Bol. Inst. Esp. Oceanogr. 22 (1-4). 2006: 83-90

BOLETÍN. INSTITUTO ESPAÑOL DE OCEANOGRAFÍA
ISSN: 0074-0195

Using time series methods for completing fisheries historical series

I. Preciado [1], *A. Punzón* [1], *J. L. Gallego* [2] *and Y. Vila* [3]

[1] Centro Oceanográfico de Santander, Instituto Español de Oceanografía, Promontorio de San Martín, s/n, Apdo. 240, E-39080 Santander, Cantabria, Spain. E-mail: ipreciado@st.ieo.es; antonio.punzon@st.ieo.es
[2] Departamento de Economía, Facultad de Ciencias Económicas y Empresariales, Universidad de Cantabria, Avda. de los Castros, s/n, E-39005 Santander, Cantabria, Spain. E-mail: gallegoj@unican.es
[3] Unidad de Biología Pesquera, Instituto Español de Oceanografía, Puerto Pesquero, Muelle de Levante, s/n, Apdo. 2609, E-11006 Cádiz, Spain. E-mail: yolanda.vila@cd.ieo.es

Received March 2004. Accepted December 2006.

ABSTRACT

The so-called *baca* trawl fleet of the port of A Coruña working in the VIIIc ICES Division Area has been used since 1985 as the tuning fleet for the analytical assessment of the southern stock of hake *Merluccius merluccius* (L., 1758). The historical series of cpue dates from 1983 to 2000 with data missing from 1998. From January 1999, landings were estimated by means of monthly samplings in port. The predictions of the arima model fit the 1983-1997 series, and estimates of the transfer function model from the 1983-2000 series made it possible to complete the series, and to detect errors and possible changes in the orientation of the fleet. The comparison of the two models confirms the series's stability over time, validating the estimates made in the last two years.

Keywords: Arima models, fisheries database, hake, time series, transfer function model.

RESUMEN

Empleo de modelos de series temporales para la mejora de la calidad de las series históricas en pesquerías

La flota de baca del puerto de A Coruña que opera en el área VIIIc del CIEM se usa desde 1985 como flota de calibración en la evaluación del stock *sur de merluza* Merluccius merluccius *(L., 1758). Se dispone de una serie histórica mensual de cpue (1983-2000), excepto para 1998, año del que no se posee información. Desde enero de 1999 los desembarcos han sido estimados a partir de muestreos mensuales en lonja. Las predicciones efectuadas por el modelo arima ajustado a la serie 1983-1997 y las estimaciones del modelo de función de transferencia, considerando la serie 1983-2000, han permitido completar la serie, así como detectar errores o cambios en el comportamiento de la flota. La comparación de ambos modelos confirma la estabilidad de la serie, validando las estimaciones efectuadas en los dos últimos años.*

Palabras clave: *Modelos arima, bases pesqueras de datos, merluza, series temporales, modelo función de transferencia.*

INTRODUCTION

A major goal of fisheries databases is to obtain a census of catches and efforts of their fleets which are considered suitable for the elaboration of representative indices of catch rate (catch per unit effort). The application of such databases has been called into question, and there have been several recommendations aimed at improving estimates (Harley, Ransom and Dunn, 2001; ICES, 2002b), since nominal effort is used in most cases. Nevertheless, these indices must be estimated in some way, as they are used in the calibration of the analytical models used in the assessment of exploited stocks in ICES Working Groups.

The so-called *baca* trawl fleet of VIIIc ICES División waters is part of a demersal mixed-species fishery. The effort of the fleet making its landings in the port of A Coruña (northwestern Spain) is used as a tuning fleet in the assessment of the southern stock of hake *Merluccius merluccius* (L., 1758) (ICES, 2002a; ICES, 2002c). The landing data supplied to the ICES Working Group by the different partners come from log-books, official organisations of each partner, and/or directly from the sector. Catches and effort of this fleet are supplied voluntarily by the sector, and so different problems arise in relation to them, such as the quality control of the data and missing data. The hake cpue time series (1983-2000) from the trawl fleet in the port of A Coruña (figure 1, table I) is incomplete, since monthly data on cpue are not available for 1998. In addition, values from January 1999 had to be estimated by trip sampling.

Historical series of cpue must be treated as time series (Hilborn and Walters, 1992), with the arima and transfer function models being fast and efficient tools for the analysis of time series of fisheries data (Stergiou and Christou, 1996; Stergiou, Christou and Petrakis, 1997) and the most commonly used to make short-term predictions (Mendelssohn, 1981; Fogarty, 1988; Stergiou, 1989; Lloret, Lleonart and Solé, 2000; Parsons and Colbourne, 2000). The aim of the present paper is to use this type of model to complete the historical series and validate the estimates of total catch obtained from sampling exercises over the last two years of the series.

MATERIALS AND METHODS

Monthly hake landings and effort (fishing days by 100 HP) at A Coruña for VIIIc ICES Division waters from 1983 to 1997 were supplied by fishermen's associations (all catch data come from landing statistics). As monthly hake landings were not available from 1998, a number of trips since 1999 were sampled, and the estimates thus obtained were later weighted to the total effort of the fleet to estimate its total monthly catch by species (for hake).

Table I. Total monthly cpue (fishing days per 100 HP) of *M. merluccius* from A Coruña trawl fleet in VIIIc ICES Division waters (1983-2000). (–): no available data; (*): data estimated by trip sampling. Number of trips sampled in 1999 and 2000 are given in brackets

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1983	18.53	33.68	48.24	59.52	52.19	41.39	33.71	42.71	35.44	29.20	20.80	20.69
1984	20.81	49.34	44.05	54.34	39.06	27.23	17.54	16.69	18.06	13.04	9.40	6.21
1985	14.72	27.95	30.95	37.30	28.41	22.83	20.43	21.28	18.05	13.42	6.77	10.28
1986	12.24	22.38	28.82	27.31	31.33	24.85	19.72	20.47	22.72	18.54	13.87	10.57
1987	14.56	27.91	28.21	31.63	32.28	21.29	18.39	14.77	12.69	14.29	11.61	11.99
1988	14.14	31.22	26.32	24.91	19.54	16.61	15.49	11.32	12.77	7.97	9.03	13.10
1989	10.53	18.51	20.30	17.56	19.76	12.36	16.91	16.95	17.77	13.29	17.51	9.65
1990	22.04	23.60	17.96	24.92	22.45	16.52	14.45	13.87	14.00	10.46	8.36	12.47
1991	13.65	19.82	20.45	19.46	22.18	10.99	6.30	5.56	4.76	7.33	4.76	6.68
1992	15.37	21.95	23.20	24.03	20.89	15.39	10.28	7.06	11.47	10.88	11.76	10.40
1993	12.89	16.91	21.37	17.05	13.73	9.68	6.74	5.81	8.96	9.78	8.59	11.50
1994	18.02	15.48	11.83	12.69	8.46	7.07	5.53	12.49	16.86	14.26	8.91	12.06
1995	19.74	26.40	22.80	22.60	17.21	18.74	16.97	16.79	23.08	21.79	20.37	13.84
1996	19.38	33.03	32.05	22.81	27.91	22.48	15.95	13.88	18.51	12.15	11.97	9.79
1997	25.80	26.12	13.15	16.75	13.70	11.77	19.18	13.04	21.81	16.72	15.38	22.65
1998	–	–	–	–	–	–	–	–	–	–	–	–
1999*	42.92 (2)	42.92 (3)	19.20 (42)	16.08 (39)	18.11 (34)	14.70 (41)	10.94 (45)	14.75 (27)	25.19 (25)	27.58 (34)	23.80 (33)	22.00 (23)
2000*	28.02 (27)	25.41 (37)	15.64 (47)	19.60 (51)	19.33 (51)	11.19 (53)	11.11 (56)	18.92 (41)	24.31 (47)	23.83 (51)	19.64 (49)	19.74 (43)
Mean	17.99	26.25	24.97	26.39	23.91	17.95	15.27	15.67	18.03	15.56	13.09	13.16

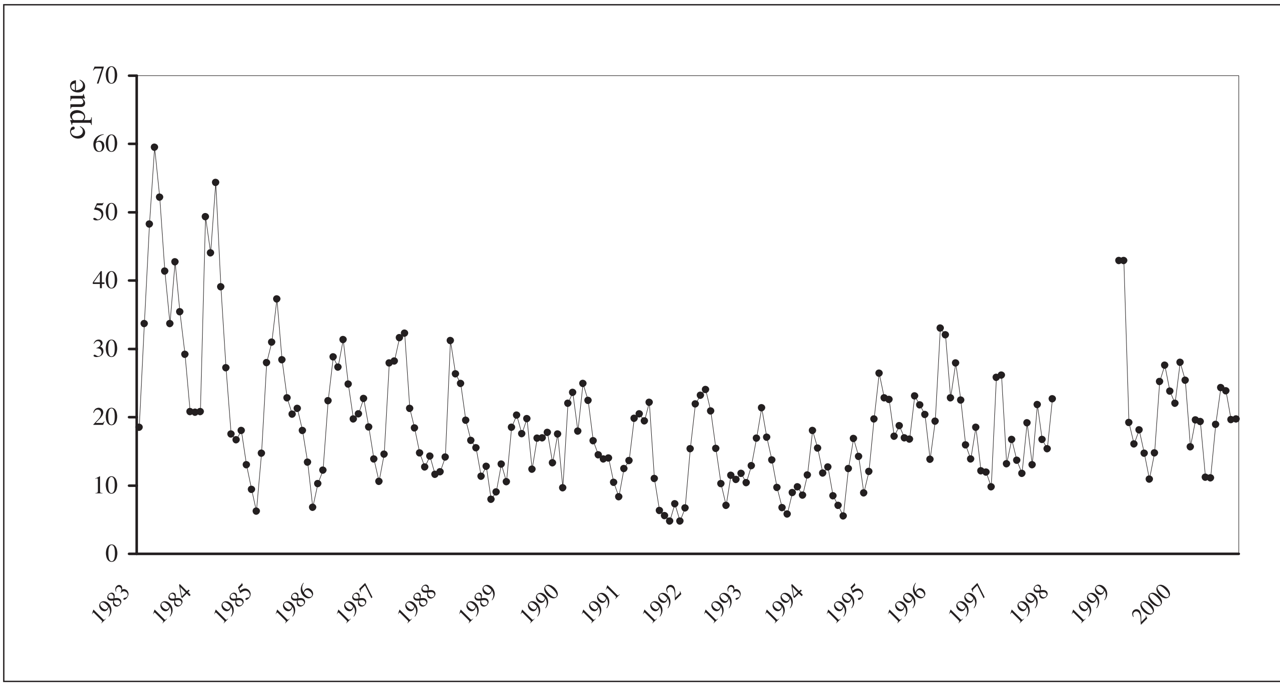

Figure 1. Hake monthly cpue time series (1983-2000) from A Coruña baca trawl fleet

Estimation of missing values with arima and transfer function models

Consecutive missing data in a seasonal time series can be estimated by computing the expectation of the unobserved random variable, given the rest of the data (Brubacher and Wilson, 1976; Damsleth, 1980; Abraham, 1981; Box, Jenkins and Reinsel, 1994; Peña, 1987, 2001). Here, we used a two-step procedure to calculate missing values. Statistical analyses were performed using Empiricus software. We started by building a multiplicative arima (p,d,q) $(P,D,Q)_s$ model

$$\phi(B)\ \Phi(B^s)\ \nabla^d\ \nabla_s^D z_t = \theta(B)\ \Theta(B^s)\ a_t$$
$$t = 1, \ldots \tau$$

where B is the backshift operator such that $Bz_t = z_{t-1}$, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ is the regular autoregressive polynomial, $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ is the regular moving-average polynomial, $\Phi(B^s) = 1 - \Phi_1 B^s - \ldots - \Phi_P B^{Ps}$ is the seasonal autoregressive polynomial, $\Theta(B) = 1 - \Theta_1 B^s - \ldots - \Theta_Q B^{Qs}$ is the seasonal moving-average polynomial, $\nabla = 1 - B$ is the regular difference, $\nabla_s = 1 - B^s$ is the seasonal difference, d and D are integers, s is the seasonal period (s = 12 for monthly series), and a_t is a sequence of independently distributed normal variables having mean zero and variance σ_a^2, usually referred to as a white noise process (Box, Jenkins and Reinsel, 1994).

This model can be used to predict unavailable data $z_{\tau+L}$ $(L \geqslant 1)$ as a linear combination of the observations $z_1, \ldots z_\tau$. The minimum mean square error (MMSE) prediction of $z_{\tau+L}$ made at the origin τ for lead time L, denoted by $\hat{z}_\tau$ (L) is the conditional expectation of $z_{\tau+L}$ at time τ and satisfies the difference equation

$$\phi(B)\ \Phi(B^s)\ \nabla^d\ \nabla_s^D\ \hat{z}_\tau(L) = 0$$
$$L > p + Ps$$

where B operates on the horizon L and $\hat{z}_\tau(L) = z_{\tau-1}$ for $L \leq 0$. These forecasts provided rough estimates for the missing values (January-December 1998). These data were introduced into the time series to predict 1999, and to compare them with observations. Similarly, predictions for 2000 were made and compared with the corresponding observations. To compare forecasted with observed values, the coefficient of determination was used (Stergiou, 1989).

Secondly, we built a transfer function model using the entire time series (1983-2000)

$$z_t = \frac{\omega(B)}{\delta(B)} x_t + N_t$$
$$\phi(B)\ \Phi(B^s)\ \nabla\nabla_s N_t = \theta(B)\ \Theta(B^s)\ a_t$$

where the output z_t and the input x_t are linked by the transfer function $\omega(B)/\delta(B)$

$$\omega(B) = \omega_0 - \omega_1 B - ... - \omega_m B^m$$
$$\delta(B) = 1 - \delta_1 B - ... - \delta_r B^r$$

are polynomials of orders m and r, and the noise N_t is described by a multiplicative arima (p,d,q) $(P,D,Q)_s$. This model was used by Box and Tiao (1975) to estimate the effect of interventions on a given response variable. Following Peña (1987), to compute missing data using the transfer function model we began by replacing the 12 consecutive missing observations (January-December 1998) with zero values, which can be thought of as additive outliers resulting from an intervention at time τ. Next, we defined $x_t = \xi_t^{p,\tau}$ as a pulse variable

$$\xi_t^{p,\tau} = \begin{cases} 1, & t = \tau + 1 \\ 0, & t \neq \tau + 1 \end{cases}$$

whose transfer function is

$$\omega(B) = \omega_0 - \omega_1 B - \Lambda - \omega_m B^m$$

and $\delta(B) = 1$. The coefficients $|\omega_j|$ provide the estimates for the missing values $z_{\tau+j}$. The estimation of this model requires an arima model to be identified for the noise, which was done in the first stage.

RESULTS

Arima modelling (1983-1997)

We described the univariate analysis of the hake cpue monthly time series from January 1983 to December 1997 using the Box-Jenkins iterative approach based on identification, estimation and diagnostic checking stages (Box, Jenkins and Reinsel, 1994). The identification of a tentative model for a seasonal time series was based mainly on the visual inspection of the plot (figure 1), sample autocorrelation function (SACF) and sample partial autocorrelation function (SPACF) of Z_t and $\nabla\nabla_s Z_t$ (figure 2). Examination of these graphs for our series indicated that both regular and seasonal differences are needed to achieve stationarity. The SACF of $\nabla\nabla_s Z_t$ showed significant correlations at lags 1, 11, 12 and 13, whereas its SPACF was dominated by damped exponentials at lags 1, 12, 24, with significant positive coefficients at lags 11 and 23, suggesting that this time series could be described by a multiplicative IMA $(1,1)(1,1)_{12}$ process. Estimation by exact maximum likelihood based on the Ansley (1976) algorithm provided the results shown in table II. These diagnostic statistics, along with the SACF and cumulative periodogram for residuals of the model (figure 3), suggest that the model is satisfactory.

We used the model to forecast missing values from January 1998 to December 1998 (table III). Furthermore, this model provides a tentative specification for the noise of the dynamic regression model. Observations and predictions for 1999 were similar ($r^2 = 0.81$, $p < 0.01$), although unusually high values could be observed for January and February (figure 4). As the number of trips sampled during those months was very low, we removed this information from the historical series and treated it in the same way as the missing data from 1998. Similarly, comparisons between observations and predictions for 2000 ($r^2 = 0.70$, $p < 0.01$) are shown in figure 4.

Transfer function model (1983-2000)

The hake cpue time series was analysed for the period from January 1983 to December 2000, excluding the period between January 1998 and February 1999. These missing data were consid-

Table II. Estimated parameters, standard error (SE) and correlation coefficient (r^2) of model for the time series 1983-1997 and 1983-2000. $\bar{a}$: mean of the residuals; σ_a: standard error of the residuals; Q(K): Ljung-Box statistic computed from the first K residual autocorrelations; H(g): likelihood ratio test statistic for heteroscedasticity; JB: Jarque and Bera test statistic for normality

	Time series (1983-1997)		Time series (1983-2000)	
	Parameters	SE	Parameters	SE
θ	0.34	0.07	0.33	0.07
Θ	0.59	0.07	0.59	0.06
r^2	0.16		0.21	
	$\nabla\nabla_{12} Z_t = (1-0.34B)(1-0.59B^{12})a_t$ $\bar{a} = 0.050\ (0.366)$ $\sigma_a = 4.73$ Q(39) = 49.1; H(3) = 2.23; JB = 0.56		$\nabla\nabla_{12} N_t = (1-0.33B)(1-0.59B^{12})a_t$ $\bar{a} = -0.014\ (0.315)$ $\sigma_a = 4.49$ Q(39) = 57.1; H(3) = 2.68; JB = 0.51	

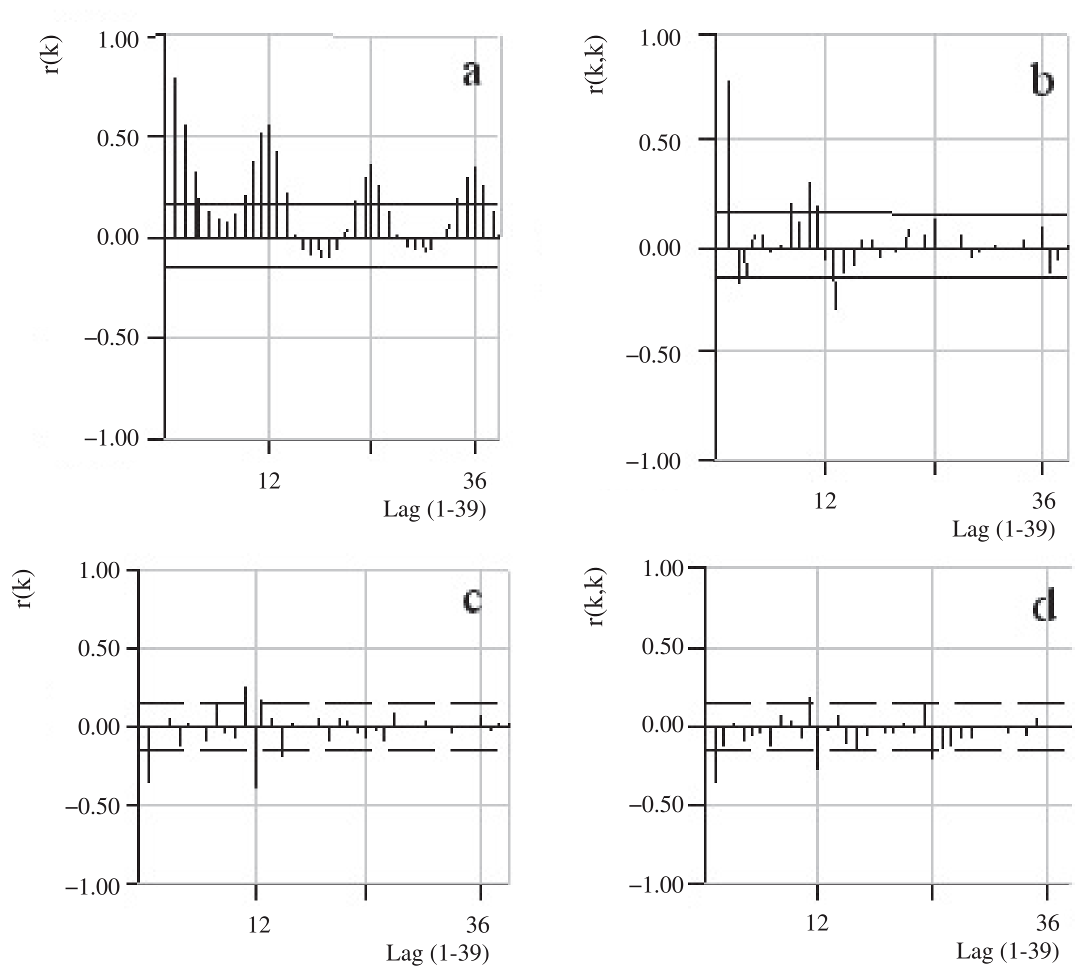

Figure 2. Identification tools for the series 1983-1997. (a): sample autocorrelation function for the series Z_t; (b): sample partial autocorrelation function for the series Z_t; (c): sample autocorrelation function for the series $\nabla\nabla_s Z_t$; (d): sample partial autocorrelation function for the series $\nabla\nabla_s Z_t$

Table III. Missing values of A Coruña cpue time series for 1998 and January and February 1999, estimated as forecasts $[\hat{z}_{\tau-1}(j+l)]$ and unit impulse response $(\hat{\omega}_j)$. (SE): standard error

j	$\hat{z}_{\tau-1}(j+l)$	SE	$(\hat{\omega}_j)$	SE
Jan-98	29.62	4.73	27.86	4.01
Feb-98	34.50	5.66	29.93	4.57
Mar-98	28.40	6.46	22.35	4.85
Apr-98	27.53	7.17	21.53	5.16
May-98	26.14	7.81	20.88	5.38
Jun-98	23.40	8.41	16.57	5.52
Jul-98	24.04	8.97	15.64	5.62
Aug-98	21.44	9.49	16.34	5.60
Sep-98	27.67	10.46	23.16	5.54
Oct-98	23.64	10.91	21.09	5.40
Nov-98	22.30	11.35	18.22	5.20
Dec-98	24.20	12.41	18.51	4.92
Jan-99	33.00	13.16	25.52	4.67
Feb-99	37.88	19.09	26.62	4.06

ered to be the effect of an intervention affecting the level for 14 consecutive months. Using the results from the previous analysis, the noise was described using an IMA$(1,1)(1,1)_{12}$ model. Estimation by exact maximum likelihood provided the results shown in table II. The estimates for the missing values were provided by the absolute values of the impulse response function (table III).

DISCUSSION

The results of the model that fit the partial series (1983-1997) and the entire series (1983-2000) are from an arima model $(0,1,1)(0,1,1)_{12}$, and are similar to the results found by Lloret, Lleonart and Solé (2000) for hake in the Mediterranean Sea.

The parameters estimated by the two models, and therefore the observations obtained by sampling trips in 1999 and 2000, do not change the series's behaviour.

The transfer function model provided estimates of monthly cpue for 1998, for which information was not available, and for the data of January and February 1999. The inclusion of these data in the historical series enabled us to make predictions for the remainder of 1999 and to compare them with the observations. The differences between the values of the months of January 1999 and February 1999 can be explained by the small number of trips sampled these months (2 and 3, respectively). During the rest of the year, both trend and seasonality are similar. The observations made in this case can be validated except for January and February, in which the data that should be used are those estimated by the model. The results corresponding to 2000 reflect a similar trend and seasonality in predictions and observations. Nevertheless, the predictions made by the model tend to overestimate cpue. As it is a mixed-species fishery, it should be taken into account that its orientation is subject to economic criteria and species accessibility among other factors, and that a monospecific approximation may be restrictive (Biseau, 1998). In addition, the abundance index used is not standardised, and so these variables, the technical characteristics of the fleet and the geographical references have not been taken into account (Hilborn and Walters, 1992; Salthaug and Godø, 2001). Moreover, part of the effort is targeting other resources (Punzón *et al.*, 2001; Punzón *et al.*, 2005) so that differences with observed values could be an indication of changes in the effort. Considering that in the development of this kind of model, the recent past has the greatest influence on predictions due to the condition of invertibility, this overestimate may indicate a change in yields and/or the orientation of the fleet in favour of other species in addition to hake. Stergiou (1989),

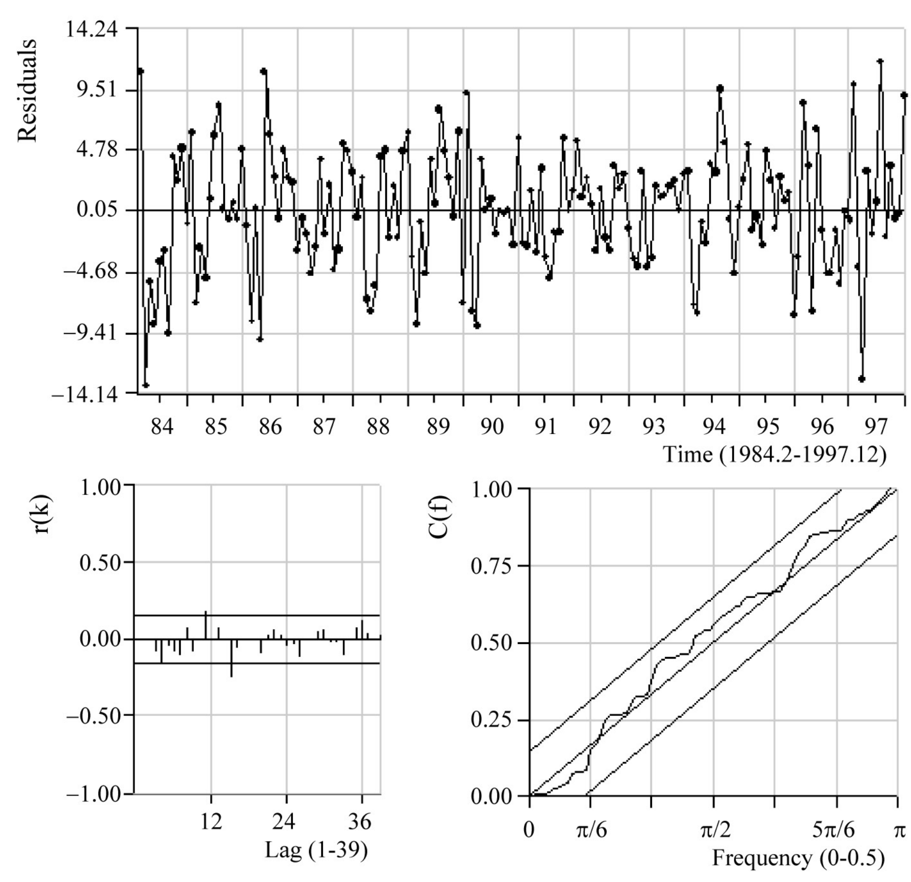

Figure 3. Diagnostic checking for residuals of the arima model. r(k): autocorrelation function; C(f): cumulative periodogram

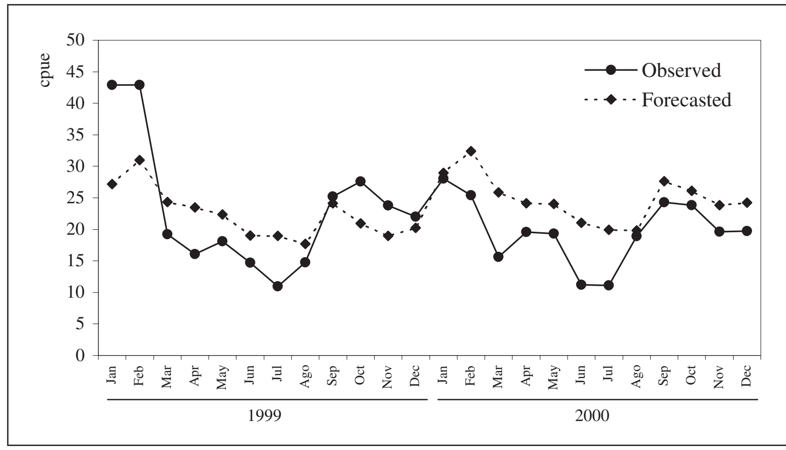

Figure 4. Comparison between observed cpue data from A Coruña baca trawl fleet and forecasted cpue according to the model (1983-1998 time series), for 1999 and 2000. January and February data were estimated with the transfer function model

Pajuelo and Lorenzo (1995) and Park (1998), among others, show how differences between predictions and observations are attributable, as in our case, to modifications in factors governing the historical series (environmental conditions, recruitment or effort) which may partially or substantially alter observations.

This leads us to the need to review the historical and current behaviour of this fleet, to improve the estimation of mixed-species fisheries' catch rate through the standardisation of effort, and to analyse and check the data observed. It is interesting to note that some deviations from the model estimated could be an indication of changes in the effort or in some of the parameters of the stock, which would require further investigation. This technique could be routinely used to check and control the quality of information in fisheries databases.

ACKNOWLEDGEMENTS

Our thanks to the entire team of the ICES Database and at the Instituto Español de Oceanografía, and particularly to Rosa Gancedo and Roberto Morlán for their collaboration on data collection. Thanks also to Pablo Abaunza and Valentín Trujillo for their useful suggestions.

REFERENCES

Abraham, B. 1981. Missing observations in time series. *Comm. Statis.* A10: 1643-1653.

Ansley, C. F. 1976. An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrics* 66: 59-65.

Biseau, A. 1998. Definition of a directed fishing effort in a mixed-species trawl fishery, and its impact on stock assessments. *Aquatic Living Resources* 11: 119-136.

Box, G. E. P., G. M. Jenkins and G. C. Reinsel. 1994. *Time series analysis. Forecasting and Control*. Prentice-Hall. New Jersey: 598 pp.

Box, G. E. P. and G. C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. *J. Am. Stat. Assoc.* 70: 70-79.

Brubacher, S. R. and G. T. Wilson. 1976. Interpolating time series with applications to the estimation of holiday effects on electricity demand. *Appl. Stat.* 25: 107-116.

Damsleth, E. 1980. Interpolating missing values in a time series. *Scand. J. Statist.* 7: 33-39.

Fogarty, M. J. 1988. Time Series Models of the Maine Lobster Fishery: the effect of temperature. *Can. J. Fish. Aquat. Sci.* 45: 1145-1153.

Harley, S. J., A. M. Ransom and A. Dunn. 2001. Is catch-per-unit-effort proportional to abundance? *Can. J. Fish. Aquat. Sci.* 58: 1760-1772.

Hilborn, R. and C. J. Walters. 1992. *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman & Hall. New York: 570 pp.

ICES. 2002a. Report of the Working Group on the Assessment of Southern Shelf Demersal Stocks. *ICES C.M.* 2002/ACFM:05: 981 pp.

ICES. 2002b. Report of the Working Group on Methods on Fish Stock Assessments. *ICES C.M.* 2002/D:08: 98 pp.

ICES. 2002c. Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. *ICES C.M.* 2002/ACFM:06: 474 pp.

Lloret, J., J. Lleonart and I. Solé. 2000. Time Series Modelling of Landings in Northwest Mediterranean Sea. *ICES. Journal of Marine Sciencie* 57: 171-184.

Mendelssohn, R. 1981. Using Box-Jenkins Models to Forecast Fishery Dynamics: Identification, Estimation and Checking. *Fish. Bull.* 78: 887-896.

Pajuelo, J. G. and J. M. Lorenzo. 1995. Análisis y predicción de la pesquería demersal de las Islas Canarias mediante un modelo ARIMA. *Scientia Marina* 59: 155-164.

Park, H. H. 1998. Analysis and prediction of walleye pollock (*Theragra chalcogramma*) landings in Korea by time series analysis. *Fish. Res.* 38: 1-7.

Parsons, D. G. and E. B. Colbourne. 2000. Forecasting Fishery Performance for Northern Shrimp (*Pandalus borealis*) on the Labrador Shelf (NAFO Divisions 2HJ). *J. Northwest Atl. Fish. Sci.* 27: 11-20.

Peña, D. 1987. Measuring the importance of outliers in ARIMA models. In: *New Perspectives in Theoretical and Applied Statistics*. M. L. Puri (ed.). Wiley. New York: 109-118.

Peña, D. 2001. Outliers, Influential Observations, and Missing Data. In: *A Course in Time Series Analysis*. D. Peña, G. C. Tiao and R. S. Tsay (eds.). Wiley. New York: 136-170.

Punzón, A., G. Costas, R. Gancedo and R. Morlan. 2001. Segmentation of the mixed "baca" trawl fishery that exploits demersal resources in the Cantabrian Sea (ICES Division VIIIc). *ICES C.M.* 2001/Q:19: 7 pp.

Punzón, A., V. Trujillo, E. Abad, C. Pampillón, F. Velasco, P. Abaunza, B. Villamor, N. Pérez and C. Hernández. 2005. Fisheries and Prestige: Review and update of studies on the effects of the Prestige oil spill. *ICES C.M.* 2005/S:11: 20 pp.

Salthaug, A. and O. R. Godø. 2001. Standardisation of commercial cpue. *Fish. Res.* 49: 271-281.

Stergiou, K. I. 1989. Modelling and forecasting the fishery for pilchard (*Sardina pilchardus*) in Greek waters Using ARIMA Time-Series Models. *J. Cons. Int. Explor. Mer.* 46: 16-23.

Stergiou, K. I. and E. D. Christou. 1996. Modelling and forecasting annual fisheries catches: comparison of regression, univariate and multivariate time series methods. *Fish. Res.* 25: 105-138.

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