment.

REFERENCES

Basamba, T. A., Sekabira, K., Kayombo, C. M. & Ssegawa, P. 2013 Application of factor and cluster analysis in the assessment of sources of contaminants in borehole water in Tanzania. Journal of Environmental Study 22 (2), 337–346.

Box, G. & Cox, D. R. 1964 An analysis of transformations. Journal of the Royal Statistical Society (Series B) 26 (2), 211–252.

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

Cha, Y., Park, S. S., Kim, K., Byeon, M. & Stow, C. A. 2014 Probabilistic prediction of cyanobacteria abundance in a Korean reservoir using a Bayesian Poisson model. Water Resources Research 50 (3), 2518–2532. DOI: 10.1002/2013WR014372.

Choi, Y. S. & Byun, H. S. 1996 Resistant singular value decomposition and its statistical applications. Journal of the Korea Statistical Society 25 (1), 67–80.

Chu, T. C. & Rienzo, M. J. 2013 Bloom forming cyanobacteria and other phytoplankton in northern New Jersey freshwater bodies. In: International Perspectives on Water Quality Management and Pollutant Control (N. W. T. Quinn, ed.). Intech Open, London, UK, pp. 1–23.

Coad, P., Cathers, B., Ball, J. E. & Kadiuczka, R. 2014 Proactive management of estuarine algal blooms using an automated monitoring buoy coupled with an artificial neural network. Environmental Modelling and Software 61, 395–409. DOI: 10.1016/j.envsoft.2014.07.011.

Davis, T. W., Berry, D. L., Boyer, G. L. & Gobler, C. J. 2009 The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms. Harmful Algae 8, 715–725.

Gibbs, R. J. 1977 Transport processes in lakes and oceans. In: Marine Science (R. J. Gibbs, ed.), Vol. 7. Springer, Boston, MA, USA, pp. 183–186.

Hamilton, G., McVinish, R. & Mengersen, K. 2009 Bayesian model averaging for harmful algal bloom prediction. Ecological Applications 19 (7), 1805–1814.

Hardy, J. T. 1973 Phytooneuston ecology of a temperate marine lagoon. Limnology and Oceanography 18, 525–533.

Heiskary, S. & Markus, H. 2001 Establishing relationships among nutrient concentrations, phytoplankton abundance, and biochemical oxygen demand in Minnesota, USA, rivers. Journal of Lake and Reservoir Management 17 (4), 251–262.
Kim, S. 2002 Configuration of hydrological processes in a steep hillslope through causality analysis of soil moisture measurements. *Hydrological Processes* 26, 3434–3447.

Kim, M. S., Chung, Y. R., Suh, E. H. & Song, W. S. 2002 Eutrophication of Nakdong river and statistical analysis of environmental factors. *Algae* 17 (2), 105–115.

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

Kormas, K. A., Gkelis, S., Vardaka, E. & Gouni, M. M. 2010 Morphological and molecular analysis of bloom-forming cyanobacteria in two eutrophic, shallow Mediterranean lakes. *Linnmologica* 41, 167–173.

Lee, J. H. W., Huang, Y., Dickman, M. & Jayawadema, A. W. 2005 Neural network modeling of coastal algal blooms. *Ecological Modelling* 159, 179–201.

Lehmann, A. & Rode, M. 2001 Long term behaviour and cross-correlation water quality analysis of the river Elbe, Germany. *Water Research* 35 (9), 2153–2160.

Maier, H. R. & Dandy, G. C. 1997 Determining inputs for neural network models of multivariate time series. *Microcomputers in Civil Engineering* 12, 353–368.

Mainstone, C. P. & Parr, W. 2002 Phosphorus in river-ecology and management. *Science of the Total Environment* 282–283, 25–47.

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

Marsili-Libelli, S. 2004 Fuzzy prediction of the algal blooms in the Orbetello lagoon. *Environmental Modelling and Software* 19 (9), 799–808.

Mohamed, I., Othman, F., Ibrahim, A. I. N., Eldin, M. E. A. & Yunus, R. M. 2015 Assessment of water quality parameters using multivariate analysis for Klang river basin, Malaysia. *Environmental Monitoring and Assessment* 187, 4182.

Mutil, N. & Lee, J. H. W. 2005 Genetic programming for analysis and real time prediction of coastal algal blooms. *Ecological Modelling* 189, 363–376.

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

Organisation for Economic Co-operation and Development 2013 *OECD Compendium of Agri-Environmental Indicators*. OECD Publishing, Paris.

Park, S. B. 2012 Algal blooms hit South Korea rivers. *Nature*. doi: 10.1038/nature 2012.11221.

Reynolds, C. S. 1984 *The Ecology of Freshwater Phytoplankton*. Cambridge University Press, Cambridge, UK.

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

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

US Environmental Protection Agency 2015 *Application of BASIN/HSPF to Data-Scarce Watersheds*. EPA/600/R-15/007, Washington, DC.

Whitehead, P. G., Howard, A. & 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.

Woo, H. 2009 Korea to launch a major project on river rehabilitation. *Journal of Hydraulic Research* 47 (5), 74–75.

World Health Organization 2005 Algae and cyanobacteria in fresh water. In: *Guidelines for Safe Recreational Water Environments. Vol. 1: Coastal and Fresh Water*. World Health Organization, Geneva, pp. 136–158.

Wrigley, C. 1957 The distinction between common and specific variance in factor theory. *British Journal of Statistical Psychology* 10, 81–98.

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