

Predictor selection for species distribution modeling in a marine environment

Bosch Samuel^{1,2}, Lennert Tyberghein², and Olivier De Clerck¹

¹ Phycology Research Group, Ghent University, Krijgslaan 281–S8, 9000 Ghent, Belgium

² Flanders Marine Institute (VLIZ), InnovOcean site, Wandelaarskaai 7. B–8400 Oostende
E-mail: samuel.bosch@ugent.be

Over the last few years increasing attention has been given to the development of species distribution models of marine species. New datasets with marine predictors like Bio-ORACLE (Tyberghein *et al.*, 2012) and MARSPEC (Sbrocco and Barber, 2013) have appeared but little research has been done on which and how many predictors should be selected in order to get a good species distribution model (SDM). This is an important issue as the predictor set directly impacts the performance, the transferability through space and time and the interpretation of the model. In order to find out which predictors are useful we will create models for a diverse set of marine species for which high quality distribution records were derived from the international Ocean Biogeographic Information System (OBIS) (www.iobis.org). These models will be created using five different distribution modelling algorithms (GLM, GAM, MaxEnt, Random Forest and Boosted Regression Trees) and two different sub-sampling methods (random and spatial). By ranking the AUC of models with and without a predictor we can find the relative importance of the predictors. The accuracy of the models with a different number of predictors allows us to derive the optimal number of predictors to use when developing SDMs in a marine environment. The results from a preliminary test on 11 predictors from Bio-ORACLE and 1756 species indicate that seemingly very good models (AUC > 0.9) can have unexpected predictors, for example pH as the most important predictor contributing to the model. Surprisingly some biologically relevant predictors like mean sea-surface temperature were on average not more important than salinity and dissolved oxygen. These first results will be further explored by running models with different predictor combinations from different environmental datasets.

References

- Sbrocco E.J. and P.H. Barber. 2013. MARSPEC: ocean climate layers for marine spatial ecology. *Ecology* 94(4):979–979. doi:10.1890/12-1358.1
- Tyberghein, L., H. Verbruggen, K. Pauly, C. Troupin, F. Mineur, and O. De Clerck. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography* 21(2):272–281.