Ecological Engineering 68 (2014) 80–89

Contents lists available at ScienceDirect

Ecological Engineering

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

Modeling eutrophication and risk prevention in a reservoir in the Northwest of Spain by using multivariate adaptive regression splines analysis

J.R. Alonso Fernández[a,*], P.J. García Nieto[b], C. Díaz Muñiz[a], J.C. Álvarez Antón[c]

[a] Cantabrian Basin Authority, Spanish Ministry of Agriculture, Food and Environment, 33071 Oviedo, Spain
[b] Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo, Spain
[c] Department of Electrical Engineering, Electronics and Computer Systems, 33204 Gijón, Spain

ARTICLE INFO

Article history:
Received 27 November 2013
Received in revised form 28 February 2014
Accepted 31 March 2014

Keywords:
Statistical learning techniques
Eutrophication
Chlorophyll
Multivariate adaptive regression splines (MARS)
Regression analysis

ABSTRACT

The aim of this study was to obtain a predictive model able to perform an early detection of eutrophication using as predictors the chlorophyll concentration of the previous days. In this research work, the evolution of chlorophyll in the Trasona reservoir (Principality of Asturias, Northern Spain) was studied with success using the data mining methodology based on multivariate adaptive regression splines (MARS) technique. For this purpose, some biological parameters (phytoplankton species expressed in biovolume) in addition to the most important physical–chemical parameters are considered. The results of the present study are two-fold. In the first place, the significance of each biological and physical–chemical variables on the eutrophication in the reservoir is presented through the model. Secondly, a model for forecasting eutrophication is obtained. The agreement between experimental data and the model confirmed the good performance of the latter. Finally, conclusions of this innovative research work are exposed.

1. Introduction

Eutrophication is a water enrichment in nutrients that generally leads to symptomatic changes when the production of algae and other aquatic vegetations are increased, degradation of fisheries and deterioration of water quality as well as all its uses in general. Eutrophication is a peculiar form of water pollution, causing major alterations: abnormal water colorations, loss of transparency and toxicity by the presence of certain algae products (Reynolds, 2006). The eutrophication process is caused by the increase in nutrients levels, particularly phosphorus and nitrogen, which exceeds the limiting threshold of the primary production and, consequently, ecosystem control mechanisms are not used (Ansari et al., 2010). The increase in primary nutrients often leads to a great development of algal populations, mainly in illuminated water layers. These algal blooms are often noticeable by the water color. These blooms are usually due to alterations in the physical–chemical conditions in water bodies (Álvarez Cobelas and Arauzo, 2006a,b).

In recent decades, one side effect of the rapid economic development and the population growth, is an increasing tendency toward the eutrophication phenomenon in some water bodies (Ansari et al., 2010). This is now one of the most prevalent global problems of this era which affects lakes, rivers, and coastal waters.

Eutrophication has caused a variety of impacts (Mäler, 2000), such as high levels of chlorophyll *a*. A recent issue of The water wheel reported that 54% of the lakes or reservoirs in Asia are impaired by eutrophication; in Europe the number is 53%; in North America 48%; in South America 41%; and in Africa 28%. This impact of eutrophication on lakes and reservoirs has attracted attention all over the world.

Additionally, the eutrophication assessment requires data collection. These data can be collected either within the framework of a monitoring program (Kitsiou and Karydis, 2011) or as targeted projects.

Furthermore, the measurement of ecological indicators (Karydis, 2009) is one of the most popular methods for condensing ecological information. Many ecological indicators have also formed the basis of a scoring system to characterize trophic status (Spatharis and Tsirtsis, 2010), especially for the implementation of the Water Framework Directive (WFD) (Directive 2000/60/EC, 2000; Borja and Dauer, 2008).

* Corresponding author. Tel.: +34 658061580; fax: +34 985103354.
E-mail addresses: jrafernandez@chcantabrico.es, jralonsof@gmail.com (J.R. Alonso Fernández).

http://dx.doi.org/10.1016/j.ecoleng.2014.03.094
0925-8574/

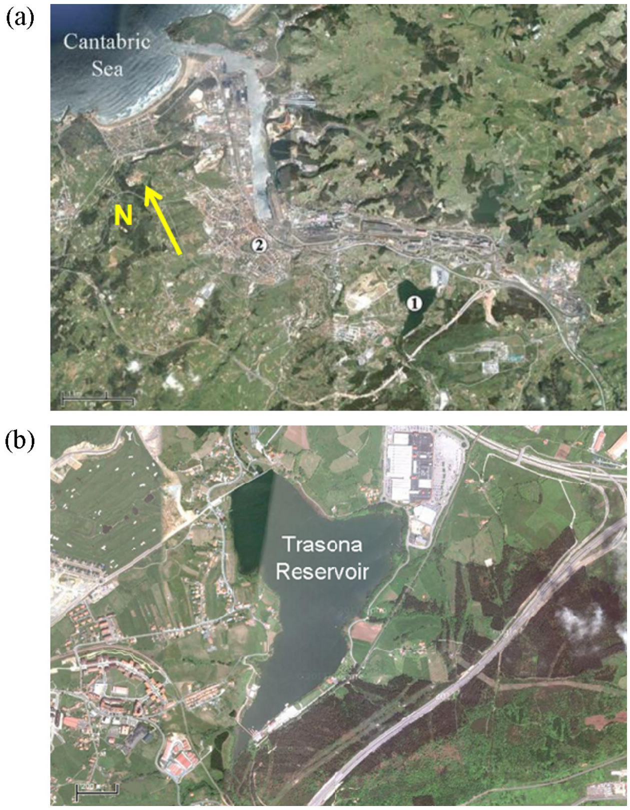

Fig. 1. (a) Aerial photograph of the Trasona reservoir (1) and the city of Avilés (Northern Spain) (2) and (b) an aerial photograph of Trasona reservoir in great detail.

In this sense, a more accurate analysis of this complex problem requires the consideration of biovolume estimations. Microalgae have a wide range of shapes and size, from submicron species such as the picoplanktonic prochlorophytes to diatoms greater than 1 mm in diameter (Hillebrand et al., 1999). Hence, cell counts are inadequate as a measure of the relative contribution of the different microalgae. In order to solve this subject, depending on the equivalent geometric shapes for each microalgae different equations sets are used to estimate their volume (Hillebrand et al., 1999).

The aim of this research is to construct a multivariate adaptive regression splines (MARS) model to identify spatial chlorophyll concentration in waterways in the Trasona reservoir (Principality of Asturias, Northern Spain) (see Fig. 1 (upper) and (lower)). Multivariate adaptive regression splines (MARS) technique is a form of regression analysis introduced by Jerome Friedman in 1991 (Friedman, 1991; Sekulic and Kowalski, 1992; Friedman and Roosen, 1995; Hastie et al., 2003). It is a non-parametric regression technique and can be seen as an extension of linear models that automatically models nonlinearities and interactions as those analyzed successfully in this innovative research work. The Trasona reservoir, which was initially destined to the industrial supply, is complemented at present with a recreational utilization as a high performance training center of canoeing. It is an eutrophic ecosystem, which has been characterized for *Cyanobacteria* outcrops in certain periods.

Eutrophication in the Trasona reservoir is a complex problem with serious effects on the state and health of the ecosystem. The consequences of the anthropogenic nutrient over-enrichment have escalated since the second half of the 20th century. Due to its characteristics, the Trasona reservoir is highly sensitive to all kind of environmental harms. Among them, eutrophication is the most serious threat due to the spreading of hypoxia and its related processes (Díaz and Rosenberg, 2011), increased harmful, or toxic, algal blooms, poorer conditions for zoobenthos and reduced fish habitats as well as the current scenarios of climate change (Karlson et al., 2002).

This innovative research work is organized as follows. Firstly, the necessary materials and methods to carry out this study are described. Secondly, the obtained results are shown and discussed. Finally, the main conclusions drawn from the results are exposed.

2. Materials and methods

2.1. Experimental data set

The dataset used for the MARS analyses was collected over 5 years (2006–2010) from a variable number of samples in Trasona reservoir. The time distribution of this dataset depends on the blooms occurrence. In this way, four samples per month were collected in July, August, September and October of 2006, 2007 and 2008 (months with blooms occurrence), respectively. However, in November of 2006, 2007 and 2008, and in June of 2007 and 2008, the sampling was carried out three times per month, respectively. However, the reservoir was sampled twice per month for the remaining months. Therefore, 33 samples were collected in 2006, 34 samples in 2007, 34 samples in 2008, 24 samples in 2009 and 24 samples in 2010.

The total number of real data processed was 149 (see Appendix A). The information is quantitative on the abundance of phytoplankton species. Specifically, this reservoir was sampled several times a month from January 16, 2006 to December 29, 2010, following the sampling protocols for lakes and reservoirs of the Spanish Ministry of Environment and Rural and Marine Affairs, which are consistent with the guidelines established by the European Union and international agencies dealing with these issues (Fogg et al., 1973; World Health Organization, 1998; Willame et al., 2005; Reynolds, 2006; Smith et al., 2008; Huisman et al., 2010). In practice, a single point of sampling is taken into account in the place where reservoir depth is the greatest, which is determined by a depth gauge (Willame et al., 2005). The samples were taken with a Niskin hydrographic bottle at different depths in the zone corresponding to the depth of the water in the reservoir that is exposed to sufficient sunlight for photosynthesis to occur called the *euphotic zone* (Willame et al., 2005). This zone is determined from the Secchi depth (depth at which the pattern on the Secchi disk is no longer visible). This value is taken as a measure of the transparency of the water in lakes, reservoirs and oceans. In this way, the quantitation of the phytoplankton and chlorophyll was carried out in a sample composed of five homogeneous subsamples obtained with the hydrographic bottle at various equidistant depths in the euphotic zone (Quesada et al., 2004; Brönmark and Hansson, 2005; Reynolds, 2006).

The physical–chemical parameters were analyzed by an ISO 17025 acredited laboratory, following the corresponding methods in the Standard Methods for the Examination of Water and Wastewater (American Public Health Association, 2005). A quality assessment program including internal laboratory control (use of standards, blanks and replicates during analysis) as well as analysis of blanks, replicates and blind samples collected in the Trasona reservoir was applied. During the sampling procedure, field blanks were also collected. A total of 10% of samples were replicated to assess variability.

The main goal of this research work was to obtain the dependence relationship of chlorophyll concentration of the Trasona reservoir (output variable), expressed in micrograms per liter, as a function of the following seven biological and fifteen