Available online at www.sciencedirect.com
ScienceDirect
Fisheries Research 86 (2007) 120–128
FISHERIES RESEARCH
www.elsevier.com/locate/fishres

Short-term forecasting of halibut CPUE: Linear and non-linear univariate approaches

Ivone Alejandra Czerwinski [a], Juan Carlos Gutiérrez-Estrada [b,*], José Antonio Hernando-Casal [a]

[a] *Dpto. Biología, Facultad de Ciencias del Mar y Ambientales, Universidad de Cádiz, Campus Río San Pedro, 11510 Puerto Real, Cádiz, Spain*
[b] *Dpto. de Ciencias Agroforestales, Escuela Politécnica Superior, Universidad de Huelva, Campus Universitario de La Rábida, 21819 Palos de la Frontera, Huelva, Spain*

Received 15 November 2006; received in revised form 11 April 2007; accepted 16 May 2007

Abstract

In the present paper, two univariate forecasting techniques were tested to evaluate the short-term CPUE capacity forecast for Pacific halibut, *Hippoglossus stenolepis* (Pleuronectidae). The first methodology, based on the Box–Jenkins approach (autoregressive integrated moving average models [ARIMA models]), assumes a linear relationship between the time series data. The second methodology, using artificial neural network models (ANNs), enables highly non-linear processes to be modelled. The best results from a seasonal ARIMA model indicated that one non-seasonal autoregressive term combined with a non-seasonal moving average term and a seasonal moving average term was suitable to explain a variance level of 32.6% in the validation phase, providing statistically acceptable but insufficiently satisfactory estimations. The configuration of the best ANN model (three autoregressive terms in the input layer and five neurons in the hidden layer) provided a significant improvement in the independent validation phase (91% of the variation explained), indicating a clear non-linear relationship between variables. Modelling of the abundance indices is a useful tool for understanding the dynamics of populations and may enable short-term quantitative recommendations for fisheries management to be made.

Keywords: Prediction; ARIMA models; Census X12; Artificial neural network (ANN); Time series; *Hippoglossus stenolepis*

1. Introduction

Considering the economic importance of the fisheries, it is essential to know as much as possible about fish populations, which can undergo many changes in response to commercial fishing. These include changes in abundance, total biomass, size-frequency distributions, age-structure and spatial distributions. A wide variety of mathematical and statistical methods are available to fisheries science for characterising these processes in order to understand the dynamics of exploited populations. The underlying assumption is that, if we can understand the response of fish populations to different perturbations, then we should be able to manage the fisheries in line with our chosen objectives (Haddom, 2001).

The Pacific halibut fishery is a good example of an exploited population. Pacific halibut were first harvested in Alaskan waters during the 1880s (Rigby et al., 1995). The Alaskan Pacific halibut fishery generates considerable value, and commercial landings for the year 2004 were almost 33 million kg, valued at 234 USD million (€182 million) (Low and Stocker, 2005).

The current health of this fishery is attested by the fact that some the highest landings on record were taken in the latter half of the 1990s. Since 1995, the fishery has been managed under an Individual Transferable Quota System (ITQS) (Clark and Hare, 2004). The International Pacific Halibut Commission (IPHC) is responsible for management of Pacific halibut fisheries for optimum yield. Each year the IPHC estimates the exploitable biomass (total Pacific halibut biomass available to the fishery), recruitment and other statistics for the entire North Pacific halibut stock. Harvest quotas are based on a strategy of constant annual exploitation rate (currently 20%). Generally, age-at-catch

* Corresponding author. Tel.: +34 959 217528; fax: +34 959 217528.
E-mail address: juanc@uhu.es (J.C. Gutiérrez-Estrada).

0165-7836/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2007.05.006

information has not been available from recreational fisheries, but in this case the sport harvest has been included as part of total removals, from 1977 to 2005 (Geernaert et al., 2005).

In spite of the economic importance of this fishery, prediction models for long-term biomass or CPUE are scarce, except for some studies such as Quinn (1985), Zhang et al. (1991), Tanaka (2000) and Clark and Hare (2004).

Forecasting biomass available for a fishery is an extremely relevant topic, because it plays a central role in management of stocks, preceding decision making (Makridakis et al., 1983). In fisheries management policy, the main goal is to establish the maximum fishing effort applicable in a defined area during a known period to keep the stock replacement rate stable. To achieve this aim, it is necessary to predict the effect of uncontrollable events on abundance. Changes in abundance can be forecast if quantitative data are available on the past catch, and if the "assumption of continuity" is accepted: that is to say, if we assume that some features of the past pattern will continue into the future (Makridakis et al., 1983; Stergiou et al., 1997). Apart from methods based on biological principles, a variety of statistical techniques have also been used in fisheries forecasting. These methods are directed towards: (a) modelling on the basis of deterministic, regression techniques that explain changes in fishery variables in terms of changes in various biotic and/or abiotic variables; (b) modelling on the basis of univariate time series techniques that treat the system as a "black-box", i.e. viewed as an unknown generating process and (c) methods that synthesise the two above mentioned general approaches (Stergiou et al., 1997).

Among the methods based on univariate techniques, the ARIMA models by Box and Jenkins (1976) stand out because of their wide range of application. These statistical models assume linear dependence between the time series data. Thus, each observation can be explained as a linear function of its past values, but with some errors. The variability found in the results from applying ARIMA models is due mostly to the fact that they are linear univariate models (the linear relationship between variables is assumed). ARIMA models have the same properties as the most simple time series models like moving average (MA) and autoregressive (AR) models. They can include cyclic/seasonal components and their mathematics is not excessively complex. These properties have favoured the application of this kind of model to predict different variables in numerous fields of engineering and the sciences, including fishery science (Phillips, 1983; Stocker and Noakes, 1988; Stergiou, 1989, 1990, 1991; Stergiou et al., 1997; Lloret et al., 2000; Becerra-Muñoz et al., 2003; Hänninen et al., 2003; Punzón et al., 2004; Gutiérrez-Estrada et al., 2004).

Stenseth et al. (2002) reported that ecological systems (like fisheries) may be heavily influenced by substantial stochastic processes (as climate conditions), as well as by non-linearities, leading to non-equilibrium systems as the results. Thus, Lek et al. (1996a,b), Guégan et al. (1998), Gozlan et al. (1999) and Gutiérrez-Estrada et al. (2000) propose a non-linear approach to the forecasting of ecological–biological variables. Therefore, these authors assume that the variables considered in the ecological systems studied are related in a non-linear way.

Significant progress in the fields of non-linear pattern recognition and system control theory have recently been made possible through advances in a branch of non-linear system theoretical modelling called artificial neural networks (ANNs) (Lek et al., 1996b). In recent years, ANNs have attracted increasing attention from both academic researchers and industrial practitioners. Basically, the reason for their popularity is their powerful pattern recognition and flexible non-linear modelling capacity (Qi and Zhang, 2001). ANNs have a great capacity to fit highly scattered data, far from normality, and produce powerful models from few data, thus providing reliable predictions (Govindajaru, 2000).

In this paper, the principal aim is evaluate and compare the ARIMA and ANN models for Pacific halibut CPUE forecasting. For this we analyse the general characteristics of fishing time series and identify the possible sources of error in the models.

2. Material and methods

2.1. Study area and data source

The Pacific halibut catches data are from the data base of the IPHC available in http://www.iphc.washington.edu/halcom/. The data show daily catches during the fishing period (May to September) from 1998 to 2003 in the IPHC regulation area (Bering Sea and Gulf of Alaska). The gear consisted of fixed hook, 45 m skates with 16/0 circle hooks spaced 45 cm apart. This standard skate is used to calculate the CPUE.

The seasonal character of this fishery implies a minimum sample size problem. Since during 8.5 months per year there is no fishing activity, most of the time series data have no information about CPUE. As a compromise, considering length of the time series (the number of data needed to calibrate and validate the models) and the prediction unit (one-step period prediction), a transformation of the time series is proposed. The daily time series data were effectively smoothed by grouping values in 23 sets of 5 consecutive days, over the same fishing period (from May 29 to September 16).

In both cases (ARIMA and ANN models), the calibration set was composed by the first 107 patterns (approximately 77% of total patterns) and the validation set was composed by the last 31 patterns (approximately 23% of the total records).

2.2. ARIMA models

ARIMA(p,d,q) models assume that a time series is a linear combination of its own past values and current and past values of an error term. The mathematical model can then be written as follows:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots + \phi_p w_{t-p} + \theta_0 - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + \varepsilon_t \quad (1)$$

where w_t is the original data series or difference of degree d of the original data at time t, a_t the random error, innovation or shock at time t, $\phi_1, \phi_2, \ldots, \phi_p$ the autoregressive parameters, p the autoregressive order, θ_0 a constant term, $\theta_1, \theta_2, \ldots, \theta_q$ the moving average parameter, q the moving average order and ε_t is

the white noise (Box and Jenkins, 1976). ARIMA models can also adjust for seasonality in the data, in which case the model is denoted by ARIMA(p,d,q) $(P,D,Q)^S$, where P is the seasonal autoregressive order, Q the seasonal moving average order, D the seasonal differencing order and S is the seasonal order.

Extraction of the lag orders (p, q, P, Q and S) of the time series can be done using the autocorrelation (ACF) and partial autocorrelation function (PCAF) (Holton-Winson and Keating, 1996). The values of p, d and q that proved to be more appropriate according to the accuracy measures presented in model selection criteria section were then used. The parameters ϕ and θ were fixed by using function minimization procedures, so that the sum of squared residuals was minimized. The level of significance of these parameters should be evaluated (acceptable if $p < 0.05$). For the seasonal orders, values of P and Q varying from one to three (with a unitary step) were proved. A seasonal differencing order $D = 1$ was considered because the first differencing was enough to eliminate stochastic seasonal non-stationarity. Also, Census X12 method was applied in order to extract trend and seasonal patterns (Yaffee and McGee, 2000). The same lag component was used with ANN models.

2.3. Artificial neural network models

Artificial neural networks are mathematical models inspired by the neural architecture of the human brain. The model neurone or node is a simple non-linear unit. The neurones collect inputs from single or multiple sources and produce an output. Interconnecting many of these single neurones or nodes in a known layer configuration creates a model neural network.

Each node j receives incoming signals from every node i in the previous layer. Associated with each incoming signal (x_i) is a weight (W_{ji}). The effective incoming signal (I_j) to node j is the weighted sum of all the incoming signals:

$$I_j = \sum_{i=1}^{q} x_i W_{ji} \tag{2}$$

The effective incoming signal, I_j, is passed through an activation function (sometimes called a transfer function) to produce the outgoing signal (y_j) of the node j. In this study, the linear function ($y_j = I_j$) will be used in the output layer and the sigmoid non-linear function will be used in the hidden layers:

$$y_j = f(I_j) = \frac{1}{1 + e^{-I_j}} \tag{3}$$

A typical ANN structure with two hidden layers is denoted by—I:FHs:SHs:Ol, where I is the number of nodes or neurons in the input layer, FH and SH the number of nodes or neurons in the first and second hidden layers, respectively, O the number of neurons in the output layer, s denotes the sigmoid transfer function and l indicates the linear transfer function (Fig. 1).

The training method used was the Levenberg–Marquardt algorithm (Shepherd, 1997) controlled by the method of internal validation (about 27% of calibration data to test the error at the end of each epoch) (Tsoukalas and Uhrig, 1997). The weights are updated at the end of each epoch.

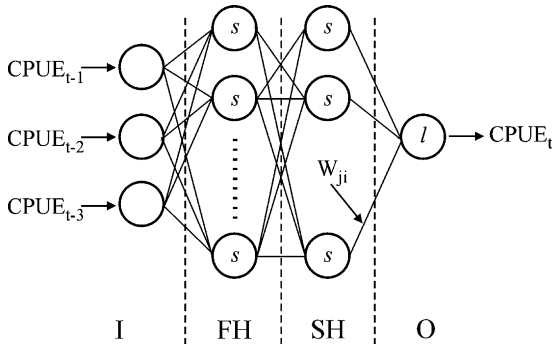

Fig. 1. Schematic representation of a four-layer feed forward artificial neural network for computing halibut CPUE.

The number of hidden layers and nodes in the hidden layers were determined by trial and error. ANNs with one hidden layer and three to nine hidden nodes were successively trained based on the calibration data set. The ANN having the best performance when applied to the validation set, within a pool of five repetitions, was selected (Anctil and Rat, 2005). The ANN models were implemented using STATISTICA 6.0.

2.4. Model selection criteria

There are many measures of forecasting accuracy that one may use to compare different models (Legates and McCabe, 1999; Abrahart and See, 2000). The correlation between observed and predicted CPUE was expressed by means of the correlation coefficient R. The coefficient of determination (R^2) describes the proportion of the total variance in the observed data that can be explained by the model. Others measures of variances applied were the percent standard error of prediction (%SEP) (Ventura et al., 1995), the coefficient of efficiency (E_2) (Nash and Sutcliffe, 1970; Kitanidis and Bras, 1980) and the average relative variance (ARV) (Griñó, 1992). These four estimators are unbiased estimators that are employed to see how far the model is able to explain the total variance of the data.

In addition, it is advisable to quantify the error in the same units as the variables. These measures, or absolute error measures, included the root of the mean square error (RMSE) and the mean absolute error (MAE), given by:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{CPUE}_t - \widehat{\mathrm{CPUE}_t})^2}{N}},$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}|\mathrm{CPUE}_t - \widehat{\mathrm{CPUE}_t}|}{N} \tag{4}$$

where CPUE_t is the observed Pacific halibut CPUE at the time step t, $\widehat{\mathrm{CPUE}_t}$ the estimated Pacific halibut CPUE at the same time step t and N is the total number of observations of the validation set.

The percent standard error of prediction, %SEP, is defined by:

$$\%\mathrm{SEP} = \frac{100}{\overline{\mathrm{CPUE}}}\mathrm{RMSE} \tag{5}$$

where $\overline{\text{CPUE}}$ is the average of the observed Pacific halibut CPUE of the validation set. The principal advantage of %SEP is its non-dimensionality, which allows forecasts given by different models to be compared on the same basis. The coefficient of efficiency E_2 and the average relative variance ARV are used to see how the model explains the total variance of the data and represent the 'proportion' of the variation of the observed data considered by the model. E_2 and ARV are given by:

$$E_2 = 1 - \frac{\sum_{i=1}^{N}(\text{CPUE}_t - \widehat{\text{CPUE}_t})^2}{\sum_{i=1}^{N}(\text{CPUE}_t - \overline{\text{CPUE}_t})^2}, \qquad \text{ARV} = 1 - E_2 \tag{6}$$

The sensitivity to outliers due to the squaring of the difference terms is associated with E_2 or, equivalently, with ARV. A value of zero for E_2 indicates that the observed average $\overline{\text{CPUE}}$ is as good predictor as the model, while negative values indicate that the observed average is a better predictor than the model (Legates and McCabe, 1999).

For a perfect match, the values of R^2 and of E_2 should be close to one and those of %SEP and ARV close to zero.

Also the persistence index, PI, was used for the model performance evaluation (Kitanidis and Bras, 1980):

$$\text{PI} = 1 - \frac{\sum_{i=1}^{N}(\text{CPUE}_t - \widehat{\text{CPUE}_t})^2}{\sum_{i=1}^{N}(\text{CPUE}_t - \text{CPUE}_{t-L})^2} \tag{7}$$

where CPUE_{t-L} is the observed CPUE at the time step $t-L$ and L is the lead-time. In the applications carried out, L was set equal to one, since a set of 5 days ahead forecasts were performed. A PI value of one reflects a perfect correspondence between predicted and observed values, and a value of zero is equivalent to saying that the model is no better than a 'naïve' model, which always gives as prediction the previous observation.

A negative PI value would mean that the model is degrading the original information, thus denoting a performance worse than the one of the naïve model (Anctil and Rat, 2005).

For each measure of accuracy the benchmark of the worst permissible error was calculated. McLaughlin (1983) suggests that a naïve model determine the forecasting accuracy benchmark of any model. The basic naïve model, known as 'Naïve Forecast I' (NFI) is defined as the next period's level will be the same as that of the preceding period. This way, if the forecasting model cannot do better than NFI, it should be disqualified.

3. Results

3.1. ARIMA models

The CPUE data show an overall negative long-term trend by comparing the year averages (least significant difference test [LSD test], $p < 0.05$). Therefore, differencing (modelling the change in CPUE from set to set rather than the CPUE series itself) was necessary. One degree of the non-seasonal differencing factor was enough to eliminate the trend. It was not necessary to log-transform the data before fitting the models, as the raw data were normally distributed and had no significant scattering (Shapiro–Wilk test: $W = 0.988$, $p = 0.894$; Levene test = 0.726, $p < 0.05$).

Table 1 shows the error terms of ARIMA models $(p,d,q)(P,D,Q)^S$ (with $S = 23$) in the validation phase ($N = 31$). The explained variance level was under 40%, reaching a maximum level of 39.8% ($R^2 = 0.398$; ARIMA(1 1 1) $(1\ 1\ 1)^{23}$) and a minimum level of 9% ($R^2 = 0.099$; ARIMA(1 1 0) $(0\ 0\ 1)^{23}$). In spite of the low values of the determination coefficient, the percent standard errors prediction was around 50%, except for the ARIMA models (0 0 1) $(0\ 0\ 1)^{23}$, (1 0 0) $(0\ 0\ 1)^{23}$ and (1 0 0) $(1\ 0\ 0)^{23}$. Globally, different behaviour is observed in function of the seasonal differentiation degree (order D). Thus the ARIMA models with order $D = 0$ provided the worst estimates for the validation data series. Individually, the best estimates in the validation process were obtained with ARIMA model (1 1 1) $(0\ 1\ 1)^{23}$. This model had the smallest error magnitudes in five of the seven error terms considered. In this case, the determination coefficient (R^2) as well as the medium absolute error (MAE) were slightly worse than that obtained for the ARIMA model (1 1 1) $(1\ 1\ 1)^{23}$.

With respect to the persistence index (PI), except in four cases, the PI values were very close to zero, with a mean value of -0.15 ± 0.7, which means a naïve overall behaviour. A PI value closer to one was observed for the ARIMA model (1 1 1) $(0\ 1\ 1)^{23}$ (PI = 0.428), which was significantly better than the second best result (ARIMA model (1 1 0) $(0\ 1\ 1)^{23}$; PI = 0.361).

Table 2 shows the estimators calculated for the two best models. In both cases the autoregressive parameters were non-significant. However, in the autocorrelation analysis of the residuals, none of the 24 autocorrelation coefficients was statistically significant, indicating that the residuals time series can be considered as similar to white noise (Fig. 2).

Results slightly worst were obtained when the ARIMA model were calibrated with Census X12 method. The best estimation was obtained with ARIMA(1,1,1) $(0,1,1)^{23}$. In this case, the explained variance in the independent validation phase ($R^2 = 0.12$) was lower than obtained with ARIMA classical calibration method. Also, worst values of the rest of accuracy measures were obtained (Fig. 3).

3.2. ANN models

The same cases used for validation of the ARIMA models were considered with the ANNs. Different network architectures, meaning different numbers of nodes in the hidden layer, have been tested. Table 3 shows the goodness of fit for all the trained ANN models. The proportion of variance explained was more than 37%, reaching a maximum proportion of 91% ($R^2 = 0.910$; ANN 3:5*s*:1*l*). Globally, the percent standard error of prediction (%SEP) was around 16% which implied CPUE mean variations of 14 kg/standard skate. Moreover, the goodness-of-fit of the models was corroborated by the values obtained for E_2 and ARV terms in the validation phase. However the PI term results indicated, in some cases, that the model

Table 1
Results of ARIMA models in the independent validation phase

Model	Error terms						
	R^2	RMSE	MAE	%SEP	E_2	ARV	PI
NFI	0.472	53.009	37.316	50.120	−0.056	1.056	0
$(0\,0\,1)\,(0\,0\,1)^{23}$	0.195	93.177	82.197	103.888	−3.316	4.316	−2.612
$(0\,0\,1)\,(0\,1\,1)^{23}$	0.301	44.687	37.071	49.824	0.007	0.993	0.169
$(0\,0\,1)\,(1\,1\,1)^{23}$	0.291	45.736	37.969	50.994	−0.040	1.040	0.130
$(0\,0\,1)\,(1\,1\,0)^{23}$	0.154	45.379	34.127	50.595	−0.024	1.024	0.143
$(0\,0\,1)\,(1\,0\,0)^{23}$	0.307	61.162	48.738	68.193	−0.859	1.859	−0.556
$(0\,1\,1)\,(0\,0\,1)^{23}$	0.097	50.568	40.962	56.381	−0.271	1.271	−0.064
$(0\,1\,1)\,(0\,1\,1)^{23}$	0.302	42.664	35.429	47.569	0.095	0.905	0.243
$(0\,1\,1)\,(1\,1\,1)^{23}$	0.312	41.819	34.942	46.626	0.131	0.869	0.272
$(0\,1\,1)\,(1\,1\,0)^{23}$	0.176	43.447	32.708	48.442	0.062	0.938	0.215
$(0\,1\,1)\,(1\,0\,0)^{23}$	0.148	49.722	40.255	55.437	−0.229	1.229	−0.028
$(1\,1\,1)\,(0\,0\,1)^{23}$	0.164	42.531	32.211	47.420	0.101	0.899	0.248
$(1\,1\,1)\,(0\,1\,1)^{23}$	0.326	37.082[a]	31.134	41.344[a]	0.317[a]	0.683[a]	0.428[a]
$(1\,1\,1)\,(1\,1\,1)^{23}$	0.398[a]	40.174	30.971[a]	44.792	0.198	0.802	0.329
$(1\,1\,1)\,(1\,1\,0)^{23}$	0.169	46.550	35.132	51.902	−0.077	1.077	0.099
$(1\,1\,1)\,(1\,0\,0)^{23}$	0.208	42.149	32.027	46.994	0.117	0.883	0.261
$(1\,1\,0)\,(0\,0\,1)^{23}$	0.099	49.607	39.939	55.309	−0.223	1.223	−0.024
$(1\,1\,0)\,(0\,1\,1)^{23}$	0.324	39.184	33.123	43.688	0.237	0.763	0.361
$(1\,1\,0)\,(1\,1\,1)^{23}$	0.323	39.390	33.177	43.918	0.229	0.771	0.355
$(1\,1\,0)\,(1\,1\,0)^{23}$	0.192	44.214	33.922	49.297	0.028	0.972	0.187
$(1\,1\,0)\,(1\,0\,0)^{23}$	0.139	49.459	39.770	55.144	−0.216	1.216	−0.018
$(1\,0\,0)\,(0\,0\,1)^{23}$	0.203	73.608	62.268	82.069	−1.693	2.693	−1.254
$(1\,0\,0)\,(0\,1\,1)^{23}$	0.305	44.189	36.695	49.268	0.029	0.971	0.188
$(1\,0\,0)\,(1\,1\,1)^{23}$	0.293	45.600	37.828	50.841	−0.034	1.034	0.135
$(1\,0\,0)\,(1\,1\,0)^{23}$	0.163	44.538	33.808	49.658	0.014	0.986	0.175
$(1\,0\,0)\,(1\,0\,0)^{23}$	0.204	71.807	61.055	80.061	−1.563	2.563	−1.145

In NFI row the benchmark values for each accuracy terms are indicated.
[a] Best results.

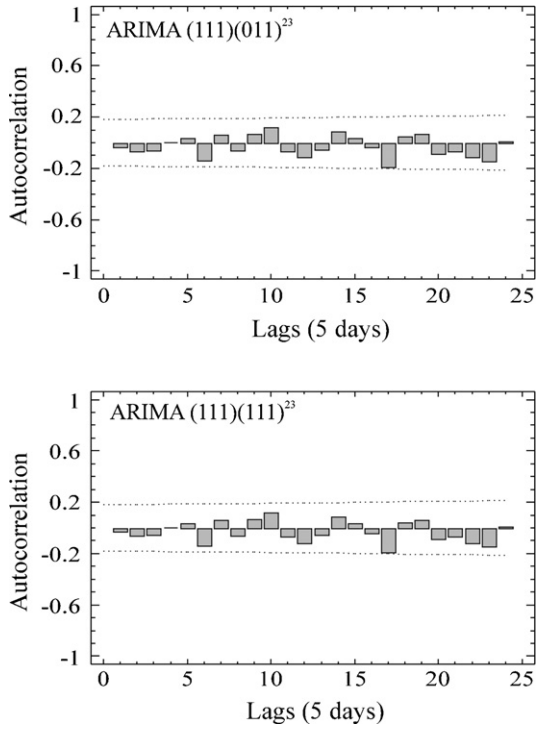

Fig. 2. Residual autocorrelations of ARIMA(1 1 1) $(0\,1\,1)^{23}$ and ARIMA(1 1 1) $(1\,1\,1)^{23}$ in the independent validation phase.

had a naïve behaviour. For example, the ANN 3:3*s*:1*l* (second repetition) presented a determination coefficient close to 0.87 and good values for RMSE, MAE, %SEP, E_2 and ARV, but PI value was close to zero.

The best estimate was obtained when the CPUE values of the three preceding sets of 5 days were used as inputs, with five nodes in the hidden layer. In this case, all the error magnitudes were significantly better than those obtained for the two best ARIMA models. Fig. 3 shows the best regressions of the ARIMA and ANN models in the validation phase. It can be seen that the ARIMA models presented a higher dispersion around the regression line than the ANN model, which was more fitted to the 1:1 line.

Table 2
Parameter values and significant levels of two best ARIMA models

Parameters	Value	Significant level
ARIMA(1 1 1) $(0\,1\,1)^{23}$		
ϕ	−0.00001	$p = 0.832$
θ	0.678	$p < 0.001$
Θ	0.805	$p < 0.001$
ARIMA(1 1 1) $(1\,1\,1)^{23}$		
ϕ	−0.00007	$p = 0.520$
θ	0.690	$p < 0.001$
Θ	0.00008	$p = 0.523$
Φ	0.807	$p < 0.001$

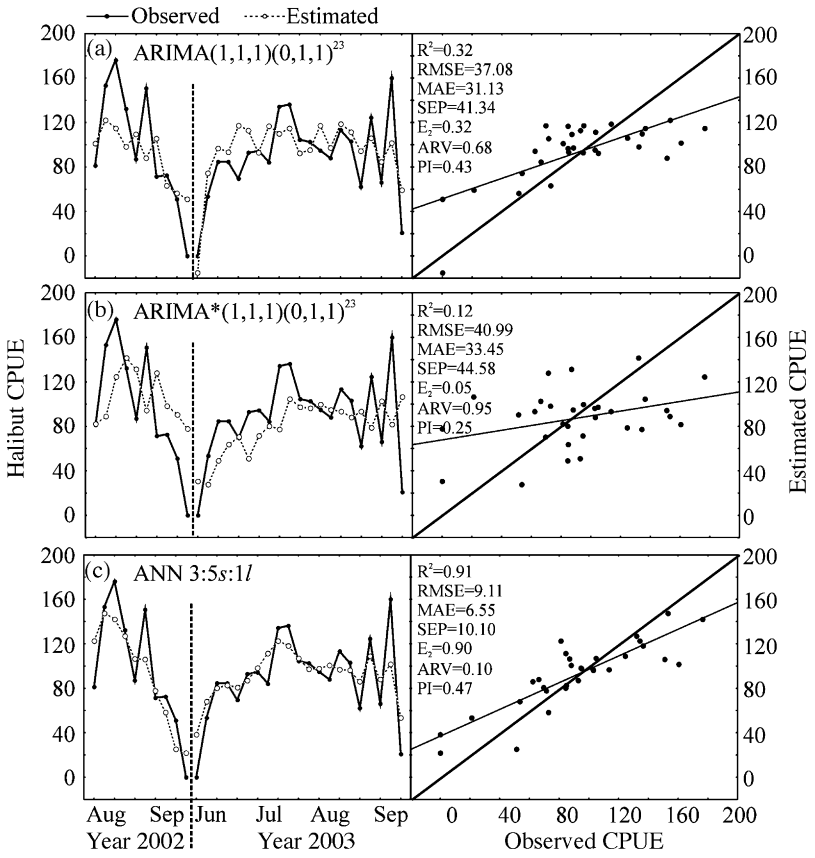

Fig. 3. (a) CPUE observed and estimated variation and linear regressions of the best ARIMA model [ARIMA(1 1 1) (0 1 1)^{23}] calibrated with a procedure of minimization of the sum of squared residuals in the independent validation phase, (b) CPUE observed and estimated variation and linear regressions of the best ARIMA model [ARIMA* (1 1 1) (0 1 1)^{23}] calibrated with X12 method in the independent validation phase and (c) CPUE observed and estimated variation and linear regressions of the best ANN model (3:5*s*:1*l*) in the independent validation phase.

Table 3
Results of ANNs in the independent validation phase with CPUE of three previous sets of 5 days

Model	Repetition	Error terms						
		R^2	RMSE	MAE	%SEP	E	ARV	PI
NFI	–	0.472	53.009	37.316	50.120	−0.056	1.056	0
3:3*s*:1*l*	1	0.752	19.656	16.201	21.786	0.534	0.466	−1.486
3:3*s*:1*l*	2	0.874	12.435	9.319	13.783	0.814	0.186	0.005
3:3*s*:1*l*	3	0.573	20.713	14.335	22.958	0.483	0.517	−1.760
3:3*s*:1*l*	4	0.908	10.813	7.940	11.985	0.859	0.141	0.248
3:3*s*:1*l*	5	0.887	11.470	9.003	12.713	0.841	0.159	0.154
3:5*s*:1*l*	1	0.101	33.119	24.935	36.708	−0.323	1.323	−6.057
3:5*s*:1*l*	2	0.900	10.834	8.078	12.008	0.858	0.142	0.245
3:5*s*:1*l*	3	0.901	10.561	8.145	11.705	0.865	0.135	0.282
3:5*s*:1*l*	4	0.910[a]	9.108[a]	6.552[a]	10.095[a]	0.900[a]	0.100[a]	0.466[a]
3:5*s*:1*l*	5	0.906	14.469	11.432	16.037	0.747	0.253	−0.347
3:7*s*:1*l*	1	0.896	9.569	6.884	10.606	0.890	0.110	0.411
3:7*s*:1*l*	2	0.372	24.404	18.740	27.048	0.282	0.718	−2.831
3:7*s*:1*l*	3	0.893	14.212	11.364	15.752	0.756	0.244	−0.299
3:7*s*:1*l*	4	0.652	17.605	13.048	19.513	0.626	0.374	−0.994
3:7*s*:1*l*	5	0.791	15.259	12.376	16.912	0.719	0.281	−0.498
3:9*s*:1*l*	1	0.788	16.702	13.715	18.512	0.664	0.336	−0.795
3:9*s*:1*l*	2	0.908	9.230	7.218	10.230	0.897	0.103	0.452
3:9*s*:1*l*	3	0.903	9.452	6.689	10.477	0.892	0.108	0.425
3:9*s*:1*l*	4	0.878	10.387	7.950	11.513	0.870	0.130	0.306
3:9*s*:1*l*	5	0.834	15.924	12.764	17.649	0.694	0.306	−0.631

In NFI row the benchmark values for each accuracy terms are indicated.
[a] Best results.

4. Discussion

In this paper, the capacity of two types of model (one linear, the other non-linear approaches) to produce accurate forecasts for Pacific halibut CPUE were compared. In total, 70 models were constructed (50 ARIMA models and 20 ANN models). The goodness of fit was tested using the seven accuracy measures most frequently used according to the bibliography (R^2, RMSE, MAE, %SEP, E_2, ARV, PI). The complementary use of different measures is highly recommended when two or more forecasting models are compared because most measures suffer certain limitation (Stergiou et al., 1997). This way, a model which one explain a high level of variance (R^2) in the validation phase can has associated a high value of absolute (RMSE, MAE) and relative (%SEP) error. In the case of time series forecasting, the persistence index (PI) is a very useful measure because provides a measure of comparison with a naïve model which one may provides explained variances and absolute and relative error statistically acceptable.

In many studies, ARIMA models are compared with other traditional statistical techniques, such as multiple linear regression (MLR), non-linear regression (NLR), seasonal time variation regression models (TVS), smoothing models, dynamic regression (DREG), harmonic regression (HREG), autoregression vector (VAR), generalized autoregressive conditional heteroscedasticity (GARCH), Thomas–Fiering models or Gausian autoregressive models (Stergiou, 1991; Stergiou et al., 1997; Romilly, 2005). These studies show that ARIMA validation errors are significantly lower, hence better than classic forecasting techniques.

The results showed that the ARIMA models produced forecasts with explained variances close to 40% and percent standard error prediction around 41%. These results, although statistically acceptable, differed significantly from those obtained by other authors for other fisheries. Stergiou et al. (1997) reported that the ARIMA models for forecasts of monthly captures of 16 species in the Hellenic marine waters were characterized by unbiased fits and forecasts, by a low level in all of the error terms used, and determination coefficient very close to 0.9. Similar results were found for forecasts of monthly landings of walleye pollock *Theragra chalcogramma* (Gadidae) in Korea. Phillips (1983) used univariate modelling techniques based on Box–Jenkins methodology to forecast the CPUE of anchovy and mackerel in the San Pedro fishery (USA). His models, in the validation phase, explained variances of 42 and 18%, respectively, which are similar to the results presented in this study using the same methodology. Also, Lloret et al. (2000) reported that the models did not explain the data variability in a satisfactory way for non-target demersal and pelagic species. Similar conclusions were obtained by Punzón et al. (2004) analysing a mackerel *Scomber scombrus* (Scombridae) fishery with time series of longline catches.

CPUE is an index of fish abundance that assumes proportionality between abundance and stock size. However, it is known that there are many parameters (including economics, geographical distributions, availability of fish to capture, vulnerability of fish and efficiency of the fishing gear), which affect CPUE but do not represent changes in abundance and which introduce non-linearity factors into the data series. Taking this in account, models like ARIMA, which assume the linearity of the process, can give statistically acceptable but insufficiently satisfactory estimations. Moreover, the accuracy of the model forecast is not only strongly influenced by the data quality used in the calibration phase and the kind of relationship between the original series data, but also by the quantity of the data available for the calibration and validation phases. Thus, the characteristics of the CPUE data series used in this study explain the lack of accuracy of the ARIMA models tested. This effect can be more important in the case of the ARIMA model calibrated with Census X12 method. In this case the model is calibrated with smoothing data by means of a linear, symmetric and centred moving average filter and validated with an original data series. This filtering process leads to very severe reduction of non-linear features in the original data series which ones cannot be recognised by the model in the validation phase.

In these conditions, methods like artificial neural networks, with the capacity to combine linear and highly non-linear relationships between variables, may be more appropriate for the characterisation and forecasting of Pacific halibut CPUE. Moreover, the design of the ANN allows each input variable to act independently but also in parallel with all the other variables, with the information being held in a network of weights. This provides the neural network with great flexibility for representing at the same time several non-linear relationships, which are very problematic for linear models.

As shown, non-linear methods such as ANNs are more appropriate for describing the univariate relationships of Pacific halibut CPUE. Thus, the proportions of variance explained and the forecast capacity are significantly higher than those obtained using ARIMA models. Gutiérrez-Estrada et al. (2004) showed the potential of autorregresive ANN models versus multiple regressions, classical smoothing models and ARIMA models, for forecasting the ammonia concentration in an intensive cultivation system for eels. These authors reported that ANN models were extremely capable of generalising from exceptional situations.

The autoregressive terms at time set t^{23} indicate that the CPUE (based on 5-day sets) exhibited a strongly cyclical seasonal character. Pacific halibut CPUE increased from a minimum in the first set (May 29–Jun 2) to a maximum in set 15 (August 7–11) and declined thereafter. It can be hypothesised that this marked seasonal cycle is due to the critical habitat of Pacific halibut during summer months being defined, at least partially, by water temperature (Loher and Seitz, 2006). This way, this persistence may indicate that the CPUE is linear dependent of the environmental conditions, catches and other factors of the previous fishing period. Forecasting accuracy is higher for data in 5-day sets than for monthly data, and improves considerably when annual fishing period forecasts are estimated from set forecasts. From non-linear point of view, the autoregressive terms at $t-1$, $t-2$ and $t-3$ also indicates a high relationship with captures in a short time in the same fishing period. This may be related with catch-sharing and catch limits plan established by the IPHC for this fishery.

In spite of the good results obtained with ANNs, it is necessary to put the utility of the model (its forecasting capacity and/or it biological/oceanographic significance) in the context of a short-medium term time period. In this framework, the model may be useful for estimating the CPUE value when the real value cannot be obtained (for example, if fishing effort values are unknown), for estimating interpolated data between two consecutives samples, and for simulating different fishery scenarios. Also, the model may be useful for detecting significant changes in the fishing effort, stock response or abundance related to various different factors (environmental, anthropogenic, etc.), since a lack of fit between observed and estimated data will indicate that a new CPUE pattern must be incorporated in the model; therefore the model must be calibrated and validated again.

Acknowledgements

The authors wish to express their gratitude to the IFAPA (Instituto de Investigación y Formación Agroalimentaria y Pesquera de Andalucía; Junta de Andalucía) for partially financing this research under Project C03-007. We are grateful to the International Pacific Halibut Commission (IPHC) to permit the use of Halibut databases. Also, we wish to express our gratitude to Royston Snart, B.Sc. for assistance with the critical revision of the English language.

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. Hydrol. Process. 14, 2157–2172.

Anctil, F., Rat, A., 2005. Evaluation of neural network stream flow forecasting on 47 watersheds. J. Hydrol. Eng. 10, 85–88.

Becerra-Muñoz, S., Buelna-Obsen, H.R., Catalán-Romero, J.M., 2003. Spatial patterns of ARIMA modeled rates of changes of atherinids (*Chirostoma* spp.) and goodeid *Chapalichthys encaustus* from Lake Chapala, México. Ecol. Model. 165, 237–250.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco, CA.

Clark, W.G., Hare, S.R., 2004. A conditional constant catch policy for managing the Pacific halibut fishery. N. Am. J. Fish. Manage. 24, 106–113.

Geernaert, T., Clark, B., Gilroy, H., 2005. Commercial halibut catch and effort for IPHC statistical areas on the B.C. coast (Regulatory Area 2B). International Pacific Halibut Commission. http://www.iphc.washington.edu/halcom/commerc/2bcpue2004.pdf.

Govindajaru, R.S., 2000. Artificial neural networks in hydrology. I. Preliminary concepts. J. Hydrol. Eng. 5, 115–123.

Gozlan, R.E., Mastrorillo, S., Copp, G.H., Lek, S., 1999. Predicting the structure and diversity of young-of-the-year fish assemblages in large rivers. Fresh Biol. 41, 809–820.

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

Guégan, J.F., Lek, S., Oberdorff, T., 1998. Energy availability and habitat heterogeneity predict global riverine fish diversity. Nature 391, 382–384.

Gutiérrez-Estrada, J.C., Pulido-Calvo, I., Prenda-Marín, J., 2000. Gonadosomatic index estimates of an introduced pumpkinseed (*Lepomis gibbosus* L. 1758) population in a Mediterranean stream, using computational neural networks (CNNs). Aquat. Sci. 62, 350–362.

Gutiérrez-Estrada, J.C., de Pedro-Sanz, E., López-Luque, R., Pulido-Calvo, I., 2004. Comparison between traditional methods and artificial neural networks for ammonia concentration forecasting in an eel (*Anguilla anguilla* L.) intensive rearing system. Aquat. Eng. 31, 183–203.

Haddom, M., 2001. Modelling and Quantitative Methods in Fisheries. Chapman & Hall/CRC, Boca Raton.

Hänninen, J., Vuorinen, I., Kornilovs, G., 2003. Atlantic climatic factors control decadal dynamics of a Baltic Sea copepod *Temora longicornis*. Ecography 26 (5), 672–678.

Holton-Winson, J., Keating, B., 1996. Previsión en los negocios. Richard D. Irwin, Madrid.

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

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, 233–241.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M., 1996a. Role of some environmental variables in trout abundance models using neural networks. Aquat. Liv. Res. 9, 23–29.

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

Lloret, J., Lleonart, J., Solé, I., 2000. Time Series modelling of landings in Northwest Mediterranean Sea. ICES J. Mar. Sci. 57, 171–184.

Loher, T., Seitz, A., 2006. Seasonal migration and environmental conditions of Pacific halibut *Hippoglossus stenolepis*, elucidated from pop-up archival transmitting (PAT) tags. Mar. Ecol. Prog. Ser. 317, 259–271.

Low, L.L., Stocker, M., 2005. Annual Report 2004. International Pacific Halibut Commission, Seattle, USA.

Makridakis, S., Wheelwrighy, S., McGee, V., 1983. Forecasting: Methods and Applications. John Wiley and Sons, New York.

McLaughlin, R.L., 1983. Forecasting models: sophisticated or naïve? J. Forecast. 2 (3), 27–276.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. I. A discussion of principles. J. Hydrol. 10, 282–290.

Phillips, G.M., 1983. Interactions between fishing effort and catch per unit effort in the San Pedro fishery: a statistical note. CalCOFI Rep. 24, 255–258.

Punzón, A., Villamor, B., Preciado, I., 2004. Analysis of the handline fishery targeting mackerel (*Scomber scombrus* L.) in the North of Spain (ICES Division VIIIbc). Fish. Res. 69, 189–204.

Qi, M., Zhang, G.P., 2001. An investigation of model selection criteria for neural network time series forecasting. Eur. J. Oper. Res. 132, 666–680.

Quinn, T.J., 1985. Catch-per-unit-effort-a statistical-model for Pacific halibut (*Hippoglossus stenolepis*). Can. J. Fish. Aquat. Sci. 42, 1423–1429.

Rigby, P.W, Ackley, D.R., Funk, F., Geiger, H.J., Kruse, G.H., Murphy, M.C., 1995. Management of marine fisheries resources of Alaska: a report to the Northern Forum. Regional Information Report: 5J95-04. ADF&G, Commercial Fisheries Management and Development Division.

Romilly, P., 2005. Time series modelling of global mean temperature for managerial decision-making. J. Env. Manage. 76, 61–70.

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

Stenseth, N.C., Lekve, K., Gjöæter, J., 2002. Modeling species richness controlled by community-intrinsic and community-extrinsic processes: coastal fish communities as an example. Popul. Ecol. 44, 165–178.

Stergiou, K.I., 1989. Modelling and forecasting the fishery for pilchard (*Sardina pilchardus*) in Greek waters using ARIMA time series models. ICES J. Mar. Sci. 46, 16–23.

Stergiou, K.I., 1990. Prediction of the Mullidae fishery in the eastern Mediterranean 24 months in advance. Fish. Res. 9, 67–74.

Stergiou, K.I., 1991. Describing and forecasting the sardine-anchovy complex in the eastern Mediterranean using vector autoregressions. Fish. Res. 11, 127–141.

Stergiou, K.I., Chritou, E.D., Petrakis, G., 1997. Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. Fish. Res. 29, 55–95.

Stocker, M., Noakes, D.J., 1988. Evaluating forecasting procedures for predicting Pacific herring (*Cuplea harengus pallasi*) recruitment in British Columbia. Can. J. Fish. Aquat. Sci. 45, 928–935.

Tanaka, E., 2000. Estimating of relative biomass of recruits using catch and effort series. Fish. Res. 66, 249–256.

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

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

Yaffee, R., McGee, M., 2000. Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS. Academic Press, San Diego, CA.

Zhang, C.I., Gunderson, D.R., Sullivan, P.J., 1991. Using data on biomass and fishing mortality in stock production modelling of flatfish. Neth. J. Sea Res. 27, 459–467.