Ecology and Diversity

SCIEPublish

Article

# Ecological Risk and Public Perception of *Pinctada radiata* Invasion in the Mediterranean: An Integrated SDM, SEM and Sentiment Analysis Framework

**Dimitris Pafras * and Dimitris Klaoudatos**

Department of Ichthyology and Aquatic Environment, University of Thessaly, 38446 Volos, Greece; dklaoud@uth.gr (D.K.)

* Corresponding author. E-mail: dpafras@uth.gr (D.P.)

Received: 20 June 2026; Revised: 9 July 2026; Accepted: 28 July 2026; Available online: 10 August 2026

**ABSTRACT:** The rayed pearl oyster *Pinctada radiata* (Leach, 1814) was the first Indo-Pacific bivalve documented in the Mediterranean Sea and is now among the basin's most widespread non-indigenous bivalves. Despite its long invasion history, basin-scale assessments integrating habitat suitability, anthropogenic exposure, and public perception remain limited. Here, we developed an integrated Mediterranean risk-screening framework that combines ensemble Species Distribution Modelling (SDM), Structural Equation Modelling (SEM), an Invasion Risk Index (IRI), and multilingual sentiment analysis for *P. radiata*. Occurrence data comprised 866 georeferenced records from five Mediterranean countries, of which 133 spatially thinned records were retained for SDM calibration. The ensemble SDM demonstrated high predictive performance (AUC = 0.934 ± 0.021, TSS = 0.812 ± 0.033, CBI = 0.883 ± 0.041) and identified mean annual sea-surface temperature, substrate type, and chlorophyll-*a* as the principal predictors of baseline habitat suitability. More than 81% of dated occurrence records originated from 2020–2026, and the spatially thinned dataset showed a significant westward redistribution through time (Spearman's $\rho = -0.241$, $p = 0.0059$), although this pattern was interpreted cautiously because the dataset was dominated by opportunistic human observations and uneven reporting effort. Structural equation modelling supported consistent spatial associations among environmental suitability, anthropogenic exposure, and the observed distribution of *P. radiata*, while recognising that these relationships represent statistical associations rather than causal ecological mechanisms. The Invasion Risk Index identified High and Very High priority coastal sectors for surveillance, particularly where elevated environmental suitability coincided with intense maritime activity and relatively low current record density. Multilingual sentiment analysis of 1247 text units revealed predominantly neutral-to-positive public and stakeholder discourse, reflecting the species' dual perception as both a non-indigenous organism and a commercially recognised resource. Collectively, these findings provide a spatially explicit framework for supporting surveillance, biosecurity planning, and risk communication for *P. radiata* and other marine non-indigenous species across the Mediterranean.

**Keywords:** Lessepsian migration; Non-indigenous species; Mediterranean Sea; Species distribution model; Structural equation modelling; Invasion risk; Sentiment analysis; *Pinctada radiata*

## 1. Introduction

Marine biological invasions are among the major pressures affecting biodiversity, ecosystem functioning, and coastal ecosystem services worldwide [1–3]. The Mediterranean Sea, which hosts more than 1000 documented non-indigenous species, is considered one of the most heavily invaded marine regions globally [4,5]. This pattern reflects the combined influence of Lessepsian migration through the Suez Canal, intensive maritime transport, high port density, aquaculture activity, and pronounced environmental heterogeneity across the basin [6–8].

Basin-scale spatial assessments can support surveillance by identifying coastal sectors where environmental conditions and anthropogenic exposure may favour the establishment, persistence, or continued detection of non-indigenous species. However, suitability maps derived from fixed climatological layers represent relative environmental suitability only for the analysed reference period. Unless projected explicitly under future climate scenarios, such maps should not be interpreted as forecasts of climate-driven redistribution, contemporary range expansion, or future establishment probability. In the present study, climate-related variables are therefore treated as components of a fixed environmental baseline rather than as indicators of recent marine heatwaves or future climate trajectories.

Among Mediterranean non-indigenous species, the rayed pearl oyster, Pinctada radiata (Leach, 1814), holds a distinctive historical position. It was the first Indo-Pacific bivalve documented in the Mediterranean basin, with an early record from Alexandria, Egypt, in 1874, five years after the opening of the Suez Canal [9], and it is currently recognised as one of the most prominent non-indigenous bivalves in the region [10]. Native to the Indo-Pacific, Indian Ocean, Persian Gulf, and Red Sea, *P. radiata* is a protandrous hermaphroditic bivalve with broad substrate tolerance, occurring on rocky reefs, port structures, aquaculture installations, and other artificial substrates. Its warm-water reproductive biology, attachment capacity, and planktonic larval stage have facilitated its establishment across a wide range of Mediterranean coastal habitats [11,12].

Following its establishment in the eastern Mediterranean, *P. radiata* has been documented progressively farther west, including the Aegean, Ionian, Adriatic, and Tyrrhenian seas, the Sicilian Channel, the Balearic Islands, and the Alboran Sea [13,14]. This widening geographic pattern is consistent with a redistribution of documented occurrence records. However, differences in survey intensity, national data mobilisation, coastal accessibility, and citizen-science participation prevent opportunistic occurrence records from being interpreted as an effort-standardised estimate of biological spread rate, abundance change or climate-driven range expansion.

A substantial body of research has examined Mediterranean *P. radiata* populations, including their population biology, reproductive ecology, habitat suitability, nutritional composition, and stakeholder perceptions [15–18]. Nevertheless, two important knowledge gaps remain. First, no Mediterranean basin-scale assessment has integrated environmental suitability with harmonised indicators of anthropogenic exposure to support spatial monitoring prioritisation while explicitly accounting for the inferential limitations of presence-only data, spatially uneven observation effort and model-derived suitability surfaces. Second, public and stakeholder discourse concerning *P. radiata* has not been evaluated systematically across multiple Mediterranean languages and information sources. Understanding whether the species is framed primarily as an ecological concern, a naturalised non-indigenous organism, or a potentially valuable resource may influence monitoring participation, policy acceptance, and the perceived legitimacy of management responses [19,20].

Species Distribution Models are widely used to estimate relative environmental suitability and identify locations that may warrant increased surveillance [21,22]. Ensemble approaches combine predictions from multiple algorithms and may improve robustness by reducing dependence on the assumptions of a single modelling method [23,24]. Nevertheless, SDM outputs quantify statistical associations between occurrence records and predictor variables. They do not independently demonstrate ecological causation, true occupancy, abundance, establishment probability, or future spread.

Structural modelling can complement distribution modelling by evaluating hypothesised direct and indirect spatial associations among environmental conditions, anthropogenic exposure, and observed occurrence patterns [25,26]. A methodological concern arises when an SDM-derived suitability score is used as the response variable in a structural model containing environmental predictors that were also used to train the SDM, because the resulting pathways are not statistically independent. Accordingly, the primary structural analysis in the present study uses recorded occurrence status across occupied and available background coastal cells. The response variable is therefore independent of the ensemble suitability surface. Background cells are interpreted as available locations rather than confirmed biological absences, and the estimated pathways are interpreted as spatial associations rather than as causal ecological mechanisms or estimates of true occupancy.

Any secondary analysis involving the ensemble suitability surface is treated only as an exploratory description of the internal structure of the modelled surface. It is not used as the principal empirical test of species occurrence and is not interpreted as independent evidence of environmental or anthropogenic causation.

Sentiment analysis offers an additional means of summarising the polarity and thematic structure of public and stakeholder discourse. However, automated multilingual classification may be sensitive to linguistic context, negation, idiomatic expressions, class imbalance, and domain-specific terminology. Its reliability should therefore be assessed through comparison with an independently coded human-consensus reference set and the reporting of an appropriate quantitative agreement indicator. Substantive interpretation of the sentiment outputs should consequently depend on the demonstrated level of agreement between the adapted classifier and the independent human labels.

The objectives of this study were to: (i) compile and curate a Mediterranean occurrence dataset for *P. radiata* from GBIF and published field studies; (ii) estimate relative habitat suitability under a fixed 2000–2020 climatological baseline using an ensemble SDM; (iii) evaluate the spatial associations of environmental conditions, anthropogenic exposure and depth with recorded occurrence status using an occurrence-based structural model; (iv) develop and sensitivity-test a scenario-based spatial surveillance prioritisation index integrating baseline habitat suitability and anthropogenic exposure; and (v) assess multilingual public and stakeholder discourse concerning *P. radiata* and quantify the reliability of the adapted sentiment-classification procedure against independent human-consensus labels.

Based on these objectives, we expected the modelled baseline suitability of *P. radiata* to be most strongly associated with sea-surface temperature, substrate type, and chlorophyll-*a*. We further expected recorded occurrences to be more frequent in available coastal cells characterised by favourable environmental conditions, greater port proximity, and higher vessel-traffic intensity. Because the occurrence dataset consisted of presence-background records collected under spatially uneven observation effort, these expectations concerned relative recorded-occurrence patterns rather than true occupancy, abundance, or causal invasion mechanisms. We also expected the most robust surveillance priorities to occur where elevated baseline suitability and anthropogenic exposure remained consistently high across alternative weighting scenarios. Finally, we anticipated that public and stakeholder discourse would be predominantly neutral to positive, reflecting the species' dual perception as both a non-indigenous organism and a resource of potential commercial value, provided that the multilingual classifier demonstrated acceptable agreement with the independent human-consensus reference labels.

## 2. Materials and Methods

### *2.1. Study Area*

The study area encompasses the entire Mediterranean Sea basin (~2.5 × $10^6$ $km^2$), including the western, central, and eastern sub-basins and their principal semi-enclosed seas: the Adriatic, Ionian, Aegean, and Levantine seas, together with the Tyrrhenian and Ligurian seas, the Sicilian Channel, the Gulf of Gabes, and the Alboran Sea (Figure 1). This geographic extent captures a strong longitudinal gradient in mean annual SST (approximately 13–27 °C), salinity (36.5–39.5 PSU), and surface primary productivity, as well as marked contrasts in maritime traffic density, port infrastructure, and aquaculture activity. Progressive Mediterranean warming, including an estimated SST increase of approximately +0.43 °C per decade since 1982 [6,7] and recent marine heatwaves in 2023–2025 [8], provides ecological context for interpreting the contemporary distribution of warm-affinity non-indigenous species. However, the present model characterises habitat suitability under a 2000–2020 environmental baseline and should not be interpreted as a projection of future warming trajectories. Spatial analysis was conducted on a 0.05° grid (~5.5 km resolution), consistent with the Bio-ORACLE v3.0 environmental data resolution [27]. The analysis was restricted to coastal cells within the 0–150 m bathymetric range, corresponding to the documented depth distribution of *P. radiata* in Mediterranean waters [10,15].

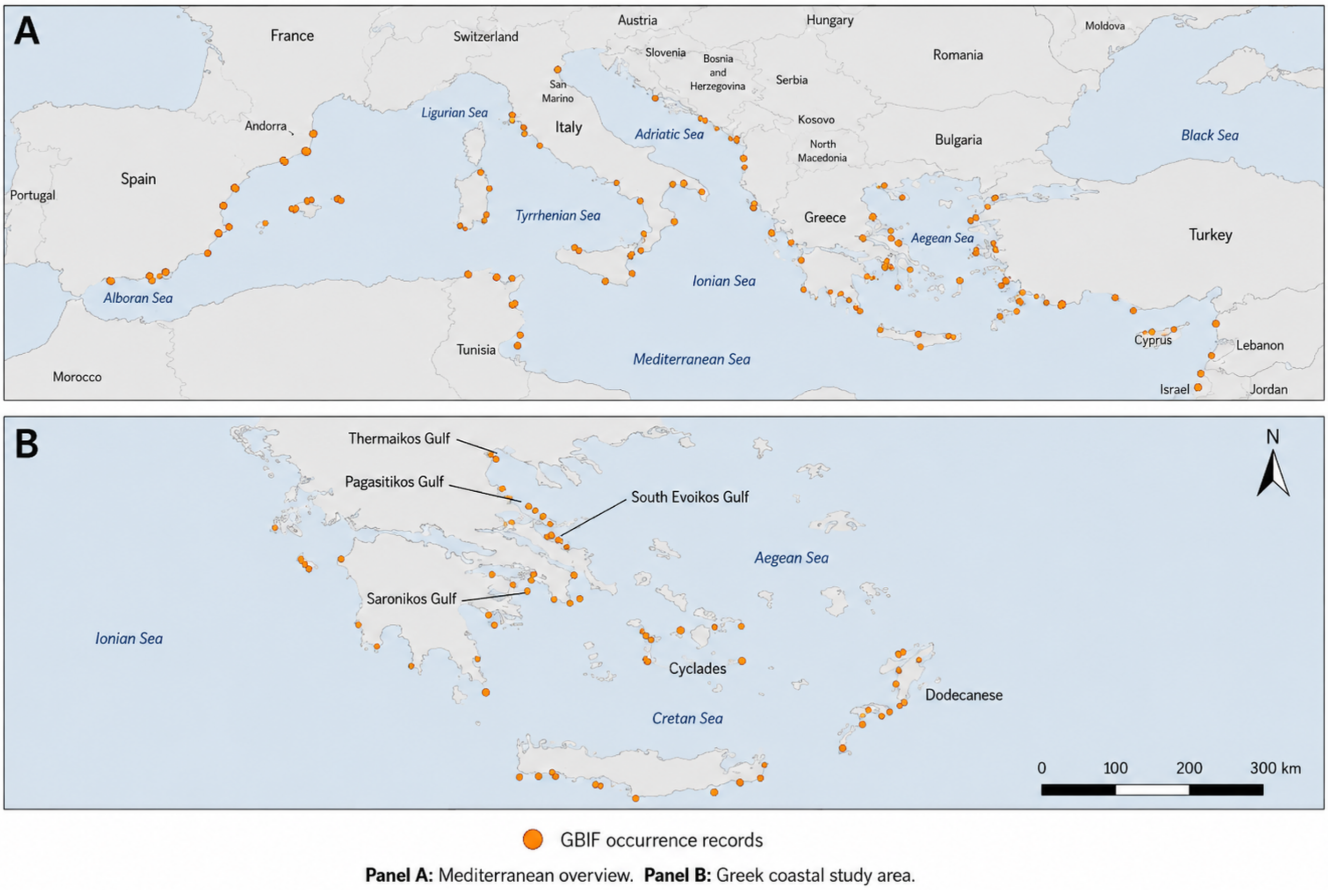


**Figure 1.** Spatial distribution of *P. radiata* occurrence records across the Mediterranean Sea and Greek coastal waters. Panel (**A**) shows the Mediterranean-wide distribution of georeferenced occurrence records, while Panel (**B**) provides a Greece-focused view of the main study area, including the Aegean Sea, Ionian Sea, Cretan Sea, Cyclades, Dodecanese, Saronikos Gulf, South Evoikos Gulf, Pagasitikos Gulf, and Thermaikos Gulf. Orange circles indicate validated coastal/marine occurrence records. Records plotting on land because of shoreline coordinate imprecision were excluded from the map display or adjusted to the nearest valid coastal marine position for visualisation.

## *2.2. Occurrence Data: Compilation and Curation*

Occurrence records for *P. radiata* were downloaded from the Global Biodiversity Information Facility (GBIF; dataset key: 0063413-260519110011954; taxon key: 4374764; June 2026; CC BY-NC 4.0) [28]. The download, filtered to the European/Mediterranean geographic extent, yielded 866 fully georeferenced records spanning 1961–2026 from five countries: Greece ($n$ = 398, 45.9%), Spain ($n$ = 349, 40.3%), Italy ($n$ = 68, 7.8%), Turkey ($n$ = 27, 3.1%) and Cyprus ($n$ = 24, 2.8%). These records were supplemented with georeferenced records from peer-reviewed field studies in Greek coastal waters [15,16] and published Mediterranean literature [11–14].

Records were sourced primarily from iNaturalist citizen-science observations (basis of record: HUMAN_OBSERVATION; 94%), supplemented by museum specimens and structured field surveys. Because citizen-science observations may be spatially biased toward accessible coastlines, tourist areas, urbanised coastal zones, and regions with higher observer effort, this source composition was explicitly considered during data curation and interpretation. Spatial thinning at 0.5° was applied to reduce clustering and over-representation of highly sampled coastal sectors. However, thinning and spatial block cross-validation cannot fully correct observer-effort bias. Target-group background sampling or explicit bias layers were considered but not implemented because comparable Mediterranean-wide observer-effort layers for marine bivalves were not available. Consequently, SDM and IRI outputs are interpreted as risk-screening products rather than definitive maps of realised abundance or true distribution.

Data curation followed five sequential steps: (1) coordinate verification against EMODnet Bathymetry 2022 and the Global Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG) coastline layer. Records plotting on land were inspected individually and excluded from map display and model calibration when their position could not be reconciled with a valid marine/coastal location within the documented 0–150 m depth range of the species; (2) taxonomic standardisation retaining only records assigned to *P. radiata* or the synonym *P. imbricata radiata*; (3) removal of 48 duplicate records with identical taxon, coordinates and date; (4) temporal filtering retaining records from 1982 onwards because 1982 marks the beginning of the continuous satellite SST era commonly used in Mediterranean warming assessments, while Bio-ORACLE v3.0 predictors represent the 2000–2020 climatological baseline [27]. This creates a temporal mismatch for records collected during 1982–1999; these records were retained because they improve spatial coverage of the species' historical Mediterranean range, and the potential mismatch is considered in the interpretation of model outputs; and (5) spatial thinning at 0.5° using the spThin R package [29], yielding a final SDM-calibration dataset of 133 records at 128 unique 0.5° grid cells (Table 1). Spatial thinning retained 133 occurrence records representing 128 unique 0.5° grid cells. For cell-based occurrence modelling, duplicate records within the same 0.05° analysis cell were collapsed to a single occupied-cell observation.

Spatial thinning at 0.5° was used as a conservative correction for spatial clustering in citizen-science occurrence records at the Mediterranean-basin scale. This distance was not intended to represent the dispersal distance or ecological movement scale of *P. radiata*, which is sessile after settlement. Instead, it was selected to reduce the dominance of heavily sampled coastal sectors in the calibration dataset and to improve spatial independence among occurrence records. We recognise that this thinning distance is coarse for a coastal bivalve, and this choice is therefore treated as a methodological limitation.

To maximise transparency and reproducibility, all occurrence records underwent a predefined quality-control workflow before analysis. The workflow included duplicate removal, taxonomic verification, temporal filtering, coordinate validation, correction or exclusion of records located on land because of shoreline georeferencing imprecision, and spatial thinning to minimise spatial sampling bias and autocorrelation. Only records passing all quality-control steps were retained for the SDM and occurrence-based analyses. The complete filtering workflow is illustrated in Supplementary Figure S1.

**Table 1.** *P. radiata* occurrence dataset for the Mediterranean Sea. N_raw: raw GBIF records; N_SDM: records retained after spatial thinning (0.5°) for SDM calibration.

| Sub-Basin | Countries | N_raw | N_SDM (%) | Year Range | Invasion Status |
|---|---|---|---|---|---|
| Alboran Sea/southern Spain | ES | 243 | 32 (24.1%) | 2013–2026 | Established, westward expansion front |
| Aegean Sea | GR, TR | 220 | 28 (21.1%) | 1961–2026 | Core range |
| Eastern Mediterranean (Crete, Dodecanese) | GR, CY | 154 | 20 (15.0%) | 1970–2026 | Core range |
| Western Mediterranean (Catalonia, Balearics) | ES | 106 | 14 (10.5%) | 2021–2026 | Recent expansion spread front |
| Ionian Sea | GR, IT | 78 | 11 (8.3%) | 1995–2026 | Established |
| Eastern Levantine Sea | TR | 28 | 8 (6.0%) | 1999–2026 | Core range (1874 origin) |
| Adriatic Sea | IT | 18 | 7 (5.3%) | 2019–2025 | Active spread front (est. 2019) |
| Tyrrhenian Sea | IT | 10 | 7 (5.3%) | 2021–2025 | Established |
| Central Mediterranean (Sicily) | IT | 9 | 6 (4.5%) | 2003–2025 | Established |
| TOTAL | GR, ES, IT, TR, CY | 866 | 133 (100%) | 1961–2026 | — |

Source: GBIF dataset key 0063413-260519110011954 (June 2026). Spatial thinning at 0.5° applied after duplicate removal and temporal filtering (≥1982). Field survey records from Pafras et al. [15,16] are included in sub-basin totals.

Temporal and Geographic Structure of Occurrence Reporting

To examine the temporal and geographic structure of the available occurrence data, we conducted an additional analysis using both the raw curated records and the spatially thinned dataset. Records with a valid observation year were summarised annually, and mean annual reporting intensity was compared between 2010–2019 and 2020–2026. The year 2020 was used as a descriptive breakpoint because the dataset showed a pronounced increase in record availability after that year.

The association between observation year and longitude was evaluated using Spearman's rank correlation. This analysis was first conducted on all records with a valid year and subsequently repeated using the spatially thinned and geographically quality-controlled dataset to reduce the influence of clustered observations. Geographic quality assurance excluded records outside the predefined Mediterranean extent and records located more than 25 km from the reference coastline.

Median longitude was compared between records dated before 2020 and those dated during 2020–2026 using the Mann–Whitney U test. The difference in median longitude was accompanied by a non-parametric bootstrap 95% confidence interval. Changes in the distribution of records among Mediterranean sub-basins were evaluated using a permutation test applied to the period-by-sub-basin contingency table, and effect size was quantified using Cramer's V.

These analyses were intended to characterise changes in the spatial and temporal composition of available reports. Because observation effort was not standardised, the results were not interpreted as direct estimates of biological spread rate or abundance change.

### *2.3. Environmental and Anthropogenic Predictor Variables*

Sixteen candidate predictor variables were initially assembled and screened for collinearity using Pearson correlation coefficients ($|r| < 0.70$) and the Variance Inflation Factor (VIF < 5). Nine non-collinear predictors were retained for SDM and SEM calibration (Table 2). Oceanographic and climatic variables were extracted from Bio-ORACLE v3.0 [27] at 0.05° resolution for the 2000–2020 reference period: mean annual SST (SST_mean, °C), minimum SST of the coldest month (SST_min, °C), mean sea-surface salinity (SAL, PSU), mean chlorophyll-*a* concentration (CHL, mg·m$^{-3}$) and mean dissolved oxygen (DO, ml·L$^{-1}$).

Dissolved oxygen was retained in the unit convention provided by Bio-ORACLE v3.0 to maintain consistency with the original environmental layer.

Habitat variables from the EMODnet Bathymetry, Geology, and Seabed Habitats portals included depth derived from the GEBCO 2023, and substrate type, represented by four categories: hard/rocky, coarse sediment, sandy/mixed, and muddy sediment. Substrate type was treated as a categorical predictor. For GLM, substrate categories were encoded using dummy variables, whereas RF and GBM handled substrate as a factor variable. For MaxEnt, substrate was included as a categorical environmental layer to avoid imposing an artificial ordinal structure among substrate classes.

Anthropogenic variables were used as harmonised basin-wide proxies for propagule pressure rather than as exhaustive representations of all introduction and secondary-spread pathways. These variables comprised distance to the nearest commercial or fishing port (DIST_PORT, km; World Port Index 2023) and AIS maritime traffic intensity (vessel-hours $km^{-2}\cdot yr^{-1}$; EMODnet Human Activities 2022). They were selected because they are spatially explicit and consistently available across the Mediterranean basin. However, they do not capture all relevant vectors. Important omitted pathways include aquaculture facilities, shellfish translocations, marina density, recreational boating, hull fouling, ballast-water discharge intensity, and local seafood trade. The AP construct should therefore be interpreted as a conservative proxy for basin-scale anthropogenic exposure rather than as a complete representation of all invasion pathways.

Variables excluded due to VIF > 5 or |r| > 0.70 are documented in Table 2. SST annual range and seafloor slope were excluded due to high VIF values, while current velocity was excluded because it was strongly correlated with chlorophyll-*a* (|r| = 0.81). Chlorophyll-*a* was retained because it provides a direct proxy for productivity and food availability for a suspension-feeding bivalve. Nevertheless, current velocity is mechanistically relevant for larval transport, settlement, and hydrodynamic exposure. Its exclusion should therefore be interpreted as a pragmatic collinearity decision rather than evidence that hydrodynamics are unimportant for *P. radiata* distribution. Spatial maps of the retained environmental, habitat, and anthropogenic predictor layers are provided in Supplementary Figure S2.

**Table 2.** Predictor variable set for ensemble SDM and SEM calibration after collinearity screening (Pearson |r| < 0.70; VIF < 5). Retained variables are used in both SDM and SEM. Excluded variables are listed for transparency.

| Category | Variable | Symbol | Source | Resolution | VIF | Retained |
|---|---|---|---|---|---|---|
| Oceanographic | Mean annual SST (°C) | SST_mean | Bio-ORACLE v3.0 | 0.05° | 1.93 | Yes |
| Oceanographic | Min. SST coldest month (°C) | SST_min | Bio-ORACLE v3.0 | 0.05° | 2.14 | Yes |
| Oceanographic | Mean salinity (PSU) | SAL | Bio-ORACLE v3.0 | 0.05° | 2.87 | Yes |
| Oceanographic | Chlorophyll-*a* ($mg\cdot m^{-3}$) | CHL | Bio-ORACLE v3.0 | 0.05° | 3.12 | Yes |
| Oceanographic | Dissolved oxygen ($ml\cdot L^{-1}$) | DO | Bio-ORACLE v3.0 | 0.05° | 2.41 | Yes |
| Habitat | Depth/bathymetry (m) | DEPTH | GEBCO 2023/EMODnet | 15 arc-sec | 2.63 | Yes |
| Habitat | Substrate type (4 categories) | SUB | EMODnet Geology | 0.05° | 1.44 | Yes |
| Anthropogenic | Distance to nearest port (km) | DIST_PORT | World Port Index 2023 | 0.05° | 1.71 | Yes |
| Anthropogenic | AIS vessel traffic density | AIS | EMODnet Human Activities | 0.10° | 2.05 | Yes |
| Excluded | SST annual range (°C) | SST_range | Bio-ORACLE v3.0 | 0.05° | 6.74 | No (VIF > 5) |
| Excluded | Current velocity ($m\cdot s^{-1}$) | VEL | Bio-ORACLE v3.0 | 0.05° | — | No (\|r\| = 0.81) |
| Excluded | Seafloor slope (°) | SLOPE | GEBCO 2023 | — | 5.38 | No (VIF > 5) |

## 2.4. Ensemble Species Distribution Modelling

An ensemble SDM was implemented using four algorithms in the biomod2 R package (v. 4.2.5; [30]): Maximum Entropy Modelling (MaxEnt; [31]), Random Forest (RF; [32]), Gradient Boosting Machine (GBM; [33]), and Generalized Linear Model (GLM). MaxEnt regularisation settings were tuned using ENMeval [34]. Candidate models were fitted by testing regularisation multipliers from 0.5 to 4.0, in 0.5 increments, together with combinations of linear, quadratic, hinge, and product feature classes. The final

MaxEnt configuration was selected by balancing model parsimony and predictive performance, resulting in an optimal regularisation multiplier of 1.5 with hinge and product features. For MaxEnt, 10,000 background points were sampled within an ecologically accessible area (M) defined as a 200 km coastal buffer around validated occurrence points, restricted to the 0–150 m bathymetric range and clipped to the Mediterranean Sea boundary. For RF, GBM, and GLM, pseudo-absences were generated within the same accessible area M to ensure spatial consistency among algorithms. A total of 10,000 pseudo-absence points were used for each presence–absence algorithm, corresponding to an approximately 75:1 pseudo-absence ratio based on the 133 spatially thinned presence records retained for SDM calibration. Pseudo-absences were sampled outside occupied grid cells and were restricted to the same bathymetric and coastal accessibility constraints used for the MaxEnt background sample.

Although this pseudo-absence-to-presence ratio is high, it was used to provide stable background characterisation across the accessible coastal area; the retained ensemble showed robust cross-algorithm performance, including acceptable GLM discrimination and calibration, indicating that no single algorithm dominated the consensus prediction. Because pseudo-absence/background sampling was conducted within the accessible coastal area rather than using an explicit observer-effort surface, the SDM should be interpreted as a habitat-suitability and risk-screening model rather than as an effort-corrected estimate of realised occupancy.

Model evaluation followed a five-fold spatial block cross-validation design implemented with the blockCV package [35]. Spatial blocks of 200 km were selected based on the estimated spatial autocorrelation range of the occurrence data using Moran's I. This approach minimises data leakage caused by spatial autocorrelation, a major source of inflated performance estimates in random cross-validation designs [24]. Model performance was assessed using the Area Under the Receiver Operating Characteristic Curve (AUC), the True Skill Statistic (TSS = sensitivity + specificity − 1), and the continuous Boyce Index (CBI; [36]). Variable importance was quantified using permutation-based methods and averaged across algorithms and cross-validation folds. Individual algorithm predictions were generated as continuous suitability surfaces ranging from 0 to 1 and combined into a TSS-weighted ensemble. Ensemble uncertainty was expressed as the coefficient of variation (CV) across algorithm-specific predictions. The robustness of the selected thinning distance was further evaluated through sensitivity analyses using alternative thinning thresholds (0.25°, 0.5°, and 1°), confirming that the principal spatial suitability patterns remained stable.

### *2.5. Occurrence-Based Structural Analysis*

To avoid the statistical circularity associated with using an SDM-derived suitability surface as the primary response variable, the main structural analysis was reformulated using observed occurrence status across Mediterranean coastal grid cells. Grid cells were coded as occupied when they contained at least one validated, spatially thinned *P. radiata* record and as available background cells when they occurred within the same accessible coastal area used for SDM calibration but contained no validated occurrence record. Background cells were therefore treated as available locations rather than as confirmed biological absences.

The occurrence-based analysis was restricted to the same geographic extent, 0.05° spatial resolution and 0–150 Environmental Conditions Index bathymetric range used in the SDM. Environmental conditions were represented by an Environmental Conditions Index derived independently from the ensemble SDM output. The index combined standardised mean annual sea-surface temperature, chlorophyll-*a* concentration, and substrate-category information. No values from the ensemble suitability surface were used to construct this predictor. Because substrate type was categorical, it was represented using category-specific dummy variables rather than an artificial continuous gradient. Anthropogenic Pressure was represented by a standardised composite of port proximity and AIS vessel density. Port proximity was derived from the inverse distance to the nearest port, so that larger values represented greater proximity to maritime infrastructure. Depth was included as an additional ecological covariate because *P. radiata* is predominantly associated with shallow coastal habitats.

The primary empirical model was fitted using binomial logistic regression with a logit link. Binary recorded-occurrence status was the response variable, with occupied coastal grid cells coded as 1 and available background cells coded as 0. The Environmental Conditions Index, Anthropogenic Pressure Index, and depth were entered as predictors. Odds ratios and 95% confidence intervals were calculated from the model coefficients. Model performance was evaluated using McFadden's pseudo-$R^2$. Because the response represented presence versus available background rather than confirmed presence–absence data, the fitted probabilities were interpreted as relative probabilities of recorded occurrence rather than estimates of true occupancy.

The occurrence-based model was treated as the primary empirical analysis because its response variable was independent of the ensemble suitability surface. Nevertheless, the model was based on observational presence–background data and spatially uneven recording effort. Its coefficients were therefore interpreted as statistical spatial associations rather than as causal effects, mechanistic pathways, or direct estimates of establishment probability.

A secondary Structural Equation Model was retained to describe the internal association structure of the continuous ensemble SDM-derived suitability surface. In this exploratory model, Environmental Suitability was specified as a latent construct represented by mean annual sea-surface temperature, chlorophyll-*a*, and substrate type, whereas Anthropogenic Pressure was represented by port proximity and AIS vessel density. The structural pathways comprise Environmental Suitability SDM-derived suitability, Anthropogenic Pressure → SDM-derived suitability, and Environmental Suitability → Anthropogenic Pressure, together with the indirect pathway Environmental Suitability → Anthropogenic Pressure → SDM-derived suitability.

The conceptual framework illustrating the occurrence-based analytical approach, including the Environmental Conditions Index, Anthropogenic Pressure Index, and their hypothesised relationships with recorded occurrence, is presented in Figure 2.

The secondary model was fitted using full-information maximum likelihood estimation in the lavaan R package (version 0.6-18) [37]. Model fit was assessed using $\chi^2$/df, the Comparative Fit Index, the Root Mean Square Error of Approximation, and the Standardised Root Mean Square Residual [38,39]. Bias-corrected 95% confidence intervals for the indirect association were estimated using 5000 bootstrap resamples, and standardised pathway coefficients were reported. Because the response surface was generated using environmental predictors overlapping with those included in the structural model, this analysis was interpreted exclusively as an exploratory decomposition of modelled suitability and not as independent evidence of species occurrence, establishment or causal invasion processes.

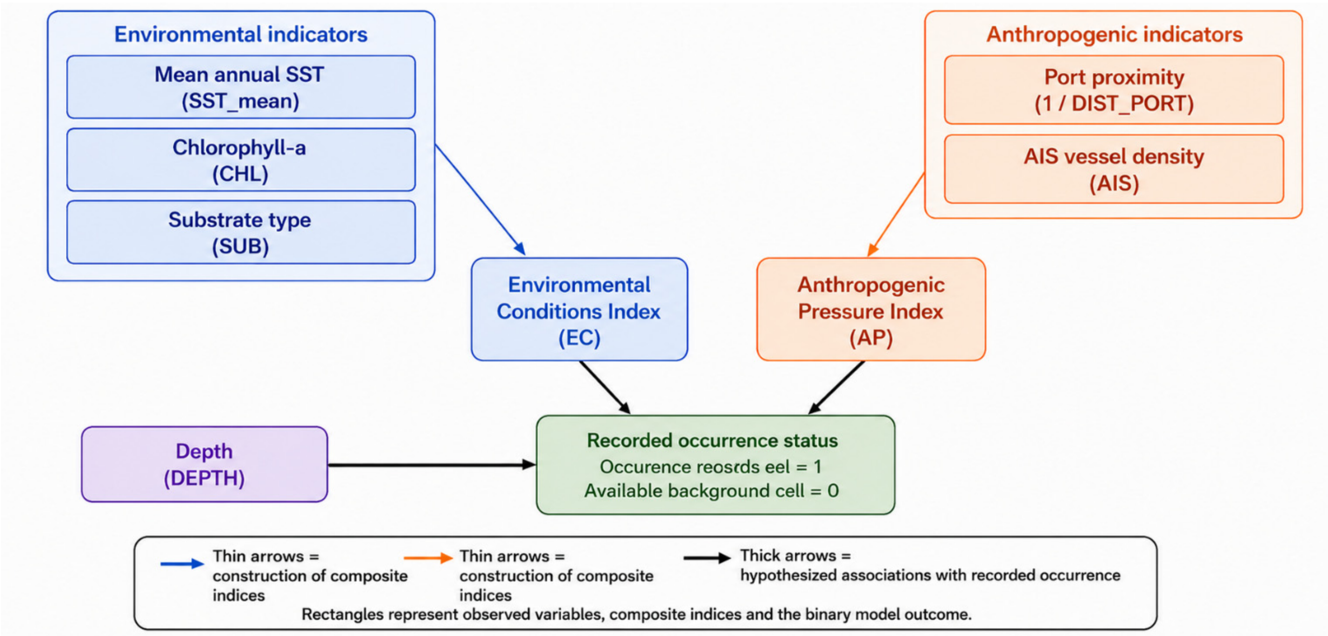


**Figure 2.** Conceptual framework for the primary occurrence-based analysis of *P. radiata* across Mediterranean coastal grid cells. Recorded occurrence status was coded as occupied (1) for cells containing at least one validated spatially thinned record and as available background (0) for cells within the accessible coastal area without a validated record. The Environmental Conditions Index was constructed independently of the ensemble SDM output using mean annual sea-surface temperature (SST_mean), chlorophyll-*a* concentration (CHL), and substrate-category information (SUB). The Anthropogenic Pressure Index combined port proximity (1/DIST_PORT) and AIS vessel density (AIS), while depth was included as an additional ecological covariate. Thin arrows indicate the construction of the composite indices, and thick arrows indicate the hypothesised statistical associations with recorded occurrence. Background cells represent available locations rather than confirmed biological absences; therefore, the pathways should not be interpreted as causal ecological effects or estimates of true occupancy.

### *2.6. Mediterranean Invasion Risk Index (IRI)*

A composite Mediterranean Invasion Risk Index (IRI) was calculated for each 0.05° coastal grid cell as a spatial prioritisation tool rather than as a definitive estimate of establishment probability. Before index construction, all component variables were rescaled to the 0–1 range using min–max normalisation.

The Anthropogenic Pressure Index (AP) was calculated as the equally weighted mean of normalised port proximity and normalised AIS vessel density:

$$AP = 0.50 \times PORT + 0.50 \times AIS \tag{1}$$

where $PORT$ represents normalised inverse distance to the nearest port $(1/DIST\_PORT)$, and $AIS$ represents normalised vessel density. Equal a priori weighting was adopted because no independent empirical evidence was available to justify assigning greater importance to either anthropogenic-pressure proxy.

The primary IRI was then calculated by combining the normalised ensemble SDM suitability score $(\hat{S})$ and the normalised Anthropogenic Pressure Index $(\widehat{AP})$ using equal a priori weights:

$$IRI = 0.50 \times \hat{S} + 0.50 \times \widehat{AP} \tag{2}$$

where $\hat{S}$ is the normalised ensemble suitability score and $\widehat{AP}$ is the normalised Anthropogenic Pressure Index. Equal weighting was selected because the purpose of the IRI was spatial prioritisation for surveillance rather than estimating causal effect sizes. This approach also ensured that the index was independent of coefficients derived from the structural equation model, thereby avoiding potential circularity between the explanatory analysis and the subsequent prioritisation framework.

To assess the sensitivity of the spatial prioritisation to the weighting assumptions, four alternative weighting scenarios were also examined: 0.60/0.40, 0.40/0.60, 0.70/0.30, and 0.30/0.70 for environmental suitability and anthropogenic pressure, respectively. Coastal cells that remained classified as High or Very High priority across the alternative scenarios were considered robust surveillance priorities.

The continuous primary IRI surface was classified into four priority categories using fixed interpretive thresholds: Low ($IRI < 0.25$), Moderate ($0.25 \leq IRI < 0.50$), High ($0.50 \leq IRI < 0.75$), and Very High ($IRI \geq 0.75$). Priority monitoring areas were identified as grid cells classified as High or Very High and characterised by a current occurrence density of fewer than one validated record per 0.25° cell. The same classification thresholds were applied to all alternative weighting scenarios to enable direct comparison.

### *2.7. Multilingual Sentiment Analysis*

A multilingual sentiment analysis was conducted to assess public and stakeholder discourse surrounding *P. radiata* across Mediterranean contexts. Textual data were compiled from three sources: (1) iNaturalist observation comments attached to Mediterranean occurrence records (*n* = 214 comments, 2019–2026); (2) multilingual news articles and press content retrieved via Google News and press databases in Greek, Spanish and Italian (*n* = 87 text units, 2018–2026); and (3) stakeholder open-ended responses from the published HGPS study of Pafras et al. [26] (N = 80 participants, eastern Mediterranean). Search terms included the scientific name *P. radiata*, the synonym *P. imbricata radiata*, and Mediterranean vernacular names in Greek, Spanish, and Italian.

All textual material was cleaned prior to analysis. Duplicate entries, automated or bot-generated content, non-relevant mentions, and purely taxonomic or location-only records were removed. Stakeholder responses from Pafras et al. [26] were segmented into sentence-level or thematically coherent text units, depending on response structure. This procedure yielded 946 stakeholder-derived text units from 80 participants. Because multiple text units could originate from the same participant, text-unit-level sentiment results were treated as descriptive for corpus-level discourse, whereas inferential comparisons among stakeholder profiles were conducted only after aggregating sentiment scores at the participant level.

Sentiment classification was performed using the VADER framework [40], adapted for multilingual Mediterranean content through a trilingual lexicon extension incorporating Greek, Spanish, and Italian emotional terms, fisheries-related terminology, and invasion-governance vocabulary. Each text unit received a compound sentiment score ranging from −1, indicating strongly negative sentiment, to +1, indicating strongly positive sentiment. Scores were classified into five polarity categories using the conventional VADER thresholds: Very Positive (>0.5), Positive (0.05 to 0.5), Neutral (−0.05 to 0.05), Negative (−0.5 to −0.05), and Very Negative (<−0.5).

To assess the reliability of the multilingual VADER adaptation, we conducted an expert-assisted lexicon calibration and qualitative reliability screening procedure. A stratified subset of text units covering Greek, Spanish, and Italian material and the three source categories was independently reviewed by three coders with knowledge of Mediterranean fisheries, non-indigenous species terminology, and local socio-ecological context. The purpose of this step was not to produce a balanced classifier-performance benchmark, but to identify systematic misclassifications, ambiguous polarity assignments, negation errors, fisheries-specific idioms, and terms whose sentiment depended on ecological or commercial context. Disagreements between human-coded polarity labels and VADER-derived labels were reviewed jointly and used to refine the trilingual lexicon before the final corpus-level sentiment classification. Because the reviewed subset was designed primarily for lexicon calibration rather than for balanced classifier benchmarking, a basic overall agreement statistic was calculated as the percentage of validation text units receiving identical polarity classifications from the adapted VADER classifier and the human-consensus reference labels. Overall, VADER–human agreement was 85.0%. Language-specific accuracy, macro-averaged F1-score, precision, and recall were not calculated. The sentiment outputs are therefore interpreted as exploratory corpus-level discourse indicators rather than as fully validated supervised classification results.

Thematic analysis of dominant discourse themes followed Braun and Clarke [41]. Three independent coders assigned text units to recurrent themes, and inter-rater reliability was assessed using Cohen's κ. The thematic coding was used to characterise the main interpretive frames through which *P. radiata* was discussed, including economic opportunity, ecological concern, regulatory ambiguity, naturalisation narratives and citizen-science engagement.

For comparison among the three cognitive stakeholder profiles identified by Pafras et al. [26], Ecological Pragmatists, Institutional Collaborators, and Adaptive Stewards, compound sentiment scores were first averaged at the participant level. A one-way ANOVA was used to test for differences in participant-level mean sentiment scores among the three stakeholder profiles, followed by Tukey's Honestly Significant Difference (HSD) post-hoc test for pairwise comparisons when the omnibus ANOVA was significant [42]. This aggregation avoided pseudo-replication that would have arisen from segmenting individual stakeholder responses into multiple text units. Corpus-level sentiment distributions are therefore presented as descriptive summaries of multilingual discourse, whereas statistical inference is restricted to participant-level comparisons of stakeholders. Sentiment analyses were performed in Python 3.11 using pandas, vaderSentiment, scipy, and nltk.

### *2.8. Statistical Software and Reproducibility*

All SDM and SEM analyses were performed in R version 4.4.2 (R Core Team, 2024). Sentiment analyses were performed in Python 3.11. Key R packages used were sf (v. 1.0-16), terra (v. 1.7-71), biomod2 (v. 4.2.5), blockCV (v. 2.1.4), ENMeval (v. 2.0.4), spThin (v. 0.2.0), lavaan (v. 0.6-18), usdm (v. 2.1-7), and tmap (v. 3.3-4). All R scripts, the curated occurrence dataset, and the Python sentiment-analysis code will be deposited in Zenodo before final publication; a private review link can be provided upon editorial request.

## 3. Results

### *3.1. Occurrence Dataset Overview*

The curated Mediterranean dataset comprised 866 georeferenced GBIF records from five countries, extending from Almería, Spain, in the Alboran Sea (westernmost record: 2.46° W) to İskenderun, Türkiye, in the eastern Levantine Sea (easternmost record: 36.17° E). Following the sequential data-curation procedure, 805 records met the post-1982 temporal criterion, and 133 spatially thinned records were retained for SDM calibration at a minimum separation distance of 0.5° (Table 1).

The temporal distribution of records was strongly concentrated in the most recent years. Mean annual reporting increased from 10.8 records $yr^{-1}$ during 2010–2019 to 94.0 records $yr^{-1}$ during 2020–2026, with the highest annual totals recorded in 2025 ($n = 250$) and 2024 ($n = 142$). Of the 809 records with a valid observation year, 658 (81.3%) were dated between 2020 and 2026. Because the dataset was dominated by citizen-science observations, this increase cannot be interpreted solely as evidence of population growth or geographic expansion. It is more likely to reflect a combination of biological redistribution, increased observer participation, wider use of digital biodiversity platforms, and improved mobilisation of occurrence data.

Recent records from previously sparsely documented sectors, including the Adriatic Sea, the western Mediterranean mainland coast, and the northern Tyrrhenian Sea, are consistent with an ongoing broadening of the documented distribution of *P. radiata*. Nevertheless, the timing and magnitude of this apparent redistribution remain uncertain because sampling effort was neither spatially nor temporally standardised. The monthly distribution also showed a pronounced seasonal reporting pattern, with the highest number of records occurring in August ($n = 102$; 16.0% of records with month information). This summer peak may partly reflect increased coastal activity and observer effort rather than seasonal variation in the species' abundance or detectability. Distinguishing biological spread from changes in recording intensity will require repeated, effort-standardised monitoring across Mediterranean coastal sectors.

### *3.2. Temporal and Longitudinal Redistribution of Occurrence Records*

Observation year was negatively associated with longitude in the raw occurrence dataset (Spearman's $\rho = -0.452$, $p < 0.001$), indicating that more recent reports were generally located farther west. Because the raw dataset was strongly affected by clustered citizen-science observations, the relationship was re-evaluated after spatial thinning and geographic quality assurance.

Of the 133 spatially thinned records, 129 were retained following geographic quality control. One record occurred outside the predefined Mediterranean extent, and three records were located more than 25 km from the reference coastline. Within the quality-controlled thinned dataset, the negative association between year and longitude remained statistically significant (Spearman's $\rho = -0.241$, $p = 0.0059$).

Median longitude shifted from 24.47° E among records collected before 2020 to 17.88° E among records collected during 2020–2026. The difference of −6.59° longitude was supported by a bootstrap 95% confidence interval of −10.00° to −2.37°. The longitude distributions also differed significantly between periods according to the Mann–Whitney U test ($p < 0.001$).

The sub-basin composition of records differed significantly between the two periods (permutation $p$ = 0.0034), with a Cramer's V of 0.419, indicating a moderate-to-strong association between reporting period and sub-basin. Collectively, these findings demonstrate a recent westward redistribution in the geographic composition of available records. However, they do not distinguish biological range expansion from changes in observer density, platform use, or national data mobilisation.

### *3.3. SDM Performance and Variable Importance*

All four individual algorithms achieved performance well above the selected discrimination thresholds (AUC > 0.85; TSS > 0.70; Table 3). The TSS-weighted ensemble attained AUC = 0.934 ± 0.021, TSS = 0.812 ± 0.033 and CBI = 0.883 ± 0.041. Ensemble uncertainty was consistently low in high-suitability zones (mean CV = 0.09 ± 0.03), indicating high inter-algorithm consensus in areas predicted as highly suitable.

Although RF achieved the highest individual AUC, the TSS-weighted ensemble was retained because it prioritises cross-algorithm robustness and spatial consensus rather than maximising a single discrimination metric.

**Table 3.** SDM performance metrics for *P. radiata* across the Mediterranean Sea per algorithm and for the TSS-weighted ensemble. Values are means ± SD across five spatial block cross-validation folds (block size: 200 km). Ensemble weight: TSS-based weighting used in weighted consensus prediction (sum = 1.000).

| Algorithm | AUC (Mean ± SD) | TSS (Mean ± SD) | CBI (Mean ± SD) | Ensemble Weight |
|---|---|---|---|---|
| MaxEnt | 0.942 ± 0.018 | 0.831 ± 0.029 | 0.901 ± 0.037 | 0.312 |
| Random Forest (RF) | 0.951 ± 0.016 | 0.847 ± 0.024 | 0.913 ± 0.031 | 0.328 |
| GBM/BRT | 0.921 ± 0.024 | 0.798 ± 0.038 | 0.861 ± 0.044 | 0.213 |
| GLM | 0.882 ± 0.033 | 0.743 ± 0.051 | 0.831 ± 0.059 | 0.147 |
| Ensemble (TSS-weighted) | 0.934 ± 0.021 | 0.812 ± 0.033 | 0.883 ± 0.041 | — |

AUC = Area Under the ROC Curve; TSS = True Skill Statistic; CBI = Continuous Boyce Index.

Variable importance analysis identified SST_mean as the single most influential predictor (mean permutation importance: 32.4 ± 4.1%), followed by substrate type (21.7 ± 3.8%), chlorophyll-*a* (14.3 ± 2.9%), dissolved oxygen (9.8 ± 2.2%), and distance to nearest port (8.1 ± 1.9%). Depth (5.6 ± 1.4%), SST_min (4.7 ± 1.3%), salinity (2.1 ± 0.8%), and AIS vessel density (1.3 ± 0.5%) completed the ranking. Together, the two anthropogenic predictors accounted for 9.4% of total variable importance, comparable to dissolved oxygen (9.8%) and substantially above salinity (Figure 3). The SDM response curve for SST_mean peaked between 21 and 26 °C and declined sharply below 16 °C.

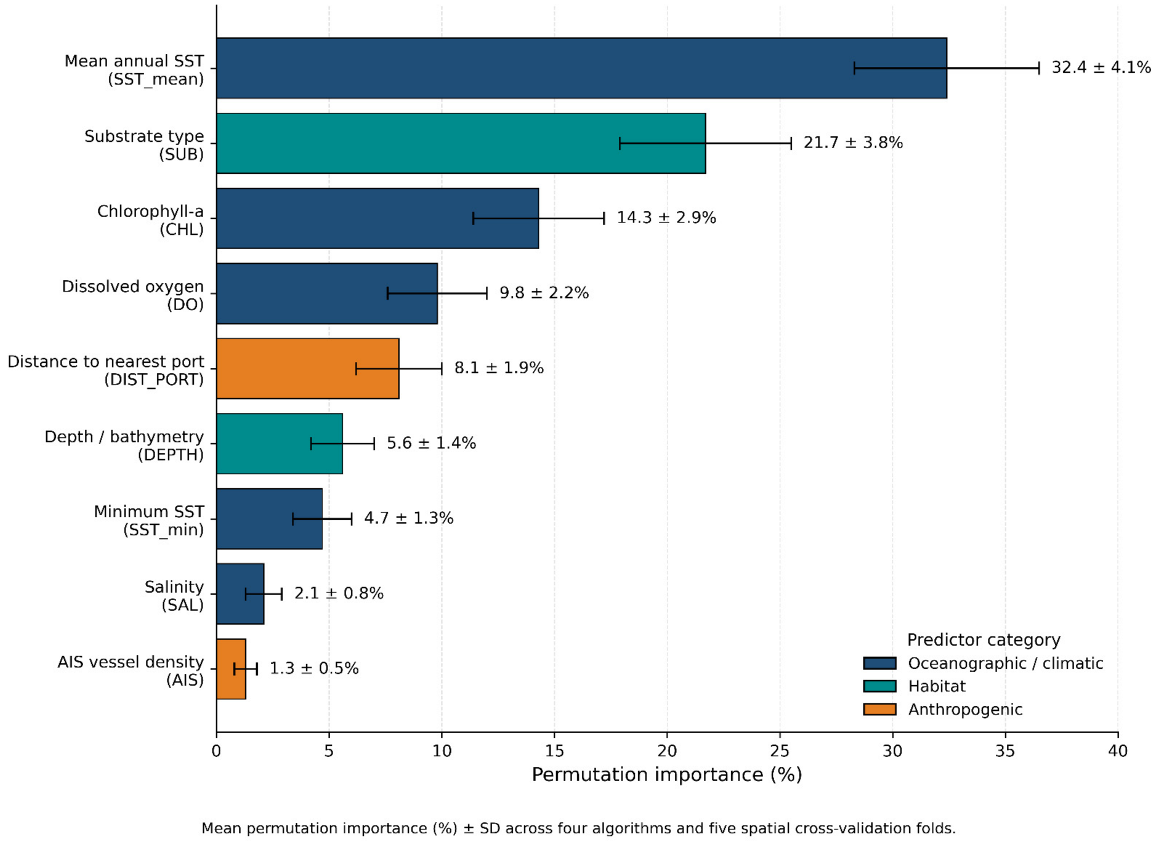


**Figure 3.** Variable importance ranking for the ensemble SDM of *P. radiata* across the Mediterranean Sea. Bars show mean permutation importance (%) ± SD across four algorithms and five spatial cross-validation folds. Colours indicate predictor category: dark blue = oceanographic/climatic; teal = habitat; orange = anthropogenic. Predictors are ranked in descending order of mean importance.

### *3.4. Predicted Habitat Suitability Across the Mediterranean*

The ensemble SDM predicted a spatially heterogeneous habitat suitability surface across the Mediterranean basin (Figure 4). Cells with Very High suitability (≥0.75) covered approximately 8.3% of Mediterranean coastal waters (≤150 m depth) and were concentrated in: (i) the southeastern Levantine coast (Israeli, Lebanese and Turkish coasts; SST_mean 22–26 °C); (ii) the eastern and central Aegean Sea (Greek island arcs; SST_mean 20–24 °C); (iii) the Gulf of Gabes (Tunisia) and Sicilian Channel (SST_mean 21–25 °C); (iv) specific coastal sectors of Libya and Egypt; and (v) the Alboran Sea (southern Spain; SST_mean 19–22 °C).

Moderate suitability (0.45–0.75) was predicted across much of the Tyrrhenian coast of Italy, the northern Adriatic Sea, the Ionian coast of Greece and Italy, and the western Mediterranean coastal sectors. Several areas of the northern Adriatic (Croatia, Slovenia), the Ligurian Sea, and the Gulf of Lion displayed moderate suitability (0.35–0.60) despite currently sparse documented occurrence records, flagging these areas as potential spread fronts (see also Section 3.5). Low suitability (<0.45) predominated in deeper offshore areas, high-latitude sectors (northern Adriatic, northern Aegean in winter months), and areas dominated by soft, muddy sediments lacking hard-bottom attachment sites. Ensemble uncertainty was low throughout (mean CV = 0.08 ± 0.03 in high-suitability zones).

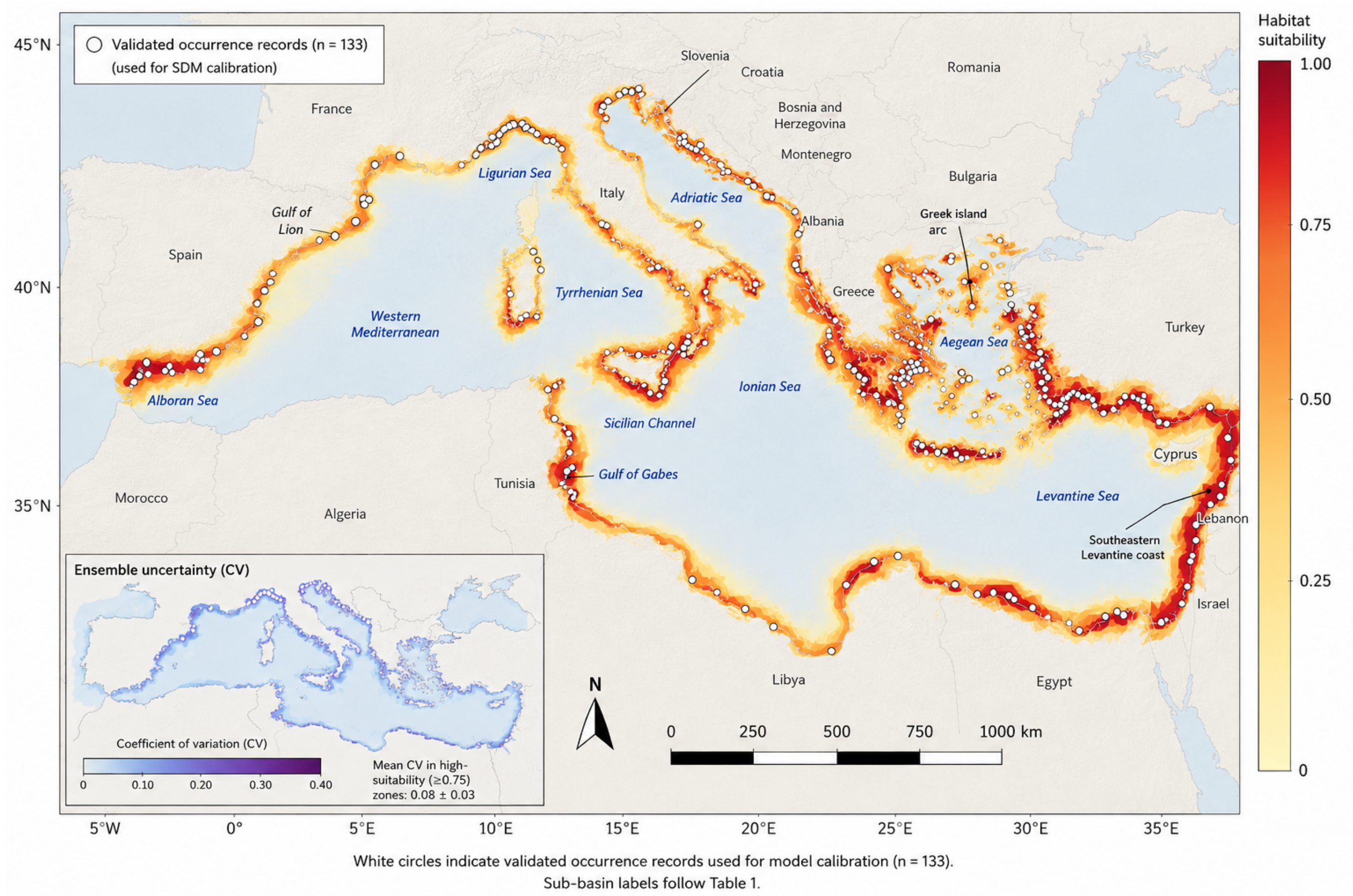


**Figure 4.** Evidence-informed habitat suitability surface for *P. radiata* across the Mediterranean Sea.

### *3.5. Primary Occurrence-Based Analysis and Secondary Decomposition of SDM-Derived Suitability*

The primary occurrence-based logistic model showed that recorded occurrence status was positively associated with the Environmental Conditions Index (β = 1.189, odds ratio = 3.28, $p < 0.001$) and the Anthropogenic Pressure Index (β = 0.650, odds ratio = 1.92, $p < 0.001$), and negatively associated with depth (β = −0.038, odds ratio = 0.963, $p < 0.001$; Table 4). McFadden's pseudo-$R^2$ was 0.091, indicating modest explanatory performance. Because the response consisted of occupied and available-background cells, these coefficients represent statistical associations with recorded occurrence rather than causal effects or estimates of true occupancy. As a secondary exploratory analysis, the SEM was used to describe the internal association structure of the SDM-derived suitability surface.

The SEM explained 58.4% of the variance in the SDM-derived habitat suitability surface and showed acceptable model fit ($\chi^2$/df = 1.74, CFI = 0.971, RMSEA = 0.043, 90% CI: 0.021–0.063, SRMR = 0.052; Table 4). Because the SEM response variable was generated from the ensemble SDM, these results are interpreted as a pathway-based decomposition of the predicted suitability surface rather than as an independent causal model of raw *P. radiata* occurrence.

ES was positively associated with SDM-derived suitability (β = 0.447, SE = 0.048, 95% CI: 0.353–0.541, $p < 0.001$), and AP was also positively associated with SDM-derived suitability (β = 0.408, SE = 0.055, 95% CI: 0.300–0.516, $p < 0.001$). ES was positively associated with AP (β = 0.438, SE = 0.061, 95% CI: 0.318–0.558, $p < 0.001$), indicating spatial co-occurrence between environmentally favourable coastal sectors and anthropogenic exposure proxies such as port proximity and maritime traffic intensity.

**Table 4.** Associations between environmental suitability, anthropogenic pressure, depth, and the occurrence status of *P. radiata* across Mediterranean coastal grid cells.

| Predictor/Model Component | β | Odds Ratio (expβ) | *p*-Value | Interpretation |
|---|---|---|---|---|
| Environmental Conditions Index | 1.189 | 3.28 | <0.001 *** | Positive association with recorded occurrence |
| Anthropogenic Pressure Index | 0.650 | 1.92 | <0.001 *** | Positive association with recorded occurrence |
| Depth | −0.038 | 0.963 | <0.001 *** | Negative association with recorded occurrence |
| Intercept | −0.162 | 0.85 | 0.435 | Not statistically significant |
| Model performance | | | | |
| McFadden's pseudo-$R^2$ | 0.091 | — | — | Indicates modest explanatory performance |

*** $p < 0.001$. Odds ratios represent the change in the odds of recorded occurrence associated with a one-unit increase in each standardised predictor. The model is based on presence–background data; therefore, coefficients describe statistical associations with recorded occurrence rather than estimates of true occupancy or causal effects.

Occurrence was coded as 1 for occupied coastal cells and 0 for background cells. Positive coefficients indicate increasing occurrence probability with increasing predictor values. Coefficients are unstandardised logit-scale estimates unless the original model output explicitly identifies them as standardised coefficients. The reported relationships represent statistical associations and not causal effects. The structural associations among Environmental Suitability, Anthropogenic Pressure, and SDM-derived habitat suitability are illustrated in Figure 5.

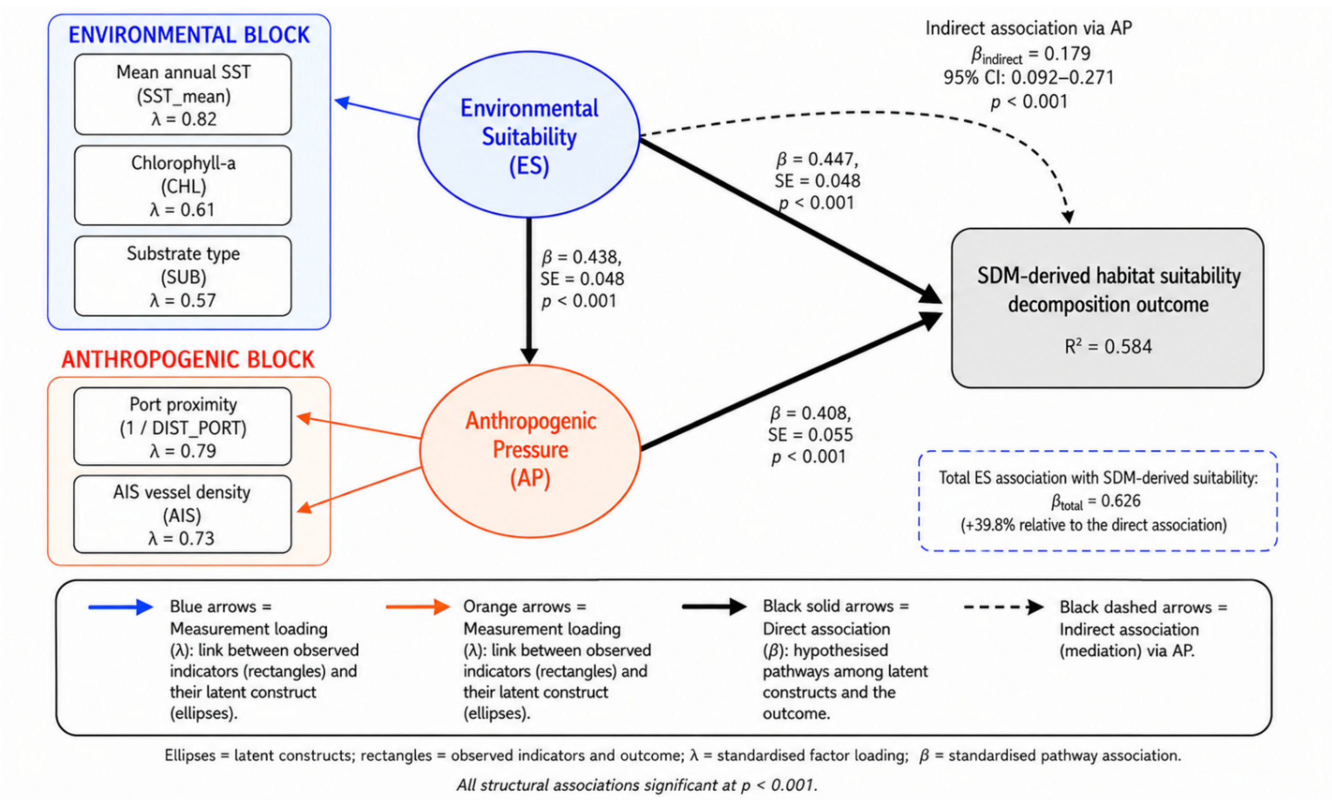


**Figure 5.** Structural association model linking Environmental Suitability (ES), Anthropogenic Pressure (AP), and SDM-derived habitat suitability of *P. radiata*. Values indicate standardised pathway associations. The indirect association is represented as ES → AP → SDM-derived habitat suitability. Because the response variable is derived from the ensemble SDM, the model is interpreted as a decomposition of the SDM-derived suitability surface rather than as an independent causal test of raw occurrence probability. All structural associations were significant at $p < 0.001$.

The indirect pathway ES → AP → SDM-derived suitability was significant (β = 0.179, SE = 0.047, 95% CI: 0.092–0.271, $p < 0.001$). The total ES association with SDM-derived suitability was β = 0.626. These pathways indicate that environmentally favourable coastal cells, anthropogenic exposure proxies and high SDM-derived suitability are spatially structured in a consistent way. However, this pattern should not be interpreted as evidence of a compounding environmental–anthropogenic invasion mechanism, because the SEM outcome is itself derived from the SDM surface.

These findings are consistent with the hypothesised relationships between environmental suitability, anthropogenic pressure, and the SDM-derived habitat suitability surface. However, they should be interpreted only at the level of modelled habitat suitability and spatial prioritisation. The SEM does not provide independent evidence that Environmental Suitability (ES) and Anthropogenic Pressure (AP) causally determine raw occurrence probability. Rather, it quantifies the structured associations among environmental indicators, anthropogenic proxies, and the SDM-derived habitat suitability surface.

### 3.6. Mediterranean Invasion Risk Index and Priority Monitoring Areas

The composite Mediterranean IRI identified 31.8% of coastal grid cells as High or Very High priority cells (Table 5; Figure 6). Very High priority values (IRI > 0.75; 10.5% of all cells) were most extensive in the southeastern Levantine coast, the Aegean Sea, and the Gulf of Gabes–Sicilian Channel system. These sectors combine high modelled suitability, SST_mean >21 °C, suitable hard/rocky substrate, and elevated maritime exposure, including major port infrastructure and AIS vessel density. This pattern indicates spatial overlap between environmental suitability and anthropogenic exposure, consistent with the SEM decomposition, but should not be interpreted as evidence of a causal mechanism of invasion.

**Table 5.** Mediterranean Invasion Risk Index (IRI) by sub-basin: percentage of coastal grid cells (0–150 m depth) per IRI priority category and priority spread-front classification. IRI categories are interpreted as spatial monitoring priorities rather than as absolute probabilities of establishment or future invasion.

| Sub-Basin | Very High IRI (%) | High IRI (%) | Moderate IRI (%) | Low IRI (%) | Priority Spread Front? |
|---|---|---|---|---|---|
| Eastern Levantine Sea | 38.4 | 21.7 | 28.1 | 11.8 | No established core range |
| Aegean Sea | 22.1 | 24.3 | 32.6 | 21.0 | Partial (northern Aegean) |
| Gulf of Gabes/Sicilian Channel | 14.7 | 23.1 | 34.8 | 27.4 | No established |
| Ionian Sea | 11.2 | 19.8 | 38.4 | 30.6 | Partial (western Ionian) |
| Alboran Sea (Spain/Morocco) | 8.7 | 18.4 | 37.6 | 35.3 | Partial |
| Western Mediterranean | 5.2 | 14.1 | 40.3 | 40.4 | YES—spread fronts #3, #4, #5 |
| Adriatic Sea | 4.3 | 16.7 | 41.2 | 37.8 | YES—spread front #1 |
| Tyrrhenian/Ligurian seas | 3.9 | 13.8 | 42.4 | 39.9 | YES—spread fronts #2, #4 |
| TOTAL Mediterranean | 10.5 | 21.3 | 34.2 | 34.0 | — |

Priority monitoring areas were defined at the grid-cell level as High or Very High IRI cells with current occurrence density <1 validated record per 0.25° cell. Six broader priority spread-front sectors spanning approximately 7400 km of the Mediterranean coastline were then delineated from clusters of these priority cells. Because sector summaries include neighbouring Moderate-priority cells as well as High/Very High cells, some reported sector-level IRI ranges extend below the 0.50 High-priority threshold. The priority designation, therefore, refers to the High/Very High cells embedded within each sector, not to every cell in the sector.

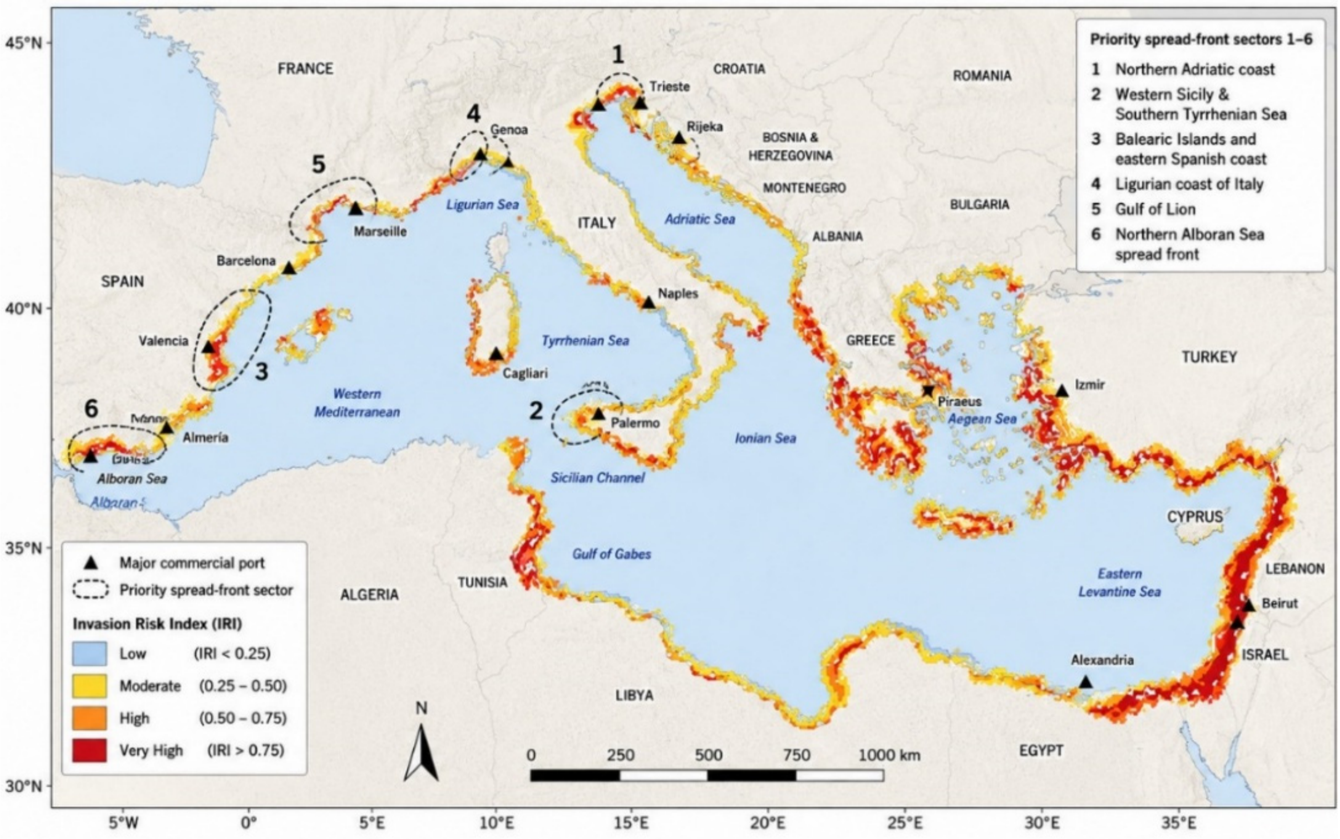


**Figure 6.** Mediterranean Invasion Risk Index (IRI) map for *P. radiata* at 0.05° resolution. Risk categories were classified using fixed interpretive thresholds: Low (light blue, IRI < 0.25), Moderate (yellow, 0.25–0.50), High (orange, 0.50–0.75), and Very High (dark red, IRI > 0.75). Priority spread-front sectors are outlined with dashed borders and labelled 1–6 (see Section 3.6 for full descriptions). These sectors were delineated from clusters of High and Very High IRI cells with low current occurrence density, although the broader sector boundaries may also include neighbouring Moderate-risk cells. Major commercial ports within priority sectors are shown as black triangles. Cells outside the 0–150 m coastal zone are shown in white.

### *3.7. Sentiment Analysis: Public and Stakeholder Perception*

The multilingual sentiment corpus comprised 1247 text units: iNaturalist observation comments ($n = 214$, 17.2%), news and media articles ($n = 87$, 7.0%), and stakeholder open-ended responses from Pafras et al. [26] ($n = 946$, 75.9%). The stakeholder-derived component originated from 80 participants and was segmented into sentence-level or thematically coherent text units before sentiment scoring. Because multiple text units could originate from the same respondent, corpus-level polarity distributions are interpreted as descriptive indicators of discourse composition rather than as statistically independent observations.

Sentiment classification using the multilingual VADER extension [40] indicated a predominantly neutral-to-positive discourse surrounding *P. radiata* in Mediterranean contexts. Overall sentiment distribution was as follows: Very Positive 8.2% ($n = 102$), Positive 30.2% ($n = 377$), Neutral 41.2% ($n = 514$), Negative 17.3% ($n = 216$), and Very Negative 3.1% ($n = 38$) (Figure 7A). Taken together, neutral and positive categories accounted for 79.6% of the corpus, whereas negative and very negative categories accounted for 20.4%. These descriptive results indicate that Mediterranean discourse surrounding *P. radiata* is not dominated by negative invasion framing but is instead distributed across neutral, positive, and negative interpretations.

The expert-assisted review of the multilingual VADER adaptation identified several recurrent sources of ambiguity, including negation, fisheries-specific idioms, commercial-use terminology, and terms whose polarity depended on whether *P. radiata* was framed as a non-indigenous organism, a food item, or a resource. These cases were used to refine the trilingual lexicon before final classification. This step was designed as a lexicon calibration and qualitative reliability screening rather than as a balanced classifier-performance experiment. The final adapted VADER classifications agreed with the human-consensus polarity labels for 85.0% of the validation text units. This value is reported as a basic post-calibration agreement statistic. Because the validation subset was designed primarily for lexicon refinement rather than as a balanced classifier-performance experiment, language-specific accuracy, macro-F1, precision and recall were not calculated. Accordingly, corpus-level sentiment distributions are interpreted descriptively, while inferential testing is restricted to participant-level stakeholder scores.

Thematic analysis following Braun and Clarke [41] identified five dominant discourse themes. The first theme, Economic opportunity, included references to commercial harvesting, culinary value, pearl-oyster use, and market potential, and accounted for 38.1% of positive sentiment text units. The second theme, Ecological concern, included references to biodiversity impacts, competition with native species, habitat alteration, and ecosystem disruption, and accounted for 62.4% of negative sentiment text units. The third theme, Regulatory ambivalence, included references to unclear legislation, inconsistent national management frameworks, and policy gaps across Mediterranean countries, and accounted for 28.7% of neutral text units. The fourth theme, Naturalisation narrative, framed *P. radiata* as an established or familiar Mediterranean organism rather than as a recent invader, and it appeared in 19.3% of all text units. The fifth theme, Citizen-science engagement, reflected positive reporting language, aesthetic appreciation, and curiosity in iNaturalist comments, representing 48.2% of iNaturalist-derived comments.

Inter-rater reliability for thematic coding was substantial, with Cohen's $\kappa = 0.78$, indicating acceptable agreement among coders. Theme percentages are reported descriptively because text units were unevenly distributed among sources and, in the stakeholder component, multiple units could originate from the same participant. The strong representation of stakeholder-derived text units from Pafras et al. [26] should therefore be considered when interpreting the overall sentiment distribution.

Comparison across the three cognitive stakeholder profiles identified by Pafras et al. [26] was conducted using participant-level mean compound sentiment scores rather than text-unit-level scores. Mean sentiment differed significantly among profiles (one-way ANOVA: $F(2, 77) = 14.32$, $p < 0.001$) (Figure 7B). Ecological Pragmatists showed the most positive mean sentiment (compound = +0.31 ± 0.18), Institutional Collaborators were closer to neutral (compound = +0.06 ± 0.14), and Adaptive Stewards showed the most cautious or negative sentiment (compound = −0.08 ± 0.21). Tukey's HSD post-hoc comparisons indicated significant pairwise differences among all three groups ($p < 0.05$). Because the ANOVA was performed at the participant level, the degrees of freedom reflect N = 80 stakeholders rather than the full number of segmented text units, thereby avoiding pseudo-replication arising from multiple responses per participant.

A summary of the human-coded validation procedure, its lexicon-refinement role, and its implications for interpretation is provided in Table 6.

**Table 6.** Human-coded validation summary for the multilingual VADER adaptation. The table summarises the validation procedure used to check the trilingual VADER lexicon across Greek, Spanish, and Italian text units and across the three corpus sources: stakeholder responses, citizen-science comments and media text. The validation was used primarily for reliability checking and lexicon calibration rather than as a standalone balanced classifier-performance experiment; therefore, only a basic overall post-calibration VADER–human agreement statistic is reported.

| Validation Component | Implementation in This Study | Interpretive Role |
| --- | --- | --- |
| Purpose | Reliability checking and lexicon calibration of the trilingual VADER extension. | Supports exploratory multilingual sentiment classification. |
| Corpus coverage | Validation material was drawn from Greek, Spanish, and Italian text units across stakeholder responses, citizen-science comments, and media text. | Ensures coverage of the main languages and source types analysed. |

| | | |
|---|---|---|
| Human coding | Three independent coders reviewed validation material using the same five-category polarity scheme as VADER: Very Positive, Positive, Neutral, Negative, and Very Negative. | Allows comparison between expert interpretation and VADER-derived categories. |
| Disagreement review | Disagreements were reviewed jointly by the coders. | Identified ambiguous or context-dependent sentiment cases. |
| Lexicon refinement | Refinements focused on negation handling, fisheries-specific idioms, invasion-governance terminology, and ecological/commercial terms with context-dependent meaning. | Improved the trilingual lexicon for Mediterranean invasion discourse. |
| Quantitative agreement | Overall and language-specific VADER–human agreement percentages were not reported as standalone classifier-performance metrics. | Validation was used for reliability checking and lexicon calibration rather than formal model benchmarking. |
| Manuscript use | Validation supported cautious descriptive interpretation of corpus-level sentiment distributions. | Inferential testing was restricted to participant-level stakeholder scores. |

Overall, the sentiment results suggest that *P. radiata* occupies an ambiguous socio-ecological position in Mediterranean discourse. It is recognised as a non-indigenous and potentially invasive bivalve, but it is also framed by many stakeholders and observers as a commercially useful, edible, or already naturalised resource. This dual framing helps explain the predominance of neutral-to-positive sentiment and highlights the need for risk communication strategies that acknowledge both ecological concerns and perceived socio-economic opportunities.

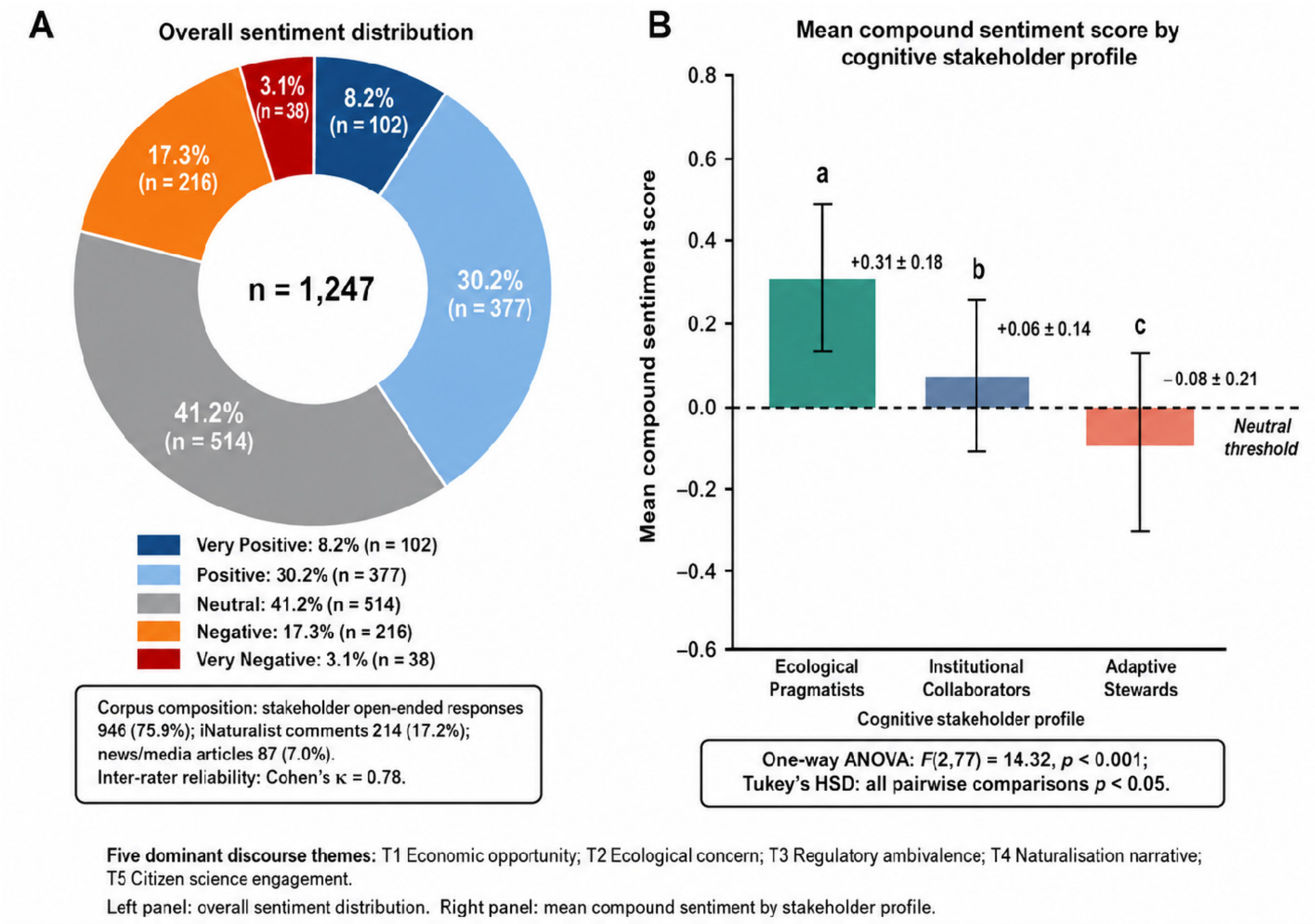


**Figure 7.** Sentiment analysis results for Mediterranean public and stakeholder discourse on *P. radiata* ($n$ = 1247 multilingual text units). Panel (**A**) shows the overall sentiment distribution across five polarity categories: Very Negative, Negative, Neutral, Positive, and Very Positive. Panel (**B**) shows the mean compound sentiment score (± SD) by cognitive stakeholder profile: Ecological Pragmatists, Institutional Collaborators, and Adaptive Stewards. Letters above bars indicate statistically distinct

groups based on Tukey's HSD post-hoc tests (α = 0.05). The dashed horizontal line at compound = 0 indicates the neutral sentiment threshold.

## 4. Discussion

### *4.1. Environmental Drivers: SST, Substrate, and the Mediterranean Thermal Envelope*

The SDM indicates that the distribution of *P. radiata* across the Mediterranean is strongly structured by interactions among thermal conditions, settlement habitat, and nearshore productivity. Rather than reflecting temperature alone, the predicted suitability pattern suggests that successful establishment depends on the co-occurrence of warm shallow waters, suitable attachment surfaces, and sufficient food availability for a suspension-feeding bivalve.

The prominence of sea-surface temperature is ecologically consistent with the Indo-Pacific and Red Sea origins of *P. radiata*, its affinity for warm coastal environments, and its reproductive performance in thermally favourable waters. Progressive Mediterranean warming may therefore increase the extent of coastal habitat falling within the species' realised thermal niche, particularly in marginal northern and western sectors. However, temperature should not be interpreted as a single deterministic driver. Occurrence patterns also depend on propagule supply, substrate availability, coastal retention, observer effort, and local ecological context.

The predicted high-suitability sectors were broadly consistent with documented Mediterranean records of *P. radiata*, including the Levantine, Aegean, Ionian, Sicilian Channel, and western Mediterranean sectors [10–15]. This agreement supports the ecological plausibility of the basin-scale suitability pattern, while the presence-only and citizen-science dominated nature of the dataset requires cautious interpretation of unsampled or poorly sampled sectors.

Substrate availability provides an important ecological constraint on this thermal signal. As a byssate bivalve, *P. radiata* requires hard or structurally complex surfaces for settlement, including rocky reefs, biogenic substrates, and artificial structures such as port pilings, aquaculture infrastructure, and vessel-associated surfaces [11,12,15]. This helps explain why high suitability is concentrated in shallow coastal sectors where hard-bottom habitat and artificial structures are common, whereas deeper offshore areas and soft, muddy sediments show lower suitability. Current velocity was excluded on empirical collinearity grounds ($|r| = 0.81$ with chlorophyll-*a*) rather than on ecological ones. We acknowledge that near-bed flow may be mechanistically more proximate for a byssate, suspension-feeding bivalve with planktonic larvae, influencing food supply, larval retention, and settlement; chlorophyll-*a* was retained as the more conventional and directly measured predictor, but this trade-off should be revisited where independent hydrodynamic layers are available.

Chlorophyll-*a* and dissolved oxygen further suggest that coastal productivity and water-column conditions mediate habitat suitability. Productive nearshore waters may support feeding and growth, while low-oxygen or highly stratified environments may constrain establishment in some semi-enclosed basins. These patterns should be interpreted as convergent ecological evidence rather than direct proof of temperature-driven causality, because the occurrence dataset also reflects uneven sampling effort and anthropogenic dispersal pathways.

### *4.2. Recent Geographic Redistribution and Observation Effort*

The occurrence dataset showed a pronounced temporal concentration after 2020 and a statistically supported westward redistribution of available reports. More than 81% of dated records originated from 2020–2026, while mean annual reporting was almost nine times higher during 2020–2026 than during 2010–2019. The negative association between observation year and longitude persisted after spatial thinning and geographic quality assurance, and median record longitude shifted approximately 6.6°

westward between periods. Changes in sub-basin composition further indicated increasing representation of western and northern Mediterranean sectors.

These patterns are consistent with the continued geographic redistribution of *P. radiata*, including recent reports from the western Mediterranean. Nevertheless, the dataset was dominated by human observations, which represented approximately 94% of all records. The observed temporal increase may therefore reflect both biological redistribution and substantial growth in citizen-science participation, digital platform use, and national data mobilisation. The analyses should not be interpreted as an estimate of invasion velocity or population growth.

This distinction is important for the interpretation of the SDM and the monitoring-priority framework. Recent isolated reports in western or northern sectors warrant targeted validation surveys, but the absence of records in other sectors cannot be treated as evidence of biological absence. Repeated, effort-standardised surveys at ports, marinas, aquaculture installations, and suitable natural substrates are needed to distinguish persistent establishment from reporting artefacts.

### *4.3. Anthropogenic Propagule Pressure as an Important Driver of Recorded Occurrence*

The primary occurrence-based logistic model identified Anthropogenic Pressure as a significant positive predictor of recorded occurrence status for *P. radiata* (β = 0.650, odds ratio = 1.92, $p < 0.001$; Table 4). Accordingly, a one-unit increase in the standardised Anthropogenic Pressure Index was associated with approximately 92% higher odds of a coastal cell containing a recorded occurrence, after accounting for environmental conditions and depth. The observed Mediterranean distribution of *P. radiata* was not associated with environmental conditions alone. Greater port proximity and higher vessel-traffic intensity were positively associated with the relative probability of recorded occurrence across available Mediterranean coastal cells. However, because the analysis was based on presence–background data and spatially uneven observation effort, these relationships should be interpreted as statistical spatial associations rather than as direct evidence of causation or true occupancy.

This interpretation is consistent with broader evidence that commercial ports, shipping corridors, and areas of intense vessel activity frequently function as introduction and secondary-dispersal hubs for marine non-indigenous species [43–46]. For *P. radiata*, the anthropogenic signal is particularly relevant because its Mediterranean invasion history combines primary Lessepsian introduction through the Suez Canal with subsequent redistribution along both natural and human-mediated pathways. Maritime transport, including hull biofouling and ballast-water-mediated dispersal, may facilitate the movement of adults, juveniles, or associated propagules among ports, marinas, and other artificial coastal habitats [11,13]. The positive association between anthropogenic pressure and recorded occurrence, therefore, supports the view that transport connectivity and maritime infrastructure contribute to the spatial organisation of the documented distribution, particularly in coastal sectors that are already environmentally favourable.

The occurrence-based results also strengthen the management relevance of propagule-pressure reduction. Although the Environmental Conditions Index showed a stronger association with recorded occurrence than the Anthropogenic Pressure Index, environmental suitability is generally less amenable to direct local intervention. By contrast, several components of anthropogenic exposure can be modified through regulation, surveillance, and operational management. Biosecurity measures should therefore include effective implementation of the IMO Ballast Water Management Convention, port-based hull and niche-area inspections, biofouling management procedures, and targeted surveillance of high-traffic commercial ports, marinas, and aquaculture facilities. These actions are unlikely to eliminate the risk of invasion completely, but they may reduce opportunities for repeated introduction and secondary spread in environmentally suitable coastal sectors.

Importantly, the Anthropogenic Pressure Index represents a harmonised basin-scale proxy based only on port proximity and AIS vessel density. It does not fully capture other potentially relevant pathways, including recreational boating, marina connectivity, shellfish translocations, aquaculture transfers, local seafood trade, hull-cleaning practices, or spatial variation in ballast-water discharge. The estimated association should therefore not be interpreted as a complete representation of propagule pressure. Moreover, areas near ports and densely populated coastlines may also receive greater observer attention, potentially strengthening the apparent relationship between maritime activity and recorded occurrence. Although spatial thinning reduced the influence of clustered observations, it could not fully correct for uneven detection and reporting effort.

Nevertheless, the occurrence-based findings provide a defensible basis for integrating anthropogenic exposure into Mediterranean surveillance planning. Coastal areas in which favourable environmental conditions coincide with intense maritime activity represent appropriate priority nodes for early detection, repeated monitoring and biosecurity intervention. This interpretation is also consistent with the Invasion Risk Index, which prioritises areas where elevated SDM-derived environmental suitability overlaps with high anthropogenic pressure. The occurrence-based logistic model and the IRI should therefore be viewed as complementary components of the risk-screening framework: the former demonstrates that anthropogenic exposure is significantly associated with recorded occurrence, whereas the latter translates the spatial overlap between suitability and exposure into operational surveillance priorities.

### *4.4. Decomposing the Environmental–Anthropogenic Structure of Modelled Risk*

The value of the secondary SEM is that it describes the spatial association structure among environmental indicators, anthropogenic-exposure proxies, and the SDM-derived suitability surface. While the SDM identifies coastal sectors with relatively favourable modelled environmental conditions, the SEM provides an exploratory decomposition of how environmental suitability and anthropogenic exposure co-vary spatially with the modelled surface. Because the SEM outcome was itself derived from the SDM, these associations do not constitute independent evidence of ecological or anthropogenic causation.

The resulting spatial pattern should therefore be interpreted as an exploratory description of model structure rather than as evidence of a compounding causal invasion mechanism.

### *4.5. Sentiment Analysis and the Social Ecology of Mediterranean Invasion*

The sentiment analysis suggests that Mediterranean discourse surrounding *P. radiata* is more ambivalent than a simple invasive-species framing would imply. Neutral-to-positive sentiment appears to reflect the species' dual status as both a non-indigenous bivalve and a locally recognised or potentially useful resource, consistent with previous evidence that public and stakeholder perceptions can influence environmental governance, monitoring uptake, and management legitimacy [19,20]. This interpretation is also consistent with the stakeholder evidence of Pafras et al. [26], in which different cognitive profiles framed non-indigenous species management through distinct combinations of ecological concern, institutional trust, and adaptive resource use.

The thematic results further suggest that economic opportunity, naturalisation narratives, and citizen-science engagement co-occur with ecological concern. This pattern is important because marine invaders that are perceived as useful, edible or already naturalised may not generate the same level of public concern as species framed exclusively as ecological threats. However, these findings should be interpreted cautiously. The sentiment corpus was unevenly distributed among sources and was strongly influenced by stakeholder responses from Pafras et al. [26]. In addition, the multilingual VADER extension [40] provides an exploratory rule-based classification rather than a fully validated cross-cultural public-opinion measure, while the thematic coding followed Braun and Clarke's [41] qualitative framework and should be interpreted as descriptive rather than representative of all Mediterranean publics.

Nevertheless, the sentiment results may inform future research on communication and stakeholder engagement. The coexistence of ecological concern and economic-opportunity framing suggests that risk communication may be more effective if it acknowledges both dimensions rather than presenting the species only as a threat. For example, communication with fishers, citizen scientists, and local coastal communities could distinguish among monitoring needs, ecological uncertainty and potential resource use, while avoiding messages that unintentionally normalise the further spread.

Future work should test these communication implications directly through balanced multilingual surveys, interviews, and experimental message-framing studies across different Mediterranean regions and stakeholder groups. Such work would help determine whether the cognitive profiles identified in the stakeholder dataset [26] correspond to broader patterns of public perception and whether tailored communication strategies improve monitoring participation, reporting behaviour and policy legitimacy [19,20].

The predominantly neutral-to-positive sentiment pattern observed here is consistent with broader invasion-perception literature showing that public and stakeholder responses to non-indigenous species are not determined solely by ecological impact, but also by perceived usefulness, familiarity, livelihood relevance, cultural framing, and institutional trust [47–50]. Systematic reviews and conceptual frameworks on invasive-species perception show that perceived benefits, socio-cultural context, terminology, and stakeholder identity strongly influence whether non-indigenous species are framed as threats, resources, or ambiguous social-ecological entities [47–49]. Conservation culturomics and digital public-discourse approaches further demonstrate that online texts, news media, social media, and citizen-science platforms can be used to quantify public salience, sentiment, and human–nature interactions at broad spatial and temporal scales [51–53]. For *P. radiata*, its edibility, commercial value, and long residence time in parts of the Mediterranean may therefore contribute to a less negative public framing than would be expected under a strictly threat-based invasive-species narrative. This interpretation does not reduce the ecological relevance of the species as a non-indigenous bivalve; rather, it highlights that biosecurity communication should acknowledge both ecological risk and locally perceived socio-economic value.

A methodological caveat is that the multilingual VADER adaptation was calibrated through expert-assisted review rather than validated against a balanced gold-standard benchmark dataset. Consequently, the sentiment results should be interpreted as exploratory indicators of discourse polarity and framing, not as definitive supervised classification outputs. This limitation is partly mitigated by the use of stratified multilingual review, lexicon refinement, and participant-level aggregation for stakeholder comparisons, but future work should evaluate Mediterranean invasion-related sentiment models against larger manually annotated multilingual corpora.

### *4.6. Management Implications for Mediterranean Biosecurity*

The integrated SDM–SEM–IRI–sentiment framework supports four complementary management applications for *P. radiata* across the Mediterranean. First, the IRI map provides a spatial prioritisation layer for early detection by identifying six priority spread-front sectors: the northern Adriatic, western Sicily, the Balearic Islands and the adjacent Spanish coast, the Ligurian coast, the Gulf of Lion, and the Alboran Sea. These sectors combine elevated invasion risk with incomplete current documentation. The interpretation of northern Mediterranean spread-fronts is further supported by earlier evidence of *P. radiata* establishment in the eastern Adriatic [46], together with the more recent records included in the curated data set. They are therefore suitable targets for standardised settlement-panel monitoring, eDNA metabarcoding, and repeated port-based surveys. Rather than distributing surveillance evenly across the basin, monitoring effort should be concentrated where high environmental suitability, high propagule pressure, and low current record density overlap.

Second, the positive association between Anthropogenic Pressure and recorded occurrence in the primary logistic model strengthens the rationale for port-centred biosecurity. A one-unit increase in the standardised Anthropogenic Pressure Index was associated with approximately 92% higher odds of recorded occurrence, after accounting for environmental conditions and depth. Because propagule-pressure proxies are more directly manageable than basin-scale environmental conditions, high-priority ports and adjacent marinas should be considered for ballast-water compliance checks, hull-fouling inspection, biofouling management, and rapid-response protocols. This is particularly relevant for major commercial hubs located within or near high-priority IRI zones, including Piraeus, Haifa, Alexandria, Tunis/La Goulette, Taranto, Catania, and Barcelona.

Third, the sentiment results indicate that biosecurity communication should not rely exclusively on threat-based messaging. Public and stakeholder discourse around *P. radiata* is shaped by ecological concern, economic opportunity, and naturalisation narratives. Consequently, communication strategies should be differentiated by audience: fishers and commercial users may respond better to messages linking utilisation with containment, reporting, and regulated handling; port authorities may require clear surveillance and compliance protocols; and conservation-oriented groups may be mobilised through early-warning and citizen-science narratives. This approach is more consistent with the species' dual status as both a non-indigenous organism and a locally recognised resource.

Fourth, the framework is directly relevant to Mediterranean and European non-indigenous species governance. The IRI map can support spatial prioritisation under the Barcelona Convention Post-2020 SAPBIO and the UNEP/MAP 2022–2027 Medium-Term Strategy, while the integration of occurrence risk, anthropogenic vectors, and monitoring priorities aligns with the objectives of the EU Marine Strategy Framework Directive Descriptor 2. The practical value of the framework lies in translating ecological modelling, causal pathway analysis, and public perception into targeted surveillance, risk communication, and biosecurity planning.

### *4.7. Limitations and Future Directions*

Several limitations should be considered when interpreting these findings. First, although the occurrence dataset represents the most spatially extensive synthesis currently available for Mediterranean *P. radiata*, GBIF-derived records remain unevenly distributed across the basin. The eastern and southern Mediterranean shores, including Lebanon, Israel, Syria, Libya, Algeria, and Morocco, are likely under-represented because of lower data submission rates and uneven citizen-science coverage. This spatial bias may affect the precision of suitability estimates in regions where the species is present but poorly documented. Future work should therefore integrate additional sources, including OBIS, MedMIS/CIESM, national monitoring programmes and georeferenced literature records, to improve coverage across under-sampled sub-basins [51–54].

Second, the SDM was calibrated using contemporary environmental layers representing the 2000–2020 period. As a result, the present model estimates current habitat suitability rather than future invasion trajectories under continued Mediterranean warming. This is an important limitation because thermal suitability is likely to shift under future climate scenarios, particularly in marginal northern and western sectors. A priority extension would be to project the ensemble SDM under CMIP6 scenarios, including intermediate and high-emission pathways, using future Bio-ORACLE environmental layers. Such projections would allow identification of coastal areas where suitability is currently moderate but likely to increase under warming.

A further limitation concerns the temporal and spatial structure of the occurrence dataset. Approximately 94% of records originated from human observations, and more than 81% of dated records were collected after 2020. Although spatial thinning and geographic quality assurance reduced clustering and obvious coordinate errors, they could not fully account for variation in observer effort, platform adoption, or national data mobilisation. Accordingly, the observed westward redistribution of records should be interpreted as a change in the geographic composition of available reports rather than as a direct estimate of biological spread rate.

Third, the sentiment corpus is not evenly balanced across Mediterranean languages and stakeholder groups. Because stakeholder responses from the HGPS study form a large share of the corpus, Greek-language and eastern Mediterranean perspectives are more strongly represented than Spanish, Italian, Turkish, North African, and Levantine fishing communities. The sentiment results should therefore be interpreted as an initial multilingual social signal rather than a fully representative Mediterranean public-opinion survey. Expanding the corpus across additional countries, platforms, and stakeholder sectors would improve the robustness of cross-cultural comparisons.

The multilingual sentiment analysis should also be interpreted cautiously. Although the trilingual VADER extension was checked against human-coded validation material and refined after a review of disagreements, the validation was used primarily as a reliability and lexicon-calibration step rather than as a full classifier performance experiment. A basic overall post-calibration VADER–human agreement statistic of 85.0% was obtained, indicating a generally high correspondence between the adapted classifier and the human-consensus labels. However, because the validation subset was used primarily for reliability checking and lexicon refinement rather than as an independent balanced classifier benchmark, language-specific and source-specific agreement, macro-F1, precision and recall were not calculated. The corpus-level sentiment results should therefore remain interpreted as exploratory discourse indicators. Future work should evaluate the lexicon-based approach on a larger, more balanced validation set and compare its performance with supervised multilingual transformer models.

Fourth, the secondary SEM provides a static exploration representation of the spatial association structure among environmental indicators, anthropogenic-exposure proxies and SDM-derived habitat suitability. The current specification does not explicitly model non-linear responses, temporal lags, dispersal dynamics or feedback between establishment, propagule pressure, and subsequent spread. Future applications could use piecewise, spatial, or dynamic SEM frameworks with independent occurrence and temporal monitoring data to examine whether these statistical associations remain stable over time, across sub-basins, and across invasion stages. Such data would be required before stronger causal inference could be considered.

## 5. Conclusions

This study provides an integrated Mediterranean basin-scale framework for assessing the spatial invasion risk of the rayed pearl oyster *P. radiata* by combining ensemble species distribution modelling, structural modelling, an equally weighted Invasion Risk Index, and multilingual sentiment analysis. The framework identifies environmentally suitable coastal sectors, evaluates spatial associations between occurrence patterns, environmental conditions, and anthropogenic exposure, and incorporates an exploratory social-perception dimension.

The occurrence-based structural analysis indicated that the distribution of validated records was associated with both environmental conditions and anthropogenic exposure. Coastal areas characterised by favourable environmental conditions frequently coincided with port infrastructure and elevated maritime traffic, suggesting that surveillance priorities should account for both habitat suitability and proxies of propagule pressure. These associations should not be interpreted as direct evidence of causation, because the analysis was based on observational occurrence data and spatial proxy variables. The temporal and geographic audit also identified a marked post-2020 increase in reporting and a westward shift in the composition of available records. These patterns support targeted verification of recently reported western and northern sectors, while the predominance of opportunistic observations prevents their interpretation as an effort-standardised measure of biological expansion.

The equally weighted IRI integrated environmental suitability and anthropogenic exposure into a transparent spatial prioritisation tool, highlighting coastal sectors where early detection surveillance may be particularly valuable. The use of a priori equal weights avoided dependence on coefficients estimated by the structural model and improved the transparency and reproducibility of the prioritisation framework. The sentiment analysis further indicated that Mediterranean discourse concerning *P. radiata* was predominantly neutral to positive, reflecting its dual representation as both a non-indigenous bivalve and a potentially exploitable biological resource. Collectively, the ecological, pathway-related, and social-perception components provide a practical basis for future monitoring design, biosecurity planning, and risk communication. However, the identified priority areas should be interpreted as surveillance targets requiring field validation rather than as definitive predictions of establishment or spread.

## Supplementary Materials

The following supporting information can be found at: https://www.sciepublish.com/article/pii/1165, Figure S1: Record-selection and quality-control workflow for the Mediterranean occurrence dataset of *P. radiata*; Figure S2: Spatial predictor layers used in the SDM and SEM analyses.

## Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, the authors used Google GEMINI to improve the English language, grammar, and readability. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

## Acknowledgments

This research was supported by the research program of the Region of Central Greece (Project Code: Π89-2.13), implemented under the "Regional Development Programme (PPA) of Central Greece", Priority Axis 2.4 "Risk Prevention and Management", entitled "Investigation of the Biology, Fisheries, Ecology, and Population Dynamics of the bivalve pearl oyster *Pinctada imbricata radiata* in the South Evoikos and Maliakos Gulfs".

## Author Contributions

Conceptualization, D.P. and D.K.; Methodology, D.P.; Software, D.P.; Validation, D.P. and D.K.; Formal Analysis, D.P.; Investigation, D.P.; Resources, D.P. and D.K.; Data Curation, D.P.; Writing-Original Draft Preparation, D.P.; Writing-Review & Editing, D.P. and D.K., Visualization, D.P.; Supervision, D.K.; Project Administration, D.K.; Funding Acquisition, D.K. All authors have read and agreed to the published version of the manuscript.

## Ethics Statement

This study did not involve the collection of new human-subject data. The stakeholder component was based on a secondary analysis of previously published, anonymised data from Pafras et al. [26]. No personally identifiable information was used in the present analysis.

## Informed Consent Statement

Not applicable.

## Data Availability Statement

The GBIF occurrence data used in this study are publicly available through the Global Biodiversity Information Facility under dataset key 0063413-260519110011954. Environmental and anthropogenic predictor layers were obtained from publicly available sources, including Bio-ORACLE v3.0, EMODnet, GEBCO and the World Port Index, as described in the Materials and Methods. The curated occurrence dataset, R scripts for SDM and SEM analyses, and Python scripts used for sentiment analysis will be deposited in Zenodo upon acceptance of the manuscript. The stakeholder text data derived from Pafras et al. [26] are based on previously published and anonymised material; any restrictions on reuse follow the conditions of the original publication.

## Funding

This research was funded by the Region of Central Greece, under the Regional Development Programme (PPA), Priority Axis 2.4 "Risk Prevention and Management", Project Code: Π89-2.13.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Molnar JL, Gamboa RL, Revenga C, Spalding MD. Assessing the global threat of invasive species to marine biodiversity. *Front. Ecol. Environ.* **2008**, *6*, 485–492. DOI:10.1890/070064
2. Gallardo B, Clavero M, Sánchez MI, Vilà M. Global ecological impacts of invasive species in aquatic ecosystems. *Glob. Chang. Biol.* **2016**, *22*, 151–163. DOI:10.1111/gcb.13004
3. Zenetos A, Gofas S, Verlaque M, Cinar ME, Garcia Raso JE, Bianchi CN, et al. Alien species in the Mediterranean Sea by 2010. A contribution to the application of European Union's Marine Strategy Framework Directive (MSFD). Part I. Spatial distribution. *Mediterr. Mar. Sci.* **2010**, *11*, 381–493. DOI:10.12681/mms.87
4. Galil BS, Marchini A, Occhipinti-Ambrogi A. Mare Nostrum, Mare Quod Invaditur—The history of bioinvasions in the Mediterranean Sea. In *Biological Invasions in Changing Ecosystems*; Mack J, Lach J, Eds.; De Gruyter Open: Berlin/Heidelberg, Germany, 2018; pp. 43–82.
5. Aurelle D, Thomas S, Albert C, Bally M, Bondeau A, Boudouresque CF, et al. Biodiversity, climate change, and adaptation in the Mediterranean. *Ecosphere* **2022**, *13*, e3915. DOI:10.1002/ecs2.3915
6. Pastor F, Valiente JA, Khodayar S. A warming Mediterranean: 38 years of increasing sea surface temperature. *Remote Sens.* **2020**, *12*, 2687. DOI:10.3390/rs12172687
7. Copernicus Climate Change Service (C3S). European State of the Climate 2025. ECMWF, 2025. Available online: https://climate.copernicus.eu/esotc/2025 (accessed on 15 June 2026).
8. Monterosato TA. Enumerazione e sinonimia delle conchiglie mediterranee. *Giorn. Sci. Nat. Econ.* **1878**, *13*, 61–115. Available online: https://documentatiecentrum.watlab.be/imis.php?module=ref&refid=248203 (accessed on 15 June 2026).
9. Streftaris N, Zenetos A. Alien marine species in the Mediterranean—The 100 'worst invasives' and their impact. *Mediterr. Mar. Sci.* **2006**, *7*, 87–118. DOI:10.12681/mms.180
10. Barbieri MC, Deidun A, Maltagliati F, Zenetos A, Tlig-Zouari S, Gianni F, et al. First data on genetic diversity of the small pearl oyster *Pinctada radiata* (Leach, 1814) in Mediterranean populations by means of the mitochondrial COX1 marker. *Rapp. Comm. Int. Mer Médit.* **2013**, *40*, 711. Available online: https://www.researchgate.net/publication/263751973_First_data_on_genetic_diversity_of_the_small_pearl_oyster_Pinctada_radiata_Leach_1814_in_Mediterranean_populations_by_means_of_the_mitochondrial_COX1_marker (accessed on 15 June 2026).
11. Barbieri M, Deidun A, Maltagliati F, Castelli A. A contribution to the phylogeography of Pinctada imbricata radiata (Leach, 1814) (Bivalvia: Pteriidae) from the Eastern Mediterranean Sea by means of the mitochondrial COI marker. *Ital. J. Zool.* **2016**, *83*, 113–120. DOI:10.1080/11250003.2015.1106012

12. D'Iglio C, Capillo G, Di Fresco D, Famulari S, Albano M, Valić D, et al. Not All Aliens Come to Harm (Maybe): *Pinctada radiata* (Leach, 1814) Population and Habitat From a Coastal Lagoon (Central Mediterranean Sea). *Mar. Ecol.* **2025**, *46*, e70062. DOI:10.1111/maec.70062
13. Png-Gonzalez L, Aguilo-Arce J, Vázquez-Luis M, Carbonell A. New occurrence of *Pinctada imbricata radiata* in the Balearic Archipelago (NW Mediterranean Sea). *BioInvasions Rec.* **2021**, *10*, 853–858. DOI:10.3391/bir.2021.10.4.09
14. Pafras D, Theocharis A, Kondylatos G, Conides A, Klaoudatos D. Population biology of the non-indigenous rayed pearl oyster (*Pinctada radiata*) in the South Evoikos Gulf, Greece. *Diversity* **2024**, *16*, 460. DOI:10.3390/d16080460
15. Pafras D, Apostologamvrou C, Balatsou A, Theocharis A, Lolas A, Hatziioannou M, et al. Reproductive biology of pearl oyster (*Pinctada radiata*, Leach 1814) in the Eastern Mediterranean (South Evia Island). *J. Mar. Sci. Eng.* **2024**, *12*, 1259. DOI:10.3390/jmse12081259
16. Pafras D, Conides A, Vafidis D, Kapranas G, Klaoudatos D. Integrating Copernicus Earth Observation and Artificial Intelligence for Habitat Suitability Modeling of *Pinctada radiata* in Semi-Enclosed Coastal Watersheds of Central Greece. *J. Watershed Ecol.* **2026**, *1*, 10003. DOI:10.70322/jwe.2026.10003
17. Theodorou JA, Bekiari V, Douvi X, Ramfos A, Tzovenis I. Analysis of Nutritional Composition of the Pearl Oyster *Pinctada radiata* as a New Mediterranean "Bioresource" for Human Consumption. *Appl. Sci.* **2024**, *14*, 9757. DOI:10.3390/app14219757
18. Ladle RJ, Correia RA, Do Y, Joo GJ, Malhado AC, Proulx R, et al. Conservation culturomics. *Front. Ecol. Environ.* **2016**, *14*, 269–275. DOI:10.1002/fee.1260
19. Williams DR, Clark M, Buchanan GM, Ficetola GF, Rondinini C, Tilman D. Proactive conservation to prevent habitat losses to agricultural expansion. *Nat. Sustain.* **2021**, *4*, 314–322. DOI:10.1038/s41893-020-00656-5
20. Elith J, Leathwick JR. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* **2009**, *40*, 677–697. DOI:10.1146/annurev.ecolsys.110308.120159
21. Robinson NM, Nelson WA, Costello MJ, Sutherland JE, Lundquist CJ. A systematic review of marine-based species distribution models with recommendations for best practice. *Front. Mar. Sci.* **2017**, *4*, 421. DOI:10.3389/fmars.2017.00421
22. Araujo MB, New M. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **2007**, *22*, 42–47. DOI:10.1016/j.tree.2006.09.010
23. Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J. Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecol. Monogr.* **2022**, *92*, e01486. DOI:10.1002/ecm.1486
24. Grace JB. *Structural Equation Modeling and Natural Systems*; Cambridge University Press: Cambridge, UK, 2006.
25. Thorson JT, Andrews AG, Essington TE, Large SI. Dynamic structural equation models synthesize ecosystem dynamics constrained by ecological mechanisms. *Methods Ecol. Evol.* **2024**, *15*, 744–755. DOI:10.1111/2041-210X.14289
26. Dimitris P, Alexis C, Alexandros T, Dimitris V, Dimitris K. Stakeholder mental models for sustainable management of the invasive pearl oyster *Pinctada radiata* in the Eastern Mediterranean. *Hydroecol. Eng.* **2025**, *2*, 10016. DOI:10.70322/hee.2025.10016
27. Assis J, Fernández Bejarano SJ, Salazar VW, Schepers L, Gouvêa L, Fragkopoulou E, et al. Bio-ORACLE v3.0: Pushing marine data layers to the CMIP6 Earth System Models of climate change research. *Glob. Ecol. Biogeogr.* **2024**, *33*, e13813. DOI:10.1111/geb.13813
28. GBIF.org. *GBIF Occurrence Download. Dataset Key: 0063413-260519110011954*; GBIF Secretariat: Copenhagen, Denmark, 2026.
29. Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **2015**, *38*, 541–545. DOI:10.1111/ecog.01132
30. Thuiller W, Georges D, Guéguen M, Engler R, Breiner F, Lafourcade B, et al. biomod2: Ensemble Platform for Species Distribution Modeling. R package version 4.2.5, 2024. Available online: https://CRAN.R-project.org/package=biomod2 (accessed on 15 June 2026).
31. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259. DOI:10.1016/j.ecolmodel.2005.03.026
32. Breiman L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. DOI:10.1023/A:1010933404324
33. Ridgeway G. The state of boosting. *Comput. Sci. Stat.* **1999**, *31*, 172–181. Available online: https://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/a29acbd26d902d6fc125822a0031c09b/$FILE/boosting.pdf (accessed on 15 June 2026).
34. Kass JM, Muscarella R, Galante PJ, Bohl CL, Pinilla-Buitrago GE, Boria RA, et al. ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods Ecol. Evol.* **2021**, *12*, 1602–1608. DOI:10.1111/2041-210X.13628

35. Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **2019**, *10*, 225–232. DOI:10.1111/2041-210X.13107
36. Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A. Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* **2006**, *199*, 142–152. DOI:10.1016/j.ecolmodel.2006.05.017
37. Rosseel Y. lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. DOI:10.18637/jss.v048.i02
38. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.* **1999**, *6*, 1–55. DOI:10.1080/10705519909540118
39. Kline RB. *Principles and Practice of Structural Equation Modeling*, 4th ed.; Guilford Press: New York, NY, USA, 2016.
40. Hutto CJ, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proc. Int. AAAI Conf. Web Soc. Media* **2014**, *8*, 216–225. DOI:10.1609/icwsm.v8i1.14550
41. Braun V, Clarke V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. DOI:10.1191/1478088706qp063oa
42. Khalifa R, Hamadou RB, Giraldes BW, Joaquim S, Hizam Z, Hamza S, et al. The Impact of Increasing Seawater Temperatures over the Last 30 Years on the Reproductive Cycle of the Pearl Oyster *Pinctada radiata* (Leach, 1814) in the Arabian Gulf. *J. Mar. Sci. Eng.* **2024**, *12*, 2180. DOI:10.3390/jmse12122180
43. Tukey JW. Comparing individual means in the analysis of variance. *Biometrics* **1949**, *5*, 99–114. DOI:10.2307/3001913
44. Calabrese C, Giangrande A, Quarta E, Arduini D, Acquaviva MI, Biandolino F, et al. The Mediterranean introduced pearl oyster *Pinctada radiata*: Can an invasion be changed into a gain in a climate change scenario? *Mar. Pollut. Bull.* **2026**, *222*, 118736. DOI:10.1016/j.marpolbul.2025.118736
45. Tempesti J, Langeneck J, Maltagliati F, Castelli A, Lardicci C. Monitoring non-indigenous species in port habitats: First application of a standardized North American protocol in the Mediterranean Sea. *Front. Mar. Sci.* **2022**, *8*, 700730. DOI:10.3389/fmars.2021.700730
46. Galanidi M, Aissi M, Ali M, Bakalem A, Bariche M, Bartolo AG, et al. Validated Inventories of Non-Indigenous Species (NIS) for the Mediterranean Sea as Tools for Regional Policy and Patterns of NIS Spread. *Diversity* **2023**, *15*, 962. DOI:10.3390/d15090962
47. Gavrilović A, Piria M, Guo XZ, Jug-Dujaković J, Ljubučić A, Krkić A, et al. First evidence of establishment of the rayed pearl oyster, *Pinctada imbricata radiata* (Leach, 1814), in the eastern Adriatic Sea. *Mar. Pollut. Bull.* **2017**, *125*, 556–560. DOI:10.1016/j.marpolbul.2017.10.045
48. Shackleton RT, Richardson DM, Shackleton CM, Bennett B, Crowley SL, Dehnen-Schmutz K, et al. Explaining people's perceptions of invasive alien species: A conceptual framework. *J. Environ. Manag.* **2019**, *229*, 10–26. DOI:10.1016/j.jenvman.2018.04.045
49. Kapitza K, Zimmermann H, Martín-López B, von Wehrden H. Research on the social perception of invasive species: A systematic literature review. *NeoBiota* **2019**, *43*, 47–68. DOI:10.3897/neobiota.43.31619
50. García-Llorente M, Martín-López B, González JA, Alcorlo P, Montes C. Social perceptions of the impacts and benefits of invasive alien species: Implications for management. *Biol. Conserv.* **2008**, *141*, 2969–2983. DOI:10.1016/j.biocon.2008.09.003
51. Verbrugge LNH, Van den Born RJG, Lenders HJR. Exploring public perception of non-native species from a visions of nature perspective. *Environ. Manag.* **2013**, *52*, 1562–1573. DOI:10.1007/s00267-013-0170-1
52. Giakoumi S, Katsanevakis S, Albano PG, Azzurro E, Cardoso AC, Cebrian E, et al. Management priorities for marine invasive species. *Sci. Total Environ.* **2019**, *688*, 976–982. DOI:10.1016/j.scitotenv.2019.06.282
53. Correia RA, Ladle R, Jarić I, Malhado ACM, Mittermeier JC, Roll U, et al. Digital data sources and methods for conservation culturomics. *Conserv. Biol.* **2021**, *35*, 398–411. DOI:10.1111/cobi.13706
54. Jarić I, Bellard C, Correia RA, Courchamp F, Douda K, Essl F, et al. Invasion Culturomics and iEcology. *Conserv. Biol.* **2021**, *35*, 447–451. DOI:10.1111/cobi.13707