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

Martin Lindegren (mli@aqu.edu)  
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
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
Identifying key processes and drivers affecting the success of non-indigenous marine species in coastal waters

Martin Lindegren1,2,*, Aurelia Pereira Gabellini2, Peter Munk2, Karen Edelvang2, Flemming Hansen2

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

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

Abstract

Non-indigenous species (NIS) pose a major threat to biodiversity and the functioning and services of ecosystems. Despite their rapid spread in coastal waters worldwide, biotic invasions are widely disregarded in marine conservation planning. To guide conservation actions, a better understanding of the underlying mechanisms determining the success of NIS are therefore needed. Here we develop 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 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.
Key words: community assembly, environmental filtering, drift modelling, species distribution modelling, marine spatial planning

Acknowledgements

This project was funded by the Danish Maritime Fund by request of the Danish Environmental Protection Agency (DEPA). We wish to thank all people involved in the NOVANA program for ensuring availability and accessibility to high-quality monitoring data.
Non-indigenous species (NIS) are considered one of the major threats to biodiversity and the functioning and services of ecosystems worldwide (Vitousek et al., 1996; Mack et al., 2000; Ehrenfeld, 2010; IPBES 2019). In order to mitigate their impacts on native species and communities, there is an urgent need to adopt management strategies capable of accounting for NIS. Despite their presence and negative effects on many aquatic ecosystems (Pimentel et al., 2005; Lindegren et al., 2012a; Albins, 2013; Katsanevakis et al., 2014), biotic invasions are widely disregarded in marine conservation planning (Giakoumi et al., 2016). Hence, the ambitious policy goals to prevent the introduction and spread of NIS (EC 2008; CBD 2021; Tittensor et al., 2014) are far from being achieved. To guide conservation actions, a better understanding of the underlying mechanisms controlling the introduction, establishment and spread of NIS (Theoharides & Dukes, 2007; Richardson & Pyšek, 2012), as well as tools to model and predict such invasions, are therefore needed (Gallien & Carboni, 2017).

The underlying mechanisms determining the presence and dynamics of NIS can conceptually be described as a multi-stage process of invasions comprised of: (i) the transport and introduction of organisms to a novel habitat from their native range; (ii) the establishment and growth of self-sustaining populations within the new environment, (iii) and the secondary spread of the organism across the land/seascape (Kolar & Lodge, 2001; Beletsky et al., 2017). The first step described above involves both a transport mechanism (i.e., vector) and a geographic path (i.e., route) along which an NIS is carried to a new location (Carlton & Ruiz 2005; Clarke Murray et al., 2014; Galil et al., 2014). In marine and freshwater environments, the main vectors and routes involve primarily aquaculture and commercial shipping through ballast water and hull fouling (Leppäkoski et al., 2002; Drake & Lodge, 2007; Ojaveer et al., 2010; Bailey, 2015). Once transported to a new location, the subsequent steps of invasions can be thought of as a series of “filters” or community assembly processes (Keddy, 1992) influencing the success of NIS outside their native range (Theoharides & Dukes, 2007;
Richardson & Pyšek, 2012; Gallien & Carboni, 2017). Firstly, the success of NIS is determined by the organism’s tolerance to the environmental conditions in a new location. This assembly processes, often termed environmental filtering, is largely conditioned on the particular characteristics (i.e., traits) of the organism in question and the degree to which its resulting environmental niche overlaps with the abiotic conditions in the new location (Ackerly, 2011; Gallien et al., 2014; Pecuchet et al., 2016; Beukhof et al., 2019a). Secondly, the degree to which NIS are able to grow and reproduce is conditioned by its competitive abilities to acquire space and/or resources relative to ecologically similar species within the recipient community and food web (Stubbs & Wilson, 2004; Shinen & Morgan, 2009; Gallien et al., 2015; David et al., 2017). This assembly process, often termed biotic filtering or limiting similarity, is also determined by the traits of organisms, such as their feeding rates, diet preferences and/or natural mortality imposed by native predators.

Finally, once established as a self-sustaining population, a range of natural or anthropogenic factors may determine the subsequent spread of NIS across the land/seascape. This final assembly process, termed dispersal filtering, is determined by both the physical environment (i.e., establishment of pathways or barriers for dispersal), as well as the movement of organisms, either through active locomotion or passive drift (Gallien et al., 2015; Beletsky et al., 2017; Li & Shen, 2020). In the latter case, movement is primarily through early-life stages such as seeds, eggs and larvae. As evident from the steps outlined above, the presence of NIS in any given location is jointly determined by multiple, non-exclusive assembly processes and mechanisms, both external and internal to the organism in question (Richardson & Pyšek, 2012). This complicated interplay of governing factors and processes means that it is challenging to understand and project the success and spread of NIS in any given area (Clarke Murray et al., 2014), particularly in open marine environments. To understand the key drivers and underlying mechanisms involved in shaping the success and dynamics of NIS, a holistic approach is needed that accounts for the effect of multiple interacting factors determining their transport, establishment and spread.
In this study, we develop and apply such a holistic approach to study the underlying mechanisms explaining the success of marine NIS, using Danish coastal waters as a case study. We focus primarily on benthic invertebrates with no or very limited locomotion as adults (e.g., bivalves, crustaceans and annelids), but with highly effective dispersal during their planktonic egg or larvae stages (Pechenik, 1999). To do so, we compiled a comprehensive set of long-term monitoring data on NIS in the coastal waters of the southeast North Sea, Kattegat and Western Baltic Sea (Fig. 1); a region highly affected by NIS (Leppäkoski et al., 2002; Ojaveer et al., 2010). We complemented these observations with data on commercial shipping, environmental conditions and estimates of larvae dispersal based on advanced drift model simulations. Together, these variables reflect a potential set of drivers affecting the introduction, establishment and spread of NIS across the study area, respectively. Using this data as joint input, we then applied a set of comparable species distribution modelling (SDM) methods to identify the key predictors and underlying community assembly processes determining the occurrence of NIS. Finally, we used a set of parameterized models to predict the probability of occurrence of NIS throughout the entire study area with the aim of providing managers with key information needed to prioritize monitoring and conservation actions, as well as develop marine spatial management plans accounting also for NIS and their potential impacts on marine ecosystems.

**Material and Methods**

*Monitoring data and standardization*

Based on a list of marine species considered as non-indigenous within the study area (Danish EPA, personal communication, 3rd July, 2020) we extracted the taxonomic information of all benthic invertebrates containing a benthic adult stage and a pelagic egg or larvae stage (Table S1). Subsequently, we extracted all records of NIS from the long-term Danish National Monitoring programme (NOVANA) hosted by the Danish Centre for Environment and Energy (https://odaforalle.au.dk/main.aspx, Accessed 15th of October 2020) (Table 1). The NOVANA program and its precursors date back to 1988 and include a set of frequently (i.e., monthly or
quarterly) sampled sites originally selected to assess the effect of anthropogenic impacts such as nutrient enrichment in Danish waters (Svendsen et al., 2005). At each station, five replicate sediment samples are typically collected for faunal analysis. The samples are rinsed over a 1mm sieve, preserved in 96% ethanol solution and subsequently sorted into individuals, identified to the lowest taxonomic level possible. To avoid bias introduced by differences in sampling protocols and gears, we included only records obtained using the standard 0.0143m² “HAPS” corer, the most commonly applied sampling method of benthic invertebrates across all years and sites within the monitoring program. Furthermore, we accounted for the marked differences in sampling effort over time and between sites (Fig. 1), which may bias the detection and probability of presence of NIS at any given site, by performing a formal data standardization. This was achieved by randomly bootstrapping (i.e., resampling without replacement) the same minimum number of unique samples from each site and season (i.e., aggregating months by quarters) over 500 random iterations. Two stations (“Hirsholmene” and “Knud001”) were excluded due to insufficient sampling coverage across years and seasons. We then extracted the mean abundances and their standard deviations for all species, sites, years and seasons across the entire set of randomized draws and used this standardized metric in the following statistical analysis. Furthermore, we restricted the analysis to the years from 2000 onwards to allow for a sufficient coverage of sampling across stations and ensure a temporal match with the drift model simulations. After omitting sporadic records based on other gear types and observations prior to 2000, the standardized data set contained 17,769 mean abundances (or presence/absences) of 11 NIS covering in total 66 sites.

Transport vectors and habitat characteristics

To represent the role of commercial shipping, which is considered to be one of the key transport vectors introducing marine NIS into the study area (Leppäkoski et al., 2002; Ojaveer et al., 2010), we collected available data on ship activity from EMODNET (https://www.emodnet-humanactivities.eu/view-data.php). The data is based on all available vessel positions retrieved from
the Automatic Identification System (AIS) in 2019, and is expressed as hours per km² (Fig. 2a). To represent key habitat characteristics allowing establishment of NIS, we also extracted available data on depth, temperature and salinity (Fig. 2b-d). Depth data were obtained from in situ measurements, supplemented with information from EMODNET if missing (https://www.emodnet.eu/en/bathymetry). Temperature (surface and bottom) and salinity (surface) were obtained from the Global Ocean Physics Reanalysis with (GLORYSs2v4) and were downloaded from the Copernicus Marine Environment Monitoring Service (http://marine.copernicus.eu/). In order to fit the spatio-temporal coverage of the observational data and the drift model simulations, temperature and salinity values were given at each of the sampling station as monthly (or seasonal) means from 2000 onwards.

Drift model simulations of larvae dispersal

To represent the dispersal phase of NIS (once having been introduced and established) we simulated larval dispersal using the agent-based modelling system IBMlib (Christensen, 2008; Christensen et al. 2018). IBMlib is a model library for individual-based modelling using a Lagrangian approach that has been specifically developed for simulating larval dispersal including biological processes affecting spawning, drift and settlement of marine larvae. We ran simulations for each of the NIS for the years 2001-2010, using a new modified version of the IBMlib software to include the mapping of the density of agent trajectories during the simulation. The model was forced by hydrodynamic data, including current velocities, salinity and temperature covering the entire study area, derived from a regionally downscaled hydrodynamic model of the North Sea-Baltic Sea (Berg & Poulsen 2012). The spatial resolution of the model is 0.5 nm in Kattegat, inner Danish straits and the western part of the Baltic Sea, and 5 nm outside this region. The vertical resolution of the model is 50 and 52 layers, respectively.
To identify a set of credible starting locations for releasing agents (i.e., larvae), as well as the timing of release and duration of simulations, available information on species traits reflecting their spawning period, pelagic larvae duration, substrate required to settle, preferred depth, as well as the temperature and salinity tolerances for adults and larvae were collected for each NIS (Table S1). Please note that most of the traits data refer to studies from other parts of the world. These could be from the native range of the species or determined on the basis of experiments. While spawning period and length of the pelagic larvae phase serve to constrain the release and duration of agents in the simulations, the remaining traits were used to create a set of species habitat maps following the approach by Hansen et al. (2020). These maps were used to constrain the seeding of agents to locations fitting some of the fundamental requirements of each species with respect to their seabed substrate, water depth and adult salinity tolerance (Supplementary text; Table S1, S2).

Taking these aspects into account, the agents in each simulation were distributed randomly in space within the maximum area coverage of the species habitat maps and randomly in time within the spawning period of each species. For each NIS, we released 200,000 agents per year from 2001 to 2010 (i.e. a total of 2 million agents per species). The mapping of the dispersal trajectories of agents was carried out by registering the presence of each agent at each time step in a regular grid (280 x 120) defined by 54°N to 60°N and 3°E to 17°E. All trajectories experiencing salinity conditions outside the larval salinity tolerance range reported for each species were ignored and not included in the mapping. Finally, we produced maps representing all end positions of successfully settled larvae (here termed “settlement densities”), taking into consideration not only currents, but the salinity range experienced in the trajectory, as well as the reported habitat requirements (depth and substrate type) at the time of larvae settlement. The maps were created by summing all the individual track counts across years, then dividing by the area of each grid cell. Furthermore, we normalized the densities by the maximum values observed for each species, so that the values range between 0 and 1. In order to fit the spatial coverage of the observational data on NIS, we estimated the average densities of settled larvae for each of the sampling station in the NOVANA data.
To assess the effect and relative importance of factors explaining the overall presence and absence of the set of NIS considered, we applied Generalized Additive Models (GAMs) and random forests (RFs). These methods represent two complementary tools commonly used in species distribution modelling (Elith & Leathwick, 2009; Goldsmit et al., 2018; Lindegren et al., 2020). The same set of explanatory variables, reflecting factors related to the introduction, establishment and spread of NIS were used as covariates. The resulting statistical relationships and derived response curves between the set of covariates and the presence and absence of NIS were compared across methods to assess the sensitivity and robustness of the results to the choice of methodology. In terms of GAM, we used the following binomial model formulation:

\[
PA_{s,l,y,q} = a + s(\text{spn}_{s,l}) + s(\text{ship}_{l}) + s(\text{temS}_{l,y,q}) + s(\text{temB}_{l,y,q}) + s(\text{sal}_{l,q}) + s(\text{depth}_{l}) + 5y + S + \varepsilon
\]

where the response variable PA is the presence/absence (0,1) of each species \(s\), at site \(l\), in year \(y\) and season \(q\) as a function (using a logit link) of the simulated settling density (spn), ship activity level (ship), surface and bottom temperature (temS, temB), salinity (sal) and depth at each site. To account for mean differences in the probability of presence between species and over time we included species identity (S) and years classified into 5-year time periods (5y) as fixed effect factors. The constant \(a\) is the overall intercept, \(s\) the thin plate smoothing function for each smooth term and \(\varepsilon\) the error term.

Although the number of regression splines is optimized (and penalized) by the generalized cross validation criterion (GCV), the degrees of freedom of the spline smoother function \(s\) was further constrained to three knots \((k=3)\) to allow for potential nonlinearities, but restrict flexibility during model fitting. Finally, we applied backwards model selection to identify the best possible set of predictors.

The second method used, i.e., RF, is a machine learning tool comprising ensembles of decision trees that rely on bagging (i.e. bootstrap aggregation). RFs are capable of reproducing complex nonlinear shapes in single and multiple dimensions, making them suitable for ecological applications in which complex shapes are to be expected (Breiman 2001). In addition, RF has fewer constraints and is able to
capture interactions between variables that cannot easily be achieved with GAMs. Individual classification trees within the random forest are trained on randomly selected subsets of the data. The final forest prediction is obtained by averaging predictions across all trees in the forest. We used the same model setup in terms of response and explanatory variables as in the GAM. In order to account for differences in the number of observations reporting absences versus presences of NIS, we accounted for such unbalanced classes by assigning different weights to each class (using the “classwt” option). Once trained on the available data, we used the final RFs (based on 10,000 individual trees) to estimate the relative importance of each predictor, as well as visualize the partial response curves of each individual explanatory variable. Finally, we used the final models to estimate the overall probability of occurrence of NIS across the entire study area, beyond the NOVANA monitoring stations used for model fitting and training. This was achieved by predicting occurrences based on a compilation of all covariate values in each grid cell, i.e., based on both the collected ship activity, depth, temperature and salinity data (Fig. 2), as well as the drift model outputs of species-specific settling densities (Fig. S2). The model predictions serve to highlight areas highly vulnerable to NIS and in need of monitoring and conservation actions. All statistical analyses were conducted using the R software, version 4.0.2 (R core Team 2020) using the following packages: “mgcv” (Wood 2017) and “randomForest” (Liaw and Wiener 2002).

Results

Spatial patterns of NIS and environmental conditions

The overall occurrences and abundances of the set of NIS demonstrated pronounced spatial heterogeneity across sites throughout the study area. This is illustrated by marked spatial differences in the number of NIS that are frequently occurring at each station based on our bootstrapping (Fig. 3a). Notably, the highest richness of NIS occurred in nearshore coastal areas, particularly in the southern part of the area (i.e., around the densely populated islands of Zealand and Fyn and in the Wadden Sea), while the lowest richness (or even absence) of NIS was found in more offshore areas in
Kattegat and the North Sea. A similar spatial pattern was also evident in the sum of median abundances across NIS (Fig. 3b), indicating that not only the number of species, but the total number of individuals of NIS are higher in nearshore and shallow, coastal waters, especially in the southern part of the area. Note however, that the individual abundance records are highly variable across species, sites and seasons and may in some samples be considerably higher (>100) than the sum of median abundances shown here (Table 1; Fig. S1). Interestingly, some of the sampling stations where NIS are entirely absent are located in deeper, offshore sites with low vessel activity (Fig. 3c) (e.g., in the North Sea).

Drift model simulations

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

Statistical analysis

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

The fixed effect (parametric) terms of the GAM indicated clear differences in the overall mean
probability of presence between species, as well as between 5-year time periods (Table 2), where in
the latter case the higher estimated coefficients for the recent decades indicate an overall increased
overall probability of presence of NIS over time. Furthermore, the partial smooth plots showed a
positive, exponentially increasing response of probability of presence to higher settling densities (Fig.
5a). However, please note that the relationship towards low values of settling density are driven by a
pronounced scatter of observations at zero values of settling density. The corresponding partial
dependence plots derived based on RF predictions showed a similar positive relationship, but with a
more linear increase across all values of settling densities (Fig. 6a). In terms of ship activity, both
GAM and RF demonstrated a positive and saturating response at higher values of vessel density (Fig.
5b; Fig. 6b). Please note that the slightly decreasing relationship based on GAMs towards the highest
values of vessel density is uncertain and driven by very few extreme observations. Hence, it should be
considered with caution. The relationships with temperature, salinity and depth indicate the partial
responses of probability of presence of NIS to environmental conditions (Fig. 5c-e; Fig. 6c-e). The
responses are largely similar between methods, except for bottom temperature, where GAM
demonstrated a dome-shaped relationship (with highest probability at medium temperatures of \(~10^\circ C\)), while the partial dependence based on RF was more erratic and highly variable across the range of values. With regards to salinity and depth both methods demonstrated a negative, linear or non-linear declining relationship, respectively, with low probability of presence at greater depths and high salinities. Among the entire set of predictors, species identity had the highest relative variable importance in the RF model, as illustrated by the largest decrease in the mean Gini index (Fig. S5).

Subsequently, the set of environmental predictors were deemed important (i.e., in the following order: salinity \(\gg\) depth \(\gg\) temperature), closely followed by the effect of shipping activity, time period and settling density. Finally, predictions based on the selected models demonstrate a high mean probability of presence of NIS in shallow, brackish and coastal areas, especially in the inner Danish waters of Kattegat and the Belt Seas (Fig. 7a, b). In general, GAMs predict higher mean probabilities compared to RF and also demonstrate a higher standard deviation of predicted probabilities across all species compared to RF (Fig. 7c, d). Conversely, RF predicted higher probabilities along the major shipping routes compared to GAM. In general, the model predictions are able to adequately characterize the overall patterns of observed probability of presence of NIS based on available monitoring data, but tend to either slightly overestimate (i.e., GAM) or underestimate (i.e., RF) probabilities (Fig. S6), particularly in the inner Danish waters (Fig. 7e, f).

Discussion

The underlying mechanisms determining the success of NIS outside their native ranges can be described as a multi-stage process of invasions involving a number of filters, or community assembly processes regulating their introduction, establishment and spread (Keddy, 1992; Kolar & Lodge, 2001; Beletsky et al., 2017). The complicated interplay between these processes means that it can be difficult to understand and project biotic invasions in any given area (Clarke Murray et al., 2014), particularly in open marine environments. To overcome this challenge, we here developed a joint modelling approach, including high-resolution monitoring data, drift model simulations and a set of
statistical SDMs to investigate the effect and relative importance of multiple factors explaining the occurrence of marine NIS in Danish coastal waters. Below we discuss the key drivers and assembly processes involved, and briefly elaborate on the broader use of our approach and the improved process understanding it generates within conservation and management.

The first step in the multi-stage process of invasions regards the initial introduction of organisms to a novel habitat, well beyond a natural range expansion from its native distribution area. In marine ecosystems, the main vectors and paths transporting NIS to distant locations are caused by commercial (long-range) shipping, primarily through carrying organisms, particularly their early-lifts stages via ballast water and hull fouling (Drake & Lodge, 2007; Bailey, 2015). Our model results, demonstrating a positive relationship between shipping activity and the probability of occurrence of NIS, lends quantitative support to previous studies identifying shipping as a primary factor determining the introduction of marine NIS into the area (Leppäkoski et al., 2002; Ojaveer et al., 2010). Moreover, the derived partial dependence plots of our models indicate that the responses to increasing shipping activity may be described as rapidly saturating. This means that potential effects of shipping for introduction of NIS may occur at even low, or relatively moderate activity levels, beyond which a further increase has only a marginal effect on the potential introduction and probability of presence of marine NIS. Once transported and introduced to a new habitat, the second step of the invasion process regards the establishment and growth of self-sustaining populations. This establishment phase is largely dictated by the particular traits and adaptations of the organisms in question, allowing them to tolerate the environmental conditions in the new locations (Gallien & Carboni, 2017). Our model results show that the type of environments where NIS are more likely to be present are shallow coastal areas, primarily in brackish waters (i.e., salinity <15 psu) of medium temperatures (~10°C). Among the environmental variables, depth and salinity are likely the key limiting factors determining the occurrence and establishment of NIS within our study area, as illustrated by their considerably higher relative importance compared to temperature. The negative, linear response to depth is unlikely due to a direct physiological constraint acting on the organisms in
question (since all should be able to tolerate the water pressure at these shallow depths). Instead, it
may partly reflect a negative response to low oxygen concentrations at greater depths. Such
permanent or seasonal bottom hypoxia are widespread and increasing throughout the region (Conley
et al., 2011) and have shown to have serious impacts on benthic communities (Rosenberg et al., 1991;
Gray et al., 2002; Conley et al., 2007), as well as the overall ecosystem dynamics in the area
(Lindegren et al., 2010, 2012b). Alternatively, the negative effect of depth may also reflect a declining
availability of food that these benthic filter- and/or deposit feeders rely upon, since the concentrations
of plankton, as well as sinking detritus reaching the seafloor decrease with depth (Suess, 1980). In
contrast to depth, the dome-shaped response to temperature and exponentially declining response to
salinity are likely due to direct, physiological constraints, primarily acting on thermo-tolerance and
osmoregulation, especially for the more sensitive larvae stages (Gosselin & Qian, 1997; Pechenik,
1999). Since temperature was found less influential and the effect uncertain and sensitive to model
choice (i.e., illustrated by the different partial effects based on GAM and RF), salinity may prove to
be the critical environmental conditions allowing them to successfully occupy and establish within the
study area. Hence, our results support the important role of environmental filtering as one of the
primary assembly process affecting the establishment of NIS and community composition at large in
both terrestrial and aquatic environments (Gallien et al., 2014; Beukhof et al., 2019b; Radinger et al.,
2019).

Once a NIS have been introduced and managed to establish a self-sustaining population, the final
stage of the invasion process involves the subsequent spread of individuals across the land/seascape.
While some marine organisms, especially fish are able to utilize both active swimming as adults and
passive drift of early-life stages as modes of dispersal, benthic invertebrates are mainly restricted to
dispersal during their planktonic egg or larvae stages (Pechenik, 1999). Our results demonstrate a
positive effect of the simulated settling densities on the observed presence and absence of NIS, thus
highlighting that larvae dispersal may allow for an effective secondary spread of marine NIS in
coastal waters. This is well in accordance with other marine and freshwater studies, some including
drift models to study the role of dispersal on NIS (Gallien et al., 2015; Beletsky et al., 2017; Li & Shen, 2020). Although this factor may seem to have a smaller relative importance compared to the other above mentioned factors (i.e., shipping and environmental conditions), the effect of settling densities was significant and robust to the choice of method (i.e., showing similar response curves in both GAM and RF). In real marine environments, as well as in our drift model simulations, the dispersal limitation of NIS are conditioned on the interplay between the hydrography of the study area (i.e., in terms of currents or other pathways and barriers for dispersal), as well as the particular traits of the organisms in question (i.e., here determined by their timing of spawning and larvae duration). Taken together, our findings illustrate the role of dispersal filtering, acting as one of the key assembly processes determining the success of invasive benthic invertebrates in coastal waters.

The holistic approach we have developed and applied using a combination of observational data, advanced drift models and multiple SDMs has generated insight into the key factors and assembly processes affecting the success of marine NIS. However, we fully acknowledge that our understanding of all the underlying mechanisms and drivers regulating their introduction, establishment and spread throughout the entire study area and beyond is still incomplete. This is primarily due to a number of limitations, such as the availability and quality of input data, as well as sources of model uncertainty. The NOVANA monitoring program provides an excellent and best available data set for studying marine NIS within the study area. But, the number and geographical coverage of stations, the choice of gear, as well as the spatio-temporal resolution of sampling may not entirely reflect the true underlying occurrences and abundances of the set of NIS considered; nor may it represent the entire range of conditions and habitats where NIS are present or absent. This is especially true for NIS in the early phases of the invasion process that have not yet reached equilibrium and managed to occupy all the potential sites and habitats with conditions fitting their reported niches (Fig. S3). Secondly, while we aimed to include a broad set of covariates, we may have missed, or not sufficiently accounted for all the key determinants regulating the introduction, establishment and spread of NIS. One such aspect is the degree to which biotic interaction, including
competition and predator-prey interactions with the native species of the recipient communities and food webs may affect the success of NIS (Shinen & Morgan, 2009; Gallien et al., 2015; David et al., 2017; Gallien & Carboni, 2017). Formally accounting for this additional assembly process (termed biotic filtering or limiting similarity) is challenging, since it requires a detailed understanding of potential competition or predator-prey interactions between NIS and native species of the recipient communities. In the absence of such information, biotic interactions may be inferred on the basis of traits and spatio-temporal patterns of co-occurrence among both native and non-native species. Promising method developments, including advanced Bayesian statistics and deep-learning algorithms are emerging that may, at least partly, take such information into considerations when studying community assembly processes (Chen et al., 2017; Ovaskainen et al., 2017; Tikhonov et al., 2020). Hence, we advocate future studies to explore and take advantage of such method developments to better understand the interplay and relative importance of both abiotic and biotic factors regulating the establishment and success of NIS. Other potential variables not included in our study are e.g., the frequency and duration of hypoxic events in coastal waters, as well as the proximity to major ports, reflecting source of early-life stages (through ballast water and hull fouling) increasing propagule pressure and the probability of successful introductions of NIS (Hansen et al., 2020). In addition to data limitations, potential sources of model uncertainty, including both the parameterization and setup may affect the results and the conclusions drawn from these. Fortunately, such sensitivity may, at least to some degree be tested and accounted for by applying several, different methods using the same input data (Norberg et al., 2019; Lindegren et al., 2020). Our approach demonstrates the value of model comparisons to identify outcomes robust to choice of method.

Conclusion and management outlook

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

Albins, M. A. 2013. Effects of invasive Pacific red lionfish Pterois volitans versus a native predator on Bahamian coral-reef fish communities. Biological Invasions, 15: 29–43.

Bailey, S. A. 2015. An overview of thirty years of research on ballast water as a vector for aquatic invasive species to freshwater and marine environments. Aquatic Ecosystem Health and Management, 18: 261–268.

Barr, L. M., and Possingham, H. P. 2013. Are outcomes matching policy commitments in Australian marine conservation planning? Marine Policy, 42: 39–48.

Beletsky, D., Beletsky, R., Rutherford, E. S., Sieracki, J. L., Bossenbroek, J. M., Lindsay Chadderton, W., Wittmann, M. E., et al. 2017. Predicting spread of aquatic invasive species by lake currents. Journal of Great Lakes Research, 43: 14–32.

Beukhof, E., Spaanheden Dencker, T., Pecuchet L., Lindegren, M. 2019a. Spatio-temporal variation in marine fish traits reveals community-wide responses to environmental change. Marine Ecology-Progress Series. 610:205-222.

Beukhof, E., Frelat, R., Pecuchet, L., Maureaud, A., Dencker, T. S., Sólmundsson, J., Punzón, A., et al. 2019b. Marine fish traits follow fast-slow continuum across oceans. Scientific reports, 9. 17878. doi:10.1038/s41598-019-53998-2

Christensen, Asbjørn. 2008, Bank resolved prognoses of sandeel fishing potential in the North Sea”.
Final report for the project ”Fiskeriudsigt for tobis i Nordsøen på bankeniveau. (FIUF, 2005-2007).

Christensen, Asbjørn, Patrizio Mariani, and Mark R. Payne. 2018. A Generic Framework for Individual-Based Modelling and Physical-Biological Interaction. PLoS ONE 13(1): e0189956. https://doi.org/10.1371/journal.pone.0189956

Chen, D., Xue, Y., Fink, D., Chen, S., and Gomes, C. P. 2017. Deep multi-species embedding. IJCAI
Clarke Murray, C., Gartner, H., Gregr, E. J., Chan, K., Pakhomov, E., and Therriault, T. W. 2014. Spatial distribution of marine invasive species: Environmental, demographic and vector drivers. Diversity and Distributions, 20: 824–836.

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

Conley, D. J., Carstensen, J., Aigars, J., Axe, P., Bonsdorff, E., Eremina, T., Hahti, B. M., et al. 2011. Hypoxia is increasing in the coastal zone of the Baltic Sea. Environmental Science and Technology, 45: 6777–6783.

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

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

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

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

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

Elith, J., and Leathwick, J. R. 2009. Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40:
Fitzsimons, J. a. 2011. Mislabling marine protected areas and why it matters-a case study of Australia. Conservation Letters, 4: 340–345.

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

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

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

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

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

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

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

Gray, J. S., Wu, R. S. S., and Or, Y. Y. 2002. Effects of hypoxia and organic enrichment on the coastal marine environment RID B-4203-2010. Marine Ecology-Progress Series, 238: 249–279.
Katsanevakis, S., Wallentinus, I., Zenetos, A., Leppäkoski, E., Çinar, M. E., Oztürk, B., Grabowski, M., et al. 2014. Impacts of invasive alien marine species on ecosystem services and biodiversity: A pan-European review. Aquatic Invasions, 9: 391–423.

Keddy, P. A. 1992. Assembly and response rules: two goals for predictive community ecology. Journal of Vegetation Science, 3: 157–164.

Kolar, C. S., and Lodge, D. M. 2001. Progress in invasion biology: Predicting invaders.

Leppäkoski, E., Gollasch, S., Gruszka, P., Ojaveer, H., Olenin, S., and Panov, V. 2002. The Baltic - A sea of invaders. Canadian Journal of Fisheries and Aquatic Sciences, 59: 1175–1188.

Li, Y., and Shen, Z. 2020. Roles of Dispersal Limit and Environmental Filtering in Shaping the Spatiotemporal Patterns of Invasive Alien Plant Diversity in China. Frontiers in Ecology and Evolution, 8: 1–14.

Lindegren, M., Diekmann, R., and Möllmann, C. 2010. Regime shifts, resilience and recovery of a cod stock. Marine Ecology Progress Series, 402: 239–253.

Lindegren, M., Vigliano, P., and Nilsson, P. A. 2012a. Alien Invasions and the Game of Hide and Seek in Patagonia. Plos One, 7: e44350–e44350.

Lindegren, M., Blenckner, T., and Stenseth, N. C. 2012b. Nutrient reduction and climate change cause a potential shift from pelagic to benthic pathways in a eutrophic marine ecosystem. Global Change Biology, 18: 3491–3503.

Lindegren, M., Thomas, M. K., Jónasdóttir, S. H., Nielsen, T. G., and Munk, P. 2020. Environmental niche separation promotes coexistence among ecologically similar zooplankton species—North Sea copepods as a case study. Limnology and Oceanography, 65: 545-556.

Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., and Bazzaz, F. A. 2000. Biotic invasions: Causes, epidemiology, global consequences, and control. Ecological Applications, 10: 689–710.
Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecological Monographs, 89: 1–24.

Ojaveer, H., Jaanus, A., Mackenzie, B. R., Martin, G., Olenin, S., Radziejewska, T., Telesh, I., et al. 2010. Status of biodiversity in the Baltic sea. PLoS ONE, 5: 1–19.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., et al. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology Letters, 20: 561–576.

Pechenik, J. A. 1999. On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. Marine Ecology Progress Series, 177: 269–297.

Pecuchet, L., Törnroos, A., and Lindegren, M. 2016. Patterns and drivers of fish community assembly in a large marine ecosystem. Marine Ecology Progress Series, 546: 239–248.

Pimentel, D., Zuniga, R., and Morrison, D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics, 52: 273–288.

Radinger, J., Alcaraz-Hernández, J. D., and García-Berthou, E. 2019. Environmental filtering governs the spatial distribution of alien fishes in a large, human-impacted Mediterranean river. Diversity and Distributions, 25: 701–714.

Richardson, D. M., and Pyšek, P. 2012. Naturalization of introduced plants: Ecological drivers of biogeographical patterns. New Phytologist, 196: 383–396.

Rosenberg, R., Hellman, B., and Johansson, B. 1991. Hypoxic Tolerance of Marine Benthic Fauna. Marine Ecology-Progress Series, 79: 127–131. ISI:A1991HF03400010.

Rosenberg, R., Agrenius, S., Hellman, B., Nilsson, H. C., and Norling, K. 2002. Recovery of marine benthic habitats and fauna in a Swedish fjord following improved oxygen conditions. Marine Ecology Progress Series, 234: 43–53.
Shinen, J. S., and Morgan, S. G. 2009. Mechanisms of invasion resistance: competition among intertidal mussels promotes establishment of invasive species and displacement of native species. Marine Ecology Progress Series, 383: 187–197.

Stubbs, W. J., and Wilson, J. B. 2004. Evidence for limiting similarity in a sand dune community. Journal of Ecology, 92: 557–567.

Theoharides, K. A., and Dukes, J. S. 2007. Plant invasion across space and time: Factors affecting nonindigenous species success during four stages of invasion. New Phytologist, 176: 256–273.

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., and Ovaskainen, O. 2020. Joint species distribution modelling with the r-package Hmsc. Methods in Ecology and Evolution, 11: 442–447.

Wood, S. N. 2017. Generalized additive models: An introduction with R, 2nd ed. Chapman and Hall/CRC.

Statements and Declarations

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

Author Contributions

FH and ML designed the study. FH and APG performed drift model simulations. APG performed data extraction. ML performed data processing and statistical analysis. All authors took part in assessing and discussing results, as well as writing the ms.
Table 1. Scientific names and summary statistics of the set of NIS present within the NOVANA monitoring program.

| Species                          | Taxa      | Id | Nr. samples present | Nr. stations present | Probability present (%) | Abundance (mean) | Abundance (max) |
|---------------------------------|-----------|----|---------------------|----------------------|-------------------------|-----------------|-----------------|
| *Amphibalanus improvisus*       | Arthropoda| Ai  | 12                  | 5                    | 0.48                    | 0.03            | 72              |
| *Alitta succinea*               | Anellida  | As  | 141                 | 14                   | 1.87                    | 0.04            | 28              |
| *Crepidula fornicata*           | Mollusca  | Cf  | 4                   | 3                    | 0.16                    | 0.01            | 8               |
| *Ensis directus*                | Mollusca  | Ed  | 54                  | 15                   | 2.15                    | 0.02            | 3               |
| *Marenzelleria neglecta*        | Anellida  | Ma  | 665                 | 32                   | 8.81                    | 0.44            | 136             |
| *Molgula manhattensis*          | Chordata  | Mn  | 1                   | 1                    | 0.04                    | <0.01           | 1               |
| *Mya arenaria*                  | Mollusca  | Mv  | 241                 | 10                   | 9.58                    | 0.38            | 35              |
| *Palaemon elegans*              | Crustacea | Pc  | 273                 | 20                   | 3.62                    | 0.26            | 128             |
| *Polydora cornuta*              | Anellida  | Pe  | 1                   | 1                    | 0.01                    | <0.01           | 1               |
| *Rhithropanopeus harrisi*       | Crustacea | Rh  | 2                   | 1                    | 0.03                    | <0.01           | 2               |
| *Streblospio benedicti*         | Anellida  | Sb  | 55                  | 8                    | 0.73                    | 0.02            | 17              |
Table 2. Summary statistics of the final GAM, including the fixed effect factors species (Sp) and 5-year time period (5y), as well as the smooth terms for the environmental predictors. Edf is the estimated degrees of freedom for the model smooth terms (s) (i.e., edf>1 indicates a non-linear relationship), Chi.sq is the associated chi square statistics and p-values indicate the significance of each predictor (sal=salinity, temB=bottom temperature, ship=vessels activity, spn=settling densities). (See Table 1 for species id and corresponding names).

| Parametric coefficients | Estimate | SE  | z   | P     |
|-------------------------|----------|-----|-----|-------|
| (Intercept)             | -7.80    | 0.48| -16.32 | <0.001|
| Sp(As)                  | 2.15     | 0.48| 4.53  | <0.00  |
| Sp(Cf)                  | -0.90    | 0.67| -1.33 | 0.18   |
| Sp(Ed)                  | 1.56     | 0.49| 3.18  | <0.001 |
| Sp(Ma)                  | 4.61     | 0.45| 10.15 | <0.001 |
| Sp(Mn)                  | -2.27    | 1.10| -2.06 | 0.04   |
| Sp(Mv)                  | 1.90     | 0.48| 3.94  | <0.01  |
| Sp(Pc)                  | 2.83     | 0.47| 6.06  | <0.01  |
| Sp(Pe)                  | 2.24     | 1.10| -2.04 | 0.04   |
| Sp(Rh)                  | -63.51   | 95.00| 0.01 | 1.00   |
| Sp(Sb)                  | 1.65     | 0.49| 3.34  | <0.001 |
| 5y[05-10]               | 0.21     | 0.18| 1.15  | 0.25   |
| 5y[10-15]               | 0.75     | 0.19| 3.91  | <0.001 |
| 5y[15-20]               | 0.61     | 0.21| 2.94  | <0.001 |

| Smooth terms           | edf      | Ref.df | Chi.sq | p-value |
|------------------------|----------|--------|--------|---------|
| s(spn)                 | 1.98     | 2.00   | 56.18  | <0.001  |
| s(ship)                | 1.91     | 1.99   | 15.90  | <0.001  |
| s(temB)                | 1.95     | 2.00   | 24.35  | <0.001  |
| s(sal)                 | 1.98     | 2.00   | 385.22 | <0.001  |
| s(Depth)               | 1.01     | 1.02   | 312.26 | <0.001  |

R-sq.(adj) = 0.447  Deviance explained = 51.9%

UBRE = -0.88238  Scale est. = 1  n = 14757
**Figure 1.** Map of the study area and its position within northern Europe. Points indicate the positions or available sampling station of the NOVANA program used in this study. The size of points represents the sample effort as the total number of available samples for each site across seasons and years.

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

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

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

**Figure 5.** GAM smooth plots showing the partial effects of settling density (a), vessel density (b), bottom temperature (c), salinity (d) and depth (e) on the probability of presence of NIS. Solid lines and shaded areas show the mean and 95% confidence intervals, respectively, and filled circles the partial residuals for each predictor.
Figure 6. Partial dependence plots based on RF showing the partial effects of settling density (a), vessel density (b), bottom temperature (c), salinity (d) and depth (e) on the probability of presence of NIS.

Figure 7. The mean (a, b) and standard deviation (c, d) of predicted probability of presence across all NIS throughout the study area based on the final GAM (left column) or RF (right column). In the lower panels the mean deviation between predicted and observed probability of presence of NIS (ΔP) are shown for each sampling station for GAM (e) and RF (f), respectively. (See Fig.S6 for an overall comparison of predicted vs observed probability of presence).
Identifying key processes and drivers affecting the success of non-indigenous marine species in coastal waters: a joint modelling approach

Martin Lindegren\textsuperscript{1,2,*}, Aurelia Pereira Gabellini\textsuperscript{2}, Peter Munk\textsuperscript{2}, Karen Edelvang\textsuperscript{2}, Flemming Hansen\textsuperscript{2}

\textsuperscript{1}Centre for Ocean Life, National Institute of Aquatic Resources, Technical University of Denmark, Kemitorvet, Building 202, 2800 Kgs. Lyngby, Denmark. \textsuperscript{2}National Institute of Aquatic Resources, Technical University of Denmark, Kemitorvet, Building 202, 2800 Kgs. Lyngby, Denmark.

*Corresponding author: mli@aqua.dtu.dk

Supplementary information

Species traits

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

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

b) Spawning season: In case of missing information of spawning season for Danish conditions, we estimated the likely months of start and end of the spawning season based on either data on minimum
temperature for reproduction or larval survival (if available), or data on spawning season from other
geographical regions and considering the difference in climatic conditions.

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

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

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

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

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

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

As an exception to the other 22 species, the orange striped green sea anemone (*Diadumene lineata*)
does not have sexual reproduction in introduced areas. Dispersal occurs by fission of fragments that
can drift and reattach to substrates at a new location. For this species we estimated “spawning season”
from studies on fission rate dependency on temperature (Flenikken 2017). And we used estimated
PLD from an expected dispersal distance of ca. 10 km (Bering Sea Marine Invasive Species
Assessment), although this fragment can principally drift for many weeks or month. For 3 of the 23 species we did not find sufficient data on traits. These include *Polydora aggregate, Hyperetone heteropoda* and *Fenestruina malusii*. For the species *Teredo navalis*, it inhabits fixed and floating wood and this type of data is not available.

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

**Definition of species habitats**

Habitat maps for each species were created as raster maps with a spatial resolution of 0.01 degrees and covering the spatial extend of 3 – 17 degree east, and 53 – 60 degree north. The classification of preferred seabed substrates into 3 classes (Mud, Sand, and Hard substrate) was based on a reclassification of the seabed substrates from EMODNET (see Table S2). Notice that “hard substrate” and “Sand” both include the EMODNET seabed substrates “Mixed sediments” and “Coarse sediments”. The rationale for this is that these two EMODNET classes represent seabeds with varying compositions of sand, gravel and stones and we make the assumption that those types of habitats support both species associated with sandy and hard substrates. In addition to seabed substrate from EMODNET, for species associated with hard substrates including mussel and oyster shells, and/or
seaweed and seagrass surfaces, we extended the hard substrate category to include all seabed within
the depth interval from 0 – 15 m. Within this depth range mussels, in particular blue mussel (*Mytilus
edulis*), are abundant throughout large parts of the Danish marine and brackish water territory.
Similarly, this depth range support both seagrass and seaweeds. Although both mussel and oyster
abundances as well as coverages of seaweed and seagrasses, may be highly fragmented depending a
number of factors including physical stress at the seabed (e.g. from waves), bottom trawling fishing
activities, light attenuation etc. the exclusion of these biogenic substrates will substantially
underestimate the extent and coverage of potential habitats for certain species. These include
*Amphibalanus improvises, Austrominius modestus, Crepidula fornicate, Diadumene lineata, Molgula
manhattensis and Styela clava.*

**Supplementary references**

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

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

Kristensen E, Banta G T, Quintana C O, Delefosse M, Flindt M, 2012. Hvad ved vi om Svovlorm,
Marenzelleria viridis? Vand og jord. Vol.19, pp 27-32. (in Danish)
Table S1. Traits of 23 benthic species considered as non-indigenous in Danish waters. Species with asterisk denote the NIS present also in the 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 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 |
|--------------------------|------------|-----------|-----------|----------------------|----------------|--------------|---------|-------|-----------------|-----------------|----------------------|-------------------|-----------------|-----------------|------------------|------------------|
| Alitta succinea*         | Anellida   | 10        | 14        | 1                    | June           | Oct          | ALL     | 30    | 0               | 36              | 0.14                 | 80                | 14              | 45              |                  |
| Amphibalanus improvises* | Arthropoda  | 11        | 42        | 1                    | June           | Sept         | HMus    | 30    | -2              | 38              | 0.5                 | 40                | 10              | 30              | 2                | 40               |
| Austrominus modestus     | Crustacea  | 10        | 15        | 1                    | 5              | 10           | HMusW   | 20    | 0               | 26              | 14                  | 40                | 6               | 25              | 25               | 32               |
| Crassostrea gigas        | Mollusca   | 21        | 28        | 1                    | 7              | 8            | A       | 15    | 3               | 35              | 12                  | 42                | 18              | 26              | 10               | 42               |
| Crepidula fornicata*     | Mollusca   | 11        | 32        | 1                    | June           | Sept         | HSMus   | 40    | 0               | 30              | 15                  | 32                | 15              | 30              | 15               | 30               |
| Diadumene lineata        | Cnidaria   | 0         | 10        | 1                    | June           | Sept         | HMusW   | 50    | 0               | 40              | 7                   | 74                | 4               | 74               |                  |
| Ensis directus*          | Mollusca   | 14        | 21        | 1                    | 3              | 4            | SM      | 12    | 0               | 26              | 7                   | 32                | 15              | 28              | 15               | 32               |
| Hemigrapsus sanguineus   | Crustacea  | 16        | 55        | 1                    | 5              | 9            | HS      | 40    | 5               | 30              | 15                  | 33                | 15              | 30              | 20               | 35               |
| Hemigrapsus takanoi      | Crustacea  | 30        | 30        | 1                    | 5              | 9            | A       | 20    | 0               | 20              | 7                   | 35                | 15              | 30              | 25               | 35               |
| Marenzelleria neglecta*  | Anellida   | 28        | 84        | 1                    | Sep            | Sept         | SM      | 50    | 0               | 25              | 0.05                | 10                | 0               | 10              |                  |
| Marenzelleria viridis*   | Anellida   | 28        | 49        | 1                    | 9              | 11           | SM      | 20    | 0               | 25              | 16                  | 32                | 15              | 25              | 16               | 30               |
| Molgula manhattensis*    | Chordata   | 1         | 3         | 3                    | May            | Sept         | HMus    | 40    | 5               | 9               | 40                  | 10                | 9               | 40              |                  |
| Mya arenaria*            | Mollusca   | 14        | 21        | 1                    | May            | June         | SM      | 30    | -2              | 28              | 4                   | 35                | 12              | 28              | 10               | 32               |
| Palaemon elegans*        | Crustacean | 18        | 45        | 1                    | April           | Sept         | ALL     | 10    | 2               | 34              | 1                   | 45                | 14              | 20              | 5                | 40               |
| Petricolaria philadiformis| Mollusca   | 10        | 14        | 1                    | June           | Aug          | ALL     | 30    | 10              | 30              | 10                  | 10                | 30              | 0               |                  |
| Polydora cornuta*        | Anellida   | 12        | 60        | 3                    | May            | Sept         | SM      | 20    | -2              | 29              | 2                   | 75                | 10              | 5               | 75               |                  |
| Rhiithropanopeus harrisi*| Crustacea  | 7         | 43        | 1                    | 6              | 9            | A       | 37    | 0               | 35              | 5                   | 30                | 14              | 27              | 5                | 30               |
| Striblospio benedicti*   | Anellida   | 5         | 40        | 2                    | May            | Oct          | SM      | 20    | -2              | 34              | 4                   | 45                | 7.5              | 30              | 4                | 45               |
| Styela clava             | Chordata   | 1         | 1         | 1                    | Jul            | Aug          | HMusW   | 20    | -2              | 27              | 10                  | 35                | 16              | 20              | 35               |                  |
| Hypereteone heteropoda   | Anellida   | SM        |          |                      |                |              |         |       |                 |                 |                     |                   |                 |                 |                  |
| Polydora aggregate       | Anellida   | 55        | 34        | May                  | Sept           |             | Wood    | 30    | 1               | 30              | 7                   | 40                | 10              | 30              | 5                | 40               |
| Teredo navalis           | Mollusca   | 20        | 34        | Aug                  | Nov            | Wood       |         | 30    | 1               | 30              | 7                   | 40                | 10              | 30              | 5                | 40               |
| Fenestrulina malusii     | Bryozoa    | 1 h       |          |                      |                |              | HMW     |       |                 |                 |                     |                   |                 |                 |                  |
Table S2. EMODNET seabed substrate classified into 3 main categories: “Mud”, “Sand” and “Hard substrates”. Notice that “hard substrate” and “Sand” both include “Mixed sediments” 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         |               |                             |
**Figure S1.** Plot showing the uneven distribution of species abundances (n) across all samples extracted from in the NOVANA data.
Figure S2. Maps of normalized settling densities for NIS derived from drift model simulations.
Figure S3. Habitat maps based on substrate preferences, depth preferences and adult salinity tolerances. Green represents habitats where salinity conditions are within the reported tolerance range of the species. Yellow represent habitats where the salinity conditions may exceed the tolerance thresholds of the species in shorter or longer periods.
Figure S4. Box plots of GAM predictions of probability of occurrence versus the observed classes (0=absence and 1= presence). Please note the high accuracy of predicting absences while predicted presence are highly scattered with a mean error of ~50%.
**Figure S5.** Variable importance of each predictors based on RF estimated as the mean decrease in the Gini index. (sp=species id (factor), sal=salinity, temB=bottom temperature, ship=vessels activity, FiveY=five year periods (factor), spn=settling densities).
**Figure S6.** Scatter plot of observed mean probability of presence across species per sampling station vs the predicted mean probabilities based on either the final GAM (black) and RF (blue). Please note that while the explained variance is similar between methods (i.e., 0.57 and 0.50, respectively), GAM tend to slightly overestimate probabilities (i.e., the slope (b) is >1), while RF tend to underestimate (i.e., b<1). The dashed grey line shows the one to one line. (See Fig. 7 for model predictions and deviations between observed and predicted mean probabilities per station for both methods).