

Species distribution modelling of the Southern Ocean benthos: a review on methods, cautions and solutions

CHARLÈNE GUILLAUMOT ^{1,2}, BRUNO DANIS ¹ and THOMAS SAUCÈDE ²

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

²UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, 6 bd Gabriel F-21000, Dijon, France
charleneguillaumot21@gmail.com

Abstract: Species distribution modelling studies the relationship between species occurrence records and their environmental setting, providing a valuable approach to predicting species distribution in the Southern Ocean (SO), a challenging region to investigate due to its remoteness and extreme weather and sea conditions. The specificity of SO studies, including restricted field access and sampling, the paucity of observations and difficulties in conducting biological experiments, limit the performance of species distribution models. In this review, we discuss some issues that may influence model performance when preparing datasets and calibrating models, namely the selection and quality of environmental descriptors, the spatial and temporal biases that may affect the quality of occurrence data, the choice of modelling algorithms and the spatial scale and limits of the projection area. We stress the importance of evaluating and communicating model uncertainties, and the most common evaluation metrics are reviewed and discussed accordingly. Based on a selection of case studies on SO benthic invertebrates, we highlight important cautions to take and pitfalls to avoid when modelling the distribution of SO species, and we provide some guidelines along with potential methods and original solutions that can be used for improving model performance.

Received 23 December 2020, accepted 8 April 2021

Key words: Antarctica, biases, limits, marine benthic invertebrates, modelling performance

Introduction

Due to its remoteness and extreme weather and sea conditions, the Southern Ocean (SO) is a challenging region in which to carry out biological studies (Kaiser *et al.* 2013, Gutt *et al.* 2017). It is also one of Earth's regions where we observe the most rapid and dramatic environmental changes in marine ecosystems, motivating the study of these marine communities (Turner *et al.* 2014, Ashton *et al.* 2017, Clark *et al.* 2019). Ecological modelling approaches are now well established and can be used to predict spatial patterns of organisms', populations' and species' distributions and assess their environmental drivers (Peterson *et al.* 2011). Based on field observations and experimental datasets, ecological modelling encompasses valuable approaches to helping to analyse biological data and interpolating our knowledge of species distributions in relation to environmental descriptors (Kennicutt *et al.* 2014).

Species distribution models (SDMs) are ecological models that study the statistical relationship between species occurrence records and environmental factors, determining the set of environmental conditions that is suitable to a species distribution (Elith *et al.* 2006, Elith & Leathwick 2009, Peterson *et al.* 2011). They represent the species realized niche (Pearson 2007, Sillero 2011),

being the ensemble of abiotic conditions in which the species survives and reproduces, adding into consideration the influence of biotic interactions (competition, predation, parasitism, symbiosis, etc.) (Hutchinson 1957). SDMs have been widely used in various fields of ecology, such as conservation biology, biogeography, palaeoecology and global change biology (Pearson 2007). In recent years, a growing number of ecological studies have used SDMs to analyse the distribution of marine pelagic and benthic species in the SO (e.g. marine invertebrates, fish, seabirds and marine mammals) and to determine species environmental preferences (Loots *et al.* 2007, Pierrat *et al.* 2012, Xavier *et al.* 2016, Nachtsheim *et al.* 2017), to compare ecological niche predictions in response to changing environments (Basher & Costello 2016, Gallego *et al.* 2017, Guillaumot *et al.* 2018b, Jerosch *et al.* 2019) or to identify diversity hotspots for conservation purposes (Pinkerton *et al.* 2010, Hibberd 2016, Thiers *et al.* 2017).

However, the quality of ocean-wide models is often limited by the heterogeneity, amount and spatial distribution of data, along with limited temporal and spatial resolutions. For all of these reasons, both modelling methods and model construction should be tested for accuracy and robustness prior to interpretation, and these indicators should be transparently communicated to ensure

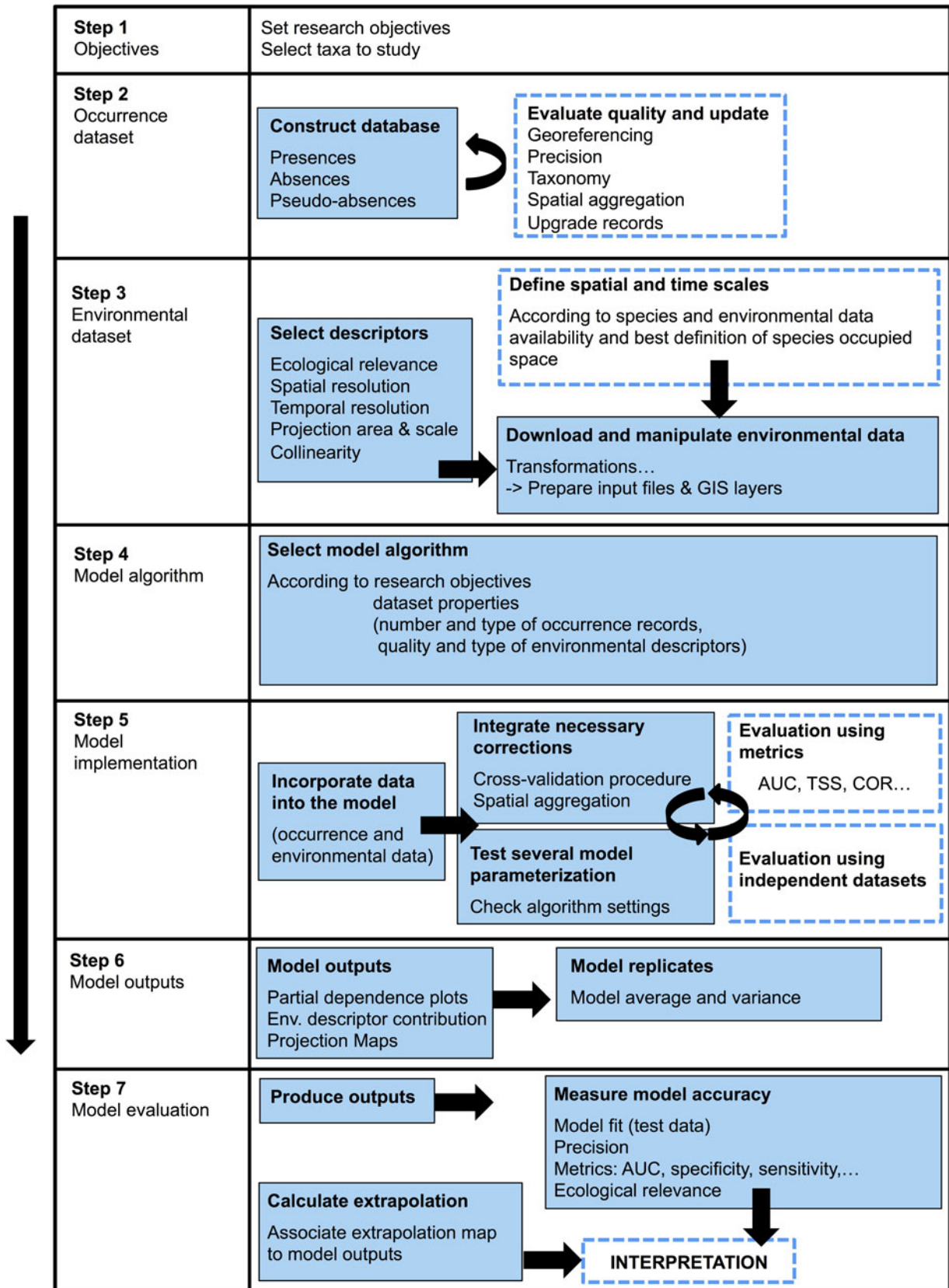


Fig. 1. Flow chart of the species distribution model construction process. Steps 1–4 concern data collection and treatment. Steps 5–7 integrate procedures for model implementation and validation. Dashed rectangles allow for a possible step backwards when assessing model uncertainties or evaluating model performance. AUC = area under the receiver operating curve; COR = Pearson correlation; GIS - Georeferencing Information System; TSS = true skill statistic.

that model outputs are relevant given the specificities of the datasets used for modelling.

In the present paper, we review the most common methodological issues encountered in species distribution modelling applied to the SO, following the flowchart in Fig. 1. Challenges regarding occurrence and environmental dataset peculiarities are described. The choice of SDM algorithm and procedures to implement and evaluate models are addressed. Based on benthic invertebrate case studies, we stress important precautions to take and pitfalls to avoid during common steps of SDM implementation. Finally, we aim to provide some guidelines with a set of potential methods and original solutions that can be used for improving model performance.

Quality of datasets

Environmental datasets: field data

Preparing environmental datasets is the first encountered challenge when generating models (Gutt *et al.* 2012, De Broyer *et al.* 2014). The SO, here defined as waters south of 45°S latitude, covers an extensive area of > 20 million km² (Breitzke 2014). Having access to environmental data with good temporal and spatial resolutions at such a broad scale is challenging, an issue common to all broad-scale oceanographic studies (Robinson *et al.* 2017). 'Broad scale' is defined here as the entire SO, 'regional scale' as smaller areas of a few hundred square kilometres and 'local scale' as a few square kilometres to square metres (Gage 2004).

Oceanographic data acquisition in the field is strongly conditioned by weather and sea conditions along with the seasonality of polar regions (polar night and dense sea-ice coverage in winter) that prompt recurring gaps in the acquisition of environmental data in the SO. Data are also much more frequently sampled close to research stations and along main sailing routes (Guillaumot *et al.* 2019). This is particularly striking in regions such as the south-western Weddell Sea, along the shores of the western Antarctic Peninsula and in the Bellingshausen and Amundsen seas (Clarke *et al.* 2007, Griffiths *et al.* 2014).

Environmental datasets: satellite-derived data

Satellite-derived data form a significant source of information for SO oceanographic studies. Providing valuable environmental indicators at broad spatial scale, they can give details about continuous and long-term measurements of water masses including sea-ice coverage, extent and duration, sea-surface temperatures and salinities, biogeochemical parameters, sea level, primary production and typical meteorological parameters (El Mahrad *et al.* 2020).

The accuracy of satellite data, however, should be considered with care, given detection limits, interpolations

that reduce the influence of atmospheric particulate scatter and the use of interpolation and gap-filling methods that smooth raw data at broad spatial and temporal scales (Pope *et al.* 2017, Stock *et al.* 2020).

Whenever possible, it is recommended to validate environmental data derived from satellite products at regional and local scales by comparing pixels on a satellite image with 'real' field observation data (Henson *et al.* 2015, Trull *et al.* 2018). Simple correlation analyses or more complex ground-truth processes are available to compare satellite and *in situ* data and to secure the interpretation of satellite-derived products (White-Newsome *et al.* 2013, Allan 2014). This, however, constitutes a huge task and is not performed generally before implementing SDMs.

Environmental datasets: access to datasets

Environmental data generated at the scale of the entire SO can be accessed for free through different web portals such as the NASA's OceanColor Web (<https://oceancolor.gsfc.nasa.gov>), where satellite-derived data, averaging different temporal measurements down to 4 km resolution, are available at the scale of the entire SO dating from 2000. These images are post-processed to characterize sea-surface temperature or ocean colour as proxies of surface productivity.

The National Oceanic and Atmospheric Administration's (NOAA) data centre (WOCE2013; <https://www.nodc.noaa.gov/OC5/woa13/woa13data.html>) also makes available post-processed data of ocean temperature, salinity, oxygen concentration and nutrients at different grid formats, down to 0.25° resolution, averaging over six decades (from 1955 to 2012). Bio-ORACLE (<https://www.bio-oracle.org>) compiles a large panel of marine data layers at 1° spatial resolution for different depth layers and time periods, for the present (2005–2012) and the future (2040–2050; 2090–2100) (Assis *et al.* 2018). Finally, GEBCO (<https://www.gebco.net>) is the reference platform for very-high-resolution bathymetry data (~500 m resolution) of the world's oceans.

Several works also make available compilations of these SO datasets dedicated to ecological modelling in the SO; they represent a valuable source of information for starting data preparation and modelling (https://data.aad.gov.au/metadata/records/Polar_Environmental_Data, https://data.aad.gov.au/metadata/records/fulldisplay/environmental_layers, https://data.aad.gov.au/metadata/records/Environmental_data_Southern_Ocean).

An increasing amount of environmental data collected during SO oceanographic campaigns have been made accessible for regional-scale studies. Several web portals aggregate all of these field measurements and provide them open access (e.g. <https://www.marine-geo.org/collections/#!/collection/USAP#summary>; <https://www.pangaea.de>).

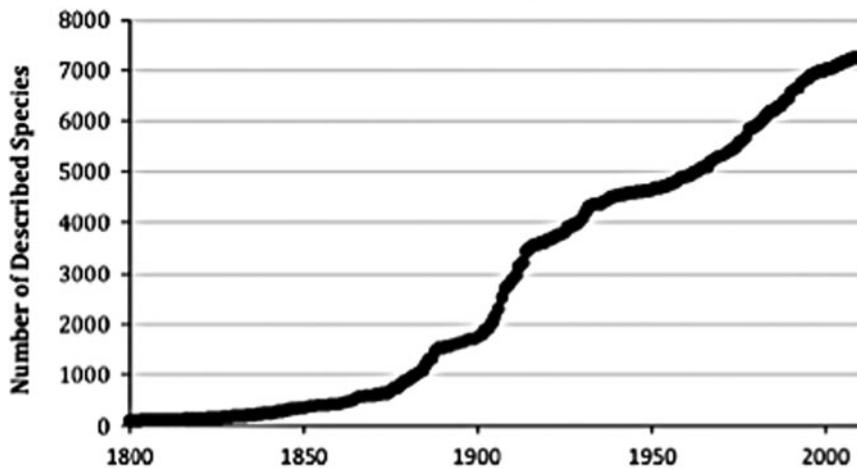


Fig. 2. Cumulative number of Antarctic species described over time, according to data available in the Register of Antarctic Marine species (until March 2010). From De Broyer & Danis (2011).

Environmental datasets: spatial and temporal resolutions

Most environmental data are accessible through broad-scale maps from the aforementioned data portals and are available with a finest spatial resolution of ~ 4 km, if not coarser (https://data.aad.gov.au/metadata/records/Polar_Environmental_Data, De Broyer *et al.* 2014, https://data.aad.gov.au/metadata/records/fulldisplay/environmental_layers, https://data.aad.gov.au/metadata/records/Environmental_data_Southern_Ocean). This low resolution strongly hampers the precise assessment of relationships between species occurrences and environmental descriptors (Pittman 2017, Staveley *et al.* 2017) and consequently the accuracy of model predictions (Connor *et al.* 2018), because the relevance of environmental descriptors represents a trade-off between their resolution and their spatial and temporal coverage (Guisan *et al.* 2007, Seo *et al.* 2009, Lauzeral *et al.* 2013, Vale *et al.* 2014). It is recommended that the resolution of environmental descriptors used in SDM should be in line with the scale of ecological processes at play and for which species ecophysiological responses show the highest variations, if models are expected to capture most species-environment relationships (Austin & Van Niel 2011).

The published environmental datasets are often averaged over relatively long periods of time (from years to decades for WOCE2013 or Bio-ORACLE). The analysis of inter-annual variations can complement the interpretation of model predictions: the absence of such information does not preclude running models, but this should be kept in mind when it comes to interpreting model outputs (Guillaumot *et al.* 2018a). Important environmental variations within a reference time period may not satisfy the equilibrium criterion between species distribution and environmental conditions, which is a strong prerequisite of SDM (Elith *et al.* 2006) and may affect the relevance and accuracy of model predictions

(Guillaumot *et al.* 2018a). In this respect, an alternative for improving modelling performance would be using seasonal averages or extreme values as environmental descriptors rather than pluri-annual to annual averages (Franklin 2009, Bradie & Leung 2017).

Environmental datasets: cartographic projections

Considering the poles in numerical analyses has long been a source of difficulty in spatial modelling as the convergence of meridians distorts shapes, surfaces, angles or distances towards high latitudes when using standard cylindrical representations such as the Mercator projection (Deleersnijder *et al.* 1993, Eby & Holloway 1994, Murray 1996). Working with conical or azimuthal projections (e.g. polar stereographic system) helps maintain the consistency of angles and shapes and therefore better meets the requirements of SO studies, although areas and distances are progressively distorted when moving away from the pole (Mulcahy & Clarke 2001).

Mapping environmental descriptors and projecting model predictions can be carried out with either square or hexagonal pixels. Each option does not alter image quality and hexagonal shapes may even offer some advantages (Kamgar-Parsi & Sander 1989, Tirunelveli *et al.* 2002). However, some contrasts may be present between images using square or hexagonal pixels, as each pixel measures the average environmental conditions in the considered surface (Vanden Berghe *et al.* 2013).

Subdividing the study area into sub-regions and using different pixel shapes can be a good solution for improving the relevance of representations (Vanden Berghe *et al.* 2013, M. Cryer, unpublished data 2015). Evaluating the accuracy of environmental values captured both in square and hexagonal pixels using baseline *in situ* field measurements can also be

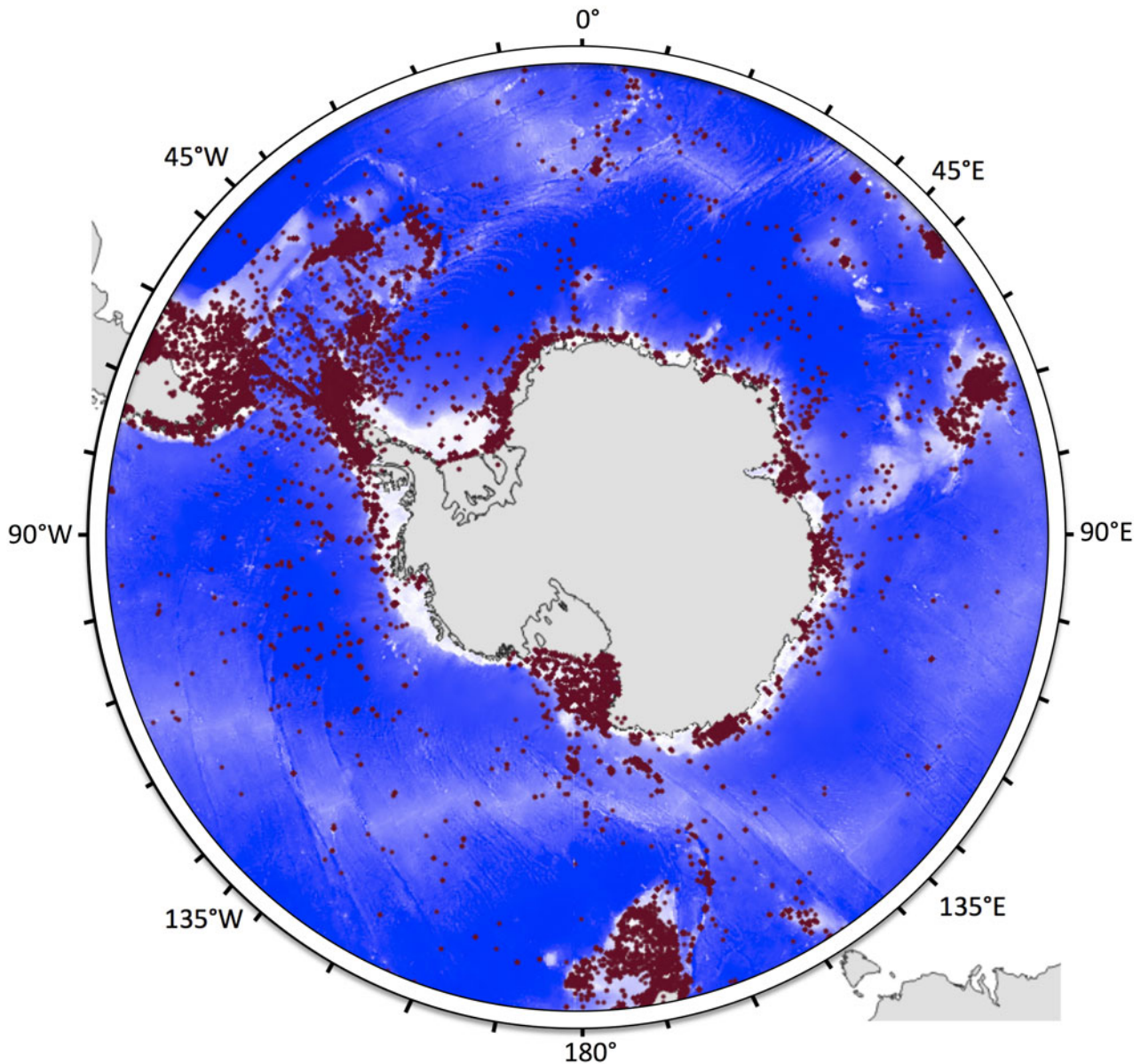


Fig. 3. Distribution of benthos sampling sites (red dots) in the Southern Ocean (SO, < 45°S). Sampling sites are not evenly distributed in the SO, showing important spatial aggregation in the Scotia Arc region and Western Antarctic Peninsula with several clusters along the Antarctic shelf, and over the Kerguelen and Campbell plateaus. In contrast, deep-sea regions and remote areas of the Antarctic shelf are under-sampled. From Guillaumot *et al.* (2019), updated from Griffiths *et al.* (2014).

suggested. This is yet to be tested for ecological modelling studies for the SO.

Environmental datasets: future forecasts

Since 1992, future climate models have been constantly updated through the efforts of the Coupled Model Intercomparison Projects (CMIP) featured by the Intergovernmental Panel on Climate Change (IPCC) Assessment Reports (ARs) with the aim of providing a

plausible representation of future climate linked to potential anthropogenic impacts (<https://www.ipcc.ch/site/assets/uploads/2018/03/sres-en.pdf>, Mearns *et al.* 2001). Recent updates (CMIP5 and CMIP6) of climate models are driven by different possible future greenhouse gas emission scenarios (Representative Concentration Pathways RCP2.6, RCP4.5, RCP6.0 and RCP8.5, from the least to the most pessimistic scenario for CMIP5, and Shared Socioeconomic Pathways SSP1 to SSP5 for CMIP6) and are built upon the average of an ensemble of simulations (Hayhoe *et al.* 2017). Future

climate models for the SO are available through two main online platforms: Bio-ORACLE (<https://www.bio-oracle.org>; Assis *et al.* 2018) and the NOAA's portal (<https://psl.noaa.gov/ipcc/ocn>).

The relevance of using future predictions based on global assessment scenarios for marine studies has been widely questioned (Flato *et al.* 2014, Frölicher *et al.* 2016, de la Hoz *et al.* 2018), including their use in SDMs, given that climate models mainly rely on untestable assumptions (Beaumont *et al.* 2008, Gotelli & Stanton-Geddes 2015, Cavanagh *et al.* 2017, Freer *et al.* 2018), future layers are not always available for oceanographic studies (Fabri-Ruiz 2018, Guillaumot *et al.* 2018a, 2018b), discrepancies between present observations and future predictions can be problematic (Jiménez-Valverde *et al.* 2021) and models are based on a representation of the climate system that has a complex cascading effect on ecological processes (Cavanagh *et al.* 2017). Cavanagh *et al.* (2017) examined how well IPCC-class models reproduced sea-ice conditions. By subsetting CMIP5 models that best describe spatial extent and temporal ice cover, they improved the precision of the projected future sea-ice distribution, which was better suited to ecological analyses. Extending this method to other key oceanographic parameters should contribute to improving the accuracy of future climate models for the SO and their relevance to ecological studies.

Occurrence datasets: historical compilation

Biological sampling in the SO began with the first expeditions of the HMS *Challenger* (1873–1876). Sampling effort has considerably increased over the second part of the twentieth century and during recent decades in particular, following technological advances that have enabled the access to remote regions and sample processing (Fig. 2).

This long-lasting and irregular effort in biogeographical (occurrence) data collection has had an impact on data compilation and has resulted in heterogeneous datasets, as observed in several data papers and associated Integrated Publishing Toolkit (IPT) databases such as Guillaumot *et al.* (2016), Fabri-Ruiz *et al.* (2017) or Moreau *et al.* (2018), or in the general platform *biodiversity.aq* web portal.

The historical compilation of biological data includes 1) taxon misidentifications and taxonomic inconsistencies due to the various taxonomic revisions published through time, 2) errors in the georeferencing of occurrence records due to contrasting nomenclatures used to report latitude and longitude, 3) the accumulation of errors in metadata through the different generations of curation and 4) errors due to the use of different coordinate projection systems. Finally, in cases where species distributions may have shifted with time, species environmental preferences may have changed or

non-contemporaneous environmental or occurrence datasets are used, discrepancies between occurrence records and environmental conditions can be present and violate the environment-occurrence equilibrium assumption necessary to generate SDMs. All of these side effects were reviewed in detail by Newbold (2010). The impacts on species niche definition and SDM predictions have been reported in many works (Ensing *et al.* 2012, Lahoz-Monfort *et al.* 2014, Monk 2014, Aguiar *et al.* 2015, Tassarolo *et al.* 2017, Guillaumot *et al.* 2018a) that all advise us to thoroughly check datasets for quality management prior to running models.

Occurrence datasets: spatial aggregation

Most species occurrence data were collected in the vicinity of research bases or their surroundings or along recurrent maritime routes, leading to clear spatial aggregation patterns in biological datasets (Fig. 3) (Griffiths *et al.* 2014, Guillaumot *et al.* 2019).

Spatial aggregation can affect model accuracy, as aggregated presence records do not fully and homogeneously represent the entire environment that is occupied by given species. This aggregation also violates an initial assumption of SDMs that requires independence between records (Araújo & Guisan 2006, Hijman 2012). This may bias model predictions (Luoto *et al.* 2005, Segurado *et al.* 2006, Dormann 2007, Kühn 2007, Crase *et al.* 2012), leading to statistical artefacts and generating inaccurate patterns (Bahn & McGill 2007, Currie 2007).

Spatial aggregation of data and the effect of this spatial aggregation on model outputs can be quantified using the Moran's *I* index, which estimates the spatial autocorrelation between the presence records used to build the model and predicted presence probabilities (Luoto *et al.* 2005). This spatial autocorrelation implies that close pixels are expected to present more similar predicted probabilities than distant ones due to the short geographical distance between records rather than environmental similarities alone. Testing and correcting for this bias should help to reduce its impact on model predictions (see the 'Correcting spatial biases' section) (Diniz-Filho *et al.* 2003, Kühn 2007).

Occurrence datasets: presence-only records

SDMs based on presence/absence data are recognized as having better predictive performance than models using presence-only data (Zaniewski *et al.* 2002, Brotons *et al.* 2004, Wisz & Guisan 2009, Lobo *et al.* 2010, Smith 2013, Carvalho *et al.* 2015, Peel *et al.* 2019). However, except for some local-scale studies (e.g. Robinson *et al.* 2011), in most oceanographic studies species absence records are usually not available for SDMs, and working with presence-only records is the only alternative (Lobo

et al. 2010). SDMs are then built by associating presence-only records with a random selection of background records that will be used to characterize the full environmental conditions (Franklin 2010, Barbet-Massin *et al.* 2012). Background records should not be mistaken for pseudo-absence records that are artificial absence data, where the species is supposed (but not confirmed) to be absent. Pseudo-absence records do not represent the overall conditions of the study area. Presence/pseudo-absence models represent another modelling approach, predicting occupied and unoccupied habitats rather than suitable and less suitable habitats for presence/background modelling (Sillero & Barbosa 2021).

Presence-only datasets may contain several uncertainties that can bias model predictions. 1) Working on rare or cryptic species is generally prone to taxonomic misidentifications that may either contract or, alternatively, expand the extent of predicted species distributions (Costa *et al.* 2015, Aubry *et al.* 2017). Such biases due to taxonomic errors were shown to be highly variable and to depend on experts identifying specimens, as suggested by Beale & Lennon (2012), who worked on a compilation of several collections. 2) Sampling gear may have an impact on species detection. Inaccurate species observations may generate false-positive results (species predicted as being present when they were not sampled or observed in the field) and false-negative results (species predicted as being absent when they were sampled or observed in the field) during model initialization (Guillera-Arroita 2016). Species presence records should be carefully scrutinized prior to modelling (Lozier *et al.* 2009), or at least records should be categorized into different subsets of data verifiability (Aubry *et al.* 2017). 3) Georeferencing errors are a frequent issue in databases (Murphey *et al.* 2004, Maldonado *et al.* 2015). This is especially the case in large databases compiling independent datasets using species presences recorded with varying levels of precision (Graham *et al.* 2008, Bloom *et al.* 2018). Several studies have simulated virtual random georeferencing errors and have shown that these errors lead to significant drops in model performance and inconsistencies in the respective contributions of environmental descriptor contributions, influencing model interpretation (Graham *et al.* 2008, Osborne & Leitão 2009, Naimi *et al.* 2011). These side effects seem to be minimized in local-scale models, here again advocating for the use of local-scale models whenever possible (Mitchell *et al.* 2017).

Occurrence datasets: dealing with small datasets

Usually, the number of species presence records available for modelling is relatively limited considering the wide geographical extent of the SO (De Broyer *et al.* 2014). Generating SDMs with small datasets may include many

pitfalls: 1) It reduces the potential of SDMs to transfer in space and time (Hernandez *et al.* 2006, Raes 2012), 2) it truncates predicted distribution and niche definition (Hortal *et al.* 2007, 2008, Rocchini *et al.* 2011, Sánchez-Fernández *et al.* 2011, Titeux *et al.* 2017, El-Gabbas & Dormann 2018), 3) it reduces modelling goodness-of-fit as the model may wrongly represent reality (Stockwell & Peterson 2002, McPherson *et al.* 2004, Pearson *et al.* 2007, Wisz *et al.* 2008, Liu *et al.* 2019), 4) it increases instability between model replicates (Guillaumot *et al.* 2018a), 5) it gives rise to methodological constraints on threshold selection (Jiménez-Valverde & Lobo 2007, Bean *et al.* 2012), 6) it gives rise to methodological constraints on the application of evaluation metrics (Pearson *et al.* 2007), 7) it complicates the identification of model optimal complexity (Galante *et al.* 2018) and 8) it leads to a reduction in model accuracy because presence and background datasets would not differ markedly (Luoto *et al.* 2005).

Alternatives are being developed to produce more accurate models based on a limited amount of presence records. One solution is generating several models performed on restricted areas and datasets with more detailed information and then averaging them with a weighted ensemble approach. This 'ensemble of small models' approach showed improved performance compared to single models (Lomba *et al.* 2010, Breiner *et al.* 2015, 2018).

Another alternative is to restrict the prediction area according to where occurrence records are found and ensuring upstream that the number of records is sufficient to precisely characterize the species environmental preferences: trivial advice that is surprisingly neglected, as recently pointed out by Morales *et al.* (2017) and Araújo *et al.* (2019).

Occurrence datasets: definition of species-occupied environmental space

Spatial aggregation, along with heterogeneity, limited size and uncertainties in datasets, can strongly bias the quantification of the species-occupied environmental space (Hortal *et al.* 2008, Newbold 2010, Tassarolo *et al.* 2017). However, accurately defining species-occupied space is the cornerstone of SDM initialization (Elith *et al.* 2006, Boulanger *et al.* 2018).

Moreover, SDMs suppose that species are in equilibrium with the environmental conditions that they inhabit. SDMs do not take into consideration potential vagrants that have dispersed out of their usual environmental range or populations that could momentarily survive in unsuitable habitats because doing so violates the equilibrium assumption between species distribution and environmental conditions (Beale & Lennon 2012). These elements should be cautiously considered when preparing datasets prior to generating models by removing any atypical records.

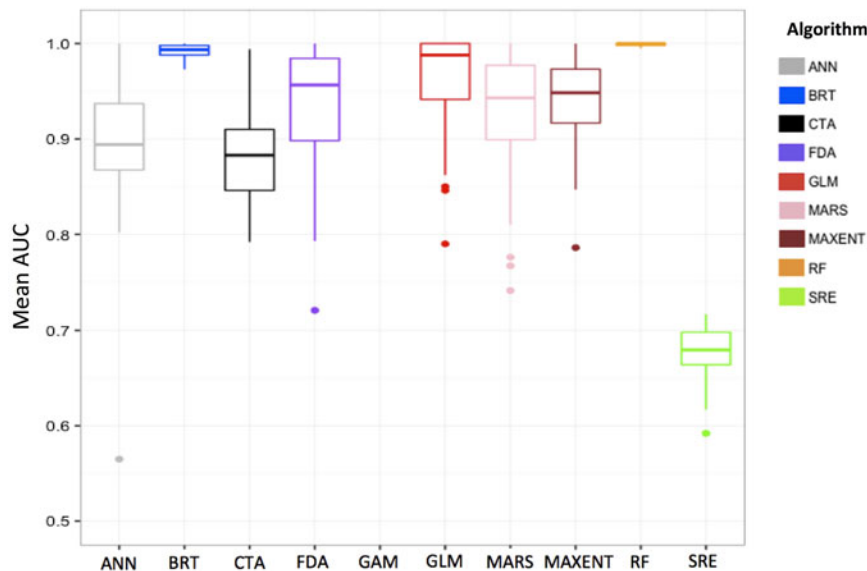


Fig. 4. Compared area under the receiver operating curve (AUC) performances of species distribution models generated with different algorithms (ANN = artificial neural network, BRT = boosted regression trees, CTA = classification tree analysis, FDA = flexible Discriminant analysis, GAM = generalized additive model, GLM = generalized linear model, MARS = multivariate adaptive regression splines, MAXENT = maximum entropy, RF = random forest, SRE = surface range envelope) to predict the distribution of the sea urchin *Sterechinus diadema* in the Southern Ocean. Results show a good performance for BRT and RF, adapted to small, historically compiled datasets (temporally heterogeneous) and spatially aggregated presence-only data. Models were calibrated with presence-only data and 200 background data randomly sampled in the study area. Average scores of 100 model replicates. See Guillaumot *et al.* (2018b) for details.

Over the last two decades, field data acquisition has expanded through the use of biologging technology with electronic devices attached to seabirds and marine mammals in order to access the positions of species all year long (Raymond *et al.* 2015, Ropert-Coudert *et al.* 2020). These data uncover the hidden behaviours of marine animals and constitute a powerful way of better estimating species-occupied space; they can also be used to validate and refine our understanding of the environmental conditions prevailing in those species distribution areas (Arthur *et al.* 2017, Nachtsheim *et al.* 2017, Hindell *et al.* 2020).

Adapting model implementation to datasets

The choice of modelling algorithms

To run performant SDMs, several assumptions must be tested and computing methods adapted to each case study (Austin 2002, de la Hoz *et al.* 2019). Among them, the choice of the modelling algorithm should be of major concern, as no algorithm works best for all species, in all areas, at all spatial scales and in all time periods (Jarnevich *et al.* 2015, Qiao *et al.* 2015). The selection and parameterization of modelling algorithms proved to be major causes of variation between SDM predictions (Dormann *et al.* 2008, Diniz-Filho *et al.* 2009, Buisson *et al.* 2010, Watling *et al.* 2015, Boulanger *et al.* 2018). Each algorithm is particularly suited for

dealing with a specific type and quality of data (Guisan & Zimmermann 2000, Austin 2002, Elith *et al.* 2006, Peterson 2011, Guisan *et al.* 2017), which will determine the final model outputs (Aguirre-Gutiérrez *et al.* 2013, Beaumont *et al.* 2016).

When modelling species distribution, it is necessary to select appropriate algorithms that have good transferability performances (i.e. have good abilities to correctly transfer predictions to other geographic space and time periods; Randin *et al.* 2006) and that they limit overfitting (i.e. mitigate model complexity) while being flexible in integrating complex environmental relationships. Machine-learning algorithms (e.g. maximum entropy (MaxEnt), boosted regression trees (BRTs), random forests (RFs), support vector machines (SVMs); Vapnik 1998, Breiman 2001, Elith *et al.* 2008, 2011) give access to important aspects of computing performance (Zhou 2012) and are relevant approaches for handling complex relationships between species occurrences and the environment (Olden *et al.* 2008, Elith & Leathwick 2009). The BRT and RF algorithms are particularly suited to complex and heterogeneous datasets (Fig. 4; Guillaumot *et al.* 2020a). They were proven to be efficient in generating performant models with limited overfitting (Elith *et al.* 2006, Wisz *et al.* 2008, Wenger & Olden 2012). They can automatically select the most informative features among a large set (Merow *et al.* 2014, García-Callejas & Araújo 2016, Guillaumot *et al.* 2020a) and perform well at

generalizing predictions in the absence of information or, conversely, at dealing with redundant information provided by correlated factors (Breiman 1984, De'ath & Fabricius 2000, Friedman 2001).

The different fields of application and the respective performance of existing algorithms have been extensively compared in various works based on 1) a single species (Pearson *et al.* 2006: plants in South Africa; Elith & Graham 2009: plant distribution in South Australian landscapes; Marmion *et al.* 2009: European butterflies; Lorena *et al.* 2011: plants in South America; Beaumont *et al.* 2016: mammals in Australia) or 2) an ensemble of worldwide distributed terrestrial (Elith *et al.* 2006) or marine species (Ready *et al.* 2010), for 3) certain regions only (Guisan *et al.* 2007: trees in Switzerland; Tsoar *et al.* 2007: snails, birds and bats in Israel; Reiss *et al.* 2011: benthic marine species in the North Sea; Bucklin *et al.* 2015: vertebrates of Florida) or 4) using virtual species (Meynard & Quinn 2007, Qiao *et al.* 2015, García-Callejas & Araujo 2016).

However, in order to generate such comparisons (Fig. 4), it is important to specifically adjust each algorithm to the case study. Algorithms all perform differently with regards to overfitting, spatial aggregation and transferability, and comparing model performances using different parameter settings is challenging (Merow *et al.* 2014) given that model parameterization has strong effects on the quality of model outputs (Anderson & Gonzalez 2011, Rodda *et al.* 2011, Warren & Seifert 2011, Yackulic *et al.* 2013, Radosavljevic & Anderson 2014, Moreno-Amat *et al.* 2015, Halvorsen *et al.* 2016, Galante *et al.* 2018, Lieske *et al.* 2018).

Initially developed in the 1990s, ensemble modelling has been increasingly used since then (Hansen & Salamon 1990, Schapire 1990). Ensemble modelling consists of combining several algorithms (Zhou 2012), input datasets (occurrence or environmental descriptors datasets) or parameterizations (Araújo & New 2007; and see Hao *et al.* 2019 for a review of applications). The approach is interesting as it can provide predictions that take into account the variability of several models (Araújo & New 2007, Hao *et al.* 2019).

Ensemble modelling has been used for various studies with SDMs (Araújo & New 2007, Marmion *et al.* 2009, Thuiller *et al.* 2009, Buisson *et al.* 2010, Luedeling *et al.* 2014, Trolle *et al.* 2014, Carvalho *et al.* 2015, Scales *et al.* 2016, Jerosch *et al.* 2019) and has benefitted from the development of R packages to implement them (*Biomod*: Thuiller *et al.* 2009; *BiodiversityR*: <https://cran.r-project.org/web/packages/BiodiversityR/index.html>; *biomod2*: <https://cran.r-project.org/package=biomod2>; *sdm*: <https://cran.r-project.org/web/packages/sdm/index.html>).

The main benefits of using ensemble models lie in the fact that the different algorithms will perform differently for various input cases (regardless of their overall

performance). The models thus complement each other, avoiding some biases that might have resulted from using a single algorithm (Marmion *et al.* 2009, Knutti 2010, Zhou 2012). However, model interpretation is much more difficult when mixing algorithms implemented differently, with contrasting ways of presenting outputs (Sillero 2011) and different definitions of thresholds for identifying habitat suitability (Perrault-Hébert 2019), requiring the normalization of predictions, which is rarely applied (Zhang & Mahadevan 2019). This is the main limitation to the approach and could offset the gains in model performance (Crimmins *et al.* 2013, Zhu & Peterson 2017, Hao *et al.* 2020). Such gains were contested, especially since model evaluation was often performed without using an independent evaluation dataset (Hao *et al.* 2019). Combining predictions of different models generated with contrasting assumptions is therefore tricky when interpreting results (Perrault-Hébert 2019). Optimizing the parameterization of a single algorithm (which could be correctly evaluated) may therefore constitute a more valuable approach (Perrault-Hébert 2019). Comparing the performance of different algorithms can be helpful in the first stage of the modelling process in order to select the most suitable algorithm and to calibrate the models (Massada *et al.* 2013).

The choice of environmental descriptors

The selection of environmental descriptors is also a crucial step in the modelling process (Franklin 2010, Austin & Van Niel 2011, Petitpierre *et al.* 2017). Ideally, environmental descriptors should be selected for their ecological relevance to the studied organisms (Austin & Van Niel 2011, Dormann *et al.* 2012, Bradie & Leung 2017), they must capture environmental discontinuities and constraints in the distribution area (Jarnevich *et al.* 2015) and they should also be detailed enough to represent the habitat complexity and variability in order to allow for good SDM accuracy and performance (Elith & Leathwick 2009, Barbet-Massin *et al.* 2012, Bucklin *et al.* 2015, Petitpierre *et al.* 2017).

In most studies, the final number of descriptors selected to depict the species environment is generally close to 10 (Pierrat *et al.* 2012, Mormède *et al.* 2014, Guillaumot *et al.* 2018a, Fabri-Ruiz *et al.* 2019). Overall, a small number of descriptors will allow for the generation of less complex models and facilitate interpretation (Austin & Van Niel 2011, Braunisch *et al.* 2013, Bucklin *et al.* 2015, Petitpierre *et al.* 2017). In contrast, increasing the number of descriptors potentially increases the effect of any collinearity between them (i.e. correlation between values of descriptors), which may lead to statistical artefacts in model predictions if the algorithms cannot handle information redundancy (Dormann *et al.* 2012,

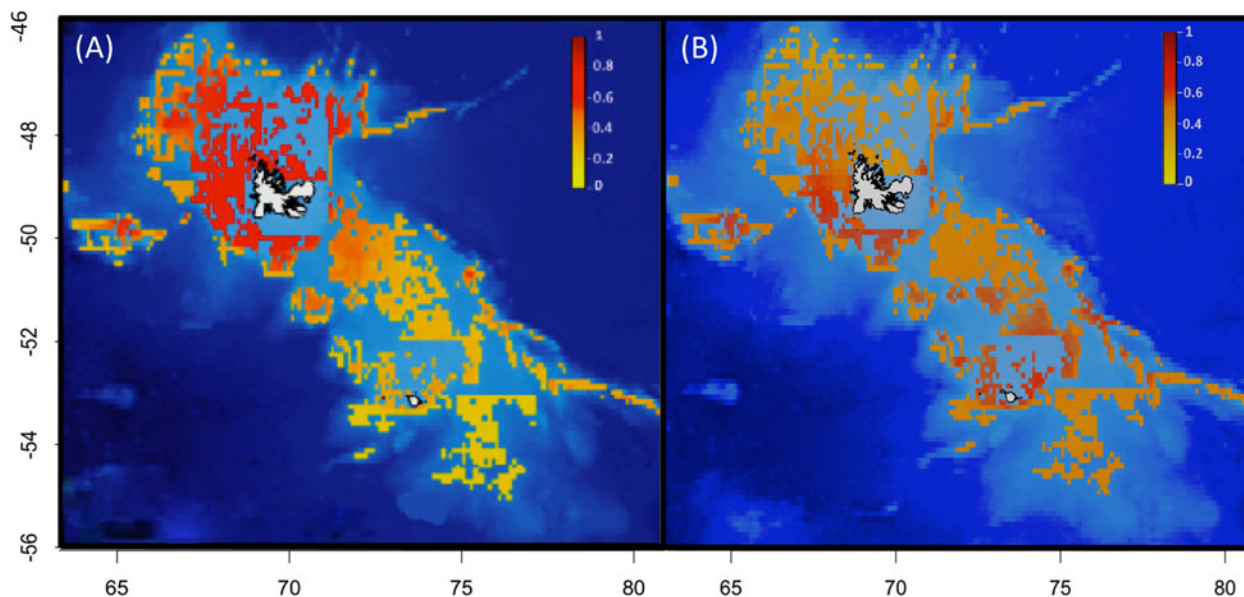


Fig. 5. Comparison of predicted distribution probabilities (between 0 and 1) of the sea urchin *Ctenocidaris nutrix* on the Kerguelen Plateau: **a.** without compensating for sampling bias; **b.** with a kernel density estimator (KDE) correction. More background data are sampled in highly sampled areas. The spatial aggregation of presence-only records near the shoreline of the Kerguelen Islands strongly biases model predictions. The KDE correction was proven to be efficient at correcting for such a bias and provides more relevant predictions. From Guillaumot *et al.* (2018a).

Merow *et al.* 2014). Therefore, collinearity is usually tested for beforehand and collinear descriptors are adjusted (in practice, one descriptor of a pair is removed) before running the model (Dormann *et al.* 2012, Merow *et al.* 2013, Fois *et al.* 2018). However, Guillaumot *et al.* (2020a) showed that model complexity, transferability and accuracy do not significantly change between models generated with different sets, including from 4 to 58 collinear descriptors when using the BRT algorithm. BRTs automatically keep the most relevant descriptors to describe species distribution and can deal with redundant information (De'ath & Fabricius 2000, Whittingham *et al.* 2006, Elith *et al.* 2008), which is not the case for all algorithms (Merow *et al.* 2014).

Selecting environmental descriptors therefore implies that several tests should be performed upstream in order to determine the best set to be used depending on research objectives. Fois *et al.* (2018) recommended first calibrating models with a large set of descriptors of various natures (proximal *vs* distal descriptors) that will be pruned stepwise after analysing their ability to accurately describe the habitat and after testing for collinearity (El-Gabbas & Dormann 2018). Generating, testing and comparing several sets of descriptors is a widespread strategy to target in a stepwise manner the set that gives the best predictive accuracy (Snickars *et al.* 2014, Bucklin *et al.* 2015, Bradie & Leung 2017, Petitpierre *et al.* 2017). Replacing environmental descriptors by principal components of a factorial

analysis also proved to be efficient because complex environmental gradients of the study area are simplified in fewer, orthogonalized components (Kühn 2007, Petitpierre *et al.* 2017). So far, this latter method has never been applied to SO case studies, and it should be tested in order to evaluate the interpretability of model results.

Correcting spatial biases

Generating a model based on spatially aggregated presence-only records may bias predictions with a higher probability of occurrence predicted in highly sampled areas (Dormann 2007, Guillaumot *et al.* 2018a). To compensate for such a bias, a first approach is to sample background records according to the spatial bias introduced by the aggregated presence records themselves (Phillips *et al.* 2009). The background dataset is used to define the environmental background: its boundaries and variability constitute essential information for building and projecting model outputs (Wisž & Guisan 2009, Barbet-Massin *et al.* 2012). The choice of the number of background records to be sampled and the extent of their distribution should be considered carefully when calibrating a model because it can strongly influence model predictions (Chefaoui & Lobo 2008, Lobo *et al.* 2010, Barbet-Massin *et al.* 2012, Jarnevich *et al.* 2017). This number should be with respect to the prevalence score, being the ratio between

the species-occupied space (represented by presence record locations) and the total surface of the study area (represented by background locations; McPherson *et al.* 2004). Some advice is provided in Barbet-Massin *et al.* (2012) for selecting the correct number of background records according to prevalence scores.

Targeting background records has been extensively tested, and several procedures have been developed to significantly improve the relevance of models (Fig. 5). Background records can be sampled within predefined areas (i.e. 'discs' or 'buffers') close to presence records (Hengl *et al.* 2009, Phillips *et al.* 2009, Fourcade *et al.* 2014, Bertrand *et al.* 2016), following the presence or absence of other species (Phillips *et al.* 2009, Syfert *et al.* 2013, Iturbide *et al.* 2015, Molloy *et al.* 2017, Phillips *et al.* 2017, Ranc *et al.* 2017), according to probabilities given by a kernel density estimator of the sampling frequency (Fourcade *et al.* 2014, Jarnevich *et al.* 2017, Guillaumot *et al.* 2018a, Fabri-Ruiz *et al.* 2019) or according to additive descriptors of accessibility and sampling effort (El-Gabbas & Dormann 2018). Once again, the selected method should be adapted to each case study and its efficiency tested prior to model interpretation (Støa *et al.* 2018).

A second method consists of filtering the available presence data to reduce the influence of the clustering of species records (Segurado *et al.* 2006, Kramer-Schadt *et al.* 2013, Boria *et al.* 2014). This is an efficient method compared to the background targeted sampling approach detailed above, but the remaining number of presence records after filtering should be sufficient to correctly determine species-occupied space (Kramer-Schadt *et al.* 2013). Reliable information should also be available to characterize the bias in species occurrence data (Aiello-Lammens *et al.* 2015, Sillero & Barbosa 2021). The filtering protocol requires meeting many prerequisites, but priority is given to keeping presence data independent and minimizing records clustering (D. Alagador, personal communication 2019).

Overall, if several methods are developed to correct for the effect of spatial aggregation on model outputs, it is recommended that one should interpret model projections performed for poorly sampled areas with great caution (Phillips *et al.* 2009, Iturbide *et al.* 2018).

Model outputs

Taxonomic bias and population variability

SDMs are usually parameterized using all presence records available for a species and all environmental conditions prevailing in the species records (Elith & Leathwick 2009). When modelling species distribution at a broad spatial scale, it is often assumed that all populations of a species have the same relationship to

environmental conditions over the entire distribution area (Pierrat *et al.* 2012, Xavier *et al.* 2016, Guillaumot *et al.* 2018b, Fabri-Ruiz *et al.* 2019). However, occurrence datasets may include a set of populations with different phenotypic plasticities (Chevin *et al.* 2010), transgenerational adaptations (Dixon *et al.* 2015) or simply different habitat selection in the case of vagile species. Therefore, the modelled species can actually present different abilities to respond to environmental changes. In particular, physiological performances of populations are likely to vary in marine species with wide distribution ranges and high dispersal capabilities over long distances (Thatje 2012). This is particularly relevant with regards to future predictions that do not integrate inter-population variability in the potential acclimation of species, and this may lead models to alternatively over- or under-estimate the distribution of species-suitable environments (Cacciapaglia & van Woesik 2017, Thyrring *et al.* 2017).

Phylogeographical studies have also regularly revealed the existence of cryptic species in the SO benthos, which show similar morphologies for distinct genotypes and potentially distinct ecological requirements and geographical distributions (Lozier *et al.* 2009). Such studies often stress the need for taxonomic revisions (González-Wevar *et al.* 2019, Ocaranza-Barrera *et al.* 2019, Moreau *et al.* 2021). SDMs can be generated based on a spatial subdivision of presence records according to the genetic structure of taxa, and in a second step, the different predictions can be merged together to the broader scale (Knowles *et al.* 2007, Marcer *et al.* 2016, Cacciapaglia & van Woesik 2017, Ikeda *et al.* 2017, Roberts *et al.* 2017, Pardo-Gandarillas *et al.* 2018). However, defining the genetic structure of benthic species in the SO is a long-term endeavour that requires a constantly renewed sampling effort, considering the extent and complexity of the study area (Moreau *et al.* 2017, Fraser *et al.* 2018, Moore *et al.* 2018). Waiting for taxonomic revisions and enhanced sampling efforts to best depict relationships between genetic units and environmental conditions (Vandersteen 2011) and combining SDMs with experimental data or mechanistic approaches can be alternatives for taking into account the possible physiological contrasts between populations (Kearney & Porter 2009, Buckley *et al.* 2010, Kearney *et al.* 2010, Fordham *et al.* 2013, Briscoe *et al.* 2016, Feng & Papes 2017, López-Farrán *et al.* 2021).

Definition of the projection area

The limitations in the current knowledge of species distribution also affect the quality of information available for estimating their potential distribution (Thuiller *et al.* 2003). When the limits of species environmental ranges are not fully captured, this uncertainty can significantly impact the accuracy of SDM predictions (Hortal *et al.* 2007, 2008, Rocchini *et al.* 2011, Sánchez-Fernández *et al.* 2011, Titeux *et al.*

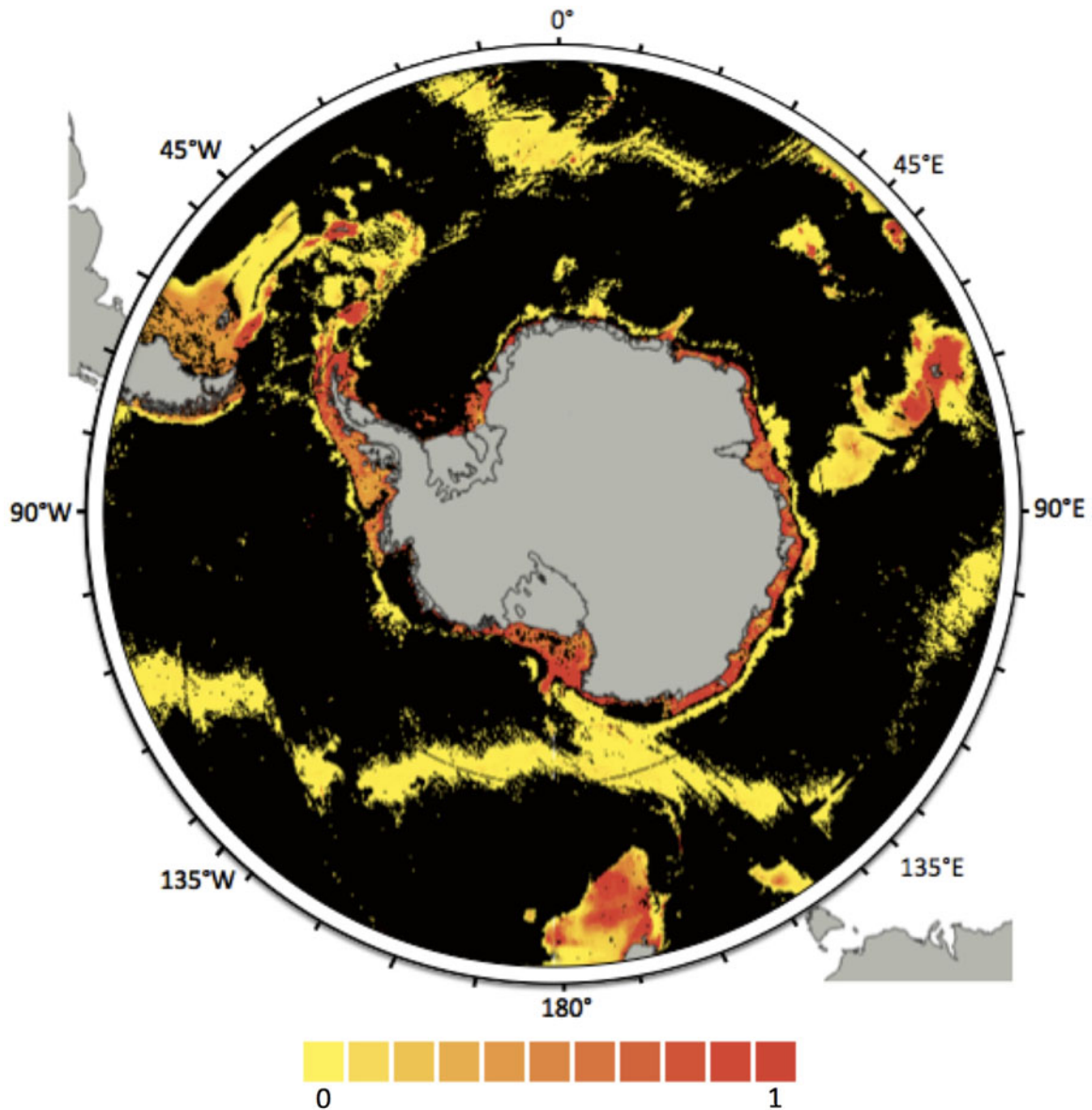


Fig. 6. Extrapolation map of the species distribution model generated for the sea star *Acodontaster hodgsoni*, with all presence-only records available. Extrapolation corresponds here to the ensemble of environmental conditions that are outside of the boundaries of the calibration range. The extrapolation area is displayed in black and covers 78.6% of the entire projection area; coloured pixels (yellow-red colour palette) show distribution probabilities (included between 0 and 1). Extracted from Guillaumot *et al.* (2020b).

2017, El-Gabbas & Dormann 2018). It reduces the applicability of models for predictive purposes (Thuiller *et al.* 2004), induces model overfitting (Tsoar *et al.* 2007, Barve *et al.* 2011, Guillaumot *et al.* 2018b) and can lead to overestimating the extent of suitable areas (Anderson & Raza 2010). This bias can be partly overcome by reducing the extent of the projection area to the known distribution of the available occurrence records (Anderson & Raza 2010) and by increasing knowledge regarding species ecology and physiology in order to identify the environmental conditions that are unsuitable for their survival or development (Byrne *et al.* 2016).

Model extrapolation

Models are said to extrapolate when a portion of the predicted area includes environmental conditions that are outside the range of values for which the model was calibrated. Model extrapolation may occur when model predictions are transferred, either in space or time. When extrapolated, model predictions are in non-analogue conditions compared to the initial calibration conditions because calibration data may not encompass the entire environmental range of each of the predictors (Guillaumot *et al.* 2020b). The set of projected

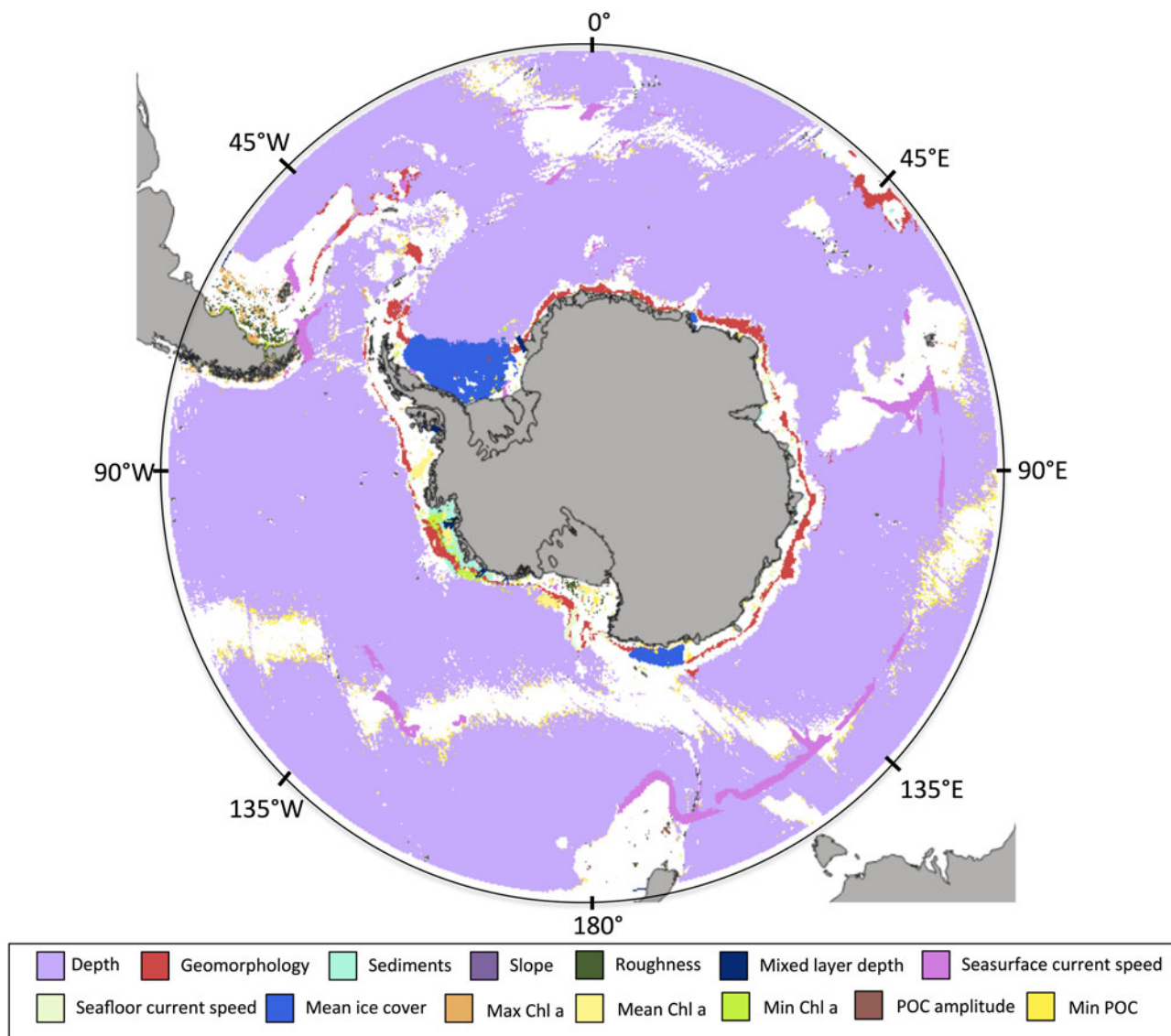


Fig. 7. Extrapolation map of the species distribution model generated for the sea star *Acodontaster hodgsoni* indicating environmental descriptors responsible for extrapolation (black pixels of Fig. 6 are here coloured according to the descriptor responsible for extrapolation; i.e. for each pixel, the predictor in question lies outside the calibration range). In this case study, 14 environmental descriptors are responsible for extrapolation, with depth being the main contributor. White pixels correspond to areas where the model does not extrapolate (the corresponding model predictions are shown in Fig. 6). POC stands for 'particulate organic carbon' and Chl *a* is the concentration in chlorophyll *a* on the sea surface. Generated from <https://cran.r-project.org/package=SDMPlay>.

environmental conditions can otherwise still be within the range of conditions, but specific combinations of environmental descriptors may be new, also leading to extrapolation (Mesgaran *et al.* 2014). In such conditions, predictions might be ecologically and statistically invalid and model interpretations inaccurate (Randin *et al.* 2006, Williams & Jackson 2007, Williams *et al.* 2007, Fitzpatrick & Hargrove 2009, Owens *et al.* 2013).

Among the different approaches, Elith *et al.* (2010) propose estimating and quantifying model extrapolation using the Multivariate Environmental Similarity Surface

(MESS) index to identify the most influential descriptors that lead to extrapolation. Grid-cell pixels for which at least one environmental descriptor has a value outside the range of environmental values defined by presence-only records (calibration range) are considered to be extrapolations. In these cases, the MESS index gives negative values and the ensemble of pixels containing negative values defines the extrapolation area (Elith *et al.* 2010, Guillaumot *et al.* 2020b). Most often, for SDMs performed at the scale of the SO, the number of records available to define the environmental space

occupied by species is limiting and the resolution of environmental descriptors relatively low (see the 'Quality of datasets' section). As a consequence, SDM projections sometimes include wide extrapolation areas that may cover > 75% of the predicted regions (Fig. 6; Guillaumot *et al.* 2020b).

In addition to quantifying the overall extrapolation area (Fig. 6), it is possible to fine-tune the analysis and define which environmental descriptors and areas are concerned with extrapolation (Fig. 7; Owens *et al.* 2013, <https://CRAN.R-project.org/package=SDMPlay>). Such information could be used to resample the environmental descriptors implemented in the model.

In any case, it has been recommended to provide information on model extrapolation and more generally to other concepts of uncertainties (species detection, errors, etc.), along with model predictions, because they are essential to accurate interpretation (Beale & Lennon 2012, Addison *et al.* 2013, Guisan *et al.* 2013).

Limiting model projections to 'realistic' depth ranges or some other environmental limiting factor based on a robust knowledge of species ecology (i.e. some expert-driven decision) was proven to be efficient at reducing extrapolation (Kearney & Porter 2009, Hare *et al.* 2012, De Villiers *et al.* 2013, Guillaumot *et al.* 2020b). Such a strategy is transitional until complementary samples and more comprehensive occurrence datasets are made available to better define the species-occupied space (Guillaumot *et al.* 2020b).

Model validation and accuracy of model predictions

Some common metrics for the evaluation of model predictions

Once models are generated, the accuracy of their predictions must be assessed in order to evaluate the validity of the models with regards to scientific issues to address, to compare different model outputs and to allow for the formulation of reliable interpretations (Zurell *et al.* 2020). Several metrics were developed in order to evaluate the performance of models (Fielding & Bell 1997, Allouche *et al.* 2006). Most of them are based on the calculation of an error matrix (or confusion matrix) that displays the proportion of presence and absence records that are correctly predicted by the model (Allouche *et al.* 2006).

In most biological studies focused on the SO benthos, absence records are usually unavailable and SDMs are generated based on a set of presence/background records (see the 'Occurrence datasets' section). As a consequence, the statistics that are commonly used for presence/absence datasets may not be appropriate for model evaluation (Wiley *et al.* 2003, Phillips *et al.* 2006,

Braunish *et al.* 2013), such as the κ statistic (Allouche *et al.* 2006). In contrast, the area under the curve or area under the receiver operating curve (AUC) is one of the most used and appropriate metrics for measuring the performance of model predictions based on presence/background data (Hand 2009). The AUC is an objective measure that remains stable with low-prevalence datasets (i.e. low frequency of occurrences with regards to the projection space) and is not sensitive to threshold effects (Manel *et al.* 2001, Hand 2009, van Proosdij *et al.* 2016). However, for presence/background models, specificity (the fraction of correctly predicted absences) might be overestimated when the number of background records is much higher than the number of presence-only records or when background and presences are associated with very different environmental values. This incidentally inflates AUC scores (Phillips *et al.* 2006, Raes & ter Steege 2007, Lobo 2008, Jiménez-Valverde 2012) and invalidates the relevance of the AUC metrics (van Proosdij *et al.* 2016).

Even when properly employed, the AUC cannot be used to compare models when SDMs are generated for different species based on different environmental descriptors or projected on distinct regions because the values depend on the relative size of suitable areas and prevalence scores may contrast (see the 'Correcting spatial biases' section) (Wisz *et al.* 2008, Anderson & Gonzalez 2011). The AUC metrics must be used as a simple measure of the relative ranking of model predictions associated with a specific dataset (El-Gabbas & Dormann 2018). Overall, each statistic is characterized by specific advantages and potential biases, so that it is recommended that one uses several statistics for evaluating model predictions (Allouche *et al.* 2006).

The accuracy of model predictions can also be evaluated by testing the classification of independent test data, where the available occurrence dataset can be split into independent subsets to train or test the model (for a review, see Fielding & Bell 1997).

Cross-validation procedures

Cross-validation procedures are aimed at evaluating model predictions using a subset of presence or absence records retrieved from the initial dataset used for modelling in order to assess how well the test data match with the modelled predictions (Bahn & McGill 2013). When working with presence-only datasets, two subsets of presence records are used: one subset is used to train the model (the training group) and the second subset is used to test the model (the test group). Test data and training data must be spatially independent from each other (Hijmans 2012, Bahn & McGill 2013). In most modelling exercises, standard cross-validation procedures are commonly used, in which the initial

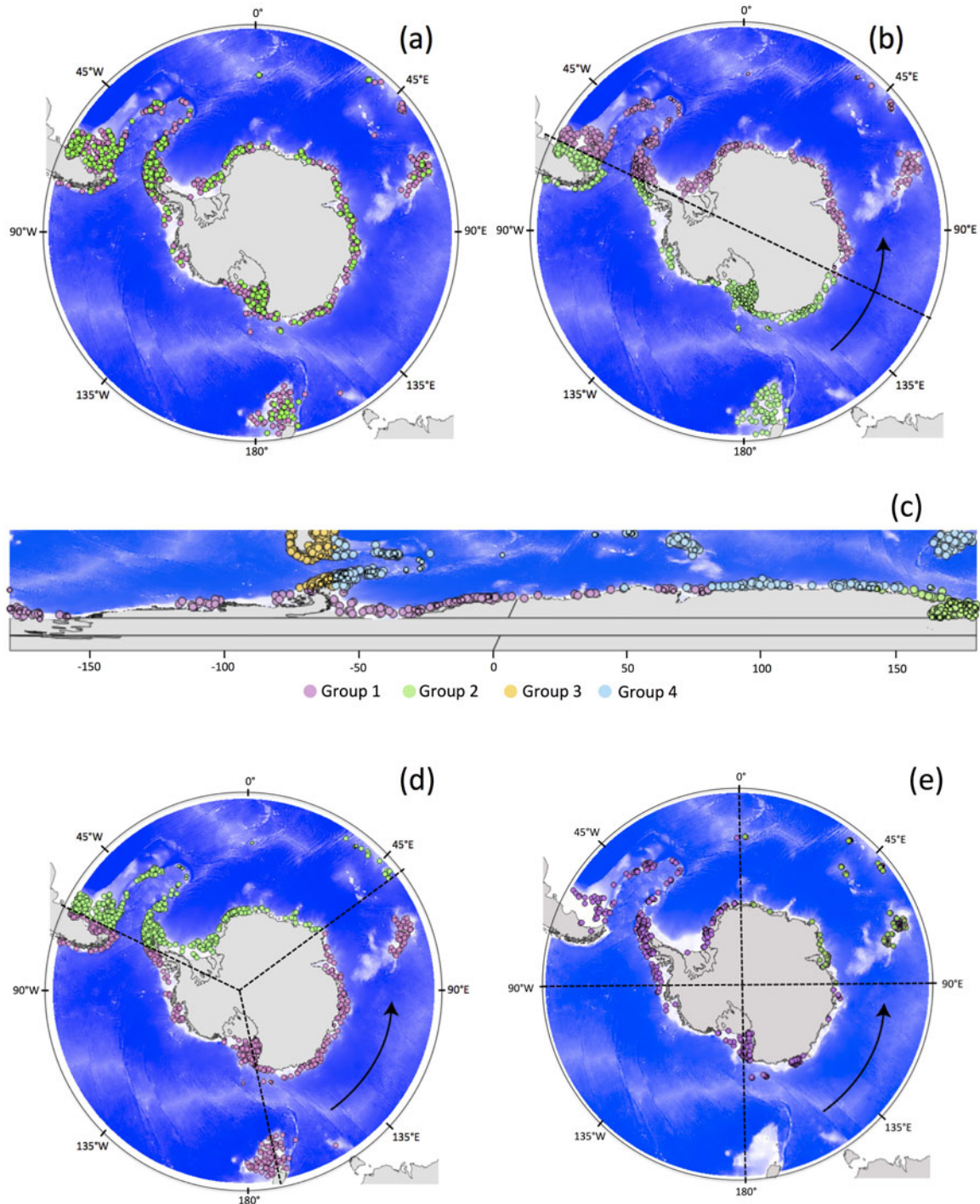


Fig. 8. Different cross-validation procedures based on the study of the sea star *Odontaster validus*, showing presence-only records and a random set of 1,000 background data selected according to a kernel density estimation weighting scheme from the dataset of Griffiths *et al.* (2014) on sampling effort of the Southern Ocean benthos. Data are split into training (pink) and test (green) subsets. The blue background corresponds to bathymetry and grey areas to emerged lands. **a.** Random cross-validation procedure, with a random split into 75% training and 25% test data. **b.** 'Twofold CLOCK' clustering by random spatial partition of the dataset into two groups (one training and one test). **c.** 'BLOCK' splitting, generated according to median latitudinal and longitudinal values (Muscarella *et al.* 2014). After the generation of four groups (corresponding to the four colours), one group is randomly defined as the test subset and the other three groups as the training subsets. A different system of projection was used to represent this map in order to highlight the latitudinal and longitudinal definition of the transects. **d.** 'Threefold CLOCK' clustering by random spatial partition of the dataset into three groups (two training and one test). **e.** 'Fourfold CLOCK' clustering by random spatial partition of the dataset into four groups (three training and one test). Extracted from Guillaumot *et al.* (2019).

presence dataset is randomly split into a training and test subset. Frequently, as previously discussed, presence data are spatially aggregated in SO datasets and the necessary condition of independence between training and test data is seldom met, making the model accuracy evaluation overly optimistic (Telford & Birks 2009, Hijmans 2012, Radosavljevic & Anderson 2014). In contrast to random procedures, spatial cross-validation procedures improve the performance of the validation step by spatially segregating the training and test subsets, ensuring the spatial independence between data even when they are spatially aggregated in the initial datasets (Dhingra *et al.* 2016, Roberts *et al.* 2017, Guillaumot *et al.* 2019; see also http://cran.rapporter.net/web/packages/blockCV/vignettes/BlockCV_for_SDM.html).

Several spatial cross-validation procedures have been proposed (Fig. 8), and the most appropriate one can be determined by comparing the different procedures in order to define the one that is most suitable for the study (Muscarella *et al.* 2014, Radosavljevic & Anderson 2014, Guillaumot *et al.* 2019, Valavi *et al.* 2019) depending on the spatial scale of the analysis, the number and spatial distribution of the presence data and the selected algorithm (and its associated complexity) used for modelling (El-Gabbas & Dormann 2018, Hao *et al.* 2020).

Conclusions and future prospects

This review summarizes some points and issues to be considered during SDM construction for modelling the distribution of SO species (Fig. 1). It shows that accurate and efficient SDMs can be produced for SO species when considering potential common biases and issues and correcting for their side effects. Proposed corrections must be adjusted to each case study: no consensus method nor implementation procedure always perform best - each case study requires proper analyses in order to generate the most relevant and accurate predictions. This means that, for each model, several procedures to implement the model should be tested in order to select the most suitable one, ideally giving priority to the availability of independent datasets for evaluating the models. We showed that SDMs perform best when the species-occupied space is accurately described, using extensive occurrence datasets with both presence and absence records, and when data are checked for positioning and georeferencing errors. A good knowledge of species' ecology, life history traits and populational variations within the overall species distribution and environmental range help to improve model quality (Fois *et al.* 2018). The compilation, examination and preparation of datasets prior to modelling are essential steps in generating efficient models. Estimating and communicating the uncertainties

associated with model predictions are also important tasks to be highlighted. This process may include a 'simple' interpretation of the ecological relevance of SDM outputs by experts (Merow *et al.* 2017) for the mapping of model extrapolations, as illustrated here. Model uncertainties are part of model outputs and should not be omitted (Guisan *et al.* 2013, Grimm *et al.* 2014, Grimm & Berger 2016).

Remaining challenges for constructing relevant SDMs for SO studies include more efforts regarding data collection outside of the main sampling hotspots and filling in knowledge gaps in SO species taxonomy. Some methodological perspectives, developed in other regions, address the integration of physiological information into SDMs. This facilitates the understanding of species environmental preferences and helps one to better estimate the niches of species (Kearney & Porter 2009, Talluto *et al.* 2016, Mathewson *et al.* 2017, Rodríguez *et al.* 2019, Gamliel *et al.* 2020). Such studies have recently been developed for SO benthic species: in López-Farrán *et al.* (2021), the combination of physiological experimental results and SDM projections allowed for the assessment of the invasive potential of the Patagonian crab *Halicarcinus planatus* (Fabricius, 1775) on Antarctic coasts, as was similarly done in Byrne *et al.* (2016) for the Arctic sea star *Asterias amurensis* (Lutken, 1871). Hybrid modelling approaches constitute another exciting approach, where information from both SDMs and physiological models are fully integrated, using the physiological information as a prior to inform the SDM (Gamliel *et al.* 2020). Recently applied to an endemic sea urchin of the Kerguelen Plateau (C. Guillaumot *et al.*, unpublished data 2021), the method allows for more precise prediction of the effects of seasonal variations on species habitat suitability.

Other interesting methodological approaches include the consideration of biotic interaction information, dispersal capacity estimates or population dynamics in complement to SDM predictions in order to generalize the understanding of the main drivers of species distribution (Pellissier *et al.* 2010, Meier *et al.* 2011, Pagel & Schurr 2012, Conlisk *et al.* 2013, Pellissier *et al.* 2013, Leach *et al.* 2016, Anderson 2017). These, however, necessitate a deep knowledge of the species ecology and of the surrounding environment, suggesting that their first applications should be expected in local- or regional-scale studies.

A final take-home message is that model outputs should be interpreted carefully and model predictions always considered with a critical eye. Models are simple representations of complex systems and should be used to complement other approaches in order to support conservation strategies or to address fundamental research objectives (Porfirio *et al.* 2014, Kampichler & Sierdsema *et al.* 2018).

Financial support

This work was supported by '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. 47 to the vERSO project (www.versoproject.be), funded by the Belgian Science Policy Office (BELSPO, contract n°BR/132/A1/vERSO). Research was also financed by the 'Refugia and Ecosystem Tolerance in the Southern Ocean' project (RECTO; BR/154/A1/RECTO) funded by the Belgian Science Policy Office (BELSPO), this study being contribution number 24.

References

- ADDISON, P.F., RUMPF, L., BAU, S.S., CAREY, J.M., CHEE, Y.E., JARRAD, F.C., *et al.* 2013. Practical solutions for making models indispensable in conservation decision-making. *Diversity and Distributions*, **19**, 490–502.
- AGUIAR, L.M., DA ROSA, R.O., JONES, G. & MACHADO, R.B. 2015. Effect of chronological addition of records to species distribution maps: the case of *Tonatia saurophila maresi* (Chiroptera, Phyllostomidae) in South America. *Austral Ecology*, **40**, 836–844.
- AGUIRRE-GUTIÉRREZ, J., CARVALHEIRO, L.G., POLCE, C., VAN LOON, E.E., RAES, N., REEMER, M. & BIESMEIJER, J.C. 2013. Fit-for-purpose: species distribution model performance depends on evaluation criteria - Dutch hoverflies as a case study. *PLoS One*, **8**, e63708.
- AIELLO-LAMMENS, M.E., BORJA, R.A., RADOSAVLJEVIC, A., VILELA, B. & ANDERSON, R.P. 2015. *spThin*: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, **38**, 541–545.
- ALLAN, M.G. 2014. *Remote sensing, numerical modelling and ground truthing for analysis of lake water quality and temperature*. Doctoral dissertation, University of Waikato.
- ALLOUCHE, O., TSOAR, A. & KADMON, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.
- ANDERSON, R.P. 2017. When and how should biotic interactions be considered in models of species niches and distributions? *Journal of Biogeography*, **44**, 8–17.
- ANDERSON, R.P. & GONZALEZ, I., JR. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecological Modelling*, **222**, 2796–2811.
- ANDERSON, R.P. & RAZA, A. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.
- ARAÚJO, M.B. & GUISAN, A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- ARAÚJO, M.B. & NEW, M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.
- ARAÚJO, M.B., ANDERSON, R.P., BARBOSA, A.M., BEALE, C.M., DORMANN, C.F., EARLY, R., *et al.* 2019. Standards for distribution models in biodiversity assessments. *Science Advances*, **5**, eaat4858.
- ARTHUR, B., HINDELL, M., BESTER, M., DE BRUYN, P.N., TRATHAN, P., GOEBEL, M. & LEA, M.A. 2017. Winter habitat predictions of a key Southern Ocean predator, the Antarctic fur seal (*Arctocephalus gazella*). *Deep-Sea Research II: Topical Studies in Oceanography*, **140**, 171–181.
- ASHTON, G.V., MORLEY, S.A., BARNES, D.K., CLARK, M.S. & PECK, L.S. 2017. Warming by 1°C drives species and assemblage level responses in Antarctica's marine shallows. *Current Biology*, **27**, 2698–2705.
- ASSIS, J., TYBERGHEIN, L., BOSCH, S., VERBRUGGEN, H., SERRÃO, E.A. & DE CLERCK, O. 2018. BioORACLE v2.0: extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, **27**, 277–284.
- AUBRY, K.B., RALEY, C.M. & MCKELVEY, K.S. 2017. The importance of data quality for generating reliable distribution models for rare, elusive, and cryptic species. *PLoS One*, **12**, e0179152.
- AUSTIN, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- AUSTIN, M.P. & VAN NIEL, K.P. 2011. Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, **38**, 1–8.
- BAHN, V. & MCGILL, B.J. 2007. Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, **16**, 733–742.
- BAHN, V. & MCGILL, B.J. 2013. Testing the predictive performance of distribution models. *Oikos*, **122**, 321–331.
- BARBET-MASSIN, M., JIGUET, F., ALBERT, C.H. & THUILLER, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, **3**, 327–338.
- BARVE, N., BARVE, V., JIMÉNEZ-VALVERDE, A., LIRA-NORIEGA, A., MAHER, S.P., PETERSON, A.T., *et al.* 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810–1819.
- BASHER, Z. & COSTELLO, M.J. 2016. The past, present and future distribution of a deep-sea shrimp in the Southern Ocean. *PeerJ*, **4**, e1713.
- BEALE, C.M. & LENNON, J.J. 2012. Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 247–258.
- BEAN, W.T., STAFFORD, R. & BRASHARES, J.S. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, **35**, 250–258.
- BEAUMONT, L.J., HUGHES, L. & PITMAN, A.J. 2008. Why is the choice of future climate scenarios for species distribution modelling important? *Ecology Letters*, **11**, 1135–1146.
- BEAUMONT, L.J., GRAHAM, E., DUURSMA, D.E., WILSON, P.D., CABRELLI, A., BAUMGARTNER, J.B., *et al.* 2016. Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges? *Ecological Modelling*, **342**, 135–146.
- BERTRAND, A., HABASQUE, J., HATTAB, T., HINTZEN, N.T., OLIVEROS-RAMOS, R., GUTIÉRREZ, M., *et al.* 2016. 3-D habitat suitability of jack mackerel *Trachurus murphyi* in the southeastern Pacific, a comprehensive study. *Progress in Oceanography*, **146**, 199–211.
- BLOOM, T.D., FLOWER, A. & DECHAIINE, E.G. 2018. Why georeferencing matters: introducing a practical protocol to prepare species occurrence records for spatial analysis. *Ecology and Evolution*, **8**, 765–777.
- BORJA, R.A., OLSON, L.E., GOODMAN, S.M. & ANDERSON, R.P. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, **275**, 73–77.
- BOULANGER, Y., PARIEN, M.A. & WANG, X. 2018. Model-specification uncertainty in future area burned by wildfires in Canada. *International Journal of Wildland Fire*, **27**, 164–175.
- BRADIE, J. & LEUNG, B. 2017. A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *Journal of Biogeography*, **44**, 1344–1361.
- 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. 2001. Random forests. *Machine Learning*, **45**, 5–32.
- BREIMAN, L., FRIEDMAN, J.H., OLSHEN, R.A. & STONE, C.J. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth International Group, 368 pp.

- BREINER, F.T., GUIBAN, A., BERGAMINI, A. & NOBIS, M.P. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, **6**, 1210–1218.
- BREINER, F.T., NOBIS, M.P., BERGAMINI, A. & GUIBAN, A. 2018. Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods in Ecology and Evolution*, **9**, 802–808.
- BREITZKE, M. 2014. Overview of seismic research activities in the Southern Ocean—quantifying the environmental impact. *Antarctic Science*, **26**, 80–92.
- BRISCOE, N.J., KEARNEY, M.R., TAYLOR, C.A. & WINTLE, B.A. 2016. Unpacking the mechanisms captured by a correlative species distribution model to improve predictions of climate refugia. *Global Change Biology*, **22**, 2425–2439.
- BROTONS, L., THUILLER, W., ARAÚJO, M.B. & HIRZEL, A.H. 2004. Presence-absence *versus* presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.
- BUCKLEY, L.B., URBAN, M.C., ANGILLETTA, M.J., CROZIER, L.G., RISSLER, L.J. & SEARS, M.W. 2010. Can mechanism inform species' distribution models? *Ecology Letters*, **13**, 1041–1054.
- BUCKLIN, D.N., BASILLE, M., BENSOTER, A.M., BRANDT, L.A., MAZZOTTI, F.J., ROMANACH, S.S., *et al.* 2015. Comparing species distribution models constructed with different subsets of environmental predictors. *Diversity and Distributions*, **21**, 23–35.
- BUISSON, L., THUILLER, W., CASAJUS, N., LEK, S. & GRENOUILLET, G. 2010. Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.
- BYRNE, M., GALL, M., WOLFE, K., & AGÜERA, A. 2016. From pole to pole: the potential for the Arctic seastar *Asterias amurensis* to invade a warming Southern Ocean. *Global Change Biology*, **22**, 3874–3887.
- CACCIAPAGLIA, C. & VAN WOESIK, R. 2017. Marine species distribution modelling and the effects of genetic isolation under climate change. *Journal of Biogeography*, **45**, 154–163.
- CARVALHO, B.M., RANGEL, E.F., READY, P.D. & VALE, M.M. 2015. Ecological niche modelling predicts southward expansion of *Lutzomyia* (*Nyssomyia*) *flaviscutellata* (Diptera: Psychodidae: Phlebotominae), vector of *Leishmania* (*Leishmania*) *amazonensis* in South America, under climate change. *PLoS One*, **10**, e0143282.
- CAVANAGH, R.D., MURPHY, E.J., BRACEGIRDLE, T.J., TURNER, J., KNOWLAND, C.A., CORNEY, S.P., *et al.* 2017. A synergistic approach for evaluating climate model output for ecological applications. *Frontiers in Marine Science*, **4**, 308.
- CHEFAOUL, R.M. & LOBO, J.M. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210**, 478–486.
- CHEVIN, L.M., LANDE, R. & MACE, G.M. 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biology*, **8**, e1000357.
- CLARK, M.S., NIEVA, L.V., HOFFMAN, J.I., DAVIES, A.J., TRIVEDI, U.H., TURNER, F., *et al.* 2019. Lack of long-term acclimation in Antarctic encrusting species suggests vulnerability to warming. *Nature Communications*, **10**, 1–10.
- CLARKE, A., GRIFFITHS, H.J., LINSE, K., BARNES, D.K. & CRAME, J.A. 2007. How well do we know the Antarctic marine fauna? A preliminary study of macroecological and biogeographical patterns in Southern Ocean gastropod and bivalve molluscs. *Diversity and Distributions*, **13**, 620–632.
- CONLISK, E., SYPHARD, A.D., FRANKLIN, J., FLINT, L., FLINT, A. & REGAN, H. 2013. Uncertainty in assessing the impacts of global change with coupled dynamic species distribution and population models. *Global Change Biology*, **19**, 858–869.
- CONNOR, T., HULL, V., VIÑA, A., SHORTRIDGE, A., TANG, Y., ZHANG, J., *et al.* 2018. Effects of grain size and niche breadth on species distribution modeling. *Ecography*, **41**, 1270–1282.
- COSTA, H., FOODY, G.M., JIMÉNEZ, S. & SILVA, L. 2015. Impacts of species misidentification on species distribution modeling with presence-only data. *ISPRS International Journal of Geo-Information*, **4**, 2496–2518.
- CRASE, B., LIEDLOFF, A.C. & WINTLE, B.A. 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography*, **35**, 879–888.
- CRIMMINS, S.M., DOBROWSKI, S.Z. & MYNSBERGE, A.R. 2013. Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling*, **266**, 126–130.
- CURRIE, D.J. 2007. Disentangling the roles of environment and space in ecology. *Journal of Biogeography*, **34**, 2009–2011.
- DHINGRA, M.S., ARTOIS, J., ROBINSON, T.P., LINARD, C., CHAIBAN, C., XENARIOS, I., *et al.* 2016. Global mapping of highly pathogenic avian influenza H5N1 and H5Nx clade 2.3. 4.4 viruses with spatial cross-validation. *Elife*, **5**, e19571.
- DE'ATH, G. & FABRICIUS, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, **81**, 3178–3192.
- DE BROYER, C. & DANIS, B. 2011. How many species in the Southern Ocean? Towards a dynamic inventory of the Antarctic marine species. *Deep-Sea Research II: Topical Studies in Oceanography*, **58**, 5–17.
- DE BROYER, C., KOUUBI, P., GRIFFITHS, H.J., RAYMOND, B., D'UDEKEM D'ACQZ, C., VAN DE PUTTE, A.P., *et al.* 2014. *Biogeographic atlas of the Southern Ocean*. Cambridge: Scientific Committee on Antarctic Research, 498 pp.
- DE LA HOZ, C.F., RAMOS, E., PUENTE, A. & JUANES, J.A. 2019. Temporal transferability of marine distribution models: the role of algorithm selection. *Ecological Indicators*, **106**, 105499.
- DE LA HOZ, C.F., RAMOS, E., ACEVEDO, A., PUENTE, A., LOSADA, Í.J. & JUANES, J.A. 2018. OCLE: a European open access database on climate change effects on littoral and oceanic ecosystems. *Progress in Oceanography*, **168**, 222–231.
- DE VILLIERS, M., HATTINGH, V. & KRITICOS, D.J. 2013. Combining field phenological observations with distribution data to model the temporal distribution of the fruit fly *Ceratitris rosa* Karsch (Diptera: Tephritidae). *Bulletin of Entomological Research*, **103**, 60–73.
- DELEERSNIJDER, E., VAN YPERSELE, J.P., & CAMPIN, J.M. 1993. An orthogonal curvilinear coordinate system for a world ocean model. *Ocean Modelling*, **100**, 7–10.
- DINIZ-FILHO, J.A., BINI, L.M. & HAWKINS, B.A. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53–64.
- DINIZ-FILHO, J.A., MAURICIO BINI, L., FERNANDO RANGEL, T., LOYOLA, R.D., HOF, C., NOGUÉS-BRAVO, D. & ARAÚJO, M.B. 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, **32**, 897–906.
- DIXON, G.B., DAVIES, S.W., AGLYAMOVA, G.V., MEYER, E., BAY, L.K. & MATZ, M.V. 2015. Genomic determinants of coral heat tolerance across latitudes. *Science*, **348**, 1460–1462.
- DORMANN, C.F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129–138.
- DORMANN, C.F., PURSCHKE, O., MÁRQUEZ, J.R., 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, C.F., ELITH, J., BACHER, S., BUCHMANN, C., CARL, G., CARRÉ, G., *et al.* 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.
- EBY, M., & HOLLOWAY, G. 1994. Grid transformation for incorporating the Arctic in a global ocean model. *Climate Dynamics*, **10**, 241–247.
- EL-GABBAS, A. & DORMANN, C.F. 2018. Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. *Ecology and Evolution*, **8**, 2196–2206.

- EL MAHRAD, B., NEWTON, A., ICELY, J.D., KACIMI, I., ABALANSA, S. & SNOUSSI, M. 2020. Contribution of remote sensing technologies to a holistic coastal and marine environmental management framework: a review. *Remote Sensing*, **12**, 2313.
- ELITH, J. & GRAHAM, C.H. 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, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.
- ELITH, J., KEARNEY, M. & PHILLIPS, S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.
- ELITH, J., LEATHWICK, J.R. & HASTIE, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.
- ELITH, J., PHILLIPS, S.J., HASTIE, T., DUDÍK, M., CHEE, Y.E. & YATES, C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.
- ELITH, J., GRAHAM, C., ANDERSON, R., DUDÍK, M., FERRIER, S., GUISAN, A., *et al.* 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- ENSING, D.J., MOFFAT, C.E. & PITHER, J. 2012. Taxonomic identification errors generate misleading ecological niche model predictions of an invasive hawkweed. *Botany*, **91**, 137–147.
- FABRI-RUIZ, S. 2018. *Modèles de distribution et changements environnementaux: application aux faunes d'échinides de l'océan Austral et écorégionalisation*. Doctoral dissertation, Université de Bourgogne Franche-Comté, Dijon, France.
- 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**, 1–20.
- FABRI-RUIZ, S., DANIS, B., DAVID, B. & SAUCÈDE, T. 2019. Can we generate robust species distribution models at the scale of the Southern Ocean? *Diversity and Distributions*, **25**, 21–37.
- FENG, X. & PAPEŠ, M. 2017. Can incomplete knowledge of species' physiology facilitate ecological niche modelling? A case study with virtual species. *Diversity and Distributions*, **23**, 1157–1168.
- FIELDING, A.H. & BELL, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- FITZPATRICK, M.C. & HARGROVE, W.W. 2009. The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, **18**, 2255.
- FLATO, G., MAROTZKE, J., ABIODUN, B., BRACONNOT, P., CHOU, S.C., COLLINS, W., *et al.* 2014. Evaluation of climate models. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, 741–866.
- FOIS, M., CUENA-LOMBRANA, 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. *Ecological Modelling*, **385**, 124–132.
- FORDHAM, D.A., MELLIN, C., RUSSELL, B.D., AKÇAKAYA, R.H., BRADSHAW, C.J., AIELLO-LAMMENS, M.E., *et al.* 2013. Population dynamics can be more important than physiological limits for determining range shifts under climate change. *Global Change Biology*, **19**, 3224–3237.
- FOURCADE, Y., ENGLER, J.O., RÖDDER, D. & SECONDI, J. 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS One*, **9**, e97122.
- FRANKLIN, J. 2009. *Mapping species distributions: spatial inference and prediction*. Cambridge: Cambridge University Press, 340 pp.
- FRANKLIN, J. 2010. Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, **16**, 321–330.
- FRASER, C.I., MORRISON, A.K., HOGG, A.M., MACAYA, E.C., VAN SEBILLE, E., RYAN, P.G., *et al.* 2018. Antarctica's ecological isolation will be broken by storm-driven dispersal and warming. *Nature Climate Change*, **8**, 704–708.
- FREER, J.J., PARTRIDGE, J.C., TARLING, G.A., COLLINS, M.A. & GENNER, M.J. 2018. Predicting ecological responses in a changing ocean: the effects of future climate uncertainty. *Marine Biology*, **165**, 7.
- FRIEDMAN, J.H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, **29**, 1189–1232.
- FRÖLICHER, T.L., RODGERS, K.B., STOCK, C.A. & CHEUNG, W.W. 2016. Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors. *Global Biogeochemical Cycles*, **30**, 1224–1243.
- GAGE, J.D. 2004. Diversity in deep-sea benthic macrofauna: the importance of local ecology, the larger scale, history and the Antarctic. *Deep-Sea Research II: Topical Studies in Oceanography*, **51**, 1689–1708.
- GALANTE, P.J., ALADE, B., MUSCARELLA, R., JANSA, S.A., GOODMAN, S.M. & ANDERSON, R.P. 2018. The challenge of modeling niches and distributions for data-poor species: a comprehensive approach to model complexity. *Ecography*, **41**, 726–736.
- GALLEGO, R., DENNIS, T.E., BASHER, Z., LAVERY, S. & SEWELL, M.A. 2017. On the need to consider multiphasic sensitivity of marine organisms to climate change: a case study of the Antarctic acorn barnacle. *Journal of Biogeography*, **44**, 2165–2175.
- GAMLIEL, I., BUBA, Y., GUY-HAIM, T., GARVAL, T., WILLETTE, D., RILOV, G. & BELMAKER, J. 2020. Incorporating physiology into species distribution models moderates the projected impact of warming on selected Mediterranean marine species. *Ecography*, **43**, 1090–1106.
- GARCÍA-CALLEJAS, D. & ARAÚJO, M.B. 2016. The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling*, **326**, 4–12.
- GONZÁLEZ-WEVAR, C.A., HÜNE, M., ROSENFELD, S., NAKANO, T., SAUCÈDE, T., SPENCER, H. & POULIN, E. 2019. Systematic revision of *Nacella* (Patellogastropoda: Nacellidae) based on a complete phylogeny of the genus, with the description of a new species from the southern tip of South America. *Zoological Journal of the Linnean Society*, **186**, 303–336.
- GOTELLI, N.J. & STANTON-GEDES, J. 2015. Climate change, genetic markers and species distribution modelling. *Journal of Biogeography*, **42**, 1577–1585.
- GRAHAM, C.H., ELITH, J., HIJMANS, R.J., GUISAN, A., TOWNSEND PETERSON, A., LOISELLE, B.A. & NCEAS Predicting Species Distributions Working Group. 2008. The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**, 239–247.
- GRIFFITHS, H.J., VAN DE PUTTE, A.P. & DANIS, B. 2014. Data distribution: patterns and implications. In DE BROYER, C., KOUUBI, P., GRIFFITHS, H.J., RAYMOND, B., UDEKEM D'ACÓZ, C.D', *et al.*, eds. *Biogeographic atlas of the Southern Ocean*. Cambridge: Scientific Committee on Antarctic Research, 16–26.
- GRIMM, V. & BERGER, U. 2016. Robustness analysis: deconstructing computational models for ecological theory and applications. *Ecological Modelling*, **326**, 162–167.
- GRIMM, V., AUGUSIAK, J., FOCKS, A., FRANK, B.M., GABSI, F., JOHNSTON, A.S., *et al.* 2014. Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. *Ecological Modelling*, **280**, 129–139.
- GUILLAUMOT, C., DANIS, B. & SAUCÈDE, T. 2020a. Selecting environmental descriptors is critical for modelling the distribution of Antarctic benthic species. *Polar Biology*, **43**, 1363–1381.
- 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. *Marine Ecology Progress Series*, **594**, 149–164.

- GUILLAUMOT, C., MOREAU, C., DANIS, B. & SAUCÈDE, T. 2020b. Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, **188**, 102438.
- 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**, 1–17.
- GUILLAUMOT, C., FABRI-RUIZ, S., MARTIN, A., ELÉAUME, M., DANIS, B., FÉRAL, J.P. & SAUCÈDE, T. 2018b. Benthic species of the Kerguelen Plateau show contrasting distribution shifts in response to environmental changes. *Ecology and Evolution*, **8**, 6210–6225.
- GUILLAUMOT, C., ARTOIS, J., SAUCÈDE, T., DEMOUSTIER, L., MOREAU, C., ELÉAUME, M., *et al.* 2019. Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography*, **175**, 198–207.
- GUILLERA-ARROITA, G. 2016. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, **40**, 281–295.
- GUISAN, A. & ZIMMERMANN, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- GUISAN, A., THUILLER, W. & ZIMMERMANN, N.E. 2017. *Habitat suitability and distribution models: with applications in R*. Cambridge: Cambridge University Press, 478 pp.
- GUISAN, A., GRAHAM, C. H., ELITH, J., HUETTMANN, F. & NCEAS Species Distribution Modelling Group. 2007a. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332–340.
- GUISAN, A., TINGLEY, R., BAUMGARTNER, J.B., NAUJOKAITIS-LEWIS, I., SUTCLIFFE, P.R., TULLOCH, A.I., *et al.* 2013. Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.
- GUTT, J., ISLA, E., BERTLER, A.N., BODEKER, G.E., BRACEGIRDLE, T., CAVANAGH, R.D., *et al.* 2017. Cross-disciplinarity in the advance of Antarctic ecosystem research. *Marine Genomics*, **37**, 1–17.
- GUTT, J., ZURELL, D., BRACEGIRDLE, T., CHEUNG, W., CLARK, M., CONVEY, P., *et al.* 2012. Correlative and dynamic species distribution modelling for ecological predictions in the Antarctic: a cross-disciplinary concept. *Polar Research*, **31**, 11091.
- HALVORSEN, R., MAZZONI, S., DIRKSEN, J.W., NÆSSET, E., GOBAKKEN, T. & OHLSON, M. 2016. How important are choice of model selection method and spatial autocorrelation of presence data for distribution modelling by MaxEnt? *Ecological Modelling*, **328**, 108–118.
- HAND, D.J. 2009. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, **77**, 103–123.
- HANSEN, L.K. & SALAMON, P. 1990. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 993–1001.
- HAO, T., ELITH, J., GUILLERA-ARROITA, G. & LAHOZ-MONFORT, J.J. 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, **25**, 839–852.
- HAO, T., ELITH, J., LAHOZ-MONFORT, J.J. & GUILLERA-ARROITA, G. 2020. Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, **43**, 549–558.
- HARE, J.A., WUENSCHER, M.J. & KIMBALL, M.E. 2012. Projecting range limits with coupled thermal tolerance-climate change models: an example based on gray snapper (*Lutjanus griseus*) along the US east coast. *PLoS One*, **7**, e52294.
- HAYHOE, K., EDMONDS, J., KOPP, R., LEGRANDE, A., SANDERSON, B., WEHNER, M. & WUEBBLES, D. 2017. Climate models, scenarios, and projections. In WUEBBLES, D.J., FAHEY, D.W., HIBBARD, K.A., DOKKEN, D.J., STEWART, B.C. & MAYCOCK, T.K. eds. *Climate science special report: a sustained assessment activity of the U.S. Global Change Research Program*. Washington, DC: US Global Change Research Program, 186–227.
- HENGL, T., SIERDSEMA, H., RADOVIĆ, A. & DILO, A. 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. *Ecological Modelling*, **220**, 3499–3511.
- HENSON, S.A., YOOL, A. & SANDERS, R. 2015. Variability in efficiency of particulate organic carbon export: A model study. *Global Biogeochemical Cycles*, **29**, 33–45.
- HERNANDEZ, P.A., GRAHAM, C.H., MASTER, L.L., & ALBERT, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.
- HIBBERD, T. 2016. *Describing and predicting the spatial distribution of benthic biodiversity in the sub-Antarctic and Antarctic*. Doctoral dissertation, University of Tasmania.
- HUJMAN, R.J. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679–688.
- HINDELL, M.A., REISINGER, R.R., ROPERT-COUDERT, Y., HÜCKSTÄDT, L.A., TRATHAN, P.N., BORNEMANN, H., *et al.* 2020. Tracking of marine predators to protect Southern Ocean ecosystems. *Nature*, **580**, 87–92.
- HORTAL, J., LOBO, J.M. & JIMÉNEZ-VALVERDE, A. 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology*, **21**, 853–863.
- HORTAL, J., JIMÉNEZ-VALVERDE, A., GÓMEZ, J.F., LOBO, J.M. & BASELGA, A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, **117**, 847–858.
- HUTCHINSON, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.
- IKEDA, D.H., MAX, T.L., ALLAN, G.J., LAU, M.K., SHUSTER, S.M. & WHITHAM, T.G. 2017. Genetically informed ecological niche models improve climate change predictions. *Global Change Biology*, **23**, 164–176.
- ITURBIDE, M., BEDIA, J. & GUTIÉRREZ, J.M. 2018. Background sampling and transferability of species distribution model ensembles under climate change. *Global and Planetary Change*, **166**, 19–29.
- ITURBIDE, M., BEDIA, J., HERRERA, S., DEL HIERRO, O., PINTO, M. & GUTIÉRREZ, J.M. 2015. A framework for species distribution modelling with improved pseudo-absence generation. *Ecological Modelling*, **312**, 166–174.
- JARNEVICH, C.S., STOHLGREN, T.J., KUMAR, S., MORISETTE, J.T. & HOLCOMBE, T.R. 2015. Caveats for correlative species distribution modeling. *Ecological Informatics*, **29**, 6–15.
- JARNEVICH, C.S., TALBERT, M., MORISETTE, J., ALDRIDGE, C., BROWN, C.S., KUMAR, S., *et al.* 2017. Minimizing effects of methodological decisions on interpretation and prediction in species distribution studies: an example with background selection. *Ecological Modelling*, **363**, 48–56.
- JEROSCH, K., SCHARF, F.K., DEREGIBUS, D., CAMPANA, G.L., ZACHER, K., PEHLKE, H., *et al.* 2019. Ensemble modelling of Antarctic macroalgal habitats exposed to glacial melt in a polar fjord. *Frontiers in Ecology and Evolution*, **7**, 207.
- JIMÉNEZ-VALVERDE, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, **21**, 498–507.
- JIMÉNEZ-VALVERDE, A. & LOBO, J.M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, **31**, 361–369.
- JIMÉNEZ-VALVERDE, A., RODRÍGUEZ-REY, M. & PEÑA-AGUILERA, P. 2021. Climate data source matters in species distribution modelling: the case of the Iberian Peninsula. *Biodiversity and Conservation*, **30**, 67–84.

- KAISER, S., BRANDÃO, S.N., BRIX, S., BARNES, D.K.A., BOWDEN, D.A., INGELS, J., *et al.* 2013. Patterns, processes and vulnerability of Southern Ocean benthos: a decadal leap in knowledge and understanding. *Marine Biology*, **160**, 2295–2317.
- KAMGAR-PARSI, B. & SANDER, W.A. 1989. Quantization error in spatial sampling: comparison between square and hexagonal pixels. In *Proceedings CVPR'89: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 604–611.
- KAMPICHLER, C. & SIERDSEMA, H. 2018. On the usefulness of prediction forecasts for local species distribution model forecasts. *Ecological Informatics*, **47**, 67–72.
- KEARNEY, M. & PORTER, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.
- KEARNEY, M.R., WINTLE, B.A. & PORTER, W.P. 2010. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, **3**, 203–213.
- KENNICUTT, M.C., CHOWN, S.L., CASSANO, J.J., LIGGETT, D., MASSOM, R., PECK, L.S., *et al.* 2014. Six priorities for Antarctic science. *Nature*, **512**, 23–25.
- KNOWLES, L.L., CARSTENS, B.C. & KEAT, M.L. 2007. Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology*, **17**, 940–946.
- KNUTTI, R. 2010. The end of model democracy? *Climatic Change*, **102**, 395–404.
- KRAMER-SCHADT, S., NIEBALLA, J., PILGRIM, J.D., SCHRÖDER, B., LINDENBORN, J., REINFELDER, V., *et al.* 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, **19**, 1366–1379.
- KÜHN, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, **13**, 66–69.
- LAHOZ-MONFORT, J.J., GUILLERA-ARROITA, G. & WINTLE, B.A. 2014. Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504–515.
- LAUZERAL, C., GRENOUILLET, G. & BROSSÉ, S. 2013. Spatial range shape affects the grain size effects in species distribution models. *Ecography*, **36**, 778–787.
- LEACH, K., MONTGOMERY, W.I. & REID, N. 2016. Modelling the influence of biotic factors on species distribution patterns. *Ecological Modelling*, **337**, 96–106.
- LIESKE, D.J., SCHMID, M.S. & MAHONEY, M. 2018. Ensembles of ensembles: combining the predictions from multiple machine learning methods. In *Machine Learning for Ecology and Sustainable Natural Resource Management*. Cham: Springer, 109–121.
- LIU, C., NEWELL, G. & WHITE, M. 2019. The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography*, **42**, 535–548.
- LOBO, J.M. 2008. More complex distribution models or more representative data? *Biodiversity Informatics*, **5**, 10.17161/bi.v5i0.40.
- LOBO, J.M., JIMÉNEZ-VALVERDE, A. & HORTAL, J. 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, **33**, 103–114.
- LOMBA, A., PELLISSIER, L., RANDIN, C., VICENTE, J., MOREIRA, F., HONRADO, J. & GUISAN, A. 2010. Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, **143**, 2647–2657.
- LOOTS, C., KOUUBI, P. & DUHAMEL, G. 2007. Habitat modelling of *Electrona antarctica* (Myctophidae, Pisces) in Kerguelen by generalized additive models and geographic information systems. *Polar Biology*, **30**, 951–959.
- LÓPEZ-FARRÁN, Z., GUILLAUMOT, C., VARGAS-CHACOFF, L., PASCHKE, K., DULIÈRE, V., DANIS, B., *et al.* 2021. Is the southern crab *Halicarcinus planatus* (Fabricius, 1775) the next invader of Antarctica? *Global Change Biology* 10.1111/gcb.15674.
- LORENA, A.C., JACINTHO, L.F., SIQUEIRA, M.F., DE GIOVANNI, R., LOHMANN, L.G., DE CARVALHO, A.C. & YAMAMOTO, M. 2011. Comparing machine learning classifiers in potential distribution modelling. *Expert Systems with Applications*, **38**, 5268–5275.
- LOZIER, J.D., ANIELLO, P. & HICKERSON, M.J. 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *Journal of Biogeography*, **36**, 1623–1627.
- LUEDELING, E., KINDT, R., HUTH, N.I. & KOENIG, K. 2014. Agroforestry systems in a changing climate - challenges in projecting future performance. *Current Opinion in Environmental Sustainability*, **6**, 1–7.
- LUOTO, M., PÖYRY, J., HEIKKINEN, R.K. & SAARINEN, K. 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography*, **14**, 575–584.
- MALDONADO, C., MOLINA, C.I., ZIZKA, A., PERSSON, C., TAYLOR, C.M., ALBÁN, J., *et al.* 2015. Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography*, **24**, 973–984.
- MANEL, S., WILLIAMS, H.C. & ORMEROD, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921–931.
- MARCEA, A., MÉNDEZ-VIGO, B., ALONSO-BLANCO, C. & PICÓ, F.X. 2016. Tackling intraspecific genetic structure in distribution models better reflects species geographical range. *Ecology and Evolution*, **6**, 2084–2097.
- MARMION, M., PARVIAINEN, M., LUOTO, M., HEIKKINEN, R.K. & THUILLER, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.
- MASSADA, A.B., SYPHARD, A.D., STEWART, S.I. & RADELOFF, V.C. 2013. Wildfire ignition-distribution modelling: a comparative study in the Huron-Manistee National Forest, Michigan, USA. *International Journal of Wildland Fire*, **22**, 174–183.
- MATHEWSON, P.D., MOYER-HORNER, L., BEEVER, E.A., BRISCOE, N.J., KEARNEY, M., YAHN, J.M. & PORTER, W.P. 2017. Mechanistic variables can enhance predictive models of endotherm distributions: the American pika under current, past, and future climates. *Global Change Biology*, **23**, 1048–1064.
- MCPHERSON, J.M., JETZ, W. & ROGERS, D.J. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811–823.
- MEARNS, L.O., HULME, M., CARTER, T.R., LEEMANS, R., LAL, M., WHETTON, P., *et al.* 2001. Climate scenario development. In *Climate change 2001: the science of climate change*. Cambridge: Cambridge University Press, 739–768.
- MEIER, E.S., EDWARDS T.C. JR, KIENAST, F., DOBBERTIN, M. & ZIMMERMANN, N.E. 2011. Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica* L. *Journal of Biogeography*, **38**, 371–382.
- MEROW, C., SMITH, M.J. & SILANDER, J.A. JR. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.
- MEROW, C., WILSON, A.M. & JETZ, W. 2017. Integrating occurrence data and expert maps for improved species range predictions. *Global Ecology and Biogeography*, **26**, 243–258.
- MEROW, C., SMITH, M.J., EDWARDS, T.C. JR, GUISAN, A., MCMAHON, S.M., NORMAND, S., *et al.* 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography*, **37**, 1267–1281.
- MESGARAN B.M., COUSENS R.D. & WEBBER B.L. 2014. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions*, **20**, 1147–1159.

- MEYNARD, C.N. & QUINN, J.F. 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**, 1455–1469.
- MITCHELL, P.J., MONK, J. & LAURENSEN, L. 2017. Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution*, **8**, 12–21.
- MOLLOY, S.W., DAVIS, R.A., DUNLOP, J.A. & VAN ETEN, E. 2017. Applying surrogate species presences to correct sample bias in species distribution models: a case study using the Pilbara population of the Northern Quoll. *Nature Conservation*, **18**, 27–46.
- MONK, J. 2014. How long should we ignore imperfect detection of species in the marine environment when modelling their distribution? *Fish and Fisheries*, **15**, 352–358.
- MOORE, J.M., CARVAJAL, J.I., ROUSE, G.W. & WILSON, N.G. 2018. The Antarctic Circumpolar Current isolates and connects: structured circumpolarity in the sea star *Glabraster antarctica*. *Ecology and Evolution*, **8**, 10621–10633.
- MORALES, N.S., FERNÁNDEZ, I.C. & BACA-GONZÁLEZ, V. 2017. MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. *PeerJ*, **5**, e3093.
- 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). *Journal of Biogeography*, **44**, 848–860.
- MOREAU, C., JOSSART, Q., DANIS, B., ELÉAUME, M., CHRISTIANSEN, H., GUILLAUMOT, C., *et al.* 2021. The overlooked diversity of Southern Ocean sea stars (Asteroidea) reveals original evolutionary pathways. *Progress in Oceanography*, **190**, 102472.
- MOREAU, C., MAH, C., AGÜERA, A., AMÉZIANE, N., BARNES, D., CROKAERT, G., *et al.* 2018. Antarctic and sub-Antarctic Asteroidea database. *ZooKeys*, **747**, 141–156.
- MORENO-AMAT, E., MATEO, R.G., NIETO-LUGILDE, D., MORUETA-HOLME, N., SVENNING, J.C. & GARCÍA-AMORENA, I. 2015. Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. *Ecological Modelling*, **312**, 308–317.
- MORMÈDE, S., IRISSON, J.O. & RAYMOND, B. 2014. Distribution modelling. In DE BROYER C., KOUUBI P., GRIFFITHS H.J., RAYMOND B., UDEKEM D'ACOS C.D', *et al.*, eds. *Biogeographic atlas of the Southern Ocean*. Cambridge: Scientific Committee on Antarctic Research, 27–29.
- MULCAHY, K.A. & CLARKE, K.C. 2001. Symbolization of map projection distortion: a review. *Cartography and Geographic Information Science*, **28**, 167–182.
- MURPHEY, P.C., GURALNICK, R.P., GLAUBITZ, R., NEUFELD, D. & RYAN, J.A. 2004. Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the Mountain and Plains Spatio-Temporal Database-Informatics Initiative (Mapstedi). *PhyloInformatics*, **3**, 1–29.
- MURRAY, R.J. 1996. Explicit generation of orthogonal grids for ocean models. *Journal of Computational Physics*, **126**, 251–273.
- MUSCARELLA, R., GALANTE, P.J., SOLEY-GUARDIA, M., BORJA, R.A., KASS, J.M., URIARTE, M. & ANDERSON, R.P. 2014. *ENM eval*: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, **5**, 1198–1205.
- NACHTSHEIM, D.A., 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 Biology*, **40**, 961–976.
- NAIMI, B., SKIDMORE, A.K., GROEN, T.A. & HAMM, N.A. 2011. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *Journal of Biogeography*, **38**, 1497–1509.
- NEWBOLD, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, **34**, 3–22.
- OCARANZA-BARRERA, P., GONZÁLEZ-WEVAR, C.A., GUILLEMIN, M.L., ROSENFELD, S. & MANSILLA, A. 2019. Molecular divergence between *Iridaea cordata* (Turner) Bory de Saint-Vincent from the Antarctic Peninsula and the Magellan Region. *Journal of Applied Phycology*, **31**, 939–949.
- OLDEN J.D., LAWLER J.J. & POFF N.L. 2008. Machine learning methods without tears: a primer for ecologists. *Quarterly Review of Biology*, **83**, 171–193.
- OSBORNE, P.E. & LEITÃO, P.J. 2009. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Diversity and Distributions*, **15**, 671–681.
- OWENS, H.L., CAMPBELL, L.P., DORNAK, L.L., SAUPE, E.E., BARVE, N., SOBERÓN, J., *et al.* 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, **263**, 10–18.
- PAGEL, J. & SCHURR, F.M. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, **21**, 293–304.
- PARDO-GANDARILLAS, M.C., IBÁÑEZ, C.M., TORRES, F.I., SANHUEZA, V., FABRES, A., ESCOBAR-DODERO, J., *et al.* 2018. Phylogeography and species distribution modelling reveal the effects of the Pleistocene ice ages on an intertidal limpet from the south-eastern Pacific. *Journal of Biogeography*, **45**, 1751–1767.
- PEARSON, R.G. 2007. Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, **50**, 54–89.
- PEARSON, R.G., RAXWORTHY, C.J., NAKAMURA, M. & TOWNSEND PETERSON, A. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.
- PEARSON, R.G., THULLER, W., ARAÚJO, M.B., MARTINEZ-MEYER, E., BROTONS, L., MCCLEAN, C., *et al.* 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.
- PEEL, S.L., HILL, N.A., FOSTER, S.D., WOTHERSPOON, S.J., GHIGLIONE, C. & SCHIAPARELLI, S. 2019. Reliable species distributions are obtainable with sparse, patchy and biased data by leveraging over species and data types. *Methods in Ecology and Evolution*, **10**, 1002–1014.
- PELLISSIER, L., ROHR, R.P., NDIRIBE, C., PRADERVAND, J.N., SALAMIN, N., GUIAN, A. & WISZ, M. 2013. Combining food web and species distribution models for improved community projections. *Ecology and Evolution*, **3**, 4572–4583.
- PELLISSIER, L., ANNE BRÄTHEN, K., POTTIER, J., RANDIN, C.F., VITTOZ, P., DUBUIS, A., *et al.* 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004–1014.
- PERRAULT-HÉBERT, M., GIRARD, F., BOUCHER, Y., FOURNIER, R., MANSUY, N. & VALERIA, O. 2019. *Evaluation of spatiotemporal transferability of wildfire probability across eastern boreal forest of North America*. PhD dissertation, University of Montréal.
- PETERSON, A.T., SOBERÓN, J., PEARSON, R.G., ANDERSON, R.P., MARTÍNEZ-MEYER, E., NAKAMURA, M. & ARAÚJO, M.B. 2011. *Ecological niches and geographic distributions (MPB-49)* (Vol. 56). Princeton, NJ: Princeton University Press, 328 pp.
- PETITPIERRE, B., BROENNIMANN, O., KUEFFER, C., DAEHLER, C. & GUIAN, A. 2017. Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, **26**, 275–287.
- PHILLIPS, N.D., REID, N., THYS, T., HARROD, C., PAYNE, N.L., MORGAN, C.A., *et al.* 2017. Applying species distribution modelling to a data poor, pelagic fish complex: the ocean sunfishes. *Journal of Biogeography*, **44**, 2176–2187.

- PHILLIPS, S.J., ANDERSON, R.P. & SCHAPIRE, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- PHILLIPS, S.J., DUDÍK, M., ELITH, J., GRAHAM, C.H., LEHMANN, A., LEATHWICK, J. & FERRIER, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.
- PIERRAT, B., SAUCÈDE, T., LAFFONT, R., DE RIDDER, C., FESTEATU, A. & DAVID, B. 2012. Large-scale distribution analysis of Antarctic echinoids using ecological niche modelling. *Marine Ecology Progress Series*, **463**, 215–230.
- PINKERTON, M.H., SMITH, A.N., RAYMOND, B., HOSIE, G.W., SHARP, B., LEATHWICK, J.R. & BRADFORD-GRIEVE, J.M. 2010. Spatial and seasonal distribution of adult *Oithona similis* in the Southern Ocean: predictions using boosted regression trees. *Deep-Sea Research I: Oceanographic Research Papers*, **57**, 469–485.
- PITTMAN, S.J., ed. 2017. *Seascape ecology*. Hoboken, NJ: John Wiley & Sons, 501 pp.
- POPE, A., WAGNER, P., JOHNSON, R., SHUTLER, J.D., BAESEMAN, J. & NEWMAN, L. 2017. Community review of Southern Ocean satellite data needs. *Antarctic Science*, **29**, 97–138.
- PORFIRIO, L.L., HARRIS, R.M., LEFROY, E.C., HUGH, S., GOULD, S.F., LEE, G., et al. 2014. Improving the use of species distribution models in conservation planning and management under climate change. *PLoS One*, **9**, e113749.
- QIAO, H., SOBERÓN, J. & PETERSON, A.T. 2015. No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution*, **6**, 1126–1136.
- RADOSAVLJEVIC, A. & ANDERSON, R.P. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, **41**, 629–643.
- RAES, N. 2012. Partial versus full species distribution models. *Natureza & Conservação*, **10**, 127–138.
- RAES, N. & TER STEEGE, H. 2007. A null-model for significance testing of presence-only species distribution models. *Ecography*, **30**, 727–736.
- RANC, N., SANTINI, L., RONDININI, C., BOITANI, L., POITEVIN, F., ANGERBJÖRN, A. & MAIORANO, L. 2017. Performance tradeoffs in target-group bias correction for species distribution models. *Ecography*, **40**, 1076–1087.
- RANDIN, C.F., DIRNBÖCK, T., DULLINGER, S., ZIMMERMANN, N.E., ZAPPA, M. & GUISAN, A. 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.
- RAYMOND, B., LEA, M.A., PATTERSON, T., ANDREWS-GOFF, V., SHARPLES, R., CHARRASSIN, J.B., et al. 2015. Important marine habitat off east Antarctica revealed by two decades of multi-species predator tracking. *Ecography*, **38**, 121–129.
- READY, J., KASCHNER, K., SOUTH, A.B., EASTWOOD, P.D., REES, T., RIUS, J., et al. 2010. Predicting the distributions of marine organisms at the global scale. *Ecological Modelling*, **221**, 467–478.
- REISS, H., CUNZE, S., KÖNIG, K., NEUMANN, H. & KRÖNCKE, I. 2011. Species distribution modelling of marine benthos: a North Sea case study. *Marine Ecology Progress Series*, **442**, 71–86.
- ROBERTS, D.R., BAHN, V., CIUTI, S., BOYCE, M.S., ELITH, J., GUILLERA-ARROITA, G., et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, **40**, 913–929.
- ROBINSON, L.M., ELITH, J., HOBDAI, A.J., PEARSON, R.G., KENDALL, B.E., POSSINGHAM, H.P. & RICHARDSON, A.J. 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, **20**, 789–802.
- ROBINSON, N.M., NELSON, W.A., COSTELLO, M.J., SUTHERLAND, J.E. & LUNDQUIST, C.J. 2017. A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Frontiers in Marine Science*, **4**, 421.
- ROCCHINI, D., HORTAL, J., LENGUEL, S., LOBO, J.M., JIMENEZ-VALVERDE, A., RICOTTA, C., et al. 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, **35**, 211–226.
- RODDA, G.H., JARNEVICH, C.S. & REED, R.N. 2011. Challenges in identifying sites climatically matched to the native ranges of animal invaders. *PLoS One*, **6**, e14670.
- RODRÍGUEZ, L., GARCÍA, J.J., CARREÑO, F. & MARTÍNEZ, B. 2019. Integration of physiological knowledge into hybrid species distribution modelling to improve forecast of distributional shifts of tropical corals. *Diversity and Distributions*, **25**, 715–728.
- ROBERT-COUDERT, Y., VAN DE PUTTE, A.P., REISINGER, R.R., BORNEMANN, H., CHARRASSIN, J.B., COSTA, D.P., et al. 2020. The retrospective analysis of Antarctic tracking data project. *Scientific Data*, **7**, 1–11.
- SÁNCHEZ-FERNÁNDEZ, D., LOBO, J.M. & HERNÁNDEZ-MANRIQUE, O.L. 2011. Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. *Diversity and Distributions*, **17**, 163–171.
- SCALES, K.L., MILLER, P.I., INGRAM, S.N., HAZEN, E.L., BOGRAD, S.J. & PHILLIPS, R.A. 2016. Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. *Diversity and Distributions*, **22**, 212–224.
- SCHAPIRE, R.E. 1990. The strength of weak learnability. *Machine Learning*, **5**, 197–227.
- SEGURADO, P., ARAUJO, M.B. & KUNIN, W.E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, **43**, 433–444.
- SEO, C., THORNE, J.H., HANNAH, L. & THUILLER, W. 2009. Scale effects in species distribution models: implications for conservation planning under climate change. *Biology Letters*, **5**, 39–43.
- SILLERO, N. 2011. What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling*, **222**, 1343–1346.
- SILLERO, N. & BARBOSA, A.M. 2021. Common mistakes in ecological niche models. *International Journal of Geographical Information Science*, **35**, 213–226.
- SMITH, A.B. 2013. On evaluating species distribution models with random background sites in place of absences when test presences disproportionately sample suitable habitat. *Diversity and Distributions*, **19**, 867–872.
- SNICKARS, M., GULLSTRÖM, M., SUNDBLAD, G., BERGSTRÖM, U., DOWNIE, A.L., LINDEGARTH, M. & MATTLA, J. 2014. Species-environment relationships and potential for distribution modelling in coastal waters. *Journal of Sea Research*, **85**, 116–125.
- STAVELEY, T.A., PERRY, D., LINDBORG, R. & GULLSTRÖM, M. 2017. Seascape structure and complexity influence temperate seagrass fish assemblage composition. *Ecography*, **40**, 936–946.
- STØA, B., HALVORSEN, R., MAZZONI, S. & GUSAROV, V.I. 2018. Sampling bias in presence-only data used for species distribution modelling: theory and methods for detecting sample bias and its effects on models. *Sommerfeltia*, **38**, 1–53.
- STOCK, A., SUBRAMANIAM, A., VAN DIJKEN, G.L., WEDDING, L.M., ARRIGO, K.R., MILLS, M.M., et al. 2020. Comparison of cloud-filling algorithms for marine satellite data. *Remote Sensing*, **12**, 3313.
- STOCKWELL, D.R. & PETERSON, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1–13.
- SYFERT, M.M., SMITH, M.J. & COOMES, D.A. 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One*, **8**, e55158.
- TALLUTO, M.V., BOULANGEAT, I., AMEZTEGUI, A., AUBIN, I., BERTEAUX, D., BUTLER, A., et al. 2016. Cross-scale integration of knowledge for

- predicting species ranges: a metamodeling framework. *Global Ecology and Biogeography*, **25**, 238–249.
- TELFORD, R.J. & BIRKS, H.J. 2009. Evaluation of transfer functions in spatially structured environments. *Quaternary Science Reviews*, **28**, 1309–1316.
- TESSAROLO, G., LADLE, R., RANGEL, T. & HORTAL, J. 2017. Temporal degradation of data limits biodiversity research. *Ecology and Evolution*, **7**, 6863–6870.
- THATJE, S. 2012. Effects of capability for dispersal on the evolution of diversity in Antarctic benthos. *Integrative and Comparative Ecology*, **52**, 470–482.
- THIERS, L., DELORD, K., BOST, C.-A., GUINET, C. & WEIMERSKIRCH, H. 2017. Important marine sectors for the top predator community around Kerguelen Archipelago. *Polar Biology*, **40**, 365–378.
- THULLER, W., BROTONS, L., ARAÚJO, M.B. & LAVOREL, S. 2004. Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165–172.
- THULLER, W., LAFOURCADE, B., ENGLER, R. & ARAÚJO, M.B. 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.
- THULLER, W., VAYREDA, J., PINO, J., SABATE, S., LAVOREL, S. & GRACIA, C. 2003. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology and Biogeography*, **12**, 313–325.
- THYRRING, J., BUNDGAARD, A. & SEJR, M.K. 2017. Seasonal acclimation and latitudinal adaptation are of the same magnitude in *Mytilus edulis* and *Mytilus trossulus* mitochondrial respiration. *Polar Biology*, **40**, 1885–1891.
- TIRUNELVELI, G., GORDON, R. & PISTORIUS, S. 2002. Comparison of square-pixel and hexagonal-pixel resolution in image processing. In *IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No. 02CH37373)*, Vol. **2**. Piscataway, NJ: IEEE, 867–872.
- TITEUX, N., MAES, D., VAN DAELE, T., ONKELINX, T., HEIKKINEN, R.K., ROMO, H., *et al.* 2017. The need for large-scale distribution data to estimate regional changes in species richness under future climate change. *Diversity and Distributions*, **23**, 1393–1407.
- TROLLE, D., ELLIOTT, J.A., MOOIJ, W.M., JANSE, J.H., BOLDING, K., HAMILTON, D.P. & JEPPESEN, E. 2014. Advancing projections of phytoplankton responses to climate change through ensemble modelling. *Environmental Modelling & Software*, **61**, 371–379.
- TRULL, T.W., PASSMORE, A., DAVIES, D.M., SMIT, T., BERRY, K. & TILBROOK, B. 2018. Distribution of planktonic biogenic carbonate organisms in the Southern Ocean south of Australia: a baseline for ocean acidification impact assessment. *Biogeosciences*, **15**, 31.
- TSOAR, A., ALLOUCHE, O., STEINITZ, O., ROTEM, D. & KADMON, R. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**, 397–405.
- TURNER, J., BARRAND, N., BRACEGIRDLE, T., CONVEY, P., HODGSON, D.A., JARVIS, M., *et al.* 2014. Antarctic climate change and the environment: an update. *Polar Record*, **50**, 237–259.
- VALAVI, R., ELITH, J., LAHOZ-MONFORT, J.J. & GUILLERA-ARROITA, G. 2019. *blockCV*: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, **10**, 225–232.
- VALE, C.G., TARROSO, P. & BRITO, J.C. 2014. Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara-Sahel transition zone. *Diversity and Distributions*, **20**, 20–33.
- VAN PROOSDIJ, A.S., SOSEF, M.S., WIERINGA, J.J. & RAES, N. 2016. Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, **39**, 542–552.
- VANDEN BERGHE, E. 2013. *Report of the project 'Turning OBIS data into information': project funded by the Census of Marine Life International Cosmos Prize Fund*. New Brunswick, NJ: Rutgers University Press, 62 pp.
- VANDERSTEEN, W. 2011. Detecting gene expression profiles associated with environmental stressors within an ecological context. *Molecular Ecology*, **20**, 1322–1323.
- VAPNIK, V. 1998. *Statistical learning theory*. New York: Wiley, 624 pp.
- WARREN, D.L. & SEIFERT, S.N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335–342.
- WATLING, J.I., BRANDT, L.A., BUCKLIN, D.N., FUJISAKI, I., MAZZOTTI, F.J., ROMAÑACH, S.S. & SPEROTERRA, C. 2015. Performance metrics and variance partitioning reveal sources of uncertainty in species distribution models. *Ecological Modelling*, **309**, 48–59.
- WENGER, S.J. & OLDEN, J.D. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, **3**, 260–267.
- WHITE-NEWSOME, J.L., BRINES, S.J., BROWN, D.G., DVONCH, J.T., GRONLUND, C.J., ZHANG, K., *et al.* 2013. Validating satellite-derived land surface temperature with in situ measurements: a public health perspective. *Environmental Health Perspectives*, **121**, 925–931.
- WHITTINGHAM, M.J., STEPHENS, P.A., BRADBURY, R.B. & FRECKLETON, R.P. 2006. Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, **75**, 1182–1189.
- WILEY, E.O., MCNYSET, K.M., PETERSON, A.T., ROBINS, C.R. & STEWART, A.M. 2003. Niche modeling perspective on geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography*, **16**, 102–127.
- WILLIAMS, J.W. & JACKSON, S.T. 2007. Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**, 475–482.
- WILLIAMS, J.W., JACKSON, S.T. & KUTZBACH, J.E. 2007. Projected distributions of novel and disappearing climates by 2100 ad. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5738–5742.
- WISZ, M.S. & GUISAN, A. 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecology*, **9**, 8.
- WISZ, M.S., HIMMANS, R.J., LI, J., PETERSON, A.T., GRAHAM, C.H., GUISAN, A. & NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.
- XAVIER, J.C., RAYMOND, B., JONES, D.C. & GRIFFITHS, H. 2016. Biogeography of Cephalopods in the Southern Ocean using habitat suitability prediction models. *Ecosystems*, **19**, 220–247.
- YACKULIC, C.B., CHANDLER, R., ZIPKIN, E.F., ROYLE, J.A., NICHOLS, J.D., CAMPBELL GRANT, E.H. & VERAN, S. 2013. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, **4**, 236–243.
- ZANIEWSKI, A.E., LEHMANN, A. & OVERTON, J.M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**, 261–280.
- ZHANG, X., & MAHADEVAN, S. 2019. Ensemble machine learning models for aviation incident risk prediction. *Decision Support Systems*, **116**, 48–63.
- ZHOU, Z.H. 2012. *Ensemble methods: foundations and algorithms*. London: Chapman and Hall/CRC Press, 236 pp.
- ZHU, G.P. & PETERSON, A.T. 2017. Do consensus models outperform individual models? Transferability evaluations of diverse modeling approaches for an invasive moth. *Biological Invasions*, **19**, 2519–2532.
- ZURELL, D., FRANKLIN, J., KÖNIG, C., BOUCHET, P.J., DORMANN, C.F., ELITH, J., *et al.* 2020. A standard protocol for reporting species distribution models. *Ecography*, **43**, 1261–1277.