Regional Studies in Marine Science 77 (2024) 103587

Contents lists available at ScienceDirect

Regional Studies in Marine Science

journal homepage: www.elsevier.com/locate/rsma

Performance of different modeling techniques in testing the impact of environmental variables on eel landing in Ichkeul Lake, a RAMSAR Wetland and UNESCO biosphere reserve

Sabrine Sahbani [a,b], Béchir Béjaoui [a,*], Ennio Ottaviani [c], Sihem Benabdallah [d], Eva Riccomagno [e], Enrico Prampolini [c], Donata Melaku Canu [f], Hechmi Missaoui [a,b], Cosimo Solidoro [f,g]

[a] National Institute of Marine Sciences and Technologies, 28 Street du 2 March 1934, Carthage Salammbò, 2025, Tunisia
[b] National Institute of Agronomy of Tunisia, 43 Av. Charles Nicolle, Tunis 1082, Tunisia
[c] OnAIR srl, 26, Via Carlo Barabino, Genoa 16129, Italy
[d] Geo-resources Laboratory, Center for Water Research and Technologies, PO Box 273, Soliman, 8020, Tunisia
[e] University of Genova, Via Dodecaneso, 35, Genova 16146, Italy
[f] National Institute of Oceanography and Applied Geophysics, OGS, Borgo Grotta Gigante, 42/c, Trieste, Sgonico 34010, Italy
[g] International Centre for Theoretical Physics, ICTP, Trieste, Strada Costiera, Italy

ARTICLE INFO

Keywords:
Random forest
Cubist model
Environmental variables
Anthropogenic pressures
Anguilla anguilla
Ichkeul Lake

ABSTRACT

Advanced modeling techniques, including Random Forest (RF) and Cubist model (CB), were used to assess the relationship between environmental factors and European eels (*Anguilla anguilla*) abundance and to provide insights into the lake's ecological status while considering climate change and anthropogenic influences. A comprehensive dataset, attained through extensive environmental and biological monitoring for the period 2010–2020, was employed. The performance of the models is carried out using key metrics including the root mean square error (RMSE), coefficient of determination (R^2), and mean estimation error (MAE). In addition, a sensitivity analysis was conducted to ascertain the relative significance of the thirteen input variables in shaping the predictions of the models. The precision of the CB and RF models in predicting eel landings surpassed that of Multiple Regression. In the training dataset, the CB model achieved R^2=0.55, RMSE=7.68 tons, and MSE=6.20 tons, and the RF model achieved R^2=0.56, RMSE=7.20 tons, and MSE=5.56 tons. High accuracy was maintained on the testing dataset, with CB achieving R^2=0.73, RMSE=5.13 tons, and MSE=5.89 tons, and RF achieving R^2=0.73, RMSE=5.81 tons, and MSE=4.67 tons. The scatter plot between predicted and measured eel landings indicated that the RF model tends to overestimate lower values and underestimate higher values of eel landings, while the CB model gave better performance in this context. Further, the carried sensitivity analysis using the CB model unequivocally identified three pivotal factors – water level, salinity, and turbidity level – as the most influential determinants governing the landing of eels in this ecosystem. Thus, the CB model is considered to be more promising for interpreting the relationship between environmental parameters and eel landings, which could be used by managers for an effective lake management strategy.

1. Introduction

Coastal wetlands are renowned for their wealth of biodiversity and biogeochemical processes, making them among the world's most ecologically rich environments (Galewski et al., 2012; Newton et al., 2018). These areas provide a plethora of invaluable services that contribute significantly to human well-being (Galewski et al., 2012; Newton et al., 2018). Nevertheless, the rapid growth of human activities poses a severe threat to these fragile ecosystems, making human impact the most significant challenge to ecological quality in these areas (Newton et al. 2018.).

Even though Mediterranean ecosystems are being recognized as one of the world's primary biodiversity hotspots (Derneği, 2010), they are, regrettably, not exempt from the concerning trend of declining

* Corresponding author.
E-mail addresses: bejaoui.bechir@gmail.com, bejaoui.bechir@instm.rnrt.tn (B. Béjaoui).

https://doi.org/10.1016/j.rsma.2024.103587
Received 5 October 2023; Received in revised form 13 May 2024; Accepted 20 May 2024
Available online 26 May 2024

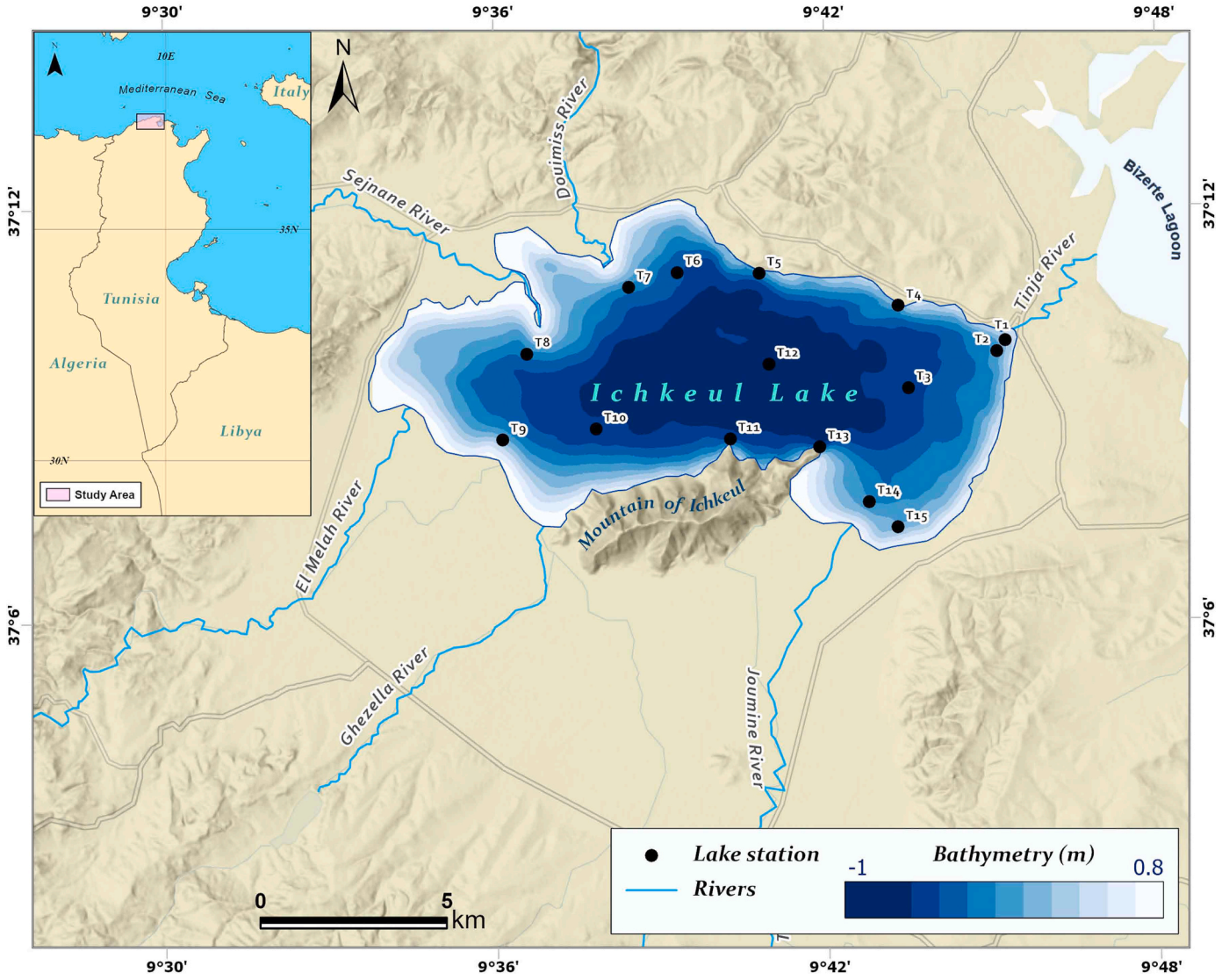

Fig. 1. Geographic localization of Ichkeul Lake and location of sampling stations.

biological diversity (Newton et al., 2018). Among the coastal wetlands in the arid and semi-arid Mediterranean region, with a particular focus on North Africa, Ichkeul Lake remains an example of a jeopardized ecosystem due to the combined effects of climate change and human activities. This lake has been included in national and international conventions including RAMSAR and UNESCO conventions (ANPE, 2008), and it serves as a vital winter refuge for Palearctic waterfowl (Sahbani et al., 2022).

However, the construction of dams along the main rivers has led to a substantial reduction in the inflow of fresh water into the lake. Additionally, the installation of sluice gates at the lake's outlet, situated at the Tinja River, has severely restricted the exchange of water and sediments within the complex Lake Ichkeul-Bizerte Lagoon system.

Previous studies (Aouissi et al., 2014; Madyouni et al., 2023) have highlighted the absence of adequate wastewater treatment systems in rural communities within the Ichkeul catchment area. Additionally, the lake is surrounded by intensive agricultural activities such as cereal and sunflower cultivation, which involve heavy use of chemical fertilizers. The extensive hydrographic system surrounding the lake exacerbates the issue by allowing untreated wastewater and chemical components to be discharged into the environment.

Recent research has underscored the vulnerability of the Ichkeul ecosystem to climate change. Previous studies, such as those by Béjaoui et al., (2022) and Sahbani et al., (2022), have pointed out the warming of the lake's water and the imbalance between cumulative evaporation and precipitation in the region, exacerbating the effects of anthropogenic pressures and further disrupting the lake's hydrological cycle.

These alterations profoundly affect the migration of various fauna species, particularly diadromous fish such as European eels. These fish species are renowned as exceptional bioindicators that serve as a barometer of the biological health of ecosystems, bridging the vital link between the sea (for reproduction) and freshwater (for growth) (Galewski et al., 2012).

The European eel (*Anguilla anguilla,* Linnaeus, 1758), referred to as 'eel' hereafter, holds a position of immense importance in the Mediterranean ecosystem, especially within the Ichkeul wetland (Derouiche, 2016). However, there has been a concerning decline in eel biomass since the 1970s, leading to its inclusion in Annex II of the Convention on International Trade in Endangered Species in 2007 (Rehof, 2021) and its classification as critically threatened in the International Union for Conservation of Nature Red List of threatened species in 2010 (Jacoby et al., 2015).

Various challenges during migration and movements are faced by eels including predation, habitat loss, and scarcity of food which contributed to this decline (Lagarde et al., 2021).

Eels begin their life cycle in the Sargasso Sea, hatching as

leptocephali larvae. They then undertake a challenging journey to the coasts of Europe and Northern Africa, where they transition into glass eels in their continental growth habitats. After completing their growth phase, they migrate back to the Sargasso Sea as silver eels (Lagarde et al. 2021). Throughout their lifecycle, eels spend years moving within freshwater, brackish, and marine environments. Despite extensive research on migratory eel movements, there is limited understanding of eel movements within habitats and how environmental factors affect non-migratory eel stocks (Lagarde et al. 2021).

For ecosystems under significant pressure, it can be challenging to distinguish between the climate-related factors driven impacts and those resulting from human activities (Leoni et al., 2021). Nonetheless, numerous studies have used variables such as precipitation, temperature, evaporation, solar energy, and wind intensity to delineate the repercussions of climate change (Cutforth, 2000; Bardossy and Van Mierlo, 2000; Dai et al., 2018).

Further, predicting the diverse impacts of environmental variables is complex due to their interactions (Béjaoui et al. 2016). Over the past decade, machine learning algorithms, particularly CB and RF models, have emerged as powerful tools. They have demonstrated their effectiveness in delineating non-linear relationships between predictor variables and response features by minimizing the loss function, primarily through an iterative training process (Zhang et al. 2021). These models are increasingly recognized as promising methods in the fields of environmental sciences (Motarjemi et al. 2020; Zhang et al. 2021).

The global objective of this study is twofold: firstly, to assess the interaction between environmental factors and the abundance of European eels under the impact of climate change and anthropogenic pressures; and secondly, to gain a deeper understanding of the ecological state of Ichkeul Lake.

Making use of a comprehensive dataset that combines historical records with collected sampling data, this research endeavors to accomplish the following key objectives: (i) characterizing the seasonal and annual variations of the environmental parameters within Ichkeul Lake, elucidating the stark contrasts that define its overall condition and health, thereby revealing its ecological state, (ii) developing two distinct machine learning models RF and CB regression techniques to estimate the impact of climatic and anthropogenic pressures on eel landing, and (iii) assessing the models' performance through a comparative analysis of their results. This evaluation aims to identify the most suitable model for our database characteristics.

2. Material and methods

2.1. Study area

Ichkeul Lake is an ecologically significant ecosystem situated in the northeastern region of Tunisia (Fig. 1). This remarkable environment encompasses approximately 30 km^2 of temporary swamps and is bounded on its southern edge by a mountain (1366 ha). The climate in the Ichkeul region is characterized as sub-humid Mediterranean, marked by distinct seasonal periods, including both hot and rainy periods (Aouissi et al., 2014).

Ichkeul Wetland is positioned at the confluence of six rivers: Douimis, Sejnane, Melah, Ghezela, Joumine, and Tine. These rivers, originating from a catchment area spanning 2600 km^2, serve as a source of freshwater inflow into the lake. Furthermore, the lake is connected to Bizerte Lagoon, which in turn links to the Mediterranean Sea via the Tinja River.

One of the distinctive features of this lake is its cyclical hydrological system dynamics. During the wet season, it receives an influx of freshwater from both its catchment area and rainfall. Consequently, this leads to an increase in water levels and a decrease in salinity. In contrast, during the dry season, the prevailing inflow of seawater, coupled with increased evaporation, results in a decline in water levels, and higher water salinity (Sahbani et al., 2022; Aouissi et al. 2014).

Concerning eel landing in the lake, they represent a significant portion of the fish biomass, contributing approximately 38% of the total production during the period 1973–2020. The average total production amounted to 131.73 tons, including 50.80 tons attributed to European eels (Fig.10).

2.2. Data collection

The data set comprised 14 variables collected from 15 stations in the lake (T1 to T15) during the period 2010–2020 (Fig. 1). The sampling stations were selected at the mouths of the tributary rivers and according to the environmental and ecological conditions of the lake.

Eel production data was obtained from the General Directorate of Fisheries and Aquaculture of Tunisia (DGPA, 2020), which controls "Société Tunisie Lagune STL", the only private operator in the Ichkeul ecosystem. STL employs various fishing gears, including trammel nets (with a mesh size of 30/40 mm), creels, and capetchades (mesh size 6 mm), as well as fixed nets (eel weirs) and fixed fisheries (weir deployed near the lock at Tinja River). STL nets are deployed in the afternoon before sunset and recovered the following day in the morning (Derouiche, 2016). The daily quantities of eels caught are recorded and then transmitted to the DGPA every month. The biotic data integrated into the models (in tons) are linked to the date of the available environmental parameters, to build up a comprehensive dataset.

The period factor (*Pr*) represents the temporal aspect, characterized by distinct wet and dry periods. The wet period spans from September to February, encompassing winter and autumn seasons, while the dry period spans from March to August, constituting the spring and summer seasons.

Meteorological parameters, including precipitation P (mm) and wind intensity W ($m.s^{-1}$), were obtained from the National Oceanic and Atmospheric Administration NOAA (http://www.meteomanz.com/index?l=1). The data specifically pertains to the Sidi Ahmed station located in the eastern part of the Ichkeul region.

The physico-chemical, chemical parameters, and chlorophyll-a were derived from the BASSIANA database, a part of the IMAS-Ichkeul project, and from a dedicated sampling campaign conducted in the summer of 2020. The BASSIANA database incorporates diverse data derived from various studies conducted on Ichkeul Lake, including contributions from Abdallah (2017); ANPE, (2017); Shaiek, (2017); Béjaoui et al., (2022); and Brik et al., (2022).

Sampling campaigns were conducted during the dry period, on August 24th and 25th, 2020, at 15 stations in the lake (Fig. 1). This campaign serves a dual purpose. Firstly, it aims to provide an overview of the current state of the lake by collecting data at 15 stations where monitoring was conducted between 2010 and 2017. Secondly, it seeks to assess the lake's condition following the significant rainfall events that occurred between 2017 and 2019 (IUCN, 2020).

The geographical coordinates of the sampling stations were recorded using a GPS device. Physico-chemical parameters such as water temperature T (°C), salinity S (psu), dissolved oxygen DO (mg. l^{-1}), and Turbidity Tur (NTU) were measured in the field using a WTW multiparameter (Multi 340i). Water level WL ($m.s^{-1}$) was assessed using a sonar-based sounder. Water samples were analyzed for ammonium ($NH4^{+}$), nitrate (NO_3^-), nitrite (NO_2^-), and phosphorous (PO_4^{3-}) using an Auto-analyser. The concentrations of these components (μM) were determined calorimetrically using a UV–visible (6400/6405) spectrophotometer (Tréguer and Le Corre, 1975; APHA, 1992). Total nitrogen TN (μM) and total phosphorus TP (μM) were determined after mineralization into ammonia and orthophosphate, respectively (Rodier et al., 1996). The chlorophyll-*a* concentrations Chl.*a* (μg. l^{-1}) were measured using the spectrophotometric method of Lorenzen (1967), after 24-hour extractions in 90% acetone at 4°C in the dark following the procedure given by Parsons et al. (1984). For the present paper, inorganic nutrients were classified into dissolved inorganic nitrogen (DIN = $NO_3^- + NO_2^- + NH_4^+$) (μM) and dissolved inorganic phosphorus (DIP = PO_4^{3-}) (μM).

Table 1
Description of the original dataset:173 observations with 5 factors (Date, Period, Station name, Longitude, Latitude) and 13 parameters at different sampling stations in Ichkeul Lake.

Year	Period	Month	Number of stations	Stations name	Number of observations
2010	Dry	July	3	T13/T16/T1	3
	Wet	January	1	T13	1
		November	3	T13/T16/T1	3
2011	Dry	March	7	T10/T12/T6/ T7/T8/T9/T10	8
		April	8	T13/T1/T10/ T6/T7/T8/T9/ T13	8
		June	1	T13	4
	Wet	September	1	T13	3
		October	8	T1/T10/T12/ T14/T6/T7/ T8/T9	8
		December	3	T13/T8/T9	5
2012	Dry	March	2	T13/T1	2
		July	1	T13	2
	Wet	Junuary	2	T13/T8	5
		February	1	T13	1
		December	4	T3/T4/T13/ T11	4
2013	Dry	June	3	T13/T14/T15	3
		July	2	T13/T5	2
	Wet	January	5	T13/T3/T4/ T5/T6	5
		November	2	T5/T6	2
		December	2	T5/T6	2
2014	Dry	March	3	T13/T1/T2	3
		April	1	T13	2
		June	2	T12/T13	2
	Wet	January	2	T11/T13	2
		December	3	T11/T12/T13	3
2015	Dry	March	2	T10/T13	2
		July	4	T10/T11/T12/ T13	4
	Wet	January	2	T8/T10	2
		November	5	T7/T10/T11/ T12/T13	5
2016	Dry	May	3	T10/T11/T13	3
		June	4	T10/T11/T13/ T15	4
	Wet	October	8	T6/T7/T8/T9/ T10/T11/T12/ T13	10
		November	8	T6/T7/T8/T9/ T10/T11/T12/ T13	9
2017	Dry	March	6	T6/T7/T8/T9/ T10/T12	6
		April	6	T6/T7/T8/T9/ T10/T12	6
		May	6	T6/T7/T8/T9/ T10/T12	6
		August	6	T6/T7/T8/T9/ T10/T12	6
	Wet	October	6	T6/T7/T8/T9/ T10/T12	6
		December	6	T6/T7/T8/T9/ T10/T12	6
2020	Dry	August	15	T1/T2/T3/T4/ T5/T6/T7/T8/ T9/T10/T11/ T12/T13/T14/ T15	15

13 parameters are Precipitation, Wind intensity, Temperature, Water level, Salinity, Dissolved oxygen, Turbidity, DIN, DIP, Total nitrogen, Total phosphorus, Chlorophyll-a, and Eels landing.

In this study, climate parameters were represented by precipitation, wind intensity, and temperature, while human-induced activities are reflected in parameters such as water level, salinity, turbidity, dissolved oxygen, total phosphorus, total nitrogen, DIN, DIP, and chlorophyll-a.

2.3. Data analysis

The original dataset constructed in this study comprises 173 observations, capturing data on 5 factors (date, Pr, station name, longitude, latitude) and 13 parameters (P, W, T, WL, S, DO, Tur, DIN, DIP, TN, TP, Chl.a, and Eels) spanning the period from 2010 to 2020.

Data were collected from a variety of sources, such as the BASSIANA Platform, NOAA's website, DGPA, and a sampling campaign. Due to the diverse origins of the data, the sampling frequency varied, occurring at intervals including weekly, monthly, or seasonally, across different stations in the lake. In Table 1, we provide details on the year and month of sampling, the stations sampled, and the corresponding number of observations.

Furthermore, this dataset contains data gaps with approximately 20% of values missing. To mitigate the impact of these missing data on model accuracy (Umar et al. 2021), two steps were taken. Firstly, the spatial information (i.e., station name, longitude, and latitude columns) was removed, and rows sharing the same date were averaged, thereby reducing the proportion of missing data to 18%. Additionally, the date factor was disregarded, treating the period as the sole temporal factor. Secondly, a machine learning algorithm called "missForest" was applied to statistically impute the missing value. This approach was applied to a dataset comprising 101 observations for 14 parameters (Pr, P, W, T, WL, S, DO, Tur, DIN, DIP, TN, TP, Chl.a, and Eels) for the period 2010–2020.

The missForest algorithm is known for its capability to handle various types of missing data, adapt to interactions and nonlinearity, and perform well with large datasets (Tang and Ishwaran, 2017). Additional information concerning this algorithm is given in the studies of Stekhoven & Bühlmann, (2012).

The performance of this imputation process is evaluated using the normalized root mean squared error (NRMSE) (Oba et al., 2003). A lower NRMSE indicates a more accurate estimation. In the present study, the missForest algorithm was performed using the R package "missForest" (Stekhoven and Bühlmann, 2012).

Additionally, the response variable undertook a Box-Cox transformation to meet the assumptions of normality and homogenous variances required for regression models. Normality was assessed using the Shapiro–Wilk test. The Box-Cox transformation was constructed using the *bcPower* function of the R package "car" version 3.1–0 (Fox et al., 2022).

To address multicollinearity issues and enhance the stability in coefficient estimates, Spearman's rank correlation matrix and Variance Inflation Factor (VIF) were examined. Spearman correlation and VIF were constructed using the R package "gplots" version 3.1.3 (Warnes et al., 2022) and "car" version 3.1–0 (Fox et al., 2022) respectively. In multiple regression analysis, multicollinearity is considered problematic when the correlation coefficient between variables exceeds the absolute value of 0.7 (Fois et al., 2018) and when VIF values are above 5 (Marcoulides and Raykov, 2019). In this study, variables that exceeded these thresholds were excluded from the dataset.

To facilitate the visualization of the data distribution, box and violin plots were performed using the R package "vioplot" version 0.3.7 (Adler et al., 2021). Finally, two machine learning algorithms were applied. The workflow of the methodology is summarized in Fig. 2.

2.4. Machine learning algorithms: Random Forest (RF) and Cubist models (CB)

The RF model, introduced by Breiman in 2001, is a statistical technique widely successful across various fields such as classification, regression, and clustering. It uses Out-of-Bag (OOB) data for model

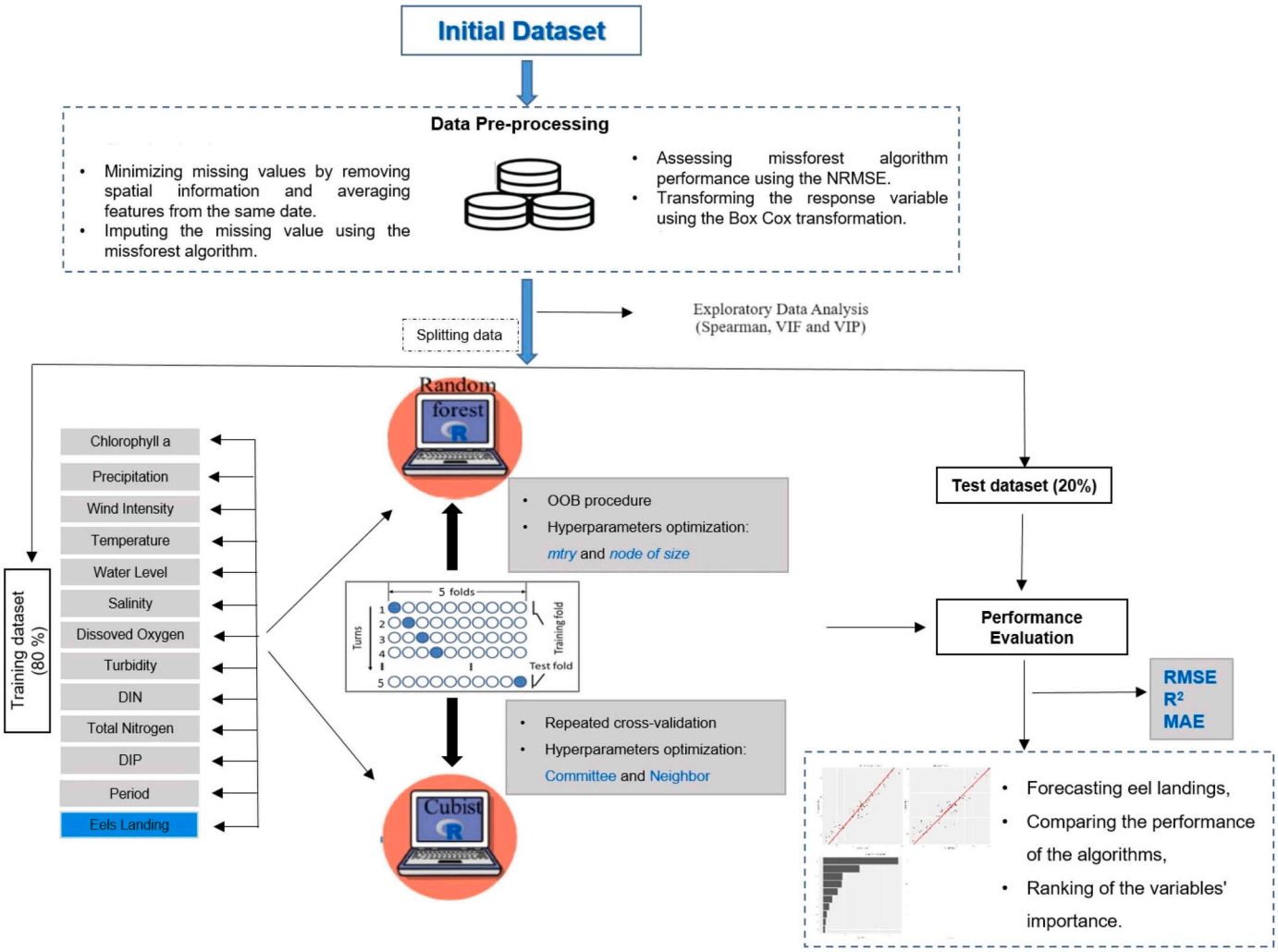

Fig. 2. Detailed illustration of the study methodology.

evaluation, providing a valuable metric of performance (Breiman, 2001).

The CB model, an extension of Quinlan's M5 tree model (John et al., 2020), operates as a regression tree-based model and offers increased interpretability by granting access to the expressions defining rule sets and regression equations (Kumar et al. 2021).

Further details concerning the RF model can be found in Cutler et al. (2007), and in Zhou et al., (2019) concerning the CB approach.

In supervised regression, the dataset is typically split into a training set (80%) and a test set (20%), with the former used for building the prediction model and the latter for evaluating its performance. In this study, a dataset of 101 observations was randomly divided into a training set (81 observations) and a testing set (20 observations) using the R package "Caret" (Kuhn et al. 2022).

2.5. Optimizing the hyperparameters of RF and CB models

To enhance the predictive performance of both RF and CB models, we employed hyperparameter tuning and five-fold cross-validation. For RF, hyperparameter optimization focused on adjusting the minnode size and m_{try} parameters to mitigate overfitting, enhance model diversity, and potentially improve the performance of trees (Breiman, 2001). In contrast, CB tuning involved optimizing the number of rule sets or *committees* and the number of *neighbors* to balance model complexity and predictive performance while avoiding overfitting (John et al., 2020). Five-fold cross-validation was applied to determine optimal hyperparameters by iteratively training the models on different subsets of the dataset and assessing their predictive accuracy (John et al., 2020).

The RF model was performed using the R package "Caret" (Kuhn al., 2022) with the "ranger" method, then it underwent cross-validation using the OOB procedure. The CB model was also executed using the R package "Caret" (Kuhn al., 2022) selecting the "cubist" method, then it underwent cross-validation using the "repeated-cv" method of the "trcontrol" function.

2.6. Performance metrics

The RF and CB models undertook evaluation using both the training and test datasets. We computed metric errors for both the training and test datasets for each model to assess their performance in an out-of-sample context, as opposed to merely in-sample evaluation.

The assessment of the models' performance hinged upon three fundamental metrics: the mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination (R^2), all of which served as robust indicators for predictive accuracy (Culter et al., 2007). An effective model prediction would yield low MAE and RMSE values, approximating 0, and an R^2 value approaching 1 (Culter et al., 2007).

MAE, as the mean of all absolute errors, stood out as a widely employed measure for evaluating model performance (Zhou et al., 2019). On the other hand, RMSE provided a quantitative gauge of the model's predictive error (Culter et al., 2007). As for R^2, it served as an indicator of the alignment between observed and predicted values (Zhou

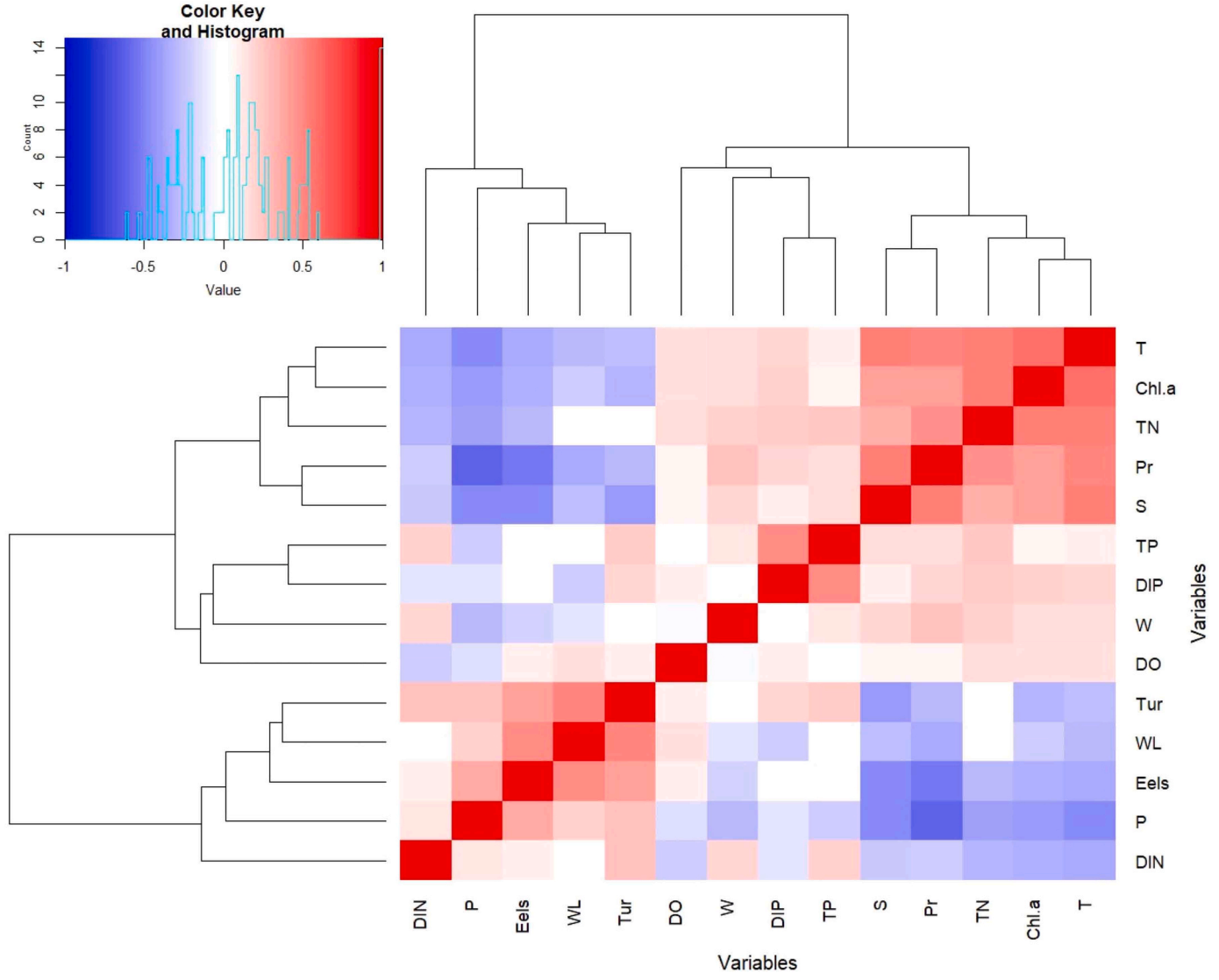

Fig. 3. Heatmap of Spearman correlation between variables for the period 2010–2020 in Ichkeul Lake. The colors represent the correlation, with red being more positive and bluer more negative. The variables included period Pr (dry or wet), precipitation P (mm), wind intensity W ($m.s^{-1}$), water temperature T (°C), water level WL ($m.s^{-1}$), salinity S (psu), dissolved oxygen DO ($mg.\ l^{-1}$), turbidity Tur (NTU), dissolved inorganic nitrogen DIN (μM), dissolved inorganic phosphorus DIP (μM), total nitrogen TN (μM), total phosphorus TP (μM), chlorophyll-*a* Chl.a ($\mu g.\ l^{-1}$), and European eels landing Eels (tons).

et al. 2019). Additionally, actual versus predicted regression plots were employed to visually represent the model's performance in capturing eel landings.

Further, features importance ranking is used to measure the degree of contribution of each variable to the estimation based on the R package "VIP" (Greenwell et al., 2020).

3. Results

3.1. Data pre-processing

The missForest algorithm produced low NRMSE (0.22), indicating a good model performance.

The Box-Cox transformation (Fox et al. 2022) of the target variable ensured that the indicator was centered and scaled. The Spearman correlation revealed a strong correlation among various predictors: DIP and total phosphorus (TP), chlorophyll-a (Ch.a) and temperature (T), salinity (S) and period (Pr), turbidity (Tur) and water level (WL), and between precipitation (P) and period (Pr) (Fig. 3). The VIF analysis confirmed the strong correlation only between DIP and TP.

Table 2
Values of the Variance Inflation Factor (VIF) for predictors variables.

Parameter	VIF values
Precipitation	1.6
Wind Intensity	1.3
Temperature	1.4
Water Level	1.8
Salinity	2.1
Dissolved Oxygen	1.7
Turbidity	2.1
Total Nitrogen	2.7
Dissolved Inorganic Nitrogen	1.8
Dissolved Inorganic Phosphorus	3.4
Total Phosphorus	4.6
Chlorophyll-a	1.9

Consequently, we excluded TP with higher VIF (Table 2), resulting in 13 variables to construct the prediction models.

Furthermore, the Spearman test revealed significant correlations between eel landings and various factors. There was a negative

Table 3
Summary of the dataset post-transformation (removal of spatial information and imputation of missing values): 101 observations with 14 parameters for the period 2010–2020. The temporal factor corresponds to the *period* (*Pr*).

		2010	2011	2012	2013	2014	2015	2016	2017	Summer 2020	2010–2020
P(mm)	Min-Max	0.00–107.42	0.00–182.62	0.00–182.89	0.00–249.93	0.00–113.03	1.27–172.20	0.00–149.09	0.00–119.12	0.00–7.37	0.00–249.39
	Mean	61.49	77.63	53.34	90.73	51.46	63.04	42.76	50.01	2.56	54.78
	SD	23.20	55.02	60.57	77.26	43.22	49.83	50.92	48.80	4.17	45.89
W ($m.s^{-1}$)	Min-Max	3.50–7.97	2.41–8.00	3.48–7.52	2.95–8.15	3.35–8.15	3.69–8.81	2.99–8.90	2.66–7.14	3.64–6.22	2.41–8.81
	Mean	5.93	4.91	5.69	5.39	5.71	5.74	5.53	5.49	5.33	5.52
	SD	1.63	1.44	1.34	1.64	1.50	1.42	1.69	1.44	1.47	1.51
T (°C)	Min-Max	12.03–15.97	10.00–27.30	10.40–25.73	10.60–40.36	11.40–26.96	12.35–62.75	10.3–26.20	11.7–28.40	23.70–28.6	10.00–62.75
	Mean	14.06	16.47	18.94	16.55	15.53	19.03	17.37	17.04	25.93	17.88
	SD	1.80	5.14	5.85	8.74	5.30	13.78	5.80	5.63	2.48	6.06
WL (cm)	Min-Max	31.80–73.00	10.00–156.00	3.00–151.00	16.14–142.11	15.32–113.00	25.21–140.00	32.98–78.00	47.89–122.00	35.45–58.61	3.00–156.00
	Mean	50.25	50.38	63.71	58.01	64.35	65.33	53.58	76.87	46.52	58.78
	SD	20.53	51.37	59.44	37.95	46.59	39.16	16.18	22.28	11.61	33.90
S (psu)	Min-Max	13.07–59.59	3.43–56.01	6.60–67.10	3.79–70.14	5.031–46.34	7.45–41.11	7.58–71.00	8.45–56.4	53.57–65.50	3.43–71.00
	Mean	31.89	30.07	29.30	24.54	29.47	21.41	29.40	31.33	59.72	31.90
	SD	15.37	16.86	20.92	19.56	12.49	11.16	20.76	15.44	5.97	15.39
DO ($mg.\ l^{-1}$)	Min-Max	5.20–11.40	4.20–14.00	4.30–9.08	3.20–13.35	4.67–13.00	4.50–8.54	3.60–10.10	3.70–11.00	7.20–8.22	3.20–14.00
	Mean	8.13	7.915	6.73	7.11	8.58	6.27	7.18	7.26	7.88	7.45
	SD	3.04	2.80	1.83	2.48	2.68	1.29	1.81	2.31	0.59	2.10
Tur (NTU)	Min-Max	16.01–33.00	11.17–45.00	11.00–40.00	11.10–46.00	12.30–33.90	16.77–44.00	17.00–32.00	10.14–40.00	11.00–11.20	10.14–45.00
	Mean	25.53	24.26	25.62	24.75	22.39	26.56	25.21	21.41	11.11	22.98
	SD	7.49	10.43	10.74	11.57	7.10	9.99	5.26	9.55	0.60	8.08
DIN (μM)	Min-Max	10.00–17.30	15.00–22.84	18.69–18.69	32.00–32.58	18.43–21.00	12.30–20.32	12.00–23.56	13.20–28.41	14.15–15.02	10.00–32.58
	Mean	13.67	18.43	18.69	32.28	20.02	16.72	18.98	20.51	14.58	19.32
	SD	0.90	3.18	1.20	0.60	1.17	3.16	4.63	4.15	1.34	2.26
TN (μM)	Min-Max	11.20–24.10	6.45–50.03	11.01–39.89	11.20–56.60	15.30–47.70	8.34–45.50	11.73–73.51	6.21–40.04	14.80–19.01	6.21–73.51
	Mean	16.73	24.20	19.47	34.86	26.93	23.86	30.93	20.66	16.94	23.84
	SD	5.21	13.37	8.98	13.68	12.64	11.16	20.33	10.65	2.11	10.90
DIP (μM)	Min-Max	0.76–2.03	0.23–7.40	0.26–3.24	0.28–1.02	0.44–2.45	0.25–6.43	0.10–7.62	0.32–2.22	2.00–6.30	0.28–7.62
	Mean	1.39	1.95	1.21	0.72	0.95	1.40	1.98	0.74	4.90	1.69
	SD	0.90	2.41	1.18	0.39	0.62	1.92	2.79	0.59	8.80	2.18
TP (μM)	Min-Max	2.22–5.34	3.65–30.1	2.10–4.29	3.00–3.56	1.32–5.76	2.12–21.3	4.54–36.42	3.43–20.30	3.00–6.50	1.32–36.4
	Mean	3.67	9.68	4.00	3.45	3.71	8.88	11.89	7.22	5.00	6.38
	SD	1.2	9.51	1.1	1.2	1.83	7.27	13.76	5.43	6.30	5.29
Chl.a ($\mu g.\ l^{-1}$)	Min-Max	1.21–8.60	1.23–7.50	2.12–9.08	1.32–8.76	2.03–8.01	1.24–8.67	2.56–9.14	1.87–9.80	8.49–8.70	1.24–9.80
	Mean	3.69	4.13	5.75	3.29	3.94	4.27	6.84	5.15	8.59	5.07
	SD	3.16	1.92	2.89	1.96	2.04	2.54	2.93	2.49	0.50	2.27
Eels (tons)	Min-Max	4.69–20.02	0.02–45.00	0.01–29.56	0.08–22.05	0.01–21.80	0.08–44.30	0.50–14.26	0.24–13.03	0.00–0.49	0.01–45.00
	Mean	12.37	7.20	7.09	6.97	7.84	6.18	5.77	4.61	0.29	6.48
	SD	7.27	11.73	11.06	6.29	6.58	11.67	4.75	4.39	3.00	7.41
Number of observations (n)		6	19	10	14	11	13	12	13	3	101

Notes: SD: standard deviation, P: precipitation, W: wind intensity, T: water temperature, WL: water level, S: salinity, DO: dissolved oxygen, Tur: turbidity, DIN: dissolved inorganic nitrogen, DIP: dissolved inorganic phosphorus, TN: total nitrogen, TP: total phosphorus, Chl.a: chlorophyll-a, and Eels: european eels landing.

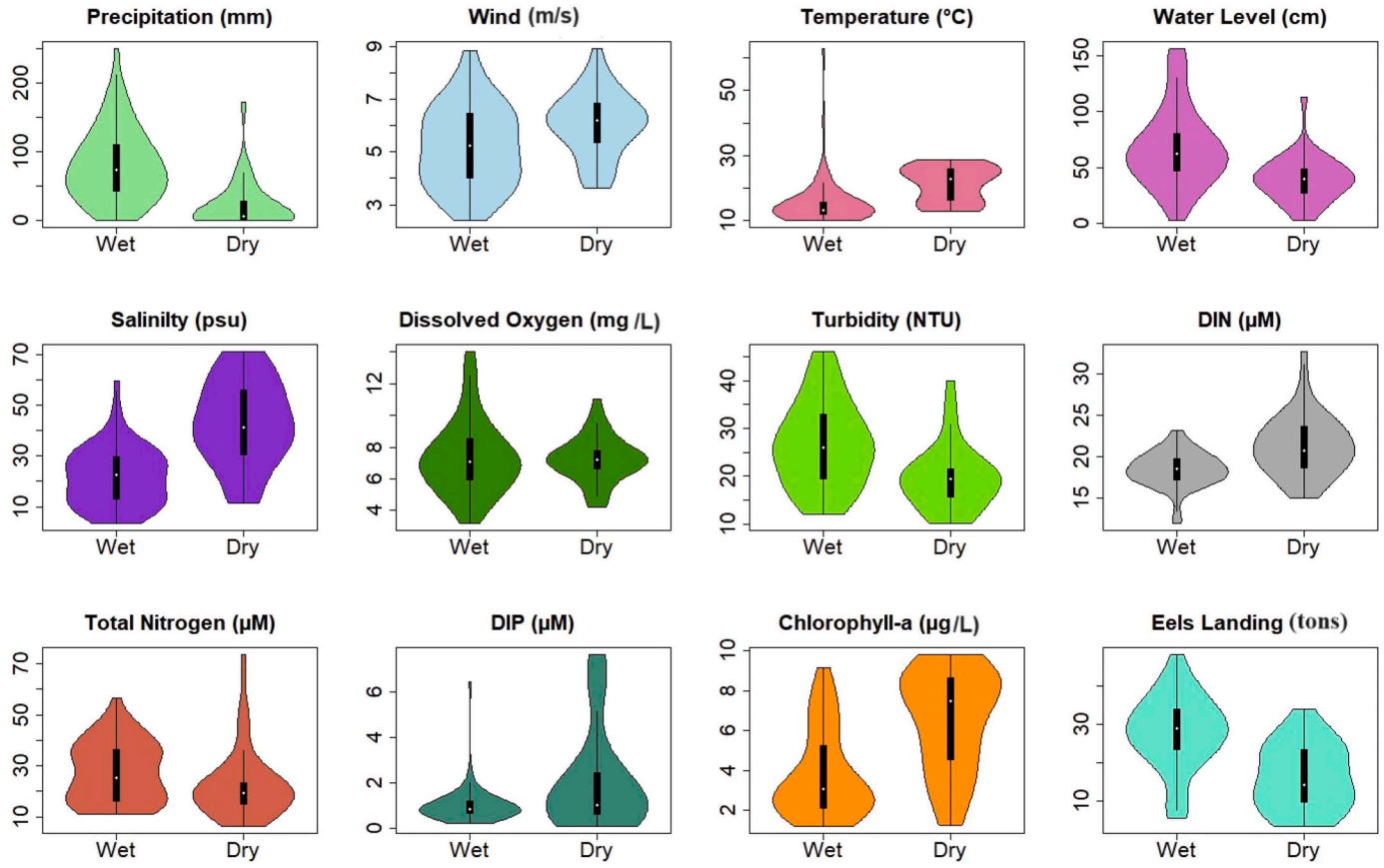

Fig. 4. Data visualization with box and violin plots VIP. The shaded area indicates quartiles and data densities. The third quartile and first quartile are indicated as the lower and upper boundaries of the thick line at the center of the VIP. A broader VIP indicates a higher data density. Within the VIP, the top and bottom extremes represent the maximum and minimum values, respectively. The white dot positioned in the center of the plot corresponds to the median. Box and whisker plots show outliers.

correlation observed between eel landings and both salinity (r=-0.50) and period (r=-0.70), indicating that eel landings decreased as salinity and in the dry period. Conversely, there was a positive correlation between eel landings and water level (r=0.50) as well as turbidity (r=0.45), suggesting that eel landings tended to increase with higher water level and turbidity level.

3.2. Parameter's properties

Table 3 provides a comprehensive summary of the dataset post-transformation, which includes the removal of spatial information and imputation of missing values, but before the exclusion of the TP parameter. It provides insights into the yearly variations in the environmental parameters of Ichkeul Lake, reflecting changes observed from 2010 to 2020. Notably, the data for 2020 covers the dry period, thus presenting mean and standard deviation (SD) values tailored to this timeframe. In contrast, Fig. 4 delves into seasonal fluctuations, providing a detailed examination of the variations experienced throughout the dry and wet periods.

From meteorological parameters, the wet period exhibits higher rainfall (90.57% between 0 and 250 mm) and wind intensity (4–6 m/s) compared to the drier period (9.42% rainfall, 5.22 m/s wind).

Concerning physicochemical parameters, the mean annual water temperature stood at approximately 17.90°C. The averages for water level (WL), dissolved oxygen (DO), and turbidity (Tur) exhibited a significant increase during the wet period compared to the dry period, primarily attributed to the influx of freshwater from the rivers. Conversely, the average salinity (S) demonstrated an inverse trend, showing an increase during the dry period (40–70 psu) owing to the sea water inflow from Bizerte lagoon.

In terms of chemical parameters, dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP), and total nitrogen (TN) exhibited distinct seasonal patterns, with two prominent episodes annually. Elevated levels during the dry period can be attributed to the extensive use of chemical fertilizers nearby. Peak values for DIN, TN, and DIP reached 18, 55, and 1 µM, respectively, during the wet season. Conversely, during the dry period they reached 30, 20, and 1.5 µM, respectively.

The chlorophyll-a concentration (Chl.a), indicating phytoplankton biomass, ranged from 0 µg. l^{-1} to 10 µg. l^{-1}, a value characteristic of eutrophic ecosystems (Vollenweider et al., 1998).

As regards the eel landings, substantial catches were observed during the wet period, reaching up to 45 tons. In contrast, during the dry period, eel catches were lower, due to eel migration to the Sargasso Sea for reproduction.

3.3. Results of hyper-parameters tuning

In both the RF and CB models, the input parameters, including meteorological (P and W), physicochemical (T, WL, S, DO, and Tur), chemical variables (DIN, DIP, and TN) and Chl.a play a significant role in predicting the target variable, which is eel landings.

The RF model's optimal tuning values were $m_{try} = 9$ and min node size = 9, whereas the final values for the CB model were *committees* = 5 and *neighbors* = 9 (Fig. 5).

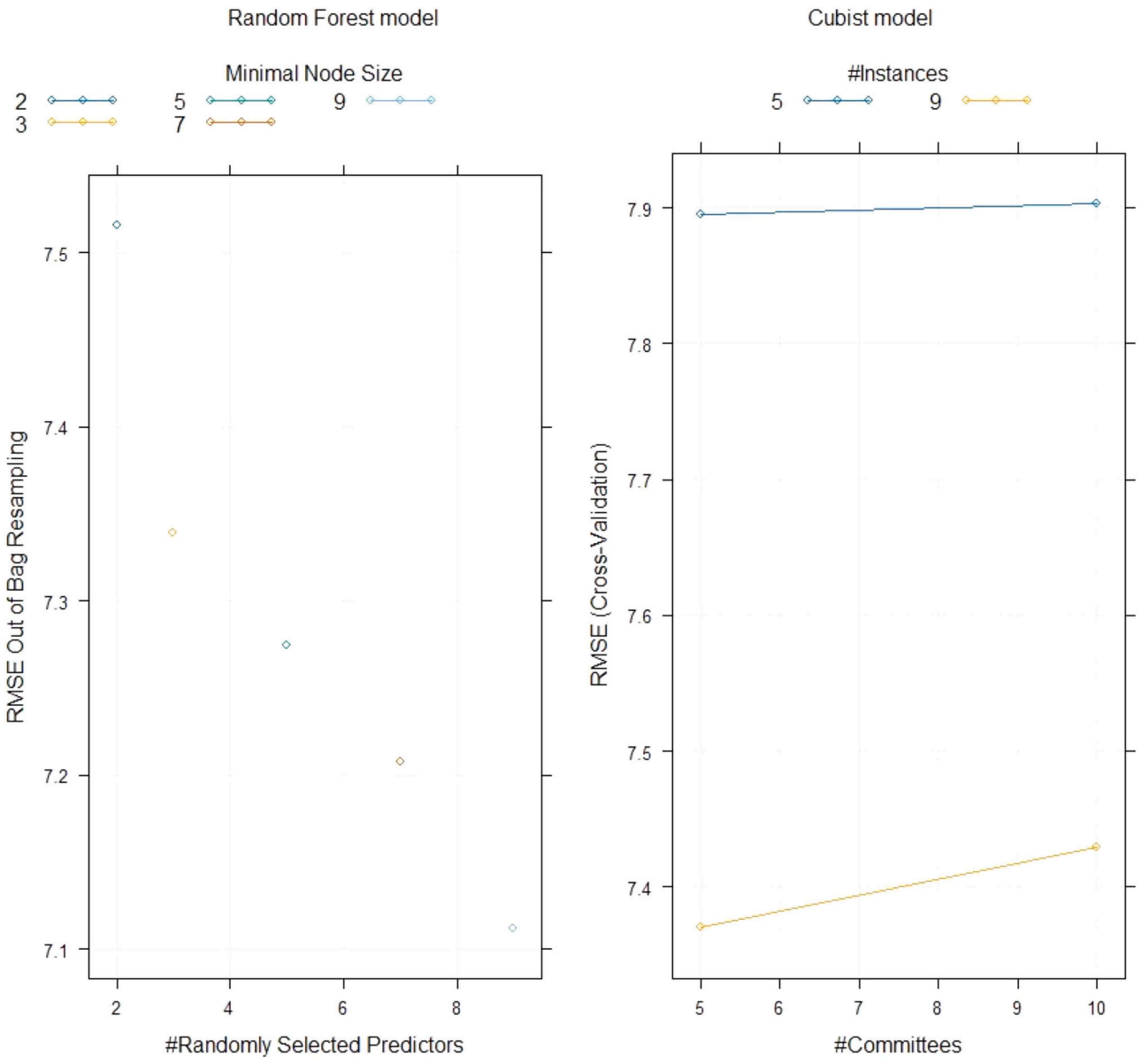

Fig. 5. Five-fold cross-validated RMSE profiles for determining the optimal tuning parameters for the Random forest (m_{try} and *minimal node size*) and the Cubist models (*committees* and *instances*).

Table 4
Performance Metrics of Random Forest, Cubist, and Multiple Regression Models for Predicting Training Dataset. Averaging Eel Landing at 24.16 tons over the Period 2010–2020.

Effectiveness metrics	Random Forest	Cubist	Multiple regression (for comparison)
RMSE (tons)	7.20	7.68	6.24
R^2	0.56	0.55	0.64
MAE (tons)	5.65	6.20	5.20

Table 5
Performance Metrics of Random Forest, Cubist, and Multiple Regression Models for Predicting Testing Dataset. Averaging Eel Landing at 24.16 tons over the Period 2010–2020.

Effectiveness metrics	Random Forest	Cubist	Multiple regression (for comparison)
RMSE (tons)	5.81	5.13	7.99
R^2	0.73	0.73	0.41
MAE (tons)	4.97	5.89	6.55

3.4. Results of performance metrics

Table 4 displays the performance metrics for the training dataset of the RF and CB models, and for comparative purposes, a Multiple Regression (MR) model was included in the analysis. The MR model demonstrates the highest R^2 values and the lowest RMSE (tons), and MAE (tons), values, while the RF and CB models exhibit similar performance metrics.

Table 5 presents the performance metrics values for the test dataset for the RF, CB, and MR models. We note that the MR model's predictions are influenced by training data, resulting in instability with the test dataset variation, while RF and CB models consistently offer more stable predictions.

Fig. 6 indicates significant disparities between the predicted values and the actual observations of eel landings for the training dataset. Specifically, for the RF model, a systematic tendency emerges with a

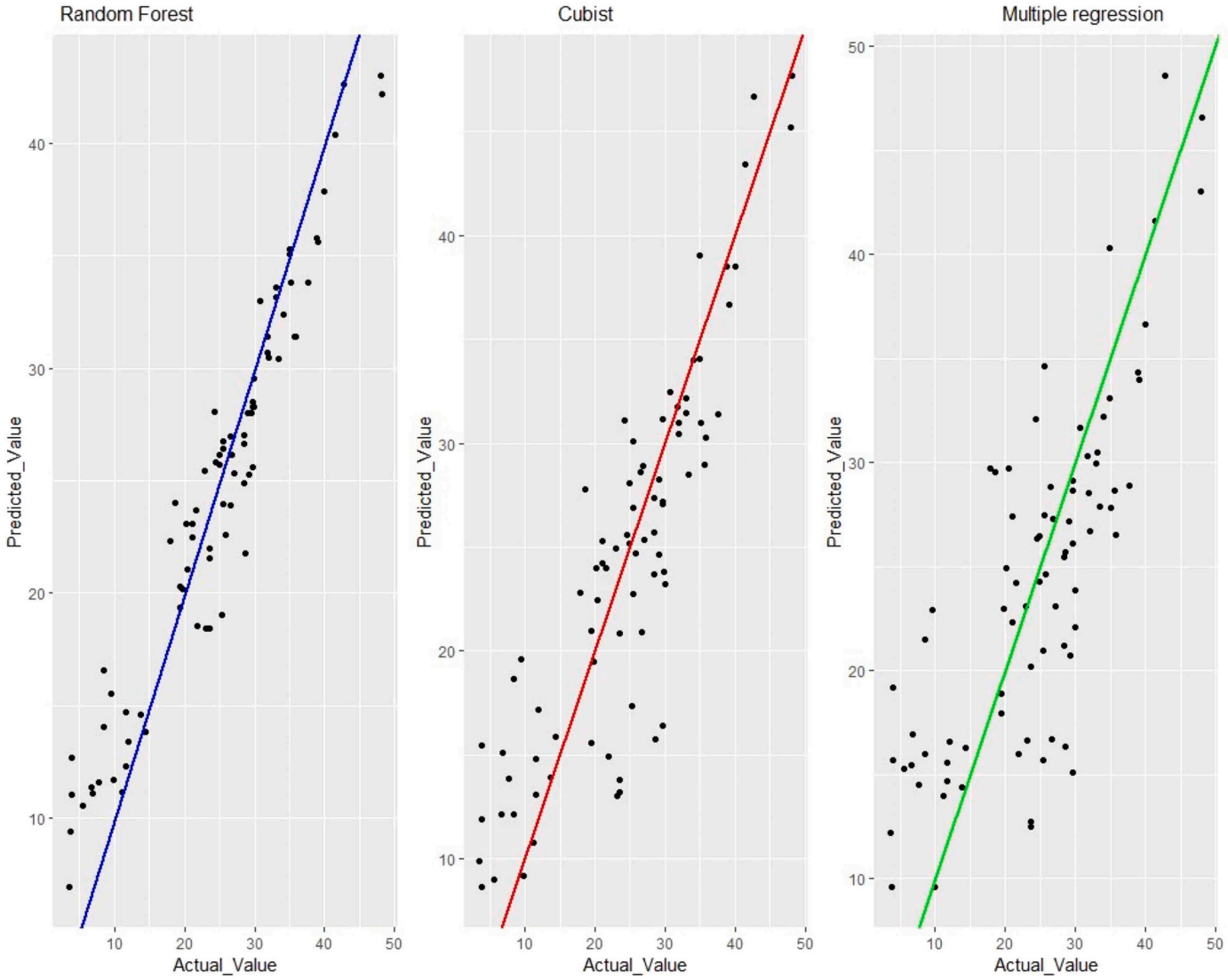

Fig. 6. Scatter plot comparing the observed and the predicted eel landing using Random Forest, Cubist, and Multiple regression models, with R-squared values of 56%, 55%, and 64%, respectively. The black points represent the model training dataset.

trend to overestimate lower values and underestimate higher values of eel landings. Remarkably, this pattern persists even though the biotic data were scaled and centered. This observation suggests that the model's residuals are not randomly distributed but rather exhibit a correlation with the magnitude of the data, as shown in Fig. 7. In stark contrast, the actual vs. predicted regression plot for the CB model showcases markedly improved prediction performance. This leads us to conclude that the CB model surpasses both the RF and MR models in predictive accuracy.

3.5. Determining variables' importance and forecasting eel landing

Given the better performance of the CB model, we applied it to determine the variable's importance, which emphasized that water level, salinity, and turbidity are the most influential predictors followed closely by nitrogen and temperature as illustrated in Fig. 8. However, wind intensity and dissolved inorganic phosphorus (DIP) did not significantly contribute to the model's predictive performance.

In a practical application of this approach, we leveraged an environmental dataset (15 observations with 12 parameters) gathered on January 24 and 25, 2022 in 15 stations in the lake. Faced with a lack of eel landing data, we chose to use the CB model developed for eel landing forecasting. According to this model, eel landings were anticipated to range between 22 and 47 tons during the period from January 24–25, 2022. Fig. 9 shows the forecasted eel landings relative to the water level.

4. Discussion

4.1. Ecological relevance of the study

In this survey, we explored variations in environmental parameters in Ichkeul Lake. Temperature trends aligned with the typical patterns in the Mediterranean coastal marine climate (Béjaoui et al., 2016; Ben Hadid, 2021; Dhib et al. 2016). Our findings unveiled a decline in water levels compared to prior studies (Hollis et al., 1986; ANPE, 2008), coupled with a notable increase in salinity, as demonstrated in Sahbani et al. (2022). These shifts are linked to anthropogenic disturbances, including dam construction on main rivers and the installation of sluice gates at the Tinja River outlet.

Field observations also indicated increased nitrogen and phosphorus levels during the study period (2010–2020) compared to the pre-dam era in both 1977 and 1992–1993 (Dridi, 1977, Ben Rejeb-Jenhani, 1992, Chaouachi et al., 2001). This could be attributed to pollutant inputs and untreated wastewater from watershed rivers (Madyouni et al.,

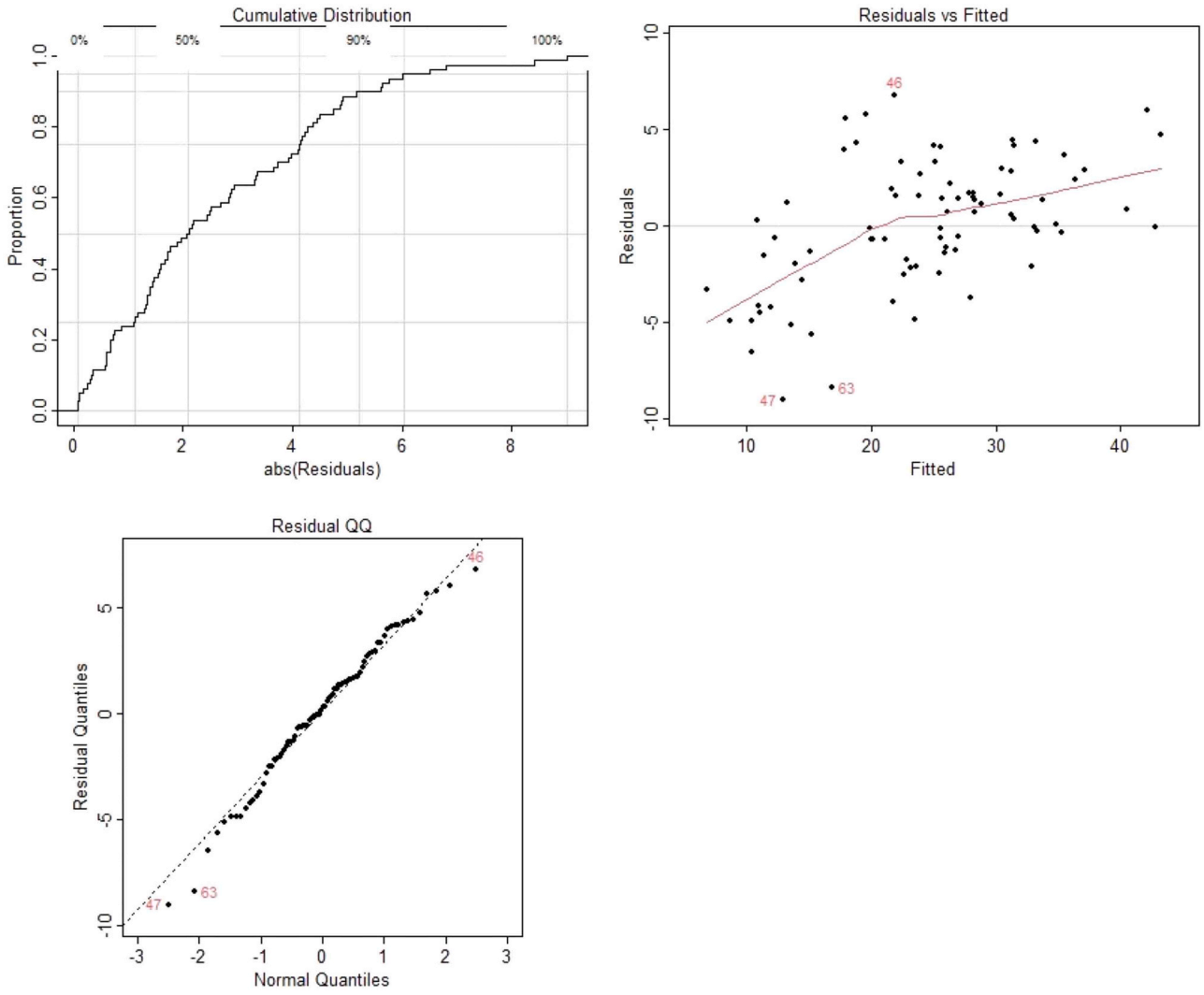

Fig. 7. Random Forest model residual plots. The top right is a standard Residuals vs Fitted plot of the training data, accompanied by a LOWESS (Locally Weighted Scatterplot Smoothing) smooth represented by the red line. Cases 46, 47, and 63 have the largest residuals with corresponding fitted values of 22, 14, and 17. The QQ plot (bottom left) reveals that these residuals deviate somewhat from expected norms. Given their status as the largest absolute residuals, they determine the right bound of the Cumulative Distribution Plot (top left).

2023). Additionally, our study highlighted seasonal fluctuations in these components, with higher levels during the dry season, attributed to the intensive use of agricultural fertilizers during this period (Aouissi et al., 2014).

The DIN/DIP ratio, which varies from 2.83 to 182 and from 2.47 to 223 during wet and dry periods respectively, underscores an imbalance in the lake's ecosystem. This higlights the primary limitation of phytoplankton growth by phosphorus availability. Concerning the eutrophication indicator (expressed as chlorophyll-a) and juxtaposed with the threshold determined by Vollenweider et al., (1998), the observed Chl.a values suggest eutrophic conditions in Ichkeul Lake.

Changes in lake conditions affected the ecosystem's flora and fauna, notably influencing eels in their continental-rearing habitats. These changes can influence eel movements directly by limiting energy expenditure or indirectly by affecting factors like food availability (Riley et al., 2011). To shed light on these impacts on the eel population in Ichkeul Lake, the results of the Random Forest and Cubist models are of particular interest.

4.2. Evaluation of the model's performance

In our analysis, we noted the Multiple Regression (MR) model's limited adaptability to variations in the test dataset, and for the RF model, non-random correlations between residuals and data magnitude were observed. These outcomes highlighted the outstanding performance of the CB model, attributed to its unique ability to demonstrate a delicate balance between predictive accuracy and interpretability (Zhou et al. 2019). Its versatility makes it highly advantageous in diverse practical applications, as it excels in capturing relationships between variables and the target property (Zhou et al. 2019).

Developing a predictive model customized for the Ichkeul ecosystem, capable of pinpointing the key factors driving eel landings and offering forecasts of future eel populations based on environmental variables, stands as the optimal approach for preserving the lake's environmental integrity over the long haul.

Machine learning techniques also offer a powerful means to forecast future eel landings by incorporating forthcoming environmental parameters as inputs. By envisioning various hypothetical scenarios of environmental change, these techniques can anticipate environmental parameters and generate response predictions for each scenario, yielding a probability distribution of future landings. In our case, a potential limitation arises when the model is trained on a limited range of observed environmental values over 10 years, compromising prediction accuracy when future observations significantly differ from the training

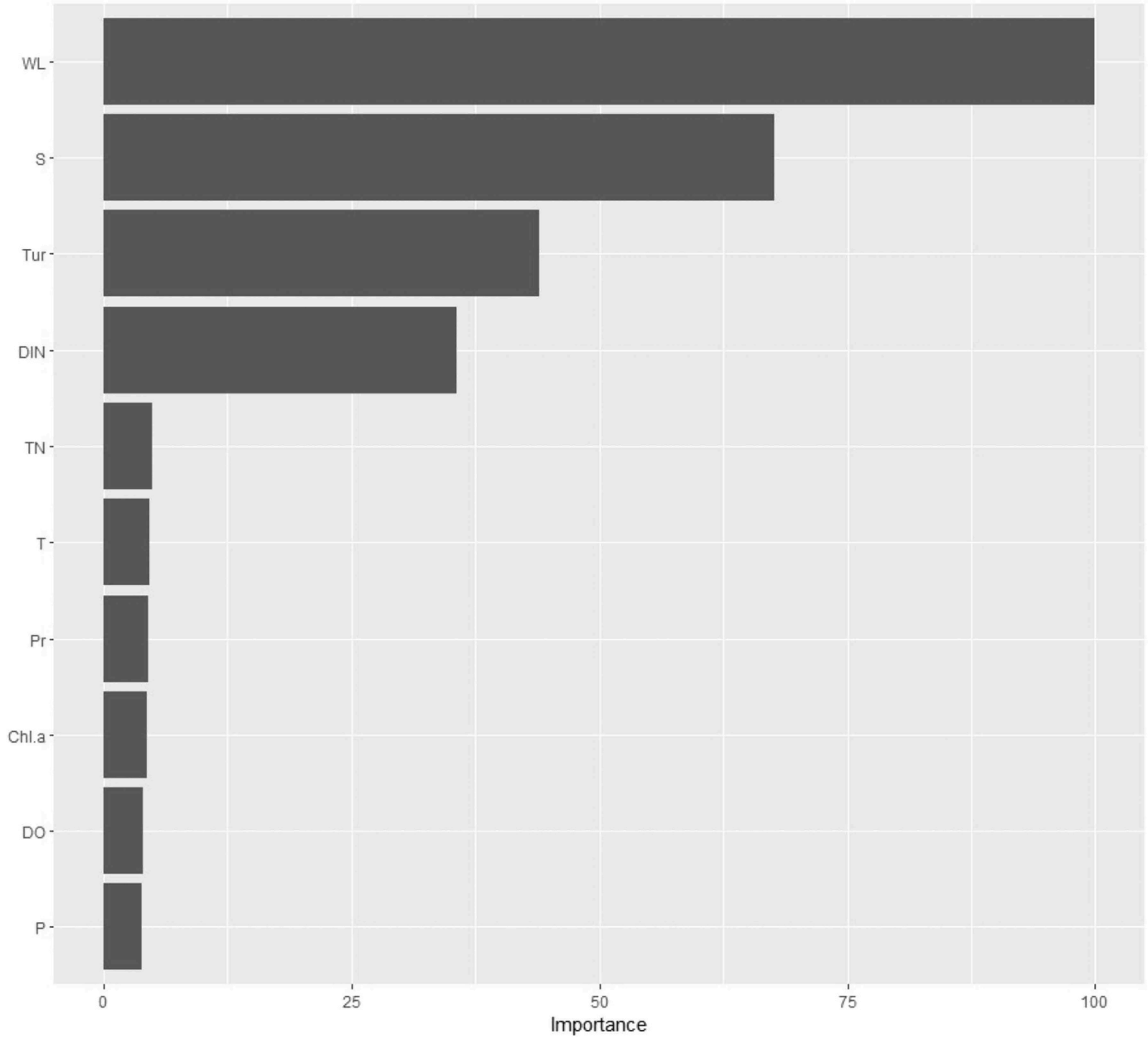

Fig. 8. Variable importance scores for the predictors in the Cubist Model for eels landing using a dataset composed of 101 observations. The 12 predictors are Pr (Period), P (precipitation), W (wind intensity), T (water temperature), WL (water level), S (salinity), DO (dissolved oxygen), Tur (turbidity), DIN (dissolved inorganic nitrogen), DIP (dissolved inorganic phosphorus), TN (total nitrogen), and Chl.a (chlorophyll-*a*). W and DIP did not significantly contribute to the model's predictive performance.

range. In recent years, a growing number of studies have sought to synergize traditional forecasting models with the power of machine learning to improve the accuracy and efficiency of predicting response variables. This fusion of methodologies has given rise to a burgeoning field known as "physics-informed machine learning (PILM)". PILM represents a crucial frontier in the field of predictive modeling, focusing on the incorporation of domain-specific physical principles into data-driven forecasting approaches (Truong et al., 2023).

The results obtained from the CB model in this study outperformed the RF model outcomes observed in Bizerte Lagoon (R^2= 0.51) (Béjaoui et al., 2016). This difference in performance can be attributed to several factors, including the specific dataset used, the unique hydrological processes of each ecosystem, and the distinct environmental stressors that each ecosystem faces.

4.3. *Effect of environmental factors on the eel population*

The findings derived from the CB model shed light on the most influential factors affecting the presence of eels among all variables. These key factors include water level, salinity, turbidity, and nutrient concentrations.

Eel movement is primarily driven by water levels, with their entry into the lake occurring between September and May when water levels rise and flow toward the Bizerte lagoon (Derouiche, 2016). Therefore, as a fundamental step in management, it is advisable to refrain from dredging activities during this timeframe, as these are the conditions in which non-migratory eels become most active and exhibit increased mobility.

In Ichkeul Lake, the eel intensity was low when the salinity increased. Several studies, such as those of Tosi et al. (1990) and Edeline

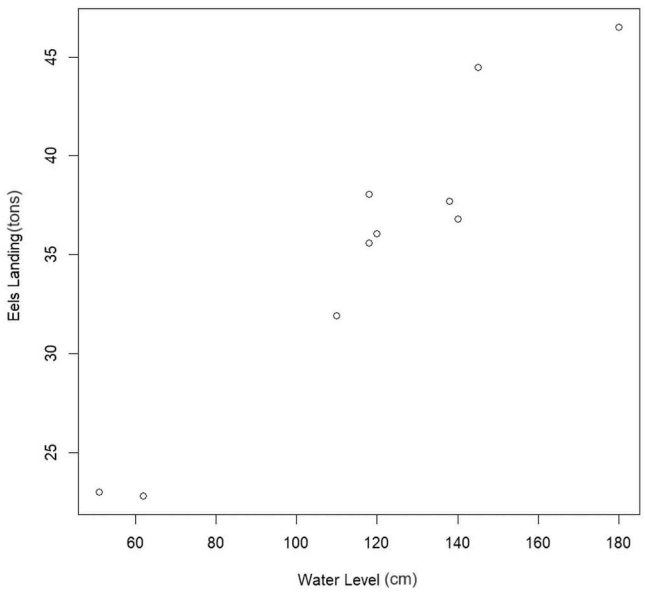

Fig. 9. Forecasted eel landing relative to water level using the Cubist Model and environmental data (n= 15 and 12 parameters) across 15 Lake stations on January 24–25, 2022. The 12 parameters are Pr (Period), P (precipitation), W (wind intensity), T (water temperature), WL (water level), S (salinity), DO (dissolved oxygen), Tur (turbidity), DIN (dissolved inorganic nitrogen), DIP (dissolved inorganic phosphorus), TN (total nitrogen), and Chl.a (chlorophyll-*a)*.

et al., (2005) have consistently emphasized the eel's preference for either high or low-salinity environments. Upon arrival from the sea, eels show a flexible pattern of colonization of continental habitats, either migrating upstream (freshwater) or settling in marine and estuarine habitats, due to endocrine controls and genetic factors. Indeed, high thyroid hormone (TH) and low allozygous heterozygosity promote a preference for freshwater (Edeline et al., 2005). These conditions foster the production of large female eels (Oliveira et al. 2001), which is illustrated in the Ichkeul Lake by the study of Derouiche (2016). At the same time, low TH and high allozygous heterozygosity favored marine preference (Tosi et al., 1990).

Regarding the period factor (the distinction between dry and wet periods), it is important to highlight its minimal impact on eel landings. This is primarily because eel species are predominantly caught during wet periods; however, they also inhabit the lake during dry periods to complete their growth phase.

The significant abundance of eels found in the lake with heightened turbidity and nutrient levels can be attributed to their foraging behavior and the avoidance of predators (Lagarde et al., 2021). However, it's important to recognize that fish activities, including swimming and searching for food within sediments, also play a substantial role in sediment resuspension. Moreover, these activities can contribute to the regulation of phytoplankton biomass and community structure in shallow lakes (Havens, 1991).

Eel species display an impressive degree of trophic plasticity and adaptability in response to changing ecosystem conditions. For instance, in eutrophic conditions, they undergo a transition in their trophic strategies. Instead of primarily preying on fish as piscivorous and pelagic predators, they shift towards an omnivorous and benthic predator mode, targeting invertebrates as their primary prey. This adaptive strategy is particularly evident in the case of Ichkeul Lake (Shaiek. 2017). Residing in eutrophic conditions presents certain challenges for eels, such as a reduction of niche overlap and, consequently, competition with other fish species (Caputi et al., 2020). Therefore, managers should prioritize conservation efforts and continuously monitor the lake's trophic and ecological dynamics. This multifaceted approach ensures the sustainable coexistence of eels and other species within these complex environments.

5. Conclusion

This paper has shown the robustness of the CB algorithm in predicting the most influential predictors affecting eel presence, notably water level and salinity, followed closely by turbidity and nutrient levels. We advocate for incorporating these variables into watershed management plans, benefiting not only Ichkeul Lake but also other Tunisian coastal ecosystems.

we also advocate for the integration of advanced technologies into local management agencies' monitoring systems to track the trophic and ecological status of the lake, ensuring its long-term ecological integrity. By providing managers with forecasts on faunal communities, the proposed approach facilitates informed decisions on stock management, fishing accessibility, and quantity assessment, leading to optimized fish stocks, species preservation, and sustainable food production. Furthermore, the model enhanced government oversight of concession companies and aids in estimating illegal production captured by residents of Tinja town, providing improved monitoring and data reliability.

It is demonstrated that the CB technique is well-suited for our dataset type. To further gauge the CB model's performance. Additional experiments with larger datasets concerning the environmental parameters are also recommended. Other biotic data, including for instance habitat structures and specific richness are needed to capture the multifaceted relationships within the ecosystem. Furthermore, we promote the integration of socio-economic data related to land use changes, fishing practices, and pollution levels, which will offer a more holistic perspective on human impacts. This interdisciplinary approach will strengthen the model's ability to anticipate the interactions between environmental and anthropogenic factors. These endeavors will contribute valuable insights for future research and management strategies.

CRediT authorship contribution statement

Sabrine Sahbani: Writing – review & editing, Writing – original draft, Software, Methodology. **Béchir Béjaoui:** Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Ennio Ottaviani:** Validation, Supervision, Software, Methodology. **Sihem Benabdallah:** Writing – review & editing, Validation, Investigation. **Eva Riccomagno:** Writing – review & editing. **Enrico Prampolini:** Software. **Donata Melaku Canu:** Writing – review & editing. **Hechmi Missaoui:** Writing – review & editing, Supervision, Resources. **Cosimo Solidoro:** Validation, Methodology, Investigation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgements

This study was carried out within the framework of IMAS-Ichkeul project (An Integrated Modeling Approach for sustainable Development of the Ichkeul Lake Eco-Tourism & Aquaculture) funded by the Partnerships for Enhanced Engagement in Research (PEER) program under USAID cooperative agreement AID-OAA-A-11–00012. We would like to thank the reviewers for their invaluable comments during the

revision process.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.rsma.2024.103587.

References

Abdallah, T., 2017. Caractérisation hydro-biologique du Lac Ichkeul et modélisation des échanges Lac Ichkeul-Lagune de Bizerte. Master de recherche. Ecole National des Ingénieurs de Tunis ENIT (83p).

Adler, D., Kelly, S.T., Elliott, T.M. 2021. Package 'Vioplot' Violin Plot, Version 0.3.7.pp. 1-10.

ANPE (National Agency for Environmental Protection). 2008. Rapport sur le suivi scientifique au Parc National d'Ichkeul, année 2006-2007. National Report. 1-84.

ANPE (National Agency for Environnemental protection). 2017. Reserve de la Biosphère: Parc National d'Ichkeul-Tunisie. National Report. 1–18.

Aouissi, J., Benabdallah, S., Chabaane, Z.L., Cudennec, C., 2014. Modeling water quality to improve agricultural practices and land management in a tunisian catchment using the soil and water assessment tool. Environ. Qual. 43, 18–25. https://doi.org/10.2134/jeq2011.0375.

APHA, 1992. Standard Methods for the Examination of Water and Wastewater (American Public Health Association). American Public Health Association, Washington, DC.

Bardossy, A., Van Mierlo, J.M.C., 2000. Scénarios régionaux de précipitations et de températures dans la perspective d'un changement climatiques. Hydrol. Sci. J. *45* (4), 559–575. https://doi.org/10.1080/02626660009492357.

Béjaoui, B., Armi, Z., Ottaviani, E., Barelli, E., Gargouri-Ellouz, E., Chérif, R., Turki, S., Solidoro, C., Aleya, L., 2016. Random Forest model and TRIX used in combination to assess and diagnose the trophic status of Bizerte Lagoon, southern Mediterranean. Ecol. Indic. 71, 293–301. https://doi.org/10.1016/j.ecolind.2016.07.010.

Béjaoui, B., Salem, H., Basti, L., Feki-Sahnoun, W., Dahmani, S., Melaku Canu, D., Benabdallah, S., Blake, R., Norouzi, H., Solidoro, C., 2022. Hydrology, biogeochemistry and metabolism in a semi-arid Mediterranean coastal wetland ecosystem. Res. Sq. https://doi.10.21203/rs.3.rs-1278908/v1.

Ben Hadid, N. 2021. Machine Learning modeling techniques for forecasting the trophic state in a restored South Mediterranean lagoon using Chlorophyll-a in connection with the physico-chemical variables. Université de Perpignan; Institut national agronomique de Tunisie, 2021. English. NNT: 2021PERP0041. Tel-03588904.

Ben Rejeb-Jenhani, A., 1992. Evaluation des éléments nutritifs dans les eaux du lac Ichkeul durant la période 1985-1988. Bull. Soc. Nat. Tunis. 20 (21), 30–35.

Breiman, L., 2001. Random forests. Random For. 1–122. https://doi.org/10.1201/9780429469275-8.

Brik, B., Shaiek, M., Trabelsi, L., Regaya, K., Ben Mbarek, N., Béjaoui, B., Martins, M.V. A., Zaaboub, N., 2022. Quality status of surface sediments of lake ichkeul (Ne Tunisia): an environmental protected area and world heritage site. Water, Air, Soil Pollut. 233 (7) https://doi.org/10.1007/s11270-022-05648-z.

Caputi, S.S., Careddu, G., Calizza, E., Fiorentino, F., Maccapan, D., Rossi, L., Costantini, M.L., 2020. Changing isotopic food webs of two economically important fish in mediterranean coastal lakes with different trophic status. Appl. Sci. 10 (8), 2756 https://doi.10.3390/app10082756.

Chaouachi, B., Ben Hassine, O.K., Lemoalle, J., 2001. Variations des teneurs en chlorophylle-a et en sels nutritifs dans la lagune d'Ichkeul. Bull. INSTM 28, 105–111.

Culter, D.R., Thomas, C., Edwards, J., Beard, K.H., Culter, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792.

Cutforth, H.W., 2000. Climate change in the semiarid prairie of southwestern Saskatchewan: Temperature, precipitation, wind, and incoming solar energy. Can. J. Soil Sci. *80* (2), 375–385. https://doi.org/10.4141/S99-074.

Dai, A., Zhao, T., Chen, J., 2018. Climate change and drought: a precipitation and evaporation perspective. Curr. Clim. Change Rep. 4 (3), 301–312. https://doi.org/10.1007/s40641-018-0101-6.

Derneği, D. 2010. Mediterranean Basin Biodiversity Hotspot for Submission to the Cepf Donor Council. Middle East, 251.

Derouiche, E. 2016. Analyse de la migration catadrome de l'anguille européenne *Anguilla anguilla* dans les lagunes septentrionales de Tunisie: caractéristiques et état de santé des individus, quantification du phénomène. Thèse de Doctorat. Faculté des Sciences de Tunis. Oceandocs.Org 385.

DGPA. 2020. Annuaire Statistique de la pêche et de l'Aquaculture en Tunisie pour 2020. General Directorate of Fishing and Aquaculture (144p).

Dhib, A., Denis, M., Barani, A., Turki, S., Aleya, L., 2016. Ultra- and microplankton assemblages as indicators of trophic status in a Mediterranean lagoon. Ecol. Indic. 64, 59–71.

Dridi, M.S. 1977. Recherche écologique sur les milieux lagunaires du nord de la Tunisie. Thèse 3ème cycle, Univ. Tunis: 88 p.

Edeline, E., Dufour, S., Elie, P., 2005. Role of glass eel salinity preference in the control of habitat selection and growth plasticity in *Anguilla anguilla*. Mar. Ecol. Prog. Ser. 304 (December), 191–199. https://doi.org/10.3354/meps304191.

Fois, M., Cuena-lombraña, A., Fenu, G., Cogoni, D., 2018. Does a correlation exist between environmental suitability models and plant population parameters? An experimental approach to measure the influence of disturbances and environmental changes. Ecol. Indic. 86 (June 2017), 1–8. https://doi.org/10.1016/j.ecolind.2017.12.009.

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboissiere, R., Maechler, M., Monette, G., Murdoch, G., Nilsson, H., Ogle, D., Ripley, B., Venables, W., Walker, S., Winsemius, D., Zeileis, A., R Core Team, 2022. Package 'Car' Companion to Applied Regression Version 3.1-0. pp 1-54.

Galewski, T., Balkız, Ö., Beltrame, C.M., Chazee, L., 2012. Biodiversity – Status and trends of species in Mediterranean wetlands. (Issue June). https://doi.org/10.13140/RG.2.1.3873.0321.

Greenwell, B., Bochmke, B., Gray, B. 2020. Package « VIP » Variable Importance Plots, Version 0.3.2. pp. 1-24.

Havens, K.E., 1991. Fish-induced sediment resuspension: effects on phytoplankton biomass and community structure in a shallow hypereutrophic lake. J. Plankton Res 13, 1163–1176.

Hollis, G.E., Agnew, C.T., Battarbee, R.W., Chisnall, N., Fisher, R.C., Flower, R., Goldsmith, F.B., Phethmean, S.J., Skinner, J., Stevenson, A.C., Warren, A., Wood, J. B. 1986. The Modeling and Management of the Internationally Important Wetland at Garaet el Ichkeul, Tunisia. Ecol. Conserv. Unit, Univ. Coll. London ministère de l'Agriculture, Tunis. IWRB Spec. Publ., Slimbridge, Glos, U.K. 4, 121 p.

IUCN. 2020. Ichkeul National Park - 2020 Conservation Outlook Assessment. World Heritage Outlook: https://worldheritageoutlook.iucn.org/explore-sites/wdpaid/4322.

Jacoby, D.M.P., Casselman, J.M., Crook, V., DeLucia, M.B., Ahn, H., Kaifu, K., Kurwie, T., Sasal, P., Silfvergrip, A.M.C., Smith, K.G., Uchida, K., Walker, A.M., Gollock, M.J., 2015. Synergistic patterns of threat and the challenges facing global anguillid eel conservation. Glob. Ecol. Conserv. 4, 321–333. https://doi.org/10.1016/j.gecco.2015.07.009.

John, K., Isong, I.A., Kebonye, N.M., Ayito, E.O., Agyeman, P.C., Afu, S.M., 2020. Using Machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. Land 9, 487. https://doi.org/10.3390/land9120487.

Kuhn,M., Wing, J., Weston, S., Williams, A., Keefer,C., Engelhardt,A., Cooper,T., Mayer, Z., Kenkel,B., R. Core Team. 2022. Package 'Caret' Classification and Regression Training, Version 6.0-93. pp 1-224.

Kumar, C., Podesta, G., Kilpatrick, K., Minnett, P., 2021. A machine learning approach to estimating the error in satellite sea surface temperature retrievals. Remote Sens. Environ. 255 (112227), 1–14 https://doi:10.1016/j.rse.2020.112227.

Lagarde, R., Peyre, J., Amilhat, E., Bourrin, F., Prellwitz, F., Simon, G., Faliex, E., 2021. Movements of non-migrant European eels in an urbanized channel linking a Mediterranean lagoon to the sea. Water (Switz.) 13 (6), 1–14. https://doi.org/10.3390/w13060839.

Leoni, B., Patelli, M., Nava, V., Tolotti, M., 2021. Cladocera paleocommunity to disentangle the impact of anthropogenic and climatic stressors on a deep subalpine lake ecosystem (Lake Iseo, Italy). Aquat. Ecol. 55 (2), 607–621. https://doi.org/10.1007/s10452-021-09850-9.

Lorenzen, C.J., 1967. Determination of chlorophyll and pheopigments by spectrophotometric equations. Limnol. Oceano. https://doi.org/10.4319/lo.1967.12.2.0343.

Madyouni, H., Almanza, V., Benabdallah, S., Joaquim-Justo, C., Romdhane, M.S., Habaieb, H., Deliege, J.-F., 2023. Assessment of water quality variations and trophic state of the joumine reservoir (Tunisia) by multivariate analysis. Water 15, 3019. https://doi.org/10.3390/w15173019.

Marcoulides, K.M., Raykov, T., 2019Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods.https://doi.org/10.1177/0013164418817803.

Motarjemi, S.K., Møller, A.B., Plauborg, F., & Iversen, B.V. 2020. Predicting tile drainage discharge using machine learning algorithms. January: 1–22.

Newton, A., Brito, A.C., Icely, J.D., Derolez, V., Clara, I., Angus, S., Schernewski, G., Inácio, M., Lilleb, A.I., Sousa, A.I., Béjaoui, B., Solidoro, C., Tosic, M., Cañedo-Argüelles, M., Yamamuro, M., Reizopoulou, S., Tseng, H.-C., Donata, C., Roselli, L., Brookes, J., 2018. Assessing, quantifying, and valuing the ecosystem services of coastal lagoons. J. Nat. Conserv. 1–2. https://doi.org/10.1016/j.jnc.2018.02.009.

Oba, S., Sato, M.A., Takemasa, I., Monden, M., Matsubara, K.I., Ishii, S., 2003. A Bayesian missing value estimation method for gene expression profile data. Bioinformatics 19 (16), 2088–2096. https://doi.org/10.1093/bioinformatics/btg287.

Oliveira, K., McCleave, J.D., Wippelhauser, G.S., 2001. Regional variation and the effect of lake: River area on sex distribution of American eels. J. Fish. Biol. 58 (4), 943–952. https://doi.org/10.1006/jfbi.2000.1503.

Parsons, T.R., Maita, Y., Lalli, C.M., 1984. A manual of chemical and biological methods for seawater analysis. Pergamon, Oxford-Sized Algae, and Natural Seston Size Fractions. Pergamon Press, Oxford Oxfordshire; New York.

Rehof, L.A., 2021. The Text of the Convention. Guide to the Travaux Préparatoires of the United Nations Convention on the Elimination of All Forms of Discrimination against Women, 16–28. https://doi.org/10.1163/9789004479449_007.

Riley, W.D., Walker, A.M., Bendall, B., Ives, M.J., 2011. Movements of the European eel (*Anguilla anguilla*) in a chalk stream. Ecol. Freshw. Fish. 20 (4), 628–635. https://doi.org/10.1111/j.1600-0633.2011.00513.x.

Rodier, J., Bazin, C., Broutin, J.P., Chambon, P., Champsur, H., Rodi, L., 1996. L'analyse de. l'eau, Eaux naturelles, Eaux Résiduaires, Eau De Mer. Dunod, Paris.

Sahbani, S., Béjaoui, B., Benabdallah, S., Toujani, R., Fathalli, A., Zaaboub, N., Aouissi, J., Kassouk, Z., Hamdi, N., Mbarek, N.Ben, Missaoui, H., Basti, L., Blake, R., Norouzi, H., 2022. Systematic review of a RAMSAR wetland and UNESCO biosphere reserve in a climate change hotspot (Ichkeul Lake, Tunisia). J. Sea Res. *190*, 15. https://doi.org/10.1016/j.seares.2022.102288.

Shaiek, M. 2017. Assemblages ichtyques, structure et fonctionnement du réseau trophique de la lagune d'Ichkeul, Thèse de Doctorat. Institut National Agronomique de Tunis. 301 p.

Stekhoven, D.J., Bühlmann, P., 2012. MissForest — non-parametric missing value imputation for mixed-type data, 28 (1), 112–118. https://doi.org/10.1093/bioinformatics/btr597.

Tosi, L., Spampanato, A., Sola, C., Tongiorgi, P., 1990. Relation of water odor, salinity, and temperature to the ascent of glass eels, Anguilla anguilla: a laboratory study. J. Fish. Biol. 36, 327–340. https://doi.org/10.1111/j.1095-8649.1990.tb05613.x.

Truong, X., N., Jan, D., Colin, J., Zoltan, N., Roland, S., Biswadip, D., Ankush, C., Stefano, D.C., Joel, A., P., Andrea, C., Melanie, Z., Wenceslaw, S., C., Draguna, L., V., 2023. Physics-Informed Machine Learning for Modeling and Control of Dynamical Systems. In: American Control Conference (ACC), 201. Mitsubishi Electric Research Laboratories, Inc, Broadway, Cambridge, Massachusetts, pp. 1–18, 02139.

Umar, A.U., Balakrishnan, U.V., Jha, P.S., 2021. A study of multicollinearity detection and rectification under missing values. Turk. J. Comput. Math. Educ. Vol.12 (2021), 399–418 https://doi/10.17762/turcomat.v12i1S.1880.

Vollenweider, R.A., Giovanardi, F., Montanari, G., Rinaldi, A., 1998. Characterization of the trophic conditions of marine coastal waters with special reference to the NW Adriatic Sea: proposal for a trophic scale, turbidity, and generalized water quality index. Environmetrics 9 (3), 329–357. https://doi.org/10.1002/(sici)1099-095x(199805/06)9:33.3.co;2-0.

Warnes, G.R., Bolker, B., Huber, W., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., 2022. Various R programming tools for plotting data version 3.1.3. Compr. R. Arch. Netw. (CRAN) *1*, 69. ⟨https://github.com/talgalili/gplots⟩.

Zhang, Z., Xin, Q., Li, W., 2021. Machine learning-based modeling of vegetation leaf area index and gross primary productivity across North America and comparison with a process-based model. e2021MS002802 J. Adv. Model. Earth Syst. 13. https://doi.org/10.1029/2021MS002802.

Zhou, J., Li, E., Wei, H., Li, C., Qiao, Q., Armaghani, D.J., 2019. Random forests and cubist algorithms for predicting shear strengths of rockfill materials. Appl. Sci. 9 (1621), 1–16. https://doi.org/10.3390/app9081621.