Machine Learning Techniques for Forecasting the Effect of Climate Change and Anthropogenic Pressures on Coastal Wetlands (Ichkeul Lake, Ramsar site)

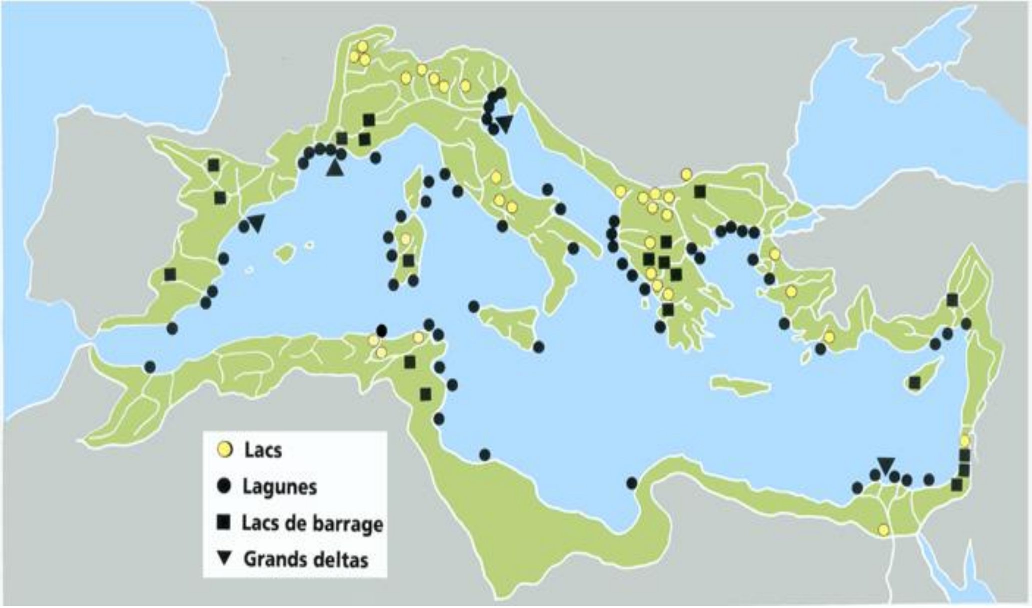
Symposium Trends, Reflections, Evolution, and Visions in Ocean Research
A celebration of the scientific life of Trevor Platt
9-11 August 2023 _ Plymouth Marine Laboratory, UK

Presented by: Sabrine SAHBANI

Supervised by: Béchir BEJAOUI, Ennio OTTAVIANI, Eva RICCOMAGNO, Enrico PRAMPOLINI, Donata MELAKU CANU, Hechmi MISSAOUI, Cosimo SOLIDORO

1. Introduction

- Coastal wetlands have undergone several disruptions, which have been threatening their ecological status , their biodiversity, and their socio-economic services.
- Necessity to assess their sensitivity to anthropogenic pressures and climate change.



Mediterranean Wetlands (Tour du Valat, 2016 - https://tourduvalat.org)

Form a question:

What data should we use?

What changes and pressures could be significant?

Is Climate Change and Anthropogenic Pressures have an effect on Coastal Wetland ?

How we predict this effect?

Are we thinking about the coastal wetland as whole, or should we focus on part of it?

Mrs. Thomas Jackson, TPS: Training 2023-Module 9-Ocean Colour and Climate
https://www.youtube.com/watch?v=c6fBjZ3YeYw

2. Study Site: Ichkeul wetland

Situation

North of Tunisia

Morphological Characteristics

An area of 133 km² with three units: Lake, marshes, and wooded massive (Djebel Ichkeul).

Continental Hydrology

The catchment (2100 km²) drains a developed network of six rivers.

Connection with the Bizerte Lagoon

The Lake communicate with the Bizerte Lagoon.

Ecological value

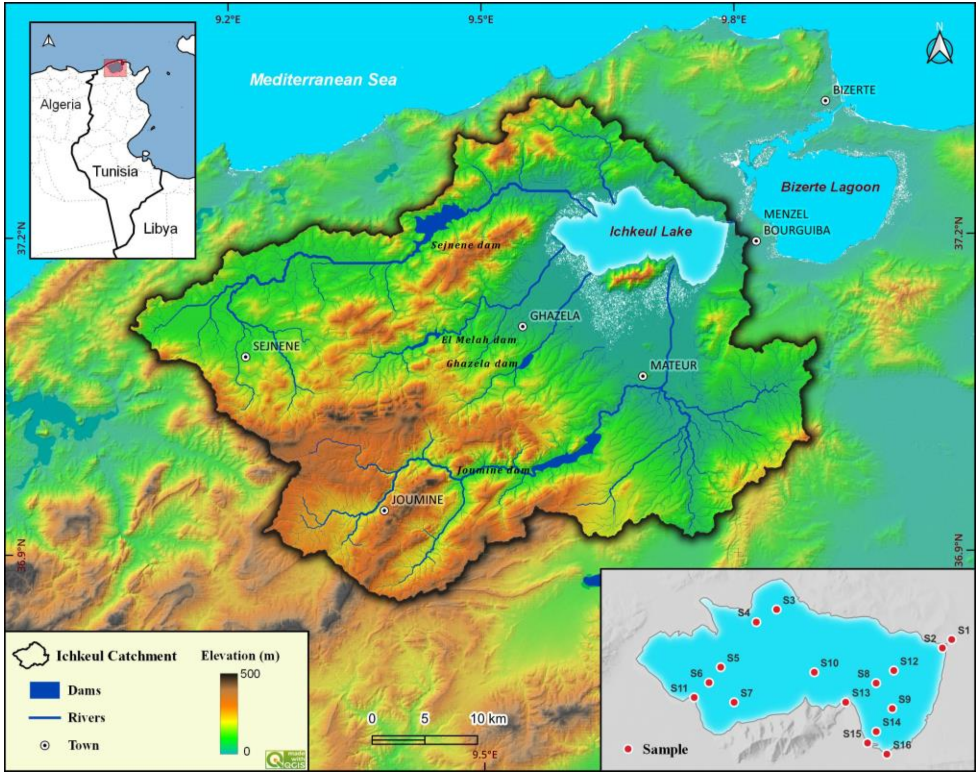
The most productive environment in Tunisia.

Registration in three international conventions.

Ecological fragility

Climate change: increase in T and decrease in P.

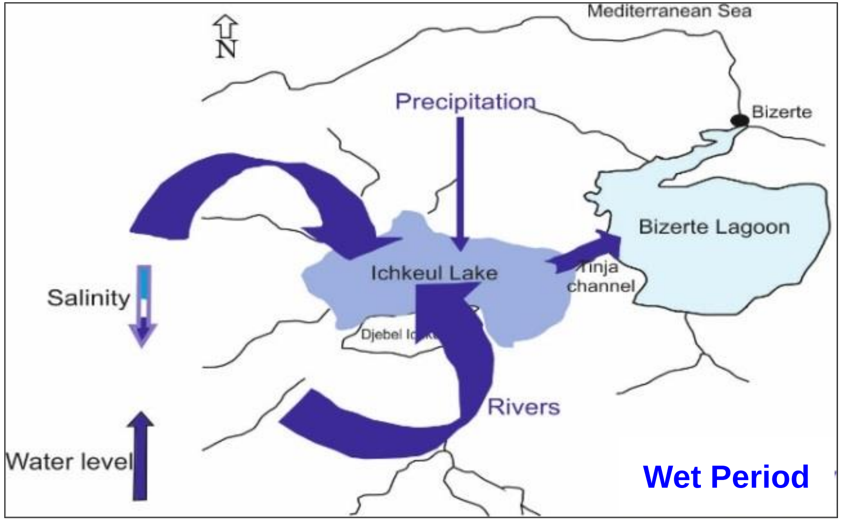
Anthropogenic pressures: construction of dams and locks, over-exploitation of resources, pollution...

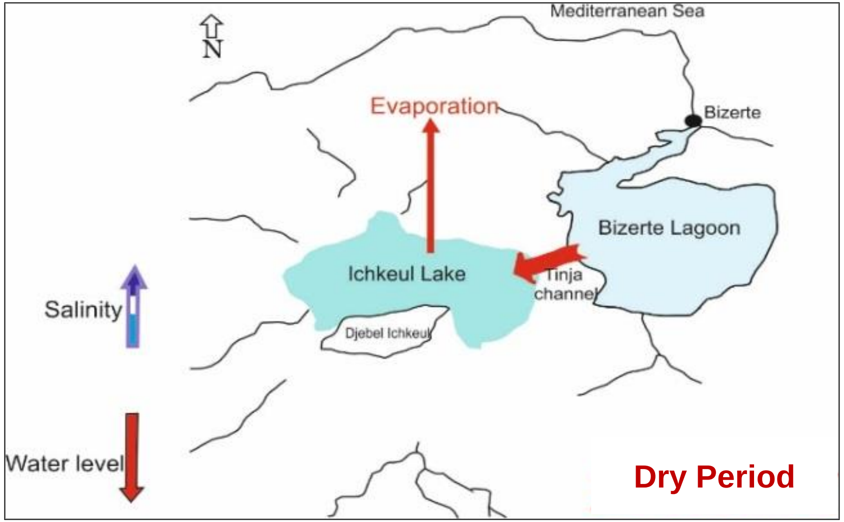


Localisation of Ichkeul Lake and sampling stations

2. Study Site: Ichkeul wetland

- The water budget in the Ichkeul Lake is characterized by seasonal variation:





In wet period:

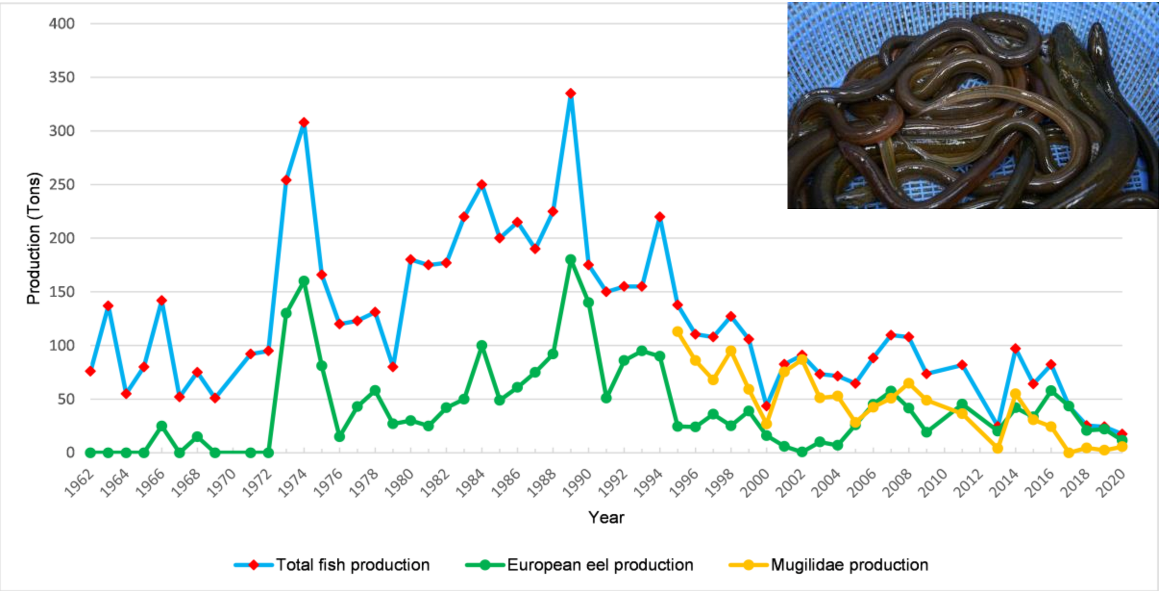
- **Water leaves Ichkeul Lake,**
- **High rainfall,**
- **High water level / low salinity.**

In dry period

- **Water spills off to the Bizerte Lagoon,**
- **High evaporation,**
- **High salinity/low water level.**

2. Study Site: Ichkeul wetland

- The fishery resources are threatened by :
- The collapse of some fish species (*Lithognathus mormyrus*...),
- Decrease in stocks of other species (Mugilidae and *A. anguilla*): Eel production has been reduced from 86 tons for the period 1985-1995 to 32 tons for the period 2010-2020.



Evolution of European eel and Mugilidae production in relation to total ichthyofauna production in Lake Ichkeul between 1962 and 2020.

3. Objectives

- Evaluate the effect of Climate Change and anthropogenic pressures on European eel yield by developing a predictive approach to determine the most important environmental factors influencing the landing of the species.
- Compare the performance of the techniques used.

4. Data Collection

- Dataset = 142 samples of 13 variables (from 2010 to 2020)
- Data sources: From BASSIANA database, and field monitoring.
- Abiotic parameters:
- Meteorological parameters: P, and W.
- Physico-chemical parameters: T, WL, S, DO and Tur.
- Chemical parameters: DIN,DIP, TN.
- Biotic parameter: Chl.a, and Eels landing

Summary of environmental parameters in Ichkeul Lake for the period 2010-2020

PARAMETER	ABREVIATION & UNIT	MEAN	
Period	Pr	DRY (Spring & Summer)	WET (Autumn & Winter)
Precipitation	P (mm)	27.09	78.51
Wind Intensity	W ($m.s^{-1}$)	6.04	5.19
Temperature	T (°C)	20.97	15.45
Water Level	WL (cm)	42.87	66.78
Salinity	S (psu)	40.88	22.61
Dissolved Oxygen	DO ($mg. l^{-1}$)	7.20	7.33
Turbidity	Tur (NTU)	20.60	26.17
Dissolved Inorganic Nitrogen	DIN (µM)	20.93	18.63
Total Nitrogen	TN (µM)	22.41	27.37
Dissolved Inorganic Phosphorus	DIP (µM)	2.046	1.022
Total Phosphorus	TP (µM)	10.27	7.12
Chlorophyll-a	Chl.a ($\mu g. l^{-1}$)	6.48	3.79
Eels Landing	Eels (kg)	2569.06	8758.55

5. Data Analysis

- missForest,
- Pearson Correlation,
- Box-Cox transformation,
- Random Forest and Cubist models.

5. Data Analysis

5.1. missForest :

- To deal with missing values NA (proportion of missingness = 5 %),
- Prediction of the NA from the non-missing value available from the dataset,
- Performance Evaluation: the Normalized Root Mean Squared Error NRMSE:

$$NRMSE = \sqrt{\frac{mean\left(\left(X^{true} - X^{imp}\right)^2\right)}{var\left(X^{true}\right)}}$$

Where: X^{true} is the complete data matrix
X^{imp} is the imputed data matrix.

5.2. Pearson Correlation:

- To avoid the problem of multicollinearity (Potential problem when correlation coefficient > |0.7|)

5.3. Box-Cox transformation:

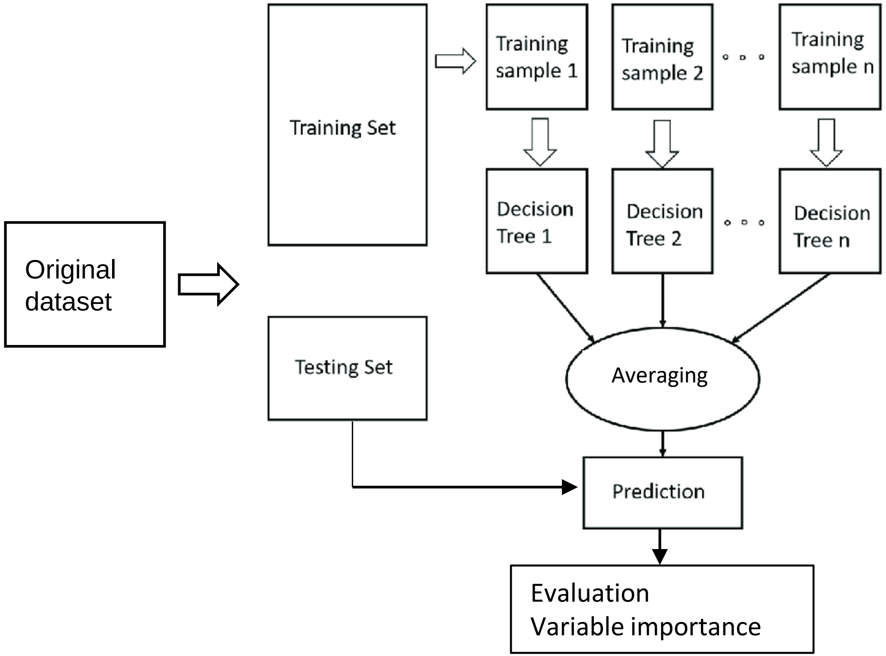
- To conform to the requirement for normality,
- Assessed using Shapiro–Wilk test.

5. Data Analysis

5.4. Machine Learning models

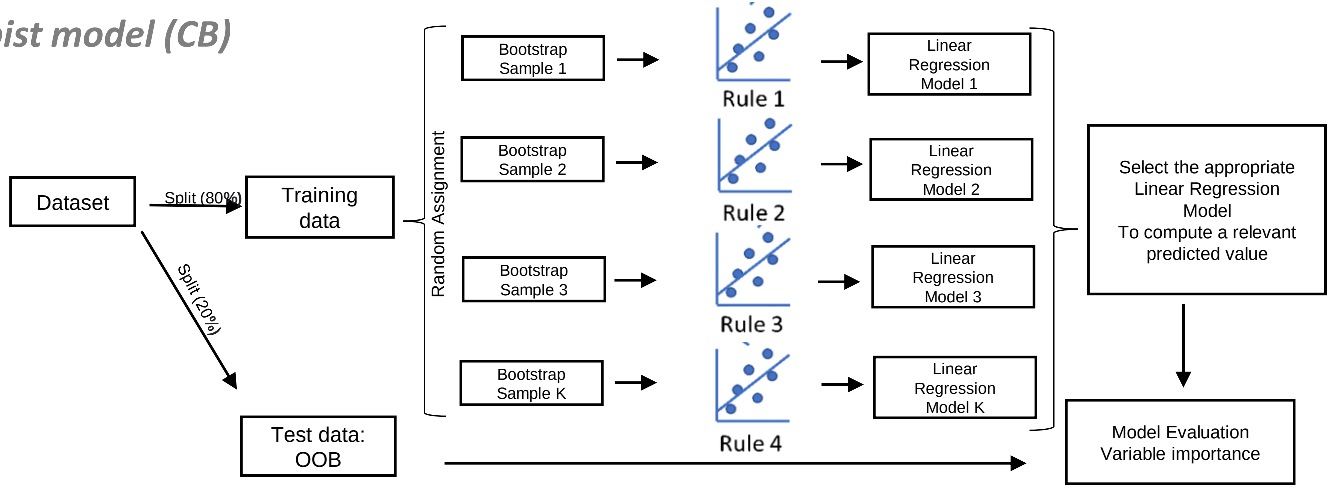
- Evaluate the relationship between the target (Eels landing) and predictors (abiotic parameters).

a. Random Forest model (RF)



5. Data Analysis

b. Cubist model (CB)



c. Models Optimizing

- Hyperparameters selected for tuning
- 3-folds Cross-Validation Method (for CB model) and OOB procedure (for RF model): used to determine the final optimum hyperparameters of the models.

Model	Hyper-parameter	Description
RF	*min.node.size*	Number of trees for the forest
	m_{try}	Number of predictors selected at each split
CB	*neighbors*	Number of instances
	committees	Number of rules set

5. Data Analysis

d. Performance metrics

- Coefficient of Determination $$R^2 = \frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y^{-obs}\right)^2 - \sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{pred}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{obs} - Y^{-obs}\right)^2} \in [0,1]$$
- Mean Absolute Error $$MAE = \frac{\sum_{i=1}^{n}\left|Y_i^{obs} - Y_i^{pred}\right|}{n} \in [0,+\infty]$$
- Root Mean Square Error $$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y^{-obs}\right)^2}{n}} \in [0,+\infty]$$

Where

n: the total number of data,
Yi^{pred} : the predicted eels landing of i observation,
Yi^{obs} : the measured eels landing of i observation,
Y^{-obs} : the mean of all observed responses.

- A good model prediction was expected to have low MAE and RMSE (close to 0) as well as an R^2 value close to 1.

6. Results & Discussions

6.1. MissForest algorithm:

- NRMSE=0.22, indicating a sufficiently good performance.

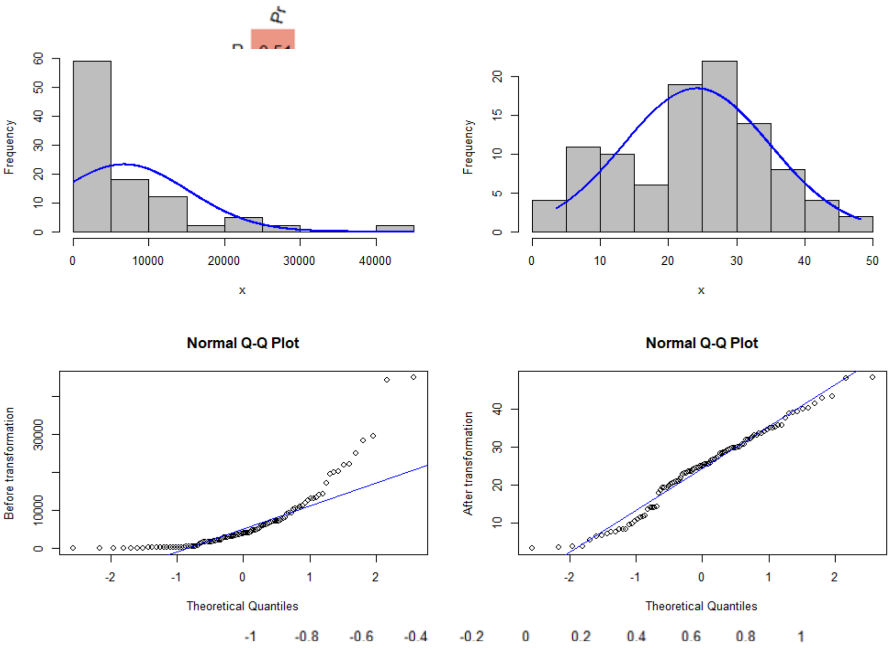
6.2. Pearson Correlation:

- TP and DIP have strong correlation (r = 0.79)

➡ We have chosen to remove TP from the dataset.

- Eel landing was positively correlated with water level and turbidity, but negatively correlated with salinity.

6.3. Box-Cox transformation:

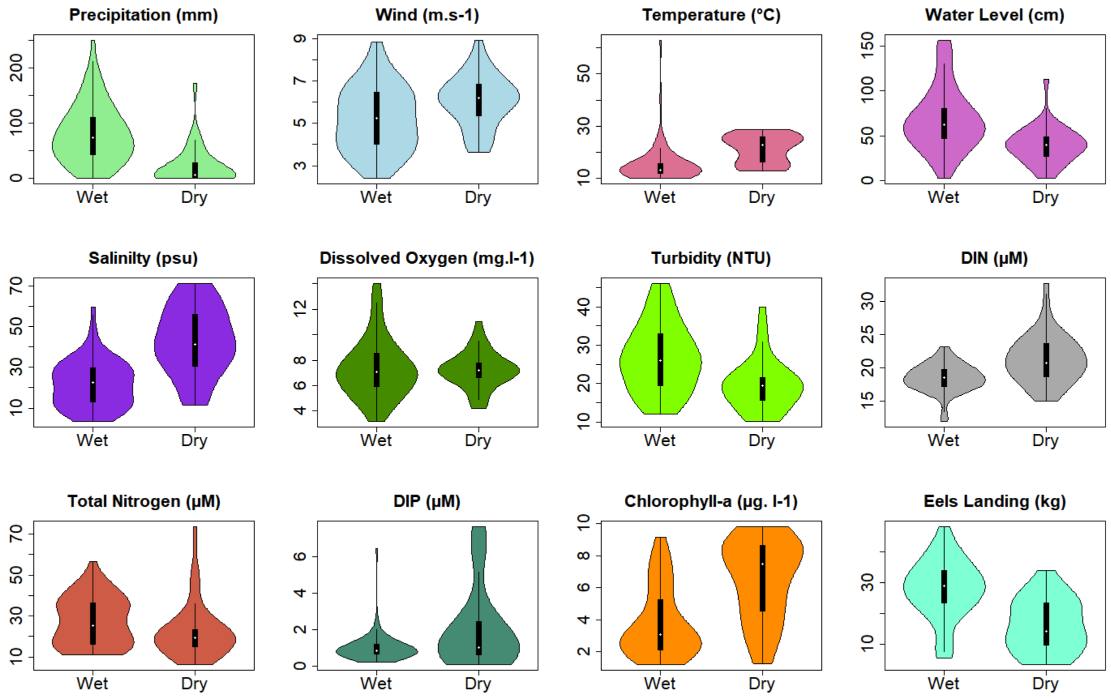


Histogram and Normal Q-Q plots before and after data transformation

6. Results & Discussions

6.4. Parameter's properties

- Seasonal variability of environmental parameters.
- The water turbidity was high throughout the year due to the effect of meteorological and morphological characteristics.
- High levels of TP, DIP,TN and DIN during the period of study and,
- A clear variations between the seasons for the Eels landing.

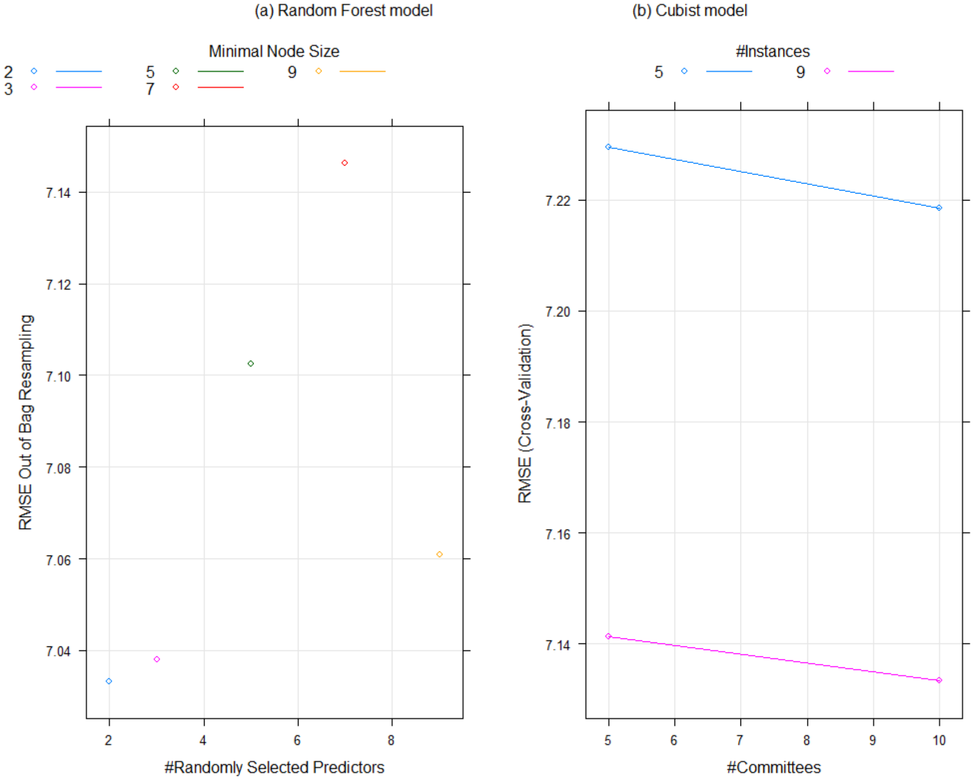


Data visualization with box and violin plots (VIP)

6. Results & Discussions

6.5. Results of hyper-parameters tuning

- For the RF model:
  - The best tuning values of *mtry* equal to 2 and *min.node.size* equal to 2,
- The CB model:
  - The optimal values were *committees* equal to 10 and *instances* equal to 9.



OOB procedure and 3-folds cross-validated RMSE profiles for determining the optimal tuning parameters for (a) RF model and (b) Cubist model

6. Results & Discussions

6.6. Results of performance metrics

- We calculated the metric errors for training and test dataset for each model to see how the models perform out-of-sample rather than in-sample.

For training dataset

Effectiveness metrics	RF	CB	Multiple regression ML (for comparison)
RMSE (kg)	7.20	7.68	**6.24**
R^2	0.56	0.55	**0.64**
MAE (kg)	5.65	6.20	**5.20**

For test dataset

Effectiveness metrics	RF	CB	Multiple regression ML (for comparison)
RMSE (kg)	5.81	**5.13**	7.99
R^2	**0.73**	**0.73**	0.41
MAE (kg)	**4.97**	5.89	6.55

Average Eel landing over the 2010-2020 period is:

Before transformation **6484.8 kg**.

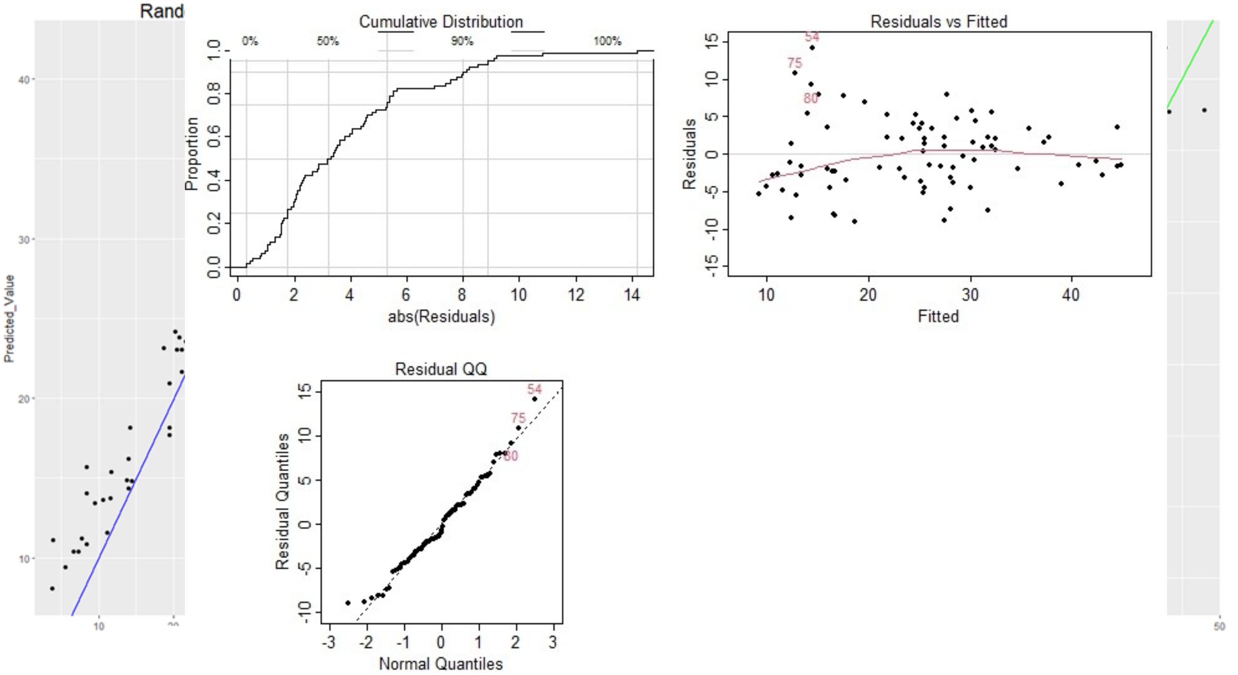
After transformation **24.16 kg** .

- CB and RF are better than MR, Because MR prediction is highly dependent on the training data, and any change in the data potentially affects the model, leading to unstable results.

6. Results & Discussions

6.6. Results of performance metrics

- The points fall quite differently, implying quite large differences between the predicted values.
- For the RF model, there is a systematic overestimation of low values, and underestimation of higher value, despite the data are scaled and centered.
- Which implies that the residuals are not random but correlated to data magnitude.
- CB model is better than RF and MR. This result was also highlighted by Delgado et al., (2019).



Residual vs. Fitted plot for the RF model

6. Results & Discussions

6.6. Results of performance metrics

- The points fall quite differently, implying quite large differences between the predicted values.
- For the RF model, there is a systematic overestimation of low values, and underestimation of higher value, despite the data are scaled and centered.
- Which implies that the residuals are not random but correlated to data magnitude.
- CB model is better than RF and MR. This result was also highlighted by Delgado et al., (2019).

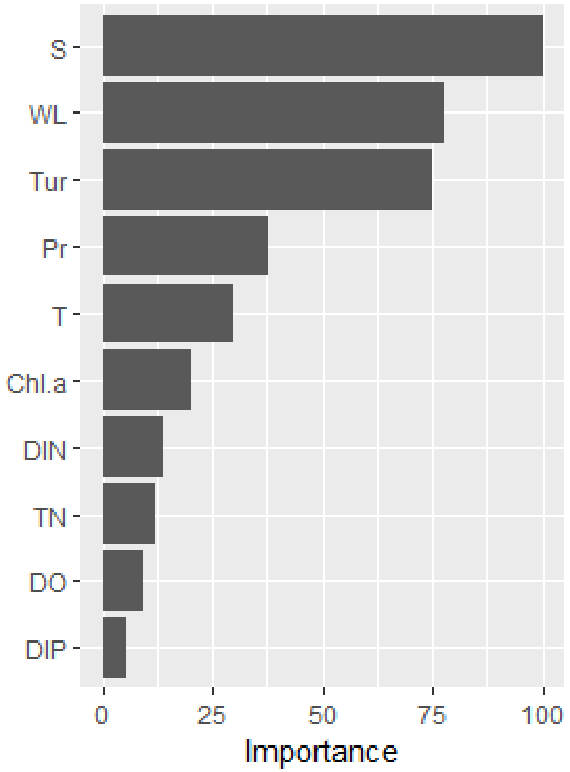
Table: Overview of best performing families of regression algorithms and for each family the best model (Delgado et al., 2019)

Algorithm	Nickname	Category
Cubist Regression	***Cubist***	**Rule-based**
Generalized Boosting Regression	*Gbm*	Boosting
Averaged Neural Network	*avNNet*	Neural Networks
Extremely Randomized Rergression Trees	*extraTrees*	Random Forests
Bayesian Additive Regression Tree	*bartMachine*	Bayesian Models
SupportVector Regression	*svr*	Support Vector Machines

6. Results & Discussions

6.7. Variable importance

- According to the Cubist model, the most important predictors are Salinity followed by Water level and Turbidity.
- The model result between Eels landing and the predictors is consistent with the relationships found with Pearson correlation.
- Water level and turbidity promote eel migration and facilitate the foraging process, while the salinity plays a key factor during the cycle life of European eels (Lagarde et al., 2021).



Variable importance scores for the 13 predictors in Cubist model for Eels Landing

7. Conclusion

- **Based on the numerical approach:**

  ● Climate change and anthropogenic pressures have affected the functioning of the lake.

  ● The construction of dams and locks in the Ichkeul basin has led to a drop in water levels and an increase in salinity. As a direct consequence, the eel population has declined significantly, and the overall abundance of biotic resources has been affected.

- **The results of this study suggest that local management agencies can use these smart technologies in the monitoring system of the trophic and ecological status of the lake, as they offer a reliable and efficient means of maintaining its ecological conditions in the future.**

Acknowledgements

We extend our heartfelt gratitude to Trevor Platt Science Foundation (TPSF) for organizing this important event, and to the Partnership for Observation of the Global Ocean (POGO) for generously fully funding my participation in this Symposium. Its contribution was invaluable in making my involvement in the event possible.

TPS
FOUNDATION

pogo
Partnership for
Observation of the Global Ocean

Thanks for your attention

Sabrine SAHBANI:
National Institute of Marine Science & Technologies.

E-mails:
sahbenisabrine0@gmail.com
sabrine.sahbani@inat.u-carthage.tn

Béchir BEJAOUI:
National Institute of Marine Science & Technologies.

E-mail:
bejaoui.bechir@gmail.com