ISSN: 2538-2152; ECOPERSIA 2026;14(3): 303-318.

Application of Geostatistical Techniques to Predict the Distributional Areas of the Common Kilka (*Clupeonella cultriventris*) in the Southern Caspian Sea, Iran

ARTICLEINFO

Article Type
Original Research

Authors
Gozal Kokabee, *M.Sc.*[1]
Mohammad Farhangi, *Ph.D.*[2*]
Hadi Raeisi, *Ph.D.*[3]
Hojjatollah Jafaryan, *Ph.D.*[4]

How to cite this article
Kokabee G., Farhangi M., Raeisi H., Jafaryan J. Application of Geostatistical Techniques to Predict the Distributional Areas of the Common Kilka (Clupeonella cultriventris) in the Southern Caspian Sea, Iran. ECOPERSIA 2026;14(3): 303-318.

DOI:
10.48311/ecopersia.2026.104019.0

[1] M.Sc. Student, Department of Fisheries, Faculty of Agricultural Science and Natural Resources, Gonbad Kavous University, Gonbad, Iran.
[2] Associate Professor, Department of Fisheries, Faculty of Agricultural Science and Natural Resources, Shahid Fallahy Ave., Gonbad Kavous University, Gonbad Kavous, Golestan Province, Iran.
[3] Assistant Professor, Department of Fisheries, Faculty of Agricultural Science and Natural Resources, Gonbad Kavous University, Gonbad, Iran.
[4] Professor, Department of Fisheries, Faculty of Agricultural Science and Natural Resources, Gonbad Kavous University, Gonbad, Iran.

*** Correspondence**
Address: Associate Professor, Department of Fisheries, Faculty of Agricultural Science and Natural Resources, Shahid Fallahy Ave., Gonbad Kavous University, Gonbad Kavous, Golestan Province, Iran.
Tel.: 09381335401
Email: farhangi@gonbad.ac.ir
ORCID: 0000-0003-1956-6453

Article History
Received: June 26, 2026
Revised: August 2, 2026
Accepted: August 3, 2026
Published: August 10, 2026

ABSTRACT

Aims: The present study aimed to evaluate the spatial distribution of common Kilka (*Clupeonella cultriventris*) in the southern Caspian Sea using geostatistical interpolation, adaptive neuro-fuzzy inference system (ANFIS), and machine-learning approaches. The study addresses the limited use of integrated spatial and predictive modeling techniques to understand Kilka distribution patterns across varying environmental conditions.
Material & Methods: Daily catch and fishing-effort data were collected from commercial fishing grounds in Mazandaran Province between 1996 and 2021 during the active fishing season (April–October). Monthly CPUE values were calculated for six fishing stations. Environmental variables included fishing depth, chlorophyll-a concentration, sea surface temperature, and wind speed. Spatial interpolation methods (IDW, Kriging, and Natural Neighbor), ANFIS models with different membership functions, and a Multi-Layer Perceptron (MLP) model with a five-month lag structure were evaluated using prediction error metrics and correlation coefficients.
Findings: Among the evaluated interpolation methods, IDW with power 3 provided the most reliable representation of CPUE-based relative abundance patterns. Common Kilka abundance increased with fishing depth up to approximately 80 m and declined beyond this range. Among the ANFIS configurations, the Gaussian membership function yielded the highest predictive performance ($r = 0.86$). The MLP model with a five-month lag structure achieved the highest predictive performance within the evaluated dataset ($R = 0.98$). Fishing depth, chlorophyll-a concentration, sea surface temperature, and wind speed were identified as important predictors of Kilka distribution patterns.
Conclusion: The integration of geostatistical interpolation, ANFIS, and machine-learning approaches provided an effective framework for predicting the spatial distribution of common Kilka in the southern Caspian Sea. IDW produced the most reliable spatial interpolation results, whereas the MLP model demonstrated the highest predictive performance. The findings indicate that environmental variability plays an important role in shaping Kilka distribution patterns, and nonlinear predictive models may provide useful tools for fisheries management and ecological monitoring. Nevertheless, further validation with independent temporal datasets is recommended to assess the model's generalizability and reduce potential temporal validation bias.

Keywords: Caspian Sea; *Clupeonella cultriventris*; CPUE, Geostatistics; Spatial Distribution.

CITATION LINKS

[1] Addis P., Secci M., Angioni A., ...[2] Mohseni F., Saba F., Mirmazloumi......[3] Trochta J.T., Branch T.A., Shelt...[4] Sun F., Sun W., Chen J., Gong P....[5] Melo-Merino S.M., Reyes-Bonilla ...[6] Gladju J., Kamalam B.S., Kanagar...[7] De Campos Souza P.V. Fuzzy neura...[8] Ghoushchi S.J., Manjili S., Mard...[9] Leroy S.A., Lahijani H.A., Créta...[10] Fazli H., Kaymaram F., Daryanaba...[11] Naseka N.M., Bogutskaya N.G. Fis...[12] Janbaz A.A., Fazli H., Pourghola... [13] Fazli H., Janbaz A.A., Rabbaniha...[14] Aazami J., Esmaili-Sari A., Abdo...[15] Hashemi S., Ghorbani R., Kymaram...[16] Franceschini S., Gandola E., Mar...[17] Siddiqui S. A., Salman A., Malik...[18] Allken V., Handegard N.O., Rosen...[19] Sangün L., Güney O.İ., Özalp P.,...[20] Muthukumaran V. Water quality in...[21] Zhang Y., Yamamoto M., Suzuki G....[22] Wackernagel M., Moran D., White ...[23] Habashi H., Hosseini S.M., Shaab...[24] Abzalov M., Abzalov M. Introduct...[25] Ahmed H., Siddique M.T., Iqbal M...[26] Jiang C., Zhang H., Ren Y., Han ...[27] Wang J., Duncan D., Shi Z., Zhan...[28] Fazli H. Some Environmental Fact... [29] Martin P. Y. Gender as social in...[30] Van Meirvenne M. Characterizatio...[31] Inniss L., Simcock A., Ajawin A....[32] Bernal, P. & A. Simcock. 2016. ...[33] Henson S.A., Beaulieu C., Ilyina...[34] Mirhasannia1 S.D., Akhoundian H....[35] Jägerbrand A.K., Bouroussis C.A....[36] Gholizadeh M.H., Melesse A.M., R...[37] Kalaji H.M., Jajoo A., Oukarroum...[38] Thanigaivel S., Vickram S., Dey ...[39] Muruganandam M., Rajamanickam S....[40] Poloczanska E.S., Burrows M.T., ...[41] Makhlough A., Nasrollahzadeh Sar...[42] Mahmoudi N., Babanezhad M., Seyf... [43] Liu S., Liu Y., Alabia I. D., Ti...[44] Spencer P.D., Hollowed A. B., Si...

Introduction

Geostatistical analysis of marine fish catch data is widely used to model the spatial distribution of biomass and to understand relationships between biomass spatial patterns and environmental characteristics [1]. Satellite sensors can measure both physical and biological attributes. This technology enables the reliable assessment of global ocean coverage for variables such as sea surface temperature (SST), chlorophyll-a (Chl-a), and turbidity, with relatively high spatial and temporal resolution achievable through space-based measurements [2]. In most observations, the relationship between fish distribution and environmental factors exhibits nonlinear or irregular patterns, resulting in a random spatial structure of fish biomass [3]. Compared with conventional field sampling, satellite remote sensing offers a more efficient approach by reducing limitations related to time, cost, and spatial coverage. In addition, Geographic Information System (GIS) techniques are extensively employed in the processing of satellite images [4]. These techniques integrate theoretical aspects of oceanography and ecology with spatial databases and statistical functions, enabling comprehensive data analysis [5]. The application of artificial neural networks and fuzzy logic techniques has expanded considerably in fish-distribution studies [6]. Artificial neural networks are characterized by their topology, learning mechanisms, and the communication between inputs and outputs. Fuzzy methods have emerged as an alternative approach to traditional regression models. In general, concepts of neural network learning have been incorporated into fuzzy inference systems [7]. The combined use of artificial neural networks (ANNs) and fuzzy logic offers distinct advantages over conventional models, particularly in their ability to handle outlier data in dynamic, nonlinear systems [8]. The Caspian Sea, known as the largest enclosed lake in the world, holds great fishery value [9]. Bony fishes constitute an important component of the Caspian Sea fisheries and regional economy [10]. However, many fish stocks have declined in recent decades owing to overexploitation, environmental change, and habitat degradation. Three species of Kilka fish, namely *Clupeonella cultriventris*, *C. engrauliformes*, and *C. grimmi*, rank among the most abundant fishes in the Caspian Sea [11]. Kilka fisheries provide an important source of animal protein and income for communities in the southern Caspian Sea region. The common Kilka is distributed throughout the Caspian Sea. Kilka primarily feeds on zooplankton and serves as an important food source for valuable species such as Caspian trout and sturgeon [12]. Furthermore, Kilka species play a crucial ecological role as a food source for sturgeons and the Caspian seal, highlighting their ecological importance [13]. Numerous studies have been conducted in Iran and elsewhere to investigate fish distribution patterns and their relationships with environmental variables using geographic information systems and neural-network approaches [14–21].

Research novelty and justification: Previous studies in the Caspian Sea have primarily focused on descriptive fisheries statistics, conventional stock assessments, or isolated environmental analyses. However, integrated geostatistical, neuro-fuzzy, and machine-learning approaches have rarely been applied simultaneously to evaluate the spatial distribution of common Kilka in the context of long-term environmental variability. Therefore, this study sought to bridge this research gap by integrating spatial interpolation techniques, ANFIS models, and MLP-based predictive modeling into a unified analytical framework.

Therefore, this study aimed to evaluate the spatial distribution of common Kilka (*Clupeonella cultriventris)* in the southern Caspian Sea by integrating geostatistical interpolation techniques, ANFIS models, and machine-learning approaches. The study further examined the relationships between environmental variables and CPUE variability to improve the prediction of Kilka distribution patterns under changing environmental conditions.

Materials & Methods

The methodological framework integrated geostatistical interpolation, Adaptive Neuro-Fuzzy Inference System (ANFIS), and Multi-Layer Perceptron (MLP) modeling approaches to evaluate and predict the distribution patterns of common Kilka in the southern Caspian Sea.

Experiment Plan

Data on Kilka catch weight (kg) and species composition (%) were collected daily from commercial fishing grounds in Mazandaran Province between 1996 and 2021. Data were collected daily from fishing grounds located in Mazandaran Province, specifically from Amirabad port to Chalus. The variable of interest was the average monthly catch per unit of fishing effort at each location. Daily catch data were standardized by fishing location and effort prior to CPUE calculation. Subsequently, the average catch per unit effort per vessel per night was calculated by dividing the total catch for vessels at a specific location by the number of fishing nights. Finally, the average monthly catch per unit effort at a specific location was determined by dividing the total catch per unit effort over a month by the number of fishing nights in that month, based on Eq. (1). CPUE was expressed as kilograms of fish caught per vessel per fishing night.

Sampling was conducted monthly during the active fishing season (April–October). Species composition data from landing sites and commercial catches were used to distinguish common Kilka from other Kilka species. Fishing depth and geographical coordinates were recorded during each fishing operation. Monthly CPUE was calculated as the mean catch per vessel per fishing night:

$$CPUE_{month} = \frac{1}{K}\sum_{j=1}^{K}\left[\frac{\sum_{j=1}^{nj} X_{ij}}{nj}\right] \qquad \text{Eq. (1)}$$

where Xij is the catch of vessel i during fishing night j, expressed in kilograms; nj is the number of vessels operating during night j and K is the total number of fishing nights in the corresponding month^{-1}. CPUE was therefore expressed as kg vessel^{-1}night. Six major fishing stations were included in the study. The stations examined in this study were Amirabad, Babolsar, Chalus, Chabokrod, Daryakenar, and Khazarabad. To address the limited number of catch records in certain fishing locations, which can hinder the implementation of neural and neuro-fuzzy modeling, and to account for similar catch rates among these locations, a clustering method was employed to merge some locations. Consequently, the fishing locations were consolidated based on fishing statistics for common Kilka and data on fishing depth, surface temperature, and surface wind speed. This was achieved using the K-means clustering algorithm, resulting in six distinct categories. The factors examined in this study encompass fishing depth, chlorophyll a, surface temperature, and surface wind speed. The optimal number of clusters was determined by minimizing within-cluster variance and maximizing similarity among observations across CPUE, fishing depth, SST, and wind speed.

Independent Variables

Fishing Depth

Echosounder devices installed on fishing vessels were used to record water depth at each fishing location. These devices measure water depth from the surface to the seabed, similar to the Kilka fishing method, which uses a conical lift net and a moonlight lamp. To determine the average monthly catch depth at a specific location, daily catch data for each month were analyzed to calculate the average.

Chlorophyll a

Level-2 chlorophyll-a data in HDF format, about chlorophyll *a* concentration (in $mg.m^{-3}$) at the surface of the Caspian Sea, was obtained from the Sea Star satellite's Sea WiFS. sensor. The data, downloaded from the website http://oceancolor.gsfc.nasa.gov, covers the period 1996-2021 and consists of raster data with a spatial resolution of 0.0833333 degrees.

Surface Temperature

The website http://oceancolor.gsfc.nasa.gov provided access to long-term daily raster data of the Caspian Sea's surface temperature (in degrees Celsius). These data were acquired from the MODIS sensor on the Aqua satellite and were downloaded in HDF format at level 2. The temporal coverage spans from 1996 to 2021, and the spatial resolution of the data is 0.0833333 degrees. Monthly averaged SST values corresponding to the timing of nighttime fishing operations were extracted from satellite imagery.

Surface Wind Speed

The website http://apdrc.soest.hawaii.edu provides access to NetCDF raster data representing long-term monthly averages of surface wind speed ($m.s^{-1}$) over the Caspian Sea. These data were acquired from the QuikSCAT sensor on the QuickBird satellite. The downloaded data covers the period from 1996 to 2021 and has a spatial resolution of 0.25 degrees. To input data into the software, a file was created containing the longitude and latitude coordinates of the sampling points, along with the corresponding values for the desired features. Subsequently, interpolation techniques, including Inverse Distance Weighting (IDW), spline, and geostatistical Kriging, were applied in ArcGIS. Spatial interpolation was applied to generate continuous CPUE surfaces from the geographically referenced fishing observations. IDW, spherical and exponential kriging, and Natural Neighbor methods were compared under identical spatial settings. It primarily relies on two deterministic methods: IDW and the geostatistical interpolation method known as Kriging. Interpolation can be applied to assess the spatial distribution of a species within an area. Interpolation performance was evaluated using mean absolute error (MAE), mean bias error (MBE), mean squared standardized error ratio (MSDR), and root mean square error (RMSE). MAE and RMSE quantify the overall magnitude of prediction errors, with lower values indicating greater predictive accuracy. MBE measures systematic prediction bias and was calculated as:

$$\text{MBE}=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{z}\text{-}z_i\right) \quad \text{Eq. (2)}$$

where zi_ is the observed value at location *i*, z^i is the corresponding predicted value obtained by interpolation, and n is the total number of validation observations. Positive MBE values indicate overall overestimation, whereas negative values indicate underestimation. Mean Square Standard Deviation Ratio (MSDR) were employed, as described by Wakernagel [22].

For the kriging models, the consistency between prediction errors and estimated prediction variances was assessed using MSDR, as described by Wackernagel [22]:

$$\text{MSDR}=\left(\frac{1}{N}\right)\sum_{i=1}^{n}\left(\frac{\hat{z}_i - z_i}{\P_i}\right)^2 \quad \text{Eq. (3)}$$

where σi is the kriging prediction standard error associated with location *i*. And MSDR value close to 1 indicates that the prediction variances are consistent with the observed interpolation errors. RMSE was calculated as:

$$\text{RMSE}=\sqrt{\frac{1}{n}}\sum_{i=1}^{n}\left(\hat{z}_i - z_i\right)^2 \quad \text{Eq. (4)}$$

where zi_ and z^i represent the observed and predicted values at validation location i, respectively, and n denotes the total number of validation observations. The interpolation method with the lowest RMSE and MAE, an MBE closest to zero, and, for kriging models, an MSDR closest to 1 was considered to provide the most reliable prediction.

Adaptive Fuzzy Neuro Modeling

ANFIS models were developed in MATLAB 2018a to estimate common Kilka CPUE from fishing depth, chlorophyll-a, SST, and wind speed. Four membership-function families—triangular, trapezoidal, Gaussian, and Double-Gaussian—were evaluated under comparable training and testing conditions. Model performance was compared using RMSE and the correlation between observed and predicted CPUE. It employs adaptive neural learning techniques to develop a fuzzy modeling procedure for extracting information from a given dataset. This system calculates the membership function parameters to ensure the fuzzy inference system aligns with the input/output dataset. The adaptive fuzzy neural network is highly efficient at modeling complex processes by leveraging the linguistic capabilities of fuzzy inference systems and the computational power and learning capacity of neural networks. The adaptive fuzzy neural network is a 5-layer network with nodes and connecting arcs. The first layer handles input data with membership degrees. All modeling operations are conducted in the second to fourth layers. The last layer represents the network output, aiming to minimize the discrepancies between the network output and the actual output. An appropriate adaptive fuzzy neural network structure is selected based on the input data, membership degrees, input and output membership functions, rules, and functions. During the training phase, the membership degrees are adjusted by modifying the parameters according to an acceptable error threshold, gradually bringing the input values closer to the actual values [26, 27]. The current investigation employed the network separation technique to develop a model for estimating fish catch quantities using fishing effort as a basis. During grid separation, the input space is partitioned into multiple equal sections. In this scenario, as the number of divisions increases, the number of agents grows exponentially. Consequently, when dealing with a large number of inputs, the network's training process can take hours or even days.

Membership Functions of the Adaptive Fuzzy Neural System

Triangular, trapezoidal, Gaussian, and double Gaussian membership functions were used to evaluate the effect of membership function types on modeling efficiency [26].

Number of Membership Functions

To evaluate the influence of the number of membership functions during the fuzzification stage, the number of membership functions for each input variable was systematically altered across four values. As a result, a total of 12 distinct scenarios were created by modifying the type of membership functions used. The entire modeling process was conducted within the MATLAB 2018a neural-fuzzy toolbox.

Validation of the Model

Model performance was evaluated using the root mean square error (RMSE), as defined in Eq. (4), together with the Pearson correlation coefficient (rrr). Lower RMSE values and higher correlation coefficients indicate better predictive performance. The dataset consisted of 450 observations collected between 1996 and 2021. A random 70:30 train-test split was used, with 70% of observations for model development and the remaining 30% for independent validation. Model performance was evaluated using RMSE and the Pearson correlation coefficient (r). Because the dataset represents a temporal sequence and the MLP model incorporates lagged inputs, random partitioning may introduce some degree of temporal information leakage. Therefore, model performance should be interpreted with caution, and future studies are encouraged to use chronological or blocked time-series validation approaches to assess the model's robustness further.

Multilayer Perceptron Neural Network

An MLP neural network was developed to evaluate temporal relationships between environmental variables and CPUE. Input variables included fishing depth, chlorophyll-a concentration, SST, and wind speed. Several lag structures were examined, and the five-month lag configuration provided the best predictive performance. The MLP architecture consisted of input variables representing fishing depth, chlorophyll-a concentration, SST, and wind speed. The network was trained with supervised learning, and model performance was evaluated on the independent test set using RMSE and the Pearson correlation coefficient (R).

Findings

This study was conducted across six major fishing grounds in the southern Caspian Sea, namely Amirabad, Babolsar, Chalus, Chabokrod, Daryakenar, and Khazarabad. A total of 450 observations collected between 1996 and 2021 were used to investigate the spatial distribution of common Kilka (*Clupeonella cultriventris*) using ArcGIS software (Figure 1).

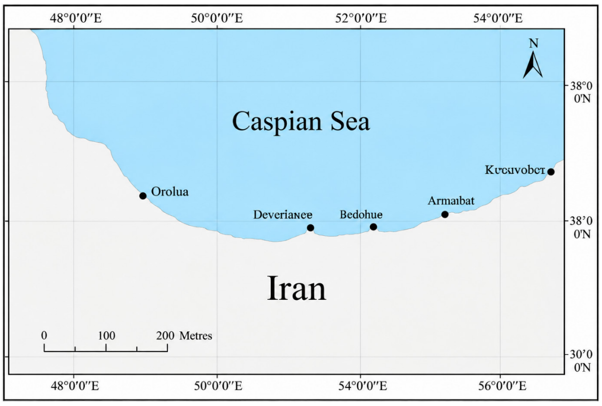

Figure 1) Location of the study area in Mazandaran Province.

Note: The circles indicate the sampling locations.

Figure 2 illustrates the spatial distribution of common Kilka estimated using the Inverse Distance Weighting (IDW) interpolation method. Comparison of interpolation performance indicated that the IDW power-3 configuration yielded the lowest prediction error among the evaluated IDW models. Consequently, this configuration provided the most reliable representation of common Kilka distribution patterns within the evaluated dataset.

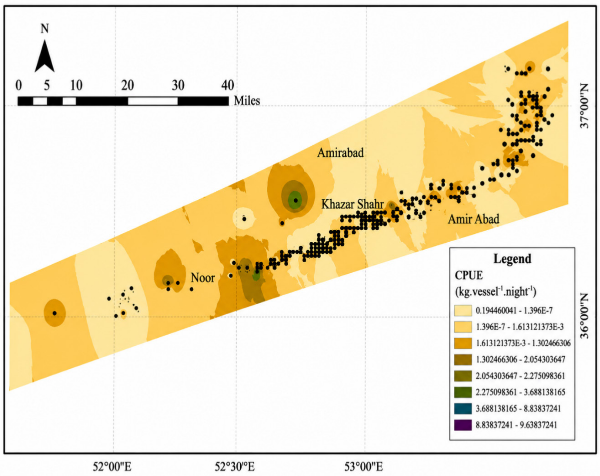

Figure 2) Figure 2. Spatial distribution of common Kilka (*Clupeonella cultriventris*) CPUE (kg. vessel^{-1}. night^{-1}) in the southern Caspian Sea predicted using the IDW interpolation method (power = 3), which produced the lowest interpolation error among the evaluated IDW configurations.

Figure 3 presents the spatial distribution of common Kilka (*Clupeonella cultriventris*) predicted using the Kriging interpolation method. Relative error comparisons indicated that the Kriging-Spherical and Kriging-Exponential models exhibited lower predictive performance than the best-performing IDW configuration and the Natural-Neighbor approach. This outcome may be related to the limited spatial density of sampling observations, which can affect the reliability of geostatistical interpolation methods.

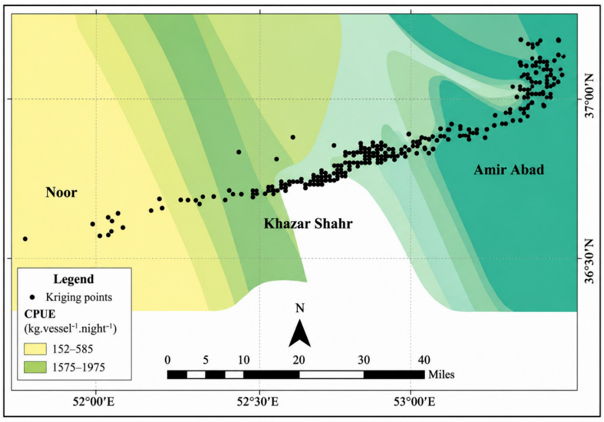

Figure 3) Spatial distribution of common Kilka (*Clupeonella cultriventris*) CPUE (kg. vessel^{-1}.night^{-1}) estimated using the Kriging-Spherical interpolation method in the southern Caspian Sea.

Figure 4 presents the spatial distribution of common Kilka (*Clupeonella cultriventris*) estimated using the Natural-Neighbor interpolation method. Relative error comparisons showed that Natural-Neighbor provided improved predictive performance compared with the kriging-based methods. Nevertheless, the IDW power-3 model exhibited the lowest interpolation error among all evaluated methods and was therefore considered the most reliable approach for subsequent spatial analyses.

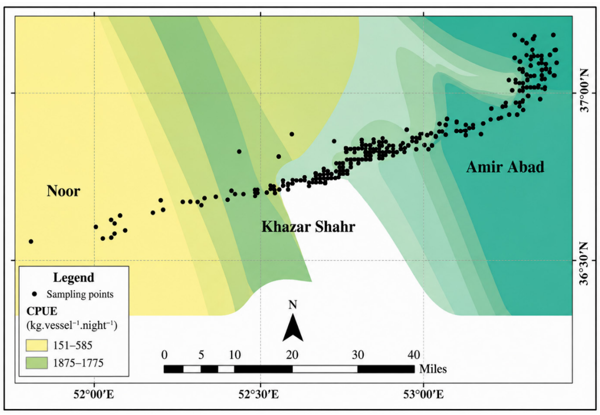

Figure 4) Spatial distribution of common Kilka (*Clupeonella cultriventris*) CPUE (kg.vessel^{-1}.night^{-1}) in the southern Caspian Sea estimated using the Natural-Neighbor interpolation method.

The ANFIS model employing the triangular membership function demonstrated good predictive performance, yielding a correlation coefficient of r = 0.86 between observed and predicted values. The close agreement between model outputs and fishery observations suggests that the model successfully captured major patterns in CPUE variability. Fishing depth, chlorophyll-a concentration, sea surface temperature, and wind speed were included as predictor variables in the ANFIS model and were associated with observed variability in CPUE—the spatial distribution and relative abundance of common Kilka (*Clupeonella cultriventris*) (Figure 5).

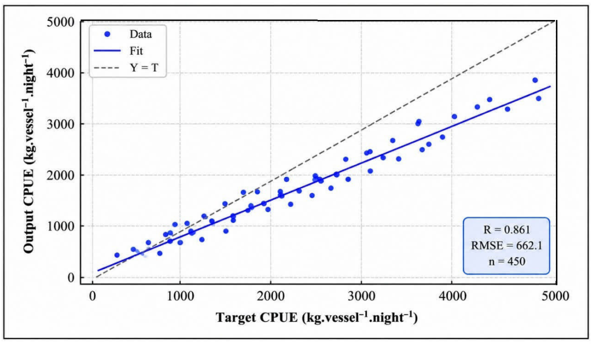

Figure 5) Correlation between observed and predicted CPUE values of common Kilka (*Clupeonella cultriventris*) generated by the ANFIS model with a triangular membership function.

The validation results of the ANFIS model with the triangular membership function showed that the residuals were approximately normally distributed and exhibited no clear systematic trend. Residual plots indicated random error dispersion without an apparent systematic pattern, suggesting that the model adequately captured the major patterns in the data without substantial bias. These findings support the model's suitability for predicting CPUE variability associated with the evaluated environmental variables (Figure 6).

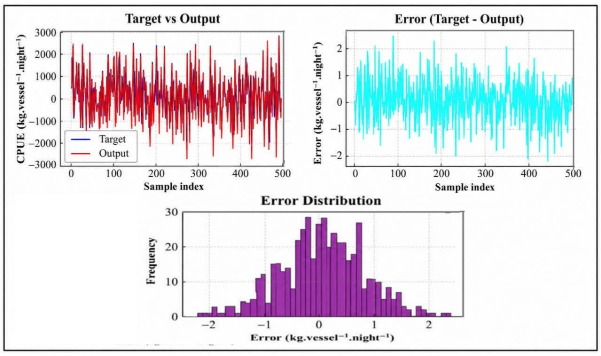

Figure 6) Validation results of the ANFIS model with a triangular membership function, including residual distribution and residual-versus-observed plots used to assess model performance and error patterns.

The ANFIS model employing the trapezoidal membership function demonstrated relatively lower predictive performance compared with the other evaluated ANFIS configurations, yielding a correlation coefficient of r = 0.49 between observed and predicted CPUE values. Although the model captured some of the variability in common Kilka CPUE, its predictive capability was weaker than that of the triangular, Gaussian, and double-Gaussian membership functions. Environmental variables, including fishing depth, chlorophyll-a concentration, sea surface temperature, and wind speed, were incorporated as predictors into the model and contributed to the prediction of CPUE patterns (Figure 7).

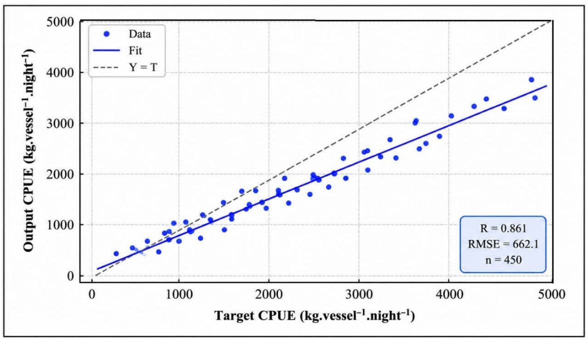

Figure 7) Correlation analysis of predicted and observed CPUE (kg.vessel^{-1}.night^{-1}) for common Kilka (*Clupeonella cultriventris*) using an Adaptive Neuro-Fuzzy Inference System (ANFIS) with a trapezoidal membership function across all model inputs.

The validation results of the ANFIS model employing the trapezoidal membership function indicated that the residuals were approximately normally distributed and exhibited no clear systematic pattern. Residual plots showed a random dispersion of errors around the observed values, suggesting that the model adequately represented the major patterns in the data without substantial bias. These findings support the model's suitability for evaluating relationships between environmental variables and common Kilka CPUE (Figure 8).

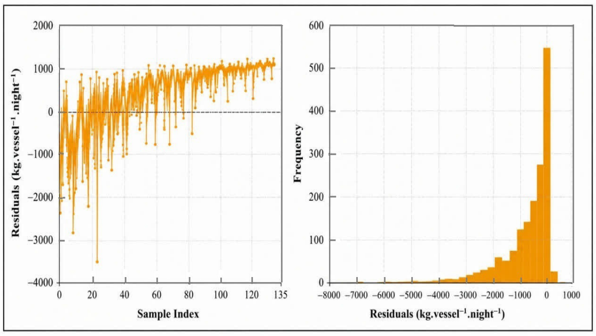

Figure 8) Validation performance of the ANFIS model with a trapezoidal membership function based on 135 testing observations, including (a) residuals (prediction − observed) versus sample index and (b) residual error distribution, used to evaluate model performance and potential prediction bias.

The ANFIS model employing the Gaussian membership function demonstrated the highest predictive performance among the evaluated ANFIS configurations (r = 0.86), substantially outperforming the trapezoidal membership function (r = 0.49) and slightly exceeding the triangular (r = 0.86 and double-Gaussian (r = 0.84) configurations. A comparison of model performances is presented in Figure 13.

The Multi-Layer Perceptron (MLP) model with a five-month lag structure demonstrated excellent predictive performance, yielding an r = 0.98 correlation between observed and predicted CPUE values. The model was trained on 70% of the available observations and validated on an independent test set comprising the remaining 30%. The close agreement between observed and predicted values indicates that the model successfully captured the temporal variability in common Kilka CPUE. Nevertheless, further validation on independent datasets is recommended to assess the model's generalizability and reduce the risk of overfitting (Figure 9).

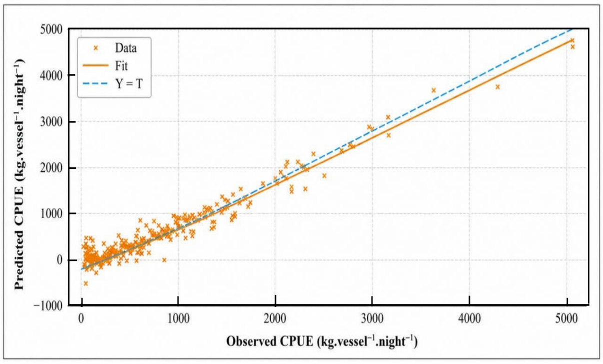

Figure 9) Relationship between observed and predicted CPUE (kg.vessel^{-1}.night^{-1}) values generated by the Multi-Layer Perceptron (MLP) model with a five-month lag structure. The model achieved a correlation coefficient of R = 0.980 on an independent test set (30% of observations). The dashed line represents the 1:1 relationship between observed and predicted values.

The validation results of the MLP model with a five-month lag structure indicated that the residuals exhibited no clear systematic pattern and were generally concentrated around lower error values. The residual plots suggest that the model adequately captured the major temporal variations in CPUE, while the distribution of residuals indicates acceptable predictive performance across the testing dataset. These findings support the model's suitability for predicting common Kilka CPUE under the evaluated environmental conditions (Figure 10).

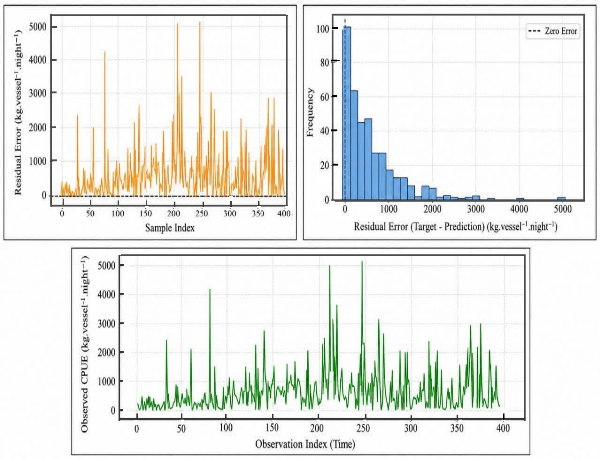

Figure 10) Validation diagnostics for the MLP model with a five-month lag structure using 400 testing observations, including (a) residual errors across samples, (b) frequency distribution of residual errors, and (c) temporal variation in observed CPUE values used for model evaluation.

The ANFIS model employing the double-Gaussian membership function demonstrated good predictive performance, yielding a correlation coefficient of $r = 0.84$ between observed and predicted CPUE values. The results indicate that the model captured a substantial proportion of the variability in common Kilka CPUE. Fishing depth, chlorophyll-a concentration, sea surface temperature, and wind speed were included as predictor variables in the model. The performance of the double-Gaussian configuration was slightly lower than that of the Gaussian and triangular membership functions, but substantially higher than that of the trapezoidal membership function (Figure 11).

The validation results of the ANFIS model with a double-Gaussian membership function indicated that the residuals were approximately normally distributed and exhibited no evident systematic pattern. Residual plots showed a random dispersion of errors around zero, suggesting that the model captured major patterns in variability in common Kilka CPUE, with limited evidence of systematic bias. These findings support the model's suitability for predicting CPUE variability associated with the evaluated environmental variables (Figure 12).

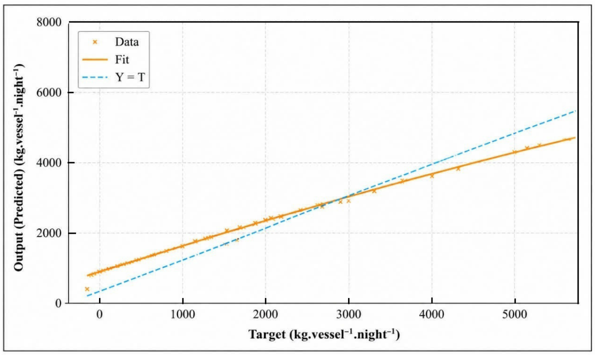

Figure 11) Relationship between observed and predicted CPUE ($kg.vessel^{-1}.night^{-1}$) values of common Kilka (*Clupeonella cultriventris*) obtained from the ANFIS model with a double-Gaussian membership function.

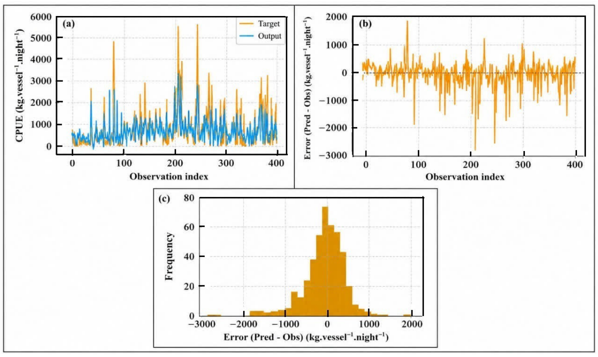

Figure 12) Validation diagnostics for the ANFIS model with a double-Gaussian membership function based on 400 testing observations, including (a) observed and predicted CPUE ($kg.vessel^{-1}.night^{-1}$) across observations, (b) residual errors across observations, and **(c)** frequency distribution of residual errors used to evaluate model performance and potential prediction bias.

The newly developed MLP model with a five-month lag structure demonstrated the highest predictive performance among the evaluated models, yielding an $r = 0.98$ correlation between observed and predicted CPUE values. The model was trained on 70% of the available observations and independently validated on the remaining 30%, indicating strong predictive performance within the evaluated dataset. The improved performance of the MLP model may be attributed to its ability to capture complex nonlinear and temporal relationships among environmental variables, including fishing depth, chlorophyll-a concentration, sea surface temperature, and wind speed.

the ANFIS configurations, the Gaussian membership function produced the highest predictive performance (r = 0.86), followed by the triangular (r = 0.86) and double-Gaussian (r = 0.84) functions. In contrast, the trapezoidal membership function exhibited substantially lower predictive performance (r = 0.49), indicating a reduced ability to represent the underlying relationships between environmental variables and CPUE. Overall, the results indicate that MLP and selected ANFIS configurations, particularly the Gaussian and triangular membership functions, provide useful tools for modeling common Kilka (*Clupeonella cultriventris*) distribution patterns and CPUE variability under dynamic environmental conditions. Nevertheless, additional validation on independent datasets is recommended to assess the model's generalizability further and reduce potential temporal validation bias and the risk of overfitting.

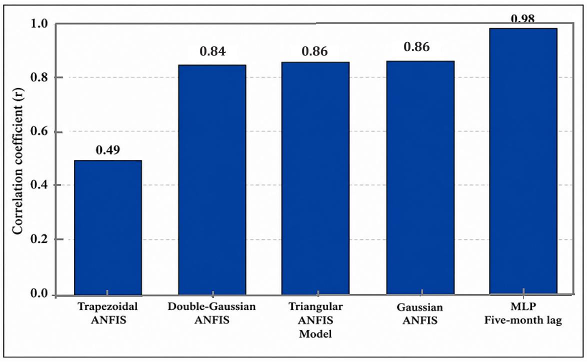

Figure 13) Comparison of predictive performance among the evaluated ANFIS membership functions (triangular, trapezoidal, Gaussian, and double-Gaussian) and the MLP model with a five-month lag structure based on correlation coefficients (r). The MLP model achieved the highest predictive performance (r = 0.98), followed by the Gaussian (r = 0.86), triangular (r = 0.86), and double-Gaussian (r = 0.84) ANFIS configurations. In contrast, the trapezoidal membership function showed substantially lower predictive performance (r = 0.49).

Discussion

The Caspian Sea supports one of the most important inland fisheries in the region and plays a major role in fish production and food security. The data gathered from the study of the distribution of common Kilka (*Clupeonella cultriventris*) revealed a clear relationship between fishing depth and CPUE variability. The results of the interpolation analysis indicated that CPUE-based relative abundance increased with fishing depth from approximately 30 to 80 m, then remained relatively stable or declined at greater depths. This pattern may reflect the vertical distribution of zooplankton prey and favorable thermal conditions within the southern Caspian Sea. The reduced abundance of common Kilka at depths greater than 80 m may explain commercial fishing operations' preference for intermediate-depth fishing grounds, where catch efficiency and economic profitability are generally higher. Furthermore, integrating catch and fishing-effort data in a GIS environment provides a useful tool for evaluating spatial fishing patterns and supporting stock assessment and fisheries management.

Stock assessment techniques can be integrated with GIS-based spatial analyses to improve the identification of fishing grounds and support more effective fisheries management strategies [29]. In the present study, the IDW interpolation approach provided the most reliable representation of the common Kilka distribution under the available sampling conditions. In contrast, kriging-based methods exhibited lower predictive performance, likely due to the limited number of observations and the heterogeneous spatial distribution of fishing activities. Similar findings were reported by Van Meirvenn [30] and Habashi et al. [23], who concluded that kriging methods may be less suitable when sample sizes are limited and spatial variability is high. These results suggest that deterministic interpolation approaches, such as IDW, may yield more stable spatial predictions under sparse sampling conditions. In contrast, kriging methods may perform better when spatial structure is stronger and more observations are available.

Phytoplankton communities are widely distributed throughout the Caspian Sea and play a fundamental role in ecosystem functioning and food-web dynamics [33]. Their abundance and diversity are strongly influenced by physical, chemical, and biological factors, including light availability, temperature, salinity, dissolved Oxygen, nutrient concentrations, and grazing pressure [33]. Consequently, environmental variables such as sea surface temperature (SST) and chlorophyll-a are widely used as indicators of oceanographic conditions and biological productivity [34-36]. Temperature fluctuations can influence the distribution, metabolism, and behavior of marine organisms, including plankton and fish species [37,38]. In the present study, SST and chlorophyll-a were included as environmental predictor variables. They were associated with variability in common Kilka CPUE, highlighting the influence of environmental conditions on species distribution patterns. Furthermore, nutrient enrichment resulting from riverine inputs and coastal pollution may enhance plankton productivity in the Caspian Sea, thereby affecting food availability for Kilka populations [39,40]. These findings highlight the potential value of incorporating environmental variables into fisheries assessments to improve stock evaluation and predictive modeling [41-43].

Additionally, the present study investigated the relationships among environmental, spatial, temporal, and biological variables and their associations with variability in common Kilka (*Clupeonella cultriventris*) CPUE in the southern Caspian Sea. By integrating environmental information with advanced modeling approaches, this study contributes to a broader understanding of the ecological factors associated with Kilka distribution patterns. It provides insights into factors linked to fluctuations in catch rates.

Among the evaluated ANFIS configurations, the Gaussian membership function yielded the highest predictive performance ($r = 0.86$), followed by the triangular ($r = 0.86$) and double-Gaussian ($r = 0.84$) membership functions. In contrast, the trapezoidal membership function produced the lowest predictive performance ($r = 0.49$). These findings suggest that Gaussian-based membership functions may provide a more effective representation of the nonlinear relationships associated with common Kilka distribution patterns than the other ANFIS configurations evaluated. The higher predictive performance of the Gaussian membership function is consistent with the findings of Jiang et al. [26], who reported improved predictive capability of Gaussian functions in environmental modeling applications.

The newly developed MLP model with a five-month lag structure produced the highest predictive performance among all interpolation and ANFIS models evaluated in the present study, yielding a correlation coefficient of $R = 0.98$. The dataset was divided into training (70%) and testing (30%) subsets, and model performance was assessed using the independent testing dataset. The improved performance of the MLP model suggests that temporal dependencies may help explain CPUE variability in common Kilka. By incorporating environmental information from multiple preceding months, the model captured temporal patterns and delayed associations between environmental variables and CPUE that interpolation methods or ANFIS configurations may not fully represent. Nevertheless, neural network models may be susceptible to overfitting, and further validation on independent datasets from other regions and time periods is recommended. In addition, because the model incorporated lagged inputs and the dataset was randomly partitioned into training and testing subsets, some degree of temporal information leakage cannot be entirely ruled out. Therefore, the reported predictive performance should be interpreted within the context of the evaluated dataset.

The results further demonstrated that increasing model complexity does not necessarily improve predictive performance. Although a larger number of membership functions can enhance model flexibility, it may also substantially increase computational requirements and training time. In the present study, relatively simple ANFIS structures, particularly those employing Gaussian and triangular membership functions, produced competitive predictive performance. These findings suggest that an appropriate balance between model complexity and predictive accuracy is essential when developing ecological prediction models.

Several limitations should be acknowledged. Spatial prediction uncertainty may increase in offshore and nearshore areas because fishing-vessel trajectories followed discontinuous fishing paths rather than systematic survey designs. Additionally, geostatistical methods rely on assumptions of stationarity and spatial continuity that may not always be satisfied under highly heterogeneous marine conditions. The absence of long-term, consistent photosynthetically active radiation (PAR) datasets prevented the inclusion of this potentially important environmental variable in the modeling process. Furthermore, although the MLP model was evaluated on an independent test set, additional external validation on independent datasets from other regions and time periods would further strengthen confidence in the model's generalizability. Future studies should incorporate additional environmental variables, broader seasonal coverage, and advanced machine-learning approaches to improve predictive performance and ecological interpretation further.

Although the MLP model achieved high predictive performance (R = 0.98), caution is warranted when interpreting these results. Because the dataset represents a long-term time series and the model incorporates a five-month lag structure, random train-test partitioning may allow temporal information leakage between training and testing subsets. Consequently, model performance may be somewhat optimistic. Future studies should evaluate model robustness using chronological validation approaches, such as rolling-origin validation or blocked time-series cross-validation, to better assess predictive generalizability.

Conclusion

The results demonstrated that the MLP model with a five-month lag structure produced the highest predictive performance among all evaluated models, achieving a correlation coefficient of approximately R = 0.98. The incorporation of temporal information improved predictions of CPUE variability and distribution patterns for *Clupeonella cultriventris* under the evaluated environmental conditions. suggesting that Gaussian-based membership functions may provide an effective representation of nonlinear relationships associated with common Kilka distribution patterns.

Despite these promising results, several limitations should be acknowledged. Limited observations at some fishing locations may have affected the reliability of spatial predictions, and neural network models may be susceptible to overfitting despite independent model validation. Because the MLP model incorporated lagged inputs and used random train-test partitioning, future studies should evaluate the model's robustness using chronological validation approaches to assess predictive generalizability further. Future studies should incorporate additional environmental variables, broader seasonal coverage, and advanced machine-learning approaches to improve predictive performance and ecological interpretation further.

Acknowledgments

The article's authors would like to express their sincere gratitude to Gonbad Kavous University for their moral support in conducting the research.

Authors' Contribution: G Kokabee: Student, M.Sc.; **M Farhangi**: Supervision, Conceptualization, Project; **H Raeisi:** Methodology, Formal analysis. **H Jafaryan:** Writing & review.

Ethical Permission: All experiments were performed in accordance with the protocol approved by the Research & Ethics Committee of Gonbad Kavous University on 20 July 2022, under approval number 206035.

Fundings/Support: This work was supported and carried out by Gonbad Kavous University as a master's project.

Conflict of Interest: The authors reported no potential conflict of interest.

AI Use Declaration: The authors used AI tools for language editing. The authors have reviewed and edited the output and take full responsibility for the publication's content.

References

1. Addis P., Secci M., Angioni A., Cau, A. Spatial distribution patterns and population structure of the sea urchin *Paracentrotus lividus* (Echinodermata: Echinoidea), in the coastal fishery of western Sardinia: a geostatistical analysis. Sci. Mar. 2012; 76(4): 733-740.
2. Mohseni F., Saba F., Mirmazloumi S.M., Amani M., Mokhtarzade M., Jamali S., Mahdavi S. Ocean water quality monitoring using remote sensing techniques: A review. Mar. Environ. Res. 2022; 180: 105701.
3. Trochta J.T., Branch T.A., Shelton A.O., Hay D.E. The highs and lows of herring: A meta-analysis of patterns and factors in herring collapse and recovery. Fish 2020; 21(3): 639-662.
4. Sun F., Sun W., Chen J., Gong P. Comparison and improvement of methods for identifying water bodies in remotely sensed imagery. Int. J. Remote Sens. 2012; 33(21): 6854-6875.
5. Melo-Merino S.M., Reyes-Bonilla H., Lira-Noriega A. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. Ecol. Model. 2020; 415: 108837.
6. Gladju J., Kamalam B.S., Kanagaraj A. Applications of data mining and machine learning framework in aquaculture and fisheries: Smart Agric. Technol. 2022; 2: 100061.
7. De Campos Souza P.V. Fuzzy neural networks and neuro-fuzzy networks: A review of the main techniques and applications used in the literature. Appl. Soft Comput. 2020; 92:106275.
8. Ghoushchi S.J., Manjili S., Mardani A., Saraji M.K. An extended new approach for forecasting short-term wind power using modified fuzzy wavelet neural network: A case study in wind power plant. EN. 2021; 223: 120052.

9. Leroy S.A., Lahijani H.A., Crétaux J. F., Aladin N.V., Plotnikov I.S. Past and current changes in the largest lake of the world: The Caspian Sea. Large Asian Lakes in a Changing World: Natural State and Human Impact. 2020; 65-107.
10. Fazli H., Kaymaram F., Daryanabard G.R. Effects of fishing and environmental parameters on the commercial bony fish assemblage in the southern Caspian Sea. Oceanol. Hydrobiol. Stud. 2022; 51(1): 90-99.
11. Naseka N.M., Bogutskaya N.G. Fishes of the Caspian Sea: zoogeography and updated check-list. Zoosyst. Ross. 2009; 18(2): 295-317.
12. Janbaz A.A., Fazli H., Pourgholam R., Kaymaram F., Bandpei A., Abdolmaleki S., Khedmati K. Fishery and biological aspects of anchovy Kilka (*Clupeonella engrauliformis*) in the southern Caspian Sea. Iran. J. Fish. Sci. 2012; 11(4): 796-806.
13. Fazli H., Janbaz A.A., Rabbaniha M., Khedmati K., Chaudhari H.S. Study of environmental and three Kilka species regime shifts in the Caspian Sea. Iran. J. Fish. Sci. 2021; 20(5): 1247-1261.
14. Aazami J., Esmaili-Sari A., Abdoli A., Sohrabi H., Van den Brink P. J. Monitoring and assessment of water health quality in the Tajin River, Iran using physicochemical, fish, and macroinvertebrate indices. J. Environ. Health Sci. Eng. 2015; 13(1): 1-12.
15. Hashemi S., Ghorbani R., Kymaram F., Hossini S.A., Eskandari G., Hedayati A. Relationship of physicochemical factors with fish biomass and production in Shadegan Wetland, Iran. B JBD. 2016; 17(2): 32-39.
16. Franceschini S., Gandola E., Martinoli M., Tancioni L., Scardi M. Cascaded neural networks improving fish species prediction accuracy: the role of the biotic information. Sci. Rep. 2018; 8(1): 4581.
17. Siddiqui S. A., Salman A., Malik M. I., Shafait F., Mian A., Shortis M. R., Harvey E. S. Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labeled data. ICES 2018; 75(1): 374-389.
18. Allken V., Handegard N.O., Rosen S., Schreyeck T., Mahiout T., Malde K. Fish species identification using a convolutional neural network trained on synthetic data. ICES 2019; 76(1): 342-349.
19. Sangün L., Güney O.İ., Özalp P., Başusta, N. Estimation of body weight of *Sparus aurata* with artificial neural network (MLP) and M5P. (nonlinear regression)–LR algorithms. Iran. J. Fish. Sci. 2020; 19(2): 541-550. URL:
20. Muthukumaran V. Water quality index process using artificial neural networks. IJITRA. 2022; 1(1): 33-37.
21. Zhang Y., Yamamoto M., Suzuki G., Shioya H. Collaborative forecasting and analysis of fish catch in Hokkaido from multiple scales by using neural network and ARIMA. Model. IEEE Access. 2022; 10(1): 7823-7833.
22. Wackernagel M., Moran D., White S., Murray M. Ecological Footprint accounts for advancing sustainability: measuring human demands on nature. SDIs in Ecolo Econo. 2006; 246p.
23. Habashi H., Hosseini S.M., Shaaban Sh., Jahangard M. Evaluation and accuracy of interpolation methods in estimating total soil nitrogen using GIS. Spatial Information Systems Conference. 2015; 3: 1-7.
24. Abzalov M., Abzalov M. Introduction to geostatistics. J. Appl.Min. Geol. 2016; 2: 233-237.
25. Ahmed H., Siddique M.T., Iqbal M., Hussain F. Comparative study of interpolation methods for mapping soil pH in the apple orchards of Murree, Pakistan J.SE., 2017; 36(1): 70-76.
26. Jiang C., Zhang H., Ren Y., Han Z., Chen K. C., Hanzo L. Machine learning paradigms for next-generation wireless networks. IEEE Wirel. Commun. 2016; 24(2): 98-105.
27. Wang J., Duncan D., Shi Z., Zhang B. WEB-based gene set analysis toolkit (Web Gestalt): update 2013. Nucleic Acid. Res. 2013; 41(W1): W77-W83.
28. Fazli H. Some Environmental Factors Effects on Species Composition, Catch and CPUE of Kilka in the Caspian Sea. ECOPERSIA 2011; 1(2); 157-164.
29. Martin P. Y. Gender as social institution. Soc. Force. 2004; 82(4): 1249-1273.
30. Van Meirvenne M. Characterization of soil spatial variation using geostatistics. Ph.D. Thesis. Academic Press, University of Gent, Belgium. 1991;168P.
31. Inniss L., Simcock A., Ajawin A. Y., Alcala A. C., Bernal P., Calumpong H. P., Węsławski J. M. The first global integrated marine assessment. United Nations. Accessed on 5 February. 2016. Bernal, P. & A. Simcock. 2016. Marine Scientific Research. Chapter 30, p 1-18. In: L. Inniss and A. Simcock (Joint Coordinators). The First Global Integrated Marine Assessment. World Ocean Assessment I. United Nations, New York.
32. Henson S.A., Beaulieu C., Ilyina T., John J.G., Long M., Séférian R., Sarmiento J.L. Rapid emergence of climate change in environmental drivers of marine ecosystems. Nat. Commun. 2017; 8(1): 14682.
33. Mirhasannia1 S.D., Akhoundian H., Taghavi H. Spatial and Temporal Distribution and Biomass Trend of Planktonic Green Microalgae (Chlorophyta) in Southeast of the Caspian Sea. ECOPERSIA 2019; 7(1):1-11.
34. Jägerbrand A.K., Bouroussis C.A. Ecological impact

of artificial light at night: effective strategies and measures to deal with protected species and habitats. Sustainability. 2021; 13(11): 5991.
35. Gholizadeh M.H., Melesse A.M., Reddi L. A comprehensive review on water quality parameters estimation using remote sensing techniques. J. Sens. 2016; 16(8): 1298.
36. Kalaji H.M., Jajoo A., Oukarroum A., Brestic M., Zivcak M., Samborska I.A., Ladle R.J. Chlorophyll a fluorescence as a tool to monitor physiological status of plants under abiotic stress conditions. Acta Physiol. Plant. 2016; 38(4): 1-11.
37. Thanigaivel S., Vickram S., Dey N., Jeyanthi P., Subbaiya R., Kim W., Karmegam N. Ecological disturbances and abundance of anthropogenic pollutants in the aquatic ecosystem: Critical review of impact assessment on the aquatic animals. Chemosphere. 2022; 137475.
38. Muruganandam M., Rajamanickam S., Sivarethinamohan S., Reddy M.K., Velusamy P., Gomathi R., Munisamy S.K. Impact of climate change and anthropogenic activities on aquatic ecosystem–A review. Environ. Res. 2023; 117233.
39. Poloczanska E.S., Burrows M.T., Brown C.J., García Molinos J., Halpern B.S., Hoegh-Guldberg O., Sydeman W.J. Responses of marine organisms to climate change across oceans. Front. Mar. Sci. 2016; 62(3): 212-220.
40. Makhlough A., Nasrollahzadeh Saravi H., Roohi A., Keyhansani A.R., Vahedi F. The study on dinocysts in the southern coastal waters of the Caspian Sea. Iran. J. Fish. Sci. 2022; 21(1): 48-62.
41. Mahmoudi N., Babanezhad M., Seyfabadi J., Ahmadi M.R. Spatiotemporal relationships between life stages of the invasive ctenophore, *Mnemiopsis leidyi*, and environmental parameters in the southern Caspian Sea. J. Gt. Lakes Res. 2020; 46(5): 1262-1276.
42. Liu S., Liu Y., Alabia I. D., Tian Y., Ye Z., Yu H., Cheng J. Impact of climate change on wintering ground of Japanese anchovy (*Engraulis japonicus*) using marine geospatial statistics. Front. Mar. Sci. 2020; 7(1): 41-47.
43. Spencer P.D., Hollowed A. B., Sigler M.F., Hermann A.J., Nelson M.W. Trait-based climate vulnerability assessments in data-rich systems: An application to eastern Bering Sea fish and invertebrate stocks. Glob. Change Biol. 2019; 25(11): 3954-3971.