Cryptic mussels in a contact zone: towards to species distribution modelling

**V. M. Khaitov, A. A. Zaychikova, P. Y. Safonov, M. V. Katolikova, M. V. Ivanov, P. P. Strelkov**

## Introduction

One of the urgent tasks of the modern marine ecology is the construction of relevant species distribution models (SDM) as numerical tools describing the relationship between species occurrence and environmental parameters (Elith, Leathwick, 2009). These models allow to predict distribution patterns of ecologically and economically important species (Lowen et al 2019) and the amount of papers devoted to SDMs for marine species has been growing significantly in recent years (Robinson et al 2017). Importantly that SDMs can serve as a basis for prognosis of species distribution both in unexplored areas (Russo et al 2014) and in the future as a consequence of environmental changes (Buisson et al 2010 but see Santini et al 2021 for critical discussion).

Although the biological unit for SDM can be either species, genetically distinct subpopulations, or even genetic lineages (Lowen et al 2019), however, in marine ecology SDMs are currently developed for “good” species whose taxonomic status is not in doubt (e.g. Reiss et al., 2011; Lindegren et al., 2022). The taxonomic certainty gives an opportunity to focus on solving actual problems related to the search for modeling algorithms (see Elith et al 2006 for methods comparative analysis) and/or identification of the most significant environmental predictors in marine realm (Bosh et al 2018). However, the increasing evidence for cryptic species among marine animals (Bastrop et al 1998; Fiser et al 2010) makes it urgent to build SDMs in such cases as well. Only few papers have directly addressed the construction of SDMs for marine animals’ within species genetic lineages and cryptic species (e.g. Dennis, Hellberg, 2010; Lowen et al 2019).

The construction of SDM for cryptic species has its own peculiarities. Dealing with cryptic species, forming parapatric or sympatric populations, both genetic processes (e.g., hybridization) and ecological interactions between closely related species potentially can influence the result of analysis and consequently model prediction quality (see Fielding&Bell, 1997). Strictly speaking, in the case when SDM is built not for a single species, but for several coexisting ones, which is the case of sympatric cryptic species, the Joint Species Distribution Modelling (JSDM) framework should be considered (Ovaskainen, Abreg, 2020). Whereas the standard SDM approach assume to use the presence/absence (or abundance) of a species as a response variable, which is under control of numerous environmental predictors, in JSDM framework some correlative assessment of interspecific interactions should be incorporated into the model as well (Ovaskainen, Abreg, 2020).

Blue mussels from “*Mytilus edulis*” complex (Gardner 1996; Varvio et al., 1988) provide a good opportunity to work out the approaches to SDM construction in the case of coexisting cryptic species.The native ranges of Atlantic *M. edulis* (thereafter *ME*) and pacific *M. trossulus* (*MT*) are differ (Riginos, Cunningham, 2005), but at North Atlantic coasts they form mixed settlements in several contact zones (Varvio et al., 1988; Hilbish et al., 2000; Mathiesen et al. 2016; Brooks, Farmen, 2013; Katolikova et al. 2016; Vainola, Strelkov 2011; Wenne et al. 2020). In these areas the proportion of *MT:ME* can vary in broad limits even in small spatial scales (Katolikova et al. 2016) that allows considering the *MT:ME* ratio as a very sensitive variable responding both to the influence of environmental factors and/or interactions between *MT* and *ME*. Thus the use of *MT:ME* ratio brings us closer to the JSDM framework but allow to consider single dependent variable describing the structure of a community consisting of two interacting species.

The SDMs are also considered as a tool of species’ ecologial niche describing (Elith, Leathwick, 2009). Thus in the case of use *MT:ME* ratio as a dependent variable in SDM the later would describe not only species’ spatial distribution but divergence of their niches as well. To date the knowledge about differences in parameters of ecological niches of these two mussel’s species is rather incomplete and contradict (see bellow) but it is known that these species differently interact with humans: the commercial value of both species as objects of aquaculture is not equal (Penney et al., 2008). The unexpected appearance of *MT* leads to dramatic decrease of commercial value of shellfish farm’s production (Beaumont et al 2008). In this case the relevant SDM with good predictive power would help to find areas for aquaculture development with a minimum probability of *MT* invasion.

It is generally assumed that the *MT:ME* ratio in the contact zone varies along certain environmental gradients associated with some key factors (Väinölä and Hvilsom 1991; Riginos and Cunningham 2005; Väinölä and Strelkov 2011). On a biogeographic scale, the distribution of *Mytilus* species is thought to be regulated by regional temperature and its correlates (Popovic & Riginos, 2019; Wenne et al., 2020). *MT* is believed to be a more northerly species, more tolerant to cold waters inhabiting areas that once existed in glaciation zones. *ME* potentially can distribute in cold waters but tend to form settlements in more temperate latitudes (Wenne et al., 2020).

Salinity was often considered as a key factor regulating the spatial distribution of species in *ME*-*MT* contact zones. *MT* are thought to gravitate toward areas of more freshened water, and *ME* toward higher salinity (Riginos & Cunningham, 2005; Ridgway & Naewdal, 2004; Kijewski et al., 2019; Ridgway & Naevdal, 2004). For example, *MT* form virtually single-species populations in the heavily freshened Baltic Sea (Gosling, 1992; Vainola Strelkov, 2011) but *ME* proportion dramatically increase when crossing the Danish straits toward the North Sea where oceanic salinity occurs (Kijewski et al., 2019). In other contact areas (Gulf of St. Lawrence, Mareau et al. 2005, Bergen area, Ridgway & Naevdal, 2004) a similar pattern is observed: *MT* settlements are represented in the most desalinated areas.

Wave action was also considered as an important factor regulated species distribution: *MT* has been shown to gravitate toward more turbulent habitats and *ME* toward quieter areas (Tam, Scrosati, 2014; Comesaña et al., 1999; Bates, Innes, 1995). The pattern of species segregation by substrate has also been attempted to be linked to surf exposure (Katolikova et al, 2016): *MT* tended toward fucoid thalli, and *ME* toward the surface of bottom stony or sediment substrates. The authors (Katolikova et al, 2016) suggested that fucoids act as shock absorbers, mitigating the impact of surf damaging *MT* which possess thinner and more fragile shells (Beaumont et al, 2008; Michalek et al, 2021). Importantly, in the last case, an inverse relationship is assumed: *MT* avoids surf-exposed areas. However, in some studies, when the material collection design was specifically adapted to study of the surf effects, the role of this factor in the regulation of the *MT:ME* ratio has not been proven (Moreau et al., 2005).

Another approach to explanation of patterns in the spatial distribution of *MT* and *ME* can be related to the history of formation of the contact zone. For example, it is supposed that in the Barents and White Seas, where originally only *ME* is thought to have been present, the *MT* settlements appeared as a result of the introduction of this species from the coast of North America by ships performing transatlantic traffic during the World War II (Vainola, Strelkov, 2011). According to this the populations of *MT* is expected to be observed in harbours (Vainola and Strelkov, 2011; Katolikova et al. 2016), where this species was originally introduced.

We can see a large spectrum of factors (salinity, wave action, proximity to harbor, and substrate type) which were considered as possible agents regulating the distribution of *ME* and *MT* but assessment which factors (when they act simultaneously) are crucial in species distribution may not be obvious. Moreover, some correlations that seem evident may have a different interpretation. For example, the relationship of high abundance of *MT* with reduced salinity can be interpreted by another way: the most desalinated areas (river mouths) are often also the quietest areas with minimal wave impact. Port structures, similarly, are often represented in the most sheltered and frequently desalinated areas. Thus, there is no strict evidence for the exclusive role of some leading factor in the regulation of species distribution and consequently no reasons for use it as singular predictor in SDM construction.

The only way to solve this problem is to include all predictors potentially affecting the composition of mixed settlements in one model and this is the main methodology in species distribution modelling (Bosch et al 2018 ???). Such approach to analysis of mussel’s species distribution have been attempted before (but without direct reference to SDM framework, see Kijewski et al., 2019; Wenne et al., 2020) and allowed to show a clear relationship of the *MT:ME* ratio primarily with salinity and temperature gradients. However, the populations involved in the works cited were scattered over a very wide geographic range, with distances between them reaching hundreds and thousands of kilometers. On the other hand, the gradients of the key factors regulating species distribution can vary on much smaller spatial scales. For example, salinity, as a prime candidate regulating the *MT:ME* ratio, can vary significantly on scales of a few kilometers in river estuaries (Van Damme, 2005). If *MT:ME* ratio is regulated by salinity does this mean that the ratio will show the same pattern at smaller geographic scales as well? In addition, will salinity play a leading role in species distribution amid the inclusion of other predictors in the model?

In this paper, we will consider the case where *ME* and *MT* come into contact forming mixed populations in the Kandalaksha Bay of the White Sea (Vainola, Strelkov, 2011; Katolikova et al. 2016; Khaitov et al, 2021). This contact zone found in the relatively small area is extremely convenient for investigation of environmental factors’ role in the regulation of the *MT:ME* ratio and therefore SDM construction.The salinity gradients are well expressed in this area (Filatov et al., 2005) which is determined by the inflow of several large and small rivers into the sea (Katolikova et al., 2016). Further, the coast of the gulf possesses numerous sheltered areas (inlets and skerries) and areas open to the surf which provide the opportunity to assess the wave impact gradient. There are several active and several abandoned ports (Katolikova et al., 2016) that give an opportunity to trace the role of vessel trafic in species distribution. And finally important is that the small size of the Kandalaksha Bay (Fig. 1) ensures that there are no clear temperature gradients in both warm and cold seasons. Therefore, the influence of temperature, which probably regulates the distribution of *Mytilus* species on a broader geographic scale (Wenne et al. 2020) can be excluded from the discussion.

Using the material collected in this contact zone we will construct model which would describe the association of *MT:ME* ratio with potentially important factors mentioned above. Basing on the model constructed we would assess the niche divergence between *MT* and *ME* coexisting sympatrically. And finally we will check the ability of the model constructed to predict species distribution pattern in other contact zones.

## Material and methods

Study area

The 185 km long Kandalaksha Bay in the north of the White Sea is oriented from southeast to northwest and is funnel-shaped (Fig. 1). Climate is continental subarctic with cold winters (4-5 months of ice cover) and relatively warm summers (August SST is 13.8°C) (Filatov et al., 2005). The bay contains hundreds of banks, skerries and islands. The bay’s shoreline is fringed by numerous smaller inlets. Because of this shoreline geometry, different areas even situated close each to other are exposed to wave action in varying degrees. The summer surface salinity is 24 ppm in most of the Bay (“normal” salinity of the White Sea is 24-30 ppm), but lower in the estuary areas (Filatov et al., 2005). A total of 24 rivers and two canals of hydropower plants (STable ++ ) with catchment area of 141 - 12830 km2 (average 240 km2) discharge fresh water into the bay creating local salinity gradients. The top of the Kandalaksha Bay is the most desalinated, due to the influence of the largest river Niva (Katolikova et al., 2016).

Historically (through most of the 20th century), several ports (Fig. 1 A) were functioning in the area operating high tonnage oceanic vessels (Sailing directions of the White Sea, 1932; Krasavtsev, 2011). Two of these ports, Kandalaksha and Vitino, both in the top of the Bay, are currently functioning while the remaining ports have not been used to serve ocean-going ships in recent decades, but are still visited by coaster ships, according to the conventional knowledge.

Mussels are ubiquitous and abundant along the coasts of the Kandalaksha Bay. On the littoral, they are most abundant at local mussel beds (Khaitov, 2013) and within the fucoid belt formed mainly by *Fucus vesiculosus* and *Ascophyllum nodosum* (Naumov, 2001)

Field collections

When sampling, we tried to describe littoral populations of mussels in the Kandalaksha Bay in as much detail as possible, also taking into account the heterogeneity of their habitats by salinity, wave exposure and proximity to ports. The material was collected on the littoral of the mainland and the islands of the Kandalaksha Bay (Fig. 1 A). Collections were carried out in July-August in 2011-2018 at 95 sites.

At each site we collected mussels from two substrates: fucoid’s thallus (thereafter “Algae”) and bottom surface (i.e. stones, pebbles, sand, mud; thereafter “Bottom”). Bottom samples were taken using a 0.025 m2 square frame placed on a ground in such a way that at least a few dozens of mussels could be sampled.The sample was placed in a plastic bag and transported to the laboratory. The abundance of mussels from Bottom samples was recalculated to standard area of 1 m2 by multiplication of mussels number by 40. To collect Algae samples, we placed a 0.5x0.5 m square frame on areas completely covered with fucoids. Inside the area of the frame we collected an individual fucoid’s bundle large enough to contain at least a few dozens of mussels (however, this condition proved to be unfeasible for several samples from sparse populations). The fucoid’s bundle (with mussels and without removing water) was weighted with electronic balance and placed into a plastic bag for the further processing. The rest of algae from the area inside the frame was taken and weighted by the same manner. Using the data on weights of fucoid’s bundle and algae rest from the frame the abundance of mussels was recalculated to standard area. To note, at 12 sites the information on fucoid’s weights was not obtained and these sites were excluded from those analyses which required data on absolute mussel’s abundance. Three samples (replications) of each type were taken in each site no more than 20 meters apart. Geographic coordinates were determined for each site (i.e., the same coordinates were assigned to each sample within the same site). The dataset described above was considered as main, training, data which will be used for construction of models (see bellow).

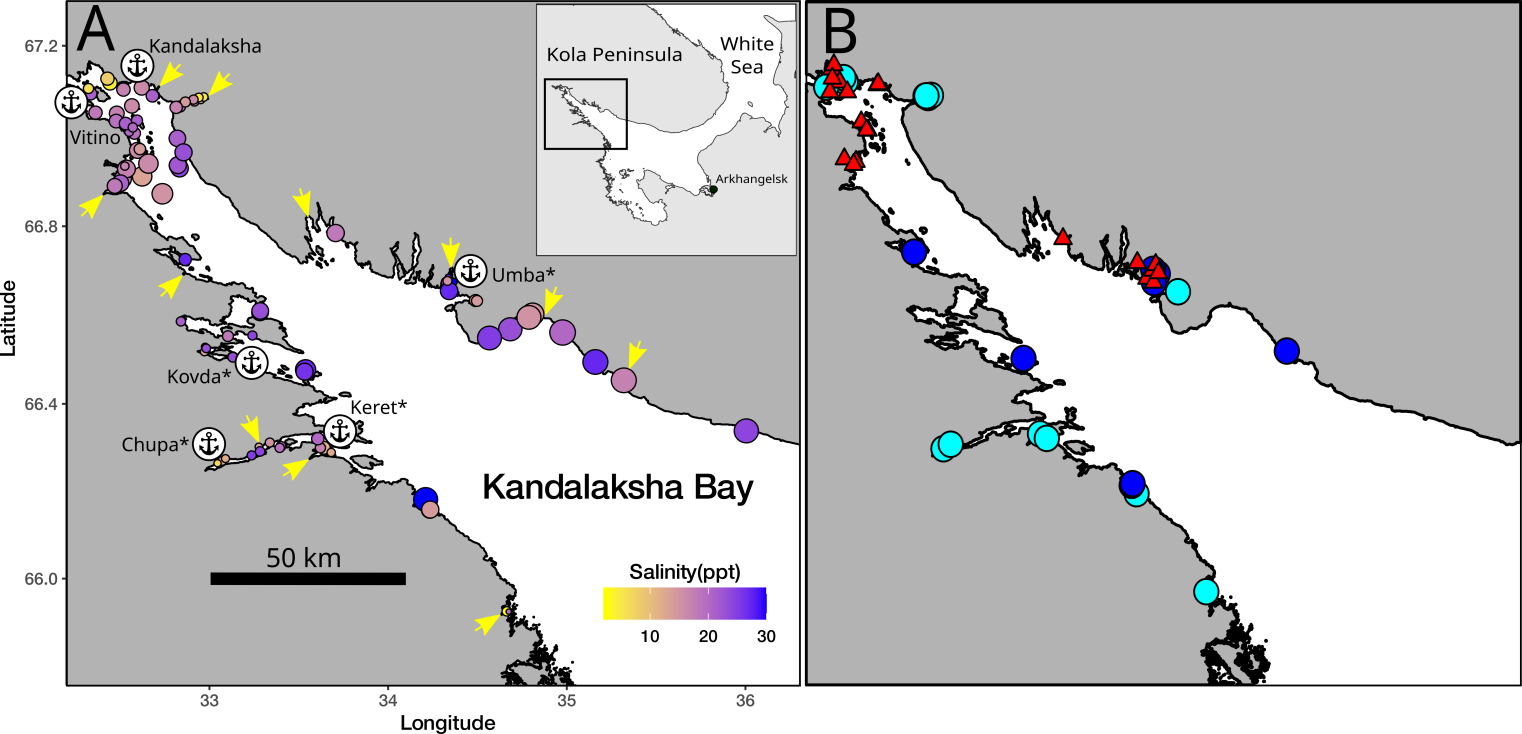


Fig. 1 Sampling sites from the main data set (A) and testing data set (B) collected in the Kandalaksha bay of the White Sea. On the pannel A: the size of points is proportional to average wind fetch, the fill of points represent salinity, arrows indicate the mouths of the largest rivers, signs of anchors indicate harbors name of which is marked by asterics if a port is abandoned. On the pannel B: red triangles represents sites from testing data set, cyan cyrcles - sites from main data set with extreme low salinity, blue - sites with extremly high salinity.

Mussel species identification

As it was shown in previous works (Katolikova et al., 2016; Khaitov et al. 2021) in the case of the White Sea the mussel’s species identification basing on the expensive genetic manipulations can be effectively replaced by a cheep “morphotype test” based on the study of an easy to observe conchological trait (Khaitov et al. 2021). The morphotype test allow identifying mussel species with sufficient accuracy and makes it possible to process large material amount. To apply the morphotype test mussels with shell length more than 10 mm were cleaned from tissues and dry shells were analysed with dissecting microscope. Shell morphotypes (E-morphotype, characteristic to *ME*, and T-morphotype, characteristic to *MT*, see Katolikova et al. 2016 and Khaitov et al. 2021) were identified as in Khaitov et al. 2021. In total we processed 5.5516^{4} specimens. The mean number of mussels collected per sample was 97(range 2, - 785).

Environmental parameters assessment

This work was aimed to construct SDM explaining the mussel’s species proportion in contact zones as function of key environmental parameters such as salinity, surf level, substrate and proximity to ports. Some of these parameters were measured directly but for others some proxies were considered.

The nature of the substrate was characterized by where the corresponding sample was taken from (*Algae* vs *Bottom*). The degree of proximity to port was estimated as the straight line distance (not accounting shoreline bends) from the sampling site to the nearest of the 6 large ports presented in the area (Fig 1 A). Additionally each port was assigned to one of two categories: *“Active”* if at the time of sampling the port was serving heavy-tonnage vessels plying out of the White Sea, or *“Abandoned”* if the port served only small vessels cruising within the White Sea, or was completely abandoned at the time of sampling.

*Wind fetch as a proxy for wave impact*. When selecting sampling sites, we assumed that collections should characterize both surf and sheltered habitats. Each site was visually assigned to one of two categories: *“Exposed”* (sites located on open shores exposed to waves coming from the open part of the bay) or *“Sheltered”* (sites located in bays or skerries). For a more formal assessment, the value of wind fetch was calculated for each site. It was calculated as the unobstructed length of water surface (in km) over which wind from a certain direction can blow over (LaBarre et al., 2023, Marchand, Gill, 2018). These values can be obtained from the analysis of geographical maps of the region. To calculate the wind fetch values, the R-package “windfetch” was used (Seers, 2022). The higher fetch values denote the more wave impact could be brought from a certain direction. Wind fetch was measured for each site from all directions (North, East, South and West, STable +) but as a value for the further analysis we used an averaged fetch value.

*Proxies for salinity assessment*. Water samples were taken at each site when mussels were collected, i.e. in the time of the low tide. The salinity was measured using an “Atago S/Mill-E” handheld salinity refractometer (with 1 ppm accuracy). However, salinity measured directly at the sampling site poorly reflects the salinity regime *per se* in the site due to its fluctuations during the tidal cycle and due to weather conditions (waving, precipitation *etc.*). Therefore, as an alternative estimate of the desalination level in a given site, we used the characteristics of fresh water discharging sources closest to the sampling site (rivers or channels of hydroelectric power plants).

In total, 24 rivers are took place in the region, information about which is present in open sources (STable +). Additionally, two powerful freshwater dischargers (HPP canals) were included in the analysis. For each river, we assessed the catchment basin area and coordinates of its mouth. For HPP canals, this information was obtained by visual analysis of geographical maps. Each of the rivers was assigned to one of two conditional categories: “Large” or “Small” depending on whether its catchment area was more or less than an median catchment area of all rivers included in the analysis (240 km2). To note, this operational classification should not be misleading with classification of rivers size proposed by hydrologists (Khublaryan, 2009).

We calculated the distance (straight line, not accounting shoreline bends) from each sampling site to the mouth of the nearest river. Thus salinity was characterized by three values: directly measured salinity, distance to the nearest river and the nearest river size.

Testing datasets

Additionally we included into analysis two other data arrays which were used as testing datasets. The first one consisted of samples from 23 sites collected in the Kandalaksha bay (STable ++, Fig. 1 B). These sites were sampled using a different protocol than in the case of main dataset (i.e. only Bottom or Algae samples were collected per each site). All predictor’s values for these samples were assessed as for main dataset.

The second testing dataset was obtained from published data on mussel populations in the Tyuva inlet situated in the Barents Sea where another *MT-ME* contact zone is represented (Marchenko et al. 2023, supplementary electronic materials). Since the algae cover was assessed in the sampling area we classified those samples where cover rank was high (rank 3-5 in presented data set) as samples from Algae substrates, but when cover ranks were lower (1-2) the samples were classified as Bottom. Other parameters were assessed as follows. The Tyuva River flowing into the upper part of the inlet having the catchment area of 351 km2 was assessed as a “Large” river (see above). The fetch values for each sites were assessed using the geographic map. The abandoned port constructions (69.190018 N, 33.622952 E) are found in the Tyuva inlet and *MinDistPort* values for each sampling site were calculated as the straight line distances to this port area.

Statistical analysis

The number of T- and E-morphotype mussels in each sample was counted and proportion of T-morphotype mussels (*PT*) was calculated. Additionally this proportion was assessed for each site (*PTSite*, the total number of T-morphotype mussels collected at this site divided by total number specimens collected). Our assumption was that *PT* value in the White Sea is strongly correlated with probability to find *MT* genes in mixed mussels’ populations (i.e. STRUCTURE score) and can be assumed as a good proxy for *MT:ME* ratio as was shown in previous works (Katolikova et al., 2016; Khaitov et al.2021). To translate *PT* value into proportion of *MT* (*Ptros*) in a population the equation from Khaitov et al.(2021) could be used as follow.

All statistical processing was performed using the statistical programming language R 4.05 (R core Team, 2021). The analysis was conducted in several steps.

At the first step, we assumed that the relationship between dependent variable and predictors can be curvelinear that’s why we constructed a generalized additive mixed model (Wood, 2017) with beta-binomial residuals distribution (hereafter *Model 1*). The fixed part of the *Model 1* included predictors as follows:

* *Salinity*: continuous, the directly measured salinity (ppt);
* *Substrate*: categorical, with two levels “Bottom” vs “Algae”;
* *MinDistRiver*: continuous, distance to the nearest river mouth (km);
* *RiverSize*: categorical, with two levels “Small” vs “Large” accordingly to size of the nearest river;
* *MinDistPort*: continuous, the distance to the nearest port (km);
* *PortStatus*: categorical, with two levels “Abandoned” vs “Active” for the neares port;
* *AverageFetch*: continuous, fetch value (km).

For all categorical predictors, the first mentioned category was used as base level in the analyses. The function gam() from the package “mgcv” (Wood 2017) was used to build the model and the restricted maximum likelihood (REML) method for parameters estimation was used. Smoothers for all continuous predictors were fitted using cubic basic splines. Categorical predictors were included as parametric terms in the model. “Site” was considered as random factor in the model.

The *Model 1* was considered as the main construction reflecting association of *PT* with environmental parameters. However, since the data on direct salinity measurement could potentially be inappropriate assessment of salinity regime (see above) we constructed the *Model 2* which was organized as the *Model 1* but *Salinity* was excluded from the predictor’s set.

We did not attempt to simplify both models by excluding predictors that do not have a statistically significant effect on the dependent variable. We were only interested in the extent to which removing salinity from the predictors would change the predictive properties of the model. For both models we calculated Akaike information criterion (AIC) and to check weather the performance of these two models differ substantially. Before AIC calculations both models were refitted using maximum likelihood (ML) method.

To check for the predictors’ collinearaity in the *Model 1* we calculated the variance inflation factor (VIF, Fox & Monette, 1992) for each of them. Additionally we calculated Pearson correlation between continuous predictors. To verify that the *Model 1* and *Model 2* meet the assumptions of sampling independence, we examined the presence of residuals’ spatial autocorrelation by means of spline correlogram construction (Bjornstad, Falck, 2001) with the function spline.correlog() from the package “ncf” (Bjornstad, 2022). No evidences of spatial autocorrelation were revealed for the models.

The second step of the analysis was associated with a more deep analysis of association between *PT* and substrate type. For each site we calculated the difference (*Dif*) between *PT* on Algae (total mussel of T-morphotype sampled from Algae substrate in the given site divided by total number of mussels collected from the Algae in this site) and *PT* on Bottom substrates (calculation was made by the same manner as for Algae). The obtained *Dif* values were regressed against two predictors reflecting two different aspects of mixed populations structure: the proportion of T-morphotype in site (*PTSite*) and the mean mussel’s abundance in the site per 1 m2 (*MeanN*). We used function gam() from the package “mgcv” for the regression model construction. Gaussian residuals’ distribution was assumed for this analysis.

On the third step we performed the assessment of model’s ability to predict the *MT-ME* distribution. We used three data sets to assess the prediction quality: (1) main training dataset, (2) testing dataset from the Kandalaksha bay and (3) testing dataset from the Tyuva bay.

To assess the models’ predictive power all sites included in the analysis were divided into two groups. We set as classification boundary *PT*-value equal to 0.5 which corresponds accordingly to *Eq 1* to *Ptros* = 0.57. All samples with *PT* above this boundary were assigned as *MT*-dominated, the others as *ME*-dominated.

Using the *Model 1* and *Model 2* we calculated predictions of the models for each of three datasets. The predicted values were considered as classifiers for detecting *MT*- or *ME*-dominated groups. The receiver operated characteristics (ROC) followed by analyzing the area under curve (AUC, Fielding, Bell,1997; Fawcett, 2006) was used to evaluate models performance. The function roc() from the package “pROC” (Xavier et al. 2011) was used.

## Results

## Environmental parameters in the Kandalaksha Bay

Two of six ports presented in the area were active at the date of material collection (Fig. 1 A, STable ++ SEM). The distance from sampling sites to the nearest ports ranged from 0.11 to 82 km (Median = 18.7 km) (SFig ++, A). Sites visually categorized as “Sheltered” on average were characterized by lower values of wind fetch than those categorized as “Exposed” (SFig. ++, B). The most exposed sites (maximum values of wind fetch) were located on the western coast of the Kandalaksha Bay, and on the open shores of the islands at the top of the bay (Fig. 1 A). Salinity measured at the time of sampling ranged from 2 to 30 ppm (Median = 19 ppm). The most desalinated areas were the top of the Bay and inlets where rivers flow into (Fig. 1 A). Salinity at sites situated close to “Large” rivers was lower than at sites closer to “Small” ones (SFig. ++, C). Salinity tended to increase with increasing of distance to the mouth of the nearest river (SFig. ++ D) but no association between distance to the nearest port and salinity was found (SFig. +++ E). Sites situated close to ports tended to have lower wind fetch (SFig. +++ F). All correlations between salinity, distance to the nearest river mouth, distance to the nearest port and wind fetch were rather low (STable +++). The largest correlation revealed was between wind fetch and distance to nearest port (r = 0.46).

## *PT* spatial distribution in the Kandalaksha Bay

The maximum proportion of the T-morphotype was observed in the top of the Kandalaksha Bay itself and in the small bays deeping into the mainland (Fig. 2). At the same time, sites with a lower proportion of the T-morphotype were located in the immediate vicinity of sites where settlements with high *PT* values were represented (Fig. 2), reflecting the high variation of population structure in small spatial scale. Settlements with the lowest frequency of T-morphotype were represented in the open areas of the North-West coast.

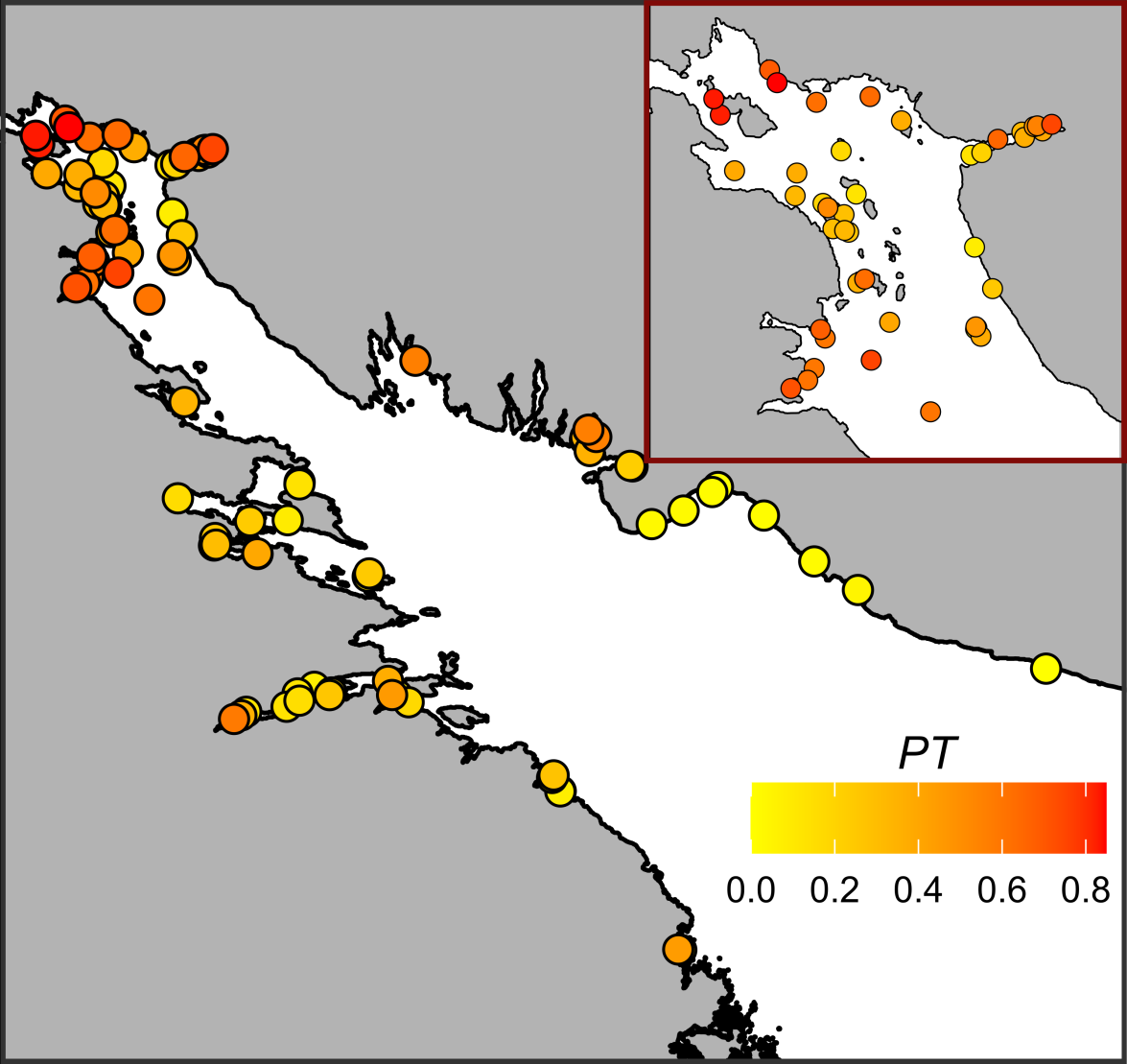


Fig. 2. Proportion of mussels with T-morphotype on sampling sites. The fill of points represents *PT* value. Insertated plot represents the detailed *PT* distribution in the top of Kandalakshs bay.

## Relationship of *PT* to environmental parameters

The values of VIF calculated for the predictors’ set included in the analysis were rather small (the maximal VIF was detected for Min\_dist\_river: 1.58). Although some pairwise correlation between environmental factors were found (see above) the low VIF values can be interpreted as a negligible collinearity between predictors (i.e. they don’t mask influence of each for other).

To assess the dependency of *PT* on all predictors considered we constructed the *Model 1* (Table 1, R2 = 0.65) which explained 85.2% of total deviance. The *Model 1* revealed statistically significant dependency of *PT* on all predictors except one.

The effective degrees of freedom (*edf*, Table 1) for most smoothers were equal to 1, reflecting the linear dependence of *PT* on such continuous predictors as distance to the nearest port (*MinDistPort*) and fetch (*AverageFetch*). The exception was *Salinity*, where a pronounced curvilinear dependence was revealed.

Table 1 . Parameters of smoothers and coefficients of parametric terms for the Model describing dependency of PT on predictors.

| Model term | edf | ref.edf | Chi.sq | p-value |
| --- | --- | --- | --- | --- |
| s(Salinity) | 5.6 | 5.7 | 27.9 | 0.0002 |
| s(MinDistRiver) | 1.0 | 1.0 | 2.4 | 0.1224 |
| s(AverageFetch) | 1.0 | 1.0 | 12.2 | 0.0005 |
| s(MinDistPort) | 1.0 | 1.0 | 5.0 | 0.0249 |
| Random effect s(Site) | 72.0 | 88.0 | 485.0 | 0.0000 |
| Model term | Parameter estimate | SE | z-statistic | p-value |
| (Intercept) | -2.4 | 0.2 | -11.5 | 0.0000 |
| Substrate(Algae) | 1.1 | 0.1 | 13.7 | 0.0000 |
| RiverSize(Large) | 0.5 | 0.2 | 2.2 | 0.0293 |
| PortStatus(Active) | 0.9 | 0.3 | 3.2 | 0.0012 |

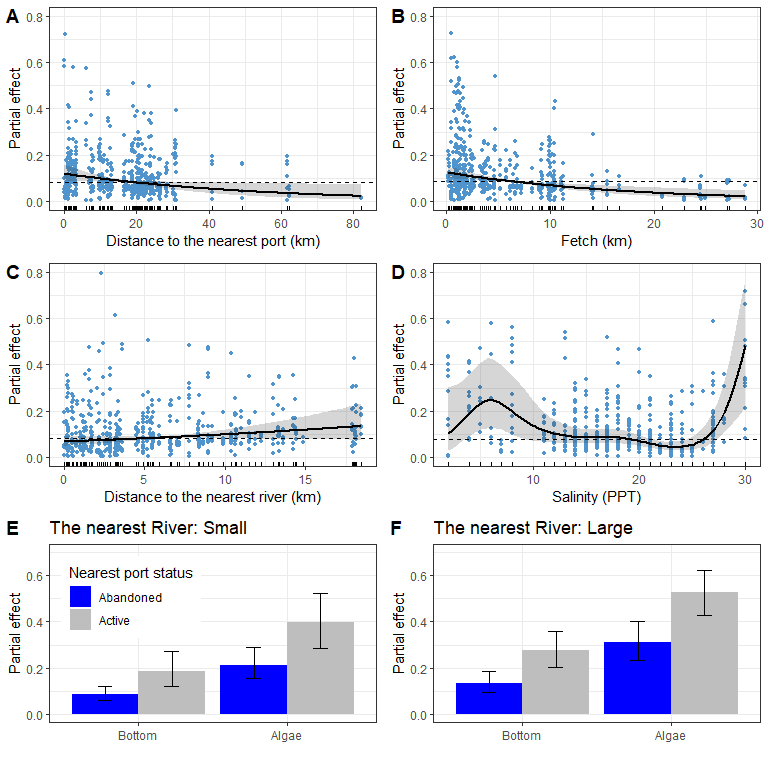


Fig. 3. Predictions of the *Model 1* describing the dependency of *PT* on distance to the nearest port (A), average fetch (B), distance to the nearest river (C), salinity (E) and substrate type, nearest port status and size of the largest river (E, F). The gray ribbons and wiskers demonstrate 95% confidence interval. The dotted lines in panels A-D represent the median predicted value, this value is given to track the trend in the change in model predictions. Points on panels A-D reflect partial residuals but not raw data for each sample.

According to the *Model 1* (Table 1, Fig. 3) *PT* has a tendency to decrease with increasing of distance to the nearest port (Fig. 1 A) starting to be in average lower than median partial effect when distance to the nearest port is higher than 20 km. The nearest port status had also significant influence on *PT*: the value was higher when port status was “Active” (Fig. 3 E, F). The dependency of *PT* on average fetch (Fig. 3, B) was negative: *PT* become lower than median partial effect when value of fetch started to be higher of 5 km. Weak tendency to *PT* increase with increasing the distance to the neares river (*MinDistRiver*) was not significant (Table 2, Fig 3 C).

The curvilinear dependence of *PT* on salinity (Fig. 3, D) can be described as follows. The predicted value of *PT* is close to the median partial effect at intermediate salinity range. However, at extremely low (less than 12 ppt) and high (more than 27 ppt) salinity, *PT* becomes higher than the median predicted value reflecting a sort of anomalies. Extreme low salinity was observed at 15 sites and extreme high salinity at only 8 sites (Fig. 1 B). For other 72 sites, with intermediate salinity, predicted *PT* was close to the median value (Fig. 3 D). All extremely low salinity sites were represented in the top parts of different inlets (Fig. 1 B) but no clear pattern revealed in the distribution of sites with extremely high salinity (Fig. 1 B).

According to the *Model 1* the value of *PT* is dramatically higher on the Algae substrate than on the Bottom one (Table 1, Fig. 3 E, F). This value is higher in those sites which is close to active ports than to abandoned ones. Finally, the *PT* value is slightly higher at sites situated close to large rivers than to small ones.

## Dependency of *PT* on substrate type

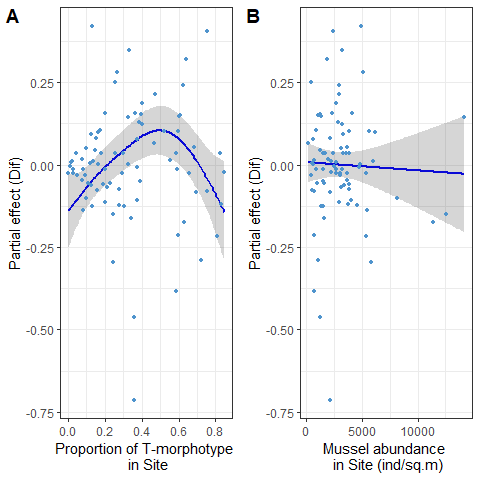


Fig. 4. The dependence of difference between *PT* on Algae and Bottom subatrates on PTSite (A) and total mussel abundance (B).

The difference between *PT* on Algae and Bottom substrates was significantly dependent on proportion of T-morphotype in site (STable +, Fig. 4 A). The dependency revealed was curvelinear (edf = 2.9, STable +) with minimal *Dif* values close to zero in sites with minimal and maximal values of *PTSite*. The maximal *Dif* values were associated with intermediate *PTSite*. This pattern reflect the prevalence of T-morphotypes at Algae substrate and prevalence of E-morphotype at Bottom substrate in those sites where both morphotypes are presented in equal amount. No significant dependency was revealed with total absolute mussel abundance (STable +, Fig. 4 B)

## Assessment of model’s predictive power

We constructed the *Model 2* which was the model of the same type as *Model 1* but *Salinity* was excluded from the predictor’s set (STable ++). The AICs for both models were close to each other (AICModel 1 = -2120.03, AICModel 2 = -2118.47, ΔAIC = 1.57).

Table 2. The values assessing models performance

| Model | AUC training data | AUC testing data Kandalaksha bay | AUC testing data Tyuva inlet |
| --- | --- | --- | --- |
| Model 1 | 0.8691 | 1.0000 | 0.6387 |
| Model 2 | 0.8687 | 0.9911 | 0.7267 |

Both models fitted were close each to other in their ability to recognize *MT*-dominated sites being applied to training and testing datasets (Table 2). In the case of the Tyuva testing data set the *Model 2* had even better performance showing AUC higher than in the case of *Model 1*.

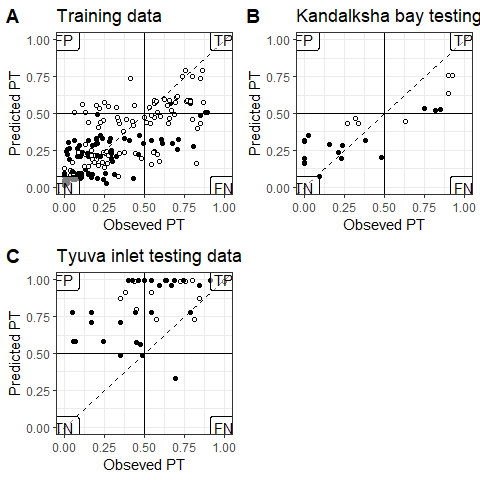


Fig. 5. Comparison of *PT*-values predicted by *Model 2* and obsreved proportion of T-morphotype for Bottom (black circles) and Algae (white circles) samples. The horizontal and vertical solid lines reflects the boundary between *MT*- and *ME*-dominated sites. Dotted lines represent the position of perfect fit when observed and predicted values are equal. Labels in the plot corners mark the quadrants with false positive (FP), true positive (TP), true negative (TN) and false negative (FN) predictions.

Since the *Model 2* is simpler and do not require directly measured salinity (a poor determined environmental parameter for intertidal) we used this variant of model to analyze the prediction’s error rate. We plotted values predicted by the *Mode 2* against observed ones (Fig. 5).The diagram constructed for training dataset reveal relatively good correspondence between predicted and observed *PT*. It is important to note that the most false positive preditcions for main, training, data set were made for Algae samples whereas false negative predictions were made both for Algae and Bottom samples.

In the case of testing dataset from the Kandalaksha bay (Fig. 1 B) the most sites were correctly classified by the *Model 2*. The AUC for this data was close to 1 (Table 2) indicating high predictive ability of the model. The *Model 2* being applied to testing dataset from the Tyuva inlet (although AUC was relatively high, Table 2, and some correlation between predicted and observed *PT* was revealed r=0.398) produced only few true negative predictions but a lot of false positive one.

## Discussion

Using the “morphotype test” (Khaitov et al. 2021) as a tool for species identification, we were able to process large amount of samples from numerous sites covering wide spectrum of habitats. This material allowed to construct the *Model 1* describing the relationship between *PT* and all the factors that are generally considered to be the main regulators of the *MT:ME* ratio in mixed mussel settlements. Our analysis showed that the predictors included in the model weakly correlated with each other and do not exhibit high collinearity. This means that we can consider all factors involved as independent entities that do not mask each other in their influence which is important for SDM construction (De Marco, Nobrega, 2011). We found that all factors, which were considered as crucial in previous works, were significant for *PT* value (and consequently proportion of *MT* in mixed populations) in a relatively small contact zone of the Kandalaksha Bay. This means that the patterns revealed on more extensive contact zones (e.g. in the Baltic - North Sea contact zone, see Kijewsky et al, 2019) with longer ecological gradients retain their influence in essentially smaller spatial scales. This fact allow to optimistically expect the ability to construct working SDM describing distribution of *MT* and *ME* in contact zones of different sizes.

The SDMs are frequently considered as models reflecting structure of species’ ecological niche (Elith, Leathwick, 2009). In this sense the significant association of *MT:ME* ratio with some environmental factors would reflect the divergency of *MT* and *ME* niches. We revealed that the *PT* (i.e. proportion of *MT*) is significantly dependent on a set of environmental factors (the salinity, the nature of the substrate, some factors associated with the proximity to active port systems, and the degree of coast exposure to wave action). To date we can only roughly characterize the mechanisms underlying the divergence of ecological niches, experimental works should be carried out to discover this mechanisms deeply.

From all environmental variables included into analysis the distance to the nearest port and its status (Abandoned VS Active) may have dual interpretation. We revealed higher *MT* proportion close to active ports. On the one hand, active ports tend to have a higher degree of anthropogenic pollution, leading to stress of mussels (Hellou & Law, 2003). Different mussel species are known to interact differently with pollutants (Beyer et al., 2017). For example, in the Halifax area *MT* are more abundant in the most polluted habitats than *ME*, but show less stress tolerance than *ME* (Hellou & Law, 2003). However, studies of the immune systems of *MT* and *ME* in the Canadian contact zone suggest a greater suppression of the *MT* immune system in areas prone to pollution (Coray et al., 2006), making them more sensitive to stress than *ME*. The last fact contradicts the hypothesis of better performance of *MT* to pollutant influence allowing them to set populations in active port harbors. Another explanation may be that there is some foci of disease associated with port areas that affect *MT* and *ME* differently. For example, an increased frequency of bivalvian transmissible neoplasia (BTN) has been shown to be associated with ports (Hammel et al., 2023). Given that this disease takes its origin from *MT* (Yonemitsu et al., 2019) one would suppose this species to be more resistant to it. The second species, *ME*, if less resistant to the disease, is expected to be less abundant near harbors. Thus, if we believe that polutants or disease are the cause of the dependence on port proximity revealed, some deep comparative ecophysiological and immunological studies in different contact zones are needed for future SDM construction.

From the other hand the dependency of *PT* on distance to the nearest port and its status can reflect not true axis in ecological space but some trace of historical processes. It is believed that *MT* was introduced into the Barents and the White Seas due to ship traffic during World War II (Vainola & Strelkov, 2011). This hypothesis was based primarily on the fact that high concentrations of *MT* were found in the vicinity of currently active or abandoned ports in the region (Vainola & Strelkov, 2011; Katolikova et al. 2016). Our study confirmed this hypothesis. The proportion of T-morphotype (and hence *MT*) in the studied area increased with approaching to harbors. Apparently, the degree of port activity also plays an important role. According to our data, if the port nearest to the mixed settlement is active, i.e. it serves the ships coming from outside the White Sea (in our case it is “Kandlaksha” and “Vitino” ports), then the *PT* value in such settlements is higher than in the settlements adjacent to the abandoned harbors. The observed relationship may be related to the fact that the currently active ports were the largest and most economically sustainable systems in the region. This may be based on the higher shipping traffic, which has led to a more massive invasion of *MT* from other regions (primarily from the ports of Murmansk in the Kola Bay, where the abundance of *MT* is very high, Vainola&Strelrov, 2013; Marchenko et al., 2023). In the abandoned ports, the introduction of the alien species occurred during the times of their active operation. However, during several decades of inactivity of abandoned ports *MT* were probably displaced by native *ME*.

In any case, our data are in line with earlier stated hypothesis that in the case of the White Sea *MT* is an invasive species (Vainola&Strelkov 2011). Unfortunately, only few data could be provided to check this hypothesis for other European *MT* -*ME* populations. However the ship traffic was successfully included into SDMs describing invasive species in European waters (Lindegren et al., 2022) this factor was never included into models describing *MT*-*ME* distribution previously. In some cases it is probably an unrealistic aim due to absence of this predictor variability. For example in the case of *MT*-*ME* populations in the Baltic Sea, where there is extremely active ship traffic and vessels visiting almost all parts of the Baltic Sea, it is not possible to identify a correlation between the distribution of *MT* and the presence of ports. At the same time, if we look at the more broad map of *MT* distribution (Vainola & Strelkov, 2011) around European coasts, the confinement of *MT* to large European ports out of the Baltic Sea (e.g. Bergen, Murmansk), rather speaks in favor of the hypothesis of *MT* introduction due to ship traffic. Thus, while proximity to ports is not an environmental factor in a direct sense, it is a parameter that should be kept in mind when building a more general SDM describing *MT*-*ME* distribution.

Our analysis revealed that the higher wind fetch is associated with lower *PT* value. That means *MT* avoid open coasts potentially exposed to the surf. The lowest *PT* value (i.e., E-morphotype mussels dominate in the settlements) is observed on open shores of the north-east coast of Kandalaksha Bay (Fig. 1 A), exposed to wave action due to southeast and south-west winds which are frequent in the White Sea region (Berger, Naumov, 2001). This result corresponds well with the data on the thickness of *MT* and *ME* shells. It is known that *MT* have thinner, more fragile shells (Beaumont et al., 2008; Michalek et al, 2021; our unpublished observations of mussels from the White Sea are in line with this data). It is logical to expect that mollusks with such characteristics should avoid places with high wave loading. In such habitats, *ME* with stronger shells are expected to benefit. A similar pattern can be seen in other contact zones. For example, in Greenland, an increased frequency of *MT* was noted at the sheltered top of the fjord (GLS and GLD sites, Fig. 1 in Wenne et al. 2016), while the population at a more open site located near fjord mouth (GLL site in Wenne et al. 2016) is dominated by *ME*.

Again, experimental evidence is needed to confirm the divergence of the niches of the two species along the surf-impact axis. It is far from certain that a thinner but more flexible shell could be a hindrance to set settlement in areas more exposed to surf. There is a need to assess the level of mortality of the two mussel species due to shell damage along the surf gradient.

The nature of the substrate was discovered as the powerful predictor regulating *PT* variation in the scale of several meters. The proportion of T-morphotype mussels appears to be significantly higher on algal than on bottom substrates (silty-sandy sediment, pebbles, boulders and rock surfaces) located in close proximity (not more than 20 m, within the same fucoid belt). This pattern was first detected in the analysis of a significantly smaller volume of material (Katolikova et al., 2016). Bringing in new data collected in a wider range of conditions confirmed the presence of this pattern.

However, a closer examination of the relationship with substrate type (Fig. 4 A) showed that the maximum *PT* divergence between Algae and Bottom substrates occurred only when both mussel species were present in approximately equal numbers at a given location. This could be interpreted as divergence of realized niches as a result of interspecific competition. This, however, is contradicted by the lack of correlation of *Dif*-value with the absolute abundance of mussels at the site. If the mentioned divergence of realized niches were a consequence of competition, one would expect that the strength of this interaction would increase with increasing settlement density, which is not observed. Thus another explanation should be considered.

As possible explanation we can propose some difference in byssus attachment strength. Perhaps *MT* are capable of attaching to underwater substrates more strongly than *ME* do. This should cause less strongly attached molluscs to break off and remain on the bottom under constant fluctuations of fucoid’s thallus. This hypothesis is contradicted by the fact that a comparison of attachment strength in *MT* and *ME* from the contact zone in Canada, showed no significant differences between the species when they were kept under laboratory conditions (Lowen et al., 2013). Thus some additional investigations of attachment strength of both species are needed.

An intriguing result of our study was the absence of clear linear dependency between taxonomic structure of mixed populations and salinity that was revealed in other contact zones (Kijewsky, 2019). It is generally considered that *MT* is more tolerant to reduced salinity, where it dominates over *ME*. As salinity increases, species change places (Riginos Cunningham 2005). However, surprisingly, in spite of the fact that salinity is considered as a leading factor determining the segregation of *MT* and *ME* in their contact zones majority of studies discussing this relationship are correlative. Only a few works have been devoted to comparison of ecophysiological responses of these two species to salinity (Gardner&Thompson 2001; Qiu et al., 2002; Knobel et al. 2021). Moreover, the direct physiological investigations did not reveal unequivocal evidence that adult *MT* are better adapted to reduced salinity than adult *ME* (Gardner&Thompson 2001; Qiu et al., 2002). It has only been shown that larval stages of *MT* survive lower salinity better than the same *ME* stages (Qiu et al., 2002). That is the prevalence of *MT* in lower salinity could be explained by early selective pressure during passive larval drift across salinity gradient (Knobel et al. 2021).

In our investigation the GAM fitted reveal a statistically significant curvelinear (close to U-shape) relationship. From one hand, our *Model 1* revealed higher *PT* (i.e. proportion of *MT*) in lower salinity which is typical for other contact zones. However, on the other hand, this proportion was remarkable inflated when salinity was unusually high. We can propose three explanation to this fact.

Firstly, although directly measured salinity seems to be a natural ecological factor, this may not be the case for littoral mussel settlements. The salinity levels significantly vary during the tidal cycle and the range of the variation is highly dependent on the distance from the fresh water source (Attrill, 2002; Marchenko et al., 2023). We do not know which salinity is more important for littoral mussels living in the estuaries: high during the ebb time or reduced in the low tide. Further, the salinity measured once during the sampling process may occasionally be anomalously low or high by accident (e.g. due to precipitation or upwelling of deep salted water, respectively). If this explanation is correct, then the observed increase in *PT* at high salinity may be a consequence of some incorrect estimates of salinity values.

A second explanation for the *PT* peak in high salinity sites may be related to the lower reliability of the morphotype-test for species identification under high salinity conditions. Khaitov et al (2021) suggested that in cold Arctic waters with high salinity, *ME* more often exhibit T-morphotype, which reduces the sensitivity of the method. However this is contradicted by the fact that in the same locations where anomalously high salinity and high *PT* values were detected (e.g. in the Umba area, Fig. 1 A, B) genetic analysis of mussel populations revealed an increased frequency of *MT* genes (Katolikova et al.2016).

And finally, it should be noted that sites with high *PT* and extreme low (<12 ppt) or unusual high (>27 ppt) salinity are not numerous. In total, the number of such anomalous sites is about 24.2 % of the total number of sites surveyed. In locations, where salinity is intermediate, *PT* does not differ significantly from the background value. That is, anomalously high *PT* values are observed in some marginal habitats. Such pattern can be expected if we assume that *MT* is a weaker competitor, which on average cannot displace *ME*. This assumption is supported by the fact that the settlements that we classify as *MT*-dominated are essentially scarcer in the studied area (TP + FN part of the Fig. 5 A) than the *ME*-dominated sites (FP + TN part of the Fig. 5 A). The roughly U-shape association between *PT* and salinity can be turned into bell-like dependency of proportion of E-morphotype (i.e. *ME* frequency) on salinity. The *PT* proportion predicted by the *Model 1* for the most frequent salinity values (12 - 27 ppt) is about 0.1. This means that the majority of mussels presented in the sites within this salinity range are mussels with E-morphotype (mostly *ME*). Such displacement of weaker competitor into marginal habitat is well known for other systems where inviders met resident species: mosquito larvae (Juliano, 1998), fishes(Herbold, Moyle, 1986) and geckos (Case et al 1994). Usually resident species persist in their optimal habitats if they have not been destroyed but introduced species prevail in marginal habitats (Juliano, 1998). To note, actively used harbors appear to be the very marginal habitats as well.

All these remarks about salinity make us doubt the reliability of salinity as a good predictor in SDM for littoral mussels. Probably the assessment of salinity influence can be improved when we use more accurate assessment of salinity variation in the site of samples. Undoubtedly salinity is the most important factor regulating the distribution of *MT* and *ME* and this factor be it properly assessed should be taken into account. However the salinity in the form used in our work may not be as high for SDM construction. After we removed salinity from the model, the *Model 2* constructed had AIC very close to AIC of the *Model 1*. The rule of thumb in the model selection paradigm (Burnham, Anderson 2004) suggests that an AIC delta less than 2 indicates high support for the reduced model, that is, the information loss when salinity was dropped from the model is not critical. The comparison of full *Model 1* and reduced *Model 2* by their predictive capacity by ROC analysis did not reveal significant decreasing of predictive power of *Model 2* when it was applied to training dataset.

The strongest argument that the constructed model is useful is its ability to predict phenomena on the dataset that was not used to build the model. We used two testing datasets. One was represented by material collected in the same area as the training samples (Fig. 5 B). The analysis of ROC-curves showed that *Model 2* has a high ability to predict the assignment of the sample to one of the two categories *MT*-dominated vs *ME*-dominated. When this model was used for predictions in a different region, the Tyuva inlet, where an independent contact zone is presented, the quality of the prediction decreased. However, two points should be emphasized. Firstly, the AUC= 0.7267 for *Model* 2 was quite high and close to a reasonable predictive performance AUC level (Elith et al. 2006). Secondly, it is important to note that the predictive power of the *Model 1* which includes salinity was lower, further emphasizing the low power of this parameter as a predictor of *MT* and *ME* distribution.

The important result of the *Model* 2 applying was the high number of false positive predictions in the case of the Tyuva inlet. The prediction errors may have numerous reasons (see Fielding & Bell, 1997). Firstly it can be consequence of a poor model tuning. In fact, applying the *Model 2* we made some assumptions which may not be entirely realistic. For example we considered samples with high algae coverage as samples from fucoid’s tallus. Additionally the small area of the inlet did not provide sufficient variation in some parameters which is based on distances (fetch measure, distance to the nearest port or river mouth). Although there may be some ecological reasons as well. If a suitable habitat is not colonized by a target species which population is small (i.e. the area is not saturated) or the population is collapsed due to displacement as a result of inter-specific competition, all this reasons could produce false positive result when a model-classifier is applied (Fielding & Bell, 1997). Indeed the *MT* population in the Tyuva inlet displayed pronounced decreasing associated with increased *ME* population (Marchenko et al 2023). We may expect the Tyuva inlet conditions are highly suitable for *MT* population but once appeared it demonstrated decrease probably due to competitive displacement by *ME*. We expect that alternative scenario may take place in the case of the Kandalaksha Bay. Let’s remember that *MT*-dominated sites were quite rare in the area and that most of false positive prediction were made for Algae substrates (which is more suitable for *MT*). It can be interpreted as evidence of underdevelopment of *MT* population in the area: suitable habitats (primarily fucoids) are not occupied still. It can be hypothesized that this is the result of recent *MT* population formation in the Kandalaksha Bay and some outbreak of *MT* invasion may happen in the area.

Whether our model can predict the distribution of *MT* and *ME* in other contact zones is an unsolved propblem still and requires additional data. At the same time, it can be noted that in the case of the famous sudden appearance of *MT* in a region previously dominated by *ME* - the Loch Etive farms (Beaumont et al. 2008), our model seems to be quite applicable at least at the conceptual level. Indeed, the aquaculture farms where unexpected invasion of *MT* was described (56° 27′ 30″ N, 5° 19′ 12″ W, Beaumont et al. 2008) are located in a narrow fjord (low fetch) with high salinity variability associated with the influx of the full-flowing Etive river (salinity ranges from 1 to 22 ppt, Beumont et al, 2008), and an active port in Oban located about 14 km along the coastline. All this correspondences indicate that this area could be predicted by our model as the “red zone” where *MT* may establish their populations with high probability.

## Conclusion

In our analysis, we did not aim to construct some perfect SDM it is the task for further investigation when information on different contact zones would be combined. That is why we did not try to build some final model (for example, by backward selection procedure) and did not try to find the most appropriate modelling method (see Elith et al 2006). It was important for us to understand whether in principle a working model can be built based on those predictors that have come to the attention of researchers studying contact zones of cryptic mussel species. Our results are optimistic about this possibility but since the contact zones between *MT* and *ME* are scattered over a very wide range of latitudes and situated in various biotic surroundings the appropriate form of the “general” model should be well considered. It is obvious that some additional predictors (e. g. temperature or its correlates, the position in the intertidal-subtidal gradient, *etc*) should be included in a future model. It is also clear that the “general” model should include some biotic interactions (see Wisz et al 2013) and be developed within the JSDM framework (see Ovaskainen & Abrego, 2021 for a theory). The dependent part of this model should include not only proportions of two coexisting mussel species but a matrix describing the community where mussel’s species coexist. Our study has shown that algae play an important role in the local distribution of species. Perhaps the spatial variation in algae abundance (and probably algae species composition) can play a significant role in the distribution of the cryptic mussel species. Thus algae characteristics in region should be somehow incorporated in the model. Perhaps some role may also be played by predators, which can regulate the proportion of mussel species (Khaitov et al.2018; Khaitov et al.2023 ). However the first step to construct the working SDM is the deep comparison of ecological features of different *MT-ME* contact zones.

## Literature cited

Лоция Белого моря / Гидрографическое управление ; общ. ред. А. Н. Рождественского. – 5-е изд. – Ленинград : Издание Гидрографического управления, 1932. – LXXII, 566 с., 22 вкл. л. ил., карт., схем.

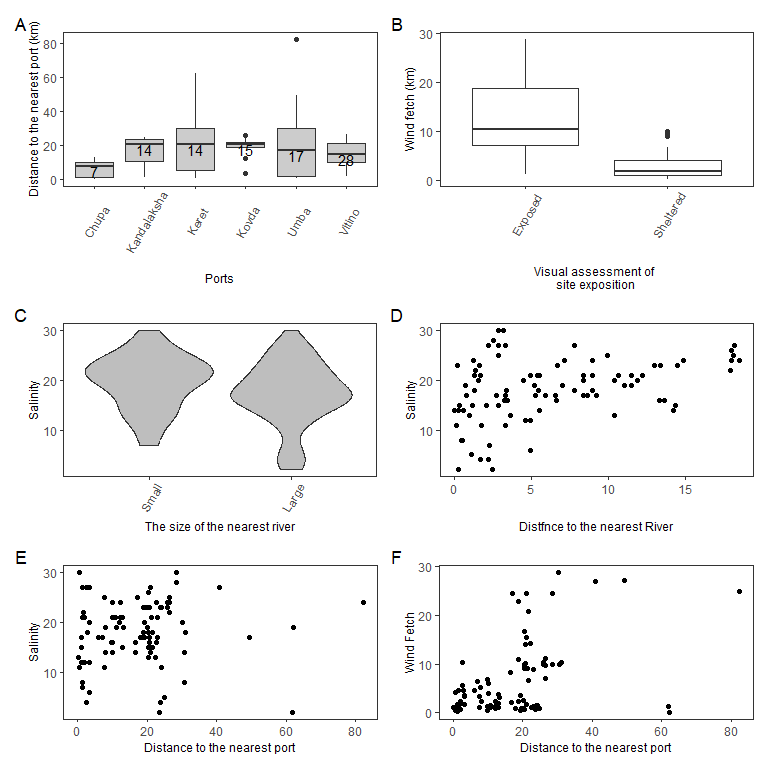
# Supplementary electronic materials (SEM)

Pearson correlation between environmental parameters

| Predictor | Min\_dist\_river | Min\_dist\_port | Average\_Fetch |
| --- | --- | --- | --- |
| Salinity | 0.368 | 0.049 | 0.261 |
| Min\_dist\_river |  | 0.153 | 0.279 |
| Min\_dist\_port |  |  | 0.462 |

Table +. Parameters of river and other fresh-water discharging source associated with Kandalaksha Bay.

| River | Drainage Area sq. km | River Size | Mouth Latitude | Mauth Longitude | Number of sampling sites near the river |
| --- | --- | --- | --- | --- | --- |
| Chernaya | 141.0 | Small | 66.51896 | 32.94966 | 6 |
| Gridina | 540.0 | Large | 65.92462 | 34.65846 | 2 |
| Keret | 3,360.0 | Large | 66.28172 | 33.58034 | 4 |
| Knyazhegubskaya HPP channel |  | Large | 66.87178 | 32.40742 | 0 |
| Kolvitsa | 1,260.0 | Large | 67.08320 | 32.99110 | 14 |
| Kovda | 26,100.0 | Large | 66.69549 | 32.87103 | 1 |
| Kuzreka | 255.0 | Large | 66.59999 | 34.81303 | 4 |
| Kyatka | 204.0 | Small | 65.88838 | 34.68294 | 0 |
| Lupche-Savino | 146.0 | Small | 67.16291 | 32.34259 | 0 |
| Luvenga | 175.0 | Small | 67.10037 | 32.69313 | 11 |
| Nilma | 167.0 | Small | 66.50577 | 33.13507 | 8 |
| Niva | 12,800.0 | Large | 67.13027 | 32.41825 | 9 |
| Nivskaya HPP channel |  | Large | 67.15676 | 32.37713 | 0 |
| Olenitsa | 403.0 | Large | 66.45276 | 35.30060 | 2 |
| Pila | 313.0 | Large | 66.78056 | 34.15745 | 0 |
| Plavezhma |  | Small | 66.25417 | 33.00065 | 3 |
| Porya | 240.0 | Large | 66.81954 | 33.55116 | 1 |
| Pulonga | 630.0 | Large | 66.30251 | 33.27729 | 6 |
| Ryasanka | 102.0 | Small | 66.73710 | 33.44663 | 0 |
| Salnitsa | 121.0 | Small | 66.37576 | 35.68487 | 1 |
| Shoguy | 103.0 | Small | 66.74761 | 34.02558 | 0 |
| Shushpanikha | 65.2 | Small | 66.85175 | 33.60705 | 0 |
| Sonreka | 167.0 | Small | 66.15655 | 34.22415 | 4 |
| Umba | 6,250.0 | Large | 66.66461 | 34.29766 | 6 |
| Varsuga | 9,840.0 | Large | 66.26487 | 36.94030 | 0 |
| Vypcha | 161.0 | Small | 66.63304 | 34.49202 | 3 |



SFigure +. Associations between environmental factors. (A) Boxplots showing the variation in distances to the nearest port. Numbers insite the box indicate the number of sites for which the corresponding port is the nearest. (B) Boxplots showing the variation of wind fetch for two groups of sites (Exposed vs Sheltered) choosen for material collection. (C) The kernel density violin-plots showing salinity variation for sites near Small and Large rivers. (D) The association of salinity with distance to the mouth of the nearest river. (E) The association of salinity with distance to the nearest port. (F) The association between wind fetch and distance to the nearest port

## Parameters of Model 2

STable + . Parameters of smoothers and coefficients of parametric terms for the Model 2 describing dependency of PT on predictors (Salinity was removed).

| Model term | edf | ref.edf | Chi.sq | p-value |
| --- | --- | --- | --- | --- |
| s(MinDistRiver) | 1.0 | 1 | 0.0 | 0.9316 |
| s(AverageFetch) | 1.0 | 1 | 6.6 | 0.0105 |
| s(MinDistPort) | 3.9 | 4 | 19.3 | 0.0010 |
| Random effect s(Site) | 76.0 | 89 | 597.9 | 0.0000 |
| Model term | Parameter estimate | SE | z-statistic | p-value |
| (Intercept) | -2.4 | 0.2 | -11.5 | 0.000 |
| Substrate(Algae) | 1.1 | 0.1 | 13.7 | 0.000 |
| RiverSize(Large) | 0.3 | 0.3 | 1.1 | 0.279 |
| PortStatus(Active) | 1.3 | 0.3 | 4.7 | 0.000 |

## Parameters of Model describing association with substrate type

STable + . Parameters of smoothers in the model describing dependency of Difference of PT on Algae and Bottom substrates on proporion of T-morphotype and total mussel abundance in sites.

| Model term | edf | ref.edf | Chi.sq | p-value |
| --- | --- | --- | --- | --- |
| s(PT) | 2.9 | 3.6 | 4.3 | 0.0043 |
| s(Total abundance) | 1.0 | 1.0 | 0.1 | 0.7528 |