



# Selecting environmental descriptors is critical for modelling the distribution of Antarctic benthic species

Guillaumot Charlène<sup>1,2</sup> · Danis Bruno<sup>1</sup> · Saucède Thomas<sup>2</sup>

Received: 8 October 2019 / Revised: 12 June 2020 / Accepted: 25 June 2020 / Published online: 6 July 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

Species distribution models (SDMs) are increasingly used in ecological and biogeographic studies by Antarctic biologists, including for conservation and management purposes. During the modelling process, model calibration is a critical step to ensure model reliability and robustness, especially in the case of SDMs, for which the number of selected environmental descriptors and their collinearity is a recurring issue. Boosted regression trees (BRT) was previously considered as one of the best modelling approach to correct for this type of bias. In the present study, we test the performance of BRT in modelling the distribution of Southern Ocean species using different numbers of environmental descriptors, either collinear or not. Models are generated for six sea star species with contrasting ecological niches and wide distribution ranges over the entire Southern Ocean. For the six studied species, overall modelling performance is not affected by the number of environmental descriptors used to generate models, BRT using the most informative descriptors and minimizing model overfitting. However, removing collinear descriptors also helps reduce model overfitting. Our results confirm that BRTs may perform well and are relevant to deal with complex and redundant environmental information for Antarctic biodiversity distribution studies. Selecting a limited number of non-collinear descriptors before modelling may generate simpler models and facilitate their interpretation. The modelled distributions do not differ noticeably between the different species despite contrasting species ecological niches. This unexpected result stresses important limitations in using SDMs for broad scale spatial studies, based on limited, spatially aggregated data, and low-resolution descriptors.

**Keywords** Species distribution models (SDMs) · Boosted regression trees (BRT) · Southern ocean · Collinearity · Asteroidea · Conservation · Environmental descriptors

## Introduction

The Southern Ocean is one of the regions on Earth that is undergoing climate change at the fastest pace (Convey et al. 2009; Turner et al. 2014; Henley et al. 2019). Predicting the response of Antarctic species and communities to environmental changes is challenging but it has become a pressing

need to address conservation issues and support guidance for the management of living resources in a dynamic context (Gutt et al. 2012; Ingels et al. 2012; Constable et al. 2014; De Broyer et al. 2014; Convey and Peck 2019). The Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) actively works for the sustainable management of Antarctic marine ecosystems and marine life (see <https://www.ccamlr.org/en/organisation>, Access August 2019). Recent proposals from CCAMLR and existing marine protected areas (MPAs), such as those newly designated around the South Orkney Islands or in the Ross Sea (CCAMLR 2009, 2016), partly rely on species distribution modelling (SDM) (Ballard et al. 2012; Anderson et al. 2016; Davis et al. 2017; Arthur et al. 2018).

SDM is a correlative approach that depicts the relationship between the distribution of species occurrence records and a set of environmental descriptors, to interpolate and predict the potential distribution of species over their entire

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00300-020-02714-2>) contains supplementary material, which is available to authorized users.

✉ Guillaumot Charlène  
charleneguillaumot21@gmail.com

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

<sup>2</sup> UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, 6 bd Gabriel, 21000 Dijon, France

distribution range (Elith et al. 2006; Peterson et al. 2011). Over the last decades, SDMs have been increasingly used to address conservation issues (Guisan et al. 2013; Ross and Howell 2013; Marshall et al. 2014; Reiss et al. 2014; Arthur et al. 2018), predict species suitable areas (Meier et al. 2011; Reiss et al. 2011; Nachtsheim et al. 2017; Phillips et al. 2017), including potential distribution shifts (Ficetola et al. 2007; Václavík and Meentemeyer 2009; Jiménez-Valverde et al. 2011; Tingley et al. 2014), and guide sustainable management plans for commercial purposes (Valavanis et al. 2008; Maxwell et al. 2009). They have particularly proved useful to improve our understanding of species distribution in poorly sampled and seldom accessed areas (Elith et al. 2006; Peterson et al. 2011) and for the conservation of Southern Ocean marine life (De Broyer et al. 2014; Basher and Costello 2016; Hogg et al. 2018; Jansen et al. 2018; Jerosch et al. 2019).

Calibration is a critical step in SDM procedures, influencing their relevance, robustness and accuracy (Barbet-Massin et al. 2012; Guisan et al. 2013; Anderson et al. 2016). The selection of environmental descriptors is also important, as it shapes model accuracy and performance (Elith and Leathwick 2009; Austin and van Niel 2011; Dormann et al. 2012; Braunisch et al. 2013; Bucklin et al. 2015; Bradie and Leung 2017; Petitpierre et al. 2017). The inappropriate selection of descriptors has been shown to cause overfitting in SDMs, especially when the number of descriptors is high compared to the number of occurrences available (Anderson and Gonzalez 2011; Synes and Osborne 2011; Braunisch et al. 2013; Kramer-Schadt et al. 2013; Petitpierre et al. 2017), leading to over-complex models, reduced transferability performances and underestimation of predicted suitable areas (Beaumont et al. 2005).

Collinearity between descriptors is another major concern when addressing the quality of SDMs (Dormann et al. 2012). Collinearity occurs when at least two descriptors are linearly related in a statistical model (Dormann et al. 2012). In regression models, multicollinearity increases variance values between independent descriptors. It can cause incorrect estimations of beta regression coefficients and bias interpretation, making it difficult to disentangle the respective contributions of independent variables to explaining the dependent variable (Hair et al. 2014). Collinear descriptors are traditionally removed from datasets to calibrate SDMs (Dormann et al. 2012; Pierrat et al. 2012; Merow et al. 2013; Fabri-Ruiz et al. 2018; Guillaumot et al. 2018b), while a recent study showed that collinear descriptors could also improve the model's fit (Freer et al. 2019).

Machine-learning algorithms can effectively model complex relationships between environmental conditions and occurrence records (Olden et al. 2008; Elith and Leathwick 2009). They can harness incomplete datasets and missing data, as well as contrasting and extreme values, and generate

predictive models with high transferability performances and low sensitivity to species niche width (Elith et al. 2006, 2008; Elith and Graham 2009; Reiss et al. 2011; Barbet-Massin et al. 2012; Heikkinen et al. 2012; Qiao et al. 2015). In machine-learning algorithms, the Boosted Regression Trees approach (BRT) has been shown to be particularly efficient when dealing with non-informative environmental descriptors or conversely, with redundant information provided by correlated factors (Breiman et al. 1984; De'ath and Fabricius 2000; Elith et al. 2008).

In the present work, we test the robustness of SDMs generated with BRT for various numbers of environmental descriptors and different collinearity values. Models are generated for six common and abundant asteroid (sea star) species that have been extensively sampled and studied; here used as representative case studies for the Antarctic benthos: *Acodontaster hodgsoni* (Bell 1908), *Bathybiaster loripes* (Sladen 1889), *Glabraster antarctica* (Smith 1876), *Labidiaster annulatus* (Sladen 1889), *Odontaster validus* (Koehler 1906) and *Psilaster charcoti* (Koehler 1906) (McClintock et al. 2008; Mah and Blake 2012; Lawrence 2013; Brandt et al. 2014; Danis et al. 2014; Moles et al. 2015; Moreau et al. 2018).

Because the Southern Ocean is scarcely accessed and sampled, spatial analyses of species distribution are usually based on aggregated and relatively small presence-only datasets, often compiled from historical records (De Broyer et al. 2014; Guillaumot et al. 2016, 2018a; Fabri-Ruiz et al. 2017; Moreau et al. 2018), which strongly hampers SDM performances (Hortal et al. 2008; Loiselle et al. 2008; Phillips et al. 2009; Costa et al. 2010; Newbold 2010; Guillera-Arroita et al. 2015; Guillaumot et al. 2018a). The objectives of this study are to assess the limits and potential of BRT to generate robust models for Southern Ocean benthic species and to provide some recommendations on the selection of environmental descriptors.

## Methods

### Selection of environmental descriptors

A set of 58 environmental descriptors was compiled from different sources (Supplementary Material #1). This set can be downloaded from the *blueant* R package (<https://github.com/AustralianAntarcticDivision/blueant>), following the procedure given in the “data\_for\_SDM\_vignette” at [https://australianantarcticdivision.github.io/blueant/articles/SO\\_SDM\\_data.html](https://australianantarcticdivision.github.io/blueant/articles/SO_SDM_data.html).

Most descriptors are average abiotic conditions taken from the WOCE database (Supplementary Material #1) and describe the average abiotic conditions for the [2005–2012] time period (*i.e.* temperature, salinity, chlorophyll-a,

particulate organic carbon flux). Some descriptors are available for longer time periods only ([1957–2017] and [1955–2012] for sea ice cover and seafloor oxygen concentration respectively). More recent or precise datasets are not available at the scale of the Southern Ocean. Raster layers were compiled with a  $0.1 \times 0.1^\circ$  pixel resolution (11 km approximately), each  $0.1 \times 0.1^\circ$  pixel being used as a single grid-cell pixel, and cropped to the extent of the Southern Ocean (herein defined as waters south of  $45^\circ\text{S}$  latitude) for a total of 1.26 million pixels. Missing values are not interpolated to avoid potential biases. Available descriptors are selected according to their ecological relevance to benthic studies and following previous recommendations provided for species distribution modelling (Franklin 2010; Anderson 2013) and Antarctic studies (Saucède et al. 2014). The selected descriptors best document the main characteristics of the species physical habitat (depth, sea water temperature, geomorphology, sediment nature, slope, roughness), geography (distance to the Antarctic continent, to canyons, to continental shelves, to the maximal sea ice extent in winter), seasonality (sea ice concentration and thickness), food resources (chlorophyll-a concentration and Particulate Organic Carbon [POC] exported on the sea bottom) and chemical environment (oxygen concentration and seafloor salinity). Minimal, maximal and range values (min–max difference) of some descriptors are computed to complement the dataset (Franklin 2010; Bradie and Leung 2017; Guillaumot et al. 2018a). Extreme weather conditions and climate events were shown to strongly impact natural environments, notably species survival and distribution (Easterling et al. 2000; Wernberg et al. 2013). Here, supplementary descriptors are specially developed for the intensity and frequency of monthly changes in seafloor temperature, salinity, oxygen and chlorophyll-a concentrations. For each pixel and 1 year, these layers document how many times monthly average values are respectively higher (‘maximal extreme event’) or lower (‘minimal extreme event’) than the yearly median value (Supplementary Material #2, codes available at <https://github.com/charleneguillaumot/THESIS>).

## Biological records

Antarctic sea stars play an important role in the structuring of benthic communities (McClintock et al. 2008; Mah and Blake 2012; Lawrence 2013), they have contrasting ecological niches and life history traits (e.g. feeding diets, reproduction and development modes) that condition habitat preferences and dispersal abilities (Moreau et al. 2017, Table 1). Here, SDMs are generated for six sea star species using presence-only records obtained from the “Antarctic and sub-Antarctic asteroid database” published by Moreau et al. (2018): *Acodontaster hodgsoni* (Bell 1908), *Bathybiaster loripes* (Sladen 1889), *Glabraster antarctica* (Smith


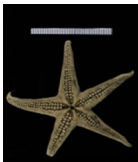


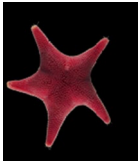

1876), *Labidiaster annulatus* Sladen 1889, *Odontaster validus* Koehler 1906 and *Psilaster charcoti* (Koehler 1906). The studied species are abundant and have been regularly sampled during benthic expeditions to the Southern Ocean, making them some of the best-documented occurrence records on database available for Southern Ocean benthic species (Moreau et al. 2018). The working database (Moreau et al. 2018) includes presence-only records obtained by trawling and scuba diving during numerous expeditions to the Southern Ocean ranging from 1872 to 2016 (Supplementary Material #3). Occurrence data collected during the last 50 years are the most abundant with an intense sampling effort carried out in the framework of the International Polar Year (IPY: 2007–2009) and the Census of Antarctic Marine Life (CAML: 2005–2010). All occurrence data are selected to ensure that a sufficient number of records are available to run the models (Stockwell and Peterson 2002; van Proosdij et al. 2016) and exhaustively cover the geographical space occupied by the considered species. Presence-only records are spatially aggregated near coastal areas and scientific stations (Supplementary Material #4 and see De Broyer et al. 2014; Guillaumot et al. 2019). Presence record duplicates found in the same grid-cell pixel are removed to reduce spatial replication as described by Segurado et al. (2006) and Boria et al. (2014). Because the considered species have different depth ranges (Moreau et al. 2018), model projection is performed for each species independently and bounded by maximal depth value defined by the species deepest record (see Table 1 for details).

## Model calibration

Boosted regression tree (BRT) is chosen as a robust method to test the influence of descriptor selection on model performance. This machine-learning algorithm has been shown to be well suited to accommodate presence-only data and incomplete datasets, to fit complex relationships between species records and environmental descriptors, to limit model overfitting and to have high transferability performances (Elith et al. 2006, 2008; Hastie et al. 2009; Ward et al. 2009; Reiss et al. 2011; Heikkinen et al. 2012; Mainali et al. 2015; Guillaumot et al. 2019), transferability being defined as the ability of models to predict in new environmental conditions (Friedman et al. 2001).

BRT models are calibrated following the procedure detailed in Guillaumot et al. (2019) and using the *gbm* R package (Elith et al. 2008; Ridgeway 2015). BRT parameters are set to minimize both the optimal number of trees used to build the model and the minimal predictive deviance (learning rate, bag fraction and tree complexity are provided for each species in Supplementary Material #5). A set of 1,000 background records are randomly sampled in the environmental space (maximal depth limit depending on the studied

**Table 1** The six studied species and their respective ecological traits

	<i>Acodontaster hodgsoni</i> (Bell 1908)	<i>Bathyiaster loripes</i> (Sladen 1889)	<i>Glabraster antarctica</i> (Smith 1876)	<i>Labidiaster annulatus</i> (Sladen 1889)	<i>Odontaster validus</i> (Koehler 1906)	<i>Psilaster charcoii</i> (Koehler 1906)
						
Feeding diet	Predator (mainly sponges) (Brueggeman 1998)	Detritivorous (Dearborn 1977)	Deposit feeder, predator, or scavenger (Brueggeman 1998)	Predator (Dearborn et al. 1991)	Opportunistic feeder (suspensivorous, deposit feeder, predator, scavenger) (Brueggeman 1998)	Deposit feeder, predator (Brueggeman 1998)
Reproduction and development modes	Broadcaster with non-feeding planktonic larvae (Bosch and Pearse 1990)	Broadcaster with non-feeding planktonic larvae (Bosch and Pearse 1990)	Broadcaster with feeding planktonic larvae (Bosch 1989)	Broadcaster and probably feeding planktonic larvae (Janosik et al. 2008)	Broadcaster with feeding planktonic larvae (Bosch and Pearse 1990)	Broadcaster with non-feeding planktonic larvae (Bosch and Pearse 1990)
Occurrence number	297	585	844	373	309	350
Model maximum depth	1500 m	4000 m	4000 m	1500 m	1500 m	4000 m

Presence-only records duplicates present on a same grid-cell pixel are removed from the count of occurrences. The model maximum depth is defined for each species independently according to the density distribution of recorded depth values. Images sources: Brueggeman (1998), BIOMAR ULB database (P. Pernet), proteker.net, B121 expedition (Q. Jossart)

species, Table 1). This number is tested sufficient enough to represent the whole spectrum of environmental conditions existing in the geographic area of interest (Guillaumot et al. 2019; Supplementary Material) while being as close as possible to the number of records used to generate the model (Barbet-Massin et al. 2012). One hundred background data samples are generated as model replicates. Spatial aggregation of occurrence records is a recurrent bias in Antarctic benthic species databases (Fabri-Ruiz et al. 2018; Guillaumot et al. 2018b, 2019). To reduce the effect of spatial aggregation on model outputs, background records are sampled following a target-group approach (Phillips et al. 2009). In this approach, background data are randomly sampled in the area of interest, following a weighting scheme defined by a Kernel Density Estimation (KDE) of sampling effort in the Southern Ocean (Guillaumot et al. 2018a, supplementary material in Guillaumot et al. 2019).

When using spatially aggregated records, standard cross-validation procedures used to evaluate modelling performances can be strongly biased (Hijmans 2012; Roberts et al. 2017; Guillaumot et al. 2019). The random selection of training and test data leads to the violation of independence between training and test subsets, which can induce an over-estimation of correctly predicted test data by the model (Hijmans 2012). Using cross-validation procedures that spatially segregate training and test data (defined based on presence and background subsets) is a good alternative to accurately evaluate the performance of SDMs based on aggregated datasets. In the present study, a “sixfold CLOCK” cross-validation approach adapted from Guillaumot et al. (2019) was applied. This procedure randomly defines six sectors around Antarctica according to longitude, three for training data and three for test data.

### Collinearity and the selected number of environmental descriptors

Collinearity between the 58 selected descriptors is analysed following a stepwise approach that eliminates layers with a Variance Inflation Factor (VIF) > 10, using the ‘vif.step’ function of the *usdm* R package (Naimi et al. 2014). VIF > 10 is defined as the threshold above which the effect of multicollinearity on model predictions is considered significant (Hair et al. 2014) and too strong to be automatically corrected by machine-learning algorithms (Dormann et al. 2012). Multicollinearity is measured on projection areas, that is the portion of the environment for which SDMs do not extrapolate. Extrapolation areas are defined for each species independently using the Multivariate Environmental Similarity Surface index (MESS, Elith et al. 2010). They correspond to all grid-cell pixels where descriptor values are not contained within the range of environmental conditions on which presence-only data are recorded. Models generated

with the 58 environmental descriptors are compared to models for which collinear descriptors are removed.

A stepwise procedure is used to test the effect of the selected number of environmental descriptors on model performance. SDMs are first generated for the six species using the total set of 58 environmental descriptors. Then, the six descriptors that contribute the least to the average model are iteratively pruned at each step of a series of SDMs successively generated with 58, 52, 46, 40, 34, 28, 22, 16, 10, and four environmental descriptors.

### Model evaluation and comparisons

The percentage of presence data correctly predicted (i.e. correctly classified test data) is computed to assess the performance of SDMs in terms of transferability. Model performances are also assessed using the Area Under the Receiver Operating Curve (AUC, Fielding and Bell 1997), the Point Biserial Correlation between predicted and observed values (COR, Elith et al. 2006) and the True Skill Statistic (TSS, Allouche et al. 2006). Suitable areas are classified using the Maximum Sensitivity plus Specificity threshold (MaxSSS), which is the most adapted index for SDMs using presence-only data (Liu et al. 2013). MaxSSS enables to split model projections into suitable (> MaxSSS value) and unsuitable areas (< MaxSSS value). The average number of regression trees produced by BRT to generate models (*gbm* R package, Elith et al. 2008) is calculated to evaluate model complexity. Scores of SDM series generated with a decreasing number of environmental descriptors are compared between each other using the Mann–Kendall non-parametric trend test to assess the presence of a monotonic trend (Hipel and McLeod 1994). Differences between model performances (AUC, TSS, COR, percentage of correctly classified test data), model properties (number of trees) and outputs (percentage of predicted suitable area) are tested using a Wilcoxon–Mann–Whitney pairwise test.

### Final SDM outputs

Six final SDMs are proposed for the six considered species for the [2005–2012] time period after selection of the optimal number of descriptors and after removing collinear descriptors. The contribution of descriptors and their marginal effects (partial dependence plots) are provided and compared between each other. Environmental conditions predicted as suitable for species distribution are plotted through a principal component analysis (PCA) to display the predicted species occupied environmental space. PCA is compared between species having the same projection depth threshold, either 1500 m or 4000 m depth (Table 1).



## Results

### Contribution of environmental descriptors

All models generated for the six species and with the total set of 58 descriptors perform well with an average AUC score value of 0.853 (min. 0.827; max. 0.883) and an average of 67.2% of correctly predicted test data (59.5–75.1%). ‘Extreme events’ descriptors specifically computed for this study (Supplementary Material #2) never contribute more than 1% to SDMs, some extreme chlorophyll-a layers excepted (Table 2). Overall, parameters that contribute the

most to all SDMs are depth, currents, ice thickness and seafloor properties (Table 2, Fig. 1). Few contrasts are obtained in contributions between species models except for the contribution of seafloor current speed and POC concentrations that, respectively, vary from 1.95 to 10.84% and 0.49 to 7.05% between SDMs (Fig. 1).

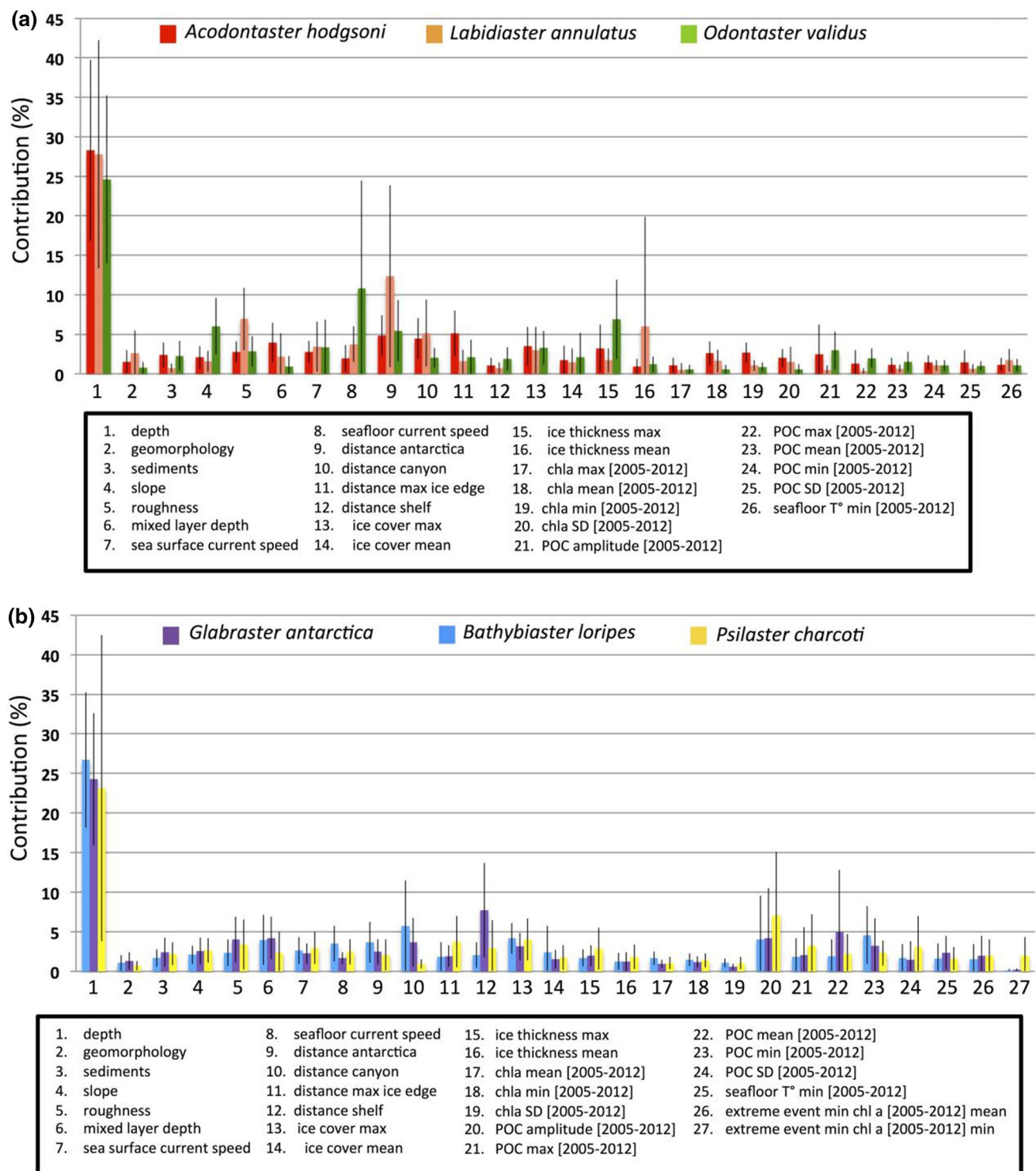
### Number of environmental descriptors

Overall, models generated with different numbers of environmental descriptors do not show significant changes in model performance (Mann–Kendall trend tests, Table 3). Models generated with four environmental descriptors only

**Table 2** Average contribution of each environmental descriptor (based on 100 model replicates) generated for the six studied species using the total set of 58 descriptors

Descriptor	Contribution	Descriptor	Contribution	Descriptor	Contribution
depth	■	ice_thickness_range	■	seafloor_sali_2005_2012_min	■
geomorphology	■A,B,G,L	chla_ampli_alltime_2005_2012	■	seafloor_sali_2005_2012_sd	■
sediments	■A,B,G,O,P	chla_max_alltime_2005_2012	■A	seafloor_temp_2005_2012_ampli	■
slope	■	chla_mean_alltime_2005_2012	■A,B,L,P	seafloor_temp_2005_2012_max	■
roughness	■	chla_min_alltime_2005_2012	■A,B,G,L,P	seafloor_temp_2005_2012_mean	■
mixed_layer_depth	■	chla_sd_alltime_2005_2012	■A,B,L,P	seafloor_temp_2005_2012_min	■B,G,L,P
seasurface_current_speed	■	POC_2005_2012_ampli	■A,B,G,O,P	seafloor_temp_2005_2012_sd	■
seafloor_current_speed	■	POC_2005_2012_max	■A,B,G,O,P	extreme_event_max_chl_2005_2012_ampli	■
distance_antarctica	■	POC_2005_2012_mean	■A,B,G,O,P	extreme_event_max_chl_2005_2012_max	■
distance_canyon	■	POC_2005_2012_min	■	extreme_event_max_chl_2005_2012_mean	■
distance_max_ice_edge	■	POC_2005_2012_sd	■A,B,G,O,P	extreme_event_max_chl_2005_2012_min	■
distance_shelf	■A,B,G,O,P	seafloor_oxy_19552012_ampli	■	extreme_event_min_chl_2005_2012_ampli	■
ice_cover_max	■	seafloor_oxy_19552012_max	■	extreme_event_min_chl_2005_2012_max	■
ice_cover_mean	■	seafloor_oxy_19552012_mean	■	extreme_event_min_chl_2005_2012_mean	■B,G,O,P
ice_cover_min	■	seafloor_oxy_19552012_min	■	extreme_event_min_chl_2005_2012_min	■P
ice_cover_range	■	seafloor_oxy_19552012_sd	■	extreme_event_min_oxy_1955_2012_nb	■
ice_thickness_max	■	seafloor_sali_2005_2012_ampli	■	extreme_event_max_sali_2005_2012_nb	■
ice_thickness_mean	■B,G,L,O,P	seafloor_sali_2005_2012_max	■	extreme_event_min_sali_2005_2012_nb	■
ice_thickness_min	■	seafloor_sali_2005_2012_mean	■	extreme_event_max_temp_2005_2012_nb	■
				extreme_event_min_temp_2005_2012_nb	■

In dark blue, descriptors always contribute more than 1% to all models. In light blue, descriptors contributing more than 1% to some species models only (A: *Acodontaster hodgsoni*, B: *Bathyiaster loripes*, G: *Glabraster antarctica*, L: *Labidiaster annulatus*, O: *Odontaster validus*, P: *Psilaster charcoti*). In red, descriptors never contributing more than 1% to all species models. The description of the different environmental descriptors is provided in Supplementary Material #1



**Fig. 1** Contribution of environmental descriptors to SDMs projected until **a** 1500 m and **b** 4000 m depth for the six species. Environmental descriptors contributing less than 1% to all models are not shown.

Error bars correspond to standard deviation values of the contribution percentages (100 replicates of background sampling and spatial cross-validation splitting)

show a significant decrease in AUC, COR, and TSS values, and in the percentage of correctly classified test data for all species but *G. antarctica* (Fig. 2, Supplementary Material

#6). Significant differences in model performance are model-specific, whatever the number of descriptors used (Fig. 2, Supplementary Material #6). Differences in the number of

**Table 3** Mann–Kendall statistic scores ( $\tau$ )

	<i>Acodontaster hodgsoni</i>	<i>Bathybiaster loripes</i>	<i>Glabraster antarctica</i>	<i>Labidiaster annulatus</i>	<i>Odontaster validus</i>	<i>Psilaster charcoti</i>
AUC	−0.111	0.022	−0.644	−0.067	−0.378	−0.289
COR	−0.111	0.156	−0.556	−0.244	−0.289	−0.289
TSS	−0.244	−0.067	−0.600	−0.067	−0.289	−0.422
Number of trees	0.205	0.675	0.303	−0.322	0.023	−0.210
% Correctly classified test data	−0.067	0.511	0.156	0.511	−0.156	0.511
Average number of suitable pixels	0.167	0.111	0.200	0.333	0.333	0.289

Models are built with 58, 52, 46, 40, 34, 28, 22, 16, 10 and 4 environmental descriptors respectively. 100 replicates are generated in each case. The Mann–Kendall trend test is realized on the median value of the 100 replicates. All tests are not significant. The direction of the monotonic trend is given by the sign of the  $\tau$  values. *AUC* area under the curve, *COR* biserial Pearson correlation, *TSS* true skill statistic. The percentage of correctly classified test data is defined by the proportion of presence test data correctly predicted by the model

trees used to generate models and in the size of suitable areas are never tested significant (Table 3).

### Collinearity

Most SDMs generated with and without collinear descriptors show similar performance statistics (AUC, TSS, COR, and percentage of correctly classified test data) and a comparable number of trees is used to build models (Table 4). However, for *A. hodgsoni* and *G. antarctica*, lower AUC, TSS and COR values are obtained for models generated without collinear descriptors. The percentage of correctly classified test data remains unchanged except in models generated without collinear descriptors for *A. hodgsoni* (−9.9%) and *O. validus* (−19.5%) in which it significantly decreases. For all species but *G. antarctica*, the proportion of predicted suitable area increases in models generated without collinear descriptors (Table 4).

### Comparison between final SDMs

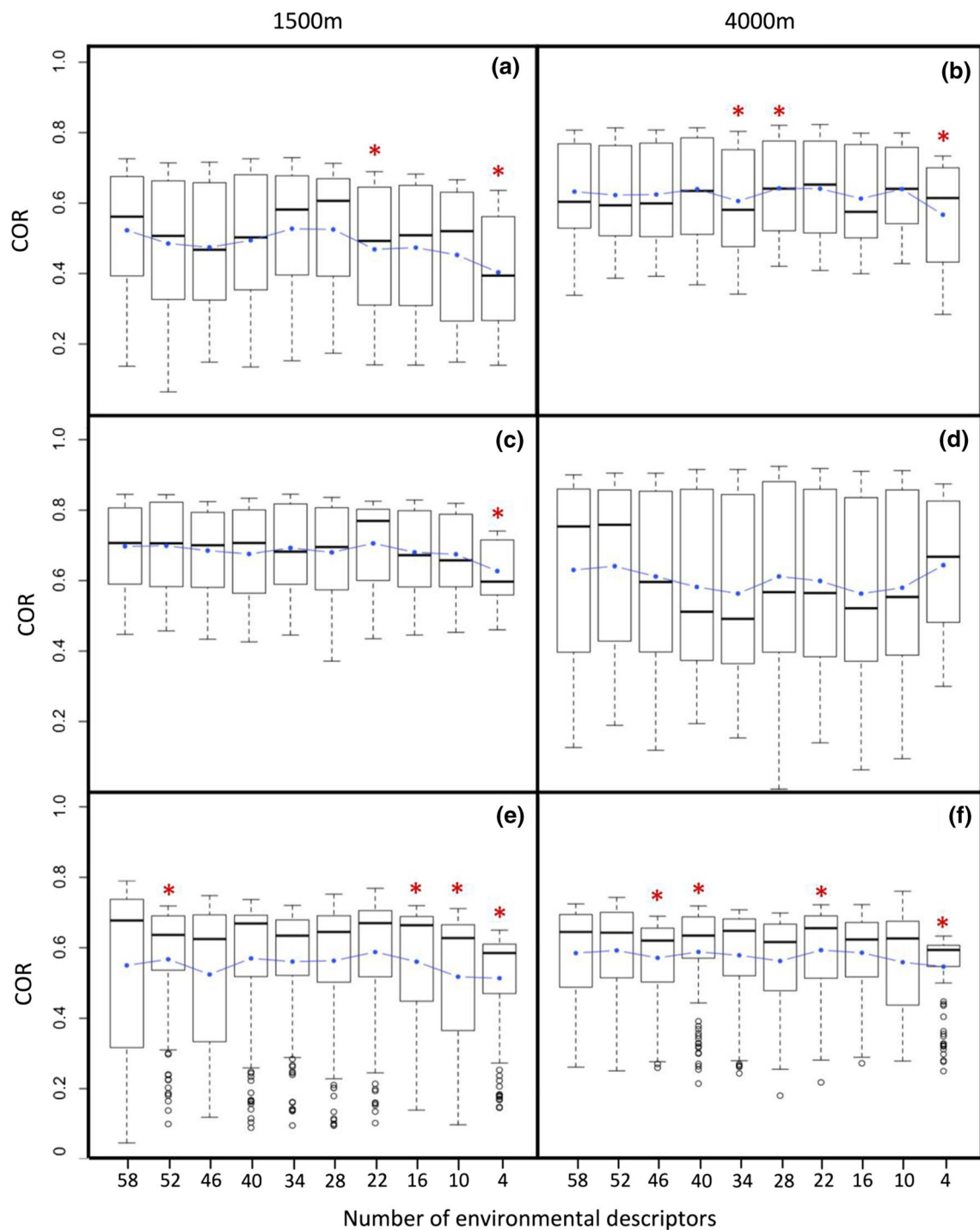
Distance layers (i.e. distance to Antarctic coasts, to shelves, to the nearest canyons, to the maximum ice edge in winter, see Supplementary Material #1) are used as descriptors in a first phase of the analysis to test the effect of collinearity and the number of descriptors on model performance because they are commonly used in SDMs performed for Southern Ocean species (Mormède et al. 2014). However, although relevant when interpolating species distribution patterns (Table 2), interpreting the contribution of such descriptors is not straightforward when it comes to describe species ecological niche. Therefore, these descriptors are excluded from analyses in the final set of SDMs. In addition, descriptors that never contribute more than 1% to SDMs (Table 2) as well as collinear descriptors (depending on species) are removed from the initial set of descriptors. Depending on the species under study, a set of 14 to 16 descriptors is used to calibrate final

models: 13 of these descriptors are common to the six studied species and for three species, additional descriptors on extreme events on chlorophyll-a concentration are used (Supplementary Material #7). The performance of final models is good for all species, with AUC values ranging from  $0.810 \pm 0.09$  to  $0.872 \pm 0.07$  (mean and standard deviation values), TSS values from  $0.461 \pm 0.121$  to  $0.546 \pm 0.08$ , COR values from  $0.503 \pm 0.136$  to  $0.656 \pm 0.121$  and correctly classified test data from  $51.6 \pm 23.7$  to  $80.7 \pm 10.1\%$  (Supplementary Material #8).

The PCA (Fig. 3) shows an important contribution of both the physical environment (slope, roughness) and food resources (chlorophyll-a concentrations) to SDMs projected down to 1500 m depth (strong correlation with PC1) and a weaker and independent contribution of mean sea ice cover and seafloor current speed (strong correlation with PC2, Fig. 3d). In contrast, food resources (chlorophyll-a and POC concentrations), sea ice cover and depth are the main contributors to SDMs projected down to 4000 m depth (high correlation with PC1) with weaker contributions of the physical environment (slope and roughness) (correlation with PC2, Fig. 3h). Major differences are obtained between "shallow" and "deep" models (Figs. 1, 3) whatever the other species ecological traits (Table 1).

Spatial projections of SDMs also show important contrasts in distribution patterns between "shallow" (1500 m) and "deep" (4000 m) models (Fig. 4). Shallow models present low probability values along the Antarctic coasts and higher probabilities in the sub-Antarctic Islands, in the Kerguelen or South Georgia archipelagos, except for *O. validus*. The three SDMs projected down to 4000 m depth show common patterns, with high probabilities predicted close to the Antarctic coasts where most occurrences are recorded (Supplementary Material #4). High probabilities are also predicted on the Kerguelen Plateau for *B. loripes* and *G. antarctica*, while low probabilities





**Fig. 2** Influence of the number of environmental descriptors on SDM performance. Boxplot of 100 model replicate scores. Changes in biserial correlation (COR) values for **a** *Acodontaster hodgsoni*, **b** *Bathybiaster loripes*, **c** *Labidiaster annulatus*, **d** *Glabaster antarctica*, **e** *Odontaster validus* and **f** *Psilaster charcoti*. Average values are indi-

cated in blue. Red asterisks indicate significant changes in median values between the series and preceding value (Wilcoxon rank paired test,  $p$ -value < 0.05). The left-side and right-side columns correspond to species for which models are respectively projected until 1500 m and 4000 m depth

**Table 4** Mann–Whitney Wilcoxon pairwise test (*W*) comparing statistics of models generated without collinear descriptors and models run with the total set of 58 environmental descriptors. Associated

p-values are summarized by asterisks (no star  $p > 0.05$ , \* $p < 0.05$ , \*\* $p < 0.01$  and \*\*\* $p < 0.001$ )

	1500 m			4000 m		
	<i>Acodontaster hodgsoni</i>	<i>Labidiaster annulatus</i>	<i>Odontaster validus</i>	<i>Bathyiaster loripes</i>	<i>Glabraster antarctica</i>	<i>Psilaster charcoti</i>
AUC	6041*	4754	5738	5578	5931*	5280
COR	5842*	4783	5867*	5596	5964*	5247
TSS	6138**	4792	5748	5596	5840*	5425
% Correctly classified test data	6234**	5546	6247.5**	5590	5145	4512.5
Number of trees	5359	4352	4811	5031.5	4641.5	4312
% suitable area	3526***	6272**	8695***	9759***	4796	8571***

AUC area under the curve, COR biserial Pearson correlation, TSS true skill statistic. The percentage of correctly classified test data is defined by the proportion of presence test data correctly predicted by the model

are predicted for *P. charcoti* in the sub-Antarctic Islands (Fig. 4).

## Discussion

### Influence of the number of descriptors on modelling performance

SDMs performed at the scale of the Southern Ocean are usually based on a limited mass of occurrence data, patchy datasets and using low-resolution environmental descriptors. Recent studies have questioned the relevance of using such SDMs considering the spatial and temporal heterogeneities of datasets and the importance of sampling biases (Fabri-Ruiz et al. 2018; Guillaumot et al. 2018a). In the present work, we focus on the selection of environmental descriptors as a critical step for model calibration (Bucklin et al. 2015; Petitpierre et al. 2017). Machine-learning algorithms such as BRT were proved efficient to deal with non-informative descriptors (De'ath and Fabricius 2000; Elith et al. 2008) and to correct for the influence of collinearity between descriptors (Dormann et al. 2012). The performance of BRT to model the distribution of Antarctic benthic species at large spatial scale is herein evaluated.

Successive models were generated from four to 58 environmental descriptors. All models have similar accuracy (AUC, TSS, COR) and transferability (percentage of correctly classified test data) performances. Models generated with four environmental descriptors only (depending on each species) show significant differences in performance values and low capacities to describe and predict species distribution. SDMs generated for the species *G. antarctica* depart from this general result with no significant differences in modelling performances between models generated with four to 58 descriptors. This may be due to the large number

of occurrence data available to describe the species distribution and conversely, the limited number of environmental descriptors contributing to the models (Supplementary Material #4, Fig. S4d, Table 1).

Many studies have stressed the risk of model overfitting when using too many descriptors (Anderson and Gonzalez 2011; Synes and Osborne 2011; Braunisch et al. 2013; Kramer-Schadt et al. 2013; Petitpierre et al. 2017) or the risk of underestimating the extent of suitable areas due to reduced transferability performances (Beaumont et al. 2005). In contrast, our results show that models generated with a different number of predictors are characterized by similar performance levels. This is congruent with results obtained by Bucklin et al. (2015) who highlighted that the random addition of descriptors has a minor influence on modelling performances when using machine-learning algorithms. The absence of significant changes in the number of trees used to build BRT models, using a different number of environmental descriptors show that BRT is not sensitive to model overfitting, and only selects the relevant information needed for model calibration, a property formulated as the stagewise selection by Elith et al. (2008). Non-informative environmental data that might complexify SDMs are automatically pruned when generating BRT trees, and the most relevant descriptors only are retained to model species distribution (De'ath and Fabricius 2000; Whittingham et al. 2006; Elith et al. 2008). However, selecting a reduced number of environmental descriptors allows the production of simpler models for which descriptor contributions can be easily interpreted (Bucklin et al. 2015).

### Influence of collinearity on modelling performance

Removing collinear descriptors from datasets has remained an usual approach in species distribution modelling (Dormann et al. 2012; Merow et al. 2013; Fabri-Ruiz et al. 2018;

Guillaumot et al. 2018b). However, this strategy has recently been questioned when SDMs are not used for extrapolation (Braunisch et al. 2013; Bucklin et al. 2015; Li et al. 2016; Petitpierre et al. 2017). In the present study, results show that modelling performances (AUC, TSS, COR and percentage of correctly classified test data) of some SDMs significantly decrease when collinear descriptors are removed (i.e. *A. hodgsoni*, *O. validus* and *G. antarctica*). Removing collinear variables that significantly contribute to SDMs may induce model instability and reduce modelling performance. The observed decrease in AUC scores may be due to the reduction of model overfitting when removing collinear descriptors (Dhingra et al. 2016).

Machine-learning algorithms are efficient modelling tools that take into account the multiple interactions among descriptors (Segurado and Araújo 2004; Araújo and Guisan 2006; Dormann et al. 2008; Elith et al. 2008; Braunisch et al. 2013) and can correct for collinearity between environmental descriptors if not too strong (Dormann et al. 2012). In the present work, this is shown by the fact that performance of SDMs produced for the species *B. loripes*, *L. annulatus* and *P. charcoti* remains unaffected when collinear descriptors are removed from the analysis.

However, using collinear descriptors in SDMs can make model outputs difficult to interpret when temporal and spatial relationships between descriptors are unknown (Dormann et al. 2012), because collinearity induces complex relationships between environmental drivers and the underlying processes (Guisan and Thuiller 2005; Elith and Leathwick 2009; Merow et al. 2013). Several methods have been documented to correct for strong collinear effects. The sequential regression approach is one of them and aims at replacing correlated variables by a linear or non-linear model (Leathwick et al. 2006; Dormann et al. 2008). A second method consists in using descriptor score values on PCA principal components rather than descriptor raw values themselves (Kühn 2007; Dormann et al. 2008). However, in this latter approach, SDMs and species ecological preferences are difficult to interpret.

### Selection of environmental descriptors

'Distance layers' (Supplementary Material #1) have been commonly used as descriptors in previous SDMs performed for Southern Ocean studies (Cheung et al. 2008; Murase et al. 2013; Mormède et al. 2014; Nachtsheim et al. 2017). In the present work, 'distance layers' were used in the first set of SDMs and they all showed strong contributions to model outputs. 'Distance layers' may be strongly correlated to environmental gradients, and especially to latitudinal gradients, or may integrate the multiple effects of diverse environmental variations (Bradie and Leung 2017; Ferrari et al. 2018). Interpreting the contribution of such descriptors to SDMs

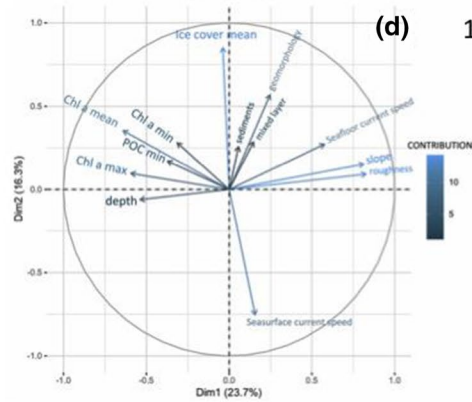
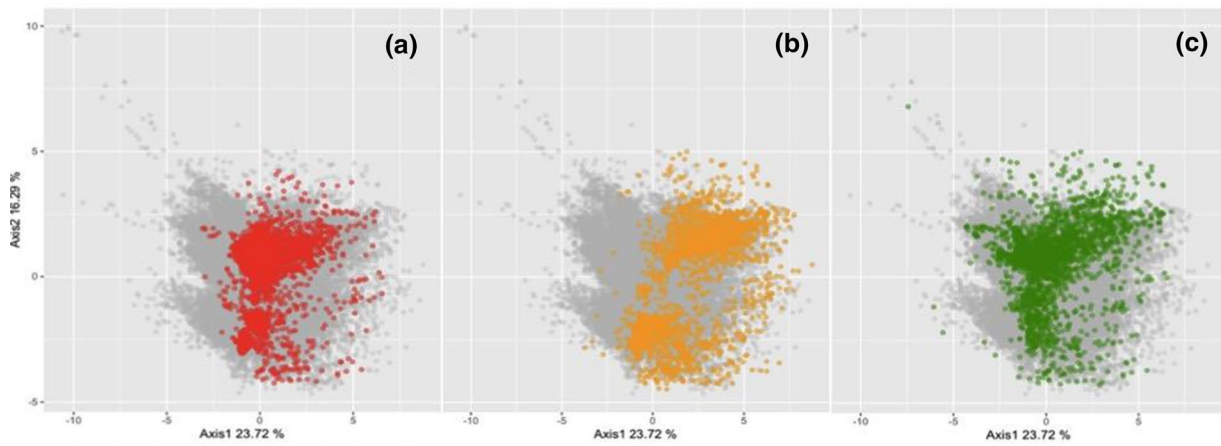
can remain problematic and depends on research objectives, especially depending on whether ecological significance or statistical contributions only are sought. The statistical contribution of a descriptor to the model is the independent contribution of the descriptor deduced from what other descriptors already bring (Dormann et al. 2012), it may not necessarily imply a direct ecological significance. Consequently, 'distance layers' were removed from the initial set of environmental descriptors along with collinear descriptors and descriptors that contributed the least to models (28 descriptors out of the 58 available, Table 2). This reduces the set to 14 or 16 descriptors only depending on the species under study (Supplementary Material #7).

### Final model outputs

In the present study, SDMs performed for *A. hodgsoni*, *L. annulatus* and *O. validus* showed lower performances (lower AUC, TSS, COR and correctly classified test data) compared to SDMs performed for *B. loripes*, *G. antarctica* and *P. charcoti*. For these last three species, a higher number of records were available and contributed to the high model performances as species niches were better described during model calibration (Qiao et al. 2015; van Proosdij et al. 2016; Guillaumot et al. 2018a). Despite these differences in model performance, descriptor contributions and species predicted distributions are mostly similar between models (Figs. 1, 3, 4). This is an unexpected result as the six studied species were initially selected for their contrasting ecological niches and life traits, which should have determined distinct occupied environments and biogeographic patterns. This unexpected result stresses the limits of SDMs performed at broad spatial scale. The low resolution (in space and time) of environmental descriptors, the heterogeneous sampling and the relative low number of occurrence records available are cumulative limitations to model accuracy and species ecological requirements were not precisely captured by models. In contrast, models are all structured by large-scale and common environmental drivers relating to broad-scale latitudinal gradients that prevail between Antarctic and sub-Antarctic regions (Clarke and Johnston 2003; Linse et al. 2006; De Broyer et al. 2014; Moreau et al. 2017).

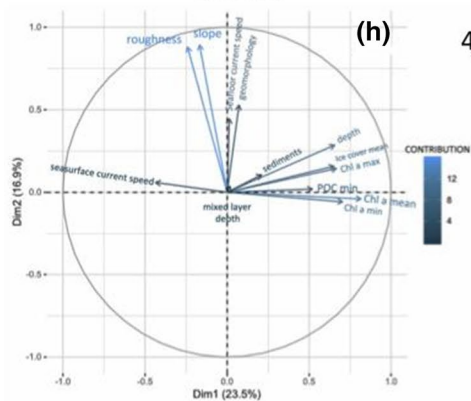
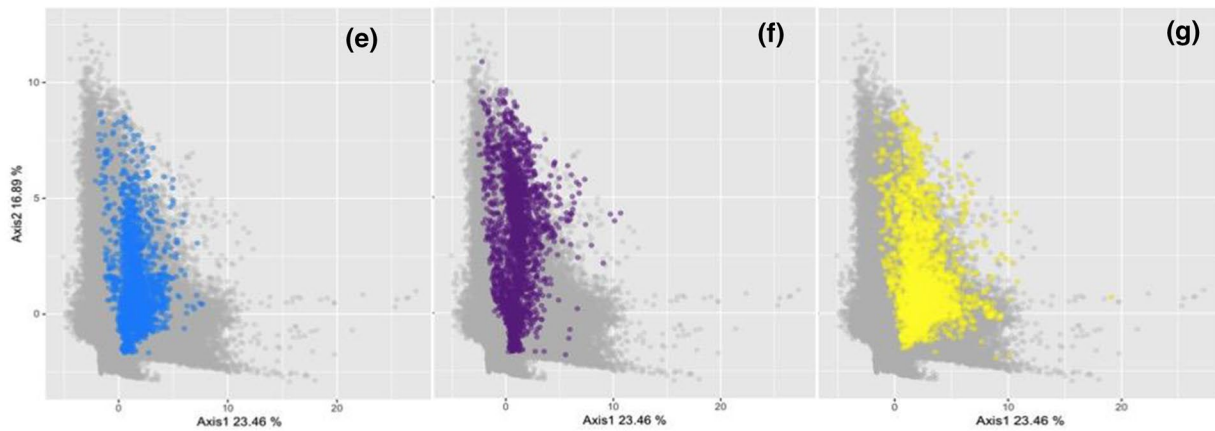
### Conclusions and recommendations

This work aimed at testing the influence of the number of selected environmental descriptors and their collinearity on model performance. Models were generated at the scale of the entire Southern Ocean using BRT. The BRT algorithm is a machine-learning approach that automatically selects descriptors that best characterize species niches (Elith et al. 2008). This matches our results that highlight that all models



1500m models

Axis 1	Correlation	Axis 2	Correlation
roughness	0.83	Ice cover mean	0.85
slope	0.81	geomorphology	0.56
seafloor current speed	0.58	sediments	0.29
Chl a min	-0.31	mixed layer	0.28
POC min	-0.37	seasurface current speed	-0.75
depth	-0.54		
Chl a max	-0.59		
Chl a mean	-0.64		



4000m models

Axis 1	Correlation	Axis 2	Correlation
Chla mean	0.82	slope	0.89
Chla max	0.70	roughness	0.88
Ice cover mean	0.67	seafloor current speed	0.52
Depth	0.65	geomorphology	0.45
POC min	0.65		
ChL a min	0.52		
sediments	0.21		
Seasurface current speed	-0.44		

**Fig. 3** PCA of environmental values (grey dots) from descriptors used in final species distribution models, and that are common between the six species (Supplementary Material #7: depth, geomorphology, sediments, slope, roughness, mixed layer depth, seafloor and seafloor current speed, ice cover mean, chlorophyll-a min, max and mean concentrations for 2005–2012, POC minimum concentrations for 2005–2012), limited to 1500 m (a–c) and 4,000 m depth (e–g) respectively. Colour dots: species suitable area (probabilities > average maxSSS scores) for **a** *Acodontaster hodgsoni*, **b** *Bathylaster loripes*, **c** *Labidiaster annulatus*, **e** *Glabraster antarctica*, **f** *Odontaster validus*, **g** *Psilaster charcoti*. PCA plot of environmental descriptors (**d**, **h**) and appended tables with the associated correlations to PC1 and PC2. All correlation values are significant ( $p < 0.05$ )

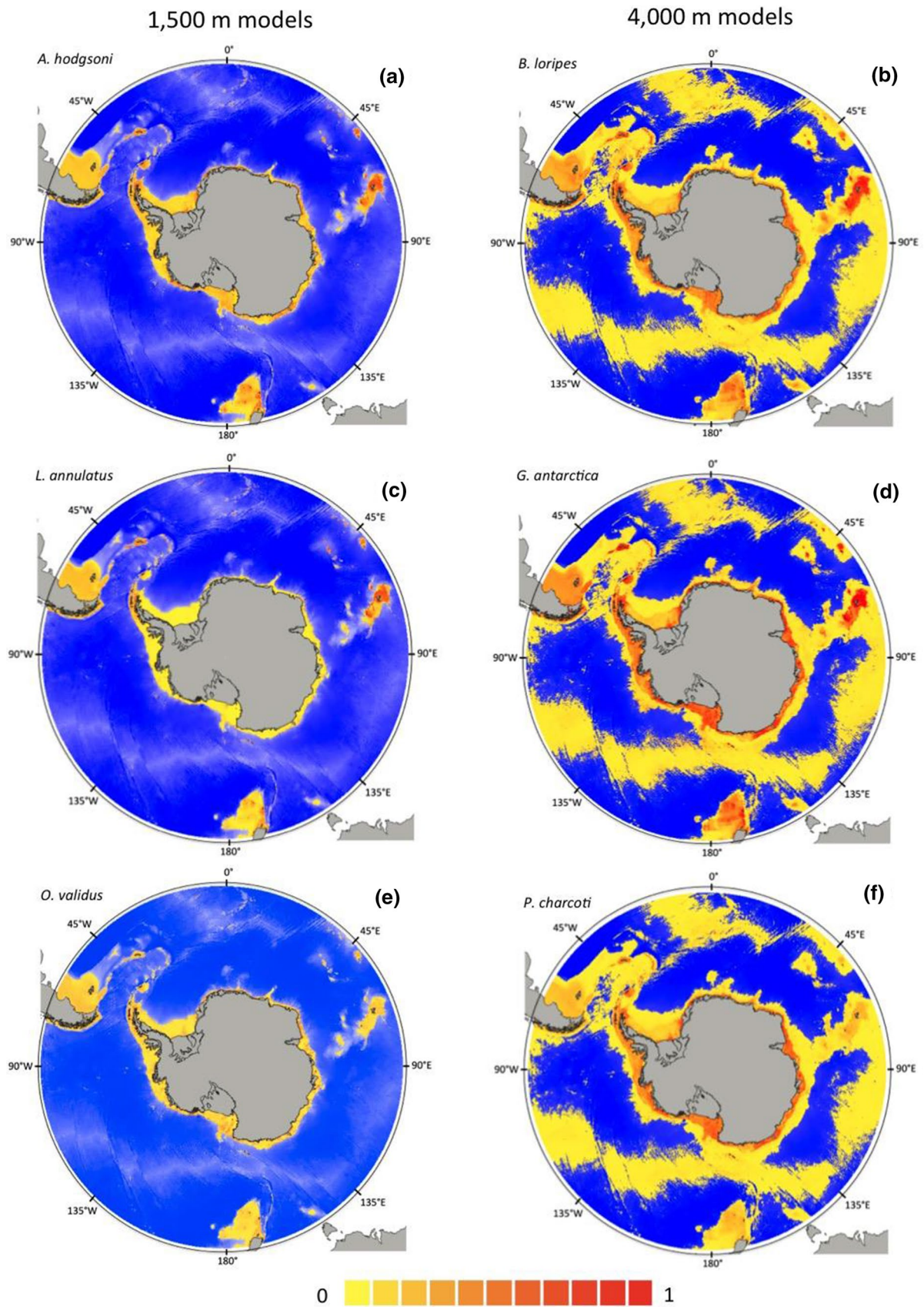
generated with different number of environmental descriptors showed similar performances. In contrast, in most SDMs generated without collinear descriptors, model overfitting tends to be minimized in comparison with models generated with the whole set of 58 descriptors. In three species only, no difference in model performance was observed between models using either collinear or non-collinear descriptors.

Final models were generated using a subset of 14 to 16 environmental descriptors that best explain species distributions. The selected descriptors are not collinear to limit interpretation errors, reduce model complexity, and favour the ecological relevance of models (Austin and van Niel 2011; Braunisch et al. 2013; Bucklin et al. 2015; Petitpierre et al. 2017). However, final SDMs are not very contrasted between species despite significant differences in species ecological niches (McClintock et al. 2008; Mah and Blake 2012; Lawrence 2013; Brandt et al. 2014; Danis et al. 2014; Moles et al. 2015). The performed SDMs are more sensitive

to the number of occurrence records available and to the extent of the projection area. This final result questions the ecological relevance of using modelling approaches at broad spatial scale when based a limited number of occurrence data, spatially aggregated and using descriptors with coarse spatial and temporal resolutions.

These results match those obtained in previous studies and suggest that the validation of model predictions should use independent data, appropriate statistics and expert-based interpretations (Guisan et al. 2013; Fois et al. 2018; Fourcade et al. 2018; Leroy et al. 2018). Combining model outputs performed at narrow spatial scale and complementary data on biotic interactions (Van der Putten et al. 2010; Wisz et al. 2013; Leach et al. 2016), habitat features (Ferrari et al. 2018) and physiological traits (Kearney and Porter 2009; Fordham et al. 2013; Wittmann et al. 2016; Feng and Papeş 2017; Mathewson et al. 2017; Pertierra et al. 2017) constitutes a good alternative. This can enhance the relevance of explanatory models and their use for ecological studies and conservation purposes. Downscaling SDM studies also has the advantage of improving model accuracy relating to particular, local to regional phenotypic or physiological traits of populations, which may differ at broader scale (Thatje 2012). Waiting for more data and ensuring the taxonomic quality of datasets, we recommend the use of SDMs for narrow scale studies using scrutinized and comprehensive occurrence datasets, as much as possible, while selecting non-collinear and ecologically relevant descriptors to minimize model overfitting (El Gabbas and Dormann 2018; Fois et al. 2018).





**Fig. 4** SDMs generated based on the final selection of environmental descriptors for the six studied species (Supplementary Material #7). Projection areas are limited to 1500 m depth (left-hand column) or 4000 m depth (right-hand column) for **a** *Acodontaster hodgsoni*, **b** *Bathyiaster loripes*, **c** *Labidiaster annulatus*, **d** *Glabraster antarctica*, **e** *Odontaster validus*, **f** *Psilaster charcoti*. Blue colours correspond to depth gradient. The colour chart indicates species presence probability comprised between 0 and 1. Polar stereographic projection

**Acknowledgements** This work was supported by a “Fonds pour la formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA) and “Bourse Fondation de la mer” grants to C. Guillaumot. This is contribution no. 31 to the vERSO project and no. 10 to the RECTO project ([www.rectoversoprojects.be](http://www.rectoversoprojects.be)), funded by the Belgian Science Policy Office (BELSPO, contracts no. BR/132/A1/vERSO and no. BR/154/A1/RECTO). This is contribution to the IPEV programs n°1124 REVOLTA and n°1044 PROTEKER. We are grateful to the crew and participants of all the cruises and research programs involved in the capture of the samples included in this dataset (see Moreau et al. 2018): POKER 2, REVOLTA 1 & 2, CEAMARC, JR144, JR179, JR230, JR262, JR275, JR287, JR15005.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43:1223–1232
- Anderson RP, Gonzalez JI (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecol Model* 222:2796–2811
- Anderson RP (2013) A framework for using niche models to estimate impacts of climate change on species distributions. *Ann N Y Acad Sci* 1297:8–28
- Anderson OF, Guinotte JM, Rowden AA, Tracey DM, Mackay KA, Clark MR (2016) Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. *Deep Sea Res Part I* 115:265–292
- Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *J Biogeogr* 33:1677–1688
- Arthur B, Hindell M, Bester M, De Bruyn PN, Goebel ME, Trathan P, Lea MA (2018) Managing for change: using vertebrate at sea habitat use to direct management efforts. *Ecol Ind* 91:338–349
- Austin MP, Van Niel KP (2011) Improving species distribution models for climate change studies: variable selection and scale. *J Biogeogr* 38:1–8
- Ballard G, Jongsomjit D, Veloz SD, Ainley DG (2012) Coexistence of mesopredators in an intact polar ocean ecosystem: the basis for defining a Ross Sea marine protected area. *Biol Conserv* 156:72–82
- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Method Ecol Evol* 3:327–338
- Basher Z, Costello MJ (2016) The past, present and future distribution of a deep-sea shrimp in the Southern Ocean. *PeerJ* 4:e1713
- Beaumont LJ, Hughes L, Poulsen M (2005) Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species’ current and future distributions. *Ecol Model* 186:251–270
- Boria RA, Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol Model* 275:73–77
- Bosch I (1989) Contrasting modes of reproduction in two Antarctic asteroids of the genus *Porania*, with a description of unusual feeding and non-feeding larval types. *Biol Bull* 177:77–82
- Bosch I, Pearce JS (1990) Developmental types of shallow-water asteroids of McMurdo Sound, Antarctica. *Mar Biol* 104:41–46
- Bradie J, Leung B (2017) A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *J Biogeogr* 44:1344–1361
- Brandt A, Van de Putte A, Griffiths HJ (2014) Southern Ocean benthic deep-sea biodiversity and biogeography. In: De Broyer C, Koubbi P, Griffiths HJ, Raymond B, d’Udekem d’Acoz C, Van de Putte A, Danis B, David B, Grant S, Gutt J, Held C, Hosie G, Huettmann F, Post A, Ropert-Coudert Y (eds) *Biogeographic atlas of the Southern Ocean*. SCAR, Cambridge, p 510
- Braunisch V, Coppes J, Arlettaz R, Suchant R, Schmid H, Bollmann K (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography* 36:971–983
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth International Group, Belmont, CA
- Brueggeman P (1998) Underwater field guide to Ross Island & McMurdo Sound, Antarctica. The National Science Foundation’s Office of Polar Programs sponsored Norbert Wu. Univ. California, San Diego.
- Bucklin DN, Basille M, Benscoter AM, Brandt LA, Mazzotti FJ, Romanach SS, Speroterra C, Watling JI (2015) Comparing species distribution models constructed with different subsets of environmental predictors. *Divers Distrib* 21:23–35
- CCAMLR (2009) Commission for the conservation of Antarctic marine living resources. Conservation measure 91–03. Protection of the South Orkney Islands southern shelf. 2pp. <https://www.ccamlr.org/sites/drupal.ccamlr.org/files/91-03.pdf>. Accessed 1 Aug 2018
- CCAMLR (2016) Commission for the conservation of Antarctic marine living resources. Conservation measure 91–05. Ross Sea region marine protected area. pp. 17. [https://www.ccamlr.org/sites/drupal.ccamlr.org/files/91-05\\_2.pdf](https://www.ccamlr.org/sites/drupal.ccamlr.org/files/91-05_2.pdf). Accessed 1 Aug 2018
- Cheung WW, Lam VW, Pauly D (2008) Modelling present and climate-shifted distribution of marine fishes and invertebrates. *Fish Center Res Rep* 16:1–76
- Clarke A, Johnston NM (2003) Antarctic marine benthic diversity. *Oceanogr Mar Biol Annu Rev* 41:55–57
- Constable AJ, Melbourne-Thomas J, Corney SP, Arrigo KR, Barbraud C, Barnes DK, Bindoff NL, Boyd PW, Brandt A, Costa DP, Davidson AT, Ducklow HW, Emmerson L, Fukuchi M, Gutt J, Hindell MA, Hofmann EE, Hosie GW, Iida T, Jacob S, Johnston NM, Kawaguchi S, Kokubun N, Koubbi P, Lea MA, Makhado A, Massom RA, Meiners K, Meredith MP, Murphy EJ, Nicol S, Reid K, Richerson K, Riddle MJ, Rintoul SR, Smith WO Jr, Southwell C, Stark JS, Sumner M, Swadling KM, Takahashi KT, Trathan PN, Welsford DC, Weimerskirch H, Westwood K, Wienecke BC, Wolf-Gladrow D, Wright SW, Xavier JW, Ziegler P (2014) Climate change and Southern Ocean ecosystems I: how changes in physical habitats directly affect marine biota. *Glob Change Biol* 20:3004–3025

- Convey P, Bindschadler R, Di Prisco G, Fahrbach E, Gutt J, Hodgson DA, Mayewski PA, Summerhayes CP, Turner J, ACCE Consortium (2009) Antarctic climate change and the environment. *Ant Sci* 21:541–563
- Convey P, Peck LS (2019) Antarctic environmental change and biological responses. *Sci Adv* 5:eaz0888
- Costa GC, Nogueira C, Machado RB, Colli GR (2010) Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodivers Conserv* 19:883–899
- Danis B, Griffiths HJ, Jangoux M (2014) Asteroidea. In: De Broyer C, Koubbi P, Griffiths HJ, Raymond B, d'Udekem d'Acoz C, Van de Putte A, Danis B, David B, Grant S, Gutt J, Held C, Hosie G, Huettmann F, Post A, Ropert-Coudert Y (eds) *Biogeographic atlas of the Southern Ocean*. SCAR, Cambridge, p 510
- Davis LB, Hofmann EE, Klinck JM, Piñones A, Dinniman MS (2017) Distributions of krill and Antarctic silverfish and correlations with environmental variables in the western Ross Sea, Antarctica. *Mar Ecol Prog Ser* 584:45–65
- Dearborn JH (1977) Foods and feeding characteristics of Antarctic asteroids and ophiuroids. *Adaptations within Antarctic ecosystems*, 293–326.
- Dearborn JH, Edwards KC, Fratt DB (1991) Diet, feeding behavior, and surface morphology of the multi-armed Antarctic sea star *Labidiaster annulatus* (Echinodermata: Asteroidea). *Mar Ecol Prog Ser* Oldendorf 77:65–84
- De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178–3192
- De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C, Van de Putte AP, Danis B, David B, Grant S, Gutt J, Held C, Hosie G, Huettmann F, Post A, Ropert-Coudert Y (eds) (2014) p. 510, Cambridge, SCAR. ISBN 978-0-948277-28-3
- Dhingra MS, Artois J, Robinson TP, Linard C, Chaiban C, Xenarios ER, Lietchi K, Xiao X, Von Dobschuetz S, Claes F, Newman SH, Dauphin G, Gilbert M (2016) Global mapping of highly pathogenic avian influenza H5N1 and H5Nx clade 23.44 viruses with spatial cross-validation. *elife* 5:e19571
- Dormann CF, Purschke O, Márquez JRG, Lautenbach S, Schröder B (2008) Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecology* 89:3371–3386
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, García Márquez JR, Gruber B, Lafourcade B, Leitao P, Münkemüller T, McClean C, Osborne PE, Reineking B, Schroder B, Skidmore AK, Zurell D, Lautenbach S (2012) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46
- Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO (2000) Climate extremes: observations, modeling, and impacts. *Science* 289:2068–2074
- El-Gabbas A, Dormann CF (2018) Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. *Ecol Evol* 8:2196–2206
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *J Anim Ecol* 77:802–813
- Elith J, Graham CH (2009) Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32:66–77
- Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Ann Rev Ecol Evol Syst* 40:677–697
- Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods Ecol Evol* 1:330–342
- Fabri-Ruiz S, Saucède T, Danis B, David B (2017) Southern Ocean Echinoids database—an updated version of Antarctic, Sub-Antarctic and cold temperate echinoid database. *ZooKeys* 697:1
- Fabri-Ruiz S, Danis B, David B, Saucède T (2018) Can we generate robust species distribution models at the scale of the Southern Ocean? *Divers Distrib* 25:21–37
- Feng X, Papeş M (2017) Can incomplete knowledge of species' physiology facilitate ecological niche modelling? a case study with virtual species. *Divers Distrib* 23:1157–1168
- Ferrari R, Malcolm H, Neilson J, Lucieer V, Jordan A, Ingleton T, Figueira W, Johnston N, Hill N (2018) Integrating distribution models and habitat classification maps into marine protected area planning. *Estuar Coast Shelf Sci* 212:40–50
- Ficetola GF, Thuiller W, Maud C (2007) Prediction and validation of the potential global distribution of a problematic alien invasive species: the American bullfrog. *Divers Distrib* 13:476–485
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38–49
- Fois M, Cuena-Lombraña A, Fenu G, Bacchetta G (2018) Using species distribution models at local scale to guide the search of poorly known species: review, methodological issues and future directions. *Ecol Model* 385:124–132
- Fordham DA, Mellin C, Russell BD, Akçakaya RH, Bradshaw CJ, Aiello-Lammens ME, Caley JM, Connell SD, Mayfield S, Shepherd SA, Brook BW (2013) Population dynamics can be more important than physiological limits for determining range shifts under climate change. *Glob Change Biol* 19:3224–3237
- Fourcade Y, Besnard AG, Secondi J (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob Ecol Biogeogr* 27:245–256
- Franklin J (2010) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge
- Freer JJ, Tarling GA, Collins MA, Partridge JC, Genner MJ (2019) Predicting future distributions of lanternfish, a significant ecological resource within the Southern Ocean. *Divers Distrib* 25:1259–1272
- Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning, vol. 1, Springer Series, New York in Statistics
- Guillaumot C, Martin A, Fabri-Ruiz S, Eléaume M, Saucède T (2016) Echinoids of the Kerguelen Plateau—occurrence data and environmental setting for past, present, and future species distribution modelling. *ZooKeys* 630:1
- Guillaumot C, Martin A, Eléaume M, Saucède T (2018a) Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau. *Mar Ecol Prog Ser* 594:149–164
- Guillaumot C, Fabri-Ruiz S, Martin A, Eléaume M, Danis B, Féral JP, Saucède T (2018b) Benthic species of the Kerguelen Plateau show contrasting distribution shifts in response to environmental changes. *Ecol Evol* 8:6210–6224
- Guillaumot C, Artois J, Saucède T, Demoustier L, Moreau C, Eléaume M, Agüera A, Danis B (2019) Species distribution models in a data-poor and broad scale context. *Prog Oceanogr* 175:198–207
- Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA (2015) Is my



- species distribution model fit for purpose? Matching data and models to applications. *Glob Ecol Biogeogr* 24:276–292
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8:993–1009
- Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AI, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney M, Possingham HP, Buckley YM (2013) Predicting species distributions for conservation decisions. *Ecol Lett* 16:1424–1435
- Gutt J, Zurell D, Bracegridle T, Cheung W, Clark M, Convey P, Danis B, David B, De Broyer C, Prisco G, Griffiths H, Laffont R, Peck LS, Pierrat B, Riddle MJ, Saucède T, Turner J, Verde C, Wang Z, Grimm V (2012) Correlative and dynamic species distribution modelling for ecological predictions in the Antarctic: a cross-disciplinary concept. *Polar Res* 31:11091
- Hair J, Black W, Babin B, Anderson R (2014) *Multivariate Data Analysis*, 7th edn. Pearson Education Limited, Pearson, p 729
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning-data mining inference and prediction*. Springer, New York
- Heikkinen RK, Marmion M, Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* 35:276–288
- Henley SF, Schofield OM, Hendry KR, Schloss IR, Steinberg DK, Moffat C, Peck LS, Costa DP, Bakker DCE, Hugues C, Rozema P, Ducklow HW, Abele D, Stefels J, Van Leeuwe MA, Brussaard CPD, Buma AGJ, Kohut J, Meredith MP (2019) Variability and change in the west Antarctic Peninsula marine system: research priorities and opportunities. *Progr Oceanogr* 173:208–237
- Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93:679–688
- Hipel K, McLeod A (1994) *Time series modelling of water resources and environmental systems*. Elsevier, Amsterdam, p 1012
- Hogg OT, Huvenne VA, Griffiths HJ, Linse K (2018) On the ecological relevance of landscape mapping and its application in the spatial planning of very large marine protected areas. *Sci Total Environ* 626:384–398
- Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117:847–858
- Ingels J, Vanreusel A, Brandt A, Catarino AI, David B, De Ridder C, Dubois P, Gooday AJ, Martin P, Pasotti F, Robert H (2012) Possible effects of global environmental changes on Antarctic benthos: a synthesis across five major taxa. *Ecol Evol* 2:453–485
- Janosik AM, Mahon AH, Scheltema RS, Halanych KM (2008) Short note: life history of the Antarctic sea star *Labidiaster annulatus* (Asteroidea: Labidiasteridae) revealed by DNA barcoding. *Antract Sci* 20:563–564
- Jansen J, Hill NA, Dunstan PK, Eléaume MP, Johnson CR (2018) Taxonomic resolution, functional traits, and the influence of species groupings on mapping Antarctic seafloor biodiversity. *Front Ecol Evol* 6:81
- Jerosch K, Scharf FK, Derigibus D, Campana GL, Zacher K, Pehlke H, Falk U, Hass HC, Quartino ML, Abele D (2019) Ensemble modelling of Antarctic macroalgal habitats exposed to glacial melt in a polar fjord. *Front Ecol Evol* 7:207
- Jiménez-Valverde A, Peterson AT, Soberón J, Overton JM, Aragón P, Lobo JM (2011) Use of niche models in invasive species risk assessments. *Biol Invasions* 13:2785–2797
- Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol Lett* 12:334–350
- Kramer-Schadt S, Niedballa J, Pilgrim JD, Schröder B, Lindenborn J, Reinfelder V, Stillfried M, Heckmann I, Scharf AK, Augeri DM, Cheyne SM, Hearn AJ, Ross J, Macdonald DW, Mathai J, Eaton J, Marshall AJ, Semiadi G, Rustam R, Bernard H, Alfred R, Samejima H, Duckworth JW, Breitenmoser-Wuersten C, Belant JL, Hofer H, Wilting A (2013) The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers Distrib* 19:1366–1379
- Kühn I (2007) Incorporating spatial autocorrelation may invert observed patterns. *Divers Distrib* 13:66–69
- Lawrence JM (2013) *Starfish: biology and ecology of the Asteroidea*. Johns Hopkins University Press, Baltimore, p 288
- Leach K, Montgomery WI, Reid N (2016) Modelling the influence of biotic factors on species distribution patterns. *Ecol Model* 337:96–106
- Leathwick JR, Elith J, Hastie T (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol Model* 199:188–196
- Leroy B, Delsol R, Huguéy B, Meynard CN, Barhoumi C, Barbet-Massin M, Bellard C (2018) Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *J Biogeogr* 45:1994–2002
- Li J, Tran M, Siwabessy J (2016) Selecting optimal random forest predictive models: a case study on predicting the spatial distribution of seabed hardness. *PLoS ONE* 11:e0149089
- Linse K, Griffiths HJ, Barnes DK, Clarke A (2006) Biodiversity and biogeography of Antarctic and sub-Antarctic mollusca. *Deep Sea Res Part II: Top Stud Ocean* 53:985–1008
- Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. *J Biogeogr* 40:778–789
- Loiselle BA, Jørgensen PM, Consiglio T, Jiménez I, Blake JG, Lohmann LG, Montiel OM (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *J Biogeogr* 35:105–116
- Mah CL, Blake DB (2012) Global diversity and phylogeny of the Asteroidea (Echinodermata). *PLoS ONE* 7:e35644
- Mainali KP, Warren DL, Dhileepan K, McConnachie A, Strathie L, Hassan G, Karki D, Shrestha BB, Parmesan C (2015) Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. *Glob Change Biol* 21:4464–4480
- Marshall CE, Glegg GA, Howell KL (2014) Species distribution modelling to support marine conservation planning: The next steps. *Mar Policy* 45:330–332
- Mathewson PD, Moyer-Horner L, Beever EA, Briscoe NJ, Kearney M, Yahn JM, Porter WP (2017) Mechanistic variables can enhance predictive models of endotherm distributions: the American pika under current, past, and future climates. *Glob Change Biol* 23:1048–1064
- Maxwell DL, Stelzenmüller V, Eastwood PD, Rogers SI (2009) Modelling the spatial distribution of plaice (*Pleuronectes platessa*), sole (*Solea solea*) and thornback ray (*Raja clavata*) in UK waters for marine management and planning. *J Sea Res* 61:258–267
- McClintock JB, Angus RA, Ho CP, Amsler CD, Baker BJ (2008) Intraspecific agonistic arm-fencing behavior in the Antarctic keystone sea star *Odontaster validus* influences prey acquisition. *Mar Ecol Prog Ser* 371:297–300
- Meier ES, Edwards TC Jr, Kienast F, Dobberty M, Zimmermann NE (2011) Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica*. *J Biogeogr* 38:371–382
- Merow C, Smith MJ, Silander JA Jr (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36:1058–1069

- Moles J, Figuerola B, Campaña-Llovet N, Monleón-Getino T, Taboada S, Avila C (2015) Distribution patterns in Antarctic and Subantarctic echinoderms. *Polar Biol* 38:799–813
- Moreau C, Saucède T, Jossart Q, Agüera A, Brayard A, Danis B (2017) Reproductive strategy as a piece of the biogeographic puzzle: a case study using Antarctic sea stars (Echinodermata, Asteroidea). *J Biogeogr* 44:848–860
- Moreau C, Mah C, Agüera A, Améziane N, Barnes D, Crokaert G, Eléaume M, Griffiths H, Guillaumot C, Hemery L, Jazdzewska A, Jossart Q, Laptikhovsky V, Linse K, Neill K, Sands C, Saucède T, Schiaparelli S, Sicinski J, Vasset N, Danis B (2018) Antarctic and sub-Antarctic Asteroidea database. *ZooKeys* 747:141
- Mormède S, Irisson JO, Raymond B (2014) Distribution modelling. In: De Broyer C, Koubbi P, Griffiths HJ, Raymond B, d'Udekem d'Acoz C, Van de Putte A, Danis B, David B, Grant S, Gutt J, Held C, Hosie G, Huettmann F, Post A, Ropert-Coudert Y (eds) *Biogeographic atlas of the Southern Ocean*. SCAR, Cambridge, p 510
- Murase H, Kitakado T, Hakamada T, Matsuoka K, Nishiwaki S, Naganobu M (2013) Spatial distribution of Antarctic minke whales (*Balaenoptera bonaerensis*) in relation to spatial distributions of krill in the Ross Sea, Antarctica. *Fish Ocean* 22:154–173
- Nachtsheim DA, Jerosch K, Hagen W, Plötz J, Bornemann H (2017) Habitat modelling of crabeater seals (*Lobodon carcinophaga*) in the Weddell Sea using the multivariate approach Maxent. *Polar Biol* 40:961–976
- Naimi B, Hamm NA, Groen TA, Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37:191–203
- Newbold T (2010) Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Prog Phys Geogr* 34:3–22
- Olden JD, Lawler JJ, Poff NL (2008) Machine learning methods without tears: a primer for ecologists. *Q Rev Biol* 83:171–193
- Pertierra LR, Aragón P, Shaw JD, Bergstrom DM, Terauds A, Olalla-Tárraga MÁ (2017) Global thermal niche models of two European grasses show high invasion risks in Antarctica. *Glob Change Biol* 23:2863–2873
- Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB (2011) *Ecological niches and geographic distributions* (MPB-49). vol. 56, Princeton University Press, Princeton
- Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A (2017) Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Glob Ecol Biogeogr* 26:275–287
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol Appl* 19:181–197
- Phillips ND, Reid N, Thys T, Harrod C, Payne NL, Morgan CA, White CA, Porter S, Houghton JD (2017) Applying species distribution modelling to a data poor, pelagic fish complex: the ocean sunfishes. *J Biogeogr* 44:2176–2187
- Pierrat B, Saucède T, Laffont R, De Ridder C, Festeau A, David B (2012) Large-scale distribution analysis of Antarctic echinoids using ecological niche modelling. *Mar Ecol Prog Ser* 463:215–230
- Qiao H, Soberón J, Peterson AT (2015) No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Method Ecol Evol* 6:1126–1136
- Reiss H, Cunze S, König K, Neumann H, Kröncke I (2011) Species distribution modelling of marine benthos: a North Sea case study. *Mar Ecol Prog Ser* 442:71–86
- Reiss H, Birchenough S, Borja A, Buhl-Mortensen L, Craeymeersch J, Dannheim J, Darr A, Glaparsoro I, Gogina M, Neumann H, Populus J, Rengstorff AM, Valle M, van Hoey G, Zettler ML, Degreer S (2014) Benthos distribution modelling and its relevance for marine ecosystem management. *ICES J Mar Sci* 72:297–315
- Ridgeway G (2015) gbm: generalized boosted regression models. R package version 2.1.1. <https://CRAN.R-project.org/package=gbm>
- Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Montfort JJ, Schroder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40:913–929
- Ross RE, Howell KL (2013) Use of predictive habitat modelling to assess the distribution and extent of the current protection of 'listed' deep-sea habitats. *Divers Distrib* 19:433–445
- Saucède T, Pierrat B, David B, In: De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C, Van de Putte AP, Danis B, David B, Grant S, Gutt J, Held C, Hosie G, Huettmann F, Post A, Ropert-Coudert Y (eds) (2014) *Biogeographic Atlas of the Southern Ocean*. Chapter 5.26. Echinoids, p. 510. Cambridge, SCAR. ISBN 978-0-948277-28-3
- Segurado P, Araujo MB (2004) An evaluation of methods for modelling species distributions. *J Biogeogr* 31:1555–1568
- Segurado P, Araujo MB, Kunin WE (2006) Consequences of spatial autocorrelation for niche-based models. *J Appl Ecol* 43:433–444
- Stockwell DR, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. *Ecol Model* 148:1–13
- Synès NW, Osborne PE (2011) Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Glob Ecol Biogeogr* 20:904–914
- Thatje S (2012) Effects of capability for dispersal on the evolution of diversity in Antarctic benthos. *Integr Comp Biol* 52:470–482
- Tingley R, Vallinoto M, Sequeira F, Kearney MR (2014) Realized niche shift during a global biological invasion. *Proc Nat Acad Sci*. <https://doi.org/10.1073/pnas.1405766111>
- Turner J, Barrand NE, Bracegirdle TJ, Convey P, Hodgson DA, Jarvis M, Jenkins A, Marshall G, Meredith MP, Roscoe H, Shanklin J, French J, Goosse H, Guglielmin M, Gutt J, Jacobs S, Kennicutt MC, Masson-Delmotte V, Mayewski P, Navarro F, Robinson S, Scambos T, Sparrow M, Summerhayes C, Speer K, Klepikov A (2014) Antarctic climate change and the environment: an update. *Polar Rec* 50:237–259
- Václavík T, Meentemeyer RK (2009) Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? *Ecol Model* 220:3248–3258
- Valavanis VD, Pierce GJ, Zuur AF, Palialexis A, Saveliev A, Katara I, Wang J (2008) Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. In: Valavanis VD (ed) *Essential fish habitat mapping in the Mediterranean*. Springer, Dordrecht, pp 5–20
- Van der Putten WH, Macel M, Visser ME (2010) Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. *Philos Trans R Soc Lond B* 365:2025–2034
- van Proosdij AS, Sosef MS, Wieringa JJ, Raes N (2016) Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39:542–552
- Ward G, Hastie T, Barry S, Elith J, Leathwick JR (2009) Presence-only data and the EM algorithm. *Biometrics* 65:554–563
- Wernberg T, Smale DA, Tuya F, Thomsen MS, Langlois TJ, De Bettignies T, Rousseaux CS (2013) An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. *Nat Clim Change* 3:78



- Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP (2006) Why do we still use stepwise modelling in ecology and behaviour? *J Anim Ecol* 75:1182–1189
- Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Grytnes JA, Guisan A, Heikkinen RK, Høye TT, Kuhn I, Luoto M, Maiorano L, Nilsson MC, Normand S, Ockinger E, Schmidt NM, Termansen M, Timmermann A, Wardle DA, Aastrop P, Svenning JC (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol Rev* 88:15–30
- Wittmann ME, Barnes MA, Jerde CL, Jones LA, Lodge DM (2016) Confronting species distribution model predictions with species functional traits. *Ecol Evol* 6:873–879

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.