UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year : 2019

Uncertainty, errors and virtual ecology: using artificial data to improve species distribution models

Fernandes Rui

Fernandes Rui, 2019, Uncertainty, errors and virtual ecology: using artificial data to improve species distribution models

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_24435437EAAB0

Droits d'auteur
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Département d'écologie et évolution

Uncertainty, errors and virtual ecology: using artificial data to improve species distribution models

Thèse de doctorat ès sciences de la vie (PhD)
Écologie et Évolution

Présentée à la
Faculté de biologie et de médecine de l'Université de Lausanne
Par

Rui FERNANDES

Biologiste diplômé et Master en écologie de l'Université de Porto, Portugal

Jury

Prof. Yves Poirier, Président
Prof. Antoine Guisan, Directeur de thèse
Dr. Laure Gallien, expert
Dr. Jonathan Lenoir, expert

Lausanne 2018

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Yves	**Poirier**
Directeur·rice de thèse	Monsieur	Prof.	Antoine	**Guisan**
Experts·es	Madame	Dre	Laure	**Gallien**
	Monsieur	Dr	Jonathan	**Lenoir**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Rui Filipe Freitas Fernandes

Master Universidade do Porto, Portugal

intitulée

**Uncertainty, errors and virtual ecology:
using artificial data to improve
species distribution models**

Lausanne, le 29 novembre 2018

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Yves Poirier

ACKNOWLEDGEMENTS

I would like to start by thanking my supervisor Antoine Guisan for giving me the chance of trying a new and challenging field that virtual ecology entails, and for helping me along the way to become a better researcher and professional. I will always be grateful to him for the greatest opportunity a student can have. I will also like to thank Daniel Scherrer, who not only was my office mate, but also like a second supervisor and a very good friend. I wouldn’t have made it without your help or at least it wouldn’t have been as fun.

To all my friends and colleagues at ECOSPAT, I will never forget how well I was received and how welcome you made me feel. A special thanks to Olivier Bröennimann and Heidi Mod for being such awesome friends. Thanks to Blaise Petitpierre for the programming support and all the talks about movies and TV shows, to Aline Buri for sharing this PhD experience with me and to Rubén, Manuela, Imanol and Frank for your friendship. Thanks also to the borrowed member of the group, Martha, who was always available to go get coffee for all of us. Finally, thanks to my extraordinary friend Valeria Di Cola for always telling me what I needed to hear, for always giving me good advice and for being the mother of one of my favourite persons in the world. To my friend Gabi, thanks for being like a big/older brother to me and thanks for all the support. Finally, thanks to all the DEE members, specially to Susana, Diana, Vera and Ana (aka As Tugas), but also to Oriane, Thomas, Guillaume, Robbie, Lucia and Rubén for all the comradery.

To all my friends in Portugal, gladly too many to name here, I thank you for always making me feel welcome back in Porto and Castelo de Paiva. We might not see each other as often as we would like but our friendship is never weakened. A special thanks to my dearest friend Simão for giving me a place to stay when I arrived in Switzerland, for sharing an apartment with me and for helping me with all things Swiss.

Thanks to all my family, and specially to my mom and dad, who always gave me everything that I needed, for constantly supporting me and for always being there for me. I’m very happy for being your son 😉

Finally, to Filipa…my favourite thing in the whole world, the best part of my day and with whom I want to share every little bit of every single day…LOVE YOU!

ABSTRACT

With the growing pressures exerted by anthropogenic activities (e.g. land-use changes, habitat fragmentation, greenhouse gas emissions) and environmental changes (e.g. climate change, biological invasions), biodiversity is being threatened worldwide. It is therefore important to sufficiently understand which factors influence the distribution and composition of species assemblages, develop tools allowing us to accurately predict them under current and future environmental conditions.

Species distribution models (SDMs) are especially useful to tackle these challenges since they allow the modelling of the distribution of species and their assemblages at different spatial and temporal scales. This is done by simply relating species observations with environmental conditions where they occur. However, different factors (e.g. sample size, modelling technique) and errors/bias (i.e. false presences/absences) were shown to affect the prediction accuracy of single species and assemblage SDMs (i.e. S-SDMs). SDMs can also provide biased projections when predicting to regions or time periods with environmental conditions outside the range of data used for model calibration (i.e. model transferability) or when that data doesn’t capture the full conditions occupied by the species (i.e. truncated datasets). While the majority of SDMs use real species data, it is important to assess their accuracy by having complete control of the data and factors influencing species distributions, hence the use of virtual or simulated species.

In the first chapter of my thesis, I used virtual species data to test SDM/S-SDMs and determine the degree to which different types and levels of errors in species data (i.e. false presences or absences) affect the predictions of individual species models, and how this is reflected in metrics that are frequently used to evaluate the prediction accuracy of SDMs. I found that interpretation of models’ performance depended on the data and metrics used to evaluate them, with model performance being more affected by false positives. In the second chapter, I assessed how different factors (sample size, sampling method, sampling prevalence, modelling technique and thresholding method) affect the prediction accuracy of S-SDMs. I found that prediction accuracy is mostly affected by modelling technique followed by sample size and that a ‘plot-like’ sampling method is recommended when sampling species data (i.e. best approximation of the species’ true prevalence).

In my third chapter I tested the potential causes that increasingly truncated datasets have on the predictive accuracy of species assemblages and if the variables used to calibrate the models also influence that accuracy, finding that the degree of truncation has more influence on species with wide realized niches. Finally, on my last main chapter, I tested and compared how accurate different modelling strategies are at predicting species assemblages under current and future climatic conditions, assessing their transferability. I found that when using presence/pseudo-absence data, all the strategies failed to predict accurate species assemblages, being better when presence-absence data is used (under current environmental conditions).

KEYWORDS: Artificial data, virtual ecology, species distribution models, assemblage modelling, stacked-SDMs, evaluation metrics, predictive accuracy, uncertainty, errors, integrating scales, hierarchical modelling, truncated datasets, Swiss Alps, Europe.

RÉSUMÉ

La biodiversité est actuellement mondialement menacée par l'augmentation de la pression due aux activités anthropiques (p. ex. changement dans l'utilisation du territoire, fragmentation des habitats, émission de gaz à effet de serre) et aux changements environnementaux (p. ex. changements climatiques, invasions biologiques). Il est donc capital de comprendre les facteurs influençant la distribution et la composition des assemblages d'espèces ainsi que de développer des outils pour les prédire précisément autant dans des conditions environnementales actuelles que future. Les modèles prédictifs de distribution (MPDs) sont des outils particulièrement utiles pour appréhender ce genre de challenges, car ils permettent de modéliser la distribution des espèces ainsi que leurs assemblages à différentes échelles spatiales et temporelles. Cela peut se faire en reliant des observations d'espèces avec les conditions environnementales dans lesquelles elles se trouvent. Cependant, il a été montré que différent facteurs (p. ex. taille d'échantillonnage, techniques de modélisation) et erreur/biais (c.-à-d. fausses présences/absences) peuvent affecter la qualité des prédictions obtenues lors de la modélisation prédictive de la distribution de simples espèces (MPD) et d'assemblages (S-SDMs). Les MPDs peuvent aussi créer des projections biaisées lorsqu'ils prédisent dans des régions ou des périodes de temps qui possèdent des conditions environnementales en dehors de la gamme de données utilisées lors de la calibration du modèle (c.-à-d. transférabilité du modèle) ou quand les données ne représentent pas l'entier des conditions occupées par l'espèce (c.-à-d. jeu de données tronqué). Bien que la majorité des MPDs utilisent des données d'espèces réelles, il est important de pouvoir évaluer leurs précisions en ayant le contrôle complet des données ainsi que des facteurs pouvant influencer la distribution des espèces. Seul l'utilisation d'espèces virtuelles ou simulées permet d'obtenir ce contrôle total. Dans le premier chapitre de ma thèse, j'ai utilisé des données d'espèces virtuelles afin de déterminer, à l'aide de MPDs/S-SDMs, dans quelle mesure différents types et niveaux d'erreurs dans les données d'espèces (c.-à-d. fausses présences ou absences) pouvaient affecter les prédictions obtenues. J'ai aussi cherché à comprendre comment cela se reflète sur les métriques habituellement utilisées pour évaluer la qualité des prédictions de ces MPDs. J'ai découvert que l'interprétation des performances des modèles dépends des données et des métriques utilisées pour les évaluer. Cette performance est particulièrement affectée par les faux positifs. Dans le second chapitre, j'ai évalué comment différents facteurs (taille d'échantillonnage, méthode d'échantillonnage, prévalence d'échantillonnage, technique de modélisation et méthode de définition des seuils) affectent la qualité des prédictions obtenues à l'aide de S-SDMs. J'ai trouvé que la qualité des prédictions est principalement affectée par les techniques de modélisation, suivie par la taille de l'échantillonnage. Une méthode d'échantillonnage dite « plot-like » est recommandée lors de la récolte de données (c.-à-d. qu'elle donne la meilleure approximation de la réelle prévalence de l'espèce).

Dans mon troisième chapitre, j'ai testé quels pouvaient être les potentiels effets de l'utilisation de jeux de données de plus en plus tronqués sur la qualité des prédictions des assemblages d'espèces ainsi que l'influence des variables utilisées lors de la calibration. Il s'avère que le degré de troncature a plus d'effet sur les espèces ayant une large niche réalisée. Finalement, dans mon dernier chapitre, j'ai testé différentes stratégies de modélisation puis j'ai comparé leur aptitude à prédire des assemblages d'espèces dans des conditions présentes et futures pour évaluer leur transférabilité. J'ai découvert que lors de l'utilisation de données de présences/pseudo-absences, toutes les stratégies échouaient à prédire de manière précise les assemblages. L'utilisation de données de présence/absences a permis, quant à elle, d'obtenir de meilleurs résultats, principalement dans des conditions environnementales présentes.

TABLE OF CONTENTS

PREFACE

The thesis presented here was completed within the scope of the SESAM'ALP project (see Figure 1), which had the objective of overcoming identified limitations when assemblages were predicted using stack species distribution models (S-SDMs). SESAM'ALP focused on four major components: 1) tackling challenges in correctly predicting species assemblages through the use of a virtual ecologist approach; 2) improve model predictions using very high-resolution environmental data as meaningful predictor variables; 3) use a hierarchical approach to integrate models from different spatial scales to try to limit the problem of truncated species responses to some environmental factors; and 4) the consideration of new global change drivers to be included in spatial models, allowing the generation of high-resolution projections over larger extents.

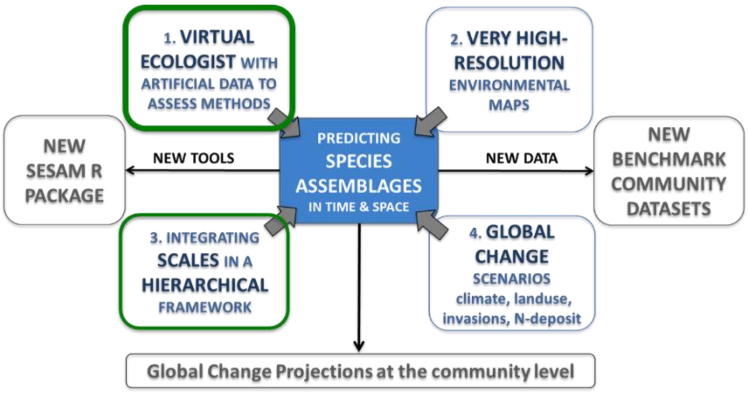


FIGURE 1 The four main components of the SESAM'ALP project and the four main outputs. In the green boxes, the two components in which the work developed in this thesis contributes to study and improve the prediction of species/assemblages in space and time: (1) Virtual ecologist with artificial data to assess methods, and (2) Integrating scales in a hierarchical framework Source: SESAM'ALP proposal, A. Guisan, 2013.

Within those main components, my thesis focused specifically on the implementation of a virtual ecologist approach to assess different aspects in species distribution modelling (i.e. different sources of uncertainty and factors affecting the accuracy of S-SDMs), and in testing the integration of large scales into regional/local scales (e.g. hierarchical models; truncated datasets) (Fig.1, green boxes). Since the research developed in this thesis encompasses different topics (e.g. species distribution modelling, community/assemblage modelling, sources of uncertainty and errors in SDM or the use of simulated data), I discuss some of the topics that I believe are important to understand the thesis chapters. I start the INTRODUCTION by giving a description on the growing impacts that affect biodiversity and what determines species' distributions. Then I provide an overview of the most useful aspects regarding species distribution models (i.e. history, known uses, main assumptions, data requirements and essential building steps). In the next section I focus on the definition of community modelling and on the different strategies used to predict assemblage distributions and composition. I then talk about the several sources of uncertainty and errors know to affect SDMs and how to deal with them. Finally, in the last section I explain why it is important to use virtual species to assess different aspects of species distribution models, making a connection to the objectives and structure of my thesis.

INTRODUCTION

The growing negative impacts affecting biodiversity

Over the past 500 years, but undoubtedly since 1800 (i.e. around the start of the Industrial Revolution), humans' impact on the global environment has become so massive that this global-scale influence has been known and referred to as the Anthropocene (Steffen *et al.*, 2011a; Steffen *et al.*, 2011b; Dirzo *et al.*, 2014). Among the most evident signs of human-driven changes to the environment, climate change is without a doubt the most mentioned and studied (e.g. Pauli *et al.*, 1996; Parmesan & Yohe, 2003; Steffen *et al.*, 2018). The evidence that Earth is warming is clear, with greenhouse gas emissions as the leading cause of that warming, at least since the middle of the past century (IPCC, 2007), with concentrations at the highest level in 800 000 years, according to the World Meteorological Organization (2006). This increase appears to be already enough to warm the planet more than 2ºC in the coming decades (Ramanathan & Feng, 2008). Other risks associated with climate change, like sea-level rise, melting of the polar ice sheets, extreme weather events and shifts in rainfall patterns are also increasingly worrisome (Richardson *et al.*, 2011). However, climate change is only one factor with strong human influence since we are also altering biogeochemical cycles like nitrogen or phosphorus and changing essential water cycles by interrupting river flow from the mountains to the sea (e.g. by building dams).

With all these human activities, we are undoubtedly transforming ecosystems for human use (Millennium Ecosystem Assessment, 2005), having modified already as much as 50% of the terrestrial land cover (McGill *et al.*, 2015), leading to an increasing loss of biodiversity (Barnosky *et al.*, 2011), habitat loss or fragmentation and biotic homogenization (McKinney & Lockwood, 1999; McGill *et al.*, 2015). Declines in biodiversity can further reduce the resilience in communities to environmental changes and alter food-web structures (Olden *et al.*, 2004). The effects of factors like land-use change (e.g. Newbold *et al.*, 2015; Newbold *et al.*, 2016a) or habitat fragmentation (e.g. Fahrig, 2003; Haddad *et al.*, 2015), were shown to be significant drivers of spatial homogenization of communities, intensely reducing biodiversity worldwide, and with additional threats also on the horizon (e.g. population growth; Tilman *et al.*, 2017). This biodiversity loss is leading to what is being called as the sixth major extinction event in Earth's history (e.g. Ceballos & Ehrlich, 2002; Wake & Vredenburg, 2008; Barnosky *et al.*, 2011), with species and populations disappearing and local species abundance declining (see Figure 1).

The observed changes in land cover, together with climate change and the continuous increase in human mobility, amplified the chances of alien species to spread into new locations and become invasive, by facilitating species migration and colonisation (Vilà & Ibáñez, 2011; Chytrý *et al.*, 2012). The impacts of invasive species are so worrying, that when Wilcove *et al.* (1998) quantified and ranked different threats to biodiversity, they found that habitat loss and the spread of alien species were the most important ones, followed by pollution, overexploitation, and diseases.

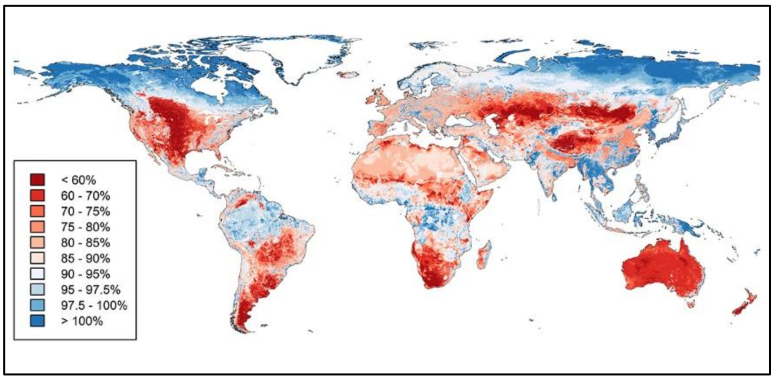


FIGURE 1 Worldwide biodiversity loss, with remaining populations of indigenous species as a percentage of their original populations. The blue areas are within proposed safe limits, and the red regions are beyond (see Newbold *et al.*, 2016b). This map was created from data available in the PREDICTS database (downloaded from http://www.predicts.org.uk/).

Several other studies also researched the effects of global changes on biodiversity, assessing their impacts. For example, Sala *et al.* (2000) developed global scenarios of biodiversity change for the year 2100, identifying land-use and climate change as having the most substantial effects on terrestrial ecosystems. Pereira *et al.* (2010) also analysed different global biodiversity scenarios, indicating that all of them show a decline of biodiversity for this century. As another example, the risk of extinction in amphibians was studied (Wake & Vredenburg, 2008), with the growing pressures of habitat destruction and climate change mostly impacting species with narrow niches, while global warming will also aggravate threats caused by infectious diseases. Finally, according to the latest version of the Red List of Threatened Species, produced by the International Union for Conservation of Nature (IUCN, 2018), it's estimated that from the all the assessed species (see Figure 2), 41% of amphibians, 35% of reptiles, 35% of dicotyledons, 34% of conifers, 25% of mammals and 13% of birds are threatened. All these facts make it clear that biodiversity loss is felt across all life forms and is increasing due to different drivers of global change. Therefore, it is important to understand the factors determining species' distributions across the globe and how the current and future distributions of those species might be affected by drivers of global change.

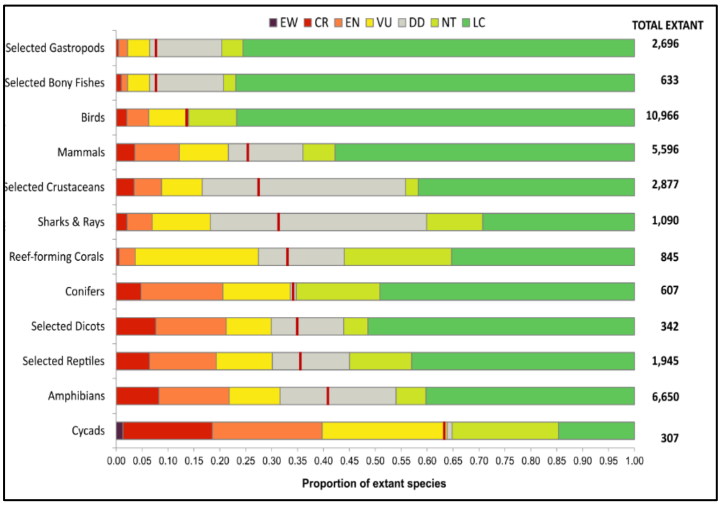


FIGURE 2 The proportion of extant species in *The IUCN Red List of Threatened Species. Version 2018-1* measured in each category for the more comprehensively assessed groups (more 150 species). The numbers to the right of each bar represent the total number of species assessed for each group. Legend: EW (Extinct in the Wild), CR (Crit. Endangered), EN (Endangered), VU (Vulnerable), NT (Near Threatened), DD (Data Deficient), LC (Least Concern). (www.iucnredlist.org/about/summary-statistics).

What determines species’ distribution?

Questions concerning how different plants and animals are distributed across the globe and why different regions of Earth have such diverse and distinct numbers of species (e.g. increase of biodiversity from the poles to the equator) have inspired researchers from different fields to understand these patterns (e.g. like the works of Von Humboldt; Figure 3).

In addition to the study of patterns of climate and vegetation, different researchers also started to focus on the links between abiotic and biotic environments. For example, Clements (1916) theorised that vegetation succession occurs with plant communities moving as a unit to an optimal state that was conditioned by environmental conditions. On the other side of the argument, Gleason (1926) understood that plants follow an individual trajectory, with species abundance optima and their limits being distributed independently across the environmental gradients. Another example is the work developed by Whittaker (1967), where he mapped the environmental gradients shaping plant communities (i.e. response curves), showing the important role taken by abiotic gradients in structuring communities and determining species distributions. Other important work worth mention is the formalization of the fundamental niche concept by Hutchinson (1957), based on work developed prior by Grinnell (1917), defined by a combination of environmental characteristics (i.e. n-dimensional hypervolume) in which species populations can grow. According to the individualistic continuum concept proposed by Gleason (1926), the species abundance optima and their limits are independently distributed along the environmental gradients, with Gause (1936) suggesting that species response functions are bell-shaped in response to environmental factors (see Franklin, 2010 for additional information). Finally, another important contribution to biogeography came from MacArthur, Wilson (1963; 1967) and Preston (1980), with the proposal that number of species in an island is in dynamic equilibrium between the species that arrive to the island and those present that go extinct. Therefore, the number of species would be constant over time, with evolution action as the main driver to increase equilibrium number of species.

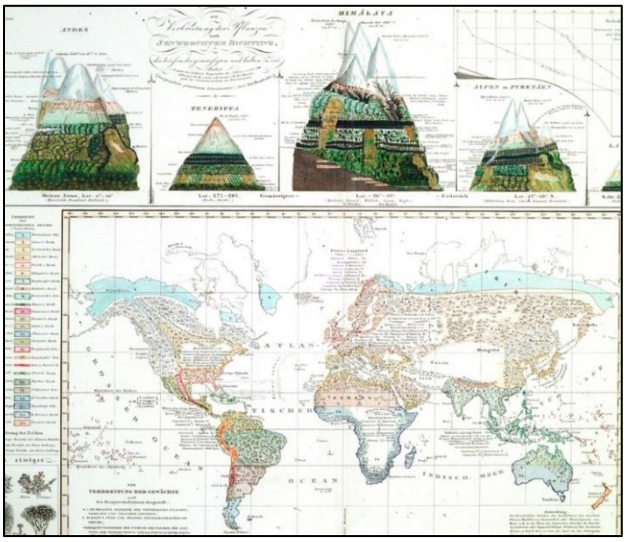


FIGURE 3 Alexander von Humboldt’s Umrisse des Pflanzengeographie (outline sketch of the Geography of Plants), with a global map of vegetation providing information of the spatial extent of different botanical types and sketches of altitudinal variations of several mountain systems (source: H. Berghaus, 1851, Physikalischer Atlas, vol.V, plate No.1. Downloaded from http://journals.openedition.org/cybergeo/docannexe/image/25478/img-12.jpg; accessed August 2018).

Different factors where then shown to affect species' distribution along a gradient or in a specific area: e.g. the type and availability of resources that are used or consumed, patterns of disturbance (like fires or diseases), ability to disperse, the presence or absence of suitable abiotic environmental conditions, and the influence of biotic interactions (Franklin, 2010). However, the main factors (Figure 4) determining the species' spatial patterns can be summarized into three important groups (see Pulliam, 2000; Soberon, 2007; Guisan *et al.*, 2017).

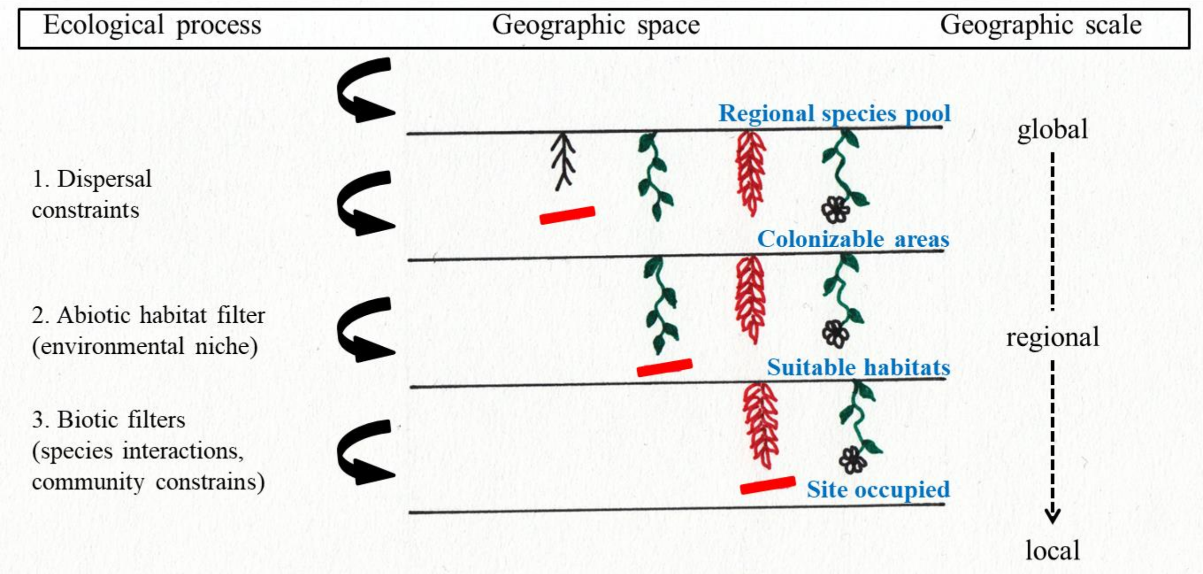


FIGURE 4 Hierarchical view of the three main influences determining species occurrence at a given site: dispersal limitation, abiotic habitat filtering, and biotic filters, with corresponding geographic space at the successive scales, from global to regional to local. Adapted from Guisan *et al.*, 2017.

First, a species needs to be able to spread into new areas and disperse there, either by itself or using external agents to disperse their seedlings. This factor determines which parts of the planet can be accessible to the species, how well it can disperse and overcome different migration barriers (e.g. mountains, rivers or oceans). Second, the abiotic environmental conditions need to be suitable for the species, for its individuals to successfully establish, survive and reproduce. This means that the habitat needs to be suitable for a species to grow and maintain populations, establishing the limits of the environmental niche of the species. Finally, the third factor concerns the biotic interactions between species and the availability of resources. These interactions can be either negative (predation, competition, and parasitism) or positive (commensalism, mutualism), which can modify the suitable conditions determined by the previous factor.

Species distribution modelling – an overview

A brief introduction

In the past 30 years, as biodiversity is under growing pressure by anthropogenic and environmental changes, a tool has become increasingly important to help address several issues in the fields of evolution, biogeography, ecology, or conservation (Guisan & Thuiller, 2005; Franklin, 2010; Peterson, 2011; Guisan *et al.*, 2013; Guisan *et al.*, 2017). That widely used tool is known as species distribution models (SDMs) but is also referred by some as habitat suitability models, ecological niche models or habitat distribution models (Elith & Leathwick, 2009; Guisan *et al.*, 2017, p. 8). Species distribution models are empirical models that statistically relate species observations, usually obtained through field observations or databases, with environmental data (Guisan & Zimmermann, 2000; Elith & Leathwick, 2009). The development of these models frequently starts with observations of species occurrences (i.e. presence, presence-absence or abundance), and with environmental variables that are thought to exert an effect on habitat suitability, ultimately influencing the species distribution (Guisan & Thuiller, 2005). Sometimes, one is unable to obtain species absence data due to limited access to an area, and artificially generated pseudo-absences are often used to circumvent this problem (Lobo & Tognelli, 2011; Barbet-Massin *et al.*, 2012).

There is another type of models, mechanistic models (sometimes called process-based models), that use detailed knowledge about factors limiting species distributions, usually physiological constraints, to make robust predictions (Robertson *et al.*, 2003; Kearney & Porter, 2009). However, in this thesis, we focus our attention primarily on SDMs, which contrary to mechanistic models are static and probabilistic and allow the creation of models and predictions with less demanding data requirements (i.e. widely available species and environmental data).

In this section, the aim is to provide an overview of the most useful aspects regarding species distribution models (i.e. history, uses, main assumptions, data requirements and essential building steps). It is by no means an exhaustive review of all the theories, techniques or approaches that are necessary to understand to build and apply distribution models, but surely enough to understand the work developed in this thesis. Additional information about this topic can be found in several textbooks (e.g. Franklin, 2010; Peterson, 2011; Guisan *et al.*, 2017) and scientific publications (e.g. Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005; Austin, 2007; Elith & Leathwick, 2009).

What are SDMs used for?

With the increasing impacts caused to biodiversity and ecosystems by the different sources of global changes - habitat fragmentation, biological invasions, climate and land-use change - (Chapin *et al.*, 2000; Sala *et al.*, 2000; Cardinale *et al.*, 2012; Steffen *et al.*, 2018), there is a need to improve the understanding of, and associated capacity to model, the main factors driving these changes in the distributions of species, communities or ecosystems (Dawson *et al.*, 2011). This need to explain, understand and even predict the distribution of species across different spaces and times for supporting conservation decisions led to the development and widespread usage of species distribution models (Guisan *et al.*, 2013).

Species distribution models have been used to model species distributions and their changes in a specific region and time for a variety of species in different biological groups: from vascular plants (Pearson *et al.*, 2004; Engler *et al.*, 2011) and lichens (Bolliger *et al.*, 2007; Hespanhol *et al.*, 2015), to birds (Brotons *et al.*, 2004; Bastos *et al.*, 2016), insects (Fleishman *et al.*, 2001; Buse *et al.*, 2007), fishes (Wisz *et al.*, 2015; Ferreira *et al.*, 2016), reptiles (Morán-Ordóñez *et al.*, 2017) or mammals (Rondinini & Boitani, 2006; Morán-Ordóñez *et al.*, 2018). Additionally, SDMs have also been used in numerous applications, such as: the quantification of the environmental niche of species (Luoto *et al.*, 2006; Broennimann *et al.*, 2012; Wisz *et al.*, 2015), the assessment of current and future biological invasions (Broennimann & Guisan, 2008; Vicente *et al.*, 2010; Gallien *et al.*, 2012; Petitpierre *et al.*, 2016), the research on the impacts of climate, land use or other environmental changes on species distributions (Álvarez-Martínez *et al.*, 2014; García-Valdés *et al.*, 2015), the support for the creation of appropriate and effective conservation and management plans (Vicente *et al.*, 2013; Fernandes *et al.*, 2014; Vicente *et al.*, 2016), to increase the accuracy of prediction of rare species (Lomba *et al.*, 2010; Breiner *et al.*, 2015) or the prediction of spatial patterns of species assemblages (Dubuis *et al.*, 2011; Guisan & Rahbek, 2011; D'Amen *et al.*, 2015b). This final application concerning the prediction of species assemblages is a topic that is going to be further developed in the next section (*"From individual species to community modelling"*).

Theories and assumptions

The species niche concept

Despite the relevance of multiple ecological theories, the species niche concept is essential in several areas of ecology (e.g. study of behavior, morphology, physiology or ecosystem functioning; Chase & Leibold, 2003), and is also the basic underlying principle on which species distribution models rely today (Guisan & Zimmermann, 2000). This concept was first defined by Grinnell (1917) as the habitat or environment that a species is capable of occupying, emphasising the role of the environment in shaping species' geographic ranges and local habitat distributions. Hutchinson (1957) improved that concept, defining the niche as a set of conditions and resources required for a species to survive (i.e. *the n-dimensional hypervolume*) and distinguishing between the fundamental and realised niche.

The fundamental niche (physiological tolerance) was defined as the response of species to the environment (resources) where conditions are suitable for a specific population to grow (Pulliam, 2000; Franklin, 2010), in the absence of biotic interactions (e.g. competition, predation or facilitation). The realised niche was described as the ecological or actual niche of a species, which means it comprises the environmental conditions in which a species can survive and reproduce, while also considering the effects of biotic interactions that affect that distribution (Chase & Leibold, 2003; Guisan *et al.*, 2017). The realised niche usually represents a sub-space of the fundamental niche where biotic interactions do not exclude the species. However, this might not always be the case if a “sink” habitat, dispersal limitations or a positive biotic interaction occur (i.e. occurrence of another species is necessary) (Franklin, 2010). This “sink” habitat concept is derived from the source-sink theory (Pulliam, 1988), where source habitat –ranges where local reproduction is higher than mortality– is distinguished from “sink” habitat –when individual is located but not contribute to population growth.

The realised niche (i.e., where a species is present) can then be shaped by biotic interactions that constrain the fundamental niche of a species (i.e. where a species has conditions to be present and is usually observed), conditions. These interactions can be either positive (facilitating a species to occur in sites where environmental conditions are unsuitable) or negative (excluding a species from sites that are suitable). If interactions are present, they can affect the predictability of a species when modelled using environmental predictors only (Araújo & Guisan, 2006; Elith & Leathwick, 2009), but they are rarely considered.

Since SDMs relate observed species presence-absence (or presence-only when absences are not available) to environmental conditions, species distribution models estimate Hutchinson’s realised niche of species (Guisan & Thuiller, 2005; Araújo & Guisan, 2006). Nevertheless, there might be cases when a species is limited by dispersal processes, and the previous definition is not adequate to characterise SDMs. In these cases, the BAM diagram proposed by Soberon (2007) is useful to better reflect the niche being fitted. Under this proposed framework, the species distribution is limited by the available conditions present in a site, the species fundamental niche, its biotic environment and the ability to disperse. One then assumes that if the models are fitted on the observed distribution of the species, the modelled niche should reflect the previously mentioned limitations. When using SDMs, it is then essential to consider their limitations and be aware of common assumptions being made.

Assumptions made when building SDMs

There is a sentence attributed to the statistician George Box, “all models are wrong, but some are useful”, that is frequently used to express the fact that models are simplified versions of the real world. The same is valid for species distribution models, meaning that in order to create and use SDMs, different methodological and theoretical assumptions need to be made (Guisan & Thuiller, 2005;

Araújo & Guisan, 2006; Elith & Leathwick, 2009). While these assumptions were reviewed before (see Franklin, 2010; Peterson, 2011; Guisan *et al.*, 2017), I briefly detail those I believe are the most important ones:

- Species-environment equilibrium: this is one of the most important theoretical assumptions made when using SDMs (see Guisan & Zimmermann, 2000; Elith & Leathwick, 2009; Araujo & Peterson, 2012). Because species data is in most of the cases sampled over a very short period (e.g., field campaign), one only has the species-environment relationship information about that period. It is then assumed that the species modelled are in pseudo-equilibrium with their environment (Guisan & Theurillat, 2000), with the models capturing the realised environmental niche of the species. This can then be used to project the distributions into another period or region while expecting that the species-environment relationship remains unaltered. One then assumes that the species is completely distributed in its entire suitable habitat by the time the data was sampled. However, this assumption does not hold situations like when an invasive species is just starting to be established and spreading into a new area (Broennimann *et al.*, 2007; Petitpierre *et al.*, 2012) or when some existing species is recolonizing an area left empty after a major disturbance (e.g., fires, glaciation; Briani *et al.*, 2004; Normand *et al.*, 2011). Because these species, and especially invasive species, are not at equilibrium with their environment in the new (invaded) range, they should be modelled using either data from their native range (Peterson, 2003) or both ranges (e.g., Gallien *et al.*, 2012; Petitpierre *et al.*, 2016).

- Appropriate statistical methods: in this methodological assumption, one assumes that once a statistical method is chosen, it is the best one to fit the data available. The choice of an incorrect method can however lead to erroneous or uncertain predictions (e.g. Guisan, 2002; Guisan *et al.*, 2002; Fernandes *et al.*, in press) since different statistical models are appropriate or designed to be used with different types of response variables (Guisan & Zimmermann, 2000). Therefore, the correct type and probability distribution of the response variables need to be identified in order to select the appropriate model (see Guisan & Zimmermann, 2000 for details).

- Unbiased species data: another important methodological assumption concerns the reliability of the species data. In SDMs, it is expected that the data represents the wider possible range of the suitable habitat of the species being modelled (Guisan *et al.*, 2017). This means that species data needs to be unbiased, which is usually not possible when the sampling design is poorly done (Graham *et al.*, 2004; Albert *et al.*, 2010). Incorrectly or insufficient sampled data typically results from the fact that data is, in most cases, sampled along communication routes (e.g. roads, rivers or railroads) or in habitats known to be occupied by the species (Kadmon *et al.*, 2004; Edwards *et al.*, 2006; Hortal *et al.*, 2007). So, any bias that might be present in species data can lead to an incomplete quantification of the niche, not accounting for the entire full distribution of species, causing models to wrongly identify all suitable habitats of a species and providing incorrect spatial predictions (Thuiller *et al.*, 2004a). The effects of a specific type of biased species data (i.e. species misidentification), was tested in Chapter 1.1, Part I of this thesis, in which I evaluated the degree to which this bias affects model predictions and how this is reflected in commonly used evaluation metrics.

Methodological steps – a summary

Depending on the type of reasoning followed, the framework and procedures used for modelling species distributions can slightly differ. For example, in her book, Franklin (2010) organized the modelling framework as presented by Austin (2002), divided in three main parts: the ecological model that includes the theories or tested hypotheses; the data model concerning the decision of how data is collected, measured and estimated; and the statistical model, where the methods to implement the calibration and validation are decided. However, in this thesis, a framework was followed using a structure initially proposed by Guisan and Zimmermann (2000) and Guisan and Thuiller (2005), and recently modified by Guisan *et al.* (2017). This framework follows five main steps (Figure 5): (i) Conceptualization; (ii) Data preparation; (iii) Model calibration or fitting; (iv) Model evaluation; and (v) Spatial predictions.

Conceptualization

Before the start of any model building, it is essential to identify all the issues that might require a methodological decision to be taken, to guarantee a smooth modelling procedure in later stages.

It is important to start by defining the questions, hypothesis, and objectives of the study, as well as the species one aims to model. It is then necessary to have some ecological knowledge of the species to model and determine the assumptions one needs to make in order to build a model (e.g. species-environment equilibrium). Other decisions, like which environmental predictors to use, the sampling design necessary to obtain species observations, what kind of models to use and how to evaluate and predict these models need also to be considered at this stage.

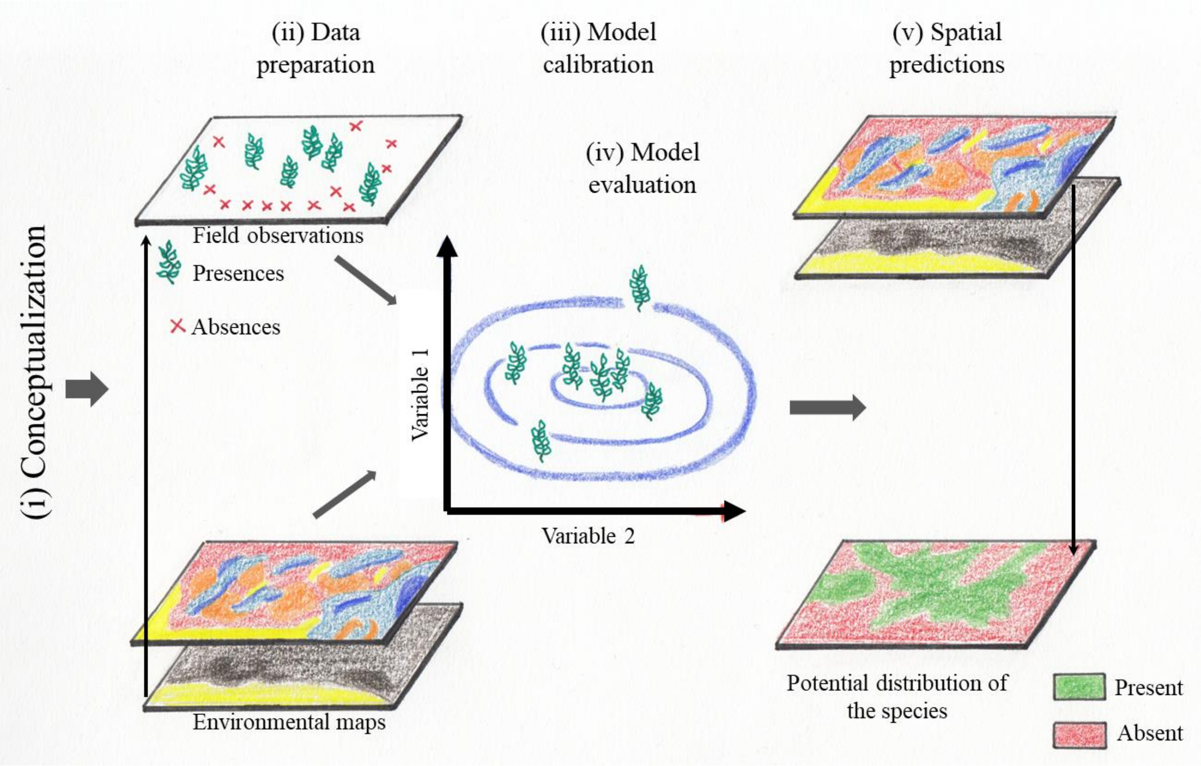


FIGURE 5 Habitat suitability modelling framework (adapted from Guisan *et al.*, 2017): (i) Conceptualization; (ii) Data preparation; (iii) Model calibration; (iv) Model evaluation; and (v) Spatial predictions. See text for details.

Data preparation

After conceptualising the primary objectives of the study and deciding how to achieve them, it is important to prepare the data accordingly. Preparing the necessary environmental predictors is usually one of the first steps taken in data preparation, as these will be used as the explanatory variables in the models. Therefore, the variables should be as proximal as possible (i.e. determine or has a direct impact on a species response), for the model to be the most robust, applicable and with as much ecological meaning as possible (Austin, 2002; Austin, 2007).

Depending on the type of study one intends to conduct and the knowledge of the factors that can explain the distribution of a species, there are a variety of different predictors that can be used (see Mod *et al.*, 2016 for a review of widely used variables for plants): (i) climatic data, that can be obtained in databases like WorldClim (Fick & Hijmans, 2017) or CHELSA (Karger *et al.*, 2017); (ii) land-use or land cover data (e.g. GlobCover - http://due.esrin.esa.int/page_globcover.php - or Corine Land Cover - www.eea.europa.eu/data-and-maps); (iii) topographic data obtained through digital elevation models; or (iv) other spatial vectors like rivers, lakes, roads or administrative areas.

Finally, one needs to determine if enough species observations are available to build the model (see Wisz *et al.*, 2008) and develop additional sampling campaigns if necessary. The sampling design should also take into consideration the appropriate resolution and geographic extents of the study. In the best-case scenario, the predictors' resolution should be the same as the species data (Guisan & Thuiller, 2005), but this might not always be the case (e.g. predictors available at a much coarser grain sizes, different grains between the predictors, species data gathered from atlas have coarser resolutions).

The design should also consider the type of data -presence, presence-absence, abundance- to be sampled and based on which type of sampling design (i.e. random, stratified; Hirzel & Guisan, 2002). Typically, it is possible to use presence-absence data when a proper sampling design is used, representing that way the full probabilistic distribution of a species along the sampled environmental gradients. However, in some cases (e.g. museum collections or GBIF database), observed or reliable absences might not be available, and models can be built using pseudo-absences (see Wisz & Guisan, 2009; Barbet-Massin *et al.*, 2012 for additional information) combined with existing presences.

Model calibration

When all the objectives are delineated, and the necessary data is obtained, one is ready to start calibrating (or fitting) the model. Since species distribution models establish a statistical relationship between species observations and environmental data, different algorithms can be used (see Franklin, 2010 for details; Guisan *et al.*, 2017).

Among the most used algorithms or modelling techniques used, one can highlight envelope approaches (e.g. BIOCLIM, ENFA; Busby, 1991; Hirzel *et al.*, 2002), which are the most straightforward and oldest methods available. Regression approaches, like generalised linear models (GLM; McCullagh & Nelder, 1989) or generalised additive models (GAM; Hastie & Tibshirani, 1990), are the most commonly used in ecology and rely on robust statistical theories. Artificial neural networks (ANN; Lek *et al.*, 1996; Ripley, 1996) and boosting and bagging approaches like random forests (RF; Breiman, 2001) and boosted regression trees (BRT, also known as GBM; Friedman *et al.*, 2000) are also viable and widely used options to calibrate models. Finally, the often used maximum entropy approach (Phillips *et al.*, 2006) also need to be mentioned, which is mostly used to create SDMs with presence-only data, but it was recently shown to be a special type of point Poisson process GLM (Renner & Warton, 2013).

It has been shown that different algorithms can produce models with different performances and predictions (e.g. Elith *et al.*, 2006; Guisan *et al.*, 2007; Graham *et al.*, 2008; Grenouillet *et al.*, 2011), emphasising the importance of correctly choosing the algorithm that better fits the aims of the study. These varying performances can also create models with different projections when trying to identify changes in species distribution under different climatic scenarios (e.g. Thuiller *et al.*, 2004b; Araújo, 2005; Pearson *et al.*, 2006). Because each algorithm has arguments for and against it (compared e.g. in Elith *et al.*, 2006; Elith & Graham, 2009), a commonly used solution is to jointly run several algorithms, using one of the hundreds of programming packages available today. A widely used solution to circumvent the variation in models and jointly run different algorithms is called the ensemble forecasting approach (Araujo & New, 2007), done by assembling (i.e. averaging) predictions from SDMs calibrated with different modelling techniques (Araujo & New, 2007; Franklin, 2010). By combining the different predictions, uncertainty among the individual models is accounted for, allowing for more robust predictions (Araujo & New, 2007; Buisson *et al.*, 2010).

Model evaluation

It is important to note that since models are simplifications of reality, all of them have predictions errors (i.e. any mistake but also statistical variability). Errors in species distribution model predictions can be a consequence of errors in data or model parameterisation (Barry & Elith, 2006; Beale & Lennon, 2012). While this topic is going to be further discussed in the section "*Uncertainty and errors in models*", it is important to know how to evaluate the quality of the models after calibration.

Although there are numerous metrics used to evaluate the performance and accuracy of models, and different choices can be made depending on the type of response variables (Guisan *et al.*, 2017 ; p.242), the most used metric is without doubt the area under the receiver-operating characteristic curve (AUC-ROC) (Fourcade *et al.*, 2018). It is a threshold-independent metric (i.e. comparing observed presence-absence to raw probabilistic predictions) calculated by plotting a model's sensitivity against its false positive rate at all possible thresholds (Hanley & McNeil, 1982), this way measuring the performance of a model in discriminating between species presences and absences (Lobo *et al.*, 2008). Due to some known limitations of AUC, like the dependency on the calibration data or the equal weighting of omission and commission errors (e.g. Lobo *et al.*, 2008; Peterson *et al.*, 2008; Jiménez-Valverde, 2012), other metrics have been proposed to evaluate models. The most used substitutes are to look for the maximum performance of threshold-dependent metrics (see Guisan *et al.*, 2017): Cohen's Kappa (Kappa; Cohen, 2016), that corrects the overall accuracy of model predictions by the accuracy expected to occur by chance; and the True Skill Statistic (TSS; Allouche *et al.*, 2006), that is based on the probability threshold for which the sum of sensitivity and specificity is maximized while also correcting Kappa's dependency on prevalence (see Liu *et al.*, 2005 for details; Freeman & Moisen, 2008).

Ideally these metrics should be calculated using an independent dataset, but since in most cases such dataset is unavailable, a currently used solution involves splitting the original dataset into two parts, one to calibrate the model (usually 70 or 80% of the data) and the other to evaluate it (the remaining % part of the data). Overall this cross-validation procedure is repeated a certain number of times, and the evaluation results averaged.

Spatial predictions

In the final step, models can either be predicted into the same area and time or projected into a new one (e.g. to determine the effects of climate change or predict the spread of invasive species). First, it is essential to determine if the same environmental conditions are observed in both spaces or times (i.e. are the conditions used to fit the model comparable to the ones used for projecting it). This leads to two complementary questions, related to the available environment in each time/space (realized environment; Jackson & Overpeck, 2000) and how analogue the two environments are (i.e. environmental analogy; Fitzpatrick & Hargrove, 2009). The realised environment is the set of possible combinations between the environmental variables that exist in a certain space/time. The original and projected areas/times should, therefore, have the same conditions. It is also important to identify the environments that exist in one specific space/time but do not occur in another (i.e. non-analogue environment; Petitpierre *et al.*, 2017), since this can have an influence on the quantification of the realised niche in each space/time, influencing the way the models are built, compared and projected (Guisan *et al.*, 2014).

Therefore, when projecting a model one implicitly assumes that (Guisan *et al.*, 2017): (i) the different environments (original and projection) are identically available and analogous, and (ii) the realised niche is fully accounted for in the model. It means for instance that if a model is built using only a part of the information about the niche of the species (i.e. due to limited extent or faulty sampling design), it may introduce errors when projected to different spaces/times, because the model was fitted using truncated or biased response curves (Thuiller *et al.*, 2004b; Barbet-Massin *et al.*, 2010).

I will not go into further details regarding the types of projections one can decide to make (i.e. in space and time, to past and new environments/areas, under different resolutions or using an ensemble approach), and additional information can be found in Guisan *et al.* 2017 (Chapter 17), for example.

From individual species to community modelling

Community modelling

Before starting to think about applying SDMs to model communities (or assemblages, used here interchangeably), it is important first to understand what a community is and what is studied in community ecology. A community can be defined as “a group of organisms representing multiple species living in a specified place and time” (Vellend, 2016). Researchers usually focus on a subset of species of the full community (e.g. plants, birds or mammals; i.e. focal community) because it is nearly impossible to study all the species that occur together at once (Mittelbach, 2012). Community ecology is then most often the study of patterns in diversity, abundance, and composition of species in specific communities (or focal communities), while also focusing on the processes that are essential to generate these patterns (see Vellend, 2010 for a review).

Since SDMs are widely used to understand and model individual species distributions and the patterns that they generate, it is reasonable to think that these can also be applied to study and model communities and their distributions in space and time (Ferrier & Guisan, 2006). To model communities with this approach, one mostly needs the same type of data that is used to model single species distributions (i.e. observations of species existing in a community and environmental predictors). The decision of using single species or community models will eventually depend on the objectives of the study and the type, quality, quantity, and availability of the data necessary to build the model, as well as on the final objective of the study.

Community modelling approaches are useful for their ability to produce spatial patterns’ information about biodiversity concisely (e.g. one species richness map instead of multiple single species maps), that is both useful and easily understandable by scientists and stakeholders alike (Ferrier & Guisan, 2006). Over the past 30 years, different community modelling approaches or strategies based on SDMs have been presented, tested and reviewed (e.g. Ferrier & Guisan, 2006; Guisan & Rahbek, 2011; Mokany *et al.*, 2012; D'Amen *et al.*, 2017). Because they differ on the type of data used, methodology and final outputs, different approaches will have different strengths and weaknesses.

How to model communities? – Different strategies: strengths and weaknesses

As previously said, different approaches or strategies have been proposed to model communities, and a thorough review of the most widely used can be found in D'Amen *et al.* (2017). Here, I focus on two strategies, as previously described by Ferrier and Guisan (2006), that I consider being the most relevant to understand the work developed in this thesis:

- *Assemble first and predict later*: in this approach, macroecological models (or MEMs) are usually used to predict species richness directly. These models are fitted by statistically relating species richness with values of environmental variables of a specific pixel or area. The number of species that can co-occur is expected to depend on factors like resource availability, disturbance levels or environmental heterogeneity. A limitation that is commonly attributed to MEMs pertains to the fact that the information about the identity of the species occurring in each location is not provided, making it unable to predict changes in species composition.

- *Predict first and assemble later*: in this approach, species distribution models are used to predict the distribution of individual species separately and then summing those predictions to obtain species assemblages (i.e. stacked species distribution models; S-SDMs). As a result, the information about species richness and the identity of the species occupying each location (i.e. composition) is given. That information can be provided either as probability (pS-SDM) or as binary predictions (bS-SDM) when the probabilities are transformed into a binary scale. However, previous studies showed that bS-SDMs tend to overpredict the number of species that occur in each location (Guisan & Rahbek, 2011; Cord *et al.*, 2014; D'Amen *et al.*, 2015b). This overprediction does not seem to be a general rule, as a recent study developed by D'Amen *et al.* (2015a) showed that it depends on the considered taxonomic group and might also depend on the quality of the individual models used.

Due to the limitations identified in both approaches, Guisan and Rahbek (2011) proposed a unifying modelling framework (i.e. SESAM - spatially explicit species assemblage modelling), consisting on the integration of the previously mentioned filters: habitat, dispersal and biotic interactions, together with macroecological constraints either derived from a separate macroecological model (MEM) or from the same pS-SDM. The reasoning behind it being that if constraints are added through macroecological models or from the same pS-SDM, the risk of overprediction of realised species richness (usually obtained when using bS-SDMs) will decrease. This will consequently improve prediction success while also keeping the identity of the species that occur in each site (for a detailed description of SESAM, see Guisan & Rahbek, 2011; D'Amen *et al.*, 2015a).

Uncertainty and errors in models

Sources of uncertainty and their effects on models

As previously stated, SDMs statistically relate species observations with environmental data to predict species distributions in a determined space or time. Because these models are widely used nowadays (e.g. conservation, invasions management; see graphical trend in Guisan *et al.*, 2013) and attempt to summarise complex patterns with limited information, it is important to understand and consider the different sources of uncertainty and errors that might restrict their applicability. Species distribution models contain uncertainties from different sources (see e.g. Barry & Elith, 2006; Beale & Lennon, 2012; Gould *et al.*, 2014 for a review), with the most common and widely studied being associated with species data, including for example:

- *Biased samples*: The best way to gather data to be used in a model is by planning a sampling strategy that is adequate for the studied species, with environmental gradients being widely sampled, as well as the full extent of the study area. Biases in samples occur because in most of the cases, these structured sampling strategies are replaced by opportunistic collections that are normally carried out at lower elevations, close to roads (Kadmon *et al.*, 2004) or other easy access routes (e.g. rivers or railroads). They are also sampled in specific habitats chosen by expert knowledge of the species in question, close to populated areas (i.e. travelling convenience) or in specific times of the year (Daru *et al.*, 2018), when a species might be suffering a contraction/expansion of the niche (Bean *et al.*, 2012). These geographical/environmental biases mean that the statistical relationships established by the model are only focused on the patterns at the sampled sites and not across the entire occupied niche, leading to prediction uncertainty.
- *The incorrect location of species records*: Samples can also be biased by errors in the observations themselves, either by wrongly identifying a species as presence/absence (see next point) or by the incorrect location of species observations (i.e. referred as locational or positional uncertainty; see e.g. Graham *et al.*, 2008; Mitchell *et al.*, 2017). These errors in occurrence data are usually caused by annotation mistakes (when transferring information from paper to electronic databases), imprecise location descriptions or georeferencing (i.e. errors associated with GPS devices that have coarse accuracy). Like other sample biases, locational errors will also make the model incorrectly characterize the species-environment relationship, leading to inaccurate predictions.
- *Imperfect detection*: another data bias commonly affecting SDMs concerns the inability to separate potentially false and true species' occurrences obtained through field surveys (Kéry *et al.*, 2009; Lahoz-Monfort *et al.*, 2014). Imperfect detection occurs mainly due to double-counting, misidentification or omission of certain species or individuals, and can lead to underestimation of species occupancy and incorrect predictions (Guillera-Arroita *et al.*, 2010; Lahoz-Monfort *et al.*, 2014). The misidentification of absences in presence-absence datasets or the omission of presences

in presence-only models usually leads to predictions that reflect where the species is detected (i.e. apparent distribution) and not the areas where it actually occurs or not (Kéry, 2011; Lahoz-Monfort *et al.*, 2014). Another consequence of these errors concerns the fact that some species-environment relationships are expected to be missed or wrongly identified (Kéry, 2011).

- *Small sample sizes*: the size of the samples used to develop SDMs has an important effect on their predictive accuracy, with a minimum number of presence-absences being required in order to obtain accurate predictions (e.g. Stockwell & Peterson, 2002; Hernandez *et al.*, 2006; Wisz *et al.*, 2008). Model performance generally decreases with sample size due to several reasons. For example, uncertainty derived from model parameterization decreases with increasing sample size (Crawley, 2002; Wisz *et al.*, 2008). Outliers also carry extra weight when sample sizes are small since less data is available to reduce their effects. A small sample can also reduce the accurate description of the full suitable conditions in which a species occurs (Barry & Elith, 2006; Wisz *et al.*, 2008).
- *Unavailable absence data*: a model can, in some cases, wrongly identify the characteristics of unsuitable locations for a certain species. This usually occurs because the model is unable to correctly discriminate between suitable and unsuitable habitats due to the unavailability of species' absence data (e.g. Elith & Leathwick, 2009; Phillips *et al.*, 2009). Datasets without absence data are commonly designated as presence-only and are widely available through museum/herbaria collections or by opportunistic field observations.
- *Errors/absence of important variables*: variables used as model predictors are also susceptible to errors, or might even be completely missing from the model (Mod *et al.*, 2016), leading to prediction inaccuracies. These errors found in variables can have different causes, like the variables being derived from interpolation of point data (Barry & Elith, 2006; Heritage *et al.*, 2009), the uncertain location of polygon boundaries (e.g. derived from land cover maps) and ecotones (Fortin *et al.*, 2000), or when transforming fine-grained information into coarser data (i.e. different spatial scales that can lead to losses of information; Keil *et al.*, 2013; Fernandes *et al.*, 2014). In the case of some variables being missing (i.e. not accounted for) from a model, this will lead to an incomplete understanding of the environmental factors that influence the distribution of a species, limiting the predictive accuracy of SDMs (see Mod *et al.*, 2016).

How to deal with uncertainty?

If we know that a model can be affected by different sources of errors/uncertainties, the next logical step is to determine where improvements can be made and to try to find methods or approaches that can deal or minimize the effects of those errors. Over the years, several of the previously mentioned impacts on SDMs have been studied, and a variety of different solutions have been proposed to deal with them.

For example, issues related with imperfect detection (i.e. misidentification of presence and absences) have received considerable attention: (i) MacKenzie *et al.* (2002) proposed a model to account for imperfect detection (i.e. site-occupancy models), which provided good occupancy estimates; (ii) the impacts of non-detection of species in models was evaluated by Gu and Swihart (2004), suggesting that multiple samplings should be standard in future studies; (iii) Guillera-Arroita *et al.* (2010) provided information to improve survey designs, advising a decrease in the number of sites and an increase of the replications per site when interested in species detection; and (iv) Lahoz-Monfort *et al.* (2014) also evaluated the impacts of imperfect detection on SDMs, showing their negative impact on model performance and suggesting that to avoid this problem one can either increase the survey effort per site or collect survey data in a way that allows for the modelling of the detection process (i.e. larger sample sizes). These negative effects were also assessed in a recent paper by Guélat and Kéry (2018), where the authors showed that spatial N-mixture models were able correctly estimate true abundances.

When dealing with sources of sample bias like the impact of small sample sizes in SDMs, solutions have been tested and suggested (e.g. Wisz *et al.*, 2008; Hanberry *et al.*, 2012; Moudrý & Šímová, 2012; Liu *et al.*, 2018). A new framework has recently been proposed to circumvent limitations when modelling rare species (Lomba *et al.*, 2010; Breiner *et al.*, 2015; Breiner *et al.*, 2018). Recommendations to use such ensemble models, increase the number of sampled sites or the used of algorithms based on maximum entropy (i.e. MAXENT) are generally suggested.

Implication of climatic and roadside bias have also been discussed (Kadmon *et al.*, 2003, 2004), with the authors observing negative effects on model performance from both biases. However, Thibaud *et al.* (2014) tested the effects of two sampling designs (random sampling or road bias) on the performance of SDMs, determining that their effect is minimal at least in the mountain landscape they studied. In any case, solutions to correct these sampling biases when present usually include the sampling of additional data through targeted surveys in poorly sampled areas (Cawsey *et al.*, 2002), environmental or spatial filtered models (Boria *et al.*, 2014; Varela *et al.*, 2014), and through the integration of occurrences from focal and related species in SDMs (Qiao *et al.*, 2017).

Finally, solutions have also been proposed to solve the problem of the incorrect location of species records (e.g. regression calibration; Hefley *et al.*, 2014; Hefley *et al.*, 2017; Mitchell *et al.*, 2017), for models developed with presence-only data (e.g. logistic descrimination, hierarchical models; Pearce & Boyce, 2006; Elith & Leathwick, 2009; Gallien *et al.*, 2012) or using different spatial scales (e.g. downscaling; Keil *et al.*, 2013; Fernandes *et al.*, 2014; Azaele *et al.*, 2015).

Virtual ecology and simulations – a brief review

Simulations, virtual species and real-world data limitations

Throughout this introduction, having already discussed how biodiversity is increasingly being threatened by anthropogenic actions and how SDMs can be important tools to help understand the factors influencing species and assemblages' distributions, and anticipate future ones. Species distribution models have recently been used to help evaluate the impacts of wind farms on birds (Bastos *et al.*, 2016), help in the creation of cost-effective monitoring networks (Vicente *et al.*, 2016) or in the modelling of rare species (Breiner *et al.*, 2015), just to name a few examples. However, despite SDMs being useful for various applications, certain precautions need to be taken or at least considered due to uncertainties or errors that can occur at different stages of the modelling process.

In addition to the limitations that reduce reliance on model predictions (e.g. species misidentification), researchers are frequently unable to obtain complete and reliable information about the full area or species studied. In most cases, one is limited to field observations obtained under specific conditions, often constrained by limited time, accessibility and budgets. These observations might also derive from different sampling strategies or observers, and performed at different spatial resolutions, introducing biases in the sampled data. Additionally, several factors cannot be controlled when sampling species occurrences, for example: the type of environmental conditions that can be either beneficial/unfavourable for the occurrence of a species in a specific period; previous disturbances that occurred without being recorded like fires or floods; historical uses of the land that might influence the species that occur in an area at different times (e.g. transition from grassland to forest); and biotic interactions like competition, that are difficult to measure.

So, considering species distribution models' uncertainties and the inability to obtain a complete representation of the real world, how can one reliably assess the accuracy of a method or model? How can one determine if the distribution of a certain species is predicted correctly? Are we able to test different methods or approaches against a complete and known reality? A way to avoid previously mentioned limitations and be able to reply these questions is to use "virtual species" (Hirzel *et al.*, 2001) or "artificial data" (Austin *et al.*, 2006). Using virtual species, all the information necessary for a study is always available in artificial or semi-artificial reality, which allows for the complete or at least partial control of the data and models/methods being tested (Hirzel *et al.*, 2001; Austin *et al.*, 2006). In a relatively recent paper, Zurell *et al.* (2010) proposed to call the use of artificial data to answer questions in ecology a "virtual ecologist" approach (see Figure 6). In such virtual approach, artificial/simulated data are used as a substitute for reality, allowing for the accurate comparison and testing of models, methods and sampling schemes with a fully known truth. This framework usually consists in four steps (see Zurell *et al.*, 2010 for details): (i) a virtual model that generates the species, landscape or ecosystem being studied, with all the relevant processes for the study being included at this stage (e.g. which factors influence the distribution of species, existence of biotic interactions or dispersal); (ii) a sampling method that simulates a real word observation process; (iii) a statistical model or approach used to predict ecological processes; and (iv) the results are evaluated against the initial truth. These steps can be adapted to fit a specific study, but it provides a clear outline to test and compare different methods.

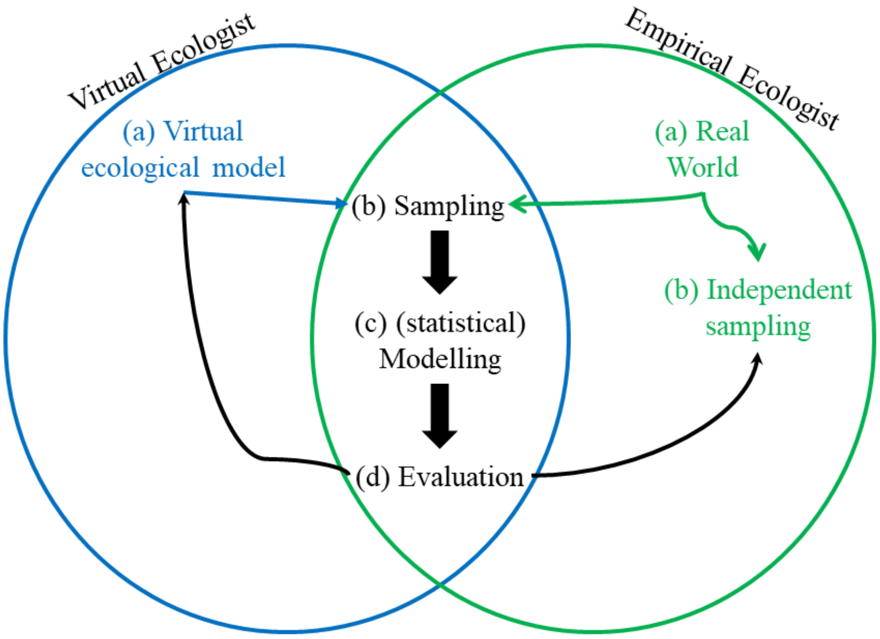


FIGURE 6 The elements of virtual ecologist approach (adapted from Zurell *et al.*, 2010). See text for details.

When applied to species distribution models, the use of virtual data instead of real species allows for the "true" distribution of the species to be completely known (as firstly stated in Hirzel *et al.*, 2001). However, not only the distribution of the species is known but also the factors that influence that distribution can be controlled. This means that contrary to the use of real species data, the effects of processes like biotic interactions, assembly rules or dispersal limitations can be accounted (i.e. by simulations) or simply removed (i.e. not considered as having an effect).

Virtual species have been used to test a variety of issues ranging from the best approaches to sample species data (e.g. Hirzel & Guisan, 2002), to being used to assess the best approach to downscale coarse-grain data (e.g. Bombi & D'Amen, 2012) or select pseudo-absences (e.g. Wisz & Guisan, 2009; Barbet-Massin *et al.*, 2012), and measure the effects of different factors affecting predictions (e.g. Thibaud *et al.*, 2014; Fernandes *et al.*, accepted). While the method used for the creation or simulation of virtual species can be different depending on the study, it usually follows two steps (adapted from Meynard & Kaplan, 2013): (i) a functional response needs to be created, often simulating the probability of a species to occur in a certain set of environmental conditions; and (ii) these probabilities are converted into presence-absence data, using either a fixed threshold value or through a random process attributing different values to the probabilities of occurrence, linked to its response in an environmental gradient (i.e. probabilistic approach). Meynard and Kaplan (2013) also provided five reasons explaining that the threshold approach can be problematic, like giving over-optimistic measures, and should be dropped in favour of the probabilistic approach.

A virtual species can be created by simply using the predicted distribution of a real species and consider it as the "true" distribution from which presence-absence data can later be sampled and be used in a virtual ecologist approach (the procedure used in chapters 1.1 and 1.2 of this thesis). Over the past three years, different packages were also proposed to facilitate the creation, sampling and application of virtual species (e.g. SDMvspecies, NicheLim or virtualspecies - Duan *et al.*, 2015; Huang *et al.*, 2016; Leroy *et al.*, 2016).

Are virtual species a solution to improve SDMs?

Having described why one might use virtual species and how they can be created, it is important to enumerate further the different uses, applications and strengths of using virtual data, specifically to improve species distribution models. Different questions can be raised about the necessity or not of improving SDMs, and if the use of artificial data is the best way to achieve it. Over the years, different frameworks concerning SDMs have been established and accepted (e.g. Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005; Elith & Leathwick, 2009; Franklin, 2010; Peterson, 2011). However, many of the studies assessing methodological issues that might affect SDMs are performed using real species observations. While valid and extremely useful, studies that for example compare different statistical methods (Elith *et al.*, 2006), how spatial errors and sample sizes affect model performance (Graham *et al.*, 2008; Wisz *et al.*, 2008) or which method to use to predict species assemblages (D'Amen *et al.*, 2015b; D'Amen *et al.*, 2015a), have limitations in assessing how good a model or method really is. This is because, in most of these studies, the complete distribution of the species and their relationship with the environment is not known or cannot be controlled.

Additionally, other confounding effects between statistical methods - data uncertainties, the accuracy of the metrics used or other factors previously discussed - make it nearly impossible for the complete generalization of rules and recommendations (Meynard & Kaplan, 2013). Therefore, virtual species can be a useful method to sort out the effects that can be attributed to different factors when modelling species distributions, since they allow for the comparison of model output with a fully known "truth". This means that one can assess how different decisions, related with data or model parameterisation, change the effectiveness of a certain method to reproduce the known "truth", determining this way the adequacy of that method.

Virtual species have been used to test a variety of issues in what are called “virtual species distribution models” (Miller, 2014). Issues related to species attributes, data characteristics, statistical methods or accuracy metrics have been widely studied using virtual species (see Miller, 2014 for recent applications). Several studies have used virtual species to validate proposed SDM methods, like comparing different approaches to model range dynamics and propose improvements in range projections (Zurell *et al.*, 2016), the testing and assessment of new R packages (Guisande *et al.*, 2017) or the proposal of a framework that helps to overcome limitations of non-equilibrium in SDMs (Hattab *et al.*, 2017). As previously stated, Hirzel *et al.* (2001) were among the first ones to use virtual species to compare two habitat suitability methods (i.e. ENFA and GLM), but artificial data was also used to test different approaches to sample species data (Hirzel & Guisan, 2002), recommending guidelines to improve sample design. Other studies also focused on comparing and assessing how different statistical methods performed. Even though a very important study was performed by Elith *et al.* (2006), the comparison of the 16 tested methods was done while using real species data. However, in a following study, Elith and Graham (2009) tested different modelling techniques while using virtual species data. In a fairly recent work, Thibaud *et al.* (2014) measured the effects of different factors affecting single species model predictions, highlighting the major influence of sample size and modelling technique on their predictive accuracy. The performance of four connectivity metrics was also evaluated using a virtual ecologist approach, suggesting that the performance of these metrics is dependent on the context in which they are used (Simpkins *et al.*, 2018). Finally, artificial species were also used to examine thirteen threshold selection methods and assess which ones can be used with presence-only data (Liu *et al.*, 2013). The authors were able to determine three methods that were not affected by pseudo-absences, which can be used for threshold selection when presence-only data is available. While not extensive, these are just a few examples that show the potentialities of working with SDMs and virtual species.

Taking all this into account, research using virtual species and the virtual ecologist approach can be employed to further develop consistent distribution models, being properly tested in a “controlled environment”. It’s well known and accepted that most of the experimental studies are oversimplifications of the real world, due to the impossibility to obtain the complete information about reality. When testing a method using virtual species, in a more or less complex but completely known reality, one can argue that if the method being tested is unsuccessful in that simplified “reality”, the probability of that method to work with real data is also reduced (Miller, 2014). On the other hand, if the method tested is successful, there might be a higher probability of also working with real data, although it will still need to be tested using real and more complex data. Something similar was also stated in Zurell *et al.* (2010), with a method tested using virtual species being more easily discredited than corroborated. Therefore, virtual species should be considered as an important tool to assess the reliability of methods and approaches (e.g. serving as a filter for those that have erroneous outputs), but not as the “silver bullet” that will solve all the problems occurring in SDMs.

Furthermore, virtual species have been rarely used at the assemblage level, likely because there is yet no unanimous method on how to correctly predict species assemblages, or because the computational requirements to study assemblages remains relatively data intensive. This is a gap I tried to fill in my thesis, using virtual species to assess different factors that influence the predictive accuracy of assemblage models.

Objectives and thesis structure

The main objectives of my thesis, within the context of the SESAM'ALP project, were to implement a virtual ecologist approach and artificial data to test species distribution modelling approaches in order to: (i) determine the degree to which different sources of uncertainty or factors used in SDMs affect the predictive accuracy of models of individual species and their assemblages; (ii) determine the degree to which truncated datasets affect the accuracy of assemblage predictions; and (iii) test the validity of strategies used to integrate information from large scales into regional/local scales (e.g. hierarchical models).

I decided to structure my thesis in three main parts (i.e. I, II and III). The chapters developed around the main objectives of the thesis are presented in PART I, also being the main the focus of the issues presented in the INTRODUCTION and debated in the SYNTHESIS AND DISCUSSION section.

More specifically, in **CHAPTER 1.1**, I sought evaluate the degree to which different types and levels of errors in species data (i.e. false presences or absences) affect model predictions, and how this is reflected in metrics that are frequently used to evaluate the prediction accuracy of SDMs. I found that the interpretation of models' performance depended on the data and metrics used to evaluate them and that some high evaluation metrics could still be obtained when large amount of error was added. This highlights the need to reconsider the interpretation scale of some metrics (Somers'D/AUC).

In **CHAPTER 1.2**, I assessed how different factors can affect the prediction accuracy of virtual assemblages obtained by stacking individual SDM predictions (stacked SDMs). Specifically, I evaluate the effects of five different factors (i.e. sample size, sampling method, sampling prevalence, modelling technique and thresholding method). I found that the S-SDM prediction accuracy is mostly affected by modelling technique followed by sample size. I also found that even with complete knowledge of the factors influence the species distribution, and with large sample sizes, one was unable to reach perfect accuracy (always residual uncertainty). I also recommended the use of a 'plot-like' sampling method (best approximation of the species' true prevalence) and not simply increasing the number of presences-absences of species.

In **CHAPTER 1.3**, I tested the potential causes that increasingly truncated datasets have on the predictive accuracy of species assemblages and if the variables used to calibrate the models also influence that accuracy. For that, I tested different realizations of virtual species niches and calibrated SDMs using either a non-truncated or highly truncated datasets. I found that the degree of truncation has more influence on species with wide realized niches, not affecting species with narrow niches.

Finally, in **CHAPTER 1.4**, I tested and compared how accurate different modelling strategies are at predicting species assemblages under current and future climatic conditions, assessing their transferability (i.e. how good the strategies are at predicting to spaces and times outside the range of data used for model calibration). I was particularly interested in assessing the validity of hierarchical models, and on determining the influence that presences/absences or presence/pseudo-absences data have on the transferability capabilities of the different strategies. I found that when using presence/pseudo-absence data, all the strategies failed to predict accurate species assemblages.

Two additional chapters within the main scope of the thesis, and in which I actively participated, are presented in PART II. In one of those manuscripts, we tested two implementations of the SESAM framework using either a “probability ranking” rule or a “trait range rule” (CHAPTER 2.1). We found that the “probability ranking” rule allowed for improvements in prediction of assemblage composition and discussed further improvements to the SESAM framework. In the second manuscript (CHAPTER 2.2), we assess which thresholding method is the best choice for predicting species assemblage composition, arguing that community-based thresholding can be the better alternative. Based on that, we provided a framework for cross-validation at the community level.

Additionally, in PART III, two additional chapters in which I’m the main author are presented. Despite these two final chapters being outside the scope of the thesis, I decided to include them in the document due to the time and effort involved on them.

In the APPENDICES section, the supplementary information of all the presented chapters is available.

Finally, I contributed to two already published manuscripts developed in collaboration with researchers predominantly based in Portugal (not presented in this thesis). The first manuscript is entitled “*Cost-effective monitoring of biological invasions under global change: a model-based framework*” and was published on the Journal of Applied Ecology. The second manuscript, entitled “*Estimating Invasion Success by Non-Native Trees in a National Park Combining WorldView-2 Very High Resolution Satellite Data and Species Distribution Models*” was published in the journal Diversity.

I also participated in writing a chapter of a book about climate change impacts on mountain biodiversity (Guisan *et al.,* In press), and supervised a master thesis that it’s currently being converted into a scientific publication. In that thesis, the master student (Cindy Ramel) developed ways to integrate ecosystem services’ monetary values within a framework for spatial conservation prioritization.

References

Albert, C.H., Yoccoz, N.G., Edwards, T.C., Graham, C.H., Zimmermann, N.E. & Thuiller, W. (2010) Sampling in ecology and evolution - bridging the gap between theory and practice. *Ecography*, **33**, 1028-1037.

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

Álvarez-Martínez, J.M., Suárez-Seoane, S., Stoorvogel, J.J., de Luis Calabuig, E. & Gilliam, F. (2014) Influence of land use and climate on recent forest expansion: a case study in the Eurosiberian-Mediterranean limit of north-west Spain. *Journal of Ecology*, **102**, 905-919.

Araujo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol*, **22**, 42-7.

Araujo, M.B. & Peterson, A.T. (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**, 1527-39.

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688.

Araújo, M.B.W., R.J.; Ladle, R.J. & Erhard, M. (2005) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology & Biogeography*, **14**, 529-538.

Austin, M. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1-19.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. & Luoto, M. (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, **199**, 197-216.

Azaele, S., Maritan, A., Cornell, S.J., Suweis, S., Banavar, J.R., Gabriel, D., Kunin, W.E. & Murrell, D. (2015) Towards a unified descriptive theory for spatial ecology: predicting biodiversity patterns across spatial scales. *Methods in Ecology and Evolution*, **6**, 324-332.

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? *Ecography*, **33**, 878-886.

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

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B. & Ferrer, E.A. (2011) Has the Earth's sixth mass extinction already arrived? *Nature*, **471**, 51-7.

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

Bastos, R., Pinhanços, A., Santos, M., Fernandes, R.F., Vicente, J.R., Morinha, F., Honrado, J.P., Travassos, P., Barros, P., Cabral, J.A. & Cadotte, M. (2016) Evaluating the regional cumulative impact of wind farms on birds: how can spatially explicit dynamic modelling improve impact assessments and monitoring? *Journal of Applied Ecology*, **53**, 1330-1340.

Beale, C.M. & Lennon, J.J. (2012) Incorporating uncertainty in predictive species distribution modelling. *Philos Trans R Soc Lond B Biol Sci*, **367**, 247-58.

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

Bolliger, J., Bergamini, A., Stofer, S., Kienast, F. & Scheidegger, C. (2007) Predicting the potential spatial distributions of epiphytic lichen species at the landscape scale. *Lichenologist*, **39**, 279-291.

Bombi, P. & D'Amen, M. (2012) Scaling down distribution maps from atlas data: a test of different approaches with virtual species. *Journal of Biogeography*, **39**, 640-651.

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. *Ecological Modelling*, **275**, 73-77.

Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5-32.

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. *Methods in Ecology and Evolution*, **9**, 802-808.

Breiner, F.T., Guisan, A., Bergamini, A., Nobis, M.P. & Anderson, B. (2015) Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, **6**, 1210-1218.

Briani, D.C., Palma, A.R.T., Vieira, E.M. & Henriques, R.P.B. (2004) Post-fire succession of small mammals in the Cerrado of central Brazil. *Biodiversity and Conservation*, **13**, 1023-1037.

Broennimann, O. & Guisan, A. (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biol Lett*, **4**, 585-9.

Broennimann, O., Treier, U.A., Muller-Scharer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecol Lett*, **10**, 701-9.

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. *Global Ecology and Biogeography*, **21**, 481-497.

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437-448.

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

Busby, J. (1991) BIOCLIM-a bioclimate analysis and prediction system. *Plant protection quarterly (Australia)*,

Buse, J., Schroder, B. & Assmann, T. (2007) Modelling habitat and spatial distribution of an endangered longhorn beetle - A case study for saproxylic insect conservation. *Biological Conservation*, **137**, 372-381.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S. & Naeem, S. (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**, 59-67.

Cawsey, E.M., Austin, M.P. & Baker, B.L. (2002) Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling *Biodiversity and Conservation*, **11**, 2239-2274.

Ceballos, G. & Ehrlich, P.R. (2002) Mammal population losses and the extinction crisis. *Science*, **296**, 904-7.

Chapin, F.S., 3rd, Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C. & Diaz, S. (2000) Consequences of changing biodiversity. *Nature*, **405**, 234-42.

Chase, J.M. & Leibold, M. (2003) *Ecological Niches: Linking Classical and Contemporary*

Approaches, Chicago, IL.

Chytrý, M., Wild, J., Pyšek, P., Jarošík, V., Dendoncker, N., Reginster, I., Pino, J., Maskell, L.C., Vilà, M., Pergl, J., Kühn, I., Spangenberg, J.H. & Settele, J. (2012) Projecting trends in plant invasions in Europe under different scenarios of future land-use change. *Global Ecology and Biogeography*, **21**, 75-87.

Clements, F.E. (1916) Plant Succession - An analysis of the development of vegetation.

Cohen, J. (2016) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**, 37-46.

Cord, A.F., Klein, D., Gernandt, D.S., de la Rosa, J.A.P., Dech, S. & McGeoch, M. (2014) Remote sensing data can improve predictions of species richness by stacked species distribution models: a case study for Mexican pines. *Journal of Biogeography*, **41**, 736-748.

Crawley, M. (2002) *Statistical computing: an introduction to data analysis using S-plus*. John Wiley & Sons, Ltd, Chichester, U.K.

D'Amen, M., Pradervand, J.-N. & Guisan, A. (2015a) Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography*, **24**, 1443-1453.

D'Amen, M., Rahbek, C., Zimmermann, N.E. & Guisan, A. (2017) Spatial predictions at the community level: from current approaches to future frameworks. *Biol Rev Camb Philos Soc*, **92**, 169-187.

D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L. & Guisan, A. (2015b) Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *Journal of Biogeography*, **42**, 1255-1266.

Daru, B.H., Park, D.S., Primack, R.B., Willis, C.G., Barrington, D.S., Whitfeld, T.J.S., Seidler, T.G., Sweeney, P.W., Foster, D.R., Ellison, A.M. & Davis, C.C. (2018) Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytol*, **217**, 939-955.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science*, **332**, 53-8.

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J. & Collen, B. (2014) Defaunation in the Anthropocene. *Science*, **345**, 401-6.

Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Wu, G.-L. & Wang, Z.-G. (2015) SDMvspecies: a software for creating virtual species for species distribution modelling. *Ecography*, **38**, 108-110.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122-1131.

Edwards, T.C., Cutler, D.R., Zimmermann, N.E., Geiser, L. & Moisen, G.G. (2006) Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling*, **199**, 132-141.

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

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677-697.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. & E. Zimmermann, N. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Engler, R., Randin, C.F., Thuiller, W., Dullinger, S., Zimmermann, N.E., AraÚJo, M.B., Pearman, P.B., Le Lay, G., Piedallu, C., Albert, C.H., Choler, P., Coldea, G., De Lamo, X., DirnbÖCk, T., GÉGout, J.-C., GÓMez-GarcÍA, D., Grytnes, J.-A., Heegaard, E., HØIstad, F., NoguÉS-Bravo, D., Normand, S., PuŞCaŞ, M., SebastiÀ, M.-T., Stanisci, A., Theurillat, J.-P., Trivedi, M.R., Vittoz, P. & Guisan, A. (2011) 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology*, **17**, 2330-2341.

Fahrig, L. (2003) Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 487-515.

Fernandes, R.F., Scherrer, D. & Guisan , A. (accepted) How much should one sample to accurately predict the distribution of species assemblages? A virtual community approach *Ecological Informatics*,

Fernandes, R.F., Scherrer, D. & Guisan , A. (in press) Effects of simulated observation errors on the performance of species distribution models. *Diversity and Distributions*,

Fernandes, R.F., Vicente, J.R., Georges, D., Alves, P., Thuiller, W. & Honrado, J.P. (2014) A novel downscaling approach to predict plant invasions and improve local conservation actions. *Biological Invasions*, **16**, 2577-2590.

Ferreira, M., Filipe, A.F., Bardos, D.C., Magalhaes, M.F. & Beja, P. (2016) Modeling stream fish distributions using interval-censored detection times. *Ecol Evol*, **6**, 5530-41.

Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393-404.

Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302-4315.

Fitzpatrick, M.C. & Hargrove, W.W. (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, **18**, 2255-2261.

Fleishman, E., Mac Nally, R., Fay, J.P. & Murphy, D.D. (2001) Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology*, **15**, 1674-1685.

Fortin, M.J., Olson, R.J., Ferson, S., Iverson, L., Hunsaker, C., Edwards, G., Levine, D., Butera, K. & Klemas, V. (2000) Issues related to the detection of boundaries. *Landscape Ecology*, **15**, 453-466.

Fourcade, Y., Besnard, A.G. & Secondi, J. (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, **27**, 245-256.

Franklin, J. (2010) *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge.

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. *Ecological Modelling*, **217**, 48-58.

Friedman, J., Hastie, T. & Tibshirani, R. (2000) Additive logistic regression:a statistical view of boosting. *The Annals of Statistics*, **28**, 337-407.

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? *Global Ecology and Biogeography*, **21**, 1126-1136.

García-Valdés, R., Svenning, J.-C., Zavala, M.A., Purves, D.W., Araújo, M.B. & Saura, S. (2015) Evaluating the combined effects of climate and land-use change on tree species distributions. *Journal of Applied Ecology*, **52**, 902-912.

Gause, G. (1936) The struggle for existence. *Soil Science*, **41**, 159.

Gleason, H.A. (1926) The individualistic concept of the plant association. *Bulletin of the Torrey botanical club*, 7-26.

Gould, S.F., Beeton, N.J., Harris, R.M., Hutchinson, M.F., Lechner, A.M., Porfirio, L.L. & Mackey, B.G. (2014) A tool for simulating and communicating uncertainty when modelling species distributions under future climates. *Ecol Evol*, **4**, 4798-811.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol Evol*, **19**, 497-503.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A. & Gro, N.P.S.W. (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**, 239-247.

Grenouillet, G., Buisson, L., Casajus, N. & Lek, S. (2011) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, **34**, 9-17.

Grinnell, J. (1917) Field tests of theories concerning distributional control. *American Naturalist*, **51**, 115-128.

Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, **116**, 195-203.

Guélat, J. & Kéry, M. (2018) Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, **9**, 1614-1625.

Guillera-Arroita, G., Ridout, M.S. & Morgan, B.J.T. (2010) Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, **1**, 131-139.

Guisan, A. (2002) Semi-quantitative models for predicting the spatial distribution of plant species. *Predicting Species Occurrences: Issues of Accuracy and Scale., Island Press, Covelo, California*,

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Guisan, A. & Theurillat, J.-P. (2000) Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia*, **30**, 353-384.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Guisan, A. & Rahbek, C. (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433-1444.

Guisan, A., Edwards, T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89-100.

Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press.

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C. & Kueffer, C. (2014) Unifying niche shift studies: insights from biological invasions. *Trends Ecol Evol*, **29**, 260-9.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007) What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615-630.

Guisan, A., Broennimann, O., Buri, A., Cianfrani, C., D'Amen, M., Di Cola, V., Fernandes, R., Gray, S., Mateo, R.G., Pinto, E., Pradervand, J.-N., Scherrer, D., von Däniken, I., Yashiro, E. & Vittoz, P. (In press) Climate change impact on mountain biodiversity. *Climate Change and Biodiversity* (ed. by T.E. Lovejoy and L. Hannah). Yale University Press.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecol Lett*, **16**, 1424-35.

Guisande, C., García-Roselló, E., Heine, J., González-Dacosta, J., Vilas, L.G., García Pérez, B.J. & Lobo, J.M. (2017) SPEDInstabR: An algorithm based on a fluctuation index for selecting predictors in species distribution modeling. *Ecological Informatics*, **37**, 18-23.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.X. & Townshend, J.R. (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci Adv*, **1**, e1500052.

Hanberry, B.B., He, H.S. & Dey, D.C. (2012) Sample sizes and model comparison metrics for species distribution models. *Ecological Modelling*, **227**, 29-33.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29-36.

Hastie, T.J. & Tibshirani, R.J. (1990) *Generalized Additive Models*. Chapman & Hall, London.

Hattab, T., Garzón-López, C.X., Ewald, M., Skowronek, S., Aerts, R., Horen, H., Brasseur, B., Gallet-Moron, E., Spicher, F., Decocq, G., Feilhauer, H., Honnay, O., Kempeneers, P., Schmidtlein, S., Somers, B., Van De Kerchove, R., Rocchini, D. & Lenoir, J. (2017) A unified framework to model the potential and realized distributions of invasive species within the invaded range. *Diversity and Distributions*, **23**, 806-819.

Hefley, T.J., Brost, B.M. & Hooten, M.B. (2017) Bias correction of bounded location errors in presence-only data. *Methods in Ecology and Evolution*, **8**, 1566-1573.

Hefley, T.J., Baasch, D.M., Tyre, A.J. & Blankenship, E.E. (2014) Correction of location errors for presence-only species distribution models. *Methods in Ecology and Evolution*, **5**, 207-214.

Heritage, G.L., Milan, D.J., Large, A.R.G. & Fuller, I.C. (2009) Influence of survey strategy and interpolation model on DEM quality. *Geomorphology*, **112**, 334-344.

Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773-785.

Hespanhol, H., Cezon, K., Felicisimo, A.M., Munoz, J. & Mateo, R.G. (2015) How to describe species richness patterns for bryophyte conservation? *Ecol Evol*, **5**, 5443-55.

Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331-341.

Hirzel, A.H., Helfer, V. & Metral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145**, 111-121.

Hirzel, A.H., J. Hausser, D. Chessel & N. Perrin (2002) Ecological-Niche Factor Analysis: How to Compute Habitat-Suitability Maps Without Absence Data? *Ecology*, **83**, 2027-2036.

Hortal, J., Lobo, J.M. & Jimenez-Valverde, A. (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conserv Biol*, **21**, 853-63.

Huang, M., Kong, X., Varela, S. & Duan, R. (2016) The Niche Limitation Method (NicheLim), a new algorithm for generating virtual species to study biogeography. *Ecological Modelling*, **320**, 197-202.

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symp Quantitative Biol.*, **22**, 415-427.

IPCC (2007) Climate Change 2007: Synthesis Report. In:

IUCN (2018) The IUCN Red List of Threatened Species. Version 2018-1. Downloaded on 05 July 2018. In, http://www.iucnredlist.org.

Jackson, S.T. & Overpeck, J.T. (2000) Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, **26**, 194-220.

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

Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, **13**, 853-867.

Kadmon, R., Farber, O. & Danin, A. (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, **14**, 401-413.

Karger, D.N., Conrad, O., Bohner, 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**, 170122.

Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol Lett*, **12**, 334-50.

Keil, P., Belmaker, J., Wilson, A.M., Unitt, P., Jetz, W. & Freckleton, R. (2013) Downscaling of species distribution models: ␟a hierarchical approach. *Methods in Ecology and Evolution*, **4**, 82-94.

Kéry, M. (2011) Towards the modelling of true species distributions. *Journal of Biogeography*, **38**, 617-618.

Kéry, M., Dorazio, R.M., Soldaat, L., Van Strien, A., Zuiderwijk, A. & Royle, J.A. (2009) Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, **46**, 1163-1172.

Lahoz-Monfort, J.J., Guillera-Arroita, G. & Wintle, B.A. (2014) Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504-515.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagner, S. (1996) Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, **90**, 39-52.

Leroy, B., Meynard, C.N., Bellard, C. & Courchamp, F. (2016) virtualspecies, an R package to generate virtual species distributions. *Ecography*, **39**, 599-607.

Liu, C., Newell, G. & White, M. (2018) The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography*, **0**

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385-393.

Liu, C., White, M., Newell, G. & Pearson, R. (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, **40**, 778-789.

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. *Journal for Nature Conservation*, **19**, 1-7.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145-151.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J. & Guisan, A. (2010) Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, **143**, 2647-2657.

Luoto, M., Heikkinen, R.K., Poyry, J. & Saarinen, K. (2006) Determinants of the biogeographical distribution of butterflies in boreal regions. *Journal of Biogeography*, **33**, 1764-1778.

MacArthur, R. & Wilson, E. (1967) The theory of island biogeography: Princeton Univ Pr. *Press, Princeton*,

MacArthur, R.H. & Wilson, E.O. (1963) Equilibrium-Theory of Insular Zoogeography. *Evolution*, **17**, 373-&.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248.

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models. 2nd edition*. Chapman and Hall, London.

McGill, B.J., Dornelas, M., Gotelli, N.J. & Magurran, A.E. (2015) Fifteen forms of biodiversity trend in the Anthropocene. *Trends Ecol Evol*, **30**, 104-13.

McKinney, M.L. & Lockwood, J.L. (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends Ecol Evol*, **14**, 450-453.

Meynard, C.N. & Kaplan, D.M. (2013) Using virtual species to study species distributions and model performance. *Journal of Biogeography*, **40**, 1-8.

Millennium Ecosystem Assessment (2005) Ecosystems and human well-being: Synthesis. In, Washington, DC.

Miller, J.A. (2014) Virtual species distribution models: Using simulated data to evaluate aspects of model performance. *Progress in Physical Geography*, **38**, 117-128.

Mitchell, P.J., Monk, J., Laurenson, L. & Chisholm, R. (2017) Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution*, **8**, 12-21.

Mittelbach, G.G. (2012) *Community ecology*. Sinauer Associates.

Mod, H.K., Scherrer, D., Luoto, M. & Guisan, A. (2016) What we use is not what we know: environmental predictors in plant distribution models. *Journal of Vegetation Science*, **27**, 1308-1322.

Mokany, K., Harwood, T.D., Williams, K.J. & Ferrier, S. (2012) Dynamic macroecology and the future for biodiversity. *Glob Chang Biol*, **18**, 3149-3159.

Morán-Ordóñez, A., Briscoe, N.J. & Wintle, B.A. (2018) Modelling species responses to extreme weather provides new insights into constraints on range and likely climate change impacts for Australian mammals. *Ecography*, **41**, 308-320.

Morán-Ordóñez, A., Lahoz-Monfort, J.J., Elith, J. & Wintle, B.A. (2017) Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions? *Global Ecology and Biogeography*, **26**, 371-384.

Moudrý, V. & Šímová, P. (2012) Influence of positional accuracy, sample size and scale on modelling species distributions: a review. *International Journal of Geographical Information Science*, **26**, 2083-2095.

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Gray, C.L., Scharlemann, J.P.W., Börger, L., Phillips, H.R.P., Sheil, D., Lysenko, I. & Purvis, A. (2016a) Global patterns of terrestrial assemblage turnover within and among land uses. *Ecography*, **39**, 1151-1163.

Newbold, T., Hudson, L.N., Arnell, A.P., Contu, S., De Palma, A., Ferrier, S., Hill, S.L., Hoskins, A.J., Lysenko, I., Phillips, H.R., Burton, V.J., Chng, C.W., Emerson, S., Gao, D., Pask-Hale, G., Hutton, J., Jung, M., Sanchez-Ortiz, K., Simmons, B.I., Whitmee, S., Zhang, H., Scharlemann, J.P. & Purvis, A. (2016b) Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science*, **353**, 288-91.

Newbold, T., Hudson, L.N., Hill, S.L., Contu, S., Lysenko, I., Senior, R.A., Borger, L., Bennett, D.J., Choimes, A., Collen, B., Day, J., De Palma, A., Diaz, S., Echeverria-Londono, S., Edgar, M.J., Feldman, A., Garon, M., Harrison, M.L., Alhusseini, T., Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Correia, D.L., Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P. & Purvis, A. (2015) Global effects of land use on local terrestrial biodiversity. *Nature*, **520**, 45-50.

Normand, S., Ricklefs, R.E., Skov, F., Bladt, J., Tackenberg, O. & Svenning, J.C. (2011) Postglacial migration supplements climate in determining plant species ranges in Europe. *Proc Biol Sci*, **278**, 3644-53.

Olden, J.D., Leroy Poff, N., Douglas, M.R., Douglas, M.E. & Fausch, K.D. (2004) Ecological and evolutionary consequences of biotic homogenization. *Trends Ecol Evol*, **19**, 18-24.

Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37-42.

Pauli, H., Gottfried, M. & Grabherr, G. (1996) Effects of climate change on mountain ecosystems–upward shifting of alpine plants. *World resource review*, **8**, 382-390.

Pearce, J.L. & Boyce, M.S. (2006) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, **43**, 405-412.

Pearson, R.G., Dawson, T.P. & Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285-298.

Pearson, R.G., Thuiller, W., Araujo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704-1711.

Pereira, H.M., Leadley, P.W., Proenca, V., Alkemade, R., Scharlemann, J.P., Fernandez-Manjarres, J.F., Araujo, M.B., Balvanera, P., Biggs, R., Cheung, W.W., Chini, L., Cooper, H.D., Gilman, E.L., Guenette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R. & Walpole, M. (2010) Scenarios for global biodiversity in the 21st century. *Science*, **330**, 1496-501.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Q Rev Biol*, **78**, 419-33.

Peterson, A.T. (2011) *Ecological niches and geographic distributions (MPB-49)*. Princeton University Press.

Peterson, A.T., Papeş, M. & Soberón, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, **213**, 63-72.

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. *Global Ecology and Biogeography*, **26**, 275-287.

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C. & Guisan, A. (2012) Climatic niche shifts are rare among terrestrial plant invaders. *Science*, **335**, 1344-8.

Petitpierre, B., McDougall, K., Seipel, T., Broenniman, O., Guisan, A. & Kueffer, C. (2016) Will climate change increase the risk of plant invasions into mountains? *Ecological Applications*,

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231-259.

Phillips, S.J., Dudik, 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**, 181-97.

Preston, F.W. (1980) Non-Canonical Distributions of Commonness and Rarity. *Ecology*, **61**, 88-97.

Pulliam, H.R. (1988) Sources, Sinks, and Population Regulation. *American Naturalist*, **132**, 652-661.

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349-361.

Qiao, H., Peterson, A.T., Ji, L., Hu, J. & Isaac, N. (2017) Using data from related species to overcome spatial sampling bias and associated limitations in ecological niche modelling. *Methods in Ecology and Evolution*, **8**, 1804-1812.

Ramanathan, V. & Feng, Y. (2008) On avoiding dangerous anthropogenic interference with the climate system: formidable challenges ahead. *Proc Natl Acad Sci U S A*, **105**, 14245-50.

Renner, I.W. & Warton, D.I. (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, **69**, 274-81.

Richardson, K., Steffen, W., Liverman, D., Barker, T., F. Jotzo, Kammen, D., Leemans, R., Lenton, T. & al., e. (2011) *Climate change: Global risks, challenges and decisions.* . Cambridge University Press., Cambridge.

Ripley, B.D. (1996) *Pattern recognition and neural networks*. Cambridge university press.

Robertson, M.P., Peter, C.I., Villet, M.H. & Ripley, B.S. (2003) Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, **164**, 153-167.

Rondinini, C. & Boitani, L. (2006) Differences in the umbrella effects of African amphibians and mammals based on two estimators of the area of occupancy. *Conserv Biol*, **20**, 170-9.

Sala, O.E., Chapin, F.S., 3rd, Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M. & Wall, D.H. (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770-4.

Simpkins, C.E., Dennis, T.E., Etherington, T.R. & Perry, G.L.W. (2018) Assessing the performance of common landscape connectivity metrics using a virtual ecologist approach. *Ecological Modelling*, **367**, 13-23.

Soberon, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecol Lett*, **10**, 1115-23.

Steffen, W., Grinevald, J., Crutzen, P. & McNeill, J. (2011a) The Anthropocene: conceptual and historical perspectives. *Philos Trans A Math Phys Eng Sci*, **369**, 842-67.

Steffen, W., Persson, A., Deutsch, L., Zalasiewicz, J., Williams, M., Richardson, K., Crumley, C., Crutzen, P., Folke, C., Gordon, L., Molina, M., Ramanathan, V., Rockstrom, J., Scheffer, M., Schellnhuber, H.J. & Svedin, U. (2011b) The anthropocene: from global change to planetary stewardship. *Ambio*, **40**, 739-61.

Steffen, W., Rockstrom, J., Richardson, K., Lenton, T.M., Folke, C., Liverman, D., Summerhayes, C.P., Barnosky, A.D., Cornell, S.E., Crucifix, M., Donges, J.F., Fetzer, I., Lade, S.J., Scheffer, M., Winkelmann, R. & Schellnhuber, H.J. (2018) Trajectories of the Earth System in the Anthropocene. *Proc Natl Acad Sci U S A*, **115**, 8252-8259.

Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1-13.

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C. & Guisan, A. (2014) Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution*, **5**, 947-955.

Thuiller, W., Brotons, L., Araújo, M.B. & Lavorel, S. (2004a) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165-172.

Thuiller, W., Araujo, M.B., Pearson, R.G., Whittaker, R.J., Brotons, L. & Lavorel, S. (2004b) Biodiversity conservation: uncertainty in predictions of extinction risk. *Nature*, **430**, 1 p following 33; discussion following 33.

Tilman, D., Clark, M., Williams, D.R., Kimmel, K., Polasky, S. & Packer, C. (2017) Future threats to biodiversity and pathways to their prevention. *Nature*, **546**, 73-81.

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. *Ecography*, **37**, no-no.

Vellend, M. (2010) Conceptual synthesis in community ecology. *Q Rev Biol*, **85**, 183-206.

Vellend, M. (2016) *The theory of ecological communities (MPB-57).* Princeton University Press.

Vicente, J., Alves, P., Randin, C., Guisan, A. & Honrado, J. (2010) What drives invasibility? A multi-model inference test and spatial modelling of alien plant species richness patterns in northern

Portugal. *Ecography*, **33**, 1081-1092.

Vicente, J.R., Fernandes, R.F., Randin, C.F., Broennimann, O., Goncalves, J., Marcos, B., Pocas, I., Alves, P., Guisan, A. & Honrado, J.P. (2013) Will climate change drive alien invasive plants into areas of high protection value? An improved model-based regional assessment to prioritise the management of invasions. *J Environ Manage*, **131**, 185-95.

Vicente, J.R., Alagador, D., Guerra, C., Alonso, J.M., Kueffer, C., Vaz, A.S., Fernandes, R.F., Cabral, J.A., Araujo, M.B. & Honrado, J.P. (2016) Cost-effective monitoring of biological invasions under global change: a model-based framework. *Journal of Applied Ecology*, **53**, 1317-1329.

Vilà, M. & Ibáñez, I. (2011) Plant invasions in the landscape. *Landscape Ecology*, **26**, 461-472.

Wake, D.B. & Vredenburg, V.T. (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *PNAS*, **105**, 11466-11473.

Whittaker, R.H. (1967) Gradient analysis of vegetation. *Biol Rev Camb Philos Soc*, **42**, 207-64.

Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. (1998) Quantifying threats to imperiled species in the United States. *Bioscience*, **48**, 607-615.

Wisz, M.S. & Guisan, A. (2009) Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecol*, **9**, 8.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & Distribut, N.P.S. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763-773.

Wisz, M.S., Broennimann, O., Grønkjær, P., Møller, P.R., Olsen, S.M., Swingedouw, D., Hedeholm, R.B., Nielsen, E.E., Guisan, A. & Pellissier, L. (2015) Arctic warming will promote Atlantic–Pacific fish interchange. *Nature Climate Change*, **5**, 261-265.

World Meteorological, O. & Kishocho (2006) WMO greenhouse gas bulletin : the state of greenhouse gases in the atomosphere based on global observations. *WMO greenhouse gas bulletin : the state of greenhouse gases in the atomosphere based on global observations.*,

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B. & Grimm, V. (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622-635.

Zurell, D., Thuiller, W., Pagel, J., Cabral, J.S., Munkemuller, T., Gravel, D., Dullinger, S., Normand, S., Schiffers, K.H., Moore, K.A. & Zimmermann, N.E. (2016) Benchmarking novel approaches for modelling species range dynamics. *Glob Chang Biol*, **22**, 2651-64.

PART I – MAIN CHAPTERS OF THE THESIS

Chapter 1.1

Effects of simulated observation errors on the performance of species distribution models

This manuscript was accepted in *Diversity and Distributions*

My contribution to the paper: I conceived the ideas and developed the methodological framework in coordination with Antoine Guisan and Daniel Scherrer. I also developed the codes to create and run the data (with important help from Daniel Scherrer), analysed that data and led the writing of the different versions of the manuscript.

Effects of simulated observation errors on the performance of species distribution models

Rui F. Fernandes[a], Daniel Scherrer[a] and Antoine Guisan[a,b]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland
[b] Institute of Earth Surface Dynamics, University of Lausanne, Geopolis, 1015 Lausanne, Switzerland

ABSTRACT

Species distribution information is essential under increasing global changes and models can be used to acquire such information, but they can be affected by different errors/bias. Here, we evaluated the degree to which errors in species data (false presences-absences) affect model predictions and how this is reflected in commonly used evaluation metrics. Using 100 virtual species and different sampling methods, we created observation datasets of different sizes (100-400-1600) and added increasing levels of errors (creating false positives or negatives; from 0% to 50%). These degraded datasets were used to fit models using generalised linear model, random forest and boosted regression trees. Model fit (ability to reproduce calibration data) and predictive success (ability to predict the true distribution) were measured on probabilistic/binary outcomes using Kappa, TSS, MaxKappa, MaxTSS and Somers'D (rescaled AUC). The interpretation of models' performance depended on the data and metrics used to evaluate them, with conclusions differing whether model fit, or predictive success were measured. Added errors reduced model performance, with effects expectedly decreasing as sample size increased. Model performance was more affected by false positives than by false negatives. Models with different techniques were differently affected by errors: models with high fit presenting lower predictive success (RFs), and vice-versa (GLMs). High evaluation metrics could still be obtained with 30% error added, indicating that some metrics (Somers'D) might not be sensitive enough to detect data degradation. Our findings highlight the need to reconsider the interpretation scale of some commonly used evaluation metrics: Kappa seems more realistic than Somers'D/AUC or TSS. High fits were obtained with high levels of error added, showing that RF overfits the data. When collecting occurrence databases, it is advisory to reduce the rate of false positives (or increase sample sizes) rather than false negatives.

KEYWORDS: Artificial data, evaluation metric, AUC, TSS, Kappa, model fit, predictive accuracy, uncertainty, ecological niche models, habitat suitability models

INTRODUCTION

As biodiversity and ecosystems are under growing pressure by global changes, we need to urgently increase our understanding of, and associated capacity to model, the main factors driving changes in the distributions of species, assemblages, and ecosystems (Dawson *et al.*, 2011). Species distribution models (SDMs; Guisan *et al.*, 2017) allow modeling the distribution of species and their assemblages at different spatial and temporal scales (Ferrier & Guisan, 2006; D'Amen *et al.*, 2017). SDMs statistically correlate species observations (presence-absence or presence-only) with environmental data (Guisan & Thuiller, 2005) and are commonly evaluated by assessing their predictive performance and accuracy (Peterson *et al.*, 2011). The most used metric is, by far, the area under the receiver-operating characteristic curve (AUC-ROC) (Fourcade *et al.*, 2018). It is calculated by plotting a model's sensitivity against its false positive rate at all possible thresholds (Hanley & McNeil, 1982), measuring the model's performance in discriminating between species presences and absences (Lobo *et al.*, 2008). Alternative metrics have also been proposed, mainly due to the known limitations of the AUC (e.g. dependence on the calibration area, ignores spatial distribution of errors, relies on the ranking of sensitivity/specificity across thresholds and ignores the probability values given by a model or equally weights omission/commission errors; Lobo *et al.*, 2008; Peterson *et al.*, 2008; Jiménez-Valverde, 2012; Jiménez-Valverde *et al.*, 2013). The most common alternatives are Cohen's Kappa (Kappa; Cohen, 2016) and the True Skill Statistic (TSS; Allouche *et al.*, 2006). Kappa corrects the overall accuracy of model predictions by the accuracy expected to occur by chance while TSS corrects Kappa's dependency on prevalence (see Table 1 for more information). Moreover, SDMs can contain uncertainty from various sources (reviewed by e.g. Barry & Elith, 2006; Beale & Lennon, 2012), including errors associated with species data (e.g. unavailable absence data, small or insufficient sample sizes, unexplored geographical bias or spatial errors; e.g. Fielding & Bell, 1997; Pearce & Ferrier, 2000; Jenkins *et al.*, 2003), environmental variables (e.g. missing important ones; Mod *et al.*, 2016) or modelling techniques (e.g. Guisan *et al.*, 2007b; Thibaud *et al.*, 2014). One problem commonly affecting SDMs concerns the inability to separate potentially false and true species' absences obtained through field surveys (Lahoz-Monfort *et al.*, 2014) leading to underestimation of species occupancy (i.e. when occupied sites are misclassified as unoccupied; Guillera-Arroita *et al.*, 2010), incorrect inference about species distributions, or inaccurate predictions (Lahoz-Monfort *et al.*, 2014). The wrongly recorded absences (false absences) in presence-absence datasets or the omission of presences in presence-only models can then lead to predictions that will reflect where the species is more or less likely to be detected instead of the locations where if should occur or not (Kéry, 2011; Lahoz-Monfort *et al.*, 2014). This means that one would eventually model what is called the "apparent distribution" and not the true distribution (Kéry, 2011). Additionally, some environmental relationships that are important to explain species occurrence and distribution might be wrongly identified or completely missed when false absences/presences are recorded (Kéry, 2011). The effect of detection errors on model performance is likely to depend on the modelling techniques used as those differ in their ability to fit complex response curves (i.e., species-environment relationships; Guisan *et al.*, 2007b; Merow *et al.*, 2014).

Several of these issues have received considerable attention in recent years, providing information to improve survey designs, proposing approaches to account for imperfect detection, and evaluating the impacts of non-detection of species in models of individual species (MacKenzie *et al.*, 2002; Gu & Swihart, 2004; Guillera-Arroita *et al.*, 2010). However, the majority of the studies focusing on uncertainties in SDMs used real species observations, putting a limit to proper assessment of model accuracy because the complete distribution and the species-environment relationships cannot be entirely known and may result from factors that can’t be controlled. A way to avoid these limitations is to use artificial data (Austin *et al.*, 2006) in a virtual ecologist approach (see Zurell *et al.*, 2010 for a review), where all the information necessary for a study can always be obtained in a fully artificial or semi-artificial world, allowing complete or at least partial control on the data and models being tested (Austin *et al.*, 2006). In one of the first application to SDMs, Hirzel *et al.* (2001) created virtual species to test different habitat suitability methods and their predictive power under different scenarios. Virtual species have been used to test different ecological models and assumptions, to test different approaches to sample species data (Hirzel & Guisan, 2002), to downscale coarse-grain data into high-resolution predictions (Bombi & D'Amen, 2012) or to measure the relative effect of different factors affecting predictions (Thibaud *et al.*, 2014).

In this study, we take a virtual ecologist approach, using 100 virtual species defined from real observations in a real mountain landscape with large environmental gradients, to investigate: (1) the effect of sample size when error is added to the data; (2) the model performance behaviour when different levels of errors are added to the training data (to presences or absences) and how different evaluation approaches influence the conclusions of that performance, (3) how different metrics traditionally used to evaluate SDM predictions perform with those errors (4) what are the implications for interpreting the performance/reliability of models when using those metrics, (5) how different modelling techniques deal with degraded training data, and (6) how different types of errors affect models and metrics. Taking into account the frequent use of SDMs in ecology, evolution and conservation, this paper provides an essential analysis of the potential effects of errors in species data on SDM reliability and on the interpretation of common evaluation metrics.

METHODS

Analytical framework

We implemented a virtual ecologist approach (see Figure 1), based initially on real data in a real landscape (i.e. which can also be considered as a semi-virtual study; Albert *et al.*, 2010) in the western

Swiss Alps (a priority research area; http://rechalpvd.unil.ch), covering approximately 700 km^2. We defined the distributions of virtual species based on predictions of models fitted on real data in this study area to keep ecological realism (see step 1 below). The approach consisted of five steps:

Step 1. *Creating virtual species*

From a set of real species data (previously sampled in the study area), we generated 100 virtual species, by fitting SDMs (initial SDMs in Fig.1) using presence-absence data against five environmental predictors: summer mean monthly temperatures (2 to 19°C), sum of winter precipitation (65 to 282 mm), annual sum of potential solar radiation (KJ), slope (°) and topographic position (unitless; indicating ridges and valleys; see Supplementary Information Appendix 1).

The models were fitted using generalized linear models (GLMs; McCullagh & Nelder, 1989), random forests (RFs; Breiman, 2001) or boosted regression trees (BRTs; Friedman *et al.*, 2000) as modelling techniques. These modelling techniques were chosen because GLMs allow hump-shaped and linear response curves that can be easily justified by ecological niche theory while RFs and BRTs have been increasingly used in recent years as they allow for more complex combinations and interactions of environmental factors, which can result in more complex species-environment relationships. This study setup allowed us to check if the complexity of those relationships could influence the outcome of our study.

The resulting probability distributions were transformed into presence-absence data (considered as our "true" virtual species distribution) using three thresholding approaches: (1) threshold that corresponded to the point on the receiver operating characteristic plot (ROC; sensitivity against 1-specificity across successive thresholds; Hanley & McNeil, 1982; Swets, 1988) with the shortest distance to the top-left corner (0,1) of that plot (Cantor *et al.*, 1999); (2) threshold maximizing Kappa (MaxKappa; Huntley *et al.*, 1995; Guisan *et al.*, 1998); and (3) threshold maximizing TSS (MaxTSS; which is equivalent to the sensitivity-specificity sum maximization described in Liu *et al.*, 2005). By using a number of different thresholding techniques, we minimize the bias of thresholding techniques on the interpretation of the results.

In this study all initial environmental and species data was available at a 25 m resolution. In real word studies, the spatial resolution can have an important influence on model predictions, with diverging results being observed between small and large-scale studies (e.g. Meyer & Thuiller, 2006; Mertes & Jetz, 2018; Record *et al.*, 2018), or when changing resolution or extent (e.g. Thuiller *et al.*, 2004; Guisan *et al.*, 2007a). This can for instance result from the scale dependency of the environmental predictors (Vicente *et al.*, 2014) and spatial stochastic effects at smaller spatial scales (Steinmann *et al.*, 2011; Scherrer *et al.*, 2018b). As a result, the distribution of real species cannot usually be fully explained by the abiotic predictors, as dispersal and biotic factors also play a role and interact with scale (Soberon & Nakamura, 2009). Here, we avoid this problem by using a virtual species approach, with the same predictors being used to create the species and fit their distribution models, and therefore the initial species distributions are fully explained by the chosen predictors at the study scale (extent and resolution). This approach guaranteed that the virtual species showed realistic response curves for our landscape resulting in realistic species assemblages. In theory, the resolution should thus not matter in our study, and should not affect our findings. All models were run in R software version 3.3.3 (R Core Team, 2017), using biomod2 default settings (Thuiller *et al.*, 2009), as in most published studies.

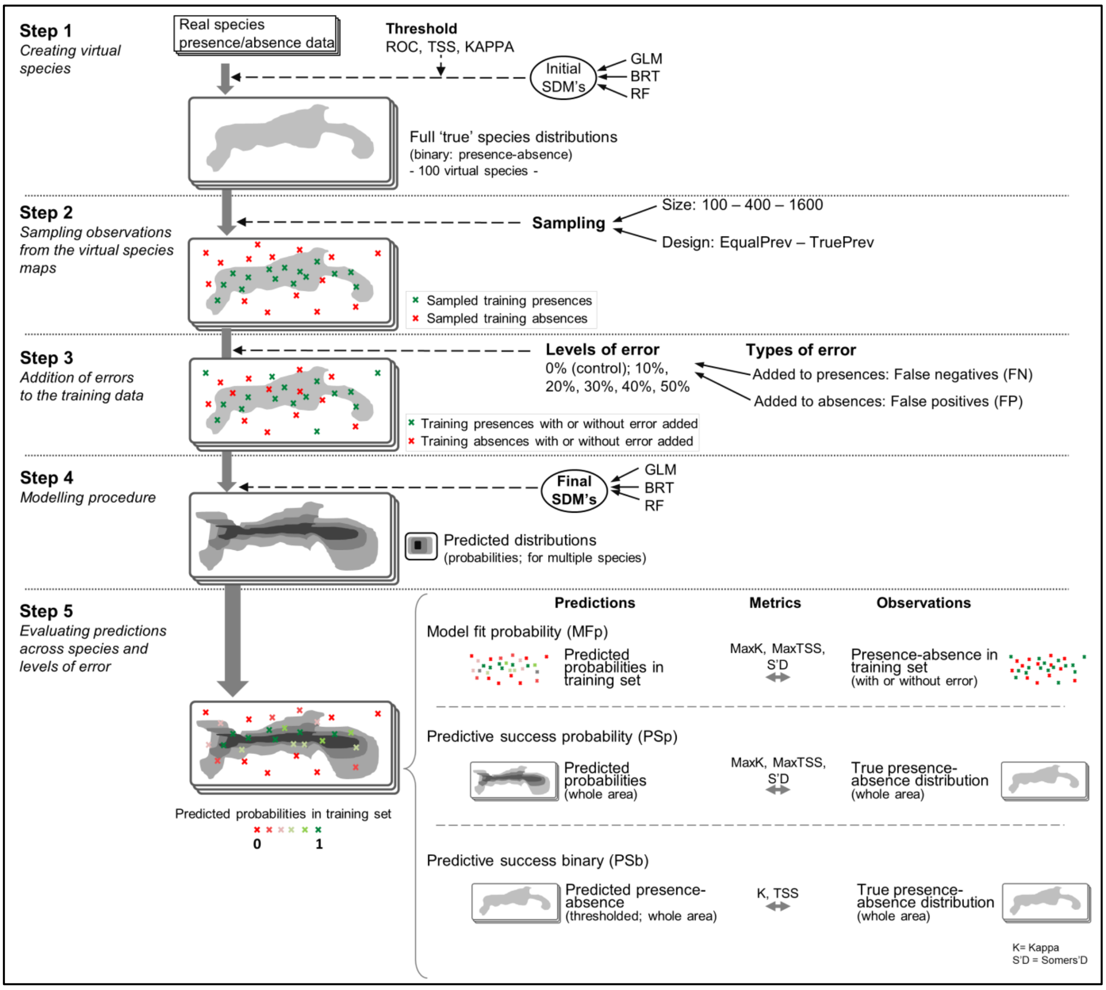


FIGURE 1 Workflow of the analytical steps followed in the study. **Step 1 –** We started by creating binary distribution maps for 100 virtual species from models based on real species' data (using either generalized linear models (GLM), boosted regression trees (BRT) or random forests (RF) as modelling techniques and the receiver operating characteristic (ROC), true skill statistic (TSS) or KAPPA as thresholding techniques). **Step 2 –** For each species, we sampled presence-absence data using three different sample sizes (100-400-1600) and two sampling designs (EqualPrev and TruePrev). **Step 3 –** To each of the sampled datasets, errors were added according to 6 different levels (0% - training data without error added, the control; 10%, 20%, 30%, 40% and 50% - training data with error added) and two different types of error (errors added to presences, creating false negatives or added to absences, creating false positives). **Step 4 –** Each occurrence dataset was used to create single species distribution models (probability and binary maps), using three different modelling techniques (GLM, BRT and RF). *Step 5* – The predictions for each species were then evaluated with three evaluation approaches: model fit probability (MFp), predictive success probability (PSp) and predictive success binary (PSb), using different metrics: maximized Kappa (MaxKappa), maximized TSS (MaxTSS) and Somers' D (rescaled measure of AUC) for MFp and PSp; Kappa and TSS for PSb.

Step 2. *Sampling observations from virtual species maps*

Virtual species presence-absence data was randomly sampled to create training datasets of different sizes (100-400-1600) using two sampling designs: (i) selection of species data with equal number of presences-absences (equal prevalence; “EqualPrev”); and (ii) selection of species data taking into account a presence-absence ratio reflecting the true prevalence of the species (species true prevalence; “TruePrev”). These datasets served as control (0% error) to establish a baseline of the potential (re-) sampling bias for the different sampling schemes, modelling techniques and species (Step 2, Fig. 1).

Step 3. *Addition of errors to the training data*

For each sampling design and size, errors were randomly added (using software R) to the training data according to six different levels (i.e. “levels of error”; 0% (no error added), 10%, 20%, 30%, 40% and 50%). These errors were added either to presences only (creating false negatives (FN) by changing presences to absences), or to absences only (creating false positives (FP) by changing absences to presences) (Step 3, Fig. 1).

Step 4. *Modelling procedure*

The control dataset (without error) and all the datasets with errors added were used to create SDMs (final SDMs in Step 4, Fig.1) using the same environmental predictors and modelling techniques employed to initially create the virtual species. This ensures that – without error added (i.e. controls) - the models can potentially replicate perfectly the distributions of our species, since all information that initially defined these distributions is available (i.e. same predictors) and the response curves could be fitted perfectly (i.e. if using the same technique). The only factors that can affect the performance are therefore the sample size, the change of modelling technique, the threshold method and the errors added, which we can untangle through the control and the known full distribution. In other words, having SDM predictions for our control and degraded datasets allowed us to distinguish decreases in model performance only caused by resampling (using the control dataset), the thresholding effect and from the effects caused by the errors added to the presences (FN) and/or the absences (FP).

Step 5. *Evaluating predictions across species and levels of errors*

Finally, we evaluated all predictions built for each sample size, sampling design, modelling technique and threshold approach by measuring model fit on probability (MFp) at sampled sites and predictive success for probabilistic (PSp) and binary predictions (PSb) across the whole area (i.e. evaluation approaches; see description below), using five widely used agreement/evaluation metrics (for more information see Table 1 and Liu *et al.*, 2005): Cohen’s Kappa (Kappa), True Skill Statistic (TSS), maximized Kappa (MaxKappa), maximized TSS (MaxTSS) and a rescaled measure of AUC, Somers’ rank correlation (Somers'D; Harrell, 2015). Somers’D was used instead of AUC, because its rescaled

between -1 and +1, making it directly comparable to the other used evaluation metrics (and is therefore also intuitively interpretable along the same scale as a correlation coefficient).
Depending on the evaluation data used (i.e. evaluation approach hereafter), different evaluation metrics were used. For MFp/PSp we calculated MaxKappa, MaxTSS and Somers'D, while for PSb only observed Kappa and TSS under a chosen threshold could be calculated (Step 5, Fig. 1):

(i) *Model fit probability* (MFp) corresponds to the ability of the model to reproduce the training data. It was measured by comparing predicted probabilities of the different models (control and the various levels of errors) to the data used to fit those models, and thus was conducted on the same set of points used to build the models (presence-absence in training dataset with errors added; and without errors for the control).

(ii) *Predictive success probability* (PSp) is the potential of the model to recreate the complete true distribution of a species when the model is trained with degraded (or not) training data. It was calculated by comparing predicted probabilities of the different models (control and various levels of errors) to the original true species distribution map (presence-absence), giving Somers'D, MaxTSS and MaxKappa across the whole study area.

(iii) *Predictive success binary* (PSb) is the ability of the model to predict the complete true distribution of the species based on the degraded (or not) training data, using only information available to the model (no information about the truth available for threshold selection). It was calculated by comparing binary predictions of the different models (control and various levels of errors), to the complete true distribution dataset. To create binary predictions, MaxTSS (for the calculation of TSS) and MaxKappa (for the calculation of Kappa) thresholds were selected based on the predicted probabilities and the training data used in each model (calibration data with error).

Evaluating model predictions with the control data (no error added) allows to measure the effect of sampling and, more particularly, since the sampling design was random, to assess the effect of sample size. Also, to assess if evaluation values decrease with increasing errors in the training data, we standardised all our degraded models with the corresponding control (0% error) to eliminate resampling effects (see Results; difference = [evaluation value of degraded model – evaluation value of control model]). Therefore, negative values indicate that model performance decreased compared to the control (i.e. the higher the decrease, the higher the effects of errors added).

TABLE 1 Detailed information about the evaluation metrics used to assess the predictive performance of SDMs (adapted from Liu *et al.*, 2005 and Allouche *et al.*, 2006), a is true positives (or presences), b is false positives (or presences), c is false negatives (or absences), d is true negatives (or absences), n (=a+b+c+d) is the total number of sites. Sensitivity is the probability that the model will correctly classify a presence (a/a+c). Specificity is the probability that the model will correctly classify an absence (d/b+d).

Metric	Acronym	Definition/Formula	Scale	Reference
Area under the receiver operating curve	AUC	Calculated by plotting a model's sensitivity against its false positive rate at all possible thresholds	0/+1	Hanley & McNeil, 1982
Somers' rank correlation	Somers'D	2*(AUC - 0.5)	-1/+1	Harrell, 2015
Cohen's Kappa	Kappa	$\frac{\left(\frac{a+d}{n}\right) - \frac{(a+b)(a+c)+(c+d)(d+b)}{n2}}{1 - \frac{(a+b)(a+c)+(c+d)(d+b)}{n2}}$	-1/+1	Cohen, 1960
True Skill Statistic	TSS	Sensitivity + Specificity – 1	-1/+1	Allouche *et al.,* 2006
Kappa maximization	MaxKappa	Kappa statistic is maximized	-1/+1	Huntley *et al.,* 1995; Guisan *et al.,* 1998
TSS maximization	MaxTSS	TSS statistic is maximized	-1/+1	Liu *et al.,* 2005

RESULTS

Model evaluation using training data without errors added: effects of sampling

Evaluation values increased with increasing sample size, regardless of the sampling design ("EqualPrev" and "TruePrev"; Figure 2) with the exception of Model fit probability (MFp) which decreased when models were fitted by GLMs/BRTs.

The MFp for the initial models (i.e. 0% with no errors added) were always above 0.75 for all modelling techniques and metrics (except MaxKappa in "TruePrev") and mostly close to 1 (which can be considered an excellent model) when species were created by GLMs or fitted using BRT/RF. In contrast, the Predictive success probability (PSp) and Predictive success binary (PSb) showed much higher variation, ranging from 0.75 to 1 for all metrics when species were created by GLMs, but from 0.25 to 1 when created by BRT/RF (Fig. 2).

Somers'D presented always the highest evaluation values, usually followed by MaxTSS and MaxKappa (Fig. 2; MFp/PSp). MaxKappa was the metric that presented the greatest range of variation, while models evaluated by Somers'D presented very similar values. When PSb was measured, TSS had the highest values and Kappa the lowest, independently of the modelling technique used (Fig. 2).

Models fitted using species created by GLMs showed the highest evaluation values (usually above 0.75 for all metrics; Fig. 2). However, models fitted using virtual species created by BRTs/RFs presented a wider range of values, with model performance being worse than when species were created by GLMs. Independently of the modelling technique used to create the species, we observed that models fitted by RFs had higher evaluation values of MFp/PSp while models fitted by GLMs presented the highest values of PSb.

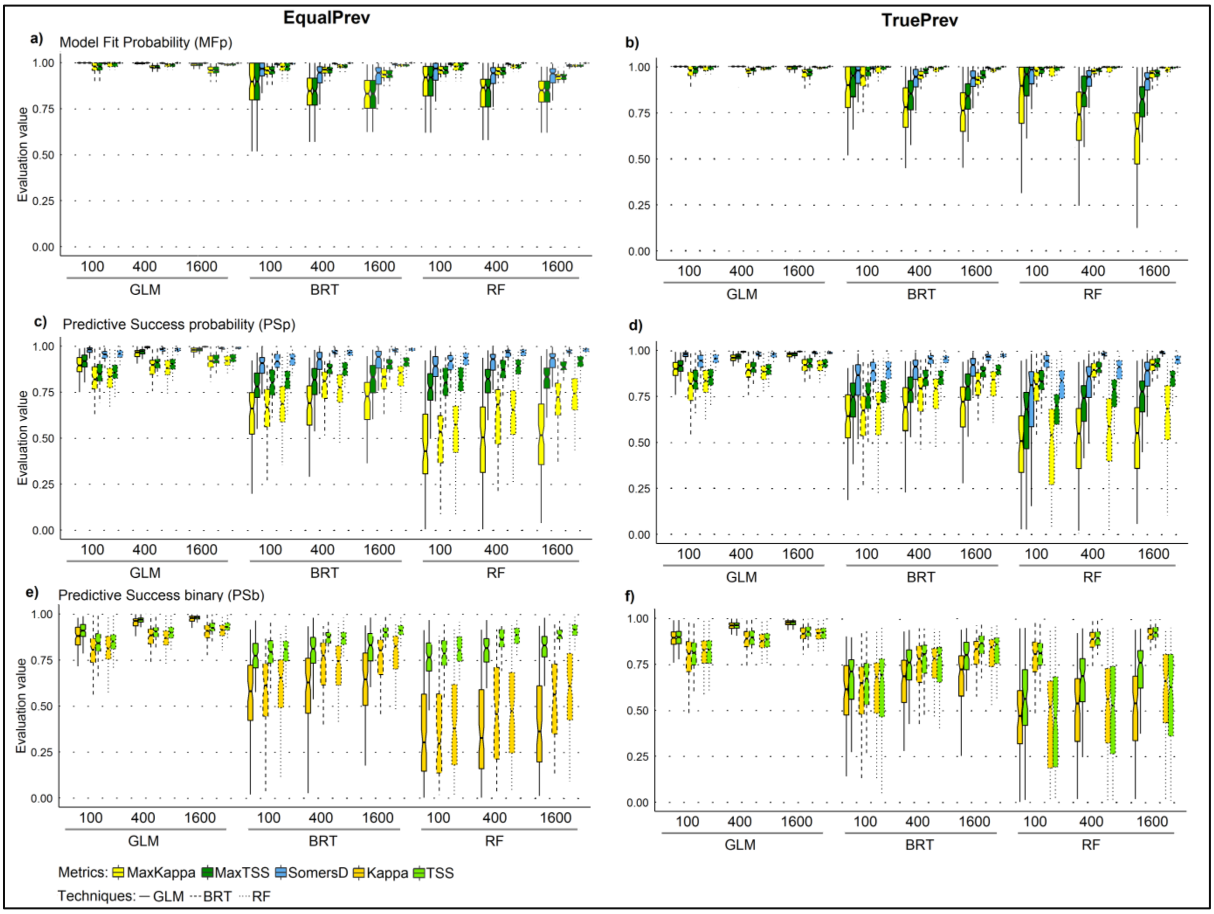


FIGURE 2 Evaluation values of control model (using training data without errors added; 0%) for “EqualPrev” (left column) and “TruePrev” (right column) sampling designs, with measured MFp, PSp and PSb for virtual species (n=100), created using generalized linear models (GLM), boosted regression trees (BRT) or random forests (RF) (initial SDMs) and with different sample sizes (100-400-1600). Model fit probability (MFp) and Predictive success probability (PSp) were measured using maximized Kappa (MaxKappa; yellow), maximized TSS (MaxTSS; green) and Somers’D (blue), while Predictive success binary (PSb) was measured using Kappa (gold) and true skill statistic (TSS; light green). For each sample size, three sets of three boxplots are displayed, corresponding to models fitted (final SDMs) using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots) and evaluated with corresponding metrics. The same applies to PSb, but only two boxplots are displayed in each of the three model sets, corresponding to the two metrics used.

Effects on model evaluation of adding errors to the training data

As the patterns observed across sample sizes were similar, we only report results on the intermediate sample size (i.e. 400; but see Appendix 2-3 for complete results on “EqualPrev” and “TruePrev” sampling designs, respectively). The effect of error added decreased with sample size, with more accurate models being observed at higher sample sizes (i.e. difference between control and degraded models was smaller).

Regardless of the evaluation approach, as errors were increasingly added to training data, evaluation values decreased when compared with the control models (Figure 3). This decrease in model performance was more pronounced in Model fit probability and Predictive success binary (MFp/PSb). Still, models whose performance decreased the most in each approach depended on the modelling technique used (Fig. 3-4). As a result, random forests (RF) presented higher model performance when MFp/PSp were measured and generalized linear models (GLM) when PSb was measured.

In general, the creation of false positives (FP) (Fig. 3, left) had a stronger negative effect on model performance than false negatives (FN) (Fig. 3, right), but in some rare cases the creation of FN could have a stronger effect on model performance (e.g. Fig. 3e-f, PSb evaluated by TSS in models fitted by RFs).

Somers'D displayed the smallest decrease in model performance when errors were added to the calibration data (for MFp/PSp), regardless the technique used to fit the models (Fig. 3) or to create the virtual species (Appendix 2). The strongest decrease in model performance (for MFp/PSp) was usually presented by MaxKappa, while MaxTSS presented intermediate values (Fig. 3). When measuring PSb, true skill statistic (TSS) usually showed a smaller decrease in model performance (Fig. 3e-f), except when creating false negatives for species generated either by BRTs (e.g. Fig. S5, Appendix 2) or RFs (e.g. Fig. S8, Appendix 2). The results obtained with sampling design "TruePrev" (Fig. 4 and Appendix 3) did not differ from those previously described in "EqualPrev" (Fig. 3 and Appendix 2), except when models were evaluated with MaxKappa (Fig. 4a-b; FN have a stronger effect on model performance when fitted by GLMs) or Kappa (Fig. 4f). However, this is most likely an artefact of the difference in the number of species with successful models, which was lower when FN were created. This difference in number was due to the presence reduction of some species when adding FN, making it impossible to correctly fit a model.

Additionally, the use of different threshold techniques (to create initial models) didn't bias the results and their interpretation, with the same patterns being observed across techniques (see Appendix 4).

After increasingly degrading the data and when MFp/PSp were measured, models fitted by GLMs (Figs 3-4) presented the highest decrease when compared with control models. On the opposite side, models fitted by RFs (Figs 3-4) were the least affected by the addition of degraded data. Still, when measuring PSb, the decrease in model performance was higher for models fitted by RFs and more stable for models fitted by GLMs (especially as the errors added increased).

We performed an additional evaluation approach, *predictive success on calibration data* (PSc), not providing the results here since it's a subset of PSb and accordingly yielded similar patterns (but see Appendix 5).

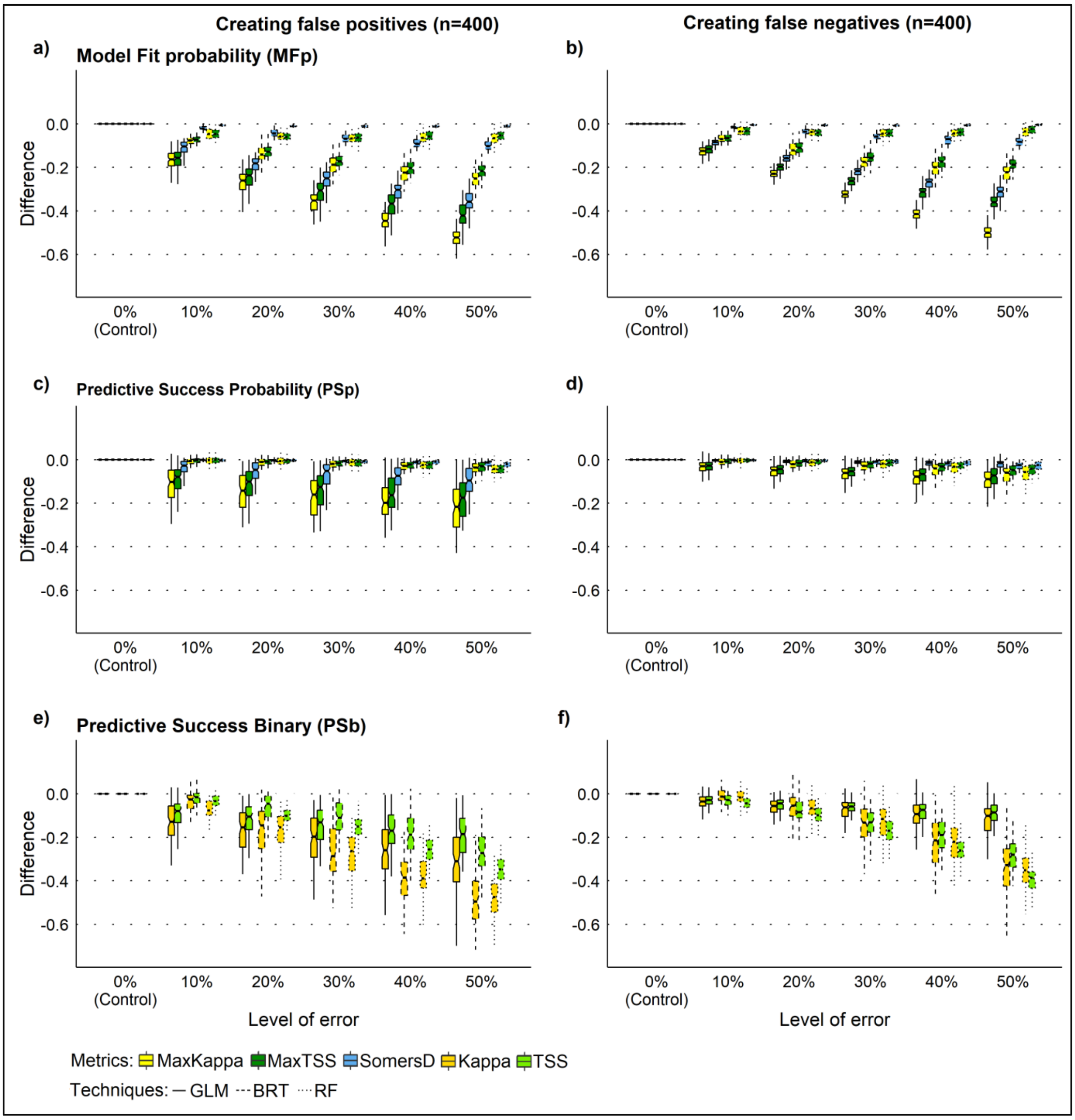


FIGURE 3 Observed difference of measured Model fit probability (**MFp)**, Predictive success probability (**PSp**) and Predictive success binary (**PSb**) between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **400**, for virtual species created using **GLM** (generalized linead models). Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using maximized Kappa (MaxKappa; yellow), maximized TSS (MaxTSS; green) and Somers'D (blue), while PSb was measured using Kappa (gold) and true skill statistic (TSS; light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets.

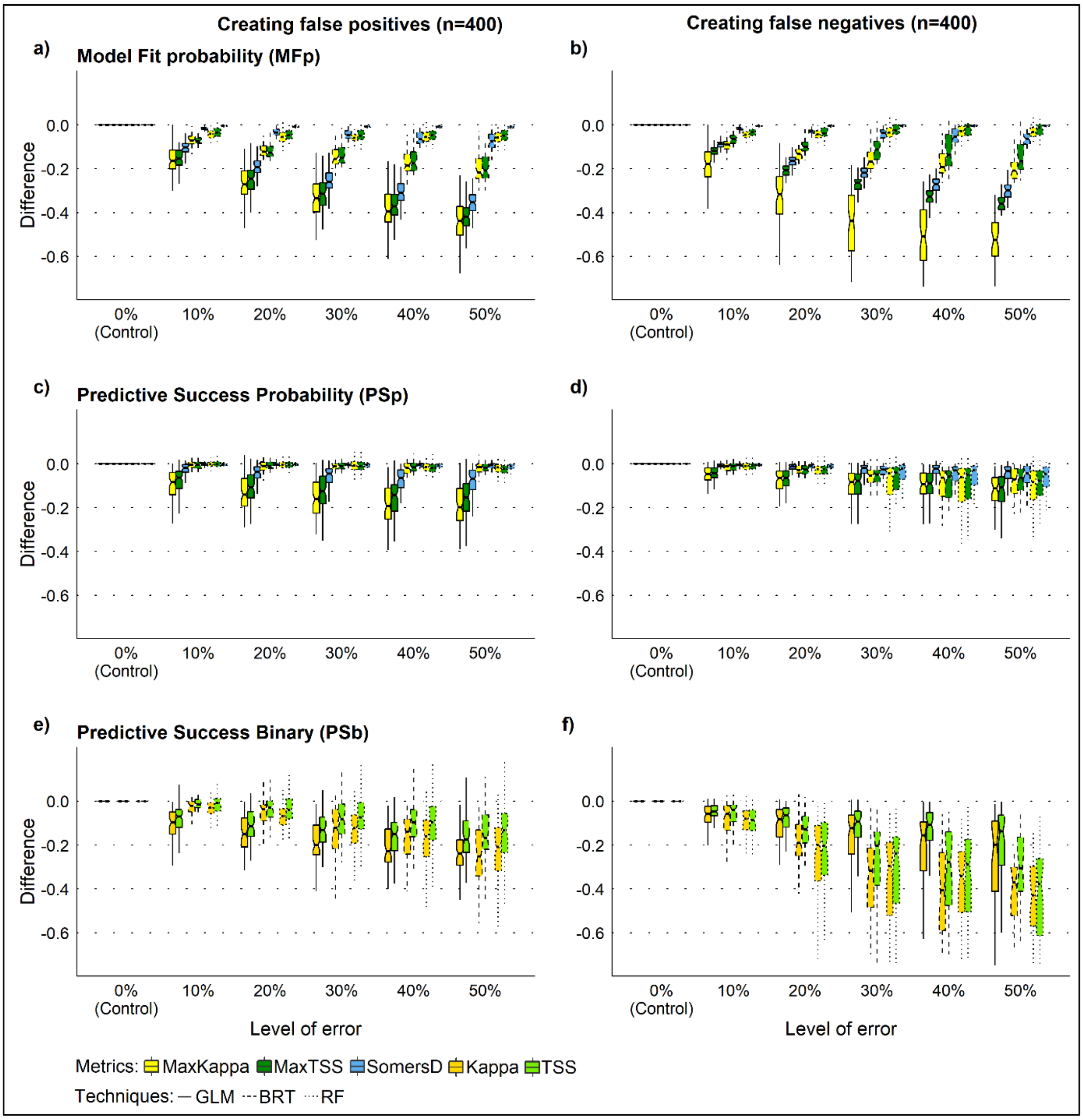


FIGURE 4 Observed difference of measured Model fit probability (**MFp**), Predictive success probability (**PSp**) and Predictive success binary (**PSb**) between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **400**, for virtual species created using **GLM** (generalized linear models). Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using maximized Kappa (MaxKappa; yellow), maximized TSS (MaxTSS; green) and Somers'D (blue), while PSb was measured using Kappa (gold) and true skill statistic (TSS; light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets.

DISCUSSION

We used a virtual ecologist approach with artificial species data to evaluate the degree to which errors in presences/absence data (see Tyre *et al.*, 2003; Graham *et al.*, 2004; Guillera-Arroita *et al.*, 2010 for examples of causes like false-negative errors or imperfect detection, taxonomic inaccuracies or biases in the spatial coverage of data) can affect SDM predictions and assess the reliability of currently used evaluations metrics. By using artificial data, we prevented limitations of real world data (most previous studies used real species data from surveys, herbaria or museums; e.g. Hernandez *et al.*, 2006; Osborne & Leitao, 2009; Mitchell *et al.*, 2017), allowing us to have complete knowledge of the full species distribution and to simulate errors in presence/absence data with complete control of the factors affecting their distribution. The models must then find a signal in the degraded (or not) training data and be able to predict to the known remaining distribution which is largely unaffected by errors. Our work revealed four main findings. First, as expected, the effect of degraded data decreased as sample size increased. Second, the classification of a model along a range of performance (e.g. poor, fair, good, excellent) strongly depended on the metric used to evaluate it. Models evaluated by Somers'D (a rescaled measure of the AUC) still corresponded to high values of predictive performance (according to the interpretation scales as in Araujo *et al*., 2005; fail: AUC<0.7, fair: >0.7, good >0.8, excellent >0.9; refined from the initial scale by Swets, 1988, note nr 11). This suggests that whatever the modelling technique used, AUC, Somers D and related metrics produce overoptimistic evaluations, potentially affecting the conclusion of studies that rely solely on it (e.g. conservation prioritization studies, assessment of future climate change impacts on plants or animals, current and future threats and spread of invasive species). However, other metrics, such as Kappa (or MaxKappa), can provide more realistic evaluations. Third, we confirmed that predictions with too good model fit (MFp) usually presented low predictive success (PSb), with data-driven techniques such as RF usually tending toward higher overfitting and lower prediction success, while model-driven techniques like GLMs showing the opposite (Randin *et al.*, 2006; Petitpierre *et al.*, 2017). Fourth, the creation of false positives had a stronger effect in decreasing model performance than the creation of false negatives. We discuss these findings below.

Confirming the effect of sample size and controlling for it

We found that effects of degraded data consistently decreased with sample size, showing sample size as an important factor affecting model performance. This relationship between model performance and sample size is well known (e.g. Stockwell & Peterson, 2002; Wisz *et al.*, 2008; Thibaud *et al*., 2014; Mitchell *et al.*, 2017). It can be partially explained by the fact that with greater number of presence/absence data, a more complete (broader) information about the occupied environmental space will likely be available. This improves parameter definition, leading to more accurate predictions (Carroll & Pearson, 1998). Our results could be useful since we showed that accurate models (i.e. when all the metrics show high evaluation values) could be generated even when substantial levels of errors (>30%) are present in the training data (if a large enough sample is collected and the adequate modelling techniques are used).

Importance of contrasting model fit and predictive success

Different conclusions about model performance can be inferred depending on how model performance is measured (i.e. our different evaluation approaches). To our knowledge, this is the first study to formally test and compare the outputs of these evaluation approaches to assess model predictions. This was possible through the use of virtual species allowing us to simultaneously assess how well models reproduced the partially degraded training data (MFp), how well they predicted the true distribution of species despite the added errors, taking into account restrictions of real world data (PSb) and without those restrictions when the evaluation with the complete distribution knowledge was available (PSp). Studies with real data have contrasted model fit, internal validation and external validations (e.g. Randin *et al.*, 2006; Wenger & Olden, 2012; Petitpierre *et al.*, 2017), which is distinct from what was done here using and only possible with artificial data.

As expected, all evaluation approaches showed a decrease in model performance with increasing degraded data (and in both sampling designs, EqualPrev and TruePrev). However, we showed that the different evaluation approaches are complementary, since predictions with good (i.e. high values) model fit (MFp) usually presented a bad (i.e. or low values) predictive success (PSb). Additionally, the same pattern is reflected in the different modeling techniques (i.e. techniques with good MFp had poor PSb and vice-versa). This reflects the classical trade-off between model (over-) fitting and model predictive performance, and is supported by previous works showing a decrease in evaluation values between model fit and independent evaluation (e.g. Randin *et al.*, 2006; transferability test, where General Additive Models -GAMs- fit better than GLMs but predict worse to independent data).

How do evaluation metrics reflect model performance?

A consistent pattern was identified, with models evaluated by Somers'D (rescaled AUC) always yielding the highest evaluation values, usually followed by MaxTSS and MaxKappa, or TSS and Kappa (for probabilistic and binary predictions respectively). Within the same model, Somers'D values had very small differences when compared with the control model (even with errors >30%). Somers'D (rescaled AUC; from -1 to 1) was used instead of the widely used AUC to allow direct comparisons to the other evaluation metrics, as they all range between -1 and +1, being interpreted roughly in a same way as correlation coefficients. This means that when considering Somers'D (or AUC, with even higher evaluation values, concentrated between 0.5 and 1), all models evaluated in this study would be considered at least fair (based on thresholds proposed by Swets (1988, note nr 11); i.e. models with AUC values above 0.7 are considered 'useful for some purpose', while models with AUC>0.9 are considered as being 'of rather high accuracy'). However, when evaluated by the other metrics, a large amount of these models would be considered poor or not different than random. Therefore, concluding whether a model is good, fair or poor partly depends on the evaluation metric used and not only on model performance. In particular, our results suggest a strong tendency of Somers'D (i.e. AUC) to yield overoptimistic evaluations. We also observed, although in a lesser measure, a tendency of TSS (resp. MaxTSS) to yield overoptimistic values, whereas Kappa (resp. MaxKappa) proved to better reflect the level of errors added to the training data. These results are supported by recent findings showing that AUC/TSS are not the most efficient metrics to assess model performance (being over-optimistic or unrealistic) and that these could be classified as having good performance even when "dummy" data (e.g. pseudo-predictors derived from paitings; Fourcade *et al.*, 2018) or wrong information (e.g. locational uncertainty; Graham *et al.*, 2008; Mitchell *et al.*, 2017) was used. As a result, many models could be considered as satisfactory despite generating partially wrong spatial predictions. Our results confirm these previous criticisms and show how important it is to take into account these drawbacks in future uses of AUC (or Somers'D) – and to a lesser extent of TSS - to assess model performance. Some suggested approaches might be to assess the spatial predictions when comparing models (Randin *et al.*, 2006; Mitchell *et al.*, 2017) or accounting for the most relevant section of the ROC curve (Peterson *et al.*, 2008; assuming that true absences and independent data exists). However, as noted by Fourcade *et al.* (2018), this "perfect" data is usually unavailable and detailed screening of ROC plots can be difficult when modelling multiple species. Therefore, the use of AUC needs to be considered with great care in future studies and the interpretation scales (Swets, 1988; Araujo *et al.*, 2005) used to assign a level of model performance to its values need to be revisited. We believe it is probably more effective and productive to investigate new ways/methods to correctly evaluate model performance and predictions, with the use of artificial data being a useful tool to completely assess the value of these new methods.

How do different modelling techniques deal with the degraded training data?

The contrasted results of predictions with high model fit (random forests) presenting low values of predictive success (i.e. higher with generalized linear models) and vice-versa, clearly show that some techniques (like RF/BRT) are good at finding a signal in the degraded training data (i.e. can fit complex responses; Merow *et al.*, 2014) and still deliver a good MFp (as seen in Figs 3-4). However, these techniques are not as good at predicting to independent data (in our case to the rest of the distribution, largely unaffected by errors). On the other hand, techniques like GLMs reflect better the errors in training data (though showing a drop in MFp), but are still fairly good (within a reasonable range of error added) at predicting the true distribution of the species across the whole study area

(PSb). Considering Predictive success binary (PSb) as the expected aim for any predictive model, it turns out that some modelling techniques (here GLM) are able to “compensate” for errors added to the training data (i.e. still fit a similar response curve, e.g. unimodal, with increase error bonds) while others are not (random forests; i.e. might fit totally different response curves, adapted to the modified training data). So, models with simpler response curves (like GLM) tend to better manage errors when present in the species data, resulting in better predictions to independent data (see e.g. Randin *et al.*, 2006 when compared to GAMs), and better fit to ecological theory (Austin, 2002; Austin, 2007). More complex methods (here RF/BRT) seem to over fit the degraded data, maintaining a good/high model fit (MFp) but at the cost of a poorer/low predictive success (PSb) (see also Merow *et al.*, 2014).

How do different types of errors affect models and metrics?

The creation of false positives (FP) had a stronger negative effect on model performance than when false negatives (FN) were created. This is especially true when species had the same number of presences-absences (“EqualPrev”), not being obvious when sampling true prevalence (“TruePrev”), possibly due to the characteristic low prevalence of some species. False positives had a stronger negative effect because presences are expected to be on average more informative. They generally occur in a unimodal and limited way along environmental gradients, contributing to a fairly clear signal that can be captured in a model. On the other hand, absences are usually less informative since they can span entire environmental gradients, and thus be found e.g. on both sides of the mode of a species’ occurrences (i.e. would need a bimodal response to be captured). Depending on the species, absences can still hold a signal in some cases (e.g. low elevations for alpine plants), but it is likely to be on average much weaker than that of presences. We can think of the creation of false negatives (in “EqualPrev”) the same way as one uses pseudo-absences (i.e. when real absences are not available), setting the weights of those pseudo-absences to 0.5 (therefore ensuring equal prevalence). As the addition of errors to presences/absences decreased model performance in both cases, it’s important to account for imperfect detection in models (see Lahoz-Monfort *et al.*, 2014; Guélat & Kéry, 2018 for recommendations).

Conclusions, recommendations and perspectives

Our study showed that much can be learned by using artificial data where truth is known, especially by contrasting model fit and predictive success, and modellers would gain much by using these virtual approaches more systematically in the future in complement to real data. Several important findings emerged specifically from this study:

- The effect of errors added to species data decreased with increasing sample size.

- Different conclusions about model performance can be inferred depending on how it is measured (i.e. which metrics and which data):
- Models with high evaluation values can be obtained even with high levels of error artificially added to the training data;
- The classical classification or interpretation of a model as excellent, good, fair or poor strongly depends on the metric used to evaluate it, and thus can be misleading;
- Evaluation metrics matter: we identified AUC as a particularly overoptimistic metric and in a lesser measure MaxTSS (TSS for binary predictions), with often high evaluation values produced even with high levels of errors in the training data, thus not necessarily translating a good predictive success; therefore, we recommend the use of MaxKappa.
- Modelling techniques were differently affected by added error, with some delivering better measures of predictive success (GLMs here) and others delivering better model fit (RFs).
- The creation of false positives had a stronger effect on the measured evaluation approaches than the creation of false negatives.

A particularly important finding in our study is thus the need to seriously reconsider the current use of AUC (here rescaled, Somers' D) and its scale of interpretation. We advise caution when models are solely evaluated with this metric (and to a lesser extent by TSS and MaxTSS) as the interpretation of their quality, reliability and transferability might be too optimistic and lead to biased conclusions. The incorrect interpretation of how good/accurate a model is might have serious consequences if not considered. For example, the prioritization of specific areas for conservation can be wrong if the models used for that prioritization are overoptimistic or biased. The same can be said if invasive species prevention/eradication efforts are occurring, with an overoptimistic prediction possibly leading to management being directed to areas where those efforts are unnecessary. Taking into account previous and current studies, the most appropriate measure might be to completely cease to use AUC and instead focus on more effective evaluation metrics. Based on our results, we recommend using MaxKappa (resp. Kappa) if one wants a metric that better reflects the actual level of errors in the predictions. As it is usually preferable to evaluate models using spatially independent data (James *et al.*, 2013; Guisan *et al.*, 2017), our results suggest that techniques that are better at reproducing ecological theory (Austin *et al.*, 2006), like GLMs here, tend to show a better overall behaviour for modelling species distributions. However, additional modelling techniques (e.g. as found in Elith *et al.*, 2006) should also be tested to determine the most suitable ones. Additionally, effort should be put in minimizing false positive rates when collecting training data (e.g. improving species identification or detectability). Finally, research using a virtual ecologist approach could also be employed to further develop more reliable evaluation metrics that could be properly tested in a "controlled environment".

In a general manner, a more systematic use of artificial data bears the potential to improve methodological developments considerably in future ecological and evolutionary research.

ACKNOWLEDGMENTS

This work was supported by SNF project “SESAM'ALP - Challenges in simulating alpine species assemblages under global change”. Computations were performed at Vital-IT Center for high-performance computing of the Swiss Institute of Bioinformatics. We thank Damaris Zurell and Catherine Graham for their valuable comments on an early stage of this work.

Chapter 1.2

How much should one sample to accurately predict the distribution of species assemblages? A virtual community approach

This manuscript is published in *Ecological Informatics*

My contribution to the paper: I conceived the ideas and developed the methodological framework in coordination with the Antoine Guisan and Daniel Scherrer. I also developed the codes to create and run and analyse the data, leading the writing of the different versions of the manuscript.

How much should one sample to accurately predict the distribution of species assemblages? A virtual community approach

Rui F. Fernandes[a], Daniel Scherrer[a] and Antoine Guisan[a,b]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland
[b] Institute of Earth Surface Dynamics, University of Lausanne, Geopolis, 1015 Lausanne, Switzerland

ABSTRACT

Correlative species distribution models (SDMs) are widely used to predict species distributions and assemblages, with many fundamental and applied uses. Different factors were shown to affect SDM prediction accuracy. However, real data cannot give unambiguous answers on these issues, and for this reason, artificial data have been increasingly used in recent years. Here, we move one step further by assessing how different factors can affect the prediction accuracy of virtual assemblages obtained by stacking individual SDM predictions (stacked SDMs, S-SDM). We modelled 100 virtual species in a real study area, testing five different factors: sample size (200-800-3200), sampling method (nested, non-nested), sampling prevalence (25%, 50%, 75% and species true prevalence), modelling technique (GAM, GLM, BRT and RF) and thresholding method (ROC, MaxTSS, and MaxKappa). We showed that the accuracy of S-SDM predictions is mostly affected by modelling technique followed by sample size. Models fitted by GAM/GLM had a higher accuracy and lower variance than BRT/RF. Model accuracy increased with sample size and a sampling strategy reflecting the true prevalence of the species was most successful. However, even with sample sizes as high as >3000 sites, residual uncertainty remained in the predictions, potentially reflecting a bias introduced by creating and/or resampling the virtual species. Therefore, when evaluating the accuracy of predictions from S-SDMs fitted with real field data, one can hardly expect reaching perfect accuracy, and reasonably high values of similarity or predictive success can already be seen as valuable predictions. We recommend the use of a 'plot-like' sampling method (best approximation of the species' true prevalence) and not simply increasing the number of presences-absences of species. As presented here, virtual simulations might be used more systematically in future studies to inform about the best accuracy level that one could expect given the characteristics of the data and the methods used to fit and stack SDMs.

KEYWORDS: Virtual community ecologist; stacked species distribution models; nested design; factors importance; relative effects; sampling effect

INTRODUCTION

Important species co-existence questions have been raised in the field of community ecology over the past years (e.g. Gotzenberger *et al.*, 2012; Munkemuller *et al.*, 2014; Mittelbach & Schemske, 2015), with particular focus given to understanding what drives the distribution of assemblages (i.e. communities, sometimes used interchangeably here) and why and how their composition and richness can change in space and time. Additionally, with the increasing impacts caused by global changes (e.g. habitat fragmentation, biological invasions, climate and land-use change), it becomes critical to develop methods and tools that allow predicting the spatial distribution of species assemblages (D'Amen *et al.*, 2015b).

Species distribution models (SDMs; also called habitat suitability or ecological niche models; see Guisan et al., 2017), which statistically relate species observations, usually obtained through field observations or databases with environmental data (Guisan & Zimmermann, 2000), are useful tools in this regard as they can be stacked to predict the distribution and composition of species assemblages (e.g. Ferrier & Guisan, 2006; Dubuis *et al.*, 2011; D'Amen *et al.*, 2015b). When dealing with species richness (SR), the simplest and most common method consists in modelling the distribution of all individual species in a pool and then summing their predictions to obtain assemblages (stacked-SDM, S-SDM; Ferrier & Guisan, 2006; Dubuis *et al.*, 2011). However, this method has some limitations, such as over-predicting species richness per site (Guisan & Rahbek, 2011) or being sensitive to methodological biases (Calabrese *et al.*, 2014; Scherrer *et al.*, 2018a). Additionally, while single species models are useful, numerous factors (e.g. sample size, sampling prevalence, sampling design, modelling techniques, imperfect detection of species or the choice of environmental variables) can lead to an increase in the uncertainty of their predictions (e.g. Kadmon *et al.*, 2003; Barry & Elith, 2006; Guisan *et al.*, 2007b; Beale & Lennon, 2012), potentially propagating into the predictions of species assemblages. Until now the majority of studies used real species data to assess the effects of different factors on SDM performance at the individual species level. A recent study (Thibaud *et al.*, 2014) proposed the use of virtual or simulated data (Hirzel *et al.*, 2001; Zurell *et al.*, 2010) to assess how a set of factors affect the predictive performance of single species models. With the use of virtual data instead of real species, the “true” distribution of the species is completely known (Hirzel *et al.*, 2001) as well as the predictors that influence that distribution. Contrary, when using real species data, biological assembly rules or dispersal limitations might prevent species from coexisting even when adequate conditions exist. Other sources of uncertainty might also occur (e.g. missing environmental variables or stochasticity), making real data more difficult to use to test the relative importance of various factors. Using virtual species, whose distributions are solely determined by a set of environmental factors, ensures that the suitability of all species in each site is strictly determined by those factors with no additional biotic (e.g. competition) or dispersal restrictions. By simulating virtual sampling of these distributions with various effects (see above) and then refitting the models, one can compare the initial “true distribution” with the predicted distributions with and without ‘effects’ and in this way determine which factors affect models the most (Zurell *et al.*, 2010). This is still rarely done at the assemblage level, likely because there is yet no unanimous method on how to correctly predict species assemblages, or because the data and computational requirements to predict assemblages remained relatively intensive.

Here, we aimed at filling this gap by assessing how different factors affect the prediction accuracy of virtual species assemblages (obtained through S-SDMs). Specifically, we wanted to analyse the effects of five different factors - sample size, sampling method, sampling prevalence (i.e. the proportion of samples in which one found the species; not to be confused with species prevalence, the number of places occupied by a species out of the total number of places available), thresholding method and modelling technique -both separately and nested within each other- to determine: (i) what overall accuracy can be expected when sampling the known distributions in various ways (i.e. nested/non-nested, different prevalences), and (ii) which factors most affect S-SDMs performance. As a direct corollary, this should allow us to estimate the best achievable accuracy in a given modelling context where multiple factors affect the models, an aspect rarely if ever assessed in S-SDM studies (assemblage models).

METHODS

Study area

To apply our approach, we used a real landscape located in the Western Alps of Switzerland (http://rechalp.unil.ch), covering an area of approximately 700 km^2. Our analyses were conducted on 762133 sites corresponding to open, non-woody vegetation (i.e. grassland, meadow, rock, and scree). This is an intensively sampled region where many high locational accuracy biological data and high-resolution environmental data is available, providing a realistic set of species observations and predictors at very high resolution (25 meters).

Analytical framework

We followed six main steps of a virtual species simulation framework (see Figure 1), to assess how multiple factors affect the prediction accuracy of species assemblages, obtained with binary stacked species distribution models (bS-SDM). To ensure ecological realism, we defined the distributions of our virtual species based on predictions of models fitted on real data in the same study area (see section 2.2.1 below; as done in Thibaud *et al*., 2014).

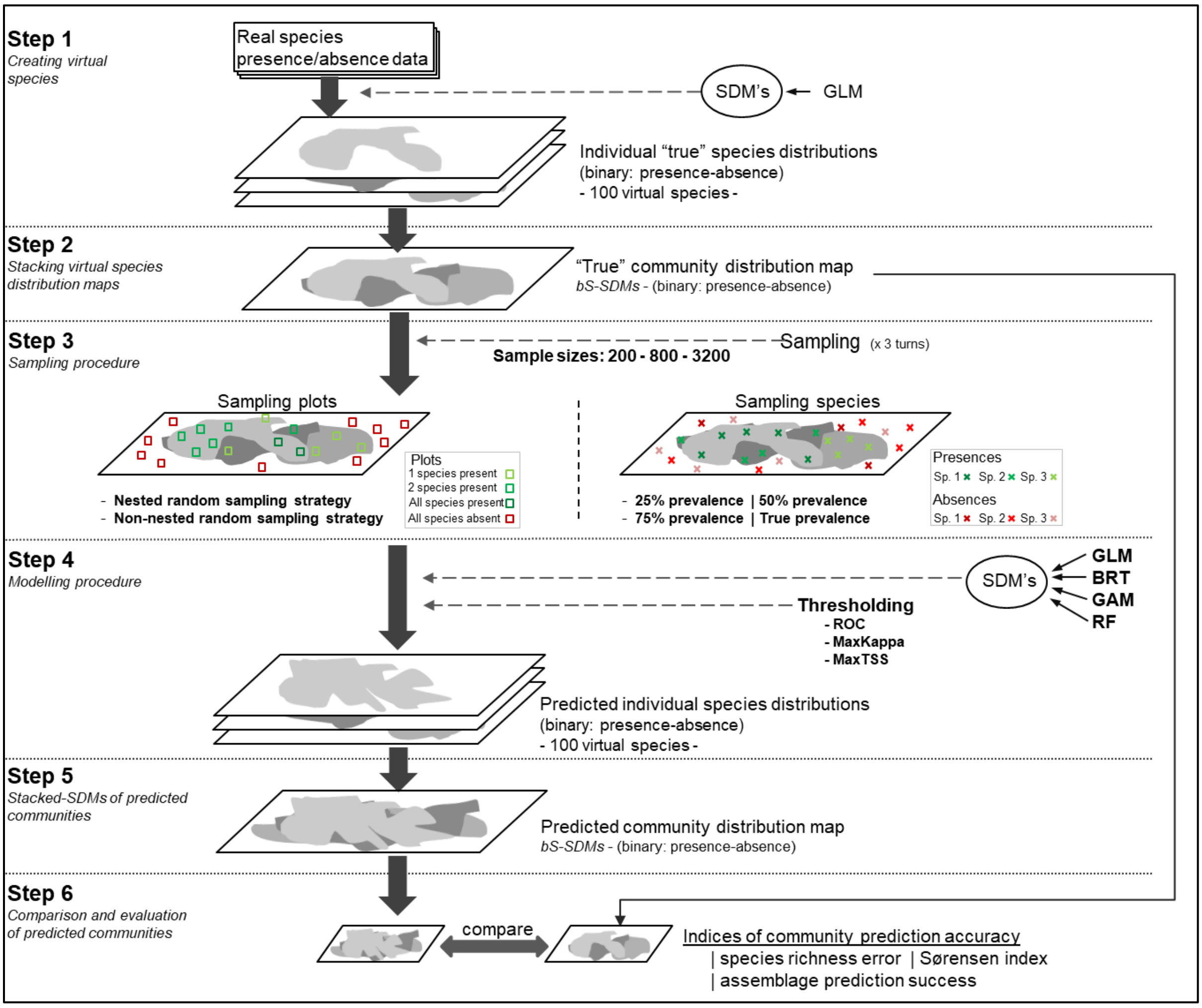


FIGURE 1 Framework of the analytical steps followed in the study (in bold the tested effects: sample size, sampling prevalence, modelling technique and thresholding method). *Step 1 – creating virtual species*: from a set of real species presence-absence data, 100 virtual species were created by fitting GLMs. *Step 2 – stacking virtual species distribution maps*: the created distribution maps were stacked together, and the resulting map was considered as the "true" assemblage distribution map. *Step 3 – sampling procedure*: the occurrence of all the virtual species present in the true assemblages were sampled using different sample sizes (200, 800 and 3200) and two different occurrence sampling methods: i) sampling plots (random nested or non-nested sampling); ii) sampling species (four different prevalence classes: 25%; 50%; 75% and a value based on the true prevalence of each species). *Step 4 – modelling procedure*: all single species models (using the occurrence data sampled in the previous step) were fitted using generalized linear models (GLM), generalized additive models (GAM), boosted regression trees (BRT) or random forests (RF). The projected species distributions were converted into binary presence/absence data using three threshold methods: ROC, MaxKappa and MaxTSS. *Step 5 – stacked-SDMs of predicted assemblages*: the binary projections created in the previous step were stacked together to create a predicted assemblage distribution map for each sample size, sampling strategy, modelling technique and thresholding method. *Step 6 – comparison and evaluation of predicted assemblages*: all the outputs resulting from the different predicted assemblages were compared with the observed true assemblage and several indices of accuracy of assemblage predictions were calculated.

Creation of virtual species

We generated presence-absence maps for 100 virtual species from a set of real species presence-absence data previously sampled in the study area. The species were randomly chosen from a set containing more than 1088 species (Dubuis *et al.*, 2011; Pottier *et al.*, 2013), but only selecting from those with more than 30 occurrences (around 627; see Appendix A for more information on the original dataset). We did this by fitting generalized linear models (GLMs; McCullagh & Nelder, 1989) to the species data as a function of five environmental predictors: summer mean monthly temperature (2 to 19° C), sum of winter precipitation (65 to 282 mm), annual sum of potential solar radiation (KJ), slope (°), and topographic position (unit-less, indicating ridges and valleys). We then used these models – in binary form – to predict the 100 species distributions across the study area. The binary presence-absence information was obtained by transforming the predicted probabilities using the threshold corresponding to the point on the receiver operating characteristic (ROC; Hanley & McNeil, 1982; Swets, 1988) plot - sensitivity against 1-specificity across successive thresholds - with the shortest distance to the top-left corner of the plot (Cantor *et al.*, 1999). We considered these binary predictions as the "true distributions". All the initial environmental and species data used in this study were available at 25 m resolution. Diverging results have been observed when different resolutions or extents were used (e.g. Thuiller *et al.*, 2004; Guisan *et al.*, 2007a), demonstrating that scale parameters can have an important influence on model predictions. This can be a result of the scale dependency of the environmental predictors (Vicente *et al.*, 2014) or the spatial stochastic effects at smaller spatial scales (Steinmann *et al.*, 2011; Scherrer *et al.*, 2018b).

Furthermore, dispersal and biotic factors can also play an important role interacting with scale (Soberon & Nakamura, 2009), with the distribution of real species not being fully explained by abiotic predictors alone. We avoid all these issues in our work by using virtual species and the same predictors to create the species and fit their distribution models, ensuring that the initial species distributions are fully explained by the chosen predictors at the study scale (extent and resolution, and using the same technique). With this approach, we tended to guarantee that the virtual species presented realistic response curves for our landscape, resulting in realistic species assemblages. The resolution should thus not matter in our study and should not affect our findings. However, we acknowledge that in real ecosystems the explanatory power of abiotic environmental factors (as used in this study) on single species distributions and assemblages (i.e. assembly rules) might strongly depend on the spatial resolution of the study, and other factors such as dispersal limitations and biotic interactions might modify the abiotic responses and interact with scale. All models were run in the R software version 3.3.3 (R Core Team, 2017), using biomod2 (Thuiller *et al.*, 2009) default settings for sake of simplicity and comparability.

Stacking virtual species distribution maps

The binary predictions for each species were then stacked to create species assemblages (i.e. binary stacked SDMs, bS-SDMs), providing both species richness and composition for each pixel in the study area. In this simplified theoretic approach, the assemblages resulting from this stacking of binary SDMs are then considered our “true” assemblages (i.e. S-SDM), meaning that all species stacked into these “true” assemblages can coexist based on the abiotic factors only, without biotic or dispersal restrictions further excluding some of them from an assemblage. This is the power and appeal of simulations, since one can restrict the niche to known factors only (in our case abiotic environment), and this way facilitate the assessment of factors affecting S-SDMs through the sampling of virtual observation sites.

Sampling procedure

Presences and absences were sampled for all species using increasing sample sizes (n=200, 800 and 3200) and according to two schemes representing the dominating types of data available to fit SDMs: 1) simulating the sampling of “vegetation plots”, to reproduce real datasets obtained in field surveys where all species were sampled in the same plots (i.e. in a ‘plot-like’ fashion), using a nested (i.e. plots sampled in the smaller sample sizes are included in the larger sample sizes) or a non-nested random sampling strategy; here, the species prevalence cannot be controlled: and 2) simulating occurrence data as typically available in biodiversity databases where species are sampled individually from each other, but here with absences also available. In this case, we used four different sampling prevalence values: 25%; 50%; 75% and the true prevalence of each species. The complete sampling procedure was repeated three times for each of the 100 virtual/simulated species. This case can also be considered as a simulation of a situation where presence-only data is available, with pseudo-absences weighted to 25, 50 or 75%.

Modelling procedure

To test the effects of different modelling techniques, single species models were fitted with four techniques (see Guisan *et al.*, 2017 for an overview): generalized linear models (GLM; McCullagh & Nelder, 1989) , generalized additive models (GAMs; Hastie & Tibshirani, 1990), boosted regression trees (BRTs; Friedman *et al.*, 2000) and random forests (RFs; Breiman, 2001), all commonly used in SDMs (e.g. Guisan *et al.*, 2002; Prasad *et al.*, 2006; Elith *et al.*, 2008). We used the same five environmental variables to calibrate the models as previously used to create the virtual species, all at a 25-meter resolution: summer mean monthly temperature, the sum of winter precipitation, the annual sum of potential solar radiation, slope, and topographic position. Each individual model was calibrated using 80% of the available data and evaluated on the remaining 20%. This cross-validation procedure was repeated 20 times and averaged using an ensemble approach (i.e. mean probabilities across predictions). The models were evaluated on the evaluation dataset using ROC (receiver operating characteristic; Hanley & McNeil, 1982; Swets, 1988), MaxKappa (Guisan *et al.*, 1998; Huntley *et al.*,

2004) and MaxTSS (i.e. equivalent to the sensitivity-specificity sum maximization described in Liu *et al.*, 2005) (see Guisan *et al.*, 2017 for details on maximization approaches). Finally, the projected species distributions were converted into binary presence/absence using the same approach as described in section 2.2.1 (ROC plot). In parallel, we also used two other thresholding methods to transform probability distributions into presence-absence data: selecting the thresholds maximizing Kappa (MaxKappa) or maximizing TSS (MaxTSS) as presented above. The whole approach was implemented in version 3.3.3 of the open-source software R (R Core Team, 2017).

Stacked-SDMs of predicted communities

Species binary predictions were stacked together to predict assemblages for each sample size, sampling method, threshold method, and modelling technique. With these predicted assemblage maps, we simultaneously obtained information on species richness and composition for each modelled site across the whole study area. We also made three repetitions of the sampling procedure (i.e. turns; T1, T2 and T3) of different sample sizes (200, 800 and 3200), different sampling prevalences (nested or non-nested sampling of random plots and four different sampling prevalence types), three thresholding methods (ROC, MaxKappa and MaxTSS) and four different modelling techniques (GLM, GAM, BRT and RF). Consequently, we ended up with a final set of more than 64 000 models.

Comparison and evaluation of predicted communities

Composition outputs and species richness resulting from the differently predicted assemblages (for each site) obtained through the previous steps were compared and evaluated to our observed assemblages (i.e. “true” assemblage map). We calculated three main indices of assemblage prediction accuracy by using the *ecospat.SSDMeval* function available in the “*ecospat*” R package (see Table A.1 for details on all the indices; Di Cola *et al.*, 2017): (i) species richness error (i.e. difference between predicted and observed species richness); (ii) the assemblage prediction success (i.e. proportion of species correctly predicted as present or absent); and (iii) a widely used metric of assemblages similarity, the Sørensen index (Sørensen, 1948). We calculated six additional indices to complement our analyses: (iv) community TSS (here measured for a site across all species, rather than for a species across all sites as in single SDM evaluation; Pottier *et al.*, 2013) and (v) community Kappa (same as for previous metric, for a site across species; Pottier *et al.*, 2013), (vi) over-prediction; (vii) under-prediction; (viii) sensitivity (i.e. the proportion of species correctly predicted as present); and (ix) specificity (i.e. the proportion of species correctly predicted as absent). Finally, some of those indices were used to assess the importance of the different studied factors following a procedure similar to the one proposed by Thibaud *et al.* (2014). However, contrary to the latter study, here we were not only interested to measure model accuracy for individual species models, but mainly to assess the predictive accuracy of species assemblages. To assess the importance of the studied factors, we analysed the variation of the previously mentioned indices via a linear mixed-effects model, adapting codes from Thibaud *et al.* (2014). To examine the relative importance of factors in the linear models, we calculated the marginal and conditional coefficients of determination (R2; Nakagawa & Schielzeth, 2013). The R package *nlme* (Pinheiro *et al.*, 2017) was used to fit these linear mixed-effects models, with each factor and all its interactions being excluded and compared using marginal R^2 (i.e. calculating the proportion of variance that is explained by fixed effects compared to that of the full model). By excluding one factor at a time, we can measure its contribution to the full model, and therefore its contribution to improve the predictive accuracy of SDMs (i.e. the lower the value of R^2 when compared with the full model, the greater the effect of the excluded factor).

RESULTS

The SDMs used to create the virtual species had scores ranging between 0.626 and 0.967 when evaluated by ROC (mean = 0.86 ± SD = 0.13) and between 0.1/0.73 and 0.2/0.9 when evaluated respectively by MaxKappa (0.35 ± 0.14) and MaxTSS (0.6 ± 0.14). The prevalence values for the virtual species' presence-absence distributions ranged between 0.02 and 0.74 (0.35 ± 0.17). The virtual species SDMs (i.e. fitted using virtual species sampled data) had very high evaluation scores (ROC: 0.999 ± 0.002; MaxKappa: 0.99 ± 0.02; MaxTSS: 0.99 ± 0.02).

Results from the SDMs based on a random sampling of 100 000 plots from the virtual species distribution maps (Fig. 2, MaxKappa; see Appendix A for the other thresholding methods results) revealed that modelling technique and sample size were the factors with the largest effect on prediction accuracy of our assemblages when all the factors are taking into account in a nested manner (Fig. 2 and 3). This can also be observed in the values of marginal and conditional R^2, calculated through a linear mixed-effects model to quantify our visual impressions from the previous mentioned figures (Table 1; i.e. the partial models when modelling technique or sample size are excluded have the lowest values of R^2 when compared with the full model, indicating the important effects of these factors). Independently of the calculated indices and taking into account the reduction in marginal R^2, modelling technique is more important than sample size (e.g. when Sørensen was used, marginal R^2 gets reduced from 0.913 in the full model to 0.366 when modelling technique is excluded and to 0.526 when sample size is excluded). However, the effects of the different factors can also be important when analysed separately. Models fitted with GAM and GLM provided the most accurate predictions on average (i.e. highest similarity of observed/predicted assemblages – Sørensen above 0.95 – and prediction success, also always above 0.95 on average; the values for the two techniques are similar, followed by BRT and RF (on average below 0.95 both for Sørensen and prediction success; Fig.2b, c). It's also noticeable that models fitted by BRT or RF presented higher variance for the different calculated metrics. Within each modelling technique, higher sample sizes decreased the difference between predicted and observed SR (i.e. species richness error; Fig. 2a). Higher sample sizes also increased prediction success (Fig. 2b) and assemblage similarity (i.e. Sørensen index; Fig. 2c). Additionally, both over- and under-prediction decreased with increasing sample size (Fig. 3a, b), while sensitivity (Fig. 3c) and specificity (Fig. 3d) increased with increasing sample size.

The level of sampling prevalence appeared to individually (i.e. taking into account one factor at a time) influence the accuracy of predicted assemblage models (Fig. 2 and 3), not being important when all factors are taking into account (see Table 1). Models calibrated with higher levels of sampling prevalence (i.e. 75%) presented higher species richness error (Fig. 2a), with a higher number of species predicted than those observed (Fig. 3a, b; over-prediction larger than under-prediction; plus 3 species on average) and a higher sensitivity than specificity (Fig. 3c, d). On the other hand, levels of sampling prevalence of 25% presented the inverse pattern, with a lower species richness error (Fig.2a) and a lower number of species predicted than observed (Fig. 3a and b; under-prediction larger than over-prediction; minus 1 species on average). When considering over- and under-prediction as well as sensitivity and specificity, similar patterns were observed with the true prevalence sampling method or when using the “plot-like” sampling methods (i.e. nested and non-nested random sampling). These three methods presented very similar results and more accurate predictions (i.e. higher values of predictions success and the same number of species predicted as observed, on average). Models calibrated with high sampling prevalence (75%) also appeared to have lower values of assemblage predictive success (Fig. 2b; 0.96 on average) and assemblage similarity (Sørensen, Fig. 2c, 0.94 on average), yet with relatively small differences when compared with other sampling prevalences (Sørensen around 0.95, depending on sample size and modelling technique). Furthermore, based on the calculated R^2 (Table 1), the nested importance, considering the other factors at once, of sampling prevalence was negligible (i.e. for each calculated index, the R^2 actually increased when compared with the full model). However, it could still be important when analysing individual cases (e.g. when species have values of sampling prevalence above 75%). The method employed to transform our probability distributions into presence-absence data (i.e. thresholding method) was also not an important factor influencing assemblages’ prediction success, with the same patterns being observed in the three tested methods (see Appendix A) and with calculated R^2 not suffering any reduction when this factor was removed (i.e. indicating that it was not important in the overall model; see Table 1).

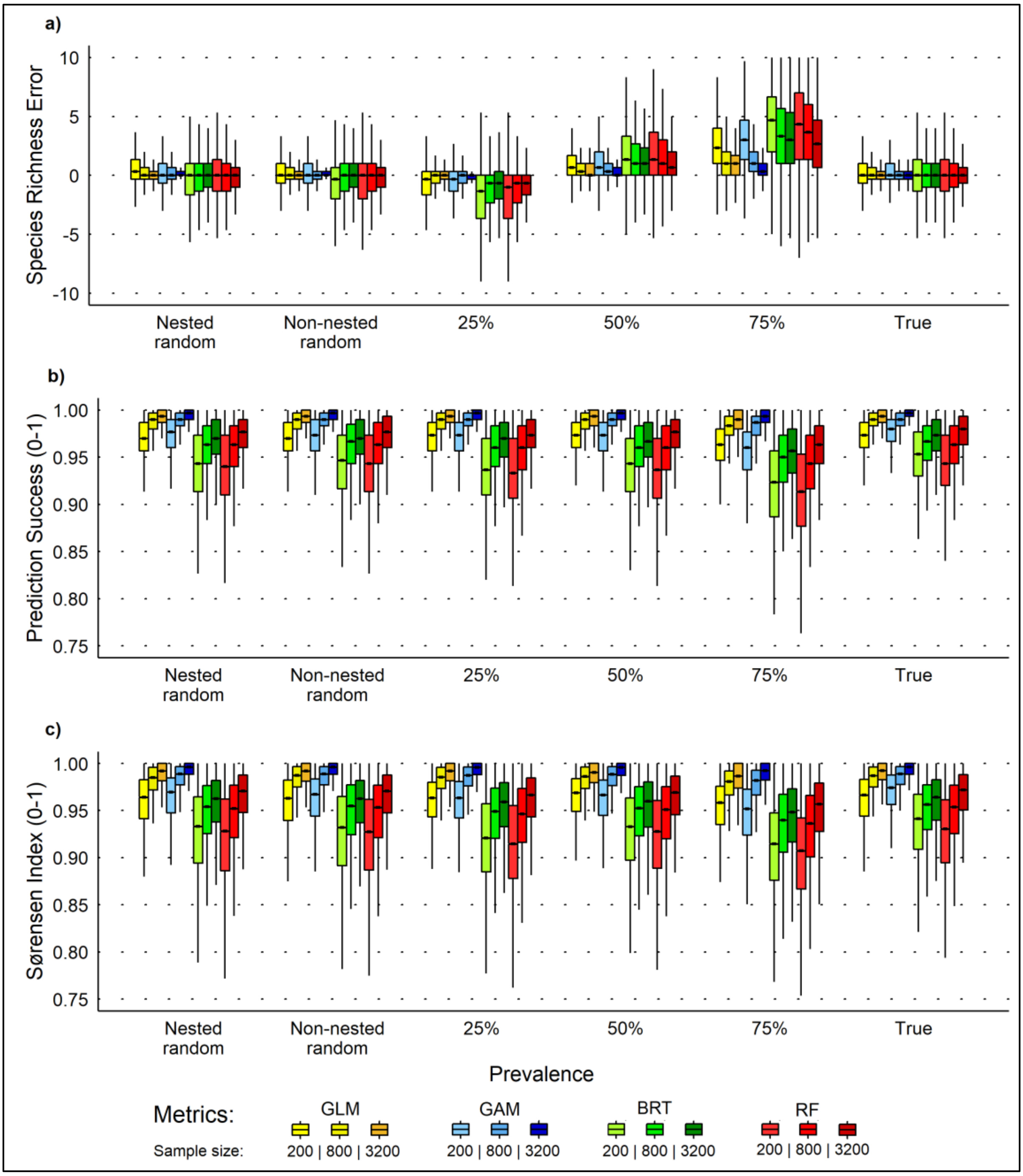


FIGURE 2 Boxplots of different indices of assemblage prediction (S-SDM) accuracy (i.e. species richness error, prediction success and Sørensen) for all the simulated species and for all the sampling strategies (based on plots (nested or not) or prevalence (25%, 50%, 75% or true) sampling; in abscissa). Each box shows the variation across all virtual species in a random subset of the study area (100 000 plots) for the binary predictions obtained using MaxKappa as thresholding technique, averaged from the three sampling turns. For each prevalence sampling, four sets of three boxplots are displayed, corresponding to models fitted using either GLMs (yellow), GAMs (blue), BRTs (green) or RFs (red), with increasing values of sample size (200, 800 and 3200).

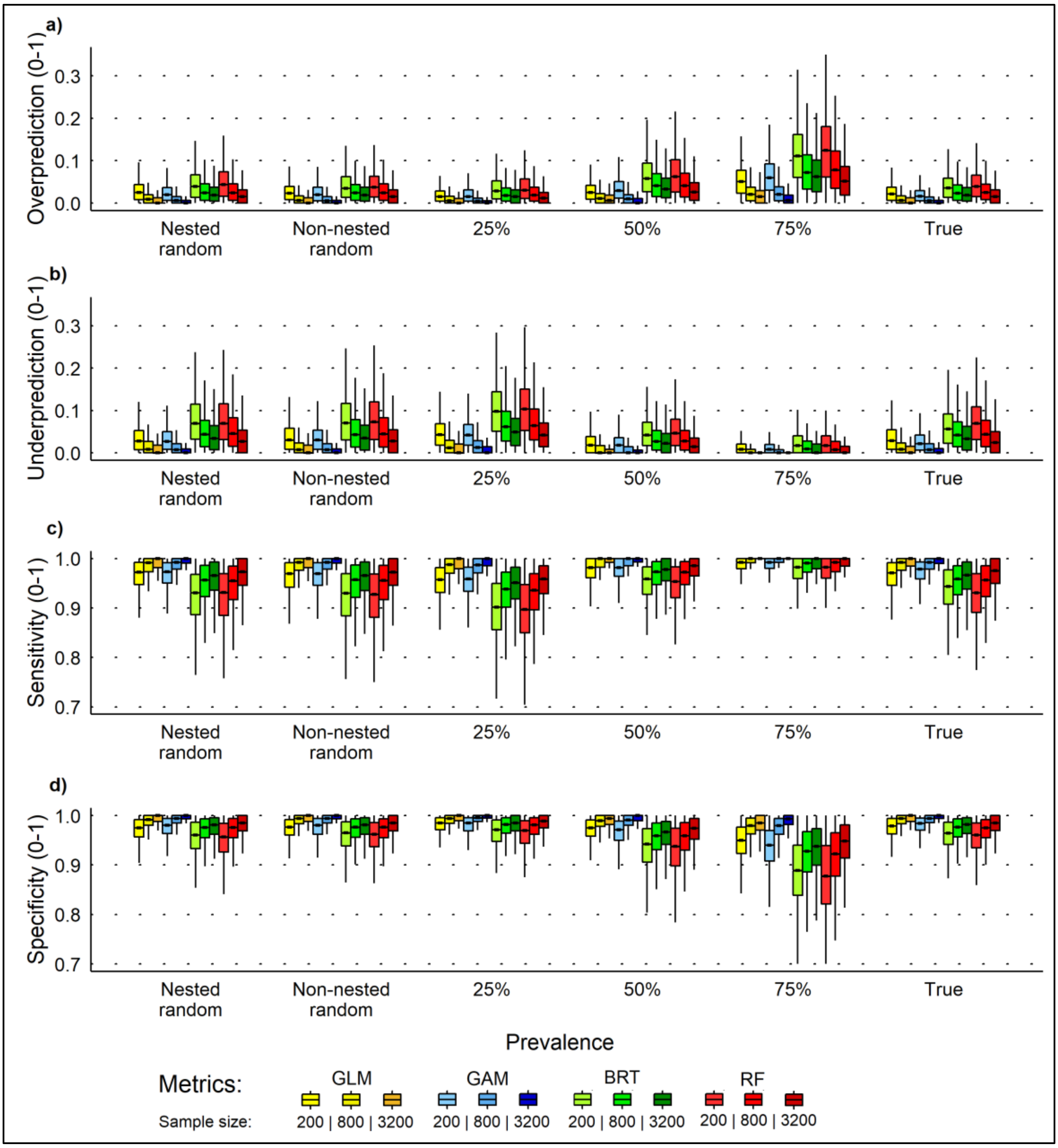


FIGURE 3 Boxplots of different indices of assemblage prediction (S-SDM) accuracy (i.e. over and underprediction, sensitivity and specificity) for all the simulated species and for all the sampling strategies (based on plots (nested or not) or prevalence (25%, 50%, 75% or true) sampling; in abscissa). Each box shows the variation across all virtual species in a random subset of the study area (100 000 plots) for the binary predictions obtained using MaxKappa as thresholding technique, averaged from the three sampling turns. For each prevalence sampling, four sets of three boxplots are displayed, corresponding to models fitted using either GLMs (yellow), GAMs (blue), BRTs (green) or RFs (red), with increasing values of sample size (200, 800 and 3200).

TABLE 1. Marginal and conditional R^2 for the five calculated indices used to evaluate our assemblage prediction (S-SDM) accuracy. Marginal and conditional R^2 are reported for each index, containing information for the full model (all factors together) and the submodels for each factor (and all its interactions with the other factors) excluded at a time. The lower the R^2 value when compared with the full model, the greater the effect of the excluded factor.

Calculated indices	Model	Marginal R^2	Conditional R^2
Sørensen	Full model	0.913	0.985
	Prevalence	0.918	0.985
	Sampling procedure	0.915	0.984
	Modelling technique	**0.366**	**0.388**
	Sample size	**0.526**	0.974
	Thresholding method	0.914	0.986
Predictive Success	Full model	0.895	0.982
	Prevalence	0.901	0.982
	Sampling procedure	0.898	0.981
	Modelling technique	**0.332**	**0.368**
	Sample size	**0.540**	0.967
	Thresholding method	0.895	0.983
Community TSS	Full model	0.942	0.988
	Prevalence	0.943	0.987
	Sampling procedure	0.943	0.987
	Modelling technique	**0.347**	**0.347**
	Sample size	**0.567**	0.973
	Thresholding method	0.943	0.989
Community Kappa	Full model	0.880	0.977
	Prevalence	0.887	0.978
	Sampling procedure	0.884	0.977
	Modelling technique	**0.324**	**0.371**
	Sample size	**0.532**	0.959
	Thresholding method	0.881	0.979

DISCUSSION

In this paper, we used a virtual ecologist approach to test the effects and importance of five factors – modelling techniques, sample size, sampling method, sampling prevalence and prediction thresholding - on the predictive accuracy of stacked binary predictions (bS-SDM) of species distribution models (SDMs). Our framework takes inspiration on the methodology first proposed by Thibaud *et al.* (2014), following a nested approach to test the relative effects of various factors on SDMs, but differing markedly from the latter by focusing here mainly on species assemblage models (S-SDMs). Furthermore, we used a much greater number of virtual species (n=100) and additionally assessed the importance of effects like sampling nestedness and prevalence on the accuracy of assemblage predictions. However, unlike Thibaud *et al.* (2014), we did not assess the effects of sampling bias and spatial autocorrelation. We took this decision because the previously mentioned paper reported that sample size and modelling technique were the factors that contributed most to the variation in prediction accuracy, while sampling bias and spatial autocorrelation had smaller and negligible effects respectively. To our knowledge, the present study is the first to use a virtual ecology framework to assess and report cumulative effects of different factors on species assemblage predictions (S-SDMs). We found that modelling technique and sample size were the most important factors, relative to the others tested (i.e. taking into account the importance of all factors at once), affecting the accuracy of SDMs and - most importantly here - of their assemblage predictions. Additionally, we found that the overall accuracy that can be expected from the S-SDMs depended on the options made when fitting models (e.g. the choice of technique, sample size, how to sample), with inaccurate assemblage predictions being obtained after simulated sampling even when the initial distribution of the species and the environmental factors determining them are completely known.

Which methodological factors most affect the performance of S-SDM?

Modelling technique

The importance of modelling technique and the fact that different algorithms can provide different predictions and predictive performances is something widely reported (e.g. Guisan *et al.*, 2007b; Graham *et al.*, 2008; Elith & Graham, 2009; Marmion *et al.*, 2009a), but the nested framework used here allowed to further discuss these differences in the light of (i.e. relative to) other factors: sample size, sampling design, thresholding criteria and sampling prevalence. Our findings partly confirmed the results obtained by Thibaud *et al.* (2014) for single SDMs, but here applied to species assemblage predictions (S-SDMs). However, we observed that modelling technique had a larger impact than sample size in affecting assemblage prediction accuracy (see Table 1), contrary to the previously mentioned study for single SDMs. Also, while the main results reported here derived from virtual species created using a GLM, the models obtained through this technique were not better than models fitted by GAM. This was contrary to the pattern observed in the aforementioned study, where GLMs clearly presented the best results when the virtual species were also created using that technique. Additionally, models fitted here both by BRT and RF presented results with higher variance and lower predictive success. This is contrary to some SDM studies (e.g. Elith *et al.*, 2006; Guisan *et al.*, 2007a; Graham *et al.*, 2008; Williams *et al.*, 2009), which found that these techniques performed better than GLM or GAM, while other studies showed no major difference in the performance of models fitted by the different techniques (e.g. Elith & Graham, 2009; Roura-Pascual *et al.*, 2009). These differences might result from the fact that in our study the virtual species' distributions were created using a regression model (GLM) and were then resampled to fit models with various techniques, whereas in other studies the same technique was always used to both create a virtual species and then fit the models and assess the effect of the different factors on it. BRT and RF would thus be good at finding a signal in the training data, but less good at predicting to independent data (i.e., in this case, the random subset of the study area -100 000 plots - used to calculate the different indices). Another reason for the discrepancy among studies might be that, in our study, the virtual species were fully explained by the predictors used and, since they were created by GLMs, showed clearly defined unimodal response curves. However, these response curves can be much more complex when real species are used, especially due to interactions between species and the environment and among themselves. More complex techniques like GAM, RF or BRT would fit these more complex curves better, but at the cost of then predicting worse to independent data. This had been shown already for single SDMs (e.g. Randin *et al.*, 2006) and discussed elsewhere (e.g. Merow *et al.*, 2014), but had never been shown so far for S-SDMs.

Sample size

The well-known impact of sample size on single SDMs (e.g. Stockwell & Peterson, 2002; Wisz *et al.*, 2008; Mitchell *et al.*, 2017) is in large part explained by the fact that a greater number of presence-absence data provides a larger amount of information about the occupied multi-dimensional environmental space, allowing to fit more reliable species response curves along all the considered environmental gradients, improving the species' niche quantification and associated predictions. However, we showed here that even when using a large amount of sampled data (>3000 sites) we can still obtain some inaccuracy in assemblage predictions, depending on the modelling and sampling technique used. However, if these other factors – modelling and sampling - are taken into account, one can achieve relatively high or very acceptable values of prediction success and assemblage similarity (Sørensen index) even with the smallest sample size assessed here (200 sites).

Sampling prevalence

Species prevalence is often a key factor affecting model performance (e.g. Manel *et al.*, 2001; Jiménez-Valverde *et al.*, 2009; Santika, 2011; Lawson *et al.*, 2014). Here, we showed that sampled prevalence also has an effect in some components of the assemblage evaluation, with higher sampling prevalence (here 75%) causing species richness over-prediction and favouring sensitivity, whereas lower sampling prevalence (here 25%) causes species richness under-prediction and favors specificity. Yet, it did not affect greatly our assemblage predictions (prediction success and Sørensen index) in our nested analysis taking into account simultaneously for the other methodological factors (see Table 1). However, if we consider the patterns observed in the different sample prevalence groups, assemblage prediction accuracy increased slightly (i.e. close to 1 at large sample sizes, around 0.95 at smaller sizes) when sampled prevalence reflected the true prevalence of the species (i.e. species prevalence) in the study area. The same occurred when low sampling prevalences were used or when the sampling was done in a 'plot-like' fashion – nested or non-nested random sampling – (Fig. 2 and 3). We observed high prediction success when using the species true prevalence mainly because, when using those values, the information given to the SDMs (presence-absence) allows unbiased estimates of species richness (Calabrese *et al.*, 2014). It can thus be expected that, if one obtains individual SDMs reflecting true prevalence, one should also get more accurate S-SDM predictions of species richness.

By sampling prevalences of 25, 50 or 75%, one wrongly defines the initial level of SR in the model. As the virtual species' true prevalence was between 25 and 50% (35% on average), this then explains why we under-predict models when sampling at the 25% prevalence level and over-predict at the 50% and 75% prevalence levels.

Previous studies like the one of Jiménez-Valverde *et al.* (2009) on single SDMs further showed that species prevalence strongly interacts with sample size, with the distribution of species being over-predicted when sample sizes are small and species prevalences high. These patterns were also observed in our data but at the assemblage level (Fig. 2 and 3). The authors of the previously mentioned study additionally showed that when the sampled presence-absences cover the entire environmental gradient, high or low species prevalences have less or no effect on model accuracy. In our study, this representative sample of the entire environmental gradient was best reflected by the true prevalence sampling method (if perfect conditions were possible), which simulated the most correct distribution of the species (considering also that the sites were randomly sampled), reducing the probability of sampling all the presences (or absences) in a reduced part of the environmental gradients. Accordingly, one should obtain minimal or no error in predicted species richness, as we observed. However, in real-world conditions where sampling the true species prevalence is impossible, the most appropriate method appears to be to randomly (or random-stratified) sample in a 'plot-like' fashion (i.e. similar results to true species prevalence; Fig.2 and 3), an approach used in many studies with real data (e.g. Dubuis *et al.*, 2011; Pottier *et al.*, 2013; D'Amen *et al.*, 2015a).

Sampling strategies

We also observed that the different sampling strategies (i.e. between sampling information in a 'plot-like' fashion - inventorying all species in each plot as sampling unit - and sampling species individually and independently of each other) is not one of the most important factors affecting assemblage predictions (i.e. considering all factors at once - i.e. relative effects; Table 1). Nevertheless, when analysing specific cases (like the individual sampling prevalence groups), we can say that sample prevalence is important to take into account, thus sampling plots is preferable to single occurrences as it is the best approximation to true species prevalence sampling (see above).

Thresholding methods

While the effects of different thresholding methods on species distribution predictions were widely studied for single SDMs (see e.g. Liu *et al.*, 2005; Lobo *et al.*, 2008; Lawson *et al.*, 2014; Vale *et al.*, 2014), our results for S-SDMs showed that this factor had negligible importance on the prediction success of our assemblages, with different methods presenting the same patterns (see Appendix A). This was surprising because, as showed by Nenzén and Araújo (2011), the choice of the threshold explained 25% of the variability in their results (with modelling technique explaining 35% and their interaction 19%). The fact that the thresholding method had no effect on the predictive success of our modelled species and assemblages might be associated with the values of the selected thresholds. In our case, these were always below 0.5 (and often <0.4), independent of the thresholding method (thus not showing a great variation in threshold values).

What overall accuracy can be expected when perfectly known species distributions are sampled?

We showed clearly here that, even with complete initial knowledge of the distribution of the species and assemblages and of the environmental factors determining those, fitting the models on samples of the data (of varying size) quickly brings some error in assemblage predictions, even with quite large samples (>3000 sites) and even if models with high evaluation values (e.g. ROC or MaxKappa > 0.9) were still obtained. This means that even if our individual species models present very high evaluation scores (i.e. on average close to 1 for the three metrics), we are unable to fully recover the initial assemblages, based on the stacking of all virtual species' distributions. This could be caused by the fact that even when one is able to obtain accurate individual species (i.e. SDMs), small errors (i.e. falsely predicted presences or absences) can occur in each of them. These errors can then accumulate and prevent us from getting accurate assemblage predictions (i.e. S-SDMs). If this is the case with virtual species, one can expect an equivalent or likely higher error accumulation with real species. Therefore, obtaining inaccurate community predictions using real species data might also occur due to methodological problems (e.g. Calabrese *et al.*, 2014) and not only because of missing dispersal or biotic constraints (e.g. Guisan & Rahbek, 2011). Similar tests could also be performed when using more mechanistic or process-based methods to determine if the same patterns are observed. Nevertheless, despite not being able to completely predict species assemblages, depending on the factors used to model these distributions (i.e. modelling technique, sample size or sampling method), one can still yield valuable accurate predictions (e.g. assemblage prediction success around or above 0.95). More particularly, we showed that one can obtain very good assemblage predictive success (i.e. predicted assemblages very similar to the observed ones across the whole area; Fig. 1) particularly when large sample sizes are available, when GLMs or GAMs are used (rather than more complex techniques like RF or BRT) to fit the models and when the sampling prevalence reflects either the true prevalence (possible with artificial data) or a 'plot-like' sampling of the species (realistic method that samples an approximation of the species true prevalence). Large sample sizes as the one used here (3200 plots) might be difficult to obtain for the majority of species and taxonomic groups, but we showed that even with smaller samples (200) we are still able to obtain very good assemblage predictions (e.g. prediction success and Sørensen similarity index above 0.95) when GLM or GAMs were used. So, more effort should be put into getting a representative sample of species distributions in a certain area using a 'plot-like' sampling method, and not simply increasing the number of presences or absences for some species. This will guarantee that the sampled data would potentially cover all the spatial and environmental gradients in a certain area while also reflecting the species true prevalence (or a very close approximation).

A critical assessment of our framework and how to go forward

Our framework, combined with the use of realistic virtual species can be used in future studies to test an even larger number of factors that might affect the accuracy of models, with the choice of those factors depending on the availability of computational power that is necessary in order to perform all the required steps. A more complex nested approach could include factors like the extent of the study area, different spatial scales, a larger set of available modelling techniques, the use of different environmental variables to create virtual species and fit the models (i.e. missing covariates), sampling bias (e.g. random, clustered, close to roads, stratified), spatial autocorrelation, the effect of using presence-absence data or only presences or the effects of different methods to create pseudo-absences.

We also recognize that the method used to create our virtual species might be considered simplistic and that other methods (threshold vs probabilistic approach; see Meynard & Kaplan, 2013) and different packages (e.g. SDMvspecies, NicheLim or virtualspecies - Duan *et al.*, 2015; Huang *et al.*, 2016; Leroy *et al.*, 2016) could be used to create the virtual species. This might be another opportunity to test if the same conclusions can be obtained when virtual species are created by different methods and further contribute to refine a specific methodological choice or approach. Even larger sample sizes might also be investigated to assess how large samples would need to be to reach near perfect predictions.

The majority of experimental studies are by essence oversimplifications of the real world, and alike our objective was to use a simplified and controlled artificial reality instead of simulations on real community data. Our goal was to assess the effects of a set of methodological factors on species assemblage modelling (S-SDMs when assembly rules are purely determined by abiotic constraints, thus ruling out effects from biotic interactions and dispersal limitations. While this is of course ecologically not fully realistic, it allowed us to test a scenario that was simple and with (nearly perfectly) known abiotic assembly rules. Considering that this study was performed to determine the influence of certain methodological factors on assemblage predictions, one can reasonably think that if the methods used were not good enough to correctly predict assemblages given such simplistic environmental drivers, then adding other factors (not considered here; e.g. dispersal limitations, biotic interactions, sampling bias or wrongly parameterized techniques) can only reduce the chances of obtaining accurate predictions.

We showed one illustration of how the use of virtual species can be helpful in spatial modelling of species assemblages, but it can also prove useful in numerous other situations in ecology. For SDMs, virtual species were already used in several instances (see Miller, 2014 for examples of recent applications) to validate proposed methods (e.g. Zurell *et al.*, 2016; Guisande *et al.*, 2017; Hattab *et al.*, 2017), but also to test the effects of observation errors on model performance and the efficiency of currently used evaluation metrics (Fernandes *et al.*, in press), to test different approaches to sample species data (Hirzel & Guisan, 2002), to test different downscaling methods (Bombi & D'Amen, 2012) or to assess the effectiveness of different hierarchical modelling frameworks when compared to more traditional methods (Fernandes *et al.*, unplub.). Other potential uses worth exploring might include the testing of methods or software used for spatial conservation planning (e.g. ConsNet, Zonation or Marxan - Ciarleglio *et al.*, 2009; Watts *et al.*, 2009; Moilanen *et al.*, 2014), to determine their real effectiveness in defining prioritization areas and identify strengths and weaknesses of different alternative approaches.

CONCLUSION AND MAIN MESSAGES

With this paper, we wished (i) to contribute to the ongoing discussion on the usefulness and validity of stacking species distribution models (SDMs) to predict species assemblages (S-SDMs), and (ii) to propose a way (taken from single SDMs; Thibaud *et al*., 2014) to analyse the relative effects and behaviour of different methodological factors potentially affecting S-SDM predictive success. We also discussed different features and potentialities that can help virtual species and simulations become a more useful tool in ecological or evolutionary research, e.g. to test the efficiency of alternative modelling frameworks in a fully controlled abiotic environment.

The main conclusions for factors affecting S-SDMs, based on our findings and given our study settings, are:

1) even when starting with the full knowledge of the species (i.e. all abiotic factors influencing its distribution being known) and sampling a large number of sites (>3000), "perfect" predictions of assemblage are difficult to attain, but very good predictions are reachable;
2) modelling technique and sample size were the most important factors, relative to the others tested (i.e. accounting for the importance of all factors at once);
3) contrary to previous studies on single SDMs, we showed that the choice of the modelling technique used to fit the models had a larger impact than sample size on S-SDM prediction success;
4) accuracy increases with sample size, but depending on the modelling technique (GLM or GAM) and sampling method (sampling 'plot-like' methods), accurate predictions could already be obtained with relatively small sample sizes (200 sites);
5) sampling species data using a 'plot-like' method is more desirable than sampling species individually, as it proves a better approximation of the true species prevalence and provides more accurate assemblage predictions.

ACKNOWLEDGEMENTS

This work was supported by the SNF project "SESAM'ALP - Challenges in simulating alpine species assemblages under global change". The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. We thank Damaris Zurell and Catherine Graham for their valuable comments on an early stage of this work.

Chapter 1.3

Using virtual species to assess the effects of truncated occurrence datasets on species distribution models

This manuscript is in preparation

My contribution to the paper: I conceived the initial idea and developed the methodological framework in coordination with Antoine Guisan and Daniel Scherrer. I developed all the necessary codes, analysed all the resulting data and led the writing of the various versions of the manuscript.

Using virtual species to assess the effects of truncated occurrence datasets on species distribution models

Rui F. Fernandes[a], Daniel Scherrer[a] and Antoine Guisan[a,b]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland
[b] Institute of Earth Surface Dynamics, University of Lausanne, Geopolis, 1015 Lausanne, Switzerland

ABSTRACT

Due to the growing threats to biodiversity, species distribution models (SDMs) are widely used to predict spatial distribution and range shifts of species under environmental changes, especially climate change. However, the accuracy of SDM predictions can be affected by different factors. An implicit assumption behind the use of these models to make projections is that available data should be covering the full range of climate conditions suitable to the modelled species, but this might not always be the case if truncated datasets (and truncated climatic information) are used. However, although it can lead to errors in projections, the influence of the level of truncation was not thoroughly analysed, especially when predicting future species assemblages. We created 20 virtual species in a real study area (Europe), ensuring that the full realized niche of each species was contained within the extent considered. Three different niche breadths were considered (wide, medium and narrow) and two training datasets were sampled across the initial extent, covering: (1) Europe (non-truncated climate) and (2) a smaller area with 500 km x 500 km (highly truncated). The models using each dataset were fitted with generalised additive models (GAM) to assess the effects of the level of truncation and variable importance, under current and future conditions. We found that truncated training datasets negatively influenced the accuracy of species distributions and respective assemblages, with that influence depending on the size of the species realised niche (stronger for wide realised niches). Predictions made under future conditions failed to produce accurate assemblages (Sørensen ≤ 0.7, Jaccard ≤ 0.5), both using non-truncated or highly truncated datasets. We also confirm the importance of variable selection since we found that when a set of predictors different from those used to create our virtual species (i.e. random predictors) were used to calibrate the models of all the species, the effects of truncation were even more pronounced (low or no similarity between observed and predicted assemblages). When using S-SDMs in a small area, we suggest that wide range species should be calibrated with information available at a larger extent (avoiding truncated datasets), while narrow range species might still be reasonably modelled with data that presents truncation, but this will depend on where the truncation occurs along the species' response to environmental gradients. We also highlight the importance of choosing predictors that correctly reflect the species environmental requirements. Finally, we also identify possible limitations and provide new perspectives on how to potentially improve assemblage predictions with truncated data information.

KEYWORDS: Species distribution models, truncated niche, virtual species, simulated realised niche, Europe, variable importance

INTRODUCTION

It is increasingly important to understand and predict the current and future distribution of species and their assemblages, especially considering the growing pressures and effects caused by global changes (e.g. habitat fragmentation, biological invasions, species extinctions).

Species distribution models (SDMs; see Guisan *et al.*, 2017) are a widely used tool to predict the distribution of species under these global changes, being applied by statistically relating species observations with environmental data (Guisan & Thuiller, 2005; Elith & Leathwick, 2009). This type of models can additionally be used to predict the distribution of species assemblages (richness and composition), with the most commonly used approach consisting in the modelling and subsequent simple stacking of the individual spatial predictions of all the species present in a regional species pool (stacked-SDMs or S-SDM; Ferrier & Guisan, 2006; Dubuis *et al.*, 2011).

Because SDMs can be easily implemented and require low amounts of species data (i.e. presence/absence data mainly from field observations), they are often used to study the response to climate change of multiple species at a time (e.g. Engler *et al.*, 2011; D'Amen *et al.*, 2015b). This type of models works by quantifying the climate space suitable to each species and projecting it in geographic space under present and future conditions. Since it is implicitly assumed when making projections that the full species' suitable climate space is captured in SDMs (Guisan et al. 2017), complications can appear when trying to project species distributions into regions/times that differ from those used to calibrate the models (Thuiller *et al.*, 2004; Hannemann *et al.*, 2016). This is because SDMs are often calibrated using data limited to a specific region or country, which might not include enough information about the species suitable conditions (i.e. truncated niche space). Therefore, it is important to take into consideration the extent of the study area where species occurrences are sampled, ensuring that the full geographical range of the species is captured (e.g. Barbet-Massin *et al.*, 2010), and/or assess the degree to which truncated datasets can affect model predictions (Thuiller *et al.*, 2004). Other factors (e.g. competition, incomplete or erroneous data) can also degrade the accuracy of distribution models (e.g. Araújo, 2001; Scherrer *et al.*, 2017), potentially leading, if part or all SDMs are biased, to wrong applied recommendations, e.g. to support conservation actions.

The causes of inaccurate model predictions when using truncated datasets that do not account for the full climatic range of the species are thought to result from the models being unable to capture the correct relationship (i.e. species' response curves; e.g. Austin, 1987; Austin & Gaywood, 1994) between the species occurrence and the climate (Thuiller, 2004a; Thuiller *et al.*, 2004; Barbet-Massin *et al.*, 2010), which is also the likely reason why some SDMs can have difficulties to predict species distributions into non-analogous climates (e.g. Harrison *et al.*, 2006; Williams & Jackson, 2007; Fitzpatrick & Hargrove, 2009; Rodríguez-Castañeda *et al.*, 2012).

A possible solution for reducing the risk of failed predictions when modelling in a restricted geographic extent (with the risk of modelling truncated niche) involves increasing the spatial extent of the study area (Thuiller *et al.*, 2004; Titeux *et al.*, 2017), provided that the spatial enlargement also results in enlarging the climate space used to fit the model. On the other hand, the effects of modelling algorithm and variable selection (i.e. variables used in SDMs should effectively capture the distributions bioclimatic range; Barbet-Massin *et al.*, 2010) on models that use truncated datasets have not been widely studied. In a recent study using real species, Hannemann *et al.* (2016) investigated the effects that the choice of modelling algorithm, variable selection and the degree of truncation have as causes of SDM prediction errors. The authors showed that for individual species, model performance decreases with truncated datasets, but the effects caused by the choice of algorithm and variable selection were more important to the observed prediction errors. Another study (Titeux *et al.*, 2017), examined if truncated models were able to predict accurate distribution under current and future environmental conditions. They found that these truncated or restricted-scale models fail to capture the full climatic space of the species, producing biased estimates of species distributions. However, no study so far assessed the effect of truncation on predictions of whole species assemblages and their community properties, e.g. when using stacked predictions from multiple SDMs (e.g. D'Amen *et al.*, 2015b).

It is therefore important to understand the limitations of SDMs and their stacking (i.e. S-SDMs) into assemblages when truncated datasets are used, and a robust way to test these limitations is the use of artificial data (Austin *et al.*, 2006) in a virtual ecologist approach (see Zurell *et al.*, 2010 for a review). Artificial or virtual species have been used in different SDM studies to test different factors affecting model performances (e.g. Hirzel *et al.*, 2001; Thibaud *et al.*, 2014), and are useful because all the information necessary for a specific study can be completely obtained/controlled in an artificial or "known" world. However, very few studies used such virtual ecologist approach to assess the effect of different factors on species assemblage predictions (i.e. S-SDMs; e.g. Fernandes *et al.*, 2018), and none to our knowledge assessed specifically the effect of truncation on S-SDMs. Here, using a virtual ecologist approach, we assess the effect of truncation on the accuracy of S-SDMs. We hypothesize that increasingly truncated datasets will decrease the predictive accuracy of assemblage predictions. We want to assess more specifically how greatly truncated datasets (i.e. level of truncation) can affect the accuracy of S-SDM predictions or if a decrease in accuracy is independent of the level of truncation, being influenced instead by variable selection, as previous studies demonstrated for single species (Hannemann *et al.*, 2016).

METHODS

Our virtual ecologist framework was implemented using virtual species and real climatic data in Europe, to test the effects that truncating datasets has on the predictive accuracy of community models. The approach is divided in five major steps (Figure 1):

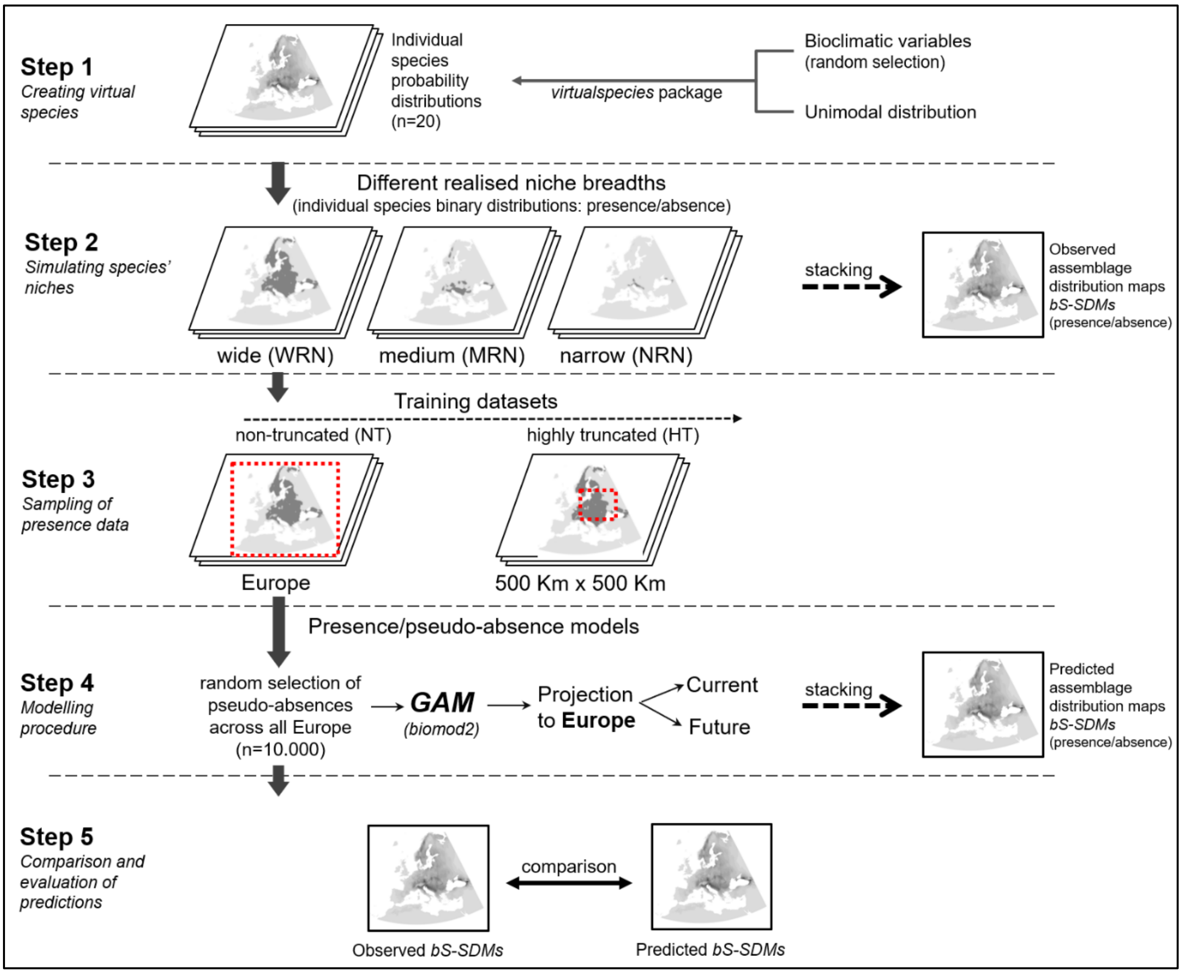


FIGURE 1 Framework of the analytical steps followed in this study. Step 1- we started by creating 20 virtual species suitability maps, ensuring that their response functions followed a unimodal distribution along environmental variables that were randomly selected for each species; Step 2- the probability of occurrence of each species was binarized using three different threshold values (i.e. 0.2, 0.5 and 0.7). These binary maps are considered as projections from wide (WRN), medium (MRN) and narrow (NRN) realised niches. For each threshold, we then stacked the single species' binary predictions to create "known" assemblage distribution maps (i.e. bS-SDMs); Step 3- species presence data (n=1000) was sampled to two training datasets covering opposing extents: (1) non-truncated –NT- Europe, and (2) 500 km x 500 km square (highly truncated – HT). The first dataset used the entire study area, while the centre of gravity of each species geographic distribution was considered for the calculation of the highly truncated areas; Step 4- the sampled datasets were used to create SDMs using the same environmental predictors selected to create the virtual species, both under current and future climatic conditions. Five pseudo-absences datasets consisting of 10.000 pseudo-absences each where randomly selected, ensuring that the pseudo-absences had the same weight as presence data in the models. The models were then fitted using generalized additive models (GAM) and the resulting predicted probabilities of occurrence (i.e. projected for Europe) were transformed into presence-absence data (considered as our known virtual species distribution) using the threshold maximizing TSS, and those binary predictions were stacked to create species assemblages; Step 5- to determine the effects of truncated *vs.* non-truncated training datasets, we evaluated and compared our predicted and observed assemblages using a random sample of four million points (across all Europe) at 1 km^2 resolution for each predicted assemblage. That data was then used to calculate two main indices of assemblage prediction similarity (i.e. Sørensen and Jaccard indices).

Step 1 – Creating virtual species: We started by generating a set of virtual species (n=20) for the region of Europe/North Africa at 1 km^2 resolution. The virtual species distributions were created using the *virtualspecies* package in the R software (Leroy *et al.*, 2016; R Core Team, 2017). To create our set of species, we ensured that their response functions followed unimodal distributions along environmental variables, with a random number of them being selected for each species (obtained from the Worldclim database; Hijmans *et al.*, 2005) to determine the initial (true) distribution of each species. This set of variables was randomly selected by choosing temperature and precipitation variables (minimum of 1 and maximum of 4 for each group). The correlation among those variables was analysed using Pearson's correlation coefficient. If the correlation between two variables was detected (Pearson's R > 0.7), only one was randomly selected. For each virtual species, we combined all unimodal response curves to obtain a probability distribution map depicting the suitability of each species in Europe (i.e. which in this case can be considered as the projection of the species' realised climatic niche).

Step 2 – Simulating species' niches: In the previous step, we were able to create a simplified realised niche of the species, by using climatic variables and ensuring that the species only occur in our study area. Since our species information was available as a probability of occurrence, we created three sets of individual species binary distributions, selecting for that different thresholds (i.e. 0.2, 0.5 and 0.7). These can be described as projections from wide (WRN), medium (MRN) and narrow (NRN) realised niches respectively. Independently of the threshold, these three projections thus share the same niche (i.e. have the same response curves to all climatic variables), and by setting the different thresholds we define different sizes of realized niches and associated geographic distributions. Finally, for each threshold, we stacked the single species' binary predictions to create "known" assemblage distribution maps (i.e. binary stacked SDMs, bS-SDMs).

Step 3 – Sampling of presence data: From the available binary virtual species distributions, we sampled species presence data (n=1000) to create two training datasets covering opposing extents: (1) non-truncated –NT- Europe, and (2) 500 km x 500 km square (highly truncated – HT). These training datasets were selected to simulate the possible occurrence of truncated climatic spaces, and while for the first dataset we used the entire study area, we considered the centre of gravity of each species geographic distribution for the calculation of the highly truncated square/areas. This centre of gravity was chosen to delimitate the truncated area to ensure that species presences were available to be sampled, eliminating the chance of sampling in an area were no species occur.

Step 4 – Modelling procedure: The sampled datasets were used to create SDMs using the same environmental predictors selected to create the virtual species, both under current and future climatic conditions (RCP8.5 - mean global warming increase of 2ºC between 2046-2065 and 3.7ºC between 2081-2100 -, for 2070; available at http://worldclim.org/CMIP5v1). This ensured that the models could perfectly replicate the distributions of each species, since all information that initially defined these distributions was made available (i.e. same climatic predictors). We randomly selected 5 pseudo-absences datasets consisting of 10.000 pseudo-absences each, and ensuring that the pseudo-absences had the same weight as presence data in the models (Barbet-Massin *et al.*, 2012). The models were then fitted using generalized additive models (GAM; Hastie & Tibshirani, 1990), with each individual model using 70% of the data as calibration and 30% as validation data (i.e. random cross-validation), with the selected data being different in each repetition (procedure repeated five times). Therefore, for each species, a total of 25 models were run (1algorithm x 5 repetitions x 5 pseudo-absences datasets). The resulting predicted probabilities of occurrence (i.e. projected across Europe) were transformed into presence-absence data (considered as our known virtual species distribution) using the threshold maximizing TSS (MaxTSS; equivalent to the sensitivity-specificity sum maximization in Liu *et al.*, 2005). All models were run in the R software version 3.3.3 (R Core Team, 2017), using biomod2 (Thuiller *et al.*, 2009). Additionally, to test the influence of variable selection in the accuracy of SDM, we followed the previous described procedure but using five environmental variables to fit the models (i.e. same variables used for all the species): annual mean temperature (º C), temperature seasonality (C of V), the mean temperature of the warmest quarter (º C), precipitation seasonality (C of V) and precipitation of warmest quarter (mm).

Finally, the binary predictions of each species were stacked to create species assemblages, providing both species richness and composition for each pixel in the study area. The predicted assemblages resulting from this stacking of binary SDMs could then be compared with our "known" initial assemblages (S-SDM).

Step 5 – Comparison and evaluation of predictions: Finally, to determine the effects of using truncated *vs.* non-truncated training datasets when modelling assemblage distributions, we evaluated and compared our predicted and observed assemblages. We sampled a set of four million points (across Europe) at 1 km^2 resolution for each predicted assemblage, which were then used to calculate two main indices of assemblage prediction similarity using the *ecospat.SSDMeval* function available in the "*ecospat*" R package (see Table S1 in Appendix A): (i) the Sørensen index (Sørensen, 1948) and (ii) the Jaccard index (Jaccard, 1912).

RESULTS

Model accuracy

The results revealed a clear pattern in the accuracy of the models, with those that were fitted using the wide realised niche of the species (WRN) presenting evaluation values lower than models calibrated using the narrow (NRN) realised niche (Figure 2). Concerning the average values of MaxTSS, we observed that models calibrated using the WRN-NT setting presented MaxTSS values ranging between 0.36 and 0.98 (mean =0.78 ± 0.18). These values were higher for WRN-HT models, ranging between 0.66 and 0.99 (mean =0.87 ± 0.09). The evaluation values obtained for the MRN and NRN models were very similar between them (MaxTSS >0.95; see Table S2 in Appendix A). The prevalence values for the virtual species' "known" binary distributions (Table S2) ranged between 0.02 and 0.63 for the WRN (0.19 ± 0.17) and between 0.003-0.08 for the NRN (0.03 ± 0.03). Finally, we observed that the species with the highest prevalence usually presented the lowest evaluation values, something especially evident for models calibrated using the wide realised niches (WRN; Table S2).

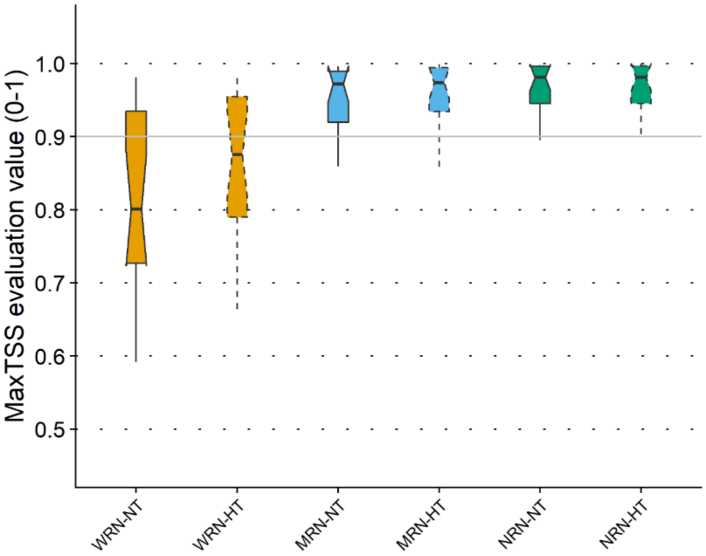


FIGURE 2 Evaluation values (MaxTSS) of individual species distribution models (i.e. calibrated using the variables used to create each virtual species) for each simulated realised niche (WRN - wide realised niche; MRN - medium realised niche; and NRN - narrow realised niche) and training dataset (NT: non-truncated and HT: highly truncated).

The effect of truncated datasets

A first visual comparison between observed (true) and predicted assemblage patterns under current and future environmental conditions (for species fitted with the same variables used to generate them) across the projection area (Europe) revealed that the overall species richness level got reduced with truncation (Fig. 3), a pattern more clearly seen for species modelled with wide realised niches (Fig. 3c and e). A slight decrease (one or two species) in species richness was also visible with increasingly truncated training datasets (MRN, Fig.3f, g and g; NRN, Fig. 3k, m and o).

These patterns were further confirmed when analysing the different indices of assemblage prediction accuracy (Figure 4). Looking at the presented similarity indices (i.e. Sørensen and Jaccard), one could see the effect that increasingly truncated datasets have on the accuracy of assemblage predictions (i.e. community similarity decreases when truncated dataset is used). However, confirming what we already observed in the spatial predictions of species distributions (Fig. 3), the truncated datasets had a much larger effect on assemblages of species with wide realised niche (WRN; Fig.4 a and b). This could be seen (Fig. 4a) in the reduction of the similarity between observed and predicted assemblages, with a drop from around 1 (high similarity between observed and predicted assemblages) when modelled with the non-truncated (WRN-NT) datasets to around 0.7 and 0.5 – for Sørensen and Jaccard respectively – when modelled with highly truncated datasets (WRN-HT). This pattern was also evident when analysing predictions of future assemblages (Fig. 4c and d), with a noticeable effect of truncation visible for assemblages of species with WRN (average Sørensen drops from around 0.65 for WRN-NT to 0.3 in WRN-HT). That reduction in similarity also occurred, but with more variation, for species with MRN (Fig. 4c and d; average Sørensen around 0.5 for NT and 0.35 for HT, Jaccard around 0.3 and 0.2 for NT and HT, respectively) and NRN (Sørensen around 0.58 for NT and 0.45 for HT, Jaccard around 0.4 and 0.28 for NT and HT, respectively).

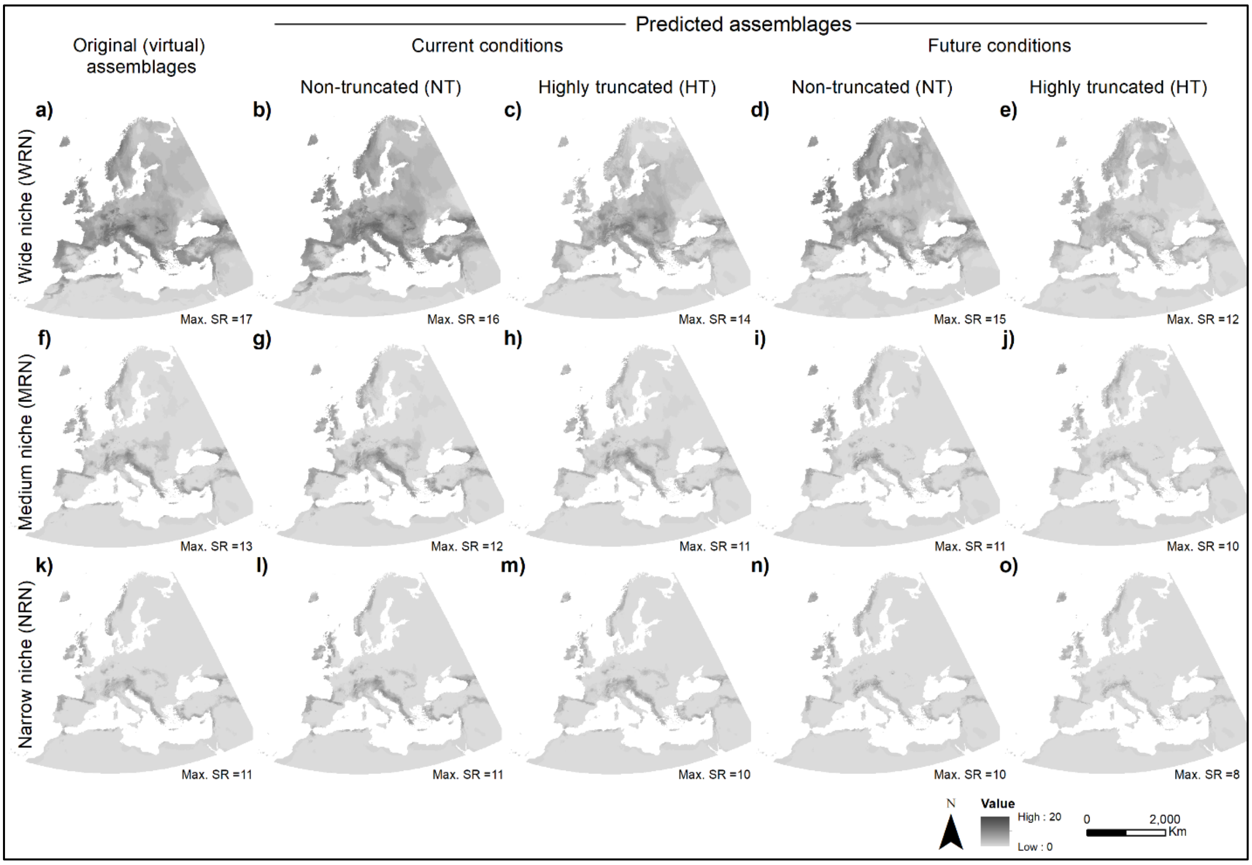


FIGURE 3 Spatial representation of observed and predicted assemblage species richness for each simulated realised niche (WRN - wide realised niche; MRN - medium realised niche; and NRN - narrow realised niche) and training dataset (NT: non-truncated and HT: highly truncated), for predictions under current and future environmental conditions. These models were fitted using, for each species, the same set of environmental variables used to create our virtual species.

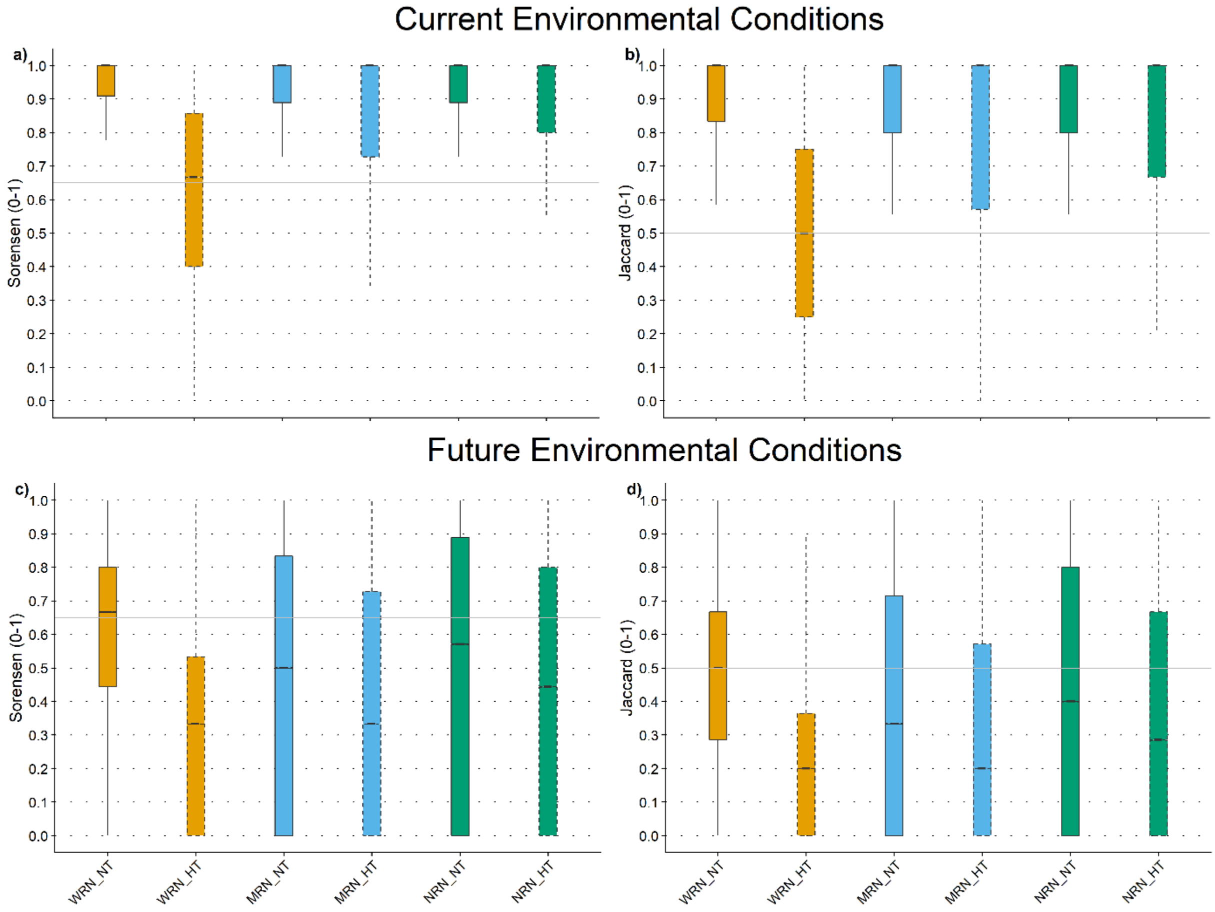


FIGURE 4 Boxplots of different indices of assemblage prediction (bS-SDM) accuracy - (a) Sørensen and (b) Jaccard - for all the simulated realized niches (WRN - wide realized niche; MRN - medium realised niche; and NRN - narrow realised niche) and training datasets (NT – non truncated and HT - highly truncated), for predictions under current and future environmental conditions. These models were fitted using, for each species, the same set of environmental variables used to create our virtual species. Each box shows the variation of each index across a random subset of Europe (four million points) for the binary predictions obtained using MaxTSS as thresholding technique.

The effect of variable selection

If we take the effect of variable selection into consideration (Figure 5), by fitting SDMs with the same variables for all the species instead of using the variables known to determine species distributions (Figure 4; only possible using virtual species), we could see that: 1) MaxTSS evaluation values were lower (around 0.8) for the models fitted using non-truncated (NT) datasets and higher (MaxTSS ≥ 0.9) for models using highly truncated (HT) datasets (see Figure S1 in Appendix A); 2) the similarity between observed and predicted assemblages was lower for all species, thus with all sizes of simulated realized niches (i.e. WRN, MRN and NRN; on average, Sørensen and Jaccard values were below 0.5 except for WRN-NT); and 3) the effect of truncation was higher, especially for WRN under current conditions (Fig. 5a and b; similarity measured by Sørensen/Jaccard is on average close to zero when using highly truncated datasets) and under future conditions for all sizes of realized niches (no similarity between observed and predicted assemblages with highly truncated datasets).

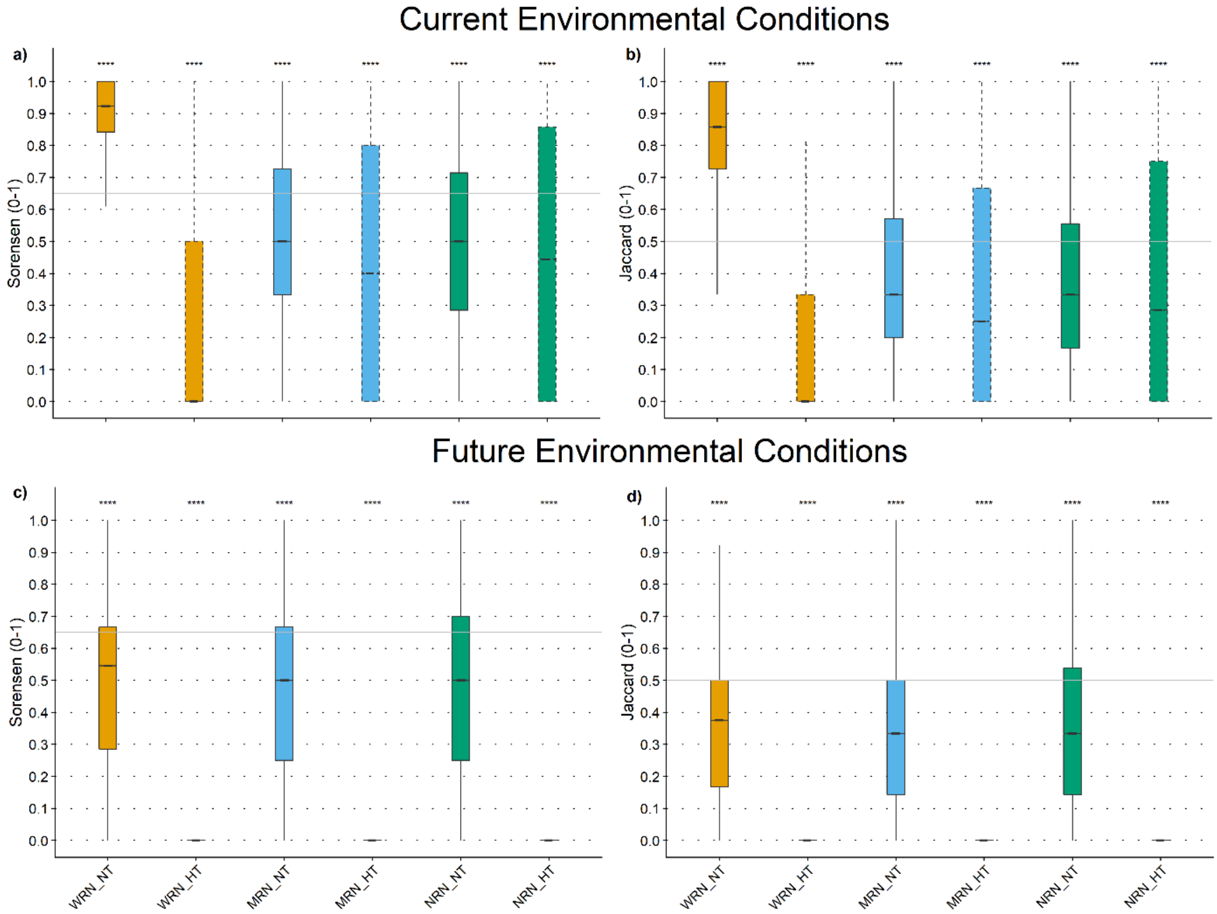


FIGURE 5 Boxplots of different indices of assemblage prediction (bS-SDM) accuracy - (a) Sørensen and (b) Jaccard - for all the simulated fundamental niches (WRN - wide realized niche; MRN - medium realized niche; and NRN - narrow realized niche) and training datasets (NT – non truncated and HT - highly truncated), for predictions under current and future environmental conditions. These models were fitted using the same set of environmental variables for all the species (random predictions; i.e. a set of predictors different from those used to create our virtual species). Each box shows the variation of each index across a random subset of Europe (four million points) for the binary predictions obtained using MaxTSS as thresholding technique.

DISCUSSION

For species distribution models to be useful for conservation or management actions (e.g. Franklin, 2013; Vicente *et al.*, 2016), it is important that they can provide accurate current and future predictions of the distribution of individual species and assemblages. SDMs are built by statistically relating information from species presence (and absence) with environmental conditions to quantify and predict spatially where suitable conditions for a species exist (Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005). However, potential bias can occur when truncated datasets are used for calibration, with the models being unable to capture the full and correct relationship between species occurrence and the environment (e.g. Thuiller, 2004a; Thuiller *et al.*, 2004; Barbet-Massin *et al.*, 2010). In this study, starting with virtual species for which the exact niche and related distribution are known (i.e. no risk of initial sampling bias or imperfect detection) and using the same environmental predictors to define their initial distributions and fit the models, we observed that the truncation of the training datasets had a clear negative influence on the accuracy of species predictions, and thus on the accuracy of the predicted assemblages (Fig. 4), but that this influence depended on the size of the species' realized niches (i.e. stronger influence for assemblages of species with wide realized niches). Since it seems statistically easier to characterize distributions of species with narrow environmental tolerances (e.g. Guisan & Hofer, 2003; Kadmon *et al.*, 2003; Wisz *et al.*, 2008), it was somewhat expected that models for species with narrow niches would produce better predictions under current environmental conditions, especially when compared with wide realised niche (WRN) species (also observed in e.g. McPherson *et al.*, 2004; McPherson & Jetz, 2007). This could be an indication that if species assemblages need to be predicted in an area with a small extent (e.g. Switzerland or a specific area in a larger country), a good solution could include the modelling of narrow range species using data sampled in that small area, with wide range species being modelled with data sampled across a larger extent (i.e. avoiding truncation; see Petitpierre *et al.*, 2016). A possible explanation for this pattern of wide range species producing less accurate predictions might be because they occupy a large range with greater environmental variability, leading to less accurate relationships between species occurrence and the environment (McPherson & Jetz, 2007). Another explanation could be due to the method used to select our truncated areas, which might favour narrow ranged species. The area was delimitated around the centre of gravity of each species geographic distribution and while this approach guaranteed that we selected an area where the species was present, it also had the tendency to always contain a large part of the species niche, likely centred around the species' climatic optimum, and thus was more likely to capture the correct relationship between species presence and the environment for species with small ranges (i.e. for which the full information about the species response curve remains inside the truncated area).

Despite the higher similarity between observed and predicted assemblages of narrow realized niche (NRN) species under current predictions (Sørensen close to 1), all models failed to produce accurate assemblages when predictions were made to future climatic conditions (Sørensen ≤ 0.7 and Jaccard ≤ 0.5). This drop in similarity between observed and predicted assemblages occurred both under models calibrated using non truncated (NT) and truncated (HT) datasets (with lower similarity with HT datasets; lower Sørensen and Jaccard, Fig. 4c and d). This indicates that even with complete knowledge of the factors influencing each species distributions, and fitting models with the same environmental variables used to create the virtual species' response curves, we were unable to produce predictions of species assemblages with similarity greater than 0.7 for Sørensen and 0.5 for Jaccard under future conditions (see Fig. 4).

We also found that when a random set of predictors (i.e. predictors different from those used to create our virtual species, and therefore not necessarily determining species suitability) were used to calibrate the models for all the species, the effects of truncation became even more pronounced (i.e. producing predicted assemblages with low similarity with the observed ones; Figure 5), especially for WRN species under current environmental conditions and for all types of realized niches (i.e. WRN, MRN, NRN). These results confirm the importance of variable selection, something also shown to be important in previous studies (Hannemann *et al.*, 2016; Fourcade *et al.*, 2018), and that selection should be made considering the species ecology (i.e. like we did here when models were fitted with the same variables used to create the species). The selection of adequate predictors should also be made considering their capability of providing SDMs with good transferability potential (i.e. ability to predict distributions into conditions not accounted for in the calibration dataset) (see Petitpierre *et al.*, 2017 for strategies on how to choose those variables).

Additionally, we found that the values expressed by the used evaluation metric (MaxTSS > 0.8) would lead one to consider all models as having good performance and be accurate, even if their predictions provide predicted assemblages that have similarity lower than 0.7 for Sørensen and 0.5 for Jaccard when compared with observed (true) assemblages, especially for models fitted with predictors that were not used to determine the virtual species' response curves (Fig. 5). This seems to indicate that some currently used evaluation metrics, such as MaxTSS in this case, are unable to correctly assess the predictive performance of distribution models (e.g. Fourcade *et al.*, 2018; Fernandes *et al.*, in press).

Our study mainly demonstrated the effect that truncated datasets can have on the prediction accuracy of species and their assemblages (particularly for wide range species), but also that it is important to correctly identify the environmental predictors explaining species distributions and improve the capacity of spatial modelling strategies to make them able to predict models into new environmental conditions (i.e. transferability), even when non truncated datasets are used to fit SDMs. We further showed that the use of artificial data in a virtual species approach can be an important tool to test and validate various methods or strategies associated with species distribution modelling. Based on our results and considering some limitations observed in our analysis, we suggest that:

- Wide range species should be calibrated using the largest extent possible, avoiding truncated datasets;
- Predictions under future environmental conditions should be considered with great care, especially those generated from truncated datasets;
- Correctly selecting predictors that reflect the species environmental requirements is important (see Petitpierre *et al.*, 2017);
- Determining if the full relationship between environmental data and species occurrence is captured (i.e. analysing species response functions), and evaluating the predictions made when using truncated data or projections into new environments, is also important;
- The use of the same approach with fixed truncated areas for all species also needs to be tested, to determine if the narrow range species would keep being less affected by truncated datasets

or if it was an artefact of our truncated area selection centred on the species' geographic centroid;

- In this study, we were able to sample the same number of presences (n=1000) for each species. A new iteration of this framework should consider testing also the effects of sample size, increasing the complexity of the study;
- It was previously shown that individual algorithms have important influence on prediction errors (Hannemann *et al.*, 2016). In this study we used only generalised additive models (GAM). The influence of individual algorithms and their ensembles should be further tested when making predictions using truncated datasets;
- Finally, while considered useful to help in conservation and management decisions, SDM predictions should be carefully employed, indicating both their strengths, limitations and uncertainties.

ACKNOWLEDGEMENTS

We thank Damaris Zurell and Catherine Graham for their valuable comments on an early stage of this work. This work was supported by the SNF project "SESAM'ALP (nr 31003A-1528661) - Challenges in simulating alpine species assemblages under global change". The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics.

Chapter 1.4

Assessing the transferability of species distribution models in time with virtual species: are hierarchical models more transferable?

This manuscript is in preparation

My contribution to the paper: I conceived the initial idea and developed the methodological framework in coordination with Antoine Guisan. I developed all the necessary codes in collaboration with Daniel Scherrer. I also analysed all the resulting data and led the writing of the various versions of the manuscript.

Assessing the transferability of species distribution models in time with virtual species: are hierarchical models more transferable?

Rui F. Fernandes[a], Daniel Scherrer[a] and Antoine Guisan[a,b]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland

[b] Institute of Earth Surface Dynamics, University of Lausanne, Geopolis, 1015 Lausanne, Switzerland

ABSTRACT

Species distributions models (SDMs) are increasingly used to project species' future distributions under novel climatic conditions, a procedure usually involving model transferability, leading some SDMs (e.g. using truncated datasets) to show limited predictability and biased projections when applied to different areas or climatic conditions. We tested and compared how accurate different strategies (i.e. based on datasets with different resolutions, truncated or using a multiscale approach) are at predicting species assemblages under current and future climatic conditions (i.e. assessing their transferability). We were particularly interested in determining the validity of hierarchical models. The strategies were run using presences/absences (PA) or presence/pseudo-absences (PO) data, since SDMs can be fitted with both types of data, and it is useful to assess their respective influence on model transferability. We created 100 virtual species at a 1 km^2 resolution over the entire world, and used then seven strategies based on various ways to sample the virtual species to fit SDMs: global and truncated downscaling models (DwSc, TDwSc), fine and truncated fine scaled models (FnSc, TFnSc), a new modelling approach mixing coarse and fine scale data (Mxd), and two hierarchical models that differ in the weights given to pseudo-absences (Hier1 and Hier2). Ensemble models were built with five techniques (GLM, GAM, ANN, BRT and RF) using either PA or PO data, as follows: (i) 1000 points with random PA points reflecting the true prevalence of the species; (ii) using the same number of presences and pseudo-absences (POequal); (iii) or the same number of presences as previously used and 10.000 randomly sampled pseudo-absences (PO10M). All models were projected onto Europe under current and future environmental conditions. As expected, PA models (MaxTSS $\geq$0.9) gave better current assemblage predictions (i.e. similarity between observed and predicted assemblages, Sørensen (S) $\geq$ 0.8) than PO models (S $\leq$ 0.6), but all PA strategies failed to give accurate future assemblage predictions (S $\leq$ 0.8). The tested PO models also presented good evaluation values for single models (MaxTSS<0.8), but this did not translate into as good predicted assemblages (S $\leq$ 0.6 under current conditions and S $\leq$ 0.5 under future conditions). The tested hierarchical models did not improve predictions over the remaining tested strategies (S $\leq$ 0.6). The transferability capabilities of the tested modelling strategies thus warn us against careless use of these models for projections in time or space. Furthermore, we identified possible limitations within the different strategies, providing perspectives/guidelines to potentially improve their accuracy and transferability.

KEYWORDS: Bioclimatic models, hierarchical approach, multi-scale, species distribution modelling, virtual species, downscaling, truncation

INTRODUCTION

As the effects of global warming become more prevalent, the impacts of climate change on species distribution and biodiversity will be increasingly high, possibly with devastating consequences (Sala *et al.*, 2000; Barnosky *et al.*, 2011). It is then important to understand the processes determining those species distributions, both in space and time, and if and which methodological factors can affect predicted patterns of assemblages (i.e. communities) under current and future climatic conditions.

Species distribution models (SDMs; see e.g. Guisan *et al.*, 2017 for a review) are one of the most used and relevant tools to model the distribution of species, also predicting range shifts caused by global changes at different spatial and temporal scales and the distribution/composition of species assemblages (e.g. Dubuis *et al.*, 2011; D'Amen *et al.*, 2015b). The simplest and most commonly used method to predict these assemblages is the stacked-SDMs (S-SDM; Ferrier & Guisan, 2006; Dubuis *et al.*, 2011). This method consists in modelling the distribution of all individual species present in a regional species pool and then summing (i.e. stacking) their predictions to obtain assemblages, giving information about species richness and composition. However, recent studies showed that S-SDMs can sometimes over-predict species richness per site (e.g. Pineda & Lobo, 2009; Dubuis *et al.*, 2011; Guisan & Rahbek, 2011; Pineda & Lobo, 2012) or can be sensitive to methodological biases (Calabrese *et al.*, 2014; Scherrer *et al.*, 2018a). Additionally, single SDMs can also contain various sources of uncertainty derived either from deficiencies in the data (e.g. unavailable absence data; incorrectly identified presences/absences, imperfect detection) or in their ecological realism (see e.g. Barry & Elith, 2006; Buisson *et al.*, 2010; Rocchini *et al.*, 2011; Beale & Lennon, 2012), which might propagate into assemblage predictions.

Species distribution models are also being increasingly used to project future distributions of species under novel environmental conditions and to estimate the potential spread of biological invasions, for example. This procedure usually involves some sort of model transfer into new regions/time (i.e. model transferability; e.g. Randin *et al.*, 2006; Peterson *et al.*, 2007). What this means is that models are calibrated based on conditions observed in a certain time/region and transferred (i.e. projected) to another. This model transferability can be problematic, especially when trying to predict species distributions under environmental changes that are outside the range of data used to calibrate a model, potentially leading to biased predictions (e.g. Owens *et al.*, 2013; Fitzpatrick *et al.*, 2018; Qiao *et al.*, 2018). Recently, a set of knowledge gaps was identified, that if resolved could increase the accuracy of models transferred to new conditions, the authors argue (see Yates *et al.*, 2018 for a review). An intuitive way to minimize transferability problems is to use data that covers large spatial extents (continental or world scales; e.g. Pearson & Dawson, 2003; Sánchez-Fernández *et al.*, 2011; Titeux *et al.*, 2017), this way capturing the entire realized niche of species (i.e. the conditions where a species can survive with the effects of biotic interactions; Pearson & Dawson, 2003), but often only climatic data are available at such extent (e.g. Pearson & Dawson, 2003; Thuiller, 2004b). Furthermore, this approach also entails some problems since broad-scale data might not be available for several species' groups (Hudson *et al.*, 2014), especially microorganisms (e.g. soil protists; Geisen *et al.*, 2017). It also often occur that, due to data limitations (e.g. misidentification of presences/absences, spatial uncertainty, atlas data only available at coarse resolutions), the knowledge about species distributions is only available at grains or resolutions that are coarser than the ones at which biological processes act, or at which meaningful management decisions can be made (Franklin, 2013; Keil *et al.*, 2013). Budgetary reasons, species misidentification or a faulty sampling strategy, can additionally lead to incomplete information about the full species' distribution. This means that a narrower niche might be sampled and modelled for a species (i.e. model fitted using truncated/biased response curves; Barbet-Massin *et al.*, 2010) than the one it actually occupies (e.g. truncated climatic niches; Hannemann *et al.*, 2016), leading to biased predictions of current species distributions or projections of future ones (Petitpierre *et al.*, 2016). Truncation usually occurs because one defines a study area according to a political or known geographic limit (e.g. countries, continents), even if the distribution of the species occurs beyond those limits. Therefore, it is reasonable to expect that under different conditions or datasets, one can encounter species data that: (i) is at a coarser resolution than what would be desirable; (ii) is only available for a portion of the species' niche, both at coarse or fine resolutions (i.e. information about the full niche unavailable; geographic or environmental niche truncation); or (iii) does not contain absence data, so that only presence data are available to fit the models.

Different solutions have been proposed to account for these problems. For example, model-based downscaling approaches have been proposed to use coarse species occurrences to predict distributions at finer resolution (Bombi & D'Amen, 2012; Fernandes *et al.*, 2014). However, most methods published so far had reported limitations, were based on some unrealistic assumptions and provided mixed or ambiguous results (McPherson *et al.*, 2006; Bombi & D'Amen, 2012; Keil *et al.*, 2013). Different hierarchical models (i.e. combining local and global scale analyses) were proposed to capture the maximum information about species' full realized niches into the models before projecting them into new areas or time periods under new environmental conditions (e.g. Pearson *et al.*, 2004; Gallien *et al.*, 2012; Talluto *et al.*, 2016). The main idea behind these models, even if the framework varies, is to build SDMs at different extents, one large enough to cover the full climatic niche of the species and a second, more local or regional one (where the study takes place) accounting for more local/regional predictors (e.g. land-use, substrate; Pearson *et al.*, 2004; Gallien *et al.*, 2012; Petitpierre *et al.*, 2016; see Guisan *et al.*, 2017). Their use is promising, but further testing is needed to assess their reliability and promote a more general use of them.

Species distribution models can be generated either using species presence/absence (PA) or presence only data (PO; recurring to the creation of random pseudo-absences). The use of real absence data was judged preferable by some authors to models fitted with presence/pseudo-absence data (e.g. Brotons *et al.*, 2004), while others found advantages in the use of PO models (e.g. Elith *et al.*, 2006; Petitpierre *et al.*, 2016), which can depend on whether sensitivity or specificity should be favoured in the predictions (Maher *et al.*, 2014), and thus the question remains partly unanswered. Furthermore, while real species have widely been used as suitable models to study the effects of climate change in their distributions (e.g. Engler *et al.*, 2011; Regos *et al.*, 2015), they convey a lot of uncertainty, whereas the use of virtual species (i.e. or artificial/simulated species; see e.g. Hirzel *et al.*, 2001; Austin *et al.*, 2006; Wisz & Guisan, 2009; Meynard & Kaplan, 2013) allows conveniently relying on a fully known "artificial" world. In other words, the use of virtual species allows for the complete control of the data and models/approaches being tested (Zurell *et al.*, 2010), making it possible to robustly test the previously mentioned limitations (e.g. Fernandes *et al.*, 2018) – e.g. to determine how accurate different strategies behave when dealing with different types of data (e.g. PA vs. PO data, multiple scales) or under different transferability or change of scale situations (e.g. new climatic conditions) – under controlled conditions.

In this study, we used virtual species data created at global scale and relatively fine resolution (i.e. the entire world at 1 km^2 resolution) to implement species distribution models built with different strategies (i.e. based on datasets with different resolutions, with truncation or using a multiscale approach). We tested and compared these strategies - that might be used in different studies depending on the available data or multiple scales – in order to: (i) determine how accurate they predict current species assemblages (i.e. are they able to predict assemblages similar to the observed ones), using both PA and PO data; (ii) assess their transferability when projecting into a new area/climatic conditions; and (iii) evaluate the value of using hierarchical models compared with non-hierarchical ones. We also proposed a new strategy, using coarse-grain species data available to randomly sample fine-grain environmental data, and combine these with high-resolution data (available in the study area) to fit a 'mixed' model. Based on our results, we discuss the advantages/disadvantages of each strategy, giving future perspectives on how they can be improved. Considering the wide and increasing use of SDMs, and particularly S-SDMs, in biogeography and conservation, this paper thus provides a comparative analysis of the potential usefulness of different modelling strategies to derive spatial projections in time and space, while discussing current limitations of available approaches and of this study and identifying possible future improvements.

METHODS

Analytical framework

We implemented a virtual ecologist approach (see Figure 1), to compare different modelling strategies and assess their transferability and accuracy at predicting assemblages when different types of data are available (presence-absence, PA, versus presence-only, PO) and environmental conditions occur (current vs. future conditions). The approach consisted of five steps:

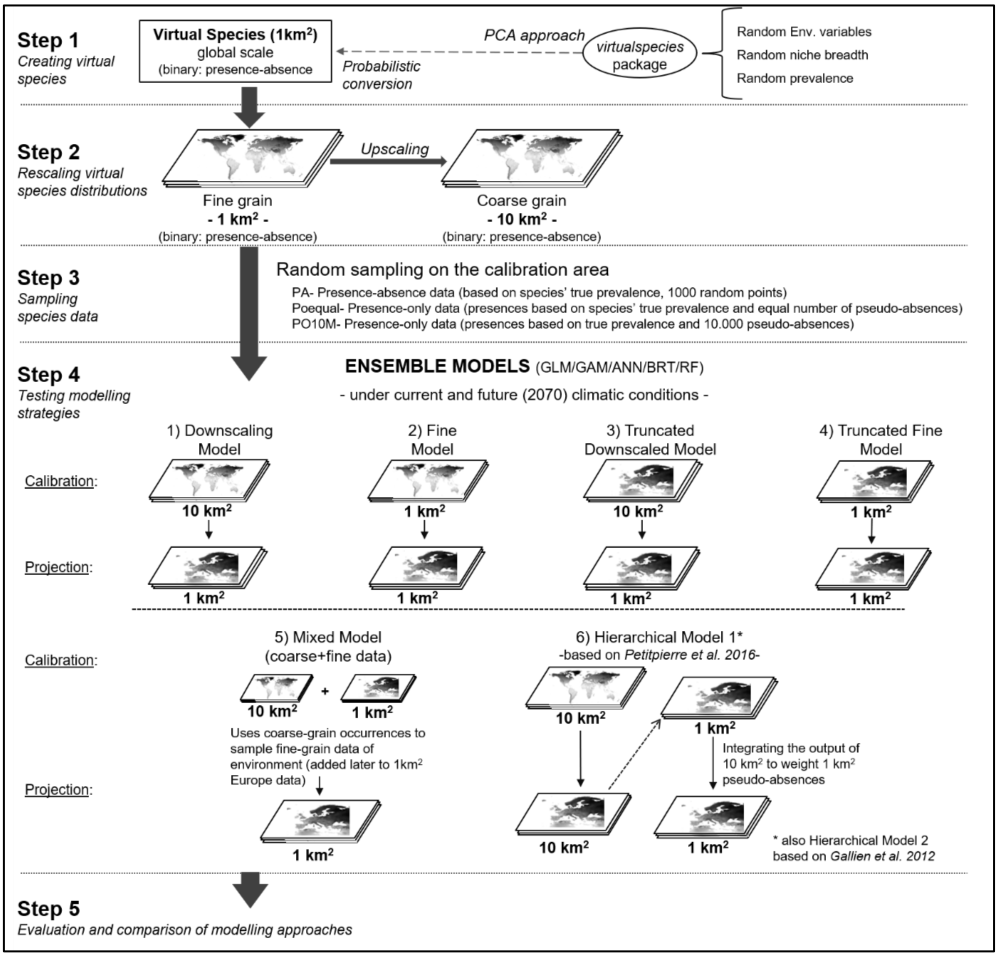


FIGURE 1 Workflow of the analytical steps followed in the study. Step 1 – We started by creating global binary distribution maps, at 1 km^2 resolution, for 100 virtual species. Each species was created using a random set on environmental variables, random niche breadth (i.e. wide/narrow) and random prevalence. Using the PCA approach of *virtualspecies* R package, species response to environmental data set, defining their suitability. That suitability was then converted in binary presence/absence data using the probabilistic approach (i.e. making a random draw of presence/absence weighted by the probability of occurrence). Step 2 – In order to be able to test approaches based on different presence-absence data resolutions, the original distribution maps were upscaled to a coarser resolution (10 km^2). Step 3 – Three sets of data were randomly sampled to fit the models: 1000 random presence/absence points (PA) ensuring the species' true prevalence, a set with an equal number of presence/pseudo-absence data (POequal) and presence/pseudo-absence data with a random set of 10.000 pseudo-absences (PO10M). Step 4 – We then used either that sampled presence/absence (PA) or presence-only data (PO) to calibrate the different modeling approaches under current and future climatic conditions (2070). We tested six different approaches : Downscaling model; Fine model; Truncated downscaling model; Truncated fine model; Mixed model; and Hierarchical model 1 and 2. Step 5- The different modelling techniques were then compared and evaluated to determine their accuracy in prediction species assemblages when using either presence-absence or presence-only data.

Step 1 - creating virtual species

To create our set of 100 virtual species, we started by simulating each species distribution at global scale and fine resolution (i.e. the entire world, 1km^2), using the *virtualspecies* package in software R (Leroy *et al.*, 2016; R Core Team, 2017). In order to obtain species with realistic environmental requirements, a set of variables was selected for each species (obtained from the Worldclim database; Hijmans *et al.*, 2005). This set of variables was selected by randomly choosing different temperature and precipitation variables (minimum of 1 and maximum of 4 for each group), for each species. The correlation among variables was analysed using Pearson's correlation coefficient. If the correlation between two variables was detected (Pearson's $R > 0.7$), only one was randomly selected. Using an approach available in the *virtualspecies* package, we generated a Principal Component Analysis (PCA) space of the selected environmental variables for each species, defining their responses in two PCA axes. This approach allowed us to generate virtual species with plausible environmental requirements, while still being initially dependent on all our variables (but with different importance in defining the axes). To increase the variability of virtual species, the width of the species niche (i.e. wide or narrow) and the species' prevalence (i.e. the number of places occupied by the species from all the available locations) was also randomly assigned to each species. The species' environmental suitability was converted into presence-absence using a probabilistic approach (i.e. the probability of getting a presence of a species in a given pixel is dependent on its suitability in that pixel; Meynard & Kaplan, 2013), available in the *virtualspecies* package. In this probabilistic approach, the environmental suitability of each pixel is converted into a probability of occurrence. That probability is then used to assign a presence or absence to each pixel (i.e. making a random binomial draw based on the probability of occurrence; see also Dubuis *et al.*, 2011; Pellissier *et al.*, 2013). This conversion into binary data can be made by adjusting specific parameters (see below) and can range from a threshold-like conversion to linear or logistic. The two main parameters are α (i.e. controls the slope of the conversion curve; low values make the conversion curve linear, medium values logistic and high values threshold-like) and β (i.e. controls the inflexion point). We decided to attribute a random value of α to each species (i.e. ranging between -0.1 and -0.01), ensuring that the conversion would be logistic, with β being chosen depending on the slope of the curve. This was done because it was shown to produce more realistic virtual species (Meynard & Kaplan, 2013; Leroy *et al.*, 2016). The species current binary distribution was then considered as our "known" distribution, and the same parameters (i.e. response of the species along the PCA) used to define the species' current distribution were applied to map the future distributions, ensuring that the species' niche is identical - and thus the species is at equilibrium - between both time periods (or areas). To obtain the future distributions, we used environmental variables corresponding to the year 2070 (i.e. our "known" future distribution; based on the representative concentration pathway "RCP 8.5" from the IPCC fifth Assessment Report -AR5). This concentration pathway projects a global warming increase of 2 ºC on average, with emissions continuing to rise throughout the 21^{st} century and ranging between 1.4 and 2.6 ºC for the period 2046-2065.

Step 2 - *Rescaling virtual species distributions*

To be able to test approaches based on different data resolutions, we rescaled our virtual species distribution maps, allowing a comparison in later steps. This was done by aggregating species presence-absence (PA) data, available at fine-grain (1 km^2), to a coarser resolution of 10 km^2. We considered that if the species is present in one of the 1 km^2 pixels, then that corresponding 10 km^2 pixel (contains 100 pixels with fine resolution) necessarily also contains the species.

Step 3 - *Sampling species data*

Because we wanted to test the different modelling strategies using either PA or PO data, we created three sets of species data (i.e. PA, POequal and PO10M). First, we randomly sampled PA data, and since we are dealing with virtual species, we considered our sampling strategy as a "best case scenario", where complete and know data is available. The sampled points (n=1000) were independently selected for each species and strategy (i.e. different sampled points were used in the different strategies), while also considering the species true prevalence. We sampled the species' true prevalence because it is possible with virtual species and was previously shown to provide more accurate assemblage predictions when compared with other methods (Fernandes *et al.*, 2018). Second, we sampled PO data while also considering the species true prevalence when sampling the presences (POequal). This means that the number of sampled presences was the same as in the previous method, but instead of true absences, three random sets of pseudo-absences (with the same number of available presences) were selected. We selected only three sets of pseudo-absences due to highly intensive computational requirements necessary to run the models. Third, we sampled PO data, with the same number of presences as in POequal but with three random sets of 10.000 pseudo-absences being selected (PO10M). This two pseudo-absence sampling methods were used because it was previously recommended to apply POequal when using classification techniques (e.g. BRT or RF) and PO10M when using regression techniques (e.g. GLM or GAM) (see Barbet-Massin *et al.*, 2012 for details).

Step 4 - *Testing modelling* strategies

We then used the sampled datasets to calibrate different types of modeling approaches, based on different combinations of grain and extents, always using the same number of presences for each species, independently of calibrating models with PA or PO data. We fitted SDMs for all the species using five environmental variables (which may or not be the same that were used to generate the virtual species): annual mean temperature (° C), temperature seasonality (C of V), the mean temperature of the warmest quarter (° C), precipitation seasonality (C of V) and precipitation of warmest quarter (mm). Models were fitted with five techniques in an ensemble forecasting approach, using *biomod2* default options (Thuiller *et al.*, 2009), as in most published studies: generalized linear models (GLM; McCullagh & Nelder, 1989), generalized additive models (GAM; Hastie & Tibshirani, 1990), artificial neural networks (ANN; Venables & Ripley, 2002), boosted regression trees (BRT; Friedman *et al.*, 2000) and random forests (RF; Breiman, 2001). A repeated split-sample procedure was used (N=5) for model evaluation, followed by a weighted (TSS) ensemble forecast across techniques and repetitions. To compare them with the original presence-absence maps, the resulting probabilistic predictions (i.e. for the present and future) were transformed into presence-absence data (binary variable) using the threshold that maximizes the True Skill Statistic (MaxTSS; equivalent to the sensitivity-specificity sum maximization described in Liu *et al.*, 2005). This corresponds to the default procedure for ensemble modelling when using *biomod2* (Marmion *et al.*, 2009b).

When fitting models that used PO data, we randomly sampled pseudo-absences reflecting either the same number of available presences (POequal) or 10.000 pseudo-absences (PO10M). When used, pseudo-absence data had the same weight as presence data (Barbet-Massin *et al.*, 2012), except for the Hierarchical models (see explanation below). This procedure was repeated three times and the described model parameterizations were applied to the seven different strategies, to predict our virtual species' current and future distributions (2070):

Downscaling Model - DwSc: This method allows the calibration of models using coarse global species data (e.g. 10 or 50 km^2, usually obtained in herbaria, museums or species atlas), while making predictions to areas where environmental data is available at finer resolutions (Araújo *et al.*, 2005). In our case, we used global virtual species data available at 10 km^2, to fit models and predict the distribution of our virtual species at the European continental scale, and at a finer resolution of 1 km^2.

Fine Model - FnSc: In the case where global fine species data is available, it is possible to calibrate the models directly at the fine resolution (unlike in the DwSc model), using the same high-resolution environmental data to predict the species' current and future distributions. We used global data available at 1 km^2 to calibrate the models and predict the species distribution in Europe at the same resolution.

Truncated Downscaling Model - TDwSc: It often happens that species data is unavailable for the entire world (data available only for Europe or Switzerland for example). It might also occur that the data is accessible only at a coarse resolution. In these cases, one might employ a methodology like the one presented in the DwSc model, with the possibility of encountering the problem of climatic niche truncation. Therefore, to assess the accuracy of those models, we used species data available for Europe at 10 km^2 to calibrate the models and predict species distributions also in Europe at a higher resolution (1 km^2). This is the most common case observed in SDM studies, where one tries to make climate change projections using a local or regional study area, without the complete information of the species niche (e.g. Engler *et al.*, 2004; Sánchez-Fernández *et al.*, 2011; Vicente *et al.*, 2013).

Truncated Fine Model - TFnSc: This model is based on the same reasoning of the TDwSc model, but with high-resolution data (1 km^2) being used to calibrate the models.

Mixed Model - Mxd: In our new proposed modelling strategy, we used coarse-grain species data available at 10 km^2 to randomly sample fine-grain environmental data (1 km^2; 3 random points selected in each). This selection was done for the entire world, except for the locations where we want to predict the distribution of the species, and where high-resolution data is available (Europe in our case). This sampled data was then added to sampled fine-grain data available at 1 km^2 for Europe. The distributions were finally projected at that fine resolution in Europe. This allowed the use of data available at different resolutions, when available at fine-grain in a certain area and coarse-grain at another, while theoretically giving broader information about the species' full niche.

Hierarchical Model - Hier: While traditionally pseudo-absences have the same weight as presences, this approach doesn't assume that all pseudo-absences represent real absences (see Gallien *et al.*, 2012; Petitpierre *et al.*, 2016). To create our hierarchical models (either for POequal and PO10M), we started by creating a global model using coarse-grain presence-only data (10 km^2), predicting the distribution of our virtual species on the continental scale (Europe). The global model projections were used to determine where the species is predicted as absent and those locations were used to sample new pseudo-absences in Europe at 1 km^2 resolution. Pseudo-absence were selected inversely proportionally to the suitability projected by the global model, and the weight was calculated according to a linear weight (Hier1; Petitpierre *et al.*, 2016) or via an inverse logistic transformation (Hier2; Gallien *et al.*, 2012). These new weighted pseudo-absences, together with presence data sampled at the same resolution, were used to calibrate local/continental models (Hier1 and 2), predicting the distribution of the species for Europe at 1 km^2.

$$Hier1: \qquad Weight\ (x) = 1 - \frac{projGlobal(x)}{1000}$$

$$Hier2: \qquad Weight\ (x) = \frac{1}{1 + (\frac{projGlobal(x)}{projGlobal(x) - 1})^2}$$

Step 5 - *Evaluation and comparison of modelling* strategies

We evaluated and compared the accuracy of the different modelling strategies, considering their species richness (SR) patterns by measuring different assembly metrics. These metrics were calculated using the *ecospat.SSDMeval* function available in the "*ecospat*" R package (see Appendix S1 for details on all the indices; Di Cola *et al.*, 2017): (i) species richness error (i.e. difference between predicted and observed species richness); (ii) the assemblage prediction success (i.e. proportion of species correctly predicted as present or absent; values close to 1 indicate that species are correctly identified); and two widely used metrics of assemblages similarity, (iii) the Sørensen index (Sørensen, 1948) and (iv) the Jaccard index (Jaccard, 1912). When the values of Sørensen and Jaccard are close to 1, this indicates that observed and predicted assemblages are similar. This was done by sampling one set of points (one million, across all Europe) at 1 km^2 resolution for each species/approach, which could then be used to compare with our "known" current and future distributions.

RESULTS

Number of modelled species in each modelling strategy

The created virtual species presented prevalence values ranging between 0.006 and 0.997 (Mean ± StDv; 0.55 ± 0.34) under current environmental conditions, varying between 0.001 and 0.998 (0.51 ± 0.35) under future climatic conditions. The number of successfully modelled species was constant among the different approaches (n=100), independently of the use of PA/PO data (Table S2, Appendix S1). However, when using PA data, the TDwSc model (truncated downscaling model) was only able to model around half of the species (only 53 species modelled). The reason that almost half of the species were not modelled in TDwSc model has to do with the fact that when the upscaling procedure is done (i.e. increasing the species' distribution resolution from 1km^2 to 10 km^2), the species with higher prevalence (in the truncated area) lose absence information. The model is therefore unable to be fitted because not enough absences are available. Because of this, it is important to note that all analyses done for the TDwSc model when using PA data only account for a total species richness of 53 and must be compared to other approaches taking that fact into consideration.

Evaluation of individual models in the different modelling strategies

We observed that when models were fitted using PA data (Figure 2; solid lines), all the strategies presented individual models with very high evaluation values (MaxTSS around 0.9). When analysing the models fitted with PO data and the same number of presences/pseudo-absences (POequal; Figure 2; dashed lines), their evaluation values were much lower, with MaxTSS values around 0.5, except for the Mixed model –Mxd– that presented values above 0.5. Finally, models fitted using PO data but with random sets of 10.000 pseudo-absences (PO10M; Figure 2, dotted lines) presented MaxTSS values above those observed with POequal (between 0.7-0.8), but still below those observed when PA data was used (i.e. around 0.9). Both hierarchical modelling strategies presented similar values of MaxTSS to the ones observed in the other techniques (around 0.5 for POequal and 0.7-0.8 in PO10M), with the Mixed model presenting MaxTSS values around 0.55 (POequal) and 0.85 (PO10M).

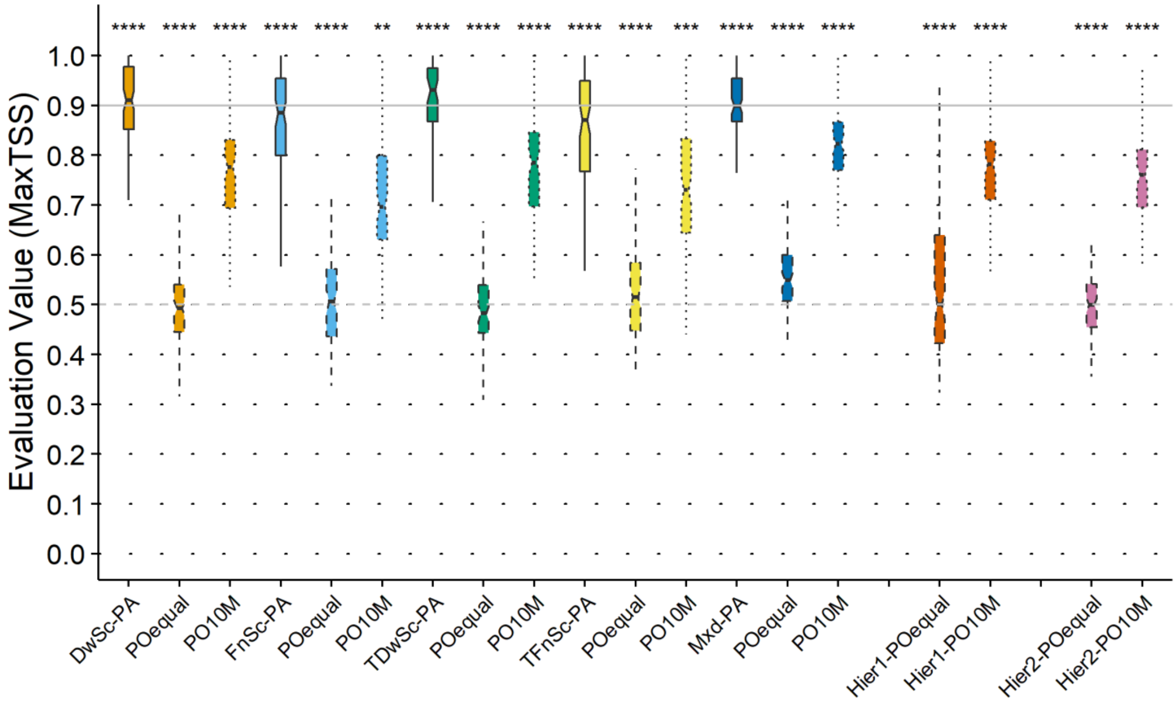


FIGURE 2 Evaluation values of ensemble models measured using maximized TSS (MaxTSS; measured in the training data) for each modelled virtual species (n=100) and all modelling strategies (DwSc-Downscaling model; FnSc-Fine model; TDwSc- Truncated Downscaling model; TFnSc- Truncated Downscaling model; Mxd- Mixed model; Hier1- Hierarchical model 1; Hier2- Hierarchical model 2). For each modelling strategy, three sets of data were randomly sampled to fit the models: presence/absence data (PA), a set with an equal number of presence/pseudo-absence data (POequal) and presence/pseudo-absence data with a random set of 10.000 pseudo-absences (PO10M). Note that all the modelling techniques were calibrated using PA data (and POequal and Po10M) except for the hierarchical models (only POequal and PO10M) and that in the results of TDwSc-PA only 53 species were considered in the analysis.

Community prediction accuracy in presence/absence (PA) models

When analysing the predictive accuracy of the modelling strategies fitted using PA data (Figure 3), we observed that binary S-SDMs based on the downscaling approaches - DwSc and TDwSc models - showed higher predicted SR than what is observed in the original assemblages, with the caveat that the analysis for TDwSc only accounts for a total species richness equal to 53 (Fig. 3a; SR error). On average, the overprediction of SR was higher than ten species per 1 km^2, increasing under future climatic conditions in the DwSc model and decreasing in the TDwSc model. The remaining models – FnSc, TFnSc and Mxd - presented SR values relatively like the ones observed in the original assemblages, only slightly over- (FnSc; 3 species on average) or under-predicting SR (TFnSc and Mxd; minus 4 species on average). For these three models, an over-prediction of SR under future climatic conditions occurs, but still lower than the one observed in the DwSc and TDwSc models (i.e. less than ten species on average).

Considering the other three calculated metrics (i.e. prediction success, Sørensen and Jaccard indices), a clear pattern was observed, where model predictions made under current conditions showed higher values than future predictions (Fig. 3b, c, d). The fine scale model -FnSc- presented the highest values of all the bS-SDMs models, with prediction success and Sørensen around 0.9 (under current conditions). The remaining models presented values (of prediction success and Sørensen) between 0.8-0.9 (i.e. indicating high similarity between observed and predicted assemblages), except the TDwSc model, with values around 0.7-0.8 (for the 53 modelled species) for both current and future conditions. Jaccard values were lower for all the tested modelling strategies, but the patterns observed before were still present (i.e. higher values for FnSc model and lower for TDwSc model).

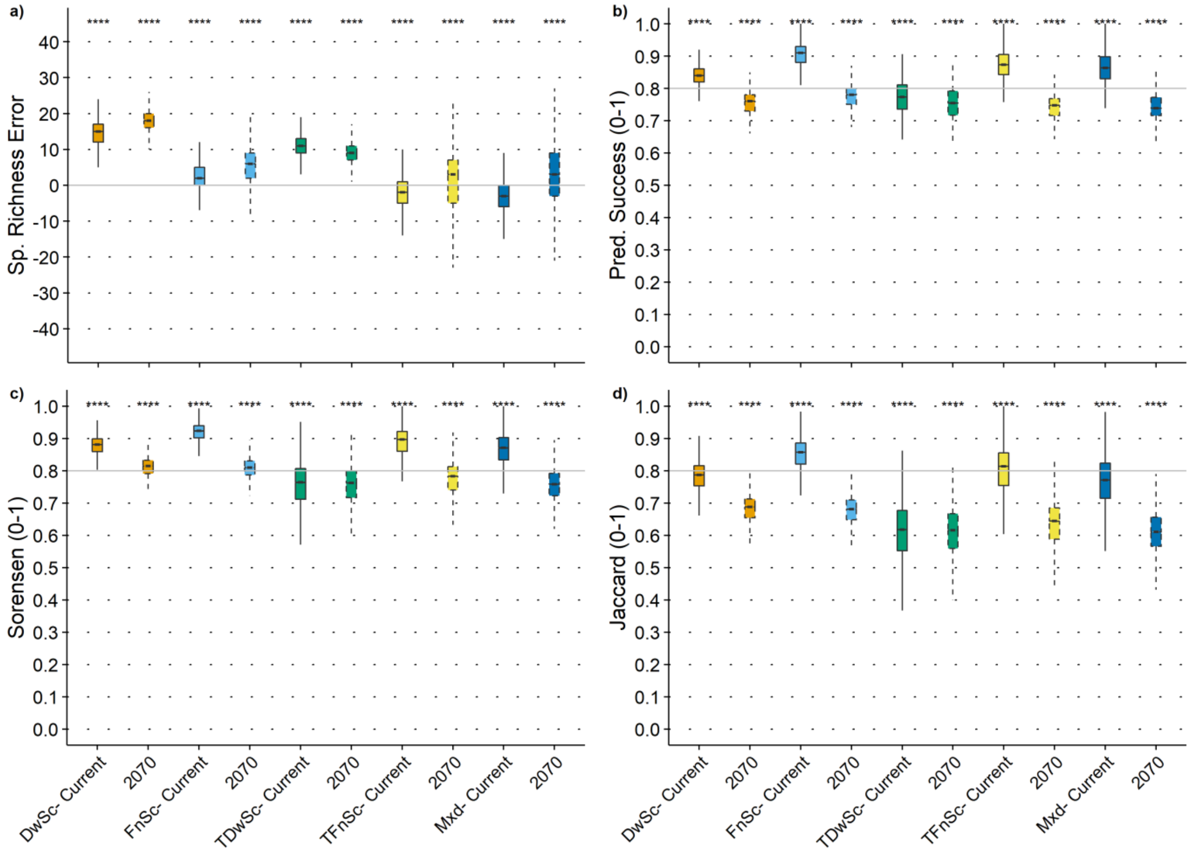


FIGURE 3 Boxplots of different indices of assemblage predictions (bS-SDM) accuracy (i.e. species richness error, prediction success, Sørensen and Jaccard similarity indices), for all the virtual species and modelling strategy (DwSc-Downscaling model; FnSc-Fine model; TDwSc- Truncated Downscaling model; TFnSc- Truncated Downscaling model; Mxd- Mixed model) under current and future climatic conditions. Each box shows the variation across all virtual species in a random subset of the study area (Europe, one million points) for the binary predictions obtained when using presence/absence data (PA) to calibrate the models and MaxKappa as thresholding technique. Note that in the results of TDwSc only 53 species were considered in the analysis.

Community prediction accuracy in presence-only (PO) models

The predictive accuracy of modelling strategies when PO data was used was substantially reduced (Figure 4 and 5) when compared with the same approaches using PA data (Figure 3). When considering strategies that used an equal number of presences/pseudo-absences (Figure 4; POequal), we observed that all the approaches predict lower species richness than occur per plot (1 km^2) in the original assemblages. On average, the different approaches predicted around 30 species less than what was initially observed (Fig. 4a), both under current and future conditions. The Mixed model is the approach with the lower difference between observed and predicted species richness, around 20 species less than originally observed, staying similar under future climatic conditions. This is also reflected in the other three metrics, where the Mixed model presented the higher values of prediction success, Sørensen and Jaccard, on average (i.e. close to 0.7 for the first two metrics, around 0.5 for Jaccard). However, all strategies had values of prediction success and Sørensen below 0.7 (Fig. 4b and c; values close to 1 indicate very high similarity between observed and predicted assemblages, and close to zero a very high dissimilarity). The values for Jaccard index were even lower, below 0.5 for predictions under current conditions and 0.4 under future conditions (Fig. 4d). The values observed for the two hierarchical modelling strategies (Hier1 and 2) were like those observed in the other approaches.

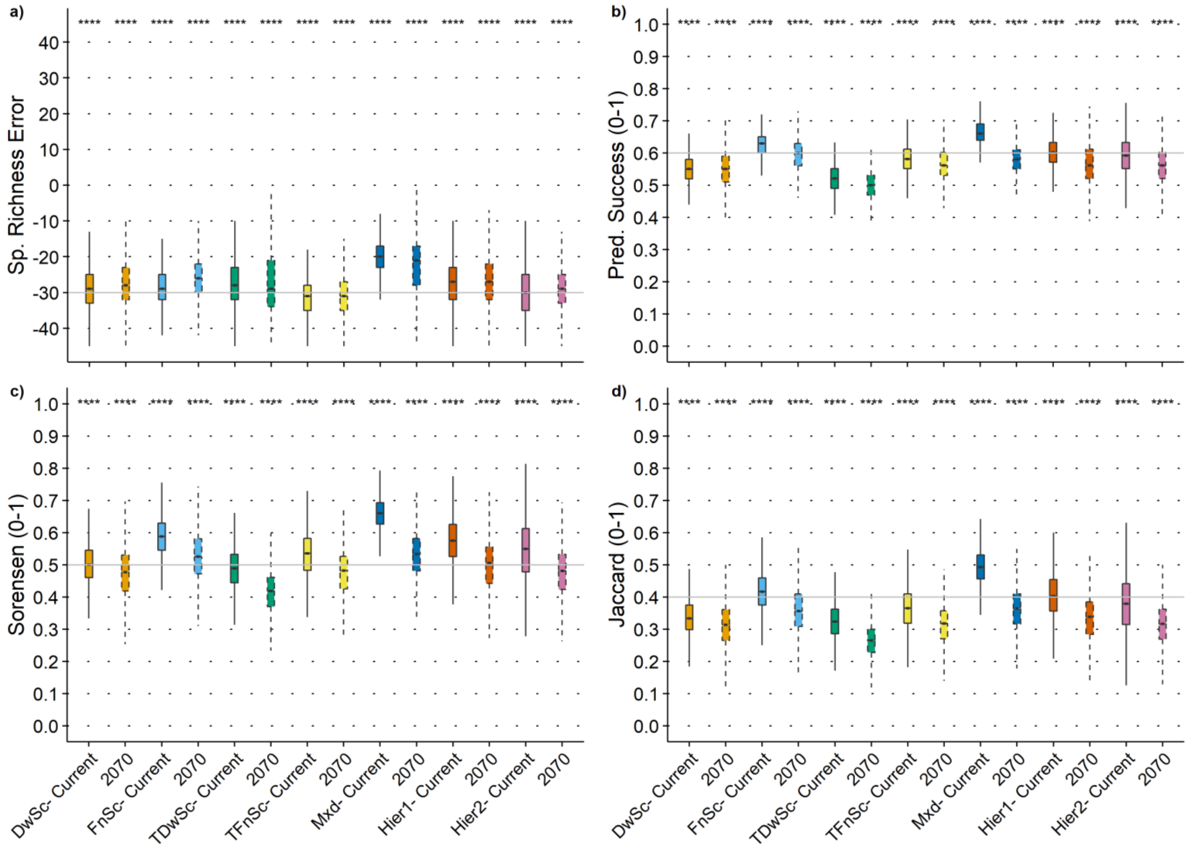


FIGURE 4 Boxplots of different indices of assemblage predictions (bS-SDM) accuracy (i.e. species richness error, prediction success, Sørensen and Jaccard similarity indices), for all the virtual species and all modelling strategy (DwSc-Downscaling model; FnSc-Fine model; TDwSc- Truncated Downscaling model; TFnSc- Truncated Downscaling model; Mxd- Mixed model; Hier1- Hierarchical model 1; Hier2- Hierarchical model 2) under current and future climatic conditions. Each box shows the variation across all virtual species in a random subset of the study area (Europe, one million points) for the binary predictions obtained when using an equal number of presence/pseudo-absence data (POequal) to calibrate the models and MaxKappa as thresholding technique.

If we analyse the patterns obtained when using PO data but with 10.000 randomly sampled pseudo-absences (Figure 5; PO10M), we observed that the values for all the calculated metrics were even lower than when POequal data was used (on average below 0.5 when using PO10M data contrary to almost always above 0.5 with POequal data) . However, we also observed that the metrics' values are higher under future predictions than under current climatic conditions (i.e. predicting higher future similarity than current one; increase around 0.2 in Sørensen, but still below 0.6 on average).

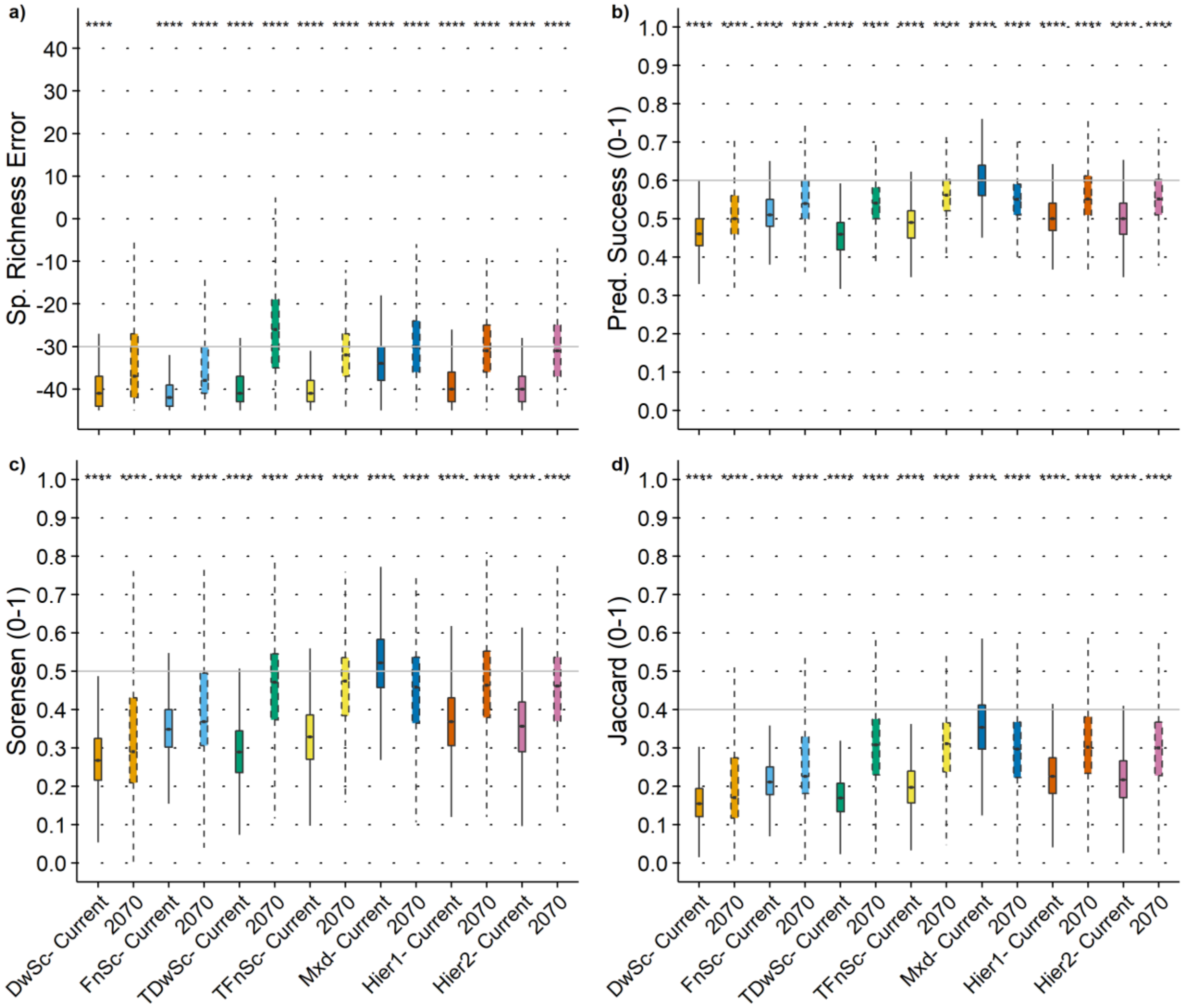


FIGURE 5 Boxplots of different indices of assemblage predictions (bS-SDM) accuracy (i.e. species richness error, prediction success, Sørensen and Jaccard similarity indices), for all the virtual species and all modelling strategy (DwSc-Downscaling model; FnSc-Fine model; TDwSc- Truncated Downscaling model; TFnSc- Truncated Downscaling model; Mxd- Mixed model; Hier1- Hierarchical model 1; Hier2- Hierarchical model 2) under current and future climatic conditions. Each box shows the variation across all virtual species in a random subset of the study area (Europe, one million points) for the binary predictions obtained when using presence/pseudo-absence data with a random set of 10.000 pseudo-absences (PO10M) to calibrate the models and MaxKappa as thresholding technique.

DISCUSSION

In this work, we used a virtual ecologist approach to assess if different modelling approaches – and especially hierarchical ones - can provide accurate assemblage predictions under current and future climatic conditions (i.e. assessing transferability), therefore preventing limitations of real-world data and allowing a complete knowledge of the full species distribution and assemblage composition. We found that hierarchical models, especially those fitted using the PO10M method, presented evaluation values that are usually considered as models with good prediction accuracy (MaxTSS> 0.7). However, this was not reflected on the similarity of predicted assemblages (Sørensen ≤ 0.6), indicating that further improvements need to made before using their predictions to make conservation decisions or anticipate future invasions (e.g. Petitpierre *et al.*, 2016). We also found that models calibrated with presence/absence (PA) data presented, in general, high prediction success (above 0.7) and similarity between observed and predicted assemblages (i.e. Sørensen/Jaccard ≥ 0.8). However, this prediction success observed when PA data was used, gets reduced when projections were made into future climatic conditions (Sørensen ≤ 0.8). Modelling approaches using initial coarse PA data (i.e. DwSc) overpredicted species richness, with alternatives like a hierarchical Bayesian modelling framework (Keil *et al.*, 2013) being an interesting alternative. Our new proposed mixed strategy (Mxd) also presented interesting results when used with PA data (i.e. high evaluation values and similarity between observed and future assemblages under current conditions), but potential improvements are discussed below. Additionally, we found that the overall accuracy that can be expected from stacked SDMs (S-SDMs) when using presence/pseudo-absence data was lower than when PA data was used (i.e. lower transferability; Sørensen ≤ 0.6). This was something already observed and discussed for single species models, with authors suggesting that methods using PA data are preferable (e.g. Brotons *et al.*, 2004; Graham & Hijmans, 2006; Lobo *et al.*, 2010). The method of selecting pseudo-absences is also known to influence prediction accuracy (e.g. Chefaoui & Lobo, 2008; Barbet-Massin *et al.*, 2012), and that pattern was also evident in our work, with models fitted using the POequal method showing higher assemblage similarity (Sørensen) than models fitted using the PO10M method. We also observed that while presenting high evaluation values (MaxTSS>0.7) when using the total species pool (n=100), all the approaches greatly under-predict species richness when PO data was used (around 30 species). Finally, and contrary to what we were expecting, hierarchical models produced very similar predictions when compared with the other PA modelling strategies.

Previous virtual species studies considering the validity of models projections/transferability focused on the effects of model complexity (Bell & Schlaepfer, 2016; García-Callejas & Araújo, 2016), on testing the accuracy of different statistical models (Meynard & Quinn, 2007; Elith & Graham, 2009; Qiao *et al.*, 2018) or on proposing new algorithms that increase transferability in presence-only models (Qiao *et al.*, 2015). Other studies also used virtual species to test the best sampling strategies to fit distribution models (Hirzel & Guisan, 2002), to assess how to define pseudo-absences (Wisz & Guisan, 2009), to compare different downscaling approaches (Bombi & D'Amen, 2012), to assess the relative effects of factors affecting species distribution models (Thibaud et al. 2014) or to use virtual predictors to study the importance of variable selection and evaluation metrics (Fourcade *et al.*, 2018), but none assessed yet model transferability. Different hierarchical or multi-scale approaches have also been proposed and tested (e.g. Pearson *et al.*, 2004; Gallien *et al.*, 2012; Bastos *et al.*, 2016; Talluto *et al.*, 2016), though all these studies used real species.

To the best of our knowledge, this study was the first to use a virtual ecology framework to compare seven different modelling strategies while using a high number of virtual species (n=100) to assess the predictive accuracy and transferability when predicting the distribution of species assemblages (not only single species distributions) under current and future climatic conditions (using either PA or PO data) and at multiple scales (varying resolution and extent). Below, we discuss our findings in additional detail and explain their importance while also critically evaluating our approach and detailing further improvements.

Do the different modelling strategies give useful assemblage predictions when using presence-absence (PA) data?

Under current climatic conditions, we were able to predict reasonably accurate assemblages (i.e. Sørensen similarity index above 0.7). Despite that fact, not all the modelling approaches predicted the correct species richness (i.e. difference between predicted and observed species richness; e.g. DwSc and TDwSc models). Downscaling methods are known to produce results with mixed quality (Bombi & D'Amen, 2012; Keil *et al.*, 2013), and especially the approach employed here, often called the direct approach (e.g. Araújo *et al.*, 2005; McPherson *et al.*, 2006). This approach tends to be problematic because it assumes that species distribution at fine resolutions have the same environmental associations as the distributions at coarser resolutions (Keil *et al.*, 2013). Despite this, even if it overpredicts species richness, the overall similarity between predicted and observed assemblages is close to the values presented in strategies that used fine resolution data (FnSc). In a relatively recent study, Keil *et al.* (2013) showed that a hierarchical Bayesian downscaling approach could be used to provide improved predictions of fine scale species distributions when compared with the downscaling direct approach. Therefore, this Bayesian approach needs to be further tested, especially using virtual species.

As we expected, the predictions obtained through the Fine scale approach (FnSc) presented the highest similarity and lowest species richness error of all the strategies. We expected this because in the FnSc strategy, data was sampled across the entire climatic conditions and at fine resolution. Nonetheless, and surprisingly, this approach, like all the others, was unable to correctly predict species assemblages under future climatic conditions. It has previously been shown that even when complete information about the species and the factors influencing their distributions is initially known, when sampling a limited number of observation sites from the initial true distribution, one can hardly obtain exactly similar observed/predicted assemblages, and that Sørensen values above 0.8 can already be considered as acceptable (Fernandes *et al.*, 2018). However, a possible explanation why the FnSc strategy was unable to completely predict accurate future assemblages may lie in the fact that in this study, the environmental variables known to influence the distribution of the species (i.e. the variables used to create the virtual species) might not be the ones finally used to fit the models. This is something that also occurs in the real word, where usually several species are modelled using the same predictors. While this limitation might reduce the power and potential of using virtual species with the complete species' knowledge, at least it replicates a real world situation. In any case, future studies could overcome this problem by constraining the model fitting to use the same initial predictors (as in chapter 3), and approaches to select predictors that maximize the transferability of SDMs have been proposed (Petitpierre *et al.*, 2017), and it would be interesting to include these approaches in our proposed framework.

Finally, because irregular or incorrect species distributions can occur due to climatic truncation (Normand *et al.*, 2009; Hannemann *et al.*, 2016), it was expected that the models using truncated climatic data would perform worse than other models, especially if no-analogue climates occur under future conditions (Williams & Jackson, 2007). Contrary to what was expected (i.e. reduced accuracy of assemblage predictions), the truncated fine scale models (TFnSc) presented similar assemblage prediction success and similarity when compared with the FnSc models (non-truncated strategy), but like all the other approaches was unable to accurately predict future changes. One possible explanation for the fact that FnSc and TFnSc approaches produced similar results might be because all our species were still well represented in the truncated region (Europe) and that the region might not have enough truncation to be relevant for comparison with the fine scale strategy. Therefore, the effect of the truncation should be tested using a region smaller than Europe to assess the TFnSc limitations.

Differences in prediction accuracy when using presence-only (PO) data

We found in this study that all the modelling approaches using PO data were unable to accurately predict species assemblages under both current and future climatic conditions (Sørensen below 0.6; high similarity is usually close to 1). Another important observed pattern was the fact that even if POequal models presented lower evaluation values (MaxTSS between 0.5-0.6) than PO10M models (MaxTSS ≥ 0.7), the predicted assemblages of the former presented higher similarity with observed assemblages (Sørensen between 0.5-06) than the latter (Sørensen between 0.25 and 5). A possible explanation for the fact that the single species model evaluation values of PO10M were high (MaxTSS>0.7) but the observed/predicted assemblage similarity (i.e. measured by Sørensen and Jaccard) was low, might be because the sampled presences might cover a range nearly as large as the pseudo-absences, making the model unfit to find a good signal for the presences. Also, species prevalence (e.g. Engler *et al.*, 2004; McPherson *et al.*, 2004; Hanberry & He, 2013) is known to influence the accuracy of PO models, with a limited number of pseudo-absences usually providing a low fit. Another possible explanation for the poor PO models performance can be either due to the choice of the number of pseudo-absences, how the pseudo-absences were selected or the method used to weighting them, something also shown to affect PO model calibration and predictions (e.g. Elith & Graham, 2009; Ward *et al.*, 2009; Wisz & Guisan, 2009; Barbet-Massin *et al.*, 2012). These factors have also been shown to differently affect the discrimination and calibration of models depending on the modelling technique used (Elith & Graham, 2009; Barbet-Massin *et al.*, 2012). In our work, we sampled random pseudo-absences across the whole study area, something done before (e.g. Peterson *et al.*, 2007) but criticized because it might use pseudo-absences from both the training and evaluation regions combined, which is not advised under climate change applications (Phillips, 2008). However, this factor needs to be further investigated, since in both hierarchical models (Hier1 and Hier2), our pseudo-absences were selected only on the ecoregions covered by the species distribution (see Petitpierre *et al.*, 2016), and the models still failed to correctly predict species assemblages. We also used different numbers of pseudo-absences, following the suggestions of Barbet-Massin *et al.* (2012). In that study, the authors suggested using the same number of pseudo-absences as available presences (i.e. POequal) when using techniques like boosted regression trees or random forests or a large number of pseudo-absences (10.000) with equal weighting for presences and absences (i.e. PO10M) when techniques like generalized linear or additive models were used. Since we used an ensemble model with a combination of regression (i.e. GLM and GAM) and classification techniques (i.e. BRT and RF), this could explain why model predictions were inaccurate for most of the species. It might be useful to further develop our proposed framework testing the modelling approaches using individual modelling techniques, perfectly tuned to what was previously suggested to work best (e.g. Wisz & Guisan, 2009; Barbet-Massin *et al.*, 2012).

In any case, the accuracy of PO models was studied exhaustively in the past, with mixed results being obtained. While some authors found that PA data is preferable (e.g. Brotons *et al.*, 2004; Graham & Hijmans, 2006; Wisz & Guisan, 2009), others reported that PO models still produced good results, but not necessarily better than PA models (e.g. Elith *et al.*, 2006; Petitpierre *et al.*, 2016). These partially ambiguous results enforce the need to thoroughly testing the different factors affecting the performance of PO models, both for individual or nested factor, being artificial data a useful solution to help determine the methods that work and the ones that don’t.

Are hierarchical models useful in predicting species assemblages?

Contrary to what was observed in studies that used correlative SDMs to hierarchically model single species distributions (e.g. Pearson *et al.*, 2004; Gallien *et al.*, 2012; Petitpierre *et al.*, 2016), the two hierarchical modelling approaches applied here were unable to predict accurate species assemblages (under current and future climatic conditions; Sørensen and Jaccard similarity indices below 0.6 on average, for both PO models). While this might be due to factors already discussed in the previous section (e.g. sampling or weighting of pseudo-absences), we would like to discuss three additional factors. First, the global model used in the hierarchical approach is based on coarse presence-only data. If our FnSc model was unable to predict accurate species distribution using fine scale information, we can assume that the coarse scale model would predict even worse distributions, potentially biasing the locations where pseudo-absences were selected in the local model. A good alternative to the use of coarse-scale models as the global model might be our proposed Mixed approach, using those predictions to further improve the hierarchical model. Although the Mixed models still failed to predict completely accurate species assemblages, its predictions presented the best results of all the tested approaches (using PO data) and further improvements to the methodology can prove useful (see following section).

Second, the global (coarse) models used in the first stage of the hierarchical modelling strategy are generally fitted with climatic predictors while the finer scale model (i.e. local model; second stage of the hierarchical strategy) other predictors are also included (e.g. land-use data, human disturbance, distance to roads or rivers; Petitpierre *et al.*, 2016). In our study and because we know that our virtual species are only determined by climatic data, this second data dimension was not included. Therefore, additional work needs to be performed taking this into consideration, creating virtual species that are also influenced by factors like land-use or human disturbance. Finally, we believe this strategy should be tested by e.g. giving different weights to the presences instead of to pseudo-absences or complementing the fine-resolution presences in the local model with coarse resolution ones from the global model.

Conclusions, limitations and perspectives

In this paper, we used virtual species to compare how accurate different modelling approaches were at predicting current and future distributions of species assemblages, using either presence-absence (PA) or presence/pseudo-absences (PO) data to fit the models. We observed that PA models can predict accurate current species assemblages but failed to predict equally accurate future assemblages (i.e. high values of Sørensen/Jaccard). On the other hand, PO models failed to correctly predict both current and future assemblages. This fact indicates that additional work is still needed to improve the predictions of PO models when the goal is to predict the spatial distribution of species assemblages, and that virtual species can be a useful tool to support those improvements. Our study was, to the best of our knowledge, the first one to use a virtual ecology framework in combination with a large set of virtual species (n=100, with different prevalences), to assess the predictive accuracy and transferability of seven different strategies modelling to model species assemblages (i.e.community modelling). We already discussed different aspects, like transferability, modelling strategy, model calibration using either PA or PO data, the type of PO data sampled (i.e. POequal and PO10M). However, important questions remain that should also be raised. Here, based on our results, we discuss some limitations to our work and possible future improvements:

1) Our virtual species were created using random environmental variables, but the models were fitted using the same set of predictors for each species. While this reflects what is usually done in

practice, with the same set of predictors being used to predict the distribution of several species at a time (e.g. Dubuis *et al.*, 2013; D'Amen *et al.*, 2015b; Petitpierre *et al.*, 2016), the different modelling strategies should also be tested with the species being fitted with the same predictors used to generate them. This will increase the power of using virtual species and remove a potential source of bias and errors in our framework.

2) We used two different methods to sample PO data (i.e. POequal and PO10M). This was done because one method was recommended (Barbet-Massin *et al.*, 2012)to be used with classification trees (POequal; with BRT or RF) and the other with regression techniques like GLM or GAM (PO10M). However, here we used an ensemble (combination) regression techniques and classification trees, potentially biasing the predictions. Therefore, our framework should be further tested using individual modelling techniques that are better suited for each type of selected PO sampled data.
3) The prediction accuracy of PO models is also influenced by species prevalence (e.g. Engler *et al.*, 2004; McPherson *et al.*, 2004; Hanberry & He, 2013) and since in this study we simulated species with different prevalences (which is then reflected in the training data), we recommend the assessment of the modelling strategies with the same number of presences for all species, when using PO data.
4) Additional work should also include the assessment and improvement of methods focusing on Bayesian modelling (e.g. Keil *et al.*, 2013) or multi-scale approaches (e.g. Bastos *et al.*, 2016; Talluto *et al.*, 2016), comparing those methods with the approaches tested in this work. Also, instead of the weighting procedure used in the hierarchical models, an interesting idea could be to use the global model to directly correct the signal (i.e. response curve) driven by presences (not pseudo-absences as currently done).
5) Caution also needs to be taken in the interpretation of the models' evaluation, since wrong conclusions can be made depending on the metric or data used (e.g. Fourcade *et al.*, 2018; Fernandes *et al.*, in press). This was something particularly evident in PO models, with evaluation values being high (MaxTSS >0.7) but assemblages being incorrectly predicted.
6) Finally, our proposed Mixed approach presented promising results in the prediction of assemblages when using PO data and when the global information about the species niche is only available at coarse resolutions. However, because we assumed that all sampled areas inside the coarse-scale presence pixels contain suitable conditions for the species, potential bias might occur. This means that a 10 km^2 pixel can contain several 1 km^2 pixels without ideal conditions for a species to survive. Further improvements and tests are necessary to fully assess the value of this approach, like combining global coarse data with fine local data directly, avoiding potential bias that might occur when fine environmental data is sampled.

ACKNOWLEDGEMENTS

This work was supported by the SNF project “SESAM'ALP - Challenges in simulating alpine species assemblages under global change” (nr 31003A-1528661). The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics.

References

Albert, C.H., Yoccoz, N.G., Edwards, T.C., Graham, C.H., Zimmermann, N.E. & Thuiller, W. (2010) Sampling in ecology and evolution - bridging the gap between theory and practice. *Ecography*, **33**, 1028-1037.

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

Araujo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species-climate impact models under climate change. *Global Change Biology*, **11**, 1504-1513.

Araújo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2005) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, **14**, 17-30.

Araújo, M.B.H., C.J.; Densham, P.J.; Lampinen, R.; Hagemeijer, W.J.M.; Mitchell-Jones, A.J.; Gasc, J.P. (2001) Would environmental diversity be a good surrogate for species diversity? . *Ecography*, **24**, 103-110.

Austin, M. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1-19.

Austin, M.P. (1987) Models for the analysis of species' response to environmental gradients. *Vegetatio*, **69**, 35-45.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.

Austin, M.P. & Gaywood, M.J. (1994) Current Problems of Environmental Gradients and Species Response Curves in Relation to Continuum Theory. *Journal of Vegetation Science*, **5**, 473-482.

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. & Luoto, M. (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, **199**, 197-216.

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? *Ecography*, **33**, 878-886.

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

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B. & Ferrer, E.A. (2011) Has the Earth's sixth mass extinction already arrived? *Nature*, **471**, 51-7.

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

Bastos, R., D’Amen, M., Vicente, J., Santos, M., Yu, H., Eitelberg, D., Gonçalves, J., Civantos, E., Honrado, J. & Cabral, J.A. (2016) A multi-scale looping approach to predict spatially dynamic patterns of functional species richness in changing landscapes. *Ecological Indicators*, **64**, 92-104.

Beale, C.M. & Lennon, J.J. (2012) Incorporating uncertainty in predictive species distribution modelling. *Philos Trans R Soc Lond B Biol Sci*, **367**, 247-58.

Bell, D.M. & Schlaepfer, D.R. (2016) On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological Modelling*, **330**, 50-59.

Bombi, P. & D'Amen, M. (2012) Scaling down distribution maps from atlas data: a test of different approaches with virtual species. *Journal of Biogeography*, **39**, 640-651.

Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5-32.

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437-448.

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

Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014) Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99-112.

Cantor, S.B., Sun, C.C., Tortolero-Luna, G., Richards-Kortum, R. & Follen, M. (1999) A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *Journal of Clinical Epidemiology*, **52**, 885-892.

Carroll, S.S. & Pearson, D.L. (1998) The effects of scale and sample size on the accuracy of spatial predictions of tiger beetle (Cicindelidae) species richness. *Ecography*, **21**, 401-414.

Chefaoui, R.M. & Lobo, J.M. (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210**, 478-486.

Ciarleglio, M., Barnes, J.W. & Sarkar, S. (2009) ConsNet: new software for the selection of conservation area networks with spatial and multi-criteria analyses. *Ecography*, **32**, 205-209.

Cohen, J. (2016) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**, 37-46.

D'Amen, M., Pradervand, J.-N. & Guisan, A. (2015a) Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography*, **24**, 1443-1453.

D'Amen, M., Rahbek, C., Zimmermann, N.E. & Guisan, A. (2017) Spatial predictions at the community level: from current approaches to future frameworks. *Biol Rev Camb Philos Soc*, **92**, 169-187.

D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L. & Guisan, A. (2015b) Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *Journal of Biogeography*, **42**, 1255-1266.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science*, **332**, 53-8.

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. *Ecography*, **40**, 774-787.

Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Wu, G.-L. & Wang, Z.-G. (2015) SDMvspecies: a software for creating virtual species for species distribution modelling. *Ecography*, **38**, 108-110.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122-1131.

Dubuis, A., Rossier, L., Pottier, J., Pellissier, L., Vittoz, P. & Guisan, A. (2013) Predicting current and future spatial community patterns of plant functional traits. *Ecography*, **36**, 1158-1168.

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677-697.

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

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *J Anim Ecol*, **77**, 802-13.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endagered species from occurence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263-274.

Engler, R., Randin, C.F., Thuiller, W., Dullinger, S., Zimmermann, N.E., AraÚJo, M.B., Pearman, P.B., Le Lay, G., Piedallu, C., Albert, C.H., Choler, P., Coldea, G., De Lamo, X., DirnbÖCk, T., GÉGout, J.-C., GÓMez-GarcÍA, D., Grytnes, J.-A., Heegaard, E., HØIstad, F., NoguÉS-Bravo, D., Normand, S., PuŞCaŞ, M., SebastiÀ, M.-T., Stanisci, A., Theurillat, J.-P., Trivedi, M.R., Vittoz, P. & Guisan, A. (2011) 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology*, **17**, 2330-2341.

Fernandes, R.F., Scherrer, D. & Guisan , A. (2018) How much should one sample to accurately predict the distribution of species assemblages? A virtual community approach *Ecological Informatics*, **48**

Fernandes, R.F., Scherrer, D. & Guisan , A. (in press) Effects of simulated observation errors on the performance of species distribution models. *Diversity and Distributions*,

Fernandes, R.F., Scherrer, D. & Guisan , A. (unplub.) Predicting current and future virtual species assemblages: are hierarchical models useful?

Fernandes, R.F., Vicente, J.R., Georges, D., Alves, P., Thuiller, W. & Honrado, J.P. (2014) A novel downscaling approach to predict plant invasions and improve local conservation actions. *Biological Invasions*, **16**, 2577-2590.

Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393-404.

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

Fitzpatrick, M.C. & Hargrove, W.W. (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, **18**, 2255-2261.

Fitzpatrick, M.C., Blois, J.L., Williams, J.W., Nieto-Lugilde, D., Maguire, K.C. & Lorenz, D.J. (2018) How will climate novelty influence ecological forecasts? Using the Quaternary to assess future reliability. *Global change biology*,

Fourcade, Y., Besnard, A.G. & Secondi, J. (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, **27**, 245-256.

Franklin, J. (2013) Species distribution models in conservation biogeography: developments and challenges. *Diversity and Distributions*, **19**, 1217-1223.

Friedman, J., Hastie, T. & Tibshirani, R. (2000) Additive logistic regression:a statistical view of boosting. *The Annals of Statistics*, **28**, 337-407.

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? *Global Ecology and Biogeography*, **21**, 1126-1136.

García-Callejas, D. & Araújo, M.B. (2016) The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling*, **326**, 4-12.

Geisen, S., Mitchell, E.A., Wilkinson, D.M., Adl, S., Bonkowski, M., Brown, M.W., Fiore-Donno, A.M., Heger, T.J., Jassey, V.E. & Krashevska, V. (2017) Soil protistology rebooted: 30

fundamental questions to start with. *Soil Biology and Biochemistry*, **111**, 94-103.

Gotzenberger, L., de Bello, F., Brathen, K.A., Davison, J., Dubuis, A., Guisan, A., Leps, J., Lindborg, R., Moora, M., Partel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel, K. & Zobel, M. (2012) Ecological assembly rules in plant communities--approaches, patterns and prospects. *Biol Rev Camb Philos Soc*, **87**, 111-27.

Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, **15**, 578-587.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol Evol*, **19**, 497-503.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A. & Gro, N.P.S.W. (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**, 239-247.

Gu, W.D. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, **116**, 195-203.

Guélat, J. & Kéry, M. (2018) Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, **9**, 1614-1625.

Guillera-Arroita, G., Ridout, M.S. & Morgan, B.J.T. (2010) Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, **1**, 131-139.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Guisan, A. & Hofer, U. (2003) Predicting reptile distributions at the mesoscale: relation to climate and topography. *Journal of Biogeography*, **30**, 1233-1243.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Guisan, A. & Rahbek, C. (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433-1444.

Guisan, A., Theurillat, J.P. & Kienast, F. (1998) Predicting the potential distribution of plant species in an Alpine environment. *Journal of Vegetation Science*, **9**, 65-74.

Guisan, A., Edwards, T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89-100.

Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press.

Guisan, A., Graham, C.H., Elith, J. & Huettmann, F. (2007a) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332-340.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007b) What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615-630.

Guisande, C., García-Roselló, E., Heine, J., González-Dacosta, J., Vilas, L.G., García Pérez, B.J. & Lobo, J.M. (2017) SPEDInstabR: An algorithm based on a fluctuation index for selecting predictors in species distribution modeling. *Ecological Informatics*, **37**, 18-23.

Hanberry, B. & He, H. (2013) Prevalence, statistical thresholds, and accuracy assessment for species distribution models. *Web Ecology*, **13**, 13-19.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29-36.

Hannemann, H., Willis, K.J. & Macias-Fauria, M. (2016) The devil is in the detail: unstable response functions in species distribution models challenge bulk ensemble modelling. *Global Ecology* *and Biogeography*, n/a-n/a.

Harrell, F.E. (2015) *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer International Publishing.

Harrison, P., Berry, P., Butt, N. & New, M. (2006) Modelling climate change impacts on species' distributions at the European scale: implications for conservation policy. *environmental science & policy*, **9**, 116-128.

Hastie, T.J. & Tibshirani, R.J. (1990) *Generalized Additive Models*. Chapman & Hall, London.

Hattab, T., Garzón-López, C.X., Ewald, M., Skowronek, S., Aerts, R., Horen, H., Brasseur, B., Gallet-Moron, E., Spicher, F., Decocq, G., Feilhauer, H., Honnay, O., Kempeneers, P., Schmidtlein, S., Somers, B., Van De Kerchove, R., Rocchini, D. & Lenoir, J. (2017) A unified framework to model the potential and realized distributions of invasive species within the invaded range. *Diversity and Distributions*, **23**, 806-819.

Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773-785.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331-341.

Hirzel, A.H., Helfer, V. & Metral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145**, 111-121.

Huang, M., Kong, X., Varela, S. & Duan, R. (2016) The Niche Limitation Method (NicheLim), a new algorithm for generating virtual species to study biogeography. *Ecological Modelling*, **320**, 197-202.

Hudson, L.N., Newbold, T., Contu, S., Hill, S.L., Lysenko, I., De Palma, A., Phillips, H.R., Senior, R.A., Bennett, D.J. & Booth, H. (2014) The PREDICTS database: a global database of how local terrestrial biodiversity responds to human impacts. *Ecology and evolution*, **4**, 4701-4735.

Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography*, **22**, 967-1001.

Huntley, B., Green, R.E., Collingham, Y.C., Hill, J.K., Willis, S.G., Bartlein, P.J., Cramer, W., Hagemeijer, W.J.M. & Thomas, C.J. (2004) The performance of models relating species geographical distributions to climate is independent of trophic level. *Ecology Letters*, **7**, 417-426.

Jaccard, P. (1912) The distribution of the flora in the alpine zone. *New Phytol*,

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An introduction to statistical learning*. Springer.

Jenkins, C.N., Powell, R.D., Bass, O.L. & Pimm, S.L. (2003) Why sparrow distributions do not match model predictions. *Animal Conservation*, **6**, 39-46.

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

Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2009) The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology*, **10**

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. *Global Ecology and Biogeography*, **22**, 508-516.

Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, **13**, 853-867.

Keil, P., Belmaker, J., Wilson, A.M., Unitt, P., Jetz, W. & Freckleton, R. (2013) Downscaling of species distribution models: a hierarchical approach. *Methods in Ecology and Evolution*, **4**, 82-94.

Kéry, M. (2011) Towards the modelling of true species distributions. *Journal of Biogeography*, **38**, 617-618.

Lahoz-Monfort, J.J., Guillera-Arroita, G. & Wintle, B.A. (2014) Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504-515.

Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014) Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, **5**, 54-64.

Leroy, B., Meynard, C.N., Bellard, C. & Courchamp, F. (2016) virtualspecies, an R package to generate virtual species distributions. *Ecography*, **39**, 599-607.

Liu, C.R., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385-393.

Lobo, J.M., Jimenez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145-151.

Lobo, J.M., Jiménez-Valverde, A. & Hortal, J. (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, **33**, 103-114.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248-2255.

Maher, S.P., Randin, C.F., Guisan, A. & Drake, J.M. (2014) Pattern-recognition ecological niche models fit to presence-only and presence-absence data. *Methods in Ecology and Evolution*, **5**, 761-770.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-931.

Marmion, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009a) The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecological Modelling*, **220**, 3512-3520.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009b) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59-69.

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models. 2nd edition*. Chapman and Hall, London.

McPherson, J.M. & Jetz, W. (2007) Effects of species' ecology on the accuracy of distribution models. *Ecography*, **30**, 135-151.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811-823.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2006) Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions—possibilities and limitations. *Ecological Modelling*, **192**, 499-522.

Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E. & Elith, J. (2014) What do we gain from simplicity versus complexity in species distribution models? *Ecography*, **37**, 1267-1281.

Mertes, K. & Jetz, W. (2018) Disentangling scale dependencies in species environmental niches and distributions. *Ecography*, **41**, 1604-1615.

Meyer, C.B. & Thuiller, W. (2006) Accuracy of resource selection functions across spatial scales. *Diversity and Distributions*, **12**, 288-297.

Meynard, C.N. & Quinn, J.F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**, 1455-1469.

Meynard, C.N. & Kaplan, D.M. (2013) Using virtual species to study species distributions and model performance. *Journal of Biogeography*, **40**, 1-8.

Miller, J.A. (2014) Virtual species distribution models: Using simulated data to evaluate aspects of model performance. *Progress in Physical Geography*, **38**, 117-128.

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. *Methods in Ecology and Evolution*, **8**, 12-21.

Mittelbach, G.G. & Schemske, D.W. (2015) Ecological and evolutionary perspectives on community assembly. *Trends Ecol Evol*, **30**, 241-7.

Mod, H.K., Scherrer, D., Luoto, M. & Guisan, A. (2016) What we use is not what we know: environmental predictors in plant distribution models. *Journal of Vegetation Science*, **27**, 1308-1322.

Moilanen, A., Pouzols, F.M., Meller, L., Veach, V., Arponen, A., Leppänen, J. & Kujala, H. (2014) ZONATION: spatial conservation planning framework and software. Version 4. User Manual. In. Atte Moilanen / Metapopulation Research Group, University of Helsinki, Finland.

Munkemuller, T., Gallien, L., Lavergne, S., Renaud, J., Roquet, C., Abdulhak, S., Dullinger, S., Garraud, L., Guisan, A., Lenoir, J., Svenning, J.C., Van Es, J., Vittoz, P., Willner, W., Wohlgemuth, T., Zimmermann, N.E. & Thuiller, W. (2014) Scale decisions can reverse conclusions on community assembly processes. *Glob Ecol Biogeogr*, **23**, 620-632.

Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133-142.

Nenzén, H.K. & Araújo, M.B. (2011) Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling*, **222**, 3346-3354.

Normand, S., Treier, U.A., Randin, C., Vittoz, P., Guisan, A. & Svenning, J.-C. (2009) Importance of abiotic stress as a range-limit determinant for European plants: insights from species responses to climatic gradients. *Global Ecology and Biogeography*, **18**, 437-449.

Osborne, P.E. & Leitao, P.J. (2009) Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Diversity and Distributions*, **15**, 671-681.

Owens, H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C.M., Myers, C.E. & Peterson, A.T. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, **263**, 10-18.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225-245.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology & Biogeography*, **12**, 361-371.

Pearson, R.G., Dawson, T.P. & Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285-298.

Pellissier, L., Espíndola, A., Pradervand, J.-N., Dubuis, A., Pottier, J., Ferrier, S., Guisan, A. & Araújo, M. (2013) A probabilistic approach to niche-based community models for spatial forecasts of assemblage properties and their uncertainties. *Journal of Biogeography*, n/a-n/a.

Peterson, A.T., Papes, M. & Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, **30**, 550-560.

Peterson, A.T., Papes, M. & Soberon, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, **213**, 63-72.

Peterson, A.T., Soberon, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araujo, M.B. (2011) Evaluating model performance and significance. *Ecological niches and geographic distributions*, pp. 150-181. Princeton University Press, Princeton, NJ.

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. *Global Ecology and Biogeography*, **26**, 275-287.

Petitpierre, B., McDougall, K., Seipel, T., Broenniman, O., Guisan, A. & Kueffer, C. (2016) Will climate change increase the risk of plant invasions into mountains? *Ecological Applications*,

Phillips, S.J. (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography*, **31**, 272-278.

Pineda, E. & Lobo, J.M. (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182-190.

Pineda, E. & Lobo, J.M. (2012) The performance of range maps and species distribution models representing the geographic variation of species richness at different resolutions. *Global Ecology and Biogeography*, **21**, 935-944.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B. & Maintainer, R. (2017) Package 'nlme'. *Linear and nonlinear mixed effects models*, 3-1.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P., Guisan, A. & Field, R. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **22**, 52-63.

Prasad, A.M., Iverson, L.R. & Liaw, A. (2006) Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181-199.

Qiao, H., Lin, C., Jiang, Z. & Ji, L. (2015) Marble Algorithm: a solution to estimating ecological niches from presence-only records. *Scientific Reports*, **5**, 14232.

Qiao, H., Feng, X., Escobar, L.E., Peterson, A.T., Soberón, J., Zhu, G. & Papeş, M. (2018) An evaluation of transferability of ecological niche models. *Ecography*, **0**

R Core Team (2017) *R: A language and environment for statistical computing.*

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689-1704.

Record, S., Strecker, A., Tuanmu, M.N., Beaudrot, L., Zarnetske, P., Belmaker, J. & Gerstner, B. (2018) Does scale matter? A systematic review of incorporating biological realism when predicting changes in species distributions. *PLoS One*, **13**, e0194650.

Regos, A., D'Amen, M., Herrando, S., Guisan, A. & Brotons, L. (2015) Fire management, climate change and their interacting effects on birds in complex Mediterranean landscapes: dynamic distribution modelling of an early-successional species—the near-threatened Dartford Warbler (Sylvia undata). *Journal of Ornithology*, **156**, 275-286.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011) Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography*, **35**, 211-226.

Rodríguez-Castañeda, G., Hof, A.R., Jansson, R. & Harding, L.E. (2012) Predicting the fate of biodiversity using species' distribution models: enhancing model comparability and repeatability. *PLoS One*, **7**, e44402.

Roura-Pascual, N., Brotons, L., Peterson, A.T. & Thuiller, W. (2009) Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions*, **11**, 1017-1031.

Sala, O.E., Chapin, F.S., 3rd, Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M. & Wall, D.H. (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770-4.

Sánchez-Fernández, D., Lobo, J.M. & Hernández-Manrique, O.L. (2011) Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. *Diversity and Distributions*, **17**, 163-171.

Santika, T. (2011) Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography*, **20**, 181-192.

Scherrer, D., D'Amen, M., Mateo, M.R.G., Fernandes, R.F. & Guisan , A. (2018a) How to best threshold and validate stacked species assemblages? Community optimisation might hold the answer. *Methods in Ecology and Evolution*, **in press**

Scherrer, D., Massy, S., Meier, S., Vittoz, P., Guisan, A. & Serra-Diaz, J. (2017) Assessing and predicting shifts in mountain forest composition across 25 years of climate change. *Diversity and Distributions*, **23**, 517-528.

Scherrer, D., Mod, H.K., Pottier, J., Litsios-Dubuis, A., Pellissier, L., Vittoz, P., Götzenberger, L., Zobel, M. & Guisan, A. (2018b) Disentangling the processes driving plant assemblages in mountain grasslands across spatial scales and environmental gradients. *Journal of Ecology*, **0**

Soberon, J. & Nakamura, M. (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proc Natl Acad Sci U S A*, **106 Suppl 2**, 19644-50.

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

Steinmann, K., Eggenberg, S., Wohlgemuth, T., Linder, H.P. & Zimmermann, N.E. (2011) Niches and noise—Disentangling habitat diversity and area effect on species diversity. *Ecological Complexity*, **8**, 313-319.

Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1-13.

Swets, J.A. (1988) Measuring the Accuracy of Diagnostic Systems. *Science*, **240**, 1285-1293.

Talluto, M.V., Boulangeat, I., Ameztegui, A., Aubin, I., Berteaux, D., Butler, A., Doyon, F., Drever, C.R., Fortin, M.-J., Franceschini, T., Liénard, J., McKenney, D., Solarik, K.A., Strigul, N., Thuiller, W. & Gravel, D. (2016) Cross-scale integration of knowledge for predicting species ranges: a metamodelling framework. *Global Ecology and Biogeography*, **25**, 238-249.

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C. & Guisan, A. (2014) Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution*, **5**, 947-955.

Thuiller, W. (2004a) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2020-2027.

Thuiller, W., Brotons, L., Araujo, M.B. & Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165-172.

Thuiller, W., Lafourcade, B., Engler, R. & Araujo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369-373.

Thuiller, W.A., M.B. & Lavorel, S. (2004b) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*, **31**, 353-361.

Titeux, N., Maes, D., Van Daele, T., Onkelinx, T., Heikkinen, R.K., Romo, H., García-Barros, E., Munguira, M.L., Thuiller, W., van Swaay, C.A.M., Schweiger, O., Settele, J., Harpke, A., Wiemers, M., Brotons, L. & Luoto, M. (2017) The need for large-scale distribution data to estimate regional changes in species richness under future climate change. *Diversity and*

Distributions, **23**, 1393-1407.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003) Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, **13**, 1790-1801.

Vale, C.G., Tarroso, P. & Brito, J.C. (2014) Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara-Sahel transition zone. *Diversity and Distributions*, **20**, 20-33.

Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistic with S*. Springer, Dordrecht, The Netherlands.

Vicente, J.R., Goncalves, J., Honrado, J.P., Randin, C.F., Pottier, J., Broennimann, O., Lomba, A. & Guisan, A. (2014) A framework for assessing the scale of influence of environmental factors on ecological patterns. *Ecological Complexity*, **20**, 151-156.

Vicente, J.R., Fernandes, R.F., Randin, C.F., Broennimann, O., Goncalves, J., Marcos, B., Pocas, I., Alves, P., Guisan, A. & Honrado, J.P. (2013) Will climate change drive alien invasive plants into areas of high protection value? An improved model-based regional assessment to prioritise the management of invasions. *J Environ Manage*, **131**, 185-95.

Vicente, J.R., Alagador, D., Guerra, C., Alonso, J.M., Kueffer, C., Vaz, A.S., Fernandes, R.F., Cabral, J.A., Araujo, M.B. & Honrado, J.P. (2016) Cost-effective monitoring of biological invasions under global change: a model-based framework. *Journal of Applied Ecology*, **53**, 1317-1329.

Ward, G., Hastie, T., Barry, S., Elith, J. & Leathwick, J.R. (2009) Presence-Only Data and the EM Algorithm. *Biometrics*, **65**, 554-563.

Watts, M.E., Ball, I.R., Stewart, R.S., Klein, C.J., Wilson, K., Steinback, C., Lourival, R., Kircher, L. & Possingham, H.P. (2009) Marxan with Zones: Software for optimal conservation based land- and sea-use zoning. *Environmental Modelling & Software*, **24**, 1513-1521.

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

Williams, J.N., Seo, C.W., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M. & Schwartz, M.W. (2009) Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, **15**, 565-576.

Williams, J.W. & Jackson, S.T. (2007) Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**, 475-482.

Wisz, M.S. & Guisan, A. (2009) Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecol*, **9**, 8.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & Distribut, N.P.S. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763-773.

Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Fielding, A.H., Bamford, A.J., Ban, S., Barbosa, A.M., Dormann, C.F., Elith, J., Embling, C.B., Ervin, G.N., Fisher, R., Gould, S., Graf, R.F., Gregr, E.J., Halpin, P.N., Heikkinen, R.K., Heinänen, S., Jones, A.R., Krishnakumar, P.K., Lauria, V., Lozano-Montes, H., Mannocci, L., Mellin, C., Mesgaran, M.B., Moreno-Amat, E., Mormede, S., Novaczek, E., Oppel, S., Ortuño Crespo, G., Peterson, A.T., Rapacciuolo, G., Roberts, J.J., Ross, R.E., Scales, K.L., Schoeman, D., Snelgrove, P., Sundblad, G., Thuiller, W., Torres, L.G., Verbruggen, H., Wang, L., Wenger, S., Whittingham, M.J., Zharikov, Y., Zurell, D. & Sequeira, A.M.M. (2018) Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*,

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Munkemuller, T., Nehrbass, N., Pagel, J., Reineking, B., Schroder, B. & Grimm, V. (2010) The virtual ecologist approach:

simulating data and observers. *Oikos*, **119**, 622-635.

Zurell, D., Thuiller, W., Pagel, J., Cabral, J.S., Munkemuller, T., Gravel, D., Dullinger, S., Normand, S., Schiffers, K.H., Moore, K.A. & Zimmermann, N.E. (2016) Benchmarking novel approaches for modelling species range dynamics. *Glob Chang Biol*, **22**, 2651-64.

PART II - ADDITIONAL CHAPTERS INSIDE THESIS' SCOPE

Chapter 2.1

Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models

This paper is published in *Journal of Biogeography*

My contribution to the paper: This paper was a collaboration with current and former members of my group to publish a previously rejected manuscript. I analysed the data, mainly re-running all the models. I also created all the figures and was actively involved in the re-writing of large portions of the manuscript.

Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models

Manuela D'Amen[a†], Anne Dubuis[a†], Rui F. Fernandes[a], Julien Pottier[b], Loïc Pellisser[a,c] and Antoine Guisan[a,d *]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland

[b] INRA, Grassland Ecosystem Research Unit (UREP), 5 Chemin de Beaulieu, 63100 Clermont-Ferrand, France

[c] Department of Biology, Ecology and Evolution, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland

[d] Institute of Earth Surface Dynamics, University of Lausanne, Géopolis, CH-1015 Lausanne, Switzerland

ABSTRACT

Modelling species distributions at the community level is required to make effective forecast of global change impacts on diversity and ecosystem functioning. Community predictions may be achieved using macroecological properties of communities (macroecological models, MEM), or by stacking of individual species distribution models (stacked species distribution models, S-SDMs). To obtain more realistic predictions of species assemblages, the SESAM (spatially explicit species assemblage modelling) framework suggests applying successive filters to the initial species source pool, by combining different modelling approaches and rules. Here we provide a first test of this framework in mountain grassland communities. Two implementations of the SESAM framework were tested: a 'probability ranking' rule based on species richness predictions and rough probabilities from SDMs, and a 'trait range' rule that uses the predicted upper and lower bound of community-level distribution of three different functional traits (vegetative height, specific leaf area and seed mass) to constrain a pool of species from binary SDMs predictions. We showed that all independent constraints contributed to reduce species richness overprediction. Only the 'probability ranking' rule allowed slight but significant improvements in the predictions of community composition. We tested various implementations of the SESAM framework by integrating macroecological constraints into S-SDM predictions, and report one that is able to improve compositional predictions. We discuss possible improvements, such as further understanding the causality and precision of environmental predictors, using other assembly rules and testing other types of ecological or functional constraints.

KEYWORDS: Community ecology, functional ecology, macroecological models, MEM, SESAM framework, species distribution models, SDM, stacked-SDM

INTRODUCTION

Understanding the distribution and composition of species assemblages and being able to predict them in space and time are important for understanding the fate of biodiversity under global change. Different approaches have been proposed to predict the composition of species assemblages, which can work on mechanistic or empirical bases. Neutral views have also been proposed to explain relative abundance patterns in communities (Hubbell, 2001), which were contrasted to niche/trait views (Wennekes *et al.*, 2012). Neutral theory has been challenged for not representing forces that actually operate in nature to shape communities and their composition (e.g. Clark, 2009). Using a more deterministic approach, Shipley *et al.* (2006) proposed the use of predicted community weighted means of functional traits to infer the assemblage composition given species traits through a maximum entropy approach (Shipley *et al*., 2006, 2011; Sonnier *et al*., 2010a; see also Laughlin *et al.*, 2012). Mokany *et al*. (2011, 2012) proposed a dynamic framework to model species richness and composition dissimilarity based on species data. A distinct approach, not requiring traits, is to use the empirical relationships between species distribution data and environmental factors to predict community types or axes of compositional variation derived from ordination techniques (Ferrier & Guisan, 2006).

One widely used method is to predict the distributions of individual species with niche-based species distribution models (SDMs; also called ecological niche models, ENMs; see Guisan *et al.,* 2013), and then to stack them to predict species assemblages (stacked-SDM, S-SDM; Dubuis *et al.,* 2011). This method pertains to the category 'predict first, assemble later' in Ferrier & Guisan's (2006) classification of community-level models, and has been tested in recent studies to draw conclusions about species richness (SR), assemblage composition or species turnover under current or future climatic conditions (Baselga & Araújo, 2009, 2010; Aranda & Lobo, 2011; Albouy *et al.*, 2012; Pottier *et al.*, 2013). Stacking individual species predictions can be applied to both rough probabilities (pS-SDM) and binary predictions from SDMs (bS-SDM) (e.g. Dubuis *et al*., 2011; Calabrese *et al*., 2014). pS-SDM currently allows the prediction of species richness only, while bS-SDM also provides information on species composition. It has been shown that bS-SDMs tend, on average, to overpredict species richness per unit area (Algar *et al*., 2009; Dubuis *et al*., 2011; Mateo *et al*., 2012), whereas pS-SDMs do not (Dubuis *et al.,* 2011; Calabrese *et al.,* 2014). Overprediction by bS-SDMs could be expected, as reconstructing communities from SDM predictions implies applying a series of species-specific abiotic filters, without consideration for macroecological constraints on the general properties of the system as a whole (Guisan & Rahbek, 2011). As an alternative explanation, it has also been suggested that overprediction could result from a mathematical artefact if the stacking process is applied to binary SDM predictions, i.e. after thresholding the rough probability of species' predictions (Calabrese *et al.*, 2014).

Guisan & Rahbek (2011) proposed a framework – SESAM: spatially explicit species assemblage modelling – that aims to improve predictions of species assemblages. The main idea of the SESAM framework is to reconstruct species assemblages by applying successive filters of the assembly process through four main conceptual steps (Hortal *et al.*, 2012). First, the species pool of each modelling unit in the study area must be defined. Second, species are filtered from the species pool according to their suitability to the environmental conditions in the modelling unit, e.g. by fitting SDMs. Third, limits previously set to one or several properties of each assemblage (e.g. richness or functional properties) are used to apply constraints on the assemblage in each unit, based on model predictions. Fourth, the species to be kept in the assemblage are chosen among the potential coexisting species (i.e. those predicted by the S-SDM), through biotic assembly rules. Macroecological constraints can be defined by macroecological models (MEMs), i.e. models of emergent properties or attributes of communities, such as species richness (SR) or other functional characteristics (e.g. functional richness) that are theoretically predictable directly from environmental variables (Francis & Currie, 2003; Moser *et al*., 2005; Sonnier *et al*., 2010b; Dubuis *et al*., 2011, 2013). MEMs, which belong to the 'assemble first, predict later' category of Ferrier & Guisan (2006)'s classification, have been shown to provide less biased predictions of SR than bS-SDMs (Dubuis *et al*., 2011). Yet, no attempt has been made to implement and test the SESAM framework.

In the SESAM framework, assemblage properties are predicted to define constraints to be applied to the assemblage in each unit. In this study, we test three macroecological constraints: (1) richness predicted by the sum of probability S-SDM (pS-SDM); (2) direct predictions of species richness (MEM) (Dubuis *et al*., 2011); and (3) predicted values of three functional traits (Dubuis *et al.*, 2013). In particular, we test the use of functional traits as macroecological constraints, as they can be predicted spatially (Dubuis *et al.*, 2013) and may provide an understanding of the functional underpinnings of plant communities, allowing generalization beyond species identities (e.g. Hooper *et al.*, 2005; McGill *et al.,* 2006). Functional traits are supposed to enable the refinement of predictions of community composition along environmental gradients, by contrasting trait values for individual species to the ones aggregated at the community level (Shipley *et al.*, 2006; Douma *et al.*, 2012). We consider extremes in trait values to represent a filtering effect, i.e. the trait values that allow a species to be included in a community in a given environment (Keddy, 1992a, b). In order to build macroecological constraints, the same rationale applies to both richness and traits extreme values: limited amount of resources or environmental conditions (e.g. heterogeneity) defines 'how many' or 'what type of' species can strive in the considered unit. Here, both species richness and the functional characteristics of the community are assumed to be mainly controlled, among other possible factors, by available energy, as expressed by climatic predictors (Wright, 1983; Currie, 1991; Hawkins *et al.*, 2003; Shipley *et al*., 2006; see Guisan & Rahbek, 2011).

By integrating over these sources of information, we set macroecological constraints on the pool of species predicted to potentially co-occur in each site according to SDM predictions only. Doing this, we test – for the first time – a simplified version of the SESAM framework (i.e. without elaborated biotic assembly rules), using outputs from MEMs or pS-SDMs as constraints to limit the number of species predicted by bS-SDMs, this way attempting to improve predictions of community composition. More specifically, we ask the following questions:

1. Does combining different modelling techniques developed for biodiversity prediction improve the predictions of community attributes such as richness, species composition, traits distribution?

2. Does the use of assembly rules (driven either by habitat suitability or functional characteristics) to select the species that enter in the predicted community from SDMs improve the predictions of community richness and composition?

MATERIALS AND METHODS

Vegetation and traits data

The study area is located in the Alps of western Switzerland (http://rechalpvd.unil.ch) and covers *c*. 700 km^2, with elevations ranging from 375 to 3210 m. The species occurrence data used in our analysis originate from fieldwork conducted between 2002 and 2009 in the study area following a random-stratified sampling design and limited to open, non-woody vegetation (for more information see Dubuis *et al*., 2011). A first dataset of 613 vegetation plots of 4 m^2 each was inventoried and used for SDM and MEM calibration ('calibration dataset'). An additional set of 298 plots was identically surveyed to evaluate S-SDMs, and test the efficiency of MEM constraints ('evaluation dataset') (Fig. 1 – Data box). This evaluation dataset was shown to be spatially independent of the first one, and thus valid for model evaluation, by calculating the spatial correlation of SDMs' residuals between the calibration and the evaluation datasets based on neighbourhood graphs and Moran's *I* coefficient (Pottier *et al*., 2013).

A total of 241 species were recorded in the study area, with traits data available for a subset of the 189 most frequent species of this pool (Fig. 1; Pottier *et al.*, 2013; Dubuis *et al.*, 2013). We selected three traits (vegetative height, specific leaf area and seed mass) that are expected to represent the key axes of plant ecological strategies following the leaf–height–seed (LHS) scheme of Westoby (1998), already widely used for studying plant assembly rules. In particular, vegetative height (H) and specific leaf area (SLA) were measured on the field (for each species between 4 and 20 individuals were sampled over its entire bioclimatic range). We used the average trait value among all sampled individuals for each species for further analyses (Dubuis *et al*., 2013). Height was measured for each species in the field as the distance between top photosynthetic tissues and the ground, expressed in mm. This trait is related to competitive ability and is correlated with above-ground biomass (Cornelissen *et al*., 2003). SLA was calculated as the ratio of leaf surface to its dry mass and expressed in mm^2 mg^{-1}. SLA is correlated with the relative growth rate and photosynthetic ability of plant species (Cornelissen *et al.*, 2003). Seed mass (SM) data originate from literature and field measurements (Pellissier *et al*., 2010) and is expressed in milligrams. This trait is a good predictor of colonization ability of the species and seedling survivorship (Moles & Westoby, 2006). To account for trait range limitation, we calculated percentiles of trait distribution in sites where the 189 species for which trait data were available represented more than 80% of the total vegetation cover (Pakeman & Quested, 2007; see Pottier *et al.,* 2013, Dubuis *et al.*, 2013).

General analytical framework

We tested different implementations of the SESAM framework to predict species composition, by applying two different types of species assembly rules:

1. ‘Probability ranking’ rule: this rule is based on the assumption that species with the highest habitat suitability are competitively superior. According to this rule, community composition is obtained by selecting the species in decreasing order of their predicted probability of presence from SDMs up to the richness prediction (i.e. predictions from MEM or pS-SDM).

2. ‘Trait range’ rule: we applied a filter based on important functional characteristics of plant species that relate to competitive and reproductive abilities. We used percentile predictions from MEMs of three functional traits, individual or in combination, as criteria to discard species that do not fall into the predicted functional range of the sites. We implemented this approach with the three percentiles boundaries.

We fitted all the models (both SDMs and MEMs) by applying three modelling techniques in R (2.14.1) with the BIOMOD package (Thuiller *et al*., 2009): generalized linear models (GLMs), generalized additive models (GAMs) and generalized boosted models (GBMs). The resulting projections were averaged to implement an ensemble forecasting approach.

We applied the SESAM framework following the four-step design described by Guisan & Rahbek (2011) and adapted to our study case (Fig. 1).

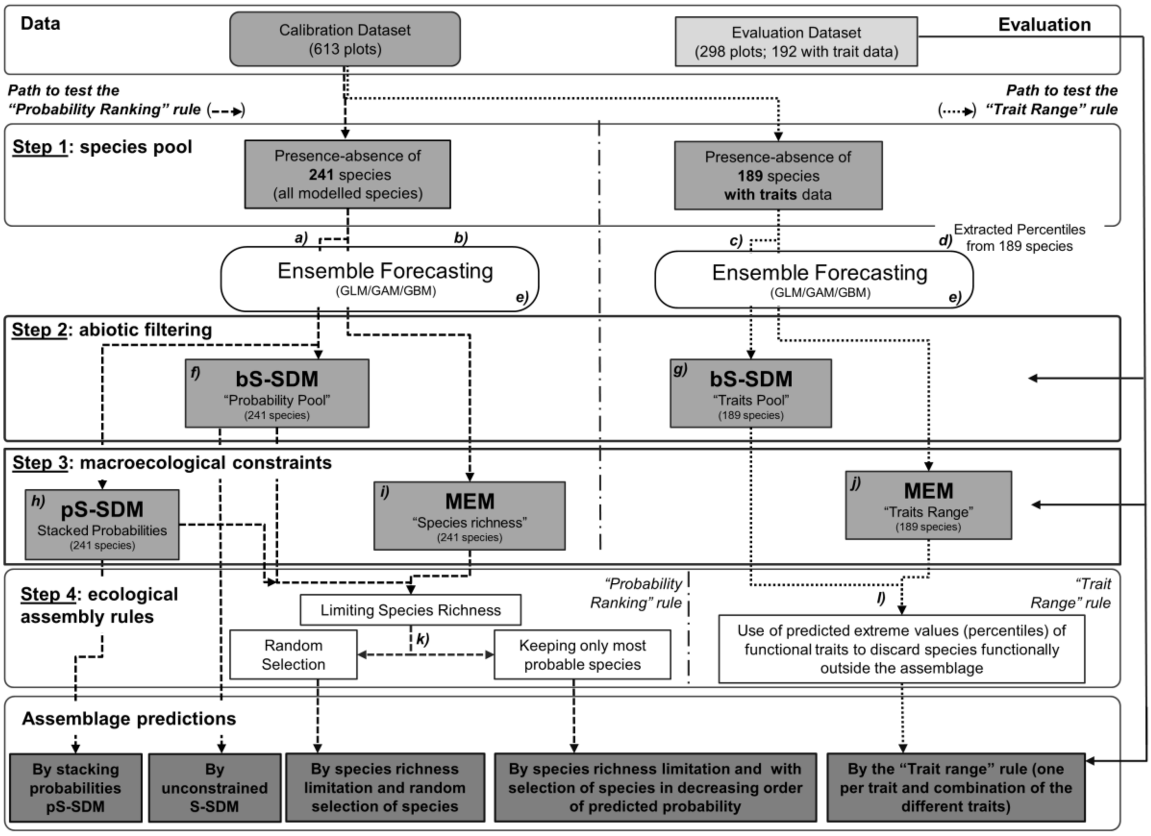


FIGURE 1 Workflow of the analytical steps followed in the study. *Data box:* We used a calibration and an evaluation dataset derived from field samplings carried out on 613 and 298 (192 with trait data) plots, respectively. These datasets were used to test the 'probability ranking' rule (left side of the figure with dashed arrows) and the 'trait range' rule (left side of the figure with dotted arrows). *Step 1 – species pool:* a total of 241 species collected in the study area were considered the 'species pool' to test the 'probability ranking' rule, (a) and (b). A subset of this species pool (189 species with trait data) was used to test the 'trait range' rule (c) and (d). All models were fitted by an ensemble forecasting approach based on the average of three techniques: generalized linear models (GLM), generalized additive models (GAM), and generalized boosted models (GBM). (e). *Step 2 – abiotic filtering*: distribution of individual species (a) and (c) were modelled and then stacked to create binary stacked species distribution model (S-SDM) predictions to represent a 'probability pool' for the 'probability ranking' rule test (f) and a 'traits pool' for the 'trait range' rule test (g). *Step 3 – macroecological constraints:* three different methods were used to define macroecological constraints, resulting in models with the stacked probabilities from SDMs (h; pS-SDM) and two different macroecological models (MEMs). These were created by modelling directly species richness values (i; SR_MEM) and three pairs of traits percentiles (j; Traits_MEM). *Step 4 – ecological assembly rules*: in the test of the 'probability ranking' rule (k) we limited species richness to fit the MEM or pS-SDM predictions and the species composition was determined (1) as a random selection from the pool or (2) selecting the species in decreasing order of predicted probability. In the test for the 'trait range' rule (l) we used the predicted values of MEM of functional traits (each trait separately and combinations of traits) to discard species functionally outside the assemblage. *Assemblage prediction box:* all the outputs resulting from the different approaches were compared and evaluate using the evaluation dataset (solid arrows).

Step 1 – Species pool

As the first component of the SESAM framework, we considered a unique species pool for all modelling units, defined as the most frequent plant species occurring in our study area (241 species). This pool was used to test the 'probability ranking' rule. A subset of this pool was used to test the 'trait range' rule (189 species).

Step 2 – Abiotic filtering

Single species models were fitted with environmental predictors calculated from temperature and precipitation data recorded by the Swiss network of meteorological stations and from a digital elevation model at 25 m resolution (see Dubuis *et al*., 2011). We used growing degree-days (above 0 °C), moisture index over the growing season (difference between precipitation and potential evapotranspiration), the sum of solar radiations over the year, slope (in degree) and topographic position (unit less, indicating the ridges and valleys). These five variables have been shown to be useful for predicting the topo-climatic distributions of plant species in mountainous environment (Dubuis *et al*., 2011). The models were evaluated on the evaluation dataset with the area under the curve (AUC) of a receiver operating characteristic plot (ROC) and the true skill statistic (TSS; Allouche *et al*., 2006). Ensemble predictions were obtained by computing the weighted average of the predictions by the three techniques. To do this, we used weights from the internal cross-validation with both AUC (Swets, 1988) and TSS (Allouche *et al*., 2006) evaluation metrics. The predictive ability of the final ensemble models was then tested with the same metrics using the external evaluation dataset. The raw predictions for the 241 species represent the 'probability pool' used in the 'probability ranking' rule test. In 'trait range' rule tests the projected species distributions for the 189 species were transformed into binary presences and absences using two threshold approaches: (1) the threshold corresponding to equal values of sensitivity and specificity (Liu *et al*., 2005), and (2) the threshold maximizing TSS. The resulting binary projections were stacked to predict assemblages in each of the evaluation plots (bS-SDM). This way, we obtained a pool of species potentially present filtered by topo-climatic factors.

Step 3 – Macroecological constraints

Three different methods were used to define macroecological constraints. First, we summed probabilities from SDMs (Dubuis *et al*., 2011) for the 241 species, obtaining a prediction of richness for each unit (pS-SDM). Second, observed species richness (SR) was calculated as the number of species (among the 241 used in this study) present in each sampling plot. Total SR was predicted with the same environmental predictors and modelling techniques used for SDMs fitted with a Poisson distribution. Also, in this case, we applied the ensemble forecasting approach (as described above) to obtain a final richness prediction ('species richness' MEM; see Dubuis *et al*., 2011). Finally, we modelled traits values, considering three pairs of percentiles limits: 1st–99th, 5th–95th and 10th–90th. We modelled each trait percentile as a function of the environmental predictors and assuming a normal distribution (‘traits range’ MEM; Dubuis *et al.*, 2013). The modelling procedure was the same used for species richness prediction. Prior to modelling, trait data were log-transformed. The predictive power of the SR and traits range models were measured by computing a Spearman rank correlation between the observed and predicted indices values for the evaluation dataset.

Step 4 – Ecological assembly rules

We applied our rules to couple results coming from previous steps. To test the ‘probability ranking’ rule, we determined the community composition by ranking the species in decreasing order of their predicted probability of presence from SDMs up to the richness prediction by pS-SDM or SR-MEM. We further compared the application of this rule with a random selection of species in the number of the richness predictions, as a null test of composition prediction success. This was performed on the full evaluation dataset of 298 plots not used in model calibration.

In the ‘trait range’ rule, for each site, among the species predicted as present by the binary SDMs (‘traits pool’), we excluded from the final community prediction those species with traits valued outside the predicted functional range predicted by MEMs. In particular, for each percentile pair (1st–99th, 5th–95th, and 10th–90th), we considered the predicted trait values and we excluded all species having traits values outside these quantiles. All seven combinations of the three functional traits were considered (taken singularly, in pairs or all together) to constraint community composition. As a result, we tested a total of 21 macroecological constraints based on traits. The ‘trait range’ rule was applied to the 192 plots of the evaluation dataset for which we had trait data for more than 80% of the vegetation cover for the second test.

Finally, species richness and composition outputs resulting from the SESAM approaches were compared to the evaluation dataset. Assemblage predictions were evaluated with several metrics based on a confusion matrix where all species (species pool: SP) are classified into: TP: the species observed as well as predicted as present (true positive), FN: the species observed as present but predicted as absent (false negative; omission error), FP: the species observed as absent but predicted as present (false positive; commission error) and TN: the species both observed and predicted as absent (true negative) (see Appendix S1 in Supporting Information). We computed the species richness error (predicted SR – observed SR, expressed as a number of species in Fig. 2), the assemblage prediction success (a), and the Sørensen index, related to Bray–Curtis dissimilarity (b).

$$\text{(a) Prediction success} = \frac{TP + TN}{SP} \qquad \text{(b) Sørensen index} = \frac{2TP}{2TP + FN + FP}$$

RESULTS

SDMs for most species had an AUC value higher than 0.7 and can therefore be considered as useful for predictions (see Appendix S2). The MEM for species' richness and pS-SDM gave similar results: both predictions showed fair correlations between observed and predicted values of richness in the evaluation dataset (ρ = 0.529 and 0.507, respectively, Spearman rank correlation test). Macroecological models for traits were all above 0.5 (ρ values, Spearman rank correlation test) except for the 1st and 5th percentiles of log(SM) (Appendix S2). The 'trait range' rule was applied by considering all couples of percentile, but as the results are consistent (see Appendix S3), in the following section we only show results coming from the 5th–95th percentiles. The S-SDM built with binary SDMs overpredicted species richness (SR) in all plots (Figs 2a & 3). All filtering types, both coming from the 'probability ranking' rule and the 'trait range' rule contributed on average to reduce SR overprediction, i.e. reduction of SR error (Figs 2a, d & 3), except when using the combination of SLA and SM trait limits as constraining rule.

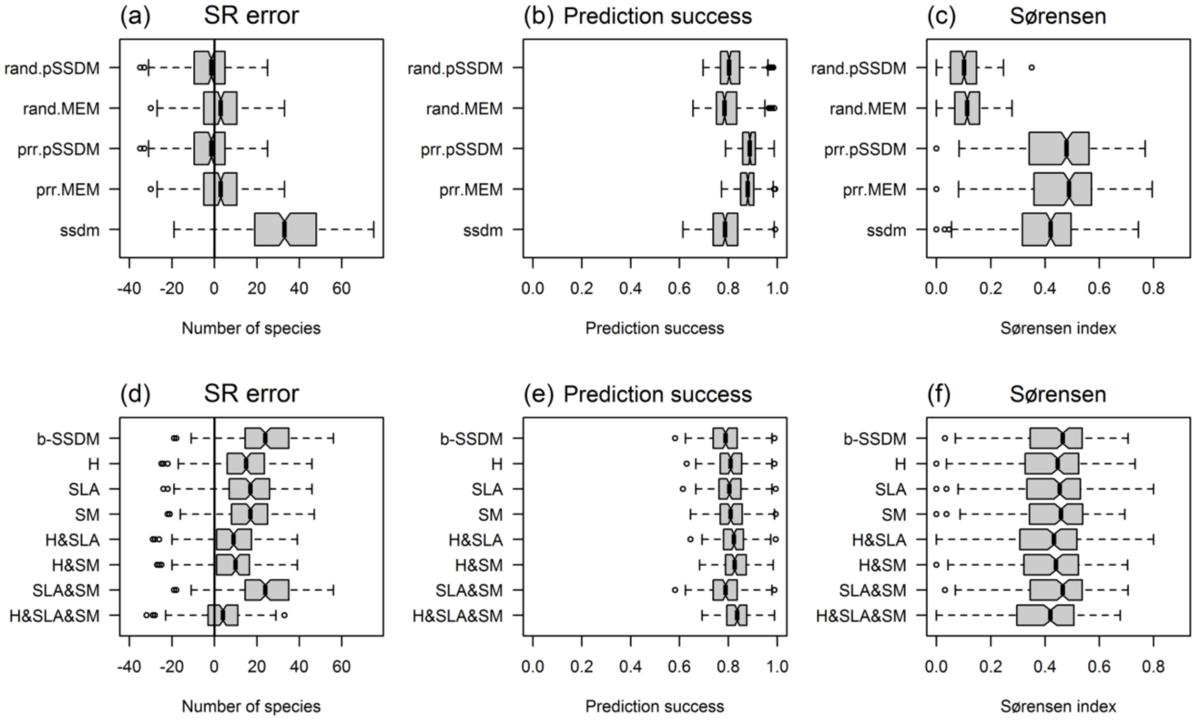


FIGURE 2 Boxplots comparing unconstrained stacked species distribution model (S-SDM) predictions to results from the 'probability ranking' rule and random tests when applied constraining richness by the sum of probabilities from SDMs (PRR.pSSDM and rand.pSSDM, respectively) or by macroecological models (PRR.MEM and rand.MEM, respectively) (a, b, c), and to results from the 'trait range' rule test for single traits and all their combinations (d, e, f). The metrics utilized in the comparison are: species richness error, i.e. predicted SR – observed SR (first column); prediction success, i.e. sum of correctly predicted presences and absences divided by the total species number (second column); and Sørensen index, i.e. a statistic used to compare the similarity of two samples (third column). Abbreviations: SR rand, a random choice of species from the probability pool to reach the number predicted by richness model; SR prob, selection of the most probable species to reach the number predicted by richness model; H, height; SLA, specific leaf area of the community; SM, seed mass.

Considering composition predictions, the prediction success was increased when applying either the ‘probability ranking’ rule or the ‘trait range’ rule (Fig. 2b,e), again with the exception of the combination of SLA and SM trait limits. Results from the Sørensen index (Fig. 2c) indicate that the ‘probability ranking’ rule increased the predictive capability by using both predicted SR from MEM and pS-SDM, as a limit, with the former slightly outperforming the latter. In both cases, the Sørensen index was significantly higher than the one of the simple bS-SDM (Wilcoxon signed rank test, *P*-value < 0.005). On average, this approach was less affected by errors of commission (false positive; Appendix S1) than other approaches and had the highest rate of correctly predicted absences (Fig. 4a). Using SR as a limit (from both MEM and pS-SDM) but choosing species randomly among those predicted yielded the worst assemblage composition predictions (Fig. 2c). We observed a decrease in the ability to correctly predict species identities when using the ‘trait range’ rule to constraints S-SDM predictions (Fig. 2f). Predicted functional traits did not provide a sufficient constraint to improve composition and did not allow for a complete reduction of the SR over-prediction. Their use allowed species richness prediction to be improved, but at the cost of slightly decreasing assemblage composition prediction success (Sørensen index) (Fig. 4b). The applications of our rules did not produce a prediction of species assemblage compositions better than an average Sørensen’s similarity of 0.5. Results for community predictions using TSS and the ‘trait range’ rule were similar to those using AUC and are thus presented in Appendix S1.

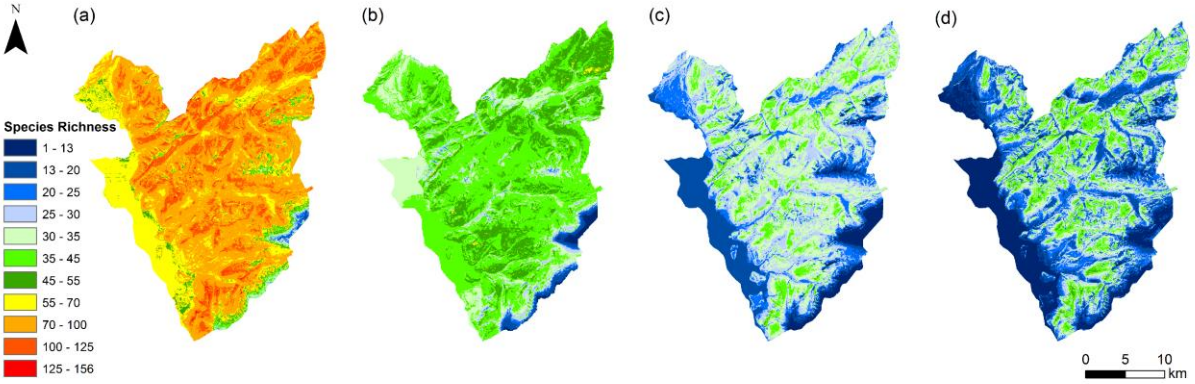


FIGURE 3 Predictions of species richness on the whole study area produced by (a) the unconstrained stacked species distribution model (S-SDM), and by the application of the SESAM framework implemented with (b) the ‘probability ranking’ rule implemented with the sum of probabilities from SDMs (pS-SDM), (c) the ‘probability ranking’ rule implemented with the richness estimation by the macroecological model (MEM) and (d) the ‘trait range’ rule (using the combination of the three traits as constraints).

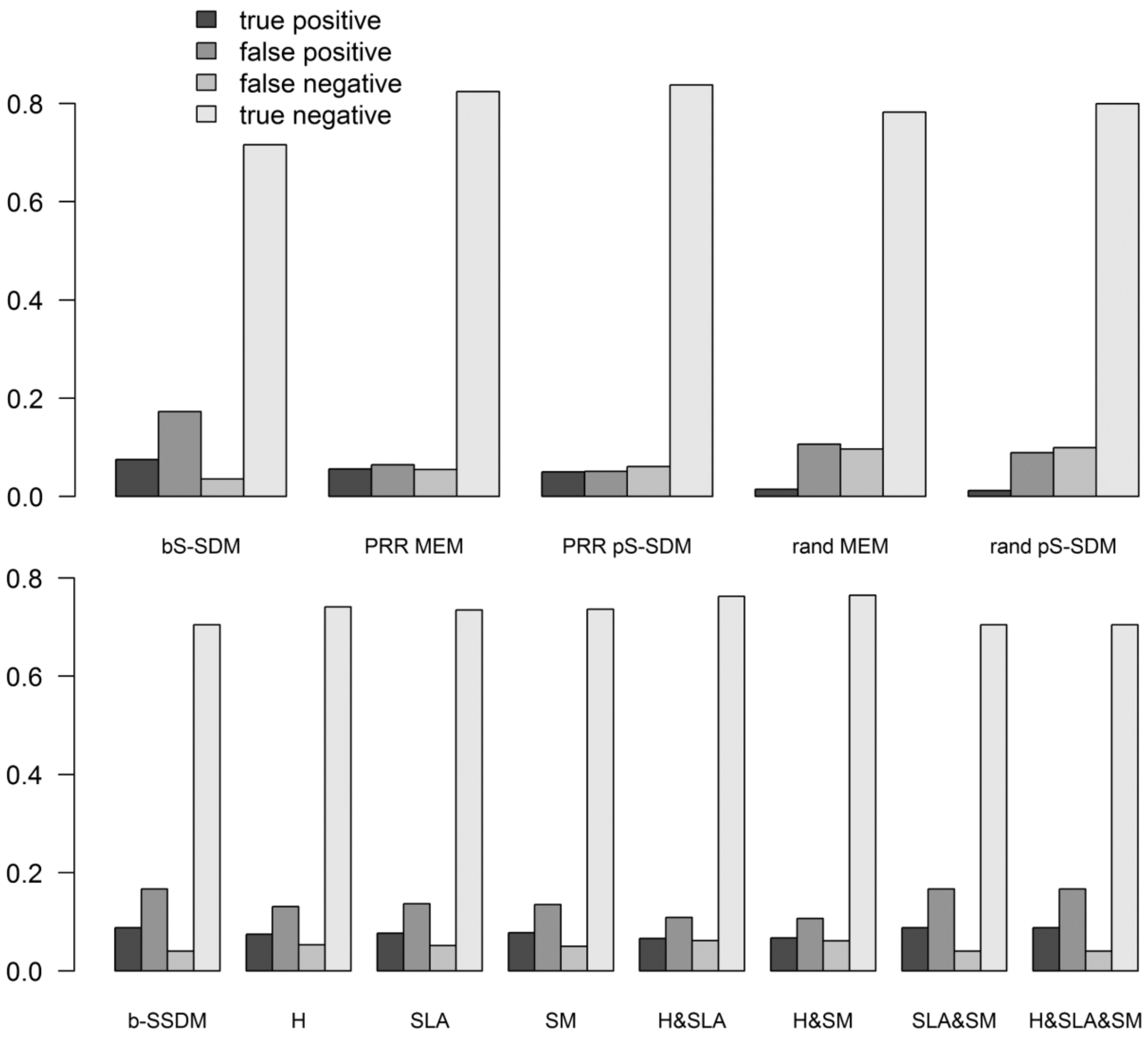


FIGURE 4 Histograms showing the proportion (mean among all plots) of true and false positive, as well as true and false negative for all the implementations of the SESAM framework, compared with the unconstrained sum of binary species distribution model (bS-SDM). In the upper plot results from the ‘probability ranking’ rule test implemented with macroecological models and sum of probabilities from SDMs (PRR MEM and PRR pS-SDM, respectively) and random selections (rand MEM and rand pS-SDM, respectively). In the lower plot results from the ‘trait range’ rule test for single traits and all their combinations (H, height; SLA, specific leaf area of the community; SM, seed mass).

DISCUSSION

This study represents the first formal test of the SESAM framework (Guisan & Rahbek, 2011). We have shown different ways to implement the SESAM framework, by integrating stacked predictions from species distribution models (S-SDMs) with richness predictions from macroecological models (MEMs) or from the sum of rough probabilities from S-SDM (pS-SDM). Our results show that the application of macroecological constraints on single species predictions from SDMs improve the overall quality of assemblage’ composition estimation. As expected, all the macroecological constraints considered reduced the overprediction of species richness. But more importantly, the sequence of steps of the framework allowed a more accurate prediction of the realized species assemblage as measured with metrics equally weighting commission (false presence) and omission (false absence) errors. This positive result encourages further developments of the SESAM framework to improve the prediction of community attributes.

Among the implementations of the SESAM framework tested here, the application of the 'probability ranking' rule improved the predictions of species richness and composition. First, both ways of producing species richness predictions, i.e. stacking of probabilities from SDMs (pS-SDM), and directly predicting species' richness (MEM), gave more reliable results than the simple binary S-SDMs, a result shown previously (e.g. Dubuis *et al.*, 2011; Calabrese *et al*., 2014). Second, this approach also produced better predictions of community composition, by selecting single species from the pool predicted by SDMs by decreasing order of predicted probability (until the predicted richness is reached). One possible explanation for this positive result is that the same species that are least likely to be present, i.e. the ones removed by the rule, are also the ones most likely to be overpredicted by bS-SDMs.

The 'trait range' rule (as applied here) proved less effective in constraining community predictions, and no specific functional trait or any percentile interval proved more efficient than another in reducing species richness overprediction. Although surprising because MEMs for traits were on average better than those for species richness (see Dubuis *et al*., 2011, 2013), we can hypothesize some explanations for this result: (i) we used trait averages for each species, whereas each of these traits is known *in situ* to exhibit intraspecific variation along environmental gradients (Albert *et al.*, 2010); (ii) the traits that we used have been shown not always to relate significantly to species' habitat suitability (Thuiller *et al*., 2010); (iii) a larger dataset of traits, as used in trait-based modelling approaches (e.g. Shipley *et al*., 2011), could have been more efficient in setting specific functional limits for the community prediction than the three traits used here. Still, the use of the combination of three traits as a constraint allowed an efficient decrease of species richness' overprediction, supporting the need to put restraints on species pools based on a simple stacking of species predictions. Roots traits, indicating below-ground competition, could be good additional candidates to complement the functional constraints. These and other possible trait types should be assessed in future studies testing the SESAM framework. A potential limitation to the use of particular functional traits is that they must relate to species' ability to cope with the environment and be reliably predicted in space by MEM (e.g. Dubuis *et al*., 2013), which may not always be possible. Finally, we used three different percentiles ranges to depict minimal and maximal trait values as functional constraints, but the results for community predictions were not significantly different, so that we can be confident that our outcomes were not dependent on the percentiles' choice.

Overall, and even after strongly reducing the species richness overprediction bias, predicted assemblage composition was improved but still remained significantly distinct from the observed ones, a result consistent with those by Aranda & Lobo (2011) and Pottier *et al.* (2013). Even if the individual SDMs have reasonably good independent evaluations, each of them nevertheless contains substantial errors that can be unevenly distributed among species and along environmental gradients (Pottier *et al*., 2013). By stacking SDMs, small errors in many individual species models can accumulate into quite large errors in the community predictions, degrading assemblage predictions accordingly (Pottier *et al.*, 2013). In this regard, the values of the Sørensen index of community similarity obtained in this first formal test of the SESAM framework – above 0.5 – can be considered a reasonable first achievement. A correction for the probability values based on the true species richness has been recently proposed by Calabrese *et al.* (2014). Their maximum likelihood approach, however, still does not allow the determination of which species in the list of probabilities will enter the final community. The error propagation could be even more severe if the single species predictions were binarized before reconstructing the community composition, because the choice of a threshold can matter (Liu *et al*., 2005). Moreover, a statistical bias was recently proposed as the main cause of the general overprediction in richness estimation showed by summing binary SDMs (e.g. Calabrese *et al*., 2014). As just discussed, we acknowledge the fact that stacking binary SDMs could add biases to the community prediction, but on the other hand it has the strength to allow an easy identification of the component species. Predicting assemblage composition over probabilities is still largely wished and applied, especially in conservation studies (e.g. Faleiro *et al.*, 2013; Leach *et al*., 2013). In order to partially control for the additional uncertainty introduced by thresholding, we ran all our analyses using both AUC and TSS threshold maximization metrics. The results of both analyses were consistent and therefore we can be confident that our outcomes are not too sensitive to this threshold choice.

The possibility of predicting species composition in a probabilistic way, without thresholding, holds the promise of reducing methodological biases, but it is still an unresolved issue that will need further developments. In the test of the ‘probability ranking’ rule, we proposed one solution, which avoids the binary transformation of SDM predictions, while still maintaining information about species composition. We did this by selecting a number of species equal to the prediction of species richness on the basis of decreasing probability of presence calculated by SDMs. Predictions of species composition is a great challenge for community ecologists and not many applicable solutions have been proposed (e.g. Webb *et al.*, 2010; Shipley *et al.*, 2011; Laughlin *et al.*, 2012). Our results thus provide new insights to achieve this goal by using SDMs, while avoiding the statistical bias potentially occurring when stacking binary SDM predictions (Calabrese *et al.*, 2014). Yet, several issues still need to be resolved; in particular, new approaches are needed to decrease rates of omission error in SDMs and in the resulting community predictions. One route to improve compositional predictions could come from producing single species models that are more efficient at predicting presences correctly (i.e. limiting omission errors by optimizing sensitivity). A source of omission errors in our case may come from limitations related to the environmental predictors and resolution used to build the SDMs

(Pradervand *et al.*, 2014). Available predictors can themselves include some level of errors (e.g. from measurement, interpolation, calculation) and other important predictors (see below) may be missing in the underlying SDMs (Austin & Van Neil, 2010). As a result, species' realized niches are likely to be incompletely described and some suitable or unsuitable situations for a species cannot be captured in the model. Two recent papers have shown similar problems of assemblage predictions in the case of butterflies and plants, respectively (Pellissier *et al.*, 2012; Pottier *et al.*, 2013). In both cases, the sensitivity (true-positive rate) of assemblage predictions was lower at higher elevations, which was probably due to the more fragmented, mosaic-like environmental conditions there and to missing substrate predictors (e.g. rock type, soil depth). Regarding our study area, snow cover and geomorphology (Randin *et al*., 2009), soil moisture and soil temperature (Le Roux *et al*., 2013), as well as edaphic conditions (Dubuis *et al*., 2012) and finer micro-climatic measurements (Pradervand *et al.*, 2014), are potential missing predictors that could contribute to improve SDMs and hence the resulting community composition predictions. Yet, these missing predictors are currently not available or only available for some plots, and none of them exist in a spatially explicit way to support the final predictions to be generalized to the whole study area.

Conclusions and future perspectives

In the last decade, the range of possible approaches to model species communities has been expanding. Remarkably, most of the very recent solutions agree on the idea of combining complementary approaches into a single framework, as we did here with SESAM (e.g. Webb *et al*., 2010; Mokany *et al*., 2012; Fernandez *et al*., 2013). A framework approach has a number of highly desirable characteristics, in particular the flexibility to integrate different drivers and processes to represent the complexity of factors that influence community assembly and the possibility to couple strengths of different pre-existing techniques in a unique workflow. Community ecology research is in continuous development and any new technical improvement coming from theoretical advances could be promptly accommodated in a framework approach. For instance, in this study we tested the integration of two types of macroecological models, but other recent implementation could also be used, such as the use of sum of predicted species probabilities (Dubuis *et al*., 2011; Calabrese *et al*., 2014). Another innovative way to model species categories would be the species archetypes model (SAM; Dustan *et al*., 2011; Hui *et al*., 2013), which predicts communities using a finite mixture of regression model, on the basis of common responses to environmental gradients. Also, macroecological models not based on correlative statistics could be included to explicitly incorporate the mechanisms responsible for the observed distributions (e.g. Gotelli *et al*., 2009).

Among the great challenges in predictive community ecology is the inclusion of biotic rules. This has been repeatedly attempted in simple SDMs (e.g. by adding other species or simple biotic variables as predictors of the modelled species) with the result of improving significantly the predictions (reviewed in Kissling *et al.*, 2012 and Wisz *et al.*, 2013). In contrast, community-level models most often incorporate the effect of biotic interactions indirectly by considering synthetic community attributes (as we did in this study), while only in a few cases were biotic interactions accounted for in an explicit fashion (e.g. Laughlin *et al*., 2012; Fernandes *et al*., 2013; Pellissier *et al*., 2013). This gap could partly be explained by the shortage of data available to characterize interactions among species in diverse communities (Araújo *et al*., 2011). A potential way to overcome the lack of biotic interaction information could be the analysis of the spatial patterns of geographical overlap in the distributions of species. These can inform about potential interactions between species but approaches to control for species habitat requirements should be applied in co-occurrence analyses to correctly infer biotic interactions from observed patterns (e.g. Gotelli *et al*., 1997; Peres-Neto *et al*., 2001; Ovaskainen *et al*., 2010). Because considering each pairwise interaction as a separate process is difficult, some alternative solutions to reduce this complexity have been also suggested, such as the analysis of separate smaller 'community modules' (as applied in food web analyses; Gilman *et al*., 2010), or the use of proxies of interactions ('interaction currencies') based on measures of non-consumable environmental conditions (described in Kissling *et al*., 2012).

The implementation of the full SESAM framework, i.e. implementing the 'step 4' through the definition of biotic assembly rules coming from empirical patterns of co-occurrence or experiments, could represent a promising route to further define the group of species that can coexist at each site, and help decreasing the rate of omission error. This fourth component of the framework has not been tested in an ecologically explicit way in this study, although using ranked probabilities of occurrence per site can be considered a form of implicit biotic rules. Identifying and quantifying other biotic assembly rules that can be applied generally along wide environmental gradients appears still to be difficult given our current state of knowledge and the heterogeneity of approaches used (Götzenberger *et al*., 2012; Kissling *et al*., 2012; Wisz *et al*., 2013), but it constitutes a necessary target if we want to improve our capacity to predict assemblages in space and time.

Further important drivers of community assembly are stochastic processes, associated with environmental disturbance and demographic dynamics within local and regional species pools (Dornelas *et al*., 2006). The potential presence of stochastic effects would deviate the community assemblage process from being fully deterministic, i.e. from yielding a specific community configuration for a given environmental combination and species pool, but instead be probabilistic so that the projections could for instance consist of a density function of various possible end compositions (Ozinga *et al*., 2005; Shipley, 2010; Pellissier *et al*., 2012; Pottier *et al*., 2013). Therefore, assemblage composition will always entail some level of prediction errors. In this regard, what would prove useful in future studies would be to understand and discern better the different sources of errors in the single techniques integrated in the SESAM framework. In particular, it would be useful to assess how errors propagate from individual SDMs to S-SDMs, and what value of the

Sørensen index (or other evaluation metric of community similarity) would qualify as a fair value of assemblage prediction. This will help estimate the level of similarity and reliability with which one can ultimately expect species assemblages to be successfully predicted, and how far the latter may contribute to a better understanding and prediction of community assembly in space and time (Hortal *et al*., 2012).

ACKNOWLEDGEMENTS

We are grateful to the numerous people who contributed to the data collection and to Pascal Vittoz and Glenn Litsios for insightful discussions and comments on the manuscript. We would like also to thank B. Cade and M. Geraci for their useful insights on the statistics of trait analyses. This study was supported by the Swiss National Science Foundation (grant no. 31003A-125145 to A. Guisan), by the FP6 Ecochange project of the European Commission (grant GOCE-CT-2007–036866). M. D'Amen was supported by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (FP7-PEOPLE-2012-IEF, SESAM-ZOOL 327987).

Chapter 2.2

How to best threshold and validate stacked species assemblages? Community optimisation might hold the answer

This paper is published in *Methods in Ecology and Evolution*

My contribution to the paper: This paper was a collaboration with other members of my group, led by Daniel Scherrer. I actively participated in discussions and was responsible for analysing the plant data, also critically contributing to the literature review and in the writing of the different versions of the manuscript.

How to best threshold and validate stacked species assemblages? Community optimisation might hold the answer

Daniel Scherrer[a], Manuela D'Amen[a], Rui F. Fernandes[a], Rubén G. Mateo[a,b] and Antoine Guisan[a,c]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland

[b] ETSI de Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040, Madrid, Spain.

[c] Institute of Earth Surface Dynamics, University of Lausanne, Géopolis, CH-1015 Lausanne, Switzerland

ABSTRACT

The popularity of species distribution models (SDMs) and the associated stacked species distribution models (S-SDMs), as tools for community ecologists, largely increased in recent years. However, while some consensus was reached about the best methods to threshold and evaluate individual SDMs, little agreement exists on how to best assemble individual SDMs into communities, i.e. how to build and assess S-SDM predictions. Here, we used published data of insects and plants collected within the same study region to test (1) if the most established thresholding methods to optimize single species prediction are also the best choice for predicting species assemblage composition, or if community-based thresholding can be a better alternative, and (2) whether the optimal thresholding method depends on taxa, prevalence distribution and/or species richness. Based on a comparison of different evaluation approaches we provide guidelines for a robust community cross-validation framework, to use if spatial or temporal independent data are unavailable. Our results showed that the selection of the "optimal" assembly strategy mostly depends on the evaluation approach rather than taxa, prevalence distribution, regional species pool or species richness. If evaluated with independent data or reliable cross-validation, community-based thresholding seems superior compared to single species optimisation. However, many published studies did not evaluate community projections with independent data, often leading to overoptimistic community evaluation metrics based on single species optimisation. The fact that most of the reviewed S-SDM studies reported over-fitted community evaluation metrics highlights the importance of developing clear evaluation guidelines for community models. Here, we move a first step in this direction, providing a framework for cross-validation at the community level.

KEYWORDS: Community cross-validation, community modeling, ecological niche models, species distribution modeling, species distribution models, thresholding methods

INTRODUCTION

Past and future environmental changes may not only lead to shifts in species distributions (e.g., Parmesan & Yohe, 2003; Thuiller *et al.*, 2005; Dullinger *et al.*, 2012), but also to changes in species assemblages and interactions (e.g., Zachos *et al.*, 2008; Van der Putten *et al.*, 2010; Nogues-Bravo & Rahbek, 2011; Blois *et al.*, 2013). Information about communities, here defined as a taxonomic assemblage of distinct populations of species that co-occur in a given space at a given time (Begon *et al.*, 1996), is therefore essential to make informed decisions for conservation prioritisation (D'Amen *et al.*, 2011; Guisan *et al.*, 2013; Mateo *et al.*, 2013) and to create biodiversity indices (e.g., Essential Biodiversity Variables; Pereira *et al.*, 2013) for policy decisions (Hancock *et al.*, 2011; Jones *et al.*, 2013).

Different approaches to model communities are available, using either correlative (e.g., Ferrier & Guisan, 2006; Guisan & Rahbek, 2011) or mechanistic techniques (e.g., Kearney & Porter, 2009; Mokany & Ferrier, 2011), with some predicting only macro-ecological properties such as species richness (e.g., Gotelli *et al.*, 2009; Dubuis *et al.*, 2011; Kremer *et al.*, 2012) and others also predicting community composition (see Manel *et al.*, 2010 for a review). In this study, we focused on correlative approaches based on individual species distribution models (SDMs), as they are the most common technique applied to conservation strategies (Guisan *et al.*, 2013), and to predict future patterns of biodiversity in the face of global change (Manel *et al.*, 2010; Nogues-Bravo & Rahbek, 2011). Niche-based SDMs quantify the relationship between available species occurrences and different environmental factors to analyse and predict distributional patterns (Guisan & Thuiller, 2005; Elith & Leathwick, 2009; Vilhena & Antonelli, 2014). By additionally stacking individual SDMs (S-SDMs), one can produce spatiotemporal projections of species richness and composition (Ferrier & Guisan, 2006; Guisan & Rahbek, 2011).

While there is a vast and now long-standing literature on advances and limitations of single species predictions (e.g., Guisan & Thuiller, 2005; Guisan *et al.*, 2006; Maggini *et al.*, 2006; Elith & Leathwick, 2009; Meier *et al.*, 2010; Zimmermann *et al.*, 2010; Merow *et al.*, 2014), studies exploring how to improve community predictions based on aggregated information from individual SDMs emerged more recently (e.g., Mateo *et al.*, 2012; Benito *et al.*, 2013; Manel & Holderegger, 2013; but see Ferrier et al. 2002; Cord *et al.*, 2014). A fundamental difference among the proposed solutions is whether to maintain the information on species composition in the final predictions. For instance, the simple sum of probabilities of individual SDM predictions usually gives better estimates of species richness, but the information on species identity is lost (Dubuis *et al.*, 2011; Calabrese *et al.*, 2013). Therefore, predictions of community composition have mainly been achieved so far by thresholding the individual continuous SDM predictions (e.g., probability or suitability index) to obtain binary maps (Kiehl, 2011) and then stacking the latter at the assemblage level (e.g., Guralnick *et al.*, 2007; Pottier *et al.*, 2013; D'Amen *et al.*, 2015).

There are several examples in the literature of optimizing thresholding methods for single species predictions (e.g., Liu *et al.*, 2005; Jimenez-Valverde & Lobo, 2007; Freeman & Moisen, 2008; Kiehl, 2011). These led to a mounting consensus about the most appropriate methods, with the majority of SDM studies published nowadays using either an approach maximising the true skills statistics (Max.TSS) or based on the curve in a receiver operating characteristic plot (Opt.ROC, related to AUC) (see Vilhena & Antonelli, 2014; Table S1). However, the threshold selection can strongly influence the reliability of the predicted richness and composition of S-SDMs assemblages (Pineda & Lobo, 2009; Benito *et al.*, 2013). It is thus relevant to explore which thresholding approach provides the best performance in assemblage estimates, and if alternatives exist that can improve the assemblage prediction from individual SDMs.

Studies focussing on S-SDMs tend to over-predict species richness when based on (thresholded) binary predictions (e.g., Herzog, 2005; Pineda & Lobo, 2009; Dubuis *et al.*, 2011; Mateo *et al.*, 2012; Pottier *et al.*, 2013), with some exceptions (e.g., D'Amen *et al.*, 2015; Distler *et al.*, 2015). Different factors have been proposed to explain this over-prediction: (1) a statistical bias in thresholding site-level occurrence probabilities for each species (Calabrese *et al.*, 2013); (2) the implicit assumption of unsaturated communities not assuming an ecological limit for species numbers in assemblages (environmental carrying capacity; Guisan & Rahbek, 2011); (3) the lack of considering different constraints on community composition (i.e., ecological, evolutionary, historical, or biological biodiversity drivers; see Mateo *et al.*, 2017).

The commonly used approach to get binary maps from continuous SDM predictions is to use a species-specific threshold, i.e. each species has a single threshold across all sites ("species threshold", Calabrese *et al.*, 2013). Recently, another community-based approach, called probability ranking rule (PRR), was proposed to predict assemblage composition from individual SDMs (Guralnick *et al.*, 2007). This method does not require a species-specific threshold, therefore preventing over-prediction, but site-by-site ecological constraints (e.g., macro-ecological models) are applied to assemblages to predict species richness (“site-threshold”).

Surprisingly, studies aiming to test and improve S-SDM have used very different approaches to evaluate the predicted assemblages (Herzog, 2005; Alvarez *et al.*, 2009; Schoville *et al.*, 2012; Cord *et al.*, 2014; Zurell *et al.*, 2016) and this evaluation aspect of the community modelling procedure has not yet received all the attention it deserves. In most studies, assemblage predictions are not adequately evaluated because the data used for the evaluation were already used for individual model fitting, not allowing anymore a correct cross-validation at the community level. Ideally, the best evaluation method should use spatial or temporal independent data (Elith *et al.*, 2006; Vilhena & Antonelli, 2014), but if not available, an appropriate cross-validation approach should at least be set up.

Here, we used published high-resolution data of insects (butterflies and grasshoppers) and plants (forests and grasslands sites), collected within the same study region to (1) test if the most established thresholding methods for optimal single species prediction (i.e., Max.TSS and Opt.ROC) are also the best choice for species assemblages, (2) investigate if the optimal thresholding method depends on taxa, prevalence distribution (Allouche *et al.*, 2006), and/or species richness and (3) provide guidelines for a correct community cross-validation framework, to be used if spatially- or temporally-independent data are unavailable.

MATERIALS AND METHODS

Community data and environmental variables

Study area

The data on all taxa were collected within the same study area located in the western Swiss Alps of the canton Vaud (Fig. 1; 46°10′ to 46°30′ N; 6°50’ to 7°10’ E), covering an area of ca. 700 km^2, with elevation ranging from 375 to 3210 m a.s.l. and forested areas up to 1900 m a.s.l. For centuries, agriculture (farming and pasturing) has maintained grasslands among forests and altered the position of the treeline. The highly variable topography and diverse land use of the study area, in combination with our high-resolution environmental data (25 x 25 m cell size), provide a huge range of complex species-environment relationships to test our modelling framework.

Plant data

The forest data were part of a forest inventory of the canton Vaud conducted between 1988 and 2002 (mostly 1990 to 1994) and consisted of 3076 sites. The forest sites were distributed on a 400 m grid all across the forested area of the canton and had a circular area of 314 m^2 (Fig. 1; for details see Hartmann *et al.*, 2009). In total, 703 plant species were recorded, but only 312 (44%) had enough occurrence data (> 20 occurrences) across the dataset for modelling purposes (see Table 1 for more detailed statistics on the datasets).

The grassland dataset was collected between 2002 and 2009 following an equal random-stratified sampling of non-forested areas in the study area. In total, 911 vegetation sites of 4 m^2 were sampled (Fig. 1; for more information see Dubuis *et al.*, 2011). A total of 905 plant species were recorded but only the 212 most frequent (>20 occurrences) were selected for modelling (Table 1).

To predict the distribution of the plant species we used five environmental variables: growing degree-day (above 0 °C), moisture index over the growing season (difference between precipitation and potential evapotranspiration), the sum of potential solar radiation over the year, slope (in degrees), and topographic position (unit-less, indicating the ridges and valleys). All these variables were at a 25 m resolution and have been shown to be useful predictors for plant species in mountain environments (see Guralnick *et al.*, 2007; Dubuis *et al.*, 2011; Jaramillo & Cárdenas, 2013 for details on predictors).

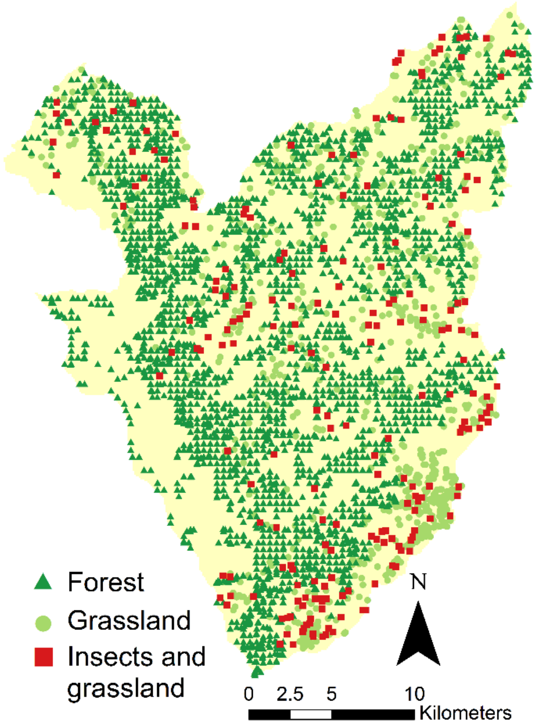


FIGURE 1 Map of the study area with the forested sites (dark green triangles, N=3076), the grassland sites (light green circles and red squares, N=903) and the insect sites (red squares, butterflies N=192, grasshoppers N=202).

Table 1 Basic statistics of the data sets used for the case study and the evaluation metrics (AUC) for the individual species distribution models using the three different community evaluation approaches.

Data set	**Number of species modelled (recorded)**	**Prevalence (mean ± sd)**	**Species richness (mean ± sd)**	**AUC SSV (mean ± sd)**	**AUC ID (mean ± sd)**	**AUC CCV (mean ± sd)**
Forest	312 (703)	0.044 ± 0.090	29.5 ± 11.8	0.80 ± 0.09	0.80 ± 0.08	0.79 ± 0.09
Grassland	212 (905)	0.098 ± 0.089	23.5 ± 13.8	0.82 ± 0.07	0.83 ± 0.06	0.81 ± 0.06
Butterflies	77 (131)	0.235 ± 0.137	18.1 ± 9.2	0.76 ± 0.10	0.75 ± 0.12	0.76 ± 0.10
Grasshoppers	20 (41)	0.256 ± 0.193	5.1 ± 3.3	0.84 ± 0.07	0.86 ± 0.08	0.84 ± 0.06

Insect data

Data on butterflies and grasshoppers were respectively collected in 192 and 202 squares of 50 m x 50 m across all the elevational range of the study area (Fig. 1; see Pellissier *et al.*, 2012; Pradervand *et al.*, 2013, for more information). In total, 131 butterfly and 41 grasshopper species were observed, but due to model limitations only the most common 77 butterfly and 20 grasshopper species (>=20 occurrences) were considered for modelling (Table 1).

For our SDMs we used the same predictors as D'Amen *et al.* (2015): four bioclimatic variables (solar radiation, summer temperature, annual degree-days and annual average number of frost days during the growing season), an index of vegetation productivity, i.e. normalized difference vegetation index (as proxies for trophic resources), and the distance to forest. These variables were selected as they are not highly correlated (<0.7; Dormann *et al.*, 2013) and considered ecologically important for insects (e.g., Thuiller, 2005; Grytnes & Beaman, 2006).

The modelling framework

Our modelling framework used three different S-SDM based community modelling pathways ("single species cross-validation", "independent data" and "community cross-validation) representing the most commonly reported practices in the literature (see Fig. 2 and "Evaluating community predictions" section).

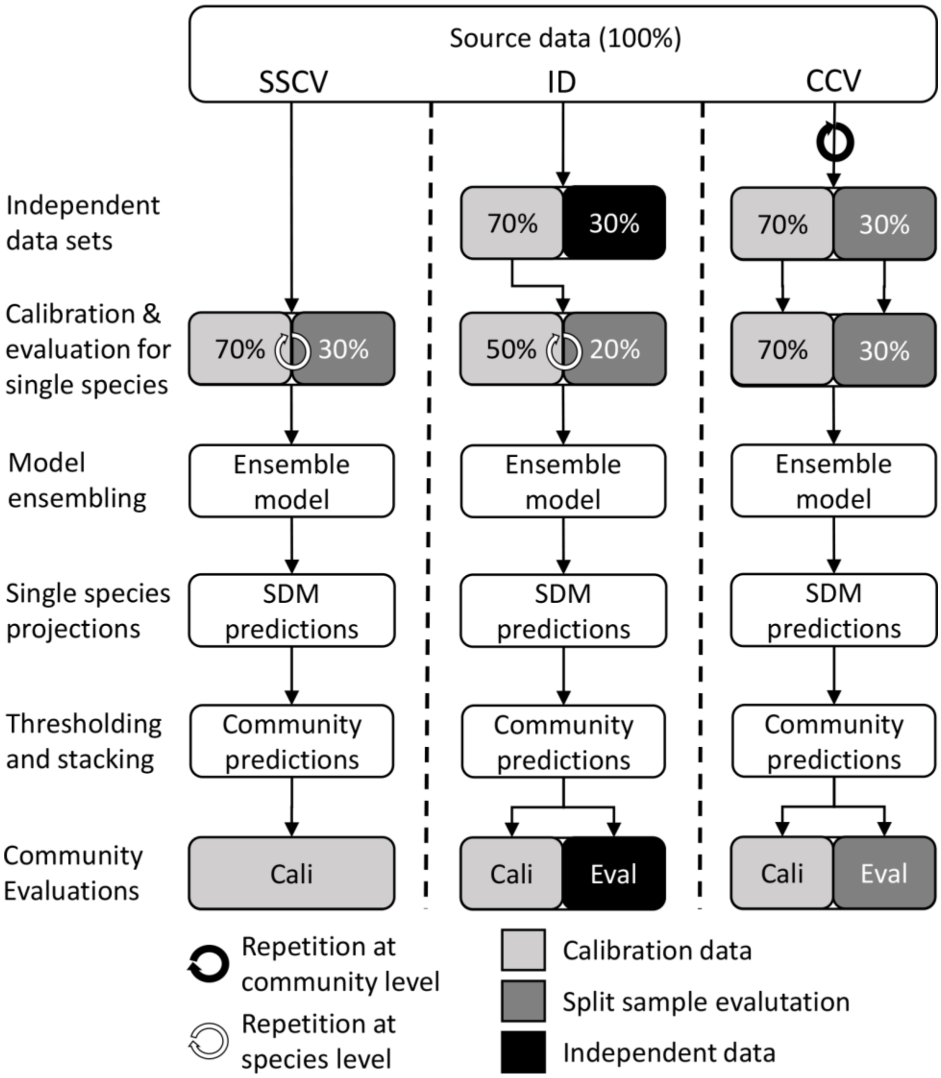


FIGURE 2 The modelling framework illustrating the three different community modelling approaches: (a) "single species cross-validation" (SSCV) on the left, (b) "independent data" (ID) in the middle and (c) "community cross-validation" (CCV) on the right.

Single species modelling, thresholding and evaluation

Individual species models were run by generalised linear models (GLM; McCullagh & Nelder, 1989), generalised additive models (GAM; Hastie & Tibshirani, 1990), random forest (RF; Breiman, 2001) and boosted regression trees (BRT; Elith *et al.*, 2008). Models for species with more than 50 occurrences were fitted by simple SDMs using all five selected predictors, followed by a weighted

(AUC) ensemble forecast (Marmion *et al.*, 2009). Species having only between 20 and 50 occurrence records were fitted by an ensemble bivariate approach optimised for rare or under-sampled species (Grytnes, 2003; Lomba *et al.,* 2010): individual models were calibrated on bivariate combinations of the selected predictors with all four modelling techniques, followed by a consensus forecast from all the resulting "small models" weighted by their AUC scores. We used a repeated split-sample procedure (N=25) for model evaluation, followed by a weighted (AUC) ensemble forecast (across techniques and split-sample runs).

The projected probability outputs of the ensemble models were binarised using two thresholding schemes: (1) *species-specific-thresholds* (a single threshold calculated for each species) and (2) *site-specific-thresholds* (differing for each site on the basis of additional community information, i.e. species richness predictions). We selected seven different species-specific-thresholding techniques, which can be classified in four major groups: single-index based, sensitivity and specificity combined, model-building data-only-based, and predicted probability-based (see Table S1; Liu *et al.*, 2005; Nenzen & Araujo, 2011 for details on classification). As the thresholding techniques showed minimal within-group variance (see Figure S1 and S2), we decided to only present the results for one thresholding technique per group in the main manuscript. The chosen techniques were: Cohen's Kappa maximization approach (*Max.Kappa*; single-index based), TSS maximization approach (*Max.TSS*, sensitivity and specificity combined), observed prevalence (*Obs.Preval*; model-building data-only-based approach), and average probability approach (*AvgProb*; predicted probability-based approach; for details on techniques see Table S1). In addition, we applied two site-thresholds (community-based approaches) using species richness (SR) predictions in combination with a probability ranking rule (PRR). These methods selected a number of species equal to the predicted SR on the basis of decreasing probabilities of presence calculated by the SDMs (Guralnick *et al.*, 2007; D'Amen *et al.*, 2015). Therefore, the species with the highest probabilities in a site are selected (considered present) in decreasing order until the SR predicted for the site is reached. The SR predictions were derived by either summing the per site probabilities of individual SDMs, obtaining a prediction of richness for each site (pS-SDM; Dubuis *et al.*, 2011) or by a macro-ecological model (MEM; see D'Amen *et al.*, 2015 for details), directly modelling the richness of the sites. As results from the two site-thresholds were concordant, we only show here the former (*pS-SDM+PRR*).

To evaluate the threshold independent performance of our individual species models, the area under the curve of a Receiver-Operating Characteristic (ROC) plot (AUC; Fielding & Bell, 1997) was calculated based on a repeated split sampling cross-validation (Thuiller *et al.*, 2013). Additionally, based on our independent/cross-validation data we calculated five threshold dependent metrics for each thresholding technique: the overall accuracy (PCC; i.e. proportion of correctly classified presence and absences; Fielding & Bell, 1997), sensitivity (proportion of correctly predicted presences), specificity (proportion of correctly predicted absences), the true skill statistic (i.e. [(sensitivity +

specificity) -1]; TSS; Allouche *et al.*, 2006) and Cohen's Kappa (Kappa; i.e., overall accuracy but corrected for chance performance; Cohen, 1968).

Evaluating community predictions

All the community predictions were built by stacking binary SDMs of individual species (S-SDMs; Dubuis *et al.*, 2011; Guisan & Rahbek, 2011). The three modelling pathways (Fig. 2) were identical regarding the modelling procedure for single species, thresholding and community assemblage and only varied in the selection of the data for community calibration and evaluation.

- The "single species cross-validation" (SSCV) approach (Fig. 2a) has not fully "unused/independent" data for community evaluation (i.e. sites not used for the calibration of any single species). Here, in the process of the cross-validation of all individual SDMs (i.e. across all species), different sites are selected at each resampling iteration and for each species, so that all sites are most likely used in at least one split-sampling run and their information incorporated in the final ensemble model. This approach cannot thus be considered based on fully independent data. The SSCV approach has been to date the most common way to model and evaluate communities' predictions based on S-SDMs (Fig. 2a; e.g., Dubuis *et al.*, 2011; Calabrese *et al.*, 2013; D'Amen *et al.*, 2015; Distler *et al.*, 2015). As no independent data is set aside for community evaluation, this approach usually gets evaluated with all the sites used for calibration. However, to avoid bias in the results due to different numbers of evaluation sites, we evaluated the SSCV approach only on 30% of the available sites (identical to the ID and CCV approach below).
- The (spatial or temporal) "independent data" (ID) approach (Fig. 2b) starts with two completely independent datasets. One is used for the calibration of the SDMs (i.e. 70% of the sites) and the other set is used (only) to evaluate the performance of the community predictions (i.e. 30% of the sites; Fig 2b; e.g., Guralnick *et al.*, 2007; Benito *et al.*, 2013; Pottier *et al.*, 2013; Cord *et al.*, 2014; Zurell *et al.*, 2016).
- The "community cross-validation" (CCV) approach (Fig. 2c) uses a repeated split sampling of sites (100 repetitions) dividing the available sites into calibration (70%) and evaluation sets (30%) to perform all the modelling procedure from the single species prediction to the community assembly (Fig. 2c). In contrast to the previous ID pathway (above), which only uses one (spatial or temporal) fixed independent evaluation dataset, in the CCV approach all SDMs are fitted at each split-sample iteration using the same training and test sets for all species, thus minimizing the risk of bias in the evaluation data (i.e. if the training and test sets differ across species, as in the ID approach). This repeated cross-validation also allows the estimation/simulation of confidence intervals for community predictions instead of just a single value per community. To our knowledge, no study used this community cross-validation method so far.

To compare the community model performance among thresholding techniques and modelling pathways, we calculated eight different community agreement metrics: 1) the deviation of the predicted from the observed species richness (SR.deviation), 2) the proportion of species correctly predicted as present (community sensitivity), 3) the proportion of species correctly predicted as absent (community specificity), 4) community accuracy (PCC; i.e. the percent correctly classified species, present or absent), 5) the community TSS (here measured for a site across all species, rather than for a species across all sites as in single SDM evaluation; Pottier *et al.*, 2013) , 6) the community kappa (same as for TSS, for a site across species; Pottier *et al.*, 2013), and 7) the Sørensen similarity (Sørensen, 1948).

Correlation of single species and community evaluation metrics

For each combination of dataset, modelling pathway and thresholding method (4 x 3 x 9 = 108) we calculated the average evaluation metric for all five single species metrics and all seven community metrics. We then calculated the Spearman correlation of all possible combinations of our five single species and seven community evaluation metrics. The resulting correlation matrix tells us if methods (modelling pathways or thresholding methods) that yield the highest scores in a certain single species metric also yield the highest score in the corresponding community evaluation metric.

RESULTS

Performance of individual SDMs

As expected the evaluation scores of the individual SDMs were similar to earlier studies published with the same data (Guralnick *et al.*, 2007; Jaramillo & Cárdenas, 2013; D'Amen *et al.*, 2015) and their performance was not affected by the chosen community evaluation approach (Table 1, Table S3). Despite their differences in site SR, prevalence distribution and species pool the average performance of individual SDMs was similar across all taxa (Table 1, Table S3). Additionally, the often-reported effect of species prevalence on model performance was only marginal in our study, with rare and common species having similar average model performance within a given taxonomic group (Fig. S3).

Correlation of single species and community evaluation metrics

The correlation between the single species and corresponding community metrics was highest (cor > 0.93; Table 2) for some combinations of metrics based on partial information from the contingency table comparing predictions to observations (i.e. PCC, specificity and sensitivity) and considerably lower for the metrics accounting for all dimensions of the contingency table, such as TSS and Cohen’s Kappa (cor = 0.73; Table 2). Correlations between non-corresponding single species and community metrics (i.e. Sørensen and SR deviation) tended to be even lower, with the exception of Kappa versus Sørensen (Table 2).

TABLE 2 Pearson Correlation of single species and community evaluation statistics. The asterisks indicate the significance level (* p<0.05, ** p<0.01, *** p<0.001)

	Community metrics						
Single species	**Accuracy**	**Sensitivity**	**Specificity**	**KAPPA**	**TSS**	**Sørensen similarity**	**SR deviation**
Accuracy	**1.00 *** **	-0.37 *	0.95 ***	0.70 ***	0.37 *	0.37 *	-0.58 ***
Sensitivity	-0.36 **	**0.93 *** **	-0.54 ***	0.01 n.s.	0.56 ***	0.18 n.s.	-0.44 ***
Specificity	0.97 ***	-0.53 ***	**0.99 *** **	0.64 ***	0.20 n.s.	0.31 *	-0.63 ***
KAPPA	0.41 **	0.50 *	0.27 *	**0.79 *** **	0.72 ***	0.82 ***	-0.3 *
TSS	0.06 n.s.	0.85 ***	-0.14 n.s.	0.35 n.s.	**0.79 *** **	0.38 **	-0.20 n.s.

Species richness and compositional similarity

The deviation in species richness between observed and predicted communities was strongly dependent on the chosen thresholding method (Fig. 3). The thresholding approach that uses the average predicted probability (*AvgProb*) showed the highest amount of over-prediction followed by the combined sensitivity and specificity approach (*Max.TSS*). The other three thresholding methods (*Preval*, *Max.Kappa* and *pS-SDM+PRR*) performed very similar and showed overall no tendency to over-predict species richness. There were no significant differences between the three modelling pathways for any of the studied taxa (Fig. 3). The absolute number of over-predicted species was strongly related to the average number of species per plot (SR) and therefore differed among the taxa (Fig. 3). However, when corrected for the differences in SR the over-prediction did not significantly vary anymore across taxa.

The compositional similarity (Sørensen similarity index) varied significantly both among thresholding techniques and modelling pathways (Fig. 4). The compositional similarity was expectedly always much higher with the “single species cross-validation” (SSCV) pathway compared to the “independent data” (ID) or the “community cross-validation” (CCV) pathways, which both performed similarly. There was also a strong interaction between modelling pathway and thresholding technique. Using the SSCV pathway, thresholding by *Obs.Preval* and by *Max.Kappa* performed better (Fig. 4). However, if independent sites were available for the community evaluations (ID and CCV pathways), the community-based approaches (*pS-SDM+PRR*) performed better than the *Obs.Preval* and *Max.Kappa* thresholds (Fig. 4). The similarity between predicted and observed communities was higher in the two insect datasets than in the two plant datasets (Fig. 4), which is likely due to the lower number of insect species compared to plant species modelled. Surprisingly, the most established thresholding methods for single species SDMs based on sensitivity and specificity (i.e. *Max.TSS*, *Opt.ROC* and *SenSpec*;

Fig. 4 and Fig. S1 and S2) never ranked highest, as one or more of the other thresholding method always ranked above them, both for community composition and for species richness.

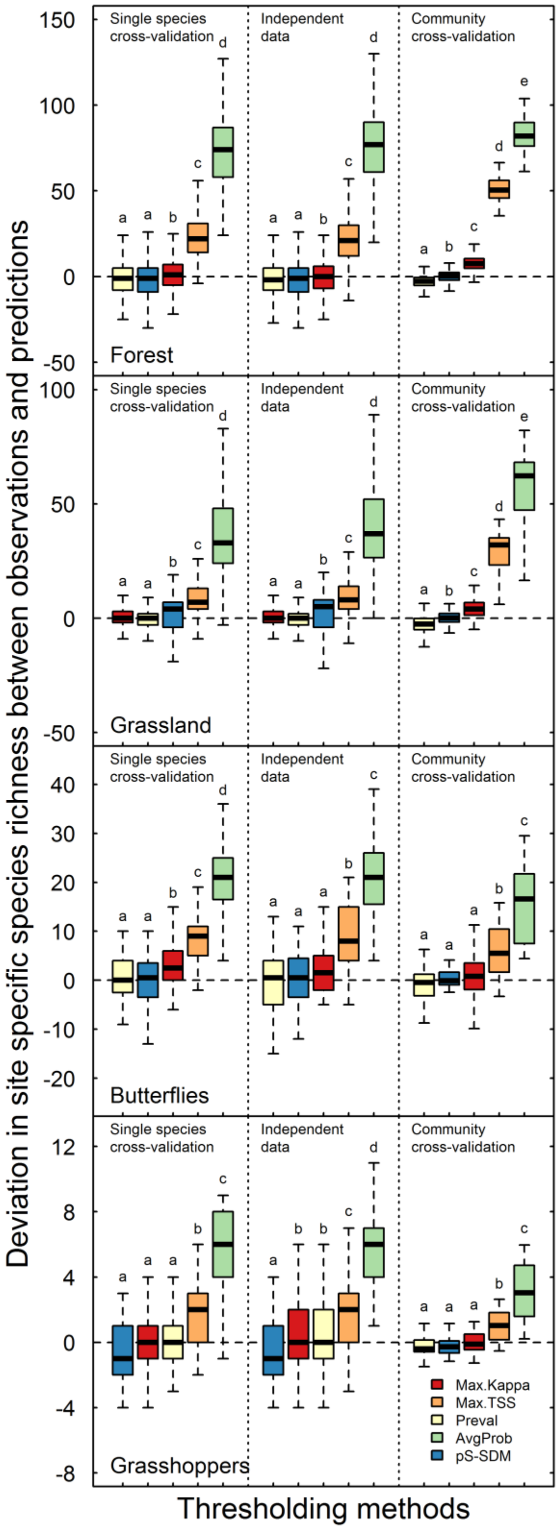


FIGURE 3 Deviation in site specific species richness between observations and predictions for the four different datasets (top to bottom) and the three different modelling pathways (left to right). The boxplots are sorted by the median and the colours indicate the different thresholding techniques used to binarise predictions. The line in the box indicates the median, boxes range from the 25th to the 75th percentile and the whiskers indicate ± 2 standard deviations. Letters above the boxplots indicate significant differences (Wilcoxon rank sum test, $p < 0.05$).

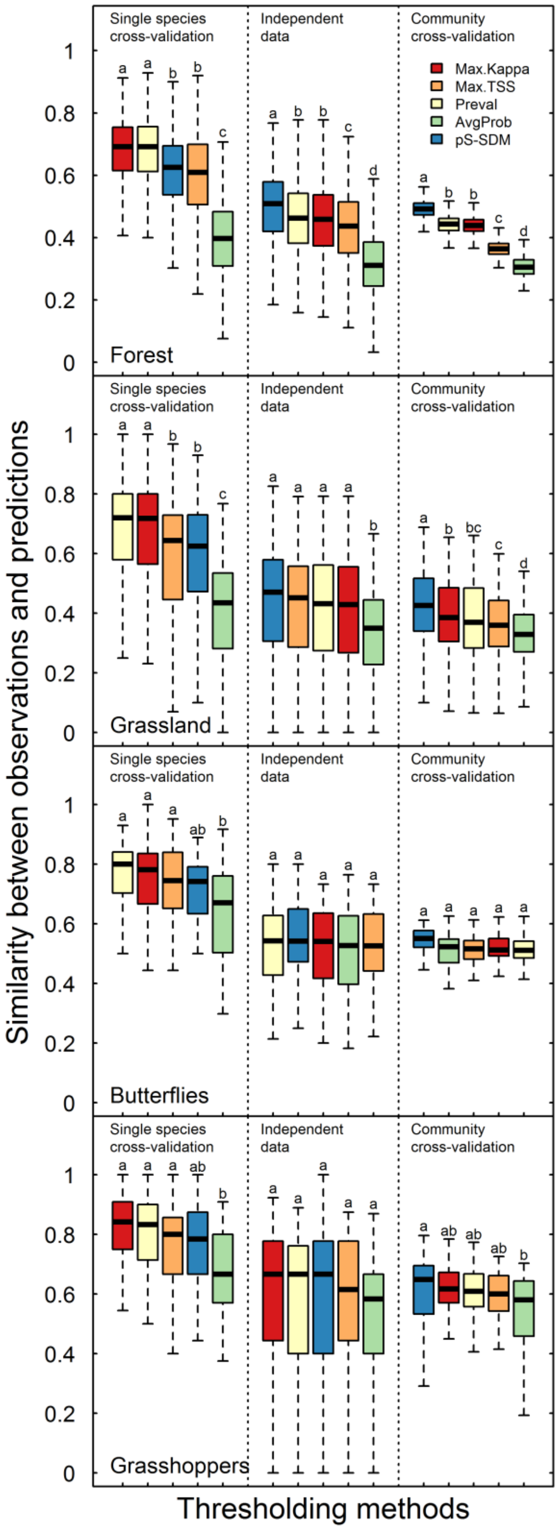


FIGURE 4 Sørensen similarity between observations and predictions for the four different datasets (top to bottom) and the three different modelling pathways (left to right). The boxplots are sorted by the median and the colours indicate the different thresholding techniques. The line in the box indicates the median, boxes range from the 25th to the 75th percentile and the whiskers indicate ± 2 standard deviations. Letters above the boxplots indicate significant differences (Wilcoxon rank sum test, $p < 0.05$).

DISCUSSION

Do the most established thresholds for single species work as well for community predictions?

In this paper, we asked if the most established methods for single species thresholding are also the optimal choice for making predictions at the community level and if there is a direct link between the individual species predictions and the corresponding community metrics. Our results confirm the existence of such a link for single-index based metrics such as sensitivity, specificity and accuracy. However, these results should be interpreted with caution as maximising sensitivity or specificity can simply be achieved by predicting the species as present or absent (respectively) everywhere. In our study system, most of the modelled species have a low prevalence (i.e. are absent at most sites), thus accuracy (PCC) can often be improved by predicting the species as “absent” nearly everywhere.

The two most commonly used community evaluation metrics, Sørensen similarity index and deviation in species richness, were only weakly correlated with most evaluation metrics used for individual species. The most established thresholding methods for individual species predictions (i.e., *Max.TSS*, *Opt.ROC*, *SenSpec*) did show lower performance when applied to community-level predictions. This is likely due to the fact that both TSS and ROC try to find the best trade-off between sensitivity and specificity (Vilhena & Antonelli, 2014). As most of the species have a prevalence far below 50% (i.e., are absent in many more sites than present), adding a few more presences might have a big effect on the sensitivity (by increasing the chance of finding the few real presences) but only marginally affects the specificity. By definition, increasing sensitivity also increases TSS, but with the drawback of a slight over-prediction. While this might not matter much on a single species basis, for community-level predictions the over-prediction will accumulate when summing binarised maps across all species, leading to the often observed over-estimation of species richness in S-SDMs (e.g., Herzog, 2005; Pineda & Lobo, 2009; Dubuis *et al.*, 2011; Mateo *et al.*, 2012; Pottier *et al.*, 2013; Zurell *et al.*, 2016). It is important to remark, that in the rare case of an ecosystem mostly comprising of widespread species (i.e., prevalence >50 %) this will turn into the opposite as TSS and ROC will optimise absences leading to an underestimation of species richness. The strength of the over/under prediction bias is therefore linked to the prevalence distribution of the modelled species assemblages. However, in the vast majority of natural systems, both the site SR and the regional species pool are driven by a large number of rare (low prevalence species) compared to a few widespread species (Preston, 1948; Magurran & Henderson, 2003).

The community-based thresholding methods based on the selection of the most probable species (through a probability ranking) up to the predicted site richness (*MEM+PRR, pS-SDM+PRR*) can overcome this problem, because they are able to constrain species predictions based on a different value of species richness in each site (i.e. making them site-specific thresholding methods). Therefore, these methods prevent over-prediction while still allowing the analyses of species composition. Our results thus support the conclusion that, when the final goal is to optimize community composition, community-thresholding methods are the best option. Yet, as discussed in the next section, two single-species thresholding methods – *maximized Kappa* and *observed prevalence* – also showed good results for predicting communities (close to the community-based approaches). However, as community-based thresholds combine the optimisation of species richness prediction and a probability ranking rule (PRR), they would always select the species with the highest predicted probabilities in each site (D'Amen *et al.,* 2015). This could seem logic and straightforward, but there might be a bias when the species in the community have varying prevalence (Körner, 2011). In fact, the maximum predicted probability is depending on the prevalence of the species, thus the common species will tend to always have greater maximum predicted probabilities than rare species and, as a result, will be considered present an over-proportionate number of time in the final community compositions. This bias will produce high similarity scores (Sørensen index) in the prediction evaluation, as the most common species are correctly predicted in most sites. However, the drawback is that the rarest species will be often omitted in the community predictions, which can be for instance problematic if the final goal of the modelling exercise has conservation implications.

Is there a "best" threshold for community S-SDMs?

We also tested if different methods for binarising community S-SDMs could be superior depending on the taxonomic group, prevalence distribution or species richness. While we observed significant differences between the different groups (i.e. taxa), there is no simple statistical way to assess if these differences are attributable to the biology of the taxa themselves or simply to the differences in site species richness and prevalence distributions. Nevertheless, when we standardized the deviation in species richness by the total number of modelled species (regional species pool), no significant difference was any more visible among the different taxonomic groups. The differences in species richness deviation seem therefore a direct cause of the regional species pool. The same also seems correct for the Sørensen similarity index, as datasets with higher species richness and species pool have lower similarity scores. This likely results from the fact that the more species need to be predicted correctly, the more difficult it becomes to predict the whole communities.

A similar ranking of thresholding methods was overall observed across taxonomic group within a given modelling pathway, while among the pathways there were clear shifts in the ranking of thresholding methods: with no independent community evaluation data (SSCV), the *Obs.Preval* and *Max.Kappa* threshold showed superior results, while the pathways using independent community evaluation data (ID and CCV) indicated the community-based thresholding to be superior (*pS-SDM+PRR*). This observation is in line with published literature, where studies not using independent community data usually report a good performance of single species optimisations methods (e.g. Alvarez *et al.*, 2009; D'Amen *et al.*, 2015; Distler *et al.*, 2015), while studies using independent data usually have better results using community constraints (e.g. Guralnick *et al.*, 2007). Yet, it is remarkable to notice that, although previously much criticized in the literature (e.g., McPherson *et al.*, 2004; Allouche *et al.*, 2006), maximized Kappa (together here with the observed prevalence) did indeed perform well as a thresholding method for predicting both single species and communities, being nearly always superior to the sensitivity-specificity thresholding methods supporting earlier findings of Manel *et al.* (2001).

It is important to notice that the shift in ranking between modelling pathways was likely due to a lower degree of overfitting and therefore a lower decrease in performance when predicting to independent data.

Summing up: How to evaluate community predictions correctly?

Our results show that the "single species cross-validation" approach (SSCV), the most commonly used in the literature to evaluate community predictions (e.g., Dubuis *et al.*, 2011; Calabrese *et al.*, 2013; Distler *et al.*, 2015), yields overoptimistic and thus not fully realistic measures of predictive power. While this approach is usually able to provide satisfying evaluation for single species, as revealed by the cross-validation of individual species runs, it shows a clear degradation of predictions when measured at the level of communities. This occurs likely because "all" sites are used at least once at some stage across all modelling runs of the split-sampling procedure, and thus no observation (or very few in the best cases) remains fully independent (i.e. unused) for the final evaluation at the community level. Additionally, the sets of training sites used at each run differ among the species, making the results not entirely comparable across species.

The second approach found in the literature builds on the first one (SSCV; thus including an internal cross-validation evaluation), but uses spatially or temporally independent data (ID) for the assessment (thus an external evaluation), thus (unlike SSCV) using the same set of evaluation sites for all species (e.g., Benito *et al.*, 2013; Pottier *et al.*, 2013; Cord *et al.*, 2014). When such independent data are available, this method provides the best possible evaluation, provided that the evaluation data are representative of the area where the models apply. This approach – with both internal and external evaluation - is also the one considered as optimal in James *et al.* (2013), and recently promoted in the field of SDMs by Vilhena and Antonelli (2014).

The third approach (CVV), newly presented here, repeats the ID approach a large number of times within a cross-validation procedure at the community-level (no example of this approach known in the literature). By doing this, the risk of bias in the evaluation data, inherent to the selection of a single evaluation data set, is minimized compared to the simple ID approach. Additionally, the repeated cross-validation allows the assessments of uncertainty and confidence intervals around the community predictions' performance metrics. However, as this approach selects the same sites for all species, its application is only possible under specific circumstances. First, all the species data need to be collected in the same sites (i.e. true 'community data'). Second, as this approach leads to an unequal number of presences/absences between different cross-validation runs for the same species, it can lead to models failing for very rare (low sample size) species in some of the cross-validation runs if not enough presence sites are selected in the training set.

According to our results and despite the potential limitations we advise the use of the proposed community cross-validation approach (CCV) to evaluate community models in future studies. In fact, we clearly showed that the common practice of evaluating the community predictions on the same dataset used for calibration process (SSCV) leads to overoptimistic estimations of model performance. In the commonest case of unavailability of truly spatial (i.e., different region) or temporal (i.e., different sampling period) independent data, often independent datasets are "created" by randomly splitting the initial dataset in two parts. However, we advocate against this practise and instead promote the community cross-validation approach, which minimizes the artefacts of randomly splitting the initial data and allows the estimation of uncertainty associated with the community evaluation metrics.

ACKNOWLEDGEMENTS

This study was supported by the Swiss national Science Foundation (SESAM'ALP project, grant nr 31003A-1528661) to AG and by the European Commission, Marie Skłodowska-Curie Research Fellowship Programme (SESAM-ZOO project) to MDA and AG. R.G.M. was funded by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (ACONITE, PIEF-GA-2013-622620). The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics.

References

Albert, C.H., Thuiller, W., Yoccoz, N.G., Soudant, A., Boucher, F., Saccone, P. & Lavorel, S. (2010) Intraspecific functional variability: extent, structure and sources of variation. *Journal of Ecology*, **98**, 604–613.

Albouy, C., Guilhaumon, F., Araújo, M.B., Mouillot, D. & Leprieur, F. (2012) Combining projected changes in species richness and composition reveals climate change impacts on coastal Mediterranean fish assemblages. *Global Change Biology*, **18**, 2995–3003.

Algar, A.C., Kharouba, H.M., Young, E.R. & Kerr, J.T. (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography,* **32**, 22–33.

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

Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., Schönswetter, P., Taberlet, P., Brodbeck, S., Gaudeul, M. & Gielly, L. (2009) History or ecology? Substrate type as a major driver of patial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632-640.

Aranda, S.C. & Lobo, J.M. (2011) How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. *Ecography*, **34**, 31–38.

Araújo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011) Using species co-occurrence networks to assess the impacts of climate change. *Ecography*, **34**, 897-908.

Austin, M.P. & Van Niel, K.P. (2010) Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, **38**, 1-8.

Baselga, A. & Araújo, M.B. (2009) Individualistic vs community modelling of species distributions under climate change. *Ecography*, **32**, 55–65.

Baselga, A. & Araújo, M.B. (2010) Do community-level models describe community variation effectively? *Journal of Biogeography*, **37**, 1842–1850.

Begon, M., Harper, J.L. & Townsend, C.R. (1996) *Ecology: individuals populations and communities*, Third edition edn. Blackwell Science Inc, Oxford.

Benito, B.M., Cayuela, L. & Albuquerque, F.S. (2013) The impact of modelling choices in the predictive performance of richness maps derived from species-distribution models: guidelines to build better diversitymodels. *Methods in Ecology and Evolution*, **4**, 327-335.

Blois, J.L., Zarnetske, P.L., Fitzpatrick, M.C. & Finnegan, S. (2013) Climate change and the past, present, and future of biotic interactions. *Science*, **341**, 499-504.

Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5-32.

Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014) Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.

Clark, J.S. (2009) Beyond neutral science. *Trends in Ecology and Evolution*, **24**, 8-15.

Cohen, J. (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *American Psycological Association*, **70**, 213-220.

Cord, A.F., Klein, D., Gernandt, D.S., la Rosa, J.A.P. & Dech, S. (2014) Remote sensing data can improve predictions of species richness by stacked species distribution models: a case study for Mexican pines. *Journal of Biogeography*, **41**, 736-748.

Cornelissen, J.H.C., Lavorel, S., Garnier, E., Díaz, S., Buchmann, N., Gurvich, D.E., Reich, P.B., ter Steege, H., Morgan, H.D., van der Heijden, M.G.A., Pausas, J.G. & Poorter, H. (2003) A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany*, **51**, 335-380.

Currie, D.J. (1991) Energy and large-scale patterns of animal- and plant-species richness. *The American Naturalist*, **137**, 27–49.

D'Amen, M., Bombi, P., Pearman, P.B., Schmatz, D.R., Zimmermann, N.E. & Bologna, M.A. (2011) Will climate change reduce the efficacy of protected areas for amphibian conservation in Italy? *Biological Conservation*, **144**, 989-997.

D'Amen, M., Pradervand, J.N. & Guisan, A. (2015) Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography*, **24**, 1443-1453.

Distler, T., G., S.J., Velasquez-Tibata, J. & Langham, G.M. (2015) Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *Journal of Biogeography*, 1-13.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Marquez, J.R.G., Gruber, B., Lafourcade, B., Leitao, P.J., Munkemuller, T., McClean, C., Osborne, P.E., Reineking, B., Schroder, 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. *Ecography*, **36**, 27-46.

Dornelas, M., Connolly, S.R. & Hughes, T.P. (2006) Coral reef diversity refutes the neutral theory of biodiversity. *Nature*, **440**, 80-82.

Douma, J.C., Witte, J.-P.M., Aerts, R., Bartholomeus, R.P., Ordoñez, J.C., Venterink, H.O., Wassen, M.J. & van Bodegom, P.M. (2012) Towards a functional basis for predicting vegetation patterns; incorporating plant traits in habitat distribution models. *Ecography*, **35**, 294–305.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122–1131.

Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P. & Guisan, A. (2012) Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science*, **24**, 593-606.

Dubuis, A., Rossier, L., Pottier, J., Pellissier, L. & Guisan, A. (2013) Predicting current and future community patterns of plant functional traits. *Ecography,* **36**, 1158-1168.

Dullinger, S., Gattringer, A., Thuiller, W., Moser, D., Zimmermann, N.E., Guisan, A., Willner, W., Plutzar, C., Leitner, M., Mang, T., Caccianiga, M., Dirnbock, T., Ertl, S., Fischer, A., Lenoir, J., Svenning, J.C., Psomas, A., Schmatz, D.R., Silc, U., Vittoz, P. & Hulber, K. (2012) Extinction debt of high-mountain plants under twenty-first-century climate change. *Nature Climate Change*, **2**, 619-622.

Dunstan, P.K., Foster, S. D. & Darnell, R. (2011) Model based grouping of species across environmental gradients. *Ecological Modelling*, **222**, 955-963.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

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

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology Evolution and Systematics*, **40**, 677-697.

Faleiro, F.V., Machado, R.B. & Loyola, R.D. (2013) Defining spatial conservation priorities in the face of land-use and climate change. *Biological Conservation*, **158**, 248-257.

Fernandes, J.A., Cheung, W.W., Jennings, S., Butenschon, M., de Mora, L., Frolicher, T.L. & Grant, A. (2013) Modelling the effects of climate change on the distribution and production of marine fishes: accounting for trophic interactions in a dynamic bioclimate envelope model. *Global Change Biology*, **19**, 2596-2607.

Ferrier, S., Drielsma, M., Manion, G. & Watson, G. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, **11**, 2309-2338.

Ferrier S. & Guisan A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.

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

Francis, A.P. & Currie, D.J. (2003) A globally consistent richness–climate relationship for angiosperms. *The American Naturalist*, **161**, 523–536.

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. *Ecological Modelling*, **217**, 48-58.

Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010) A framework for community interactions under climate change. *Trends in Ecology and Evolution*, **25**, 325-331.

Gotelli, N.J., Buckley, N.J. & Wiens, J.A. (1997) Co-occurrence of Australian land birds: Diamond's assembly rules revisited. *Oikos*, **80**, 311-324.

Gotelli, N.J., Anderson, M.J., Arita, H.T., Chao, A., Colwell, R.K., Connolly, S.R., Currie, D.J., Dunn, R.R., Graves, G.R., Green, J.L., Grytnes, J., Jiang, Y., Jetz, W., Lyons, S.K., McCain, C.M., Magurran, A.E., Rahbek, C., Rangel, T.F.L.V.B., Soberón, J., Webb, C.O. & Willig, M.R. (2009) Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, **12**, 873-886.

Götzenberger, L., De Bello, F., Anne Bråthen, K., Davison, J., Dubuis, A., Guisan, A., Lepš, J., Lindborg, R., Moora, M., Pärtel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel, K. & Zobel, M. (2012) Ecological assembly rules in plant communities—approaches, patterns and prospects. *Biological Reviews*, **87**, 111–127.

Grytnes, J.A. (2003) Species-richness patterns of vascular plants along seven altitudinal transects in Norway. *Ecography*, **26**, 291-300.

Grytnes, J.A. & Beaman, J.H. (2006) Elevational species richness patterns for vascular plants on Mount Kinabalu, Borneo. *Journal of Biogeography*, **33**, 1838-1849.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Guisan, A., Lehmann, A., Ferrier, S., Aspinall, R., Overton, R., Austin, M. & Hastie, T. (2006) Making better biogeographic predictions of species distribution. *Journal of Applied Ecology*, **43**, 386-392.

Guisan, A. & Rahbek, C. (2011) SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424-1435.

Guralnick, R.P., Hill, A.W. & Lane, M. (2007) Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters*, **10**, 663-672.

Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G., Toomajian, C., Roux, F. & Bergelson, J. (2011) Adaptation to climate across the Arabidopsis thaliana genome. *Science*, **334**, 83-86.

Hartmann, P., Fouvy, P. & Horisberger, D. (2009) L'Observatoire de l'écosystème forestier du canton de Vaud: espace de recherche appliquée| The Forest Ecosystem Observatory in Canton Vaud: a field of applied research. *Schweizerische Zeitschrift für Forstwesen*, 160, s2-s6.

Hastie, T.J. & Tibshirani, R. (1990) *Generalized additive models*. Chapman & Hall, London.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guégan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.

Herzog, F. (2005) *Evaluation des mesures écologiques: domaine de l'azote et du phosphore*. Agroscope FAL Reckenholz.

Hooper, D.U., Chapin, F.S., III, Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A.J., Vandermeer, J. & Wardle, D.A. (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs*, **75**, 3–35.

Hortal, J., De Marco, P., Santos, A.M.C. & Diniz-Filho, J.A.F. (2012) Integrating biogeographical processes and local community assembly. *Journal of Biogeography*, **39**, 627–628.

Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ.

Hui, F.C.K., Warton, D.I., Foster, S.D. & Dunstan, P.K. (2013) To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models. *Ecology*, **94**, 1913–1919.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An introduction to statistical learning*. Springer.

Jaramillo, C. & Cárdenas, A. (2013) Global warming and neotropical rainforests: A historical perspective. *Annual Review of Earth and Planetary Sciences*, **41**, 741-766.

Jimenez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica-International Journal of Ecology*, **31**, 361-369.

Jones, M.R., Forester, B.R., Teufel, A.I., Adams, R.V., Anstett, D.N., Goodrich, B.A., Landguth, E.L., Joost, S. & Manel, S. (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455-3468.

Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334-50.

Keddy, P.A. (1992a) A pragmatic approach to functional ecology. *Functional Ecology*, **6**, 621–626.

Keddy, P.A. (1992b) Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science*, **3**, 157–164.

Kiehl, J. (2011) Lessons from Earth's past. *Science*, **331**, 158-159.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E. & O'Hara, R.B. (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Körner, C. (2011) Coldest places on earth with angiosperm plant life. *Alpine Botany*, **121**, 11-22.

Kremer, A., Ronce, O., Robledo-Arnuncio, J.J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J.R., Gomulkiewicz, R., Klein, E.K. & Ritland, K. (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, **15**, 378-392.

Laughlin, D.C., Joshi, C., van Bodegom, P.M., Bastow, Z.A. & Fulé, P.Z. (2012) A predictive model of community assembly that incorporates intraspecific trait variation. *Ecology Letters*, **15**, 1291-1299.

Leach, K., Zalat, S. & Gilbert, F. (2013) Egypt’s Protected Area network under future climate change. *Biological Conservation*, **159**, 490-500.

Le Roux, P. C., Lenoir, J., Pellissier, L., Wisz, M. S. & Luoto, M. (2013) Horizontal, but not vertical, biotic interactions affect fine-scale plant distribution patterns in a low-energy system. *Ecology*, **94**, 671-682.

Liu, C.R., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography,* **28**, 385–393.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J. & Guisan, A. (2010) Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, **143**, 2647-2657.

Maggini, R., Lehmann, A., Zimmermann, N.E. & Guisan, A. (2006) Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography*, **33**, 1729-1749.

Magurran, A.E. & Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature,* **422**, 714-716.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-931.

Manel, S., Joost, S., Epperson, B.K., Holderegger, R., Storfer, A., Rosenberg, M.S., Scribner, K.T., Bonin, A. & Fortin, M..J. (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760-3772.

Manel, S. & Holderegger, R. (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*, **28**, 614-621.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59-69.

Mateo, R.G., Felicísimo, Á.M., Pottier, J., Guisan, A. & Muñoz, J. (2012) Do stacked species distribution models reflect altitudinal diversity patterns? *PloS ONE,* **7**, e32586.

Mateo, R.G., de la Estrella, M., Felicisimo, A.M., Munoz, J. & Guisan, A. (2013) A new spin on a compositionalist predictive modelling framework for conservation planning: A tropical case study in Ecuador. *Biological Conservation*, **160**, 150-161.

Mateo, R.G., Mokany, K. & Guisan, A. (2017) Biodiversity Models: What If Unsaturation Is the Rule? *Trends in Ecology & Evolution*, **32**, 556-566.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models. 2nd edition*. Chapman and Hall, London.

McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution,* **21**, 178–185.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811-823.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.C., Thuiller, W., Araujo, M.B., Guisan, A. & Zimmermann, N.E. (2010) Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33**, 1038-1048.

Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E. & Elith, J. (2014) What do we gain from simplicity versus complexity in species distribution models? *Ecography,* **37**, 1267-1281.

Mokany, K. & Ferrier, S. (2011) Predicting impacts of climate change on biodiversity: a role for semi-mechanistic community-level modelling. *Diversity and Distributions*, **17**, 374-380.

Mokany, K., Harwood, T.D., Overton, J.M., Barker, G.M. & Ferrier, S. (2011) Combining α - and β -diversity models to fill gaps in our knowledge of biodiversity. *Ecology Letters*, **14**, 1043–1051.

Mokany, K., Harwood, T.D., Williams, K.J. & Ferrier, S. (2012) Dynamic macroecology and the future for biodiversity. *Global Change Biology*, **18**, 3149–3159.

Moser, D., Dullinger, S., Englisch, T., Niklfeld, H., Plutzar, C., Sauberer, N., Zechmeister, H.G. & Grabherr, G. (2005) Environmental determinants of vascular plant species richness in the Austrian Alps. *Journal of Biogeography*, **32**, 1117–1127.

Moles, A.T. & Westoby, M. (2006) Seed size and plant strategy across the whole life cycle. *Oikos*, **113**, 91-105.

Nenzen, H.K. & Araujo, M.B. (2011) Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling*, **222**, 3346-3354.

Nogues-Bravo, D. & Rahbek, C. (2011) Communities under climate change. *Science*, **334**, 1070-1071.

Ovaskainen, O., Hottola, J. & Siitonen, J. (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, **91**, 2514–2521.

Ozinga, W.A., Schaminée, J.H.J., Bekker, R.M., Bonn, S., Poschlod, P. & Tackelberg, O. (2005) Predictability of plant species composition from environmental conditions is constrained by dispersal limitation. *Oikos*, **108**, 555–561.

Pakeman, R.J. & Quested, H.M. (2007) Sampling plant functional traits: what proportion of the species need to be measured? *Applied Vegetation Science*, **10**, 91–96.

Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37-42.

Pellissier, L., Fournier, B., Guisan, A. & Vittoz, P. (2010) Plant traits co-vary with altitude in grasslands and forests in the European Alps. *Plant Ecology*, **211**, 351–365.

Pellissier, L., Pradervand, J.-N., Pottier, J., Dubuis, A., Maiorano, L. & Guisan, A. (2012) Climate-based empirical models show biased predictions of butterfly communities along environmental gradients. *Ecography*, **35**, 684–692.

Pellissier, L., Rohr, R. P., Ndiribe, C., Pradervand, J.-N., Salamin, N., Guisan, A. & Wisz, M. (2013) Combining food web and species distribution models for improved community projections. *Ecology and Evolution*, **3**, 4572-4583.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M. & Wegmann, M. (2013) Essential biodiversity variables. *Science*, **339**, 277-278.

Peres-Neto, P.R., Olden, J.D. & Jackson, D.A. (2001) Environmentally constrained null models: site suitability as occupancy criterion. *Oikos*, **93**, 110–120.

Pineda, E. & Lobo, J.M. (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182-190.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P. & Guisan, A. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **22**, 52–63.

Pradervand, J.N., Dubuis, A., Reymond, A., Sonnay, V., Gelin, A. & Guisan, A. (2013) Quels facteurs influencent la richesse en orthoptères des Préalpes vaudoises? *Bulletin de la Société Vaudoises des Sciences Naturelles*, **93**, 155-173.

Pradervand, J.-N., Dubuis, A., Pellissier, L., Guisan, A. & Randin, C.F. (2014) Very high-resolution environmental predictors in species distribution models: moving beyond topography? *Progress in Physical Geography,* **38**, 79-96.

Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology,* **29**, 254-283.

Randin, C.F., Vuissoz, G., Liston, G.E., Vittoz, P. & Guisan, A. (2009) Introduction of snow and geomorphic disturbance variables into predictive models of alpine plant distribution in the western Swiss Alps. *Arctic, Antarctic, and Alpine Research*, **41**, 347–361.

Schoville, S.D., Bonin, A., François, O., Lobreaux, S., Melodelima, C. & Manel, S. (2012) Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution, and Systematics,* **43**, 23-43.

Shipley, B., Vile, D. & Garnier, E. (2006) From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science*, **314**, 812–814.

Shipley, B. (2010) Community assembly, natural selection and maximum entropy models. *Oikos*, **119**, 604-609.

Shipley, B., Laughlin, D.C., Sonnier, G. & Ottinowski, R. (2011) A strong test of a maximum entropy model of trait-based community assembly. *Ecology*, **92**, 507–517.

Sonnier, G., Shipley, B. & Navas, M. (2010a) Plant traits, species pools and the prediction of relative abundance in plant communities: a maximum entropy approach. *Journal of Vegetation Science*, **21**, 318–331.

Sonnier, G., Shipley, B. & Navas, M.L. (2010b) Quantifying relationships between traits and explicitly measured gradients of stress and disturbance in early successional plant communities. *Journal of Vegetation Science*, **21**, 318–331.

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

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293.

Thuiller, W. (2005) Modélisation empirique de la biodiversité : un outil pour évaluer l'impact des changements globaux. . *Les biodiversités entre sciences et politiques : objets, théories, pratiques* (ed. by Cnrs-Editions), pp. 113-123, Paris.

Thuiller, W., Lavorel, S., Araujo, M.B., Sykes, M.T. & Prentice, I.C. (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the united States of America*, **102**, 8245-50.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo M.B. (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.

Thuiller, W., Albert, C.H., Dubuis, A., Randin, C. & Guisan, A. (2010) Variation in habitat suitability does not always relate to variation in species' plant functional traits. *Biology Letters*, **6**, 120–123.

Thuiller, W., Georges, D. & Engler, R. (2013) biomod2: Ensemble platform for species distribution modeling. *R package version*, 2, r560.

Van der Putten, W.H., Macel, M. & Visser, M.E. (2010) Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365**, 2025-2034.

Vilhena, D.A. & Antonelli, A. (2014) Beyond similarity: A network approach for identifying and delimiting biogeographical regions. *arXiv preprint arXiv:1410.2942*,

Webb, C.T., Hoeting, J.A., Ames, G.M., Pyne, M.I. & LeRoy Poff, N. (2010) A structured and dynamic framework to advance traits-based theory and prediction in ecology. *Ecology Letters*, **13**, 267–283.

Wennekes, P., Rosindell, J. & Etienne, R. (2012) The neutral–niche debate: a philosophical perspective. *Acta Biotheoretica*, **60**, 257-271.

Westoby, M. (1998) A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil*, **199**, 213–227.

Wisz, M.S., Pottier, J., Kissling, W.D. *et al*. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews of the Cambridge Philosophical Society*, **88**, 15–30.

Wright, D.H. (1983) Species-energy theory: an extension of species-area theory. *Oikos*, **41**, 496–506.

Zachos, J.C., Dickens, G.R. & Zeebe, R.E. (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, **451**, 279-283.

Zimmermann, N.E., Edwards, T.C., Graham, C.H., Pearman, P.B. & Svenning, J.-C. (2010) New trends in species distribution modelling. *Ecography*, **33**, 985-989.

Zurell, D., Zimmermann, N.E., Sattler, T., Nobis, M.P. & Schröder, B. (2016) Effects of functional traits on the prediction accuracy of species richness models. *Diversity and Distributions*, **22**: 905-917.

PART III - CHAPTERS OUTSIDE THESIS’ SCOPE, WITH MAIN CONTRIBUTION

Chapter 3.1

Climate and land-use changes reshuffle politically-weighted priority areas of mountain biodiversity

This manuscript is in review at Global Ecology and Conservation

My contribution to the paper: I conceived the initial idea and developed the methodological framework in coordination with Antoine Guisan and Claire Vincent. Claire Vincent analysed all the data and wrote the first draft (master thesis). Claire and I reformulated the thesis into a manuscript form, leading the writing of the various versions of the manuscript.

Climate and land-use changes reshuffle politically-weighted priority areas of mountain biodiversity

Claire Vincent[a †], Rui F. Fernandes[a* †], Ana Cardoso[a,c], Olivier Broennimann[a,b], Valeria Di Cola[a], Manuela D'Amen[a], Sylvain Ursenbacher[d,e], Benedikt R. Schmidt[d,f.], Jean-Nicolas Pradervand[g], Loïc Pellissier[h,i] and Antoine Guisan[a,b]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland

[b] Institute of Earth Science dynamics, Geopolis, University of Lausanne, 1015 Lausanne, Switzerland

[c] Faculdade de Ciências Sociais e Humanas – Universidade Nova de Lisboa, Lisboa, Portugal

[d] Info Fauna Karch, UniMail, Bâtiment G, Bellevaux 51, 2000 Neuchâtel, Switzerland

[e] Department of Environmental Sciences, Section of Conservation Biology, University of Basel, St. Johanns-Vorstadt 10, 4056 Basel, Switzerland

[f] Institut für Evolutionsbiologie und Umweltwissenschaften, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[g] Swiss Ornithological Institute, Valais Field Station, Rue du Rhône 11, CH-1950 Sion, Switzerland

[h]Landscape Ecology, Institute of Terrestrial Ecosystems, ETH Zürich, CH-8092 Zürich, Switzerland [i]Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland

ABSTRACT

Protected areas (PA) play a critical role in conserving biodiversity and maintaining viable populations of threatened species. Yet, as global change could reduce the future effectiveness of existing PA in covering high species richness, updating the boundaries of existing PA or creating new ones might become necessary to uphold conservation goals. Modelling tools are increasingly used by policy makers to support the spatial prioritization of biodiversity conservation, enabling the inclusion of scenarios of environmental changes to achieve specific targets. Although mountain regions are considered particularly threatened biodiversity sanctuaries, only few quantifications of the efficiency of protected areas have been conducted in mountains. Here, using the Western Swiss Alps as a case study, we show how integrating species richness derived from species distribution model predictions for four taxonomic groups under present and future climate and land-use conditions into two conservation prioritization schemes can help optimize extant and future protected areas. We found that existing mountain PA are currently not situated in the most environmentally nor politically suitable locations when maximizing alpha diversity for the studied taxonomic groups, and that current PA could become even less optimum under the future climate and land-use change scenarios. This analysis has focused on general areas of high species richness and did not account for special habitats or functional groups that were used to create the existing network. We identified possible ways forward to improve conservation planning under scenarios of environmental changes. We conclude that such an integrated framework could support more efficient conservation planning and can be similarly applied to other landscapes or other biodiversity conservation indices.

KEYWORDS: Conservation Prioritization, Decision Support Tools, Land Costs, Species Distribution Models, Switzerland, Zonation

INTRODUCTION

Growing anthropogenic activities are causing increasing threats to biodiversity even in pristine areas (Lambin et al. 2001). Land-use change, human induced climate change, or the introduction of invasive species (Millennium Ecosystem Assessment 2005) are affecting natural ecosystems, reducing the habitat available for species and therefore causing species extinctions and biodiversity loss (Steffen et al. 2007). The effects of climate change have already influenced species ranges, shifting them northwards and to higher altitudes (Walther et al. 2002; Hickling et al. 2006; Chen et al. 2011), causing some species to move out of current protected areas (PA; Araújo 2004). The fast rate of climate change might limit the capacity of species to adapt locally or to migrate to suitable areas (Welch 2005). Land-use change can have further substantial and severe negative effects on biodiversity (Struebig et al. 2015), stressing some vulnerable species and pushing them closer to extinction (Gibson et al. 2011). These changes can lead to an overall decrease in species diversity in previously species rich areas.

The international conservation community has identified PA as the cornerstone of biodiversity conservation (CBD 2010), protecting vulnerable species and habitats, with mountains playing a special role as species sanctuaries (Bugmann et al. 2007; Guisan et al. In Press). Due to legal constraints, terrestrial PA are usually designed to be static, not accounting for ecosystem changes or shifts in species ranges, potentially limiting their effectiveness under climate or land-use change (Pressey et al. 2007; Alagador et al. 2014).

Systematic conservation planning is a spatial process designed to optimize the delimitation of PA, by identifying a limited set of unique and/or complementary areas that maximize biodiversity conservation (Pressey et al. 2007). While financial resources are limited, considering land costs makes conservation planning more effective, as it balances the costs of conservation with the benefits to wildlife (Newburn et al. 2005). Thus, incorporating land-use change and land costs into conservation prioritization analyses has the power to provide an added priority ranking of PA (Naidoo et al. 2006). However, a PA network should be designed in such a way that it can be adapted to changing conditions, since the effects of climate change are now inevitable and cause species distributions to be non-static within landscapes (Heller & Zavaleta 2009).

To include future scenarios of species distributions, predictions based on species distribution models (SDMs; Guisan et al. 2017) can be incorporated into the analyses. A recent study reported that conservation planners could still make better use of SDMs, supporting further development of such quantitative tools in conservation planning (Tulloch et al. 2016). Prioritization tools for conservation planning, such as “Zonation” (Moilanen et al. 2005), support the identification of important areas while considering a cost-benefit ratio. Zonation can incorporate SDM predictions with cost maps, future landscape changes, species interactions, and habitat connectivity to find the optimal PA network based on pre-defined conservation goals. It has been used for conservation planning on a variety of projects around the globe (e.g. Summers et al. 2012; Faleiro et al. 2013; Pouzols et al. 2014), but we know only three examples where it was used in mountain areas (Fleishman et al. 2014; Wan et al. 2016; Zhang et al. 2011), none of which were in Europe.

Here, we used an existing prioritization method and compared it to a new multi-driver framework within the Zonation software to assess existing PA and propose a new PA network through a case study in the Western Swiss Alps. The goal of this analysis was to illustrate an improved framework to identify the most cost-effective PA network that will protect the highest level of species richness across several taxonomic groups, and to investigate whether the areas identified by these two prioritization methods provide the same hotspots for protection. The analysis included species data for insects, amphibians, reptiles, and plants together with environmental, landscape and human socio-economic data. We then investigated the effects of climate and land-use changes on optimal conservation solutions and determined potential drawbacks and improvements to these prioritization methods. Additionally, based on this study, we identified possible gaps and ways forward to improve conservation planning in mountain regions.

METHODS

Study area

The Western Swiss Alps of the Vaud canton were used for this case study as they represent a priority area for research of the new Center for Mountain Studies (CIRM) at the University of Lausanne (see https://www.unil.ch/centre-montagne). Due to its high diversity of species and of key conservation habitats, it was declared a priority area for conservation by WWF, Birdlife International and Pro Natura in 2015 (http://www.leregional.ch/N67958/la-position-du-wwf-et-de-pro-natura-en-detail.html). The elevation ranges from 375 m to 3210 m with an annual mean temperature between 3.5°C and 8°C, and the annual sum of rainfall between 1400 mm and 2400 mm (Randin et al. 2006). Full description of the different biological and environmental characteristics of this area can be found at http://rechalp.unil.ch.

In the study area, PA were assigned to one of four tiers based on their international commitments to protect biodiversity (Table S1, Appendix 1). PA designated with the International Union for Conservation of Nature (IUCN) Category Ia “Strict Nature Reserve” are assigned to Tier 1, and IUCN Category IV “Habitat/Species Management Areas” to Tier 2. Tier 3 PA do not align with these IUCN categories, even though they may have stronger national regulations. Tier 4 includes protected areas that are not designed for biodiversity conservation, but which may indirectly protect nature. In the study area, 18.12% of the landscape is protected by either Tier 1 (0.02%) or 2 (18.1%) areas (Fig 1). Although some of the existing protected areas were created to protect a specific habitat (e.g. fens, bogs and alluvial zones) or a taxonomic group (e.g. amphibians), the goal of this analysis was to protect the highest number of species within the most cost-effective area of land (i.e. alpha diversity; the analysis optimizes a different goal than the existing set of PA).

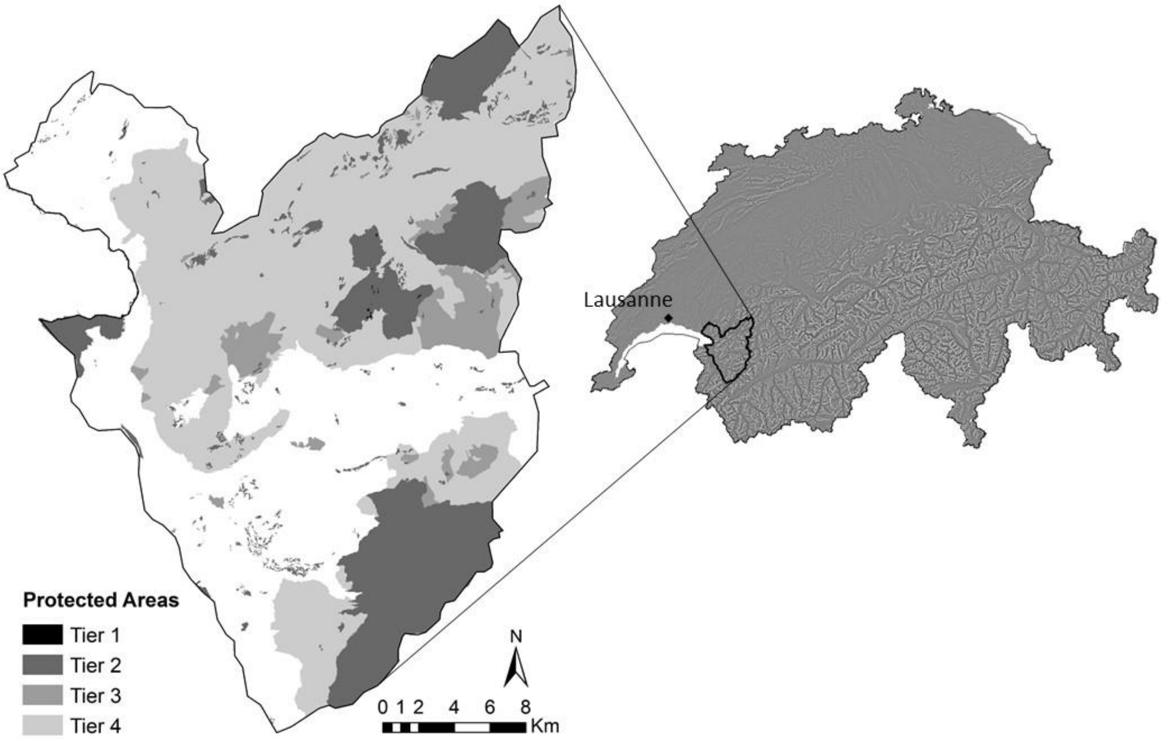


FIGURE 1 The study site and its position in Switzerland. The current protected areas are shaded in grey with the highest level of protection (Tier 1) given the darkest color, and the lowest level of protection (Tier 4) shown with the lightest grey.

Data sources

We used predictions from species distribution models (SDMs) available for four different taxonomic groups of species: insects, amphibians, reptiles, and plants (Table 1). Different sampling strategies were used to sample the different taxonomic groups: random stratified sampling was used to sample plants and insects, while occurrence records for amphibians and reptiles are from opportunistic observations (for a detailed description see Appendix 3). SDMs predictions were based on 25 m resolution topoclimatic variables including monthly maximum and minimum temperature and sum of precipitations, annual growing degree days, annual evapotranspiration, topographic position, aspect and slope (for a detailed description see Appendix 3). These variables have been shown to be useful predictors in mountain environments (D’Amen et al. 2015; Dubuis et al. 2011; Scherrer et al. 2017; 2018). Predictions for the future were based on variables translating the A1B climate storyline (IPCC 2001) for the years 2045 – 2074 (from here on referred to as 2060) and calculated using the climatic anomalies for all Swiss weather stations (Bosshard et al. 2011). This future climatic scenario predicts the global average temperature to rise by 1.8°C by 2060 (for a detailed description see Appendix 3).

The baseline land-use map was taken from the Swiss land-use statistics data from 2009 (SFSO 2013; Fig S1a, Appendix 1). The future land-use scenario map was modelled by Price et al. (2015) under the A1 climate change scenario (the scenario family in which A1B is found) using the Dynamic Conversion of Land Use and its Effects Model framework (Dyna-CLUE; Verburg & Overmars 2009) for the year 2035 (Fig S1b, Appendix 1). This model predicts high levels of agricultural land abandonment, especially in mountain pastures, and urbanization in lower elevation areas near roads and existing towns. Under this climate scenario there is more afforestation, defined as a transition from any land use to forest, than any other scenario.

Species distributions models

We used an ensemble of small models (ESMs; Breiner et al. 2015) to predict the current and future distributions of 767 species (Table 1) based on environmental data at high resolution (25 m). ESMs were implemented to avoid overfitting and to improve the modelling of rare species as they fit and average many small models each with few predictors at a time (typically 2) weighted by their cross-validated predictive performance (Lomba et al. 2010; Breiner et al. 2015). While ESMs were designed to model species with small sample sizes, they also work well for common species (Breiner et al. 2015). Only species with at least 10 presences were modelled (see Appendix 2 for the number of presences/absences for each species). A weighted mean of the following three modelling techniques was used to predict species distributions based on observational data and environmental variables: Generalized Linear Models (GLM; McCullagh & Nelder 1989), Random Forests (RF; Breiman 2001) and Maximum Entropy (MaxEnt; Phillips et al. 2006; Phillips & Dudik 2008).

TABLE 1 A summary of the species list (see Appendix 2 for details), showing the number of species for each taxonomic group used in this analysis.

Taxonomic Group	No. of species	Threatened species (CR, EN, VU)[a]	Protected species[b]
Amphibians	5	2	5
Insects	123	3	4
Plants	627	4	47
Reptiles	12	9	12
Total	767	18	68

[a] Threatened species are based on the Swiss national Red Lists.

[b] Protected species are explicitly referenced in Swiss legislative documents.

The data for each species was randomly partitioned into 70% for calibration and 30% for validation and this procedure was repeated 5 times. For each combination of environmental predictors, the different models were evaluated using a maximisation of the True Skill Statistic (maxTSS; see Appendix 2 for values), taking both omission and commission errors into account (see Appendix 3 for details on ESM parameterization). Predictions were turned into binary presence-absence data using the same thresholding approach that maximises the TSS (Liu et al. 2005) for both current and future distributions. Another map was saved for each species after applying a land-use filter which removed any projected presence that occurred in a land-use type in which no observations were recorded.

Spatial prioritization methods

Two methods of spatial prioritization for biodiversity conservation were compared (Fig 2). First, the “Priority Scores Method” (Fig 2a; Jenkins et al. 2015) identifies priority areas for the expansion of the existing protected area network. It enables planning to be focused on target species having the greatest need for protection. Second, the “Zonation Land-Use Filter Method” (Fig 2b) uses the Zonation software to combine the effects of climate change, land-use change, and land costs to propose optimal conservation areas and to assess the current protected areas ability to protect species richness. The priority scores were calculated with both the current and future predicted species distributions. These methods both generated maps with continuous ranking of the landscape to identify conservation priorities.

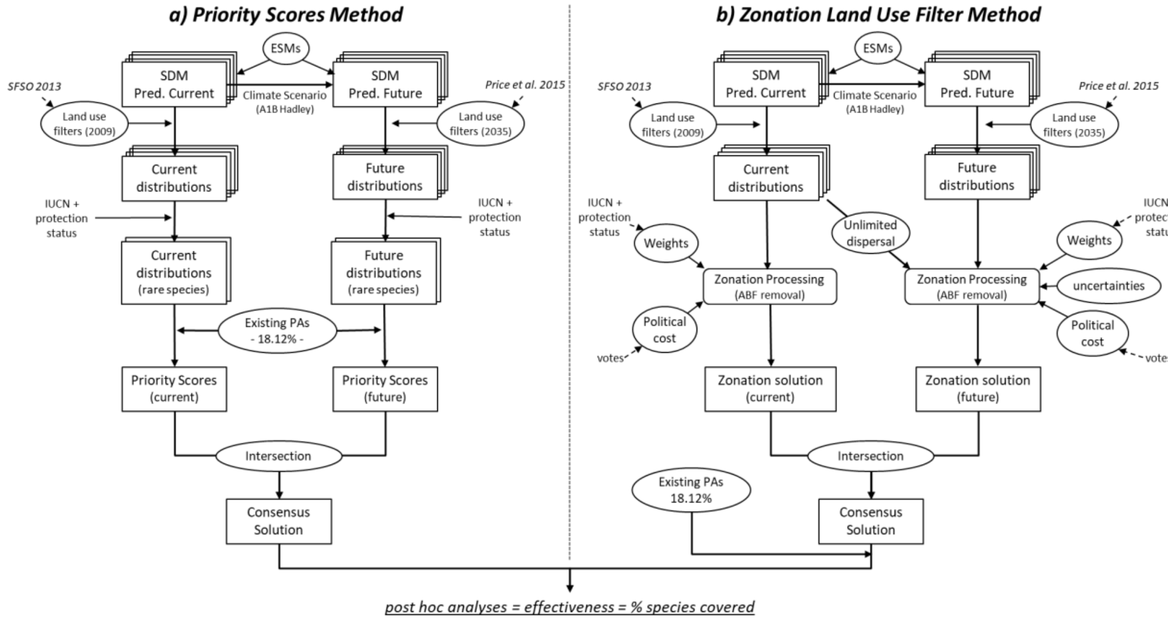


FIGURE 2 A schematic representation of the framework used for this analysis: a) the “Priority Scores Method” and b) the “Zonation Land-Use Filter Method”, both described in detail in the methods. The Zonation method with the land-use mask and negative weighting differs from the Land-Use Filter Method since it used unfiltered species distributions and excluded urban areas while also negatively weighting less favorable land-use types (not in figure).

The priority scores method is based on the proportion of each species range that is currently protected under the existing conservation network. It is calculated here by dividing the area of each species distribution that does not fall in a Tier 1/2 area by the total area of the species range. For each $25m^2$

pixel, this species-specific score was then summed across all species present in each pixel to get the total priority scores across the landscape. The analysis was done first with all species distributions and then with only species of conservation concern as they have the greatest need for protection. Species were classified as being of conservation concern if they were threatened (VU, EN, or CR category on the Swiss National Red Lists) or protected by Swiss legislation (see Appendix 2).

The Zonation Method uses an algorithm with an iterative ranked cell removal method (Moilanen et al. 2005). The additive benefit function (ABF) was used as the removal rule because it is optimal when the spatial extent of the case study is relatively small when compared to species ranges, and when the interest is conserving overall species richness. In addition, core-area Zonation (CAZ) is not recommended when incorporating land costs to an analysis that identifies biologically important areas (Moilanen et al. 2014).

Each feature (e.g. species, land-use type) that was added to a Zonation analysis was assigned a weight, and the aggregate weight of features in each cell defines its priority during the cell removal process with a higher weight given a higher priority (Moilanen et al. 2005). In this analysis, species feature weights were first assigned based on species IUCN status with critically endangered species given the highest weight (Appendix 2). Uncertainty in the modelling was addressed by giving lower weights to the future species distributions than to the present, and the lowest in the connectivity between these time steps, because future projections carry higher uncertainties.

Zonation solutions were calculated by dividing the total weight of a cell, in this case species richness, by the cost associated with protecting that cell. Here, costs were defined as political costs, as calculated by Cardoso (2015), by averaging the opposition results from two votes on legislature with potentially strong effects on biodiversity conservation. The regions that had higher opposition to beneficial measures for biodiversity were given higher political cost (Fig S2, Appendix 1). This was included as it is more difficult to implement changes to expand PA in regions with higher social opposition to such changes. Monetary costs were not incorporated as they were directly related to land-use.

To prioritize for future distributions under climate change scenarios in the Zonation method, the distribution interaction component of Zonation was used (Moilanen et al. 2014; Rayfield et al. 2009). This provides solutions that transform one conservation target based on its proximity to another represented as an ecological interaction. Here, this was represented as the connectivity on each species distribution between time steps, characterized by the dispersal ability of each species. Engler et al. (2009) found that, for the same study area, simulations of plant distributions with limited dispersal gave similar results to those with unlimited dispersal, being significantly different from those with no dispersal. Therefore, here we only present the results with unlimited dispersal.

Land-use change was added to Zonation by two separate analyses: (i) land-use filter on SDMs (LU-filter; Fig 2), and (ii) removal mask layer with negative weighting (LU-mask). The LU-filter removed any predicted presence from a land-use type in which no observations were recorded. The LU-mask used unfiltered species distributions, excluded urban areas from the analysis using a mask, and negatively weighed other less favorable land-use types. We consider that a conservative estimate of the amount of land-use change predicted in the study area exists, when using future land-use maps for 2035 and climate scenarios to 2060.

Creating consensus solutions and measuring effectiveness

To identify key areas to protect rare species or species with small ranges, we summed the priority scores across all species of conservation concern. Locations with priority scores within the top quartile were selected as priority areas. The current and future scores for species of conservation concern were overlaid to create a consensus map to select the most important areas for conservation.

In the Zonation Method, post-hoc analyses were done automatically in two ways. First, to identify the top 18.12% (same area as the current Tier 1/2 PA) of the landscape to protect. Second, to assess the effectiveness of both the current Tier 1/2 PA (included post-hoc as a mask) and the Zonation solutions in protecting modelled species ranges. The Zonation consensus solution is the spatial overlap between current and future proposed networks with the LU-mask method.

RESULTS

Priority scores method

The results from the priority scores follow a similar trend as the diversity maps (Fig S3, Appendix 1) and areas with high diversity have higher scores (Fig 3). Priority scores are predicted to decrease in the future (Fig 3). The current distributions provide similar priority areas for species of conservation concern as for all species (Fig 3a, 3c). The future distributions indicate a mismatch in priority scores between common species and those of conservation concern (Fig 3b, 3d). The priority scores differ spatially for different taxa (see Figure S4 and S5, Appendix 1). At both time steps, amphibians have high priority near waterways (Fig S4). Reptiles, *Bombus* spp. and orthopterans have higher priorities at lower elevations (Fig S4, S5), and lepidopterans at higher elevations (Fig S5).

Zonation solutions and their effectiveness

The Zonation solutions presented here used the political land costs, unlimited species dispersal, ABF cell removal rule, and the A1B climate change scenario. The Zonation solutions have over 90% spatial similarity between the land-use methods and the current and future time steps (Fig S6, Appendix 1). Although the LU-filter and LU-mask methods protect an almost equal proportion of species distributions (Fig 4), the LU-filter selects more unfavorable urban areas for converting to PA. Therefore, the LU-mask was used to create the final solution.

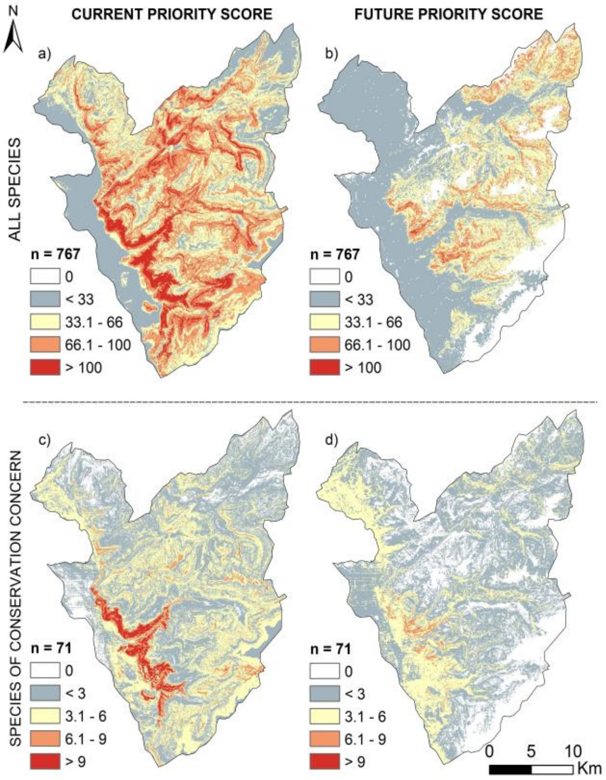


FIGURE 3 Priority maps showing the sum of species priority scores of the modelled species with their current (a, c) and projected future (b, d) distributions. Future refers to the distributions predicted to 2060 using the A1B climate change scenario. A subset of the species was used to create maps of species of conservation concern which were identified as being threatened and/or protected (c, d).

The solutions created by Zonation for the new PA networks performed better than the Tier 1/2 protected areas at both time steps (Fig 4). Existing PA cover on average 25% of all species ranges and new PA cover an average of 44%. Concerning the future scenario, existing PA would become less efficient protecting a mean of 23% of a species' range compared with 49% by the new PA. The existing PA would protect 23 fewer species than the new PA (i.e. 23 species would have their range completely outside of the PA), of which one species is nationally threatened (*Polyommatus damon*, VU), and three are protected by regional or national law (*Traunsteinera globosa*, *Monotropa hypopitys* and *Epipactis purpurata*). The difference in effectiveness between the new and existing PA is more notable with species of conservation concern (Fig 4).

Spatial overlap and final solutions

The composite map of the top priority scores identifies areas at the transition zone between the Rhone valley and the Alps in the South-West of the study area (Fig 5a). This zone is predominantly south facing dry open grasslands above 1500 m asl. The Zonation consensus solution identifies the areas where current and future LU-mask solutions overlap (Fig 5b) and were selected independently of the existing network. The Zonation solution identifies the same transitional zone as the priority scores as key areas to expand the PA network. The Zonation solution, however, also identifies many other areas that the priority scores method omits, predominantly in the South East of the study site, where the Grand Muveran hunting reserve is located.

The supplemental Zonation analyses found that the spatial overlap between new and existing PA was low, at 34% for current and 32% for future solutions. The analysis found no difference between solutions with unlimited and no dispersal (100% spatial overlap).

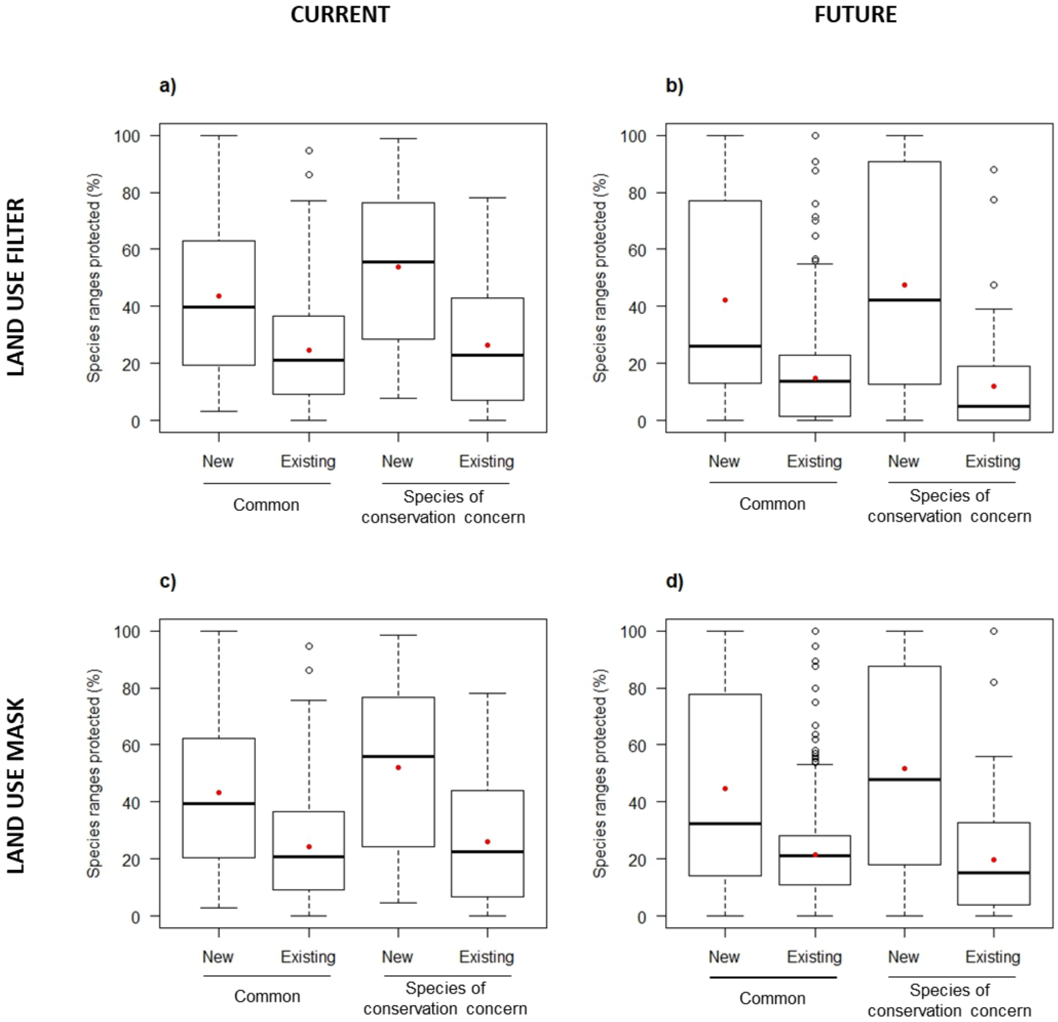


FIGURE 4 The proportions of species ranges protected by the Zonation solutions (New) created with LU-filter (a, b) and LU-mask (c, d) compared with the existing protected areas (Existing). Species of conservation concern are all of those listed as threatened and/or protected, and common species are all other species. Future refers to the distributions predicted to 2060 with the A1B climate change scenario. Open circles indicate outliers, black bars show medians, and solid dots indicate means. The same data is subdivided by taxonomic group and can be found in Appendix 1 (Fig S7).

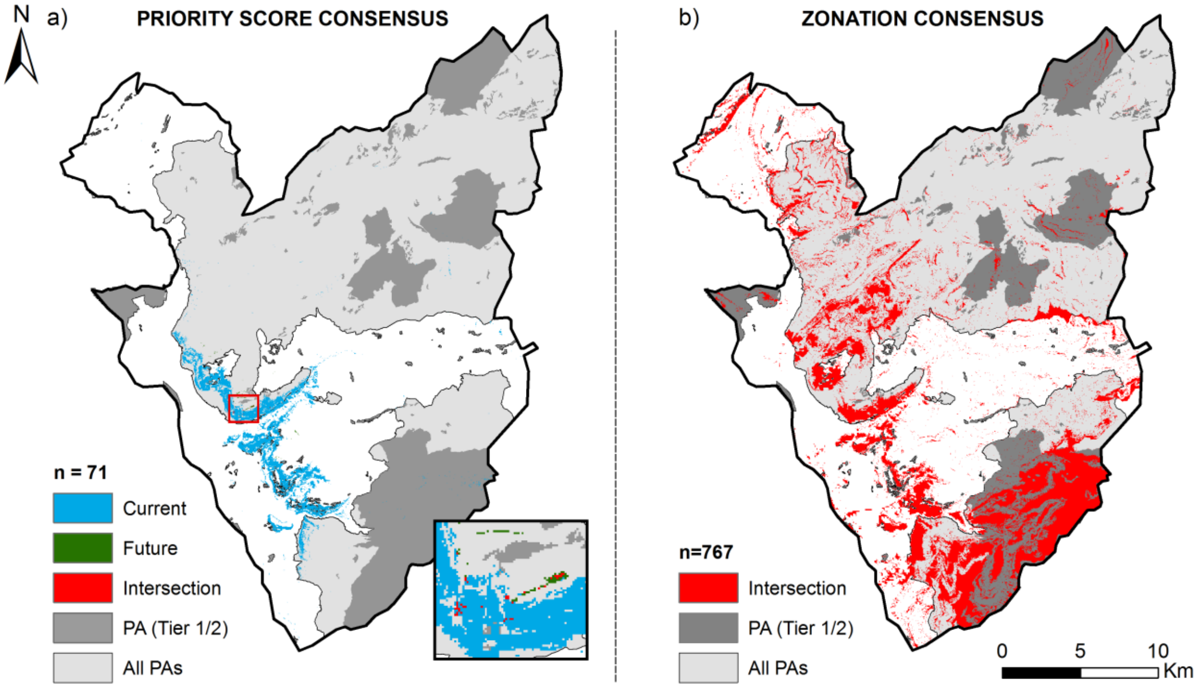


FIGURE 5 Final solutions with spatial overlaps (i.e. intersections). a) The priority score consensus solution created by overlaying the current (blue) and future (green) priorities for threatened and/or protected species. The overlapping regions are shown in red and represent areas that are priorities in both time steps (see inset for more detail). b) The Zonation consensus solution created by overlaying the current and future LU-mask solutions. The overlapping regions are also shown in red and represent the areas that are priorities at both time steps. The full extent of protected areas is shaded in light grey (all 4 tiers) and the top two tiers are shaded in dark grey.

DISCUSSION

We presented a multi-driver spatial prioritization method – including current and future predictions of species distributions by SDMs of multiple taxonomic groups (amphibians, reptiles, plants and three groups of insects), land-use, and political costs – for assessing the joint impacts of climate and land-use changes on the optimal locations of nature reserves for the preservation of maximal species richness in a region with wide elevational and environmental gradients. We used two schemes of spatial conservation prioritization at fine resolution (25 x 25 m). The solutions from both priority scores and Zonation schemes suggest that the current PA are not optimally located to protect high levels of species richness in the Western Swiss Alps. However, some existing PA were established for other purposes, e.g. the conservation of species-poor habitats such as mires or of specific (e.g. red-listed) species. It is important to note that implementing an optimum PA network might be difficult to achieve, and that simple decision rules (like protecting available sites with the highest species richness) may be more effective in some cases (Meir et al. 2004). PA are predicted to cover less species richness under future scenarios of climate and land-use change, with 60 of 767 species having their future range outside PA according to the predicted range shifts. These species might thus face higher threats of local extinction due to habitat fragmentation or degradation outside PA (Wilcox & Murphy 1985; Gibson et al. 2011). There is a predicted loss of suitability of existing PA networks for terrestrial fauna and flora (Araújo et al. 2011; Françoso et al. 2015), but in our study area it is likely stronger due to the mountain landscapes with steep climatic gradients along elevation. This emphasizes the need for using spatial decision support tools for conservation prioritization in areas with steep elevation gradients, like mountains. One limitation of this case study is that it did not account for new species that could colonize the study area following climate change (e.g. from warmer lowland areas in Switzerland, or from outside Switzerland), which could complement species losses within the study area. Because our solutions used the same species pool, it should not affect our comparisons, but future studies using the same framework should include this dimension.

The priority scores and Zonation methods provide contrasting solutions that can be useful to address different conservation questions. Both methods identified the transition zone between the low elevation Rhone valley and the Alps as a priority area to conserve biodiversity. This zone is predominantly dry grassland, which holds one of the most diverse plant communities among European ecosystems (Janišová et al. 2011). This is likely explained by the high number of plant species used for this analysis as well as our goal of protecting the highest level of species richness from our set of species. The Zonation solution also highlighted the importance of the Grand Muveran hunting reserve in the south-east of the study area that the priority scores method did not identify. The priority scores method provides solutions that can be more easily implemented as it proposes priority areas to be added to existing networks (Fig 5a). The Zonation method (Fig 5b) supplies key regions where human impacts should be minimized, or where future protection should be focused, to maintain high species richness. The results from the priority scores show a mismatch in the priority areas between common species and those of conservation concern. This highlights the importance of setting clear goals with policy makers and scientists before undertaking a similar analysis in other areas (e.g. protecting threatened species versus biodiversity in general; Grant et al. 2013).

Land-use change has only recently been incorporated into Zonation analyses (Faleiro et al. 2013; Struebig et al. 2015; Zwiener et al. 2017; Verhagen et al. 2018). In this work, two methods of accounting for land-use change – LU-filter and LU-mask – were used. The LU-mask produced solutions for lower conflicts between human uses and biodiversity conservation as urban areas were excluded from solutions. Future analyses could also add other non-convertible lands to this mask (e.g. one might exclude areas where agricultural production is given priority over other uses by the government). Globally, the increasing demand for food by the growing human population (Foley et al. 2011) will likely increase the risk of conflict over land-uses. Because of the trade-off between minimizing conflict and increasing the proportion of species ranges that can be protected, we would recommend combining these land-use methods in future studies by first including a filter and then adding a mask to remove non-convertible lands.

The goal of this analysis was to protect the highest possible level of diversity across all species in the four taxonomic groups and those of conservation concern using decision support tools. Applying these tools towards conservation prioritization is less common in European countries than in more economically developing or biodiversity rich countries (e.g. Brazil, Australia). This infrequent recourse to spatial decision tools could be due to the well-established (even if not sufficient) existing PA network in Europe, likely considered already optimal by the public and governments. However, existing PA have been more frequently created in places that are unsuitable for economic activities, and not necessarily in areas with high biodiversity (Margules & Pressey 2000). Human density is high in Europe, creating a trade-off between minimizing conflict and increasing the proportion of species ranges that can be protected. Additionally, it could be due to an emphasis placed on land sharing rather than land sparing in many densely populated countries (Phalan et al. 2011). For example, many European countries use set-asides in agricultural areas to protect biodiversity (Van Buskirk & Willi 2004). This strategy allows the focus to be on conserving irreplaceability ahead of high biodiversity, highlighting the challenges of conservation and the need to set specific conservation targets.

Identified gaps, suggestions and ways forward in conservation planning

While working through this analysis, five typical problems were observed. First, although we opted for a maximization of species richness, other dimensions of biodiversity could be considered, such as irreplaceability and complementarity, special habitats (e.g. bogs), functional groups (e.g. national priority species), ecosystem services (i.e. nature's contributions to society; Díaz et al. 2015), and various genetic facets of biodiversity. Similarly, other threats to biodiversity could have been included, such as biological invasions (Vicente et al. 2013). Second, data for other important taxonomic groups, in this case especially mammals and birds, were not available with the same coverage, quality or reliability in a spatially-explicit way (e.g. required to build models). This limited the conservation priorities that could be set, but it did not prevent illustrating the approach for these groups. Future analyses could also differentially weight taxonomic groups based on pre-defined conservation goals. Third, mismatches in the data from different academic fields, such as different spatial and temporal scales for the biological, climate and land-use data and scenarios, can make their joint use in integrated analyses more difficult. Fourth, uncertainties associated with SDMs (Barry & Elith 2006; Rocchini et al. 2011), and when using decision tools like Zonation – especially when considering land-use and climate change predictions (Moilanen et al. 2006) – are likely to occur, yet these remain difficult to account for and quantify. Uncertainty issues within Zonation could be attenuated through distribution discounting (Moilanen et al. 2006), measuring the variance between species distributions predicted from multiple SDMs (Faleiro et al. 2013; Lemes & Loyola 2013), or by differentially weighing current and future species distribution layers based on uncertainty, as future distributions have higher uncertainty (Kujala et al. 2013). It is also important to evaluate if sufficient data is available for all the species to be used in the analysis (e.g. Canessa et al. 2015). Fifth, although we did not define conservation targets in concert with stakeholders in this scientific study, doing so would be crucial if our framework was to be applied to revise the network of protected areas in this region. In this regard, there is still limited communication between science and policy on the use of modelling approaches to support conservation decisions (Guisan et al. 2013, Tulloch et al. 2016). Ideally, scientists, stakeholders and the public would need to come together and share ideas to define targets and achieve a consensus on what might be considered as achievable (Wilson 2008, Dicks et al. 2014). Improving this science-society link remains especially important to resolve the 'implementation crisis' wherein the scientific ability to create these results outweighs the ability to apply them (Knight & Cowling 2003; Arlettaz et al. 2010). A step was already taken in this direction through the implementation of a science-policy 'forum' ("*bourses aux questions*") set-up on the RECHALP web geoportal to support transdisciplinary research in this study area (rechalp.unil.ch/bourseauxquestions).

Nevertheless, this study illustrates a framework for conservation prioritization with decision support tools that improved upon previous methods of conservation planning by incorporating data from multidisciplinary sources and by adding data from multiple taxonomic groups at fine resolution in a mountain region. A framework as presented here but improved with the suggestions above could be applied to create solutions for future conservation goals in this region, but also for other regions with similar data. However, its suitability needs to be tested when using different extents or resolutions. Utilizing spatial decision support tools is just one component of conservation planning (Guisan et al. 2013), which also needs to be addressed through policy, education, and economics (Wilson et al. 2005). As species are expected to continually move with global change, setting conservation targets needs to be an adaptive process that will not create one final optimal answer but several dynamic optima, which vary with changing threats and conservation targets (Singh & Milner-Gulland 2011). A model-based framework, as proposed here, has the additional advantage that it can be implemented in an adaptive manner (Guisan et al. 2006).

ACKNOWLEDGEMENTS

This study was supported by the project "SESAM'ALP - Challenges in simulating alpine species assemblages under global change" (Swiss NSF grant 31003A-1528661), within the scope of a master's thesis by C. Vincent supervised by R. Fernandes and A. Guisan, at the University of Lausanne. We would like to thank former members of the Spatial Ecology group (ECOSPAT) for their great contribution to the collection of the plant and insect species data (especially C. Randin, A. Dubuis, J. Pottier, P. Vittoz), and all the providers of the data obtained from the database of the Info Fauna - Karch (Swiss Amphibian and Reptile Conservation Programme). We would also like to thank B. Barca and S. R. Vincent for their useful comments on early versions of the manuscript.

Chapter 3.2

Species distribution models support the need of international cooperation towards successful management of plant invasions

This manuscript is in review at Journal for Nature Conservation

My contribution to the paper: I conceived the initial idea and developed the methodological framework in coordination with Joana Vicente, João Honrado and Antoine Guisan. I created the codes, analysed the data an also led the writing of the various versions of the manuscript.

Species distribution models support the need of international cooperation towards successful management of plant invasions

Rui F. Fernandes[a*], João P. Honrado[b,c], Antoine Guisan[a,d], Alice Roxo[c], Paulo Alves[b,c], João Martins[c], Joana R. Vicente[b,c]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland
[b] Faculdade de Ciências da Universidade do Porto (Departamento de Biologia), Edifício FC4, Rua do Campo Alegre, S/N, 4169-007 Porto, Portugal and
[c] InBIO - Rede de Investigação em Biodiversidade e Biologia Evolutiva, Laboratório Associado; CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos, Predictive Ecology (PRECOL) research group, Campus Agrário de Vairão, Universidade do Porto 4485-601 Vairão, Portugal
[d] Institute of Earth Science dynamics, Geopolis, University of Lausanne, 1015 Lausanne, Switzerland

ABSTRACT

To protect native biodiversity and habitats from the negative impacts of biological invasions, comprehensive studies and measures to anticipate invasions are required, especially across countries in a transfrontier context. Species distribution models (SDMs) can be particularly useful to integrate different types of data and predict the distribution of invasive species across borders, both for current conditions and under scenarios of future environmental changes. We used SDMs to test whether predicting invasions and potential spatial conflicts with protected areas in a transfrontier context, under current and future climatic conditions, would provide additional insights on the patterns and drivers of invasion when compared to models obtained from predictions for individual regions/countries (different modelling strategies). The framework was tested with the invasive alien plant *Acacia dealbata* in North of Portugal/NW Spain Euro-region, where the species is predicted to increase its distribution under future climatic conditions. While SDMs fitted in a transfrontier context and using "the national strategy (with Portugal calibration data) presented similar patterns, the distribution of the invasive species was higher in the former. The transfrontier strategy expectedly allowed to capture a more complete and accurate representation of the species' niche. Predictions obtained in a transfrontier context are therefore more suitable to support resource prioritisation for anticipation and monitoring impacts of biological invasions, while also providing additional support for international cooperation when tackling issues of global change. Our proposed framework provided useful information on the potential patterns of invasion by *Acacia dealbata* in a transfrontier context, with an emphasis on protected areas. This information is crucial for decision-makers focusing on the prevention of invasions by alien species inside protected areas in a transfrontier context, opening a new way for collaborative management of invasions.

KEYWORDS: Distribution, Invasions, Portugal, SDMs, Spain, Transfrontier context

INTRODUCTION

The global transportation of plant species has occurred for centuries, but due to the intensification of commercial trade and travelling, plant displacement assisted by humans is accelerating (Rejmánek *et al.*, 2013). Many of these species have been introduced in areas where they did not exist before (e.g. for agriculture, forestry or ornamental use), often causing landscape modifications (Vilà & Ibáñez, 2011). Some of the species have become naturalised and even invasive (sensu Richardson *et al.*, 2000) in many areas where they were introduced (Caplat *et al.*, 2013; Rejmánek *et al.*, 2013) and some are now recognised as major threats to native biodiversity (Richardson & Rejmánek, 2011). Invasive species, and especially woody invasive plant species (WIPS), can cause severe negative impacts on invaded ecosystems, such as potentially decrease the number of native species, affect the functions and services provided by an ecosystem, and reshape ecosystems and their interactions at the landscape level (e.g. Vilà *et al.*, 2010; Le Maitre *et al.*, 2011; Vila *et al.*, 2011). These negative impacts can occur mainly because WIPS potentially become dominant in the invaded plant communities (Haugo *et al.*, 2011; Caplat *et al.*, 2013; Pyšek *et al.*, 2013), often originating what has been called ‘novel ecosystems’ (Hobbs *et al.*, 2006).

Due to the strong impacts of plant invasions, increasing attention has been given to the identification of susceptible geographical regions in order to anticipate future invasions (Gundale *et al.*, 2013). Complete control or eradication of established invasive species is difficult and unlikely due to logistical problems and/or financial limitations (Genovesi, 2005; Gallien *et al.*, 2012; Hulme, 2012). Therefore, anticipating the introduction and preventing the expansion of invasive species into a specific region is considered the most cost-effective way of managing biological invasions (Hulme, 2006; Broennimann & Guisan, 2008; Gallien *et al.*, 2012; Petitpierre *et al.*, 2012). Predicting the spatial distribution of invasive species, understanding the ecological requirements of those species and the different environmental drivers that influence their distribution can thus have important impacts in management choices (Gallien *et al.*, 2012; Donaldson *et al.*, 2013; Vicente *et al.*, 2013). Species distribution models (hereafter SDM) are useful tools to address and optimize the management of invasive species (Guisan *et al.*, 2013; Vicente *et al.*, 2016), as they rely on the establishment of statistical relationships between environmental conditions and the occurrence of a given species, thereby providing an effective approach to predict current or future invasive species distributions (Elith & Leathwick, 2009; Hulme, 2012; Guisan *et al.*, 2013).

Predicting and managing invasive species across national borders holds several advantages (Martins *et al.*, 2016), but it can be especially complicated, because, as mentioned by Jaksic *et al.* (2002), different challenges may occur: 1) political, when the firstly invaded country is not concerned by the invasion in some neighbouring country, and the problem becomes the concern of the newly invaded country; 2) ecological, when the habitats invaded in different countries differ in their susceptibility or resilience to invasions; and 3) geographical, because landscape structure determines the corridors and barriers for invasions, which can be different between neighboring countries. Additionally, truncated occurrence datasets are often used to predict species distributions, which might lead to prediction errors resulting from an incomplete representation of the species geographical range (Hannemann *et al.*, 2016).

The invasion of protected areas threatens native biodiversity and causes changes in ecosystem structure and function (Pyšek *et al.*, 2002; Theoharides & Dukes, 2007; Thuiller *et al.*, 2007), representing a serious problem due to the static nature of protected areas (Pressey *et al.*, 2007). The establishment of national parks and natural reserves all over the world, of which the Natura 2000 network in Europe is a remarkable example, has provided areas where native biodiversity can have some protection from several threats, but biological invasions are now threatening their integrity. The protected areas can occur inside a country or across borders therefore calling for cooperation between different countries (Hanks, 2003). Currently, more than 10% of the world's network of protected areas corresponds to areas shared by different countries (Hanks, 2003).

Therefore, it is important to predict current and future invasive species distribution and their potential conflicts with the protected areas not only in a given country (e.g. Vicente *et al.*, 2013), but considering the neighbouring countries together. In such transfrontier context, cooperation between different countries is essential to improve prevention and management of current and future invasions. Since 2014, new European legislation (the Invasive Alien Species EU Regulation No 1143/2014) provides a set of measures that should be taken across European Union countries. A hierarchical approach to combat invasive species was proposed covering prevention, early detection and rapid eradication, management measures that the different countries should apply while also accounting for relevant transboundary impacts and features.

Here, we used species distribution models to predict current and future potential distribution (at 1 km^2 resolution) of the woody invasive plant species *Acacia dealbata* in the northwest Iberian Peninsula (Portugal and Spain), as well as its potential conflicting invasion of protected areas under current and future climatic conditions. We assess what is the most appropriate approach to predict invasive species distribution and anticipate potential conflicts with protected areas in a transfrontier context, to improve management and monitoring actions. Specifically, we tested different modelling strategies to 1) determine how different the forecasts of potential conflicts with protected areas are, and 2) provide additional insights into the patterns and drivers of invasion in a transfrontier context. To address these objectives, three different modelling strategies to fit and project the models were used: (i) national models and predictions, in which the distribution of the species was modelled for the two different countries separately and the predicted patterns were merged afterwards; (ii) national models projected for the whole area, i.e. the distribution of the species was modelled for the full study area, but fitting the models using only the occurrence data available for either Portugal or Spain, and (iii) transfrontier model and prediction, where the distribution of the species was predicted for the full study area using occurrence data for the full study area to fit the models.

METHODS

Study area

The study area is located in the Northwest of the Iberian Peninsula and includes the North of Portugal and the South of Galicia (northwest Spain) (Fig.1), covering an area of 35.017 km^2. It includes the transition between the Mediterranean and the Atlantic biogeographic regions (Rivas-Martínez *et al.*, 2004), and it is a topographically heterogeneous area, with elevation ranging from sea level (at west) to 2050 m in the eastern mountains, resulting in marked variations of environmental conditions. The annual mean temperature ranges from ca. 5ºC to ca. 16ºC, and the total annual precipitation varies between ca. 500 mm and ca. 3000 mm (Mónica & Santos, 2011). These topographic and climatic ranges reflect on a marked ecological heterogeneity, expressed by a complex vegetation cover and a large diversity of land cover and land use types (Caetano *et al.*, 2009; Calvo-Iglesias *et al.*, 2009).

About 27% of the study area is covered by protected areas (conservation of native biodiversity and habitats), most of them (89%) included in the EU's Natura 2000 network. More than 3000 plant taxa are recorded in this area, some of which (>100) are endemic/sub-endemic or have a limited distribution in Iberia (project BIODIV_GNP; visit *www.biodiversidade.eu* for additional information). Despite the presence of high levels of plant biodiversity and protected areas, this region is experiencing a high intensity of environmental disturbances, including an increasing presence and expansion of invasive species (Vicente *et al.*, 2011; Vicente *et al.*, 2013) and the occurrence of large wildfires (Alonso-Betanzos *et al.*, 2003; Fuentes-Santos *et al.*, 2013).

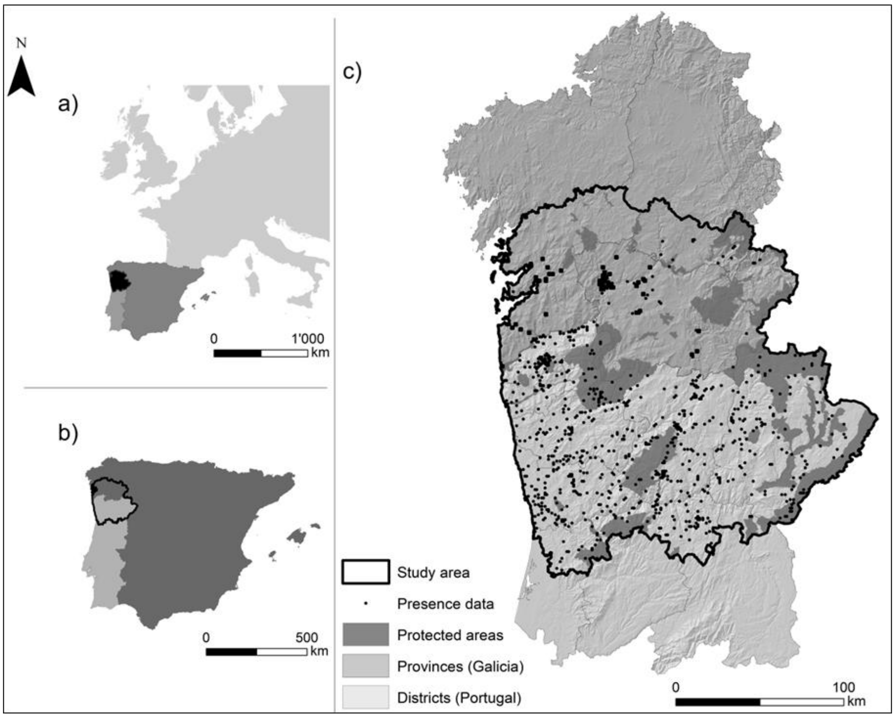


FIGURE 1 Study area location in Europe (a) and in the Iberian Peninsula (b). Representation of the protected areas, presence data available in the study area and its provinces (Galicia) and districts (Portugal) (c).

Test species and occurrence data

Acacia dealbata Link., also known as silver wattle, is a woody plant species of the *Fabaceae* family, native to Australia and Tasmania (Lorenzo *et al.*, 2010). Several species of genus *Acacia* are currently widespread causing severe negative impacts all over the world (Le Maitre *et al.*, 2011). *Acacia dealbata* has become very common in Mediterranean countries after its introduction in Europe (around the 1820s; Carballeira and Reigosa (1999). This species presents a high colonising capability and can be found invading disturbed forests and scrublands, on the edge of rivers and roads (corridors), or grown as an ornamental plant (Lorenzo *et al.*, 2010). The species can replace native vegetation due to the ability to produce a high number of seeds, the seed germination stimulated by fire, and the re-sprouting after cutting, fire or frost (Lorenzo *et al.*, 2010). As a result, the species is often present as dense populations that prevent native vegetation from regenerating and growing (e.g. through competition for resources; Lorenzo *et al.* (2010) and by allelopathic interference; Marchante *et al.* (2011).

The species occurrence dataset applied in this study was collected and harmonized as part of the transfrontier cooperation project BIODIV_GNP (0479_BIODIV_GNP_1_E). Data for Portugal resulted from previous research conducted in the region (Vicente *et al.*, 2010; Vicente *et al.*, 2011; Vicente *et al.*, 2013) and data for Spain was provided by the “Direccion Xeral de Conservacion da Natureza da Xunta da Galicia”. The spatial resolution of the species occurrences was set at 1 km^2.

Environmental data

We initially selected predictors (at 1 km^2 resolution) that, according to expert knowledge and based on previous reports in the scientific literature (e.g. Lorenzo *et al.*, 2010; Vicente *et al.*, 2011; Vicente *et al.*, 2013; Vicente *et al.*, 2016), could act as determinants of the distribution of the test species. Then, using Spearman’s *rho* correlation coefficient, we tested pair-wise correlations to avoid colinearity between those predictors (Dormann *et al.*, 2012). Only predictors with pairwise correlations lower than 0.5 were considered (Elith *et al.*, 2006), thus much lower than the 0.7 maximum value recommended by (Dormann *et al.*, 2012). When correlated pairs of predictors occurred, we kept the predictor with the most direct ecological impact on the species distribution, based on expert judgment or previous findings in the scientific literature (Guisan & Thuiller, 2005). This analysis yielded a final set of 11 environmental predictors to fit SDMs, grouped into four environmental types that reflect distinct ecological drivers of species distributions (climate; landscape composition; soil types; and ecosystem productivity) (see Table S1 in Supplementary Information).

Predictors expressing current climatic conditions were obtained from the Worldclim database (Hijmans *et al.*, 2005); available at http://worldclim.org/download). Predictors related to future climatic conditions were also used to assess future invasion dynamics under climate change scenarios. These predictors were derived from a global circulation model (HadGEM2-ES) and based on the representative concentration pathways RCP4.5 (mean global warming increase of 1ºC) and RCP8.5

(mean global warming increase of 2ºC between 2046-2065 and 3.7ºC between 2081-2100), for the years 2050 and 2070 (available at http://worldclim.org/CMIP5v1). The mentioned pathways were adopted by the IPCC - Intergovernmental Panel on Climate Change - in its Fifth Assessment Report, AR5 (available at http://www.ipcc.ch/report/ar5/index.shtml). Remotely-sensed ecosystem productivity data was obtained from the Numerical Terradynamic Simulation Group database (http://www.ntsg.umt.edu/project/mod17). Finally, predictors describing current land cover were computed based on data acquired from the European Environment Agency database (EEA; available at http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-raster-2).

Analytical framework

In order to analyse the conflicts between the potential distribution of *Acacia dealbata* and the protected areas present in the study area (under current and future climatic conditions, at 1 km^2 resolution), in a transfrontier context, we developed an analytical framework divided into six major sections, as illustrated in Fig. 2:

Data gathering - First, we gathered environmental and species occurrence (presence-only) datasets (see *Environmental data* section for a description of how important predictors were selected). The final occurrence dataset for *A. dealbata* used for model fitting included 967 presence records for the full study area.

Modelling strategies - To assess the differences and potential improvements of predicting invasions in a transfrontier context, we used three different strategies to fit the models: (i) National models and predictions (Fig. 2a), where the distribution of the species was modelled and projected for the two different countries separately, and projections were spatially merged afterwards; (ii) National models projected for the whole area (Fig. 2b), where the distribution of the species was projected for the full study area, but models were calibrated using the occurrence data available only for Portugal or for Spain; and (iii) Transfrontier model and prediction (Fig. 2c), where the distribution of the species was modelled and projected for the whole study area using occurrence data also for the whole area. To fit the models in each approach, we used only data available for the specific areas, so that data limitation due to lack of cooperation between entities of the different countries could be simulated (in the first two approaches) and compared with a scenario where cooperation between countries exists, in order to follow the EU regulation for the prevention, early detection, rapid eradication and management (transfrontier context; third approach). Thus, the occurrence dataset used for the calibration of the individual models consisted of 746 presences (only Portugal) and 221 presences (only Spain). The occurrence data set used to fit the models for the full area consisted of 967 presences.

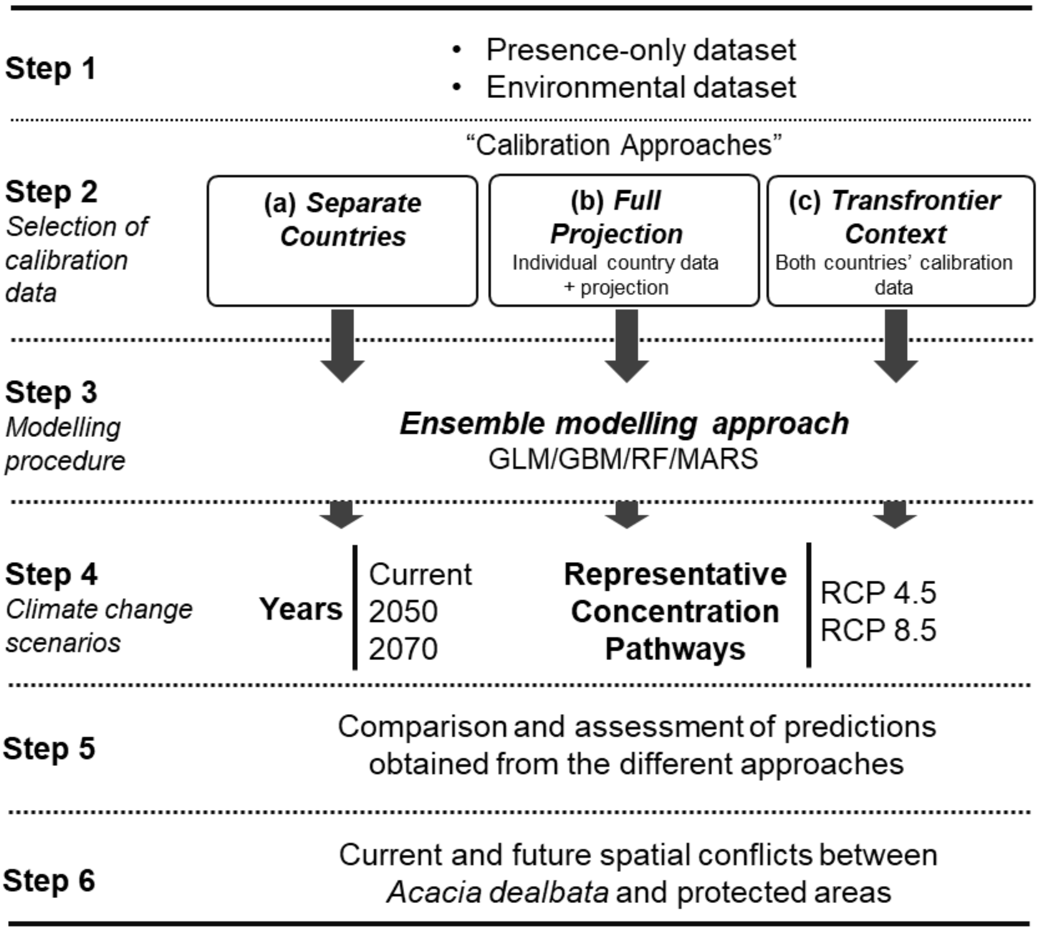


FIGURE 2 Analytical framework for modelling the distribution of the species *Acacia dealbata* and assess the spatial conflicts with the protected areas present in the study area. First, we collected environmental variables and presence-only occurrence data for the test species - Step 1. Then we applied different "calibration approaches" to calibrate several models for the test species - Step 2: (a) Separate Countries data (calibration and projection individually for each one of the two different countries and spatially combined to obtain the final output); (b) Full Projections (projection of the model for the full study area, but using either the occurrence data only for Portugal or for Galicia for calibration purposes), and (c) Transfrontier Context (calibration and projection for the full study area with occurrence data from the full area). The models were calibrated using an ensemble modelling approach available in biomod2, by fitting generalised linear models (GLM), generalised boosted models (GBM), random forests (RF) and multiple adaptive regression splines (MARS) - Step 3. For each "calibration approach", models were then projected for the entire study area for current and future (2050 and 2070, under the representative concentration pathways RCP4.5 and RCP8.5) climatic conditions - Step 4. Finally, we compared and assessed the predictions obtained from the different "calibration approaches" - Step 5 - and analysed the potential conflicts between the predicted distribution of *A. dealbata* and the protected areas located in the study area - Step 6.

Modelling procedure - We implemented a modelling procedure using the *biomod2* package (Thuiller *et al.*, 2009); available at http://cran.r-project.org/web/packages/biomod2/index.html) in the *R* statistical software. SDMs for the current distribution of *A. dealbata* were fitted and then projected under current and future environmental conditions to predict the species' potential distribution, at 1 km^2 resolution. The models were fitted using an ensemble modelling approach available in *biomod2*, by fitting generalised linear models (GLM), generalised boosted models (GBM), random forests (RF) and multiple adaptive regression splines (MARS) (for more details see *biomod2* help files and vignettes). As the occurrence dataset consisted only of confirmed presences for the study area (and the actual number of presence records was dependent on the "calibration approach" used; see Step 2), 10 pseudo-absences datasets consisting of 1000 pseudo-absences each were randomly generated (i.e. using the user-defined option in *biomod2* to create pseudo-absences) to fit the models (Barbet-Massin *et al.*, 2012). Also, we ensured a prevalence of 0.5, which means that the presences and the pseudo-absences have the same importance in the model calibration process. Each individual model was fitted using 70% of the available data and evaluated on the remaining 30%. This cross-validation procedure was repeated 10 times, and a total of 400 models were run (4 algorithms x 10 repetitions x 10 pseudo-absences datasets). For each model, a single ensemble model was obtained by applying the weighted mean of probabilities consensus method (i.e. estimates the weighted sum of probabilities; see (Marmion *et al.*, 2009). This consensus method provides more robust predictions than single models or other consensus methods (Marmion *et al.*, 2009). Finally, model projections were reclassified into presence-absence using a threshold that corresponds to the point on the Receiver Operating Characteristic (ROC) plot (sensitivity against 1-specificity) with the shortest distance to the top-left corner (0,1) of the plot (Swets, 1988; Fielding & Bell, 1997). The final ensemble models were evaluated primarily using the area under the curve (AUC) of the ROC plot while also calculating the True Skill Statistics (TSS; Allouche *et al.*, 2006), Cohen's Kappa (Cohen, 1968) and a "presence-only" evaluation metric, the Boyce index (Hirzel *et al.*, 2006).

Model predictions - For each modelling strategy, spatial predictions of the species' distribution were made over the full geographical extent of the study area (i.e. individual regions – North of Portugal, NW Spain - or the full area). The outcomes from climate change scenarios were obtained projecting the models for 2050 and 2070, using either RCP4.5 or RCP8.5. In the merged strategy, the spatial predictions of the two individual regions/countries were spatially merged (using ArcGIS 10). By comparing the predictions obtained using different calibration strategies, we were able to assess the best method to predict invasions and conflicts with protected areas (Step 5) in a transfrontier context.

Conflicts with protected areas - To analyse the potential conflicts between the predicted distribution of *A. dealbata* and the protected areas located in the full study area, we spatially overlapped the distribution maps (obtained in the previous step) with a spatial mask containing only protected areas (using ArcGIS 10; ESRI, 2011).

Comparison of modelling strategies - Finally, we used model outputs to compare the three modelling strategies according to the following criteria: (i) model accuracy (AUC, TSS, Boyce); (ii) raking of predictor importance; (iii) current and future area predicted as suitable for *A. dealbata*; and (iv) current and future predicted conflicts inside protected areas.

RESULTS

Model accuracy

All final ensemble models, independently of the modelling strategy, presented AUC values higher than 0.8 (Table 1). This means that all models can be considered useful for predictions, according to the interpretation scales described in Araujo et al. (2005; fail: AUC<0.7, fair: >0.7, good >0.8, excellent >0.9). The highest AUC values were observed in the national models of NW Spain (both for individual predictions or when projected to the whole area). Additionally, the Boyce index results were also higher than 0.96 for all the strategies, with the transfrontier model presenting the highest value (0.999). Model evaluations based on TSS and Kappa varied among approaches (Table 1). For both metrics, the highest values were observed for the national model and prediction for NW Spain (TSS=0.791; Kappa=0.435). The transfrontier model presented the lowest value of Kappa (0.194).

TABLE 1 Evaluation metrics (i.e., accuracy measures; AUC, TSS, Kappa and Boyce index) for the ensemble models of the different "calibration approaches". In bold, the highest values in each metric.

Approach	**AUC**	**TSS**	**Kappa**	**Boyce index**
Separate Countries (Portugal)	0.829	0.510	0.304	0.996
Separate Countries (Galicia)	**0.953**	**0.791**	**0.435**	0.978
Full projection (Portugal)	0.832	0.516	0.302	0.996
Full projection (Galicia)	**0.953**	0.786	0.426	0.962
Transfrontier context	0.840	0.530	0.194	**0.999**

Ranking of predictor importance

For all the different modelling strategies, the most important variable was the minimum temperature of the coldest month (Table S2). This predictor was followed by annual precipitation and temperature annual range, except in the transfrontier model, where temperature annual range and precipitation seasonality were the second and third most important variables, respectively.

Current and future potential distribution of Acacia dealbata

The distribution of *Acacia dealbata* is predicted to increase under both future climatic scenarios, with this tendency expectedly more pronounced under the climatic scenario RCP8.5 (Table S3). When considering the distribution in the whole study area (Fig. 3), suitable areas are predicted to increase in the future regardless of the modelling strategy and for both climatic scenarios.

Depending on the modelling strategy, the predicted presence areas can differ, with ca. of 30% in the national models and predictions strategy and 34% in the transfrontier model, under current conditions. This difference further increases under the projections for 2070, with a presence area of 63% (RCP4.5) and 78% (RCP8.5) for the national models and predictions, and 75% (RCP4.5) and 83% (RCP8.5) in the transfrontier model.

Despite all the models presented AUC values higher than 0.8 (as well as Boyce index above 0.9; Table 1), important differences in the distribution and invasion patterns were observed when considering the spatial projections of the modelling approaches (Figure 3). The predictions obtained through the combination of individual models (national models and predictions) presented the most contrasting patterns between North of Portugal and NW of Spain (Fig. 3, first row). Considering the prediction obtained by the national models projected into the whole area, using either North of Portugal or NW of Spain calibration datasets, it is possible to denote very different patterns of invasion (Fig. 3). Predictions obtained by the transfrontier strategy and the national model of Portugal projected for the full area exhibit similar patterns of distribution of *A. dealbata* (especially under climate change scenarios), with a higher predicted distribution in the transfrontier model (Fig. 3).

The transfrontier strategy and the Portugal national model projected into the full area predict an increase of the species' distribution towards more interior areas (Fig. 3), while the opposite occurs in the Spain national model projected to the whole area, with a big spatial gap in the distribution of the species between coastal areas and the interior of NW Spain. This gap is particularly evident for year 2070 in RCP4.5, and similar patterns can be observed for year 2050 under both RCPs.

Current and future predict conflicts inside protected areas

When considering the distribution of the species only inside protected areas (Table 2), the same previous patterns are observed, with additional presence areas predicted by the transfrontier model compared to the other approaches under both climate change scenarios. While only around 15% of protected areas are predicted as suitable for the invasive species under current conditions, under future conditions and especially in the transfrontier model, it can reach 49% (RCP4.5) or 65% (RCP8.5) in 2070. Taking into account the spatial distribution of the species inside protected areas, specifically for the transfrontier model (Figure 4), it was possible to predict an increase in the distribution of *A. dealbata* regardless the climatic scenario. The predicted future suitable areas are mainly located in the east and centre of Portugal but also at the north of Spain, being this pattern consistent both for 2050 and 2070.

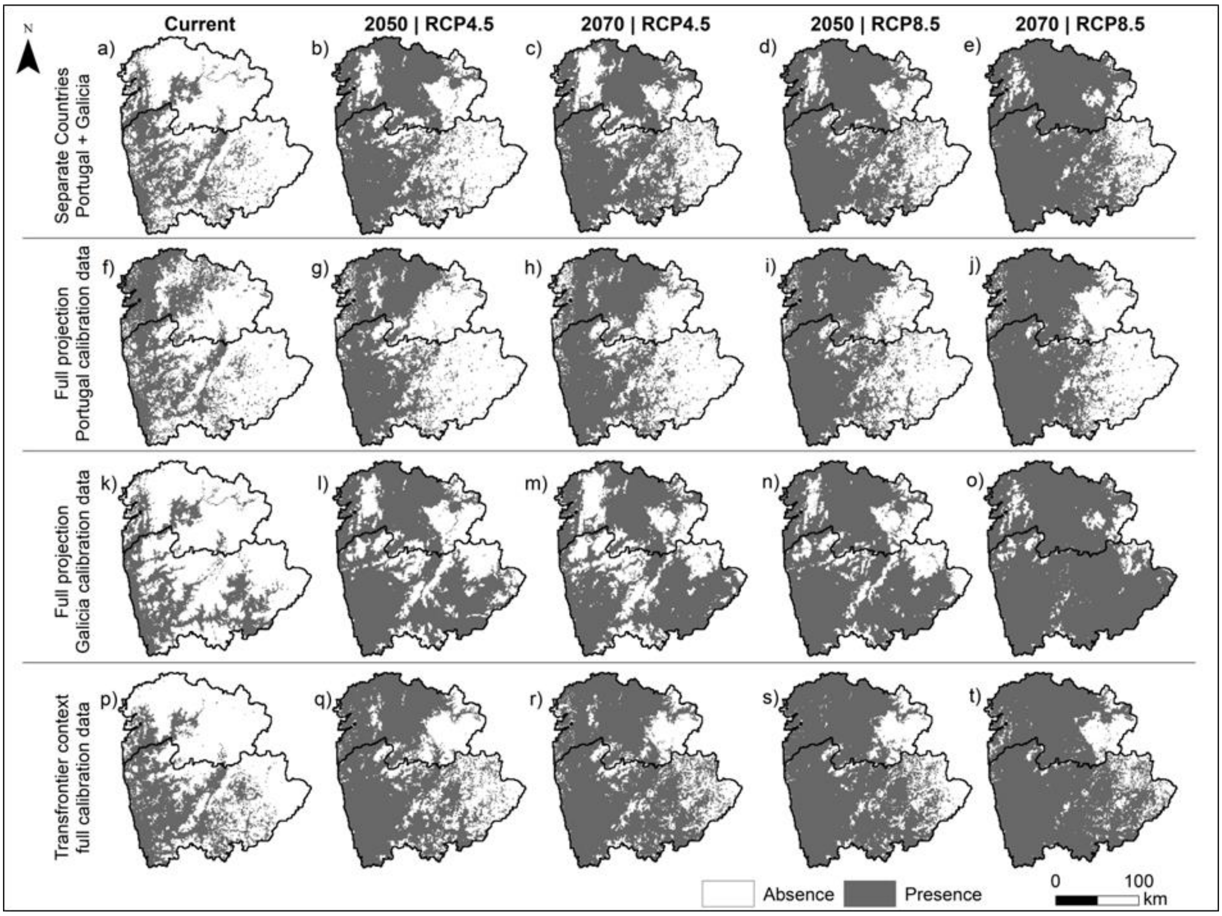


FIGURE 3 Spatial potential distribution of *A. dealbata* obtained from the different “calibration approaches”: using separate countries calibration data (first row, a), b), c), d), e)), full projections using data available for Northern Portugal only (second row, f), g), h), i), j), full projections using data available for Galicia only (third row) and the transfrontier context, where the distribution of the species was modelled and projected for the full study area using occurrence data also for the full area (bottom row). The models were calibrated for current climatic conditions (first column), for 2050 (second and fourth columns) and 2070 (third and fifth columns) for the two selected representative concentration pathways (RCP4.5 and RCP8.5).

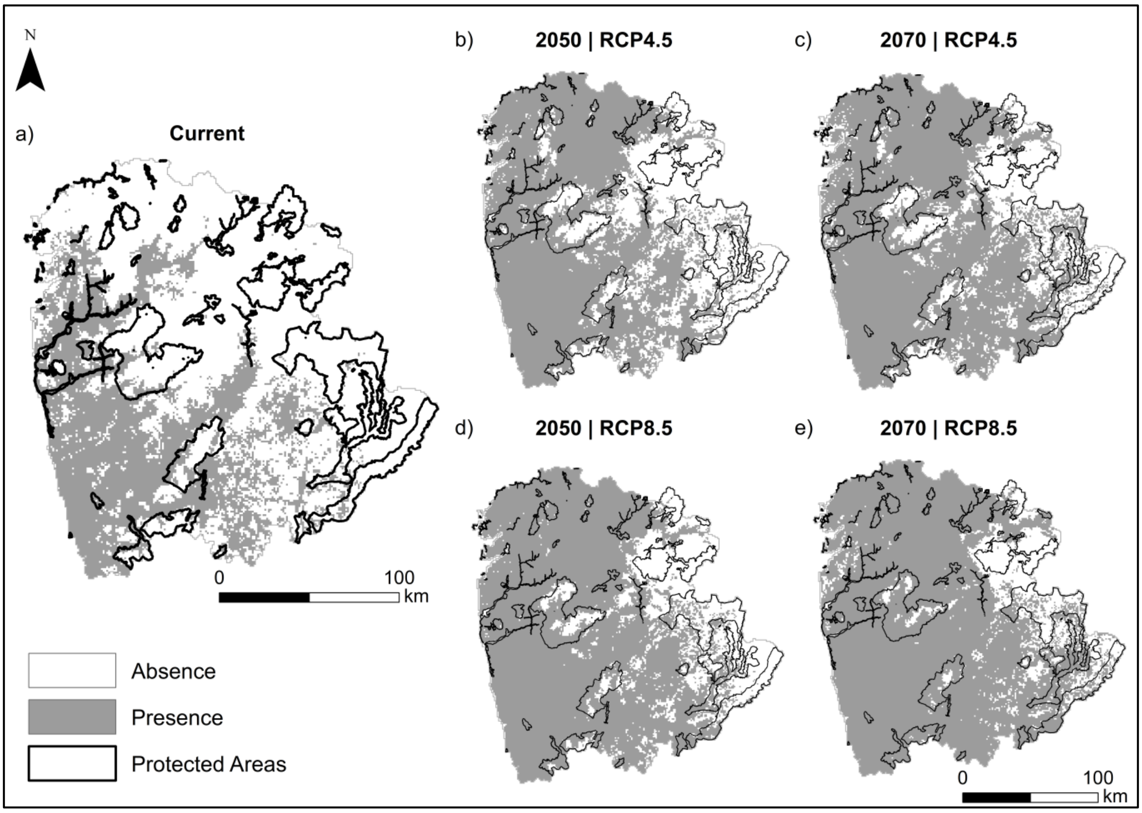


FIGURE 4 Potential distribution of *A. dealbata* across the full area and inside protected areas, applying the "transfrontier context" approach, under current climatic conditions (a) model projections for 2050 (b; d) and for 2070 (c; e), for the two selected representative concentration pathways RCP 4.5 (less extreme; b) and c)) and RCP 8.5 (more extreme; d) and e)).

DISCUSSION

Current and future distribution of Acacia dealbata, and conflicts with protected areas in a transfrontier context

Under current conditions, our projections show that the area where *Acacia dealbata* is predicted as present reflects the majority of the study area. This is particularly alarming considering the amount of protected areas suitable for the species invasion (especially under extreme climate change scenarios). These results are in line with the patterns of invasion previously reported by Vicente *et al.* (2013) for *A. dealbata* in the North of Portugal, who considered this species to be one of the most problematic invasive species in the region, taking into account its distribution pattern and dynamics of invasion. The projections of the current distribution of the species might also suggest that the main areas of introduction of *A. dealbata* are in the west, with highways and rivers apparently acting as the main dispersal corridors (see Fig. 4). Although this was not something directly studied here, it is consistent with previous literature showing that rivers and roads can enhance the spread of invasive species (Säumel & Kowarik, 2010),

Our predictions also stress that the spatial conflicts between the species and protected areas will mainly occur in the western part of the study area, in agreement with previous studies developed by

Vicente *et al.* (2010) and by Vicente *et al.* (2013). We can also observe that some protected areas in the inland part of the Spain region are predicted to be occupied by the invasive species. This may relate to the fact that additional areas of introduction can be located outside the study area (e.g. in the coastal region of Coruña, North of Spain) and also to the high concentration of dispersal corridors in that region (see e.g. Simberloff, 1988; Saunders & Hobbs, 1991; Proches *et al.,* 2005). Predictions of future conflicts support the recommendation that prevention measures should be applied in protected areas located in the eastern part of the study area, to avoid further establishment or expansion of *A. dealbata*, as we have predicted an even larger future area of conflict with protected areas under both climate change scenarios. These measures, fulfilling the previously mentioned EU regulation on Invasive Alien Species, would certainly be more effective if useful models (in a transfrontier context) could be used to aid in the decision-making process. Prevention and control measures should also be conducted in protected areas located in the western part of the region, to avoid additional and more severe invasions by *A. dealbata* and other alien invasive species. These invasions will be enhanced by future climatic and fire conditions, with an expected increase in the frequency of wildfires due to climate change (IPCC, 2013), which in turn will boost seed germination of *A. dealbata* (Lorenzo *et al.*, 2010; Souza-Alonso *et al.*, 2017), increasing their distribution areas.

The added value of predicting invasions and conflicts with protected areas in a transfrontier context

It is widely accepted that the most efficient way to manage biological invasions is through effective prevention and anticipation of invasions (Hulme, 2006; Broennimann & Guisan, 2008; Gallien *et al.*, 2012; Guisan *et al.*, 2013). To do so, the best available tools and frameworks should be applied to understand the influence that different environmental drivers can have on the distribution of invasive species and thus to improve or support management decisions (Gallien *et al.*, 2012; Donaldson *et al.*, 2013; Guisan *et al.*, 2013). International cooperation is of the utmost importance in invasive species management, as exemplified by callings for a coordinated strategy on invasive species at the European level (Commission of the European Communities, 2008; Hulme *et al.*, 2009). Here we addressed the added value of analysing invasions in a transfrontier context, providing evidence that such approach allows for additional insights on the patterns and drivers of invasion, resulting in different forecasts of conflicts when compared to the analyses of the two countries separately (or using calibration data of a single region and projecting for the full area), and thereby having the potential to improve management and monitoring decisions.

When trying to predict and prevent invasions, it is important to produce accurate forecasts, able to support and improve management actions. The ability to obtain accurate predictions from SDMs can be improved by increasing the sample size used to fit the models (Wisz *et al.*, 2008). Here, we have shown that analysing invasion in a transfrontier context has important advantages when compared to predicting and analysing invasions of countries that share a common border (and consequently can share a common invasion) individually. This may be because, by including the full records available for the area of interest into a single model calibration, instead of multiple calibrations with data from each individual country, the sample size of the calibration dataset considerably increased. At the same time, the larger extent may allow the model to capture a more accurate representation of the species' niche (Barbet-Massin *et al.*, 2010; Suárez-Seoane *et al.*, 2014), as areas with a particular set of suitable conditions may be present in only one of the areas. While the effect of sample size and larger extents are widely known, our results suggest that the prediction, prevention, and management of possible conflicts in protected areas should be improved with transfrontier analysis. However, even with all knowledge this is not commonly done, and our work can, therefore, encourage international cooperation in data sharing as well as policy and management initiatives.

Also, due to its increased exactitude when compared to predictions obtained from models for individual regions, predictions made in a transfrontier context could allow additional insights into the factors that are driving each invasion process. This will also permit more accurate forecasting of potential distributions under future conditions, and therefore for improved predictions of future dynamics and conflicts. Combining improved model predictions with robust analytical frameworks, it is possible to obtain better information regarding the threat of invasive species for protected areas (Vicente *et al.*, 2013).

Our work also raises an important challenge for SDM research and application. In our transfrontier context, we used data from different countries, resulting from previous separate research and management activities, and thus collected in different ways. As a result, we had available presence-absence data for one region and presence-only data for another, which needed to be combined in a same model fitting procedure. We employed one possible solution to use these data, namely to use sets of pseudo-absences for the full area and removing the recorded absences from the analyses, but there may be other approaches that need to be tested (i.e. hierarchical modelling approach; Pearson *et al.*, 2004; Petitpierre *et al.*, 2015). However, it is important to note that neighbouring countries would benefit from joining efforts and establishing common monitoring/prevention measures regarding invasive species control. This was also recognized under the EU regulation on Invasive Alien Species, which provides guidance and financial support for trans-border cooperation to prevent the spread, eradication and control of invasive species.

Current and future challenges for the conservation of protected areas in a transfrontier context

Designating and managing protected areas is one of the most important approaches to preserve global biodiversity, but as mentioned in Pyšek *et al.* (2002), invasive species are present in many protected reserves worldwide. In some cases, protected areas are a last refuge for endemic, rare or endangered species, capturing important biodiversity and mitigating possible negative effects of external pressures (i.e. climate change, habitat fragmentation, invasive species; Gaston *et al.*, 2008). Assessing these pressures, especially the conflicts and impacts of invasive species, so they can be effectively anticipated and addressed, has become an important effort for the scientific community (Davis *et al.*, 2005). A review of 199 articles describing the impacts of 135 alien plant taxa assessed that the diversity and abundance of native species decreased in invaded sites, providing evidence that invasive plants can cause important ecological impacts (see Vila *et al.*, 2011). Vicente *et al.* (2013) also showed that protected areas are predicted to suffer impacts caused by invasive species under future climatic conditions.

Here, our framework provided useful information on the potential patterns of invasion by *Acacia dealbata* in a transfrontier context, with an emphasis on protected areas. This information can be crucial for decision makers focusing on the prevention of invasions by alien species inside protected areas in a transfrontier context, opening a new way for collaborative management of invasions. Fairly recent policy initiatives also highlight the need for international cooperation to tackle this global change driver (e.g. G8, 2009). Despite this, additional efforts are needed to improve monitoring and management actions on threatened protected areas, mainly because of the lack of cooperation/communication between the scientific community, stakeholders and decision makers (Guisan *et al.*, 2013).

CONCLUSIONS

In this work, we expect to contribute to the improvement of management and monitoring actions in a transfrontier context using SDMs, thereby helping to implement recent EU regulation for the prevention, eradication and management of woody invasive plant species. With our proposed framework, we assessed the most appropriate approach to predict the distribution of Acacia dealbata, and their potential conflicts with protected areas in a transfrontier context. Therefore, considering our framework and results we would like to highlight that:

1) the distribution of *Acacia dealbata* was predicted to increase under both climatic change scenarios, regardless of the “calibration approach”
2) the species distribution areas are higher under the transfrontier context, both for the full and protected areas
3) all the “calibration approaches” presented high evaluation values, with the models being considered as having excellent predictive accuracy
4) the minimum temperature of the coldest month and annual precipitation were the most important variables to explain the distribution of the species
5) differences in spatial patterns were observed under the different “calibration approaches”
6) spatial conflicts between the species and protected areas were predicted to mainly occur in the western part of the study area

7) the transfrontier approach allows for additional insights on the current and future patterns of invasion, having important advantages when compared to predicting and analysing invasions of individual countries that share a common border

ACKNOWLEDGEMENTS

We thank Blaise Petitpierre and Oliver Broennimann for the productive discussions and the anonymous reviewers that helped to improve this manuscript. This research was developed as part of the project “Biodiversidad Vegetal Amenazada Galicia-Norte de Portugal. Conocer, gestionar e implicar” (0479_BIODIV_GNP_1_E), funded by "2º Eixo prioritário – Cooperação e Gestão Conjunta e Meio Ambiente, Património e Gestão de Riscos" of the "Programa Operacional de Cooperação Transfronteiriça Espanha – Portugal (POCTEP) 2007-2013". Joana R. Vicente is supported by the Portuguese Science Foundation (FCT) through Post-doctoral grant SFRH/BPD/84044/2012.

References

Alagador, D., Cerdeira, J.O., Araújo, M.B. & Saura, S. (2014) Shifting protected areas: scheduling spatial priorities under climate change. *Journal of Applied Ecology*, **51**, 703-713.

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

Alonso-Betanzos, A., Fontenla-Romero, O., Guijarro-Berdiñas, B., Hernández-Pereira, E., Inmaculada Paz Andrade, M.a., Jiménez, E., Luis Legido Soto, J. & Carballas, T. (2003) An intelligent system for forest fire risk prediction and fire fighting management in Galicia. *Expert Systems with Applications*, **25**, 545-554.

Araujo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species-climate impact models under climate change. *Global Change Biology*, **11**, 1504-1513.

Araujo, M.B., Cabeza, M., Thuiller, W., Hannah, L. & Williams, P.H. (2004) Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, **10**, 1618-1626.

Araújo, M.B., Alagador, D., Cabeza, M., Nogues-Bravo, D. & Thuiller, W. (2011) Climate change threatens European conservation areas. *Ecol Lett*, **14**, 484-92.

Arlettaz, R., Schaub, M., Fournier, J., Reichlin, T.S., Sierro, A., Watson, J.E. & Braunisch, V. (2010) From publications to public actions: when conservation biologists bridge the gap between research and implementation. *BioScience*, **60**, 835-842.

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? *Ecography*, **33**, 878-886.

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

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

Bosshard, T., Kotlarski, S., Ewen, T. & Schär, C. (2011) Spectral representation of the annual cycle in the climate change signal. *Hydrology and Earth System Sciences*, **15**, 2777-2788.

Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5-32.

Breiner, F.T., Guisan, A., Bergamini, A., Nobis, M.P. & Anderson, B. (2015) Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, **6**, 1210-1218.

Broennimann, O. & Guisan, A. (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biol Lett*, **4**, 585-9.

Bugmann, H., A.B. Gurung, F. Ewert, W. Haeberli, A. Guisan, D. Fagre, A. Kääb, & GLOCHAMORE participants (2007) Modeling the Biophysical Impacts of Global Change in Mountain Biosphere Reserves. *Mountain Research and Development*, **27**, 66-77.

Caetano, M., Nunes, V. & Nunes, A. (2009) CORINE Land Cover 2006 for Continental Portugal, Relatório técnico. In. Instituto Geográfico Português.

Calvo-Iglesias, M.S., Fra-Paleo, U. & Diaz-Varela, R.A. (2009) Changes in farming system and population as drivers of land cover and landscape dynamics: The case of enclosed and semi-openfield systems in Northern Galicia (Spain). *Landscape and Urban Planning*, **90**, 168-177.

Canessa, S., Guillera-Arroita, G., Lahoz-Monfort, J.J., Southwell, D.M., Armstrong, D.P., Chadès, I., Lacy, R.C., Converse, S.J. & Gimenez, O. (2015) When do we need more data? A primer on calculating the value of information for applied ecologists. *Methods in Ecology and Evolution*,

6, 1219-1228.

Caplat, P., Hui, C., Maxwell, B.D. & Peltzer, D.A. (2013) Cross-scale management strategies for optimal control of trees invading from source plantations. *Biological Invasions*, **16**, 677-690.

Carballeira, A. & Reigosa, M.J. (1999) Effects of natural leachates of Acacia dealbata link in Galicia (NW Spain). *Botanical Bulletin of Academia Sinica*, **40**, 87-92.

Cardoso, A. (2015) *Spatial data and modelling for the prioritisation of conservation areas in the alpine region of the canton of Vaud*. Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa, Portugal.

CBD. (2010) Global Biodiversity Outlook 3. Secretariat of the Convention on Biological Diversity (CBD). In, p. 94 pp, Montreal.

Chen, I.C., Hill, J.K., Ohlemuller, R., Roy, D.B. & Thomas, C.D. (2011) Rapid range shifts of species associated with high levels of climate warming. *Science*, **333**, 1024-6.

Cohen, J. (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*, **70**, 213-20.

Commission of the European Communities (2008) Towards an EU strategy on Invasive Species. In: *COM2008 (789)*, Brussels.

D'Amen, M., Pradervand, J.-N. & Guisan, A. (2015) Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography*, **24**, 1443-1453.

Davis, M.A., Thompson, K. & Philip Grime, J. (2005) Invasibility: the local mechanism driving community assembly and species diversity. *Ecography*, **28**, 696-704.

Díaz, S., Demissew, S., Carabias, J., Joly, C., Lonsdale, M., Ash, N., Larigauderie, A., Adhikari, J.R., Arico, S. & Báldi, A. (2015) The IPBES Conceptual Framework—connecting nature and people. *Current Opinion in Environmental Sustainability*, **14**, 1-16.

Dicks, L.V., Walsh, J.C. & Sutherland, W.J. (2014) Organising evidence for environmental management decisions: a ‘4S’hierarchy. *Trends in ecology & evolution*, **29**, 607-613.

Donaldson, J.E., Richardson, D.M. & Wilson, J.R.U. (2013) Scale-area curves: a tool for understanding the ecology and distribution of invasive tree species. *Biological Invasions*, **16**, 553-563.

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. *Journal of Biogeography*, **39**, 2119-2131.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122-1131.

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677-697.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. & E. Zimmermann, N. (2006) Novel methods improve prediction of species’ distributions from occurrence data. *Ecography*, **29**, 129-151.

Engler, R., Randin, C.F., Vittoz, P., Czáka, T., Beniston, M., Zimmermann, N.E. & Guisan, A. (2009) Predicting future distribution of mountain plants under climate change: does dispersal and landscape matter? *Ecography*, **32**, 34-45.

ESRI (2011) *ArcGIS Desktop: Release 10*.

Faleiro, F.V., Machado, R.B. & Loyola, R.D. (2013) Defining spatial conservation priorities in the face of land-use and climate change. *Biological Conservation*, **158**, 248-257.

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

Fleishman, E., Thomson, J., Kalies, E., Dickson, B., Dobkin, D. & Leu, M. (2014) Projecting current and future location, quality, and connectivity of habitat for breeding birds in the Great Basin. *Ecosphere*, **5**, 1-29.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K. & West, P.C. (2011) Solutions for a cultivated planet. *Nature*, **478**, 337.

Françoso, R.D., Brandão, R., Nogueira, C.C., Salmona, Y.B., Machado, R.B. & Colli, G.R. (2015) Habitat loss and the effectiveness of protected areas in the Cerrado Biodiversity Hotspot. *Natureza & Conservação*, **13**, 35-40.

Fuentes-Santos, I., Marey-Perez, M.F. & Gonzalez-Manteiga, W. (2013) Forest fire spatial pattern analysis in Galicia (NW Spain). *J Environ Manage*, **128**, 30-42.

G8 (2009) Carta di Siracusa on Biodiversity, adopted by the G8 Environment ministers at theirmeeting in Siracusa, 22–24 April 2009. Available at: www.g8.utoronto.ca/environment/env090424-biodiversity.pdf. Accessed 10 January 2015.

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? *Global Ecology and Biogeography*, **21**, 1126-1136.

Gaston, K.J., Jackson, S.F., Cantú-Salazar, L. & Cruz-Piñón, G. (2008) The Ecological Performance of Protected Areas. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 93-113.

Genovesi, P. (2005) Eradications of invasive alien species in Europe: a review. *Biological Invasions*, **7**, 127-133.

Gibson, L., Lee, T.M., Koh, L.P., Brook, B.W., Gardner, T.A., Barlow, J., Peres, C.A., Bradshaw, C.J., Laurance, W.F. & Lovejoy, T.E. (2011) Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, **478**, 378.

Grant, E.H.C., Zipkin, E.F., Nichols, J.D. & Campbell, J.P. (2013) A strategy for monitoring and managing declines in an amphibian community. *Conservation biology*, **27**, 1245-1253.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) *Habitat Suitability and Distribution Models: With Applications in R.* Cambridge University Press.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. & Zimmermann, N.E. (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**, 501-511.

Guisan, A., Broennimann, O., Buri, A., Cianfrani, C., D'Amen, M., Di Cola, V., Fernandes, R., Gray, S., Mateo, R.G., Pinto, E., Pradervand, J.-N., Scherrer, D., von Däniken, I., Yashiro, E. & Vittoz, P. (In press) Climate change impact on mountain biodiversity. *Climate Change and Biodiversity* (ed. by T.E. Lovejoy and L. Hannah). Yale University Press.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecol Lett*, **16**, 1424-35.

Gundale, M.J., Pauchard, A., Langdon, B., Peltzer, D.A., Maxwell, B.D. & Nuñez, M.A. (2013) Can model species be used to advance the field of invasion ecology? *Biological Invasions*, **16**, 591-607.

Hanks, J. (2003) Transfrontier Conservation Areas (TFCAs) in Southern Africa. *Journal of Sustainable Forestry*, **17**, 127-148.

Hannemann, H., Willis, K.J. & Macias-Fauria, M. (2016) The devil is in the detail: unstable response functions in species distribution models challenge bulk ensemble modelling. *Global Ecology and Biogeography*, **25**, 26-35.

Haugo, R.D., Halpern, C.B. & Bakker, J.D. (2011) Landscape context and long-term tree influences shape the dynamics of forestmeadow ecotones in mountain ecosystems. *Ecosphere*, **2**, art91.

Heller, N.E. & Zavaleta, E.S. (2009) Biodiversity management in the face of climate change: a review of 22 years of recommendations. *Biological conservation*, **142**, 14-32.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, **199**, 142-152.

Hobbs, R.J., Arico, S., Aronson, J., Baron, J.S., Bridgewater, P., Cramer, V.A., Epstein, P.R., Ewel, J.J., Klink, C.A., Lugo, A.E., Norton, D., Ojima, D., Richardson, D.M., Sanderson, E.W., Valladares, F., Vila, M., Zamora, R. & Zobel, M. (2006) Novel ecosystems: theoretical and management aspects of the new ecological world order. *Global Ecology and Biogeography*, **15**, 1-7.

Hulme, P.E. (2006) Beyond control: wider implications for the management of biological invasions. *Journal of Applied Ecology*, **43**, 835-847.

Hulme, P.E. (2012) Weed risk assessment: a way forward or a waste of time? *Journal of Applied Ecology*, **49**, 10-19.

Hulme, P.E., Pysek, P., Nentwig, W. & Vila, M. (2009) Ecology. Will threat of biological invasions unite the European Union? *Science*, **324**, 40-1.

IPCC (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jaksic, F.M., Iriarte, J.A., Jiménez, J.E. & Martínez, D.R. (2002) Invaders Without Frontiers: Cross-border Invasions of Exotic Mammals. *Biological Invasions*, **4**, 157-173.

Janišová, M., Bartha, S., Kiehl, K. & Dengler, J. (2011) Advances in the conservation of dry grasslands: Introduction to contributions from the seventh European Dry Grassland Meeting. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*, **145**, 507-513.

Jenkins, C.N., Van Houtan, K.S., Pimm, S.L. & Sexton, J.O. (2015) US protected lands mismatch biodiversity priorities. *Proc Natl Acad Sci U S A*, **112**, 5081-6.

Knight, A. & Cowling, R. (2003) Conserving South Africa’s “lost” biome: a framework for securing effective regional conservation planning in the Subtropical Thicket Biome. *Port Elizabeth: University of Port Elizabeth. Terrestrial Ecology Research Unit report*,

Kujala, H., Moilanen, A., Araújo, M.B. & Cabeza, M. (2013) Conservation planning with uncertain climate change projections. *PloS one*, **8**, e53315.

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Fischer, G. & Folke, C. (2001) The causes of land-use and land-cover change: moving beyond the myths. *Global environmental change*, **11**, 261-269.

Le Maitre, D.C., Gaertner, M., Marchante, E., Ens, E.-J., Holmes, P.M., Pauchard, A., O’Farrell, P.J., Rogers, A.M., Blanchard, R., Blignaut, J. & Richardson, D.M. (2011) Impacts of invasive Australian acacias: implications for management and restoration. *Diversity and Distributions*, **17**, 1015-1029.

Lemes, P. & Loyola, R.D. (2013) Accommodating species climate-forced dispersal and uncertainties in spatial conservation planning. *PloS one*, **8**, e54323.

Liu, C.R., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385-393.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J. & Guisan, A. (2010) Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, **143**, 2647-2657.

Lorenzo, P., González, L. & Reigosa, M.J. (2010) The genus Acacia as invader: the characteristic case of Acacia dealbata Link in Europe. *Annals of Forest Science*, **67**, 101-101.

Marchante, H., Freitas, H. & Hoffmann, J.H. (2011) Post-clearing recovery of coastal dunes invaded by Acacia longifolia: is duration of invasion relevant for management success? *Journal of Applied Ecology*, **48**, 1295-1304.

Margules, C.R.P., R.L (2000) Systematic conservation planning. *Nature*, **405**, 243-253.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59-69.

Martins, J., Richardson, D.M., Henriques, R., Marchante, E., Marchante, H., Alves, P., Gaertner, M., Honrado, J.P. & Vicente, J.R. (2016) A multi-scale modelling framework to guide management of plant invasions in a transboundary context. *Forest Ecosystems*, **3**

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models. 2nd edition*. Chapman and Hall, London.

Meir, E., Andelman, S. & Possingham, H.P. (2004) Does conservation planning matter in a dynamic and uncertain world? *Ecology Letters*, **7**, 615-622.

Millennium Ecosystem Assessment (2005) Ecosystems and human well-being: Synthesis. In, Washington, DC.

Moilanen, A., Wintle, B.A., Elith, J. & Burgman, M. (2006) Uncertainty analysis for regional-scale reserve selection. *Conservation Biology*, **20**, 1688-1697.

Moilanen, A., Franco, A.M., Early, R.I., Fox, R., Wintle, B. & Thomas, C.D. (2005) Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 1885-1891.

Moilanen, A., Pouzols, F.M., Meller, L., Veach, V., Arponen, A., Leppänen, J. & Kujala, H. (2014) ZONATION: spatial conservation planning framework and software. Version 4. User Manual. In. Atte Moilanen / Metapopulation Research Group, University of Helsinki, Finland.

Mónica, S. & Santos, F. (2011) Trends in extreme daily precipitation indices in Northern of Portugal. *Geophysical Research Abstracts*, **13**

Naidoo, R., Balmford, A., Ferraro, P.J., Polasky, S., Ricketts, T.H. & Rouget, M. (2006) Integrating economic costs into conservation planning. *Trends in ecology & evolution*, **21**, 681-687.

Newburn, D., Reed, S., Berck, P. & Merenlender, A. (2005) Economics and land-use change in prioritizing private land conservation. *Conservation Biology*, **19**, 1411-1420.

Pearson, R.G., Dawson, T.P. & Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285-298.

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C. & Guisan, A. (2012) Climatic niche shifts are rare among terrestrial plant invaders. *Science*, **335**, 1344-8.

Petitpierre, B., McDougall, K., Seipel, T., Broennimann, O., Guisan, A. & Kueffer, C. (2015) Will climate change increase the risk of plant invasions into mountains? *Ecological Applications*,

Phalan, B., Onial, M., Balmford, A. & Green, R.E. (2011) Reconciling food production and biodiversity conservation: land sharing and land sparing compared. *Science*, **333**, 1289-1291.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161-175.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231-259.

Pouzols, F.M., Toivonen, T., Di Minin, E., Kukkala, A.S., Kullberg, P., Kuusterä, J., Lehtomäki, J., Tenkanen, H., Verburg, P.H. & Moilanen, A. (2014) Global protected area expansion is compromised by projected land-use and parochialism. *Nature*, **516**, 383.

Pressey, R.L., Cabeza, M., Watts, M.E., Cowling, R.M. & Wilson, K.A. (2007) Conservation planning in a changing world. *Trends Ecol Evol*, **22**, 583-92.

Price, B., Kienast, F., Seidl, I., Ginzler, C., Verburg, P.H. & Bolliger, J. (2015) Future landscapes of Switzerland: Risk areas for urbanisation and land abandonment. *Applied Geography*, **57**, 32-41.

Proches, S., Wilson, J.R.U., Veldtman, R., Kalwij, J.M., Richardson, D.M. & Chown, S.L. (2005) Landscape Corridors: Possible Dangers? *Science*, **310**, 779-783.

Pyšek, P., Jarošík, V., Pergl, J., Moravcová, L., Chytrý, M. & Kühn, I. (2013) Temperate trees and shrubs as global invaders: the relationship between invasiveness and native distribution depends on biological traits. *Biological Invasions*, **16**, 577-589.

Pyšek, P., Jarošík, V. & Kučera, T. (2002) Patterns of invasion in temperate nature reserves. *Biological Conservation*, **104**, 13-24.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689-1704.

Rayfield, B., Moilanen, A. & Fortin, M.-J. (2009) Incorporating consumer–resource spatial interactions in reserve design. *Ecological Modelling*, **220**, 725-733.

Rejmánek, M., Richardson, D.M. & Pyšek, P. (2013) Plant invasions and invasibility of plant communities. *Vegetation Ecology, Second Edition*, 387-424.

Richardson, D.M. & Rejmánek, M. (2011) Trees and shrubs as invasive alien species - a global review. *Diversity and Distributions*, **17**, 788-809.

Richardson, D.M., Pysek, P., Rejmánek, M., Barbour, M.G., Panetta, F.D. & West, C.J. (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions*, **6**, 93–107.

Rivas-Martínez, S., Penas, A. & Díaz, T.E. (2004) Biogeographic Map of Europe.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011) Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography*, **35**, 211-226.

Säumel, I. & Kowarik, I. (2010) Urban rivers as dispersal corridors for primarily wind-dispersed invasive tree species. *Landscape and Urban Planning*, **94**, 244-249.

Saunders, D.A. & Hobbs, R.J. (1991) The role of corridors in conservation: what do we know and where do we go?

Scherrer, D., Massy, S., Meier, S., Vittoz, P., Guisan, A. & Serra-Diaz, J. (2017) Assessing and predicting shifts in mountain forest composition across 25 years of climate change. *Diversity and Distributions*, **23**, 517-528.

Scherrer, D., Mod, H.K., Pottier, J., Litsios-Dubuis, A., Pellissier, L., Vittoz, P., Götzenberger, L., Zobel, M. & Guisan, A. (2018) Disentangling the processes driving plant assemblages in mountain grasslands across spatial scales and environmental gradients. *Journal of Ecology*, **0**

SFSO (2013) Swiss Federal Statistical Office. Land use in Switzerland: Results of the Swiss land use statistics. Neuchatel.

Simberloff, D. (1988) The contribution of population and community biology to conservation science. *Annual review of ecology and systematics*, **19**, 473-511.

Singh, N.J. & Milner-Gulland, E.J. (2011) Conserving a moving target: planning protection for a migratory species as its distribution changes. *Journal of Applied Ecology*, **48**, 35-46.

Souza-Alonso, P., Rodríguez, J., González, L. & Lorenzo, P. (2017) Here to stay. Recent advances and perspectives about Acacia invasion in Mediterranean areas. *Annals of Forest Science*, **74**

Steffen, W., Crutzen, P.J. & McNeill, J.R. (2007) The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature. *AMBIO: A Journal of the Human Environment*, **36**, 614-621.

Struebig, M.J., Wilting, A., Gaveau, D.L., Meijaard, E., Smith, R.J., Abdullah, T., Abram, N., Alfred, R., Ancrenaz, M. & Augeri, D.M. (2015) Targeted conservation to safeguard a biodiversity hotspot from climate and land-cover change. *Current Biology*, **25**, 372-378.

Suárez-Seoane, S., Virgós, E., Terroba, O., Pardavila, X. & Barea-Azcón, J.M. (2014) Scaling of species distribution models across spatial resolutions and extents along a biogeographic gradient. The case of the Iberian moleTalpa occidentalis. *Ecography*, **37**, 279-292.

Summers, D.M., Bryan, B.A., Crossman, N.D. & Meyer, W.S. (2012) Species vulnerability to climate change: impacts on spatial conservation priorities and species representation. *Global Change Biology*, **18**, 2335-2348.

Swets, J.A. (1988) Measuring the Accuracy of Diagnostic Systems. *Science*, **240**, 1285-1293.

Theoharides, K.A. & Dukes, J.S. (2007) Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. *New Phytol*, **176**, 256-73.

Thuiller, W., Richardson, D.M. & Midgley, G.F. (2007) Will Climate Change Promote Alien Plant Invasions? *Biological Invasions*, pp. 197-211. Springer.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369-373.

Tulloch, A.I.T., Sutcliffe, P., Naujokaitis-Lewis, I., Tingley, R., Brotons, L., Ferraz, K.M.P.M.B., Possingham, H., Guisan, A. & Rhodes, J.R. (2016) Conservation planners tend to ignore improved accuracy of modelled species distributions to focus on multiple threats and ecological processes. *Biological Conservation*, **199**, 157-171.

Van Buskirk, J. & Willi, Y. (2004) Enhancement of farmland biodiversity within set-aside land. *Conservation Biology*, **18**, 987-994.

Verburg, P.H. & Overmars, K.P. (2009) Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape Ecol.*, **24**, 1167-1181.

Verhagen, W., van Teeffelen, A.J. & Verburg, P.H. (2018) Shifting spatial priorities for ecosystem services in Europe following land use change. *Ecological Indicators*, **89**, 397-410.

Vicente, J., Alves, P., Randin, C., Guisan, A. & Honrado, J. (2010) What drives invasibility? A multi-model inference test and spatial modelling of alien plant species richness patterns in northern Portugal. *Ecography*, **33**, 1081-1092.

Vicente, J., Randin, C.F., Gonçalves, J., Metzger, M.J., Lomba, Â., Honrado, J. & Guisan, A. (2011) Where will conflicts between alien and rare species occur after climate and land-use change? A test with a novel combined modelling approach. *Biological Invasions*, **13**, 1209-1227.

Vicente, J.R., Fernandes, R.F., Randin, C.F., Broennimann, O., Goncalves, J., Marcos, B., Pocas, I., Alves, P., Guisan, A. & Honrado, J.P. (2013) Will climate change drive alien invasive plants into areas of high protection value? An improved model-based regional assessment to prioritise the management of invasions. *J Environ Manage*, **131**, 185-95.

Vicente, J.R., Alagador, D., Guerra, C., Alonso, J.M., Kueffer, C., Vaz, A.S., Fernandes, R.F., Cabral, J.A., Araujo, M.B. & Honrado, J.P. (2016) Cost-effective monitoring of biological invasions under global change: a model-based framework. *Journal of Applied Ecology*, **53**, 1317-1329.

Vila, M., Espinar, J.L., Hejda, M., Hulme, P.E., Jarosik, V., Maron, J.L., Pergl, J., Schaffner, U., Sun,

Y. & Pysek, P. (2011) Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecol Lett*, **14**, 702-8.

Vilà, M. & Ibáñez, I. (2011) Plant invasions in the landscape. *Landscape Ecology*, **26**, 461-472.

Vilà, M., Basnou, C., Pyšek, P., Josefsson, M., Genovesi, P., Gollasch, S., Nentwig, W., Olenin, S., Roques, A., Roy, D. & Hulme, P.E. (2010) How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. *Frontiers in Ecology and the Environment*, **8**, 135-144.

Walther, G.R.P., E.; Convey, P.; Menzel, A.; Parmesan, C.; Beebee, T.J.C.; Fromentin, J.M.; Hoegh-Guldberg, O. & Bairlein, F. (2002) Ecological responses to recent climate change. *Nature*, **416**, 389-395.

Wan, J., Wang, C., Yu, J., Nie, S., Han, S., Liu, J., Zu, Y. & Wang, Q. (2016) Developing conservation strategies for Pinus koraiensis and Eleutherococcus senticosus by using model-based geographic distributions. *Journal of forestry research*, **27**, 389-400.

Welch, D. (2005) What should protected areas managers do in the face of climate change? *The George Wright Forum* (ed by, pp. 75-93.

Wilcox, B.A. & Murphy, D.D. (1985) Conservation strategy: the effects of fragmentation on extinction. *The American Naturalist*, **125**, 879-887.

Wilson, K., Pressey, R.L., Newton, A., Burgman, M., Possingham, H. & Weston, C. (2005) Measuring and incorporating vulnerability into conservation planning. *Environmental Management*, **35**, 527-543.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & Distribut, N.P.S. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763-773.

Zhang, K., Laffan, S.W., Ramp, D. & Webster, E. (2011) Incorporating a distance cost in systematic reserve design. *International Journal of Geographical Information Science*, **25**, 393-404.

Zwiener, V.P., Padial, A.A., Marques, M.C., Faleiro, F.V., Loyola, R. & Peterson, A.T. (2017) Planning for conservation and restoration under climate and land use change in the Brazilian Atlantic Forest. *Diversity and Distributions*, **23**, 955-966.

SYNTHESIS & DISCUSSION

SYNTHESIS & DISCUSSION

It has become well recognised that species distribution models (SDMs) are a very important tool in different studies of biogeography, evolution, macroecology or conservation science. They have been used to study a vast amount of different subjects ranging from simply predicting the current and future distributions of species under climate change (see e.g. Franklin, 2010; Guisan *et al.*, 2017 for review), to the support of effective conservation plans (e.g. Vicente *et al.*, 2016) or the prediction of species assemblages through stacked SDMs (S-SDMs; e.g. Dubuis *et al.*, 2011; D'Amen *et al.*, 2015b).

However, through my thesis, I also established that the accuracy of SDMs is often affected by different factors (e.g. incorrect detection, modelling technique or truncated datasets) that can introduce bias into their predictions. It, therefore, became important to assess the reasons, effects and solutions for those biases, something that is difficult to achieve when dealing with real-world data where a large number of factors cannot be controlled (e.g. biotic interactions, environmental constraints or the full knowledge of species presence/absence).

The solution found to overcome these limitations in this thesis was to use artificial data in the form of simulations of virtual species (e.g. Hirzel *et al.*, 2001; Austin *et al.*, 2006; Zurell *et al.*, 2010), with all the information necessary for a study fully available in a controlled environment. My thesis used these virtual species to assess different factors that were thought to be affecting the accuracy of individual SDMs and their assemblage (S-SDM).

The implications of the work developed in this thesis are important because they further enhance the scrutiny of methodologies involved in species distribution modelling, giving answers to some questions and raising new important ones. The main objectives of my thesis were 1) to determine the degree to which different sources of uncertainty or factors used in SDMs affect the predictive accuracy of models of individual species and their assemblages (chapters 1.1 and 1.2), 2) determine the degree to which truncated datasets affect the accuracy of assemblage predictions (chapter 1.3), and 3) test the validity of strategies used to integrate information from large scales into regional/local scales (e.g. hierarchical models; chapter 1.4).

In the following sections, I discuss the developments made in this PhD work but also the limitations of my results and frameworks, also elaborating on potential alternatives and future developments that I believe are important to improve in order to have SDMs and S-SDMs giving more robust and accurate predictions.

Synthesis, main findings and their significance

The main chapters of this PhD thesis presented original and, in my view, important work on the use of virtual data to assess the effects that different sources of uncertainty and bias have on species distributions models, particularly when predicting species assemblages. The two additional chapters in which I was actively involved that fall inside the scope of this thesis will also be briefly discussed here (mainly Chapter 2.1) since they help to understand the motivations for the work developed in this thesis.

Chapter 2.1 gives a useful framing to the current status of using species distributions models to predict species assemblages, testing the SESAM framework in mountain grassland communities of the western Swiss Alps. It shows that the tested “probability ranking” rule allowed for improvements in the prediction of community composition. With the implementation of this rule, which consisted on building community composition by selecting species in decreasing order of their predicted probability of presence until the maximum predicted richness for a certain plot was reached, improving the prediction of species richness, but also of species composition. This approach was able to give more reliable results than by simply stacking binary S-SDM, something also suggested or shown in other studies (e.g. Dubuis *et al.*, 2011; Calabrese *et al.*, 2014; D'Amen *et al.*, 2015a). Additionally, using the “probability ranking“ rule, the predictions of community composition were improved, mainly because it was recently shown to be a thresholding method optimized for communities (chapter 2.2; Scherrer *et al.*, 2018), but also because the species that are most likely to be overpredicted in bS-SDMs are also the ones less likely to be present and therefore removed by the rule. However, despite the improvement in reducing overprediction of species richness, the improved assemblage composition predictions were still different from those observed. While this tendency to overpredict species richness was also observed in other studies (e.g. Aranda & Lobo, 2011; Dubuis *et al.*, 2013; Pottier *et al.*, 2013; Calabrese *et al.*, 2014), it appears to have exceptions. D'Amen *et al.* (2015a), used a similar approach to the one developed in this study and applied it to the analysis of the assemblage prediction of two insect groups but found that overprediction of richness by bS-SDMs is not a general rule, being circumvented if good individual species models are generated and/or when the total species pool remains small. In another study, Distler *et al.* (2015) used S-SDMs to assess species richness patterns of North American birds, showing that stacked distribution models were able to predict accurate species richness patterns at broad scales.

On the other hand, Calabrese *et al.* (2014) suggested that the overprediction of species richness observed in S-SDM could be explained by the use of thresholding methods that transform continuous probability predictions of species distributions into binary ones. In that study, the authors also developed a maximum-likelihood approach to adjust S-SDM. Identifying the optimal thresholding method and ways to correctly evaluate assemblage predictions would then be a logic next step, something that was developed in **Chapter 2.2** (Scherrer *et al.*, 2018). Furthermore, due to the presence of stochastic effects, assemblage composition will always include some level of prediction errors, making it useful to understand and determine how different sources of error affect single species distribution models and how those errors can propagate from SDMs to S-SDMs. It was then reasonable to think that if one were able to improve these sources of uncertainty present in SDMs, then the assemblage predictions provided not only from S-SDMs but also by the SESAM framework would be more accurate (i.e. not overpredicting species richness). This was the basic reasoning behind the development of the main chapters of this thesis.

Chapter 1.1 illustrates the application of virtual species in evaluating the degree to which certain errors in species data, namely false presences and false absences, affect model predictions and how this is reflected in commonly used evaluation metrics. It shows that the interpretation of model's performance strongly depended on the data and metric used to evaluate those models, with different conclusions being made depending on whether model fit or predictive success was measured. It also showed that added errors were less important with large sample sizes and that model performance was more affected by the occurrence of false positives in the species data. It also provided evidence that the interpretation of widely used evaluation metrics needs to be reconsidered since high (over-optimistic) evaluation values were obtained even when high levels of error were added. This is an important finding and can be problematic since SDMs are widely evaluated by the metrics tested in this study (Somers'D/AUC, TSS and Kappa). The debate about which metric to use and if they are good enough to evaluate models is not new, with Lobo *et al.* (2008) and Peterson *et al.* (2008) criticizing AUC as a misleading measure and identifying drawbacks of their use like weighting omission and commission errors equally or the fact that if does not provide information about the performance of a model. Allouche *et al.* (2006), argued for the use of TSS instead of Kappa since the latter's prevalence dependency introduces bias on the estimates of accuracy. McPherson *et al.* (2004) also studied Kappa statistics, concluding that it is inappropriate for comparisons of model accuracy and recommending the use of AUC instead as a more reliable metric of model performance. Our results showed a strong tendency of Somers'D/AUC (and to a certain point TSS) to produce overoptimistic evaluations whereas Kappa better reflected the actual levels of error in the training data. Our findings were supported in two very recent studies, also suggesting that AUC and TSS are not the most efficient metrics to assess model performance (i.e. over-optimistic or unrealistic metrics). In the first study, Mitchell *et al.* (2017) evaluated the effects of locational uncertainty across several sample sizes for kelp, noting that it is important not only to consider model performance (i.e. using AUC and TSS) but also the spatial predictions when comparing models. In a second study, Fourcade *et al.* (2018) used pseudo-predictors (i.e. using digitized and georeferenced classical paintings instead of real environmental data; see also Elith, 2002) to test model performance based on AUC and TSS, comparing those results with values obtained when real predictors were used. Their findings also suggest that the used evaluations metrics were unable to correctly evaluate the significance of SDMs since most models calibrated with pseudo-predictors were classified as having good performances (sometimes even better evaluations than when using real data). Our findings together with these two studies highlight the need to reconsider the use of widely used evaluation metrics, since wrongly evaluated models can be considered as valid, which can have serious consequences if then used in conservation or management actions, for example. Another novelty in this study was the fact that we tested and compared how conclusions about model performance are inferred if model fit (i.e. how well models reproduced the data) or predictive success (i.e. how well models predict the species' true distribution) is calculated. We showed that depending on which one is considered, different conclusions can be made since predictions with good model fit usually presented bad predictive success, and vice-versa, a pattern also reflected in the modelling technique performance (see also Randin *et al.*, 2006).

Chapter 1.2 illustrates a virtual ecologist application for the assessment of how different methodological factors can affect the prediction accuracy of S-SDM assemblages. It demonstrates that the accuracy of S-SDMs is mostly affected by modelling technique followed by sample size (i.e. from the factors tested in the study). We confirmed that model accuracy increased with sample size, something that was already shown in previous studies focusing on SDMs (e.g. Stockwell & Peterson, 2002; Wisz *et al.*, 2008; Mitchell *et al.*, 2017), but demonstrated that even when using large amounts of training data one still obtained some inaccuracy in assemblage predictions. Most of that inaccuracy depended on the modelling and sampling technique used. Because the sampling strategy that reflected the species true prevalence was the most successful, we recommended that future studies using real species data would implement what we called the "plot-like" sampling method (i.e. sampling all species in the same plots; best approximation to species' true prevalence). This study was not the first to test the effects of different factors affecting SDMs, taking inspiration on a methodology proposed by Thibaud *et al.* (2014), but it was the first to use a virtual ecologist approach to assess the cumulative effects of different factors affecting species assemblage predictions (S-SDMs). Contrary to the previously mentioned study, we found that the choice of the modelling technique used had a larger impact than sample size on the accuracy of assemblage predictions. We also found that techniques like BRT and RF presented higher variance and lower predictive success than GLM and GAM, something that was contrary to other studies focusing on single species SDMs (e.g. Elith *et al.*, 2006; Guisan *et al.*, 2007; Graham *et al.*, 2008; Williams *et al.*, 2009), but that indicate that BRT/RF are good at finding a signal in the training data while being less good at predicting to independent data. However, this pattern could also be an artefact from the fact that our virtual species were created by GLMs, and this methodological step is prone to some criticism (threshold vs probabilistic approach; see Meynard & Kaplan, 2013) with further tests being needed, perhaps recurring to specialized packages to create virtual species (e.g. SDMvspecies, NicheLim or virtualspecies - Duan *et al.*, 2015; Huang *et al.*, 2016; Leroy *et al.*, 2016).

An important analysis in this chapter was the assessment of what overall accuracy could be expected when known species data was used to predict assemblages. This is important because if we can determine the level of uncertainty that can be expected when “perfect” data is used, we can evaluate real data predictions accordingly. What we found is that even with the full knowledge of the species and sampling large amounts of data, “perfect” predictions of assemblage are difficult to attain, but very good predictions are reachable. If this is the case when using virtual species, one can expect an equivalent or even higher inaccuracy when using real species. This means that even if perfect similarity scores are not reached when using real species, values close to that maximum value can already be considered as satisfactory. Using artificial data in a virtual ecologist approach to complement analyses on real data can thus help to assess the best predictions one can get in the real context. In any case, this study was a first attempt to assess the most important factors affecting assemblage predictions, and although future more complex nested approaches could include a higher number of factors, good assemblage prediction could already be obtained if a representative sample of species distribution is obtained (using a “plot-like” sampling method).

Chapter 1.3 illustrates an attempt to analyse the influence of the level of truncation in the prediction of future species assemblages. This study analyses how accurate assemblages comprised of 20 virtual species are predicted under current and future conditions when using truncated training datasets. We showed that truncated training datasets have a negative effect on the accuracy of species distributions and their assemblages, especially when modelling wide range species. Predictions made under future conditions failed to produce accurate assemblages both using non-truncated or highly truncated datasets. This was the first study to analyse the effects of truncated datasets on the accuracy of species assemblages (S-SDMs), but the study of truncations in by no means a new field. For example, Thuiller *et al.* (2004) examined the consequences of using truncated environmental conditions to project the future distributions of an endemic European tree, finding that the restriction of the environmental range of data had a strong influence on the estimation of response curves, leading models to capture only part of the full species’ environmental range and fitting wrong response curves. This means that the applicability of models to be used for predictions is reduced and as our results also suggest, projections of future distributions can only be accurate if the entire bioclimatic envelope of the species is considered. In another study, Barbet-Massin *et al.* (2010) compared projections of the distribution of bird species when using truncated occurrence datasets. The authors found that the use of this truncated datasets led to an underestimation of species richness (something also observed in chapter 1.3) and that misleading conclusions can be made when models are fitted with truncated data.

Another example is the work of Titeux *et al.* (2017) where species distribution models were tested using geographically restricted data (truncated occurrences) of butterfly species, with models also being projected under future climate change scenarios. The authors reached results like those observed in this chapter, with truncated models failing to predict accurate species richness under new climatic conditions. Their recommendation to use distribution data beyond the boundaries of the study area in order to capture the full species response curves is complementary to our idea that species assemblages should be modelled with wide range species sampled using the largest extents possible to avoid truncation. A useful strategy to be further tested to predict accurate species assemblages could be to join the sampling of wide range species in larger extents with the modelling of narrow range species present in a smaller study area (since our results suggest they might be less susceptible to truncated datasets, but see possible limitations of our approach there). Finally, we also showed the importance of correctly selecting predictors that reflect the species environmental requirements. This is because if not used, as we proved using virtual species, the effects of truncation are even more pronounced, and the predicted assemblages have little similarity with observed ones. Therefore, the effects of truncation need to be further tested with new methods, and until then, the predictions obtained by SDMs should be carefully considered.

Finally, in **Chapter 1.4** different modelling strategies were tested and compared to assess how accurate they were at predicting species assemblages under current and future climatic conditions (i.e. assessing their transferability). Particular attention was given in determining the validity of two hierarchical modelling frameworks to improve model transferability. For that, models sampled with presences/absences or presences/pseudo-absences of 100 virtual species were used, showing that hierarchical models were not able in our setting to improve assemblage predictions when compared with the other tested strategies. When using presence/absence data, all the tested strategies were able to predict better species assemblages (i.e. more similar with observed ones) than models using presence/pseudo-absence data, while all strategies failed to correctly predict future species assemblages. These results are significant because it warns us against the use of SDMs to make projections in space and time and that the transferability capabilities of these models need to be further tested and improved. Hierarchical modelling approaches have been previously tested and applied with real species data. Pearson *et al.* (2004) used a framework that integrated land-cover data into a bioclimatic model in a hierarchical scale-dependent way, with the model being first fitted using European bioclimatic data and then incorporating land-cover data the local (Britain) scale. The authors found good predictive performance for each species and that the multi-scale approach had potential to help understand environmental limitations to species' distributions. In another study focused on invasive species, Gallien *et al.* (2012) provided a new methodological framework to improve the regional modelling of invasive species, using the output of a global model to weight pseudo-absences in a regional model. They found that the predictive performance of the regional SDMs was significantly improved, something also found by Petitpierre *et al.* (2016) when using a similar approach but differed in the weighting method. However, contrary to the presented studies, we were unable to achieve accurate assemblage predictions when using hierarchical models (under current and future conditions). We need to also consider that if virtual species were not available, we would have reached similar conclusions to the ones presented before (Pearson *et al.*, 2004; Gallien *et al.*, 2012; Petitpierre *et al.*, 2016) if only accounting for the evaluation values obtained by our models (high evaluation values; MaxTSS>0.7). This once again is an argument in favour of caution when interpreting models' evaluation, since wrong conclusions can be made depending on the metric or data used, with assemblages being incorrectly predicted (as in our case). We recommended that additional work should include the assessment and improvement of methods focusing on Bayesian frameworks (e.g. Keil *et al.*, 2013) or multi-scale approaches (e.g. Bastos *et al.*, 2016; Talluto *et al.*, 2016), and that hierarchical models could be improved if different weights are given to presences rather than to pseudo-absences, or complementing the fine-resolution presences in the local model with coarse resolution ones from the global model.

Can simulations improve species distribution models?

A recent study analysed the number of articles related to SDMs from the period of 2000-2009 and found 2118 articles with a total of 37.854 citations (Brotons, 2014). Consequently, there is no doubt that species distribution models are a widely used tool that has been intensively used to study the potential distribution of species under climate change, assess invasion risk or in conservation planning and prioritization (e.g. Guisan & Zimmermann, 2000; Broennimann & Guisan, 2008; Alagador *et al.*, 2014; Vicente *et al.*, 2016). However, SDMs are also known to have limitations and uncertainties associated with them (e.g. Araújo & Guisan, 2006; Barry & Elith, 2006; Dormann *et al.*, 2008).

A way to avoid these limitations is to use artificial data (Austin *et al.*, 2006) and implement a virtual ecologist approach (Zurell *et al.*, 2010) to test different methods or ecological processes that are impossible or very difficult to verify or compare in reality (e.g. limited in terms of available data, time period or spatial extent). The use of virtual data removes this real data limitation, allowing for the complete control of the data and models being tested in an artificial world where all the necessary information is available (e.g. Zurell *et al.*, 2010). It is, therefore, reasonable to think that artificial data could be used more systematically to study different processes in ecology and species distribution modelling. The first use of this approach is probably the one by Swan (1970), using artificial data to evaluate the performance of the Bray-Curtis ordination technique. Artificial data was greatly used in studies related to ordination methods (e.g. Kenkel & Orloci, 1986; Dray *et al.*, 2003; Whittaker, 2012). Austin *et al.* (2006) noted that the use of artificial data requires knowledge about vegetation theory, data measurement errors, ecological process models, evaluation methods and sampling designs to be successful at comparing different statistical methods. In perhaps the first application with SDMs, Hirzel *et al.* (2001), who actually coined the term "virtual species", created those kind of species to test different habitat suitability methods (i.e. ENFA and GLM) and assess their predictive power under different historic scenarios (i.e. spreading, equilibrium and overabundant species). In a following study, Hirzel and Guisan (2002) used the same virtual species to compare different sampling strategies, determining that the regular and equal-stratified sampling strategies were the most robust and accurate. Based on those results the authors were also able to provide guidelines to improve sample design. Virtual species have also been used to test different statistical models (Meynard & Quinn, 2007), comparing GAMs, GLMs, classification trees and GARP with virtual species that varied in their response to environmental gradients. The authors were able to recommend the use of GAM or GLM over the other techniques. Railsback *et al.* (2003) tested different ecological models and assumptions that were very difficult to test in reality and evaluated factors like habitat quality and selection using a virtual trout population, identifying reasons why animal density may not reflect habitat quality. Another useful example of virtual species use was the study developed by Thibaud *et al.* (2014), where different factors affecting the prediction of single species SDMs were measured and compared in term of their relative effects, and illustrating that spatial autocorrelation was not an important factor in the tested landscape, unlike reported in other studies, with most of the variation in prediction accuracy being due to sample size and modelling technique.

The usefulness of virtual species was also demonstrated in this thesis, where I showed (in chapter 1.1) that when errors in species presence/absence exist, model performance is most affected by false positives and that the interpretation of widely used metrics to evaluate SDMs can be misleading towards erroneous conclusions (i.e. a model being considered as good when in fact its predictions are inaccurate). Additionally, by using virtual species to test the effects of different factors affecting the accuracy of S-SDMs (chapter 1.2) I was able to determine that modelling technique and sample size were the factors mostly affecting that accuracy. Based on the results from that work I recommended that the "plot-like" sampling method is more efficient that simply increasing the number of sampled data. Other work developed in this thesis allowed for a better understanding of the effect of truncated datasets on the accuracy of future assemblage predictions (chapter 1.3) or identify challenges in models that are projected into new environmental conditions (i.e. transferability) and the validity of hierarchical models (chapter 1.4).

I therefore believe that virtual or artificial data and simulations are useful tools to help in the development and improvement of species distribution modelling, contributing to a better understanding of the behaviour of the methods and metrics used, and allowing to develop more accurate and reliable predictions of future species distributions, risk assessments (i.e. biological invasions) and conservation or monitoring actions. Recent efforts were also made to facilitate the use of virtual species, with R packages being already available (e.g. SDMvspecies, NicheLim or virtualspecies - Duan *et al.*, 2015; Huang *et al.*, 2016; Leroy *et al.*, 2016). Additionally, other R packages like *ecolottery* (Munoz *et al.*, 2018) or *VirtualCom* (Münkemüller & Gallien, 2015) allow for the simulation of communities with environmental filtering and neutral dynamics (in the first example) or the simulation of trait evolution and community assembly processes (in the second example). Other recent examples of tools being created to simulate metapopulation dynamics and range expansion (Mestre *et al.*, 2016), range shifts in response to climate (Midgley *et al.*, 2010) or to test the efficacy of functional diversity indices (McPherson *et al.*, 2018), lead me to believe that the use of artificial data is a step in the right direction if we want to obtain more useful and more reliable species distribution models, and they should be used in parallel with the proposition of new methods. We can then say that if a method fails in the simplified simulated world that is being used to test it, the chance that it fails in the real world is high, and therefore is not worth to further develop it. However, it is important to note that the opposite does not occur, because of the added complexity that exists in the real world.

Moving forward with species assemblage modelling - perspectives

In the studies developed in this thesis, I confirmed that even when one has complete knowledge of the factors determining species distribution and when the sampling of species occurrence was done without bias, I was unable to obtain “perfect” predictions of species assemblages. This is important because it confirms the major influence of the sampling process, and particularly of sample size, and potentially other confounding factors that one cannot control when modelling species distributions. It is then reasonable to think that if one is unable to obtain perfect predictions with virtual species unless a very large number of sites (unrealistic with real field sampling) is sampled, then this procedure is even more complicated when real species are considered. Biotic interactions, for example, can play a substantial role in determining the responses of real species and communities in changing environments (Araújo & Luoto, 2007; Kissling *et al.*, 2012). However, biotic interactions and how it might determine species co-occurrence is still rarely taken into account within species distribution models (Araújo & Luoto, 2007; Wisz *et al.*, 2013), which might explain the failure of some models to predict the correct distribution of species assemblages (Kissling *et al.*, 2012). Virtual simulations could also focus on trying to include different types of biotic interactions when predicting species assemblages (e.g. mutualism, commensalism, competition). Of course, this will involve large amounts of work and time because as Dodds and Nelson (2006) showed, as the number of species that can interact increase, so does the number of possible direct and indirect interactions. SDMs are used and been around for some time, but there is still much we do not know about how to correctly predict species distribution and assemblages. I believe that future species or community modelling frameworks would gain to include virtual species simulations in their preliminary assessments. Only this way would one be able correctly determining the amount of certainty a framework contains, based on what it can predict in a virtual world, then transferred into a more complex reality where informed decisions must be made.

In my thesis, I have shown different research questions where virtual species could be successfully used and showing that it can be a tool worth developing. I showed in chapters 1.3 and 1.4 that virtual species could be used to test the accuracy of SDMs and S-SDMs when projecting distributions under scenarios of climate change. It would be interesting to test also the effects of several other factors that might be important and influence current and future predictions. Testing the effects of different biotic interactions (like previously discussed), the simulation of different invasions in a native assemblage to assess community dynamics or the influence that land-use change and habitat fragmentation have on future species distributions is also something to investigate further. Associated with S-SDM/SESAM, virtual species can also be useful to test and refine different ecological assembly rules (Gotzenberger *et al.*, 2012), used to restrict the observed patterns of species assemblages (Guisan & Rahbek, 2011). Another important aspect that should gain more attention is the collaboration between different research fields. As noted by Zurell *et al.* (2010), empiricists and theoreticians rarely work together, and their collaboration could be beneficial since it would help applied scientists and practitioners to better plan their work and modellers to increase the value of the framework they develop.

Furthermore, since I showed in my studies that two key model evaluation metrics used (AUC and TSS) provide overoptimistic values, another important development would be to propose and test more effective metrics for model evaluation, both for single species and communities, and how to threshold species distributions into binary data (e.g. Liu *et al.*, 2016). Tighter coupling of the study focusing on transferability (chapter 1.4) and the one focusing on truncation (chapter 1.3) could also be done to further assess the effects of truncation on the transferability of model predictions. Improvements in the hierarchical modelling framework also need to be tested, with the idea of using the global model to directly correct the signal (i.e. response curve) driven by presences and not weighting pseudo-absences, as currently done. Finally, the way of correctly selecting environmental variables in a model (variable selection) could also be further analyzed with virtual species. This means that not only the effects of different variables in the prediction of species should be assessed, but also new methods efficiently select those that are more suitable. As I showed in chapter 1.3, the choice of variables that reflect the species environmental requirements is important, with inaccurate predictions potentially occurring otherwise.

While artificial or virtual data can be used in a variety of applications, it is always important to remember that when working with virtual data one necessary works with a simplification of reality. This means that the complexities of real-world data like landscape structure, biotic interactions, organism behaviour or other biases are usually not considered in a simulated analysis. Simplifications in the case of virtual ecologist approaches are often made in purpose, to avoid confounding factors that might hide the effects of a factor we are trying to assess. Some studies (Meynard & Kaplan, 2013; Miller, 2014) have also criticized certain methods of creating virtual species (like the one employed in chapter 1.1 and 1.2), suggesting that a better alternative might be the probabilistic approach of converting suitability into binary data (Meynard & Kaplan, 2013; applied in chapter 1.4). However, while this might be considered as the most appropriate approach for the creation of virtual species, other methods to create virtual species can also work depending on the study purposes and objectives. For example, in chapters 1.1 and 1.2, I wanted to avoid adding stochasticity as much as possible from the entire process. This means that the appropriateness of a given method depends on the study objective, and therefore the correct method should be explained and defined before the study starts.

Other limitations can be identified in the work developed in this thesis. For example, the effects of truncation in Chapter 1.4 were analyzed using a truncated area (Europe) that might not be sufficiently truncated (i.e. still covering quite large environmental gradients), and therefore other truncated extents could have been tested. Additionally, strategies to sample pseudo-absences that are better suited to be used with classification trees were also used with regression techniques, and vice-versa. This can potentially lead to bias predictions and should be further investigated. Another potential limitation of my work concerns another aspect of the selection of the truncated areas in chapter 1.3. I opted for the selection of individual truncated areas for each species, which could also have potentially biased the results, if it prevented species' response curves to be too severely truncated, e.g. truncating only the tails of unimodal curves rather than their mode. It would thus be important to test the same framework but using the same truncated area (or several truncated areas) for all species. Supplementary analyses to assess changes in the response curves of the species between the truncated and non-truncated models could also be conducted to fully understand the effects of truncation.

Nevertheless, despite the limitations discussed here and, in each chapter, virtual species still appear to be a useful tool to evaluate methods and assess the effects of factors influencing the predictive accuracy of species and community distribution models, and their more systematic use should be encouraged.

References

Alagador, D., Cerdeira, J.O., Araújo, M.B. & Saura, S. (2014) Shifting protected areas: scheduling spatial priorities under climate change. *Journal of Applied Ecology*, **51**, 703-713.

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

Aranda, S.C. & Lobo, J.M. (2011) How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. *Ecography*, **34**, 31-38.

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688.

Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743-753.

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. & Luoto, M. (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, **199**, 197-216.

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? *Ecography*, **33**, 878-886.

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

Bastos, R., D'Amen, M., Vicente, J., Santos, M., Yu, H.R., Eitelberg, D., Goncalves, J., Civantos, E., Honrado, J. & Cabral, J.A. (2016) A multi-scale looping approach to predict spatially dynamic patterns of functional species richness in changing landscapes. *Ecological Indicators*, **64**, 92-104.

Broennimann, O. & Guisan, A. (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biol Lett*, **4**, 585-9.

Brotons, L. (2014) Species distribution models and impact factor growth in environmental journals: methodological fashion or the attraction of global change science. *PLoS One*, **9**, e111996.

Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014) Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99-112.

D'Amen, M., Pradervand, J.N. & Guisan, A. (2015a) Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography*, **24**, 1443-1453.

D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L. & Guisan, A. (2015b) Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *Journal of Biogeography*, **42**, 1255-1266.

Distler, T., Schuetz, J.G., Velasquez-Tibata, J. & Langham, G.M. (2015) Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *Journal of Biogeography*, **42**, 976-988.

Dodds, W.K. & Nelson, J.A. (2006) Redefining the community: a species-based approach. *Oikos*, **112**, 464-472.

Dormann, C.F., Purschke, O., Garcia Marquez, J.R., Lautenbach, S. & Schroder, B. (2008) Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. *Ecology*, **89**, 3371-86.

Dray, S., Chessel, D. & Thioulouse, J. (2003) Co-inertia analysis and the linking of ecological data tables. *Ecology*, **84**, 3078-3089.

Duan, R.Y., Kong, X.Q., Huang, M.Y., Wu, G.L. & Wang, Z.G. (2015) SDMvspecies: a software for creating virtual species for species distribution modelling. *Ecography*, **38**, 108-110.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122-1131.

Dubuis, A., Rossier, L., Pottier, J., Pellissier, L., Vittoz, P. & Guisan, A. (2013) Predicting current and future spatial community patterns of plant functional traits. *Ecography*, **36**, 1158-1168.

Elith, J. (2002) *Predicting the distribution of plants*. The University of Melbourne, Melbourne, Australia.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Fourcade, Y., Besnard, A.G. & Secondi, J. (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, **27**, 245-256.

Franklin, J. (2010) *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge.

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? *Global Ecology and Biogeography*, **21**, 1126-1136.

Gotzenberger, L., de Bello, F., Brathen, K.A., Davison, J., Dubuis, A., Guisan, A., Leps, J., Lindborg, R., Moora, M., Partel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel, K. & Zobel, M. (2012) Ecological assembly rules in plant communities--approaches, patterns and prospects. *Biol Rev Camb Philos Soc*, **87**, 111-27.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A. & Gro, N.P.S.W. (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**, 239-247.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Guisan, A. & Rahbek, C. (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433-1444.

Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) *Habitat Suitability and Distribution Models: With Applications in R.* Cambridge University Press.

Guisan, A., Graham, C.H., Elith, J. & Huettmann, F. (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332-340.

Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331-341.

Hirzel, A.H., Helfer, V. & Metral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145**, 111-121.

Huang, M.Y., Kong, X.Q., Varela, S. & Duan, R.Y. (2016) The Niche Limitation Method (NicheLim), a new algorithm for generating virtual species to study biogeography. *Ecological Modelling*, **320**, 197-202.

Keil, P., Belmaker, J., Wilson, A.M., Unitt, P., Jetz, W. & Freckleton, R. (2013) Downscaling of species distribution models: a hierarchical approach. *Methods in Ecology and Evolution*, **4**, 82-94.

Kenkel, N.C. & Orloci, L. (1986) Applying Metric and Nonmetric Multidimensional-Scaling to Ecological-Studies - Some New Results. *Ecology*, **67**, 919-928.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kuhn, I., McInerny, G.J., Montoya, J.M., Romermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.C., Zimmermann, N.E. & O'Hara, R.B. (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163-2178.

Leroy, B., Meynard, C.N., Bellard, C. & Courchamp, F. (2016) virtualspecies, an R package to generate virtual species distributions. *Ecography*, **39**, 599-607.

Liu, C., Newell, G. & White, M. (2016) On the selection of thresholds for predicting species occurrence with presence-only data. *Ecol Evol*, **6**, 337-48.

Lobo, J.M., Jimenez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145-151.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811-823.

McPherson, J.M., Yeager, L.A., Baum, J.K. & Price, S. (2018) A simulation tool to scrutinise the behaviour of functional diversity metrics. *Methods in Ecology and Evolution*, **9**, 200-206.

Mestre, F., Canovas, F., Pita, R., Mira, A. & Beja, P. (2016) An R package for simulating metapopulation dynamics and range expansion under environmental change. *Environmental Modelling & Software*, **81**, 40-44.

Meynard, C.N. & Quinn, J.F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**, 1455-1469.

Meynard, C.N. & Kaplan, D.M. (2013) Using virtual species to study species distributions and model performance. *Journal of Biogeography*, **40**, 1-8.

Midgley, G.F., Davies, I.D., Albert, C.H., Altwegg, R., Hannah, L., Hughes, G.O., O'Halloran, L.R., Seo, C., Thorne, J.H. & Thuiller, W. (2010) BioMove - an integrated platform simulating the dynamic response of species to environmental change. *Ecography*, **33**, 612-616.

Miller, J.A. (2014) Virtual species distribution models: Using simulated data to evaluate aspects of model performance. *Progress in Physical Geography*, **38**, 117-128.

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. *Methods in Ecology and Evolution*, **8**, 12-21.

Münkemüller, T. & Gallien, L. (2015) VirtualCom: A simulation model for eco-evolutionary community assembly and invasion. *Methods in Ecology and Evolution*, n/a-n/a.

Munoz, F., Grenié, M., Denelle, P., Taudière, A., Laroche, F., Tucker, C., Violle, C. & Chisholm, R. (2018) ecolottery

: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution*, **9**, 693-703.

Pearson, R.G., Dawson, T.P. & Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285-298.

Peterson, A.T., Papeş, M. & Soberón, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, **213**, 63-72.

Petitpierre, B., McDougall, K., Seipel, T., Broenniman, O., Guisan, A. & Kueffer, C. (2016) Will climate change increase the risk of plant invasions into mountains? *Ecological Applications*, Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P., Guisan, A. & Field, R. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **22**, 52-63.

Railsback, S.F., Stauffer, H.B. & Harvey, B.C. (2003) What can habitat preference models tell us? Tests using a virtual trout population. *Ecological Applications*, **13**, 1580-1594.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689-1704.

Scherrer, D., D'Amen, M., Mateo, M.R.G., Fernandes, R.F. & Guisan , A. (2018) How to best threshold and validate stacked species assemblages? Community optimisation might hold the answer. *Methods in Ecology and Evolution*, **in press**

Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1-13.

Swan, J.M.A. (1970) An examination of some ordination problems by use of simulated vegetation data. *Ecology*, 89-102.

Talluto, M.V., Boulangeat, I., Ameztegui, A., Aubin, I., Berteaux, D., Butler, A., Doyon, F., Drever, C.R., Fortin, M.-J., Franceschini, T., Liénard, J., McKenney, D., Solarik, K.A., Strigul, N., Thuiller, W. & Gravel, D. (2016) Cross-scale integration of knowledge for predicting species ranges: a metamodelling framework. *Global Ecology and Biogeography*, **25**, 238-249.

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C. & Guisan, A. (2014) Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution*, **5**, 947-955.

Thuiller, W., Brotons, L., Araujo, M.B. & Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165-172.

Titeux, N., Maes, D., Van Daele, T., Onkelinx, T., Heikkinen, R.K., Romo, H., García-Barros, E., Munguira, M.L., Thuiller, W., van Swaay, C.A.M., Schweiger, O., Settele, J., Harpke, A., Wiemers, M., Brotons, L. & Luoto, M. (2017) The need for large-scale distribution data to estimate regional changes in species richness under future climate change. *Diversity and Distributions*, **23**, 1393-1407.

Vicente, J.R., Alagador, D., Guerra, C., Alonso, J.M., Kueffer, C., Vaz, A.S., Fernandes, R.F., Cabral, J.A., Araujo, M.B. & Honrado, J.P. (2016) Cost-effective monitoring of biological invasions under global change: a model-based framework. *Journal of Applied Ecology*, **53**, 1317-1329.

Whittaker, R.H. (2012) *Ordination of plant communities*. Springer Science & Business Media.

Williams, J.N., Seo, C.W., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M. & Schwartz, M.W. (2009) Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, **15**, 565-576.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & Distribut, N.P.S. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763-773.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Hoye, T.T., Kuhn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Ockinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol Rev Camb Philos Soc*, **88**, 15-30.

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Munkemuller, T., Nehrbass, N., Pagel, J., Reineking, B., Schroder, B. & Grimm, V. (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622-635.

APPENDICES

Supplementary Information

Supplementary material of chapter 1.1

Appendix 1: Description of species data and environmental predictors used to create virtual species

Plant data

We created virtual species distribution maps using a real species presence/absence dataset. This dataset of grassland species was collected following an equal random-stratified sampling of non-forested areas in the study area (i.e. Alps of western Switzerland; http://rechalpvd.unil.ch). The data was collected between 2002 and 2009 and consisted in a total of 911 vegetation plots of 4 m^2 (Fig. 1; for more information see Dubuis *et al.* 2011). While a total of 905 plant species were recorded, only the 212 most frequent (>20 occurrences) were selected for modelling. Then, a set of 100 of those species were randomly selected to be used has our studied virtual species.

Environmental predictors

In order to both generate our virtual species and create the models using the "degraded" training data (see *Modelling procedure*), a set of 5 environmental predictors were used as explanatory variables: summer mean monthly temperatures, sum of winter precipitation, solar radiation, slope (in degrees) and topographic position (unit-less, indicating the ridges and valleys). The first three variables were derived from temperature and precipitation values interpolated from a network of meteorological stations and from GIS-derived solar radiation. A digital elevation model was used to derive slope and topographic position. All these variables were at a 25 m resolution and have been shown to be useful predictors for plant species in mountain environments (see Dubuis *et al*., 2011; D'Amen *et al*., 2015; Scherrer *et al*., 2017 for details on predictors).

References:

D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L. & Guisan, A. (2015) Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *Journal of Biogeography*, **42**, 1255-1266.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122-1131.

Scherrer, D., Massy, S., Meier, S., Vittoz, P., Guisan, A. & Serra-Diaz, J. (2017) Assessing and predicting shifts in mountain forest composition across 25 years of climate change. *Diversity and Distributions*, **23**, 517-528.

Appendix 2: Results obtained using sampling design “EqualPrev”

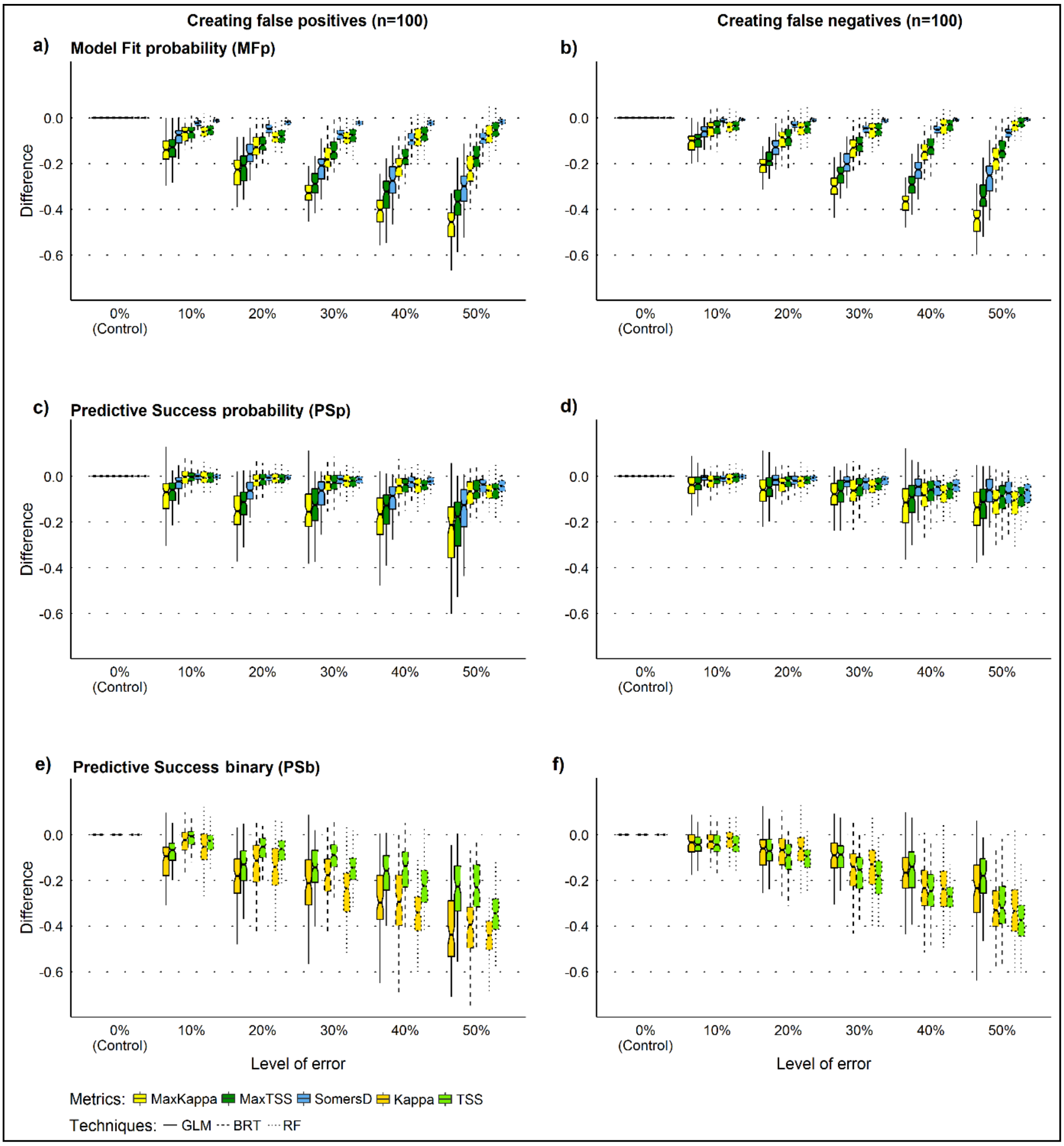


Figure A1 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **100**, for virtual species created using **GLM**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers’D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

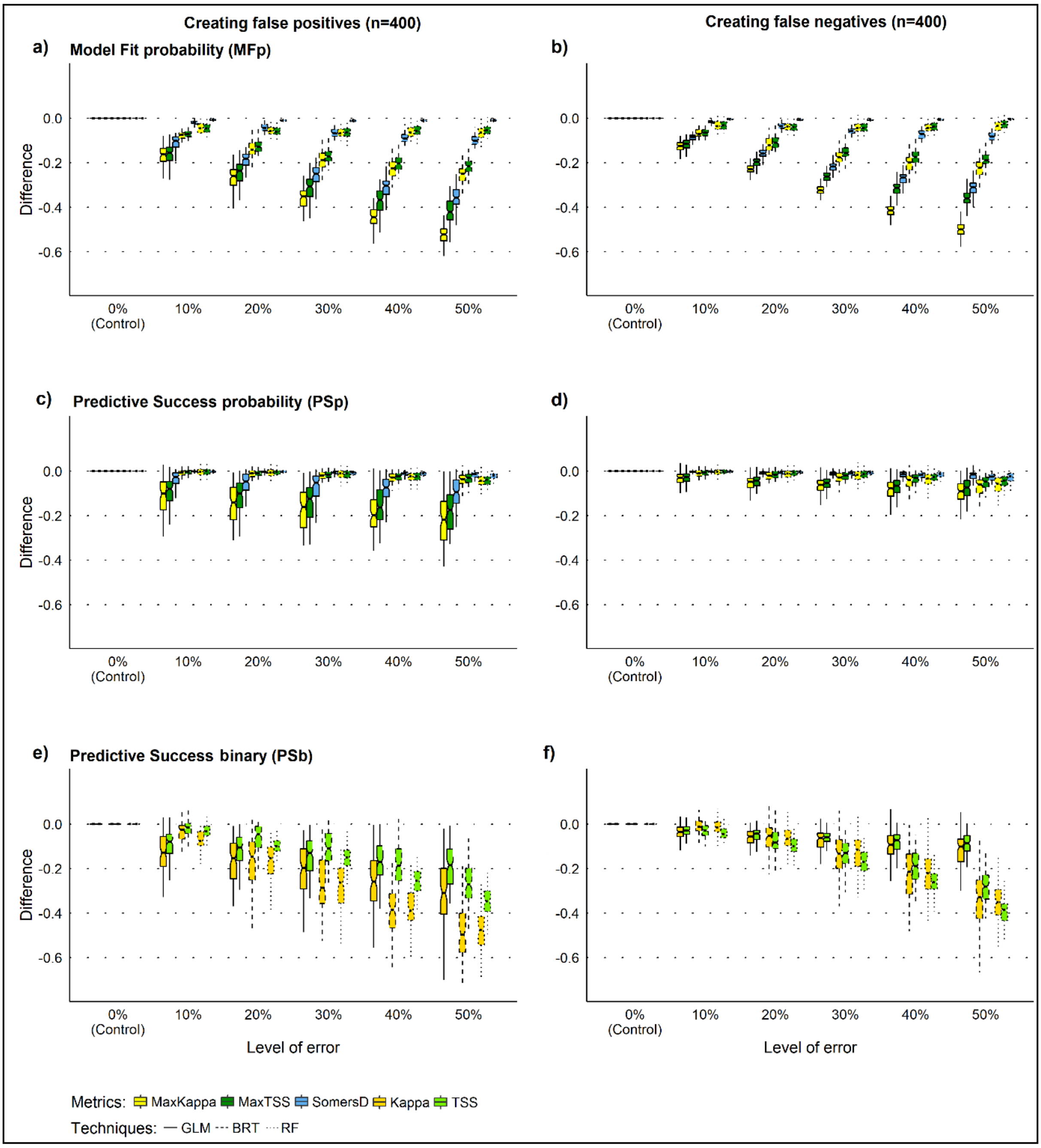


Figure A2 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **400**, for virtual species created using **GLM**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

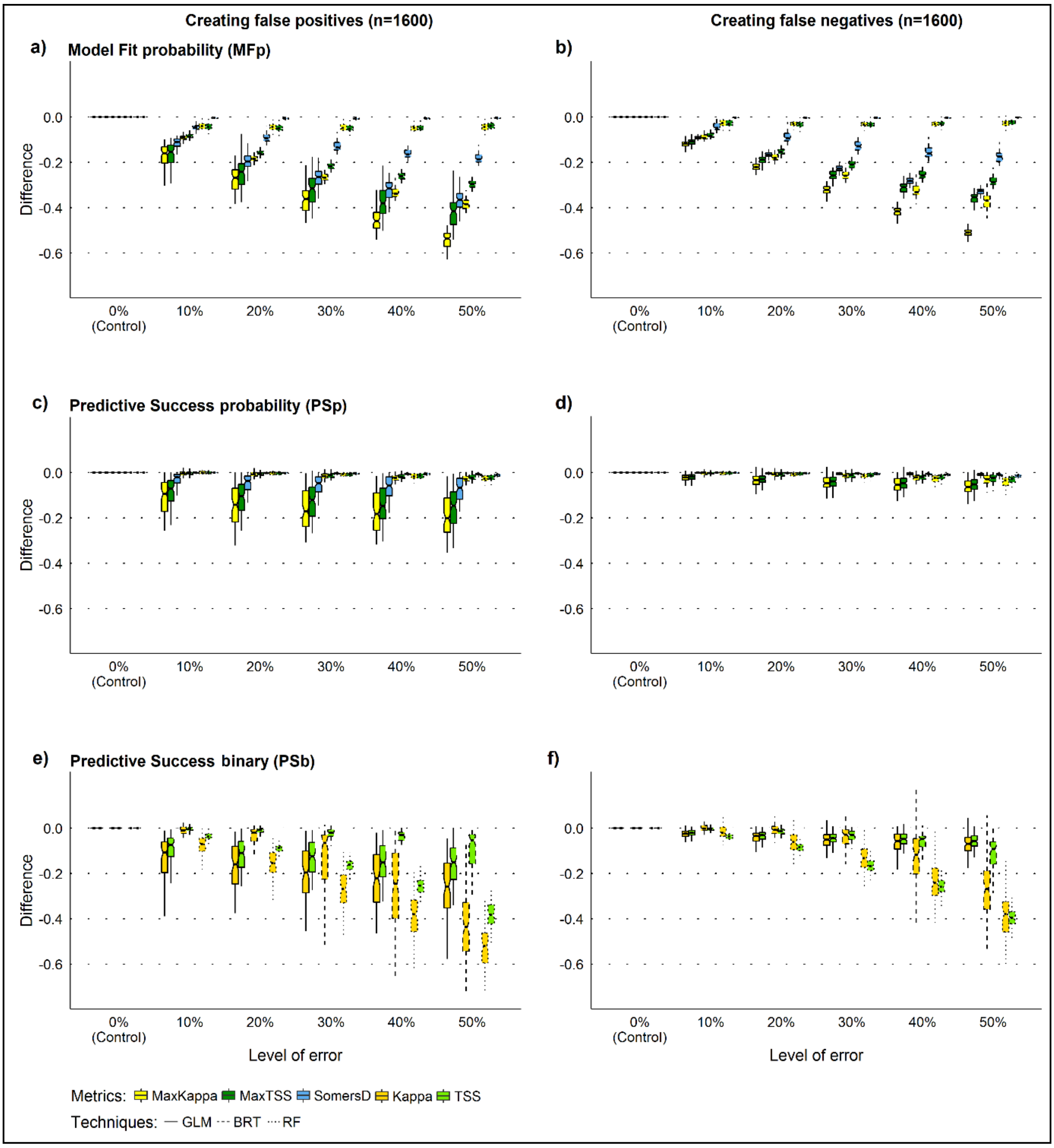


Figure A3 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **1600**, for virtual species created using **GLM**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

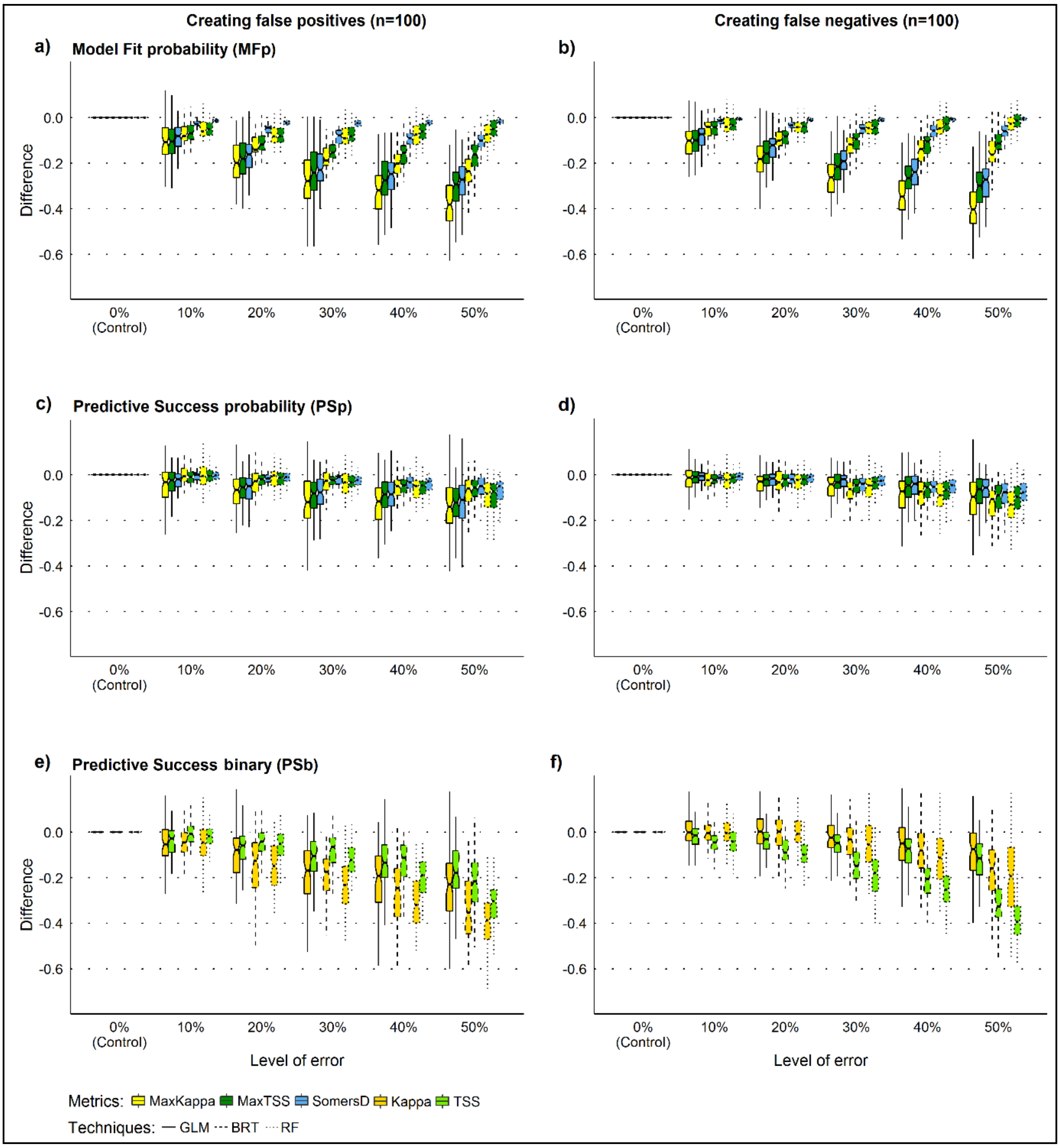


Figure A4 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **100**, for virtual species created using **BRT**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

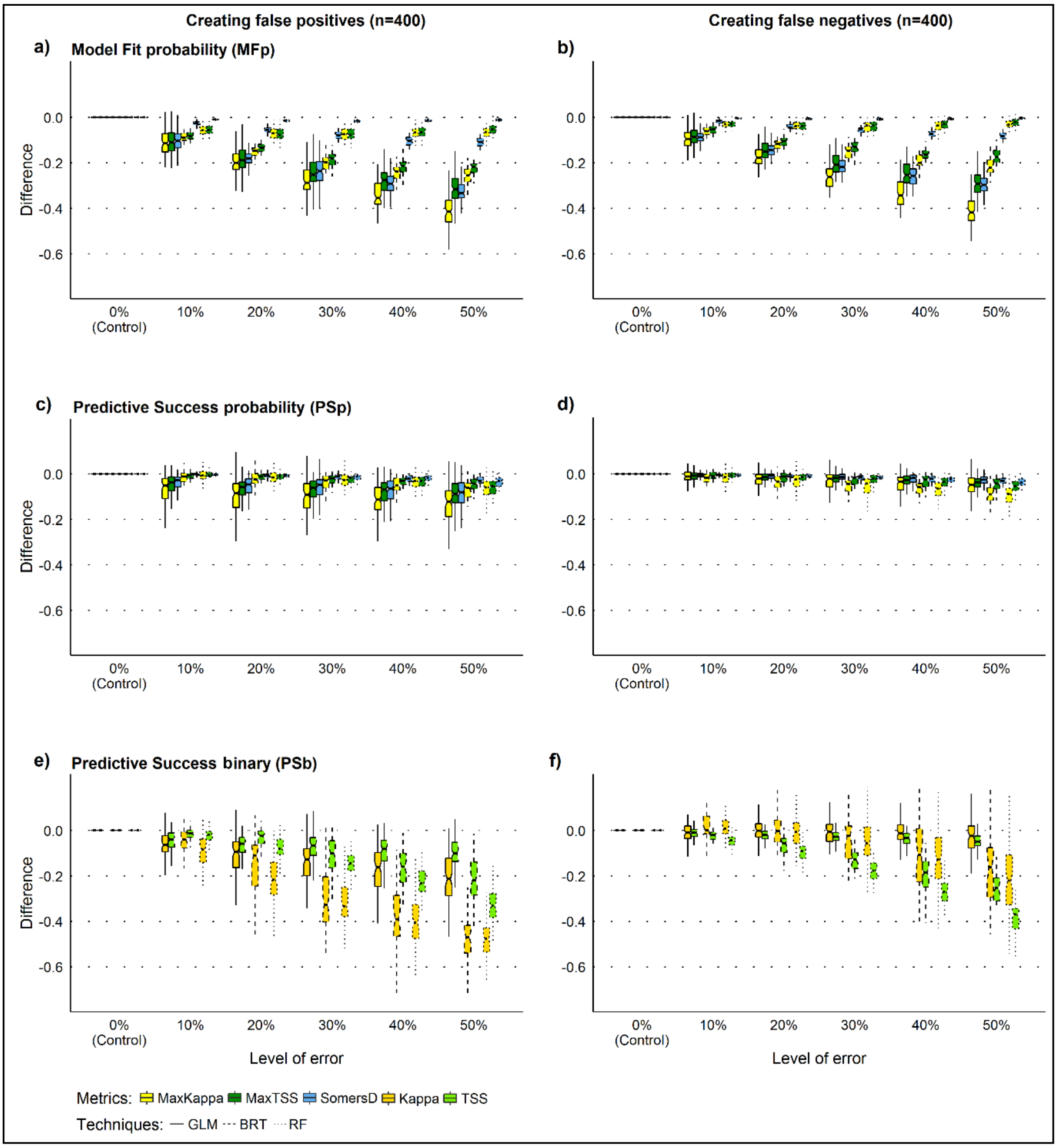


Figure A5 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **400**, for virtual species created using **BRT**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

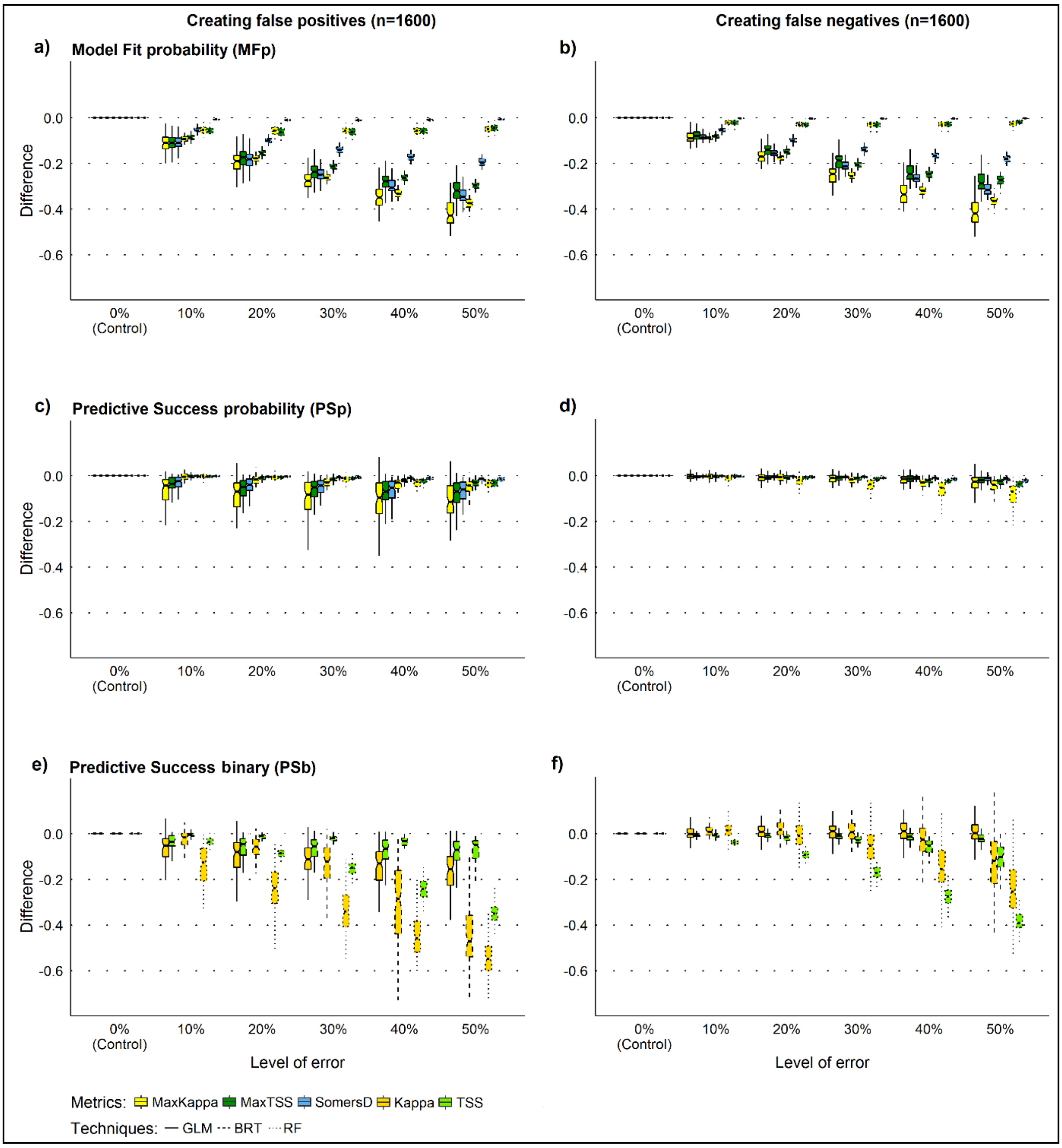


Figure A6 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **1600**, for virtual species created using **BRT**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

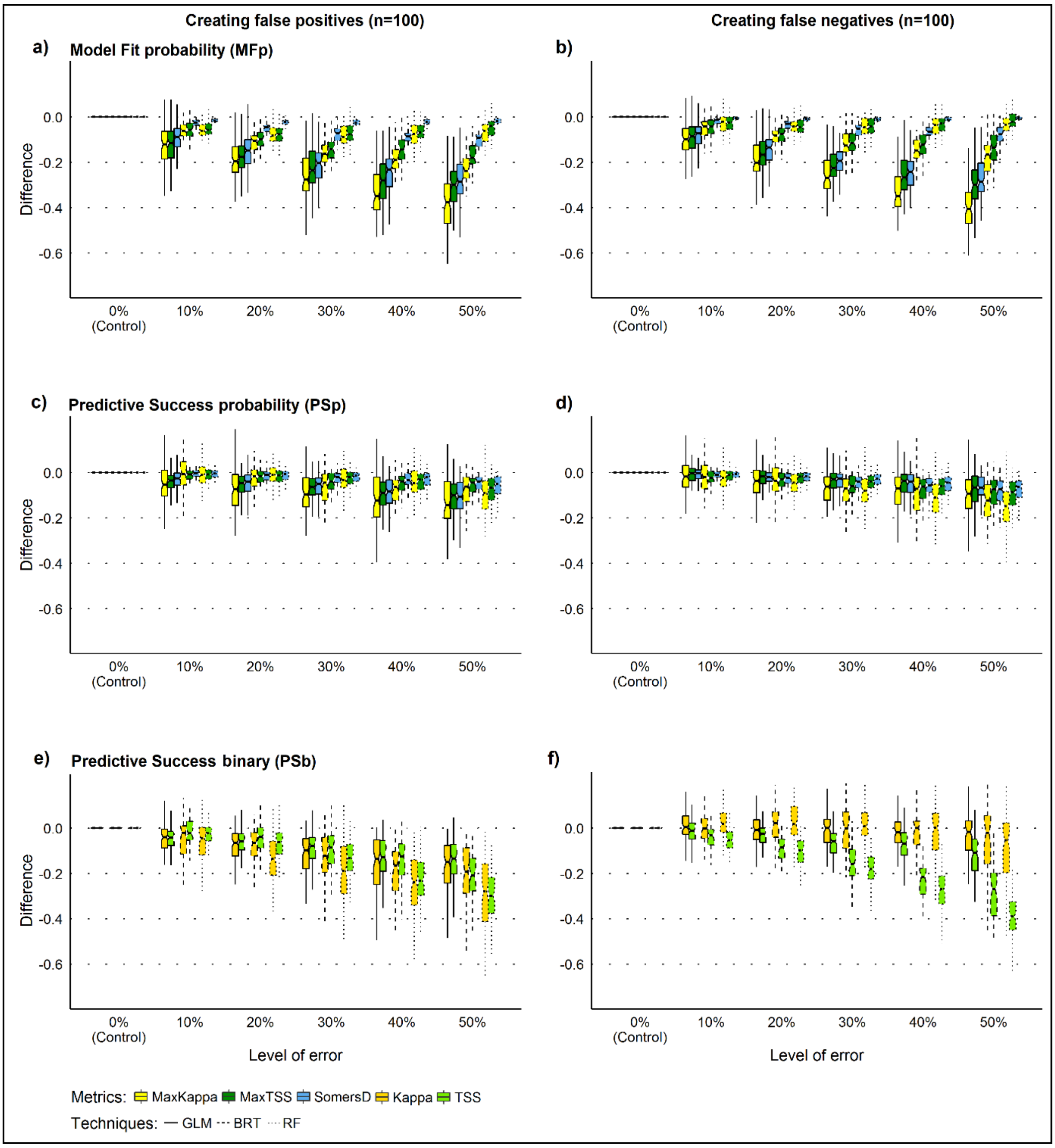


Figure A7 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **100**, for virtual species created using **RF**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

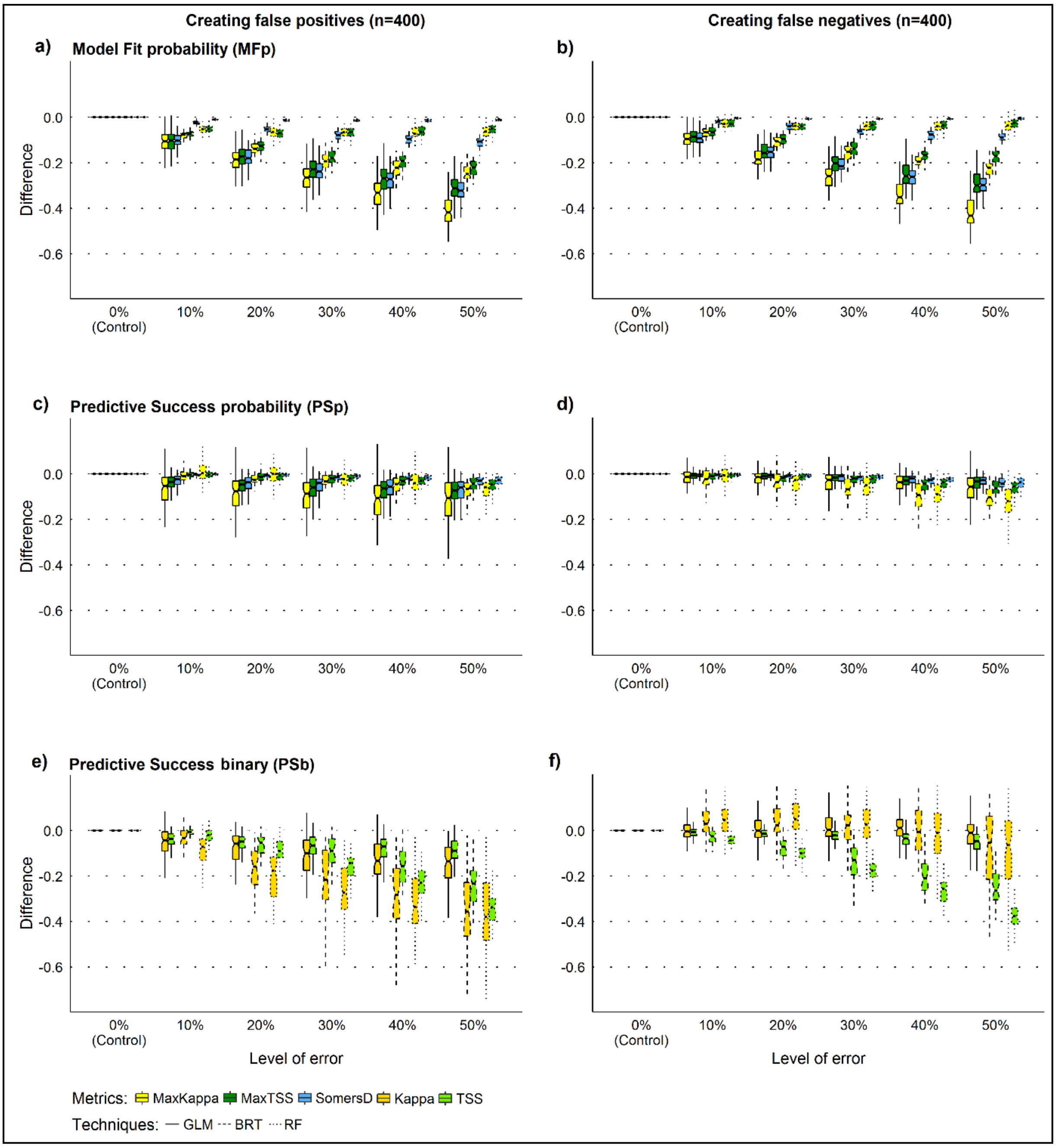


Figure A8 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **400**, for virtual species created using **RF**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

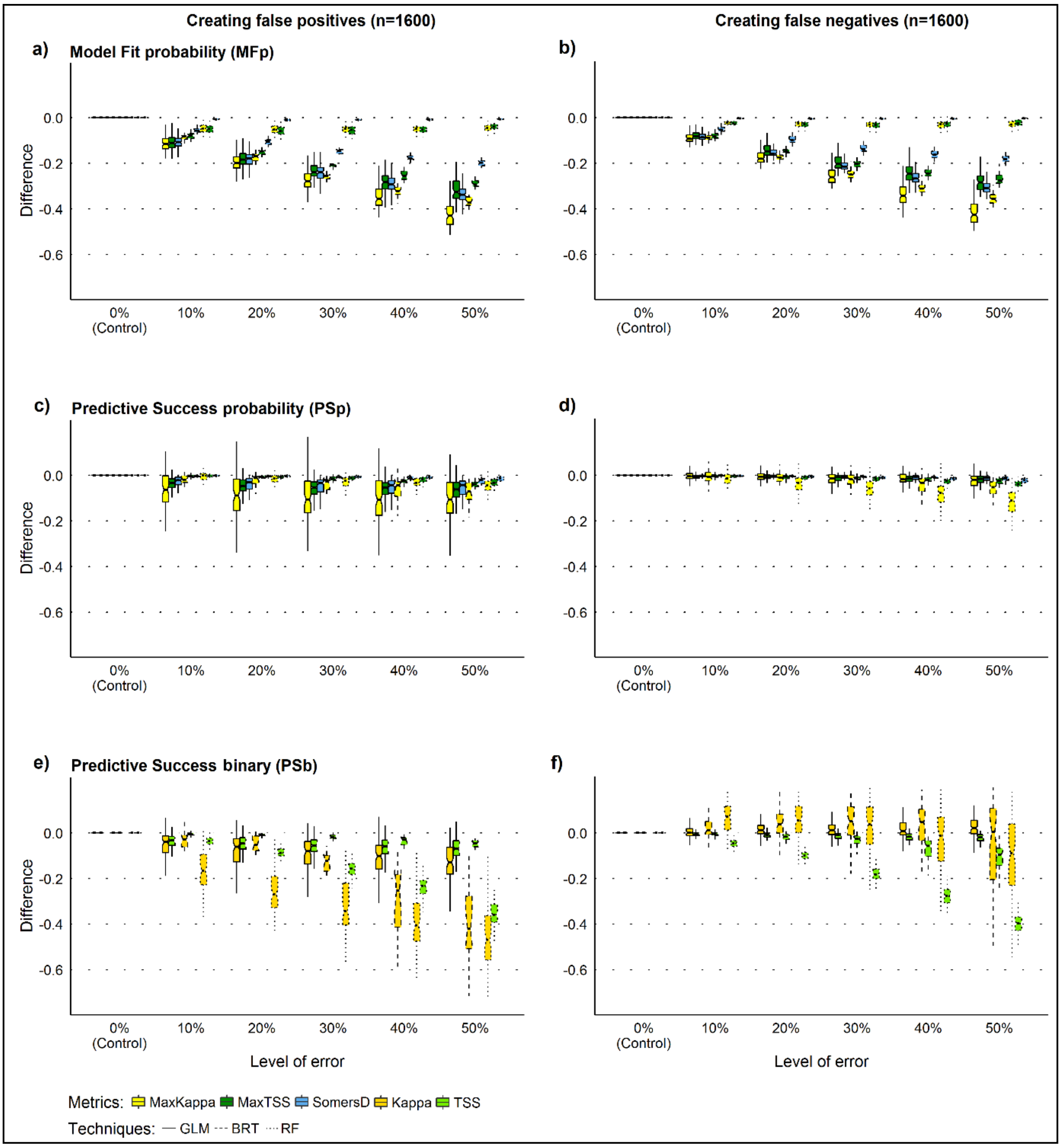


Figure A9 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size **1600**, for virtual species created using **RF**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

Appendix 3: Results obtained using sampling design "TruePrev"

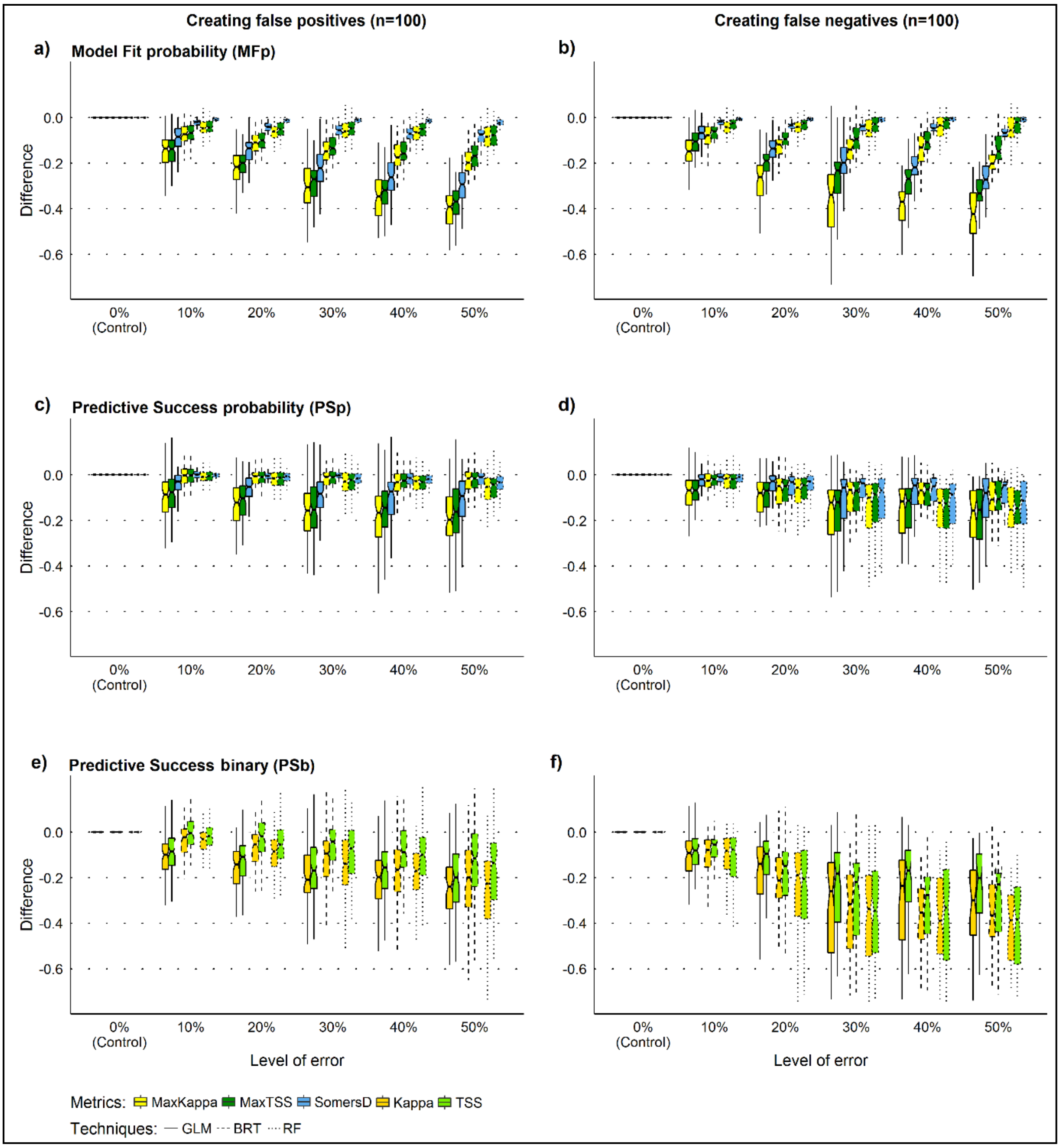


Figure A10 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **100**, for virtual species created using **GLM**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

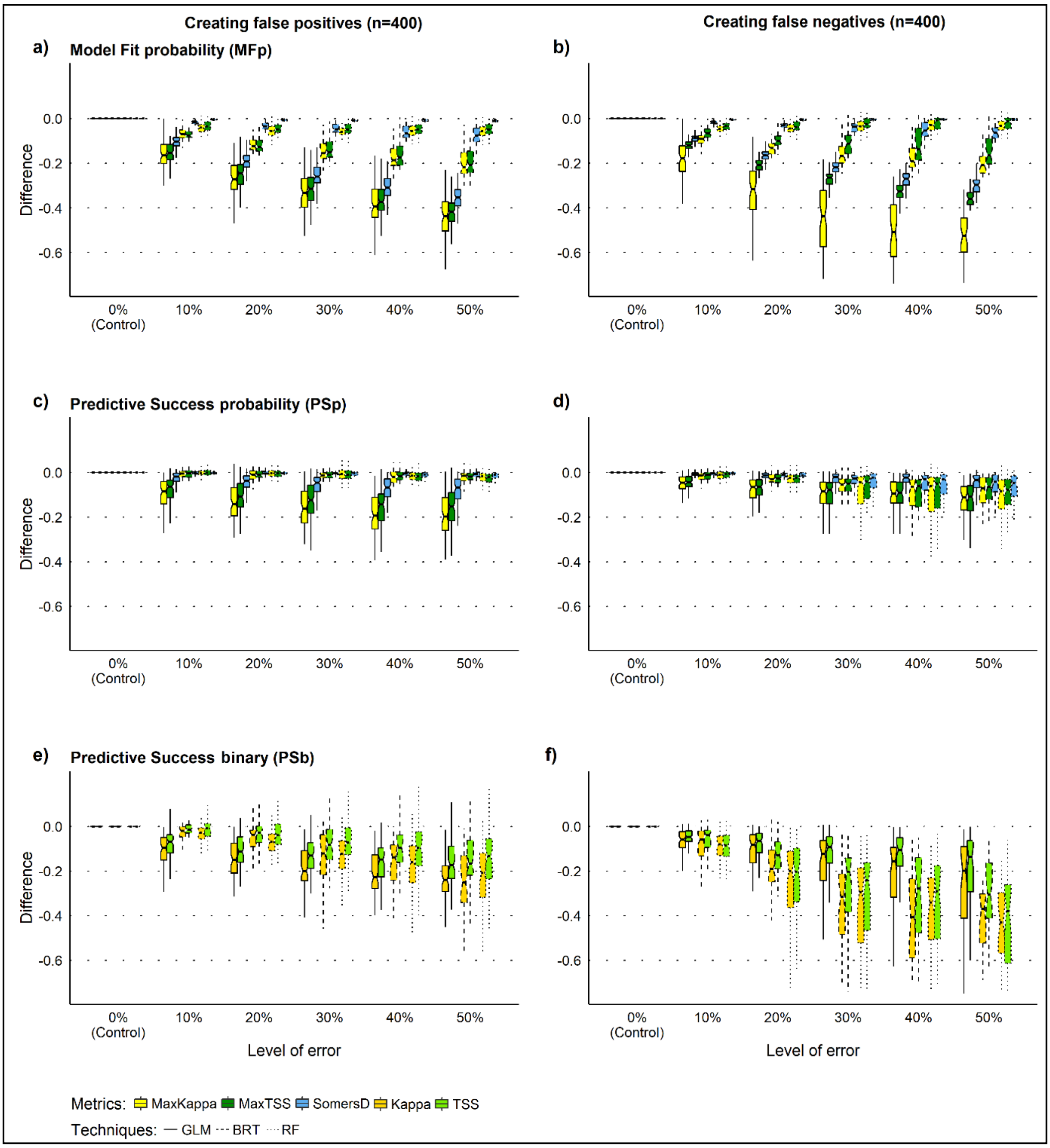


Figure A11 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **400**, for virtual species created using **GLM**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

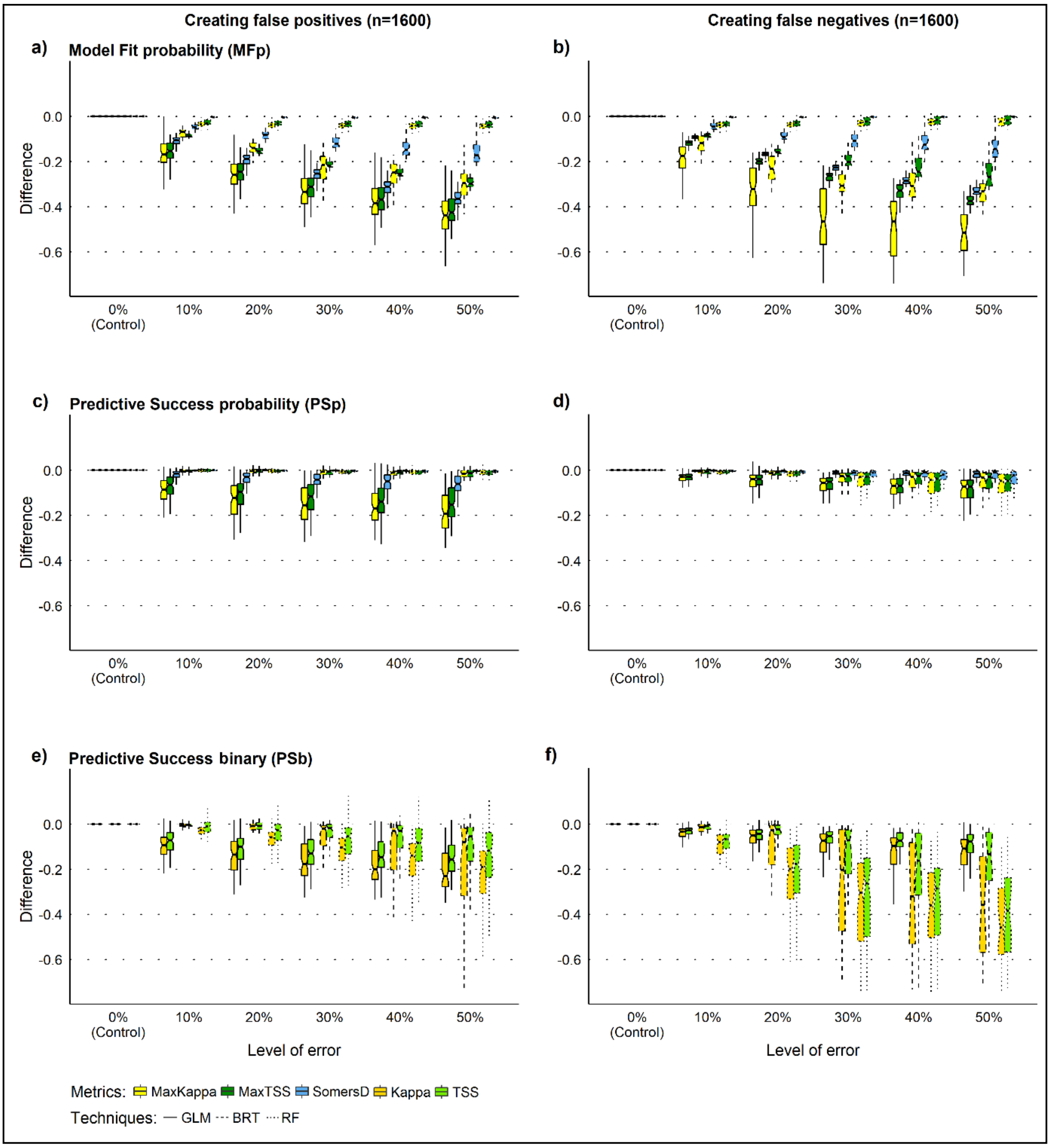


Figure A12 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **1600**, for virtual species created using **GLM**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

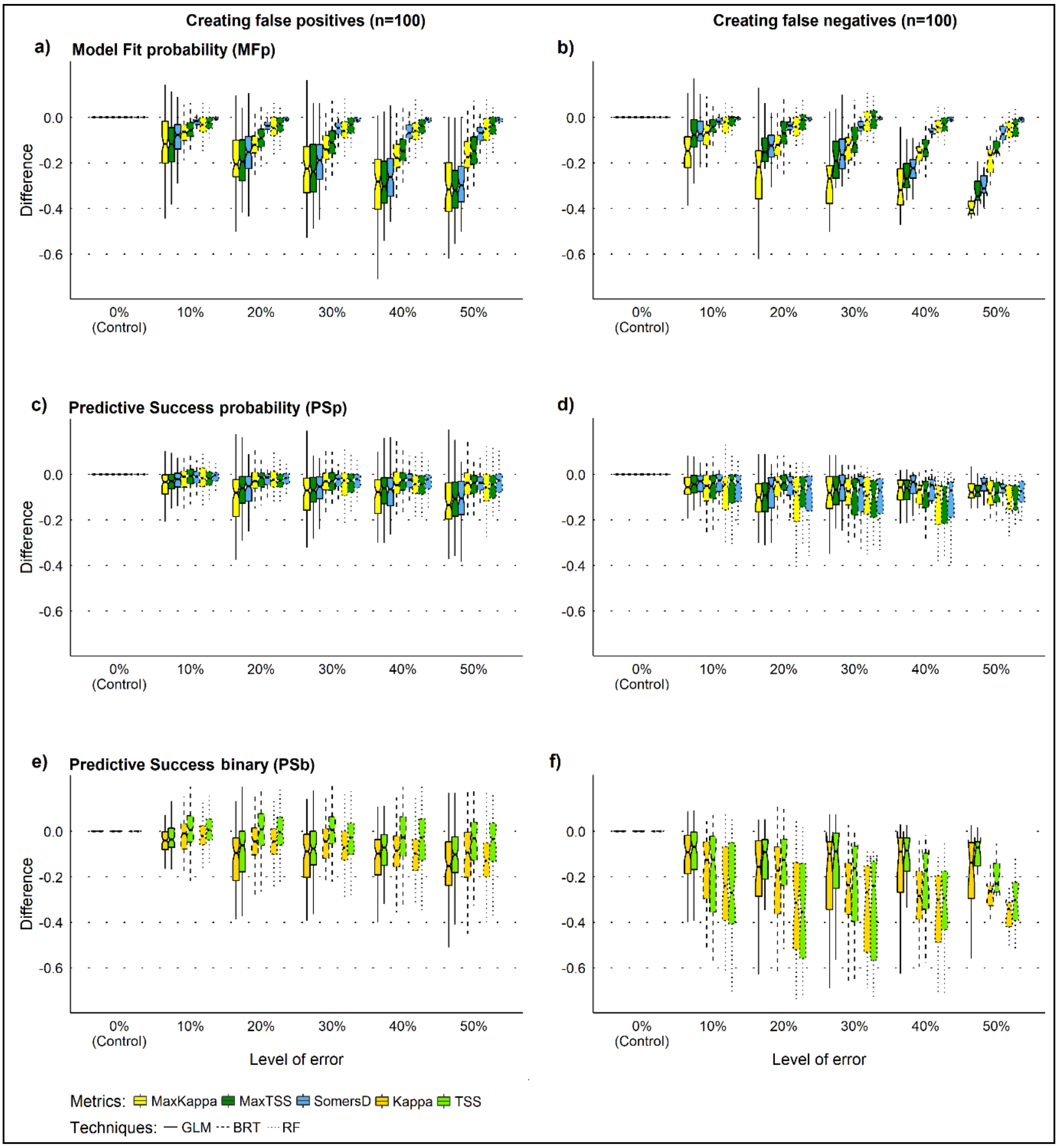


Figure A13 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **100**, for virtual species created using **BRT**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

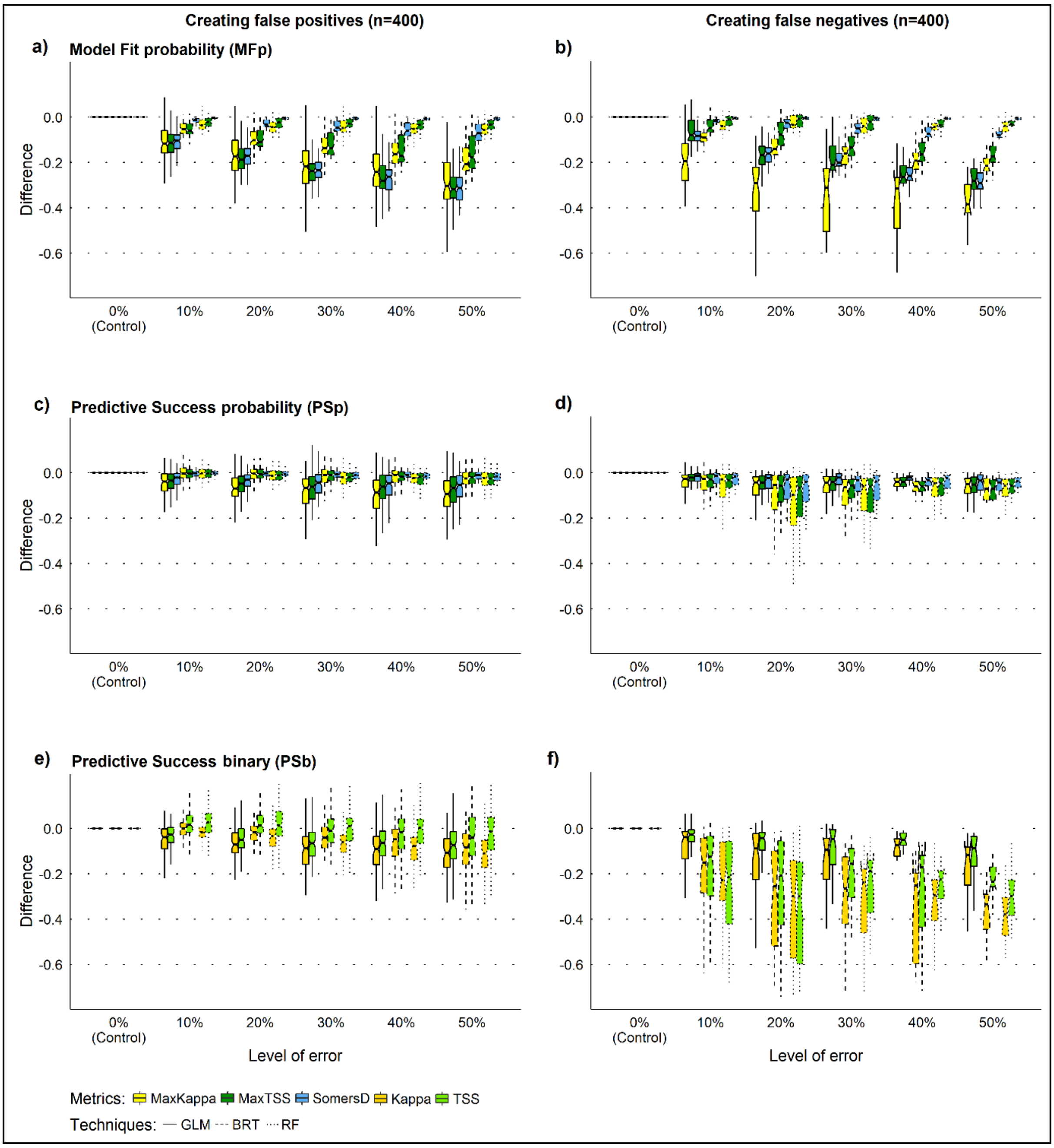


Figure A14 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **400**, for virtual species created using **BRT**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

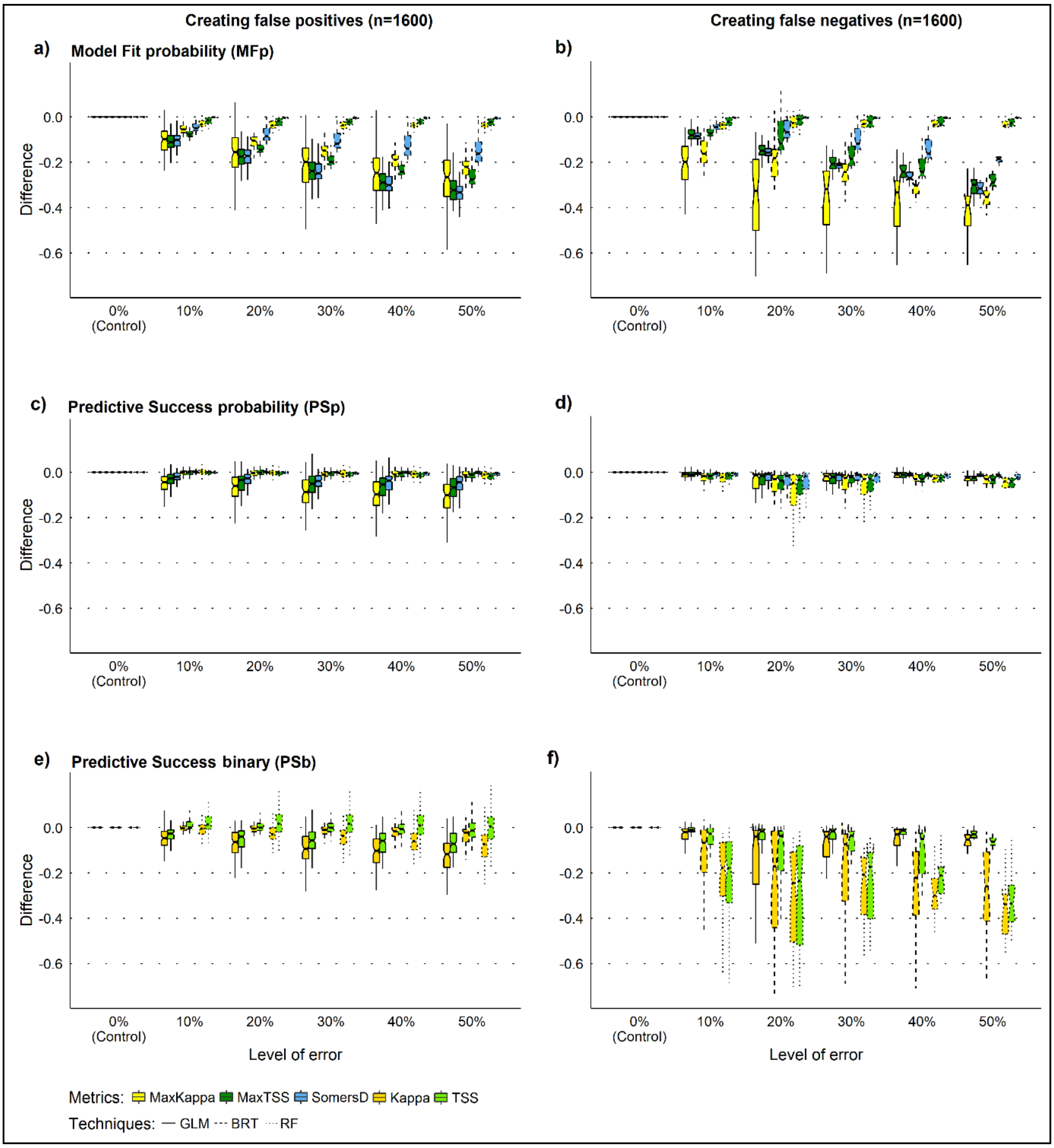


Figure A15 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **1600**, for virtual species created using **BRT**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

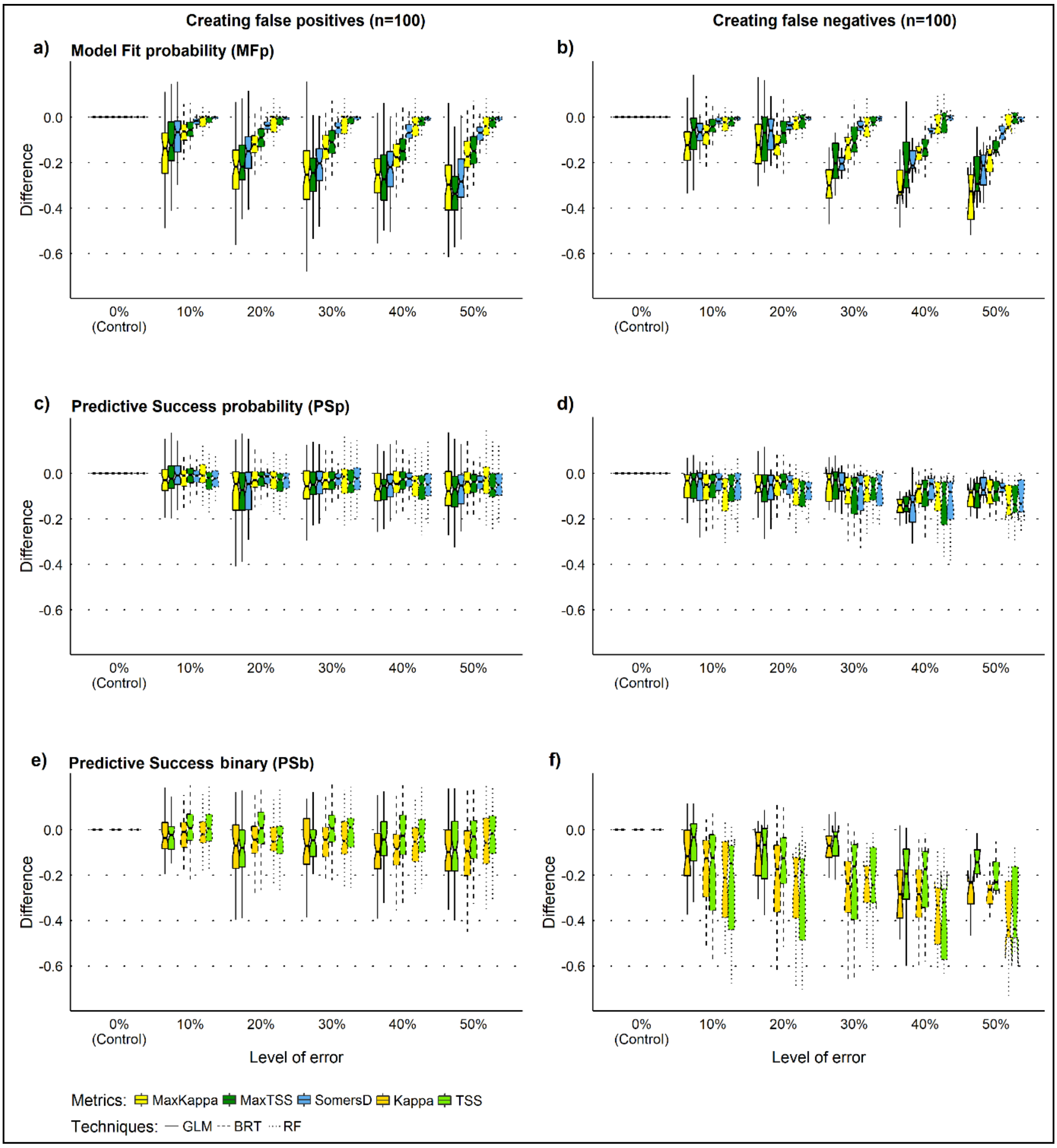


Figure A16 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **100**, for virtual species created using **RF**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

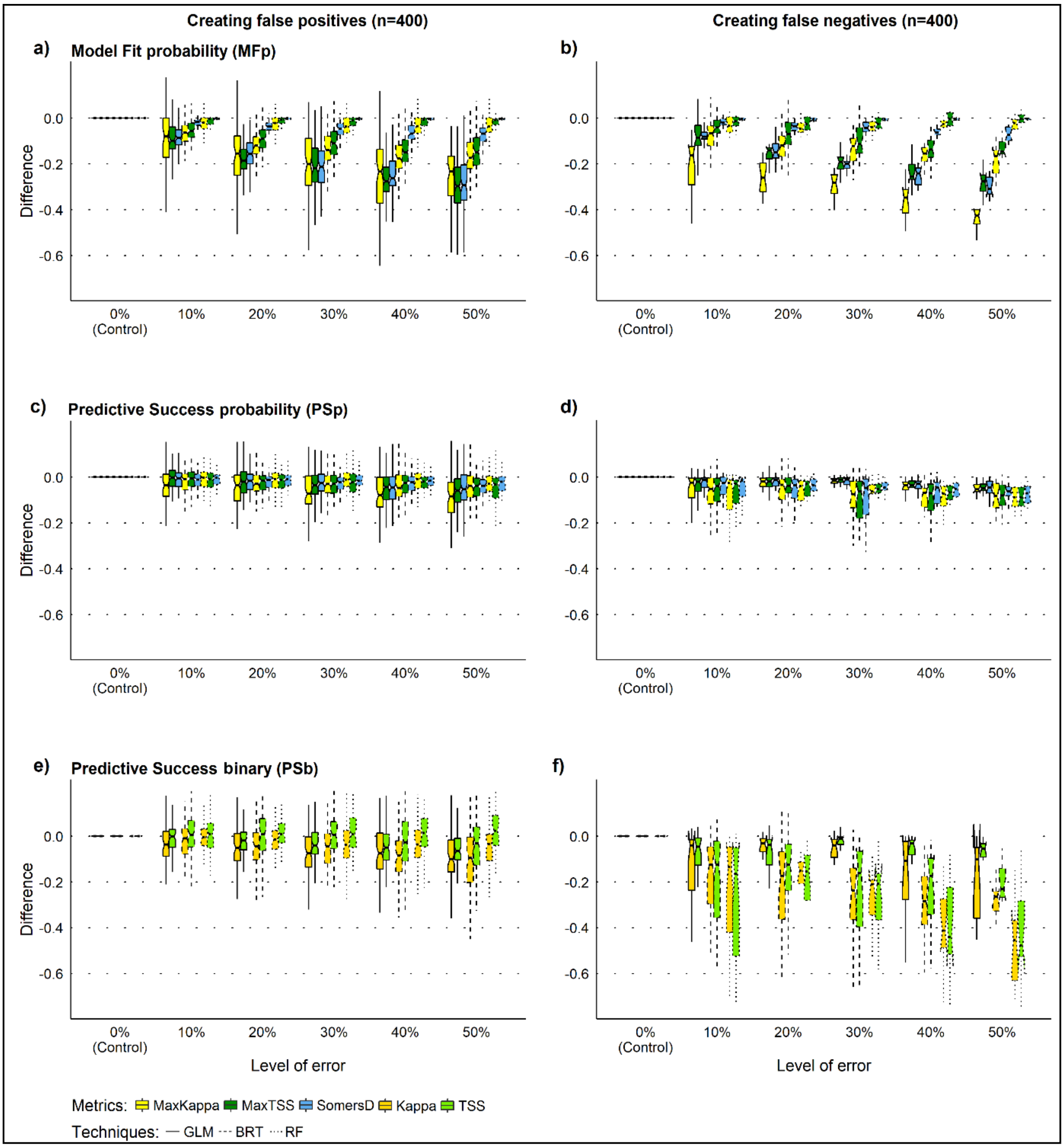


Figure A17 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **400**, for virtual species created using **RF**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

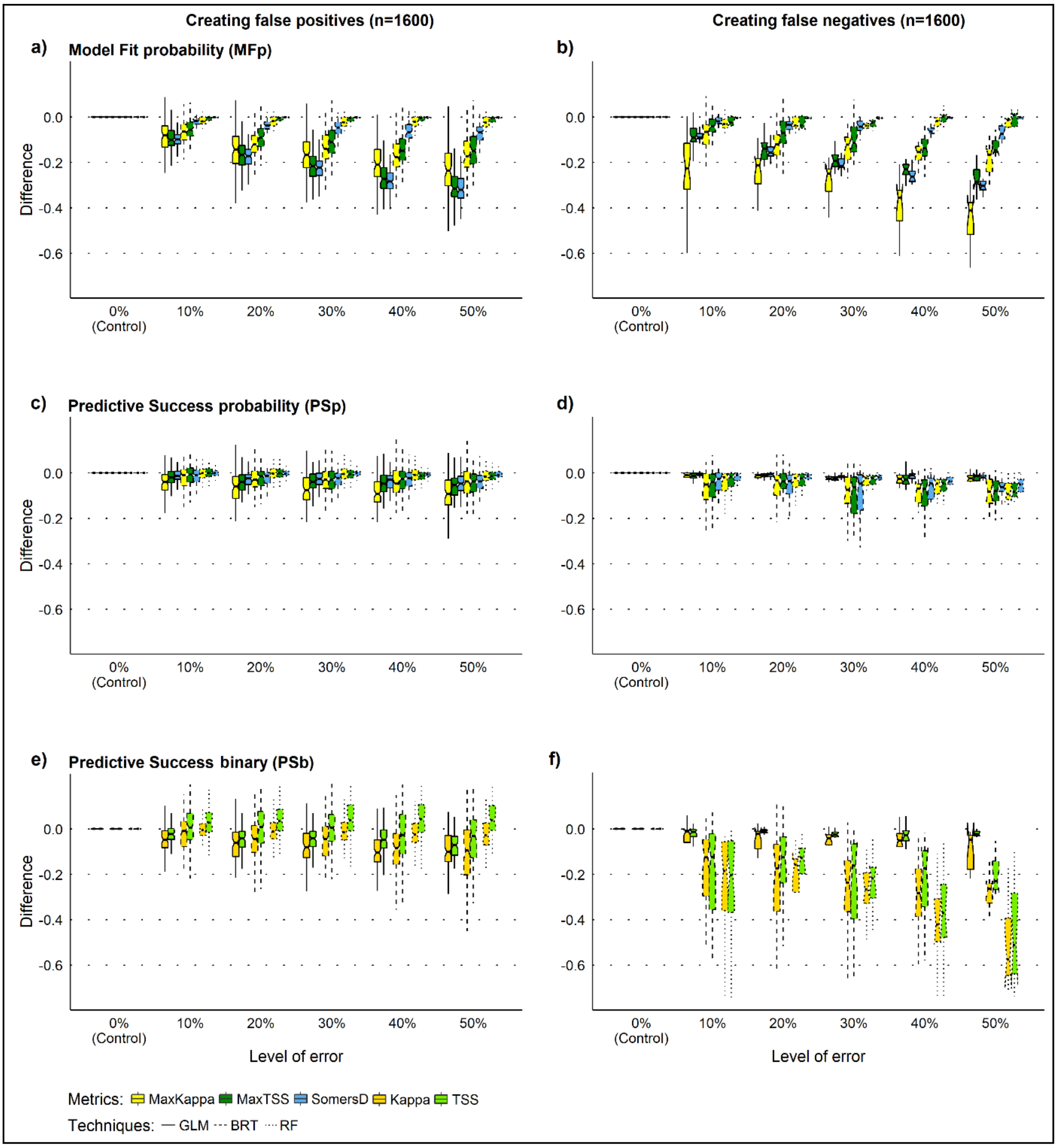


Figure A18 Observed difference of measured **MFp**, **PSp** and **PSb** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size **1600**, for virtual species created using **RF**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp and PSp were measured using MaxKappa (yellow), MaxTSS (green) and Somers'D (blue), while PSb was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). For PSb, only two plots are present in each of the three sets. See Fig. 1 for the explanation of the different analyses.

Appendix 4: Comparison of results obtained using different thresholding techniques (ROC, Kappa and TSS) to create initial SDMs

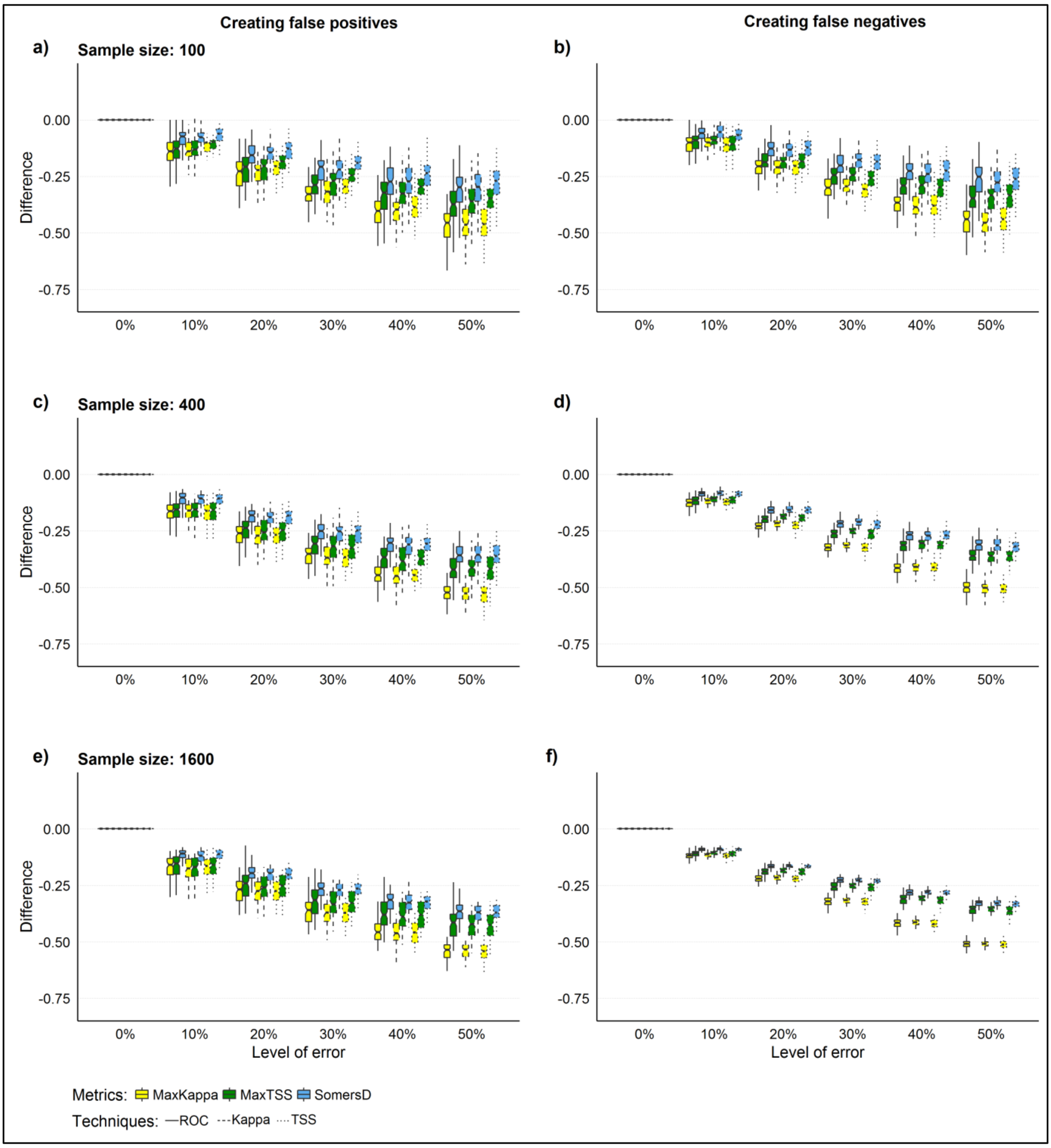


Figure A19 Graphical representation of measured **MFp**, under the sampling design **EqualPrev** and for virtual species created using **GLM**. Using different sample sizes (100, 400 and 1600), errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). MFp was measured by different metrics: MaxKappa (yellow), MaxTSS (green) and Somers'D (blue). For each level of error, three sets with three plots are observed, corresponding to models binarized using either **ROC** (solid plots), **Kappa** (dashed plots) or **TSS** (dotted plots). **These models were all fitted using GLM**.

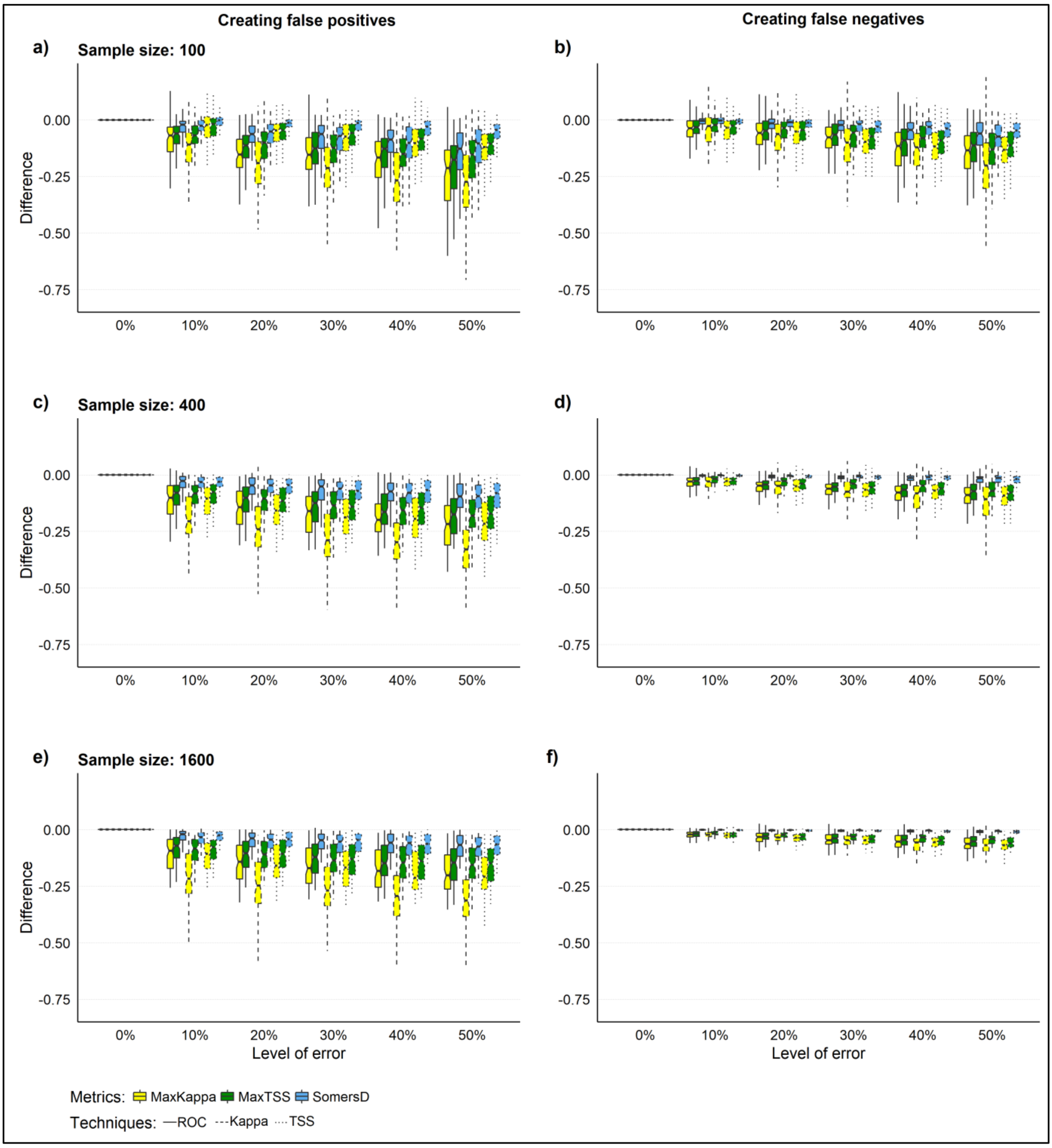


Figure A20 Graphical representation of measured **PSp**, under the sampling design **EqualPrev** and for virtual species created using **GLM**. Using different sample sizes (100, 400 and 1600), errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). PSp was measured by different metrics: MaxKappa (yellow), MaxTSS (green) and Somers'D (blue). For each level of error, three sets with three plots are observed, corresponding to models binarized using either **ROC** (solid plots), **Kappa** (dashed plots) or **TSS** (dotted plots). **These models were all fitted using GLM**.

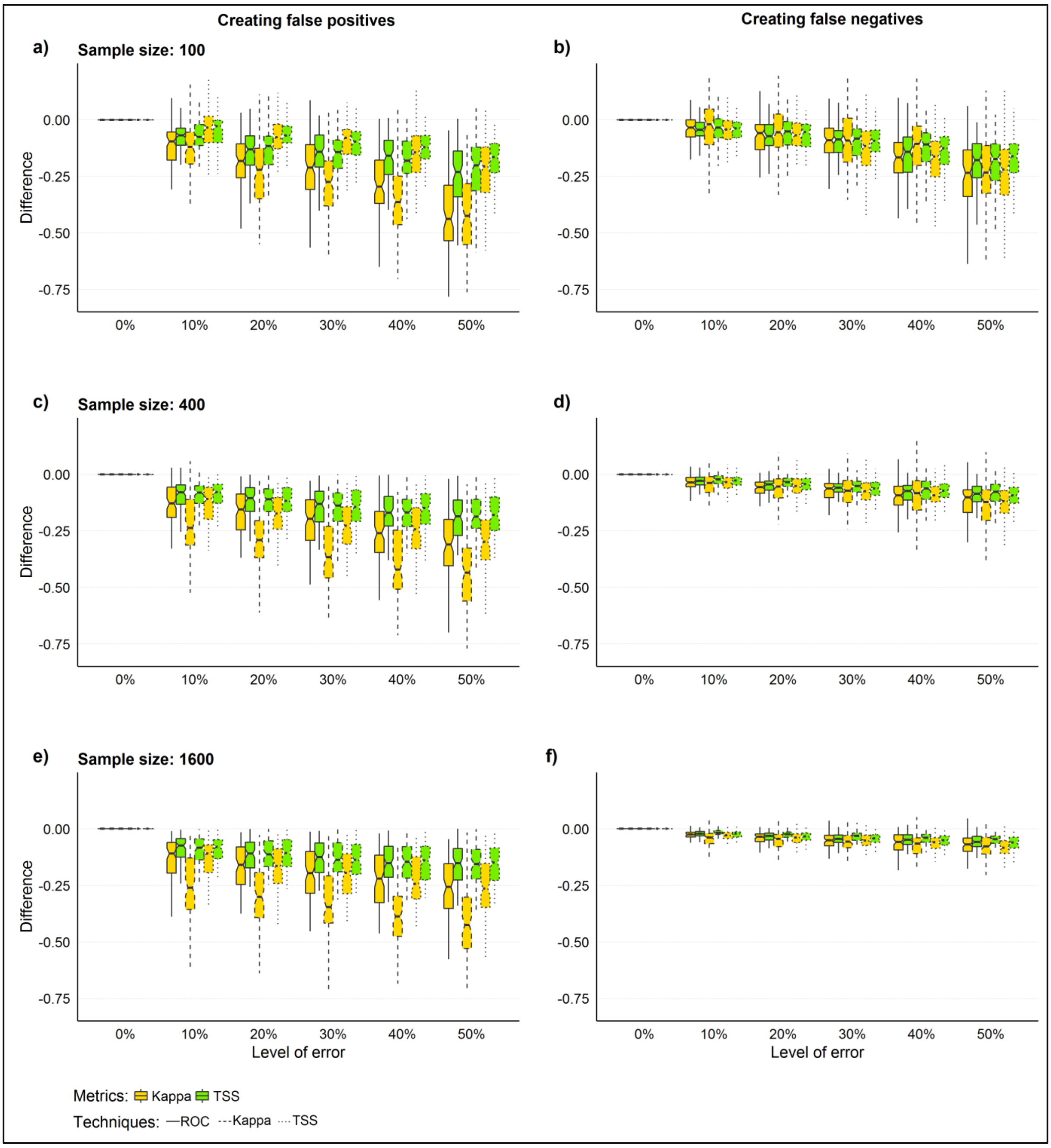


Figure A21 Graphical representation of measured **PSb**, under the sampling design **EqualPrev** and for virtual species created using **GLM**. Using different sample sizes (100, 400 and 1600), errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). PSb was measured by different metrics: Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models binarized using either **ROC** (solid plots), **Kappa** (dashed plots) or **TSS** (dotted plots). **These models were all fitted using GLM.**

Appendix 5: Description and results of evaluation approach Predictive Success on Calibration data (PSc)

We initially evaluated all predictions built for each sample size, sampling design and modelling technique by measuring model fit on probability (model fit probability) at sampled sites and predictive success for probabilistic (predictive success probability) and binary predictions (predictive success binary) across the whole area (Fig. 1). Additionally, we also evaluated the predictions measuring the predictive success calibration. For this extra evaluation approach, we calculated observed Kappa and TSS under a chosen threshold (Step 5, Fig. A22):

- Predictive Success Calibration (PSc) is the ability of the model to predict the true binary species distribution for the points used to calibrate the model despite having the various levels of errors added to the training data). It was calculated by comparing binary predictions of the different models (control and various levels of errors), to the initially sampled true distribution dataset (i.e. control without error added). To create the binary predictions, MaxTSS (for the calculation of TSS) and MaxKappa (for the calculation of Kappa) thresholds were selected based on the training data (with or without error) and the predicted probabilities corresponding to each model.

However, we opted for not including this additional measure in the main manuscript since it just represents a subset of the results presented by predictive success binary (see below for results obtained when virtual species were created by GLMs).

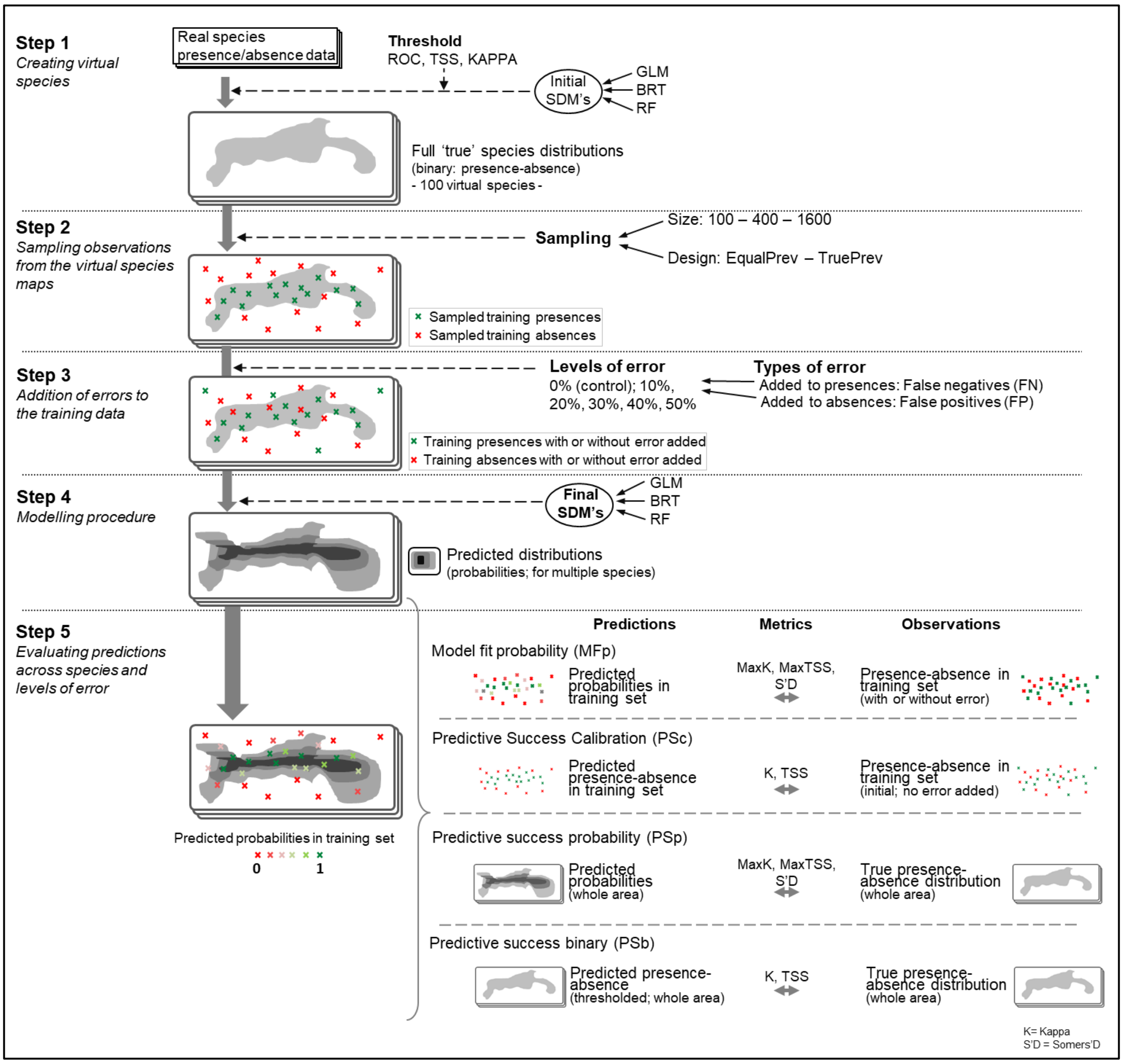


Figure A22 Workflow of the analytical steps followed in the study with additional evaluation approach. **Step 1 –** We started by creating binary distribution maps for 100 virtual species from models based on real species' data (using either GLM, BRT or RF as modelling techniques and ROC, TSS or KAPPA as thresholding techniques). **Step 2 –** For each species, we sampled presence-absence data using three different sample sizes (100, 400 and 1600) and two sampling designs (EqualPrev and TruePrev). **Step 3 –** To each of the sampled datasets, errors were added according to 6 different levels (0% - training data without error added, used as control; 10%, 20%, 30%, 40% and 50% - training data with error added) and two different types of error (errors added to presences, creating false negatives or errors added to absences, creating false positives). **Step 4 –** Each occurrence dataset was used to create single species distribution models (probability and binary maps), using three different modelling techniques (GLM, BRT and RF). *Step 5* – The predictions for each species were then evaluated with four evaluation approaches: model fit probability (MFp), predictive success calibration (PSc), predictive success probability (PSp) and predictive success binary (PSb), using different metrics:MaxKappa, MaxTSS and Somers' D for MFp and PSp; Kappa and TSS for PSc and PSb.

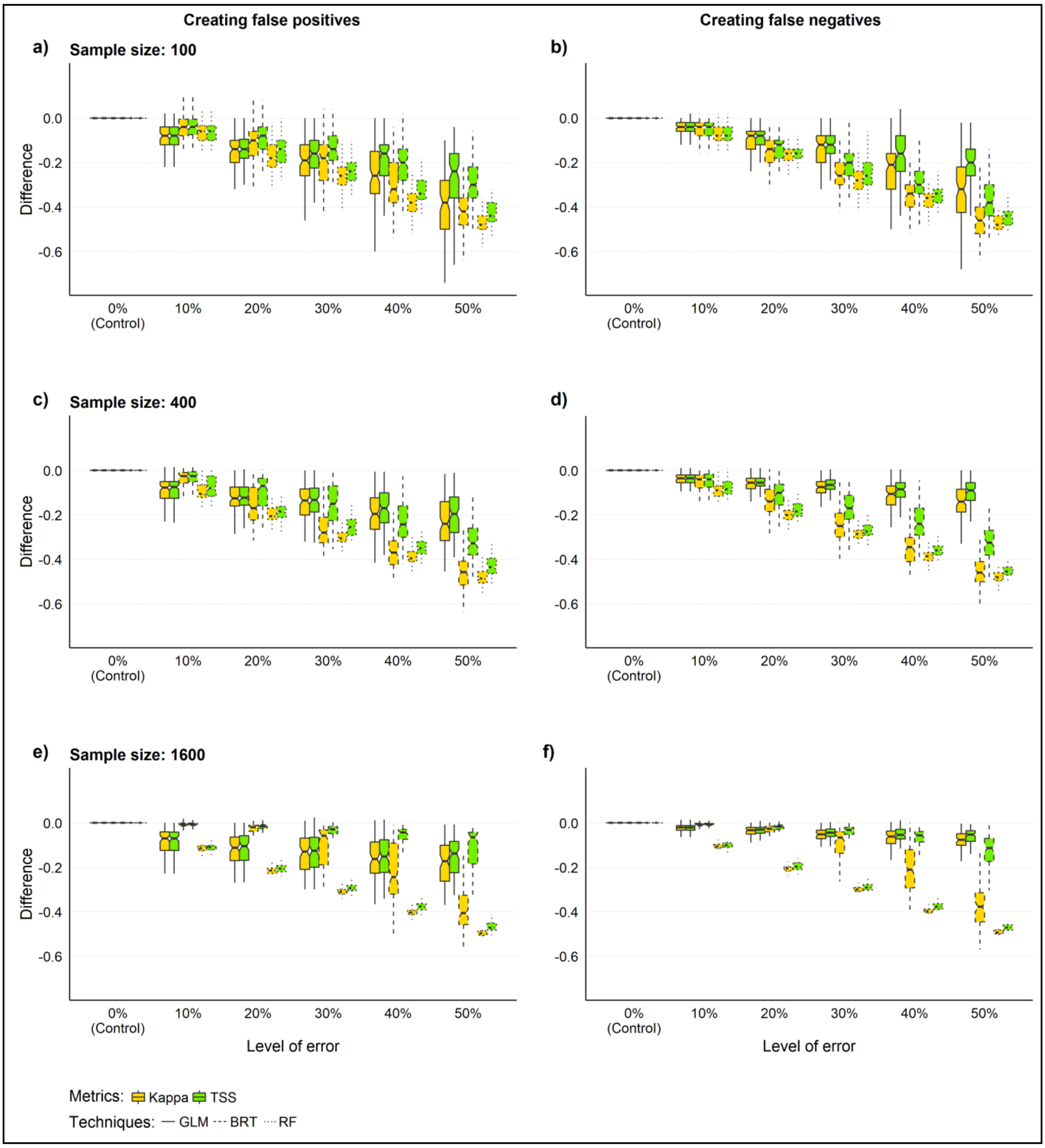


Figure A23 Observed difference of measured **PSc** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **EqualPrev** and sample size 100, 400 and 1600, for virtual species created using **GLM**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). PSc was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). See Fig. A22 for the explanation of the different analyses.

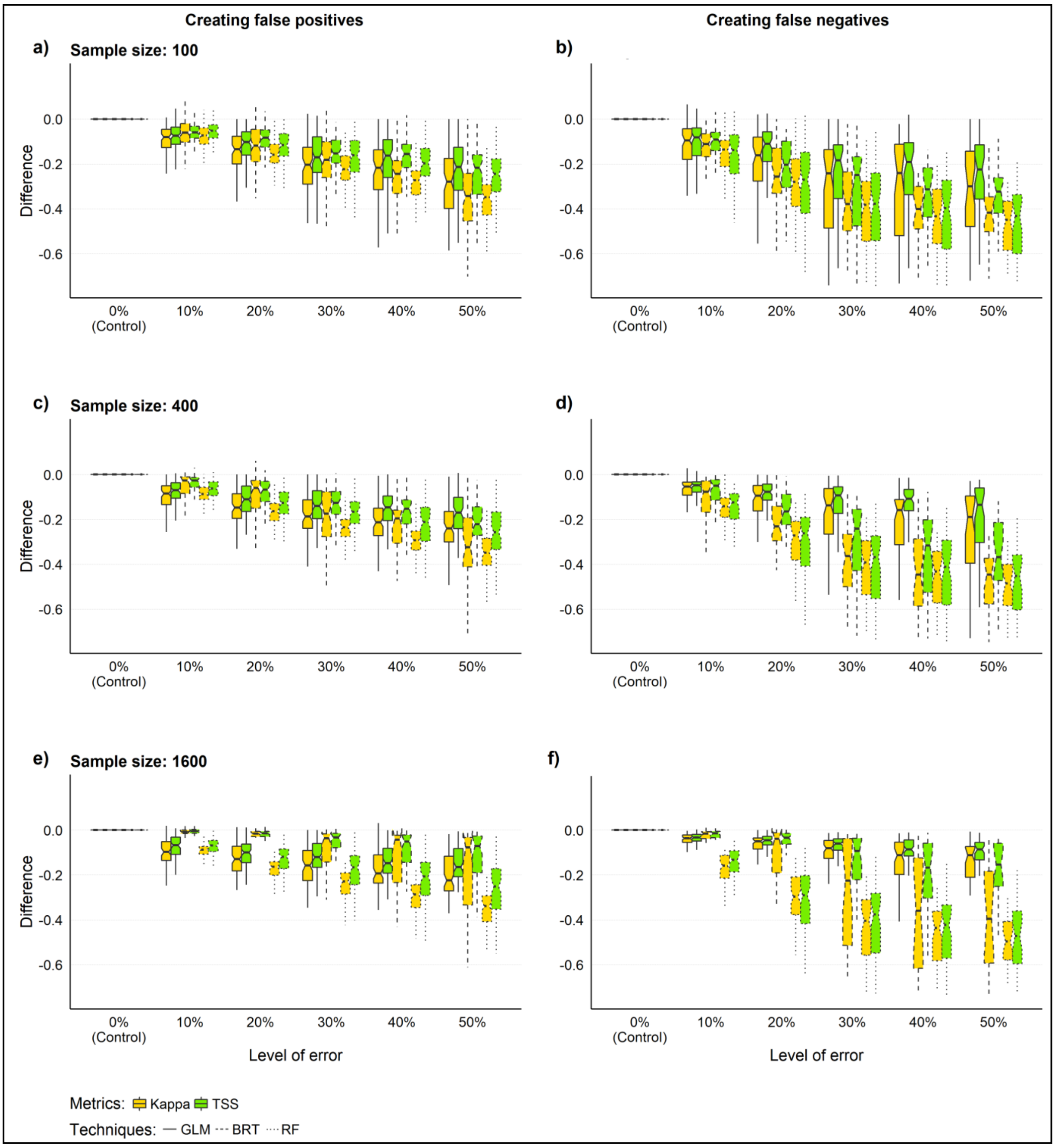


Figure A24 Observed difference of measured **PSc** between control (training data without errors added; 0% - sampled data) and degraded data (training data with errors added) models, under the sampling design **TruePrev** and sample size 100, 400 and 1600, for virtual species created using **GLM**. Errors were added to the occurrence dataset, creating either false positives (errors added only to absences; left column) or false negatives (errors added only to presences; right column). PSc was measured using Kappa (gold) and TSS (light green). For each level of error, three sets with three plots are observed, corresponding to models fitted using either GLMs (solid plots), BRTs (dashed plots) or RFs (dotted plots). See Fig. A22 for the explanation of the different analyses.

Supplementary material of chapter 1.2

Appendix A

Table A.1. Community evaluation metrics used in this study (see Di Cola et al., 2017 for details)

Metric	Definition	Description	References
Species Richness Error	$SRe = n_{pred} - n_{obs}$	Difference between predicted and observed species richness	(Pottier et al., 2013)
Prediction Success	$PredSuc = \frac{(TP + TA)}{N}$	Proportion of species correctly predicted as present or absent	(Pottier et al., 2013)
Sensitivity	$Sens = \frac{TP}{TP + FA}$	The proportion of species correctly predicted as present	(Pottier et al., 2013)
Specificity	$Spec = \frac{TA}{TA + FP}$	The proportion of species correctly predicted as absent	(Pottier et al., 2013)
Over-prediction	$OverPred = \frac{FP}{FP + TA}$	The proportion of species predicted as present but not observed among the species predicted as present	(Pottier et al., 2013)
Under-prediction	$UnderPred = \frac{FA}{TP + FA}$	The proportion of species predicted as absent but observed among the species observed as present	(Pottier et al., 2013)
Community TSS	$TSS = Sens + Spec - 1$	Same as TSS but measured for a site across all species, rather than for a species across all sites	(Pottier et al., 2013)
Community Kappa	$K = \frac{Acc - p_e}{1 - p_e}$	Same as Kappa but measured for a site across all species, rather than for a species across all sites	(Pottier et al., 2013)
Sørensen	$S = \frac{2 * TP}{2 * TP + FP + FA}$	Similarity index (compares similarity between observed and predicted assemblages)	(Sørensen, 1948)

Original dataset information

To obtain the full information about the original plant data, two datasets where combined (Dubuis et al., 2011; Pottier et al., 2013). The first one with 912 vegetation plots of 4 m^2 were selected following a random-stratified sampling design and the presence-absence of each species was recorded in each plot. Only vascular species in open and non-woody vegetation were sampled. This dataset consisted of 795 species. The second dataset with 3076 vegetation plots, from a grid of 400 m over all of the study area; therefore, a point was recorded every 400 m. If a point was falling into a forest, a field sampling was made. The field sampling was done in a circle with center at the coordinates of the point and with a radius of 10 m. All vascular plant species were recorded. This dataset consisted of 667 species. The final dataset of plants has 3967 plots composed by 1088 species (627 after removing the species with less than 30 occurrences).

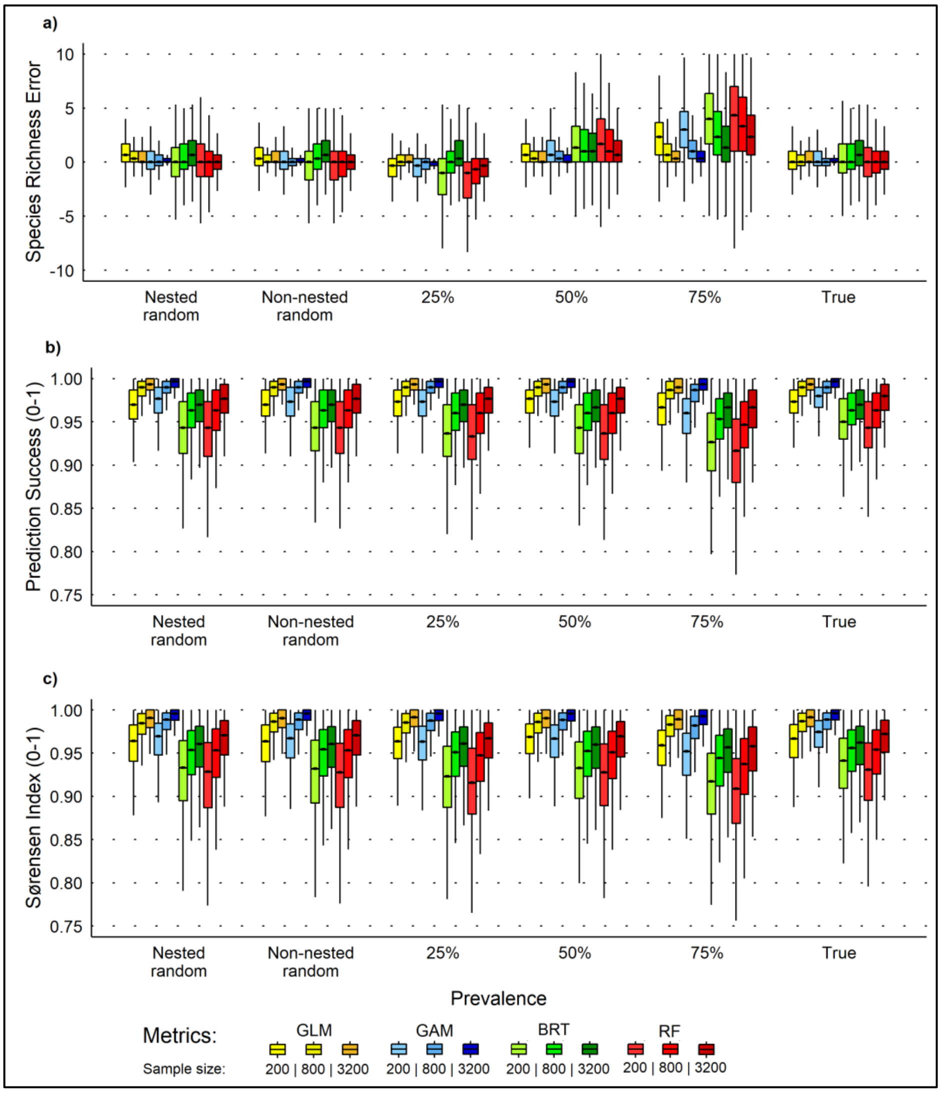


Fig. A.1 Boxplots of different indices of community prediction (S-SDM) accuracy (i.e. species richness error, prediction success and Sørensen) for all the simulated species and for all the sampling strategies (based on plots (nested or not) or prevalence (25%, 50%, 75% or true) sampling; in abscissa). Each box shows the variation across all virtual species in a random subset of the study area (100 000 plots) for the binary predictions obtained using ROC as thresholding technique, averaged from the three sampling turns. For each prevalence sampling, four sets of three boxplots are displayed, corresponding to models fitted using either GLMs (yellow), GAMs (blue), BRTs (green) or RFs (red), with increasing values of sample size (200, 800 and 3200).

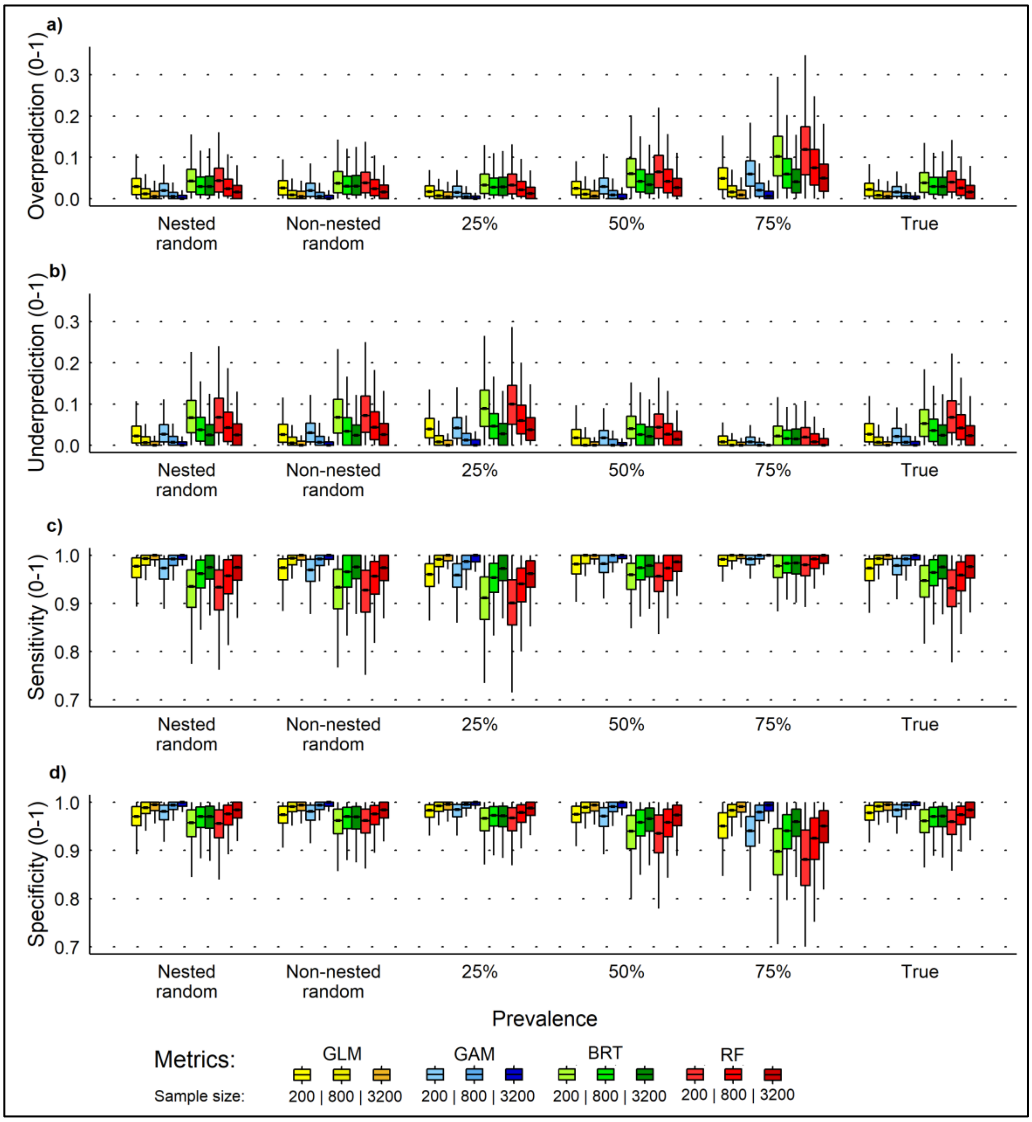


Fig. A.2 Boxplots of different indices of community prediction (S-SDM) accuracy (i.e. over and underprediction, sensitivity and specificity) for all the simulated species and for all the sampling strategies (based on plots (nested or not) or prevalence (25%, 50%, 75% or true) sampling; in abscissa). Each box shows the variation across all virtual species in a random subset of the study area (100 000 plots) for the binary predictions obtained using ROC as thresholding technique, averaged from the three sampling turns. For each prevalence sampling, four sets of three boxplots are displayed, corresponding to models fitted using either GLMs (yellow), GAMs (blue), BRTs (green) or RFs (red), with increasing values of sample size (200, 800 and 3200).

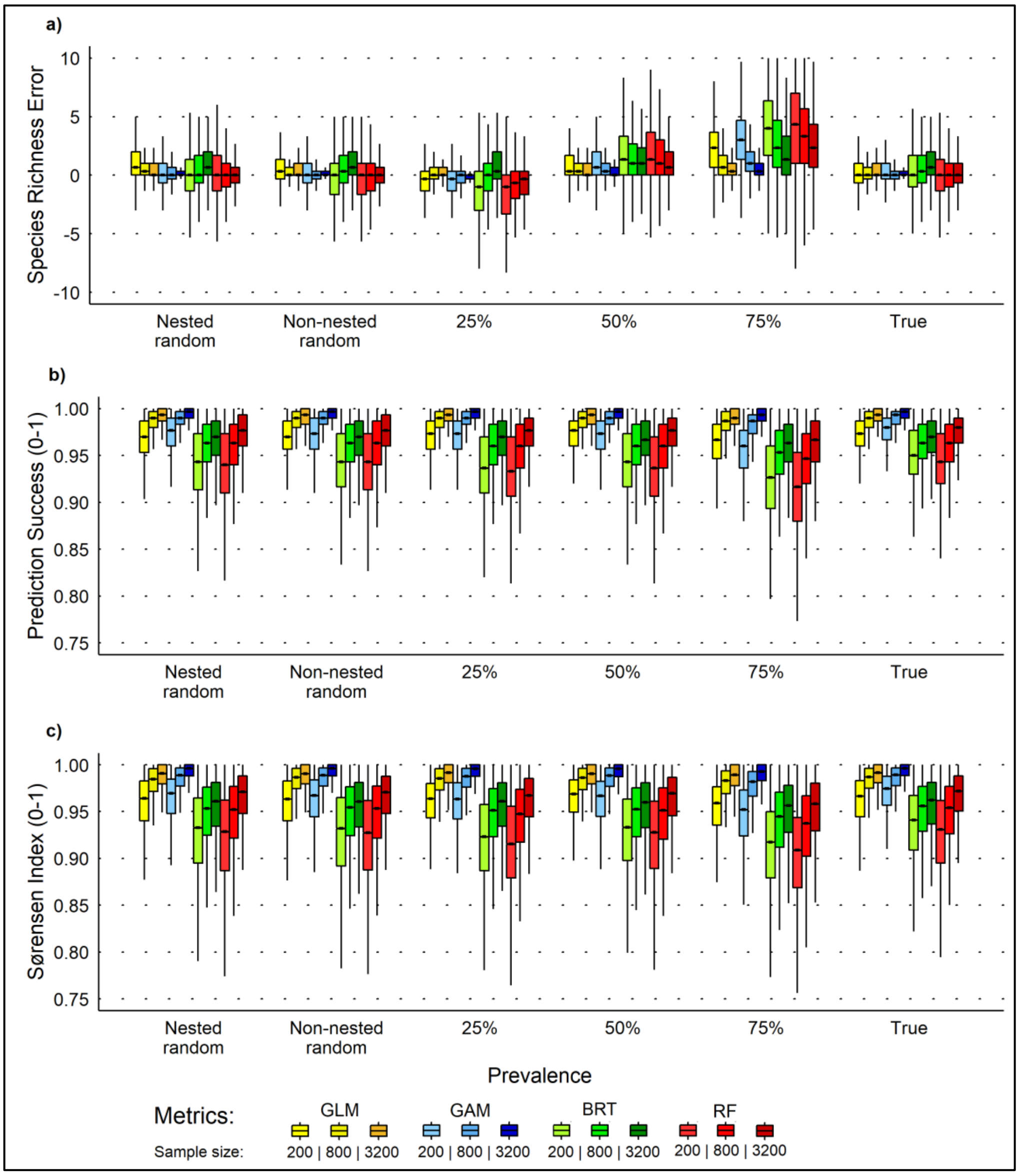


Fig. A.3 Boxplots of different indices of community prediction (S-SDM) accuracy (i.e. species richness error, prediction success and Sørensen) for all the simulated species and for all the sampling strategies (based on plots (nested or not) or prevalence (25%, 50%, 75% or true) sampling; in abscissa). Each box shows the variation across all virtual species in a random subset of the study area (100 000 plots) for the binary predictions obtained using MaxTSS as thresholding technique, averaged from the three sampling turns. For each prevalence sampling, four sets of three boxplots are displayed, corresponding to models fitted using either GLMs (yellow), GAMs (blue), BRTs (green) or RFs (red), with increasing values of sample size (200, 800 and 3200).

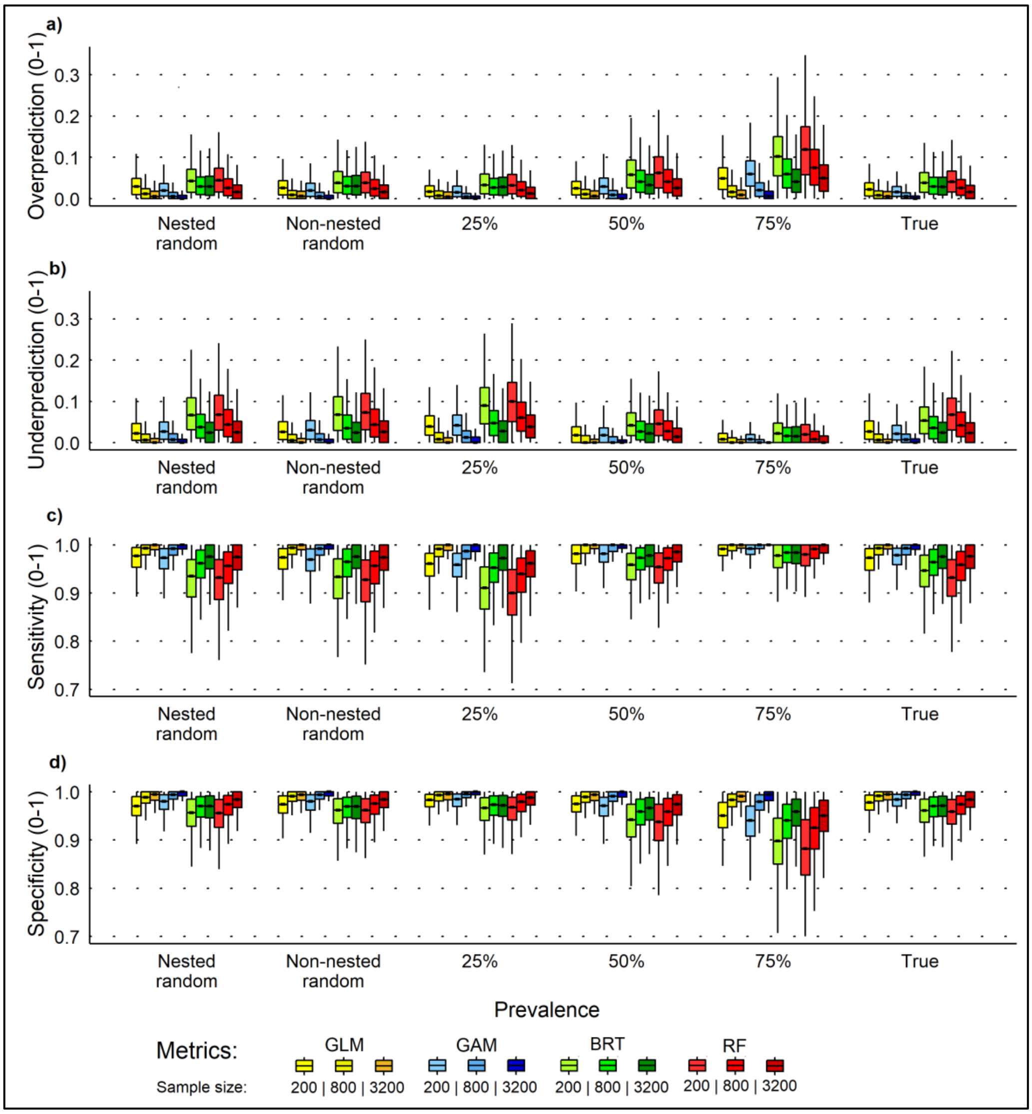


Fig. A.4 Boxplots of different indices of community prediction (S-SDM) accuracy (i.e. over and underprediction, sensitivity and specificity) for all the simulated species and for all the sampling strategies (based on plots (nested or not) or prevalence (25%, 50%, 75% or true) sampling; in abscissa). Each box shows the variation across all virtual species in a random subset of the study area (100 000 plots) for the binary predictions obtained using MaxTSS as thresholding technique, averaged from the three sampling turns. For each prevalence sampling, four sets of three boxplots are displayed, corresponding to models fitted using either GLMs (yellow), GAMs (blue), BRTs (green) or RFs (red), with increasing values of sample size (200, 800 and 3200).

Supplementary material of chapter 1.3

Appendix A

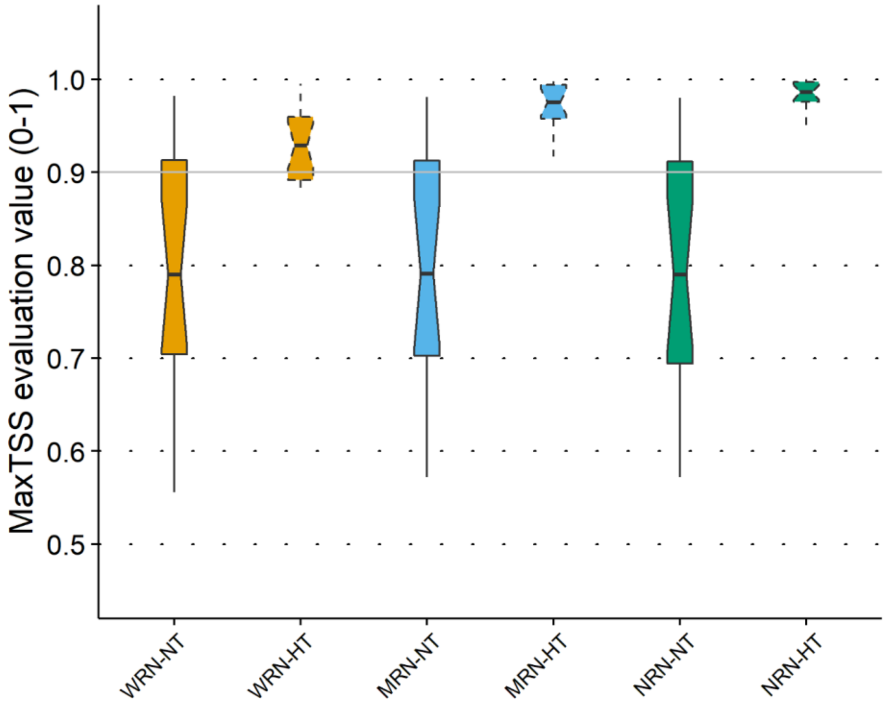


Figure S1 Evaluation values (MaxTSS) of individual species distribution models for each simulated realised niche (WRN - wide realised niche; MRN - medium realised niche; and NRN - narrow realised niche) and training dataset (NT: non-truncated and HT: highly truncated), for models fitted using the same set environmental variables for all the species.

Table S1 Community evaluation metrics used in this study (see Di Cola *et al.*, 2017 for details)

Metric	Definition	Description	References
Jaccard	$J = \dfrac{TP}{TP + FP + FA}$	Similarity index (compares similarity between observed and predicted assemblages)	(Jaccard, 1912)
Sørensen	$S = \dfrac{2 * TP}{2 * TP + FP + FA}$		(Sorensen, 1948)

n_{pred} =Number of species predicted
n_{obs} = Number of species observed
N = Number of events
TP = Correctly predicted present species
TA = Correctly predicted absent species
FP = Falsely predicted present species
FA = Falsely predicted absent species

Table S2 List of all virtual species modelled in the analysis, the evaluation values (MaxTSS) for the final ensemble model of each species, presented for each simulated realised niche (WRN - wide realised niche; MRN - medium realised niche; and NRN - narrow realised niche) and training dataset used to calibrate the models (NT: non-truncated and HT: highly truncated), and their prevalence for each simulated realised niche (WRN and NRN).

	Evaluation value (MaxTSS)						**Virtual Species Prevalence**	
Species	**WRN-NT**	**WRN-HT**	**MRN-NT**	**MRN-HT**	**NRN-NT**	**NRN-HT**	**WRN**	**NFN**
VS1	0.698	0.777	0.975	0.974	0.988	0.989	0.265	0.01
VS2	0.941	0.952	0.991	0.996	0.996	0.999	0.046	0.004
VS3	0.916	0.933	0.985	0.99	0.993	0.997	0.068	0.005
VS4	0.79	0.901	0.974	0.982	0.982	0.995	0.162	0.014
VS5	0.955	0.962	0.996	0.999	0.998	0.981	0.032	0.002
VS6	0.913	0.982	0.968	0.973	0.975	-	0.069	0.019
VS7	0.784	0.894	0.985	0.989	0.997	0.999	0.18	0.003
VS8	0.726	0.781	0.88	0.885	0.91	0.915	0.239	0.078
VS9	0.69	0.79	0.889	0.899	0.931	0.932	0.26	0.063
VS10	0.979	0.982	0.996	0.995	0.997	0.997	0.017	0.002
VS11	0.37	0.79	0.859	0.893	0.922	0.938	0.603	0.062
VS12	0.812	0.856	0.93	0.943	0.95	0.953	0.158	0.041
VS13	0.981	0.98	0.993	0.994	0.996	0.996	0.016	0.004
VS14	0.36	0.662	0.773	0.851	0.895	0.896	0.629	0.082
VS15	0.591	0.759	0.937	0.947	0.96	0.965	0.322	0.032
VS16	0.899	0.914	0.97	0.973	0.98	0.979	0.083	0.017
VS17	0.978	0.986	0.995	0.999	0.997	0.999	0.017	0.003
VS18	0.778	0.857	0.989	0.996	0.995	1	0.207	0.004
VS19	0.722	0.834	0.903	0.908	0.929	0.936	0.236	0.059
VS20	0.729	0.812	0.925	0.951	0.955	0.98	0.222	0.033
Mean	**0.78**	**0.87**	**0.95**	**0.96**	**0.97**	**0.97**	**0.19**	**0.03**
Standard Dev.	0.18	0.093	0.059	0.046	0.033	0.033	0.17	0.03
Minimum	0.36	0.662	0.773	0.851	0.895	0.896	0.02	0.002
Maximum	0.981	0.986	0.996	0.999	0.998	1	0.63	0.08

Additional references

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. Ecography, 40, 774-787.

Jaccard, P. (1912) The distribution of the flora in the alpine zone. New Phytol,

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

Supplementary material of chapter 1.4

Appendix S1

Table S1 Community evaluation metrics used in this study (see Di Cola et al., 2017 for details)

Metric	Definition	Description	References
Species Richness Error	$SRe = n_{pred} - n_{obs}$	Difference between predicted and observed species richness	(Pottier et al., 2013)
Assemblage prediction Success	$PredSuc = \frac{(TP + TA)}{N}$	Proportion of species correctly predicted as present or absent	(Pottier et al., 2013)
Sørensen	$S = \frac{2 * TP}{2 * TP + FP + FA}$	Similarity indices (compares similarity between observed and predicted assemblages)	(Sørensen, 1948)
Jaccard	$J = \frac{TP}{TP + FP + FA}$		(Jaccard, 1912)

n_{pred} =Number of species predicted
n_{obs} = Number of species observed
N = Number of events
TP = Correctly predicted present species
TA = Correctly predicted absent species
FP = Falsely predicted present species
FA = Falsely predicted absent species

Table S2 Number of modelled species in each modelling strategy (DwSc-Downscaling model; FnSc-Fine model; TDwSc- Truncated Downscaling model; TFnSc- Truncated Downscaling model; Mxd- Mixed model; Hier1- Hierarchical model 1; Hier2- Hierarchical model 2) and for each set of the sampled calibration data used to fit the models: presence/absence data (PA), an equal number of presence/pseudo-absence data (POequal) and presence/pseudo-absence data with a random set of 10.000 pseudo-absences (PO10M).

Model	Data used in calibration	Number of species modelled
DwSc model	PA	100
	POequal	100
	PO10M	100
FnSc model	PA	100
	POequal	100
	PO10M	100
TDwSc model	PA	**53**
	POequal	98
	PO10M	98
TFnSc model	PA	95
	POequal	98
	PO10M	98
Mxd model	PA	88
	POequal	100
	PO10M	100
Hier1 model	POequal	100
	PO10M	100
Hier2 model	POequal	100
	PO10M	100

Additional references

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. Ecography 40, 774-787.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.P., Guisan, A., 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. Diversity and Distributions 17, 1122-1131.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P., Guisan, A., Field, R., 2013. The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. Global Ecology and Biogeography 22, 52-63.

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

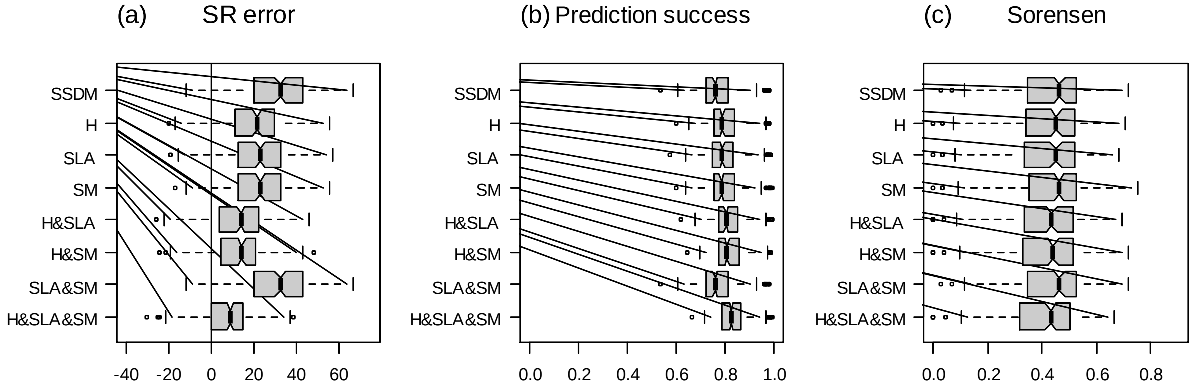
Supplementary material of chapter 2.1

Appendix S1 Assemblage evaluation metrics and supplementary results.

Table S1 Confusion matrix used to compute the assemblage evaluation metrics.

		observed	
		0	**1**
predicted	**0**	*TN*	*FN*
	1	*FP*	*TP*

FIGURE S1 Results from true skill statistic (TSS) thresholding criterion: the boxplots compare results from the ‘trait range’ rule test for single traits and all their combinations when using TSS to binarize the SDM predictions. The metrics utilized in the comparison are: (a) species richness error, i.e. predicted SR – observed SR (first column), (b) prediction success, i.e. sum of correctly predicted presences and absences divided by the total species number (second column), and (c) Sørensen index, i.e. a statistic used to compare the similarity of two samples (third column). Abbreviations: H, height; SLA, specific leaf area of the community; SM, seed mass.



Appendix S2 Evaluation results for SDMs and MEMs.

TABLE S2 Summary of the SDMs evaluation results.

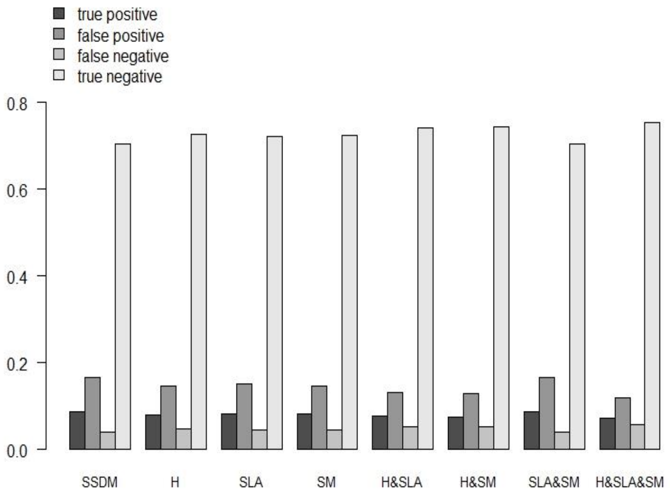
	GAM	**GBM**	**GLM**
AUC mean	0.803	0.783	0.799
AUC stdev	0.078	0.081	0.077
TSS mean	0.541	0.507	0.537
TSS stdev	0.142	0.137	0.142

TABLE S3 Values of Spearman correlation test between observed and predicted values of trait percentiles.

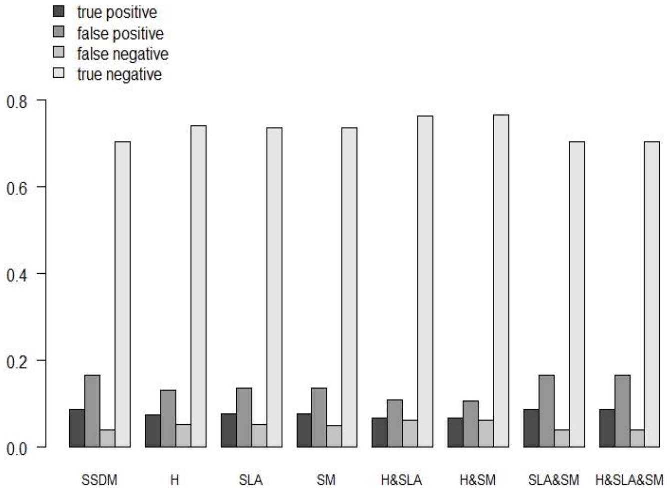
Percentile	**Trait**	**ρ**
1st	H	0.711
1st	SLA	0.759
1st	SM	0.152
99th	H	0.859
99th	SLA	0.584
99th	SM	0.514
5th	H	0.825
5th	SLA	0.803
5th	SM	0.350
95th	H	0.887
95th	SLA	0.652
95th	SM	0.528
10th	H	0.848
10th	SLA	0.814
10th	SM	0.550
90th	H	0.867
90th	SLA	0.677
90th	SM	0.645

Appendix S3 Comparison of the assemblage predictions coming from the application of the trait range rule with three pairs of percentiles. Abbreviations: SSDM, sum of binary SDMs; H, height; SLA, specific leaf area of the community; SM, seed mass.

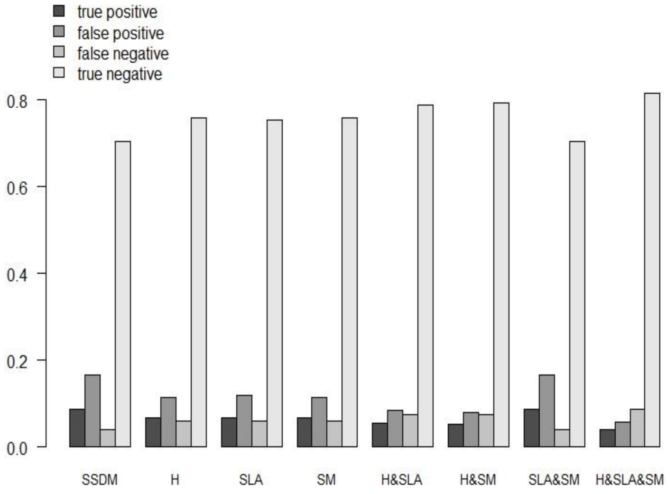
1st–99th



5th–95th



10th–90th



Supplementary material of chapter 2.2

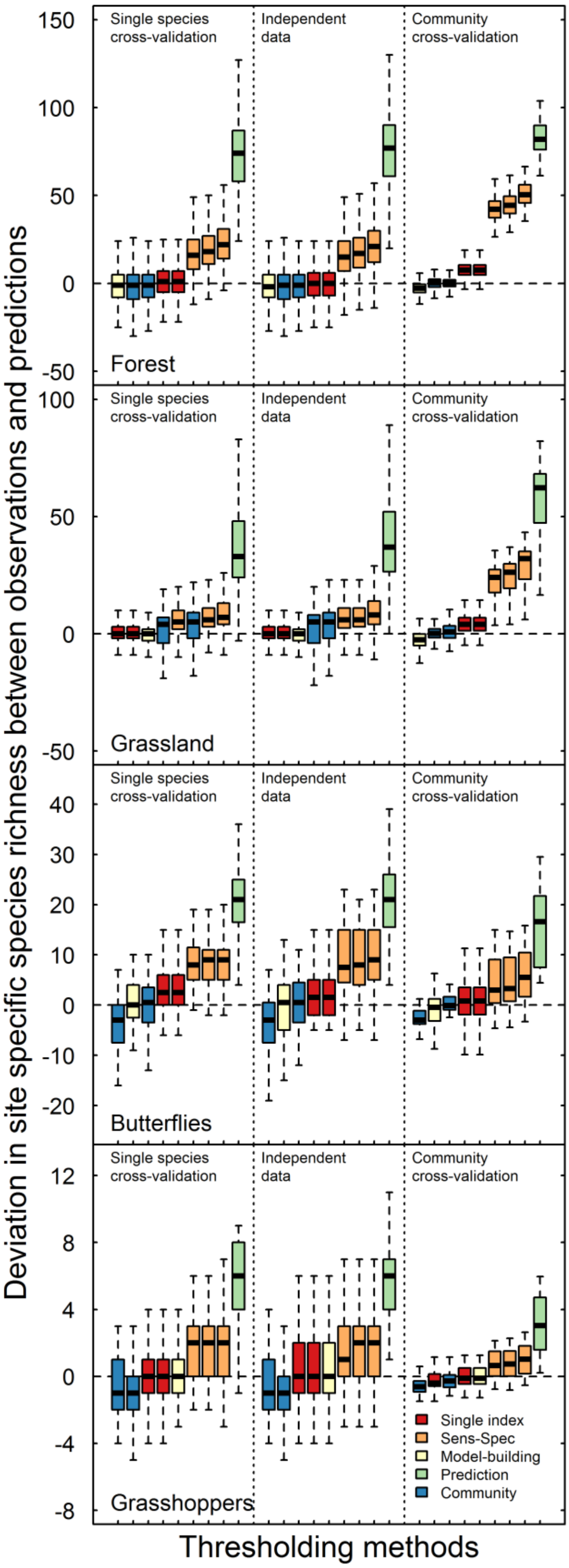


Figure S1 Deviation in site specific species richness between observations and predictions for the four different datasets (top to bottom) and the three different modelling pathways (left to right). The boxplots are sorted by the median and the colours indicate the different thresholding techniques. The line in the box indicates the median, boxes range from the 25th to the 75th percentile and the whiskers indicate ± 2 standard deviations. For details on the method used within each threshold group see Table S1.

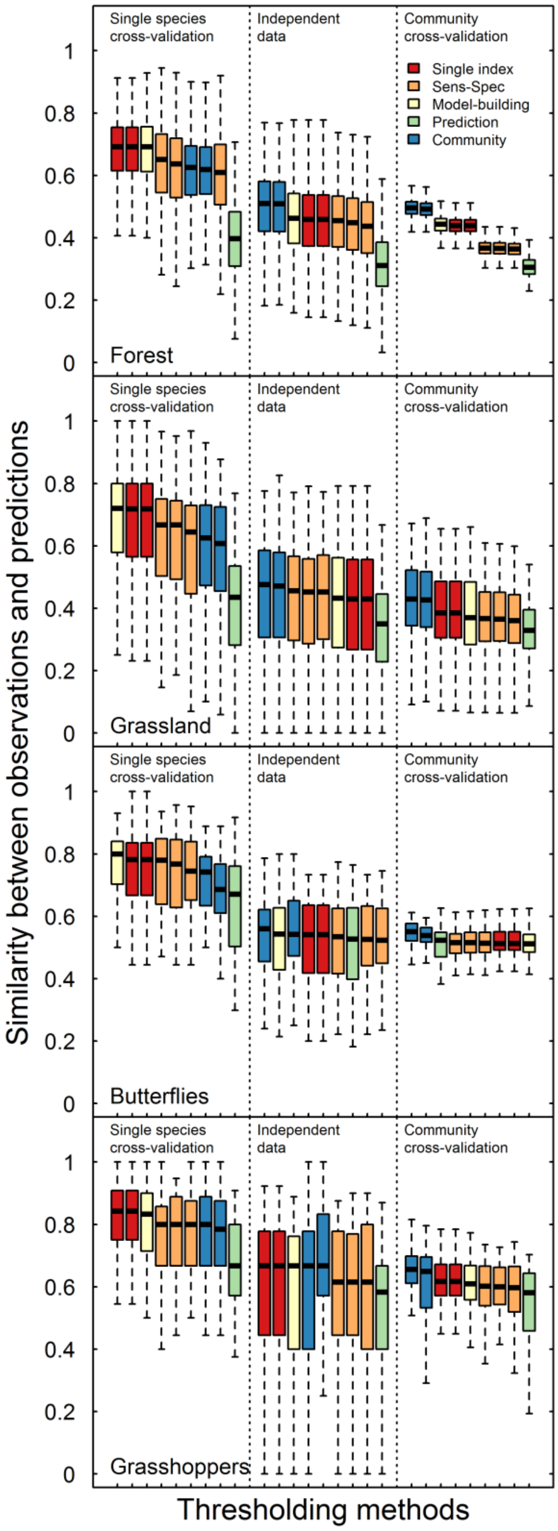


Figure S2 Sørensen similarity between observations and predictions for the four different datasets (top to bottom) and the three different modelling pathways (left to right). The boxplots are sorted by the median and the colours indicate the different thresholding techniques. The line in the box indicates the median, boxes range from the 25th to the 75th percentile and the whiskers indicate ± 2 standard deviations. For details on the method used within each threshold group see Table S1.

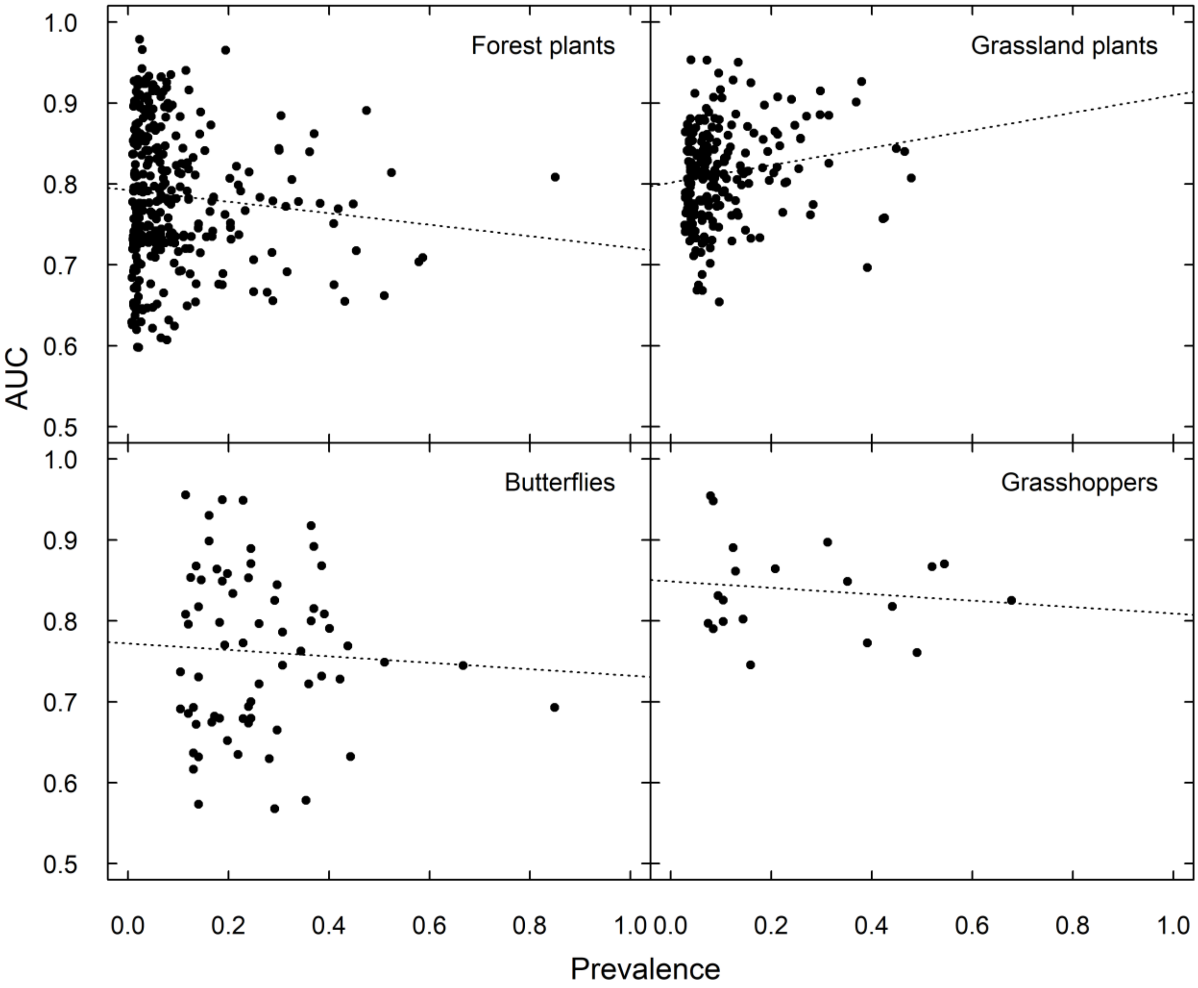


Figure S3 The relationship of the prevalence of a species (i.e., percentage of sites inhabited) to the performance of the SDMs (i.e., as measured by AUC) for the four studied data sets (taxa).

Table S1 Description of the ten thresholding methods based on Liu *et al.* (2005) and Nenzen and Araujo (2011).

Approach	**Accronym**	**Definition**	**Reference**
Single index-based approaches			
1. Kappa maximization approach	Max.Kappa	Kappa statistic is maximized	(Huntley *et al.* 1995; Guisan, Theurillat & Kienast 1998)
2. Maximum commission error	MCE05	Allowed a maximum commission error of 5%	(Mateo et al. 2012)
Sensitivity and specificity-combined approaches			
3. TSS maximization approach	Max.TSS	TSS statistic is maximized	(Allouche, Tsoar & Kadmon 2006)
4. Sensitivity-specificity equality approach	SensSpec	Difference of sens-spec is minimized	(Cantor *et al.* 1999)
5. ROC plot-based approach	Opt.ROC	ROC statistic is maximized	(Cantor *et al.* 1999)
Model-building data-only-based approach			
6. Prevalence approach	Preval	Prevalence of the calibration data	(Cramer 2003)
Predicted probability-based approaches			
7. Average probability approach	AvgProb	Taking the average predicted probability of the model-building data as threshold	(Cramer 2003)
Community based approaches			
8. pS-SDM+PRR	pS-SDM+PRR	Probability stacked SDM	(Dubuis *et al.* 2013)
9. MEM+PRR	MEM+PRR	Macroecological model for SR	(Guisan & Rahbek 2011)

Table S2 Community evaluation metrics used in this study.

Metric	Definition
Species richness	
Deviation in species richness	$Dev.SPR = n_{pred} - n_{obs}$
Prediction success	
Sensitivity	$Sens = \frac{TP}{TP + FA}$
Specificity	$Spec = \frac{TA}{TA + FP}$
Community accuracy	$Acc = \frac{TP + TA}{N}$
Community TSS	$TSS = Sens + Spec - 1$
Community Kappa	$K = \frac{Acc - \ p_e}{1 - p_e}$
Community composition	
Sørensen	$S = \frac{2 * TP}{2 * TP + FP + FA}$

n_{pred} =Number of species predicted
n_{obs} = Number of species observed
N = Number of events
TP = Correctly predicted present species
TA = Correctly predicted absent species
FP = Falsely predicted present species
FA = Falsely predicted absent species
$p_e = \frac{(TP+FA)(TP+FP)+(TA+FP)(TA+FA)}{N^2}$

Table S3 Evaluation scores of individual SDMs by TSS (A), Kappa (B), PCC (C), Sensitivity (D) and Specificity (E) for the three community evaluation approaches and four datasets. SSCV = Single species cross-validation, ID = Independent data, CCV = Community cross-validation, FO = Forest plants, GL = Grassland plants, BF = Butterflies, GH = Grasshoppers.

(A) TSS

Thresholding Approach	**SSCV**				**ID**				**CCV**			
	FO	**GL**	**BF**	**GH**	**FO**	**GL**	**BF**	**GH**	**FO**	**GL**	**BF**	**GH**
Max.Kappa	0.2 ± 0.14	0.27 ± 0.2	0.31 ± 0.23	0.42 ± 0.17	0.21 ± 0.17	0.25 ± 0.21	0.3 ± 0.22	0.43 ± 0.26	0.23 ± 0.13	0.28 ± 0.14	0.31 ± 0.18	0.37 ± 0.15
MCE05	0.3 ± 0.17	0.32 ± 0.17	0.27 ± 0.18	0.43 ± 0.16	0.28 ± 0.17	0.34 ± 0.17	0.27 ± 0.23	0.45 ± 0.21	0.25 ± 0.17	0.34 ± 0.12	0.31 ± 0.17	0.42 ± 0.12
Max.TSS	0.35 ± 0.2	0.38 ± 0.14	0.34 ± 0.24	0.47 ± 0.12	0.33 ± 0.22	0.38 ± 0.21	0.34 ± 0.23	0.5 ± 0.26	0.35 ± 0.14	0.39 ± 0.11	0.35 ± 0.18	0.44 ± 0.12
SensSpec	0.32 ± 0.16	0.36 ± 0.14	0.33 ± 0.18	0.51 ± 0.18	0.31 ± 0.2	0.37 ± 0.21	0.34 ± 0.24	0.51 ± 0.26	0.35 ± 0.13	0.38 ± 0.11	0.35 ± 0.18	0.45 ± 0.11
Opt.ROC	0.34 ± 0.21	0.36 ± 0.19	0.33 ± 0.17	0.44 ± 0.26	0.32 ± 0.21	0.37 ± 0.21	0.34 ± 0.23	0.49 ± 0.25	0.35 ± 0.13	0.38 ± 0.11	0.35 ± 0.18	0.44 ± 0.12
Preval	0.18 ± 0.15	0.27 ± 0.15	0.3 ± 0.19	0.41 ± 0.23	0.2 ± 0.16	0.26 ± 0.2	0.31 ± 0.22	0.4 ± 0.23	0.18 ± 0.14	0.23 ± 0.15	0.3 ± 0.17	0.37 ± 0.17
AvgProb	0.43 ± 0.16	0.5 ± 0.12	0.41 ± 0.21	0.55 ± 0.14	0.47 ± 0.16	0.53 ± 0.13	0.38 ± 0.23	0.56 ± 0.15	0.44 ± 0.16	0.49 ± 0.11	0.4 ± 0.18	0.54 ± 0.14
pS-SDM+PRR	0.12 ± 0.19	0.17 ± 0.24	0.28 ± 0.24	0.28 ± 0.22	0.14 ± 0.18	0.2 ± 0.24	0.24 ± 0.24	0.29 ± 0.28	0.14 ± 0.17	0.19 ± 0.21	0.27 ± 0.23	0.27 ± 0.22
MEM+PRR	0.16 ± 0.17	0.2 ± 0.23	0.25 ± 0.24	0.3 ± 0.24	0.14 ± 0.18	0.2 ± 0.24	0.22 ± 0.24	0.32 ± 0.28	0.14 ± 0.17	0.2 ± 0.22	0.25 ± 0.22	0.3 ± 0.22

(B) KAPPA

Thresholding Approach	**SSCV**				**ID**				**CCV**			
	FO	**GL**	**BF**	**GH**	**FO**	**GL**	**BF**	**GH**	**FO**	**GL**	**BF**	**GH**
Max.Kappa	0.2 ± 0.14	0.24 ± 0.18	0.28 ± 0.22	0.35 ± 0.19	0.21 ± 0.15	0.24 ± 0.19	0.29 ± 0.22	0.42 ± 0.25	0.2 ± 0.14	0.24 ± 0.15	0.29 ± 0.18	0.36 ± 0.15
MCE05	0.21 ± 0.13	0.23 ± 0.15	0.24 ± 0.22	0.31 ± 0.16	0.21 ± 0.13	0.28 ± 0.15	0.27 ± 0.22	0.41 ± 0.18	0.11 ± 0.12	0.16 ± 0.14	0.24 ± 0.17	0.32 ± 0.16
Max.TSS	0.19 ± 0.12	0.27 ± 0.17	0.3 ± 0.19	0.4 ± 0.14	0.21 ± 0.14	0.27 ± 0.16	0.3 ± 0.21	0.41 ± 0.23	0.17 ± 0.13	0.22 ± 0.15	0.3 ± 0.17	0.36 ± 0.15
SensSpec	0.21 ± 0.14	0.21 ± 0.17	0.29 ± 0.19	0.41 ± 0.17	0.22 ± 0.14	0.28 ± 0.17	0.3 ± 0.21	0.42 ± 0.23	0.17 ± 0.13	0.23 ± 0.15	0.3 ± 0.17	0.37 ± 0.14
Opt.ROC	0.15 ± 0.15	0.22 ± 0.14	0.32 ± 0.18	0.43 ± 0.17	0.22 ± 0.14	0.27 ± 0.16	0.3 ± 0.21	0.42 ± 0.23	0.17 ± 0.13	0.23 ± 0.15	0.3 ± 0.18	0.37 ± 0.14
Preval	0.2 ± 0.13	0.22 ± 0.18	0.3 ± 0.2	0.34 ± 0.17	0.21 ± 0.15	0.25 ± 0.18	0.3 ± 0.22	0.39 ± 0.22	0.19 ± 0.14	0.23 ± 0.15	0.3 ± 0.17	0.36 ± 0.16
AvgProb	0.17 ± 0.12	0.21 ± 0.14	0.26 ± 0.17	0.38 ± 0.15	0.17 ± 0.13	0.22 ± 0.15	0.26 ± 0.19	0.37 ± 0.18	0.16 ± 0.13	0.22 ± 0.15	0.29 ± 0.16	0.37 ± 0.16
pS-SDM+PRR	0.15 ± 0.16	0.17 ± 0.18	0.26 ± 0.23	0.28 ± 0.25	0.14 ± 0.17	0.18 ± 0.2	0.22 ± 0.23	0.29 ± 0.29	0.14 ± 0.16	0.17 ± 0.19	0.26 ± 0.21	0.28 ± 0.22
MEM+PRR	0.14 ± 0.16	0.17 ± 0.21	0.24 ± 0.22	0.32 ± 0.26	0.14 ± 0.17	0.18 ± 0.2	0.22 ± 0.23	0.34 ± 0.29	0.15 ± 0.16	0.19 ± 0.19	0.25 ± 0.21	0.32 ± 0.22

C) Percentage correct classified (PCC)

Thresholding	**SSCV**				**ID**				**CCV**			
Approach	**FO**	**GL**	**BF**	**GH**	**FO**	**GL**	**BF**	**GH**	**FO**	**GL**	**BF**	**GH**
Max.Kappa	0.91 ± 0.09	0.9 ± 0.07	0.78 ± 0.1	0.83 ± 0.08	0.9 ± 0.09	0.89 ± 0.07	0.77 ± 0.09	0.83 ± 0.1	0.88 ± 0.09	0.87 ± 0.06	0.76 ± 0.08	0.82 ± 0.07
MCE05	0.85 ± 0.15	0.79 ± 0.07	0.66 ± 0.11	0.82 ± 0.08	0.88 ± 0.07	0.87 ± 0.05	0.77 ± 0.09	0.82 ± 0.08	0.59 ± 0.14	0.68 ± 0.09	0.64 ± 0.1	0.73 ± 0.08
Max.TSS	0.85 ± 0.08	0.84 ± 0.07	0.73 ± 0.11	0.83 ± 0.09	0.86 ± 0.1	0.85 ± 0.08	0.73 ± 0.11	0.81 ± 0.09	0.77 ± 0.07	0.79 ± 0.05	0.72 ± 0.08	0.79 ± 0.06
SensSpec	0.79 ± 0.1	0.84 ± 0.04	0.73 ± 0.09	0.81 ± 0.08	0.87 ± 0.09	0.86 ± 0.07	0.73 ± 0.1	0.81 ± 0.08	0.79 ± 0.07	0.81 ± 0.05	0.73 ± 0.07	0.8 ± 0.06
Opt.ROC	0.86 ± 0.08	0.86 ± 0.06	0.72 ± 0.11	0.82 ± 0.05	0.87 ± 0.1	0.86 ± 0.07	0.74 ± 0.1	0.81 ± 0.09	0.79 ± 0.07	0.81 ± 0.05	0.73 ± 0.08	0.8 ± 0.06
Preval	0.92 ± 0.08	0.91 ± 0.05	0.79 ± 0.08	0.83 ± 0.06	0.9 ± 0.08	0.89 ± 0.06	0.77 ± 0.1	0.83 ± 0.09	0.9 ± 0.08	0.89 ± 0.06	0.77 ± 0.08	0.82 ± 0.07
AvgProb	0.71 ± 0.07	0.67 ± 0.08	0.64 ± 0.11	0.74 ± 0.08	0.69 ± 0.08	0.69 ± 0.07	0.64 ± 0.11	0.73 ± 0.08	0.69 ± 0.07	0.69 ± 0.07	0.66 ± 0.09	0.73 ± 0.07
pS-SDM+PRR	0.93 ± 0.09	0.88 ± 0.07	0.76 ± 0.08	0.86 ± 0.09	0.91 ± 0.1	0.89 ± 0.08	0.77 ± 0.11	0.83 ± 0.1	0.91 ± 0.1	0.89 ± 0.08	0.78 ± 0.09	0.84 ± 0.1
MEM+PRR	0.92 ± 0.1	0.9 ± 0.09	0.8 ± 0.09	0.84 ± 0.09	0.91 ± 0.1	0.89 ± 0.09	0.79 ± 0.1	0.86 ± 0.08	0.91 ± 0.1	0.89 ± 0.08	0.79 ± 0.08	0.86 ± 0.09

D) Sensitivity

Thresholding	**SSCV**				**ID**				**CCV**			
Approach	**FO**	**GL**	**BF**	**GH**	**FO**	**GL**	**BF**	**GH**	**FO**	**GL**	**BF**	**GH**
Max.Kappa	0.31 ± 0.18	0.36 ± 0.24	0.47 ± 0.21	0.51 ± 0.26	0.27 ± 0.21	0.32 ± 0.25	0.46 ± 0.24	0.56 ± 0.26	0.32 ± 0.18	0.37 ± 0.18	0.5 ± 0.2	0.52 ± 0.22
MCE05	0.45 ± 0.13	0.61 ± 0.11	0.66 ± 0.18	0.73 ± 0.12	0.35 ± 0.17	0.42 ± 0.17	0.42 ± 0.23	0.57 ± 0.19	0.68 ± 0.05	0.67 ± 0.07	0.7 ± 0.09	0.71 ± 0.11
Max.TSS	0.44 ± 0.28	0.55 ± 0.15	0.62 ± 0.14	0.67 ± 0.15	0.46 ± 0.28	0.52 ± 0.27	0.59 ± 0.22	0.66 ± 0.25	0.57 ± 0.11	0.59 ± 0.11	0.61 ± 0.14	0.63 ± 0.15
SensSpec	0.55 ± 0.25	0.49 ± 0.15	0.58 ± 0.18	0.65 ± 0.13	0.42 ± 0.26	0.49 ± 0.25	0.58 ± 0.22	0.67 ± 0.25	0.54 ± 0.1	0.55 ± 0.11	0.58 ± 0.14	0.63 ± 0.12
Opt.ROC	0.44 ± 0.2	0.54 ± 0.14	0.57 ± 0.21	0.67 ± 0.14	0.44 ± 0.26	0.49 ± 0.26	0.58 ± 0.22	0.65 ± 0.24	0.55 ± 0.1	0.56 ± 0.11	0.59 ± 0.14	0.62 ± 0.14
Preval	0.28 ± 0.17	0.28 ± 0.21	0.46 ± 0.21	0.55 ± 0.25	0.26 ± 0.2	0.32 ± 0.24	0.49 ± 0.22	0.53 ± 0.25	0.24 ± 0.18	0.29 ± 0.18	0.47 ± 0.19	0.5 ± 0.23
AvgProb	0.76 ± 0.11	0.82 ± 0.1	0.77 ± 0.12	0.83 ± 0.1	0.79 ± 0.11	0.85 ± 0.1	0.78 ± 0.17	0.85 ± 0.13	0.76 ± 0.1	0.81 ± 0.06	0.78 ± 0.1	0.84 ± 0.07
pS-SDM+PRR	0.21 ± 0.28	0.26 ± 0.32	0.43 ± 0.36	0.42 ± 0.39	0.21 ± 0.28	0.28 ± 0.33	0.44 ± 0.36	0.45 ± 0.4	0.21 ± 0.27	0.26 ± 0.3	0.45 ± 0.33	0.41 ± 0.36
MEM+PRR	0.21 ± 0.28	0.3 ± 0.31	0.38 ± 0.36	0.41 ± 0.38	0.21 ± 0.29	0.28 ± 0.33	0.38 ± 0.35	0.43 ± 0.38	0.21 ± 0.27	0.28 ± 0.3	0.39 ± 0.32	0.42 ± 0.34

E) Specificity

Thresholding	SSCV				ID				CCV			
Approach	FO	GL	BF	GH	FO	GL	BF	GH	FO	GL	BF	GH
Max.Kappa	0.95 ± 0.09	0.91 ± 0.07	0.82 ± 0.12	0.87 ± 0.1	0.94 ± 0.07	0.93 ± 0.07	0.84 ± 0.12	0.88 ± 0.1	0.91 ± 0.09	0.93 ± 0.07	0.81 ± 0.12	0.85 ± 0.1
MCE05	0.87 ± 0.15	0.81 ± 0.04	0.78 ± 0.11	0.79 ± 0.09	0.93 ± 0.03	0.91 ± 0.03	0.86 ± 0.1	0.88 ± 0.08	0.57 ± 0.15	0.91 ± 0.03	0.61 ± 0.13	0.7 ± 0.1
Max.TSS	0.88 ± 0.09	0.79 ± 0.07	0.73 ± 0.1	0.83 ± 0.09	0.87 ± 0.1	0.86 ± 0.09	0.75 ± 0.13	0.84 ± 0.09	0.78 ± 0.08	0.86 ± 0.09	0.74 ± 0.1	0.8 ± 0.08
SensSpec	0.89 ± 0.07	0.82 ± 0.06	0.77 ± 0.1	0.81 ± 0.09	0.89 ± 0.09	0.88 ± 0.07	0.75 ± 0.12	0.84 ± 0.09	0.81 ± 0.07	0.88 ± 0.07	0.76 ± 0.09	0.82 ± 0.07
Opt.ROC	0.84 ± 0.08	0.87 ± 0.08	0.74 ± 0.11	0.82 ± 0.09	0.88 ± 0.09	0.88 ± 0.08	0.76 ± 0.13	0.84 ± 0.09	0.8 ± 0.07	0.88 ± 0.08	0.76 ± 0.09	0.82 ± 0.07
Preval	0.92 ± 0.07	0.95 ± 0.05	0.85 ± 0.13	0.85 ± 0.09	0.94 ± 0.08	0.94 ± 0.06	0.82 ± 0.13	0.87 ± 0.11	0.94 ± 0.08	0.94 ± 0.06	0.84 ± 0.11	0.87 ± 0.09
AvgProb	0.69 ± 0.08	0.68 ± 0.06	0.63 ± 0.12	0.72 ± 0.09	0.68 ± 0.08	0.68 ± 0.07	0.6 ± 0.12	0.71 ± 0.09	0.68 ± 0.08	0.68 ± 0.07	0.62 ± 0.09	0.69 ± 0.08
pS-SDM+PRR	0.92 ± 0.15	0.91 ± 0.12	0.81 ± 0.19	0.86 ± 0.19	0.93 ± 0.15	0.92 ± 0.12	0.8 ± 0.23	0.84 ± 0.21	0.93 ± 0.14	0.92 ± 0.12	0.82 ± 0.19	0.86 ± 0.18
MEM+PRR	0.93 ± 0.14	0.93 ± 0.1	0.85 ± 0.18	0.91 ± 0.15	0.93 ± 0.15	0.92 ± 0.13	0.85 ± 0.2	0.89 ± 0.16	0.93 ± 0.14	0.92 ± 0.13	0.86 ± 0.17	0.88 ± 0.15

Supplementary material of chapter 3.1

Appendix 1 - Supplementary Figures and Tables

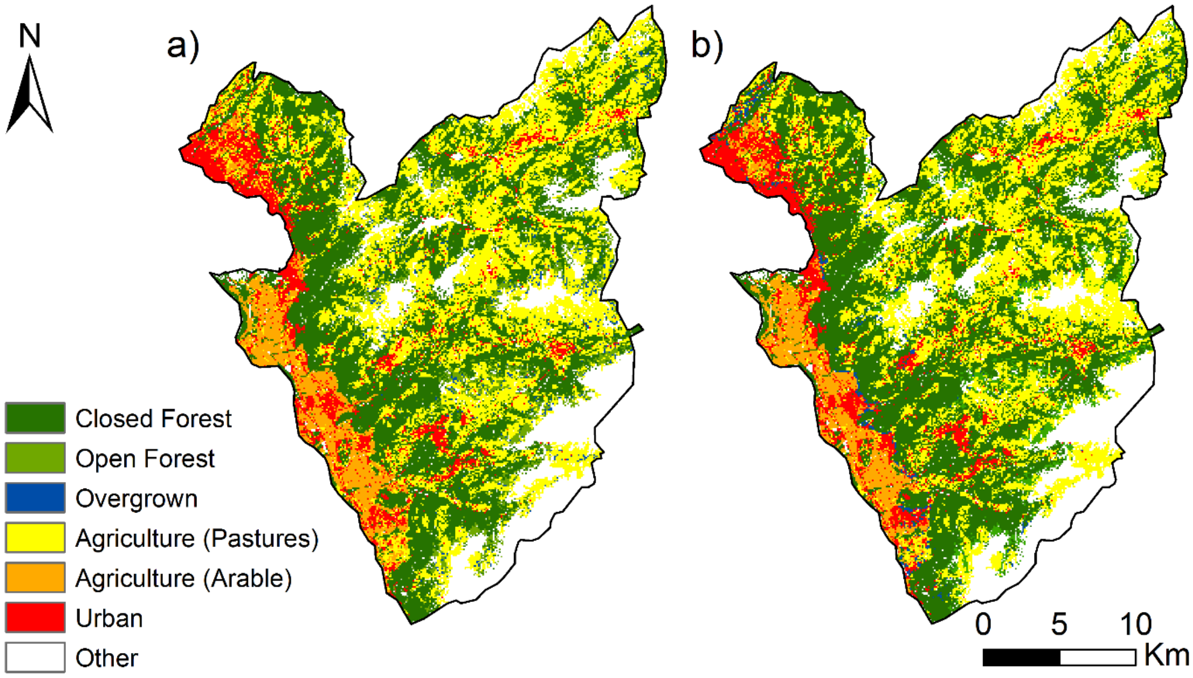


Figure S1 Land use maps (100x100m) acquired from a) Swiss land use statistics data for 2009 (SFSO 2013) and b) the predicted 2035 land use change modelled by Jenkins et al. (2015). The seven land categories are derived from the 72 categories defined by the SFSO.

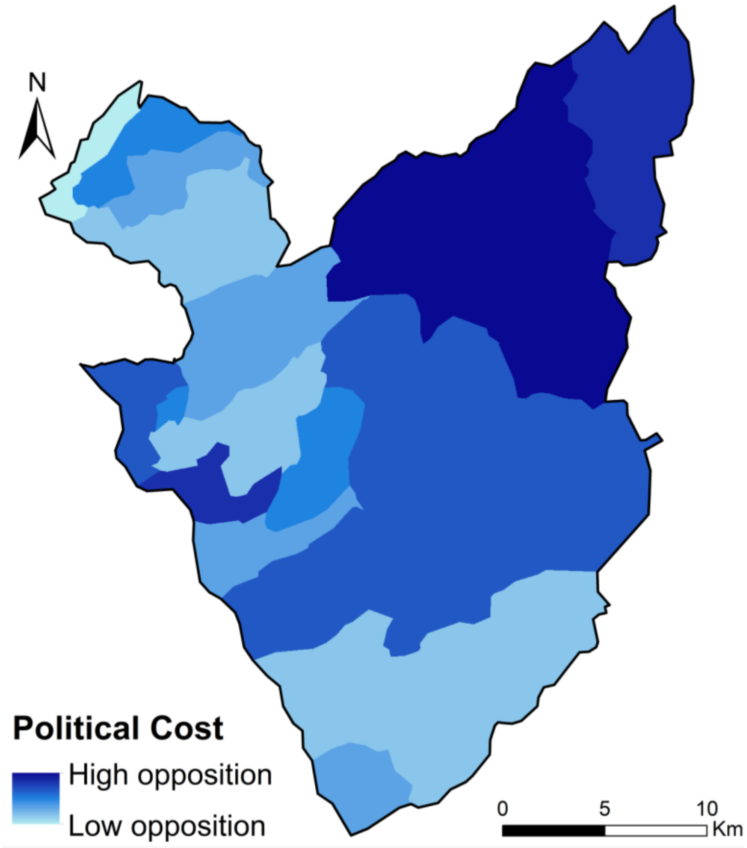


Figure S2 Political costs acquired from Cardoso (2015). The darker regions indicate communes that have a higher political cost as they voted against policy that would support biodiversity in the two most recent elections.

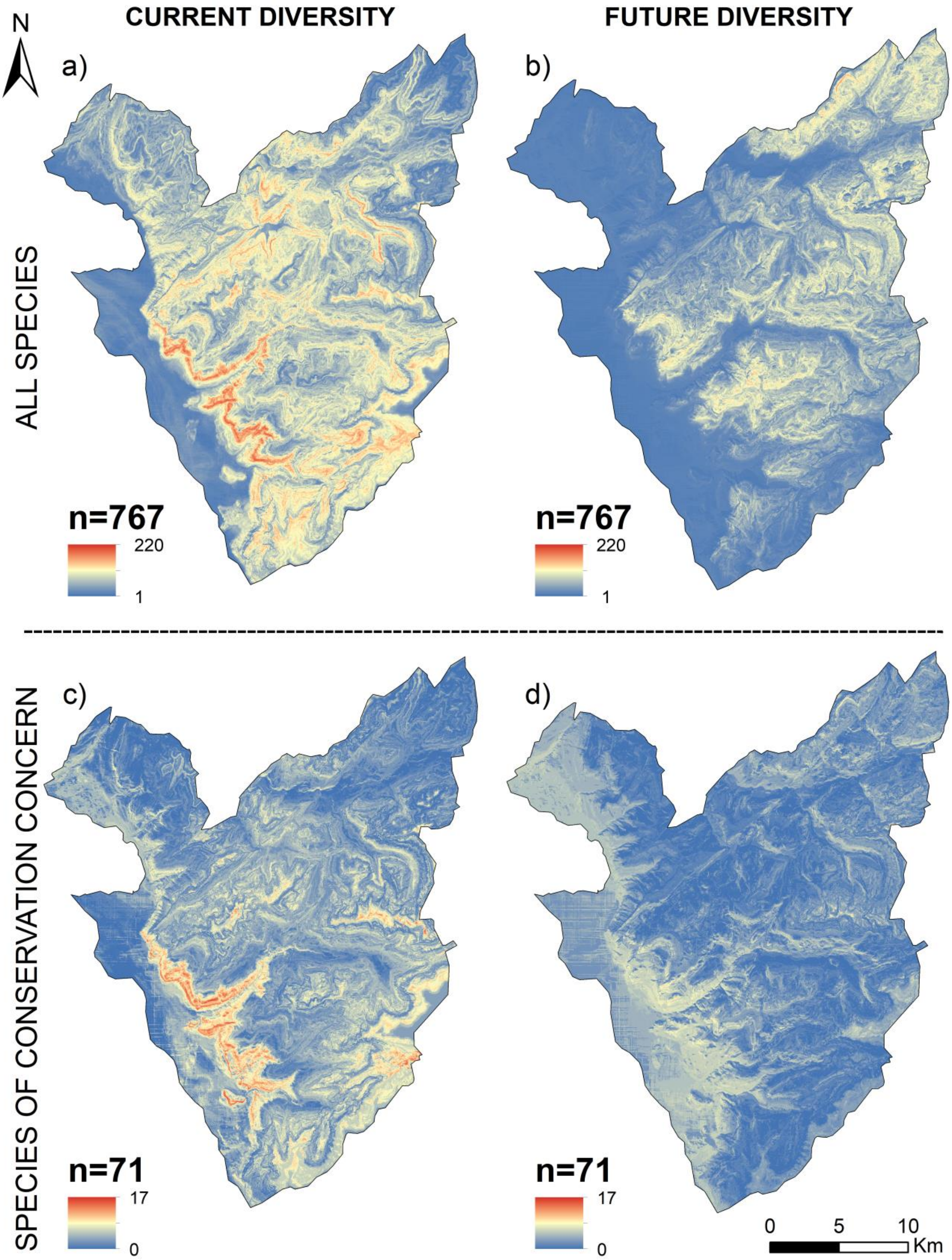


Figure S3 Biodiversity maps represent richness of the modelled species with their current (a, c) and projected future (b, d) distributions. Future refers to the distributions predicted to 2060 using the A1B climate change scenario. A subset of the species was used to create maps of rare species which were identified as being threatened and/or protected (c, d). Species richness is represented here as the number of species with overlapping ranges.

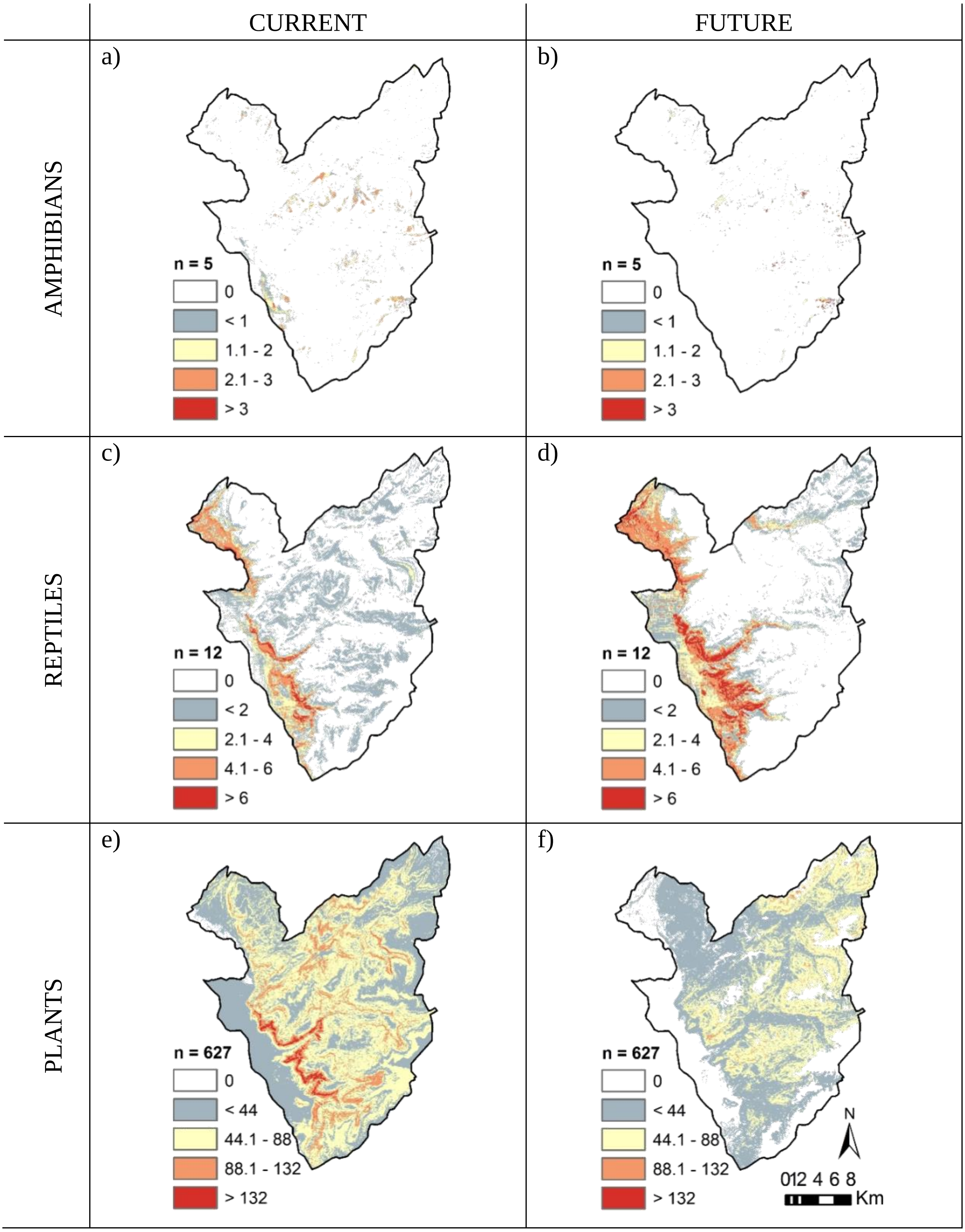


Figure S4 Priority maps showing the sum of species priority scores within each taxonomic group. These were calculated using species ranges from the current time step (a, c, e) and using the predicted 2060 ranges using A1B climate change scenario (b, d, f).

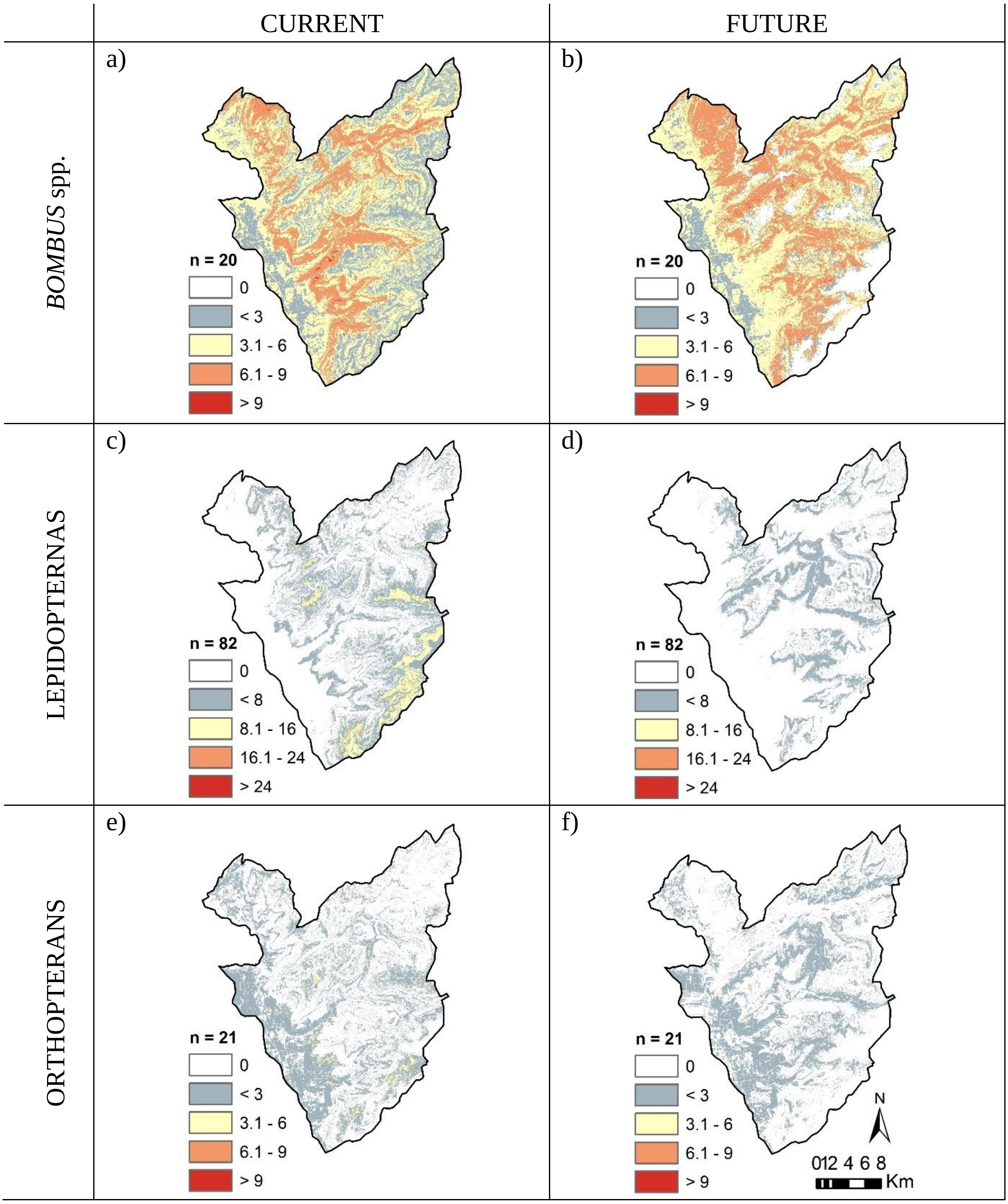


Figure S5 Continuation of Fig S5. Priority maps showing the sum of species priority scores within each taxonomic group. These were calculated using species ranges from the current time step (a, c, e) and using the predicted 2060 ranges using A1B climate change scenario (b, d, f).

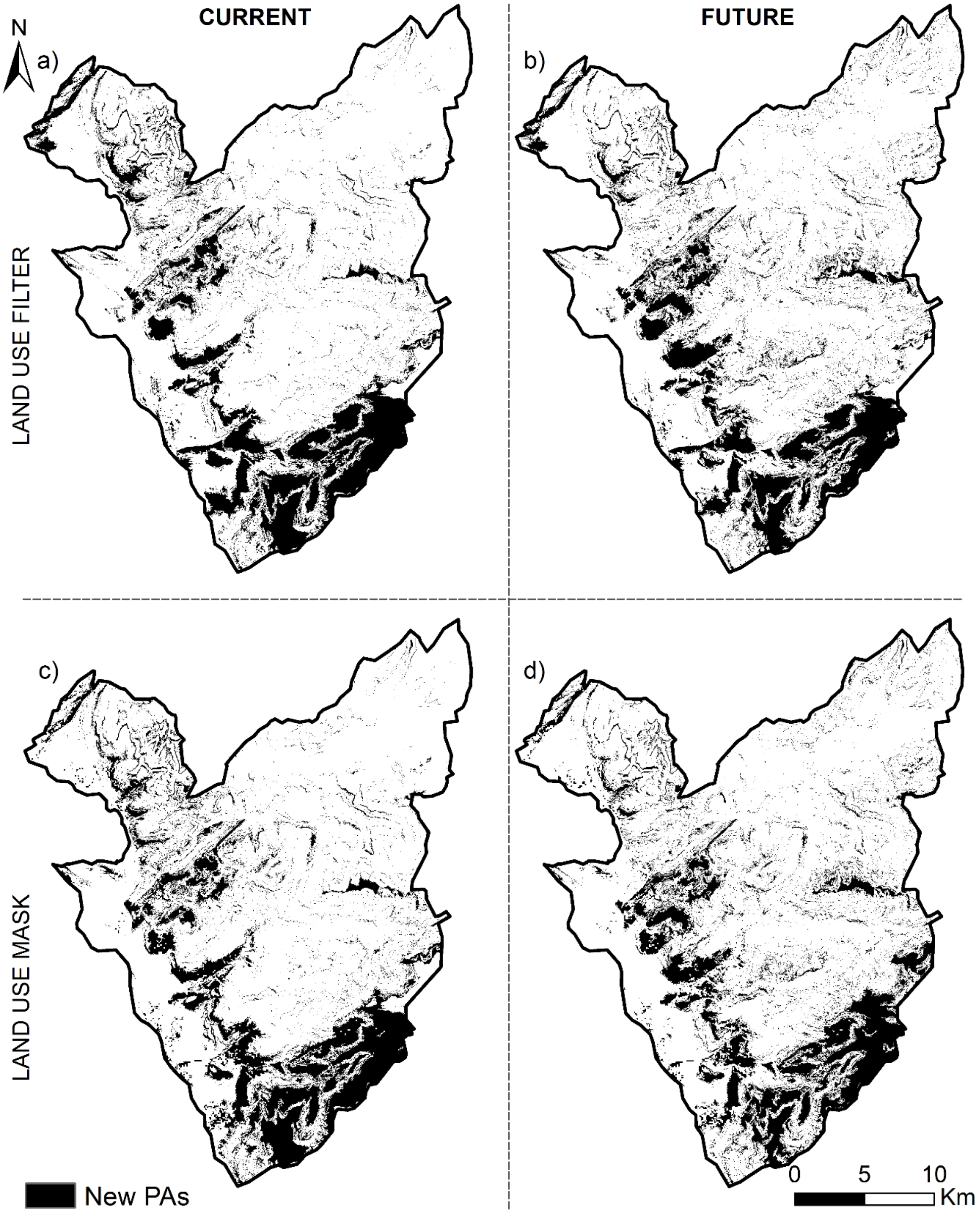
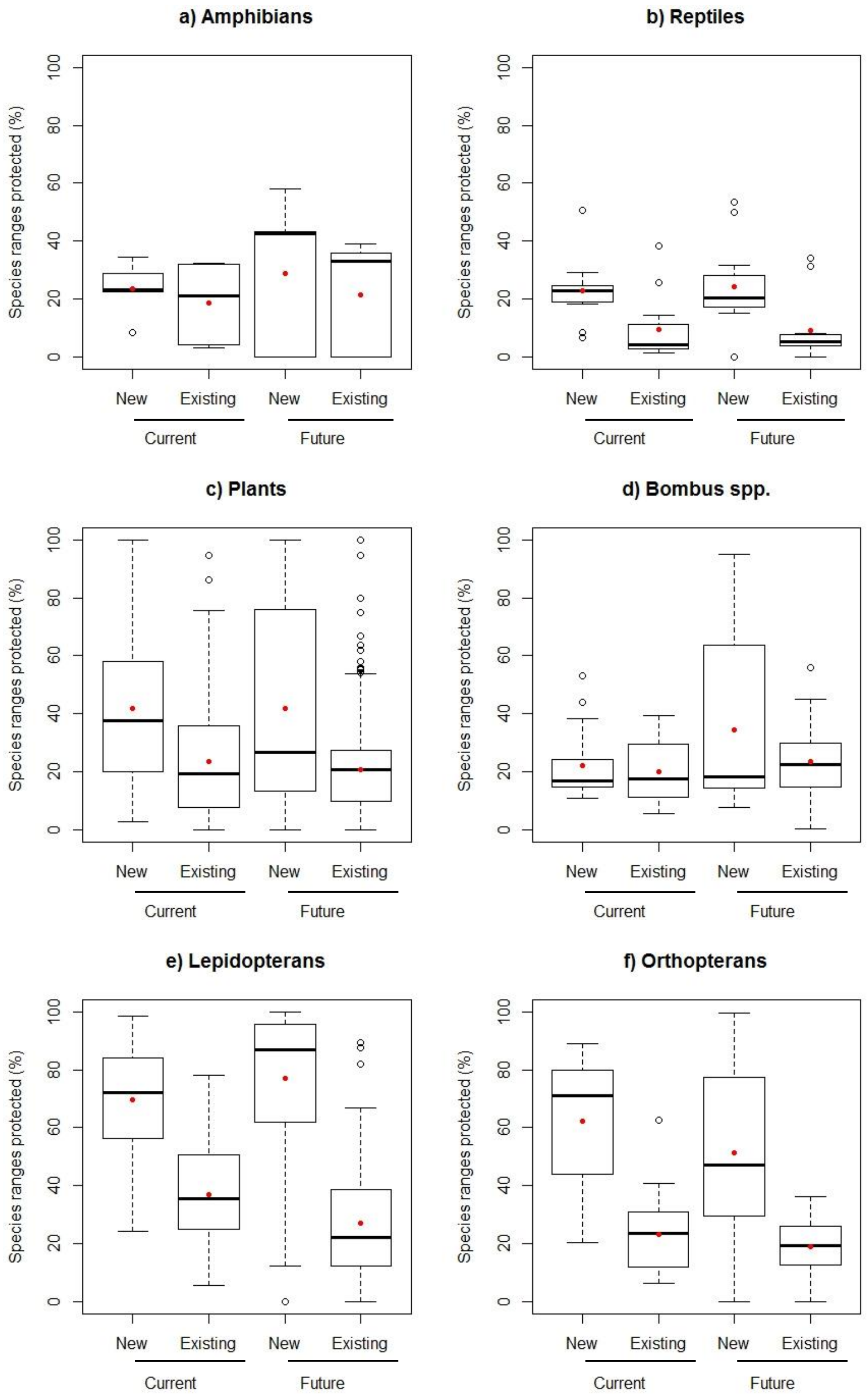


Figure S6 Spatial representation of the Zonation solutions created using the LU-filter (a, b) and LU-mask (c, d). Future refers to the distributions predicted to 2060 using the A1B climate change scenario. The black areas represent the top priority areas and cover the same proportion of the landscape as the existing Tier 1/ 2 protected areas.



a) Amphibians

b) Reptiles

c) Plants

d) Bombus spp.

e) Lepidopterans

f) Orthopterans

Figure S7 The proportions of species ranges protected by the new Zonation solutions compared with the existing protected areas using LU-mask. Future refers to the distributions predicted to 2060 using the A1B climate change scenario. Open circles indicate outliers, black bars show medians, and red dots indicate means.

Table S1 The types of protected areas found within the Western Swiss Alps study site.

Level of Protection[a]	**Type of Protected Area**	**Land Cover %**[b]
Tier 1	UNESCO-MAB Biosphere Reserve	0.02
	Federal Inventory of Raised and Transitional Mires of National Importance	
Tier 2	Ramsar Sites	18.10
	Emerald Sites	
	Federal Hunting Reserves	
	Federal Inventory of Alluvial Zones of National Importance	
	Federal Inventory of Fenlands of National Importance	
	Federal Inventory of Dry Grasslands and Pastures of National Importance	
	Federal Inventory of Reserves for Waterbirds and Migratory Birds of International and National Imp.	
	Federal Inventory of Amphibian Spawning Areas of National Importance	
Tier 3	Forestry Reserves	5.69
	Pro-Natura Preserves	
Tier 4	Natural Landscapes and Monuments of National Importance	39.19
	Gruyère Regional Park	

[a] Tier 1 represents the protected areas with the highest level of protection, and Tier 4 the lowest.

[b] The proportion of the study area covered by each tier is shown as a percentage of land cover.

Appendix 2 – Species Information: Available in this link: **https://goo.gl/pKNSgD**

Appendix 3 - Supplementary Methods

Species Data

Species occurrence data from field surveys in the study area for plants (n = 627), amphibians (n = 5), reptiles (n = 12), and insects (n = 123; of which lepidopterans n = 82, orthopterans n = 21, and *Bombus* spp. n = 20) were used in this study (Table 1; Appendix 2). Each species was assigned a threat status per the Swiss Red Lists from the Federal Office for the Environment (FOEN) at the national scale, except for *Bombus* spp. which used the IUCN European Red List of Threatened Species (IUCN, 2016). The Swiss protection status of each species was outlined by Cardoso (2015) and included species that are explicitly referenced in legislative documents. While some bird and mammal data would have also been available, we could not use them to build spatial models because these data are only available as occurrences (presence-only) in heterogeneous national databases and at a resolution too coarse for our analyses.

To obtain the full information about the plant data, two datasets where combined. The first, included 912 vegetation plots of 4 m^2 which were selected following a random-stratified sampling design (Hirzel & Guisan 2002) and the presence-absence of each species was recorded in each plot (Dubuis *et al*. 2011). Only vascular species in open and non-woody vegetation were sampled. This dataset consisted of 795 species. The second dataset included 3076 vegetation plots, from a grid of 400 m over the entire study area; therefore, a point was recorded every 400 m (Hartmann *et al*. 2009). If a point fell on forested land, a field sampling was carried out. The plot was a circle with a radius of 10m from the centre point where the grid point was located. All vascular plant species were recorded. This dataset consisted of 667 species. The final dataset of plants has 3967 plots composed by 1088 species (627 after removing the species with less than 10 occurrences).

Insect species were sampled in 50 m x 50 m plots selected with a balanced stratified random sampling design (Hirzel & Guisan 2002) and centered on the coordinates of previous 4 m^2 vegetation sampling sites (Dubuis *et al*. 2011). A total of 202 sites were investigated for bumblebees and orthopteran species (Pradervand *et al*. 2012; Pradervand *et al*. 2014a) and 208 sites for lepidopteran species (Pellissier *et al*. 2013). The sampling identified 28 species of bumblebees, 138 species of Lepidoptera and 41 species of Orthoptera. The sampling was performed during the insects' active hours (10:00 to 17:00) under good weather conditions (i.e. little wind, sunny, and temperatures above 15°C). Presence-absence of species was first recorded in four sub-squares of 10 m x 10 m located at the four cardinal points of the large square and subsequently in the complete 50 m square was assessed. After removing the species with less than 15 occurrences, a total of 123 insect species were retained (lepidopterans n = 82, orthopterans n = 21, and *Bombus* spp. n = 20).

The amphibians and reptiles data originated from direct observations gathered in the database of *info species* (Neuchâtel, Switzerland) from KARCH (Koordinationsstelle für Amphibien und Reptilienschutz in der Schweiz) at 1 km resolution. For amphibians, a total of 268 observations of 5 amphibian species were included. For reptiles, the time frame of recorded observations was limited between 1965 and 2014, and 1426 observations of 13 reptile species were used.

Environmental Variables

Environmental variables for temperature and precipitation were calculated from the mean during the growing season (March to August) from 1981 to 2009 at a resolution of 25 m. This data was then used to project the mean monthly temperatures for the whole study area, a method previously used by Pradervand *et al*. (2014b). From this, the maximum temperature and precipitation, the minimum temperature and precipitation, and the sum precipitation were selected to use in the models. Evapotranspiration was calculated based on these precipitation variables and the amount of solar radiation. Aspect was calculated based on the digital elevation model of the study area. Topography, growing degree days, and slope were retrieved from the Swiss Federal Institute for Forest, Snow and Landscape Research WSL database at 25 m resolution for vegetation, and 50 m for the insect groups.

Species Distribution Models

To model the current and future distributions of the studied species, an ensemble of small models (ESMs; Breiner *et al*. 2015) was used: The data for each species was randomly partitioned into 70% for calibration and 30% for validation. For each combination of environmental predictors, GLM, RF, and MaxEnt were evaluated using a maximisation of the True Skill Statistic (maxTSS), a balanced view evaluation method which takes both omission and commission errors into account (Hanssen & Kuipers 1965). These evaluations were used to create a weighted mean of the predictions to create an ensemble model. Each ensemble model for each combination of predictors was again evaluated by maxTSS, and these evaluations give the best combination of predictors to obtain the final ensemble model which was used for the final evaluation. The full dataset was then used for the projection step where the evaluators for each of the three techniques produce an average weight of the predictions within each of them. Finally, a last prediction for the ensemble model was calculated using the weighted mean across the techniques. A binary presence-absence distribution was obtained for each species. These distributions were then summed, using the *raster* package (Hijmans 2015) in R, across all species of an interest group to create different diversity maps.

Climate Scenarios

As future climate is difficult to predict, numerous climate change scenario storylines have been created by several researchers (Moss *et al*. 2010). For this study, we focused on the A1B scenario which predicts rapid economic growth and the global population peaking around 2050 and subsequently declining (IPCC 2001). It also forecasts the introduction of more efficient technology and a balance of fossil fuels and renewable energy sources.

The HadCM3Q0-CLM (Hadley) climate change model was used to model future species distributions. This climate model was developed by the Swiss Climate Change Scenario CH2011 project through the Centre for Climate Systems Modelling (www.c2sm.ethz.ch/). The model is based on high resolution data which incorporates anomalies from all Swiss weather stations (Bosshard *et al*. 2011). It is important to include regional data as Switzerland is warming at a higher rate compared to the rest of the Northern Hemisphere, and this difference is most notable in the Alps (Rebetez & Reinhard 2007). HadCM3 (Hadley Centre Coupled Model, version 3) is a global climate model that couples atmospheric and oceanic general circulation models and was developed by the Hadley Centre (Gordon *et al.* 2000) with standard sensitivity (Q0). The Community Land Model (-CLM) is the regional climate model that is coupled with the HadCM3Q0 to add regional land-surface atmosphere interactions and was developed by the Swiss Federal Institute of Technology Zurich. Both climate models were projected using seasonal temperature and precipitation values to the years 2045 – 2074 (from here on referred to as 2060) under the A1B scenario.

Supplementary References

Bosshard T, Kotlarski S, Ewen T, Schär C. 2011. Spectral representation of the annual cycle in the climate change signal. Hydrology and Earth System Sciences **15**:2777–2788.

Breiner FT, Guisan A, Bergamini A, Nobis MP. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. Methods in Ecology and Evolution **6**:1210–1218.

Cardoso ARP. 2015. Spatial data and modelling for the prioritisation of conservation areas in the alpine region of the canton of Vaud. Masters. Universidade do Porto.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J. P. & Guisan, A. 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions,* 17**,** 1122-1131.

Gordon C, Cooper C, Senior CA, Banks H, Gregory JM, Johns TC, Mitchell JFB, Wood RA. 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics **16**:147–168.

Hanssen AJ, Kuipers WJ. 1965. On the relationship between the frequency of rain and various meteorological parameters. Meded Verhand:2–15.

Hartmann, P., Fouvy, P. & Horisberger, D. 2009. L'Observatoire de l'écosystème forestier du canton de Vaud: espace de recherche appliquée| The Forest Ecosystem Observatory in Canton Vaud: a field of applied research. *Schweizerische Zeitschrift für Forstwesen,* 160**,** s2-s6.

Hijmans RJ. 2015. Raster: geographic data analysis and modelling. R package version 2.5-2. https://CRAN.R-project.org/package=raster

Hirzel, A. & Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling,* 157**,** 331-341.

IPCC. 2001. Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA. Pages 88.

IUCN. 2016. The IUCN Red List of Threatened Species. Version 2016-3. Available from www.iucnredlist.org (accessed May 2016).

Moss RH et al. 2010. The next generation of scenarios for climate change research and assessment. Nature **463**:747–756.

Pellissier, L., Alvarez, N., Espindola, A., Pottier, J., Dubuis, A., Pradervand, J. N. & Guisan, A. 2013. Phylogenetic alpha and beta diversities of butterfly communities correlate with climate in the western Swiss Alps. *Ecography,* 36**,** 541-550.

Pradervand, J.-N., Dubuis, A., Reymond, A., Sonnay, V., Gelin, A. & Guisan, A. 2012. Quels facteurs influencent la richesse en orthoptères des Préalpes vaudoises? *Bulletin de la Société Vaudoise des Sciences Naturelles,* 93**,** 155-173.

Pradervand, J.-N., Pellissier, L., Randin, C. F. & Guisan, A. 2014a. Functional homogenization of bumblebee communities in alpine landscapes under projected climate change. Climate Change Responses. BioMed Central Ltd, 1.

Pradervand J-N, Pellissier L, Randin CF, Guisan A. 2014b. Functional homogenization of bumblebee communities in alpine landscapes under projected climate change. Climate Change Responses **1**:1.

Rebetez M, Reinhard M. 2007. Monthly air temperature trends in Switzerland 1901–2000 and 1975–2004. Theoretical and Applied Climatology **91**:27–34.

Supplementary material of chapter 3.2

Table S2 Environmental predictors grouped into environmental types that reflect their ecological meaning.

Ecological type	Environmental predictor	References
Climate	Minimum Temperature of Coldest Month (Bio6)	(Arévalo et al., 2005)
	Temperature Annual Range (Bio7)	(Vicente et al., 2013a)
	Annual Precipitation (Bio12)	(Wang & Wang, 2006)
	Precipitation Seasonality (Bio15)	(Vicente et al., 2013a)
Soil types	Percentage cover of cambisoils (%CamSoils)	(Vicente et al., 2013a)
Productivity	Mean gross annual primary productivity (GPP)	(Vicente et al., 2014; Williams et al., 2004)
Landscape composition	Percentage cover of broadleaf forests (%BlFor)	(Pino et al., 2005)
	Percentage cover of conifer forests (%ConFor)	(Vicente et al., 2013b)
	Percentage cover of mixed forests (%MixFor)	(Vicente et al., 2013a)
	Percentage cover of artificial areas (%Art)	(Chytrý et al., 2008)
	Percentage cover of arable land (%Arab)	(Vicente et al., 2013a)

Table S2 Variable importance for the invasive species *Acacia dealbata*, assessed using the ensemble modelling approach for the different "calibration approaches".

Variables| Approach	Separate Countries (Portugal)	Separate Countries (Galicia)	Full projection (Portugal)	Full projection (Galicia)	Transfrontier context
Minimum Temperature of Coldest Month (Bio6)	**0.57±0.32**	**0.68±0.22**	**0.53±0.31**	**0.69±0.23**	**0.61±0.27**
Temperature Annual Range (Bio7)	0.22±0.32	0.24±0.18	0.19±0.17	0.26±0.19	**0.22±0.17**
Annual Precipitation (Bio12)	**0.27±0.19**	**0.28±0.16**	**0.22±0.28**	**0.27±0.16**	0.1±0.07
Precipitation Seasonality (Bio15)	0.09±0.07	0.23±0.16	0.07±0.07	0.22±0.16	0.15±0.08
Percentage cover of cambisoils (%CamSoils)	0.01±0.02	0.02±0.02	0.02±0.02	0.02±0.02	0.01±0.02
Mean gross annual primary productivity (GPP)	0.05±0.05	0.02±0.03	0.05±0.05	0.03±0.03	0.04±0.04
Percentage cover of broadleaf forests (%BlFor)	0.01±0.01	0±0.01	0.02±0.02	0±0.01	0.01±0.01
Percentage cover of conifer forests (%ConFor)	0.07±0.05	0.01±0.01	0.06±0.04	0.01±0.01	0.04±0.02
Percentage cover of mixed forests (%MixFor)	0.02±0.02	0.04±0.05	0.02±0.03	0.04±0.05	0.04±0.03
Percentage cover of artificial areas (%Art)	0.2±0.07	0±0.01	0.2±0.07	0.01±0.01	0.06±0.03
Percentage cover of arable land (%Arab)	0.07±0.04	0±0	0.05±0.04	0±0	0.06±0.03

Table S3 Percentage of areas predicted as presence or absence, under current conditions, 2050, and 2070 (RCP4.5 and RCP 8.5) for the different "calibration approaches".

Year	Approach	Scenario	Prediction (0/1)	All study area (%)	Only protected areas (%)
Current	Separate Countries	-	Absences	70	87
			Presences	**30**	13
	Full projection (Portugal)		Absences	60	85
			Presences	40	15
	Full projection (Galicia)		Absences	70	86
			Presences	30	14
	Transfrontier context		Absences	66	85
			Presences	**34**	**15**
2050	Separate Countries	RCP 4.5	Absences	42	71
			Presences	57	28
		RCP 8.5	Absences	21	57
			Presences	69	42
	Full projection (Portugal)	RCP 4.5	Absences	51	78
			Presences	48	21
		RCP 8.5	Absences	40	67
			Presences	60	32
	Full projection (Galicia)	RCP 4.5	Absences	31	59
			Presences	68	41
		RCP 8.5	Absences	24	52
			Presences	75	48
	Transfrontier context	RCP 4.5	Absences	32	60
			Presences	67	39
		RCP 8.5	Absences	23	49
			Presences	77	51
2070	Separate Countries	RCP 4.5	Absences	36	66
			Presences	**63**	34
		RCP 8.5	Absences	21	45
			Presences	**78**	54
	Full projection (Portugal)	RCP 4.5	Absences	44	73
			Presences	55	26
		RCP 8.5	Absences	37	62
			Presences	62	38
	Full projection (Galicia)	RCP 4.5	Absences	32	59
			Presences	67	40
		RCP 8.5	Absences	9	24
			Presences	90	76
	Transfrontier context	RCP 4.5	Absences	24	50
			Presences	**75**	**49**
		RCP 8.5	Absences	16	35
			Presences	**83**	**65**

References

Arévalo, J. R., Delgado, J. D., Otto, R., Naranjo, A., Salas, M., & Fernandez-Palacios, J. M. (2005). Distribution of alien vs. native plant species in roadside communities along an altitudinal gradient in Tenerife and Gran Canaria (Canary Islands). *Perspectives in Plant Ecology Evolution and Systematics, 7,* 185-202.

Chytrý, M., Maskell, L. C., Pino, J., Pyšek, P., Vilà, M., Font, X., & Smart, S. M. (2008). Habitat invasions by alien plants: a quantitative comparison among Mediterranean, subcontinental and oceanic regions of Europe. *Journal of Applied Ecology, 45,* 448-458.

Pino, J., Font, X., Carbo, J., Jove, M., & Pallares, L. (2005). Large-scale correlates of alien plant invasion in Catalonia (NE of Spain). *Biological Conservation, 122*, 339-350.

Vicente, J. R., Fernandes, R. F., Randin, C. F., Broennimann, O., Goncalves, J., Marcos, B., Pocas, I., Alves, P., Guisan, A., & Honrado, J. P. (2013a). Will climate change drive alien invasive plants into areas of high protection value? An improved model-based regional assessment to prioritise the management of invasions. *Journal of Environmental Management, 131*, 185-195.

Vicente, J. R., Pereira, H. M., Randin, C. F., Goncalves, J., Lomba, A., Alves, P., Metzger, J., Cezar, M., Guisan, A., & Honrado, J. (2014). Environment and dispersal paths override life strategies and residence time in determining regional patterns of invasion by alien plants. *Perspectives in Plant Ecology Evolution and Systematics, 16,* 1-10.

Vicente, J. R., Pinto, A. T., Araujo, M. B., Verburg, P. H., Lomba, A., Randin, C. F., Guisan, A., & Honrado, J. P. (2013b). Using Life Strategies to Explore the Vulnerability of Ecosystem Services to Invasion by Alien Plants. *Ecosystems, 16,* 678-693.

Wang, R., & Wang, Y. Z. (2006). Invasion dynamics and potential spread of the invasive alien plant species Ageratina adenophora (Asteraceae) in China. *Diversity and Distributions, 12*, 397-408.

Williams, J. W., Seabloom, E. W., Slayback, D., Stoms, D. M., & Viers, J. H. (2004). Anthropogenic impacts upon plant species richness and net primary productivity in California. *Ecology Letters, 8*, 127-137.