Journal of Hazardous Materials 195 (2011) 414–421

Contents lists available at SciVerse ScienceDirect

Journal of Hazardous Materials

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

Study of cyanotoxins presence from experimental cyanobacteria concentrations using a new data mining methodology based on multivariate adaptive regression splines in Trasona reservoir (Northern Spain)

P.J. Garcia Nieto[a,*], F. Sánchez Lasheras[b], F.J. de Cos Juez[c], J.R. Alonso Fernández[d]

[a] Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo, Spain
[b] Department of Construction and Manufacturing Engineering, University of Oviedo, 33204 Gijón, Spain
[c] Mining Exploitation and Prospecting Department, University of Oviedo, 33004 Oviedo, Spain
[d] Water Planning Office, Cantabrian Basin Authority, 33071 Oviedo, Spain

ARTICLE INFO

Article history:
Received 9 June 2011
Received in revised form 13 August 2011
Accepted 22 August 2011
Available online 27 August 2011

Keywords:
Statistical learning techniques
Cyanobacteria
Cyanotoxins
Multivariate adaptive regression splines (MARS)

ABSTRACT

There is an increasing need to describe cyanobacteria blooms since some cyanobacteria produce toxins, termed cyanotoxins. These latter can be toxic and dangerous to humans as well as other animals and life in general. It must be remarked that the cyanobacteria are reproduced explosively under certain conditions. This results in algae blooms, which can become harmful to other species if the cyanobacteria involved produce cyanotoxins. In this research work, the evolution of cyanotoxins in Trasona reservoir (Principality of Asturias, Northern Spain) was studied with success using the data mining methodology based on multivariate adaptive regression splines (MARS) technique. The results of the present study are two-fold. On one hand, the importance of the different kind of cyanobacteria over the presence of cyanotoxins in the reservoir is presented through the MARS model and on the other hand a predictive model able to forecast the possible presence of cyanotoxins in a short term was obtained. The agreement of the MARS model with experimental data confirmed the good performance of the same one. Finally, conclusions of this innovative research are exposed.

1. Introduction

Cyanobacteria also known as blue-green algae, blue-green bacteria, and cyanophyta is a phylum of bacteria that obtain their energy through photosynthesis. Cyanobacteria can be found in almost every conceivable environment: in oceans, lakes and rivers as well as on land. Even they flourish in Arctic and Antarctic lakes [1], hotsprings and wastewater treatments plants. Aquatic cyanobacteria is probably best known for the extensive and highly visible blooms that can form in both freshwater and the marine environment. The association of toxicity with such blooms has frequently led to the closure of recreational waters when blooms are observed. Some cyanobacteria produce toxins, called *cyanotoxins* [2], and in freshwater ecosystems are the most common cause of eutrophication. The blooms are not always green [3]. They can be blue, and some cyanobacteria species are coloured brownish-red. The water can become malodorous when the cyanobacteria in the bloom die.

Cyanotoxins are an important environmental problems in reservoirs [4]. Water is never perfectly clean and polluted water is also a continuing threat to human health and welfare [5]. The toxins include potent neurotoxins, hepatotoxins, cytotoxins, and endotoxins [6]. Most reported incidents of poisoning by microalgal toxins have occurred in freshwater environments, and they are becoming more common and widespread [7].

Generally these blooms are harmless, but if not they are called harmful algal blooms (HABs) [8]. HABs can contain toxins which result in fish kill and can also be fatal to humans [9].

The aim of this research is to construct a multivariate adaptive regression splines (MARS) model to identify spatial cyanotoxins in waterways in the Trasona reservoir (Principality of Asturias, Northern Spain) (see Fig. 1(a) and (b)). Multivariate adaptive regression splines (MARS) technique is a form of regression analysis introduced by Friedman in 1991 [10–13]. It is a non-parametric regression technique and can be seen as an extension of linear models that automatically models non-linearities and interactions as those analyzed in this innovative research work successfully. The Trasona reservoir, which was initially destined to the industrial supply, is complemented at present with a recreational utilization as a high performance training centre of canoeing. It is an eutrophic ecosystem, which has been characterized for cyanobacteria

* Corresponding author. Tel.: +34 985 103417; fax: +34 985 103354.
E-mail address: lato@orion.ciencias.uniovi.es (P.J. Garcia Nieto).

0304-3894/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2011.08.061

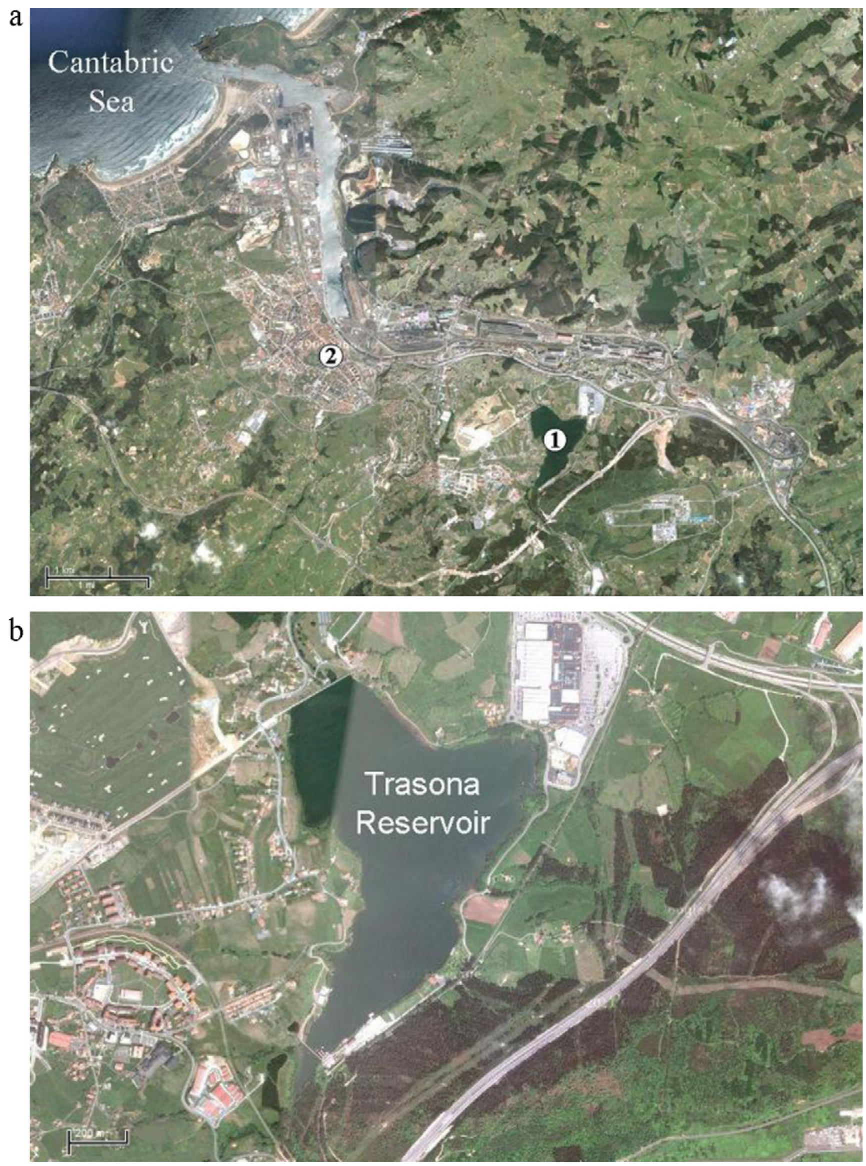

Fig. 1. (a) Aerial photograph of the city of Avilés (Northern Spain) (2) and Trasona reservoir (1); and (b) an aerial photograph of Trasona reservoir in great detail (lower).

outcrops in certain periods, which sometimes has produced variable concentrations of cyanotoxinas, mainly *mycrocistins*.

This innovative research work is structured as follows. In the first place, the necessary materials and methods are described to carry out this study. Next 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 data used for the MARS analysis were collected over five years (2006–2010) from lots of samples in Trasona reservoir and the total number of data processed was about five hundred and eleven values. The supplementary site-specific experimental data associated with this article can be found at the following online link: http://dl.dropbox.com/u/36679320/Trasona_reservoir_data.xls. The information is quantitative on the abundance of phytoplankton species. Specifically, this reservoir was sampled several times a month from January 1, 2006 to December 31, 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 [4–9]. In practice, a single point of sampling is taken into account in the place of greater depth of the reservoir, which is determined with a depth gauge [9]. The samples were taken with a Niskin hydrographic bottle (see Fig. 2(a)) 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* [5]. This zone is determined from the Secchi depth which is the depth at which the pattern on the Secchi disk (see Fig. 2(b)) is no longer visible and it is taken as a measure of the transparency of the water in lakes, reservoirs and oceans. The values of phytoplankton and concentrations of cyanotoxins and chlorophyll were determined from a sample composed of five homogeneous subsamples obtained with the hydrographic bottle at various equidistant depths in the euphotic zone [14–16].

The main goal of this research work is to obtain the dependence relationship of the cyanotoxins (output variable) of the Trasona reservoir as a function of the following input variables [17]: