

[Research]

## Forecasting the catch of kilka species (*Clupeonella* spp.) using Time Series SARIMA models in the Southern Caspian Sea

Amiri K.<sup>1</sup>, Shabanipour N.<sup>1,2\*</sup>, Eagderi S.<sup>3</sup>

1. Department of Biology, Faculty of Science, University of Guilan, Rasht, Iran

2. Department of Marine Sciences, The Caspian Sea Basin Research Centre, University of Guilan, Rasht, Iran

3. Department of Fisheries, Faculty of Natural Resources, University of Tehran, Karaj, Iran

\* Corresponding author's E-mail: shabani@guilan.ac.ir

(Received: May 31. 2018 Accepted: Nov. 26. 2018)

### ABSTRACT

Fisheries management receives assistance by prediction of events to evaluate fluctuating values for a target species to formulate proper policies and actions particularly for threatened and endangered species. This study aimed to predict 7 years Catch Per Unit Effort (CPUE) of kilka fishes as at-risk population in southern regions of the Caspian Sea. The former catch data from the Fisheries Organization of Iran (IFO) archives (1997 to 2014) were analyzed using ARIMA and SARIMA models. The data were divided into four parts (quarters) addressing one-fourth of a year to represent time and expressed as "Q". According to periodic changes of ACF and PACF indices, seasonal ARIMA (SARIMA) models were used. The appropriate SARIMA models were examined using BIC, RMSE, R<sup>2</sup>, MSE and Ljung-Box indices. SARIMA (0, 1, 1) × (0, 1, 1)<sub>4</sub> process was the selected final model which met the criterion of model parsimony according to BIC of 31.91, RMSE of 7195193, MAE of 4372178, R<sup>2</sup> of 0.82 and Ljung-Box index < 0.05. Based on selected SARIMA model, the forecasts indicated that if the fishing fleet and efforts remain at the present level, the performance of kilka fishing will likely have gentle rise by 2021.

**Key words:** SARIMA model, Kilka, Time series forecasting, Fishing effort, South Caspian Sea.

### INTRODUCTION

The Caspian Sea (36° to 47° N and 46° to 54° E) is a landlocked body of water on the Euro-Asian continent. Modern autochthonous Caspian fauna evolved from limited number of marine species since 1.8 million years ago as an isolated brackish water body without competition enforcement by marine species (Karpinsky 2002).

The Caspian Sea kilka group (*Clupeonella* spp.) are small pelagic fishes belonging to the family Clupeidae including common kilka (*Clupeonella cultriventris caspia* Bordin, 1904), big eye kilka (*C. grimmi* Kessler, 1877) and anchovy kilka (*C. engrauliformis* Svetovidov, 1941). The three

mentioned species also constitute the main catch of the sea comprising 80% of total catches. They are crucial high protein part of the food web in the Caspian Sea. The Iranian kilka related firms impose important economic effects on various feed producers for poultry, livestock and aquaculture industries. However, the sustainable management of kilka stocks and catches is not only vital to the Iranians but also to the countries benefited by the Caspian Sea such as Azerbaijan, Russia and Turkmenistan (Strukova *et al.* 2015).

Over the past two decades, the species composition on one part and density of the Caspian Kilka on the other part have been

negatively changed (Karimzadeh *et al.* 2010). According to Iranian Fisheries Organization (IFO) reports, the total catch of kilka decreased by almost 50% ranging from 45000 ton in 2001 to 22000 ton in 2014 relating mainly to anthropogenic pressures. Pollution, domestic and industrial run-off, development of vast oil and gas fields, uncontrolled fishing and inefficient management has escalated the deterioration of the Caspian Sea ecosystem and trophic chains (Ivanov 2000).

Forecasting based on historical time series data, has become an efficient tool for future fisheries planning. Efficient models can provide accurate operational forecasts of annual commercial landings in coastal waters (Stergiou & Christou 1996). Planners can predict commercial landings for the next year and season using time series approaches where the data are updated periodically (Czerwinski *et al.* 2007). Such estimations help to predict fluctuations of target species biomass and make realistic policies for effective fisheries management (Stergiou & Christou 1996). Time series analyses, such as Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model, has been used for numerous scientific fields including fishery science such as short-term forecasting of landings and also to evaluate forecast efficiencies (Georgakarakos *et al.* 2006; Czerwinski *et al.* 2007). Where periodic components are included in this model, it is called Seasonal Autoregressive Integrated Moving Average (SARIMA). The SARIMA model deals with seasonality in a more implicit manner, while ARIMA model is deficient in dealing with seasonal data. SARIMA model acts better than ARIMA when the seasonal pattern is both strong and stable over time. Several studies have employed SARIMA model to forecast fish catch amounts around the world (Christou 1996; Prista *et al.* 2011; Bako *et al.* 2013; Lazaro & Lazaro Jere 2013; Kim *et al.* 2015). The total catch of the Caspian Sea kilka has declined to a critical point in recent years. To recover the stock and achieve sustainable yield, fishery policies need

a reform and sufficient information about future situation of kilka catches.

Therefore, the present study tried to forecast the monthly and also seasonal CPUE (Catch Per Unit Effort) of the kilka fishes in the southern part of the Caspian Sea employing SARIMA model. The efficiency of the forecasting techniques applicable to the Caspian fisheries was discussed.

## MATERIALS AND METHODS

### Data collection and analysis

The analyses were based on the daily CPUE data obtained from fishing vessels during 1997 to 2014 at two main Kilka fishing areas in the southern part of the Caspian Sea i.e. Anzali (37°28'N, 49°25'E) and Babolsar (36°42'N, 52°39'E) ports (Fig. 1). CPUE is calculated as the catch divided by the effort. The CPUE data collected by the Iranian Fisheries Organization (IFO) were divided by four quarters. A quarter refers to one-fourth of a year and is typically expressed as "Q". The four quarters that made the year were: March to June (Q1); June to September (Q2); September to December (Q3); December to March (Q4) representing four seasons in Iran. CPUE of kilka ranged between 93802 kg to 14239979 kg, reaching its highest value of 30993544 kg in winter (Q4) of 1998 and falling to 93802 kg in spring (Q1) of 2014.

### Modeling

Box and Jenkins (1976) presented Seasonal Autoregressive Integrated Moving Average (SARIMA) of a time series model as (1):

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \dots - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad [1]$$

Where  $X_t$  is the value of a stationary time series  $\{X_t\}$  at time  $t$ . Clearly  $p$  and  $q$  are integers and  $\{\varepsilon_t\}$  is a white noise process. The  $\alpha$ 's and  $\beta$ 's are constants such that model (1) is both stationary and invertible.

If  $q = 0$  then (1) is an autoregressive model of order  $p$ , denoted by AR ( $p$ ). If  $p = 0$  the model is a moving average model of order  $q$ , denoted by MA ( $q$ ). Model (2) may be put as:

$$A(L)X_t = B(L)\epsilon_t \quad [2]$$

where  $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$  and  $B(L) = 1 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q$  and  $L^k X_t =$

$X_{t-k}$ .  $A(L)$  and  $B(L)$  are the autoregressive (AR) operator and the moving average (MA) operator respectively.

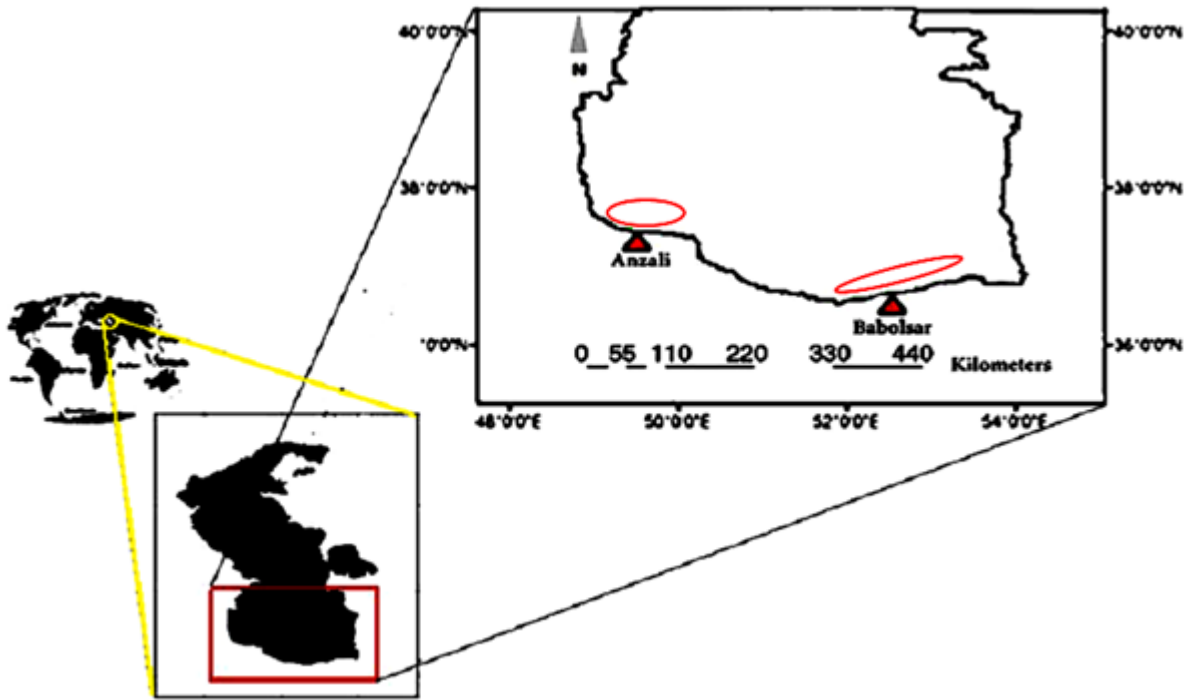


Fig 1. Approximate kilka fishing areas (red ovals) in the South Caspian Sea (Amiri et al. 2017).

If the time series  $\{X_t\}$  is non-stationary, Box and Jenkins proposed that differencing of the series to an appropriate degree might make it stationary. Suppose  $d$  is the least positive integer such that the  $d^{\text{th}}$  difference of  $\{X_t\}$  denoted by  $\{\nabla^d X_t\}$  is stationary where  $\nabla = 1 - L$ .  $\{X_t\}$  is said to be  $I(d)$ , and replacement of  $X_t$  by its  $d^{\text{th}}$  difference in (1) yields an autoregressive integrated moving average of order  $p, d, q$ , denoted by  $ARIMA(p, d, q)$ , in the original series. If  $\{X_t\}$  is in addition seasonal, Box and Jenkins further proposed that, in order to capture the seasonality, it might be modeled by (3):

$$A(L)\Phi(L^s)\nabla^d \nabla_s^D X_t = B(L)\theta(L^s)\epsilon_t \quad [3]$$

where  $s$  is the seasonality period,  $\Phi(L) = 1 + \phi_1 L + \phi_2 L^2 + \dots + \phi_p L^p$  and  $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$  are the seasonal AR and MA operators respectively, the  $\phi$ 's and  $\theta$ 's being constants such that the entire model is

stationary and invertible.  $\nabla_s^D$  represents the  $D^{\text{th}}$  seasonal difference operator. Model (2) is called a seasonal autoregressive integrated moving average model of order  $p, d, q, P, D, Q, s$  denoted by  $SARIMA(p, d, q) \times (P, D, Q)^s$ . Box-Jenkins modeling involves first of all the determination of the orders in (1) and (2). The AR orders  $p$  and  $P$  are determined by the non-seasonal and the seasonal cut-off lags of the autocorrelation function (ACF). Their counterparts,  $q$  and  $Q$  may be estimated by the non-seasonal and the seasonal cut off lags of the partial autocorrelation function (PACF). It is sufficient to choose the differencing orders  $d$  and  $D$  such that their sum is at most for stationary to be attained. The seasonality period often suggestive naturally or by an inspection of the series. Stationary and the best alternative models compared according to ADF, smallest BIC, RMSE & MSE maximum R-square and the  $p$ -values above 0.05 of Ljung-Box Test.

### Stationary test

A test of stationary (or non - stationary) that has become widely popular over the past several years is the unit root test. This is the test that is used to carry out or to know the order of integration. It is important to know the order of integration of non-stationary variables, so they may be different before being included in a regression equation. The most common unit root tests are ADF test (Dickey & Fuller 1979).

### Ljung-Box Test

Ljung and Box (1978) proposed a Q-Test called Ljung-Box test which is commonly used in linear models following Box-Jenkins methodology. This test is applied to the residuals of a fitted model, not the original series, and in such applications the hypothesis to be tested is that the residuals from the model have no autocorrelation. Perhaps it performs a lack-of-fit hypothesis test for model misspecification based on the Q-statistic given in (4):

$$Q = N(N + 2) \sum_{j=1}^L \frac{\rho_j^2}{(N-j)} \quad [4]$$

Where N = sample size, L = number of autocorrelation lags included in the statistic, and  $\rho_j^2$  is the squared sample autocorrelation at lag j. Under the null hypothesis of no serial correlation, the Q-test statistic is asymptotically Chi-Square distributed. The p-values above 0.05 indicate the acceptance of the null hypothesis of model accuracy under 95% significant levels (Fadhilah & Ibrahim 2012).

### Choice of best model

First, the stationary and seasonality have been addressed, and then the order (the p and q) of the autoregressive and moving average terms were found. The primary tools for doing so are the autocorrelation plot and the partial autocorrelation plot. However, according to Box & Jenkins (1976) the model should be parsimonious, having as few parameters as possible and fulfill all the diagnostic checks. The BIC, RMSE, MAE, R-square, and Ljung-Box

test suggested that the parsimony criteria of the model building as information criteria for the purpose of selecting an optional model fits to a given data. Also the model adequacy by examining the sample autocorrelation as function of the residual (ACF) and the sample partial autocorrelation as function of the residual (PACF) was checked. We can conclude that the model is adequate if there are no spikes in the ACF and PACF.

### RESULTS

The first step in developing a Box-Jenkins model is to determine if the series are stationary. So that, Root test of stationary ADF was carried out. The P-value of ADF test was 0.03 and smaller than 0.05. As the result, original time series was not stationary (Fig. 2 a,b and c) and quarterly differencing was necessary (Fig. 3 a,b and c).

The ACF had spikes of multiple of 4, 8, 12, 16, 20 and 24 (Fig. 3 a, b and c), indicating that seasonal differencing was necessary. The seasonal difference on the stationed quarterly series (ADF test P-value>0.05) was used to compute various SARIMA models (Fig. 3 & 4 a, b and c).

### Model selection

Nine candidate models were selected (Table 1). The best model was found as SARIMA (0, 1, 1) × (0, 1, 1)<sup>4</sup> according to BIC, RMSE, MAE and R<sup>2</sup> indices.

The Ljung-Box statistic indicated that there was no significant departure from white noise for the residuals as the P-values of the test statistic clearly exceeds the 5% significant level for all lag orders (Table 2; Fig. 5). The estimation of the model as summarized in Table 3 yields (5). Seasonal prediction was done according to the selected model for 2015 to 2021 (Table 4; Fig. 6). The actual and fitted data are indicated by solid and dash lines respectively. The forecasted CPUE (in kg) are shown by the thicker line in yellow shaded area.

The bounded dash line shows 80% (lower line) and 95% (upper line) prediction intervals respectively.

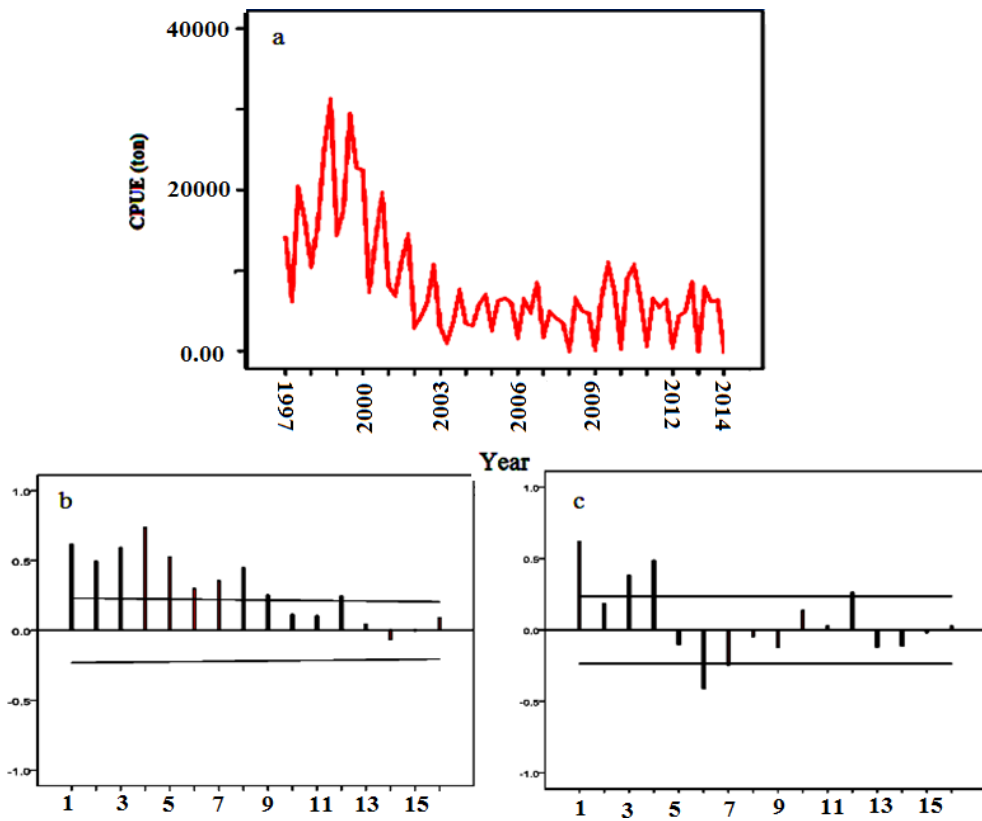


Fig 2. Time series of CPUE (a) and correlograms (ACF and PACF) (b,c) for original data set.

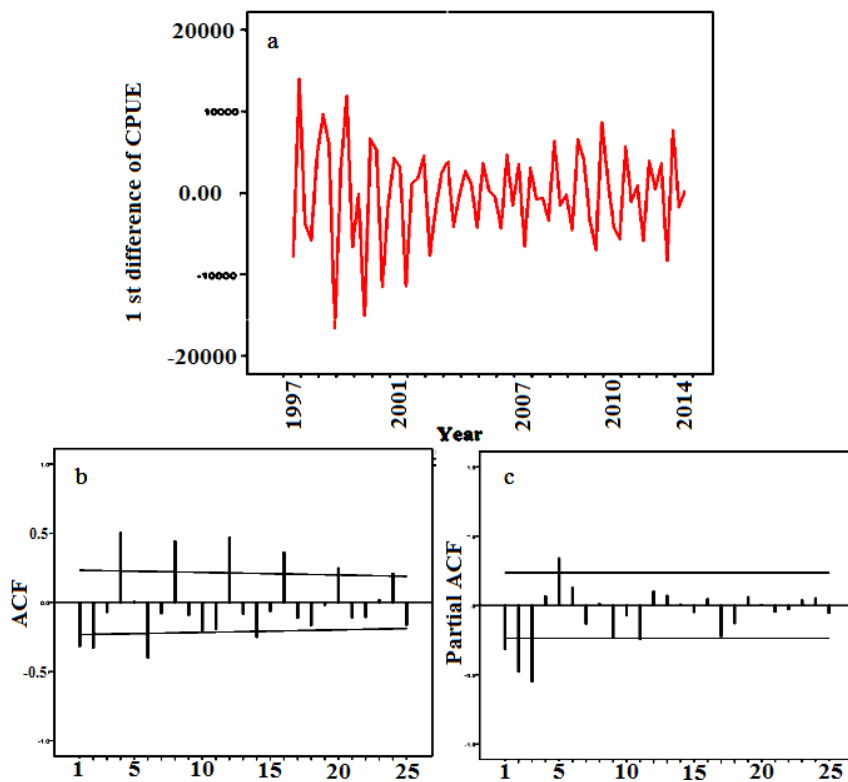


Fig. 3. Kilka CPUE (ton) after 1<sup>st</sup> difference (a) and related correlograms (ACF and partial ACF) (b,c).

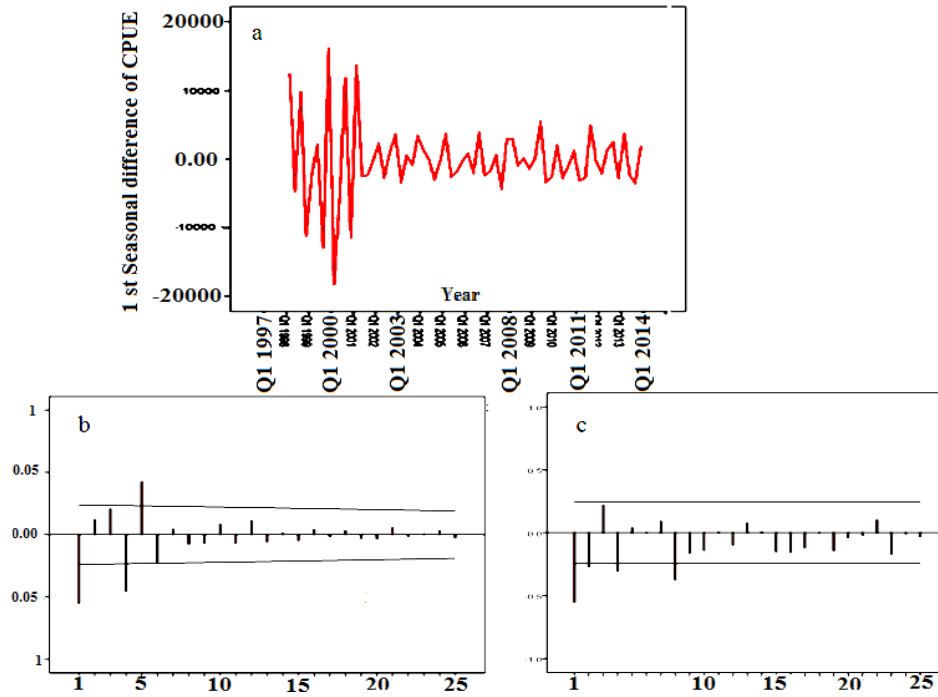


Fig.4. Kilka CPUE (ton) after 1<sup>st</sup> Seasonal (Q) difference (a) and related correlograms (ACF and partial ACF) (b,c).

Table 1. Results of several models for model identification.

| Model - Type             | BIC    | RMSE    | MAE     | R <sup>2</sup> |
|--------------------------|--------|---------|---------|----------------|
| SARIMA (1,1,1) (1,1,1)4  | 32.017 | 7339396 | 4520576 | 82             |
| SARIMA (1,1,1)(1,1,0)4   | 32.04  | 7708253 | 4905617 | 80             |
| *SARIMA (0,1,1) (0,1,1)4 | 31.91  | 7195193 | 4372178 | 82             |
| SARIMA (1,1,0) (1,1,0)4  | 32.84  | 9651076 | 6633747 | 68             |
| SARIMA (1,1,0) (0,1,1)4  | 32.23  | 8738546 | 5562030 | 74             |
| SARIMA (1,1,0)(1,1,1)4   | 30.73  | 8748802 | 5363459 | 74             |
| SARIMA (0,1,1)(1,1,0)4   | 32.37  | 9390823 | 5678808 | 70             |
| SARIMA (1,1,1)(0,1,1)4   | 32.20  | 8598984 | 4705856 | 74             |
| SARIMA (0,1,1)(1,1,1)4   | 32.29  | 8701065 | 4871968 | 74             |

\* = Best model

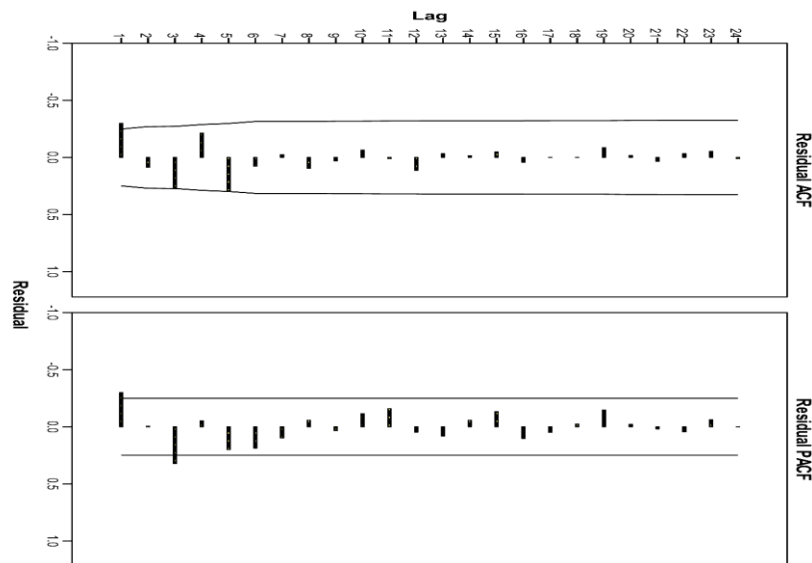


Fig. 5. Noise residual ACF and PACF.

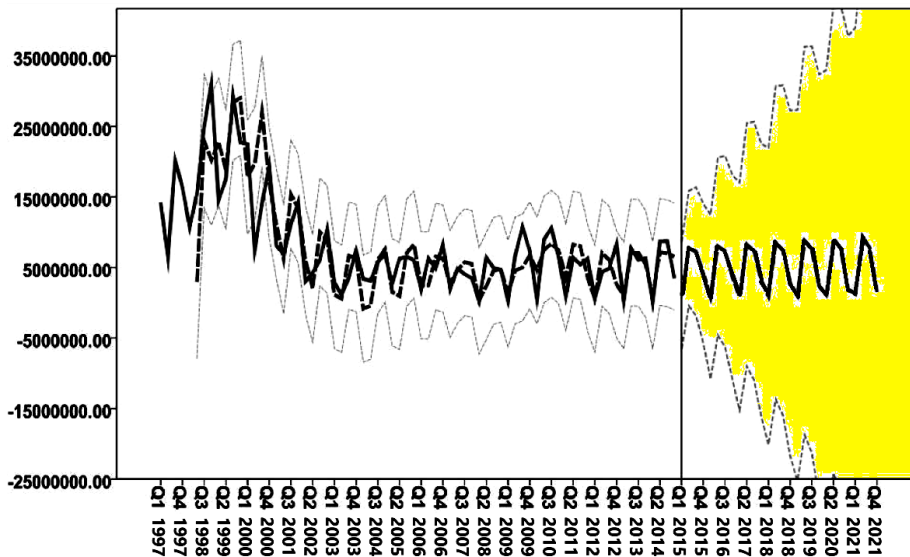
**Table 2.** Estimates of parameters for SARIMA (0, 1, 1) × (0, 1, 1)<sup>4</sup>.

| Model parameters   | Estimate | SE    | T-Ratio | Sig.   |
|--------------------|----------|-------|---------|--------|
| MA 1               | 0.532    | 0.112 | 4.740   | 0.0000 |
| Non-seasonal Diff. | 1        |       |         |        |
| SMA 1              | 0.623    | 0.110 | 5.682   | 0.000  |
| Seasonal Diff.     | 1        |       |         |        |
| Liung-Box Q test   | 0.369    |       |         |        |

**Table 3.** Estimation of the SARIMA (0, 1, 1) × (0, 1, 1)<sup>4</sup> model.

| Model parameters | Estimate | SE    | Sig.   |
|------------------|----------|-------|--------|
| MA 1             | -0.553   | 0.119 | 0.0000 |
| MA 4             | -0.456   | 0.116 | 0.0000 |
| MA 5             | 0.425    | 0.116 | 0.0000 |

$$Kilka\ seasonal\ CPUE = -0.553\varepsilon_{t-1} - 0.456\varepsilon_{t-4} + 0.425\varepsilon_{t-5} + \varepsilon_t \quad [5]$$



**Fig. 6.** The actual and predicted amounts of seasonal kilka CPUE by SARIMA (0, 1, 1) × (0, 1, 1)<sup>4</sup> in southern part of the Caspian Sea.

**Table 4.** Actual and forecasted seasonal CPUE (kg) for 2014 to 2021.

| Season | Year              | Spring  | Summer  | Autumn  | Winter  |
|--------|-------------------|---------|---------|---------|---------|
|        | 2014 (actual)     | 1126211 | 7188542 | 7012349 | 6546864 |
|        | 2015 (forecasted) | 1042650 | 7707830 | 7244714 | 4137857 |
|        | 2016 (forecasted) | 805873  | 8005441 | 7282515 | 3671137 |
|        | 2017 (forecasted) | 866468  | 8247064 | 7350819 | 3213749 |
|        | 2018 (forecasted) | 941577  | 8501932 | 7432679 | 2769841 |
|        | 2019 (forecasted) | 1030186 | 8770295 | 7528036 | 2339429 |
|        | 2021 (forecasted) | 1132290 | 9052153 | 7636887 | 1922513 |
|        | 2021 (forecasted) | 1247889 | 9347507 | 7759234 | 1519091 |

**CONCLUSION**

This study aimed to identify a time series model to forecast the kilka CPUE in the southern part of the Caspian Sea from 1997 to 2014 employing of Box-Jenkins fundamental approach. The model was developed in three

stages. In the first stage, the model was identified, where the series was not non stationary at level form based on the result provided by ADF test, correlogram and time plot. It was found out that the series was stationary at the 1<sup>st</sup> difference. More so,

seasonal difference was made due to significant spikes at certain lags of the quarterly stationary series. Based on simultaneous criteria selection for BIC and RMSE, SARIMA (0, 1, 1) × (0, 1, 1)<sup>4</sup> was the best model to fit the data.

Then, CPUE was forecasted indicating that the kilka CPUE will have gentle rise by 2021 (Table. 3; Fig. 6).

The forecast indicated that in current fishing effort, the CPUE of southern part of the Caspian Sea kilka will be almost stable for few coming years ahead and the present seasonal condition will be retained. In addition, the first difference ACF correlogram showed that Q4 fishing season in Iran established the key role in kilka CPUE fluctuations.

#### ACKNOWLEDGEMENTS

We would like to thank Fisheries Statistics and Economy Group of Iran Fisheries Organization especial Sabah Khorshidi for providing the data, which has been used in this study.

The authors wish to thank Ette Etuk (Rivers state University, Nigeria) and Hadi Poorbagher (University of Tehran) for their review and scientific guidance.

#### REFERENCES

- Amiri, K, Shabanipour, N & Eagderi S 2017, Using kriging and co-kriging to predict distributional areas of Kilka species (*Clupeonella* spp.) in the southern Caspian Sea. *International Journal of Aquatic Biology*, 5: 108-113.
- Bako, HY, Rusiman, MS, Kane, IL, Matias & Peralta, HM 2013, Predictive modeling of pelagic fish catch in Malaysia using seasonal ARIMA models. *Agriculture, Forestry and Fisheries*, 2: 136-140.
- Czerwinski, IA, Gutierrez Estrada, JC & Hernando Casal, JA 2007, Short-term forecasting of halibut CPUE: Linear and non-linear univariate approaches. *Fisheries Research*, 86: 120-128.
- Dickey, DA & Fuller, WA 1979, Distributions of the estimators for autoregressive Time series with a unit root. *Journal of the American Statistical Association*, 74: 427-431.
- Fadhilah, Y & Ibrahim, LK 2012, Modeling monthly rainfall time series using ETS state space and sARIMA models. *International Journal of Current Research*, 4: 195-200.
- Fazli, H, Zhang, CI, Hay, DE, Lee, CW, Janbaz, AA & Borani MS 2007, Population ecological parameters and biomass of anchovy kilka (*Clupeonella engrauliformis*) in the Caspian Sea. *Fisheries Science*, 73: 285-294.
- Georgakarakos, S, Koutsoubos, D & Valavanis, V 2006, Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters. *Fisheries Research*, 78: 55-71.
- Ivanov, PI 2000, Biological resources of the Caspian Sea. KaspNIRKH, Astrakhan, 130 P.
- Karimzadeh, G, Gabrielian, B & Fazli, H 2010, Population dynamics and biological characteristics of kilka species (Pisces: *Clupeidae*) in the south eastern coast of the Caspian Sea. *Iranian Journal of Fisheries Sciences*, 9: 422-433.
- Karpinsky, MG 2002, Ecology of the Benthos of the Middle and Southern part of the Caspian. VNIRO, Moscow. 283 P.
- Kim, JY, Jeong, HC, Kim, H & Kang, S 2015, forecasting the monthly abundance of anchovies in the South Sea of Korea using a univariate approach. *Fisheries Research*, 161: 293-302.
- Lazaro, M & Lazaro Jere, WW 2013, The status of the commercial chambo (*Oreochromis Species*) fishery in Malawi: A time series approach. *International Journal of Science and Technology*, 3: 322-327.
- Prista, N, Diawara, N, José Costa, M & Jones, C 2011, Use of SARIMA models to assess data-poor fisheries: a case study with a sciaenid fishery off Portugal. *Fishery Bulletin*, 109: 170-185.
- Stergiou, KI & Christou, ED 1996, Modeling and forecasting annual fisheries catches: comparison of regression, univariate and



- multivariate time series methods. *Fisheries Research*, 25: 105-138.
- Strukova, E, Guchgeldiyev, O, Evans, A, Katunin, D, Khodorevskaya, R, Kim, Y, Akhundov, M, Mammadli, T, Shahivar, R & Muradov, O 2015, Exploitation of the Caspian Sea bioresources (with focus on economics of bioresources utilization). *The Handbook of Environmental Chemistry*, pp. 1-44.

## پیش‌بینی صید کیلکا ماهیان (*Clupeonella spp.*) با استفاده از سری‌های زمانی مدل‌های SARIMA در جنوب دریای خزر

امیری ک.ا.<sup>۱</sup>، شعبانی پور ن.<sup>۲</sup>، ایگدري س.<sup>۳</sup>

۱- گروه زیست‌شناسی، دانشکده علوم، دانشگاه گیلان، رشت، ایران

۲- گروه علوم دریایی، پژوهشکده حوضه آبی دریای خزر دانشگاه گیلان، رشت، ایران

۳- گروه شیلات، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران

(تاریخ دریافت: ۹۷/۰۳/۱۰ تاریخ پذیرش: ۹۷/۰۹/۰۵)

### چکیده

مدیریت شیلاتی از پیش‌بینی وقایع در جهت ارزیابی مقادیر در نوسان خصوصاً در رابطه با گونه‌های نشاندار یاری طلبیده تا سیاست‌ها و اقدامات مقتضی خصوصاً برای گونه‌های در خطر اتخاذ کند. مطالعه حاضر با هدف پیش‌بینی ۷ سال صید در واحد تلاش (CPUE) کیلکا ماهیان به عنوان جمعیت ریسک‌پذیر در نواحی جنوبی دریای خزر انجام شد. اطلاعات صید بین ۱۹۹۷ الی ۲۰۱۴ از آرشیو سازمان شیلات ایران با استفاده از مدل‌های ARIMA و SARIMA تجزیه و تحلیل شد. اطلاعات به چهار یک قسمت نشان‌دهنده یک چهارم هر سال بوده و با "Q" نشانه گذاری شد. طبق تغییرات زمانی ACF و شاخص‌های PACF، مدل ARIMA (SARIMA) فصلی به کار گرفته شد. مدل‌های ARIMA مناسب با استفاده از شاخص‌های BIC، RMSE، R2، MSE و Ljung-Box آزمایش شدند. فرایند SARIMA (۱، ۱، ۰) × (۱، ۱، ۰) ۴ مدل نهایی انتخاب شده بود که با معیارهای مثبت مدل بر اساس BIC ۳۱/۹۱، MSE ۷۱۹۵۱۹۳، MAE ۴۳۷۲۱۷۸، R2 ۰/۸۲ و Ljung-Box با شاخص > ۰/۰۵ منطبق بود. بر اساس مدل انتخابی SARIMA پیش‌بینی نشان از آن دارد که اگر ناوگان ماهی‌گیری و کوشش آنها در حد فعلی باقی بماند، عملکرد صید کیلکا تا سال ۲۰۲۱ اندکی افزایش می‌یابد.

\*مؤلف مسئول