



Trends in marine fish production in Tamil Nadu using regression and autoregressive integrated moving average (ARIMA) model

A. Anuja, V. K. Yadav*, V. S. Bharti and N. R. Kumar

ICAR, Central Institute of Fisheries Education, Mumbai-400061, INDIA

*Corresponding author. E-mail: vinodkumar@cife.edu.in

Received: May 9, 2016; Revised received: January 9, 2017; Accepted: April 2, 2017

Abstract: Tamil Nadu is situated in the south eastern coast of the Indian peninsula with a coastal line of 1076 km (13% of the country's coast line), 0.19 million sq.km of EEZ (9.4 % of total national EEZ) and a continental shelf of about 41,412 sq. km. This is one of the country's leading state in marine fish production and ranks third in marine fish production. In Tamil Nadu, Ramanathapuram district is a leading maritime district followed by Nagapattinam and Thoothukudi. The objective of this study was to investigate the trends in marine fish production in Tamil Nadu. Yearly fish production data for the period of 1988-1989 to 2012-2013 were analyzed using time-series method called Autoregressive Integrated Moving Average (ARIMA) model and Regression analysis (curve estimation). In our study, the developed best ARIMA model for Tamil Nadu marine fish production was found to be ARIMA (1, 1, 1) which have the minimum BIC (Bayesian Information Criterion). ARIMA model had got a slightly higher forecasting accuracy rate for forecasting marine fish production of Tamil Nadu than Regression trend analysis. The independent sample test showed there was no significant difference between the two models. The limitations of ARIMA model include its requirement of a long time series data for better forecast. It is basically linear model assuming that data are stationary and have a limited ability to capture non-stationarities and nonlinearities in series data. Both the models indicated that Tamil Nadu marine fish production has plateaued and fishermen should be encouraged to adopt sustainable fishing practices.

Keywords: ARIMA, BIC, Marine Production, Sustainable fishing, Trend line regression, Tamil Nadu

INTRODUCTION

Tamil Nadu is situated in the south eastern coast of the Indian peninsula with a coastal line of 1076 km (13 % of the country's coast line) (TNFD, 2016) and is comprising the Coramandel coast, Palk Bay, Gulf of manar and West coast. The west coast offers excellent scope for fishing throughout the year. The fishery resources spread over the 0.19 million sq.km of EEZ (9.4 % of total national EEZ) and a continental shelf of about 41,412 sq. km (<http://www.fisheries.tn.gov.in/marine-main.html>). The state ranks second in India's marine fish production with a total catch landing of 6.88 lakh tonnes in 2013-14 (CMFRI, annual report-2013-14). The state has huge fishery resources with 608 fishing villages and 13 coastal districts (Marine Fisheries Census, 2010). The total marine fisher folk population in the state was 7, 87,474 in 2010 (Tamil Nadu marine fisher folk census 2010). Ramanathapuram district was the leading maritime district in production in the state followed by Nagapattinam and Thoothukudi. The estimated marine fish production was 432265.03 tonnes in 2013-14 and the most of the fresh catch fish was consumed within the state itself. The state's fishery resources were presently exploited

jointly by traditional crafts and mechanized fishing boats. The state had exported marine fish and fish products to the level of 96429 tonnes, valued at 5316.29 crores in 2013-2014 (Tamil Nadu State Fisheries Department (TNFD), 2014). The research carried out in market had revealed the increasing demands for marine products in foreign countries. The forecasting of the marine production will be of immense help to the Export Development Authority, as it will facilitate better planning and export strategy (Srinath and Data, 1985). An effort was made in this paper to develop an Autoregressive Integrated Moving Average (ARIMA) model and Regression analysis (curve estimation) model for marine production data of Tamil Nadu for the period of 1988-1989 to 2012-2013 and to apply the same in forecasting marine production for the five leading years.

MATERIALS AND METHODS

Trend line curve estimation (Regression analysis): When the data distribution is linear, the linear equation/models can be derived by using slope-intercept formula but in the real world most data is not linear. One of the methods to handle this type of data is trend line. This is also known as a line of best fit and least

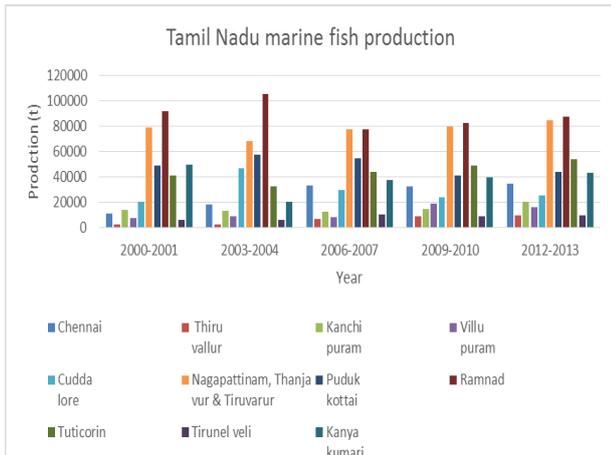


Fig. 1. District wise Tamil Nadu marine fish production (source- Tamil Nadu State Fisheries Department (TNFD), 2014).

squares line. In statistics,, trend analysis often refers to techniques for finding an underlying pattern of behaviour of observed data in a series. This pattern would be hidden by noise (error) partly or nearly completely. A simple description of these techniques is trend estimation, which is basically a formal regression analysis(https://en.wikipedia.org/wiki/Linear_trend_estimation). This is the technique of estimating future values of a time series data by extending the trend line into future. Creating a trend line and calculating its coefficients allows for the quantitative analysis of the underlying data and the ability to both interpolate and extrapolate the data for forecast purposes. The use of this approach assumes that errors have zero mean and

Table 1. Total Fish Production (tonn) in Tamil Nadu (1998-2013).

Year	Production (Tonn)
1998-1999	377483
1999-2000	373926
2000-2001	372402
2001-2002	373861
2002-2003	379214
2003-2004	381148
2004-2005	307693
2005-2006	389713.07
2006-2007	392191.32
2007-2008	393266.22
2008-2009	397117.22
2009-2010	401128
2010-2011	424823.85
2011-2012	426735.44
2012-2013	429641.24

Table 2. Constructed ARIMA model for marine fish production in Tamil Nadu.

ARIMA (p, d, q)	BIC
111	20.923
211	21.141
212	21.422
112	21.249

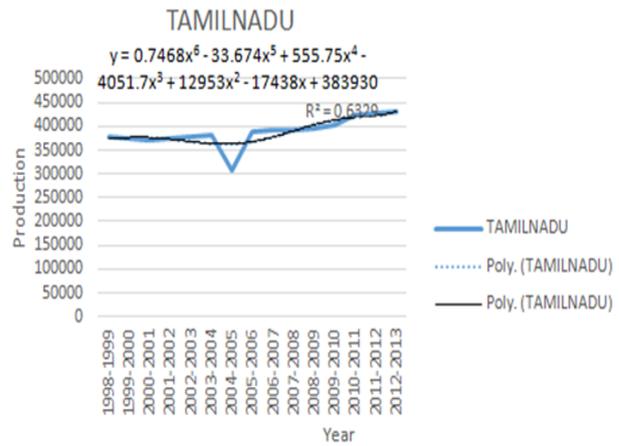


Fig. 2. Trends in Tamil Nadu marine fish production (in tonn) (1998-2013).

constant variance. Unfortunately, this is not the case most of the time.

Autoregressive integrated moving average (ARIMA) model: ARIMA model has been popular and widely chosen for modelling fisheries science’s time series data, since 1970 (Gutiérrez-Estrada *et al*, 2007, Stergiou *et al*, 1997, Bako *et al*, 2013). The ARIMA model is a linear combination of time-lagged variables and error terms. Autoregressive Integrated Moving Average (ARIMA) model was introduced by Box and Jenkins (1970) (hence also known as Box-Jenkins model) in 1960s for forecasting a variable. ARIMA method is an extrapolation method for forecasting and, like any other such method, it requires only the historical time series data on the variable under forecasting Box and Jenkins (1970). Among the extrapolation methods, this is one of the most sophisticated methods, for it incorporates the features of all such methods, does not require the investigator to choose the initial values of any variable and values of various parameters a priori and it is robust to handle any data pattern (Mandal, 2005).

Trend and prediction of time series can be computed by using ARIMA model. ARIMA (p,d,q) model is a complex linear model. *p* is order of process AR, *q* is the order of process MA and *d* is the order of

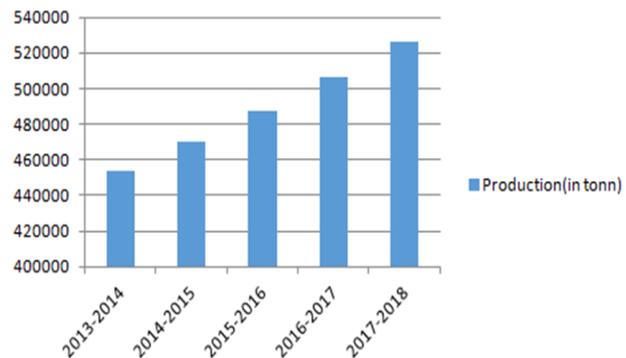


Fig. 3. Forecasted marine fish production in Tamil Nadu.

Table 3. ARIMA Model Parameter Estimation.

			ARIMA Model Parameters				
				Estimate	SE	t	Sig.
Tamil Nadu-Model_1	Tamil Nadu	Natural Log	Constant	-0.014	0.019	-0.770	0.459
			AR Lag 1	-0.140	0.380	-0.368	0.721
			Difference	1.00			
			MA Lag 1	0.983	5.819	0.169	0.869
			Numerator Lag 0	0.003	0.002	1.410	0.189
	Year	No Transformation					

difference. There are three parts (they do not have to contain always all of these): AR (Autoregressive)–linear combination of the influence of previous values; I Integrative) – random walk; MA (Moving average) – linear combination of previous errors. These models are very flexible, quite hard for computing and for the understanding of the results.

They are basically linear models assuming that data are stationary and have a limited ability to capture non-stationarities and nonlinearities in series data (Khashei and Bijari, 2010 and 2011). If the data are highly non-linear & have fuzziness, advance methods like artificial neural network (ANN), fuzzy inference system etc. can be taken for better prediction of marine fish production.

RESULTS AND DISCUSSION

Tamil Nadu marine fish production: The fish catch records during the year 1998-2004 showed a slow decline in growth irrespective of increased fishing capacity during that period and the reason maybe mainly due to the decline in oil sardine production. Further, by the year 2004-2005, there was heavy decline in marine fish production in the state due to the occurrence of the natural catastrophe “Tsunami” in the state by the end of the 2004 year. However, the year

2005-2010 showed an encouraging revival in fish production with unexpected growth in oil sardine production due to targeted fishing for the specific fish species. Table 1 shows the total marine fish production data in Tamil Nadu (1998-2013).

The district wise marine fish production during the period 2000-2013 in the state had been led by Ramanathapuram district followed by Nagapattinam and Thoothukudi as shown in Fig 1.

Trends in Tamil Nadu marine fish production:

Using the above Table 1 data, a trend line and forecast of marine fish production was constructed as shown in Fig 2 using regression analysis (curve estimation). The figure showed the trends in fish production in Tamil Nadu for the indicated respective periods of time. The severities in fluctuations in production were pronounced and were underscored by the best fit given by the polynomial of degree 6. The sudden drop in production during 2004-05 was due to tsunami while 2010-2013 production trends shows a plateau indicating the stagnant production during that particular period of time.

The forecasted marine fish production for the period 2013-2018 was shown in figure 3 using regression analysis (curve estimation) and the graph indicates the marginal increasing trend in production.

Tamil Nadu marine fish production by ARIMA model:

As ARIMA model requires a large data set, the data for Tamil Nadu marine fish production during 1998-2013 was used from Table 1 so as to fit the model. ARIMA model for any variable involves three steps such as identification, estimation and verification. ARIMA model was estimated only after transforming

Table 4. Forecasted Tamil Nadu marine fish production (in tonn) by using regression analysis and ARIMA model.

Year	Forecasted by regression	Forecasted by ARIMA
1998-1999	378526.4	
1999-2000	373913.6	376090.3
2000-2001	370678.3	373332.3
2001-2002	368820.6	372323.2
2002-2003	368340.4	372854.5
2003-2004	369237.9	375014.7
2004-2005	371512.9	378663.8
2005-2006	375165.5	382488.7
2006-2007	380195.6	376321.9
2007-2008	386603.4	384551.6
2008-2009	394388.7	393546.5
2009-2010	403551.6	403144.3
2010-2011	414092.1	413845.2
2011-2012	426010.1	424542.3
2012-2013	439305.7	439507.9
2013-2014	453979	455528.8
2014-2015	470029.8	471164
2015-2016	487458.1	490564.1
2016-2017	506264.1	512047.2
2017-2018	526447.6	536124.6

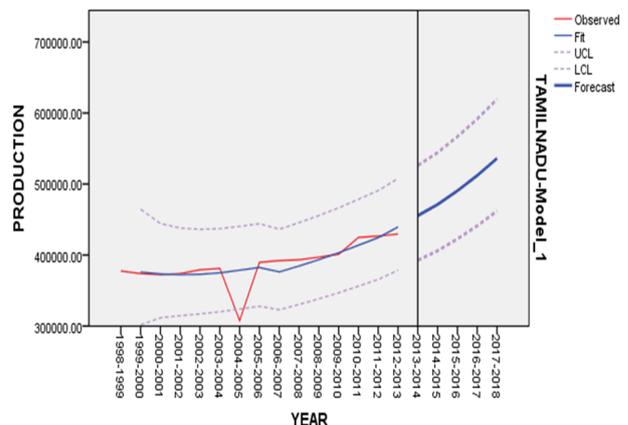


Fig. 4. Tamil Nadu marine fish production (in tonn) by ARIMA model.

Table 5. Independent sample test of forecasted value with regression analysis and ARIMA model.

		Independent Samples Test									
		Levene's Test for Equality of Variances				t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
									Lower	Upper	
Forecast Equal variances assumed	Equal variances not assumed	0.001	0.970	-0.143	36	0.887	-2403.17904	16788.2038	-36451.234	31644.876	
				-0.143	35.989	0.887	-2403.17904	16788.2038	-36451.582	31645.224	

the variable under forecasting into a stationary series. The stationary series is the one whose values vary over time only around a constant mean and constant variance. If the given data is non-stationary in mean, it is corrected through appropriate differencing of the data; the next step is to identify the values of p and q. The ACF and PACF of the order of p and q can at most be 1. We constructed four tentative ARIMA models (Table 2) and chose that model which has minimum BIC (Bayesian Information Criterion). The models and corresponding BIC values was listed in Table 2.

Model parameters were estimated (Table 3) using SPSS package, and with low BIC values, the most suitable model was ARIMA (1, 1, 1).

$$y_t = C + \phi_1 y_{t-1} - \phi_1 y_{t-2} + e_t - \theta_1 e_{t-1}$$

Where,

C = Constant term;

ϕ_j = jth auto regression parameter;

θ_1 = Moving average parameter

e_t = error term at time t

So the fitted ARIMA model for fish production is

$$y_t = -.014 - .140y_{t-1} - \phi_1 y_{t-2} + e_t - 0.983e_{t-1}$$

The forecasted marine fish production in the state as shown in figure 4 using ARIMA model indicates the increasing trend in production during the selected period of time.

Forecasted marine fish production values of Tamil Nadu using regression analysis (curve estimation) and ARIMA model are shown in Table 4.

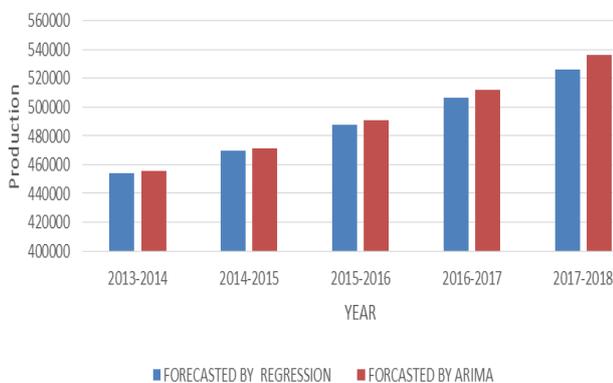


Fig. 5. Comparative forecasted value (in ton) with regression analysis and ARIMA model.

In time series forecasting, the forecasting accuracy of a model is commonly measured in terms of Mean Square Error (MSE) or in terms of Average Error. Lower the MSE or average error, better the forecasting method. MSE is defined as

Mean Square Error =

$$\frac{\sum_{i=1}^n (actual\ value_i - forecasted\ value_i)^2}{n}$$

and forecasting error as

Forecasting error (in percent) =

$$\frac{|forecasted - actual\ value|}{actual\ value} \times 100$$

Average forecasting error (in percent) =

$$\frac{sum\ of\ forecasting\ error}{numbers\ of\ errors}$$

With the above comparison of actual production of marine fish production of Tamil Nadu with the forecasted production by regression analysis and ARIMA model, average forecasting error (%) are 3.09 and 3.03, respectively. Hence, the ARIMA model can get a slightly higher forecasting accuracy rate for forecasting marine fish production of Tamil Nadu than Regression trend analysis.

The comparative forecasted value analysed with regression analysis and ARIMA model for the year 2013-2018 are being displayed below in figure 5.

Independent samples test of forecasted value by Regression and ARIMA: The independent samples test (Table 5.) showed that there was no significant difference between the two models; regression and ARIMA models.

Conclusion

ARIMA model offers a better technique for predicting the magnitude of any variable than the regression analysis. Its limitations include its requirement of a long time series data for better forecast. They are basically linear models assuming that data are stationary and have a limited ability to capture non-stationarities and nonlinearities in series data. In our study, the developed model for Tamil Nadu marine fish production was found to be ARIMA (1, 1, 1). ARIMA model had got slightly higher forecasting accuracy rate for forecasting marine fish production of Tamil Nadu than

Regression trend analysis. The independent samples test showed that there was no significant difference between the two models; regression and ARIMA models. From the forecast available by using the developed ARIMA model and regression analysis, indicated that Tamil Nadu marine fish production has plateaued. From this we can suggest that Fishermen should be encouraged to adopt sustainable fishing practices. It is necessary to increase awareness on sustainable fisheries. The focus of the Government should be towards promotion of sustainable fisheries rather than on welfare aspects. If the relationships between different variables in fisheries are highly non-linear and also data have fuzziness, advance methods like artificial neural network (ANN), fuzzy inference system etc. can be taken for better prediction of marine fish production.

REFERENCES

- Bako, H. Y., Rusiman, M. S., Kane, I. L., and Matias-Peralta, H. M. (2013), Predictive modelling of pelagic fish catch in Malaysia using seasonal ARIMA models, *Agriculture, Forestry and Fisheries*, 2(3): 136–140
- Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day.
- Central Marine Fisheries Research Institute (CMFRI) annual report, 2013- 2014
- CMFRI (2010). *Marine Fisheries Census, Part II.4, Tamil Nadu*
- Gutiérrez-Estrada, J. C., Silva, C., Yáñez, E., Rodríguez, N. and Pulido-Calvo, I. (2007). Monthly catch forecasting of anchovy *Engraulis ringens* in the north area of Chile: non-linear univariate approach. *Fisheries Research*, 86 (2-3):188–200
- Khashei, M. and Bijari, M. (2010). An artificial neural network (p, d, q) model for time series forecasting. *Expert Systems with Applications*, 37(1): 479–489
- Khashei, M. and Bijari, M. (2011). A new hybrid methodology for nonlinear time series forecasting. *Modelling and Simulation in Engineering*, 2011, Article ID 379121, 5 pages
- Mandal, B. N. (2005). Forecasting sugarcane production in India with ARIMA model, Pp.1-13. Retrieved January 9, 2017 from <https://www.researchgate.net/publication/256079794>
- Srinath, M. and Datta, K. K. (1985). Forecasting marine product exports time-Series analysis. *Indian Journal of Fisheries*, 32(2): 264-267
- Stergiou, K. I., Christou, E. D. and Petrakis, G. (1997). Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. *Fisheries Research*, 29(1): 55–95
- Tamil Nadu State Fisheries Department (TNFD) (2014). 'Fisheries at a Glance, Marine-2013-14', Pp. 1-2. Retrieved January 9, 2017 from <http://www.fisheries.tn.gov.in/marine-main.html>
- Tamil Nadu State Fisheries Department (TNFD) (2016). 'Marine Fisheries Development', Pp. 1. Retrieved January 9, 2017 from <http://www.fisheries.tn.gov.in/marine-main.html>