Artificial neural networks as empirical models for estimating phytoplankton production

Michele Scardi*

Laboratorio di Oceanografia Biologica, Stazione Zoologica 'A. Dohrn' di Napoli, Villa Comunale, I-80121 Napoli, Italy

ABSTRACT: Many empirical models have been developed in order to obtain phytoplankton production estimates from other variables that are easier to measure. These empirical models are usually based on regression of phytoplankton production against biomass and other variables. They are particularly useful to fully exploit data sets acquired by both *in situ* instrumental measurements and remote sensing. Two conventional empirical models were compared with a new approach, based on artificial neural networks. Although very simple neural networks were used, they provided a much better fit to observed data than conventional models do and they seem a very promising tool for phytoplankton production modeling.

KEY WORDS: Empirical models · Neural networks Phytoplankton · Production Estuaries

INTRODUCTION

Empirical models of phytoplankton production are an important tool in biological oceanography. They provide reasonably accurate production estimates on the basis of widely available variables, whereas direct measurements are expensive and difficult to carry out on a routine basis.

Even though many empirical models share common features, such as mathematical formulation or independent variables, each of them has been optimized with respect to a particular data set or environment. In other words, empirical models usually trade generality for simplicity.

Most empirical models are linear, since they assume that primary production is a linear function either of a few independent variables or of a single composite variable. Phytoplankton biomass is of course always taken into account, whereas other variables are not (irradiance, photic depth or water transparency, temperature, photoperiod, etc.).

The first empirical model was probably the one that Ryther & Yentsch (1957) applied to their own data as well as to data from Riley (1939, 1956) and Conover

(1956). They described photosynthetic rate at light saturation, p(sat.), as a function of chlorophyll concentration [3.7 g C fixed h^{-1} (g chl)$^{-1}$] and found a good agreement between predicted and observed production. They also took into account relative photosynthesis (R, computed as a function of surface irradiance) and light extinction (k). In the simplest case, i.e. when phytoplankton was homogeneously distributed with depth, the model had the following formulation:

$$PP = \frac{R}{k} \; p(\text{sat.})$$

where PP is phytoplankton primary production.

More recent empirical models that focus on ocean data range from very simple formulations, based on chlorophyll concentrations only (e.g. Smith et al. 1982), to more complex, multiple regression approaches (e.g. Eppley et al. 1985). Most of these models are specifically aimed at the exploitation of phytoplankton biomass data estimated from satellite images (e.g. Lorenzen 1970).

In estuarine and coastal systems the relationships between phytoplankton production and environmental factors are probably simpler. Primary production, for instance, is usually independent of inorganic nutrient concentrations, which are seldom growth-limiting. Moreover, the water column is often well mixed and

*E-mail: mscardi@mclink.it

the phytoplankton biomass is homogeneously distributed with depth.

In these situations mean biomass and light availability are the most relevant factors that control phytoplankton production and they can be easily used to build up empirical models.

An example of this kind of empirical model, both simple and effective, is the one proposed by Cole & Cloern (1984), which is based on the assumption that phytoplankton primary production is directly proportional to phytoplankton biomass, surface irradiance and photic zone depth.

$$PP = a + b \cdot BZ_p I_0$$

where PP is the phytoplankton primary production, B is the phytoplankton biomass, Z_p is the photic zone depth, I_0 is the surface irradiance and a and b are the linear regression intercept and slope. This model was applied to different estuaries (Cole & Cloern 1987) and to mesocosms (Keller 1988). In other cases the model formulation was formally different, as the inverse of the light extinction coefficient, k^{-1}, was used instead of Z_p (Pennock & Sharp 1986).

Some attempts have been made to improve the performances of empirical models by adding more detail to their formulation (e.g. Balch et al. 1989). However, these semi-analytical models did not show a significant improvement with respect to much simpler empirical models.

Empirical models of phytoplankton production are not only a postprocessing tool. In fact, they are also essential for instrumental estimates of primary production (e.g. by pump and probe fluorometers), as these usually rely on algorithms that are based on empirical models.

Therefore, effective empirical models could significantly improve the quality of data sets acquired by both *in situ* instrumental measurements and remote sensing.

In this paper the application of artificial neural networks (a brief introduction to error back-propagation neural networks and their terminology is presented in Appendix 1) to empirical modeling is compared with a conventional approach. Neural networks are a very promising tool for empirical modeling of complex systems: networks with at least 1 hidden layer can accurately model non-linear systems even though the underlying causal links are unknown or not fully understood.

The use of neural networks has been hindered by computational problems for many years, because no learning rule was available to adjust the weights of the hidden layer connections before Rumelhart et al. (1986) developed the error back-propagation algorithm.

Even though their seminal paper was issued in a very interdisciplinary journal (*Nature*), it had virtually no impact outside the artificial intelligence and computer science field. In the last few years, however, neural networks have spread in many different fields, as their computational requirements have been matched by the capability of the latest generation of personal computers and their application has become feasible even for non-specialist scientists.

Neural networks seem to be a very promising tool for the empirical modeling of phytoplankton production. The most important conceptual advantage over conventional empirical (and semi-analytical) models is probably the possibility of collating heterogeneous information in a single computational framework, even though no theoretical guidelines are provided.

METHODS

In order to compare the performance of artificial neural networks with that of other empirical models, an estuarine data set was chosen.

The main reason for this choice was that estuaries are characterized by a very broad range of environmental conditions. Estimating phytoplankton productivity in estuaries is therefore a challenging test for both empirical and analytical models. Moreover, several recent papers have dealt with empirical models of phytoplankton productivity in estuaries and they could provide effective benchmarks for evaluating the performances of neural-network-based models.

The test data set (Table 1) was taken from Harding et al. (1986). It included 27 samples from Chesapeake Bay and Delaware Bay (USA) and was collected during 6 cruises that were carried out in 1982 and 1983. This particular data set was selected because the authors also provided extensive data analysis, discussed the relationships between phytoplankton production and other variables and presented an empirical model of phytoplankton production.

Values for the following variables were extracted from different tables of the cited paper:

- Phytoplankton production (PP, g C m^{-2} d^{-1})
- Surface irradiance (I_0, E m^{-2} d^{-1})
- Mean chlorophyll concentration (B, mg m^{-3})
- Depth of photic zone (Z_p, m)
- Light extinction coefficient (k_t, m^{-1})
- Station depth (H, m).

A new binary variable, *Bay*, was added to this data set to indicate the sampling station location (1 for Chesapeake Bay and 0 for Delaware Bay).

Table 1 The data set from Harding et al. (1986) includes primary production (*PP*), surface irradiance (I_0), mean chlorophyll concentration (*B*), depth of photic zone (Z_p), light extinction coefficient (k_t) and station depth (*H*). The binary variable in the last column (*Bay*) represents the location of each station [Chesapeake (CB) or Delaware Bay (DB)] and was added in order to improve the performance of the artificial neural network. In Harding et al. (1986) the same result was achieved by calculating a different slope of the linear empirical model for each bay

Cruise	Stn	PP $(\text{g C m}^{-2}\text{ d}^{-1})$	I_0 $(\text{E m}^{-2}\text{ d}^{-1})$	B (mg chl m^{-3})	Z_p (m)	k_t (m^{-1})	H (m)	Bay $(1 = \text{CB}, 0 = \text{DB})$
CB-1	I	0.101	28.0	1.3	2.6	1.80	6.0	1
Mar 82	II	0.877	3.1	5.6	6.0	0.77	7.5	1
	III	0.372	3.3	4.7	12.0	0.30	12.0	1
	IV	1.430	36.0	7.3	12.0	0.24	12.0	1
	V	0.317	33.0	4.2	11.0	0.15	11.0	1
CB-2	IV	0.476	16.0	13.0	1.7	2.80	5.0	1
Jul 82	III	1.780	58.0	9.2	3.4	1.40	7.0	1
	II	2.650	55.0	9.9	4.6	1.00	9.0	1
	I	0.611	52.0	4.7	8.1	0.57	11.0	1
CB-3	V	0.804	26.0	13.0	2.4	1.90	5.0	1
Oct 82	IV	0.954	18.0	10.0	4.7	0.98	8.0	1
	III	0.962	21.0	6.6	6.1	0.75	11.0	1
	II	0.603	20.0	3.7	9.5	0.49	12.0	1
	I	0.770	19.0	3.2	8.6	0.54	15.0	1
CB-4	V	0.089	36.0	3.7	1.3	3.40	6.0	1
Mar 83	IV	0.638	18.0	8.7	3.7	1.20	8.0	1
	III	1.910	41.0	12.0	4.6	1.00	9.0	1
	II	0.140	11.0	4.3	7.5	0.61	13.0	1
	I	0.369	34.0	3.9	6.6	0.70	12.0	1
DB-1	III	0.564	18.0	9.3	2.3	2.00	11.0	0
Nov 82	IV	0.110	10.0	6.2	1.3	2.70	12.0	0
	II	0.868	20.0	21.0	4.6	1.00	8.0	0
	I	0.323	24.0	2.9	5.3	0.86	25.0	0
DB-2	IV	0.021	6.6	5.5	1.1	4.30	8.0	0
Apr 83	III	0.171	12.0	7.8	1.8	2.50	10.0	0
	II	0.758	17.0	20.0	3.9	1.20	15.0	0
	I	0.326	23.0	6.4	6.1	0.68	19.0	0

The empirical model proposed by Harding et al. (1986) is based on surface irradiance (I_0) and the proportion of light attenuated by chlorophyll (Bk_c/k_t):

$$PP = a + b \ B\frac{k_c}{k_t} I_0$$

where k_c is a constant [0.015 m^{-1} (mg chl a m^{-3})$^{-1}$] and k_t is the light extinction coefficient, measured independently of Z_p.

Both this model and the one by Cole & Cloern (1984) were compared with models based on artificial neural networks.

Since the main concern of this paper is to point out a possible improvement in empirical models of phytoplankton production rather than in artificial neural networks, very basic networks and training algorithms were selected.

In fact, feed-forward networks with 1 hidden layer were used, all hidden and output nodes had sigmoid activation functions, and training was performed by error back-propagation. The learning rate was not

optimized (i.e. a constant value, $\eta = 1$, was used) and no momentum term was considered.

Five different networks were trained, all with only 1 output node that returned the *PP* estimate and bias nodes both in input and hidden layers.

The number of input nodes ranged from 3 to 5. Three input nodes were always fed with *B*, Z_p and I_0 data, whereas the fourth input node, when present, was associated to the *Bay* binary variable. A network with 5 input nodes was developed to show how additional information, even though not directly related to *PP*, can improve the accuracy of a model: in this case the fifth input node was fed with station depth (*H*).

The number of hidden layer nodes was determined by comparing the performances of different networks, with 1 to 12 nodes in the hidden layer The lowest mean squared errors were obtained by networks with 5 to 8 nodes in the hidden layer, depending on the input variables and on the number of input nodes.

In this paper the networks will be cited according to their structure, i.e. as x-y-z (input-hidden-output nodes). In other words, a 3-5-1 network is a network with 3 input nodes, 5 hidden nodes and 1 output node.

In order to be used as training patterns for the neural networks, all non-binary raw data were scaled into a [0,1] interval. As the neural networks have sigmoid activation functions, their training becomes easier if the training patterns do not contain values that are too close to the limits of this interval: therefore, data were scaled by dividing them by arbitrary maximum values somewhat larger than the maximum observed values (Table 2), rather than by their range. A small positive offset (0.15 g C m^{-2} d^{-1}) was added to raw PP data to avoid scaled values too close to 0.

The neural network training was carried out according to the following scheme:

· 14 patterns were randomly selected out of the 27 samples in the data set (the same sample could be selected more than once);
· White noise (i.e. a small random value, ±5% of the input value) was added to each input (Györgyi 1990);
· After each 14 pattern training cycle ('epoch', in neural network jargon) the network was validated on the whole data set and a mean squared error was computed;
· After 50000 epochs, training was stopped and the weights corresponding to the lowest mean squared error were saved.

This training procedure avoided overfitting, i.e. undesired reproduction of data set peculiarities and loss of the 'regularity' that is needed for good model generalization. This is particularly relevant when data sets include only a relatively small number of patterns (i.e. samples).

Finally, it should be stressed that the performances of the neural networks that are presented in this paper are to be considered as minimal estimates. In fact, network training was intentionally limited and further improvement is certainly possible (e.g more epochs, different network initialization, etc.).

RESULTS

The empirical model presented by Harding et al. (1986) provided a fairly good fit to the observed data set, even though they had to sacrifice its simplicity to achieve this result. In fact, instead of a single linear model (Fig. 1a.1), 2 different linear models were defined for Chesapeake and Delaware Bays and 2 outliers were excluded from the Chesapeake Bay data subset (Fig. 1b.1).

Each linear component of the model accounted for about 70% of the PP variance (Chesapeake Bay· r^2 = 0.69; Delaware Bay· r^2 = 0.77) and the overall fitting of the composite model was much better than in the case of a single linear model without outliers (r^2 = 0.745 instead of r^2 = 0.271; Fig. 1b.2, a.2).

However, it is interesting to note that both models tended to overestimate small PP values and to underestimate large PP values. Of course, this depends on the fact that the assumption of linearity was valid only for intermediate values of the independent composite variable $Bk_c/k_t I_0$ and of PP. This is also the reason why the non-zero intercept of the linear model is biologically meaningless.

Finally, although the composite model had a more symmetrical error distribution than the single linear model (Fig. 1a.3, b.3), the maximum error, excluding outliers, was still very large (1.034 g C m^{-2} d^{-1}).

A 3-6-1 and a 4-6-1 neural network were also trained on this data set (Fig. 1c.1, d.1). Chlorophyll concentration (B), light extinction coefficient (k_t) and surface irradiance (I_0) were used as inputs of both networks, whereas station location (Bay) was added as an input to the second network. The network output, of course, was always PP.

PP values predicted by neural network are plotted against observed PP values in Fig. 1c.2, d.2. The performances of these networks should be compared, respectively, with the single linear model and with the composite linear model.

The 3-6-1 neural network provided a better fit than the best linear model (r^2 = 0.900), whereas the 4-6-1 network, which takes station location into account as does the composite linear model, achieved a further

Table 2. Arbitrary maximum values were used to scale the data set into the [0,1] interval. Scaled data were obtained by dividing raw data by the arbitrary maximum. A small offset (0.15 g C m^{-2} d^{-1}) was added to raw PP data to avoid scaled values too close to interval limits

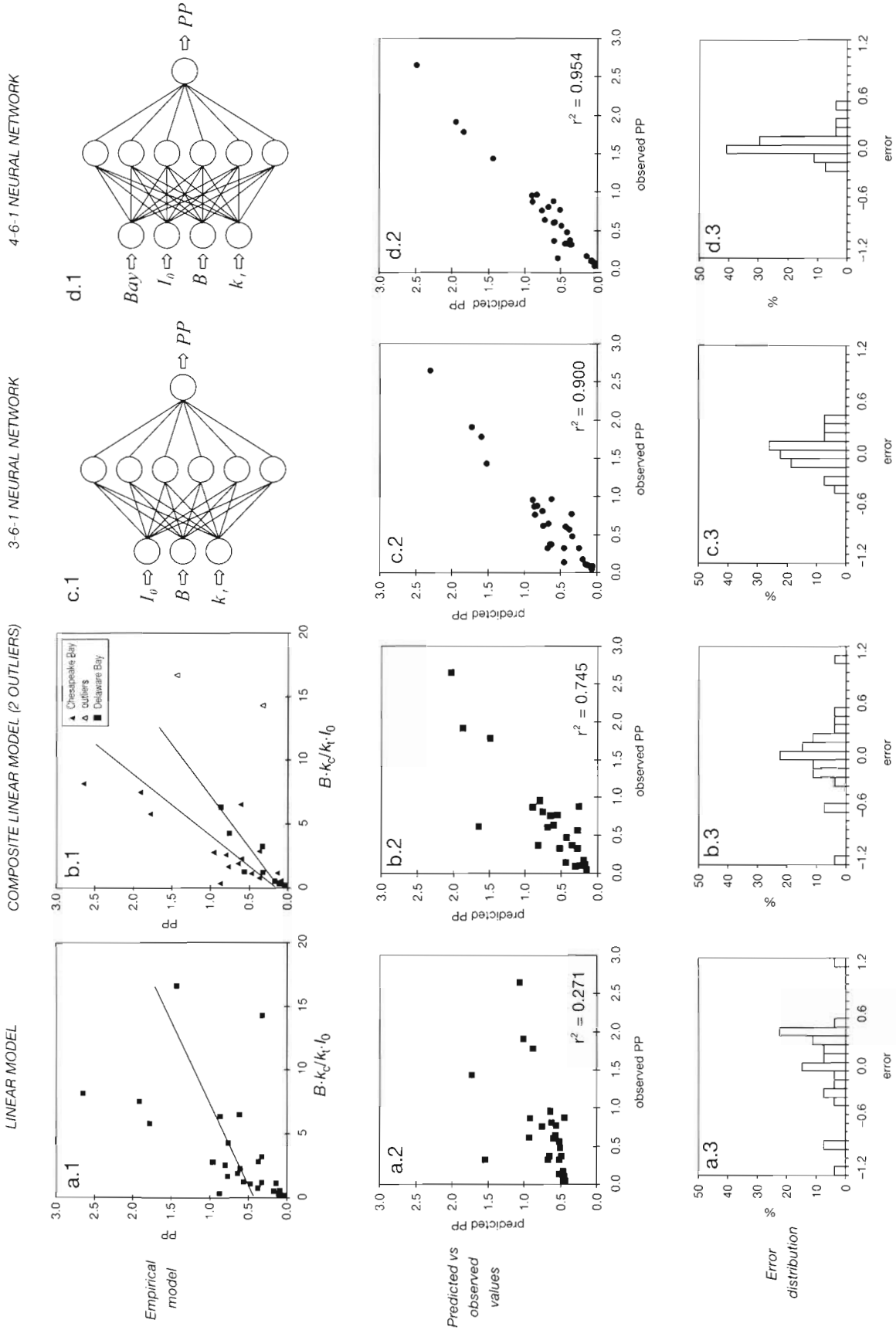|	PP (g C m^{-2} d^{-1})	I_0 (E m^{-2} d^{-1})	B (mg chl m^{-3})	Z_p (m)	k_t (m^{-1})
Arbitrary maximum	3	70	30	15	5

Fig. 1. Comparison between different empirical models of phytoplankton production (*PP*, g C m^{-2} d^{-1}). Models are based on surface irradiance (I_0, E m^{-2} d^{-1}), biomass (*B*, mg chl m^{-3}), light extinction coefficient (k_t, m^{-1}), light absorption by chlorophyll [$k_c = 0.015$ m^{-1}(mg chl m^{-3})$^{-1}$]. The second and fourth model (b & d) also take into account station location (*Bay*, Chesapeake Bay = 1 and Delaware Bay = 0). The neural network structures are simplified, as bias neurons are not shown. In the error distribution histograms, labels indicate the upper limit of each class

fitting improvement ($r^2 = 0.954$), although no outliers were excluded from the data set.

It should be stressed, however, that the meaning of r^2 as a measure of goodness of fit is not the same when dealing with linear models or neural networks. In fact, in the former case it has a unique, exact value, whereas in the latter case it is affected by the random (e.g. noise summed to input values) and arbitrary factors (e.g. maximum number of epochs) involved in the network training procedure. Therefore, some improvement is probably still possible, even while preserving the same level of model generality.

The error distributions of the neural-network-based models were satisfactory (Fig. 1c.3, d.3), as they were almost symmetrical and more leptokurtic than those of the linear models. Moreover, the neural network PP estimates were not systematically biased, as error means were negligible: -0.016 g C m^{-2} d^{-1} for the 3-6-1 network and -0.010 g C m^{-2} d^{-1} for the 4-6-1 network. The maximum errors with respect to the training set were also much smaller than those of linear models (0.424 and 0.403 g C m^{-2} d^{-1}).

The empirical model proposed by Cole & Cloern (1984, 1987), which had been selected as an example of a more generalized formulation, was also applied to the same data set. In order to preserve this generalized nature, no outliers were excluded from computations (Fig. 2a.1). This model accounted for almost 60% of the PP variance ($r^2 = 0.574$). A comparison between observed and predicted PP is shown in Fig. 2a.2. It is interesting to note that, as in the previous linear models, almost all the predicted values were overestimated when the observed PP values were smaller than 0.5 g C m^{-2} d^{-1} and underestimated when they were larger than 1.5 g C m^{-2} d^{-1}

The error distribution was nearly symmetrical (Fig. 2a.3), but the maximum error of the PP estimate exceeded 1 g C m^{-2} d^{-1} (1.071 g C m^{-2} d^{-1}).

A 3-5-1 neural network was trained on the same data set (Fig. 2b.1). The network outputs matched the observed PP data much better than the linear model ($r^2 = 0.940$), with no systematic error for low and high PP values (Fig 2b.2).

The error distribution was much more leptokurtic than that of the linear model (Fig. 2b.3), with a smaller maximum error (0.362 g C m^{-2} d^{-1}). The PP estimates were not systematically biased, as the error mean was very close to 0 (-0.009 g C m^{-2} d^{-1})

When the station location was added as a fourth input, a 4-8-1 neural network (Fig. 2c.1) achieved a further performance improvement.

The predicted PP values matched very closely the observed ones ($r^2 = 0.975$; Fig. 2c.2), with no evidence of systematic estimate errors.

The error distribution was symmetrical (Fig 2c.3) and its mean was negligible (0.008 g C m^{-2} d^{-1}). Moreover, the maximum error of the PP estimate for this model was 0.203 g C m^{-2} d^{-1}, i.e. 1/5 of the largest error of a comparable linear model.

An example of the calculations that are needed to obtain PP estimates from input variables using the 3-5-1 network is presented in Appendix 2.

DISCUSSION

Neural-network-based empirical models of phytoplankton production are far more effective than linear empirical models and the higher r^2 values bear witness to their superiority.

However, even though r^2 is an important criterion, it is obvious that a non-linear model provides an improvement in data fitting in comparison with a linear model. Other factors have to be taken into account when evaluating an empirical model of a complex process such as phytoplankton production.

The distribution of the predicted value errors is probably the most relevant, because a good model should not provide systematically biased estimates. In other words, the mean error should be close to 0 and the error distribution should be as symmetrical and leptokurtic as possible.

Neural-network-based models achieved good results even in this respect. In fact, their mean error was very small (± 0.008 to 0.016 g C m^{-2} d^{-1}) if compared with the mean error of the composite linear model (-0.065 g C m^{-2} d^{-1}) and negligible if compared with the null mean error of single linear models. Moreover, neural networks always had a better error distribution than conventional models.

The main processes that determine phytoplankton production can be approximated by linear or simple non-linear (e.g. logarithmic) functions only to a limited extent. Therefore, such models are not able to reproduce the behaviour of real systems when very low or high values of the independent variables are considered. On the other hand, neural networks with at least 1 hidden layer can model non-linear systems independently of their complexity. Of course, complex systems need complex networks, adequate training and a large data set to be modeled.

The differences between a neural-network-based model and a conventional linear model of phytoplankton production are summarized in Fig. 3. The surfaces correspond to estimated PP values for every combination of surface irradiance (I_0) and depth of photic zone values (Z_p) at a 15 mg m^{-3} chlorophyll concentration.
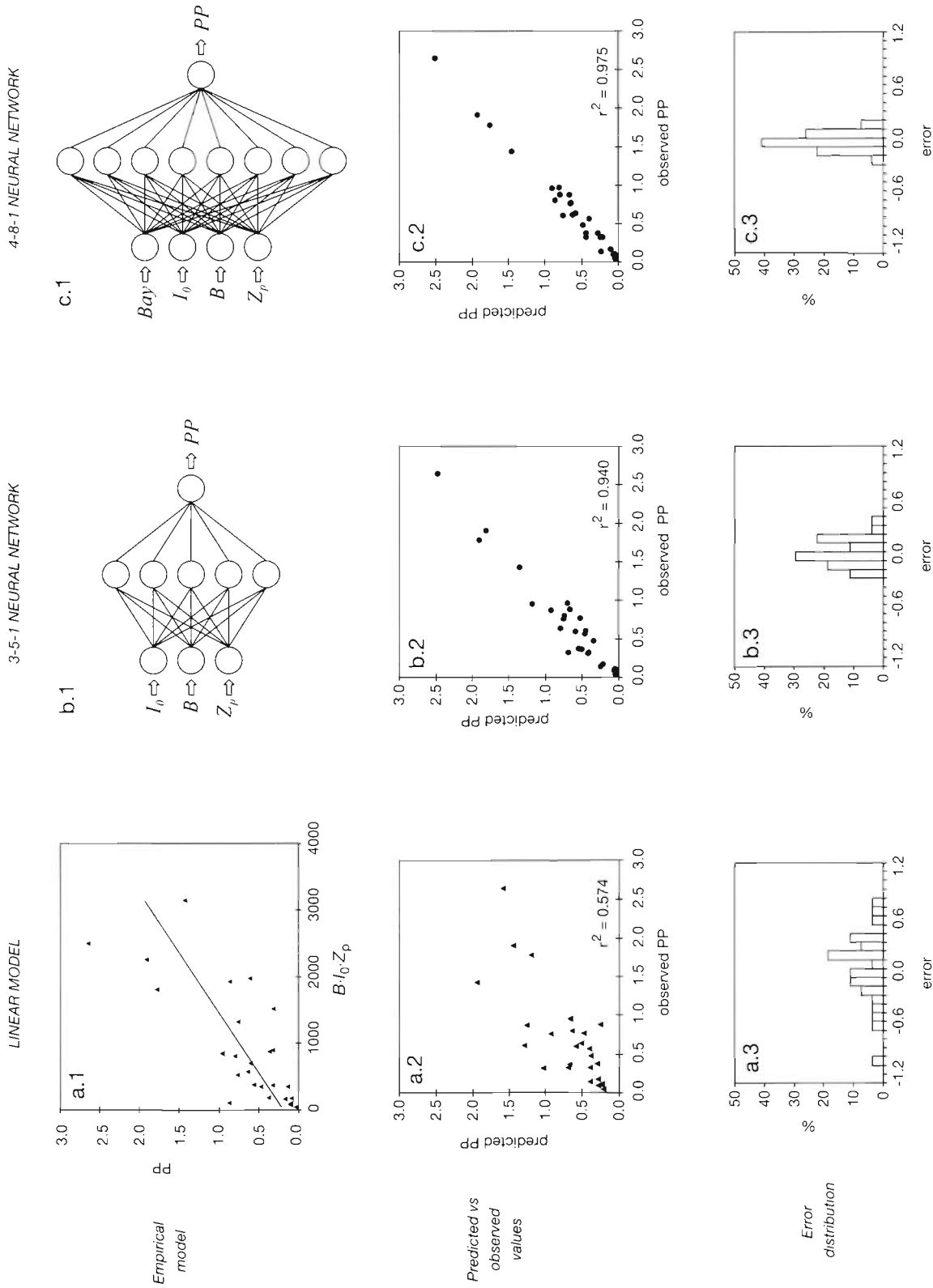
Fig. 2. Comparison between different empirical models of phytoplankton production (PP, g C m^{-2} d^{-1}). All models are based on surface irradiance (I_0, E m^{-2} d^{-1}), biomass (B, mg chl m^{-3}) and depth of photic zone (Z_p, m). In the 4-8-1 neural network station location was also taken into account (Bay, Chesapeake Bay = 1 and Delaware Bay = 0). The neural network structures are simplified, as bias nodes are not shown. In the error distribution histograms, labels indicate the upper limit of each class
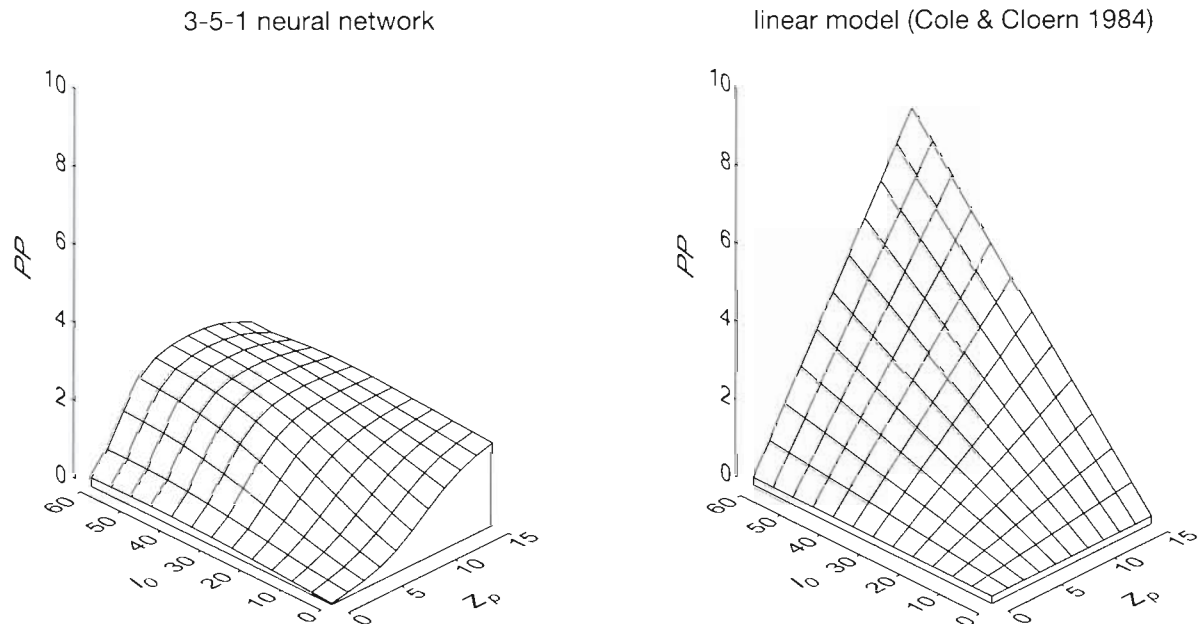
Fig. 3. Estimated phytoplankton production (PP) as a function of surface irradiance (I_0) and depth of photic zone (Z_p) for a 15 mg m^{-3} chlorophyll concentration. Outputs from the 3-5-1 neural network (left) and the Cole & Cloern (1984) linear model (right) are compared

The output of the neural network (see Fig. 2b) never exceeds likely PP values and shows interesting details of the relationships between PP and the independent variables. For example, it can be seen that there is little variation in PP values for different I_0 levels if Z_p is very low or high. Moreover, the steepest gradient of the surface slope in the Z_p direction (i.e. the maximum of the partial derivative with respect to Z_p) is not the same for different I_0 values: when I_0 is high it corresponds to low Z_p values, whereas when I_0 is low it corresponds to intermediate Z_p values. On the other hand, none of the abovementioned details can be reproduced by the linear model (see Fig. 2a), which always returns the same output provided that $I_0 \cdot Z_p$ is constant.

There are of course some inconsistent features even in the neural network output (e.g. significant PP values are predicted for very low I_0 levels, especially when Z_p is high), but they depend on the small number of patterns in the training data set rather than on intrinsic limitations, as in the case of linear models.

Another advantage of neural networks over conventional empirical models is the possibility of adding new independent variables even though their relationships with PP are not known or are difficult to be formally defined. In order to provide an example of this capability, a new network was trained by adding station depth (H) to the input variables that were used in the neural network shown in Fig. 2c.1 (i.e. Bay, I_0, B and Z_p).

The rationale for this choice is that in shallow stations Z_p has a particular meaning when it coincides with H, i.e. when water column transparency is high enough to make the photic zone depth limited by station depth (this typically happens when a Secchi disk is used to assess photic zone depth, so that it is not possible to extrapolate a theoretical value greater than station depth). Therefore, if Z_p is equal to H then Z_p is certainly underestimated and the output of the model should take this fact into account.

The resulting network worked optimally with 7 nodes in the hidden layer, hence its final structure was 5-7-1. This neural network worked better than the 4-8-1 network, as r^2 increased from 0.975 to 0.989 (Fig. 4a). Moreover, the error distribution was improved (Fig. 4b) and the maximum error was only 0.137 g C m^{-2} d^{-1}, i.e. 2/3 of the 4-8-1 network maximum error.

It is obvious that it would be almost impossible to obtain a similar result if a conventional approach was used. At least 3 linear models would be needed: 1 for Delaware Bay stations (always $Z_p < H$) and 2 for Chesapeake Bay stations ($Z_p < H$ or $Z_p = H$). However, each linear model would be less accurate than a comparable neural network and their combination could never work better than an empirical model based on a single neural network.

The only theoretical advantage of conventional empirical models over neural networks is that their
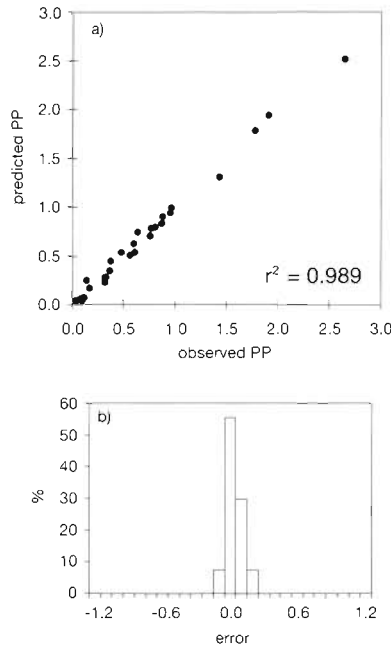
Fig. 4. Station depth (H) was added as a new input to the neural network based on B, Z_p, I_0 and station location (Bay). The performance of the resulting 5-7-1 neural network was improved with respect to the 4-input network, as (a) r^2 increased from 0.975 to 0.989 and (b) error distribution was more leptokurtic, with a 33% smaller maximum error (0.137 g C m^{-2} d^{-1})

parameters provide information about the relative importance of the independent variables (although this is not true when composite variables are used). However, the same results can be obtained by performing a sensitivity analysis of the neural-network-based models. Moreover, in this case it is possible to detect the strength of the actual link between each input variable and PP, rather than the relative weight of independent variables given a simplified theoretical model.

Finally, it should be stressed that the error back-propagation neural networks that were presented in this paper are very basic and widely applied to different problems. As research in this field is continuously providing new types of both networks and training algorithms, accurate and general empirical models of phytoplankton production are probably not out of reach.

As Balch et al. (1989) pointed out, an idealized algorithm for the prediction of PP should account for 100% of its variance, providing a 1:1 relationship between predicted and observed PP values. Provided that adequate data sets are available and that deterministic relationships exist between some input variables and PP, neural-network-based empirical models are as close as possible to this definition.

Appendix 1. An introduction to error back-propagation neural networks

An error back-propagation neural network consists of several layers of nodes somehow analogous to neurons: an input layer (i), 1 or more hidden layers (h) and an output layer (o).

Each node in a layer receives its input from the output of the previous layer nodes or from the network input. The connections between nodes are associated with synaptic weights (**W**, **Z**) that are iteratively adjusted during the training process. A simple example of such a network is shown in Fig. A1. It corresponds to the network that was used with I_0, B and Z_p as inputs (see Fig. 2b.1).

An additional node with a constant output (usually 1) is often added to the input and hidden layers. These nodes are known as bias nodes. Their role in neural networks is very similar to that of the constant term in multiple regression, i.e. they permit shifting of the origin of the hyperspace defined by the input variables.

Each hidden and output node is associated with an activation function, i.e. a differentiable function of the node total input. Several functions can be used as activation functions, but the most common choice is the sigmoid function:

$$f(a) = \frac{1}{1+e^{-a}}$$

Provided that the activation function of the hidden layer nodes is non-linear, an error back-propagation neural network with an adequate number of hidden nodes is able to approximate every non-linear function.

A neural network works at its best if all its synaptic weights have been properly adjusted. The error back-propagation
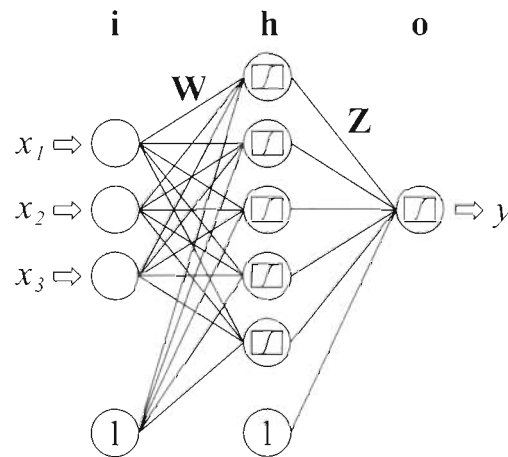

Fig. A1. A 3-5-1 error back-propagation neural network. Three input nodes (i), 5 hidden layer nodes (h) and 1 output node (o) are shown. Connections between nodes are shown by solid lines: they are associated with synaptic weights (**W**, **Z**) that are adjusted during the training. Two bias nodes are also shown, labelled with 1 (i.e. their output). The sigmoid activation functions are plotted within the nodes

algorithm is a way to compute these weights and involves 4 steps: (1) the network is initialized by assigning random values to synaptic weights; (2) a training pattern is fed and

Appendix 1 (continued)

propagated forward through the network to compute an output value for each output node; (3) computed outputs are compared with the expected outputs; (4) a backward pass through the network is performed, changing the synaptic weights on the basis of the observed output errors. Steps 2 through 4 are iterated for each pattern in a training set, then the network performance is checked (usually on the basis of a mean squared error) and a new set of training patterns is submitted to the network (i.e. a new epoch is started) if it needs further optimization.

In the case of neural networks with a single hidden layer, like the ones that were used for phytoplankton production modeling, the forward propagation step is carried out as follows:

$$h_k = \sum_{j=1}^{n_i+1} i_j\, w_{jk} \qquad (k = 1, \ldots, n_h)$$

where i_j is the output of the input layer (i.e. the network inputs and 1 for the bias node) and w_{jk} is the weight of the connections between input and hidden layers. To compute the outputs of the hidden layer, these weighted sums are passed to the activation function, but for the bias node, which is forced to have an output equal to 1:

$$h_k = f(h_k)$$
$$h_{n_h+1} = 1$$

Then, the network outputs are computed in the same way:

$$o_l = \sum_{k=1}^{n_h+1} h_k\, z_{kl} \qquad (l = 1, \ldots, n_o)$$
$$o_l = f(o_l)$$

After the forward propagation, estimated outputs o_l are compared with expected outputs y_l and a mean quadratic error for the current pattern is computed as.

$$E = \frac{1}{n_0} \sum_{l=1}^{n_o} (y_l - o_l)^2$$

Then, in the back-propagation step, all the synaptic weights are adjusted in order to follow a gradient descent on the error surface.

For the connections between hidden and output layers, z_{kl} are changed as:

$$z_{kl} = z_{kl} + \eta\, \delta_l^o\, h_k \qquad (k = 1, \ldots, n_h + 1;\, l = 1, \ldots, n_o)$$

where η is a constant (learning rate) and:

$$\delta_l^o = (y_l - o_l) \cdot f'(o_l) = (y_l - o_l) \cdot o_l(1 - o_l)$$

The weights w_{jk} of the connections between hidden and input layer are also adjusted:

$$w_{jk} = w_{jk} + \eta\, \delta_k^h\, i_j \qquad (k = 1, \ldots, n_h + 1;\, j = 1, \ldots, n_i + 1)$$

where δ_k^h is computed as:

$$\delta_k^h = f'(h_k) \sum_{l=1}^{n_o} \delta_l^o z_{kl} = h_k(1 - h_k) \cdot \sum_{l=1}^{n_o} \delta_l^o z_{kl}$$

The network training is iterated until a given condition is met. Minimization of the quadratic error is usually involved, but other criteria can also be used.

It has to be stressed, however, that the weight adjustment process does not provide a unique optimized result, since many non-deterministic factors (e.g. different starting values of the synaptic weights) can affect the network training. Moreover, the gradient descent on the error surface might find a local minimum.

An up-to-date, more general and comprehensive introduction to the fundamental neural network architectures can be found in Abdi (1994).

Error back-propagation neural networks are available in many commercial, shareware and public domain software packages. However, the FORTRAN source code of the very simple implementation that was used for *PP* modeling can be requested from the author.

Appendix 2. A worked example of feed-forward calculation for the 3-5-1 neural network

In the particular case of the trained 3-5-1 network (see Fig. 2b.1) the **W** and **Z** matrices were:

$$\mathbf{W} = \begin{pmatrix} 5.147 & -18.184 & -1.412 & -2.172 & 33.078 & 0.120 \\ -4.870 & 1.946 & 4.283 & -3.169 & -2.157 & 0.072 \\ -0.289 & -0.302 & -0.247 & 0.380 & 0.376 & 0.397 \end{pmatrix}$$

$$\mathbf{Z} = (-5.578 \quad -10.916 \quad -5.260 \quad -8.984 \quad -6.957 \quad 13.735)^t$$

Network inputs were scaled in a [0,1] interval, assuming a fixed arbitrary maximum value for each input variable. These values are:

$$\mathbf{x}_{max} = \begin{pmatrix} 70 \\ 30 \\ 15 \end{pmatrix}$$

where the first element corresponds to surface irradiance, the second to chlorophyll concentration and the third to photic zone depth. For example, the input values:

$$\mathbf{x}_{raw} = \begin{pmatrix} 55.0 \\ 9.9 \\ 4.6 \end{pmatrix} \quad \text{i.e.} \quad \begin{pmatrix} I_0, \mathrm{E\ m^{-2}\ d^{-1}} \\ B, \mathrm{mg\ chl\ m^{-3}} \\ Z_p, \mathrm{m} \end{pmatrix}$$

are scaled and passed to the input nodes i_j as:

$$i_j = \frac{x_{raw_j}}{x_{max_j}} \qquad (j = 1, \ldots, n_i)$$

Hence, the input vector is:

$$\mathbf{i} = \begin{pmatrix} 0.786 \\ 0.330 \\ 1.000 \end{pmatrix}$$

where the last element is constant and corresponds to the bias node.

The network output (i.e. the predicted *PP*) is computed by carrying out the forward propagation as described in Appendix 1.

Since in the training patterns the output variable (*PP*) was also scaled into a [0,1] interval, the network output has to be scaled back to the original units:

$$PP = o \cdot PP_{max} - PP_{offset}$$

With $PP_{max} = 3$ and $PP_{offset} = 0.15$, the predicted *PP* value is 2.479 g C m^{-2} d^{-1}.

In particular, this is the predicted value that corresponds to the highest *PP* value in the data set (see the rightmost point in Fig. 2b.2).

Other predicted values can be easily computed following the same procedure with different $\mathbf{x}_{raw}$ input vectors.

LITERATURE CITED

Abdi H (1994) A neural network primer. J Biol Syst 2(3): 247–281

Balch WM, Eppley RW, Abbott MR (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 BE, Cloern JE (1984) Significance of biomass and light-availability to phytoplankton productivity in San Francisco Bay. Mar Ecol Prog Ser 17:15–24

Cole BE, Cloern JE (1987) An empirical model for estimating phytoplankton productivity in estuaries. Mar Ecol Prog Ser 36:299–305

Conover SAM (1956) Oceanography of Long Island Sound, 1952–1954. IV. Phytoplankton. Bull Bingham Oceanogr Coll 15:62–112

Eppley RW, Stewart E, Abbott MR, Heyman U (1985) Estimating ocean primary production from satellite chlorophyll. Introduction to regional differences and statistics for the Southern California Bight. J Plankton Res 7:57–70

Györgyi G (1990) Inference of a rule by a neural network with thermal noise. Phys Rev Lett 64:2957–2960

Harding LW Jr, Meeson BW, Fisher TR 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

Keller AA (1988) Estimating phytoplankton productivity from light availability and biomass in the MERL mesocosms and Narragansett Bay. Mar Ecol Prog Ser 45:159–168

Lorenzen CJ (1970) Surface chlorophyll as an index of the depth, chlorophyll content, and primary productivity of the euphotic layer. Limnol Oceanogr 15:479–480

Pennock JR, Sharp JH (1986) Phytoplankton production in the Delaware Estuary: temporal and spatial variability. Mar Ecol Prog Ser 34:143–155

Riley GA (1939) Plankton studies II. The Western North Atlantic, May–June, 1939. J Mar Res 2:145–162

Riley GA (1956) Oceanography of Long Island Sound, 1952–1954. IX. Production and utilization of organic matter. Bull Bingham Oceanogr Coll 15:324–334

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323: 533–536

Ryther JH, Yentsch CS (1957) The estimation of phytoplankton production in the ocean from chlorophyll and light data. Limnol Oceanogr 2:281–286

Smith RC, Eppley RW, Baker KS (1982) Correlation of primary production as measured aboard ship in Southern California coastal waters and as estimated from satellite chlorophyll images. Mar Biol 66:281–288