Indian Journal of Geo Marine Sciences
Vol. 51 (02), February 2022, pp. 194-203

# Comparison between different modeling techniques for assessing the role of environmental variables in predicting the catches of major pelagic fishes off India's north-west coast

V K Yadav*[,a,b], S Jahageerdar[b] & J Adinarayana[a]

[a]Centre of Studies in Resource Engineering (CSRE), Indian Institute of Technology, Bombay, Maharashtra – 400 076, India

[b]Central Institute of Fisheries Education (CIFE), Panch Marg, Off Yari Road, Versova, Andheri (W), Mumbai,
Maharashtra – 400 061, India

*[E-mail: vinodkumar@cife.edu.in]

*Received 22 April 2019; revised 20 December 2021*

The contribution of four variables, namely Chlorophyll-*a* (Chl-*a*), Sea Surface Temperature (SST), diffuse attenuation coefficient (Kd_490 or Kd), and Photosynthetically Active Radiation (PAR), in predicting the catches of major pelagic fish species (Indian mackerel, horse mackerel, Bombay duck, oil sardine, and other sardines) was evaluated using Canonical Correlation Analysis (CCA). The outcome of the analysis was compared with those obtained by using the following models and methods: the Generalized Linear Model (GLM), the Generalized Additive Model (GAM), connection weight methods, and the explanatory methods of Artificial Neural Networks (ANNs). Both the sets of results were in agreement. Neither the GAM nor the ANNs method showed any clear advantage over each other, although the GAM performed better than the GLM.

[**Keywords**: Artificial neural networks, Canonical correlation analysis, Fish resources, Generalized additive model, Generalised linear model, Sensitivity analysis]

## Introduction

The marine fisheries play a significant role in employment generation, food, nutritional security, and as a source of income. There are about four million people in India who depend for their livelihood on the marine fisheries sector, which provides employment to nearly one million fishermen and contributes significantly to the country's export. In 2017, the estimated marine fish landings for peninsular India were 3.83 million tonnes. Particular interest is placed on the north-west coast (Gujarat & Maharashtra) as it stood first in production in India. Gujarat is the first position, and Maharashtra is at 5th position[1]. Pelagic fishes live predominately in the upper layer of the sea and forms roughly 54 % of the total annual marine fish landings of India[1]. Approximately 40 % of total production is coming from pelagic fishes in both the states, Maharashtra and Gujarat. Understanding how climate/ environment variability may affect the marine population, especially pelagic resources, within the objective of proposing a sustainable fisheries management plan is a challenge. The interaction between environmental factors and the spatiotemporal dynamics of living organisms is an important aspect of ecology. Once we understand the sensitivity of variables affecting the marine system, particularly fish catch, the variables can help in the prediction of major pelagic fishes.

The various ecological models were used for understanding the relationship between environmental variables and the fish species abundance or prediction[2]. Traditionally, for prediction ecology, linear models with environmental variables[3] are used, and normal errors are found in the data using linear regressions, multiple regressions, and multiple discriminant analyses[4] errors that often raise statistical and theoretical concerns[5]. New paradigms in modelling have been developed to address such concerns and include non-linear models like Generalized Linear Models (GLM), and Generalized Additive Models (GAM), which are widely used[6-8].

The environmental pattern that affects the catches of fish are particularly complex and highly non-linear, which is why many researchers prefer Artificial Neural Networks (ANNs), which are robust in dealing with non-linear relationships[3,9-14] to linear statistical models[3,15,16]. However, ANNs fail to explain the relationships between the independent (explanatory) variables and the dependent variables[11]. In the present study, the connection weight method of ANNs is used to obtain such explanations because the connection weight method has been shown to be a better performer than another method in ascertaining the significance of independent variables[17]. More specifically, the GLM, GAM, and ANNs singly or in combination to rank different environmental variables in terms of their importance in predicting the catches of a wide variety of fish species off India's north-west coast were used as has been used by many researchers to study a range of species in different regions[6,9,18,19].

While, the ANNs, GAM, and GLM have been used in many different studies involving many different variables, the three approaches have rarely been compared in the fisheries sector with such variables as chlorophyll-*a*, diffusion attenuation coefficient at 490 nm, photosynthetically active radiation, and sea surface temperature. The significance of these variables in fisheries is discussed in the latter part of the study.

The present study sought to rank, using GLM, GAM, and connection weight methods of ANNs, the above four variables concerning their contribution to predicting the catches of major fish species in two states in India, Gujarat, and Maharashtra, along India's north-west coast. Given the multiple species of fish and the four environmental variables, the Canonical Correlation Analysis (CCA) was also carried out to arrive at the rankings and compare the results as obtained by these two approaches, namely the methods mentioned above and the CCA, and also estimated the models for their accuracy in determining the comparative importance of the four variables in forecasting the catches.

## Materials and Methods

### Data on variables and catches of fish

Mean values of Chl-*a*, SST, PAR and Kd for the study area were obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) level 3 binned images recorded by the Ocean Biology Processing Group (OBPG). The data were in the form of ASCII files and covered the period of 1997 to 2013. Because the data on the catches of fish had been aggregated quarterly, the values of the above four variables were also taken as average values for each quarter. Data on the catches of fish were taken from the National Marine Fisheries Data Centre (NMFDC) of the Central Marine Fisheries Research Institute (CMFRI), Kochi, Kerala. The four quarters for each year were January to March, April to June, July to September, and October to December.

### Study variables and their importance

#### *Concentration of chlorophyll-a (Chl-a)*

Because Chl-*a* is the key pigment for photosynthesis by the phytoplankton and marine algae, which is used as food for the fish and thus determines the assemblages of fish in a given area or the potential fishing zone, was taken as one of the inputs into the forecasting models. Chl-*a* is measured as mg/m³.

#### *Sea Surface Temperature (SST)*

Movement, feeding, and reproduction of fish is affected by SST (°C), and therefore it has been taken as one of the inputs in the model.

#### *Photosynthetically Active Radiation (PAR)*

Photosynthetically active radiation is the amount of light available for photosynthesis. It is defined as the quantum energy flux from sunlight in the wavelength band of 400 – 700 nm. As some fish species (for example, mackerels) are herbivores[18], PAR, that is, the number of photons received by a unit area over a specified amount of time, or the photosynthetic photon flux density (PPFD, expressed per square meter per day) is considered as one of the significant biophysical parameters.

#### *Diffuse attenuation coefficient (Kd)*

The diffuse attenuation coefficient is the measurement of the transparency of the bottom water, and it is important because some fish species (for example, tuna) need light to locate their prey and thus affect the availability of food. It is measured as $m^{-1}$.

The above four independent environmental variables were used as inputs into the models to predict the fish catch more consistently. For reference, the summary of some of the environmental variables used by different authors for different purposes in fish prediction is displayed in Table S1.

### Methodology

#### *Statistical methods and Artificial Neural Networks (ANNs)*

In the present study, univariate analysis of the statistical parameters consisted of determining the minimum, maximum, median, quartile, and mean values, the standard deviation (SD), and the coefficient of variation. In the multivariate analysis, the associations between the independent variables and the dependent variable, were examined namely the estimated quarterly catches of fish, using the GLM, GAM, and the ANNs technique. Because the estimates of the catches were highly variable, those data were subjected to logarithmic transformation before analysis.

### *Generalized Linear Model (GLM) and Generalized Additive Model (GAM)*

Generalized linear models (GLM) are generalizations of linear regression models and admit the non-linearity and non-constant variance in the data[20]. The data presume to fall in any of the probability distributions of normal, Poisson, binomial, negative binomial, and gamma[21]. Because ecological relationships are inadequately represented by the Gaussian distributions, GLMs are better suited and more flexible for analysing ecological relationships[22].

Generalized additive models[23,20] are semi-parametric extensions of GLMs, and the basic assumption is that the functions are additive and that the components are smooth. The strength of GAMs is their capability to deal with non-linear and non-monotonic relationships between the response and the explanatory variables[8]. More details about GLMs and GAMs are given by Guisan *et al.*[7]. The package "glm2" and "mgcv" are used for GLM and GAM analysis, respectively.

The data with different error distributions and spline functions (Cubic regression splines, Duchon splines, and thin plate regression splines) was used for GAM model building and found that cubic regression splines was the best over others.

GAM model for any fish species (for example, Bombay duck) catch prediction was estimated in the following way:

log(Bombay_duck) ~ s(Chl-a, bs = "cr") + s(SST, bs = "cr") + s(PAR, bs = "cr") + s(Kd, bs = "cr")

Where, cr = Cubic regression splines, and bs = B-splines

*Or*

log (Bombay_duck) = c) + $f_1$(Chl-a) + $f_1$(SST) + $f_3$(PAR) + $f_4$(kd) + ε

Where, fi are smoothers, c - a constant, and ε - a random error term

The software package R 3.6.1 was used for the analysis.

### *Artificial Neural Networks (ANNs)*

In the present study, the multi-layer feed-forward neural network architecture and back propagation for training the network[24] was used. The weights were adjusted using the back propagation error, that is, the difference between the observed result and the estimated result[11]. The supervised learning procedure was used for minimizing the difference between the desired outputs and the predicted outputs. The neural network that we used consisted of three layers (Fig. 1). The first layer, the input layer, is connected to the input variables and comprises four neurons (four input variables). The third layer, the output layer, connected to the output variable, and comprised only one neuron (the output variable). The second layer, or the intermediate layer, was referred to as the hidden layer and lay between the first and the third layer. At hidden and output layer of ANN, sigmoid and linear activation function respectively were used.

The selection of neurons in the hidden layer is the main criterion of ANN. The approach to selecting a network consists of testing many distinct probable designs and choosing the one, offers the minimum bias and variance and the training that gives a better generalization[26]. In the present study, the network with one hidden layer of 3 or 4 neurons was preserved. The connection weight method was used for analysing (using MATLAB R2012a) the contribution of each of the four variables to the already calibrated ANN model.

The association between the inputs ($y_{t-1}$, $y_{t-2}, \ldots, y_{t-p}$) and the output ($y_t$) are represented as follows:

$$y_t = f\left(\sum_{j=0}^{q} \omega_j g\left(\sum_{i=0}^{p} \omega_{ij} y_{t-i}\right)\right) \qquad \text{...(1)}$$

Where, $\omega_j (j = 0,1,2, \ldots\ldots, q)$ and $\omega_{ij} (i = 0,1,2, \ldots\ldots, p, j = 0,1,2, \ldots\ldots q)$, are called as the connection weights, $p$ and $q$ are the number of input and hidden nodes, respectively, $g$ and $f$ denote the activation function at the hidden and output layer, respectively. The commonly used activation function at the hidden layer is a logistic (sigmoid) function.

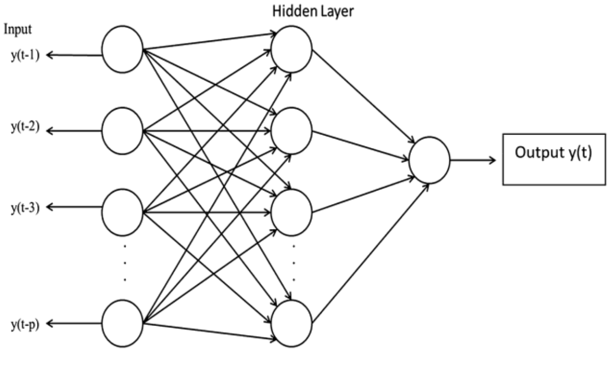


Fig. 1 — Neural network architecture (Yadav et al.25)

***Connection weights algorithm: ranking of variables using an ANN***

The method of connection weights calculates the product of weights between input-hidden and hidden-output and the connection through each input and output neuron and sums the products across all hidden neurons[27].

The relative importance of a given input variable can be defined as follows:

$$RI_x = \sum_{y=1}^{m} w_{xy} w_{yz} \quad \text{(ref. 28)}$$

Where, $RI_x$ is the relative importance of input neuron "*x*", $\sum_{y=1}^{m} w_{xy} w_{yz}$ is the sum of the product of final weights obtained by training the network of the connection from input neuron to hidden neurons with the connection from hidden neurons to output neuron, '*y*' is the total number of hidden neurons, and '*z*' is the number of output neurons.

***Canonical Correlation Analysis (CCA)***

The objective of canonical correlation is to correlate simultaneously several metric-dependent variables and several metric-independent variables[29]. Individual set can have various variables, and CCA calculates a linear from individual set, called a canonical variable, with the objective to maximize the correlation between two canonical variables[30].

CCA is an extension of multiple regression analysis with more than one set of dependent variables and helps in finding the complex interactions between two sets of variables and in estimating the extent to which variance in one set is common to, or predictable from, the variance in the other set[31].

Hotelling, in 1935, was the first to introduce the CCA[32], a powerful multivariate technique that has since found application in diverse fields including psychology, the social sciences, political science, ecology, education, sociology-communication, and marketing[33]. In fisheries, however, only a few studies have used the CCA[34,35].

The correlation between the canonical variate obtained from the estimated catches and each of the four environmental variables can be called canonical correlation. Squared canonical correlation (canonical roots or eigenvalues) represents the amount of variance in one canonical variate accounted for by the other canonical variate[29]; however, a detailed account of the CCA is beyond the scope of this paper.

## Results and Discussion

**Generalized linear model and generalized additive model**

A summary of fitting the two models relating the fish catch (in tonnes) to the figures predicted from the environmental factors is given in Tables 1 & 2.

The two tables show that the GAM performed better than the GLM for both the states, as evident from the higher adjusted R2 value. Damalas et al.6, in their study of the catch of swordfish, expressed as CPUE, or catch per unit effort, calculated the variance at 36.1 % and 46.7 % using the GLM and the GAM model, respectively; the input variables were the month, year, gear type, latitude, longitude, lunar index, bathymetry, and SST. Usman *et al.*[36], in estimating the CPUE with reference to the Skipjack tuna, attributed 26 % of the variance in CPUE to Chl-*a*.

**Artificial neural networks**

***Predictive capacity***

The mean percentages of recognition (the training and validation set) and prediction (the test set) change quickly with the number of neurons in the hidden layer. Seeing the values of MSE (Mean Square Error) and *R* (correlation coefficient) of the training, validation and testing data from 66 data points, different numbers of hidden neurons were selected for different species (resources) wherever the variation in MSE and *R* among all the three sets (training,

Table 1 — Comparison of GLM and GAM model for Gujarat region

| Species (resources) | Adjusted R2 in model | |
|---|---|---|
| | GLM (Gaussian distribution with identity link function) | GAM (Gaussian distribution with identity link function) |
| Indian mackerel | 0.272 | 0.305 |
| Horse mackerel | 0.146 | 0.212 |
| Bombay duck | 0.707 | 0.713 |
| Other sardine | 0.354 | 0.370 |

Table 2 — Comparison of GLM and GAM model for Maharashtra region

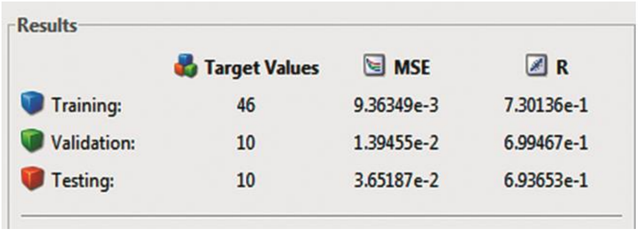
| Species (resources) | Adjusted R2 in model | |
|---|---|---|
| | GLM (Gaussian distribution with identity link function) | GAM (Gaussian distribution with identity link function) |
| Indian mackerel | 0.486 | 0.544 |
| Horse mackerel | 0.200 | 0.240 |
| Bombay duck | 0.474 | 0.580 |
| Oil sardine | 0.662 | 0.671 |

validation, and testing) was minimal to handle the overfitting and poor generalization (Figs. 2, 3, and 4 for each of the three major species off the Gujarat coast). The optimal weight between input to hidden and hidden to the output layer of the neural network taking 3 hidden neurons selected are listed in Table S2 (a to c) (for illustration, Indian mackerel, Horse mackerel and Bombay duck of Gujarat region were selected).

The results of the correlation coefficient were as good in the learning set as in the testing set. For the next step, the ANN and the full database for sensitivity analysis of the variables (input parameters) for different species were used.

***Comparison between the generalized linear model and generalized additive model***

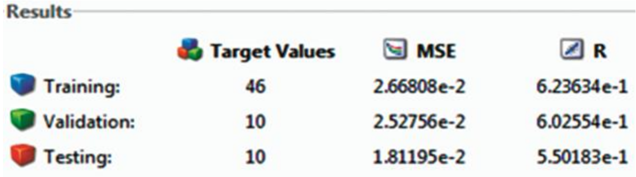
Judged on the basis of the adjusted R² (Tables 3 and 4), the ANN performed better than the GAM in predicting the catches of Indian mackerel and horse mackerel and worse than the GAM in predicting the catches of Bombay duck, oil sardine, and other sardines from the coasts of both Maharashtra and Gujarat. Thus, neither of the two models emerged as clearly superior to the other, probably because the data set used for training the network was small37.



Results

| | Target Values | MSE | R |
|---|---|---|---|
| Training: | 46 | 9.36349e-3 | 7.30136e-1 |
| Validation: | 10 | 1.39455e-2 | 6.99467e-1 |
| Testing: | 10 | 3.65187e-2 | 6.93653e-1 |

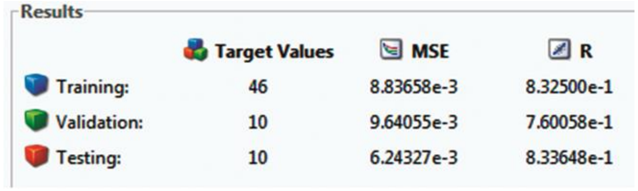
Fig. 2 — MSE and R on 66 data points for Indian mackerel



Results

| | Target Values | MSE | R |
|---|---|---|---|
| Training: | 46 | 2.66808e-2 | 6.23634e-1 |
| Validation: | 10 | 2.52756e-2 | 6.02554e-1 |
| Testing: | 10 | 1.81195e-2 | 5.50183e-1 |

Fig. 3 — MSE and R on 66 data points for Horse mackerel



Results

| | Target Values | MSE | R |
|---|---|---|---|
| Training: | 46 | 8.83658e-3 | 8.32500e-1 |
| Validation: | 10 | 9.64055e-3 | 7.60058e-1 |
| Testing: | 10 | 6.24327e-3 | 8.33648e-1 |

Fig. 4 — MSE and R on 66 data points for Bombay duck

***Connection weight***

The results after optimal training of the ANN are shown in Figures 2, 3, and 4. The different weights obtained for different species were used for ranking the variables for their predictive value by deploying the connection weight algorithm of the ANN (detailed methodology can be seen in Yadav *et al*.[38]). The relative importance of each variable in predicting the catch is given in parenthesis as a percentage in Table 5 (Gujarat coast) and Table 6 (Maharashtra coast).

**Canonical Correlation Analysis (CCA)**

***Gujarat***

The estimated canonical correlation values between pairs of canonical variates were 0.74, 0.47, 0.31, and 0.26, respectively (Table S3). The correlation between the first pair of the canonical variates was significant ($P < 0.01$) as judged by the likelihood ratio test (Table S4). The first test of significance, tests all the four canonical roots of significance. The remaining canonical correlations were not statistically significant ($P > 0.05$). The value of significance from the likelihood ratio test was also equal to the value of significance of Wilks' lambda. However, the value of the redundancy measure (squared correlation) of 0.55 for the first canonical variate suggests that about 55 % of the variance in Y variables was accounted for by X variables, whereas the corresponding figure for the second canonical variate was only about 22 %.

As can be seen from Tables S5 and S6, the first canonical function of the dependent and independent

Table 3 — Comparison between ANN and GAM model for Gujarat region

| Species (resources) | Adjusted R2 in model | |
|---|---|---|
| | ANN model | GAM (Gaussian distribution with identity link function) |
| Indian mackerel | 0.480 | 0.305 |
| Horse mackerel | 0.330 | 0.212 |
| Bombay duck | 0.640 | 0.713 |
| Other sardine | 0.354 | 0.370 |

Table 4 — Comparison between ANN and GAM model for Maharashtra region

| Species (resources) | Adjusted R2 in model | |
|---|---|---|
| | ANN model | GAM (Gaussian distribution with identity function) |
| Indian mackerel | 0.55 | 0.544 |
| Horse mackerel | 0.27 | 0.24 |
| Bombay duck | 0.55 | 0.58 |
| Oil sardine | 0.67 | 0.671 |

Table 5 — The relative importance of each input variable (in %) in predicting fish species catch for Gujarat coast landing data

| Species input → ↓ | Indian mackerel | Horse mackerel | Bombay duck | Other sardine |
|---|---|---|---|---|
| SST | 4 (4 %) | 3 (10.09 %) | 1 (45.10 %) | 1 (35.82 %) |
| Chl-a | 1 (44.36 %) | 1 (57.52 %) | 4 (7.10 %) | 2 (27.26 %) |
| Kd | 3 (11.26 %) | 2 (29.28 %) | 3 (16.74 %) | 3 (22.83 %) |
| PAR | 2 (42.39 %) | 4 (3.09 %) | 2 (25.54 %) | 4 (14.07 %) |

Table 6 — The relative importance of each input variable (in %) in predicting fish species catch for Maharashtra coast landing data

| Species input → ↓ | Indian mackerel | Horse mackerel | Bombay duck | Oil sardine |
|---|---|---|---|---|
| SST | 4 (3.6 %) | 4 (3.63 %) | 1 (62.36 %) | 2 (25.68 %) |
| Chl-a | 1 (43.6 %) | 1 (33 %) | 4 (5.09 %) | 1 (43.94 %) |
| Kd | 3 (25.97 %) | 2 (32.93 %) | 3 (11.79 %) | 3 (21.60 %) |
| PAR | 2 (26.8 %) | 3 (30.37 %) | 2 (20.79 %) | 4 (8.76 %) |

variables represents 27.52 % of the variance in the dependent variable and 27.38 % of the variance in the independent variable. Similarly, the second, third, and fourth canonical functions of the dependent variables represent 9.19, 18.12, and 23.81 % of the variance, respectively, in the dependent variables and 20.97, 28.38, and 23.26 % of the variance, respectively, in the independent variables.

The correlation between the variables and the related canonical variates (canonical loading) are shown in Table S7. When the first canonical functions of the dependent and the independent variables were taken, Bombay duck was found to be highly and positively loaded on the canonical function of the dependent variable (V1), whereas SST and PAR were highly and negatively loaded on the canonical function of the independent variable (U1). Madhavan *et al.*[18] also reported a positive correlation between SST and PAR. The negative relationship between the catches of Bombay duck and SST shows that the catches decrease as the temperature increases. Because the fourth canonical function of the dependent variable represents greater percentage variation in the dependent variable than that represented by the second and the third function (Table S5). The fourth canonical component (V4) can be considered, and when it was considered, the catches of Indian mackerel, horse mackerel, and other sardines were found to be loaded highly (Table S7) and the catches of these species were influenced by Chl-*a* and Kd, because the catches were highly loaded on the fourth canonical function (U4) of the two independent variables, followed by SST and PAR (Table S7).

The third and the fourth canonical functions of the independent variables represent 28.38 and 23.26 % of the variation, respectively in the independent variable (Table S6) because the third canonical function of the independent variable represents a higher percentage variation in an independent variable. The third canonical component (U3) can be considered whenever the contributions of both Kd and Chl-*a* were higher than those of SST and PAR (Table S7).

To distinguish between the variability in the catches of horse mackerel from those of Indian mackerel and other sardines as influenced by the environmental variables, the third canonical function was considered. When the third canonical functions of the dependent and independent variables were taken, catches of horse mackerel were found to be positively loaded (although less so when compared to those from the fourth canonical dependent variate) on the canonical functions of the dependent variables. Also, Chl-*a* and Kd were highly negatively loaded, with the loading of Kd being more than that of Chl-*a*, when the third canonical function of an independent variable was taken. At the same time, the catches of Indian mackerel and other sardines were loaded negatively. Hence, from the third and fourth canonical variate it can be inferred that both Chl-*a* and Kd have a positive influence on the catches of Indian mackerel and other sardines and a negative influence on the catches of horse mackerel. The reason for this difference is obvious: Indian mackerel and other sardines are planktivores and higher concentrations of Chl-*a* encourage the species to assemble in larger numbers. Given the high positive correlation between Chl-*a* and Kd (Table S8), Chl-*a* can be considered an important variable in predicting the catches of those species.

Also, the negative influence (or weak correlation) of those two variables (either Kd or Chl-*a*) and the catches of Horse Mackerel (Table S8) although, indirectly agreed by CCA, the reason is obvious: the

Horse Mackerel is a carnivore and needs clearer waters to spot its prey, and that clarity or transparency is a function of Kd (the lower the Kd, the greater the transparency).

#### *Maharashtra*

The estimated canonical correlation values between pairs of canonical variates were 0.58, 0.33, 0.19, and 0.07 (Table S9), and the canonical correlations between the first pair of canonical variates were found to be significant ($P < 0.01$) (Table S10).

As can be seen from Tables S11 and S12, the first canonical function of the dependent variables and the independent variables represents 26.90 and 24.00 % of the variance, respectively. The second, third, and fourth canonical functions of the dependent variable represent 21.14, 16.44, and 19.09 % of the variance in the dependent variables, and 28.91, 26.14, and 20.94 % variance in the independent variables, respectively.

Correlations between the variables (or resources) and the related canonical variates (canonical loading) are shown in Table S13. When the first canonical function of the dependent variable (V1) and that of the independent variable (U1) were taken, Bombay duck was highly and positively loaded on the dependent canonical function. At the same time, if considered the first canonical function of an independent variable, SST and PAR were highly and negatively loaded. The negative relationship between the catches of Bombay duck and SST shows that the catch decreases as the temperature increases.

When the third canonical function of the dependent variable (V3) was taken, Indian mackerel was highly loaded on the dependent canonical function. At the same time, when the third canonical function of an independent variable (U3) was taken, Chl-*a* and Kd were loaded highly.

When the fourth canonical function of the dependent and the independent variables (V4 and U4, respectively) was taken, horse mackerel was loaded positively on the dependent canonical function and negatively on the independent canonical functions (Chl-*a* and Kd), with the loading of Kd being higher than that of Chl-*a*. The catches of oil sardines were highly loaded on the second canonical variate of the dependent variable (V2) and SST and on the independent variable Chl-*a* (U2). Thus the catches of oil sardines were influenced by both SST and Chl-*a*.

### The relative importance of variables in predicting catches

The rankings arrived at by all the methods of the four variables in terms of predicting the catches of fish along the coasts of Gujarat (Table 7), and Maharashtra (Table 8) were broadly similar for both the states at least for the two most important variables.

The oil sardine (*Sardinella longiceps*, Clupeidae) and Indian mackerel (*Rastrelliger kanagurta*, Scombridae) are the major commercial species on India's west coast[39]. The oil sardine, primarily an herbivore that feeds on phytoplankton, has replaced the lesser sardine or other sardines, which are mid-level carnivores[40]. The populations of oil sardine can explode and dominate other fish species in terms of

Table 7 — Comparative ranking of the relative importance of variables in predicting the fish catch for Gujarat coastal area.

| Species (resources) | Significant variables in different models | | | |
|---|---|---|---|---|
| | GLM | GAM | Connection wt of ANN (Variables importance in decreasing order) | CCA |
| Indian mackerel | Chl-a | Chl-a | Chl-a >PAR >Kd >SST | Chl-a, Kd |
| Horse mackerel | Chl-a | Chl-a, Kd | Chl-a >Kd >SST >PAR | Chl-a, Kd |
| Bombay duck | SST, PAR, Chl-a, Kd | SST, PAR | SST >PAR >Kd >Chl-a | SST, PAR |
| Other sardine | SST | SST, Chl-a | SST >Chl-a >Kd >PAR | Chl-a, Kd, SST |

Table 8 — Comparative ranking of the relative importance of variables in predicting the fish catch for Maharashtra coastal area

| Species (resources) | Significant variables in different models | | | |
|---|---|---|---|---|
| | GLM | GAM | Connection wt of ANN (Variables importance in decreasing order) | CCA |
| Indian mackerel | Chl-a, Kd, PAR | Chl-a, Kd, SST | Chl-a >PAR >Kd >SST | Chl-a, Kd |
| Horse mackerel | Chl-a, Kd | Chl-a, Kd | Chl-a >Kd >PAR >SST | Kd, Chl-a |
| Bombay duck | PAR | PAR, SST, Chl-a, Kd | SST >PAR >Kd >Chl-a | SST, PAR |
| Oil sardine | PAR, SST, Chl-a, Kd | PAR, SST Chl-a, Kd | Chl-a >SST >Kd >PAR | SST, Chl-a |

abundance under such favorable conditions as an abundance of phytoplankton[35]. The models used in the study indicated that Chl-*a* and SST are the two most important factors for predicting the catches of the oil sardine, whereas SST is a major predicting factor for other sardines (except that CCA identified Chl-*a* is the most important variable). Also, Chl-*a* can predict whether the Indian mackerel is distributed densely and evenly. The scatter plot of other sardines and SST (Fig. S1) clearly shows that other sardines are more abundant when the temperatures range from 27.5 to 29.5 °C. No such pattern was seen in the case of Chl-*a* (Fig. S1). In the case of the oil sardine for Maharashtra, both Chl-*a* and SST are important for predicting the catches (Fig. S2).

During the south-west monsoon (June to September), due to coastal upwelling along with Somalia, the concentration of Chl-*a* increases (Table S14).

Manjusha *et al*.[35] also reported a similar pattern. Usually, spawning in such pelagic fishes as in oil sardine, Indian mackerel, and horse mackerel peaks during the south-west monsoon[41] as a result, the catches are higher during the post-monsoon period (October to January). The higher average concentrations of Chl-*a* during the south-west monsoon (Table S14) also increase the catches during the same period (Figs. S3 – S6). Manjusha *et al*.[35] also reported greater catches of oil sardine and Indian mackerel from October to January along the south-west coast because of higher concentrations of Chl-*a* during the post-monsoon period. The main predictor of the catches of the above- mentioned species is, therefore, is Chl-*a*, given that the species are herbivores and feed on phytoplankton.

The models showed that SST and PAR are the main predictors of the catches of Bombay duck, which is particularly sensitive to high temperatures[42]. The catches of Bombay duck in Gujarat and Maharashtra were higher in the fourth quarter, *i.e*. from October to December (Figs. S7 and S8), during which the average values of SST and PAR were low (Table S15), and the two variables are known to have positive correlation[18].

The model also indicated Kd and Chl-*a* to be the main predictors of the catches of the horse mackerel, a carnivore. As mentioned earlier, the species requires clear water for sighting its prey, and the availability of the prey depends on the availability of phytoplankton which, in turn, is governed by Chl-*a*. The scatter plots for the catches of horse mackerel plotted against the values of Kd and Chl-*a* for Gujarat and Maharashtra are shown in Figures S9 and S10, respectively, and, as can be seen in Figures S11 and S12, Chl-*a* and Kd are highly correlated.

## Conclusion and Summary

Four independent variables, namely Chl-*a*, SST, PAR and Kd_490(or Kd), were ranked for their ability to predict the catches of five major pelagic fish species (Indian mackerel, horse mackerel, Bombay duck, oil sardine, and other sardines) off the coasts of Gujarat and Maharashtra. Two models, namely the GLM and GAM, and connection weight methods of ANN were used to arrive at the rankings. For both the states, Gujarat and Maharashtra, the GAM performed better than the GLM and showed higher adjusted R². Compared to the GAM, ANN performed better in predicting the catches of Indian mackerel and horse mackerel and worse in predicting those of Bombay duck, oil sardine, and other sardines. Thus, neither of the two models was superior to the other, probably because the data set used for training the ANN was small.

The rankings arrived at by all the methods of the four variables in terms of predicting the catches of fish along the coasts of Gujarat and Maharashtra were broadly similar, at least for the two most important variables.

For forecasting the catches of Indian mackerel (*Rastrelliger kanagurta*), Chl-*a* was a particularly important factor, and both Chl-*a* and SST were important in the case of the oil sardine; both SST and PAR were significant for Bombay duck, whereas, for other sardines, SST alone was important. The models showed Kd and Chl-*a* to be the main predictors for horse mackerel. Canonical correlation analysis was also performed between the sets of resources (the fish species) and the four environmental variables to understand their joint impacts on the catches. The results from the two models and from the ANN agreed with the analysis.

One limitation of the study was that the data set was small: a larger data set will help in firmer rankings of the variables and in determining which of the three, the two models and the ANN is best suited for such a study.

## Supplementary Data

Supplementary data associated with this article is available in the electronic form at http://nopr.niscpr.res.in/jinfo/ijms/IJMS_51(02)194-203_SupplData.pdf

## Acknowledgments

The authors are grateful to the Director, ICAR-CIFE, Mumbai, and Director, IIT Bombay, for providing the facilities to carry the work.

## Conflict of Interest

The authors would like to declare that there are no conflicts of interest to publish this research papers in the journal.

## Author Contributions

The authors like to certify that, the first author (VK) had contributed towards the preparation of the paper such as conceptualization, data collection, data analysis, and drafting of the manuscript; the second author (SJ) contributed in guidance and editing of the contents; and the third author (JA) contributed in overall supervision and guidance on the manuscript revision.

## References

1 CMFRI, *Central Marine Fisheries Research Institute (CMFRI) annual report*, 2017- 2018, pp. 9.

2 Razaei R & Sengul, Development of Generalized Additive Models (GAMs) for *Salmorizeensis* Endemic to North-Eastern Streams of Turkey, *Turkish J Fish Aquat Sci*, 19 (4) (2018) 29-39.

3 Manel S Dias J M & Ormerod S J, Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird, *Ecol Model*, 120 (1999) 337-347.

4 Yadav V K, Jahageerdar S, Ramasubramanian V, Bharti V S & Adinarayana J, Use of different approaches to model catch per unit effort (CPUE) abundance of fish, *Indian J Geo-Mar Sci*, 45 (12) (2016) 1677-1687.

5 Austin M P & Meyers J A, Current approaches to modeling the environmental niche of eucalyptus: implications for management of forest biodiversity, *Forest Ecol Manag*, 85 (1996) 95–106.

6 Damalas D, Megalofonou P & Apostolopoulou M, Environmental, spatial, temporal and operational effects on swordfish (*Xiphias gladius*) catch rates of eastern Mediterranean Sea longline fisheries, *Fish Res*, 84 (2007) 233–246.

7 Guisan A, Edwards T C, Thomas C & Hastie T, Generalized linear and generalized additive models in studies of species distributions: setting the scene, *Ecol Model*, 157 (2002) 89-100.

8 Maunder M N & Punt A, Standardizing catch and effort data: a review of recent approaches, *Fish Res,* 70 (2004) 141–159.

9 Georgakarakos S, Koutsoubas D & Valavanis V, Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters, *Fish Res*, 78 (2006) 55–71.

10 Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, *et al.*, Application of neural networks to modelling nonlinear relationships in ecology, *Ecol Model*, 90 (1996) 39-52.

11 Muriel G, Dimopoulos I & Lek S, Review and comparison of methods to study the contribution of variables in artificial neural network models, *Ecol Model*, 160 (2003) 249-264.

12 Yadav V K, Krishnan M, Biradar R S, Kumar N R & Bharti V S, A comparative study of neural-network & fuzzy time series forecasting techniques — Case study: Marine fish production forecasting, *Indian J Geo-Mar Sci*, 42 (6) (2013) 707-716.

13 Bharti V S, Inamdar A B, Purusothaman C S & Yadav V K, Soft Computing and Statistical Technique - Application to Eutrophication Potential Modelling of Mumbai Coastal Area, *Indian J Geo-Mar Sci*, 47 (2) (2018) 365-377.

14 Yadav V K, Jahageerdar S & Adinarayana J, A comparison of different Fuzzy Inference Systems for prediction of Catch per Unit Effort (CPUE) of Fish, *Indian J Geo-Mar Sci*, 48 (01) (2019) 60-69.

15 Paruelo J M & Tomasel F, Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models, *Ecol Model*, 98 (1997) 173-186.

16 Ramos-Nino M E, Ramirez-Rodriguez C A, Clifford M N & Adams M R, A comparison of quantitative structure-activity relationships for the effect of benzoic and cinnamic acids on Listeria monocytogenes using multiple linear regression, artificial neural network and fuzzy systems, *J Appl Micro*, 82 (1997) 168-176.

17 Olden J D, Joy M K & Death R G, An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data, *Ecol Model*, 178 (2004) 389-397.

18 Madhavan N, Thirumalai V D, Ajith J K & Sravani K, Prediction of Mackerel Landings Using MODIS Chlorophyll-a, Pathfinder SST, and SeaWiFS PAR, *Indian J Nat Sci*, 5 (29) (2015) 4858-4871.

19 Stergiou K I, Christou E D & Petrakis G, Modeling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods, *Fish Res*, 29 (1) (1997) 55-95.

20 Hastie T J & Tibshirani R J, *Generalized Additive Models*, 1st edn, (Routledge), 1990, pp. 352. https://doi.org/10.1201/9780203753781

21 Nelder J A, Wedderburn R W M, Generalized linear models, *J R Statist Soc A*, 137 (1972) 370–384.

22 Austin M P, Models for the analysis of species' response to environmental gradients, *Vegetatio*, 69 (1987) 35-45.

23 Hastie T J & Tibshirani R J, Generalized additive models, *Stat Sci*, 1 (1986) 97-318.

24 Rumelhart D E, Hinton G E & Williams R J, Learning representations by backpropagation error, *Nature*, 323 (1986) 533-536.

25 Yadav V K, Jahageerdar S & Adinarayana J, Forecasting quarterly landings of total fish and major pelagic fishes and modelling the impacts of climate change on Bombay duck along India's north-western Gujarat coast, *Indian J Geo-Mar Sci*, 50 (07) (2021) 557-565.

26 Geman S, Bienenstock E & Doursat R, Neural networks and the bias/valance dilemma, *Neural Comput*, 4 (1992) 1-58.

27 Olden J D & Jackson D A, Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks, *Ecol Model*, 154 (2002) 135–150.

28 Ibrahim O M, A comparison of methods for assessing the relative importance of input variables in artificial neural networks, *Journal Appl Sci Res*, 9 (11) (2013) 5692-5700.

29 Hair J F, Anderson R E, Tatham R H & William C B, *Multivariate Data Analysis*, 5th Edn, (Prentice Hall, Inc), 1998, pp. 442–462.

30 Lebart L, Morineau A & Warwick K M, *Multivariate descriptive statistical analysis*, (Wiley, New York), 1984.

31 Weiss D J, Canonical correlation analysis in counselling psychology research, *J Couns Psychol*, 19 (1972) 241–252.

32 Wood D A & Erskine J A, Strategies in canonical correlation with application to behavioral data, *Educ Psychol Meas*, 36 (1976) 861–878.

33 Jaiswal U C, Poonia J S & Kumar J, Canonical correlation analysis for studying the relationship among several traits: An example of calculation and interpretation, *Indian J Anim Sci*, 65 (1995) 765–769.

34 Goes J I, Thoppil P G & Gomes H D R, Warming of the Eurasian landmass is making the Arabian Sea more productive, *Science*, 308 (2005) 545-547.

35 Manjusha U, Jayasankar J, Remya R, Ambrose T V & Vivekanandan E, Influence of coastal upwelling on the fishery of small pelagic off Kerala, south-west coast of India, *Indian J Fish*, 60 (2) (2013) 37-42.

36 Usman T, Ersti Y S & Syaifuddin A, The relationship between concentration of chlorophyll-a with skipjack (*Katsuwonus pelamis*, Linnaeus 1758) production at West Sumatera waters, Indonesia, IOP Conf Series, *Environ Earth Sci*, 54 (2017) p. 012072.

37 Raudys S & Jain A K, Small sample size problems in designing artificial neural networks, *Machine Intelli Patter Recog*, 11 (1991) 33-50.

38 Yadav V K, Jahageerdar S & Adinarayana J, Use of Different modeling approach for Sensitivity analysis in predicting Catch per Unit Effort (CPUE) of fish, *Indian J Geo-Mar Sci*, 49 (11) (2020) 1729-1741.

39 Vivekanandan E, Mohamed K S, Kuriakose S, Sathianandan T V, Ganga U, *et al*., Status of marine fish stock assessment in India and development of sustainability index, *Second Workshop on Assessment of Fishery Stock Status in South and South-east Asia*, Bangkok, WPO2h, 2009, pp. 15.

40 Vivekanandan E, Hussain A, Jasper B & Rajagopalan M, The vulnerability of corals to warming of the Indian seas: a projection for the 21st century, *Curr Sci*, 97 (2009) 1654-1658.

41 Nair R V, Synopsis of the biology and fishery of the Indian sardine, *Proceedings of FAO World Science Meetings on the biology of sardine and related species*, 2 (1960) 329- 414.

42 Rao K, Bombay duck: iconic fish fast disappearing from city's coastal waters (22 March 2013). Retrieved from https://www.theguardian.com/environment/blog/2013/mar/22/bombay-duck-mumbai-fish