



HAL
open science

Broad-scale species distribution models applied to data-poor areas.

Charlène Guillaumot, Jean Artois, Thomas Saucède, Laura Demoustier, Camille Moreau, Marc Eléaume, Antonio Agüera, Bruno Danis

► **To cite this version:**

Charlène Guillaumot, Jean Artois, Thomas Saucède, Laura Demoustier, Camille Moreau, et al.. Broad-scale species distribution models applied to data-poor areas.. Progress in Oceanography, 2019, 175, pp.198-207. 10.1016/j.pocean.2019.04.007 . hal-02270076

HAL Id: hal-02270076

<https://hal.science/hal-02270076>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Broad-scale species distribution models applied to data-poor areas

Guillaumot Charlène^{1,3*}, Artois Jean², Saucède Thomas³, Demoustier Laura¹, Moreau Camille^{1,3}, Eléaume Marc⁴, Agüera Antonio^{1,5}, Danis Bruno¹

¹ Université Libre de Bruxelles, Marine Biology Lab. Avenue F.D. Roosevelt, 50. CP 160/15 1050 Bruxelles, Belgique

² Université Libre de Bruxelles, Spatial Epidemiology Lab. (SpELL). Avenue F.D. Roosevelt, 50. CP 160/15 1050 Bruxelles, Belgique

³ UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, 6 bd Gabriel F-21000 Dijon, France

⁴ Muséum national d'Histoire naturelle, Département Systématique et Évolution, UMR ISYEB 7205, 57 rue Cuvier, F-75231 Paris Cedex 05, France

⁵ Danish Shellfish Center, DTU-aqua, Øroddevej 80, 7900 Nykøbing Mors, Denmark.

* charleneguillaumot21@gmail.com

Running headline: broad scale SDM in data-poor areas

34 **Abstract**

35 Species distribution models (SDMs) have been increasingly used over the past decades to characterise the
36 spatial distribution and the ecological niche of various taxa. Validating predicted species distribution is
37 important, especially when producing broad-scale models (i.e. at continental or oceanic scale) based on
38 limited and spatially aggregated presence-only records. In the present study, several model calibration
39 methods are compared and guidelines are provided to perform relevant SDMs using a Southern Ocean
40 marine species, the starfish *Odontaster validus* Koehler, 1906, as a case study. The effect of the spatial
41 aggregation of presence-only records on modelling performance is evaluated and the relevance of a target-
42 background sampling procedure to correct for this effect is assessed. The accuracy of model validation is
43 estimated using k-fold random and spatial cross-validation procedures. Finally, we evaluate the relevance of
44 the Multivariate Environmental Similarity Surface (MESS) index to identify areas in which SDMs accurately
45 interpolate and conversely, areas in which models extrapolate outside the environmental range of
46 occurrence records.

47 Results show that the random cross-validation procedure (i.e. a widely applied method, for which training
48 and test records are randomly selected in space) tends to over-estimate model performance when applied to
49 spatially aggregated datasets. Spatial cross-validation procedures can compensate for this over-estimation
50 effect but different spatial cross-validation procedures must be tested for their ability to reduce over-fitting
51 while providing relevant validation scores. Model predictions show that SDM generalisation is limited when
52 working with aggregated datasets at broad spatial scale. The MESS index calculated in our case study show
53 that over half of the predicted area is highly uncertain due to extrapolation. Our work provides
54 methodological guidelines to generate accurate model assessments at broad spatial scale when using
55 limited and aggregated presence-only datasets. We highlight the importance of taking into account the
56 presence of spatial aggregation in species records and using non-random cross-validation procedures.
57 Evaluating the best calibration procedures and correcting for spatial biases should be considered ahead the
58 modelling exercise to improve modelling relevance.

59

60 **Key-words**

61 Boosted Regression Trees (BRTs), presence-only, cross-validation, extrapolation, modelling evaluation

62

63

64

65 1. INTRODUCTION

66 Species Distribution Models (SDMs) have been increasingly used during the past decades. The diversity of
67 applications has widened to include a vast panel of topics from studies of invasive species distribution range
68 shifts to assessment of species responses to environmental drivers and conservation issues from local to
69 global scales ([Guisan and Thuiller 2005](#), [Ficetola et al. 2007](#), [Guisan et al. 2013](#), [Beaumont et al. 2016](#),
70 [Phillips et al. 2017](#)). In vast and remote areas such as the Southern Ocean, modelling species distributions is
71 challenged by (1) the paucity of biotic data available (a serious constraint when describing species realised
72 niche), (2) by the heterogeneous quality of environmental data describing environmental conditions (*e.g.*
73 missing data in coastal areas, low resolution of environmental layers, limited number of environmental
74 descriptors available), and (3) by the sampling bias (spatial and temporal aggregation of data collection)
75 ([Barry and Elith 2006](#), [Robinson et al. 2011](#), [Hortal et al. 2012](#), [Tessarolo et al. 2014](#), [Guillaumot et al. 2018](#)).
76 Sampling effort has mostly been carried out offshore or in the vicinity of research stations during the austral
77 summer while remote shallow areas are seldom accessed and dense winter sea ice conditions limit
78 oceanographic studies ([Gutt et al. 2012](#)).

79 Several studies have proposed model corrections or alternatives to separately mitigate the induced impacts
80 of spatial and temporal biases on modelling performance ([Phillips et al. 2009](#), [Newbold 2010](#), [Barbet-Massin
81 et al. 2012](#), [Hijmans 2012](#), [Tessarolo et al. 2014](#), [Guillera-Aroita et al. 2015](#), [Guillaumot et al. 2018](#), [Valavi et
82 al. 2018](#)). However, to our knowledge, no study has yet proposed methodological guidelines to address such
83 issues when dealing with data-poor and broad spatial areas (*i.e.* at continental or oceanic scales).

84
85 Several statistical tools such as the Area Under the Curve of the Receiver Operating characteristic (AUC),
86 the True Skill Statistic, or the Point Biserial Correlation are commonly used to evaluate the relevance of
87 SDMs predictions ([Fielding and Bell 1997](#), [Allouche et al. 2006](#)). Using these indices for models performed
88 with presence-only data has been widely discussed because background-data are usually considered as
89 absences, leading to confusion in model interpretation and violating most test assumptions (*i.e.* computing
90 AUC and TSS statistics requires the use of true absences) ([Jimenez-Valverde 2012](#), [Li and Guo 2013](#)).
91 These methods can also be biased when applied to limited and broadly distributed data. Machine-learning
92 algorithms are widely used in SDMs to fit complex relationships between species occurrences and
93 environmental data ([Elith et al. 2006](#)). The resulting models may be highly complex and poorly efficient under
94 changing environmental conditions as they may fit a response to any variation including the random noise
95 (=model overfitting),([Wenger and Olden 2012](#)). Models' ability to predict in new environmental conditions is
96 described as the generalisation performance.by [Friedman et al. \(2001\)](#).

97 Producing reliable SDMs implies finding a good trade-off between model complexity and predictive and
98 generalisation performances (Anderson and Gonzalez 2011, Radosavljevic and Anderson 2014). The
99 relevance of modelling and generalisation performance, and the optimal level of model complexity can be
100 tested using independent data. The method has been commonly applied and referred to as the cross-
101 validation procedure (Araujo and Guisan 2006, Valavi et al. 2018). The cross-validation procedure uses a
102 training subset of occurrence data to fit the model and a separate test subset to validate the predictions and
103 the statistical relationships between the studied variables (Fielding and Bell 1997). 'Random cross-validation'
104 procedures are widely used and randomly split the occurrence dataset into training and test subsets.
105 However, the spatial aggregation of occurrence data can lead to the violation of the independence
106 assumption between training and test data randomly sampled, and in turn to false confidence in modelling
107 validation performances (Hijmans 2012). The violation of the independence assumption can also lead to
108 generate highly complex and overfitted models (Boria et al. 2014, Merow et al. 2014, Radosavljevic and
109 Anderson 2014). Therefore, the cross-validation procedure should be adapted to each given dataset and
110 case study, so that, different 'spatial cross-validation' procedures have been developed and compared in this
111 study. The spatial cross-validation procedures aim at spatially splitting the occurrence dataset into a training
112 and a test subset by increasing the geographical distance between the two subsets (Veloz 2009, Brenning
113 2012, Muscarella et al. 2014, Radosavljevic and Anderson 2014, Brown et al. 2017, Valavi et al. 2018). The
114 spatial cross-validation reduces spatial correlation between training and test data in situations where spatial
115 autocorrelation is significant in the occurrence dataset, a common issue in ecology (Roberts et al. 2017).

116 Uncertainties in SDMs represent another limitation to model usage that should be quantified and the effects
117 must be specifically assessed or taken into account during model interpretation (Barry and Elith 2006,
118 Carvalho et al. 2011, Beale and Lennon 2012, Guisan et al. 2013). Model extrapolation outside the range of
119 the known species environmental conditions leads to misinterpretation of SDM outputs and can be a real
120 issue when using SDM predictions as a support tool for conservation decisions. Therefore, areas of optimal
121 predictions and limited uncertainties must be identified. This can be achieved using indicators such as the
122 Multivariate Environmental Similarity Surface (MESS). Developed for SDMs, the MESS index highlights
123 areas where environmental conditions are outside the range of conditions observed in data (Elith et al.
124 2010).

125 In the present study, model uncertainties and the performance of several spatial cross-validation procedures
126 were analysed using the case study of the sea star *Odontaster validus* Koehler, 1906. Distributed over the
127 entire Southern Ocean (< 45°S), *O. validus* is a common and abundant species in shallow-water benthic
128 habitats (McClintock et al. 2008, Lawrence 2013), characterised by an opportunistic feeding behaviour (from

129 suspension-feeding to algivory, deposit-feeding and predation). It has been shown to play a significant role in
130 structuring benthic communities and regulating populations of other benthic taxa (McClintock et al. 2008).
131 The species physiology was recently modelled using the Dynamic Energy Budget approach (Agüera et al.
132 2015) which allows for the assessment of the metabolic performance of the species under different
133 environmental conditions. Here, SDMs were produced to interpolate the known distribution of *O. validus* over
134 its entire geographic range using an available dataset of environmental descriptors. The influence of spatial
135 data aggregation on model outputs was analysed and the performance of correction procedures evaluated.
136 In a second step, several cross-validation procedures were assessed and compared to test for modelling
137 accuracy, optimal level of complexity and predictive performance. A final 'optimum' model is proposed, which
138 takes into account uncertainty estimates. Results are generalised and formalised as guidelines for further
139 SDM works, showing the relevance of the approach when working at broad spatial scale with a limited
140 number of spatially aggregated presence-only records.

141

142 **2. Material and methods**

143 *2.1. Model selection and calibration procedures*

144 SDMs were generated using the Boosted Regression Trees (BRTs) algorithm. BRTs were selected for their
145 ability to fit complex relationships between species records and the related environment, while guarding
146 against over-fitting (Elith et al. 2008, Reiss et al. 2011). BRTs are also adapted to deal with incomplete
147 datasets (Elith et al. 2008), can perform well with low prevalence datasets (Barbet-Massin et al. 2012), are
148 weakly sensitive to species niche width (Qiao et al. 2015) and were recognised to transfer well in space and
149 time (Elith et al. 2006, Elith and Graham 2009, Heikkinen et al. 2012).

150 BRTs were calibrated using the method proposed by Elith et al. (2008) to select the optimal number of trees
151 in the final model (Appendix A). The combination of parameters that minimises the optimal number of trees
152 to build the model (reduction of complexity) while reaching a minimum predictive deviance to the test data
153 (reduction of error) was selected. The following parameters were used to calibrate the models: tree
154 complexity= 4, bag fraction= 0.75 and learning rate= 0.007 (Fig. S2). The number of background data
155 sampled in the area was set at 1,000 sampled points after evaluating the optimal number of data points to be
156 sampled (see Appendix A for details). This number constitutes the best trade-off between describing
157 environmental conditions and being as close as possible to the number of species presence records
158 available (Barbet-Massin et al. 2012). All background sampling was restricted in space to areas shallower
159 than 1,500 m depth, which corresponds to the species deepest record, in order to avoid model extrapolation

160 at depths known as unsuitable for the species survival based on knowledge of the species ecology
161 (McClintock et al. 2008, Lawrence 2013). Sampling was restricted to a single background data per pixel.
162 Similarly, presence records falling on a same 0.1° grid-cell pixel were filtered before model calibration in
163 order to reduce spatial over-weighting (Segurado et al. 2006, Boria et al. 2014).

164

165 2.2. Occurrence dataset

166 SDMs were generated using presence-only data made available for the sea star *O. validus* by Moreau et al.
167 (2018). Presence-only records of *O. validus* are strongly aggregated in space (*i.e.* concentrated in “easily”
168 accessible and frequently visited areas characterised by relatively low sea ice concentrations), a condition
169 also prevailing in the total dataset available for Southern Ocean benthic taxa (updated from Griffiths et al.
170 (2014), Fig. S3), making *O. validus* a representative case study for Southern Ocean benthic studies.

171 Models were generated using the environmental descriptors published as raster layers by Fabri-Ruiz et al.
172 (2017). They were collected from different sources and modified to fit modelling requirements at the scale of
173 the Southern Ocean (from 45°S latitude to Antarctica coasts). Collinearity between environmental descriptors
174 was tested using the Variance Inflation Factor (VIF) stepwise procedure of the ‘usdm’ R package (Naimi et
175 al. 2014) and Spearman correlations (r_s) (R Core Team, 2017). Surface temperature and roughness, a
176 depth-derived variable, were respectively correlated to ice cover and depth. They were omitted according to
177 the commonly used thresholds of $VIF > 5$ and $r_s > 0.85$ (Pierrat et al. 2012, Dormann et al. 2013, Duque-
178 Lazo et al. 2016). A final set of 16 environmental descriptors at 0.1° resolution was compiled to build the
179 models (Table S5).

180

181 2.3. Evaluation and correction of spatial aggregation

182 The significance of spatial aggregation of occurrence data was tested by measuring spatial autocorrelation
183 (Legendre and Fortin 1989) on model residuals using the Moran’s I index (Segurado et al. 2006, Dormann
184 2007, Crase et al. 2012). A positive Moran’s I value (between 0 and 1) indicates that spatially close residuals
185 will share similar values. A negative (close to -1) or null value respectively indicates a maximal dispersion or
186 a random dispersion of residuals in space (Cliff and Ord 1981). Detecting significant spatial autocorrelation in
187 presence-only records will assess the degree of aggregation of species records in the studied area.

188 Two null models were generated and their respective outputs compared to each other in order to evaluate
189 the importance of spatial aggregation in the total Southern Ocean benthic dataset (Fig. S3). Null model #1
190 was produced to evaluate the overall spatial aggregation of benthic records in the Southern Ocean due to
191 sampling effort. It was generated by randomly sampling $n = 309$ occurrence records (corresponding to the

192 number of non-duplicate presence-only data available for *O. validus*) in the total Southern Ocean benthic
193 dataset (Fig. S3). 1,000 background records were randomly sampled in the entire Southern Ocean. The
194 Moran's I score was calculated by comparing model #1 predictions to the distribution of the total Southern
195 Ocean benthic dataset (Fig. S3). Null model #2 was built to compute a reference Moran's I score for a model
196 generated with randomly distributed records. 309 presence data and 1,000 background data were randomly
197 sampled in the entire Southern Ocean. Null model #2 would provide a reference value for spatial
198 autocorrelation scores due to the intrinsic structure of environmental data. It will serve as a reference model
199 for comparison with Moran's I scores of model null #1 and to assess the degree of spatial aggregation due to
200 sampling effort.

201 To correct for the effect of spatial aggregation on modelling performance, a target-background correction
202 method was applied (Phillips et al. 2009). The total Southern Ocean benthic dataset (Fig. S3) was used to
203 create a Kernel Density Estimation layer that provides an estimate of the probability to find a benthic
204 presence data for each pixel. The Kernel Density Estimation was calculated with the 'kde2d' function of the
205 MASS R package (Ripley, 2015) on the extent of the Southern Ocean (n and lims parameters defined to fit a
206 raster layer of extent (-180,180,-80,-45) and 0.1° resolution). Null model #1 was corrected by randomly
207 sampling 1,000 background records according to the weighting scheme of the Kernel Density Estimation
208 layer.

209
210 After evaluating spatial aggregation in the total Southern Ocean benthic dataset, spatial autocorrelation was
211 specifically assessed for *O. validus*. Spatial autocorrelation was measured for two models generated without
212 (model A) and with (model B) Kernel Density Estimation correction. Comparison between the two models
213 aimed at assessing the efficiency of the Kernel Density Estimation correction for *O. validus*. Model A (without
214 correction) was built using all presence-only data available for *O. validus* and 1,000 background records
215 randomly sampled in the Southern Ocean. Model B (with correction) was built using all presence-only data
216 available for *O. validus* and 1,000 background records that were sampled following the weighting scheme of
217 the Kernel Density Estimation layer. Each model was generated 100 times and the two averaged models
218 (average models A and B) were compared to each other. Differences between models A and B quantify the
219 importance of spatial aggregation on model outputs.

220 Finally, model relevance was assessed using three statistics: the Area Under the Receiver Operating Curve
221 (AUC) (Fielding and Bell 1997), the Point Biserial Correlation between predicted and observed values (COR,
222 Elith et al. 2006) and the True Skill Statistic (TSS, Allouche et al. 2006).

223

224 2.4. Testing different cross-validation procedures

225 SDMs validation was performed using different cross-validation procedures. Background data were first
226 sampled in the entire area following the Kernel Density Estimation scheme and the compilation of presence-
227 only and background data was then split into a training and a test subset to build the cross-validation
228 procedure. Two splitting procedures were followed; they differ between each other in the spatial
229 independence between the training and the test subset.(1) The random cross-validation procedure,
230 commonly used in SDMs, aims at randomly splitting the dataset into training and test subsets (Fielding and
231 Bell 1997, Hijmans 2012) which may lead to close spatial vicinity between the two datasets (Hijmans 2012),
232 and, (2) the spatial cross-validation procedure that aims at spatially spitting the dataset in order to reduce
233 spatial correlation and may improve independence between the two subsets (Hijmans 2012, Muscarella et
234 al. 2014).

235 The random procedure was therefore compared to four different spatial cross-validation procedures. (1) In
236 the 'BLOCK' method developed by Muscarella et al. (2014), different subsets of equal occurrence numbers
237 are created. For each replicate, this k-fold procedure divides the dataset into four equal subsets according to
238 the mean latitude and mean longitude positions of occurrence data (Fig. 1C), then three of these four
239 subsets are randomly selected to train the model (75%) and the last one is used to test the model (25%). (2)
240 In the 'CLOCK' methods, the dataset was divided according to random longitudinal transects, splitting the
241 Antarctic Circle into two parts (2-fold 'CLOCK' method, Fig. 1B), (3) three parts (3-fold 'CLOCK' method, Fig.
242 1D) or (4) four parts (4-fold 'CLOCK' method, Fig. 1E). In the 2-fold 'CLOCK' method, one subset was
243 considered as the training subset, the second one as the test subset; in the 3-fold 'CLOCK' method, two
244 subsets were defined for training and the third one for testing; in the 4-fold 'CLOCK' method, three subsets
245 were considered for training and one for testing (Fig. 1). Different cross-validation procedures were tested
246 using the 'gbm.step' procedure available in the *dismo* R package (Elith et al. 2008, Hijmans et al. 2016).
247 Once the dataset is split in different folds, Elith et al. (2008) apply an iterative procedure that enable to find
248 the minimum deviance to the test data, and relates it to the optimal number of trees (optimal model
249 complexity) to generate the model. If test and training data are spatially correlated, the number of trees
250 required to build BRTs will be overestimated. Therefore, the use of Elith et al. (2008) procedure will enable to
251 accurately interpret and compare optimal complexity and performance scores of models calibrated with
252 either randomly or spatially segregated folds (i.e. with contrasting distances between training and test
253 subsets), and thus will help explain the influence of occurrence spatial aggregation on model complexity and
254 performance.

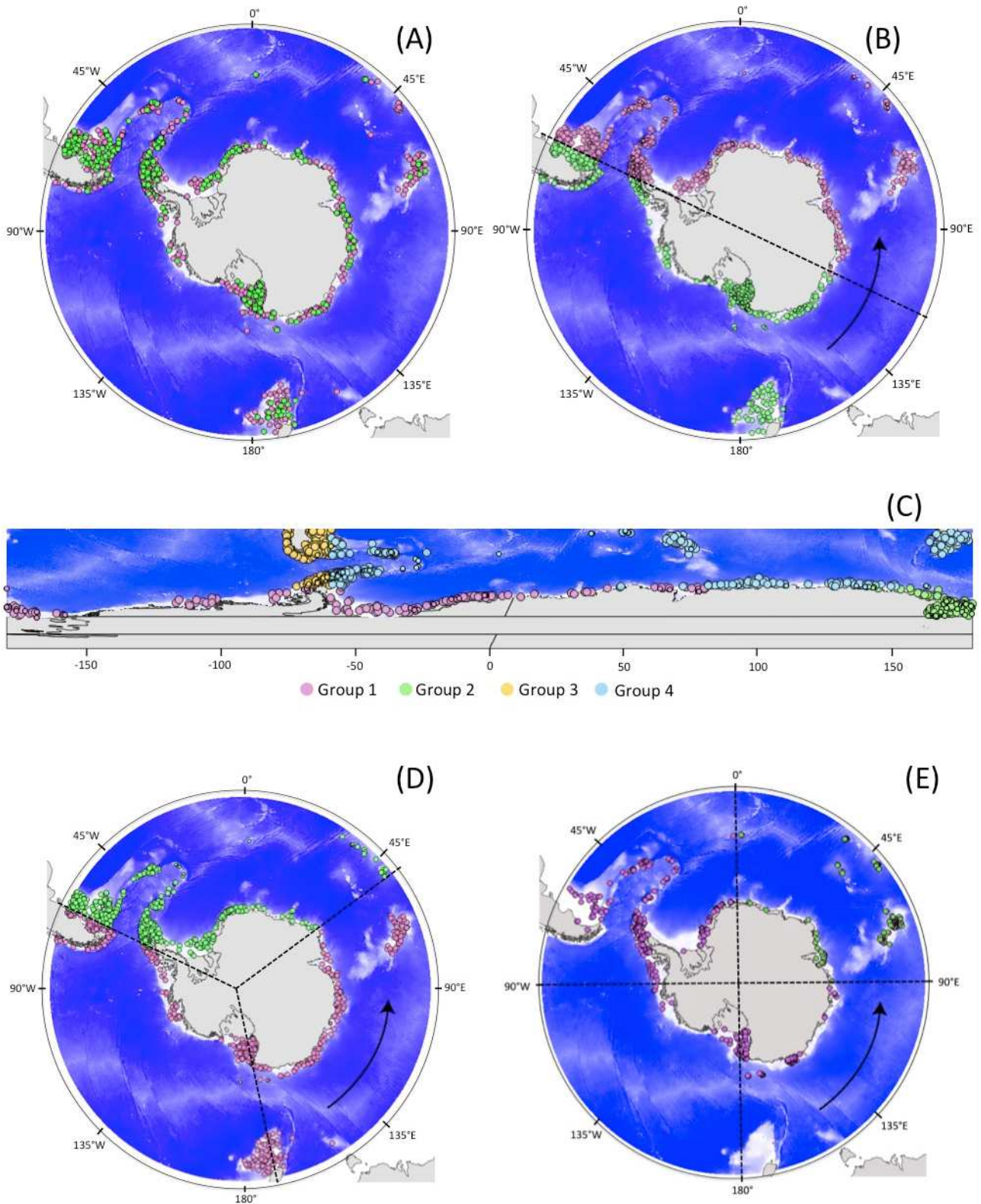
255 R scripts written to generate the models and the different cross-validation procedures are provided online at:
256 <https://github.com/charleneguillaumot/THESIS/>.

257 Independence between training and test subsets was evaluated using the Spatial Sorting Bias index (SSB)
258 (Hijmans 2012). SSB compares the distance between training-presence and testing-presence data with the
259 distance between training-presence and training-background. SSB~0 (non independence) means that the
260 "*distance between training-presence and test-presence sites will tend to be smaller than the distance*
261 *between training-presence and test-background sites*" (Hijmans 2012). SSB ~1 indicates that the two
262 distances are comparable (enough independent) (Hijmans 2012). SSB was calculated with the *dismo* R
263 package (Hijmans et al. 2016).

264 SDMs evaluation was generated by computing the percentage of test data that fall on grid-cell pixels
265 predicted as suitable. Suitable pixels were defined using the Maximum sensitivity plus specificity threshold
266 (MaxSSS) that splits models into suitable (>MaxSSS value) and unsuitable areas (<MaxSSS value).
267 MaxSSS is accepted as a relevant threshold for presence-only SDMs (Liu et al. 2013). The averaged optimal
268 number of trees required to generate BRTs was compared between models and used as a proxy of model
269 complexity.

270 Statistical differences between models generated with the different cross-validation procedures (AUC, TSS,
271 COR, percentage of correctly classified test data, number of trees) were tested using the non-parametric
272 Mann-Whitney Wilcoxon pairwise comparison.

273
274
275
276
277
278



279

280 **Figure 1:** Comparison of the different cross-validation procedures. Dots represent *Odontaster validus* presence-only
 281 records and a random set of 1,000 background data, sampled according to the Kernel Density Estimation weighting
 282 scheme. Colors indicate data splitting into training (pink) and test (green) subsets. Blue background corresponds to
 283 bathymetry and grey areas to emerged lands. For each case, 100 replicates of random background-data sampling and

284 transects partitioning are performed. (A) Random cross-validation procedure, with a random splitting into 75% training
285 and 25% test data. (B) '2-fold CLOCK' clustering by random spatial partition of the dataset into two groups (one training,
286 one test). (C) 'BLOCK' splitting, generated according to the median latitudinal and longitudinal values (Muscarella et al.
287 2014). After generation of four groups (corresponding to the four colors), one group is randomly defined as the test
288 subset, the other three groups as the training subset. A different system of projection was used to represent this map to
289 highlight the latitudinal and longitudinal definition of the transects. (D) '3-fold CLOCK' clustering by random spatial
290 partition of the dataset into three groups (2 training, 1 test). (E) '4-fold CLOCK' clustering by random spatial partition of
291 the dataset into four groups (3 training, 1 test).

292

293 2.5. Assessment of model uncertainty

294 The Multivariate Environmental Similarity Surface (MESS) index was estimated following the procedure
295 described by Elith et al. (2010) using the *dismo* R package (Hijmans et al. 2016). The MESS calculation
296 consists in extracting the environmental conditions where presence-only data were recorded and
297 determining for each pixel of the model projection layer if environmental conditions are covered by presence-
298 only records. Negative MESS values indicate areas of model extrapolation in which the value of at least one
299 environmental descriptor is beyond the environmental range covered by available presence-only records.
300 Conversely, positive MESS values indicate areas of model projection in which values of environmental
301 descriptors are within the environmental range covered by presence-only records. According to the number
302 of environmental descriptors that are not included inside the range of presence records values, MESS
303 outcome can strongly vary. The MESS evaluation deals with each environmental descriptor equally
304 (unweighted analysis) and in this study, a pixel was considered as unsuitable as soon as a single descriptor
305 value does not match the environmental range of presence-only records. On a projection map, SDMs
306 prediction was darkened according to the MESS extrapolation range to visualise the uncertain area due to
307 extrapolation. Extrapolation performance of SDMs was assessed by comparing the proportion of the
308 environment predicted as suitable by the model with the total set of environmental conditions.

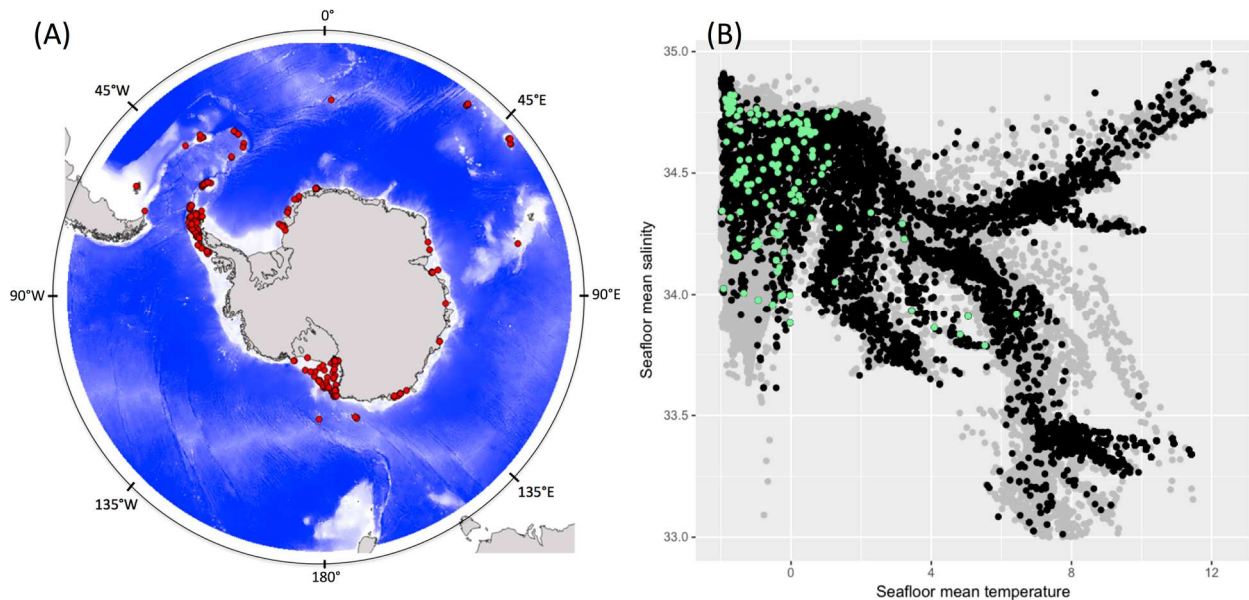
309

310 3. RESULTS

311 3.1. Available data and spatial autocorrelation

312 Distribution records available for *Odontaster validus* display a circumpolar and patchy spatial pattern (Fig.
313 2A). The niche occupied by *O. validus* does not cover the entire range of environmental conditions prevailing
314 in the projection area (Fig. 2B). *O. validus* is recorded in conditions close to zero and sub-zero seafloor
315 temperatures (Fig. 2B) and is mainly distributed in shallow and coastal areas. Most of *O. validus* presence

316 records are aggregated in regions where scientific benthic surveys are most often led and where sampling
317 effort was privileged due to access facilities (e.g. the Ross Sea and the Antarctic Peninsula). Overall, this
318 holds true for presence records of all benthic Southern Ocean taxa as well (Fig. S3), although, in this case,
319 most environmental conditions are covered by the total benthic samples (Fig. 2B).



320
321 **Figure 2:** (A) Presence-only records of the sea star *Odontaster validus* in the Southern Ocean. Duplicates (occurrences
322 falling on a same 0.1° resolution pixel) were removed from the display. (B) Values of the environmental range covered by
323 the entire benthos sampling dataset presented in Fig. S3 (black dots), by presence-only records of *O. validus* (green
324 dots) in comparison with a 1,000 background dots randomly sampled according to the Kernel Density Estimation scheme
325 (grey dots) for two environmental descriptors: mean seafloor temperature (°C) and mean seafloor salinity (PSU). A part
326 of the environment (grey dots) does not contain benthic occurrence samples (black dots), illustrating that sampling effort
327 is not geographically exhaustive.

328
329 Spatial autocorrelation was measured for both the total Southern Ocean benthic dataset (null models) and
330 for *O. validus* alone (models A and B) (Table 1). Moran's I scores were tested significant for all models, null
331 model #2 excepted. The absence of spatial autocorrelation ($I=0.005 \pm 0.004$; $p=0.19$) in null model #2 shows
332 that environmental data are not strongly aggregated in space. In contrast, presence-only records of the total
333 Southern Ocean benthic dataset are spatially aggregated. The degree of spatial aggregation due to sampling
334 effort is evidenced by the comparison between null model #1 and #2, scores of model #1 being 10 times
335 higher than those of null model #2 (Moran's $I=0.050 \pm 0.011$ and 0.005 ± 0.004 , respectively).

336 Values of Moran's I computed for models of *O. validus* (models A and B) are higher than those computed for
 337 the total Southern Ocean benthic dataset (null model #1 and #1 with Kernel Density Estimation). The
 338 sampling bias is therefore more pronounced for *O. validus* than for the majority of other benthic species.
 339 Model correction by the Kernel Density Estimation procedure was shown to reduce spatial autocorrelation
 340 with Moran's I values decreasing from 0.050 to 0.034 for null model #1, and from 0.085 to 0.069 for *O.*
 341 *validus* models A and B (Table 1). However, although lower, Moran's I values remain significant after
 342 correction, indicating that the applied corrections do not entirely remove the presence of spatial
 343 autocorrelation.

344

345 **Table 1:** Comparison between models of spatial autocorrelation values measured on model residuals (average and
 346 standard deviation of Moran's I values computed for 100 model replicates). Moran's I significance is indicated by p-
 347 values; for $p < 0.05$, the absence of spatial autocorrelation (null hypothesis) is rejected. Null model #1: 309 presence
 348 records were randomly sampled among occurrences of the total Southern Ocean benthic dataset (Fig. S3) and
 349 background data are composed of 1,000 points randomly sampled in the entire Southern Ocean; model #2: 309 records
 350 (to define presence records) and 1,000 background data both randomly sampled in the entire Southern Ocean; model #1
 351 with Kernel Density Estimation: similar to model null #1 but with 1,000 background data randomly sampled following the
 352 Kernel Density Estimation weighting scheme; model A: 309 presence records of *Odontaster validus* and 1,000
 353 background data were randomly sampled in the entire Southern Ocean; model B: similar to model A but with the 1,000
 354 background data sampled following the Kernel Density Estimation weighting scheme. AUC: Area Under the Receiver
 355 Operating Curve, TSS: True Skill Statistic, COR: Point Biserial Correlation.

356

	Null model #1	Null model #2	Null model #1 with Kernel Density Estimation	Model A	Model B
Spatial autocorrelation (Moran's I)	0.050 ± 0.011 <i>p</i> <0.001	0.005 ± 0.004 <i>p</i> =0.19	0.034 ± 0.011 <i>p</i> <0.001	0.085 ± 0.009 <i>p</i> <0.001	0.069 ± 0.006 <i>p</i> <0.001
AUC	0.976 ± 0.010	0.710 ± 0.014	0.964 ± 0.015	0.997 ± 0.001	0.948 ± 0.003
TSS	0.674 ± 0.013	0.331 ± 0.020	0.660 ± 0.019	0.698 ± 0.002	0.696 ± 0.003
COR	0.850 ± 0.028 <i>p</i> <0.001	0.336 ± 0.018 <i>p</i> <0.001	0.801 ± 0.037 <i>p</i> <0.001	0.944 ± 0.011 <i>p</i> <0.001	0.923 ± 0.015 <i>p</i> <0.001

357

358 *3.2. Comparison of cross-validation procedures*

359 For the BRT fitted with the random cross-validation procedure, all overall goodness-of-fit metrics (AUC, TSS,
360 COR) were good with predictive accuracy Area Under the Curve (AUC) values higher than 0.9 (Table 2).
361 However, when evaluated through spatial cross-validation procedure, the AUC scores decreased in all
362 BRTs. These results show that BRTs tend to overfit the data if the independence between training and test
363 data is not ensured. Indeed, the random cross-validation procedure presents SSB values close to zero,
364 indicating that training and test subsets may be highly correlated (Fig. 1A). In contrast, all spatial cross-
365 validation procedures have SSB values close to 1, indicating a better spatial independence between training
366 and test data (Table 2).

367 The generalisation performance (AUC and correctly classified test data) are very high for the random cross-
368 validation procedure, with more than 89.4% of test-presence records falling correctly in areas predicted as
369 suitable by the model (Table 2).

370 The random cross-validation procedure generates more complex BRTs compared to the spatial methods
371 (significantly higher number of trees for the random cross-validation procedure compared to the spatial
372 cross-validation procedures). As the model closely fits the dataset used for its construction, high AUC, TSS
373 and COR scores were obtained but these results may be misleading and overestimated. In contrast, spatial
374 cross-validation procedures generate less complex models (more general), which could account for lower
375 AUC, TSS and COR scores.

376

377

378

379

380

381

382

383

384

385

386

387

388

389 **Table 2:** Average Spatial Sorting Bias (SSB) and standard deviation values for the 100 model replicates (background
390 sampling + test/training clustering). AUC: Area Under the Receiver Operating Curve; Correctly classified test data (%):
391 percentage of presence-test and background-test records falling on predicted suitable areas (prediction > maximum
392 sensitivity plus specificity (maxSSS) threshold); TSS: True Skill Statistic; COR: Point Biserial Correlation; ntrees:
393 averaged optimal number of trees required to generate BRTs. Stars are indicated for spatial cross-validation groups
394 significantly different from the random cross-validation procedure (non-parametric pairwise Mann-Whitney Wilcoxon test,
395 p-value <0.01).

	Random cross-validation Random splitting	Spatial cross-validation Block method	Spatial cross-validation 2-fold Clock method	Spatial cross-validation 3-fold Clock method	Spatial cross-validation 4-fold Clock method
Mean SSB	0.101 ± 0.04	0.802 ± 0.37	0.832 ± 0.09	0.803 ± 0.23	0.848 ± 0.32
AUC	0.947 ± 0.013	0.854* ± 0.06	0.811* ± 0.053	0.818* ± 0.078	0.824* ± 0.089
Correctly classified test data (%)	89.452 ± 1.523	80.946* ± 7.504	80.039* ± 3.489	80.713* ± 5.421	79.471* ± 8.538
Test data (% of total dataset)	25%	[13-38]%	[19-81%]	[1-68%]	[1-66%]
TSS	0.715 ± 0.041	0.542* ± 0.188	0.465* ± 0.088	0.490* ± 0.136	0.576* ± 0.165
COR	0.792 ± 0.029	0.632* ± 0.126	0.584* ± 0.089	0.591* ± 0.12	0.483* ± 0.197
ntrees	1580 ± 251.058	543.5* ± 88.9	375* ± 91.9	424.5* ± 131.1	379* ± 98.5

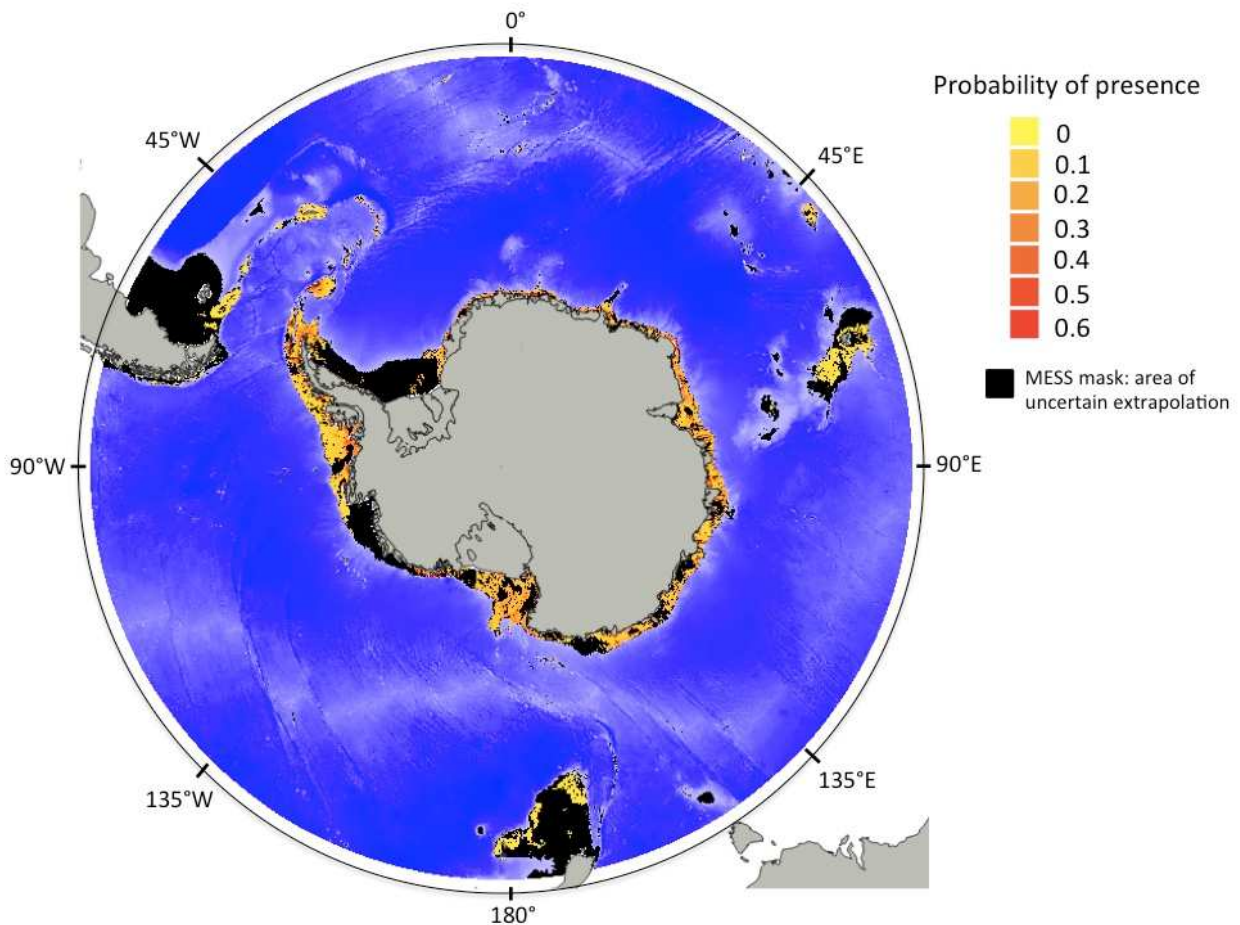
396

397 3.3. Proposed model and uncertainty map

398 We decided to maximise the spatial independence between training and test subsets, minimise model
399 complexity and optimise generalisation performances in *O. validus* model. Using these criteria, we found that
400 the ‘2-fold CLOCK’ modelling method was well adapted to *O. validus* dataset (second highest TSS and COR
401 scores; high proportion of test data being correctly classified, with the lowest standard deviation score (80.04
402 ± 3.49%); an important proportion of the total dataset used a test subset [19-81%] and the lowest model
403 complexity (ntrees = 375 ± 91.9).

404 The MESS index was calculated in order to define the part of this extrapolated area, that is, the part of the
405 geography for which at least one environmental descriptor is outside the environmental conditions of the

406 sampled presence records. The MESS index was compiled as a raster layer and projected on the probability
407 distribution map by darkening uncertain areas (Fig. 4). Uncertain areas due to extrapolation represent 64.2%
408 of the entire projected surface, the major part being also predicted by the model as unsuitable (Table 3).
409 Almost 9.5% of the area was however predicted as suitable by the model although considered as an
410 extrapolated area.
411



412
413 **Figure 4:** SDMs performed with the spatial cross-validation '2-fold CLOCK' method. Average of 100 model replicates.
414 Distribution probabilities are darkened according to the Multivariate Environmental Similarity Surface (MESS) layer, with
415 dark pixels corresponding to regions where the model extrapolates outside of the environmental conditions in which the
416 species was sampled. Dark pixels represent 64.2% of the entire projected area. Probabilities of presence are contained
417 between 0 and 1 but the colorbar was scaled until 0.6 to enhance visual contrast.
418
419
420
421
422

423 **Table 3:** Proportion of interpolated and extrapolated pixels according to the averaged SDMs predictions. Interpolation (or
 424 uncertain extrapolation respectively) refers to areas where environmental conditions within the pixel are inside (or
 425 outside, respectively) of the species ecological range, as defined by the Multivariate Environmental Similarity Surface
 426 (MESS). Suitable pixels were defined using the MaxSSS threshold that splits model predictions into suitable (>maxSSS
 427 mean score) or unsuitable areas (<maxSSS mean score).

MESS classification	Model prediction	
	Suitable pixels	Unsuitable pixels
Interpolation	10.24%	25.57%
Uncertain extrapolation	9.42%	54.77%

428

429 4. DISCUSSION

430 4.1. Evaluating SDM performance

431 Using independent datasets to test SDM performance is a prerequisite for relevant validation analyses
 432 (Peterson et al 2011). At broad spatial scale and in data-poor areas, the number of available data is limited
 433 and data distribution often patchy, which really challenges the success of validation procedures. Estimating
 434 the performance of SDMs predictions and the level of extrapolation in such areas is a necessity. The cross-
 435 validation procedure has been proposed as a reliable approach to evaluate modelling performances
 436 (Fielding and Bell 1997, Hijmans 2012, Dhingra et al. 2016, Roberts et al. 2017). Cross-validation
 437 procedures must however be adapted to spatially aggregated data because training and test subsets may be
 438 sampled in close areas, violating the independence assumption (Segurado et al. 2006, Hijmans 2012). Such
 439 a potential bias is rarely taken into account. In the present work, we compared SDM performance using five
 440 different cross-validation procedures for modelling, at broad spatial scale, the distribution of a species for
 441 which available data are limited in number and are spatially aggregated. Results show strong differences
 442 between procedures, which highlights the importance of testing and selecting the most appropriate method
 443 when evaluating model performance.

444

445 4.2. Correction for spatial autocorrelation and spatial bias

446 Strong significant Moran's I scores were measured on model residuals, revealing the presence of spatial
 447 autocorrelation in the total Southern Ocean benthic dataset (Fig. S3). The difference between null models #1
 448 and #2 evidences the influence of sampling aggregation on spatial autocorrelation values (Table 1) as

449 discussed by [Guillaumot et al. \(2018\)](#). *O. validus* presence-only dataset follows the same pattern, with
450 records aggregated in coastal areas where sampling effort has been mostly concentrated (Table 1, Fig. 2). A
451 target-group background sampling was applied and proved to be efficient to reduce spatial autocorrelation
452 (as assessed using Moran's I statistic), although it still remains at a significant level. Spatial autocorrelation
453 scores are strongly dependent on the resolution of environmental raster layers. The coarse resolution of
454 environmental data used in the present study may be responsible for the over-estimation of spatial
455 autocorrelation scores. This could account for spatial autocorrelation remaining significant even after the
456 Kernel Density Estimation correction.

457

458 4.3. Selection of cross-validation procedures

459 The random cross-validation procedure has been widely used in ecological modelling to evaluate model
460 predictions ([Fielding and Bell 1997](#), [Merow et al. 2013](#), [Mainali et al. 2015](#), [Torres et al. 2015](#), [Phillips et al.](#)
461 [2017](#)) but the method has been rarely compared to alternative procedures. The present study shows that
462 contrasting model assessments are obtained when using different cross-validation procedures
463 ([Radosavljevic and Anderson 2014](#), [Roberts et al. 2017](#)). Applying a random cross-validation to an
464 aggregated dataset at a broad spatial scale can result in training and test subsets being sampled in the
465 same area, and leads to an inflation of modelling performances ([Veloz 2009](#), [Hijmans 2012](#), [Radosavljevic](#)
466 [and Anderson 2014](#), [Wenger and Olden 2012](#)). In the context of this study, SDMs produced with a broad-
467 scale and spatially aggregated occurrence dataset and a random cross-validation procedure are more
468 complex and likely over-fit the training dataset. This also may account for the high evaluation scores
469 obtained (AUC, TSS, COR) and may also explain the apparent high generalisation performance of BRTs
470 fitted with random cross-validation. The lack of model generality can *a posteriori* lead to strong caveats and
471 unreliable models with poor transferability performance when projected on a new environmental space
472 ([Wenger and Olden 2012](#), [Crimmins et al. 2013](#)). Methods that select the most parsimonious BRT, combine
473 low model complexity and high modelling performance should therefore be preferred.

474

475 The spatial cross-validation procedures tested in this study were shown to produce less complex models
476 than the random cross-validation procedure. Increased model generality (i.e. decrease in model over-fitting)
477 and forced spatial segregation between training and test subsets result in decreasing SDMs validation
478 scores. These results show that applying a random cross-validation procedure for a patchy dataset can lead
479 to over-estimation of SDMs predictive performance if training and test subsets are not independent. This is in
480 line with several works ([Brenning et al., 2005](#), [Elith et al. 2010](#), [Andersen 2013](#), [Muscarella et al. 2014](#)) in

481 which a decrease of AUC scores can be reported when using a spatial cross-validation procedure instead of
482 a random procedure. Machine-learning algorithms have been reported to be the best approaches to
483 generate SDMs but the influence of over-fitting on model evaluation are underestimated ([Reiss et al. 2011](#),
484 [Duan et al. 2014](#), [Beaumont et al. 2016](#), [Thuiller et al. 2016](#), [Guillaumot et al. 2018](#)) although its effect has
485 been pointed out in several works ([Elith et al. 2008](#), [Jimenez-Valverde 2008](#), [Wenger and Olden 2012](#)). Our
486 results show that the evaluation of SDMs performance can be strongly influenced by the choice of the
487 evaluation procedure.

488

489 In this work, several spatial cross-validation procedures were compared with each other but no single and
490 best procedure emerged, a common case in ecological modelling ([Qiao et al. 2015](#)). The appropriate method
491 to be used is highly dependent on the species and dataset under study. For instance, the 'BLOCK' method
492 introduced by [Muscarella et al. \(2014\)](#) should not be used at broad spatial scale, where too important
493 latitudinal contrasts in environmental conditions are present. In this study, such contrasting environmental
494 conditions (due to the presence of an environmental latitudinal gradient between sub-Antarctic and Antarctic
495 regions, with occurrence aggregation in the two regions) lead to higher variability in generalisation
496 performance during model projection, depending on the data subsets selected to train and test the model
497 ([Roberts et al. 2017](#)). The 'BLOCK' method favors the independence between training and test subsets but
498 models are slightly more complex because they are calibrated on contrasting environmental conditions (sub-
499 Antarctic vs. Antarctic areas) and over-fit the training dataset that could also present a patchy distribution.
500 The 'BLOCK' method is therefore more adapted to case studies without strong patchy and contrasting
501 environmental conditions. The 'CLOCK' procedures developed in this study helped reduce the effect of
502 latitudinal patchy occurrences distribution by mixing presence records sampled in Antarctic and sub-Antarctic
503 regions to define training and test subsets. The 'CLOCK' methods generate less complex models and were
504 proved more efficient to define spatially independent training and test subsets. However, the number of
505 training and test records sampled between model replicates is not constant, which contributes to an
506 important variability in validation performance scores. The selection of the different 'CLOCK' methods also
507 depends on the importance of data aggregation and patchy patterns within environmental conditions. For
508 strong data aggregation, the '2-fold CLOCK' approach will help reduce the influence of patchy patterns
509 during model calibration and will help generalise the model and decrease its complexity. '3 or 4-fold CLOCK'
510 methods present close modelling performances but the proportion of occurrence records used to test the
511 model might be very low.

512 Alternative SDM evaluation procedures can be found in the literature: for instance, calibrated cross-validation
513 procedures aim at removing occurrences from the test subset when considered too close to the training
514 subset (and considered as non-informative according to a statistical threshold) (Hijmans 2012). For limited
515 presence-only datasets, removing a part of the available occurrence data may lead to the removal of a
516 proportion of informative records, which does not constitute a reasonable option (Bean et al. 2012, van
517 Proosdij et al. 2016). The leave-one-out method can also provide a relevant estimate of model goodness-of-
518 fit, even for spatially aggregated datasets (Olden et al. 2002, Wenger and Olden 2012). The method aims at
519 randomly excluding a single record from the total dataset. The model is trained on the remaining data and
520 predicts the model response on the single removed point to test for model prediction. The procedure is
521 replicated several times, providing a powerful evaluation of model accuracy. However, assessment of
522 generalisation performances is not permitted with this approach (Wenger and Olden 2012).

523

524 In addition to cross-validation procedures, the relevance of model validation performance is also strongly
525 dependent on the quality of environmental descriptors available. The number of no-data pixels as well as
526 grid-cell resolution can critically affect model evaluation. This is especially true in the present study because
527 environmental variables, measured or interpolated, rarely extend to coastal areas, and resolution in the
528 Southern Ocean can rarely be better than 10 km². Good quality datasets are needed and such limitations
529 must be taken into account when interpreting model outputs.

530

531 *4.4. Uncertainty assessment in SDMs predictions*

532 SDM uncertainty assessment has been a widely discussed topic (Barry and Elith 2006, Carvalho et al. 2011,
533 Beale and Lennon 2012, Guisan et al. 2013). Uncertainty in model predictions has been often assessed as
534 the variation among the predicted distribution probabilities (Buisson et al 2010) but this approach does not
535 provide precise information on the origin of uncertainty (Tessarolo et al. 2014).

536 The MESS metric is a relevant indicator of SDMs extrapolation performance (Elith et al. 2010, Dhingra et al.
537 2016). The Mobility Oriented Parity (MOP) introduced by Owens et al. (2013) was recently proposed as an
538 alternative to the MESS index. MESS considers extrapolation on a pixel as uncertain when at least one
539 environmental value falls outside the environmental range of presence records. In contrast, MOP offers more
540 flexibility by defining an extrapolated area when all environmental values fall outside the sampled
541 environmental range. Therefore, MESS is more conservative than MOP to define species ecological
542 envelope.

543 Here, MESS was used to assess the proportion of the projected area for which models extrapolate. Our
544 results show that more than half of the area corresponds to environmental conditions for which presence
545 records have not been sampled. 9.5% of this extrapolated area is even predicted as a suitable environment.
546 This highlights the weakness of SDMs for spatial generalisation and the risk of providing inaccurate SDMs
547 for conservation purposes, especially if the communication between modellers and environmental managers
548 is neglected (Guisan et al. 2013). Our results show the importance of providing uncertainty maps along with
549 SDM outputs in order to help interpret models with the necessary caution.

550

551 5. Conclusion

552 This work highlights the importance of assessing the relevance of SDM evaluation procedures. When
553 applied to occurrence datasets, spatially autocorrelated and broad-scale presence-only datasets, the random
554 cross-validation procedure may over-estimate model validation scores due to the violation of independence
555 between training and test subsets. Applying a spatial cross-validation procedure that spatially segregates
556 training and test data was shown to be effective to provide a reliable analysis of model performance. Spatial
557 cross-validation methods also help reduce model complexity and therefore improve generalisation
558 performances. The 'CLOCK' methods developed in this paper were proved to be appropriate to our Southern
559 Ocean case study and could be applied to other non-polar case studies. This study proves the importance of
560 testing and comparing several spatial cross-validation procedures to identify the procedure most adapted to
561 each case study.

562 The MESS index was used to visualise areas where SDMs extrapolate outside the range of the
563 environmental conditions where presence records were sampled. Such results show the importance of
564 providing information on model uncertainty to correctly interpret SDM outputs.

565

566 6. Acknowledgements

567 This work was supported by a "Fonds pour la formation à la Recherche dans l'Industrie et l'Agriculture"
568 (FRIA) grant to C. Guillaumot. This is contribution no. 26 to the vERSO project (www.versoproject.be),
569 funded by the Belgian Science Policy Office (BELSPO, contract n°BR/132/A1/vERSO). We are also thankful
570 to the anonymous reviewers that help improve this work with relevant remarks and advices.

571

572 7. Authors' contributions

573 CG, JA, TS conceived the ideas and designed the methodology; LD provided a part of the data; CM, ME, AA
574 and BD contributed to the writing of the manuscript. All authors contributed critically to the drafts and gave
575 final approval for publication.

576

577 8. References

578

579 [Allouche](#), O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models:
580 prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232.

581

582 [Agüera](#), A., Collard, M., Jossart, Q., Moreau, C., & Danis B. (2015) Parameter estimations of Dynamic
583 Energy Budget (DEB) model over the life history of a key Antarctic species: the Antarctic sea star *Odontaster*
584 *validus* Koehler, 1906. *Plos One*, 10(10), e0140078.

585 [Anderson](#), R. P., & Gonzalez Jr, I. (2011). Species-specific tuning increases robustness to sampling bias in
586 models of species distributions: an implementation with Maxent. *Ecological Modelling*, 222(15), 2796-2811.

587

588 Anderson, R. P. (2013). A framework for using niche models to estimate impacts of climate change on
589 species distributions. *Annals of the New York Academy of Sciences*, 1297(1), 8-28.

590

591 [Araujo](#), M.B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of*
592 *Biogeography*, 33(10), 1677-1688.

593

594 [Barbet-Massin](#), M., Jiguet, F., Albert, C.H., & Thuiller, W. (2012). Selecting pseudo-absences for species
595 distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), 327-338.

596

597 [Barry](#), S., & Elith, J. (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology*, 43(3), 413-
598 423.

599

600 [Beale](#), C.M., & Lennon, J.J. (2012). Incorporating uncertainty in predictive species distribution modelling.
601 *Philosophical Transaction of the Royal Society B*, 367(1586), 247-258.

602

603 [Bean](#), W.T., Stafford, R., & Brashares, J.S. (2012). The effects of small sample size and sample bias on
604 threshold selection and accuracy assessment of species distribution models. *Ecography*, 35(3), 250-258.

605

606 [Beaumont](#), L.J., Graham, E., Duursma, D.E., Wilson, P.D., Cabrelli, A., Baumgartner, J.B., ... & Laffan, S.W.
607 (2016). Which species distribution models are more (or less) likely to project broad-scale, climate-induced
608 shifts in species ranges? *Ecological Modelling*, 342, 135-146.

609

610 [Boria](#), R.A., Olson, L.E., Goodman, S.M., & Anderson, R.P. (2014). Spatial filtering to reduce sampling bias
611 can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73-77.

612

613 [Buisson](#), L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting
614 of species distribution. *Global Change Biology*, 16(4), 1145-1157.

615

616 Brenning, A. (2005). Spatial prediction models for landslide hazards: review, comparison and evaluation.
617 *Natural Hazards and Earth System Science*, 5(6), 853-862.

618

619 [Brenning](#), A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote
620 sensing: The R package *sperrorest*. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE*
621 *International* (pp. 5372-5375). IEEE.

622

623 [Brown](#), J. L., Bennett, J. R., & French, C. M. (2017). SDMtoolbox 2.0: the next generation Python-based GIS
624 toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, 5, e4095.

625

626 [Carvalho](#), S.B., Brito, J.C., Crespo, E.G., Watts, M.E., & Possingham, H.P. (2011). Conservation planning
627 under climate change: toward accounting for uncertainty in predicted species distributions to increase
628 confidence in conservation investments in space and time. *Biological Conservation*, 144(7), 2020-2030.

629

630 [Cliff](#), A., Ord, J.K. (1981). *Spatial Processes Models and Applications*. Pion Ltd.

631

632 [Cruse](#), B., Liedloff, A.C. & Wintle, B.A. (2012) A new method for dealing with residual spatial autocorrelation
633 in species distribution models. *Ecography*, 35(10), 879-888.

634

635 [Crimmins](#), S. M., Dobrowski, S. Z., & Mynsberge, A. R. (2013). Evaluating ensemble forecasts of plant
636 species distributions under climate change. *Ecological modelling*, 266, 126-130.

637

638 [Dhingra](#), M.S., Artois, J., Robinson, T.P., Linard, C., Chaiban, C., Xenarios, I., ... & Von Dobschuetz, S.
639 (2016). Global mapping of highly pathogenic avian influenza H5N1 and H5Nx clade 2.3. 4.4 viruses with
640 spatial cross-validation. *elife*, 5, e19571.

641

642 [Dormann](#), C.F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution
643 data. *Global Ecology and Biogeography*, 16(2),129-138.

644

645 [Dormann](#), C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Münkemüller, T. (2013).
646 Collinearity: a review of methods to deal with it and a simulation study evaluating their performance.
647 *Ecography*, 36(1), 27-46.

648

649 [Duan](#), R.Y., Kong, X.Q., Huang, M.Y., Fan, W.Y., & Wang, Z.G. (2014). The predictive performance and
650 stability of six species distribution models. *PLoS One*, 9(11), e112764.

651

652 [Duque-Lazo](#), J., Van Gils, H.A., Groen, T.A. & Navarro-Cerrillo, R.M. (2016) Transferability of species
653 distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia.
654 *Ecological Modeling*, 320, 62-70.

655

656 [Elith](#), J., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R., & Loiselle, B. (2006). Novel methods
657 improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.

658

659 [Elith](#), J., Leathwick, J.R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal*
660 *Ecology*, 77(4), 802-813.

661

662 [Elith](#), J., & Graham, C.H. (2009). Do they? How do they? Why do they differ? On finding reasons for differing
663 performances of species distribution models. *Ecography*, 32(1), 66-77.

664

665 [Elith](#), J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology*
666 *and Evolution*, 1(4), 330-342.

667

668 [Fabri-Ruiz](#), S., Saucède, T., Danis, B., & David, B. (2017). Southern Ocean Echinoids database—An updated
669 version of Antarctic, Sub-Antarctic and cold temperate echinoid database. *ZooKeys*, (697), 1.

670

671 [Ficetola](#), G.F., Thuiller, W., & Miaud, C. (2007). Prediction and validation of the potential global distribution of
672 a problematic alien invasive species—the American bullfrog. *Diversity and Distributions*, 13(4), 476-485.

673

674 [Fielding](#), A.H., & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in
675 conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49.

676 Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New
677 York: Springer series in statistics.

678

679 [Griffiths](#), H.J., Van de Putte, A.P., & Danis, B. (2014). CHAPTER 2.2. Data distribution: Patterns and
680 implications. In: De Broyer, C., Koubbi, P., Griffiths, H.J., Raymond, B., Udekem d'Acoz, C. d', et al. (eds.).
681 Biogeographic Atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge, pp. 16-
682 26.

683 [Guillaumot](#), C., Martin, A., Eléaume, M. & Saucède, T. (2018). Methods for improving species distribution
684 models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau. in press.
685 *Mar. Ecol. Prog. Ser.* 594:149-164

686

687 [Guillera-Arroita](#), G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., ... & Wintle, B.A.
688 (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global
689 Ecology and Biogeography*, 24(3), 276-292.

690

691 [Guisan](#), A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models.
692 *Ecology letters*, 8(9), 993-1009.

693

694 [Guisan](#), A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A. I., ... & Martin, T.
695 G. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16(12), 1424-1435.

696

697 [Gutt](#), J., Zurell, D., Bracegridle, T., Cheung, W., Clark, M., Convey, P., ... & Griffiths, H. (2012). Correlative
698 and dynamic species distribution modelling for ecological predictions in the Antarctic: a cross-disciplinary
699 concept. *Polar Research*, 31(1), 11091.

700

701

702 [Heikkinen](#), R.K., Marmion, M., & Luoto, M. (2012). Does the interpolation accuracy of species distribution
703 models come at the expense of transferability? *Ecography*, 35(3), 276-288.

704

705 [Hijmans](#), R.J, Phillips, S., Leathwick, J. and Elith, J. (2016), Package 'dismo'. Available online at:
706 <http://cran.r-project.org/web/packages/dismo/index.html>.

707

708 [Hijmans](#), R.J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and
709 calibration with a null model. *Ecology*, 93(3), 679-688.

710

711 [Hortal](#), J., Lobo, J. M., & Jiménez-Valverde, A. (2012). Basic questions in biogeography and the (lack of)
712 simplicity of species distributions: putting species distribution models in the right place. *Natureza &*
713 *Conservação*, 10(2), 108-118.

714

715 [Jiménez-Valverde](#), A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: the importance of
716 concepts in species distribution modelling. *Diversity and distributions*, 14(6), 885-890.

717

718 [Jiménez-Valverde](#), A. (2012). Insights into the area under the receiver operating characteristic curve (AUC)
719 as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21(4), 498-
720 507.

721

722 [Lawrence](#), J.M. (2013) *Starfish: Biology and ecology of the Asteroidea*. JHU Press, Baltimore, 267p.

723 Legendre, P., & Fortin, M.J. (1989). Spatial pattern and ecological analysis. *Vegetatio*, 80(2), 107-138.

724

725 [Li](#), W., & Guo, Q. (2013). How to assess the prediction accuracy of species presence–absence models
726 without absence data? *Ecography*, 36(7), 788-799.

727

728 [Liu](#), C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with
729 presence-only data. *Journal of Biogeography*, 40(4), 778-789.

730

731 [Mainali](#), K.P., Warren, D.L., Dhileepan, K., McConnachie, A., Strathie, L., Hassan, G., ... & Parmesan, C.
732 (2015). Projecting future expansion of invasive species: comparing and improving methodologies for species
733 distribution modeling. *Global Change Biology*, 21(12), 4464-4480.

734

735 [McClintock](#), J.B., Angus, R.A., Ho, C., Amsler, C.D., & Baker, B.J. (2008) A laboratory study of behavioral
736 interactions of the Antarctic keystone sea star *Odontaster validus* with three sympatric predatory sea stars.
737 *Marine Biology*, 154(6), 1077-1084

738

739 [Merow](#), C., Smith, M.J., & Silander, J.A. (2013). A practical guide to MaxEnt for modeling species'
740 distributions: what it does, and why inputs and settings matter. *Ecography*, 36(10), 1058-1069.

741

742 [Moreau](#), C., Mah, C., Agüera, A., Améziiane, N., Barnes, D., Crokaert, G., ... & Jażdżewska, A. (2018).
743 Antarctic and sub-Antarctic Asteroidea database. *ZooKeys*, (747), 141.

744

745 [Muscarella](#), R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., & Anderson, R.P.
746 (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal
747 model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5(11), 1198-1205.

748

749 [Naimi](#), B., Hamm, N.A., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014). Where is positional
750 uncertainty a problem for species distribution modelling? *Ecography*, 37(2), 191-203.

751

752 [Newbold](#), T. (2010). Applications and limitations of museum data for conservation and ecology, with
753 particular attention to species distribution models. *Progress in Physical Geography*, 34(1), 3-22.

754

755 [Olden](#), J.D., Jackson, D.A., & Peres-Neto, P.R. (2002). Predictive models of fish species distributions: a note
756 on proper validation and chance predictions. *Transactions of the American Fisheries Society*, 131(2), 329-
757 336.

758

759 [Owens](#), H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J., ... & Peterson, A.T. (2013).
760 Constraints on interpretation of ecological niche models by limited environmental ranges on calibration
761 areas. *Ecological Modelling*, 263, 10-18.

762

763 [Peterson](#), A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo,
764 M. B. (2011). *Ecological niches and geographic distributions* (MPB-49). Princeton University Press.
765

766 [Phillips](#), S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample
767 selection bias and presence-only distribution models: implications for background and pseudo-absence
768 data. *Ecological Applications*, 19(1), 181-197.
769

770 [Phillips](#), N.D., Reid, N., Thys, T., Harrod, C., Payne, N.L., Morgan, C.A., ... & Houghton, J.D. (2017).
771 Applying species distribution modelling to a data poor, pelagic fish complex: The ocean sunfishes. *Journal of*
772 *Biogeography*, 44(10), 2176-2187.
773

774 [Pierrat](#), B., Saucède, T., Laffont, R., De Ridder, C., Festeau, A., & David, B. (2012). Large-scale distribution
775 analysis of Antarctic echinoids using ecological niche modelling. *Marine Ecology Progress Series*, 463, 215-
776 230.
777

778 [Qiao](#), H., Soberón, J., & Peterson, A.T. (2015). No silver bullets in correlative ecological niche modelling:
779 insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and*
780 *Evolution*, 6(10), 1126-1136.
781

782 [R Core Team](#), (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical
783 Computing, Vienna, Austria. 2016.
784

785 [Radosavljevic](#), A., & Anderson, R.P. (2014). Making better Maxent models of species distributions:
786 complexity, overfitting and evaluation. *Journal of Biogeography*, 41(4), 629-643.
787

788 [Reiss](#), H., Cunze, S., König, K., Neumann, H., & Kröncke, I. (2011). Species distribution modelling of marine
789 benthos: a North Sea case study. *Marine Ecology Progress Series*, 442, 71-86.
790

791 [Ripley](#), B. (2015). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. 2015. *URL*
792 *https://CRAN.R-project.org/package=MASS.R.package.version,7-3*.
793

794 [Roberts](#), D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... & Warton, D. I. (2017).
795 Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*,
796 40(8), 913-929.
797
798 [Robinson](#), L. M., Elith, J., Hobday, A. J., Pearson, R. G., Kendall, B. E., Possingham, H. P., & Richardson, A.
799 J. (2011). Pushing the limits in marine species distribution modelling: lessons from the land present
800 challenges and opportunities. *Global Ecology and Biogeography*, 20(6), 789-802.
801
802 [Segurado](#), P.A., Araújo, M.B., & Kunin, W.E. (2006). Consequences of spatial autocorrelation for niche-
803 based models. *Journal of Applied Ecology*, 43(3), 433-444.
804
805 [Tessarolo](#), G., Rangel, T.F., Araújo, M.B., & Hortal, J. (2014). Uncertainty associated with survey design in
806 Species Distribution Models. *Diversity and Distributions*, 20(11), 1258-1269.
807
808 [Thuiller](#) W, Georges D, Engler R and Breiner F (2016) biomod2: Ensemble Platform for Species Distribution
809 Modeling. R package version 3.3-7. <https://CRAN.R-project.org/package=biomod2>
810
811 [Torres](#), L.G., Sutton, P.J., Thompson, D.R., Delord, K., Weimerskirch, H., Sagar, P.M., ... & Phillips, R.A.
812 (2015). Poor transferability of species distribution models for a pelagic predator, the grey petrel, indicates
813 contrasting habitat preferences across ocean basins. *PLoS One*, 10(3), e0120014.
814
815 [Valavi](#), R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2018). blockCV: an R package for
816 generating spatially or environmentally separated folds for k-fold cross-validation of species distribution
817 models. *bioRxiv*, 357798.
818
819 [van Proosdij](#), A.S., Sosef, M.S., Wieringa, J.J., & Raes, N. (2016). Minimum required number of specimen
820 records to develop accurate species distribution models. *Ecography*, 39(6), 542-552.
821
822 [Veloz](#), S.D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-
823 only niche models. *Journal of Biogeography*, 36(12), 2290-2299.
824

825 [Wenger](#), S.J., & Olden, J.D. (2012). Assessing transferability of ecological models: an underappreciated
826 aspect of statistical validation. *Methods in Ecology and Evolution*, 3(2), 260-267.
827