

# The role of genetic observatory networks in the detection and forecasting of marine non-indigenous species

Justine Pagnier

[justine.pagnier@gu.se](mailto:justine.pagnier@gu.se)

Department of Marine Sciences, SciLifeLab, University of Gothenburg <https://orcid.org/0000-0002-6531-1374>

Tobias Andermann

Uppsala University <https://orcid.org/0000-0002-0932-1623>

Mats Andersson

Swedish Veterinary Agency (SVA)

Matthias Obst

Department of Marine Sciences, University of Gothenburg, Sweden

---

## Article

## Keywords:

**Posted Date:** January 30th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-8702791/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** There is **NO** Competing Interest.

---

# Abstract

Marine biological invasions threaten ocean health, yet management remains reactive rather than proactive. Predictive tools, such as species distribution models, have the potential to provide indications about which areas are particularly likely to be colonised by non-indigenous species, allowing a more proactive management approach. Here we introduce an integrated framework utilising DNA-based monitoring data from genetic observatory networks to identify non-indigenous species for modelling and to independently validate the species distribution models forecasting invasion risk areas across European seas. We modelled habitat suitability for 69 marine non-indigenous species using global occurrence data and an ensemble modelling approach based on 6,555 individual species distribution models built with five different algorithms. Model validation against independent DNA-based detections from observatory networks showed 90% of observations occurred in predicted suitable habitat, confirming robust predictive capacity. Under current conditions, models identify invasion hotspots in the North Sea, North Atlantic, Mediterranean, and Black Sea. Climate projections to 2100 reveal pronounced vulnerability in Arctic and subarctic regions (up to 300% increase in habitat suitability under SSP5-8.5 scenario), while Mediterranean regions show modest change. We further demonstrate how the models can be applied in preventive action by supporting decisions in ballast water management. By coupling standardised and spatiotemporally consistent molecular monitoring with predictive modelling, we provide a scalable approach for marine biosecurity forecasts in a rapidly changing ocean.

## Introduction

In a rapidly changing world, the ability to predict ecological dynamics and potential disruptions is essential. Marine ecosystems are increasingly threatened by rising temperatures<sup>1–3</sup>, ocean acidification<sup>2,4</sup>, salinity changes<sup>5</sup>, and intensifying human activities<sup>6</sup>. Among these pressures, biological invasions by non-indigenous species (NIS) represent a persistent and accelerating driver of ecological change. These species, also called *alien*, *non-native*, or *introduced* species, are often transported through shipping, aquaculture, or drifting marine debris<sup>7</sup>. Some become *invasive*, outcompeting native organisms, restructuring food webs, and altering ecosystem services<sup>8</sup>. Biological invasions across marine, terrestrial, and freshwater ecosystems are now recognized as one of the top five drivers of global biodiversity loss<sup>9–11</sup>. Their global economic costs have risen by 702% from 1980–1999 to 2000–2019<sup>12,13</sup>.

Anticipating and mitigating invasions has become a priority for managing and preserving ocean ecosystems<sup>14</sup>. Biological invasions happen through distinct, yet coupled, stages of transport, introduction, establishment, and spread<sup>8</sup>, each presenting a separate opportunity for intervention. Yet, current marine biosecurity remains largely reactive rather than preventive, with countries typically deploying expensive post-invasion responses that rarely prove effective<sup>15,16</sup>.

Detection efforts, although effective, operate after introduction has already occurred. They allow responses during the establishment and spread phases, when eradication becomes exponentially more difficult and costly. However, species detection in marine environments presents significant logistical challenges. The ocean's vast scale and dynamic nature make systematic surveillance difficult, and traditional survey methods remain spatially patchy, resource-intensive, and challenging to standardise across regions. In this context, DNA-based monitoring with metabarcoding offers a solution, enabling rapid, standardised detection of biodiversity from small environmental samples. The ARMS-MBON network (Autonomous Reef Monitoring Structures - Marine Biodiversity Observation Network) illustrates this approach: through standardised deployments of ARMS units at marine observatories across Europe and adjacent regions since 2018, this network provides comparable DNA-based biodiversity data that capture a wide spectrum of benthic organisms<sup>17</sup>. Recent studies demonstrate that ARMS-MBON can detect marine NIS effectively and at early stages<sup>18–20</sup>. Additionally, many of the standards and methods developed by ARMS-MBON are now implemented in national monitoring programs (*e.g.*, Swedish Ports Monitoring program<sup>21,22</sup>). Yet, while improved detection accelerates response time, it remains inherently reactive.

Prediction, on the other hand, enables proactive intervention before introduction occurs. Species Distribution Models (SDMs)<sup>23,24</sup>, also known as Habitat Suitability or Ecological Niche-based Models, provide a way to forecast where species are likely to establish under current and future conditions. By identifying which regions provide suitable habitat for potential invaders, we can target prevention efforts at the transport stage. For example, this enables strengthening biosecurity at high-risk ports, prioritising surveillance in vulnerable ecosystems, and allocating management resources where they will be most effective. This is particularly crucial in marine systems, where dispersal can span hundreds of kilometers<sup>7,25</sup>. SDMs have been increasingly used to forecast climate-driven range shifts<sup>26–29</sup>, identify suitable habitats in data-poor regions<sup>30</sup>, and assess invasion risk<sup>31–34</sup>. One key application of spatial risk forecasts is the designation of Ballast Water Contingency Areas (BWCAs), which are zones where ships can perform ballast water exchange in case of contingencies. These areas must be carefully defined not to favour the spread of NIS into new areas.

However, current marine biosecurity lacks integrated analytical frameworks that couple these predictive models with sensor networks and management decision tools. This disconnect, where monitoring data, ecological forecasts, and policy interventions remain poorly integrated, creates time lags that prevent proactive NIS management. The convergence of genetic monitoring and species distribution modelling offers a solution: a continuous feedback loop from detection to risk assessment to intervention. Such integrated frameworks align with emerging digital infrastructures like Digital Twins of the Ocean (DTOs)<sup>35</sup>, which combine oceanographic, biological, and socioeconomic data into dynamic systems for scenario exploration and early warning.

In this study, we demonstrate the validity of coupling DNA-based monitoring with species distribution modelling for NIS risk assessment and management. Specifically, we address three questions: (i) can

SDMs trained on global occurrence data accurately predict current distributions of NIS detected by genetic observatories? (ii) can ensemble models identify spatial hotspots with elevated establishment potential (i.e., areas where multiple NIS find suitable habitat simultaneously) under current and future ocean conditions? and (iii) how can such spatially explicit forecasts support preventive NIS management?

Here, we modelled habitat suitability for 69 marine NIS detected by genetic observatories across European seas using global occurrence records from the Global Biodiversity Information Facility ([www.gbif.org](http://www.gbif.org)) and the Ocean Biodiversity Information System ([www.obis.org](http://www.obis.org)), coupled with environmental predictors from Bio-ORACLE<sup>36</sup>. We then validate predictions against available independent DNA-based detections. Our models predict where NIS can establish based on environmental conditions. This provides a foundation for risk assessment that can be integrated with species-specific impact evaluations to prioritise management actions. By projecting suitability under multiple climate scenarios, we assess how alien spread may shift as ocean conditions change, providing a blueprint for a sensor-model-decision framework for evidence-based marine biosecurity in a changing ocean. We illustrate how these spatial forecasts can be integrated with existing marine and digital infrastructure to support ecosystem-based management in the future.

## Results

We successfully modelled habitat suitability across current and future ocean conditions for various marine NIS detected by genetic observatory networks (ARMS-MBON<sup>17</sup>, Swedish Ports Monitoring<sup>21,22</sup>). These species span 10 major taxonomic groups (22 arthropods, 11 annelids, 10 molluscs, 7 rhodophytes, 7 chordates, 4 bryozoans, 3 cnidarians, 3 ochrophytes, one ctenophore, and one dinoflagellate, see Supplementary File 1), revealing species-specific predictions of suitable habitat as well as invasion pressure range-shift under climate change across European waters.

## Model assessments

### Individual models

Of the 81 NIS initially selected, 69 passed quality criteria and produced viable models, with 95 individual models built per species (6,555 SDMs total): 75 cross-validation models (5 algorithms × 5 k-folds × 3 pseudo-absences datasets), 15 fold-averaged models per pseudo-absences dataset, and 5 full-dataset models, trained on global occurrence data from GBIF and OBIS. Models were evaluated using 5-fold cross-validation, with performance assessed using TSS and ROC-AUC metrics.

All algorithms demonstrated strong discrimination ability on validation data (Table 1), with AUC values ranging from  $0.859 \pm 0.132$  (GAM) to  $0.917 \pm 0.0827$  (RF) and  $0.916 \pm 0.0747$  (MAXNET). Threshold-dependent performance showed greater variability, with TSS scores ranging from  $0.611 \pm 0.244$  (GAM) to  $0.687 \pm 0.188$  (MAXNET).

As expected, calibration metrics were uniformly higher than validation metrics across all algorithms (Table 1), indicating train-test gaps of varying magnitude between algorithms. For example, RF showed potential overfitting with near-perfect calibration ( $AUC \approx 1.00$ ,  $TSS = 0.989 \pm 0.00983$ ) compared to validation scores ( $AUC = 0.917 \pm 0.0827$ ,  $TSS = 0.647 \pm 0.213$ ). In contrast, MAXNET exhibited the smallest train-test gap (calibration  $TSS = 0.826 \pm 0.0644$  vs. validation  $TSS = 0.687 \pm 0.188$ ), while other algorithms showed intermediate patterns. Despite these gaps, validation metrics remained high across all algorithms (mean  $AUC = 0.892 \pm 0.101$ , mean  $TSS = 0.644 \pm 0.214$ ), indicating robust discriminatory ability. The ensemble modelling framework, which combines predictions from algorithms with complementary strengths and weaknesses, provides additional regularisation for final projections (see next section).

Table 1

**Models evaluation metrics.** Average ( $\pm$  standard deviation) values across all species, pseudo-absences' datasets, and model runs, for each algorithm and for all combined.

|                 | GAM                | MARS               | MAXNET             | RF                  | XGBOOST            | All                |
|-----------------|--------------------|--------------------|--------------------|---------------------|--------------------|--------------------|
| Calibration TSS | 0.813 $\pm$ 0.118  | 0.841 $\pm$ 0.0877 | 0.826 $\pm$ 0.0644 | 0.989 $\pm$ 0.00983 | 0.861 $\pm$ 0.088  | 0.866 $\pm$ 0.104  |
| Validation TSS  | 0.611 $\pm$ 0.244  | 0.648 $\pm$ 0.201  | 0.687 $\pm$ 0.188  | 0.647 $\pm$ 0.213   | 0.625 $\pm$ 0.210  | 0.644 $\pm$ 0.214  |
| Calibration AUC | 0.948 $\pm$ 0.0397 | 0.961 $\pm$ 0.0288 | 0.961 $\pm$ 0.0209 | 1.00 $\pm$ 0.000673 | 0.968 $\pm$ 0.0247 | 0.968 $\pm$ 0.0314 |
| Validation AUC  | 0.859 $\pm$ 0.132  | 0.883 $\pm$ 0.102  | 0.916 $\pm$ 0.0747 | 0.917 $\pm$ 0.0827  | 0.888 $\pm$ 0.0893 | 0.892 $\pm$ 0.101  |

The number of occurrence records used per species ranged from 10 (minimum allowed) to 847 (median: 118), with a clear positive relationship between sample size and model performance. Species with fewer than 50 occurrence records showed the highest variability and lowest mean performance ( $AUC: 0.822 \pm 0.15$ ;  $TSS: 0.491 \pm 0.3$ ), while species with 50–200 occurrences demonstrated substantial improvement ( $AUC: 0.906 \pm 0.06$ ;  $TSS: 0.67 \pm 0.14$ ) (Fig. 1a). Model performance continued to improve with increasing sample size, reaching optimal and stable performance in species with more than 100 occurrences ( $AUC: >0.94$ ;  $TSS: >0.75$ ).

Beyond improving performance, larger occurrences datasets substantially enhanced model performance consistency. Standard deviation for TSS decreased 84% (from 0.300 to 0.048) and for AUC decreased 87% (from 0.149 to 0.019) across the sample size gradient (Fig. 1b), indicating that models built with > 100 occurrences produce more consistent predictions across different algorithms and validation folds.

### Ensemble models

Given the variable performance across individual algorithms, we constructed ensemble models to use the strengths of multiple approaches. To create robust ensemble predictions for each species, we

retained only individual models that met minimum performance thresholds (TSS > 0.6 and AUC > 0.85).

Four species had fewer than 10 models available for ensemble construction and were flagged for cautious interpretation: *Apionsoma (Apionsoma) misakianum* (n = 9 models) *Herdmania momus* (n = 5), *Tharyx setigera* (n = 4), and *Pseudocalanus acuspes* (n = 4). Figures showing the performance distribution across species and algorithms are available in Supplementary Figs. 1 and 2.

Ensemble models showed improved performance compared to individual algorithms, with average scores of AUC =  $0.976 \pm 0.0137$  and TSS =  $0.863 \pm 0.055$  across all 69 species (calibration values, as ensemble models do not have validation values). Performance varied by species, with AUC ranging from 0.927 to 0.997 and TSS from 0.707 to 0.970.

Overall, ensemble models improved predictive performance, providing robust suitability estimates for subsequent spatial analyses.

#### External validation with independent DNA-based detections

We used metabarcoding-based species detections from genetic observatory networks as an independent dataset to evaluate our model predictions. These detections from ARMS-MBON deployments (2020 and 2021) and Swedish Ports Monitoring (2023 and 2025) were unpublished in GBIF/OBIS at the time of modelling and therefore excluded from model training. An independent validation dataset was therefore available for 33 species.

Our model predictions showed strong agreement with independent ARMS observations (Table 2) as approximately 90% of detections occurred in areas predicted as suitable (suitability > 0.5), and around 47% in areas predicted as highly suitable (suitability > 0.8) by our ensemble models (Supplementary Fig. 3). Species-specific and combined validation results are provided in Supplementary File 2. Absences were not tested as DNA-based absences do not confirm true absences.

## Individual suitability maps and uncertainty assessment

Weighted ensemble model predictions provide habitat suitability estimates ranging from 0 (unsuitable) to 1 (highly suitable) for each species. We assessed prediction uncertainty using two complementary metrics: coefficient of variation (*EMcv* in BIOMOD2, renamed CoV here) across algorithm predictions, and committee averaging (*EMca* in BIOMOD2, renamed CA here) showing the proportion of models agreeing on presence/absence. However, coefficients of variation values are inflated in areas of very low habitat suitability due to division by near-zero mean values, and we therefore recommend interpreting uncertainty with care in these areas.

All species-specific habitat suitability, CoV and CA maps are available on Figshare (see Data Availability Statement), for current ocean conditions and three 2100 climate scenarios (SSP1-2.6, SSP2-4.5, SSP5-8.5).

Here we present detailed results for two example species spanning different taxonomic groups and geographic patterns: the mollusc *Crepidula fornicata* and the copepod *Acartia (Acanthacartia) tonsa* (Fig. 2).

*C. fornicata* shows widespread suitable habitat throughout the North Atlantic, North Sea, and Kattegat/Skagerrak regions, with particularly high suitability along the southern and eastern coastline of the North Sea (Fig. 2a). Despite this broad predicted range, uncertainty remains relatively low across most suitable areas (Fig. 2b), with CoV values predominantly below 0.5, indicating good consensus among modelling algorithms. Model agreement is high ( $CA > 0.75$ ) throughout suitable areas (Fig. 2c), suggesting robust predictions in these areas.

*A. tonsa* displays a different geographical pattern, with suitable habitat concentrated in the Baltic Sea, and discrete areas in the Mediterranean and Black Sea (Fig. 2d). Uncertainty patterns for this species show moderate CoV values (0.25–0.75) in key suitable areas, while model agreement remains high ( $CA > 0.75$ ) in core suitable habitats such as the southern and central Baltic Sea (Fig. 2f).

For both species, the coefficient of variation increases with distance to the coast, and these areas are also areas where the models do not fully agree on absences or presences.

Climate projections under the moderate emissions scenario (SSP2-4.5, 2100) reveal species-specific patterns of habitat redistribution (Fig. 3).

*C. fornicata* exhibits the most important redistribution, with substantial decreases in suitability ( $< -0.15$ ) predicted throughout its current core areas in the North Sea, English Channel, and Irish Sea (Fig. 3a). In parallel, we observe suitability gains in previously unsuitable Arctic waters, particularly the Barents Sea and northern Norwegian coast, suggesting a northward displacement of suitable habitat. For *A. tonsa*, projected changes are more geographically restricted (Fig. 3c). The species faces moderate to substantial decreases across much of the Baltic Sea, currently a core suitable habitat, as well as scattered decreases in the Mediterranean and Black Sea.

All species-specific projections for all three SSP scenarios (SSP1-2.6, SSP2-4.5, SSP5-8.5) and associated uncertainty metrics are available in the data repository.

## Current priority areas for NIS management

Beyond species-specific assessments, identifying geographic areas where multiple NIS may co-occur is critical for prioritising limited biosecurity resources. To identify priority areas for biosecurity alert and intervention, we aggregated suitability predictions across all 69 modelled species, revealing geographic hotspots where environmental conditions support multiple NIS (Fig. 4). Mean suitability values reflect the average habitat favorability across all 69 modelled species in each grid cell, with higher values indicating that, on average, environmental conditions are more favorable for the species pool. Note that

this metric captures average suitability rather than species richness. Areas with high mean values may reflect either many species with moderate suitability or fewer species with very high suitability.

Severe priority areas ( $S \geq 0.6$ , dark red) are spatially restricted to small, fragmented coastal patches primarily concentrated along both sides of the English Channel and the northern part of the Bay of Biscay. High priority zones ( $S = 0.5-0.6$ , red-orange) extend these hotspots along the North Sea coastlines and up to the Kattegat/Skagerrak region. These zones also comprise the northern Adriatic Sea, the Atlantic side of the Strait of Gibraltar, the northern Black Sea coast, the Tunisian coast, the Egyptian coast around the Suez Canal, parts of the southern French Mediterranean coast and of the northern parts of the Norwegian coastline. Moderate priority areas ( $S = 0.3-0.5$ ) show broader geographic extent, covering additional Scandinavian coastal regions (including the Baltic), Icelandic coasts, and appearing in scattered locations throughout the Mediterranean and along Atlantic-facing shores. Arctic waters and offshore areas show consistently low suitability ( $S < 0.3$ ), reflecting limited NIS establishment under current conditions.

## Future changes in NIS suitability and regional patterns

We projected habitat suitability to 2100 using environmental data from Bio-ORACLE v3.0 based on CMIP6 climate models under three emission scenarios: SSP1-2.6 (high mitigation), SSP2-4.5 (moderate mitigation), and SSP5-8.5 (high emissions). Change in mean suitability was calculated as the difference between future (2090–2100 average) and current (2000–2020 average) conditions across all 69 species.

Projections under future climate scenarios revealed predominantly northward increases in mean habitat suitability, with magnitude strongly dependent on emission pathway (Fig. 5).

Under the low emission scenario (SSP1-2.6), changes in mean suitability were relatively modest with 96% of the total study area (European continental shelf) experiencing minimal change (between  $-0.05$  and  $+0.05$ , white on maps). Only 2.8% of the total study area showed slight to moderate increases, primarily in northern latitudes (Fig. 5a), while less than 1% exhibited slight decreases along the coast of the North Sea and the Kattegat.

The intermediate emission scenario (SSP2-4.5) projected more pronounced increases, with 16% of the total study area experiencing slight to substantial increases, particularly in the Norwegian and Barents Seas (Fig. 5b), and only 0.4% showing slight decreases in small coastal areas.

The high emission scenario (SSP5-8.5) produced the most dramatic changes: 55% of the total study area showed slight to moderate increases, while 3.6% exhibited important to major increases (Fig. 5c). These increasingly suitable areas were concentrated in northern regions, especially in Arctic and sub-Arctic waters.

We analysed these changes across 18 Marine Ecoregions of the World (MEOW)<sup>37</sup> within European waters to identify regional patterns of climate-driven alien spread risk. Regional analysis revealed a pronounced latitudinal gradient (Fig. 6, Supplementary Table 1). Arctic and subarctic ecoregions showed both the largest relative (Fig. 6a) and absolute (Fig. 6b) increases. A detailed table with the computed relative and absolute changes can be found in Supplementary File 3.

Arctic regions exhibited the most substantial relative changes under SSP5-8.5: the North and East Barents Sea showed a  $301\% \pm 135\%$  increase (mean  $\pm$  SD across pixels within the ecoregion), followed by Northern Norway and Finnmark ( $104\% \pm 71\%$ ), and Icelandic waters (North and East Iceland:  $129\% \pm 50\%$ ; South and West Iceland:  $132\% \pm 70\%$ ). Even under the conservative SSP1-2.6 scenario, the Barents Sea showed a  $65\% \pm 57\%$  increase.

Mid-latitude Atlantic regions (North Sea, Celtic Seas, South European Atlantic Shelf) demonstrated more moderate responses, with increases of 70–90% under SSP5-8.5 and 10–15% under SSP1-2.6.

Mediterranean ecoregions generally showed modest increases of 10–30% under SSP5-8.5, with the Adriatic and Ionian Seas (28–29%) at the higher end and the Alboran Sea showing minimal change. Under lower-emission scenarios, Mediterranean increases were around 10% or below.

The Black Sea exhibited near-zero or slightly negative changes across all scenarios despite currently high habitat suitability.

## Ballast water contingency areas (BWCA)

We then explored a potential application of the above results for biosecurity. Using the aggregated suitability map for current conditions, we identified potential Ballast Water Contingency Areas in the eastern North Sea and Baltic Sea by applying spatial criteria: we kept areas with mean habitat suitability below 0.2 and located > 7 kilometers from Marine Protected Areas, offshore wind farms, and coastlines. These represent invasion cold spots where NIS establishment risk is minimised (Fig. 7).

Ships operating in these regions could be directed to these zones during contingency situations requiring ballast water discharge. Note that these thresholds are illustrative and may be adjusted based on specific contingency scenarios and regional biosecurity priorities.

## Discussion

The application of species distribution models (SDMs) to non-indigenous species (NIS) has been contested due to potential violations of the niche-environment equilibrium assumption<sup>38–40</sup>. For example, recent work on the Pacific oyster invasion in Swedish waters has shown that species-habitat associations at invasion fronts may differ from established populations<sup>41</sup>. However, Jiménez-Valverde et al.<sup>42</sup> argue that when appropriately constructed (*i.e.* using distributional data from all invaded regions,

predictors linked to physiological requirements, and careful model evaluation), SDMs provide robust predictions for invasion risk assessment. We show here that a dynamic integration of SDMs with standardised genetic observatories addresses these limitations allowing for continuously updated species distributions and thereby constantly refining estimates of the realised environmental preferences in invaded regions.

By modelling the habitat suitability of 69 non-indigenous marine metazoan and algal taxa (spanning ten phyla) detected on Autonomous Reef Monitoring Structures across Europe, we obtained species-specific habitat suitability maps (current and future scenarios) and maps of priority areas for NIS management. These species were systematically and manually curated to confirm their introduced status, distinguishing genuine non-indigenous taxa from native species and potential contamination<sup>19</sup>. This curation step is essential because while DNA-based monitoring efficiently captures cryptic invasions and early-stage establishments in near real-time<sup>43,44</sup>, genetic detections alone cannot distinguish introduced from native occurrences.

Our validation demonstrates that ensembles of models trained on global occurrence data effectively predict the distributions of non-indigenous species detected through genetic observatories. Our ensemble modelling approach combines predictions from algorithms with complementary strengths and weaknesses, effectively regularising final projections by balancing conservative and flexible model types with different overfitting tendencies. Across 33 tested species, 90% of their independent DNA-based detections from the ARMS-MBON network and the Swedish Ports Monitoring occurred within areas predicted as suitable habitat ( $> 0.5$ ). Importantly, this framework allows continuous model validation: as new DNA-based detections emerge, they can be used to assess model accuracy before being incorporated into updated training datasets, creating an iterative cycle of prediction, validation, and refinement.

It is important to note that our habitat suitability projections are impact-neutral: we model environmental suitability for non-indigenous species regardless of whether individual species have demonstrated ecological or economic impacts. Not all non-indigenous species cause negative effects, and impacts vary substantially across species and contexts<sup>45</sup>. Our framework provides spatially explicit predictions of where species are likely to establish based on environmental conditions, but decisions about management priorities require integration of species-specific impact assessments alongside habitat suitability. This approach enables evidence-based, targeted resource allocation rather than categorical responses to all non-indigenous species.

Stacking individual species projections allows the delineation of priority areas for NIS management. Under current conditions, these concentrate in the North Sea, with high mean habitat suitability extending along both sides of the English Channel, the northern Bay of Biscay, and through coastal regions to the Kattegat/Skagerrak. In contrast, the Mediterranean Sea exhibits patchier suitability despite experiencing high propagule pressure from shipping traffic<sup>46</sup> and connectivity to tropical source regions via the Suez Canal<sup>47</sup>. These areas of elevated invasion risk reflect current environmental constraints on

NIS establishment, while the species-based predictions (e.g., *Crepidula fornicata* and *Acartia (Acanthacartia) tonsa*) illustrate the underlying individual responses to environmental conditions driving species-specific range expansions.

Climate change is rapidly reshaping these biogeographic dynamics. Future projections show scenario-dependent changes in habitat suitability across European waters, with high-emissions scenarios (SSP5-8.5) predicting increases across 58.79% of the study region by 2100 (over three times the extent under moderate emissions). A pronounced latitudinal gradient emerges across all scenarios, with Arctic and subarctic regions experiencing the greatest increases. The Barents Sea exhibits a three-fold increase in habitat suitability under SSP5-8.5. These projections indicate potential weakening of thermal barriers that currently constrain poleward expansion<sup>1</sup>, suggesting that climate-driven range shifts will define future invasion dynamics in European seas. This pattern aligns with observed increases in NIS discovery rates in Arctic systems globally<sup>48</sup>, with hotspots emerging in the Iceland Shelf, Barents Sea, Norwegian Sea, Hudson Bay, and Chukchi/Eastern Bering seas<sup>32,49</sup>. Together, these findings indicate that Arctic and subarctic marine ecosystems face heightened invasion pressure as warming relaxes historical thermal constraints. Concurrent expansion of Arctic shipping routes driven by sea ice reduction may further increase introduction opportunities through ballast water and biofouling vectors<sup>50</sup>. Hence, the identified sub-/arctic regions should be actively protected from potential introductions in the future.

Our habitat suitability projections provide spatially explicit information for multiple biosecurity applications. Ballast water discharge represents one of the primary vectors for marine biological invasions globally, with vessels transporting thousands of species across biogeographic barriers annually<sup>51,52</sup>. We demonstrate one decision-support application through identification of ballast water contingency areas in Northern European waters. This approach builds upon recent regulatory developments in the OSPAR commission, where the Intra-North Sea Ballast Water Contingency Area was implemented in 2025 and will remain effective until at least 2030<sup>53</sup>. By overlaying aggregated current habitat suitability map with Marine Protected Areas and Offshore Wind Farms in HELCOM regions, we delineate zones where ballast water discharge presents reduced invasion risk, i.e. areas characterised by low mean suitability and enough distance from sites that could facilitate secondary spread<sup>54,55</sup>.

This approach offers a practical complement to existing ballast water management regulations, particularly for vessels unable to perform mid-ocean exchange due to safety constraints or operational limitations<sup>56</sup>. Our contingency zone framework provides port authorities and vessel operators with spatially explicit alternatives: if mid-ocean exchange is not feasible, discharge in designated low-risk areas minimises the probability that propagules encounter suitable habitat or colonise ecologically sensitive sites. As climate scenarios shift the geography of habitat suitability, these contingency zones can be updated, ensuring management strategies remain responsive to changing species' spreads. Additionally, risk assessments at port-level can help prioritise inspection and monitoring efforts by cross-referencing vessel arrivals with projected suitability for species known to be present in source regions<sup>57</sup>.

Surveillance optimisation represents another key application. Deploying genetic observatories in areas with high projected suitability but no current detections enables interception during establishment phases when eradication remains feasible<sup>43</sup>. This proactive approach contrasts with reactive monitoring in areas where invasions have already occurred, potentially reducing management costs and ecological impacts.

Model projections and genetic monitoring both carry inherent uncertainties that must be acknowledged. First, our taxonomic scope is constrained to benthic fauna and flora detectable on ARMS units using COI and 18S metabarcoding, excluding most fish and pelagic species. This is particularly relevant for the Mediterranean, where our species list likely underrepresents Lessepsian migrants from the Suez Canal, many of which are mobile taxa not always detected on ARMS. Additionally, the geographic distribution of ARMS deployments shapes our species inventory, with undersampled regions potentially harboring undetected NIS.

Genetic monitoring faces inherent challenges that may affect detection accuracy. False negatives can occur due to low species abundance, seasonal variability in DNA shedding rates, or limited DNA persistence in the environment<sup>58,59</sup>. Metabarcoding identifications are further constrained by incomplete reference databases, cryptic species complexes, and potential sequence misassignments, though our use of manually curated species lists, verified against WRiMS records, substantially mitigates these taxonomic uncertainties.

SDMs project habitat suitability based on realised environmental niches captured in available occurrence data, which may not fully represent a species' fundamental niche or account for biotic interactions, dispersal limitations, or rapid evolutionary adaptation<sup>38</sup>. Recent methodological advances combining population genomics with SDMs show promise for incorporating genetic connectivity and local adaptation signals into niche models<sup>60</sup>, though such approaches require genetic data types beyond the scope of metabarcoding-based observatories. Similarly, joint species distribution models (JSDMs) offer complementary approaches by explicitly accounting for species co-occurrences and potential biotic interactions through residual correlation structures<sup>61,62</sup>. Therefore, our projections represent potential suitability rather than realised distributions, as we do not explicitly model dispersal, biotic interactions, or propagule pressure. Additionally, our reliance on pseudo-absences rather than confirmed absences introduces additional uncertainty, as background sampling points may occur in locations where species are present but undetected. However, as ARMS-MBON time series lengthen with continued deployment, the accumulation of temporal sampling will allow distinction between true absences and non-detection events, improving the reliability of absence inference from DNA-based datasets and enabling more robust model development<sup>63</sup>.

Environmental projections derived from CMIP6 climate models carry inherent uncertainty that increases with projection distance into the future, primarily due to scenario divergence and inter-model variability<sup>64,65</sup>. While sea surface temperature projections for European shelf seas have medium-high confidence in IPCC assessments, other oceanographic variables show greater inter-model

disagreement. Importantly, projections to 2100 may involve novel climate combinations outside the environmental space captured in our training data, requiring model extrapolation into conditions with no modern analogue<sup>24</sup>. This is particularly relevant under high-emission scenarios (SSP5-8.5), where the magnitude of environmental change increases uncertainty in predicted responses. We partially address these uncertainties through ensemble modelling and coefficient of variation analyses, though our focus on relative spatial patterns (identifying where risk increases most) should be more robust than absolute distribution predictions.

Forecasting ecological futures requires humility: marine biological systems are complex and subject to contingencies our models cannot anticipate<sup>66</sup>. However, our modelling approach can accommodate much of the complexity when continuously updated with emerging data and validated against empirical observations.

Integrating anthropogenic pressure layers (including port traffic, shipping networks, and aquaculture facilities) would transform habitat suitability into comprehensive invasion risk maps by accounting for introduction pathways alongside environmental suitability. Network analysis linking source populations to European coasts via shipping routes or currents could refine arrival probabilities. In the future, it will also be possible to include oceanographic models to incorporate species dispersal dynamics in this framework, which would enable prediction of spread rates and invasion corridors after establishment. This, however, requires more data on the studied species, especially when considering marine invertebrates which can often be overlooked<sup>67</sup>, and is therefore beyond the scope of this study.

Integration with genetic observatory infrastructure represents a critical next step. Recent advances in multi-omics-driven frameworks for invasive species management emphasise the value of integrating molecular detection with predictive modelling to enable proactive, rather than reactive, biosecurity responses<sup>68</sup>. The ARMS-MBON network and similar DNA-based monitoring networks can generate continuous species observations data layers that can validate and refine model predictions on a regular basis. Standardising data pipelines between genetic observatories will enable automated model updates as new detections accumulate, transforming static forecasts into dynamic early warning systems. Our framework supports iterative updating, i.e., as new NIS arrive or established NIS shifts its range, incorporating observations and rerunning projections will track the invasion frontiers. Automated workflows integrating occurrence databases, environmental data streams, and modelling pipelines could enable annual or event-triggered model updates. This adaptive approach treats forecasting as an automated process rather than a one-time analysis.

The integration of standardised genetic monitoring with ensemble modelling provides actionable spatial information for proactive biosecurity. Integration into the European Digital Twin of the Ocean infrastructure would enable real-time decision support, connecting invasion forecasts to marine spatial planning and climate adaptation strategies. Such systems require sustained investment in cyberinfrastructure and data standardisation across monitoring networks, but foundational components

demonstrated here confirm this vision is achievable and point toward marine biosecurity infrastructures fit for an era of accelerating global change.

## Material and Methods

### Data

#### Species selection

Species were included if detected in any processed ARMS data from 2018–2024, comprising: (1) the ARMS-MBON network (19 observatories across 14 countries, 2018–2021)<sup>19</sup>, or (2) the Swedish Ports Monitoring program (SPM; 23 sites, 2023–2024)<sup>21,22</sup>.

Candidate NIS species were initially detected using standardised genetic protocols based on DNA metabarcoding of mitochondrial COI and nuclear 18S rRNA markers, following the methodology established by the ARMS-MBON consortium<sup>17,18,69</sup>. Initial DNA-based species detections were cross-referenced against records from the World Register of Introduced Marine Species (WRiMS)<sup>70</sup>. Ambiguous or mis-assigned operational taxonomic units (OTUs) were removed through manual curation, retaining only confident species-level identifications and confirmed alien status in Europe. The detailed methodology and finalised ARMS-MBON NIS list has been published in Pagnier et al.<sup>19</sup>. The same method has been applied on ARMS metabarcoding data obtained from the SPM campaigns from 2023 and 2024. All steps and code used to extract NIS occurrences from ARMS metabarcoding datasets have been published by Daraghmeh<sup>71</sup>. This process yielded 81 marine NIS with confirmed genetic detections and verified alien status in at least one European marine region (Supplementary File 1).

This complete dataset (all years) was used only for species selection to maximise detection of non-indigenous species present in European waters. However, publication status in GBIF/OBIS at the time of modelling determined whether specific ARMS detections were included in model training or reserved for independent validation (Table 2). In fact, our modelling workflow is based on GBIF/OBIS data; published ARMS detections were therefore included in training by default (see Occurrence data compilation and cleaning).

Table 2

Genetic observatory datasets and their role in modelling workflow. ARMS-MBON stands for Autonomous Reef Monitoring Structures - Marine Biodiversity Observatory Network, which is a pan-European network sampling and sequencing in a standardised way. SPM stands for Swedish Ports Monitoring, which is a program using similar methods as ARMS MBON for ARMS sampling, sequencing and bioinformatics, but focused on ports located in Sweden. ARMS-MBON and SPM data selected for model training were complemented by other GBIF and OBIS occurrences (see next section). The SPM dataset from 2025 had not been fully processed at the time of species selection and was therefore only used for validation.

| Dataset   | Years     | N sites | Publication status (August 2025) | Species selection | Model training | Independent validation |
|-----------|-----------|---------|----------------------------------|-------------------|----------------|------------------------|
| ARMS-MBON | 2018–2019 | 15      | Published in GBIF/OBIS           | ✓                 | ✓              | —                      |
| ARMS-MBON | 2020–2021 | 13      | Unpublished                      | ✓                 | —              | ✓                      |
| SPM       | 2023      | 6       | Unpublished                      | ✓                 | —              | ✓                      |
| SPM       | 2024      | 11      | Published in GBIF/OBIS           | ✓                 | ✓              | —                      |
| SPM       | 2025      | 10      | Unpublished                      | —                 | —              | ✓                      |

#### Occurrence data compilation and cleaning

Global occurrence records of the 81 selected species were obtained from two databases to maximise spatial coverage: the Global Biodiversity Information Facility (GBIF, [www.gbif.org](http://www.gbif.org)) using the *rgbif* R package v.3.8.1<sup>72</sup>, and the Ocean Biodiversity Information System (OBIS, [www.obis.org](http://www.obis.org)), using *robis* package v.2.11.3<sup>73</sup>. ARMS data availability in GBIF/OBIS determined their use in our modelling workflow (Table 2).

We applied a multi-step filtering procedure to ensure data quality. First, occurrence points on land were removed to exclude specimens stored in research facilities or records with erroneous georeferencing. Duplicates and points with invalid coordinates or geospatial issues were also removed. We only kept occurrences from January 2000 to July 2025 to match the temporal resolution of the environmental predictors used in modelling (see Environmental predictors section). This initial filtering resulted in 226,478 occurrence points across 81 species.

To reduce spatial sampling bias, we applied spatial thinning using the *spThin* R package v.0.2.0<sup>74</sup> to all species with sufficient records, retaining only one occurrence per 10 km × 10 km grid cell using the *thin* function. After spatial thinning, 25,745 occurrences remained across the 81 species.

During model formatting with BIOMOD2<sup>75</sup>, occurrences falling outside the extent of environmental predictors or coinciding with cells containing missing values were excluded. This resulted in 12,300 occurrence points across all species. This important loss was due to the continental shelf restriction of

the environmental predictors (see next section). Twelve species with fewer than 10 remaining occurrences were excluded from further analysis. The final dataset comprised 69 species with sufficient data for modelling.

A table summarising the number of occurrences per species at each stage can be found in Supplementary File 4. All raw and thinned/cleaned occurrences can be accessed on Figshare (see Data availability statement).

Environmental predictors

## Predictors selection

We selected 34 environmental variables from Bio-ORACLE v.3.0<sup>36</sup> representing key physiological drivers (temperature, salinity, oxygen, nutrients) and biological productivity (chlorophyll, primary productivity). Variables were downloaded using the *biooracler* R package v.0.0.0.9000 (<https://github.com/bio-oracle/biooracler>). We generated an additional distance-to-coast layer by calculating the distance from each marine cell to the nearest coastline using the Bio-ORACLE bathymetry raster as reference. All layers had a resolution of 0.05°.

To reduce multicollinearity, we applied Variance Inflation Factor (VIF) analysis using the *vifstep* function from the *usdm* R package v.2.1–7<sup>76</sup>, iteratively removing variables with VIF > 10. This reduced the predictor set to 19 variables (Supplementary Table 2). To ensure consistency between current and future projections, we retained surface chlorophyll concentration rather than mean depth chlorophyll concentration, as the latter was unavailable for future climate scenarios.

## Temporal scope

We obtained the same 19 environmental predictors for both current conditions (2000–2020 average) and end-of-century projections (2090–2100 average) under three CMIP6-based scenarios: SSP1-2.6 (high mitigation pathway), SSP2-4.5 (moderate mitigation scenario), and SSP5-8.5 (high emission scenario). Bathymetry and distance-to-coast layers remained constant across all scenarios.

## Spatial processing

Bio-ORACLE layers contained missing values in shallow coastal cells where many occurrence records were located. To retain these ecologically relevant areas, we applied focal interpolation using nearest-neighbor values to estimate environmental conditions for coastal cells, with coastlines defined using the medium-scale world map from the *rnaturalearth* R package v.1.1.0<sup>77</sup>.

We then restricted all environmental stacks to the continental shelf (0–200 m depth) based on the bathymetry layer. This spatial constraint was critical for two reasons: (1) it focused pseudo-absence selection on environmentally accessible habitats, reducing bias; and (2) it prevented models from simply learning to distinguish coastal from open-ocean environments, as our target species are predominantly

coastal taxa. The final processed stacks comprised 19 environmental layers for four temporal scenarios (current + 3 SSP scenarios).

### Pseudo-absences generation

We used a hybrid pseudo-absence (PA) selection strategy to balance spatial realism with geographic coverage in our models. We generated three independent PA datasets per species, each using a 50:50 mixture of two sampling strategies, to reduce sensitivity to pseudo-absence configuration. Half of the pseudo-absences were sampled using a disk-based approach: points were selected at distances between 20 and 100 km from observed presences, ensuring that pseudo-absences represented geographically accessible environments where species were not detected. The remaining half were randomly distributed across the study area (continental shelf only), providing broader environmental coverage and reducing spatial bias inherent in presence-only data.

The number of pseudo-absences was scaled to three times the number of presence records per species, following BIOMOD2 recommendations<sup>78</sup>. A summary of the number of pseudo-absences generated per species can be found in Supplementary File 4.

Before model fitting, we implemented a deduplication procedure to remove spatially redundant pseudo-absence points. When multiple pseudo-absences from different replicates overlapped spatially, we retained unique coordinates while tracking which replicates used each location.

## Modelling

### Model construction

Species distribution models (SDMs) were constructed using functions from the *BIOMOD2* R package v.4.2-6-2<sup>75</sup>. For each species, we implemented an ensemble modelling approach, combining outputs from multiple algorithms including Generalized Additive Models (GAMs), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy (MAXNET), Random Forest (RF), and Extreme Gradient Boosting (XGBOOST). These algorithms were selected to represent diverse modelling approaches (parametric vs. machine learning; linear vs. non-linear) and capture different aspects of species-environment relationships, with ensemble predictions reducing bias associated with any single algorithm<sup>79</sup>.

Occurrence data were randomly divided into five folds, with each fold serving once as validation data while the remaining four folds trained the model. This process was repeated for all three PA datasets, resulting in 75 cross-validated models per species (5 algorithms × 5 folds × 3 PA datasets). Additionally, 15 fold-averaged models (5 algorithms × 3 PA datasets) and 5 full-dataset models (trained on all data) were produced, yielding 95 total models per species.

Models were then evaluated using two complementary metrics: the True Skill Statistic (TSS), which accounts for both sensitivity and specificity while remaining independent of prevalence<sup>80</sup>, and the Area

Under the ROC Curve (AUC), which measures discrimination ability across all thresholds<sup>81</sup>. TSS and AUC were calculated on validation data (20% of occurrences in each fold) to assess model generalisation.

To assess the relationship between sample size and model performance, we binned species into five occurrence categories (< 50, 50–100, 100–200, 200–500, > 500 records) and calculated mean performance metrics and standard deviations for each bin. LOESS smoothing was applied with the *geom\_smooth* function to visualize performance trends across the sample size gradient.

### Ensemble models and current projections

For each species, individual models with TSS > 0.6 and AUC > 0.85 (indicating reliable predictive performance) were kept. Ensemble models were produced in *BIOMOD2* using the weighted mean approach (*EMwmean*) to combine predictions from individual algorithms. In this ensemble type, each algorithm's contribution is weighted by TSS during model calibration. This means that models with higher predictive accuracy contribute more strongly to the final ensemble, while models with similar scores receive comparable weights.

Current projection maps were built using the *BIOMOD\_EnsembleForecasting* function, the ensemble model and the stack of environmental predictors for the current scenarios. The resulting maps therefore represent the consensus probability of suitable habitat across all algorithms, emphasising the most reliable models while retaining balanced representation of the ensemble.

### Assessing ensemble models' uncertainty

In addition to the ensemble weighted average projections, we also produced coefficients of variation maps (named *EMcv* by *BIOMOD2*, renamed CoV here) and committee averaging maps (named *EMca* by *BIOMOD2*, renamed CA here) for each species and scenario (current and future).

The CoV represents the coefficient of variation (standard deviation divided by mean) of predicted habitat suitability across all selected algorithms within the ensemble. High CoV values indicate areas where models disagree strongly on species suitability, whereas low values indicate higher inter-model agreement. Note that CoV is inflated in areas with very low mean suitability (< 0.3) due to near-zero denominators. For visualisation, we therefore masked CoV values to areas with suitability > 0.3, displaying low-suitability areas (< 0.3) in white to indicate where CoV estimates are unreliable for uncertainty assessments. CoV values from *BIOMOD2* (given as percentages) were divided by 100 for consistency.

The CA (committee averaging) represents the proportion of models agreeing on species presence or absence after transforming continuous predictions into binary classifications using thresholds that maximise TSS on validation data. This metric serves both as a prediction and an uncertainty measure: values near 1.0 indicate strong consensus for presence, values near 0.0 indicate consensus for absence, and values near 0.5 indicate high disagreement (approximately half the models predict presence while the other half predict absence). CA was rescaled from 0-1000 to 0–1 for consistency with other outputs.

## Future habitat suitability projections

Weighted ensemble models were projected onto three future climate scenarios (SSP1-2.6, SSP2-4.5, SSP5-8.5) using *BIOMOD\_EnsembleForecasting* with 2090–2099 environmental predictors. Projections used the same spatial extent as model training (0–200m continental shelf).

## Processing of projections maps

To restrict analyses to ecologically relevant marine areas and species' non-indigenous ranges, we applied a multi-step spatial masking procedure on all BIOMOD's projections outputs. First, all terrestrial cells were converted to NA. Second, we restricted projections to European waters using Marine Ecoregions of the World (MEOW)<sup>37</sup>. Importantly, species were masked to only those MEOWs where they are considered non-indigenous. This species-specific regional masking ensures that hotspot analyses reflect true invasion risk rather than including native populations. The MEOW assignments per species are provided in Supplementary File 5.

Finally, we rescaled all current and future habitat suitability maps from BIOMOD2's native 0–1000 scale to a 0–1 range for intuitive interpretation.

## External validation with independent ARMS detections

To assess model performance against spatially and temporally independent data, we compared ensemble predictions to ARMS detections that were excluded from model training. Specifically, we used: (1) ARMS-MBON detections from 2020–2021, and (2) SPM detections from 2023 and 2025 (Table 2). These datasets were unpublished in GBIF/OBIS at the time of modelling (August 2025) and therefore represent truly independent validation data, whereas ARMS-MBON 2018–2019 and SPM 2024 data had been published and were included in the training dataset along with other global occurrences. While these data did not allow formal statistical validation due to species-level imbalances in sample sizes and the lack of confirmed absences (DNA-based absences do not confirm true absence<sup>82</sup>), they provided a valuable real-world test of model predictions.

For each independent detection, we extracted the predicted habitat suitability value (0–1) at that location from our current-day ensemble models. Of the 69 modelled species, 51 had independent DNA-based detections available, of which 33 had detections falling within the projection extent (continental shelf). We calculated the proportion of these detections located in areas predicted suitable ( $> 0.5$ ) and areas predicted highly suitable ( $> 0.8$ ). Details can be found in Supplementary File 2.

# Analysis of model predictions

We analysed habitat suitability projections using R v.4.3.1 with the *terra* v.1.8–54<sup>83</sup>, *raster* v.3.6–32<sup>84</sup>, and *dplyr* v1.1.4<sup>85</sup> packages for spatial processing, and *ggplot2* v.3.5.2<sup>86</sup> for visualisation. Analyses were conducted at two scales: individual species and aggregated across all 69 species.

## Individual species analyses

For each species, we generated five map types to characterise current distributions, uncertainty, and climate-driven changes:

1. Current habitat suitability: Weighted ensemble predictions showing continuous suitability values (0–1 scale)
2. Prediction uncertainty (CoV): Coefficient of variation maps masked to areas with suitability > 0.3, as CoV values are inflated in low-suitability areas due to near-zero denominators. Areas with  $S \leq 0.3$  were displayed in white to indicate unreliable CoV estimates.
3. Model agreement (CA): Committee averaging maps showing the proportion of models agreeing on presence or absence (see Ensemble Models section)
4. Projected change: Difference maps ( $\Delta S = S_{\text{future}} - S_{\text{current}}$ ) for SSP2-4.5 as an illustrative scenario, classified into seven categories: moderate decrease ( $< -0.10$ ), slight decrease ( $-0.10$  to  $-0.05$ ), no change ( $-0.05$  to  $0.05$ ), slight increase ( $0.05$  to  $0.10$ ), moderate increase ( $0.10$  to  $0.15$ ), substantial increase ( $0.15$  to  $0.20$ ), and major increase ( $> 0.20$ ). The  $\pm 0.05$  threshold excludes typical modelling noise while capturing ecologically meaningful shifts.
5. Suitability transitions: Binary classification of pixels using  $S = 0.5$  as threshold into four categories: (a) remains unsuitable (current < 0.5, future < 0.5), (b) becomes suitable (current < 0.5, future  $\geq 0.5$ ), (c) remains suitable (current  $\geq 0.5$ , future  $\geq 0.5$ ), and (d) becomes unsuitable (current  $\geq 0.5$ , future < 0.5).

## Aggregated multi-species analyses

To identify broad-scale NIS spread patterns, we calculated mean ensemble suitability across all 69 species for each grid cell, producing three analysis types:

1. Priority areas for biosecurity: Mean suitability values across species were classified into management priority categories using 0.1 intervals: negligible ( $S < 0.1$ ), low (0.1–0.2), low-moderate (0.2–0.3), moderate (0.3–0.4), moderate-high (0.4–0.5), high (0.5–0.6), severe ( $> 0.6$ ). Higher values indicate areas where environmental conditions support a greater proportion of the modelled NIS.
2. Scenario-based change maps: For all three SSP scenarios (SSP1-2.6, SSP2-4.5, SSP5-8.5), we calculated change in mean suitability ( $\Delta S = S_{\text{future}} - S_{\text{current}}$ ) and classified values using the same seven categories as individual species maps. We quantified the percentage of the study area falling into each change category for each scenario.
3. Regional analysis by ecoregion: We used Marine Ecoregions of the World (MEOW)<sup>37</sup> to analyse 18 ecoregions within European waters: North and East Barents Sea, Northern Norway and Finnmark, Southern Norway, North and East Iceland, South and West Iceland, Faroe Plateau, North Sea, Celtic Seas, South European Atlantic Shelf, Baltic Sea, Black Sea, Adriatic Sea, Ionian Sea, Aegean Sea, Alboran Sea, Western Mediterranean, Levantine Sea, and Tunisian Plateau/Gulf of Sidra. For each

ecoregion, we extracted pixel values using *terra::extract* and calculated both absolute change ( $\Delta S_{\text{mean}}$ ) and percentage change ( $100 \times \Delta S_{\text{mean}} / S_{\text{current}}$ ) for all three scenarios. Statistics reported as mean  $\pm$  standard deviation across all pixels within each ecoregion. Ecoregions were ordered by centroid latitude for visualisation.

## Ballast water contingency areas (BWCA)

Spatial data on Offshore Wind Farms (OWFs) and Marine Protected Areas (MPAs) were downloaded from EMODnet ([www.emodnet.ec.europa.eu](http://www.emodnet.ec.europa.eu)). Distance to OWFs and MPAs were computed as new rasters, with the same method as the one employed above for distance to land.

To identify potential Ballast Water Contingency Areas (BWCAs), we applied four spatial criteria: (1) low mean habitat suitability across all 69 modelled NIS, (2) minimum distance  $d$  from MPAs, (3) minimum distance  $d$  from OWFs, and (4) minimum distance  $d$  from coastline. Binary raster layers were created for each criterion using user-defined thresholds. For examples shown in Fig. 7, thresholds were selected as: suitability = 0.2,  $d$  = 7 km.

All code used to generate these analyses is fully accessible, allowing users to adjust thresholds, incorporate additional constraints, or tailor the approach to regional management needs and local environmental conditions. See Code Availability Statement.

## Declarations

## Data availability

All species distribution model outputs, including individual species ensemble projections and uncertainty metrics (CoV and CA), are publicly available on Figshare at <https://figshare.com/s/ab27e1dcaee11ba59e88>. The repository includes: (i) habitat suitability maps (GeoTIFF format) for 69 marine non-indigenous species under current conditions and three future climate scenarios (SSP1-2.6, SSP2-4.5, SSP5-8.5 for 2090–2099); (ii) ensemble uncertainty metrics (coefficient of variation and committee averaging maps) for all species and all scenarios; (iii) aggregated mean suitability maps across all species and scenarios; and (iv) processed environmental predictor layers (Bio-ORACLE v.3.0 derived) used for modelling.

Species occurrence data used for model training were obtained from the Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>) and Ocean Biodiversity Information System (OBIS, <https://obis.org/>) and are accessible through these public repositories. Raw GBIF downloads can be accessed via the DOIs found in Supplementary File 4 and are also provided on the Figshare repository alongside the raw OBIS downloads. Final cleaned and spatially thinned occurrence datasets for all 69 modelled species ( $n = 12,300$  occurrence points) are provided in the Figshare repository as CSV files with coordinates.

Model performance metrics (TSS and AUC scores) for all 6,555 individual species distribution models (69 species  $\times$  5 algorithms  $\times$  5 folds  $\times$  3 pseudo-absence datasets, plus fold-averaged and full-dataset

models) are provided as a comprehensive CSV file in the Figshare repository.

ARMS-MBON detection data (2018–2019) are available through GBIF (18S: <https://www.gbif.org/dataset/6f2f07f3-1ef4-4f82-b3d4-d3bd1406650f>; COI: <https://www.gbif.org/dataset/b9afe2d0-b264-4422-bf8c-096b2a53c18f> ). ARMS-MBON detection data (2020–2021) got published in December 2025 (post-modelling) and are available through GBIF (18S: <https://www.gbif.org/dataset/6f2f07f3-1ef4-4f82-b3d4-d3bd1406650f> ; COI: <https://www.gbif.org/dataset/542b31ca-712f-43dd-8f1d-8eeefb88c9a3>).

Swedish Ports Monitoring data (2024) are available through [https://www.gbif.se/ipt/resource?r=nis\\_monitor\\_2024&v=1.1](https://www.gbif.se/ipt/resource?r=nis_monitor_2024&v=1.1). Unpublished ARMS validation data (ARMS-MBON 2020–2021; SPM 2023, 2025) are available in Supplementary File 2.

## Code availability

All scripts for data collection, species distribution modelling, visualisation and analyses are publicly available via GitHub (<https://github.com/JustinePa/ARMS-marine-alien-SDM/tree/main>) and Figshare (<https://figshare.com/s/ab27e1dcaee11ba59e88>).

## Conflict of interest statement

The authors declare no conflict of interest.

## Authors contributions

J.P. and M.O. conceptualised the study and designed the research framework. J.P. performed all data collection, species distribution modelling, spatial analyses, and visualisation. T.A. provided critical feedback on modelling methodology and model evaluation approaches. M.G.A. developed the computational framework for identifying ballast water contingency areas based on model outputs and created Fig. 7. J.P. wrote the original manuscript draft. M.O. and T.A. supervised the research. All authors contributed to manuscript revision and approved the final version.

## Acknowledgements

This work was supported by the SciLifeLab & Wallenberg Data Driven Life Science Program (grant: KAW2024.0159). The project was also supported financially by a grant from the DTO-bioflow project (grant no. 101112823), the EU project MARCO BOLO (grant no. 101082021) and Swedish Biodiversity Data Infrastructure (grant no. 2019 – 00242). Computations and data processing was enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022–06725, enabling computing using the Dardel high-performance computing system at the PDC Center for High

Performance Computing, KTH Royal Institute of Technology, Stockholm, Sweden (<https://www.pdc.kth.se>). We thank the Biodiversity Data Lab at the University of Uppsala for providing feedback during project development, particularly on the statistical modelling approach. Support and feedback were also received from the Swedish Transport Agency, the Swedish Agency for Marine and Water Management, and the OSPAR/HELCOM Joint Task Group on Ballast Water Management Convention (BWMC) and Biofouling.

## References

1. Venegas RM, Acevedo J, Treml EA (2023) Three decades of ocean warming impacts on marine ecosystems: A review and perspective. *Deep Sea Res Part II Top Stud Oceanogr* 212:105318
2. Alter K et al (2024) Hidden impacts of ocean warming and acidification on biological responses of marine animals revealed through meta-analysis. *Nat Commun* 15:2885
3. Wernberg T et al (2025) Marine heatwaves as hot spots of climate change and impacts on biodiversity and ecosystem services. *Nat Rev Biodivers* 1:461–479
4. Shi Y, Li Y (2024) Impacts of ocean acidification on physiology and ecology of marine invertebrates: a comprehensive review. *Aquat Ecol* 58:207–226
5. Röthig T et al (2023) Human-induced salinity changes impact marine organisms and ecosystems. *Glob Change Biol* 29:4731–4749
6. Korpinen S et al (2021) Combined effects of human pressures on Europe's marine ecosystems. *Ambio* 50:1325–1336
7. Katsanevakis S, Zenetos A, Belchior C, Cardoso AC (2013) Invading European Seas: Assessing pathways of introduction of marine aliens. *Ocean Coast Manag* 76:64–74
8. Haubrock PJ et al (2025) The spread of non-native species. *Biol Rev*
9. Vilà M et al (2010) How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. *Front Ecol Environ* 8:135–144
10. IPBES. IPBES Invasive Alien Species Assessment: Summary for Policymakers (2023) 10.5281/zenodo.11254974
11. Seebens H et al (2025) Biological invasions: a global assessment of geographic distributions, long-term trends, and data gaps. *Biol Rev Camb Philos Soc* 100:2542–2583
12. Turbelin AJ et al (2023) Biological invasions are as costly as natural hazards. *Perspect Ecol Conserv* 21:143–150
13. Diagne C et al (2021) High and rising economic costs of biological invasions worldwide. *Nature* 592:571–576
14. Pyšek P et al (2020) Scientists' warning on invasive alien species. *Biol Rev* 95:1511–1534
15. Ahmed DA et al (2022) Managing biological invasions: the cost of inaction. *Biol Invasions* 24:1927–1946

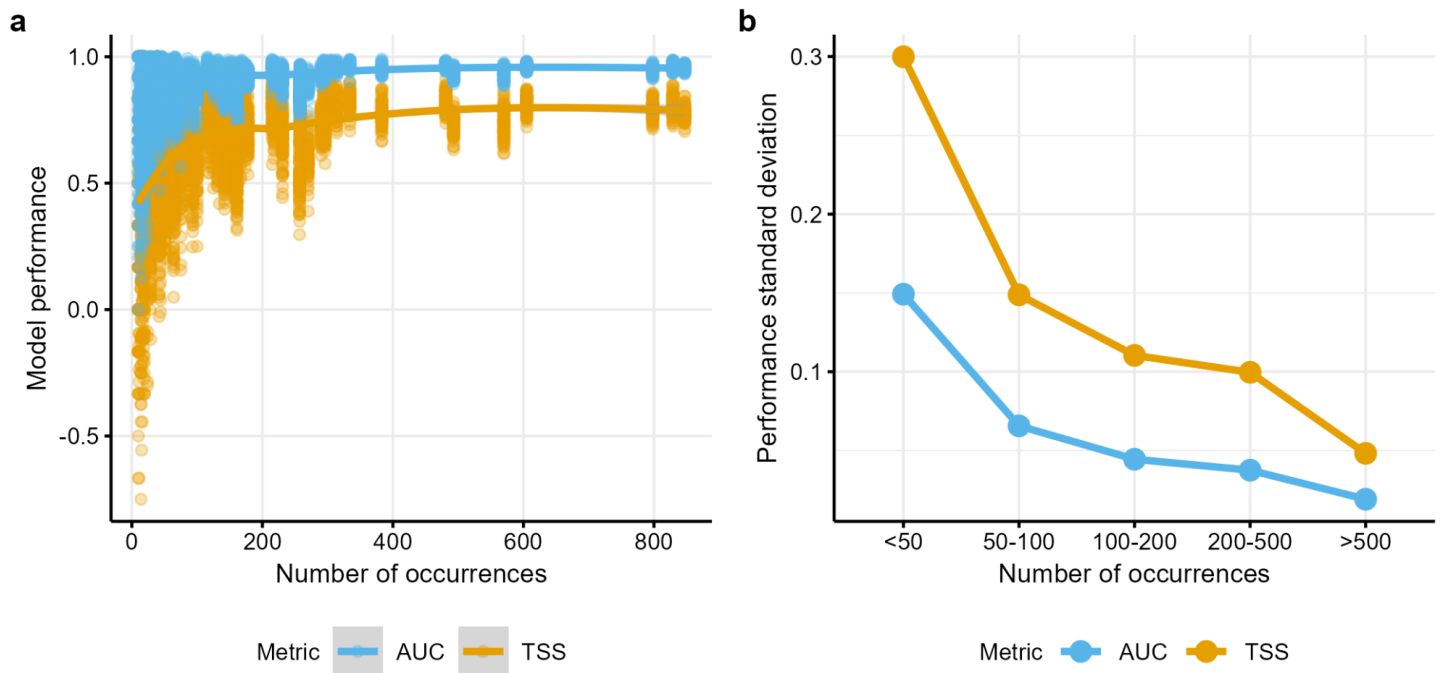
16. Cuthbert RN et al (2022) Biological invasion costs reveal insufficient proactive management worldwide. *Sci Total Environ* 819:153404
17. Obst M et al (2020) A Marine Biodiversity Observation Network for Genetic Monitoring of Hard-Bottom Communities (ARMS-MBON). *Front Mar Sci* 7:572680
18. Daraghmeh N et al (2025) A Long-Term Ecological Research Data Set From the Marine Genetic Monitoring Program ARMS-MBON 2018–2020. *Mol Ecol Resour* 25:e14073
19. Pagnier J, Daraghmeh N, Obst M (2025) Using the long-term genetic monitoring network ARMS-MBON to detect marine non-indigenous species along the European coasts. *Biol Invasions* 27:1–26
20. Piazza A, Mikac B, Colangelo MA, Costantini F (2026) ARMS in ports: monitoring non-indigenous species through Autonomous Reef Monitoring Structures. *Mar Pollut Bull* 222:118545
21. Obst M (2024) National monitoring program for non-indigenous species (NIS), Sweden. <https://doi.org/10.15468/ckcffa> (2025)
22. Sunberg P, Eriksson A-K, Breidenbach M, Panova M, Obst M (2025) *Övervakning av främmande marina arter med eDNA i södra Bohuslän 2024*. <https://www.lansstyrelsen.se/vastra-gotaland/om-oss/vara-tjanster/publikationer/2025/overvakning-av-frammande-marina-arter-med-edna-i-sodra-bohuslan-2024.html>
23. Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135:147–186
24. Elith J, Leathwick JR (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu Rev Ecol Evol Syst* 40:677–697
25. Katsanevakis S et al (2023) Marine invasive alien species in Europe: 9 years after the IAS Regulation. *Front Mar Sci* 10:1271755
26. Yang T, Liu X, Han Z (2022) Predicting the Effects of Climate Change on the Suitable Habitat of Japanese Spanish Mackerel (*Scomberomorus niphonius*) Based on the Species Distribution Model. *Front Mar Sci* 9:927790
27. Sun Y et al (2024) Simulating the changes of the habitats suitability of chub mackerel (*Scomber japonicus*) in the high seas of the North Pacific Ocean using ensemble models under medium to long-term future climate scenarios. *Mar Pollut Bull* 207:116873
28. Gouvêa LP, Krause-Jensen D, Duarte CM, Assis J (2025) Projected impacts of future climate change on the aboveground biomass of seagrasses at global scale. *Sci Total Environ* 966:178680
29. Leitão F, Cánovas F (2025) Predicting climate change impacts on marine fisheries, biodiversity and economy in the Canary/Iberia current upwelling system. *J Environ Manage* 384:125537
30. De Wysiecki A et al (2025) Using global occurrence data to predict suitable habitats for widely distributed marine species in data-scarce regions. *Biodivers Conserv* 34:1497–1523
31. Robinson NM, Nelson WA, Costello MJ, Sutherland JE, Lundquist CJ (2017) A Systematic Review of Marine-Based Species Distribution Models (SDMs) with Recommendations for Best Practice. *Front Mar Sci* 4:421

32. Goldsmit J et al (2020) What and where? Predicting invasion hotspots in the Arctic marine realm. *Glob Change Biol* 26:4752–4771
33. Melo-Merino SM, Reyes-Bonilla H, Lira-Noriega A (2020) Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecol Model* 415:108837
34. Kumschick S et al (2025) Mapping potential environmental impacts of alien species in the face of climate change. *Biol Invasions* 27:43
35. Obst M, Huertas C, De Carlo F, DTO-BioFlow (2025) DUC 1 - Data-driven Strategies for Invasive Species Management. <https://doi.org/10.5281/zenodo.17484964>
36. Assis J et al (2024) Bio-ORACLE v3.0. Pushing marine data layers to the CMIP6 Earth System Models of climate change research. *Glob Ecol Biogeogr* 33:e13813
37. Spalding MD et al (2007) Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *Bioscience* 57:573–583
38. Václavík T, Meentemeyer RK (2012) Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Divers Distrib* 18:73–83
39. Early R, Sax DF (2014) Climatic niche shifts between species' native and naturalized ranges raise concern for ecological forecasts during invasions and climate change. *Glob Ecol Biogeogr* 23:1356–1365
40. Srivastava V, Lafond V, Griess VC (2019) Species distribution models (SDM): applications, benefits and challenges in invasive species management. *CABI Rev* 1–13. 10.1079/PAVSNNR201914020
41. Hedensjö A, Strand Å, Laugen AT (2025) Habitat Preferences at the Leading Edge of a Marine Bioinvasion. *Ecol Evol* 15:e72475
42. Jiménez-Valverde A et al (2011) Use of niche models in invasive species risk assessments. *Biol Invasions* 13:2785–2797
43. Holman LE et al (2019) Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Sci Rep* 9:11559
44. Morissette J et al (2021) Strategic considerations for invasive species managers in the utilization of environmental DNA (eDNA): steps for incorporating this powerful surveillance tool. *Manag Biol Invasions Int J Appl Res Biol Invasions* 12:747–775
45. Sax DF, Schlaepfer MA, Olden JD (2022) Valuing the contributions of non-native species to people and nature. *Trends Ecol Evol* 37:1058–1066
46. Ulman A et al (2019) A Hitchhiker's guide to Mediterranean marina travel for alien species. *J Environ Manage* 241:328–339
47. Katsanevakis S et al (2014) Invading the Mediterranean Sea: biodiversity patterns shaped by human activities. *Front Mar Sci* 1:32
48. Chan FT et al (2019) Climate change opens new frontiers for marine species in the Arctic: Current trends and future invasion risks. *Glob Change Biol* 25:25–38

49. Goldsmit J, McKindsey C, Archambault P, Howland K (2019) L. Ecological risk assessment of predicted marine invasions in the Canadian Arctic. *PLoS ONE* 14:e0211815
50. Aksenov Y et al (2017) On the future navigability of Arctic sea routes: High-resolution projections of the Arctic Ocean and sea ice. *Mar Policy* 75:300–317
51. Bailey SA (2015) An overview of thirty years of research on ballast water as a vector for aquatic invasive species to freshwater and marine environments. *Aquat Ecosyst Health Manag* 18:261–268
52. Seebens H, Schwartz N, Schupp PJ, Blasius B (2016) Predicting the spread of marine species introduced by global shipping. *Proc. Natl. Acad. Sci. U. S. A.* 113, 5646–5651
53. OSPAR. Intra North Sea Ballast Water Contingency and Compliance Area in accordance with BWM.2/Circ.62 and MEPC.387(81) (2025)
54. Burfeind DD, Pitt KA, Connolly RM, Byers JE (2013) Performance of non-native species within marine reserves. *Biol Invasions* 15:17–28
55. Adams TP, Miller RG, Aleynik D, Burrows MT (2014) Offshore marine renewable energy devices as stepping stones across biogeographical boundaries. *J Appl Ecol* 51:330–338
56. Gollasch S, David M (2019) Ballast Water: Problems and Management. *World Seas: An Environmental Evaluation*. Elsevier, pp 237–250. doi:10.1016/B978-0-12-805052-1.00014-0.
57. Seebens H, Gastner MT, Blasius B (2013) The risk of marine bioinvasion caused by global shipping. *Ecol Lett* 16:782–790
58. Blackman R et al (2024) Environmental DNA: The next chapter. *Mol Ecol* 33:e17355
59. Çevik T, Çevik N, Environmental (2025) DNA (eDNA): A review of ecosystem biodiversity detection and applications. *Biodivers Conserv* 34:2999–3035
60. Rius M, Pascual M (2025) Genomics-informed Modelling: Advancing Our Understanding of Non-indigenous Species' Colonization and Spread. in *Invasion Genomics* 162–174
61. Pollock LJ et al (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol Evol* 5:397–406
62. Tikhonov G, Abrego N, Dunson D, Ovaskainen O (2017) Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. *Methods Ecol Evol* 8:443–452
63. Guillera-Aroita G (2017) Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* 40:281–295
64. Hawkins E, Sutton R (2009) The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bull Am Meteorol Soc* 90:1095–1108
65. Frölicher TL, Rodgers KB, Stock CA, Cheung W (2016) W. L. Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors. *Glob Biogeochem Cycles* 30:1224–1243
66. Vucetich JA, Hoy SR, Peterson R (2025) O. More reason for humility in our relationships with ecological communities. *Bioscience* 75:163–171

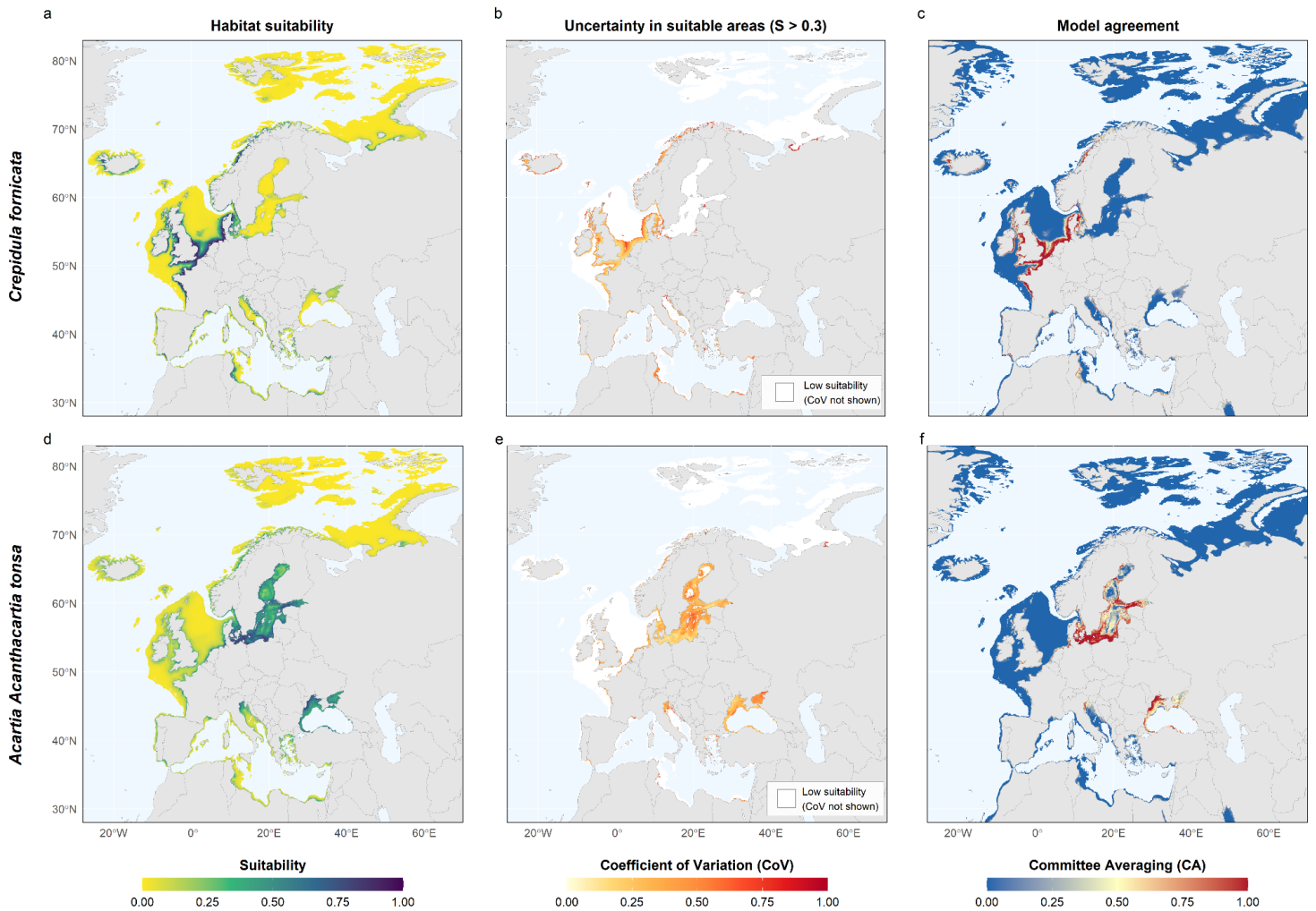
67. Chen EY-S (2021) Often Overlooked: Understanding and Meeting the Current Challenges of Marine Invertebrate Conservation. *Front Mar Sci* 8:690704
68. Zhan A (2025) Multi-Omics-Driven Adaptive Management of Biological Invasions: Toward a Proactive, Predictive, and Integrative Framework. *Biol Divers* 2
69. Pagnier J et al (2025) A long-term ecological research dataset from the marine genetic monitoring programme ARMS-MBON 2020–2021. *Biodivers Data J* 13:e148981
70. Costello MJ et al (2026) World Register of Introduced Marine Species (WRiMS). <https://doi.org/10.14284/347>
71. Daraghmeh N (2024) ARMS-MBON 18S rRNA and COI gene metabarcoding: scanning for non-indigenous species
72. Chamberlain S et al (2024) rgbif: Interface to the Global Biodiversity Information Facility API
73. Provoost P, Bosch S, Appeltans W, OBIS (2022) &. robis: Ocean Biodiversity Information System (OBIS) Client
74. Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38:541–545
75. Thuiller W, Lafourcade B, Engler R, Araújo M (2009) B. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32:369–373
76. Naimi B, usdm (2023) Uncertainty Analysis for Species Distribution Models
77. Massicotte P, South A, Hufkens K (2025) rnaturalearth: World Map Data from Natural Earth
78. Guéguen M, Blancheteau H, Lemaire-Patin R, Thuiller W (2025) Pseudo-absences. [https://biomodhub.github.io/biomod2/articles/vignette\\_pseudoAbsences.html](https://biomodhub.github.io/biomod2/articles/vignette_pseudoAbsences.html)
79. Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol* 22:42–47
80. Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43:1223–1232
81. Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38–49
82. Duarte S, Vieira PE, Lavrador AS, Costa FO (2021) Status and prospects of marine NIS detection and monitoring through (e)DNA metabarcoding. *Sci Total Environ* 751:141729
83. Hijmans RJ et al (2025) terra: Spatial Data Analysis
84. Hijmans RJ et al (2025) raster: Geographic Data Analysis and Modeling
85. Wickham H et al (2023) dplyr: A Grammar of Data Manipulation
86. Wickham H et al (2025) ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics

## Figures



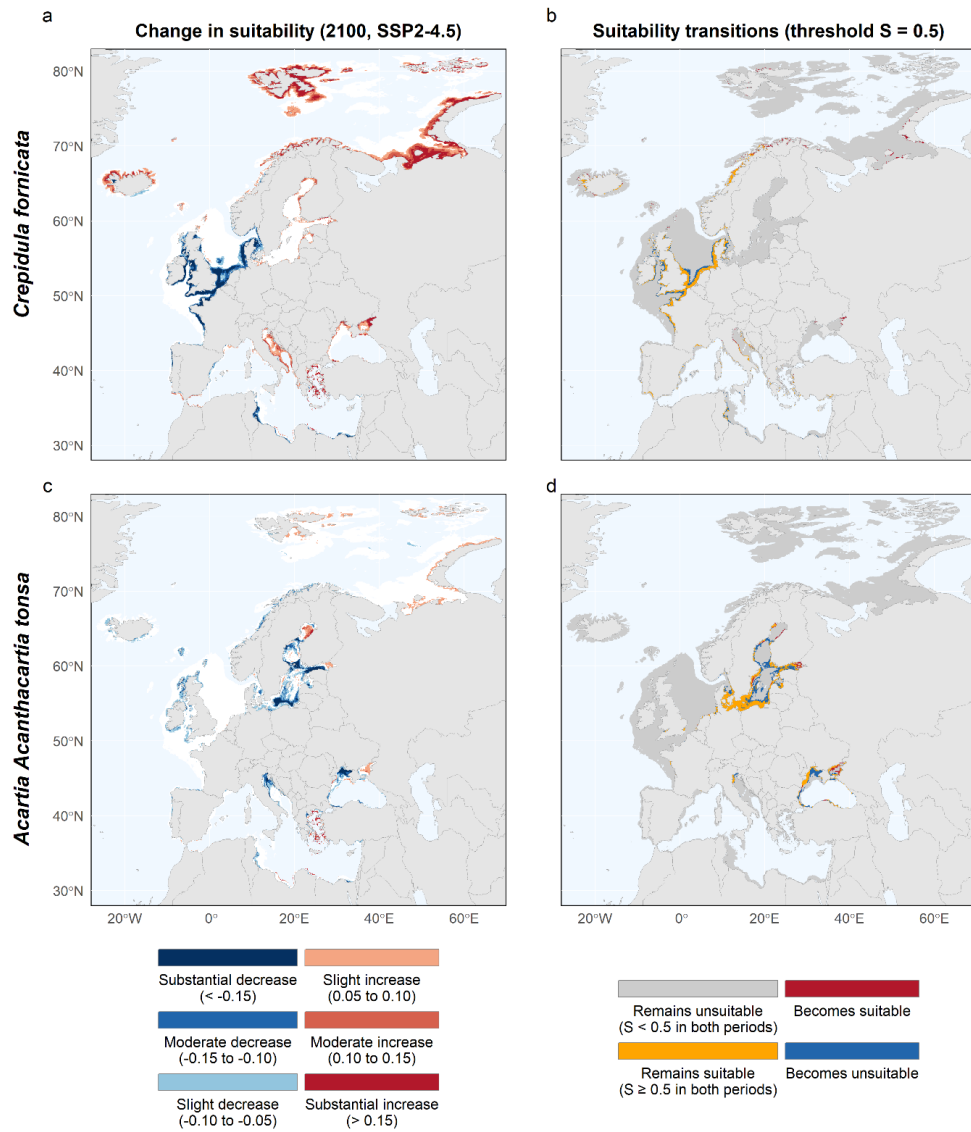
**Figure 1**

**Models' performance metrics across varying sample sizes. (a)** Relationship between model performance (AUC and TSS) and number of occurrence records for individual species ( $n = 69$ ). Each point represents a single SDM. LOESS smoothing was applied to visualise performance trends across the sample size gradient. **(b)** Standard deviation of model performance across binned occurrence categories, as a measure of performance consistency.



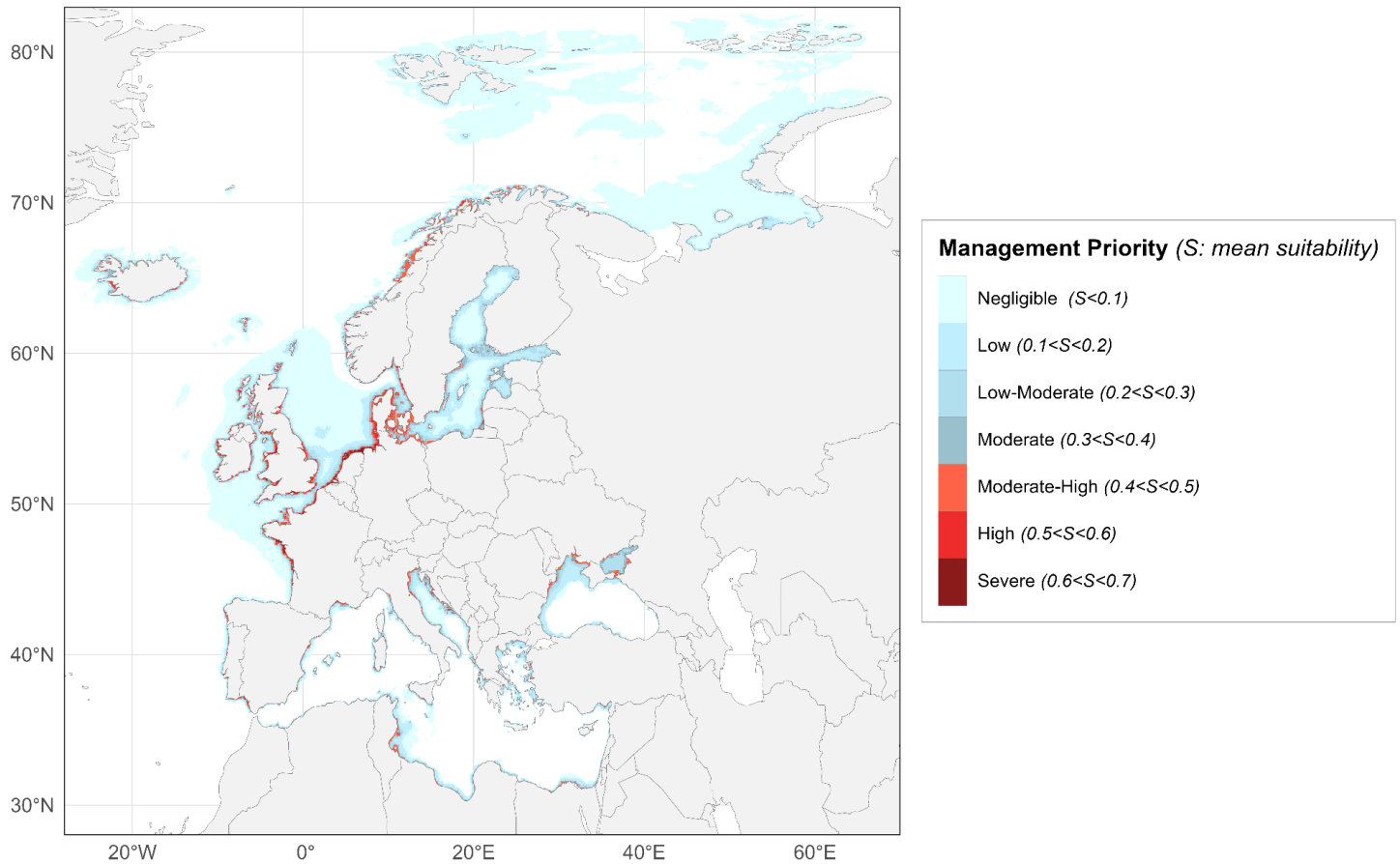
**Figure 2**

**Habitat suitability, coefficients of variation (CoV), and model agreement (CA) for two non-indigenous species across European seas.** Panels show ensemble model predictions for *Crepidula fornicata* (a-c) and *Acartia (Acanthacartia) tonsa* (d-f). (a, d) Habitat suitability from ensemble models, with values ranging from 0 (unsuitable, yellow) to 1 (highly suitable, dark blue). (b, e) Coefficient of variation (CoV) displaying prediction uncertainty only in areas with suitability > 0.3, with white areas indicating low suitability where CoV is not shown. Lower CoV values (light colors) indicate higher model consensus, while higher CoV values (dark red) indicate greater disagreement among algorithms. (c, f) Committee averaging showing the proportion of models agreeing on habitat suitability, with blue indicating high agreement among models and red indicating low agreement.



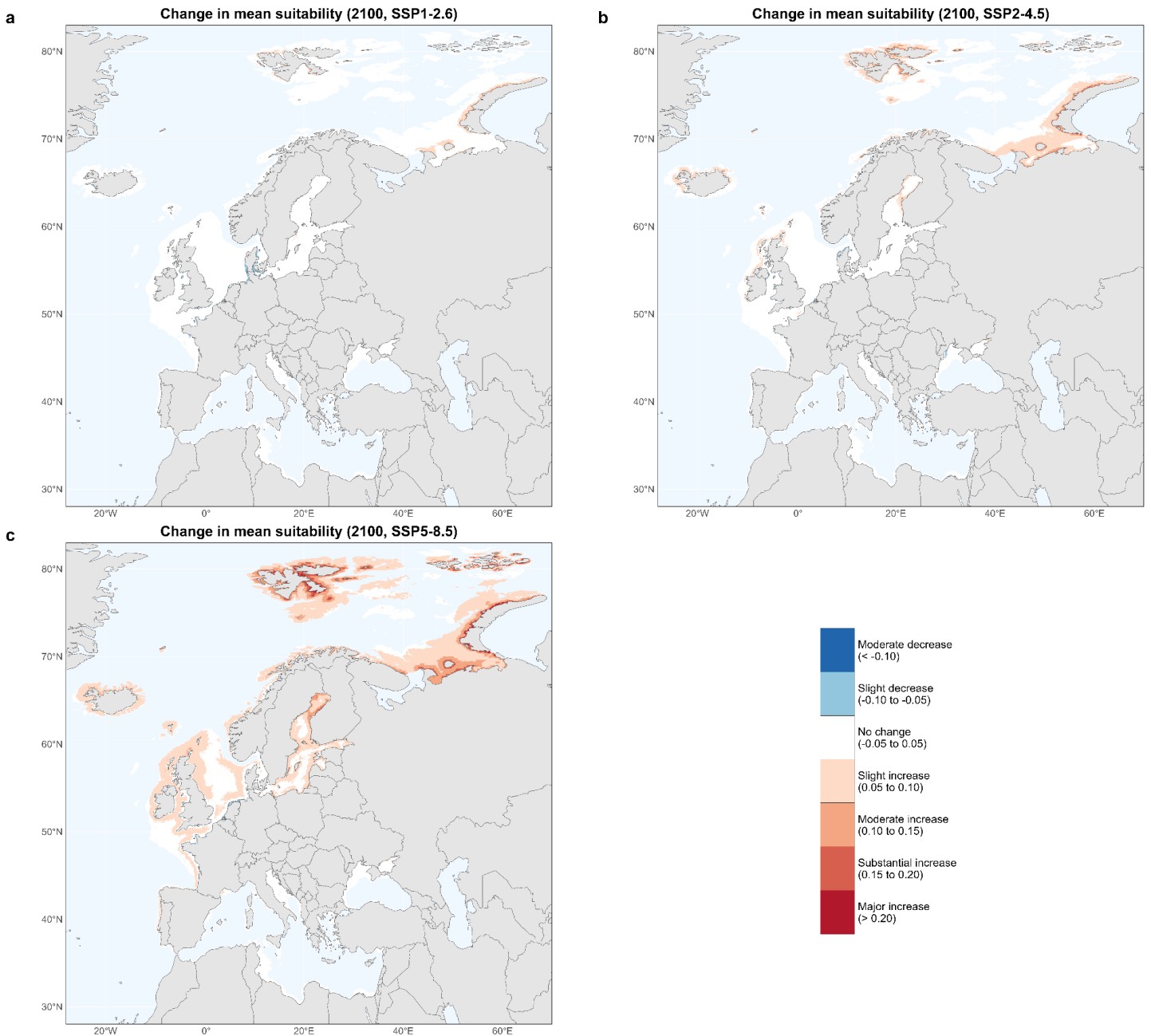
**Figure 3**

**Projected changes in habitat suitability and transitions for two marine non-indigenous species in European waters under SSP2-4.5 climate scenario by 2100.** (a, c) Projected change in habitat suitability between present conditions and 2100 for *Crepidula fornicata* (a) and *Acartia (Acanthacartia) tonsa* (c). Blue colours indicate decreasing suitability, while red/orange colours indicate increasing suitability. (b, d) Suitability transitions based on a threshold of  $S = 0.5$  for *C. fornicata* (b) and *A. tonsa* (d). Grey areas remain unsuitable in both periods ( $S < 0.5$ ), red areas become newly suitable, orange areas persist as suitable, and blue areas become unsuitable.



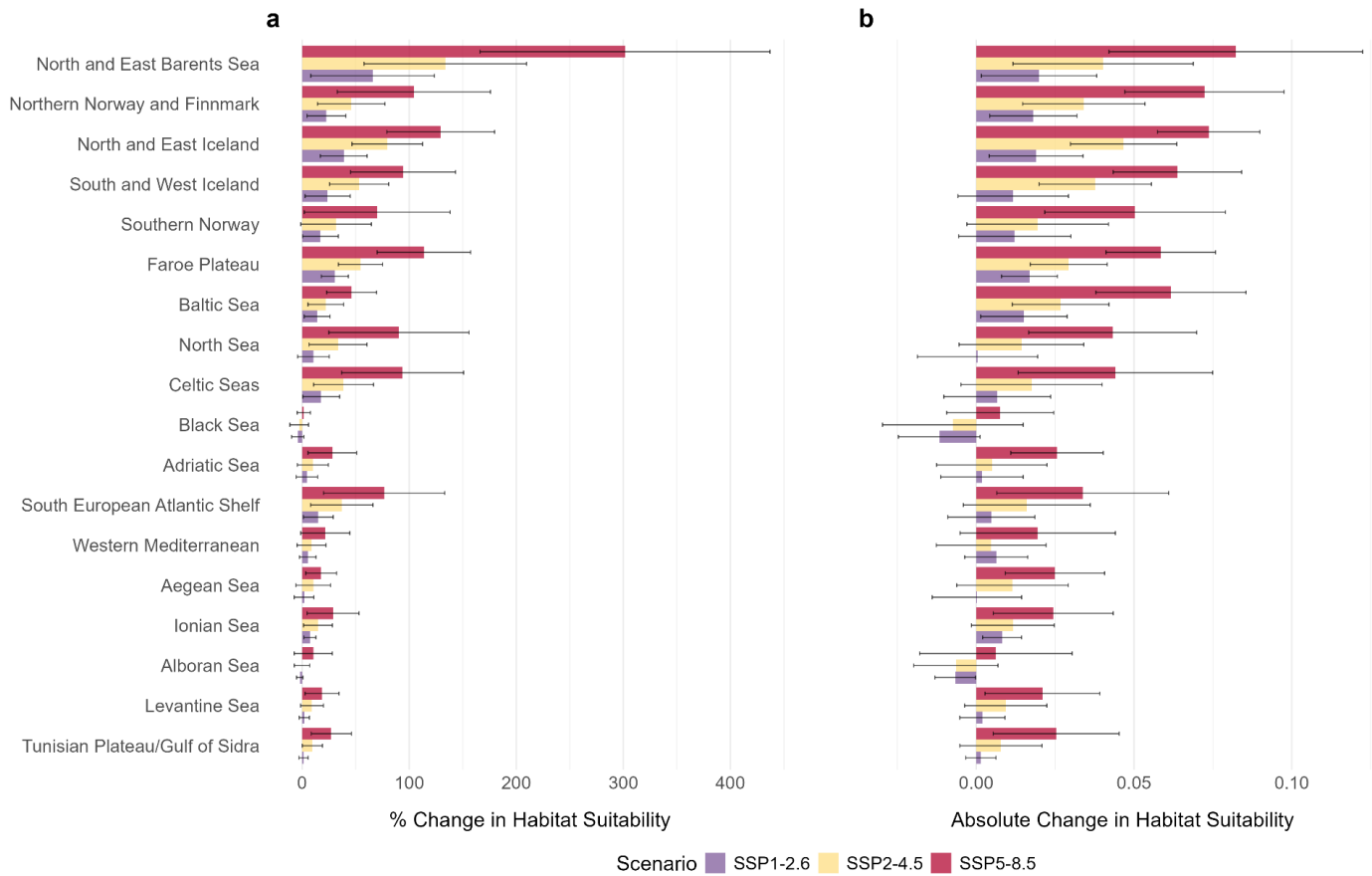
**Figure 4**

**Priority areas for non-indigenous species management in European seas under current climate conditions.** This map shows classified mean habitat suitability based on aggregated ensemble predictions from 69 marine non-indigenous species (NIS). Colors represent management priority levels derived from average ensemble suitability values ( $S$ ), ranging from Negligible ( $S < 0.1$ , light blue) to Severe ( $S \geq 0.6$ , dark red).



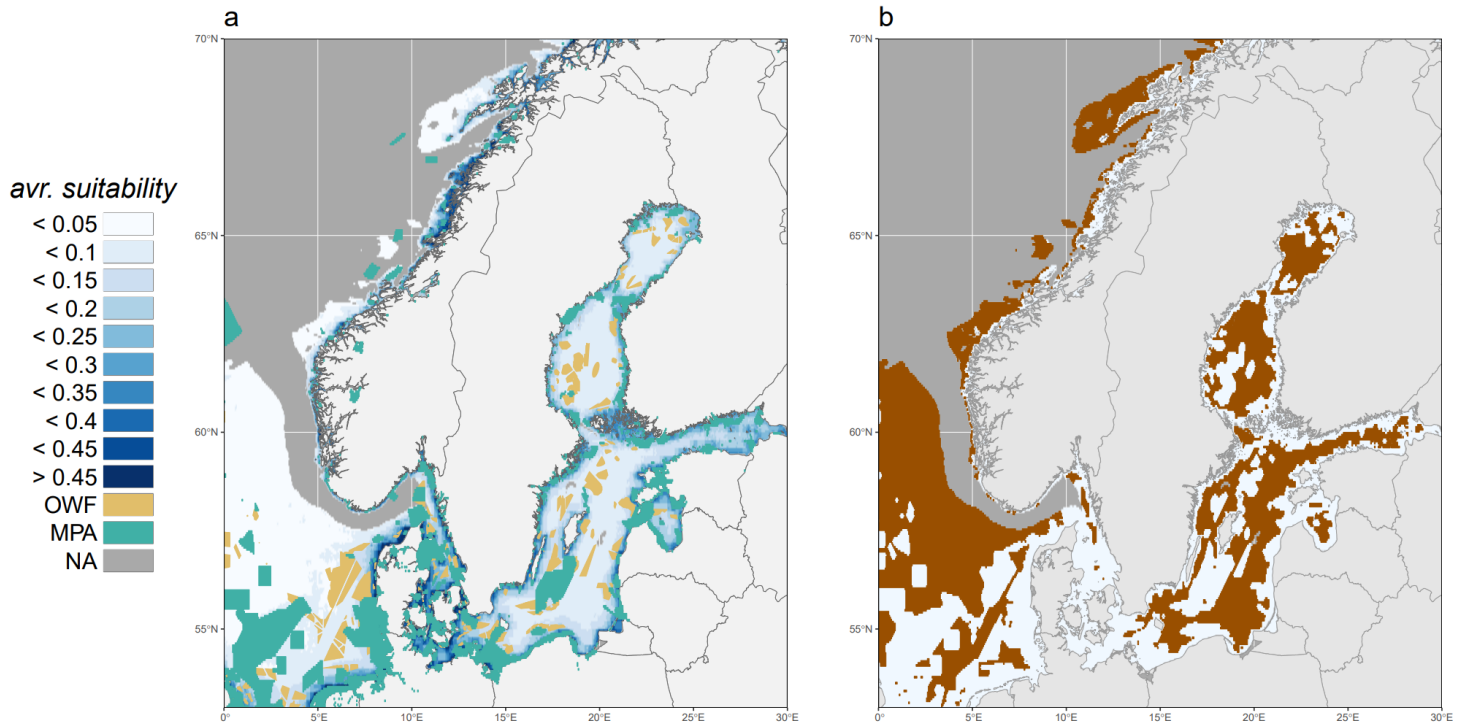
**Figure 5**

**Projected changes in mean habitat suitability (69 NIS) in European seas.** Projected changes in mean habitat suitability by 2100 under three climate scenarios: (a) SSP1-2.6, (b) SSP2-4.5, and (c) SSP5-8.5. Changes are calculated as the difference between future and current mean suitability (positive values indicate increased suitability, negative values indicate decreased suitability). Colors represent magnitude of change, classified as: moderate decrease ( $< -0.10$ ), slight decrease ( $-0.10$  to  $-0.05$ ), no change ( $-0.05$  to  $0.05$ ), slight increase ( $0.05$  to  $0.10$ ), moderate increase ( $0.10$  to  $0.15$ ), substantial increase ( $0.15$  to  $0.20$ ), and major increase ( $> 0.20$ ).



**Figure 6**

**Regional changes in mean habitat suitability (69 marine NIS) under three climate change scenarios. (a)** Relative change in mean habitat suitability from current to 2100 across three emission scenarios (SSP1-2.6, SSP2-4.5, SSP5-8.5). **(b)** Absolute change in mean habitat suitability for the same regions and scenarios. Error bars represent spatial variability (standard deviation) within each ecoregion. Ecoregions are ordered by latitude from north (top) to south (bottom).



**Figure 7**

**Identification of ballast water contingency areas based on marine alien species' habitat suitability and potential stepping stones (Marine Protected Areas and Offshore WindFarms).** (a) Overlay of current mean habitat suitability for non-indigenous species with offshore wind farms (OWF, yellow) and Marine Protected Areas (MPA, teal). (b) Proposed ballast water contingency areas (brown) corresponding to identified cold spots where low habitat suitability for alien species suggests reduced ecological risk for ballast water discharge or exchange operations.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile1Specieslistanddetails.xlsx](#)
- [SupplementaryFile2Externalvalidationdataandresults.xlsx](#)
- [SupplementaryFile3Futurechangesperecoregions.xlsx](#)
- [SupplementaryFile5MarineEcoregionsperspecies.xlsx](#)
- [PagnieretalSupplementaryInformation.docx](#)
- [SupplementaryFile4TrackingofoccurrencesPAnumberpremodelling.xlsx](#)