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

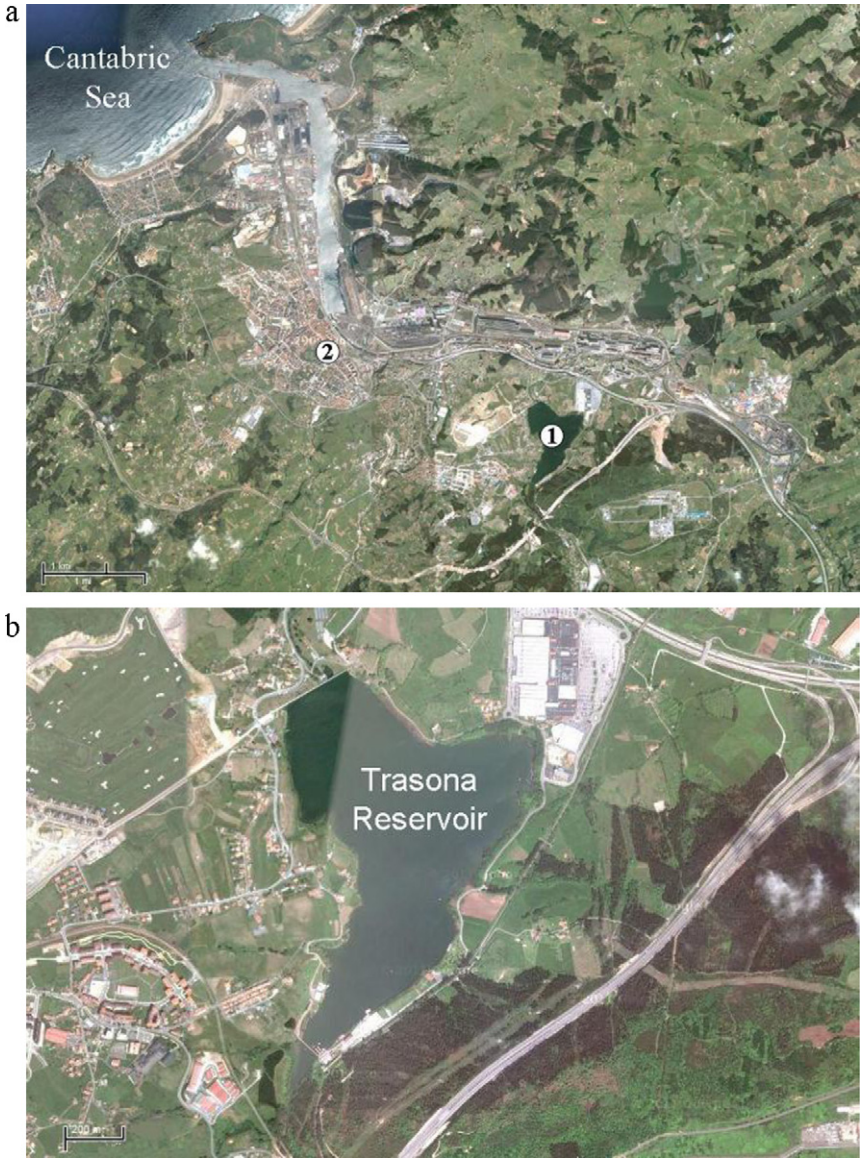

Fig. 1. (a) Aerial photograph of the city of Avilés (Northern Spain) (2) and Trasona reservoir (1); and (b) an aerial photograph of Trasona reservoir in great detail (lower).

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]:

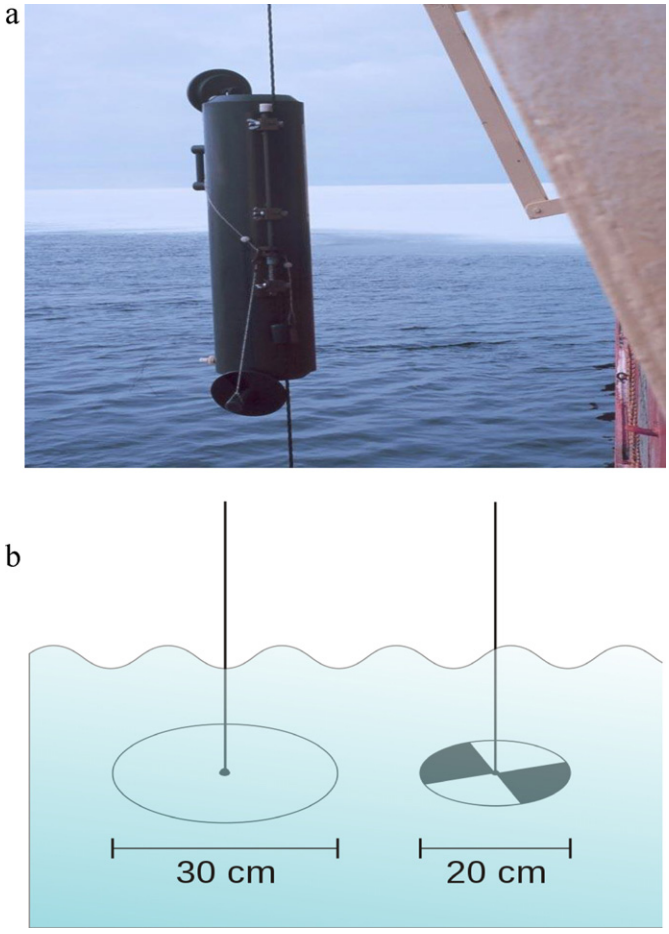

Fig. 2. (a) A Niskin hydrographic bottle about to be lowered into the water; and (b) different kinds of Secchi disks.

- *Microcystis aeruginosa*: Is a type of harmful blue-green algae which is also referred to as colonial cyanobacteria.
- *Woronichinia naegeliana*: Is a kind of cyanobacteria present in waters of a lower trophic status.
- *Other cyanobacteria*: They represent the rest of cyanobacteria excluding the two previous ones.
- *Diatoms*: Are a major group of algae, and are one of the most common types of phytoplankton.
- *Chrysophytes*: Are small flagellates that are a yellowish brown colour. They can also be found singly or in a colony.
- *Chlorophytes*: Refer to a highly paraphyletic group of all the green algae within the green plants.
- *Other species of the phytoplankton*: They represent the rest of the phytoplankton excluding all the previous ones.

All the input variables are measured in number of cells per milliliter and the output variable (cyanotoxins) in micrograms per liter. To fix ideas, in this research work, physical–chemical parameters normally used in limnological studies have been measured [7]. Analyses of chlorophyll have been carried out to study of the phytoplankton. Fig. 3(a) shows the evolution of chlorophyll concentration and cyanobacteria cell number per milliliter in the Trasona reservoir from January of 2006 to December of 2010. Higher levels of both variables are observed at certain periods of the years 2006–2008, which are significantly greater than the values obtained in the years 2009 and 2010. The peaks in Fig. 3(a) correspond to the cyanobacteria blooms: summer and fall of those years. However, there are no cyanobacteria blooms in years 2009 and 2010. Fig. 3(b) shows the evolution of cyanotoxins concentration and cyanobacteria cell number per milliliter in the Trasona reservoir from January of 2006 to December of 2010. Similarly, the peaks in Fig. 3(b) correspond to the cyanobacteria blooms and large concentrations of cyanotoxins.

Specifically, cyanobacteria cell number per milliliter was less than 50,000 and cyanotoxins concentration was always zero in 2009 and 2010.

In fact, the Trasona reservoir is an eutrophic ecosystem [15] which has been characterized for the presence of cyanobacteria. These last ones sometimes have produced variable concentrations of cyanotoxins, mainly microcystins [16]. Microcystins are cyclic nonribosomal peptides produced by cyanobacteria. They are cyanotoxins and can be very toxic for plants and animals including humans [17]. Their hepatotoxicity may cause serious damage to the liver [18]. Once the problem has been identified, civil works have been carried out in order to diminish the nutrients contributions to the reservoir although a part of spillages still reaches the same one. The guideline values for safe recreational water quality raises *alert level 2* [19] with values greater than 100,000 cells per milliliter and a microcystin concentration greater than 20.0 μg/l (see Fig. 3(a) and (b)).

The inventories of cells were taken through an inverted microscope on settled samples. The cyanotoxins have been analyzed by means of the high-performance liquid chromatography (HPLC) technique [20]. High-performance liquid chromatography (or high-pressure liquid chromatography) is a chromatographic technique that can separate a mixture of compounds and is used in biochemistry and analytical chemistry to identify, quantify and purify the individual components of the mixture. With the HPLC technique, a pump (rather than gravity) provides the higher pressure required to move the mobile phase and analyte through the densely packed column. The increased density arises from smaller particle sizes. This allows for a better separation on columns of shorter length when compared to ordinary column chromatography.

The Trasona reservoir is located near the industrial city of Avilés (Principality of Asturias, Northern Spain). Practically chained to the Trasona reservoir, it is possible to observe a wetland created artificially in order to shelter one changeable aquatic avifauna. This lagoon is able to store approximately 50,000 m^3 of water and the almost constant level of the water sheet of this lagoon allows the building of nests of different species of birds. Both the Trasona reservoir and the wetland belong to a ZEPA (zone of special protection for the birds) area [21–23].

2.2. Multivariate adaptive regression splines (MARS) method

Multivariate adaptive regression splines (MARS) is a multivariate nonparametric classification/regression technique introduced by Friedman [10–13,24,25]. The theoretical model that is explained below has already been presented by the authors in previous researches [26,27]. In spite of this fact and due to its interest for the reader in order to achieve a full understanding of the research that is presented in this paper. Its main purpose is to predict the values of a continuous dependent variable, $\vec{y}(n \times 1)$, from a set of independent explanatory variables, $\vec{X}(n \times p)$. The MARS model can be represented as:

$$\vec{y} = f(\vec{X}) + \vec{e} \quad (1)$$

where f is a weighted sum of basis functions that depend on $\vec{X}$ and $\vec{e}$ is an error vector of dimension $(n \times 1)$.

MARS does not require any a priori assumptions about the underlying functional relationship between dependent and independent variables. Instead, this relation is uncovered from a set of coefficients and piecewise polynomials of degree q (basis functions) that are entirely "driven" from the regression data $(\vec{X}, \vec{y})$. The MARS regression model is constructed by fitting basis functions to

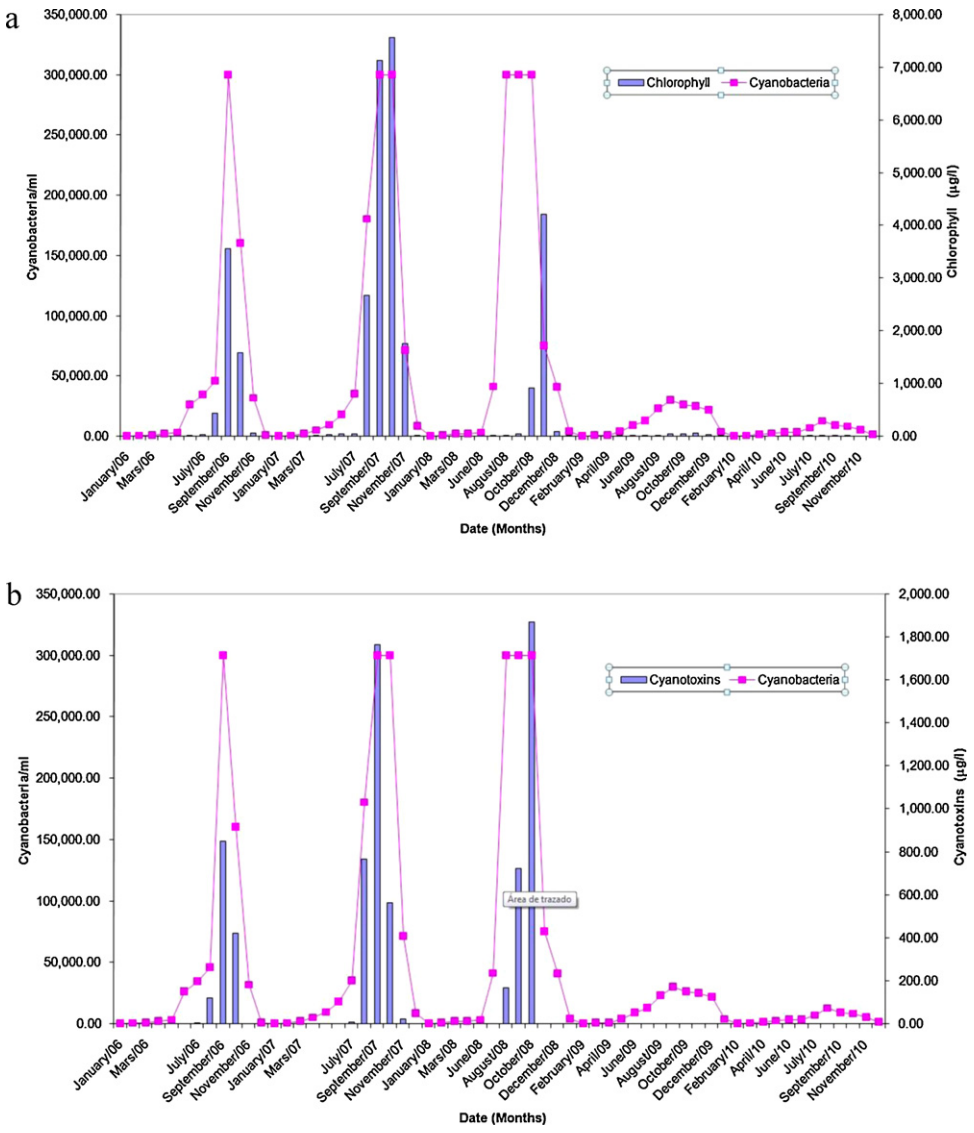

Fig. 3. (a) Evolution of chlorophyll concentration and cyanobacteria cell number per milliliter as a function of time in the Trasona reservoir from January of 2006 to December of 2010; and (b) evolution of cyanotoxins concentration and cyanobacteria cell number per milliliter as a function of time in the Trasona reservoir from January of 2006 to December of 2010.

distinct intervals of the independent variables. Generally, piecewise polynomials, also called splines, have pieces smoothly connected together. In MARS terminology, the joining points of the polynomials are called knots, nodes or breakdown points. These will be denoted by the small letter t. For a spline of degree q each segment is a polynomial function. MARS uses two-sided truncated power functions as spline basis functions, described by the following equations [10–13]:

$$[-(x-t)]_+^q = \begin{cases} (t-x)^q & \text{if } x < t \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$[+(x-t)]_+^q = \begin{cases} (t-x)^q & \text{if } x \geq t \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $q(\geq 0)$ is the power to which the splines are raised and which determines the degree of smoothness of the resultant function estimate. When $q = 1$, which is the case in this study, only simple linear splines are considered.

The MARS model of a dependent variable $\vec{y}$ with M basis functions (terms) can be written as [24–27]:

$$\hat{\vec{y}} = \hat{f}_M(\vec{x}) = c_0 + \sum_{m=1}^{M} c_m B_m(\vec{x}) \quad (4)$$

where $\hat{\vec{y}}$ is the dependent variable predicted by the MARS model, c_0 is a constant, $B_m(\vec{x})$ is the mth basis function, which may be a single spline basis functions, and c_m is the coefficient of the mth basis functions.

Both the variables to be introduced into the model and the knot positions for each individual variable have to be optimized. For a data set $\vec{X}$ containing n objects and p explanatory variables, there are $N = n \times p$ pairs of spline basis functions, given by Eqs. (2) and (3), with knot locations x_{ij} ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, p$).

A two-step procedure is followed to construct the final model. First, in order to select the consecutive pairs of basis functions of the model, a two-at-a-time forward stepwise procedure is implemented [25,28,29]. This forward stepwise selection of basis function leads to a very complex and overfitted model. Such a model, although it fits the data well, has poor predictive abilities

Table 1
Set of input variables used in this study.

Input variables (cell/ml)	Name of the variable
Microcystis aeruginosa	Microcystis_aeruginosa
Woronichinia naegeliana	Woronichinia_naegeliana
Other cyanobacteria	Other_species_Cyanobacteria
Diatoms	Diatoms
Chrysophytes	Chrysophytes
Chlorophytes	Chlorophytes
Other species of the phytoplankton	Other_phyto
Microcystis aeruginosa × Woronichinia naegeliana (synergistic interaction variable)	Microcys_×_Worochinia

for new objects. To improve the prediction, the redundant basis functions are removed one at a time using a backward stepwise procedure. To determine which basis functions should be included in the model, MARS utilizes the generalized cross-validation (*GCV*) [10–13,25,30,31]. In this way, the *GCV* is the mean squared residual error divided by a penalty dependent on the model complexity. The *GCV* criterion is defined in the following way [10–13,25]:

$$GCV(M) = \frac{(1/n)\sum_{i=1}^{n}(y_i - \hat{f}_M(\vec{x}_i))^2}{(1 - C(M)/n)^2} \quad (5)$$

where $C(M)$ is a complexity penalty that increases with the number of basis functions in the model and which is defined as [10–13]:

$$C(M) = (M + 1) + dM \quad (6)$$

where M is the number of basis functions in Eq. (4), and the parameter d is a penalty for each basis function included into the model. It can also be regarded as a smoothing parameter. Large values of d lead to fewer basis functions and therefore smoother function estimates. For more details about the selection of the d parameter, see the references [10–13,25]. In our studies, the parameter d equals 2, and the maximum interaction level of the spline basis functions is restricted to 3.

2.3. The importance of the variables in the MARS model

Once the MARS model is constructed, it is possible to evaluate the importance of the explanatory variables used to construct the basis functions. Establishing predictor importance is in general a complex problem which in general requires the use of more than one criterion. In order to obtain reliable results, it is convenient to use the *GCV* parameter explained before together with the parameters Nsubsets (criterion counts the number of model subsets in which each variable is included) and the residual sum of squares *RSS* [32].

Table 3
Evaluation of the importance of the variables that form the model according to criteria Nsubsets, *GCV* and *RSS*.

Variable	Nsubsets	*GCV*	*RSS*
Microcys_×_Worochinia	15	100	100
Other_species_Cyanobacteria	13	51.25753	52.41336
Microcystis_aeruginosa	12	23.66865	24.81192
Woronichinia_naegeliana	10	12.10512	13.28461

3. Analysis of results and discussion

The list of input variables taken into account is the research work shown in Table 1 [33–35]. As it can be observed one of the variables is formed by the product of the variable *M. aeruginosa* multiplied by the variable *W. naegeliana* due to the coexistence of these two species of cyanobacteria in order to reproduce their dynamics without interference from external factor. This mathematical formulation adds a multiplicative additional term to account for the two species' interactions according to a more realistic mathematical modelling in Biology [36,37]. This kind of interaction (synergistic interaction) will be explained later in more detail. All the input variables are measured in number of cells per milliliter and the output variable (cyanotoxins) in micrograms per liter. The total number of prediction variables used to build the MARS model was 8.

In this work, a second-order MARS model has been used, so that the basis functions of the model consist of linear and second-order splines and the maximum number of terms was not limited (no pruning). The results of the MARS model computed using all the available data observations is shown in Table 2. Table 2 shows a list of the 16 main basis functions of the MARS models and their coefficients. Please note that $h(x) = x$ if $x > 0$ and $h(x) = 0$ if $x \leq 0$. Therefore, the MARS model is a form of non-parametric regression technique and can be seen as an extension of linear models that automatically models non-linearities and interactions as a weighted sum of basis functions called *hinge functions* [10–13]. The predicted response or cyanotoxins presence is now a better fit to the original values since the MARS model has automatically produced a kink in the predicted dependent variable to take into account non-linearities. A graphical representation of the terms that constitute the model can be seen in Fig. 4.

In this research work, the fitted MARS model has a coefficient of determination R^2 equal to 0.84 and a correlation coefficient equal to 0.91. These results indicate an important goodness of fit, that is

Table 2
List of basis functions of the MARS model and their coefficients c_i.

B_i	Definition	c_i
B_1	1	1.8×10^{3}
B_2	h(0,Microcystis_aeruginosa-135,000)	0.042
B_3	h(0,135,000-Microcystis_aeruginosa)	0.025
B_4	h(0,Microcys_×_Worochinia-16,900)	1.9
B_5	h(0,16,900-Microcys_×_Worochinia)	0.011
B_6	h(0,Microcystis_aeruginosa-135,000) × h(0,Woronichinia_naegeliana-110,000)	-5.2×10^{-6}
B_7	h(0,Microcystis_aeruginosa-135,000) × h(0,110,000-Woronichinia_naegeliana)	-8.6×10^{-7}
B_8	h(0,Microcystis_aeruginosa-135,000) × h(0,Woronichinia_naegeliana-120,000)	1.1×10^{-5}
B_9	h(0,Microcystis_aeruginosa-70,000) × h(0,16,900-Microcys_×_Worochinia)	3.3×10^{-6}
B_{10}	h(0,70,000-Microcystis_aeruginosa) × h(0,16,900-Microcys_×_Worochinia)	-1.5×10^{-6}
B_{11}	h(0,Microcystis_aeruginosa-140,000) × h(0,Microcys_×_Worochinia-16,900)	2.5×10^{-5}
B_{12}	h(0,Woronichinia_naegeliana-130,000) × h(0,Microcys_×_Worochinia-10,377)	-7.4×10^{-6}
B_{13}	h(0,130,000-Woronichinia_naegeliana) × h(0,Microcys_×_Worochinia-10,377)	-7.9×10^{-6}
B_{14}	h(0,Other_species_Cyanobacteria-39,617) × h(0,Microcys_×_Worochinia-10,377)	2.6×10^{-5}
B_{15}	h(0,Other_species_Cyanobacteria-60,000) × h(0,Microcys_×_Worochinia-10,377)	1.4×10^{-5}
B_{16}	h(0,60,000-Other_species_Cyanobacteria) × h(0,Microcys_×_Worochinia-10,377)	2.2×10^{-5}

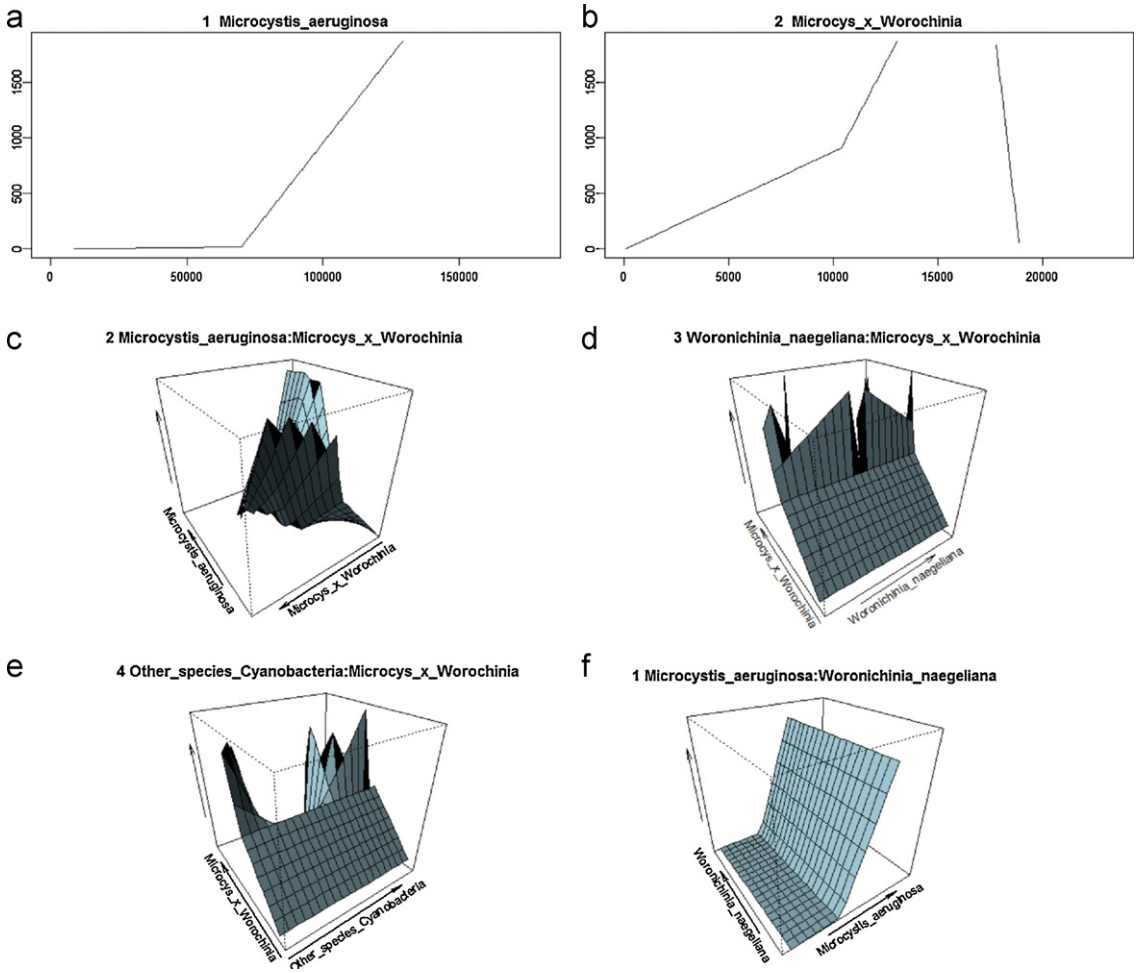

Fig. 4. Graphical representation of the terms that constitute the MARS model: (a) first order term of the variable *Microcystis aeruginosa*; (b) first order term of the product of the variables *Microcystis aeruginosa* and *Woronichinia naegeliana*; (c) second order term of the variables *Microcystis aeruginosa* and the synergistic interaction variable *Microcystis aeruginosa Woronichinia naegeliana*; (d) second order term of the variables synergistic interaction variable *Microcystis aeruginosa Woronichinia naegeliana* and *Woronichinia naegeliana*; (e) second order term of the variables synergistic interaction variable *Microcystis aeruginosa Woronichinia naegeliana* and other species of cyanobacteria; and (f) second order term of the variables *Woronichinia naegeliana* and *Microcystis aeruginosa*.

to say, a good agreement is obtained between our model and the observed data. It must be taken into account that the goodness of fit should not be considered as a proof of the predictive ability of the MARS model.

According to the results shown in Table 3, the most important variables for the prediction of the cyanotoxins (output variable) are as follows: *M. aeruginosa* multiplied by *W. naegeliana* (Microcys_×_Worochinia), Other species of cyanobacteria, the Microcystis_aeruginosa and finally the Woronichinia_naegeliana on its own.

In order to guarantee the ability prediction of the MARS model an exahustive cross-validation algorithm is used. The referred algorithm consists on the creation of 511 different MARS model (one model for each observation). Each of this model was trained using all the data except the observarion for which it was created as the validation was performed predicting its corresponding output value. The results obtained by means of this procedure are shown in Fig. 5.

The main finding of this study is the interaction between input variables *M. aeruginosa* and *W. naegeliana* not considered in previous works [38–40] and it is the result of the exhaustive work carried out on the Trasona reservoir for five years and presented here. This led to the consideration of a new input variable equal to the product of the concentrations of the two above input variables in addition to other input variables empirically measured in Trasona reservoir. The consideration of this interaction is known as *synergy* or *synergistic behavior* and it has not been considered in previous research works.

It is well known that *M. aeruginosa* is potentially toxic and produces a type of toxin known as microcystin. Up to now, there is no evidence of the toxicity of the *W. naegeliana* in Spain and there is only a partial evidence of its toxicity outside Spain [38]. This synergistic behavior is the result of joint action of two or more causes, but characterized by having a greater effect than that resulting from the sum of these causes, that is to say, the production of cyanotoxins from *M. aeruginosa* can be increased by combined presence of both species: *M. aeruginosa* and *W. naegeliana*. Synergy has been advanced as a hypothesis on how complex systems operate. Environmental systems may react in a nonlinear way to perturbations, so that the outcome may be greater than the sum of the individual component alterations. Synergistic responses are a complicating factor in environmental modelling.

Finally, this research work was able to estimate the presence of cyanobacteria blooms from 2006 to 2010 in agreement to the actual cyanobacteria blooms observed with great accurateness and success (see Fig. 5).

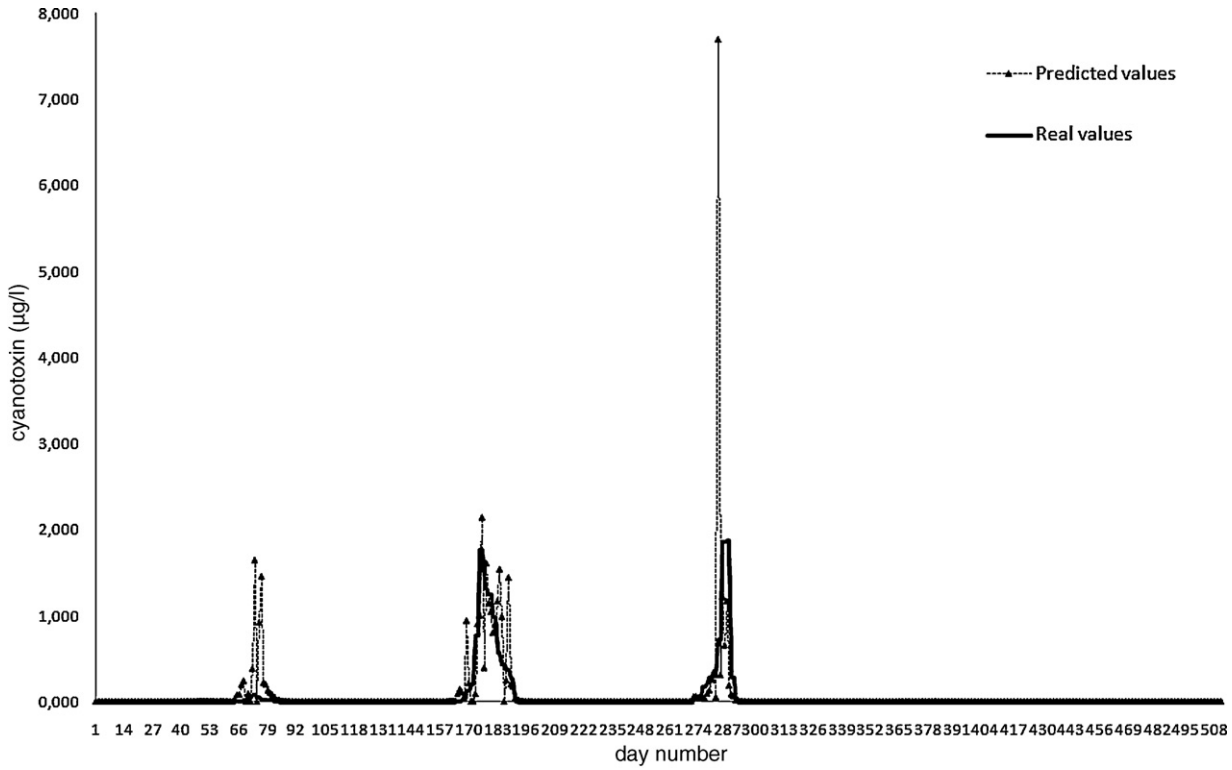

Fig. 5. Comparison between the three blooms of cyanobacteria observed and predicted by the model MARS on the Trasona reservoir from 2006 to 2010.

4. Conclusions

In the first place, the main purpose of this research was to build a cyanotoxin diagnostic model by using MARS technique in Trasona reservoir with the site-specific experimental data and this goal was achieved in this work successfully. Future researches may aim at collecting more important physical, chemical and biological variables that will increase the calculation accuracies.

Secondly, the predicted results for the MARS model have demonstrated to be consistent with the observed actual cyanobacteria blooms history from 2006 to 2010. In this way, this original and innovative methodology can be applied to other reservoirs with similar or different sources of pollutants, but it is always necessary to take into account the specificities of each location.

Finally, one of the main findings of this study is the existence of synergistic behavior between two cyanobacteria: specifically, *M. aeruginosa* and *W. naegeliana*. Synergy, in general, may be defined as two or more things functioning together to produce a result not independently obtainable. This performance is the result of joint action of the two cyanobacteria in the production of cyanotoxins: the efficiency can be increased by combined action of both species.

Acknowledgements

Authors wish to acknowledge the computational support provided by the Departments of Mathematics, Construction and Mining Exploitation at University of Oviedo as well as pollutant data in the Trasona Reservoir of Avilés (Northern Spain) supplied by the Cantabrian Basin Authority (Spanish Ministry of Environment, Rural and Marine Affairs). Furthermore, authors would like to express their gratitude to the Department of Education and Science of the Principality of Asturias for its partial financial support (Grant reference FC-11-PC10-19). The English grammar and spelling of the manuscript have been revised by a native person.

References

[1] L. Spoof, K.A. Berg, J. Rapala, K. Lahti, L. Lepisto, J.S. Metclaf, G.A. Codd, J. Meriluoto, First observation of cylindrospermopsin in *Anabaena lapponica* isolated from the boreal environment (Finland), Environ. Toxicol. 21 (6) (2006) 552–560.
[2] A. Quesada, E. Moreno, D. Carrasco, T. Paniagua, L. Wormer, C. de Hoyos, A. Sukenik, Toxicity of *Aphanizomenon ovalisporum* (cyanobacteria) in a spanish water reservoir, Eur. J. Phycol. 41 (1) (2006) 39–45.
[3] M.J. Smith, G.R. Shaw, G.K. Eaglesham, L. Ho, J.D. Brookes, Elucidating the factors influencing the biodegration of cylindrospermopsin in drinking water sources, Environ. Toxicol. 23 (3) (2008) 413–421.
[4] V. Vasconcelos, Eutrophication, toxic cyanobacteria and cyanotoxins: when ecosystems cry for help, Limnetica 25 (1–2) (2006) 425–432.
[5] M.J. Dasí, M.R. Miracle, A. Camacho, J.M. Soria, E. Vicente, Summer phytoplankton assemblages across trophic gradients in hard-water reservoirs, Hydrobiologia 369–370 (1998) 27–43.
[6] A. Dixit, R.K. Dhaked, S.I. Alam, L. Singh, Military potential of biological neurotoxins, Toxin Rev. 24 (2) (2005) 175–207.
[7] A.I. Negro, C. de Hoyos, J.C. Vega, Phytoplankton structure and dynamics in Lake Sanabria and Valparaíso reservoir (NW Spain), Hydrobiologia 424 (2000) 25–37.
[8] G.E. Fogg, W.D.P. Stewart, P. Fay, A.E. Walsby, The Blue-green Algae, London, Academic Press, 1973.
[9] C. de Hoyos, A. Negro, J.J. Aldasoro, Cyanobacteria distribution and abundance in the Spanish water reservoirs during thermal stratification, Limnetica 23 (1-2) (2004) 119–132.
[10] J.H. Friedman, Multivariate adaptive regression splines (with discussion), Ann. Stat. 19 (1991) 1–141.
[11] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer-Verlag, New York, 2003.
[12] J.H. Friedman, C.B. Roosen, An introduction to multivariate adaptive regression splines, Stat. Methods Med. Res. 4 (1995) 197–217.
[13] S.S. Sekulic, B.R. Kowalski, MARS: a tutorial, J. Chemom. 6 (1992) 199–216.
[14] C. Pérez-Martínez, P. Sánchez-Castillo, Temporal occurrence of *Ceratium hirundinella* in spanish reservoirs, Hydrobiologia 452 (1–3) (2004) 101–107.
[15] A. Quesada, D. Sanchis, D. Carrasco, Cyanobacteria in spanish reservoirs. How frequently are they toxic? Limnetica 23 (1-2) (2004) 109–118.
[16] M. Alvarez Cobelas, M. Arauzo, Phytoplankton responses to varying time scales in a eutrophic reservoir, Arch. Hydrobiol. Ergebn. Limnol. 40 (2006) 69–80.
[17] C.S. Reynolds, Ecology of Phytoplankton, Cambridge University Press, New York, 2006.
[18] I. Chorus, J. Bartram, Toxic Cyanobacteria in Water. A Guide to their Public Health Consequences, Monitoring and Management, Word Health Organization, London, 1999.

[19] World Health Organization, Guidelines for Drinking-water Quality: Health Criteria and Other Supporting Information, vol. 2, World Health Organization, Geneva, 1998.
[20] American Public Health Association, American Water Works Association, Water Environment Federation, Standard Methods for the Examination of Water and Wastewater, no. 20, APHA/AWWA/WEF, Washington, 1998.
[21] C. Brönmark, L.-A. Hansson, The Biology of Lakes and Ponds, Oxford University Press, New York, 2005.
[22] M. Scheffer, Ecology of Shallow Lakes, Springer, New York, 2005.
[23] A.G. van der Valk, The Biology of Freshwater Wetlands, Oxford University Press, New York, 2006.
[24] Q.S. Xu, M. Daszykowski, B. Walczak, F. Daeyaert, M.R. de Jonge, J. Heeres, L.M.H. Koymans, P.J. Lewi, H.M. Vinkers, P.A. Janssen, D.L. Massart, Multivariate adaptive regression splines—studies of HIV reverse transcriptase inhibitors, Chemom. Intell. Lab. 72 (1) (2004) 27–34.
[25] F.J. De Cos Juez, F. Sánchez Lasheras, P.J. García Nieto, M.A. Suárez Suárez, A new data mining methodology applied to the modelling of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women, Int. J. Comput. Math. 86 (10) (2009) 1878–1887.
[26] D. Guzmán, F.J. de Cos Juez, R. Myers, A. Guesalaga, F. Sánchez Lasheras, Modeling a MEMS deformable mirror using non-parametric estimation techniques, Opt. Express 18 (20) (2010) 21356–21369.
[27] J. De Andrés, P. Lorca, F.J. de Cos Juez, F. Sánchez-Lasheras, Bankruptcy forecasting: a hybrid approach using Fuzzy c-means clustering and multivariate adaptive regression splines (MARS), Expert Syst. Appl. 38 (3) (2011) 1866–1875.
[28] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.
[29] V. Vapnik, Statistical Learning Theory, Wiley-Interscience, New York, 1998.
[30] F.J. de Cos Juez, P.J. García Nieto, J. Martínez Torres, J. Taboada, Analysis of lead times of metallic components in the aerospace industry through a supported vector machine model, Math. Comput. Model. 52 (7–8) (2010) 1177–1184.
[31] J.R. Leathwick, J. Elith, T. Hastie, Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions, Ecol. Model. 199 (2006) 188–196.
[32] D. Guzmán, F.J. de Cos Juez, F. Sánchez Lasheras, R. Myers, L. Young, Deformable mirror model for open-loop adaptive optics using multivariate adaptive regression splines, Opt. Express 18 (2010) 6492–6505.
[33] P.M. Gault, H.J. Marler, Handbook on Cyanobacteria: Biochemistry, Biotechnology and Applications, Nova Science Publishers, New York, 2009.
[34] B.A. Whitton, M. Potts, The Ecology of Cyanobacteria: Their Diversity in Time and Space, Springer, New York, 2000.
[35] J. Huisman, H.C.P. Matthijs, P.M. Visser, Harmful Cyanobacteria, Springer, New York, 2010.
[36] E.S. Allman, J.A. Rhodes, Mathematical Models in Biology: An Introduction, Cambridge University Press, New York, 2003.
[37] D.J. Barnes, D. Chu, Introduction to Modeling for Biosciences, Springer, New York, 2010.
[38] R. Willame, T. Jurckzak, J.F. Iffly, T. Kull, J. Meriluoto, L. Hoffman, Distribution of hepatotoxic cyanobacterial blooms in Belgium and Luxembourg, Hydrobiologia 551 (2005) 99–117.
[39] I. Chorus, J. Bartram, Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management, Spon Press, New York, 1999.
[40] J. Seckbach, Algae and Cyanobacteria in Extreme Environments, Springer, New York, 2007.