ELSEVIER

Ecological Modelling 146 (2001) 33–45

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

Advances in neural network modeling of phytoplankton primary production

Michele Scardi

Department of Zoology, University of Bari, Via Orabona 4, 70125 Bari, Italy

Abstract

Neural networks are powerful tools for phytoplankton primary production modeling, even though their application might be hindered by the limited amount of available data. Some new approaches that could enhance neural network models to overcome this problem are presented and discussed in this paper. For instance, co-predictors allow to improve neural network estimates when additional inputs from a broader range of variables are selected. Theoretical knowledge about biological processes can be easily embedded into neural network models by means of a constrained training procedure. Finally, information derived from both existing models and real data can be effectively exploited by a metamodeling approach. Since the underlying rationale applies to a wide spectrum of problems, the proposed approaches are not confined to phytoplankton primary production modeling, but they can also play a role in other ecological applications. © 2001 Elsevier Science B.V. All rights reserved.

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

1. Introduction

Empirical modeling of phytoplankton primary production has always been based on predictive variables (mainly phytoplankton biomass and irradiance) that are more easily available and cheaper to measure than primary production. Since direct primary production measurements are not only expensive, but also difficult and time consuming, the role of empirical models in oceanographic research has always been a major one both at global scale (Bidigare et al., 1992; Behrenfeld and Falkowski, 1997) and at smaller spatial scales (e.g. Cole and Cloern, 1984, 1987; Keller, 1988; Scardi and Harding, 1999). This is especially true if the huge amount of information about predictive variables that can be obtained by remote sensing is taken into account.

In this framework artificial neural networks provide an effective alternative to conventional modeling techniques (Scardi, 1996, 2000) and their application to regional scale phytoplankton primary production modeling has already been presented (Scardi and Harding, 1999).

The context of ecological modeling is quite different from that of most neural network applications, as data sets and knowledge are often very limited with respect to the complexity of the real world processes. Therefore, relationships between variables are only partly known and understood,

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

PII: S0304-3800(01)00294-0

as they are usually studied by analyzing correlations rather than by defining causal pathways in a strictly deterministic framework.

However, in order to fully exploit the flexibility and the power of neural networks in ecological modeling, which is mainly due to their ability in modeling complex non-linear ecological systems (Lek et al., 1996), additional information or new training strategies should be used by means of unconventional approaches.

For instance, since neural network models are quite robust with respect to redundant inputs, *co-predictors* can play an important role in improving existing models, exploiting non-linear, even non-monotonic relationships between primary production and other variables that are not directly involved in photosynthetic processes.

Even though one of the major advantages of neural networks over conventional non-linear techniques is the absence of an a priori definition of the model structure, this flexibility can be difficult to control when data are not enough to effectively train and validate a model. In order to obtain biologically sound models in data-limited situations, theoretical knowledge can be used to perform a *constrained training* of the neural networks.

From a more general viewpoint, empirical models are as good as the data sets they have been built upon, and neural network models obviously comply to this rule. However, phytoplankton primary production data are seldom as abundant as needed, because of the already mentioned acquisition problems. In order to extend the predictive ability of a neural network model, other existing models can be used as a source of baseline information for time and/or space regions where real data are not available. In other words, merging real and modeled information by means of a *metamodeling* procedure is a very effective way to enhance and generalize neural network models as much as possible, even in absence of large data sets.

These approaches will be presented and discussed in this paper, focusing on applications of neural networks to phytoplankton primary production modeling. However, co-predictors, constrained training and metamodeling can effectively play a role also in other applications, because the underlying rationale apply to a much wider spectrum of problems.

2. Materials and methods

All data sets have been rescaled into a [0, 1] interval before training the neural networks, whereas neural network output, i.e. phytoplankton primary production, has always been scaled back to its original units before plotting the results or performing the error calculations.

All the neural network models presented in this paper have been trained using the error back-propagation (EBP) algorithm (Rumelhart et al., 1986), even though it has been tuned in different ways according to each data set. Unit bias nodes as well as sigmoid activation functions

$$f(a) = 1/(1 + e^{-a}),$$

where a is the activation value (i.e. the scalar product of the synaptic weight and input vector), have been used in all the neural networks, both in hidden and output layers.

The neural network training has always been performed according to the 'learning per pattern' paradigm. Moreover, training patterns were submitted in random order at each learning epoch in order to avoid that memorization of the submission order could adversely affect the training.

When needed, the neural network training has been carried out according to an early stopping strategy. Even though the term early stopping sometimes refers to a training procedure stopped after a given number of epochs, in this paper it will be used to indicate a training procedure stopped as soon as the validation error begins to increase.

Finally, jittering, i.e. addition of a small amount of noise to input patterns at each epoch (Györgyi 1990), has been performed during all the training procedures. Gaussian noise with $\mu = 0$ and $\sigma = 0.01$ has always been used. Jittering helps neural network generalization by providing a virtually unlimited number of artificial training patterns that are closely related, even though not exactly identical, to the original ones.

The neural network structures have been defined by a heuristic approach: given a single hidden layer, the number of neurons in it was allowed to vary from one half to the double of the number of input neurons and the best performing structure was chosen.

Mean square error (MSE), i.e. the sum of the square deviations of the neural network outputs from the target values, has been used as the main neural network performance criterion both during training and when comparing different neural networks. Error distributions were also taken into account in order to check whether the neural network estimates were unbiased or not, but they were not considered as a primary criterion for neural network structure selection.

Phytoplankton primary production has been always considered as integrated within the euphotic zone, i.e. from the surface to the depth where irradiance was 1% of the surface value. Neural network input (i.e. predictive) variables, on the contrary, included only surface or depth independent measurements. When sampling date and station longitude were used, they were both coded as two derived variables ($date_1$ and $date_2$ and $long_1$ and $long_2$, respectively), using a trigonometric transformation that mapped them onto a unit diameter circle. The following transformations were used:

$$\text{date}_1 = \frac{1}{2}\left[\cos\left(\frac{2\pi \ \text{day-of-the-year}}{365}\right)+1\right]$$

$$\text{date}_2 = \frac{1}{2}\left[\sin\left(\frac{2\pi \ \text{day-of-the-year}}{365}\right)+1\right]$$

$$\text{long}_1 = \frac{1}{2}\left[\cos\left(\frac{2\pi \ (\text{longitude}+180)}{360}\right)+1\right]$$

$$\text{long}_2 = \frac{1}{2}\left[\sin\left(\frac{2\pi \ (\text{longitude}+180)}{360}\right)+1\right]$$

Information about specific data sets and training procedures are provided in the following sections for each example presented in this paper.

2.1. Co-predictors

The data set selected to test the role of co-predictors included data from all the oceans. It has been derived from the one used by Scardi (2000), but 304 new patterns were also added, so its size increased from 2218 to 2522 patterns. Input variables included sampling date (two derived variables), latitude, longitude (two derived variables), average depth (log-transformed), standard deviation of the average depth, sea surface temperature, day length, surface irradiance and phytoplankton biomass (as chlorophyll *a*).

Bathymetric information was obtained from a global grid with a quarter degree mesh size. Average depth and standard deviation of depth were computed within cells spanning five meshes in longitude (i.e. 1.25°) and three meshes in latitude (i.e. 0.75°). Average depth was then taken as its base 10 logarithm, since it can be assumed that any depth variation is more relevant in shallower than in deeper regions. The size of the windows within which these statistics were computed was defined after empirical tests. A wider size in longitude makes sense even from a theoretical point of view, because most continental margins are aligned from North to South and therefore high longitudinal variability in bathymetric data provide a good criterion for spotting those regions. In fact, standard deviation of depth allows to identify, given similar log average depths, regions where steep bathymetrical gradients are present and regions where depth is more uniform.

The new neural network global model, which included also day length as a new input variable, had an 11-14-1 structure (input-hidden-output nodes). Both variable learning rate and momentum were used. In particular, training was performed in three steps, with the learning rate set to 0.9, 0.5 and 0.1, respectively. In the meantime the momentum term was set to 0.1, 0.5 and 0.9.

The available data set ($n = 2522$) was divided into four subsets. One of them was used as validation set ($n = 630$), another one was only used to test the performance of the final model ($n = 631$) and two were combined to obtain a larger training set ($n = 1261$).

2.2. Constrained training

A small data set was used to demonstrate the constrained training of a neural network model.

This approach is particularly useful when the available data are not as abundant as needed and overfitting is likely to happen. In particular, only 97 patterns from Western Mediterranean were used, considering sea surface temperature, surface irradiance and phytoplankton biomass (as chlorophyll *a*) as input variables and depth-integrated phytoplankton primary production as output variable.

A very simple neural network structure (3-4-1) was used, with unit learning rate and null momentum term, in order to allow a straightforward comparison of the different training strategies.

In particular, three different models were obtained: (1) an overtrained model, trained on all the available patterns; (2) a generalized model, trained using two thirds of the available patterns ($n = 65$) for training and one third ($n = 32$) for validation, according to an early stopping strategy; (3) a constrained model, trained on all the availabe patterns with an additional penalty term in the MSE calculation, which depended on the deviations of the primary production vs. surface irradiance and phytoplankton biomass surfaces from a given theoretical shape.

The penalty term in the MSE calculation was obtained by examining the response surfaces of primary production vs. surface irradiance and phytoplankton biomass. These response surfaces were allowed to have no more than one maximum and four minima within the given range of irradiance and biomass values (0–60 E m^{-2} per day and 0–2 mg Chl m^{-3}, respectively). This condition, if checked throughout the whole sea surface temperature range, insures that the model is biologically meaningful.

In particular, primary production surfaces were generated for each sea surface temperature in the (12–23°C) range with steps of 1°C. Primary production surfaces were approximated by 10×10 grids, so the mesh size was 1/10 both of the irradiance and of the biomass range. Minima and maxima of the surfaces could be easily found by looking for those nodes where primary production estimates were larger (or lower) than all the neighbouring nodes. If more than one maximum or more than four minima were found, then the penalty term was increased by one unit for each maximum or minimum exceeding these limits. The resulting penalty term was then multiplied by the MSE and added to the actual MSE in order to compute the corrected error E^*:

$$E^* = \text{MSE} + \text{penalty} \times \text{MSE}.$$

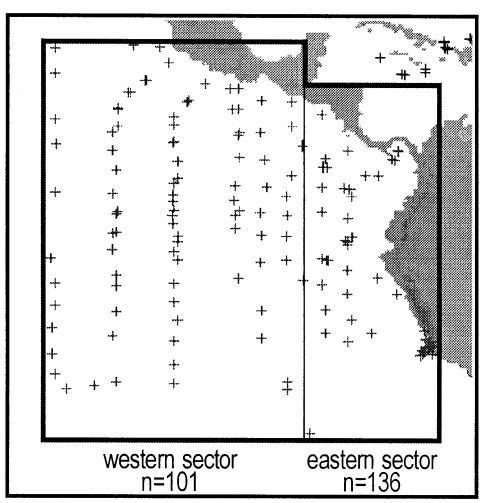

Fig. 1. Two subsets of a global data set from the Eastern Equatorial Pacific Ocean were used to demonstrate the metamodeling approach. Phytoplankton biomass and primary production are lower in the western sector than in the eastern one, which includes more coastal stations.

2.3. Metamodeling

Two subsets of the global data set from Eastern Equatorial Pacific have been selected to train a neural network metamodel. One of these subsets included 101 patterns in a rectangular sector whose oppostite corners were located at 25°S, 120°W and at 20°N, 90°W, whereas the other included 136 patterns and was located East of the former, in the rectangular sector whose opposite corners were located at 25°S, 90°W and at 15°N, 75°W (Fig. 1). The main difference between these two sectors was in their average phytoplankton biomass and primary production, that were larger in the eastern sector, which is closer to the west coast of South America.

As an additional source of information a global phytoplankton primary production model (the Vertically Generalized Production Model, aka

VGPM, by Behrenfeld and Falkowski, 1997) was used, even though it was necessary to perform a linear correction of the original formulation (i.e. $PP = 0.523\ PP_{VGPM}$) in order to improve its performance within the selected geographic area.

At first, a neural network model was trained on patterns from the western sector only, then a metamodeling approach was tested using real training patterns from the western sector and real input patterns with output target primary production modeled according to VGPM from the eastern sector. In other words, this example simulated a very likely scenario, i.e. a study in which values for predictive variables were available everywhere (e.g. by remote sensing), whereas target values for the NN output were not available for a whole geographic region. Therefore, NN output target values in the training and validation sets were arranged by mixing known data (western sector, $n = 101$) and estimates from an existing model (eastern sector, $n = 136$). Of course, only real primary production data from both sectors were used to test the metamodel performance.

Neural network input variables included sampling date (two derived variables, see Section 2.1), latitude, longitude, average depth (log-transformed), standard deviation of the average depth, sea surface temperature, surface irradiance and phytoplankton biomass (as log-transformed chlorophyll *a*). Phytoplankton primary production was also log-transformed. Both the neural network model for the western sector and the metamodel had a 9-7-1 structure. They were trained using early stopping as well as jittering to improve generalization, with one third of the available patterns in the validation set and the remaining patterns in the training set.

3. Results

3.1. Co-predictors

Co-predictors are variables not directly related to the ones that are to be predicted, even though they provide information that allows to improve the accuracy of the estimates. In other words, they are correlated to the variables that are to be predicted, rather than linked to them by causal relationships. This kind of information cannot be exploited by a strictly deterministic approach, but it comes handy if an empirical one has been chosen, as in the case of neural network applications.

For instance, bathymetric information, which is significantly correlated to phytoplankton primary production (Spearman's $r = -0.318$, $P < 0.01$, $n = 2522$), allowed to improve primary production estimates when added to a previous neural network global model (Scardi, 2000), even though it is obvious that depth doesn't affect the photosynthetic activity of phytoplankton directly.

Apart from the statistical issues, the rationale for the choice of bathymetric information as a source of co-predictors is that primary production is usually higher in coastal areas, on the continental shelf and in upwelling areas. That is true independently of its latitudinal variations at global scale and for any given level of phytoplankton biomass, because of the availability of a richer nutrient pool. From a more general point of view, water column depth, although not directly related to phytoplankton primary production, affects water column dynamics, which in turn is related to nutrient availability, that certainly drives primary production.

Log-transformed average depth and standard deviation of depth within a given area can help to identify coastal, continental shelf and potential upwelling areas, so these new variables were added to the NN model, that was subsequently retrained.

The comparison between the model with co-predictors and the existing one (Scardi, 2000) is shown in Fig. 2. It is interesting to notice that both the training/validation data set and the testing data set provided better estimates with the new neural network model than with the older one. The improvement in mean square error was not dramatic ($MSE = 330\,233$ for the new model and $MSE = 405\,117$ for the older one), but it was certainly noticeable, especially if it is taken into account the fact that the older NN model was already much more effective than other empirical models.

It is interesting to notice that both models were not very accurate when low observed primary production values were involved. This problem,

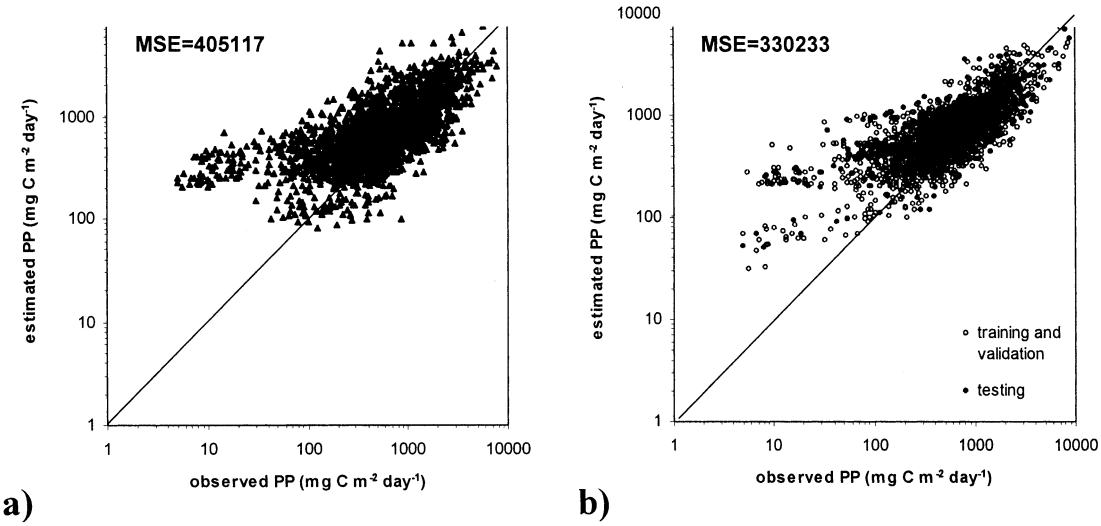

Fig. 2. Estimated vs. observed phytoplankton primary production: (a) from the 7-7-1 global model in Scardi (2000); (b) from the new 11-14-1 model that included bathymetric variables as co-predictors. Mean square error (MSE) was reduced by almost 20% in the latter model.

however, affects almost every primary production model and it is related to the quality of the observed data. In fact, observed primary production values should not be considered as 'true' values, because they are often affected by large measurement errors (Eppley, 1980). Nevertheless, they were treated as if all of them were 'true' values during model development (i.e. training, validation and testing of neural networks), because there was no way to distinguish real low primary production data from underestimated measurements. This implies that when low primary production values are taken into account, the noise to signal ratio is usually quite high and, consequently, predictability is very poor. Of course, NN models are not exceptions to this rule.

The different behavior in the low range of primary production estimates, as shown in Fig. 2, was the most evident difference between the old 7-7-1 model and the new 11-14-1 model with co-predictors because of the log scale. In spite of this the improvement in MSE that was obtained by the latter model mainly depends on more accurate estimates in mid-range primary production values. The contribution to the sum of the square errors of different classes of primary production values is shown in Fig. 3. The 11-14-1 NN model with co-predictors provided a lower error level for all classes, but for very high primary production values. In the latter case the old 7-7-1 NN model performed slightly better. The largest differences between the two models were observed in the 100–2000 mg C m^{-2} per day range, which includes about 90% of the available cases (2257 out of 2522).

The distribution of the sum of square errors clearly shows that high production values play a very important role in driving the NN model training, as their contribution to the sum of square errors (and to the MSE, obviously) is much more relevant than the one provided by low production values (<100 mg C m^{-2} per day). This difference, however, makes sense from a biological point of view, as high production values are affected by smaller relative errors if compared to low production values.

Other evaluation criteria provided further insights into the role of bathymetric information. For instance, a sensitivity analysis of the neural network model revealed that log average depth was the second most influential predictive variable. Feeding all the available patterns (training,

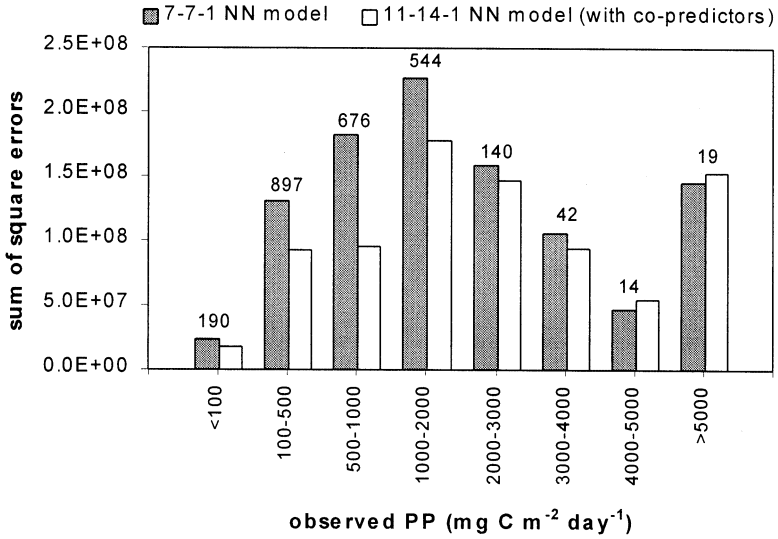

Fig. 3. Distributions of the sum of square errors for primary production estimates among different classes of observed phytoplankton primary production values. The figure above each couple of bars indicates the number of primary production values within the range indicated by the label below the bars. The grey bars represent the sum of square errors for the 7-7-1 NN model in Scardi (2000), whereas the white ones refer to the 11-14-1 NN model with bathymetric co-predictors.

validation and testing) after the addition of white noise in the $[-0.5, +0.5]$ range to all the input variables, one at time, returned perturbated primary production estimates. The maximum increase in MSE after this kind of perturbation was obviously recorded in the case of phytoplankton biomass data (1601% MSE increase with respect to original data), but second came log average depth (181% MSE increase), whereas only latitude produced an increase in MSE larger than 100% among the other input variables.

Another interesting point of view about the role of bathymetric information was provided by the relationship between MSE and average depth (Fig. 4). In fact the variation in MSE made much more ecological sense when bathymetric information was taken into account. In the latter case MSE increased at the margin of the continental shelf and then decreased with depth, i.e. when distance from continent and shelf margin increased and more homogeneous oceanographic conditions were found. In the case of the older model, such pattern was only found in shallower regions (depth < 3000 m), whereas MSE tended to vary in an apparently random way in deeper regions.

3.2. Constrained training

Several techniques may induce regularization in neural network training, but all of them rely on heuristic approaches. Early stopping, jittering (i.e. noise addition) or weight decay are some of such techniques. However, when neural networks are

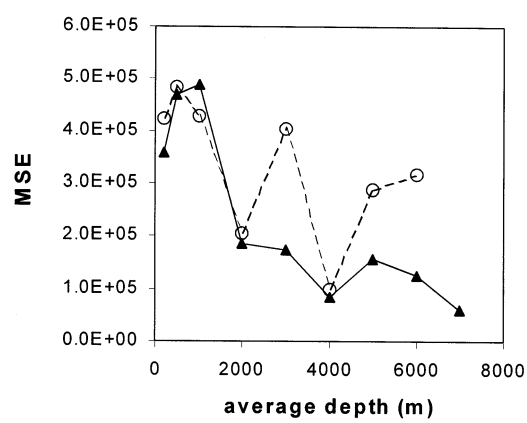

Fig. 4. Effect of co-predictors: MSE vs. average depth. White circle and dashed lines refer to the 7-7-1 NN model in Scardi (2000), whereas black triangles and solid line refer to the 11-14-1 NN model with bathymetric co-predictors.

used for ecological modeling, theoretical knowledge of the biological processes, even though non-quantitative, can be used to constrain the neural network training in order to enhance the 'learning quality'.

In the case of primary production modeling, for instance, it is obvious that, for a given temperature level, only a single maximum can be found in each primary production vs. biomass and irradiance surface. In fact, primary production should not be always maximum when phytoplankton biomass in the water column is maximum, because of selfshading, i.e. of the increased light attenuation in the upper layer of the water column. Moreover, maximum primary production might be attained at an irradiance level lower than maximum because of photoinhibition mechanisms, that protect the photosynthetic apparatus of the phytoplankton cells from excessive irradiation. Therefore, a single primary production maximum is a good quality criterion to test the biological soundness of a primary production surface generated by the model with respect to given ranges of biomass and irradiance values. As a consequence, such a surface cannot have more than four relative minima (that will be located at its corners).

A constrained training can be performed by applying to the error computation a penalty term which depends on the deviations from the above mentioned shape constraints of the primary production response surface. This strategy is somewhat related to weight decay (because of the penalty term) and its overall result is a better generalized learning. Simpler response surfaces are favoured, as well as smaller weights, which limit the variance of the output, as in the case of conventional weight decay (Geman et al., 1992).

An example of the constrained training results with respect to other training strategies is shown in Fig. 5. Each surface describes the variation of primary production as a function of phytoplankton biomass (B, as chlorophyll *a* concentration) and surface irradiance (I_0) at a constant temperature (i.e. 12°C, which is the minimum temperature in the Western Mediterranean data set that was used for this example).

The surface in Fig. 5a was obtained after an error backpropagation training with no generalization tricks, i.e. by iterating the NN training to convergence, without a validation set. It is clear that it is devoid of biological meaning, as primary production seems to be related only to biomass. The only exception is located in a small region of the surface, where midrange to high irradiance values make the primary production estimate dramatically increase in a narrow interval of biomass values.

When usual regularization techniques, such as early stopping and jittering, were used, the resulting primary production surface was much more regular and primary production was positively related to both biomass and irradiance (Fig. 5b).

The constrained approach, however, allowed to define a neural network model that was able to reproduce even more subtle features, providing results that were more sound from a biological viewpoint (Fig. 5c). For instance, primary production slightly decreased when biomass was very large (as in case of selfshading), whereas it was always dependent on irradiance, even though the role of the latter variable became more important for midrange to high biomass levels, which certainly affect the turbidity of the water column.

It is interesting to notice that the constrained training provided better results than the conventional training procedure not only from a theoretical point of view, but also in terms of mean square error. Constrained training mean square error was MSE = 70 759, whereas early stopping returned MSE = 89 480 (the overtrained neural network, which acted as a memory rather than as a model, had MSE = 57 932). The better MSE obtained by constrained training is not only due to the algorithm, but it also depends on the fact that this training strategy does not require a validation set, so that all the available patterns can be used for training. In fact, the validation set is not needed because the NN training can be iterated to convergence and the penalty term associated to the biological constraints prevents the NN from overtraining.

The different properties of the three neural

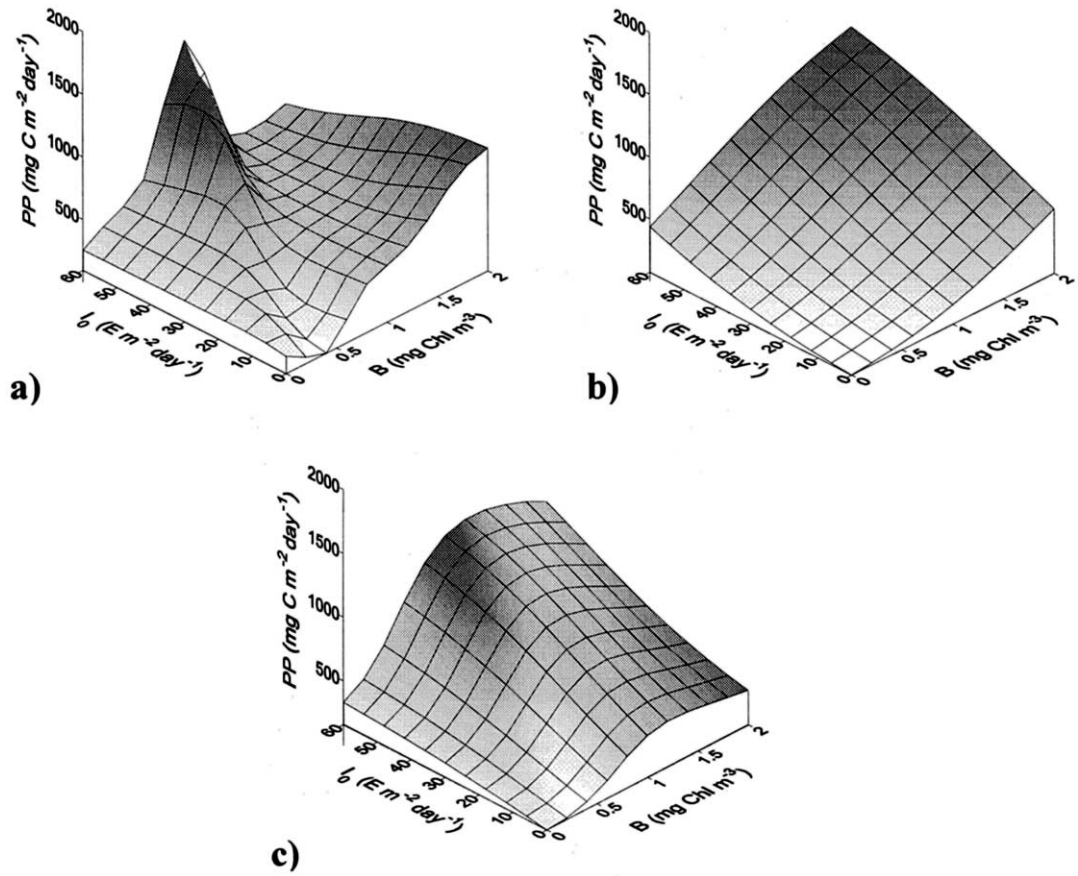

Fig. 5. Phytoplankton primary production (PP) vs. biomass (B) and irradiance (I_0), given a 12°C sea surface temperature. The response surfaces were obtained by conventional and constrained neural network training procedures on a Western Mediterranean data set: (a) overtrained model; (b) generalized model (early stopping and jittering); (c) constrained model.

network models in the above example can be summarized by the average magnitude of their weights: 5.71 in the overtrained case, 0.89 with early stopping and 1.47 with constrained training. It is obvious that lower weights are associated with stiffer models, whereas large weights produce unpredictable results when generalization is needed. Constrained training, on the other hand, allows not to stop the training procedure too early, so that the resulting model is free to find a better compromise between suggestions that come from data structure and constraints that are imposed by the theoretical knowledge of the underlying processes.

3.3. Metamodeling

Empirical models are usually not able to extrapolate, as they rely on the inner structure of the available data sets rather than on the comprehension of the modeled processes. Neural networks are not an exception to this rule. If properly generalized, they can effectively interpolate the available information, but they will almost certainly fail to make accurate predictions out of the range of the training and validation data sets.

However, if other models are already available, then it is possible to use them as an additional

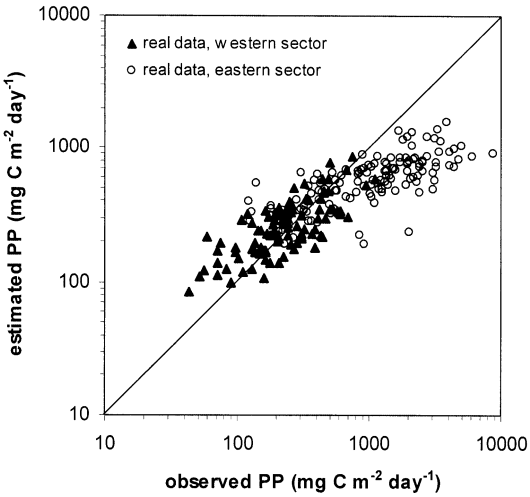

Fig. 6. Estimated vs. observed phytoplankton primary production in the Eastern Equatorial Pacific Ocean. A 9-7-1 neural network model trained on data from the western sector (black triangles) systematically underestimated primary production in the eastern sector (white circles).

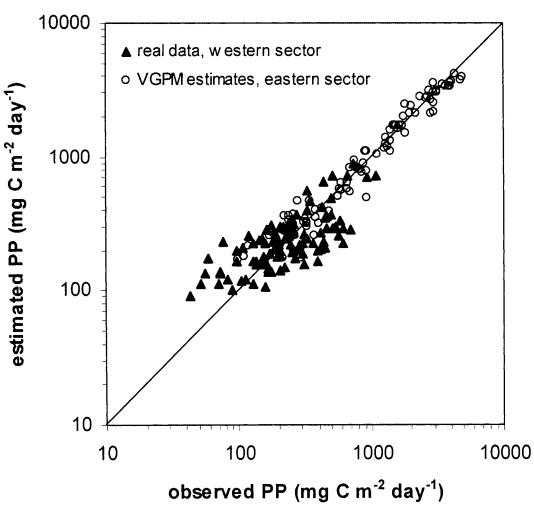

Fig. 7. Estimated vs. observed and VGPM (Behrenfeld and Falkowski, 1997) modeled phytoplankton primary production in the Eastern Equatorial Pacific Ocean. Real data from the western sector and VGPM modeled data from the eastern sector were used to train a new 9-7-1 neural network metamodel: it is evident that the neural network metamodel was more effective in estimating the VGPM modeled primary production data than the real ones.

source of information for training, validating and testing the neural networks. Since such a combination would model not only data, but also the behaviour of another model, it can be regarded as a metamodel.

A simple metamodeling example will demonstrate this procedure and will allow to compare its results with the ones that were provided by a more conventional neural network approach.

A neural network model of phytoplankton primary production was trained and validated on a data set from the western sector in the Eastern Equatorial Pacific Ocean (Fig. 1). Then, data from both the western and the eastern sector were used to check the ability of the neural network model to predict primary production in the eastern sector, i.e. to extrapolate to this sector what it learned from the western sector data. The comparison between primary production estimates and observed values is shown in Fig. 6, where black triangles represent data from the western sector and white circles represent data from the eastern one.

As expected, the neural network model was not able to correctly reproduce the relationships among predictive variables and primary production in the eastern sector, where both phytoplankton biomass and primary production were higher than in the western sector. In particular, the neural network model tended to underestimate primary production, especially when the observed values were larger than the ones that were recorded in the western sector.

In order to help the neural network model to correctly extrapolate, phytoplankton primary production values were computed by means of an existing model (VGPM by Behrenfeld and Falkowski, 1997) on the basis of available predictive data and the resulting patterns were added to the neural network training and validation sets. Thus, these artificial patterns drove the neural network estimates where extrapolation was needed, while the neural network was still able to learn from real data where enough of them were available.

The result of this procedure was that the neural network metamodeled both the real patterns from the western sector and the model-generated patterns from the eastern sector. Of course, the second subset was easier to reproduce, because of the much simpler underlying structure (Fig. 7).

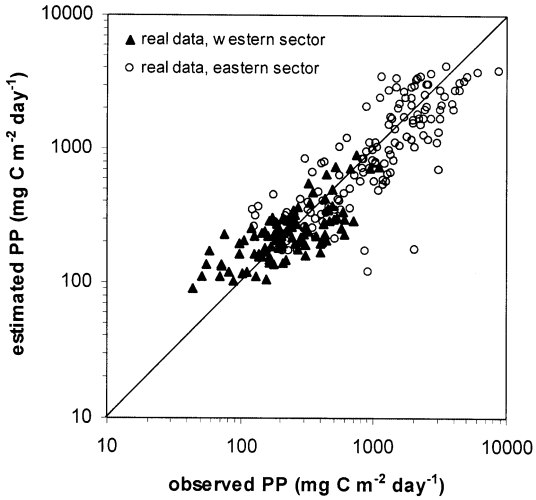

Fig. 8. Estimated vs. observed phytoplankton primary production in the Eastern Equatorial Pacific Ocean. Using a metamodeling approach it was possible to obtain reasonably accurate primary production estimates even for the eastern sector (white circles), where no real primary production data were used for neural network training.

When tested with respect to real data from both sectors, the neural network metamodel provided primary production estimates for the eastern sector that were much improved with respect to the first attempt, i.e. with respect to extrapolated estimates (Fig. 8). The mean square error dropped from MSE = 1 290 622 (neural network trained on western sector data only) to MSE = 488 911 (neural network metamodel trained on western sector data and eastern sector VGPM estimates). It is interesting to notice that if only VGPM had been used, then the MSE would have been much larger, i.e. MSE = 2 071 114. Moreover, the metamodel estimates were almost unbiased, with a mean error much closer to zero than in the case of the basic neural network model ($\bar{E} = -130$ and -508 mg C m^{-2} per day, respectively).

4. Discussion

The training strategies that have been described in this paper provided some hints for enhancing the capabilities of neural networks (namely multilayer perceptrons) as ecological models, but a lot of place is obviously left for experimentation.

Using co-predictors to improve neural network models is not an original development (basically it's a fancy name for an obvious option), but a creative use of additional, co-varying predictive variables can effectively improve existing models that were originally designed according to strictly deterministic blueprints.

As far as primary production is concerned, a lot of variables can play this role. For instance, rainfall, wind stress, cloud cover, distance from coastline and other variables might be considered, at both global and regional scale.

Adding co-predictors is certainly worth trying if suitable information is available, because neural networks are very robust with respect to redundant, even non-relevant inputs, but it is important to bear in mind that the potential benefits provided by this strategy do not come for free. In fact, more inputs, whether they are relevant or not, imply an increase in the number of synaptic weights, which in turn requires more patterns for training the neural network and makes overfitting much likely to happen.

Constrained training provides a more solid approach to optimization of existing models or to development of new ones, as it allows to embed theoretical knowledge into typical empirical models. It has not only theoretical advantages over other generalization strategies, but it also allows to push the training phase to its limits, while keeping synaptic weights under control. In particular, since no validation set is needed, all the available patterns can be used for training and a better learning can be achieved. Finally, constrained training produces smaller synaptic weight than conventional training. This implies, as in weight decay, a small output variance, which is certainly a desirable property.

On the flip side, it is obvious that constrained training works effectively when only a limited number of input variables is linked to neural network outputs by known relationships. As the amount of computations that are needed to check the compliance of the neural network to the constraints is a power function of the number of involved variables, computing time easily becomes a limiting factor. Moreover, since such compliance can only be checked on discrete grids (and

smaller mesh sizes imply longer computations), it is not possible to train an absolutely constrained model, because small scale deviations might elude the minima and maxima search grids. Therefore, constrained training can be more difficult to implement in the case of models with many input and/or output variables.

Ecological models are always hindered by the limited availability of data, and primary production modeling is not an exception to this rule. For instance, most of the available phytoplankton primary production data are concentrated around North America, with a very limited number of observations in other areas (e.g. in the Indian Ocean). Thus, a metamodeling approach provides an easy way to stretch the limits of the available information by exploiting the predictive capabilities of existing models. Thanks to the flexibility of neural networks, metamodeling allows to mimic these 'parent' models when/where no data were recorded (i.e. while extrapolating), but also to improve them when/where data are available (i.e. while interpolating).

In fact, when a neural network is trained using such a mixture of real and modeled data, it acts as a hybrid whose fitness with respect to its 'knowledge environment' is certainly better than the one of its parents.

From a more general point of view, neural network metamodeling can be regarded as a particular case of constrained training, in that theoretical knowledge can be embedded in parent models, from which the neural network metamodel learns how to behave when no real data are available. The balance between the number of real and modeled training patterns, as well as the way they are distributed in training and validation sets, determines the relative influence of data and knowledge over the metamodel.

The bottom line about the role of neural networks in phytoplankton primary production modeling (as well as in other kinds of ecological modeling) is that there is plenty of space for experimentation and for creative use of computational tools.

Ecologists and other practitioners should be aware that neural networks are not just black boxes: they can open the hood, see what is in and try some tricks. Ecological applications of neural networks have not to comply with the need for real time performances or for extreme training accuracy as, for instance, engineering applications. They have to provide ecologically sound results, helping ecologists to bridge the gap between the increasing amount of available information and their limited comprehension of the underlying laws.

Acknowledgements

I wish to thank Vincenzo Saggiomo and Maurizio Ribera d'Alcalà (Stazione Zoologica 'A. Dohrn', Napoli, Italy) for their primary production data from the Gulf of Napoli, the Ocean Primary Productivity Team (Institute of Marine and Coastal Sciences at Rutgers University, New Brunswick, NJ, USA) for their on-line data base at http://marine.rutgers.edu/opp/Database/Database2.html, which significantly contributed to the training and validation sets of my models, and the N.O.A.A. National Oceanographic Data Center (NODC) at Silver Spring (MD, USA) for their F029 and F049 data sets.

References

Behrenfeld, M.J., Falkowski, P.G., 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnol. Oceanogr. 42 (1), 1–20.

Bidigare, R.R., Prezelin, B.B., Smith, R.C., 1992. Bio-optical models and the problems of scaling. In: Falkowski, P.G. (Ed.), Primary Productivity and Biogeochemical Cycles in the Sea. Plenum Press, New York, pp. 175–212.

Cole, B.E., Cloern, J.E., 1984. Significance of biomass and light availability to phytoplantkon productivity in San Francisco Bay. Mar. Ecol. Prog. Ser. 17, 15–24.

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., 1980. Estimating phytoplankton growth rates in the central oligotrophic oceans. In: Falkowski, P.G. (Ed.), Primary Productivity in the Sea. Plenum Press, New York, pp. 231–242.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Comput. 4, 1–58.

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

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

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.

Rumelhart, D.E., Hinton, G.E., Williams, G.E., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.

Scardi, M., 1996. Artificial neural networks as empirical models of phytoplankton production. Mar. Ecol. Prog. Ser. 139, 289–299.

Scardi, M., Harding, L.W. Jr., 1999. Developing an empirical model of phytoplankton primary production: a neural network case study. Ecol. Model. 120, 213–223.

Scardi, M., 2000. Neural network models of phytoplankton primary production. In: Lek, S., Guegan, J.-F. (Eds.), Artificial Neuronal Networks: Application to Ecology and Evolution. Springer-Verlag, Berlin/Heidelberg, pp. 115–129.