Andrea Sozzi

Mapping and predicting coralligenous habitats with a single-beam echosounder along the coast of the Fourni Islands, Aegean Sea, Greece

Bachelor's thesis

Bachelor of Engineering

Environmental Engineering

2024

South-Eastern Finland
University of Applied Sciences

South-Eastern Finland University of Applied Sciences

Degree title	Bachelor of Engineering
Author(s)	Andrea Sozzi
Thesis title	Mapping and predicting coralligenous habitats with a single-beam echosounder along the coast of the Fourni Islands, Aegean Sea, Greece
Commissioned by	Dr. Tim Grandjean - Archipelagos Institute of Marine Conservation
Time	2024
Pages	89 pages, 6 pages of appendices
Supervisor	Dr. Shakil Regmi

ABSTRACT

The aim of the study was to provide information about the potential distribution of coralligenous habitats by introducing a methodology to assess, classify, predict, and map the potential presence of such formations of the seabed from single-beam echosounder (Simrad EK 80). The data was collected during a two-day boat survey along the coastline of the Fourni Islands, Greece, on the 25th of September 2023 covering east, south and west of the area of interest, and on the 26th of September 2023, covering the northern part of the area.

Features of the bottom of the sea were first extrapolated from the acquired sonar data, by running a bottom classification module in Echoview 13 licensed software. Subsequently, from the derived selected features, a Random Forest (RF) classifier was developed in Python environment (and Scikit-learn library) to determine prediction of potential presence “Yes” class, and non-presence “No” class. The RF model was created and trained with the dataset of the first-day survey by reviewing topographic characteristics of coralligenous assemblages, such as depth and structure, together with the echograms from the echosounder data. Citizen science data, images, and videos from Remote Operated Vehicle (ROV) dives performed in the area of interest were also considered when developing the model, as complementary data. The RF classifier was then applied on the second-day survey dataset, and its performance tested. The *accuracy*, *precision*, *recall*, and *F1 score* of the model were 90, 73, 93, 81 percent, respectively. The predicted potential coralligenous presence segments were then mapped along the transect in GIS software for cartographical representation.

Ground truthing should be employed to ultimately confirm the predictions of the Random Forest classifier. However, considering the intrinsic challenges to explore these marine habitats, the developed methodology and the RF model provide the commissioner with more tools to define coralligenous formations in the area of research, in the effort to optimize conservation endeavors and protection activities of these hidden underwater gems.

Keywords: coralligenous habitat, sonar data, GIS, random forest

CONTENTS

1 INTRODUCTION

Oceans and seas across the globe are heavily under stress. Disturbances, mainly derived by anthropogenic activities, are affecting the habitats and intricate ecosystems present within their waters (Halpern et al. 2019, 1). The Mediterranean Sea and its several marginal seas are no different. The Aegean Sea (Figure 1), situated in the northeastern part of the Mediterranean, between the coasts of Greece and Turkey, hosts a rich marine biodiversity. It constitutes a unique ecoregion characterized by distinctive oceanographic, topographic, and biological features, allowing different species to thrive (Sini et al. 2017, 3).

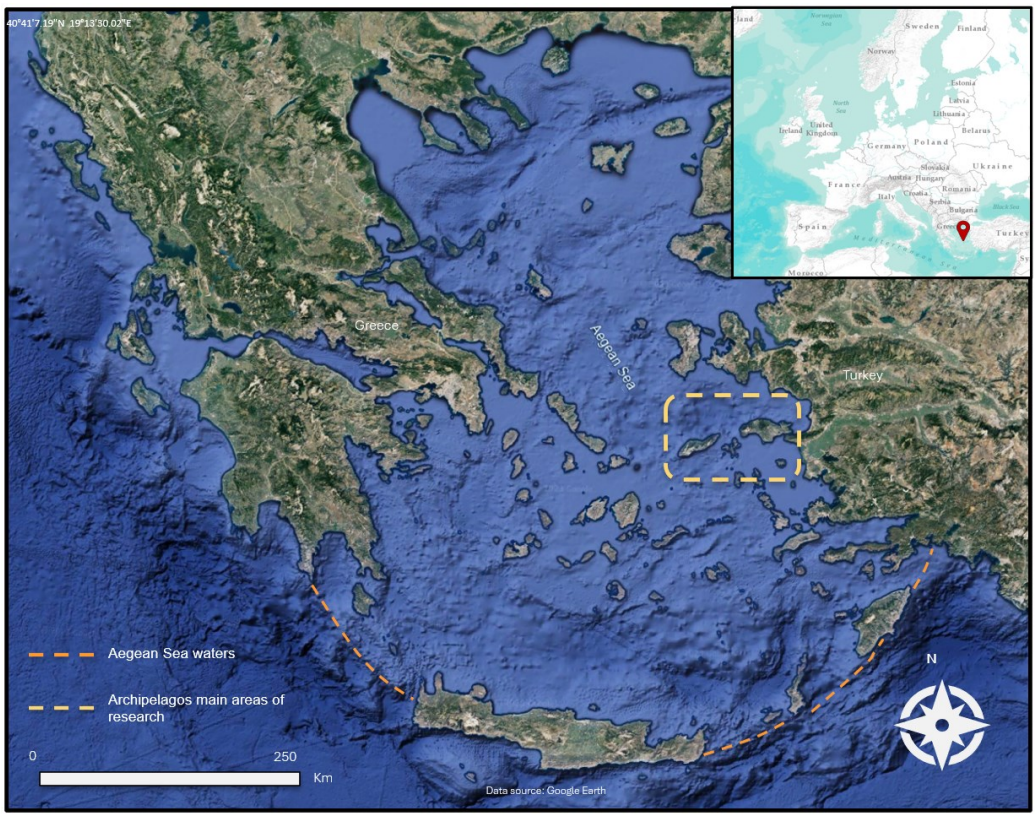


Figure 1. Aegean Sea waters, between the coasts of Greece and Turkey

As well as other numerous marine and coastal environments, the endemic coralligenous formations are facing growing pressures from a wide array of human-induced disruptions associated with activities such as pollution, overfishing, littering, changes in sedimentation patterns, invasive species, and the impacts of climate change (Ballesteros 2006, 174-178; Piazzi et al. 2012, 2623; Salomidi et al. 2012; Bevilacqua et al. 2018, 1). However, thanks to their

ecological significance, several initiatives towards their protection and conservation have been prompted through international agreements (Giakoumi et al. 2013, 2).

In this light, over the past four decades, the European Union has implemented a series of protective measures aimed at minimizing the impact on ecosystems and increasing public awareness (Fakiris et al. 2023, 2). Designating specific ecosystems as priority habitats with the European Union's Habitat Directive (EU 92/43/CEE 1992) was the first step. Coralligenous formations were included in the Habitat Type 1170, commonly referred to as Reefs. They were also included in the network of the Natura 2000 sites. The second milestone was ratified within the Barcelona Convention (UNEP-MAP-RAC/SPA 2008), where various species of corals received special attention under the "Protocol concerning Specially Protected Areas and Biological Diversity in the Mediterranean". With the introduction of the Marine Strategy Framework Directive 2008/56/EC, the European Union added one more layer to address member states on how to safeguard marine environments and their habitats, also by introducing the “Good Environment Status” (GES) concept, which entails oceans being ecologically diverse, clean, healthy, and productive (EU 2008/56/EC 2008).

The lack of cartographic data regarding fine-scale distribution of coralligenous habitats, along with the absence of consistent and comparable information on the ecological health of its communities, represents a relevant obstacle to the actual formulation of effective management and conservation strategies and implementation actions (UNEP-MAP-RAC/SPA 2008). In this light, spatial distribution of coralligenous habitats and species becomes imperative to the success of implementing the right conservation measures also in the Aegean Sea (Sini et al. 2017, 3).

The present thesis aims to provide information about the potential distribution of coralligenous habitats by developing a methodology that assesses, classifies, and predicts the presence of such formations on the bottom of the sea, using single-beam echosounder data, collected on a two-day boat survey. The

methodology combines sonar data collection, the use of licensed software for deriving features of the bottom, and the development of a Random Forest (RF) model for prediction of the presence of the habitat. The results will then be mapped in GIS (Geographical Information Systems). The main question to be addressed in the study is:

Can single-beam echosounder data be used to map and predict the presence of coralligenous formations in a surveyed area?

This research will help conservation in the area of interest. Archipelagos Institute of Marine Conservation - the commissioner of the thesis - by researching, finding, and bringing habitat locations to the attention of policymakers with reports and scientific documents, encourages to take actions serving as a backbone for marine protection laws, and supporting the process of establishing further Marine Protected Areas in the Aegean Sea.

2 LITERATURE REVIEW

2.1 Coralligenous outcrops

The term coralligenous dates back to 1883, when Marion (1883) first introduced it, literally meaning “producer of coral”, to distinguish those mesobenthic habitats from seagrass meadows and muddy bottoms (e.g. dependent on depth, light availability, substratum). According to Ballesteros (2006, 123), coralligenous assemblages are “hard substratum of biogenic origin and built by the accumulation of encrusting algae”, which can grow in low light conditions. They can be mostly found in the circalittoral zone, even though their growth can occur also within the infralittoral zone, where dim light conditions allow it. The biota thriving within the coralligenous habitats are referred to as sciaphilic: algae and invertebrates that can grow in low light environments and are distinguished in different communities (Ballesteros 2006, 123): living algae that are mostly present in the upper side of the assemblage, suspension feeders in the lower part and cavities, borers within the concretions, and soft-bottom fauna, present in the sediment inside the holes of the assemblage (Laborel 1961).

Overall, coralligenous habitat (Figure 2) should be viewed as a complex underwater landscape, a mosaic of different species rather than as one single community, in which builders and eroders shape its form, and several environmental factors affect its composition (Ballesteros 2006, 140).

Figure 2. Section of a coralligenous bank, showcasing its heterogeneous habitats (Ballesteros 2006, by Corbera)

2.1.1 Environmental factors

The development and growth of coralligenous outcrops require light, but without too high levels of irradiance (Peres & Picard 1964) as their principal builder, macroalgae, require a specific irradiance range to thrive, amounting to a minimum of 1.3 megajoules of light energy per square meter annually, up to a maximum of 50-100 megajoules per square meter annually. This translates between 0.05 percent and 3 percent of the total energy reaching the surface (Ballesteros 1992, 557). In terms of dissolved nutrients, coralligenous assemblages appear to have adapted to low nutrient levels in seawater (Ballesteros 1992, 520). When nutrient levels rise, they significantly alter the specific composition of these communities, hindering the formation of coralligenous structures, and leading to higher rates of destruction (Ballesteros 2006, 176).

Normal seasonal temperature ranges, found in different parts of the Mediterranean waters, are supported by coralligenous communities, varying from 10-23 degrees Celsius on the coast of southern France (Laubier 1966) as well as

on the coast of Spain from wintertime to the beginning of summer (13 to 16 degrees Celsius), also with an increase of temperature in summer of up to 22 degree Celsius at 40 meters depth (Ballesteros 1992, 168).

Currents occur at the depths where the coralligenous formations thrive, as noted by Riedl in 1966, and hydrodynamics of waves play a relevant role, even at depths of 50 meters, for wave heights exceeding 1 meter (Ballesteros 2006, 129). However, due to the intricate structure of the coralligenous formations, hydrodynamics can vary significantly across different microenvironments (Ballesteros 2003, 11).

Coralligenous outcrops can be found at different depths, depending on the areas of the Mediterranean, ranging from 20 meters up to 140 meters (Laborel 1961; Laubier 1966; Gill & Ros 1984; Ballesteros 1992; Georgiadis et al. 2009; Giakoumi et al. 2013; Martin et al. 2014; Ingrassia et al. 2019; Pierdomenico et al. 2021; Fakiris et al. 2023). In the Aegean Sea, the maximum coverage of coralligenous concretions was found to be in the range of 63 meters to 110 meters (Fakiris et al. 2023, 13). Moreover, in previous studies based on Remote Operated Vehicle (ROV) dives conducted during the period 2021-2023 by Archipelagos in the study area, the habitat depth distribution for coralligenous concretions was found in the range between 20 meters and 100 meters; however, with the majority found in the range of 50 to 90 meters (Figure 3 and 4) (internal communication at Archipelagos 2023).

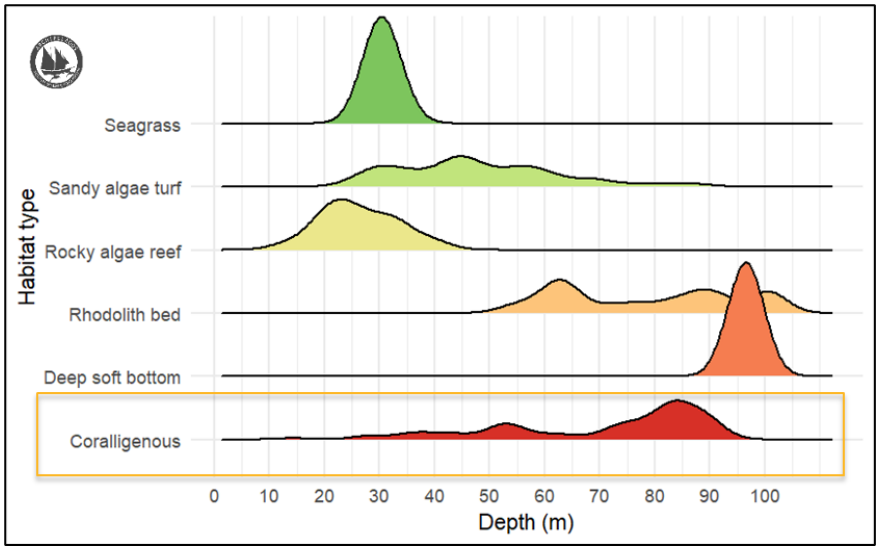


Figure 3. Habitat types and their distribution on depth based on ROV dives in 2021-2023 (Internal communication at Archipelagos, by Cao Sánchez)

This range differs from the other main habitat in the study area, seagrass meadows, which are predominantly found up to a depth of approximately 40 meters (internal communication at Archipelagos 2023).

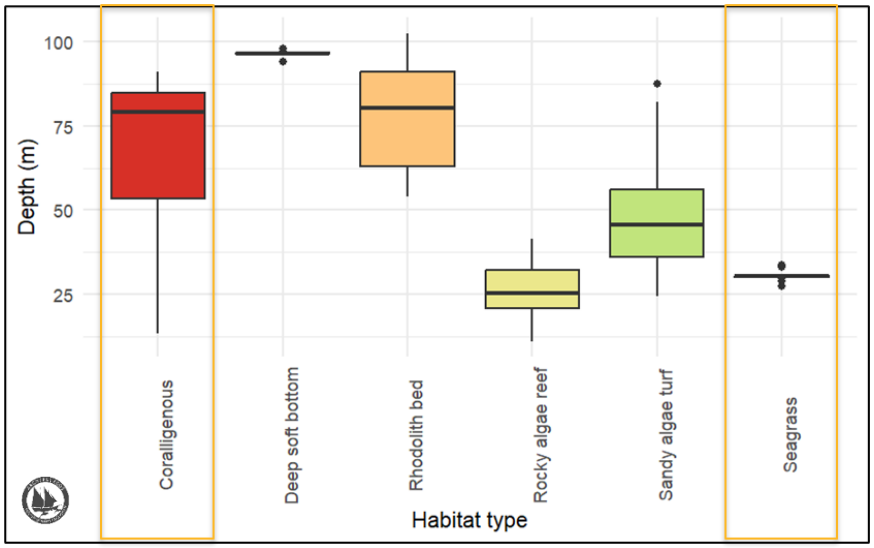


Figure 4. Boxplots for habitat types and their distribution on depth on ROV dives during 2021-2023 (Internal communication at Archipelagos, by Cao Sánchez)

2.1.2 Geomorphology and structure

According to Laborel (1961), the geomorphology and structure of coralligenous assemblages depend not only on the depth but also on the topography and the prevailing algal constructors. Specifically, two main morphologies are found,

banks and rims (Peres & Picard 1964; Laborel 1986). Banks are relatively level structures with a thickness that can vary between 0.5 to several meters, typically around 3 to 4 meters in height. They are primarily constructed on relatively flat substrates and possess a highly porous structure, characterized by numerous cavities and holes. Rims instead tend to develop on vertical cliffs, and external sections of caves, and their thickness varies, spanning from 20 to 25 centimeters to over 2 meters. From shallower to deeper waters, the thickness of the rims tends to increase. (Laborel 1986.) Coralligenous formations in the Aegean Sea tend to form structures in “sub-cropping or out-cropping areas” (Fakiris et al. 2023, 13), on underwater slopes surrounding islands, coastal slopes, and ridges (Georgiadis et al. 2009; Martin et al. 2014; Fakiris et al. 2023). Moreover, ROV data collected in the study area by Archipelagos during the period 2021-2023 provide a visual representation of the type of structure surrounding the Fourni Islands, in the Aegean waters as rim structures (Figure 5) (internal communication at Archipelagos 2023).

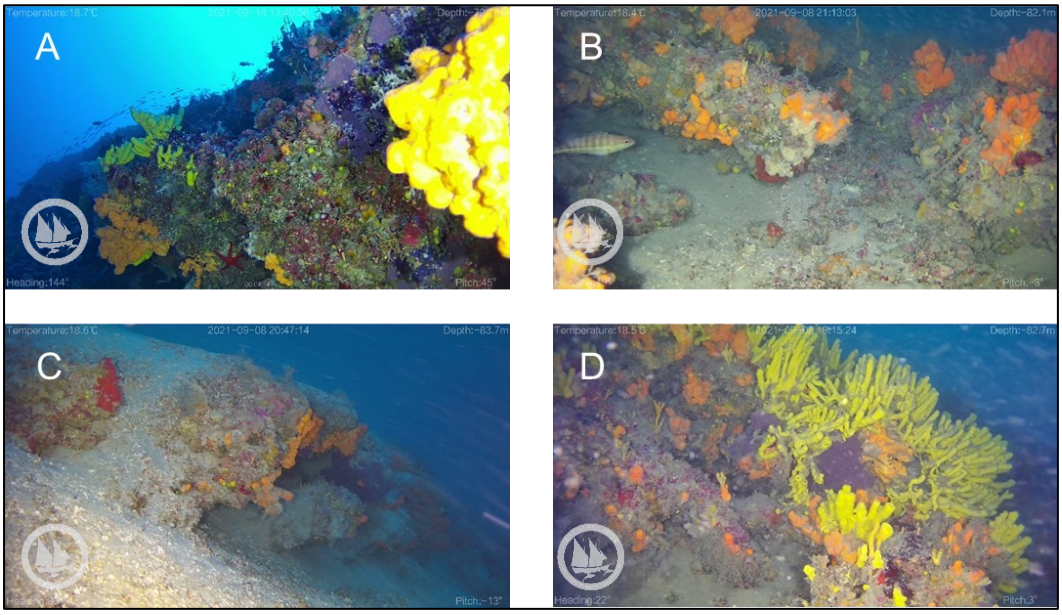


Figure 5. Different examples of coralligenous concretions in the study area during 2021-2023. A. Rim on the side of a vertical cliff. B. Rims forming an overhang. C. Concretion on the external section of a small cave. D. Concretion formed on the side of a seamount slope (Internal communication Archipelagos 2023)

The holes and hollow spaces within the coralligenous structure consistently support a complex community that is predominantly composed of a variety of

suspension-feeding organisms such as sponges, anthozoans, bryozoans, hydrozoans, serpulids, tunicates, and mollusks. Even the tiniest crevices and gaps within the coralligenous formation host an exceptionally diverse and abundant group of biotas, including polychaetes and crustaceans. Moreover, numerous organisms, whether attached or unattached, populate the primary macroalgae and macrofauna, existing on the surface of the concretion or inside cavities. (Ballesteros 2006, 133.)

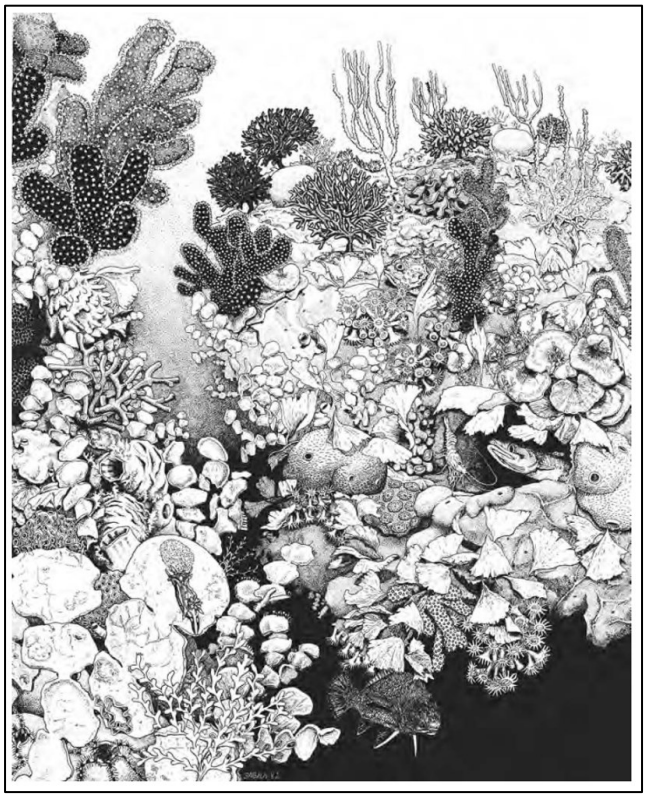


Figure 6. Coralligenous assemblage dominated by algae (Ballesteros 2006)

As discussed in Ballesteros (2006, 134), invertebrates in the coralligenous structure can be categorized into four groups based on their position and ecological role. The first category is composed of buildup-contributing fauna. These organisms assist in the development and solidification of the framework formed by calcareous algae, and they include various bryozoans, serpulids, corals, and sponges, making up 24 percent of the total species count. The second category includes cryptofauna. They inhabit the small gaps and crevices within the coralligenous structure. They account for roughly 7 percent of the species and encompass various crustaceans, mollusks, and polychaetes.

Epifauna and endofauna represent the third category, living on top of the assemblage or inside the sediments present in it. A significant majority of the species are part of this group, amounting to nearly 67 percent. The last category concerns eroding species, amounting to about 1 percent of the species present in the concretion. (Hong 1982.)

2.1.3 Threats and disturbances to coralligenous habitats

The combination of low recruitment rates, extended lifespan, and the immobile nature of species that shape coralligenous assemblages (Teixidó et al. 2011, 1) make them particularly susceptible to anthropogenic impacts. According to Ballesteros (2006, 174-178), coralligenous habitats are threatened by different types of stressors, namely: fishing, diving, large scale events (such as ocean acidification and climate change), waste waters, and invasive species.

Fishing equipment and lines have the potential to harm the fauna by unintended removal, breakage of their branches, and tissue abrasion, resulting in necrosis (Angiolillo et al. 2015, 150; Ferrigno et al. 2017, 42; Consoli et al. 2019, 478; Enrichetti et al. 2019, 114). Trawling, especially, poses a significant threat as it not only physically damages the coralligenous structure irreversibly, but also leads to increased turbidity and sedimentation rates, which affect the level of photosynthesis produced by the species inhabiting the coralligenous outcrops. Overall, these factors have a detrimental impact on the growth of algae and the feeding habits of suspension feeders. (Ballesteros 2006, 174; Linders et al. 2018, 786-787.) Diving activities recently also became a considerable threat and potential disturbance to coralligenous habitats (Boudouresque 2004, 128; Ballesteros 2006, 177). The zones mainly affected by diving are those “stormed” by recreational divers (those sites in which coralligenous outcrops are present at shallower depth of less than 40 meters). Moreover, as coralligenous outcrops are based primarily on calcareous organisms, they are affected by sea surface temperature changes, ocean acidification and climate change, threats occurring currently at a global level (Steller et al. 2007, 451; Zunino et al. 2019, 2; Ceccherelli et al. 2020, 2; Gomes-Graz et al. 2021).

Another disturbance affecting coralligenous communities is waste waters which block coralline algae development while increasing, on the one hand, the rates at which bioerosion occurs, and on the other hand produces a decrease in species diversity and abundance (Ballesteros 2006, 180). Finally, invasive alien species are developing as the next potential disturbance, increasing their population across the Mediterranean. In particular, the red turf alga, *Womersleyella setacea*, has been identified as a detrimental species for coralligenous habitats. This is attributed to its formation of a thick carpet that covers and hinders the growth of essential coralline organisms, thereby inhibiting the development of key coralligenous builders. (Ballesteros 2006, 178.)

2.1.4 Previous approaches to coralligenous distribution in the Aegean

The coralligenous communities rank as the second most significant “hotspot” of species diversity in the Mediterranean, following the Posidonia Oceanica meadows (Boudouresque 2004, 118). Given their abundant fauna (Laubier 1966), intricate structure (Pérès & Picard 1964), and the scarcity of research specifically focusing on coralligenous biodiversity, it is likely that they host a greater number of species compared to any other Mediterranean community (Ballesteros 2006, 147).

In the Aegean Sea, the geographical distribution of coralligenous outcrops has been based mainly on previous mapping studies, with different spatial resolutions (Sini et al. 2017, 7). Georgiadis et al. (2009) studied the morphology and distribution of coralligenous formation in the southern Aegean Sea on the Cyclades plateau, by using a combination of sonars (e.g. single-beam echo sounder, sidescan sonar), sub-bottom profiler recordings, together with ground-truthing techniques, both biological and from sediment analysis.

Giakoumi et al. (2013) produced basin-scale and ecoregional scale distribution maps, through a systematic planning approach based on the identification of priority areas and opportunity costs. Specifically, spatial priorities for the conservation of Posidonia Oceanica, marine caves, and coralligenous outcrops (Figure 7) were identified and compared with prior identified Marine Protected

Areas (MPAs), Natura 2000 sites, and SPAMIs (Specially Protected Areas of Mediterranean Importance), of the Mediterranean, including the Aegean Sea. Based on the whole-basin scale scenario, 23 percent of the areas within the Aegean Sea, with special emphasis on the Cyclades Archipelago and along the Turkish coast, were identified as high priority for conservation (second highest after the Ionian Sea, with 25 percent of areas) (Figure 8). (Giakoumi et al. 2013, 7-8.)

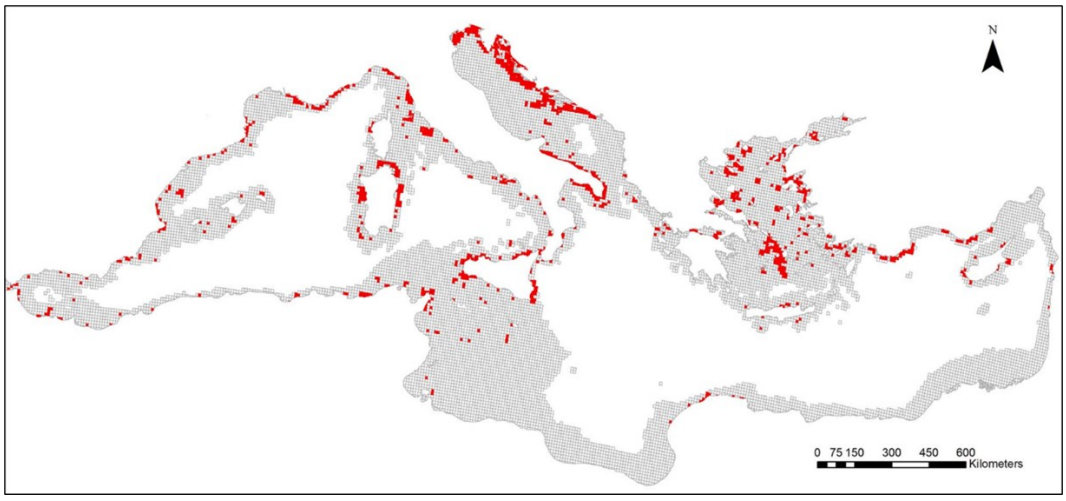


Figure 7. Coralligenous habitat distribution in the Mediterranean Sea (Giakoumi et al. 2013)

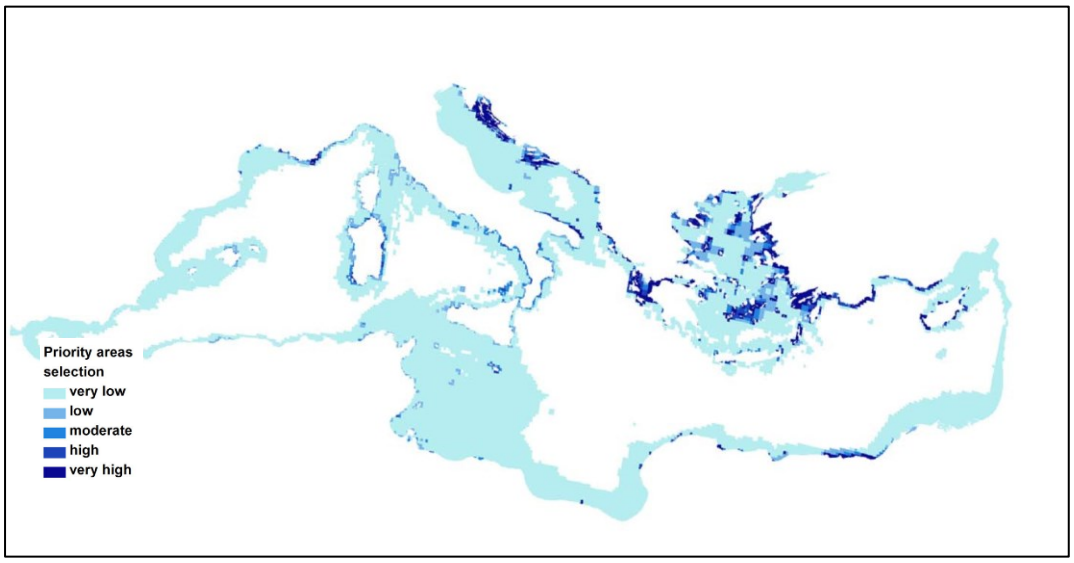


Figure 8. Mediterranean-basin planning scenario, priority areas (Giakoumi et al. 2013)

Martin et al. (2014) reviewed existing spatial data of coralligenous outcrops across the Mediterranean, in the form of various types of datasets and file formats. The surface area of the available coralligenous polygons analyzed in the

study amounted to 2763.4 square kilometers, and, when also added lines and points data, the mapped coastline of the Mediterranean basin represented around 30 percent of the total coasts (Figure 9) (Martin et al. 2014, 2).

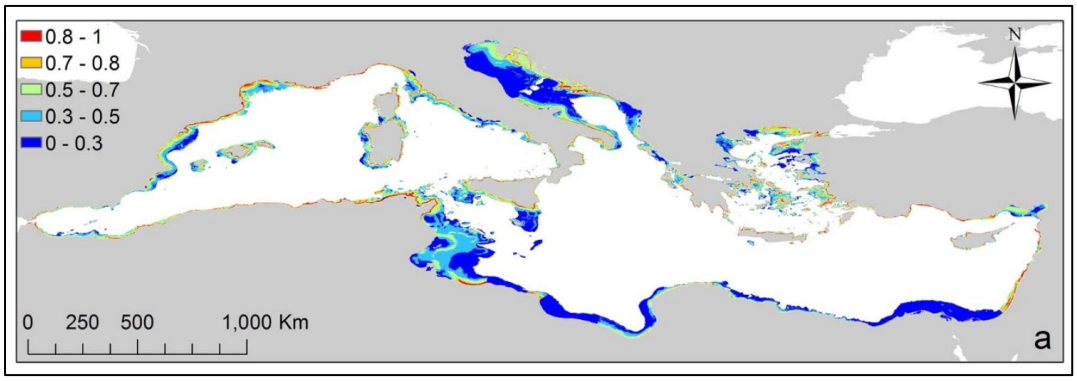


Figure 9. Occurrence probabilities for coralligenous outcrops in the Mediterranean basin (Martin et al. 2014)

In the distribution model used for the occurrence probabilities of coralligenous outcrops, the predictor variables with the highest contribution (in percentage) were the bathymetry, which contributed to its development for 37,4 percent, followed by the slope of the seafloor, with a contribution of 31,9 percent (Martin et al. 2014, 4).

In the Aegean Sea, a research presented by Sini et al. (2017) combined several types of information available from different data sources to map the distribution of several different ecological components, including also coralligenous assemblages (and rhodoliths). The distribution encompassed 49 percent of coralligenous platforms, 30 percent of coralligenous of the littoral rock, and 21 percent was unspecified. The coralligenous platforms were mainly mapped in the South Aegean, specifically on the Cyclades Plateau while the coralligenous of the littoral rock, amounted more in the North Aegean, especially along the coast of Chalkidiki Pelio, and the North Sporades Islands (Figure 10). (Sini et al. 2017, 12.)

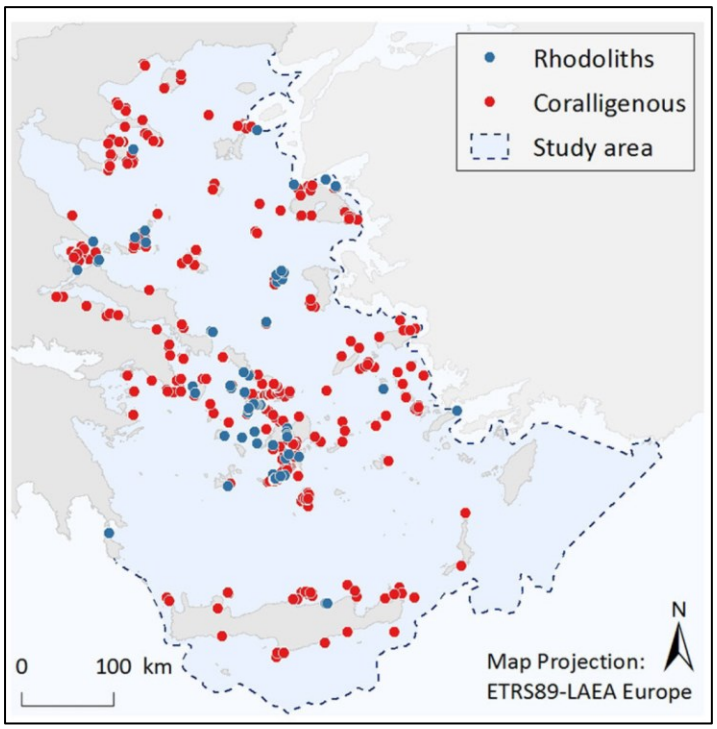


Figure 10. Coralligenous (and Rhodolith) data from previous studies in the Aegean Sea (Sini et al. 2017)

Understanding the distribution of coralligenous habitats is key for conservation planning (Giakoumi et al.2013, 3), considering that 95 percent of coralligenous habitats are still to be mapped, in the Mediterranean, specifically in lower depths (Martin et al. 2014, 5). Further studies in the Aegean should be performed to achieve a higher completeness of the available data, for improved mapping of coralligenous presence in the area (Sini et al. 2017, 15). Towards this direction, Fakiris et al. (2023) developed an atlas of presence-absence of coralligenous formation, based 30 years record of hydroacoustic data, which helped create probability maps of their occurrence in the Aegean Sea (for a total seafloor area of 3197,68 square kilometers) (Figure 11) that should improve the capability of proper spatial planning and monitoring activities. Topographic features such as depth, and especially slope, bBPI, fBPI, proved to be important indices for suitability modeling when considering coralligenous outcrops distribution in the Aegean basin. (Fakiris et al. 2023, 7-10.)

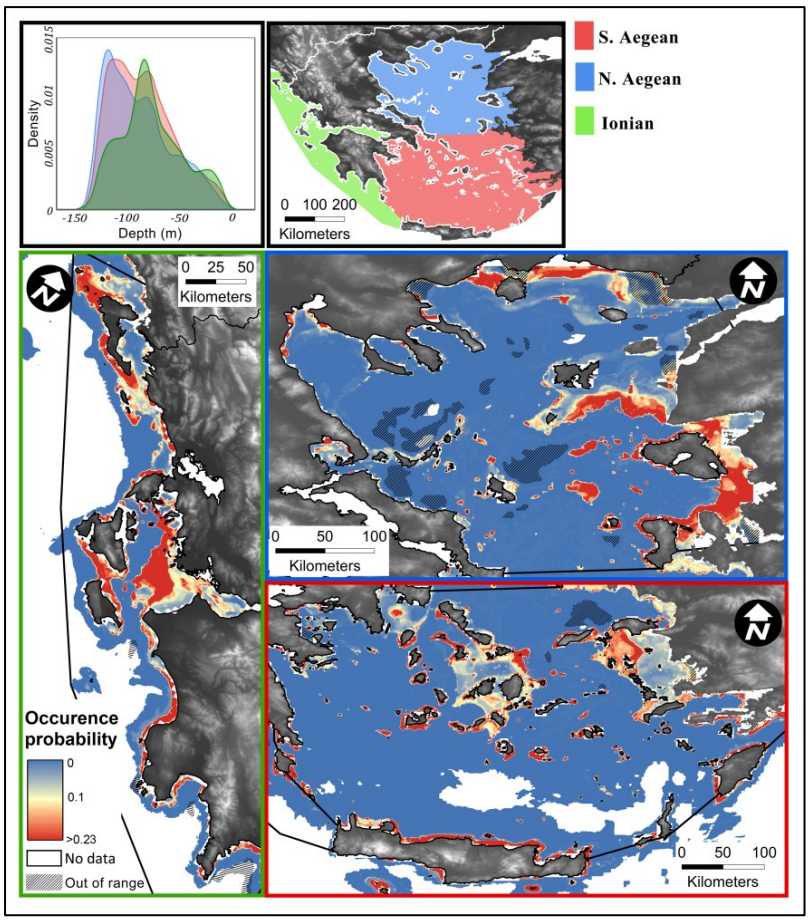


Figure 11. ”Spatial output” presented by a predictive model, divided in the three ecoregion, Ionian, South Aegean and North (Fakiris et al. 2023)

2.2 Acoustic Seabed Classification (ASC)

Acoustic Seabed Classification (ASC) is an acoustic remote sensing technique which enables the visualization of the seabed's material composition and topographical features by using “pseudo-colors”, using different types of instruments, such as single-beam echosounders (SBES), sidescan sonars (SSS), or multibeam sonar systems (MBES). Classification techniques of single-beam echosounders are based on the “available acoustic backscatter data” (in absolute or relative units). Sidescan and multibeam instead present the backscatter intensity (also referred as amplitude) as raster from a series of pings. (Anderson 2007, 29.)

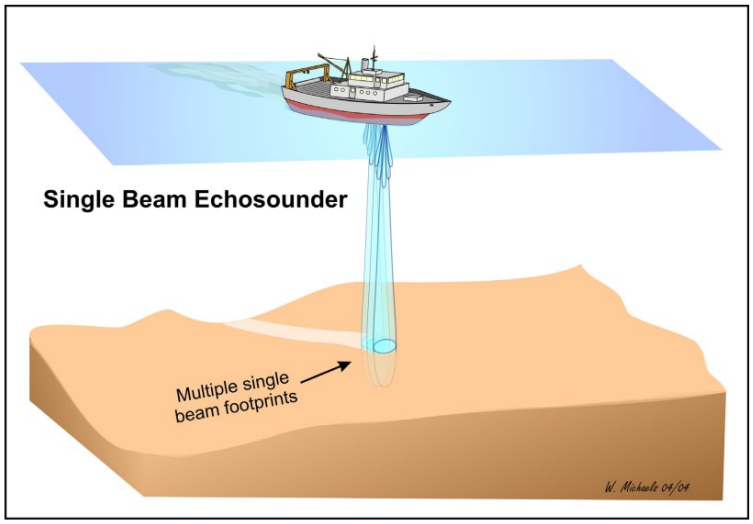


Figure 12. Single-beam Echosounders - SBES (Anderson 2007, by W. Michaels 04/04)

Several steps are performed to achieve seabed classification. They entail the generation, acquisition, and preparation of acoustic data. Following Anderson's (2007, 30) process, firstly an acoustic pulse is emitted towards the seabed either as a vertical beam or a narrow fan. To emit the pulses, a single downward-looking transducer is used in a single-beam echosounder (Figure 12), while the MBES is provided with an array of downward-looking transducers. SSS is instead equipped with a pair of sidewards-looking transducers. The sonar transducer produces "a pressure wave that propagates radially through the water" (Anderson 2007, 31-33) as sound can travel through various homogenous or inhomogeneous mediums, such as freshwater or saltwater. Constant sound speed and absorption are defined within a homogenous medium, leading to sound propagation along straight beams. These properties are influenced by temperature, salinity, and pressure. In oceanic environments, these factors notably change with depth, creating horizontal layers of sound speed, thus resulting in an inhomogeneous medium, while in coastal regions it is common for water temperatures to decrease with increasing depth. Also, the overall temperature and saline properties of the medium are relatively important, especially for SSS and MBES systems, as sound propagation occurs at least partially at substantial angles from the nadir, producing refraction. In SBES systems instead, refraction effects are almost "negligible" as the rays produced from the transducer are directed downwards and refracted nearly vertically, according to Snell's Law. Transducer motion (such as the ones that may occur on

vessels like pitch, roll, yaw, heave, and heading) aeration, turbulence, as well as high transmit power, produce relevant impacts on sound generation and reception. Thus, establishing proper reference values and recording measurements for sea state and the conditions in which the vessel is operating, while the sonar systems are active, are key features to decrease those adverse effects, or at least to account for, in the data processing phase at later stages. In this “controlled” environment, the acoustic data represent the sum of different features of seabed characteristics, including material composition, grain size, surface texture, and slope, in addition to depth. (Anderson 2007, 31-33.)

The pulse (Figure 13) then traverses through the water column, and interacts with the seabed, causing scattering. Assuming a flat substrate and the acoustic beam hitting directly, the amplitude and nature of the backscattered signal alter based on the type of seabed. Backscatter signal is significantly influenced by factors like surface roughness, slope (angle of beam incidence), and the area covered by the beam. (Anderson 2007, 31-33.)

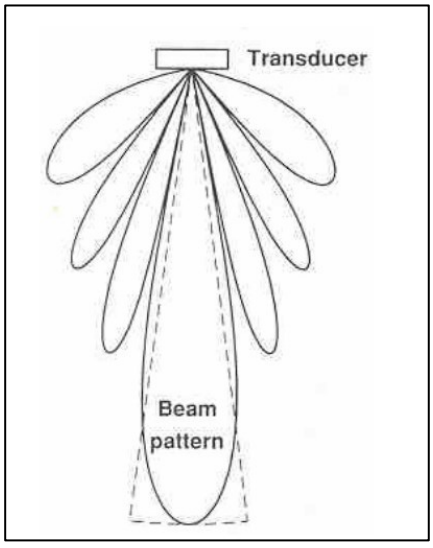


Figure 13. Beam Pattern of a transducer (Anderson 2007)

Subsequently, it retraces its path, as it is received back by the transducer and transformed into electrical signal. The last step involves data processing, which entails several sub steps, such as amplification, filtering, and compensating for non-seabed-related effects, which ensures “constant sensitivity for the observation of a specified target at any range”. (Anderson 2007, 31-33.)

Feature extraction is then performed, measured from single or series of adjacent echoes signals, depending on the sonar device utilized and resulting in a numerical representation of the corrected seabed echoes. In this phase, the main objective is to extract characteristics that will allow a relatively homogenous categorization of the classified bottom. Several properties of the bottom may influence the seabed echo, such as for example the water content, the roughness and hardness of the bottom, the benthic flora and fauna present in it, providing "an acoustic signature" that can be used to derive the bottom properties accordingly. High-resolution relief performed by multibeam sonars is also a reconstructed property of the sea bottom which is used to develop features used for classification. Moreover, additional information, such as depth, temperature, wind and waves, shore distance, sea mounts, reefs, and other bio-geographical characteristics of the area may be relevant to improve classification results and they are commonly referred as ancillary or additional features. (Anderson 2007, 40-41.)

Consistent data of the seabed, combined with phenomenological and statistical methods, are required to achieve proper acoustic seabed classification. Classification can be defined as "the segmentation of the whole dataset into homogenous subsets of objects" and can be divided into two main categories, supervised and unsupervised. The first, refers to human users setting *a priori* classes, "supervising" the classification process. On the other hand, the unsupervised classification occurs when algorithms provide the segmentation before knowing any defined classes. While the results of the first approach are linked to the training dataset to determine the classes beforehand, the results of the unsupervised approach are related to the difference within the dataset and variability in the sea bottoms, which in turn is a function on how wide the surveyed area is. (Anderson 2007, 66-67.)

Scientific echosounder has been used for studying and classify the bottom of the sea combined with software (Siwabessy et al. 2000; Fajaryanti & Kang 2019; Fauziyah et al. 2020). Siwabessy et al. (2000) collected data of the Southeast continental shelf of Australia, using a stand-alone Simrad EK500 scientific

echosounder in combination with statistical tools to classify bottom features, including Principal Component Analysis (PCA) and Cluster Analysis (CA). Fajaryanti and Kang (2019) conducted a preliminary study of bottom classification with the combination of a split-beam echosounder (Simrad EK60 sonar system), Echoview, and sediment sampling. They investigated the seabed along the coast of Tongyeong, South Korea, and after collecting sonar data from surveys in the area, they derived values of E1 (bottom roughness) and E2 (bottom hardness) for three different soil types by using Principal Component Analysis and K-mean clustering provided by Echoview bottom classification module, and then validated the results with ground-truthing sampling. Fauziyah et al. (2020) analyzed the bottom substrate types in the east of the Banyuasin waters, in Indonesia, using a single-beam echosounder (Simrad EK15) together with Echoview, and a Ekman grab for sediment collection. In the conducted analysis, E1 and backscatter strength were first obtained, and then associated with the type of soil in each station sampled.

2.3 Remotely operated vehicles (ROVs)

Remotely operated vehicles (ROVs) are underwater robotic devices, controlled by an operator on surface with a controlling board, connected to the ROV via cable (also referred as umbilical), which also provides power to it (Ludvigsen et al. 2007, 141; Sørensen et al. 2020, 4). The capability of the ROVs fits into the overall strategy, with continuous development in technology, of exploration of extreme environments, such as lower depth in the seas and oceans (Sørensen et al. 2020, 4-5). They can be used to perform a series of tasks underwater, such as collection of physical, sediment, and water samples, as well as recording of images and videos (Nevstad 2022, 4). The connection by umbilical allows “almost unlimited power and high bandwidth communication”; however, it also constrains its available “spatial coverage”, limited by the length and durability of the cable (Sørensen et al. 2020, 4). Also, ROVs are susceptible to currents and waves, which increase the difficulty of operating it (Sørensen et al. 2020, 4), especially in challenging environments, such as nearby caves, slopy terrains, and ridges. Operators are required to perform the piloting with extra care, for both avoiding the loss or entanglement of the ROV, to produce proper sample of the study

area, and avoiding any extra stress posed to the analyzed environment, for example by mistakenly touching and modifying its surroundings (internal communication at Archipelagos 2023).

2.4 Citizen science for marine species and habitats

In the Mediterranean Sea, citizen science programs have been used for species and habitats identification and distribution (Vlachopoulou et al. 2013; Giovos et al. 2019; Naasan Aga Spyridopoulou et al. 2020; Gatti et al. 2022; Pietroluongo et al. 2022). Citizen science, defined as a form of “open collaboration” in which single persons can actively participate as volunteers in the scientific process (Bonney et al. 2021, 519), can represent a way to furthering the knowledge on a specific topic, area, habitat, helping to add relevant information about fisheries, their management, and the ecosystem, above and underwater, they operate in. The collaboration between fishermen, divers, and other inhabitants of coastal areas with scientists by using citizen science projects has the potential to become a key milestone, not only for knowledge gathering, but also for the involvement of local and indigenous communities in the decision-making processes (Bonney et al. 2021, 519) and acting also as a motor towards reaching UN Sustainable Development Goals (Fritz et al 2016, 922).

In the area of interest, analyses of habitat distribution of Posidonia Oceanica seagrass meadows, along the coast of Samos and the Fourni Islands in Greece, were conducted by Vlachopoulou et al. (2013), and with the help of Archipelagos, by leveraging citizen science data. In the research, semi-structured interviews were conducted with fishermen and other stakeholder groups. Moreover, GIS workshops with public participation were organized involving fishermen from both islands. They were asked to highlight on maps the locations of Posidonia Oceanica meadows, trawling routes, current and depleted fish stocks, and other details regarding local fishing activities (Figure 14).

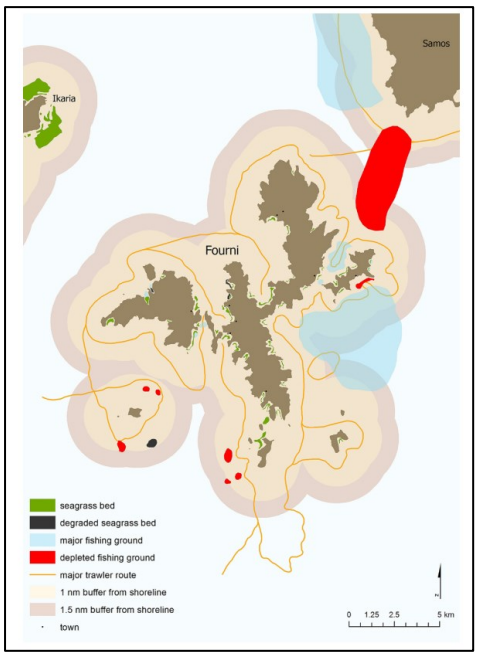


Figure 14. Map of Fourni, Greece, with citizen science GIS workshop (Vlachopoulou et al. 2013)

Other projects involving citizen science have occurred in the Mediterranean and are currently ongoing. The project named “Is it Alien to you? Share it!!!”, developed by the Environmental Organization iSea in 2016, sets a clear example on how, in Greece, species identification and distribution was enhanced by a citizen science platform. Giovos et al. (2019) used the same project with the main goal to demonstrate how its available data could contribute to furthering knowledge of species.

Naasan Aga Spyrodopoulou et al. (2020), instead, made use of the recordings from “Is it Alien to You? Share it!!!” project, specifically by analyzing citizen science information about only five rare species in Greece, native of the Aegean and Levantine Sea. Citizen science was also employed to gather information about vulnerable species, such as the Mediterranean monk seal (*Monachus Monachus*). For example, in the research conducted by Pietroluongo et al. (2022), monitoring data about living and stranded seals were also obtained from local and port authorities, tourists, and local communities in the island of Samos, Greece, to increase the understanding of the species status. Regarding coralligenous habitat distribution, Gatti et al. (UNEP-MAP-SPA/RAC 2022) at the *4th Mediterranean Symposium on the conservation of Coralligenous & other Calcareous Bio-Concretions*, presented results of a program named

“CIGESMED”, aimed at divers. The program run between 2016 and 2021, and thanks to the help of local communities facilitating the task and considering the different levels of skillsets and knowledge of divers on the subject matter, 150 observations were gathered by volunteer divers around the area of Calanque National Park, in France. These recordings allowed researchers to define the abundance of the main coralligenous taxa, the pressures sustained by their habitats, and their variation through time.

3 MATERIAL AND METHODOLOGY

3.1 Study area

The study area (Figure 15) is the waters around the coasts of Fourni, a group of Greek islands located on the east side of the Aegean Sea, part of the North Aegean administrative area, between the islands of Ikaria and Samos, and comprising 17 islets (Christodoulakis et al. 2001).

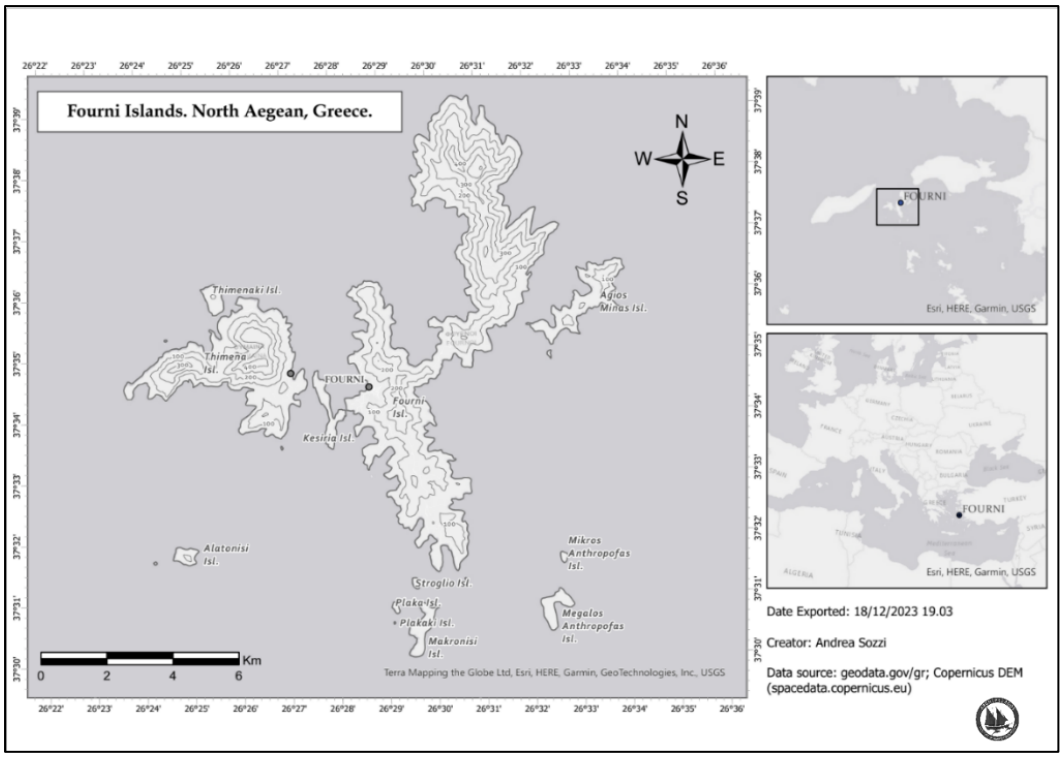


Figure 15. The area of interest (AOI) of the present study

The surveyed transects are mainly along the coastline of the main island, named Fourni Korseon, and for a short tract, also along the south and east coastline of

Agios Minas, south of Makronisi, west of Plakaki, Plaka, and Stroglio, east of Thimena, and west of Kesiria.

3.2 Single-beam echosounder (SBES): Simrad EK80 sonar system

Single-beam echosounder has previously been used for gathering data for classification of seabed (Siwabessy et al. 1999; McCauley & Siwabessy 2006; Fajaryanti & Kang 2019; Fauziyah et al. 2020). The Simrad EK80 echosounder by Kongsberg was the system used for the data collection onboard of one of the commissioner vessels, the *Aegean Explorer*, Archipelagos' research boat mainly used for GIS work and mapping surveys. The sonar system is composed of three main units: the transducer, attached to a pole and lying underwater, at 3 meters below the sea surface, which sends the ping signal down the water column, receives it back, and transforms it into an electrical signal. A transceiver, the signal processor, in the GIS room of the boat, connected to the transducer on the one side, and on the other side connected to the third part, a processing unit (Figure 16).

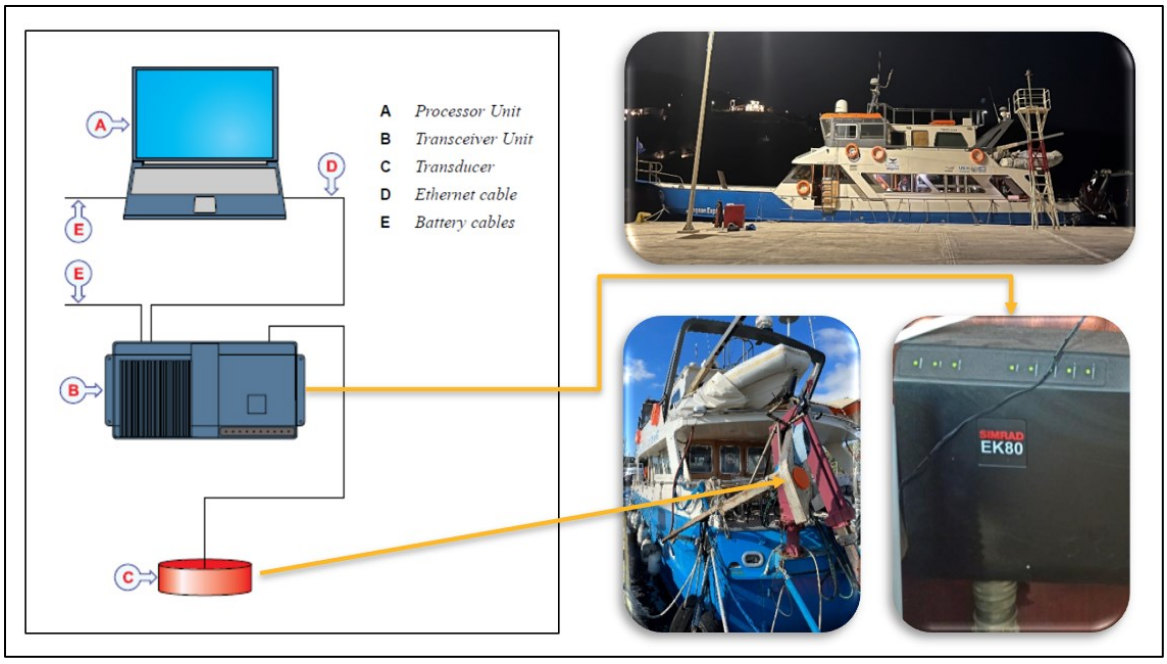


Figure 16. Sonar schema, and sonar parts on board of the *Aegean Explorer* (Kongsberg 2023)

The specification and settings of the transducer on board of the *Aegean Explorer*, as provided also in Fajaryanti and Kang (2019, 41), are shown in Table 1.

Table 1. Transducer specification

Parameter	Value
Operating frequency (kHz)	120.00
Sound speed (m/s)	1528.95
Transmission power (W)	250.00
Transmitted pulse length (m/s)	0.256
Equivalent beam angle (dB re 1 Steradian)	-20.70
Absorption coefficient (dB/m)	0.0450

The operating frequency was set at 120 kHz, because previous surveys done by Archipelagos in the research area demonstrated its usefulness at identifying mesobenthic habitats (Archipelagos 2022).

3.3 Echoview software

Echoview is a licensed software package for hydroacoustic data processing. It provides robust and versatile capabilities for analyses of water column, bathymetric echosounder, and sonar data. It is recognized globally as a tool used by fisheries and environmental scientists for monitoring and managing marine and freshwater environments. The software is divided into several packages and modules, which can be added afterwards as expansion packs, on top of the base product. (Echoview 2023a.)

During the study, Echoview license was provided by the commissioner, and it included the baseline of the software with some base operators, and one extra module, named *Bottom Classification*, to perform operations for classification of the seabed from the collect .raw data, and to obtain several features of the bottom as result of the classification.

3.3.1 Depth normalization and acoustic data analysis

Seabed features are dependent on echo shapes and energy. The echoes from the sea bottom contain “encoded time and energy information” from a single transmitted pulse interacting with the seabed substrate. When the pulse is transmitted at a normal incidence, it travels $c\tau/2$, where c is the speed of sound in

water (m/s) and τ is the transmitted pulse duration of the transducer (s). However, in the off-axis region of the beam, the pulse covers a distance greater than $c\tau/2$. Consequently, the seabed echo spreads over this effective pulse length (Penrose et al. 2005; Echoview 2023b). Features of the bottom are normalized based on a normalization reference depth, which is used to normalize pulse length. Referring to Echoview (2023b), the best value for reference depth is the average depth of the bottom. (Fajaryanti & Kang 2019, 41-42.)

The effective pulse length, defined as "the total distance that the off-axis part of the beam needs to travel one whole pulse length in the substrate" (Fajaryanti & Kang 2019, 41; Echoview 2023c) is equivalent to:

$$Effective\ pulse\ length = \left[d_0\ + \frac{c\tau}{2}\right]\ - d_0\ cos\left(\frac{\theta}{2}\right) \quad (1)$$

Where:

c	the speed of sound in water (m/s).
τ	the transmitted pulse duration of the transducer (s).
θ	the major-axis 3 dB beam angle of the transducer.
d_0	normal incidence start depth (m) of the first echo at a ping P.

And defining off axis pulse length in the equations below (2) (3):

$$OffAxisPulseLength_{Ref} = \frac{c\tau}{2}\ +\ Reference_{Depth}\ -Reference_{Depth}\ cos\left(\frac{\theta}{2}\right) \quad (2)$$

$$OffAxisPulseLength_{Act} = \frac{c\tau}{2}\ +\ Actual_{Depth}\ -Actual_{Depth}\ cos\left(\frac{\theta}{2}\right) \quad (3)$$

Where:

$OffAxisPulseLength_{Ref}$	the pulse length off-axis where the normal incidence start depth is specified by the *Depth normalization reference depth*.

$Reference_{Depth}$ the normalization reference depth.

$OffAxisPulseLength_{Act}$ the pulse length off-axis of the first bottom echo, where the normal incidence start depth of the first bottom echo is given by $Actual_{Depth}$ for a certain ping.

$Actual_{Depth}$ Depth of the Bottom line (at normal incidence) for the first bottom echo at a certain ping.

c the speed of sound in water (m/s).

τ the transmitted pulse duration of the transducer (s).

θ the major-axis 3 dB beam angle of the transducer.

And from (2) and (3), deriving then the normalization of the depth equation for each feature (4):

$$Value_{DepthNorm} = \left(\frac{OffAxisPulseLength_{Ref}}{OffAxisPulseLength_{Act}}\right) Value_{Act} \quad (4)$$

Where:

$Value_{DepthNorm}$ the value of a feature of the bottom, normalized by depth.

$Value_{Actual}$ the value of the bottom feature.

$OffAxisPulseLength_{Ref}$ the pulse length off-axis where the normal incidence start depth is specified by the *Depth normalization reference depth*.

$OffAxisPulseLength_{Act}$ the pulse length off-axis of the first bottom echo, where the normal incidence start depth of the first bottom echo is given by $Actual_{Depth}$ for a certain ping.

In the case of single-beam echosounders, the energy emitted by seabed reflections is quantified across distinct time intervals once the echosounder leading edge of the pulse touches the seabed. Initially, the amplitude of the seafloor echo reaches its highest peak, due to a coherent reflection occurring near a perpendicular angle. This often produces saturation in the majority of echosounders. Subsequently, it diminishes to a certain level at a rate which is contingent on the beam pattern and characteristics of the seabed. At the same time, the pulse is progressively encompassing a larger annulus area at oblique angles relative to the beam's central axis (Figure 17). The characteristics of the seabed echo, specifically its energy distribution and the duration of the tail following the initial peak at orthogonal angle, are frequently employed to define seabed roughness resulting from both surface scattering and volume backscattering, also commonly defined as E1 feature. (Anderson 2007, 48.)

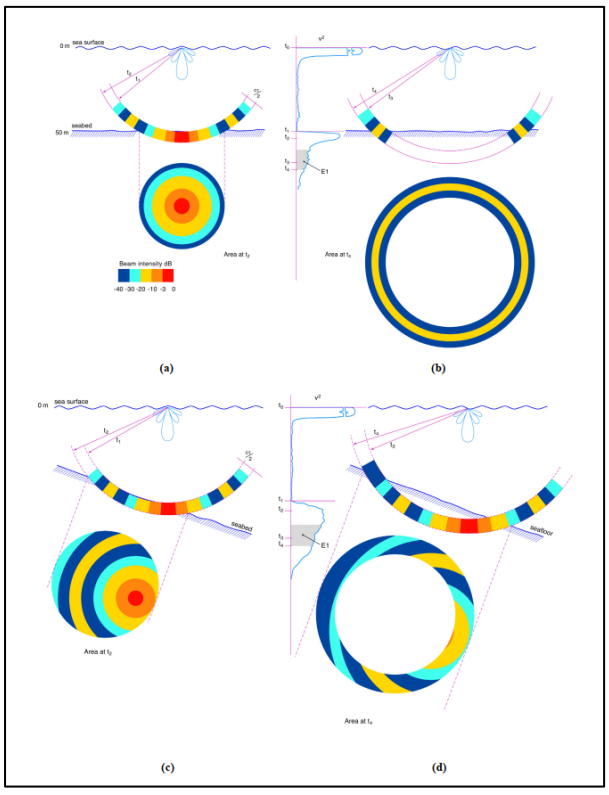


Figure 17. SBES insonification area as function of time, on different inclination of the seabed (Anderson 2007)

E1 “is derived from an integration of the tail of the first acoustic bottom return” (Siwabessy et al. 1999, 2; McCauley & Siwabessy 2006, 13-14; Fajaryanti & Kang 2019, 42). In cases of echosounders not saturating completely the bottom

echo, the energy content can be gauged from the initial bottom echo. However, often, it is approximated using the second bottom return after the initial reflection off the sea surface and back to the sea bottom (E2 feature), which represents the hardness of the sea bottom (Figure 18). (Siwabessy et al. 1999, 2; Anderson 2007, 48-49.) Roughness on the seabed surface causes scattering and reflection of the sound waves, generating the energy of the tail of the first acoustic bottom return (E1), while impedance mismatch between the seabed and the water column (the difference, causes some energy to be reflected), results in the hardness of the seafloor (E2 component) (Siwabessy et al. 1999, 2).

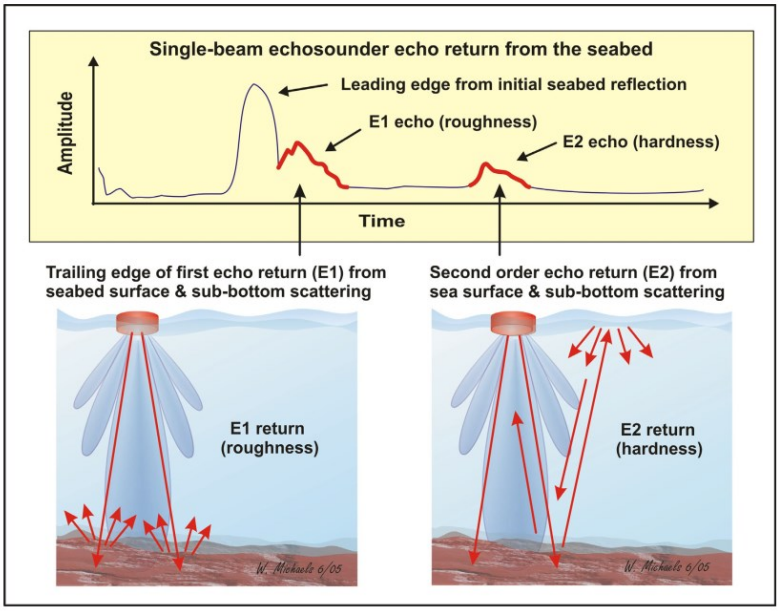


Figure 18. Example of single-beam echosounder echo returns E1, and E2 (Anderson 2007)

Acoustically different bottom of the sea types can be determined by clustering the backscatter signals by E1 and E2 parameters. Siwabessy et al. (1999, 2) provide the equation below (5), for determining E1 and E2 values:

$$E = 4\pi(1852)^2 \frac{\sum_{l=1}^{p}\left(\sum_{k=1}^{d} s_{v(k,l)}\right)}{p} \quad (5)$$

Where:

d	number of s_v values within the tail of the first acoustic bottom return (roughness), and that within the complete second acoustic bottom return (hardness).
S_v	is the backscattering coefficient.
p	number of pings within a horizontal interval of 0,05 nmi.

A bottom feature consists of a measured characteristic of the first and second bottom echoes (E1 and E2), and it is computed within a specific feature extraction interval, which uses an average of similar echoes for more stability in the data for classification. Indeed, each feature is considered as “a mean of the measured ping characteristics in the feature interval” (Fajaryanti & Kang 2019, 42; Echoview 2023b). Different factors may influence the feature interval, such as ping rate, ping footprint (considering beam width and seabed depth), and the speed of the vessel where the sonar is located. (Echoview 2023b.)

Table 2. Dimensions and features of the bottom, provided by Echoview (2023i)

Type	**Name**	**Specification**
Dimension	Point index	ID number
Dimension	Point date	Date
Dimension	Point time	Time
Dimension	Point time in milliseconds	Time (ms)
Dimension	Latitude	GPS coord.
Dimension	Longitude	GPS coord.
Dimension	Depth	Distance (m)
Feature	Bottom roughness (normalized)	Backscat. strength (Sv) (dB re 1m2/m3) – from E1
Feature	Bottom hardness (normalized)	Backscat. strength (Sv) (dB re 1m2/m3) – from E2
Feature	First bottom length (normalized)	Distance (m) – from E1
Feature	Second bottom length (normalized)	Distance (m) – from E2
Feature	Bottom rise time (normalized)	Distance (m) – from E1
Feature	Bottom line depth mean (normalized)	Distance (m) – from E1
Feature	Bottom max Sv (normalized)	Backscat. strength (Sv) (dB re 1m2/m3) – from E1

In this study, the default value of the ping rate was set to ten pings (ten emitted sonar pulses) and used as the feature extraction interval. Several dimensions and features are provided by the software, other than bottom roughness and hardness of the bottom in equation 5 (Table 2).

First bottom length normalized (dB re 1m2/m3) is defined as “the mean of the depth normalized length of the first bottom echo, in the feature extraction interval for a bottom classification”, summarized by the equation below (6), provided by Echoview (2023d):

$$First\ bottom\ length\ norm_j = \frac{1}{n}\sum_{i=1}^{n} First\ bottom\ length_{DepthNorm,i,j} \quad (6)$$

Where:

j	bottom point in a bottom points variable, assigned sequentially in time. Pings of the echogram are partitioned according to the number of pings in the specific feature extraction interval.
i	ping *i* in the feature extraction interval *j*.
First bottom length$_{DeptNorm,i,j}$	Depth normal. length of first bottom echo in ping *i* in the feature extraction interval *j* and based on (4) with the actual value of the feature (Bottom line depth to the end of the first bottom echo at ping *i* in Feature extraction interval *j*).
n	last ping in the feature extraction interval *j*.

Bottom rise time normalized is defined as “the mean depth of the bottom line rise time of the first bottom echo, in the feature extraction interval for a bottom classification”, summarized by the equation below (7), provided by Echoview (2023e):

$$Bottom\ rise\ time\ norm_j = \frac{1}{n}\sum_{i=1}^{n} Bottom\ rise\ time_{DepthNorm,i,j} \quad (7)$$

Where:

j	bottom point in a bottom points variable, assigned sequentially in time. Pings of the echogram are partitioned according to the number of pings in the specific feature extraction interval.
i	ping *i* in the feature extraction interval *j*.
Bottom rise time$_{DeptNorm,i,j}$	Depth normal. rise time of first bottom echo in ping *i* in the feature extraction interval *j* and based on (4) with the actual value of the feature (Bottom line sample peak of the first bottom echo at ping *i* in Feature extraction interval *j*).
n	last ping in the feature extraction interval *j*.

Bottom line depth mean is defined as “the mean depth of the bottom line in the feature extraction interval for a bottom classification” and summarized by the equation below (8), provided by Echoview (2023f):

$$Bottom\ line\ depth\ mean_j = \frac{1}{n}\sum_{i=1}^{n} Bottom\ line\ depth_i \quad (8)$$

Where:

j	bottom point in a bottom points variable, assigned sequentially in time. Pings of the echogram are partitioned according to the number of pings in the specific feature extraction interval.
i	ping *i* in the feature extraction interval *j*.

Bottom line depth$_i$	Depth of the bottom line at ping *i* in the feature extraction interval *j*.
n	number of pings as specified by the feature extraction interval *j*.

Second bottom length normalized (dB re 1m2/m3) is defined as “the mean of the depth normalized length of the second bottom echo in the feature extraction interval for a bottom classification”, summarized by the equation below (9), provided by Echoview (2023g):

$$Second\ bottom\ length\ norm_j = \frac{1}{n}\sum_{i=1}^{n} Second\ bottom\ length_{DepthNorm,i,j} \quad (9)$$

Where:

j	bottom point in a bottom points variable, assigned sequentially in time. Pings of the echogram are partitioned according to the number of pings in the specific feature extraction interval.
i	ping *i* in the feature extraction interval *j*.
Second bottom length$_{DeptNorm,i,j}$	Depth normal. length of first bottom echo in ping *i* in the feature extraction interval *j* and based on (4) with the actual value of the feature (From the beginning of the second bottom echo to the end of the second bottom echo at ping *i* in the feature extraction interval *j*).
n	last ping in the feature extraction interval *j*.

Bottom max Sv (dB re 1m2/m3) is defined as “the mean maximum Sv (dB) of the first bottom echo in the feature extraction interval for a bottom classification”, summarized by the equation below (8), provided by Echoview (2023h):

$$Bottom\ max\ Sv_j = \frac{1}{n}\sum_{i=1}^{n} maximum\ Sv_i \quad (7)$$

Where:

j	bottom point in a bottom points variable, assigned sequentially in time. Pings of the echogram are partitioned according to the number of pings in the specific feature extraction interval.
i	ping *i* in the feature extraction interval *j*.
maximum Sv_i	Maximum Sv (dB) of the first bottom echo in the ping *i* in the feature extraction interval *j*.
n	number of pings as specified by the feature extraction interval *j*.

3.3.2 Principal component analysis and k-mean clustering

Echoview bottom classification module is based on Principal Component Analysis (PCA) and K-mean clustering. PCA is used for classification of the seabed, when multiple features are involved and can be summarized into fewer perpendicular components, which provide an explanation of the decreasing proportion of the total variance of the whole data (Siwabessy et al. 1999, 4; Anderson 2007, 68-69; Echoview 2023i). There is no strict number amount of component needed for the classification, however it is paramount that overall, they would represent a major part of the variance (Anderson 2007, 68-69). Moreover, in case highly correlated attributes are available, only one of them should be implemented in the PCA. Principal Components scores are mapped so that the specific aspect of each component of the seabed echo properties is spread out in space, and through classification, interpretable patterns may be found within the dataset, usually by also visualizing the scores with color coding. PCA analysis hence provides a first

spatial structure of seabed variance, with unsupervised classification, as bottom features represent dimensions of the dataset (Anderson 2007, 69; Echoview 2023i).

Once PCA is performed, different clustering methods can be implemented to identify homogenous subcategories within the whole dataset. K-mean partitioning, which differentiates all dataset into K non-overlapping subsets is often chosen for seabed classification. The optimal number of divisions into subgroups is performed either by a user with knowledge assisted by guiding statistics or according to statistical methods. (Anderson 2007, 70.)

Upon running and completion of the bottom classification module in the Echoview software, the obtained unsupervised classification provided Class ID and Class Name dimensions, together with several dimensions and the normalized features of the bottom, as shown in Table 2.

3.4 Random forest classifier

Random Forest (RF) classifier is a type of machine learning algorithm based on multiple decision trees (forest) for the classification of a dataset, which is based on vote majority (Figure 21).

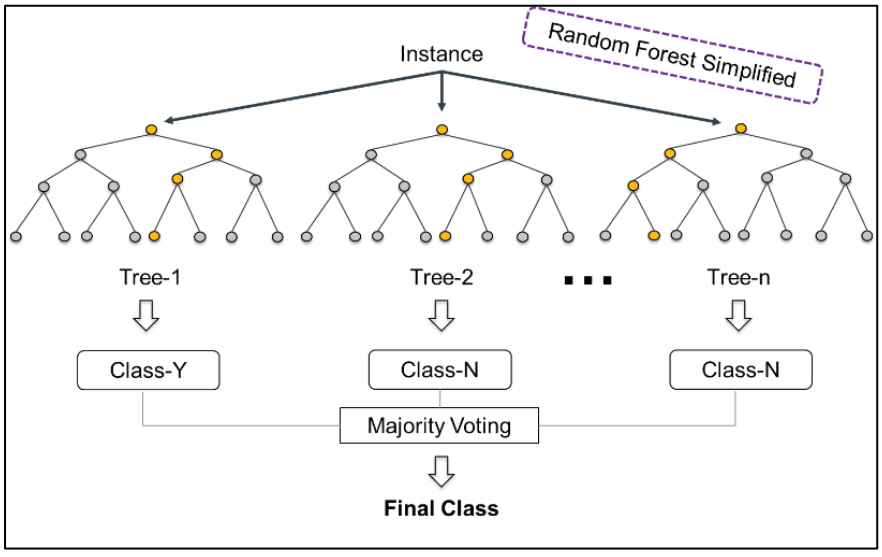


Figure 19. Random forest classifier schema

Because of this process, the potential risk occurring on a single tree of overfitting is reduced by using multiple trees of random variables. Moreover, RF fits appropriately for the presence / non-presence type of classification, and for habitat suitability modeling. (Pearman et al. 2020, 6; Jackson-Bué et al. 2022, 5; Fakiris et al. 2023, 6.) Each decision tree in a RF starts with a root node, which represents the starting point of the tree and provides information about the variables (features) for splitting the data, and it helps understanding how many variables (N) of the total (K) will be evaluated at each split, based on the N = log2 K +1 (Fakiris et al. 2023, 7). The sum of all the levels needed to reach a final decision in the decision tree, reaching a leaf node, represents the depth of the tree.

In the current study, the RF model is developed in Python environment using the Scikit-Learn library. The model default settings were used for its implementation, as provided by the Scikit-Learn library guidelines (Scikit-Learn 2023).

3.5 ROV data from the study area

ROV dives were performed during prior surveys by Archipelagos throughout the period 2021-2023. Five specific dives were performed in the area of interest around the Fourni Islands in 2021. One dive occurred close to the east tip of Agios Minas island (37.5941N, 26.5716E) on the 8th of September 2021; a second and a third, on a seamount named by Archipelagos as Grandjean-Foster (37.5492N, 26.5321E) on the 9th of September 2021; a fourth one in the north of Makronisi island (37.5183N, 26.505E) on the 16th of September 2021; a fifth one on the west of Alatonisi island (37.52516N, 26.399902E) on the 18th of September 2021.

Two more ROV dives were then performed during a two-day boat survey in the summer of 2023 (between the 3rd and 4th of August 2023), the first one again in the surrounding of the east tip of Agios Minas, and the second one in the surrounding of the Grandjean-Foster seamount. (Internal communication at Archipelagos 2023.)

All dives confirmed the presence of coralligenous distribution (Figure 20), amongst other marine habitats (internal communication at Archipelagos 2023). The analysis conducted by Archipelagos about the depth distribution of the habitat, based on ROV data, together with the location of the dives, and footage confirming the presence of the outcrops will be used as complementary information in the analysis of the data collected in the transects of the current study. Regarding the mapped location of the ROV dives as shapefile in ArcGIS Pro, a buffer zone of a 100-meter radius was created from the GPS coordinates of each dive (Figure 21).

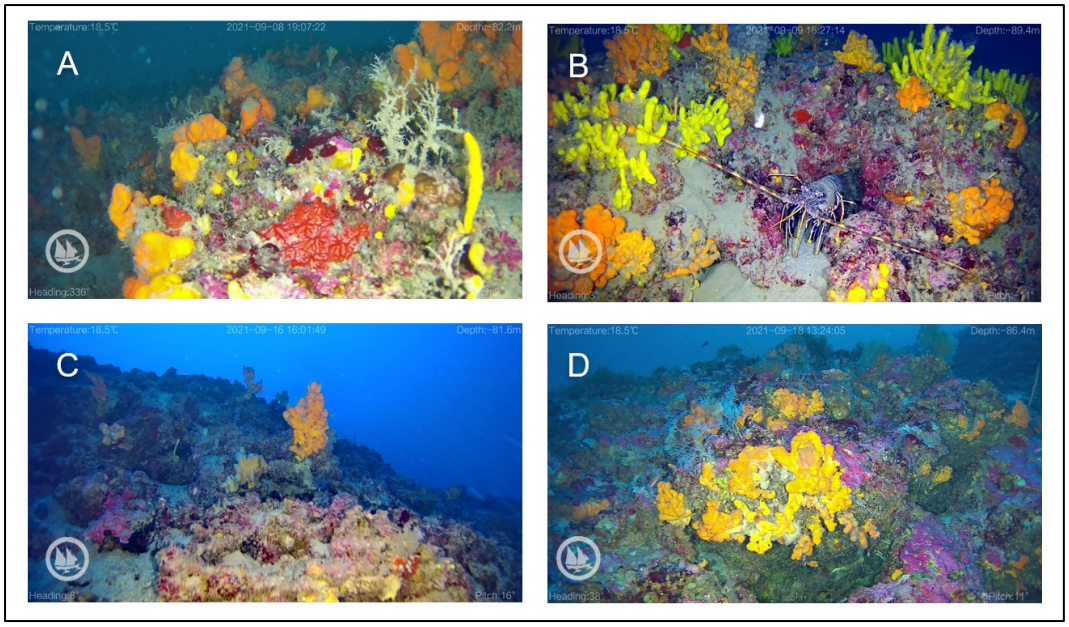


Figure 20. ROV dives image examples from expedition during September 2021 in the area of study of the Fourni Islands. A. East tip of Agios Minas (82.2-meter depth). B. Seamount Grandjean-Foster (89.4-meter depth). C. North Makronisi (81.6-meter depth). D. West Alatonisi (86.4-meter depth) (internal communication at Archipelagos 2023)

3.6 Citizen science data from the study area

Archipelagos Institute of Marine Conservation, the commissioner of the study, throughout years of operation in the waters and coasts of Samos, and around the area of interest of the Fourni Islands, has conducted activities of citizen science with the island's local communities. Specifically, during public participation with the GIS team for mapping marine habitats, trawling areas, and species presence, fisheries and local experts were listened to also for coralligenous distribution in the area of interest. The identified zones by the stakeholders were first drawn on

paper maps, then digitalized and georeferenced in GIS software, and afterward, their equivalent was created as shapefile (.shp). The available data are from 2012, 2022, and 2023 (Figure 21). The habitat distribution of 2023 represents an upgrade of the previously defined zones during the GIS public participation conducted in the year 2012. (internal communication at Archipelagos 2023.)

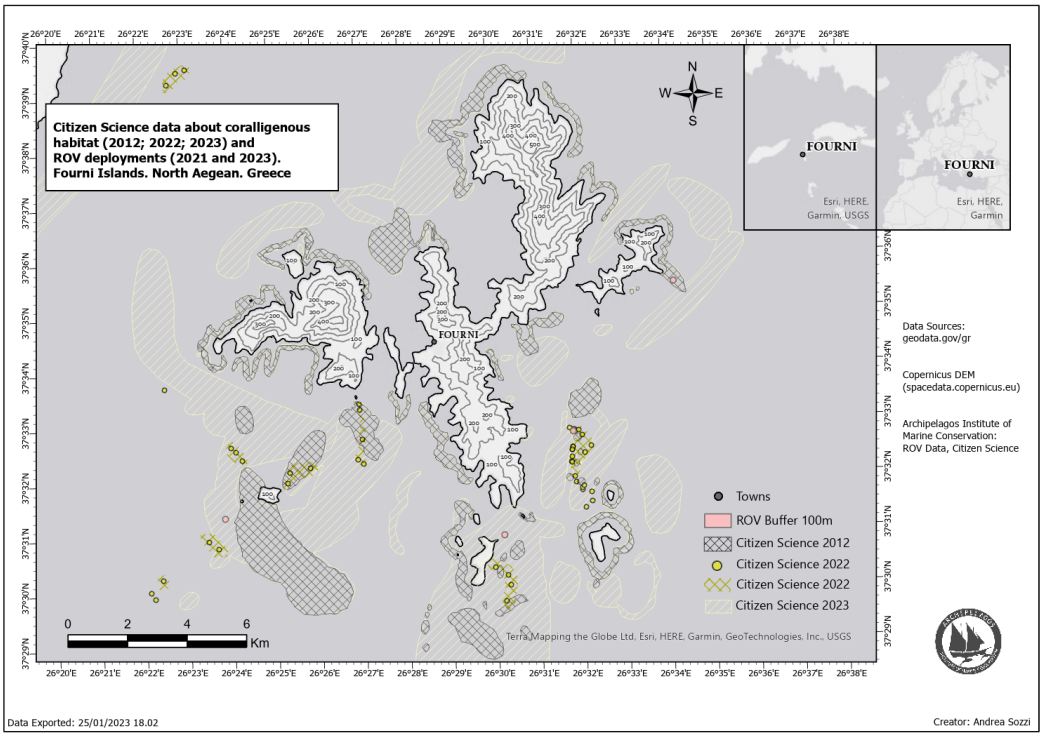


Figure 21. Citizen science data about coralligenous habitats years 2012, 2022, and 2023, and ROV deployments occurred in 2021 and 2023 in the AOI

3.7 Data collection

The data was collected on a two-day boat survey, which took place on the 25th and 26th of September 2023 on the Archipelagos research boat, the *Aegean Explorer* (Figure 22).

On the first day (survey ID F_ESW_25092023), data from east, south, and west were gathered (ESW dataset), followed on the second day (survey ID F_N_26092023) by surveying the northern part of the Fourni coastline (N dataset), for a total of 295Gb (195Gb on the first day, 100Gb on the second day). The ESW survey length amounted to 54.36 kilometers, while the N one to 34.78 kilometers.

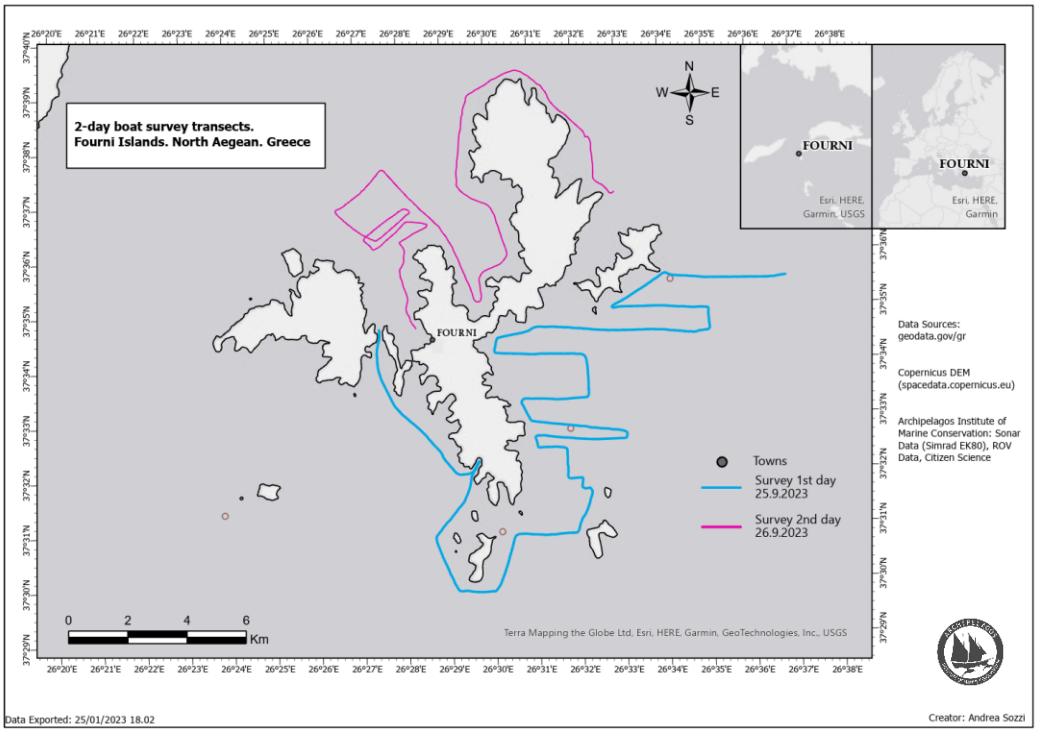


Figure 22. Two-day survey transects in the area of interest

3.8 Data processing and analysis

The echosounder data collected during the 25th and 26th of September required processing with specific software, Echoview, provided by Archipelagos, which can read Simrad EK80 .raw files, visualize them with echograms, and save them into a proprietary format *.ev*. Processing of the data was divided into five distinct phases (Figure 23).

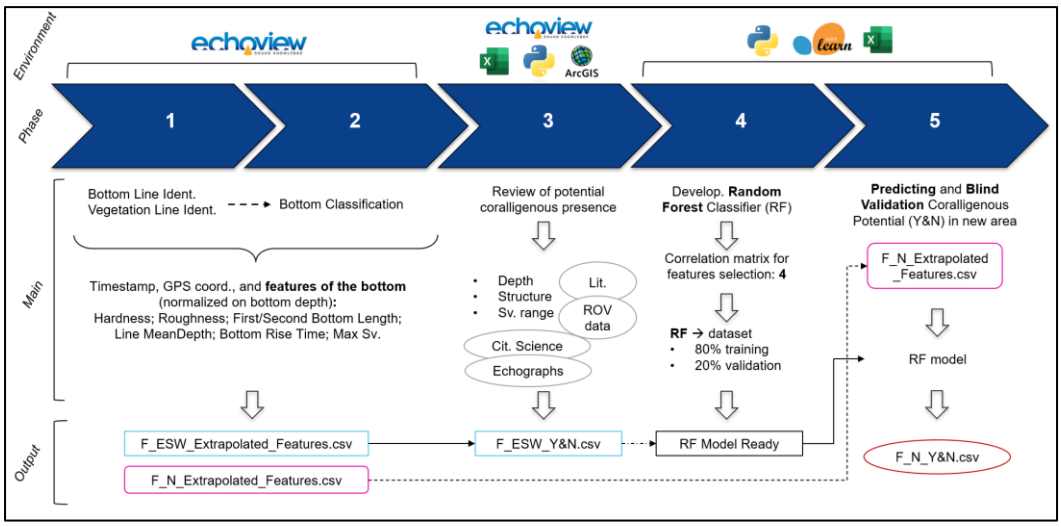


Figure 23. Diagram of the five phases of the process

The first two phases were applied to the data collected on both surveys, during the first and second days, for both ESW and N datasets. Firstly, bottom and vegetation lines were identified in Echoview on the volume backscattering strength echogram (sv1) (Figure 24). Bottom line was identified by following the methodology of Fajaryanti and Kang (2019, 41), firstly by using the proprietary algorithm within Echoview, “Best bottom candidate” using a 0.0 back-step range, start depth at 5 meter, and minimum backscatter strength for good pick at -70.00 db. Subsequently, the line was edited directly in the echogram (sv), ensuring continuity of the whole line across the whole analyzed transect, in both ESW and N datasets. Identifying the bottom line was a key action because, on the one hand, it allowed to properly distinguish between the water column and seabed, and on the other hand, Echoview mandates to have the bottom line derived, before the utilization of the bottom classification module.

A vegetation line was then added (Figure 24), to provide further information to the software, before the start of the bottom classification. According to Echoview (2023j), a threshold offset operator can be used for adding a vegetation line into the echogram. Specifically, based on literature values of where rim structures are usually located, 0.3 meters from the bottom was the minimum accepted distance, and, following a manual analysis of the echogram, the threshold of backscatter strength (sv), averaging backscatter strength values within the echogram directly,

and where potential biomass was present on top of the seabed, was set up to -55db.

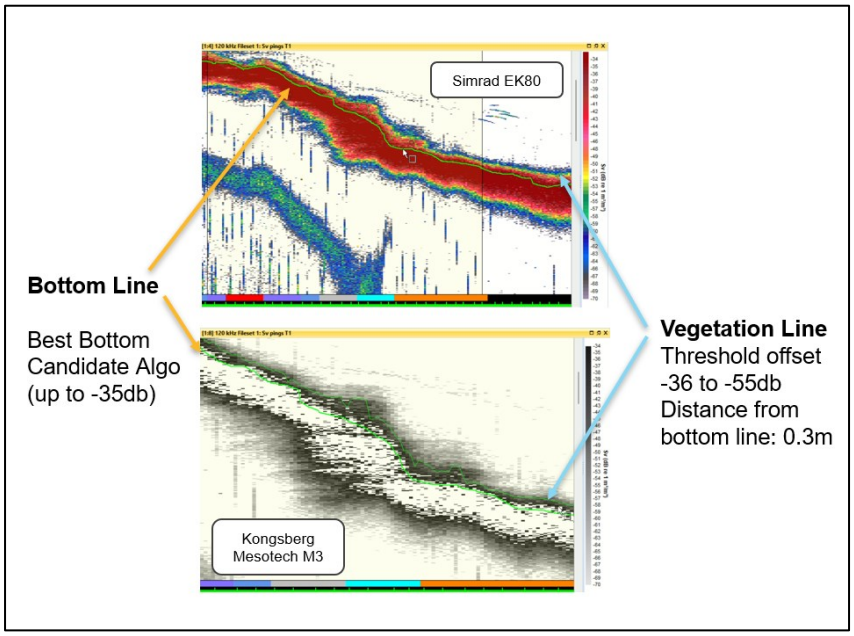


Figure 24. Echograms examples with bottom and vegetation lines

Secondly, running the bottom classification operator in the software, which uses Principal Component Analysis and k-mean clustering (Fajaryanti & Kang 2019; Echoview 2023i), generated a new file per survey (25th of September - File ID F_ESW_Extrapolated_Features – 6413 datapoints; 26th of September - File ID F_N_Extrapolated_Features – 4382 datapoints), with normalized features of the bottom, from first (E1) and second (E2) echoes (Siwabessy et al. 1999; McCauley & Siwabessy 2006). The generated features: *bottom roughness*, *bottom hardness*, *first* and *second bottom length*, *bottom rise time*, *bottom maximum backscattering strength,* and *bottom line mean depth*. Moreover, the main dimensions of the echograms, such as *Ping ID*, *Timestamp*, and *GPS coordinates*, were also provided, enabling a series of analyses based on time and spatial positioning.

During the third phase, the potential presence of coralligenous formation was reviewed and added to the newly created file for the ESW dataset, (F_ESW_Extrapolated_Features) only. To perform this phase, a combination of actions in Echoview, Excel, Python, and ArcGIS Pro was required. The

parameters found in the literature on the typical geomorphology, structure, and depth of coralligenous formations, combined with the visual representation of the volume backscattering strength echograms (Figure 25 and 26), and the values obtained by Archipelagos internal research for structure and habitat depth distribution were used to review the potential presence of the formations.

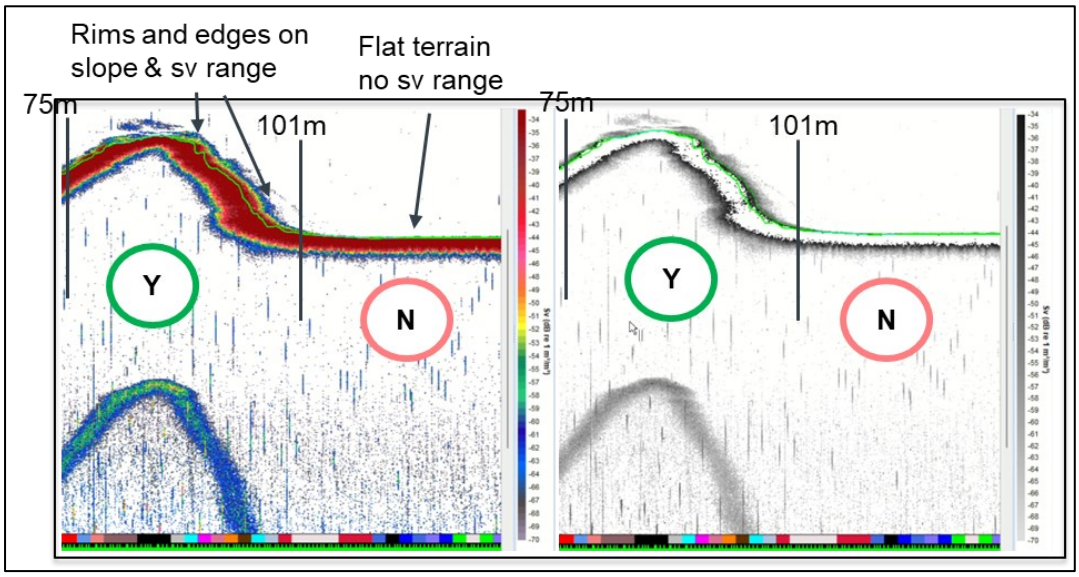


Figure 25. Phase 3 example of the reviewing of potential coralligenous presence on same echogram but different color palette

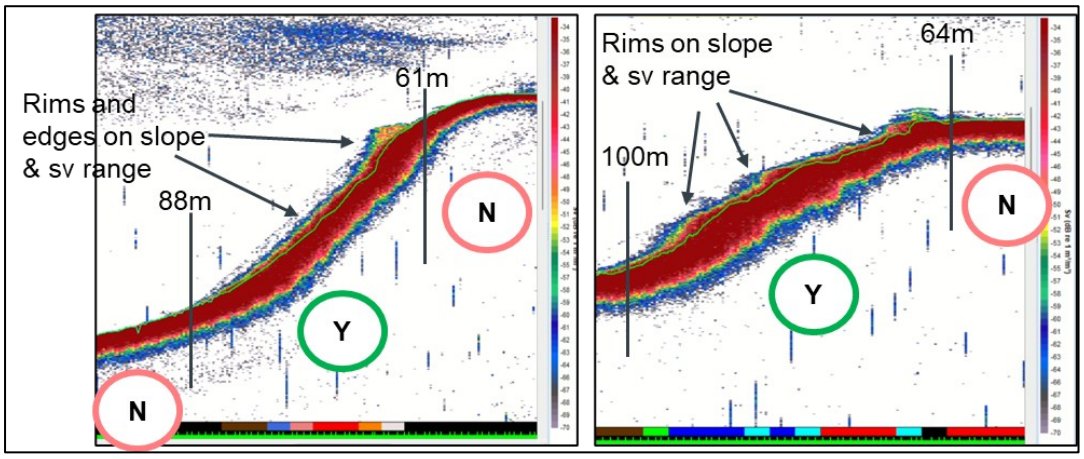


Figure 26. Phase 3, another example of the reviewing of potential coralligenous presence on different echograms

In terms of depth, 50 meters below the sea surface was considered. The threshold was indeed planned to exclude the other main endemic underwater habitat in the area, seagrass meadows (Posidonia Oceanica). The structure of the formation considered were rims on sloped terrains, and the backscattering

strength, based on the analysis of each single part of the available echogram, was within the range between -36db and -55db (Figures 25 and 26). ROVs and Citizen Science data were used as complementary in this stage of the process to enhance the selection of points for potential presence of coralligenous concretion.

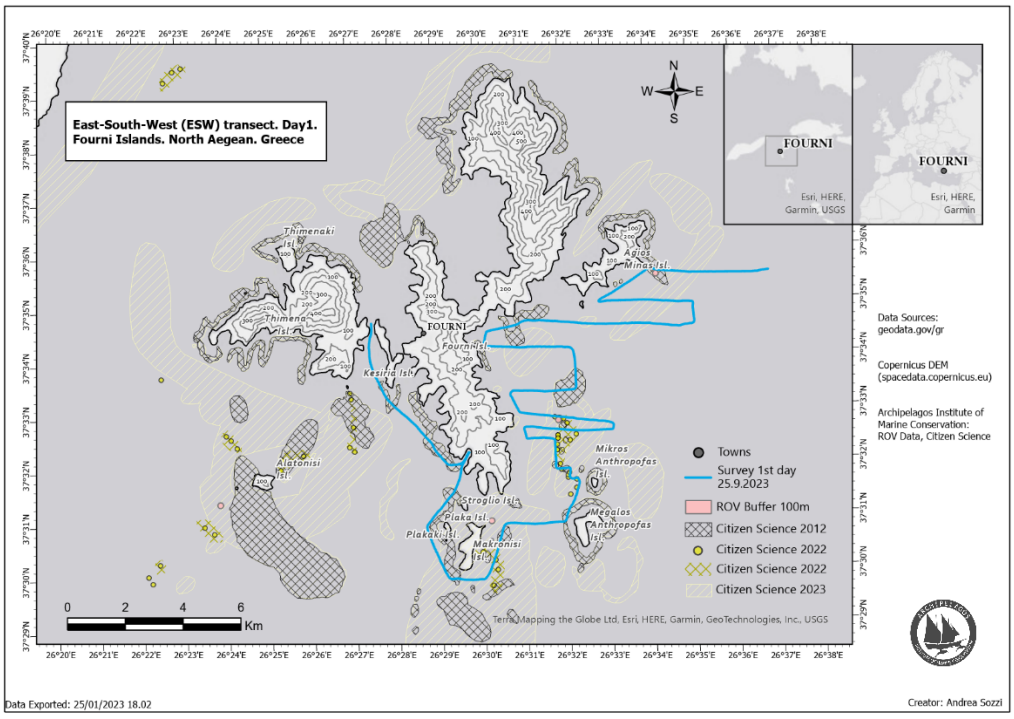


Figure 27. Map of the transect during the first day of survey (ESW) in the AOI, together with ROV and Citizen Science data.

Vector files (.shp), were implemented in ArcGIS Pro, to visualize ROVs and Citizen Science Data, together with the first day ESW transect (Figure 27). Lastly, the selected potential areas were added to the extrapolated features dataset file, as new class, into a new column (Potential_Coralligenous_Presence) and saved into a new file (F_ESW_Y&N.csv) (Appendix 1). The file contained “Yes” and “No” values for presence, and they were plotted into a map (Figure 28), providing geographical visualization.

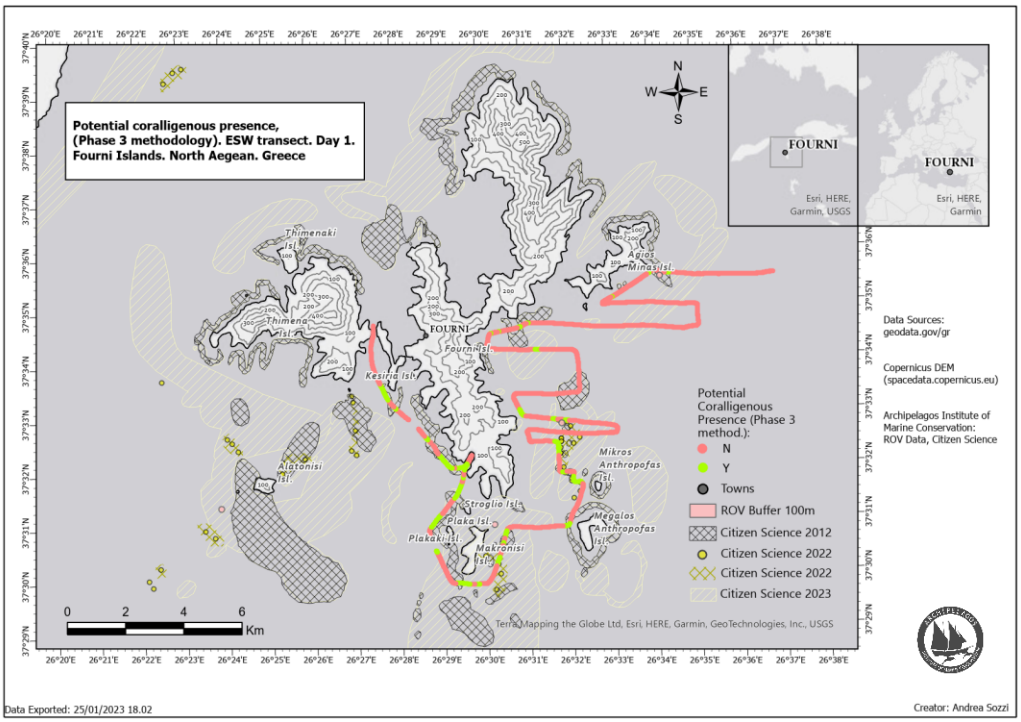


Figure 28. Map of the transect during the first day of survey (ESW) with potential coralligenous potential based on phase 3 of methodology

During the fourth phase, the Random Forest classifier for predicting the potential coralligenous presence was developed in a Python environment, with Scikit-learn library and based on the extrapolated features provided in the F_ESW_Y&N.csv (6413 datapoints) dataset. To decide which features were going to be used to implement the RF model, correlation matrix was first developed to highlight the potential high correlation amongst certain features (Figure 29). The *first bottom length* had an extremely high correlation value with the *bottom rise time* (0.97) and with the *second bottom length* (0.89). Also, these last two features were highly correlated with each other (0.89). For this reason, they were excluded from the development of the model, while the *first bottom length* was kept. Same applied for the *bottom max backscattering strength (max sv)*, which was excluded from the features used to develop the model after showing high correlation with *bottom roughness* (0.87).

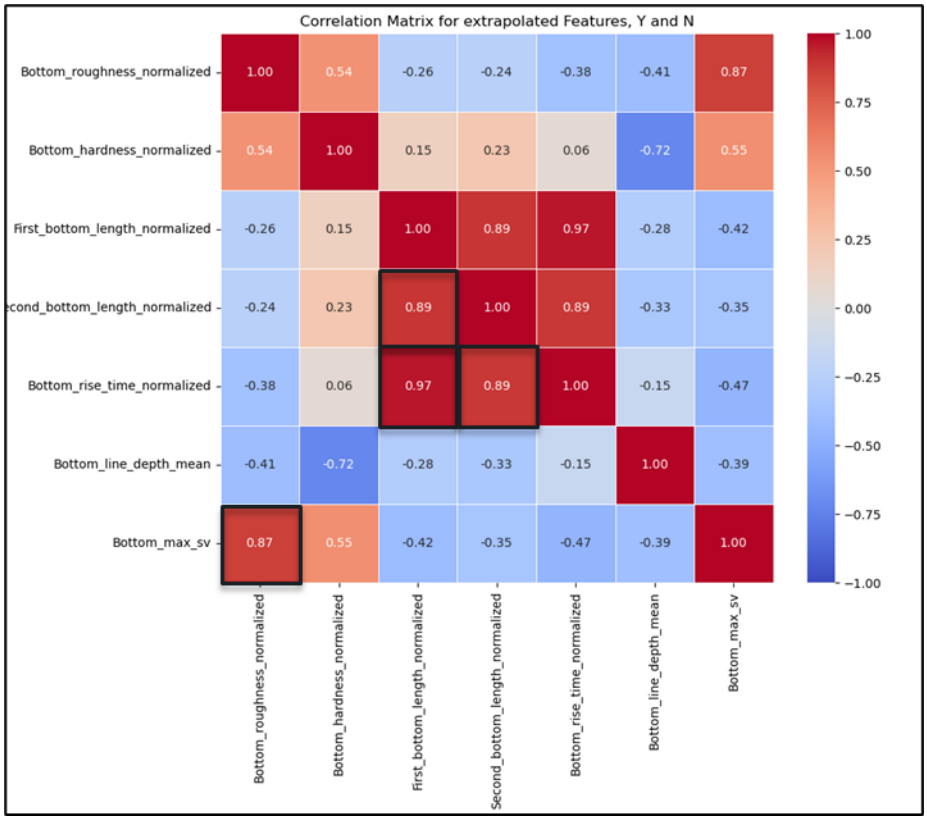


Figure 29. Correlation matrix with all available features from the ESW dataset

After the correlation matrix, and a process of trial and error, the RF model was implemented based on the combination of four features leading to the maximization of its performance: *bottom roughness*, *bottom hardness*, *first bottom length*, and *bottom-line mean depth*. The ESW dataset of the first day of the survey with the reviewed “Yes” and “No” values (F_ESW_Y&N.csv with 6413 datapoints), was split during the RF model creation, by default percentages, between training (80 percent of datapoints - 5130) and validation (20 percent datapoints - 1283).

The fifth and last phase involved the N dataset extrapolated data of the 26th of September (F_N_Extrapolated_Features with 4382 datapoints), which were used to test how the RF classifier could accurately and precisely predict the presence of the potential habitat in a different area, in the studied case, the north of Fourni, from a different day of the survey, the second day, in which no prior presence classification had occurred (Figure 30). Once the N dataset was injected into the RF classifier, a new file (F_N_Y&N.csv with 4382 datapoints) was created, with

the added columns “Coralligenous_Potential”, representing the class of potential coralligenous presence, and “Yes” and “No” as values (Appendix 2).

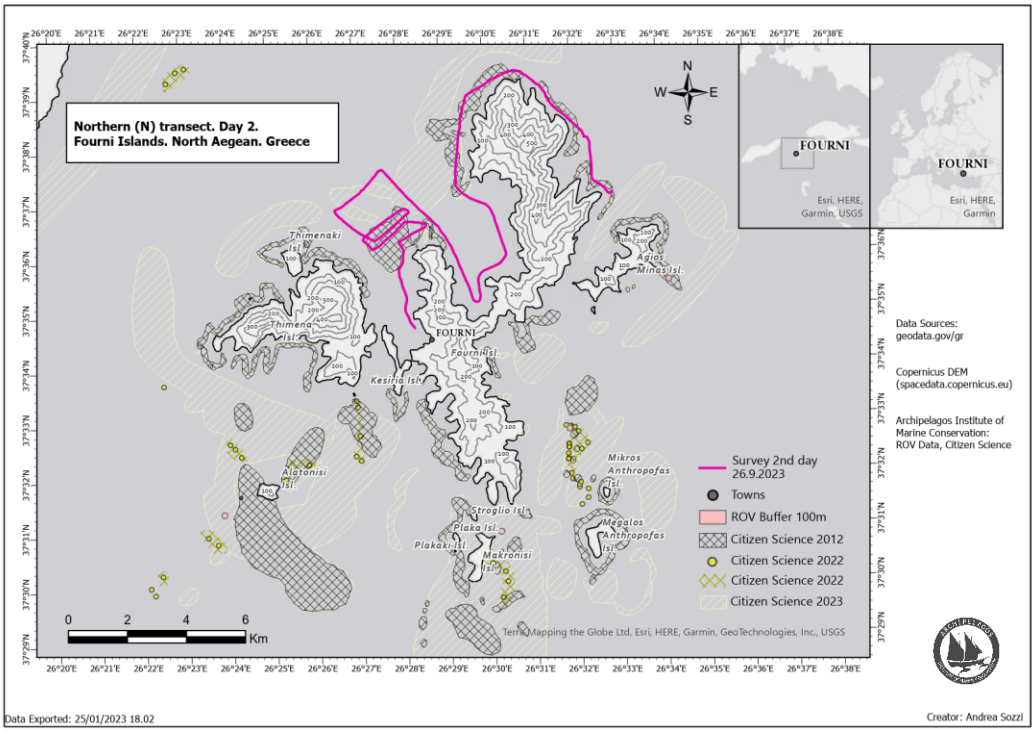


Figure 30. Map of the northern (N) transect during the second day of survey

Blind validation, based on the Third Phase approach, was then implemented on the same survey N dataset (F_N_Extrapolated_Features) without knowing the RF model results to compare and evaluate how the RF would have accurately and precisely predicted the outcomes. Specifically, several metrics and assessment tools were considered to evaluate the performance of the RF classifier and deployed with the Python library Scikit-Learn. These included confusion matrix, accuracy, precision, recall, and F1 score, metrics usually considered when evaluating RF performance (Rodriguez-Galiano et al. 2011; Sabat-Tomala et al. 2020; Fakiris et al. 2023), and calculated on both datasets, ESW (on the 20 percent, 1283 validation datapoints), and N (4382 predicted datapoints against blind validation).

Furthermore, for both datasets, ESW, and N, the Levene statistical test was then performed to understand whether or not homogeneity of variance (Cleves 1996; Boutros et al. 2015; Derrick et al. 2018) occurred between “Yes” and “No” values

in every single feature with a statistically significant difference. Subsequently, the Mann-Whitney U (Gilbertson et al. 1985; Serandour et al. 2010; Salek et al. 2013) test was also performed to discover whether statistically significant differences between “Yes” and “No” classes were present on each feature.

4 RESULTS

Statistical analysis and performance of the developed Random Forest (RF) Classifier were evaluated for both developing (F_ESW_Y&N.csv dataset with 6413 datapoints) and predicted (F_N_Y&N.csv dataset with 4382 datapoints) datasets separately.

4.1 Descriptive statistics for two-day survey datasets

The ESW dataset (F_ESW_Y&N.csv) used to develop the RF Classifier consisted of 6413 datapoints. The mean value of the normalized *bottom roughness* (dB re $1m^2/m^3$) was 7.64 (standard deviation 0.29), of normalized *bottom hardness* (dB re $1m^2/m^3$) 3.32 (standard deviation 0.70), of normalized *first bottom length* (m) 4.50 (standard deviation 4.98), and of the normalized *line depth mean* (m) 93.61 (standard deviation 22.61) (Appendix 3).

The N dataset (F_N_Y&N.csv) dataset created from the prediction made by the previously developed RF classifier consisted of 4382 datapoints. The mean value of the normalized *bottom roughness* (dB re $1m^2/m^3$) was 7.76 (standard deviation 0.34), of normalized *bottom hardness* (dB re $1m^2/m^3$) 3.35 (standard deviation 0.64), of normalized *first bottom length* (m) 4.83 (standard deviation 1.75), and of the normalized *line depth mean* (m) 98.18 (standard deviation 20.71) (Appendix 3).

4.2 Statistical analysis for the potential coralligenous presence

Statistical analysis between the “Yes” (Y) and “No” (N) values from the potential coralligenous presence was performed for each of the four selected normalized features, used to build the RF model, to further understand the datasets, and to

determine statistical significance, for both ESW (used to develop the RF model), and for the N datasets (predicted by the RF model).

In the ESW dataset used to develop the RF classifier, for the normalized *bottom roughness* (dB re $1m^2/m^3$), the mean of "Yes" values was 7.79 (with a median of 7.81 and standard deviation of 0.24), while for "No" values was 7.61 (with a median of 7.61 and standard deviation of 0.29). For the normalized *bottom hardness* (dB re $1m^2/m^3$), the mean of "Yes" values was 3.54 (with a median of 3.51 and standard deviation of 0.56), while for "No" values was 3.29 (with a median of 3.03 and standard deviation of 0.71). For the normalized *first bottom length* (m), the mean of "Yes" values was 6.40 (with a median of 6.02 and standard deviation of 2.51), while for "No" values was 4.19 (with a median of 3.73 and standard deviation of 5.21). For the normalized *line mean depth* (m), the mean of "Yes" values was 82.44 (with a median of 80.04 and standard deviation of 24.36), while for "No" values was 95.43 (with a median of 100.39 and standard deviation of 21.24) (Table 3, Appendix 4).

Table 3. Mean, median, and standard deviation, grouped by "Yes" and "No" values for coralligenous potential presence, for the selected normalized features used to develop the RF classifier

Stat.	**Dataset**	**Bottom roughness** (dB re 1m2/m3)		**Bottom Hardness** (dB re 1m2/m3)		**First bottom length** (m)		**Line depth Mean** (m)	
Potential Presence		**Y**	**N**	**Y**	**N**	**Y**	**N**	**Y**	**N**
Mean	ESW	7.79	7.61	3.54	3.29	6.40	4.19	82.44	95.43
Median	ESW	7.81	7.61	3.51	3.03	6.02	3.73	80.04	100.39
Std	ESW	0.24	0.29	0.56	0.71	2.51	5.21	24.36	21.24
Mean	N	8.00	7.64	3.32	3.37	6.48	4.05	87.96	103.04
Median	N	8.04	7.67	3.27	3.19	5.96	3.99	88.60	105.75
Std	N	0.29	0.29	0.45	0.71	2.00	0.86	16.38	20.79

In the N dataset automatically predicting presence of coralligenous potential by the RF classifier, for the normalized *bottom roughness* (dB re $1m^2/m^3$), the mean of "Yes" values was 8.00 (with a median of 8.04 and standard deviation of 0.29), while for "No" values was 7.64 (with a median of 7.67 and standard deviation of 0.29). For the normalized *bottom hardness* (dB re $1m^2/m^3$), the mean of "Yes" values was 3.32 (with a median of 3.27 and standard deviation of 0.45), while for

“No” values was 3.37 (with a median of 3.19 and standard deviation of 0.71). For the normalized *first bottom length* (m), the mean for “Yes” values was 6.48 (with a median of 5.96 and standard deviation of 2.00), while for “No” values was 4.05 (with a median of 3.99 and standard deviation of 0.86). For the normalized *line mean depth* (m), the mean of “Yes” values was 87.96 (with a median of 88.60 and standard deviation of 16.38), while for “No” values was 103.04 (with a median of 105.75 and standard deviation of 20.79) (Table 3, Appendix 4).

In the ESW dataset, the Levene test for *bottom roughness* resulted in a p-value of 0.002 (less than the 0.05 threshold), while for all other three tested features, the obtained p-value for each of them was lower than 0.001 (less than the 0.05 threshold), thus concluding that all four features presented statistically significant variances between “Yes” and “No” classes, hence excluding t-test from the analysis. The same results occurred as well in the N dataset, in which, running the Levene test resulted in a p-value of less than 0.001 (less than the 0.05 threshold) for all the analyzed features (Appendix 5). Furthermore, a p-value lower than 0.001 (less than the 0.05 threshold) was achieved by the Mann-Whitney U test in each feature for both datasets, hence confirming the statistically significant difference between “Yes” and “No” classes, on all selected features for both the ESW and the N datasets (Appendix 6).

4.3 Random forest classifier performance

Performance of the Random Forest classifier was evaluated for both datasets. In the ESW dataset, the 20 percent validation datapoints (1283) were used to understand the RF model performance and how the selected features impacted the model, while secondarily, on the N dataset, the full datapoints (4382) were evaluated against blind validation (developed by leveraging “phase 3” of the methodology).

To understand how each single feature contributed to the development of the RF classifier, feature importance was run on Python, via Scikit-Learn library. Feature importance provides insights on how the model learned to make prediction during the training phase. The highest importance was attributed to *the line depth mean*,

contributing 0.36 (36 percent) in learning to predict the potential coralligenous presence and assigning “Yes” or “No” values. *First bottom length* followed as second, with 0.34 (34 percent) contribution, and *bottom hardness* and *roughness*, afterward, with 0.16 (16 percent) and with 0.14 (14 percent) respectively (Figure 31).

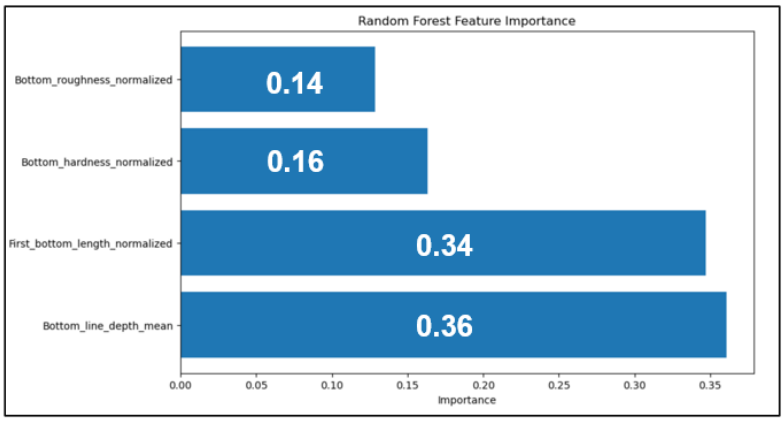


Figure 31. Feature importance based on the ESW training dataset

In the confusion matrix built from the ESW validation (20 percent) dataset, the *True Label* data are defined by phase 3 of the methodology, while the *Predicted Label* data are identified by the automated prediction of the RF model. Of the 1283 datapoints, 1079 were identified as True Negatives (TN), 136 as True Positives (TP), 40 as False Negatives (FN), and 28 as False Positives (FP) (Figure 32). The accuracy reached by the RF model on the ESW 20 percent datapoints was 0.95 (95 percent), Precision 0.83 (83 percent), Recall 0.77 (77 percent), and F1 score 0.80 (80 percent) (Table 4).

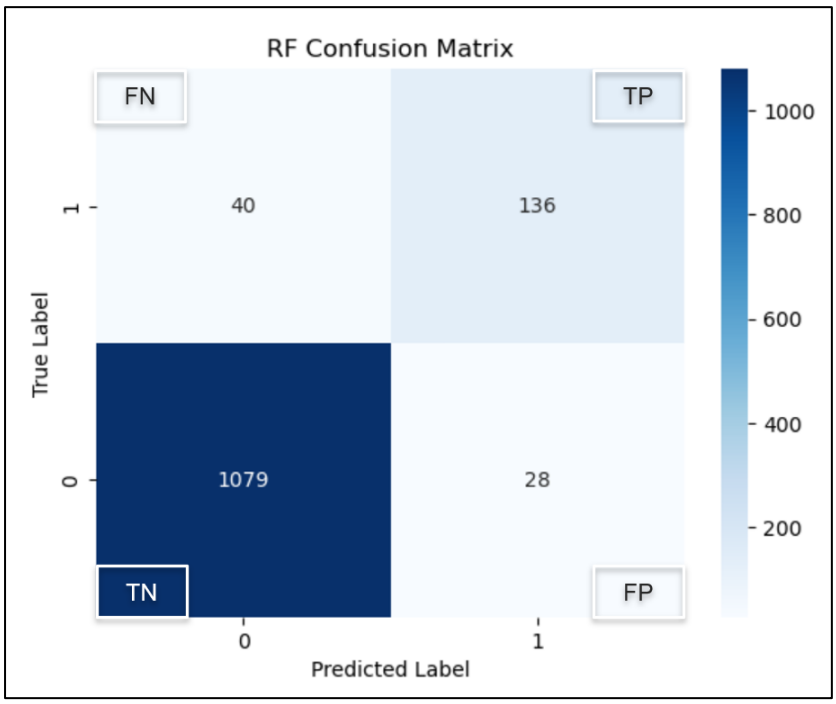


Figure 32. Confusion matrix of the first-day ESW survey based on validation set, 20 percent of the whole (1283 datapoints) dataset

In the confusion matrix of the N dataset instead, of the 4382 datapoints (Figure 33), 2891 were identified as True Negatives (TN), 1035 as True Positives (TP), 79 as False Negatives (FN), and 377 as False Positives (FP) (Figure 37). The accuracy reached by the RF model, on the N predicted datapoints was 0.90 (90 percent), Precision 0.73 (73 percent), Recall 0.93 (93 percent), and F1 score 0.81 (81 percent) (Table 4).

Table 4. Summary of the Random Forest Classifier results obtained for both first- and second-day survey dataset

Day of survey	*First*	*Second*
Date	25.9.2023	26.9.2023
File name	*F_ESW_Y&N.csv*	*F_N_Y&N.csv*
Total datapoints (dp)	6 413	4 382
Dp. used for prediction	1 283	4 382
Accuracy	0.95	0.90
Precision	0.83	0.73
Recall	0.77	0.93
F1 Score	0.80	0.81
True Negative (TN)	1 079	2 891
True Positive (TP)	136	1 035
False Negative (FN)	40	79
False Positive (FP)	28	377

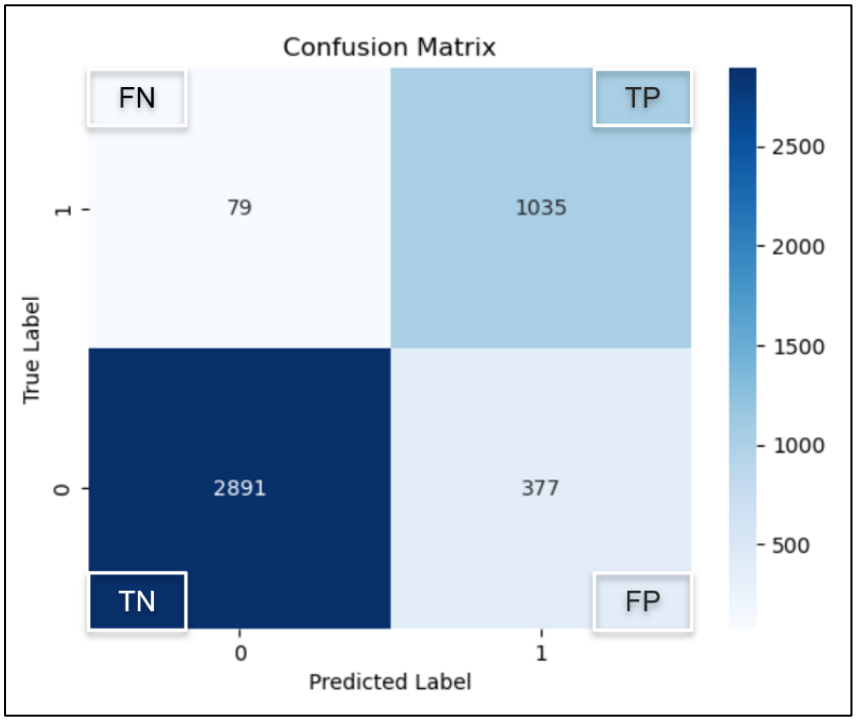


Figure 33. Confusion matrix of the second-day survey based on the whole (4382 datapoints) predicted N dataset

4.4 Mapping the random forest classifier results

Along the Northern (N) survey transect, amounting to a total of 34.7 kilometers in length, the prediction provided by the RF model, identified 128 and 130 segments for the “Yes” and “No” classes respectively, while 26 and 28 segments were defined during blind validation. In the prediction made by the RF model, the “Yes” class represented the 32 percent of the total length of the survey transect, while in the blind validation 25 percent. Specifically, the length of the “Yes” class was 11 048.58 (m) for the RF prediction, and 8 775.91 (m) for the blind validation, while the length of the “No” class was 23 732.39 for the RF prediction, and 26 004.21 (m) for the blind validation (Table 5).

Table 5. “Yes” and “No” class values of the segments identified in the survey with their total length (in meters), for both the RF prediction and Blind validation datasets

	Segments		**Length (m)**		**Length (%)**		**Total (m)**
Class	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	
RF Prediction	128	130	11 048.58	23 732.39	32	68	34 780.97
Blind Validation	26	28	8 775.91	26 004.21	25	75	34 780.12

The identified “Yes” and “No” classes resulting from the Random Forest classification and from the Blind validation dataset were then plotted and visualized in ArcGIS Pro (Figure 34).

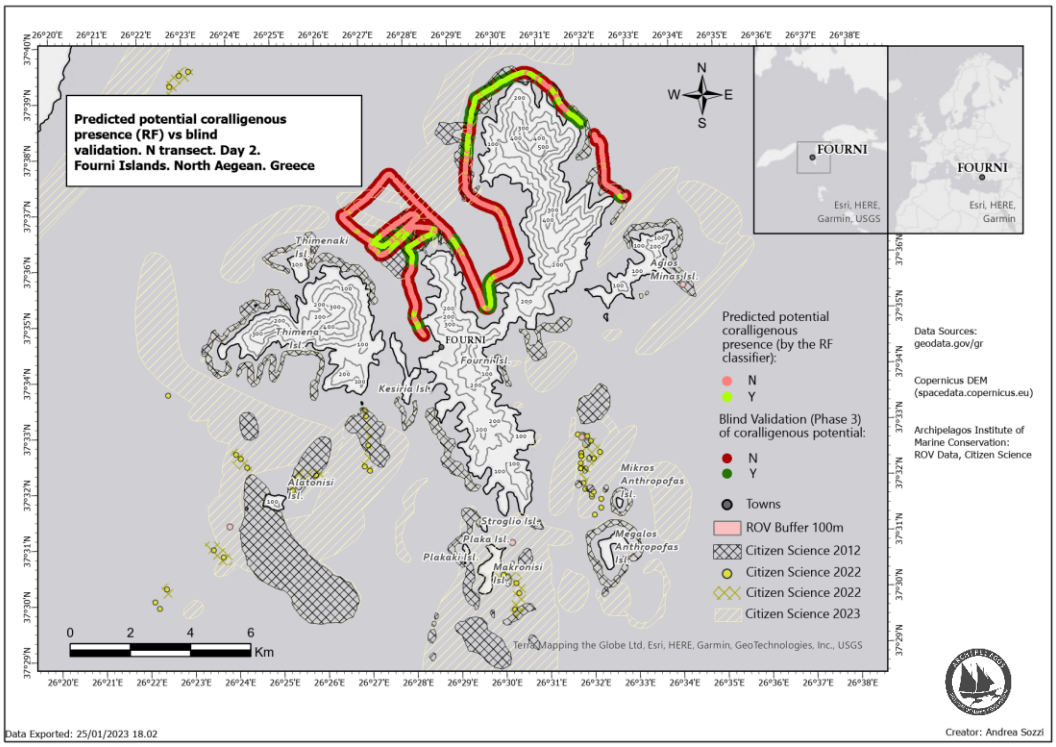


Figure 34. Map of the predicted potential coralligenous presence by the RF classifier

5 DISCUSSION

The main objective of the study was to develop a methodology to assess, classify, and predict potential coralligenous presence along the coastline of the area of interest, using single-beam echosounder data. Statistical analysis and development of the Random Forest (RF) classifier based on features of the bottom extracted from bottom classification with Echoview provided the means to reach the objective, as well as gather the information required for proper mapping in GIS software.

5.1 Random forest model results

The key to providing a streamlined methodology was to evaluate whether or not the creation of the RF model from one dataset of a certain coastal area (in the

specific case east, south, and west) was able to provide accurate and precise measurements on a second dataset (northern coastal area). Once the RF model was constructed, testing it in a brand-new zone would indeed showcase its potential for prediction. In the present study, the RF model developed from the ESW dataset values (F_ESW_Y&N.csv, first day of survey), when applied to the N dataset (second day of survey), delivered consistent performance in terms of accuracy (90 percent) and precision (73 percent).

5.1.1 Selected features for the RF development

In both datasets, ESW and N, the selected features of the bottom were evaluated to understand their statistical differences. As seen in Table 3, pattern of similarities is apparent between the values of ESW and N datasets, on both “Yes” and “No” classes, in which “Yes” tend to be higher than “No” for all the features, but the *line depth mean,* in which the opposite occur, being the “No” values higher than “Yes”. Overall, the correspondence amongst features and classes between the ESW and the N datasets can be considered a sign of consistency between both learning and predicted data. Moreover, statistical differences occurred between “Yes” and “No” values for each evaluated feature, tested with the Mann-Whitney U, provided further reliability in the separation between the “Yes” and “No” classes on both datasets.

The RF model (developed on the ESW dataset) was constructed upon four selected features (Table 3). Their subsequent importance was then evaluated to assess the relevancy of each feature in predicting the targeted variable. Feature importance is based on the decrease in impurity (or Gini index) resulting from the splitting of nodes according to a particular feature during the development of the RF model (Hong et al. 2016; Fakiris et al. 2023). The findings indicate that *Bottom roughness* and *hardness* (contributing for 0.14 and 0.16 respectively), even though they do characterize the RF model and are essential parts in identifying bottom characteristics, played a minor role in the definition of the RF model. The *line depth mean*, and *first bottom length* resulted instead in the highest importance scores (0.36 and 0.34). This suggests that variations in these two features play a significant part in determining the outcome of the model. On

the one hand, even though *line mean depth* is a feature provided by the bottom classification, its values are considered directly as one of the main parameters for reviewing the potential presence of coralligenous concretions, based on literature review, and based on previous knowledge about the area of interest (as defined in the phase 3 of the methodology). The mean values of the depth obtained in both ESW and N datasets of 82.44 and 87.96 m fall into the range provided by previous scientific studies and by the commissioner's own research. A higher difference between the "Yes" and "No" classes was then pursued, and identified, as depicted in the boxplots in Appendix 4. On the other hand, the *first bottom length*, is a feature extracted exclusively during bottom classification. As seen in equation (6), it refers to the average length of the first bottom echo, in the analyzed interval, capturing the extent of the reflection from the seabed to bottom surface (Echoview 2023d). In the case of its importance to the model, higher variability between "Yes" and "No" classes can be seen also seen when looking at the boxplots in Appendix 4. This discriminant difference allows the RF model to differentiate between "Yes" and "No" classes more effectively, consistently contributing to reduce impurity across decision trees within the forest, and hence assigning its higher importance score. These results obtained on the importance of *depth* and *first bottom length* features are in line with previous research on coralligenous habitat distribution determined by predictive modeling.

In the model developed by Martin et al. (2014, 4-5), bathymetry turned out to be the highest importance feature when determining the distribution presence, followed by the slope of the seafloor. Also, in Fakiris et al. (2023, 9-10), both depth and slope are important characteristics when defining their suitability model for coralligenous distribution. The slope of the seafloor identified by Martin et al. (2014, 4-5) as the second most important feature to determine the model, and a relevant feature as well by Fakiris et al. (2023, 9-10) could be interpreted as congruence with the current study's second most important feature as well, the *first bottom length*.

Upon running an empirical analysis on random segments of the N transect, from latitude 37.624067 to the latitude 37.649783 (Figure 35), by looking at the

echograms and at the *first bottom length* values, the average value of the “Yes” segments is higher compared to the average value of the “No” class (Table 6), as per statistical findings in Table 3 of the results. And higher values corresponded to steeper incline when visualizing the echogram of the segments (Figure 36).

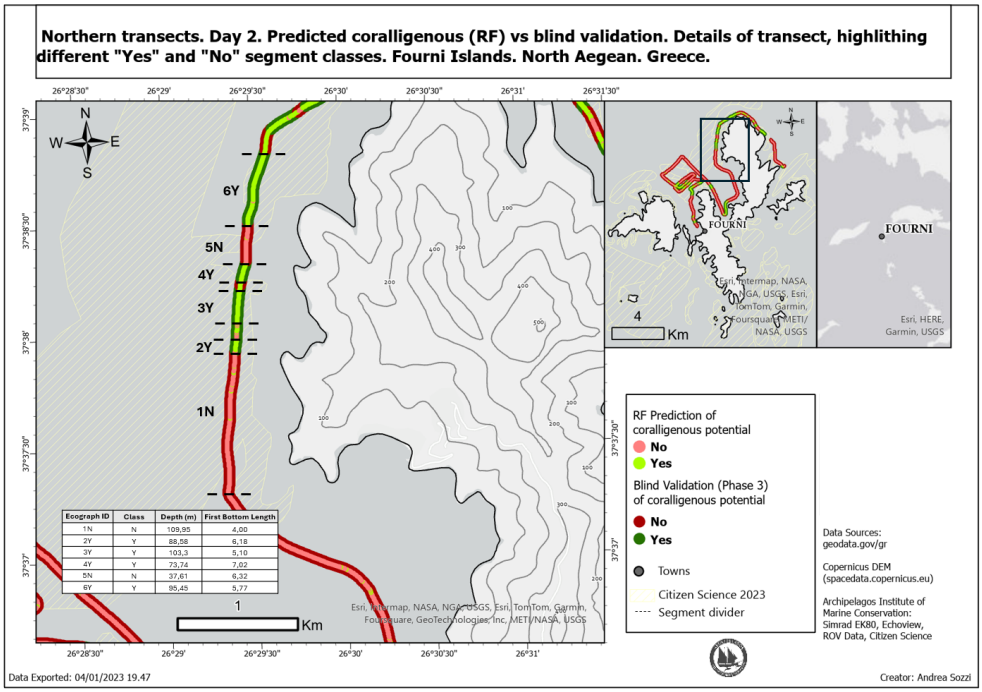


Figure 35. Map highlighting details of the transect from segments provided in Table 6

The “1N” image is the only “No” class that matches the depth required for the RF model, however the incline being almost flat, matches with a lower value of *first bottom length* than the other “Yes” class, and in line with the average value of the “No” class (4.00).

Table 6. Avg. *line depth mean* and avg. *first bottom length* values for specific segments of the N dataset

Ecograph ID	Index initial	Index final	Lat. initial	Lat. final	Class	Count	Depth (m)	First Bottom Length
1N	2939	3097	37.624067	37.634900	N	158	109.95	4.00
2Y	3097	3108	37.634900	37.635705	Y	11	88.58	6.18
3Y	3128	3165	37.636709	37.639665	Y	37	103.3	5.10
4Y	3174	3192	37.640259	37.641450	Y	19	73.74	7.02
5N	3194	3234	37.641585	37.644334	N	42	37.61	6.32
6Y	3237	3315	37.644545	37.649783	Y	79	95.45	5.77

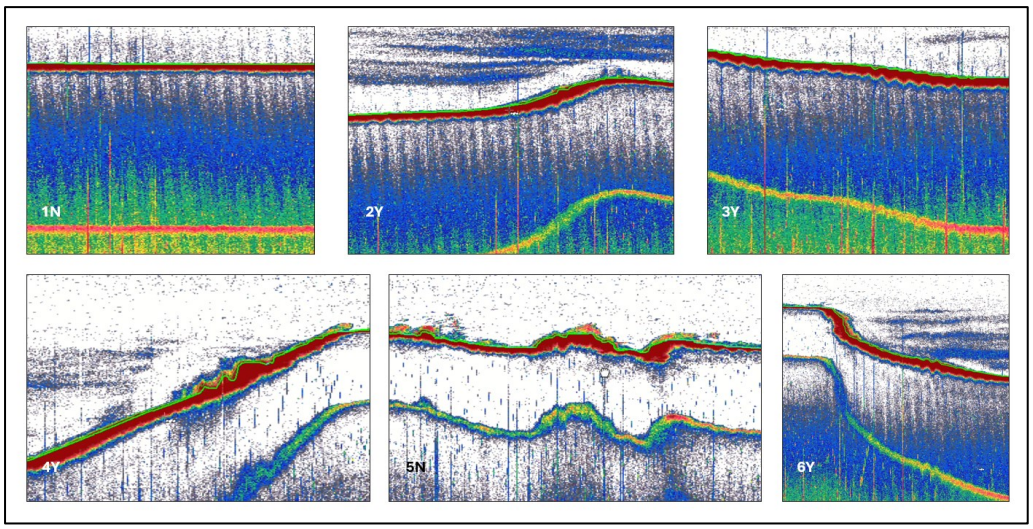


Figure 36. Echograms of specific segments from the N dataset based on Table 6

The other segments are all considered “Yes” class, higher values of *first bottom length* matching with sloped terrain. The exception is for the segment “5N”, in which even though sloped terrain is present in its echogram, its depth of 37.61 meter (average) does not match to the minimum depth requirement for the RF model to classify it as “Yes” class. These empirical findings, in which sea bottom segments with a higher inclination translate into higher value of extrapolated feature *first bottom length* from the bottom classification in Echoview, should be the center of further studies which would determine accurately the relation between *first bottom length* and slope, a type of research outside of the scope of the current study, though providing the background for further research. For example, the values of *first bottom length* produced when running the classification in Echoview could be compared with previous data about the slope of the area of interest, from publicly available databases, such as EMODnet, or with other data collected on surveys by the commissioner. In the past, at Archipelagos several projects were conducted with spatial analysis in GIS software to derive the slope of the seafloor, for example by using multibeam sonar data (internal communication at Archipelagos 2023).

5.1.2 Assessment of the RF metrics performance

The metrics assessed for both RF classifiers, the one developed with the 20 percent validation datapoints from the first-day bottom survey (ESW dataset), and

the one automatically predicting the “Yes” and “No” classes on the data of the second-day bottom survey (N dataset), provided an overview of the RF achieving consistent performance with class prediction.

Accuracy can be defined as “the ratio between the number of correct predictions to the total number of predictions” (Olson & Delan 2008, 32; Google 2024a). Considering the ESW and N datasets, *accuracy* levels of 95 percent and of 90 percent were achieved respectively, translating into an overall high level of correct classification by the RF classifier in identifying “Yes” and “No” classes of potential coralligenous presence, compared to the training dataset and the blind validation (through phase three of the methodology).

Precision, *Recall*, and *F1 score* were identified as performance metrics for evaluating the RF model, in tandem with the *confusion matrices*, for both datasets, providing a detailed breakdown of the model’s predictions. *Precision* is defined as “the ratio of True Positives to the sum of True Positives and False Positives”. *Recall* is “the ratio of True Positives to the sum of True Positives and False Negative”, while the *F1 score* is the combination of Precision and Recall, more specifically calculated as “the ratio of twice the product of precision and recall to the sum of precision and recall” (Olson & Delan 2008, 138-139; Google 2024b). In Fakiris et al. (2023), the threshold representing the minimum predicted probability or score required for an instance to be classified as a positive class used by the researchers in their study was 0.23 and developed by analyzing the Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values. Their developed RF model achieved a *Precision* score of 64.47 percent on the South Aegean and 76.54 percent on the North Aegean (the Fourni Islands being geographically in the South Aegean however belonging to the North Aegean administrative area), and overall, of 70 percent for the Aegean basin. Even though different geographical scales, thresholds for predicted probability, and different features for developing the RF model than the ones in the present study, these values can be accounted for reference for a broader first comparison. And they provide a starting benchmark of the RF model in terms of *Precision* performance. In the current study, the threshold value for predicted

probability was the default 0.50, provided by Scikit-learn (Stackoverflow 2022; Scikit-learn 2024). *Precision* was higher on the validation datapoints of the ESW dataset, compared to the predicted dataset N, 83 percent versus 73 percent, while it was the opposite for the *Recall*, consisting of 77 percent for the ESW dataset versus 93 percent for the N dataset. *F1 score* in both RF model performances was harmonized upon these discrepancies, setting at 80 and 81 percent for the ESW and N datasets respectively.

In light of the usefulness of the RF model in developing binary predictions for potential coralligenous presence, *Precision* should be considered on a higher domain than *Recall,* as the main goal of the model is to find *True Positives* and *Negatives* and minimize *False Positives*, those events which truly were a “No” class, however predicted by the model as a “Yes”. In this regard, even though relevant overall, minimizing *False Negatives* is less critical than *False Positives*, as it is imperative to find first areas of presence, to enable further protection activity. Predicting a “Yes” class wrongly could indeed translate into allocating resources for further research in areas of no interest. (internal communication at Archipelagos 2023.)

The RF model predicted in the northern area of the Fourni Islands (N dataset) 377 *False Positives*, leading to the 73 percent *Precision* score (Figure 37).

As described in Table 5, different numbers of segments and different lengths (in meters) were identified between the RF prediction and the blind validation. The RF model predicted much more, and much shorter instances of both “Yes” and “No” classes, compared to the classes derived via blind validation. Further analysis should be conducted to better understand the reasoning behind this phenomenon, for example by employing cross-validation methods, and ROC curves and AUC values for both ESW and N dataset, as well as investigating every single point of discrepancy, trying to find specific information or pattern determining the origin of such determination of the RF model.

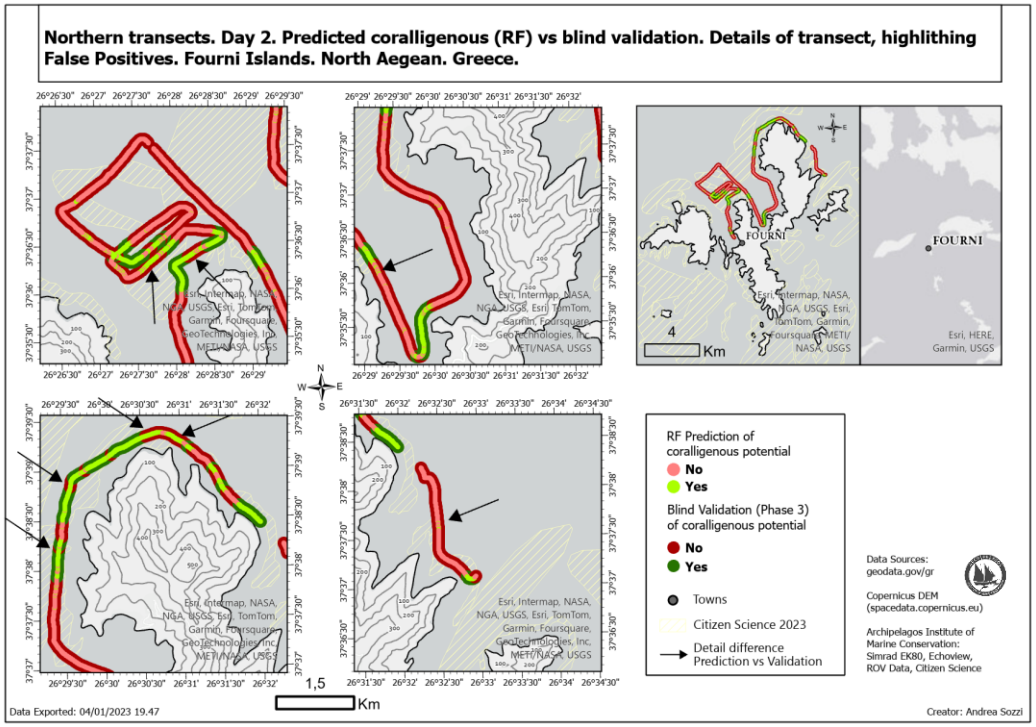


Figure 37. Maps with details of the N transect highlighting False Positives

5.1.3 RF results and citizen science data

During phase 3 of the methodology, when building the model on the ESW dataset, citizen science (CS) data was used as complementary information, to support literature values, echograms analysis, and previous research conducted by Archipelagos on the habitat distribution. During the second-day of survey, the northern part of the main island of Fourni was explored. When considering only the transect passing through the CS areas defined within the 2023 citizen science data (Figure 21), the mapped results (Figure 38) obtained by the RF predicted dataset (N) of the “Yes” and “No” classes, have a higher “Yes” to “No” class ratio *inside* the CS patches (70.4 percent) compared to the areas *outside* (35.6 percent). Indeed, the potential distribution described by fishermen prior to the current study, and mapped in GIS software, are corresponding to areas of higher prevalence of potential coralligenous presence compared to the areas *outside* the CS patches. Specifically, a positive difference between the percentage of the “Yes” class *inside* the areas highlighted by the fishermen and the “Yes” class *outside*, amounting to 14.9 percent, compared to the negative 15.5 percent difference between *inside* and *outside* *“*No*”* class, suggests that potential

coralligenous presence is more relevant within the highlighted areas. This is a key finding, for two reasons: it provides verification, a first acknowledgement that the methodology to construct the RF model succeeded to find ”Yes” classes also into the area that local communities highlighted beforehand. And additionally, as highlighted in other studies, it serves as further evidence of the significance in engaging and gathering information from indigenous communities when running scientific research (Vlachopoulou et al. 2013; Bonney et al. 2021). However, it is crucial to also understand that due to the intrinsic nature of the single-beam echosounder, the data available from the defined transects are determined exclusively by the size of the beam hitting the seabed, while the areas highlighted by the locals represent a much wider extent. Further research would require planning and defining transect lines crossing through multiple times CS patches, or the use and subsequent analysis of multibeam sonar data, which, with a fan-shape beam, would scan a wider area per passage, compared to a single-beam sonar.

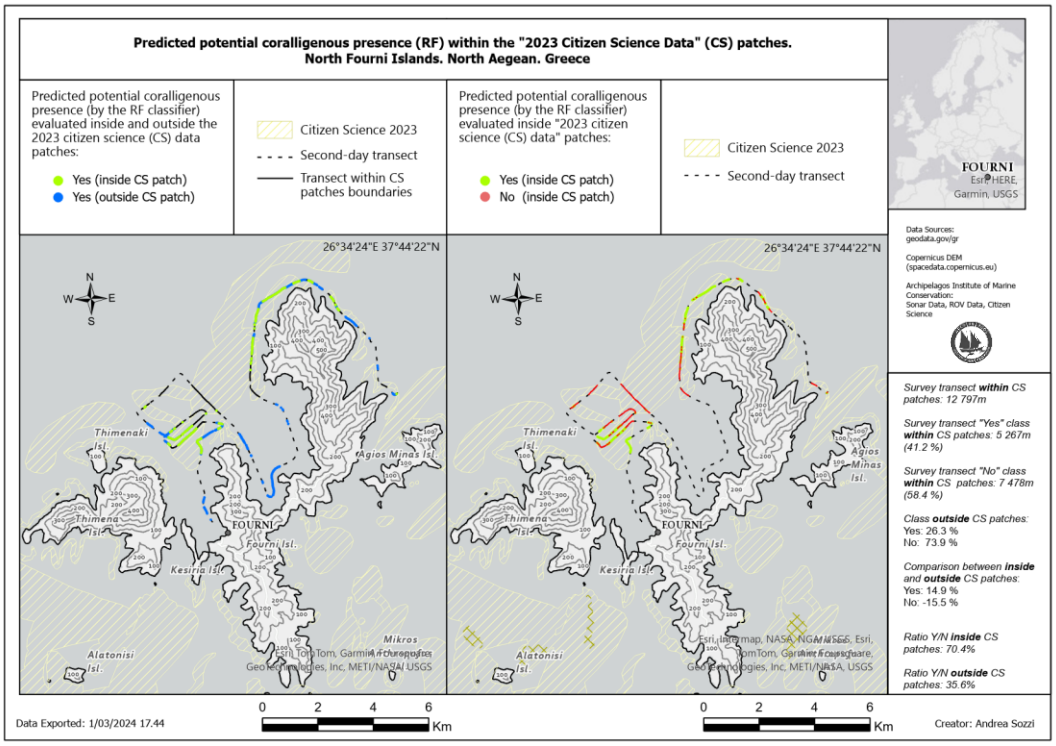


Figure 38. Predicted potential coralligenous presence (RF) within the “2023 Citizen Science Data” (CS) patches

5.1.4 RF and coralligenous habitat distribution

Coralligenous habitat is present in the Aegean Sea at various degrees. When comparing side by side the cartographic representations provided by prior research (Figures 7, 8, 9, 10, 11) with the findings delineated in the present study (Figure 34), a variable level of correspondence is noticeable, in terms of areas highlighting coralligenous distribution, most likely because of a different scale used in the other studies. Giakoumi et al. (2013) used previously available sources to extract zones of presence of coralligenous formations. In their map showcasing the distribution (Figure 7), the southwest part of the Fourni Islands is highlighted as presence of the habitat. The same results are visible also in the map (Figure 10) provided by Sini et al. (2017), also combining previous sources together to plot the distribution. Based on the produced map (Figure 9) in Martin et al. (2014), the occurrence probability of coralligenous outcrops in the area of interest of the current study, falls between 0.3 to 0.7, highlighting similarities with the findings of the current study about the potential presence of the habitats. When considering the spatial output of the South Aegean ecoregion from the predictive model developed by Fakiris et al. 2023 in Figure 11, the area of the Fourni Islands is highlighted as well, however with lower than their set threshold occurrence probability of 0.23. Specifically, the area of interest falls in the color coding corresponding to the values between 0.1 and 0.23, which suggests a potential presence however not high enough to the minimum probability threshold. These diverse results from previous literature in the area of interest explain why further research is needed to improve the knowledge of habitat distribution. In this sense, the current study, as well as other prior research produced by the commissioner, such as citizen science data, surveying for bottom mapping with multibeam sonar, and ROV dives, are the key tools to provide further information on potential and proven presence (for example with video footage) of coralligenous habitats in the area of interest. These findings are the milestone to further the knowledge of habitat distribution and help improve the accuracy at which coralligenous concretions could be spotted along the coastline of the Fourni Islands, and in other places across the Aegean Sea.

The coralligenous habitat distribution was evaluated at depth below 50 meters onwards only, to ensure the exclusion of the other main habitat of Posidonia Oceanica present in the area of interest up to 40 meters depth. As described in literature, even though more often than not coralligenous habitats are present on the depth used on the methodology of the current study, it may also occur at depths shallower than 50 meters. Further investigation should be carried out in the area of interest, and around the zones of operation of the commissioner, to explore to which extent coralligenous habitats are present in shallow waters, and how to eventually differentiate them with the use of sonar data, against seagrass meadows habitats.

5.2 Methodological considerations, limitations, and future perspectives

The created methodology provided an innovative way for processing, analyzing, and mapping on GIS software, single-beam sonar data. The methodology was split in different phases within different environments (Echoview, Python, Excel, ArcGIS Pro), to overcome certain limitations when running the .raw data in Echoview software. The bottom classification tool in Echoview was the only module available during the processing of the data which provided the selected extracted features of the bottom. As per the name of the software module, the tool indeed finds different bottom classes, and name them with ID numbers, however, to properly identify what classes belongs to what soil type and habitat, ground truthing of some sort, such as ROV dives, or sediment sampling collection, is required. In the previous research using Echoview for bottom classification, the studies combined the classes provided by the software with ground truthing by sediment sampling (Fajaryanti & Kang 2019; Fauziyah et al. 2020). The innovativeness of the current study is that, instead of using the automatically unidentified classes provided by the software, it used the features of the bottom to develop a prediction of habitat distribution, based on certain intrinsic characteristics of coralligenous formations, described in literature, and further type of information in regards of their presence in the area of interest (such as citizen science data and ROV dives). The idea of creating such a methodology stems from the fact that ground truthing is time-consuming and more expensive in terms of resources (internal communication at Archipelagos

2023). In this context, providing a first detailed small-scale and high-resolution maps of potential habitat distribution derived from sonar data assists the commissioner to better decision-making when planning future ground truthing activities, and surveys for mapping the seabed.

Defining the right boundaries of both data collection and also data processing is key to the successful repeatability of the model performance. Navigation and sonar settings should be replicated as close as possible, to increase the chances of a perfect replicate of the conditions that occurred during the two-day surveys. Especially the velocity of the vessel, which during the data collection was of approximately 4 knots, should be maintained, as well as certain sonar settings, such as the ones described in Table 1. During the data collection process, is also crucial to always collect both echoes of the bottom, *first* and *second*. To validate so, the operators in the GIS office onboard the *Aegean Explorer*, must at all times ensure that the depth at which the transducer will send the ping is enough to cover both echoes (Figure 39). The failure to keep on acquiring data from both two echoes may compromise the extrapolation of the features of the bottom on the Echoview bottom classification tool, hence jeopardizing the whole scope of the survey.

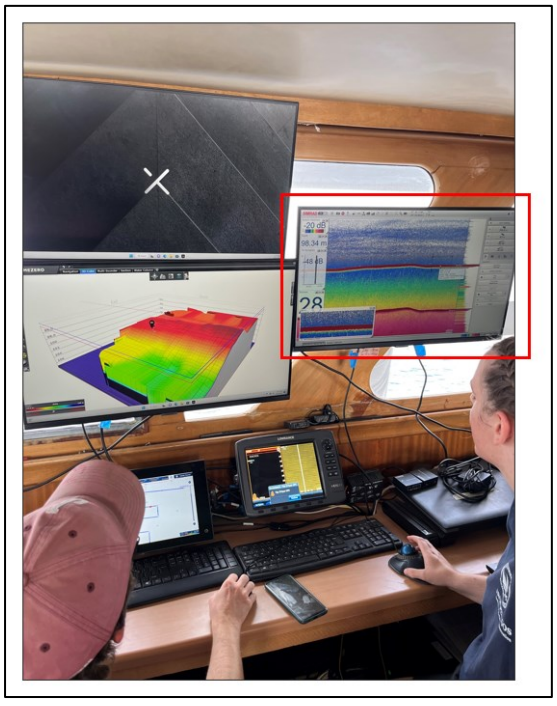


Figure 39. Operators onboard of the *Aegean Explorer*, ensuring to collect echosounder data from both echoes

Another relevant factor involves the planning of the survey transects. Empirical findings on previous survey data collected during August 2021 on the east side of the Fourni Islands suggest that multiple passages on a specific area, for example next to the east tip of Agios Minas island (Figure 15), followed by a conjunction route to another area in which other multiple passages were made around the island of Megalo Anthropofagas, led to skewed bottom features values after performing bottom classification in Echoview. In this case the data should be split into two separated datasets, in the specific example, one for Agios Minas, and one for Megalos Anthropofagas, either during the data collection phase directly, by turning off the sonar while navigating towards the next research spot, and turning it back on once at destination, or eventually during the data analysis phase, in Echoview. The second option was performed to the dataset and proved to solve the issue, by producing values of the features of the bottom in line with the ones produced by the ESW survey of the 25th of September 2023.

The applicability and replicability of the methodology is two-fold. The first application leverages phases 1, 2, and 5 of the process as described in Figure 23. It consists of injecting newly collected single-beam sonar data in the already developed RF model based on the ESW dataset. In this case, the .raw data acquired on boat surveys is imported in Echoview for bottom classification and extrapolation of the values of the features of the bottom (phase 1, and phase of the method). Once the features are extracted and available as .csv file, loading it in the RF model will automatically produce the “Yes” and “No” coralligenous potential classes, as a new column in the new file. The classes will then be ready to be plotted in GIS for mapping and analysis. The other application requires the use of each phase described in the process (Figure 23). This would occur in the eventuality that, even though the methodological considerations mentioned above are considered accordingly, the extraction of the features of the seabed delivered completely different values than the ones obtained in the ESW dataset. In this case, phases 3 and 4 of the process are required to build a new RF model for the new area appropriately. However, before suggesting this route, further trials of the RF model developed in the current study should be performed for

proper testing of prediction, within the area of interest but especially outside of it, for example in other zones of research scanned by Archipelagos. In this way, the suitability and the consistency of the model performance will be tested and proved for greater reliability. In this light, a brief test was conducted on data collected on another boat survey in the surroundings of Agathonisi island on the 12th and 13th of December 2023. A specific segment was previously reviewed by employing phase 3 of the process and identified completely as an area including only “No” classes of potential coralligenous presence. Implemented onto this dataset, the RF model achieved a 100 percent accuracy in attributing the “No” class of coralligenous presence to all the points belonging to the segment (internal communication at Archipelagos 2023).

The ultimate proof of reliability producing tangible information about the accuracy and usefulness of the model would be ground truthing. Planning ROV dives and sediment sampling in the Northern area of Fourni above the transect line of the current study would allow to test the accuracy of the model, by actually revealing the concrete presence of the coralligenous habitat. Thus, further research in this direction is highly suggested.

6 CONCLUSIONS

Marine ecosystems, and coralligenous habitats are globally under threats (Giakoumi et al. 2013; Martin et al. 2014; Sini et al. 2017; Halpern et al. 2019; Fakiris et al. 2023). One of the main known challenges for coralligenous formations and their preservation, amongst many identified around the world, and especially in the area of interest, is trawling (Vlachopoulou et al. 2013, 106; McConnaughey et al. 2019, 2-4). The lack of consistent cartographic data of their distribution, and the difficulty in obtaining homogeneous results from various datasets to define their location, bring further uncertainties on how to properly address conservation activities for coralligenous outcrops (Sini et al. 2017, 3). The collaborative nature of the commissioner research activities showcases the relevance of interdisciplinary cooperation in addressing marine conservation challenges, by bringing together scientists, conservationists, and local communities.

Exploring, investigating, and reporting of coralligenous habitat presence and their status, also through smaller scale research, are key factors to improve the chances of better protection and proper conservation, as these activities represent the bridge between the knowledge of previously unknown areas and proper intervention. In this light, the activities conducted by the commissioner in its area of operation between the island of Samos, Ikaria, the Fourni Islands, and the northern part of the Dodecanese archipelago, become of paramount importance to ensure that marine and environmental related issues become relevant to policymakers and other stakeholders, raising the awareness about their conservation challenges, and trying to push forwards for concrete milestones within the legislative framework.

The current study is the exact expression of this mandate: producing one more tool in the hand of the commissioner, to address the challenges of discovery and exploration of these hidden underwater gems. The developed predictive model automatically identified classes of coralligenous presence with an *accuracy*, *precision*, *recall*, and *F1 score* of 90, 73, 83, 81 percent respectively, and should enable Archipelagos with more capabilities to define these hotspot areas, optimizing conservation initiatives for maximum impacts. Another key factor to the success of the Random Forest classifier is its applicability to future data collected from future surveyed areas. Ground truthing would be the ultimate type of research for confirming the presence of coralligenous concretions, confirming the classes attributed by the predictions. However, due to the intrinsic difficulties in exploring these habitats, predictive modeling remains a more than viable method to efficiently allocate resources, prioritize conservation efforts, and guide future research endeavors to further understand and better protect these complex, endangered, yet fabulous, marine ecosystems.

REFERENCES

Angiolillo, M., Di Lorenzo, B., Farcomeni, A., Bo, M., Bavestrello, G., Santangelo, G., Cau, A., Mastascusa, V., Cau, A., Sacco, F. & S. Canese. 2015. Distribution and assessment of marine debris in the deep Tyrrhenian Sea (NW Mediterranean Sea, Italy). *Marine Pollution Bulletin*. 92, 1–2, 149-159. E-journal. Available: https://www.sciencedirect.com/science/article/pii/S0025326X15000041 [Accessed 8 December 2023].

Anderson, J., T. 2007. Acoustic Seabed Classification of Marine Physical and Biological Landscapes. *ICES Cooperative Research Report*. 286. PDF Document. Available: https://www.vliz.be/imisdocs/publications/125836.pdf [Accessed 19 December 2023].

Archipelagos. 2022. Protecting Aegean Coralligenous – Overall Activities. *Archipelagos Institute of Marine Conservation*. PDF Document. [Accessed 20 December 2023].

Ballesteros, E. 1992. Els vegetals i la zonació litoral: espècies, comunitats i factors que influeixen en la seva distribució. *Arxius Secció Ciències*. 101, 1–616. PDF Document. Available: https://digital.csic.es/bitstream/10261/22696/1/1992_Tesis.Els%20vegetals.%20i%20la%20zonaci%C3%B3%20litoral...1-616.pdf [Accessed 12 December 2023].

Ballesteros, E. 2003. The coralligenous in the Mediterranean Sea. Definition of the coralligenous assemblage in the Mediterranean, its main builders, its richness and key role in benthic ecology as well as its threats. *RAC/SPA- Regional Activity Centre for Specially Protected Areas*. PDF Document. Available: https://wedocs.unep.org/bitstream/handle/20.500.11822/812/coralligeneeng.pdf?sequence=1&isAllowed=y [Accessed 12 December 2023].

Ballesteros, E. 2006. Mediterranean coralligenous assemblages: a synthesis of present knowledge. *Oceanography and Marine Biology: An Annual Review*. 44, 123-195. E-journal. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=81492ed388163b4d1bd543f14b4d6b6fc751cc15 [Accessed 15 December 2023].

Bevilacqua, S., Guarnieri, G., Farella, G., Terlizzi, A. & Fraschetti, S. 2018. A regional assessment of cumulative impact mapping on Mediterranean coralligenous outcrops. *Sci Rep,* 8, 1757. E-journal. Available: https://www.nature.com/articles/s41598-018-20297-1#citeas [Accessed 10 December 2023].

Bonney, R., Byrd, J., Carmichael, j., T., Cunningham, I., Oremland, l., Shirk, J., Von Harten, A. 2021. Sea Change: Using Citizen Science to Inform Fisheries Management. *BioScience*, 71, 519–530. E-journal. Available:

https://academic.oup.com/bioscience/article/71/5/519/6162973 [Accessed 19 December 2023].

Boudouresque, C. 2004. Marine biodiversity in the mediterranean: status of species, populations and communities. *Sci-Rep*. 20, 97-146. E-journal. Available: https://www.researchgate.net/profile/Charles-Boudouresque/publication/256843455_Marine_biodiversity_in_the_Mediterranean_Status_of_species_populations_and_communities/links/58dc9df092851c611d3aff21/Marine-biodiversity-in-the-Mediterranean-Status-of-species-populations-and-communities.pdf Accessed 14 December 2023]

Boutros, N., Shortis, M., R. & Harvey, E., S. 2015. A comparison of calibration methods and system configurations of underwater stereo-video systems for applications in marine ecology. *Limnol. Oceanogr*. 13, 224–236. E-journal. Available: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.1002/lom3.10020 [Accessed 15 January 2024].

Ceccherelli, G., Pinna, F., Pansini, A., Piazzi, L. & La Manna, G., 2020. The constraint of ignoring the subtidal water climatology in evaluating the changes of coralligenous reefs due to heating events. *Sci. Rep*. 10, 1, 1–13. E-journal. Available: https://www.nature.com/articles/s41598-020-74249-9 [Accessed 18 December 2023].

Christodoulakis, D., Artelari, R., Georgiadis, T. & Tzanoudakis, D. 2001. New records to the flora of Fourni (E. Aegean islands, Greece). *Bocconea*. 13, 491-494. E-journal. Available: https://www.herbmedit.org/bocconea/13-491.pdf [Accessed 3 January 2024].

Cleves, M. A. 1996. Robust tests for the equality of variances. *Stata Technical Bulletin*. 25. PDF Document. Available: https://www.researchgate.net/publication/24137168_Robust_tests_for_the_equality_of_variances [Accessed 15 January 2024].

Consoli, P., Romeo, T., Angiolillo, M., Canese, S., Esposito, V., Salvati, E., Scotti, G., Andaloro, F. & Tunesi, L., 2019. Marine litter from fishery activities in the Western Mediterranean sea: The impact of entanglement on marine animal forests. *Environ. Pollution*. 249, 472-481. E-journal. Available: https://www.sciencedirect.com/science/article/abs/pii/S0269749118353685 [Accessed 13 December 2023].

Derrick, B., Ruck, A., Toher, D. & White, P. 2018. Tests for equality of variances between two samples which contain both paired observations and independent observations. *Journal of Applied Quantitative Methods*. 13, 2, 36-47. E-journal. Available: https://www.jaqm.ro/issues/volume-13,issue-2/3_BEANDEPA.PHP [Accessed 13 January 2024].

Echoview. 2023a. About Echoview. Web page. Available: https://support.echoview.com/WebHelp/_Introduction/About_Echoview.htm [Accessed 14 December 2023].

Echoview. 2023b. Configuring a bottom classification. Web page. Available: https://support.echoview.com/WebHelp/How_To/Classify_Bottoms/Configuring_a_bottom_classification.htm#The_beam_and_seabed [Accessed 14 December 2023].

Echoview. 2023c. Depth normalization reference depth algorithms. Web page. Available: https://support.echoview.com/WebHelp/Reference/Algorithms/Bottom_Classifcation/Bottom_classification_algorithms.htm#Depth_normalization_algorithms [Accessed 14 December 2023].

Echoview. 2023d. First_bottom_length_normalized. Web page. Available: https://support.echoview.com/WebHelp/Analysis_Variables/Bottom/First_bottom_length_normalized.htm [Accessed 14 December 2023].

Echoview. 2023e. Bottom_rise_time_normalized. Web page. Available: https://support.echoview.com/WebHelp/Analysis_Variables/Bottom/Bottom_rise_time_normalized.htm [Accessed 14 December 2023].

Echoview. 2023f. Bottom_line_depth_mean. Web page. Available: https://support.echoview.com/WebHelp/Analysis_Variables/Bottom/Bottom_line_depth_mean.htm [Accessed 14 December 2023].

Echoview. 2023g. Second_bottom_length_normalized. Web page. Available: https://support.echoview.com/WebHelp/Analysis_Variables/Bottom/Second_bottom_length_normalized.htm [Accessed 14 December 2023].

Echoview. 2023h. Bottom_max_Sv. Web page. Available: https://support.echoview.com/WebHelp/Analysis_Variables/Bottom/Bottom_max_Sv.htm [Accessed 14 December 2023].

Echoview. 2023i. Algorithms – Bottom Classification. Web page. Available: https://support.echoview.com/WebHelp/Reference/Algorithms/Bottom_Classifcation/Bottom_classification_algorithms.htm#PCA_dimensions [Accessed 14 December 2023].

Echoview. 2023j. Detecting a vegetation boundary. Web page. Available: https://support.echoview.com/WebHelp/How_To/Use_The_Threshold_Offset_Operator/Detecting_a_vegetation_boundary.htm [Accessed 14 November 2023].

Enrichetti, F., Bava, S., Bavestrello, G., Betti, F., Lanteri, L. & Bo, M., 2019. Artisanal fishing impact on deep coralligenous animal forests: a Mediterranean case study of marine vulnerability. *Ocean Coast. Manag*. 177, 112–126. E-journal. Available: https://www.sciencedirect.com/science/article/abs/pii/S0964569118309530 [Accessed 12 December 2023].

EU 92/43/EEC. 1992. Council directive on the conservation of natural habitats and of wild fauna and flora. PDF Document. Available: https://eur-

lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31992L0043 [Accessed 10 December 2023].

EU 2008/56/EC. Directive the European Parliament and of the Council establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). PDF Document. Available: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32008L0056 [Accessed 10 December 2023].

Fajaryanti, R. & Kang, M. 2019. A preliminary study on seabed classification using a scientific echosounder. *Journal of the Korean Society of Fisheries Technology*. 55, 1, 39–49. E-Journal. Available: https://koreascience.kr/article/JAKO201912964894514.pdf [Accessed 10 December 2023].

Fakiris, E., Dimas, X., Giannakopoulos, V., Geraga, M., Koutsikopoulos, C., Ferentinos, G. & Papatheodorou, G. 2023. Improved predictive modelling of coralligenous formations in the Greek Seas incorporating large-scale, presence–absence, hydroacoustic data and oceanographic variables. *Frontiers in Marine Science*. 10. E-journal. Available: https://www.frontiersin.org/articles/10.3389/fmars.2023.1117919/full [12 December 2023].

Fauziyah, A., Purwiyanto, I., S, Agustriani, F., Putri, W., A., E., Liyani, M., Aryawati, R., Ningsih, E., N. & Suteja, Y. 2020. Detection of bottom substrate type using singlebeam echo sounder backscatter: a case study in the east coastal of Banyuasin. *IOP Conf. Ser.: Earth Environ. Sci*. 404, 012004. E-journal. Available: https://iopscience.iop.org/article/10.1088/1755-1315/404/1/012004 [Accessed 17 December 2023].

Ferrigno, F., Appolloni, L., Russo, G.F., Sandulli, R., 2017a. Impact of fishing activities on different coralligenous assemblages of Gulf of Naples (Italy). *J. Mar. Biol. Assoc.* E-journal. Available: https://www.researchgate.net/publication/318024292_Impact_of_fishing_activities_on_different_coralligenous_assemblages_of_Gulf_of_Naples_Italy [Accessed 13 December 2023]

Fritz, S., See, L., Carlson, T., Haklay, M., Oliver, J., L., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L., A., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., He, S., Hepburn, L., Hsu, A., Long, D., Masó, J., McCallum, I., Muniafu, M., Moorthy, I., Obersteiner, M., Parker, A., J., Weisspflug, M. & West, S. 2019. Citizen science and the United Nations Sustainable Development Goals. *Nat Sustain*. 2, 922–930. E-journal. Available: https://www.nature.com/articles/s41893-019-0390-3 [Accessed 20 December 2023].

Georgiadis, M., Papatheodorou, G., Tzanatos, E., Geraga, M., Ramfos, A., Koutsikopoulos, C. & Ferentinos, G. 2009. Coralligène formations in the eastern Mediterranean Sea: Morphology, distribution, mapping and relation to fisheries in the southern Aegean Sea (Greece) based on high-resolution acoustics. *Journal*

of Experimental Marine Biology and Ecology. 368, 1, 44-58. E-journal. Available: https://www.sciencedirect.com/science/article/pii/S0022098108004905 [Accessed 10 December 2023].

Giakoumi, S., Sini, M., Gerovasileiou, V., Mazor, T., Beher, J., Possingham, H., P., Abdulla, A., Çinar, M., E., Dendrinos, P., Gucu, A., C., Karamanlidis, A., A., Rodic, P., Panayotidis, P., Taskin, E., Jaklin, A., Voultsiadou, E., Webster, C., Zenetos, A. & Katsanevakis, S. 2013. Ecoregion-Based Conservation Planning in the Mediterranean: Dealing with Large-Scale Heterogeneity. *PLoS ONE*. 8, 10. E-journal. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0076449 [Accessed 10 December 2023].

Gili, J., M. & Ros, J. 1985. Study and cartography of the benthic communities of Medes islands (NE Spain). *Pubblicazioni della Stazione Zoologica di Napoli I: Marine Ecology*. 6, 219–238. E-journal. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1439-0485.1985.tb00323.x [Accessed 10 December 2023].

Gilbertson, D., D., Kent, M. & Pyatt, F., B. 1985. Data analysis and interpretation I: introduction and the Mann- Whitney U test. *Practical Ecology for Geography and Biology*. 197-207. E-book. Available: https://link.springer.com/chapter/10.1007/978-1-4684-1415-8_11 [Accessed 15 January 2024].

Giovos, I., Kleitou, P., Poursanidis, D., Batjakas, I., Bernardi, G., Crocetta, F., Doumpas, N., Kalogirou, S., E., Kampouris, T., E., Keramidas, I., Langeneck, J., Maximiadi, M., Mitsou, E., Stoilas, V., Tiralongo, F., Romanidis-Kyriakidis, G., Xentidis, N., Zenetos, A. & Katsanevakis, S. 2019. Citizen-science for monitoring marine invasions and stimulating public engagement: a case project from the eastern Mediterranean. *Biol. Invasions*. 21, 3707–3721. E-journal. Available: https://bernardi.eeb.ucsc.edu/wp-content/uploads/2019/09/2019_Citizen_Science.pdf [Accessed 19 December 2023].

Gómez-Gras, D., Linares, C., Dornelas, M., Madin, J.S., Brambilla, V., Ledoux, J.-B., López-Sendino, P., Bensoussan, N. & Garrabou, J. 2021. Climate change transforms the functional identity of Mediterranean coralligenous assemblages. *Ecology Letters*. 24, 1038-1051. E-journal. Available: https://onlinelibrary.wiley.com/doi/10.1111/ele.13718 [Accessed 16 December 2023].

Google. 2024a. Classification: Accuracy. Web page. Available: https://developers.google.com/machine-learning/crash-course/classification/accuracy [Accessed 4 February 2024].

Google. 2024b. Classification: Precision and Recall. Web page. Available: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall [Accessed 4 February 2024].

Halpern, B., S., Frazier, M., Afflerbach, J., Lowndes, J., S., Micheli, F., O'Hara, C., Scarborough, C. & Selkoe, K., A. 2019. Recent pace of change in human impact on the world's ocean. *Scientific Reports*. 9, 11609. E-journal. Available: https://www.nature.com/articles/s41598-019-47201-9 [Accessed 10 December 2023].

Hong, J., S. 1982. Contribution à l'étude des peuplements d'un fond coralligène dans la région marseillaise en Méditerranée Nord-Occidentale. *Bulletin of Korea Ocean Research and Development Institute*. 4, 27–51.

Hong, H., Xiaoling, G. & Hua Yu. 2016. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. *7th IEEE International Conference on Software Engineering and Service Science*. PDF Document. Available: https://ieeexplore.ieee.org/abstract/document/7883053 [Accessed 23 January 2024].

Ingrassia, M., Martorelli, E., Sañé, E., Falese, F., G., Bosman, A., Bonifazi, A., Argenti, L. & Chiocci, F., L. 2019. Coralline algae on hard and soft substrata of a temperate mixed siliciclastic-carbonatic platform: Sensitive assemblages in the Zannone area (western Pontine Archipelago; Tyrrhenian Sea). *Marine Environmental Research*. 147, 1–12. E-journal. Available: https://support.echoview.com/WebHelp/Reference/Algorithms/Bottom_Classifcation/Bottom_classification_algorithms.htm#PCA_dimensions [Accessed 14 December 2023].

Jackson-Bué, T., Williams, G., J., Whitton, T., A., Roberts, M., J., Brown, A., G. & Amir, H. 2022. Seabed morphology and bed shear stress predict temperate reef habitats in a high energy marine region. *Estuar. Coast. Shelf Sci*. 274, 107934. E-journal. Available: https://www.sciencedirect.com/science/article/pii/S0272771422001925 [Accessed 10 December 2023].

Kongsberg. 2023. EK80 - End user manual. PDF Document. Available: https://www.simrad.net/ek80/documents.htm [Accessed 21 December 2023]

Laborel, J. 1961. Le concretionnement algal "coralligène" et son importance géomorphologique en Méditerranée. *Recueil des Travaux de la Station Marine d'Endoume*. 23 (37), 37–60. E-journal. Available: http://paleopolis.rediris.es/benthos/REF/som/R-pdf/1961-23-37_37.pdf [Accessed 13 December 2023].

Laborel, J. 1986. Marine biogenic constructions in the mediterranean. A review. *Trav. Sci. du Parc Natl. Port-Cros*. 281–310. E-journal. Available: https://portals.iucn.org/library/node/12349 [Accessed 10 December 2023].

Linders, T., Nilsson, P., Wikström, A. & Sköld, M. 2018. Distribution and fate of trawling-induced suspension of sediments in a marine protected area. *ICES Journal of Marine Science*. 75, 2, 785–795. E-journal. Available:

https://academic.oup.com/icesjms/article/75/2/785/4683659 [Accessed 12 December 2023]

Ludvigsen, M., Sortland, B., Johnsen, G. & Singh, H. 2007. Applications of Geo-Referenced Underwater Photo Mosaics in Marine Biology and Archaeology. *Oceanography*. 20, 4. Available: https://www.researchgate.net/publication/33549713_Applications_of_Geo-Referenced_Underwater_Photo_Mosaics_in_Marine_Biology_and_Archaeology [Accessed 17 December 2023].

Marion, A., F. 1883. Esquisse d'une topographie zoologique du Golfe de Marseille. *Annales Musée d'Histoire Naturelle Marseille*. 1, 1–108. PDF Document. Available: https://books.google.fi/books?hl=en&lr=&id=QAgLAQAAIAAJ&oi=fnd&pg=PA3&dq=Marion,+A.F.+1883.+Esquisse+d%E2%80%99une+topographie+zoologique+du+Golfe+de+Marseille.+Annales+Mus%C3%A9e+d%E2%80%99Histoire+Naturelle+Marseille+1,+1%E2%80%93108&ots=SAmCT1GZR1&sig=DYBwsuDUNHY9AilRv6n98MTrehY&redir_esc=y#v=onepage&q&f=false [Accessed 13 December 2023].

Martin, C., S., Giannoulaki, M., De Leo, F., Scardi, M., Salomidi, M., Knitweiss, L., Pace, M., L., Garofalo, G., Gristina, M., Ballesteros, E., Bavestrello, G., Belluscio, A., Cebrian, E., Gerakaris, V., Pergent, G., Pergent-Martini, C., Schembri, P.J., Terribile, K., Rizzo, L., Ben Souissi, J., Bonacorsi, M., Guarnieri, G., Krzelj, M., Macic, V., Punzo, E., Valavanis, V. & Fraschetti, S. 2014. Coralligenous and maërl habitats: Predictive modelling to identify their spatial distributions across the mediterranean sea. *Scientific Reports - Nature*. 4, 5073. E-journal. Available: https://www.nature.com/articles/srep05073.pdf [Accessed 13 December 2023].

McCauley, R., D. & Siwabessy, P., J., W. 2006. Practical guide to acoustic techniques for benthic habitat classification. *Cooperative Research Centre for Coastal Zone, Estuary & Waterway Management*. E-book. Available: https://ozcoasts.org.au/wp-content/uploads/pdf/CRC/84_practical_benthic_classification_screen.pdf [Accessed 10 December 2023].

McConnaughey, R., A., Hiddink, J., G., Jennings, S., Pitcher, C., R., Kaiser, M., J., Suuronen, P., Sciberras, M., Rijnsdorp, A., D., Collie, J., S., Mazor, T., O Amoroso, R., Parma, A., M. & Hilborn R. 2019. Choosing best practices for managing impacts of trawl fishing on seabed habitats and biota. *Fish and Fisheries*. 21, 2, 319-337. E-journal. Available: https://onlinelibrary.wiley.com/doi/10.1111/faf.12431 [Accessed 25 January 2024].

Naasan Aga Spyridopoulou, R., Langeneck, J., Bouziotis, D., Giovos, I., Kleitou, P. & Kalogirou, S. 2020. Filling the Gap of Data-Limited Fish Species in the Eastern Mediterranean Sea: A Contribution by Citizen Science. *Journal of Marine Science and Engineering*. 8, 2, 107. E-journal. Available: https://www.mdpi.com/2077-1312/8/2/107 [Accessed 19 December 2023].

Nevstad, M., B. 2022. Use of different imaging systems for ROV-based mapping of complex benthic habitats - Master's thesis in Ocean Resources. *Norwegian University of Science and Technology*.PDF Document. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3047578 [Accessed 20 December 2023].

Olson, D., L. & Delen, D. 2008. Advanced Data Mining Techniques. DBLP. E-book. Available: https://www.researchgate.net/publication/220695151_Advanced_Data_Mining_Techniques [Accessed 30 January 2024].

Pearman, T., R., R., Robert, K., Callaway, A., Hall, R., Lo Iacono, C. & V., A. I. Huvenne. 2020. Improving the predictive capability of benthic species distribution models by incorporating oceanographic data – Towards holistic ecological modelling of a submarine canyon. *Progress in Oceanography*. 184, 102338. E-journal. Available: https://www.sciencedirect.com/science/article/pii/S007966112030077X [Accessed 10 December 2023].

Penrose, J., D., Siwabessy, P., J., W., Gavrilov, A., N. & Parnum, I. 2006. Acoustic techniques for Seabed Classification. *CRC for Coastal Zone, Estuary & Waterway Management*. E-book. Available: https://www.researchgate.net/publication/236952024_Acoustic_Techniques_for_Seabed_Classification [Accessed 13 December 2023].

Pérès, J. & Picard, J., M. 1964. Nouveau manuel de bionomie benthique de la mer Méditerranée. *Recueil des Travaux de la Station Marine d'Endoume*. 31(47), 1–131. E-book. Available: http://paleopolis.rediris.es/benthos/REF/som/R-pdf/Manuel_bionomie_benthique_1964.pdf [Accessed 10 December 2023].

Pierdomenico, M., Bonifazi, A., Argenti, L., Ingrassia, M., Casalbore, D., Aguzzi, L., Viaggiu, E., Le Foche, M. & Chiocci, F., L. 2021. Geomorphological characterization, spatial distribution and environmental status assessment of coralligenous reefs along the Latium continental shelf. *Ecological Indicators*. 131. E-journal. Available: https://www.sciencedirect.com/science/article/pii/S1470160X21008840 [Accessed 14.12.2023].

Pietroluongo, G., Martín-Montalvo, B.,Q., Ashok, K., Miliou, A., Fosberry, J., Antichi, S., Moscatelli, S., Tsimpidis, T., Carlucci, R. & Azzolin, M. 2022. Combining Monitoring Approaches as a Tool to Assess the Occurrence of the Mediterranean Monk Seal in Samos Island, Greece. *Hydrobiology.* 1, 4, 440-450. E-journal. Available: https://www.mdpi.com/2673-9917/1/4/26 [Accessed 20 December 2023].

Riedl, R. 1966. Biologie der Meereshöhlen. *Limnology and Oceanography*. 12, 4, 725-726. E-book. Available: https://aslopubs.onlinelibrary.wiley.com/doi/10.4319/lo.1967.12.4.0725 [Accessed 10 December 2023].

Rodriguez-Galiano V.,F., Ghimire, B., Rogan, J., Chica-Olmo, M. & Rigol-Sanchez, J., P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*. 67, 93-104. E-journal. Available: https://www.sciencedirect.com/science/article/abs/pii/S0924271611001304 [Accessed 17 January 2024].

Sabat-Tomala, A., Raczko, E. & Zagajewski, B. 2020. Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data. *Remote Sens*. 12, 516. E-journal. Available: https://www.mdpi.com/2072-4292/12/3/516 [Accessed 25 January 2024].

Sálek, M., Cervinka, J., Banea, O., Krofel, M. 2014. Population densities and habitat use of the golden jackal (Canis aureus) in farmlands across the Balkan Peninsula. *European Journal of Wildlife Research*. 60, 2, 193-200. Available: https://www.researchgate.net/publication/257916882_Population_densities_and_habitat_use_of_the_golden_jackal_Canis_aureus_in_farmlands_across_the_Balkan_Peninsula [Accessed 17 January 2024].

Salomidi, M., Katsanevakis, S., Borja, A., Braeckman, U., Damalas, D., Galparsoro, I., Mifsud, R., Mirto, S., Pascual, M., Pipitone, C., Rabaut, M., Todorova, V., Vassilopoulou, V. & Vega Fernandez, T. 2012. Assessment of goods and services, vulnerability, and conservation status of European seabed biotopes: a stepping stone towards ecosystem-based marine spatial management. *Mediterranean Marine Science*, 13, 1. E-journal. Available: https://ejournals.epublishing.ekt.gr/index.php/hcmr-med-mar-sc/article/view/11996 [Accessed 10 December 2023].

Scikit-Learn. 2024. RandomForestClassifier. Web page. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html [Accessed 15 November 2023]

Serandour, J., Willison, J., Thuiller, W. & Ravanel, P. 2010. Environmental drivers for Coquillettidia mosquito habitat selection: A method to highlight key field factors. *Hydrobiologia*. 652, 1. E-journal. Available: https://www.researchgate.net/publication/48417744_Environmental_drivers_for_Coquillettidia_mosquito_habitat_selection_A_method_to_highlight_key_field_factors [Accessed 17 January 2024].

Sini, M., Katsanevakis, S., Koukourouvli, N., Gerovasileiou, V., Dailianis, T., Buhl-Mortensen, L., Damalas, D., Dendrinos, P., Dimas, X., Frantzis, A., Gerakaris, V., Giakoumi, S., Gonzalez-Mirelis, G., Hasiotis, T., Issaris, Y., Kavadas, S., G., Koutsogiannopoulos, D., D., Koutsoubas, D., Manoutsoglou, E., Markantonatou, V., Mazaris, A., D., Poursanidis, D., Papatheodorou, G., Salomidi, M., Topouzelis, K., Trygonis, V., Vassilopoulou, V. & Zotou, M. 2017. Assembling ecological pieces to reconstruct the conservation puzzle of the aegean sea. *Frontiers in Marine Science*. 4. E-journal. Available:

https://www.frontiersin.org/articles/10.3389/fmars.2017.00347/full [Accessed 10 December 2023].

Siwabessy, J., Penrose, J., Kloser, R. & Ross, D.1999. Seabed Habitat Classification. Shallow Survey '99 - International Conference on High Resolution Surveys in Shallow Water. Sydney, Australia. PDF Document. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b30825676a8f9e1bf6b2a65029d7710bf49b2592 [Accessed 10 December 2023].

Sørensen A.,J, Ludvigsen M., Norgren P., Ødegård Ø. & Cottier, F. 2020. Chapter 9. Sensor Carrying Platforms - In: Berge J., Johnsen G., Cohen J.,H. Polar Night Marine Ecology. 4, 241-275. E-book. Available: https://link.springer.com/chapter/10.1007/978-3-030-33208-2_9 [Accessed 19 December 2023].

Steller, D., L., Hernandez-Ayon, J., M., Riosmena-Rodriguez, R.& Cabello-Pasini, A. 2007. Effect of temperature on photosynthesis, growth and calcification rates of the free-living coralline alga Lithophyllum margaritae. *Ciencias Marinas*. 33, 4. E-journal. Available: https://cienciasmarinas.com.mx/index.php/cmarinas/article/view/1255 [Accessed 9 December 2023].

Teixidó N, Garrabou J, Harmelin JG (2011) Low Dynamics, High Longevity and Persistence of Sessile Structural Species Dwelling on Mediterranean Coralligenous Outcrops. *PLOS ONE* .6, 8, e23744. E-journal. Available: https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0023744&type=printable [Accessed 10 December 2023].

UNEP-MAP-RAC/SPA. 2008. Action plan for the conservation of the coralligenous and other calcareous bio-concretions in the Mediterranen Sea. *RAC/SPA*. PDF Document. Available: https://racspa.org/sites/default/files/action_plans/pacoralligene.pdf [Accessed 10 December 2023].

UNEP-MAP-RAC/SPA. 2022. Proceedings of the 4th Mediterranean Symposium on the conservation of Coralligenous & other Calcareous Bio-Concretions (Genova, Italy, 20-21 September 2022). *RAC/SPA*. PDF Document. Available: https://www.racspa.org/sites/default/files/proceedings/proceedings_mscc_2022_f.pdf [Accessed 20 December 2023].

Vlachopoulou, E., I., Meriwether Wilson, A. & Miliou, A. Disconnects in EU and Greek fishery policies and practices in the eastern Aegean Sea and impacts on Posidonia oceanica meadows. *Ocean & Coastal Management*, 76, 105-113. E-journal. Available: https://www.sciencedirect.com/science/article/abs/pii/S0964569113000331?via%3Dihub [Accessed 19 December 2023].

Zunino, S., Melaku Canu, D., Zupo, V. & Solidoro, C. 2019. Direct and indirect impacts of marine acidification on the ecosystem services provided by coralligenous reefs and seagrass systems. *Global Ecology and Conservation.* 18,

e00625. E-journal. Available: https://www.sciencedirect.com/science/article/pii/S2351989418305328 [Accessed 16 December 2023].

LIST OF FIGURES

LIST OF TABLES

Example of the ESW dataset (F_ESW_Y&N.csv)

Point_index	Coralligeno	Point_millisecond	Latitude	Class_name	Bottom_roughness_normalized	Bottom_hardness_normalized	First_bott	Second_b	Bottom_rise_time_normalize	Bottom_li	Bottom_m	Bottom_ku	Bottom_sk
0	N	382.05.00	37.5944852681	Bottom class 1	7.865734	3.048437	4.026018	0.154935	0.892408	97.739116	-18.53298	9.033163	2.795524
1	N	139.05.00	37.5944756125	Bottom class 1	7.881616	3.476885	3.970665	0.153979	0.893450	97.842363	-18.12393	9.536189	2.905447
2	N	896.05.00	37.5944662040	Bottom class 1	7.851703	3.709951	4.077405	0.156967	0.953992	97.968581	-18.64215	9.995604	2.824754
3	N	641.00.00	37.5944582690	Bottom class 1	7.860806	2.878844	4.237862	0.084355	0.933152	97.837293	-18.93627	8.670569	2.754603
4	N	388.00.00	37.5944502724	Bottom class 1	7.795693	3.286642	4.152780	0.149835	0.918614	98.009849	-19.78370	7.913313	2.648367
5	N	150.00.00	37.5944440741	Bottom class 1	7.844697	3.669939	4.078363	0.111050	0.953562	98.355621	-19.01349	8.219043	2.651524
6	N	893.05.00	37.5944308599	Bottom class 1	7.840456	3.224186	4.111589	0.109242	0.835055	98.290291	-18.57747	10.792076	2.975684
7	N	650.00.00	37.5944287449	Bottom class 1	7.801456	3.051083	4.037269	0.175960	1.000468	98.313197	-19.59126	8.400011	2.665797
8	N	401.05.00	37.5944225878	Bottom class 1	7.788088	2.965799	4.299961	0.137101	1.122288	98.368172	-20.01718	8.097575	2.654506
9	N	154.00.00	37.5944141745	Bottom class 1	7.847591	2.702177	4.028091	0.072685	0.964513	98.423042	-18.85712	9.406637	2.837608
10	N	912.05.00	37.5944069079	Bottom class 1	7.875592	2.876447	4.114594	0.113309	0.797493	98.477995	-18.12604	12.661696	3.190146
11	N	660.00.00	37.5943988524	Bottom class 1	7.861781	3.095977	4.153757	0.172565	1.029411	98.532818	-18.29445	11.594156	3.084212
12	N	407.00.00	37.5943891173	Bottom class 1	7.883109	3.241725	4.174970	0.226283	1.093861	98.587669	-18.49758	11.386691	2.942217
13	N	164.05.00	37.5943794709	Bottom class 1	7.829426	3.022562	4.107771	0.166949	1.043303	98.642602	-19.06336	9.293597	2.774067
14	N	911.00.00	37.5943710243	Bottom class 1	7.874069	2.962997	4.081539	0.150592	1.068365	98.697417	-18.53565	10.553411	2.959199
15	N	663.00.00	37.5943616818	Bottom class 1	7.833769	3.078305	4.144610	0.178276	1.104816	98.752292	-18.44586	13.560903	3.309944
16	N	413.05.00	37.5943537912	Bottom class 1	7.879233	3.417613	4.029555	0.248353	1.206603	98.807163	-18.35697	10.364763	2.970599
17	N	161.00.00	37.5943450381	Bottom class 1	7.873525	2.957436	3.987792	0.140881	1.157005	98.861971	-18.38458	10.275045	2.978523
18	N	920.00.00	37.5943346482	Bottom class 1	7.817832	2.997848	4.183558	0.179482	1.245460	98.917081	-19.21902	11.658469	3.041664
19	N	692.05.00	37.5943250526	Bottom class 1	7.842290	2.951305	3.970549	0.100887	1.362024	98.972157	-19.11132	9.200528	2.740282
20	N	440.00.00	37.5943112861	Bottom class 2	7.687680	2.873642	4.162756	0.112893	1.143761	99.026952	-20.90980	7.685408	2.581607
21	N	200.05.00	37.5943057243	Bottom class 1	7.752456	3.124135	4.105566	0.090512	1.207659	99.063119	-19.93793	9.576929	2.809776
22	N	949.00.00	37.5942962168	Bottom class 1	7.794584	2.830269	3.943748	0.065491	0.941644	99.084308	-20.06981	6.902201	2.443659
23	N	696.05.00	37.5942870894	Bottom class 1	7.767889	3.069615	4.021125	0.102677	0.856558	99.105517	-19.66007	9.358127	2.804358
24	N	453.05.00	37.5942771939	Bottom class 1	7.822408	2.896673	4.096160	0.087661	1.007382	99.126731	-19.37899	7.313036	2.516299
25	N	203.00.00	37.5942673276	Bottom class 1	7.830150	2.926700	3.994287	0.066089	0.963170	99.147925	-18.85807	9.150006	2.735848
26	N	961.00.00	37.5942570464	Bottom class 2	7.729405	2.925405	4.049814	0.063267	0.896451	99.169156	-20.20971	9.454042	2.677768
27	N	709.05.00	37.5942482248	Bottom class 1	7.788199	3.151059	3.813584	0.071224	0.786808	99.190360	-19.21448	8.634815	2.720194
28	N	457.00.00	37.5942388206	Bottom class 1	7.796332	3.213403	3.769089	0.080272	1.024834	99.211543	-19.31411	8.002477	2.579989
29	N	218.05.00	37.5942289205	Bottom class 1	7.859588	3.145845	3.669144	0.063553	0.908001	99.232800	-18.64196	7.190967	2.529293
30	N	965.00.00	37.5942225781	Bottom class 1	7.847507	3.870470	3.671092	0.130991	0.774035	99.253984	-18.54436	8.897910	2.670390
31	N	710.00.00	37.5942133744	Bottom class 1	7.868541	3.103050	3.914779	0.088201	0.831955	99.275173	-17.78161	13.034980	3.256750
32	N	473.00.00	37.5942055268	Bottom class 1	7.868857	3.212163	3.815826	0.104579	0.674598	99.293648	-18.51349	7.623862	2.604339

Example of the N dataset (F_N_Y&N.csv)

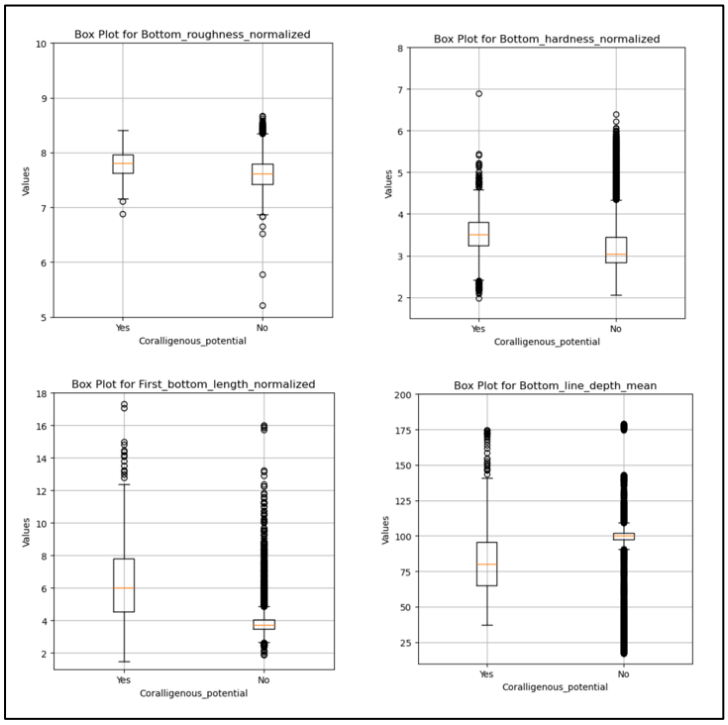
Point_index	Coralligenous_potential	Point_date	Latitude	Longitude	Bottom_roughness_normalized	Bottom_hardness_normalized	First_bottom_length_normalized	Second_b	Bottom_ri	Bottom_line_depth_mean	Bottom_m	Bottom_ku	Bottom_sk
0	N	26/09/2023	37.5808719423	26.4737614829	7.773079	4.226377	4.587075	1.524698	0.916105	50.352399	-19.73770	9.878164	2.962957
1	N	26/09/2023	37.5809046406	26.4737347639	7.722435	4.185074	4.236967	1.151538	0.831055	50.690521	-19.68451	11.646005	3.116698
2	N	26/09/2023	37.5809289488	26.4737103789	7.686025	4.33744	4.324232	1.450060	0.876286	50.692891	-20.46177	9.006593	2.812189
3	N	26/09/2023	37.580963872	26.4736830996	7.734793	4.262814	4.385763	1.555774	0.928611	50.748361	-20.05053	10.114434	2.924670
4	N	26/09/2023	37.5809941466	26.473659016	7.66597	4.390345	4.413362	1.730058	1.055182	50.778569	-20.97525	8.002331	2.720180
5	N	26/09/2023	37.5810261614	26.4736326152	7.720892	4.223492	4.360497	1.860455	1.219485	50.722322	-20.06372	12.023135	3.148643
6	N	26/09/2023	37.581053866	26.4736100821	7.724497	4.094649	4.357391	1.333647	1.061713	51.061815	-19.81484	10.411462	3.038559
7	N	26/09/2023	37.5810878048	26.4735799143	7.585707	3.690883	4.142312	0.945078	1.058360	51.165771	-21.31448	10.820326	2.978524
8	N	26/09/2023	37.5811173615	26.4735567898	7.535936	3.909997	3.821861	1.114048	0.860252	51.450403	-22.52293	7.980536	2.685799
9	N	26/09/2023	37.5811496474	26.4735287468	7.590993	4.065161	4.454063	1.444281	0.975326	51.189744	-21.23770	10.025200	2.971021
10	N	26/09/2023	37.5811798811	26.4735020951	7.580635	4.082938	4.056132	1.376348	0.945218	51.399381	-22.30008	9.722975	2.790696
11	N	26/09/2023	37.581210873	26.4734757937	7.632898	3.914196	4.702839	1.297245	0.943452	51.163296	-21.80903	7.143156	2.566582
12	N	26/09/2023	37.5812405263	26.4734487915	7.636876	3.94335	4.480707	1.426295	0.987205	51.293124	-21.04501	8.714165	2.840486
13	N	26/09/2023	37.5812726033	26.4734204867	7.588711	3.974955	4.532045	1.339124	0.862065	51.513303	-21.52873	10.447389	3.027963
14	N	26/09/2023	37.5812986913	26.4733968149	7.626372	4.10598	4.644253	1.652226	0.970162	51.407783	-21.75671	10.063529	2.922107
15	N	26/09/2023	37.5813355556	26.4733665556	7.703502	4.143516	4.883653	1.477324	1.144871	51.539001	-21.22594	7.899634	2.692126
16	N	26/09/2023	37.5813616386	26.4733429018	7.822352	4.118717	4.830351	1.220225	1.047571	51.48287	-19.94643	7.744867	2.642475
17	N	26/09/2023	37.5813982883	26.4733111512	7.818803	4.15666	4.716159	1.390713	1.060973	51.570752	-19.13449	11.846809	3.174822
18	N	26/09/2023	37.581426903	26.4732864304	7.862409	4.243017	4.937389	1.312404	1.134602	51.425012	-18.96138	10.987577	3.122154
19	N	26/09/2023	37.5814591524	26.4732569295	7.698855	4.31542	4.98761	1.412201	1.047339	51.572658	-21.13647	9.398173	2.822541
20	N	26/09/2023	37.5814860224	26.4732335154	7.820866	4.362523	4.553831	1.337258	0.899730	51.543459	-18.58859	13.720680	3.371396
21	Y	26/09/2023	37.5815213229	26.4732030091	7.925867	4.567018	5.021121	1.625571	1.266411	51.407122	-18.19779	12.184599	3.206945
22	N	26/09/2023	37.5815474385	26.4731775503	7.942091	4.588151	4.922571	1.455214	1.015377	51.507861	-17.62488	13.902466	3.355504
23	N	26/09/2023	37.5815806444	26.4731467481	7.874131	4.529529	4.633282	1.830938	1.096863	51.349283	-18.58279	11.439406	3.180055
24	N	26/09/2023	37.5816059467	26.4731243386	7.763067	4.44619	4.721391	1.573334	1.004316	51.469597	-19.77551	10.636904	3.060989
25	N	26/09/2023	37.5816387196	26.4730912804	7.745459	4.299801	4.486719	1.607234	1.203458	51.28374	-20.19919	9.223007	2.793916
26	N	26/09/2023	37.5816684545	26.4730632121	7.812884	4.359922	4.378675	1.560017	1.065025	51.35169	-19.24738	9.356028	2.883680
27	N	26/09/2023	37.5817015603	26.473031773	7.543107	4.184835	4.28806	1.604095	0.740627	51.41273	-22.41855	7.265432	2.592136
28	N	26/09/2023	37.5817583156	26.4729766844	7.624228	3.915848	4.527311	1.370413	0.818241	51.051722	-21.40663	8.587040	2.750332
29	N	26/09/2023	37.5818160123	26.4729189877	7.601069	4.033832	4.348874	1.690437	1.007344	50.645211	-21.68951	9.181624	2.814405
30	N	26/09/2023	37.5818761486	26.472860518	7.625858	4.010301	4.561811	1.471949	0.777537	50.335947	-21.23891	9.115085	2.862233
31	N	26/09/2023	37.5819321274	26.4728040349	7.56404	3.868364	4.419227	1.526545	1.307722	49.697521	-23.09444	8.133302	2.541626
32	N	26/09/2023	37.5819946517	26.4727449801	7.266455	3.561836	3.980897	0.875293	0.656562	49.559619	-23.92983	9.563775	2.927598
33	N	45195	37.582054247	26.4726890864	7.552959	3.778578	5.110758	1.646826	1.201858	47.657149	-22.53582	8.435794	2.834668

Descriptive statistics for the selected normalized features of the bottom used for developing the RF model (first-day survey dataset, F_ESW_Y&N.csv).

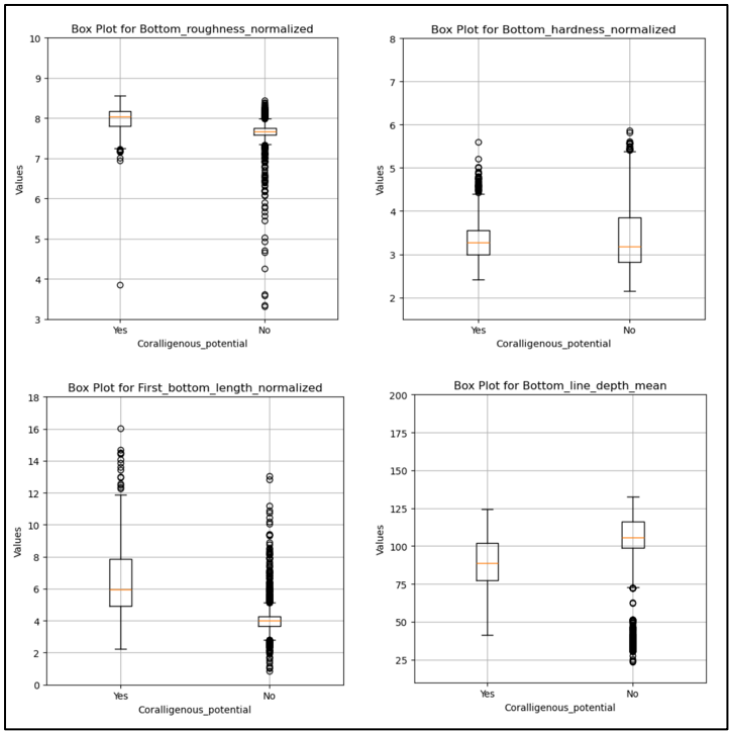
Metric	Bottom roughness (sv)	Bottom hardness (sv)	First bottom length (m)	Line depth mean (m)
count	6413	6413	6413	6413
mean	7.64	3.32	4.50	93.61
std	0.29	0.70	4.98	22.16
min	3.76	1.97	1.47	17.01
25%	7.44	2.86	3.51	95.53
50%	7.64	3.10	3.80	100.01
75%	7.82	3.56	4.23	102.03
max	8.67	6.89	18.86	179.01

Descriptive statistics for the selected normalized features of the bottom used to run the RF model for predicting potential coralligenous presence (second-day survey dataset, F_N_Y&N.csv).

Metric	Bottom roughness (sv)	Bottom Hardness (sv)	First bottom length (sv)	Line depth mean (sv)
count	4382	4382	4382	4382
mean	7.76	3.35	4.83	98.18
std	0.34	0.64	1.75	20.71
min	3.31	2.14	0.85	23.68
25%	7.61	2.88	3.84	88.62
50%	7.71	3.22	4.21	103.48
75%	7.91	3.70	5.21	110.59
max	8.56	5.86	18.58	132.81



Boxplots for the selected normalized features. Highlighting Coralligenous Potential Yes and No values, used to develop the RF Classifier with the first-day survey dataset (F_ESW_Y&N.csv, 6413 datapoints).



Boxplots for selected features. Highlighting Coralligenous Potential Yes and No values which were predicted by the RF Classifier when used on the second-day survey dataset (F_N_Y&N.csv, 4382 datapoints).

Levene test results for both ESW and N datasets

Feature	Dataset	Test statistics	p-value
Bottom Roughness	ESW	5.353930274862484	0.02
Bottom Hardness	ESW	8.924208543169655	<0.001
First Bottom Length	ESW	38.48143890406763	<0.001
Bottom line depth mean	ESW	149.0243602447999	<0.001
Bottom Roughness	N	120.09655462051957	<0.001
Bottom Hardness	N	221.88642226202148	<0.001
First Bottom Length	N	1962.4283324930132	<0.001
Bottom line depth mean	N	79.91344883278796	<0.001

Mann-Whitney U test results for both ESW and N datasets

Feature	Dataset	Test statistics	p-value
Bottom Roughness	ESW	3456307.5	<0.001
Bottom Hardness	ESW	3451716.0	<0.001
First Bottom Length	ESW	4275129.5	<0.001
Bottom line depth mean	ESW	1217997.5	<0.001
Bottom Roughness	N	3272499.5	<0.001
Bottom Hardness	N	2087463.5	<0.001
First Bottom Length	N	3672829.5	<0.001
Bottom line depth mean	N	719267.0	<0.001