This article has been accepted for publication in a future workshop of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/MetroSea.12345.2025.1234567, 2025 IEEE International Workshop on Metrology for the Sea

Implementation of Machine – Learning Models to study Habitat Suitability for Deep-Diving Cetaceans in the Ikarian Basin Paper

1st Beatriz Tintoré
Brunel University of London
Department of Mechanical Engineering & Institute for Health, Environmental & Societies.
Archipelagos Institute of Marine Conservation

London, United Kingdom
b.tintore@archipelago.gr

2nd Gera Troisi
Brunel University Lomdon
Department of Mechanical Engineering & Institute for Health, Environmental & Societies.
London, United Kingdom
Gera.Troisi@brunel.ac.uk

3rd Marie Cigarroa
Archipelagos Institute of Marine Conservation

Samos, Greece
marie.cigarroa@gmail.com

4th Thodoris Tsimpidis
Archipelagos Institute of Marine Conservation

Samos, Greece
t.tsimpidis@archipelago.gr

***Abstract*— The Ikarian Basin, located in the eastern Mediterranean Sea, is home to three deep-diving cetacean species: sperm whales (*Physeter macrocephalus*), goose-beaked whales (*Ziphius cavirostris*), and Risso's dolphins (*Grampus griseus*). These species, classified as endangered or vulnerable in the Mediterranean by the IUCN, face multiple threats, including overexploitation of fish stocks, habitat degradation, and the effects of climate change. Understanding their biology and ecology - migrations, habitat use, behavioural patterns, and residency dynamics - is essential for the development of effective conservation strategies. Modelling is a key tool for predicting species movements and habitat suitability. This study analyses the habitat suitability of 41 observations of *Grampus griseus*, 87 of *Physeter macrocephalus*, and 95 of *Ziphius cavirostris*, using depth, slope, and distance from shore as predictor variables. Using a Random Forest model, species were accurately classified, achieving an out-of-bag (OOB) error rate of 10.76%. Results indicate all environmental variables significantly influence species occurrence. These findings underscore the importance of the Ikarian Basin as a suitable habitat for deep-diving cetaceans, providing a foundation for proposing the area as a potential Important Marine Mammal Area (IMMA). Further research incorporating logged behavioural data and acoustic detections is needed to refine conservation strategies and gain deeper insights into habitat utilization.**

Keywords*— *Ikarian Basin, deep-diving cetaceans, Habitat Suitability Models, Important Marine Mammal Areas; IMMA

I. Introduction

Understanding cetacean abundance and distribution is essential to ensuring that conservation measures are based on scientific evidence [1]. Cetaceans are influenced by a combination of static and dynamic environmental variables. This includes temperature, productivity, geographic features and human activities [2], [10]. The study of physiographical parameters (depth, slope, and distance to the coast) enables the correlation between species occurrence and the geographical features of an area. Deep-diving species are particularly known to inhabit deep waters, around the continental slope or submarine canyons [4]. The study of species distribution models (SDMs) is essential for cetacean conservation management and the choice of the model changes depending on the variables available [4], [6]. Studies on deep-diving cetacean distribution in the Mediterranean indicate a higher density in the western regions (Alboran Sea, Ligurian Sea, Tyrrhenian Sea, Adriatic Sea) compared to the central and eastern sections [2]. This may be influenced by unequal data collection across these areas [2]. In Greek waters, the Hellenic Trench is considered an area of high concentration for deep-diving cetaceans [7], [8], [9]. Nevertheless, other Greek marine regions also gather suitable characteristics for deep-diving species occurrences, such as the Aegean Sea or the Myrtoan Sea, where deep-diving cetacean presence has been demonstrated, but little studied [7], [11].

Therefore, understanding the distribution and behaviour of deep-diving species - sperm whale (*Physeter macrocephalus - Pm*), goose-beaked whale (*Ziphius cavirostris - Zc*), and Risso's dolphin (*Grampus griseus - Gg*) - is crucial to identify key habitats and implement effective conservation strategies. Globally the IUCN Red List classifies the sperm whales as a "vulnerable" species, and the Risso's dolphin and goose-beaked whale as "least concern" species [10]. However, in the Mediterranean, sperm whales and Risso's dolphins are both classified as "endangered", and the goose-beaked whale as "vulnerable" [10], due to intensified anthropogenic pressures. Consequently, they are protected under international, European, and Mediterranean frameworks. For instance, the Bonn Convention [12] (Appendices I, II),

This article has been accepted for publication in a future workshop of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/MetroSea.12345.2025.1234567, 2025 IEEE International Workshop on Metrology for the Sea

Habitats Directive [13] (Article 12 and Annex IV), and Barcelona Convention [14] (Article 12 and Appendix II, SPA/BD protocol). While conventions and directives offer a strategic framework for effective marine spatial planning, promoting deeper research, and understanding on the species habitats and the factors that determine their environment. An Important Marine Mammal Area (IMMA) offers a more targeted and scientifically grounded tool to support conservation efforts guiding the establishment of Marine Protected Areas (MPAs) [15]. The designation of IMMAs considers different criteria, including the presence of vulnerable species, and the evidence of small, resident populations or aggregations of individuals [16]. Unfortunately, our understanding of the abundance and distribution of these species, particularly in the Eastern Mediterranean, is limited [2].

Thus, this study aims to describe the Ikarian Basin as a suitable habitat for deep-diving cetaceans by examining the influence of key physiographical parameters (depth, slope and distance from coast) on species occurrence. It also aims to contribute to the data deficit on deep-diving cetacean distribution in the study area and support the future application of an IMMA.

II. Materials and Methods

A. Study Area

The Ikarian Basin is located in the north of Ikaria and Samos Island in the east-central Aegean Sea, eastern Mediterranean Sea. It is a deep-water region characterized by steep slopes and submarine canyons, making it an ideal habitat for deep-diving cetaceans. Moreover, the bathymetric features of this area reach up to 1,570 meters in depth [17]. Different studies report the presence of deep-diving species in the Ikarian Basin [18], [19], as well as the occurrence of other protected species such as the common dolphin (*Delphinus delphis*), the common bottlenose dolphin (*Tursiops truncatus*) and the striped dolphin (*Stenella coeruleoalba*).

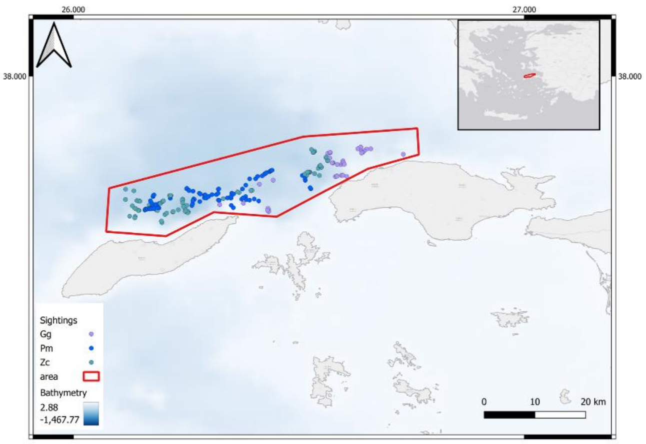

Fig. 1. Map of the study Area (Eastern Aegean Sea, Eastern Mediterranean Sea) and related bathymetry. The red box encompasses the Ikarian Basin.

B. Sighting Data

Sighting data on deep-diving cetaceans was collected during systematic boat-based surveys from January 2021 to December 2024 in Beaufort Sea State $\leqslant 4$. Surveys followed a standardized methodology with a minimum of four trained observers always scanning the water using binoculars (7x50) and covering the 360º of the boat. The GPS position of the boat was recorded every 30 minutes together with information on sea state, visibility, marine traffic, temperature and depth. When a sighting was recorded, coordinates were logged with a 0.001º x 0.001º decimal degree resolution. During the period of the study there were 41 observations of *Grampus griseus*, 87 of *Physeter macrocephalus* and 95 for *Ziphius cavirostris*.

C. Data Processing & Analysis

QGIS software [20] was used to plot sighting data and extract the depth (EMODnet [17]) information for every observation point, as well as the slope and the distance from shore using the "Distance to nearest Hub" vector analysis. Minimum and maximum values on depth, slope and distance from shore were extracted for all the species using descriptive statistics on R environment software [21]. Canonical Correspondence Analysis (CCA; [22]) with the Vegan package [23] in Rstudio was used to explore the species-geomorphology associations as well as to identify the overlap between the three different deep-diving species. The variables selected for the CCA analysis were Depth, Slope and Distance from Shore. As Latitude and Longitude ranked high Variance Inflation Value (VIF > 7). The significance of the CCA model was tested using a permutation-based ANOVA (analysis of variance) in Rstudio. As a complementary analysis, a Random Forest model was performed to predict species based on the morphological and spatial variables. The Random Forest model was trained to predict deep-diving cetaceans occurrence using five variables: Latitude, Longitude, Depth, Slope and Distance from Shore and the evaluation of the model was based on the Out of Bag (OOB) Error.

III. Results

The results on maximum and minimum depth for the three deep-diving species, showed a wider range of depths for *Physeter macrocephalus* compared to *Ziphius cavirostris* and *Grampus griseus*, but their occurrence also reached a large range of slope values and a greater distance from shore. In contrast, *Ziphius cavirostris* showed a specific preference for deeper waters and *Grampus griseus* an intermediate habitat preference between these two species (Table 1).

TABLE I. Summary Table of Depth, Slope and Distance from Shore for *Grampus griseus* (Gg), *Physeter macrocephalus* (Pm) and *Ziphius cavirostris* (Zc).

Variables	*Gg*	*Pm*	*Zc*
Depth - m (Min)	238	73.23	500.49
Depth - m (Max)	1,511	1,541.94	1,554.80
Depth - m (Mean)	696	1,044	1124

This article has been accepted for publication in a future workshop of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/MetroSea.12345.2025.1234567, 2025 IEEE International Workshop on Metrology for the Sea

Variables	*Gg*	*Pm*	*Zc*
Depth - m (Sd)	244	275	216
Slope - % (Min)	190,350	51,408.87	166,817.9
Slope- % (Max)	4,802,852	3,882,822	3,477,144
Slope - % (Mean)	1,666,294	1,204,753	1,156,617
Slope - % (Sd)	1,206,067	1,003,390	788,720
Distance – m (Min)	5,336.35	6,299.63	4,826.78
Distance – m (Max)	16,215.17	20,412.70	15,824.86
Distance – m (Mean)	8,753	11,050	10,181
Distance – m (Sd)	1,151	2,924	2,861

A CCA was tested to assess the association between deep-diving species distribution and physiographical variables (depth, slope and distance from shore). Excluding Latitude and Longitude since the VIF value was of 7.06 for Latitude and 7.23 for Longitude. The ANOVA showed that all three variables have significant influence on the deep-diving cetacean distribution (Table 2). To validate the findings a permutation test with 999 permutations was used and deep-diving species were only significantly associated with CCA1 (41.76% of total variance, $p<0.001$). The CCA biplot (Fig. 2) shows that most of the observations clustered near the origin, suggesting a homogeneous response of the species towards the variables. The arrows indicate the strengths and type of correlation between the variables in the CCA1 model. Overall, the CCA suggests that slope and distance from shore have the strongest influence on the occurrence of the species followed by the depth.

TABLE II. PERMUTATION ANOVA RESULTS FOR CCA MODEL

Covariates	Df	Chi2	F-value	P-value	
Depth	1	0.00084	17.390	0.001	***
Slope	1	0.00464	95.612	0.001	***
Distance	1	0.00236	48.703	0.001	***
Residual	219	0.01063			

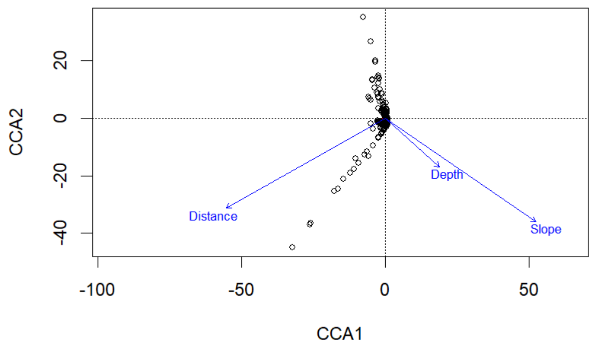

Fig. 2. CCA biplot showing the relationship between deep-diving species occurrences and Depth, Slope and Distance from Shore

The Random Forest model built with 700 trees showed an OOB error rate of 10.76%, which represents a strong predictive performance. The model included Latitude, Longitude, Depth, Slope and Distance to shore as variables. After evaluating the variable importance score we found that Latitude, Longitude and Depth were the most influential predictors. As well, the confusion matrix (Fig. 3) reveals the lowest misclassification rate in Zc, followed by Pm which is occasionally confused with Zc and the highest misclassification rate (17.1%) is for Gg which is often confused by Pm and Zc.

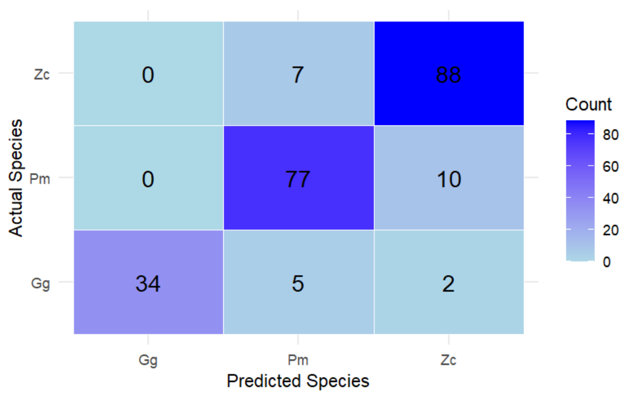

Fig. 3. Confusion matrix Heatmap of Random Forest model predictions for deep-diving species classification

IV. Discussion

The Habitat suitability models combined with machine learning tools are used in this study to identify crucial habitats of the studied species, confirming that the Ikarian Basin is indeed a key habitat for deep-diving species, reinforcing its ecological significance. The CCA results fall on the habitat preference described by many other studies, deep waters with a high slope percentage [9], [24], [25], [26]. This topographic complexity supports prey availability –

such as mesopelagic squid, deep-sea fish, and crustaceans – highlighting the role of these regions as ecological hotspots for deep-diving species like the goose-beaked whale and sperm whale [27]. The Random Forest model provided a classification outline to investigate the species influence of covariates, and the small misclassification encountered could be due to these species regularly occurring in similar areas [28], [29]. The most significant misclassification was found for Risso's dolphin – distribution in the Mediterranean showed low densities and wide-ranging movement – as also seen in the results of this study, makes their occurrence unpredictable [25]. After understanding habitat preferences, it is critical for conservation to understand distribution patterns [26]. For the sperm whales, in the eastern Mediterranean their occurrence is significantly lower than the western Mediterranean, with limited population mixing between basins [1]. Distinctly, the Hellenic trench is considered as one of the most significant hotspots of sperm whales in the eastern Mediterranean and previous re-sightings with the Aegean Sea suggest a migration within individuals of these sub-population [8]. On the other hand, migrations and movement information of goose-beaked whales and Risso's dolphins in the Mediterranean are limited [6], [27]. However, in certain areas of the Mediterranean both species have shown a significant long-term site fidelity [28], [30]. In the study area, results have indicated that the region occupied is a suitable habitat. This allows for the long-term survival of these species, suggesting the potential site fidelity that these species could display in the Ikarian Basin. Further research should incorporate photo identification data and acoustic monitoring to enhance our understanding of abundance, residency patterns and behavioural dynamics. As well, other variables such as sea surface current which can influence the Shelf Slope Front (SFF) and therefore the vertical movement of nutrients which also influences the migration of deep-diving species preys [33].

This study highlights the ecological importance of the Ikarian Basin, and it supports the application for an *Important Marine Mammal Area* (IMMA) based on evidence of a small, potentially resident population (criteria B1). This would constitute a significant first step towards its recognition as a deep-diving cetacean hotspot and subsequently trigger implementation of conservation measures. Like the *Hellenic Trench*, research in the Ikarian Basin emphasises the presence of vulnerable species facing anthropogenic threats [11] (criterion A). Moreover, the subpopulation of deep-diving cetacean species of the Mediterranean Sea is genetically unique and isolated from the global population [32] (criterion D1). The Ikarian basin also hosts diverse species such as common dolphins, bottlenose dolphins, striped dolphins, which are protected species [3] and the Mediterranean monk seal for which an IMMA has already been designated [31] (criterion D2). Under criterion C, the key life cycle activity can be further discussed with the presence of calves or juvenile individuals, or with potential migration routes. For example, Frantzis et al. [8] documented movements of one individual between the Hellenic Trench and the Aegean Sea. More knowledge on the behaviour, group composition and movements of the species in this specific area is requested to compose a solid and relevant application for an IMMA and thus target better marine protection.

V. Conclusion

The findings of this research give valuable insights into the spatial ecology of deep-diving cetaceans, offering guidance for targeted conservation efforts. In this case, since the spatial variables – depth, slope and distance from shore played a significant role on the occurrence prediction of these species, this highlights the need to incorporate these variables into the design of a potential IMMA. For a more robust understanding, future research in the Ikarian Basin should include the study of prey availability, chlorophyll concentrations, sea surface current velocity and changes on environmental variables. Additionally, further research to understand the movement of the deep-diving species between the Hellenic Trench and the Ikarian Basin, could help identify and safeguard potential migration corridors.

Acknowledgment

The authors would like to thank the Archipelagos Institute of Marine Conservation team for their support with data collection.

References

[1] A. Frantzis, S. Airoldi, G. Notarbartolo-di-Sciara, C. Johnson, and S. Mazzariol, "Inter-basin movements of Mediterranean sperm whales provide insight into their population structure and conservation", Deep Sea Res 1 Oceanogr Res Pap, vol. 58, no. 4, pp. 454–459, Apr. 2011, doi: 10.1016/j.dsr.2011.02.005.

[2] T. Awbery *et al.*, "Spatial Distribution and Encounter Rates of Delphinids and Deep Diving Cetaceans in the Eastern Mediterranean Sea of Turkey and the Extent of Overlap With Areas of Dense Marine Traffic", Front Mar Sci, vol. 9, no. July, pp. 1–15, 2022, doi: 10.3389/fmars.2022.860242.

[3] M. Ingrosso *et al.*, "Environmental variables influencing occurrence and distribution of Delphinus delphis in the eastern Aegean Sea (eastern Mediterranean Sea)", Aquat Conserv, vol. 34, no. 1, Jan. 2024, doi: 10.1002/aqc.4031.

[4] G. Gnone *et al.*, "Cetaceans in the Mediterranean Sea: Encounter Rate, Dominant Species, and Diversity Hotspots", Diversity (Basel), vol. 15, no. 3, Mar. 2023, doi: 10.3390/d15030321.

[5] E. Pasanisi, D. S. Pace, A. Orasi, M. Vitale, and A. Arcangeli, "A global systematic review of species distribution modelling approaches for cetaceans and sea turtles", Ecol Inform, vol. 82, Sep. 2024, doi: 10.1016/j.ecoinf.2024.102700.

[6] L. Cardona, N. Amigó, J. Ouled-Cheikh, M. Gazo, and C. A. Chicote, "Cetaceans and sea turtles in the northern region of the Mediterranean Cetacean Migration Corridor: abundance and multi-model habitat suitability analysis", Front Mar Sci, vol. 12, 2025, doi: 10.3389/fmars.2025.1496039.

[7] K. F. Thompson, T. Webber, L. Karantzas, J. Gordon, and A. Frantzis, "Summer and winter surveys of deep waters of the Hellenic Trench, Greece, provide insights into the spatial and temporal distribution of odontocetes", Endanger Species Res, vol. 52, pp. 163–176, 2023, doi: 10.3354/ESR01265.

[8] A. Frantzis, P. Alexiadou, and K. C. Gkikopoulou, "Sperm whale occurrence, site fidelity and population structure along the Hellenic Trench (Greece, Mediterranean Sea)", Aquat Conserv, vol. 24, no. SUPPL.1, pp. 83–102, 2014, doi: 10.1002/aqc.2435.

[9] M. Podestà *et al.*, "Cuvier's Beaked Whale, *Ziphius cavirostris,* Distribution and Occurrence in the Mediterranean Sea: High-Use Areas and Conservation Threats", in Advances in Marine Biology, vol. 75, Academic Press, 2016, pp. 103–140. doi: 10.1016/bs.amb.2016.07.007.

[10] IUCN, "The conservation status of cetaceans in the Mediterranean Sea: trends and changes after a decade of conservation efforts". Gland, Switzerland: IUCN, 2023. [Online]. Available: https://portals.iucn.org/library/sites/library/files/documents/RL-262-005-En.pdf

[11] I. Foskolos, K. C. Gkikopoulou, and K. C. Gkikopoulou, “Current State of Knowledge and Conservation Perspectives on the Cetaceans of the Aegean Sea”, Handbook of Environmental Chemistry, vol. 103, 2020, doi: 10.1007/698_2020_653.

[12] Convention on the Conservation of Migratory Species of Wild Animals, Bonn, 1979.[Online]. Available: https://www.cms.int/en/convention-text

[13] European Commission, Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and fora, (Habitats Directive) The Council of the European Communities, Brussels, 1992 [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A01992L0043-20130701

[14] UNEP/MAP, Convention for the Protection of the Marine Environment and the Coastal Region of the Mediterranean and Its Protocols. Nairobi, 2019.[Online]. Available: https://wedocs.unep.org/bitstream/handle/20.500.11822/31970/bcp2019_web_eng.pdf

[15] T. Agardy, “Effects of fisheries on marine ecosystems: a conservationist’s perspective”, ICES Journal of Marine Science, vol. 57, no. 3, pp. 761–765, Jun. 2000, doi: 10.1006/jmsc.2000.0721.

[16] IUCN Marine Mammal Protected Areas Task Force, “The IMMA Handbook”, Version: March 2024, 67 pp. [Online]. Available: https://www.marinemammalhabitat.org/download/handbook-on-the-use-of-selection-criteria-for-the-identification-of-important-marine-mammal-areas-immas/

[17] EMODnet Bathymetry Consortium, “EMODnet Digital Bathymetry (DTM 2018)”, EMODnet Bathymetry Consortium, September 2018.

[18] P. Hostetter, A. Koroza, T. Tsimpidis, G. Pietroluongo, R. Carlucci, and G. Cipriano, “Occurrence of *Physeter macrocephalus* and *Ziphius cavirostris* in the North Ikaria Basin, Aegean Sea”, International Workshop on Metrology for the Sea, 2020.

[19] G. Pietroluongo, A. Miliou, B. Tintore, T. Tsimpidis, and M. Azzolin, “Vessel-based surveys as a tool to fill critical knowledge gap on *Grampus griseus* in the northeastern Aegean Sea”, in 2021 IEEE International Workshop on Metrology for the Sea: Learning to Measure Sea Health Parameters, MetroSea 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 116–120. doi: 10.1109/MetroSea52177.2021.9611614.

[20] QGIS Development Team, "QGIS Geographic Information System," version 3.34.13, 2022. Available: QGIS Website.

[21] RStudio 2024.12.0+467 "Kousa Dogwood" Release (cf37a3e5488c937207f992226d255be71f5e3f41, 2024-12-11)

[22] J. V Redfern *et al.*, “Techniques for cetacean-habitat modeling”, 2006. [Online]. Available: www.int-res.com

[23] https://CRAN.R-project.org/package=vegan

[24] A. Moulins, M. Rosso, B. Nani, and M. Würtz, “Aspects of the distribution of Cuvier’s beaked whale (*Ziphius cavirostris*) in relation to topographic features in the Pelagos Sanctuary (north-western Mediterranean Sea)”, Journal of the Marine Biological Association of the United Kingdom, vol. 87, no. 1, pp. 177–186, Feb. 2007, doi: 10.1017/S0025315407055002.

[25] A. Frantzis and D. L. Herzing, “Mixed-species associations of striped dolphins (*Stenella coeruleoalba*), short-beaked common dolphins (*Delphinus delphis*), and Risso’s dolphins (*Grampus griseus*) in the Gulf of Corinth (Greece, Mediterranean Sea)”, Aquat Mamm, vol. 28, pp. 188–197, 2002, Accessed: Sep. 06, 2018. [Online]. Available: https://www.aquaticmammalsjournal.org/share/AquaticMammalsIssueArchives/2002/AquaticMammals_28-02/28-02_Frantzis.pdf

[26] G. Notarbartolo-Di-Sciara, “Sperm whales, *Physeter macrocephalus*, in the Mediterranean Sea: A summary of status, threats, and conservation recommendations”, Aquat Conserv, vol. 24, no. SUPPL.1, pp. 4–10, 2014, doi: 10.1002/aqc.2409.

[27] A. Cañadas *et al.*, “The challenge of habitat modelling for threatened low density species using heterogeneous data: The case of Cuvier’s beaked whales in the Mediterranean”, Ecol Indic, vol. 85, pp. 128–136, Feb. 2018, doi: 10.1016/j.ecolind.2017.10.021.

[28] A. Arcangeli *et al.*, “Testing indicators for trend assessment of range and habitat of low-density cetacean species in the Mediterranean Sea”, Front Mar Sci, vol. 10, 2023, doi: 10.3389/fmars.2023.1116829.

[29] R. Maglietta *et al.*, “Environmental variables and machine learning models to predict cetacean abundance in the Central-eastern Mediterranean Sea”, Scientific Reports 2023 13:1, vol. 13, no. 1, pp. 1–14, Feb. 2023, doi: 10.1038/s41598-023-29681-y.

[30] V. Corrias, F. Filiciotto, and F. Giardina, “Sightings of Risso’s Dolphin (*Grampus griseus*) off the Southern Coast of Linosa Island (South-Central Mediterranean Sea)”, Mediterr Mar Sci, vol. 22, no. 2, pp. 387–392, 2021, doi: 10.12681/mms.23648.

[31] G. Pietroluongo *et al.*, “Combining Monitoring Approaches as a Tool to Assess the Occurrence of the Mediterranean Monk Seal in Samos Island, Greece”, Hydrobiology, vol. 1, no. 4, pp. 440–450, Oct. 2022, doi: 10.3390/hydrobiology1040026.

[32] G. Notarbartolo di Sciara, “Marine Mammals in the Mediterranean Sea: An Overview”, in Advances in Marine Biology, vol. 75, Academic Press, 2016, pp. 1–36. doi: 10.1016/bs.amb.2016.08.005.

[33] I. M. Belkin and P. C. Cornillon, *Fronts in the World Ocean’s Large Marine Ecosystems*, ICES CM 2007/D:21, Theme Session D: Comparative Marine Ecosystem Structure and Function: Descriptors and Characteristics, International Council for the Exploration of the Sea, 2007