

Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling

Neftalí Sillero^{a,*}, Salvador Arenas-Castro^{a,b}, Urtzi Enriquez-Urzelai^{c,d}, Cândida Gomes Vale^e, Diana Sousa-Guedes^a, Fernando Martínez-Freiría^c, Raimundo Real^f, A.Márcia Barbosa^a

^a CIGCE - Centro de Investigação em Ciências Geo-Espaciais, Alameda do Monte da Virgem, 4430-146 Vila Nova de Gaia, Portugal

^b Escola Superior Agrária, Instituto Politécnico de Viana do Castelo, 4990-706 Ponte de Lima, Portugal

^c CIBIO - Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus de Vairão, Rua Padre Armando Quintas, N.º 7. 4485-661 Vairão, Vila do Conde, Portugal

^d Czech Academy of Sciences, Institute of Vertebrate Biology, Květná 8, 60365 Brno, Czech Republic

^e CIIMAR - Interdisciplinary Centre of Marine and Environmental Research of the University of Porto. Terminal de Cruzeiros do Porto de Leixões. Avenida General Norton de Matos, S/N. 4450-208 Matosinhos, Portugal

^f Grupo de Biogeografía, Diversidad y Conservación. Departamento de Biología Animal. Facultad de Ciencias. Universidad de Málaga, 29071 Málaga, España

ARTICLE INFO

Keywords:

Species distribution models
Best practices
Model assumptions and limitations
Model preparation and parametrization
Model validation
Model performance

ABSTRACT

The use of correlative ecological niche models has highly increased in the last decade. Despite all literature and textbooks in this field, few practical guidelines exist on the correct application of these techniques. We present here a step-by-step guideline explaining best practices for calculating correlative ecological niche models considering their conceptual and statistical assumptions and limitations. We divided the modelling process into four stages: 1) data collection and preparation; 2) model calculation; 3) model evaluation and validation; 4) and model application. Based on ecological niche theory, we review the concepts of ecological niche and how they can be modelled; classes of correlative models; modelling software; selection of study area; data sources for species records and environmental variables; types of species records and their influence on correlative models; errors in species records; minimum number of species records and environmental variables; effects of prevalence, sampling design, biases, and collinearity between variables; model calculation; model projection to different scenarios in time and space; ensemble modelling; model validation; classification, discrimination and calibration metrics; calculation of null models; analysis of model results; and model thresholding. This guideline is expected to help potential users to obtain better results when using correlative ecological niche models.

1. What is ecological niche modelling?

Ecological niche models (ENMs) are empirical or mathematical approximations to the ecological niche of a species (Barbosa et al., 2012). ENMs relate physiological or chorological (i.e. species location) data to environmental variables using statistical methods or theoretically derived response surfaces, with the aim of describing, understanding and/or predicting the distribution of species (Franklin, 2010; Peterson et al., 2011; Sillero, 2011; Barbosa et al., 2012; Guisan et al., 2017). The mathematical output of ENMs can be an equation relating the expected distribution of the species (the dependant variable) to a set of environmental predictors (the independent variables) (Olivera-Gomez and Mellink, 2005) or the response curves describing how each

environmental predictor relates to the dependant variable (Chapman et al., 2019). This mathematical model can be spatialised into a cartographic model, i.e. a map representing habitat suitability, probability of the species occurrence, or the favourability (see definition in Table 1) for species occurrence (Acevedo et al., 2012; Hatten, 2014; Ørsted and Ørsted, 2019). Therefore, ENMs are predicted in the environmental space and projected to the geographical space (see definitions in Table 1). ENMs have wide applicability in numerous fields (see Table 2).

Authors use different terms to refer to a family of related concepts and modelling approaches (Franklin, 2010; Peterson et al., 2011; Guisan et al., 2017). Species Distribution Model (SDM) is a more biogeographical term, which emphasises the distribution of species, and it reflects an interest in historical, anthropogenic and population dynamics

* Corresponding author. CIGCE – Centro de Investigação em Ciências Geo-Espaciais, Alameda do Monte da Virgem, 4430-146 Vila Nova de Gaia, Portugal.
E-mail address: neftali.sillero@gmail.com (N. Sillero).

<https://doi.org/10.1016/j.ecolmodel.2021.109671>

Received 12 April 2021; Received in revised form 6 July 2021; Accepted 8 July 2021

Available online 19 July 2021

0304-3800/© 2021 Elsevier B.V. All rights reserved.

Table 1
Definitions for favourability, environmental space, and geographical space.

Term	Definition
Favourability	Contribution of the environment to the presence probability after removing the effect of prevalence (or proportion of presences). In other words, it is the degree to which a particular combination of environmental variables favours the presence of the species over and above its general prevalence in the study area.
Environmental space	The space defined by the environmental variables that influence the species' distribution, i.e. the multidimensional space of environmental variables where the species' ecological niche exists. The axes of this n-dimensional hypervolume define the environmental space.
Geographical space	Any geo-referenced space, real or virtual, where the ecological niche models are projected.

Table 2
List of possible applications of ecological niche models with some (non-exhaustive) example references.

Application	References
Identification of species ecological niches	(Chefaoui et al., 2016)
Niche identification with spatial representation	(Ørsted and Ørsted, 2019)
Niche identification without spatial representation	(Olivera-Gomez and Mellink, 2005)
Comparison of species niches with spatial representation	(Warren et al., 2010; Ahmadzadeh et al., 2016)
Comparison of species niches without spatial representation	(Broennimann et al., 2012; Blonder et al., 2014; Di Cola et al., 2017; Martínez-Freiría et al., 2020)
Prediction of past and future species distributions	(Martínez-Freiría et al., 2013, 2015, 2017, 2020; Sillero and Carretero, 2013)
Prediction of fossil species distributions	(Nogués Bravo, 2009; Chiarenza et al., 2019)
Prediction of disease prevalence and/or spread	(Cumming, 2002; Wint et al., 2002; Dicko et al., 2014)
Analyses of climate change effects on species distributions	(Real et al., 2017; Estrada et al., 2018)
Analyses of species conservation status	(Carvalho et al., 2010a)
Species conservation planning	(Barbosa et al., 2003; Regos et al., 2021)
Identification of hybrid zones	(Martínez-Freiría et al., 2008; Tarroso et al., 2014)
Integration of genetic data	(Espregueira Themudo et al., 2009; Tarroso et al., 2014; Ikeda et al., 2017)
Analyses of species expansions	(Sullivan et al., 2012; Franch et al., 2015)
Analyses of species invasions	(Ficetola et al., 2007; Silva-Rocha et al., 2015)
Prediction of species richness	(Sillero et al., 2009; Martínez-Freiría et al., 2013; Vale et al., 2016)
Analyses of dispersion hypotheses	(Engler and Guisan, 2009)
Definition of protected areas	(Carvalho et al., 2010b; 2011)
Prediction of animal road-kills	(Ramp et al., 2005; Sillero, 2008)
Prediction of windmills effects on fauna	(Santos et al., 2013)
Prediction of snake bites	(Yañez-Arenas et al., 2014, 2016; Zacarias and Loyola, 2019)
Prediction of wildfire occurrences	(Batllori et al., 2013)
Analysis of niche partitioning and competing species occurrence	(Martínez-Freiría et al., 2008; Hemami et al., 2018)
Analysis of temporal range dynamics	(Martínez-Freiría et al., 2016; Khalatbari et al., 2018)
Prediction of the inter-annual abundance dynamics for rare plants	(Arenas-Castro et al., 2019)
Crop projections for climate change assessment and adaptation	(Arenas-Castro et al., 2020; Arenas-Castro and Gonçalves, 2021)

aspects of the distribution, in addition to its environmental drivers. SDM has been preferentially used for algorithms that predict occurrence probability, thus using species presences and absences (Guisan and Zimmermann, 2000; Elith and Leathwick, 2009). Climate Envelope

Model (CEM) is specifically focused on the relationship of the species distribution with climate, as it uses only climate-related predictors and a type of algorithm called 'envelope' (Hampe, 2004; Araújo and Peterson, 2012; Booth et al., 2014). Habitat Suitability Model (HSM) normally refers to algorithms predicting how suitable habitats are (e.g. ENFA; Hirzel et al., 2002). The term ENM is more inclusive, as it embraces most types of algorithms to model species distributions (mechanistic and correlative methods, see below), linking them to the ecological niche theory (Sillero, 2011). It helps to keep in mind the importance of ecological niche theory when modelling species distributions (Sillero, 2011).

The arrival of easy-to-use correlative techniques such as Maxent (Phillips et al., 2006) revolutionised the field of ENMs. Researchers were able to model species distributions without providing explicit observations of absence. The outputs were very similar to those provided by presence-absence methods: variable contributions to the model, variable response curves, and a map representing a continuous predicted response – as opposed to the categorical maps provided by climate envelope algorithms (e.g. Bioclim; Booth et al., 2014), with only two classes: suitable and unsuitable. However, what is being modelled is frequently misunderstood (Sillero, 2011). Consequently, a thorough, updated and generic step-by-step guideline on how to model a species' niche is much needed (Sillero et al., 2010; Barbosa et al., 2012; Anderson, 2015; Jarnevich et al., 2015). Recent literature has provided sets of standards for niche modelling under different purposes (Araújo et al., 2019; Feng et al., 2019b; Sofaer et al., 2019b; Zurell et al., 2020). Other studies focus on a specific modelling method (Wilson, 2009; Merow et al., 2013). However, an outline of the sequence of key modelling steps, including best practices, is still lacking. Misunderstandings and bad practices are still frequent in ENMs (Sillero and Barbosa, 2021). Here, we present an updated and detailed guide on how to perform correlative ecological niche models, aiming to highlight two key points: 1) use a modelling method that correctly addresses the research question; and 2) interpret ENM results correctly.

2. Ecological niche background

Many definitions of ecological niche have been proposed over time. The "Grinnellian niche" was the first to be defined (Grinnell, 1917). It considers the ecological niche as a subset of the environment containing the habitat conditions that allow the individuals of a species to survive and reproduce (Grinnell, 1917). The habitat is therefore composed of variables (the so-called scenopoetic variables, e.g. variables able to create scenes) that cannot be modified by the species (e.g. climate; Soberón, 2007). Grinnell's niche is a property of the environment. The "Eltonian niche" was proposed by Elton (1927). Here, the ecological niche is the functional role of a species inside a community (Elton, 1927), mainly according to consumed resources (bionomic variables, e.g. nutrients; Soberón, 2007). Elton's niche is a property of the ecosystem. The "Hutchinsonian niche" provided the first mathematical description of the ecological niche (Hutchinson, 1957). It splits the niche into two categories: fundamental and realised niche (Fig. 1). The fundamental niche is the 'n-dimensional hypervolume' on the environmental space where a species can maintain a viable population and persist over time without immigration. Each dimension is an environmental variable. The realised niche is the subset of the fundamental niche where the species is not excluded by competition (Hutchinson, 1957). Posteriorly, other biotic interactions (e.g. predation, parasitism, symbiosis), dispersal ability, as well as geographical and historical constraints, were also considered limiting the realised niche, in what was called occupied niche (Pearson, 2007). Hutchinson's niche is a property of the species. This niche is a space volume defined by its position, boundaries, size, and shape, but it lacks a formal definition of internal heterogeneity. According to Hutchinson, the fundamental niche supposed equal probability of species' persistence in all localities, and zero survival probability in all points outside the niche. Given that

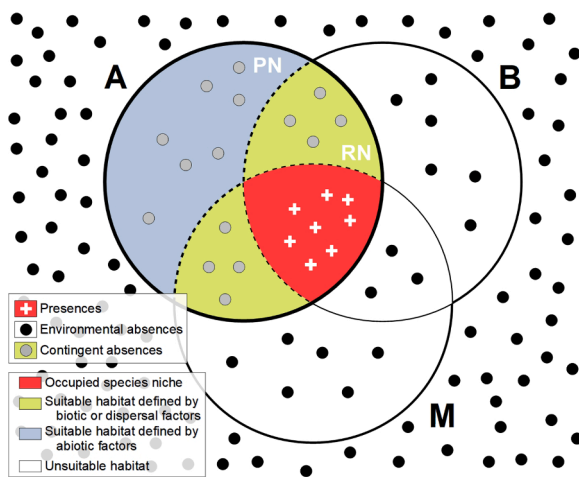


Fig. 1. BAM (Biotic-Abiotic-Movement) diagram. The circle A represents the Potential niche, the abiotic part of the environmental space where a species can survive and persist over time. The circle B represents the part without detrimental biotic interactions. The circle M is the areas of the environmental space where the species can disperse. The Realised niche is the overlap of the three circles. Mechanistic models, when projected to the geographical space, forecast A, i.e. the Potential niche. Correlative models, depending if they use absences or not, and what type of absences are considered (environmental or contingent), provide a closer forecast to the Realised niche.

habitat suitability is a gradient of conditions (from optimal to suboptimal), a graduality in the niche definition that relates niche to habitat suitability is needed. By addressing the task of describing an internal structure of Hutchinson's niche, niches can be considered as the responses of species to habitat conditions (Maguire, 1973), thus treating the niche as an interaction between the species and the environment. Maguire's niche is thus a property of the species–environment complex.

Some expansions to the niche concept were later introduced, with implications for ENM interpretation. First, the potential niche was defined as the part of the fundamental niche that exists on Earth at a given time (Jackson and Overpeck, 2000). Thus, part of the fundamental niche is not expressed in the current conditions, although it may have been present in the past or may appear in the future. This is important, for example, in predictions of future distributions under climate change scenarios. Pulliam (1988) applied the source-sink theory to the ecological niche: populations of a species may occur in unsuitable areas (sinks) if immigrants are constantly arriving from healthy populations (sources). Therefore, species may use a range of environmental conditions larger than their fundamental niche (Pulliam, 1988). This concept is thus crucial for understanding ENM results (see, for example, Santos et al., 2009; Silva-Rocha et al., 2015; Pulido-Pastor et al., 2018). The occurrence of species in areas deemed as unsuitable by an ENM is not necessarily a case of erroneous prediction: some populations may live in suboptimal conditions (Santos et al., 2009; Silva-Rocha et al., 2015). Similarly, species can also be absent from suitable areas. Historical or dispersal reasons, for example, may prevent the species from arriving at suitable habitats (Holt, 2003). In metapopulation theory (Levins, 1969), many species are considered to live in fragmented suitable habitat patches connected by migration, which can be temporarily unoccupied due to demographic stochasticity (Pulido-Pastor et al., 2021). This is also an important concept to consider when interpreting ENM results. Models can identify suitable habitats that are not occupied by the species, and this is not necessarily an error of the model (Pulido-Pastor et al., 2021).

The conceptual differences between the potential and the realised niche can be represented in the Biotic-Abiotic-Movement diagram (BAM; Fig. 1) (Soberón and Peterson, 2005). The niche is represented by the overlap of three circles in the environmental space, each one

describing different driving factors. The circle A represents the abiotic environmental space where the species populations can survive and persist over time, corresponding therefore to the potential niche. The circle B is the environmental space where the species is free of biotic exclusion (e.g. by competition). The circle M is the environmental space where the species can disperse without limitations. The realised niche corresponds to the intersection of the three circles.

3. Main categories of ENMs and the species niche they forecast

ENMs may be classified into three main categories, mechanistic, correlative and hybrid models (Jiménez-Valverde et al., 2008; Dormann et al., 2012), depending on how explicitly biological processes (e.g. metabolism) are incorporated. While mechanistic models typically incorporate physiological, morphological, and behavioural data (Kearney and Porter, 2009), correlative and hybrid models use geographical occurrence data. Mechanistic models tend to predict the fundamental niche (but when expressed in a map, it is the potential niche that is forecasted), and correlative and hybrid models the realised niche (Sillero, 2011). Ideally, we should choose the modelling algorithm depending on the research question and not based on the available data (Guillera-Arroita et al., 2015).

3.1. Mechanistic models

Mechanistic niche models allow us to model, depict, and interrogate species' fundamental and potential niches. Typically, phenotypic trait values (i.e. morphology, physiology and behaviour) of a given organism are translated into performance or fitness components using sets of equations – usually mass and energy balance equations – that define how these traits interact with the environment (Porter et al., 1973; Kearney and Porter, 2009; Kearney et al., 2010; Buckley et al., 2018). That way it is possible to map species' potential niches into the landscape, at the same time knowing the mechanisms or processes encapsulated by the model.

Explicitly incorporating processes and functional traits are advantages of mechanistic models. In doing so they 1) reduce the risk of extrapolation under novel environmental conditions (e.g. when projecting in space or time; Elith and Leathwick, 2009; Elith et al., 2010); (2) help to model the very processes limiting distribution ranges and driving range-shifts (Briscoe et al., 2016; Enriquez-Urzelai et al., 2019); and (3) allow us to account for geographic phenotypic variation, phenotypic plasticity, and evolutionary adaptation (Kearney et al., 2009; Kolbe et al., 2010; Moran et al., 2016; Enriquez-Urzelai et al., 2020). As any hypothetical-deductive approach, mechanistic models can balance conflicting evidence given the theory and, thus, help to improve theoretical knowledge. Unfortunately, the amount of required data on species' traits and the scarcity of available tools (e.g. open-source code) to fit mechanistic niche models have dragged the use of these models (Buckley et al., 2018). Still, species trait data and open-source code for mechanistic modelling are becoming more readily available (e.g. NicheMapR [<https://github.com/mrke/NicheMapR>] and trenchR [<https://github.com/trenchproject/TrenchR>]). This, together with the current appreciation of the need for accounting for processes in ecology, biogeography, and climate change biology (Cabral et al., 2017), is opening new horizons for mechanistic niche modelling.

3.2. Correlative models: presence-absence, presence-background, and presence-only models

Correlative methods can be classified into three main groups depending on the type of species' occurrence data used: presence-absence, presence-background (or profile), and presence-only methods.

Presence-absence methods use observations of species occurrences and absences. Absences correspond to places where the species did not occur upon observation (i.e. upon field sampling). In theory, these

methods distinguish the environmental conditions between occupied and non-occupied habitats, providing the probability of finding the species at each place (Sillero, 2011). The probability of species occurrence in a given habitat represents the proportion of expected occurrence at those specific habitat conditions (Feller, 1968; Guisan and Zimmermann, 2000). The capacity of presence-absence methods in correctly distinguishing between probably occupied and probably unoccupied habitats will depend on what types of absences are used (see below). When georeferenced observations of absences are not available, it is possible as well to lattice the study area using a grid of spatial units, either arbitrary (e.g. UTM cells), natural (e.g. river basins) or administrative (e.g. municipalities), and compute presences as the cells with at least one observation and absences as the cells with no observation of the species (Cressie, 1993). Pseudo-absences (absence points created artificially) can also be used as surrogates of true observation of absences. Presence-absence modelling methods include logistic regression (Brito et al., 1999), generalized linear models and generalized additive models (Guisan et al., 2002), random forest (Breiman, 1999), boosted regression trees (Elith et al., 2008), multivariate adaptive regression splines (Moisen and Frescino, 2002), and artificial neural networks (Tarroso et al., 2012). These methods can also be used with presence-only data when comparing, for example, occurrence localities of two parapatric species or two species with similar ecology (Brito and Crespo, 2002; Real et al., 2005; Arntzen and Espregueira Themudo, 2008; Romero et al., 2014; Torreblanca et al., 2019). Here, the presence of one species does not mean necessarily the absence of the other species; rather, these models distinguish where a species is more likely to occur regarding its parapatric or ecological counterpart.

Presence-background methods compare the available environmental conditions in the study area (i.e. background) with the conditions used by the species, as represented by its occurrences (Hirzel et al., 2002; Phillips et al., 2006). All locations in the background where the species did not occur are considered as available and unused conditions. In theory, presence-background methods distinguish between suitable and unsuitable habitats. These methods do not provide the probability of finding the species in a given place, but an index of habitat suitability (i.e. the quality of the habitat for the survival and persistence of the species; Sillero, 2011), which is specific to each modelling method (Acevedo et al., 2012). In general terms, habitat suitability does not assure the presence of the species, nor habitat unsuitability assure the absence of the species. Presence-background modelling methods include Ecological Niche Factor Analysis (ENFA; Hirzel et al., 2002), Genetic Algorithm for Rule-set Production (GARP; Stockwell and Noble, 1992), and Maximum Entropy (Maxent; Phillips et al., 2006; S.J. 2017). Despite being erroneously called presence-only methods as they do not use absences or pseudo-absences, they require both presences and background information to compute habitat suitability. It is also very frequent to consider the background as equal to pseudo-absences, although these concepts are not equivalent. The background is a representation of the entire study area (Phillips et al., 2009; Guillera-Aroita et al., 2014), including presences and absences.

Presence-only methods do not use absence nor background data: species occurrences are the only species data in the analysis. Presence-only methods include envelope algorithms (Booth et al., 2014), i.e. methods that define habitat suitability as an n-dimensional volume (the so-called envelope) enclosed by the environmental limits inferred by species occurrence in the study area. In envelope models, suitable habitats can be delimited with different geometric forms, such as a rectangle, the minimum convex polygon, or more complex polygons covering all (or most of) the occurrences. In theory, these correlative methods might provide the closest approximation to the potential niche, as they may consider suitable any condition in which the species has been observed (Jiménez-Valverde et al., 2008; Sillero, 2011). However, they may fail to identify sinks (Pulliam, 1988), where conditions do not allow self-sustaining populations and presence is only maintained by immigration. Mahalanobis distance explicitly considers the intensity of

occurrence in different conditions, but not the relative availability of habitats in the study area (Clark et al., 1993). Consequently, a widely available habitat that is only partially used can be considered more suitable than a very scarce habitat that is always used. Presence-only methods include Overlap analysis (Brito et al., 1999), Bioclim (Nix, 1986; Booth et al., 2014), Domain (Carpenter et al., 1993), Habitat (Walker and Cocks, 1991), and Mahalanobis distance (Clark et al., 1993).

In summary, presence-absence methods are in theory generally closer to species' distributions, while presence-background and presence-only methods tend to provide larger predictions of the realised niche (see, however, Sánchez-Montes et al., 2019). We can say that conceptually the outputs of the latter are intermediate between the realised and potential niches (Jiménez-Valverde et al., 2008; Sillero, 2011). For example, presence-absence models might be conservative to identify new possible locations prone to invasion by an alien species, as they aim at predicting occupied niches, rather than those suitable areas where the species is not present (see, however, Muñoz and Real, 2006; Pereira et al., 2020). How close a model is to the species distribution will depend on the biases in the species records, presences, and absences (see below).

3.3. Hybrid models

A third and less explored category involves process-orientated models, which estimate occupied distribution areas and associated environments from assumptions about niche dimensions, biotic interactions, and dispersal abilities (Heisey et al., 2010; Peterson et al., 2015). Thus, some models can be clearly placed as correlative or process based (Dormann et al., 2012), while most models usually fall somewhere in between that continuum. In this sense, unlike integrated models where correlative and process-based modelling strategies are fitted simultaneously to data (e.g. demography within suitable habitats; Pagel and Schurr, 2012), hybrid models use process-derived explanatory variables, such as dispersal, to fit later correlative models (Zurell, 2017; Enriquez-Urzelai et al., 2019).

4. How to build correlative ecological niche models

First, we should identify the "target" ecological niche that will answer the research question of the study, in order to select the algorithm accordingly. If the research question is to forecast the fundamental or potential niches, mechanistic methods should be selected; to forecast the realised niche, correlative methods should be preferred (Sillero, 2011). To analyse the species' observed distribution to understand why the species occurs in some places and not in others, due to habitat suitability and other factors such as history, anthropogenic effects, population dynamics or biotic interactions, presence-absence methods could be the most helpful. However, to identify habitat suitability for a given species, i.e. those locations that fit the environmental requirements of the species in the study area, presence-background or presence-only methods may be preferable.

Then, the modelling process will be composed of four stages (Table 3): preparation of all necessary data, calculation of models, evaluation and validation of models, and application of models. Data preparation is usually the most time-consuming.

As any statistical method, ENMs make some assumptions (Guisan and Thuiller, 2005; Wiens et al., 2009; Gallien et al., 2012; Anderson, 2012):

- 1) The species is in equilibrium with the environment, i.e. all suitable habitats are occupied by the species. This rarely happens, as the capacity of the species to occupy habitats is limited mainly by its dispersal characteristics and biogeographic history (i.e. suitable habitats may occur beyond a geographical barrier which is impossible to cross for the species). Therefore, what should be guaranteed

Table 3

List of items to consider when calculating a correlative ecological niche model, their impacts, and possible solutions.

	Aspects to consider	Their impact(s)	Possible solution(s)
Study area	How big? Does it include the whole species' range?	It affects model predictions and measures of goodness of fit. Models on partial distributions may not capture the whole range of conditions in which species can survive.	<ul style="list-style-type: none"> - Use biogeographical regions - Avoid areas where the species cannot disperse - Avoid areas where the frequency distribution curves of the values of critical environmental variables are truncated
	Errors?	Most typical errors include erroneous identifications, lack of one or both coordinates, coordinates set as zero, switched coordinates, positional errors, and duplicated records. The modelling algorithm will use these records as long as they fall inside the study area, thus, leading to biased niche estimates.	<ul style="list-style-type: none"> - Check occurrence points by plotting and visually inspecting species records on a map - Remove duplicated records - Remove occurrences without coordinates
Occurrence data	Are all points inside the study area?	If many occurrences are available this will have a negligible effect. In contrast, when only a few presences are available, this could utterly change niche estimates.	<ul style="list-style-type: none"> - Change limits of the study area to include all points - Move points to the nearest valid pixel
	How many species records?	It is one of the largest sources of uncertainty in prediction accuracy.	<ul style="list-style-type: none"> - Check algorithm-specific requirements - Check how accuracy changes with increasing numbers of records - In general, the accuracy of the models increases until an asymptote is reached, frequently around, 20–30 presence records
Absence/ Pseudoabsence/ Background data	Is it biased?	Geographic biases (e.g. partial distributions) provide weak results when the aim is to model the realized niche of the species. Environmental biases can result in niche estimates with under-represented subsets of the environmental space. The spatial location of “absences” will determine the results of the model.	<ul style="list-style-type: none"> - Reduce record clusters through a process called thinning or filtering of records - Select a background / pseudo-absence sample with the same bias than the presence data - Report sampling bias if possible (e.g. Maxent)
	How to choose absences/ pseudoabsences/ background?		<ul style="list-style-type: none"> - Real absences are only available through fieldwork. - Species atlases can be used to select (pseudo-) absences from grid cells without presences. - Even with intensive and even sampling, some secretive species might not be found
Environmental data	What is being modelled with different absence choices?	By locating “absences” in different regions (e.g. using different strategies) the model captures different subsets of the niche of any given species.	<ul style="list-style-type: none"> - Species' realised distribution: it is captured when absences are located in unoccupied areas due to biotic factors. - Species' potential distribution: it is obtained when absences are located in unoccupied regions due to historical factors.
	How many variables?	Although model accuracy improves with increasing numbers of variables, collinearity also increases.	<ul style="list-style-type: none"> - The number of variables should be less than the number of records. - As a rule, the fewer the species records, the fewer the variables introduced in the model. - For regression-based algorithms the recommended number is: $k = (n-50)/8$ or $k = n-104$ (k is the number of variables and n the number of species records). - For the remaining algorithms no more than $n/10$ predictors are recommended.
	Are the environmental variables correlated?	On the one hand results will be over-fitted and on the other response curves for each variable will not represent exclusively that variable, but its interaction with other correlated variables.	<ul style="list-style-type: none"> - Excluded variable with a correlation higher than (-) 0.7 or (-)0.8. - Exclude variables with a variance inflating factor (VIF) value higher than 10. - Perform a PCA a use (orthogonal) PCA axes as predictors. This, however, greatly hinders biological interpretation. - Select the environmental variables based on biological meaning, given the ecology of the species of interest.

is that the species is in pseudo-equilibrium, i.e. the species occupies all suitable habitats where it can disperse. Thus, the study area should not include areas where the species cannot disperse (Anderson and Raza, 2010). However, Chamorro et al. (2020) demonstrated that models based on fuzzy set theory can measure the spatial and temporal disequilibrium between a species' breeding area and its climatic favourability for breeding. Although a species is in equilibrium with all the combinations of forces resulting in the current configuration of its distribution, including historical ones, it does not have to be in equilibrium with each one of them. Models can assess how the changes in particular favourable conditions attract the species and lead to changes in species distributions (Chamorro et al., 2020).

- 2) The biases in the modelling system are minimal (Fithian et al., 2015). The sampling design should not be focused on particular regions of the study area, habitat types, or activity periods. Unfortunately, most

species surveys are biased to some degree, as species occurrences are normally collected near roads or at easily accessible locations. Also, the sampling effort should be homogeneous across the study area. The selection of variables for modelling may also be biased: some specific variable is selected actively or some important variable for the species is lacking. Four types of biases should be avoided within the study region (Anderson, 2012):

- Niche Space Assumption: the study area holds all the adequate environmental conditions for the species occurrence.
- Dispersal/demographic Noise Assumption: the species occupies all suitable areas and is not excluded from some areas due to dispersal, establishment, and persistence.
- Biotic Noise Assumption: the species is not excluded from suitable areas due to biotic interactions.
- Human Noise Assumption: the species is not excluded from suitable areas due to human activities.

- 3) All variables included in the model are related to species occurrence. This is essential in correlative algorithms: if the variable does not have a relationship with the species occurrence (e.g. it does not show spatial variability), it should not be included in the model. The relationship between the species presences and variables can be analysed with Pearson or Spearman Rank correlations.
- 4) The species niche is conserved across space and time. When a species is modelled in a given study area, we are assuming that the relationships established between the environmental variables and the species are constant across the study area. When a model is projected to another scenario in time or space, we are assuming that the conditions defined by the model will be equally maintained. If the species has adapted to the new scenario, it may have created new environmental relationships, and the model will fail to predict them.

4.1. Preparing the data

4.1.1. Available software for modelling

Several computer programmes are available to run correlative ecological niche models (see Supplementary Text S1), such as Maxent (Phillips et al., 2006; 2017), Biomapper (Hirzel et al., 2002), ModEco (Guo and Liu, 2010), OpenModeller (Enrique et al., 2009), and R (RCoreTeam, 2019). R is the software with more modelling applications (see Supplementary Text S2).

4.1.2. Selection of the study area

Once you have determined how to address your objective, it is necessary to define the study area. Models including the entire species' range usually provide better results than models of partial distributions (Barbet-Massin et al., 2010; Raes, 2012; Carretero and Sillero, 2016), as the study area encloses all the environmental variability of the species range. Models from partial distributions can fail to forecast the complete species range (Carretero and Sillero, 2016). However, partial distributions may be better to identify other distribution constraints, which may differ in different parts of the distribution range (Castro et al., 2008; Martínez-Freiría et al., 2008; Vale et al., 2015). In any case, the exact definition of the limits of the study area is not straightforward (Jarnevich et al., 2017). The study area can strongly affect model results even if Jarnevich et al. (2017) did not find noticeable differences between four background methods (rectangle around presence points, minimum convex polygon, a binary kernel, and a continuous kernel) in a specific case study. The study area should cover relevant ranges for the variables that are important for the modelled species. This can vary between regions, as it depends on the spatial arrangement of the relevant variables relative to species occurrences. This subject has not yet received enough attention (i.e. Acevedo et al., 2012).

The size of the study area also affects metrics of model performance. The discrimination capacity of the models (i.e. the ability to correctly distinguish presence from absence localities; see Section 3: Validating the model), for example, typically increases with the size of the study area. This is mostly because larger areas tend to include absences more ecologically distant from presences, which are easier to distinguish (Lobo et al., 2008; VanDerWal et al., 2009). The calibration capacity (how well the model predicts the gradual occurrence probability) decreases with the size of the area because larger surfaces tend to include areas far away from presence locations, which are not relevant to infer the interaction between the species and the environment (Acevedo et al., 2012).

It is important therefore to define the extent and limits of the study area based on scientific criteria. For this, three main criteria can be followed: 1) use biogeographical regions (Comrie and Glenn, 1999; Sillero, 2010); 2) avoid areas where the species cannot disperse (Anderson and Raza, 2010); and 3) avoid areas where the frequency distribution curves of the values of critical environmental variables are truncated (Guisan and Thuiller, 2005). Modelling inside a

biogeographical region might be the simplest way, as they work as relatively contained systems (Sillero et al., 2009; Sillero, 2010; Barve et al., 2011). Another option is to exclude those areas where the species cannot disperse (Anderson and Raza, 2010), not only in current times but also in the past (Barve et al., 2011). In this way, we increase the probability of the species being at equilibrium inside the study area, and thus the quality of model predictions. But how can we define the areas where the species cannot disperse? One option is to perform a preliminary spatial model over a large study area, to identify those areas outside the species reach (Barve et al., 2011; Acevedo et al., 2012). Another alternative is to reduce the study area to a buffer around the species records (Phillips et al., 2009; Báez et al., 2020). The size of the buffer is somewhat subjective, and the best distance should be selected from different trials or using information about the species' dispersal capacity, including long-distance dispersal events (Barve et al., 2011). When the frequency distribution curves of the values of critical environmental variables are truncated, the extent of the study should be increased, as it is not including all environmental variability of the species (Guisan and Thuiller, 2005).

Species distributions are affected by biotic interactions, human actions, or dispersal capacities. Consequently, the study area should contain all environmental conditions required by the species and without significant bias (Anderson, 2013).

4.1.3. Obtaining occurrence records: data sources

Species occurrence records are indispensable to calculate a correlative niche model. Presences and, depending on the algorithm, absences are necessary. There are several sources of species presence records. amongst the most frequent sources are distribution atlases (Sillero et al., 2014), online databases such as the citizen science-based GBIF-Global Biodiversity Information Facility (Yesson et al., 2007), and field-collected data. Recording species locations with a GPS will avoid errors and save time (Sillero et al., 2005).

The spatial location of absences will determine the results of the model (Lobo et al., 2010). There are three types of absences (Fig. 1) according to the type of information they provide (Lobo et al., 2010):

- 1) **Contingent absences:** They correspond to environmentally suitable areas not occupied due to biotic, historical and dispersal reasons. Contingent absences are outside the occupied niche but inside the realised or potential niches. They occur at the overlap of two of the three circles represented in the Biotic-Abiotic-Movement (BAM) diagram (Fig. 1). These absences are informative of historical or dispersal constraints, and they are paramount to distinguish between the occupied and the realised niches (Pearson, 2007). They are more likely in localities with suitable environmental conditions that are spatially distant from current populations.
- 2) **Environmental absences:** They correspond to environmentally unsuitable areas. Environmental absences are outside both the realised and potential niches (Fig. 1). These absences define the niche limits. If all absences were of this type, even presence-background methods could infer these environmental conditions directly from presence records. They are likely to correspond to locations where environmental conditions are vastly different from those found in the localities of presence. If the study area is exceptionally large, a huge proportion of the pseudo-absences may be in fact environmental absences.
- 3) **Methodological absences:** These are false absences created by errors in species surveys. Methodological absences can be located inside both the realised and the potential niches. Most modelling methods can correctly deal with methodological absences, as far as they are the result of an insufficient but well-distributed sampling effort. Indeed, they infer the ecological requirements of the species according to the variation in the proportion of used habitat in the dataset, so unbiased false absences do not affect this variation. On the other hand, spatially or environmentally biased false absences

can introduce errors in the models. They will reflect the distribution of survey effort, rather than the distribution of species. Unbiased methodological absences are more likely in the environmentally suitable localities placed near the known presence points, while biased absences could be anywhere. It is important to highlight that a bias in the absences also implies a bias in the presences, as false absences are true presences that are missing from the dataset.

Consequently, pseudo-absences must be created with caution, as their placement may strongly affect the results of the models (Chefaoui and Lobo, 2008; VanDerWal et al., 2009; Barbet-Massin et al., 2012). Pseudo-absences can be created from four sources (Barbet-Massin et al., 2012; Senay et al., 2013; Hertzog et al., 2014):

- 1) **Allopatric species:** from the presence of sister species with an excluding distribution. Niamir et al. (2016) noticed that the probability of an undetected presence of the target species at localities in which other taxonomically related species were recorded is lower than in those localities without such recordings. Phillips et al. (2009) also explored the use of the occurrences of a target taxonomic group as biased background data, as an entire target group of species observed by similar methods will likely share similar bias. They found that target-group background improved average performance for all the modelling methods they considered.
- 2) **Environmental absence areas:** from verified climatically unsuitable areas, e.g. using a model calculated with a mechanistic or a presence-background algorithm (Engler et al., 2004).
- 3) **Random:** random points across the study area. These pseudo-absences are similar to the background but excluding locations with presences. This is the reason why presence-background methods often provide results very similar to models calculated with (many) random pseudo-absences.
- 4) **Presence areas:** within and/or outside a specified buffer distance around presence locations, so that pseudo-absences are neither too close nor too far from recorded presences (Phillips et al., 2009). Pseudo-absences should be close enough to represent areas to where the species could have dispersed (if environmental conditions were favourable) and to reflect the same spatial bias of the presences, therefore avoiding regions that were not surveyed; but not so close that they fall within the home range of the recorded presences.

The best source of pseudo-absences depends on the modelling technique (Barbet-Massin et al., 2012; Kanagaraj et al., 2013): random selection should be preferred when using classical regression methods like generalized linear models and generalized additive models, and geographically and environmentally stratified random selections for machine-learning techniques like boosted regression trees, classification trees, and random forest. When biases are present in species surveys, selecting pseudo-absences considering these biases provides better results (Hertzog et al., 2014). However, methods that incorporate environmental stratification of the background are generally better (Iturbide et al., 2015).

Model results are more affected by sources of bias and by the number of pseudo-absences than by the distribution of pseudo-absences (Lobo and Tognelli, 2011). The best number of pseudo-absences depends on the modelling technique (Barbet-Massin et al., 2012; Kanagaraj et al., 2013): recommendations include using a large number (e.g. 10,000) of pseudo-absences with regression techniques (e.g. generalised linear models and generalised additive models); averaging several runs (e.g. 10) with fewer pseudo-absences (e.g. 100) for multiple adaptive regression splines and discriminant analysis; and using the same number of pseudo-absences as the number of available presences (averaging several runs if this number is small) for classification techniques such as boosted regression trees, classification trees and random forests. Despite all studies analysing how to create pseudo-absences, true observations of absences always provide better results (Wisz and Guisan, 2009).

4.1.4. Checking species records for errors

Published occurrence datasets are not free of errors (Samy et al., 2013; Beck et al., 2014; Sillero et al., 2014; Meyer et al., 2016). Errors in coordinates must be always checked by plotting and visually inspecting species records on a map. There are several other possible sources of error, such as:

- 1) **Species may be erroneously identified.** If the modeller is not familiar with the taxonomic group, these errors are almost impossible to spot, especially for widespread species. Errors can be detected when the species is reported far from its known or plausible distribution range. Species misidentification may strongly influence ENM results (Costa et al., 2015; Aubry et al., 2017). Indeed, false presences affect models more than false absences (Fernandes et al., 2019).
- 2) **Species records may lack spatial coordinates (one or both)** or their values may be erroneously zero. This is very frequent in herbaria and museum records, but also in online databases such as GBIF. If references to localities are available, it is possible to associate coordinates to those localities. Online gazetteer services provide coordinates for the localities. However, if coordinates are obtained from originally coarse locality data, a similar accuracy should be maintained in the geo-referencing for all records: this spatial accuracy should be concordant with the spatial resolution of the environmental variables used for calculating the models. This means that the locality reference needs to be accurate, unequivocal and appropriately sized. Localities such as a country, or even a county can be misleading unless they are used on an accordingly latticed study area (Chipperfield et al., 2020).
- 3) **X and Y coordinates may be switched.** Geographical coordinates are traditionally presented as latitude-longitude, that is, as Y-X. Thus, they can be introduced in the wrong order. As the resulting presences can be located very far from their correct positions, this error can usually be spotted when mapping them.
- 4) **Coordinates may have positional errors** (e.g. errors in reading coordinates). This error is almost impossible to detect if additional information is not provided. For example, GBIF data usually include a column with a measure of positional accuracy, although it is not always filled in. Positional uncertainty can lead to a reduction in prediction accuracy in ENMs, depending on the spatial autocorrelation in predictors and the levels of positional uncertainty. The reduction in prediction accuracy is small when the range of spatial autocorrelation in predictors is larger than three times the standard deviation of the positional error (Naimi et al., 2011). Some methods such as boosted regression trees and Maxent, can be more robust to positional errors in presence coordinates (Graham et al., 2007).

Finally, some modelling methods are sensitive to the abundance of species records per pixel (e.g. Maxent), while habitat suitability and abundance are not directly related (Dallas and Hastings, 2018). Thus, it is necessary to exclude these pseudo-replicates or duplicated records from the analysis (Phillips et al., 2006). Some modelling programs do this automatically.

4.1.5. Checking that all points fall inside the study area

Once the study area is defined, it is necessary to check if all the species records are included in the study area. Some points will often be placed just outside the limits of the study area or located in a region for which the variables have no values. For example, when modelling species with coastal presences, some points may be placed in “no data” pixels (i.e. on the wrong side of the coastline), even if they were correctly recorded with a GPS, as the limits of the continents used for mapping the variables may not be very precise. These points cannot be used for modelling. Some modelling applications are not able to work with missing values, while others exclude these points automatically. If a high number of records with good coverage exist, losing some points

might not constitute a problem. In other cases, this is not acceptable. Hence, to avoid losing a point, we can either manually move it to the nearest valid pixel (Sillero and Carretero, 2013), or add a pixel covering the point and assign it the value of the nearest valid pixel (which can be done with some GIS software or with R).

4.1.6. Number of species records

Sample size is amongst the largest sources of variability in prediction accuracy (Buisson et al., 2010; Tassarolo et al., 2014; Thibaud et al., 2014). How many species records are necessary to run an ENM? There is no fixed rule, as some algorithms need more records than others. For example, envelope algorithms such as Bioclim can run with as few as three records. Obviously, the prediction may be very large or very small and it will depend completely on the records included. Maxent generally provides good results with very small sample sizes (Papes and Gaubert, 2007; Hernandez et al., 2008): three records for specialist species and 13 for generalists (van Proosdij et al., 2016). The necessary sample size depends on the extent of the species distribution: widespread species will need more samples, as their ranges encompass higher environmental variability. Conversely, specialist species require fewer points to get good predictions, as all their environmental variability is restricted to particular environmental conditions (Seoane et al., 2005; Hernandez et al., 2006; Tsoar et al., 2007; van Proosdij et al., 2016). For generalised linear models (GLM) and related methods, a recommended number of records is $50 + 8k$ or $104 + k$, being k the number of variables (Field et al., 2012). According to this, a GLM will need at least 58 records (presences and absences) for only one predictor variable. Another criterion is provided by Franklin (2010): a minimum of ten presence records per predictor variable. In general, model accuracy increases with sample size until an asymptote is reached (Hirzel and Guisan, 2002; Stockwell and Peterson, 2002; Hernandez et al., 2006; Wisz et al., 2008), often between, 20 and 30 presence records (Stockwell and Peterson, 2002).

When the sample size is very small, the determination of variables' importance can be affected (Mitchell et al., 2017). An option would be to use Ensembles of Small Models (ESM), modelling all possible combinations of only two variables (Breiner et al., 2015, 2018). As for validation, a good option would be the leave-one-out method, i.e. the test dataset is composed of only one occurrence record set aside that changes in each model replicate (Hernandez et al., 2008; Shcheglovitova and Anderson, 2013).

4.1.7. Effect of prevalence

Prevalence is the proportion of presences in the dataset, i.e. the number of presences divided by the total number of modelled localities (Jimenez-Valverde and Lobo, 2006). Prevalence strongly affects model results (Hosmer and Lemeshow, 2000; Jimenez-Valverde et al., 2009; Royle et al., 2012; Merow et al., 2013). When prevalence is highly biased (e.g. very low or very high proportion of presences), it also influences the suitability values. However, probability may be transformed into favourability (Real et al., 2006), which represents the response of the species to the environmental conditions irrespective of species prevalence. Therefore, the best option is to use the maximum possible number of relevant absences (Jimenez-Valverde and Lobo, 2006; Meynard and Quinn, 2007; Jimenez-Valverde et al., 2009), and then convert the predicted probability to favourability (e.g. Acevedo and Real, 2012).

4.1.8. Sampling design

The type of sampling design has an intermediate influence on models, although its absence may have strong effects (Tassarolo et al., 2014; Thibaud et al., 2014). For presence-background algorithms, the best sampling strategy tends to be Equal-stratified, followed by Regular, Random and Proportional-stratified; for presence-absence algorithms, the Regular and Equal-stratified strategies are the best, followed by proportional stratified and Random (Hirzel and Guisan, 2002).

4.1.9. Checking for biases in species records

Species records can be geographically or environmentally biased. Both biases can strongly influence the results of the model (Phillips et al., 2009; Hijmans, 2012; Fourcade et al., 2014). Geographic biases such as partial distributions provide weak results when the aim is to model the realised niche of the species (Barbet-Massin et al., 2010). However, biases are often originated during sampling, creating artificial clusters of points around those areas that are most easily accessible for field workers (Barbosa et al., 2012; Merow et al., 2013). Those clusters should be mitigated through a process called thinning or filtering of records (Veloz, 2009; Boria et al., 2014; Varela et al., 2014). Filtering may be better than restricting the background area (Kramer-Schadt et al., 2013). Another approach is to select a background or pseudo-absence sample that reproduces the same bias or clustering pattern as the presence data (Phillips et al., 2009; Merow et al., 2013). Maxent also allows including a raster map representing sampling bias (Phillips et al., 2009), and sampling effort can also be included as a covariate in regression-based models (see section below). Filtering species records and biased backgrounds can be used at the same time to reduce bias effects.

How to reduce the clustering level on species locations? There is not a standard procedure. The most widely used method to reduce the clustering of species occurrences is to manually remove points until no artificial clusters are apparent and the species distribution looks homogeneous. However, spatial statistics can help with this issue. The clustering level can be measured with the nearest neighbourhood index (NNI; Clark and Evans, 1954). The NNI measures the spatial distribution from 0 (clustered pattern) to 2.15 (regularly dispersed /uniform pattern): values of 1 represent randomly dispersed patterns. Therefore, any time a point or a group of points is deleted from the species' locations, the NNI can be measured until it reaches the value of 1. However, deciding what points are deleted is a subjective procedure. Currently, the SDMtoolbox package for ArcMap (Brown, 2014) and at least three R packages help to filter species records, spThin (Aiello-Lammens et al., 2015), usdm (Naimi et al., 2014), and spatialEco (Evans, 2021). See Supplementary Text S2 for a list of available R packages for ENMs. Nevertheless, we should take care not to remove clusters that reflect areas where species occurrences (not just sampled localities) are actually more frequent. Although filtering is necessary when artificial (i.e. sampling-driven) clusters are present in the species dataset, this process reduces the number of points available for modelling. Thus, the number of points deleted must be balanced to obtain a homogeneous intensity of species records that still allows good model performance and accuracy.

Filtering is often confused with "reducing autocorrelation". Spatial autocorrelation is a property of the First Law of Geography, i.e. Tobler's Law (Tobler, 1970): "Everything is related to everything else, but near things are more related than distant things". This means that closer locations will have similar values. Environmental variables, like temperature and precipitation, will generally be more similar between two near locations than between two distant ones. This relationship between the values of a variable in different locations is the spatial autocorrelation, a "special case of correlation in space, when there is a relationship between nearby spatial units of the same variable" (Getis, 2007, 2008). Therefore, spatial autocorrelation is an important property of both species' distributions and the environmental variables, as it reflects the spatial structure of their distribution (Legendre, 1993). Artificial clustering due to uneven sampling should be reduced on the species' locations, but the remaining spatial autocorrelation in both presences and environmental variables is a natural property, rather than a problem (Legendre, 1993). Notwithstanding, spatial autocorrelation should be the same for training and test data (see below). Spatial autocorrelation may inflate the estimation of variable importance when the species' response to the environmental gradients is linear. This inflation effect is larger when the species' environmental preferences coincide with the dominant environmental conditions in the study area (Harisena et al., 2021). Autocorrelation can be measured with Moran's I (Moran, 1950),

which ranges from -1 (negative autocorrelation) through 0 (no autocorrelation) to $+1$ (positive autocorrelation).

If there are biases in sampling effort, even after filtering the species records, a biased raster can be included in Maxent (Phillips et al., 2009) or as a covariate in regression-based models (Báez et al., 2007). Maxent will decrease habitat suitability in those areas with biases. The bias file is a representation of the relative sampling intensity across the whole study area. The sampling bias file needs to be a continuous or categorical raster grid (ASCII format), with any value above zero. For instance, if roads are the source of sampling, we may calculate a bias raster as a map of proximity to roads. Another alternative is to compute higher weights around occurrence points. The bias file can also be a categorical raster representing sampled and non-sampled areas. However, models with biased rasters are not always better than models without them (Syfert et al., 2013). Furthermore, proximity to roads may not represent correctly the sampling bias if roads are not climatically biased (Kadmon et al., 2004). Another approach is to select a background or a pseudo-absence sample showing the same spatial or environmental bias as the sample of presences, or as the sample of a target taxonomic group which is surveyed by similar observers with similar methods (Dudík et al., 2005; Phillips et al., 2009; Merow et al., 2013). In this way, both presences and background or pseudo-absences have the same clustered or biased pattern, so the model is less misled by the sampling bias, and it focuses on the differences between the presence data and the equally biased background or pseudo-absence data (Ferrier et al., 2002; Phillips et al., 2009). This procedure cannot be used with presence-only methods.

4.1.10. Splitting between training and test records

Ideally, the model should be validated with an independent dataset. Consequently, the origin of that dataset should be different, collected with different sampling methods, by different people, or during a different period. As an independent dataset would have a different source of biases, it can efficiently validate the model. Unfortunately, independent datasets are seldom available. An alternative is to split the available species records between training and test records, where training records will be used to calculate the model, and test records to validate it (even if the final model, e.g. for projection or explanation purposes, is built on the complete dataset in the end). It is necessary to ensure the independence of training and test datasets, meaning that they should not have the same spatial autocorrelation structure (Peterson and Soberón, 2012). A way of creating two databases not affected by spatial autocorrelation is to choose the points separated by enough distance to minimise autocorrelation. This distance can be determined by a semi-variogram, which represents autocorrelation against distance. Thus, the semi-variogram will indicate the distance above which the points are not related to a specific variable in the geographical space, although this distance can be different for different variables. Recent methods for this purpose include block cross-validation, implemented in the 'blockCV' R package, which can generate spatially or environmentally separated folds for cross-validation of models (Valavi et al., 2019).

4.1.11. Obtaining environmental data

ENMs have been traditionally limited to the use of abiotic variables, such as climate, elevation, land use, or soil type. Many different sources of these data are freely available (see Supplementary Text S3), which has contributed to an explosion of ecological niche modelling studies in the last decades. There are two main types of data: environmental ground data and remote sensing data.

Ground data are collected with fieldwork or by ground-based stations, such as meteorological stations. These data are then transformed into interpolated surfaces. Packages such as 'dismo' allow interpolating data from ground stations. Examples of such datasets are WorldClim (Hijmans et al., 2005; Fick and Hijmans, 2017), CHELSA (Karger et al., 2017), EuMedClim (Fréjaville and Benito Garzón, 2018), CRU TS (New et al., 2002; Harris et al., 2014), CliMond (Kriticos et al., 2012),

Terraclimate (Abatzoglou et al., 2018), MERRAclim (Vega et al., 2017), or Bio-ORACLE (Tyberghein et al., 2012). Distinct portals provide datasets for past (WorldClim, Paleoclim - <http://www.paleoclim.org/>) and future (WorldClim, CHELSA) climate scenarios for land environments, and Bio-ORACLE for marine environments. Some wide-scale variables may show inaccurate values on a local or regional level (Bedia et al., 2013) or large discrepancies amongst different climate models (Varela et al., 2015), so their use is not recommended, particularly when models are to be extrapolated in time.

When using elevation and its derivatives (slope, aspect, topographical indexes) as variables, some issues must be considered. As topographical variables are generally derived using distances, they should be calculated in projected coordinate systems (in meters). Geographical coordinate systems (e.g. WGS84) are spherical and thus distances correspond to arc-distances and not Euclidean distances (Burrough and McDonnell, 1998). Aspect defines the cardinal position of any pixel: north, south, east, west, and intermediate combinations (Burrough and McDonnell, 1998). Aspect is measured in degrees, thus 360° and 0° represent the same value, north. As modelling algorithms will assume that those two values represent very different positions, a better option is to use Northness ($\cos(\text{aspect})$) or Eastness ($\sin(\text{aspect})$) which range between 1 (full north/east respectively) and -1 (full south/west respectively) (Guisan et al., 1999). Aspect needs to be calculated in radians (degrees $\times \pi/180$).

Remote sensing (RS) data offer continuous and cost-effective measures of both abiotic and biotic factors across space and time, hardly quantifiable by other means (Cord et al., 2013; Pasetto et al., 2018; Arenas-Castro et al., 2018; Arenas-Castro and Sillero, 2021). While recent sensors provide better and more diverse measurements at finer spatial, temporal and spectral scales (e.g. Sentinel satellite with 10–20 m pixel size; Berger and Aschbacher, 2012), many other sensors (e.g. MODIS) of free accessibility and with global coverage provide reliable products from local to landscape scales, and for marine to terrestrial ecosystems (He et al., 2015). The most widely used remote sensors and tools in ENMs are spaceborne and airborne systems, targeting both passive (i.e. using the reflected sunlight) and active sensors (i.e. using an artificial light source). Some examples of passive sensors are AVIRIS (20 m pixel size), Landsat (15 to 60 m pixel size) or MODIS (250 m to 1 km pixel size), while active sensors include LiDAR and microwave RADARS. However, because not all species or habitats can be detected by remote sensors, the use of lightweight unmanned aerial vehicles (UAVs), or drones, can be a particularly cost-efficient way to collect data for model calibration and validation (Anderson and Gaston, 2013), although they typically cover smaller extents and their use can be limited by local regulations. In Supplementary Table S1 we provide examples of potential abiotic and biotic RS predictors that have been used widely in ENMs.

4.1.12. Number of environmental variables

Prediction accuracy on model training data generally improves with increasing numbers of variables, until an asymptote is reached (Peterson and Cohoon, 1999; Cumming, 2000). However, the number of variables should be smaller than the number of species records, and an excessively large number of variables will make the model overfit to the training data, therefore decreasing its capacity to predict outside the modelled sample. For regression-based methods, the maximum number of variables should be $k=(n-50)/8$ or $k=n-104$, k being the number of variables and n the number of species records (Field et al., 2012). For the remaining methods, there is no generally agreed rule. Franklin (2010) recommended no more than $n/10$ predictors, being n the number of occurrence records. The fewer species records are available, the fewer variables should be included in the model. In the case of the Bioclim algorithm, the size of the rectangle used to define the model decreases with the number of combinations of variables (Beaumont et al., 2005; Booth et al., 2014).

4.1.13. Clipping variables

Variables should be clipped or masked by the study area extent, to avoid inaccessible or unsurveyed areas being included in the model. If the variables have different spatial resolutions, maps should normally be aggregated to the coarsest spatial resolution. 'Aggregate' is a GIS function that groups rectangular areas to create larger pixels by applying a mathematical function (mean, maximum, minimum, sum, range). Thus, the resulting rasters will have fewer pixels. The correct way of downscaling climate variables is complex (Kriticos and Leriche, 2010), requiring dynamical downscaling for regional models (Rummukainen, 2010). As an alternative, microclimatic layers can be derived using mechanistic models (Enriquez-Urzelai et al., 2019; Kearney and Porter, 2020).

If raster maps are to be used directly as input data (rather than a table of values) but they have a different origin of coordinates (the coordinates of the bottom-left pixel), they will normally need to be resampled so that they perfectly match. 'Resample' is a GIS function that transfers values between non-matching rasters (in terms of origin and resolution). Therefore, the number of pixels does not change. The new properties of the pixels can be transferred from the nearest pixel, by bilinear interpolation, or by a cubic convolution.

4.1.14. Checking the correlation between environmental variables

Most modelling algorithms are sensitive to high levels of correlation amongst predictor variables (Dormann et al., 2013; De Marco and Nóbrega, 2018; Feng et al., 2019a). In general terms, a predictor variable is more likely to be erroneously selected (higher risk of a type 1 error) when high correlations exist between variables. In the case of ENMs, high correlations can have two main consequences (Dormann et al., 2013; De Marco and Nóbrega, 2018; Feng et al., 2019a): 1) results will be over-fitted, and 2) response curves will not be independent, i.e. the response curve of a variable will not represent exclusively that variable, but it will include interactions with other correlated variables. Yet, everything in nature is correlated: temperature and precipitation are dependant on each other, they are influenced by elevation, vegetation depends on those variables, and so on. Therefore, it is practically impossible to reduce correlation to zero in any natural set of variables. Normally, variables with a correlation stronger than a certain absolute value, such as (-)0.7 or (-)0.8, should be excluded before modelling (Dormann et al., 2013). Lower absolute thresholds of correlation may yield a small number of variables, hampering to obtain a good model if too many variables are discarded. An alternative is to transform the set of variables using principal component analysis (PCA; Vale et al., 2016; Pérez i de Lanuza et al., 2018). PCA summarises the environmental variables into orthogonal, i.e. completely uncorrelated factors, and these factors can be used to model the species' niche. However, the interpretation of the model will be more difficult. The most important factor for the model will likely be the first component, as it is the factor explaining most of the variability between the predictors. Then, the contributions of each variable to the first component must be analysed to have an idea of which variables are driving the species' distribution, and in which directions. The same for the rest of the components. However, if the majority of included variables is not relevant for the modelled species, the first component might not be the most important for that species, but simply describe most of the variation in the candidate variables.

Using any variable from a pair of highly correlated variables provides similar results. Thus, when selecting a variable, authors should opt for the one with a more relevant biological meaning, such as its capacity to explain the species distribution, given its ecology (Petitpierre et al., 2017). Correlation can be measured by parametric (e.g. Pearson) or non-parametric (e.g. Spearman) coefficients (Field et al., 2012). Parametric coefficients assume that the residuals follow a normal distribution, which can roughly be the case when the study area encompasses an adequate range of environmental variation, but it is worth checking it. The variance inflation factor (VIF) measures the correlation of each

variable with a combination of all the other variables in the model together (Zuur et al., 2010; Dormann et al., 2013). Thus, VIF measures multicollinearity, or the non-independence of predictor variables. Nevertheless, the parameters will continue to be unbiased, consistent and efficient. When forecasting the species niche in a different region or period, the model will be efficient only if the same multicollinearity pattern persists in the projected conditions. High multicollinearity exists with a VIF value higher than 10 (Kutner et al., 2004). More recent studies claim that the limit should be stated between, or below as possible as, 3–5 even when standard multiple regression and machine-learning approaches are usually applied, as the case of ENMs (Dormann et al., 2013). However, it is not necessarily good practice to exclude a variable just because it has a high VIF, i.e. a high proportion of variation already accounted for other variables in the dataset; it could be better to exclude those other variables instead, which would also reduce the VIF. Indeed, the selection of variables to be included in the model should be based on a trade-off between how relevant the variable may be and the expected consistency of that relevance over time and space where the models are fitted.

4.2. Calculating the model

4.2.1. Obtaining mean and standard deviation of replicates

If the algorithm selected for running the models has a probability or a random component (e.g. randomly generated pseudo-absences), any time the models are calculated, the results will be slightly different. Hence, it is necessary to run the models several times to get a sample of model variability and get the mean and standard deviation from that set of models.

4.2.2. Projecting models to other scenarios in time and space

Models can be transferred to other space and time scenarios. The algorithms apply the formula describing the species' niche to another set of environmental variables, which can correspond to the past (e.g. WorldClim offers past climate scenarios for the Holocene, the Last Glacial Maximum, and the Last Interglacial Period), to the future (e.g. climate change scenarios for, 2050 and, 2070 in WorldClim), or to another geographical region or resolution scale in the present conditions. Current records cannot be modelled directly against past or future environmental data. Models should be built using contemporaneous data, and then projected to different periods, even if we cannot guarantee that the relationships driving the species' niche nowadays will be the same in the future or were the same in the past. Past occurrence data such as historical records or fossils of well-preserved taxa can be modelled directly with past environmental variables (Martínez-Meyer et al., 2004; Martínez-Meyer and Peterson, 2006; Nogués-Bravo et al., 2008; Nogués Bravo, 2009; Varela et al., 2009; Martínez-Freiría et al., 2016; Chiarenza et al., 2019).

4.3. Validating the model

The validation of ecological niche models is probably the least developed step, particularly in the case of presence-only and presence-background modelling algorithms. The predictive performance of ecological niche models can be validated by three components (Hosmer and Lemeshow, 2000; Pearce and Ferrier, 2000; Leroy et al., 2018):

- **Classification capacity:** the ability of the model to correctly classify occupied sites as suitable or probable or favourable, and unoccupied sites as the opposite, based on a threshold value.
- **Discrimination capacity:** the ability of a model to generally separate or distinguish between occupied and unoccupied sites, regardless of any threshold value.
- **Calibration:** the agreement between predicted probabilities of occurrence and observed proportions of sites occupied. Calibration is the most faithful assessment of the reliability of the models, but it is

not easy to measure in all algorithms, and thus many authors have focused on classification and discrimination metrics.

Classification is measured by cross-classifying observations and predictions of presences and absences in a confusion matrix using a classification threshold, and calculating indices of classification performance, such as sensitivity, specificity, precision, over-prediction, under-prediction, Cohen's kappa or the True Skill Statistic (TSS) (Fielding and Bell, 1997; Barbosa et al., 2013; Allouche et al., 2016). In the case of presence-only or presence-background methods, absences are not available. Here, classification metrics are relative rather than absolute, and it is essential to compare the evaluation metrics with null models (see below). In addition, model predictions represent a gradient that varies continuously, and defining a crisp limit between suitable and unsuitable is difficult and artificial. Although relatively objective thresholds can be defined (e.g. maximum kappa, maximum TSS, equal sensitivity and specificity; Liu et al., 2005; Jiménez-Valverde and Lobo, 2007), the criteria to select a threshold remain rather subjective. As stated in the section 'Application of the model', applying a threshold is not mandatory and implies a loss of information.

Discrimination capacity, as measured with the Area Under the Curve (AUC) of the receiver operating characteristic (ROC) plot (Lobo et al., 2008) (see below), is not related to any particular threshold and it yields the probability that a randomly selected presence will have a higher predicted value than a randomly selected absence. A statistical significance of the AUC may be also obtained (Mason and Graham, 2002). However, a good model is not expected to discriminate beyond the discrimination that the species itself makes in the environment. Therefore, models of generalist species tend to have lower discrimination capacity than those of specialist species, which does not imply that they are worse models.

Goodness of calibration measures how well the estimated probability of presence represents the observed proportion of presences, or the agreement between observations and the continuous predicted values (Pearce and Ferrier, 2000; Jiménez-Valverde et al., 2013; Arelas-Guerreiro et al., 2016). Calibration may be measured with the Hosmer and Lemeshow goodness-of-fit test (Hosmer and Lemeshow, 2000) for the training data; with Miller's statistics (Miller et al., 1991) for the validation data; with the Boyce index, for presence-background methods (Boyce et al., 2002); and with a presence-only calibration (POC) plot (Phillips and Elith, 2010) for presence-only methods. Calibration assumes probability values, not suitability or favourability. Probability values can be derived directly from favourability (Real et al., 2006; Acevedo et al., 2012), but not from suitability (Royle et al., 2012).

4.3.1. Obtaining classification and discrimination metrics

All classification metrics are based on the confusion matrix (Fielding and Bell, 1997; Manel et al., 2001; Liu et al., 2011; Barbosa et al., 2013). amongst those, the most prominent metrics are Sensitivity and Specificity (i.e. the numbers of correctly classified presences and absences, respectively) because they are the basis for the most widely used evaluation metrics in ENMs: AUC and TSS.

AUC stands for the area under the curve of the ROC (Receiver Operating Characteristic) plot (Hanley and McNeil, 1982; Zweig and Campbell, 1993; Bradley, 1997; Fawcett, 2006), which depicts the true positive rate (Sensitivity) against the false positive rate (1-Specificity) at all possible classification thresholds. AUC ranges from 0 to 1, being 0.5 the value for a model with random discrimination. Therefore, any value above 0.5 corresponds to a model with discrimination better than random, and any value below 0.5 corresponds to a model with discrimination worse than random (inverted predictions), probably due to strong biases. The value of the AUC is mathematically equivalent to the Wilcoxon-Mann-Whitney U-Statistic comparing the predicted values for occupied vs. unoccupied sites (Mason and Graham, 2002).

The AUC has been criticised as an overall measure of model performance, as it is dependant on the extent of the study area (Lobo et al.,

2008). If the study area is big enough to have a large proportion of unoccupied habitats, very different environmentally from the occupied area, absences will be more easily distinguished from presences and thus a not-so-good model will still obtain a high AUC. On the contrary, if the study area is too small, with presences and absences in more similar and even overlapping environmental conditions, the AUC will have a low value even if the model is good and predicts correctly the distribution of the species. Let's imagine a model for a species in the Mediterranean area: it is very easy to increase the AUC by adding North Africa to the study area, as the algorithm will be able to determine that the species does not occur for example in deserts. The maximum AUC value in a presence-background algorithm is $1 - a/2$, being a the species prevalence (Phillips et al., 2006; Smith, 2013). For example, if a species occupies 60% of the study region, the maximum AUC will be $1 - 0.6/2 = 0.7$. Unfortunately, a is rarely known: it can be only inferred from the relative extent of the study area occupied by presences and background. In conclusion, the AUC is only comparable between models based on the same study area and dataset (Wiley et al., 2003; VanDerWal et al., 2009).

To avoid unrealistically high AUC values for rare species or large extents, we can use the area under the precision-recall curve (AUC-PR) instead. Precision is the proportion of localities predicted as suitable that are actually occupied, and recall is the proportion of occupied localities predicted as suitable (i.e. the same as sensitivity). As it does not incorporate correctly predicted absences, the AUC-PR is robust to species rarity and, unlike the AUC of the ROC curve, it is not affected by large geographic extents (Sofaer et al., 2019a).

The True Skill Statistic (TSS) is equal to sensitivity + specificity - 1 (Allouche et al., 2006). TSS ranges from -1 to 1, with 0 corresponding to random classification. TSS also increases with the extent of the study area and it tends to be correlated with AUC (Leroy et al., 2018).

4.3.2. Obtaining calibration metrics

Miller's statistic is based on the hypothesis that the calibration line – perfect calibration – has an intercept of zero and a slope of one, and it evaluates the capacity of a model to correctly predict the proportion of sites with a particular environmental profile that will be occupied (Miller et al., 1991; Pearce and Ferrier, 2000; Wintle et al., 2005). The presence-only calibration (POC) plot was designed for presence-only algorithms and determines whether predictions are proportional to conditional probability of presence (Phillips and Elith, 2010). Model predictions are divided into several bins, and the fraction of true presences is plotted against the average model value in each bin. If the plotted points are lined up close to the diagonal, the model is well-calibrated.

The Boyce index is the only calibration metric specifically designed for presence-background algorithms. It measures how much model predictions differ from a random distribution of the observed presences across the prediction gradients (Boyce et al., 2002). It ranges between -1 and +1. Positive values indicate a model in which predicted presences are consistent with the distribution of presences in the evaluation dataset, values close to zero mean that the model is not statistically different from a random model, and negative values indicate counter predictions, i.e. predicting lower suitability in areas where presences are more frequent (Hirzel et al., 2006).

4.3.3. Calculating null models

Another way of assessing model performance is to calculate and evaluate null models for comparison (Raes and ter Steege, 2007; Beale et al., 2008). The procedure consists of creating a set of pseudo-presence points randomly distributed across the study area, running the models, and calculating an evaluation metric, e.g. AUC or TSS. The evaluation metrics for the null models can be calculated with the species' training and test records as well (Bohl et al., 2019). The objective is to compare evaluation metrics for the species model and the null models, with a statistic such as the t -test. Validation metrics should be significantly

better for the species models. However, species prevalence will always affect the result of the evaluation metrics. The difference in the evaluation metrics between species models and null models will be smaller as prevalence increases. In other words, if the species is a generalist and is widely distributed throughout the study area, the distribution of its records will be more similar to a set of random points, so smaller differences between the species and the null models will be expected.

Independently of how statistically good the model is, it may still be biologically meaningless. As Table 2 shows, ecological niche models can be used to analyse the distribution of other entities rather than species. It is possible to model road-kills (Sillero, 2008), snake bites (Yañez-Arenas et al., 2014), or bat mortality due to collisions with windmills (Santos et al., 2013). But it is also possible to model any point process (biological or not) and obtain a statistically good statistical result. Model results should be checked always for their ecological meaning.

4.3.4. Comparing models with aic

The performance of presence-absence and presence-background models can be ranked with the Akaike Information Criterion (AIC) (Warren and Seifert, 2011; Warren et al., 2014). AIC is an estimate of the relative information lost when implementing a model, a trade-off between the goodness-of-fit and the simplicity (e.g. number of parameters) of the model. Hence, AIC is not a test of a model: it does not have a null hypothesis, and it does not provide information about the absolute quality of a model. AIC is a measure of quality relative to other models based on the same species occurrence dataset with different sets of predictor variables or parameters. AIC increases with information loss: thus, the best model for a set of occurrence data is the one with the lowest AIC (Warren and Seifert, 2011; Muscarella et al., 2014).

4.3.5. Analysing model results

After evaluating and checking the reliability of the results, it is possible to analyse the information provided by the model. First, we can analyse the contribution of the variables to the model, or how much variance each variable explains. Regression-based algorithms provide explained deviance or D^2 , as well as pseudo- R^2 values. The remaining algorithms provide only the importance of the variables. Most of the non-regression-based algorithms do not provide information about the statistical significance of the variables, although they can provide variable importance metrics.

Some algorithms provide response curves of the variables. These curves show how environmental suitability, or presence probability, changes as a function of each variable: it is possible to know which specific environmental conditions the species is preferably inhabiting. Some algorithms, such as Maxent, calculate the importance of the variables through a jack-knife procedure (Phillips et al., 2006). Maxent presents two sets of response curves. The first one shows the changes in the habitat suitability as each environmental variable is varied, keeping all other environmental variables at their average sample value. In the second set, each curve corresponds to a model calculated with only that variable. In the absence of correlation effects between variables, both sets of curves should be similar. If variables are strongly correlated, the response curve does not represent how habitat suitability changes exclusively with that variable, as the curve includes the contribution of other correlated variables. “Pure” curves are very difficult to obtain because variables always have some degree of correlation. Only when using the components of a principal component analysis, the curves will correspond exclusively to each component, but each component will be a linear combination of all variables (Pérez i de Lanuza et al., 2018).

4.3.6. Analysing model projections with clamping maps and extrapolation analysis

The next step is to analyse the model projections, i.e. predictions into the same set of variables used to train/perform models but expressing other areas or time frames. Projections can be restricted through clamping maps or carefully inspected for extrapolation through MESS

(Multivariate Environmental Similarity Surface; Elith et al., 2010; 2011) or ExDet (extrapolation detection; Mesgaran et al., 2014). Clamping maps indicate the values of the variables for the projection scenarios which are outside the modelled range of values. Let us imagine a linear regression model: the regression line of the model will be calibrated for a specific set of environmental values. The line (i.e. the model) will only be correctly extrapolated if the relationship between occurrences and environmental variables remains the same beyond those values. If the relationship changes (e.g. if there is an asymptote or an inflection), the model is not fitted for these new values of the variables, and results could be wrong (Elith and Leathwick, 2009). This rationale is the same for an ENM. The mathematical part of an ENM is a formula relating the species records to a set of environmental variables. This formula is applied to the environmental variables to obtain the final model. This is the way of projecting a model to other scenarios in time and/or space. If a model map is produced for the current environmental variables by applying the model formula, the same can be done for those variables from other scenarios in time and space. However, some values of these variables can be outside the range of values analysed by the model. The clamping maps will thus indicate the areas where the model is incurring in extrapolation. MESS and ExDet analyses provide an advanced clamping map. They indicate what variables and (combinations of) values are outside the modelled range. MESS measures the similarity of any given point to a reference set of points, in relationship to the variables (Elith et al., 2010; 2011). It reports the closeness of the point to the distribution of reference points, gives negative values for dissimilar points, and maps these values across the whole prediction region. ExDet measures the similarity between the reference and projection domains by accounting for both the deviation from the mean and the correlation between variables (Mesgaran et al., 2014).

4.4. Application of the model

4.4.1. Ensemble models

If several models are calculated with different algorithms, they can be merged in an ensemble by e.g. averaging them (Araújo and New, 2007; Marmion et al., 2009). Several R packages (Supplementary Text S2), such as ‘biomod2’, ‘sdm’ and ‘ENMTools’, allow to calculate ensemble models in several ways (Thuiller, 2003; Thuiller et al., 2009; Naimi and Araújo, 2016; Andrade et al., 2020). Although ensemble models are often calculated, performing an ensemble of good and bad models may not result in the best (or even a good) model. Bates and Granger (1969) demonstrated that combining forecasts may yield lower mean errors than any of the constituent individual forecasts, but only if 1) each of the forecasts was based on variables or information that the other forecasts had not considered, and 2) all forecasts were unbiased. Thus, ensemble forecasting should not be blindly used for combining different models based on the same variables. In addition, ENMs may yield either probability, suitability or favourability values, which differ in meaning and are not commensurable (Royle et al., 2012). Even suitability values obtained with different algorithms lack commensurability. Therefore, it is advisable not to merge models from different types of algorithms (or at least from different types of algorithm families), as noise will be introduced in the final ensemble model (Hao et al., 2019). In addition, recent studies show no particular advantage of model ensembles over individually tuned models (Hao et al., 2020). Alternatively, individual models can be overlapped to represent prediction agreement or consensus (Martínez-Freiría et al., 2017, 2020).

4.4.2. Thresholding the model

Frequently, decision-makers and stakeholders prefer to work with model predictions composed of only two categories: presence/absence, or suitable/unsuitable. It is then necessary to transform the actual model predictions, which typically range continuously from 0 to 1, to a binary prediction map, with only two classes. For that, a threshold must be applied to the continuous predictions: everything below the threshold

will be considered as predicted absence (or unfavourable/unsuitable areas); everything above the threshold will be considered as predicted presences (or favourable/suitable areas). There are many possible thresholds (Liu et al., 2005; Jiménez-Valverde and Lobo, 2007; Freeman and Moisen, 2008; Nenzén and Araújo, 2011; Liu et al., 2016), and the choice is largely arbitrary. Most thresholds are based on some metric from the confusion matrix, such as sensitivity and/or specificity. Some of the most used thresholds are the '10 percentile training presence threshold' and the 'Maximum training sensitivity plus specificity' (Vale et al., 2014). The first one accepts to classify 10% of the presences as false presences. The second one maximises the number of presences and absences correctly classified. Note that the thresholds for presence-absence and presence-background algorithms are of different power. In theory, presence-absence algorithms compare occupied to unoccupied habitats, to predict probable presence and probable absence areas. Therefore, thresholds often take values close to the species prevalence (or sometimes mean predicted probability) in the modelled sample (Nenzén and Araújo, 2011; Liu et al., 2013). Presence-background algorithms ideally distinguish suitable from unsuitable habitats based on the presences and a generally large sample of background points. Thus, the selected classification threshold will generally be lower, frequently around 0.3 or even lower, as suitable habitats include occupied and unoccupied habitats, and the ratio of presences/background points is usually low. For favourability values, which are calculated from probability obtained from presence-(pseudo) absence algorithms, a threshold of 0.5 will always separate favourable from unfavourable locations, regardless of species prevalence (Jiménez-Valverde and Lobo, 2006; Real et al., 2006; Acevedo and Real, 2012).

The transformation of model output values into binary predictions should only be done if necessary, as the actual continuous predictions contain important quantitative information. Binary model predictions may be useful to calculate potential species richness by adding up binary model outputs for a set of species (Martínez-Freiría et al., 2013; 2017, 2020; Sillero et al., 2009). Stacking species ENMs tends to overestimate species numbers; methods like the SESAM model can correct this over-estimation (Guisan and Rahbek, 2011). However, potential species richness can also be obtained without binary transformation, e.g. by summing up predicted favourabilities for a group of species using fuzzy logic (Real et al., 2017; Estrada et al., 2018).

4.4.3. Considering dispersal scenarios

When a model is projected to another period, species are usually considered to be able to disperse under two contrasting dispersal scenarios: no dispersal or unlimited dispersal (Carvalho et al., 2010a). The no dispersal model supposes that the species is not able to disperse to new areas. Thus, the projection excludes any newly suitable areas without current observed species' presences. On the other hand, the unlimited dispersal model considers that the species has an unrestricted capacity to disperse to new areas. The projection will include all newly suitable areas. The reality must be somewhere between those two dispersal scenarios. There are tools, such as the MigClim R package (Engler and Guisan, 2009; Engler et al., 2012), to model the dispersal of a species from the current situation to the projected one. It is also possible to calculate the required dispersal rate for a species and compare that to its observed dispersal rates (Yu et al., 2019).

5. Final remarks

ENMs, like any other modelling tool, have limitations and conditions to be considered in order to obtain reliable results. Currently, the most important recurrent issue in correlative ENMs is probably the lack of consistency in guaranteeing that models fulfil all statistical assumptions when applied. This is essential for employing ENMs with certainty in conservation, biogeographical, and ecological studies. In this sense, this work aims to provide step-by-step guidelines for improving the quality

of ENMs studies. However, there are some issues that still need improvement, where future research should focus on. One is model validation. New validation methods not based on the confusion matrix are necessary, especially for algorithms that do not include absences. Null models should become a standard procedure in the validation process. More studies are also needed to clarify the differences between modelling methods, as a 'best algorithm' may not exist if each algorithm is a different approximation to species' ecological niche.

Another topic in need of further development is how to define the size and shape of the study area, through easy-to-use and objective criteria. As stated in this review, some of the current criteria are not easy to implement due to the dependence on expert knowledge. Also, further research is necessary for the application of true spatial methods (e.g. auto-logistic regressions, geographically weighted regressions), where the species distribution is modelled directly in geographical- rather than in environmental- space.

Finally, more standard methods are necessary to obtain better species physiological datasets relevant to mechanistic models. Further, combining correlative and mechanistic models will make ENMs more process-based, a strong necessity for better establishing cause-effect relationship between the environment and species distributions. In summary, the future of ENMs is promising, but it will only be successful if best practices are reviewed, updated, and provided to beginners and also more experienced users.

6. CRediT author statement

Neftalí Sillero: Conceptualization. Neftalí Sillero, Salvador Arenas-Castro, Urtzi Enriquez-Urzelai: Writing- Original draft preparation. Neftalí Sillero, Salvador Arenas-Castro, Urtzi Enriquez-Urzelai, Cândida Gomes Vale, Diana Sousa-Guedes, Fernando Martínez-Freiría, Raimundo Real, A. Márcia Barbosa: Writing- Reviewing and Editing.

Declaration of Competing Interest

We declare that the manuscript titled 'Want to model a species niche? A step by step guideline on correlative ecological niche modelling', by Neftalí Sillero, Salvador Arenas-Castro, Urtzi Enriquez-Urzelai, Cândida Gomes Vale, Diana Sousa-Guedes, Fernando Martínez-Freiría, Raimundo Real, and A. Márcia Barbosa, has not been published previously and is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

Acknowledgements

NS is supported by a CEEC2017 contract (CEECIND/02213/2017). SAC is supported by the PORBIOTA—E-Infraestrutura Portuguesa de Informação e Investigação em Biodiversidade fellowship (POCI-01-0145-FEDER-022127). FM-F is supported by a contract from FCT (ref. DL57/2016/CP1440/CT0010).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ecolmodel.2021.109671](https://doi.org/10.1016/j.ecolmodel.2021.109671).

References

- Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from, 1958-2015. *Sci. Data* 5, 1–12. <https://doi.org/10.1038/sdata.2017.191>.

- Acevedo, P., Jiménez-Valverde, A., Lobo, J.M., Real, R., 2012. Delimiting the geographical background in species distribution modelling. *J. Biogeogr.* 39, 1383–1390. <https://doi.org/10.1111/j.1365-2699.2012.02713.x>.
- Acevedo, P., Real, R., 2012. Favourability: concept, distinctive characteristics and potential usefulness. *Naturwissenschaften*. 99, 515–522. <https://doi.org/10.1007/s00114-012-0926-0>.
- Ahmadzadeh, F., Flecks, M., Carretero, M.A., Bohme, W., Ihlow, F., Kapli, P., Miraldo, A., Rodder, D., 2016. Separate histories in both sides of the mediterranean: phylogeny and niche evolution of ocellated lizards. *J. Biogeogr.* 43, 1242–1253. <https://doi.org/10.1111/jbi.12703>.
- Aiello-Lammens, M.E., Boria, 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. *Ecograph. (Cop.)* 38, 1–35. <https://doi.org/10.1111/ecog.01132>.
- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (tss). *J. Appl. Ecol.* 6, 1223–1232.
- Anderson, K., Gaston, K.J., 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Front. Ecol. Environ.* 11, 138–146. <https://doi.org/10.1890/120150>.
- Anderson, R.P., 2015. El modelado de nichos y distribuciones: no es simplemente “clíc, clic, clic. *Biogeografía* 8, 4–27.
- Anderson, R.P., 2013. A framework for using niche models to estimate impacts of climate change on species distributions. *Ann. N. Y. Acad. Sci.* 1297, 8–28. <https://doi.org/10.1111/nyas.12264>.
- Anderson, R.P., 2012. Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Ann. N. Y. Acad. Sci.* 1260, 66–80. <https://doi.org/10.1111/j.1749-6632.2011.06440.x>.
- 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. *J. Biogeogr.* 37, 1378–1393. <https://doi.org/10.1111/j.1365-2699.2010.02290.x>.
- Andrade, A.F.A., de Velasco, S.J.E., De Marco Júnior, P., 2020. ENMTML: an R package for a straightforward construction of complex ecological niche models. *Environ. Model. Softw.* 125, 104615. <https://doi.org/10.1016/j.envsoft.2019.104615>.
- Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., García, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. *Sci. Adv.* 5, eaat4858.
- Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trend. Ecol. Evol.* 22, 42–47.
- Araújo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. *Ecol.* 93, 1527–1539. <https://doi.org/10.1890/11-1930.1>.
- Areias-Guerreiro, J., Mira, A., Barbosa, A.M., 2016. How well can models predict changes in species distributions? a 13-year-old otter model revisited data gathering. *Hystrix, Ital. J. Mammal.* 1–5. <https://doi.org/10.4404/hystrix-27.1-1867>.
- Arenas-Castro, S., Gonçalves, J., 2021. SDM-CropProj – A model-assisted framework to forecast crop environmental suitability and fruit production. *MethodsX* 8, 101394. <https://doi.org/10.1016/j.mex.2021.101394>.
- Arenas-Castro, S., Gonçalves, J., Alves, P., Alcaraz-Segura, D., Honrado, J.P., 2018. Assessing the multi-scale predictive ability of ecosystem functional attributes for species distribution modelling. *PLoS ONE* 13, 1–31. <https://doi.org/10.1371/journal.pone.0199292>.
- Arenas-Castro, S., Gonçalves, J.F., Moreno, M., Villar, R., 2020. Projected climate changes are expected to decrease the suitability and production of olive varieties in southern Spain. *Sci. Tot. Environ.* 709, 136161. <https://doi.org/10.1016/j.scitotenv.2019.136161>.
- Arenas-Castro, S., Regos, A., Gonçalves, J.F., Alcaraz-Segura, D., Honrado, J., 2019. Remotely sensed variables of ecosystem functioning support robust predictions of abundance patterns for rare species. *Rem. Sens.* 11, 2086. <https://doi.org/10.3390/rs11182086>.
- Arenas-Castro, S., Sillero, N., 2021. Cross-scale monitoring of habitat suitability changes using satellite time series and ecological niche models. *Sci. Tot. Environ.* 784. <https://doi.org/10.1016/j.scitotenv.2021.147172>.
- Arntzen, J.W., Espregueira Themudo, G., 2008. Environmental parameters that determine species geographical range limits as a matter of time and space. *J. Biogeogr.* 7, 1177–1186.
- 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, 1–17. <https://doi.org/10.1371/journal.pone.0179152>.
- Báez, J., Real, R., García-Soto, C., de la Serna, J., Macías, D., Camiñas, J., 2007. Loggerhead turtle by-catch depends on distance to the coast, independent of fishing effort: implications for conservation and fisheries management. *Mar. Ecol. Prog. Ser.* 338, 249–256. <https://doi.org/10.3354/meps338249>.
- Báez, J.C., Barbosa, A.M., Pascual, P., Ramos, M.L., Abascal, F., 2020. Ensemble modeling of the potential distribution of the whale shark in the atlantic ocean. *Ecol. Evol.* 10, 175–184. <https://doi.org/10.1002/ece3.5884>.
- Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Meth. Ecol. Evol.* 3, 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>.
- Barbet-Massin, M., Thuiller, W., Jiguet, F., 2010. How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? *Ecograph. (Cop.)* 33, 878–886. <https://doi.org/10.1111/j.1600-0587.2010.06181.x>.
- Barbosa, A.M., Real, R., Muñoz, A.R., Brown, J.A., 2013. New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Dive.. Distrib.* 19, 1333–1338. <https://doi.org/10.1111/ddi.12100>.
- Barbosa, A.M., Real, R., Olivero, J., Vargas, J.M., 2003. Otter (lutra lutra) distribution modeling at two resolution scales suited to conservation planning in the iberian peninsula. *Biol. Conserv.* 3, 377–387.
- Barbosa, M.A., Sillero, N., Martínez-Freiria, F., Real, R., 2012. Ecological niche models in mediterranean herpetology: past, present and future. *Ecological Modelling. WenJun Zhang*, pp. 173–204.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Modell.* 222, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>.
- Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. *J. Oper. Res. Soc.* 20, 451–468. <https://doi.org/10.1057/jors.1969.103>.
- Battlori, E., Parisien, M.A., Krawchuk, M.A., Moritz, M.A., 2013. Climate change-induced shifts in fire for mediterranean ecosystems. *Glob. Ecol. Biogeogr.* 22, 1118–1129. <https://doi.org/10.1111/geb.12065>.
- Beale, C.M., Lennon, J.J., Gimona, A., 2008. Opening the climate envelope reveals no macroscale associations with climate in european birds. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14908–14912. <https://doi.org/10.1073/pnas.0803506105>.
- Beaumont, L.J., Hughes, L., Poulsen, M., 2005. Predicting species distributions: use of climatic parameters in bioclim and its impact on predictions of species' current and future distributions. *Ecol. Modell.* 2, 251–270.
- Beck, J., Böller, M., Erhardt, A., Schwanghart, W., 2014. Spatial bias in the gbif database and its effect on modeling species' geographic distributions. *Ecol. Inform.* 19, 10–15. <https://doi.org/10.1016/j.ecoinf.2013.11.002>.
- Bedia, J., Herrera, S., Gutiérrez, J.M., 2013. Dangers of using global bioclimatic datasets for ecological niche modeling. limitations for future climate projections. *Glob. Planet. Chang.* 107, 1–12. <https://doi.org/10.1016/j.gloplacha.2013.04.005>.
- Berger, M., Aschbacher, J., 2012. The sentinel missions-new opportunities for science. *Rem. Sens. Environ.* 120, 1–2. <https://doi.org/10.1016/j.rse.2011.12.026>.
- Blonder, B., Lamanna, C., Violle, C., Enquist, B.J., 2014. The n-dimensional hypervolume. *Glob. Ecol. Biogeogr.* 23, 595–609. <https://doi.org/10.1111/geb.12146>.
- Bohl, C.L., Kass, J.M., Anderson, R.P., 2019. A new null model approach to quantify performance and significance for ecological niche models of species distributions. *J. Biogeogr.* 46, 1101–1111. <https://doi.org/10.1111/jbi.13573>.
- Booth, T.H., Nix, H.A., Busby, J.R., Hutchinson, M.F., 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current maxent studies. *Divers. Distrib.* 20, 1–9. <https://doi.org/10.1111/ddi.12144>.
- Boria, 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. *Ecol. Modell.* 275, 73–77. <https://doi.org/10.1016/j.ecolmodel.2013.12.012>.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. *Ecol. Modell.* 2–3, 281–300.
- Bradley, A.P., 1997. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Patte. Recognit.* 30, 1145–1159.
- Breiman, L., 1999. Random forest. *Mach. Learn.* 45, 1–35. <https://doi.org/10.1023/A:1010933404324>.
- Breiner, F.T., Guisan, A., Bergamini, A., Nobis, M.P., 2015. Overcoming limitations of modelling rare species by using ensembles of small models. *Meth. Ecol. Evol.* 6, 1210–1218. <https://doi.org/10.1111/2041-210X.12403>.
- Breiner, F.T., Nobis, M.P., Bergamini, A., Guisan, A., 2018. Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Meth. Ecol. Evol.* 2018, 1–7. <https://doi.org/10.1111/2041-210X.12957>.
- 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. *Glob. Chang. Biol.* 22, 2425–2439. <https://doi.org/10.1111/geb.13280>.
- Brito, J.C., Crespo, e.g., 2002. Distributional analysis of two vipers (*vipera latastei* and *v. seaneii*) in a potential area of sympatry in the northwestern iberian peninsula. In: Schuett, G.W., Hoggren, M., Douglas, M.E., Greene, H.W. (Eds.), *Biology of the Vipers*. Eagle Mountain Publishing, LC, Eagle Mountain, Utah, pp. 129–138.
- Brito, J.C., Crespo, e.g. Paulo, O.S., 1999. Modelling wildlife distributions: logistic multiple regression vs overlap analysis. *Ecograph. (Cop.)* 3, 251–260.
- Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz, N.G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N.E., Graham, C.H., Guisan, A., 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob. Ecol. Biogeogr.* 21, 481–497. <https://doi.org/10.1111/j.1466-8238.2011.00698.x>.
- Brown, J.L., 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Meth. Ecol. Evol.* 5, 694–700. <https://doi.org/10.1111/2041-210X.12200>.
- Buckley, L.B., Cannistra, A.F., John, A., 2018. Leveraging organismal biology to forecast the effects of climate change. *Integr. Comp. Biol.* 58, 38–51. <https://doi.org/10.1093/icb/icy018>.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. *Glob. Chang. Biol.* 16, 1145–1157. <https://doi.org/10.1111/j.1365-2486.2009.02000.x>.
- Burrough, P.A., McDonnell, R., 1998. *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford.
- Cabral, J.S., Valente, L., Hartig, F., 2017. Mechanistic simulation models in macroecology and biogeography: state-of-art and prospects. *Ecograph. (Cop.)* 40, 267–280. <https://doi.org/10.1111/ecog.02480>.

- Carpenter, G., Gillison, A.N., Winter, J., 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers. Conserv.* 6, 667–680.
- Carretero, M.A., Sillero, N., 2016. Evaluating how species niche modelling is affected by partial distributions with an empirical case. *Acta Oecologic.* 77, 207–216. <https://doi.org/10.1016/j.actao.2016.08.014>.
- Carvalho, S.B., Brito, J.C., Crespo, E., Watts, M.E., Possingham, H.P., 2011. Conservation planning under climate change: toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time. *Biol. Conserv.* 144, 2020–2030. <https://doi.org/10.1016/j.biocon.2011.04.024>.
- Carvalho, S.B., Brito, J.C., Crespo, E.J., Possingham, H.P., 2010a. From climate change predictions to actions: conserving vulnerable animal groups in hotspots at a regional scale. *Glob. Chang. Biol.* 16, 3257–3270.
- Carvalho, S.B., Brito, J.C., Pressey, R.L., Crespo, E., Possingham, H.P., 2010b. Simulating the effects of using different types of species distribution data in reserve selection. *Biol. Conserv.* 143, 426–438.
- Castro, A., Muñoz, A.R., Real, R., 2008. Modelling the spatial distribution of the tengmalm's owl *aegolius funereus* in its southwestern palaeartic limit (ne Spain). *Ardeola* 55, 71–85.
- Chamorro, D., Real, R., Muñoz, A.R., 2020. Fuzzy sets allow gaging the extent and rate of species range shift due to climate change. *Sci. Rep.* 10, 1–14. <https://doi.org/10.1038/s41598-020-73509-y>.
- Chapman, D., Pescott, O.L., Roy, H.E., Tanner, R., 2019. Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. *J. Biogeogr.* 46, 1029–1040. <https://doi.org/10.1111/jbi.13555>.
- Chefaoui, R.M., Assis, J., Duarte, C.M., Serrão, E.A., 2016. Large-Scale prediction of seagrass distribution integrating landscape metrics and environmental factors: the case of *cymodocea nodosa* (mediterranean-atlantic). *Estuari. Coast.* 39, 123–137. <https://doi.org/10.1007/s12237-015-9966-y>.
- Chefaoui, R.M., Lobo, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecol. Modell.* 4, 478–486.
- Chiarenza, A.A., Mannion, P.D., Lunt, D.J., Farnsworth, A., Jones, L.A., Kelland, S.-J., Allison, P.A., 2019. Ecological niche modelling does not support climatically-driven dinosaur diversity decline before the cretaceous/paleogene mass extinction. *Nat. Commun.* 10, 1091. <https://doi.org/10.1038/s41467-019-08997-2>.
- Chipperfield, J.D., Benito, B.M., O'Hara, R., Telford, R.J., Carlson, C.J., 2020. On the Inadequacy of Species Distribution Models For Modelling the Spread of SARS-CoV-2: Response to Araújo and Naimi. *EcoEvoRxiv*.
- Clark, J.D., Dunn, J.E., Smith, K.G., 1993. A multivariate model of female black bear habitat use for a Geographic Information System. *J. Wildl. Manage.* 57, 519–526.
- Clark, P.J., Evans, F.C., 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecol.* 35, 445–453.
- Comrie, A.C., Glenn, E.C., 1999. Principal components-based regionalization of precipitation regimes across the southwest united states and northern mexico, with an application to monsoon precipitation variability. *Clim. Res.* 10, 201–215. <https://doi.org/10.3354/cr010201>.
- Cord, A.F., Meentemeyer, R.K., Leitão, P.J., Václavík, T., 2013. Modelling species distributions with remote sensing data: bridging disciplinary perspectives. *J. Biogeogr.* 40, 2226–2227. <https://doi.org/10.1111/jbi.12199>.
- Costa, H., Foody, G.M., Jiménez, S., Silva, L., 2015. Impacts of species misidentification on species distribution modeling with presence-only data. *ISPRS. Int. J. Geo-Information* 4, 2496–2518. <https://doi.org/10.3390/ijgi4042496>.
- Cressie, N.A.C., 1993. *Statistical For Spatial Data*. Wiley and Sons, New York. Revised Edition.
- Cumming, G.S., 2002. Comparing climate and vegetation as limiting factors for species ranges of african ticks. *Ecol.* 1, 255–268.
- Cumming, G.S., 2000. Using between-model comparisons to fine-tune linear models of species ranges. *J. Biogeogr.* 2, 441–455.
- Dallas, T.A., Hastings, A., 2018. Habitat suitability estimated by niche models is largely unrelated to species abundance. *Glob. Ecol. Biogeogr.* 27, 1448–1456. <https://doi.org/10.1111/geb.12820>.
- De Marco, P., Nóbrega, C.C., 2018. Evaluating collinearity effects on species distribution models: an approach based on virtual species simulation. *PLoS ONE* 13. <https://doi.org/10.1371/journal.pone.0202403>.
- Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F.T., D'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R.G., Hordijk, W., Salamin, N., Guisan, A., 2017. Ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecograph. (Cop.)* 40, 774–787. <https://doi.org/10.1111/ecog.02671>.
- Dicko, A.H., Lancelot, R., Seck, M.T., Guerrini, L., Sall, B., Lo, M., Vreysen, M.J.B., 2014. Using species distribution models to optimize vector control in the framework of the tsetse eradication campaign in senegal. *Proc. Natl. Acad. Sci* 1–6. <https://doi.org/10.1073/pnas.1407773111>.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münckmüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecograph. (Cop.)* 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B., Singer, A., 2012. Correlation and process in species distribution models: bridging a dichotomy. *J. Biogeogr.* 39, 2119–2131. <https://doi.org/10.1111/j.1365-2699.2011.02659.x>.
- Dudík, M., Schapire, R.E., Phillips, S.J., 2005. Correcting sample selection bias in maximum entropy density estimation. *Adv. Neural Inf. Process. Syst.* 323–330.
- Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. *Meth. Ecol. Evol.* 1, 330–342.
- 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. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of maxent for ecologists. *Divers. Distrib.* 17, 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>.
- Elton, C., 1927. *Animal Ecology*. London.
- Engler, R., Guisan, A., 2009. MigClim: predicting plant distribution and dispersal in a changing climate. *Divers. Distrib.* 15, 590–601.
- Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *J. Appl. Ecol.* 41, 263–274.
- Engler, R., Hordijk, W., Guisan, A., 2012. The migclim R package - seamless integration of dispersal constraints into projections of species distribution models. *Ecograph. (Cop.)* 35, 872–878. <https://doi.org/10.1111/j.1600-0587.2012.07608.x>.
- Enrique, M., Muñoz, D.S., Giovanni, R.De, Siqueira, M.F.De, Sutton, T., Brewer, P., 2009. OpenModeller: a generic approach to species ' potential distribution modelling. *Geoinformati.* 1–25. <https://doi.org/10.1007/s10707-009-0090-7>.
- Enriquez-Urzelai, U., Tingley, R., Kearney, M.R., Sacco, M., Palacio, A.S., Tejedo, M., Nicieza, A.G., 2020. The roles of acclimation and behaviour in buffering climate change impacts along elevational gradients. *J. Anim. Ecol.* 89, 1722–1734. <https://doi.org/10.1111/1365-2656.13222>.
- Enriquez-Urzelai, U., Kearney, M.R., Nicieza, A.G., Tingley, R., 2019. Integrating mechanistic and correlative niche models to unravel range-limiting processes in a temperate amphibian. *Glob. Chang. Biol.* 1–15. <https://doi.org/10.1111/gcb.14673>.
- Espregueira Themudo, G., Wielstra, B., Arntzen, J.W., 2009. Multiple nuclear and mitochondrial genes resolve the branching order of a rapid radiation of crested newts (*triturus*, salamandridae). *Mol. Phylogenet. Evol.* 2, 321–328.
- Estrada, A., Barbosa, A.M., Real, R., 2018. Changes in potential mammal diversity in national parks and their implications for conservation. *Curr. Zool.* 64, 671–679. <https://doi.org/10.1093/cz/zoy001>.
- Evans, J., 2021. SpatialEco. R package version 1.3-6, <https://github.com/jeffrejevans/spatialEco>.
- Fawcett, T., 2006. An introduction to roc analysis. *Patte. Recognit. Lett.* 27, 861–874.
- Feller, W., 1968. *An Introduction to Probability Theory and Its Applications*. Popul. (French Ed). 10.2307/1526422.
- Feng, X., Park, D.S., Liang, Y., Pandey, R., Papeš, M., 2019a. Collinearity in ecological niche modeling: confusions and challenges. *Ecol. Evol.* ece3.5555 <https://doi.org/10.1002/ee3.5555>.
- Feng, X., Park, D.S., Walker, C., Peterson, A.T., Merow, C., Papeš, M., 2019b. A checklist for maximizing reproducibility of ecological niche models. *Nat. Ecol. Evol.* 3 <https://doi.org/10.1038/s41559-019-0972-5>.
- Fernandes, R.F., Scherrer, D., Guisan, A., 2019. Effects of simulated observation errors on the performance of species distribution models. *Divers. Distrib.* 25, 400–413. <https://doi.org/10.1111/ddi.12868>.
- Ferrier, S., Watson, G., Pearce, J., Drielsma, M., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast new southwales. i. species-level modelling. *Biodivers. Conserv.* 2275–2307.
- Ficetola, G.F., Thuiller, W., Maud, C., 2007. Prediction and validation of the potential global distribution of a problematic alien invasive species: the american bullfrog. *Divers. Distrib.* 13, 476–485.
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>.
- Field, A., Miles, J., Field, Z., 2012. *Discovering Statistics using R*. Sage Publications.
- 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., Elith, J., Hastie, T., Keith, D.A., 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. *Meth. Ecol. Evol.* 6, 424–438. <https://doi.org/10.1111/2041-210X.12242>.
- 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. <https://doi.org/10.1371/journal.pone.0097122>.
- Franch, M., Montori, A., Sillero, N., Llorente, G.A., 2015. Temporal analysis of mauremys leprosa (testudines, geoemydidae) distribution in northeastern iberia: unusual increase in the distribution of a native species. *Hydrobiologia* 757, 129–142. <https://doi.org/10.1007/s10750-015-2247-8>.
- Franklin, J., 2010. *Mapping Species Distributions*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324.004>.
- Freeman, E.A., Moisen, G.G., 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecol. Modell.* 1–2, 48–58.
- Fréjaville, T., Benito Garzón, M., 2018. The eumedclim database: yearly climate data (1901–2014) of 1 km resolution grids for europe and the mediterranean basin. *Front. Ecol. Evol.* 6, 1–5. <https://doi.org/10.3389/fevo.2018.00031>.
- Gallien, L., Douzet, R., Pratte, S., Zimmermann, N.E., Thuiller, W., 2012. Invasive species distribution models - how violating the equilibrium assumption can create new insights. *Glob. Ecol. Biogeogr.* 21, 1126–1136. <https://doi.org/10.1111/j.1466-8238.2012.00768.x>.
- Getis, A., 2008. A history of the concept of spatial autocorrelation: a geographer 's perspective. *Geogr. Anal.* 40, 297–309.

- Getis, A., 2007. Reflections on spatial autocorrelation. *Reg. Sci. Urban Econ.* 37, 491–496. <https://doi.org/10.1016/j.regsciurbeco.2007.04.005>.
- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Townsend Peterson, A., Loiselle, B.A., 2007. The influence of spatial errors in species occurrence data used in distribution models. *J. Appl. Ecol.* 45, 239–247. <https://doi.org/10.1111/j.1365-2664.2007.01408.x>.
- Grinnell, J., 1917. The niche-relationships of the California thrasher. *Auk* 34, 427–433.
- Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., 2014. Maxent is not a presence-absence method: a comment on Thibaud et al. *Meth. Ecol. Evol.* 5, 1192–1197. <https://doi.org/10.1111/2041-210X.12252>.
- Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R., Wintle, B.A., 2015. Is my species distribution model fit for purpose? matching data and models to applications. *Glob. Ecol. Biogeogr.* 24, 276–292. <https://doi.org/10.1111/geb.12268>.
- Guisan, A., Edwards, T.C.J., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Modell.* 89–100.
- Guisan, A., Rahbek, C., 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *J. Biogeogr.* 38, 1433–1444. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 9, 993–1009.
- Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. Habitat Suitability and Distribution Models: With applications in R. *Habitat Suitability and Distribution Models: With Applications in R*. 10.1017/9781139028271.
- Guisan, A., Weiss, S.B., Weiss, A.D., 1999. GLM versus cca spatial modeling of plant species distribution. *Plant Ecol.* 107–122.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Modell.* 135, 147–186.
- Guo, Q., Liu, Y., 2010. ModEco: an integrated software package for ecological niche modeling. *Ecograph. (Cop.)* 4, 637–642.
- Hampe, A., 2004. Bioclimate envelope models: what they detect and what they hide. *Glob. Ecol. Biogeogr.* 13, 469–471. <https://doi.org/10.1111/j.1466-822X.2004.00112.x>.
- Hanley, J.A., McNeil, V.J., 1982. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiol.* 29–63.
- 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. *Divers. Distrib.* 25, 839–852. <https://doi.org/10.1111/ddi.12892>.
- 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. *Ecograph. (Cop.)* 43, 549–558. <https://doi.org/10.1111/ecog.04890>.
- Harisena, N.V., Groen, T.A., Toxopeus, A.G., Naimi, B., 2021. When is variable importance estimation in species distribution modelling affected by spatial correlation? *Ecograph. (Cop.)* 44, 778–788. <https://doi.org/10.1111/ecog.05534>.
- Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations - the crs ts3.10 dataset. *Int. J. Climatol.* 34, 623–642. <https://doi.org/10.1002/joc.3711>.
- Hatten, J.R., 2014. Mapping and monitoring mount graham red squirrel habitat with lidar and landsat imagery. *Ecol. Modell.* 289, 106–123. <https://doi.org/10.1016/j.ecolmodel.2014.07.004>.
- He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.N., Schmidtlein, S., Turner, W., Wegmann, M., Pettorelli, N., 2015. Will remote sensing shape the next generation of species distribution models? *Rem. Sens. Ecol. Conserv.* 1, 4–18. <https://doi.org/10.1002/rse2.7>.
- Heisey, D.M., Osnas, E.E., Cross, P.C., Joly, D.O., Langenberg, J.A., Miller, M.W., 2010. Linking process to pattern: estimating spatiotemporal dynamics of a wildlife epidemic from cross-sectional data. *Ecol. Monogr.* 80, 221–240. <https://doi.org/10.1890/09-0052.1>.
- Hemami, M.R., Esmaeili, S., Brito, J.C., Ahmadi, M., Omid, M., Martínez-Freiría, F., 2018. Using ecological models to explore niche partitioning within a guild of desert felids. *Hystrix* 29. <https://doi.org/10.4404/hystrix-00042-2017>.
- Hernandez, P.A., Franke, I., Herzog, S.K., Pacheco, V., Paniagua, L., Quintana, H.L., Soto, A., Swenson, J.J., Tovar, C., Valqui, T.H., Vargas, J., Young, B.E., 2008. Predicting species distributions in poorly studied landscapes. *Biodivers. Conserv.* 17, 1353–1366. <https://doi.org/10.1007/s10531-007-9314-z>.
- 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. *Ecograph. (Cop.)* 29, 773–785.
- Hertzog, L.R., Besnard, A., Jay-Robert, P., 2014. Field validation shows bias-corrected pseudo-absence selection is the best method for predictive species-distribution modelling. *Divers. Distrib.* 20, 1403–1413. <https://doi.org/10.1111/ddi.12249>.
- Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecol.* 93, 679–688. <https://doi.org/10.1890/11-0826.1>.
- Hijmans, R.J., Cameron, E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.
- Hirzel, A.H., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modelling? *Ecol. Modell.* 2–3, 331–341.
- Hirzel, A.H., Hausser, J., Chesler, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat suitability maps without absence-data? *Ecol.* 7, 2027–2036.
- Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences: predicting species distributions - results from a second workshop on advances in predictive species distribution models, held in riederalp, switzerland, 2004. *Ecol. Modell.* 2, 142–152.
- Holt, R.D., 2003. On the evolutionary ecology of species' ranges. *Evol. Ecol. Res.* 5, 159–178.
- Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*, 2nd Edition. John Wiley & Sons, Inc.
- Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* 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. *Glob. Chang. Biol.* 23, 164–176. <https://doi.org/10.1111/gcb.13470>.
- 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. *Ecol. Modell.* 312, 166–174. <https://doi.org/10.1016/j.ecolmodel.2015.05.018>.
- Jackson, S.T., Overpeck, J.T., 2000. Responses of plant populations and communities to environmental changes of the late quaternary. *Paleobiol.* 4, 194–220.
- Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morisette, J.T., Holcombe, T.R., 2015. Caveats for correlative species distribution modeling. *Ecol. Inform.* 29, 6–15. <https://doi.org/10.1016/j.ecoinf.2015.06.007>.
- Jarnevich, C.S., Talbert, M., Morisette, J., Aldridge, C., Brown, C.S., Kumar, S., Manier, D., Talbert, C., Holcombe, T., 2017. Minimizing effects of methodological decisions on interpretation and prediction in species distribution studies: an example with background selection. *Ecol. Modell.* 363, 48–56. <https://doi.org/10.1016/j.ecolmodel.2017.08.017>.
- Jiménez-Valverde, A., Acevedo, P., Barbosa, A.M., Lobo, J.M., Real, R., 2013. Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Glob. Ecol. Biogeogr.* 22, 508–516. <https://doi.org/10.1111/geb.12007>.
- Jiménez-Valverde, A., Lobo, J., Hortal, J., 2009. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Commun. Ecol.* 10, 196–205. <https://doi.org/10.1556/ComEc.10.2009.2.9>.
- Jiménez-Valverde, A., Lobo, J.M., 2006. The ghost of unbalanced species distribution data in geographical model predictions. *Divers. Distrib.* 12, 521–524.
- Jiménez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologic.* 31, 361–369.
- Jiménez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Divers. Distrib.* 6, 885–890.
- Kadmon, R., Farber, O., Danin, A., 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol. Appl.* 14, 401–413. <https://doi.org/10.1890/02-5364>.
- Kanagaraj, R., Wiegand, T., Mohamed, A., Kramer-Schadt, S., 2013. Modelling species distributions to map the road towards carnivore conservation in the tropics. *Raffl. Bull. Zool.* 85–107.
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4, 1–20. <https://doi.org/10.1038/sdata.2017.122>.
- Kearney, M., Porter, W.P., 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol. Lett.* 4, 334–350.
- Kearney, M., Porter, W.P., Williams, C., Ritchie, S., Hoffmann, A.A., 2009. Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Funct. Ecol.* 3, 528–538.
- Kearney, M., Simpson, S.J., Raubenheimer, D., Helmuth, B., 2010. Modelling the ecological niche from functional traits. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 3469–3483. <https://doi.org/10.1098/rstb.2010.0034>.
- Kearney, M.R., Porter, W.P., 2020. NicheMapR - an R package for biophysical modelling: the ectotherm and dynamic energy budget models. *Ecograph. (Cop.)* 43, 85–96. <https://doi.org/10.1111/ecog.04680>.
- Khalatbari, L., Yusefi, G.H., Martínez-Freiría, F., Jowkar, H., Brito, J.C., 2018. Availability of prey and natural habitats are related with temporal dynamics in range and habitat suitability for asiatic cheetah. *Hystrix, Ital. J. Mammal.* 29, 145–151. <https://doi.org/10.4404/hystrix-00080-2018>.
- Kolbe, J.J., Kearney, M., Shine, R., 2010. Modeling the consequences of thermal trait variation for the cane toad invasion of Australia. *Ecol. Appl.* 20, 2273–2285. <https://doi.org/10.1890/09-1973.1>.
- Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schroder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A., 2013. The importance of correcting for sampling bias in maxent species distribution models. *Divers. Distrib.* 19, 1366–1379. <https://doi.org/10.1111/ddi.12096>.
- Kriticos, D.J., Leriche, A., 2010. The effects of climate data precision on fitting and projecting species niche models. *Ecograph. (Cop.)* 1, 115–127.
- Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J., Scott, J.K., 2012. CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Meth. Ecol. Evol.* 3, 53–64. <https://doi.org/10.1111/j.2041-210X.2011.00134.x>.
- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2004. *Applied linear statistical models* fifth edition.
- Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecol.* 74, 1659–1673.
- Leroy, B., Delsol, R., Huguency, B., Meynard, C.N., Barhoumi, C., Barbet-Massin, M., Bellard, C., 2018. Without quality presence-absence data, discrimination metrics such as tss can be misleading measures of model performance. *J. Biogeogr.* 45, 1994–2002. <https://doi.org/10.1111/jbi.13402>.

- Levins, R., 1969. Some demographic and genetic consequences of environmental heterogeneity for. *Biologic. Contr. Bull. Entomol. Soc. Am.* 15, 237–240. <https://doi.org/10.1093/besa/15.3.237>.
- Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecograph. (Cop.)* 28, 385–393.
- Liu, C., Newell, G., White, M., 2016. On the selection of thresholds for predicting species occurrence with presence-only data. *Ecol. Evol.* 6, 337–348. <https://doi.org/10.1002/ece3.1878>.
- Liu, C., White, M., Newell, G., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* 40, 778–789. <https://doi.org/10.1111/jbi.12058>.
- Liu, C., White, M., Newell, G., 2011. Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecograph. (Cop.)* 34, 232–243. <https://doi.org/10.1111/j.1600-0587.2010.06354.x>.
- Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecograph. (Cop.)* 33, 103–114.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–151.
- 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.* 19, 1–7. <https://doi.org/10.1016/j.jnc.2010.03.002>.
- Maguire, B., 1973. Niche response structure and the analytical potentials of its relationship to the habitat. *Am. Nat.* 107, 213–246. <https://doi.org/10.1086/282827>.
- Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence/absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* 38, 921–931.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* 15, 59–69.
- Martínez-Freiría, F., Argaz, H., Fahd, S., Brito, J.C., 2013. Climate change is predicted to negatively influence moroccan endemic reptile richness. implications for conservation in protected areas. *Naturwissenschaften.* 100, 877–889. <https://doi.org/10.1007/s00114-013-1088-4>.
- Martínez-Freiría, F., Crochet, P.A., Fahd, S., Geniez, P., Brito, J.C., Velo-Antón, G., 2017. Integrative phylogeographical and ecological analysis reveals multiple pleistocene refugia for mediterranean dabia vipers in north-west africa. *Biol. J. Linn. Soc.* 122, 366–384. <https://doi.org/10.1093/biolinnean/blx038>.
- Martínez-Freiría, F., Freitas, I., Zuffi, M.A.L., Golay, P., Ursenbacher, S., Velo-Antón, G., 2020. Climatic refugia boosted allopatric diversification in western mediterranean vipers. *J. Biogeogr.* 47, 1698–1713. <https://doi.org/10.1111/jbi.13861>.
- Martínez-Freiría, F., Sillero, N., Lizana, M., Brito, J.C., 2008. GIS-based niche models identify environmental correlates sustaining a contact zone between three species of european vipers. *Divers. Distrib.* 14, 452–461.
- Martínez-Freiría, F., Tarroso, P., Rebelo, H., Brito, J.C., 2016. Contemporary niche contraction affects climate change predictions for elephants and giraffes. *Divers. Distrib.* 22, 432–444. <https://doi.org/10.1111/ddi.12406>.
- Martínez-Freiría, F., Velo-Antón, G., Brito, J.C., 2015. Trapped by climate: interglacial refuge and recent population expansion in the endemic iberian adder viper *seaneis*. *Divers. Distrib.* 21, 331–344. <https://doi.org/10.1111/ddi.12265>.
- Martínez-Meyer, E., Peterson, A.T., 2006. Conservatism of ecological niche characteristics in north american plant species over the pleistocene-to-recent transition. *J. Biogeogr.* 10, 1779–1789.
- Martínez-Meyer, E., Peterson, A.T., Hargrove, W.W., 2004. Ecological niches as stable distributional constraints on mammal species, with implications for pleistocene extinctions and climate change projections for biodiversity. *Glob. Ecol. Biogeogr.* 4, 305–314.
- Mason, S.J., Graham, N.E., 2002. Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: statistical significance and interpretation. *Q. J. R. Meteorol. Soc.* 128 <https://doi.org/10.1256/003590002320603584>.
- Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to maxent for modeling species' distributions: what it does, and why inputs and settings matter. *Ecograph. (Cop.)* 36, 1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>.
- Mesgaran, M.B., 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. *Divers. Distrib.* 20 <https://doi.org/10.1111/ddi.12209>.
- Meyer, C., Jetz, W., Guralnick, R.P., Fritz, S.A., Kreft, H., 2016. Range geometry and socio-economics dominate species-level biases in occurrence information. *Glob. Ecol. Biogeogr.* 25, 1181–1193. <https://doi.org/10.1111/geb.12483>.
- Meynard, C.N., Quinn, J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *J. Biogeogr.* 8, 1455–1469.
- Miller, M.E., Hui, S.L., Tierney, W.M., 1991. Validation techniques for logistic regression models. *Stat. Med.* 10, 1213–1226. <https://doi.org/10.1002/sim.4780100805>.
- Mitchell, P.J., Monk, J., Laurenson, L., 2017. Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Meth. Ecol. Evol.* 8, 12–21. <https://doi.org/10.1111/2041-210X.12645>.
- Moisen, G.G., Frescino, T.S., 2002. Comparing five modelling techniques for predicting forest characteristics. *Ecol. Modell.* 2–3, 209–225.
- Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. *Biometrik.* 37, 17–23.
- Moran, E.V., Hartig, F., Bell, D.M., 2016. Intraspecific trait variation across scales: implications for understanding global change responses. *Glob. Chang. Biol.* 22, 137–150. <https://doi.org/10.1111/gcb.13000>.
- Muñoz, A.R., Real, R., 2006. Assessing the potential range expansion of the exotic monk parakeet in spain. *Divers. Distrib.* 6, 656–665.
- Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P., 2014. ENMeval: an r package for conducting spatially independent evaluations and estimating optimal model complexity for maxent ecological niche models. *Meth. Ecol. Evol.* 5, 1198–1205. <https://doi.org/10.1111/2041-210X.12261>.
- Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible r platform for species distribution modelling. *Ecograph. (Cop.)* 39, 368–375. <https://doi.org/10.1111/ecog.01881>.
- Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., 2014. Where is positional uncertainty a problem for species distribution modelling? *Ecograph. (Cop.)* 37, 191–203. <https://doi.org/10.1111/j.1600-0587.2013.00205.x>.
- Naimi, B., Skidmore, A.K., Groen, T.A., Hamm, N.A.S., 2011. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *J. Biogeogr.* 38, 1497–1509. <https://doi.org/10.1111/j.1365-2699.2011.02523.x>.
- Nenzen, H.K., Araújo, M.B., 2011. Choice of threshold alters projections of species range shifts under climate change. *Ecol. Modell.* 222, 3346–3354. <https://doi.org/10.1016/j.ecolmodel.2011.07.011>.
- New, M., Lister, D., Hulme, M., Makin, I., 2002. A high-resolution data set of surface climate over global land areas. *Clim. Res.* 1, 1–25.
- Niamir, A., Skidmore, A.K., Toxopeus, A.G., Real, R., 2016. Use of taxonomy to delineate spatial extent of atlas data for species distribution models. *Glob. Ecol. Biogeogr.* 25, 227–237. <https://doi.org/10.1111/geb.12405>.
- Nix, H.A., 1986. A biogeographic analysis of australian elapid snakes. In: Longmore, R. (Ed.), *Atlas of Elapid Snakes of Australia*, Australian Flora and Fauna Series Number 7. Australian Government Publishing Service, Canberra, pp. 4–15.
- Nogués-Bravo, D., Rodríguez, J., Hortal, J., Batra, P., Araújo, M.B., 2008. Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biol.* 4, e79.
- Nogués-Bravo, D., 2009. Predicting the past distribution of species climatic niches. *Glob. Ecol. Biogeogr.* 5, 521–531.
- Olivera-Gomez, L.D., Mellink, E., 2005. Distribution of the antillean manatee (*trichechus manatus*) as a function of habitat characteristics, in bahia de chetumal, mexico. *Biol. Conserv.* 1, 127–133.
- Ørsted, I.V., Ørsted, M., 2019. Species distribution models of the spotted wing drosophila (*drosophila suzukii*, diptera: drosophilidae) in its native and invasive range reveal an ecological niche shift. *J. Appl. Ecol.* 56, 423–435. <https://doi.org/10.1111/1365-2664.13285>.
- Pagel, J., Schurr, F.M., 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Glob. Ecol. Biogeogr.* 21, 293–304. <https://doi.org/10.1111/j.1466-8238.2011.00663.x>.
- Papes, M., Gaubert, P., 2007. Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (mammalia, carnivora) across two continents. *Divers. Distrib.* 6, 890–902.
- Pasetto, D., Arenas-Castro, S., Bustamante, J., Casagrandi, R., Chrysoulakis, N., Cord, A. F., Dittich, A., Domingo-Marimon, C., El Serafy, G., Karnieli, A., Kordelas, G.A., Manakos, I., Mari, L., Monteiro, A., Palazzi, E., Poursanidis, D., Rinaldo, A., Terzago, S., Ziemba, A., Ziv, G., 2018. Integration of satellite remote sensing data in ecosystem modelling at local scales: practices and trends. *Meth. Ecol. Evol.* 9, 1810–1821. <https://doi.org/10.1111/2041-210X.13018>.
- Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Modell.* 225–245.
- Pearson, R.G., 2007. Species' distribution modeling for conservation educators and practitioners. *Synth. Am. Museum Nat. Hist.*
- Pereira, P.F., Barbosa, A.M., Godinho, C., Salgueiro, P.A., Silva, R.R., Lourenço, R., 2020. The spread of the red-billed leiothrix (*leiothrix lutea*) in europe: the conquest by an overlooked invader? *Biol. Invasion.* 22, 709–722. <https://doi.org/10.1007/s10530-019-02123-5>.
- Pérez de Lanuza, G., Sillero, N., Carretero, M.Á., 2018. Climate suggests environment-dependent selection on lizard colour morphs. *J. Biogeogr.* 45, 2791–2802. <https://doi.org/10.1111/jbi.13455>.
- Peterson, A.T., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecol. Modell.* 117, 159–164.
- Peterson, A.T., Papes, M., Soberón, J., 2015. Mechanistic and correlative models of ecological niches. *Eur. J. Ecol.* 1, 28–38. <https://doi.org/10.1515/eje-2015-0014>.
- Peterson, A.T., Soberón, J., 2012. Species distribution modeling and ecological niche modeling: getting the concepts right. *Nat. Conservação* 10, 102–107.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. *Ecological Niche and Geographical Distributions*. Princeton University Press, New Jersey.
- Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., Guisan, A., 2017. Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Glob. Ecol. Biogeogr.* 26, 275–287. <https://doi.org/10.1111/geb.12530>.
- Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., Blair, M.E., 2017. Opening the black box: an open-source release of maxent. *Ecograph. (Cop.)* 40, 887–893. <https://doi.org/10.1111/ecog.03049>.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 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. *Ecol. Appl.* 19, 181–197.
- Phillips, S.J., Elith, J., 2010. POC plots: calibrating species distribution models with presence-only data. *Ecol.* 91, 2476–2484.
- Porter, W.P., Mitchell, J.W., Beckman, W.A., DeWitt, C.B., 1973. Behavioral implications of mechanistic ecology. *Oecologi.* 13, 1–54. <https://doi.org/10.1007/BF00379617>.

- Pulido-Pastor, A., Márquez, A.L., García-Barros, E., Real, R., 2018. Identification of potential source and sink areas for butterflies on the iberian peninsula. *Insect Conserv. Divers.* 11, 479–492. <https://doi.org/10.1111/icad.12297>.
- Pulido-Pastor, A., Márquez, A.L., Guerrero, J.C., García-Barros, E., Real, R., 2021. Metapopulation patterns of iberian butterflies revealed by fuzzy logic. *Insects* 12, 1–15. <https://doi.org/10.3390/insects12050392>.
- Pulliam, H.R., 1988. Sources, sinks, and population regulation. *Am. Nat.* 132, 652–661.
- Raés, N., 2012. Partial versus full species distribution models. *Nat. Conserv.* 10, 127–138. <https://doi.org/10.4322/natcon.2012.020>.
- Raés, N., ter Steege, H., 2007. A null-model for significance testing of presence-only species distribution models. *Ecograph. (Cop.)* 30, 727–736.
- Ramp, D., Caldwell, J., Edwards, K.A., Warton, D., Croft, D.B., 2005. Modelling of wildlife fatality hotspots along the snowy mountain highway in new south wales. *Aus. Biol. Conserv.* 4, 474–490.
- Real, R., Barbosa, A.M., Martínez-Solano, I., García-Paris, M., 2005. Distinguishing the distributions of two cryptic frogs (anura: discoglossidae) using molecular data and environmental modeling. *Can. J. Zool.* 4, 536–545.
- Real, R., Barbosa, A.M., Vargas, J.M., 2006. Obtaining environmental favourability functions from logistic regression. *Environ. Ecol. Stat.* 2, 237–245.
- Real, R., Márcia Barbosa, A., Bull, J.W., 2017. Species distributions, quantum theory, and the enhancement of biodiversity measures. *Syst. Biol.* 66, 453–462. <https://doi.org/10.1093/sysbio/syw072>.
- Regos, A., Arenas-Castro, S., Tapia, L., Domínguez, J., Honrado, J.P., 2021. Using remotely sensed indicators of primary productivity to improve prioritization of conservation areas for top predators. *Ecol. Indic.* 125 <https://doi.org/10.1016/j.ecolind.2021.107503>.
- Romero, D., Olivero, J., Márquez, A.L., Báez, J.C., Real, R., 2014. Uncertainty in distribution forecasts caused by taxonomic ambiguity under climate change scenarios: a case study with two newt species in mainland Spain. *J. Biogeogr.* 41, 111–121. <https://doi.org/10.1111/jbi.12189>.
- Royle, J.A., Chandler, R.B., Yackulic, C., Nichols, J.D., 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Meth. Ecol. Evol.* 3, 545–554. <https://doi.org/10.1111/j.2041-210X.2011.00182.x>.
- Rummukainen, M., 2010. State-of-the-art with regional models. *WIREs Clim. Chang.* 1, 82–96. <https://doi.org/10.1002/wcc.008>.
- Samy, G., Chavan, V., Arino, A.H., Otegui, J., Hobern, D., Sood, R., Robles, E., 2013. Content assessment of the primary biodiversity data published through gbif network: status, challenges and potentials. *Biodivers. Informatics* 8, 94–172. <https://doi.org/10.17161/bi.v8i2.4124>.
- Sánchez-Montes, G., Recuero, E., Barbosa, A.M., Martínez-Solano, I., 2019. Complementing the pleistocene biogeography of european amphibians: testimony from a southern atlantic species. *J. Biogeogr.* 46, 568–583. <https://doi.org/10.1111/jbi.13515>.
- Santos, H., Rodrigues, L., Jones, G., Rebelo, H., 2013. Using species distribution modelling to predict bat fatality risk at wind farms. *Biol. Conserv.* 157, 178–186. <https://doi.org/10.1016/j.biocon.2012.06.017>.
- Santos, X., Brito, J.C., Caro, J., Abril, A.J., Lorenzo, M., Sillero, N., Pleguezuelos, J.M., 2009. Habitat suitability, threats and conservation of isolated populations of the smooth snake (*Coronella austriaca*) in the southern iberian peninsula. *Biol. Conserv.* 142, 344–352.
- Senay, S.D., Worner, S.P., Ikeda, T., 2013. Novel three-step pseudo-absence selection technique for improved species distribution modelling. *PLoS ONE* 8. <https://doi.org/10.1371/journal.pone.0071218>.
- Seoane, J., Carrascal, L.M., Alonso, C.L., Palomino, D., 2005. Species-specific traits associated to prediction errors in bird habitat suitability modelling. *Ecol. Modell.* 2–4, 299–308.
- Shcheglovitova, M., Anderson, R.P., 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. *Ecol. Modell.* 269, 9–17. <https://doi.org/10.1016/j.ecolmodel.2013.08.011>.
- Sillero, N., 2011. What does ecological modelling model? a proposed classification of ecological niche models based on their underlying methods. *Ecol. Modell.* 222, 1343–1346.
- Sillero, N., 2010. Modelling new suitable areas for *hyla meridionalis* in a current and future expansion scenario. *Amphibia-Reptilia* 31, 37–50.
- Sillero, N., 2008. Amphibian mortality levels on spanish country roads: descriptive and spatial analysis. *Amphibia-Reptilia* 29, 337–347. <https://doi.org/10.1163/156853808785112066>.
- Sillero, N., Barbosa, A.M., 2021. Common mistakes in ecological niche models. *Int. J. Geogr. Inf. Sci.* 35, 213–226. <https://doi.org/10.1080/13658816.2020.1798968>.
- Sillero, N., Barbosa, A.M., Martínez-Freiría, F., Real, R., 2010. Los modelos de nicho ecológico en la herpetología ibérica: pasado, presente y futuro. *Bol. la Asoc. Herpetol. Esp.* 21, 2–24.
- Sillero, N., Bonardi, A., Corti, C., Creemers, R., Crochet, P., Ficetola, G.F., Kuzmin, S., Lymberakis, P., Pous, P.De, Sindaco, R., Speybroeck, J., Toxopeus, B., Vieites, D.R., Vences, M., 2014. Updated distribution and biogeography of amphibians and reptiles of europe. *Amphibia-Reptilia* 35, 1–31. <https://doi.org/10.1163/15685381-00002935>.
- Sillero, N., Carretero, M.A., 2013. Modelling the past and future distribution of contracting species. the iberian lizard *Podarcis carbonelli* (squamata: lacertidae) as a case study. *Zool. Anz.* 252, 289–298. <https://doi.org/10.1016/j.jcz.2012.08.004>.
- Sillero, N., Celaya, L., Martín-Alfageme, S., 2005. Using gis to make an atlas: a proposal to collect, store, map and analyse chorological data for. *Herpetofauna. Rev. Española Herpetol.* 19, 87–101.
- Sillero, N., Skidmore, A.K., Toxopeus, A.G., Brito, J.C., 2009. Biogeographical patterns derived from remote sensing variables: the amphibians and reptiles of the iberian peninsula. *Amphibia-Reptilia* 30, 185–206. <https://doi.org/10.1163/156853809788201207>.
- Silva-Rocha, I., Salvi, D., Sillero, N., Mateo, J.A., Carretero, M.A., 2015. Snakes on the balearic islands: an invasion tale with implications for native biodiversity conservation. *PLoS ONE* 10, e0121026. <https://doi.org/10.1371/journal.pone.0121026>.
- Smith, A.B., 2013. On evaluating species distribution models with random background sites in place of absences when test presences disproportionately sample suitable habitat. *Divers. Distrib.* 19, 867–872. <https://doi.org/10.1111/ddi.12031>.
- Soberón, J., 2007. Grinnellian and eltonian niches and geographic distributions of species. *Ecol. Lett.* 12, 1115–1123.
- Soberón, J., Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species distributional areas. *Biodivers. Informatic.* 2005, 1–10.
- Sofaer, H.R., Hoeting, J.A., Jarnevich, C.S., 2019a. The area under the precision-recall curve as a performance metric for rare binary events. *Meth. Ecol. Evol.* 10, 565–577. <https://doi.org/10.1111/2041-210X.13140>.
- Sofaer, H.R., Jarnevich, C.S., Pearse, I.S., Smyth, R.L., Auer, S., Cook, G.L., Edwards, T.C., Guala, G.F., Howard, T.G., Morissette, J.T., Hamilton, H., 2019b. Development and delivery of species distribution models to inform decision-making. *Biosci.* 69, 544–557. <https://doi.org/10.1093/biosci/biz045>.
- Stockwell, D.R.B., Noble, I.R., 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Math. Comput. Simul.* 5–6, 385–390.
- Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Modell.* 1, 1–13.
- Sullivan, M.J.P., Davies, R.G., Reino, L., Franco, A.M.A., 2012. Using dispersal information to model the species-environment relationship of spreading non-native species. *Meth. Ecol. Evol.* 3, 870–879. <https://doi.org/10.1111/j.2041-210X.2012.00219.x>.
- 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. <https://doi.org/10.1371/journal.pone.0055158>.
- Tarros, P., Carvalho, S.B., Brito, J.C., 2012. Simapse - simulation maps for ecological niche modelling. *Meth. Ecol. Evol.* 3, 787–791. <https://doi.org/10.1111/j.2041-210X.2012.00210.x>.
- Tarros, P., Pereira, R.J., Martínez-Freiría, F., Godinho, R., Brito, J.C., 2014. Hybridization at an ecotone: ecological and genetic barriers between three iberian vipers. *Mol. Ecol.* 23, 1108–1123. <https://doi.org/10.1111/mec.12671>.
- Tessarolo, G., Rangel, T.F., Araújo, M.B., Hortal, J., 2014. Uncertainty associated with survey design in species distribution models. *Divers. Distrib.* 20, 1258–1269. <https://doi.org/10.1111/ddi.12236>.
- Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C., Guisan, A., 2014. Measuring the relative effect of factors affecting species distribution model predictions. *Meth. Ecol. Evol.* 5, 947–955. <https://doi.org/10.1111/2041-210X.12203>.
- Thuiller, W., 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Glob. Chang. Biol.* 10, 1353–1362.
- Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD-a platform for ensemble forecasting of species distributions. *Ecograph. (Cop.)* 32, 369–373.
- Tobler, A.W.R., 1970. A computer movie simulating urban growth in the detroit region. *Econ. Geogr.* 46, 234–240.
- Torreblanca, E., Caminàs, J.A., Macías, D., García-Barcelona, S., Real, R., Báez, J.C., 2019. Using opportunistic sightings to infer differential spatio-temporal use of western mediterranean waters by the fin whale. *PeerJ* 2019, 1–20. <https://doi.org/10.7717/peerj.6673>.
- Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., Kadmon, R., 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Divers. Distrib.* 4, 397–405.
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., De Clerck, O., 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Glob. Ecol. Biogeogr.* 21, 272–281. <https://doi.org/10.1111/j.1466-8238.2011.00656.x>.
- 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. *Meth. Ecol. Evol.* 10, 225–232. <https://doi.org/10.1111/2041-210X.13107>.
- Vale, C.G., Campos, J.C., Silva, T.L., Gonçalves, D.V., Sow, A.S., Martínez-Freiría, F., Boratynski, Z., Brito, J.C., 2016. Biogeography and conservation of mammals from the west sahara-sahel: an application of ecological niche-based models and gis. *Hystrix, Ital. J. Mammal.* 27, 1–10. <https://doi.org/10.4404/HYSTRIX.27.1-11659>.
- Vale, C.G., Ferreira da Silva, M.J., Campos, J.C., Torres, J., Brito, J.C., 2015. Applying species distribution modelling to the conservation of an ecologically plastic species (papio papio) across biogeographic regions in west africa. *J. Nat. Conserv.* 27, 26–36. <https://doi.org/10.1016/j.jnc.2015.06.004>.
- Vale, C.G., Tarros, 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. *Divers. Distrib.* 20, 20–33. <https://doi.org/10.1111/ddi.12115>.
- van Proosdij, A.S.J., Sosef, M.S.M., Wieringa, J.J., Raes, N., 2016. Minimum required number of specimen records to develop accurate species distribution models. *Ecograph. (Cop.)* 39, 542–552. <https://doi.org/10.1111/ecog.01509>.
- VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecol. Modell.* 220, 589–594.
- Varela, S., Anderson, R.P., García-Valdés, R., Fernández-González, F., 2014. Environmental filters reduce the effects of sampling bias and improve predictions of

- ecological niche models. *Ecograph. (Cop.)* 37, 1084–1091. <https://doi.org/10.1111/j.1600-0587.2013.00441.x>.
- Varela, S., Lima-Ribeiro, M.S., Terribile, L.C., 2015. A short guide to the climatic variables of the last glacial maximum for biogeographers. *PLoS ONE* 10, e0129037. <https://doi.org/10.1371/journal.pone.0129037>.
- Varela, S., Rodríguez, J., Lobo, J.M., 2009. Is current climatic equilibrium a guarantee for the transferability of distribution model predictions? a case study of the spotted hyena. *J. Biogeogr.* 9, 1645–1655.
- Vega, G.C., Pertierra, L.R., Olalla-Tárraga, M.Á., 2017. Data descriptor: merraclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling background & Summar. *Sci. Data* 4, 1–11. <https://doi.org/10.1038/sdata.2017.78>.
- Veloz, S.D., 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *J. Biogeogr.* 36, 2290–2299. <https://doi.org/10.1111/j.1365-2699.2009.02174.x>.
- Walker, P.A., Cocks, K.D., 1991. Habitat - a procedure for modeling a disjoint environmental envelope for a plant or animal species. *Glob. Ecol. Biogeogr. Lett.* 1, 108–118. <https://doi.org/10.2307/2997706>.
- Warren, D.L., Glor, R.E., Turelli, M., 2010. ENMTTools: a toolbox for comparative studies of environmental niche models. *Ecograph. (Cop.)* 33, 607–611. <https://doi.org/10.1111/j.1600-0587.2009.06142.x>.
- Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in maxent: the importance of model complexity and the performance of model selection criteria. *Ecol. Appl.* 21, 335–342.
- Warren, D.L., Wright, A.N., Seifert, S.N., Shaffer, H.B., 2014. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 california vertebrate species of concern. *Divers. Distrib.* 20, 334–343. <https://doi.org/10.1111/ddi.12160>.
- Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. *Proc. Natl. Acad. Sci.* 106, 19729–19736.
- Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R., Stewart, A.M., 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanogr.* 16, 120–127.
- Wilson, P.D., 2009. Guidelines for computing maxent model output values from a lambdas file. *J. Mach. Learn. Resear.*
- Wint, G.R.W., Robinson, T.P., Bourn, D.M., Durr, P., Hay, S.I., Randolph, S.E., Rogers, D. J., 2002. Mapping bovine tuberculosis in great britain using environmental data. *Trend. Microbiol.* 10, 441–444.
- Wintle, B.A., Elith, J., Potts, J.M., 2005. Fauna habitat modelling and mapping: a review and case study in the lower hunter central coast region of nsw. *Aus. Ecol.* 719–738.
- 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. <https://doi.org/10.1186/1472-6785-9-8>.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Group, N.P.S. D.W., 2008. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 5, 763–773.
- Yañez-Arenas, C., Peterson, A.T., Mokondoko, P., Rojas-Soto, O., Martínez-Meyer, E., 2014. The use of ecological niche modeling to infer potential risk areas of snakebite in the mexican state of veracruz. *PLoS ONE* 9, e100957. <https://doi.org/10.1371/journal.pone.0100957>.
- Yañez-Arenas, C., Townsend Peterson, A., Rodríguez-Medina, K., Barve, N., 2016. Mapping current and future potential snakebite risk in the new world. *Clim. Chang.* 134, 697–711. <https://doi.org/10.1007/s10584-015-1544-6>.
- Yesson, C., Brewer, P.W., Sutton, T., Caithness, N., Pahwa, J.S., Burgess, M., Gray, W.A., White, R.J., Jones, A.C., Bisby, F.A., Culham, A., 2007. How global is the global biodiversity information facility? *PLoS ONE* 11, e1124.
- Yu, F., Wang, T., Groen, T.A., Skidmore, A.K., Yang, X., Ma, K., Wu, Z., 2019. Climate and land use changes will degrade the distribution of rhododendrons in china. *Sci. Tot. Environ.* 659, 515–528. <https://doi.org/10.1016/j.scitotenv.2018.12.223>.
- Zacarias, D., Loyola, R., 2019. Climate change impacts on the distribution of venomous snakes and snakebite risk in mozambique. *Clim. Chang.* 152, 195–207. <https://doi.org/10.1007/s10584-018-2338-4>.
- Zurell, D., 2017. Integrating demography, dispersal and interspecific interactions into bird distribution models. *J. Avian. Biol.* 48, 1505–1516. <https://doi.org/10.1111/jav.01225>.
- Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Aroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitão, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., Merow, C., 2020. A standard protocol for reporting species distribution models. *Ecograph. (Cop.)*. ecog 04960. <https://doi.org/10.1111/ecog.04960>.
- Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. *Meth. Ecol. Evol.* 1, 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>.
- Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* 4, 561–577.