

KLAIPĒDA UNIVERSITY
COASTAL RESEARCH AND PLANNING INSTITUTE

ANDRIUS ŠIAULYS

**EMPIRICAL MODELLING OF MACROZOOBENTHOS
SPECIES DISTRIBUTION AND BENTHIC HABITAT
QUALITY ASSESSMENT**

Doctoral dissertation

Biomedical sciences, ecology and environmental sciences (03B)

Klaipėda, 2013

Dissertation prepared 2007–2013 at Klaipėda University Coastal
Research and Planning Institute

Supervisor:

prof. habil. dr. Sergej Olenin (Coastal Research and Planning Institute,
Klaipėda University; Biomedical Sciences, Ecology and
Environmental Sciences – 03B)

Adviser:

prof. dr. Darius Daunys (Coastal Research and Planning Institute,
Klaipėda University; Biomedical Sciences, Ecology and
Environmental Sciences – 03B)

TABLE OF CONTENTS

1	INTRODUCTION	6
1.1	Relevance of the thesis	6
1.2	Objectives and main tasks of the study.....	7
1.3	Novelty of the study	8
1.4	Scientific and applied significance of the results	8
1.5	Defensive statements	9
1.6	Scientific approval.....	10
1.7	Thesis structure.....	11
1.8	Acknowledgements	12
1.9	Abbreviations.....	13
2	LITERATURE REVIEW	14
2.1	From Petersen grabs to empirical modelling	14
2.2	Assessing the quality of benthic habitats.....	16
2.3	From anthropocentric to biocentric point of view: valuation of the marine benthic habitats	18
2.4	Application of species distribution models in benthic ecology	21
3	MATERIAL AND METHODS.....	25
3.1	Case study sites, data and sampling.....	25
3.1.1	Lithuanian marine area (LMA).....	25
3.1.2	Lake Drūkšiai	28
3.2	Empirical models and modelling procedure	31
3.2.1	Species distribution models	31

3.2.2	Predictive performance, model variation and effects of data traits.....	33
3.3	Assessment of seabed quality for feeding of fish	34
3.3.1	Scheme of the assessment.....	34
3.3.2	Occurrence and importance of prey items	35
3.3.3	Development of seabed quality maps for the feeding of fish	36
3.3.4	Accuracy of seabed quality maps	37
3.4	Biological valorisation approach	38
3.4.1	Assigning biological value to benthic communities	38
3.4.2	Assignment of biological values for research sites.	42
4	RESULTS	44
4.1	Comparative analysis of macrozoobenthos species distribution modelling	44
4.1.1	Application of occurrence models	44
4.1.2	Application of biomass distribution models	45
4.1.3	Effects of data traits	48
4.2	Assessment of benthic habitat quality in terms of fish food provision	53
4.2.1	Diet composition of cod, flounder and eelpout... ..	53
4.2.2	Importance of environmental predictors.....	54
4.2.3	Validation of models	56
4.2.4	Seabed quality for feeding of fish.....	57
4.2.5	Accuracy of fish feeding ground maps.....	63
4.3	Biological valorisation of the seabed in LMA.....	64

4.3.1	Biological values of benthic communities.....	64
4.3.2	Biological valorisation model of the Lithuanian economic zone	68
4.4	Predictive mapping of benthic invasive species <i>Dreissena polymorpha</i> in the lake Drūkšiai based on video survey data	70
5	DISCUSSION.....	76
5.1	The use of species distribution models in benthic ecology and their limitations	76
5.2	Valuating the quality of benthic habitats: the added value of predictive modelling.....	78
5.3	Future perspectives	84
	CONCLUSIONS.....	86
	REFERENCES.....	87

1 INTRODUCTION

1.1 Relevance of the thesis

Marine spatial planning and ecosystem-based management involves methods of assessing the quality of habitats to deliver the high-level objective of achieving healthy ecosystems (Tillin et al., 2008; Borja et al., 2012). “Habitat” in this context is defined as a distinctive seabed area with relatively homogeneous physico-chemical conditions and matching biological features (*sensu* Hiscock and Tyler-Walters, 2003; Olenin and Ducrottoy, 2006). Historically, the quality of benthic habitats was valued in connection with production of fish food (Petersen, 1914, 1918). Such fishery based approach dominated in marine benthic studies until 1970s. Later, several seminal papers appeared where the benthic species distribution was used to assess the environmental quality of marine habitats (Leppäkoski, 1975; Pearson and Rosenberg, 1978; Järvekülg, 1979).

Valuation of seabed for fishery production became an integral part of the ecosystem goods and services assessment in marine environment (Duarte, 2000). Ecosystem services, the processes whereby ecosystems render benefits to people, are the principal means for communicating ecological change in terms of human benefits (Chan and Ruckelshaus, 2010). Whereas concepts of ecosystem goods and services in marine environments are rapidly developing (Kremen and Ostfeld, 2005; Rönnbäck et al., 2007) the quantitative approaches or assessments are rare. Furthermore, many of them focus on evaluation of socio-economic aspects only (Troy and Wilson, 2006; Sanchirico and Mumby, 2009), not reflecting the quality of benthic habitats in terms of, e.g. a fish feeding ground service.

On another hand, the assessment of the goods and services provided by marine ecosystems almost always attempts to attach a monetary value to the biodiversity in an area, or the value of an area in terms of importance for human use (Derosus et al., 2007), i.e. it is assessing the quality of habitats from the anthropocentric point of view. The concept, developed by S. Derosus et al. (2007) takes another

approach: it considers the “biological value” defined as the intrinsic value of marine biodiversity without reference to anthropogenic use. In other words, it focuses on the features of species and communities themselves and not on the contamination or the extractable/usable part of the ecosystem: such biocentric approach is still very rare in assessments of the habitat quality (Węśławski et al., 2009).

Both anthropocentric and biocentric approaches in habitat quality assessments rely on spatial information on environmental characteristics. However, such assessments are usually based on the point sampling data, especially for the distribution of marine biota. Sampling sites rarely are dense and evenly distributed within study area to use simple interpolation techniques for the creation of spatial maps (Li and Heap, 2008). Empirical, or species distribution models (SDMs) relate the occurrence or abundance of organisms with the environment factors that limit their distribution. Such models can estimate the potential habitat of a species using environmental data as predictors. SDMs are gaining increasing attention in aquatic ecology (Robinson et al., 2011) in many applications, from global predictions of the seafloor biomass (Wei et al., 2010) to species distributions at regional scale (Gogina and Zettler, 2010; Vincenzi et al., 2011), including the projection of future biological invasions.

In this study empirical modelling is being used for the assessment of the seabed quality in terms of fish food provision, for biocentric biological valuation of benthic habitats in Lithuanian marine area (LMA) and for the assessment of the spread of an invasive benthic species based on video survey data.

1.2 Objectives and main tasks of the study

The aim of this study was to assess the quality of benthic habitats by applying the empirical modelling of the macrozoobenthos species distribution.

The following tasks were raised for this work:

- 1) to perform the comparative analysis of different species distribution models and select the most fit method for the prediction of macrozoobenthos distribution;
- 2) to create a system integrating the empirical modelling of macrozoobenthos and data on fish diet for the assessment of benthic habitat quality in terms of the fish feeding ground provision;
- 3) to assess the habitat quality in terms of food provision for commercial fish species in the Lithuanian marine area;
- 4) to implement the biological valorisation approach for the benthic habitats in the Lithuanian marine area;
- 5) to assess the range of benthic habitat quality alteration caused by spread of the invasive bivalve *Dreissena polymorpha* based on video survey data and the empirical modelling.

1.3 Novelty of the study

This study provides the first results of empirical modelling of occurrence and biomass distribution of the most common 23 macrozoobenthos species or higher taxa in LMA. Some of the species were modelled in the Baltic Sea region for the first time. An original procedure is proposed for the quality assessment of fish feeding grounds based on the modelling of fish prey items and benthic habitat quality maps for feeding of Baltic cod, European flounder and viviparous eelpout were produced. A biocentric biological valorisation of benthic communities in LMA was performed and the most valuable habitats outlined. The spread of the invasive bivalve *Dreissena polymorpha* in the lake Drūkšiai was innovatively modelled based on the coverage data acquired from the remote underwater video surveys.

1.4 Scientific and applied significance of the results

The results of this study increase understanding of the spatial distribution of macrozoobenthos in Lithuanian Baltic Sea. Comparison of several empirical modelling methods can prove very useful for the selection of modelling techniques for future studies. The proposed

procedure to assess the quality of fish feeding grounds can provide useful information for ecosystem-based management in LMA and could be adjusted for other purposes, such as feeding grounds for marine birds or sensitivity assessment of species and biotopes. Biological valorisation of benthic habitats reveals the biocentric quality of benthic habitats in LMA and is also important for planning activities in marine areas.

Empirical modelling of zebra mussel *Dreissena polymorpha* increased the knowledge about the extent of the invasion in the lake Drūkšiai. Estimations of the total biomass of *D. polymorpha* can be a basis for quantifying its functional role in the ecosystem. The prediction of benthic species coverage based on remote underwater video surveys is an effective method for the assessment of species distribution at a large scale in comparison of grab sampling and SCUBA diver surveys. This could be applied in marine studies, i.e. estimating the distribution of blue mussel *Mytilus trossulus* in offshore areas.

1.5 Defensive statements

- 1) Out of tested empirical modelling methods, Random forests is the most fit method to model the occurrence and biomass distribution of macrozoobenthos in LMA.
- 2) Empirical modelling of macrozoobenthos and the data on fish diet composition can be successfully integrated for the benthic habitat quality assessment in terms of food provision for fishes.
- 3) The most valuable habitats based on biological valorisation are coastal stony bottoms situated in the northern part of LEZ.
- 4) Empirical modelling based on video survey data can be applied for the assessment of the habitat alteration range caused by invasive bivalve *Dreissena polymorpha*.

1.6 Scientific approval

Results of this study were presented in 7 international and 9 Baltic Sea regional conferences and seminars:

2nd scientific-practical conference „Marine and coastal researches – 2009“, Palanga, Lithuania, April 2008;

3rd regional student conference “Biodiversity and Functioning of Aquatic Ecosystems in the Baltic Sea Region”, Juodkrantė, Lithuania, October 2008;

Seminar at Faculty of Natural Sciences and Mathematics, Klaipėda University, Klaipėda, Lithuania, October 2008.

3rd scientific-practical conference „Marine and coastal researches – 2009“, Nida, Lithuania, April 2009;

8th international symposium “GeoHab 2009”, Trondheim, Norway, May 2009;

Regional workshop “Systematic planning tools for the conservation of biodiversity in the European seas”, Strömstad, Sweden, September 2009;

4th regional student conference “Biodiversity and Functioning of Aquatic Ecosystems in the Baltic Sea Region”, Dubingiai, Lithuania, October 2009;

International scientific conference “Cultural dialogue and personality”, Klaipėda, Lithuania, November 2009;

4th scientific-practical conference „Marine and coastal researches – 2010“, Palanga, Lithuania, April 2010;

5th international student conference “Biodiversity and Functioning of Aquatic Ecosystems in the Baltic Sea Region”, Palanga, Lithuania, October 2010;

Seminar on the results from the project PREHAB, Klaipėda, Lithuania, January 2011;

Regional workshop “Ecological, economical and institutional challenges for spatial planning in the Baltic – a multidisciplinary

introductory course on ecological mapping and economic valuation of coastal areas”, Åland, Finland, February 2011;

5th scientific-practical conference „Marine and coastal researches – 2011“, Palanga, Lithuania, April 2011;

10th international symposium “GeoHab 2011”, Helsinki, Finland, May 2011;

6th international student conference “Aquatic environmental research”, Palanga, Lithuania, October 2012.

7th scientific-practical conference „Marine and coastal researches – 2013“, Klaipėda, Lithuania, April 2013.

The material of this study was presented in 4 original publications, published in peer-reviewed scientific journals:

Šiaulys, A., Daunys, D., Bučas, M., Bacevičius, E., 2012. Mapping an ecosystem service: a quantitative approach to derive fish feeding grounds. *Oceanologia*, 54 (3), 491-505.

Šiaulys, A., Bučas, M., 2012. Species distribution modelling of benthic invertebrates in the south-eastern Baltic Sea. *Baltica*, 25 (2), 163-170.

Bučas, M., Bergström, U., Downie, A.L., Sundblad, G., Gullström, M., Numers, M., **Šiaulys, A.**, Lindegarth, M., 2013. Empirical modelling of benthic species distribution, abundance, and diversity in the Baltic Sea: evaluating the scope for predictive mapping using different modelling approaches. *ICES Journal of Marine Science* [*in press*].

Zaiko, A., Šaškov, A., **Šiaulys, A.**, Paškauskas, R., 2013. Bioinvasion impact assessment of an aquatic ecosystem engineer: a combinative method approach. *Biological Invasions* [*submitted*].

1.7 Thesis structure

The dissertation includes seven chapters: introduction, literature review, material and methods, results, discussion, conclusions and references. In the appendices, predictive maps of macrozoobenthos

occurrence and biomass distribution in LMA based on the empirical modelling is provided. The material is presented in 123 pages, 28 figures and 19 tables. The dissertation refers to 123 literature sources. Dissertation is written in English with an extended summary in Lithuanian language.

1.8 Acknowledgements

First of all I would like to thank my supervisor Sergej Olenin and research adviser Darius Daunys for introducing me to the science of benthic marine ecology and guiding me from early student days to the final straight of the PhD studies. I also thank my friend and colleague Martynas Bučas for my involvement into species distribution modelling and all the help only the best friend could offer. Many thanks for Anastasija Zaiko, Aleksej Šaškov, Egidijus Bacevičius, Viačeslav Jurkin, Kelvin D. Fowler, Jūratė Lesutienė, Diana Vaičiūtė, Loreta Kelpšaitė, Erikas Visakavičius, Ingrida Bagdanavičiūtė, Aleksas Narščius for providing data, assisting in the field work, GIS and statistical analyses, scientific insights. I greatly appreciate Marine Research Department for providing monitoring data.

I also would like to deeply thank Artūras Baziukas-Razinkovas, Nerijus Blažauskas and Georg Umgieser for the review of the manuscript, Dan Minchin and Chingiz Nigmatullin for constructive comments.

A special thanks for my parents for giving me a possibility to reach for a high education and all the love throughout good times and hard ones. The greatest indebtedness to my wife Ieva and firstborn daughter Ugnė for many evenings, weekends and holydays without a husband and father.

This study was partly supported by BONUS PREHAB (Spatial prediction of benthic habitats in the Baltic Sea: incorporating anthropogenic pressures and economic evaluation) and Norwegian Financial Mechanism project EEE (A system for the sustainable

management of Lithuanian marine resources using novel surveillance, modeling tools and an ecosystem approach).

1.9 Abbreviations

Abbreviation	Explanation
AUC	Area under the receiver operating characteristic curve
BVM	Biological valuation map
CV	Coefficient of variation
GAM	Generalized additive models
GES	Good Environmental Status
LMA	Lithuanian marine area
MAD	Mean absolute deviation
MARS	Multivariate adaptive regression splines
MaxEnt	Maximum entropy
MSFD	Marine Strategy Framework Directive
NRMSE	Root mean square error normalized by range
RF	Random forests
RMSE	Root mean squared error
SDM	Species distribution model

2 LITERATURE REVIEW

2.1 From Petersen grabs to empirical modelling

It is already a hundred years since fishery biologist and hydrologist Carl Georg Johannes Petersen introduced a device for a quantitative sampling of macrozoobenthos, simply called Petersen grab. He carried out first studies on quantitative description of typical benthic communities since 1914 (Gray and Elliott, 2009). Eventually, grab sampling became a standard for macrofauna studies up until now, nevertheless a rapid development of sciences in the last century. The idea was to quantitatively assess the characteristics of feeding grounds of benthophagous fish species, however Petersen's studies led to first quantitative descriptions of benthic communities. Petersen classified benthic communities by the characterizing species, which was not seasonal and which, because of its numerical or biomass dominance, could be regarded as typical of a given assemblage (Gray and Elliott, 2009). He also explained that the distribution of these communities was driven by two main factors: substrate type and depth. It is no surprise, that these two environmental parameters are the main or even default predictors in today's modelling of macrozoobenthos (Gogina and Zettler, 2010; Reiss et al., 2011). Other Danish scientist Gunnar A. W. Thorson continued Petersen's work scheme and applied it outside southern Baltic and the North Sea thus triggering the occurrence of similar studies worldwide (Gray and Elliott, 2009).

Based on grab sampling data the first distribution maps of the macrozoobenthos were produced for the entire Baltic Sea (Ekman, 1953; Segerstråle, 1957; Zenkevich, 1963). However, arguably the first attempt to quantitatively analyse spatial distribution of macrozoobenthos in relation to abiotic factors was performed by the Estonian hydrobiologist and zoologist Arvi Järvekülg (1979). He used a matrix of depth/salinity and depth/temperature to plot data on biomass of benthic invertebrates important for fish feeding in the Eastern part of the Baltic Sea (Figure 1 and Figure 2).

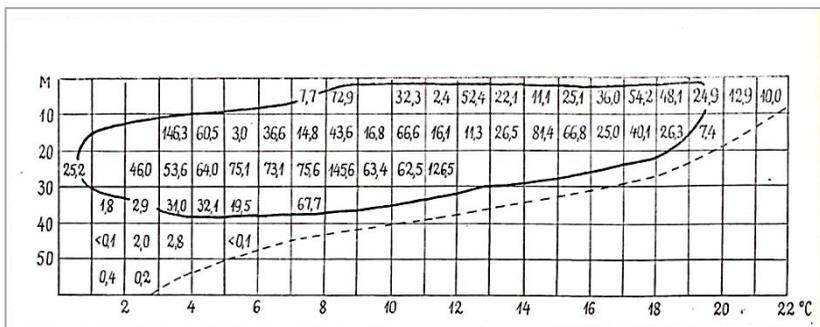


Figure 1. Summer biomass distribution (g m^{-2}) of a bivalve *Macoma balthica* in the Gulf of Riga and Vänameri depending on the depth and temperature (after Järvekülg, 1979).

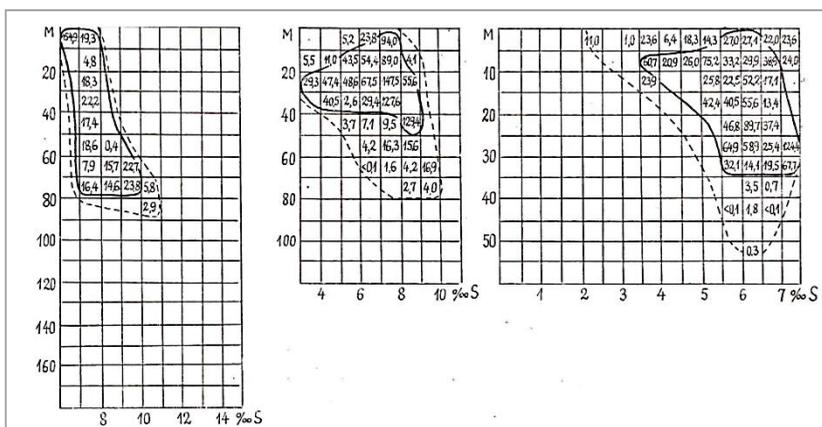


Figure 2. Biomass distribution (g m^{-2}) of a bivalve *Macoma balthica* in Baltic Proper (left), Gulf of Finland (middle), Gulf of Riga and Vänameri (right) depending on the depth and salinity (after Järvekülg, 1979).

Years later, the development of scientific equipment and integration of statistical methods in benthic ecology shifted marine research from descriptive studies to analytical research relating the distribution of species to a number of physical, chemical, geological and biological parameters using regression and other statistical

methods. Eventually, this led to tens of empirical models, which are currently called empirical models or species distribution models SDMs (Section 2.4).

2.2 Assessing the quality of benthic habitats

As mentioned above (Section 2.1), the first attempts to assess the quality of benthic habitats were based on the food they provide for different fish species. With the development of the concept of “ecosystem goods and services”, the provision of food resources started to be considered a part of the goods and services provided by ecosystems, including the marine ones (Figure 3).

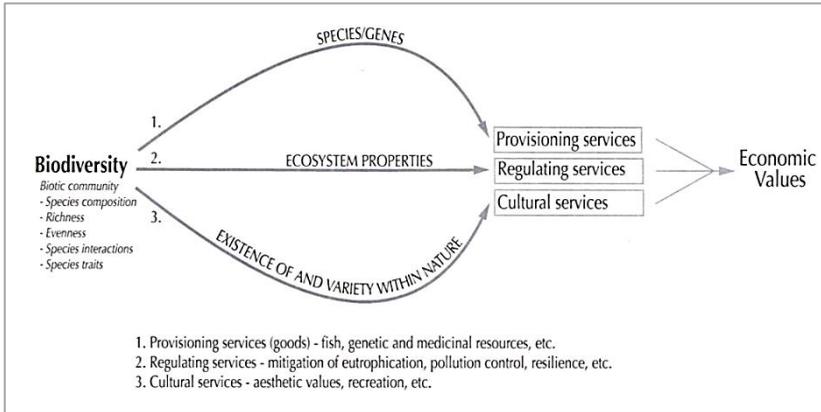


Figure 3. Linking biodiversity to the provision of ecosystem goods and services (after Rönnbäck et al., 2007).

Ecosystem services are defined as “the processes whereby ecosystems render benefits to people” (Chan and Ruckelshaus, 2010) or “components of nature, directly enjoyed, consumed, or used to yield human well-being” (Boyd and Banzhaf, 2007). Due to the anthropocentric perspective of the concept, it is only natural, that ecosystem goods and services provided by benthic habitats are assessed in terms of economic values. For example, Rönnbäck et al.

(2007) performed the monetary valuation of the Swedish coastal habitats in terms of market value of the most important commercial and recreational fish species with documented coastal habitat association.

Another aspect in quality assessment is the valuation of the environmental status of benthic habitats. In the 1970s, several papers were published on the Baltic Sea macrozoobenthos, where the benthic species distribution was used to assess the environmental quality of marine habitats. E. Leppäkoski (1975) proposed the scheme to assess the degree of pollution on the basis of macrozoobenthos in marine and brackish-water environments. In another paper, a model of the macrozoobenthos succession within the organic pollution gradient was proposed (Pearson and Rosenberg, 1978). L. Zmudzinski (1978) invented a notion of the “macrobenthic deserts” occupying the deep water, oxygen deficient areas of the Sea. He was the first to predict that if the antropogenic eutrophication will continue, the benthic deserts may occupy the entire sub-halocline area. In many cases his prognosis has been confirmed (Olenin, 1997).

Currently, the assessment of benthic habitat quality became a keystone concept in the Good Environmental Status (GES) Descriptor 6 “Sea floor integrity” of the Marine Strategy Framework Directive (EC, 2008). It is important that sea floor integrity can be affected not only by eutrophication, benthic dredging, and other stressors. It was indicated that parameters by which this GES descriptor is being evaluated, also can be changed by invasive habitat engineering species (Olenin et al., 2010). One of such habitat engineering effects was thoroughly studied by A. Zaiko (2009) in the Curonian lagoon.

Assessment of GES is the basis of the ecosystem-based approach stated in MSFD: “By applying an ecosystem-based approach to the management of human activities while enabling a sustainable use of marine goods and services, priority should be given to achieving or maintaining good environmental status in the Community’s marine environment, to continuing its protection and preservation, and to preventing subsequent deterioration” (EC, 2008). However, even such sustainable approach to the management of

marine environment implies environmental quality assessments which are, eventually, anthropocentric.

In this respect, it is important to mention another concept in habitat quality assessments, which emerged in the recent decade and aims to establish the intrinsic quality of marine biodiversity (Derous et al., 2007 and references therein), see Section 2.3.

2.3 From anthropocentric to biocentric point of view: valuation of the marine benthic habitats

Search in Google scholar service (scholar.google.com) on “marine biocentric value” exposes 1880 hits, almost one third references (548 hits) are published in 2009 and later. This shows the increasing attention to the intrinsic, biocentric values of marine environment.

As described by Derous et al. (2007), what is meant by ‘value’ is directly linked to the objectives behind the process of valuation (e.g. conservation, sustainable use, preservation of biodiversity, etc.). Discussions on the value of marine biodiversity almost always refer to the socio-economic value of biodiversity (i.e. the so-called value of the goods and services provided by marine ecosystems, or the value of an area in terms of importance for human use), and attempts to attach a monetary value to the biodiversity in an area.

The biocentric value could be defined as “the intrinsic value of marine biodiversity, without reference to anthropogenic use” (Derous et al., 2007). It empathises that the value must be based on the properties of species from the ecosystem point of view, not regarding the provision of goods and services for human uses.

As illustrated by Derous et al. (2007) (Figure 4), the intrinsic valuation of marine habitats should include criteria such as rarity (degree to which an area is characterized by unique, rare or distinct features for which no alternatives exist), aggregation (degree to which an area is a site where most individuals of a species are aggregated for some part of the year or a site which most individuals use for some important function in their life history or a site where some structural

property or ecological process occurs with exceptionally high density), fitness consequences (degree to which an area is a site where the activity(ies) undertaken make a vital contribution to the fitness of the population or species present), naturalness (the degree to which an area is pristine and characterized by native species) and proportional importance (importance in global, regional and national scale).

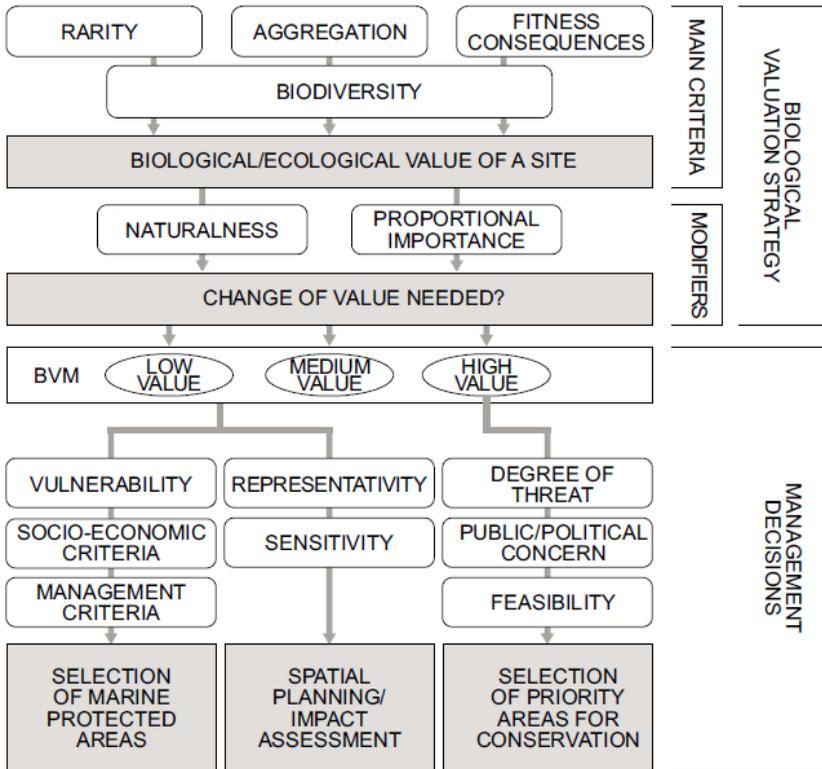


Figure 4. Conceptual scheme of the biological valuation method and possible future steps in developing decision support tools for managers (after Derous et. al., 2007).

The conceptual approach was recently applied in several regional studies in Europe using modern statistical approaches. The delineation of the biologically most valuable areas of the seabed in the Polish Exclusive Economic Zone was performed by Węślawski et al. (2009). It was concluded, that the highest values correspond to habitat-forming algae *Fucus vesiculosus* (Figure 5).

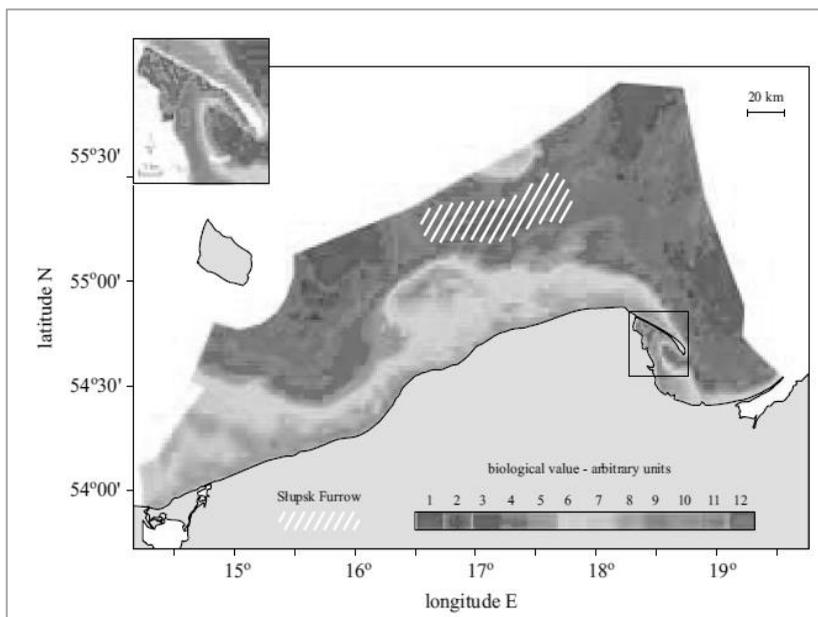


Figure 5. The biological valuation of the Polish Marine Areas. Shaded area indicates inadequate sampling for biological valuation (after Węślawski et al., 2009).

Another implementation of biocentric marine biological valuation mapping was done in the Basque continental shelf (Bay of Biscay) by Pascual et al. (2011). The biological value of habitats was determined by integrating the distribution of zooplankton, macrophytobenthos, macrozoobenthos, fishes and marine birds (Figure 6).

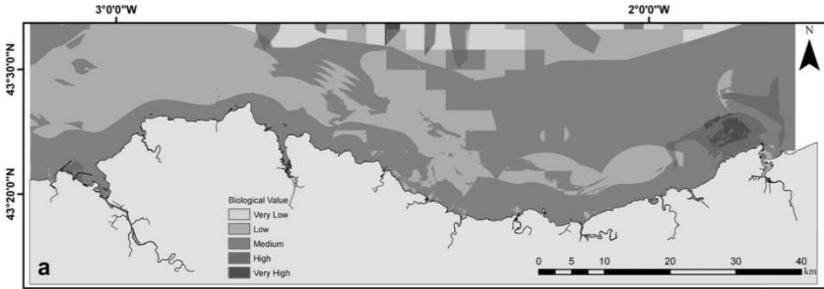


Figure 6. *Integrated Biological Valuation Map (BVM) of the Basque continental shelf and estuaries (after Pascual et al., 2011).*

Biological valuation assessment and the maps it provides are a good offset for other more common valuations of the seabed. As discussed by Węśławski et al. (2009), these maps do not reflect ecosystem health (the most valuable areas could be most degraded and *vice versa*), nor it corresponds to socio-economic valuation in terms of goods provided (fishery, gas, extractible sands, recreational areas). Finally, marine biological valuation is not a strategy for protecting all habitats and marine communities that have some ecological significance, but is a tool for calling attention to subzones that have particularly high ecological or biological significance and to facilitate provision of a greater-than-usual degree of risk aversion in spatial planning activities in these subzones (Derosus et al., 2007).

2.4 Application of species distribution models in benthic ecology

Since the first attempts to quantitatively define distribution of marine macrozoobenthos based on abiotic data (e.g. Järvekülg, 1979, see Section 2.1), the development of new remote sensing technique and integration of statistical methods in benthic ecology shifted this discipline from descriptive studies to analytical investigation. The distribution of species is related to physical, chemical, geological and biological parameters using different statistical methods (Warwick and

Clarke, 1991). Eventually, this led to tens of empirical models, which are currently called empirical or species distribution models.

Increasing availability of large-scale environmental data, better understanding of the relationship between environmental variables and species distribution, and the need for prognostic tools to predict changes in species distribution in response to environmental or climatic changes stimulated recent boom in application of species distribution models in the marine environment (Degraer et al., 2008; Glockzin et al., 2009; Gogina et al., 2010; Reiss et al., 2011; Robinson et al., 2011).

Logistic regression is a frequently used regression method for modelling species distributions and generalized linear models (GLMs – McCullagh and Nelder, 1989) and have been recognized in ecology for some time as having great advantages for dealing with data with different error structures particularly presence/absence data that is the common type of data available for spatial modelling of species distributions (Austin, 2007). Later, GLM was extended into generalised additive models (GAM – Hastie and Tibshirani, 1990) that are increasingly used for species modelling (Leathwick et al., 2006; Austin, 2007).

With the introduction of another empirical modelling method – multivariate adaptive regression splines (MARS – Friedman, 1991), GAM and MARS predictive performance were tested in several studies (e.g. Leathwick et al., 2006; Marmion et al., 2009). No significant differences in the results were detected, however it was concluded that MARS has computational advantages because of better handling interactions between the predictors (Austin, 2007).

Another group of models that are getting an increasing attention in ecology and predictive modelling are tree-based machine learners, such as Bagging (Breiman, 1996), Random forests (Breiman, 2001) and Boosted regression trees (Friedman et al., 2000). They are all ensemble techniques used to stabilize predictions from multiple individual regression trees. While retaining the ability of the underlying regression-tree base learners to deal with non-linearities, non-smooth functions, and interaction effects, the ensemble

techniques increase the precision of predictions by combining several trees in one ensemble model (Knudby et al., 2010). Several studies concludes, that the improved predictive performance can be achieved by machine-learning algorithms such as the tree-based ensemble techniques, compared to the less flexible linear models (e.g. Knudby et al., 2010).

For the presence only data models maximum entropy modelling (MaxEnt – Phillips et al., 2006, Phillips and Dudik, 2008) is very promising. In many studies, this method produced the best predictive performance in comparison to other techniques (e.g. Elith et al., 2006; Reiss et al, 2011). Since MaxEnt is a program with a user friendly interface and emphasis on the ecology studies (Elith et al., 2011) it is getting quite common among researchers.

Given the need for ecosystem-based fisheries management, most recent studies using SDMs in marine ecosystems have focused on the distribution of commercial fish species (Venables and Dichmont, 2004; Maxwell et al., 2009; Lenoir et al.; 2012). In contrast, few studies have applied SDMs to the distribution of marine invertebrate benthos. Ysebaert et al. (2002) used logistic regression to model benthic species distribution in the Westerschelde estuary in the Netherlands, Thrush et al. (2003) and Ellis et al. (2006) modelled the distribution of macrozoobenthos species in New Zealand estuaries with a similar approach. In the North Sea, most studies applying SDMs were carried out on a local scale, such as polychaete distribution in the German Bight using multivariate adaptive regression splines (Meißner et al., 2008), as well as community type and polychaete distribution in Belgium waters using discriminant function analysis, and artificial neural networks plus generalized linear models (GLMs), respectively (Degraer et al., 2008, Willems et al., 2008). In the Baltic, Gogina et al. (2010) and Gogina and Zettler (2010) modelled distribution patterns of macrozoobenthos species using GLMs, Downie et al. (2013) predicted the distribution of the seagrass *Zostera marina* using GAM and MaxEnt, Sundblad et al. (2013) modelled nursery habitat availability of predatory coastal fish using GAM, RF and MaxEnt.

There are plenty of different empirical modelling methods with their advantages and disadvantages. Since it is difficult to compare them due to different datasets and data traits, it is often recommended to use several modelling techniques (Araujo and New, 2007; Grenouillet et al., 2011).

3 MATERIAL AND METHODS

3.1 Case study sites, data and sampling

Case studies were carried out in two ecosystems: Lithuanian waters of the Baltic Sea and the largest lake in Lithuania – the lake Drūkšiai.

3.1.1 Lithuanian marine area (LMA)

3.1.1.1 General overview of environmental conditions

The Lithuanian marine area is located in the south-eastern part of the Baltic Sea. The following description of its physical-geographical conditions is based on several reviews: Žaromskis (1996), Olenin (1997), Gelumauskaitė et al. (1999), Dippner et al. (2008), Olenin et al. (2012) and references therein.

The LMA occupies ca. 6426.6 km², the maximum depth is 125 m while the average depth is approximately 50 m.

The near-bottom salinity above the halocline ranges from 6 PSU in the plume area of the Curonian lagoon and 8 PSU in open waters. The centre of the halocline is found at 74 m with mean boundaries of 64-90 m, where salinity increases from 7.7 to 10.4 PSU. The salinity in the active sub-halocline water layer (90-130 m) can reach over 11 PSU.

The coast of LMA is very exposed to any western winds with the fetch distance over 2000 km. This results in active hydrodynamic environment in coastal area, thus no oxygen depletion is present. However, a strong vertical saline stratification results the lack of efficient oxygenation in the sub-halocline area. The oxygen concentration drops significantly from 6-9.5 ml l⁻¹ (saturation 70-100 %) to 2 ml l⁻¹ (< 20 % saturation) going down from the upper water layer to the halocline range. Hypoxic conditions with salinity around 1 ml l⁻¹ are present in the active sub-halocline layer.

Vertical distribution of temperature is season dependent. The water is cold and homogeneous from December through March due to the intensive convection. In the summer, the thermocline is formed at

20-30 m depths, which separates warm upper water layer and relatively cold deep water. The temperature gradient between near-bottom waters in coastal and offshore areas can reach 12-15 °C.

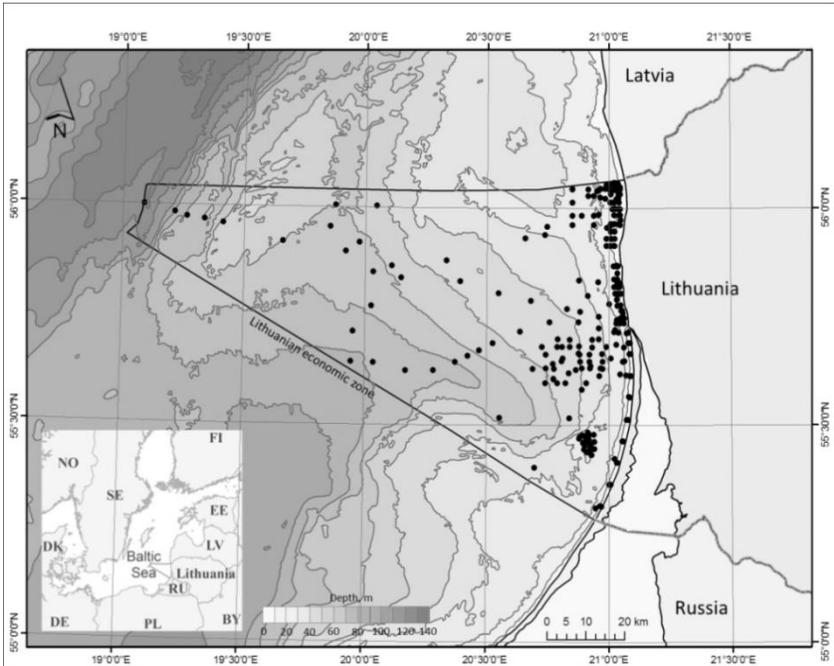


Figure 7. Distribution of the sampling sites in Lithuanian marine area in 1998-2010 (bathymetry acquired after L. Ž. Gelumbauskaitė, 2009).

Coastal slope, extending from the shore down to 25-30 m, is characterised by most diverse bottom types. The uppermost part (0-6 m), is covered by a thin layer of quartz sand, movable during storms. The morainic bench (pebble-gravel deposits with large boulders) lies beneath the sand strip, extending down to 25-30 m depth. Patches of pebble/gravel deposits occur on sites down to 60 m, but in general, this type of bottom is common only for the coastal slope. Soft bottoms change from the mixture of sand and gravel in the coastal area

affected by waves to aleurites and pelitic muds in deeper areas. In general, coarse and medium sands occur from the shore down to 20-40 m depth, and fine sands – down to 50 m. Patches of coarse aleurites are found already at 20-40 m, they extend down to 70-90 m depth. Fine aleurite and aleurite-pelitic mud encircle the slopes of the Eastern Gotland Basin at depths between 80 and 100 m, and finally, below 90-100 m the main type of bottom sediment is pelitic mud, which covers the slopes and floor of the Basin.

3.1.1.2 Environmental predictors used in models

Of the available environmental predictors known to be important for the distribution of benthic invertebrates (Olenin, 1997; Bučas et al., 2009; Gogina and Zettler, 2010; Reiss et al., 2011), eight were used for the models of species occurrence and biomass distribution: sediment types, Secchi depth, minimum near-bottom oxygen concentration, near-bottom current velocity, wave generated orbital near-bottom velocity, slope and roughness of the seabed, areas of above and below the thermocline. Quantitative environmental parameters were tested for collinearity and predictors were removed from models if the variance inflation factors were > 3 (Quinn and Keough, 2002). The layers of sediments, slope and roughness were derived from geological and bathymetrical charts (Repečka et al., 1997; Gelumbauskaitė et al., 1999; Bitinas et al., 2004). Sediments were classified into four types: boulders, cobbles/gravel, sand and silt (Wentworth, 1922). The wind wave orbital velocity data layer was derived using SWAN model (Booij et al., 1999) based on 2008-2009 wind data (Kelpšaitė, unpublished). National marine monitoring data was used to derive Secchi depth and thermocline layers (MRC, unpublished 1998-2006). The mean annual minimum near-bottom oxygen concentration (2000-2006) and bottom current velocity layers were derived from datasets produced by BALANCE project (Hansen et al., 2007; Bendtsen et al., 2007).

3.1.1.3 Sampling data

In total, data from 640 benthic samples taken at 224 sampling sites during 1998-2010 were used to model the biomass distribution of the macrozoobenthos (Figure 7). Soft-bottom samples were taken with a Van-Veen grab, while hard bottoms were sampled by SCUBA divers with a 0.2 x 0.2 m frame. Shallow areas (< 6 m) along the Curonian spit were sampled by SCUBA divers with a plastic tube with the diameter of 10 cm. Samples were taken and treated following standard guidelines for bottom macrofauna sampling (HELCOM, 1988).

Data on the feeding habits of Baltic cod (*Gadus morhua* Linnaeus, 1758), flounder (*Platichthys flesus* Linnaeus, 1758) and viviparous eelpout (*Zoarces viviparus* Linnaeus, 1758) of different body length were collected and analysed by E. Bacevičius (Šiaulytė et al., 2012).

3.1.2 Lake Drūkšiai

Lake Drūkšiai is located in the north-eastern part of Lithuania on the borderline with Belarus (Figure 8). Lake Drūkšiai belongs to the Dauguva river catchment area and outflows to the Baltic Sea via the 550 km long river continuum. The total area of the lake is nowadays about 49 km² (6.7 km² in Belarus, 42.3 km² in Lithuania). The greatest depth of the lake is 33.3 m, and the average is 7.6 m (Jurgelevičienė et al., 1983; Mažeika et al., 2006).

The lake has well developed littoral zone extended by eight islands. It constitutes up to 30% of the total lake area. The natural bottom sediments used to be dominated by sand and gravel (in the upper littoral), at depths more than 7 m – by saprogenic mud (Lithuanian State Scientific..., 1998). The functioning of the natural hydrochemical-sedimentation system of Lake Drūkšiai has been largely effected by the Ignalina Nuclear Power Plant (NPP) and the infrastructure of Visaginas town which was settled in 1975 for more than 4 000 NPP workers and their families. After two decades of the INPP operation, the muddy sediments constituted more than 65% of the bottom and this area tended to expand further (Lithuanian State Scientific..., 1998).

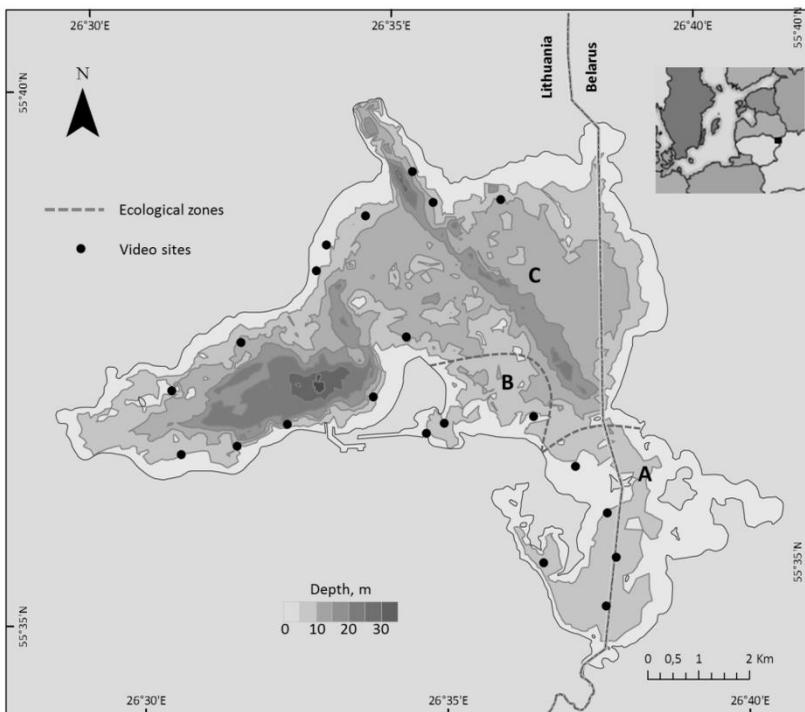


Figure 8. Distribution of underwater video sites in the lake Drūkšiai.

Due to the complex (thermal and chemical) anthropogenic impact the following ecological zones have developed in Lake Drūkšiai (Mažeika et al., 2006; Nedveckaitė et al., 2011):

- Zone A: The most eutrophicated south-eastern part of the lake, where the main source of eutrophication was the household effluents of the INPP and Visaginas town with an elevated amount of nutrients (N, P). Increased amount of plankton as well as enhanced activity of production-decomposition processes were observed in this area. Sometimes BOD_5 reached $12.5 \text{ mg O}_2 \text{ l}^{-1}$ in this most polluted area;
- Zone B: The cooling water outflow zone was the area of the greatest thermal impact, where water temperature in many cases

exceeded 28°C. The lowest abundance and variety of most planktonic organisms (phytoplankton and zooplankton) as well as lower rates of primary production and more intensive decomposition processes of organic matter were observed in this area;

- Zone C: The rest of the lake, including the deep and mediate-deep zones, where the various impact factors have affected the ecosystem occasionally, depending on the INPP operation, wind direction, waves.

3.1.2.1 Underwater video survey data

A drop-down type underwater video system was used for the bottom video survey. System comprised of the underwater and control units. The underwater unit was equipped with depth sensor, laser pointers for scale estimation, 4x50 watt xenon light bulbs, color (540 TVL) and black-white video cameras. Video from the underwater unit in real-time was transmitted into the control unit equipped with GPS antenna and overlay block, which allowed superimposing of GPS coordinates, depth sensor readings, current data and time into the video stream.

During the filming process, the underwater unit was descended vertically and hovered freely over the bottom not touching it. Therefore the difference between position of the control unit with GPS antenna and actual position of underwater unit was within GPS accuracy error (5-15 meters according to the used GPS module specifications), no additional correction of geo reference data needed. Camera altitude was regulated manually from video stream to ensure the best video quality and varied between approximately 30-50 centimetres above the bottom surface.

Overall, video footage consisted from 21 video transect and 490 video segments at the depth range from 0.5 to 10 m. Video analysis was performed by A. Šaškov (Zaiko et al., 2013).

3.1.2.2 Environmental predictors used in models

Three environmental predictors were selected for the model of the coverage of the invasive bivalve *Dreissena polymorpha*: depth,

wave exposure and slope. The exposure referred to the wind generated wave impact on the bottom expressed as Relative Exposure Index (REI), was calculated by wave exposure model WEMo 4.0 (Malhotra and Fonseca, 2007).

3.1.2.3 Sampling of macrozoobenthos

In total, 42 macrozoobenthos samples were taken in 2008 and 2010. An Ekman-type grab (catch area 0.0225 m²) was used for benthic macrofauna sampling. At least one sample was taken randomly at every video sampling site. Each sample was sieved (0.5 mm mesh size) and preserved (4% formaldehyde solution) on board. Once in the laboratory, the macrozoobenthos species were sorted out, identified and counted. Wet weight was determined to within 0.001g. Samples were collected and treated following standard guidelines for bottom macrofauna sampling (HELCOM, 1988). Analysis of samples was performed by A. Zaiko (Zaiko et al., 2013).

3.2 Empirical models and modelling procedure

In this work four species distribution models were selected in accordance to the most recent ecological studies (Wei et al., 2010; Reiss et al., 2011; Vincenzi et al., 2011; Downie et al., 2013; Sundblad et al., 2013): generalized additive models, multivariate adaptive regression splines, maximum entropy and random forests.

3.2.1 Species distribution models

Generalized additive models (GAM)

GAMs are semi-parametric extensions of generalized linear models with the assumption that the functions are additive and that the components are smooth. This method deals well with the highly non-linear and non-monotonic relationships between the set of explanatory and response variables (Guisan and Zimmerman, 2000). Model selection was based on penalized regression splines with default gamma-values and a maximum four degrees of freedom for continuous predictor variables in order to maintain ecologically

interpretable models (Wood and Augustin, 2002). The “mgcv” 1.7-9 package (Wood, 2006) within R environment was used for occurrence and biomass distribution models.

Multivariate adaptive regression splines (MARS)

MARS algorithm is a nonparametric method for multiple regression, which uses adaptively selected spline functions (Hansen and Kooperberg, 2002) developed by Friedman (1991). MARS is based on linear relationships, however it identifies and estimates a model which coefficients differ depending on the level of the predictor variable (Reiss et al., 2011). Models were built using the GLM approach and specified to include first order interactions, where significant. Both occurrence and biomass distribution models were built using the “earth” package (Milborrow, 2012) under R environment.

Maximum entropy (MaxEnt)

MaxEnt is a general-purpose machine learning method which estimates a target probability distribution by finding the probability distribution of maximum entropy and constraining the expected value of each environmental variable to match its empirical average (Phillips et al., 2006; Reiss et al., 2011). In this study MaxEnt program version 3.3.3e (Phillips et al., 2006; Phillips et al., 2008) was used. The convergence threshold was set at 10^{-5} and the maximum number of iterations at 500 to allow the algorithm to get close to convergence (Phillips et al., 2006). Although, MaxEnt works well with presence-only datasets (Elith et al., 2011), absence data was also used in MaxEnt models to be more consistent with other methods. MaxEnt was used only for modelling the occurrence probability of species.

Random Forests (RF)

RF is a classification and regression model developed by Breiman (2001) that generates multiple classification trees with a randomised subset of predictors (Reiss et al., 2011). A large number of trees are grown and the number of predictors used to find the best

split at each node is a randomly chosen subset of the total number of predictors (Prasad et al., 2006). In this study the number of trees was set to 1000, the number of variables randomly selected at each node and minimum node size were set to default values. The “randomForest 4.6-2” package (Liaw and Wiener, 2002) within the R environment was used for predictions of presence probability and biomass distribution of benthic species.

3.2.2 Predictive performance, model variation and effects of data traits

Several estimates are calculated for model validation: area under the receiver operating characteristic curve (AUC), root mean squared error (RMSE), root mean squared error normalised by range (NRMSE), mean absolute deviation (MAD), coefficients of variation (i.e. CV_{MAD}), r – correlation between observed and predicted values, coefficient of determination (R^2).

Predictive performance of the species occurrence models was estimated by AUC measures. The AUC values range between 0 and 1. According Hosmer and Lemeshow (2000) “excellent” prediction performance is achieved when $AUC > 0.9$, “good” performance – $AUC 0.7-0.9$, “poor” performance – $AUC < 0.7$. If $AUC \leq 0.5$ then predictions are no better than random. For biomass distribution models two measures were estimated: NRMSE and R^2 .

The accuracy of fish prey item biomass models was calculated by dividing CV_{MAD} from 100. The accuracy of 100% means that predictions are without errors (impossible to achieve), 0% means that prediction error is equal to the sample average.

The initial dataset was split into train set used for model build-up (70% of data) and test set used for validation (rest 30% of data) ensuring that species prevalence (the ratio between sites where a particular species is present and total number of sites) would be in equal proportions in train and test datasets. The variance of model performance was assessed by mean values of 100 iterations of splits for each species. Variation is expressed by the coefficient of variation

CV_{AUC} for occurrence models and CV_{NRMSE} and CV_R^2 for biomass distribution models.

The effect of data traits, species prevalence and distribution range, on both predictive performance and model variation were tested using Pearson's correlation between AUC, NRMSE, R^2 . The species distribution range was determined for each species using convex hull algorithm in Quantum GIS 1.7.4 (Quantum GIS Development Team, 2010). The species prevalence and distribution range significantly correlated ($r = 0.70$, $p < 0.01$), therefore the effect of the species prevalence was tested on the predictive performance of models.

3.3 Assessment of seabed quality for feeding of fish

3.3.1 Scheme of the assessment

The assessment procedure includes three parts: modelling of macrozoobenthos biomass (service provider module), analysis of fish prey items (service user module) and the output of the assessment: the seabed quality map for fish feeding (Figure 9). The first step is data acquisition: fish and macrofauna samples are gathered and processed, and then GIS layers of environmental factors (predictors) are created. The diets of the separate fish species are identified from an analysis of fish digestive tracts, after which biomass distribution models of prey items are set up on the basis of macrofauna sample analysis and layers of environmental predictors.

In the next step, weights for prey items are assigned, depending on their importance to the diet of a particular fish species, and in parallel, model predictions are transferred into the GIS environment, where biomass distribution maps are developed. Finally, biomass maps of separate prey items with assigned weights are overlain and quality maps of fish feeding grounds are generated. In addition, the accuracy is assessed to identify the reliability of the maps. The data in the service user module are not directly related to the service provider module and can be modified in accordance with the aims of the study (i.e. feeding grounds of a single fish species).

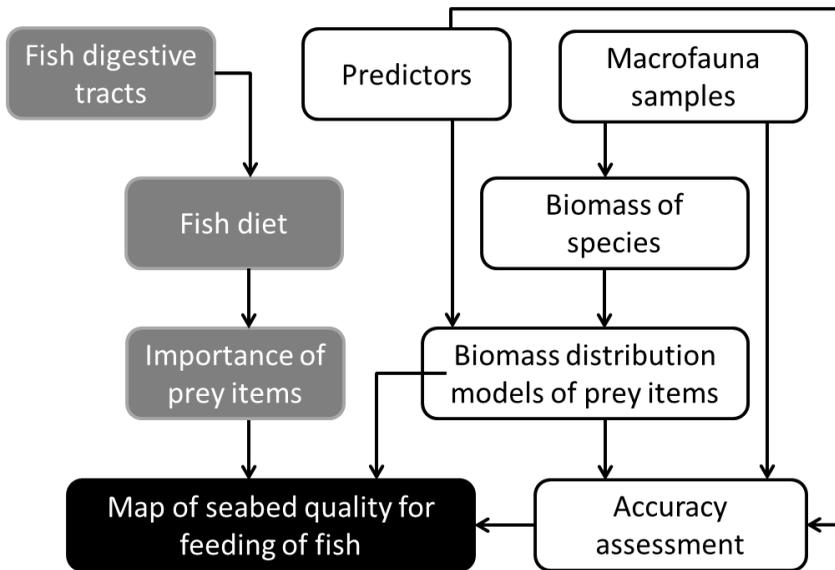


Figure 9. General scheme of the assessment of seabed quality for feeding of fish. White – service provider module, grey – service user module, black – output of the approach.

3.3.2 Occurrence and importance of prey items

The occurrence and importance of prey items were inferred from the analysis of fish digestive tracts. The former describes the relative frequency of a particular prey in all digestive tracts, while the latter indicates how much a particular prey item contributes to the total content in a discrete digestive tract. Both parameters were divided into three categories: high, moderate and low. A ‘high’ occurrence means that a particular benthic animal is found in more than 50 % of samples, ‘moderate’ – in 20-50 % of samples and ‘low’ in < 20 % of samples. A ‘high’ importance means that most of the digestive tract can be filled with a particular prey species (more than 50 % of tract content), ‘moderate’ – 20-50 % of tract content, while ‘low’ means that a particular item is only a small addition to the whole tract content

(< 20 % of tract content). The occurrence and importance of prey items are shown in Table 1. As the study aimed to evaluate the quality of the seabed for the feeding of fish, the assessment was based only on benthic invertebrates, excluding nectobenthic species and small pelagic fish.

Table 1. Occurrence in digestive tracts (first letter) and importance (second letter) of prey items for cod, flounder and eelpout. Empty cells indicate that fish do not prey on that particular item. H – high, M – moderate, L – low.

		Occurrence/Importance		
		Cod	Flounder	Eelpout
Prey items	Gammaridea	H/H	H/H	H/M
	<i>Halicryptus spinulosus</i>		M/M	
	<i>Hediste diversicolor</i>	M/L	M/L	M/L
	<i>Macoma balthica</i>		H/H	M/M
	<i>Marenzelleria neglecta</i>	L/L	L/L	L/L
	<i>Mya arenaria</i>		L/L	
	<i>Mytilus trossulus</i>		M/M	L/L
	<i>Saduria entomon</i>	H/H	M/L	M/L

3.3.3 Development of seabed quality maps for the feeding of fish

To predict the biomass distribution of prey species the Random forests regression modelling technique (Breiman, 2001) implemented in the “randomForest 4.6-2” package (Liaw and Wiener 2002) within the R environment was chosen. After the creation of biomass distribution rasters, maps of those prey items that a particular fish species feeds on were added up with different weights (Table 2). Weights are given according to the occurrence and importance shown in Table 1. Initial biomass values were multiplied by the weight in order to better reflect the important feeding items in the feeding ground map. As different multipliers were used, biomass units were

no longer suitable, so scores of weighted biomass was categorized into five levels of quality: very high, high, moderate, low and very low, where very high quality indicates the highest biomass aggregations of prey items with respect to their importance to fish diets. Finally, the maps for different fish species were combined and the map of overall seabed quality for the feeding of a given fish was produced.

Table 2. Raster weights according to the occurrence and importance of prey items for fish feeding. H –high, M – moderate, L – low.

Occurrence/Importance	Weight
H/H	1
H/M or M/H	0.75
M/M	0.5
M/L or L/M	0.37
L/L	0.25

3.3.4 Accuracy of seabed quality maps

Three levels of accuracy were generated for the quality map of fish feeding grounds. The accuracy indicated how well or badly different quartiles of a predictor range were covered by macrofauna samples. First of all, the accuracy of biomass distribution of each prey item was estimated. In relation to partial plots, every predictor was split into four intervals/categories (predictors with presence/absence data were split into two) and the number of macrofauna samples was counted for each interval/category. Since 171 samples were used for the model build up, 171 was the total point pool split between intervals/categories of a single predictor. Then the ‘Reclassify’ function was used to reclassify the predictor layer assigning these points for all intervals/categories. These point scores were multiplied by the mean decrease accuracy value (Table 13) produced by the model. In this way the accuracy of the most important predictor receives the highest weight and minor predictors had a proportionally lower impact on overall accuracy. Finally, the accuracy layers of every prey item were added up, then split into three categories (high,

moderate, low) using the geometrical interval classification method; ultimately, an accuracy layer for the feeding grounds was produced.

A ‘*high*’ accuracy is interpreted as the best possible area modelled with the current dataset, though validation errors must still be taken into account. Areas of ‘*moderate*’ accuracy should be treated as trustworthy, although they should be studied more closely before making a decision. A ‘*low*’ accuracy indicates areas that are modelled on the basis of just a few samples and should be treated with caution.

3.4 Biological valorisation approach

The concept of biological valuation of the marine environment was developed by Derous et al. (2007) and was followed by the biological valorisation methodology for the assessment of the seabed introduced by Węśławski et al. (2009).

3.4.1 Assigning biological value to benthic communities

Biological value for benthic species is assigned according ten features given in the questionnaire (Table 3). Highest feature value (3 points) is given for habitat builders and long-living species with relatively low regeneration time. Moderate value (2 points) is given for key species, rare species and species specific to a single habitat. Lowest value (1 point) is assigned for bioturbators, filter feeders, native species and species sensitive to mechanical disturbance and oxygen depletion.

The scores of biological value for benthic species according the questionnaire (Table 3) are given in Table 4. The scores are adapted from Węśławski et al. (2009). For these species which were not included in the original study, thus were evaluated independently.

Table 3. A questionnaire for biological valuation of species (after Węśławski et al. 2009).

Question	Explanation	Weight
1. Is the species a habitat builder?	A perennial, encrusting, erect, large species that provides shelter for other species	3
2. Is the generation time over 2 years?	A long-lived species will regenerate slowly after disturbance	3
3. Is it a key species?	A species that is a major predator or an important food item, placed centrally in the food web or primary production	2
4. Is the species rare?	A species that is encountered in single locations and/or as single specimens only	2
5. Is the species specific to one habitat only?	A species that has a narrow niche and is strongly linked to specific physical conditions	2
6. Is the species a bioturbator?	A species that stirs up the sediment, enhancing oxygen exchange	1
7. Is the species a filter feeder?	A species that removes particles from the water, thereby enhancing transparency	1
8. Is the species native?	A native species is natural to the area, unlike a non-indigenous one	1
9. Is the species sensitive to mechanical disturbance?	A sessile, crustose species is vulnerable to siltation and rapid water dynamics	1
10. Is the species sensitive to oxygen depletion?	A species that is sensitive to oxygen depletion is most vulnerable to environmental stress	1

Table 4. Scores of species biological valuation according the questionnaire (Table 3). Q1-Q10 represent questions of the questionnaire. * indicates species that were evaluated independently from the original methodology.

Taxon	Score	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
<i>Furcellaria lumbricalis</i> *	12	3	3	2	0	2	0	0	1	1	0
<i>Mytilus trossulus</i>	11	3	3	2	0	0	0	1	1	1	0
<i>Mya arenaria</i>	10	0	3	2	0	0	1	1	1	1	1
<i>Saduria entomon</i>	10	0	3	2	0	2	1	0	1	1	0
<i>Polysiphonia nigrescens</i> *	10	3	3	2	0	0	0	0	1	1	0
<i>Cladophora rupestris</i> *	10	3	3	2	0	0	0	0	1	1	0
<i>Macoma balthica</i>	9	0	3	2	0	0	1	1	1	1	0
<i>Cerastoderma lamarcki</i>	7	0	3	0	0	0	0	1	1	1	1
<i>Ceramium spp.</i> *	7	3	0	2	0	0	0	0	1	1	0
<i>Balanus improvisus</i>	6	3	0	0	0	0	0	1	0	1	1
<i>Bathyporeia pilosa</i>	6	0	0	2	0	2	0	0	1	0	1
Pontoporeiids	6	0	0	0	2	2	0	0	1	0	1
<i>Corophium volutator</i>	6	0	0	2	0	0	0	1	1	1	1
<i>Idotea balthica</i>	5	0	0	2	0	0	0	0	1	1	1
<i>Pilayella littoralis</i> *	5	3	0	0	0	0	0	0	1	1	0
Ostracoda undet.*	5	0	0	0	0	0	1	1	1	1	1
<i>Fabricia sabella</i>	4	0	0	2	0	0	0	1	1	0	0
<i>Gammarus spp.</i>	4	0	0	2	0	0	0	0	1	0	1
<i>Halicryptus spinulosus</i>	4	0	0	2	0	0	1	0	1	0	0
<i>Hediste diversicolor</i>	4	0	0	2	0	0	1	0	1	0	0
<i>Bylgides sarsi</i> *	3	0	0	2	0	0	0	0	1	0	0
<i>Jaera albifrons</i>	3	0	0	0	0	0	0	0	1	1	1
<i>Marenzelleria neglecta</i>	3	0	0	2	0	0	1	0	0	0	0
Oligochaeta undet.	3	0	0	2	0	0	0	0	1	0	0
<i>Pygospio elegans</i>	3	0	0	2	0	0	0	0	1	0	0
<i>Hydrobia sp.</i>	2	0	0	0	0	0	0	0	1	0	1
<i>Streblospio shrubsolii</i> *	2	0	0	0	0	0	1	0	1	0	0
<i>Theodoxus fluviatilis</i>	2	0	0	0	0	0	0	0	1	0	1

Biological valorisation of LMA was made using the same dataset described in section 3.1.1.3 (Figure 7). Additional data on macrofitobenthos were added for the relevant sites. Benthic communities were distinguished based on biomass dominance (Olenin, 1997). Every community gained a specific biological value based on three criteria (score range 1-4): species richness, biological value of present species and the size of a habitat. The scores of these three criteria are added up and the overall biological value of a particular community is estimated.

The score of species richness was estimated in the following way. The total number of species in the research area is divided into quartiles. Scores ranging from 1 to 4 are assigned for intervals separated by these quartiles. The score for a particular benthic community depends on the total number of species in the community (Table 5).

Similarly to species richness, the total biological value of all the species in the research area is also divided into quartiles. The scores from 1 to 4 are assigned depending on the total value of all the species occurring in the particular community. For example, if the sum of biological values of all species in the community is more than 75 % of the total biological value of all the species in the research area, then the community scores 4 points. Respectively, 50-75 % scores 3 points, 25-50 % scores 2 points and less than 25 % scores 1 point (Table 5).

Table 5. Scores for three components of biological valorisation: number of species, sum of biological values of species and habitat size. Each component can contribute 1-4 points to the value of the community.

Score points	Total number of species in the community	The sum of species biological values	Habitat size, km ²
4	21-28	120-159	< 50
3	14-20	80-119	50-200
2	7-13	40-79	200-500
1	≤ 6	≤ 39	> 500

According to the original methodology, smaller habitats contribute to the higher biological value for associated communities. The size of a habitat was estimated combining results from the questionnaire for Baltic marine experts on the Red List assessments of the Baltic Sea biotopes/habitats and calculation of areas between the same community sites using GIS. The categories of scores are given in Table 5.

3.4.2 Assignment of biological values for research sites

Three components of the biological value are estimated based on all samples that belong to the specific community. In other words, it is a maximum value that a particular sample can acquire depending on the overall score of the community. However, different samples of the same community vary in both species richness and biomass, meaning that some of them represent healthy and prosperous community, while others reflect a poor state. For this reason the quality of communities is classified into four types: optimal, good, fair and poor. Only those sites which represent the optimal state of the community get the maximum value estimated based on the three components mentioned above. Values for good, fair and poor state are corrected according to the ratio between species richness in the whole community and the particular site, together with ratio between maximum biomass of the community and the biomass at the research site. Quality classes were distinguished based on quartiles as explained in section 3.4.1. Corrected biological values for different quality classes are given in Table 6. The final biological value is the average of species richness and biomass values.

Table 6. Assignment of biological values for sites depending on the quality of the community.

Max value of the community	Species richness of the community				Maximum biomass in the community			
	>75 %	50-75 %	25-50 %	<25 %	>75 %	50-75 %	25-50 %	<25 %
	Optimal	Good	Fair	Poor	Optimal	Good	Fair	Poor
12	12	9	6	3	12	9	6	3
11	11	8.25	5.5	2.75	11	8.25	5.5	2.75
10	10	7.5	5	2.5	10	7.5	5	2.5
9	9	6.75	4.5	2.25	9	6.75	4.5	2.25
8	8	6	4	2	8	6	4	2
7	7	5.25	3.5	1.75	7	5.25	3.5	1.75
6	6	4.5	3	1.5	6	4.5	3	1.5
5	5	3.75	2.5	1.25	5	3.75	2.5	1.25
4	4	3	2	1	4	3	2	1
3	3	2.25	1.5	0.75	3	2.25	1.5	0.75

4 RESULTS

4.1 Comparative analysis of macrozoobenthos species distribution modelling

4.1.1 Application of occurrence models

All four compared methods on average achieved “good” predictive performance for occurrence models (Figure 10).

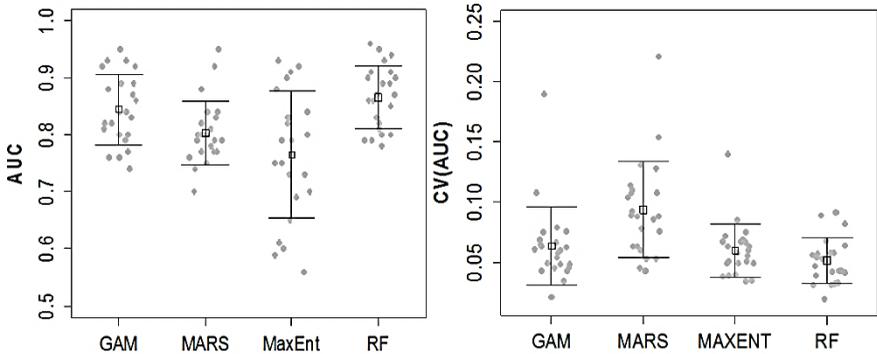


Figure 10. Performance of four predictive methods for modelling of the occurrence of benthic invertebrates according area under the curve (AUC) values (left) and coefficients of variation of AUC during 100 of iterations (right).

The highest performance was achieved by RF (AUC = 0.87 ± 0.06), followed by GAM (AUC = 0.84 ± 0.06), MARS (AUC = 0.80 ± 0.06) and MaxEnt (AUC = 0.77 ± 0.11). RF models were also the most consistent ranging from “good” to “excellent” performance (AUC = 0.78-0.96), closely followed by GAM (AUC = 0.74-0.95) and MARS (AUC = 0.70-0.95), while MaxEnt (AUC = 0.56-0.93) had six cases of “poor” predictive performance. According to coefficients of variation of AUC the most consistent method was again RF ($CV_{AUC} = 0.05 \pm 0.02$), closely followed by MaxEnt ($CV_{AUC} = 0.06 \pm 0.02$) and

GAM ($CV_{AUC} = 0.06 \pm 0.03$), while MARS ($CV_{AUC} = 0.09 \pm 0.04$) varied the most.

RF models of occurrence achieved top performance among all methods for 19 species, followed by GAM (6 species), MARS and MaxEnt (1 species per each). MaxEnt and MARS showed the worst performance for 14 and 10 species respectively), whereas for GAM and RF that was never the case (Table 7). Overall predictions of the occurrence were excellent ($AUC > 0.9$) for eight species or higher taxa: *Halicryptus spinulosus*, *Mytilus trossulus*, *Saduria entomon*, *Fabricia sabella*, *Idotea balthica*, *Jaera albifrons*, Ostracoda, *Monoporeia affinis* and *Theodoxus fluviatilis*, however only ostracods were predicted excellent by all four methods.

4.1.2 Application of biomass distribution models

Figure 11 indicates that the mean prediction error of all three methods for the biomass distribution was very similar ($NRMSE = 0.08 \pm 0.04$) among the methods. According to the coefficient of determination (R^2) the best mean performance was achieved by RF ($R^2 = 0.32 \pm 0.19$), followed by MARS ($R^2 = 0.13 \pm 0.14$) and GAM ($R^2 = 0.12 \pm 0.12$).

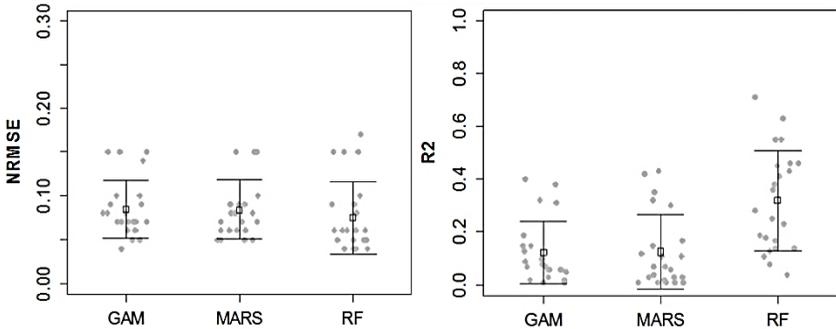


Figure 11. Performance of three predictive methods for modelling of the distribution of benthic invertebrates biomass according root mean square error normalized by range (NRMSE) and coefficient of determination (R^2) values.

Table 7. Validation results of four predictive modelling methods for occurrence and biomass distribution of 23 benthic species or higher taxa in Lithuanian marine area. AUC – area under the curve, NRMSE – root mean square error normalized by range, R^2 – coefficient of determination, Prev. – prevalence.

Phylum, class, order, family	Species, taxa	GAM			MARS			RF			MaxEnt	Prev.
		AUC	NRMSE	R^2	AUC	NRMSE	R^2	AUC	NRMSE	R^2	AUC	
Priapulida	<i>Halicryptus spinulosus</i>	0.89	0.08	0.19	0.88	0.06	0.32	0.91	0.06	0.46	0.79	0.33
Polychaeta												
	<i>Nereidae</i> <i>Hediste diversicolor</i>	0.76	0.15	0.08	0.79	0.15	0.11	0.79	0.15	0.41	0.60	0.68
	<i>Polynoidae</i> <i>Bylgides sarsi</i>	0.80	0.07	0.01	0.74	0.08	0.01	0.79	0.06	0.14	0.75	0.19
	<i>Sabelliidae</i> <i>Fabricia sabella</i>	0.92	0.07	0.07	0.84	0.08	0.11	0.95	0.06	0.19	0.91	0.10
	<i>Spionidae</i> <i>Marenzelleria neglecta</i>	0.80	0.10	0.15	0.77	0.10	0.12	0.81	0.10	0.28	0.56	0.81
	<i>Pygospio elegans</i>	0.83	0.08	0.01	0.79	0.08	0.01	0.80	0.08	0.14	0.61	0.69
	<i>Streblospio shrubsolii</i>	0.87	0.05	0.03	0.78	0.05	0.03	0.87	0.06	0.11	0.83	0.07
Oligochaeta	Oligochaeta undet.	0.77	0.04	0.05	0.77	0.05	0.01	0.82	0.04	0.08	0.65	0.59
Crustacea												
	<i>Amphipoda</i> <i>Bathyporeia pilosa</i>	0.82	0.05	0.06	0.79	0.05	0.02	0.86	0.04	0.46	0.79	0.22
	<i>Corophium volutator</i>	0.84	0.10	0.07	0.84	0.09	0.17	0.86	0.04	0.25	0.73	0.39
	<i>Gammarus spp.</i>	0.81	0.09	0.13	0.76	0.08	0.15	0.83	0.09	0.45	0.75	0.29
	<i>Monoporeia affinis</i>	0.88	0.08	0.06	0.83	0.09	0.03	0.90	0.04	0.18	0.80	0.15

<i>Cirripediae</i>	<i>Balanus improvisus</i>	0.89	0.06	0.31	0.81	0.05	0.30	0.89	0.05	0.55	0.84	0.22
<i>Isopoda</i>	<i>Idotea balthica</i>	0.79	0.07	0.06	0.77	0.07	0.01	0.87	0.06	0.04	0.93	0.03
	<i>Jaera albifrons</i>	0.93	0.06	0.10	0.78	0.06	0.03	0.93	0.04	0.36	0.88	0.15
	<i>Saduria entomon</i>	0.92	0.14	0.32	0.92	0.15	0.35	0.94	0.15	0.71	0.83	0.31
<i>Ostracoda</i>	Ostracoda undet.	0.95	0.06	0.15	0.95	0.07	0.07	0.96	0.05	0.38	0.92	0.10
Gastropoda												
<i>Hydrobiidae</i>	<i>Hydrobia sp.</i>	0.76	0.09	0.02	0.75	0.09	0.04	0.78	0.09	0.13	0.70	0.38
<i>Neritidae</i>	<i>Theodoxus fluviatilis</i>	0.93	0.07	0.09	0.82	0.06	0.01	0.91	0.05	0.23	0.90	0.14
Bivalvia												
<i>Cardiidae</i>	<i>Cerastoderma lamarcki</i>	0.74	0.07	0.06	0.70	0.06	0.07	0.80	0.05	0.17	0.73	0.13
<i>Myidae</i>	<i>Mya arenaria</i>	0.82	0.07	0.02	0.80	0.07	0.06	0.89	0.05	0.43	0.69	0.48
<i>Mytilidae</i>	<i>Mytilus trossulus</i>	0.86	0.15	0.38	0.80	0.15	0.42	0.90	0.17	0.55	0.82	0.25
<i>Tellinidae</i>	<i>Macoma balthica</i>	0.84	0.15	0.40	0.80	0.15	0.43	0.85	0.15	0.63	0.59	0.76

As shown in Figure 12 the R^2 of biomass distribution models by RF was the most consistent ($CV_{R^2} = 0.45 \pm 0.29$), but with the highest variance of NRMSE ($CV_{NRMSE} = 0.80 \pm 0.39$). GAMs were relatively consistent according to R^2 ($CV_{R^2} = 0.88 \pm 0.41$) and NRMSE ($CV_{NRMSE} = 0.52 \pm 0.25$), while the models of MARS were consistent in respect of NRMSE ($CV_{NRMSE} = 0.51 \pm 0.23$), but with the high variance in R^2 ($CV_{R^2} = 1.30 \pm 0.77$).

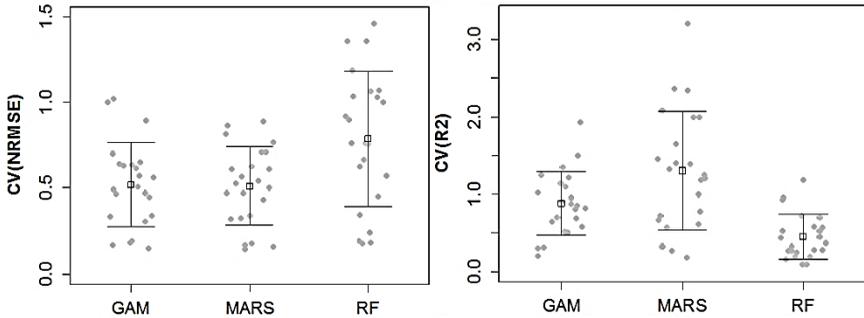


Figure 12. Coefficients of variation of root mean square error normalized by range (NRMSE) and coefficient of determination (R^2) of four predictive modelling methods during 100 iterations.

According coefficient of determination RF was the best in predicting biomass distribution for all species except one, while GAM and MARS had the worst predictive performance for 14 and 12 species respectively (Table 7). The most accurate predictions for the distribution of biomass were recorded for *Balanus improvisus*, *Macoma balthica*, *M. trossulus* and *S. entomon* ($R^2 > 0.5$).

4.1.3 Effects of data traits

The correlation between the prevalence and AUC values were negative for all methods, whereas correlation between prevalence and both NRMSE and R^2 were always positive (Table 8).

Table 8. A correlation matrix of the performance of four predictive modelling methods and species prevalence. AUC – area under the curve, NRMSE – root mean square error normalized by range, R² – coefficient of determination.

	AUC				NRMSE			R ²		
	GAM	MARS	RF	MaxEnt	GAM	MARS	RF	GAM	MARS	RF
GAM	1				1			1		
MARS	0.76	1			0.97	1		0.93	1	
RF	0.90	0.76	1		0.92	0.92	1	0.81	0.79	1
MaxEnt	0.66	0.39	0.75	1						
Prevalence	-0.41	-0.13	-0.51	-0.92	0.47	0.45	0.45	0.15	0.23	0.20

This indicates that methods tend to predict occurrence better with less occasions of species presence. While this effect was very weak for MARS ($r = -0.13$, $p > 0.05$) and moderate for GAM and RF ($r = -0.41$ and $r = -0.51$ respectively, $p < 0.05$), MaxEnt models had a very strong negative correlation with prevalence. On the contrary, performance of biomass distribution models tend to get better with increasing prevalence, however this effect was moderate in case of NRMSE and only weak-very weak in case of R². Detailed results for the separate macrozoobenthos species are given in Table 9.

Table 9. The performance of the occurrence models (based on AUC values) and their relation to the species prevalence. Figures in bold – “excellent” performance, normal – “good” performance, bold italic – “poor” performance.

	GAM	MARS	RF	MaxEnt	Prevalence
<i>M. neglecta</i>	0.80	0.77	0.81	0.56	0.81
<i>M. balthica</i>	0.84	0.80	0.85	0.59	0.76
<i>P. elegans</i>	0.83	0.79	0.80	0.61	0.69
<i>H. diversicolor</i>	0.76	0.79	0.79	0.60	0.68
Ologochaeta undet.	0.77	0.77	0.82	0.65	0.59
<i>M. arenaria</i>	0.82	0.80	0.89	0.69	0.48
<i>C. volutator</i>	0.84	0.84	0.86	0.73	0.39
<i>Hydrobia</i> sp.	0.76	0.75	0.78	0.70	0.38
<i>H. spinulosus</i>	0.89	0.88	0.91	0.79	0.33
<i>S. entomon</i>	0.92	0.92	0.94	0.83	0.31
<i>Gammarus</i> spp.	0.81	0.76	0.83	0.75	0.29
<i>M. trossulus</i>	0.86	0.80	0.90	0.82	0.25
<i>B. improvisus</i>	0.89	0.81	0.89	0.84	0.22
<i>B. pilosa</i>	0.82	0.79	0.86	0.79	0.22
<i>H. sarsi</i>	0.80	0.74	0.79	0.75	0.19
<i>M. affinis</i>	0.88	0.83	0.90	0.80	0.15
<i>J. albifrons</i>	0.93	0.78	0.93	0.88	0.15
<i>T. fluviatilis</i>	0.93	0.82	0.91	0.90	0.14
<i>C. lamarcki</i>	0.74	0.70	0.80	0.73	0.13
<i>F. sabella</i>	0.92	0.84	0.95	0.91	0.10
Ostracoda undet.	0.95	0.95	0.96	0.92	0.10
<i>S. shrubsolii</i>	0.87	0.78	0.87	0.83	0.07
<i>I. balthica</i>	0.79	0.77	0.87	0.93	0.03

Low to moderate negative correlations ($r \leq -0.51$) were determined between prevalence and coefficients of variation of models except MaxEnt (Table 10), meaning that consistency of predictions during iterations increases with decreasing occasions of species occurrence. MARS was the most sensitive in case of occurrence models, whereas biomass distribution models showed similar results for all SDMs.

Table 10. Correlation matrix of the variation of performance of four predictive modelling methods and species prevalence. AUC – area under the curve, NRMSE – root mean square error normalized by range, R^2 – coefficient of determination.

	AUC				NRMSE			R^2		
	CV_{GAM}	CV_{MARS}	CV_{RF}	CV_{MaxEnt}	CV_{GAM}	CV_{MARS}	CV_{RF}	CV_{GAM}	CV_{MARS}	CV_{RF}
CV_{GAM}	1				1			1		
CV_{MARS}	0.77	1			0.95	1		0.70	1	
CV_{RF}	0.76	0.62	1		0.71	0.73	1	0.59	0.82	1
CV_{MaxEnt}	0.34	0.12	0.68	1						
Prevalence	-0.17	-0.46	-0.01	0.09	-0.41	-0.43	-0.51	-0.22	-0.48	-0.46

Detailed results for the separate macrozoobenthos species are given in Table 11.

Table 11. The performance of the biomass distribution models (based on R^2 values) and their relation to the species prevalence. Figures in bold – $R^2 > 0.5$, normal – $R^2 = 0.1-0.5$, bold italic – $R^2 < 0.1$.

	GAM	MARS	RF	Prevalence
<i>M. neglecta</i>	0.15	0.12	0.28	0.81
<i>M. balthica</i>	0.40	0.43	0.63	0.76
<i>P. elegans</i>	0.01	0.01	0.14	0.69
<i>H. diversicolor</i>	0.08	0.11	0.41	0.68
Ologochaeta undet.	0.05	0.01	0.08	0.59
<i>M. arenaria</i>	0.02	0.06	0.43	0.48
<i>C. volutator</i>	0.07	0.17	0.25	0.39
<i>Hydrobia</i> sp.	0.02	0.04	0.13	0.38
<i>H. spinulosus</i>	0.19	0.32	0.46	0.33
<i>S. entomon</i>	0.32	0.35	0.71	0.31
<i>Gammarus</i> spp.	0.13	0.15	0.45	0.29
<i>M. trossulus</i>	0.38	0.42	0.55	0.25
<i>B. improvisus</i>	0.31	0.30	0.55	0.22
<i>B. pilosa</i>	0.06	0.02	0.46	0.22
<i>H. sarsi</i>	0.01	0.01	0.14	0.19
<i>M. affinis</i>	0.06	0.03	0.18	0.15
<i>J. albifrons</i>	0.10	0.03	0.36	0.15
<i>T. fluviatilis</i>	0.09	0.01	0.23	0.14
<i>C. lamarcki</i>	0.06	0.07	0.17	0.13
<i>F. sabella</i>	0.07	0.11	0.19	0.10
Ostracoda undet.	0.15	0.07	0.38	0.10
<i>S. shrubsolii</i>	0.03	0.03	0.11	0.07
<i>I. balthica</i>	0.06	0.01	0.04	0.03

4.2 Assessment of benthic habitat quality in terms of fish food provision

4.2.1 Diet composition of cod, flounder and eelpout

Eight macrozoobenthos species or higher taxa were identified during the analysis of fish stomach contents (Table 1). Cod mainly preyed upon isopods *S. entomon* and gammarideans, while polychaete worms were of minor importance. Preferred prey items for flounder and eelpout were gammarideans and bivalves *M. balthica*, while priapulids *H. spinulosus* and soft-shell clams *M. arenaria* were eaten only by flounder. Flounder had the most diverse diet composition (a total of eight prey items), while eelpout and cod preyed upon six and four prey items respectively. Half of the prey items were eaten by all three species, while two items (*H. spinulosus* and *M. arenaria*) were exclusively fed on by flounder. Different weights were assigned to every fish species separately according to the occurrence and importance of prey items (Table 12).

Table 12. Weight multipliers of prey items assigned according to occurrence and importance. Empty cells indicate that fish do not prey upon that particular item.

	Cod	Flounder	Eelpout
Gammaridea	1	1	0.75
<i>Halicryptus spinulosus</i>		0.5	
<i>Hediste diversicolor</i>	0.37	0.37	0.37
<i>Macoma balthica</i>		1	0.5
<i>Marenzelleria neglecta</i>	0.25	0.25	0.25
<i>Mya arenaria</i>		0.25	
<i>Mytilus trossulus</i>		0.5	0.25
<i>Saduria entomon</i>	1	0.37	0.37

4.2.2 Importance of environmental predictors

In general, for the predictions of fish prey items, the most important predictor was near-bottom oxygen level. Orbital velocity, salinity and sediments were also important, near-bottom current velocity were less important, temperature had only a minor importance or no importance at all in some cases.

The mean decrease accuracy (%IncMSE) was calculated for each predictor in order to evaluate its importance to the response variable (Table 13). The most important predictor was near-bottom oxygen concentration, especially for deep-living species like *M. balthica*, *S. entomon* and *H. spinulosus* (28.7, 12.1 and 24.6 %IncMSE respectively). Orbital velocity, salinity and sediments were also important: the biomasses of amphipods and *M. trossulus* were mostly dependent on sediments (9.3 and 34.8 %IncMSE respectively), while salinity had a major influence on both polychaete worms and *M. balthica*, and orbital velocity on *H. spinulosus* and *S. entomon* (12.7 and 18.9 respectively).

Table 13. Mean decrease accuracy (%IncMSE) of environmental predictors.

A higher value indicates greater importance. H.s. – *H. spinulosus*, Gam. – *Gammaridea*, M.n. – *M. neglecta*, H.d. – *H. diversicolor*, M.t. – *M. trossulus*, M.a. – *M. arenaria*, M.b. – *M. balthica*, S.e. – *S. entomon*.

Predictors	Gam.	H.d.	H.s.	M.a.	M.b.	M.t.	M.n.	S.e.
Current velocity	3.8	6.4	3.5	7.6	22.4	3.9	0.5	7.2
Orbital velocity	2.4	12.0	12.7	6.9	18.0	9.6	7.9	18.9
Salinity	6.7	16.3	3.8	0.2	25.1	17.0	7.4	15.0
Oxygen concentration	7.1	10.7	12.1	9.2	28.7	16.1	3.9	24.6
Sediment types	9.3	3.8	7.7	0.7	22.2	34.8	4.7	10.1
Halocline	0.4	3.2	5.5	4.6	-1.4	1.4	1.3	6.3
Thermocline	1.8	2.7	0.4	-4.2	14.4	10.4	0.7	5.8

Detailed dependencies between predictors and response variables are shown in Figure 13.

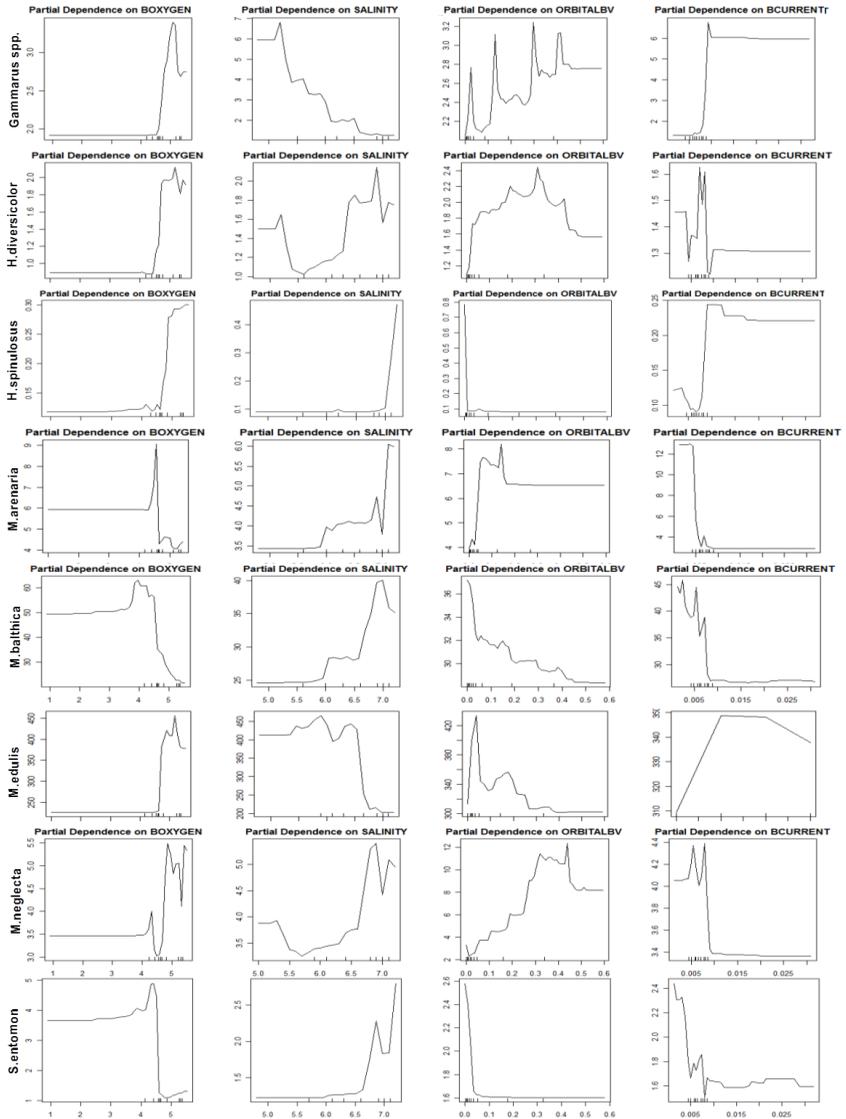


Figure 13. Partial plots between numerical environmental predictors and modelled species. X-axis: predictor values; Y-axis: biomass, g m⁻².

4.2.3 Validation of models

Output validation values of all models are given in Table 14. Highest prediction errors (MAD) were estimated for bivalves *M. trossulus* and *M. balthica* (223.5 and 26.4 respectively, lowest for priapulid and polychaete worms *H. spinulosus* and *H. diversicolor* (0.1 and 1.4 respectively). This is expected because higher average or range values generate bigger MAD, so MAD says little about model performance in the context of other models if we do not take the average into account. For this reason CV_{MAD} are calculated.

Table 14. Validation results of prey item biomass models. Columns from left to right: prey items; average sample biomass \pm standard deviation; mean absolute deviation (MAD); coefficient of variation of MAD (CV_{MAD}); correlation of observations and predictions.

Prey items	Mean biomass $g\ m^{-2}$	Validation		
		MAD $g\ m^{-2}$	CV_{MAD} %	r
<i>Gammaridea</i>	7.8 \pm 16.3	2.6	33.2	0.67
<i>Hediste diversicolor</i>	2.0 \pm 3.2	1.4	71.2	0.64
<i>Halicryptus spinulosus</i>	0.3 \pm 1.1	0.1	38.1	0.68
<i>Mya arenaria</i>	6.5 \pm 17.4	3.0	46.3	0.65
<i>Macoma balthica</i>	43.4 \pm 53.8	26.4	60.9	0.79
<i>Mytilus trossulus</i>	1385.4 \pm 1398.9	223.5	16.1	0.74
<i>Marenzelleria neglecta</i>	3.8 \pm 9.4	2.7	70.1	0.53
<i>Saduria entomon</i>	5.6 \pm 5.8	1.8	32.6	0.84

If we put CV_{MAD} values of all models together (Figure 14) it is possible to compare their errors/accuracies. The most accurate model was of *M. trossulus*. This is probably because this mollusc strictly occurs only on hard substrate and usually in high biomass thus making an easier task for the predictive model. Models of *S. entomon*, *Gammarus spp.*, *H. spinulosus* and *M. arenaria* were also quite accurate (accuracy >50%). The model of *M. balthica* was less accurate

(<40%), the lowest accuracy was estimated for polychaete models (<30%) probably due to their patchy distribution.

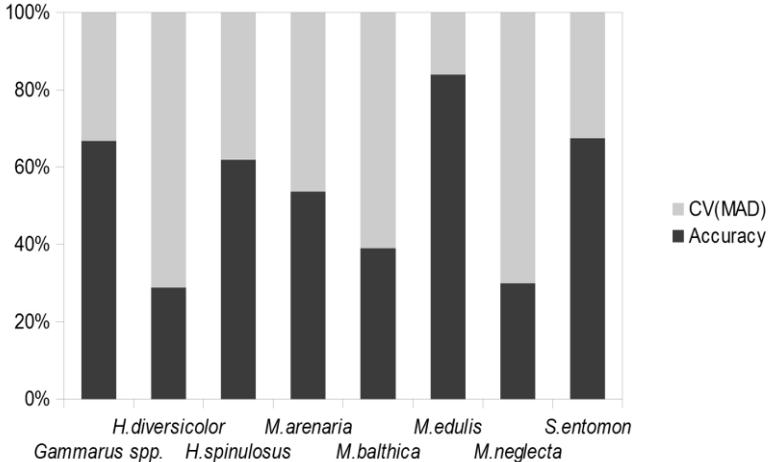


Figure 14. Accuracy and coefficients of variation of mean absolute deviation (CV_{MAD}) for all models. Accuracy of 100% means that predictions are without errors (impossible to achieve), 0% means that prediction error is equal to the sample average.

4.2.4 Seabed quality for feeding of fish

The high quality areas of the seabed for the feeding of Baltic cod (Figure 15) are distributed down to 50 meter depth, while 60 meter isobath almost perfectly separates important areas from unimportant. One of the most important areas in the northern part of the coastal zone is determined by hard substrate seafloor. Other two very important areas are situated in the central and westernmost part of LMA. These two are mostly determined by high biomass of glacial relict *S. entomon*.

The most important area for the feeding of flounder is situated in the Northern part of the coastal zone (Figure 16). This area is determined by high biomass of hard-substrate associated blue mussel *M. trossulus*, even if this mollusc is moderately important for the feeding of flounder. However, flounder is present down to 80 meters or more, so in order to delineate important areas in the offshore zone *M. trossulus* were excluded from the map (small map in Figure 16). This map shows that the most important soft bottom areas are distributed from 20 down to 60 meter depth. Shallower areas are mostly determined by infaunal clams and polychaetes, while deeper ones by *M. balthica* and *S. entomon*.

The most important areas for feeding of eelpout (Figure 17) are very similar to the flounder case, probably because of similar diet composition. However, eelpout is a coastal fish so the focus must be on near-shore zone before delineating areas of importance. It is obvious that most important areas for eelpout are coastal hard substrate area from 5 down to 20 meter depth situated in the North-eastern part of LMA. It is hard to delineate important areas in soft-bottom seafloor (small map in Figure 17) due to more or less homogeneous distribution of prey items.

The highest quality feeding grounds for all three fish species is the stony bottom in the coastal area situated in the northern part of LMA. Other high quality areas are located in the offshore zone: one in an offshore bank with heterogeneous sediments at 50 m depth (western part of LMA), another in the soft bottom at 40-50 m depths (central part of LMA). The total area of these habitats comprise 7.6 % of LMA.

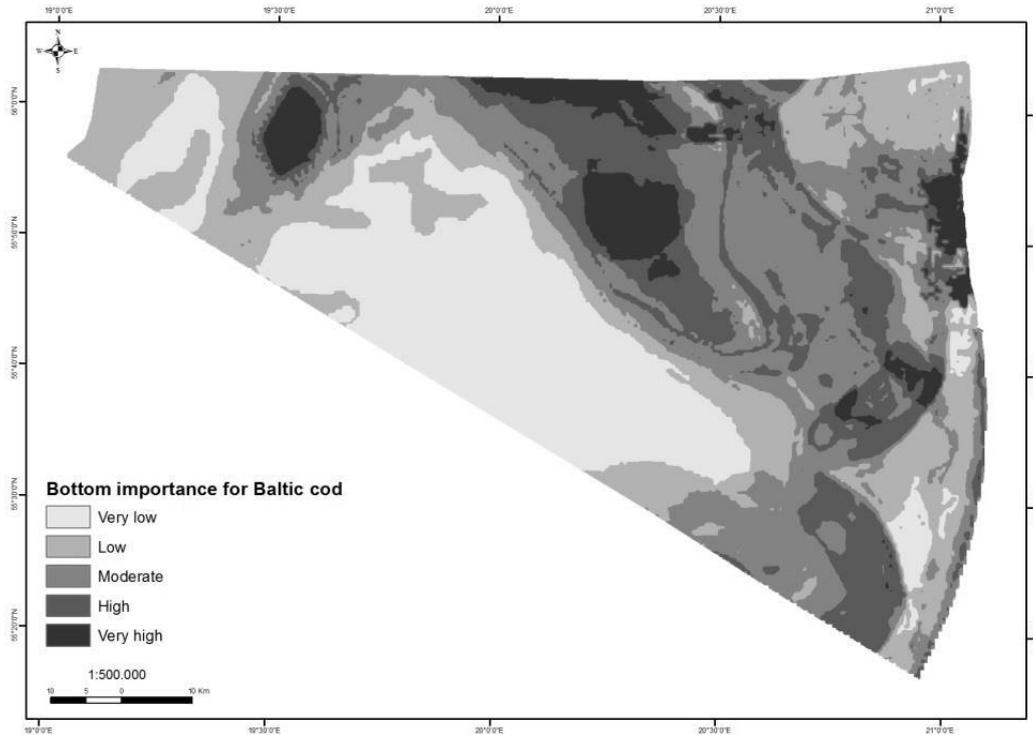


Figure 15. Seabed quality for feeding of Baltic cod based on the biomass distribution of prey items.

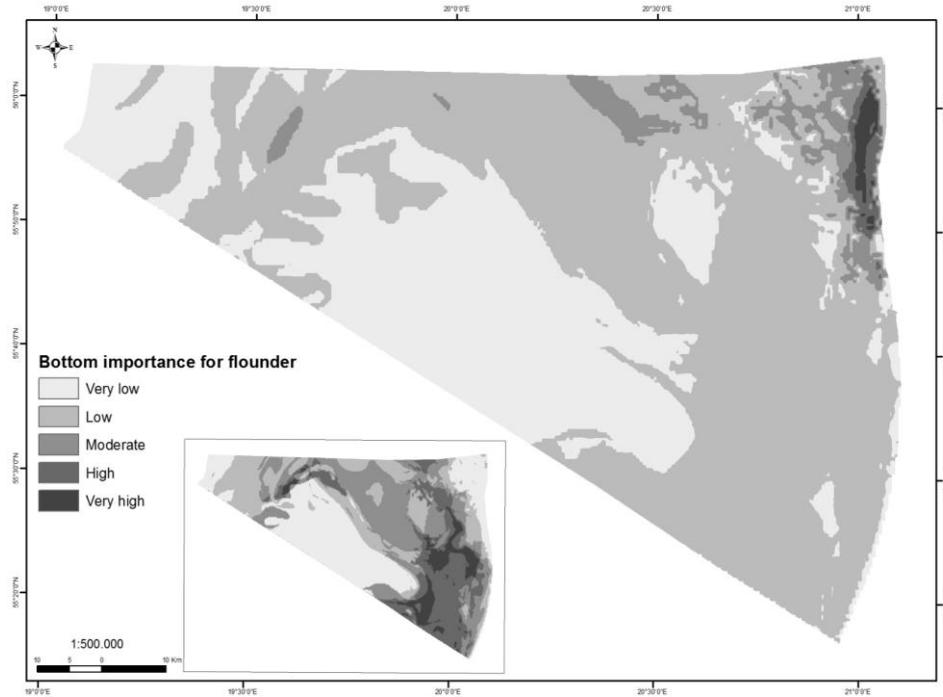


Figure 16. Seabed quality for feeding of European flounder based on the biomass distribution of prey items. Smaller map shows the same importance excluding *Mytilus trossulus*.

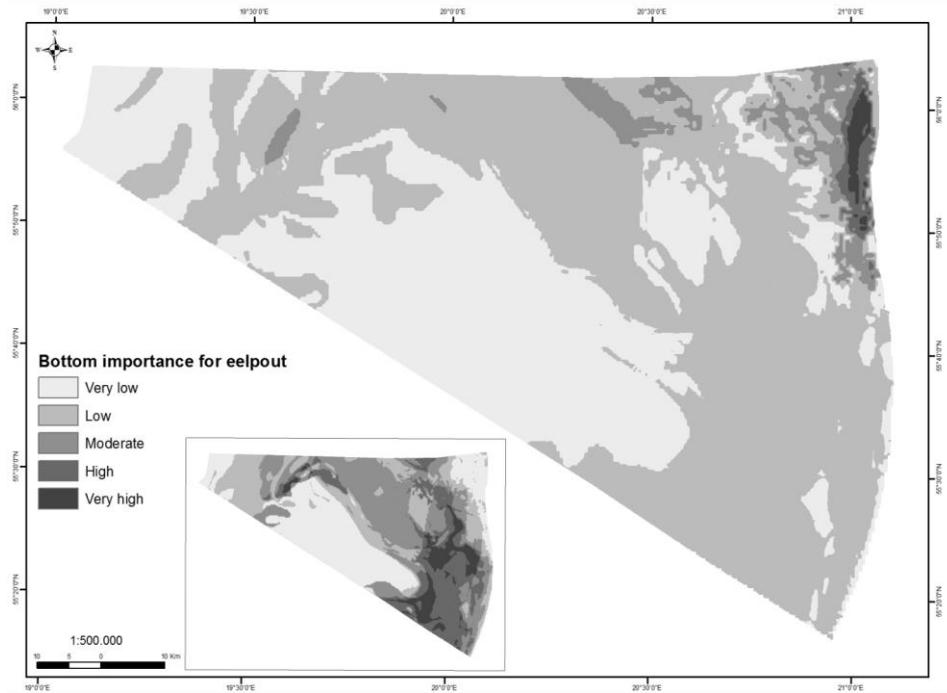


Figure 17. Seabed quality for feeding of viviparous eelpout based on the biomass distribution of prey items. Smaller map shows the same importance excluding *Mytilus trossulus*.

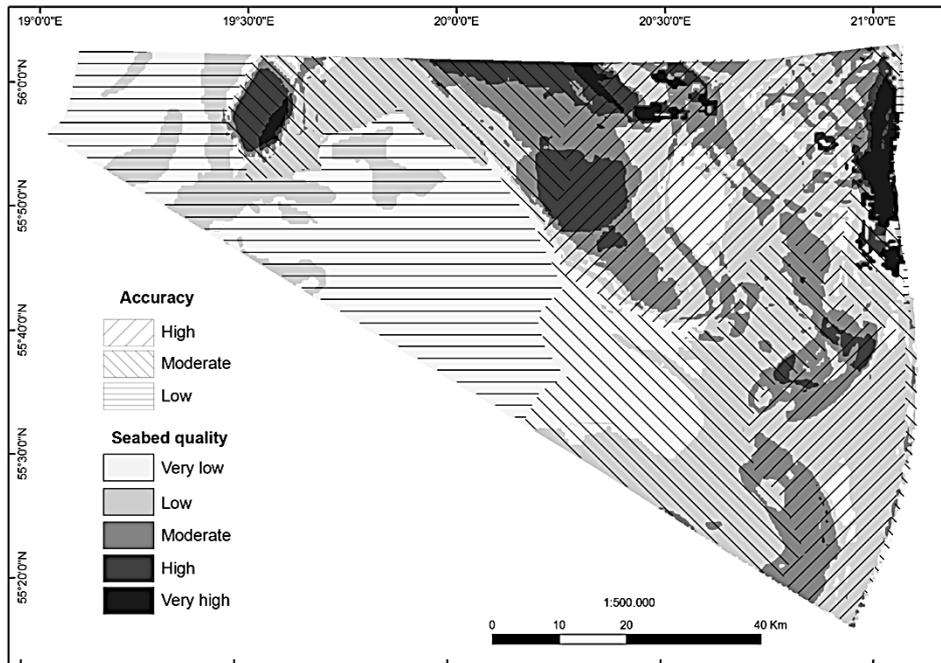


Figure 18. Map of seabed quality in the Lithuanian Economic Zone for the feeding of Baltic cod, European flounder and viviparous eelpout (with an accuracy assessment).

4.2.5 Accuracy of fish feeding ground maps

The accuracy assessment of fish feeding ground maps based on the number of field samples per category or interval of every environmental predictor is shown in Table 15.

Table 15. Number of field samples per category or interval of environmental predictor

Category/interval		Halicryptus	Gammarus	Marenzelleria	Hediste	Mytilus	Mya	Macoma	Saduria
SEDIMENTS	boulders	21	24	24	22	23	21	22	23
	cobbles/gravel	16	12	16	14	14	18	15	16
	sand	96	95	88	99	83	88	96	96
	silt	38	40	43	36	45	44	38	36
THERMOCLINE	absence	140	142	137	145	131	131	136	135
	presence	31	29	34	26	34	40	35	36
HALOCLINE	absence	164	159	159	159	153	158	160	47
	presence	7	12	12	12	12	13	11	124
NEAR-BOTTOM	0	26	33	33	33	33	38	36	35
CURRENT	0-0.01	141	135	135	134	129	130	131	131
VELOCITY	0.01-0.02	1	0	1	1	1	1	1	1
	0.02-0.03	3	3	2	3	2	2	3	4
	>0.03	0	0	0	0	0	0	0	0
ORBITAL VELOCITY	0	14	13	17	15	32	16	11	11
	0-0.2	121	124	125	124	106	131	127	128
	0.2-0.4	21	19	15	18	14	15	20	20
	0.4-0.6	15	15	14	14	13	9	13	12
	>0.6	0	0	0	0	0	0	0	0
SALINITY	<4.8	0	0	0	0	0	0	0	0
	4.8-5.4	12	14	7	10	6	12	11	11
	5.4-6.0	19	20	15	18	15	17	19	16
	6.0-6.6	39	36	36	39	42	38	37	38
	6.6-7.2	101	101	113	104	102	104	104	106
	>7.2	0	0	0	0	0	0	0	0
NEAR-BOTTOM OXYGEN	<0.9	0	0	0	0	0	0	0	0
0.9-2.06	3	2	2	2	1	2	2	3	
2.06-3.22	3	4	4	5	6	6	4	4	
3.22-4.38	19	27	24	26	22	24	26	22	
4.38-5.53	146	138	141	138	136	139	139	142	
>5.53	0	0	0	0	0	0	0	0	

The accuracy assessment indicates that the most accurate areas of the approach are at 10-40 m depths. The low accuracy areas were justified by only 18% of total samples and were set in very shallow areas (down to 3 m depth) and for the deepest areas (> 60-70 m).

Accuracy was moderate for offshore areas in the central part of LMA and for the coastal area. More than half the samples were taken in the coastal area, but because of the rapid changes in some environmental parameters (especially salinity and near-bottom orbital velocity) the quartiles of these predictors were only moderately justified in terms of accuracy.

4.3 Biological valorisation of the seabed in LMA

4.3.1 Biological values of benthic communities

In total, 10 benthic communities were identified in the research area (Table 16 and Table 17). Nine of them are dominated by macrozoobenthos species and one by marcophytobenthos. Two communities were associated with hard bottoms (*F. lumbricalis* and *M. trossulus/B. improvisus*), while other eight were occurring in soft bottoms. Biological values of benthic communities were estimated according 3 components: species richness, biological value of species in the community and the habitat size (Table 16).

Table 16. Biological value of benthic communities based on three components: species richness, the sum of biological values of species and the size of a habitat.

Community	Species richness	Component value	Species value	Component value	Habitat size, km ²	Component value	Total value
<i>F. lumbricalis</i>	23	4	115	4	15	4	12
<i>C. lamarcki</i>	13	2	55	2	25	4	8
<i>M. balthica</i>	21	4	81	3	3000	1	8
<i>M. trossulus/ B. improvisus</i>	18	3	77	3	300	2	8
<i>M. arenaria</i>	12	2	49	2	180	3	7
<i>M. neglecta/ B. pilosa</i>	12	2	49	2	80	3	7
<i>H. diversicolor</i>	11	2	32	2	250	2	6
<i>P. elegans</i>	9	2	51	2	250	2	6
Pontoporeiids	4	1	15	1	150	3	5
<i>B. sarsi</i>	3	1	12	1	300	2	4

According to the species richness, the most valuable communities were *F. lumbricalis* and *M. balthica* with 23 and 21 species or higher taxa respectively. The lowest species richness was in the communities of Pontoporeiids and *B. sarsi*, where less than 4 species were present. The highest biological value of species was in the community of *F. lumbricalis*, while the lowest for the communities of Pontoporeiids and *B. sarsi*. The smallest habitats (thus the most valuable) were estimated for the communities of *F. lumbricalis* (15 km²) and *C. lamarcki* (25 km²), while the largest was of *M. balthica* (3000 km²).

The highest biological value was assigned to the community of red algae *F. lumbricalis* (12 points), scoring maximum points in all three components of valuation. High biological value was estimated for the communities of bivalve molluscs *C. lamarcki* (8 points), *Macoma balthica* (8 points), *M. trossulus*/*B. improvisus* (8 points) and *M. arenaria* (7 points), also 8 points were assigned for the community of amphipods *B. pilosa* and polychaete *M. neglecta* (8 points). The lowest values (4-6 points) were scored for the communities of polychaetes *H. diversicolor*, *P. elegans*, *B. sarsi* and Pontoporeiids (*Monoporeia affinis* and *Pontoporeia femorata*).

Table 17. Biological values of communities (3-12 points) based on three components (1-4 points): species richness, species biological values and habitat size. Community values are divided into four states (from poor to optimal) depending on number of species and biomass in the research site. Non-italic figures indicate the range of number of species and biomass representing the condition of the community, italic figures – biological values.

Community	Max value	Species, ind. m ⁻²				Biomass, g m ⁻²			
		Optimal	Good	Fair	Poor	Optimal	Good	Fair	Poor
<i>F. lumbricalis</i>	12	> 14 <i>12</i>	10-13 <i>9</i>	5-9 <i>6</i>	< 4 <i>3</i>	> 4500 <i>12</i>	3000-4500 <i>9</i>	1500-3000 <i>6</i>	< 1500 <i>3</i>
<i>C. lamarcki</i>	8	> 9 <i>8</i>	6-8 <i>6</i>	3-5 <i>4</i>	< 2 <i>2</i>	> 45 <i>8</i>	30-45 <i>6</i>	15-30 <i>4</i>	< 15 <i>2</i>
<i>M. balthica</i>	8	> 11 <i>8</i>	8-10 <i>6</i>	4-7 <i>4</i>	< 3 <i>2</i>	> 225 <i>8</i>	150-225 <i>6</i>	75-150 <i>4</i>	< 75 <i>2</i>
<i>M. trossulus</i> / <i>B. improvisus</i>	8	> 11 <i>8</i>	8-10 <i>6</i>	4-7 <i>4</i>	< 3 <i>2</i>	> 6000 <i>8</i>	3000-6000 <i>6</i>	1500-3000 <i>4</i>	< 1500 <i>2</i>
<i>M. neglecta</i> / <i>B. pilosa</i>	7	> 7 <i>7</i>	5-6 <i>5.25</i>	3-4 <i>3.5</i>	< 2 <i>1.75</i>	> 65 <i>7</i>	45-65 <i>5.25</i>	22-45 <i>3.5</i>	< 22 <i>1.75</i>
<i>M. arenaria</i>	7	> 8 <i>7</i>	6-7 <i>5.25</i>	3-5 <i>3.5</i>	< 2 <i>1.75</i>	> 150 <i>7</i>	100-150 <i>5.25</i>	50-100 <i>3.5</i>	< 50 <i>1.75</i>
<i>H. diversicolor</i>	6	> 6 <i>6</i>	4-5 <i>4.5</i>	2-3 <i>3</i>	1 <i>1.5</i>	> 13 <i>6</i>	9-13 <i>4.5</i>	4.5-9 <i>3</i>	< 4.5 <i>1.5</i>
<i>P. elegans</i>	6	> 5 <i>6</i>	3-4 <i>4.5</i>	2 <i>3</i>	1 <i>1.5</i>	> 0.75 <i>6</i>	0.5-0.75 <i>4.5</i>	0.25-0.5 <i>3</i>	< 0.25 <i>1.5</i>
Pontoporeiids	5	> 4 <i>5</i>	3 <i>3.75</i>	2 <i>2.5</i>	1 <i>1.25</i>	> 1.5 <i>5</i>	1-1.5 <i>3.75</i>	0.5-1 <i>2.5</i>	< 0.5 <i>1.25</i>
<i>B. sarsi</i>	4	3 <i>4</i>	2 <i>3</i>	1 <i>2</i>	0 <i>1</i>	> 0.2 <i>4</i>	0.15-0.2 <i>3</i>	0.07-0.15 <i>2</i>	< 0.07 <i>1</i>

4.3.2 Biological valorisation model of the Lithuanian economic zone

Validation results of biological valorisation model showed that the predictions were accurate. Mean absolute deviation was 1.07 points what could be expressed as 8.9 % considering that the range of score points was 12. The correlation between observed and predicted values was strong and significant ($r = 0.73$, $p < 0.001$).

The most important environmental predictors for the model of biological valorisation were the Secchi depth, sediments, wave generated orbital velocity and near-bottom oxygen concentration (Table 18). Less important was near-bottom current velocity while slope and curvature were not important at all.

Table 18. Mean decrease accuracy (%IncMSE) of environmental predictors. Higher values indicate greater importance of the particular predictor.

Environmental predictors	%IncMSE
Secchi depth	25.46
Sediments	24.02
Orbital velocity	23.01
Near-bottom oxygen concentration	22.59
Near-bottom current velocity	15.98
Slope	-2.80
Curvature	-3.43

The map of biological valuation reveals that the highest biological value was estimated for the northern coastal area where stony bottoms vegetated by *F. lumbricalis* are present (Figure 19). High value soft bottom areas are situated at 10-35 m depths, where rich communities of bivalves *M. balthica*, *M. arenaria* and *C. lamarcki* are present. The lowest values are associated to sub-halocline areas with poor communities of *M. balthica*, Pontoporeiids and *B. sarsi*.

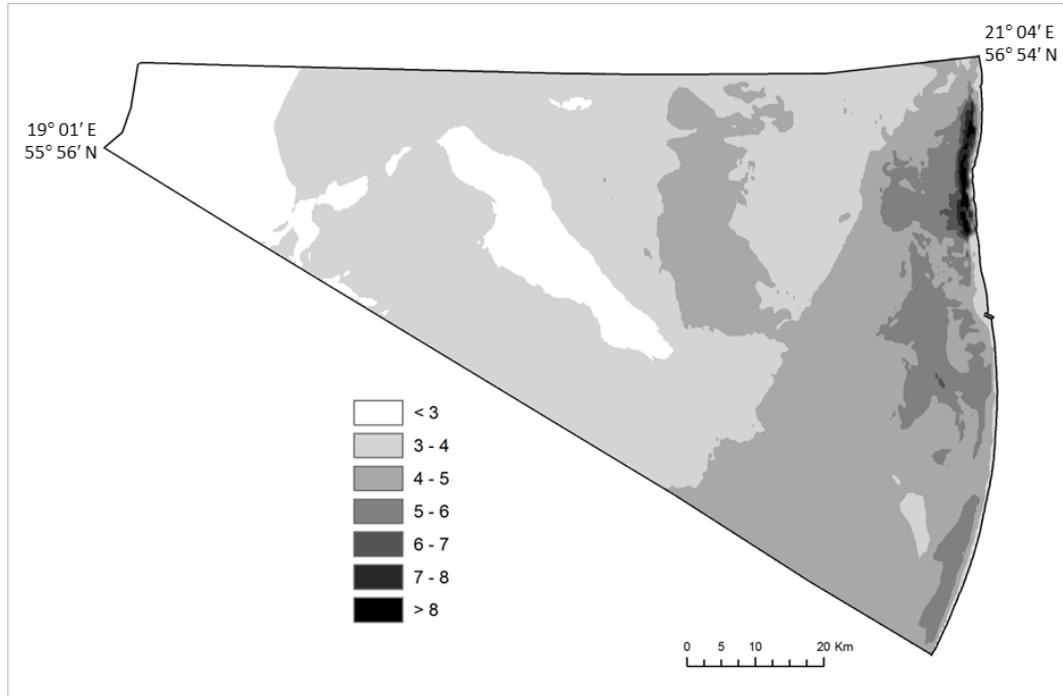


Figure 19. The results of the biological valuation of the Lithuanian marine area. Black color indicates highest values, white color – lowest values.

4.4 Predictive mapping of benthic invasive species *Dreissena polymorpha* in the lake Drūkšiai based on video survey data

Analysis of the video samples has shown that within the studied area the zebra mussel coverage varied from 0 up to 100 %. There were sites with single mussels, mussel clumps and clusters of clumps mixed with empty shells captured on the video (Figure 20). However, in most cases, live mussels were well identifiable, providing rather good quality of the coverage assessments. Their presence (> 0% of the coverage) was registered in more than 50% of the analyzed video samples.

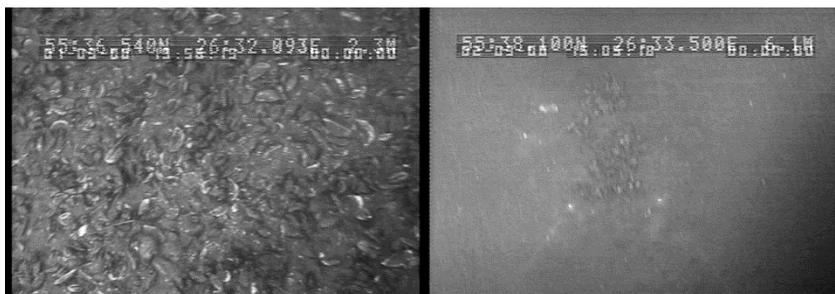


Figure 20. clusters of clumps mixed with empty shells (left) and separate mussel clumps (right) capture in the video footage.

The results of the validation have shown that the model of zebra mussel coverage was accurate. Correlation between observed and predicted values were very strong ($r = 0.87$, $p < 0.001$) and mean absolute deviation was 10.4, meaning that the average error of the predictions was around 10 % (comparable with video sample assesment accuracy). To avoid overestimations, predictions less than 10 % of coverage were marked as unlikely and excluded them from further analysis.

According to the modelling results, the area where *D. polymorpha* was present covered about 10.6 square kilometers and comprised about 24% of the total lake bottom area (Fig. 4). The highest coverage was modelled at 5-6 meter depths in the more exposed areas. Deeper than 6 meters the coverage drops significantly, and deeper than 9-10 meters *D. polymorpha* was no longer present. These findings were consistent with the results of the video analysis and earlier underwater SCUBA surveys.

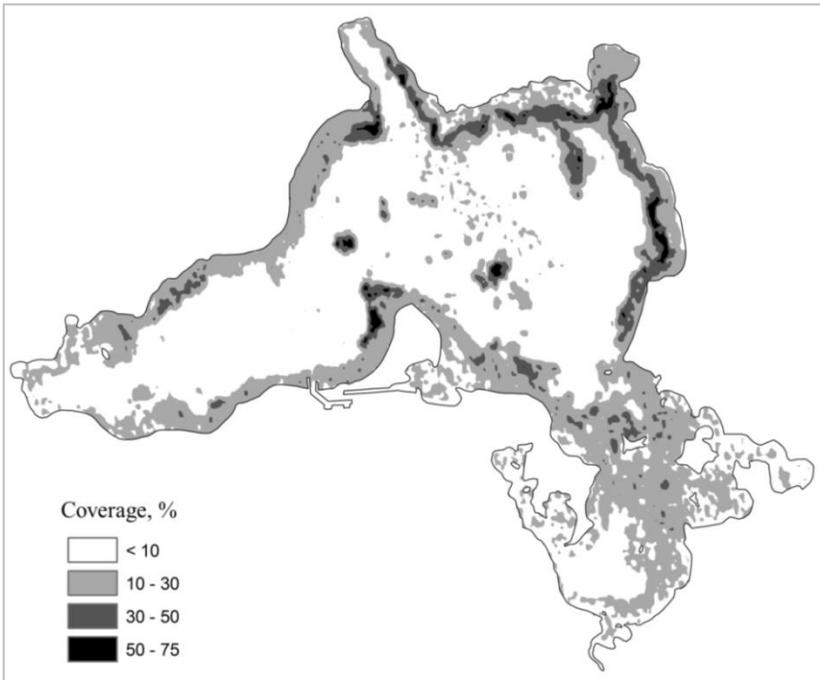


Figure 21. Modelled coverage of *Dreissena polymorpha* in the lake Drūkšiai.

There was a strong correlation ($r = 0.74$, $p < 0.001$) between predicted coverage and the sampled biomass of zebra mussel (Figure 22). There was also a strong correlation ($r = 0.79$, $p < 0.001$) between

predicted coverage and the sampled abundance of zebra mussel (Figure 23).

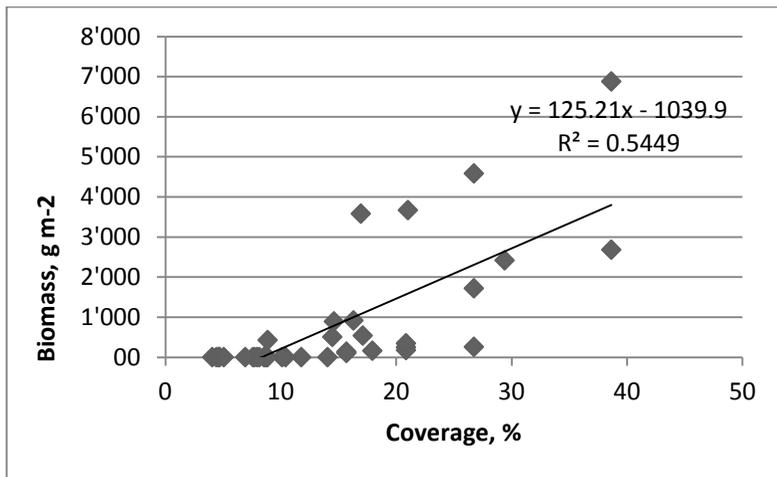


Figure 22. Relationship between modelled coverage and sampled biomass of *Dreissena polymorpha*.

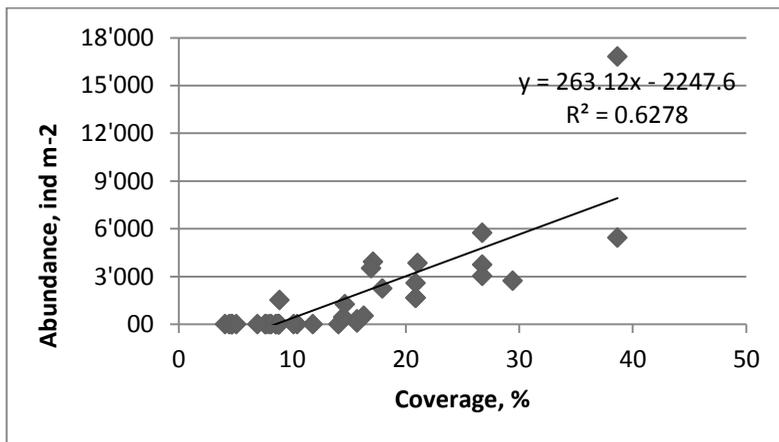


Figure 23. Relationship between modelled coverage and sampled abundance of *Dreissena polymorpha*.

Based on the linear regression equation ($y = 125.21 \cdot x - 1039.9$) the biomass in the whole lake was estimated (Figure 24).

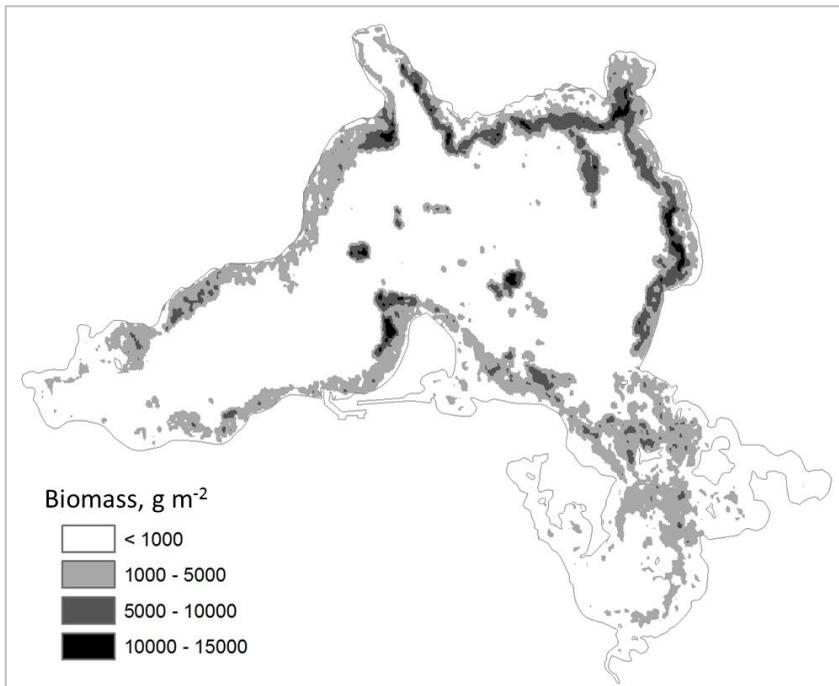


Figure 24. The biomass of *Dreissena polymorpha* in the lake Drūkšiai based on the linear regression model between predicted coverage and sampled biomass.

According to relationships between the modelled coverage of *D. polymorpha* and its sampled abundance and biomass, the total abundance and biomass of zebra mussel in the lake was estimated. In total, over 60 billion of individuals weighting more than 45 thousand tons was present in the area.

The highest densities of zebra mussel was estimated in the north-eastern part of the lake where bottom habitats are highly exposed to waves (Figure 25).

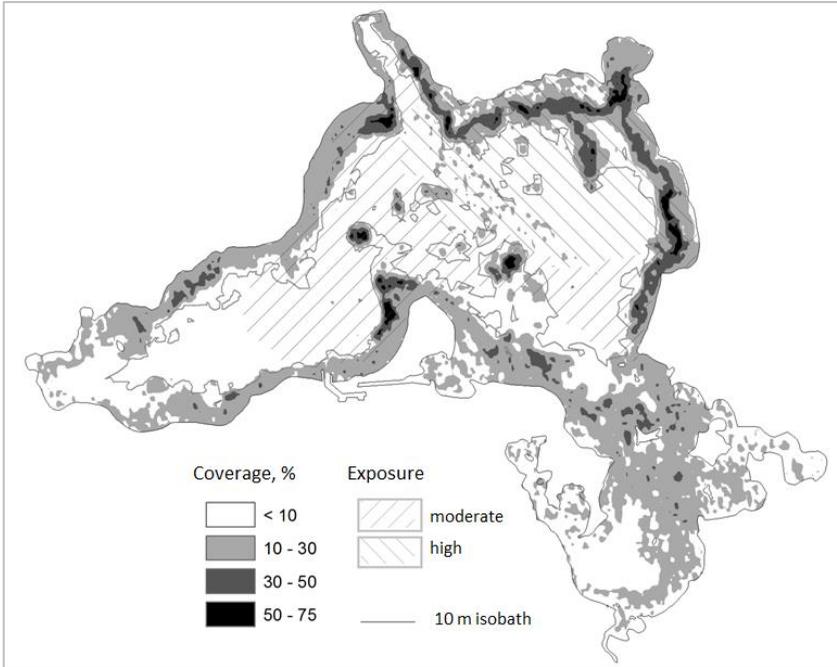


Figure 25. The modelled coverage of *Dreissena polymorpha* and the wave exposure to benthic habitats (areas with no stripes are of low exposure).

The mean coverage of *D. polymorpha* was the highest in highly exposed (41.1 ± 14.3) and moderately exposed (32.7 ± 17.1) habitats at 3-7 m depths (Figure 26, Table 19).

Table 19. Results of two-way ANOVA: dependent variable – modelled coverage of *D. polymorpha*, factors – depth (<3 m, 3-7 m, 7-10 m) and exposure (low, moderate, high).

Sources of variation	d.f.	Mean square	F-ratio	P-value
Depth	2	318495.0	3529.03	<0.001
Exposure	2	126430.0	1400.89	<0.001
Depth x Exposure	4	41603.6	460.98	<0.001
Residual	11645	90.2		

The lowest impact of zebra mussel was estimated for the deep sheltered (5.8 ± 4.7) and moderately exposed (7.9 ± 7.4) habitats. There was a significant increase in coverage going from sheltered to exposed habitats at the depths of 3-10 m, while the most exposed shallow habitats were significantly less impacted than moderately exposed and did not differ from low exposed areas (respectively 14.0 ± 9.7 , 22.2 ± 11.3 , 12.7 ± 6.9).

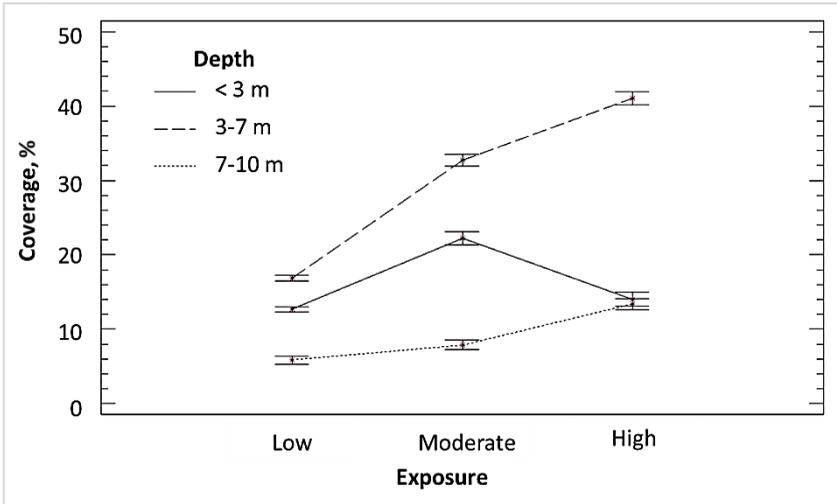


Figure 26. The mean cover of *D. polymorpha* at different depths within low, moderate and high wave exposure on bottom sediments. Homogeneity of groups was tested with Tukey's HSD test ($N = 11\ 654$, Mean \pm 0.95 Conf. intervals).

5 DISCUSSION

5.1 The use of species distribution models in benthic ecology and their limitations

All predictive modelling techniques, in general, provided useful models, while the machine learning RF method achieved the best predictive performance on both, occurrence and biomass data. This is in accordance with other recent studies (Gislason et al., 2006; Cutler et al., 2007; Collin et al., 2011).

Nevertheless, the predictive performance of models by GAM, MARS and MaxEnt was close to RF. Performance of RF occurrence models was relatively “good” (AUC > 0.8) for the most of the species, while other modelling methods, especially MaxEnt, were “good” only for few species. The MaxEnt case can be explained by a very strong negative correlation between AUC and prevalence, indicating that MaxEnt was relatively inaccurate for the widespread species (AUC = 0.56-0.61, prevalence \geq 0.68), such as bivalve *M. balthica* and polychaete worms: *M. neglecta*, *H. diversicolor* and *P. elegans* (Table 7 and Table 9). On the other hand, predictive performance of MaxEnt and other methods was relatively good for less dispersed species (AUC = 0.90-0.93, prevalence \leq 0.14), such as coastal hard-bottom associated *F. sabella*, *I. balthica* and *T. fluviatilis*.

In biomass models RF was superior over GAM and MARS according to the coefficient of determination in most of the cases and explained up to 40 % of variance more than others. This was most notable in *M. arenaria* and *B. pilosa* models, where the coefficients of determination of GAM and MARS were < 10 %, while RF achieved > 40 %.

As it was stressed in several studies (Araujo and New, 2007; Grenouillet et al., 2011; Bučas et al., 2013), even if the performance of different methods (measured as AUC, r^2 , NRMSE or other metrics) is comparable, the predictions produced by these models will inevitably be somewhat different and there is no way to identify the best method.

Thus it is recommended to apply more than a single modelling method (ensemble approach).

This study showed that the data splitting into train and test datasets can play a significant role on the performance of models. Depending on the random split the models achieved from “poor” (AUC < 0.7) to “perfect” (AUC > 0.9) performance. In this respect the most sensitive method was MARS for both, occurrence and biomass distribution models. For example, on the average of 100 iterations, the models of *M. balthica* achieved “good” prediction performance, but iterations ranged from nearly coin-flip accurate to perfect (AUC = 0.56-0.94). The most consistent method was again RF in respect of AUC and R². On the other hand, data traits, such as number of samples, sampling density, response type, number of predictors included in the model, data response prevalence (for occurrence models) and variance in the response (for abundance models) can also significantly affect the performance of modelling methods (Manel et al., 2001; Hernandez et al., 2006; McPherson and Jetz, 2007; Li and Heap, 2008; Bean et al., 2011; Bučas et al., 2013).

Although MaxEnt showed the worst overall results in comparison to other SDMs, it can be very useful for the modelling of species which are not well sampled with the Van-Veen grab sampler, since MaxEnt can manage only with presence data (Elith et al., 2011). If the density of species individuals is less than one individual per square meter or more, then even several replicates of grab samples (sampled area 0.1 m²) can easily miss the species and falsely provide absence information for the model. MaxEnt in this particular case would ignore possible false absence data and build a model only on presence data which is always true. Thus apart from less disperse species mentioned above, MaxEnt could be the most effective modelling tool for species such as *S. entomon* and nectobenthos such as mysids and shrimps.

In this study the same set of predictors was used for all the species to be more consistent when comparing different modelling techniques. However, as these species are associated with different environments, some of them were modelled less accurately most

likely because of the limitations of the used parameters. For instance, a model performance of small isopods *I. balthica* would increase significantly should the data on the macroalgae, to which they are associated (Vetter et al., 1999) were available and included in the model. Similarly, the data on the total organic content is very important for benthic species distribution modelling (Gogina and Zettler, 2010), especially for predictions of deposit feeders such as polychaetes *H. diversicolor*, *M. neglecta* and *P. elegans* (Olenin, 1997).

On the other hand, uneven distribution of sampling sites can result in different spatial accuracy, thus in the case of LMA more dense sampling in deeper areas should provide better models for widespread and deep living species. Nevertheless, both occurrence and biomass distribution models of several species were relatively good (i.e. *M. balthica*, *M. trossulus*) and can provide reliable spatial maps for further ecological studies or marine spatial planning and management.

5.2 Valuating the quality of benthic habitats: the added value of predictive modelling

The main advantage of the predictive modelling is the possibility to acquire spatial information from point based data with statistical background. Spatial information can be moulded into maps which can further provide basis for biological valuation maps (BVMs), crucially needed for the management for sustainable use and conservation of the sea (Derous et al., 2007).

The proposed method for the mapping of benthic habitat quality in terms of the fish feeding ground provision are one of the examples of species distribution modelling and the creation of BVMs. Such maps provide a different view of the spatial distribution of valuable seabed areas as they do not necessarily coincide with the high catch areas of selected fish species.

It is known that it can take more than 30 hours for prey to be digested (Macdonald et al., 1982), depending on the size of both

predator and prey (Santos and Jobling, 1991; Bromley, 1994) as well as on water temperature (Tyler, 1970). Furthermore, the sustained speed of cod can reach 0.6-0.9 BL s⁻¹ (He, 1991; Björnsson, 1993), meaning that 60 cm cod can swim for 38-58 km before their prey are digested. This shows that high catch areas of mobile fish whose stomachs are filled with benthic invertebrates do not necessarily correspond to the good quality of the seabed, for there is no proof that the fish were caught in an actual feeding ground. Certainly, this is not the case with low mobility species like flounder and eelpout.

On the other hand, these maps do not evaluate the suitability of a given environment for fish species apart from the biomass distribution of prey items and their importance to the diet. It may happen that a prey biomass is very high but the fish has limited access to this environment or the environment may be unsuitable in the context of factors other than feeding. For instance, the eelpout is exclusively associated with coastal hard bottoms, so other areas (even of the highest quality) are irrelevant to this species. Nevertheless, if the quality map of feeding grounds were combined with fish distribution maps, it would elevate our knowledge to a different level.

As in many other modelling approaches the outcome of our method is dependent on the quality of the initial data. The type of data for the service user module can be selected according to the aim of a study (in our case relatively robust data were sufficient) and could range from several categories of importance based on expert knowledge to exact figures of prey numbers and their weight. For the service provider module of the best available data on both macrozoobenthos and predictors it would be advisable, for instance, to add other environmental parameters such as organic content and nutrient supply, which could obviously enhance the quality and applicability of models (Gogina and Zettler, 2010). Furthermore, accuracy assessments have stressed that the different quartiles of a predictor range may be unevenly justified by macrofauna data, so the sampling strategy should take into account the spatial peculiarities of important predictors, especially that part of a range where significant changes in the characteristics of macrofauna occur.

It is important that the method may have many other applications. The data in the user module (in this case the feeding of cod, flounder and eelpout) could be easily replaced by different objects like the feeding of other fish species or even birds. On the other hand, it could serve not only to map feeding grounds, but also other types of services or assessments, for instance, habitat sensitivity, implemented by Hiscock and Tyler-Walters (2006). Finally, if biomass data were replaced with abundance of macrozoobenthos in the provider module, the method could be used, e.g. to assess seabed quality according to the Benthic Quality Index introduced by Rosenberg et al. (2004).

The application of biological valorisation methodology provides benthic habitat quality maps from a different perspective. The approach is based on the assessment of individual species and habitat/species assemblages and considers the value of an area in terms of its resilience and the stability of species and species assemblages, and not from the anthropocentric (goods and services) point of view (Węsławski et al., 2009).

Based on the application of biological valorisation approach the most valuable areas of LMA are a northern coastal areas where stony bottoms vegetated by macrophytes are present (Olenin et al., 1996; Olenin, Daunys, 2004; Bučas et al., 2009). As illustrated in Figure 27, the highest values of biological valorisation well correspond with the modelled coverage of the red algae *F. lumbricalis*. So it is fair to state, that the community of the habitat engineer red algae both in terms of species richness and the biological value of species is the most valuable in LMA.

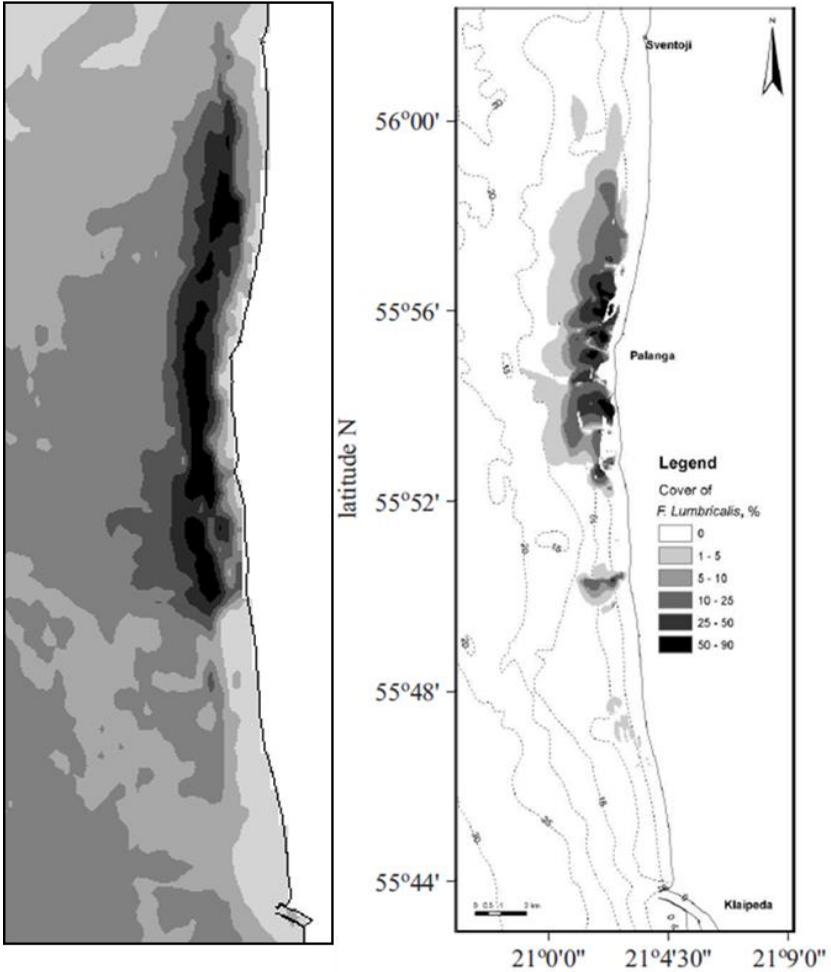


Figure 27. Biological valorisation of the northern coast of Lithuanian Baltic Sea (left) and the modelled coverage of red algae *Furcellaria lumbricalis* (right) by Bučas et al., 2009.

It is fair to notice, that both assessments (quality of fish feeding grounds and biological valorisation) slightly underestimate hard

substrate areas in the offshore of LMA, since no field sampling was performed due to the restrictions for SCUBA diving. This suggests that the integration of remote methods is needed to cover areas inaccessible for divers or grab sampling.

Application of species distribution models in mapping bioinvasions is another example of benefits from the predictive modelling. It is well known that invasive engineers in aquatic ecosystems are of a particular scientific and managerial concern, since many of them are extremely successful, become very abundant in a short time and have multiple effects on the ecosystem (Dame, 1993, 1996; Karatayev et al., 2002; Burlakova et al., 2012; Zaiko et al., 2013).

In the case of *D. polymorpha* in the lake Drūkšiai, over more than 30 years since the invasion, zebra mussel drastically shifted benthic communities in the ecosystem. According to this study, in total, over 60 billion of individuals weighting more than 45 thousand tons was present in the area during 2008-2010. These estimations are significantly higher than in previous studies. In 1985 the estimated biomass of zebra mussels in the lake was approximately 1300 tons and in 1989 it reached 5600 tons (Grigelis, 1993). However, it is hard to state if these differences are determined by different methodologies or by time gap separating studies.

As presented in Zaiko et al. (2013), there is a significant linear relationship between the biomass of *D. polymorpha* and the abundance of other benthic invertebrates (Figure 28). Furthermore, many species were found only in presence of zebra mussels (*e.g.* most of the gastropods, leeches, crustaceans, and insects), and abundances of those present also in the bare sediments were remarkably lower on average.

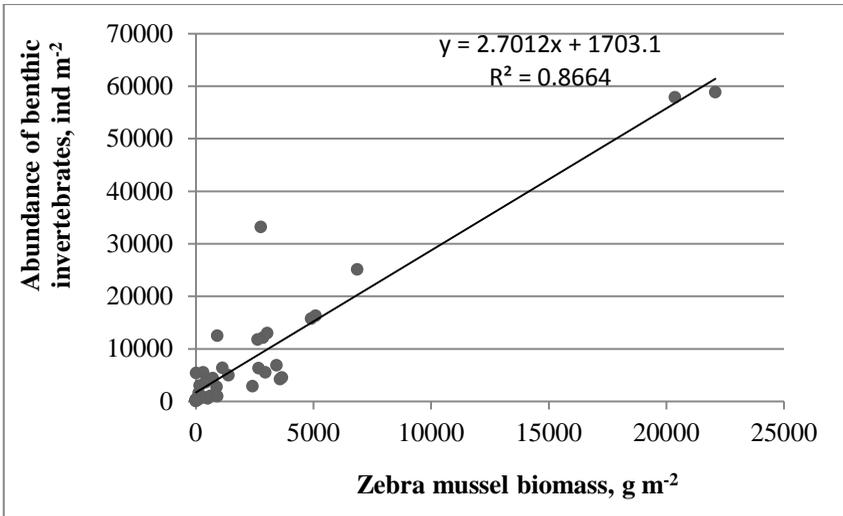


Figure 28. The relationship between the biomass of zebra mussel and the abundance of other benthic invertebrates in the lake Drūkšiai (after Zaiko et al., 2013).

A massive expansion of the zebra mussel in the shallow areas of the lake significantly altered benthic habitats. Relatively flat soft bottom habitats are shifted by mussel clumps and shell deposits that form 3 dimensional reef-like structures. These peculiar structures elevate the abundance of both epifaunal and infaunal invertebrates by providing substrate, shelter and food or even enhance the aeration of sediments by promoting microcirculation flows (Zaiko et al., 2009). On the other hand, it is known that the zebra mussel can drastically decrease the population size of the local unionid species (Schloesser and Nalepa, 1994; Ricciardi et al., 1996). Similarly, the decrease in abundance of unionids was observed in the lake Drūkšiai few years later after the occurrence of *D. polymorpha* (Grigelis, 1993), nevertheless unionids are still present in recent times (Baršienė and Rybakovas, 2008), however no information of its distribution in spatial scale is available.

To ensure the robustness of the modelling results of the distribution of *D. polymorpha* in the lake Drūkšiai, ground-truthing in some particular areas is recommended for the future studies. In this study the Belorussian part of the lake was not sampled, where high densities of mussels were modelled. Also there were a number of not-sampled shallows in the central part of the lake, identified by the model as zebra mussel beds. These results should be verified by additional surveys.

Bioinvasion assessments, such as biopollution assessment (Olenin et al., 2007) in the scale of whole ecosystem require spatial information on the distribution of the invader. Relying only on grab or SCUBA diver sampling, that would require very intensive and costly sampling, since the area of the lake is about 49 km² and the coastline of more than 70 km. In these circumstances predictive modelling provides many advantages, especially if the model can be constructed on remote video survey data. Furthermore, the same approach could be applied for the hard substrate offshore habitats in LMA for the assessment of blue mussel beds.

5.3 Future perspectives

The increase of spatial data on environmental parameters with the development of remote sensing equipment and advanced modelling of abiotic factors in recent years has opened new opportunities for benthologists.

For areas covered by multibeam and side-scan sonars a high resolution data on many environmental characteristics can be extracted, for instance, depth, sediment types, roughness and slope of the seabed and others (Gulbinskas et al., 2009; Lamarche et al., 2011). Large Baltic scale projects such as BALANCE provide open source results for other important parameters (i.e. near-bottom oxygen concentration, current velocity) shaping the distribution on benthic species. Another important factor – wind induced exposure to the bottom sediments can be generated by wave exposure models such WEMo (Malhotra and Fonseca, 2007) based only on the depth and

average wind velocities. All this spatial information is essential for the predictions of the distribution of benthic macrofauna or other organisms and it is getting more and more accessible for scientists.

Finally, the integration of species distribution modelling methods (such as random forests, DOMAIN, MaxEnt) into geographic information systems (element distribution modelling tool in ArcGIS software) and especially to open source software like R, Quantum GIS, openModeller and others (Ghisla et al., 2012) probably indicates that it is only a matter of time before empirical modelling will become a casual spatial analysis for ecologists.

CONCLUSIONS

- 1) Among the empirical models compared, the RF method showed the best results in predicting the occurrence and biomass distribution of benthic invertebrates. Predictive performance of GAMs and MARS followed RF, whereas MaxEnt accurately predicted occurrence only for the species with relatively low distribution range.
- 2) The proposed system, integrating empirical modelling of macrozoobenthos distribution and fish diet composition, allows mapping the benthic habitat quality in terms of provision of food for fishes. The highest quality feeding grounds for cod, flounder and eelpout are determined in the coastal stony bottom area, the soft bottom habitats in the central and mixed substrates in the western parts of the LMA. The total estimated area of these habitats comprises 7.6 % of the LMA.
- 3) The most valuable habitat within the Lithuanian marine area according to the biological valorisation approach based on biological values of species, species richness and the spatial extent of a habitat is the coastal stony bottom vegetated by red algae *Furcellaria lumbricalis*.
- 4) The empirical modelling based on video survey data can be used for the assessment of the distribution range of the invasive bivalve *Dreissena polymorpha*. It was estimated that the invasion caused benthic habitat quality alteration in ca. 24 % of the lake Drūkšiai (10.6 km²). The most altered habitats (with the coverage of zebra mussel being 41.1 ± 14.3 %) are located in the highly exposed areas within the depth range of 3-7 m.

REFERENCES

- Araujo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22: 42–47.
- Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *ecological modelling*, 200(1), 1-19.
- Baršienė, J., Rybakovas, A., 2008. Cytogenetic damage in gill and gonad cells of bivalve molluscs. *Ekologija* 54(4): 245–250.
- Bean, W., Stafford, R., Brashares, J., 2011. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, 34: 001–009.
- Bendtsen, J., Söderkvist, J., Dahl, K., Hansen, J.L.S., Reker, J., 2007. Model simulations of blue corridors in the Baltic Sea, BALANCE Interim Report no. 9.
- Bitinas, A., Aleksa, P., Damušytė, A., Gulbinskas, S., Jarmalavičius, D., Kuzavinis, M., Minkevičius, V., Pupienis, D., Trimonis, E., Šečkus, R., Žaromskis, R., Žilinskas, G., 2004, Baltijos jūros Lietuvos krantų geologinis atlasas, Lietuvos geologijos tarnyba, Vilnius, 95 pp.
- Björnsson, B., 1993. Swimming Speed and Swimming Metabolism of Atlantic Cod (*Gadus morhua*) in Relation to Available Food: A Laboratory Study, *Canadian Journal of Fisheries and Aquatic Sciences*, 50 (12), 2542-2551.
- Booij, N., Ris, R.C., Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions. 1. Model description and validation, *J. Geophys. Res.*, 104 (C4), 7649–7666.

- Borja, Á., Dauer, D. M., Grémare, A., 2012. The importance of setting targets and reference conditions in assessing marine ecosystem quality. *Ecological Indicators*, 12(1), 1-7.
- Boyd, J., Banzhaf, S., 2007. What are ecosystem services? The need for standardized environmental accounting units. *Ecological Economics*, 63(2), 616-626.
- Breiman, L., 2001. Random forests *Machine Learning* 45: 5-32.
- Breiman, L., 1996. Bagging predictors. *Machine Learning*, 24, pp. 123–140.
- Bromley, P.J., 1994. The role of gastric evacuation experiments in quantifying the feeding rates of predatory fish, *Reviews in Fish Biology and Fisheries*, 4, 36-66.
- Bučas, M., Daunys, D., Olenin, S., 2009. Recent distribution and stock assessment of the red alga *Furcellaria lumbricalis* on an exposed Baltic Sea coast: combined use of field survey and modelling methods, *Oceanologia*, 51 (3), 1-19.
- Bučas, M., 2009. Distribution patterns and ecological role of the red alga *Furcellaria lumbricalis* (Hudson) J.V. Lamouroux off the exposed Baltic Sea coast of Lithuania. Doctoral degree theses of Ph.D. in ecology and environmental studies, Klaipeda University. Klaipeda, 124 p.
- Burlakova, L.E, Karatayev, A.Y, Karatayev, V.A., 2012. Invasive mussels induce community changes by increasing habitat complexity. *Hydrobiologia* 685: 121-134.
- Chan, K.M., Ruckelshaus, M., 2010. Characterizing changes in marine ecosystem services. *F1000 biology reports*, 2.
- Collin, A., Archambault, P., Long, B., 2011. Predicting Species Diversity of Benthic Communities within Turbid Nearshore Using Full-Waveform Bathymetric LiDAR and Machine Learners, *PLoS ONE* 6 (6), e21265.
- Cutler et al., 2007. Random forests for classification in ecology. *Ecology* 88:2783-2792.

- Dame, R.F., 1993. The role of bivalve filter feeder material fluxes in estuarine ecosystems. In Dame RF (ed) *Bivalve filter feeders in estuarine and coastal ecosystem processes*. Springer-Verlag, Berlin.
- Dame, R.F., 1996. *Ecology of marine bivalves. An ecosystem approach*. CRC Marine Science Series, Boca Raton.
- Degraer, S., Verfaillie, E., Willems, W., Adriaens, E., Vincx, M., Van Lancker, V., 2008. Habitat suitability modelling as a mapping tool for macrobenthic communities: an example from the Belgian part of the North Sea. *Continental Shelf Research*, 28(3), 369-379.
- Derous, S., Agardy, T., Hillewaert, H., Hostens, K., Jamieson, G., Lieberknecht, L. et al., 2007. A concept for biological valuation in the marine environment. *Oceanologia*, 49 (1), pp. 99–128.
- Dippner, J.W., Vuorinen, I., Daunys, D., Flinkman, J., Halkka, A., Köster, F.W., Lehikoinen, W., MacKenzie, B.R., Möllmann, C., Møhlenberg, F., Olenin, S., Schiedek, D., Skov, H., Wasmund, N., 2008. Climate related marine ecosystem change. The BACC Author Team, *Assessment of Climate Change for the Baltic Sea Basin*, Springer Verlag, Berlin, 309–377 pp.
- Downie, A., von Numers, M., Boström, C., 2013. Influence of model selection on the predicted distribution of the seagrass *Zostera marina*. *Estuarine, Coastal and Shelf Science* 121-122:8-19.
- Duarte, C.M., 2000. Marine biodiversity and ecosystem services: an elusive link. *Journal of experimental marine Biology and Ecology*, 250(1), 117-131.
- EC, 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). *Official Journal of the European Union* L164, 19.

- Ekman, L., S., 1953. Zoogeography of the Sea. Sidgwick and Jackson Ltd. London. 417 pp.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.
- Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43-57.
- Ellis, J., Ysebaert, T., Hume, T., Norkko, A., Bult, T., Herman, P., Thrush, S., Oldman, J., 2006. Predicting macrofaunal species distributions in estuarine gradients using logistic regression and classification systems, *Mar. Ecol. Prog. Ser.*, 316, 69–83.
- Friedman, J.H., 1991. Multivariate adaptive regression splines. *The Annals of Statistics*, 19: 1–67.
- Friedman, J.H., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: A statistical view of boosting *Annals of Statistics*, 28, pp. 337–407.
- Gelumbauskaitė, L.Ž., Grigelis, A., Cato, I., Repečka, M., Kjellin, B., 1999. Bottom sediment maps of the central Baltic Sea, Scale 1:500,000. A short description, LGT Series of Marine Geological Maps No. 1, SGU Series of Geological Maps Ba No. 54, Vilnius-Uppsala.
- Gelumbauskaitė, L.Ž., 2009. Character of sea level changes in the subsiding South-Eastern Baltic Sea during Late Quaternary. *Baltica* 22 (1), 23–36.
- Ghisla, A., Rocchini, D., Neteler, M., Frster, M., Kleinschmit, B., 2012. Species distribution modelling and open source GIS:

why are they still so loosely connected? In: Seppelt, R., Voinov, A. A., Lange, S., Bankamp, D. (Eds.), International Environmental Modelling and Software Society (iEMSs) 2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting. pp. 1481-1488.

- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294-300.
- Glockzin, M., Gogina, M., Zettler, M.L., 2009. Beyond salty reins—modeling benthic species’ spatial response to their physical environment in the Pomeranian Bay (southern Baltic Sea). *Baltic Coast. Zone*, 13(3), 79-95.
- Gogina, M., Zettler, M.L., 2010. Diversity and distribution of benthic macrofauna in the Baltic Sea. Data inventory and its use for species distribution modelling and prediction. *Journal of Sea Research*, 64: 313–321.
- Gogina, M.A., Glockzin, M., Zettler, M.L., 2010. Distribution of benthic macrofaunal coenoses in the western Baltic Sea with regard to near-bottom environmental parameters. 2. Modelling and prediction. *J. Mar. Syst.* 80: 1–2.
- Gray, J.S., Elliott, M., 2009. *Ecology of marine sediments: from science to management*. OUP Oxford.
- Grenouillet, G., Buisson, L., Casajus, N. Lek, S., 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34: 9-17.
- Grigelis, A., 1993. Species composition abundance and biomass of zoobenthos. *Thermal Power Generation and Environment* 10 (2): 105-109 [in Russian].
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Modell.* 135: 147-186.

- Gulbinskas, S., Trimonis, E., Blažauskas, N., Michelevičius, D., 2009. Sandy deposits study offshore Lithuania, SE Baltic Sea. *Baltica*, 22(1), 1-9.
- Hansen, I.S., Keul, N., Sorensen, J.T., Erichsen, A., Andersen J.H., 2007. Baltic Sea Oxygen Maps, BALANCE Interim Report no. 17.
- Hansen, M.H., Kooperberg, C., 2002. Spline Adaptation in Extended Linear Models (with comments and a rejoinder by the authors. *Statistical Science*, 17(1), 2-51.
- Hastie, T., Tibshirani, R., 1990. Generalized additive models, Chapman and Hall/CRC, New York, p. 335,
- He, P., 1991. Swimming endurance of the Atlantic cod, *Gadus morhua* L., at low temperatures, *Fisheries research*, 12 (1), 65-73.
- HELCOM, 1984. Guidelines for the Baltic Monitoring Programme. Helsinki.
- HELCOM, 1988. Guidelines for the Baltic monitoring programme for the third stage. Part D, Biological determinants, pp 23-28.
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29: 773–785.
- Hiscock, K., Tyler-Walters H., 2006. Assessing the sensitivity of seabed species and biotopes - the Marine Life Information Network (MarLIN), *Hydrobiologia*, 555, 309-320.
- Hiscock, K., Tyler-Walters, H., 2003. Assessing the sensitivity of seabed biotopes to human activities and natural events. Marine Life Information Network: Biology and Sensitivity Key Information Sub-Programme, Marine Biological Association of the United Kingdom, Plymouth (2003) Available from: http://www.marlin.ac.uk/PDF/Biotope_sens_brochure.pdf.

- Hosmer, D.W., Lemeshow, S., 2000. Model-Building Strategies and Methods for Logistic Regression. Applied Logistic Regression, Second Edition, 91-142.
- Järvekülg, A., 1979. Bottom fauna of the Eastern part of the Baltic Sea. (Donnaja fauna vosochnoj chasti Baltijskogo morja), Valgus, Tallinn (in Russian), 324 pp.
- Jurgelevičienė, L., Lasinskas, M., Tautvydas, A., 1983. Hydrography of Lake Drūkšiai. Mokslas, Vilnius [in Lithuanian].
- Karatayev, A.Y., Burlakova, L.E., Padilla, D.K., 2002. Impacts of zebra mussels on aquatic communities and their role as ecosystem engineers. In: Leppäkoski E, Gollasch S, Olenin S (eds) Invasive aquatic species of Europe — distribution, impact and management. Kluwer, Dordrecht, pp 433–446.
- Knudby, A., LeDrew, E., Brenning, A., 2010. Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. Remote Sensing of Environment, 114(6), 1230-1241.
- Kremen, C., Ostfeld, R.S., 2005. A call to ecologists: measuring, analyzing, and managing ecosystem services, Front. Ecol. Environ., 3 (10), 540–548.
- Lamarche, G., Lurton, X., Verdier, A. L., Augustin, J. M., 2011. Quantitative characterisation of seafloor substrate and bedforms using advanced processing of multibeam backscatter—application to Cook Strait, New Zealand. Continental Shelf Research, 31(2), S93-S109.
- Leathwick, J.R., Elith, J., Hastie, T., 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological modelling, 199(2), 188-196.
- Lenoir, J., Virtanen, R., Oksanen, J., Oksanen, L., Luoto, M., Grytnes, J.A., Svenning, J.C., 2012. Dispersal ability links to

- cross-scale species diversity patterns across the Eurasian Arctic tundra. *Global Ecology and Biogeography*, 21(8), 851-860.
- Leppäkoski, E., 1975. Assessment of degree of pollution on the basis of macrozoobenthos in marine and brakish-water environments. *Acta Acad. Aabo Ser. B.*, 35(2), 1-96.
 - Li, J., Heap, A.D., 2008. A Review of Spatial Interpolation Methods for Environmental Scientists. Canberra: Geoscience Australia, 137 p. http://www.ga.gov.au/image_cache/GA12526.pdf
 - Liaw, A., Wiener, M., 2002. Classification and regression by random forest, *R News*, 2, 18–22.
 - Lithuanian State Scientific Research Programme “Ignalina Nuclear Power Plant and the environment” (1993-1997) (1998) The state and development of the ecosystems in the region of Ignalina Nuclear Power Plant, Final collection of scientific reports. Vilnius [in Lithuanian].
 - Macdonald, J.S., Waiwood, K. G., Green, R. H., 1982. Rates of Digestion of Different Prey in Atlantic Cod (*Gadus morhua*), Ocean Pout (*Macrozoarces americanus*), Winter Flounder (*Pseudopleuronectes americanus*), and American Plaice (*Hippoglossoides platessoides*), *Canadian Journal of Fisheries and Aquatic Sciences*, 39 (5), 651-659.
 - Malhotra, A., Fonseca, M.S., 2007. WEMo (Wave Exposure Model): Formulation, Procedures and Validation. NOAA Technical Memorandum NOS NCCOS 65, 28 p.
 - Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38: 921–931.
 - Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and distributions*, 15(1), 59-69.

- Maxwell, D.L., Stelzenmuller, V., Eastwood, P.D., Rogers, S.I., 2009. Modelling the spatial distribution of plaice (*Pleuronectes platessa*), sole (*Solea solea*) and thornback ray (*Raja clavata*) in UK waters for marine management and planning. *Journal of Sea Research*, 61, 258–267.
- Mažeika, J., Taminskas, J., Paškauskas, R., Bodoyan, A., Baghdassaryan, H., Tozalakyan, P., Davtyan, V., Grillot, J.C., Travi, Y., 2006. Ecohydrological evolution in the catchment of Lake Drūkšiai, Lithuania, under anthropogenic pressure. *Ekologija* 4: 40-50.
- McCullagh, P., Nelder, J.A., 1989. *Generalized linear models (Monographs on statistics and applied probability 37)*. Chapman Hall, London.
- McPherson, J.M., Jetz, W., 2007. Effects of species ecology on the accuracy of distribution models. *Ecography*, 30: 135–151.
- Meißner, K., Darr, A., Rachor, E., 2008. Development of habitat models for Nephthys species (Polychaeta: Nephthyidae) in the German Bight (North Sea). *Journal of Sea Research*, 60(4), 276-291.
- Milborrow, M.S., 2012. Package ‘earth’.
- Nedveckaitė, T., Marciulionienė, D., Mazeika, J., Paskauskas, R., 2011. Radiological and environmental effects in Ignalina Nuclear Power Plant cooling pond – Lake Druksiai: from plant put in operation to shut down period of time. In: Tsvetkov P (ed) *Nuclear power – operation, safety and environment*. INTECH, pp. 261-286.
- Olenin, S., Minchin, D., Daunys, D., 2007. Assessment of biopollution in aquatic ecosystems. *Marine Pollution Bulletin* 55(7-9): 379-394.
- Olenin, S., Alemany, F., Cardoso, A.C., Gollasch, S., Gouletquer, P., Lehtiniemi, M., McCollin, T., Minchin, D., Miossec, L., Occhipinti-Ambrogi, A., et al., 2010. *Marine Strategy Framework Directive Task Group 2 Report: Non-*

- indigenous species. JRC Scientific and Technical Reports, 0, 52p.
- Olenin, S., 1997. Benthic zonation of the Eastern Gotland Basin, *Netherlands Journal of Aquatic Ecology*, 30 (4), 265-282.
 - Olenin, S., Daunys, D., 2004. Coastal typology based on benthic biotope and community data: the Lithuanian case study. *Coastline Reports*, 4, 65-84.
 - Olenin, S., Ducrotoy, J.P., 2006. The concept of biotope in marine ecology and coastal management. *Marine Pollution Bulletin*, 53(1), 20-29.
 - Olenin, S., Daunys, D., Labanauskas, V., 1996. Classification principles of the Lithuanian coastal biotopes. *Annals of Geography (Vilnius)* 29, 218-231 (in Lithuanian).
 - Olenin, S., Daunys, D., Bučas, M., Bagdanavičiūtė, I. (editors), 2012. Environmental status of the Lithuanian Baltic Sea: preliminary assessment. Klaipėda University press, 74 pp. [In Lithuanian].
 - Pascual, M., Borja, A., Eede, S.V., Deneudt, K., Vincx, M., Galparsoro, I., Legorburu, I., 2011. Marine biological valuation mapping of the Basque continental shelf (Bay of Biscay), within the context of marine spatial planning. *Estuarine, Coastal and Shelf Science*, 95(1), 186-198.
 - Pearson, T.H., Rosenberg, R., 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and marine biology: an annual review*, 16.
 - Petersen, C.G.J., 1914. Valuation of the sea. II. The animal communities of the sea bottom and their importance for marine zoogeography. *Reports of the Danish Biological Station* 21 [Cited by Gray and Elliott, 2009].
 - Petersen, C.G.J., 1918. The sea bottom and its production of fish food. A survey of work done in connection with the valuation of the Danish waters from 1883–1917. *Reports of*

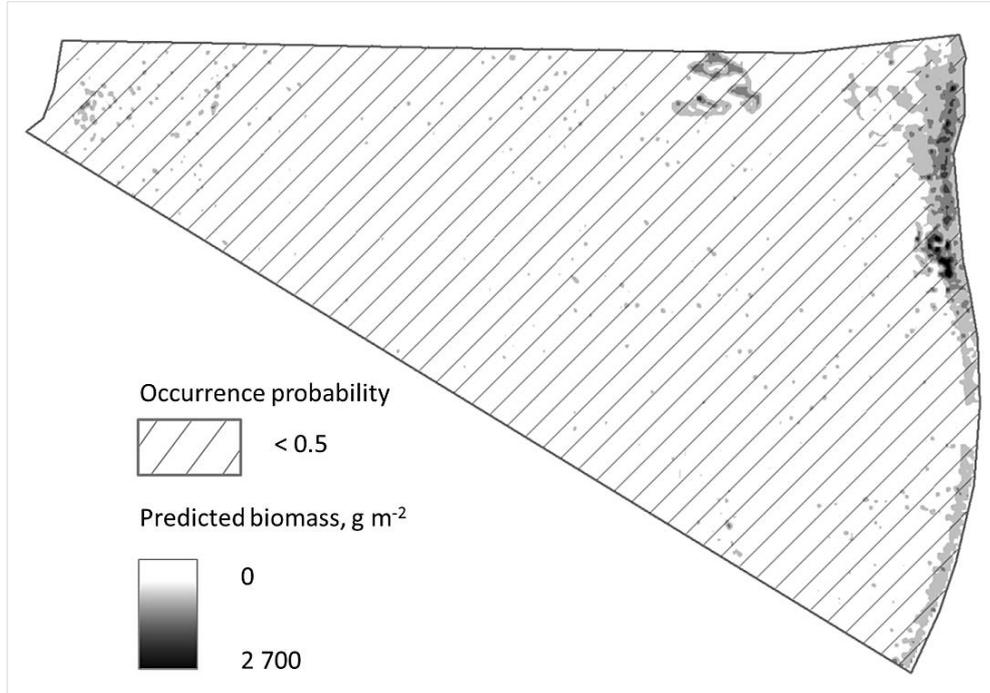
- the Danish Biological Station 25, 1–62 [Cited by Gray and Elliott, 2009].
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31(2), 161-175.
 - Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*. 190: 231–259.
 - Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* 9: 181–199.
 - Quinn, G. P., Keough, M.J., 2002. Experimental design and data analysis for biologists, Cambridge: Cambridge University Press, 520 pp.
 - Reiss, H., Cunze, S., König, K., Neumann, H., Kröncke1, I., 2011. Species distribution modelling of marine benthos: a North Sea case study. *Marine Ecol. Progr. Series* 442: 71-86.
 - Repečka, M., Gelumbauskaitė, Ž., Grigelis, A., Šimkevičius, P., Radzevičius, R., Monkevičius, A., Bubinas, A., Kasperovičienė, J., Gadeikis, S., 1997. Valstybinis jūrinis geologinis kartografavimas 1:50 000 masteliu Klaipėdos–Šventosios akvatorijoje. I objektas, Lietuvos geologijos tarnyba, Lietuvos geologijos institutas; Vilnius, 227 pp.
 - Ricciardi, A., Whoriskey, F.G., Rasmussen, J.B., 1996. Impact of the (*Dreissena*) invasion on native unionid bivalves in the upper St. Lawrence River. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(6), 1434-1444.
 - Robinson, L., Elith, J., Hobday, A., Pearson, R.G., Kendall, B.E., Possingham, H.P., Richardson, A.J., 2011. Pushing the limits in marine species distribution modeling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography* 20:789-802.

- Rönnbäck, P., Kautsky, N., Pihl, L., Troell, M., Söderqvist, T., Wennhage, H., 2007. Ecosystem goods and services from Swedish coastal habitats: Identification, valuation, and implications of ecosystem shifts. *AMBIO: A Journal of the Human Environment*, 36(7), 534-544.
- Rosenberg, R., Blomqvist, M., Nilsson, H.C., Cederwall, H., Dimming, A., 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive, *Marine Pollution Bulletin*, 49, 728–739.
- Sanchirico, J.N., Mumby, P., 2009. Mapping ecosystem functions to the valuation of ecosystem services: implications of species–habitat associations for coastal land-use decisions, *Theor. Ecol.*, 2, 67–77.
- Santos, J., Jobling, M., 1991. Factors affecting gastric evacuation in cod, *Gadus morhua* L., fed single-meals of natural prey, *Journal of Fish Biology*, 38 (5), 697-713.
- Schloesser, D.W., Nalepa, T.F., 1994. Dramatic decline of unionid bivalves in offshore waters of western Lake Erie after infestation by the zebra mussel, *Dreissena polymorpha*. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(10), 2234-2242.
- Segerstråle, S.G., 1957. Baltic Sea. In *Treatise on marine ecology and paleoecology. I. Ecology*. Edited by J.W. Hedgpeth. *Geol. Soc. Am. Mem.* 67: 751–800.
- Sundblad, G., Bergström, U., Sandström, A., Eklöv, P., 2013 Nursery habitat availability limits adult stock sizes of predatory coastal fish. – *ICES Journal of Marine Science*, doi:10.1093/icesjms/fst056.
- Thrush, S.F., Hewitt, J.E., Norkko, A., Nicholls, P.E., Funnell, G.A., Ellis, J.I., 2003. Habitat change in estuaries: predicting broad-scale responses of intertidal macrofauna to sediment mud content. *Marine Ecology Progress Series*, 263, 101-112.

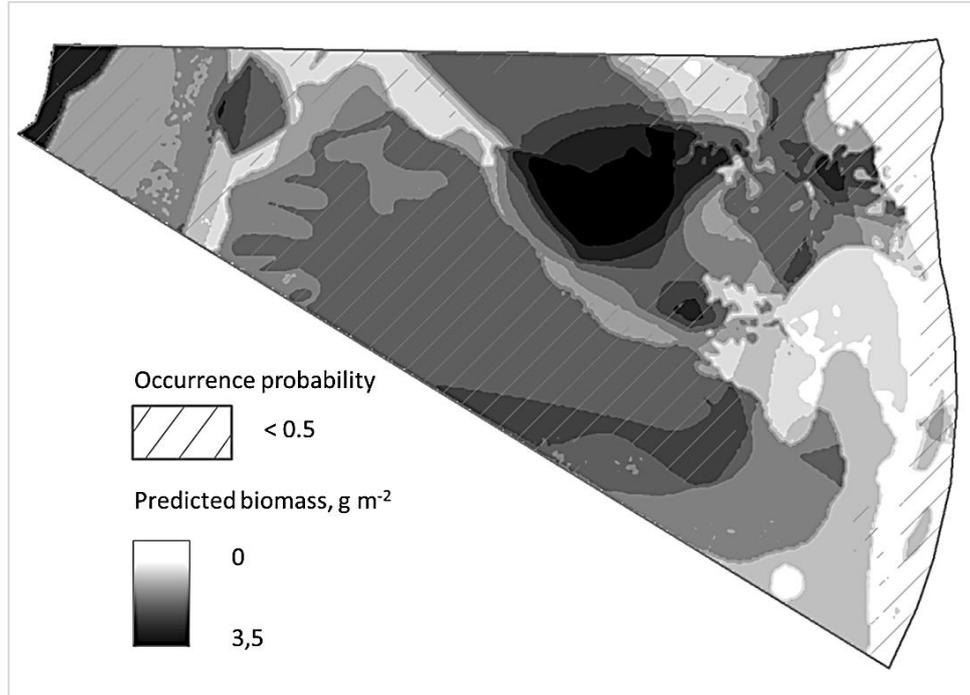
- Tillin, H.M., Rogers, S.I., Frid, C.L.J., 2008. Approaches to classifying benthic habitat quality. *Marine Policy*, 32(3), 455-464.
- Troy, A., Wilson, M.A., 2006. Mapping ecosystem services: Practical challenges and opportunities in linking GIS and value transfer, *Ecological Economics*, 60, 435–449.
- Tyler, A.V., 1970. Rates of Gastric Emptying in Young Cod, *Journal of the Fisheries Research Board of Canada*, 27 (7), 1177-1189.
- Venables, W. N., Dichmont, C.M., 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fisheries research*, 70(2), 319-337.
- Vetter, R.A.H., Franke, H.D., Buchholz, F., 1999. Habitat-related differences in the responses to oxygen deficiencies in *Idotea balthica* and *Idotea emarginata* Isopoda, Crustacea). *Journal of experimental marine biology and ecology*, 239(2), 259-272.
- Vincenzi, S., Zucchetta, M., Franzoi, P., Pellizzato, M., De Leo, G.A., Torricelli, P., 2011. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecological Modelling* 222: 1471–1478.
- Warwick, R.M., Clarke, K.R., 1991. A comparison of some methods for analysing changes in benthic community structure. *Journal of the Marine Biological Association of the United Kingdom*, 71(01), 225-244.
- Wei, C.L., Rowe, G.T., Escobar-Briones, E., Boetius, A., Soltwedel, T. et al., 2010. Global Patterns and Predictions of Seafloor Biomass Using Random Forests, *PLoS ONE* 5(12), e15323.
- Wentworth, C.K., 1922. A scale grade and class terms for clastic sediments, *Journal of Geology*, 30, 377-392.
- Węśławski, J. M., Warzocha, J., Wiktor, J., Urbanski, J., Radtke, K., Kryla, L., Tatarek, A., Kotwicki, L.,

- Piwowarczyk, J., 2009. Biological valorisation of the southern Baltic Sea (Polish Exclusive Economic Zone), *Oceanologia*, 51 (3), 415-435.
- Willems, W., Goethals, P., Van den Eynde, D., Van Hoey, G., Van Lancker, V., Verfaillie, E., Vincx, M., Degraer, S., 2008. Where is the worm? Predictive modelling of the habitat preferences of the tube-building polychaete *Lanice conchilega*. *Ecol. Model.*, 212, 74–79.
 - Wood, S.N., Augustin, N.H., 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling* 157: 157-177.
 - Wood, S.N., 2006. Generalized additive models: An introduction with R. Chapman and Hall/CRC.
 - Ysebaert, T., Meire, P., Herman, P.M., Verbeek, H., 2002. Macrobenthic species response surfaces along estuarine gradients: prediction by logistic regression. *Marine Ecology Progress Series*, 225, 79-95.
 - Zaiko, A., Daunys, D., Olenin, S., 2009. Habitat engineering by the invasive zebra mussel *Dreissena polymorpha* (Pallas) in a boreal coastal lagoon: impact on biodiversity. *Helgoland Marine Research* 63: 85-94.
 - Zenkevich, L.A., 1963. *Biology of the Seas of the USSR*. Interscience Publishers, New York. 955 pp. [Original work published in Russian in 1947].
 - Zmudzinski, L., 1978. The Evolution of Macrobenthic Deserts in the Baltic Sea. XI Conference of Baltic Oceanographers, Vol. 2, Rostock, p. 780-794.
 - Žaromskis, R., 1996. *Oceans, Seas, Estuaries*. Vilnius, 293 pp. [In Lithuanian].

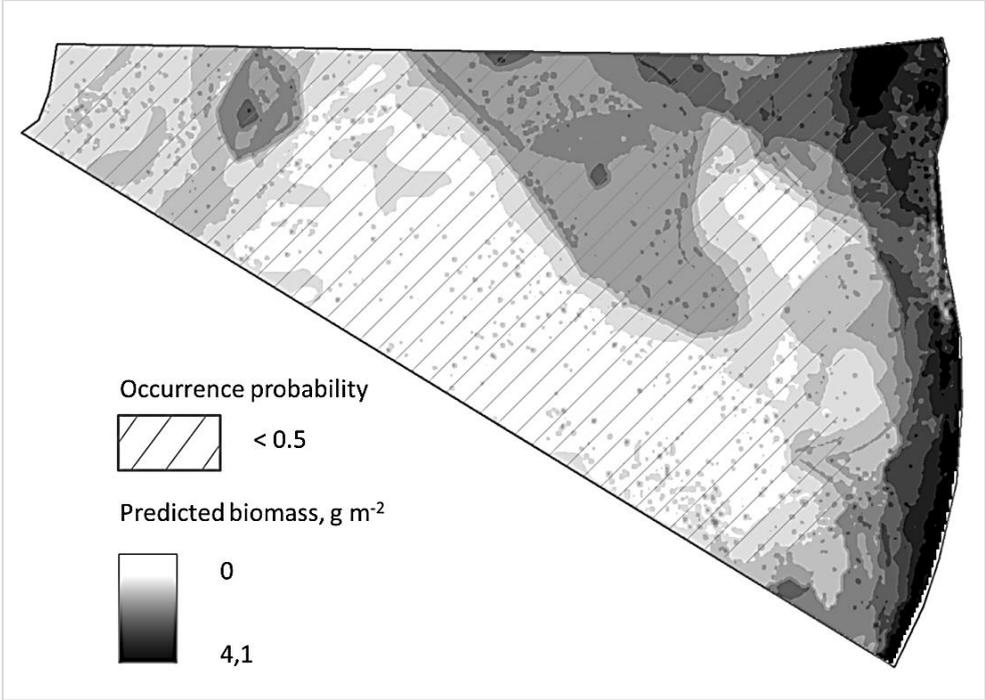
Appendix 1. Prediction map of *Balanus improvisus* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



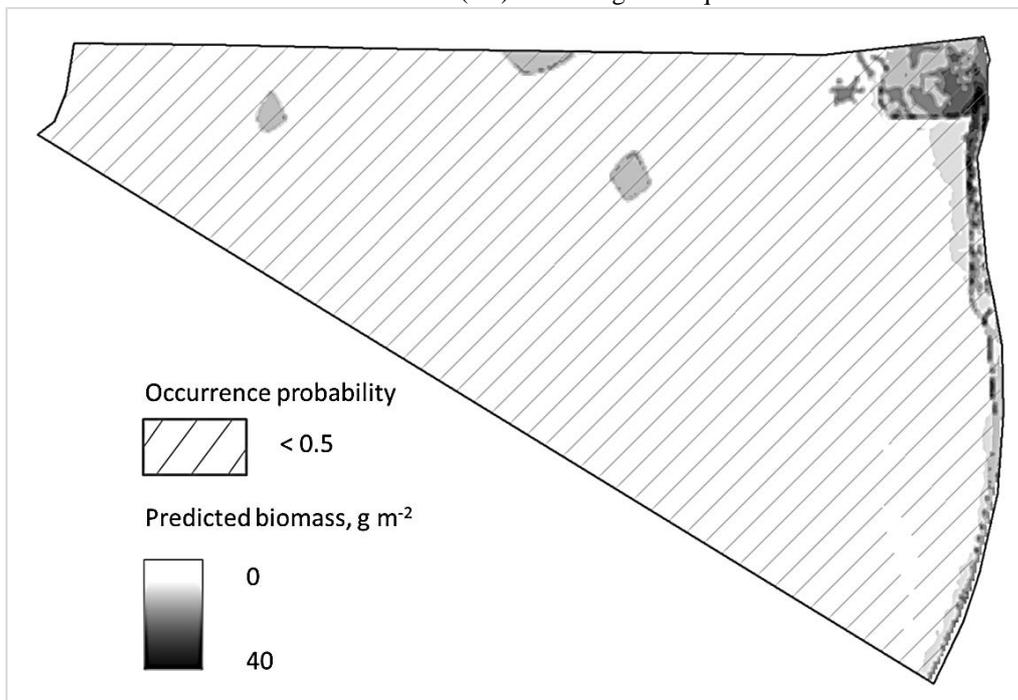
Appendix 2. Prediction map of *Halieryptus spinulosus* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



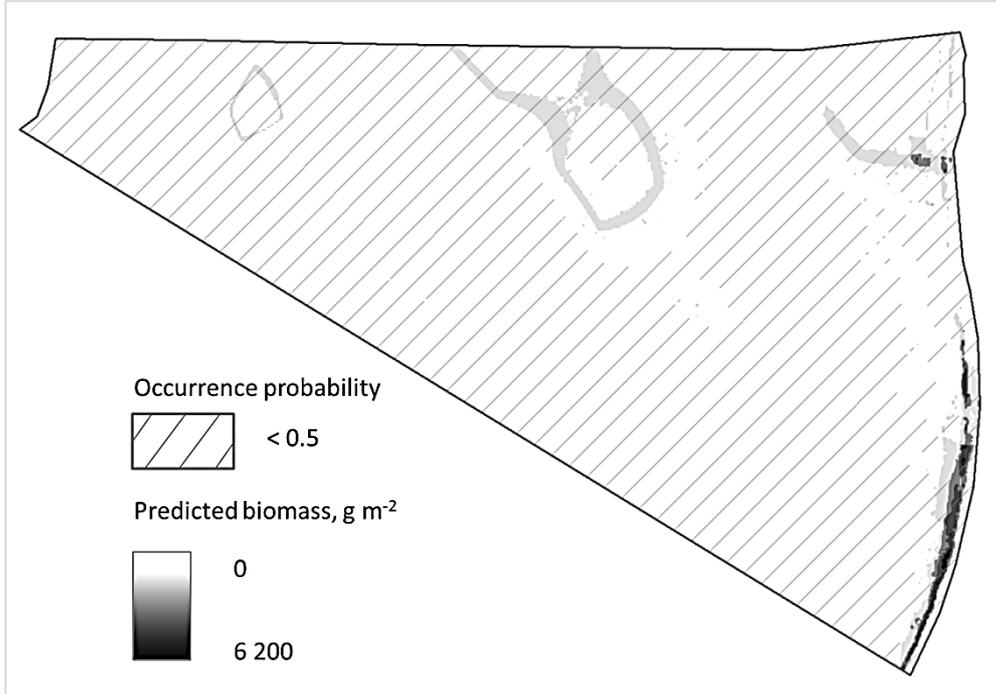
Appendix 3. Prediction map of *Hediste diversicolor* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



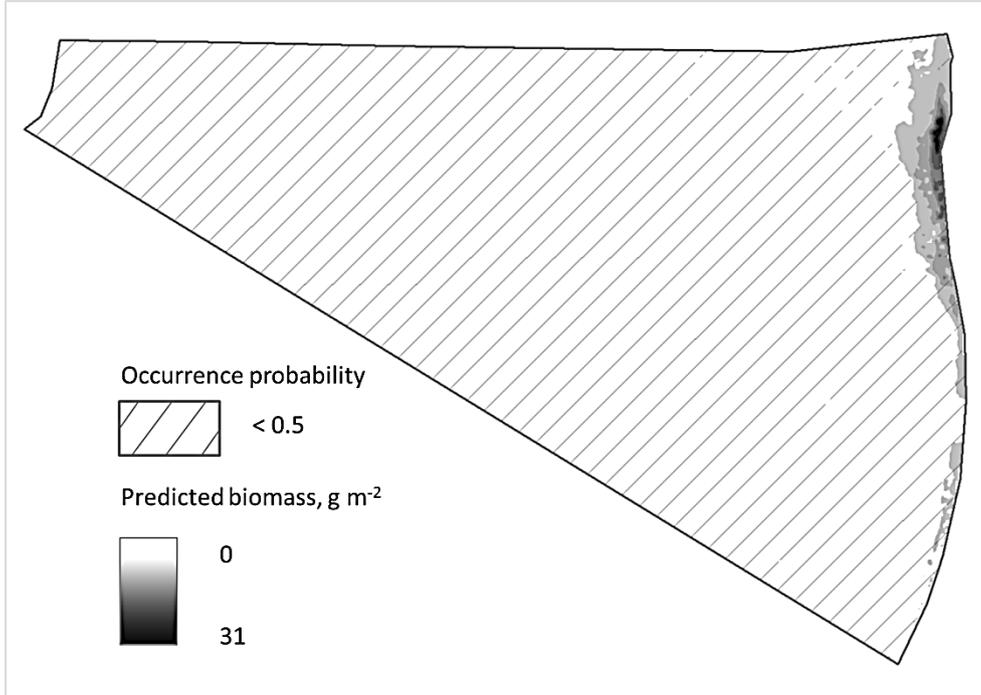
Appendix 4. Prediction map of *Bathyporeia pilosa* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



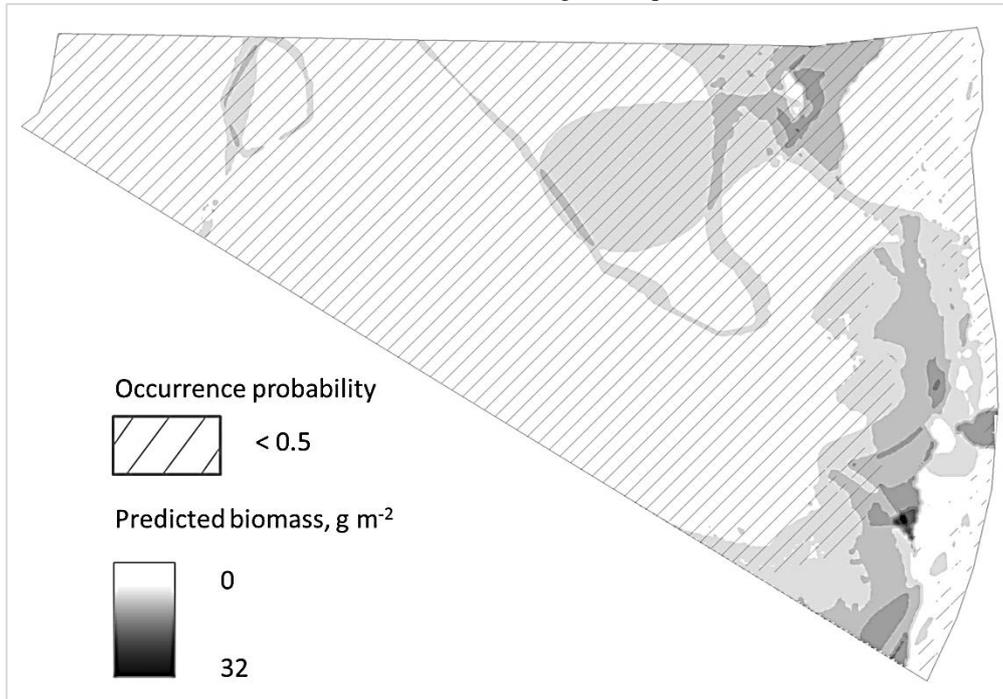
Appendix 5. Prediction map of *Cerastoderma lamarcki* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



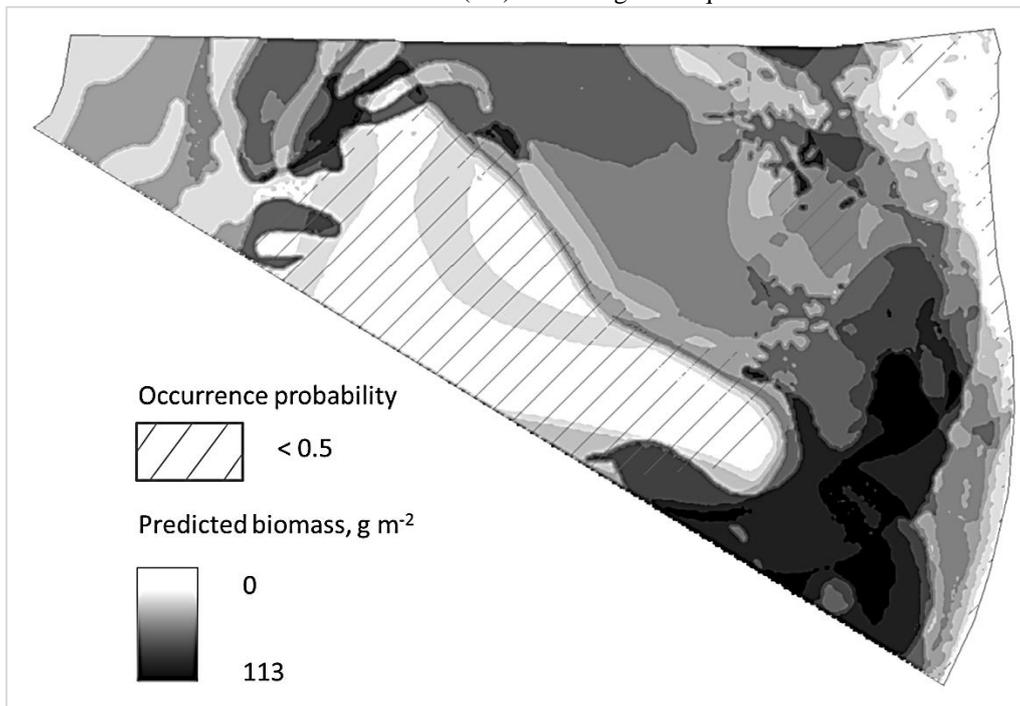
Appendix 6. Prediction map of *Gammarus spp.* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



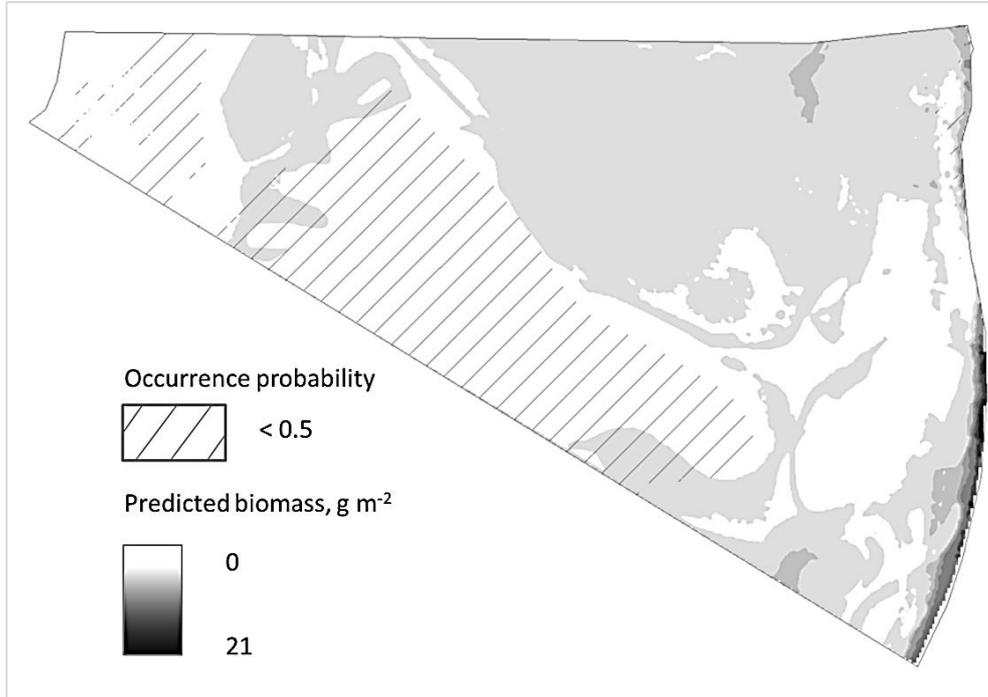
Appendix 7. Prediction map of *Mya arenaria* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



Appendix 8. Prediction map of *Macoma balthica* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.

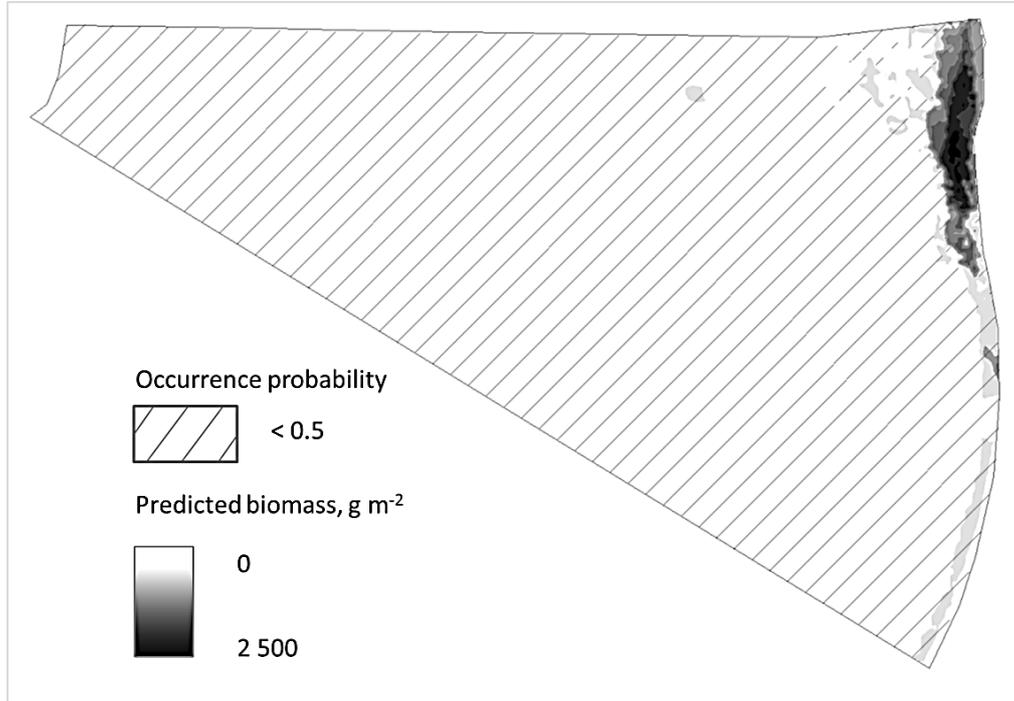


Appendix 9. Prediction map of *Marenzelleria neglecta* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.

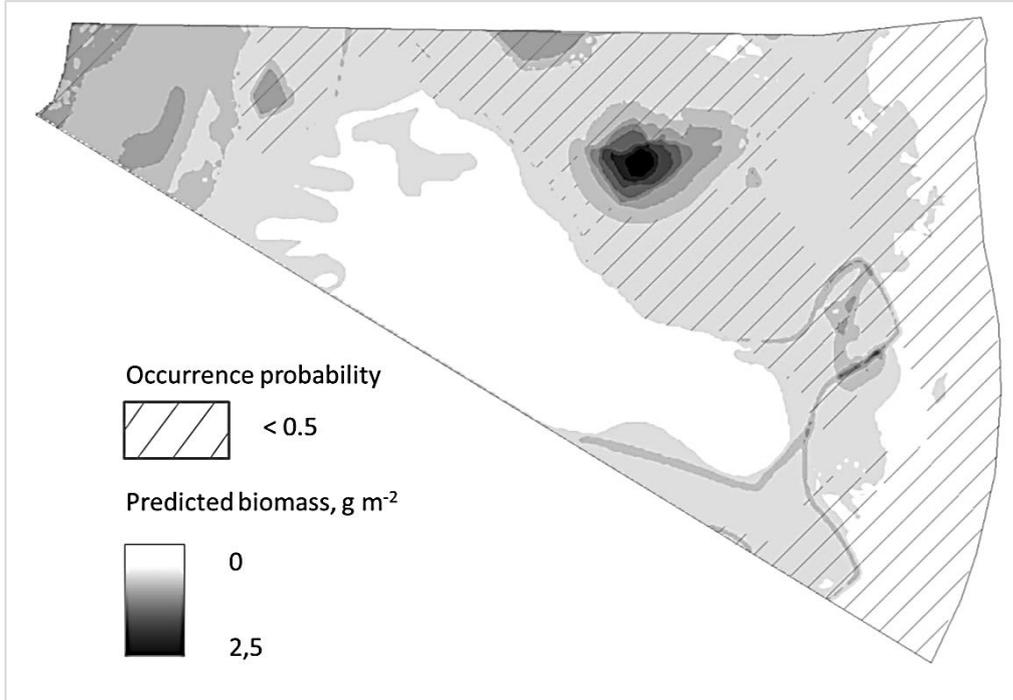


Appendix 10. Prediction map of *Mytilus trossulus* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.

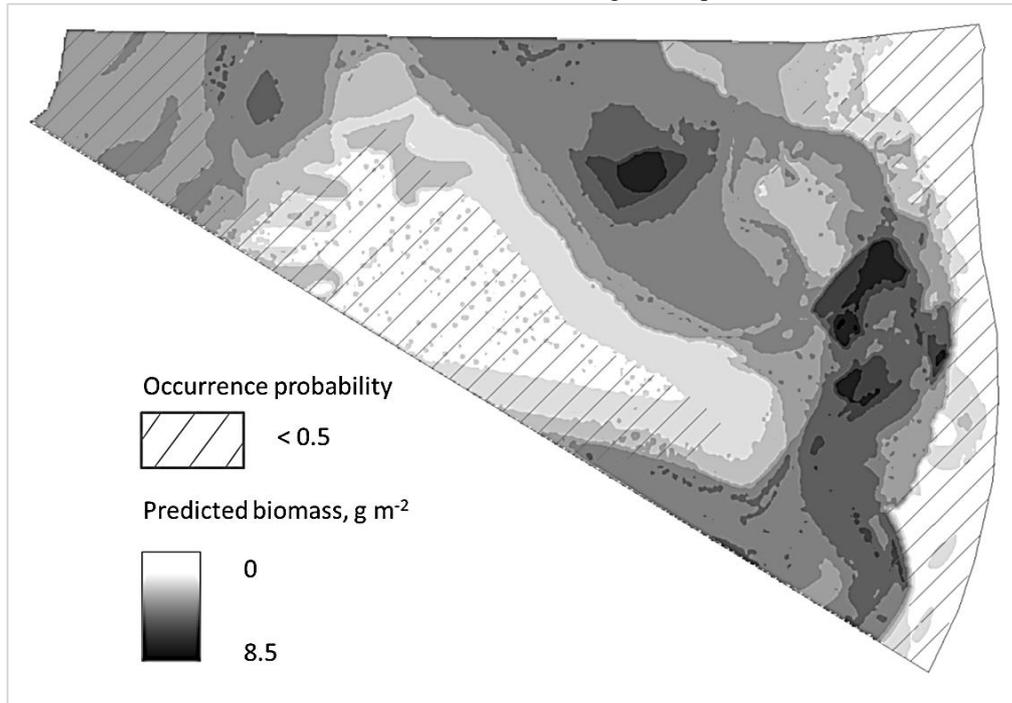
110



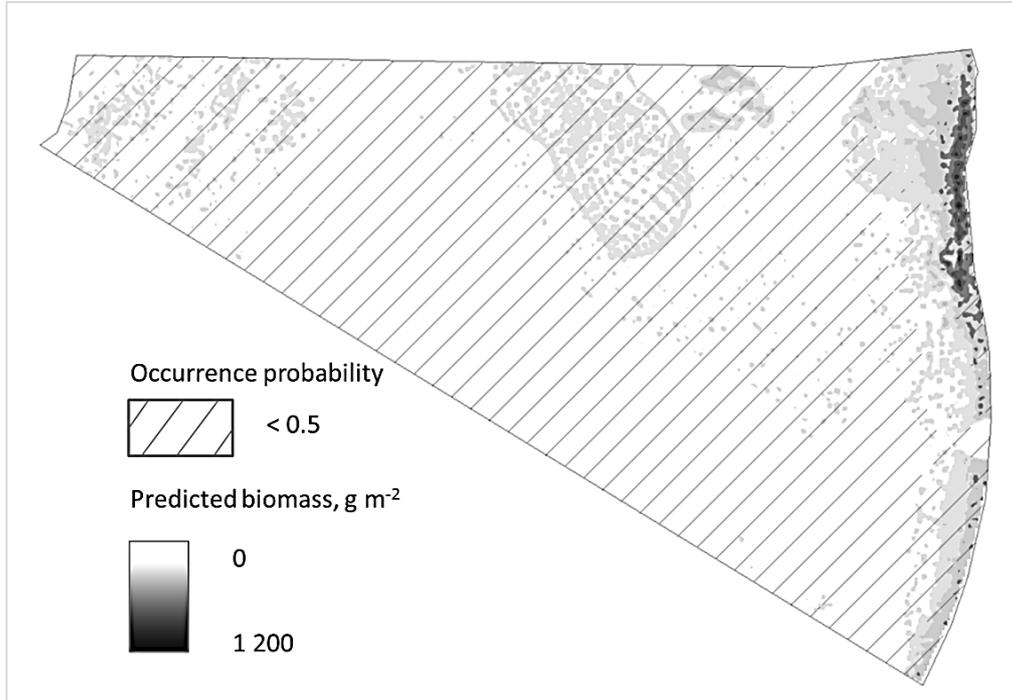
Appendix 11. Prediction map of Pontoporeiids occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



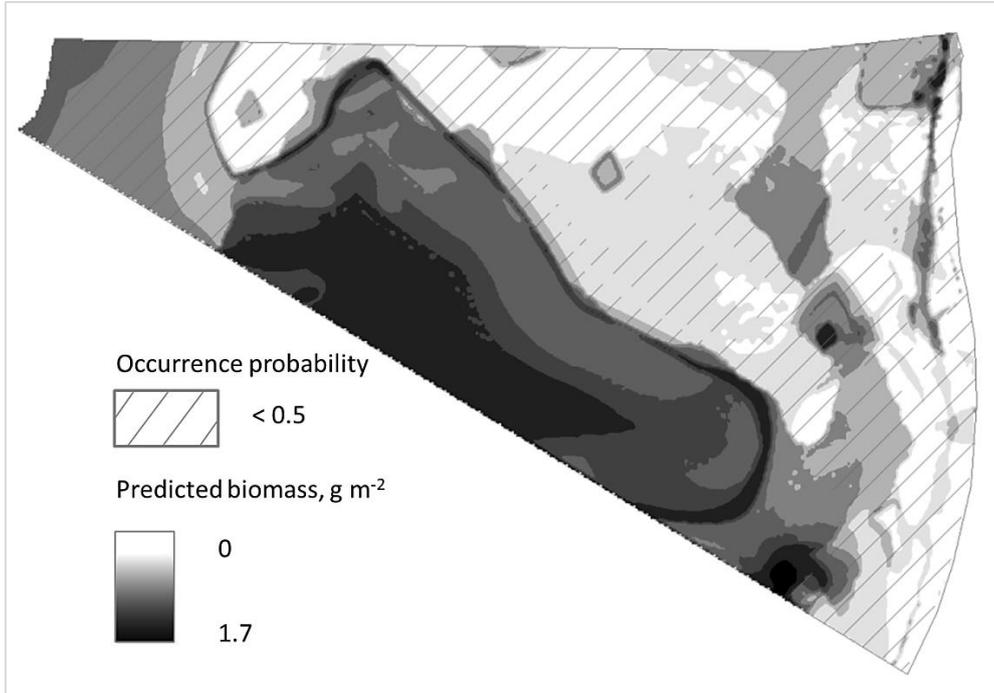
Appendix 12. Prediction map of *Saduria entomon* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



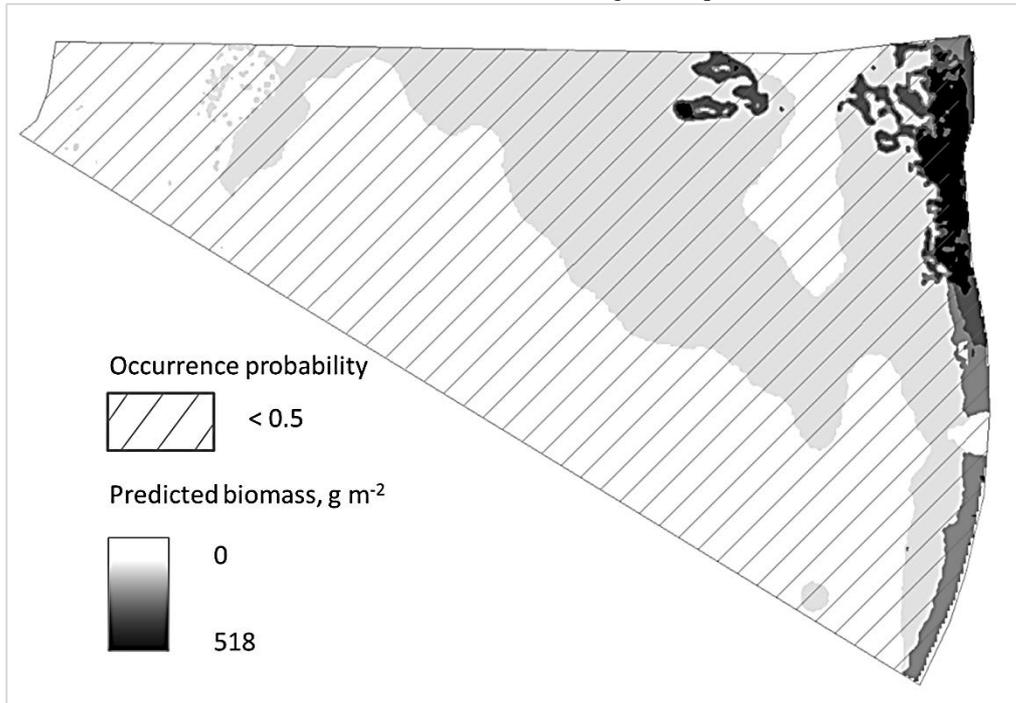
Appendix 13. Prediction map of *Corophium spp.* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



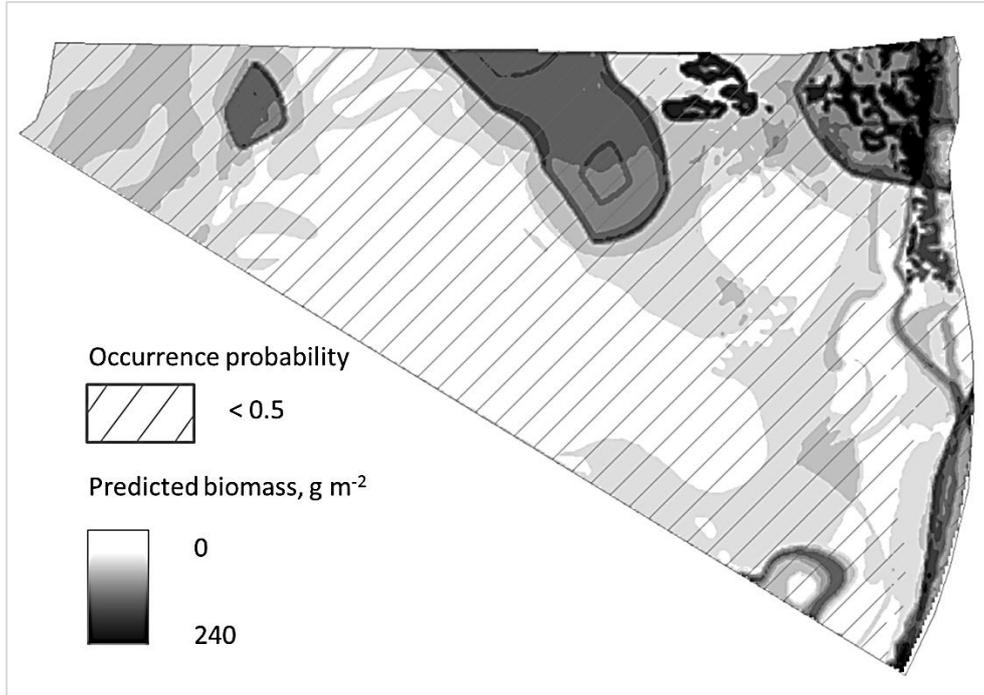
Appendix 14. Prediction map of *Bylgides sarsi* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



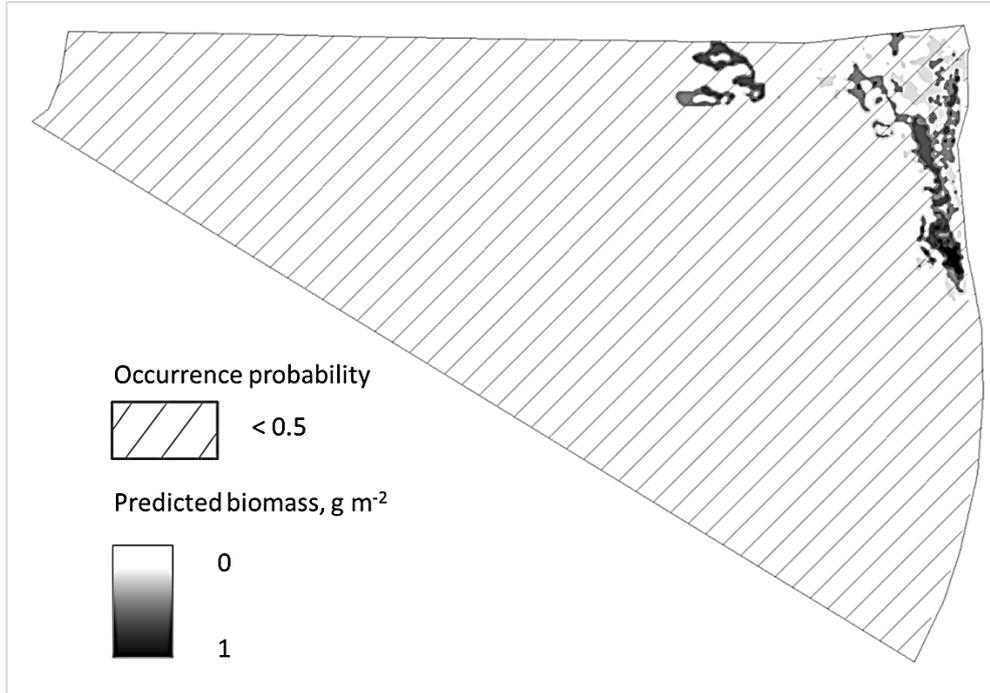
Appendix 15. Prediction map of *Fabricia sabella* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



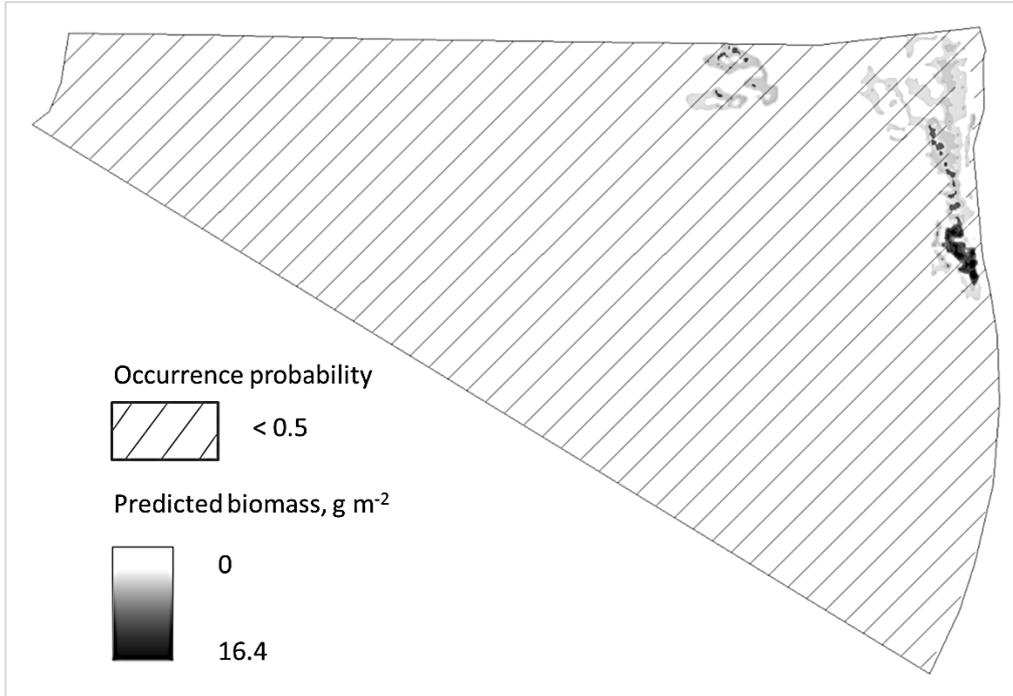
Appendix 16. Prediction map of *Hydrobia sp.* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



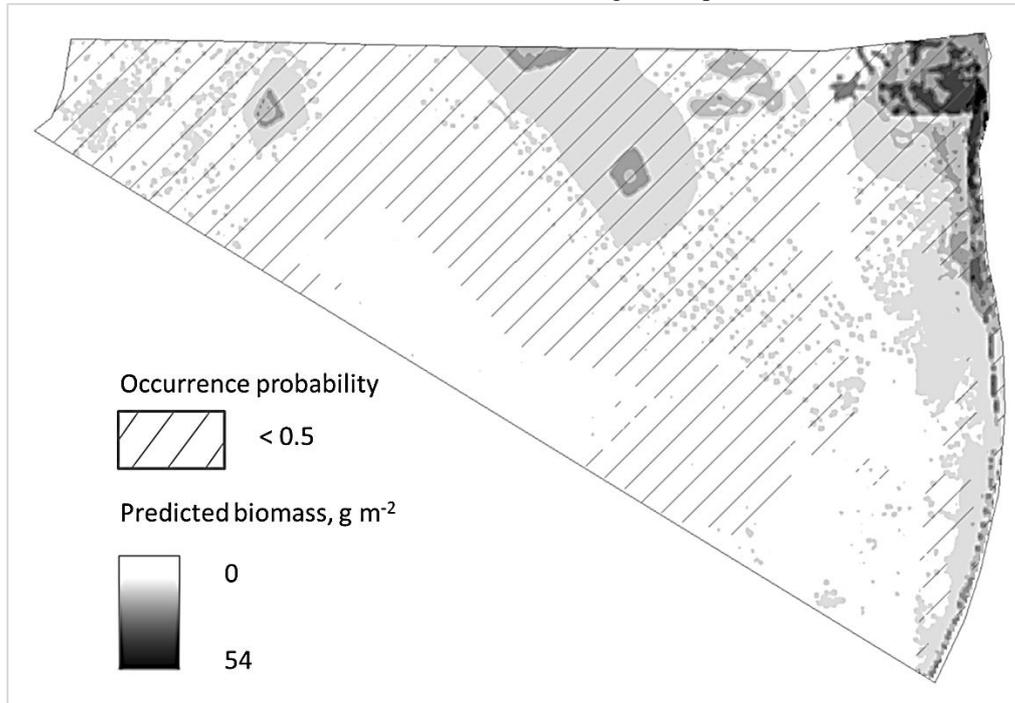
Appendix 17. Prediction map of *Idotea balthica* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



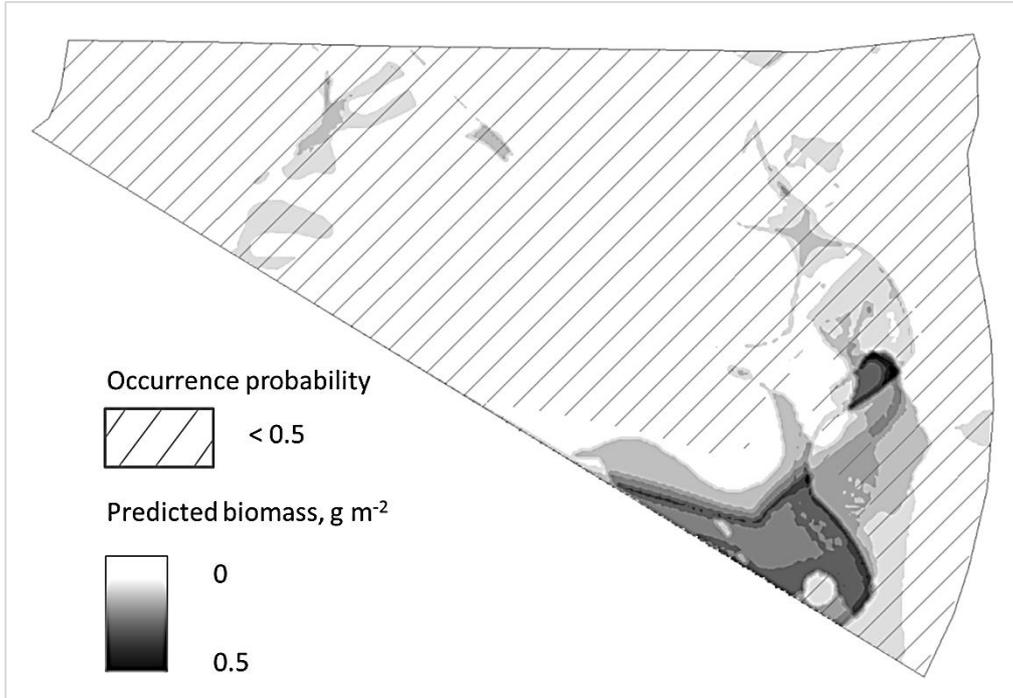
Appendix 18. Prediction map of *Jaera albifrons* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



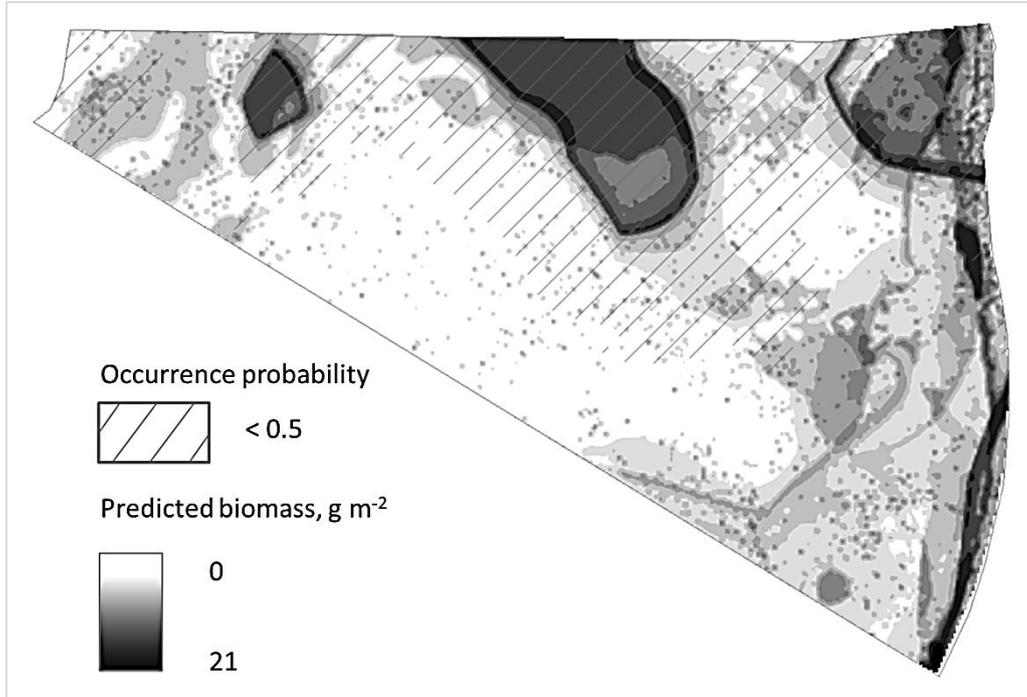
Appendix 19. Prediction map of *Oligochaeta* undet. occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



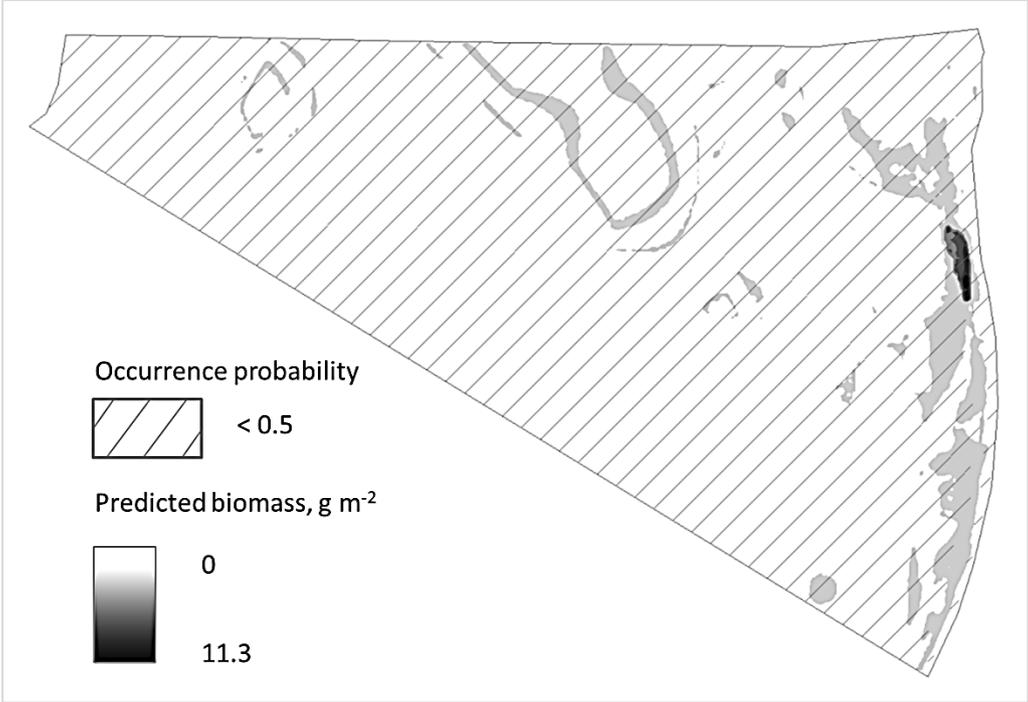
Appendix 20. Prediction map of Ostracoda undet. occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



Appendix 21. Prediction map of *Pygospio elegans* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



Appendix 22. Prediction map of *Streblospio shrubsolii* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.



Appendix 23. Prediction map of *Theodoxus fluviatilis* occurrence probability and biomass distribution modelled random forests (RF) modelling technique.

