

BRT models with optimal tree complexity are built. Modelling is done in the statistical software R 4 (R Core Team 2020) using the packages *gbm* (Ridgeway 2007) and *dismo* (Elith and Leathwick 2017) for BRT. Alternatively, random forest method can be applied using a package *randomForest* in R (Liaw and Wiener 2002).

RF is a machine learning method that generates a large number of regression trees, each calibrated on a bootstrap sample of the original data (Breiman et al., 2018). Each node is split using a subset of randomly selected predictors and the tree is grown to the largest possible extent without pruning. For predicting the value of a new data point, the data are run through each of the trees in the forest and each tree provides a value. The model prediction is then calculated as the average value over the predictions of all the trees in the forest (Breiman et al., 2018). Two parameters must be set in RF models: the number of predictor variables to be randomly selected at each node (*mtry*) and the number of trees in a forest (*ntree*). *mtry* was set to one third of the number of predictor variables as suggested by Liaw and Wiener (2002). *ntree* was set to 1000 as 500 trees usually yield stable results (Liaw and Wiener 2002).

The input data is randomly partitioned into calibration and validation datasets. The validation dataset contains data that was not included in model calibration. Calibrated models are used to predict the patterns of nature values with a grid size of 1000 meters covering the whole Gulf of Finland area. Raster layers of predictions are visually assessed to identify possible overfitting and other model- or data-driven artifacts that may not be directly reflected in mathematical validation. Based on both mathematical validation and visual expert assessment, the best performing modeling algorithm are selected. Importance of environmental predictor variables are assessed using percentage relative influence in BRT and RF.

Final modelling products include predicted values of key species, species groups and habitats per each grid cell as well as their uncertainties (standard error).

Data management and software

Data management, GIS analysis and modelling is done using MS Excel, MS Access, ArcGIS and R. Modeling is done in the statistical software R 4 (R Core Team 2020): the package *gbm* (Ridgeway 2007) and *dismo* (Elith and Leathwick 2016) are used for BRT and *randomForest* (Breiman et al 2018) for RF.

During the Adrienne project modelling workshop through Skype (18.03.2020) it was decided and agreed that the easiest and most reasonable way to share data between the partners is to use Google Drive, Excel, and web-based mapping services. To ensure the harmonization of different datasets and simplify the data management and analysis, template of data format was also uploaded on Google Drive.

Challenges and solutions related with data and modelling process

To get better overview what kind of biological and environmental data each partner country has, on what resolution, and what data is possible to add or get during the project a table that consisted such an information was circulated and filled among partner countries at the beginning of the project. Biggest challenges with data were related with the harmonizing process between different countries datasets.

During the project modelling workshop in March 2020 it was decided that cross-border modeling was not feasible at the current stage of data compilation as the number of biotic observations in the Russian waters were too low. Currently, the data compilation is in progress and more Russian macrobenthos data will be added soon when the results of extensive sampling of the summer 2020 campaign will become available.

Storage of modelling products

The map layers of the spatial prediction of key species, species groups and habitats will be published as Web Map Services and publicly displayed in the GIS Assessment Portal (Adrienne section/PlanWise4Blue portal).

References

- Antsulevich A, Välipakka P, Vaittinenc J (2003) How are the zebra mussels doing in the Gulf of Finland? Proc Estonian Acad Sci Biol Ecol 52:268-283.
- Boström C, Baden S, Bockelmann AC, Dromph K, Fredriksen S, Gustafsson C et al. (2014) Distribution, structure and function of Nordic eelgrass (*Zostera marina*) ecosystems: implications for coastal management and conservation. Aquatic Conserv: Mar Freshw Ecosyst 24:410-434.
- Breiman L, Cutler A, Liaw A, Wiener M (2018) randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4.6-14. <http://cran.r-project.org/web/packages/randomForest/>
- Darr A and Zettler ML (2000) Malakologishe abhandlungen Staatliches Museum fur Tiekunde Dresden. Band 20, Nr 19, Ausgegeben 5: 197-200.
- Eleftheriou A and McIntyre AD (2005) Methods for the Study of Marine Benthos, 3rd edn. Blackwell, Oxford.
- Elith J and Leathwick JR (2009) Species distribution models: Ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677-697.
- Elith J and Leathwick JR (2017) Boosted Regression Trees for ecological modeling. <https://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf> Accessed 30 June 2019.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151.
- Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802-813.
- Friedman JH, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting. Ann Stat 28:337-407.
- Forsström T, Fowler AE, Lindquist M, Vesakoski O (2016) The introduced dark false mussel, *Mytilopsis leucophaeata* (Conrad, 1831) has spread in the northern Baltic Sea. Bio Invasions Records. 5(2): 81-84.
- Guisan A and Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8:993-1009.
- Guisan A and Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147-186.

Herkül K, Kotta J, Kutser T, Vahtmäe E (2013) Relating remotely sensed optical variability to marine benthic biodiversity. PLoS ONE 8:e55624.

Kersen P, Kotta J, Bučas M, Kolesova N, & Değere Z (2011) Epiphytes and associated fauna on the brown alga *Fucus vesiculosus* in the Baltic and the North Seas in relation to different abiotic and biotic variables. Marine Ecology 32: 87-95.

Kautsky H, Kautsky L, Kautsky N, Kautsky U, Lindblad C (1992) Studies on the *Fucus vesiculosus* community in the Baltic Sea. Acta Phytogeogr Suec 78:33-48.

Kiirikki M and Lehvo A (1997) Life strategies of filamentous algae in the Northern Baltic Proper. Sarsia 82:259-267.

Koivisto ME and Westerbom M (2010) Habitat structure and complexity as determinants of biodiversity in blue mussel beds on sublittoral rocky shores. Mar Biol 157:1463 -1474.

Kotta J, Futter M, Kaasik A, Liversage K, Rätsep M, Barboza FR, Bergström L, et al (2020) Cleaning up seas using blue growth initiatives: Mussel farming for eutrophication control in the Baltic Sea. Science of the Total Environment 709: 136144.

Kotta J, Orav H, Kotta I (1998) Distribution and filtration activity of the zebra mussel, *Dreissena polymorpha*, in the Gulf of Riga and the Gulf of Finland. Proceedings of the Estonian Academy of Sciences, Biology Ecology 47(1): 32-41.

Kovtun A, Torn K, Martin G, Kullas T, Kotta J, Suursaar Ü (2011) Influence of abiotic environmental conditions on spatial distribution of charophytes in the coastal waters of West Estonian Archipelago, Baltic Sea. Journal of Coastal Research SI64: 412-416.

Laine AO, Mattila J, Lehtikoinen A (2006) First record of the brackish water dreissenid bivalve *Mytilopsis leucophaeata* in the Northern Baltic Sea. Aquatic Invasions. V. 1(1): 38-41.

Liaw a, Wiener M (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R. News 2/3, 18e22.

Martin G, Paalme T, Torn K (2006) Growth and production rates of loose-lying and attached forms of the red algae *Furcellaria lumbricalis* and *Coccotylus truncatus* in Kassari Bay, the West Estonian Archipelago Sea. Hydrobiologia 554:107-115.

Maximov AA, Bonsdorff E, Eremina T, Kauppi L, Norkko A, Norkko J (2015) Context-dependent consequences of *Marenzelleria* spp. (Spionidae: Polychaeta) invasion for nutrient cycling in the Northern Baltic Sea. Oceanologia 57: 341-348.

- Möller T and Martin G (2007) Distribution of the eelgrass *Zostera marina* L. in the coastal waters of Estonia, NE Baltic Sea. *Proceedings of the Estonian Academy of Sciences* 56(4): 270-277.
- Nejrup L and Pedersen M (2008) Effects of salinity and water temperature on the ecological performance of *Zostera marina*. *Aquat Bot* 88:239-246.
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Råberg S and Kautsky L (2007) A comparative biodiversity study of the associated fauna of perennial fucoids and filamentous algae. *Estuar Coast Shelf S* 73:249-258.
- Revermann R, Schmid H, Zbinden N, Spaar R, Schröder B (2012) Habitat at the mountain tops: how long can Rock Ptarmigan (*Lagopus muta helvetica*) survive rapid climate change in the Swiss Alps? A multi-scale approach. *J Ornithol* 153:891-905.
- Ridgeway G (2007) Generalized Boosted Models: A guide to the gbm package. <http://www.saedsayad.com/docs/gbm2.pdf> Accessed 30 June 2019.
- Therriault TW, Docker MF, Orlova MI, Heath DD, MacIsaac HJ (2004) Molecular resolution of Dreissenidae (Mollusca: Bivalvia) including the first report of *Mytilopsis leucophaeata* in the Black Sea basin. *Mol. Phyl. Evol.* № 30: 479-489.
- Thomsen MS, Wernberg T, Engelen AH, Tuya F, Vanderkluft MA, Holmer M, McGlathery KJ, Arenas F, Kotta J, Silliman BR (2012) A meta-analysis of seaweed impacts on seagrasses: generalities and knowledge gaps. *PloS one* 7(1):e28595.
- Torn K, Krause-Jensen D, Martin G (2006) Present and past depth distribution of bladderwrack (*Fucus vesiculosus*) in the Baltic Sea. *Aquat Bot* 84:53-62.
- Vahteri P and Vuorinen I (2016) Continued decline of the bladderwrack, *Fucus vesiculosus*, in the Archipelago Sea, northern Baltic proper. *Boreal Environ Res* 21:373-386.
- Verween A, Vinex M, Degraer S (2010) *Mytilopsis leucophaeata*: The brackish water equivalent of *Dreissena polymorpha*? A review. In G. van der Velde, S Rajagopal and A. bij de Vaate (eds.) *The Zebra Mussel in Europe*. Leiden, The Netherlands: Backhuys. P. 29-43.
- Westerbom M, Kilpi M, Mustonen O (2002) Blue mussels, *Mytilus edulis*, at the edge of the range: population structure, growth and biomass along a salinity gradient in the north-eastern Baltic Sea. *Mar Biol* 140:991-999.
- Wikström SA and Kautsky L (2007) Structure and diversity of invertebrate communities in the presence and absence of canopy-forming *Fucus vesiculosus* in the Baltic Sea. *Estuar Coast Shelf S* 72:168-176.