**Journal of the Marine Biological Association UK (JMBA)**

**SUPPORTING INFORMATION**

**Climate and environmental changes driving idiosyncratic shifts in the distribution of tropical and temperate worm reefs**

**Larisse Faroni-Perez1,2**

*Predictors*

 Environmental layers acquired from HadGEM2-ES were divided into the time period (current, middle of the century and end of the century) and the future impacts of the simulated climate change (RCP2.6 and RCP8.5). Each predictor was rescaled, and the interannual monthly mean was calculated following the annual values: average, minimum and maximum (Table 1). The set of layers in the ‘current’ time period were subdivided into two; global spatial layers were masked and cut onto the biogeographic provinces surrounding region for each species known occurrence (i.e. strictly to the occurring latitudinal span and longitudinally delimited by the closest cell grid representing the continental shoreline). Correlation matrices were analysed to exclude highly correlated predictors (Pearson, r > ±0.60), therefore remained predictors are shown in Table 1. Afterward, remained predictors were analysed by generalized linear models (GLMs) using a total number of four predictors in each model, considering all possible combinations of single predictor per niche (Table 2). A predictor was scored whilst significant in the model. Then, for each niche, based on scored rank across models the most influential predictive capacity of predictor was selected to final model (Table 2). For both the species, the final set of predictors selected by ranking were also significant by the GLM (Table 2). Statistical analysis of environmental predictors used in the final model is shown in Table 3.

**Table 1:** Environmental predictors layers acquired from IPCC, AR5 model HadGEM2-ES. Time period represent: present days (1986–2005), 50’s (middle of century: 2040–2059), and the 90’s (late of century: 2080–2099). Future scenarios represent: RCP2.6 (low emissions) and RCP8.5 (high emissions). Codes: ... = layers used for correlation matrices, **...** = layers remained and used in GLM models, and **... =** final layers selected.

****

**Table 2:** Results of GLMmodels considering all possible combinations of single predictor per niche. Predictor used in the model was scored as 1 if significant (p < 0.05). Rank = sum of predictors’ score. \*:final model selected. ****

**Table 3:** Analyses of environmental predictor used in the model encompassing biogeographic provinces where occur the topical (*Phragmatopoma caudata*) and temperate (*P. virgini*) worm reefs. Future scenarios represent RCP2.6 (low emissions) and RCP8.5 (high emissions). Time period (T.P.) represent: present days (1986–2005), 50’s (middle of century: 2040–2059), and the 90’s (late of century: 2080–2099). Future projections are based on IPCC, AR5 from model HadGEM2-ES. 

\*Predictors code: detoc=Detrital Organic Carbon Concentration (max); si=Dissolved Silicate Concentration (max); pH=Negative log of hydrogen ion concentration (min); tos=Sea Surface Temperature (min); pr = Precipitation (max); friver=Water Flux into Sea Water From Rivers (max). In brackets, the specific layer for predictor selected after peer-to-peer correlation and GLM selection processes.

*Modelling (procedure)*

Previous studies assessed the sensitivity of MaxEnt using several options and described several modelling complexities (Anderson & Raza, 2010; Elith *et al*., 2010; Anderson & Gonzalez, 2011; Shcheglovitova & Anderson, 2013). However, the spatial scales of these previous studies were mainly regional, unlike the continental scale in this present study. Although evaluating the sensitivity of MaxEnt was not the main objective of this study, I carefully checked for the different approaches for modelling based on the present data. Assessing species-specific model tuning instead of using default features is highly recommended (Merow *et al*., 2013). Following Elith *et al*., (2010), Anderson & Gonzalez, (2011), Elith *et al*., (2011) and Shcheglovitova & Anderson, (2013), the complexity of models was checked for testing the regularization multiplier (RM) values of 1 (default) and 2.5, and the following feature classes Hinge (H) or Linear+Quadratic (LQ). The fitted model has regularization value of 1 (currently default value) and Hinge feature class, due to low omission rate and higher AUC values respectively (Tables 4-5). This metric based on the omission rate tends to select simpler models (Radosavljevic & Anderson, 2014). The final MaxEnt settings used in fitting the model of the present study are likely in Elith *et al*., (2010).

Table 4 Results for *Phragmatopoma caudata* of models for the two tested feature classes: hinge (H) and linear and quadratic (LQ), with the regularization multiplier (RM) values tested 1(default) and 2.5. Fixed cumulative value 10: Training omission rate.



Table 5 Results for *Phragmatopoma virgini* of models for the two tested feature classes: hinge (H) and linear and quadratic (LQ), with the regularization multiplier (RM) values tested 1(default) and 2.5. Fixed cumulative value 10: Training omission rate.



*Clamping*

Aditionally to regular analysis, I run *do clamping* as false and calculated the relative difference in the frequency of potential suitability classes compared to the obtained results to setting *do clamping* as true, as this MaxEnt setting needs to be tested (Anderson & Raza, 2010). The *do clamping* set as true in MaxEnt means projection is made using data range found only within the training data set (Phillips et al., 2006; Elith et al., 2011), that is recommended when occurring non-analogous relationships of the predictors across the time (Fitzpatrick & Hargrove, 2009) (Table 6). The comparison between the two clamping settings demonstrated that suitable area projections for *P. virgini* were almost not sensitive, since minor differences in the confidence of suitability occurred (Table 6). For *P. caudata*, projections for the end of century under both scenarios showed the most differences in relative entropy, and the maximum value were 12% in suitability. Considering the continental scale and the metric for potential habitat suitability, it is very likely that in regional or local scale, the difference in clamping results will be more sensitive, and further analysis are recommended.

*Post Modelling*:

The technique of using threshold values to recalibrate the maps into binary showing an unsuitable or suitable area is widely used, it has been recently criticized (Guillera-Arroita *et al*., 2015). Applying the technique of a threshold conversion map results in omitting the continuous prediction and thus a gradient of environmental suitability, which real species face. The procedure to transform the continuous results into a binary map of assumed presence or absence of suitable areas is a coarse way to assess species occurrence probabilities, due to the loss of some information provided by continuous model output. Consequently, the application of this technique is not recommended (Guillera-Arroita *et al*., 2015). Here, for the practical purposes of fit, the present SDM to assess quantitative changes for each species, the categorical value of 0.2, was set, and the subsequent index system was applied. Although 0.2 is not statistically justified, to set this value as clear-cut, I considered either present-day distribution of suitable habitats and known occupancy areas for both the species. Therefore, this value is fine-tuned for present models and species, very likely it will differ according to species and modelling approaches. The index system gradients refer to the areas of predicted environmental suitability based on continuous model outputs and are not the probability of species distribution *sensu stricto*. This analysis attempted solely to characterize and quantify ‘potential’ shifts in the distribution of each species based on their areas of environmental suitability.

Table 6: Projected changes in the potential suitability for topical (*Phragmatopoma caudata*) and temperate (*P. virgini*) worm reefs with contrast to their current habitat suitability. Δ represents the difference in pixel suitability between scenarios RCP2.6 (low emissions) and RCP8.5 (high emissions) and between the 50s (mid-century: 2040–2059) and the 90s (late-century: 2080–2099) time periods. In brackets, the total frequencies (F%) given in percentage of grid cells per projection. \*ΔnotClamping: values represent the difference between results obtained when running models with clamping.



**Software used**

*Model Processing*

Primarily environmental envelopes were processed using the Climate Data Operators (CDO), version 1.4.0 available at https://code.zmaw.de/projects/cdo/files. For all experimental models and the final modeling, I used the software MaxEnt version 3.3.3k downloaded from https://www.cs.princeton.edu/~schapire/maxent/. Model accuracy were calculated using NicheToolBox (Osorio-Olvera L. 2016). Also, I executed additional analyse with R (R Core Team 2012) using the following packages:

* ‘*raster*’ (Hijmans & Van Etten, 2011)
* ‘*rgdal’* (Bivand *et al*., 2015)
* *‘maptools’* (Bivand & Lewin-Koh, 2015)
* *‘sp’* (Pebesma & Bivand, 2005)
* ‘*MuMIn’* (Barton, 2015)
* ‘*dismo*’ (Hijmans *et al*., ‎2012)

References

**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.

**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.

**Barton, K.** (2015) MuMIn: multi-model inference Available at *https://cran.r-project.org/web/packages/MuMIn/index.html*

**Bivand, R. and Lewin-Koh, N.** (2015) maptools: Tools for Reading and Handling Spatial Objects. Available at http://CRAN.R-project.org/package=maptools

**Bivand, R., Keitt, T. and Rowlingson, B.** (2015) rgdal: Bindings for the Geospatial Data Abstraction Library. Available at *https://*cran.r*-project.org/package=rgdal*

**Elith, J., Kearney, M. and Phillips, S.** (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1, 330-342.

**Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E. and Yates, C.J.** (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17, 43-57.

**Fitzpatrick, M.C. and Hargrove, W.W.** (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation* 18, 2255-2261.

**Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M.A., Tingley, R. and Wintle, B.A.** (2015) Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* 24, 276–292.

**Hijmans, R.J. and Van Etten, J.** (2011) Raster: Geographic data analysis and modeling. R package version 2.0-12. Available at *http://cran.r-project.org/package=raster*

**Hijmans, R.J., Phillips, S., Leathwick, J., and Elith, J.** (2012) dismo: Species distribution modeling. R package version 0.6-3. R Foundation for Statistical Computing, Vienna, Austria*. http://cran.r-project. org/web/packages/dismo/*

**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.

**Osorio-Olvera, L.** (2016) *NicheToolbox: A Web Tool for Exploratory Data Analysis and Niche Modeling*.

**Pebesma, E.J. and Bivand, R.S.** (2005) Classes and methods for spatial data in R. R News 5 (2). Available at *http://cran.r-project.org/doc/Rnews/*

**Phillips, S.J., Anderson, R.P. and Schapire, R.E.** (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231-259.

**Radosavljevic, A. and Anderson, R.P.** (2014) Making better MAXENT models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography* 41, 629-643.

**Shcheglovitova, M. and Anderson, R.P.** (2013) Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. *Ecological Modelling*, 269, 9-17.