



## OPEN ACCESS

## EDITED BY

Xuelei Zhang,  
Ministry of Natural Resources, China

## REVIEWED BY

Francesco Tiralongo,  
University of Catania, Italy  
John Iacozza,  
University of Manitoba, Canada

## \*CORRESPONDENCE

Rosalia Maglietta  
✉ [rosalia.maglietta@cnr.it](mailto:rosalia.maglietta@cnr.it)

RECEIVED 09 September 2024

ACCEPTED 29 October 2024

PUBLISHED 19 November 2024

## CITATION

Maglietta R, Caccioppoli R, Piazzolla D, Saccotelli L, Cherubini C, Scagnoli E, Piermattei V, Marcelli M, De Lucia GA, Lecci R, Causio S, Dimauro G, De Franco F, Scuro M and Coppini G (2024) Habitat suitability modeling of loggerhead sea turtles in the Central-Eastern Mediterranean Sea: a machine learning approach using satellite tracking data. *Front. Mar. Sci.* 11:1493598. doi: 10.3389/fmars.2024.1493598

## COPYRIGHT

© 2024 Maglietta, Caccioppoli, Piazzolla, Saccotelli, Cherubini, Scagnoli, Piermattei, Marcelli, De Lucia, Lecci, Causio, Dimauro, De Franco, Scuro and Coppini. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Habitat suitability modeling of loggerhead sea turtles in the Central-Eastern Mediterranean Sea: a machine learning approach using satellite tracking data

Rosalia Maglietta<sup>1,2\*</sup>, Rocco Caccioppoli<sup>2</sup>, Daniele Piazzolla<sup>2</sup>, Leonardo Saccotelli<sup>2</sup>, Carla Cherubini<sup>1,2,3</sup>, Elena Scagnoli<sup>4</sup>, Viviana Piermattei<sup>2</sup>, Marco Marcelli<sup>4,2</sup>, Giuseppe Andrea De Lucia<sup>5</sup>, Rita Lecci<sup>2</sup>, Salvatore Causio<sup>2</sup>, Giovanni Dimauro<sup>3</sup>, Francesco De Franco<sup>6</sup>, Matteo Scuro<sup>2</sup> and Giovanni Coppini<sup>2</sup>

<sup>1</sup>Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, National Research Council, STIIMA-CNR, Bari, Italy, <sup>2</sup>CMCC Foundation - Euro-Mediterranean Center on Climate Change, Lecce, Italy, <sup>3</sup>Department of Informatics, University of Bari 'Aldo Moro', Bari, Italy, <sup>4</sup>Laboratory of Experimental Oceanology and Marine Ecology, Department of Ecological and Biological Sciences (DEB), Università degli Studi della Tuscia, Civitavecchia, Italy, <sup>5</sup>Institute of Anthropic Impact and Sustainability in Marine Environment, National Research Council (IAS-CNR), Oristano, Italy, <sup>6</sup>Consorzio di gestione di Torre Guaceto, Brindisi, Italy

Understanding how sea turtle species move through the environment and respond to environmental features is fundamental for sustainable ecosystem management and effective conservation. This study investigates the habitat suitability of the loggerhead sea turtle (*Caretta caretta*) in the Adriatic and Northern Ionian Seas (Central-Eastern Mediterranean) by developing and validating a multidisciplinary framework that leverages machine learning to investigate movement patterns collected by satellite tags Argos satellite tags. Satellite tracking data, enriched with sixteen environmental variables from the Copernicus Marine Service and EMODnet-bathymetry, were analyzed using Random Forest models, obtaining an accuracy of 80.9% when classifying presence versus pseudo-absence of loggerhead sea turtles. As main findings, sea bottom depth, surface chlorophyll (chl-a), and mixed layer depth (MLD) were identified as the most influential features in the habitat suitability of these specimens. Moreover, statistically significant differences, evaluated using t-test statistics, were found between coastal and pelagic locations, for the different seasons, in mixed layer depth, chl-a, 3D-chlorophyll, salinity and phosphate. Although based on a limited sample of tagged animals, this study demonstrates that the distribution patterns of loggerhead sea turtles in

Mediterranean coastal and pelagic areas are primarily influenced by sea water features linked to productivity and, consequently, to potential prey abundance. Additionally, this multidisciplinary framework presents a replicable approach that can be adapted for various species and regions.

#### KEYWORDS

machine learning, random forest, satellite tag, Argos system, Copernicus marine service (CMS), *Caretta caretta*

Despite the limited number of tagged animals, this study confirms that the distribution pattern of loggerhead sea turtles in the Mediterranean coastal and pelagic areas is influenced mainly by sea water features and productivity, which are related to the potential abundance of prey species. This multidisciplinary framework offers a replicable approach applicable to different species and geographical areas. Furthermore, it will be valuable for analyzing data from a larger number of tagged animals in future studies and establishes a solid foundation for such analyses.

## 1 Introduction

Understanding how endangered species move through the environment and select the habitat in which to live is a pivotal aspect for developing sustainable ecosystem management and effective conservation measures (Mazor et al., 2016; Almpanidou et al., 2021; O'Hara et al., 2019). In particular, sea turtles often travel long distances (even hundreds of kilometers) between nesting and foraging sites, and are thus exposed to several threats that can negatively impact their populations (e.g., shipping, fishing, and marine litter; Luschi et al., 2003; Casale et al., 2018; Ashford et al., 2022; Baruffaldi et al., 2023). Hence, to deepen the current knowledge about sea turtle spatial ecology, it is crucial to investigate the relationship between their movements and physical environment (Lambardi et al., 2008; Luschi, 2013; Chambault et al., 2016).

To date, the behavioral plasticity in the use of different habitats, i.e., coastal within the continental shelf (<200 m; Ravaioli et al., 2003) and pelagic (>200 m), has made it difficult to understand the relationship between the environmental characterization of those areas where sea turtles live. However, several studies have already reported the effects of sea productivity on sea turtles' overwintering (Hochscheid et al., 2007; Hochscheid, 2014), nesting phenology (Mazaris et al., 2004) (Mazaris et al., 2009), reproductive performance (Mazaris et al., 2008), and general ecological niche (Zampollo et al., 2022). Considering such evidence, it becomes important to investigate the role that environmental variables have on the movements of sea turtles, especially in a changing scenario (McMahon and Hays, 2006) involving both climate and other environment challenges (Azzola et al., 2024). To this end, satellite

tracking provides us with detailed information on sea turtles' movement ecology, behavior and habitat use. Nowadays the performance capabilities of satellite tags are well known, and a good number of studies show their applicability to movement study, both in terrestrial (Kays et al., 2015) and aquatic environments, with a great effort in investigating sea turtle ecology (Timko and Kolz, 1982; Arnold and Dewar, 2001; Bentivegna, 2002; Bentivegna et al., 2007; Revelles et al., 2007; Zbinden et al., 2011; Rees et al., 2013; Hussey et al., 2015; Luschi et al., 2017; Abalo-Morla et al., 2018, 2022).

Moreover, combining tracking data with environmental data describing the habitat in which tagged animals move is an advanced technological approach that can address the scientific needs for a deeper understanding of such complex ecological aspects, as discussed in the recent literature (Abalo-Morla et al., 2022; Cavender-Bares et al., 2022). A study using satellite tracking, environmental variables, and machine learning algorithms examined the spatial distribution of two sea turtle species in Chesapeake Bay (Atlantic Ocean) by selecting three key environmental factors (bathymetry, sea surface temperature, and salinity) to create habitat suitability models with boosted regression trees (Di Matteo et al., 2022). Another study (Hazen et al., 2018) applied a multispecies dynamic approach, utilizing daily satellite data to track ocean features and align management scales with species movement and fisheries. This research developed habitat suitability models for three non-target species, including Leatherback turtles, within the California Current domain, using ten environmental variables such as bathymetry, chlorophyll a, and sea surface temperature. To the best of our knowledge, the original contribution presented in this paper is based on four key elements: 1) it offers a detailed description of environmental variables defining loggerhead sea turtle (*Caretta caretta*, Linnaeus 1758) habitats, surpassing prior work in modern literature; 2) it uses a novel machine learning algorithm to integrate tracking data and environmental factors in a statistically robust way; 3) although loggerhead sea turtles have been studied quite extensively in the Mediterranean Sea (Bentivegna, 2002; Mencacci et al., 2006; Bentivegna et al., 2007; Casale et al., 2007, 2008; Lazar and Gračan, 2011; Zbinden et al., 2011; Rees et al., 2013; Zampollo et al., 2022; Abalo-Morla et al., 2022), no solid machine learning-based habitat model has yet been developed for this area; 4) this

study is the first to integrate tracking data, environmental variables, and machine learning applying it to loggerhead sea turtles in the Mediterranean Sea, highlighting the importance of using this technology for habitat modeling in the region. The three technological components have been applied to loggerhead sea turtles in previous cited studies (Hazen et al., 2018; Di Matteo et al., 2022) in a different geographic area and this is the first attempt in the Mediterranean Sea.

Finally, this paper presents a multidisciplinary framework, based on machine learning algorithms, devoted to the characterization of loggerhead sea turtle movement through the analysis of satellite tracking data enriched by environmental variables, extracted by the Copernicus Marine Service (CMS) (<https://marine.copernicus.eu/it>) and EMODnet-bathymetry dataset (<https://www.emodnet-bathymetry.eu/data-products>). Specifically, the proposed framework uses a Random Forest (RF) classifier to compare the conditions of areas in which loggerhead sea turtles are present (identified by the tag data) with the conditions of pseudo-absence areas. RF is one of the most powerful and widely employed machine learning algorithms (Breiman, 2001; Maglietta et al., 2016) (Inglese et al., 2015; Maglietta et al., 2018) to have been successfully applied to several ecological studies (Jeantet et al., 2018; Maglietta et al., 2020, 2022), such as cetacean population modeling (Marini et al.,

2015; Carlucci et al., 2018; Maglietta et al., 2023). In this study, seven loggerhead sea turtles, equipped with marine Argos satellite transmitters, were released in the Adriatic Sea (Central-Eastern Mediterranean Sea), and tracking data were collected from July 2020 to February 2022. According to recent studies, the loggerhead sea turtle is considered the most abundant and widely distributed species of sea turtle in the Mediterranean Sea (Pierantonio et al., 2023). This species is globally considered Vulnerable by the IUCN Red List, whereas the Mediterranean population has been recently assessed as that of Least Concern (Casale and Tucker, 2017). However, this status depends on the conservation strategies applied within this area, thus, the loggerhead sea turtle population needs to be continuously monitored to be preserved. In addition, t-test statistics were used to assess the statistical significance of the environmental variables using presence data.

## 2 Materials and methods

### 2.1 Study area

This study was conducted by examining data acquired, through satellite tags, from seven specimens of loggerhead sea turtle along

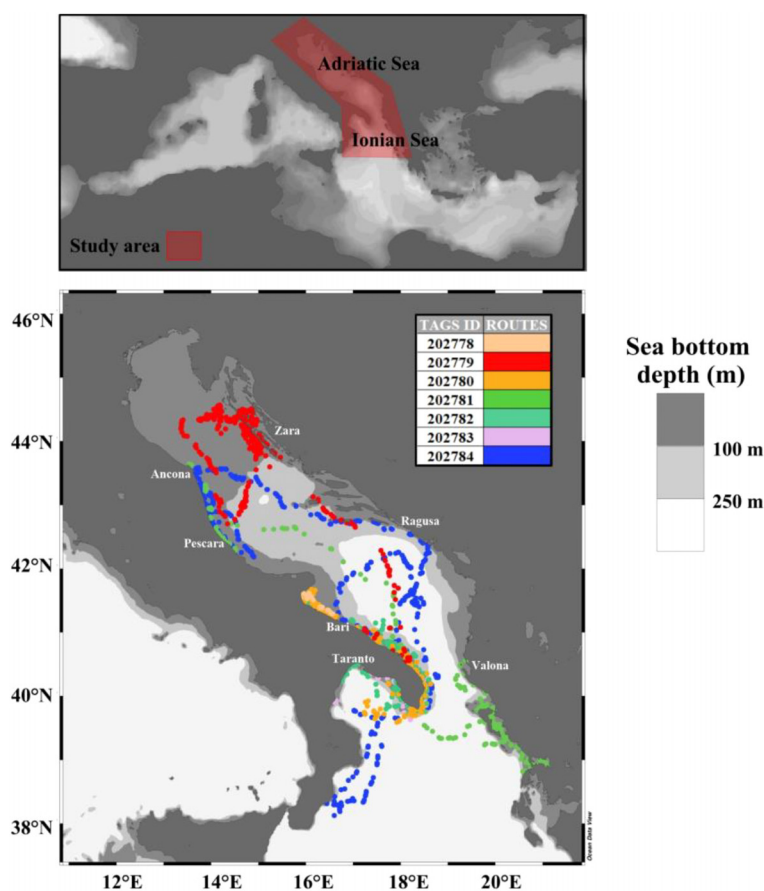


FIGURE 1

Study area and loggerhead sea turtles' routes. The colors reflect each of the seven satellite tags. The maps are displayed in geographical coordinates (Ocean Data View software).

TABLE 1 Table shows tag ids, first transmitted data in water (First Data), last transmitted data (Last Data), number of days elapsed (days), number of filtered records (Valid Rows) for each tag.

Tag ID	First Data	Last Data	Days	Valid Rows
202778	23/07/2020	10/03/2021	230	193
202779	29/07/2020	22/10/2021	450	3081
202780	14/11/2020	15/02/2022	458	993
202781	24/12/2020	27/08/2021	246	1429
202782	04/01/2021	07/09/2021	246	1018
202783	13/03/2021	16/10/2021	217	386
202784	12/03/2021	16/09/2021	188	1751
<b>TOTAL</b>				<b>8851</b>

The total number of valid rows across all tags is indicated in bold at the bottom.

different routes in the waters of the Adriatic and Northern Ionian seas (Figure 1).

Loggerhead sea turtles have a high ecological plasticity, and their life cycles are characterized by the shift in diet and habitat use between oceanic and coastal/neritic waters (Hawkes et al., 2006; McClellan and Read, 2007; Casale et al., 2008; Lazar and Gračan, 2011). Adriatic and Northern Ionian waters are widely considered important habitats for the loggerhead sea turtle. In particular, the Northern and Central Adriatic coastal waters host one of the most important neritic feeding grounds for the loggerhead sea turtle (Margaritoulis et al., 2003; Lazar et al., 2004; Mencacci et al., 2006; Zbinden et al., 2007; Lazar et al., 2010; Lazar and Gračan, 2011), while the Southern Adriatic is an oceanic developmental habitat for this turtle species (Casale et al., 2007). Moreover, fishery by-catch data showed relatively high abundances of this species in the Northern Ionian/Southern Adriatic Sea (Casale et al., 2018; Arcangeli et al., 2019).

## 2.2 Turtles capture, satellite tag description and deployment

The seven loggerhead sea turtle specimens (adults and subadults, Curved Carapace Length - CCL - between 50 and 70 cm) considered in this study come from accidental catches by fishing gear along the coastal areas of the Apulia region (Italy), recovered by the Management Consortium of Torre Guaceto (Carovigno, BR, Italy) (Table 1).

The loggerhead sea turtles were transported to the “Luigi Cantoro” Sea Turtle Rescue Center, where they underwent the necessary treatment and rehabilitation for subsequent reintroduction into the marine environment. None of the individuals suffered any significant injuries that could potentially affect their behavior. The specimens received only the essential care required to facilitate their prompt release back into the sea. Then, before being returned to the sea, satellite tags were applied to their carapace using epoxy adhesive. The tags selected for this study were Argos satellite transmitters, KiwiSat KS202 series (Lotek, 2024), designed with specific characteristics such as compact shape and

abrasion protection, ideal for animals such as the loggerhead sea turtle. The location data (including date, Coordinated Universal Time (UTC), latitude, and longitude) was derived post-dive through the analysis of Doppler shift in the received signals, an integrated process within Argos. (Argos).

## 2.3 Dataset description

In this study two datasets are considered:

- a dataset P collecting only the 8,851 points of presence of the loggerhead sea turtles (see 2.4 Coordinates selection section), each one enriched by the one hundred environmental features (see 2.5 Variables description section);
- a dataset D collecting a total of 17,702 entries, made by the sum of 8,851 presence points (dataset P) and 8,851 pseudo-absence points of the loggerhead sea turtles, obtained as described in the 2.4 Coordinates selection section. Each entry has been enriched by one hundred environmental features, described in the 2.5 Variables description section. A label is assigned to each entry of the dataset D, indicating its class of presence or pseudo-absence.

## 2.4 Coordinates selection

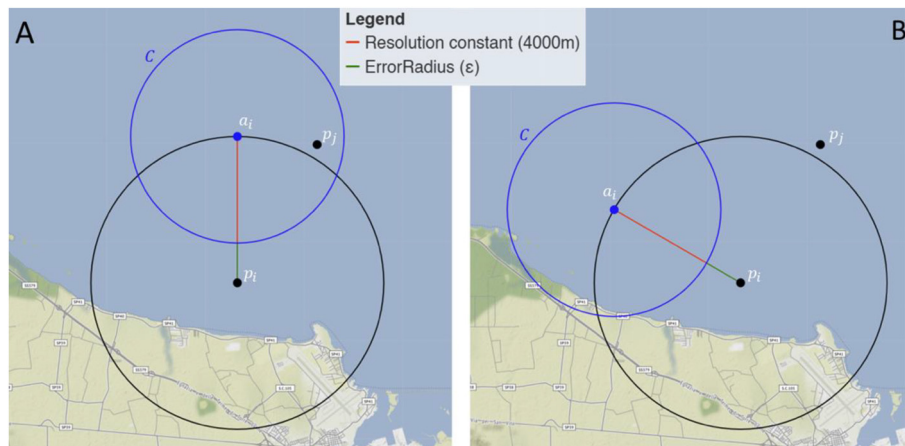
The dataset employed in this study covers the period between July 2020 and February 2022. The locations of the loggerhead sea turtles (geographical coordinates), estimated by the contact of each tag with Argos satellites, were monitored through the Lotek web service. The locations were available in different Argos accuracy classes: Z, B, A, 0, 1, 2, 3 (Douglas et al., 2012). Considering the EMODnet resolution, equal to 4,000 meters, the following analysis was performed using class 0, 1, 2, and 3 Argos data, namely with a precision of the recorded coordinates up to 1,500 meters.

A CSV file containing geographical information was downloaded for each specimen (for a total of seven CSV files, see Table 1). Filtering data, devoted to the selection of valid coordinates, is required (Coyne and Godley, 2005; Douglas et al., 2012) and the main steps were the following:

1. all points on land were discarded.
2. the ranges of the distances travelled by the loggerhead sea turtles between one point and the next, were computed; the ranges of speeds at which the loggerhead sea turtles moved were computed as the ratio between the travelled distance and the time spent, as recorded by tags. If the distance exceeded 100 km, or the speed exceeded 5 km/h (Nagelkerken et al., 2003; Arendt et al., 2012), then the coordinates were discarded.

As shown in Figure 2, the filtering process produced 8,851 valid points, i.e. latitude and longitude coordinates of loggerhead sea turtle presence:  $p_i = (\text{latitude}_i, \text{longitude}_i)$ , with  $i = 1, 8851$ .

Satellite tags only provide presence records, thus, to overcome the lack of absence data, pseudo-absence coordinates (i.e., random locations with no loggerhead sea turtle records) (Barbet-Massin



**FIGURE 2** Schematic representation of the selection strategy of pseudo-absence coordinates. **(A)**  $a_i$  is discarded if a presence point  $p_j$  is within  $C$ ; **(B)**  $a_i$  is saved if no presence point  $p_j$  is within  $C$ .

et al., 2012) were generated following an original algorithm (see Figure 2). From each point  $p_i$ , with  $i = 1, \dots, 8,851$ , a point of pseudo-absence  $a_i$  of latitude and longitude in the water was randomly selected over the circumference centered in  $p$  with radius equal to  $r = 4,000 + \epsilon$ , where 4,000 meters is the Resolution constant required to exceed the satellite’s resolution limit, and  $\epsilon$  (meters) is equal to the ErrorRadius reported by the sensor. This is a measure of positional accuracy and represents the radius around the recorded position within which the loggerhead

sea turtle was finally located (Argos, s.d.). The algorithm then considered the circumference  $C$  centered in  $a_i$  with radius equal to 4,000 meters, and verified the following condition of presence or absence of another point  $p_j$ , with  $j = 1, \dots, 8851$  and  $j \neq i$ :

- if at least one point  $p_j$  was present inside the circumference  $C$  (see Figure 2A), then point  $a_i$  is discarded; for the same point  $p_i$ , a new point  $a_i$  is randomly selected, as previously described, and the condition is verified again.

**TABLE 2** List of the sixteen variables considered in this study, the relative source, the depth levels (meters) at which they were calculated, and the total number of features for each variable (N. Features). The model horizontal grid resolution is  $1/24^\circ$  (ca. 4 km, see Materials and Methods section).

Name	Class	Units	Source	Levels (m)	N. Features
Temperature	PHY	°C	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Salinity	PHY	PSU	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Nitrate	PHY	mmol/m <sup>3</sup>	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
U Currents Component	PHY	m/s	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
V Currents Component	PHY	m/s	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Current Intensity	PHY	m/s	Computed - model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Current Direction	PHY	degree	Computed - model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Density	PHY	kg/m <sup>3</sup>	Computed - model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Primary Production	BIO	mg/m <sup>3</sup> /day	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Phosphate	BIO	mmol/m <sup>3</sup>	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
3D-chlorophyll	BIO	mg/m <sup>3</sup>	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Phytoplankton	BIO	mmol/m <sup>3</sup>	Model	Surface, 5, 10, 15, 20, 30, 60, 100	8
Surface chlorophyll (chl-a)	BIO	mg/m <sup>3</sup>	Satellite	Surface	1
Squared Brunt Väisälä Frequency	PHY	cycle/h	Computed - model	/	1
Mixed Layer Depth (MLD)	PHY	m	Model	/	1
Sea Bottom Depth	AUX	m	EMODnet 2020 bathymetry	/	1
<b>Total</b>					<b>100</b>

- if no point  $p_j$  is present inside the circumference  $C$  (see [Figure 2B](#)), point  $a_i$  is collected and saved, and the following coordinates of presence are considered.

This algorithm produced 8,851 pseudo-absence coordinates of the animals.

## 2.5 Variables description

For each of the 17,702 entries (equal to the sum of 8,851 presence points, and 8,851 pseudo-absence points of the loggerhead sea turtles), a total of sixteen environmental variables, provided by different sources, were obtained: temperature, salinity, nitrate, U currents component, V currents component, current intensity, current direction, density, primary production, phosphate, 3D-chlorophyll, phytoplankton, surface chlorophyll (chl-a), squared Brunt Väisälä frequency, mixed layer depth (MLD), sea bottom depth (see [Table 2](#)).

Physical variables were provided by the Mediterranean Sea Physics analysis ([Clementi et al., 2020](#)), produced by the Euro-Mediterranean Center on Climate Change (CMCC), and delivered by the Copernicus Marine Service. The product was generated by a coupled hydrodynamic-wave model including tides. The hydrodynamic was supplied by the Nucleus for European Modelling of the Ocean (NEMOV3.6), while the wave component was provided by Wave Watch-III; the model solutions were corrected by a 3DVAR assimilation scheme (OceanVar) of temperature and salinity vertical profiles, and along track satellite sea level anomaly observations. The model horizontal grid resolution was  $1/24^\circ$  (ca. 4 km) with 141 unevenly spaced vertical levels. In addition, the water density and the Brunt-Väisälä frequency were computed by means of the simulated temperature and salinity, according to international seawater equations ([Mcdougall et al., 2009](#); [Roquet et al., 2015](#)).

Simulated biogeochemical features were provided by the Mediterranean Sea biogeochemical analysis ([Bolzon et al., 2019](#)), produced by OGS (IT), and delivered by the Copernicus Marine Service. The data at  $1/24^\circ$  of horizontal resolution (ca. 4 km) were produced by means of the MedBFM4 model system, consisting of the coupling of the multi-stream atmosphere radiative model OASIM, the multi-stream in-water radiative and tracer transport model OGSTM\_BIOPTIMOD v4.3, and the biogeochemical flux model BFM v5. Additionally, MedBFM4 features the 3D variational data assimilation scheme 3DVAR-BIO v3.3 with the assimilation of chl-a (CMS-OCTAC NRT product) and of vertical profiles of chl-a, nitrate, and oxygen (BGC-Argo floats provided by CORIOLIS DAC). The biogeochemical MedBFM system was forced by the NEMO-OceanVar model, while the product was run by the CMCC. The ESA-CCI database of chl-a concentration (CMS-OCTAC REP product) was assimilated with a weekly frequency.

The chl-a, observed via satellite, was provided by the multi-year product Mediterranean Sea Ocean Colour Plankton L4 daily gap free, and produced by the Italian National Research Council (CNR, IT) within the Copernicus Marine Service. The chl-a was evaluated via region-specific algorithms (Case 1 waters ([Volpe et al., 2019](#)), with new coefficients; Case 2 waters ([Berthon and Zibordi, 2004](#))),

and the interpolated gap free chl-a concentration (to provide a “cloud free” product) was estimated by means of a modified version of the DINEOF algorithm ([Volpe et al., 2018](#)). The Level-4 product included the daily interpolated chlorophyll field on a 1 km spatial resolution grid, starting from the multi-sensor (SeaWiFS, MODIS, MERIS, VIIRS-SNPP & JPSS1, OLCI-S3A).

In the process of data analysis, a sea-overland extrapolation was used to prevent the presence of missing values by interpolating the oceanic fields over each loggerhead sea turtle location record. This procedure uses a diffusive boundary layer approach that extrapolates the field values on the areas near the coastline, where the Copernicus Marine Service solutions are not defined. The procedure iteratively computes the ocean quantities on the land grid points, so that such quantities can be interpolated on the location records that are very close to the coast.

Among the auxiliary variables, high-resolution bathymetry was derived from the EMODnet-bathymetry dataset 2020-DTM (<https://www.emodnet-bathymetry.eu/data-products>).

Among the sixteen variables, twelve are three-dimensional: temperature; salinity; primary production; nitrate; phosphate; 3D-chlorophyll; phytoplankton; U currents component; V currents component; currents intensity; currents direction, and density. For these variables, the water column was analyzed in correspondence to the positions of the loggerhead sea turtles (geographical coordinates), and eight levels of depth were considered: 0 (surface), 5, 10, 15, 20, 30, 60, and 100 meters. From the sixteen variables, a total number of one hundred environmental features were considered and employed.

## 2.6 Classification model and statistical analysis

Random Forest (RF) ([Breiman, 2001](#)) is an ensemble method that uses multiple de-correlated decision trees, which are merged to perform regression or classification binary tasks: each tree is built using a random subset of features and examples, while results on the test set are obtained by computing the average of the results of each tree. Here RF has been used to classify presence vs pseudo-absence classes, using a Cross Validation (CV) strategy ([Berrar, 2018](#)).

Dataset D was divided into training and validation sets, preserving the balance between presence and pseudo-absence data: 70% of examples, randomly extracted from dataset D, were collected as a training set, and the remaining 30% of examples were collected as the validation set. Subsequently, model training was performed using a stratified K-Fold CV, repeated  $n$  times on the training set ([Pedregosa et al., 2011](#)). Parameter optimization of RF classifiers was automatically performed using the *Bayesian optimization* algorithm ([Snoek et al., 2012](#)). Finally, the classifiers performance was validated on the independent validation set. By utilizing cross-validation, hyperparameter tuning, and an independent validation set, we were able to mitigate the risk of overfitting and enhance the model’s generalization capabilities.

To ensure greater explainability of the model, feature importance analysis was conducted, as it measures how variables influence the model when predicting the response variable ([Saarela and Jauhiainen,](#)

2021). The influence of a given variable increases with the value of this measure. The idea underlying the features importance computed by RF models is that if a variable is influential in prediction, then permuting its values should affect the model error; if a variable is not influential, then permuting its values should have little to no effect on the model error (Maglietta et al., 2023).

Afterwards, the data were sorted by separating pelagic observations from coastal ones: points of geographical coordinates of loggerhead sea turtles locations were identified as pelagic if the animals were found in areas with a bottom depth greater than 200 m; otherwise, these were considered coastal. Finally, to characterize the pelagic and the coastal habitats of the studied loggerhead sea turtles, a two-sample t-test has been used to investigate any potential differences in environmental variables between coastal and pelagic locations for the different seasons. For each of the remaining ninety-nine environmental features (all except sea bottom depth), the null hypothesis was that the coastal and pelagic data come from independent random samples from normal distributions with equal means and equal, yet unknown variances. The significance level was set at 0.05.

## 2.7 Evaluation of experimental performance

Statistical measures for the performance of a binary classifier, used in this work, are accuracy, sensitivity, and specificity. These metrics can be easily derived by the confusion matrix, as indicated below:

- Accuracy is the percentage of predictions that were correct:

$$accuracy = \frac{(TP + TN)}{(TP + FN + FP + TN)}$$

- Sensitivity is the percentage of positive labeled instances that were predicted as positive:

$$sensitivity = \frac{TP}{(TP + FN)}$$

- Specificity is the percentage of negative labeled instances that were predicted as negative:

$$specificity = \frac{TN}{(FP + TN)}$$

where TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives and FN represents the number of false negatives.

## 3 Results

All the experiments were performed using Python 3.9. In particular, the SciPy libraries were employed for statistical analysis (Virtanen et al., 2020), while the Sklearn libraries were used for the application of machine learning algorithms (Pedregosa et al., 2011).

RF classifiers were built using the dataset D (see 2.3 Dataset description section), to classify presence vs pseudo-absence coordinates classes. This dataset collects a total of 17,702 entries (presence + pseudo-absence data), balanced between the two classes.

Model training was performed using a repeated stratified K-Fold of the training set, with K = 5 and n = 5 (see Classification model and statistical analysis section). RF shows good performance in classifying presence vs pseudo-absence classes over the validation set: accuracy of 80.9%, sensitivity of 82.3% and specificity of 79.3%. Moreover, Figure 3 shows the first twenty most influential features

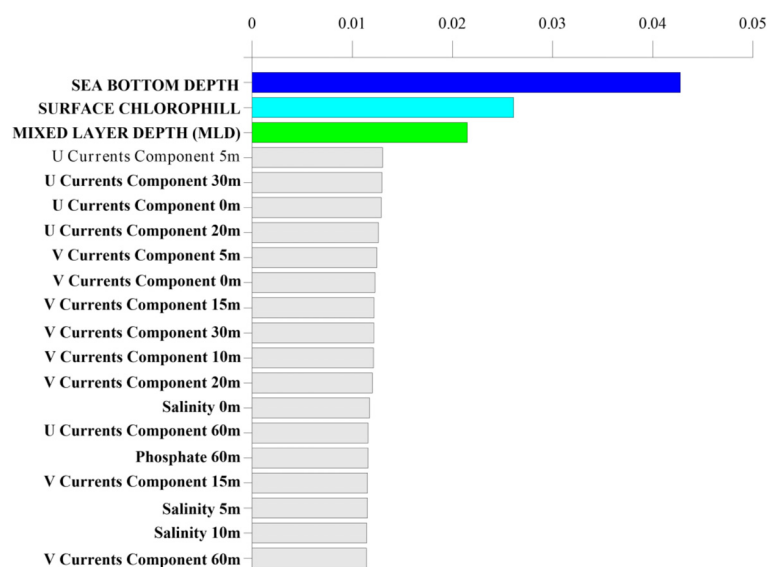
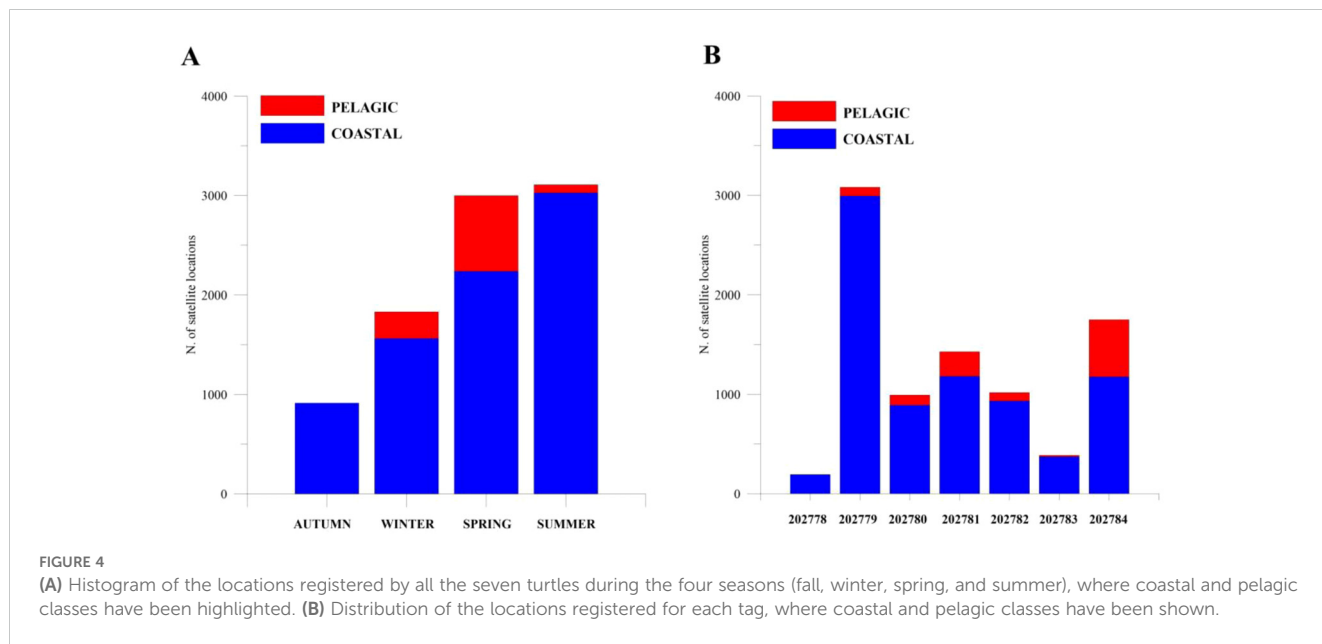


FIGURE 3

The first twenty most influential features, ordered by features importance computed by RF classifier.



in the comparison of presence vs pseudo-absence classes, evaluated using the concept of features importance as computed by the RF classifier. The most influential features in classifying presence vs pseudo-absence of loggerhead sea turtles are sea bottom depth (bathymetry max ranging from 0 meters to 2011 meters); chl-a (values range from 0.03 mg/m<sup>3</sup> to 20.5 mg/m<sup>3</sup>), and MLD (with values between 6.16 meters and 516.14 meters) (Figure 3), which clearly suggest the influence of physical and biological variations of seawater and ocean dynamics on loggerhead sea turtle distribution.

Since the RF model identified bottom depth as the most influential feature in classifying presence vs pseudo-absence of loggerhead sea turtles, a statistical analysis of pelagic and coastal data of loggerhead sea turtles' locations was developed. In particular, the 8851 entries of the dataset P were labeled, based on bottom depth values, of which 7,738 were coastal examples and 1,113 pelagic examples (see *Classification model and statistical analysis* section). Figure 4A illustrates the histogram of 8,851 positions registered by the seven tags during the four seasons

TABLE 3 Results of t-test analysis of the fifteen environmental variables are shown.

Variable	N. Features	Winter	Spring	Summer
Temperature	8	100%	88%	100%
Salinity	8	100%	100%	100%
Nitrate	8	100%	100%	88%
U Currents Component	8	100%	75%	0%
V Currents Component	8	88%	75%	75%
Currents Intensity	8	100%	88%	100%
Currents Direction	8	0%	50%	100%
Density	8	88%	88%	100%
Primary Production	8	100%	88%	100%
Phosphate	8	100%	100%	100%
3D-chlorophyll	8	100%	100%	100%
Phytoplankton	8	100%	88%	100%
Surface chlorophyll (chl-a)	1	100%	100%	100%
Squared Brunt Väisälä Frequency	1	100%	0%	100%
Mixed Layer Depth (MLD)	1	100%	100%	100%

N. Features is the total number of features for each variable (see Table 2). Winter, Spring and Summer columns show the percentage of features, for each variable, having a p-value higher than 0.05 in the season.

The values represent the percentage of features and are color-coded as follows: green represents 100% of features with a p-value > 0.05, orange and light red indicate that 50% to 90% of features have a p-value > 0.05, and dark red indicates 0%, meaning no features have a p-value > 0.05.



(fall, winter, spring, and summer), highlighting coastal and pelagic classes. The 87.4% of available data comes from the coastal area (fall = 913 locations, winter = 1563 locations, spring = 2236 locations, summer = 3026 locations) and the remaining 12.6% referred to pelagic area (winter = 269 locations, spring = 762 locations, summer = 82 locations). In fall, there are no records identified as pelagic. The distribution of the location registered by each tag, highlighting pelagic and coastal classes, is illustrated in [Figure 4B](#). Again, for all tags, the coastal class contains more examples than the pelagic class.

T-test results for the coastal and pelagic classes are shown in

[Table 3](#). Here, we cannot analyze data acquired during fall, because, as illustrated earlier, no locations from pelagic habitat were collected by the tags. Each row in the table indicates a variable and the corresponding number of features available. Moreover, for each variable, the percentage of features that has passed the statistical test (i.e.,  $p$ -value  $\leq 0.05$ ), thus rejecting the null hypothesis, is shown in winter, spring, and summer. In the case of three-dimensional variables, which have multiple features (e.g., temperature with eight features, each at different depths), a percentage of 100% indicates that the means of all eight features show significant differences, evaluated by eight statistical tests, between coastal and pelagic classes. Lower percentages, such as 88%, indicate that only some of the features of a variable (e.g., seven out of eight) result to be significantly different. In case of one-dimensional variables, having only one feature (see MLD in [Table 3](#)), there are only two possible states: 100%, indicating that the feature shows a significant difference between coastal and pelagic locations, or 0%, indicating that the feature does not show a significant difference (i.e., the null hypothesis is not rejected). Considering winter, twelve variables out of a total of fifteen showed a statistically significant difference between coastal and pelagic locations; during spring, six variables showed a statistically significant difference between coastal and pelagic locations; in summer, twelve variables showed a statistically significant difference between coastal and pelagic locations (see [Table 3](#)). To sum up, five variables (salinity, phosphate, 3D-chlorophyll, chl-a, and MLD) were found to be statistically significant, consistently showing percentages of 100% across all three seasons (highlighted in bold in [Table 3](#)). Notably, these variables exhibited distinct average values between coastal and pelagic zones at all depths.

## 4 Discussion

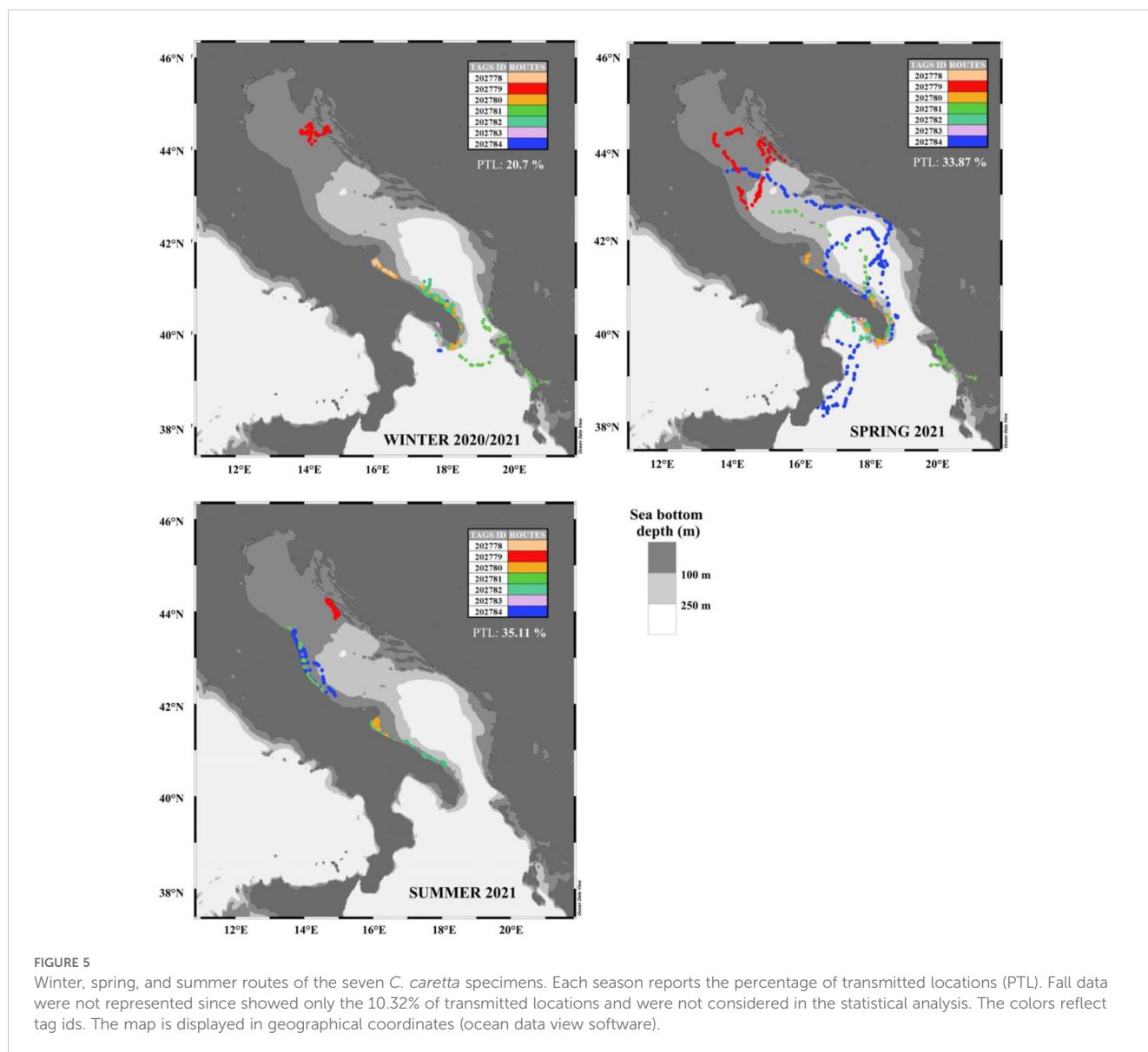
The outcome of the RF analysis revealed that sea bottom depth, chl-a, and MLD were the most important features in classifying presence vs pseudo-absence of loggerhead sea turtles. This result aligns with existing literature, underscoring its consistency with previous research findings. In fact, in ([Hazen et al., 2018](#)) bottom depth was one of the most important predictors in modeling habitat of several species over the Californian Current domain, among them Leatherback turtles. Sea turtles have complex life-history patterns, and they utilize a wide range of areas throughout their life, from pelagic zones to extremely coastal areas ([Haywood et al., 2020](#)), depending on several ecological factors ([Arcangeli et al., 2019](#)). Sea turtles, as previously noted, often make long journeys to

reach sites with optimal ecological conditions for foraging. In this regard, it is known that there is a positive relationship between high chl-a water masses and loggerhead sea turtle routes; frequently, these animals move from areas with low chl-a concentrations to higher ones for foraging ([McCarthy et al., 2010](#)), as their prey may be more abundant in these latter regions ([McCarthy et al., 2010](#)). The ocean mixed layer is a quasi-homogeneous area of the upper ocean, characterized by small temperature and density variations with depth ([Kara et al., 2000](#)), and the importance of the MLD in classifying the presence vs pseudo-absence of loggerhead sea turtles is strictly connected to the chl-a concentration. MLD influences the phytoplankton dynamics through light and nutrient control, thus affecting biological productivity in the ocean ([Sverdrup, 1953](#); [Yentsch, 1990](#); [Jang et al., 2011](#)).

Once the results of the machine-learning model were obtained and examined, we sorted our data by selecting the 200 m bathymetry (continental shelf boundary). Immediately, it became clear that a greater number of satellite locations referred to coastal areas ([Figures 4A, B](#)) and this evidence is consistent with what is reported in the literature, concerning the habit of adult and subadult loggerhead sea turtles. In fact, early-stage loggerheads sea turtles are opportunistic feeders, consuming soft prey living in the water column of open oceans ([Parker et al., 2005](#); [Marshall et al., 2012](#)); this pelagic life stage is then followed by an ontogenetic shift to coastal habitats, where loggerhead sea turtles continue to grow and sexually mature ([Marshall et al., 2012](#)). The coast provides important foraging areas ([Foley et al., 2014](#)), with an abundance and diversity of prey, suitable for loggerhead sea turtles. Moreover, coastal areas are also essential for reproduction and nesting activities ([Casale et al., 2018](#)). To this end, loggerhead sea turtles equipped with tags no. 202780, 202781, 202783, and 202784 traveled the migratory corridor of the Otranto channel ([Casale et al., 2018](#)) in winter and spring.

As for the movements of loggerhead sea turtle specimens in the pelagic environment, these were recorded most in spring ([Figure 4](#) and [Figure 5](#)), and least in winter and summer. Although our data showed significantly fewer loggerhead sea turtle satellite locations in the pelagic environment, it is well known that adult and subadult loggerhead sea turtles move frequently through migratory corridors, located both in the open ocean and along the coasts, during reproductive migrations from and to breeding sites ([Casale et al., 2018](#)), as well for post-breeding migrations from the breeding areas to foraging grounds ([Schofield et al., 2013a](#); [Dujon et al., 2014](#); [Luschi and Casale, 2014](#); [Patel et al., 2015](#); [Mingozzi et al., 2016](#); [Snape et al., 2016](#); [Casale et al., 2018](#)). Loggerhead sea turtle movements may be correlated also to post-nesting ([Zbinden et al., 2011](#); [Schofield et al., 2013b](#)) and seasonal migrations ([Bentivegna, 2002](#); [Zbinden et al., 2011](#); [Casale et al., 2012](#); [Luschi and Casale, 2014](#); [Hochscheid et al., 2005](#)).

Our results highlighted that the monitored loggerhead sea turtle specimens inhabited a coastal rather than a pelagic habitat. However, we should consider that the imbalance between coastal and pelagic classes could partially depend also on the scarcity of data for each tag (for example, ID 202778 has fewer entries compared to the others), or on the scarcity of data in different seasons, as shown in [Figure 5](#). Specifically, in fall, only 10.32% of the



locations were registered, against the 35.11% acquired during summer, 33.87% in spring and 20.7% during winter. This different distribution of data among seasons also depends on the different release data and monitoring period of the loggerhead sea turtles.

Many of the environmental variables considered in this study showed statistically significant differences between coastal and pelagic satellite locations during the considered seasons (winter, spring, and summer) (see Table 3). Among them, it is particularly interesting that chl-a and MLD showed the above significant differences in all three seasons, and that these two variables were among the three most important selected by the RF results, thus reflecting the usefulness of predictive algorithms, as well as descriptive statistics tests, for the study of habitat characterization and selection of marine species.

Above all, our results found that the areas closest to the coast differed from the offshore ones in environmental variables that clearly indicate nutrient-rich (phosphate) and highly productive

waters (chl-a, MLD, 3D-chlorophyll), thus potentially associated with an abundance of prey for sea turtles. It therefore cannot be excluded that these variables may have substantially contributed to outlining the distribution patterns of the seven specimens of loggerhead sea turtle (Figure 5).

Unlike previous studies, this is the first to apply machine learning, integrated with tracking data and environmental variables, in the Mediterranean Sea for habitat modeling of loggerhead sea turtles. While the results align with findings from previous studies, this work consolidates existing knowledge by providing a statistically robust and comprehensive model, demonstrating the reliability of machine learning for habitat modeling. Importantly, the proposed methodology could be easily applicable to the study of large datasets and of other marine species in the Mediterranean Sea and beyond. Furthermore, it could integrate additional information and data based on study needs. In our specific case, it should still be considered that the main limitation was the small sample size of tagged animals. Despite this,

the study highlighted the usefulness of the proposed methodology, offering a strong foundation for future research focused on advancing conservation efforts on endangered species in the marine environment.

## 5 Conclusions

The loggerhead sea turtle is a species threatened on a global scale that typically travels thousands of kilometers. To improve conservation policies and plans for this species, it is essential to have a thorough knowledge of the habitats it visits during its life cycle. The results of this study underlined that the environmental variables that may have greater weight in the distribution patterns of this species are proxies for nutrient-rich and highly productive waters, therefore potentially associated with the presence of prey species (e.g., chl-a and MLD). These variables have seasonal variations, are profoundly influenced by regional and local drivers, and are likely to be altered by climate change. Considering the movements of the seven loggerhead sea turtle specimens in the Adriatic and Ionian Seas, the results obtained suggest that productive and nutrient-rich seawater, potentially associated with the presence of multiple prey species, appeared to have an important role in the distribution of these animals, highlighting consistency with the existing literature. However, migratory animals inhabiting vast areas could be influenced also by aspects on a strictly local or seasonal scale (e.g., local and seasonal coastal anthropic impact, presence of specific conservation measurements such as Marine Protected Areas); this aspect needs to be confirmed with further analyses, useful for studying both the behavior of the individual animal and the influence of the anthropic impact. To this end, the continuous monitoring of many specimens using cutting-edge and minimally invasive technologies such as satellite tags could be highly useful. The use of the RF algorithm, together with the support of statistical techniques for the analysis of the remote sensor data, numerical models and satellite tag data, resulted in a functioning, replicable and safe framework that could be applied on a larger spatial scale, also considering different geographical areas in further studies. Moreover, in future studies, this framework could be adapted and applied to the monitoring of different species.

This stage of our study is currently limited by the low number of tagged animals (seven specimens, as indicated earlier). Evidently, a larger quantity of tags could prove strategic in providing further information on loggerhead sea turtle behavior. In fact, collecting more tracking data on loggerhead sea turtle movements could fill the gap caused by the data missing from the fall, also providing new and valuable information on pelagic areas. Furthermore, being able to track a greater number of animals could reveal that some of them explore larger areas, thus expanding the area of study investigated in this paper.

Another important challenge for future research is to investigate the impact of environmental variable autocorrelation on machine learning models. In this study, fully aware of the

significance of this issue, we implemented a robust Random Forest-based methodology incorporating k-fold cross-validation and design-based validation, as suggested by the relevant literature (Ambroise and McLachlan, 2002; Mushagalusa et al., 2024; Moudrý et al., 2024). That said, it is clear that with a larger number of tags, a key future challenge will be to address the design, construction, and validation of a statistically robust methodology that thoroughly assesses the effect of environmental variable autocorrelation on model performance.

Future studies will focus on an extensive analysis of the diving patterns of loggerhead sea turtle specimens recorded by satellite tags, which could, in turn, provide new and important information on the behavior of this species. In fact, the combination of surface information and dive patterns, obtained through an extended application of the methodology proposed in this study, could become an essential tool for the validation of numerical models, as well as offer important insights into the behavior of loggerhead sea turtles in relation to the effects of climate change.

## Data availability statement

Data and software have been deposited in Github and are freely available (see <https://github.com/CMCC-Foundation/SeaTurtleModelingFramework>).

## Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because all of the procedures were carried out by the specialized staff of the Marine Turtle Recovery Centre of Torre Guaceto.

## Author contributions

RM: Supervision, Methodology, Conceptualization, Writing – review & editing, Writing – original draft, Visualization, Validation. RC: Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal Analysis, Data curation. DP: Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal Analysis, Data curation. LS: Writing – review & editing, Investigation. CC: Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal Analysis, Data curation. ES: Writing – original draft, Visualization, Resources, Data curation. VP: Writing – review & editing, Visualization, Resources, Data curation. MM: Writing – review & editing, Visualization, Resources, Data curation. GL: Writing – review & editing, Writing – original draft, Investigation. RL: Writing – review & editing, Resources, Investigation, Data curation. SC: Writing – review & editing, Software. GD:

Methodology, Writing – review & editing. FF: Writing – review & editing, Resources, Investigation, Data curation. MS: Writing – review & editing, Software. GC: Writing – review & editing, Project administration, Conceptualization.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was carried out as part of three projects: MYSEA Torre Guaceto – “Osservatorio sui mari di Puglia per la tutela della Caretta caretta”, convenzione esecutiva n.1, programma di finanziamento P.O.R. PUGLIA FESR/FSE, CUP C54I20000120002; MONITORAGGIO PER LA VALUTAZIONE DEGLI AREALI DI DISTRIBUZIONE DELLE TARTARUGHE MARINE, contratto d'appalto Prot. 0001469-BA-23 DEL 26/04/2023, Service contract, CUP H83I2200027006; PNRR BIODIVERSITÀ-Centro Nazionale Ricerca per la Bioversità - National Biodiversity Future Center NBFC – N. progetto CN00000033, NBFC. CUP C83C22000550007.

## References

- Abalo-Morla, S., Belda, E. J., March, D., Revuelta, O., Cardona, L., Giralt, S., et al. (2022). Assessing the use of marine protected areas by loggerhead sea turtles (*Caretta caretta*) tracked from the western Mediterranean. *GECCO* 38, e02196. doi: 10.1016/j.gecco.2022.e02196
- Abalo-Morla, S., Marco, A., Tomás, J., Revuelta, O., Abella Pérez, E., Marco-Cabedo, V., et al. (2018). Survival and dispersal routes of head-started loggerhead sea turtle (*Caretta caretta*) post-hatchlings in the Mediterranean Sea. *Mar. Biol.* 165, 1–17. doi: 10.1007/s00227-018-3306-2
- Almpanidou, V., Tsapalou, V., Chatzimitor, A., Cardona, L., Claro, F., Hostetter, P., et al. (2021). Foraging grounds of adult loggerhead sea turtles across the Mediterranean Sea: key sites and hotspots of risk. *Biodivers. Conserv.* 31, 143–160. doi: 10.1007/s10531-021-02326-0
- Ambrose, C., and McLachlan, G. J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *PNAS* 99, 6562–6566. doi: 10.1073/pnas.102102699
- Arcangeli, A., Maffucci, F., Atzori, F., Azzolin, M., Campana, I., Carosso, L., et al. (2019). Turtles on the trash track: loggerhead turtles exposed to floating plastic in the Mediterranean Sea. *Endanger. Species Res.* 40, 107–121. doi: 10.3354/esr00980
- Arendt, M. D., Segars, A. L., Byrd, J. I., Boynton, J., Schwenter, J. A., Whitaker, J. D., et al. (2012). Migration, distribution, and diving behavior of adult male loggerhead sea turtles (*Caretta caretta*) following dispersal from a major breeding aggregation in the Western North Atlantic. *Mar. Biol.* 159, 113–125. doi: 10.1007/s00227-011-1826-0
- Argos *Argos user's manual*. Accessed April 7, 2024. Available online at: <https://www.argos-system.org/manual/>.
- Arnold, G., and Dewar, H. (2001). “Electronic tags in marine fisheries research: A 30-year perspective,” in *Electronic tagging and tracking in marine fisheries: proceedings of the symposium on tagging and tracking marine fish with electronic devices*. Eds. J. R. Sibert and J. L. Nielsen (East-West Center, University of Hawaii, Dordrecht: Springer Netherlands), 7–64. doi: 10.1007/978-94-017-1402-0\_2
- Ashford, M., Watling, J. I., and Hart, K. (2022). One shell of a problem: cumulative threat analysis of male sea turtles indicates high anthropogenic threat for migratory individuals and gulf of Mexico residents. *Remote Sens.* 14, 3887. doi: 10.3390/rs14163887
- Azzola, A., Bianchi, C. N., Merotto, L., Nota, A., Tiralongo, F., Morri, C., et al. (2024). The changing biogeography of the ligurian sea: seawater warming and further records of southern species. *Diversity* 16, 159. doi: 10.3390/d16030159
- Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *MEE* 3, 327–338. doi: 10.1111/j.2041-210x.2011.00172.x
- Baruffaldi, M., Rubini, S., Ignoto, S., Angelini, V., and Tiralongo, F. (2023). Learning from *Caretta caretta* (Linnaeus 1758) epibionts: a study from the Adriatic Sea. *Front. Mar. Sci.* 10, 1243153.
- Bentivegna, F. (2002). Intra-Mediterranean migrations of loggerhead sea turtles (*Caretta caretta*) monitored by satellite telemetry. *Mar. Biol.* 141, 795–800. doi: 10.1007/s00227-002-0856-z
- Bentivegna, F., Valentino, F., Falco, P., Zambianchi, E., and Hochscheid, S. (2007). The relationship between loggerhead turtle (*Caretta caretta*) movement patterns and Mediterranean currents. *Mar. Biol.* 151, 1605–1614. doi: 10.1007/s00227-006-0600-1
- Berrar, D. (2018). Cross-validation. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 542–545). Oxford, UK: Elsevier.
- Berthon, J.-F., and Zibordi, G. (2004). Bio-optical relationships for the northern Adriatic Sea. *Int. J. Remote Sens.* 25, 1527–1532. doi: 10.1080/01431160310001592544
- Bolzon, G., Cossarini, G., Lazzari, P., Salon, S., Teruzzi, A., Feudale, L., et al. (2019). Mediterranean sea biogeochemical analysis and forecast (CMEMS MED-biogeochemistry). *Copernicus Monit. Environ. Mar. Service (CMEMS)*. doi: 10.25423/CMCC/MEDSEA\_ANALYSIS\_FORECAST\_BIO\_006\_014
- Breiman, L. (2001). Random forests. *Mach. Learning* 45, 5–32. doi: 10.1023/A:1010933404324
- Carlucci, R., Cipriano, G., Paoli, C., Ricci, P., Fanizza, C., Capezzuto, F., et al. (2018). Random Forest population modelling of striped and common-bottlenose dolphins in the Gulf of Taranto (Northern Ionian Sea, Central-eastern Mediterranean Sea). *Estuar. Coast. Shelf Sci.* 204, 177–192. doi: 10.1016/j.ecss.2018.02.034
- Casale, P., Abbate, G., Freggi, D., Conte, N., Oliverio, M., and Argano, R. (2008). Foraging ecology of loggerhead sea turtles *Caretta caretta* in the central Mediterranean Sea: evidence for a relaxed life history model. *Mar. Ecol. Prog. Ser.* 372, 265–276. doi: 10.3354/meps07702
- Casale, P., Affronte, M., Scaravelli, D., Lazar, B., Vallini, C., and Luschi, P. (2012). Foraging grounds, movement patterns and habitat connectivity of juvenile loggerhead turtles (*Caretta caretta*) tracked from the Adriatic Sea. *Mar. Biol.* 159, 1527–1535. doi: 10.1007/s00227-012-1937-2
- Casale, P., Broderick, A. C., Camiñas, J. A., Cardona, L., Carreras, C., Demetropoulos, A., et al. (2018). Mediterranean Sea turtles: current knowledge and priorities for conservation and research. *Endanger. Species Res.* 36, 229–267. doi: 10.3354/esr00901
- Casale, P., Freggi, D., Basso, R., Vallini, C., and Argano, R. (2007). A model of area fidelity, nomadism, and distribution patterns of loggerhead sea turtles (*Caretta caretta*) in the Mediterranean Sea. *Mar. Biol.* 152, 1039–1049. doi: 10.1007/s00227-007-0752-7
- Casale, P., and Tucker, (2017). *Caretta caretta*. The IUCN Red List of Threatened Species (amended version of 2015 assessment). IUCN. doi: 10.2305/IUCN.UK.2017-2.RLTS.T3897A119333622.en
- Cavender-Bares, J., Schneider, F. D., Santos, M. J., Armstrong, A., Carnaval, A., Dahlin, K. M., et al. (2022). Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nat. Ecol. Evolution* 6, 506–519. doi: 10.1038/s41559-022-01702-5

## Acknowledgments

In addition, the authors want to thank Michele Attolico for his technical assistance.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Chambault, P., de Thoisy, B., Heerah, K., Conchon, A., Barrios, S., Reis, V. D., et al. (2016). The influence of oceanographic features on the foraging behavior of the olive ridley sea turtle *Lepidochelys olivacea* along the Guiana coast. *Prog. Oceanogr* 142, 58–71. doi: 10.1016/j.poccean.2016.01.006
- Clementi, E., Pistoia, J., Escudier, R., Delrosso, D., Drudi, M., Grandi, A., et al. (2020). Mediterranean Sea Analysis and Forecast (CMEMS MED-Currents, EAS5 system). Mediterranean Sea Analysis and Forecast (CMEMS MED-Currents, EAS5 system). *Copernicus Monit. Environ. Mar. Service (CMEMS)*. doi: 10.25423/CMCC/MEDSEA\_ANALYSIS\_FORECAST\_PHY\_006\_013\_EAS5
- Coyne, M. S., and Godley, B. J. (2005). Satellite Tracking and Analysis Tool (STAT): an integrated system for archiving, analyzing and mapping animal tracking data. *Mar. Ecol. Prog. Ser* 301, 1–7. doi: 10.3354/meps301001
- DiMatteo, A., Lockhart, G., and Barco, S. (2022). Habitat models and assessment of habitat partitioning for Kemp's ridley and loggerhead marine turtles foraging in Chesapeake Bay (USA). *Endanger. Species Res* 47, 91–107. doi: 10.3354/esr01168
- Douglas, D. C., Weinzierl, R., Davidson, S. C., Kays, R., Wikelski, M., and Bohrer, G. (2012). Moderating Argos location errors in animal tracking data. *MEE* 3, 999–1007. doi: 10.1111/j.2041-210x.2012.00245.x
- Dujon, A. M., Lindstrom, R. T., and Hays, G. C. (2014). The accuracy of Fastloc-GPS locations and implications for animal tracking. *MEE* 5, 1162–1169. doi: 10.1111/2041-210X.12286
- Foley, A. M., Schroeder, B. A., Hardy, R., MacPherson, S. L., and Nicholas, M. (2014). Long-term behavior at foraging sites of adult female loggerhead sea turtles (*Caretta caretta*) from three Florida rookeries. *Mar. Biol* 161, 1251–1262. doi: 10.1007/s00227-014-2415-9
- Hawkes, L. A., Broderick, A. C., Coyne, M. S., Godfrey, M. H., Lopez-Jurado, L.-F., Lopez-Suarez, P., et al. (2006). Phenotypically linked dichotomy in sea turtle foraging requires multiple conservation approaches. *Curr. Biol* 16, 990–995. doi: 10.1016/j.cub.2006.03.063
- Haywood, J. C., Casale, P., Freggi, D., Fuller, W. J., Godley, B. J., Lazar, B., et al. (2020). Foraging ecology of Mediterranean juvenile loggerhead turtles: insights from C and N stable isotope ratios. *Mar. Biol* 167, 28. doi: 10.1007/s00227-020-3647-5
- Hazen, E. L., Scales, K. L., Maxwell, S. M., Briscoe, D. K., Welch, H., Bograd, S. J., et al. (2018). A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. *Sci. Adv* 4, eaar3001. doi: 10.1126/sciadv.aar3001
- Hochscheid, S. (2014). Why we mind sea turtles' underwater business: A review on the study of diving behavior. *J. Exp. Mar. Biol. Ecol* 450, 118–136. doi: 10.1016/j.jembe.2013.10.016
- Hochscheid, S., Bentivegna, F., Bradai, M. N., and Hays, G. C. (2007). Overwintering behaviour in sea turtles: dormancy is optional. *Mar. Ecol. Prog. Ser* 340, 287–298. doi: 10.3354/meps340287
- Hochscheid, S., Bentivegna, F., and Hays, G. C. (2005). First records of dive durations for a hibernating sea turtle. *Biol. Lett* 1, 82–86. doi: 10.1098/rsbl.2004.0250
- Hussey, N. E., Kessel, S. T., Aarestrup, K., Cooke, S. J., Cowley, P. D., Fisk, A. T., et al. (2015). Aquatic animal telemetry: A panoramic window into the underwater world. *Science* 348, 1255642. doi: 10.1126/science.1255642
- Inglese, P., Amoroso, N., Boccardi, M., Bocchetta, M., Bruno, S., Chincarini, A., et al. (2015). Multiple RF classifier for the hippocampus segmentation: Method and validation on EADC-ADNI Harmonized Hippocampal Protocol. *Physica Medica* 31, 1085–1091. doi: 10.1016/j.ejmp.2015.08.003
- Jang, C. J., Park, J., Park, T., and Yoo, S. (2011). Response of the ocean mixed layer depth to global warming and its impact on primary production: a case for the North Pacific Ocean. *IJMS* 68, 996–1007. doi: 10.1093/icesjms/fsr064
- Jeanet, L., Dell'Amico, F., Forin-Wiart, M. A., Coutant, M., Bonola, M., Etienne, D., et al. (2018). Combined use of two supervised learning algorithms to model sea turtle behaviours from tri-axial acceleration data. *J. Exp. Biol.* 221 (10), p.jeb177378. doi: 10.1242/jeb.177378
- Kara, A. B., Rochford, P. A., and Hurlburt, H. E. (2000). An optimal definition for ocean mixed layer depth. *J. Geophys. Res* 105, 16803–16821. doi: 10.1029/2000JC900072
- Kays, R., Crofoot, M. C., Jetz, W., and Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science* 348, aaa2478-9. doi: 10.1126/science.aaa2478
- Lambardi, P., Lutjeharms, J. R., Mencacci, R., Hays, G. C., and Luschi, P. (2008). Influence of ocean currents on long-distance movement of leatherback sea turtles in the Southwest Indian Ocean. *Mar. Ecol. Prog. Ser* 353, 289–301. doi: 10.3354/meps07118
- Lazar, B., and Gračan, R. (2011). Ingestion of marine debris by loggerhead sea turtles, *Caretta caretta*, in the Adriatic Sea. *Mar. Poll. Bull* 62, 43–47. doi: 10.1016/j.marpolbul.2010.09.013
- Lazar, B., Margaritoulis, D., and Tvrtković, N. (2004). Tag recoveries of the loggerhead sea turtle *Caretta caretta* in the eastern Adriatic Sea: implications for conservation. *JMBA* 84, 475–480. doi: 10.1017/S0025315404009488h
- Lazar, B., Zuljevic, A., and Holcer, D. (2010). Diet composition of a green turtle, *Chelonia mydas*, from the Adriatic Sea. *Nat. Croat* 19, 263.
- Lotek. *KiwiSat® Glue on series* (Lotek). (2024). Available at: <http://www.lotek.com/products/kiwisat-glue-on-series/>. (accessed April 7, 2024).
- Luschi, P. (2013). Long-distance animal migrations in the oceanic environment: orientation and navigation correlates. *ISRN Zoology* 2013, 1–23. doi: 10.1155/2013/631839
- Luschi, P., and Casale, P. (2014). Movement patterns of marine turtles in the Mediterranean Sea: a review. *Ital. J. Zool* 81, 478–495. doi: 10.1080/11250003.2014.963714
- Luschi, P., Hays, G. C., and Papi, F. (2003). A review of long-distance movements by marine turtles, and the possible role of ocean currents. *Oikos* 103, 293–302. doi: 10.1034/j.1600-0706.2003.12123.x
- Luschi, P., Mencacci, R., Cerritelli, G., Papetti, L., and Hochscheid, S. (2017). Large-scale movements in the oceanic environment identify important foraging areas for loggerheads in central Mediterranean Sea. *Mar. Biol* 165, 4. doi: 10.1007/s00227-017-3255-1
- Maglietta, R., Amoroso, N., Boccardi, M., Bruno, S., Chincarini, A., Frisoni, G. B., et al. (2016). Automated hippocampal segmentation in 3D MRI using random undersampling with boosting algorithm. *PAA* 19, 579–591. doi: 10.1007/s10044-015-0492-0
- Maglietta, R., Carlucci, R., Fanizza, C., and Dimauro, G. (2022). Machine learning and image processing methods for cetacean photo identification: A systematic review. *IEEE Access* 10, 80195–80207. doi: 10.1109/access.2022.3195218
- Maglietta, R., Milella, A., Caccia, M., and Bruzzone, G. (2018). A vision-based system for robotic inspection of marine vessels. *Signal Image Video P.* 12, 471–478. doi: 10.1007/s11760-017-1181-9
- Maglietta, R., Reno, V., Caccioppoli, R., Seller, E., Bellomo, S., Santacesaria, F. C., et al. (2020). Convolutional Neural Networks for Risso's dolphins identification. *IEEE Access* 8, 80195–80206. doi: 10.1109/access.2020.2990427
- Maglietta, R., Saccotelli, L., Fanizza, C., Telesca, V., Dimauro, G., Causio, S., et al. (2023). Environmental variables and machine learning models to predict cetacean abundance in the Central-eastern Mediterranean Sea. *Sci. Rep* 13, 2600. doi: 10.1038/s41598-023-29681-y
- Margaritoulis, D., Argano, R., Baran, I., Bentivegna, F., Bradai, M., Camiñas, J., et al. (2003). "Loggerhead turtles in the Mediterranean Sea: present knowledge and conservation perspectives," in *Loggerhead sea turtles (editors: AB bolten, BE witherington)* (Smithsonian Institution Press, Washington DC).
- Marini, C., Fossa, F., Paoli, C., Bellingeri, M., Gnone, G., and Vassallo, P. (2015). Predicting bottlenose dolphin distribution along Liguria coast (northwestern Mediterranean Sea) through different modeling techniques and indirect predictors. *J. Environ. Manage* 150, 9–20. doi: 10.1016/j.jenvman.2014.11.008
- Marshall, C. D., Guzman, A., Narazaki, T., Sato, K., Kane, E. A., and Sterba-Boatwright, B. D. (2012). The ontogenetic scaling of bite force and head size in loggerhead sea turtles (*Caretta caretta*): implications for durophagy in neritic, benthic habitats. *J. Exp. Biol* 215, 4166–4174. doi: 10.1242/jeb.074385
- Mazaris, A., Kornaraki, E., Matsinos, Y., and Margaritoulis, D. (2004). Modelling the effect of sea surface temperature on sea turtle nesting. *Nat. Resour. Model* 17, 445–465. doi: 10.1111/j.1939-7445.2004.tb00145.x
- Mazaris, A. D., Kallimanis, A. S., Sgardelis, S. P., and Pantis, J. D. (2008). Do long-term changes in sea surface temperature at the breeding areas affect the breeding dates and reproduction performance of Mediterranean loggerhead turtles? Implications for climate change. *J. Exp. Mar. Biol. Ecol* 367, 219–226. doi: 10.1016/j.jembe.2008.09.025
- Mazaris, A. D., Kallimanis, A. S., Tzanopoulos, J., Sgardelis, S. P., and Pantis, J. D. (2009). Sea surface temperature variations in core foraging grounds drive nesting trends and phenology of loggerhead turtles in the Mediterranean Sea. *J. Exp. Mar. Biol. Ecol* 379, 23–27. doi: 10.1016/j.jembe.2009.07.026
- Mazor, T., Beger, M., McGowan, J., Possingham, H. P., and Kark, S. (2016). The value of migration information for conservation prioritization of sea turtles in the Mediterranean. *Glob. Ecol. Biogeogr* 25, 540–552. doi: 10.1111/geb.12434
- McCarthy, A. L., Heppell, S., Royer, F., Freitas, C., and Dellinger, T. (2010). Identification of likely foraging habitat of pelagic loggerhead sea turtles (*Caretta caretta*) in the North Atlantic through analysis of telemetry track sinuosity. *Prog. Oceanogr* 86, 224–231. doi: 10.1016/j.poccean.2010.04.009
- McClellan, C. M., and Read, A. J. (2007). Complexity and variation in loggerhead sea turtle life history. *Biol. Lett* 3, 592–594. doi: 10.1098/rsbl.2007.0355
- McDougall, T. J., Feistel, R., Millero, F. J., Jackett, D. R., Wright, D. G., King, B. A., et al. (2009). "The international thermodynamic equation of seawater 2010 (TEOS-10): Calculation and use of thermodynamic properties," in *Global ship-based repeat hydrography manual, IOCCP report no. 14* (Paris: UNESCO).
- McMahon, C. R., and Hays, G. C. (2006). Thermal niche, large-scale movements and implications of climate change for a critically endangered marine vertebrate. *Glob. Change Biol* 12, 1330–1338. doi: 10.1111/j.1365-2486.2006.01174.x
- Mencacci, R., Vallini, C., Rubini, S., Luschi, P., Sarti, A., Benvenuti, S., et al. (2006). "Movements of a male loggerhead sea turtle (*Caretta caretta*) tracked by satellite in the Adriatic sea," in *In Societas herpetologica italiana: atti del V Congresso nazionale: Calci (PI), 29 settembre-3 ottobre 2004.-(Atti; 27)* (Italy: Firenze University Press), 1000–1005.
- Mingozzi, T., Mencacci, R., Cerritelli, G., Giunchi, D., and Luschi, P. (2016). Living between widely separated areas: Long-term monitoring of Mediterranean loggerhead turtles sheds light on cryptic aspects of females spatial ecology. *J. Exp. Mar. Biol. Ecol* 485, 8–17. doi: 10.1016/j.jembe.2016.08.007

- Moudry, V., Bazzichetto, M., Remelgado, R., Devillers, R., Lenoir, J., Mateo, R. G., et al. (2024). Optimising occurrence data in species distribution models: sample size, positional uncertainty, and sampling bias matter. *Ecography* p, 07294. doi: 10.1111/ecog.07294
- Mushagalusa, C. A., Fandohan, A. B., and Glèlè Kakaï, R. (2024). Random forest and spatial cross-validation performance in predicting species abundance distributions. *Environ. Syst. Res* 13, 23. doi: 10.1186/s40068-024-00352-9
- Nagelkerken, I., Pors, L. P., and Hoetjes, P. (2003). Swimming behaviour and dispersal patterns of headstarted loggerhead turtles *Caretta caretta*. *Aquat. Ecol* 37, 183–190. doi: 10.1023/A:1023924631480
- O'Hara, C. C., Villaseñor-Derbez, J. C., Ralph, G. M., and Halpern, B. S. (2019). Mapping status and conservation of global at-risk marine biodiversity (Hoboken, NJ, USA: Wiley). doi: 10.1111/conl.12651
- Parker, D., Cooke, W., and Balazs, G. (2005). Diet of oceanic loggerhead sea turtles (*Caretta caretta*) in the central North Pacific. *Fishery Bulletin* 103 (1), 142–153.
- Patel, S. H., Morreale, S. J., Panagopoulou, A., Bailey, H., Robinson, N. J., Paladino, F. V., et al. (2015). Change-point analysis: a new approach for revealing animal movements and behaviors from satellite telemetry data. *Ecosphere* 6, 1–13. doi: 10.1890/ES15-00358.1
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res* 12, 2825–2830.
- Pierantonio, N., Panigada, S., and Lauriano, G. (2023). Quantifying abundance and mapping distribution of loggerhead turtles in the mediterranean sea using aerial surveys: implications for conservation. *Diversity* 15, 1159. doi: 10.3390/d15121159
- Ravaoli, M., Alvisi, F., and Vitturi, L. M. (2003). Dolomite as a tracer for sediment transport and deposition on the northwestern Adriatic continental shelf (Adriatic Sea, Italy). *Continental Shelf Res* 23, 1359–1377. doi: 10.1016/S0278-4343(03)00121-3
- Rees, A. F., Margaritoulis, D., Newman, R., Riggall, T. E., Tsaros, P., Zbinden, J. A., et al. (2013). Ecology of loggerhead marine turtles *Caretta caretta* in a neritic foraging habitat: movements, sex ratios and growth rates. *Mar. Biol* 160, 519–529. doi: 10.1007/s00227-012-2107-2
- Revelles, M., Cardona, L., Aguilar, A., San Felix, M., and Fernández, G. (2007). Habitat use by immature loggerhead sea turtles in the Algerian Basin (western Mediterranean): Swimming behaviour, seasonality and dispersal pattern. *Mar. Biol* 151, 1501–1515. doi: 10.1007/s00227-006-0602-z
- Roquet, F., Madec, G., McDougall, T. J., and Barker, P. M. (2015). Accurate polynomial expressions for the density and specific volume of seawater using the TEOS-10 standard. *Ocean Model* 90, 29–43. doi: 10.1016/j.ocemod.2015.04.002
- Saarela, M., and Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Appl. Sci* 3, 272. doi: 10.1007/s42452-021-04148-9
- Schofield, G., Dimadi, A., Fossette, S., Katselidis, K. A., Koutsoubas, D., Lilley, M. K., et al. (2013a). Satellite tracking large numbers of individuals to infer population level dispersal and core areas for the protection of an endangered species. *Diversity Distributions* 19, 834–844. doi: 10.1111/ddi.12077
- Schofield, G., Scott, R., Dimadi, A., Fossette, S., Katselidis, K. A., Koutsoubas, D., et al. (2013b). Evidence-based marine protected area planning for a highly mobile endangered marine vertebrate. *Biol. Conserv* 161, 101–109. doi: 10.1016/j.biocon.2013.03.004
- Snape, R. T., Broderick, A. C., Çiçek, B. A., Fuller, W. J., Glen, F., Stokes, K., et al. (2013b). Shelf life: neritic habitat use of a turtle population highly threatened by fisheries. *Diversity Distributions* 22, 797–807. doi: 10.1111/ddi.2016.22.issue-7
- Snoek, J., Larochelle, H., and Adams, R. P. (2012). “Practical bayesian optimization of machine learning algorithms,” in *Advances in neural information processing systems* (Red Hook, NY, USA: Curran Associates, Inc.). Available at: <https://papers.nips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html>.
- Sverdrup, H. U. (1953). On conditions for the vernal blooming of phytoplankton. *ICES J. Mar. Science* 18, 287–295. doi: 10.1093/icesjms/18.3.287
- Timko, R. E., and Kolz, A. L. (1982). “Satellite sea turtle tracking,” in *Mar fish rev* (USA: Silver Spring, MD), vol. 44., 19–24. Available at: <https://spo.nmfs.noaa.gov/sites/default/files/pdf-content/MFR/mfr444/mfr4442.pdf>.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2
- Volpe, G., Buongiorno Nardelli, B., Colella, S., Pisano, A., and Santoleri, R. (2018). An operational interpolated ocean colour product in the mediterranean sea. *New Frontiers in Operational Oceanography*. 227–244. doi: 10.17125/gov2018.ch09
- Volpe, G., Colella, S., Brando, V. E., Forneris, V., La Padula, F., Di Cicco, A., et al. (2019). Mediterranean ocean colour Level 3 operational multi-sensor processing. *Ocean Sci* 15, 127–146. doi: 10.5194/os-15-127-2019
- Yentsch, C. S. (1990). Estimates of ‘new production’ in the mid-north atlantic 1. *J. Plankton Res* 12, 717–734. doi: 10.1093/plankt/12.4.717
- Zampollo, A., Arcangeli, A., Costantino, M., Mancino, C., Crosti, R., Pietrolungo, G., et al. (2022). Seasonal niche and spatial distribution modelling of the loggerhead (*Caretta caretta*) in the Adriatic and Ionian seas. *Aquat. Conserv.: Mar. Freshw. Ecosyst* 32, 1141–1155. doi: 10.1002/aqc.3815
- Zbinden, J. A., Aebischer, A., Margaritoulis, D., and Arlettaz, R. (2007). Important areas at sea for adult loggerhead sea turtles in the Mediterranean Sea: satellite tracking corroborates findings from potentially biased sources. *Mar. Biol* 153, 899–906. doi: 10.1007/s00227-007-0862-2
- Zbinden, J. A., Bearhop, S., Bradshaw, P., Gill, B., Margaritoulis, D., Newton, J., et al. (2011). Migratory dichotomy and associated phenotypic variation in marine turtles revealed by satellite tracking and stable isotope analysis. *Mar. Ecol. Prog. Ser* 421, 291–302. doi: 10.3354/meps08871