

Developing an empirical model of phytoplankton primary production: a neural network case study

Michele Scardi ^{a,*}, Lawrence W. Harding Jr. ^b

^a *Stazione Zoologica 'A. Dohrn' di Napoli, Villa Comunale, 80121 Napoli, Italy*

^b *University of Maryland, Horn Point Laboratory (UMCES) and Maryland Sea Grant, Box 775, Cambridge, MA 21613, USA*

Abstract

We describe the development of a neural network model for estimating primary production of phytoplankton. Data from an enriched estuary in the eastern United States, Chesapeake Bay, were used to train, validate and test the model. Two error backpropagation multilayer perceptrons were trained: a simpler one (3-5-1) and a more complex one (12-5-1). Both neural networks outperformed conventional empirical models, even though only the latter, which exploits a larger suite of predictive variables, provided truly accurate outputs. The application of this neural network model is thoroughly discussed and the results of a sensitivity analysis are also presented. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Artificial neural networks; Empirical models; Phytoplankton; Primary production; Chesapeake Bay

1. Introduction

Estimates of phytoplankton primary production based on empirical models are increasingly used as an alternative to direct data acquisition that can be both expensive and time consuming. This is particularly true in the era of satellite oceanography because remote measurements of ocean color that provide global coverage of phytoplankton biomass can serve as inputs to models that estimate production. Although empirical models of primary production are usually based on simple linear relationships (e.g. Cole and Cloern, 1987), the estimates they provide are reason-

ably accurate because primary production is largely regulated by variables that are simple to measure, i.e. downwelling irradiance and phytoplankton biomass.

Despite the usefulness of linear relationships for estimating production, other factors that affect photosynthetic carbon assimilation are related to production in a non-linear manner, such as photosynthetic efficiency of the phytoplankton cells. Therefore, more flexible empirical models that are both simple and capable of reproducing these relationships can theoretically play an important role in improving our ability to estimate production.

Conventional models that attempted to address this problem by means of multiple linear regression (e.g. Eppley et al., 1985), or the use of

* Corresponding author. Fax: +39-81-7641355.

E-mail address: mscardi@mclink.it (M. Scardi)

semi-analytic formulations (e.g. Balch et al., 1989), did not perform significantly better than much simpler empirical models. An alternative approach, involving the use of neural networks has recently generated significant improvement in estimating production (Scardi, 1996) or other complex non-linear ecological processes (Lek et al., 1996) where sufficient training data were available. Moreover, neural networks are also able to exploit the heterogeneous information that is provided by other variables that may be correlated to primary production on a regional scale only, and to use this information to achieve refinement of primary production estimates.

The first neural network that was trained as an empirical model of phytoplankton primary production (Scardi, 1996) was essentially a toy model, because of the limited number of training patterns. It was developed with a small data set that was reported in a comprehensive study of phytoplankton photosynthesis in Chesapeake and Delaware Bays (Harding et al., 1986). These data were used in initial efforts because the data on pertinent variables were assembled and readily usable, and comparisons with linear models could be made rather easily. The main purpose of that work was to show that a simple error back-propagation neural network had the potential to outperform conventional empirical models of phytoplankton primary production. Since that initial report, further research has been carried out on primary production and ancillary data for Chesapeake Bay spanning over a decade (Harding et al., in prep.) and on the application of neural networks both to phytoplankton production modelling (Scardi, in prep.) and to other related topics (Recknagel et al., 1996; Recknagel, 1997).

In this paper, we present new results of a case study that focused on developing a reliable modelling tool for Chesapeake Bay. Contemporary studies of trophic dynamics and remotely sensed observations providing synoptic biomass fields in the Bay are components of ongoing research that entail a need for accurate estimates of phytoplankton primary production. Beyond the specific use of neural network analysis to estimate primary production in Chesapeake Bay, however, this approach has general ecological relevance. It

is successful with data from this very complex estuarine ecosystem in which the principal variables regulating primary production are characterized by variability on a wide range of time and space scales, the likelihood of a broader application to other marine systems is enhanced.

2. Materials and methods

The 1982–1983 data that were used in the initial attempt to develop a neural network model of primary productivity were collected on a series of five cruises in Chesapeake and Delaware Bays. Further development of this model, however, was focused on Chesapeake Bay only. Chesapeake Bay, in Maryland and Virginia, is the largest bay on the Atlantic coast of the US (Fig. 1). It is about 320 km long from north to south and from 5 to 40 km wide. The Susquehanna and the Potomac are the largest of its many tributary rivers and creeks. The bay is a shipping artery, and the bay cities of Norfolk, VA., and Baltimore, MD., are among the nation's leading ports. Waterfowl, fish, oysters, and crabs, long abundant, have been threatened by pollution in recent years. Chesapeake Bay is characterized by strong gradients in salinity, turbidity, dissolved nutrients and chlorophyll as a measure of phytoplankton biomass.

Integral, daily primary production was measured using ^{14}C assimilation in simulated in situ sunlight incubations. Neutral density screens were used to attenuate sunlight and generate a light series, and surface seawater was circulated for cooling. Downwelling irradiance was measured continuously with a LiCor quantum sensor positioned in an unobstructed location on the ship, and vertical profiles were made throughout the day using an underwater LiCor quantum sensor to ascertain the diffuse attenuation coefficient for photosynthetically available radiation (PAR). Further details of the methods are contained in Harding et al. (1986).

Measurements of chlorophyll concentrations were made using standard fluorometric methods (Strickland and Parsons, 1968), nutrient concen-

trations were determined by wet chemistry on a Technicon AutoAnalyzer II, and ancillary data on other properties were collected at the same times and locations as samples were collected for measuring primary productivity. The reference to the original data source (Harding et al., 1986), contains most of the detailed methods and other aspects of the data collection are presented by Fisher et al. (1988).

Measurements of primary production made from 1987–1996 used the same methods as were used in the 1982–1983 cruises. Stations were predominantly located within Chesapeake Bay along the mainstem axis from the limit of salt to the mouth and plume regions nearly 300 km seaward. Approximately ten stations were occupied on each cruise, with the exception of 1995–1996 when more than double this number of stations was

occupied per cruise and included sampling lateral to the mainstem axis. Each measurement of primary production was accompanied by collection of a full set of ancillary data.

The most recently collected data used in this analysis were from 1995–1997 and were collected on a series of cruises addressing Trophic Interactions in Estuarine Systems (TIES) sponsored by the US National Science Foundation. Of these data, the 1997 measurements were used to test the NN model and not to develop it.

All the neural networks we used as empirical models were multilayer perceptrons with one hidden layer and only one neuron in the output layer (i.e. phytoplankton primary production). This is by far the most common and flexible kind of neural network and it provides good performances in a wide range of applications.

Our applications aimed at training the most generalized neural network, rather than the one that optimally fitted the training test. Therefore, the error backpropagation training algorithm was used in its simplest version, as learning rate, set to a unit value, was not allowed to vary during training and no momentum term was used.

The training procedure was based on a subset of the 1982–1996 data set, which consisted of 326 patterns. In fact, in the case of our final model, only 100 patterns were randomly selected and used as training set, whereas the remaining 226 patterns were used as validation set. Even though a large validation set usually prevents overtraining, other techniques were also applied in order to obtain the most generalized model.

In particular, a small amount of Gaussian noise ($\mu = 0$, $\sigma = 0.01$) was added to the input patterns (Györgyi, 1990) and only a subset ($n = 50$) of the training set was randomly selected for each training epoch. The random selection of the training subset was also needed because a learning per pattern strategy was chosen and therefore it was necessary not to always submit the training patterns in the same order. Moreover, an early stopping strategy was used in the training procedure (i.e. training was stopped as soon as the validation set error started to increase).

The best structure of the neural network models was determined on the basis of empirical tests,

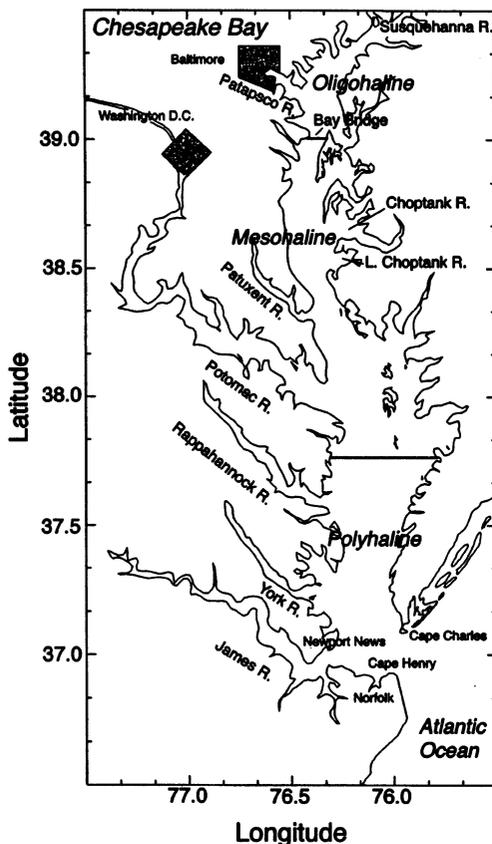


Fig. 1. Chesapeake Bay is the largest bay on the Atlantic coast of the US. It is about 320 km long and from 5 to 40 km wide.

Table 1
Neural network input and output variables^a

Variable	units	min	max
<i>Input</i>			
$\frac{1}{2} \left[\cos\left(\frac{2\pi \cdot \text{day}}{365}\right) + 1 \right]$	None	0.0	1.0
$\frac{1}{2} \left[\sin\left(\frac{2\pi \cdot \text{day}}{365}\right) + 1 \right]$	None	0.0	1.0
Latitude	Degrees	36.8	39.5
Longitude	Degrees	75.6	76.6
Station depth	m	0.0	45.0
Water temperature	°C	0.0	32.0
Salinity	PSU	0.0	32.0
Surface chlorophyll concentration (\log_{10})	mg m ⁻³	-0.8	1.9
Total chlorophyll in the photic zone (\log_{10})	mg m ⁻²	-0.3	2.7
Surface downwelling irradiance	E m ⁻² day ⁻¹	0.0	80.0
Light extinction coefficient	m ⁻¹	0.0	6.0
Photic zone depth	m	0.0	25.0
<i>Output</i>			
Phytoplankton primary production (\log_{10})	mg C m ⁻² day ⁻¹	0.9	3.9

^a Units and the minimum and maximum values that were used to scale raw data to [0, 1] intervals are also shown. Variable names followed by (\log_{10}) indicate that raw values have been log-transformed before scaling them to a [0, 1] interval.

where hidden layers with three to 15 neurons were used. The best performance was obtained with five neurons in the hidden layer both in the case of the simpler model (three inputs) and in the case of the more complex one (12 inputs). However, the differences among neural networks with different structures were not dramatic and only the performance of the 3- x -1 model was perceptibly degraded when more than ten hidden neurons were used.

The simpler 3-5-1 model used surface chlorophyll concentration, surface downwelling irradiance and depth of the photic zone as input variables, whereas nine more variables were selected as additional inputs for the more complex neural network. Input and output variables of this neural network are listed in Table 1, where the units and values that were assumed as limits to scale variables into [0, 1] intervals are also given. Inputs for both phytoplankton biomass and primary production were \log_{10} -transformed before scaling them to a [0, 1] interval. The log transformation was performed on the basis of both a theoretical assumption and an empirical test. The theoretical assumption was that the mean square

error of the neural network output will be biased when raw data are used. This pertains because training patterns containing high values for biomass and primary production, containing proportionately greater sampling and measurement errors, may unduly dominate the output. The empirical test was carried out by comparing the performance of neural networks trained with transformed data to performance with raw data. In the case of the final 12-5-1 neural network, training on log-transformed data outperformed training on raw data, as it allowed the neural network to explain almost 20% more variance (the determination coefficients were $R^2 = 0.546$ and $R^2 = 0.353$, respectively).

The serial number of the day of the year was transformed using sine and cosine functions (see Table 1) that map the date onto a circle. Two inputs—the total chlorophyll in the photic zone and the photic zone depth—were computed on the basis of other input variables. The total chlorophyll in the photic zone (Table 1) was obtained as the product of surface chlorophyll concentration and photic zone depth, assuming that the phytoplankton biomass is homogeneously

distributed in the upper water column. The photic zone depth (Table 1), i.e. the depth where the available downwelling irradiance is the 1% of the surface downwelling irradiance, was obtained as 4.605 (i.e. $\ln 0.01$) divided by the light extinction coefficient (Table 1). If the resulting value was larger than the station depth (Table 1), then the latter was assumed as photic zone depth.

3. Results

The toy model presented by Scardi (1996) performed well using the 1982–1983 data set on which it was trained ($R^2 = 0.940$). When used with a much larger data set spanning 1982–1996 and encompassing a wide range of environmental conditions, this model did not perform nearly as well ($R^2 = 0.156$), as shown in Fig. 2. This relatively simple approach was based on a 3-5-1 neural network that used surface downwelling irradiance, surface chlorophyll concentration, and photic zone depth as inputs. We found that the toy model was unable to reproduce primary production values that were larger than the ones on which it was trained, and that large errors were also obtained even within the range of observations contained in its own training set ($0\text{--}3 \text{ g C m}^{-2} \text{ day}^{-1}$).

To ascertain the performance of the toy model compared to other, more conventional approaches, we also used a common model based on linear regression of primary production on a composite variable obtained from the product of the same three variables used as neural network inputs (see Cole and Cloern, 1987). Despite what we term poor performance of the toy model with the larger data set from Chesapeake Bay, the estimates of primary production were significantly better than those obtained using conventional linear models that returned a mean square error almost twice as high.

To overcome the shortcomings of the toy model, a new 3-5-1 neural network was trained on the basis of the entire 1982–1996 data set. Two training procedures were carried out, one on raw data and the other on log-transformed biomass and primary production data, but none produced a synaptic weight set that showed a significant improvement over the toy model.

As in the case of the toy model, these networks were not able to cope with high primary production values, even though they were trained on a quite large data set. This result was not unexpected because primary production in Chesapeake Bay is clearly not regulated by phytoplankton biomass, irradiance and photic zone depth alone; there is a strong landward to seaward gradient in

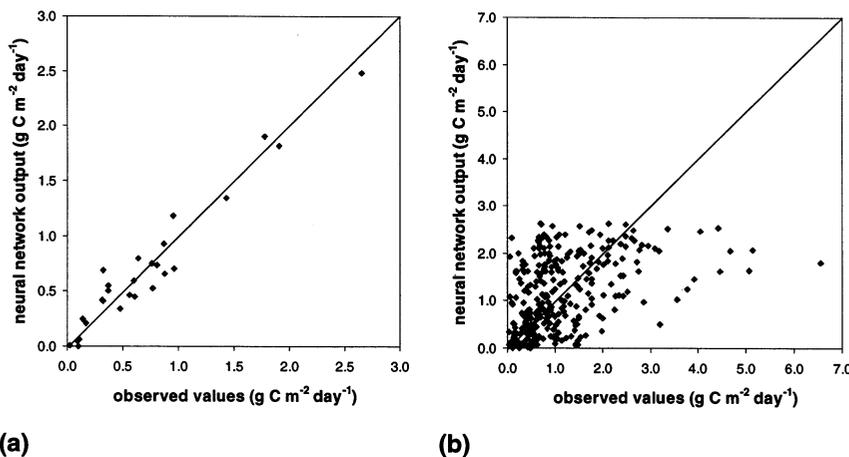


Fig. 2. Scatter plots of neural network outputs versus observed values for the toy model described in Scardi (1996). The 1982–1983 training subset was accurately fitted (a), whereas the whole 1982–1996 data set showed poor generalization (b).

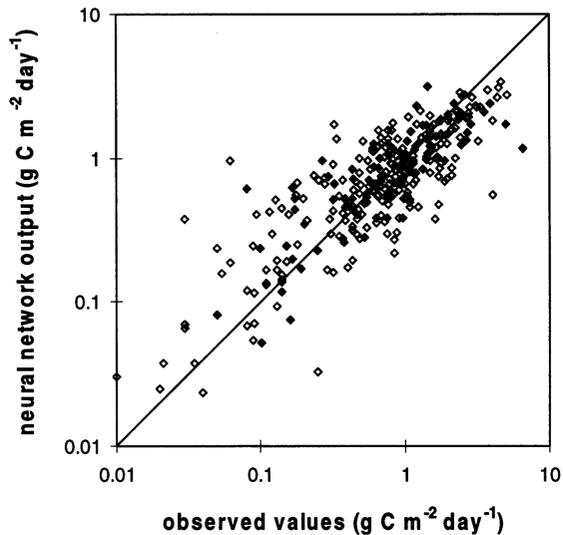


Fig. 3. Scatter plot of the neural network outputs versus observed values for the 12-5-1 model. Both training and validation data are shown. The overall agreement between observed and simulated data was satisfactory ($R^2 = 0.546$). The validation set values (white diamonds, $R^2 = 0.614$) were reproduced even better than the training set ones (black diamonds, $R^2 = 0.420$).

dissolved nutrients and much of the Bay is nutrient limited for at least part of the year. Accordingly, we surmised that additional information was needed to improve the model.

The neural network model using a 12-5-1 structure, i.e. a larger suite of predictive variables, shows improved estimates over previous approaches (Fig. 3). This finding pertains both to the training set shown as black diamonds, and to the validation set shown as white diamonds. Primary production values are predicted with greater accuracy ($R^2 = 0.546$) than in the case of linear or other simpler models. The neural network also showed good generalization properties, in that the validation set was fitted even better than the training set ($R^2 = 0.614$ and $R^2 = 0.420$, respectively).

Although data for phytoplankton biomass and primary production were log-transformed before training, the model outputs need to be transformed back to raw data in most applications. Therefore, the error distribution of the model, which was unbiased in log units, was also checked after back-transforming data to raw units and a small

bias was detected ($m_{\text{error}} = -0.14248$, in the validation set). Obviously, this systematic error depended on the different impact that very large and very small values exerted with or without log transformation.

In order to obtain unbiased primary production estimates, a simple linear correction was defined by least square optimization and applied to the neural network output. The corrected estimates were then computed by multiplying neural network outputs by 1.15575 (this correction could be visualized as a small vertical shifting of all the points in the log–log plot in Fig. 3).

The resulting error distribution was virtually unbiased ($m_{\text{error}} = -0.00174$, in the validation set) and almost symmetrical, as shown in Fig. 4. It should be noted that more than 80% of the errors of the primary production estimates in the validation set were within the $\pm 0.6 \text{ g C m}^{-2} \text{ day}^{-1}$, i.e. less than 1/10 of the observed data range. Moreover, the accuracy of the model was also slightly improved, as corrected outputs explained 3% more variance ($R^2 = 0.578$) than the uncorrected ones ($R^2 = 0.546$).

The accuracy of the neural network model and its generalization capabilities were also tested on an independent data set ($n = 52$), which was collected during 1997 and therefore was not available during the training phase. In the scatter plot in Fig. 5 predicted versus observed values are shown for both this new testing set (large black circles) and the original training and validation sets (small

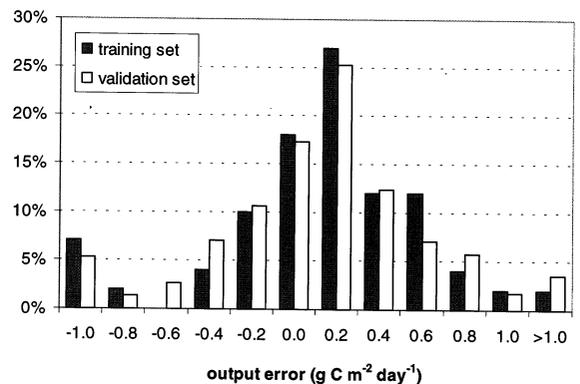


Fig. 4. Error distribution of the corrected neural network outputs. The labels on the error axis indicate the upper limit of each class.

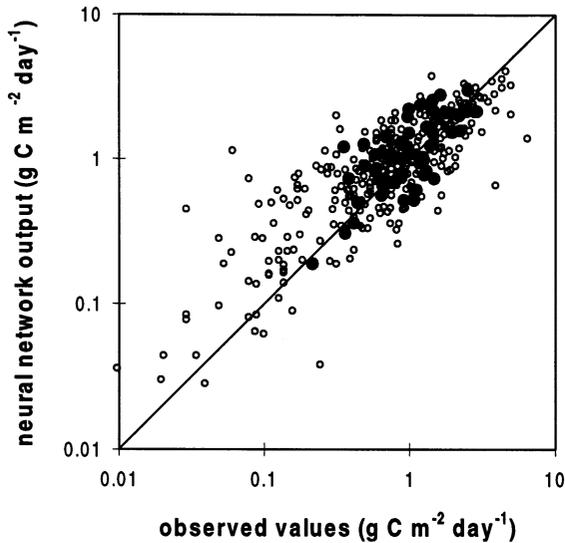


Fig. 5. Scatter plot of the neural network outputs versus observed values. Both the new independent testing set (1997, large black circles) and the original training and validation sets (1982–1996, small white circles) are shown.

white circles). The new primary production values were reproduced by the model with the same accuracy as original data and were almost unbiased, as their mean error was negligible ($m_{\text{error}} = 0.082$).

The error distribution of the new testing set is shown in Fig. 6, where it is compared to the error

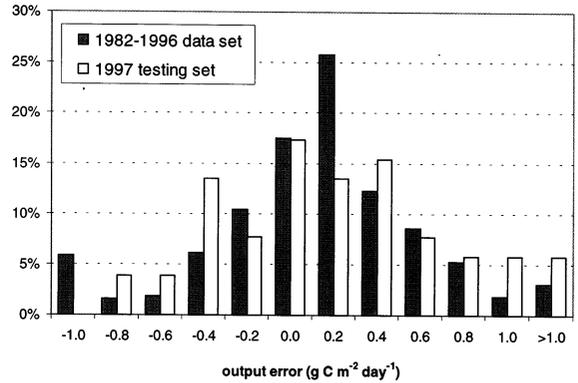


Fig. 6. Error distribution of the corrected neural network outputs. The labels on the error axis indicate the upper limit of each class.

distribution of the original data set. Even though the latter is more regular and symmetrical, the differences between the two distributions are minor and are probably influenced by the smaller number of patterns in the new testing set.

Finally, a sensitivity analysis was carried out using the whole 1982–1996 data set to assess the effect of small changes in each input on the neural network output. The results of this analysis provide a useful insight into the neural network model, but they also help to understand the underlying ecological processes, i.e. the relative importance of the predictive variables to

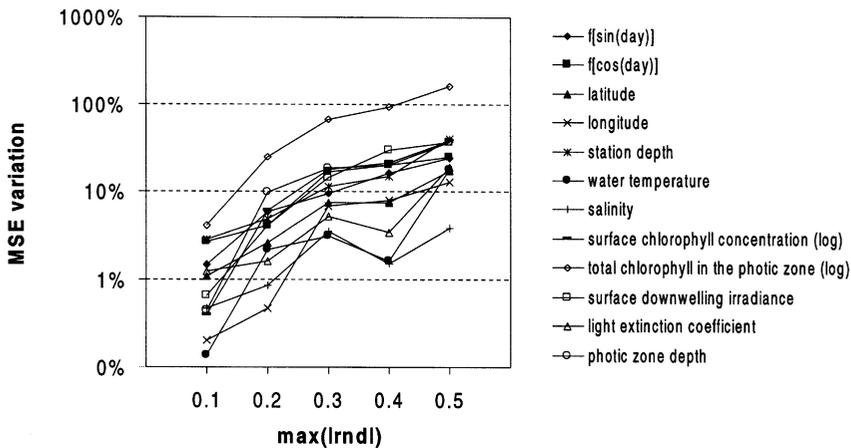


Fig. 7. Percentage variation of the mean square error of the neural network output at increasing levels of input perturbation. White noise ranging from $[-0.1, 0.1]$ to $[-0.5, 0.5]$ was added to each input variable in the whole 1982–1996 data set and the resulting increase in mean square error was expressed as a percentage of the original mean square error.

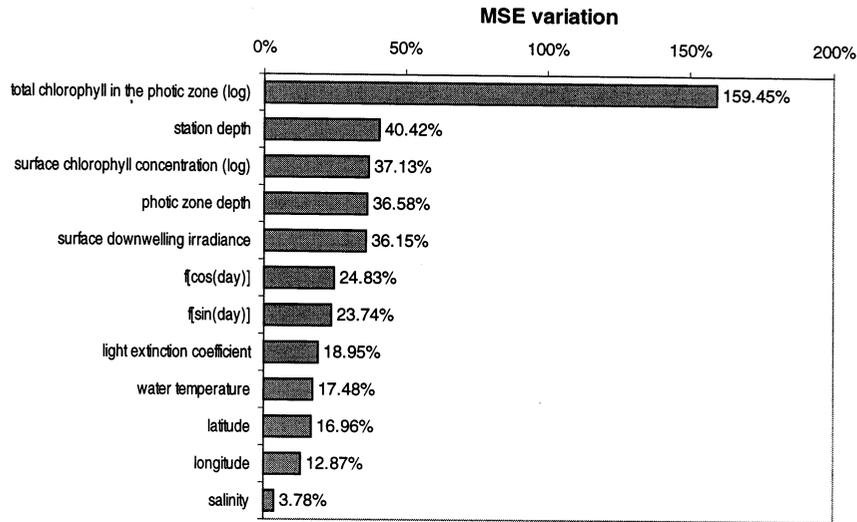


Fig. 8. Percentage variation of the mean square error of the neural network output after the addition of $[-0.5, 0.5]$ white noise. The input variables are ranked according to their sensitivity.

phytoplankton primary production in Chesapeake Bay.

In the sensitivity analysis, the mean square error of the neural network output is expected to increase as a larger amount of white noise is added to the selected input variable. The mean square error variations that were observed after white noise additions varying from $[-0.1, 0.1]$ to $[-0.5, 0.5]$, i.e. from 20 to 100% of the input range, are shown in Fig. 7.

The minimum level of input perturbation was similar in magnitude to the measurement error of the oceanographic data and so were the changes it induced in the mean square error of the neural network output ($< 5\%$). Increasing white noise additions caused increasing mean square errors in the output, even though this relationship was not absolutely monotonic, because less sensitive variables, that did not affect the neural network output very much, showed a few negative increments. However, the relative sensitivity of the input variables did not vary significantly when very large amounts of white noise were added. These results suggest that the primary production model that was embedded in the neural network was probably consistent with the ecological processes as it was not misled by unlikely input patterns.

The most influential variable among the neural network inputs in affecting output was by far the total chlorophyll in the photic zone. It was the only input variable that caused an increase in the mean square error larger than 100% when $[-0.5, 0.5]$ white noise was added, as it is clearly shown in Fig. 8. As expected for a primary production model, the predictive variables that were related to light availability and phytoplankton biomass had the largest effects on output among the remaining variables. The least influential variable was salinity, despite that it may be viewed as a proxy for freshwater inflow and often covaries with nutrient concentrations.

4. Discussion

The neural network provided accurate and unbiased estimates of phytoplankton primary production for a system that is characterized by high spatial and temporal variability. This is a satisfying result, given the shortcomings of linear models that fail to perform acceptably with the same data and that contain biases that are particularly pronounced at low and high primary productivity rates. In most cases, the error of primary production estimates obtained with the neural network

was within the range of the measurement error. We believe the neural network approach outperformed conventional empirical models because it is inherently much more flexible in dealing with the influences of a number of variables that regulate phytoplankton primary productivity in estuaries.

Our results were obtained using a very conservative approach as far as generalization is concerned, because most of the available data were used for neural network validation and only a restricted subset, i.e. less than one third of the entire data set, was used for neural network training. We also tested our neural network model using an independent data set that was not available during development of the model. The success of this approach implies that the present form of our model is probably less than optimal and that further improvements are still possible. A further consideration is that the training procedure was not optimized (a constant learning rate

and no momentum were used), and improvements in this area may refine the model further.

The neural network model of Chesapeake Bay phytoplankton primary production can play an important role in monitoring and research activities, because it may permit reduction of the number of direct primary production measurements that are needed to reconstruct large scale spatial patterns or high frequency time series. An example of such an application is shown in Fig. 9, in which the distribution of phytoplankton primary production in the mainstem area of Chesapeake Bay is presented as a grayscale image. These estimates were based on discrete data collected during a summer cruise (23–28 July 1995) that were interpolated to generate complete input grids. In future applications some of these input grids could be replaced with remotely sensed data. An aircraft remote sensing program (cf. Harding et al., 1992, 1994, 1995) provides high resolution estimates of chlorophyll for the estuary, using sensors designed to replicate band of the satellite ocean color instrument, SeaWiFS, that is now providing global coverage. Data from this source can be substituted for shipboard observations and fill time and space gaps that accompany more routine sampling. Given that the most influential variables on model output are related to the accuracy of biomass inputs, we envision improved performance of the neural network model when remotely sensed data are used.

We believe the principal explanation for the superior performance of the neural network model to that of more conventional approaches is that the complexity of factors that regulate phytoplankton primary production are better captured. From a purely theoretical viewpoint, other empirical models might also obtain this result, provided that their formulation is carefully defined and sufficiently complex to incorporate the data structure. However, in these cases the model formulation has to be explicitly defined by the modeler, who usually opts for an empirical approach when his/her understanding of the processes to be modeled is not complete or he/she thinks that accuracy can be traded for simplicity.

An example of the degree of complexity of the relationships that can be reproduced by a neural

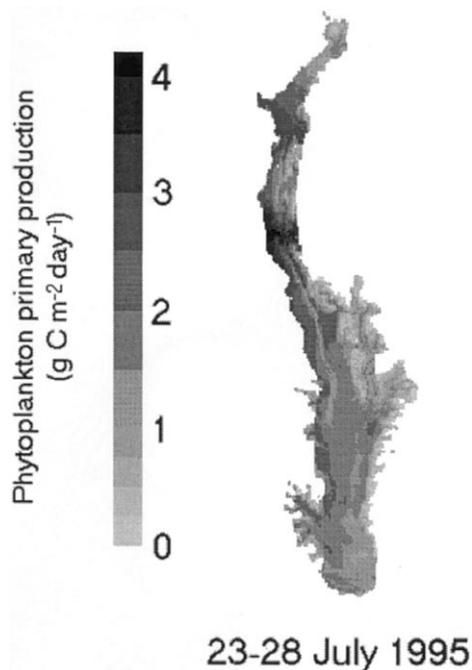


Fig. 9. An example of application of the neural network. The phytoplankton primary production was estimated over the whole Chesapeake Bay mainstem area using interpolated input data. Each pixel in the image corresponds to a 1 km² square.

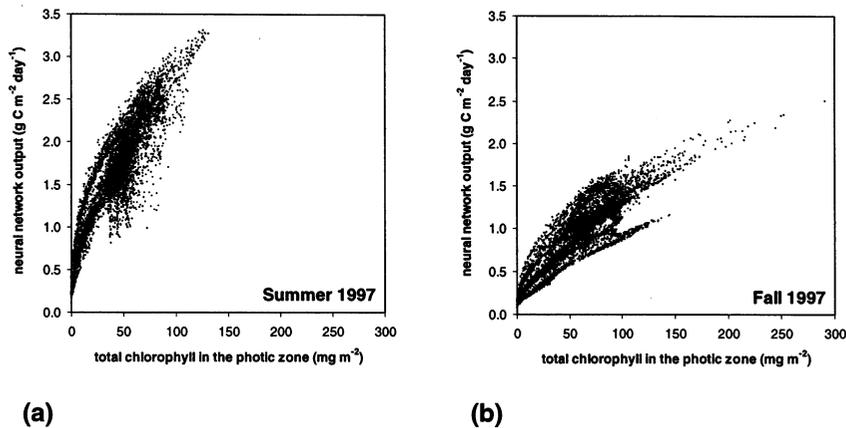


Fig. 10. Neural network output versus total chlorophyll in the photic zone in different seasonal scenarios: summer (a) and fall (b). More than 7000 points, corresponding to pixels in images similar to Fig. 9, are shown in each plot. The position of each point depends on the photosynthetic efficiency of the whole water column and a steeper overall slope implies a higher photosynthetic efficiency, as in the summer 1997 plot (a). It is interesting to notice that the neural network model was able to reproduce a range of different area-specific non-linear relationships.

network model is shown in Fig. 10. The neural network outputs (i.e. estimates of phytoplankton primary production) that were obtained for more than 7000 input patterns were plotted against one of the inputs, the total chlorophyll in the photic zone. Therefore, the position of each point is determined by the photosynthetic efficiency of the whole water column. It is very clear that the two seasonal scenarios that were considered were completely different because the overall photosynthetic efficiency varies in time. However, it is also clear that the biomass/production relationship is also variable within each sub-plot, because all the points are arranged as to form a set of curves, each one having a different slope, that reproduce the spatial variation of the biomass/production relationship in Chesapeake Bay. Even though the sensitivity analysis showed that the total chlorophyll concentration was probably the most relevant input variable, plots obtained with other variables also showed similar patterns.

Finally, it has to be stressed that sensitivity analysis might play an important role in both the optimization of the neural network models and in understanding the processes to be modeled. Sensitivity analysis is not a simple and straightforward task when analytical models are taken into account, but it is even more challenging when neural net-

works are considered. However, the procedure we used was able to analyze the first-order effects of input perturbation on the neural network output and the results provided a useful insight both into the neural network mechanics and primary production processes. The neural network outputs were almost invariant when small perturbations, similar to those that depend on sampling errors, were introduced. On the other hand, when more noise was added to the inputs, the role of each variable could be defined in terms of relative importance in determining phytoplankton primary production.

The total chlorophyll in the photic zone, i.e. the total biomass that is photosynthetically active, was clearly the most important predictive variable. Other variables, such as salinity, were less sensitive to the addition of white noise and therefore seem to play a less important role. Of course, excluding these variables might help prune the neural network structure. This kind of optimization is not very important from a computational point of view, but could reduce the cost of data acquisition without a significant loss in accuracy of the model.

The neural network model of Chesapeake Bay phytoplankton primary production has been implemented in Java and can be tested at the following URL: <http://www.mare-net.com/mscardi/work/nn/cbjavann.htm>.

References

- Balch, W.M., Eppley, R.W., Abbott, M.R., 1989. Remote sensing of primary production-II. A semi-analytical algorithm based on pigments, temperature and light. *Deep-Sea Res.* 36 (8), 1201–1217.
- Cole, B.E., Cloern, J.E., 1987. An empirical model for estimating phytoplankton productivity in estuaries. *Mar. Ecol. Prog. Ser.* 36, 299–305.
- Eppley, R.W., Stewart, E., Abbott, M.R., Heyman, U., 1985. Estimating ocean primary production from satellite chlorophyll. Introduction to regional differences and statistics for the Southern California bight. *J. Plank. Res.* 7, 57–70.
- Fisher, T.R., Harding, L.W., Stanley, D.W., Ward, L.G., 1988. Phytoplankton, nutrients and turbidity in the Chesapeake, Delaware and Hudson estuaries. *Estuar. Coast. Shelf Sci.* 27, 61–93.
- Gyögyi, G., 1990. Inference of a rule by a neural network with thermal noise. *Phys. Rev. Lett.* 64, 2957–2960.
- Harding, L.W. Jr., Meeson, B.W., Fisher, T.R. Jr., 1986. Phytoplankton production in two east coast estuaries: photosynthesis-light functions and patterns of carbon assimilation in Chesapeake and Delaware Bays. *Estuar. Coast. Shelf Sci.* 23, 773–806.
- Harding, L.W. Jr., Itsweire, E.C., Esaias, W.E., 1992. Determination of phytoplankton chlorophyll concentrations in the Chesapeake Bay with aircraft remote sensing. *Rem. Sens. Environ.* 40, 79–100.
- Harding, L.W. Jr., Itsweire, E.C., Esaias, W.E., 1994. Estimates of phytoplankton biomass in the Chesapeake Bay from aircraft remote sensing of chlorophyll concentrations, 1989–1992. *Rem. Sens. Environ.* 49, 41–56.
- Harding, L.W. Jr., Itsweire, E.C., Esaias, W.E., 1995. Algorithm development for recovering chlorophyll concentrations in the Chesapeake Bay using aircraft remote sensing, 1989–1991. *Photogr. Eng. Remote Sens.* 61, 177–185.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Model.* 90 (1), 39–52.
- Recknagel, F., 1997. ANNA—artificial neural network model predicting blooms and succession of blue-green algae. *Hydrobiology* 349, 47–57.
- Recknagel, F., French, M., Harkonen, P., Yabunaka, K.I., 1996. Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Model.* 96 (1–3), 11–28.
- Scardi, M., 1996. Artificial neural networks as empirical models of phytoplankton production. *Mar. Ecol. Prog. Ser.* 139, 289–299.
- Strickland, J.D.H., Parsons, T.R., 1968. A practical handbook of seawater analysis. *Bull. Fish. Res. Bd. Can.* 167, 1–311.