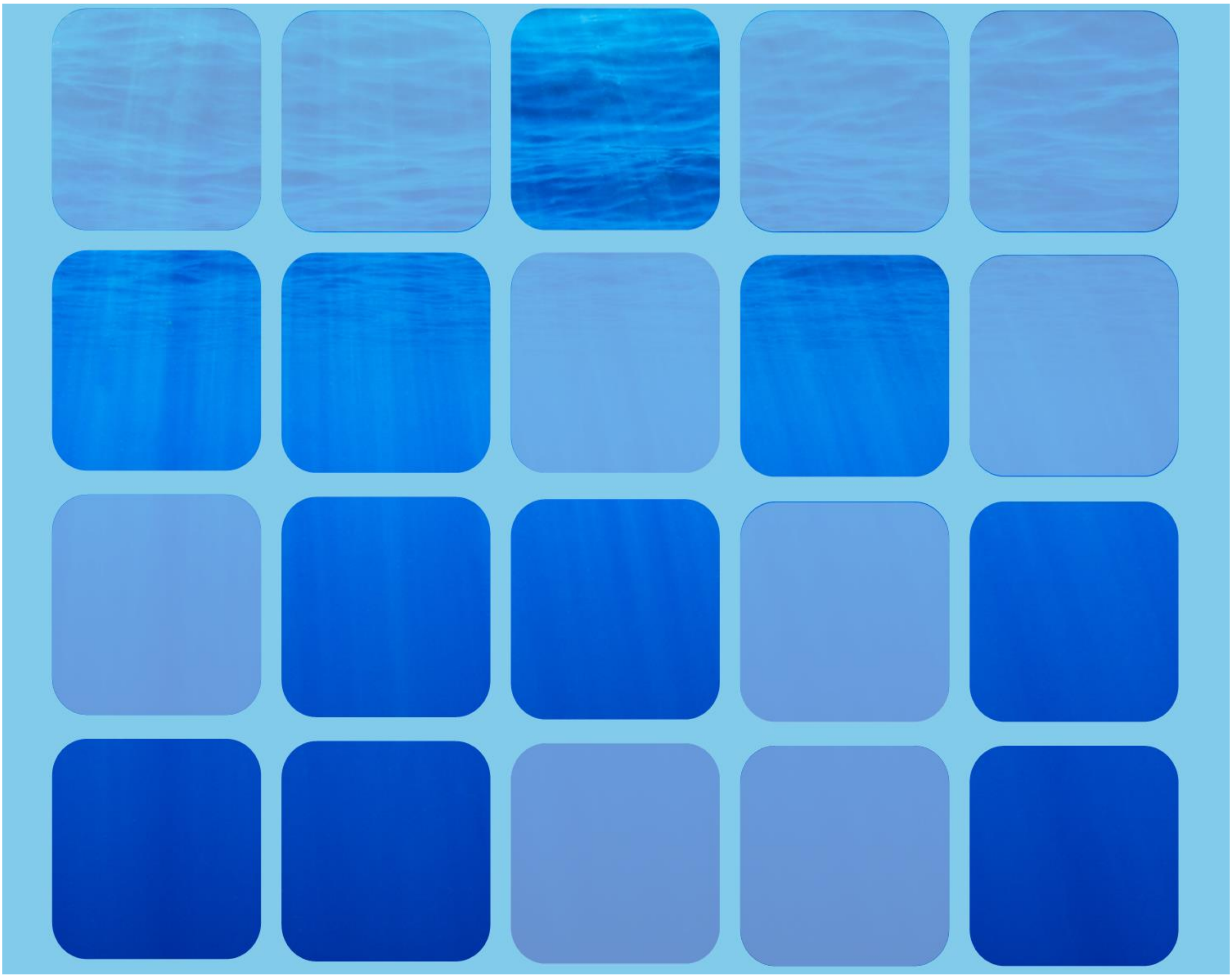
MARITIME

A Machine Learning Cumulative Impact Assessment model



Christian Simeoni, Elisa Furlan, Hung Vuong Pham, Angelica Bianconi

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869710

Suggested citation:

Simeoni, C. , Furlan, E., Pham, H.V., Bianconi, A. (2024). MARITIME: A machine learning cumulative impact assessment model. Guidelines prepared by the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) for MaCoBioS.

Acknowledgements:

The authors are grateful to the MaCoBioS Team for their support in developing MARITIME.

Design:

Bethan O'Leary

MARITIME

MARITIME serves as a decision-support tool, aiding in the strategic prioritisation of areas requiring management to mitigate local stressors. Additionally, it facilitates discussions with stakeholders to pinpoint shared objectives. Utilising a Random Forest-based model, MARITIME integrates essential variables influencing ecosystem presence or absence and forecasts potential changes over time across various climate change scenarios.

The development of MARITIME was led by researchers at the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) through the EU-funded MaCoBioS project (www.macobios.org).

KEY INFORMATION

- This document provides guidance on the datasets you will require, format of input data, assumptions of the model, and instructions for how to run the code.
- MARITIME is a methodological approach that is designed to cope with complex spatial data and inform predictions about likely changes of ecosystem distributions under climate change scenarios. Understanding future changes in an ecosystem will help the development of management strategies whilst meeting stakeholder expectations.
- The output is a series of spatial data and visualisations for the examined ecosystem, climate change scenarios, and targeted region. MARITIME can be applied for different ecosystems and climate change scenarios provided spatial data at the desired study scale and appropriate resolution are available.
- To use MARITIME the following are required:
    - A clear notion of the key variables that determine the ecosystem's distribution;
    - Spatial data at the relevant scale and resolution for key endogenic and exogenic environmental variables and predictions over time under different climate change scenarios;
    - Spatial ecosystem distribution data at the relevant scale for the reference period; and
    - MARITIME model code.
- Be aware that the most difficult aspect of applying MARITIME is obtaining good spatial data at a relevant temporal scale and resolution and its interpretation. As with all models that use future predictions, MARITIME should be used as a decision-support tool recognising the uncertainty that comes with predicting the future.

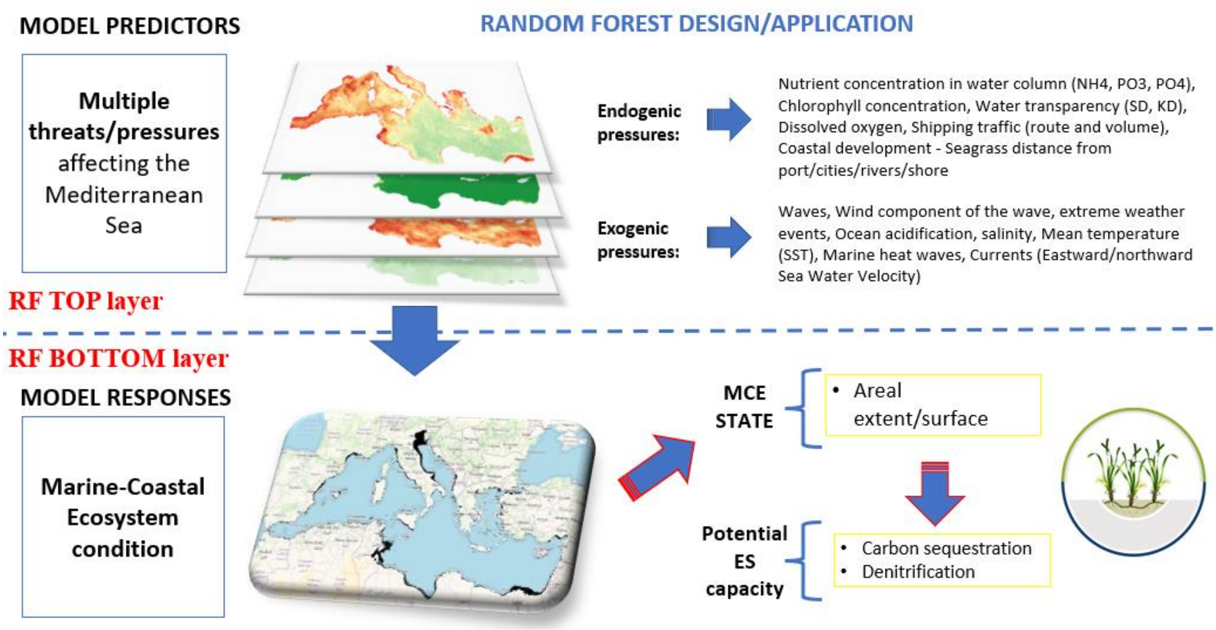
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869710

MODEL ASSUMPTIONS

MARITIME is a Random Forest-based model, thus, it inherits the assumptions underpinning RF model such as independence of trees, random sampling, feature randomness, bootstrap aggregating, and homogeneous trees.

Fundamentally, the MARITIME model is structured into two primary layers as shown below. The top/input layer comprises proxy indicators representing key threats and pressures, such as nutrient input, sea surface temperature, and salinity. The bottom/output layer integrates an indicator specifically focused on seagrass distribution. This bottom layer serves as the model's response, providing an explanation for the overall ecosystem condition. Therefore, the model utilises environmental data as input to predict or elucidate ecosystem distribution, offering insights into the broader state of the ecosystem.

The below figure illustrates the Random Forest (RF) conceptual model for MARITIME based on the Mediterranean Sea, a MaCoBioS study ecoregion: a) TOP layer of the model: multiple threats/pressures affecting focus ecoregions; b) BOTTOM layer of the model: Marine-Coastal Ecosystem (MCE) condition integrating state indicators. ES refers to ecosystem services.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869710

MODEL CODE AVAILABILITY & STRUCTURE

MaCoBioS has made the code underpinning MARITIME together with sample data for *Posidonia oceanica* in the Mediterranean Sea available on their website (www.macobios.org).

Once downloaded, the folder is comprised of 3 subfolders, namely Data_preprocessing, RF_table, and MARITIME_model.

Inside each folder, there are 3 separate subfolders called Input, Output, and Code. The Input folder contains all the raw input data supplying to run the Python code in the Code folder while all the outputs are saved in the Output folder. Importantly, the Python code are written in Jupyter Notebook which are fully annotated and can run in Google Colab. The function of each Jupyter Notebook code is described as the following:

Data_preprocessing

This notebook contains the code to calculate the annual metrics of each indicator. It is organised in several blocks. The first block contains the required libraries for the functioning of the notebook (Library and import required). The second block is dedicated to the working space setting (Setup the working space). The third block contains the descriptions of the functions necessary to calculate the metrics (Functions). The last block is an example of using the code (ML model predictor variables - indicators - metrics). The outputs of this code will serve as the inputs for the RF_table code.

RF_table

This notebook contains the code to compute the input table for the Random Forest. It is organised in several blocks. The first block contains the required libraries for the functioning of the notebook (Library and import required). The second block is dedicated to the working space setting (Setup the working space). The third block contains the description of the functions necessary to compute the table (Functions). The last block contains the application of the code, then the creation of the tables (Create tables for RF). The outputs of this code will serve as the inputs for the MARITIME_model code.

MARITIME_model

This notebook contains the code to implement the MARITIME model. It is organised in several blocks. The first block contains the required libraries for the functioning of the notebook (Library and import required). The second block is dedicated to the working space setting (Setup the working space). The third block contains the description of the functions necessary to implement the RF model (Functions). The other block (Model) is related to the development and implementation of the RF model (i.e., training, validation and testing).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869710

INSTRUCTIONS FOR USE

Follow these instructions to set up and run the model on your system:

- Folder Structure:
- Extract the downloaded ZIP file to a location of your choice.
- Inside the main folder, you will find three subfolders: Code, Input, and Output.
- Run the Code:
- The codes must be run as the following sequence: Data_preprocessing, RF_table, and finally, MARITIME_model.
- Navigate to the "Code" folder in each folder.
- Ensure that you have the required dependencies installed and you have a google account.
- Run the main Jupyter Notebook in each folder or execute the model according to the provided annotated guidelines in the main script.
- Input Data:
- Place your input data, including all considered environmental variables, in the Input folder.
- Ensure that the input data is properly formatted as reported in the section "Format of Input data".
- Output:
- The model results will be generated and saved in the Output folder.
- Ensure that the working directory is changed in the right location.

Input data format

The inputs data of each module is as the follows:

- Data_preprocessing: all the input data must be in netCDF format (e.g., variable.nc). Each variable is placed in a separate folder.
- RF_table: all the input data must be in raster format (e.g., variable.tiff) and in the same folder
- MARITIME_model: table data must be in Excel format (i.e., table.csv).

CONTACT

For further questions please contact CMCC team at christian.simeoni@cmcc.it; elisa.furlan@cmcc.it; vuong.pham@cmcc.it; angelica.bianconi@cmcc.it.

Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) – a leading research center focused on understanding the interaction between climate change and society.

Marine Coastal Ecosystems Biodiversity and Services in a Changing World (MaCoBioS) was a four year research project running between 2020 and 2024 funded by the European Union's Horizon 2020 research and innovation programme. Its objective was to inform efficient and integrated management and conservation strategies for European marine and coastal ecosystems to face climate change by: (a) advancing the scientific evidence base on ecosystem functioning, and (b) developing tools to assess vulnerability and advancing understanding of potential management options.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869710