



Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) abundance forecast off northern Chile: A multivariate ecosystemic neural network approach

Eleuterio Yáñez^{a,*}, Francisco Plaza^a, Juan Carlos Gutiérrez-Estrada^b, Nivaldo Rodríguez^a, M.A. Barbieri^a, Inmaculada Pulido-Calvo^b, Cinthya Bórquez^a

^a Pontificia Universidad Católica de Valparaíso, Casilla 1020, Valparaíso, Chile

^b Universidad de Huelva, 21819 Palos de la Frontera, Huelva, España, Spain

ARTICLE INFO

Article history:

Available online 25 September 2010

ABSTRACT

An evaluation of the performance of artificial neural networks (ANNs) to forecast monthly anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) catches in northern Chile (18°21'S–24°S) is presented, using environmental variables, anchovy and sardine CPUE, fishing effort and catches between 1963 and 2007. An analysis of previous data was carried out, consisting of a non-linear cross-correlation analysis to estimate the time lags for each input variable. A multi-layer perceptron architecture model was used, calibrated with the Levenberg–Marquardt algorithm, thus obtaining anchovy landings and sardine CPUE forecast models. An ecosystemic approach was conducted for both models, considering local and global environmental variables, the anthropogenic effect, and the interaction between species as inputs. The variance explained by both models was slightly higher than 82% and the standard error of prediction was lower than 45%. The strong correlation between the estimated and observed series on the anchovy and sardine models suggests that ANN models capture the trend of the historical data. Furthermore, the generalization capacity along with the sensitivity analysis allowed the identification of high-weight variables in the model, as well as the partial interpretation of the statistical functional relationships between the input variables and the abundance.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The Humboldt Current System (HCS) is very productive due to nutrient transport by large-scale horizontal advection and persistent coastal upwelling (Carr, 2002). The average annual landings off Chile have totaled 4.8 million tons over the last three decades. Forty percent of such landings involve pelagic resources from the northern zone (18°21'S–24°S). Northern fisheries have been successively based on anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*), and both species have been affected by fishing effort and environmental fluctuations (Yáñez et al., 2008).

This anchovy-sardine succession has been documented for 250 years (Valdés et al., 2008), and it has been associated with environmental changes. In fact, anchovy fisheries, which significantly decrease during El Niño events, collapsed after the 1972–1973 event, a situation that promoted its replacement by the sardine fisheries. However, this fishery also suffered a major decline after 1985, when a strong recovery of the anchovy fisheries

was also occurred (Yáñez et al., 2001). In 1975, an increase of sardines after a warm period was observed. After 1985, however, rather cold conditions were re-established thus promoting the anchovy recovery and the sardine decline, despite the El Niño events in 1987, 1991–1992, and 1997–1998 (Yáñez et al., 2003).

The connection between the variations in anchovy and sardine abundance and environmental changes on different spatial-temporal scales allow short-medium term predictions of landing fluctuations, which is one of the main objectives of fishery management. Marine ecosystems are constantly in a state of imbalance and they are characterized by high levels of stochasticity and non-linear relationships between variables (Murdoch, 1994; Stenseth et al., 2002). Although both species vary according to environmental changes, predators, competitors, and parasites also affect their composition and abundance.

The use of artificial neural networks (ANNs) in ecosystem modeling began in the early 1990s, particularly for situations where data do not fit statistical suppositions and non-linear relationships prevail. ANNs behave better than linear models and have a good capacity for generalizing when entering new data (Lek et al., 1996; Lek and Guégan, 1999; Özesmi et al., 2006). Non-linear modeling of pelagic fisheries off northern Chile was first done by Gutiérrez-Estrada et al. (2007), using an univariate ANN model that

* Corresponding author. Address: Pontificia Universidad Católica de Valparaíso, Facultad de Recursos Naturales, Escuela de Ciencias del Mar, Altamirano 1480, Valparaíso, Chile.

E-mail address: eyanez@ucv.cl (E. Yáñez).

only considered anchovy catches lagged up to 6 months, and Gutiérrez-Estrada et al., (2009), where an ecosystemic approach to forecast sardine landings was considered.

In this study, the performance of ANN models to forecast monthly anchovy and sardine abundance is evaluated. This multi-variate ecosystemic approach considers local and global environmental variables, anchovy and sardine catches, fishing effort and catch per effort unit during from 1963 to 2007.

2. Materials and methods

2.1. Information analysis

Total monthly anchovy and sardine landings (CA and CS, respectively) off northern Chile (18°21'S–24°S) were taken from Statistical Fishery Yearbooks¹. The fishing effort (*f*) was calculated by the C/CPUE relationship, where CPUE is the monthly average of the standardized catch per unit effort for both fisheries (Espíndola et al., 2005).

Monthly averages (1950–2007) of the following variables were analyzed, which were recorded at the Antofagasta meteorological and oceanographic stations (23°26'S): sea surface temperature of Antofagasta (SSTA), average sea level (SL), air temperature (AT), wind direction and magnitude for estimating the Ekman transport (ET) (Bakun et al., 1974) and the turbulence index (TI) (Elsberry and Garwood, 1978). Monthly averages were also considered for the Southern Oscillation Index (SOI; Allan et al., 1991); the Cold Tongue Index (CTI), an SST anomaly from 6N–6S and 180W–90W (Deser and Wallace, 1990); and the SST in El Niño region 1 + 2² (SSTN12; 0°S–10°S and 90°W–80°W) and El Niño region 3 + 4² (SSTN34; 5°N–5°S and 170°W–120°W).

According to Yáñez et al. (2008), anchovy landings are suggested to be proxy for the abundance of the species in northern Chile, as landings show clear interdecadal, interannual and intra-seasonal fluctuations that may be directly related to the resource abundance. Moreover, these authors reject CPUE as an abundance index, since it differs from the direct biomass estimations. The same authors also suggest sardine CPUE as an abundance proxy for sardine, as it shows a good correlation with biomass estimations and landings. In this work, the abundance of both resources is analyzed, i.e. landings and CPUE are considered as abundance indexes for anchovy and sardine, respectively.

2.2. Data treatment

Previous data analyses are crucial for the calibration of a neural network; consequently, the variables representing the nature of the problem to be modeled must be selected carefully. The inclusion of “noise” in the calibration of the network weight makes learning more difficult and, therefore, diminishes the predictive capacity of the model (Özesmi et al., 2006; Lek and Guégan, 1999). Thus, a correlation analysis among variables was first performed in order to determine their association to discriminate among highly correlated variables. Then, a principal component analysis was carried out to establish the most representative variables of the environmental variability.

A cross-correlation analysis was done between the input variables and the predicted variable, using non-linear functions. This method considers a series of non-linear functions to estimate the lag correlation in time (Table 1). The estimation of the cross-correlation was made using the specially designed program CORN 1.0

Table 1

Models used for the determination of linear and non-linear cross-correlation. The constant term is designated by β_0 , β_1 and β_2 are parameters of the model and *Y* and *X* is the dependent and independent variables respectively.

Model	Equation
(1) Linear	$Y = \beta_0 + \beta_1 X$
(2) Reciprocal linear	$Y = 1/(\beta_0 + \beta_1 X)$
(3) Rectangular hyperbola I	$Y = (\beta_0 + \beta_1 X)/(1 + \beta_1 X)$
(4) Reciprocal rectangular hyperbola I	$Y = (1 + \beta_1 X)/(\beta_0 + \beta_1 X)$
(5) Rectangular hyperbola II	$Y = X/(\beta_0 X + \beta_1)$
(6) Reciprocal rectangular hyperbola II	$Y = (\beta_0 X + \beta_1)/X$
(7) Parabola	$Y = \beta_0 + \beta_1 X + \beta_2 X^2$
(8) Power	$Y = \beta_0 X^{\beta_1}$
(9) Modified power	$Y = \beta_0 \beta_1^X$
(10) Root	$Y = \beta_0^{(1/\beta_1)}$
(11) Geometric	$Y = \beta_0 X^{\beta_1 X}$
(12) Modified geometric	$Y = \beta_0 X^{\beta_1 / X}$
(13) Exponential	$Y = \beta_0 \exp(\beta_1 X)$
(14) Modified exponential	$Y = \beta_0 \exp(\beta_1 / X)$
(15) Logarithm	$Y = \beta_0 + \beta_1 \log(X)$
(16) Reciprocal logarithm	$Y = 1/[\beta_0 + \beta_1 \log(X)]$
(17) Hoerl	$Y = \beta_0 \beta_1^X X^{\beta_2}$
(18) Modified Hoerl	$Y = \beta_0 \beta_1^{(1/X)} X^{\beta_2}$
(19) Normal	$Y = \beta_0 \exp[-(X - \beta_1)^2 / (2\beta_2^2)]$
(20) Normal logarithm	$Y = \beta_0 \exp[-(\log(X) - \beta_1)^2 / (2\beta_2^2)]$
(21) Cauchy	$Y = \beta_0 / [1 + ((X - X_0) / \beta_1)^2]$
(22) Beta	$Y = \beta_0 X^{\beta_1} a(1 - X)^{\beta_2}$

(Gutiérrez-Estrada et al., 2009). The highest significant absolute correlation value (*R*) for every 12 months lag was chosen as the lags for each variable. A maximum delay of 24 months (2 years) for anchovy, and 72 months (6 years) for sardine was considered, based on the biology of each species, i.e. the anchovy catches are concentrated in the early year classes, while the catches of sardine, a more long-lived species, are concentrated in the classes 5 and 6 (Yáñez and Barbieri, 1989; Yáñez et al., 2001; Alheit and Niquen, 2004).

Freón et al. (2003) recommend smoothing the data series in order to reduce the high frequency noise to achieve a clear observation of the tendency. In this case, a moving average of three centered months was used for such effect.

2.3. Artificial neural networks (ANNs)

ANNs are mathematical models inspired by the human brain. They are able to recognize behavioral patterns and learn from their interactions with the environment. The most common models used in ecology are those known as feed forward multi-layer perceptrons (Suryanarayana et al., 2008; Lek et al., 1996). A typical three-layer feed forward ANN has *g*, *n*, and *s* nodes or neurons in the input, hidden and output layers, respectively. In this paper, the ANN architectures will be defined as *g*:*n*:*s*. The parameters associated with each of the connections between nodes are called weights. An iterative process, known as neural network training or learning, defines the interconnections between the neurons, and when carried out using known inputs and outputs (training data or patterns) varied iterations are done until the desired exit values are obtained. The value of the weights is adjusted using an error convergence method, such that the desired output value is generated by an input value. There are several training algorithms and, herein, a variation of the back propagation algorithm (Rumelhart et al., 1986), known as the Levenberg–Marquardt algorithm (Hagan and Menhaj, 1994), is used, which is a second order, non-linear optimization algorithm with a very rapid convergence and is recommended by various authors (Tan and Van Cauwenberghe, 1999; Anttil and Rat, 2005).

All iterations (epochs) are performed during the training period for all patterns used in the training. Overtraining is probably the

¹ SERNAPESCA. 1963–2007. Anuarios Estadísticos de Pesca. Servicio Nacional de Pesca, Ministerio de Economía, Fomento y Reconstrucción, Chile.

² <http://www.cpc.ncep.noaa.gov/data/indices>.

most common problem when training neural networks; the network may be adjusted to noise when calculating the weights. The best method for ensuring that overtraining does not occur is periodical monitoring of the sum squared error for the training data and another group called the selected data (this process is called internal validation). The selected data are a subset of the training data (usually 15–20%). When the sum squared error of the selected data begins to increase, the training must be stopped and the weights of this epoch must be used in the external validation with the test set (also called generalization phase or external validation), which represents data the network has never encountered before.

2.4. General procedure

The monthly data for anchovy and sardine were divided into three groups (randomly selected): 60% of the data were used for training (or calibration of the neural network parameters), 20% for internal validation, and 20% for external validation (EV) of the network.

The selection of the best models was based on the fitness of the EV data, because the fitness of the training data may not perform well in cases based on the unseen data (EV group). This procedure is also used by Makkearsorn et al. (2008) and Gutiérrez-Estrada et al. (2009).

An inherent problem of ANNs is their tendency to get stuck in local minima. In order to solve this problem, the same architecture of neural networks was trained more than once, beginning with a series of random weights (Anctil and Rat, 2005). Thus, the neural network with the lowest error index in the external validation phase was chosen. In this work, all architectures were calibrated 30 consecutive times. This level of repetitions implies that the chosen model was distributed within 14% of the best possible models, with 99% confidence (Iyer and Rhinehart, 1999). This procedure allowed the determination of the best ANN architecture.

Once the external validation for each model was performed, a sensitivity analysis was done considering each input variable to the neural network as if it was not available in the model. Hunter et al. (2000) defined a procedure for substituting the missing values. First, the network is run with a set of test cases and the resulting error is saved. Later, the same network is used but the observed values are replaced with the value estimated by the missing-value substitution procedure and the error of the network is saved again. Given the effective removal of information used by the network (one of the input variables), there will be a decline in the error level. The basic sensitivity measure is the quotient (ratio) between the network error without the input variable and the original error. If the ratio is less than or equal to one, then the variable removal or adding will have no effect, or will eventually improve, the error level of the network.

2.5. Error indexes

Several accuracy measures were applied to assess the model performances during the validation phases, as no unique and more suitable performance evaluation test has been developed (Yapo et al., 1996; Legates and McCabe, 1999; Abrahart and See, 2000), such as the coefficient of determination (R^2), the percentage of standard error of prediction (%SEP) (Ventura et al., 1995), the coefficient of efficiency (E) (Nash and Sutcliffe, 1970; Kitanidis and Bras, 1980), and the average relative variance (ARV) (Griño, 1992). These estimators are not influenced by the variation range of their elements and they are used to see to what extent the model is able to explain the total data variation. They are also appropriate to quantify the error in terms of the units of the variable to be estimated. These measures of absolute errors include the square

root of the mean squared error (RMSE). In order for the goodness of fit to be acceptable, the values of R^2 and E should approach one and the values of %SEP must be close to 0.

The persistence index (PI) was used to evaluate the performance of the model (Kitanidis and Bras, 1980). The PI index is well-designed for the evaluation of the models applied in this work, since the landings in previous months represent the input variables of the neural networks. A PI = 1 indicates a perfect fit between the estimated and observed values. On the other hand, a PI = 0 is equivalent to saying that the model provides the previously observed values as a prediction, which is known as a naïve model. A negative PI value means the model is altering the original information, performing worse than a naïve model (Anctil and Rat, 2005).

Other indexes used for the identification of the best models are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Qi and Zhang, 2001). These indexes measure the goodness of fit of the model and a sanction for the excess of parameters (weights). Thus, the model with the lowest AIC and BIC values is selected. Weights are not strictly considered as parameters. Nevertheless, as in Qi and Zhang (2001) and Chen and Hare (2006), they have been considered as such when calculating AIC and BIC in order to compare this work with models obtained by other authors.

3. Results

3.1. Data treatment

The correlation analysis among variables showed high correlations between SSTA and AT (0.96), ET and TI (0.92) and SSTN34 and CTI (0.80) (Table 2). Although it is considered that two variables are highly correlated when the value of the Pearson coefficient is above or equal to 0.8 (Real et al., 2008), this paper also considers those variables where the level of correlation was marginally close to this value (SSTA-SSTN12, SSTN12-AT); these results can be also seen on the PCA (Fig. 1). The highest correlation for the PCA (and the highest negative correlation) with PC1 was SSTA and SOI (negative); and for PC2 the variables were ET and SSTN34 (negative). The aforementioned variables were selected to continue with the non-linear cross-correlation analyses.

Based on previous information analyses, two local and two basin environmental variables were chosen to be included into the ANN models: SSTA, ET, SOI and SSTN34, respectively. For the anchovy ANN model, only SSTA was considered as an environmental variable; the other three variables do not present any significant lag. The maximum non-linear cross-correlation between the landing and SSTA was –7 months (Fig. 2). Although this correlation value is slightly higher than that obtained between the lagged variables and the linear function, no difference was detected in the lag between the linear and non-linear functions. The anchovy landings (CA) from previous periods were also included to improve the prediction based on the work carried out by Gutiérrez-Estrada et al. (2007), who used an univariate model considering anchovy catches, with a lag of up to 6 months; this lag was found using an autocorrelation analysis. Sardine landings (CS) from previous periods were also included.

Two local (SSTA and ET) and one basin (SOI) environmental variables were selected for the sardine ANN model (the lags for SSTN34 were insignificant). The fishing effort (fS) and previous anchovy landings (CA) were also included as input variables. A pattern was observed for the maximum cross-correlation lags for the local environmental variables: there was a peak every 12 months approximately from the first maximum lag. On the other hand, the maximum lag for SOI was very clear, at

Table 2
Correlation matrix of environmental variables, highly correlated variables in bold.

	TI	ET	SSTA	AT	SL	PDO	SSTN12	SSTN34	SOI
ET	0.92**								
SSTA	0.38**	0.40**							
AT	0.37**	0.40**	0.96**						
SL	0.00	0.02	0.38**	0.37**					
PDO	0.11*	0.09*	0.12*	0.08	0.17**				
SSTN12	-0.01	-0.01	0.69**	0.69**	0.46**	0.21**			
SSTN34	-0.18**	-0.18**	0.00	0.00	0.35**	0.37**	0.42**		
SOI	-0.06	-0.07	-0.15**	-0.17**	-0.31**	-0.31**	-0.25**	-0.65**	
CTI	0.16**	0.14**	0.14**	0.14**	0.35**	0.34**	0.30**	0.80**	-0.65**

Values are Pearson correlation coefficients.

* $P < 0.05$.

** $P < 0.01$.

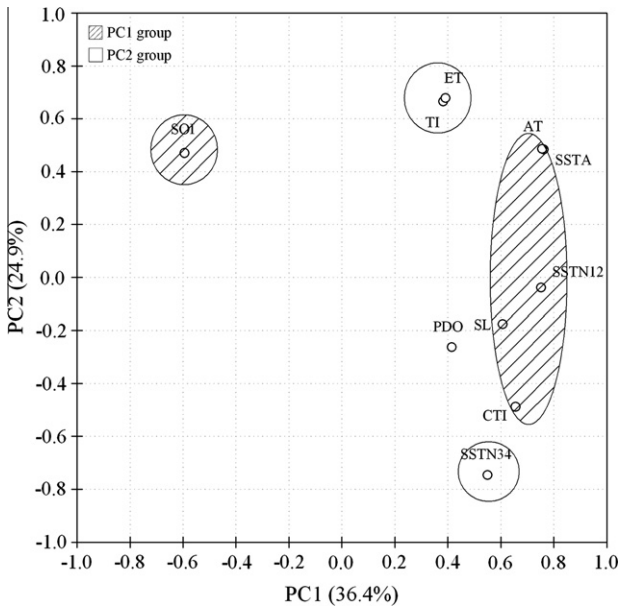


Fig. 1. Grouped environmental variables after a principal components analysis (PCA).

–39 months. The maximum cross-correlation index for fS was at lag –1 and the cross-correlation lag for CA was not as clear as the other variables (Fig. 3). Therefore, the selected lags were based on the maximum lag per year (every 12 months lag).

3.2. Abundance prediction models

The same procedure in terms of data percentage for training (60%), internal validation (20%), and external validation (20%) was used for both anchovy and sardine models. The final input configuration for both models is shown in Table 3.

A total of 420 ANNs were calibrated and validated: 180 for anchovy and 240 for sardine, each composed of six hidden and eight hidden configurations with 30 repetitions each, respectively.

An error analysis in the external validation phase of the anchovy and sardine ANN modeling was conducted. Thus, the best model was selected (from a pool of 30 repetitions) from each hidden configuration, ending up with a selection of six and seven ANN models for anchovy and sardine, respectively. The anchovy and sardine best ANN model selection (with 18 and 22 neurons in the input layer, respectively, and one in the output layer) exceeded 73% and 91% of the explained variance and showed a standard error of prediction below 44% and 31%, respectively (Table 4).

The best error measures for anchovy were associated with the 18:20:1 architecture (with 20 neurons in the hidden layer), except for the AIC and BIC indexes. This model explained 85% of the variance, which suggested an efficiency of 0.84. The persistence index is 0.83, which implies a very low lag between the observed and estimated series. The standard error of the prediction (34.15%) and RMSE (19022.07 tons) values showed the scatter between the observed and estimated series is relatively low for the low and high landing values (Fig. 4).

The best error measures for the ANN sardine model corresponded to the 22:20:1 architecture (with 20 neurons in the hidden layer), which showed an explained variance of 95%, while

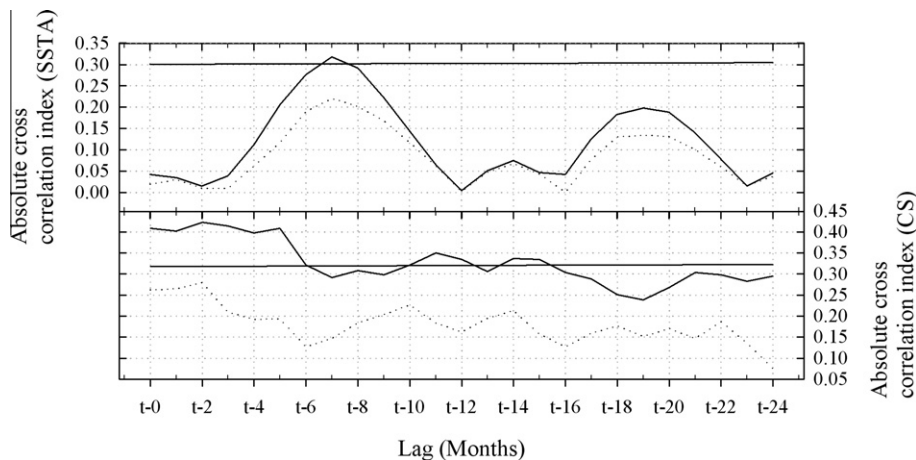


Fig. 2. Linear (dashed) and non-linear (solid) cross-correlation for anchovy catches and sea surface temperature of Antofagasta (SSTA), sea surface temperature on the Niño 34 region (SSTN34), and sardine catches (CS). Significance level also included ($P = 0.05$).

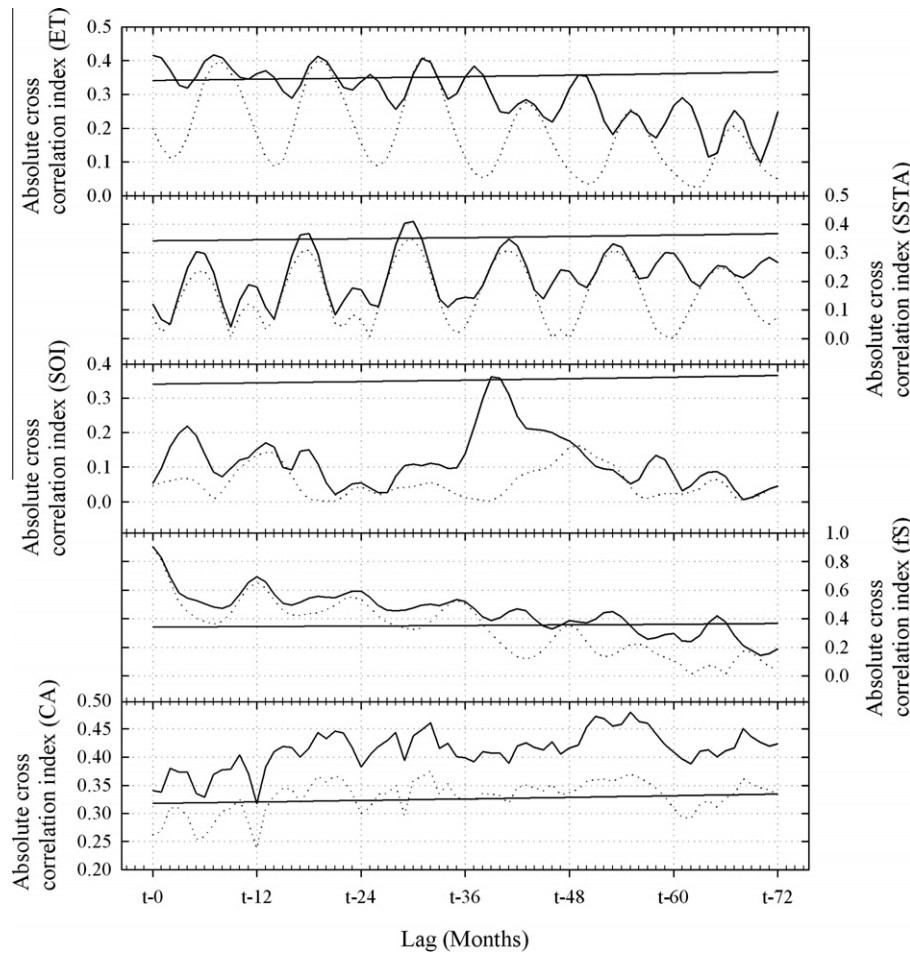


Fig. 3. Linear (dashed) and non-linear (solid) cross-correlation for sardine CPUE and Ekman transport (ET), sea surface temperature of Antofagasta (SSTA), Southern Oscillation Index (SOI), sardine fishing effort (fS) and anchovy catches (CA). Significance level also included ($P = 0.05$).

Table 3
Input configurations for each anchovy and sardine abundance forecast models.

Input variable	Lags selected
<i>Anchovy model (CA)</i>	
SSTA	-7
CA	-1, -2, -3, -4, -5, -6
CS	-1, -2, -3, -4, -5, -6, -10, -11, -12, -14, -15
<i>Sardine model (CPUES)</i>	
SSTA	-18, -30
TE	-1, -7, -19, -31, -37, -49
SOI	-39
fS	-1, -12, -23, -35, -42, -53, -65
CA	-10, -21, -32, -46, -55, -68

the AIC and BIC indexes were lower than the anchovy ANN model (12.32 and 12.14 respectively). Furthermore, the persistence index was 0.73, which showed a low lag although it was not as low as the lag presented on the anchovy ANN model (Fig. 5).

The sensitivity analysis (Table 5) identified three variables which showed a greater-than-average ratio (CA_{t-1} , CA_{t-4} , CA_{t-3} and CA_{t-2}) for the anchovy ANN models. Furthermore, none of the variables presented a ratio below 1; therefore, none could be eliminated. The most important variables (greater-than-average ratio) for the sardine ANN model sensitivity analysis were fishing effort (fS_{t-1} , fS_{t-42} , fS_{t-53} , fS_{t-12} , fS_{t-23}), Ekman transport (ET_{t-1} , ET_{t-19} , ET_{t-31}) and anchovy landings (CA_{t-68} , CA_{t-55} , CA_{t-21}). Like the anchovy models, none of the variables could be eliminated.

4. Discussion

This paper was developed based on the work of Gutiérrez-Estrada et al. (2007), where univariate ANN models dependent only on anchovy catches from previous periods were used. Moreover, Gutiérrez-Estrada et al. (2009) used a multivariate ecosystemic approach to forecast sardine landings off northern Chile. In this paper, a non-linear multivariate approach predicting monthly abundance of anchovy and sardine off northern Chile was made using an integrated ecosystemic modeling with environmental and fisheries data and the relation between species. As suggested by Legates and McCabe (1999), a multicriteria performance assessment based on different accuracy measures was appropriate to select the best models. The comparison of the results of this work revealed that the multivariate methodology was characterized by a higher accuracy in terms of all evaluation measures. It is necessary to highlight that the method is still experimental and highly dependent on expert criteria.

4.1. Anchovy abundance models

In relation to the analysis of non-linear cross-correlations for anchovy, the lags encountered for SSTA could be explained in terms of the environment-resource interaction. The most significant lag was at -7 months, the sea surface temperature off Antofagasta has a 7-month in-advance effect on landings. According to Braun et al. (1995) and Castillo et al. (2002) this is a biolog-

Table 4

Anchovy and sardine artificial neural networks (ANN) models accuracy measures. Presented are the values of the best model for each hidden neuronal architecture following the external validation (EV).

Accuracy measures	Number of weights	R^2	RMSE	SEP (%)	E	PI	AIC	BIC
<i>Anchovy ANN architecture</i>								
18:5:1	95	0.81	22775.23	39.46	0.80	0.77	10.54	10.56
18:10:1	190	0.73	28690.70	43.77	0.72	0.72	12.57	12.60
18:15:1	285	0.80	20095.75	33.02	0.79	0.77	14.09	14.13
18:20:1	380	0.85	19022.07	34.15	0.84	0.83	15.87	15.93
18:25:1	475	0.81	23302.16	40.05	0.81	0.78	17.87	17.95
18:30:1	570	0.84	22616.94	39.38	0.82	0.74	19.67	19.76
<i>Sardine ANN architecture</i>								
22:5:1	115	0.93	17.01	24.22	0.93	0.35	4.96	4.92
22:10:1	230	0.92	16.73	30.22	0.92	0.46	7.45	7.36
22:15:1	345	0.92	18.79	29.94	0.92	0.36	10.05	9.91
22:20:1	460	0.95	14.48	21.99	0.95	0.73	12.32	12.14
22:25:1	575	0.91	20.53	29.84	0.91	0.21	15.12	14.90
22:30:1	690	0.95	15.17	24.57	0.94	0.54	17.36	17.09
22:35:1	805	0.91	18.90	28.85	0.91	0.36	21.90	21.55

The bold values indicate the best validated models.

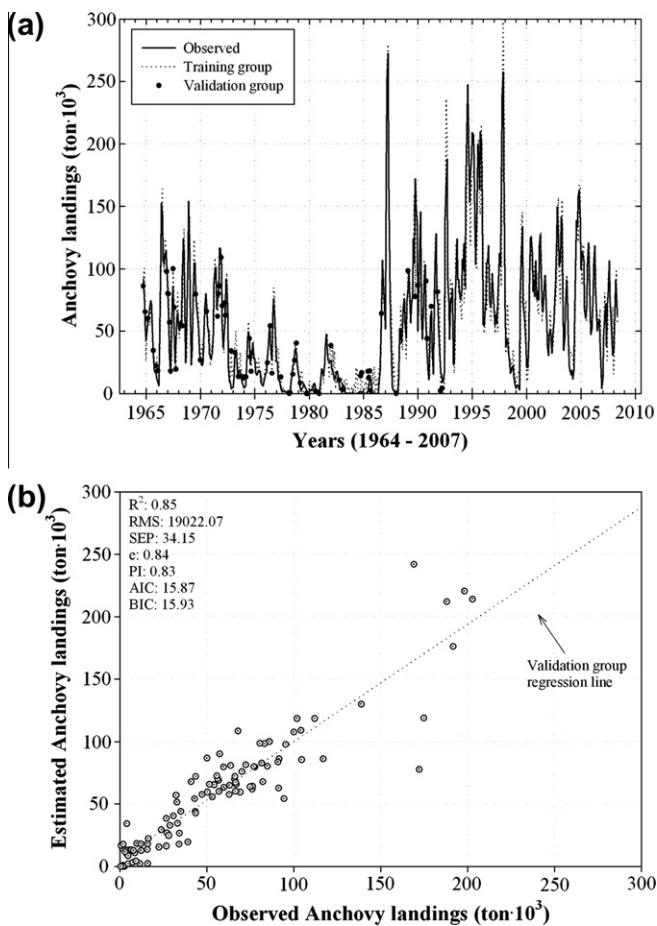


Fig. 4. Forecast of the best anchovy artificial neural network (ANN) model, and monthly observed and estimated catches for external validation (EV) data. Monthly observed and predicted landings for calibration and validation sets (a), scatter plot between observed and predicted anchovy landings (b).

ical relationship, as they indicate that anchovy recruitment occurs 5–6 months after spawning. This is added to the fact that variations in the fecundity of clupeiform species are probably affected by temperature 2–3 months before spawning (Winters et al., 1993). Hence, when adding the lags suggested by the previous authors, we obtain a range similar to that of this work through the non-linear crossed correlation analysis. The SST measured in the coastal stations in northern Chile is closely re-

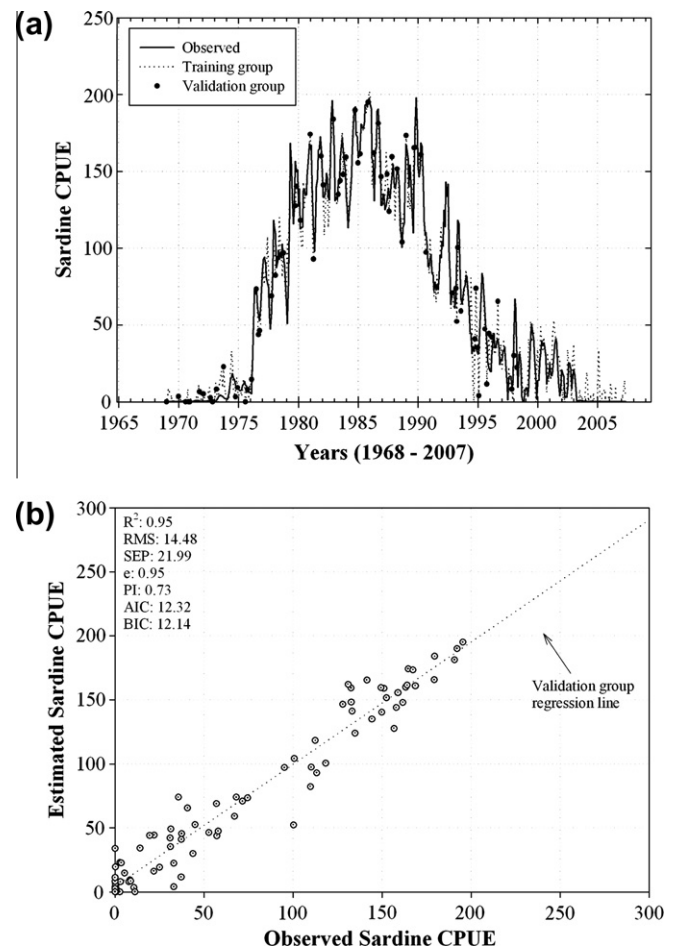


Fig. 5. Forecast of the best sardine artificial neural network (ANN) model, and monthly observed and estimated CPUE for external validation (EV) data. Monthly observed and predicted CPUE for calibration and validation sets (a), scatter plot between observed and predicted sardine CPUE (b).

lated to the temperature recorded by satellite images (Barbieri et al., 1995), and large-scale physical processes (Hormazabal et al., 2001; Montecinos et al., 2003). SST is also considered as a proxy of ecosystem variability (Yáñez et al., 2008), which would indicate changes in primary productivity, food, fertility, egg and larvae survival due to upwelling turbulence, and consequently in recruitment and landings.

Table 5

Sensitivity analysis for the anchovy and sardine artificial neural networks (ANN) models. Average of the ratios between the network error without the input variable and the original error (see text) for the best models. Ranking and mean ratio are shown. Bold indicates higher-than-mean ratio.

Number	Anchovy		Sardine	
	Variable	Ratio	Variable	Ratio
1	CA_{t-1}	3.22	fS_{t-1}	1.53
2	CA_{t-4}	1.91	ET_{t-1}	1.51
3	CA_{t-3}	1.46	fS_{t-42}	1.37
4	CA_{t-2}	1.45	fS_{t-53}	1.34
5	CS _{t-11}	1.27	fS_{t-12}	1.34
6	CS _{t-2}	1.24	ET_{t-19}	1.33
7	CS _{t-4}	1.23	CA_{t-68}	1.33
8	CA _{t-5}	1.22	CA_{t-55}	1.33
9	CS _{t-10}	1.20	fS_{t-23}	1.33
10	CS _{t-1}	1.19	ET_{t-31}	1.33
11	CS _{t-12}	1.18	CA_{t-21}	1.31
12	CS _{t-3}	1.16	ET _{t-37}	1.29
13	CS _{t-14}	1.14	SSTA _{t-30}	1.28
14	CS _{t-5}	1.13	fS _{t-35}	1.27
15	CS _{t-15}	1.13	fS _{t-65}	1.27
16	CS _{t-6}	1.12	ET _{t-49}	1.26
17	CA _{t-6}	1.08	CA _{t-32}	1.24
18	SSTA _{t-7}	1.03	ET _{t-7}	1.23
19			SSTA _{t-18}	1.22
20			CA _{t-46}	1.22
21			CA _{t-10}	1.14
22			SOI _{t-39}	1.12
	Mean ratio	1.35		1.30

The cross-correlation analysis of SSTN34 showed no significant lags with respect to CPUE. On the other hand, an 8-month lag between SSTA and SSTN34 was found, which could be indicate a remote effect from the equator on anchovy when considering the coastal trapped waves as a source of variability in the landings. The dynamics of coastal trapped waves has been studied by different authors (Pizarro et al., 1994, 2001; Hormazabal et al., 2002; Yáñez et al., 2008). The coastal trapped waves influence the dynamics of the sea level, sea surface temperature, and currents on an interseasonal scale (Shaffer et al., 1997; Hormazabal et al., 2001, 2002). Morales et al. (1999) pointed out a direct relationship between the coastal trapped waves and the variability of the oxygen minimum layer, which is directly associated to the thermocline depth in the area off Antofagasta (Morales et al., 1999; Ulloa et al., 2001). The direct relationship between the minimum oxygen layer and the vertical distribution of the anchovy was documented by Morales et al. (1996). Therefore, it is suggested that the effect of SSTN34 on landings is associated with changes in the resource accessibility–vulnerability: accessibility is related to a geographical area (fishing ground), and the vulnerability depends on the interaction between fishing gear (in this case purse seine) and the behavior of the resource. Therefore, anchovy would cause accessibility and vulnerability changes to fishing gear to migrating against long-term environmental changes in response to warm–cold–warm periods, while vertical migration in response to deepening of the thermocline would result in an increase in the vertical distribution of the anchovy.

Landings integrate all the environmental variability and any other factors not considered among the input variables. This explains the great importance of landings in the sensitivity analysis of calibrated models. Regarding the autocorrelation study of anchovy landings performed by Gutiérrez-Estrada et al. (2007), in which they show that the anchovy landing series is autocorrelated, CA_{t-1} is the most important variable. Moreover, according to Yáñez et al. (2008), anchovy landings are the longest analyzed fishing series and an empirical orthogonal function (EOF) analysis conducted between environmental fluctuations and landings, found the best

simulation was obtained from the landing series. Moreover, landings show clear interdecadal, interannual and intraseasonal fluctuations, which may be directly related to the abundance of the resource. This is why landings are proposed as a proxy for the abundance of anchovy in northern Chile.

4.2. Sardine abundance models

Regarding the non-linear cross-correlation analysis of the sardine model, the most representative ages caught for the species were 5, 6, 7 and 8 years (Yáñez and Barbieri, 1989), which are mainly concentrated at 5–6 year periods (69–72 months). The latter was considered to lag the sardine series to 72 months.

The lags for environmental variable can be gathered into two groups. The first group corresponds to values beyond –36 months that may be related to the sardine reproduction and recruitment conditions, considering the spawning that it takes between 2 and 3 years for this species to be recruited (Serra and Tsukayama, 1988; Alheit and Ñiquen, 2004). The second group involves those with lags less than 36 months, which may be associated with availability of the resource under different environmental conditions.

Sardines show a increased production during warm periods, except for the 1982–83 El Niño event when a significant decrease was observed. On the other hand, cold conditions may cause unfavorable conditions for sardines (Yáñez et al., 2001). Sardine recruitment and SST show a dome-shaped relationship with a maximum at 17.7 °C (Yáñez et al., 2001). In fact, the relationship with SST has a quadratic form that explains the increase and decrease of the sardine CPUE (Yáñez et al., 2001, 2003, 2008); this justifies the search for non-linear relations to find lags and the subsequent modeling through ANN. SSTA lags (SSTA_{t-18} and SST_{t-30}), below 36 months, may be related to accessibility and vulnerability changes of the resource associated with changes in the thermal structure, such as El Niño events, affecting the anchovy and sardine distribution, not only horizontally but also vertically with depth (Yáñez et al., 2001). Historical time series suggest that El Niño events affect sardine populations with lag of 6–36 months.

Based on the sensitivity analysis, SSTA was not one of the most important variables of the sardine model (above-mean ratio). Other variables as fS, ET or CA seemed to play a more relevant role in sardine abundance, as in the sensitivity analysis they are identified as top variables. However, the importance ratio of the other variables was not high, so that all variables contributed in a similar manner to the model.

During long warm periods, turbulence (dispersing food) and Ekman transport (advecting the reproductive elements towards the offshore west) increase. The proximity and duration of warm water in this area supports the abundance increases of sardine and, consequently, fisheries develop (Parrish et al., 1983). According to Yáñez et al. (2008), the sardine CPUE showed positive non-linear relations with fS and ET, a fact that supports the search of non-linear relations through cross-correlation and subsequent ANN modeling. On the other hand, a high correlation (0.73) has been estimated between the ET index and the parental biomass of pre-recruit sardine (average i to i-3). Cury and Roy (1989) have suggested the wind stress causes significant mixing of the water column when speeds are 5–6 m/s; this implies dissemination of plankton concentrations. The ET lags higher than 36 months were ET_{t-37}, ET_{t-49}, while those lower than 36 months were ET_{t-1}, ET_{t-7}, ET_{t-19}, ET_{t-31}.

The change of climate regime observed in 1976 agrees with the decrease of anchovy fisheries and the development of sardine fisheries (Yáñez et al., 2001; Yáñez et al., 2008), a fact that is related to the long-term SOI variability, which decreased as the local variables increased (Yáñez et al., 2008). However, the SOI has not been clearly associated with the SST change since mid-1980 (Yáñez

et al., 2001). The maximum significant lag detected for this variable was SOI_{t-39} , which may be considered as a proxy for basin-scale changes of environmental conditions, such as El Niño events or regime changes.

In the case of succession between anchovy and sardine off the coasts of Peru and Chile, climatic fluctuations can cause predation or prey size and abundance variations, thus promoting the dominance of one species or another (Yáñez et al., 2003; Alheit and Ñiquen, 2004). In this sense, sardine is a great predator of anchovy eggs (Santander et al., 1983).

4.3. Final considerations

The approach of Yáñez et al. (2008), who suggested sea surface temperature at Antofagasta as an environmental proxy, is validated for both models. This variable, may represent the environmental effects on both species, particularly recruitment and its spatial distribution against environmental phenomena such as El Niño events.

The predictive lag capacity between species abundance and the input variables of these models should be noted. We have up to 2 and 6 years lag on the input variables for the anchovy and sardine models, respectively. It should be noted that the lags between the environmental variables and landings or CPUE are not always exact, due to the occurrence of anomalous periods, such as El Niño events.

Climatic variations associated with El Niño events have an effect on the abundance and distribution of pelagic fish and fishing effort. However, long-term climatic variability is mainly affects fishing activity; such changes may be favorable or unfavorable for some species. Therefore, the suggested predictive models must be used with care. In fact, besides the difficulties of climate prediction, uncertainty must be regarded as a highly important factor considering the behavior and evolution of the population changes in atmospheric conditions.

Finally, it is important to mention that this method is clearly a first step, heavily dependent on the individuals conducting the analyses and future work should improve (from a statistical point of view) and automate the process.

In conclusion, this work focused on the prediction of anchovy and sardine abundance indexes using a multivariate approach. The low error levels in the validation phase of the models suggest that the objective was accomplished. However, it must be pointed out that difficulty of a good prediction depends not only on the relationships between environment, the resource and the spatial and temporal complexity of the environmental effects, but also on the type of relationships that are assumed between the variables and the modeling processes employed.

Acknowledgements

The authors wish to express their gratitude to the Agencia Española de Cooperación Internacional and the Pontificia Universidad Católica de Valparaíso for their support with this research under the funded Projects “Aplicación de métodos clásicos y heurísticos para la predicción de pesquerías pelágicas en la costa norte de Chile” and “Predicción de pesquerías pelágicas de la zona norte de Chile a través de redes neuronales, Fase I y II”.

References

Abrahart, R.J., See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological Processes* 14, 2157–2172.

Alheit, J., Ñiquen, M., 2004. Regime shift in the Humboldt Current ecosystem. *Progress in Oceanography* 60, 201–222.

Allan, R., Nicholls, N., Jones, P., Butterworth, I., 1991. A further extension of the Tahiti-Darwin SOI, early ENSO events and Darwin pressure. *Journal of Climate* 4, 743–749.

Anctil, F., Rat, A., 2005. Evaluation of neural network streamflow forecasting on 47 watersheds. *Journal of Hydrological Engineering* 10 (1), 85–88.

Bakun, A., McLain, D., Mayo, F., 1974. The mean annual cycle of coastal upwelling off western North America as observed from surface measurements. *Fishery Bulletin* 72 (3), 843–844.

Barbieri, M.A., Bravo, M., Farías, M., González, A., Pizarro, O., Yáñez, E., 1995. Fenómenos asociados a la estructura térmica superficial del mar observados a través de imágenes satelitales en la zona norte de Chile. *Investigaciones Marinas* 23, 99–122.

Braun, M., Castillo, J., Blanco, J., Lillo, S., Reyes, H., 1995. Monitoreo hidroacústico y oceanográfico de recursos pelágicos en I y II regiones. Informe Final FIP/93-15, 1–172.

Carr, M.E., 2002. Estimation of potential productivity in Eastern Boundary Currents using remote sensing. *Deep Sea Research Part II: Topical Studies in Oceanography* 49, 59–80.

Castillo, J., Córdova, J., Saavedra, A., Espejo, M., Gálvez, P., Barbieri, M.A., 2002. Evaluación acústica de la biomasa, abundancia, distribución espacial y caracterización de cardúmenes de anchoveta en el periodo de reclutamiento. Primavera 2001. Evaluación del reclutamiento de anchoveta en I y II regiones, temporada 2001–2002. Informe Final FIP-IT/2001–11, 1–207.

Chen, D., Hare, S., 2006. Neural network and fuzzy logic models for pacific halibut recruitment analysis. *Ecological Modelling* 195, 11–19.

Cury, P., Roy, C., 1989. Optimal environmental window and pelagic fish recruitment success in upwelling areas. *Canadian Journal of Fisheries and Aquatic Sciences* 46, 670–680.

Deser, C., Wallace, J., 1990. Large-scale atmospheric circulation features of warm and cold episodes in the tropical Pacific. *Journal of Climate* 3, 1254–1281.

Elsberry, R., Garwood, R., 1978. Sea surface temperature anomaly generation in relation to atmospheric storms. *Bulletin of the American Meteorological Society* 59, 786–789.

Espíndola, F., Yáñez, E., Bohm, G., 2005. Standardization of match rates of the pelagic fish resources in the northern Chile. *Gayana (Concepción)* 69 (2), 300–318.

Freón, P., Mullon, C., Voisin, B., 2003. Investigating remote synchronous patterns in fisheries. *Fisheries Oceanography* 12 (4), 443–457.

Griñó, R., 1992. Neural networks for univariate time series forecasting and their application to water demand prediction. *Neural Network World* 2 (5), 437–450.

Gutiérrez-Estrada, J.C., Silva, C., Yáñez, E., Rodríguez, N., Pulido-Calvo, I., 2007. Monthly catch forecasting of anchovy (*Engraulis ringens*) in the north area of Chile: non-linear univariate approach. *Fisheries Research* 86, 188–200.

Gutiérrez-Estrada, J.C., Yáñez, E., Pulido-Calvo, I., Silva, C., Plaza, F., Bórquez, C., 2009. Pacific sardine (*Sardinops sagax*, Jenyns 1842) landings prediction. A neural network ecosystemic approach. *Fisheries Research* 100 (2), 116–125.

Hagan, M., Menhaj, M., 1994. Training feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks* 5 (6), 989–993.

Hormazabal, S., Shaffer, G., Letelier, J., Ulloa, O., 2001. Local and remote forcing of sea surface temperature in the coastal upwelling system off Chile. *Journal of Geophysical Research* 106, 16657–16672.

Hormazabal, S., Shaffer, G., Pizarro, O., 2002. Tropical Pacific control of intraseasonal oscillations off Chile by way of oceanic and atmospheric pathways. *Geophysical Research Letters* 29 (6), 5.1–5.4.

Hunter, A., Kennedy, L., Henry, J., Ferguson, I., 2000. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Computer Methods and Programs in Biomedicine* 62, 11–19.

Iyer, M., Rhinehart, R., 1999. A method to determine the required number of neural-network training repetitions. *IEEE Transactions on Neural Networks* 10 (2), 427–432.

Kitanidis, P.K., Bras, R.L., 1980. Real time forecasting with a conceptual hydrological model. 2: Applications and results. *Water Resources Research* 16 (6), 1034–1044.

Legates, D., McCabe, G., 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research* 35 (1), 223–241.

Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* 120, 65–73.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulancier, S., 1996. Application of neural networks to modelling non-linear relationships in ecology. *Ecological Modelling* 90, 39–52.

Makkearsorn, A., Chang, N.B., Zhou, X., 2008. Short-term streamflow forecasting with global climate change implications – a comparative study between genetic programming and neural network models. *Journal of Hydrology* 352, 336–354.

Montecinos, A., Purca, S., Pizarro, O., 2003. Interannual-to-interdecadal SST variability along the western coast of South America. *Geophysical Research Letters* 30, 1570.

Morales, C., Blanco, J.L., Braun, M., Reyes, H., Silva, N., 1996. Chlorophyll-a distribution and associated oceanographic variables in the upwelling region off northern Chile during the winter and spring 1993. *Deep Sea Research* 43 (3), 267–289.

Morales, C., Hormazabal, S., Blanco, J.L., 1999. Interannual variability in the mesoscale distribution of the depth of the upper boundary of the oxygen minimum layer off northern Chile (18–24S): implications for the pelagic system and biogeochemical cycling. *Journal of Marine Research* 57, 909–932.

Murdoch, W.W., 1994. Population regulation in theory and practice. *Ecology* 75, 271–287.

- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. I: A discussion of principles. *Journal of Hydrology* 10, 282–290.
- Özesmi, S., Tan, C., Özesmi, U., 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecological Modelling* 195, 83–93.
- Parrish, R.H., Bakun, A., Husby, D.M., Nelson, C.S., 1983. Comparative climatology of selected environmental processes in relation to Eastern boundary current pelagic fish reproduction. *FAO Fisheries Report* 3 (291), 731–777.
- Pizarro, O., Hormazabal, S., González, A., Yáñez, E., 1994. Variabilidad del viento, nivel del mar y temperatura en la costa norte de Chile. *Investigaciones Marinas* 22, 85–101.
- Pizarro, O., Clarke, A., Van-Gorder, S., 2001. El Niño sea level and currents along the South American Coast: comparison of observations with theory. *Journal of Physical Oceanography* 31, 1891–1903.
- Qi, M., Zhang, G., 2001. An investigation of model selection criteria for neural network time series forecasting. *European Journal of Operational Research* 132, 666–680.
- Real, R., Márquez, A.L., Estrada, A., Muñoz, A.R., Vargas, J.M., 2008. Modelling chorotypes of invasive vertebrates in mainland Spain. *Diversity and Distribution* 14, 364–373.
- Rumelhart, D., Hinton, G., Williams, R., 1986. “Learning” representations by backpropagation errors. *Nature* 323, 33–36.
- Santander, H., Alheit, J., McCall, A.D., 1983. Egg mortality of the Peruvian anchovy (*Engraulis ringens*) caused by cannibalism and predation by sardines (*Sardinops sagax*). *FAO Fisheries Report* 291 (3), 1011–1025.
- Serra, R., Tsukayama, I., 1988. Sinopsis de datos biológicos y pesqueros de *Sardinops sagax* en el Pacífico suroriental. *FAO Sinopsis Pesca* 13 (1), 60pp.
- Shaffer, G., Pizarro, O., Djurfeldt, L., Salinas, S., Rutllant, J., 1997. Circulation and low-frequency variability near the Chilean coast: remotely forced fluctuations during the 1991–92 El Niño. *Journal of Physical Oceanography* 27, 217–235.
- Stenseth, N.C., Lekve, K., Gjøsæter, J., 2002. Modeling species richness controlled by community-intrinsic and community-extrinsic processes: coastal fish communities as an example. *Population Ecology* 44, 165–178.
- Suryanarayana, I., Braibanti, A., Rao, R.S., Ramam, V.A., Sudarsan, D., Rao, G.N., 2008. Neural networks in fisheries research. *Fisheries Research* 92, 115–139.
- Tan, Y., van Cauwenberghe, A., 1999. Neural-network-based *d*-step-ahead predictors for nonlinear systems with time delay. *Engineering Applications of Artificial Intelligence* 12 (1), 21–25.
- Ulloa, O., Escribano, R., Hormazabal, S., Quiñones, R., González, R., Ramos, M., 2001. Evolution and biological effects of the 1997–98 El Niño in the upwelling ecosystem off northern Chile. *Geophysical Research Letters* 28, 1591–1594.
- Valdés, J., Ortlieb, L., Gutiérrez, D., Marinovic, L., Vargas, G., Sifeddine, A., 2008. 250 years of sardine and anchovy scale deposition record in Mejillones Bay, northern Chile. *Progress in Oceanography* 79, 198–207.
- Ventura, S., Silva, M., Pérez-Bendito, D., Hervás, C., 1995. Artificial neural networks for estimation of kinetic analytical parameters. *Analytical Chemistry* 67 (9), 1521–1525.
- Winters, G.H., Wheeler, J.P., Stansbury, D., 1993. Variability in the reproductive output of spring-spawning herring in the north-west Atlantic. *ICES Journal of Marine Science* 50, 15–25.
- Yáñez, E., Barbieri, M.A., 1989. Principal pelagic resources exploited in northern Chile and their relationship to environmental variations. In: Wyatt, T., Larrañeta, M.G. (Eds.), *Long Term Changes in Marine Fish Populations*, Instituto de Investigaciones Marinas de Vigo, Spain, pp. 197–219.
- Yáñez, E., Barbieri, M.A., Silva, C., Nieto, K., Espíndola, F., 2001. Climate variability and pelagic fisheries in northern Chile. *Progress in Oceanography* 49, 581–596.
- Yáñez, E., Barbieri, M.A., Silva, C., 2003. Fluctuaciones ambientales de baja frecuencia y principales pesquerías pelágicas chilenas. In: Yáñez, E. (Ed.), *Actividad pesquera y de acuicultura en Chile*, Escuela de Ciencias del Mar. PUCV, Valparaíso, pp. 109–122.
- Yáñez, E., Hormazabal, S., Silva, C., Montecinos, A., Barbieri, M.A., Valdenegro, A., Órdenes, A., Gómez, F., 2008. Coupling between the environmental and the pelagic resources exploited off North Chile: ecosystem indicators and conceptual model. *Latin American Journal of Aquatic Research* 36 (2), 159–181.
- Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology* 181, 23–48.