Estimating microalgae *Synechococcus nidulans* daily biomass concentration using neuro-fuzzy network

**Vitor Badiale FURLONG**, Renato Dutra PEREIRA FILHO, Ana Cláudia MARGARITES, Pâmela Guder GOULARTE, Jorge Alberto Vieira COSTA

**Abstract**

In this study, a neuro-fuzzy estimator was developed for the estimation of biomass concentration of the microalgae *Synechococcus nidulans* from initial batch concentrations, aiming to predict daily productivity. Nine replica experiments were performed. The growth was monitored daily through the culture medium optic density and kept constant up to the end of the exponential phase. The network training followed a full 3³ factorial design, in which the factors were the number of days in the entry vector (3, 5 and 7 days), number of clusters (10, 30 and 50 clusters) and internal weight softening parameter (Sigma) (0.30, 0.45 and 0.60). These factors were confronted with the sum of the quadratic error in the validations. The validations had 24 (A) and 18 (B) days of culture growth. The validations demonstrated that in long-term experiments (Validation A) the use of a few clusters and high Sigma is necessary. However, in short-term experiments (Validation B), Sigma did not influence the result. The optimum point occurred within 3 days in the entry vector, 10 clusters and 0.60 Sigma and the mean determination coefficient was 0.95. The neuro-fuzzy estimator proved a credible alternative to predict the microalgae growth.

**Keywords:** black-box; cellular concentration; predictive microbiology.

**Resumo**

Neste trabalho, foi construído um estimator *neuro-fuzzy* da concentração de biomassa da microalga *Synechococcus nidulans* a partir de concentrações iniciais de baetela, visando possibilitar a predição da produtividade. Nove experimentos em réplica foram realizados. O crescimento foi acompanhado diariamente pela transmittância do meio e mantido até o final da fase exponencial de crescimento. O treinamento das redes ocorreu segundo delineamento experimental 3³, os fatores foram o número de dias no vetor de entrada (3, 5 e 7 dias), o número de clusters (10, 30 e 50 clusters) e o valor de abrandamento do filtro interno (Sigma) (0,30, 0,45 e 0,60). A variável resposta foi o somatório do erro quadrático das validações. Estas possuíam 24 (A) e 18 (B) dias de crescimento. As validações demonstraram que, em experimentos de longo período (Validação A), é necessário usar poucos clusters e Sigma altos. Já, em curtos períodos (Validação B), o Sigma não gera alterações. O ponto ótimo ocorreu com 3 dias na entrada, com 10 clusters e Sigma de 0,60, cujo coeficiente de determinação médio foi 0,95. O estimator *neuro-fuzzy* mostrou-se uma alternativa robusta para predição do crescimento desta microalga.

**Palavras-chave:** black-box; concentração celular; microbiologia preditiva.

**1 Introduction**

Microalgae are widely studied due to their photosynthetic properties. Photosynthesis provides them with the ability to use carbon dioxide as carbon source, process known as dioxide biofixation, reducing CO₂ emissions to the atmosphere (CHEN et al., 2009). Biofixation enables the microalgae to accumulate carbon, which promotes the synthesis of energy storage compounds. Among these compounds are micronutrients, carotenoids, vitamins and sterols, and macronutrients such as high biological value proteins, carbohydrates, and important fatty acids with 12-22 carbon atoms, often essentials, like linoleic and linolenic acids (CHACÓN-LEE; GONZÁLEZ-MARÍNÓ, 2010).

The biomass obtained from the cultures may also be used in the production of second-generation biofuels, which are defined as those produced from non-alimentary biomass. Usually, second generation biofuels substrates are agriculture byproducts such as cereals straw, sugar cane bagasse, and generic effluents. Other substrates may also be generated biofuels production, such as modified gramineae, rapid growth forests and microalgae (SIMS et al., 2010).

Among these microalgae that may serve to the described purposes is the cyanobacteria *Synechococcus nidulans*. This species was identified over 30 years ago and is found in aquatic environments, mostly in well-lit surface water. Naturally occurring, *Synechococcus* biomass is generally abundant, with cellular densities that can vary from hundreds to millions of cells per milliliter of sea water (SIX et al., 2007).

In order to use microalgae full potential it is necessary to establish optimum culture parameters. This goal is accomplished by changing the nutritional and physical characteristics of the organism, as well as the configuration of the reactor.
Materials and methods

2.1 Microalgae culture

Nine identical experiments were performed using the microalgae *Synechococcus nidulans* in Zarrouk medium (ZARROUK, 1966) with 50% of its original nitrogen source (1.25 g.L\(^{-1}\) of NaNO\(_3\)). This species belongs to the Biochemical Engineering Laboratory collection, in the Federal University of Rio Grande. The cultures were prepared in closed 2 L photobioreactors, under constant stirring using a diaphragm pump, sterile air at a flow of 0.480 L.min\(^{-1}\) and 2500 lux luminance with 12h day/night photoperiod. The initial cellular concentration was 0.2 g.L\(^{-1}\).

Cellular growth determination

Cellular concentration was determined each 24 h by measuring the absorbance of the culture medium in spectrophotometer at 670 nm (COSTA et al., 2002). The absorbance was then correlated with a previously established dry weight standard curve at the same wavelength. The cultures were kept under the experimental conditions until the end of the exponential growth phase. This period was established by three days of similar cellular concentrations, which demonstrated that the growth had ceased.

2.2 Digital filter

To correct the noise effects deriving from the large number of factors associated with microalgae growth, which may possess a negative effect on the repeatability and accuracy of the measurement, a noise reduction stage is necessary. To deal with these characteristics, a double exponential digital filter was employed. This utilizes a series of measurements \((x_{n-1}, x_n, x_{n+1}, \ldots)\), and their corresponding filtered values \((y_{n-1}, y_n, y_{n+1}, \ldots)\), where \(n\) is the current sample value. The filter application follows Equation 1.
Estimating microalgae *Synechococcus nidulans* daily biomass concentration using neuro-fuzzy network

where \( a \) is the internal softening parameter. As this parameter value approaches a unitary value, the output filtered value is closer to the unfiltered value. The established value for this study was 0.6 (SEBORG; EDGAR; MELLICHAMP, 1989).

### 2.3 Numeric procedure

The processing procedure used in the state estimator was establishing subsequent biomass concentrations with previous points. In order to generate more stable connections between the neurons, thus leveling different period experiments and improving processing abilities, the biomass concentration values were standardized between 0 and 1 by dividing all concentrations by the highest biomass concentration throughout the experiment.

An example of this procedure is feeding the system during the firsts three days as an entry vector in order estimate the culture growth. The response of the firsts three days was the fourth day. The fourth day biomass concentration data was then fed back into the input vector after excluding the first day, thus making the entry vector the second, third and fourth days. The response for this vector is the fifth daily biomass concentration. This procedure was repeated until the estimation reached the final experiment biomass concentration.

Therefore, during the network training it was necessary to provide to the algorithm all the biomass concentrations of the training group in the described fashion. This generated a matrix containing the entry concentration in the first three columns and the response in the last column. Hence, this group did not take part in the prediction, but it was responsible for composing the system’s relations, enabling the prediction that was possible with another data set.

The validation group was used to determine the network performance and liability. To achieve the estimation, the initial days of the validation experiment was fed into the network in order to predict the following concentration in the described manner. Such fashion was repeated until the complete culture profile was estimated.

### 2.4 Topology optimization

For assigning the correct architecture to the neuro-fuzzy network, a wide variety of design alternatives must be evaluated. Therefore, an experimental design may be applied to diminish and generate simpler array of essays. In this study, a full \( 3^n \) factorial design was applied to the neuro-fuzzy network, whose parameters were: number of the days in the entry vector, number of clusters and the internal weight softening parameter (\( \sigma, \Sigma \)).

Changing the number of concentration points during the training and network utilization the number of days in the entry vector varied. Architectures were constructed with 3, 5 and 7 days in the entry vector.
Table 1. Biomass Concentrations in the Experiments.

| Time (d) | Training experiments (g.L\(^{-1}\)) | Validation experiments (g.L\(^{-1}\)) |
|----------|------------------------------------|-------------------------------------|
|          | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 24 Days | 18 Days |
| 1        | 0.20 | 0.22 | 0.20 | 0.21 | 0.21 | 0.22 | 0.21 | 0.22 | 0.22 |
| 2        | 0.26 | 0.26 | 0.32 | 0.32 | 0.30 | 0.27 | 0.29 | 0.30 | 0.29 |
| 3        | 0.28 | 0.27 | 0.34 | 0.35 | 0.33 | 0.35 | 0.34 | 0.31 | 0.36 |
| 4        | 0.31 | 0.28 | 0.36 | 0.36 | 0.35 | 0.42 | 0.40 | 0.32 | 0.42 |
| 5        | 0.34 | 0.29 | 0.38 | 0.38 | 0.36 | 0.50 | 0.49 | 0.32 | 0.49 |
| 6        | 0.38 | 0.30 | 0.42 | 0.43 | 0.40 | 0.54 | 0.57 | 0.33 | 0.52 |
| 7        | 0.41 | 0.31 | 0.45 | 0.49 | 0.41 | 0.57 | 0.64 | 0.35 | 0.56 |
| 8        | 0.45 | 0.33 | 0.47 | 0.55 | 0.41 | 0.60 | 0.68 | 0.36 | 0.59 |
| 9        | 0.49 | 0.42 | 0.49 | 0.57 | 0.42 | 0.64 | 0.73 | 0.40 | 0.64 |
| 10       | 0.51 | 0.48 | 0.52 | 0.58 | 0.44 | 0.67 | 0.81 | 0.42 | 0.69 |
| 11       | 0.52 | 0.63 | 0.53 | 0.58 | 0.45 | 0.74 | 0.86 | 0.47 | 0.74 |
| 12       | 0.52 | 0.68 | 0.55 | 0.63 | 0.46 | 0.79 | 0.86 | 0.50 | 0.79 |
| 13       | 0.53 | 0.74 | 0.59 | 0.70 | 0.47 | 0.83 | 0.86 | 0.52 | 0.83 |
| 14       | 0.56 | 0.75 | 0.61 | 0.71 | 0.48 | 0.80 | 0.86 | 0.54 | 0.86 |
| 15       | 0.60 | 0.85 | 0.59 | 0.73 | 0.49 | 0.77 |          | 0.63 | 1.01 |
| 16       | 0.62 | 0.88 | 0.58 | 0.77 | 0.46 | 0.75 |          | 0.70 | 1.05 |
| 17       | 0.70 | 0.95 | 0.56 | 0.81 | 0.45 |          |          | 0.75 | 1.06 |
| 18       | 0.74 | 1.00 |          |          |          |          |          | 0.77 | 1.06 |
| 19       | 0.76 | 1.07 |          |          |          |          |          | 0.79 |          |
| 20       | 0.78 | 1.19 |          |          |          |          |          | 0.82 |          |
| 21       | 0.84 | 1.28 |          |          |          |          |          | 0.83 |          |
| 22       | 0.97 |          |          |          |          |          |          | 0.92 |          |
| 23       | 0.97 |          |          |          |          |          |          | 0.97 |          |
| 24       | 1.01 |          |          |          |          |          |          |          | 1.02 |

Figure 1. Biomass Prediction Profiles. (a) Validation with 24 days: (■) Experimental Data, (→) 50 Clusters, Sigma 0.60, 3 Entry Vector, (→) 10 Clusters, Sigma 0.60, 5 Entry Vector (⋯) 10 Clusters, Sigma 0.60, 7 Entry Vector; (b) Validation with 14 Days, (■) Experimental Data, (→) 30 Clusters, Sigma 0.45, 3 Entry Vector; (⋯) 10 Clusters, Sigma 0.60, 5 Entry Vector (⋯) 10 Clusters, Sigma 0.60, 7 Entry Vector.
pattern recognition can be realized. This is supported by the fact that in Validation A, a network that used 3 days in the entry vector, 10 clusters and Sigma 0.60 reached the sum of quadratic error approximately half the essays with the same number of clusters and entry vector.

To perform further analysis of the experimental design is necessary to discuss the model generated by the design and its coupled response surface. Nevertheless, the calculated Fisher F-value for the regression (1.021) was smaller than the standard (2.040) for this situation (18 degrees of freedom for residual and 8 for regression); thus, this model is not predictive, precluding the model and response surface analysis.

Validation B (18 Days) otherwise presented significant parameters for the interaction, the linear cluster and days in the entry vector effects, the interactions between linear Sigma and linear clusters, linear Sigma and days in the entry vector in its quadratic portion and linear clusters and quadratic days portion. The calculated Fisher F-value for the regression (2.883) was larger than the standard (2.040); therefore, the response surface model generation process is possible. The coded model for this design is presented in Equation 2:

$$SGQ = 2.26 + 1.40 \times C + 0.81 \times D - 1.36 \times S \times C - 1.54 \times S \times D^2 - 1.80 \times C \times D^2$$

where $C$ is the number of clusters, $D$ is the number of days in the entry vector and $S$ is the internal weight softening parameter. Since the presented model poses three coefficients to enable the graphical representation, the postulation of a given parameter is needed. Given that the Sigma parameter, alone, had no influence in interaction with other variables, at this confidence level it was determined that this value would be established in the experimental design in order to construct the response surfaces of the adjacent parameters. The responses surfaces are presented in Figure 2.

Despite the profiles changes according to each Sigma factor, the surfaces analysis establishes that lower values of quadratic error are achieved primarily in small entry vector, and, as the Sigma increases, in larger number of clusters. The smallest quadratic error was found in the experiment within 3 days in the entry vector, 0.45 Sigma and 10 clusters, in agreement with the surface and effects found in short-term experiments.

Consequently, in order to contemplate both experiment periods, the number of clusters must be small, 10 clusters, since the architecture was improved in both validations. The second parameter is the use of a small entry vector, 3 days, as it did not influenced the long-term culture validation and improved the network performance in the short-term experiment. The last parameter is the use of a large Sigma factor (0.60) as it improves assertion in long-term experiments and does not influence greatly brief cultures prediction. Also the relatively large Sigma factor improves the network robustness even in the presence of large noise variance. In this situation, the mean determination coefficient ($R^2$) between the present experiment data and the estimated biomass concentrations was 0.95 for both validations.

### Table 2. Full $3^3$ Factorial Design Summary.

| Validation | Parameter | Estimated effect | Standard error | T value | Regression coefficient |
|------------|-----------|-----------------|----------------|---------|------------------------|
| A          | Intercept | 1.857           | 0.366          | 5.075   | 1.857                  |
|            | Sigma (L) | -1.265          | 0.634          | -1.996  | -0.632                 |
|            | Cluster (L)| 1.199           | 0.634          | 1.891   | 0.599                  |
| B          | Intercept | 2.262           | 0.422          | 5.36    | 2.262                  |
|            | Cluster (L)| 2.809           | 0.731          | 3.843   | 1.405                  |
|            | Days (L)  | 1.630           | 0.731          | 2.229   | 0.815                  |
|            | Sigma (L): Cluster (L)| 2.724| 1.266| 2.151| 1.365|
|            | Sigma (L): Days (Q)| -3.093| 1.266| -2.443| -1.546|
|            | Cluster (L): Days (Q)| -3.604| 1.266| -2.845| -1.801|

Validation A: 24 days experiment; Validation B: 18 days experiment. (L) Linear Effect; (Q) Quadratic Effect.

**Figure 2.** Response Surfaces for Sum of Quadratic Error with Postulated Sigma. (a) Response Surface at Sigma 0.30; (b) Response Surface at Sigma 0.45; (c) Response Surface at Sigma 0.60.
4 Conclusion

The neuro-fuzzy estimator architecture is more reliable and generates better estimated profiles when three initial biomass concentrations are used in the entry vector. The network used 10 clusters iterations under an internal weight softening parameter of 0.60. With this architecture, the network accomplished estimated profiles with a mean determination coefficient of 0.95 in relation to the present biomass concentration data.

Thus, the estimator composed in the present study is a viable alternative tool to estimate future biomass concentration of microalgae, such as *Synechococcus nidulans*, from initial batch growth data, enabling a more extensive control over the autotrophic bioprocess and an optimization in generating biomass.

Acknowledgements

The authors would like to thank Nanofotobioc Network supported by CAPES for sponsoring the equipments, as well as undergraduate and doctorate scholarships.

We also would like to thank the INRIA research group, especially Yann Collete, author of the LOLIMOT package, as well as the R Project contributors and the entire Open Source community for providing its knowledge and dedication to other peers.

References

BORDEIANU, C. C. et al. Scilab software package for the study of dynamical systems. *Computational Physics Community*, v. 178, n. 10, p. 788-793, 2008. http://dx.doi.org/10.1016/j.cpc.2008.01.002

ÇELEKLI, A.; YAVUZATMACA, M. Predictive modeling of biomass production by *Spirulina platensis* as a function of nitrate and NaCl concentrations. *Bioresource technology*, v. 100, n. 4, p. 1847-1851, 2009. PMid:18993057. http://dx.doi.org/10.1016/j.biortech.2008.09.042

CHACÓN-LEE, T. L.; GANZÁLEZ-MARINÓ, G. E. Microalgae for “healthy” food-possibilities and challenges. *Comprehensive Reviews in Food Science and Food Safety*, v. 9, n. 1, p. 655-675, 2010. http://dx.doi.org/10.1111/j.1541-4337.2010.00132.x

CHEN, P. et al. Review of the biological and engineering aspects of algae to fuels approach. *International Journal of Agriculture and Biochemical Engineering*, v. 2, n. 1, p. 1-30, 2009.

COSTA, J. A. V. et al. Modeling of *Spirulina platensis* growth in fresh water using rezone surface methodology. *World Journal of Microbiology and Biotechnology*, v. 18, n. 7, p. 603-607, 2002. http://dx.doi.org/10.1023/A:1016822717583

GADKAR, K. G.; MEHRA, S.; GOMES, J. On-line adaptation of neural network for bioprocess control. *Computers and Chemical Engineering*, v. 29, p. 1047-1057, 2005. http://dx.doi.org/10.1016/j.compchemeng.2004.11.004

GRÖMPING, U. *Tutorial for Designing Experiments Using R Package RcmdrPlugin.DoE*. Berlin: Fachbereich Ii Beuth Hochschule Für Technik Berlin, 2011. 54 p.

HU, D. et al. The design and optimization for light-algal bioreactor controller based on artificial neural network-model predictive control. *Acta Astronautica*, v. 63, p. 1067-1075, 2008. http://dx.doi.org/10.1016/j.actaastro.2008.02.008

KARAKUZU, C.; TÜKER, M.; ÖZTÜRK, S. Modelling, on-line state estimation and fuzzy control of production scale fed-batch baker’s yeast fermentation. *Control Engineering Practice*, v. 14, p. 959-974, 2006. http://dx.doi.org/10.1016/j.conengprac.2005.05.007

KHATAEE, A.R. et al. Neural network modeling of biotreatment of triphenylmethane dye solution by a green macroalgae. *Chemical Engineering Research And Design*, v. 89, p. 172-178, 2011. http://dx.doi.org/10.1016/j.cherd.2010.05.009

MARSILI-LIBELLI, S. Fuzzy prediction of the algal blooms in Orbetello lagoon. *Enviromental Modeling & Software*, v. 19, p. 799-808, 2004. http://dx.doi.org/10.1016/j.envsoft.2003.03.008

MARZOOGHI, H. et al. Dynamic modeling of solid oxide fuel cell stack based on local linear model tree algorithm. *International Journal of Hydrology Engineering*, v. 37, n. 1, p. 4367-4376, 2012. http://dx.doi.org/10.1016/j.ijhydene.2011.11.149

MASTERS, T. *Advanced Algorithms for Neural Networks – A C++ Sourcebook*. 2nd ed. New York: John Wiley & Sons Inc., 1995. 430 p.

SCHERER, C. *Métodos Computacionais da Física*. São Paulo: Editora Livraria Da Física, 2010. 358 p.

SEBORG, D. E.; EDGAR, T. F.; MELLICHAMP , D. A. *Process Dynamics and Control*. New York: John Wiley & Sons Inc., 1989. 700 p.

SIMS, R. E. H. et al. An overview of second generation biofuels technologies. *Bioresource Technology*, v. 101, n. 1, p. 1570-1580, 2010. PMid:19963372. http://dx.doi.org/10.1016/j.biortech.2009.11.046

SIX, C. et al. Diversity and evolution of phycobilisomes in marine *Synechococcus spp.*: a comparative genomics study. *Genetic Biology*, v. 8, p. R259, 2007. PMid:18062815 PMCID:2246261. http://dx.doi.org/10.1186/gb-2007-8-12-r259

TAYLOR, B. J. *Methods and Procedures for Verification and Validation of Artificial Neural Networks*. New York, United States of America: Springer Science; Business Media Inc., 2006. 277 p.

ZARROUK, C. *Contribution à l’étude dune Cyanophycée*: influence of divers facteurs physiques et chimiques sur la croissance et la photosynthèse de spirulina máxima. 1966. Tese (Doctorat)-Université Des Paris, Paris, 1966.