MODELLING THE ECOLOGICAL NICHE OF MACROPHYTOBENTHOS SPECIES ALONG THE BULGARIAN BLACK SEA COAST

Dessislava Ganeva

University of Southampton - UK
e-mail: dessislava.ganeva@gmail.com

Keywords: *Ecological niche modeling, Maxent, Cystoseira barbata, Marine monitoring network*

Abstract: *This work is part of the effort to build statistically sound marine monitoring network to support the implementation of Article 11 (coordinated monitoring programmes) of the EU Marine Strategy Framework Directive (MSFD) in Bulgaria for the period 2014-2018. It is known that the existing monitoring is insufficient to meet the requirements of MSFD. A statistically sound sampling strategy is needed to provide adequate data for environmental status assessment. The first step in order to build a marine monitoring network is to develop an ecological niche model (ENM) of each ecosystem of the region.*

Because of the nature of the available field data, a presence-only ENM was selected and Maxent modelling tool was chosen. A Maxent model of Cystoseira barbata (brown algae) was developed and instructions are provided in order to develop an ENM for the rest of the ecosystems. The environment predictor variables used to develop the Maxent model are existing monitoring information combined with satellite-derived information.

МОДЕЛИРАНЕ НА ЕКОЛОГИЧНАТА НИША НА МАКРОФИТОБЕНТОСИ ПО БЪЛГАРСКОТО ЧЕРНОМОРСКО КРАЙБРЕЖИЕ

Десислава Ганева

University of Southampton
e-mail: dessislava.ganeva@gmail.com

Ключови думи: *Модел на екологичната ниша, Maxent, Cystoseira barbata, Морска мониторингова мрежа*

Резюме: *Тази разработка е част от усилията да се създаде статистически стабилна морска мониторингова мрежа, за да се подкрепи прилагането на член 11(координирани мониторингови програми) от Европейската Рамкова директива за морска стратегия (РДМС) в България за периода 2014-2018. Съществуващият мониторинг е недостатъчен за постигане на изискванията на РДМС. Статистически стабилна стратегия за пробонабиране е необходима, която да предостави достатъчно данни за оценка на състояние на околната среда. Първата стъпка при изграждане на морска мониторингова мрежа е разработката на модел на екологичната ниша (МЕН) за всяка екосистема в района.*

Моделът Maxent, който е „presence-only”, бе избран поради натурата на достъпните полеви данни. За Cystoseira barbata (вид кафяви водорасли) де изработен Maxent модел. Предоставени са инструкции за изработване на МЕН за останалите видове от екосистемите. Променливите, които определят околната среда използвани в Maxent модела, са съществуващи мониторингови данни комбинирани със спътникови данни

Introduction

This study is the first step in order to develop a monitoring network of the Bulgarian Black Sea by modelling the spatial distribution of the different ecosystems (Amorim *et al.* 2014, Fyhr *et al.* 2013). Existing monitoring information will be combined with satellite-derived information to enable the design of statistically sound species distribution of the ecosystems using GIS tools.

This study focuses on modelling the ecological niche of Cystoseira barbata (macrophytobenthos species) along the Bulgarian Black Sea coast using Maxent tool and gives practical steps in order to perform a Maxent model for the rest of the identified ecosystems. The present report is a summary of this study that could be downloaded from www.academia.edu.

Methodology

In the literature there is sometimes confusion between species distribution model (SDM) and ecological niche model (ENM) (Anderson 2012), however it is important question for the interpretation of the results (Araujo and Guisan 2006). In this study the ENM is considered as described by Anderson 2012, because only the suitable conditions (abiotically) will be modeled.

Modelling techniques are especially useful when there is a lack of biological surveys, as is the case in the Black sea. While ENM cannot replace the actual monitoring, its predictions can be used to construct effective marine monitoring strategies for impact and ecological status assessments needed for the implementation of ecosystem-oriented management regulations such as the European Marine Strategy Framework Directive (Fyhr *et al.* 2013, Reiss *et al.* 2014).

In order to establish the best sample sites for the monitoring program the species ecological niche should be estimated. Therefore a correlative ENM will be developed (Amorim *et al.* 2014, Stohlgren *et al.* 2011). Correlative ENM are empirical models relating field observations (sampling data) to environmental predictor variables, based on statistically or theoretically derived response surfaces (Guisan A. and Thuiller W. 2005). Therefore the quality of the input data is of course of great importance for the model (Lozier *et al.* 2009, Soberon and Peterson 2004), as Pearson, 2009 states: “Garbage in, garbage out”. In our case this applies for the sampling data and the environment data.

There are two main groups of ENM that are divided by the type of data they use: presence-absence and present-only data (Brotons *et al.* 2004). The available data for seabed habitats in the Bulgarian part of the Black sea is a presence-only data. Therefore only those types of ENM will be considered. Another specificity of the available data is that the sample size is very small. For example there are 20 sampling locations for *Cystoseira barbata* in 2012 monitoring.

The Maximum Entropy model was chosen as the most appropriate, considering the specificity of the available data and its predictive capabilities even for small sample size (Anderson and Gonzalez 2011, Magris and Déstro 2010, Meißner *et al.* 2014, Merow *et al.* 2013, Reiss *et al.* 2011, Stockwell and Peterson 2002, Wisz *et al.* 2008) and Maxent v3.3.3.3k software (Phillips et al. 2006) was used.

Maxent relies on an unbiased data, however our data is highly biased therefore special care was taken in order to limit the effect of the sampling bias (Barnes *et al.* 2014, Elith *et al.* 2010, Fourcade *et al.* 2014, Kramer *et al.* 2013, Phillips *et al.* 2009, Syfert *et al.* 2013). The split correction was used when creating the bias file, in the sense that for the strip 0-3m a Gaussian function was applied and for the strip 3-end of the distribution area of the species a value of almost 0 density was applied.

The choice of the environment predictor variables is a very important step in modelling species ecological niche because those models use correlative approaches that use the environmental variables to explain patterns of species ecological niche (Reiss *et al.* 2011). Maintaining a small number of predictor variables, when having small sampling size is a sound strategy (Warren *et al.* 2014). Along with substrate characteristics and wave regime the environment predictor variables from remote sensing in Table 1 were used.

Table 1. Products from GlobColor (www.globcolor.info) used in the study

Environmental predictor variables from remote sensing	**Product from ACRI-ST GlobColour Team 2014**	**Remarks**
Total suspended matter concentration	TSM	Chosen because validated for water type II
The Secchi Disk depth	ZSD-DORON	Chosen because validated for water type II
Chlorophyll-a concentration	CHL-2	Chosen because validated for water type II
Photosynthetically Available Radiation	PAR	

Another important aspect of a model is the spatial extend of the study area because the model performance relies on it (Anderson & Raza 2010, Barbet-Massin *et al.* 2012, Barve *et al.* 2011). The study area should be the geographical area accessible to the specie over given time period (Fourcade *et al.* 2014, Anderson & Raza, 2010). After discussion with Bulgarian macrophytobenthos experts (Elitza HINEVA from Institute of Oceanology Varna and Dimitar BEROV from the Institute of Biodiversity and Ecosystem Research at the Bulgarian Academy of Sciences in Sofia), it was agreed that the most probable habitat distribution of *Cystoseira barbata* along the Bulgarian Black Sea coast is between 0 and 15m water depth.

With different scale and resolutions of the data, different patterns are visible (Austin 2007, Guisan and Thuiller 2005, Haegeman and Etienne 2010, Pearson and Dawson 2003, Pearson *et al.* 2004), it is why a great care was taken in adjusting the right spatial scale for the application. On one hand, the samples were taken from a square of 15m X 15m and their exact location is unknown. Therefore a coarser resolution that 15m X 15m would better take into account this uncertainly. On the other hand the substrate is the predictor variable that is the most important and the resolution should be fine enough to distinguish correctly the hard substrate. A resolution of about 50m correctly represents the substrate. On the third hand the study area for each species is a narrow strip of less than 2km wide along the Bulgarian Black Sea coast and is less than 400km long. Therefore a spatial resolution of about 50m per cell size was chosen for all the variables in the Maxent model.

The calibration and selection of the suitable model is done following Scheglovitova and Anderson (2013) and Pearson *et al.* 2007 for small sampling size. The model was run with different types of features and regularization and optimising first the omission rate and then the AUC. The lowest presence threshold rule (Pearson *et al.* 2007) or the equivalent minimum training presence threshold (in Maxent) was used in order to determine if a test sample is within or out of the predicted area. Because of the small sample size the chosen threshold represents a conservative rule preventing to overestimate the area suitable for the species. Therefore the omission rate is the number of times a test sample is outside of the predicted area, the lower the omission rate the better. The types of feature for which the model was tested are L (linear), Q (Quadratic) and H (Hinge) following Philipps and Dudik 2008 recommendations because of the sample size of 20. Using complex features increase the complexity of the model and could lead to an overfitting. The following features or group of features will be used for the test: L, Q (LQ and QH are the same as Q), H (LH is the same as H), LQH. Regularization of the other hand gives more or less tolerance in the selected variable value and therefore smaller regularization value will probably result in more restricted predictions and larger regularisation values in less discriminating predictions. The regularization multipliers 0.25, 0.5, 0.75, 1, 1.25, 1.50, 1.75, 2 were used.

This study utilise two tests to evaluate the model: (i) Area Under the Receiver Operating Characteristic Curve (AUC) (ii) test for significance of the results. An AUC value of 1 is considered a perfect prediction and a value of 0.5 or less is consider a prediction no better than random. In other words a model with AUC = 0.89 ranks the suitability of the site correctly 89% of the time. The “leave-one-out” approach for model evaluation was applied as described by Pearson *et al.* 2007.

Results

The tests with lower omission rate are the tests for every tested feature class with the regularization multiplier of 1.75 and 2. For each feature class, higher regularization multipliers correspond to lower omission, which correspond to the results from Scheglovitova and Anderson (2013). The tests with the regularization multiplier 1.75 were selected as optimal settings because of the downscale resampling of the remote sensing data from 1000m to 50m spatial resolution. Therefore more value tolerance for those variables is not necessary.

For the tests with lower omission rate, an average AUC was calculated from the 20 “leave-one-out” runs of the model, results in Table 2. For each model a P-value was computer with the tool proposed by Pearson *et al.* 2007 in their Appendix. All the models are statistically significant (P-value<0.05).

Table 2. AUC and P-value for the tests with the lowest omission rate

Test	AUC mean test	P-value
Test H, 1.75	0,8426	0,000019
Test LQH, 1.75	0,84885	0,000027
Test Q, 1.75	0,84835	0,000027
Test L, 1.75	0,8303	0,000015

Conclusions

The calibration of the Maxent model shows 4 models with similar performances, see Table 2. Although it is still under consideration which of those models to choose, the less complex model was selected (L1.75, Linear feature and regularization multiplier 1.75). Only 3 out of the 6 variables contribute to the model (substrate, waves, chlorophyll-a).

The proposed *Cystoseira barbata* ecological niche as an input for modelling the monitoring network is in Figure 1.

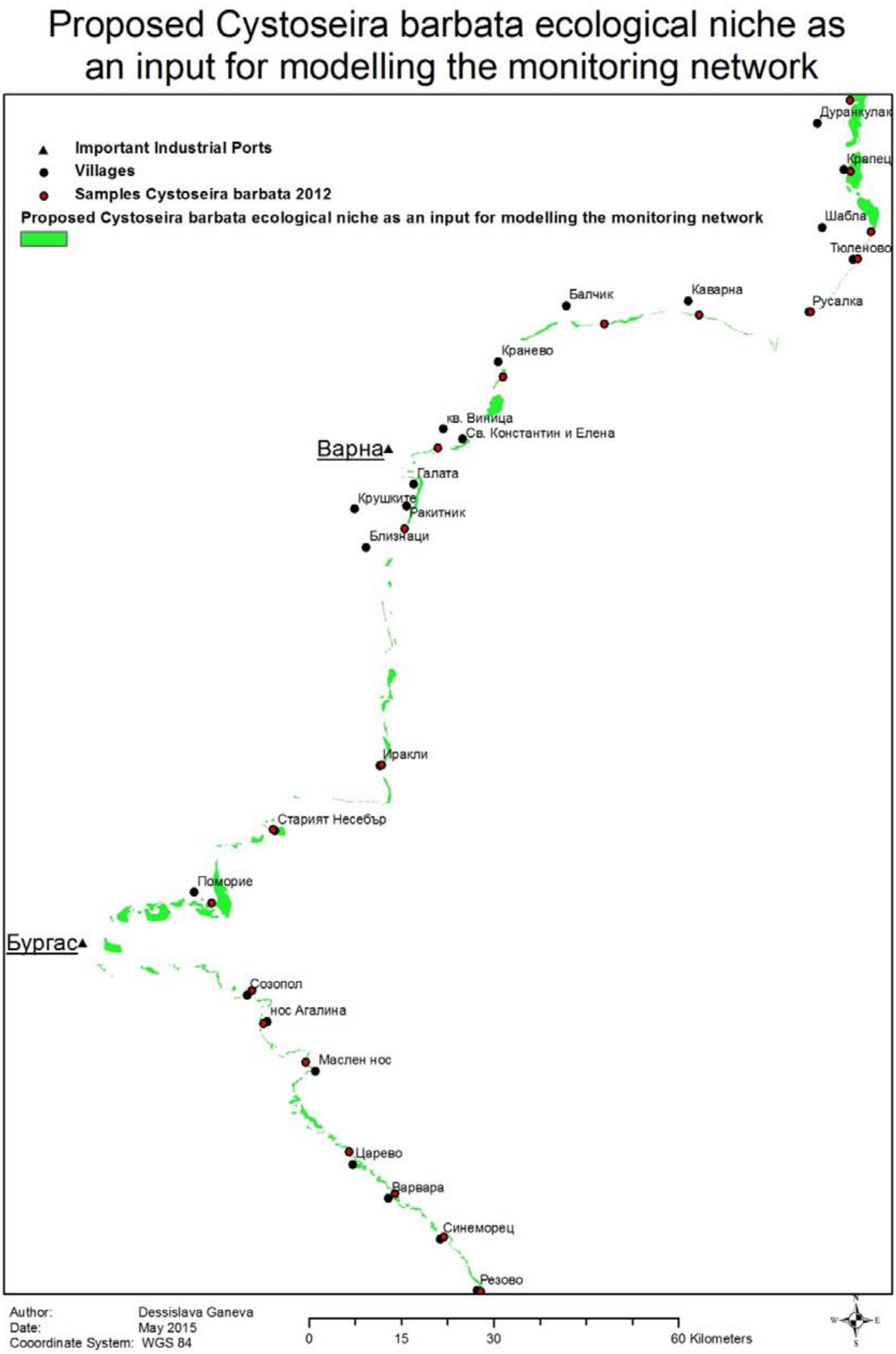

Fig. 1. Proposed *Cystoseira barbata* ecological niche as an input for modelling the monitoring network

References:

1. Amorim, F., Carvalho, S.B., Honrado, J., Rebelo, H., 2014. Designing Optimized Multi-Species Monitoring Networks to Detect Range Shifts Driven by Climate Change: A Case Study with Bats in the North of Portugal, *PLoS ONE*, 9(1), e87291
2. Anderson, R.P., 2012. Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Science*, 1260, 66-80
3. Anderson, R.P. and Raza, A., 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. *Journal of Biogeography*, 37, 1378–1393
4. Anderson, R.P. and Gonzalez Jr., I., 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling*, 222, 2796–2811
5. Araujo, M.B. and Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677–1688
6. Austin, M., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological modelling*, 200, 1–19
7. Barbet-Massin, M., Jiguet, F., Albert C.A. and Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolutio,* 1-12
8. Barnes, M.A., Jerde, C.L., Wittmann, M.E., Chadderton, W. L., Ding, J., Zhang, J., Purcell, M., Budhathoki, M. & Lodge D.M., 2014. Geographic selection bias of occurrence data influences transferability of invasive Hydrilla verticillata distribution models. *Ecology and Evolution*, 4(12), 2584–2593
9. Barve, N., Barve, V., Jimnez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberyn, J. and Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222, 1810–1819
10. Brotons, L., Thuiller, W., Araujo, M.B. and Hirzel, A.H., 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability, *ECOGRAPHY*,27, 437-448
11. Elith, J., Kearney, M. and Phillips, S., 2010. The art of modelling range-shifting species, *Methods in Ecology and Evolution,* 1, 330–342
12. Fourcade, Y., Engler, J.O., Rodder, D., Secondi, J., 2014. Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. *Plos One*, 9(5), e97122
13. Fyhr, F., Nilsson, Å., Nyström, S.A., 2013. A review of Ocean Zoning tools and Species distribution modelling methods for Marine Spatial Planning. Aquabiota Water Research paper
14. Guisan, A. and Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models, *Ecology Letters*, 8, 993–1009
15. Haegeman, B. and Etienne, R.S., 2010. Entropy Maximization and the Spatial Distribution of Species, *The American Naturalist*, 175(4), E74-E90
16. Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schroder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J. Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J. W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer H. and Wilting, A., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, 19, 1366–1379
17. Lozier, J. D., Aniello, P. and Hickerson, M. J., 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling, *Journal of Biogeography*, 36, 1623–1627
18. Magris, R.A. and Déstro, G.F.G., 2010. Predictive Modelling Of Suitable Habitats For Threatened Marine Invertebrates and implications for Conservation Assessment In Brazil, *Brazilian Journal of Oceanography*, 58, 57-68
19. Meißner, K., Fiorentino, D., Schnurr, S., Martinez Arbizu, P., Huettmann, F., Holst, S., Brix, S., Svavarsson, J., 2014. Distribution of benthic marine invertebrates at northern latitudes — An evaluation applying multi-algorithm species distribution models, *Journal of Sea Research*, 85, 241–254
20. Merow, C., Smith, M. J. and Silander, J.A. Jr., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter, *Ecography,* 36, 1058–1069
21. Pearson, R.G. and Dawson, T.E., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12, 361–371
22. Pearson, R.G, Dawson, T.P and Liu, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, 27, 285-298
23. Pearson R.G., Raxworthy C.J., Nakamura M. and Peterson A.T., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34, 102–117
24. Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259
25. Phillips, S.J. and Dudık, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175
26. Phillips, S.J., Dudik, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Feffier, F., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181-197
27. Reiss, H., Cunze S., König, K., Neumann, H., Kröncke, I., 2011. Species distribution modelling of marine benthos: a North Sea case study, *MARINE ECOLOGY PROGRESS SERIES*, 442, 71–86

28. Reiss, H., Cunze S., König, K., Neumann, H., Kröncke, I., 2011. Species distribution modelling of marine benthos: a North Sea case study, *MARINE ECOLOGY PROGRESS SERIES*, 442, 71–86
29. Reiss, H., Birchenough, S., Borja, A., Buhl-Mortensen, L., Craeymeersch, J., Dannheim, J., Darr, A., Galparsoro, I., Gogina, M., Neumann, H., Populus, J., Rengstorf, A.M., Valle, M., van Hoey, G., Zettler, M.L. and Degraer, S., 2014. Benthos distribution modelling and its relevance for marine ecosystem management. *ICES Journal of Marine Science*
30. Shcheglovitova, M. and Anderson, R.P., 2013. Estimating optimal complexity for ecological niche models: Ajackknife approach for species with small sample sizes. *Ecological Modelling*, 269, 9–17
31. Soberon, J. and Peterson, A.T., 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Le Royal Society B*, 359, 689–698
32. Stockwell, D.R.B. and Peterson, A. T., 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148, 1–13
33. Stohlgren, T.J., Kumar, S., Barnett, D.T. and Evangelista, P.H., 2011. Using Maximum Entropy Modeling for Optimal Selection of Sampling Sites for Monitoring Networks, *Diversity*, 3, 252-261
34. Syfert, M.M., Smith, M.J., Coomes, D.A., 2013. The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. *Plos One,* 8(2), e55158
35. Warren, D.L., Wright, A.N., Seifert S.N. and Shaffer H.B., 2014. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Diversity and Distributions*, 20, 334–343
36. Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14, 763–773