STATISTICAL MODELLING OF INLAND FISH PRODUCTION IN INDIA

Ranjit Kumar Paul
Indian Agricultural Statistics Research Institute

Manas. Kumar. Das
Central Inland Fisheries Research Institute

J. Inland Fish. Soc. India, **42**(2) : 1-7, 2010

STATISTICAL MODELLING OF INLAND FISH PRODUCTION IN INDIA

R. K. PAUL AND M. K. DAS
Central Inland Fisheries Research Institute, Barrackpore, Kolkata - 700 120

(Received : 14.07.2010; Accepted : 24.12.2010)

Fish production in India has increased at a higher rate compared to food grains, milk, egg and other food items. The most widely used time series model *i.e.* autoregressive integrated moving average (ARIMA) model is applied for modelling and forecasting of total inland fish production in India. The annual inland fish production data from 1951 to 2000 were used for building the model and data from 2001 to 2008 were used for validation of the model. To this end, evaluation of forecasting of inland fish production was carried out with dynamic one step ahead forecast error variance along with mean absolute error (MAE), mean absolute prediction error (MAPE) and relative mean absolute prediction error (RMAPE). The forecast of inland fish production in India for the year 2009 and 2010 have been found out as 4360 and 4610 thousand tonnes.

Keywords: ARIMA, inland fish production, forecasting, statistical modelling.

Introduction

In many parts of the world, fish stock are currently overexploited or have not been adequately managed. As a result catches are declining. An essential component of successful fisheries management is an ongoing assessment programme to monitor the condition of the fish stock in the context of the aquatic ecosystem and the fishing activities that sustain the fishing community. Trends in the size of fish catches can be an important indicator of the status of the fishing industry. The 8,000 km coastline from both inland and marine resources, 3 million hectares of reservoirs, 1.4 million hectares of brackish water, 50,600 sq km of continental shelf area and 2.2 million sq km of exclusive economic zone supplement India's vast potential for fishery.

Fish production in India has increased at a higher rate compared to food grains, milk, egg and other food items. India ranks second in the world fish production with an annual fish production of about 6.9 million metric tonnes. Fisheries sector contributed Rs. 34,758 crores to the GDP during 2005-06, which was 1.2% of the national GDP and 5.3% of the agricultural GDP. There has been a gradual shift in the production scenario from marine to inland fisheries in recent years. More than 50% of the production in the fishery sector comes from the inland fishery sector, which has grown nearly fivefold from 0.67 MT in 1970-71 to 3.2 MT in 2002-2003 and is expected to further grow at a rate of 6% per annum. The inland fisheries resources of India include a length of 0.17 million kilometers rivers and canals, 2.05 million ha of reservoir area, 2.86 million ha area of ponds and tanks and 0.8 million ha of beels, oxbow lakes and derelict water. Two decades back most of the inland fish production was obtained in the capture fishery mode, concentrated in the rivers, reservoirs and wetlands. But, the fish production from natural waters like rivers, wetlands, canals, *etc.*, followed a declining trend, primarily due to proliferation of water control structures, indiscriminate fishing and habitat degradation (Katiha, 2000).

Statistical modelling plays a very important role in comprehending underlying relationships among crucial variables in the fishery sector determining fish production. The data help both fishermen as well as fish processing plants to successfully plan for the future (Mendelssohn, 1981; Mendelssohn and Cury, 1987; Cook *et al.*, 1991; Shepherd, 1991). In view of globalization and setting up of the World Trade Organization, it is imperative to study the trend and rate of growth in inland fish production by employing sound statistical modelling techniques that, in turn will be beneficial to the planners in formulating suitable policies to face the challenges ahead. In this paper our purpose is to apply autoregressive integrated moving average (ARIMA) time series model for modelling as well forecasting of inland fish production in India.

Materials and methods

A fundamental problem in statistics is to develop models based on a sample of observations and inferences using the model so developed. In fisheries research, data are usually collected over time. One characteristic of such data is that the successive observations are dependent. Each observation of the observed data series, Y_t, may be considered as a realization of a stochastic process $\{Y_t\}$, which is a family of random variables $\{Y_t,\ t \in T\}$, where $T = \{0, \pm 1, \pm 2, \ldots\}$, and apply standard time-series approach to develop an ideal model which will adequately represent the set of realizations and also their statistical relationships in a satisfactory manner.

There are two types of fishery forecasting models: deterministic models and stochastic models. The deterministic models do not have a random variable and each prediction is made under a specific set of conditions that are always the same (William, 1986). These models include the surplus production model (Schaefer, 1954, 1957; Pella and Tomlinson, 1969; Fox, 1970, 1975) and the classic regression model (Hanson and Leggett, 1982; Prepas, 1983). The stochastic models, in contrast, have a random variable that represents error terms of random factor(s). These models include the autoregressive integrated moving average (ARIMA) model and the transfer function noise model (Box *et al.*, 2007; Liu and Hanssens, 1982). They have both been used in forecasting the commercial catches of Atlantic Menhaden (Jensen, 1976, 1985), New Zealand rock lobster (Saila *et al.*, 1979), Hawaiian skipjack tuna (Mendelssohn, 1981), as well as Mediterranean Mullidae, sardine and sardine-anchovy complex (Stergiou, 1989, 1990, 1991).

Description of the model

Autoregressive Moving Average (ARMA) Model

To achieve greater flexibility in fitting of actual time-series data, it is sometimes advantageous to include both autoregressive and moving average processes. This leads to the mixed autoregressive-moving average model

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad (1)$$

or

$$\varphi(B) y_t = \theta(B) \varepsilon_t \quad \text{where}$$

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

WN indicating White Noise. B is the backshift operator such that $By_t = y_{t-1}$

This is written as ARMA(p, q) model. In practice, it is frequently true that adequate representation of actually occurring stationary time-series can be obtained with autoregressive, moving average, or mixed models, in which p and q are not greater than 2 and often less than 2.

Autoregressive Integrated Moving Average (ARIMA) Model

A generalization of ARMA models which incorporates a wide class of non-stationary time-series is obtained by introducing the differencing into the model. The simplest example of a non-stationary process which reduces to a stationary one after differencing is Random Walk. A process $\{y_t\}$ is said to follow an Integrated ARMA model, denoted by ARIMA (p, d, q), if $\nabla^d y_t = (1-B)^d \varepsilon_t$ is ARMA (p, q). The model is written as

$$\varphi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \qquad (2)$$

where $\varepsilon_t \sim WN(0,\sigma^2)$

WN indicating White Noise. The integration parameter d is a nonnegative integer. When $d = 0$, ARIMA (p, d, q) model reduces to ARMA (p, q) model.

The ARIMA methodology is carried out in three stages, *viz*. identification, estimation and diagnostic checking. Parameters of the tentatively selected ARIMA model at the identification stage are estimated at the estimation stage and adequacy of tentatively selected model is tested at the diagnostic checking stage. If the model is found to be inadequate, the three stages are repeated until satisfactory ARIMA model is selected for the time-series under consideration. An excellent discussion of various aspects of this approach is given in Box *et al*. (2007). Most of the standard software packages, like SAS, SPSS and EViews contain programs for fitting of ARIMA models.

Estimation of Parameters

Estimation of parameters for ARIMA model is generally done through Nonlinear least squares method. Several software packages are available for fitting of ARIMA models. To this end, in this paper, SPSS software package is used. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for ARIMA model are computed by:

$$AIC = T'\log(\sigma^2) + 2(p+q+1) \qquad (3)$$

and $$BIC = T'\log(\sigma^2) + (p+q+1)\log T' \qquad (4)$$

where T' denotes the number of observations used for estimation of parameters and σ^2 denotes the Mean square error.

Results and discussion

All-India data of inland fish production during the period 1951 to 2008 are obtained from Handbook of Fishery, Ministry of Agriculture, Government of India and the website www.indiastat.com and the same are exhibited in Fig. 1. From the total 56 data points, first 50 data points corresponding to the period 1951 to 2000 are used for building the model and remaining are used for validation purpose. A

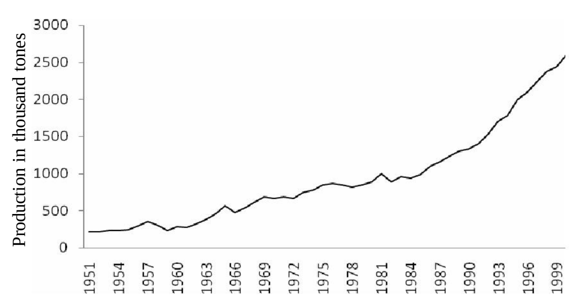

Fig. 1 Inland fish production

Table 1. Sample autocorrelation functions (Acf) and partial autocorrelation functions (Pacf) of the original and differenced series

Lag	Acf of the series	Pacf of the series	Acf of the differenced series	Pacf of the differenced series	Acf of the double differenced series	Pacf of the double differenced series
1	0.912	0.912	0.227	0.227	-0.564	-0.564
2	0.829	-0.013	0.346	0.31	0.131	-0.275
3	0.743	-0.065	0.221	0.112	-0.077	-0.219
4	0.661	-0.026	0.203	0.058	-0.075	-0.336
5	0.584	-0.018	0.328	0.231	0.106	-0.243
6	0.509	-0.035	0.242	0.107	0	-0.134
7	0.447	0.024	0.196	-0.017	-0.048	-0.194
8	0.383	-0.048	0.157	-0.02	0.182	0.122
9	0.325	-0.013	-0.101	-0.299	-0.203	0.081
10	0.277	0.018	-0.037	-0.206	0.038	-0.013

perusal of the data shows that, there is a linear trend in the inland fish production.

Fitting of ARIMA Model

From the estimated autocorrelation function (acf), reported in Table 1, it is found that it decays very slowly thereby requires to be differenced so that the resulting series depicts a pattern for a possible ARMA modelling. Further, in this situation it becomes difficult for selection of order of ARIMA model. The test for unit root proposed by Dickey and Fuller (1979) is applied for the parameter ρ in the auxiliary regression $\Delta_1 y_t = \rho y_{t-1} + \alpha_1 \Delta_1 y_{t-1} + \varepsilon_t$ The relevant null hypothesis is $\rho = 0$ and the alternative is $\rho < 0$. In the present situation

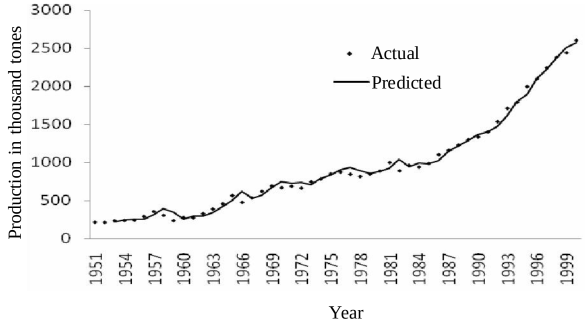

Fig. 2. Fitted ARIMA(1,2,1) model along with the data points

the estimate of ρ is 0.061 with calculated t-statistic is 5.55 which is greater than the critical value of t at 5% level of significance *i.e.* -1.95 (Franses, 1998) resulting the acceptance of null hypothesis. Thus, there is presence of unit root and so differencing is required. Usually, differencing is applied until the acf shows an interpretable pattern with only a few significant autocorrelations. On taking the second difference of the original series, it is seen that only a few acfs, reported in Table 1, are high making it easier to select the order of the model.

The appropriate model is chosen on the basis of minimum Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. Using eqs.(3) and (4), the AIC and BIC values are respectively computed and listed in Table 2. A perusal of Table 2 shows that the AIC and BIC values are minimum for ARIMA (1,2,2) but the corresponding values for ARIMA(1,2,1) model do not differ much from that of ARIMA(1,2,2). As because ARIMA (1,2,1) is more parsimonious than ARIMA (1,2,2), the ARIMA(1,2,1) model is selected for modelling and forecasting of the inland fish production in India. The estimates of

Table 2. AIC and BIC values for different ARIMA models

Criteria	ARIMA(1,2,0)	ARIMA(1,2,1)	ARIMA(2,2,0)	ARIMA(2,2,1)	ARIMA(2,2,2)	ARIMA(1,2,2)
AIC	425.87	413.22	422.97	414.32	412.93	414.27
BIC	443.34	430.69	440.44	431.79	430.41	431.75

parameters of above model are reported in Table 3.

The graph of fitted model along with data points is exhibited in Fig. 2. A perusal of Fig. 2 indicates that the fitted ARIMA(1,2,1) model is able to capture the trend present in the inland fish production in India very well.

One-step ahead forecasts of inland fish production along with their corresponding standard errors, upper confidence interval and lower confidence interval for the year, 2001 to 2008 in respect of above fitted model are reported in Table 4. The attractive feature for fitted ARIMA model is that all the forecast values except for 2008, lie within one standard error of forecasts.

The out of sample forecast of inland fish production in India for the year 2009 and 2010 have been found out as 4360 and 4610 thousand tonnes. For measuring the accuracy in fitted time series model, Mean absolute error (MAE), Mean absolute percentage error (MAPE) and Relative mean absolute prediction error (RMAPE) are computed by using the formulae given in eqs. 5, 6 and 7. The MAE, MAPE and RMAPE values for fitted ARIMA(1,2,1) model are respectively computed as 160.64, 0.044 and 4.43.

Table 3. Estimates of parameters along with their SE for fitted ARIMA(1,*2,1)* model

Parameter	Estimate	Standard error
AR1	-0.141	0.171
MA1	0.823	0.107
Constant	2.623	1.469

$$\text{MAE} = 1/8\sum_{i=1}^{8}\left|y_{t+i} - \hat{y}_{t+i}\right| \tag{5}$$

$$\text{MAPE} = 1/8\sum_{i=1}^{8}\left\{\left|y_{t+i} - \hat{y}_{t+i}\right| / y_{t+i}\right\} \tag{6}$$

$$\text{RMAPE} = 1/8\sum_{i=1}^{8}\left\{\left|y_{t+i} - \hat{y}_{t+i}\right| / y_{t+i}\right\}\times 100 \tag{7}$$

Conclusion

The ARIMA models being stochastic in nature emphasized variations in data using empirically based methods to determine the proper form

Table 4. Forecasts of inland fish production (in tonnes) for fitted models

Years	Actual	Forecasts by ARIMA(*1*,2,*1*)	SE of Forecast	Lower Confidence Limit	Upper Confidence Limit
2001	2823.0	2727.09	59.037	2609.3	2844.87
2002	2845.0	2860.68	84.010	2691.15	3030.2
2003	3126.0	2996.05	108.299	2774.81	3217.3
2004	3210.0	3134.17	132.228	2861.08	3407.27
2005	3458.0	3274.9	156.400	2948.71	3601.09
2006	3525.0	3418.25	181.033	3037.4	3799.11
2007	3755.0	3564.23	206.247	3126.96	4001.5
2008	4200.0	3712.83	232.103	3217.32	4208.33

of the model that is best suited for short-term forecasting. The more realistic forecast intervals for India's inland fish production data obtained through ARIMA approach could be of immense help to planners in formulating appropriate strategies. These in turn would also benefit the farmers in production of optimum quantities of fish. All this would ultimately result in efficient management of India's inland fish production scenario through sound statistical technique.

Acknowledgement

The authors are indebted to the Director, Central Inland Fisheries Research Institute for his encouragement.

References

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. 2007. *Time-Series Analysis: Forecasting and Control*. 3rd edition. Pearson Education, India.

Cook, R. M., Kunzlik, P. A. and Fryer, R. J. 1991. On the quality of North Sea cod stock forecasts. *ICES J. Mar. Sci*., **48**: 1-13.

Dickey, D. A. and Fuller, W. A. 1979. Distribution of the estimators for autoregressive time series with a unitroot. *J. Amer. Stat. Assoc*., **74** : 427-431.

Fox, W. W. 1970. An exponential surplus-yield model for optimizing exploited fish populations. *Trans. Am. Fish. Soc.*, **99**(1): 80-88.

Fox, W. W. 1975. Fitting the generalized stock production model by least-squares and equilibrium approximation. *U. S. Fish Bull*., **73**: 23-36.

Franses, P. H. 1998. *Time series models for business and economic forecasting*. p 82. Cambridge University Press, U. K.

Hanson, J. M. and Leggett, W. C. 1982. Empirical prediction of fish biomass and yield. *Can. J. Fish. Aquat. Sci*., **39**: 257-263.

Jensen, A. L. 1976. Time series analysis and forecasting of Atlantic menhaden catch. *Chesapeake Science*, **17**: 305-307.

Jensen, A. L. 1985. Time series analysis and the forecasting of menhaden catch and CPUE. *N. Am. J. Fish. Manage*., **5**: 78-85.

Katiha, P. K. 2000. Freshwater Aquaculture in India: Status, Potential and Constraints. pp 98-108. *In* Krishnan, M. and Birthal, P. S. (eds) *Aquacultural Development of India: Problems and Prospects*. Workshop Proceedings-7, NCAP, New Delhi.

Liu, L. M. and D. M. Hanssens 1982. Identification of multiple-input transfer function models. *Communications in Statistics - Theory and Methods*, **11**(3): 297-314.

Mendelssohn, R. 1981. Using Box-Jenkins models to forecast fishery dynamics: Identification, estimation and checking. *U.S. Fish. Bull*., **78**: 887-896.

Mendelssohn, R. and P. Cury 1987. Fluctuation of a fortnightly abundance index of the Ivorian coastal pelagic species and associated environmental conditions. *Can. J. Fish. Aquat. Sci*., **44**: 408-421.

Pella, J. J. and Tomlinson, R. K. 1969. A generalized stock production model. *Bull. Inter-Amer Trop. Tuna. Commn*, **13**: 421-496.

Prepas, E. E. 1983. Total dissolved solids as a predictor of lake biomass and productivity. *Can. J. Fish. Aquat. Sci*., **40**: 92-95.

Saila, S. B., Wigbout, M. and Lermit, R. J. 1979. Comparison of some time series models for the analysis of fisheries data. *J. Cons. int. Explor. Mer*. **39**: 44-52.

Schaefer, M. B. 1954. Some aspects of the dynamics of populations important to the management of marine fisheries. *Bull. Inter-Amer. Trop. Tuna Comm.*, **1**: 25-56.

Schaefer, M. B. 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical Pacific

Ocean. *Bull. Inter-Amer. Trop. Tuna Comm.*, **2**: 245-285.

Shepherd, J. G. 1991. Simple methods for short-term forecasting of catch and biomass. *ICES J. Mar. Sci.*, **48**: 67-78.

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. 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.

William, E. G., 1986. Systems analysis and simulation in wildlife and fisheries sciences. *John Wiley and Sons, New York, NY*, 338 pp.