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

Species distribution models (SDM) being a numerical tools describing the relationship between species occurrence and environmental parameters allow to predict distribution patterns of species both in space and time (Elith, Leathwick, 2009). The SDM framework is very popular in terrestrial ecology (+++) but the amount of papers devoted to SDMs for marine species has also been increasing in recent years (Robinson et al 2017). The range of approaches to building such models is very wide: from regular multiple regressions up to advanced machine learning methods (+++).

SDMs are mostly applied for «good», morphologically distinct species (e.g. Reiss et al., 2011; Lindegren et al., 2022) that can be involved in routine studies which in the case of marine ecology require numerous samples. The increasing evidence for cryptic species (Bastrop et al 1998; Fiser et al 2010) and polytipic species (++++REF) among marine animals makes it urgent to find out approaches to build SDMs for sympatric cryptic species, as well as to infraspecific taxa such as subspecies or phylogenetic lineages within polytypic species. Only few studies have been conducted in this direction so far in marine ecology (Lowen et al 2019; Dennis, Hellberg, 2010). Since the relationship between species and environmental factors is the essence of the ecological niche concept (+++) SDMs are considered as a tool of species’ ecologial niche assessment (Elith, Leathwick, 2009). Such interpretation is particularly useful in the context of cryptic species coexisting in sympatry since their niches’ divergence is frequently in focus of investigation (++++). Strictly speaking, in the case of several coexisting species, not SDM but JSDM (Joint Species Distribution Modelling) framework should be considered (Ovaskainen, Abreg, 2020). The JSDM paradigm implies the modelling of both ecological niche of coexisting species and their interactions as well (Ovaskainen, Abreg, 2020).

In the marine realm, the longest scientifically known and best studied complex of cryptic species is the blue mussel “*Mytilus edulis*” complex (Knowlton 1993, ???), which includes a number of species that are better distinguished genetically than morphologically and hybridize in sympatry (++++++). Blue mussels are powerful ecosystem engineers in the temperate and subpolar seas playing an important role in coastal communities (+++). They are also important objects of aquaculture (+++). In the North Atlantic sector, the dominant species are *M. edulis* (thereafter *ME*) and *M. trossulus* (*MT*), which form zones of sympatry (thereafter contact zones) in different parts of the region from Scotland and the Gulf of Man in the south to Greenland and Svalbard in the north (Wenne et al. 2020 and references therein). In contact zones, *ME*, *MT* and their hybrids are often found in the same samples from mussel settlements; such settlements are hereafter referred to as “mixed”. Scientists generally agree that *ME* and *MT* are ecologically distinct in sympatry (++++ RC, Katolikova et al.), and have different economic values in aquaculture (+++), but the data on the factors of their ecological segregation i.e. ecological niche partitioning of species is fragmentary and frequently contradictory.

On a biogeographic scale, the distribution of *Mytilus* species is thought to be regulated primarily by temperature and its correlates (Hayhurst, Rawson, 2009; Wenne et al., 2020). Both species occur both in temperate in parts of Atlantic, but *MT* being more stenothermic, cold-loving species does not penetrate as far south to the temperate/boreal seas as ME (Wenne et al., 2020).

The greatest progress in comparative ecological studies of *ME* and *MT* in sympatry has been made in the contact zones in the Baltic Sea, in the waters around of the Kola Peninsula (White and Barents Seas) and in the West Atlantic. In the Baltic Sea, only *MT* inhabits the brackish areas of the inner Sea parts, but the more saline areas closer to the North Sea inhabited by *ME* only. In the contact zone hybrids usually predominate in mixed settlements and the frequency of *MT* genes increases towards the inner Baltic (VS 11, Stuckas et al. 2017). As a result, species distribution is strongly correlated with salinity (++++), which acts as a selective factor affecting primarily larval stages (Knöbel et al., 2021). However such simple and understandable pattern in divergence of *MT* and *ME* ecological niches can hardly be revealed in other contact zones. For example, in the Kola region and in West Atlantic zones hybridization is limited with less than 20% of hybrids in mixed settlements (Wenne et al. 2020). Species are distributed in space in a mosaic fashion, both in regional and local scales (RC, +++++ ). The relationship between distribution of species and salinity is nowhere obvious (RC 05, Katolikova et al. 2016; Marchenko et al. 2023), but a number of other factors of ecological segregation have been proposed.

In the White and Barents Seas, the frequency of *MT* is elevated in port areas, possibly due to introduction of this species into the region with ship traffic (VS11++++). The only one factor of species segregation explicitly tested in the White Sea was the substrate to which littoral mussels attach (Katolikova et al. 2016). It turned out that *MT* is more common on fucoid algae while *ME* mostly lives directly on the bottom (mud, sand, stones, gravel). The non-random distribution of species across substrates, however, cannot explain the entire local-scale mosaic in their distribution (Katolikova et al. 2016). Parallel studies in the Barents Sea did not consider mussel substrates, but verified the hypotheses of non-random distribution of species with depth. It turned out that on vertical transects, the proportion of *ME* increases with depth. As a result, *MT* appears to be a more littoral species and *ME* a more sublittoral one (Marchenko et al. 2023). In the western Atlantic, depth, level of anthropogenic pollution and surf exposure has been considered as possible factors in species segregation. It was hypothesized that *MT* may be more abundant than *ME* in littoral (++++), polluted (++++), and surf-exposed habitats (+++), but none of these hypotheses have been tested directly on sufficient material.

Summarizing the above, in the non-Baltic *MT*-*ME* contact zones no clear pattern of species distribution along ecological gradients was revealed and some factors involved into analysis were potentially collinear masking each other. For example, ports are often located in storm-protected areas, frequently in estuaries. This makes it difficult to isolate the effects of shipping and other anthropogenic factors, surf and salinity. The same could be said about the effects of depth and fouling substrate since the littoral fucoids are rare in the sublittoral.

This state of knowledge is not surprising given that for most of the time of blue mussel research, scientists have used labor-intensive genotyping methods to identify cryptic species and therefore have been unable to precede large amounts of material needed for properly organized ecological studies. In addition, there were no reliable statistical methods available to model the distribution of sympatric taxa in the space of multiple factors, i.e. no SDM/JSDM approach was realized. To our knowledge, in the history of *ME* and *MT* studies the SDM approach (but without direct reference to this) have been applied only twice, by Kijewski et al.(+++) and by Wenne et al. (+++). Both times the machine learning techniques were used to model the macro-geographic distribution of species (technically, of allele frequencies at taxonomically informative genes) in the space of multiple climatic and oceanographic characteristics available from public databases. The conclusions of these studies are summarized by recognizing temperature and salinity as important factors influencing the geographical distribution with *MT* tolerate lower salinities and temperatures than *ME* (Kijewski et al. Wenne et al., see also above).

Finding a simple semi-diagnostic shell trait for *ME* and *MT* - the presence or absence of a continuous prismatic strip under the ligament on the inner side of the shell (Zolotarev +++, Katolikova et al. 2016), allowed us to reliably interpret the taxonomic structure of mixed settlements without genotyping. In the White Sea 74% of *MT*, but only 4% of *ME* have this trait (Katolikova et al., 2016), and using the regression models the frequencies of “striped” morphotypes in samples can be recalculated into proportion of *ME* and *MT* (thereafter *ME:MT* ratio; see equations in Khaitov et al., 2021). Note that hybrids are not considered as separate category under this approach.

In this study, we used the SDM approach to simultaneously assess the effects of substrate, salinity, proximity to harbors, and surf on the *ME:MT* ratio in the Kandalaksha Bay of the White Sea. All these factors have been proposed in previous studies as determinants of species’ ecological niche partitioning in sympatry. The last of the previously discussed factors, depth, was not studied by us, but was controlled by sampling at the same littoral level. To assess the predictive value of the obtained model, we also analyzed independent testing data sets from the same Kola contact zone.