

Identifying Key Processes and Drivers Affecting the Success of Non-indigenous Marine Species in Coastal Waters

Martin Lindegren (✉ mli@aqua.dtu.dk)

DTU Aqua: Danmarks Tekniske Universitet Institut for Akvatiske Ressourcer <https://orcid.org/0000-0002-9185-951X>

Aurelia Pereira Gabellini

DTU Aqua: Danmarks Tekniske Universitet Institut for Akvatiske Ressourcer

Peter Munk

DTU Aqua: Danmarks Tekniske Universitet Institut for Akvatiske Ressourcer

Karen Edelvang

DTU Aqua: Danmarks Tekniske Universitet Institut for Akvatiske Ressourcer

Flemming Hansen

DTU Aqua: Danmarks Tekniske Universitet Institut for Akvatiske Ressourcer

Research Article

Keywords: community assembly, environmental filtering, drift modelling, species distribution modelling, marine spatial planning

Posted Date: January 12th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1135935/v1>

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

[Read Full License](#)

2nd December 2021

Please find enclosed our manuscripts entitled **“Identifying key processes and drivers affecting the success of non-indigenous marine species in coastal waters”** which we would like to have evaluated for publication in *Biological Invasions* as a Research Paper.

Despite the rapid spread of non-indigenous species (NIS) in coastal waters worldwide, biotic invasions are widely disregarded in marine conservation planning. To guide conservation actions, a better understanding of the underlying mechanisms and drivers determining the success of NIS are therefore needed. In this study, we develop and apply a joint modelling approach to identify the key drivers and community assembly processes determining the occurrence of invasive benthic invertebrates, using Danish coastal waters as a case study. To reflect factors affecting the introduction, establishment and spread of NIS throughout the area, we compiled long-term monitoring data on NIS, as well as information on commercial shipping, environmental conditions and estimates of larvae settling densities derived from advanced drift model simulations (informed by species traits). We then applied a set of species distribution models to identify the key drivers determining the occurrence of NIS. Our results demonstrate a significant positive effect of vessel activity, a negative effect of depth and bottom salinity, as well as a positive effect of the simulated settling densities on the probability of presence. Taken together, our results highlight the role of commercial shipping, habitat characteristics and passive advection of early-life stages on the success of NIS. Our joint modelling approach provide improved process understanding on the key community assembly processes determining the presence of NIS and may serve to guide monitoring, management and conservation planning in order to limit future invasions and their negative consequences on coastal ecosystems.

Yours sincerely,



Dr. Martin Lindegren (Senior scientist)

Centre for Ocean Life, National Institute of Aquatic Resources, Technical University of Denmark,

E-mail: mli@aqua.dtu.dk

1 **Identifying key processes and drivers affecting the success of non-indigenous marine species in**
2 **coastal waters**

3 Martin Lindegren^{1,2*}, Aurelia Pereira Gabellini², Peter Munk², Karen Edelvang², Flemming Hansen²

4 *¹Centre for Ocean Life, National Institute of Aquatic Resources, Technical University of Denmark,*
5 *Kemitorvet, Building 202, 2800 Kgs. Lyngby, Denmark. ²National Institute of Aquatic Resources,*
6 *Technical University of Denmark, Kemitorvet, Building 202, 2800 Kgs. Lyngby, Denmark.*

7 *Corresponding author: mli@aqua.dtu.dk ORCID: 0000-0002-9185-951X

8 **Abstract**

9 Non-indigenous species (NIS) pose a major threat to biodiversity and the functioning and services of
10 ecosystems. Despite their rapid spread in coastal waters worldwide, biotic invasions are widely
11 disregarded in marine conservation planning. To guide conservation actions, a better understanding of
12 the underlying mechanisms determining the success of NIS are therefore needed. Here we develop a
13 joint modelling approach to identify the key drivers and community assembly processes determining
14 the occurrence of invasive benthic invertebrates, using Danish coastal waters as a case study. To
15 reflect factors affecting the introduction, establishment and spread of NIS throughout the area, we
16 compiled long-term monitoring data on NIS, as well as information on commercial shipping,
17 environmental conditions and estimates of larvae settling densities derived from drift model
18 simulations informed by species traits. We then applied a set of species distribution models to identify
19 the key drivers determining the occurrence of NIS. Our results demonstrate a significant positive
20 effect of vessel activity, a negative effect of depth and bottom salinity, as well as a positive effect of
21 the simulated settling densities on the probability of presence. Taken together, our results highlight
22 the role of commercial shipping, habitat characteristics and passive advection of early-life stages on
23 the success of NIS. Our joint modelling approach provide improved process understanding on the key
24 community assembly processes determining the presence of NIS and may serve to guide monitoring,
25 management and conservation planning in order to limit future invasions and their negative
26 consequences on coastal ecosystems.

27 **Key words:** community assembly, environmental filtering, drift modelling, species distribution
28 modelling, marine spatial planning

29

30 **Acknowledgements**

31 This project was funded by the Danish Maritime Fund by request of the Danish Environmental
32 Protection Agency (DEPA). We wish to thank all people involved in the NOVANA program for
33 ensuring availability and accessibility to high-quality monitoring data.

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48 **Introduction**

49 Non-indigenous species (NIS) are considered one of the major threats to biodiversity and the
50 functioning and services of ecosystems worldwide (Vitousek et al., 1996; Mack et al., 2000;
51 Ehrenfeld, 2010; IPBES 2019). In order to mitigate their impacts on native species and communities,
52 there is an urgent need to adopt management strategies capable of accounting for NIS. Despite their
53 presence and negative effects on many aquatic ecosystems (Pimentel et al., 2005; Lindegren et al.,
54 2012a; Albins, 2013; Katsanevakis et al., 2014), biotic invasions are widely disregarded in marine
55 conservation planning (Giakoumi et al., 2016). Hence, the ambitious policy goals to prevent the
56 introduction and spread of NIS (EC 2008; CBD 2021; Tittensor et al., 2014) are far from being
57 achieved. To guide conservation actions, a better understanding of the underlying mechanisms
58 controlling the introduction, establishment and spread of NIS (Theoharides & Dukes, 2007;
59 Richardson & Pyšek, 2012), as well as tools to model and predict such invasions, are therefore needed
60 (Gallien & Carboni, 2017).

61

62 The underlying mechanisms determining the presence and dynamics of NIS can conceptually be
63 described as a multi-stage process of invasions comprised of: (i) the transport and introduction of
64 organisms to a novel habitat from their native range; (ii) the establishment and growth of self-
65 sustaining populations within the new environment, (iii) and the secondary spread of the organism
66 across the land/seascape (Kolar & Lodge, 2001; Beletsky et al., 2017). The first step described above
67 involves both a transport mechanism (i.e., vector) and a geographic path (i.e., route) along which an
68 NIS is carried to a new location (Carlton & Ruiz 2005; Clarke Murray et al., 2014; Galil et al., 2014).
69 In marine and freshwater environments, the main vectors and routes involve primarily aquaculture
70 and commercial shipping through ballast water and hull fouling (Leppäkoski et al., 2002; Drake &
71 Lodge, 2007; Ojaveer et al., 2010; Bailey, 2015). Once transported to a new location, the subsequent
72 steps of invasions can be thought of as a series of “filters” or community assembly processes (Keddy,
73 1992) influencing the success of NIS outside their native range (Theoharides & Dukes, 2007;

74 Richardson & Pyšek, 2012; Gallien & Carboni, 2017). Firstly, the success of NIS is determined by the
75 organism's tolerance to the environmental conditions in a new location. This assembly processes,
76 often termed *environmental filtering*, is largely conditioned on the particular characteristics (i.e.,
77 traits) of the organism in question and the degree to which its resulting environmental niche overlaps
78 with the abiotic conditions in the new location (Ackerly, 2011; Gallien et al., 2014; Pecuchet et al.,
79 2016; Beukhof et al., 2019a). Secondly, the degree to which NIS are able to grow and reproduce is
80 conditioned by its competitive abilities to acquire space and/or resources relative to ecologically
81 similar species within the recipient community and food web (Stubbs & Wilson, 2004; Shinen &
82 Morgan, 2009; Gallien et al., 2015; David et al., 2017). This assembly process, often termed *biotic*
83 *filtering* or *limiting similarity*, is also determined by the traits of organisms, such as their feeding
84 rates, diet preferences and/or natural mortality imposed by native predators.

85

86 Finally, once established as a self-sustaining population, a range of natural or anthropogenic factors
87 may determine the subsequent spread of NIS across the land/seascape. This final assembly process,
88 termed *dispersal filtering*, is determined by both the physical environment (i.e., establishment of
89 pathways or barriers for dispersal), as well as the movement of organisms, either through active
90 locomotion or passive drift (Gallien et al., 2015; Beletsky et al., 2017; Li & Shen, 2020). In the latter
91 case, movement is primarily through early-life stages such as seeds, eggs and larvae. As evident from
92 the steps outlined above, the presence of NIS in any given location is jointly determined by multiple,
93 non-exclusive assembly processes and mechanisms, both external and internal to the organism in
94 question (Richardson & Pyšek, 2012). This complicated interplay of governing factors and processes
95 means that it is challenging to understand and project the success and spread of NIS in any given area
96 (Clarke Murray et al., 2014), particularly in open marine environments. To understand the key drivers
97 and underlying mechanisms involved in shaping the success and dynamics of NIS, a holistic approach
98 is needed that accounts for the effect of multiple interacting factors determining their transport,
99 establishment and spread.

100 In this study, we develop and apply such a holistic approach to study the underlying mechanisms
101 explaining the success of marine NIS, using Danish coastal waters as a case study. We focus primarily
102 on benthic invertebrates with no or very limited locomotion as adults (e.g., bivalves, crustaceans and
103 annelids), but with highly effective dispersal during their planktonic egg or larvae stages (Pechenik,
104 1999). To do so, we compiled a comprehensive set of long-term monitoring data on NIS in the coastal
105 waters of the southeast North Sea, Kattegat and Western Baltic Sea (Fig. 1); a region highly affected
106 by NIS (Leppäkoski et al., 2002; Ojaveer et al., 2010). We complemented these observations with
107 data on commercial shipping, environmental conditions and estimates of larvae dispersal based on
108 advanced drift model simulations. Together, these variables reflect a potential set of drivers affecting
109 the introduction, establishment and spread of NIS across the study area, respectively. Using this data
110 as joint input, we then applied a set of comparable species distribution modelling (SDM) methods to
111 identify the key predictors and underlying community assembly processes determining the occurrence
112 of NIS. Finally, we used a set of parameterized models to predict the probability of occurrence of NIS
113 throughout the entire study area with the aim of providing managers with key information needed to
114 prioritize monitoring and conservation actions, as well as develop marine spatial management plans
115 accounting also for NIS and their potential impacts on marine ecosystems.

116

117 **Material and Methods**

118 *Monitoring data and standardization*

119 Based on a list of marine species considered as non-indigenous within the study area (Danish EPA,
120 personal communication, 3rd July, 2020) we extracted the taxonomic information of all benthic
121 invertebrates containing a benthic adult stage and a pelagic egg or larvae stage (Table S1).
122 Subsequently, we extracted all records of NIS from the long-term Danish National Monitoring
123 programme (NOVANA) hosted by the Danish Centre for Environment and Energy
124 (<https://odaforalle.au.dk/main.aspx>, Accessed 15th of October 2020) (Table 1). The NOVANA
125 program and its precursors date back to 1988 and include a set of frequently (i.e., monthly or

126 quarterly) sampled sites originally selected to assess the effect of anthropogenic impacts such as
127 nutrient enrichment in Danish waters (Svendsen et al., 2005). At each station, five replicate sediment
128 samples are typically collected for faunal analysis. The samples are rinsed over a 1mm sieve,
129 preserved in 96% ethanol solution and subsequently sorted into individuals, identified to the lowest
130 taxonomic level possible. To avoid bias introduced by differences in sampling protocols and gears, we
131 included only records obtained using the standard 0.0143m² “HAPS” corer, the most commonly
132 applied sampling method of benthic invertebrates across all years and sites within the monitoring
133 program. Furthermore, we accounted for the marked differences in sampling effort over time and
134 between sites (Fig. 1), which may bias the detection and probability of presence of NIS at any given
135 site, by performing a formal data standardization. This was achieved by randomly bootstrapping (i.e.,
136 resampling without replacement) the same minimum number of unique samples from each site and
137 season (i.e., aggregating months by quarters) over 500 random iterations. Two stations
138 (“Hirsholmene” and “Knud001”) were excluded due to insufficient sampling coverage across years
139 and seasons. We then extracted the mean abundances and their standard deviations for all species,
140 sites, years and seasons across the entire set of randomized draws and used this standardized metric in
141 the following statistical analysis. Furthermore, we restricted the analysis to the years from 2000
142 onwards to allow for a sufficient coverage of sampling across stations and ensure a temporal match
143 with the drift model simulations. After omitting sporadic records based on other gear types and
144 observations prior to 2000, the standardized data set contained 17,769 mean abundances (or
145 presence/absences) of 11 NIS covering in total 66 sites.

146

147 *Transport vectors and habitat characteristics*

148 To represent the role of commercial shipping, which is considered to be one of the key transport
149 vectors introducing marine NIS into the study area (Leppäkoski et al., 2002; Ojaveer et al., 2010), we
150 collected available data on ship activity from EMODNET ([https://www.emodnet-](https://www.emodnet-humanactivities.eu/view-data.php)
151 [humanactivities.eu/view-data.php](https://www.emodnet-humanactivities.eu/view-data.php)). The data is based on all available vessel positions retrieved from

152 the Automatic Identification System (AIS) in 2019, and is expressed as hours per km² (Fig. 2a). To
153 represent key habitat characteristics allowing establishment of NIS, we also extracted available data
154 on depth, temperature and salinity (Fig. 2b-d). Depth data were obtained from in situ measurements,
155 supplemented with information from EMODNET if missing
156 (<https://www.emodnet.eu/en/bathymetry>). Temperature (surface and bottom) and salinity (surface)
157 were obtained from the Global Ocean Physics Reanalysis with (GLORYS2v4) and were downloaded
158 from the Copernicus Marine Environment Monitoring Service (<http://marine.copernicus.eu/>). In order
159 to fit the spatio-temporal coverage of the observational data and the drift model simulations,
160 temperature and salinity values were given at each of the sampling station as monthly (or seasonal)
161 means from 2000 onwards.

162

163 *Drift model simulations of larvae dispersal*

164 To represent the dispersal phase of NIS (once having been introduced and established) we simulated
165 larval dispersal using the agent-based modelling system IBMlib (Christensen, 2008; Christensen et al.
166 2018). IBMlib is a model library for individual-based modelling using a Lagrangian approach that has
167 been specifically developed for simulating larval dispersal including biological processes affecting
168 spawning, drift and settlement of marine larvae. We ran simulations for each of the NIS for the years
169 2001-2010, using a new modified version of the IBMlib software to include the mapping of the
170 density of agent trajectories during the simulation. The model was forced by hydrodynamic data,
171 including current velocities, salinity and temperature covering the entire study area, derived from a
172 regionally downscaled hydrodynamic model of the North Sea-Baltic Sea (Berg & Poulsen 2012). The
173 spatial resolution of the model is 0.5 nm in Kattegat, inner Danish straits and the western part of the
174 Baltic Sea, and 5 nm outside this region. The vertical resolution of the model is 50 and 52 layers,
175 respectively.

176

177 To identify a set of credible starting locations for releasing agents (i.e., larvae), as well as the timing
178 of release and duration of simulations, available information on species traits reflecting their
179 spawning period, pelagic larvae duration, substrate required to settle, preferred depth, as well as the
180 temperature and salinity tolerances for adults and larvae were collected for each NIS (Table S1).
181 Please note that most of the traits data refer to studies from other parts of the world. These could be
182 from the native range of the species or determined on the basis of experiments. While spawning
183 period and length of the pelagic larvae phase serve to constrain the release and duration of agents in
184 the simulations, the remaining traits were used to create a set of species habitat maps following the
185 approach by Hansen et al. (2020). These maps were used to constrain the seeding of agents to
186 locations fitting some of the fundamental requirements of each species with respect to their seabed
187 substrate, water depth and adult salinity tolerance (Supplementary text; Table S1, S2).

188

189 Taking these aspects into account, the agents in each simulation were distributed randomly in space
190 within the maximum area coverage of the species habitat maps and randomly in time within the
191 spawning period of each species. For each NIS, we released 200.000 agents per year from 2001 to
192 2010 (i.e. a total of 2 million agents per species). The mapping of the dispersal trajectories of agents
193 was carried out by registering the presence of each agent at each time step in a regular grid (280 x
194 120) defined by 54°N to 60°N and 3°E to 17°E. All trajectories experiencing salinity conditions
195 outside the larval salinity tolerance range reported for each species were ignored and not included in
196 the mapping. Finally, we produced maps representing all end positions of successfully settled larvae
197 (here termed “settlement densities”), taking into consideration not only currents, but the salinity range
198 experienced in the trajectory, as well as the reported habitat requirements (depth and substrate type) at
199 the time of larvae settlement. The maps were created by summing all the individual track counts
200 across years, then dividing by the area of each grid cell. Furthermore, we normalized the densities by
201 the maximum values observed for each species, so that the values range between 0 and 1. In order to
202 fit the spatial coverage of the observational data on NIS, we estimated the average densities of settled
203 larvae for each of the sampling station in the NOVANA data.

204 *Statistical analysis of underlying assembly processes acting on NIS*

205 To assess the effect and relative importance of factors explaining the overall presence and absence of
206 the set of NIS considered, we applied Generalized Additive Models (GAMs) and random forests
207 (RFs). These methods represent two complementary tools commonly used in species distribution
208 modelling (Elith & Leathwick, 2009; Goldsmit et al., 2018; Lindegren et al., 2020). The same set of
209 explanatory variables, reflecting factors related to the introduction, establishment and spread of NIS
210 were used as covariates. The resulting statistical relationships and derived response curves between
211 the set of covariates and the presence and absence of NIS were compared across methods to assess the
212 sensitivity and robustness of the results to the choice of methodology. In terms of GAM, we used the
213 following binomial model formulation:

$$214 \text{PA}_{s,l,y,q} = a + s(\text{spn}_{s,l}) + s(\text{ship}_l) + s(\text{tempS}_{l,y,q}) + s(\text{tempB}_{l,y,q}) + s(\text{sal}_{l,d,q}) + s(\text{depth}_l) + 5y + S + e$$

215 where the response variable PA is the presence/absence (0,1) of each species s , at site l , in year y and
216 season q as a function (using a logit link) of the simulated settling density (spn), ship activity level
217 (ship), surface and bottom temperature (temS, temB), salinity (sal) and depth at each site. To account
218 for mean differences in the probability of presence between species and over time we included species
219 identity (S) and years classified into 5-year time periods (5y) as fixed effect factors. The constant a is
220 the overall intercept, s the thin plate smoothing function for each smooth term and ε the error term.
221 Although the number of regression splines is optimized (and penalized) by the generalized cross
222 validation criterion (GCV), the degrees of freedom of the spline smoother function (s) was further
223 constrained to three knots ($k=3$) to allow for potential nonlinearities, but restrict flexibility during model
224 fitting. Finally, we applied backwards model selection to identify the best possible set of predictors.

225

226 The second method used, i.e., RF, is a machine learning tool comprising ensembles of decision trees
227 that rely on bagging (i.e. bootstrap aggregation). RFs are capable of reproducing complex nonlinear
228 shapes in single and multiple dimensions, making them suitable for ecological applications in which
229 complex shapes are to be expected (Breiman 2001). In addition, RF has fewer constraints and is able to

230 capture interactions between variables that cannot easily be achieved with GAMs. Individual
231 classification trees within the random forest are trained on randomly selected subsets of the data. The
232 final forest prediction is obtained by averaging predictions across all trees in the forest. We used the
233 same model setup in terms of response and explanatory variables as in the GAM. In order to account
234 for differences in the number of observations reporting absences versus presences of NIS, we accounted
235 for such unbalanced classes by assigning different weights to each class (using the “classwt” option).
236 Once trained on the available data, we used the final RFs (based on 10 000 individual trees) to estimate
237 the relative importance of each predictor, as well as visualize the partial response curves of each
238 individual explanatory variable. Finally, we used the final models to estimate the overall probability of
239 occurrence of NIS across the entire study area, beyond the NOVANA monitoring stations used for
240 model fitting and training. This was achieved by predicting occurrences based on a compilation of all
241 covariate values in each grid cell, i.e., based on both the collected ship activity, depth, temperature and
242 salinity data (Fig. 2), as well as the drift model outputs of species-specific settling densities (Fig. S2).
243 The model predictions serves to highlight areas highly vulnerable to NIS and in need of monitoring and
244 conservation actions. All statistical analyses were conducted using the R software, version 4.0.2 (R core
245 Team 2020) using the following packages: “mgcv” (Wood 2017) and “randomForest” (Liaw and
246 Wiener 2002).

247

248 **Results**

249 *Spatial patterns of NIS and environmental conditions*

250 The overall occurrences and abundances of the set of NIS demonstrated pronounced spatial
251 heterogeneity across sites throughout the study area. This is illustrated by marked spatial differences
252 in the number of NIS that are frequently occurring at each station based on our bootstrapping (Fig.
253 3a). Notably, the highest richness of NIS occurred in nearshore coastal areas, particularly in the
254 southern part of the area (i.e., around the densely populated islands of Zealand and Fyn and in the
255 Wadden Sea), while the lowest richness (or even absence) of NIS was found in more offshore areas in

256 Kattegat and the North Sea. A similar spatial pattern was also evident in the sum of median
257 abundances across NIS (Fig. 3b), indicating that not only the number of species, but the total number
258 of individuals of NIS are higher in nearshore and shallow, coastal waters, especially in the southern
259 part of the area. Note however, that the individual abundance records are highly variable across
260 species, sites and seasons and may in some samples be considerably higher (>100) than the sum of
261 median abundances shown here (Table 1; Fig. S1). Interestingly, some of the sampling stations where
262 NIS are entirely absent are located in deeper, offshore sites with low vessel activity (Fig. 3c) (e.g., in
263 the North Sea).

264

265 *Drift model simulations*

266 The drift model simulations demonstrated a pronounced spatial heterogeneity throughout the area.
267 This is illustrated by high overall estimates of larvae settling densities, primarily in shallow, nearshore
268 waters, while the deeper, offshore areas, or waters with high current velocities (such as in the Sound
269 and southern Kattegat) receive very few number of larvae (Fig. 4a). These patterns are broadly
270 consistent with the underlying habitat maps used to constrain the initial locations when seeding the
271 agents in our simulations (Fig. 4b). However, the areal extent of habitats classified as suitable
272 according to species traits are clearly larger than the areas receiving a high number of settled larvae,
273 particularly by extending further offshore. Please note that the simulated settling densities, as well as
274 habitat maps differ between individual species (Fig. S2-S3). These differences in turn, arise from
275 species-specific differences in the reported values of key traits reflecting their spawning period, larvae
276 duration and habitat requirements, respectively (Table S1).

277

278 *Statistical analysis*

279 Among the candidate set of predictors considered in the statistical modelling, we omitted surface
280 temperature since it had a high degree of co-linearity with other predictors, i.e., as demonstrated by a

281 high variance inflation factor ($VIF > 5$). The full model fitted on the remaining covariates demonstrated
282 significant predictors overall. Consequently, no further model reduction was performed. The model
283 demonstrated a reasonable degree of explained deviance in observed presence and absence across
284 species, sites and time periods (51.9%; Table 2). Furthermore, the model demonstrated a high overall
285 predictive performance, as demonstrated by an AUC (i.e., area under the receiver-operator
286 characteristic curve) value equal to 0.95 (i.e., 95% overall accuracy). However, please note that the
287 accuracy of predicting absences was considerably greater than predicting presences (Fig. S4). In terms
288 of RF, the final model demonstrated a slightly better overall predictive accuracy compared to GAM
289 (i.e., out-of-bag (OOB) error rate = 2.3%). More importantly, it showed more equal prediction skills
290 for both classes, demonstrated by mean error rates amounting to 2.2% and 6.6% for absences and
291 presences, respectively.

292

293 The fixed effect (parametric) terms of the GAM indicated clear differences in the overall mean
294 probability of presence between species, as well as between 5-year time periods (Table 2), where in
295 the latter case the higher estimated coefficients for the recent decades indicate an overall increased
296 overall probability of presence of NIS over time. Furthermore, the partial smooth plots showed a
297 positive, exponentially increasing response of probability of presence to higher settling densities (Fig.
298 5a). However, please note that the relationship towards low values of settling density are driven by a
299 pronounced scatter of observations at zero values of settling density. The corresponding partial
300 dependence plots derived based on RF predictions showed a similar positive relationship, but with a
301 more linear increase across all values of settling densities (Fig. 6a). In terms of ship activity, both
302 GAM and RF demonstrated a positive and saturating response at higher values of vessel density (Fig.
303 5b; Fig. 6b). Please note that the slightly decreasing relationship based on GAMs towards the highest
304 values of vessel density is uncertain and driven by very few extreme observations. Hence, it should be
305 considered with caution. The relationships with temperature, salinity and depth indicate the partial
306 responses of probability of presence of NIS to environmental conditions (Fig. 5c-e; Fig. 6c-e). The
307 responses are largely similar between methods, except for bottom temperature, where GAM

308 demonstrated a dome-shaped relationship (with highest probability at medium temperatures of
309 ~10°C), while the partial dependence based on RF was more erratic and highly variable across the
310 range of values. With regards to salinity and depth both methods demonstrated a negative, linear or
311 non-linear declining relationship, respectively, with low probability of presence at greater depths and
312 high salinities. Among the entire set of predictors, species identity had the highest relative variable
313 importance in the RF model, as illustrated by the largest decrease in the mean Gini index (Fig. S5).
314 Subsequently, the set of environmental predictors were deemed important (i.e., in the following order:
315 salinity > depth >> temperature), closely followed by the effect of shipping activity, time period and
316 settling density. Finally, predictions based on the selected models demonstrate a high mean
317 probability of presence of NIS in shallow, brackish and coastal areas, especially in the inner Danish
318 waters of Kattegat and the Belt Seas (Fig. 7a, b). In general, GAMs predict higher mean probabilities
319 compared to RF and also demonstrate a higher standard deviation of predicted probabilities across all
320 species compared to RF (Fig. 7c, d). Conversely, RF predicted higher probabilities along the major
321 shipping routes compared to GAM. In general, the model predictions are able to adequately
322 characterize the overall patterns of observed probability of presence of NIS based on available
323 monitoring data, but tend to either slightly overestimate (i.e., GAM) or underestimate (i.e., RF)
324 probabilities (Fig. S6), particularly in the inner Danish waters (Fig. 7e, f).

325

326 **Discussion**

327 The underlying mechanisms determining the success of NIS outside their native ranges can be
328 described as a multi-stage process of invasions involving a number of filters, or community assembly
329 processes regulating their introduction, establishment and spread (Keddy, 1992; Kolar & Lodge,
330 2001; Beletsky et al., 2017). The complicated interplay between these processes means that it can be
331 difficult to understand and project biotic invasions in any given area (Clarke Murray et al., 2014),
332 particularly in open marine environments. To overcome this challenge, we here developed a joint
333 modelling approach, including high-resolution monitoring data, drift model simulations and a set of

334 statistical SDMs to investigate the effect and relative importance of multiple factors explaining the
335 occurrence of marine NIS in Danish coastal waters. Below we discuss the key drivers and assembly
336 processes involved, and briefly elaborate on the broader use of our approach and the improved
337 process understanding it generates within conservation and management.

338

339 The first step in the multi-stage process of invasions regards the initial introduction of organisms to a
340 novel habitat, well beyond a natural range expansion from its native distribution area. In marine
341 ecosystems, the main vectors and paths transporting NIS to distant locations are caused by
342 commercial (long-range) shipping, primarily through carrying organisms, particularly their early-lifts
343 stages via ballast water and hull fouling (Drake & Lodge, 2007; Bailey, 2015). Our model results,
344 demonstrating a positive relationship between shipping activity and the probability of occurrence of
345 NIS, lends quantitative support to previous studies identifying shipping as a primary factor
346 determining the introduction of marine NIS into the area (Leppäkoski et al., 2002; Ojaveer et al.,
347 2010). Moreover, the derived partial dependence plots of our models indicate that the responses to
348 increasing shipping activity may be described as rapidly saturating. This means that potential effects
349 of shipping for introduction of NIS may occur at even low, or relatively moderate activity levels,
350 beyond which a further increase has only a marginal effect on the potential introduction and
351 probability of presence of marine NIS. Once transported and introduced to a new habitat, the second
352 step of the invasion process regards the establishment and growth of self-sustaining populations. This
353 establishment phase is largely dictated by the particular traits and adaptations of the organisms in
354 question, allowing them to tolerate the environmental conditions in the new locations (Gallien &
355 Carboni, 2017). Our model results show that the type of environments where NIS are more likely to
356 be present are shallow coastal areas, primarily in brackish waters (i.e., salinity <15 psu) of medium
357 temperatures (~10°C). Among the environmental variables, depth and salinity are likely the key
358 limiting factors determining the occurrence and establishment of NIS within our study area, as
359 illustrated by their considerably higher relative importance compared to temperature. The negative,
360 linear response to depth is unlikely due to a direct physiological constraint acting on the organisms in

361 question (since all should be able to tolerate the water pressure at these shallow depths). Instead, it
362 may partly reflect a negative response to low oxygen concentrations at greater depths. Such
363 permanent or seasonal bottom hypoxia are widespread and increasing throughout the region (Conley
364 et al., 2011) and have shown to have serious impacts on benthic communities (Rosenberg et al., 1991;
365 Gray et al., 2002; Conley et al., 2007), as well as the overall ecosystem dynamics in the area
366 (Lindgren et al., 2010, 2012b). Alternatively, the negative effect of depth may also reflect a declining
367 availability of food that these benthic filter- and/or deposit feeders rely upon, since the concentrations
368 of plankton, as well as sinking detritus reaching the seafloor decrease with depth (Suess, 1980). In
369 contrast to depth, the dome-shaped response to temperature and exponentially declining response to
370 salinity are likely due to direct, physiological constraints, primarily acting on thermo-tolerance and
371 osmoregulation, especially for the more sensitive larvae stages (Gosselin & Qian, 1997; Pechenik,
372 1999). Since temperature was found less influential and the effect uncertain and sensitive to model
373 choice (i.e., illustrated by the different partial effects based on GAM and RF), salinity may prove to
374 be the critical environmental conditions allowing them to successfully occupy and establish within the
375 study area. Hence, our results support the important role of environmental filtering as one of the
376 primary assembly process affecting the establishment of NIS and community composition at large in
377 both terrestrial and aquatic environments (Gallien et al., 2014; Beukhof et al., 2019b; Radinger et al.,
378 2019).

379

380 Once a NIS have been introduced and managed to establish a self-sustaining population, the final
381 stage of the invasion process involves the subsequent spread of individuals across the land/seascape.
382 While some marine organisms, especially fish are able to utilize both active swimming as adults and
383 passive drift of early-life stages as modes of dispersal, benthic invertebrates are mainly restricted to
384 dispersal during their planktonic egg or larvae stages (Pechenik, 1999). Our results demonstrate a
385 positive effect of the simulated settling densities on the observed presence and absence of NIS, thus
386 highlighting that larvae dispersal may allow for an effective secondary spread of marine NIS in
387 coastal waters. This is well in accordance with other marine and freshwater studies, some including

388 drift models to study the role of dispersal on NIS (Gallien et al., 2015; Beletsky et al., 2017; Li &
389 Shen, 2020). Although this factor may seem to have a smaller relative importance compared to the
390 other above mentioned factors (i.e., shipping and environmental conditions), the effect of settling
391 densities was significant and robust to the choice of method (i.e., showing similar response curves in
392 both GAM and RF). In real marine environments, as well as in our drift model simulations, the
393 dispersal limitation of NIS are conditioned on the interplay between the hydrography of the study area
394 (i.e., in terms of currents or other pathways and barriers for dispersal), as well as the particular traits
395 of the organisms in question (i.e., here determined by their timing of spawning and larvae duration).
396 Taken together, our findings illustrate the role of dispersal filtering, acting as one of the key assembly
397 processes determining the success of invasive benthic invertebrates in coastal waters.

398

399 The holistic approach we have developed and applied using a combination of observational data,
400 advanced drift models and multiple SDMs has generated insight into the key factors and assembly
401 processes affecting the success of marine NIS. However, we fully acknowledge that our
402 understanding of all the underlying mechanisms and drivers regulating their introduction,
403 establishment and spread throughout the entire study area and beyond is still incomplete. This is
404 primarily due to a number of limitations, such as the availability and quality of input data, as well as
405 sources of model uncertainty. The NOVANA monitoring program provides an excellent and best
406 available data set for studying marine NIS within the study area. But, the number and geographical
407 coverage of stations, the choice of gear, as well as the spatio-temporal resolution of sampling may not
408 entirely reflect the true underlying occurrences and abundances of the set of NIS considered; nor may
409 it represent the entire range of conditions and habitats where NIS are present or absent. This is
410 especially true for NIS in the early phases of the invasion process that have not yet reached
411 equilibrium and managed to occupy all the potential sites and habitats with conditions fitting their
412 reported niches (Fig. S3). Secondly, while we aimed to include a broad set of covariates, we may have
413 missed, or not sufficiently accounted for all the key determinants regulating the introduction,
414 establishment and spread of NIS. One such aspect is the degree to which biotic interaction, including

415 competition and predator-prey interactions with the native species of the recipient communities and
416 food webs may affect the success of NIS (Shinen & Morgan, 2009; Gallien et al., 2015; David et al.,
417 2017; Gallien & Carboni, 2017). Formally accounting for this additional assembly process (termed
418 biotic filtering or limiting similarity) is challenging, since it requires a detailed understanding of
419 potential competition or predator-prey interactions between NIS and native species of the recipient
420 communities. In the absence of such information, biotic interactions may be inferred on the basis of
421 traits and spatio-temporal patterns of co-occurrence among both native and non-native species.
422 Promising method developments, including advanced Bayesian statistics and deep-learning
423 algorithms are emerging that may, at least partly, take such information into considerations when
424 studying community assembly processes (Chen et al., 2017; Ovaskainen et al., 2017; Tikhonov et al.,
425 2020). Hence, we advocate future studies to explore and take advantage of such method developments
426 to better understand the interplay and relative importance of both abiotic and biotic factors regulating
427 the establishment and success of NIS. Other potential variables not included in our study are e.g., the
428 frequency and duration of hypoxic events in coastal waters, as well as the proximity to major ports,
429 reflecting source of early-life stages (through ballast water and hull fouling) increasing propagule
430 pressure and the probability of successful introductions of NIS (Hansen et al., 2020). In addition to
431 data limitations, potential sources of model uncertainty, including both the parameterization and setup
432 may affect the results and the conclusions drawn from these. Fortunately, such sensitivity may, at
433 least to some degree be tested and accounted for by applying several, different methods using the
434 same input data (Norberg et al., 2019; Lindegren et al., 2020). Our approach demonstrates the value of
435 model comparisons to identify outcomes robust to choice of method.

436

437 *Conclusion and management outlook*

438 Besides generating improved process understanding of the key mechanisms and assembly rules
439 determining the success of marine NIS, our findings have clear implications for conservation and
440 management purposes. Firstly, the demonstrated effect of shipping on the probability of presence of

441 NIS, even at rather low to moderate activity levels, illustrates the need to drastically limit, or divert
442 shipping, especially within, or adjacent to marine protected areas (MPAs). This in order to minimize
443 the risk of unintentional introductions of NIS, especially if the primary objective of protection is to
444 support biodiversity and ensure a pristine composition of species. This is particularly urgent since
445 relatively few MPAs are strictly protected and therefore open to a broad range of human activities,
446 including fishing and shipping (Dudley et al., 2010; Fitzsimons, 2011; Barr & Possingham, 2013;
447 Edgar et al., 2014). Secondly, the significant effect of environmental conditions, notably salinity and
448 depth serves to identify the type of habitats and areas particularly vulnerable to marine NIS, as well as
449 the range of traits of NIS likely to inhabit and establish in such habitats. Thirdly, the significant effect
450 of settling densities emphasize also the role of dispersal, particularly of planktonic early-life stages as
451 a key factor promoting, or hindering spread of marine NIS across the sea scape. These insights may
452 provide key guidance for improving and prioritising existing monitoring programs and sampling
453 protocols to ensure a timely detection and robust status assessment of marine NIS. As an example, our
454 identification of vulnerable habitats and areas demonstrating high predicted probability of presence of
455 NIS may serve to aid in the planning and prioritization of limited conservation efforts aiming to avoid
456 or combat marine NIS. In our case study, this is particularly evident in the shallow, brackish and
457 coastal areas in the inner Danish waters where both predicted probabilities of NIS (Fig. 7a, b) and
458 human impacts are high. Implementing conservation efforts, either through protected areas and/or
459 other direct interventions (e.g., removals) are indeed critical, especially in the early stages of the
460 invasion process, since regulating an already established, self-supporting population is considerably
461 more difficult than avoiding them getting to a location in the first place. Finally, it is our hope and
462 ambition that the model development, knowledge and maps provided here will contribute to future
463 conservation planning and ecosystem-based management (Giakoumi et al., 2016) in order to achieve
464 the ambitious goals to prevent the introduction and spread of NIS and its dire consequences on the
465 biodiversity, functioning and services of ecosystems, particularly under climate change.

466

467

468 **References**

- 469 Albins, M. A. 2013. Effects of invasive Pacific red lionfish *Pterois volitans* versus a native predator
470 on Bahamian coral-reef fish communities. *Biological Invasions*, 15: 29–43.
- 471 Bailey, S. A. 2015. An overview of thirty years of research on ballast water as a vector for aquatic
472 invasive species to freshwater and marine environments. *Aquatic Ecosystem Health and*
473 *Management*, 18: 261–268.
- 474 Barr, L. M., and Possingham, H. P. 2013. Are outcomes matching policy commitments in Australian
475 marine conservation planning? *Marine Policy*, 42: 39–48.
- 476 Beletsky, D., Beletsky, R., Rutherford, E. S., Sieracki, J. L., Bossenbroek, J. M., Lindsay Chadderton,
477 W., Wittmann, M. E., *et al.* 2017. Predicting spread of aquatic invasive species by lake currents.
478 *Journal of Great Lakes Research*, 43: 14–32.
- 479 Beukhof, E., Spaanheden Dencker, T., Pecuchet L., Lindegren, M. 2019a. Spatio-temporal variation in
480 marine fish traits reveals community-wide responses to environmental change. *Marine Ecology-*
481 *Progress Series*. 610:205-222.
- 482 Beukhof, E., Frelat, R., Pecuchet, L., Maureaud, A., Dencker, T. S., Sólmundsson, J., Punzón, A., *et*
483 *al.* 2019b. Marine fish traits follow fast-slow continuum across oceans. *Scientific reports*, 9.
484 17878. doi:10.1038/s41598-019-53998-2
- 485 Christensen, Asbjørn. 2008, Bank resolved prognoses of sandeel fishing potential in the North Sea”.
486 Final report for the project "Fiskeriudsigt for tobis i Nordsøen på bankeniveau. (FIUF, 2005-
487 2007).
- 488 Christensen, Asbjørn, Patrizio Mariani, and Mark R. Payne. 2018. A Generic Framework for
489 Individual-Based Modelling and Physical-Biological Interaction. *PLoS ONE* 13(1): e0189956.
490 <https://doi.org/10.1371/journal.pone.0189956>
- 491 Chen, D., Xue, Y., Fink, D., Chen, S., and Gomes, C. P. 2017. Deep multi-species embedding. *IJCAI*

492 International Joint Conference on Artificial Intelligence, 0: 3639–3646.

493 Clarke Murray, C., Gartner, H., Gregr, E. J., Chan, K., Pakhomov, E., and Therriault, T. W. 2014.
494 Spatial distribution of marine invasive species: Environmental, demographic and vector drivers.
495 Diversity and Distributions, 20: 824–836.

496 Conley, D. J., Carstensen, J., Aertebjerg, G., Christensen, P. B., Dalsgaard, T., Hansen, J. L. S., and
497 Josefson, A. B. 2007. Long-term changes and impacts of hypoxia in Danish coastal waters RID
498 B-5428-2008. Ecological Applications, 17: S165–S184.

499 Conley, D. J., Carstensen, J., Aigars, J., Axe, P., Bonsdorff, E., Eremina, T., Haahti, B. M., *et al.*
500 2011. Hypoxia is increasing in the coastal zone of the baltic sea. Environmental Science and
501 Technology, 45: 6777–6783.

502 David, P., Thébault, E., Anneville, O., Duyck, P. F., Chapuis, E., and Loeuille, N. 2017. Impacts of
503 Invasive Species on Food Webs: A Review of Empirical Data. Advances in Ecological
504 Research, 56: 1–60.

505 Drake, J. M., and Lodge, D. M. 2007. Hull fouling is a risk factor for intercontinental species
506 exchange in aquatic ecosystems. Aquatic Invasions, 2: 121–131.

507 Dudley, N., Parrish, J. D., Redford, K. H., and Stolton, S. 2010. The revised IUCN protected area
508 management categories: the debate and ways forward. Oryx, 44: 485–490.

509 Edgar, G. J., Stuart-Smith, R. D., Willis, T. J., Kininmonth, S., Baker, S. C., Banks, S., Barrett, N. S.,
510 *et al.* 2014. Global conservation outcomes depend on marine protected areas with five key
511 features. Nature, 506: 216–20.

512 Ehrenfeld, J. G. 2010. Ecosystem consequences of biological invasions. Annual Review of Ecology,
513 Evolution, and Systematics, 41: 59–80.

514 Elith, J., and Leathwick, J. R. 2009. Species distribution models: Ecological explanation and
515 prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40:

516 677–697.

517 Fitzsimons, J. a. 2011. Mislabeling marine protected areas and why it matters-a case study of
518 Australia. *Conservation Letters*, 4: 340–345.

519 Galil, B. S., Marchini, A., Occhipinti-Ambrogi, A., Minchin, D., Naršćius, A., Ojaveer, H., and
520 Olenin, S. 2014. International arrivals: Widespread bioinvasions in European Seas. *Ethology
521 Ecology and Evolution*, 26: 152–171.

522 Gallien, L., Carboni, M., and Münkemüller, T. 2014. Identifying the signal of environmental filtering
523 and competition in invasion patterns - a contest of approaches from community ecology.
524 *Methods in Ecology and Evolution*, 5: 1002–1011.

525 Gallien, L., Mazel, F., Lavergne, S., Renaud, J., Douzet, R., and Thuiller, W. 2015. Contrasting the
526 effects of environment, dispersal and biotic interactions to explain the distribution of invasive
527 plants in alpine communities. *Biological Invasions*, 17: 1407–1423.

528 Gallien, L., and Carboni, M. 2017. The community ecology of invasive species: where are we and
529 what's next? *Ecography*, 40: 335–352.

530 Giakoumi, S., Guilhaumon, F., Kark, S., Terlizzi, A., Claudet, J., Felling, S., Cerrano, C., *et al.* 2016.
531 Space invaders; biological invasions in marine conservation planning. *Diversity and
532 Distributions*, 22: 1220–1231.

533 Goldsmit, J., Archambault, P., Chust, G., Villarino, E., Liu, G., Lukovich, J. V., Barber, D. G., *et al.*
534 2018. Projecting present and future habitat suitability of ship-mediated aquatic invasive species
535 in the Canadian Arctic. *Biological Invasions*, 20: 501–517.

536 Gosselin, L. A., and Qian, P. Y. 1997. Juvenile mortality in benthic marine invertebrates. *Marine
537 Ecology Progress Series*, 146: 265–282.

538 Gray, J. S., Wu, R. S. S., and Or, Y. Y. 2002. Effects of hypoxia and organic enrichment on the
539 coastal marine environment RID B-4203-2010. *Marine Ecology-Progress Series*, 238: 249–279.

- 540 Katsanevakis, S., Wallentinus, I., Zenetos, A., Leppäkoski, E., Çinar, M. E., Oztürk, B., Grabowski,
541 M., *et al.* 2014. Impacts of invasive alien marine species on ecosystem services and biodiversity:
542 A pan-European review. *Aquatic Invasions*, 9: 391–423.
- 543 Keddy, P. A. 1992. Assembly and response rules: two goals for predictive community ecology.
544 *Journal of Vegetation Science*, 3: 157–164.
- 545 Kolar, C. S., and Lodge, D. M. 2001. Progress in invasion biology: Predicting invaders.
- 546 Leppäkoski, E., Gollasch, S., Gruszka, P., Ojaveer, H., Olenin, S., and Panov, V. 2002. The Baltic - A
547 sea of invaders. *Canadian Journal of Fisheries and Aquatic Sciences*, 59: 1175–1188.
- 548 Li, Y., and Shen, Z. 2020. Roles of Dispersal Limit and Environmental Filtering in Shaping the
549 Spatiotemporal Patterns of Invasive Alien Plant Diversity in China. *Frontiers in Ecology and*
550 *Evolution*, 8: 1–14.
- 551 Lindegren, M., Diekmann, R., and Möllmann, C. 2010. Regime shifts, resilience and recovery of a
552 cod stock. *Marine Ecology Progress Series*, 402: 239–253.
- 553 Lindegren, M., Vigliano, P., and Nilsson, P. A. 2012a. Alien Invasions and the Game of Hide and
554 Seek in Patagonia. *Plos One*, 7: e44350–e44350.
- 555 Lindegren, M., Blenckner, T., and Stenseth, N. C. 2012b. Nutrient reduction and climate change cause
556 a potential shift from pelagic to benthic pathways in a eutrophic marine ecosystem. *Global*
557 *Change Biology*, 18: 3491–3503.
- 558 Lindegren, M., Thomas, M. K., Jónasdóttir, S. H., Nielsen, T. G., and Munk, P. 2020. Environmental
559 niche separation promotes coexistence among ecologically similar zooplankton species—North
560 Sea copepods as a case study. *Limnology and Oceanography*, 65: 545–556.
- 561 Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., and Bazzaz, F. A. 2000. Biotic
562 invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications*, 10:
563 689–710.

- 564 Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., *et*
565 *al.* 2019. A comprehensive evaluation of predictive performance of 33 species distribution
566 models at species and community levels. *Ecological Monographs*, 89: 1–24.
- 567 Ojaveer, H., Jaanus, A., Mackenzie, B. R., Martin, G., Olenin, S., Radziejewska, T., Telesh, I., *et al.*
568 2010. Status of biodiversity in the Baltic sea. *PLoS ONE*, 5: 1–19.
- 569 Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T.,
570 *et al.* 2017. How to make more out of community data? A conceptual framework and its
571 implementation as models and software. *Ecology Letters*, 20: 561–576.
- 572 Pechenik, J. A. 1999. On the advantages and disadvantages of larval stages in benthic marine
573 invertebrate life cycles. *Marine Ecology Progress Series*, 177: 269–297.
- 574 Pecuchet, L., Törnroos, A., and Lindegren, M. 2016. Patterns and drivers of fish community assembly
575 in a large marine ecosystem. *Marine Ecology Progress Series*, 546: 239–248.
- 576 Pimentel, D., Zuniga, R., and Morrison, D. 2005. Update on the environmental and economic costs
577 associated with alien-invasive species in the United States. *Ecological Economics*, 52: 273–288.
- 578 Radinger, J., Alcaraz-Hernández, J. D., and García-Berthou, E. 2019. Environmental filtering governs
579 the spatial distribution of alien fishes in a large, human-impacted Mediterranean river. *Diversity*
580 *and Distributions*, 25: 701–714.
- 581 Richardson, D. M., and Pyšek, P. 2012. Naturalization of introduced plants: Ecological drivers of
582 biogeographical patterns. *New Phytologist*, 196: 383–396.
- 583 Rosenberg, R., Hellman, B., and Johansson, B. 1991. Hypoxic Tolerance of Marine Benthic Fauna.
584 *Marine Ecology-Progress Series*, 79: 127–131. ISI:A1991HF03400010.
- 585 Rosenberg, R., Agrenius, S., Hellman, B., Nilsson, H. C., and Norling, K. 2002. Recovery of marine
586 benthic habitats and fauna in a Swedish fjord following improved oxygen conditions. *Marine*
587 *Ecology Progress Series*, 234: 43–53.

- 588 Shinen, J. S., and Morgan, S. G. 2009. Mechanisms of invasion resistance: competition among
589 intertidal mussels promotes establishment of invasive species and displacement of native
590 species. *Marine Ecology Progress Series*, 383: 187–197.
- 591 Stubbs, W. J., and Wilson, J. B. 2004. Evidence for limiting similarity in a sand dune community.
592 *Journal of Ecology*, 92: 557–567.
- 593 Theoharides, K. A., and Dukes, J. S. 2007. Plant invasion across space and time: Factors affecting
594 nonindigenous species success during four stages of invasion. *New Phytologist*, 176: 256–273.
- 595 Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., and
596 Ovaskainen, O. 2020. Joint species distribution modelling with the r-package Hmsc. *Methods in*
597 *Ecology and Evolution*, 11: 442–447.
- 598 Wood, S. N. 2017. *Generalized additive models: An introduction with R*, 2nd ed. Chapman and
599 Hall/CRC.

600

601 **Statements and Declarations**

602 This project was funded by the Danish Maritime Fund by request of the Danish Environmental
603 Protection Agency (DEPA). The authors declare they have no financial interests.

604

605 **Author Contributions**

606 FH and ML designed the study. FH and APG performed drift model simulations. APG performed data
607 extraction. ML performed data processing and statistical analysis. All authors took part in assessing
608 and discussing results, as well as writing the ms.

609

610

611 **Table 1.** Scientific names and summary statistics of the set of NIS present within the NOVANA
 612 monitoring program.

| Species | Taxa | Id | Nr. samples present | Nr. stations present | Probability present (%) | Abundance (mean) | Abundance (max) |
|--------------------------------|------------|----|---------------------|----------------------|-------------------------|------------------|-----------------|
| <i>Amphibalanus improvisus</i> | Arthropoda | Ai | 12 | 5 | 0.48 | 0.03 | 72 |
| <i>Alitta succinea</i> | Anellida | As | 141 | 14 | 1.87 | 0.04 | 28 |
| <i>Crepidula fornicata</i> | Mollusca | Cf | 4 | 3 | 0.16 | 0.01 | 8 |
| <i>Ensis directus</i> | Mollusca | Ed | 54 | 15 | 2.15 | 0.02 | 3 |
| <i>Marenzelleria neglecta</i> | Anellida | Ma | 665 | 32 | 8.81 | 0.44 | 136 |
| <i>Molgula manhattensis</i> | Chordata | Mn | 1 | 1 | 0.04 | <0.01 | 1 |
| <i>Mya arenaria</i> | Mollusca | Mv | 241 | 10 | 9.58 | 0.38 | 35 |
| <i>Palaemon elegans</i> | Crustacea | Pc | 273 | 20 | 3.62 | 0.26 | 128 |
| <i>Polydora cornuta</i> | Anellida | Pe | 1 | 1 | 0.01 | <0.01 | 1 |
| <i>Rhithropanopeus harrisi</i> | Crustacea | Rh | 2 | 1 | 0.03 | <0.01 | 2 |
| <i>Streblospio benedicti</i> | Anellida | Sb | 55 | 8 | 0.73 | 0.02 | 17 |

613

614

615

616

617

618

619

620

621 **Table 2.** Summary statistics of the final GAM, including the fixed effect factors species (Sp) and 5-
622 year time period (5y), as well as the smooth terms for the environmental predictors. Edf is the
623 estimated degrees of freedom for the model smooth terms (s) (i.e., edf>1 indicates a non-linear
624 relationship), Chi.sq is the associated chi square statistics and p-values indicate the significance of
625 each predictor (sal=salinity, temB=bottom temperature, ship=vessels activity, spn=settling densities).
626 (See Table 1 for species id and corresponding names).

| 627 Parametric coefficients | | | | |
|-----------------------------|----------|--------|--------|---------|
| | Estimate | SE | z | P |
| 628 (Intercept) | -7.80 | 0.48 | -16.32 | <0.001 |
| 630 Sp(As) | 2.15 | 0.48 | 4.53 | <0.00 |
| 631 Sp(Cf) | -0.90 | 0.67 | -1.33 | 0.18 |
| 632 Sp(Ed) | 1.56 | 0.49 | 3.18 | <0.001 |
| 633 Sp(Ma) | 4.61 | 0.45 | 10.15 | <0.001 |
| 634 Sp(Mn) | -2.27 | 1.10 | -2.06 | 0.04 |
| 635 Sp(Mv) | 1.90 | 0.48 | 3.94 | <0.01 |
| 636 Sp(Pc) | 2.83 | 0.47 | 6.06 | <0.01 |
| 637 Sp(Pe) | 2.24 | 1.10 | -2.04 | 0.04 |
| 638 Sp(Rh) | -63.51 | 95.00 | 0.01 | 1.00 |
| 639 Sp(Sb) | 1.65 | 0.49 | 3.34 | <0.001 |
| 640 5y[05-10] | 0.21 | 0.18 | 1.15 | 0.25 |
| 641 5y[10-15] | 0.75 | 0.19 | 3.91 | <0.001 |
| 642 5y[15-20] | 0.61 | 0.21 | 2.94 | <0.001 |
| 643 | | | | |
| 644 Smooth terms | | | | |
| | edf | Ref.df | Chi.sq | p-value |
| 645 s(spn) | 1.98 | 2.00 | 56.18 | <0.001 |
| 647 s(ship) | 1.91 | 1.99 | 15.90 | <0.001 |
| 648 s(temB) | 1.95 | 2.00 | 24.35 | <0.001 |
| 649 s(sal) | 1.98 | 2.00 | 385.22 | <0.001 |
| 650 s(Depth) | 1.01 | 1.02 | 312.26 | <0.001 |

651 R-sq.(adj) = 0.447 Deviance explained = 51.9%
652 UBRE = -0.88238 Scale est. = 1 n = 14757

653
654
655

656

657

658

659

660 **Figure 1.** Map of the study area and its position within northern Europe. Points indicate the positions
661 or available sampling station of the NOVANA program used in this study. The size of points
662 represents the sample effort as the total number of available samples for each site across seasons and
663 years.

664

665 **Figure 2.** Maps showing the vessel activity (a), depth (b), as well as the mean annual temperature (c)
666 and salinity (d) throughout the study area. The inland fjord shaded in grey (“Limfjorden”) was
667 excluded in the analysis.

668

669 **Figure 3.** The total number of NIS frequently occurring at each NOVANA station (after having
670 excluded sporadic species with median abundances below zero) (a), as well as the sum of median
671 abundances across NIS (b).

672

673 **Figure 4.** Mean of normalized settling densities of larvae based on drift model simulations averaged
674 across all NIS (a), as well as the potential habitat range shown as the total number of NIS
675 experiencing suitable environmental conditions fitting their reported tolerances to depth, temperature
676 and salinity (Table S1). (Maps for each species are shown in Fig. S2-S3).

677

678 **Figure 5.** GAM smooth plots showing the partial effects of settling density (a), vessel density (b),
679 bottom temperature (c), salinity (d) and depth (e) on the probability of presence of NIS. Solid lines
680 and shaded areas show the mean and 95% confidence intervals, respectively, and filled circles the
681 partial residuals for each predictor.

682

683 **Figure 6.** Partial dependence plots based on RF showing the partial effects of settling density (a),
684 vessel density (b), bottom temperature (c), salinity (d) and depth (e) on the probability of presence of
685 NIS.

686

687 **Figure 7.** The mean (a, b) and standard deviation (c, d) of predicted probability of presence across all
688 NIS throughout the study area based on the final GAM (left column) or RF (right column). In the
689 lower panels the mean deviation between predicted and observed probability of presence of NIS (ΔP)
690 are shown for each sampling station for GAM (e) and RF (f), respectively. (See Fig.S6 for an overall
691 comparison of predicted vs observed probability of presence).

692

693



Lat

56

55

54

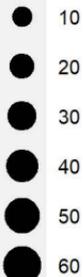
4

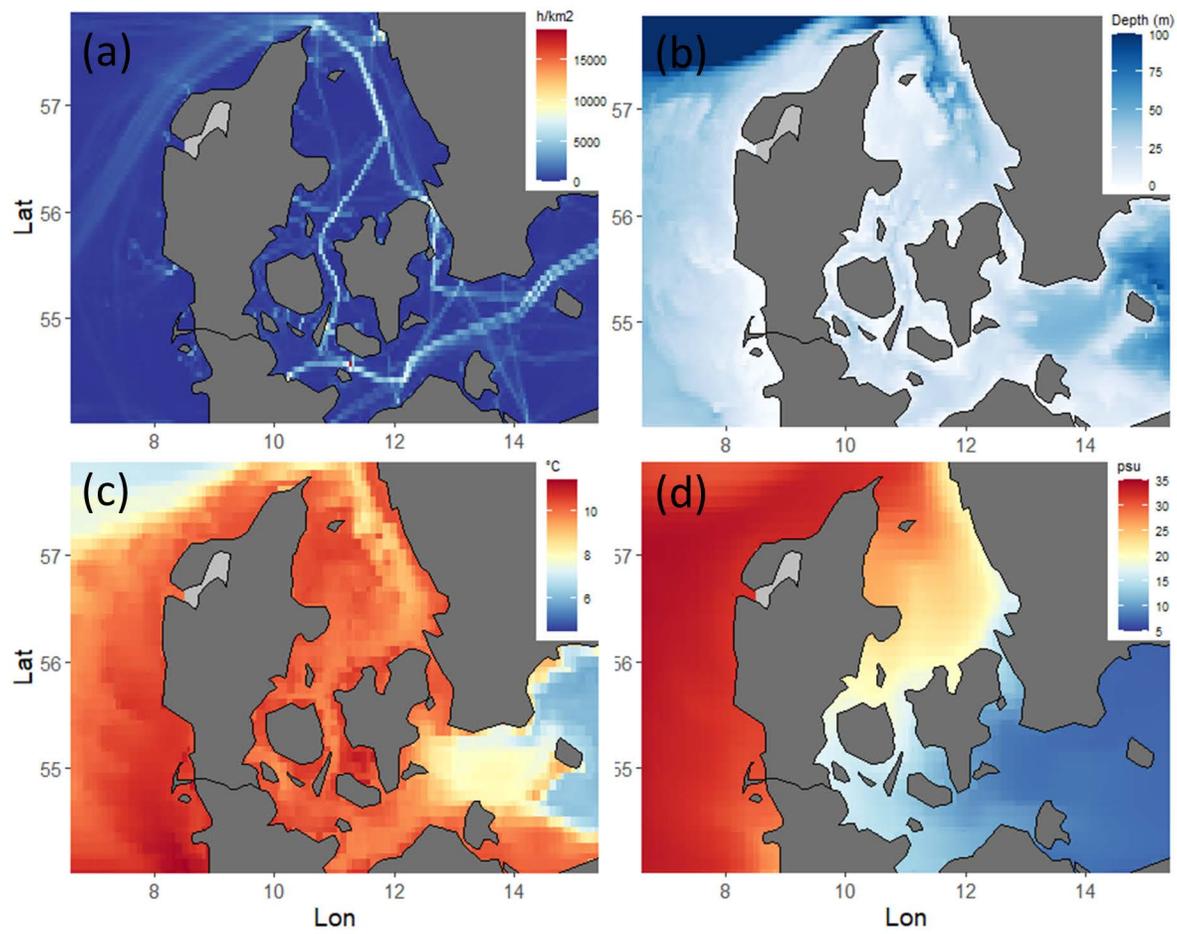
8

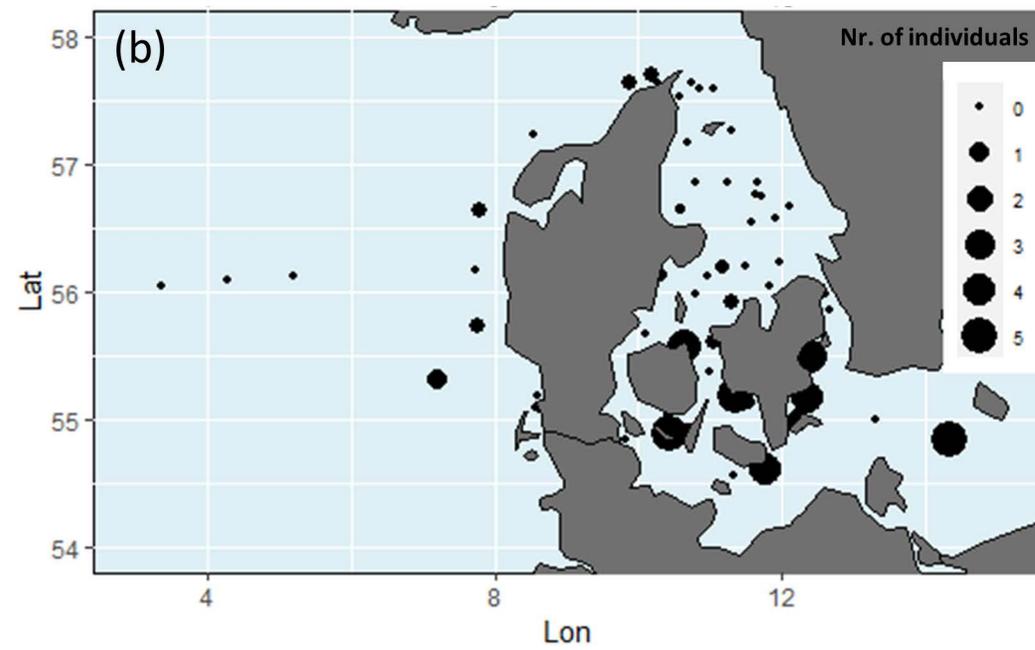
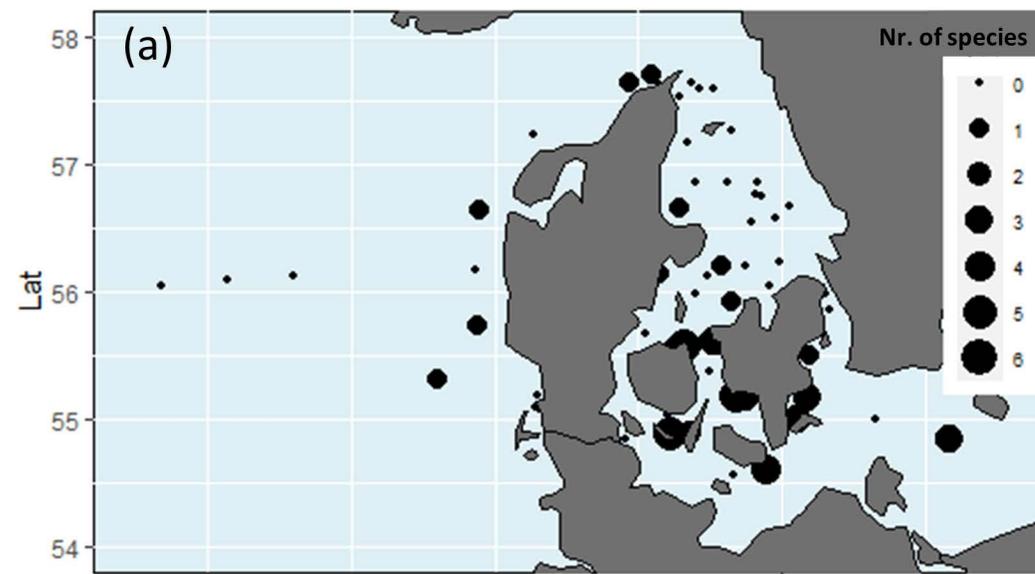
Lon

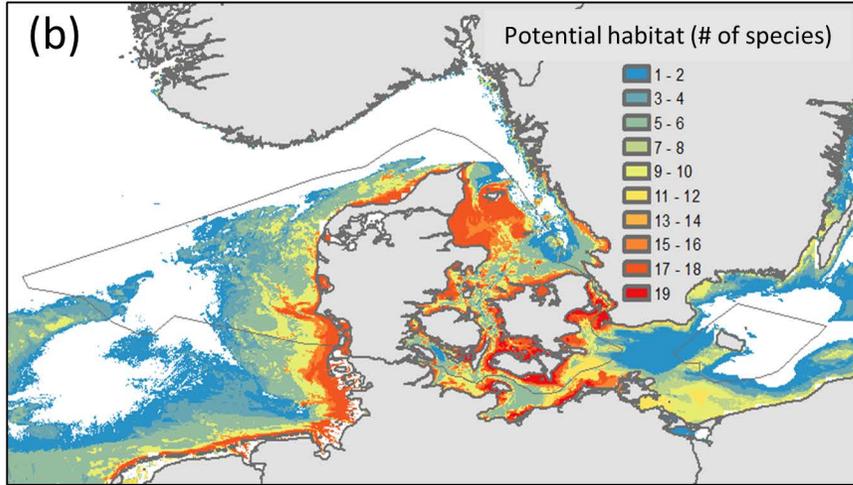
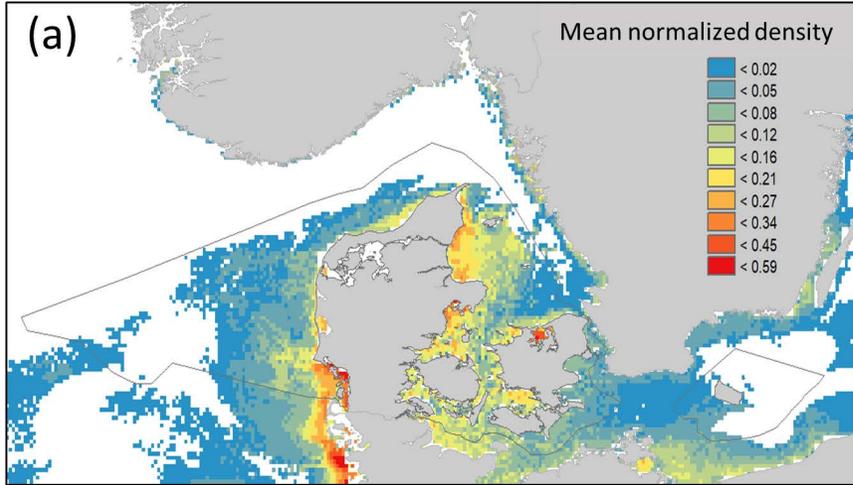
12

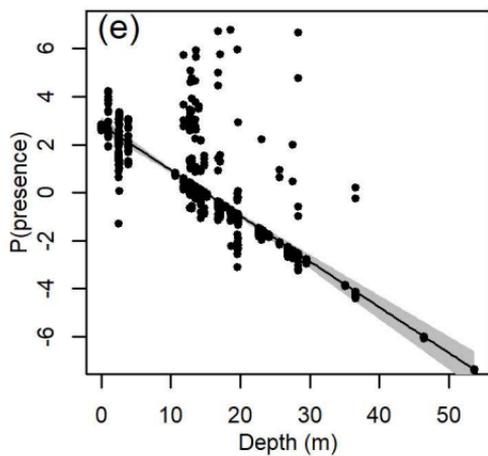
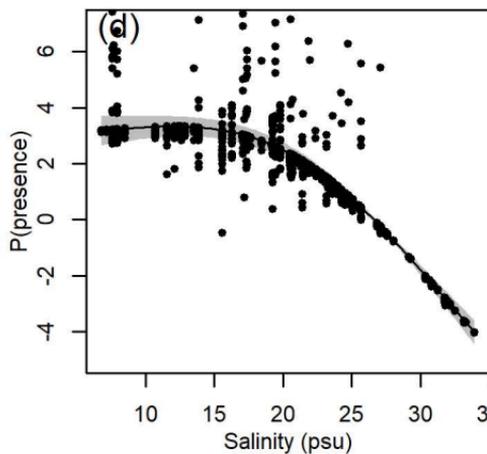
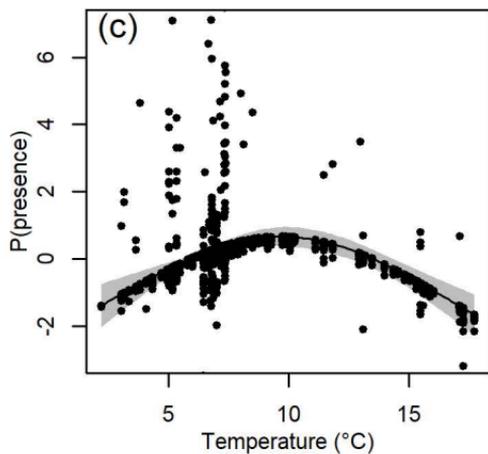
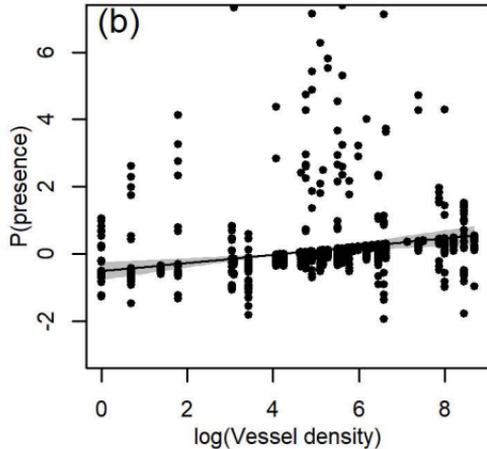
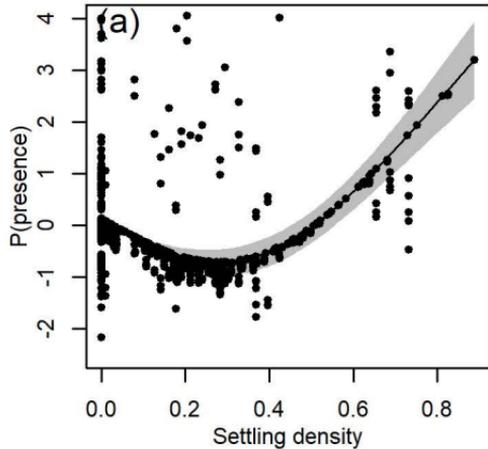
Nr. samples

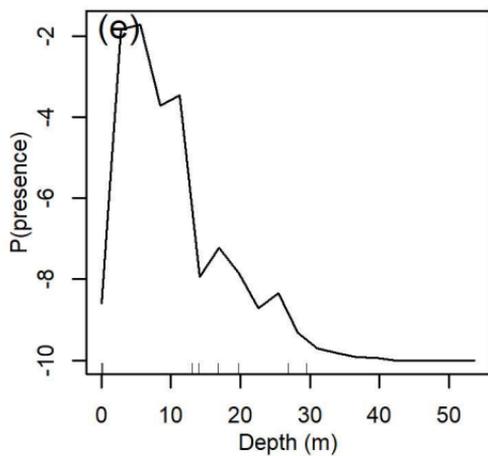
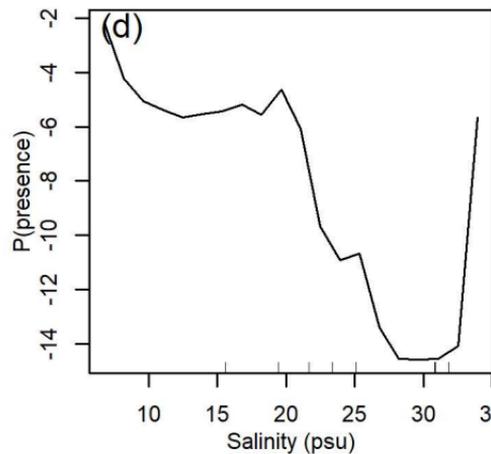
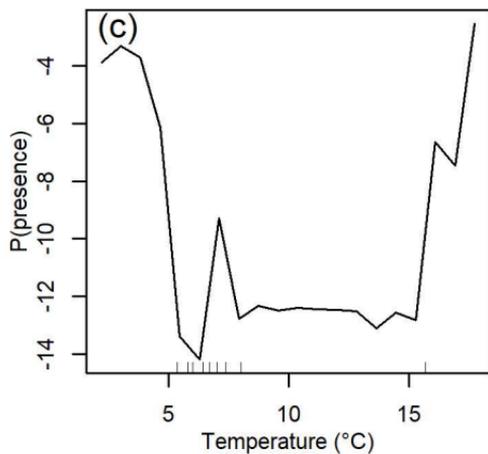
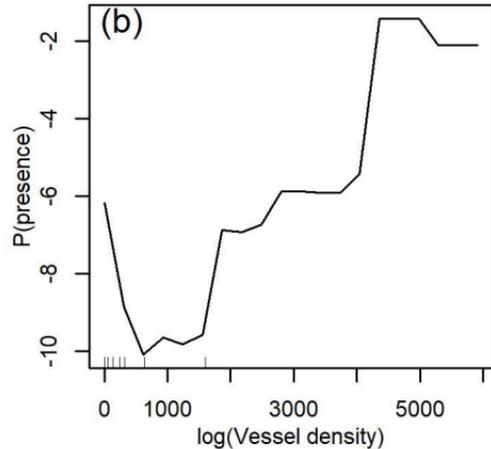
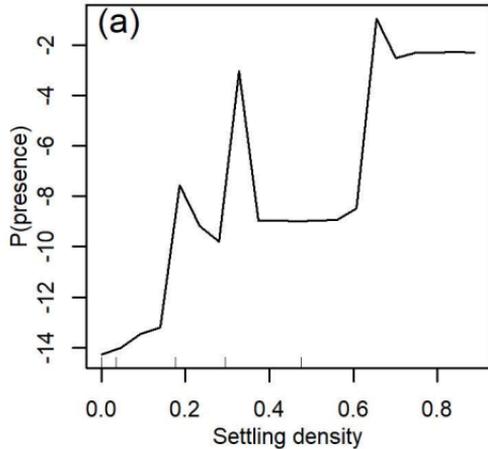


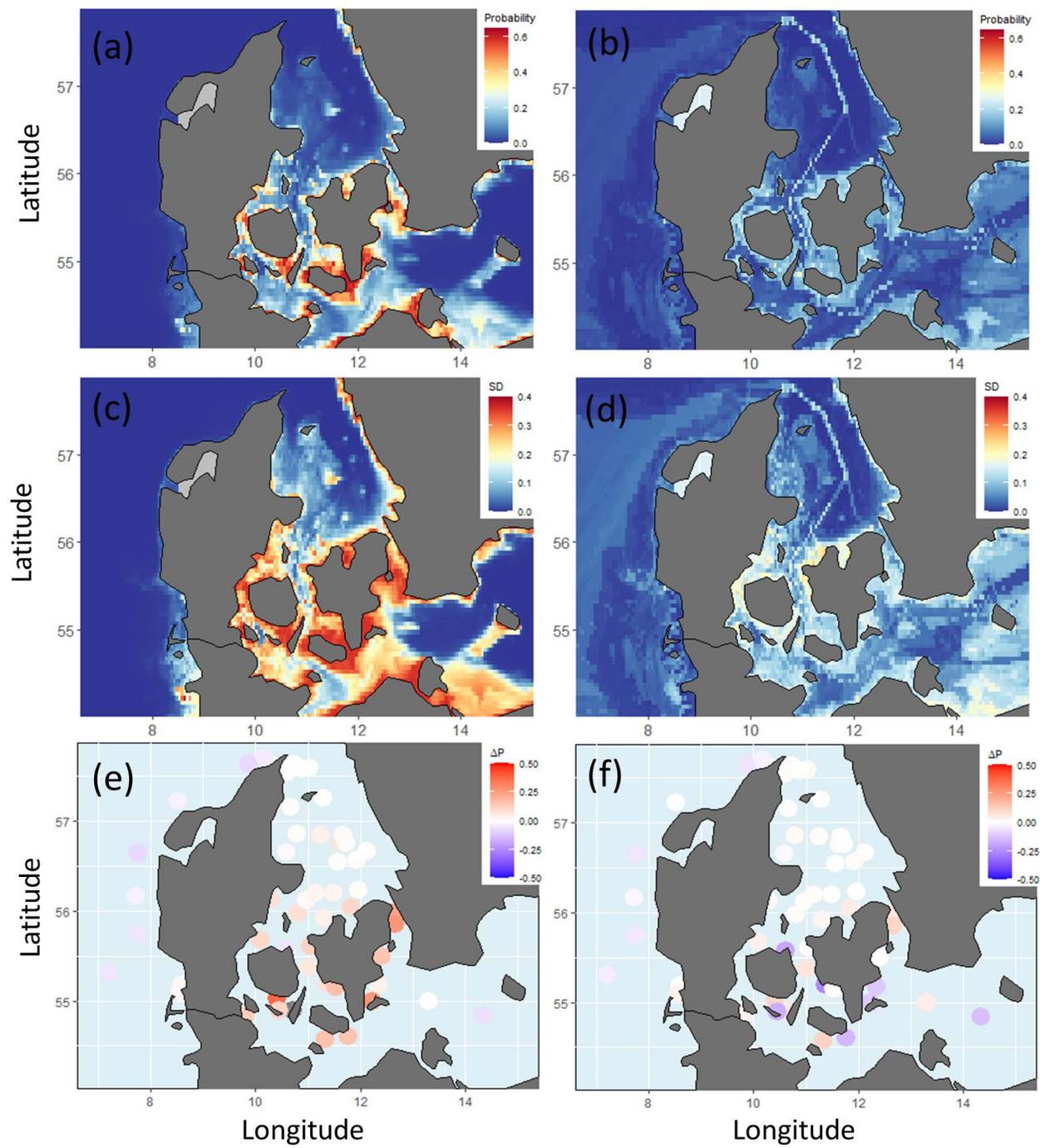












1 **Identifying key processes and drivers affecting the success of non-indigenous marine species in**
2 **coastal waters: a joint modelling approach**

3 Martin Lindegren^{1,2*}, Aurelia Pereira Gabellini², Peter Munk², Karen Edelvang², Flemming Hansen²

4 *¹Centre for Ocean Life, National Institute of Aquatic Resources, Technical University of Denmark,*
5 *Kemitorvet, Building 202, 2800 Kgs. Lyngby, Denmark. ²National Institute of Aquatic Resources,*
6 *Technical University of Denmark, Kemitorvet, Building 202, 2800 Kgs. Lyngby, Denmark.*

7 *Corresponding author: mli@aqua.dtu.dk

8

9 **Supplementary information**

10 *Species traits*

11 Data on species traits was collected for all NIS. Traits included are the spawning period, pelagic
12 larvae duration, substrate required to settle, preferred depth, and temperature and salinity tolerances
13 for adults and larvae (Table S1). Information on traits was collected from various data sources
14 including data portals (e.g. NEMESIS, CABI, AQUANIS, NOBANIS, OBIS) as well peer review
15 publications and reports. Much of the traits data refer to studies from other parts of the world. These
16 could be from the native range of the species or at a given area of introduction, or they could be from
17 laboratory test, which may however not be directly applicable to Danish conditions. If we did not find
18 sufficient or applicable information in literature the traits needed to be estimated. Such estimations
19 were based on the following assumptions:

20 a) Pelagic larval duration (PLD): We used minimum and maximum values found in the data sources,
21 and then we used the mean of the minimum and maximum PLD as the basis for the drift model
22 simulations.

23 b) Spawning season: In case of missing information of spawning season for Danish conditions, we
24 estimated the likely months of start and end of the spawning season based on either data on minimum

25 temperature for reproduction or larval survival (if available), or data on spawning season from other
26 geographical regions and considering the difference in climatic conditions.

27 c) Habitat substrates: In general data on substrate preferences were available for all species. We
28 assigned each species to one or more of 5 classes of habitat substrates including 3 seabed substrates:
29 Mud, Sand and Hard substrate, and 2 biogenic substrates “mussels and oyster shells” and “Seaweed
30 and/or seagrasses.”)

31 d) Water depth: Preferred water depth was primarily collected and inferred from the OBIS (Ocean
32 biodiversity information system) supplemented by other available data sources.

33 e) Temperature, adults: In general, data was available for the range of temperature where the
34 populations of the species have been found together with information published on temperature
35 tolerances from laboratory experiments.

36 f) Temperature, larvae: For some species data was found on larval temperature tolerance range, and/or
37 lower temperature thresholds for onset of reproduction. In cases where this type of data was not
38 available we used data on temperature tolerances/preferences for adult life stages.

39 g) Salinity, adults: In general data was available for the range of salinities where the populations of
40 the species have been found together with information published on salinity tolerances from
41 laboratory experiments.

42 h) Salinity, larvae: For some species data was found on larval salinity tolerance range from laboratory
43 experiments. In cases where data was not found, we used data on salinity tolerances of adult life
44 stages.

45 As an exception to the other 22 species, the orange striped green sea anemone (*Diadumene lineata*)
46 does not have sexual reproduction in introduced areas. Dispersal occurs by fission of fragments that
47 can drift and reattach to substrates at a new location. For this species we estimated “spawning season”
48 from studies on fission rate dependency on temperature (Flenikken 2017). And we used estimated
49 PLD from an expected dispersal distance of ca. 10 km (Bering Sea Marine Invasive Species

50 Assessment), although this fragment can principally drift for many weeks or month. For 3 of the 23
51 species we did not find sufficient data on traits. These include *Polydora aggregate*, *Hypereteone*
52 *heteropoda* and *Fenestrulina malusii*. For the species *Teredo navalis*, it inhabits fixed and floating
53 wood and this type of data is not available.

54 Of the 7 species where data on traits were collected as part of the Same-Risk-Area Case Study for
55 Kattegat and Øresund (Hansen & Christensen 2018), we revised the traits for 2 of the species
56 according the assumption and criteria as described above. For *Austrominius modestus* we extended the
57 depth range from a maximum of 5 m to 20 meters to comply with depth recordings of registrations
58 (OBIS) and we extended the habitat preference to include the coastal zone (0-15m) where mussels,
59 oysters and seaweed are considered abundant. For *Marenzelleria viridis* we changed the lower
60 salinity tolerance 1 to 16 PSU, due to some confusion in previous reportings distinguishing between
61 *M. viridis* and *M. neglecta*. We also changed the preferred depth from 63 m to 20 (to comply with the
62 OBIS database). Finally we change the habitat substrate preference from “mud” to “mud and sand” to
63 comply with observation of this species along Danish and other European shores and shallow areas
64 (Kristensen et al. 2012). The final assembly of trait information on NIS shown in Table S1.

65

66 *Definition of species habitats*

67 Habitat maps for each species were created as raster maps with a spatial resolution of 0.01 degrees
68 and covering the spatial extend of 3 – 17 degree east, and 53 – 60 degree north. The classification of
69 preferred seabed substrates into 3 classes (Mud, Sand, and Hard substrate) was based on a
70 reclassification of the seabed substrates from EMODNET (see Table S2). Notice that “hard substrate”
71 and “Sand” both include the EMODNET seabed substrates “Mixed sediments” and “Coarse
72 sediments”. The rationale for this is that these two EMODNET classes represent seabeds with varying
73 compositions of sand, gravel and stones and we make the assumption that those types of habitats
74 support both species associated with sandy and hard substrates. In addition to seabed substrate from
75 EMODNET, for species associated with hard substrates including mussel and oyster shells, and/or

76 seaweed and seagrass surfaces, we extended the hard substrate category to include all seabed within
77 the depth interval from 0 – 15 m. Within this depth range mussels, in particular blue mussel (*Mytilus*
78 *edulis*), are abundant throughout large parts of the Danish marine and brackish water territory.
79 Similarly, this depth range support both seagrass and seaweeds. Although both mussel and oyster
80 abundances as well as coverages of seaweed and seagrasses, may be highly fragmented depending a
81 number of factors including physical stress at the seabed (e.g. from waves), bottom trawling fishing
82 activities, light attenuation etc. the exclusion of these biogenic substrates will substantially
83 underestimate the extent and coverage of potential habitats for certain species. These include
84 *Amphibalanus improvises*, *Austrominius modestus*, *Crepidula fornicate*, *Diadumene lineata*, *Molgula*
85 *manhattensis* and *Styela clava*.

86

87 **Supplementary references**

88 Flenikken M M 2017. Understanding the Invasion Success and Spread of the Globally Introduced
89 Marine Invertebrate, *Diadumene Lineata*. State University of New York at Stony Brook, ProQuest
90 Dissertations Publishing, 2017. 10602467.

91 Hansen, F. T., & Christensen, A. (2018). Same Risk Area Case-study for Kattegat and Øresund. Final
92 report. DTU Aqua Report, No. 335-2018.

93 Kristensen E, Banta G T, Quintana C O, Delefosse M, Flindt M, 2012. Hvad ved vi om Svovlorm,
94 *Marenzelleria viridis*? Vand og jord. Vol.19, pp 27-32. (in Danish)

95 **Table S1.** Traits of 23 benthic species considered as non-indigenous in Danish waters. Species with asterisk denote the NIS present also in the
 96 NOVANA data set (see Table 1). The four species in “red” were not included in the study due to lack of data. Habitat abbreviations: ALL=all
 97 habitat types; S= sandy; M= muddy; H=Hard; Mus= mussel and oyster beds; W = seaweed and seagrasses.

| Species | Taxon | PLD (min) | PLD (max) | Generations per year | Spawning start | Spawning end | Habitat | Depth | Temp. Min Adult | Temp. Max Adult | Salinity Min. Adult | Salinity Max Adult | Temp. Min Larvae | Temp. Max Larvae | Salinity Min. Larvae | Salinity Max Larvae |
|-----------------------------------|------------|-----------|-----------|----------------------|----------------|--------------|---------|-------|-----------------|-----------------|---------------------|--------------------|------------------|------------------|----------------------|---------------------|
| | | Days | Days | | no. | Month | | | Month | C | C | PSU | PSU | C | C | PSU |
| <i>Alitta succinea</i> * | Anellida | 10 | 14 | 1 | June | Oct | ALL | 30 | 0 | 36 | 0.14 | 80 | | | 14 | 45 |
| <i>Amphibalanus improvises</i> * | Arthropoda | 11 | 42 | | June | Sept | HMus | 30 | -2 | 38 | 0.5 | 40 | 10 | 30 | 2 | 40 |
| <i>Austrominius modestus</i> | Crustacea | 10 | 15 | 1 | 5 | 10 | HMusW | 20 | 0 | 26 | 14 | 40 | 6 | 25 | 25 | 32 |
| <i>Crassostrea gigas</i> | Mollusca | 21 | 28 | 1 | 7 | 8 | A | 15 | 3 | 35 | 12 | 42 | 18 | 26 | 10 | 42 |
| <i>Crepidula fornicate</i> * | Mollusca | 11 | 32 | 1 | June | Sept | HSMus | 40 | 0 | 30 | 15 | 32 | 15 | 30 | 15 | 30 |
| <i>Diadumene lineata</i> | Cnidaria | 0 | 10 | | June | Sept | HMusW | 50 | 0 | 40 | 7 | 74 | | | 7 | 40 |
| <i>Ensis directus</i> * | Mollusca | 14 | 21 | 1 | 3 | 4 | SM | 12 | 0 | 26 | 7 | 32 | 15 | 28 | 15 | 32 |
| <i>Hemigrapsus sanguineus</i> | Crustacea | 16 | 55 | 1 | 5 | 9 | HS | 40 | 5 | 30 | 15 | 33 | 15 | 30 | 20 | 35 |
| <i>Hemigrapsus takanoi</i> | Crustacea | 30 | 30 | 1 | 5 | 9 | A | 20 | 0 | 20 | 7 | 35 | 15 | 30 | 25 | 35 |
| <i>Marenzelleria neglecta</i> * | Anellida | 28 | 84 | 1 | Sep | Sept | SM | 50 | 0 | 25 | 0.05 | 10 | | | 0 | 10 |
| <i>Marenzelleria viridis</i> * | Annelida | 28 | 49 | 1 | 9 | 11 | SM | 20 | 0 | 25 | 16 | 32 | 15 | 25 | 16 | 30 |
| <i>Molgula manhattensis</i> * | Chordata | 1 | 3 | 3 | May | Sept | HMus | 40 | 5 | | 9 | 40 | 10 | | 9 | 40 |
| <i>Mya arenaria</i> * | Mollusca | 14 | 21 | 1 | May | June | SM | 30 | -2 | 28 | 4 | 35 | 12 | 28 | 10 | 32 |
| <i>Palaemon elegans</i> * | Crustacean | 18 | 45 | 1 | April | Sept | ALL | 10 | 2 | 34 | 1 | 45 | 14 | 20 | 5 | 40 |
| <i>Petricolaria pholadiformis</i> | Mollusca | 10 | 14 | 1 | June | Aug | ALL | 30 | | | 10 | 30 | | | 10 | 30 |
| <i>Polydora cornuta</i> * | Anellida | 12 | 60 | 3 | May | Sept | SM | 20 | -2 | 29 | 2 | 75 | 10 | | 5 | 75 |
| <i>Rhithropanopeus harrisi</i> * | Crustacea | 7 | 43 | 1 | 6 | 9 | A | 37 | 0 | 35 | 5 | 30 | 14 | 27 | 5 | 30 |
| <i>Streblospio benedicti</i> * | Anellida | 5 | 40 | 2 | May | Oct | SM | 20 | -2 | 34 | 4 | 45 | 7.5 | 30 | 4 | 45 |
| <i>Styela clava</i> | Chordata | 1 | 1 | 1 | Jul | Aug | HMusW | 20 | -2 | 27 | 10 | 35 | 16 | | 20 | 35 |
| <i>Hypereteone heteropoda</i> | Anellida | | | | | | SM | 20 | | | 18 | 30 | | | | |
| <i>Polydora aggregate</i> | Anellida | 55 | | | May | Sept | | 20 | | | | 40 | 6 | | | |
| <i>Teredo navalis</i> | Mollusca | 20 | 34 | | Aug | Nov | Wood | 30 | 1 | 30 | 7 | 40 | 10 | 30 | 5 | 40 |
| <i>Fenestrulina malusii</i> | Bryozoa | 1 h | | | | | HMW | 110 | | | | | | | | |

99 **Table S2.** EMODNET seabed substrate classified into 3 main categories: “Mud”, “Sand” and
 100 “Hard substrates”. Notice that “hard substrate” and “Sand” both include “Mixed sediments”
 101 and “Coarse sediments”.

| Mud | Sand | Hard |
|-------------------------|-----------------|------------------------------|
| Fine mud | Sand | Rock or Other hard substrata |
| Mud to muddy sand | Coarse Sediment | Coarse Sediment |
| Sandy mud to Muddy sand | Mixed Sediment | Mixed sediments |
| Muddy Sand | | |
| Sandy Mud | | |

102

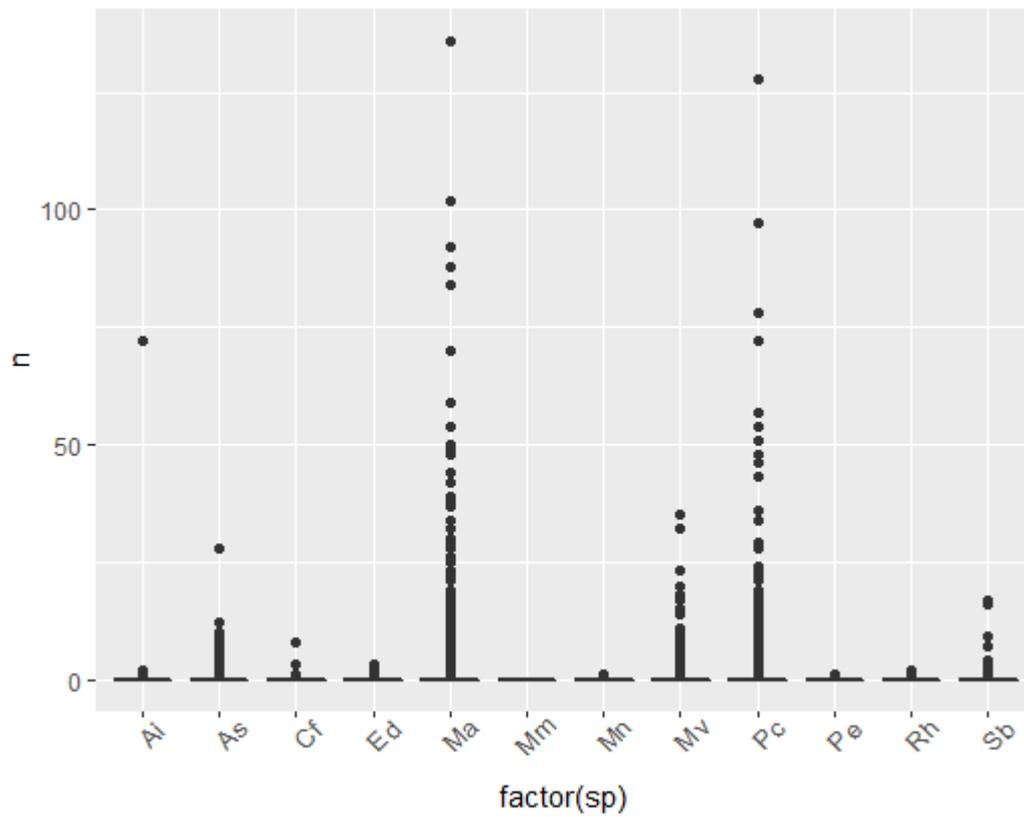
103

104

105

106

107



108

109 **Figure S1.** Plot showing the uneven distribution of species abundances (n) across all samples
 110 extracted from in the NOVANA data.

111

112

113

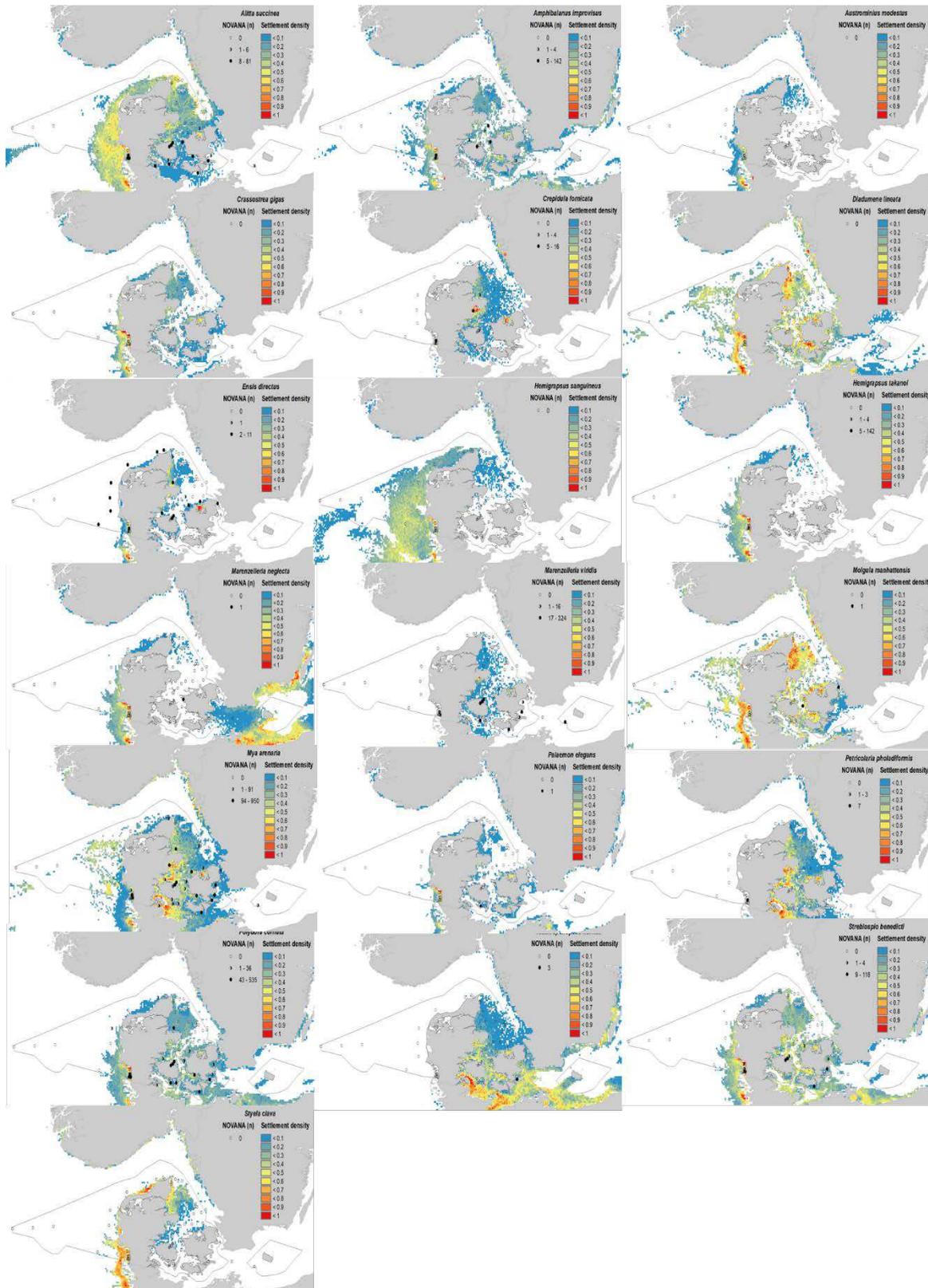
114

115

116

117

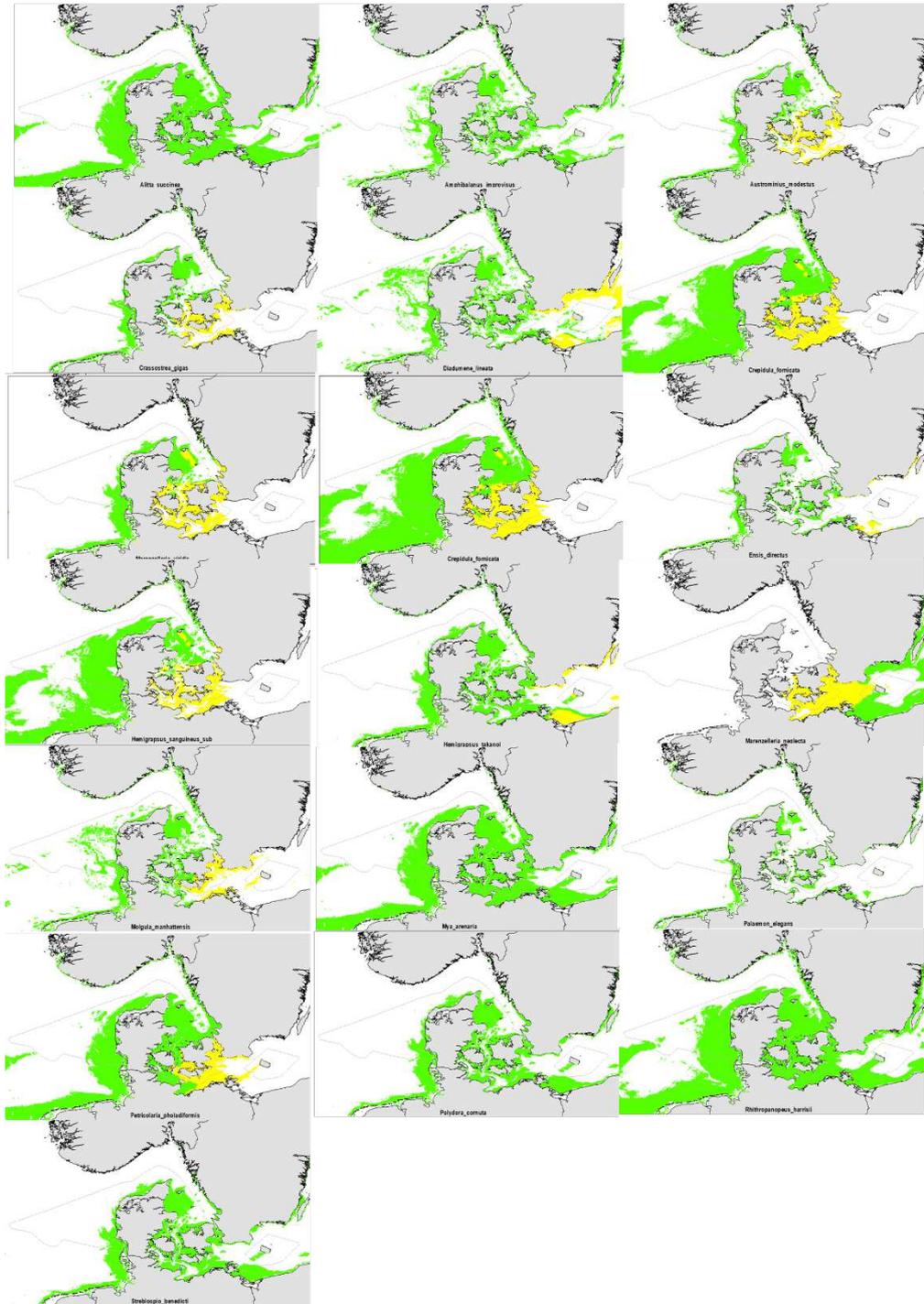
118



119

120 **Figure S2.** Maps of normalized settling densities for NIS derived from drift model

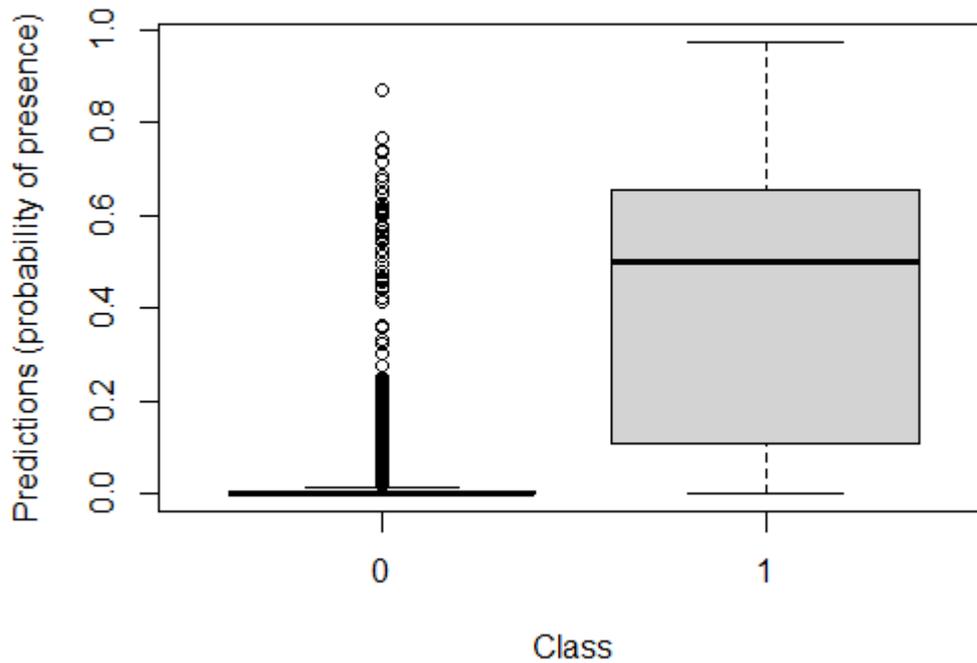
121 simulations.



122

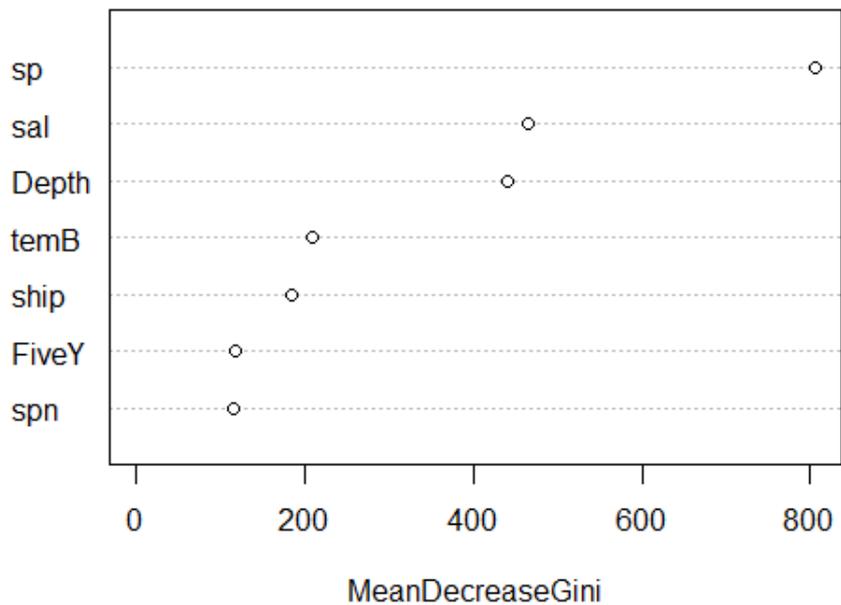
123 **Figure S3.** Habitat maps based on substrate preferences, depth preferences and adult salinity
 124 tolerances. Green represents habitats where salinity conditions are within the reported
 125 tolerance range of the species. Yellow represent habitats where the salinity conditions may
 126 exceed the tolerance thresholds of the species in shorter or longer periods.

127



128

129 **Figure S4.** Box plots of GAM predictions of probability of occurrence versus the observed
130 classes (0=absence and 1= presence). Please note the high accuracy of predicting absences
131 while predicted presence are highly scattered with a mean error of ~50%.

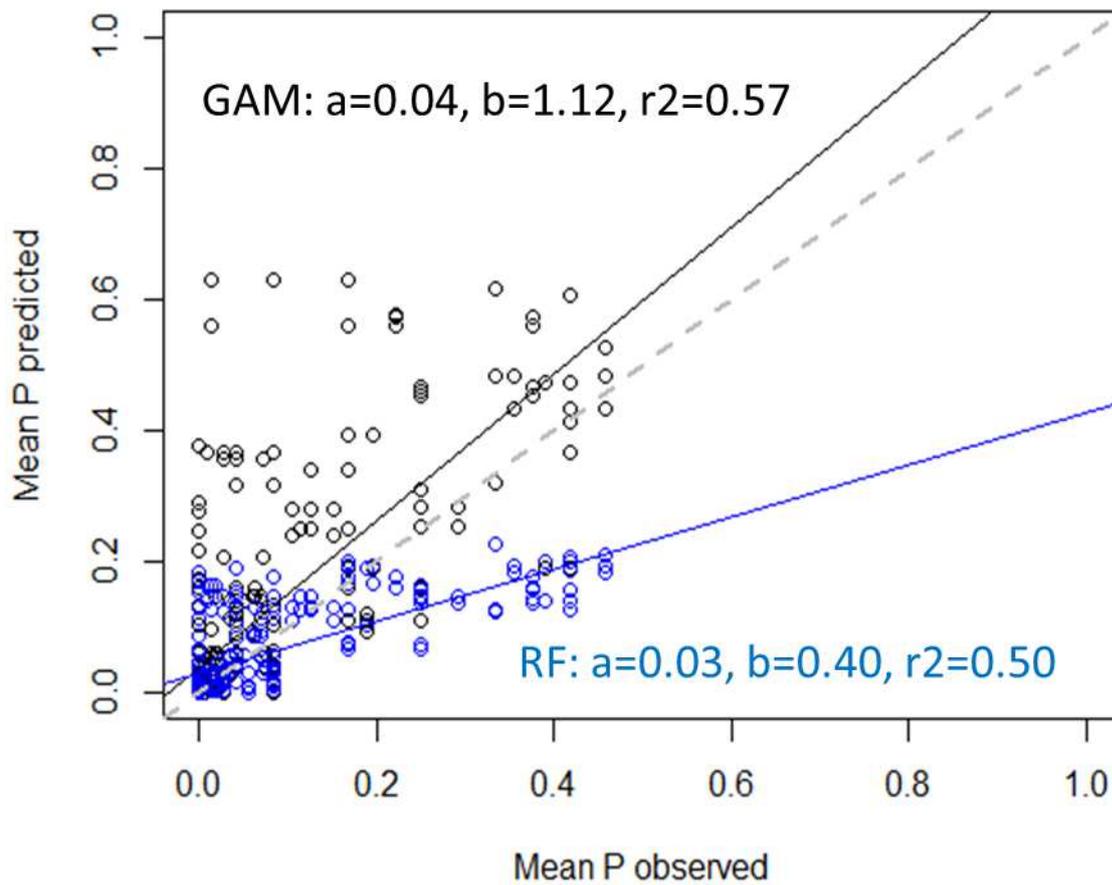


132

133 **Figure S5.** Variable importance of each predictors based on RF estimated as the mean
 134 decrease in the Gini index. (sp=species id (factor), sal=salinity, temB=bottom temperature,
 135 ship=vessels activity, FiveY= five year periods (factor), spn=setling densities).

136

137



138

139 **Figure S6.** Scatter plot of observed mean probability of presence across species per sampling
 140 station vs the predicted mean probabilities based on either the final GAM (black) and RF
 141 (blue). Please note that while the explained variance is similar between methods (i.e., 0.57
 142 and 0.50, respectively), GAM tend to slightly overestimate probabilities (i.e., the slope (b) is
 143 >1), while RF tend to underestimate (i.e., $b < 1$). The dashed grey line shows the one to one
 144 line. (See Fig. 7 for model predictions and deviations between observed and predicted mean
 145 probabilities per station for both methods).

146

147

148