

Biological response neural network prediction in coastal upwelling field

G.C. Pereira¹, R. Coutinho², N.F.F. Ebecken¹.

¹ *Civil Engineering Program, University of Rio de Janeiro - COPPE-UFRJ*

² *Almirante Paulo Moreira Study Institute - IEAPM*

Abstract

Modeling water quality, forecasting population dynamics to define the ecosystem health is a function of some indices (resilience, diversity, production etc.) and yield of many environmental variables interactions. The assessment of properties and processes of costal zone is a major issue in aquatic system management. The aim of this paper is to test and analyse neural network capability in predicting changes of chlorophyll-a as a phytoplankton biomass. Available data concern to a weekly medium-term time-series ranging from the end of 1994 to 2001, and it was collected at the Arraial do Cabo near shore upwelling, southeast of Rio de Janeiro state (Brazil). Dealing in particular with high non-linearity among variables, a genetic algorithm package was used to settle the best neural architecture and the specific pre-processing method presented. The results show the spread of predictions given by the models in short, medium and long-term periods. Thus, the conclusions allow us to say that neural networks can successfully be used to improve forecasting of complex impact factors and ecological system behavior.

1 Introduction

The main subject of this paper is to understand and model how marine phytoplankton community varies at a natural eutrophication process. An important problem in aquatic ecosystem is to recognize what is natural and anthropic variability. The knowledge of biogenic elements dynamic, the biological response and the election of indicators connecting low and higher trophic levels have become a real need for a sustainable management of the marine resource in coastal zone. The Arraial do Cabo upwelling system (23°S, 42°W) is an anomaly in that it is on the west side of the Atlantic Ocean on the coast of Rio de Janeiro state, southeastern of Brazil (Fig. 1).

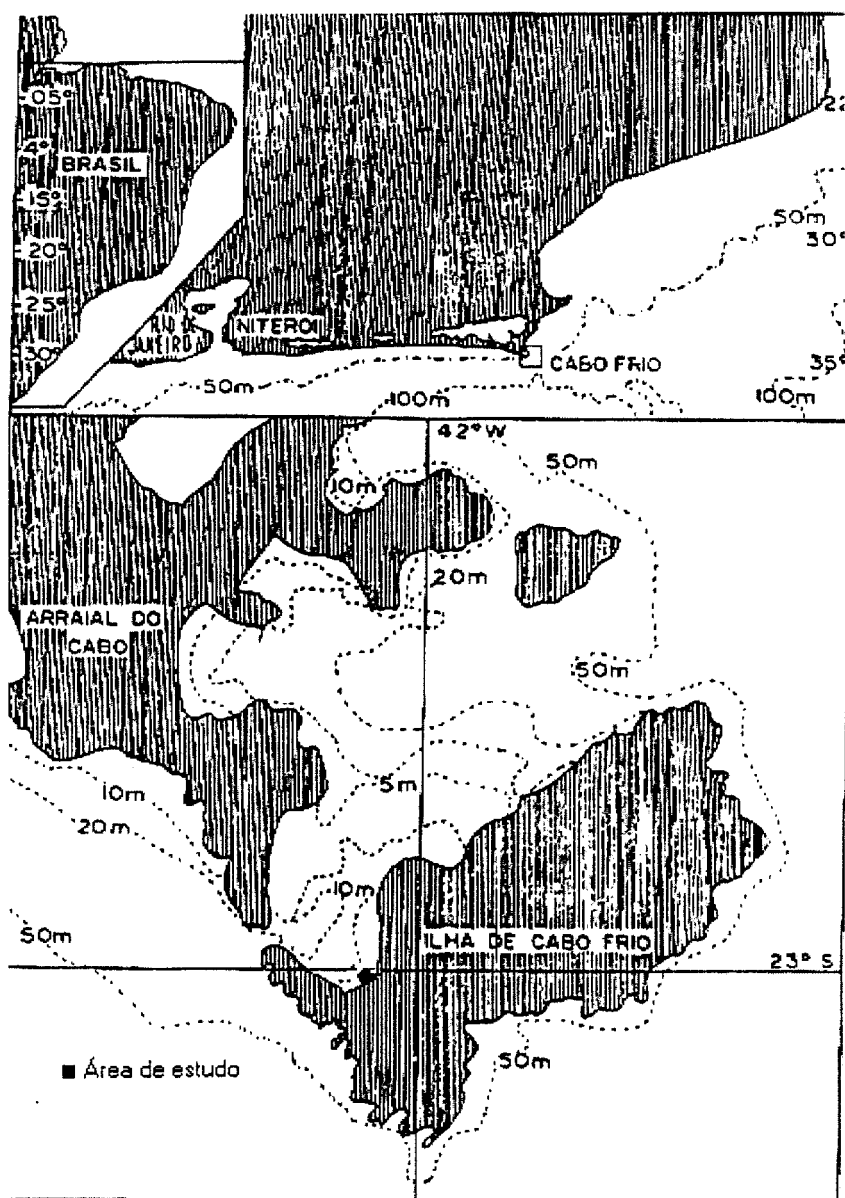


Figure 1 - Map of studied area

A change in coastal orientation from north-south to east-west at Cabo Frio region and the proximity of the 100-m isobath lead to a topography which promotes the upflow of the cold (18°C) South Atlantic Central Waters (SACW) rich in nutrients ($18\mu\text{M-NO}_3\text{-N}$ and $1.3\mu\text{M-PO}_4\text{-P}$ Gonzalez-Rodriguez et al. 1992) depending on direction and intensity of north-northeast wind. This phenomenon takes place into two steps (Moreira da Silva, 1973).

The first one begins in mid-August/September (later winter) and depends on the position of the flow of Brazil Current in relation to the continental slope and on variation in frequency and intensity of NE winds. The SACW advances beyond the continental slope and invades the shelf to depths of about 50-80 m where it remains until April (mid-fall). The second step of upwelling, mainly from September to March, is influenced by changes in intensity and direction of local winds, which cause alternating warming and cooling of near-surface waters. Two or three days of northeasterly winds (10m^{-1}) bring the deep water to the surface and cold front passages (8 to 10 days cycles), which cause the wind to shift to the west, reverse the process (downwelling). During the upwelling phase there is a thermocline formation where accumulation of organic matter favors remineralization. The dissolved oxygen levels (annual mean $4\text{-}5\text{ ml/l}^{-1}$) are stable but it can be extended to 8 when algal blooms happen.

The upwelling event has a direct impact on the composition of species and trophic structure. The vertical distribution and intensity of phytoplankton blooms is controlled by the depth and intensity of the thermocline in the euphotic layer, dispersed by advection currents, strong winds and reduced by high zooplankton grazing pressure (50%, Druehl and Yoneshigue-Braga, 1976). Although biomass levels are similar during post-upwelling phase in summer and non-upwelling in winter, they are still high when compared to oligotrophic tropical waters of the Brazil currents and the abundant, probably resuspended organic detrital matter is an important energy source for higher trophic levels. The biggest biomass is registered in summer where chlorophyll-a values are up to 1 mg/m^3 and can reach 7 mg/m^3 during strong upwelling. At these periods the phytoplankton is dominated by populations of *Asterionella glacialis*, *Skeletonema costatum* and *Nitzschia closterium*. In March-April, the end of upwelling periods, the maximum belongs to genus *Chaetoceros* and *Pleurosigma*. The supremacy of all winter peaks concern to genus *Rhizosolenia* followed by many diatoms species (fig. 2a and 2b).

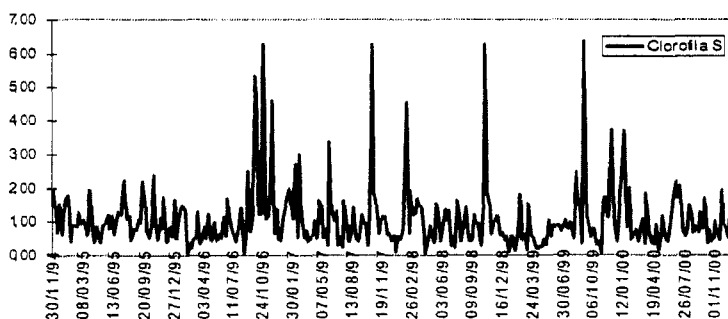


Figure 2a – Surface chlorophyll profile.

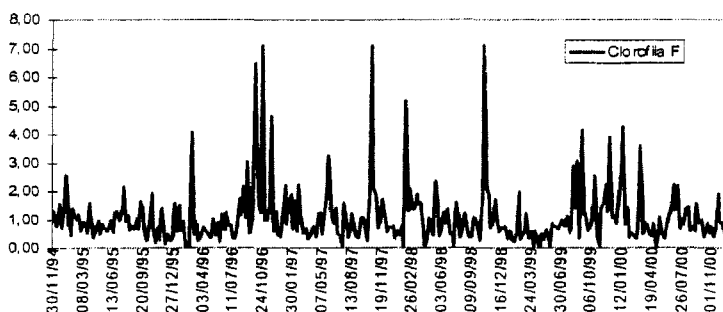


Figure 2b – Bottom chlorophyll profile.

Phytoplankton is of fundamental importance in marine ecosystem as they form the base of marine food chains. They also play an important role in global climate through their draw-down of carbon-dioxide into oceans for photosynthesis and production of gases which contribute to the formation of clouds condensation nuclei and acid rain. They are the major factor that affects the quality of coastal waters and provides an useful tracer for the investigation on dynamic process. How every these relations works? A step towards addressing these issues is to find techniques that accurately estimate or predict primary production in the oceans.

2 Problem Definition

As any other problem to be solved by artificial neural network models, eutrophication phenomenon was reduced to an input-output problem and handled accordingly. Eutrophication and associated algal blooms are caused by a number of physical, chemical and biological factors. For a successful simulation of the eutrophication these factors should be directly included in the system.

3 Data and Target Variable

The original data used in this work was yielded and gently conceded by Instituto de Estudos Almirante Paulo Moreira (IEAPM), it is part of an environment monitoring program of Cabo Frio island. The available data concern to a weekly medium-term time series of physical, chemical and biological variable ranging from the end of 1994 to 2001 as shown in table I and measured at the surface and bottom of the studied area.

Initially this data was joined to a single matrix. A preliminary pre-processing was crucial. The variable salinity show two small intervals of missing data and to solve this problem, a linear regression were performed accordingly Pyle (1999) equations. The deep nitrate variable had two lack values and at this case a medium was interpolated. The usage of raw data leads to slow training and frequent convergence at very bad local minima.

Table I – Available data

Data	Unit	Interval
Chlorophyll	Mg/m ³	Weekly
Temperature	°C	Weekly
Salinity	‰	Weekly
Oxygen	Mg/l	Weekly
Nitrate	µg/l	Weekly
Nitrite	µg/l	Weekly
Amonium	µg/l	Weekly
Phosphate	µg/l	Weekly

In order to get better model performance all variables was normalized with the next equation.

$$X = \frac{\text{actual} - \text{min.}}{\text{max.} - \text{min.}}$$

Where X is the normalized value, “actual” refers to the actual value; “min.” and “max.” refer to the minimum and maximum of the variable values. Chlorophyll-a is the most direct indicator of phytoplankton biomass since all marine planktonic algae contain this photosynthetic pigment. However, chlorophyll concentrations are not an exact assessment of phytoplankton abundance. Several factors influence this pigment abundance, include amount of solar radiation, physiological state, nutrients etc.

4 Genetic algorithm for function optimization

Forecasting complex impact factors and ecological system behavior is a difficult task, so that any environmental manager has to deals with a multivariate time series. Usually, there is a huge non-linearity among them and its interactions can not be so easy to explain. Many authors (Lek, 1996, Scard, 1996, Maier, 1998, etc.) have demonstrated that artificial neural networks have a good ability to handle with this situation. However, the problem in building a neural network model is not so much to define the local learning rule, but to find out how to arrange the neurons in the net and how to choose their synaptic couplings in order to obtain a desired learning behavior. To find the right and best neural architecture to fit the data of a studied case and the internal model parameters is an empirical process and time consumption. Because of these above facts an automation of this model development phase is recommended, for the present work a genetic algorithm (GA) approach was used. These algorithms are a computational abstraction of biological evolution that can be used to solve optimization problems. These models consist on three basic elements. A fitness measure which governs an individual’s ability to influence future generations, a selection and reproduction process which produces offspring for next generation and genetic operators which determine the genetic makeup of the offspring. The individuals,

also called chromosomes, represent possible solutions of the problem. Chromosomes are chains of bits or binary code vectors which takes the genetic information, it means, the number of layers of neural architecture, number of neurons of such layer, its connectivity, weights of synaptic values parameters or the best input configuration for a given output. The power of GA derives largely from the concept of "implicit parallelism", the simultaneous allocation of trials to many regions of the search space. For any selection algorithm, the allocation of trials to individuals induces a corresponding allocation of hyperplanes or substrings represented by individuals. The search is not directionless but makes use of the probabilistic generation of control parameters to direct the search. The main control parameters of a GA are the population size, the selection mechanism, the crossover rate, the mutation rate and the number of generations allowed for the evolution of required structure. In its simplest form a GA is a cycle following the steps accordingly.

- 1- Construct randomly an initial population of chromosomes,
- 2- Calculate and evaluate each chromosome fitness,
- 3- If the result is satisfactory, stop, if not,
- 4- Select the best chromosomes for reproduction based on its fitness (higher),
- 5- Create new offspring by application of crossover and mutation operators,
- 6- Form a population for next generation,
- 7- If process has converged, return the best chromosome as a solution, otherwise go to step two.

Three kind of neural models was left to the genetic algorithm evolve, Multilayer Perceptron (MLP), Time Delay Neural Network (TDNN), Continuous Adaptive Time (CATNN). Curiously every run always results in a Multilayer Perceptron Network.

5 Results

A supervised training strategy was adopted. It consists into three stages. The first one refers to the knowledge acquisition of models (80% of data from 1994 to 2000, randomly chosen) and it is concluded when a fixed number of training samples have been presented or when the error function is sufficiently small or there is no further improvement. At the second step the model is tested to check the acquired knowledge (20% of remainder data). Once both training and test have finished, the model is read to make predictions (2001 data), this is the validation phase. A total of six models were chosen to forecast, three of them for surface and another three to deep. The predictions periods are as follows:

short-term – 1/10/01 to 4/18/01,

medium-term – 1/10/01 to 6/18/01

long-term – 1/10/01 to 11/14/01

Initially all models runs in long term and the figure 3 and 4 show the graphic of the spread of chlorophyll-a at the studied site and the responses of each model respectively.

In such a way, it can be seen that the models in general had a good performance as their generalization capability accompanying the chlorophyll curve. Every surface model (fig.3) simulates an event in 7/25/01 around 1mg/m^3 of chlorophyll but the desired value is much smaller.

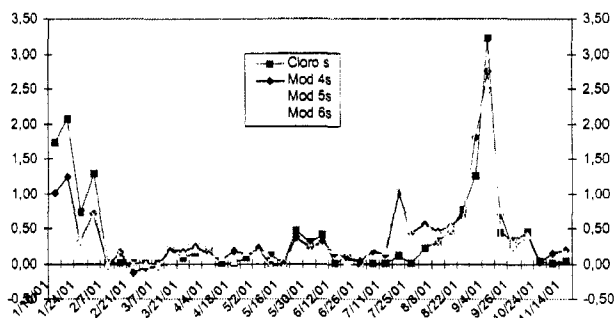


Figure 3 – Surface chlorophyll x models 4S, 5S and 6S.

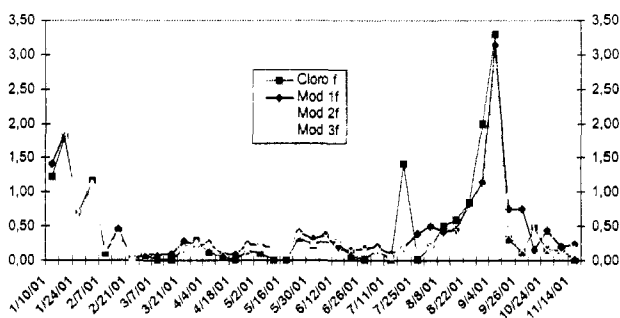


Figure 4 – Bottom chlorophyll x models 1F, 2F and 3F

It can be explained by the fact that the models are maybe sensitive to environment conditions which indicate a possibility of increase primary biomass, but in stead of these a high zooplankton herbivory pressure, advections currents or a limiting nutrient situation should be working. Despite of this, its models had good predictions in the biggest event in 9/4/01 while model 5S has the better approximation in short period forecast. When we look at the deep models (fig.4) it turns evident that they have better performance than the surface ones mainly to short and medium periods. For long term predictions only model 1F and 3F are able to simulate the big event of 9/04/01 while model 2F decay at this point. Curiously and unexpected, all three deep models are unable to predict the small peak in 7/25.01. This event is formed by only one value out of curve, therefore difficult to get anyway it remains unclear. The table II shows accuracy of all models. It is easy to see that both surface and deep models increase their precisions as much as the forecast period decrease. It is still easy to see that model 1F had the better root means square error ($RMSE = 0,08839$) and determination coefficient ($R_2 = 0,81222$) making them the winner, because of it chosen to make

predictions in medium and short-term as shown in figure 5 and figure 6 respectively.

Table II - Error Statistics of Model, Surface -S, and Botton -F.

Models N°	Prediction	Inputs	MSE	RMSE	AAE	R ²
1-F	4 mounths	12	0,00185	0,04303	0,03136	0,95246
1-F	8 mounths	12	0,00230	0,04797	0,03466	0,94470
1-F	11 mounths	12	0,00781	0,08839	0,05609	0,81222
2-F	11 mounths	10	0,00850	0,09218	0,05312	0,79669
3-F	11 mounths	11	0,00784	0,08857	0,04912	0,81668
4-S	11 mounths	11	0,04404	0,20985	0,12743	0,18889
5-S	11 mounths	13	0,01155	0,10745	0,07585	0,71829
6-S	11 mounths	14	0,01824	0,13504	0,09015	0,57999

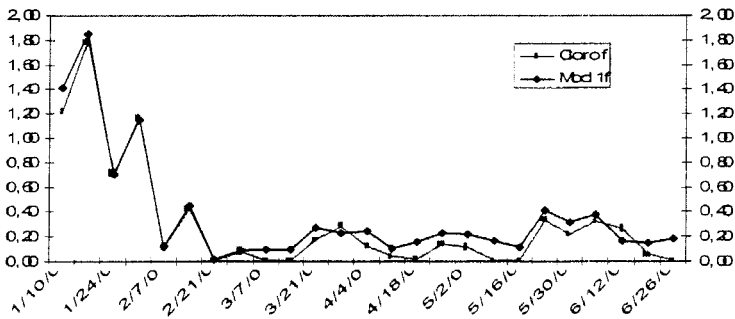


Figure 5 – Desired x Predicted Bottom chlorophyll model 1F, medium-term forecast.

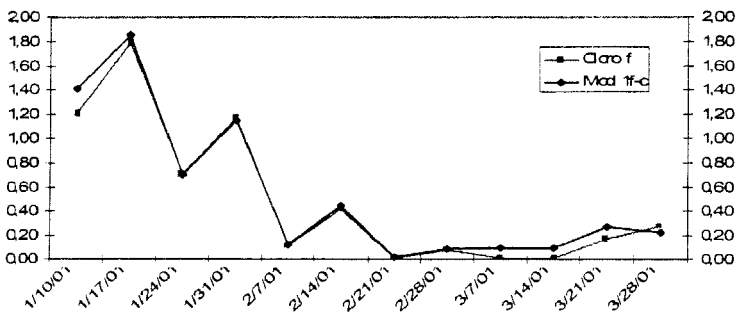


Figure 6 – Desired x Predicted Bottom chlorophyll model 1F, short-term forecast.

6 Conclusion

This study showed that long, medium and short-term variability of phytoplankton biomass could be successfully forecasted by generating mapping functions between environmental variables and algal biomass. To retain the network's ability to generalize, it may be not an optimal strategy to retain it to perfection (Haykin, 1994). The neural net risks overfitting data, and way to avoid this overfitting is limiting the number of hidden units (Smith, 1996). In our experience, the most important basis for successful neural modeling is sound and representative data base beyond topology determination. To avoid high computational effort the genetic algorithm was restrict to evolve at maximum two hidden layers and sixteen units per each layer. To improve the models response and prediction capability we need to consider a time lag between changes in phytoplankton biomass and chemical driving environmental variables. The time lag probably differs among variables, and a number of combinations of input variables with different time lags must be assumed. It is important to note that the models are not so precise in relation on very low chlorophyll-a values. Really, this value can not be considered as "true values" because another natural feedback mechanism (biochemical and physiological) control its concentration inside the cells. The fact of neural network has better performance with high values demonstrate the important role of these values in training the models. The coastal zone and estuarine waters can be considered "plankton-dominated" which implies that this assemblage are good biological indicators and provide valuable information in an assessment of ecosystem condition. As with other assemblages, phytoplankton has disadvantages associated in its biosurvey usage, one of them is high zooplankton grazing rate. This reason argues for investigating these two populations together.

References

- Biocomp Systems Inc., 1999. Neurogenetic Optimizer, Redmond, WA. USA.
- Druhl, L.D., Yoneshigue-Braga, Y., 1976. Growth and succession of tropical phytoplankton cultured in deep water. Publ. Inst. Pesq. Marinha, Rio de Janeiro 93..1-13.
- Gonzalez-Rodriguez, E., Valantin, J.L., Andre, DL., Jacob,S.A.,Upwelling and downwelling at Cabo Frio (Brazil) comparison of biomass and primary production responses, *J Plank. Res*, 14,pp. 289-306
- Hayking, S., *Neural Networks: a comprehensive foundations*, Macmiollan, N.Y.,pp..696, 1994.
- Lek, S., Delacoste, M., Baran, P., Dinopoulus, I., Lauga, J., Aulagnier, S., Aplication of neural networks to nonlinear relationship in ecology, *Ecol. Model*, 90, pp. 39-52, 1996
- LLuch-Cota, E.S., Coastal upwelling in the eastern Gulf of California., *Oceanologica Acta*, 23(6), pp 731-740, 2000

- Maier, H.R., Graeme, C.D., Michael, D.B., Use of artificial neural networks for modelling cyanobacteria *Anabena* spp.in the river Murray, South Australia, *Ecol. Model.*, **105**, pp. 257-272, 1998
- Moreira da Silva, P.C., 1973. A ressurgência em Cabo Frio. Inst. Pesq. Mar. Rio de Janeiro, 78, 56 pp.
- Murtagh, F., Zheng, G., Campbell, J.G., Aussem, A., Neural network modelling for environmental prediction, *Neurocomputing*, **30 (1-4)**, pp. 65-70, 2000.
- Pyle, D., *Data preparation for data mining*, Morgan Kaufmann Pub., Inc., San Francisco, California, 1999.
- Refenes,A.N., Azema-Barac, M., Chen, L., Karoussos, S.A., Currency Exchange Rate Prediction and Neural Networks Design Strategies, *Journal of Neuralcomputing and Applications*, **4 (2)**, pp. 417-427, 1992
- Rumelhart, D.E., Hilton, G.E., Willians, R..J., Learning representations by back-propagation error. *Nature*, **323**, pp. 533-536, 1986
- Scard, M., Artificial neural network as empirical models for estimating phytoplankton production, *Mar. Ecol. Prog. Ser.*, 139.. 289-299, 1996.
- Smith, M., *Neural network for statistical modelling*. International Thompson Computer Press, Boston. p.235, 1996.
- Valantin, J.L., Coutinho, R.,Modelling maximum chlorophill in the Cabo Frio (Brazil) upwelling..a preliminary approach, *Ecol. Modelling*, **52**,pp.103-113,1990.