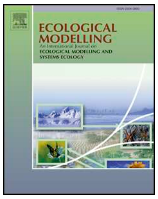
Ecological Modelling 416 (2020) 108902

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel



Modelling European small pelagic fish distribution: Methodological insights

Alexandre Schickele[a,*], Boris Leroy[b], Gregory Beaugrand[c,d], Eric Goberville[b], Tarek Hattab[e], Patrice Francour[a,1], Virginie Raybaud[a,1]

[a] Université Côte d'Azur, CNRS, UMR 7035 ECOSEAS, Nice, France
[b] Unité Biologie des Organismes et Ecosystèmes Aquatiques (BOREA), Muséum National d'Histoire Naturelle, Sorbonne Université, Université de Caen Normandie, Université des Antilles, CNRS, IRD, Paris, France
[c] CNRS, Univ. Lille, Univ. Littoral Côte d'Opale, UMR 8187, LOG, Laboratoire d'Océanologie et de Géosciences, Wimereux, France
[d] Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, Plymouth, United Kingdom
[e] MARBEC, Univ. Montpellier, CNRS, Ifremer, IRD, Avenue Jean Monnet, Sète, France

ARTICLE INFO

Keywords:
Species distribution models
Pseudo-absence
Sampling bias
Convex hull
Uncertainty
Small pelagic fish

ABSTRACT

The distribution of marine organisms is strongly influenced by climatic gradients worldwide. The ecological niche (*sensu* Hutchinson) of a species, *i.e.* the combination of environmental tolerances and resources required by an organism, interacts with the environment to determine its geographical range. This duality between niche and distribution allows climate change biologists to model potential species' distributions from past to future conditions. While species distribution models (SDMs) have been intensively used over the last years, no consensual framework to parametrise, calibrate and evaluate models has emerged. Here, to model the contemporary (1990–2017) spatial distribution of seven highly harvested European small pelagic fish species, we implemented a comprehensive and replicable numerical procedure based on 8 SDMs (7 from the Biomod2 framework plus the NPPEN model). This procedure considers critical issues in species distribution modelling such as sampling bias, pseudo-absence selection, model evaluation and uncertainty quantification respectively through (i) an environmental filtration of observation data, (ii) a convex hull based pseudo-absence selection, (iii) a multi-criteria evaluation of model outputs and (iv) an ensemble modelling approach. By mitigating environmental sampling bias in observation data and by identifying the most ecologically relevant predictors, our framework helps to improve the modelling of fish species' environmental suitability. Not only average temperature, but also temperature variability appears as major factors driving small pelagic fish distribution, and areas of highest environmental suitability were found along the north-western Mediterranean coasts, the Bay of Biscay and the North Sea. We demonstrate in this study that the use of appropriate data pre-processing techniques, an often-overlooked step in modelling, increase model predictive performance, strengthening our confidence in the reliability of predictions.

1. Introduction

Fish species distribution and assemblages are strongly influenced by both climatic and physical gradients (Ben Rais Lasram et al., 2010; Beaugrand et al., 2011; Raybaud et al., 2017). Temperature is known as a master parameter driving fish distribution at a macroecological level (Lenoir et al., 2011; Beaugrand et al., 2018). This parameter influences a large range of biological processes such as growth, reproduction, larval development, recruitment, and act as a major stressing factor depending on species thermal tolerance (psychrophile or thermophile species; Angilletta, 2011; Beaugrand and Kirby, 2018). Salinity, oxygen concentration, primary production (that are indirectly influenced by changes in temperature; *e.g.* Kirby and Beaugrand, 2009) or the physical habitat (*e.g.* sediment type; Poloczanska et al., 2013) may also highly influence marine fish species at different spatial scales.

Hutchinson (1957) conceptualised the ecological niche as the "n-dimensional ensemble of environmental conditions that enable a species to live and reproduce" and subsequently made a distinction between the fundamental and the realised niche (Hutchinson, 1978). Due to biotic interactions, dispersal limitation and/or historical factors

---

* Corresponding author.

E-mail addresses: alex.schickele@gmail.com, alexandre.schickele@univ-cotedazur.fr (A. Schickele), boris.leroy@mnhn.fr (B. Leroy), gregory.beaugrand@univ-lille1.fr (G. Beaugrand), eric.goberville@upmc.fr (E. Goberville), tarek.hattab@ifremer.fr (T. Hattab), patrice.francour@univ-cotedazur.fr (P. Francour), virginie.raybaud@univ-cotedazur.fr (V. Raybaud).

[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.ecolmodel.2019.108902
Received 11 July 2019; Received in revised form 19 November 2019; Accepted 23 November 2019

(Soberon and Peterson, 2005), species generally occupy only their realised niche, *i.e.* the subset of their fundamental niche that represents the response of all physiological processes of a species to the synergistic effects of environmental conditions (Helaouet and Beaugrand, 2009; Beaugrand et al., 2013). By defining the niche as an attribute of species instead of a portion of the environment, the Hutchinson's concept enables duality between niche and distribution (Pulliam, 2000; Colwell and Rangel, 2009). Such a relationship is of major interest in biogeography as each georeferenced species occurrence, *i.e.* where a given species has been observed, can be related to several environmental parameters such as temperature, salinity and primary production. When species are in equilibrium with their environment, associating environmental conditions and observed distributions permits climate change biologists to estimate species' potential niche (Jiménez-Valverde et al., 2008).

The relationship between species occurrences, environmental conditions and species' potential niche has become intensively studied over the last two decades, using a wide range of modelling techniques - hereafter referred to as Species Distribution Models (SDMs) to assess past, contemporary and future species distribution in both marine and terrestrial ecosystems (*e.g.* Cheung et al., 2009; Bellard et al., 2016; Cristofari et al., 2018). SDMs rely on several ecological assumptions, such as species distribution in equilibrium or habitat saturation (Soberon and Peterson, 2005), niche conservatism (Crisp et al., 2009), unlimited dispersal abilities (Wiens et al., 2009) or the non-influential role of biotic interactions in shaping large-scale distributions (*i.e.* the Gleasonian vision of biotic communities; Gleason, 1926; Guisan and Thuiller, 2005; Wiens et al., 2009). Superimposed to these assumptions, several sources of errors and uncertainties may lead to variation – sometimes conflicting – in the outputs of SDMs for a given species (Beaumont et al., 2008): (i) accuracy of observation data and (ii) lack of true absences (Proosdij et al., 2016), (iii) identification of ecologically meaningful environmental predictors with high explanatory power (Guisan and Thuiller, 2005), (iv) choice of the modelling algorithm (Buisson et al., 2010) and (v) SDMs' evaluation processes (Leroy et al., 2018). While tremendous progresses have been made on both the building and evaluation of SDMs in recent years with a plethora of new methods for modelling species' distribution (Araújo and Guisan, 2006; Leroy et al., 2018; Støa et al., 2018), the development of further procedures is still required for improving the quality of SDMs.

Species distribution models are known to be very sensitive to different sources of uncertainties and sustained attention should be devoted to each step of the modelling procedure, from the pre-processing of species occurrences data to model evaluation. Such an approach is essential to increase confidence in model outputs (Porfirio et al., 2014): for most areas of the world and species, survey effort often exhibits strong spatial and temporal bias, occurrence records being frequently too scarce, constrained to presence-only data or both. Working with biased observation datasets may result in under- or over-estimated species distributional ranges (Araújo and Guisan, 2006; Dormann et al., 2007), leading therefore to inaccurate modelled contemporary distributions, which are inadequate for assessing potential future range shifts or for defining conservation measures. Similarly, biased pseudo-absence datasets (*e.g.* multiple pseudo-absences selected in the same environmental conditions or coinciding with environmental conditions where the species is observed) may lead to a distorted estimation of species distributional ranges (*e.g.* Wisz and Guisan, 2009; Lobo and Tognelli, 2011). A modelling framework that includes a preliminary stage devoted to the construction of a representative calibration dataset – as well as its associated level of uncertainty assessment – is therefore essential (*e.g.* Varela et al., 2014).

Here, we developed a framework that encompasses recent advances on the building, calibration and evaluation of SDMs with the aim of (i) selecting the most relevant environmental parameters, (ii) generating consistent pseudo-absence data and (iii) validating representative model outputs (Cornwell et al., 2004; Varela et al., 2014; Leroy et al., 2018).

We applied this framework on seven economically important European Small Pelagic Fish (SPF) species (Mediterranean horse mackerel *Trachurus mediterraneus*, Atlantic horse mackerel *Trachurus trachurus*, European pilchard *Sardina pilchardus*, round sardinella *Sardinella aurita*, European sprat *Sprattus sprattus*, European Anchovy *Engraulis encrasicolus* and bogue *Boops boops*). These seven SPF species are widely distributed planktonic feeders known for their central role in marine food webs (Cury, 2000; Checkley et al., 2009). Moreover, they are of major economic importance and represent a large part of the Mediterranean and Black Sea commercial landings (more than 50 % between 2000 and 2013; FAO, 2016). However, while SPFs are ideal candidates for SDMs because of their sensitivity to environmental factors (Perry et al., 2005), their European distribution is far from being exhaustively documented and available records originated from diverse and/or non-standardised monitoring surveys (FAO, 2016).

2. Material and methods

2.1. Biological and environmental data

2.1.1. Small pelagic fish occurrence data

Occurrence records (*e.g.* fisheries independent trawl surveys, discrete research samplings) for the seven SPF species (Mediterranean horse mackerel, Atlantic horse mackerel, European pilchard, round sardinella, European sprat, European Anchovy and Bogue) were compiled from three available public databases: the Ocean Biogeographic Information System Mapper (OBIS, http://www.iobis.org/mapper/), the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/) and Fishbase (http://www.fishbase.org/). When possible, we included observations retrieved from the literature to construct the most up-to-date datasets encompassing their entire distribution range (see Supplementary material 1). Biological data retrieved for our study ranged from 1950 to 2017, recent records (since 1990) prevailing (83.2 ± 6.7 %) over both past (1950–1990; 12.2 ± 8.7 %) and undated observations (4.6 ± 3.6 %). Past or undated records were only considered along the distribution edge when the species presence was confirmed by recent records. This precautionary approach avoided over- or under-predictions of the model due to low quality presence data (Kramer-Schadt et al., 2013). The observation records pre-processing consisted in a data cleaning procedure applied on each species dataset to (i) remove unreliable observations (*e.g.* preserved specimen; Newbold, 2010) and false identifications (*e.g.* taxonomic confusion), (ii) discard duplicate occurrences and (iii) ensure the temporal and locational reliability at the edge of the observed distribution (*e.g.* data on land, longitudinal and/or latitudinal inversion, historical or undated data). According to the ecology of SPFs – species cannot be observed below 300 m depth (Checkley et al., 2009) – while remaining permissive, a precautionary bathymetry threshold (−1000 m) was applied to remove inconsistent occurrences. Following this pre-processing, we obtained seven clean datasets, with a number of observations ranging from 1314 (for Mediterranean horse mackerel) to 24,806 (for European sprat). For the seven SPFs, occurrences were aggregated on a 0.1° × 0.1° spatial grid (from 70 °N to 70 °S and 180 °E to 180 °W) that corresponds to that of environmental parameters.

2.1.2. Environmental data

To calculate the ecological niche (*sensu* Hutchinson, 1957) of each SPF, we collected environmental parameters from different databases (see Table 1 for details). Environmental parameters values for each spatial grid cell were first calculated for each year and then averaged on the 1990–2017 period. The environmental parameters presented in Table 1 were retrieved in different spatial resolutions ranging from 0.1° to 0.5°. For modelling purpose, all variables were therefore interpolated to a 0.1° × 0.1° grid using a bilinear interpolation in the geographical domain available for all environmental parameters, ranging from 70 °N

Table 1
Environmental parameters used to model SPF distribution.

Name (Period)	Description	Reference
Bathymetry	Spatial seafloor depth (m)	Global seafloor topography (Smith and Sandwell, 1997)
Distance to coast	Distance to the nearest coast (km)	NASA Goddard Space Flight Center (2009) (https://oceancolor.gsfc.nasa.gov/docs/distfromcoast/)
SSS (1990–2017)	Sea Surface Salinity	Levitus' climatology (Levitus, 2011) completed with (http://www.ices.dk/)
PP (1990–2017)	Sea Surface Primary Production (mol. $m^{-2}.s^{-1}$). Averaged from five general circulation models (IPSL, MPI, CNRM, HadGEM and GISS).	IPSL (Dufresne et al., 2013; Hourdin et al., 2013), MPI (Stevens et al., 2013; Giorgetta et al., 2013), CNRM (Voldoire et al., 2013), HadGEM (Jones et al., 2011) and GISS (Schmidt et al., 2014) models.
Log_PP (1990–2017)	Log10-transformed Sea Surface Primary Production	
SST (1990–2017)	Mean annual Sea Surface Temperature (°C)	AVHRR Very High Resolution Radiometer (Casey et al., 2010)
SSTmax (1990–2017)	Mean sea surface temperature of the hottest month (°C)	
SSTmin (1990–2017)	Mean sea surface temperature of the coldest month (°C)	
SSTr (1990–2017)	Mean annual sea surface temperature range (°C). Difference between SSTmax and SSTmin.	
SSTvar (1990–2017)	Mean monthly sea surface temperature variance (°C). Calculated using monthly SST data.	

to 70 °S and 180 °E to 180 °W.

2.2. Description of the models

We used two approaches to model the potential environmental suitability (*i.e.* spatialised index between 0 and 1, defined as a probability of presence based on environmental parameters) of each SPF species over the 1990–2017 period: (i) the Non-Parametric Probabilistic Ecological Niche (NPPEN; Beaugrand et al., 2011) model and (ii) seven modelling algorithms available within the BIOMOD2 package (Thuiller et al., 2016). The NPPEN model is a presence only model based on the Mahalanobis generalised distance (Mahalanobis, 1936) and on a modified version of the Multiple Response Permutation Procedure (MRPP; Mielke et al., 1981). The BIOMOD2 framework allows ensemble modelling of species distribution (*i.e.* an average model of a wide range of algorithms; Thuiller et al., 2009). Here, seven algorithms were considered: (i) Generalised Linear Model (GLM), (ii) Generalised Additive Model (GAM), (iii) Generalised Boosting Model (GBM), (iv) Artificial Neural Network (ANN), (v) Flexible Discriminant Analysis (FDA), (vi) Multiple Adaptive Regression Splines (MARS) and (vii) Random Forest (RF). Because the models used in this study have been already described and discussed in several publications (*e.g.* Beaugrand et al., 2011; Lenoir et al., 2011; Raybaud et al., 2015 for NPPEN, *e.g.* Thuiller et al., 2009; Albouy et al., 2012; Bellard et al., 2013 for BIOMOD2), we refer the reader to this literature for further information. The algorithms were calibrated using the default parameters in BIOMOD2, optimised for species distribution modelling (details in Thuiller et al., 2016). By including this large range of algorithms within an ensemble model approach, we quantified the uncertainty related to the selection of SDMs (Pearson et al., 2006; Buisson et al., 2010) by calculating the standard deviation (SD) and the coefficient of variation (CV) among SDM outputs.

2.3. Data preparation and ensemble model selection

2.3.1. Pre-selection of the environmental parameters and assessment of multicollinearity

To model the ecological niche of the seven SPFs, we first constructed the full dataset of environmental parameters based on our knowledge of the ecology of SPFs. A variable selection process (Fig. 1, step 1) was then applied to identify, at the species level, the most parsimonious dataset that explained each species distribution. This process follows the procedure described in Leroy et al. (2014) and Bellard et al. (2016). Because most of the algorithms (especially regression-based models such as GLM) are sensitive to multicollinearity – that may distort model estimation (Dormann et al., 2013) – relations among environmental parameters were assessed by means of the Pearson correlation coefficient, using a threshold $r > 0.7$ to reduce the initial environmental matrix. When two or more environmental parameters showed correlation values above this threshold, only one variable was retained (details in Supplementary material 2).

We subsequently assessed the relative importance of each environmental parameter by sequentially randomising each variable and by calculating the resulting current distribution (Leroy et al., 2014). The variables that best predicted SPF distribution were sea surface temperature annual mean (SST), temperature variability (sea surface temperature annual range or monthly variance, depending on the targeted species), bathymetry and distance to coast (see Supplementary material 2). In order to avoid model over-parameterisation (that affects model performance, model transferability and assessment of variable importance), we chose not to include bathymetry and distance to coast directly in the models, but in a hierarchical filtering approach (Hattab et al., 2014): for a given geographical cell, environmental conditions were considered as suitable for a marine species only if a probability of occurrence coincided with a distance to coast less than 50 km or up to a 300 m depth for oceanic cells, i.e. outside the 50 km wide coastal area. Concerning environmental predictors, we systematically considered temperature (mean and variability) in our models. Finally, we tested the relevance of including sea surface salinity (SSS) and/or primary production (log_PP) as a potential third explanatory environmental parameter in the models. Each run is detailed in Supplementary material 3.

2.3.2. Environmental filtration and pseudo-absence selection

Because sampling effort is neither homogeneous and nor standardised over marine regions, occurrence data may not be representative of the whole populations, a requirement to increase the reliability of SDMs (Lobo and Tognelli, 2011). While under-sampling is commonly observed at the edge of species range (Varela et al., 2014), observation datasets can also be biased toward regions more comprehensively investigated due to an easy access or a long tradition of monitoring (Fithian et al., 2015).

To consider the risk of over-sampling, and the ensuing over-representation of environmental features (Kramer-Schadt et al., 2013), we first homogenised species datasets to assign the same weight to over- and under-sampled regions (Fig. 1, step 2). A multidimensional matrix was designed for each species and each combination of environmental parameters, a dimension reflecting an environmental factor. Each cell of the homogenised matrix was considered as an environmental stratum, *i.e.* a combination of a set of parameters, with the following resolution: 0.5 °C for temperature-related parameters, 0.5 for SSS and 0.5 $mol.m^{-2}.s^{-1}$ (in log) for primary production. In case an environmental stratum contained multiple occurrences, only one occurrence (*i.e.* one 0.1° × 0.1° geographical cell with the corresponding

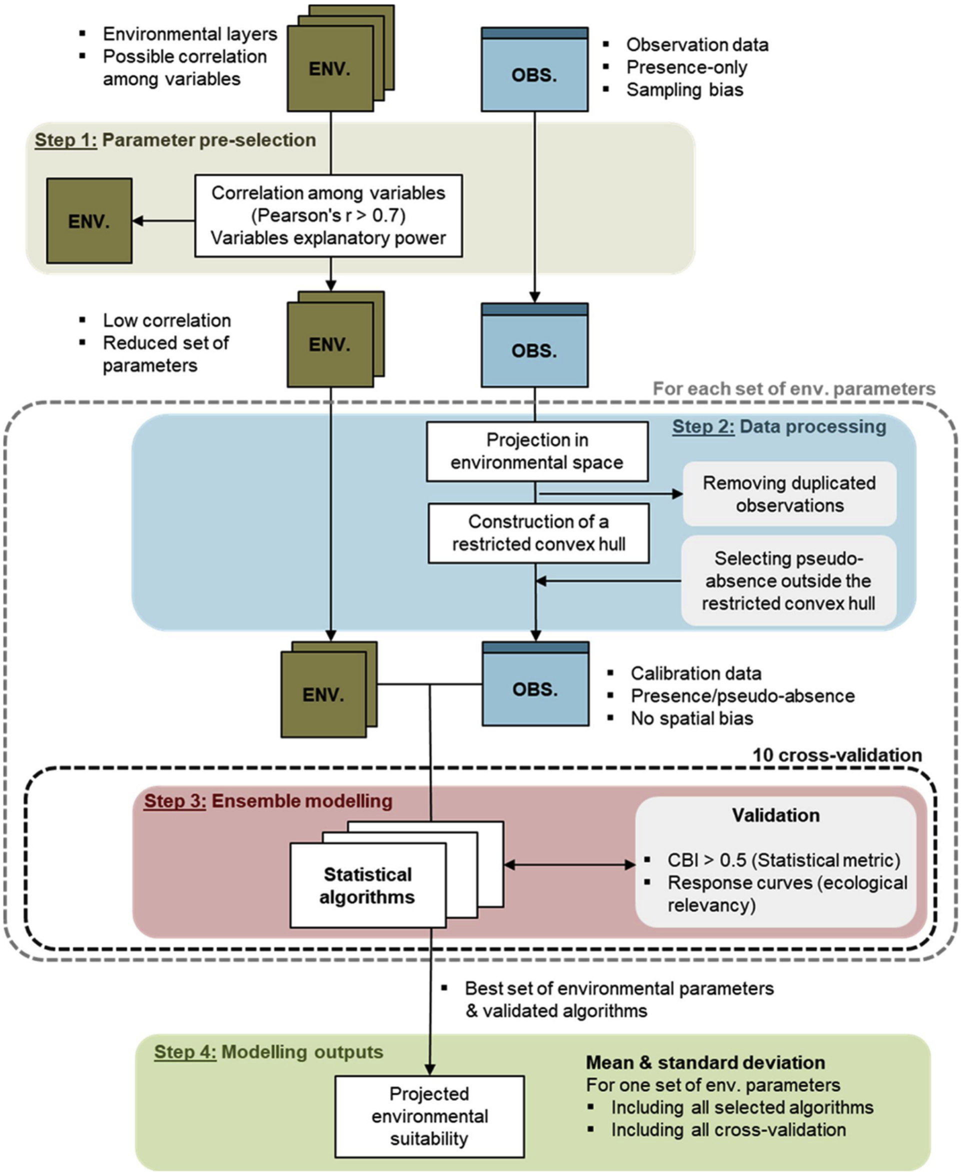


Fig. 1. Sketch diagram of the modelling framework applied to model SPFs species. “ENV.” = environmental parameters and “OBS.” = georeferenced presence data.

environmental conditions) was kept in the homogenised dataset.

We also considered the lack of absence data. To assess this gap, we generated pseudo-absences using the convex hull method (Cornwell et al., 2004; Getz and Wilmers, 2006). The convex hull was defined here as the smallest convex hyper-volume in the environmental space containing all species observation records. A restricted convex hull (see Fig. 2) has been defined as a convex hull excluding occurrence points within the 2.5 % and 97.5 % percentiles for each environmental parameter (*i.e.* excluding observations in the most extreme environmental conditions). This restricted convex hull is considered as a proxy of the suitable environmental conditions outside which, pseudo-absences were randomly generated in equal number to the filtered occurrences as recommended by the “D-designs”theory (Montgomery, 2005): the optimal design to minimise prediction variance is when an equal number of observations are at opposite value extremes (Montgomery, 2005; Hengl et al., 2009) and when there is a high spreading in the feature space. Finally, for each species, pseudo-absence were projected back in geographical cells showing environmental conditions outside SPF species' environmentally favourable areas (Fig. 2; Varela et al., 2014). Finally, model outputs obtained from our environmental filtration approach were compared with outputs for which neither environmental filtration nor the convex hull pseudo-absence selection method was applied (Fig. 3).

2.3.3. *Validation and selection of the best models*

We then quantified the performance of our models using five commonly used evaluation metrics: (i) the Continuous Boyce Index (CBI; Hirzel et al., 2006), a metric specifically designed for presence-only models and insensitive to pseudo-absences, (ii) the Area Under the Curve (AUC; Swets, 1988; Fielding and Bell, 1997), (iii) the True Skill Statistic (TSS; Allouche et al., 2006), (iv) the Jaccard and (v) the Sørensen similarity indices (Jaccard, 1908; Sørensen, 1948). However, because all evaluation metrics – except the CBI – require both presence and absence data (see discussion in Leroy et al., 2018 about the use of pseudo-absence to evaluate the performance of models) and because some may be affected by prevalence (*i.e.* the ratio between the number of observed presence and generated pseudo-absence; Leroy et al., 2018) we based our selection process of the best models on CBI values only. We considered models to be wrong when CBI values were below -0.5, “average to random” for values ranging from -0.5 to 0.5, and good for values above 0.5 (Faillettaz et al., 2019).

For each model, we computed evaluation metrics by performing a

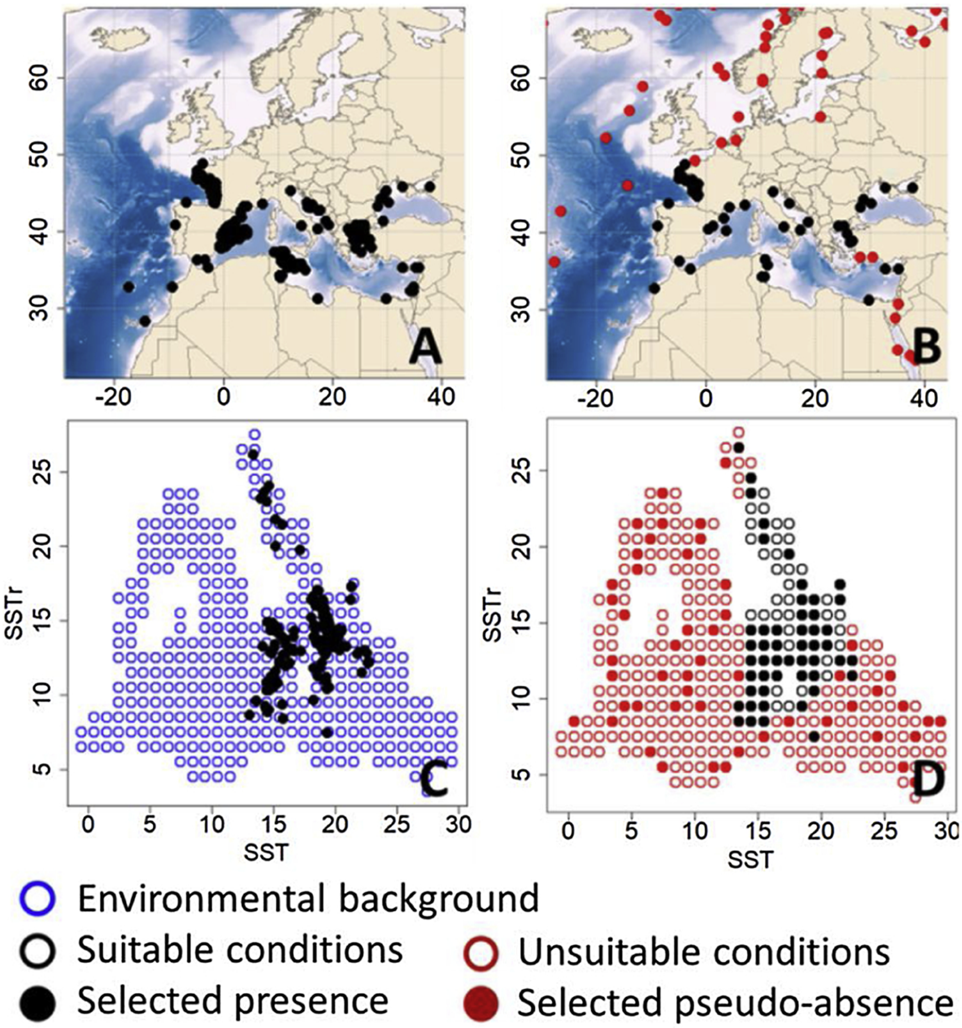


Fig. 2. Example of pseudo-absences generation for the Mediterranean horse mackerel (environmental parameters = SST + SSTr, 1 °C resolution). **A-C:** Species occurrences (black dots) in (**A**) the geographical domain and (**C**) the environmental space. **B-D:** Species occurrences (black dots) and pseudo-absences (red dots) generated from the restricted convex hull method in (**B**) the geographical domain and (**D**) the environmental space (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

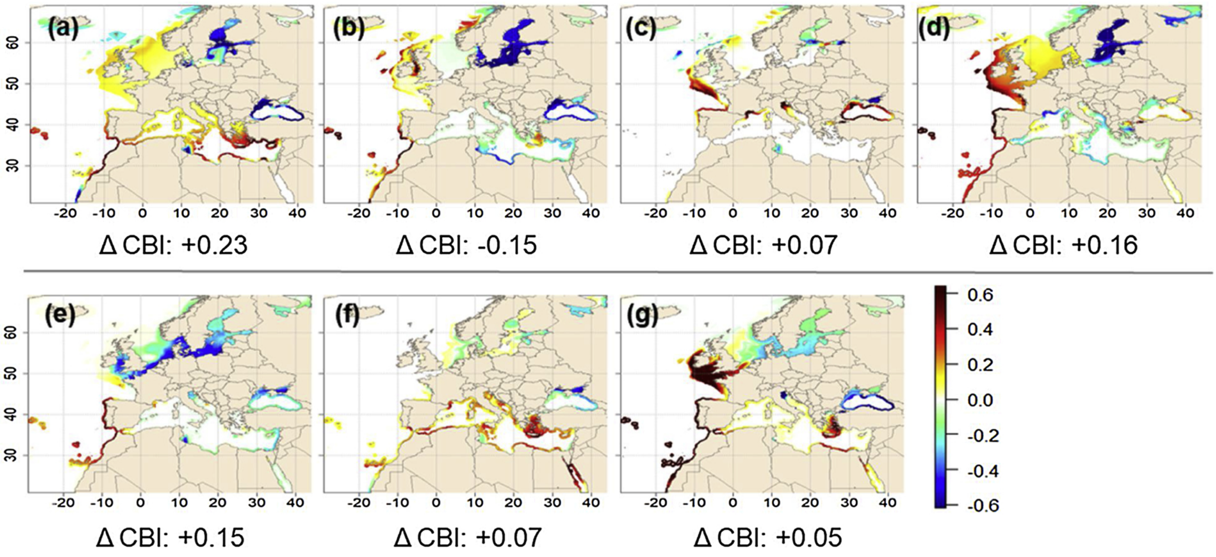


Fig. 3. Environmental suitability index and CBI differences between ensemble models originating from our modelling framework and ensemble models constructed without data filtration and random pseudo-absence selection for (a) Atlantic horse mackerel, (b) European pilchard, (c) European sprat, (d) European anchovy, (e) Mediterranean horse mackerel, (f) round sardinella and (g) bogue.

cross-validation procedure with 10 repetitions. We randomly sampled 70 % of the occurrence data to calibrate the model and kept the remaining 30 % for model validation (Merow et al., 2013). Following the "evaluation strip method" detailed by Elith et al. (2005), the adequacy between observed and modelled spatial distributions was also assessed by means of response curves. For a given environmental parameter, the corresponding response curve was calculated, while keeping the other parameters constant (*i.e.* at the mean value corresponding to their occurrence points). By doing this, we identified spurious results (*e.g.* we do not expect bimodal responses to temperature) and/or unexpected distribution ranges (*e.g.* large portions of predicted range in regions where the species has never been observed, and vice-versa; Supplementary material 4).

3. Results

3.1. SDMs and parameters selected in the ensemble models

Based on the calculation of the CBI values and the examination of species response curves (Supplementary material 3 and 4), we identified the best models for each SPF species. Our results showed that both GLM and NPPEN models were almost always selected in the ensemble model, except for the European anchovy.

Ensemble models showed that temperature-related variables were essential to assess the spatial distribution of SPFs'. For virtually all species, the models that considered mean temperature and variability showed high ability to reproduce the overall SPFs distributions (Table 2, Supplementary material 3) with CBI values always above 0.5 (Failletaz et al., 2019). However, some discrepancies were observed among species. While Mediterranean horse mackerel, Atlantic horse mackerel and European anchovy distributions were more related to mean monthly temperature variance (SSTvar), European pilchard, round sardinella, European Sprat and bogue distributions were better reproduced when mean annual temperature range (SSTr) was considered. Despite the high correlation between SSTr and SSTvar (r = 0.80, Supplementary material 2), both variables have dissimilar ecological influences (seasonality *versus* short-term climatic variability respectively). Primary production also emerged as important to model species' spatial distribution. Finally, we highlighted the important role of sea surface salinity (SSS) for both European pilchard and European anchovy, by discriminating both the Baltic and the Black seas from other regions (Table 2).

Table 2
Environmental parameters and SDMs selected by our procedure.

Mediterranean horse mackerel	Parameters:	SST, SSTvar, log_PP
	Models:	GLM, RF, NPPEN
	CBI (mean):	**0.71**
Atlantic horse mackerel	Parameters:	SST, SSTvar, log_PP
	Models:	GLM, NPPEN
	CBI (mean):	**0.95**
European pilchard	Parameters:	SST, SSTr, SSS
	Models:	GLM, GAM, NPPEN
	CBI (mean):	**0.75**
Round sardinella	Parameters:	SST, SSTr, log_PP
	Models:	GLM, RF, FDA, NPPEN
	CBI (mean):	**0.88**
European sprat	Parameters:	SST, SSTr, log_PP
	Models:	GLM, MARS, NPPEN
	CBI (mean):	**0.92**
European anchovy	Parameters:	SST, SSTvar, SSS
	Models:	GLM, FDA, MARS
	CBI (mean):	**0.88**
Bogue	Parameters:	SST, SSTr
	Models:	GLM, ANN, NPPEN
	CBI (mean):	**0.65**

The selected SDMs had a CBI > 0.5 and satisfying response curves. **Parameters:** (**SST**) Sea Surface Temperature, (**SSTr**) annual range of Sea Surface Temperature, (**SSTvar**) monthly variance of Sea Surface Temperature, (**log_PP**) log-transformed Primary Production and (**SSS**) Sea Surface Salinity. **Models**: (**GLM**) Generalised Linear Model, (**GAM**) Generalised Additive Model, (**GBM**) Generalised Boosting Model, (**ANN**) Artificial Neural Network, (**FDA**) Flexible Discriminant Analysis, (**MARS**) Multiple Adaptive Regression Splines, (**RF**) Random Forest and (**NPPEN**) Non Parametric Probabilistic Ecological Niche model.

By applying our environmental filtration framework, we substantially improved the modelling of most of the SPFs spatial distributions (Fig. 3, individual contributions of the filtration process and the convexhull are presented in Supplementary material 5), except for the European pilchard (Fig. 3b). Specifically, we observed an increase in mean CBI values that ranged from +0.05 to +0.23 (Fig. 3). For most SPFs, lower Environmental Suitability Index (ESI) values were obtained (-0.2 without filtration to –0.6 with filtration), suggesting that our procedure alleviated the risk of over-prediction, especially in the Black and Baltic seas, and beyond 60 °N where species have never been observed (Fig. 4, left panels). By increasing ESI values from +0.4 to +0.6, environmental filtration also emphasised regions known to be highly suitable for SPF species, but in which occurrences were only scarcely available (*e.g.* in the eastern Mediterranean Sea for Atlantic horse mackerel, round sardinella and bogue; Fig. 4a, f and g).

3.2. Contemporary (1990–2017) environmental suitability of small pelagic fishes

We then represented the contemporary (1990–2017) spatial distribution of the seven SPFs in the spatial domain ranging from 10 to 70 °N and from 30 °W to 45 °E (Fig. 4, middle panel) Environmental suitabilities at the calibration range (*i.e.* the entire distribution range) are provided in Supplementary material 6.

According to the observed and modelled distributions (Fig. 4, left and middle panels), two species groups were identified with respect to their environmental suitability along the European coasts. The first group encompassed temperate-to-cold water species (hereafter "temperate-cold" species; *i.e.* Atlantic horse mackerel, European pilchard, European sprat and European anchovy; Fig. 4a–d) that were more likely to be present in northern Europe. The second grouped temperate-to-warm water species (hereafter "temperate-warm" species; *i.e.* Mediterranean horse mackerel, round sardinella and bogue; Fig. 4e–g) located along the Mediterranean coasts down, to North Africa.

The four temperate-cold species showed the highest ESI values in the North Sea, in the Celtic Sea, in the Bay of Biscay (ESI values > 0.8) and to a lesser extent along Norwegian coasts (ESI values ranging from 0.2 to 0.8). For all temperate-cold species, but European pilchard, high ESI values (from 0.4 to 0.8) were expected in the western and central regions of the Baltic Sea (Fig. 4), suggesting that these species can tolerate a wide salinity range (from 8 to 38) and a high thermal variability (up to 20 °C annual range). All temperate-cold species, but European sprat, showed high ESI values (from 0.6 to 0.8) in the north-western part of the Mediterranean basin (Fig. 4). For all temperate-cold species, the modelled ESIs are in accordance with the observation data except in southern Iceland, western Norway and to a lesser extent in the eastern Black Sea where positive ESI values (between 0.05 to 0.6) are predicted while no observed distribution is available.

The three temperate-warm species showed the highest ESI values (from 0.4 to 0.8) in nearly all the regions of the Mediterranean Sea and medium to low ESI values (from 0.2 to 0.7) in the Black Sea and along the north-western African coasts. However, some discrepancies among species were detected (Fig. 4). Round sardinella appears as the most southern SPF species with no suitable conditions north of the Portuguese coast. On the contrary, Mediterranean horse mackerel and bogue showed high ESI values (from 0.6 to 0.8) along the Atlantic coasts from the Celtic sea down to northern Africa, up to 0.8 in the Bay of Biscay. While bogue showed maximum ESI values (> 0.8) in the whole Mediterranean Sea, only the north-western regions of the Mediterranean Sea were highly suitable for Mediterranean horse mackerel and round

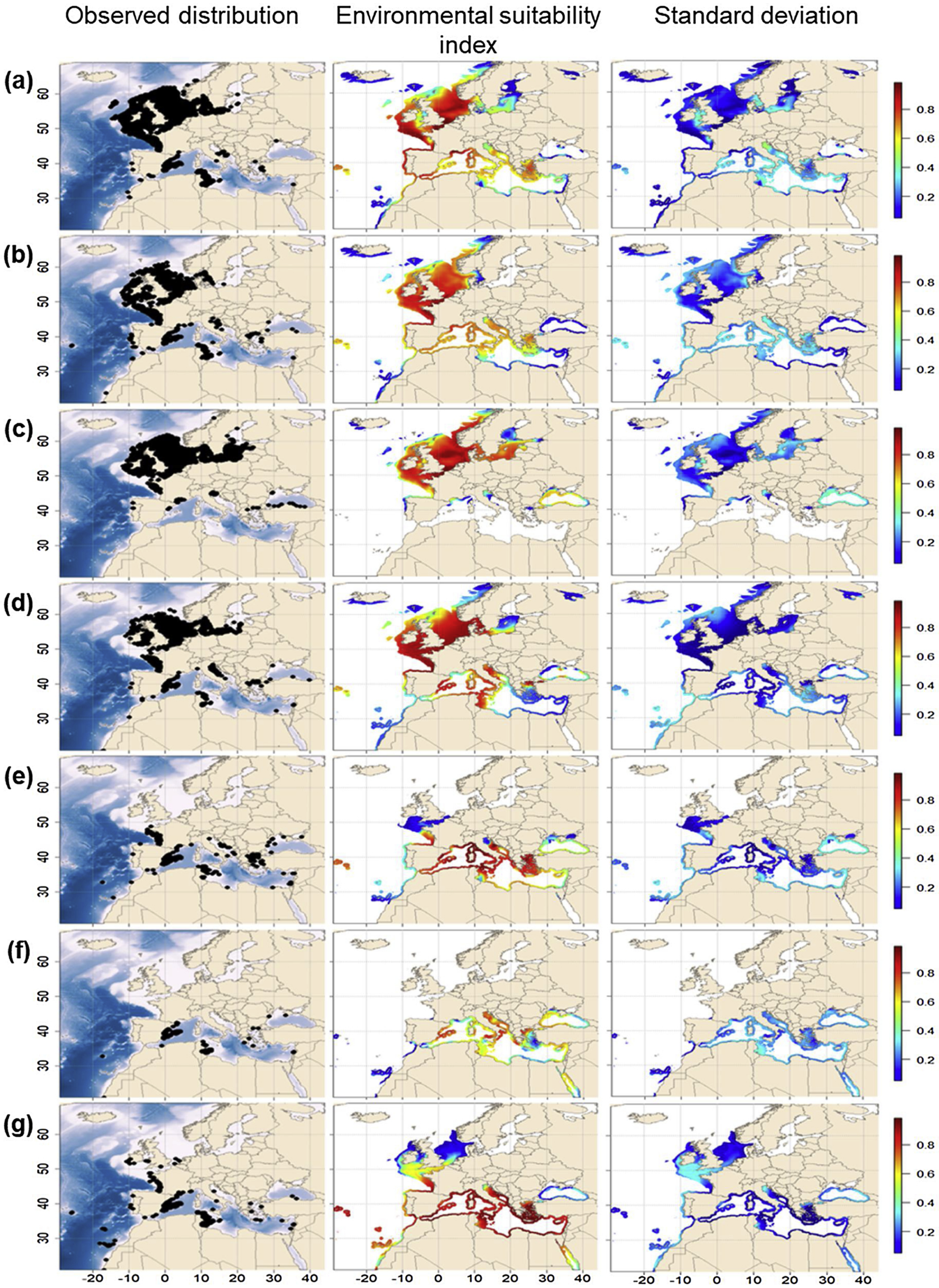


Fig. 4. Contemporary (1990–2017) observed distribution (left panels), modelled environmental suitability index (0–1, middle panels) and its associated standard deviation (0–1, based on all validated SDMs and cross-validation runs, right panels) for (a) Atlantic horse mackerel, (b) European pilchard, (c) European sprat, (d) European anchovy, (e) Mediterranean horse mackerel, (f) round sardinella and (g) bogue.

sardinella. The modelled ESIs are in accordance with the observation data except in the North Sea for Mediterranean horse mackerel and Bogue and to a lesser extent in the eastern Black Sea for all temperate-warm species. These regions highlight positive ESI values (between 0.05 and 0.6) while no observations are available. These discrepancies may result from an absence of sampling in these regions or external factors hindering species establishment despite suitable environmental conditions.

3.3. Model uncertainties

Two main sources of uncertainties in projected species distributions were considered in our study: (i) biological uncertainties, related to the quality of occurrence datasets and (ii) numerical uncertainties, inherent to the selection of different modelling algorithms (Pearson et al., 2006; Buisson et al., 2010). Standard deviations (SD) – computed, for each species, from outputs that originated from both selected algorithms and cross-validation runs – ranged from 0.1 to 0.4, indicating a convergence between models (Fig. 4, right panels). The lowest SD values (close to 0.2) were found in the north-western Mediterranean Sea for virtually all SPFs, and in the Bay of Biscay and in the North Sea when temperate-cold species were studied (Fig. 4, a–d). The highest SD values (close to 0.4) were observed in the Mediterranean Sea for Mediterranean horse mackerel, European pilchard and round sardinella (Fig. 4, e–g). For all species, the coefficient of variation (CV; Supplementary material 7) highlighted very low CV variations (< 20 %) towards their centre of distribution (in the Mediterranean Sea for all species and North Sea for temperate-cold species) while showing high variations at the leading or the trailing edge of their distribution (up to 100 % in the Black, Baltic and the Norwegian seas).

4. Discussion

By combining several numerical techniques such as the convex hull method, the ensemble models approach and an examination of species response curves in a comprehensive modelling framework, we modelled the contemporary (1990–2017) environmental suitability of seven of the most commercially and ecologically important European small pelagic fish. By relying on both an understanding of the ecological requirements of species and on the use of innovative statistical tools, our framework allowed us to focus only on the best models, to improve the way species distribution modelling is carried out, and therefore to produce more robust ecological scenarios.

At a macroecological level, thermal-induced effects have been frequently related to latitudinal mean temperature gradients (Angilletta, 2011). While our analysis showed that mean temperature (SST) had a major influence on species distributions, we also revealed the key role of temperature seasonality (SSTr) and short-term temperature variations (SSTvar) in shaping distributional ranges (Table 2). Small pelagic fishes are marine ectotherms, that mainly depend on external heat sources, their body temperature being directly controlled by environmental conditions directly (Checkley et al., 2009). Changes in temperature may therefore affect SPFs' physiological performances (*i.e.* their fitness; Perry et al., 2005; Payne et al., 2016). Because the relationship between temperature and fitness occurred through species' thermal optimum and range, and because SPFs are short lifespan species (Checkley et al., 2009), annual temperature changes may affect several life stages (especially reproduction and larval development; *e.g.* Peck et al., 2013), with long-term consequences on population dynamics (Fréon et al., 2005). Small pelagic fishes may also experience ontogenetic shifts in thermal tolerance during their development (Peck et al., 2013) and temperature seasonality (here SSTr) may either favour or perturb species development, with potential consequences on distributional patterns (Fig. 4, middle panels; Peck et al., 2013). This is especially noticeable in regions characterised by an important thermal variability, such as in the Black and Azov seas, in the Northern Adriatic Sea, in the Baltic Sea and to a lesser extent in the eastern part of the North Sea. Considering thermal variability in SDMs (*e.g.* the monthly SST variance) may therefore help to better define species environmental suitability and to minimise the risk of over-prediction at the leading and the trailing edges of their distributions (Lenoir et al., 2011).

When used in distribution modelling, regression-based algorithms such as GLM, are known to be rather sensitive to environmental sampling bias, which may induce type I errors (*i.e.* false positive), with consequences on projected species environmental suitability (Araújo and Guisan, 2006; Dormann et al., 2007). However, as for many other species (*e.g.* Boakes et al., 2010), commonly available databases of SPFs provide a distorted view of their actual distribution because of spatial and temporal bias in species observations (*e.g.* Beck et al., 2014). When the time comes to evaluate the quality of biodiversity datasets, three major issues have been raised in the literature (*e.g.* Kramer-Schadt et al., 2013; Guillera-Arroita et al., 2015): the influence of (i) prevalence, *i.e.* the proportion of sites in which the species was recorded as present, (ii) imperfect species detection and (iii) sampling bias. Despite an increasing availability of information, the biogeographic distribution of most species remain still frequently incomplete (Bini et al., 2006); a shortcoming explained, inter alia, by heterogeneous sampling effort among surveys, or the inaccessibility of some areas. For all SPF datasets, this effect is undeniable when comparing the north-western Mediterranean Sea, the Bay of Biscay, the North Sea with other European regions. (Fig. 4, left panels). To lower this issue, a plethora of data sources (*e.g.* standardised scientific surveys, biodiversity portals) are now available in collaborative databases (*e.g.* GBIF), offering more cohesive summaries of species' distributions although leading – sometimes – to enhanced spatial and environmental biases (Kramer-Schadt et al., 2013; Beck et al., 2014). Considering independent distributional data (*i.e.* from private collections or from the literature; Beck et al., 2013) along with the associated pre-processing (*e.g.* Kramer-Schadt et al., 2013; Varela et al., 2014; Aiello-Lammens et al., 2015; Fithian et al., 2015), can contribute to cover the ecological niches of species more comprehensively and to improve model accuracy. By coupling these procedures with our restricting convex hull pseudo-absence selection, we (i) assigned the same weight to environmental conditions independently of the observation density (*i.e.* alleviating observation sampling bias), (ii) lowered the weight of presence records at the distribution edge (*i.e.* avoiding the risk of over-prediction) and (iii) selected unbiased pseudo-absence (*i.e.* independent of the observation bias).

Applying environmental filtering and the restricted convexhull pseudo-absence selection method resulted in ensemble models characterised by a reduced ESI in over-sampled areas and an increased ESI in undersampled areas. Our results are consistent with our expectations and in line with previous studies that suggested that random generation of pseudo-absences and/or a selection process based on geographical criterion may lead to lower predictability (*e.g.* Wisz and Guisan, 2009; Hattab et al., 2014). Although real absences lead to higher model accuracy (Wisz and Guisan, 2009), they are rarely available (Boakes et al., 2010) and determining the location of pseudo-absences on the basis of a statistical analysis such as the convex hull is a reliable alternative (Hattab et al., 2013). Finally, our approach limits spurious species response curves (*i.e.* overfitted or bimodal curves; Supplementary material 4) and decreases the risk of over-predictions towards the edge of the species range. We acknowledge that we may have slightly under-predicted the European pilchard distribution in Kattegat (*i.e.* strait between Denmark and Sweden); the high amount of occurrence records slightly outside the modelled distribution in this region may have biased the calculation of the CBI. Despite the well-known robustness of this index (Breiner et al., 2015; Faillettaz et al., 2019), our result highlight that no evaluation metric is optimal and that both comparison between observed and modelled distributions and examination of species responses curves are essential for assessing the reliability of model outputs.

While the assessment of the environmental suitability for a given species may differ – slightly or markedly – from one SDM to another (Pearson et al., 2006; Buisson et al., 2010), it is still challenging to identify the most appropriate model (see discussion in Buisson et al., 2010). Even if several methods have been recently proposed, no consensus has emerged (see discussion in Leroy et al., 2018) and the use of different – well-fitted and evaluated – SDMs may help to better simulate potential species distributions, for past, contemporary and future environmental conditions (Araújo and New, 2007). In complementarity with a multi-SDM approach, we think that researchers should examine species response curves during the evaluation process (*e.g.* Elith and Leathwick, 2009; Jarnevich et al., 2018; Erauskin-Extramiana et al., 2019). As observed for Mediterranean horse mackerel (see details in Supplementary material 4), we invalidated response curves that were statistically significant but not in agreement with the ecological niche theory. Without this complementary evaluation method, the corresponding algorithms would have been considered in the ensemble model, therefore potentially resulting in spurious patterns of ESIs. Therefore, this multi-criteria evaluation procedure is of great interest from a (i) numerical (*i.e.* metric adapted to presence-only datasets) and an ecological (*i.e.* validation of the species-environment relationships) perspective. Note that the seven SPFs we chose are representative of a large spectrum of environmental conditions, from temperate-to-cold waters (*e.g.* European sprat) to temperate-to-warm waters (*e.g.* bogue and round sardinella).To conclude, our framework has been faced with a wide range of environmental conditions, allowing us to better evaluate its robustness, sensitivity and possible transferability to other species and ecosystems.

In this work, we have estimated species' potential niche and not the realised niche (Soberón and Nakamura, 2009). We caution that additional environmental parameters, biological interactions and species life traits (*e.g.* dispersal abilities) may explain why we detected environmentally suitable conditions in regions where SPFs were not observed (*e.g.* the Norwegian Sea; Pulliam, 2000; Pearman et al., 2008). Considering the role of biotic interactions in shaping species distributions (Chaalali et al., 2016) would improve the reliability of SDMs outputs by better estimating and simulating the realised niche of species (Wisz et al., 2013; Louthan et al., 2015). Including dispersal mechanisms while accounting for oceanic currents and physical barriers after the potential distribution modelling step may help to refine the distributional range of species (Engler and Guisan, 2009). These approaches require an exhaustive ecological understanding of the interaction process at a macroecological scale and a deep knowledge of species life traits to implement metrics that simulate the ability of species to disperse (*e.g.* Petitgas et al., 2012). Moreover, it is important to notice that no direct correlations between ESI (potential or realised) and spatialised biomass or official catches have been established in the literature although temporal correlations have been suggested however (*e.g.* Chaalali et al., 2016). Therefore, discrepancies between SPF's ESI, biomass and official catches (*e.g.* FAO, 2016) may be explained by population-related parameters (*e.g.* recruitment, growth, biotic interaction) or management policies and stock status (*e.g.* under or overfishing), respectively. Finally, inter-specific absolute ESI comparison is challenging because of the monospecific nature of SDMs.

Our study presents a detailed environmental suitability assessment of seven of the most heavily harvested European SPFs. By focusing on the most common sources of errors and uncertainties in SDMs, we designed a comprehensive - fully transferable to other species and ecosystems - modelling framework which is intended to elaborate more robust ecological scenarios. Our framework addressed several critical steps in SDMs, *i.e.* the treatment of sampling biases in observation records, the generation relevant pseudo-absences and a dual assessment of model outputs that proposes to evaluate models from both a numerical and an ecological perspective. In a conservation decision-making perspective, these different steps are essential to increase confidence in SDMs, a prerequisite to propose effective resource management measures (*e.g.* accounting for environmental stress) or to measure the effectiveness of protected areas (*e.g.* regarding environmental resilience). Moreover, when used in combination with scenarios of future environmental conditions (*i.e.* IPCC climate scenarios), this framework provides robust contemporary predictions to assess possible changes in species distribution in the context of global climate change. Despite the growing literature on the development and testing of new modelling and evaluation processes, the use of SDMs in quantitative resource management and scientific surveys is still a great challenge.

Funding

This work was supported by the Prince Albert II of Monaco foundationthrough the project CLIM-ECO^2. AS's PhD is funded by the Provence-Alpes-Côte-d'Azur (PACA) Regionin partnership with the Comité Régional des Pêches Maritimes et des Elevages Marins (CRPMEM) de PACA.

Author contribution

VR and PF conceived and supervised the study. AS, VR and EG collected the data. AS performed the numerical analysis. BL, GB, TH, EG and VR helped in the modelling process. AS and EG wrote the first draft. BL provided the code to use BIOMOD2. All authors made substantial contributions in the successive versions of the manuscript.

Declaration of Competing Interest

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

Acknowledgements

This paper is dedicated to the memory of Prof. Patrice Francour who passed away on October 13th, 2019. He devoted his life to the protection and understanding of Mediterranean ecosystems, initiated and co-supervised this work. We thank J.O. Irisson for his support on the NPPEN R code. We also acknowledge the Ocean Biogeographic Information System (OBIS), the Global Biodiversity Information Facility (GBIF) and Fishbase for providing species observation data. Finally, we thank the IPCC Coupled Model Intercomparison Project (CMIP) and the climate modelling groups for making available their model output.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ecolmodel.2019.108902.

References

Aiello-Lammens, M.E., et al., 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38, 541–545.

Albouy, C., et al., 2012. Combining projected changes in species richness and composition reveals climate change impacts on coastal Mediterranean fish assemblages. Glob. Change Biol. 18, 2995–3003.

Allouche, O., et al., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS): assessing the accuracy of distribution models. J. Appl. Ecol. 43, 1223–1232.

Angilletta, M.J., 2011. Thermal Adaptation: a Theoretical and Empirical Synthesis. Oxford University Press.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33, 1677–1688.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47.

Beaugrand, G., Kirby, R.R., 2018. How do marine pelagic species respond to climate change? Theories and observations. Ann. Rev. Mar. Sci. 10, 169–197.

Beaugrand, G., et al., 2011. A new model to assess the probability of occurrence of a species, based on presence-only data. Mar. Ecol. Prog. Ser. 424, 175–190.

Beaugrand, G., et al., 2013. Applying the concept of the ecological niche and a macroecological approach to understand how climate influences zooplankton: advantages, assumptions, limitations and requirements. Prog. Oceanogr. 111, 75–90.

Beaugrand, G., et al., 2018. Marine biodiversity and the chessboard of life. PLoS One 13, e0194006.

Beaumont, L.J., et al., 2008. Why is the choice of future climate scenarios for species distribution modelling important? Ecol. Lett. 11, 1135–1146.

Beck, J., et al., 2013. Online solutions and the 'Wallacean shortfall': what does GBIF contribute to our knowledge of species' ranges? Divers. Distrib. 19, 1043–1050.

Beck, J., et al., 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecol. Inform. 19, 10–15.

Bellard, C., et al., 2013. Will climate change promote future invasions? Glob. Change Biol. 19, 3740–3748.

Bellard, C., et al., 2016. Major drivers of invasion risks throughout the world. Ecosphere 7, e01241.

Ben Rais Lasram, F., et al., 2010. The Mediterranean Sea as a 'cul-de-sac' for endemic fishes facing climate change: a marine endemic hotspot under threat. Glob. Change Biol. 16, 3233–3245.

Bini, L.M., et al., 2006. Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. Divers. Distrib. 12, 475–482.

Boakes, E.H., et al., 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol. 8, e1000385.

Breiner, F.T., et al., 2015. Overcoming limitations of modelling rare species by using ensembles of small models (B Anderson, Ed.). Methods Ecol. Evol. 6, 1210–1218.

Buisson, L., et al., 2010. Uncertainty in ensemble forecasting of species distribution. Glob. Change Biol. 16, 1145–1157.

Casey, K.S., et al., 2010. The past, present, and future of the AVHRR pathfinder SST program. Oceanography From Space. Springer, Dordrecht, pp. 273–287.

Chaalali, A., et al., 2016. From species distributions to ecosystem structure and function: a methodological perspective. Ecol. Modell. 334, 78–90.

Checkley, D., et al., 2009. Climate Change and Small Pelagic Fish. Cambridge University Press.

Cheung, W.W.L., et al., 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish Fish. 10, 235–251.

Colwell, R.K., Rangel, T.F., 2009. Hutchinson's duality: the once and future niche. PNAS 106, 19651–19658.

Cornwell, W.K., et al., 2004. A trait-based test for habitat filtering: convex hull volume. Ecology 87, 1465–1471.

Crisp, M.D., et al., 2009. Phylogenetic biome conservatism on a global scale. Nature 458, 754–756.

Cristofari, R., et al., 2018. Climate-driven range shifts of the king penguin in a fragmented ecosystem. Nat. Clim. Change 8, 245–251.

Cury, P., 2000. Small pelagics in upwelling systems: patterns of interaction and structural changes in "wasp-waist" ecosystems. ICES J. Mar. Sci. 57, 603–618.

Dormann, C.F., et al., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628.

Dormann, C.F., et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.

Dufresne, J.-L., et al., 2013. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. Clim. Dyn. 40, 2123–2165.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.

Elith, J., et al., 2005. The evaluation strip: a new and robust method for plotting predicted responses from species distribution models. Ecol. Modell. 186, 280–289.

Engler, R., Guisan, A., 2009. MigClim: predicting plant distribution and dispersal in a changing climate. Divers. Distrib. 15, 590–601.

Erauskin-Extramiana, M., et al., 2019. Historical trends and future distribution of anchovy spawning in the Bay of Biscay. Deep. Sea Res. Part II: Top. Stud. Oceanogr. 159, 169–182.

Faillettaz, R., et al., 2019. Atlantic Multidecadal Oscillations drive the basin-scale distribution of Atlantic bluefin tuna. Sci. Adv. 5, eaar6993.

FAO, 2016. The State of Mediterranean and Black Sea Fisheries. General Fisheries Commission for the Mediterranean.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.

Fithian, W., et al., 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol. Evol. 6, 424–438.

Fréon, P., et al., 2005. Sustainable exploitation of small pelagic fish stocks challenged by environmental and ecosystem changes: a review. Bull. Mar. Sci. 76, 79.

Getz, W.M., Wilmers, C.C., 2006. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. Ecography 27, 489–505.

Giorgetta, M.A., et al., 2013. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5: climate changes in MPI-ESM. J. Adv. Model. Earth Syst. 5, 572–597.

Gleason, H.A., 1926. The individualistic concept of the plant association. Bull. Torrey Bot. Club 53, 7–26.

Guillera-Arroita, G., et al., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Glob. Ecol. Biogeogr. 24, 276–292.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.

Hattab, T., et al., 2013. The use of a predictive habitat model and a fuzzy logic approach for marine management and planning. PLoS One 8, e76430.

Hattab, T., et al., 2014. Towards a better understanding of potential impacts of climate change on marine species distribution: a multiscale modelling approach. Glob. Ecol. Biogeogr. 23, 1417–1429.

Helaouet, P., Beaugrand, G., 2009. Physiology, ecological niches and species distribution. Ecosystems 12, 1235–1245.

Hengl, T., et al., 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecol. Modell. 220, 3499–3511.

Hirzel, A.H., et al., 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecol. Modell. 199, 142–152.

Hourdin, F., et al., 2013. Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model. Clim. Dyn. 40, 2167–2192.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harb. Symp. Quant. Biol. 22, 415–427.

Hutchinson, G.E., 1978. An Introduction to Population Ecology. Yale University Press, New Haven.

Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. Bulletin de la Societe Vaudoise des Sciences Naturelles 44, 223–270.

Jarnevich, C.S., et al., 2018. Forecasting an invasive species' distribution with global distribution data, local data, and physiological information. Ecosphere 9, e02279.

Jiménez-Valverde, A., et al., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. Divers. Distrib. 14, 885–890.

Jones, C.D., et al., 2011. The HadGEM2-ES implementation of CMIP5 centennial simulations. Geosci. Model Dev. Discuss. 4, 689–763.

Kirby, R.R., Beaugrand, G., 2009. Trophic amplification of climate warming. Proc. R. Soc. B: Biol. Sci. 276, 4095–4103.

Kramer-Schadt, S., et al., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Divers. Distrib. 19, 1366–1379.

Lenoir, S., et al., 2011. Modelled spatial distribution of marine fish and projected modifications in the North Atlantic Ocean. Glob. Change Biol. 17, 115–129.

Leroy, B., et al., 2014. Forecasted climate and land use changes, and protected areas: the contrasting case of spiders. Divers. Distrib. 20, 686–697.

Leroy, B., et al., 2018. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. J. Biogeogr. 45, 1994–2002.

Levitus, S., 2011. Climatological atlas of the world ocean. Eos Trans. Am. Geophys. Union 64, 962–963.

Lobo, J.M., Tognelli, M.F., 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. J. Nat. Conserv. 1, 1–7.

Louthan, A.M., et al., 2015. Where and When do Species Interactions Set Range Limits? Trends Ecol. Evol. 30, 780–792.

Mahalanobis, P., 1936. On the generalised distance in statistics. In: Proceedings of the National Institute of Science. India. pp. 49–55 2.

Merow, C., et al., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36, 1058–1069.

Mielke, P.W., et al., 1981. Application of multi-response permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns. Mon. Wea. Rev. 109, 120–126.

Montgomery, D.C., 2005. Design and Analysis of Experiments. Wiley.

Newbold, T., 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress Phys. Geogr.: Earth Environ. 34, 3–22.

Payne, N.L., et al., 2016. Temperature dependence of fish performance in the wild: links with species biogeography and physiological thermal tolerance. Funct. Ecol. 30, 903–912.

Pearman, P.B., et al., 2008. Niche dynamics in space and time. Trends Ecol. Evol. 23, 149–158.

Pearson, R.G., et al., 2006. Model-based uncertainty in species range prediction. J. Biogeogr. 33, 1704–1711.

Peck, M.A., et al., 2013. Life cycle ecophysiology of small pelagic fish and climate-driven changes in populations. Prog. Oceanogr. 116, 220–245.

Perry, A.L., et al., 2005. Climate change and distribution shifts in marine fishes. Science 308, 1912–1915.

Petitgas, P., et al., 2012. Anchovy population expansion in the North Sea. Mar. Ecol. Prog. Ser. 444, 1–13.

Poloczanska, E.S., et al., 2013. Global imprint of climate change on marine life. Nat. Clim. Change 3, 919–925.

Porfirio, L.L., et al., 2014. Improving the use of species distribution models in conservation planning and management under climate change (L Kumar, Ed.). PLoS One 9, e113749.

Proosdij, A.S.J. van, et al., 2016. Minimum required number of specimen records to develop accurate species distribution models. Ecography 39, 542–552.

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecol. Lett. 3, 349–361.

Raybaud, V., et al., 2015. Climate-induced range shifts of the American jackknife clam Ensis directus in Europe. Biol. Invasions 17, 725–741.

Raybaud, V., et al., 2017. Forecasting climate-driven changes in the geographical range of the European anchovy (Engraulis encrasicolus). ICES J. Mar. Sci. 74, 1288–1299.

Schmidt, G.A., et al., 2014. Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive: GISS MODEL-E2 CMIP5 SIMULATIONS. J. Adv. Model. Earth Syst. 6, 141–184.

Smith, W.H.F., Sandwell, D.T., 1997. Global sea floor topography from satellite altimetry and ship depth soundings. Science 277, 1956–1962.

Soberon, J., Peterson, A.T., 2005. Interpretation of models of fundamental ecological Niches and Species' distributional areas. Biodivers. Inform. 2, 1–10. https://doi.org/10.17161/bi.v2i0.4.

Soberón, J., Nakamura, M., 2009. Niches and distributional areas: concepts, methods, and assumptions. Proc. Natl. Acad. Sci. U. S. A. 106 (Suppl. 2), 19644–19650.

Sørensen, T., 1948. {A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons}. Biol. Skr. 5, 1–34.

Stevens, B., et al., 2013. Atmospheric component of the MPI-M earth system model: ECHAM6. J. Adv. Model. Earth Syst. 5, 146–172.

Støa, B., et al., 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.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

Thuiller, W., et al., 2009. BIOMOD - a platform for ensemble forecasting of species distributions. Ecography 32, 369–373.

Thuiller, W., Georges, D., Engler, R., Breiner, F., 2016. biomod2: Ensemble platform for species distribution modeling. R package version 3.3-7.

Varela, S., et al., 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37, 1084–1091.

Voldoire, A., et al., 2013. The CNRM-CM5.1 global climate model: description and basic evaluation. Clim. Dyn. 40, 2091–2121.

Wiens, J.A., et al., 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. PNAS 106, 19729–19736.

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 Ecol. 9, 8.

Wisz, M.S., et al., 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol. Rev. 88, 15–30.