Fisheries Research 100 (2009) 116–125

Contents lists available at ScienceDirect

Fisheries Research

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

Pacific sardine (*Sardinops sagax*, Jenyns 1842) landings prediction. A neural network ecosystemic approach

Juan Carlos Gutiérrez-Estrada[a,*], Eleuterio Yáñez[b], Inmaculada Pulido-Calvo[a], Claudio Silva[b], Francisco Plaza[b], Cinthya Bórquez[b]

[a] *Departamento de Ciencias Agroforestales, EPS, Campus La Rábida, Universidad de Huelva, 21819 Palos de la Frontera, Huelva, Spain*
[b] *Escuela de Ciencias del Mar, Pontificia Universidad Católica de Valparaíso, casilla 1020, Valparaíso, Chile*

ARTICLE INFO

Article history:
Received 30 October 2008
Received in revised form 22 June 2009
Accepted 23 June 2009

Keywords:
Time series
Anchovy
Engraulis ringens
Computational neural network
Multiple linear regressions
Generalised additive model

ABSTRACT

In this study the performances of computational neural networks (CNNs), multiple linear regressions (MLRs) and generalised additive models (GAMs) to predict Pacific sardine (*Sardinops sagax*) landings and to analyse their relationships with environmental factors in the north area off Chile were studied. For this purpose several local and global environmental variables and indexes (sea surface temperature, sea level and Ekman transport index in the Chilean coast and, sea surface temperature in the area Niño 3 + 4 and Niño 1 + 2, and the south oscillation index) were considered as inputs or independent variables. Additionally, several CNNs were calibrated and validated adding the anchovy (*Engraulis ringens*) landings in the same area as model inputs. The time lags of the variables considered were selected through analysis of the non-linear cross-correlation functions and an alternative form of sensitivity analysis based on the approach of the missing value problem. The analysis of error measures with validation data set showed that the best results were obtained when local and global variables were used separately and combined with anchovy landings. Globally, the best result was given by a CNN with 18 input variables (model CNN 6(II) which only considered global variables and anchovy landings) and 10 neurons in a hidden layer. For this configuration the explained variance was slightly higher to 86% which supposed a standard error of prediction of 7.66%. These results were significantly better than those obtained with MLRs and GAMs. The strong correlation between predicted and observed sardine landings suggests that CNNs captured the trend of the historical data. Also, the generalisation capacity together the sensitivity analysis allowed us to identify the variables with a high weight in the model and partially to interpret the statistical functional relationships between these environmental variables and sardine landings.

1. Introduction

The Pacific sardine (*Sardinops sagax*) is one of the dominant clupeoid species of the Humbolt boundary current system in the north area off Chile. The same as California, Kuroshio and Benguela systems, Humbolt current system is a highly productive ecosystem which allowed extracting a mean of 1.8 millions of tons per year of this pelagic species from 1976 to 1996, which globally positions the sardine as the second more important pelagic resource of Chile. In spite of the economical importance of this fishery, studies of quantitative predictions of sardine landings and the analysis of the strengths of the environment–resource relationships are scarce. Yáñez et al. (2001, 2003) indicated that the abundance changes associated with strong environmental fluctuations is one of the main problems that hinders reaching these objectives.

Fluctuations of the abundance of *S. sagax* and other clupeoid species as *Engraulis ringens* and other *Engraulis* species have been observed in California, Japan, Perú, South Africa and Australia (Lluch-Belda et al., 1989; Lluch-Cota et al., 1997; Schwartzlose et al., 1999; Ward et al., 2001). In the north area off Chile, these fluctuations have been related to changes of the marine ecosystems in different spatial–temporal scales (Alheit and Bernal, 1993; Hare et al., 2000; Chávez et al., 2003; Yáñez et al., 2003; Alheit and Ñiquen, 2004). Yáñez et al. (2001) indicated that changes in the composition of the pelagic fisheries of the north area off Chile in the period 1950–1998 were associated with changes of sea surface temperature (SST) measured near the coast line. Also, changes of this marine ecosystem have been associated with mesoscale structures as coastal-trapped waves and Rossby waves (Hormazabal et al., 2002, 2004; Pizarro et al., 2002). In the inter-annual scale, Chávez et al. (2002) showed the relationships between fisheries resources in different development states and temperature levels, salinity, chlorophyll concentration, sea level and zooplanckton. Bernal et al. (1983), Sharp and McLain (1993), Bakun and Broad (2001), Yáñez et

---

[*] Corresponding author.
E-mail address: juanc@uhu.es (J.C. Gutiérrez-Estrada).

0165-7836/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2009.06.014

al. (2001, 2003), Chávez et al. (2003), Montecinos et al. (2003), and Bertrand et al. (2004) reported changes of sardines and anchovies abundances in a long time scale and related the abundance fluctuations with the global condition of Pacific Ocean as El Niño-southern oscillation (ENSO) events.

Linking between sardine abundance variability and environmental changes in different spatial–temporal scales opens the possibility to understanding and to predict the sardine landings fluctuations. To reach this objective requires the selection of adequate statistical and mathematical tools because, as suggested Steele (1985), Murdoch (1994) and Stenseth et al. (2002), the marine ecosystems are non-equilibrium systems characterized by the presence of a high level of stochasticity and non-linear relationships between environmental, biological and ecological variables. James and McCulloch (1990) reported that conventional models (like a wide range of classic multivariate techniques) assume continuity, steady-states and linear or simple polynomial relationships between variables which is a serious shortcoming when highly non-linear relationships must be modelled (Lek et al., 1996). Particularly, in clupeoid fisheries the traditional management techniques for estimating sustainable yields and population size from catches statistics have proved to be of little worth (Butterworth, 1983; Pitcher and Hart, 1982).

In this sense the heuristic techniques, like computational neural networks (CNNs), may be more adequate since these models are non-linear mathematical structures capable of representing complex non-linear processes which are not constrained by assumptions about the type of relationship between studied variables (Rumelhart et al., 1986). These features have favoured the development of applications to management and planning of fisheries including the modelling of abundance, recruitment, biomass stock, distribution and catches of different fisheries (Komatsu et al., 1994; Huse and Gjøsæter, 1999; Huse and Ottersen, 2003; Hyun et al., 2005; Arregui et al., 2006; Gutiérrez-Estrada et al., 2007).

Therefore, this paper evaluates the performance of CNNs to predict sardine landings in the north area off Chile and to analyse their relationships with environmental factors. For this modelling, two hypotheses were investigated: (*i*) environmental changes in different time scale explains sardine landings fluctuation; (*ii*) the Pacific sardine landings are modulated by the anchovy landings. Additionally, the accuracy of the CNN models was compared with more classic linear and non-linear fit techniques such as multiple linear regressions (MLRs) and generalised additive models (GAMs). As the performance of data driven techniques, such as CNNs or GAMs, is highly sensitive to the selected input variables (Fernando et al., 2009), this study explored procedures for selecting an optimal model input vector from a set of candidates at different time lags.

2. Methods

2.1. Study area and data source

The present study was carried out with *Sardinops sagax* landings fished in the stock unit located in the Chilean north coast. Catch area included the operational area of the purse seine fleet of the Region I and II of Chile (18°S–24°S), from the coast line (first five marine miles) to the Chilean exclusive economic zone (200 marine miles) (73°W). Monthly sardine (LSAR) and anchovy (LANC) landings data were recovered from the annual fisheries reports of the Fishery National Service of Chile (SERNAPESCA, 1978–1997). The basic statistics of sardine and anchovy time series are shown in Table 1.

Several local and global environmental factors were used as input or independent variables. Time series from 1978 to 1997 of wind magnitude and direction were used to estimate the monthly Ekman transport (EK) (Bakun, 1973). For this same period, monthly averages of: (a) sea level (SL) and sea surface temperature (SST) in the coastal station of Antofagasta (23°39′S); (b) south oscillation index (SOI); (c) sea surface temperature in areas Niño 3 + 4 (N34) (5°N–5°S, 120°W–170°W) and Niño 1 + 2 (N12) (0°–10°S, 80°W–90°W) were available (Table 1).

2.2. Modelling techniques

The prediction of sardine landings was performed using three different techniques: (a) computational neural networks (CNNs); (b) multiple linear regressions (MLR); (c) generalised additive models (GAM). CNNs are mathematical models inspired by the neural architecture of the biological nervous systems. The most widely studied and used structures are multilayer feed forward networks or multilayer perceptrons (Rumelhart et al., 1986). These models 'learn' in an iterative way in which the data set are introduced to the neural network the necessary times until to reach a determined error level (one iteration where all the data set are introduced to the CNN is named epoch). These supervised CNNs allow the analysis of complex data sets and the determination of non-linear relationships between dependent and independent variables. A detailed description of multilayer perceptrons CNNs performance can be found in Tsoukalas and Uhrig (1997) and Gutiérrez-Estrada et al. (2008).

There are many CNN calibration or learning methods. In this work, the Levenberg–Marquardt algorithm was applied (Shepherd, 1997). This is a second-order non-linear optimization algorithm with a very fast convergence which is recommended by several authors (Tan and van Cauwenberghe, 1999; Anctil and Rat, 2005).

In order to test the 'good-of-fit', the CNNs results were compared with MLRs and GAMs models. MLRs models estimate $q + 1$ parameters of the linear combination of q factors where these parameters or regression coefficients represent the independent contributions of each q independent variable to the prediction of one dependent variable e. The global statistical significance of the relationship between e and the independent variables is analysed by means of a variance analysis to ensure the validity of the model in a quantified manner (Hair et al., 1998).

On the other hand, GAM models represent a generalisation of a multiple regression. Specifically, in linear regression, a linear least-squares fit is computed for a set of predictors or q variables, to predict a dependent e variable. A generalisation of a multiple regres-

Table 1
Mean, standard deviation, maximum, minimum and variation coefficient (%) of the dependent (d) and independents (i) variables.

Variables	Mean	Standard deviation	Maximum	Minimum	Variation coefficient (%)
(i) Ekman transport ($m^3\ s^{-1}$)	662.31	216.18	1353.07	256.27	32.64
(i) Sea surface temperature (°C)	17.69	2.07	24.50	12.20	11.72
(i) Sea level (cm)	0.72	0.06	1.00	0.58	8.13
(i) Sea surface temperature (°C), area 1 + 2	23.25	2.40	29.15	19.12	10.31
(i) Sea surface temperature (°C), area 3 + 4	27.06	1.03	29.36	24.27	3.82
(i) South oscillation index	−0.37	1.10	2.03	−4.14	−298
(i) Anchovy monthly landings (tons)	57308.50	73471.80	422013	0	128.20
(d) Sardine monthly landings (tons)	77570.10	96574.70	457611	0	124.50

sion model implies maintenance of the additive nature of the model, but replacing the simple terms of the linear equation by a non-parametric function. Thus, instead of a single coefficient for each variable (additive term) in the model, in additive models an unspecified (non-parametric) function is estimated for each predictor, to achieve the best prediction of the dependent variable values (Hastie and Tibshirani, 1990). In this study Normal and Gamma distribution models with identity, inverse and logit link functions and cubic spline smooth function with three degrees of freedom were fitted.

2.3. General procedure

Previously to the models calibrations, data series of sardine and anchovy landings were smoothed. Fréon et al. (2003) reported that this procedure is recommended to remove high-frequency noise which can be caused by serious under-reports of the true landings (Alheit and Ñiquen, 2004). In this work, the smooth function was a mean of 3 months centred. In this way, the sardine and anchovy landings ($v_{wn,t}$) were smoothed as:

$$v_{won,t} = \frac{v_{wn,t-1} + v_{wn,t} + v_{wn,t+1}}{3} \qquad (1)$$

where $v_{won,t}$ is the monthly smoothed value of $v_{wn,t}$ and $v_{wn,t-1}$, and $v_{wn,t+1}$ are the previous and posterior monthly values to $v_{wn,t}$, respectively.

Later on, the data set was divided in two subsets: the first one (training set = calibration set + selection set, CS + SS) was composed of 85% sardine landings data and the second one (validation set, VS) was composed of 15% remaining data. The 75% of the training set was used for the CNNs calibration or learning (CS set). The 25% of data of the training set (SS set) was used for internal validation which allows the CNN training process to stop avoiding over-learning effects.

An inherent problem associated with CNNs is their tendency to get stuck in local minima. To solve this problem, the same CNN architecture was trained 30 times, re-sampling the training and validating sets and re-starting with a random set of weights (Anctil and Rat, 2005). This repetition level allows selecting a CNN which is among the best 14% of the distribution of all possible CNN with this configuration at the 99% confidence level (Iyer and Rhinehart, 1999). The best hidden configuration was determined comparing the mean error for the validation set by means of a Kruskal-Wallis ANOVA by ranks test (Gutiérrez-Estrada et al., 2008). Once determined the best hidden configuration, the CNN with the lowest error using the validation data set was selected.

Once the external validation for each model has been carried out, an alternative form of sensitivity analysis based on the approach of the missing value problem was used. The sensitivity analysis was carried out by replacing each variable by missing values and assessing the effect upon the output error. Then, the new error calculated was compared with the original error to obtain a ratio value (ratio = error of the model with a variable with missing values/error of the model with all variables). In this way, for a variable *x*, a ratio with a value equal to 1 or very close to 1 indicates that this variable has a very low weight in the general structure of the model (Hunter et al., 2000).

The number of input variables was determined by means a trial and error process, using as a guide the first sensitivity analyses carried out in a CNN which considered as inputs all environmental variables (SST, EK, SL, N34, N12, SOI). Independently of the time lags number (see Section 2.4) considered for each variable, the number of input variables oscillated between one and six. In all cases, the output or dependent variable was LSAR. CNN configurations with one or two hidden layers with a minimum of five nodes and a maximum of twenty nodes for each layer were successively trained.

Table 2

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* are 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	$Y = \beta_0{}^{(1/X)}$
(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	$Y = \beta_0[\beta_1^X][X_2^\beta]$
(18) Modified Hoerl	$Y = \beta_0[\beta_1^{(1/X)}][X_2^\beta]$
(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	$Y = \beta_0[X_1^\beta](1 - X)_2^\beta$

In the case of MLR and GAM models, the same number of repetitions and the same training and validation sets were used. All models were implemented using STATISTICA 7.0 (StatSoft Inc., 1984–2005).

2.4. Model order determination

A key aspect in time series modelling is the correct determination of the model order. That is to say, how much and what lags of each input variable explains the highest variance level and provides the lowest error in the external validation phase? An incorrect selection of the input variables (and their time lags) facilitates the presence non-desired effects such as naïve effect, increases the presence of local optima and hinders the extraction of physical meaning from calibrated models. Many of the neural approaches used in the specialised literature, have adopted methodologies such as autocorrelation and partial autocorrelation functions analysis, linear cross-correlation analysis or spectral analysis (Jain and Kumar, in press; Shrestha et al., 2005; Elgaali and García, 2007; Pulido-Calvo and Portela, 2007; Gutiérrez-Estrada et al., 2007) to solve this problem. However, seems evident that these methodologies are designed to capture the linear correlation between different lags. This can mask the strength of the non-linear relationships, favouring the omission of relevant inputs (model will be under specified).

To mitigate this problem, in the study the models orders were established through the analysis of the linear and non-linear cross-correlation functions calculated between LSAR and all input or independent variables. In this way, 22 models (Table 2) were fitted in order to extract the linear and non-linear cross-correlations for each possible two-dimensional relationship. The fitting for each pair of variables was carried out considering the influence of the independent variable lagged in the time. Thus, lags from *t* = 0 (no lagged) to *t* = 36 (lagged 3 years) were analysed. The maximum lag (*t* = 36) was selected introducing the following practical and biological approaches: (a) a compromise between the length of the time series, the prediction unit (step of period prediction) and the smoothed function proposed; (b) the annual recruitment of the sardine in the study area is given at age class 2+ and 3+ (Yáñez et al., 2001; Alheit and Ñiquen, 2004).

In order to facilitate the fittings comparison, the 22 models showed in Table 2 were implemented in a computer program called CORN 1.0 which was specifically designed for this study. For each pair of variables and for each time lag, the program fits all non-linear functions by mean minimum least square method obtaining the total explained variance (TEV). The absolute cross-correlation index (R_n) was obtained as $\text{TEV}^{1/2}$. The selected non-linear function maximized the summation of R_n for all lags considered.

Once the non-linear function was selected, the time lags choice of each variable was carried out using the following criteria: (a) if the selected non-linear fit and linear fit have a similar trend then the lags with very high R_n were chosen; (b) if a different trend between linear and non-linear fits were found then the time lags associated to lack of trend between both functions was selected.

2.5. Error measures

Important aspects of any model are the capture of mapping data and its capacity to generalise from examples. Generalisation refers to the model capacity to provide a correct response with patterns (of the past or future) that have not been employed in its calibration. The generalisation capacity can be evaluated by means different error or accuracy measures. In this study seven accuracy measures were calculated in the calibration and validation phases of CNN, MLR and GAM models: (a) the correlation coefficient (R); (b) the determination coefficient (R^2); (c) the square root of the mean square error (RMSE); (d) the standard error of prediction (SEP); (e) the average relative variance (ARV); (f) the efficiency coefficient (E_2); (g) the persistence index (PI) (Nash and Sutcliffe, 1970; Kitanidis and Bras, 1980; Ventura et al., 1995; Legates and McCabe, 1999).

For each accuracy measure, the benchmark of the worst permissible error was calculated. McLaughlin (1983) suggests that a naïve model determines the forecasting accuracy benchmark of any model. The basic naïve model is defined as the next period's level will be the same as that of the preceding period.

Additionally, it is interesting to analyse the model generalisation capacity at different levels of the dependent variable, that is to say, what is the model response when the sardine landings are low. Thus, all accuracy measures were re-calculated after 1990 because from this year the landings level was significantly lower than the recorded between 1978 and 1989.

3. Results

3.1. Cross-correlation function analyses

Fig. 1 shows the values of absolute cross-correlation index (R_n) between LSAR and the environmental variables SST, EK, SL, N34, N12, SOI and anchovy landings (LANC) for the 36 lags considered (cross-correlation functions). For all cases, it is possible to observe that the best non-linear fit explains a higher variance level than the linear cross-correlation index which is equivalent to Pearson correlation index (R).

Four variables (LANC, N12, SL and SST) showed a similar trend in the linear and non-linear fits. For the remaining variables (N34, SOI and EK) different partial trends were found. For example, in the case of N34 a parabolic fit provided a slight higher R_n for all lags. From the lag 0 to lags 14–15, the parabolic and linear fits had a similar trend. However, between the lags 16 and 24 months the parabolic fit exhibited a different behaviour in relation to the linear cross-correlation function. Similar effect was found for SOI. For this variable, a clear non-linear trend was found between 8 and 16 lagged months. On the other hand, EK cross-correlation function showed four periods of high R_n with respect to the linear fit. The first from lag 0 to lags 3–4 and the remaining from 11–12 to 15–16 lagged months, from 23–24 to 27–28 lagged months and from 34–35 to the last lag considered.

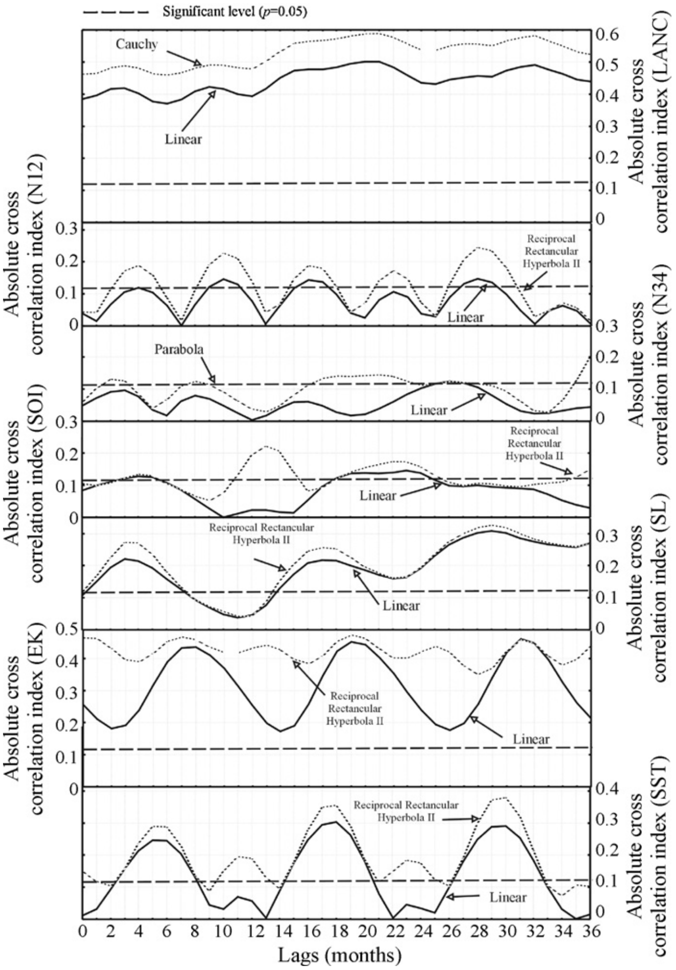

Fig. 1. Linear and non-linear cross-correlation functions for each possible two-dimensional relationship.

The visual analysis of these cross-correlations functions allowed us to detect *a priori* the lags with a high weight on the system. Table 3 shows the lags selected for each variable as well as those rejected after the first sensitivity analyses.

3.2. Pacific sardine landings prediction

A total of 5280 (22 input configurations × 8 hidden configurations × 30 repetitions) CNNs were calibrated and validated. The 22 input configurations included combinations of local and global variables together with anchovy landings. Table 4 shows the CNNs configurations used in this study. Additionally, the input variables that provided the best CNN results were used as independent variables for the MLR and GAM models. Thus, 30 MLRs (1 independent variables configuration × 30 repetitions) and 90 GAMs (1 independent variables configuration × 2 distributions × 3 link functions × 30 repetitions) were calibrated and validated.

Globally, the best fits were obtained when local and global scale variables were independently used. Thus, the CNN 6 (global scale variables as inputs) and CNN 8 (local scale variables as inputs) configurations provided the best accuracy measures. Good fits were

Table 3

Lags selected *a priori* for each variable after cross-correlation functions analysis and rejected after the first sensitivity analyses. SST = sea surface temperature in Antofagasta; EK = Ekman transport; SL = sea level in Antofagasta; N34 = sea surface temperature in area Niño 3 + 4; N12 = sea surface temperature in area Niño1 + 2; SOI = south oscillation index; LANC = anchovy landings.

Variable	Lags selected	Lags rejected
LANC	18, 19, 20, 21, 22, 23, 30, 31, 32 and 33	18, 19, 22, 23, 30 and 33
N12	3, 4, 5, 9, 10, 11, 15, 16, 17, 21, 22, 23, 27, 28 and 29	3, 5, 9, 11, 15, 17, 21, 23, 27 and 29
N34	17, 18, 19, 20, 21 and 22	
SOI	11, 12, 13, 14 and 15	11 and 15
SL	2, 3, 4, 16, 17, 18, 19, 28, 29 and 30	2, 4, 16, and 19
EK	0, 1, 2, 3, 12, 13, 14, 15, 23, 24, 25, 26, 27, 35 and 36	0, 3, 14, 15, 23, 24, 25, 26, 27, 35 and 36
SST	0, 5, 6, 7, 17, 18, 29, 30 and 31	0, 5, 29, 30 and 31

Table 4

Input CNN configurations considered. Each input configuration was proved with 8 different hidden architectures and the training of each CNN was repeated 30 times. SST = sea surface temperature in Antofagasta; EK = Ekman transport; SL = sea level in Antofagasta; N34 = sea surface temperature in area Niño 3 + 4; N12 = sea surface temperature in area Niño1 + 2; SOI = south oscillation index; LANC = anchovy landings.

CNN number	Input CNN configuration[a]
1	SST, EK, SL, N34, N12, SOI
2	SST, N34
3	SST, N34, N12
4	SST, N12
5	N34, SOI
6	N34, N12, SOI
7	SST, EK
8	SST, EK, SL
9	EK
10	EK, SL
11	N34, N12, SOI, SST

[a] Each configuration was re-calibrated together with LANC.

obtained using EK, SST and SL variables as input variables [CNN 8(I)], although the prediction capacity of this model was slightly improved when four anchovy landings lags ($LANC_{t-20}$, $LANC_{t-21}$, $LANC_{t-31}$, $LANC_{t-32}$) were added [CNN 8(II)] (Table 5). Also, good predictions were obtained by CNNs that had global scale variables as inputs. Concretely, CNNs with 10 neurons in one hidden layer with different lags of N34, N12 and SOI provided the best results [CNN 6(I and II)] (Table 5). CNN 6(I) had as inputs six different lags for N34 ($N34_{t-17}$, $N34_{t-18}$, $N34_{t-19}$, $N34_{t-20}$, $N34_{t-21}$, $N34_{t-22}$), five lags for N12 ($N12_{t-4}$, $N12_{t-10}$, $N12_{t-16}$, $N12_{t-22}$, $N12_{t-28}$) and three lags for SOI (SOI_{t-12}, SOI_{t-13}, SOI_{t-14}) which were selected from the non-linear cross-correlation analysis. The same as CNN8(II), CNN 6(II) was calibrated adding four anchovy landings lags ($LANC_{t-20}$, $LANC_{t-21}$, $LANC_{t-31}$, $LANC_{t-32}$).

Fig. 2a shows the linear regression between observed and estimated sardine landings obtained with CNN 6(I). This regression shows a good fit for the training data set. However, accuracy measures for the validation data set were significantly worse. The level of explained variance only was slightly higher than 55%. Also, this lack of fit was observed in the remaining error terms. This way, the RMSE was clearly higher than 65,000 tons, which supposed SEP value around of 14%. The PI value was significantly lower than 0.6 (PI = 0.3) as a consequence of the slight displacement between observed and estimated sardine landings and the presence of some negative predictions (Fig. 2b).

The fit capacity was significantly improved when the anchovy landings ($LANC_{t-20}$, $LANC_{t-21}$, $LANC_{t-31}$, $LANC_{t-32}$) were added (CNN 6(II); Fig. 3a). The incorporation of four anchovy landings lags was enough to adjust the model order. In the validation, the determination coefficient indicated that the 86% of the variability was

Table 5

Accuracy measures of naïve, best CNNs with local variables [CNN 8(I) and CNN 8(II)], best CNNs with global variables [CNN 6(I) and CNN 6(II)] and the best multiple linear regression and generalised additive models. Calculus for full and validation sets in the periods 1978–1996 and 1990–1996 are shown. SST = sea surface temperature in Antofagasta; EK = Ekman transport; SL = sea level in Antofagasta; N34 = sea surface temperature in area Niño 3 + 4; N12 = sea surface temperature in area Niño1 + 2; SOI = south oscillation index; LANC = anchovy landings; LSAR = sardine landings.

Model	All data set (1978–1996)							Validation set (1978–1996)						
	R	R^2	RMSE (tons)	SEP (%)	E_2	ARV	PI	N	R	R^2	RMSE (tons)	SEP (%)	E_2	ARV
Naïve	0.90	0.81	34729	7.35	0.80	0.19	0	–	–	–	–	–	–	–
CNN 8(I)	0.94	0.89	26104	5.54	0.89	0.11	0.45	30	0.88	0.78	42070	8.98	0.78	0.22
CNN 8(II)	0.97	0.95	20476	4.56	0.95	0.05	0.66	30	0.89	0.80	37374	8.52	0.79	0.21
CNN 6(I)	0.93	0.86	30084	6.34	0.85	0.15	0.27	30	0.74	0.55	65232	14.10	0.37	0.63
CNN 6(II)[a]	0.98	0.96	16384	3.46	0.96	0.04	0.78	30	0.92	0.86	35586	7.66	0.85	0.15
MLR[b]	0.69	0.48	69715	14.76	0.48	0.52	−1.66	30	0.70	0.49	66003	14.22	0.48	0.52
GAM[b,c]	0.87	0.76	38836	8.22	0.76	0.24	−0.25	30	0.86	0.74	49664	10.70	0.71	0.29

Model	All data set (1990–1996)							Validation set (1990–1996)						
	R	R^2	RMSE (tons)	SEP (%)	E_2	ARV	PI	N	R	R^2	RMSE (tons)	SEP (%)	E_2	ARV
Naïve	0.87	0.76	6485	4.55	0.96	0.04	0	–	–	–	–	–	–	–
CNN 8(I)	0.73	0.54	10941	7.68	0.89	0.11	−1.84	12	0.22	0.05	17106	12.21	0.68	0.31
CNN 8(II)	0.59	0.35	5736	4.26	0.96	0.04	−5.61	9	0.20	0.04	10961	8.21	0.81	0.18
CNN 6(I)	0.75	0.56	11659	7.99	0.87	0.13	−1.76	11	0.54	0.29	20871	14.64	0.35	0.64
CNN 6(II)[a]	0.82	0.68	7201	5.05	0.95	0.05	0.27	14	0.61	0.37	10751	7.50	0.85	0.14
MLR[b]	0.25	0.06	22737	15.94	0.53	0.47	−11.29	14	0.43	0.19	18186	12.67	0.57	0.43
GAM[b,c]	0.34	0.11	14526	10.19	0.81	0.19	−4.01	14	0.50	0.25	11694	8.15	0.82	0.18

CNN 8(I) model: $LSAR_t = f(EK_{t-1}, EK_{t-2}, EK_{t-12}, EK_{t-13}; SL_{t-3}, SL_{t-4}, SL_{t-16}, SL_{t-17}, SL_{t-18}, SL_{t-28}, SL_{t-29}, SL_{t-30}; SST_{t-6}, SST_{t-7}, SST_{t-17}, SST_{t-18})$; CNN 8(II) model: $LSAR_t = f(EK_{t-1}, EK_{t-2}, EK_{t-12}, EK_{t-13}; SL_{t-3}, SL_{t-4}, SL_{t-16}, SL_{t-17}, SL_{t-18}, SL_{t-28}, SL_{t-29}, SL_{t-30}; SST_{t-6}, SST_{t-7}, SST_{t-17}, SST_{t-18}, LANC_{t-20}, LANC_{t-21}, LANC_{t-31}, LANC_{t-32})$; CNN 6(I) model: $LSAR_t = f(N34_{t-17}, N34_{t-18}, N34_{t-19}, N34_{t-20}, N34_{t-21}, N34_{t-22}, N12_{t-4}, N12_{t-10}, N12_{t-16}, N12_{t-22}, N12_{t-28}, SOI_{t-12}, SOI_{t-13}, SOI_{t-14})$; CNN 6(II) model: $LSAR_t = f(N34_{t-17}, N34_{t-18}, N34_{t-19}, N34_{t-20}, N34_{t-21}, N34_{t-22}, N12_{t-4}, N12_{t-10}, N12_{t-16}, N12_{t-22}, N12_{t-28}, SOI_{t-12}, SOI_{t-13}, SOI_{t-14}, LANC_{t-20}, LANC_{t-21}, LANC_{t-31}, LANC_{t-32})$.

[a] Best result.

[b] The same independent variables that CNN 6(II).

[c] Distribution: Gamma; Link function: inverse.

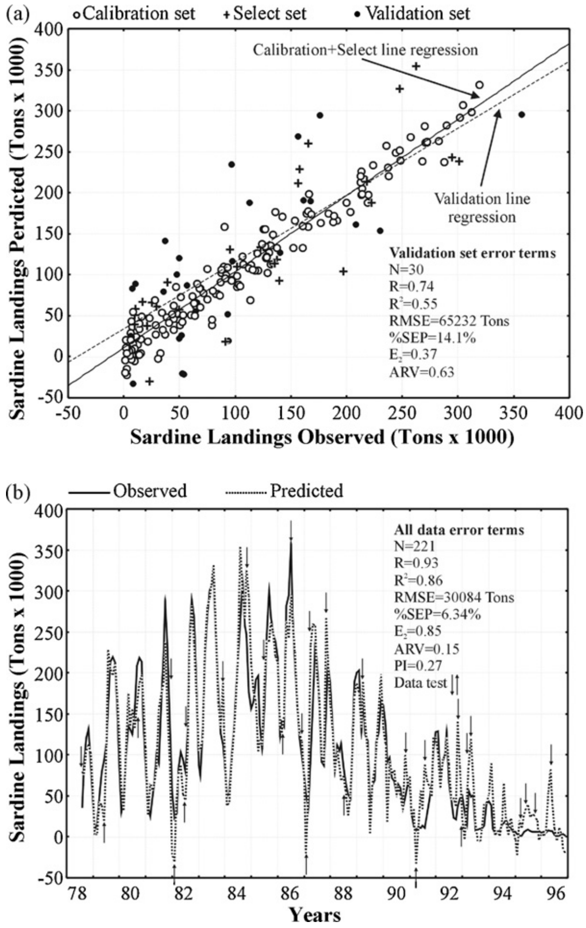

Fig. 2. CNN 6(I) prediction results. Input variables: N34$_{t-17}$, N34$_{t-18}$, N34$_{t-19}$, N34$_{t-20}$, N34$_{t-21}$, N34$_{t-22}$, N12$_{t-4}$, N12$_{t-10}$, N12$_{t-16}$, N12$_{t-22}$, N12$_{t-28}$, SOI$_{t-12}$, SOI$_{t-13}$, SOI$_{t-14}$: (a) scatter plot between observed and predicted sardine landings. The accuracy measures of validation data set are showed; (b) monthly observed and predicted sardine landings for calibration, select and validation sets.

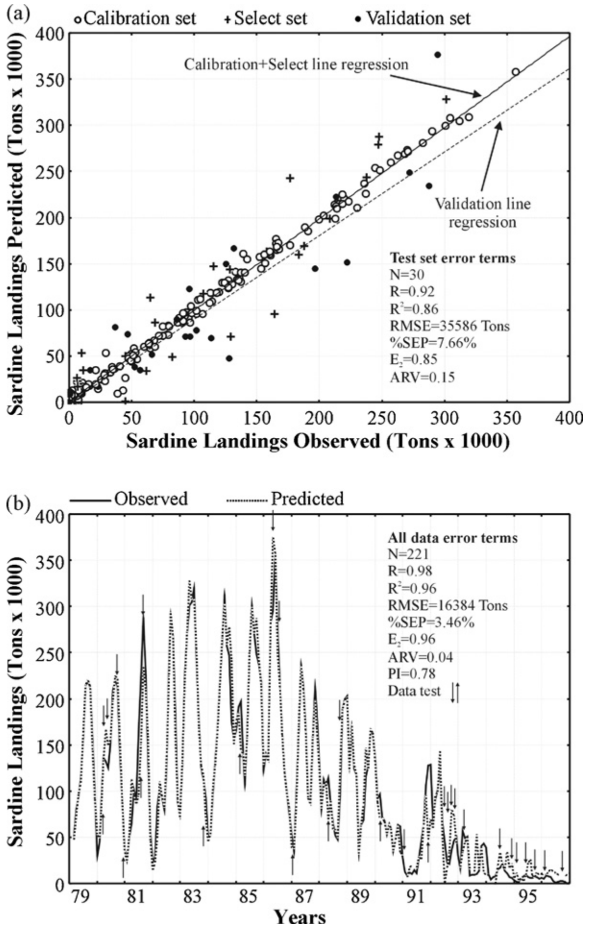

Fig. 3. CNN 6(II) prediction results. Input variables: N34$_{t-17}$, N34$_{t-18}$, N34$_{t-19}$, N34$_{t-20}$, N34$_{t-21}$, N34$_{t-22}$, N12$_{t-4}$, N12$_{t-10}$, N12$_{t-16}$, N12$_{t-22}$, N12$_{t-28}$, SOI$_{t-12}$, SOI$_{t-13}$, SOI$_{t-14}$, CANC$_{t-20}$, CANC$_{t-21}$, CANC$_{t-31}$, CANC$_{t-32}$: (a) scatter plot between observed and predicted sardine landings. The accuracy measures of validation data set are showed; (b) monthly observed and predicted sardine landings for calibration, select and validation sets.

explained ($R^2 = 0.86$). The standard error of prediction was slightly higher than 7% which supposed a RMSE value of 35,586 tons. Also, the persistence index was significantly improved (PI = 0.78) (Fig. 3b).

Globally, the response of all CNNs worsened when a detail analysis from 1990 to 1996 was carried out. In this period only the CNN 6(II) model provided satisfactory results. In this case, only the persistence index (PI) was improved in relation to the calculated for the naïve model. However, the remaining accuracy measures of CNN 6(II) for full data and validation sets were very close to the acceptance threshold (Table 5).

The CNN 6(II) sensitivity analysis indicated that nine of the input variables had ratio values higher than mean ratio (Table 6). Among these nine variables, eight were composed by lags of N34 or N12 and only one was a LANC lag (LANC$_{t-21}$). All SOI lags were considered less important by the model. In spite of this, the individual ratio levels indicated that to remove any variable would produce an important decrease of the model estimation capacity. Likewise, the sensitivity analysis showed that the most important variable was N34$_{t-20}$. However, the examination of the statistical functional relationships of N34$_{t-20}$ with sardine landings indicated that the inclusion of this variable in the model was only necessary to explain the sardine landings in presence of El Niño event. In absence of El Niño event, the effect of N34$_{t-20}$ on sardine landings was only significant when in austral winter months (July and August) the temperature in the Niño 3 + 4 area oscillated between 25.2 and

Table 6
CNN 6(II) sensitivity analysis. Ratio, ranking and mean ratio are shown. Bold indicates a variable with ratio higher than mean ratio. SST = sea surface temperature in Antofagasta; EK = Ekman transport; SL = sea level in Antofagasta; N34 = sea surface temperature in area Niño 3 + 4; N12 = sea surface temperature in area Niño1 + 2; SOI = south oscillation index; LANC = anchovy landings.

Variable	Ratio	Ranking
N34$_{t-20}$	**5.69**	**1**
N12$_{t-22}$	**5.22**	**2**
N34$_{t-18}$	**4.96**	**3**
N12$_{t-16}$	**4.89**	**4**
N12$_{t-28}$	**4.59**	**5**
N34$_{t-17}$	**4.59**	**6**
N12$_{t-4}$	**4.52**	**7**
LANC$_{t-21}$	**3.87**	**8**
N34$_{t-21}$	**3.85**	**9**
N12$_{t-10}$	3.79	10
N34$_{t-22}$	3.76	11
SOI$_{t-14}$	3.21	12
LANC$_{t-32}$	3.00	13
N34$_{t-19}$	2.89	14
LANC$_{t-20}$	2.77	15
LANC$_{t-31}$	2.72	16
SOI$_{t-13}$	2.18	17
SOI$_{t-12}$	2.14	18
Mean ratio	3.81	

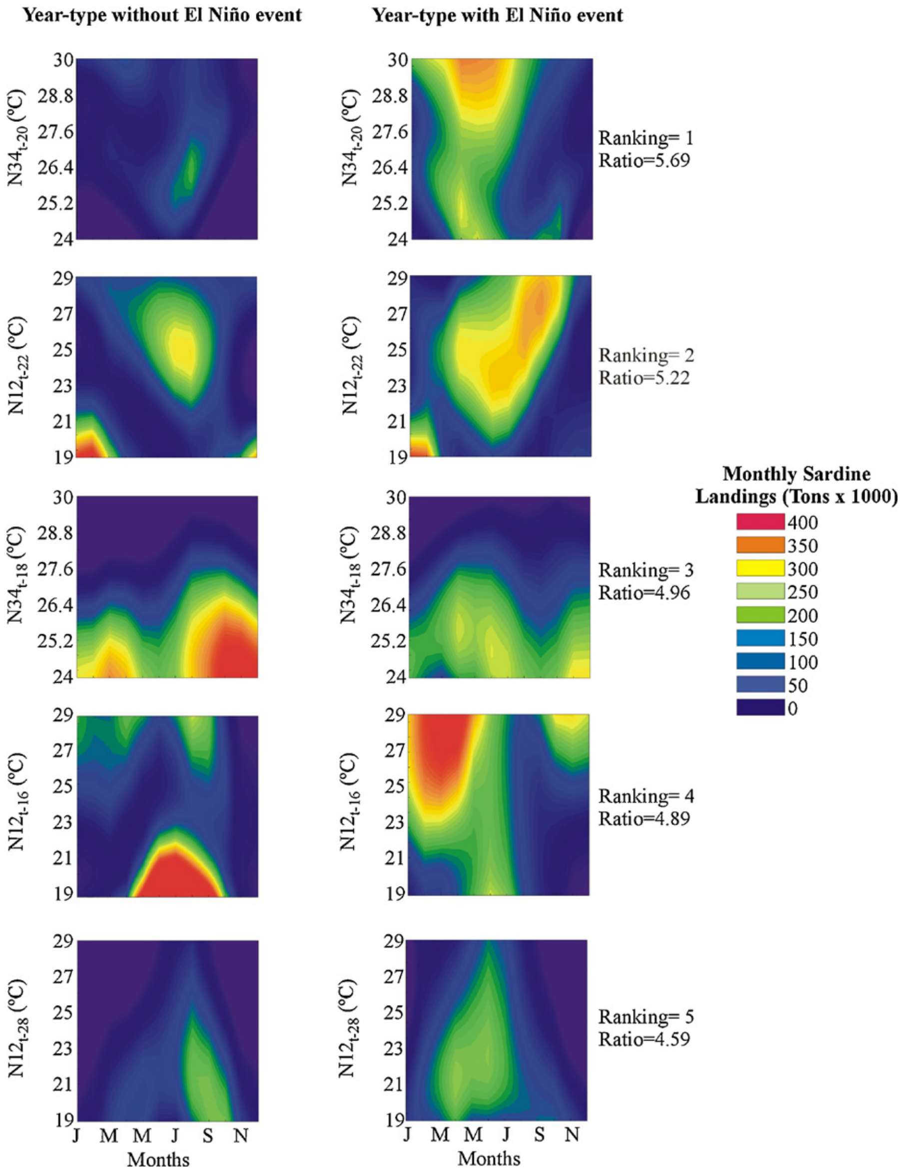

Fig. 4. CNN 6(II) statistical functional relationships between sardine landings and the input variables ranked in positions 1–5 for a year-type without and with El Niño event. A year-type with El Niño event was calculated averaging the functional relationships of the years 1982, 1983, 1987, 1990 and 1991.

27.6 °C (Fig. 4). Also, this effect was observed for $N34_{t-17}$ (ranking = 6; ratio = 4.59). On the other hand, the opposite effect was obtained for $N34_{t-18}$ (ranking = 3; ratio = 4.96). In this case, the hot focus was displaced toward the austral winter months in a year-type without El Niño event. Other variables as $N12_{t-16}$ had a significant effect in year-type with and without El Niño events, whereas variables as $LANC_{t-21}$ only showed a slight significant contribution in year-type with El Niño event in a very short-time period (April–June).

The CNN 6(II) results contrast with those obtained with the more classical approach using the same independent variables configuration and the same training and validation sets. MLR and GAM models had a fit capacity significant lower than CNNs (Table 5). Although in the validation the explained variance for both models ($N = 30$; $R^2(\text{MLR}) = 0.49$; $R^2(\text{GAM}) = 0.74$; $p < 0.01$) was statistically significant, their global behaviours on the full data and validation sets indicated that MLR and GAM models did not reach the minimum benchmark established by a naïve model in any accuracy measures. This lack of fit was very strong when only sardine landings predictions from 1990 to 1996 were considered (Table 5).

4. Discussion

There is no ideal form of fisheries resources management that is universally applicable. For this reason, it is fundamental a better understanding of the environment–resource relationships which allow to identify the sustainable exploitation alternatives and to evaluate the ecological consequences over different climatic conditions. Thus, this paper presents a study aimed at predicting sardine landings in the north area off Chile and at identifying conceptual analogies in trained and validated neural approaches in terms of the physics of the problem by the analyses of statistical functional relationships between environmental variables and sardine landings.

One of the significant steps in CNN models development has been the selection of an appropriate set of input variables from the available candidates. The strengths of the relationships between potential model inputs and outputs were examined using linear and non-linear cross-correlation functions and sensitivity analysis. The results indicated that the CNN models developed with the selected input variables performed well. The inclusion of additional variables as inputs did not increase the generalisation capabilities of the trained CNN models, as too Fernando et al. (2009) concluded, using the modified PMI algorithm (partial mutual information algorithm), to forecast water salinity in the River Murray (South Australia).

The results obtained with CNNs were clearly better than those obtained with MLR and GAM models. The GAM models performed better than the MLR models, which along with the results of the neural approaches, support the hypothesis pointed by Stenseth et al. (2002) about the stochasticity and high non-linear dependence between the abundance and ecosystem characteristics. Similar results can be found in other ecological applications that compare the prediction capacity of MLR with CNN (Lek-Ang et al., 1999; Gutiérrez-Estrada et al., 2008). Also, Schlink et al. (2003) reported that CNN models performed better than other types of models such as transfer function models, dynamic regressions implemented with Kalman filter, wavelet approaches, deterministic models and GAM models for predictions of short-time ozone concentration.

Cury et al. (2000), Boyer et al. (2001) and Hardmand-Mountford et al. (2003) indicated that although overfishing has played an important role in the decline of small pelagic stocks in different boundary current upwelling systems, the environmental variability is a key factor to explain the fluctuation of abundances of species as the Pacific sardine. Although the calibrated models in this study have been carried out on the base of the variation of sardine landings, according to Serra and Canales (2003) and Yáñez (2005) a strong correlation exists between sardine landings and sardine CPUE ($R = 0.95$). Therefore, the models developed in this paper not only explain a coupling between the landing level and environmental conditions, it would also be indicating that a strong relationship exists between sardine abundance fluctuation and the variation of environmental factors considered.

Among all CNNs calibrated, the best accuracy measures were obtained in those models that only considered local and global variables separately. In both cases the results were improved when the anchovy landings were added as models input. The effect of this variable was clearly significant on estimation capacity of CNN 6(I) model which did not achieve an explained variance level higher than 55% with the validation data set. In this way, the CNN 6(II) model explained the highest level of the variability with the validation data set (86%) and was the only model that had a correct behaviour after 1990. These results in relation to anchovy landings inclusion could be interpreted as: (a) classical biological processes such as inter-specific competence and predation would be regulation mechanisms which would play an important role in the modulation of sardine abundance changes; (b) inclusion of anchovy landings could have a proxy effect on the fit capacity of CNNs.

In relation to first interpretation, some authors have postulated the competition hypothesis for the complex sardine-anchovy (Pauly, 1987; Serra and Tsukayama, 1988; Ward et al., 2001) and other pelagic species (Cury and Fontana, 1988). Cubillos et al. (2002) analysing the monthly abundances of common sardine (*Strangomera bentincki*) and anchovy (*E. ringens*) in the central-southern area off Chile observed that when recruitment of common sardine tended to be low, the recruitment of anchovy tended to be higher, indicating that the dynamics of these two species depends on the yearly pulse of recruitment. These authors concluded that although the relationship between the recruitments of both species was not significant, one probable hypothesis that could explain the inverse fluctuation of the abundances of anchovy and common sardine would be the inter-specific competition.

On the other hand, the decrease of *S. sagax* landings associated with the setting-up of cold conditions from 1985 to 1986 (Yáñez et al., 2001) could be modulated by the increment of anchovy abundances. Santander et al. (1983) studying stomach contents of sardines from Peruvian coast indicated that 62% of stomachs analysed contained anchovy eggs. Also Miller et al. (2000) and Van der Lingen (2002) found that the main carbon contribution of the sardine diet in the California and Benguela ecosystems was from crustacean zooplankton and anchovy eggs. In this way, anchovy eggs could be an important energy source for the sardine of the Humboldt ecosystem which would explain the significant levels of sardine landings until 1992 in spite of that clear unfavourable environmental conditions for the sardine were observed 2 or 3 years before. This could to explain the weight of LANC_{t-31} on model CNN 6(II).

But, on the other hand, the inclusion of anchovy landings could have a proxy effect on the fit capacity of the CNN models. In this way, the fluctuation of anchovy landings could be a reflex of the variation of other environmental variables not considered in the model and even could be indicating the influence of different frequencies of the environmental variables already included. For example, the dynamics of anchovy and meso-zooplankton seem to be well correlated. Ayon and Guevara (2003) and Alheit and Ñiquen (2004) indicated that zooplankton volumes off central Peruvian coast decreased drastically around 1968. That is to say, 24 months previous to the drop of anchovy landings in 1970. In 1974, 2 years previous to the strong anchovy landings decrease, zooplanckton biomass decreased to very low levels and stayed low until the late-1980s.

The effects of remote physics forcing have been described for some clupeoid species. For example, Alheit and Ñiquen (2004) described minutely in three phases the events that cause the decrease, collapse and recovery of anchovy population of the Humboldt Current system. In a first phase, these authors indicated that around 1968–1971, the input of oceanic subtropical water masses to the Peruvian and Chilean coasts originated a series of biological and ecological events which caused the decrease of anchovy abundance. In the second phase, they highlight that the climate event called North Pacific regime shift occurred in winter 1976–1977 (Hare and Mantua, 2000) probably strengthened the processes described in the first phase which caused the anchovy biomass collapse. Likewise, these authors indicated that in the case of the sardine, the abundance changes cannot be a consequence of the climatic changes in this oceanic spatial scale. However, the generalisation capacity of the CNN 6(II) seems to indicate that a strong relationship exists between the sardine landings and the water temperature in the area Niño 3 + 4 (5°N–5°S, 120°W–170°W) 17–22 previous months to the sardine landings in the north area off Chile (18°S–24°S, 73°W), which is particularly significant during El Niño events.

Since an El Niño event lasts approximately 18–22 months, the results obtained allow us to suggest that the conditions of adult biomass and the reproductive success in a pre-Niño period is key factor to estimate accurately the landings 2 years later. However, in spite of the good fit of CNN 6(II), it is difficult to associate a causal effect with the temperature in the area Niño 3 + 4, a symptomatic effect being more probable. In this way, the same as LANC, N34 could be considered as a proxy variable.

In conclusion, the results of this paper support that the quantitative prediction of sardine landings can be carried out from environmental variables, although the predictive capacity of the CNNs is significantly improved if the anchovy landings are added as input variable. Also, various components of the relationships between the environmental variables and sardine landings could be identified. Additionally, it is possible to affirm that the difficulty of a good estimation is not only dependent on the nature of inter-relationships between environmental and fisheries variables and the spatial and temporal complexity of environmental effects but also on the type of relationship among variables assumed and the modeling procedures used.

On the other hand, although the good generalisation capability of the neural approaches are partially a consequence of the proposed method for the selection of an appropriate set of input variables from the available candidates (Bowden et al., 2005; May et al., 2008), it is necessary to remark that this method is even rudimentary and highly dependent of the human expert criterion, which is the reason why it should be perfected. In this sense, in future works the application of non-linear dependence measures, such as mutual information (Sharma, 2000; Fernando et al., 2009), and/or soft-computing technologies, such as evolutionary computing (Doglioni et al., 2008), should be proved in terms of the inputs identified as significant and the accuracy of the resulting models.

Acknowledgements

The authors wish to express their gratitude to the AECI (Agencia Española de Cooperación Internacional) for financing this research under Project ‘Heupes: Aplicación de métodos clásicos y heurísticos para la predicción de pesquerías pelágicas en la costa norte de Chile A/4857/06’.

References

Alheit, J., Bernal, P., 1993. Effects of physical and biological changes on the biomass yield of the Humboldt current system. In: Sherman, K., Alexander, L.M., Gold, B. (Eds.), Large Marine Ecosystem—Stress, Mitigation, and Sustainability. American Association for the Advancement of Science, pp. 53–68.

Alheit, J., Ñiquen, M., 2004. Regime shift in the Humboldt current ecosystem. Prog. Oceanogr. 60, 201–222.

Anctil, F., Rat, A., 2005. Evaluation of neural network streamflow forecasting on 47 watersheds. J. Hydrol. Eng. 10 (1), 85–88.

Arregui, I., Arrizabalaga, H., Kirby, D.S., Martín-González, J.M., 2006. Stock-environment-recruitment models for North Atlantic albacore (*Thunnus alalunga*). Fish. Oceanogr. 15 (5), 402–412.

Ayon, P., Guevara, R., 2003. Zooplankton trends off Peru between 1964 and 2002. In: Abstract, 3rd International Zooplankton Production Symposium, Gijón, Spain, May 20–23.

Bakun, A., 1973. Coastal upwelling indices, west coast of North America; 1967–73. U.S. Dept. NOAA Tech. Rep. NMFS SSRF-693. 114 pp.

Bakun, A., Broad, K., 2001. Climate and fisheries: interacting paradigms, scales, and policy approaches. In: IRI-IPRC Pacific Climate-Fisheries Workshop, Honolulu, November 14–17, p. 67.

Bernal, P., Robles, F., Rojas, O., 1983. Variabilidad física y biológica en la región meridional del sistema de corrientes Chile-Perú. FAO Fish. Rep. 291 (3), 683–711.

Bertrand, A., Segura, M., Gutiérrez, M., Vásquez, L., 2004. From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off Peru. Fish Fish. 5, 296–316.

Bowden, G.J., Dandy, G.C., Maier, H.R., 2005. Input determination for neural network models in water resources applications. Part 1. Background and methodology. J. Hydrol. 301, 75–92.

Boyer, D.C., Boyer, H.J., Fossen, I., Kreiner, A., 2001. Changes in abundance of the northern Benguela sardine stock during the decade 1990 to 2000, including a discussion of the relative importance of fishing and the environment. S. Afr. J. Mar. Sci. 23, 67–84.

Butterworth, D.S., 1983. Assessment and management of pelagic stocks in the southern Benguela region. FAO Fish. Rep. 291 (2), 329–405.

Chávez, F., Ryan, J., Lluch-Cota, S., Ñiquen, M., 2003. From anchovies to sardines and back: multidecadal change in the Pacific Ocean. Science 299, 217–221.

Chávez, F.P., Pennington, J.T., Castro, C.G., Ryan, J.P., Michiaski, R.P., Schlining, B., Walz, P., Buck, K.R., McFadyen, A., Collins, C.A., 2002. Biological and chemical consequences of the 1997–1998 El Niño in central California waters. Prog. Oceanogr. 54, 205–232.

Cubillos, L.A., Bucarey, D.A., Canales, M., 2002. Monthly abundance estimation for common sardine *Strangomera bentincki* and anchovy *Engraulis ringens* in the central-southern area off Chile (34–40°S). Fish. Res. 57, 117–130.

Cury, P., Fontana, A., 1988. Compétition et strategies démographiques compares de dues espèces de sardinelles (*Sardinella aurita* et *Sardinela maderensis*) des côtes ouest-africaines. Aquat. Living Resour. 1, 165–180.

Cury, P., Bakun, A., Crawford, R.J.M., Jarre, A., Quiñones, R.A., Shannon, L.J., Verheye, H.M., 2000. Small pelagic in upwelling systems: patterns of interaction and structural changes in “wasp-waist” ecosystems. ICES J. Mar. Sci. 57, 603–618.

Doglioni, A., Giustolisi, O., Savic, D.A., Webb, B.W., 2008. An investigation on stream temperature analysis based on evolutionary computing. Hydrol. Process. 22 (3), 315–326.

Elgaali, E., García, L.A., 2007. Using neural networks to model the impacts of climate change on water supplies. J. Water Resour. Plan. Manage. 133 (3), 230–243.

Fernando, T.M.K.G., Maier, H.R., Dandy, G.C., 2009. Selection of input variables for data driven models: an average shifted histogram partial mutual information estimator approach. J. Hydrol. 367, 165–176.

Fréon, P., Mullon, C., Voisin, B., 2003. Investigating remote synchronous patterns in fisheries. Fish. Oceanogr. 12 (4/5), 443–457.

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. Fish. Res. 86, 188–2000.

Gutiérrez-Estrada, J.C., Vasconcelos, R., Costa, M.J., 2008. Estimating fish community diversity from environmental features in the Tagus estuary (Portugal): multiple linear regression and artificial neural network approaches. J. Appl. Ichthyol. 24, 150–162.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis, Fifth ed. Prentice Hall International, Inc., London, United Kingdom.

Hardmand-Mountford, N.J., Richardson, A.J., Boyer, D.C., Kreiner, A., Boyer, H.J., 2003. Relating sardine recruitment in the Northern Benguela to satellite-derived sea surface height using a neural network pattern recognition approach. Prog. Oceanogr. 59, 241–255.

Hare, S.R., Mantua, N.J., 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. Prog. Oceanogr. 47, 103–145.

Hare, S.R., Minobe, S., Wooster, W.S., McKinnell, S., 2000. An introduction to the PICES symposium on the nature and impacts of North Pacific climate regime shifts. Prog. Oceanogr. 47, 99–102.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall/CRC, Boca Raton.

Hormazabal, S., Shaffer, G., Leth, O., 2004. The coastal transition zone off Chile. J. Geophys. Res. 109, C01021, doi:10.1029/2003JC001956.

Hormazabal, S., Shaffer, G., Pizarro, O., 2002. Tropical Pacific control of intraseasonal oscillations off Chile by way of oceanic and atmospheric pathways. Geophys. Res. Lett. 29 (6), doi:10.1029/2001GL013481.

Hunter, A., Kennedy, L., Henry, J., Ferguson, I., 2000. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. Comput. Meth. Prog. Bio. 62, 11–19.

Huse, G., Gjøsæter, H., 1999. A neural network approach for predicting stock abundance of the Barents Sea capelin. Sarsia 84 (5–6), 457–464.

Huse, G., Ottersen, G., 2003. Forecasting recruitment and stock biomass of Northeast Artic cod using neural networks. Sci. Mar. 67, 325–335.

Hyun, K., Song, M.Y., Kim, S., Chon, T.S., 2005. Using an artificial neural network to patternize long-term fisheries data from South Korea. Aquat. Sci. 67, 382–389.

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

Jain, A, Kumar, S., in press. Dissection of trained neural network hydrologic model architectures for knowledge extraction. Water Resour. Res. doi:10.1029/2008WR007194.

James, F.C., McCulloch, C.E., 1990. Multivariate analysis in ecology and systematic: panacea or Pandora’s box? Annu. Rev. Ecol. Syst. 21, 129–166.

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

Komatsu, T., Aoki, I., Mitani, I., Ishii, T., 1994. Prediction of the catch of Japanese sardine larvae in sagami bay using a neural-network. Fish. Sci. 60 (4), 385–391.

Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of ‘goodness-of-fit’ measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241.

Lek-Ang, S., Deharveng, L., Lek, S., 1999. Predictive models of collembolan diversity and abundance in a riparian habitat. Ecol. Model. 120, 247–260.

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

Lluch-Belda, D., Crawford, R.J.M., Kawasaki, T., MacCall, A.D., Parrish, R.H., Schawarzlose, R.A., Smith, P.E., 1989. Worldwide fluctuations of sardine and anchovy stocks: the regime problema. S. Afr. J. Mar. Sci. 8, 195–205.

Lluch-Cota, D.B., Hernández-Vázquez, S., Lluch-Cota, S.E., 1997. Empirical investigation on the relationship between climate and small pelagic global regimes and El Niño-southern oscillation (ENSO). FAO Fish. Circ. 934, 1–48.

May, R.J., Maier, H.R., Dandy, G.C., Fernando, T.M.K.G., 2008. Non-linear variable selection for artificial neural networks using partial mutual information. Environ. Model. Softw. 23 (10–11), 1312–1326.

McLaughlin, R.L., 1983. Forecasting models: sophisticated or naïve? J. Forecasting 2 (3), 274–276.

Miller, T.W., Brodeur, R.D., Emmett, R.L., 2000. Distribution and diet of the Pacific sardine (*Sardinops sagax*) off Oregon and Washington. In: Program and Abstract of the CalCOFI Annual Conference, Lake Arowhead, CA, October.

Montecinos, A., Purca, S., Pizarro, O., 2003. Interannual-to-interdecadal sea surface temperature variability along the western coast of South America. Geophys. Res. Lett. 30, 1570, doi:10.1029/2001GL013717.

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. J. Hydrol. 10, 282–290.

Pauly, D., 1987. In: Pauly, D., Tsukayama, I. (Eds.), Managing the Peruvian upwelling ecosystem: a synthesis. The Peruvian anchoveta and its upwelling ecosystem: three decades of change. ICLARM Stud. Rev. 15, 351.

Pitcher, T.J., Hart, P.J.B. (Eds.), 1982. Fisheries Ecology. Chapmand and Hall, p. 416.

Pizarro, O., Shaffer, G., Dewitte, B., Ramos, M., 2002. Dynamics of seasonal and interannual variability of the Peru–Chile undercurrent. Geophys. Res. Lett. 29 (12), doi:10.1029/2002GL014790.

Pulido-Calvo, I., Portela, M.M., 2007. Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds. J. Hydrol. 332 (1–2), 1–15.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. ‘Learning’ representations by back-propagation errors. Nature 323, 533–536.

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 Fish. Rep. 291 (3), 1011–1025.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M., Doyle, M., 2003. A rigorous inter-comparison of ground-level ozone predictions. Atmos. Environ. 37, 3237–3253.

Schwartzlose, R.A., Alheit, J., Bakun, A., Baumgartner, T., Cloete, R., Crawford, R.J.M., Fletcher, W.J., Green-Ruiz, Y., Hagen, E., Kawasaki, T., Lluch-Belda, D., Lluch-Cota, S.E., MacCall, A.D., Matsuura, Y., Nevares-Martínez, M.O., Parrish, R.H., Roy, C., Serra, R., Shust, K.V., Ward, N.M., Zuzunaga, J.Z., 1999. Worldwide large scale fluctuations of sardine and anchovy populations. S. Afr. J. Mar. Sci. 21, 289–347.

SERNAPESCA, 1978–1997. Anuarios Estadísticos de Pesca. Servicio Nacional de Pesca, Ministerio de Economía, Fomento y Reconstrucción, Chile.

Serra, R., Canales, C., 2003. Evaluación del stock de anchoveta y sardina española, 2003 (I y II región). Informe Final IFOP-SUBPESCA.

Serra, R., Tsukayama, I., 1988. Sinopsis de datos biológicos y pesqueros de *Sardinops sagax* en el Pacífico suroriental. FAO Sinop. Pesca 13 (1), 60.

Sharma, A., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply. Part 1. A strategy for system predictor identification. J. Hydrol. 239, 232–239.

Sharp, G.D., McLain, D.R., 1993. Fisheries, El Niño-Southern Oscillation and upper-ocean temperature records: an eastern Pacific example. Oceanography 6, 13–22.

Shepherd, A.J., 1997. Second-Order Methods for Neural Networks. Springer, New York.

Shrestha, R.R., Theobald, S., Nestmann, F., 2005. Simulation of flood flow in a river system using artificial neural networks. Hydrol. Earth Syst. Sci. 9 (4), 313–321.

Steele, J.H., 1985. Comparison of marine and terrestrial ecological systems. Nature 313, 355–358.

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. Popul. Ecol. 44, 165–178.

Tan, Y., van Cauwenberghe, A., 1999. Neural-network-based *d*-step-ahead predictors for nonlinear systems with time delay. Eng. Appl. Artif. Intel. 12 (1), 21–25.

Tsoukalas, L.H., Uhrig, R.E., 1997. Fuzzy and Neural Approaches in Engineering. Wiley Interscience, New York, 587 pp.

Van der Lingen, C.D., 2002. Diet of sardine *Sardinops sagax* in the southern Benguela upwelling ecosystem. S. Afr. J. Mar. Sci. 24, 301–316.

Ventura, S., Silva, M., Pérez-Bendito, D., Hervás, C., 1995. Artificial neural networks for estimation of kinetic analytical parameters. Anal. Chem. 67 (9), 1521–1525.

Ward, T.M., Hoedt, F., McLeay, L., Dimmlich, W.F., Jackson, G., Rogers, P.J., Jones, K., 2001. Have recent mass mortalities of the sardine *Sardinops sagax* facilitated an expansion in the distribution and abundance of the anchovy *Engraulis australis* in South Australia? Mar. Ecol. Prog. Ser. 220, 241–251.

Yáñez, E., 2005. Análisis integrado histórico ambiente-recursos, I–II regiones. Informe Final FIP 2003-33.

Yáñez, E., Barbieri, M.A., Silva, C., Nieto, K., Espíndola, F., 2001. Climate variability and pelagic fisheries in northern Chile. Prog. Oceanogr. 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. 109–121.