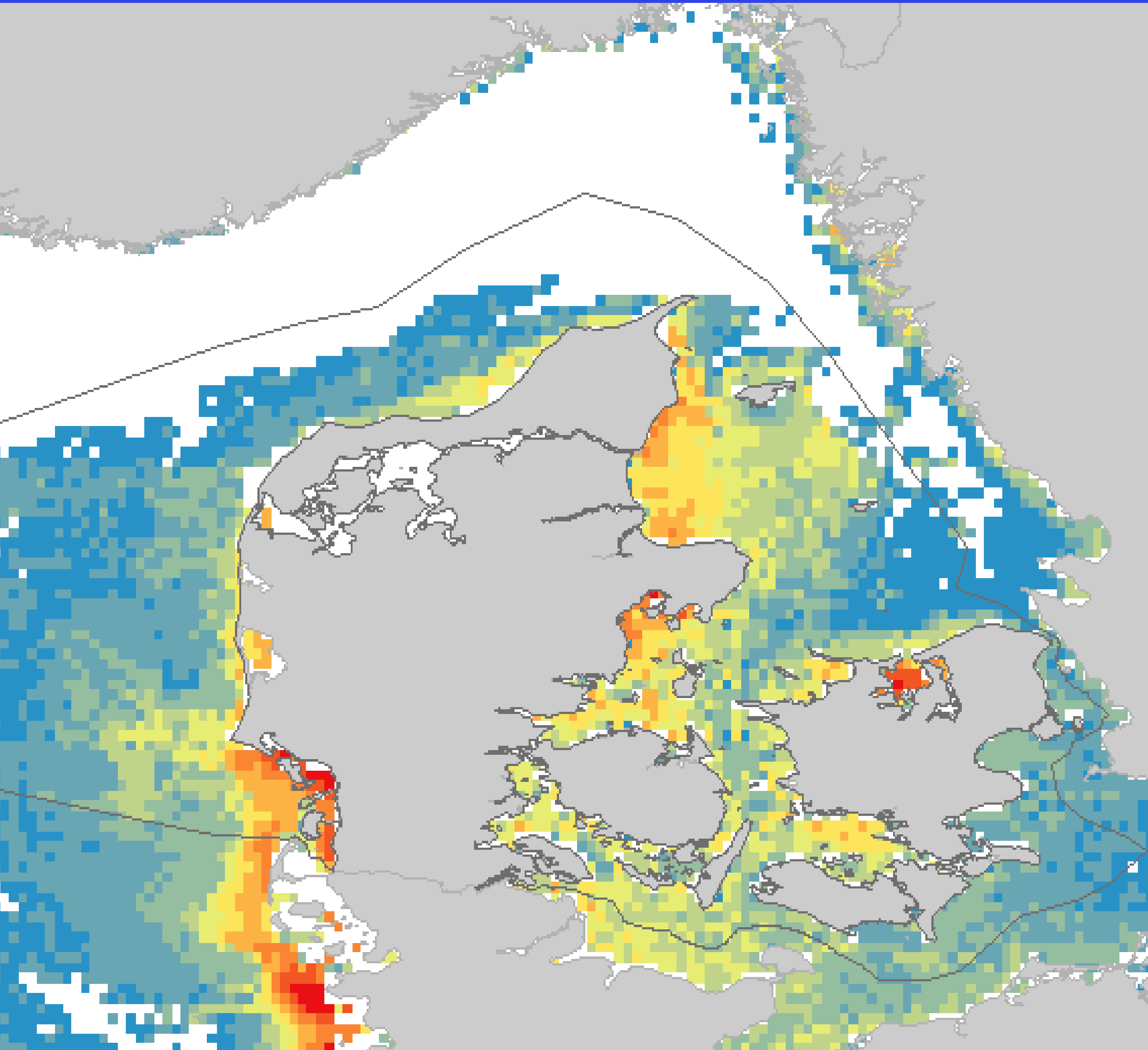
**DTU Aqua**
National Institute of Aquatic Resources

# Dispersal corridors for marine non-indigenous species in Danish waters – analysing modelled settling areas and observed occurrences

Flemming Thorbjørn Hansen, Aurelia Pereira Gabellini, Martin Lindegren, Peter Munk, Asbjørn Christensen and Karen Edelvang

DTU Aqua Report no. 386-2021



DTU Aqua
National Institute of Aquatic Resources

# Dispersal corridors for marine non-indigenous species in Danish waters – analysing modelled settling areas and observed occurrences

Flemming Thorbjørn Hansen, Aurelia Pereira Gabellini, Martin Lindegren, Peter Munk, Asbjørn Christensen and Karen Edelvang

DTU Aqua Report no. 386-2021

## Colophon

Title: Dispersal corridors for marine non-indigenous species in Danish waters – analysing modelled settling areas and observed occurrences

Authors: Flemming Thorbjørn Hansen, Aurelia Pereira Gabellini, Martin Lindegren, Peter Munk, Asbjørn Christensen and Karen Edelvang

DTU Aqua Report no.: 386-2021

Year: May 2021

Reference: Hansen, F.T., Gabellini, A.P., Lindegren, M., Munk, P., Christensen, A. & Edelvang, K. (2021). Dispersal corridors for marine non-indigenous species in Danish waters – analysing modelled settling areas and observed occurrences. DTU Aqua Report no. 386-2021. National Institute of Aquatic Resources, Technical University of Denmark. 43 pp. + appendices

Cover: Part of map showing average settling densities of all 19 species covered by the report.

Published by: National Institute of Aquatic Resources, Kemitorvet, 2800 Kgs. Lyngby, Denmark

Download: www.aqua.dtu.dk/publikationer

ISSN: 1395-8216

ISBN: 978-87-7481-309-5

**DTU Aqua Reports** contain results from research projects, reviews of specific topics, expositions for authorities etc. Unless stated in the colophon, the reports are not peer reviewed, which means that the content has not been reviewed by researchers outside the project group.

# Contents

# 1 Abstract

In the marine environment, we find a wide range of non-indigenous species. Some of these have stages of pelagic drift and the dispersal of such “introduced non-indigenous species (NIS) with pelagic stages” is dependent on both species characteristics, habitat availability and hydrographic conditions. The NIS may be highly concentrated in some marine areas while their presence is diminished or completely absent in others. Their spread between areas is determined by the corridors through which their pelagic stages are dispersed. There is a need for further understanding of these corridors in order to implement protective actions against the spread of NIS in Danish marine waters. This is specifically relevant for ballast water management with regard to ship traffic. In this respect, the Danish Environmental Protection Agency (DEPA) is assessing exemption applications under the Ballast Water Management Convention's (BWMC) exemptions rules. Information on the potential occurrences of NIS as well as on the transport corridors of their pelagic stages is thus important in support of the risk assessment.

The estimation of potential primary dispersal corridors for marine NIS may be used in an evaluation of existing national monitoring stations in Danish waters, and whether these are located in areas that are relevant for monitoring of changes in the overall NIS abundances. If the monitoring stations are outside areas of major NIS presence, the monitoring is not adequate to effectively be used for early warnings on changes in abundances, thereby having consequences for the treatment of exemption applications. Finally, the modelled dispersal corridors may contribute to the understanding of the presence/absence of NIS in the national monitoring program by an analysis of distances between national monitoring stations and main shipping routes through Danish marine waters.

In the present study, we aim at identifying the potential dispersal corridors for selected marine non-indigenous species (NIS) in the Danish marine waters, subsequently examining whether these dispersal corridors coincide with areas where NIS observations are conducted during the national monitoring programme, NOVANA or where the main shipping routes are positioned. We specifically investigate whether dispersal corridors and simulated dispersal- and settling densities of NIS (estimated from our drift model simulations) can explain the observed presence or absence of 19 selected species recorded across the NOVANA stations.

The results show clear differences in the overall mean probability of presence between species as well as between decades of analyses, where estimates for the recent decades indicate an overall increased probability of presence of NIS over time. This in turn indicates a geographical spread of NIS across the area in recent decades.

Our analyses also show that the presence of NIS in Danish waters is affecting a range of factors regarding both their colonization and their establishment. The colonization process is related to the drift and settlement of larvae, described by our drift model simulations and/or the density of vessels acting as potential mediators of dispersal. The establishment process is determined primarily by the environmental conditions and the sea bed habitat at each site, depending on its capacity to allow for survival and reproduction. The modelled settling density is related to the observed NIS occurrences, thus strengthening the potential information value by combining drift modelling and actual observations in future risk assessments.

In conclusion, the presented model studies of NIS dispersal have shown that there is a large potential for improving the general understanding of NIS presence and dispersal in Danish waters, demonstrated by combining the analysis of model output and different environmental variables and comparing them to observed NIS occurrences. We recommend continuing the analysis on these data, specifically with regard to the assessment of environmental linkages when identifying sites of primary importance for future NIS monitoring.

# 2 Dansk resume

Marine ikke-hjemmehørende arter (her kaldet NIS) spredes dels ved direkte introduktion, ofte via skibes ballastvand eller fastsiddende på deres skrog, dels ved naturlig spredning, bl.a. drevet af strømforhold, arternes biologiske karakteristika og deres krav til levesteder. Den naturlige spredning af marine organismer med havstrømmene er dog ikke ensartet fordelt i de marine områder. Kombinationen af varierende havstrømme, arternes spredningsadfærd samt deres krav til fysiske levesteder og tolerance overfor f.eks. saltholdighed og temperatur betyder, at spredningen vil være særligt intens i nogle dele af de marine områder, i de såkaldte spredningskorridorer.

Viden om hvordan NIS spredes i de danske farvande har stor betydning for Miljøstyrelsens vurdering af dispensationsansøgninger i henhold til ballastvandkonventionens dispensationsregler (A4 og Guideline 7), hvor en skibsejer kan søge dispensation for kravet om at rense ballastvand. Blandt forudsætningerne er viden om NIS' forekomst i farvande og havne samt kendskab til spredningskorridorerne.

I nærværende projekt kortlægges de potentielle spredningskorridorer for udvalgte NIS arter i danske farvande, og vi undersøger, om der er en sammenhæng mellem disse og 1) de primære skibsruter og 2) de observationer af NIS, der er gjort i perioden 1979-2020 gennem det nationale overvågningsprogram af marine arter, NOVANA. Specifikt undersøger vi, om modellerede spredningskorridorer og udbredelsesområder for 19 af de almindeligste NIS arter er samstemmende med skibsruter og/eller observerede forekomster bestemt under NOVANA programmet.

I vores simuleringer fandt vi tydelige forskelle i forekomsten af de enkelte NIS arter samt forskellige mønstre for, hvorledes deres forekomst har ændret sig gennem tiden. Generelt var forekomsterne af NIS dog stigende, og vi så også større udbredelse af NIS hen over den undersøgte 40 års periode. Vores resultater viser tydeligt, hvorledes forekomsten af NIS i danske farvande styres af en række forhold, der både har betydning for deres kolonisering af et område og for den måde, hvorpå de etablerer sig mere permanent. Selve koloniseringen er forbundet med transport og settling af NIS' larver, som beskrives ved vores modelsimuleringer, og/eller ved beregninger af skibenes funktion som medie for spredning. Etablering af NIS i givne områder er derimod hovedsagelig bestemt af de specifikke miljøforhold og bundhabitaternes egnethed for arternes vækst og overlevelse. Vi finder en sammenhæng mellem den modellerede spredning og de observerede forekomster af NIS, hvilket sandsynliggør at en kombination af driftmodellering og observationer kan være af stor værdi ved fremtidige risikovurderinger.

Sammenfattende fandt vi, at de præsenterede modelstudier af NIS' spredning gav bedre forståelse for NIS forekomst i de danske farvande, og at modelleringen i kombination med aktuelle observationer fra det nationale overvågningsprogram kan være en stor forbedring af den eksisterende risikovurdering. Vi ser derfor et stort potentiale i den demonstrerede metode, der inddrager modellering, statistisk analyse samt aktuelle observationer, og vi anbefaler yderligere analyser, der specifikt fokuserer på identifikation af de miljøforhold, der kan pege på særligt vigtige områder for den fremtidige NIS monitering.

# 3 Background

This project was funded by the Danish maritime fund “Den Danske Maritime Fond” on request from the Danish Environmental Protection Agency, *Miljøstyrelsen* (MST). The project was designed to analyse how marine non-indigenous species (NIS) are spread in the Danish marine environment through direct introduction via e.g. ballast water or attached to the ships' hulls, or by natural dispersal, e.g. driven by the flow conditions, the biological characteristics of the species as well as their habitat requirements.

The natural spread of marine organisms with sea currents depends on the combination of varying sea currents, the individual species' dispersal behaviour as well as the requirements for physical habitats and tolerance towards e.g. salinity and temperature. This means that the dispersal of NIS will be more intense in some parts of the Danish marine areas than others - in some areas it may even be completely absent. The hypothesis is that marine NIS most likely are present in or around marine dispersal corridors. Therefore, an in-depth knowledge of how NIS spread in the Danish waters is of great importance for the Danish Environmental Protection Agency's assessment of exemption applications in accordance with the Ballast Water Management Convention's exemption rules.

Knowledge of the potential primary dispersal corridors for marine NIS would further be of great value for an evaluation of the existing monitoring stations (NOVANA program) in Danish waters, in order to assess whether these are adequately located in or near these dispersal corridors. If not, then there is a risk that new invasive species are not registered, even though they may already be introduced to the Danish waters. In this way, the basis for a possible rejection an exemption application could be inadequate. A mapping of the potential primary dispersal corridors will also show the extent to which the observations of marine NIS in previous decades have been adequate in monitoring areas near the primary shipping routes through Danish waters.

# 4 Introduction

The dispersal of many marine benthic invertebrates is limited to the pelagic larval stage (plankton stage), and knowledge on processes and larval distributional patterns is crucial for the understanding of their population dynamics and connectivity. The dispersal has great implications for marine conservation initiatives (Jones, Srinivasan & Almany, 2007; Lipcius et al., 2008; Gaines et al., 2010; Toonen et al., 2011).

Plankton organisms are dispersed by sea currents in a complex manner. The combination of the sea current variation and the behaviour of the pelagic life stages, guided by their requirements for physical habitats and their tolerances to e.g. salinity and temperature, leads to a patchily distribution of species. Dispersal of “introduced NIS with pelagic stages” will also be dependent on species characteristics and actual oceanographic conditions; therefore enhanced concentrations are observed in some parts of the marine waters while their abundance in other areas are less or they are completely absent. Knowledge of corridors through which the larvae of NIS are dispersed is essential for the implementation of protective actions by governmental agencies.

The Danish Environmental Protection Agency (DEPA) is obliged to act in relation to the assessment of exemption applications under the Ballast Water Management Convention's (BWMC) exemptions rules (A4 and Guideline 7). According to the exemptions rules, the owner of a ship operating exclusively in fixed route(s) between two or more international harbours can apply for an exemption to the *de facto* requirement to treat ballast water, and thus get a permission to release untreated ballast water into the marine environment. Any exemption is to be granted by the national authorities of the involved countries. Prerequisites for granting an exemption is that a monitoring program has been conducted in each harbour, to ascertain whether specific NIS have been found in either all or none of the ports included in an exemption application. Alternatively, an exemption can be granted within the geographical area where the harbours in question are located. This on the condition that a risk assessment has concluded showing that the natural dispersion is so that the risk of dispersal of NIS would not be significantly increased by uptake and release of ballast water inside the area. Such area is referred to as a same-risk-area or SRA. An example is the exemption of Øresund as a SRA between Denmark and Sweden.

Estimates of potential primary dispersal corridors for marine NIS may also be used in an evaluation of whether the existing national monitoring stations in Danish marine waters are located in areas that effectively allow for monitoring of changes in NIS abundances. If the stations fail to do so, early warnings about changes will be missed, with consequences for the treatment of exemption applications. Further, information on dispersal corridors and observations of marine NIS may be used in an evaluation of the distances between observations and main shipping routes through Danish marine waters.

**Project aims**

The aims of the present project are to identify the potential dispersal corridors for selected marine NIS in Danish marine waters; and to examine whether these dispersal corridors coincide areas of main shipping routes and observations of marine NIS carried out through the national monitoring programme, NOVANA.

Specifically, we aim to:

1. **Develop** new algorithms for mapping planktonic organism densities in Danish waters using DTU Aqua's agent-based model system (ABM)

2. **Calculate** the dispersal of NIS and identify potential dispersal corridors, using the developed algorithms

3. **Evaluate** dispersal corridor estimations by statistical comparison of model-estimates and actual observations reported in the National Monitoring Programme (NOVANA) and in data from Ship traffic (AIS)

4. **Present** conclusions and recommendations for the future strategy on monitoring NIS in Danish waters using the outcome of analyses 1-3

# 5 Methods

## 5.1 Species selection and description

**Species selection**
The potential species for this study were selected using the following criteria:

- They must be present in the list of NIS found in Danish waters in the marine monitoring program NOVANA (data sheet provided by the Danish EPA)
- They must possess a benthic adult stage
- They must possess a pelagic dispersal phase

We focused on NIS already identified in Danish waters due to a recommendation from DEPA. Thus, this study can provide more information on NIS that have successfully invaded Danish waters or have a great potential to do so since some specimens have been found in the area. In the agent-based model (ABM) used in this study, particles (drifters) are considered passive drifters and disperse naturally with the sea currents. Therefore, we do not model fish as they can choose direction and can have speed greater than the currents. Hence, we targeted species that depend on one life cycle to disperse naturally, considering eggs and larvae passive drifters.

Based on the above criteria, 23 benthic NIS with a pelagic phase were found in the marine monitoring program NOVANA provided by DEPA. Seven of these were already studied in the "*Same Risk Area Case Study Kattegat and Øresund*" (Hansen and Christensen, 2018). The seven species already analyzed are *Austrominius modestus, Crassostrea gigas, Ensis directus, Hemigrapsus sanguineus, Hemigrapsus takanoi, Marenzelleria viridis* and *Rhithropanopeus harrisii.*

**Species traits**
Data on species traits was collected for all 23 species. Traits included are:

- Spawning period
- Pelagic larvae duration
- Substrate required to settle
- Preferred depth
- Temperature and salinity tolerances for adults and larvae

The information on traits was collected from various data sources including data portals - NEMESIS (Fofonoff et al., 2003), CABI (2020), AquaNIS (2020), NOBANIS (2020) and OBIS (2020) - as well as from peer review publications and reports. Much of the traits' data refers to studies from other parts of the world from the native range of the species or at a given area of introduction, or from laboratory test. Therefore, they may not in all cases be directly applicable to the Danish conditions.

In case sufficient or applicable information was not found in the literature, the traits were estimated based on a range of assumptions. These included:

- **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:
  - 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.

- **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 ABM simulation.

- **Habitat substrates**: In general, data on substrate preferences were available for all species. We assigned each species to one or more of four selected classes of habitat substrates including three geological substrates: *Mud, Sand and Hard* substrate, and two biogenic substrates "*mussels and oyster shells"* and "*Seaweed and/or seagrasses*" merged into a fourth category.

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

- **Temperature, adults:** In general, data was available for the range of temperature where the species are found together with information published on temperature tolerances from laboratory experiments.

- **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.

- **Salinity, adults**: In general, data was available for the range of salinities where the species are found together with information published on salinity tolerances from laboratory experiments.

- **Salinity, larvae**: For some species, data was found on larval salinity tolerance range from laboratory experiments. In cases where data was not available, 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 undergo sexual reproduction in the introduced areas. Dispersal occurs by fission of fragments that can drift and reattach to substrates at a new location. For this species, we therefore estimated "spawning season" from studies on fission rate dependency on temperature (Flenikken, 2017) and used an estimated PLD from an expected dispersal distance of ca. 10 km (Alaska Center for Conservation Science, 2017), although this fragment can principally drift for many weeks or month.

For three of the 23 species, we did not find sufficient data on traits, these were *Polydora aggregata, Hypereteone heteropoda* and *Fenestrulina malusii.* Furthermore, the species *Teredo navalis* inhabits fixed and floating wood and this type of data is not available. The remaining 19 species were subsequently used in the analyses.

In relation to the seven species where data on traits was collected as part of the Same-Risk-Area Case Study for Kattegat and Øresund (Hansen and Christensen, 2018), we revised the traits for two of the species according to the assumption and criteria as described above. Regarding *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 (0-15 m) where mussels, oysters and seaweed are abundant. Regarding *Marenzelleria viridis,* we changed the lower salinity tolerance from 1 to 16 PSU, due to some confusion in previous reports distinguishing between *M. viridis* and *M. neglecta*. We also changed the preferred depth from 63 to 20 m to comply with the OBIS database.

Finally, we changed 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 array of trait information on NIS species is given in Table 1.

Species with an upper larval salinity tolerance of 30 PSU or more and with a minimum salinity tolerance of 10 or higher were considered genuine marine species; these species were assumed not to be limited by the higher salinity conditions in the North Sea.

**Table 1. Overview of the 23 species' trait information. Habitat abbreviations: ALL=all habitat types; M= muddy; S= sandy; H=Hard; Mus= mussel and oyster beds; W = seaweed and seagrasses. The four species numbers 20-23 are not included in the study due to lack of data.**

| | Species | Taxon | PLD (min) | PLD (max) | Generations | Spawning start | Spawning end | Habitat | Depth | Temp. Min | 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 | | m | C | C | PSU | PSU | C | C | PSU | PSU |
| 1 | Alitta succinea | Annelida | 10 | 14 | 1 | June | Oct | ALL | 30 | 0 | 36 | 0.14 | 80 | | | 14 | 45 |
| 2 | Amphibalanus improvisus | Arthropoda | 11 | 42 | | June | Sept | H, Mus | 30 | -2 | 38 | 0.5 | 40 | 10 | 30 | 2 | 40 |
| 3 | Austrominius modestus * | Crustacea | 10 | 15 | 1 | May | Oct | H, Mus, W | 20 | 0 | 26 | 14 | 40 | 6 | 25 | 25 | 32 |
| 4 | Crassostrea gigas * | Mollusca | 21 | 28 | 1 | July | Aug | ALL | 15 | 3 | 35 | 12 | 42 | 18 | 26 | 10 | 42 |
| 5 | Crepidula fornicata | Mollusca | 11 | 32 | 1 | June | Sept | H, S, Mus | 40 | 0 | 30 | 15 | 32 | 15 | 30 | 15 | 30 |
| 6 | Diadumene lineata | Cnidaria | 0 | 10 | | June | Sept | H, Mus, W | 50 | 0 | 40 | 7 | 74 | | | 7 | 40 |
| 7 | Ensis directus | Mollusca | 14 | 21 | 1 | Mar | April | S, M | 12 | 0 | 26 | 7 | 32 | 15 | 28 | 15 | 32 |
| 8 | Hemigrapsus sanguineus | Crustacea | 16 | 55 | 1 | May | Sept | H, S | 40 | 5 | 30 | 15 | 33 | 15 | 30 | 20 | 35 |
| 9 | Hemigrapsus takanoi | Crustacea | 30 | 30 | 1 | May | Sept | ALL | 20 | 0 | 20 | 7 | 35 | 15 | 30 | 25 | 35 |
| 10 | Marenzelleria neglecta | Annelida | 28 | 84 | 1 | Sept | Sept | S, M | 50 | 0 | 25 | 0.05 | 10 | | | 0 | 10 |
| 11 | Marenzelleria viridis | Annelida | 28 | 49 | 1 | Sept | Nov | S, M | 20 | 0 | 25 | 16 | 32 | 15 | 25 | 16 | 30 |
| 12 | Molgula manhattensis | Chordata | 1 | 3 | 3 | May | Sept | H, Mus | 40 | 5 | | 9 | 40 | 10 | | 9 | 40 |
| 13 | Mya arenaria | Mollusca | 14 | 21 | 1 | May | June | S, M | 30 | -2 | 28 | 4 | 35 | 12 | 28 | 10 | 32 |
| 14 | Palaemon elegans | Crustacea | 18 | 45 | 1 | April | Sept | ALL | 10 | 2 | 34 | 1 | 45 | 14 | 20 | 5 | 40 |
| 15 | Petricolaria pholadiformis | Mollusca | 10 | 14 | 1 | June | Aug | ALL | 30 | | | 10 | 30 | | | 10 | 30 |
| 16 | Polydora cornuta | Annelida | 12 | 60 | 3 | May | Sept | S, M | 20 | -2 | 29 | 2 | 75 | 10 | | 5 | 75 |
| 17 | Rhithropanopeus harrisii | Crustacea | 7 | 43 | 1 | June | Sept | ALL | 37 | 0 | 35 | 5 | 30 | 14 | 27 | 5 | 30 |
| 18 | Streblospio benedicti | Annelida | 5 | 40 | 2 | May | Oct | S, M | 20 | -2 | 34 | 4 | 45 | 7.5 | 30 | 4 | 45 |
| 19 | Styela clava | Chordata | 1 | 1 | 1 | Jul | Aug | H, Mus, W | 20 | -2 | 27 | 10 | 35 | 16 | | 20 | 35 |
| 20 | *Hypereteone heteropoda* | *Annelida* | | | | | | *S, M* | *20* | | | *18* | *30* | | | | |
| 21 | *Polydora aggregata* | *Annelida* | *55* | | | *May* | *Sept* | | *20* | | | | *40* | *6* | | | |
| 22 | *Teredo navalis* | *Mollusca* | *20* | *34* | | *Aug* | *Nov* | *Wood* | *30* | *1* | *30* | *7* | *40* | *10* | *30* | *5* | *40* |
| 23 | *Fenestrulina malusii* | *Bryozoa* | *1 h* | | | | | *H, M, W* | *110* | | | | | | | | |

### Definition of species habitats

Species habitat maps for the 19 selected species were prepared as described in Hansen et al. (2020) based on species requirements on seabed substrate, water depth and adult salinity tolerance. 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° East, and 53° – 60° North.

### Sea bed substrate

The classification of preferred seabed substrates into three classes (Mud, Sand, and Hard substrate) was based on a reclassification of the seabed substrates from EMODNET given in Figure 1, where “Hard substrate” and “Sand” both include the EMODNET seabed substrates “Mixed sediments” and “Coarse substrate” (see Table 2). The rationale for this is that since these two EMODNET classes represent seabed with varying compositions of sand, gravel and stones, we assume that both types of habitats support species associated with sandy and hard substrates based on the 23 species analyzed in this study. We used the map produced in October 2016 with 1:250 000 scale (EMODnet-Geology, 2020).

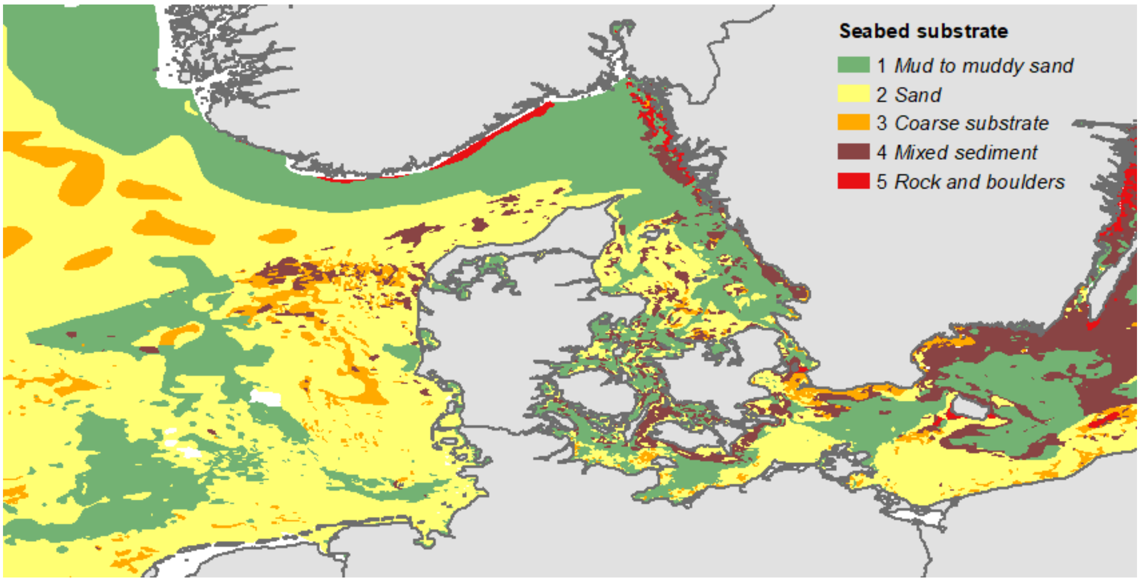


**Figure 1. Seabed substrate of the study area extracted from EMODnet with 1:250 000 scale (EMODnet-Geology, 2020).**

In addition to the given seabed substrate derived from EMODNET, we further extended the hard substrate category to include all seabed within the depth interval from 0 – 15 m for species associated with hard substrates including mussel and oyster shells, and/or seaweed and seagrass surfaces. The reason for this is that 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 is relevant to both seagrass and seaweeds. Both mussel and oyster abundances, as well as coverages of seaweed and seagrasses, may be highly fragmented. This fragmentation depends on several factors including physical stress at the seabed (e.g. from waves), bottom trawling fishing activities, light attenuation etc.

If we were to exclude these biogenic substrates, we would substantially underestimate the extent and coverage of potential habitats for certain species including *Amphibalanus improvisus, Austrominius modestus, Crepidula fornicata, Diadumene lineata, Molgula manhattensis* and *Styela clava.*

**Table 2. Definition of seabed types used to identify species habitats.**

| **Mud** | **Sand** | **Hard substrate** | **Mussel/Sea weed** |
|---|---|---|---|
| Mud | Sand | Rock and boulders | *Depth less than 15 m* |
| Sandy mud | Coarse substrate | Coarse substrate | |
| Muddy sand | Mixed sediment | Mixed sediment | |

In short, based on the EMODnet seabed substrates, we developed one map for mud, which corresponds to species that inhabit fine mud to muddy sand; another for sand, which corresponds to species that inhabit sand, coarse and mixed sediment; and a third for hard substrates, which include coarse, mixed and rock and other hard substrate. Furthermore, based on depth (defined as the maximum penetration depth for light), we created a fourth map for mussels and seaweed, which corresponds to areas shallower than 15 m. Hence, there is an overlap with our definition of sand and hard substrate, since both categories include coarse and mixed sediment, and there is an overlap when considering the mussels (based on depth) and the categories related to the mud, sand and hard substrate. In Figure 2, we show each substrate map derived from Figure 1 individually. In case species had more than one substrate, we merged these maps to create their specific substrate requirements.

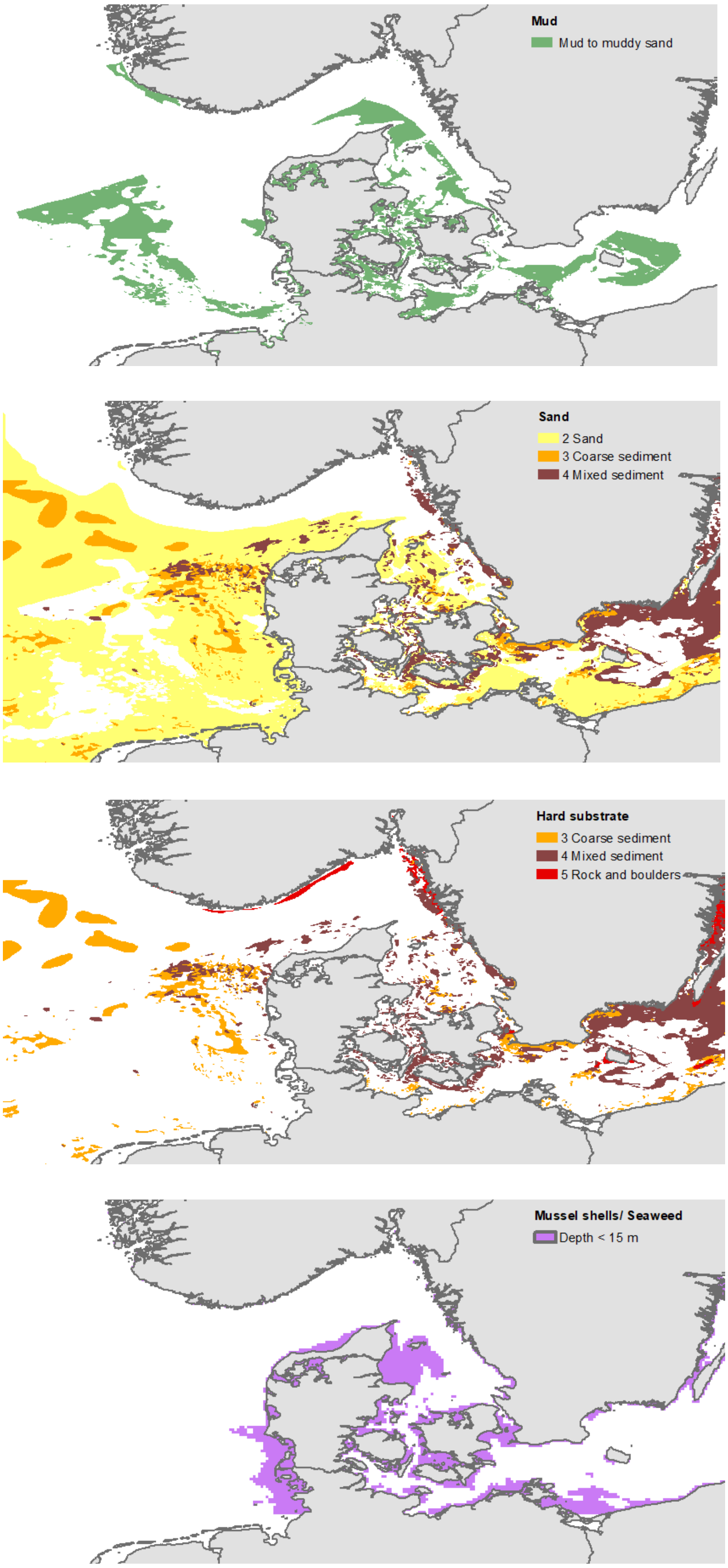


**Figure 2. Seabed substrates used to identify species habitats based on EMODNET (mud, sand and hard substrate) and depth (less than 15 m) used to represent seabed requirement of mussel shells, seaweed and sea grass.**

## Depth

The water depth was also used in the identification of potential habitats for the 19 species. Data on water depth was based on GEBCO bathymetry data set (GEBCO, 2020) (see Figure 3). All the species selected are limited to 50 m (Table 1).

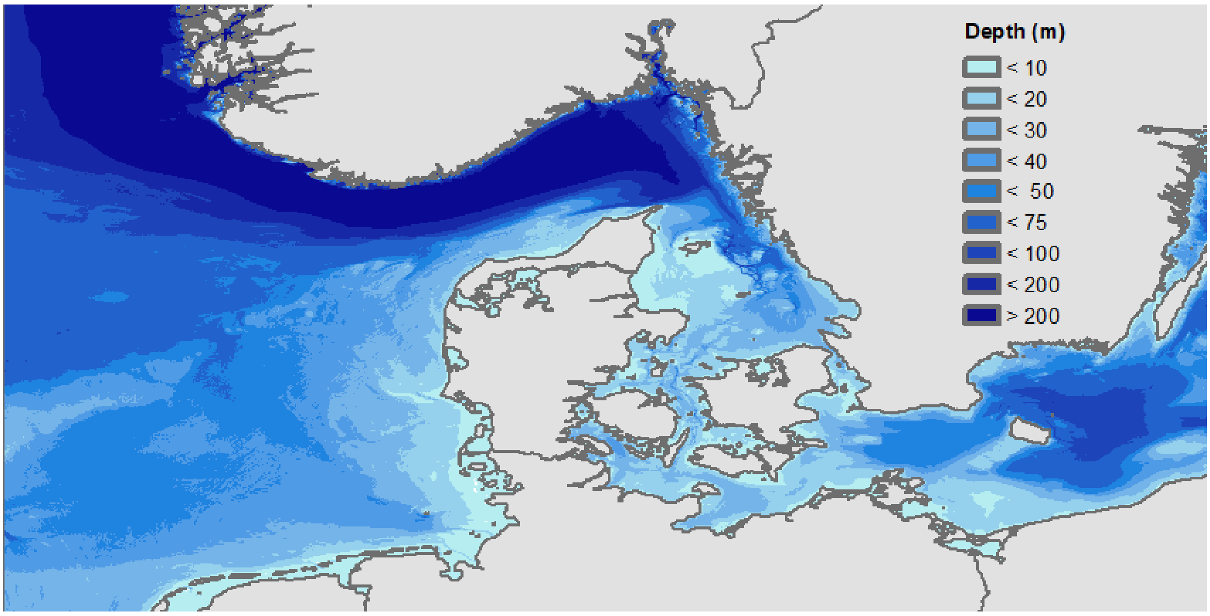


**Figure 3. Depth contours for Eastern North Sea and Western Baltic. Data extracted from GEBCO (2020).**

### Salinity tolerance

In order to assess to which extent the salinity tolerance was limiting the species distribution, we extracted minimum and maximum values of bottom salinity for the years 2005, 2010 and 2012 from the hydrographic HBM model (see Chapter 5.2.1) and interpolated between extraction points using inverse distance weighted interpolation. The three years were selected based on the (salinity) patterns related to North Atlantic Oscillation index (NAO index), namely 2005, 2010 and 2012 representing a “neutral”, a “negative” and a “positive” NAO index respectively, to reflect the expected range of salinity in the region (Hansen and Christensen 2018; Hansen et al, 2020).

## 5.2 Larval dispersal modelling

### Setup of the dispersal model

The modelling approach for simulating larval dispersal in this study is similar to the one used in a previous study by Hansen et al. (2020) *Ranking of Danish ports according to shipping activities and to the potential of natural dispersal of non-indigenous species.* Therefore, only a brief description of the model setup will be presented in the current study. For more detailed information please refer to Hansen et al. (2020).

The model is set up using the hydrodynamic data set from the HBM model (Berg and Poulsen, 2012) covering the North Sea, The Baltic Sea, Kattegat and the inner Danish straits. Data include current velocities, salinity and temperature. The spatial resolution of the model is 0.5 nm[1] 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.

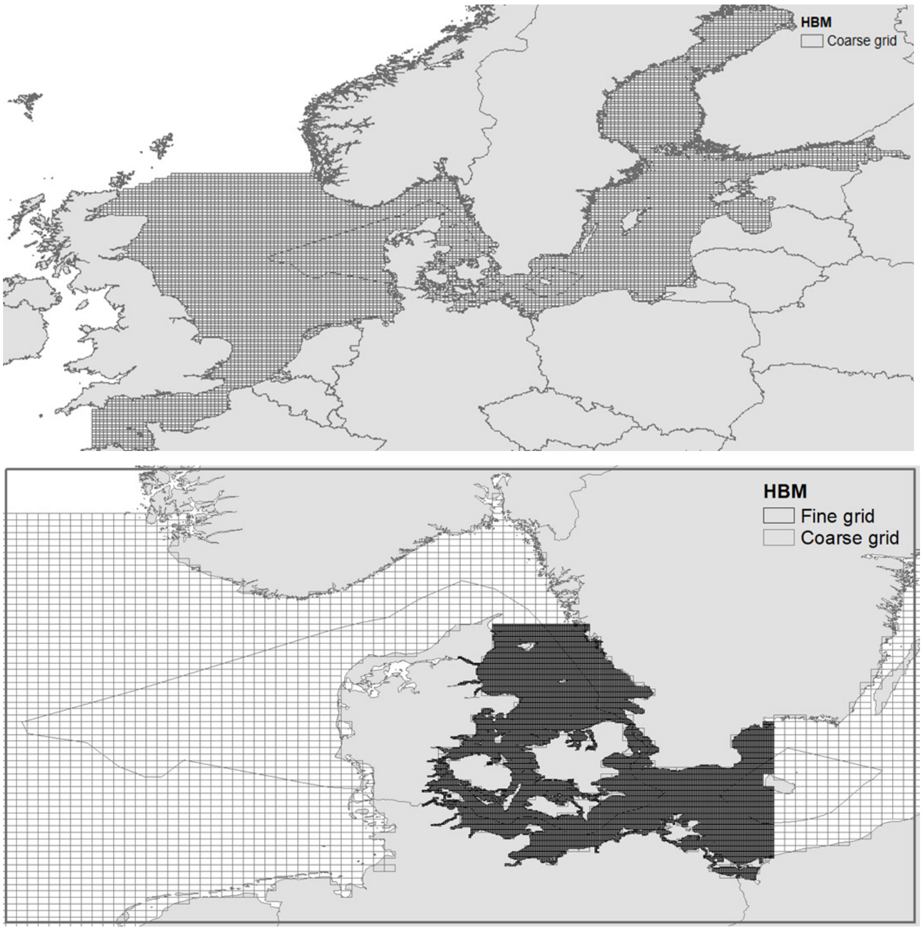


**Figure 4. Top: The full model extension. Bottom: the study area showing both the coarse and the fine grid.**

Figure 4 (top) shows the extent of the whole model area, and Figure 4 (bottom) the boundaries in the model set up for this study. The model used is the agent-based modelling system IBMlib (Christensen, 2008, Christensen et al. 2018). Simulations were conducted for the 19 species (Table 1) during a ten-year period 2001-2010 using a new and modified version of the IBM lib software to include the mapping of the density of agent trajectories during the simulation. Species habitat maps for the 19 species were prepared as described in Hansen et al. (2020) based on species requirements on seabed substrate, water depth and adult salinity tolerance.

---

[1] 1 nm is 1,852 m

The setup for each of the 19 species and for each year (in total 190 simulations) included introducing a number of agents (to simulate the spreading of eggs or larvae), 200,000 agents per year (in total 2,000,000 agents for all 10 years per species). In each simulation, the agents were distributed randomly in space within the maximum area coverage of the species habitat map, and randomly in time within the spawning period using a time step of 1,800 seconds.

**Table 3. Parameters in model setup.**

| | |
|---|---|
| **Number of species simulated** | 19 |
| **Years simulated** | 10 (2001-2010) |
| **Number of simulations per species** | 10 (one per year) |
| **Number of simulations** | 190 (19 species, 10 years each) |
| **Number of agents per simulation** | 200,000 |
| **Time step model** | 1,800 seconds |

### Tracking densities

The mapping of the dispersal trajectories of all 200,000 agents per simulation was carried out by registering the presence of each agent each time step in a grid dividing the study area into smaller subareas. The study area (see Figure 5), between 3° to 17°E and 54° to 60°N, was divided into  280 columns and 120 rows. The sum of agents in any given sub-areas illustrates the "*dispersal densities*" indicating the total number of agents in a given area.

All trajectories outside the larval salinity tolerance range for each species were ignored and not included in the mapping. Species with an upper larval salinity tolerance of 30 PSU or more and with a minimum salinity tolerance of 10 or higher were considered genuine marine species and these species were assumed not to be limited by the higher salinity conditions in the North Sea.

Maps were also produced illustrating the sum of the track densities of all the 19 species for all years. All maps were created using track counts translated into counts per area unit (km$^2$) using the Lambert Azimuthal Equal Area projection.

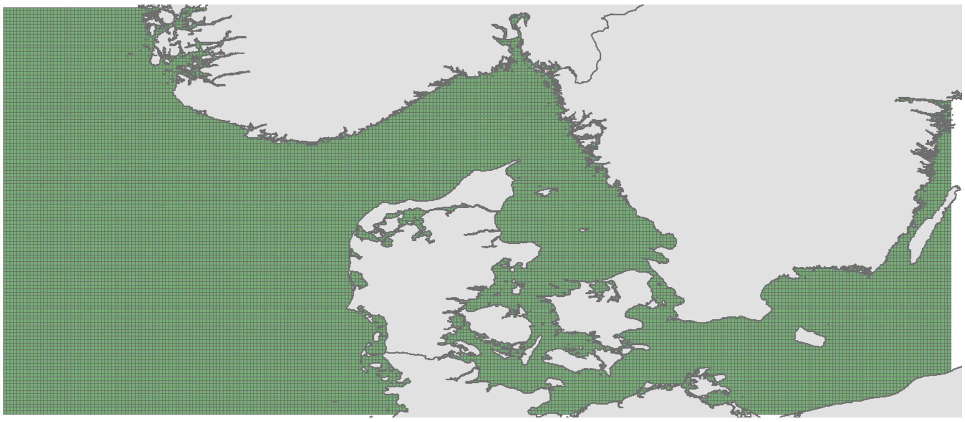


**Figure 5. Study area with the grid used in the study for the track densities.**

**Settling densities**

Maps illustrating all end positions of successful settled larvae were produced. Successfully settled larvae are defined as larvae able to settle within their natural habitat taking into account not only the salinity range experienced in the trajectory, but also the habitat requirements (depth and substrate type) at the time of settlement. This summed to "*settlement densities*".

All maps were created using track counts calculated as counts per area unit (km$^2$) using the Lambert Azimuthal Equal Area projection.

## 5.3 Field observations (NOVANA data)

In order to compare the modelled trajectories to the NOVANA data observations at the individual stations, available data were extracted describing the abundance of the 19 NIS registered in the NOVANA program from the ODA database (Danish Centre for Environment and Energy, accessed 15 October 2020). Records were extracted from all months and years available covering the time period from 1979 to 2020. Data on 15 out of the 19 species were found in the database. Since the hydrographic model is not reliable in fjords, data was only extracted from sites outside the fjords, A total of 68 sites were used for the statistics (Figure 6).

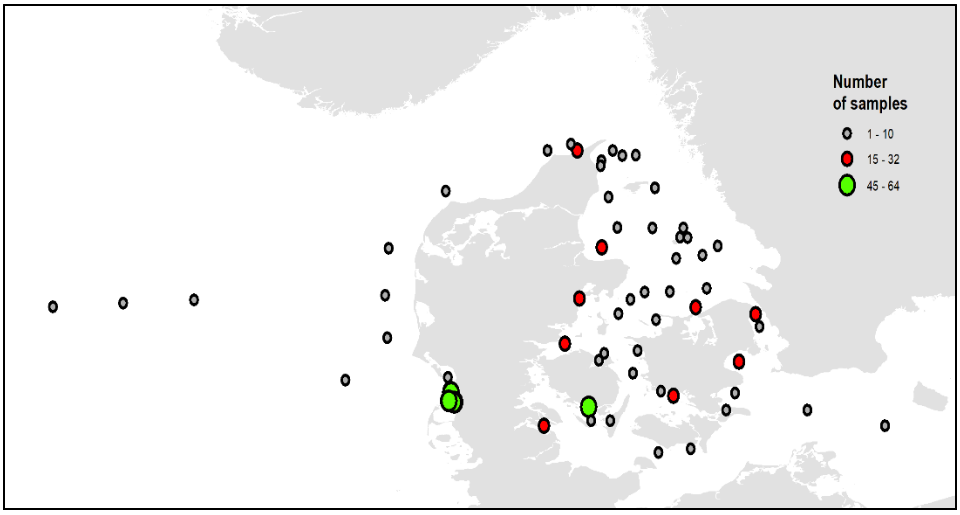


**Figure 6. Positions of the 68 sites from NOVANA used in the statistics. The areas of circles illustrate the number of samples on each site.**

## 5.4 Method for statistical analyses

### Input data

For each of the species represented in the NOVANA database, the simulated *dispersal densities* and *settling densities* were extracted from the larval dispersal model. In addition, the occurrence of NIS found at each station were plotted as an overlay to the *dispersal density* and *settling density* maps for comparison.

Shipping activities were extracted from EMODnet for all type of vessels in 2019 (EMODNET Human activities, 2020). Data on shipping activities from AIS data were also extracted from locations in the 100 meters radius from the NOVANA stations Figure 6 together with monthly data on water depth, temperature and salinity at each sampling station. Temperature and salinity data were obtained from the Global Ocean Physics Reanalysis with (GLORYSs2v4) and downloaded from the Copernicus Marine Environment Monitoring Service (http://marine.copernicus.eu/ ).

In order to account for differences in the number of available samples between stations (Figure 6) which may bias the detection and probability of presence of NIS at any given site, a formal data standardization was performed. This was achieved by randomly bootstrapping (i.e., resampling without replacement) the same number of unique samples from each site based on 100 random iterations. To ensure that a sufficient number of samples were present at each site, while maintaining the temporal structure of the data, the mean abundance (as well as presence and absence) for all species and sites were extracted for each decade and season (i.e., aggregating 3 months into quarters). Furthermore, the analysis was restricted to the years from the year 2000 onwards to ensure a sufficient coverage of sampling across stations.

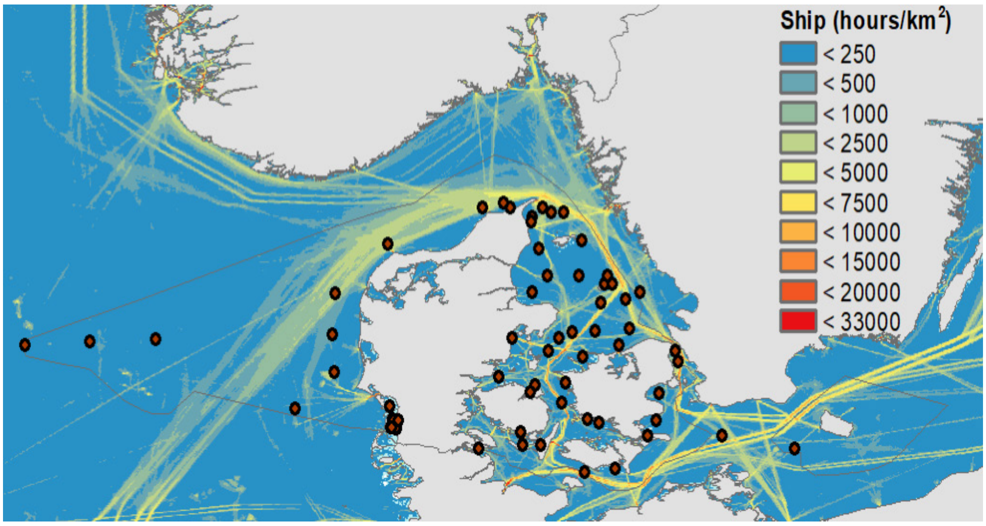


**Figure 7. Shipping activities along the major shipping routes with the NOVANA stations as overlay. Ship density of all type of vessels in 2019; ship density data extracted from EMODnet.**

The standardized dataset thus ended up consisting of 15 species and 68 sites recorded at each combination of site, decade and season with the following variables. Results are described in the chapter:

- **Presence/absence:** This reflects whether a determined species has been identified in a particular site in the ODA database, between January 1st, 1900 to October 15, 2020.

- **Species**: 15 non-indigenous species found in National Monitoring and Assessment Programme (NOVANA: *Amphibalanus improvisus, Alitta succinea, Crepidula fornicata, Ensis directus, Hemigrapsus takanoi, Marenzelleria neglecta, Marenzelleria viridis, Molgula Manhatensis, Mya arenaria, Polydora cornuta, Palaemon elegans, Petricola pholadiformis, Streblospio benedicti, Styela clava, Rhithropanopeus harrissii*.

- **Track density:** This is the sum of the *dispersal density* of each species in a determined site over 10 years (2001-2010).

- **Settling density**: This is sum of the number of agents of each species able to settle in a specific site over the 10 years (2001-2010).

- **Standardized and normalized dispersal density and settling density**: These are the *dispersal density* and *settling density* normalized or standardized as

  - $X_{norm} = \frac{X-X_{min}}{X_{\max}-X_{\_min}}$    $X_{stand} = \frac{X-X_{mean}}{Standard\ deviation}$

- **Ship activity level**: Covers 2019 vessel density of all types of ships from EMODnet extracted for the 68 sites. Ship activity is expressed as hours per square kilometer. Some of the NOVANA sites are located in these major shipping routes, but most are outside the major shipping routes.

- **Mean depth**: The mean sampling depth recorded at each sampling station.

- **Mean surface and bottom temperature**: The mean monthly temperature at the surface and bottom at each sampling station and year.

- **Mean salinity**: The mean monthly surface salinity (no bottom salinities available) at each sampling station and year.

**Models used**

Once the data collection and drift-model simulations were completed, two statistical models were applied to the established datasets in order to study the underlying mechanisms explaining the success of marine NIS taking into account the environmental conditions of the habitats, settling densities and shipping routes.

The models used were a Generalized Additive Model (GAM) and a Random Forest (RF) applied to the established datasets. These methods represent two complementary tools commonly used in species distribution modelling (Peterson 2003; Elith et al. 2011; Goldsmit et al. 2017; Lindegren et al., 2019). The same set of explanatory variables, reflecting factors related to the introduction, establishment and spread of NIS were used as co-variates.

The resulting statistical relationships and derived response curves between the set of co-variates and the presence/absence of NIS were compared across the two methods to assess the sensitivity and robustness of the results in relation to the choice of method.

In the GAM, we used the following binomial model formulation:

PAs, l, y,q= a+s(spn,s,l)+s(ship,l)+s(tempS,l,y,q)+s(tempB,l,y,q)+s(sal,l,y,q)+s(depth,l)+5y+S+/-ε

where **PA** is the presence/absence of each species **s,** at site **l**, in year **y** and season **q** as a function (using a logit link)[2] of the simulated settling density (**spn**), ship activity level (**ship**), surface and bottom temperature (**temS, temB**), salinity (**sal**) and depth at each site.
Since some NIS may have been introduced earlier and therefore had time to spread more widely throughout the area we included species identity (**S**) and years classified into 5-year time periods (**5y)** as fixed effect factors. The fitted coefficients therefore account for mean differences in the probability of presence between species over time that are not explained by the other set of covariates. The constant **a** is the overall intercept, **s** the thin plate smoothing function for each smooth term and **ε** the error term.

The number of regression splines is optimized (and penalized) by the generalized cross validation criterion (**GCV**). Furthermore, 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, RF is a machine learning tool combining 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 2002). In addition, RF has fewer constraints and is able to capture interactions between variables not easily achieved with GAMs.

We used the same model setup as in the GAM with regard to response and explanatory variables. In order to account for differences in the number of observations reporting absence/presence of NIS, we accounted for such unbalanced classes by assigning different weights to each class. 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 selected set of GAMs and RFs to predict the overall probability of occurrence of NIS across the entire study area, beyond the NOVANA monitoring stations used for model fitting and training. 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).

---

[2] The logit function is the natural log of the odds that a NIS settles in a specific location – used to ensure normally distributed residuals on the estimates.

# 6 Results and discussion

The potential habitat of the nineteen NIS in the Danish waters were modelled based on salinity tolerance, water depth and substrate requirements. Based on the definition of potential habitats, Lagrangian simulations were conducted and the modelled particle tracks and settlement densities shown in the section 1.2. We present results of two species, *Alitta succinea* and *Petricolaria pholadiformis,* to better explain the results. The maps for all the species are included in the appendix. Statistical analysis were conducted, based on field observations, tracking densities and settlement densities, and are presented in section 6.3.

## 6.1 Modelled habitats

We estimate individual habitat maps representing each of the 19 species of NIS based on substrate preference, expected depth distribution range, and expected salinity tolerance range for adult life stages. For improved insight into species habitats, we defined habitats as either optimal or suboptimal. Optimal habitats are areas where salinity conditions are within the tolerance range of the species all of the time, while suboptimal habitats are areas, where the salinity conditions may exceed the tolerance thresholds of the species in shorter or longer periods. In the habitat maps, we represent the optimal in green and suboptimal in yellow. These maps for all 19 species are shown in Appendix 3

For a more detailed presentation of habitats– and below for tracking and settling densities (in 6.2) - we will here exemplify findings for individual species by the conditions for two selected species *Alitta succinea* and *Petricolaria pholadiformis* (Figure 7). We found that both species are able to settle in all types of substrates and have the same depth range of 30 meter, but they differ with regard to tolerance of salinity range. *Alitta succinea* can tolerate a very wide range of salinities (0.14 to 80), while *P. Pholadiformis* only tolerates salinities from 10 to 30. Therefore, *A. succinea* is manly restricted by depth, but not by substrate requirements nor salinity. A comparison between the potential habitats of *A. succinea,* which is restricted by the depth 30 m, to *P. Pholadiformis* potential habitat, demonstrate the effect of salinity tolerance. The model simulations demonstrate suboptimal conditions in the areas of lower salinity south of Zeeland and further east of Bornholm. These two areas are thus not seen as potential habitats.

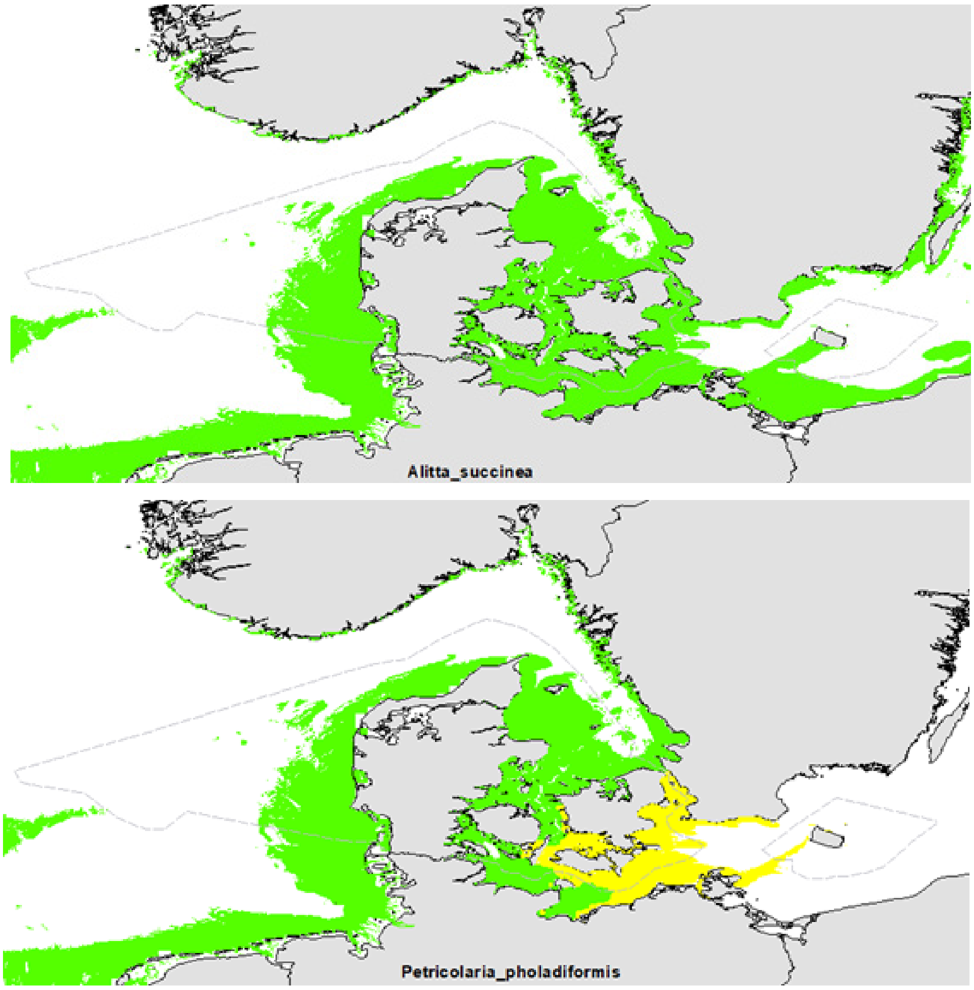


**Figure 8. Top: Potential habitat of *Alitta succinea*. Bottom: potential habitat for *Petricolaria pholadiformis*. Green colour represents habitats where salinity conditions are within the tolerance range of the species all of the time. Yellow colour represents habitats where the salinity conditions may exceed the tolerance thresholds of the species in shorter or longer periods.**

Figure 8 illustrates the modelled abundances of the 19 NIS in the Danish waters.  Top map shows the total accumulated habitats, whereas bottom map only shows habitats where salinity conditions for adult life stages are assumed optimal. In the North Sea and other areas with large water depths, abundances are low and with relatively little variability. The largest variability is seen in the Danish straits and inner areas in the vicinity of the Baltic Sea, generally in shallow coastal areas where most of the 19 species are observed. This indicates that under given conditions the number of species will increase with declining water depth.

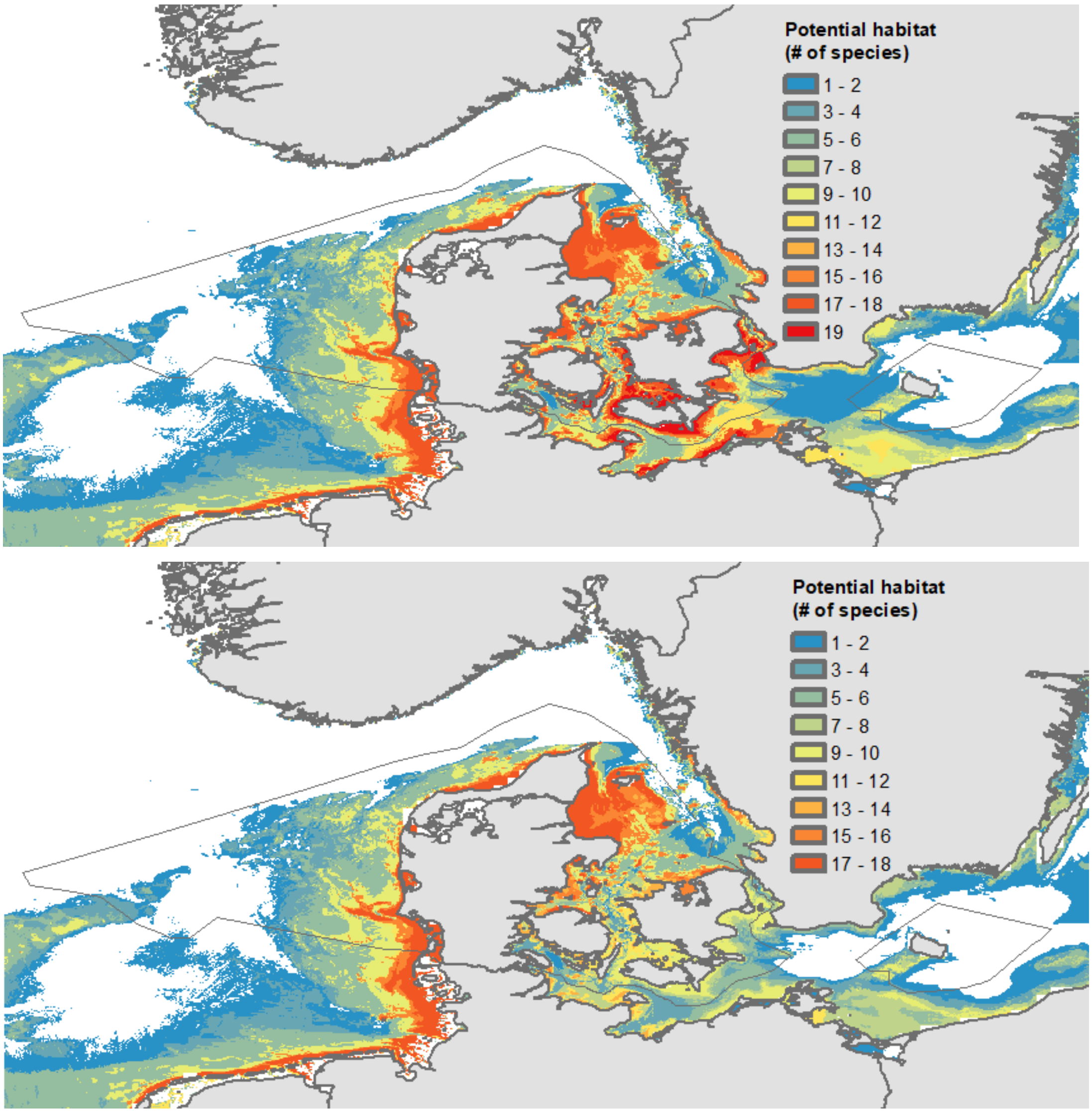


**Figure 9. Potential habitats of the 19 NIS included in the study. Top: Habitats with sub optimal and optimal salinity conditions of adult life stages are included. Bottom: Habitats where sub optimal conditions are ignored and only habitats with optimal salinity conditions of adult life stages are included.**

Based on the collected information on the 19 species regarding substrate preferences and depth range, the cumulative habitat maps show a larger concentration of species in shallower areas. Considering only optimal salinity conditions, Figure 9, bottom), in the regions south of the Belts, the number of species that are able to survive decrease. Hence, salinity might limit invasion of some of NIS in the region south of the Belts.

## 6.2 NIS particle trajectories

**Tracking densities**

In order to describe the pathways of the simulated particles, tracking densities were modelled for the individual species. The tracking density is defined by the number of times one particle has been inside a grid during the whole simulation divided by its area. The range is normalised to values between 0 to 1 as follows:

$$Tracking\ densities_{norm} = \frac{Tracking\ density - Tracking\ density_{min}}{Tracking\ density_{max} - Tracking\ density_{min}}$$

As we have a division with terms with the same unit, then tracking densities normalized is unitless. A tracking density 1 represents the maximum tracking density found, while 0.5 represents half of the maximum density when the minimum value for the tracking density is zero.

We used an average of the 10 years' simulations, producing one map per species representing 10 years' simulation. The results of all the simulations are given in Appendix 4 and in Figure 11. Examples of tracking densities are given for the same species illustrated in the previous section (Figure 10).

The model simulations show that for example *Alitta succinea* particles drift more in the Kattegat and west coast of Jutland than the Baltic Sea. This is probably because *A. succinea* larvae have a minimum salinity tolerance of 14. In contrast to this, larvae from *P. pholadiformis* are able to disperse closer to Bornholm and could not disperse through the North Sea. This is explained by the salinity restrictions for the larvae, since *P. pholadiformis* salinity limits are 10 to 30 while *A. succinea* limits are 14 to 45.

The modelled tracking densities are visually compared to data from the NOVANA program. At a NOVANA site near Bornholm, samples of *A. succinea* were observed (Figure 10) whereas this is not reflected in the modelling results. This is probably due to the salinity tolerance of the larvae given in the model. Although the adults have a wide tolerance (0 to 80), larvae have tolerance values between 14 to 45, so the modelling of the larvae released in the area near Bornholm will not "survive"; and larvae released from other areas would not reach this location. Thus, it seems likely that in the Baltic Sea the larvae of this species can tolerate a wider range of salinity than documented in the literature (and hence used in our model simulations). In general actual tolerance ranges have to be investigated further in order to improve future assessments.

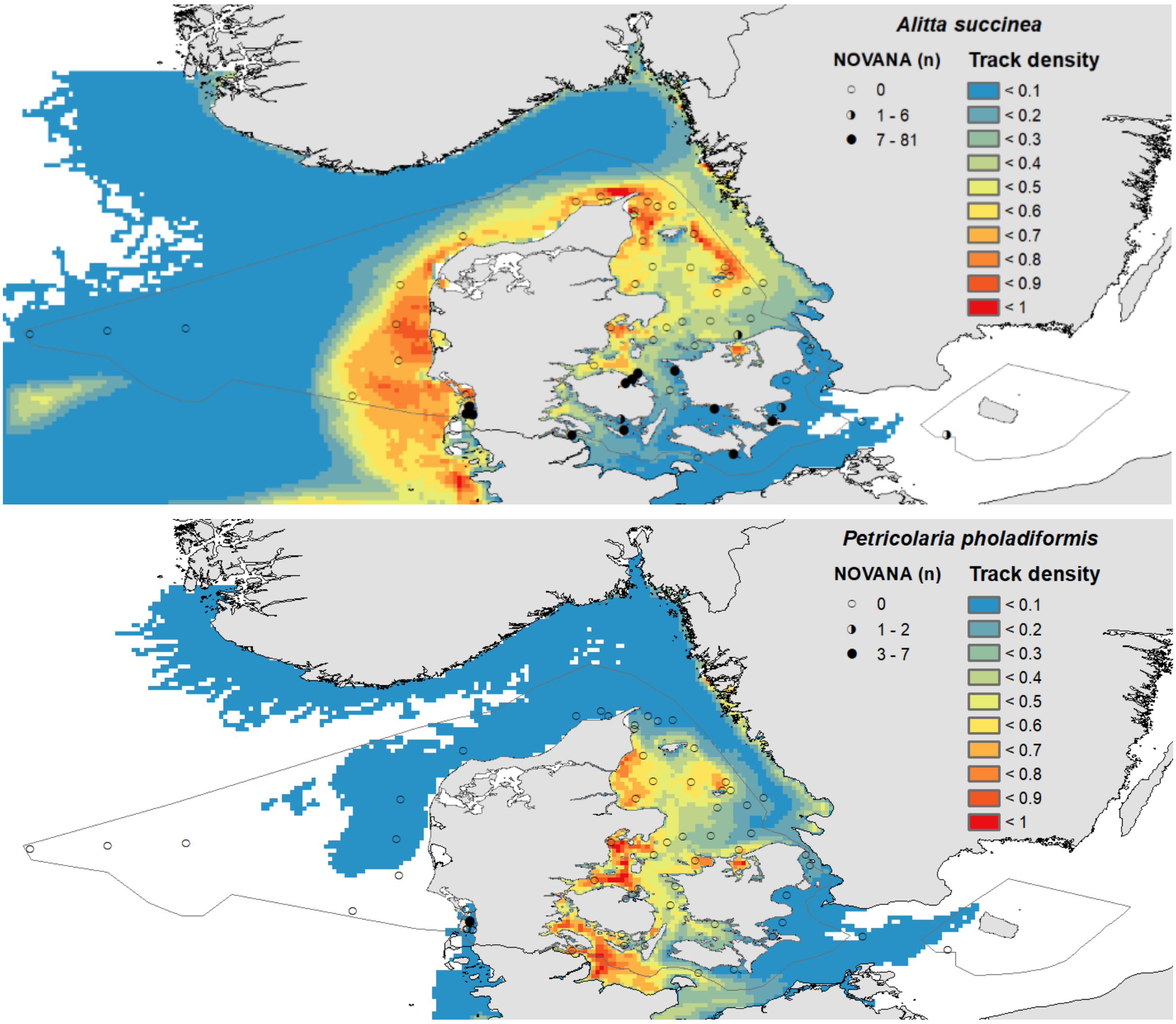


**Figure 10. Normalized track Density maps: color legends show normalized track densities; dot legend illustrates absence (white) and presences (black) of NIS at NOVANA stations. Top: Track density map of Alitta succinea. Bottom: Track density map of Petricolaria pholadiformis.**

In Figure 11, the average tracking densities for all 19 species are shown illustrating that the tracking densities generally are higher in Kattegat and Wadden Sea and lower in deeper areas in the North Sea and Skagerrak.

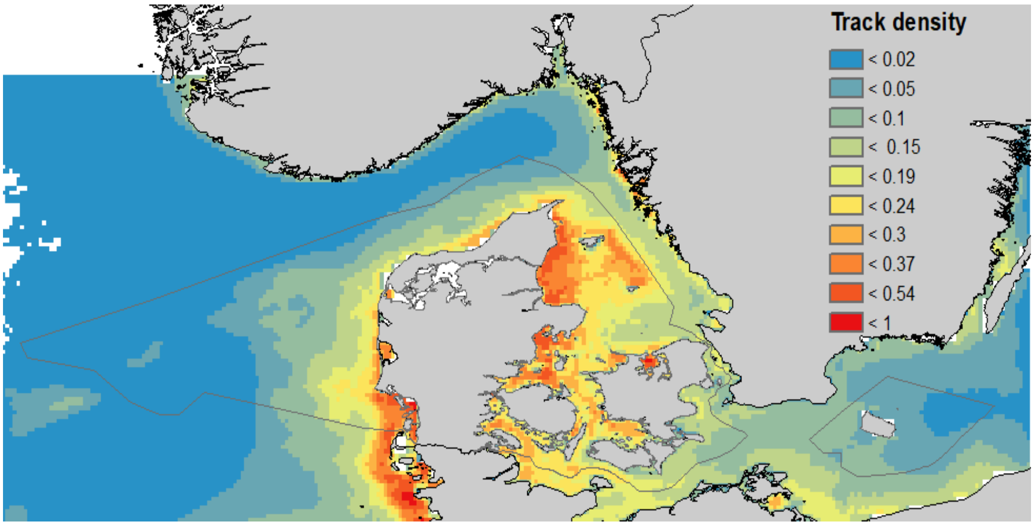


**Figure 11. Average of normalised tracking densities of the 19 species simulated.**

**Settling densities**

The settling density is defined as the number of particles that settles inside a grid cell divided by its area. The data is normalised within the range 0 to 1, and as the normalised tracking densities has no unit. The settling density maps for each of the 19 species of NIS are shown in Appendix 5, each map represents the average of 10 years for each species. Tracking density is used to understand particles trajectories (i.e. where the particles drift), and settling density is used to show the areas where particles can settle. Thus, the first describe the journey while the latter describe the destination.

In Figure 12, we illustrate the pattern by examples of settling densities, again with *Alitta succinea* and *Petricolaria pholadiformis*. Here we see that *Alitta succinea* is be able to settle in the Wadden Sea and on the west coast of Jutland, whereas *Petricolaria pholadiformis*, does not settle on the West and North coast of Jutland, probably because it is restricted by low salinity. In the Wadden Sea, the simulations do not indicate that *P. pholadiformis* is able to settle here even though specimens of this species have been observed in the NOVANA program,

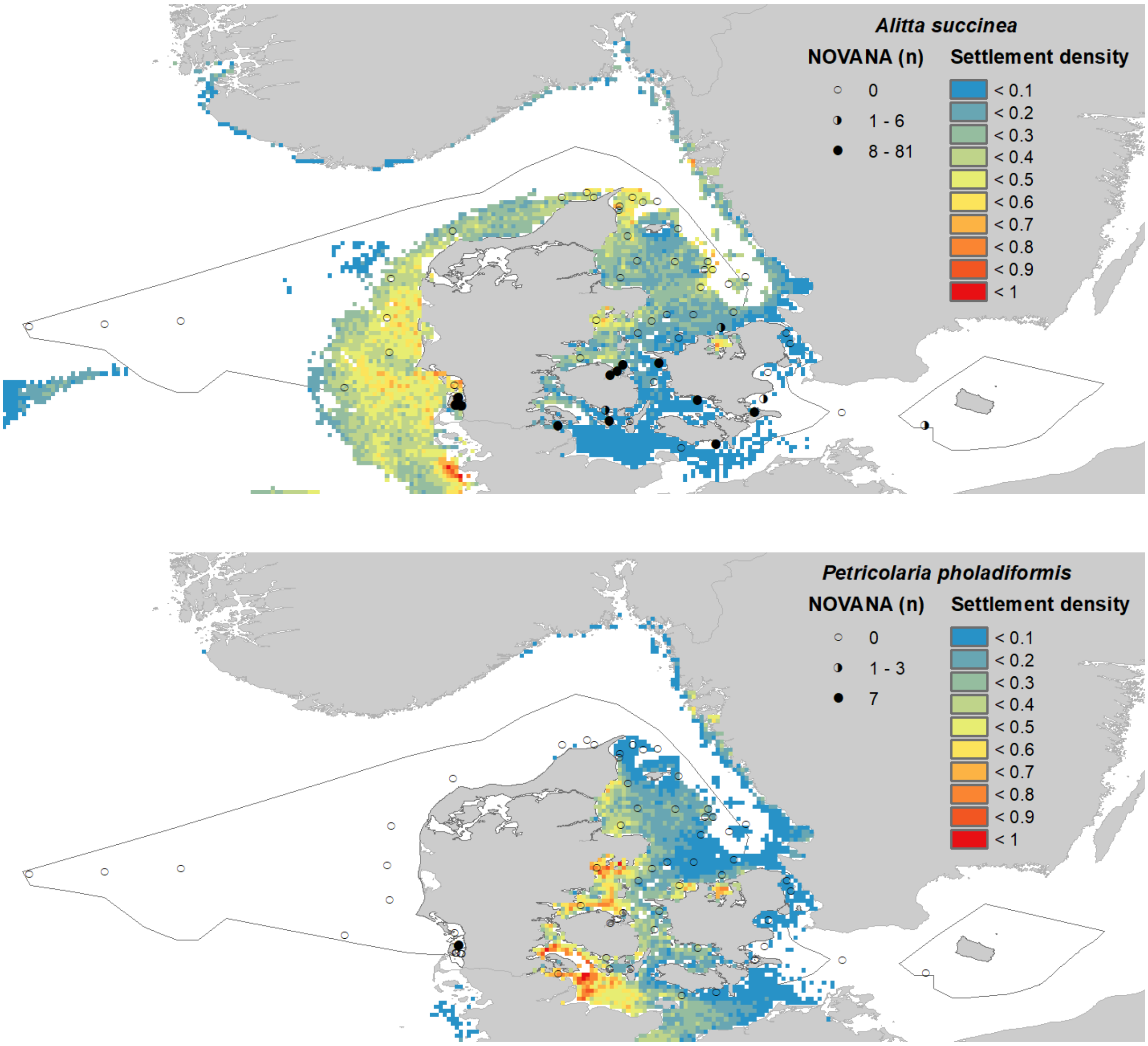


**Figure 12. Settling Density of Alitta succinea (top) and Petricolaria pholadiformis (bottom): color legends show normalized settling densities; dot legend illustrates absence (white) and presences (black) of NIS at NOVANA stations.**

In Figure 13, the average settling densities of all 19 species are shown. Settling densities were higher in the areas with higher optimal salinity habitats in Kattegat and the Wadden Sea .

Similar to the average tracking densities of 19 species (Figure 11), high settling densities would be identified in the Wadden Sea, and the coastal areas of the north-western Kattegat.

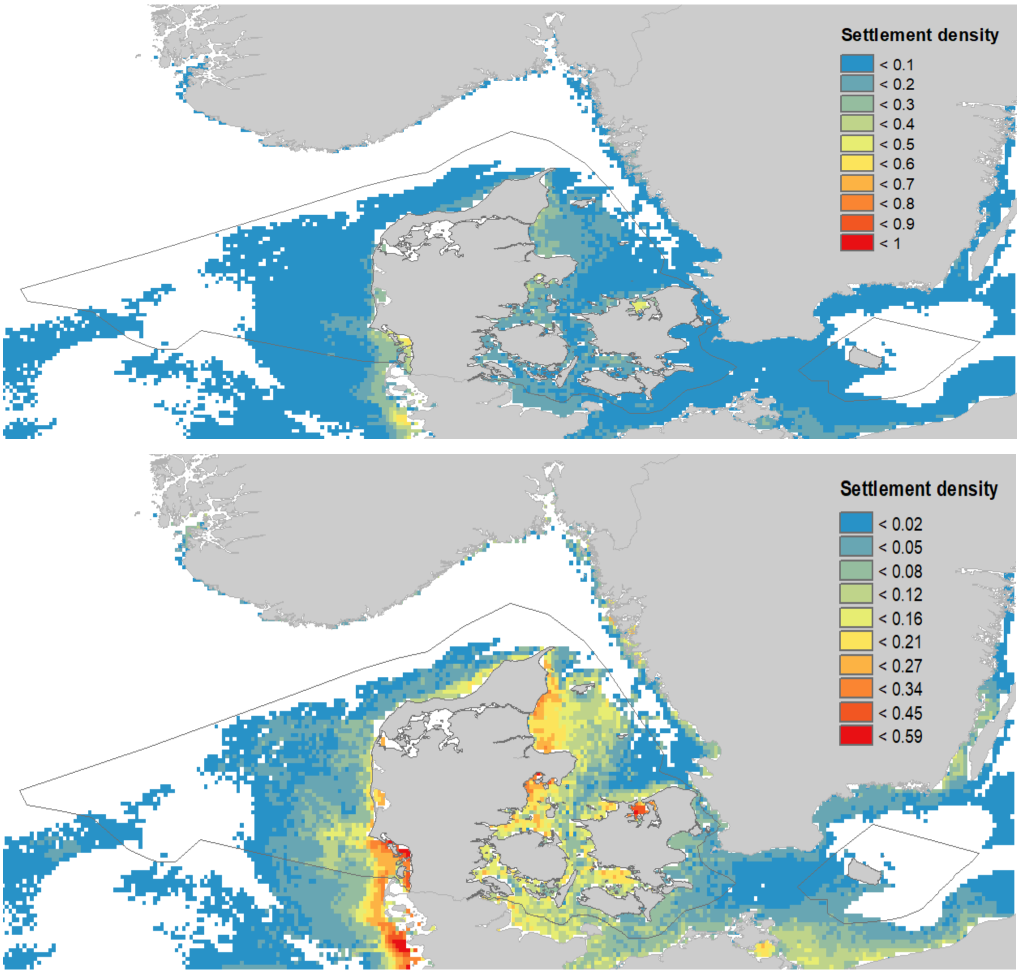


**Figure 13. Average settling densities of all the 19 species. Top: same scale as Figures 10-12. Bottom: new scale to a better visualization.**

## 6.3 Statistical analysis

A statistical analyses has been performed relating the in situ data from the NOVANA monitoring program with a set of predictors, including environmental conditions, shipping information, and the results of settling density simulations. This was done in order to identify the key drivers explaining the observed presence and absence of NIS throughout the study area, as well as characterize the statistical relationships (i.e., response curves) between NIS and the set of predictors.

The models fitted on the selected variables demonstrated statistically significant predictors overall. This means that no further removal of insignificant predictors (i.e., model reduction) was necessary. The GAM demonstrated a reasonable degree of explained variation in observed presence and absence across species, sites and decades (51.9%; Table 4). However, the accuracy of explaining the observed *absences* was considerably greater than *presence,* the latter showing an overall error rate of ~50%. The higher accuracy of explaining absences was due to the considerably higher number of observations of absences vs presences of NIS across all species and sites.

The 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 4). The higher estimated coefficients for the recent decade (2010-2020) compared to the time period 2005-2010 indicate an overall increased probability of presence of NIS over time. Furthermore, the partial smooth plots showed a positive, exponentially increasing response to higher settling densities (Figure 14A). However, please note that the relationship towards low values of settling density is uncertain since it is driven by a pronounced scatter of observations at zero values of settling density.

**Table 4. Summary statistics of parametric coefficients (i.e., species and time period) and smooth terms for the GAM.**

```
Parametric coefficients
               Est     SE      z       P
(Intercept)   -7.80    0.48  -16.32  <0.001
Species: As    2.15    0.48    4.53  <0.00
Species: Cf   -0.90    0.67   -1.33   0.18
Species: Ed    1.56    0.49    3.18  <0.001
Species: Ma    4.61    0.45   10.15  <0.001
Species: Mm  -63.81    95.00   0.01   1.00
Species: Mn   -2.27    1.10   -2.06   0.04
Species: Mv    1.90    0.48    3.94  <0.01
Species: Pc    2.83    0.47    6.06  <0.01
Species: Pe    2.24    1.10   -2.04   0.04
Species: Rh  -63.51    95.00   0.01   1.00
Species: Sb    1.65    0.49    3.34  <0.001
5y[2005,2010]  0.21    0.18    1.15   0.25
5y[2010,2015]  0.75    0.19    3.91  <0.001
5y[2015,2020]  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
```

The corresponding response plots derived based on RF showed a similar positive relationship with settling density but with a more linear increase in the probability of presence across all values of settling densities (Fig. 15A). In terms of ship activity, both GAM and RF demonstrated a positive and saturating response at higher values of vessel density (Figure 14B, Figure 15B). However, please note that the slightly decreasing relationship based on GAMs towards the highest values is uncertain and driven by very few extreme observations. Hence, it should be considered with caution. The remaining relationships with temperature, salinity and depth indicated the responses of NIS to environmental conditions.

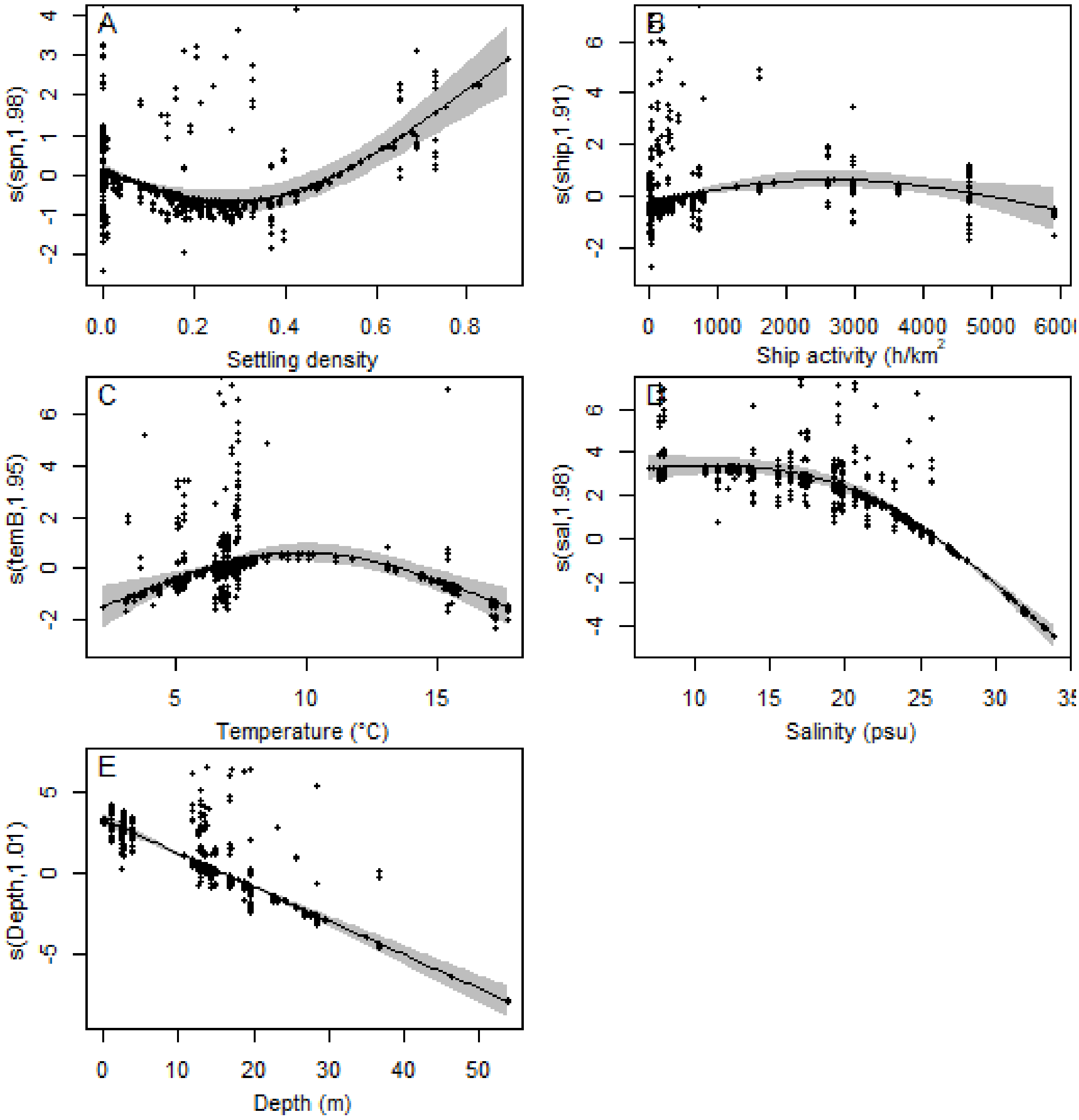


**Figure 14. GAM smooth plots showing the partial effects of settling density (A), vessel density (B), bottom temperature (C), salinity (D) and depth (E) on PA. Solid lines and shaded areas show the mean and 95% confidence intervals, respectively, and filled circles the partial residuals for each predictor.**

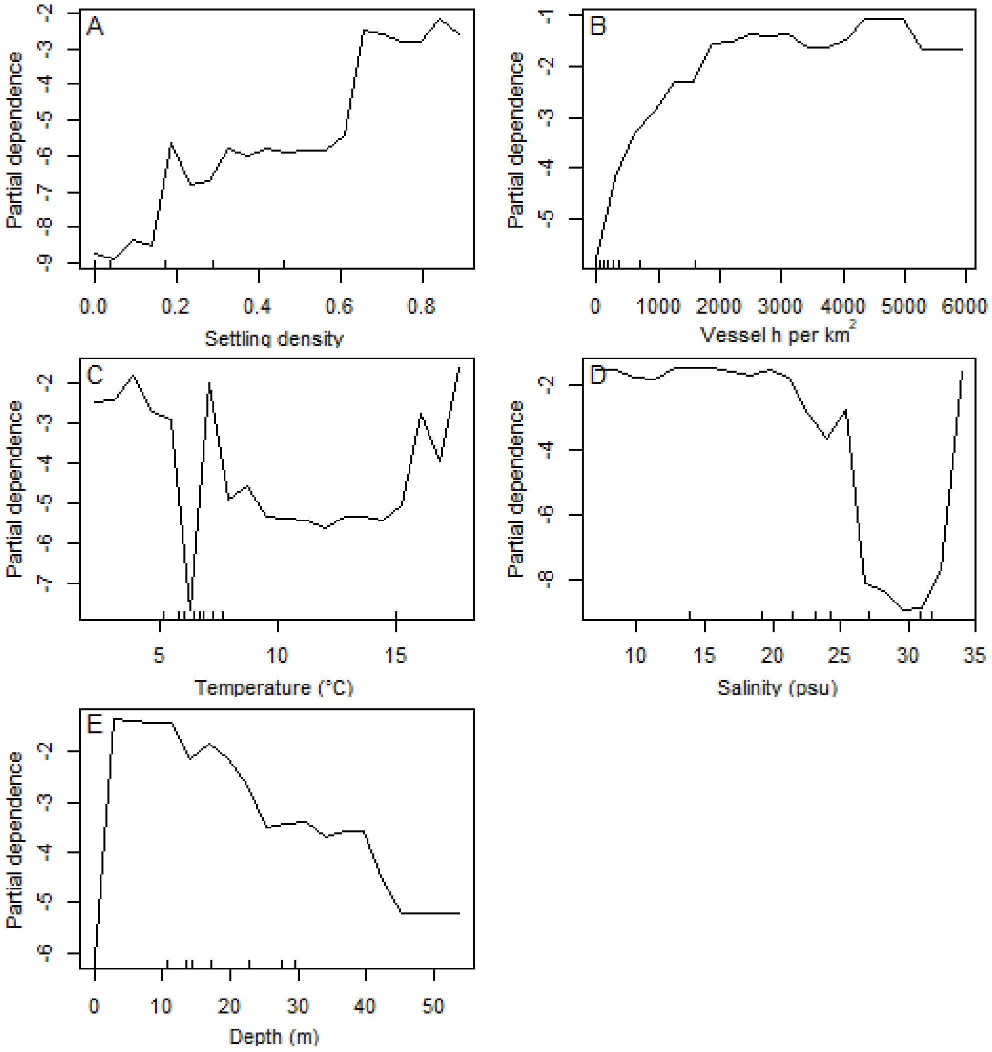


**Figure 15. Partial dependence plots based on RF showing the partial effects of settling density (A), vessel density (B), bottom temperature (C), salinity (D) and depth (E) on the presence.**

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° (Fig. 14C), while the response based on RF was more erratic and highly variable across the range of values (Fig. 15C).

With regard to salinity and depth, both methods demonstrated a negative, linear or non-linear declining relationship, respectively, with low probability of presence at higher salinities (Fig. 14D, Fig. 15D) and at greater depths (Fig. 14E, Fig. 15E). Among the set of predictors, the factor species had the highest relative variable importance in the RF, as illustrated by the largest decrease in classification error, here represented by the mean Gini index (Figure 16). Subsequently, the set of environmental predictors (salinity>depth>>temperature) were deemed important, closely followed by the effect of shipping activity, time-period and settling density.

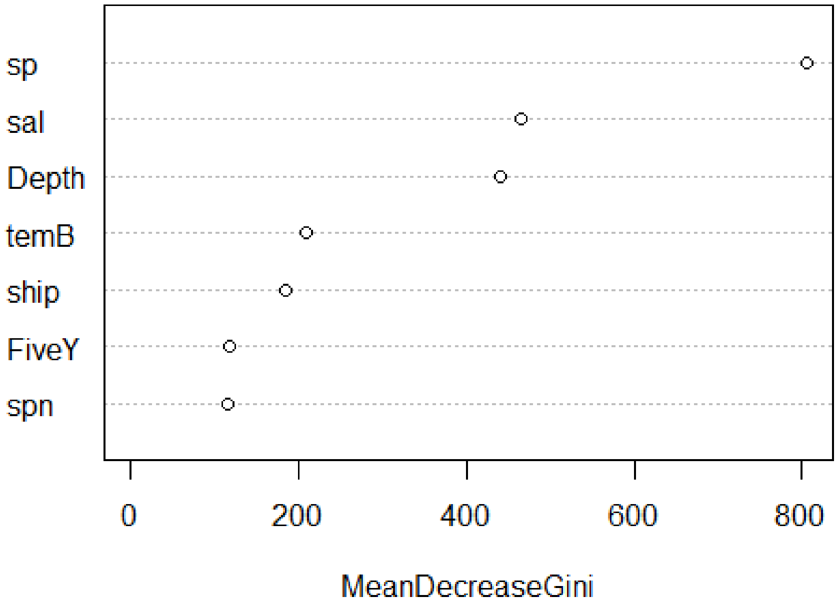


**Figure 16. Variable importance of each predictor (see abbreviations and explanations below) based on RF estimated as the mean decrease in the Gini index. (Sp=factor of species identities, FiveY=factor of 5-year time periods, sal=bottom salinity, temB=bottom temperature, spn=settling densities, ship=vessel densities).**

# 7 Conclusion

The present study has led to improved understanding of the key mechanisms and assembly rules regulating the marine NIS, which is an insight of importance for the conservation and management.

Firstly, the study demonstrated a significant effect of environmental conditions affecting the probability of occurrence of NIS throughout the study area. More specifically, our statistical analysis shows that the type of environments and habitats 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 considered in this study, depth and salinity are likely the key limiting factors determining the occurrence and establishment of NIS within our study area.

Secondly, the significant effect of settling densities emphasize also the role of dispersal on the presence and spread of NIS. 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, the benthic invertebrates considered within this study are mainly restricted to dispersal during their planktonic egg or larvae stages. Hence, our results demonstrate that larvae dispersal may allow for an effective secondary spread of marine NIS in Danish coastal waters. Our study also emphasize the value of advanced drift models taking into account not only physical transport but the particular traits of NIS during model simulations (e.g., timing of spawning and duration of larvae phase).

Thirdly, our findings demonstrate a statistical relationship between vessel activity and the probability of occurrence of NIS. This means that all kinds of shipping activities may act as a primary factor for introducing marine NIS. This may include ballast water, hull fouling etc.

To conclude, the details behind these insights may provide key guidance for improving and prioritizing existing monitoring programs and sampling protocols to ensure a timely detection and robust status assessment of marine NIS. Furthermore, the identification of vulnerable habitats and potential pathways for introductions and spread may serve to aid in the planning and prioritization of limited conservation efforts aiming to avoid or combat marine NIS. This is important 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.

# 8 Recommendations

For future assessments and impact evaluations of NIS in Danish waters, we recommend that further work is carried out on combining model output and different environmental variables for their effect on observed NIS occurrences.

We further recommend continued work on possibilities for assessing environmental linkages and for identification of sites of importance for NIS monitoring. In this context, we recommend that an extensive monitoring campaign targeting NIS in Danish marine waters is carried out. This should create the basis for improved analyses of model simulations and the relationships between environmental variables and NIS occurrences.

We suggest the following priorities in dedicated surveys with a broader focus designed to investigate the coastal nearshore areas identified in this study and close to identified shipping routes. The idea is to investigate in poorly studied areas e.g. along the Jutland coast to give snap shots across the coastal zone:

1. To use standardized sampling methods by e.g. Van Veen grabs or similar that better represent the larger benthic species and/or the use of e-DNA monitoring to register presence/absence

2. 
    a. coastal areas <10 m
    b. stations both close to and far from major ports
    c. stations both far from and close to high shipping activities.
    d. stations with variety of bottom substrates including vegetation and biogenic reefs.
    e. stations representing a variety of modelled track and settling densities , e.g. using the both the accumulated and individual density maps generated in the present study

3. To use the statistical model developed in this study to create predictive maps of the current (potential) distribution of each of the 19 NIS included in the study. It is recommended that these maps are considered in order to highlight the hotspots for NIS occurrences, and together with eventual additional criteria, used for prioritizing the future location of monitoring efforts on NIS in Danish marine waters ( e.g. in the NOVANA program).

4. To use the statistical model to understand how future changes in habitats and/or environmental conditions induced by climate change may affect the predicted distribution of NIS and the

location of hotspots in the coming years. Climate change is expected to affect both water currents, water temperature and salinities in Danish waters, in turn affecting the pelagic larvae duration, the drift patterns and mortality of larvae. For such future predictions, it is recommended to update larval dispersal models accordingly, prior to any statistical predictions.

5. To introduce joint species distribution models accounting for potential biotic interactions (competition/predation) between native and non-native species.

# 9 References

Alaska Center for Conservation Science, 2017. Bering Sea Marine Invasive Species Assessment. https://accs.uaa.alaska.edu/wp-content/uploads/diadumenelineata.pdf. Accessed on 5 June 2020.

AquaNIS. Editorial Board, 2015. Information system on Aquatic Non-Indigenous and Cryptogenic Species. World Wide Web electronic publication. www.corpi.ku.lt/databases/aquanis. Version 2.36+. Accessed 2020-06-10.

Berg P, Poulsen J W, 2012. Implementation Details for HBM. DMI Technical Report, No. 12-11, ISSN: 1399 – 1388, Copenhagen.

Breiman, L.. 2002.Manual on setting up, using, and understanding random forests v3.1, http://oz.berkeley.edu/users/breiman/ Using_random_forests_V3.1.pdf.

CABI, 2020. Invasive Species Compendium. Wallingford, UK: CAB International. www.cabi.org/isc.

Christensen A, 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.” P L O S One, edited by João Miguel Dias, vol. 13, no. 1, Public Library of Science.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

EMODnet-Geology. Seabed Substrate Dynamic. Available online: https://www.emodnet-geology.eu/dataproducts/seabed-substrates/ Accessed on 1 June 2020

EMODnet Human Activities: Vessel Density Map. Available online https://www.emodnet-human-activities.eu/search-results.php?dataname=Vessel+Density+ Accessed on 10 Aug 2020

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.

Fofonoff PW, Ruiz GM, Steves B, Simkanin C, & Carlton JT. 2020. National Exotic Marine and Estuarine Species Information System. http://invasions.si.edu/nemesis/.Access Date: 10-June -2020

Gaines SD, Lester SE, Grorud-Colvert K, Costello C, Pollnac R., 2010. Evolving science of marine reserves: new developments and emerging research frontiers. Proceedings of the National Academy of Sciences, 64:47-57.

GEBCO. 2020. Gridded bathymetry data. https://www.gebco.net/data_and_products/gridded_bathymetry_data/. Accessed on 1 June 2020.

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

Hansen, F. T., Gabellini, A. P., & Christensen, A. (2020). *Ranking of Danish ports according to shipping activities and to the potential of natural dispersal of nonindigenous species*. DTU Aqua. DTU Aqua-rapport, No. 369-2020.

Jones GP, Srinivasan M, Almany GR. 2007. Population connectivity and conservation of marine biodiversity. Oceanography, 20 (3), 100-111.

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.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 23: 18–22.

Lipcius RN, Eggleston DB, Schreiber SJ, Seitz RD, Shen J, Sisson M, Stockhausen ,2008. Importance of metapopulation connectivity to restocking and restoration of marine species. Reviews in Fisheries Science, 16:101-110.

NOBANIS. Available from http://www.NOBANIS.org. Access Date: 10-June -2020

Ocean Biographic Information System (OBIS). 2020. Retrieved from https://obis.org/.

ODA. 2020. Surface Water Monitoring, Aarhus University. Available online: https://odaforalle.au.dk/main.aspx Access Date 15 October 2020.

Toonen RJ, Andrews KR, Baums IB, Bird CE, Concepcion GT, Daly-Engel TS, Eble JA, Faucci A, Gaither MR, Iacchei M, Puritz JB, Schultz JK, Skillings DJ, Timmers MA, Bowen BW. 2011. Defining boundaries for ecosystem-based management: a multispecies case study of marine connectivity across the Hawaiian Archipelago. Journal of Marine Biology, 1-13.

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

# APPENDICES

## Appendix 1. Supplementary on GAM analyses

**Table 5. Variance inflation scores (VIF) showing the degree of multi-colinearity among the selected set of predictors where a score >5 indicates collinearity. The lower table show VIF scores after having excluded surface temperature.**

```
  Variables      VIF
1       tdn 2.102712
2       spn 1.950102
3      ship 1.199907
4     Depth 1.694267
5       tem 7.076659
6      temB 6.995684
7       sal 1.597322
```

```
Variables      VIF
1       tdn 2.078216
2       spn 1.910232
3      ship 1.229622
4     Depth 1.725610
5      temB 1.078997
6       sal 1.642807
```

**Table 6. Summary statistics of parametric coefficients (i.e., species and decade) and smooth terms for the full GAM. 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.**

```
Parametric coefficients:
                        Estimate Std. Error z value Pr(>|z|)    
(Intercept)           -8.497e+00  4.318e-01 -19.680  < 2e-16 ***
as.factor(sp)As        2.564e+00  4.253e-01   6.029 1.65e-09 ***
as.factor(sp)Cf       -5.932e-01  6.444e-01  -0.921 0.357304    
as.factor(sp)Ed        1.990e+00  4.359e-01   4.566 4.98e-06 ***
as.factor(sp)Ma        5.172e+00  4.126e-01  12.535  < 2e-16 ***
as.factor(sp)Mm       -5.081e+01  1.507e+06   0.000 0.999973    
as.factor(sp)Mn       -2.014e+00  1.080e+00  -1.865 0.062141 .  
as.factor(sp)Mv        2.331e+00  4.290e-01   5.433 5.53e-08 ***
as.factor(sp)Pc        3.251e+00  4.178e-01   7.779 7.28e-15 ***
as.factor(sp)Pe       -2.014e+00  1.080e+00  -1.865 0.062141 .  
```

```
as.factor(sp)Rh       -8.922e-01  7.062e-01  -1.263 0.206427
as.factor(sp)Sb        1.642e+00  4.453e-01   3.688 0.000226 ***
as.factor(Decade)2010  6.156e-01  1.279e-01   4.813 1.49e-06 ***
as.factor(Decade)2020  7.819e-01  1.538e-01   5.085 3.67e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
           edf Ref.df  Chi.sq  p-value
s(tdn)   1.847  1.975   8.993   0.0183 *
s(spn)   1.996  2.000  79.511  < 2e-16 ***
s(ship)  1.954  1.998  35.488 3.54e-08 ***
s(temB)  1.986  2.000  48.216 3.38e-11 ***
s(sal)   1.975  1.999 374.574  < 2e-16 ***
s(Depth) 1.881  1.985 254.418  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.425   Deviance explained = 51.9%
UBRE = -0.88866  Scale est. = 1         n = 23784
```

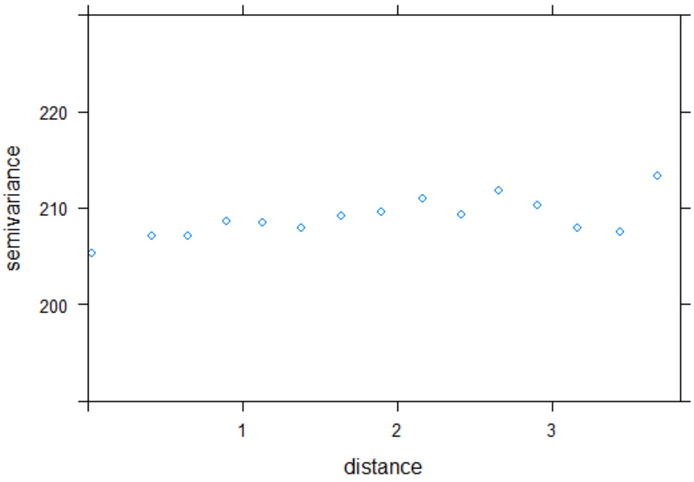


**Figure 17. Variogram of model residuals across spatial distances indicating no sign of spatial autocorrelation.**

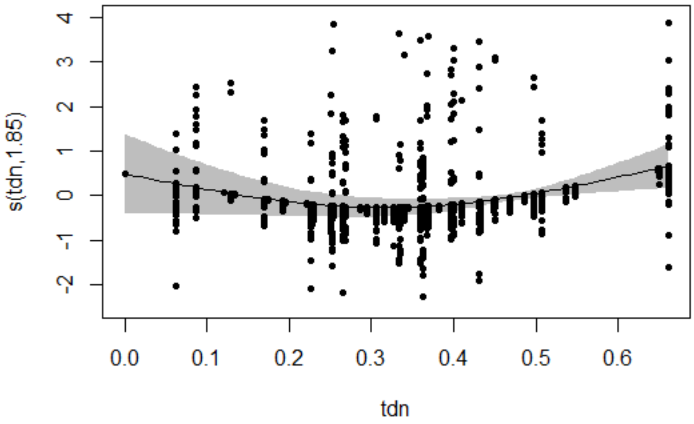


**Figure 18. GAM smooth plots showing the partial effects of dispersal density (tdn) showing an ecologically unrealistic u-shaped relationship. Solid lines and shaded areas show the mean and 95% confidence intervals, respectively, and filled circles the observed values.**

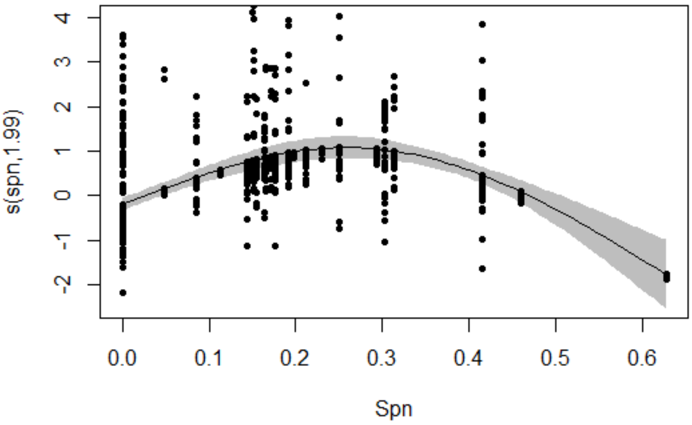


**Figure 19. GAM smooth plots showing the partial effects of settling density (spn) showing a highly uncertain and ecologically unrealistic decline towards the extreme values of spn. Solid lines and shaded areas show the mean and 95% confidence intervals, respectively, and filled circles the observed values. If omitting the relationship is robustly characterized**

# Appendix 2. List of possible constraints to the dispersal and settling density modelling

The estimation of *dispersal densites* and *settling densities* could be constrained by:

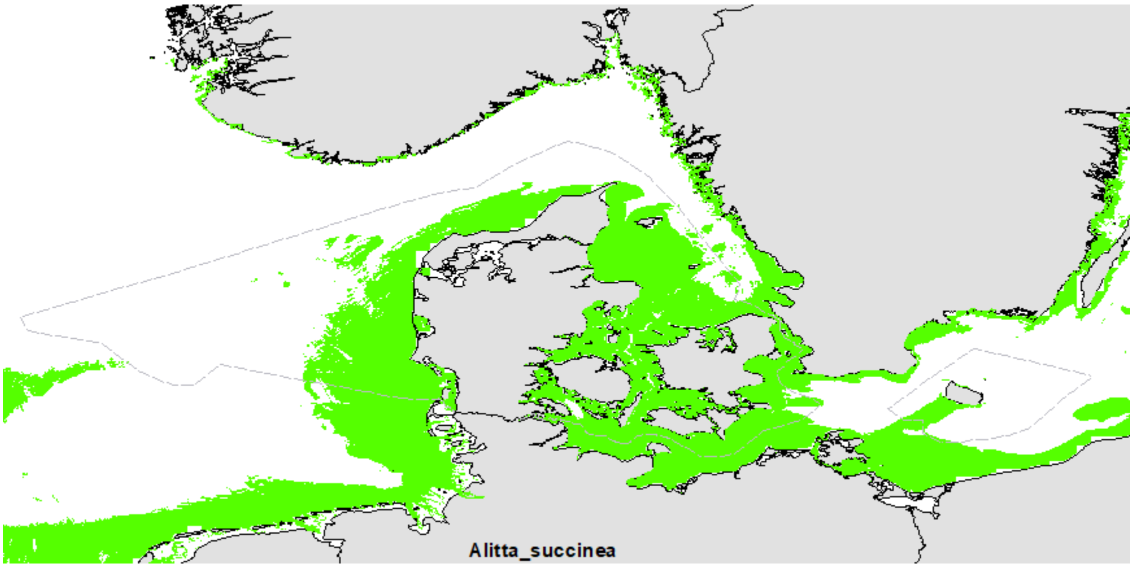
- Species not being ideally distributed in suitable habitats contrary to the assumption used for the ABM modelling
- Locations of species recordings may to large extent be determined by random introduction
- Species not being equally represented in samples due to sampling techniques using small sampling areas (i.e. Haps) underrepresenting larger species
- Different frequencies and in different periods/years in observations by stations
- The interpretation of absence of NIS at given station
- Limited data for a range of species in NOVANA
- The presence data being highly influenced by the introduction by chance, and the succession state of the NIS populations.

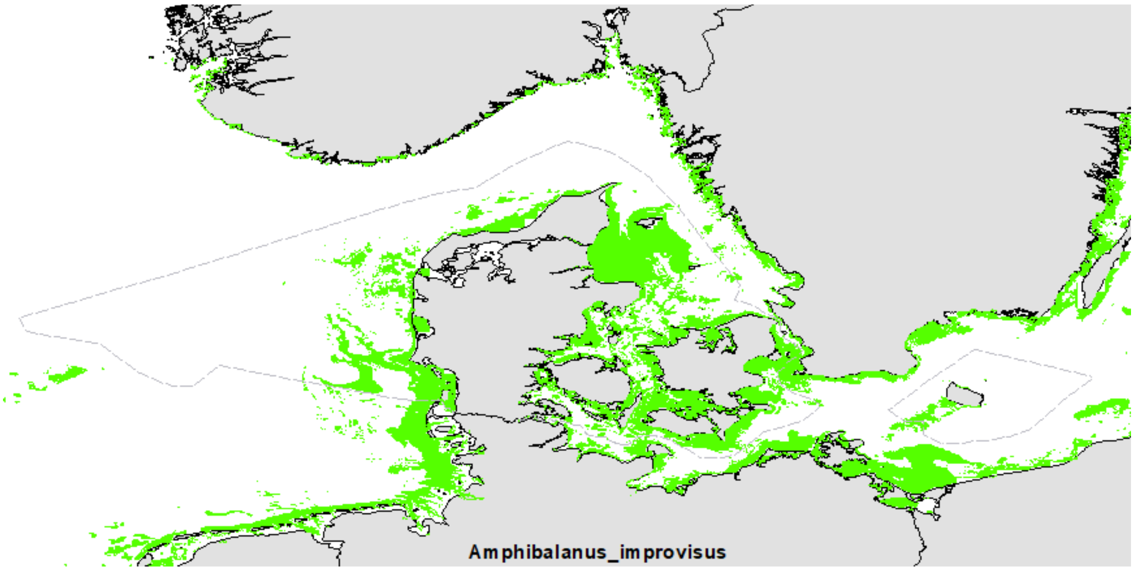
Further, the analyses of shipping data could be constrained by

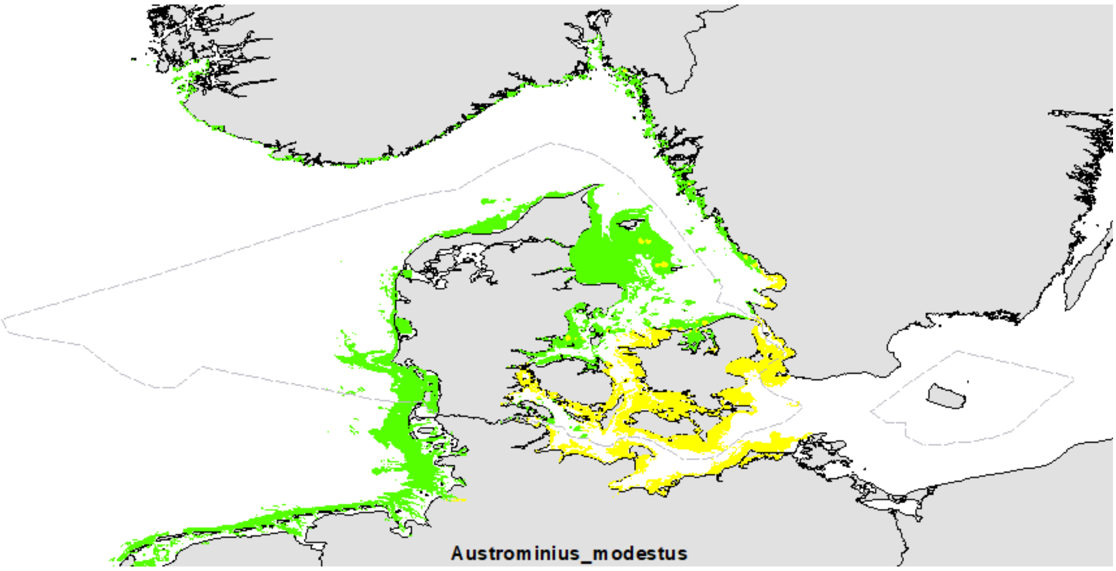
- Ballast water being primarily exchanged in port and harbour locations
- Possibility of considerable secondary spread via recreational vessels which is not reflected in the AIS data
- The occurrences of NIS would seldom represent the location of introduction (by ship) but rather areas defined by the dispersal and/or settlement of larvae.
- Potential influence from fouling agents is not considered

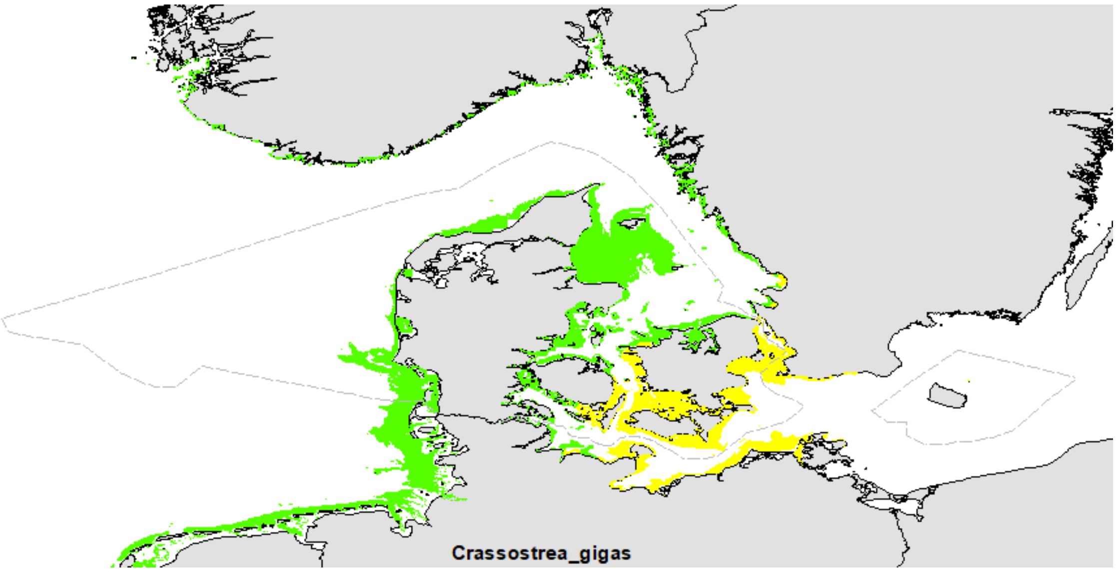
# Appendix 3. Habitat maps

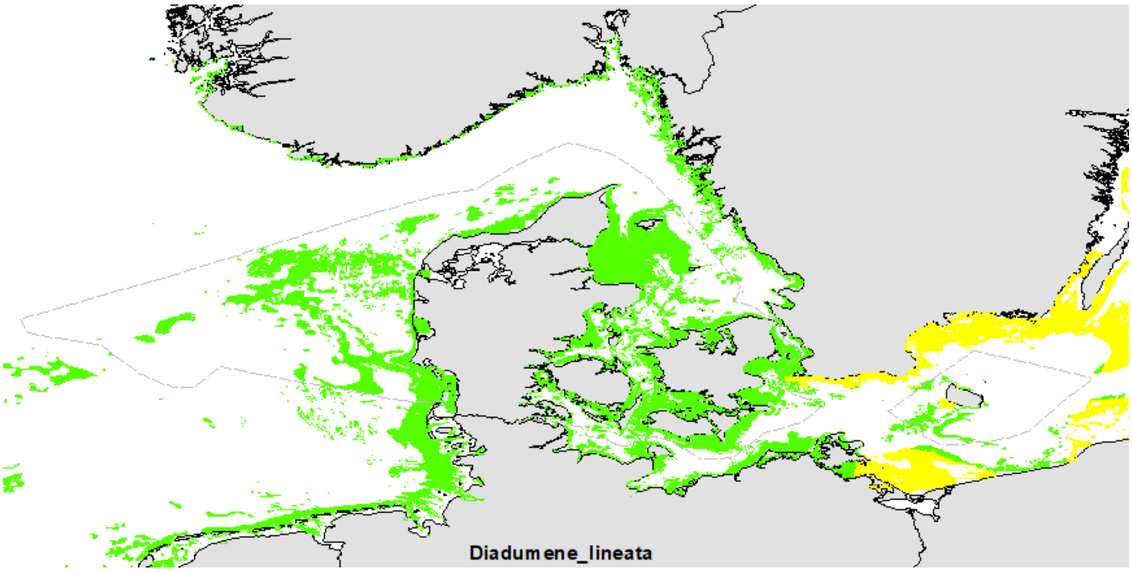
Habitat maps based on substrate preferences, depth preferences and adult salinity tolerances. Green represents habitats where salinity conditions are within the tolerance range of the species all of the time. Yellow represent habitats where the salinity conditions may exceed the tolerance thresholds of the species in shorter or longer periods.

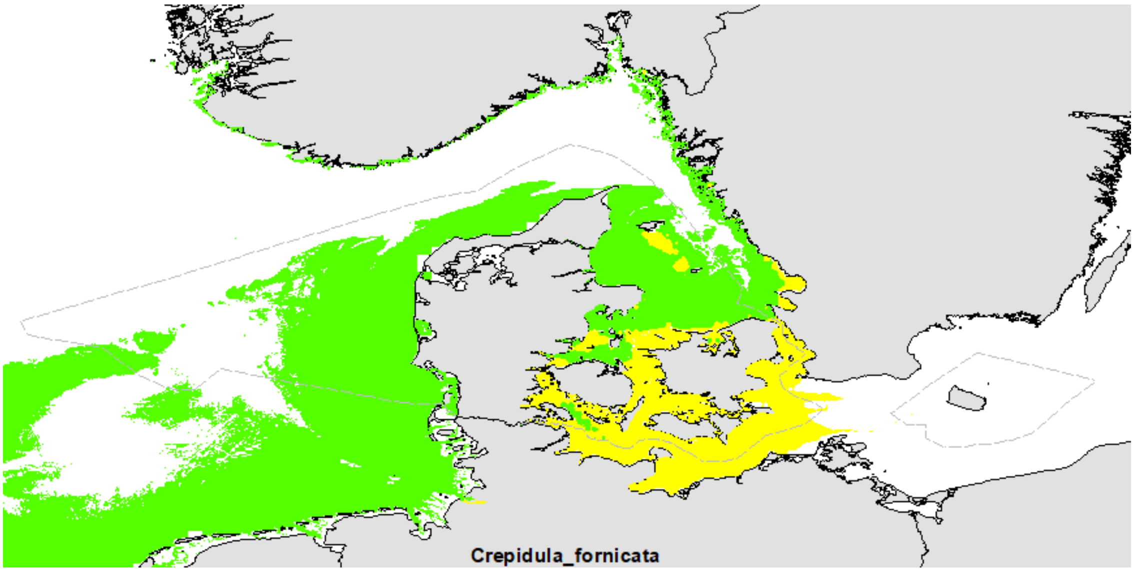


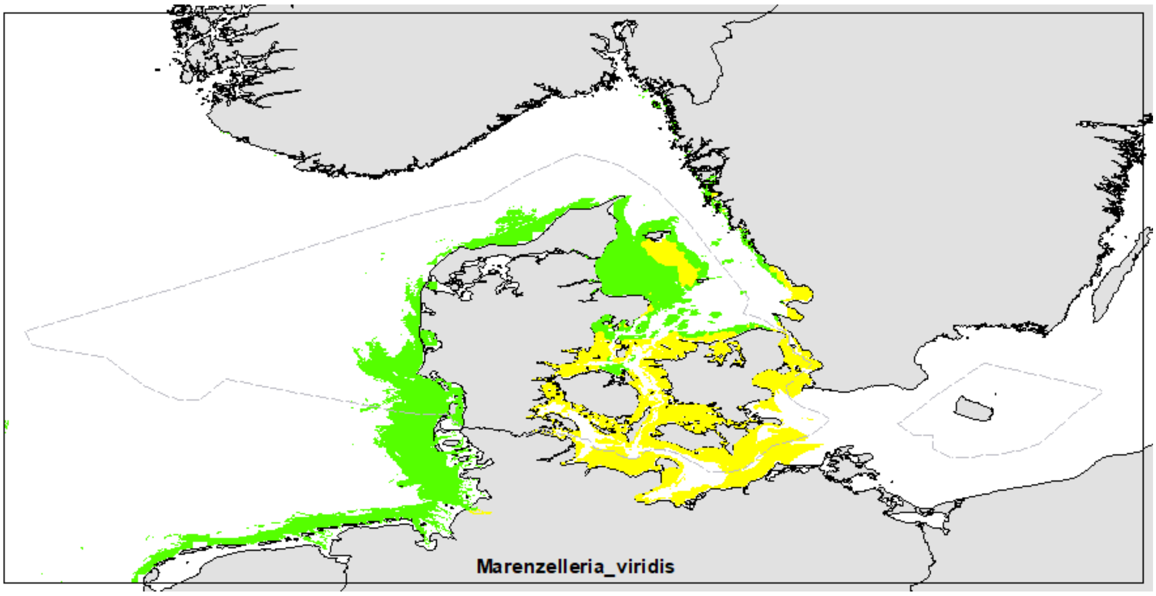


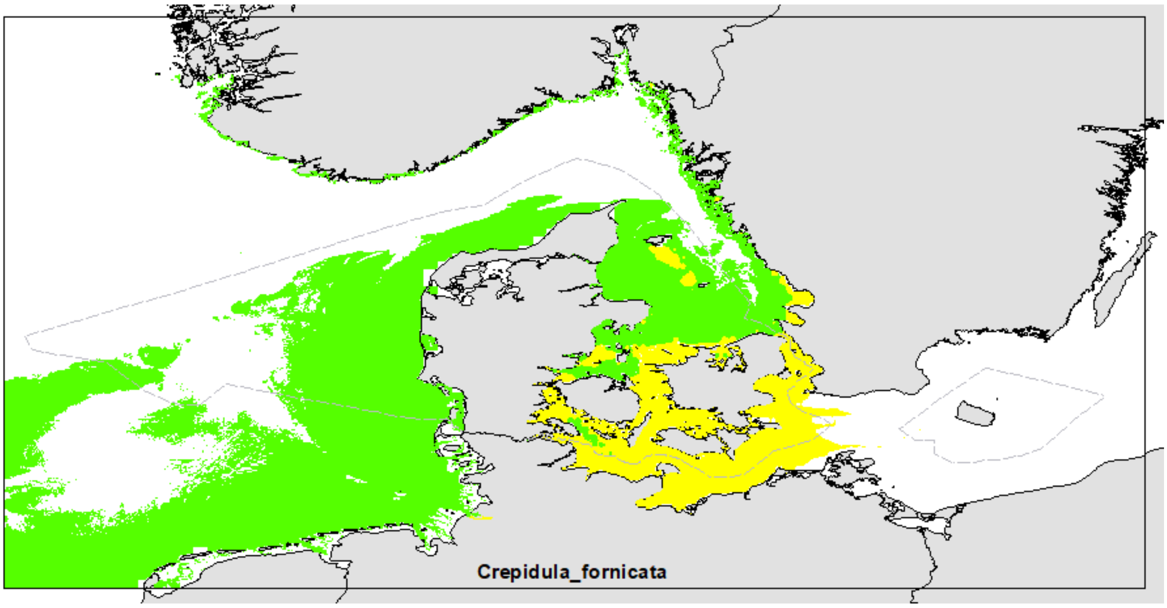

Austrominius_modestus

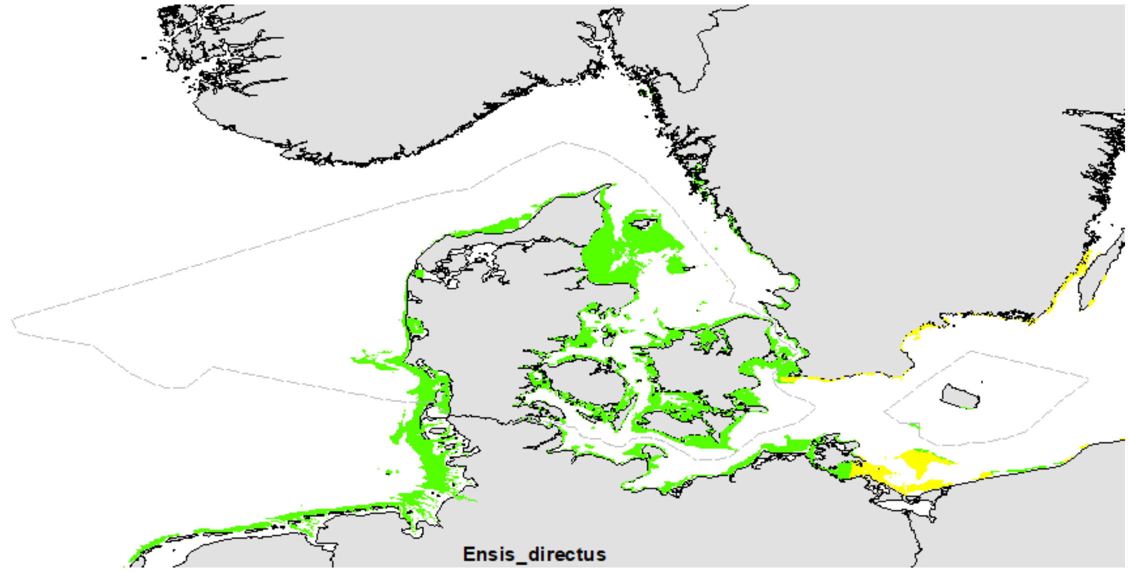

Crassostrea_gigas

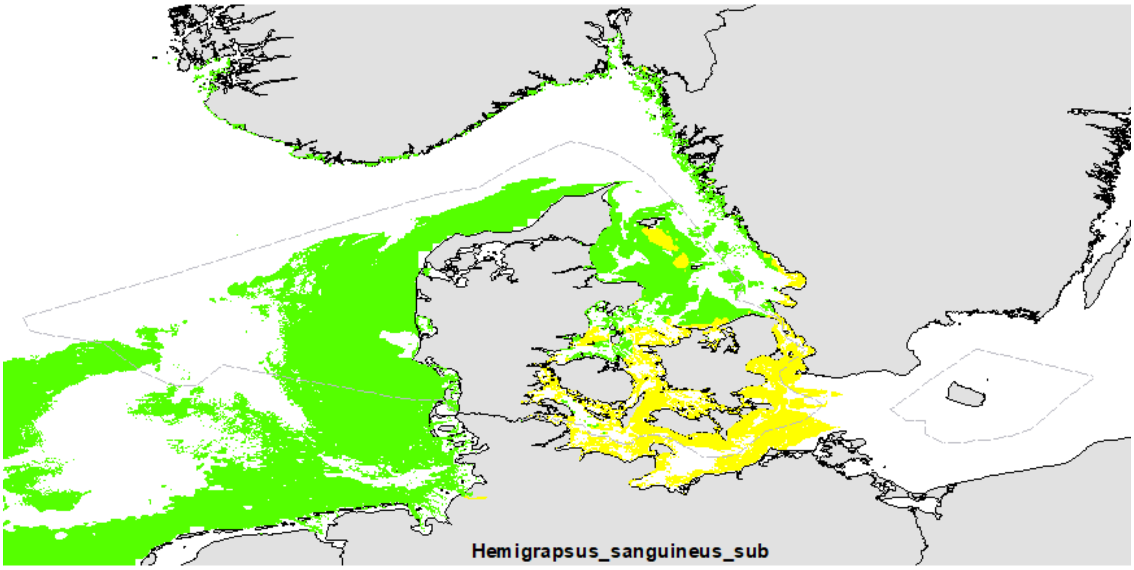

Diadumene_lineata

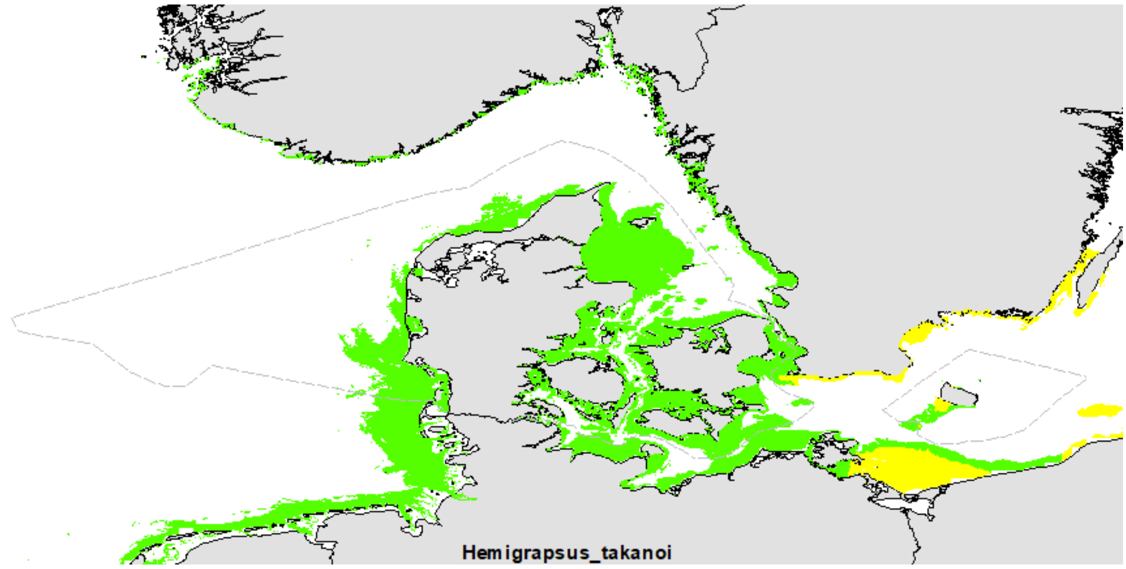

Crepidula_fornicata

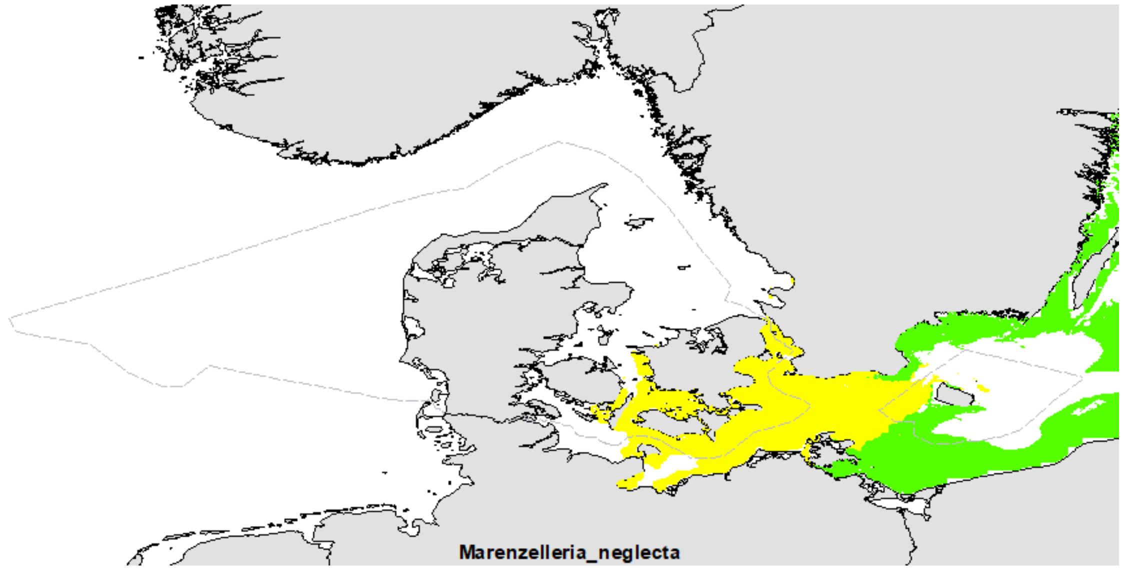

Marenzelleria_viridis

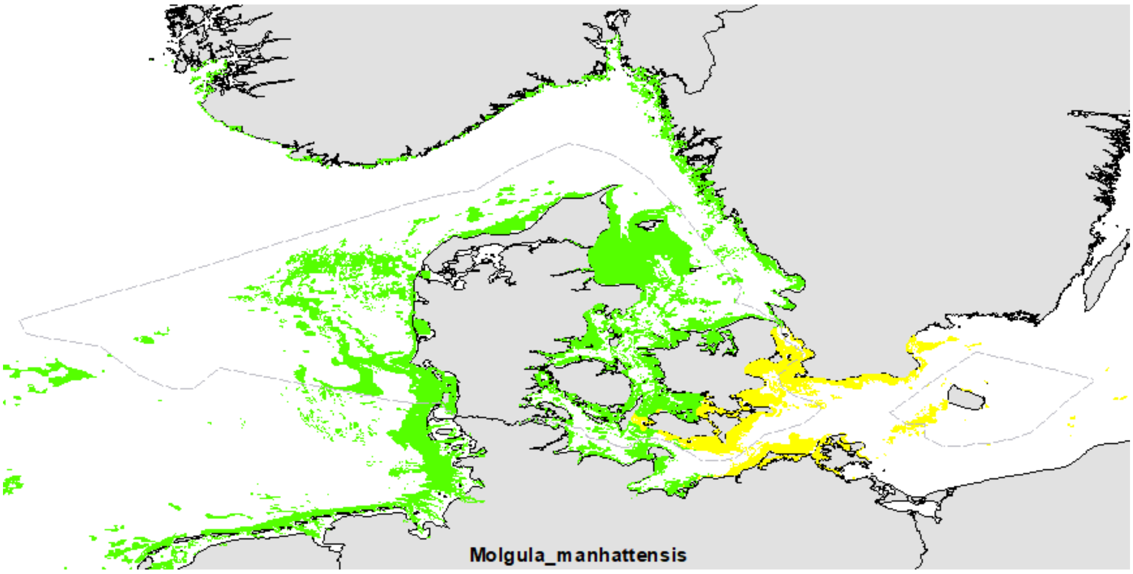

Crepidula_fornicata

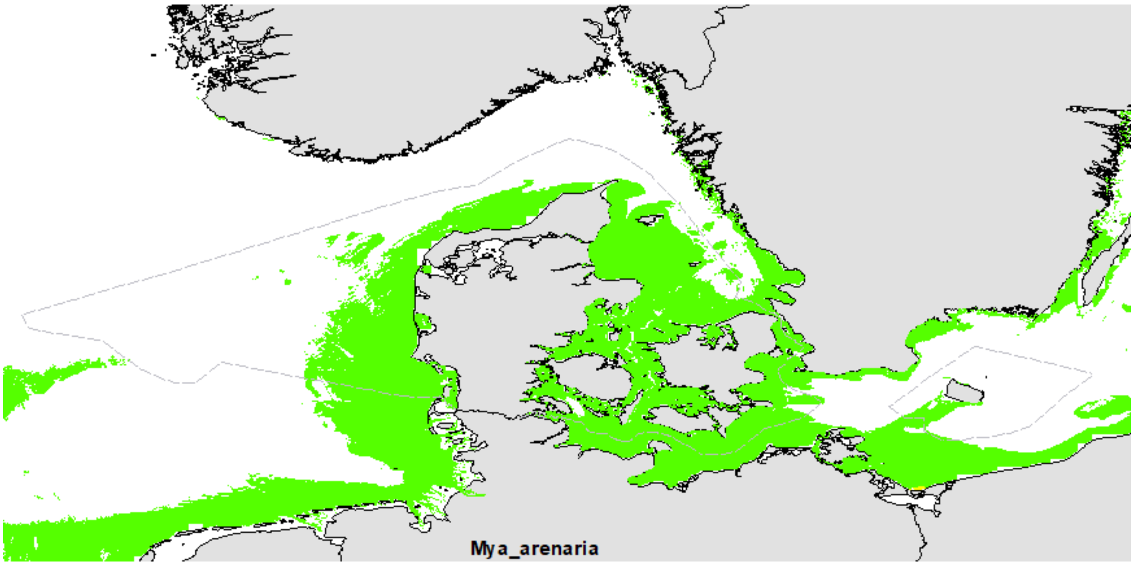

Ensis_directus

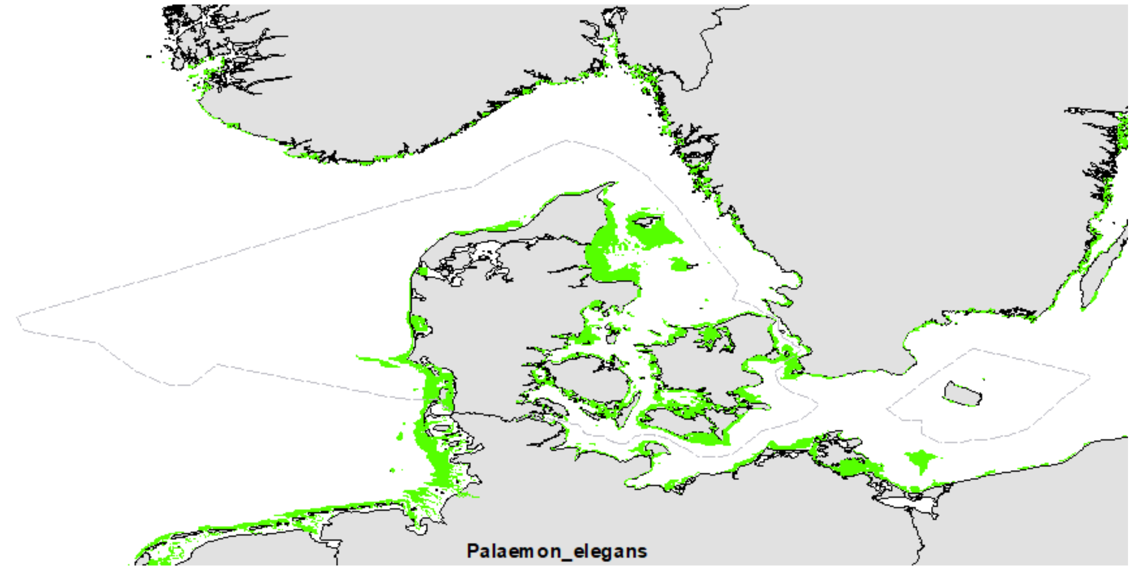

Hemigrapsus_sanguineus_sub

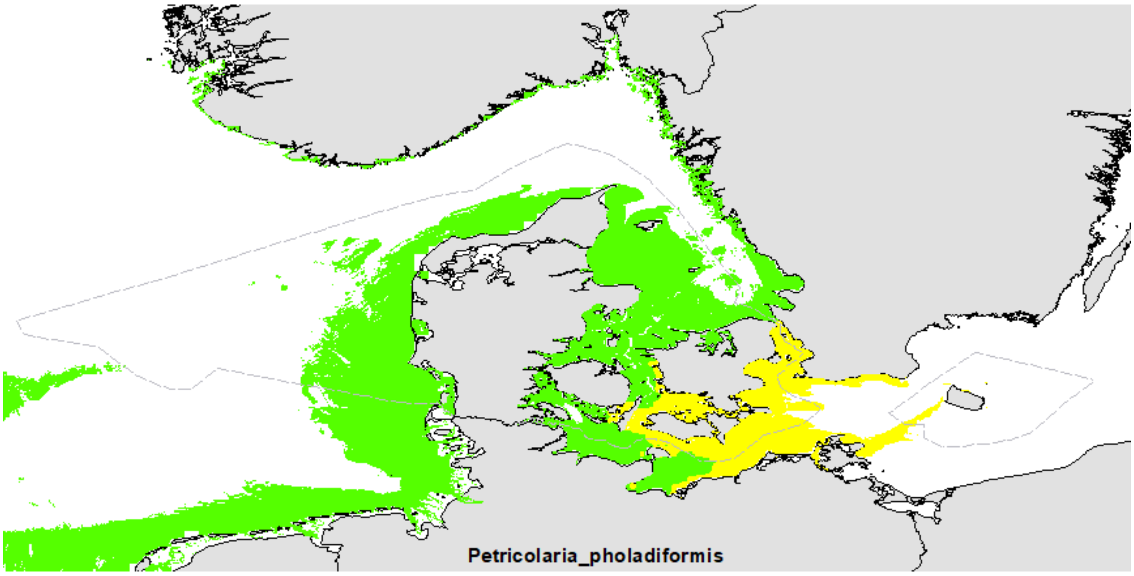

Hemigrapsus_takanoi

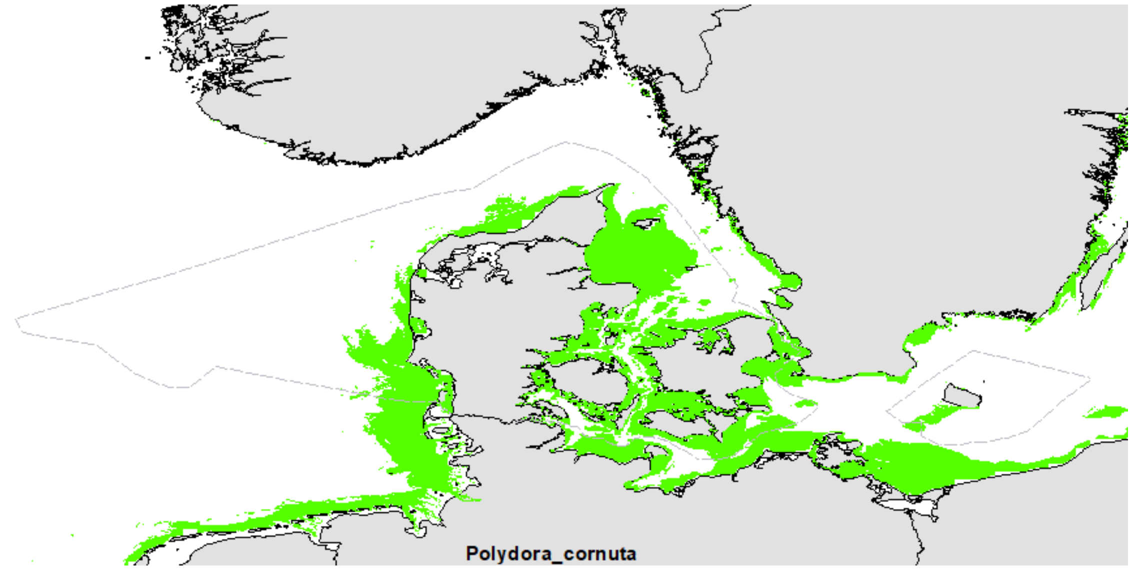

Marenzelleria_neglecta

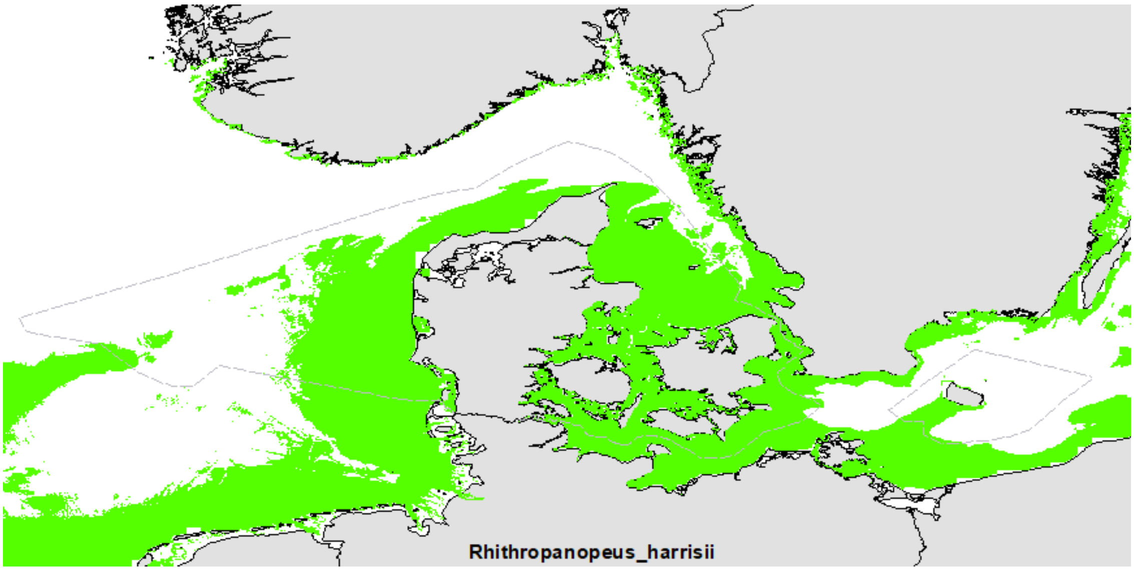

Molgula_manhattensis

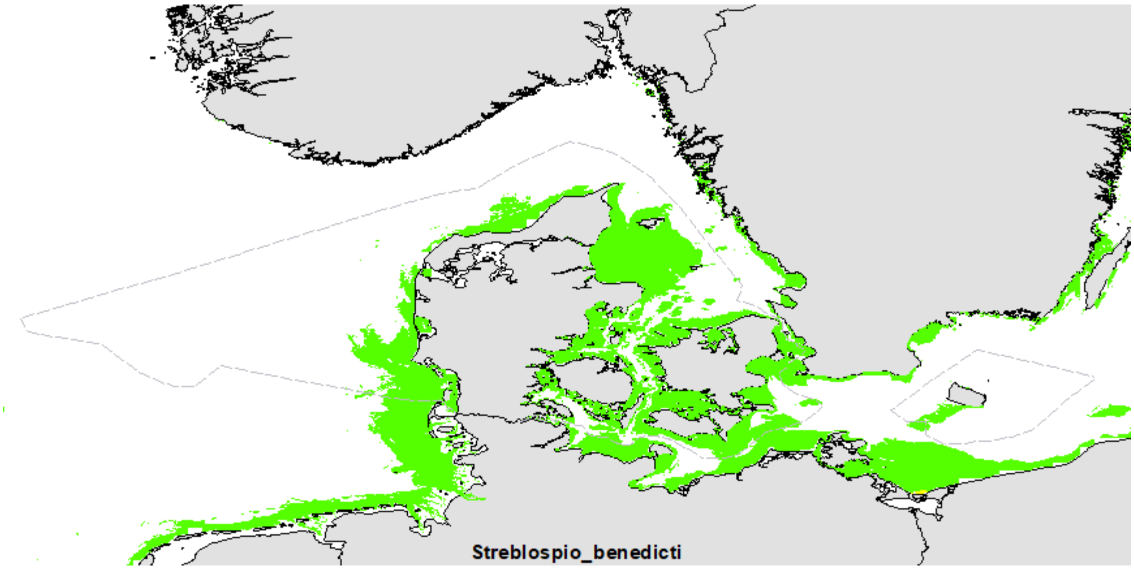

Mya_arenaria

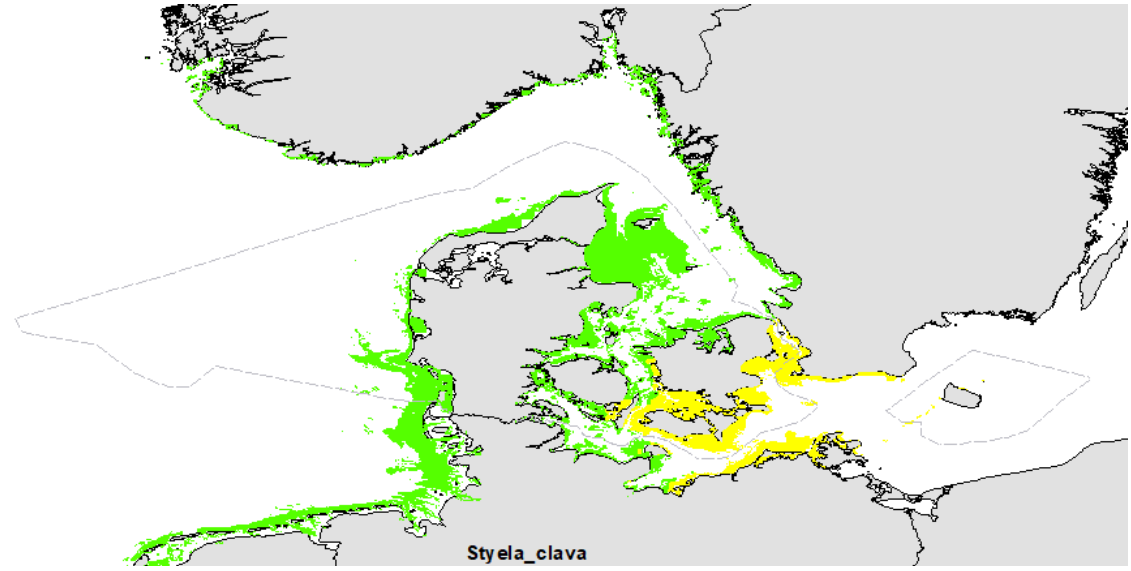

Palaemon_elegans


Petricolaria_pholadiformis


Polydora_cornuta

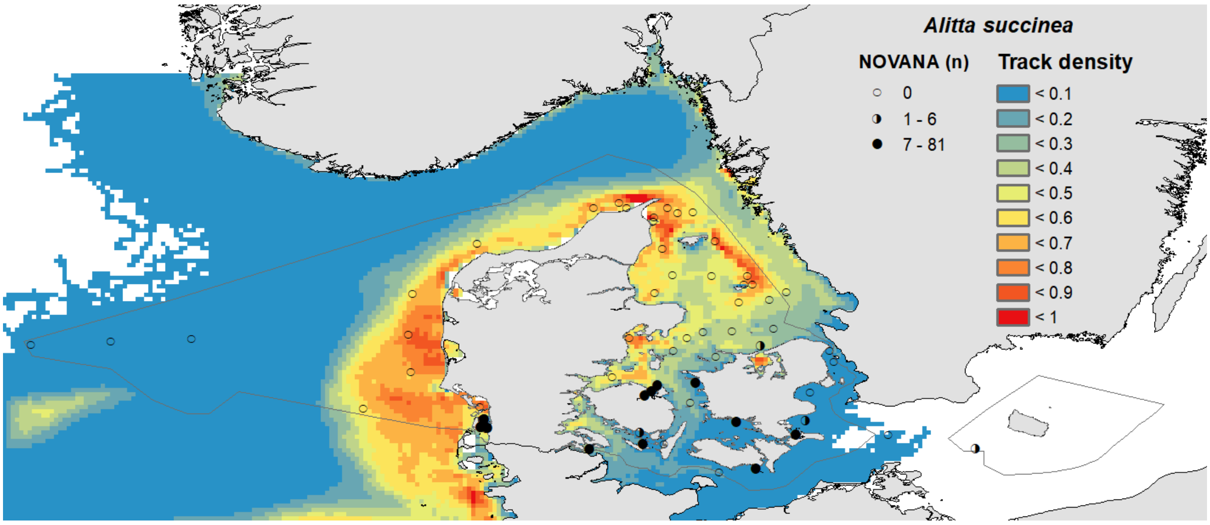

Rhithropanopeus_harrisii

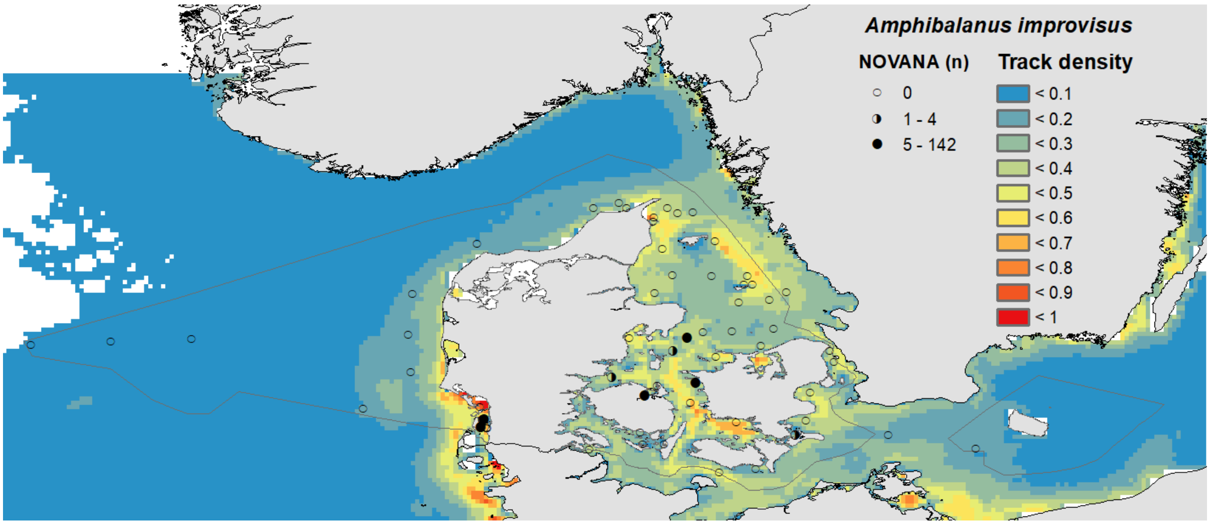

Streblospio_benedicti

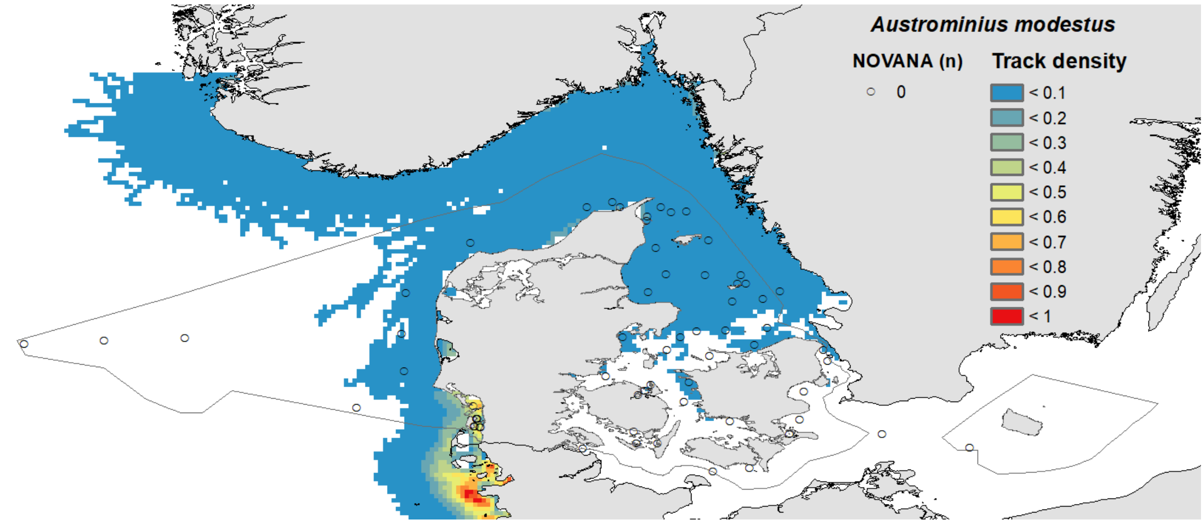

Styela_clava

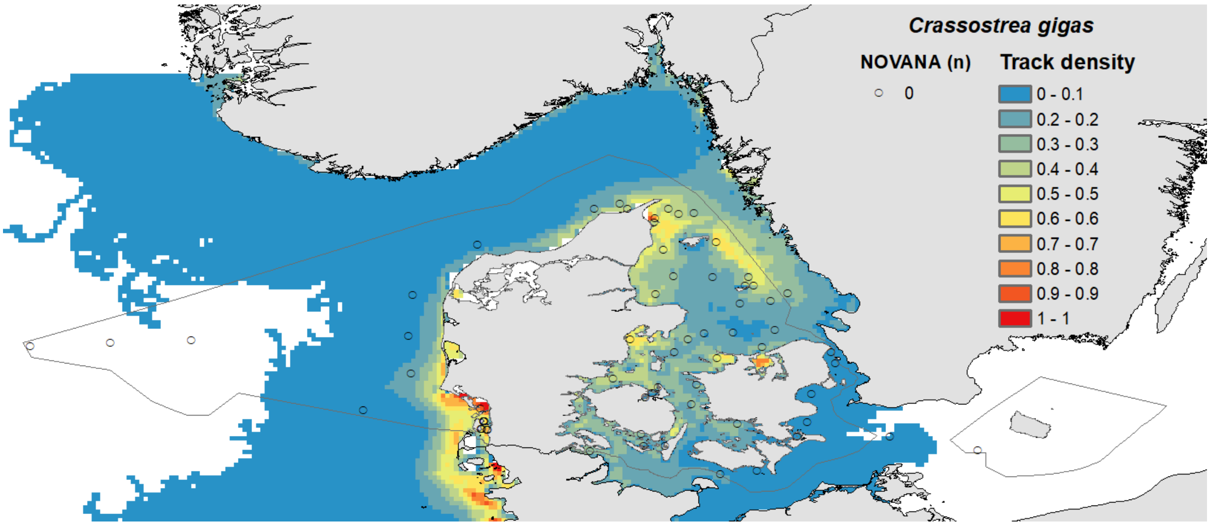
# Appendix 4. Tracking density maps

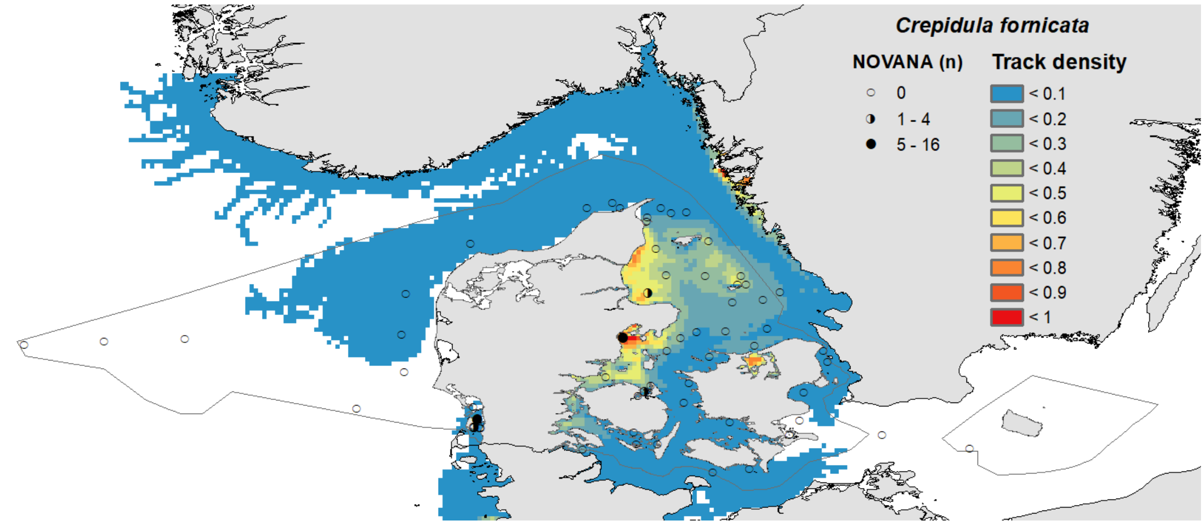
Color legends show track densities normalized. Dot legend represent absence (white) and presences (black) of NIS at NOVANA stations.

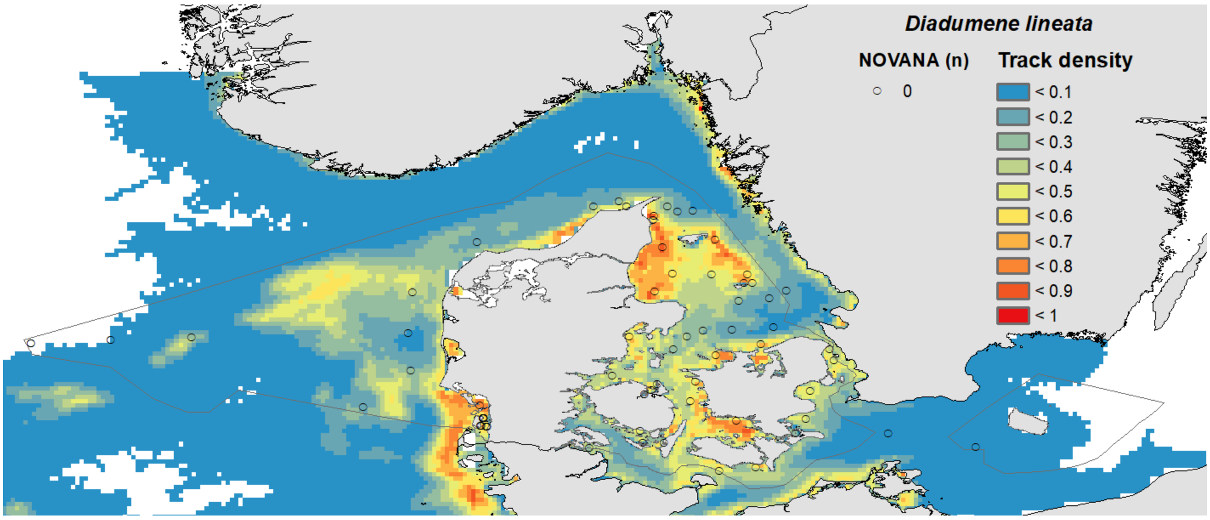


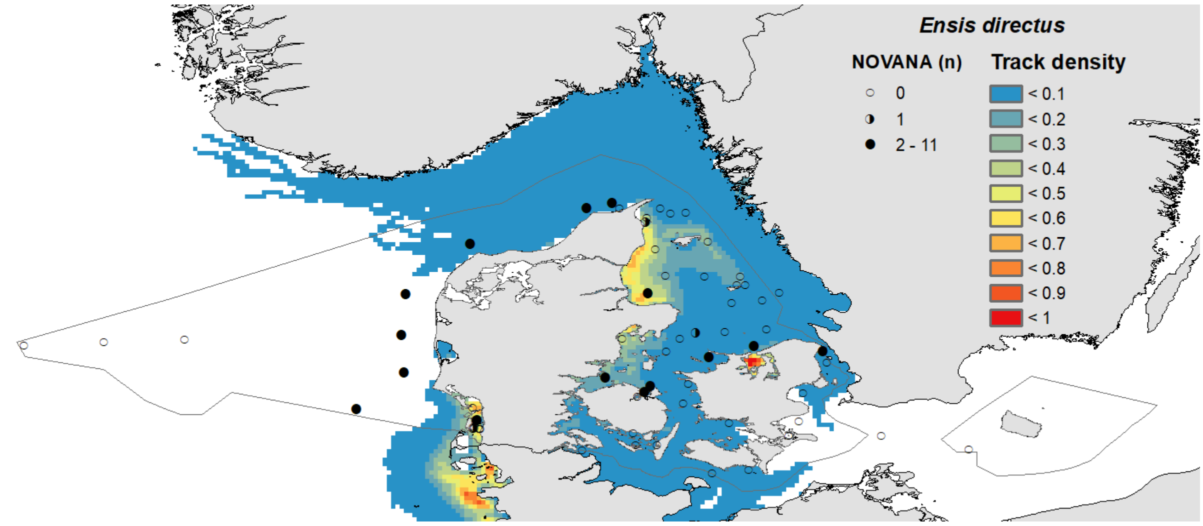


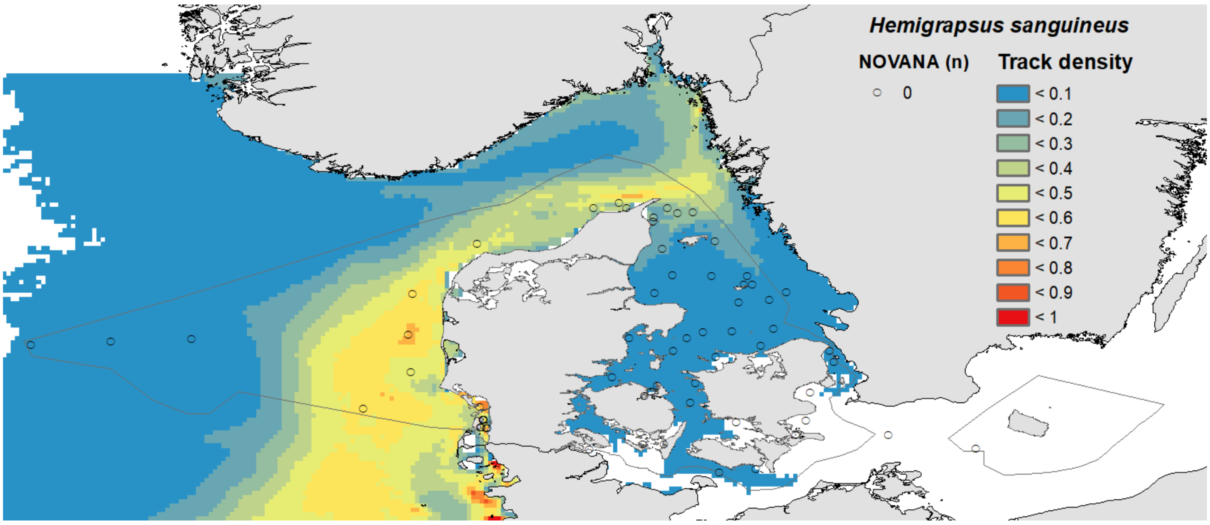

*Austrominius modestus*
NOVANA (n)
○ 0
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

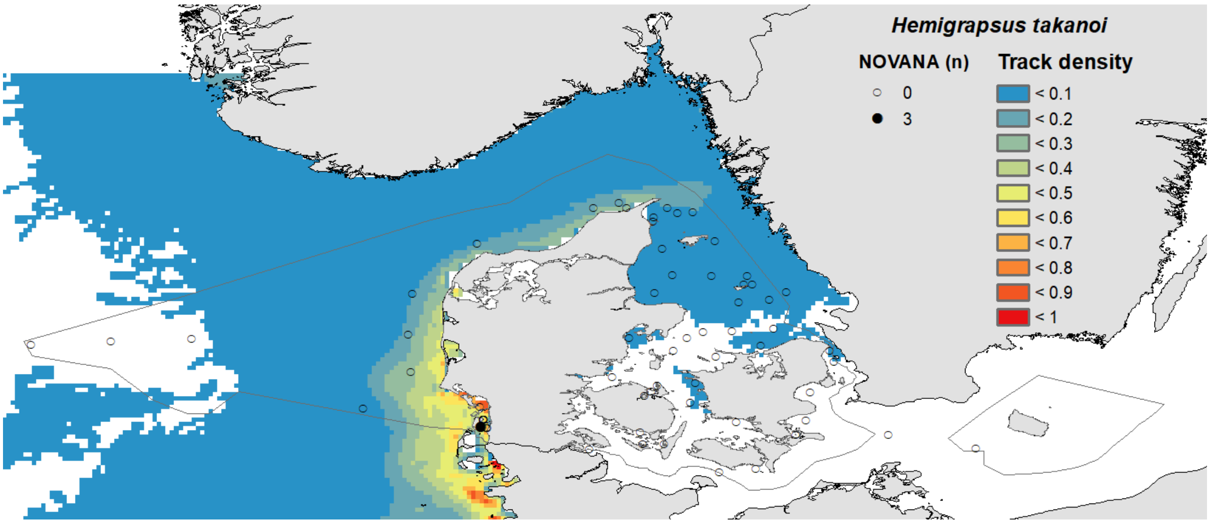

*Crassostrea gigas*
NOVANA (n)
○ 0
Track density
0 - 0.1
0.2 - 0.2
0.3 - 0.3
0.4 - 0.4
0.5 - 0.5
0.6 - 0.6
0.7 - 0.7
0.8 - 0.8
0.9 - 0.9
1 - 1

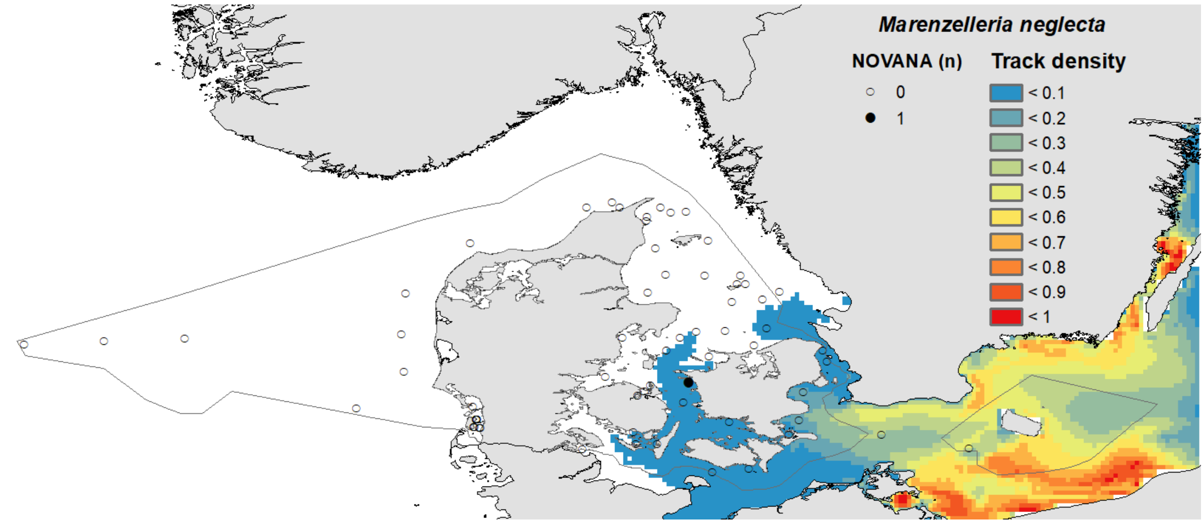

*Crepidula fornicata*
NOVANA (n)
○ 0
◑ 1 - 4
● 5 - 16
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

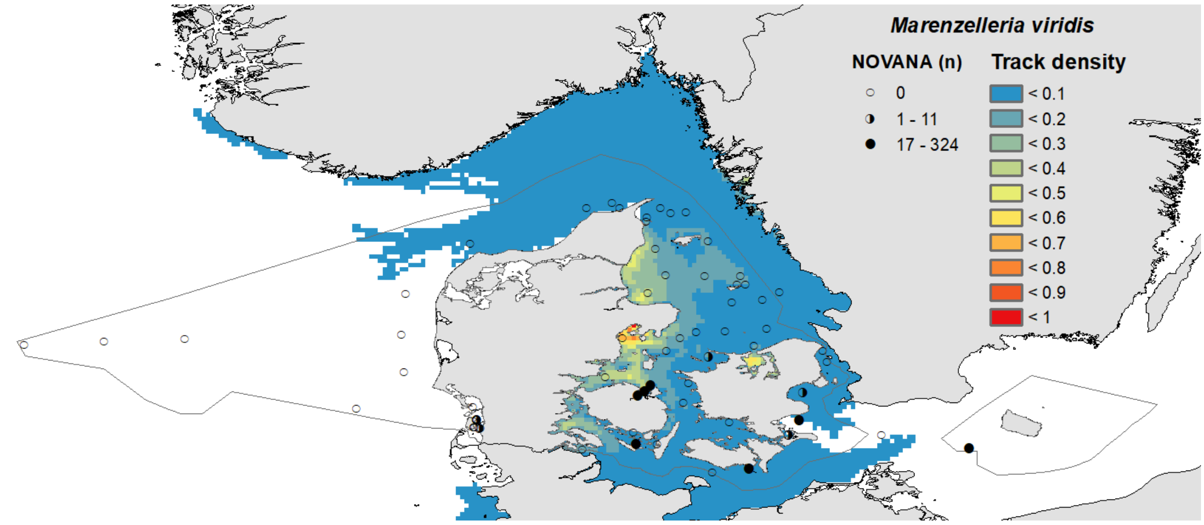

Diadumene lineata
NOVANA (n)
○ 0
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

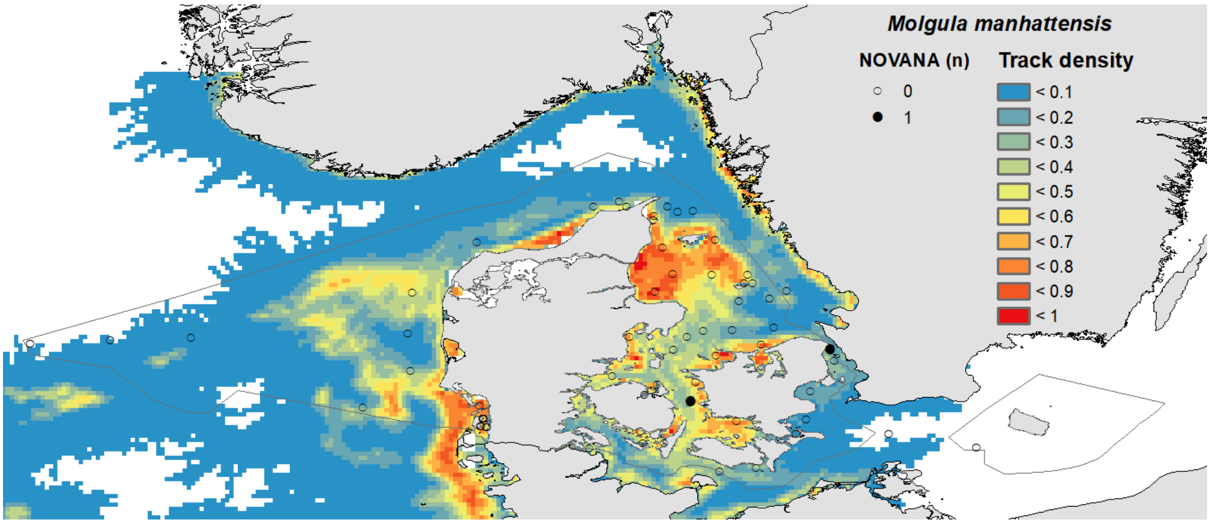

Ensis directus
NOVANA (n)
○ 0
◑ 1
● 2 - 11
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

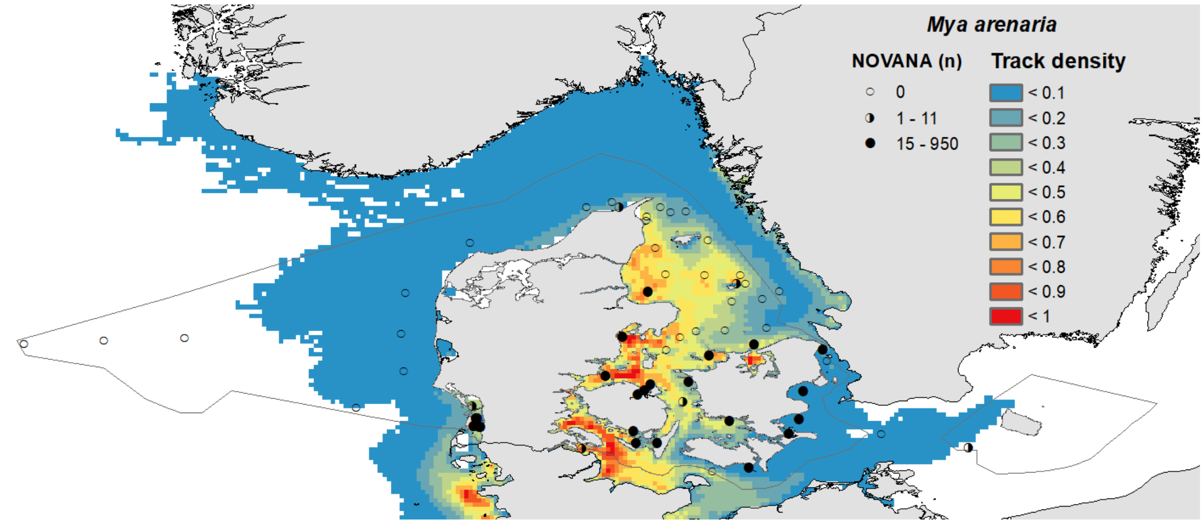

Hemigrapsus sanguineus
NOVANA (n)
○ 0
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

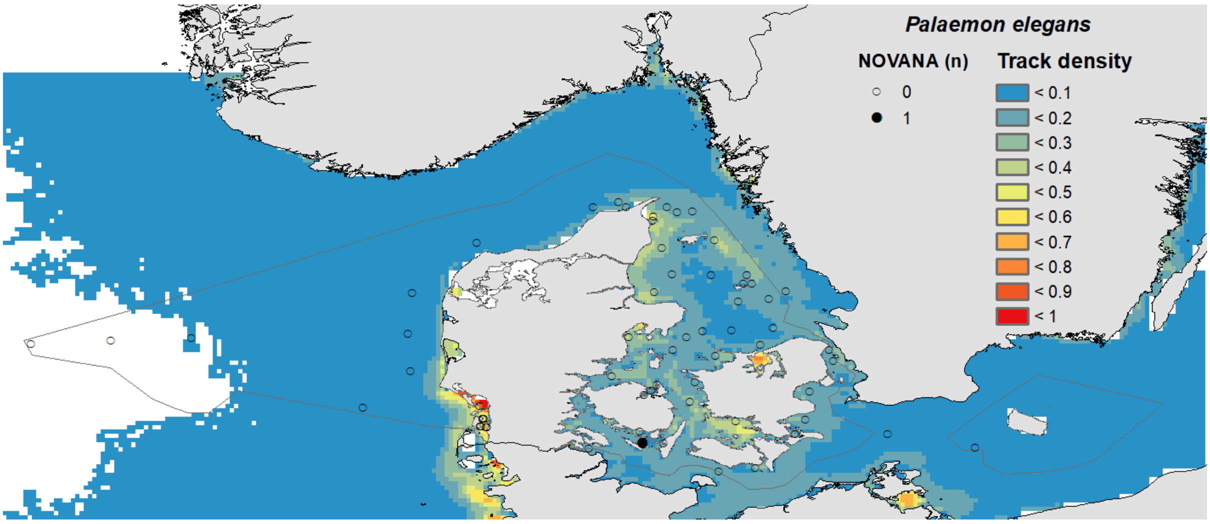

*Hemigrapsus takanoi*
NOVANA (n)
0
3
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

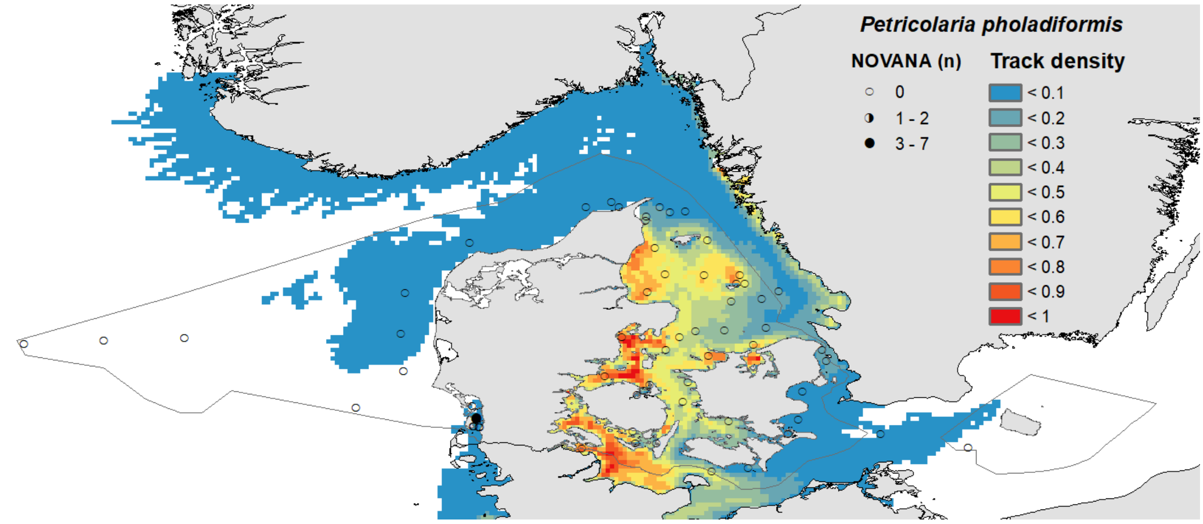

*Marenzelleria neglecta*
NOVANA (n)
0
1
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

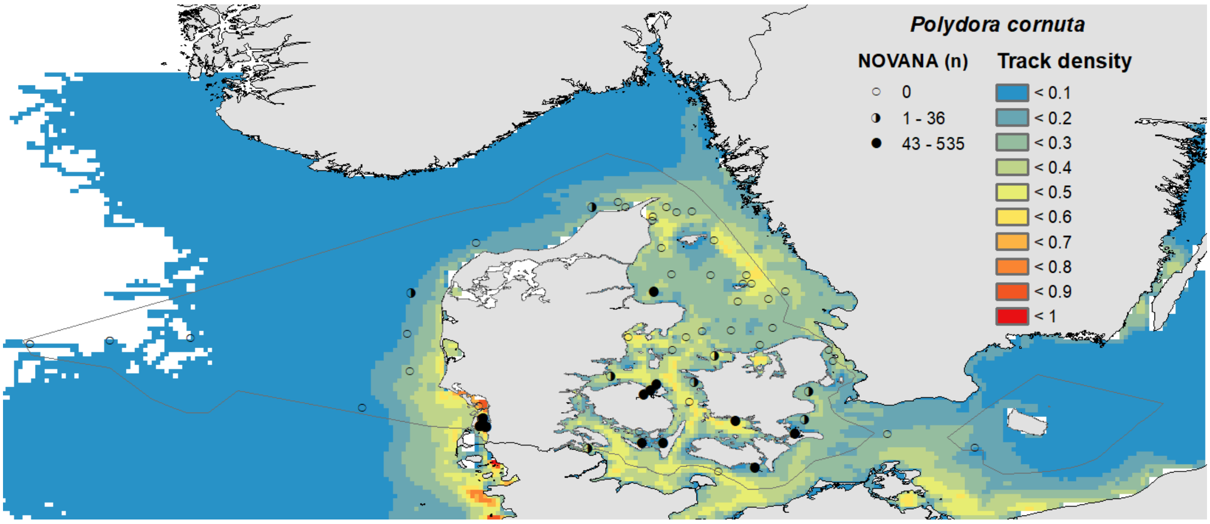

*Marenzelleria viridis*
NOVANA (n)
0
1 - 11
17 - 324
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

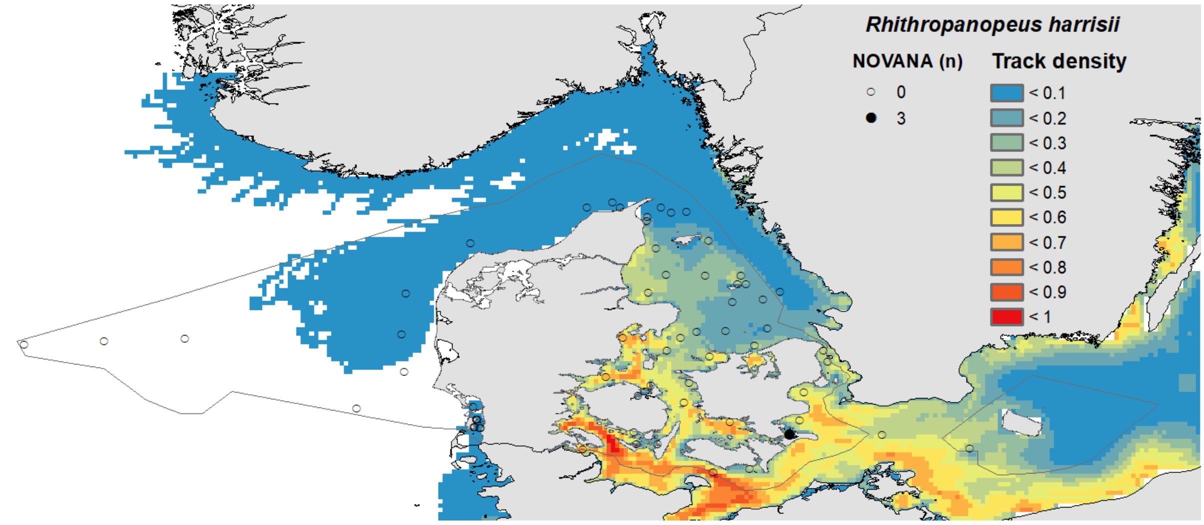

Molgula manhattensis
NOVANA (n)
○ 0
● 1
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

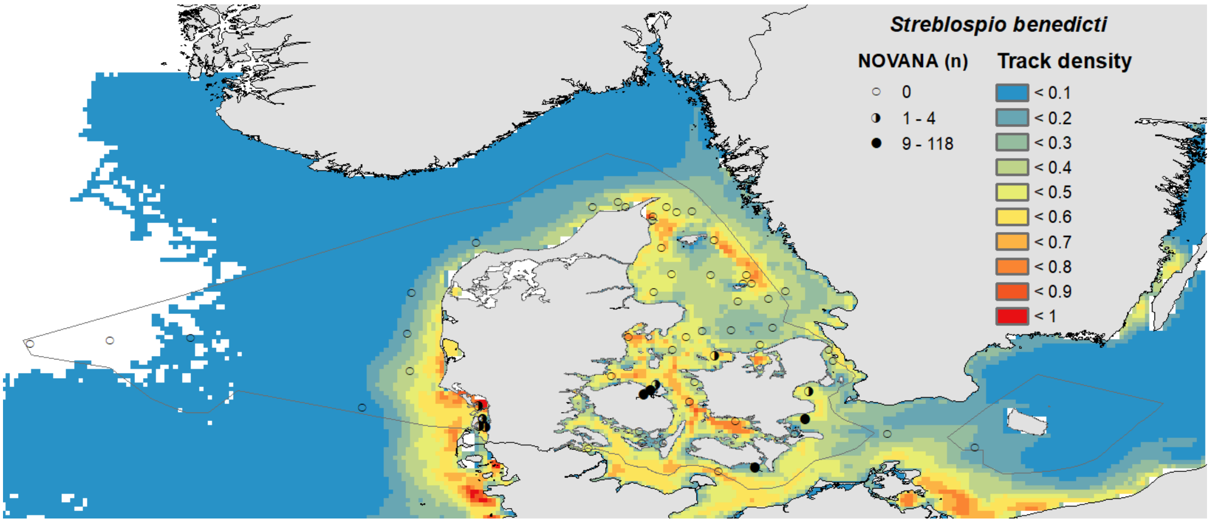

Mya arenaria
NOVANA (n)
○ 0
◑ 1 - 11
● 15 - 950
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

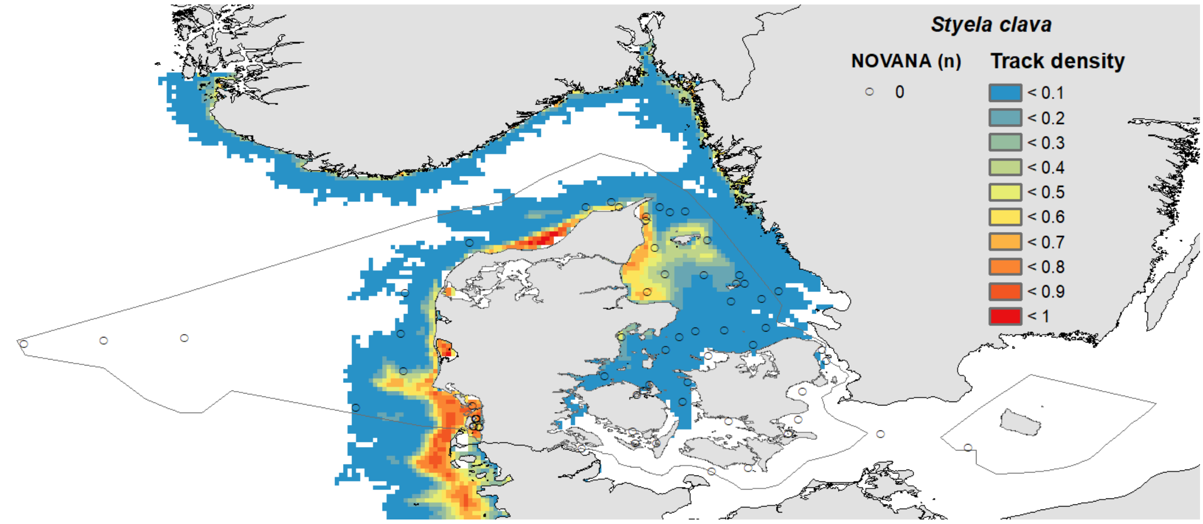

Palaemon elegans
NOVANA (n)
○ 0
● 1
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1


*Petricolaria pholadiformis*
NOVANA (n)
○ 0
◑ 1 - 2
● 3 - 7
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1


*Polydora cornuta*
NOVANA (n)
○ 0
◑ 1 - 36
● 43 - 535
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

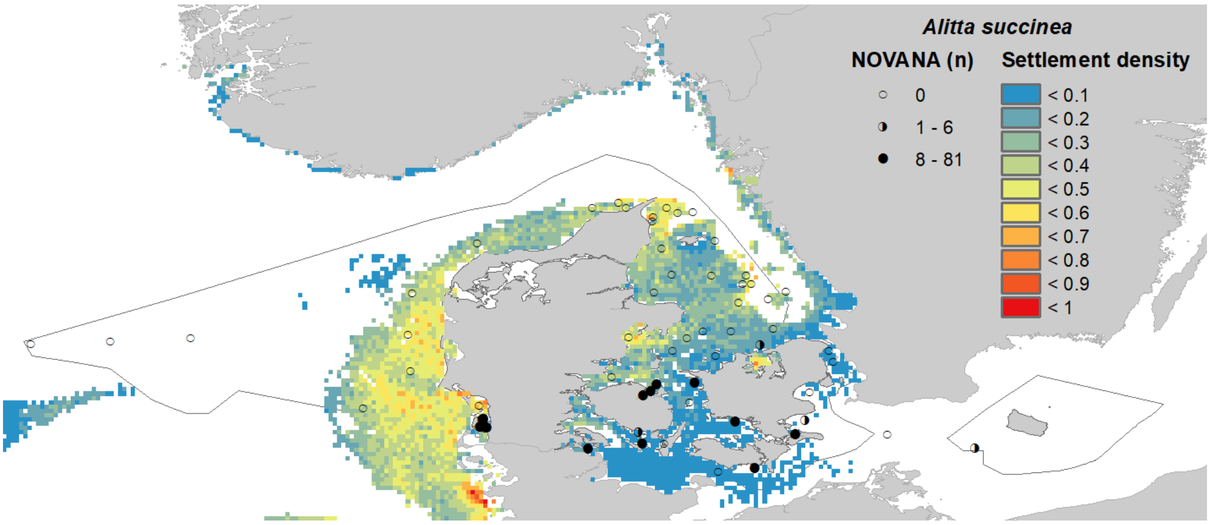

*Rhithropanopeus harrisii*
NOVANA (n)
○ 0
● 3
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

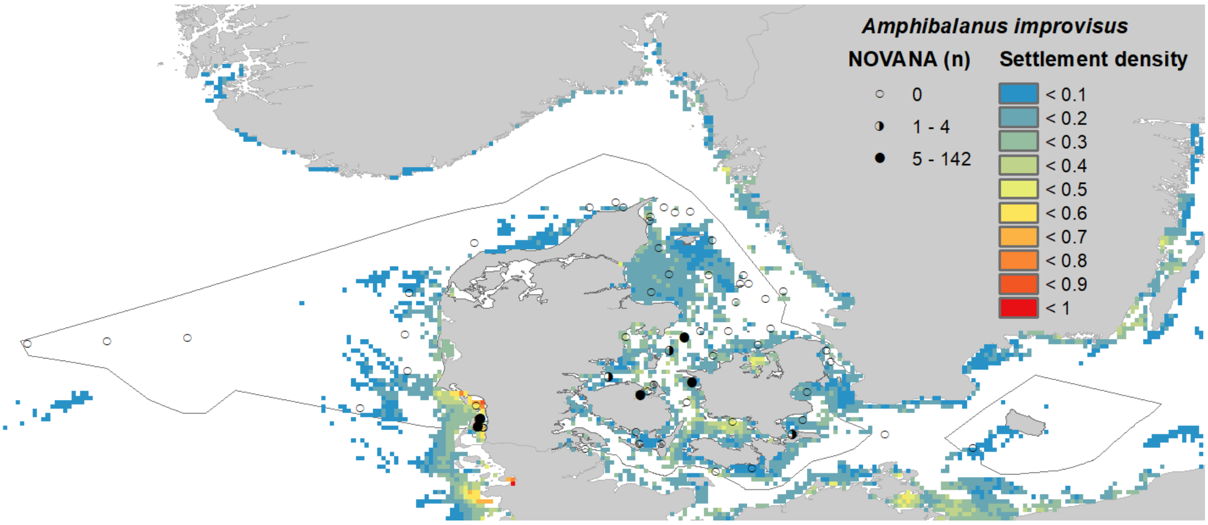

*Streblospio benedicti*
NOVANA (n)
○ 0
◑ 1 - 4
● 9 - 118
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

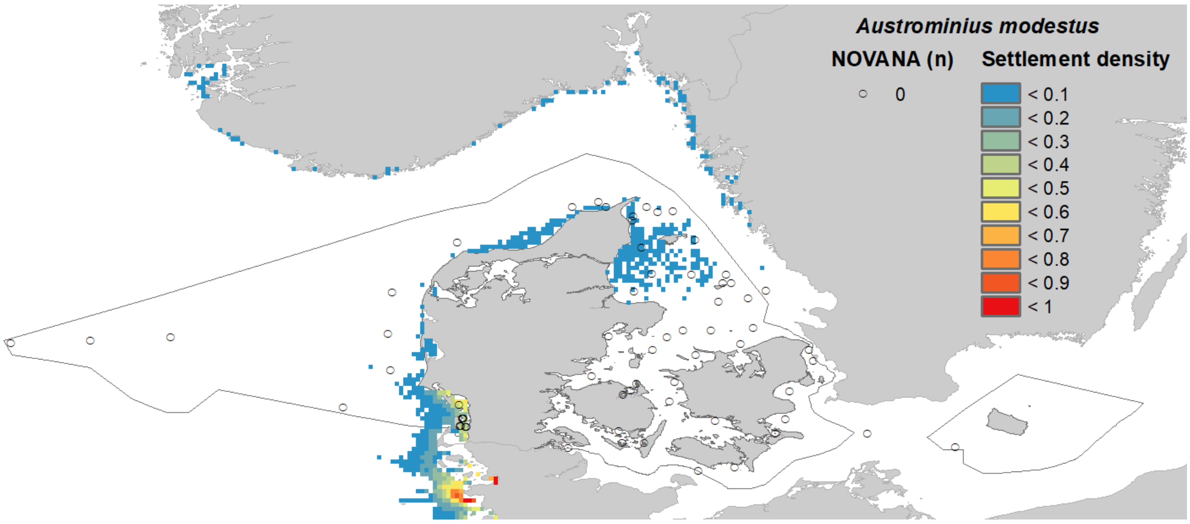

*Styela clava*
NOVANA (n)
○ 0
Track density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

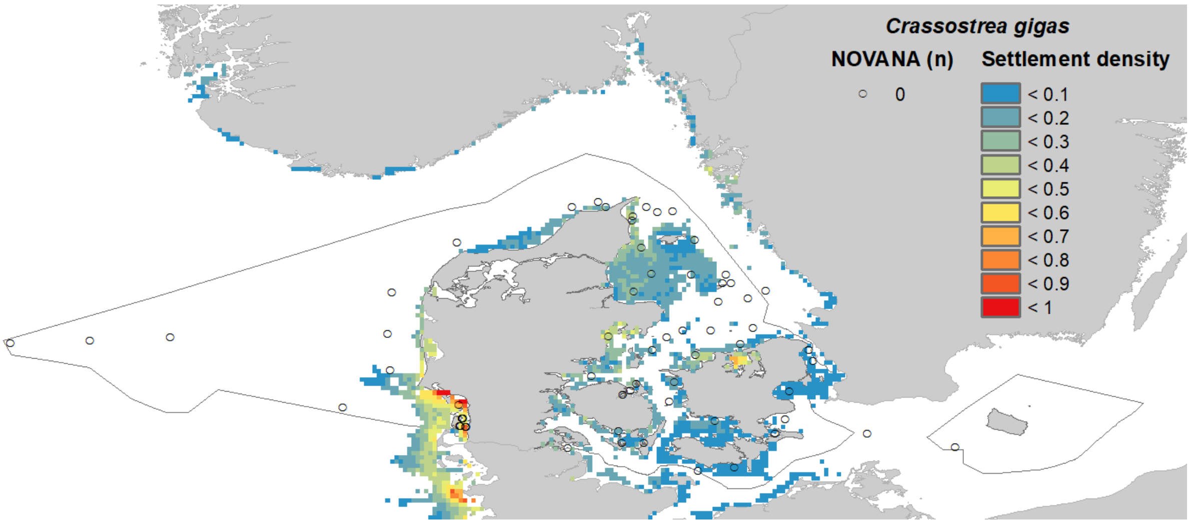
# Appendix 5. Settling density maps

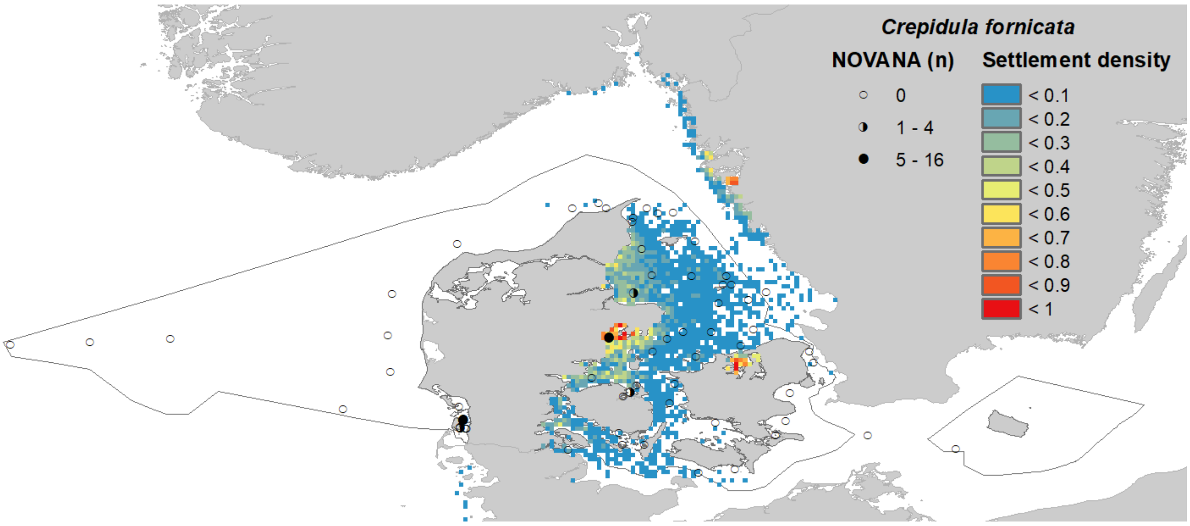
Color legends show settlement densities normalized. Dot legend represent absence (white) and presences (black) of NIS at NOVANA stations.

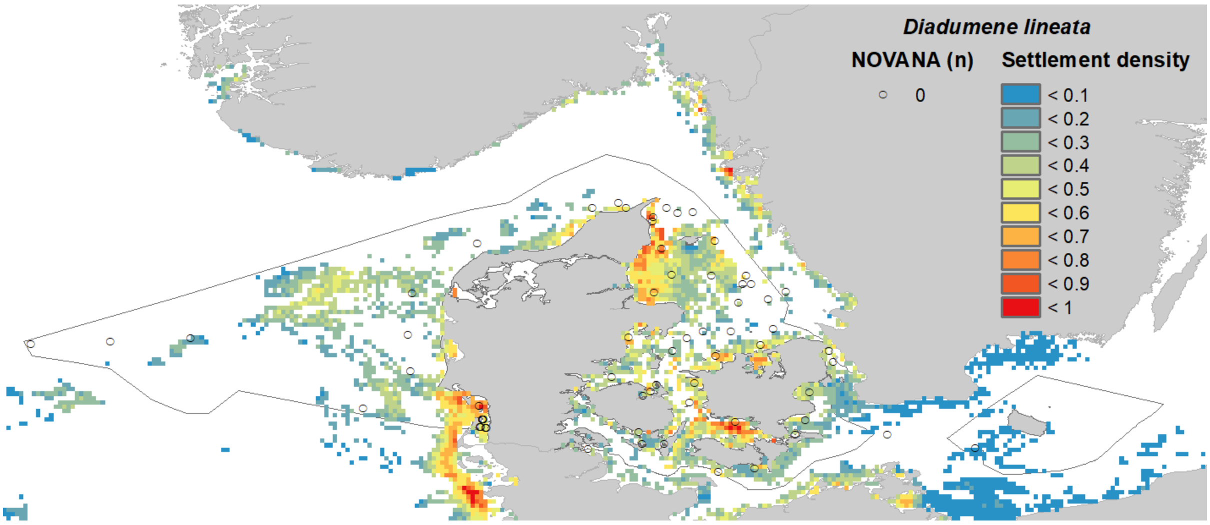


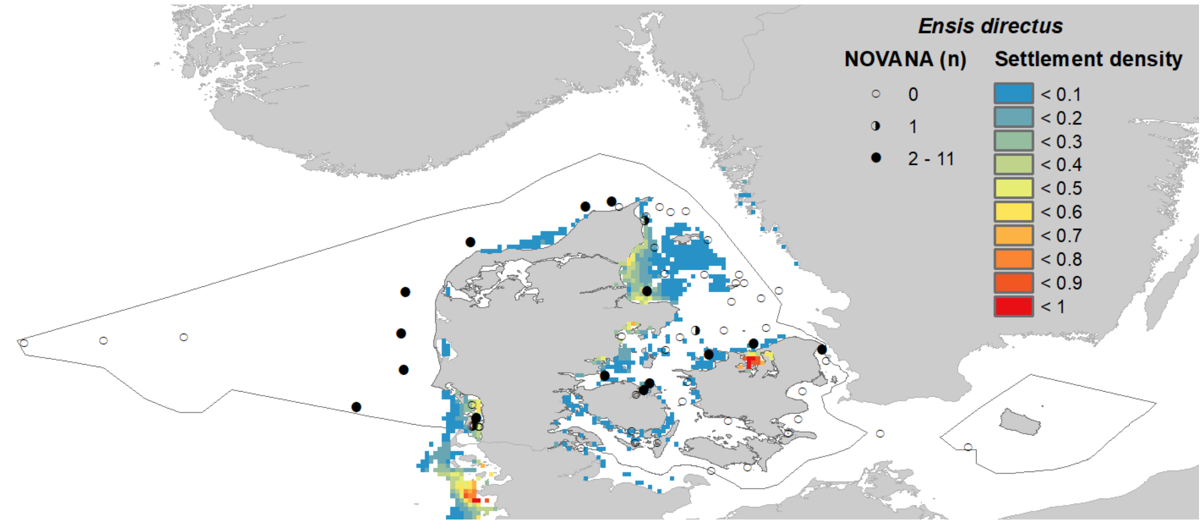


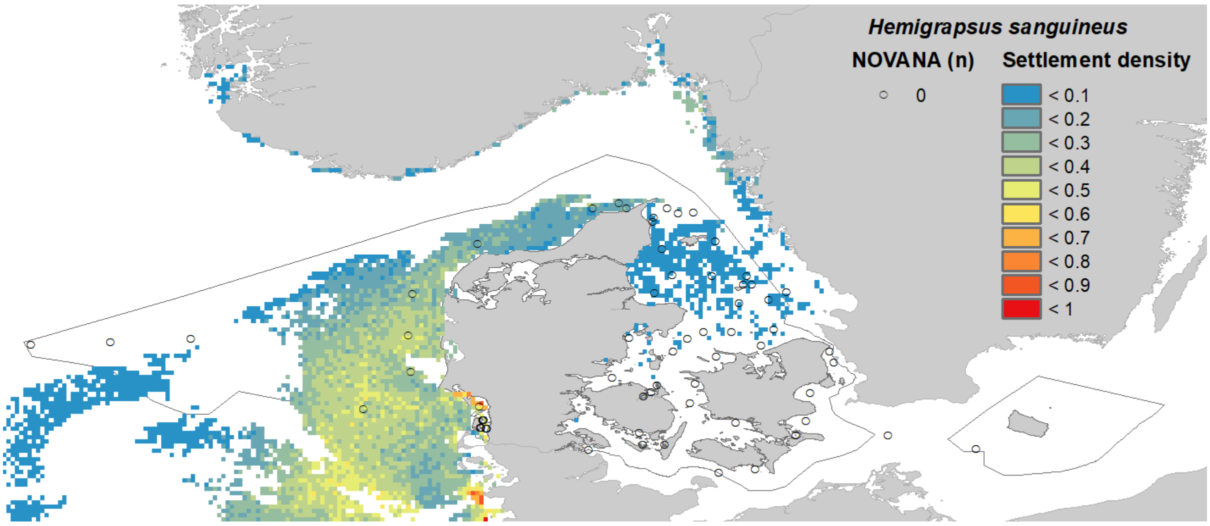

Austrominius modestus
NOVANA (n)
0
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

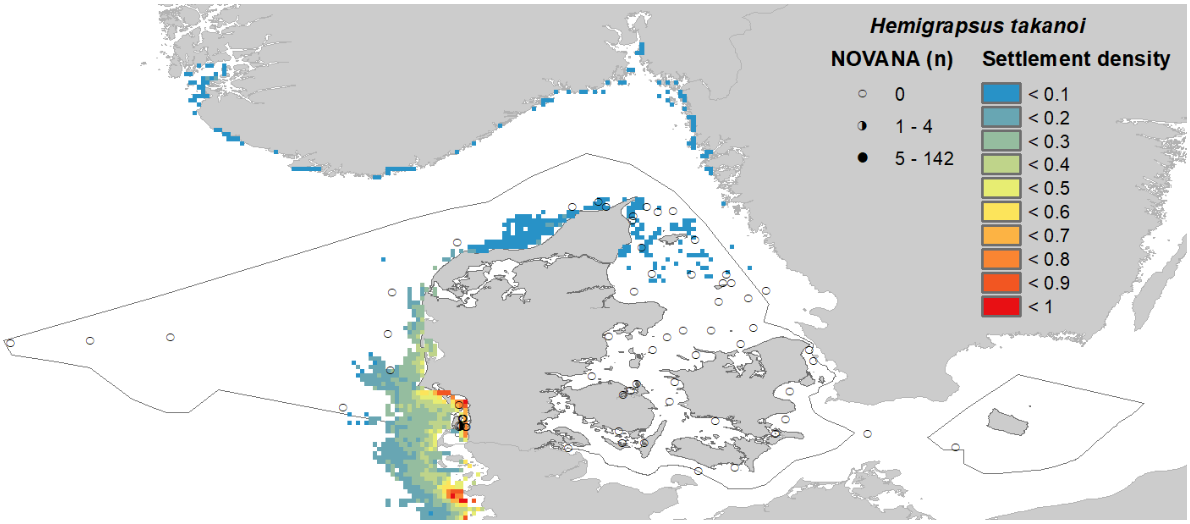

Crassostrea gigas
NOVANA (n)
0
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

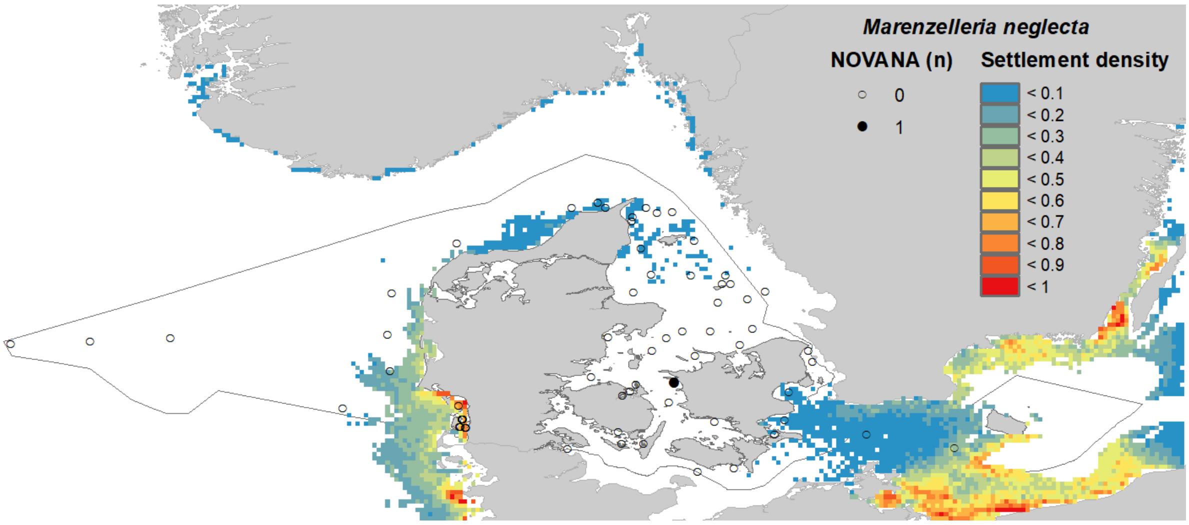

Crepidula fornicata
NOVANA (n)
0
1 - 4
5 - 16
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

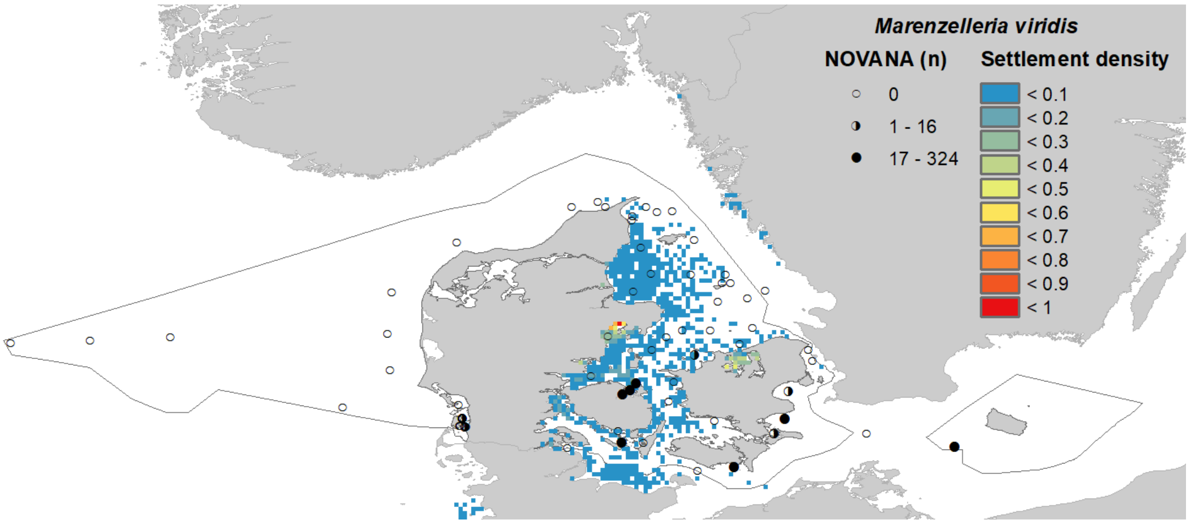

*Diadumene lineata*
NOVANA (n)
Settlement density
○ 0
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

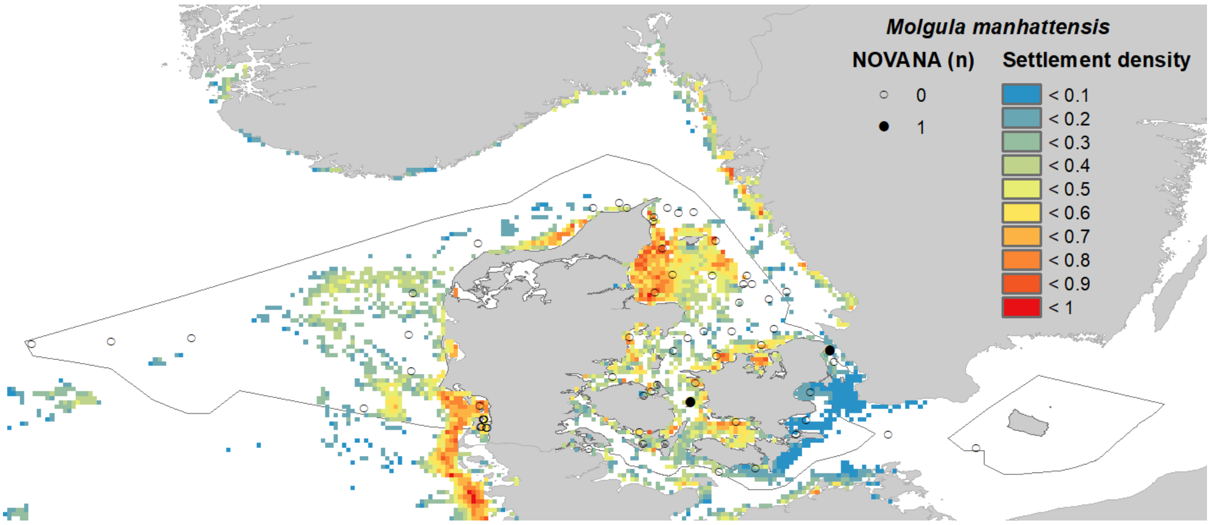

*Ensis directus*
NOVANA (n)
Settlement density
○ 0
◑ 1
● 2 - 11
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

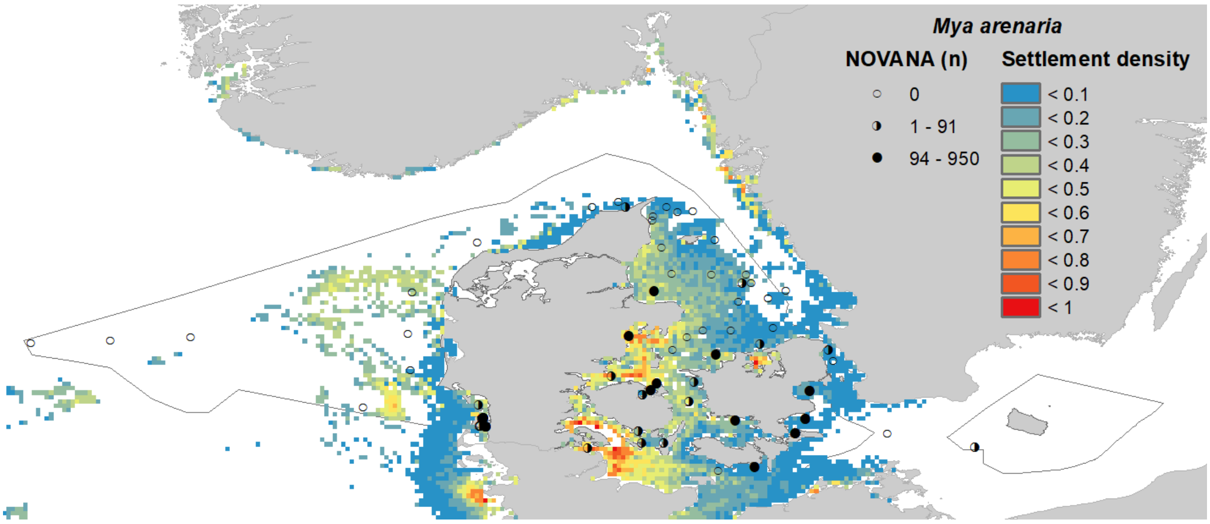

*Hemigrapsus sanguineus*
NOVANA (n)
Settlement density
○ 0
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

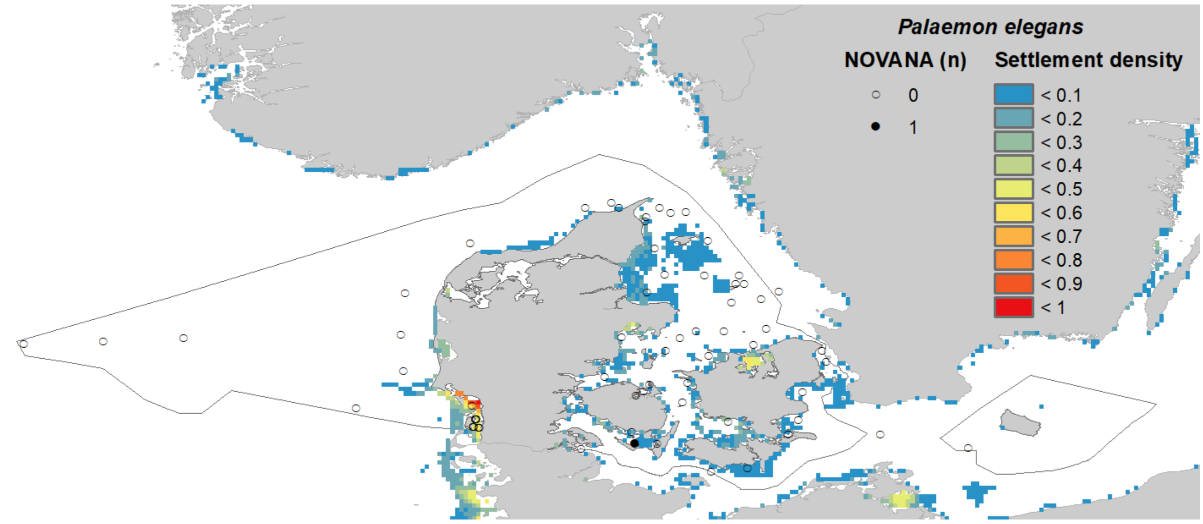

*Hemigrapsus takanoi*
NOVANA (n)
Settlement density
○ 0
◑ 1 - 4
● 5 - 142
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

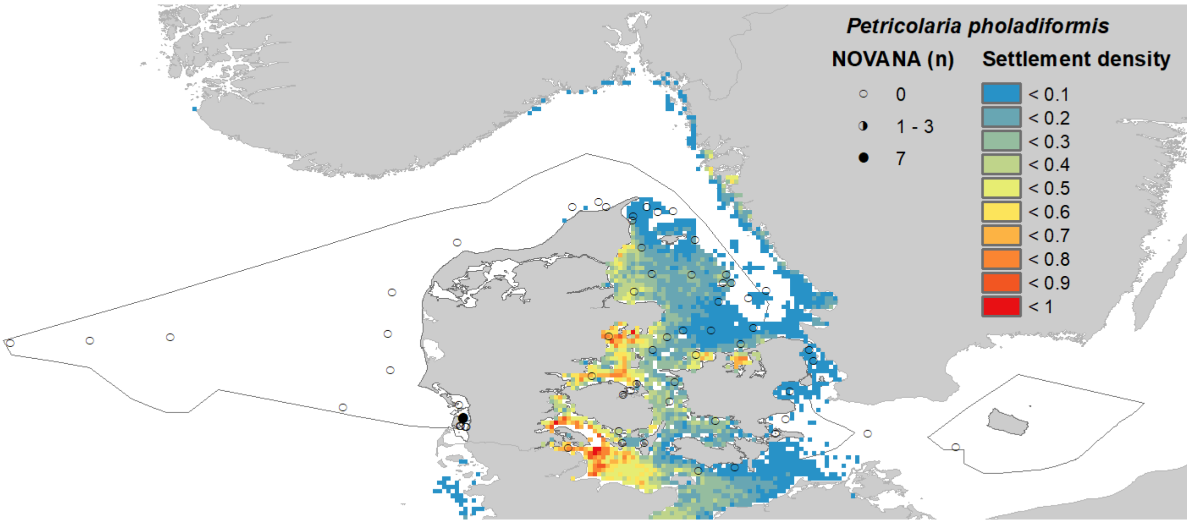

*Marenzelleria neglecta*
NOVANA (n)
Settlement density
○ 0
● 1
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

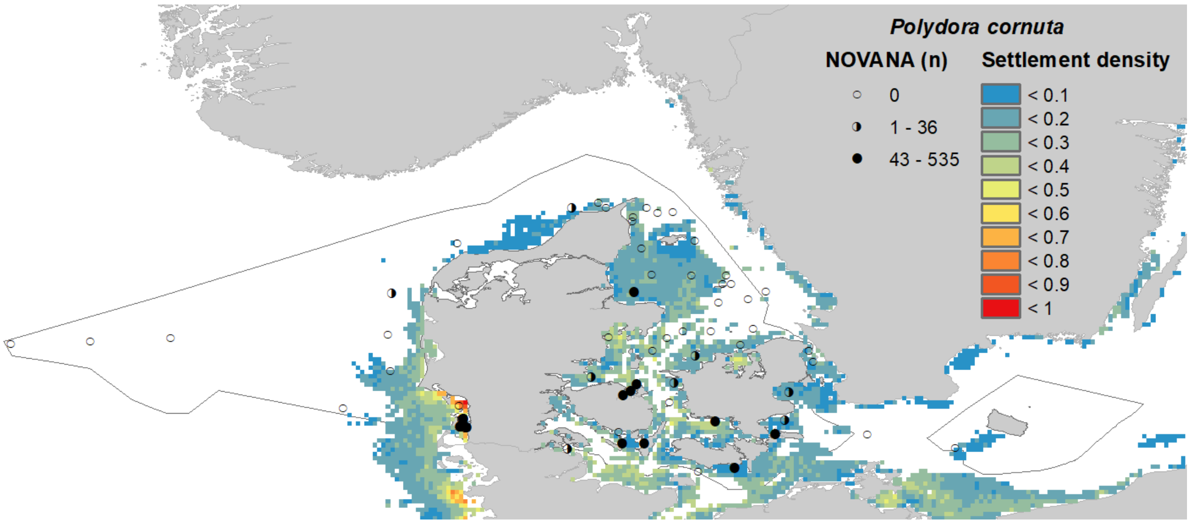

*Marenzelleria viridis*
NOVANA (n)
Settlement density
○ 0
◑ 1 - 16
● 17 - 324
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

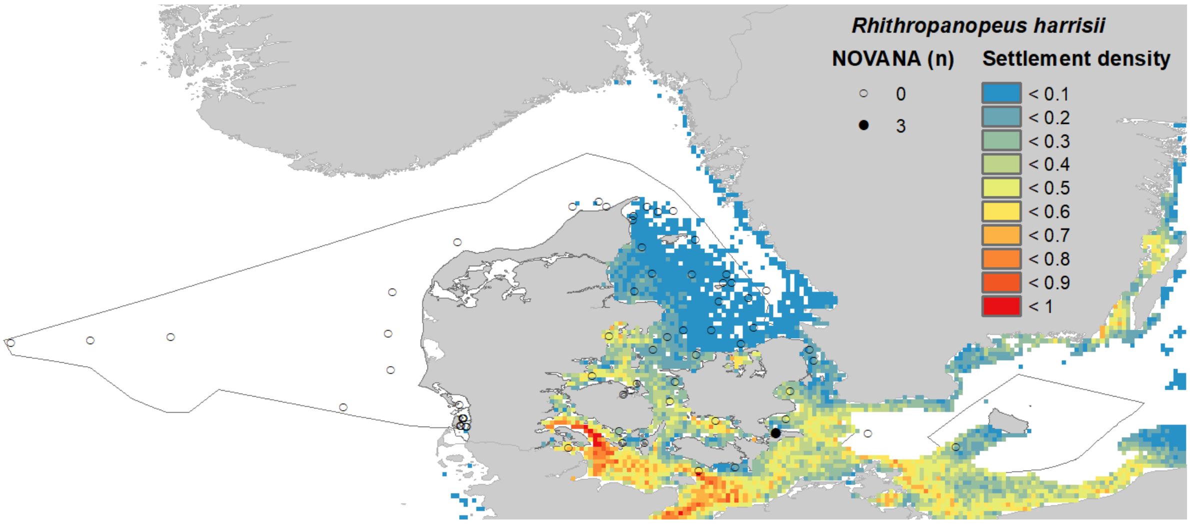

*Molgula manhattensis*
NOVANA (n)
○ 0
● 1
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

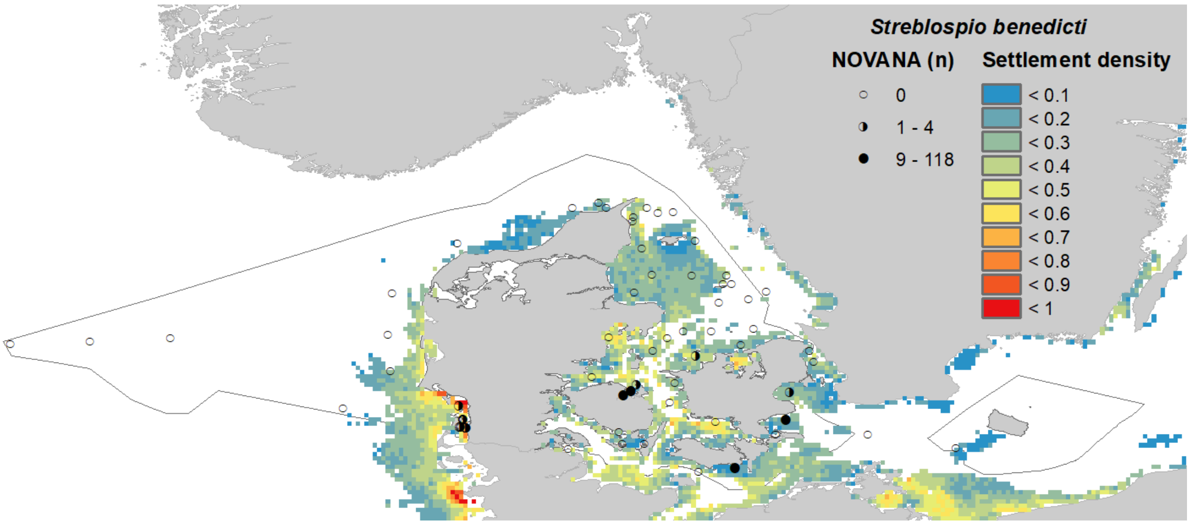

*Mya arenaria*
NOVANA (n)
○ 0
◑ 1 - 91
● 94 - 950
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

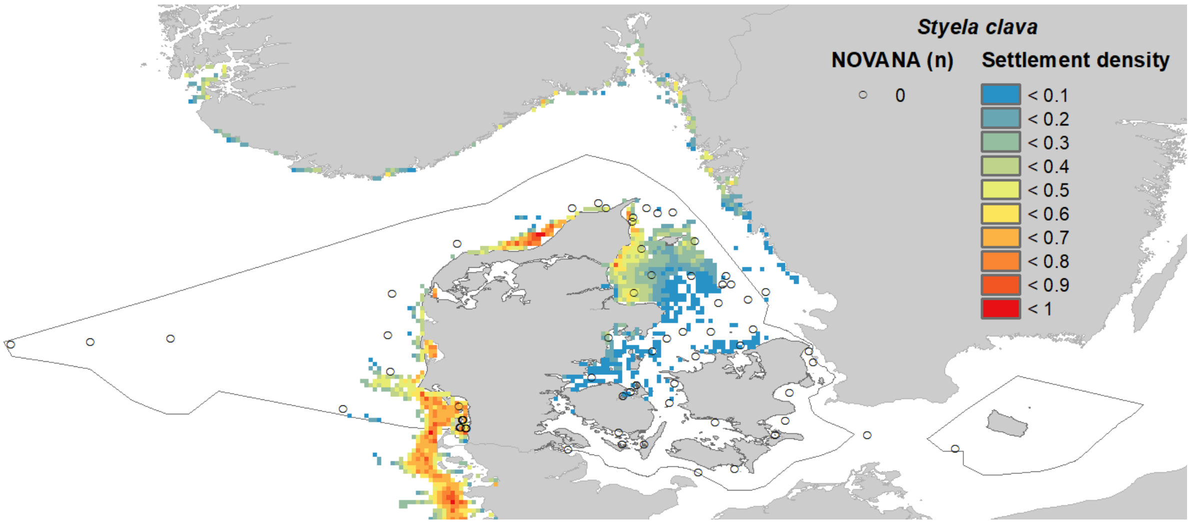

*Palaemon elegans*
NOVANA (n)
○ 0
● 1
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1


*Petricolaria pholadiformis*
NOVANA (n)
○ 0
◑ 1 - 3
● 7
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1


*Polydora cornuta*
NOVANA (n)
○ 0
◑ 1 - 36
● 43 - 535
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1


*Rhithropanopeus harrisii*
NOVANA (n)
○ 0
● 3
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1


*Streblospio benedicti*
NOVANA (n)
○ 0
◑ 1 - 4
● 9 - 118
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1


*Styela clava*
NOVANA (n)
○ 0
Settlement density
< 0.1
< 0.2
< 0.3
< 0.4
< 0.5
< 0.6
< 0.7
< 0.8
< 0.9
< 1

Technical
University
of Denmark

DTU Aqua
Kemitorvet
DK-2800 Kgs. Lyngby

www.aqua.dtu.dk