Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

multi-Risk sciEnce for resilienT commUnities undeR a changiNgclimate

Codice progetto MUR: **PE00000005** – CUP Lead Partner: F83C22001180002

Deliverable title: "Novel protocols for observation and ensuing modeling of space distribution"

Deliverable ID: DV3.8

Due date: 15/02/2026

Submission date: 15/02/2026

AUTHORS

Marco Fianchini (OGS); Donata Canu (OGS)

1. Technical references

Project Acronym	RETURN
Project Title	multi-Risk sciEnce for resilienT commUnities undeR a changiNg climate
Project Coordinator	Domenico Calcaterra UNIVERSITA DEGLI STUDI DI NAPOLI FEDERICO II domcalca@unina.it
Project Duration	December 2022 – November 2025 (36 months)

Deliverable No.	DV3.8
Dissemination level*	CO
Work Package	WP3 - Enhancing capability to observe, model, and assess environmental hazards
Task	T3.5 - Pathogens and biological invasion
Lead beneficiary	OGS
Contributing beneficiary/ies	OGS

* PU = Public

PP = Restricted to other programme participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only for members of the consortium (including the Commission Services)

Document history

Version	Date	Lead contributor	Description
0.1	20/11/2025	Marco Fianchini (OGS)	First draft
0.2	20/11/2025	Donata Canu (OGS)	Critical review and proofreading
0.3	02/02/2026	Viviana Fonti (OGS)	Edits for approval
1.0	15/02/206	Viviana Fonti (OGS)	Final version

2. Abstract

Invasive Alien Species (IAS) present an increasing threat to coastal ecosystems, exacerbated by climate change and maritime transport. Addressing this challenge requires forecasting tools that not only describe relationships between species and the environment, but also feed directly into early warning systems, adaptive monitoring, and targeted prevention measures. Ecological niche modelling (ENM), particularly when integrated with machine learning (ML), offers promising potential – provided that methodological rigour and operational relevance are ensured.

This work introduces a modular, decision-oriented ENM framework based on MaxEnt (but extensible to other presence-only algorithms) explicitly designed to support IAS monitoring and prevention strategies. Key methodological innovations include site-weighted evaluation metrics to mitigate spatial biases, model overfitting diagnostics, multi-criteria model selection aligned with intervention objectives, and objective extrapolation risk analysis as a proxy for uncertainty in projections under changing (novel) conditions.

The results are presented as actionable tools: spatial risk maps to set monitoring priorities; uncertainty overlays to identify data gaps where additional sampling would provide the greatest benefit; and temporal projections to synchronise monitoring and eradication with expected "swings" in suitability.

The framework was applied to Caulerpa cylindracea, one of the most dangerous invasive species in the Mediterranean. The selected model identified areas of persistent ecological suitability likely to sustain future spread, as well as zones where limited data and changing environmental conditions make reliable prediction difficult. Regions where high suitability coincides with high predictive uncertainty can be used as strategic targets for intensified sampling, as new observations in these areas would contribute most to improving model confidence. Simultaneously, annual suitability projections enable temporal prioritisation of control efforts.

Beyond this case study, the framework provides a transferable approach for integrating robust ecological modelling into proactive biosecurity planning.

3. Table of contents

List of Tables

List of Figures

4. Novel protocols for observation and ensuing modeling of space distribution

4.1 Introduction

4.1.1 Invasive Alien Species

Invasive alien species (IAS), defined in EU Regulation 1143/2014 as alien species whose introduction or spread threatens biodiversity and ecosystem services, are a major pressure on European seas (Coll et al., 2010; Haubrock et al., 2021). In the Mediterranean, a recognised marine biodiversity hotspot, IAS have already caused substantial ecological and socio-economic impacts (Katsanevakis et al., 2014). These impacts include competitive exclusion of native species, habitat modification, and alteration of ecosystem functioning, especially when invasions become persistent and widespread (Boudouresque & Verlaque, 2012). From an ecological risk-assessment perspective, anticipating where and when environmental conditions will be suitable for IAS establishment is central to problem formulation, early detection, containment, and long-term mitigation.

Climate change is reshaping invasion dynamics by warming and acidifying the ocean and altering circulation patterns, thereby shifting or expanding the environmental space available to many invaders and generating novel combinations of conditions (Pyšek et al., 2020). Consequently, spatially explicit forecasts of current and future IAS suitability have become a core component of marine biosecurity and MSFD Descriptor 2 implementation. Correlative ecological niche models (ENMs), which approximate the realised niche by relating species occurrences to environmental predictors and then projecting potential distributions, are now standard tools for this purpose (Marcelino & Verbruggen, 2015; Elith et al., 2011). Machine-learning implementations such as MaxEnt are particularly attractive for presence-only data and are widely used to generate risk maps and climate change scenarios for IAS (Thuiller et al., 2009).

However, conventional ENM applications to IAS have several structural limitations that reduce their value for operational risk assessment. Occurrence data are typically spatially biased, heterogeneous in quality, and affected by uneven sampling effort, which can inflate classical performance metrics and undermine model transferability (Yackulic et al., 2013). Presence-only algorithms require explicit choices about feature classes, regularisation strength, background sampling, and probability thresholds; these hyperparameters control model complexity, overfitting, and the definition of "at-risk" areas, but are often tuned using a narrow set of standard metrics and informal expert judgement, with limited exploration of trade-offs (Muscarella et al., 2014; Liu et al., 2016). Under climate change, projections are further complicated by non-analogous environmental conditions: models must balance goodness-of-fit in the calibration domain with low extrapolation risk if predictions are to remain credible in novel climates (Zhang et al., 2020). In practice, sampling bias, spatial autocorrelation, overfitting, and classification uncertainty are often treated qualitatively, which weakens the interpretation of ENM outputs as quantitative indicators of ecological risk.

4.1.2 Challenges and objectives

Using ENMs as operational tools for IAS management therefore remains challenging. Marine invasion data are noisy and unevenly sampled; key processes operate at different spatial and temporal scales; and management decisions must be made under deep uncertainty and limited resources. A critical challenge is to ensure that models which perform well according to statistical criteria also remain ecologically interpretable and decision-relevant when transferred beyond their calibration domain, particularly under changing climate. From an ecological risk-assessment perspective, this requires: (i) explicit control of sampling bias and model complexity, (ii) evaluation metrics that reflect site-specific information content and management priorities, and (iii) formal treatment of extrapolation and classification uncertainty.

The framework developed in this work addresses these needs by proposing a modular, MaxEnt-based ENM protocol tailored for IAS risk assessment. First, statistical assumptions are made explicit and upheld: sampling design, construction of the calibration region, background selection, model complexity, and cross-validation are chosen to account for spatial autocorrelation and to minimise artefacts from sampling bias and overfitting. Second, modelling decisions are ecologically constrained: variable selection, spatial and temporal resolution, and interpretation of response curves are guided by current understanding of species ecology and Mediterranean environmental gradients, favouring biologically plausible functional responses over purely numerical fits. Third, principles of ecological risk assessment are embedded in the workflow: site-weighted evaluation metrics increase the informative value of performance scores, multi-criteria decision analysis is applied to select model configurations consistent with the desired balance between sensitivity and specificity, and extrapolation-risk analysis serves as a proxy for uncertainty in projections under changing conditions.

4.1.3 Case study

The framework is demonstrated using Caulerpa cylindracea (Figure 1), one of the most dangerous invasive marine macroalgae in the Mediterranean Sea. Since its first records in the basin, C. cylindracea (historically also referred to as C. racemosa var. cylindracea) has rapidly expanded, forming dense mats that overgrow native algal and seagrass communities (Klein & Verlaque, 2008; Piazzi et al., 2016). Its invasive success is associated with fast clonal growth, efficient vegetative dispersal via fragmentation, and rapid nutrient uptake (Uyà et al., 2018; Gennaro et al., 2015). The species preferentially colonises disturbed and urbanised coasts and structurally complex substrates such as rocky reefs and artificial structures, often in interaction with local stressors and nutrient enrichment (Cantasano et al., 2017; Gennaro et al, 2015). At the same time, its long invasion history, extensive monitoring, and well-documented distribution (Piazzi et al., 2016) make C. cylindracea both a high-priority management concern and an appropriate model IAS for testing advanced ML-ENM protocols aimed at delivering spatially explicit risk and uncertainty maps for decision support.

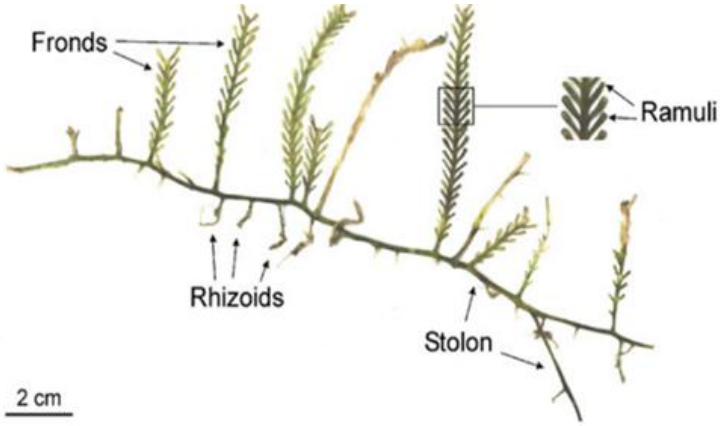

Figure 1. Thallus of the invasive *Caulerpa cylindracea* from the Gulf of Marseille (-30 m). Herbarium specimen, J. Klein (Klein & Verlaque, 2008).

4.2 Modelling approach

4.2.1 Concept and design

The modelling framework is a modular ecological niche modelling (ENM) workflow based on MaxEnt and tailored for invasive alien species (IAS) risk assessment (Fianchini et al., 2025). Its objective is to produce spatially and temporally explicit suitability patterns that are reliable enough to support early warning, monitoring design and prioritisation of management actions, while making the main sources of uncertainty explicit.

The framework is decision-oriented: model structure, evaluation criteria and output formats are derived from the management question, the invasion stage and the spatial–temporal domain. In practice, this entails defining a calibration region and resolution, specifying baseline and projection periods (or scenarios), and identifying which indicators are needed for risk assessment, such as continuous suitability, suitability classes, temporal trends and uncertainty layers. Although the case study uses MaxEnt, the workflow is transferable to other presence-only ENM algorithms with similar input–output structure.

4.2.2 ENM modelling framework

At its core, the framework fits a correlative ENM that links observed presences to environmental predictors and returns a continuous suitability surface over the target domain (Elith et al., 2011). MaxEnt is adopted as the reference algorithm because of its theoretical grounding for presence-only data and its extensive use in invasion ecology (Phillips & Dudík, 2008).

Model complexity is explored systematically rather than fixed a priori. A grid of candidate configurations combines alternative sets of feature classes with a range of regularisation multipliers, spanning models from flexible to strongly penalised, in line with recommendations on tuning complexity to model purpose (Warren & Seifert, 2011; Fianchini et al., 2025). Each configuration is calibrated on the same dataset and evaluated under a consistent resampling scheme. To reflect spatial dependence and avoid overly optimistic performance estimates, cross-validation is based on spatially structured blocks, while independent monitoring data, when available, are reserved exclusively for external validation and transferability assessment (Peterson et al., 2008).

Background points are sampled within a carefully defined calibration region that includes the realised niche under current conditions, captures the environmental gradients relevant to the invasion process and limits unnecessary extrapolation, following guidance on environmentally coherent calibration areas (Owens et al., 2013). See Figure 2 for a schematic representation of the process.

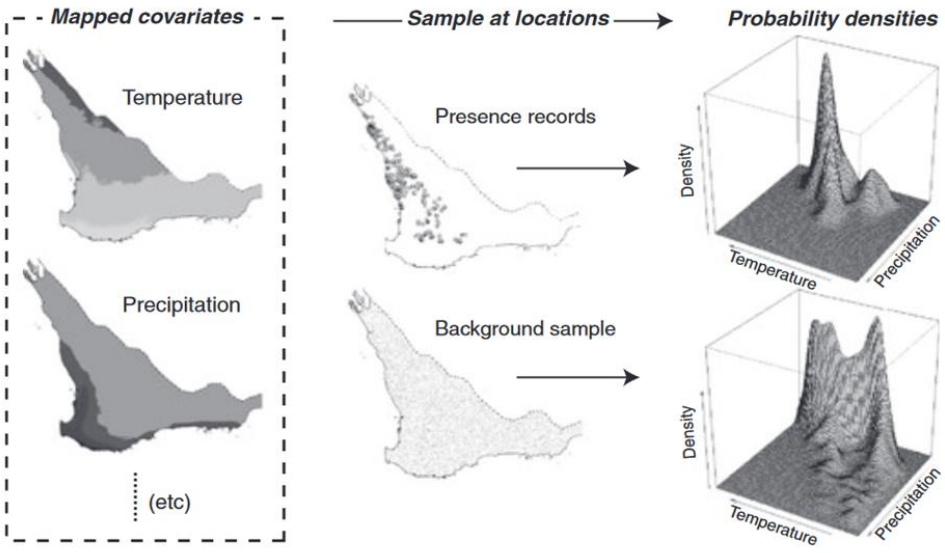

Figure 2. A schematic representation of the probability densities. The maps on the left show two examples of mapped covariates. In the centre are the locations of the presence and background samples. The density estimates on the right are not in geographic space (map), but show the distributions of values in covariate space for the presence (top right) and background (bottom right) samples. Taken from Elith et al., (2011).

4.2.3 Model evaluation

4.2.3.1 A novel set of metrics

Each candidate model is evaluated with a suite of complementary metrics that describe discrimination, statistical significance, complexity and transferability, computed for both cross-validation and independent test data. For every metric, two versions are calculated. Unweighted metrics treat all evaluation sites equally and capture raw predictive performance. Site-weighted metrics apply weights that down-weight spatially clustered or weakly informative observations and up-weight ecologically representative or higher-confidence records, thereby reducing the influence of opportunistic sampling and reinforcing the ecological signal (Fianchini et al., 2025).

To diagnose overfitting and sensitivity to sampling bias, the framework introduces delta metrics, defined as the absolute difference between unweighted and site-weighted values for each performance metric. Large delta values indicate that apparent performance is strongly driven by the spatial configuration and weighting of sites, rather than by genuine ecological signal (Townsend Peterson et al., 2007), whereas small deltas indicate robustness across weighting schemes (Fianchini et al., 2025). Considering unweighted, weighted and delta metrics together provides a multi-dimensional characterisation of each configuration's behaviour across datasets and spatial structures.

4.2.3.2 Multi criteria model selection

Model selection is formulated as a multi-criteria decision problem (Fianchini et al., 2025). After fitting the full grid of configurations, the set of candidates is first reduced by retaining all models that perform near-optimally for at least one evaluation metric, ensuring that configurations favoured by different criteria remain in the pool. These candidates are then ranked with a modified TOPSIS procedure, which orders models by their distance to a positive ideal solution that simultaneously maximises desirable metrics (e.g. discrimination, transferability) and minimises undesirable ones (e.g. complexity, delta metrics) (Chakraborty, 2022; Warren & Seifert, 2011).

Criterion weights are derived objectively using the CRITIC method, which exploits the variance and correlation structure of metrics to assign higher weight to criteria that are both informative and non-redundant (Krishnan et al., 2021). The resulting TOPSIS score provides a single, transparent measure of overall performance and robustness for each model, reducing subjectivity in the final choice and supporting defensible decision-making in an invasion-risk context (Fianchini et al., 2025).

4.2.4 Projection and classification

The selected MaxEnt configuration is projected across the spatial domain and for each year in the baseline and scenario periods, generating a series of continuous suitability surfaces (Fianchini et al., 2025). For risk communication, these values are converted into a limited number of suitability classes (for example, not suitable, low, medium, high) using thresholds such as the value that maximises the sum of sensitivity and specificity. Thresholding is treated as a decision layer: by exploring a plausible range around the reference threshold and mapping where class membership changes, the framework provides a spatially explicit assessment of classification uncertainty

Annual projections are summarised as time series of the area occupied by each suitability class. These trajectories are used to detect trends and "swings" in suitability and to contrast baseline and future periods, yielding indicators that are directly interpretable for planning surveillance (where and when to intensify monitoring) and control (how long high-risk conditions are expected to persist) (Fianchini et al., 2025).

4.2.5 Extrapolation Risk Analysis and uncertainty assessment

Because projections under climate change or in poorly sampled regions inevitably involve environmental extrapolation, the framework incorporates an explicit extrapolation-risk analysis based on the mobility-

oriented parity (MOP) metric (Cobos et al., 2023; Fianchini et al., 2025). Environmental conditions in each projection cell are compared with those in the calibration domain in multidimensional predictor space, and the minimum distance to the calibration envelope is used, together with objective thresholds, to classify cells into interpolation, borderline and strict extrapolation zones.

The resulting MOP map is interpreted as a spatially explicit uncertainty layer and overlaid on suitability projections. Areas of strict extrapolation are flagged as unreliable for primary decision support, whereas regions that combine high suitability with high extrapolation risk are highlighted as priorities for additional sampling and model refinement (Cobos et al., 2023; Fianchini et al., 2025).

4.3 Data preparation

4.3.1 Environmental data

Environmental predictors were derived from a bias-corrected, high-resolution bioclimatic dataset for the Mediterranean Sea (1/24° grid, ~4 km², annual resolution, 2000–2050, RCP8.5), built by downscaling and harmonising CMEMS and CMCC physical–biogeochemical products (Fianchini et al. 2024a,b). The dataset includes temperature, salinity, pH, currents, nutrients, oxygen, net primary production and related variables for surface and bottom layers. For each quantitative variable, five annual statistics (minimum, maximum, range, mean and coefficient of variation) were computed at both depths, providing a bioclimatic characterisation of coastal habitats. Two additional categorical layers—EUNIS biozone and substrate (Vasquez et al. 2021)—were added from EMODnet to represent depth-related habitat structure and seabed type, which are known to constrain the distribution of C. cylindracea (Piazzi et al. 2016). To ensure spatial consistency, all layers were projected onto an equal-area Eckert IV coordinate system, as required by area-dependent algorithms such as MaxEnt (Renner & Warton 2013). Categorical rasters (biozone and substrate) were resampled using an area-preserving method to avoid distortions in habitat proportions (Johnson & Clarke 2021). The environmental domain was constrained to marine cells shallower than 200 m and to biozones corresponding to infralittoral and shallow and deep circalittoral, that match the typical depth range of the species. To limit multicollinearity and retain ecologically interpretable signals, a stepwise variance inflation factor (VIF) analysis (correlation threshold 0.5, VIF threshold 2.5) identified a final set of eight predictors (Table 1) capturing thermal regime, hydrodynamics, nutrient availability, productivity and habitat characteristics (Cobos et al., 2019). A 'present' and a 'future' scenario were created as average values for each bioclimatic variable for the periods 2000–2020 and 2030–2050, respectively. The 'present' scenario has been used to train the models.

Table 1. Layers selected after VIF analysis and used in the modelling process.

Name	Var	unit	Statistic	Values	VIF score
Temperature	thetao	$°C$	max	surface	1.40
			min	surface	1.36
Salinity	so	PSU	max	bottom	1.20
Water velocity	KE	m/s	min	bottom	1.01
Phosphate	po4	$[mmolm-3]$	mean	bottom	1.69
Net Primary Production	nppv	$[mgm-2day-1]$	mean	water column sum	1.49
Biozone	-	-	-	categorical	-
Substrate	-	-	-	categorical	-

4.3.2 Occurrences

Occurrence data for C. cylindracea were compiled for 2000–2020 from EASIN, GBIF, targeted literature sources and the Reef Check Mediterranean (RCMED) citizen-science monitoring programme (Ould-Ahmed & Meinesz 2007; Zoheïr 2010; Turicchia et al. 2021; GBIF.org 2024). Records with incomplete or obviously erroneous metadata (e.g. missing coordinates or dates) were removed, and all observations were aggregated to the 1/24° grid used for the environmental layers. A conservative, threshold-based labelling was then applied: a grid cell was considered a presence only if it contained at least three unique occurrence records, thereby reducing the influence of isolated or uncertain reports (Fianchini et al. 2025). RCMED data, which include semi-quantitative abundance scores from 0 (species searched but not found) to 6 (very dense meadows), were processed separately. Cells were labelled as confirmed absences only when at least five independent RCMED dives reported abundance zero and no conflicting presence records existed; when multiple abundance values occurred in the same cell, the maximum value was retained as a precaution (Turicchia et al. 2021; Fianchini et al. 2025). Figure 3 shows the distribution of the training and testing points.

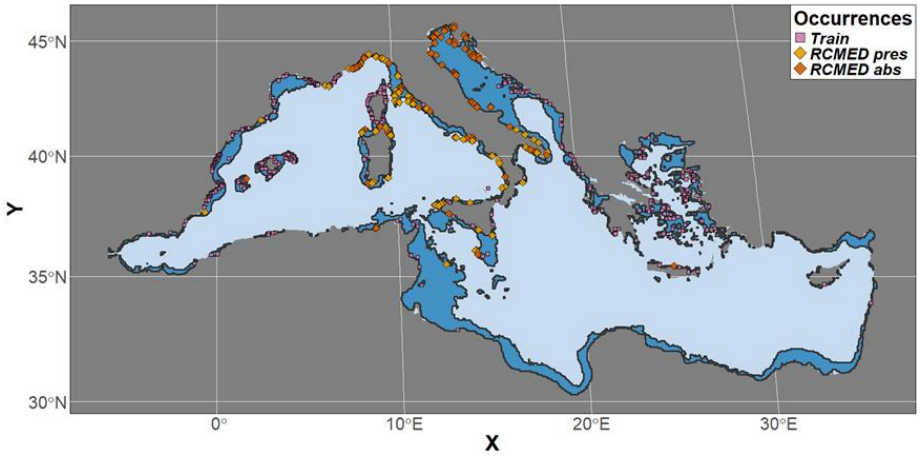

Figure 3. Presence (and absence) sites identified after the cleaning process. Dark blue areas identify the domain of the study (Mediterranean waters with depths < 200 m.

4.4 Model training and evaluation

4.4.1 Model calibration and setup

The occurrence–background dataset was calibrated on the accessible shelf domain (depth < 200 m and within 200 km of any presence), consistent with the observed expansion rate of 1–10 km year^{-1} for C. cylindracea (Iveša et al. 2015). Within this domain, the RCMED presence/absence cells (81 presences with abundance and 85 absences) were reserved as a curated validation set. To preserve spatial independence, all cells within 20 km of any RCMED site were excluded from the calibration pool, so that training and validation data are geographically separated (Fianchini et al. 2025). Figure 4 shows the domain of the study (purple) and highlights regions used for calibration (green), and regions excluded for testing purposes (red).

The occurrence–background data used for calibration were partitioned with a 5-fold masked geographically structured approach (Radosavljevic & Anderson 2014). The distance band used to separate folds was derived from semi-variograms of the selected predictors (drange = 399106.5 m), ensuring that training and test folds were separated by more than the main autocorrelation range. MaxEnt's internal maximum number of iterations was increased to 100,000 to guarantee convergence for all configurations.

Hyperparameter tuning followed an exhaustive grid search over MaxEnt model complexity. All combinations of regularisation multipliers from 0.5 to 20 (step 0.5) and feature classes (linear, quadratic, product, threshold and hinge) were tested, yielding 1240 models fitted without clamping, that is allowing

explicit extrapolation beyond the calibration environmental range (Street 2020). For each configuration, performance was summarised by a suite of standard metrics—partial ROC, AUC, CBI, SEDI, OR10p, AUCdiff and AICc—and their site-weighted counterparts, computed in cross-validation and on the independent RCMED dataset (Zhang et al. 2020; Smith et al. 2023). Differences between weighted and unweighted scores ("delta metrics") provided an additional indication of overfitting. The full matrix of metrics across the 1240 models forms the basis for the multi-criteria ranking and model selection described in the next section.

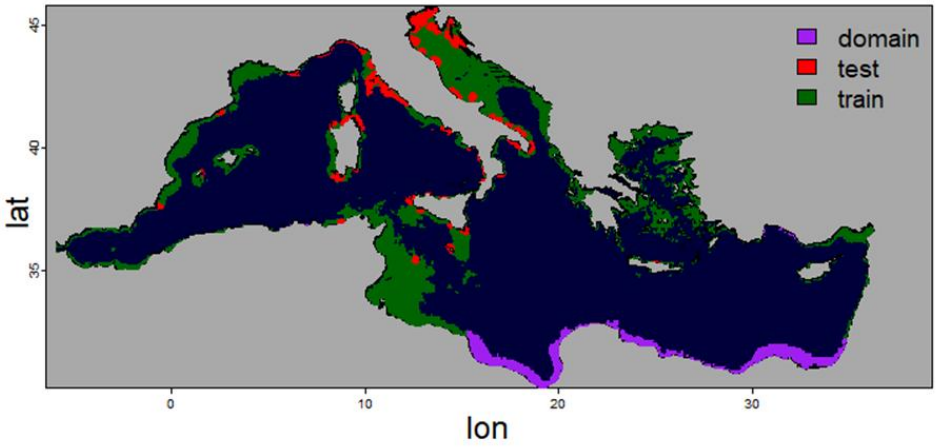

Figure 4. Geographical representation of the calibration ("train") and testing ("test") regions in relation to the full study domain.

4.4.2 Model evaluation

To reduce the influence of spatial sampling bias and heterogeneous data quality on validation, we evaluated models using site-weighted performance metrics. See Table 2 for a brief description of the metrics selected to measure model performance. IAS occurrence and citizen-science data are typically uneven in space, with presences and absences clustered around well-surveyed coastal segments (Fourcade et al., 2014). If untreated, this structure can inflate standard metrics such as AUC and lead to over-optimistic assessments of transferability (Yackulic et al., 2013).

Following Fianchini et al. (2025), each evaluation site (presence or absence) was assigned a weight reflecting its spatial configuration and ecological relevance. For the RCMED presence/absence dataset, presence weights were scaled by semi-quantitative abundance scores, so that persistent, highly invaded sites contributed more to performance estimates than marginal occurrences. Absence weights were scaled inversely with distance to the nearest known presence, giving higher importance to absences close to the invasion front than to distant, clearly uninvaded areas. This scheme down-weights dense clusters and emphasises isolated occurrences and frontier sites that are most informative for delineating the niche.

For every candidate model, we computed weighted and unweighted versions of key metrics (AUC, sensitivity, specificity and related statistics). We then derived "delta metrics", defined as the absolute difference between weighted and unweighted scores ($\Delta AUC = |AUC_{unw} - AUC_w|$; Fianchini et al., 2025). Large deltas indicate strong dependence on the weighting scheme and therefore potential overfitting to sampling artefacts, whereas small deltas suggest that performance is robust to spatial bias correction.

Table 2. Summary of the performance metrics used in this work. SEDI and OR10p are threshold-dependent metrics, while AUC, pROC and CBI are threshold-independent metrics.

Metric	Rationale	Description	Reference
AUC	Predictive ability	Measures the probability that a randomly selected positive example will be ranked higher than a randomly selected negative one	Fawcett, 2006
pROC	Statistical significance	A modified ROC analysis that adjusts the AUC to focus on the proportion of area correctly predicted as present, accounting for model-specific prediction spans and prioritizing omission errors.	Peterson et al., 2008
SEDI	Predictive ability	Measures the accuracy of deterministic forecasts of rare binary events, robust to low prevalence	Wunderlich et al., 2019
OR.10p	Predictive ability	Percentage of test sites where presence was not predicted, using the 10th percentile of training presence values as the threshold	Muscarella et al., 2014
AUCdiff	Overfitting	Measures the risk of model overfitting by comparing AUC values between training and test datasets	Warren and Seifert, 2011
AICc	Complexity	Relative measure of model fit that considers the number of parameters and penalizes overfitting	Burnham and Anderson, 2003
CBI	Transferability	Measures the accuracy and the reliability in predicting presence based on the Spearman rank correlation between predicted suitability and observed presence	Hirzel et al., 2006

4.4.3 Model selection

To select an optimal configuration from the 1,240 candidate MaxEnt models, we approached model selection as a multi-criteria decision-making (MCDM) problem. Rather than choosing the model with the highest AUC or lowest AIC – which can favour overfitted configurations and depend on ad hoc judgement (Muscarella et al., 2014) – we combined multiple, partly conflicting criteria describing discrimination, overfitting, transferability, and complexity into a single, transparent ranking (Fianchini et al., 2025).

Model performances were organised into a decision matrix comprising six families of criteria: (1) unweighted cross-validation metrics, (2) site-weighted cross-validation metrics, (3) cross-validation delta metrics, (4) unweighted independent test metrics, (5) weighted test metrics, and (6) test delta metrics. Together, these summarise how each model behaves under internal versus external validation, with and without bias correction, and how sensitive it is to the evaluation scheme (Fianchini et al., 2025).

We ranked models using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) (Chakraborty, 2022). All criteria were normalised and classified as either "benefit" (to be maximised, e.g. AUC, sensitivity) or "cost" (to be minimised, e.g. omission rate, delta metrics). TOPSIS defines an ideal solution (best value for every criterion) and a negative-ideal solution (worst values), then scores each model by its relative closeness to the ideal and distance from the negative-ideal.

To avoid subjective criterion weights, we combined TOPSIS with CRITIC (Criteria Importance Through Intercriteria Correlation) (Krishnan et al., 2021). For each metric, CRITIC quantifies information content as the product of its standard deviation and its overall dissimilarity to the other metrics (1 – correlation). Metrics that vary strongly among models and are weakly correlated with others receive higher weights; nearly constant or highly collinear metrics are down-weighted. These objective weights are then used in TOPSIS, yielding a data-driven aggregation of criteria (Fianchini et al., 2025).

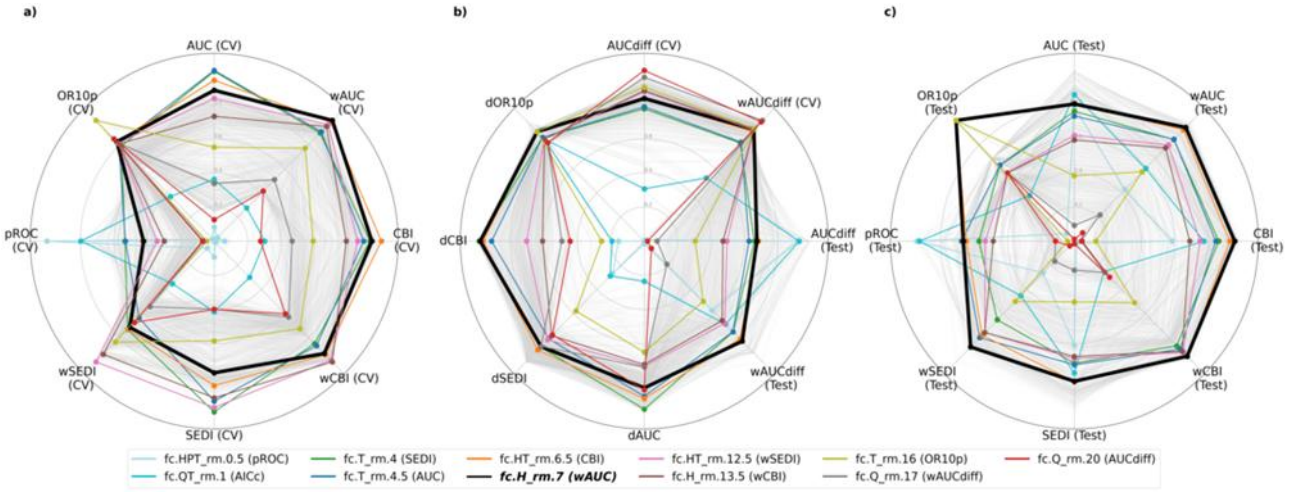

Figure 5. Kiviat diagram of model rankings for unweighted (a), weighted (c), AICc and delta (b) metrics. The delta metrics are calculated as follows: dMETRIC = | METRICunw − METRICw Coloured lines represent model candidates, while grey lines represent all other model configurations. The model selected by TOPSIS, highlighted in bold, shows the most consistent overall performance.

Model selection was conducted hierarchically: CRITIC–TOPSIS was first applied within each of the six metric groups to obtain partial scores, which were then treated as higher-level criteria in a final CRITIC–TOPSIS step. This ensures that the best model must perform consistently well across all dimensions, not just excel in a single aspect (Fianchini et al., 2025). The procedure identified a MaxEnt model with hinge (H) features and an intermediate regularisation multiplier (~7) as the overall best configuration. This model showed high and stable performance across unweighted, weighted, and delta-based metrics, with delta values close to zero on both cross-validation and independent tests, indicating low sensitivity to sampling-bias correction and robust generalisation, while more complex or simpler alternatives were downgraded once all criteria were considered. Figure 5 shows model performances on Kiviat ('spider') plots for unweighted, weighted and delta metrics.

4.5 Results

Once the CRITIC–TOPSIS procedure had identified the best-performing configuration (hinge features, intermediate regularisation), this MaxEnt model was used to generate annual projections for the Mediterranean shelf (depth < 200 m) over 2000–2050, driven by the downscaled RCP8.5 bioclimatic dataset described in Section 4.3.1. The model's output, after logistic transformation, is interpreted as the probability of occurrence p for each grid cell and year. These annual probability-of-occurrence rasters constitute the primary quantitative output of the ENM and are the basis for all subsequent risk indicators and summaries.

For decision support, continuous probabilities were translated into discrete suitability classes. Following Fianchini et al. (2025), the reference decision threshold was the value that maximises the sum of sensitivity and specificity (maxSSS; Liu et al., 2016), estimated on the independent RCMED presence–absence dataset. The corresponding maxSSS value th was then used to define four suitability classes: high suitability (HS, $p > th$), medium suitability (MS, $2th/3 < p \leq th$), low suitability (LS, $th/3 < p \leq 2th/3$) and non-suitable (NS, $0 < p \leq th/3$). This preserves a direct link between the probability scale and the classification, while distinguishing truly high-risk areas from marginally suitable conditions.

The probability density functions of predicted occurrence (Figure 6) summarise how suitability is distributed across present (2000–2020) and future (2030–2050) conditions for all 1,240 MaxEnt configurations. In the future scenario, all configurations show a coherent left-shift: density increases near zero and decreases at medium–high probabilities. This behaviour is consistent across the entire ensemble, indicating a

generalised trend towards less suitable conditions for C. cylindracea rather than an artefact of a particular model configuration.

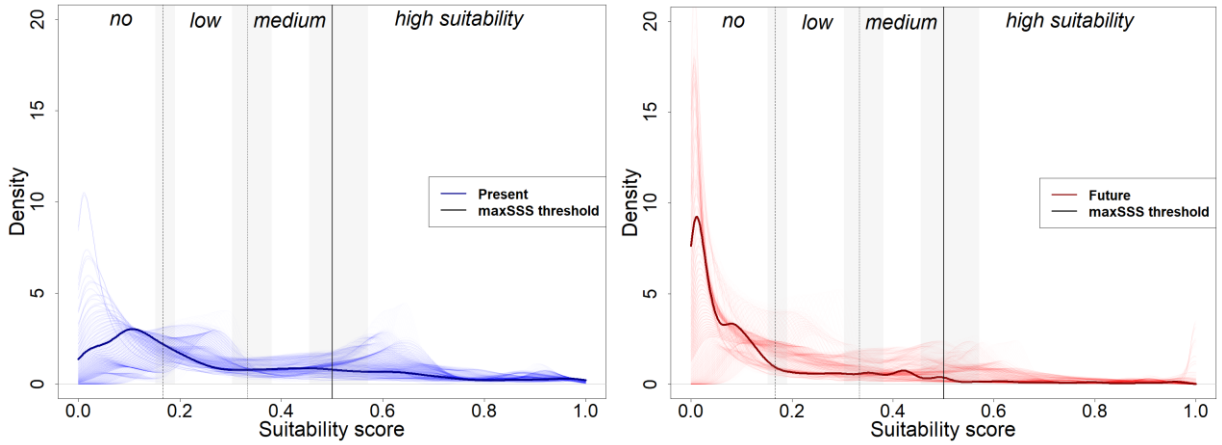

Figure 6. Probability Density Functions for the predicted probability of occurrence by the 1240 different MaxEnt models for present (left) and future (right) conditions. Thick lines highlight the selected model. Vertical lines represent maxSSS based thresholds used to classify probability of occurrence into suitability. Grey shaded areas indicate threshold range.

To characterise threshold-related uncertainty, classification was not based on a single value of th. Instead, we exploited the fact that MaxEnt computes one maxSSS threshold per cross-validation partition. The empirical range of these k thresholds was used to derive two additional, equally plausible cut-offs around the reference value, yielding three alternative threshold sets that span conservative to more permissive detection policies. For each of these, the annual probability maps were reclassified into HS, MS, LS and NS, producing a small ensemble of "classification scenarios" that reflect uncertainty in where the HS–MS–LS boundaries should be placed rather than uncertainty in the underlying probabilities themselves.

At the spatial level, the three threshold sets produce consistent large-scale patterns but differ in how sharply they delineate high-risk areas (Figure 7). Under the reference ("normal") threshold, HS cells are concentrated in already invaded regions of the Western Mediterranean, Aegean and parts of the Adriatic shelf, with MS and LS forming a surrounding transition belt. When the most permissive threshold ("best" scenario) is used, HS contracts to the core of these invasion foci, with many marginal cells in the northern and central Adriatic reclassified as MS or LS. Conversely, the most conservative threshold ("worst") expands HS into neighbouring MS/LS zones, especially along the northern Adriatic and some western coastal segments. This confirms that threshold uncertainty primarily affects the location of class boundaries rather than the underlying probability patterns, which remain robust across scenarios. Spatial sensitivity is clearly non-uniform: recently invaded regions, particularly the northern Adriatic, show the largest swings between HS, MS and LS as thresholds change, whereas long-established invasion areas in the Western Mediterranean and Aegean retain a stable classification across all three threshold sets.

To summarise temporal dynamics, for each year and threshold set we computed the total shelf area (km^2) classified as HS, MS, LS and NS and analysed trends over 2000–2050 using the Mann–Kendall test. Across all three threshold scenarios, HS areas show a statistically significant negative trend ($p < 0.05$), whereas NS exhibits a significant positive trend, indicating a progressive expansion of completely unsuitable conditions (Figure 8).

Short-term oscillations in all classes mirror interannual variability in the driving bioclimatic variables, but the sign and magnitude of the long-term trends are virtually unchanged under conservative, reference and permissive thresholds. This robustness indicates that, under RCP8.5 and assuming no species adaptation, the main signal is a gradual contraction of strongly and moderately suitable habitats, combined with the persistence of well-established core hotspots in already invaded regions and a growing fraction of the shelf where establishment becomes unlikely.

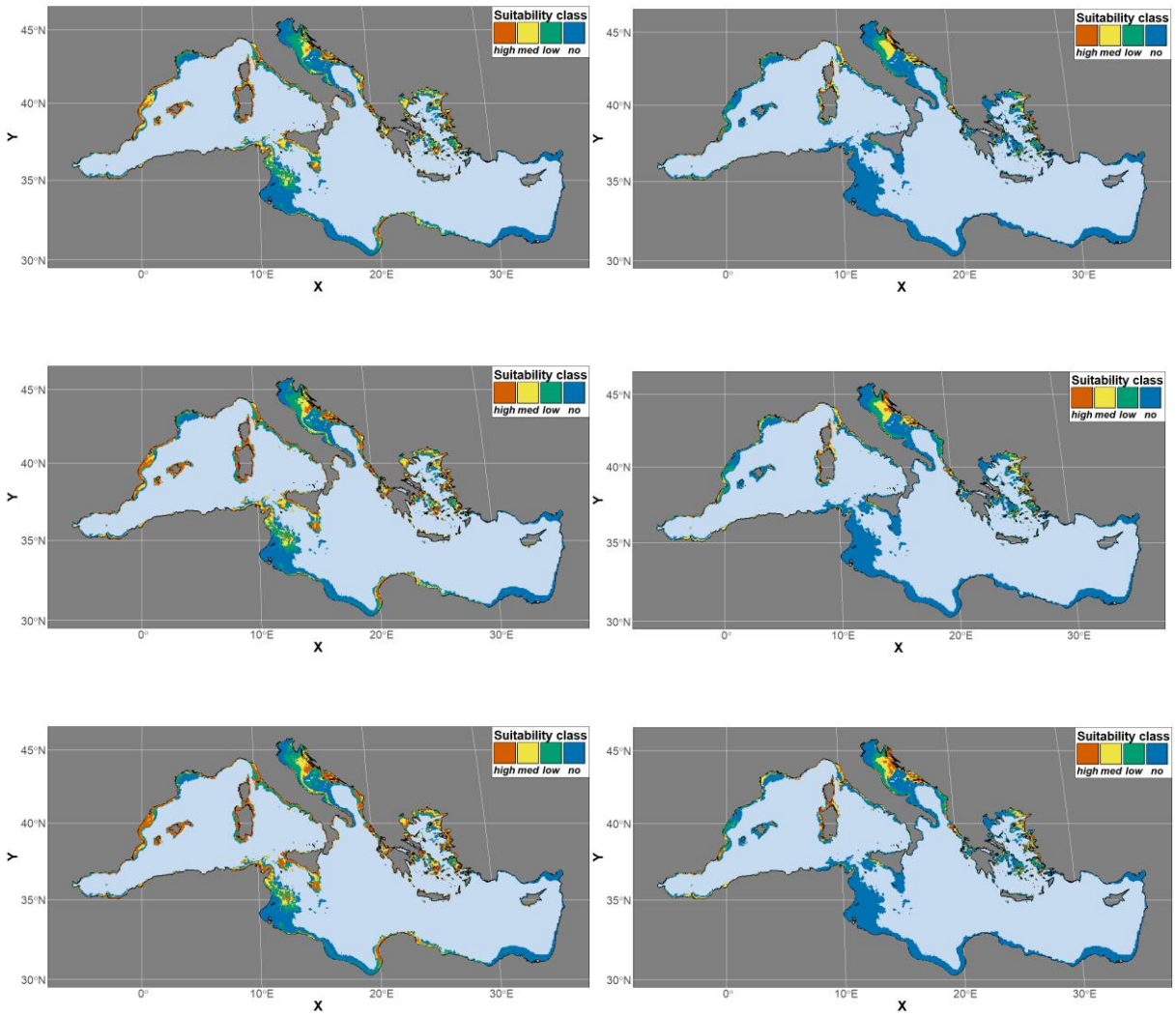

Figure 7. Present (left) and future (right) suitability maps for Caulerpa cylindracea under three alternative maxSSS-based thresholds. From top to bottom: permissive ("best" scenario, maxSSSmax), reference ("normal" scenario, maxSSS) and conservative ("worst" scenario, maxSSSmin).

In parallel, extrapolation risk was quantified with a mobility-oriented parity (MOP) analysis (Owens et al., 2013; Cobos et al., 2019). Environmental conditions in each projection cell were compared to those in the calibration region in multivariate predictor space, and the Euclidean distance to the closest point in the calibration envelope was used as an index of environmental novelty. Distances were classified using chi-squared quantiles into interpolation, borderline and strict extrapolation, with distances above the 90th percentile treated as strict extrapolation.

The resulting MOP rasters form a second key output: mobility-oriented maps that can be overlaid on suitability projections to distinguish robust predictions from those relying on extrapolation beyond the calibration space. At the scale of the Mediterranean shelf, strict extrapolation accounts for only about 0.21% of projected cells, indicating that most forecasts fall within well-known environmental conditions.

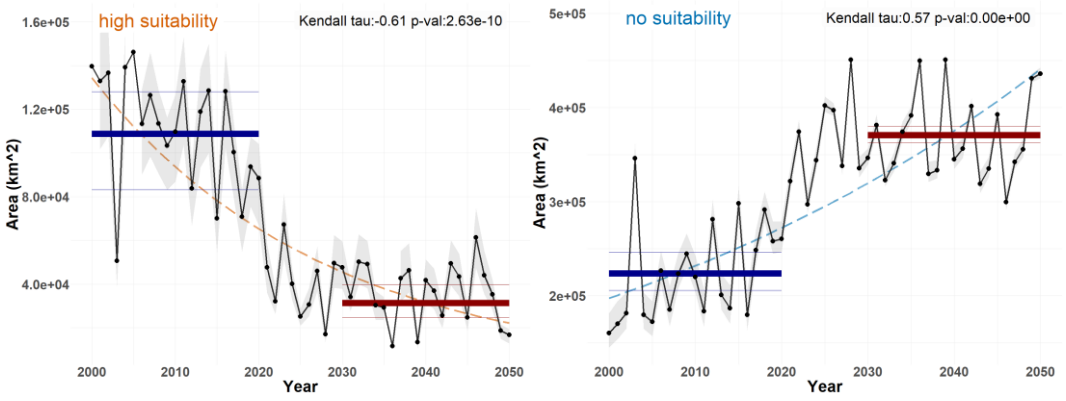

Figure 8. Area ($km2$) for HS and NS classes, as predicted by projecting the model to each available year. The grey area shows the variability associated with threshold uncertainty. Red and blue horizontal lines indicate the area (thick) predicted for present and future.

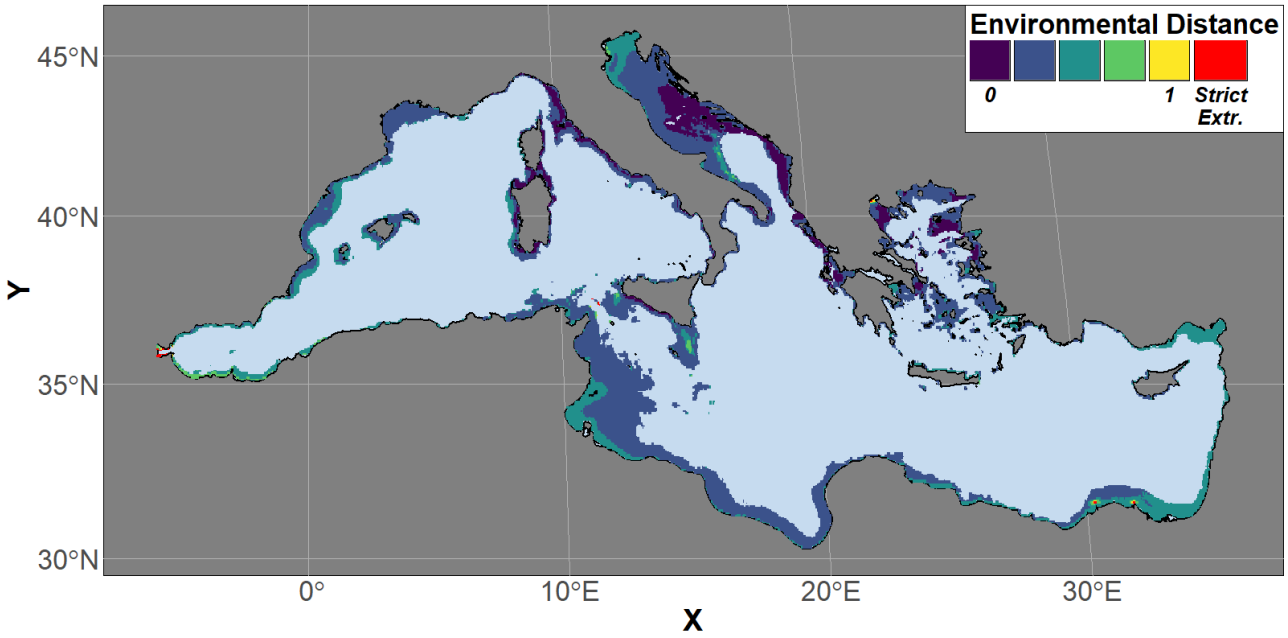

Figure 9. Mobility oriented parity analysis for the future scenario: The blue to yellow gradient indicates the increasing Euclidean distance between the calibration and the future (2030–2050) environmental space. Red signals areas where future conditions deviate strongly from the environmental combinations used to train the models ('strict extrapolation'). It is noteworthy that there are no visible (significant) 'strict extrapolation' areas that account for less than 0.21% of the domain surface. This reflects the quality of the training process and the relative lack of 'climate novelties'.

5. Conclusions

The framework developed here transforms MaxEnt-based ecological niche modelling into a transparent decision-support tool for invasive alien species risk assessment. By combining site-weighted performance metrics with a formal CRITIC–TOPSIS multi-criteria selection, the final model is selected as an explicit compromise between discrimination, transferability, parsimony, and stability, rather than relying on a single metric or informal judgement (Fianchini et al., 2025). This renders the modelling process auditable and defensible in regulatory contexts, where authorities must justify why a particular model and its projections inform management decisions. In this context, the Caulerpa cylindracea case study demonstrates how robust ML–ENM can support evidence-based implementation of IAS monitoring and management.

A second key contribution is the way the framework organises information on risk and uncertainty in space and time for monitoring planners. Continuous probability-of-occurrence outputs are translated into suitability classes using the variability of cross-validated maxSSS thresholds, while MOP-based extrapolation analysis distinguishes areas of reliable interpolation from environmentally novel conditions. Annual projections are further summarised through temporal trend analysis for each suitability class. Together, these products identify (i) core areas of persistently high suitability where long-term surveillance and containment are most urgent, (ii) zones where high suitability overlaps with high uncertainty and should be prioritised for additional sampling, and (iii) regions where suitability is expected to decline or remain marginal, which may require only opportunistic monitoring. The result is a coherent basis for risk-based allocation of limited monitoring resources, rather than uniform or ad hoc sampling strategies.

The workflow is deliberately modular and transferable. Each component – data preparation, site weighting, hyperparameter tuning, multi-criteria ranking, extrapolation analysis, and temporal summarisation – can be adapted to other presence-only algorithms, species, and regions, provided suitable data are available (Fianchini et al., 2025). Although full implementation is data- and computation-intensive, even partial adoption (for example, introducing site-weighted metrics or CRITIC–TOPSIS-based model selection into existing ENM pipelines) would already enhance the robustness and policy relevance of many applications. In the context of "novel protocols for observation and ensuing modelling of space distribution", this framework offers a concrete pathway to integrate state-of-the-art ML–ENM into proactive, adaptive biosecurity planning, helping shift marine IAS management from reactive response to anticipatory, risk-informed monitoring and control.

5.1 Limitations

The proposed ENM framework also has important limitations that should be considered when interpreting its outputs and transferring it to other systems. First, it relies on rich, high-quality data, particularly independent test data with abundance information, which enabled a rigorous, site-weighted evaluation in this case but may restrict the full framework to well-studied species and regions; alternative validation strategies and ancillary data sources will be needed where such datasets are not available.

Second, the exhaustive exploration of the hyperparameter space combined with a complex validation scheme is computationally demanding, so more efficient optimisation strategies are required to maintain scalability for very large datasets.

Finally, like all presence-only, machine-learning ENMs, the framework inherits strong ecological and statistical assumptions (e.g. equilibrium with the environment, no link between predictors and detection) that are often violated in invasion contexts, where adaptation, evolutionary change, human activities and detection biases (such as cryptic life stages or imperfect abundance–suitability relationships) can distort both the data and the inferred niche; projections should therefore be treated as structured hypotheses rather than deterministic forecasts.

6. References

Boudouresque, C., & Verlaque, M. (2012). An overview of species introduction and invasion processes in marine and coastal lagoon habitats. Cahiers de Biologie Marine, 53.

Burnham, K. P., & Anderson, D. R. (2003). Model selection and multimodel inference: A practical information-theoretic approach. Journal of Wildlife Management, 67, 655.

Cantasano, N., Pellicone, G., & Di Martino, V. (2017). The spread of Caulerpa cylindracea in Calabria (Italy) and the effects of shipping activities. Ocean & Coastal Management, 144, 51–58. https://doi.org/10.1016/j.ocecoaman.2017.04.014

Chakraborty, S. (2022). TOPSIS and Modified TOPSIS: A comparative analysis. Decision Analytics Journal, 2, 100021. https://doi.org/10.1016/j.dajour.2021.100021

Cobos, M. E., Peterson, A. T., Barve, N., & Osorio-Olvera, L. (2019). kuenm: An R package for detailed development of ecological niche models using Maxent. PeerJ, 7, e6281. https://doi.org/10.7717/peerj.6281

Coll, M., Piroddi, C., Albouy, C., Lasram, F., Cheung, W., & Christensen, V. (2011). The Mediterranean Sea under siege: Spatial overlap between marine biodiversity, cumulative threats and marine reserves. Global Ecology and Biogeography, 20, 1–16.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fianchini, M., Solidoro, C., & Canu, D. (2024a). 2005-2099 High resolution bioclimatic variables for the surface and bottom of the Mediterranean Sea. (Versione 1.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.12780161

Fianchini, M., Solidoro, C., & Canu, D. (2024b). Downscaled and bias-corrected bioclimatic dataset for the Mediterranean Sea (2005–2099). Data in Brief, 57, 110846. https://doi.org/10.1016/j.dib.2024.110846

Fianchini, M., Solidoro, C., & Melaku Canu, D. (2025). Improving MaxEnt reliability with multi-criteria analysis and site weighting: A case study on Caulerpa cylindracea. Ecological Solutions and Evidence, 6, e70074. https://doi.org/10.1002/2688-8319.70074

GBIF.org. (2024). GBIF Occurrence Download [Dataset]. https://doi.org/10.15468/dl.ctg7ht

Gennaro, P., Piazzi, L., Persia, E., & Porrello, S. (2015). Nutrient exploitation and competition strategies of the invasive seaweed Caulerpa cylindracea. European Journal of Phycology, 50(4), 384–394. https://doi.org/10.1080/09670262.2015.1055591

Haubrock, P. J., Turbelin, A. J., Cuthbert, R. N., Novoa, A., Taylor, N. G., Angulo, E., Ballesteros-Mejia, L., Bodey, T. W., Capinha, C., Diagne, C., Essl, F., Golivets, M., Kirichenko, N., Kourantidou, M., Leroy, B., Renault, D., Verbrugge, L., & Courchamp, F. (2021). Economic costs of invasive alien species across Europe. NeoBiota, 67, 153–190. https://doi.org/10.3897/neobiota.67.58196

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling, 199(2), 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Iveša, L., Djakovac, T., & Devescovi, M. (2015). Spreading patterns of the invasive Caulerpa cylindracea Sonder along the west Istrian Coast (northern Adriatic Sea, Croatia). Marine Environmental Research, 107, 1–7. https://doi.org/10.1016/j.marenvres.2015.03.008

Johnson, J. M., & Clarke, K. C. (2021). An area preserving method for improved categorical raster resampling. Cartography and Geographic Information Science, 48(4), 292–304. https://doi.org/10.1080/15230406.2021.1892531

Klein, J., & Verlaque, M. (2008). The Caulerpa racemosa invasion: A critical review. Marine Pollution Bulletin, 56(2), 205–225. https://doi.org/10.1016/j.marpolbul.2007.09.043

Krishnan, A. R., Kasim, M. M., Hamid, R., & Ghazali, M. F. (2021). A Modified CRITIC Method to Estimate the Objective Weights of Decision Criteria. Symmetry, 13(6), Articolo 6. https://doi.org/10.3390/sym13060973

Liu, C., Newell, G., & White, M. (2016). On the selection of thresholds for predicting species occurrence with presence-only data. Ecology and Evolution, 6(1), 337–348. https://doi.org/10.1002/ece3.1878

Marcelino, V. R., & Verbruggen, H. (2015). Ecological niche models of invasive seaweeds. Journal of Phycology, 51(4), 606–620. https://doi.org/10.1111/jpy.12322

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in Ecology and Evolution, 5(11), 1198–1205. https://doi.org/10.1111/2041-210X.12261

Moreno-Amat, E., Mateo, R. G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.-C., & García-Amorena, I. (2015). Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. Ecological Modelling, 312, 308–317. https://doi.org/10.1016/j.ecolmodel.2015.05.035

Ould-Ahmed, N., & Meinesz, A. (2007). First record of the invasive alga Caulerpa racemosa (Caulerpales, Chlorophyta) on the coast of Algeria. Cryptogamie Algologie, 28, 303–305.

Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E., & Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecological Modelling, 263, 10–18. https://doi.org/10.1016/j.ecolmodel.2013.04.011

Peterson, A. T., Papeş, M., & Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological Modelling, 213(1), 63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography, 31(2), 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

Piazzi, L., Balata, D., Bulleri, F., Gennaro, P., & Ceccherelli, G. (2016). The invasion of Caulerpa cylindracea in the Mediterranean: The known, the unknown and the knowable. Marine Biology, 163, 163:161. https://doi.org/10.1007/s00227-016-2937-4

Pyšek, P., Hulme, P. E., Simberloff, D., Bacher, S., Blackburn, T. M., Carlton, J. T., Dawson, W., Essl, F., Foxcroft, L. C., & Genovesi, P. (2020). Scientists' warning on invasive alien species. Biological Reviews, 95(6), 1511–1534.

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. Journal of Biogeography, 41(4), 629–643. https://doi.org/10.1111/jbi.12227

Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. Biometrics, 69(1), 274–281. https://doi.org/10.1111/j.1541-0420.2012.01824.x

Street, G. M. (2020). Habitat Suitability and Distribution Models with Applications in R. Antoine Guisan, Wilfried Thuiller, and Niklaus E. Zimmermann. 2017. Cambridge University Press, Cambridge, United Kingdom. 462 pp. $49.99 paperback. ISBN: 978-0-521-75836-9. The Journal of Wildlife Management, 84(6), 1212–1213. https://doi.org/10.1002/jwmg.21868

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. Ecography, 32(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Turicchia, E., Ponti, M., Rossi, G., & Cerrano, C. (2021). The Reef Check Med Dataset on Key Mediterranean Marine Species 2001–2020. Frontiers in Marine Science, 8. https://www.frontiersin.org/articles/10.3389/fmars.2021.675574

Uyà, M., Bulleri, F., & Gribben, P. E. (2018). Propagules are not all equal: Traits of vegetative fragments and disturbance regulate invasion success. Ecology, 99(4), 957–965. https://doi.org/10.1002/ecy.2168

Vasquez, M., Allen, H., Manca, E., Castle, L., Lillis, H., Agnesi, S., Al Hamdani, Z., Annunziatellis, A., Askew, N., Bekkby, T., Bentes, L., Doncheva, V., Drakopoulou, V., Duncan, G., Gonçalves, J., Inghilesi, R., Laamanen, L., Loukaidi, V., Martin, S., … Virtanen, E. (2021). EUSeaMap 2021. A European broad-scale seabed habitat map. https://archimer.ifremer.fr/doc/00723/83528/

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. Ecological Applications: A Publication of the Ecological Society of America, 21(2), 335–342. https://doi.org/10.1890/10-1171.1

Wunderlich, R. F., Lin, Y.-P., Anthony, J., & Petway, J. R. (2019). Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. Nature Conservation, 35, 97–116. https://doi.org/10.3897/natureconservation.35.33918

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., & Veran, S. (2013). Presence-only modelling using MAXENT: When can we trust the inferences? Methods in Ecology and Evolution, 4(3), 236–243. https://doi.org/10.1111/2041-210x.12004

Zhang, Z., Mammola, S., & Zhang, H. (2020). Does weighting presence records improve the performance of species distribution models? A test using fish larval stages in the Yangtze Estuary. Science of The Total Environment, 741, 140393. https://doi.org/10.1016/j.scitotenv.2020.140393

Zoheïr, T. (2010). First record of the invasive alga Caulerpa racemosa (Caulerpales, Chlorophyta) in the Gulf of Arzew (western Algeria). Aquatic Invasions, Volume 5, S97–S101. https://doi.org/10.3391/ai.2010.5.S1.020