



Modelling of multi-nutrient interactions in growth of the dinoflagellate microalga *Protoceratium reticulatum* using artificial neural networks



L. López-Rosales, J.J. Gallardo-Rodríguez, A. Sánchez-Mirón, A. Contreras-Gómez, F. García-Camacho*, E. Molina-Grima

Chemical Engineering Area, University of Almería, 04120 Almería, Spain

HIGHLIGHTS

- Modelling multinutrient interactions in microalgal growth using neural networks.
- Feedforward backpropagation neural network works well with two training algorithms.
- Garson's algorithm is a valuable tool to analyze relative importance of nutrients.

ARTICLE INFO

Article history:

Received 11 June 2013

Received in revised form 27 July 2013

Accepted 29 July 2013

Available online 6 August 2013

Keywords:

Microalga

Dinoflagellate

Artificial neural network

Growth modelling

ABSTRACT

This study examines the use of artificial neural networks as predictive tools for the growth of the dinoflagellate microalga *Protoceratium reticulatum*. Feed-forward back-propagation neural networks (FBN), using Levenberg–Marquardt back-propagation or Bayesian regularization as training functions, offered the best results in terms of representing the nonlinear interactions among all nutrients in a culture medium containing 26 different components. A FBN configuration of 26-14-1 layers was selected. The FBN model was trained using more than 500 culture experiments on a shake flask scale. Garson's algorithm provided a valuable means of evaluating the relative importance of nutrients in terms of microalgal growth. Microelements and vitamins had a significant importance (approximately 70%) in relation to macronutrients (nearly 25%), despite their concentrations in the culture medium being various orders of magnitude smaller. The approach presented here may be useful for modelling multi-nutrient interactions in photobioreactors.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Artificial neural networks (ANNs) have been shown to be useful for a wide range of applications, ranging from energy, chemistry, biology, agriculture, food, health and medicine etc., to chemical and bioprocess engineering (Himmelblau, 2008; Paliwal and Kumar, 2009; Vats and Negi, 2013; Nath and Das, 2011; Sivapathasekaran et al., 2010). ANNs are particularly effective in the representation of highly nonlinear bioprocesses. This is the case of microorganism culture in which metabolic responses depending on multiple interactions among numerous factors, such as environmental variables, inoculum level and concentration of medium components, are involved (Babu and Rao, 2010).

The production of high-priced bioactives from dinoflagellate microalgae for different purposes requires a medium of complex composition to ensure satisfactory growth (Gallardo-Rodríguez et al., 2012). In a recent study, an optimized culture medium for

dinoflagellates containing up to 26 different components (macronutrients, trace elements, and vitamins) was reported (García-Camacho et al., 2011). Multiple-variable analysis of the results demonstrated that it was virtually impossible to find a suitable n -dimensional model to represent the effect of combinatorial interactions between the individual medium components on the response of the cells.

It is widely recognised that many first-principles, rule-based or regression models present important restrictions for generalization, or even fail, when highly complex biological systems are addressed. Nevertheless, an extended predictive model for multinutrient-limiting growth under optimal environmental conditions must include interactive effects between the nutrients. Thus, concentration intervals for the medium components covering the range from zero to the highest relevant concentration should be incorporated into the framework of experimental input variables to obtain a model that simulates the largest number of extreme medium compositions that cells may be exposed to. When the number of nutrients is high, the use of mechanistic models is essentially ruled out. Although models based on mechanistic principles

* Corresponding author. Tel.: +34 950015303; fax: +34 950015491.

E-mail address: fgarcia@ual.es (F. García-Camacho).

reveal more details regarding the fundamentals guiding microorganism growth than ANN-based models, they are often based on a number of questionable assumptions, i.e. use of multiplicative or additive effects to take interactions into account, except for effects entirely governed by physical and/or chemical principles (Gunvig et al., 2013). In contrast, ANNs are adaptive and are able to provide a model regardless of the type of interactions concerned. Moreover, several methods for understanding the mechanics of ANNs applied in ecological sciences have been comprehensively reviewed, thus shedding light on the black box label assigned to ANNs until recently (Olden and Jackson, 2002).

No microalgal growth modelling studies that address the complete interaction of all the culture medium components have been published to date. Indeed, many of the existing models have been developed using a reduced number of components, mainly macronutrients. However, micronutrients are also exhausted during culture and their premature disappearance from the broths could well modify the cell response to an unknown extent.

In this study, ANNs were evaluated as predictive tools for growth of the dinoflagellate microalga *Protoceratium reticulatum* under laboratory conditions. The results of hundreds of experiments, generated from a previous genetic algorithm-based medium optimization process, were used to develop, train, validate and test the ANNs. Thus, numerous culture medium formulations were the input variables and the corresponding growth measures relative to a control culture were the output variables of ANN models.

2. Methods

2.1. Database

A total of 544 experiments were used in this study. These experiments formed part of a previous work in which a genetic algorithm (GA) search strategy allowed us to optimize a culture medium for growing the marine dinoflagellate microalga *P. reticulatum* GG1AM. The database comprised 16 different batches (generations) of 40 experiments each. The culture and growth conditions are reported elsewhere (García-Camacho et al., 2011). Briefly, inocula were grown under a 12:12 h light–dark cycle at 18 ± 1 °C. Filter-sterilized (0.22 µm) L1 medium prepared using natural Mediterranean seawater was used as control medium in all experiments. Cells in their exponential growth phase were used as inoculum. Erlenmeyer flasks (100-mL) shaken at 40 rpm were used as growth vessels. Cultures were illuminated from the top using Philips TLD 36W/54 fluorescent lamps at an average irradiance of 200 µmol photons $m^{-2} s^{-1}$. Growth was measured by cell counts.

2.2. ANN model formulation

The ANN input variable consisted of a matrix whose columns were vectors with the concentrations of 26 components selected from common seawater-based media (i.e. ES, West&McBr (ES), ESNW, f/2, K, L1, MNK, PC, Pro99, SN, von Stosch (Grund), and Walne's). These vectors were generated during the GA application (García-Camacho et al., 2011). The target variable representative of growth was the final cell concentration of the cultures relative to the control culture ($C_{relative}$) grown in L1 medium. Input and target variables were automatically scaled to the same range [−1,1] using Matlab. This procedure is essential for the training process when the input scales are significantly different. Differences of up to eight orders of magnitude were found for the concentrations of some nutrients in the culture media compositions assayed (García-Camacho et al., 2011).

The implementation of an ANN model initially requires the selection of a number of network properties that affect performance, such as network topology, training algorithm, adaption learning function, performance function, number of hidden layers, number of neurons in each hidden layer, and transfer function. The following back-propagation training algorithms were tested: Levenberg–Marquart, Bayesian regularization, BFGS quasi-Newton, conjugate gradient with Powell–Beale restarts, conjugate gradient with Fletcher–Reeves updates, conjugate gradient with Polak–Ribière updates, gradient descent, gradient descent with adaptive learning rate, gradient descent with momentum, gradient descent with momentum and adaptive learning rate, Resilient, and scaled conjugate gradient.

The choice of design parameters for the ANNs was the result of certain widely accepted empirical rules combined with trial and error (Nagata and Chu, 2003). For example, the following rule to find the optimum number of nodes in hidden layers ensured that the network did not overfit the training data (Maier and Dandy, 2001):

$$N \leq \frac{N^{TR}}{(N^I + 1)} \quad (1)$$

where N is the upper limit for the number of hidden layer nodes, N^{TR} is the number of training samples, and N^I is the number of inputs. Eq. (1) provided a value of 14 for N ($N^{TR} = 544$; $N^I = 26$).

There is no minimum size for the experimental data set required to statistically validate a certain ANN model, and overfitting is a major problem for small data sets. A data set is considered large if the number of coefficients in the network is smaller than the total number of points in the training set, in which case there is little or no chance of overfitting. Nonetheless, the method known as early stopping by cross-validation was applied to prevent overfitting. Briefly, the input data were randomly divided so that 70% of samples were assigned to the training set (380), 15% to the validation set (82), and 15% to the test set (82). During adjustment of the ANN to the experimental data, a cross-validation process continuously compares test data with validation data. To prevent bias during selection of the validation sub-dataset, different test sub-sets are selected each time and training is carried out again (Almeida, 2002). The number of validation checks, representing the number of successive iterations that the validation performance fails to decrease, was fixed at 30. Generalization by regularization was done automatically when the Bayesian regularization training function was used. Bayesian regularization does not require a validation data set to be separate from the training data set. The mean squared error (MSE) and correlation coefficients (R) for training, testing, and validation sets, and the best validation performance, provided by MATLAB, were mainly used to evaluate the ANN models.

2.3. Indices of ANN performance

Predicted ($C_{relative}^p$) and observed ($C_{relative}^o$) relative growth values were compared by calculating the bias (BF) and accuracy (AF) factors, as reported previously (Ross, 1996) when modelling the growth of food microorganisms:

$$\log(BF) = \frac{\sum \log\left(\frac{C_{relative}^p}{C_{relative}^o}\right)}{n}; \log(AF) = \frac{\sum \left|\log\left(\frac{C_{relative}^p}{C_{relative}^o}\right)\right|}{n} \quad (2)$$

BF indicates a systematic over- or underestimation of targets and AF is a measure of the average difference between observed and predicted $C_{relative}$ values. The following interpretation of the values of both factors has been proposed (Ross, 1996): (i) from 0.95 to 1.11, good performance; (ii) 0.87–0.95 or 1.11–1.43, acceptable; and (iii) <0.87 or >1.43, unacceptable.

Download English Version:

<https://daneshyari.com/en/article/7080990>

Download Persian Version:

<https://daneshyari.com/article/7080990>

[Daneshyari.com](https://daneshyari.com)