Journal of Hydrology 529 (2015) 805–815

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Stressor–response modeling using the 2D water quality model and regression trees to predict *chlorophyll-a* in a reservoir system

Yongeun Park [a], Yakov A. Pachepsky [a], Kyung Hwa Cho [b], Dong Jin Jeon [c], Joon Ha Kim [c,*]

[a] USDA-ARS, Environmental Microbial and Food Safety Laboratory, 10300 Baltimore Ave., Building 173, BARC-EAST, Beltsville, MD 20705, USA
[b] School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 689-798, Republic of Korea
[c] School of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Republic of Korea

ARTICLE INFO

Article history:
Received 23 May 2015
Received in revised form 31 August 2015
Accepted 1 September 2015
Available online 8 September 2015
This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Stressor–response
Machine learning
Temperature sensitivity
Chlorophyll-a

SUMMARY

To control algal blooms, the stressor–response relationships between water quality metrics, environmental variables, and algal growth need to be better understood and modeled. Machine-learning methods have been suggested as means to express the stressor–response relationships that are found when applying mechanistic water quality models. The objective of this work was to evaluate the efficiency of regression trees in the development of a stressor–response model for *chlorophyll-a* (Chl-a) concentrations, using the results from site-specific mechanistic water quality modeling. The 2-dimensional hydrodynamic and water quality model (CE-QUAL-W2) model was applied to simulate water quality using four-year observational data and additional scenarios of air temperature increases for the Yeongsan Reservoir in South Korea. Regression tree modeling was applied to the results of these simulations. Given the well-expressed seasonality in the simulated Chl-a dynamics, separate regression trees were developed for months from May to September. The regression trees provided a reasonably accurate representation of the stressor–response dependence generated by the CE-QUAL-W2 model. Different stressors were then selected as split variables for different months, and, in most cases, splits by the same stressor variable yielded the same correlation sign between the variable and the Chl-a concentration. Compared to physical variables, nutrient content appeared to better predict Chl-a responses. The highest Chl-a temperature sensitivities were found for May and June. Regression tree splits based on ammonium concentration resulted in a consistent trend of greater sensitivity in the groups of samples with higher ammonium concentrations. Regression tree models provided a transparent visual representation of the stressor–response relationships for Chl-a and its sensitivity. Overall, the representation of relationships using classification and regression tools can be considered a useful approach to assess the state of aquatic ecosystems and effectively determine significant stressor variables.

Published by Elsevier B.V.

1. Introduction

Excessive algal growth in freshwaters is globally recognized as a detrimental phenomenon. Excess algae can hamper navigation, deplete the oxygen stock in water, obliterate water clarity, cause the appearance of toxins, result in fish kills, promote growth of invasive algae species, ruin quality of surface waters for recreational use, and substantially decrease property values. As such, the monitoring and modeling of both algal growth and its physical and chemical controls constitutes an important part of environmental protection activities. Photosynthetic pigment content, including *chlorophyll-a* (Chl-a) concentrations, are measured as a surrogate for algal biomass because the cost and time required for Chl-a measurement is less than that for measurements of algal biomass. Chl-a is a response variable that is commonly used to measure biotic productivity that reflects the nutrient enrichment of a system. It is used in current numeric US EPA-approved criteria to indicate water impairment by contaminant levels of nitrogen and phosphorus (US EPA, 2013).

Trends and fluctuations of Chl-a concentrations can be used to reflect the corresponding trends and fluctuations of both the chemical parameters of water quality and physical environmental parameters. For example, inter-annual variations of Chl-a were about one order of magnitude in Lake Taihu in China (Xu et al., 2013), with oscillations greater than 1.5 orders of magnitude that were recorded during the 'June through October' periods in Oregon (Hoilman et al., 2008). Strong seasonality in Chl-a has also been observed in a multitude of lake monitoring studies, which is partially attributable to different limiting factors (Conley et al., 2009; Elser et al., 1990;

* Corresponding author. Tel.: +82 62 715 3277; fax: +82 62 715 2434.
E-mail address: joonkim@gist.ac.kr (J.H. Kim).

http://dx.doi.org/10.1016/j.jhydrol.2015.09.002
0022-1694/Published by Elsevier B.V.

Hecky and Kilham, 1988). Nitrogen and phosphorus, for example, were both found to be limiting factors in four shallow lakes in Germany; there was a trend of phosphorus limitation in the spring, yetnitrogen or light limitations later in the year (Kolzau et al., 2014). Kim et al. (2014) cited empirical evidence that the abundance and composition of algal assemblages in Lake Erie are determined by resource competition for water temperature, irradiance, and nutrient availability (i.e., nitrogen, phosphorus).

Development of stressor–response models has been proposed as a way to quantify the relationships between Chl-a and its chemical and physical controls, and to establish data-driven numerical criteria on nutrient loads for receiving water bodies (Lamon and Qian, 2008). Because these relationships are very complex, a single statistical method that is fully applicable to a stressor–response model at a specific site (US EPA 2010), with linear regression relationships often providing relatively low or no correlations (Huszar et al., 2006; Napiórkowska-Krzebietke, et al., 2013), has yet to be developed. In lieu of the relationship complexity, the establishment of 'Chl-a-nutrient' relationships was recently attempted with more complex statistical methods, such as principal component regression (Cho et al., 2009a), Bayesian networks (Mutshinda et al., 2013), and artificial neural networks (Millie et al., 2006).

Results of the mechanistic modeling of Chl-a dynamics have also been used in attempts to develop stressor–response models. Liu et al. (2014) applied an orthogonal test analysis and linear regression to results obtained from Chl-a modeling under various scenarios to distinguish the contributions of various driving forces on the quality of lake water. Zou et al. (2010) suggested the use of artificial neural networks to simulate stressor–response relationships among water quality parameters derived from a mechanistic model.

The sensitivity of Chl-a dynamics to climate change has been studied based on mechanistic models and then generalized using stressor–response modeling. The sensitivity to climate scenarios under the same nutrient loads was investigated by Pätynen et al. (2014) who observed relatively small changes in Chl-a concentrations as temperature changed. Elliott (2012) noted that, in some models, warmer water in the spring was related to an increase in nutrient consumption by the phytoplankton community at some lakes, which, to the advantage of some nitrogen-fixing cyanobacteria, caused nitrogen limitations later in the year.

Given the complexity of the relationships between Chl-a responses and leading stressors, different limiting factors can be expected to dominate at different time periods, under different temperature regimes and nutrient levels. Therefore, site-specific dependencies between independent parameters and Chl-a concentrations may be needed to describe stressor–response relationships and sensitivities of Chl-a concentrations in a given environment. It was suggested that classification algorithms be applied to distinguish components of a database on eutrophication, in which different predictive stressor–response models may be needed (US EPA, 2010).

Regression trees are a powerful statistical methodology suitable for building predictions based on preliminary classifications. Regression trees have become widely used in a number of fields, including environmental sciences (Kuhn and Johnson, 2013), they are an efficient way to analyze and model of water quality data (Jones et al., 2013; Martin et al., 2011), and they are well-suited to identify the limiting factors (Sorrell et al., 2013). For example, Sass et al. (2008) successfully applied regression trees to understand Chl-a dynamics that were dependent on changes in precipitation and evapotranspiration in shallow lakes. However, to the best of our knowledge, regression trees have not yet been used to develop stressor–response models that can predict Chl-a concentrations.

The objectives of this work were to evaluate the efficiency of regression trees in the development of a stressor–response model for Chl-a concentrations based on results from site-specific mechanistic water quality modeling and to determine the sensitivity of these results to changes in temperature.

2. Materials and methods

2.1. Site description

The Yeongsan Reservoir (YSR), built in 1981 by damming the downstream end of the Yeongsan River, is an estuarine reservoir located in the southwestern region of Korea (Fig. 1). Located 23.5 km from the Mongtan Bridge, the Yeongsan Estuarine Dam has a surface area of 34.6 km^2 and an average depth of 10.1 m (maximum depth: 21.9 m) (Lee et al., 2009). The annual freshwater inflow to the YSR is, on average, 2.19×10^9 m^3 and the annual discharge though the dam gate is, on average, 1.65×10^9 m^3. About 500 million m^3 of water is annually used as a channel flow that supplies freshwater to the Yeongam Reservoir (Park et al., 2014). The main freshwater resource of the YSR is the Yeongsan River, which flows through vast agricultural areas (1161 km^2) and urbani zation/industrialization areas (304 km^2). The YSR was created to prevent flooding in the surrounding region, supply agricultural water, and facilitate recreation.

The water quality of the YSR has deteriorated since the dam was constructed due to the accumulation of pollutants, such as organic matter, nutrients, and heavy metals (Cho et al., 2009a; Kang et al., 2009; Ki et al., 2007). Lee et al. (2009, 2010) reported that contamination of water in the YSR has resulted in areas with hypoxic conditions, and areas of high sediment accumulations, which have led to a general decrease in biodiversity. Cho et al. (2009b) and Kim et al. (2001) further noted that the YSR maintained eutrophic conditions.

2.2. Modeling

The general flowchart of the modeling in this work is shown in Fig. 2. The stressor–response model is comprised of a water quality

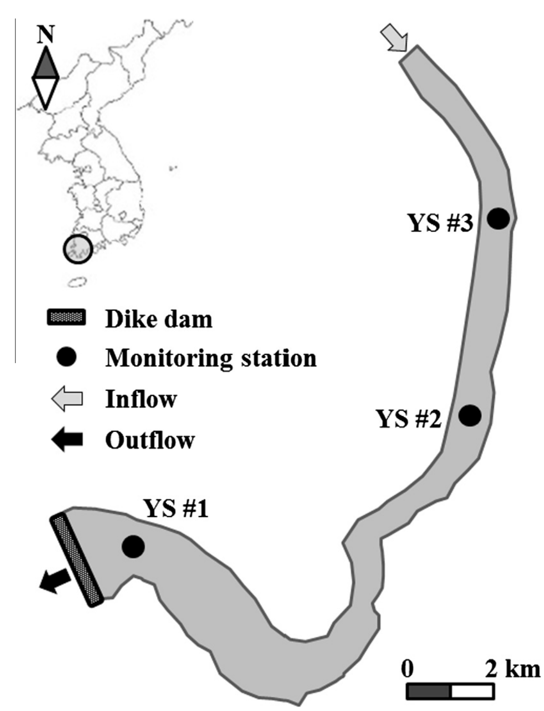

Fig. 1. Map of the Yeongsan Reservoir showing the location of water quality monitoring stations.