Facoltà di Scienze Matematiche, Fisiche e Naturali
Dipartimento di Biologia e Biotecnologie "C. Darwin"

Dottorato in Biologia Ambientale ed Evoluzionistica
Curriculum Biologia Animale (XXXVII Ciclo)

Optimizing modelling tools to anticipate biological invasions – on and beyond Species Distribution Models

PhD Candidate:	Tutor:
Bart Steen	Prof. Luigi Maiorano
Matricola: 2025594	Prof. Antoine Guisan

Anno Accademico 2024/2025

Abstract

Invasive Alien Species (IAS) are non-native taxa that cause ecological and/or economic damage. They are one of the primary causes of loss of biodiversity and ecosystem services, though they can be managed. However, conservation efforts are often short on manpower and resources. It is therefore crucial to computationally model the potential distributions of IAS. This thesis focuses on optimizing existing modelling techniques for this purpose. One such technique is known as Species Distribution Modelling (SDM), which works by relating the environmental conditions where the species occurs to those where it does not, or, more commonly, to those at randomly sampled locations, which are known as background points (BP). This allows SDMs to simulate the realized niche of and habitat suitability for a species.

Because non-native species can become invasive without our knowledge, the first chapter of this thesis compares which environmental variables most strongly predict the distributions of non-native species and their native congenerics. Subsequently, as SDMs quantify an environmental niche, we designed novel ways of environmentally sampling these variables using BP and compared them to the typical geographic way in terms of SDM performance. The clues gathered in the previous two chapters were then applied in the third to a well-known IAS, *Heracleum mantegazzianum* Sommier & Levier. Finally, the probability of invasion success in a new area was studied using the Niche Margin Index (NMI), which has been shown in previous studies to be positively correlated with establishment probability and measures the (dis)similarity between the native climatic niche and the local environmental conditions. We researched if different ways of quantifying the native niche affected the NMI values.

It was found that non-native species are more strongly associated with anthropogenic environmental variables and less pristine land cover types than their native counterparts, though climate was the strongest predictor in both species groups. Sampling BP stratified randomly and fully randomly in environmental space yielded more stable and accurate SDMs than the classic method. The use of environmentally stratified BP may facilitate SDM overfitting, however. Finally, the highest NMI values were obtained when the native niche was quantified using whole maps of the native range, followed by those obtained when a thresholded SDM of the native range was used and finally by those obtained using only native occurrences.

These findings provide novel guidelines and considerations for using SDMs and analyzing the probability of IAS establishment in order to manage biological invasions.

Table of Contents

Introduction

The conservation crisis

Nature and ecosystems underpin all life on this planet, including that of humans. All estimations regarding the contribution of nature to economy and society are likely underestimations, as many values remain hidden (see e.g., Cavender-Bares et al. 2022 for forest ecosystem services). Nevertheless, even when limiting ourselves to nature's contributions to the economy, e.g., through analysis of provisional Nature's Contributions to People (Ramel et al. 2020), evidence for the importance of the natural world is plentiful. One estimation of nature's direct contribution to global economy is half the world's GDP (World Economic Forum, 2020). Yet, anthropogenic impacts on nature progress. The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) lists five major drivers of biodiversity loss: land use change, climate change, overexploitation, pollution and invasive alien species (IAS).

The ever shifting and shrinking of species' distributions and populations due to the first four causes, or, conversely, the spreading of IAS (the fifth cause), can result in extinction and failure of ecosystem functions (IPBES 2023). Knowledge of species' current and future potential distributions is therefore pivotal. However, holistic enough knowledge is difficult to acquire. Conservation efforts are typically short on funds and manpower (Guisan et al. 2013) and it is impossible to gather sufficient information through scientific fieldwork (Araújo & Guisan 2006). The International Union for the Conservation of Nature (IUCN) has long published expert-based maps of the native distributions of endangered species, but species' actual distributions likely only cover parts of their native range (certainly in the case of plant species and insects (Díaz-Calafat et al. 2024)).

Therefore, computational simulations of species' distributions are pivotal. Sophisticated algorithms that allow fitting so-called Species Distribution Models (SDMs) have been developed by ecologists and mathematicians specifically for this purpose. The core of this thesis is on how to attune SDMs to the central problematic of IAS.

Invasive Alien Species

Even within the scientific community and literature, the definitions of alien or invasive species are often inconsistent and can vary widely (Clarke et al. 2021, Haider 2011, Kourantidou et al. 2022). In this work, we adhere to the following definition: IAS are taxa that 1) are non-native in parts of the world and 2) have been proven to cause economic and/or ecological and/or health damage (Polce et al. 2023). We also consider the term "alien" synonymous to "non-native" and emphasize that not all non-native species are invasive. A species must also cause demonstrable ecological and/or economic damage to be labeled as such (Steen et al. 2024a).

IAS cause an estimated $423 billion of economic damage every year, worldwide (IPBES 2023), through direct impediment of ecosystem services, posing health hazards (Bhowmik & Chandran 2015), overturning nutrient regimes (Steen et al. 2019a), upsetting ecological balance, outcompeting local flora and fauna, forming dense stands and monocultures, etc. It is estimated that IAS have directly caused or contributed to 60% of global species extinctions (IPBES 2023).

International efforts by, e.g. the European Alien Species Information Network (EASIN) team help identify dangerous IAS and facilitate early warning systems for other countries (EASIN, n.d.). Still, the observation of the damage often comes too late, and in such cases, remediation is the only option. Prevention is, however, a far preferable alternative (Schwindt et al. 2024, Steen et al. 2019a). Therefore, in the struggle against IAS, SDMs are invaluable, and it is important to not (always) limit research to those species already known to be invasive, but to also investigate non-native species that do not (by our knowledge) cause damage.

The IAS crisis: pathways of invasion and importance of SDMs

The EASIN team provides a comprehensive overview of the numerous modes by which IAS spread from one place to another (EASIN, n.d.).

They have subdivided these modes into six main categories: escape from confinement, release in nature, stowaway, contaminants, ecological corridors, and unaided (see also Saul et al. 2016). Indeed, some IAS still cause damage without any demonstrable interference from humans. Nevertheless, the pathways all remain greatly problematic, due to insufficient funds or means to control them at source, as well as insufficient public awareness. For instance, the species *Heracleum mantegazzianum* (giant hogweed) has spread to high altitudes in Switzerland after escaping form botanical gardens (Ecospat 2005).

However, at present, high mountains and some other areas (like the polar regions) have remained mostly free of invasion by IAS (Pauchard et al. 2009, Petitpierre et al. 2016). Prevention of damage by IAS is still possible there. However, their invasions must be foreseen. SDMs are critical tools for this.

Species Distribution Models: definition, uses and how to use them

SDMs (also known as ecological niche models (ENMs) and by many other names; Guisan et al. 2013) typify the environmental niche of a species by identifying statistical correlations between the environmental conditions at the locations of species presence (occurrence data) and those at locations where either the species does not occur (absence data) or that are randomly sampled throughout the study area (background points). Therefore, SDMs can only model the known part of the realized niche of a species. The model's outputs can then be used to create maps of habitat suitability for the modelled species all over the study area, which can range from local scale (e.g., the Alps in the Swiss contain of Vaud (RechAlp, n.d.)) to world scale. It can also vary on the temporal scale, if environmental data for the appropriate time are available. Once a species' niche is quantified, it can be projected in geographic space, as well as to the future or the past (Franklin 2023).

The general steps included in running presence-only SDMs are given in figure 1:

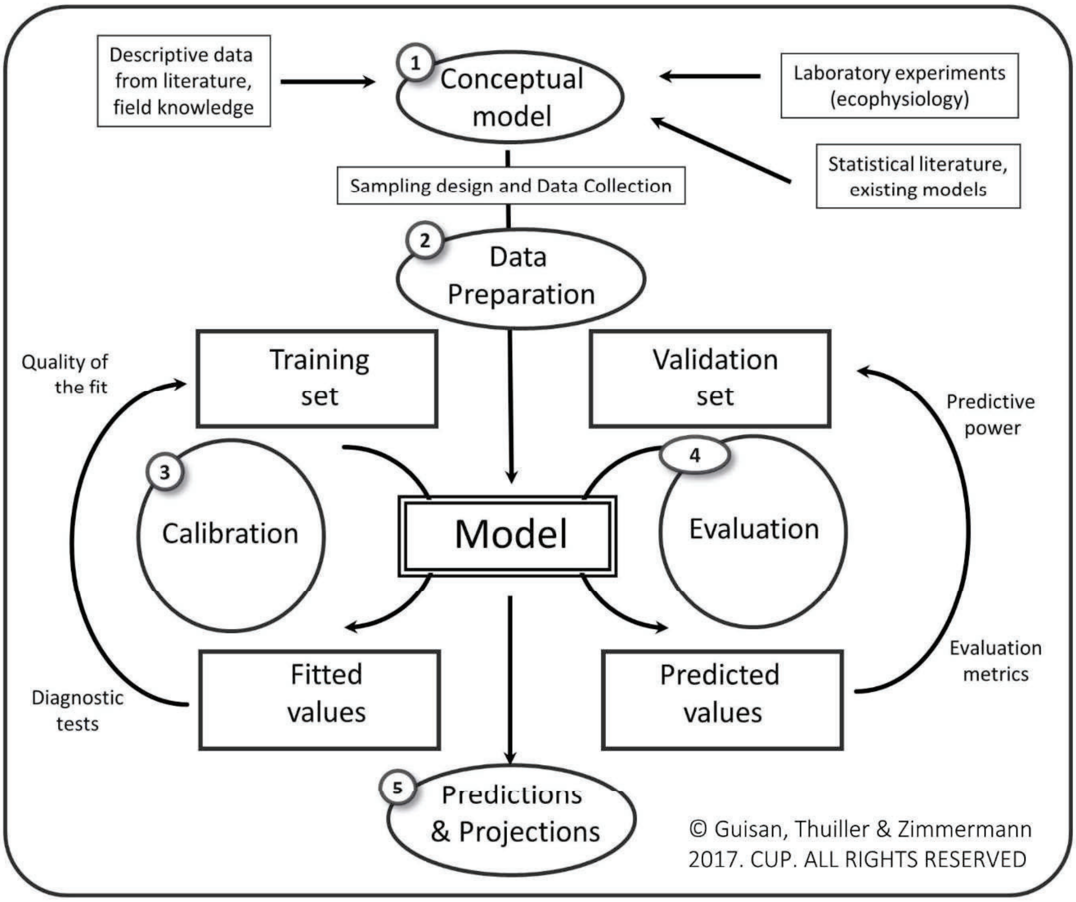


Figure 1: Sequence of the SDM analysis process. Used with permission from Guisan et al. 2017. Courtesy of Prof. Antoine Guisan.

The use of SDMs is increasingly advised in studies to guide and underpin the limited resources available for conservation measures. Though they remain underused in practice (Steen et al. 2024, Tulloch et al.

2016), they are becoming more and more popular. For instance, IUCN is using and publishing SDMs to 1) estimate the prevalence of species within its range as an indicator of extinction risk, and 2) to assess the potential impacts of climate change on the distributions of species (Cassini 2011).

SDMs can serve many purposes, and their use is a rapidly progressing field of science. Many studies use SDMs to model one or a few species of high conservation importance, such as keystone species (Song et al. 2024), umbrella species (Maslo et al. 2016), flagship species (Dejene et al. 2021), or IAS (Steen et al. 2019a, Steen et al. 2024b). However, for other purposes (e.g., assessing the integrity of national ecological infrastructure (Rey et al. 2022), SDMs on more species are needed. Recent advances have been made to both model many species (Adde et al. 2023a) and to automatically select environmental variables (Adde et al. 2023).

SDMs have been used successfully to guide allocation of conservation resources on multiple occasions, (Rathore & Sharma 2023), such as predicting the distributions of invasive alien species (Barbet-Massin et al. 2018), identifying potential protected areas (Araújo & Williams 2000) and predicting endangered species distributions (Engler et al. 2004).

Limitations of SDMs

However, despite their usefulness and successes, there are both practical and theoretical challenges to using SDMs. The list below is not meant to be an exhaustive overview of these challenges, but to address some core limitations in the SDM process that will matter in this thesis.

SDMs and their assumptions

Models are approximations of reality. As a result, one must make assumptions when using them (Franklin 2010, Guisan et al. 2017), first and possibly foremost:

- Equilibrium with the environment

SDMs being static, they implicitly assume that the relationships of a species with its environment are in equilibrium in a given time period (Franklin 1995, Guisan & Zimmermann 2000). In reality, the environment is always changing, and species are evolving, so that one rather speaks about pseudo-equilibrium at a given snapshot in time (Guisan & Zimmermann 2000). SDMs are often used in the context of climate change, in which case this assumption can become violated, e.g. if species cannot track climate change velocity (Jump & Peñuelas 2005, Radchuck et al. 2019).

The assumption is, therefore, probably not satisfied for all species. Notable among the species for which the assumption of equilibrium very likely does not hold up are IAS (Gallien et al. 2012), for which the niche can expand in some case as invasion progresses (e.g., through both niche filling and expansion; Guisan et al. 2014), though this process does halt at a certain stage, when the species has colonized all conditions of its native niche (Foster et al. 2022).

- Everywhere a species can live, it will live

Species distributions are driven by more than what SDMs can account for. Since SDMs are based on environmental data, they can only provide estimations of abiotic habitat suitability. Whilst this does probably influence the likeliness that an IAS could settle in a specific area, it does not include considerations of dispersal capacity limitations, or of barriers to establishment (e.g., soil suitability, biotic interactions, etc.). In the case of plants, natural dispersal capacity is often very limited, ranging between 0.1 and 5000m (Vittoz & Engler 2007). They rely often on the wind to spread their spores (Zanatta et al. 2020) and animals eating their fruits to spread their seeds far. It must however also be said that many plants nowadays are spread by human movement, which probably increases the chances of dispersal for many taxa (Saul et al. 2017).

There is another impediment that SDMs cannot account for, again mostly relevant to plants: the possibility of establishment. Dispersal and habitat suitability often do not suffice to grow in new places due to biotic interactions. All niches may already be occupied, and natural enemies of the species may be present. There is also the possibility of insufficient propagule pressure. For a more exhaustive overview, see Enders et al. 2020.

Probability of establishment is a complex subject. There are numerous studies focusing on the specific traits that help plants to survive (e.g., high seed production, allelopathy, phenotypic plasticity, etc. (Skočajić & Nešić 2021)), and on invasion probability at different life stages of the plant (e.g., Gioria et al. 2018, Moles & Westoby 2006, Stohlgren et al. 2005). Whilst many questions remain on this topic (Crooks & Rilov 2009), this thesis makes no attempt to research the exact mechanisms that determine invasive success, but the probability of establishment is a meaningful addition to SDM outputs which will be addressed again later.

- Niche stability

Even if a species is at equilibrium with its environment at a certain time, the relationships are assumed to be static over time and space (Barbet-Massin et al. 2018). SDM outputs are namely often projected into the future, the past, or another geographic area. The assumption of niche stability may not hold true. For instance, after their first introduction to a new area, the niches of IAS expand. This does not yet imply disequilibrium as their niche is in theory stable until they have filled all conditions within their native niche, but the native niche is possibly an unreliable estimate of their invaded niche (Broennimann et al. 2007). However, Petitpierre et al. (2012) have shown that in general niche shifts are rare in major plant invaders, and Broennimann et al. (2021) have shown that probability of invasion success by mammal species is positively correlated to the degree of similarity between the native and invaded range climatic niches. This new avenue of science requires further research – see below.

Correlation is not causality

SDMs are correlative models, and thus the species-environment relationships captured do not necessarily reflect ecological causation (Austin 2007, Dormann et al. 2012).

This limitation of SDMs links to three different aspects of SDMs: environmental variable selection, background point sampling, and occurrence data.

- Environmental data

It is possible to get a well-performing SDM whilst choosing variables that are not associated with the species' niche at all (Porfirio et al. 2014).

This is undesirable. Even if the model can explain the distribution of the occurrence data provided to it, it is unlikely that this data provides a holistic enough view of the species' actual distribution, meaning that the model's results might be misleading when being projected spatially or temporally (Randin et al. 2020). Insight into the species' niche and the causality of its distributions are needed as a reality check.

In addition, the (non-)availability of environmental covariates is often a limiting factor for SDM-based studies. For instance, world-scale climatic models do not account for local-scale weather anomalies. Still, coarse-scale world climate models are available (Karger et al. 2017) and local-scale environmental data have been used in many studies, especially in well-documented areas, such as the Swiss Western Alps of Vaud (Von Daniken et al. 2014)) or the whole area of Switzerland (Külling et al. 2024).

Nevertheless, variable selection is a critical part of properly quantifying the niche of a species.

- Background points: sampling the environment

Selection of the appropriate variables is not enough. There is also the problematic of sampling the background environment when only presences are available to fit the model (Phillips et al. 2009, Barbet-Massin et al. 2012, Wisz & Guisan 2009). But herein also lies an opportunity to introduce ecological theory and (possibly) reliability into SDMs.

The environment can be sampled with absence data, meaning locations where the species is not, but these are often not available (Václavík & Meentemeyer 2009), perhaps mostly because they are not included in opportunistic datasets, which rules out the use of citizen science data (Morera-Pujol et al. 2023). A demonstrably effective alternative are background points (Barbet-Massin et al. 2012), randomly selected locations that serve to inform the model about the environment available in the study area.

For the sake of computational efficiency, whole datasets covering every single pixel of the study area are never used (or desired) in SDMs. Instead, a random number of pixels (typically 10 000) are sampled to select the environmental values that should be contrasted with those of the locations where the species is present. Typically, in the large majority of papers, they are sampled entirely randomly in geographic space, i.e., each pixel in the map has an equal chance of being sampled (Steen et al. 2024b).

If the method of background point selection gives the model a biased view of the environment, the effect might be the same as when selecting the wrong environmental variables. Sampling an environmental niche in a geographic manner may pose this risk. Sampling the background points environmentally may therefore be more desirable (Steen et al. 2024).

In what way should the environment then be sampled? Fully random in the environmental hyperspace (as defined by the environmental variables inserted in the model) is an option. However, when looking at the optimal design for sampling species occurrences in the field, Hirzel & Guisan (2002) have hypothesized that stratified random in environmental space is more desirable. This principle is transferable to background points. If the strata are chosen well, sampling happens on both sides of the environmental limits of where the species can occur, and SDMs are given optimal information to determine where precisely the limit is. This should lead to optimal typification of the species' environmental niche and therefore to accurate spatial and temporal predictions of habitat suitability.

However, considering that the aforementioned assumption of equilibrium with the environment is often not satisfied, the limit between suitable and unsuitable habitat may be incorrectly identified also when sampling the background points in an environmentally stratified fashion. Therefore, it is interesting to study those rare cases when environmental equilibrium can be assumed. For instance, Alpine plants likely occupy their full realized environmental niches, since they can grow at lower altitudes (Steen et al. 2024b), but are outcompeted (Vetaas 2002, Lyu and Alexander 2022). The habitat of these species is at the environmental edge of the study area. Therefore, sampling the extremes is prudent in this case. This would be done by sampling stratified randomly in environmental space.

The opposite is likely true for generalist species. Here, it would be desirable to distribute the sampling points across the most common environmental values, meaning fully randomly in environmental space (Bazzichetto et al. 2023). Using this method, the species' niche would be best typified by the most common values. For instance, in Europe, a species that favors average temperatures is more likely to spread to temperate areas than to high latitudes and mountain tops, as the former are more abundant. Sampling fully randomly in environmental space would produce less accurate models for the edge of the species' potential distribution, but one could hypothesize that the practical use of sampling fully randomly would be greater in the case of generalist species.

Therefore, it is possible to tailor background points to the studied species, using ecological insights. Background points are the first stochastic step in the SDM process. Therefore, tailoring them to the species is the first step on the way to introducing causal insight into the correlative process of SDMs. At this stage, we refer again to Figure 1.

Step 2, data preparation, is where background point selection takes place in presence-only species distribution models. They must be selected with a degree of stochasticity, which is necessary to avoid random associations and have models be unbiased (Zhu 2023). Some stochasticity is needed for background sampling. As such, the outputs of SDMs are subject to statistical probability. Therefore, insensitivity of SDM predictions to stochasticity (i.e., stability of predictions) is crucial, as well as their accuracy.

However, introducing ecological knowledge and theory whilst still retaining the necessary randomness can also be done by sampling background points (Chapman et al. 2019). Sampling background points environmentally can allow for connecting SDMs to ecological theory. For instance, sampling them stratified randomly in environmental space would be prudent for specialist species at the edge of the available environment in the area (such as mountains species or Arctic species) and sampling fully randomly in environmental space might be better in the case of generalist species and IAS.

- Occurrence data: citizen science and sampling bias

A final important limitation is the availability and reliability of occurrence data. Scientists typically sample according to standardized and non-biased protocols, but these are insufficient for large-scale conservation efforts (Theobald et al. 2015). However, in recent decades, increasingly large bodies of species occurrence data collected by nonscientists, i.e., citizen science data, have become available (Bird et al. 2014). The emergence of citizen science datasets has opened the possibilities of much larger scale SDM studies (Bird et al. 2014).

Citizen science datasets, however, suffer from many shortcomings (Graham et al. 2004, Anderson et al. 2020). For instance, the competence of the collectors is not guaranteed. Moreover, the sampling is biased in many regards. Citizen science occurrence data are typically sampled close to infrastructure, are very rare in many countries, are more abundant in low-altitude areas, favor specific times of the week and year, and oversample some taxa whilst undersampling others (Bird et al. 2014, Callaghan et al. 2019, Anderson et al. 2020, Hughes et al. 2021, Díaz-Calafat et al. 2024).

When the nature of the sampling bias is known, it is usually straightforward to eliminate it by thinning the data accordingly (Steen et al. 2019b). For instance, if the data are oversampled close to roads and cities, the data can be thinned close to those areas and not further away from there. However, when using citizen science data, the exact nature of the bias is almost never known (Di Cecco et al. 2021), and as a result, compensating for the bias might also eliminate relevant ecological information (Steen et al. 2021). For instance, some plant species occur close to and inside of cities because they were kept as decorative plants and subsequently escaped (Saul et al. 2017). In this case, thinning the occurrence data close to the cities would eliminate this important mechanism of spreading.

As a result, undersampling or false treatment of occurrence records can truncate the niche of a species, potentially causing SDMs, as fully correlative models, to identify false patterns.

Beyond SDMs: Probability of establishment and the Niche Margin Index

To address the paucity of available methods to assess probability of invasion, a dimension not included in SDMs, Broennimann et al. (2021) designed a niche-based approach called the Niche Margin Index (NMI). This measure quantifies the distance of the climatic conditions in an invaded site to the closest margin of the climatic niche measured in the native range of an IAS. Positive values mean the conditions at the site are within the native niche (i.e niche innerness), a value of zero means that the site is at the native niche margin, whilst negative values mean the conditions at the site are outside the native niche (i.e., niche outerness). SDMs typically predict value within the niche, and thus account for the former, but they always predict zero outside the niche, and thus cannot predict the latter. Therefore, the NMI can provide novel insights into how much a site can be invaded even if it lies outside the native niche.

Broennimann et al. (2021) proved that the NMI is positively correlated with invasion success of mammals when the native range is quantified using expert maps of the native range of the species published by IUCN. However, such maps may overestimate the native range for plant species. Therefore, whilst the NMI is a niche-based approach that allows for assessing the probability of establishment of IAS for many species at once in large ranges, additional research is required to assess its usability for plants (and other species groups).

This thesis: the knowledge gaps addressed

The work presented here focuses on addressing several knowledge gaps remaining in those scientific methods that have the unifying goal of creating conservation tools to aid the struggle against IAS:

1) What are the differences in environmental variables most effective in building SDMs for native and non-native species?
2) Does sampling background points in environmental space improve SDM performance?
3) Does the optimal method of background point sampling differ between species that are at equilibrium and species that are at disequilibrium (e.g., IAS)?
4) How can we quantify the susceptibility of an area to invasion? In other words, how do we add the dimension of native niche outerness, and not just innerness, to the probability that a certain site will be invaded once an IAS arrives there?

Thesis outline

Following the above theory and remaining unanswered questions and data/method gaps, this thesis is structured as follows:

Chapter 1 addresses the issue that before the problems caused by IAS can be tackled, it is important to realize that many such species do not start out as invasive. A species can only be labeled as such when the damage it causes has been demonstrated and therefore, remediation is often the only option (Steen et al. 2024a). Therefore, this chapter does not limit itself to only invasive species, but studies a variety of non-native plant species, regardless of invasiveness. In this first chapter, SDMs were used to compare the environmental variables that most powerfully predict the distributions of 141 plant species non-native to Switzerland to those of a same number of their native congeners. This was made possible through the development of a specialized R package (CovSel; Adde et al. 2023). We found that variables associated with human alteration of the environment and proximity to human infrastructure more powerfully predicted the distributions of the non-native species, though climatic variables were the most powerful predictors in both native and non-native SDMs.

Chapter 2 deals with how varying the strategy of background point sampling affects the accuracy and stability of SDM outputs. Background points allow the model to make a statistical approximation of the realized niche of the species by contrasting the environmental values at the locations where the species is present to the environmental values available in the study area. Since a species' niche is an environmental hyperspace, sampling environmentally is theoretically a preferable alternative of the typical geographically random sampling. In this chapter, we present novel ways of sampling background points stratified randomly and fully randomly in environmental space and investigate their effects on model accuracy and stability.

Chapter 3 is a practical application of the clues that have been gathered over the previous two chapters for a single IAS, *Heracleum mantegazzianum*. This chapter also employs the hierarchical/nested approach to avoid niche truncation and elaborates specifically on the steps of ecological variable selection, dealing with sampling bias in the occurrence data, the stratified random selection of background points, tuning the SDMs, performing overfitting analyses, and making future projections.

Chapter 4 seeks to add a dimension to SDM outputs, namely the probability of establishment by IAS when they arrive. The chapter uses the NMI to assess if invasion success of plants can be predicted by the degree of similarity between the native realized climatic niche and the realized climatic niche in the invaded range. We investigated multiple methods to quantify the native range, namely using only occurrence points in the native range of the species and the values within a thresholded SDM, and compared the acquired NMI values between the treatments.

References

Adde, A., Rey, P.-L., Brun, P., Külling, N., Fopp, F., Altermatt, F., Broennimann, O., Lehmann, A., Petitpierre, B., Zimmermann, N. E., Pellissier, L., & Guisan, A. (2023). N-SDM: A high-performance computing pipeline for Nested Species Distribution Modelling. Ecography, 2023(6), e06540. https://doi.org/10.1111/ecog.06540

Adde, A., Rey, P.-L., Fopp, F., Petitpierre, B., Schweiger, A. K., Broennimann, O., Lehmann, A., Zimmermann, N. E., Altermatt, F., Pellissier, L., & Guisan, A. (2023). Too many candidates: Embedded covariate selection procedure for species distribution modelling with the covsel R package. Ecological Informatics, 75, 102080. https://doi.org/10.1016/j.ecoinf.2023.102080

Anderson, R. P., Araújo, M. B., Guisan, A., Lobo, J. M., Martínez-Meyer, E., Peterson, A. T., & Soberón, J. M. (2020). Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. Frontiers of Biogeography, 12(3). https://doi.org/10.21425/F5FBG47839

Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33(10), 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x

Araújo, M.B., Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. Biological Conservation 96(3):331–345. https://doi.org/10.1016/S0006-3207(00)00074-4

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

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(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barbet-Massin, M., Rome, Q., Villemant, C., & Courchamp, F. (2018). Can species distribution models really predict the expansion of invasive species? PLOS ONE, 13(3), e0193085. https://doi.org/10.1371/journal.pone.0193085

Bhowmik, P. C., Chandran, R. S. (2015) 'Biology, ecology, distribution and current status of Heracleum mantegazzianum Sommier & Levier', Journal of Crop and Weed, Vol. 11(1), pp. 1–17

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., & Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. Biological Conservation, 173, 144–154. https://doi.org/10.1016/j.biocon.2013.07.037

Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T., & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. Ecology Letters, 10(8), 701–709. https://doi.org/10.1111/j.1461-0248.2007.01060.x

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(4), 481–497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Broennimann, O., Petitpierre, B., Chevalier, M., González-Suárez, M., Jeschke, J. M., Rolland, J., Gray, S. M., Bacher, S., & Guisan, A. (2021). Distance to native climatic niche margins explains establishment success of alien mammals. Nature Communications, 12(1), Article 1. https://doi.org/10.1038/s41467-021-22693-0

Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B., & Major, R. E. (2019). Improving big citizen science data: Moving beyond haphazard sampling. PLOS Biology, 17(6), e3000357. https://doi.org/10.1371/journal.pbio.3000357

Cavender-Bares, J. M., Nelson, E., Meireles, J. E., Lasky, J. R., Miteva, D. A., Nowak, D. J., Pearse, W. D., Helmus, M. R., Zanne, A. E., Fagan, W. F., Mihiar, C., Muller, N. Z., Kraft, N. J. B., & Polasky, S. (2022). The hidden value of trees: Quantifying the ecosystem services of tree lineages and their major threats across the contiguous US. PLOS Sustainability and Transformation, 1(4), e0000010. https://doi.org/10.1371/journal.pstr.0000010

Chapman, D., Pescott, O. L., Roy, H. E., & Tanner, R. (2019). Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. Journal of Biogeography, 46(5), 1029–1040. https://doi.org/10.1111/jbi.13555

Clarke, D. A., Palmer, D. J., McGrannachan, C., Burgess, T. I., Chown, S. L., Clarke, R. H., Kumschick, S., Lach, L., Liebhold, A. M., Roy, H. E., Saunders, M. E., Yeates, D. K., Zalucki, M. P., & McGeoch, M. A. (2021). Options for reducing uncertainty in impact classification for alien species. Ecosphere, 12(4), e03461. https://doi.org/10.1002/ecs2.3461

Crooks, J. A., & Rilov, G. (2009). The Establishment of Invasive Species. In G. Rilov & J. A. Crooks (Eds.), Biological Invasions in Marine Ecosystems: Ecological, Management, and Geographic Perspectives (pp. 173–175). Springer. https://doi.org/10.1007/978-3-540-79236-9_9

Dejene, S. W., Mpakairi, K. S., Kanagaraj, R., Wato, Y. A., & Mengistu, S. (2021). Modelling continental range shift of the African elephant (Loxodonta africana) under a changing climate and land cover: Implications for future conservation of the species. African Zoology, 56(1), 25–34. https://doi.org/10.1080/15627020.2020.1846617

Definitions: Wild, native or alien? – Botanical Society of Britain & Ireland. (n.d.). Retrieved December 13, 2024, from https://bsbi.org/definitions-wild-native-or-alien

Díaz-Calafat, J., Jaume-Ramis, S., Soacha, K., Álvarez, A., & Piera, J. (2024). Revealing biases in insect observations: A comparative analysis between academic and citizen science data. PLOS ONE, 19(7), e0305757. https://doi.org/10.1371/journal.pone.0305757

Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., & Hurlbert, A. H. (2021). Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. BioScience, 71(11), 1179–1188. https://doi.org/10.1093/biosci/biab093

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(12), 2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.x

EASIN - European Alien Species Information Network—EASIN in a Nutshell. (n.d.). Retrieved December 13, 2024, from https://easin.jrc.ec.europa.eu/easin/EASINNutshell

Enders, M., Havemann, F., Ruland, F., Bernard-Verdier, M., Catford, J. A., Gómez-Aparicio, L., Haider, S., Heger, T., Kueffer, C., Kühn, I., Meyerson, L. A., Musseau, C., Novoa, A., Ricciardi, A., Sagouis, A., Schittko, C., Strayer, D. L., Vilà, M., Essl, F., Hulme, P.E., Van Kleunen, M., Kumschick, S., Lockwood, J.L., Mabey, A.L., McGeoch, M.A., Palma, E., Pyšek, P., Saul, W.-C., Yannelli, F. A., Jeschke, J. M. (2020). A conceptual map of invasion biology: Integrating hypotheses into a consensus network. Global Ecology and Biogeography, 29(6), 978–991. https://doi.org/10.1111/geb.13082

Engler, R., Guisan, A., Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J Appl Ecol 41(2):263–274. https://doi.org/10.1111/j.0021-8901.2004.00881.x

Foster, S. L., Kharouba, H. M., & Smith, T. W. (2022). Testing the assumption of environmental equilibrium in an invasive plant species over a 130 year history. Ecography, 2022(12), e12933. https://doi.org/10.1111/ecog.06284

Franklin, J. (1995) Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography,* **19,** 474-499.

Franklin, J. (2010). Moving beyond static species distribution models in support of conservation biogeography. Diversity and Distributions, 16(3), 321–330. https://doi.org/10.1111/j.1472-4642.2010.00641.x

Franklin, J. (2023). Species distribution modelling supports the study of past, present and future biogeographies. Journal of Biogeography, 50(9), 1533–1545. https://doi.org/10.1111/jbi.14617

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(11), 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x

Gioria, M., Pyšek, P., & Osborne, B. A. (2018). Timing is everything: Does early and late germination favor invasions by herbaceous alien plants? Journal of Plant Ecology, 11(1), 4–16. https://doi.org/10.1093/jpe/rtw105

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 in Ecology & Evolution, 19, 497-503.

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: Insights from biological invasions. Trends in Ecology & Evolution, 29(5), 260–269. https://doi.org/10.1016/j.tree.2014.02.009

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., Schwartz, M.W., Kearney, M.R., Possingham, H.P., Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. Ecology Letters, 16(12), 1424–1435. https://doi.org/10.1111/ele.12189

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). Habitat Suitability and Distribution Models: With Applications in R. https://doi.org/10.1017/9781139028271

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. Ecological Modelling, 135(2), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Jump, A. S., & Peñuelas, J. (2005). Running to stand still: Adaptation and the response of plants to rapid climate change. Ecology Letters, 8(9), 1010–1020. https://doi.org/10.1111/j.1461-0248.2005.00796.x

Haider, Sylvia. (2011). Altitudinal distribution of non-native plants: the effects of climate, habitat and introduction history. Doctoral thesis, Technical University München.

Half of World's GDP Moderately or Highly Dependent on Nature, Says New Report. (n.d.). World Economic Forum. Retrieved December 13, 2024, from https://www.weforum.org/press/2020/01/half-of-world-s-gdp-moderately-or-highly-dependent-on-nature-says-new-report/

Heracleum—Ecospat UNIL. (n.d.). Retrieved December 9, 2024, from https://www.unil.ch/ecospat/en/home/menuinst/research/projects-completed/heracleum.html

Hirzel, A., & Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling, 157(2), 331–341. https://doi.org/10.1016/S0304-3800(02)00203-X

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. Ecography, 44(9), 1259–1269. https://doi.org/10.1111/ecog.05926

Invasive Neophytes. (n.d.). Retrieved December 13, 2024, from https://www.infoflora.ch/en/neophytes/neophytes.html

IPBES (2023). Summary for Policymakers of the Thematic Assessment Report on Invasive Alien Species and their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Roy, H. E., Pauchard, A., Stoett, P., Renard Truong, T., Bacher, S., Galil, B. S., Hulme, P. E., Ikeda, T., Sankaran, K. V., McGeoch, M. A., Meyerson, L. A., Nuñez, M. A., Ordonez, A., Rahlao, S. J., Schwindt, E., Seebens, H., Sheppard, A. W., and Vandvik, V. (eds.). IPBES secretariat, Bonn, Germany. https://doi.org/10.5281/zenodo.7430692

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. Scientific Data, 4(1), 170122. https://doi.org/10.1038/sdata.2017.122

Kourantidou, M., Haubrock, P. J., Cuthbert, R. N., Bodey, T. W., Lenzner, B., Gozlan, R. E., Nuñez, M. A., Salles, J.-M., Diagne, C., & Courchamp, F. (2022). Invasive alien species as simultaneous benefits and burdens: Trends, stakeholder perceptions and management. Biological Invasions, 24(7), 1905–1926. https://doi.org/10.1007/s10530-021-02727-w

Külling, N., Adde, A., Fopp, F., Schweiger, A. K., Broennimann, O., Rey, P.-L., Giuliani, G., Goicolea, T., Petitpierre, B., Zimmermann, N. E., Pellissier, L., Altermatt, F., Lehmann, A., & Guisan, A. (2024). SWECO25: A cross-thematic raster database for ecological research in Switzerland. Scientific Data, 11(1), 21. https://doi.org/10.1038/s41597-023-02899-1

Lyu, S., Alexander, J.M., 2022. Competition contributes to both warm and cool range edges. Nature Communications 13 (1), 1 https://doi.org/10.1038/s41467-022-30013-3.

Maslo, B., Leu, K., Faillace, C., Weston, M. A., Pover, T., & Schlacher, T. A. (2016). Selecting umbrella species for conservation: A test of habitat models and niche overlap for beach-nesting birds. Biological Conservation, 203, 233–242. https://doi.org/10.1016/j.biocon.2016.09.012

Moles, A. T., & Westoby, M. (2006). Seed size and plant strategy across the whole life cycle. Oikos, 113(1), 91–105. https://doi.org/10.1111/j.0030-1299.2006.14194.x

Morera-Pujol, V., Mostert, P. S., Murphy, K. J., Burkitt, T., Coad, B., McMahon, B. J., Nieuwenhuis, M., Morelle, K., Ward, A. I., & Ciuti, S. (2023). Bayesian species distribution models integrate presence-only and presence–absence data to predict deer distribution and relative abundance. Ecography, 2023(2), e06451. https://doi.org/10.1111/ecog.06451

Pauchard, A., Kueffer, C., Dietz, H., Daehler, C.C., Alexander, J., Edwards, P.J., Arévalo, J.R., Cavieres. L.A., Guisan, A., Haider, S., Jakobs, G., McDougall, K., Millar, C.I., Naylor, B.J., Parks, C.G., Rew, L.J., Seipel, T. (2009). Ain't no mountain high enough: Plant invasions reaching high elevations. Frontiers in Ecology and the Environment, 7, 479-486.

Petitpierre, B., McDougall, K., Seipel, T., Broennimann, O., Guisan, A., & Kueffer, C. (2016). Will climate change increase the risk of plant invasions into mountains? Ecological Applications, 26(2), 530–544. https://doi.org/10.1890/14-1871

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecological Applications, 19(1), 181–197. https://doi.org/10.1890/07-2153.1

Polce, C., Cardoso, A. C., Deriu, I., Gervasini, E., Tsiamis, K., Vigiak, O., Zulian, G., & Maes, J. (2023). Invasive alien species of policy concerns show widespread patterns of invasion and potential pressure across European ecosystems. Scientific Reports, 13(1), 8124. https://doi.org/10.1038/s41598-023-32993-8

Porfirio, L. L., Harris, R. M. B., Lefroy, E. C., Hugh, S., Gould, S. F., Lee, G., Bindoff, N. L., & Mackey, B. (2014). Improving the Use of Species Distribution Models in Conservation Planning and Management under Climate Change. PLOS ONE, 9(11), e113749. https://doi.org/10.1371/journal.pone.0113749

Radchuk, V., Reed, T., Teplitsky, C., van de Pol, M., Charmantier, A., Hassall, C., Adamík, P., Adriaensen, F., Ahola, M. P., Arcese, P., Miguel Avilés, J., Balbontin, J., Berg, K. S., Borras, A., Burthe, S., Clobert, J., Dehnhard, N., de Lope, F., Dhondt, A. A., … Kramer-Schadt, S. (2019). Adaptive responses of animals to climate change are most likely insufficient. Nature Communications, 10(1), 3109. https://doi.org/10.1038/s41467-019-10924-4

Ramel, C., Rey, P.-L., Fernandes, R., Vincent, C., Cardoso, A. R., Broennimann, O., Pellissier, L., Pradervand, J.-N., Ursenbacher, S., Schmidt, B. R., & Guisan, A. (2020). Integrating ecosystem services within spatial biodiversity conservation prioritization in the Alps. *Ecosystem Services*, *45*, 101186. https://doi.org/10.1016/j.ecoser.2020.101186

Randin, C. F., Ashcroft, M. B., Bolliger, J., Cavender-Bares, J., Coops, N. C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., … Payne, D. (2020). Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. Remote Sensing of Environment, 239, 111626. https://doi.org/10.1016/j.rse.2019.111626

Rathore, M. K., & Sharma, L. K. (2023). Efficacy of species distribution models (SDMs) for ecological realms to ascertain biological conservation and practices. Biodiversity and Conservation, 32(10), 3053–3087. https://doi.org/10.1007/s10531-023-02648-1

RechAlp. (n.d.). Retrieved December 13, 2024, from https://rechalp.unil.ch/

Rey, P.-L., Külling, N., Adde, A., Lehmann, A., & Guisan, A. (2022). Mapping linkages between biodiversity and nature's contributions to people: A ValPar.CH perspective. ValPar.CH – Values of the Ecological Infrastructure in Swiss Parks, 2022.2, Article 2022.2. https://doi.org/10.5167/uzh-213594

Saul, W.-C., Roy, H. E., Booy, O., Carnevali, L., Chen, H.-J., Genovesi, P., Harrower, C. A., Hulme, P. E., Pagad, S., Pergl, J., & Jeschke, J. M. (2017). Assessing patterns in introduction pathways of alien species by linking major invasion data bases. Journal of Applied Ecology, 54(2), 657–669. https://doi.org/10.1111/1365-2664.12819

Schwindt, E., August, T. A., Vanderhoeven, S., McGeoch, M. A., Bacher, S., Galil, B. S., Genovesi, P., Hulme, P. E., Ikeda, T., Lenzner, B., Nuñez, M. A., Ordonez, A., Pauchard, A., Rahlao, S. J., Renard Truong, T., Roy, H. E., Sankaran, K. V., Seebens, H., Sheppard, A. W., … Meyerson, L. A. (2024). Overwhelming evidence galvanizes a global consensus on the need for action against Invasive Alien Species. Biological Invasions, 26(3), 621–626. https://doi.org/10.1007/s10530-023-03209-x

Skočajić, D., & Nešić, M. (2021). Invasive Species: Routes of Introduction, Establishment, and Expansion. In W. Leal Filho, A. M. Azul, L. Brandli, A. Lange Salvia, & T. Wall (Eds.), Life on Land (pp. 571–582). Springer International Publishing. https://doi.org/10.1007/978-3-319-95981-8_66

Song, Y., Xu, G.-B., Long, K.-X., Wang, C.-C., Chen, R., Li, H., Jiang, X.-L., & Deng, M. (2024). Ensemble species distribution modeling and multilocus phylogeography provide insight into the spatial genetic patterns and distribution dynamics of a keystone forest species, Quercus glauca. BMC Plant Biology, 24(1), 168. https://doi.org/10.1186/s12870-024-04830-1

A Steen, B., Cardoso, A.C., Tsiamis, K., Nieto, K., Engel, J., Gervasini, E. (2019) Modelling hot spot areas for the invasive alien plant Elodea nuttallii in the EU. Management of Biological Invasions 10(1): 151–170, https://doi.org/10.3391/mbi.2019.10.1.10

B Steen, V. A., Elphick, C. S. & Tingley, M. W. (2019) An evaluation of stringent filtering to improve species distribution models from citizen science data. Diversity and Distributions, 25, 1857-1869.

A Steen, B., Adde, A., Schlaepfer, M. A., Guisan, A., & Maiorano, L. (2024). Distributions of non-native and native plants are not determined by the same environmental factors. *Ecological Solutions and Evidence*, *5*(4), e12374. https://doi.org/10.1002/2688-8319.12374

B Steen, B., Broennimann, O., Maiorano, L., & Guisan, A. (2024). How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels. Ecological Modelling, 493, 110754. https://doi.org/10.1016/j.ecolmodel.2024.110754

Steen, V. A., Tingley, M. W., Paton, P. W. C., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. Methods in Ecology and Evolution, 12(2), 216–226. https://doi.org/10.1111/2041-210X.13525

Stohlgren, T. J., Crosier, C., Chong, G. W., Guenther, D., & Evangelista, P. (2005). Life-history Habitat Matching in Invading Non-native Plant Species. Plant and Soil, 277(1), 7–18. https://doi.org/10.1007/s11104-005-4893-5

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M. A., & Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation, 181, 236–244. https://doi.org/10.1016/j.biocon.2014.10.021

Tulloch, A. I. T., Sutcliffe, P., Naujokaitis-Lewis, I., Tingley, R., Brotons, L., Ferraz, K., Possingham, H. P., 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. https://doi.org/10.1016/j.biocon.2016.04.023

Václavík, T., & Meentemeyer, R. K. (2009). Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? Ecological Modelling, 220(23), 3248–3258. https://doi.org/10.1016/j.ecolmodel.2009.08.013

Vetaas, O.R., 2002. Realized and potential climate niches: a comparison of four Rhododendron tree species. Journal of Biogeography 29 (4), 545–554.

Vittoz, P., & Engler, R. (2007). Seed dispersal distances: A typology based on dispersal modes and plant traits. Botanica Helvetica, 117(2), 109–124. https://doi.org/10.1007/s00035-007-0797-8

Von Daniken, I., Guisan, A. & Lane, S. (2014) Une nouvelle plateforme UNIL de support pour la recherche transdisciplinaire dans les Alpes vaudoises. Bulletin des société vaudoise des sciences naturelles, pp. 175-178.

Which Taxa Are Alien? Criteria, Applications, and Uncertainties | BioScience | Oxford Academic. (n.d.). Retrieved December 13, 2024, from https://academic.oup.com/bioscience/article/68/7/496/5050532

Wisz, M. S., & Guisan, A. (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecology, 9(1), 8. https://doi.org/10.1186/1472-6785-9-8

Zanatta, F., Engler, R., Collart, F., Broennimann, O., Mateo, R. G., Papp, B., Muñoz, J., Baurain, D., Guisan, A. & Vanderpoorten, A. (2020) Bryophytes are predicted to lag behind future climate change despite their high dispersal capacities. Nature Communications, 11, 5601.

Zhu, J. (2023). Automated Machine Learning: Intellient Binning Data Preparation and Regularized Regression Classfier. Electronic Theses and Dissertations, 2020-2023. https://stars.library.ucf.edu/etd2020/1706

CHAPTER I

Distributions of non-native and native plants are not determined by the same environmental factors

Bart Steen*[1,2], Antoine Adde[3], Martin A. Schlaepfer**[4], Luigi Maiorano**[2], Antoine Guisan**[1,5]

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

[2] Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, viale dell'Università 32, 00185 Roma, Italy

[3] Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133 CH-8600 Dübendorf

[4] Institute for Environmental Sciences, Université de Genève, 66 Boulevard Carl-Vogt, CH-1211 Genève 5

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

* Corresponding author: bart.steen@unil.ch / b.r.l.steen2@gmail.com

** Shared last authorship

This chapter was published in Ecological Solutions and Evidence.

DOI: 10.1002/2688-8319.12374

My contribution to the project: I analyzed the data, drew the figures and led the writing of the manuscript.

Abstract

1. Global environmental change will cause shifts in species communities, with non-native species likely replacing native ones at an unprecedented rate. This will have consequences for biodiversity and ecosystem services, in addition to the ecological and economic damage caused by those non-native species that are invasive. Understanding general patterns driving distributions of native and non-native species is therefore vital, but no study has compared yet whether environmental variables that correlate with a species' presence differ between the two groups other than at local scale and often with very limited sample size.
2. In this study, we focus on 141 native and non-native congeneric plant species pairs at the scale of Switzerland. In the framework of correlative species distribution models, we used newly developed methods for efficient automated selection of a parsimonious number of predictor environmental variables to determine which ones, out of a large candidate set in eight classes, have the strongest explanatory power for both species groups.
3. Our results indicated that variables influence the two groups in significantly different ways. Climate was by far the strongest determinant of both native and non-native species distributions, although it had significantly more explanatory power in native species models.
4. Models for non-native species were significantly more influenced by anthropogenic factors, land use variables, and forest cover. The presence of non-native species was also associated with habitats with a significantly lower mean naturality value than native species.
5. These findings provide novel empirical evidence for the different environmental factors driving native and non-native plants' distributions and guidance for non-native species management.
6. Practical implications: SDMs are an increasingly frequently advised tool for conservation management and our results provide guidelines on which covariates should be specifically considered to assess the habitat suitability of non-native versus native species. The distributions of the former group of species are particularly important to research, as, in time, they may turn invasive. In addition, areas close to infrastructure should be scanned regularly for incipient colonizations by non-native species, especially in as yet uninvaded areas, such as high mountains.

Introduction

Understanding the environmental factors that determine a species' distribution is a foundational step in ecology (Brown et al. 1996). Species' distributions are shaped by abiotic and biotic factors, and their spatio-temporal variations (Pulliam 2000). Humans have significantly modified both the biosphere and geosphere (Steffen et al. 2018), for example by modifying land-cover, altering the climate, and moving species outside of their historical ranges (IPBES 2019). This begs the question whether these factors, and which ones, shape current species distributions (Thuiller et al. 2008, Guisan et al. 2013). In this regard, the introduction of species outside of their historical ranges can be used as a large-scale "natural experiment" to investigate what drives changes in species distributions (Yoshida et al. 2007, Colautti & Lau 2016). It can provide insight into the different pressures that are being put on native communities, and the pressures experienced by non-native species in a new environment. In addition, it can reveal how and where populations will change in the modern, heavily globalized world. Species are labelled as *non-native* (alternatively as introduced, alien, or exotic) when they have spread beyond their natural home range, i.e., whose colonization of new regions has been influenced by humans (Soto et al. 2024). Non-native species are further labelled as invasive if their presence is associated with clear economic losses, undesirable human health impacts (e.g., allergenicity) or are perceived to harm native biodiversity (Pyšek et al. 2020, FOEN 2022, InfoSpecies, n.d.-a, n.d.-b).

Biologists generally study invasive species either to observe evolutionarily novel situations, or to understand how to anticipate and mitigate their undesirable effects (Guisan et al. 2014, Van Kleunen et al. 2018). The overwhelming majority of non-native species, however, are not considered invasive (European Environment Agency 2023). For instance, they represent only 15% of non-native plant species in Switzerland (Infoflora 2022). As much focus is put on the invasive fraction, less is generally known about introduced non-native species ecology (where they are established) and their functions (how they interact with other biotic and abiotic components of their ecosystem). Such questions are important both to better predict the probability of a non-native species becoming invasive but also to consider their potentially harmful effects on ecosystems or, reversely, their potentially desirable contributions to ecosystem resilience and the contributions to human well-being (Davis et al. 2011, Sax et al. 2022). However, it is important to note that species that are invasive don't start out as such. It can take decades before a taxon displays the typical characteristics of an invasive species (Robeck et al. 2024). A species therefore is labeled as invasive only when the damage is already ongoing, and thus remediation is often the only management option on the table. Investigating whether non-native and native species share (or not) similar properties can thus be instrumental to shift from remediation to prevention, which is a preferable alternative. It is therefore important to investigate whether non-native and native species share (or not) similar properties (e.g., climatic niche, colonization capacity, ability to tolerate human disturbances) (Davis et al. 2011; Pyšek et al. 2020), and to not limit ourselves to only researching non-native species that are already labeled as invasive. In this regard, Species Distribution Models (SDMs; Guisan et al. 2017) are important tools to identify, in geographic space, relationships between species observations (presence and sometimes also absence; i.e., the response variable) and attributes of the abiotic environment (i.e., the covariates), such as mean annual temperature or land use types. They are popular tools in conservation planning and ecological studies (Barbet-Massin et al. 2012, Guisan et al. 2013, Araujo et al. 2019, Low et al. 2021), despite being underused in practice (Tulloch et al. 2016). By identifying key environmental factors associated with a species' distribution, SDMs can be used as explanatory tools to understand the relationship between the distributions of species and the environment, for instance to assess whether environmental drivers differ between native and non-native species.

Bioclimatic factors have a strong influence on both native and non-native species occurrence and establishment (Zimmermann, 2009, Petitpierre et al. 2016, Broennimann et al. 2021). In addition, many studies have established a statistical association between presence of non-native species and anthropogenic factors (Pyšek et al. 2020), including human population density, global trade network connectivity and gross domestic product (Chapman et al. 2017, Lembrechts et al. 2017, Zhou et al. 2020, Bonnamour et al. 2021). Furthermore, anthropogenically disturbed areas are more likely to be colonized by non-native species (Oshima & Takahashi 2020, Rojas-Sandoval et al. 2024), particularly by invasive plants (Meyer et al. 2021), due to opened niches and probable excess of nutrient availability, among other things. At a local scale, the occurrence for a non-native population is further conditioned by the presence of nearby population sources (Froese et al. 2019, Cook et al. 2021). All these latter factors, however, can also potentially determine the distribution of native species, so that it remains unclear whether similar or different environmental factors determine the distribution of native versus non-native species.

However, despite the proven effectiveness of SDMs in conservation planning (Guisan et al. 2013, Eckert et al. 2020), they suffer from a well-known potential drawback: the niche of the modelled species is assumed to be at equilibrium with the environment. This assumption is not satisfied for at least some non-native species, since their colonization of new areas is progressing and their niche is still expanding (Broennimann & Guisan 2008, Carlin et al. 2022). The assumption may however hold up better for some native species (Normand et al. 2011, Steen et al. 2024). This is another reason why it is important to compare the processes driving the distributions of native and non-native species. Such changes can be analyzed under the COUE (centroid shift, overlap, unfilling, expansion) framework postulated by Guisan et al. (2014). This bring us to another reason SDMs are interesting tools to study this concept: they allow for assessing whether there is niche overlap between native and introduced species. Niche overlap would imply competition, whilst niche divergence would mean possible co-existence.

Here, we take advantage of a large set of existing SDMs previously generated by a modelling pipeline that contains an automatic selection of the environmental variables (Adde et al. 2023a,b) to test, whether the set of factors finally selected in the model differs between non-native exotic species and their native congeners. One congener was selected randomly for each exotic species. We focused on congeneric species pairs in order to mitigate as much as possible all confounding factors existing between the target native and non-native species. Using congenerics, the major difference between the two species groups should be in their degree of nativeness. Other factors like differing growth speeds, different ability to generate propagules, different shade tolerance, etc. are theoretically comparable in our paired sampling design.

Specifically, we formulate and test the following hypotheses:

1) Native and non-native species' distributions are driven by different environmental factors.

2) Non-native species favor areas that are in close proximity to human infrastructures.

3) Non-native species' distributions are more strongly driven by anthropogenically altered natural environments.

Materials and Methods

SDMs of Swiss plant species were fit using the new N-SDM multi-species modelling software (Adde et al. 2023a) that includes an automated procedure to select the best covariates (i.e., environmental factors) out of a panel of numerous candidate ones (Adde et al. 2023b). Of these outputs, pairs of congeneric native and non-native species were chosen for comparative analysis and the environmental variables were divided into eight different classes (Table 1). We analyzed in a paired way how often these variable classes were selected in both species groups, and for the selected variables what was their average importance. In addition, we made an ordinal measure of naturality of land use/land cover factors and compared the naturality scores between both species groups.

Study area

Switzerland is home to approximately 2’800 known plant species (Flora Helvetica), of which 730 are considered non-native (FOEN 2022). The country is characterized by sharp altitudinal gradients, and a large variability in climate (Petitpierre et al. 2016). Major habitats include grasslands, montane and subalpine forests (with coniferous forests spreading to higher altitudes than deciduous ones) and fully unvegetated areas at the highest elevations (Schwarz et al. 2004, Scherrer et al. 2020).

Species distribution models

SDMs were previously constructed for all possible plant species in Switzerland with enough occurrence data for model fitting and using an efficient automatic selection process to identify the subset of the 833 initial environmental variables that have the strongest association with a species’ presence (Adde et al. 2023b). For each species, SDMs were fitted at two levels (continental and regional) in a nested design to prevent niche truncation (Chevalier et al. 2021, Adde et al. 2023a). At the continental level (here Europe), a bioclimatic SDM was fitted to capture the whole species’ climatic niche (Chevalier et al. 2022). This output was then clipped to the regional-level extent (here Switzerland) and used it as covariate for the regional-level model, as similarly done by Chevalier et al. (2022). The continental model output (i.e. bioclimatic envelope) extracted for fitting the regional-level model was the only covariate that was always forced to be included in the final set of covariates. All other variables were candidate predictors and selected automatically using the *covsel* R package (Adde et al. 2023b; see next section).

For the continental-level model, species observation data were retrieved from the Global Biodiversity Information Facility (GBIF; https://doi.org/10.15468/dl.zwp3dx; accessed on the 27th of October 2021). For the regional-level model, data from the Infospecies database (www.infospecies.ch), the Swiss center of species information, were used. At both levels, data were retrieved for the time period of 1980-2020 and the occurrence records were spatially disaggregated using a minimal distance of one km between two points at the continental level, and 200 m at the regional level.

Environmental covariates

Environmental covariates used to fit SDMs were different at each level. For the continental level, we used CHELSA version 2.1 (Karger et al. 2018) bioclimatic variables available at a resolution of one km^2. For the regional level, 67 candidate covariates were retrieved from the SWECO25 database (Külling et al. 2024).

The variables fall into eight main classes (Table 1). The full list of covariates is presented in Supporting Information (Table S2).

Table 1: The classes of environmental variables used to fit the SDMs at the Swiss-scale (Kulling et al. 2024).

Class	**Examples**	**Level**	**Number of candidates**
Bioclimatic envelope	Output from continental scale model	Continental	19
Edaphic	Nitrogen and phosphorus content, soil moisture variability	Regional	8
Hydrological (hydro)	Distance to lake, distance to river, slope. Baseflow index	Regional	10
Topographic (topo)	Aspect, hill shade, slope	Regional	3
Human population density (popdensity)	Humans per 25m.	Regional	1
Transport	Noise pollution index, distance to road/ topographic distance to road	Regional	3

Vegetation (vege)	canopy cover, proportion of different forest types	Regional	6
Land use / land cover (LULC)	forest, agriculture, farm pastures, scattered fruit trees	Regional	39

Edaphic factors are soil related variables, such as pH, nutrient content, moisture, and aeration. They were included into the study because they strongly influence the distributions of plants species and have particularly been proven to affect alien plant abundance (Bigelow & Canham 2002, Collette & Pither 2015, but Ender et al. 2017). The effect may be particularly strong at our local scale with a spatial resolution of 25X25 m, since soil variable vary strongly over small distances. Hydrological factors relate the lake/river flow and ground water. They include factors related to river morphology, river catchment discharge, distance to water and snowmelt. They are important to consider, since some of our target species are water-dependent (see Supporting Information Table D).Topographic factors are mean, median, minimum and maximum hillshade, aspect (orientation) and steepness of the slope. They are related to important determinants of habitat suitability for plants, like light availability (Charbonneau & Fahrig 2004) and potential erosion. The latter has been proven to affect species composition in the Swiss Alps (Huck et al. 2013). Human population density is both a single variable and a class. It is the number of inhabitants per 25 m pixel; an important variable to consider, since human disturbances are strong possible drivers of non-native species distributions (Pyšek et al. 2020, Zhou et al. 2020). The different transport variables are noise pollution index, Euclidean distance to roads and topographic distance to roads (meaning the shortest topographic path). These are very important variables, as they represent the main avenues for seeds of non-native plants to spread (Pyšek et al. 2020, Lembrechts et al. 2017), though we use noise pollution as a proxy for anthropogenic activties. Vegetation variables are: canopy cover (minimum, maximum and medium), proportion of coniferous forest and proportion of deciduous forest. These proportions are calculated using a moving window (Kulling et al. 2024), the size of which can range from 25 m to 5000 m and is chosen automatically based on highest explanatory power by the *covsel* R package (Adde et al. 2023b). These moving windows were also used for the land use/ land cover (LULC) class (Adde et al. 2023). Examples of the many variables in the LULC class are devastated forest, forest fresh cuts and mountain meadows and pastures. All these environmental factors affect light conditions, which has been shown to be a strong influencing factor of alien plant abundance (Charbonneau & Fahrig 2004). In addition, forest disturbances have been shown to be positively correlated with alien plant invasion (Oshima & Takahashi, 2020).

Automated covariate selection was done by using the *covsel* R package (Adde et al. 2023b), whose outputs include a list of environmental variables that are statistically associated with each species' presences (across all species), a score of their relative importance in the model (table 2). The variable that best explains each species distribution is assigned an importance value of 1, and the others are assigned the appropriate fraction. The full dataset used in this study is available online (Adde, 2024).

Selection of species pairs

We established pairs of congeneric plant species composed of a native and a non-native to control for ecological characteristics that could be attributed to a shared evolutionary history and to obtain an equal number of units in each group. First, we identified non-native species with at least one native congener. A random draw (function 'sample' in R) selected only one native species in cases where multiple native congeners existed. This resulted in 141 native – non-native species pairs, of which 15 (11%) are invasive, a percentage similar to the ratio in all non-native plant specie sin Switzerland (15%) (Infoflora 2022).

Table 2: Covariates selected for the species pairs Bidens cernua *(native) and* Bidens frondosa *(non-native)*, Heracleum mantegazzianum *(non-native) and* Heracleum sphondylium, *(native),* Vinca major *(non-native), and* Vinca minor *(native) their average importance and the naturality score of the selected variables belonging to the LULC class*.

Species modelled	**Variables selected**	**Average variable importance**	**Naturality score of LULC variables**
Bidens cernua	bioclimatic envelope	0.96	
	distance to lake	0.46	
	settlement urban amenities cover	0.32	1
	alpine pastures cover	0.31	4
	soil moisture	0.28	
	slope	0.24	
	soil ph	0.2	
Bidens frondosa	bioclimatic envelope	1.00	
	agriculture cover	0.19	2

	settlement urban amenities cover	0.18	1
	other arable land and meadows cover	0.18	3
	soil moisture	0.11	
	soil aeration	0.11	
	unproductive grass and shrubs cover	0.09	4
	euclidean distance to road	0.07	
Heracleum mantegazzianum	bioclimatic envelope	1.00	
	euclidean distance to road	0.22	
	favourable arable land and meadows cover	0.19	3
	grassland meadows cover	0.16	3
	soil moisture	0.14	
	deciduous forest cover	0.13	
	alpine pastures cover	0.13	4
	path distance to road	0.08	
	other arable land and meadows cover	0.07	3
	distance to lake	0.04	
	wetlands cover	0.03	
	remote and steep alpine meadows and pastures cover	0.02	4
Heracleum sphondynium	bioclimatic envelope	1.00	
	soil nutrients	0.25	
	soil moisture variability	0.20	
	hillshade	0.20	
	slope	0.15	
	soil ph	0.12	
	distance to river	0.11	
	soil humus	0.10	

	brush forest cover	0.09	5
	soil moisture	0.08	
	deciduous forest cover	0.08	
	clusters of trees cover	0.03	3
Vinca minor	bioclimatic envelope	1.00	
	other arable land and meadows cover	0.29	3
	soil aeration	0.27	
	soil ph	0.26	
	canopy height	0.24	
	euclidean distance to road	0.21	
	agriculture cover	0.12	2
	intensive agriculture cover	0.10	2
Vinca major	bioclimatic envelope	0.96	
	euclidean distance to road	0.35	
	soil humus	0.21	
	agriculture cover	0.15	2
	coniferous forest cover	0.14	
	deciduous forest cover	0.11	
	forest cover	0.09	5
	settlement urban amenities cover	0.09	1
	soil ph	0.07	
	noise pollution index	0.05	
	open forest cover	0.03	5

Comparing SDM predictor variables between native and non-native plant species

Two approaches were used to compare the environmental requirements of native versus non-native species. First, we tested within each class of variables for differences in the variables automatically selected in the models of the two groups. We analyzed what class the selected variables belonged to and their relative importance (Smith & Santos 2020). In the cases where covariates of a specific environmental variable class were only selected in one of the two congenerics in a pair, the pair was dropped from that particular analysis. This was necessary to keep our paired samples design. The "var.test" function in R was used to verify equality of variances and based on this result, the comparison between pairs was done using the "t.test" function in R, with either equal or unequal variances.

We investigated the three vegetation variables (minimum/maximum/medium canopy cover, proportion of coniferous forest and proportion of deciduous forest) more closely, by analyzing how often they were selected as a covariate and what their average variable importances were for native and non-native species.

A second analysis was devised to test for an association with anthropogenic habitats. Each land use/land cover (LULC) class was assigned an ordinal naturality score. Naturality captures the extent to which a habitat has been modified by human activity. Our definition for each ordinal value (Table 3) and the specific scores attributed to each habitat category (Supporting Information table C) is based on similar attempts in prior studies (Kowarik 1988, Kowarik 1995, Brown & Vivas 2005, Radford 2019).

Table 3: Ordinal classification of naturality score of the land use/land cover variables. 1 is the highest degree of anthropogenic interference, i.e., the least natural and 5 is the lowest degree of anthropogenic interference. Derived from Kowarik 1988, Kowarik, 1995, Brown 2005, Radford 2019.

Naturality Score	Definition	Example
1	Very heavily modified	Impermeable surfaces, such as asphalt or concrete roads, mineral roof tops, artificial turf (sports field)
2	Heavily modified	Permeable surfaces, such as roads and parking lots made of gravel and dirt; intensive agriculture with tilling and use of herbicides and pesticides
3	Moderately modified	Semi-natural habit, such as agro-forestry, urban parks with abundant vegetation and water, gardens, green roofs, cemeteries.
4	Lightly modified	Areas with low density of trails, roads and buildings, or managed for nature-oriented goals such cultural and regulating ecosystem services or biodiversity targets

5	Very lightly or not modified	Areas considered “wild” or lightly managed for unique biodiversity (high species richness, rare and endemic species)

For each SDM, the naturality value was calculated as the mean of all significant LULC categories times their respective naturality scores. Naturality values across species groups were compared with a paired t-test, as with other environmental variables.

As the naturality score of a habitat type likely reflects human activity at least at a coarse scale, and given that the naturality score was established independently of the modelled presence of non-native species, we assumed that the numerical values could be used for testing differences between groups of species.

In order to make our results easier to interpret considering the ecology of the species, we split up our non-native species into several categories considering different characteristics. In particular, we considered invasiveness, degree of naturalization in Switzerland (introduced more or less than 40 years ago), region of origin, growth form according to Raunkiaer, and habitat that the species are bound to (according to Infospecies (2024)).

We tested the significance of the differences in AUC values for each subdivision using a t-test.

Results

SDMs were fitted for 141 non-native plant species that had at least one native and likewise modelled congener (i.e., native species within the same genus) (see Supporting Information Table S1). The number of covariates in each SDM ranged between 5 and 12. Example SDM outputs from a species pair are provided in the supporting information, Figure S1.

After eliminating species pairs in which only one congeneric selected a variable in the class of interest, we had the following numbers of species pairs per class: 141 for the bioclimatic envelope (since it was forced into each model to avoid niche truncation), 139 for land use/land cover variables, 96 for edaphic variables, 53 for transportation variables, 50 for topographic variables, 41 for vegetation variables and 24 for hydrological variables.

Overall, bioclimatic factors had by far the highest variable importance in both species groups (Figure 1). The bioclimatic envelope, i.e., the output of the European-scale model based on bioclimatic variables only, was significantly more important for native species (paired t-test: p = 0.009, Figure 1). Predictor variables from other classes were less important (i.e., had lower significance scores) but could be significantly associated with the distribution of at least some species (both native and non-native). Three classes of predictor variables have significantly greater average importance in non-native species than in native species models: transportation, vegetation, and land use/cover (paired t-tests: p = 0.018, p = 0.007, p = 0.007, respectively; Figure 1), though no noticeable difference in the frequency of selection of these variable groups between native and non-native SDMs could be found (Supporting Information, Table S5).

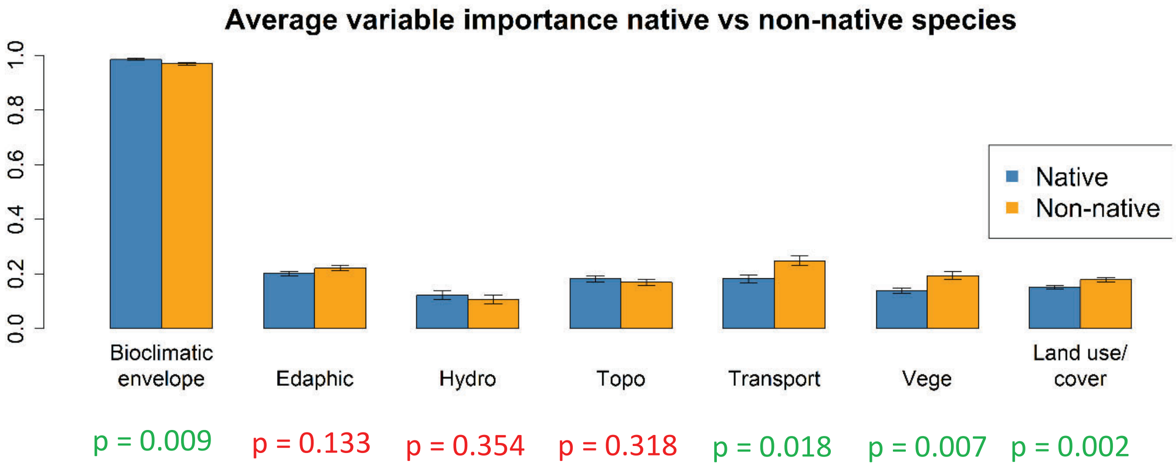


Figure 1: Comparison of average importance of variable classes selected by models built for native species vs models built for non-native species and p-values of paired t-tests. Green values indicate significant differences, red values indicate non-significant differences.

We found no significant differences between native and non-native species models for topographic, or water-related factors (t-tests: p = 0.055, p = 0.716, respectively, Figure 1). No model selected human population density as predictor. However, some models did select proxies for human population density, like the transport variables (Figure 1).

There are only a few variables in vegetation and transportation-based classes. This allows us to use our pairwise analyses to further investigate these variables. The other classes have too many variables and would lead to too many species pairs being eliminated. Even with the topographic distance to roads variable, there were only two species pairs that included the variable (see Figure 2). For the same reason, we had to consider the three canopy cover variables as only one. The only variables that had significantly higher variable importance in one class were the proportion of deciduous and coniferous forest, with a higher average for non-native species (see Figure 2).

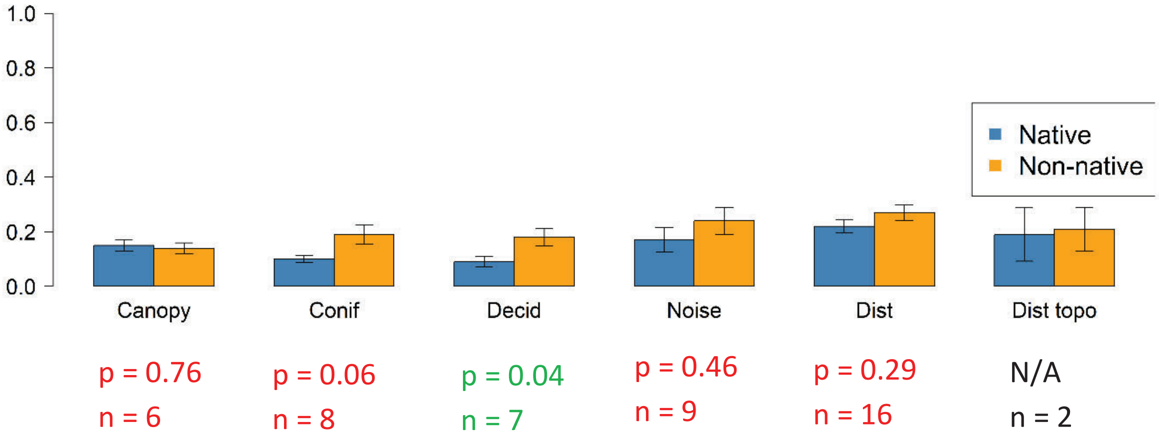


Figure 2: Average variable importance of the three vegetation-based factors and the three transport based factors by both native and non-native models, plus respective sample sizes, with paired t-tests for sample differences. Canopy = percentage of canopy cover, Conif = proportion of coniferous forest, Decid = Proportiono f deciduous forest, Noise = noise pollution index, Dist = Euclidian distance to road,

Dist topo = topographic distance to road. Green values indicate significant differences, red values indicate non-significant differences.

Finally, the average naturality scores of land use/cover variables were significantly higher in native species models (t-test, p = 2.03^{e}-05, see Figure 3).

Naturality score	*Native*	*Non-native*	*p-value (Welch t-test)*
Frequency of selection	*2912*	*2196*	
Avg per species	*3.101171*	*2.826255*	*1.65E-05*

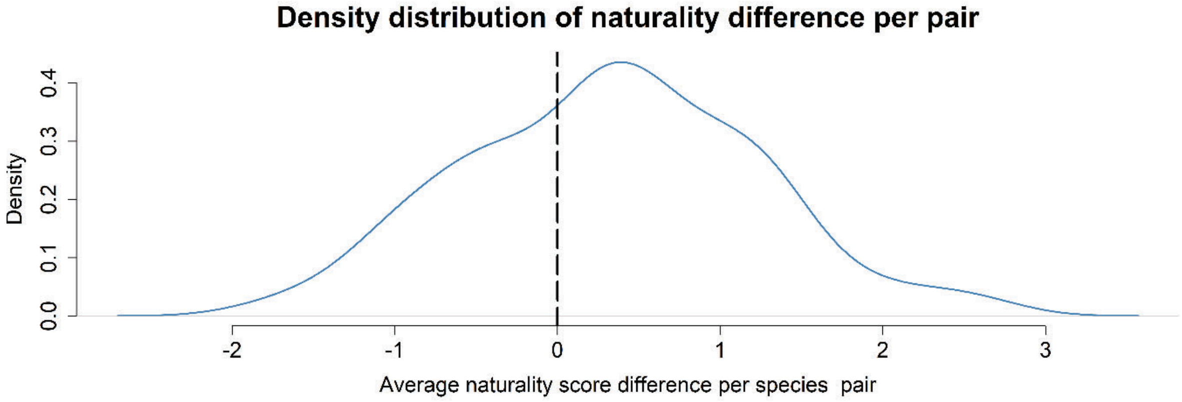


Figure 3: Kernel density distribution of average naturality score of LULC variables per species pair. Calculated by subtracting the average naturality score of the non-native species from that of their matching native congenerics. The average naturality score was 3.10 for native species and 2.77 for non-native species, which was significantly different (paired t-test: p = 2.03e-05)

When dividing the non-native species into invasive and non-invasive species, we found that barring climate, which was of course the most important variable everywhere, edaphic and vegetation factorts were the most important for invasive species, but no significant differences were found between the variable importances of these latter factors. For non-native species, transportation variables had the highest variable importance, again after climate.

When splitting the non-native species into the taxa that have been in Switzerland for over 40 years and the ones that were not, we found that in both cases transportation variables had the most powerful predictors after climate, and the importance was significantly higher for the more recent species (non-paired t-test: p = 0.04381).

We found that vegetation factors had on average the highest variable importance of all groups of environmental factors in models of non-native species from Asia (again after climate). The difference with the second-most important variable group, transportation factors, was not found to be significant (p = 0.5511). For all species from South America, edaphic variables have the highest predictive power (with topographic next and transportation third) (again, after climate). For species from all the other continents (excluding the ones with fewer than 6 species), including Europe itself, the transportation variables had the highest predictive power after climate, but the difference with it and the second most important variable group was never significant.

When analyzing variable importance in different growth forms, we find the same pattern for chamaephytes, phanerophytes, hemicryptophytes, and therophytes: after climate, transportation variable importance is highest, however, it is not significantly higher than other important variables. This is likely a consequence of too low sample size. However, we did find two growth forms that showed a different pattern: For geophytes, the edaphic variables were the most potent predictors, then vegetation variables and finally transportation variables. Again, no significant difference was found between these variable groups. Finally, when putting all woody plants together (woody chamaephytes + phanerophytes), vegetation variables were most important, with transportation variables second (but there was again no significant difference).

Splitting up the non-native species that are bound to certain habitats (76 total) revealed that 40 of them occur in (and many are common or dominant in) pionier vegetation of human-disturbed areas. It was the largest group. Second largest was plantations, fields and crops (16 total) and third were heaths, margins and meadows (12 total).

Discussion

Our results show that, overall, the environmental factors that determine the distribution of both native and non-native species are largely similar, yet with interesting differences. Hypothesis 1 (the distributions of native and non-native plants are determined by different environmental factors) is therefore partly confirmed. In fact, the distribution of species from both groups are correlated first and foremost with bioclimatic factors, suggesting that climatic factors represent a key filter in determining where species are established. However, bioclimatic factors appear slightly less important in non-native species. The significantly higher importance of the bioclimatic envelope for native species (meaning the output of the Europe-scale model of the species, which was constructed using only bioclimatic variables) is what is expected from species that are more likely to be at equilibrium with the environment (Guisan et al. 2014, Early & Sax 2014, Perret et al. 2019). The most likely explanation is that some non-native species are still expanding their range and may not yet have occupied their entire realized niche in the new ranges (i.e. ‘unfilling’ in the COUE framework; Guisan et al. 2014). Anyway, climate matching was still shown to be an important factor for non-native species (e.g. Broennimann et al. 2021 for mammals, Petitpierre et al. 2012 for plants).

Hypothesis 2 (non-native species distributions are more strongly associated with areas close to human infrastructure than native species distributions) is fully confirmed. Variables associated with transportation (e.g., proximity to roads, LULC categories and vegetation (e.g., extent of deciduous canopy cover within a given radius of the species’ observation) indeed are correlated with non-native species’ distributions more than native species’ distributions. This seems to indicate that factors that are not under the influence of human activities (soil characteristics, topography, hydrological characteristics) affect native and non-native psecies distributions equally, and that, by contrast, environmental variables shaped by anthropogenic activities are more likely to be associated with the presence of non-native species . This is consistent with literature on the distribution of non-native species (Essl et al. 2015, Pyšek et al. 2020). Firstly, human transportation plays a great role in the dispersal of non-native species. There are many examples of non-native and invasive species spreading via ship, airplane, train, or motorway (Essl et al. 2015, Bertelsmeier et al. 2017, Mang et al. 2018). This may also explains why human population density was never chosen as a predictor in any model, whilst the transportation variables (distance to roads and noise pollution index) were. This implies that human density itself is a less powerful predictor, which seems logical, since non-native species are often spread by human movement (Capinha et al. 2015). The importance of transportation variables is largely reflected when subdividing the target species by region of origin; species from Europe, North America and Africa all favored transportation variables (though no significant differences were found with other variable groups). Species of Asian origin, conversely, favored vegetation variables more, indicating that these species have already spread beyond the proximity of human infrastructure.

The observation that anthropogenic variables more strongly predict non-native species distributions also conforms to the propagule pressure hypothesis cluster, which is one of the five major ones synthesized by Enders et al. (2020) on what can make non-native species invasive (see also Blackburn et al. 2020, Pyšek et al. 2010). Many of the hypotheses presented by Enders et al. focus on what drives non-native species success and probability of establishment and are therefore relevant to be considered here. The propagule pressure cluster centers around the idea that the more individuals of a non-native species arrive in a new place, the higher the chance that the species will establish itself is.

The naturality values of LULC categories in SDMs of non-native species were, on average, 0.3 lower than LULC categories SDMs of native species. This is supported by the fact that the majority of non-native species were bound to pioneer vegetations in human-disturbed areas, and that the second most represented habitat among them is plantations, fields and crops. These results are consistent with the human commensalism hypothesis (Jeschke and Strayer 2006), which posits a positive association between non-native species with human activities. It also concurs with several case studies on the habitat preferences of invasive species (Steen et al. 2019), with terrestrial invasive plants in particular favoring abandoned agricultural ground (Kuhman et al. 2010, Kuhman et al. 2011).

The breakdown of the non-native species highlighted the the effect of naturalization: variables relating to transportation have significantly more predictive power in species that have been in Switzerland for less than 40 years. This may indicate that whilst niche shifts often happen during colonizations (Broennimann et al. 2008, Carlin et al. 2022), eventually, non-native species move closer to equilibrium with the environment. It might likewise mean that the inherent assumption of SDMs that rely on this equilibrium becomes more justified, the longer a non-native species is resident in a certain range.

This might underline the importance of managing invasive species along human infrastructure, however, in our study, we found that invasive species distributions were more powerfully predicted by edaphic and vegetation-related factors. This may be because our sample size for these species was too small (15 total), but a more likely explanation is that these have long been naturalized; all but two of the species have been in Switzerland for longer than 40 years. As a result, they have already distributed themselves beyond the proximity of human infrastructure.

Whilst no significant differences could be found when analysing variable importance within growth forms, the fact that geophytes and woody plant species show different patterns from anything else is interesting. Geophytes store nutrients and water into subterranean parts (bulbs/corms/rhizomes/tubers; Encyclopedia Brittanica (n.d.)), and therefore, the soil factors are logically powerful predictors. Woody plants are strongly associated with vegetation variables, which include proportion of forests, which is their habitat. Again, it might indicate non-native species' predilection to colonize, as the proportion of forests is indicative of how many edges there are and thus of gap dynamics.

Hypothesis 3 was also confirmed. Indeed, our results show that although overall the same type of variables predict the spatial distribution of both native and non-native plants in Switzerland, small yet significant differences between groups exist in the environmental factors that appear in the SDMs. Specifically, several lines of evidence suggest that non-native species are more likely to occupy habitat types that have been altered by human activities, which is again in agreement with the propagule pressure hypothesis cluster.

The significantly higher average importance of deciduous forest cover variables in non-native species SDMs (proportion of coniferous forest and proportion of deciduous forest; see Figure 2) may be caused by the fact that many non-native species favor gaps in the canopy and the proportion of forest cover in a certain area may be a proxy of the number of edges, which are easier to colonize (Komarul Huda et al. 2022). This is supported by the significantly higher average variable importance of the proportion of deciduous forest in non-native species models (Table 6). A possible reason why this pattern is not observed for coniferous forests is that coniferous forests tend to occur at higher altitudes than deciduous forests (Scherrer et al. 2020), and may therefore have not been colonised by non-native species to the same degree yet.

In addition, land use/cover might also reflect the non-native species colonization dynamics. Some of the many variables in this class are devastated forest, forest fresh cuts and mountain meadows and pastures. In short, our results show that anthropogenic factors play a more important role in explaining the distributions of non-native species than those of native species, though the general patterns are rather similar for both species groups. We see that the bioclimatic variables are by far the most important explanatory variables for both native and non-native species groups.

Interesting new research questions arise from these results. For instance, do invasive non-native species favor different variables than their non-invasive counterparts? The sample size in this study was likely insufficient to accurately make this assessment. In addition, the probability of successful establishment of non-native species could be researched. For instance, Broennimann et al. (2021) have shown that the similarity of the climatic niche between the native and non-native range of a species significantly influences invasion success. Such findings could provide more insight in distribution patterns of non-native species and guide conservation actions more effectively.

Furthermore, the COUE framework could be adapted to congeneric species pairs in order to add perspective on niche comparison analysis. SDMs rely on niche conservation between native and non-native range (Guisan et al. 2014). This could lead to interesting conclusions. For instance, if the niches in native and non-native range largely overlap, SDM predictions can effectively encapsulate potential distributions of non-native species. Conversely, if the niches shift (as they have been shown to do), patterns in such a shift should be identified. Our pairwise comparison of congeners is a first step in this direction and could be supported by quantifying the degree in niche similarity resulting from species distribution maps, as was for example done by Warren et al. (2008).

From a more practical perspective, our results also show that different sets of environmental predictors could be used for modelling the distribution of non-native versus native species for practical applications (e.g. pre- and post-border evaluation of invasions; Pheloung et al.1999; Gallien et al. 2012; Schackleton et al. 2020). In particular, it would seem prudent to use transportation-related variables and ones with a low naturality score/high degree of disturbance to build SDMs for non-native species, even if the latter have not (yet) become invasive. Conversely, native species could be best modelled using covariates representing relatively undisturbed habitat and climatic preferences. It must also be noted that the climate covariates must be used at the appropriate scale, as niche truncation must be avoided. This has great implications for management, since SDMs are increasingly frequently advised method for planning conservation measures (Barbet-Massin et al. 2012, Guisan et al. 2013, Tulloch et al. 2016, Araujo et al. 2019, Low et al. 2021).

Another important implication for practitioners would be aimed at foreseeing the impacts of non-native species, whether positive or negative, originating in proximity to human infrastructure. These areas would have to be scanned for incipient or ongoing colonization by non-native species and the identified taxa should be researched on their invasiveness recorded in different parts of the world. Such efforts are already ongoing in Switzerland (www.infospecies.ch) and in the EU (EASIN). In addition to scanning, the trends of the populations should be monitored, especially for strong declines or growth. This approach would be most critical to apply in the areas that have not yet been overly exposed to non-native species colonization, such as (high) mountains. Our results show that as human infrastructure expands, non-native species will probably follow. There, it would be especially important to detect colonization by non-native species early.

Finally, a word of caution to practitioners. Previous studies have shown that a correlation exists between habitat suitability (as calculated by SDMs) and species abundance/population density (Weber et al. 2017, but see Monnier-Corbel et al. 2023 and Brambilla et al. 2024). This work however suggests that population densities of non-native species near roads and in disturbed areas should not be extrapolated to other habitat types, as there, the densities are expected to be lower.

References

Adde, A. (2024). Covsel output summary of 141 native-non- native plant species pairs in Switzerland [data set]. Zenodo. https://doi.org/ 10.5281/zenodo.13325636

Adde, A., Rey, P. L., Brun, P., Külling, N., Fopp, F., Altermatt, F., Broennimann, O., Lehmann, A., Petitpierre, B., Zimmermann, N.E., Pellissier, L., Guisan, A. (2023). N-SDM: a high-performance computing pipeline for Nested Species Distribution Modelling. Ecography. https://doi.org/10.1111/ecog.06540

Adde, A., Rey, P.-L., Fopp, F., Petitpierre, B., Schweiger, A. K., Broennimann, O., Lehmann, A., Zimmermann, N. E., Altermatt, F., Pellissier, L., & Guisan, A. (2023). Too many candidates: Embedded covariate selection procedure for species distribution modelling with the covsel R package. Ecological Informatics, 75, 102080. https://doi.org/10.1016/j.ecoinf.2023.102080

Barbet-Massin, M., Jiguet, F., Albert, C.H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? https://doi.org/10.1111/j.2041-210X.2011.00172.x

Bergamin, R. S., Ascensão, F., Capinha, C., Bastazini, V. A. G., Andrade, B. O., Boldrini, I. I., Lezama, F., Altesor, A., Perelman, S., & Overbeck, G. E. (2022). Native and alien grassland diversity respond differently to environmental and anthropogenic drivers across spatial scales. Journal of Vegetation Science, 33(3), e13133. https://doi.org/10.1111/jvs.13133

Bertelsmeier, C., Ollier, S., Liebhold, A., & Keller, L. (2017). Recent human history governs global ant invasion dynamics. Nature Ecology & Evolution, 1(7), s41559-41017-40184. Doi:10.1038/s41559-017-0184

Bigelow, S. W., & Canham, C. D. (2002). Community organization of tree species along soil gradients in a north-eastern USA forest. Journal of Ecology, 90(1), 188–200. https://doi.org/10.1046/j.0022-0477.2001.00655.x

Bonnamour, A., Gippet, J. M. W., & Bertelsmeier, C. (2021). Insect and plant invasions follow two waves of globalization. Ecology Letters, 24(11), 2418-2426. Doi:10.1111/ele.13863

Brambilla, M., Bazzi, G., & Ilahiane, L. (2024). The effectiveness of species distribution models in predicting local abundance depends on model grain size. Ecology, 105(2), e4224.

Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T., & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. Ecology Letters, 10(8), 701–709. https://doi.org/10.1111/j.1461-0248.2007.01060.x

Broennimann, O., & Guisan, A. (2008). Predicting current and future biological invasions: both native and invaded ranges matter. Biology Letters, 4(5), 585-589.

Broennimann, O., Petitpierre, B., Chevalier, M., González-Suárez, M., Jeschke, J. M., Rolland, J., Gray, S. M., Bacher, S., & Guisan, A. (2021). Distance to native climatic niche margins explains establishment success of alien mammals. Nature Communications, 12(1), Article 1. https ://doi.org/10.1038/s41467-021-22693-0

Broennimann, O. (2023). Chclim25 - bioclimatic variables (biovars) (1.1) [Data set]. Zenodo.

Brown, J. H., Stevens, G. C., & Kaufman, D. M. (1996). The geographic range: size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics, 27, 597-623.

Brown, M. T., & Vivas, M. B. (2005). Landscape development intensity index. Environmental Monitoring and Assessment, 101(1), 289–309. https://doi.org/10.1007/s10661-005-0296-6

Capinha, C., Essl, F., Seebens, H., Moser, D., & Pereira, H. M. (2015). The dispersal of alien species redefines biogeography in the Anthropocene. Science, 348(6240), 1248-1251. doi:10.1126/science.aaa8913

Carlin, T. F., Bufford, J. L., Hulme, P. E., & Godsoe, W. K. (2022). Global assessment of three Rumex species reveals inconsistent climatic niche shifts across multiple introduced ranges. Biological Invasions. doi:10.1007/s10530-022-02893-5

Chapman, D., Purse, B. V., Roy, H. E., & Bullock, J. M. (2017). Global trade networks determine the distribution of invasive non-native species. Global Ecology and Biogeography, 26(8), 907-917. Doi:10.1111/geb.12599

Charbonneau NC, Fahrig L (2004) Influence of canopy cover and amount of open habitat in the surrounding landscape on proportion of alien plant species in forest sites. Ecoscience 11:278–281. https://doi.org/10.1080/11956860.2004.11682833

Chevalier M., Broennimann O., Cornuault J., Guisan A. 2021. Data integration methods to account for spatial niche truncation effects in regional projections of species distribution. Ecological Applications 31(7): e02427.

Chevalier M., Zarzo-Arias A., Guélat J., Mateo R.G., Guisan A. 2022. Accounting for niche truncation to improve spatial and temporal predictions of species distributions. Frontiers in Ecology and Evolution 10: 944116.

Colautti & Lau 2016. In Barrett et al. Invasion Genetics: The Baker and Stebbins Legacy. Chapt. 6

Collette, L. K. D., & Pither, J. (2015). Modeling the potential North American distribution of Russian olive, an invader of riparian ecosystems. Plant Ecology, 216(10), 1371–1383. https://doi.org/10.1007/s11258-015-0514-4

Cook, R. T., Ward, S. F., Liebhold, A. M., & Fei, S. (2021). Spatial dynamics of spotted lanternfly, Lycorma delicatula, invasion of the Northeastern United States. NeoBiota, 70, 23–42. https://doi.org/10.3897/neobiota.70.67950

Davis, M. A., Chew, M. K., Hobbs, R. J., Lugo, A. E., Ewel, J. J., Vermeij, G. J., Brown, J. H., Rosenzweig, M. L., Gardener, M. R., Carroll, S. P., Thompson, K., Pickett, S. T. A., Stromberg, J. C., Tredici, P. D., Suding, K. N., Ehrenfeld, J. G., Philip Grime, J., Mascaro, J., & Briggs, J. C. (2011). Don't judge species on their origins. Nature, 474(7350), 153–154. https://doi.org/10.1038/474153a

Early, R., & Sax, D. F. (2014). Climatic niche shifts between species' native and naturalized ranges raise concern for ecological forecasts during invasions and climate change. Global Ecology and Biogeography, 23(12), 1356-1365.

EASIN – European Alien Species Information Network. (n.d.). Retrieved June 14, 2023, from https://easin.jrc.ec.europa.eu/easin/Documentation/FAQ

Eckert, S., Hamad, A., Kilawe, C. J., Linders, T. E. W., Ng, W.-T., Mbaabu, P. R., Shiferaw, H., Witt, A., & Schaffner, U. (2020). Niche change analysis as a tool to inform management of two invasive species in Eastern Africa. Ecosphere, 11(2), e02987. https://doi.org/10.1002/ecs2.2987

Encyclopedia Britannica | Britannica. (n.d.). Retrieved May 8, 2024, from https://www.britannica.com/

Ender CL, Christian CE, Cushman JH (2017) Native herbivores and environmental heterogeneity as mediators of an exotic grass invasion. Ecol Evol 7:1561–1571. https://doi.org/10.1002/ece3.2727

Enders, M., Havemann, F., Ruland, F., Bernard-Verdier, M., Catford, J. A., Gómez-Aparicio, L., Haider, S., Heger, T., Kueffer, C., Kühn, I., Meyerson, L. A., Musseau, C., Novoa, A., Ricciardi, A., Sagouis, A., Schittko, C., Strayer, D. L., Vilà, M., Essl, F., … Jeschke, J. M. (2020). A conceptual map of invasion biology: Integrating hypotheses into a consensus network. Global Ecology and Biogeography, 29(6), 978–991. https://doi.org/10.1111/geb.13082

Essl, F., Bacher, S., Blackburn, T. M., Booy, O., Brundu, G., Brunel, S., Cardoso, A.-C., Eschen, R., Gallardo, B., Galil, B., García-Berthou, E., Genovesi, P., Groom, Q., Harrower, C., Hulme, P. E., Katsanevakis, S., Kenis, M., Kühn, I., Kumschick, S., Martinou, A.F., Nentwig, W., O'Flynn, C., Pagad, S., Pergl, J., Pyšek, P., Rabitsch, W., Richardson, D.M., Roques, A., Roy, H.E., Scalera, R., Schindler, S., Seebens, H., Vanderhoeven, S., Vilà, M., Wilson, J.R.U., Zenetos, A., Jeschke, J. M. (2015). Crossing

Frontiers in Tackling Pathways of Biological Invasions. BioScience, 65(8), 769–782. https://doi.org/10.1093/biosci/biv082

Froese, J.G., Pearse, A.R., Hamilton, G. (2019). Rapid spatial risk modelling for management of early weed invasions: Balancing ecological complexity and operational needs. Methods in Ecology and Evolution 00:1-13

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(11), 1126-1136. doi:DOI 10.1111/j.1466-8238.2012.00768.x

Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (n.d.). https://doi.org/10.5281/zenodo.6417333

Grunewald, K., Herold, H., Marzelli, S., Meinel, G., Richter, B., Syrbe, R.-U., & Walz, U. (2016). Assessment of ecosystem services at the national level in Germany—Illustration of the concept and the development of indicators by way of the example wood provision. Ecological Indicators, 70, 181–195. https://doi.org/10.1016/j.ecolind.2016.06.010

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: insights from biological invasions. Trends in Ecology & Evolution, 29(5), 260-269. Doi:10.1016/j.tree.2014.02.009

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). Habitat Suitability and Distribution Models: With Applications in R. https ://doi.org/10.1017/9781139028271

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(12), 1424-1435. Doi: 10.1111/Ele.12189

InfoSpecies. (n.d.-a). Retrieved April 7, 2023, from https://www.infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte

InfoSpecies. (n.d.-b). Retrieved May 17, 2024, from https://www.infospecies.ch/de/neobiota/neobiota-in-der-schweiz.html

Listes et fiches d’information. (n.d.). Retrieved June 4, 2024, from https://www.infoflora.ch/fr/neophytes/listes-et-fiches.html#liste-des-n%C3%A9ophytes-envahissantes-et-potentiellement

Huck, C., Körner, C., & Hiltbrunner, E. (2013). Plant species dominance shifts across erosion edge–meadow transects in the Swiss Alps. Oecologia, 171(3), 693–703. https://doi.org/10.1007/s00442-012-2583-6

Invasive alien species: A growing problem for environment and health — European Environment Agency. (n.d.). [News]. Retrieved October 16, 2023, from https://www.eea.europa.eu/highlights/invasive-alien-species-a-growing

Jeschke, J. M., & Strayer, D. L. (2006). Determinants of vertebrate invasion success in Europe and North America. Global Change Biology, 12(9), 1608–1619. https://doi.org/10.1111/j.1365-2486.2006.01213.x

Karger D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E, Linder, H.P., Kessler, M. (2018): Data from: Climatologies at high resolution for the earth’s land surface areas. EnviDat. https://doi.org/10.16904/envidat.228.v2.1Kowarik I (1988) Zum menschlichen Einfluß auf Flora und Vegetation. Theoreti-sche Konzepte und ein Quantifizierungsansatz am Beispiel von Berlin (West), Landschaftsentwicklung und Umweltforschung 56:1–280

Komarul Huda, M., Nursahara Pasaribu, H., Syamsuardi, S. & Siregar, E.S. (2022). Diversity, risk and management feasibility of invasive alien plants in the border zone of Sicike-cike Nature Tourism Park, North Sumatra, Indonesia. Biodiversitas, 23(6). https://doi.org/10.13057/biodiv/d230643

Kowarik I (1995) On the role of alien species in urban flora and vegetation. In: Pysek P, Prach K, Rejmanek M, Wade M (eds) Plant invasions. General as-pects and special problems. SPB Academic Publ, Amsterdam, pp 85–103

Kowarik, I. (2008). On the Role of Alien Species in Urban Flora and Vegetation. In: , et al. Urban Ecology. Springer, Boston, MA.

Kowarik, I. (2014). Natürlichkeit, Naturnähe und Hemerobie als Bewertungskriterien. In Handbuch der Umweltwissenschaften (pp. 1–18). John Wiley & Sons, Ltd. https://doi.org/10.1002/9783527678525.hbuw2006004

Külling, N., Adde, A., Fopp, F., Schweiger, A. K., Broennimann, O., Rey, P.-L., Giuliani, G., Goicolea, T., Petitpierre, B., Zimmermann, N. E., Pellissier, L., Altermatt, F., Lehmann, A., & Guisan, A. (2024). SWECO25: A cross-thematic raster database for ecological research in Switzerland. Scientific Data, 11(1), 21. https://doi.org/10.1038/s41597-023-02899-1 Lauber, K., Wagner, G., Gygax, A. Flora Helvetica – Flore illustrée de Suisse 5e edition, 2018. 1686 pages, ISBN 978-3-258-08050-5

Lembrechts, J. J., Alexander, J. M., Cavieres, L. A., Haider, S., Lenoir, J., Kueffer, C., McDougall, K., Naylor, B.J., Nuñez, M.A., Pauchard, A., Rew, L.J., Nijs, I., Milbau, A. (2017). Mountain roads shift native and non-native plant species' ranges. Ecography, 40(3), 353-364. Doi:10.1111/ecog.02200

Kuhman, T. R., Pearson, S. M., & Turner, M. G. (2010). Effects of land-use history and the contemporary landscape on non-native plant invasion at local and regional scales in the forest-dominated southern Appalachians. Landscape Ecology, 25(9), 1433–1445. https://doi.org/10.1007/s10980-010-9500-3

Kuhman, T. R., Pearson, S. M., & Turner, M. G. (2011). Agricultural land-use history increases non-native plant invasion in a southern Appalachian forest a century after abandonment. Canadian Journal of Forest Research, 41(5), 920–929. https://doi.org/10.1139/x11-026

Lists and infosheets to accompany the new FOEN publication on alien species in Switzerland. (2022). Retrieved November 9, 2023, from https://www.infoflora.ch/en/home/news/2022/11/30/die-neuen-listen-der-invasiven-neophyten/

Low, B. W., Zeng, Y., Tan, H. H., & Yeo, D. C. J. (2021). Predictor complexity and feature selection affect Maxent model transferability: Evidence from global freshwater invasive species. Diversity and Distributions, 27(3), 497–511. https://doi.org/10.1111/DDI.13211

Mang, T., Essl, F., Moser, D., Kleinbauer, I., & Dullinger, S. (2018). An integrated, spatio-temporal modelling framework for analysing biological invasions. Diversity and Distributions, 24(5), 652-665. Doi:10.1111/ddi.12707

Mapping of land cover continuous fields using MODIS data in Switzerland. (2004). Botanica Helvetica, 114(2), 151–167.

Meyer, S. E., Callaham, M. A., Stewart, J. E., & Warren, S. D. (2021). Invasive species response to natural and anthropogenic disturbance. Invasive species in forests and rangelands of the United States: A comprehensive science synthesis for the United States forest sector, 85-110.

Monnier-Corbel, A., Robert, A., Hingrat, Y., Benito, B. M., & Monnet, A. C. (2023). Species Distribution Models predict abundance and its temporal variation in a steppe bird population. Global Ecology and Conservation, 43, e02442.

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. Proceedings of the Royal Society B-Biological Sciences, 278(1725), 3644-3653. Doi:DOI 10.1098/rspb.2010.2769

OFEV (éd.) 2022 : Espèces exotiques en Suisse. Aperçu des espèces exotiques et de leurs conséquences. Office fédéral de l'environnement, Berne. Connaissance de l'environnement no 2220 : 62 p.

Oshima, K., & Takahashi, K. (2020). Forest disturbances promote invasion of alien herbaceous plants: A comparison of abundance and plant traits between alien and native species in thinned and unthinned stands. Biological Invasions, 22(9), 2749–2762. https://doi.org/10.1007/s10530-020-02283-9

Pergl, J., Brundu, G., Harrower, C. A., Cardoso, A. C., Genovesi, P., Katsanevakis, S., Lozano, V., Perglova, I., Rabitsch, W., Richards, G., Roques, A., Rorke, S. L., Scalera, R., Schonrogge, K., Stewart, A., Tricarico, E., Tsiamis, K., Vannini, A., Vil, M., … Roy, H. E. (2020). Applying the Convention on Biological Diversity Pathway Classification to alien species in Europe. NeoBiota 62, 333-363. http ://hdl.handle.net/10019.1/120840

Perret, D. L., Leslie, A. B., & Sax, D. F. (2019). Naturalized distributions show that climatic disequilibrium is structured by niche size in pines (Pinus L.). Global Ecology and Biogeography, 28(4), 429-441. Doi:10.1111/geb.12862

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., & Guisan, A. (2012). Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. Science, 335(6074), 1344–1348. https://doi.org/10.1126/science.1215933

Petitpierre, B., McDougall, K., Seipel, T., Broennimann, O., Guisan, A., & Kueffer, C. (2016). Will climate change increase the risk of plant invasions into mountains? Ecological Applications, 26(2), 530–544. https://doi.org/10.1890/14-1871

Pheloung, P. C., Williams, P. A., & Halloy, S. R. (1999). A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. Journal of Environmental Management, 57(4), 239-251. doi:DOI 10.1006/jema.1999.0297

Pulliam, H. R. (2000). On the relationship between niche and distribution. Ecology Letters, 3(4), 349-361.

Pyšek, P., Jarosik, V., Hulme, P. E., Kuhn, I., Wild, J., Arianoutsou, M., . . . Winter, M. (2010). Disentangling the role of environmental and human pressures on biological invasions across Europe. Proceedings of the National Academy of Sciences of the United States of America, 107(27), 12157-12162. Doi:DOI 10.1073/pnas.1002314107

Pyšek, P., Hulme, P. E., Simberloff, D., Bacher, S., Blackburn, T. M., Carlton, J. T., Dawson, W., Essl, F., Foxcroft, L. C., Genovesi, P., Jeschke, J. M., Kühn, I., Liebhold, A. M., Mandrak, N. E., Meyerson, L. A., Pauchard, A., Pergl, J., Roy, H. E., Seebens, H., … Richardson, D. M. (2020). Scientists' warning on invasive alien species. Biological Reviews, 95(6), 1511–1534. https://doi.org/10.1111/brv.12627

Radford, S. L., Senn, J., & Kienast, F. (2019). Indicator-based assessment of wilderness quality in mountain landscapes. ECOLOGICAL INDICATORS, 97, 438–446. https://doi.org/10.1016/j.ecolind.2018.09.054

Raunkiaer, C. (1977). The life forms of plants and statistical plant geography.: Arno Press. .

Robeck, P., Essl, F., van Kleunen, M., Pyšek, P., Pergl, J., Weigelt, P., & Mesgaran, M. B. (2024). Invading plants remain undetected in a lag phase while they explore suitable climates. Nature Ecology & Evolution, 8(3), 477–488. https://doi.org/10.1038/s41559-023-02313-4

Rojas-Sandoval, J., Ackerman, J. D., Dueñas, M.-A., Velez, J., & Díaz-Soltero, H. (2024). Habitat affiliation of non-native plant species across their introduced ranges on Caribbean islands. Biological Invasions. https://doi.org/10.1007/s10530-024-03307-4

Sax, D. F., Schlaepfer, M. A., & Olden, J. D. (2022). Valuing the contributions of non-native species to people and nature. Trends in Ecology & Evolution, 37(12), 1058–1066. https://doi.org/10.1016/j.tree.2022.08.005

Scherrer, D., Vitasse, Y., Guisan, A., Wohlgemuth, T., & Lischke, H. (2020). Competition and demography rather than dispersal limitation slow down upward shifts of trees' upper elevation limits in the Alps. Journal of Ecology, 108(6), 2416–2430. https://doi.org/10.1111/1365-2745.13451

Seaton, F. M., George, P. B. L., Lebron, I., Jones, D. L., Creer, S., & Robinson, D. A. (2020). Soil textural heterogeneity impacts bacterial but not fungal diversity. Soil Biology and Biochemistry, 144, 107766. https://doi.org/10.1016/j.soilbio.2020.107766

Shackleton, R. T., Petitpierre, B., Pajkovic, M., Dessimoz, F., Brönnimann, O., Cattin, L., Čejková, Š., Kull, C. A., Pergl, J., Pyšek, P., Yoccoz, N., & Guisan, A. (2020). Integrated Methods for Monitoring the Invasive Potential and Management of Heracleum mantegazzianum (giant hogweed) in Switzerland. Environmental Management, 65(6), 829–842. https://doi.org/10.1007/s00267-020-01282-9Soto, I., Balzani, P., Carneiro, L., Cuthbert, R.N., Macêdo, R., Serhan Tarkan, A., Ahmed, D.A., Bang, A., Bacela-Spychalska, K., Bailey, S.A., Baudry, T., Ballesteros-Mejia, L., Bortolus, A., Briski, E., Britton, J.R., Buřič, M., Camacho-Cervantes, M., Cano-Barbacil, C., Copilaș-Ciocianu, D., Coughlan, N.E., Courtois, P., Csabai, Z., Dalu, T., De Santis, V., Dickey, J.W.E., Dimarco, R.D., Falk-Andersson, J., Fernandez, R.D., Florencio, M., Franco, A.C.S., García-Berthou, E., Giannetto, D., Glavendekic, M.M., Grabowski, M., Heringer, G., Herrera, I., Huang, W., Kamelamela, K.L., Kirichenko, N.I., Kouba, A., Kourantidou, M., Kurtul, I., Laufer, G., Lipták, B., Liu, C., López-López, E., Lozano, V., Mammola, S., Marchini, A., Meshkova, V., Milardi, M., Musolin, D.L., Nuñez, M.A., Oficialdegui, F.J., Patoka, J., Pattison, Z., Pincheira-Donoso, D., Piria, M., Probert, A.F., Rasmussen, J.J., Renault, D., Ribeiro, F., Rilov, G., Robinson, T.B., Sanchez, A.E., Schwindt, E., South, J., Stoett, P., Verreycken, H., Vilizzi, L., Wang, Y.-J., Watari, Y., Wehi, P.M., Weiperth, A., Wiberg-Larsen, P., Yapıcı, S., Yoğurtçuoğlu, B., Zenni, R.D., Galil, B.S., Dick, J.T.A., Russell, J.C., Ricciardi, A., Simberloff, D., Bradshaw, C.J.A., Haubrock, P.J. (2024). Taming the terminological tempest in invasion science. Biological Reviews, https://doi.org/10.1111/brv.13071(n/a).

Staude, IR., Waller, DM., Bernhardt-Römermann, M., Bjorkman, A.D., Brunet, J., De Frenne, P., Hédl, R., Jandt, U., Lenoir, J., Máliš, F., Verheyen, K., Wulf, M., Pereira, H.M., Vangansbeke, P., Ortmann-Ajkai A., Pielech, R., Berki, I., Chudomelová, M., Decocq, G., Dirnböck, T., Durak, T., Heinken, T., Jaroszewicz, B., Kopecký, M., Macek, M., Malicki, M., Naaf, T., Nagel, T.A., Petřík, P., Reczyńska, K., Schei, F.H., Schmidt, W., Standovár, T., Świerkosz, K., Teleki, B., Van Calster, H., Vild, O. & Baeten L. (2020). Replacements of small- by large-ranged species scale up to diversity loss in Europe’s temperate forest biome. Nature Ecology & Evolution, 4(6), 802-808. Doi:10.1038/s41559-020-1176-8

Steen, B., Cardoso, A.C., Tsiamis, K., Nieto, K., Engel, J., Gervasini, E. (2019) Modelling hot spot areas for the invasive alien plant Elodea nuttallii in the EU. Management of Biological Invasions 10(1), 151–170, https://doi.org/10.3391/mbi.2019.10.1.10

Steen, B., Broennimann, O., Maiorano, L., & Guisan, A. (2024). How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels. Ecological Modelling, 493, 110754. https://doi.org/10.1016/j.ecolmodel.2024.110754

Steffen, W., Rockström, 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. Proceedings of the National Academy of Sciences of the United States of America, 115(33), 8252–8259. https://doi.org/10.1073/pnas.1810141115

Stohlgren, T. J., Pyšek, P., Kartesz, J., Nishino, M., Pauchard, A., Winter, M., Pino, J., Richardson, M. D., Wilson, J. R. U., Murray, B. R., Phillips, M. L., Ming-yang, L., Celesti-Grapow, L. & Font, X. (2011). Widespread plant species: natives versus aliens in our changing world. Biological Invasions, 13(9), 1931-1944. Doi:10.1007/s10530-011-0024-9

Svenning, J. C., Normand, S., & Skov, F. (2008). Postglacial dispersal limitation of widespread forest plant species in nemoral Europe. Ecography, 31(3), 316-326. Doi:DOI 10.1111/j.0906-7590.2008.05206.x

Treier, U. A., Broennimann, O., Normand, S., Guisan, A., Schaffner, U., Steinger, T., & Müller-Schärer, H. (2009). Shift in cytotype frequency and niche space in the invasive plant Centaurea maculosa. Ecology, 90(5), 1366-1377. Doi:10.1890/08-0420.1

Tulloch, A. I. T., Sutcliffe, P., Naujokaitis-Lewis, I., Tingley, R., Brotons, L., Ferraz, K., Possingham, H.P., 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. Doi:10.1016/j.biocon.2016.04.023

Van Kleunen, M., Bossdorf, O., & Dawson, W. (2018). The Ecology and Evolution of Alien Plants. In D. J. Futuyma (Ed.), Annual Review of Ecology, Evolution, and Systematics, Vol 49 (Vol. 49, pp. 25-47). Palo Alto: Annual Reviews.

Wang, Y.-J., Liu, Y.-Y., Chen, D., Du, D.-L., Müller-Schärer, H., & Yu, F.-H. (n.d.). Clonal functional traits favor the invasive success of alien plants into native communities. Ecological Applications, n/a(n/a), e2756. https://doi.org/10.1002/eap.2756

Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F., & Grelle, C. E. V. (2017). Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. Ecography, 40(7), 817-828.

Yoshida, T., Goka, K., Ishihama, F., Ishihara, M., & Kudo, S. (2007). Biological invasion as a natural experiment of the evolutionary processes: Introduction of the special feature. Ecological Research, 22(6), 849–854. https://doi.org/10.1007/s11284-007-0435-3

FOEN (Ed.) 2022: Alien Species in Switzerland. An inventory of alien species and their impact. 1st updated edition 2022. 1st edition 2006. Federal Office for the Environment, Bern. Environmental studiesNo 2220: 62 pp.

Zhou, Q., Wang, Y., Li, X., Liu, Z., Wu, J., Musa, A., Ma, Q., Yu, H., Cui, X., & Wang, L. (2020). Geographical distribution and determining factors of different invasive ranks of alien species across China. Science of The Total Environment, 722, 137929. https://doi.org/10.1016/j.scitotenv.2020.137929

Zimmermann, N. E., Yoccoz, N. G., Edwards, T. C. J., Meier, E. S., Thuiller, W., Guisan, A., Schmatz, D., Pearman, P. B. (2009). Climatic extremes improve predictions of spatial patterns of tree species. Proc Natl Acad Sci U S A, 106 Suppl 2, 19723-19728. doi:10.1073/pnas.0901643106

0901643106 [pii]

CHAPTER II

How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels

Bart Steen*[1,2], Olivier Broennimann[1,3], Antoine Guisan**[1,3], Luigi Maiorano**[2]

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

[2] Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, viale dell'Università 32, 00185 Roma, Italy

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

* Corresponding author: bart.steen@unil.ch / b.r.l.steen2@gmail.com

** Shared last authorship

This chapter was published in Ecological Modelling.

DOI: 10.1016/j.ecolmodel.2024.110754

My contribution to the project: I collected, generated, analyzed and cured the data, wrote the programs, co-designed the methodology (in cooperation with Luigi Maiorano and Olivier Broennimann), drew the figures and led the writing of the manuscript.

Abstract

Species distribution models (SDMs) have become central tools in ecology and biogeography. Although they can be fitted with different types of species data (e.g. presence-absence, abundance), the most common approach, based on data from large species repositories, is to use simple occurrences (i.e. presence-only) combined with background points (BP; also called pseudo-absences). But how should we sample these background points, and how does this choice affect SDMs? In most studies so far, BP were sampled randomly in geographic space, yet theory rather suggests, if a species is at equilibrium, that it is better to sample them in a stratified way in environmental space. However, this potential improvement of SDM predictions has never been tested. Furthermore, a typical assumption behind SDMs is that the modelled species are at equilibrium with their environment. But how do these models perform when species are in disequilibrium, as is the case for most invasive species? To answer these questions, we selected 30 different species (10 insects, 10 mammals and 10 plants; for each group 5 were invasive and 5 were considered at equilibrium) and for each we calibrated SDMs with different types of background selections: random in environmental space, random-stratified in environmental space, random in geographic space, and random-stratified in geographic space. For each SDM we assessed both predictive performance using standard metrics and their stability using a new approach that compares the model's habitat suitability projection with those of a SDM calibrated with virtual occurrence data generated from the most suitable areas. Finally, we compared the predictive performance of species distribution models of invasive alien (disequilibrium) species versus native (equilibrium) species by comparing model stability and performance metrics of the two groups. We found that sampling BP in a stratified-random way in environmental space yields the highest performance metrics, and that sampling fully randomly in environmental space yields the most stable models. This has implications for the use of SDMs in conservation, as the classical and frequently used fully random in geographic space BP are found to produce both less accurate and less stable models. Our results indicate that the best approach is to use stratified random in environmental space BP sampling if accuracy is essential, and fully random in environmental space BP sampling if model stability is essential.

Introduction

Species Distribution Models (SDMs), also known as Ecological Niche Models (ENMs), constitute a widespread method for ecological, biogeographic, and conservation studies (Franklin 2010, Peterson et al. 2011, Guisan et al. 2013, Guisan et al. 2017, Araujo et al. 2019), often being particularly useful when dealing with invasive alien species (IAS) (Gallien et al. 2012, Slodowicz et al. 2018, Vicente et al. 2019). SDMs take geographic observations of species and georeferenced environmental datasets in a study area to statistically identify the conditions in which the species can occur (i.e., a part or all of the species' realized environmental niche) and map the occurrence probability of the species based on these conditions back in geographic space (Colwell & Rangel 2009). The usefulness of SDMs depends both on model accuracy, i.e., how well a model is able to make predictions, and model stability, i.e., the degree of model sensitivity to stochasticity (Duan et al. 2014, Grimett et al. 2020, Mateo et al. 2010). Therefore, consistency of geographical predictions is a pivotal factor to research.

The vast majority of SDMs use presence-only data, likely because occurrence data in national or global databases (e.g., the Global Biodiversity Information Facility (GBIF); Anderson et al. 2016, Fletcher et al 2019, Chapman et al. 2020) have become numerous and easily accessible (Chauvier et al. 2021, Valavi et al. 2021, Nolan et al. 2022). In contrast, absence data are more rarely available and hard to generate, especially in the case of animal species (Gormley et al. 2011, Lobo et al. 2010). Presence-background SDMs function by contrasting the species occurrences to points sampled (often randomly) in the study area (Valavi et al. 2022), ideally along the main environmental gradients used to define the species niches (Barve et al. 2011; but see Iturbide et al. 2018 for other possibilities). This set of points representing the environment are called background points (BP; often also pseudo-absences, but see Senay et al. 2013) and are most typically sampled considering a geographically random scheme (Liu et al 2019, Chauvier et al 2021, Barber et al. 2022). Therefore, presence-background SDMs make the implicit assumption that BP sampled in geographic space represents the available environment (measured as all environmental combinations within the study area) appropriately (Barbet-Massin et al 2012, Wisz & Guisan 2009, Liu et al 2019, Cengic et al. 2020, Valavi et al. 2022).

However, since SDMs fit response curves along environmental predictors, this random selection of BP should theoretically be conducted in environmental rather than geographical space and in a stratified way (Fragnière et al. 2022). This is also supported by the fact that, in order to produce accurate and robust

SDM results, species occurrence data should optimally be sampled in the field in a random-stratified in environmental space manner (Hirzel & Guisan 2002).

Whilst there have been studies that used background points selected in environmental space, this was usually done in a variety of stratified fashions, arranging the background points along environmental gradients (e.g. Fragnière et al. 2022). It has also recently been shown that geographically constraining background point selections to areas that are geographically or environmentally close to the occurrence points can strongly influence SDM predictive accuracy (Whitford et al. 2024; Schartel & Cao 2024). However, it was rarely done fully randomly in environmental space. Here, we present the first study that also uses a fully random selection of background points in environmental space, allowing us to test the four possible types of selection based on geographic versus environmental spaces, and stratified or not, and thus whether the implicit assumption that the geographic space reflects the environmental space is acceptable or not when fitting SDMs. As Hutchinson's niche-biotope duality tells us that the geographic space does not reflect the environmental space (Colwell & Rangel 2009; Guisan et al. 2014), the question is whether it can still give a good approximation.

Another major assumption made when using SDMs is that the modelled species is at equilibrium with its environment. This is already not true for all native species (Normand et al. 2011, Svenning & Sandel 2013), but it is almost never true in the case of invasive alien species (IAS), as their spread leads them to progressively occupy new habitats and territories (Gallien et al. 2012, Petitpierre et al. 2012). Therefore, statistical species-environment relationships usually only represent a snapshot "pseudo-equilibrium" in time (Guisan & Zimmermann 2000). The question is, then, how much a disequilibrium causes problems to SDM fitting and predictions (Svenning & Sandel 2013)? Therefore, the method of sampling background points may be even more important when a species is not at equilibrium with the environment.

Here, we address this second dimension of (dis)equilibrium with the environment together with our assessment of background point selection, by using species with different putative degrees of equilibrium. More specifically, we make the simplified assumption that well-established native species, i.e., those for which we know that they have a dynamic that has not changed recently, are likely at equilibrium inside their home range and fill their realized niche. In this study, we want to answer the two following research questions around the two assumptions previously exposed:

(1) Which method of BP sampling produces the best-performing SDMs for invasive and equilibrium species in terms of model accuracy and stability?
(2) Do the optimal BP sampling strategy differ between native and invasive alien species?

Materials and Methods

Species selection and study area

In our analyses, we considered both invasive alien species (IAS), i.e., species not at equilibrium with the environment, and species which we assume to be at the equilibrium (see Table 1). Both invasive and equilibrium species included 5 insect species, 5 plant species and 5 vertebrate species, making a total of 30 target species.

All equilibrium species are mountain species, which are naturally distributed along steep elevational and environmental gradients, thus making their distribution constrained in geographic space (i.e., where mountains are). Also, this limited our study area to mountain biomes, which was advantageous for computational considerations. A final reason to select non-generalist mountain species for equilibrium studies is because whilst many of these species might physiologically be able to occur at lower altitudes, as demonstrated by growing them in botanical gardens (Vetaas 2002), competition with other species drives them away from there. For instance, the Alpine aster (*Aster alpinus*) usually occurs at elevation of over 2000 m because it gets outcompeted at lower altitudes (Lyu & Alexander, 2022). High-altitude species therefore are likely to occupy their entire realized niche, and therefore they are at (pseudo-)equilibrium with their realized environment in mountains.

We selected 15 IAS from the EU list of Union Concern (https://ec.europa.eu/environment/nature/invasivealien/list/index_en.htm) and the factsheets on invasive alien species in Switzerland provided by the Swiss Federal Office for the Environment (FOEN)(OFEV, invasive_alien_speciesinswitzerlandfactsheets.pdf). The Swiss lists were chosen because of the high abundance of high-altitude species and extensive work on the effect of non-native flora and fauna on biodiversity. It was ensured that these species at least occur in mountains in their home range and that most of their occurrence data (see section 2.4) were inside the study area (see section 2.1). The list of species is given in table 1 (see Appendix 1, table S5 for the list along with the rationale to include the species).

Table 1: Study species and matching rationale to insert them into the studies and number of occurrences (in the study area, after removal of non-native range occurrences for equilibrium species and thinning). Note that the invasive alien species all at least occur in mountain areas in their home range.

Category	**Species group**	**Species name**	**Number of occurrences**
Invasive	Plants	*Asclepias syriaca*	1342
		Heracleum mantegazzianum	833
		Impatiens glandulifera	769
		Lespedeza cuneata	236
		Lupinus polyphyllus	820
	Vertebrates	*Branta canadensis*	920
		Myocastor coypus	400
		Ondatra zibethicus	281
		Sciurus niger	686
		Tamias sibiricus	77
	Insects	*Aedes albopictus*	95
		Cameraria ohridella	225
		Harmonia ayridis	1377
		Ips typographus	77
		Popillia japonica	794
Equilibrium	Plants	*Alnus alnobetula*	164
		Aster alpinus	86
		Helictrotrichon pratense	70
		Leontopodium nivale	423
		Ranunculus glacialis	107
	Vertebrates	*Lagopus muta*	64
		Lepus timidus	286
		Marmota marmota	123
		Nucifraga caryocatactes	433
		Tetrao urogallus	445
	Insects	*Boloria pales*	75

		Colias phicomone	87
		Erebia meolans	140
		Parnassius apollo	212
		Parnassius mnemosyne	93

We defined our study area using the biomes from Olson et al. (2001) where our species occur (Fig. 1): temperate broadleaf and mixed forests, temperate coniferous forests, temperate grasslands, savannahs and shrublands, montane grasslands and shrublands, boreal forests/taiga and tundra. We excluded biomes where the presence of our species was marginal (see Appendix 3, table S31 for occurrence points omitted per species). For an additional analysis on the potential effect of the omitted occurrence points on the model outcomes, showing that the environmental values of the occurrence points off the biomes are represented into the environmental values of the background points and of the occurrence points included in the study, see Appendix 3 (Figure S32, S33).

Algorithm choice and settings

MaxEnt (which stands for 'maximum entropy'), a type of Poisson point process regression model (Renner & Warton 2013), is among the most widespread algorithms used to fit SDMs (Elith et al. 2011, Gomes et al. 2018). It also draws upon presence-only (and background) data, which is much more widely available than data on presence-absence of species (Fletcher et al. 2019, Guillera-Arroita et al. 2015). To avoid model overfitting, the ENMeval package version 2.0.3 (Kass et al. 2021) was used to find the most optimal model settings. MaxEnt uses several mathematical function types (linear, quadratic, hinge, product, threshold) and the Regularization Multiplier (RM), a controller of model complexity, to identify the relationship between data occurrence and the underlying environmental variables (Phillips et al. 2009). ENMeval calibrates a MaxEnt model by pointing out the optimal settings of the function types and RM. Since ecological interactions are never accurately described by only one type of function, e.g., only linear functions, we only chose three combinations of functions: 1) linear, quadratic and hinge (LQH), 2) linear, quadratic and product (LQP), and 3) linear, quadratic product and hinge (LQPH). The RM was varied from 1 to 5. Threshold functions were never used because hinge functions have been shown to be a better replacement (Elith et al. 2011).

Environmental variables

Bioclimatic variables were downloaded from CHELSA, version 1.2 (Karger et al. 2016), at a resolution of 0.0083 degrees. This version was chosen because the variables are composed of the average values from the time series of 1979 – 2013 and species responses to climate change often lag behind the actual climate at a given time (Menendez et al. 2006). Though this may vary among species groups, we make this assumption in this study, for the sake of comparability. An average for a time period is therefore more informative than choosing the values at one specific time point. The Ecospat R package (Di Cola et al. 2017, Broennimann et al. 2023) was used to investigate correlation among the variables (i.e., collinearity) and to select the best set of (partly) uncorrelated variables to insert in the models. A correlation limit of 0.70 between variables was used (Dormann et al. 2013). In the end, seven variables were used: mean annual temperature, diurnal temperature range, total temperature range, annual precipitation, precipitation seasonality, precipitation of driest quarter, precipitation of coldest quarter.

Species occurrence data

We downloaded data from the Global Biodiversity Information Facility (GBIF; Anderson et al. 2016) on 30 species (4 GBIF.org downloads). Only occurrence records from the year 2000 or later were selected, in order to match the CHELSA version 1.2 timeline. The delay from 1979-2013 to 2000-2022 is assumed to make up for the delayed species response (see section 2.3). To properly focus on our question of equilibrium state, only the native range of species is modeled. Conversely, IAS are expected to be at disequilibrium with the environment (Gallien et al. 2012). For these species, all the known range (in the study area) was modeled in this study. This is firstly to research the effect of this disequilibrium on model stability, and secondly because more is often known about those species in their invaded range than in their native range (Poland et al. 2021).

Many species that are not listed as invasive are still naturalized alien in some countries, which might mean that they are not at equilibrium in part of their range. Therefore, the DASCO R package (Hseebens 2022) was used to identify the native distribution range of the species at equilibrium and all occurrence records from outside that native range were removed. Then, occurrence records that were outside the study area were removed. For an additional analysis on the potential effect of this on niche truncation and model's ability to accurate characterize the niche, see supplementary material, Appendix 3, Figure S32 and S33. The data were subsequently 'cleaned'. This implied removal of unlikely coordinates, occurrences with coordinate uncertainty over 50 m, observations with zero abundance and/or an occurrence status of “absent” (in many languages) and visual inspection of the cleaned data set. This was done using the scrubr R package (R/scrubr-package.R ; R-bloggers). For species with more than 20’000 occurrence records, 20’000 occurrence points were selected at random, for reasons of computing efficiency. This number of occurrence points was considered big enough to represent the species’ distribution range effectively, thus we lost no information in this step. Finally, the occurrence data for each species were thinned (Steen et al. 2021, Inman et al. 2021) as follows. First, a Minimum Convex Polygon (MCP) was drawn around the (remaining) points. This is a technique to draw the borders of the home range of the species. Within this MCP, as many points as there are occurrence data were randomly drawn, following Chiocchio et al. (2021). This process was repeated 50 times and each time, the average nearest neighbor distance between the randomly drawn points was recorded. The average of these averages was calculated, and the species occurrence data was thinned by this distance using the spThin R package (Aiello-Lammens et al. 2015). This method matches the clustering of the occurrence records to clustering expected for a random distribution, as it returns the number of points expected in the species’ home range and therefore limits sampling bias of occurrence points. The downloaded occurrence data were then thinned to 1 point per grid cell of our environmental rasters (Steen et al. 2021) because our version of MaxEnt only accepts one occurrence point per grid cell. To limit the risk of sample size influencing our results, we kept only species that had at least 50 occurrence records remaining for further analyses, to be well above the limit of 30 occurrences under which no algorithm predicts consistently (Wisz et al. 2008).

Originally, many more species were considered to be included in the study, however, this method of thinning often removed too many records.

Background points sampling

We sampled 10,000 background points for each sampling strategy. It is generally acknowledged that this is a sufficient background point sample size for SDMs built using MaxEnt (Barbet-Massin et al. 2012). Each of the 4 methods of background point sampling – fully random in environmental space (randenv), stratified random in environmental space (stratenv), fully random in geographic space (randgeo) and stratified random in geographic space (presented by checkerboard 1 (cb1) and checkerboard 2 (cb2)) – is described below. The background points were sampled from the study area described under section 2.1. See figure 2 for a schematic explanation. All the background sample sizes were downweighted to the number of occurrences.

Fully random in environmental space (randenv)

The simplest approach would be to draw all points randomly in environmental space at once, but when the available environment of the study area is very restricted (as here with mountains), it becomes difficult to make such random points match actual conditions in the field. For this reason, we had to develop a specific procedure. First, a PCA analysis of all the environmental variables was performed using the “PCA” function of the R package “FactoMineR” (Husson et al. 2020). The first two PCA axes explain 74.4% of the variation. Then, a two-dimensional kernel density estimate was performed on the two axes using the “kde2D” function of the R package “MASS” (Venables et al 2002). Based on this, we divided the environmental dataset into 10 quantiles, each of which contained roughly 10% of all the values in the PCA axes, see Appendix 1, supplementary figure S3. Then, the values of the axes were rounded to 1 decimal. Finally, 1000 background points were sampled randomly across all 10 quantiles, resulting in a final dataset of 10000 BP, as follows: First, random numbers ranging between the minimum and the maximum value of the first PCA axis were generated. The same was done for the second PCA axis. Then, to perform the random sampling, both random numbers were matched to the values of both axes and when there was a match, the corresponding longitude and latitude were recorded into the randenv background point dataset. When performing the random sampling, two conditions had to be met: 1) the randomly sampled values on both the PC1 axis and the PC2 axis had to match the values in the matching cell and 2) no duplicate values of either PC1 or PC2 were allowed. The latter condition was used in order to prevent oversampling of more common values. Thus, we made sure that the final dataset of background points was one that at least approached a representative, fully random sample of all the environmental conditions available. For additional details, see Appendix 1, supplementary figure S3.

Stratified random in environmental space (stratenv)

The 7 CHELSA raster layers were each divided into 3 equal environmental strata as follows. First, the difference between the minimum and maximum value in a layer was calculated. Then, all the cells with values that fell within a one-third quantile of the group were selected, thus making three strata per layer. Each of these strata was assigned a unique numeric code which was not included in any of the strata of other layers. Therefore, each matching pixel across raster layers has seven different numeric codes. Then, a new raster layer was created in which all the numeric codes were summed, creating a unique new code for each combination of strata. Finally, the *sampleStratified* function of the raster package was used to sample across this new layer and create the stratenv background points. This function makes an effort to sample across all the cells in the summed layer that have the same code. Therefore, all strata were sampled (more or less) equally. Since, in environmental space, the number of the strata (and thus their resolution) should not affect the results, we used only one here (i.e. 3 strata per climatic layer).

Fully random in geographic space

This option is the most commonly used and therefore many options exist to conduct it. Here, the *randomPoints* function of the *dismo* (Hijmans et al. 2019) package was used to generate randgeo background points. This function samples fully randomly across the rows and columns of the geographic raster variables (omitting cells with no data), with weights on the rows, so not too many cells can be selected near the poles. Also, no cell can be selected more than once.

Stratified random in geographic space

The checkerboard method of the *ENMeval* package (Kass et al. 2021) was used to divide the randgeo background data (as well as the occurrence data) according to square “checkerboard” blocks (Steen et al. 2019). As, in geographic space (unlike in environmental space), the grain of the checkerboard could influence the result, we tested two different block sizes. The size of the checkerboard blocks was set at 5 km x 5 km to create the first set of bins (checkerboard 1 (cb1)) in one dataset and at 10 km x 10 km to create the second set (checkerboard 2 (cb2)). All the analyses described under section 2.7 were done for all species, with both sets.

Background points comparison

The datasets of 10’000 background points used for each treatment are shown in figure 1. In order to investigate if the background point (BP) datasets are, as one would expect, spatially different from each other, we calculated the Moran’s I for each dataset. See figure 2 for a finer illustration of the comparison between the classic fully random sampling in geographic space and the stratified random sampling in environmental space.

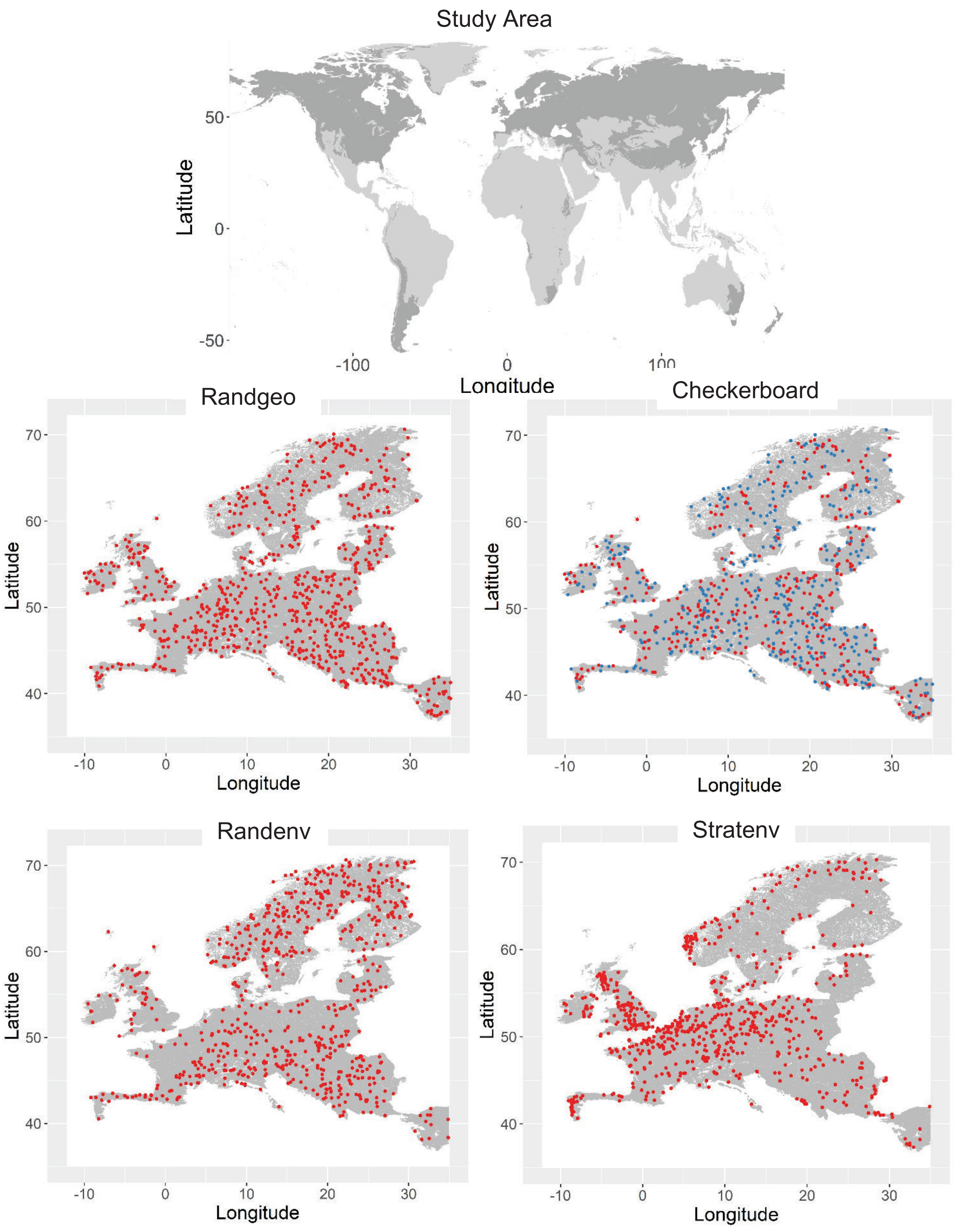

Study Area
Latitude
50
0
-50
-100
0
100
Longitude
Randgeo
Latitude
70
60
50
40
-10
0
10
20
30
Longitude
Checkerboard
Latitude
70
60
50
40
-10
0
10
20
30
Longitude
Randenv
Latitude
70
60
50
40
-10
0
10
20
30
Longitude
Stratenv
Latitude
70
60
50
40
-10
0
10
20
30
Longitude

Figure 1: Study area and comparison of methods of background creation. The background point datasets are shown for Europe only, in order to facilitate visual comparisons: Study area (plotted in dark grey); Fully random in geographic space (randgeo); Checkerboard treatment. The different bins of the data are shown in red and blue. Note that the grid size of the checkerboard blocks have been enlarged to 100km in order to show the checkered division of the BP data across the bins; Fully random in environmental space (randenv); Stratified-random in environmental space (stratenv). Note we did not display the stratified random in geographic space method here because the checkerboard models also use the random in geographic space background points (in B).

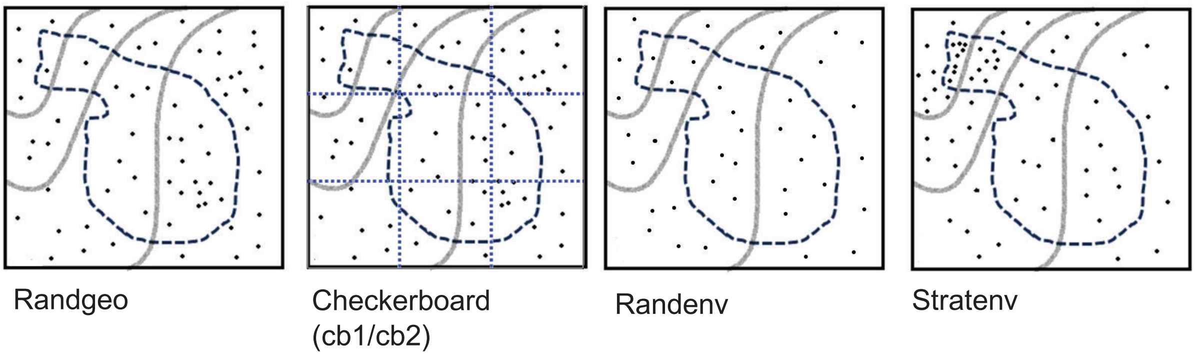


Figure 2: Geographic representation of the classic fully random in geographic space (randgeo) background sampling, the stratified random in geographic space (i.e., checkerboard; cb1/cb2) approach, the fully random in environmental space (randenv) approach, where the number of BP is proportional to the size of the strata and the stratified random in environmental space (stratenv) background sampling, where the same number of points are sampled in any combination of environmental variables. The dots represent sampling points, plain lines represent isolines of some gradually varying environmental variable (e.g., temperature changing along elevation) and the dashed line polygon represents a patch of some landscape feature (e.g. a forest). Modified from Hirzel & Guisan (2002).

Optimal model selection and general modelling framework

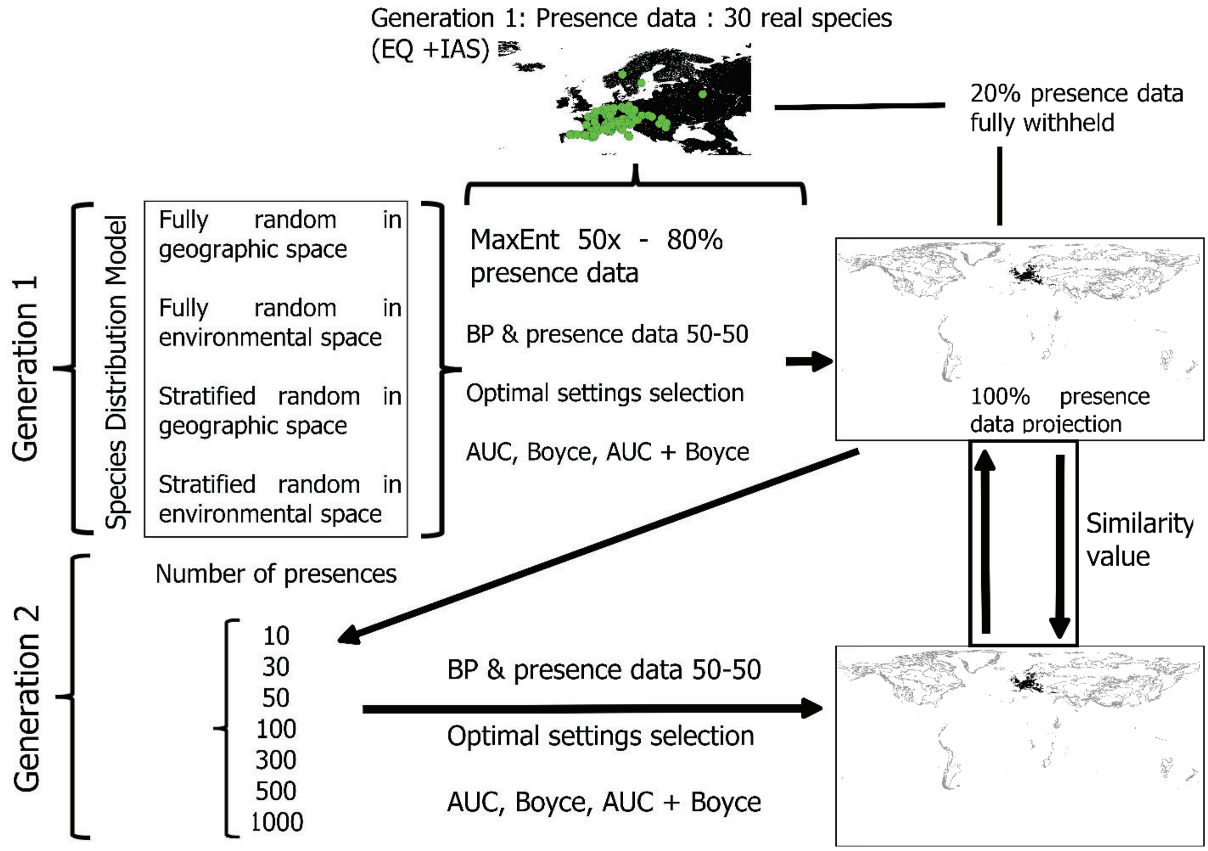


Figure 3: Workflow and design of the study. .

The thinned occurrence datasets for the target invasive (non-equilibrium) and equilibrium species were randomly divided into two bins: one that contained 80% of their occurrence data and one that contained 20%, a common choice of partition sizes for SDMs (Cosentino et al. 2023). Only the 80% bin was used for the selection of the optimal models. For each of these 80% bins, all pertaining to a different target species, 50 maxent models were run using the ENMeval package, with the MaxEnt.jar model method selected and using the aforementioned settings (see section 2.2). Before running each maxent model, the 80% bin of data was randomly split 50-50 across a training bin and a validation bin. We diverted here from the more typically used 80-20 or 70-30 division because the bins used in the checkerboard analyses would split the data roughly evenly across training and validation datasets. The best model settings were selected based on their average value of 3 predictive performance metrics: the continuous Boyce index (CBI; Hirzel et al. 2006), the Area Under the ROC Curve (AUC, Jimenez et al. 2020) and the sum of both (see Figure 4). The last one was calculated in order to not neglect models that had both (relatively) high Boyce (i.e., good calibration in predictions) and (relatively) high AUC (i.e., good discrimination in predictions). This is necessary, because it is possible for a model to have a high AUC and a low Boyce index or vice versa (i.e. balancing discrimination versus calibration in the evaluation; Guisan et al. 2017), and in both cases that would be a bad model from at least one perspective. Before the sum was calculated, the AUC was recalculated to the gini AUC (2*AUC – 1) ranging between -1 and +1 in order to give it the same "weight" as the Boyce index. This therefore yielded 3 optimal model settings (Boyce index, AUC and the sum of both) for each species. Subsequently, the full 100% of the occurrence data for each species was used to fit a final 'full' model using the identified optimal settings (James et al. 2013, Guisan et al. 2017) and again splitting the occurrence data 50-50 across the training and the validation bin in order to ensure that the checkerboard analyses are comparable to the other three treatments. This full model was finally projected in geographic space to obtain a habitat suitability map. This procedure was repeated for each species and all 4 methods of background selection. Note that only one dataset of background points for each method was used for the creation of SDMs in this entire study.

Model similarity analysis: a measure for model stability

To assess model stability, we used a form of virtual simulations based on the habitat suitability maps of the species generated from SDMs at the previous step. We refer to these SDMs and the matching distribution maps as being of generation 1. For this, we artificially considered the generation 1 prediction maps as "true", meaning a perfect presentation of habitat suitability for the species (as in Thibaud et al. 2014). New occurrence points were then generated from each of these true habitat suitability maps, using the values of the map as probability of occurrence, i.e., the higher the habitat suitability was, the higher the likelihood that an occurrence point got generated there. For each species, 7 new occurrence datasets were created: ones of 10, 30, 50, 100, 300, 500 and 1000 occurrence points. For each of these datasets, the model fitting and selection procedures described in section 2.6 were performed again, minus the splitting up of data into bins of 80% and 20%, i.e., the generation 2 data were divided 50-50, used to train 50 model replicates to identify optimal settings, then re-run with 100% of the data using those optimal settings. We underline that the virtual occurrence datasets were never resampled, i.e., the same occurrence record dataset was used for the 50 MaxEnt runs. These are the models of generation 2. The resulting habitat suitability maps were then compared to the original habitat suitability maps for each species as follows: first, the raster datasets were converted to vectors, and then their similarity was determined using the cor() function in the R language. Our measure of model stability is consequently the measure of spatial similarity between the original map and the map created using the virtual occurrences, i.e., through describing spatial autocorrelation. An overview of the study design is given in figure 3. See

Appendix 1, figure S4 for an idealized presentation of maximum model stability and Appendix 1, Figure S1 and Table S2 for the effect of the number of generated virtual points on stability. Many models, especially generation 2 models with low numbers of generated occurrences (10, 30, sometimes 50) yielded no performance metrics (see Appendix 1, supplementary table S4), which is logical, given the very low sample size and 50-50 random split of data. In such a case, the model was not included in the similarity assessments, because no optimal settings could be selected. As a result, the amount of similarity values differed across background point creation strategies: there were 1032 for stratenv, 887 for randenv, 931 for the classic randgeo, 885 for checkerboard 1 (cb1) and 920 for checkerboard 2 (cb2). We compared all the values of the strategies with Welch t-tests. Since we tested no additional explanatory variables, the t-tests were not corrected for multiple comparisons.

In order to address the different sample sizes, we also did an additional analysis in which we only used Welch t-tests after random elimination of values until the two compared strategies had the same number of values. This method was repeated 100 times over, with a new random sample of eliminated values each time. The Welch t-test was chosen because no equal variance could be assumed among both samples.

Results

The moran’s I was -3.967e-05 for the randgeo dataset, 3.573e-05 for the randenv background points and 0.001 for stratenv background points. Note that the same BP dataset was used for randgeo, cb1 and cb2; it was merely split into different bins for the latter two.

Model performance metrics comparison

Models calibrated for species at the equilibrium yielded significantly higher AUC and AUC+CBI in generation 2 for the stratenv and CB1 treatments than invasive species models. In generation 1, no significant differences were found when comparing invasive species models to equilibrium species models.

When comparing performance metrics between background point selection strategies (regardless of level of species equilibrium with the environment), stratenv yielded higher AUC values than any other strategy in generation 1 models. No additional significant effects were found in this generation. In generation 2, more significant differences were found. Stratenv had higher AUC than any other strategy, and randenv had a higher AUC than all the geographically sampled BP datasets. The same pattern was observed for AUC+CBI. Stratenv also had higher Boyce index than all other treatments, except for randenv, with which there was no significant difference.

In generation 1, when comparing AUC, CBI and AUC+CBI between IAS and EQ species, within each background point selection strategy, no significant differences in performance metrics were found. When comparing results for equilibrium species and invasive alien species within each BP sampling strategy, significant differences in average AUC and AUC+CBI measures were found in generation 2 for stratenv background points, with a slightly higher average for equilibrium species. The same pattern was found in cb1 treatments. See figure 4 for a graphical presentation of these results and Appendix 1, table S1 for the same results in tabular format.

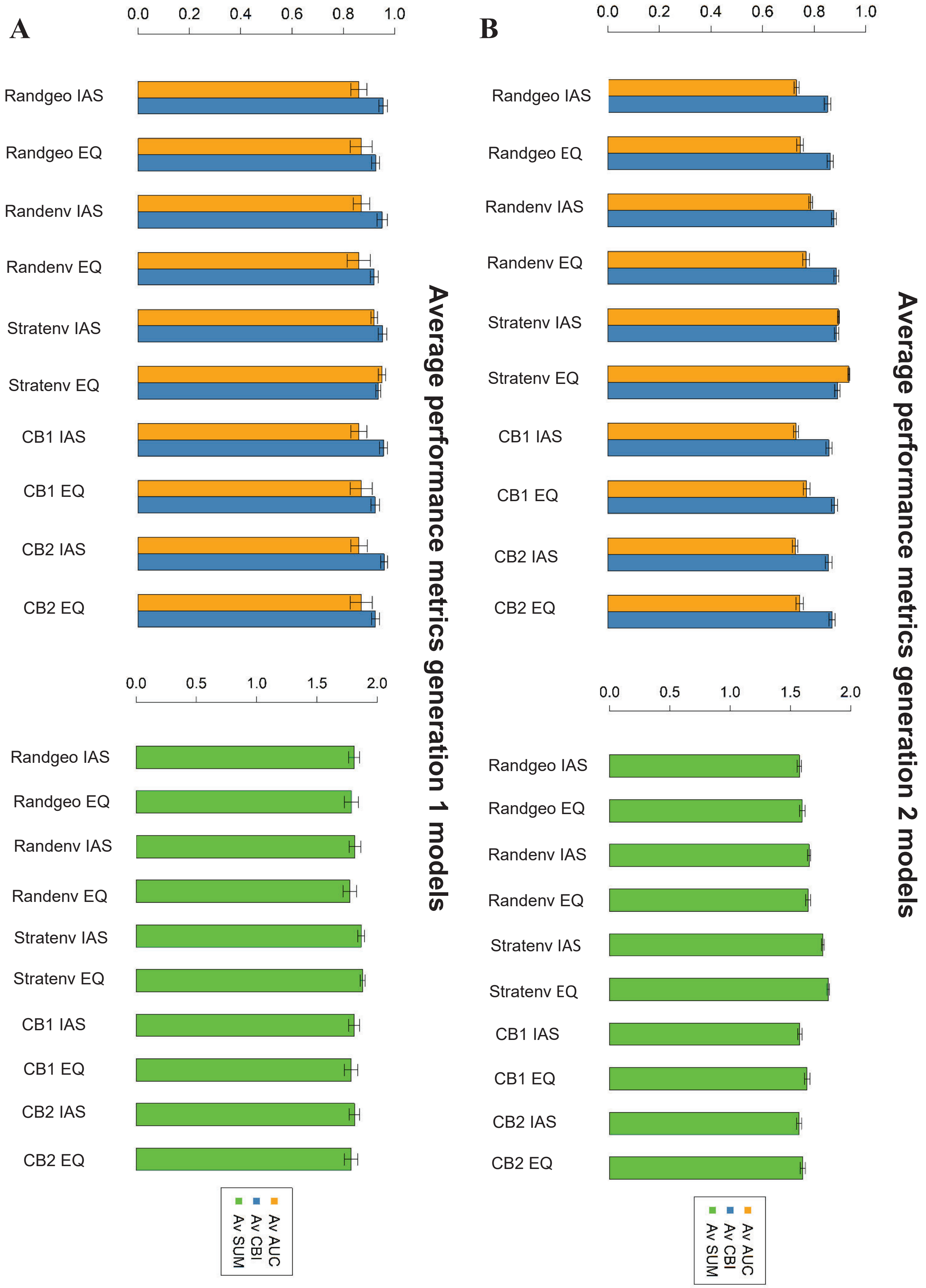

A
0.0
0.2
0.4
0.6
0.8
1.0
Randgeo IAS
Randgeo EQ
Randenv IAS
Randenv EQ
Stratenv IAS
Stratenv EQ
CB1 IAS
CB1 EQ
CB2 IAS
CB2 EQ
0.0
0.5
1.0
1.5
2.0
Randgeo IAS
Randgeo EQ
Randenv IAS
Randenv EQ
Stratenv IAS
Stratenv EQ
CB1 IAS
CB1 EQ
CB2 IAS
CB2 EQ
Av AUC
Av CBI
Av SUM
Average performance metrics generation 1 models
B
0.0
0.2
0.4
0.6
0.8
1.0
Randgeo IAS
Randgeo EQ
Randenv IAS
Randenv EQ
Stratenv IAS
Stratenv EQ
CB1 IAS
CB1 EQ
CB2 IAS
CB2 EQ
0.0
0.5
1.0
1.5
2.0
Randgeo IAS
Randgeo EQ
Randenv IAS
Randenv EQ
Stratenv IAS
Stratenv EQ
CB1 IAS
CB1 EQ
CB2 IAS
CB2 EQ
Av AUC
Av CBI
Av SUM
Average performance metrics generation 2 models

Figure 4: Average performance metrics with error bars for invasive species (IAS) and equilibrium species (EQ) for models of generation 1 (A) and generation 2 (B).

Effects of background points on model stability

When comparing all of the model stability values (i.e., the measure of similarity between a generation 2 model and the matching generation 1 model) between two treatments (i.e., not using the random elimination method), the results of the Welch t-tests indicate that the randenv strategy yields far higher model similarity (i.e., has far higher model stability) than any other strategy, for both invasive species and equilibrium species. Stratenv yields higher similarity values than the classic randgeo (table 2, table 3) and both checkerboard treatments. When randomly eliminating models, the pattern changes slightly. Randenv is still by far the highest in terms of model similarity and stratenv still yields significantly higher results than randgeo, but not higher than either checkerboard treatment anymore (see table 4).

Table 2: Comparison of model similarity (i.e., as a measure of model stability) per background sampling method across all species for method 1 (in which all models were kept.). Stratenv = stratified random in environmental space, randenv = fully random in environmental space and randgeo = fully random in geographic space, cb1 = checkerboard 5 km, cb2 = checkerboard 10 km.

	Stratenv	Randgeo	Randenv	cb1	cb2
Std error	0.0067	0.00748	0.0030	0.0083	0.0075
Average	0.4802	0.4420	0.8113	0.4534	0.4500

Table 3: P-values of t-tests comparing similarity score yielded for models per background point treatments for method 1.

	Stratenv	Randgeo	Randenv	cb1	cb2
Stratenv		0.00014	< 2.2e-16	0.0120	0.0028
Randgeo	0.00014		< 2.2e-16	0.3065	0.4482
Randenv	< 2.2e-16	< 2.2e-16		< 2.2e-16	< 2.2e-16

cb1	0.01202	0.3065	< 2.2e-16		0.7634
cb2	0.0028	0.4482	< 2.2e-16	0.7634	

Table 4: results of background point sampling strategy comparison when randomly eliminating values, to ensure equal sample size (i.e., method 2)

Strategy comparison					
	Stratenv	Randgeo	Randenv	cb1	cb2
Stratenv		100%	100%	0%	0%
Randgeo	100%		100%	100%	0%
Randenv	100%	100%		100%	100%
cb1	0%	100%	100%		0%
cb2	0%	0%	100%	0%	

Effects of equilibrium species vs invasive alien species on model stability

The results were analyzed by both Welch t-tests of all the values and again by random elimination, just as described under section 2.7. Only the stratenv resulted in higher average model stability for equilibrium species than for invasive alien species, but this result was not found to be statistically significant. The other four background treatments (i.e. including two options for the stratified in geographic space: cb1 and cb2) demonstrated the opposite effect, but this time with significant results, except for cb2 (figure 5). The random elimination method however yielded significantly higher similarity values for equilibrium species in the stratenv treatment in 100% of the cases. The strategies randenv, randgeo and cb1 also yielded 100% significant results, but with a higher average for invasive species. For cb2, no difference was found in similarity values for invasive species and equilibrium species. See Appendix 1, table S3 for the same results in tabular format.

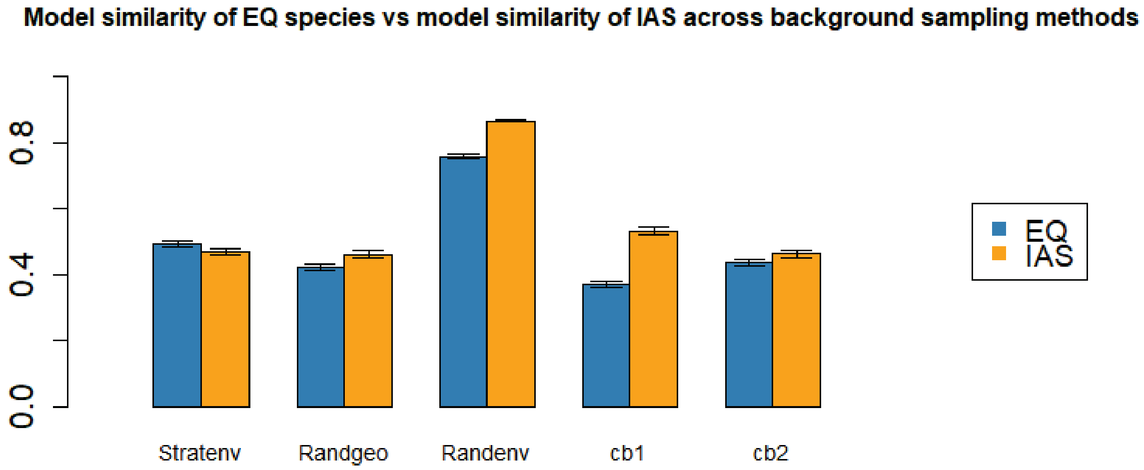


Figure 5: Comparison of model similarity for equilibrium species and invasive alien species per background sampling method. .

Model stability vs performance metrics

The stratified random in environmental space background points models yielded the highest performance metrics, particularly AUC, for both generation 1 and generation 2 models, and the fully random in environmental space treatment yielded the highest model similarity (see figure 6). See Appendix 1, Figure S2 for the same analyses on the CBI.

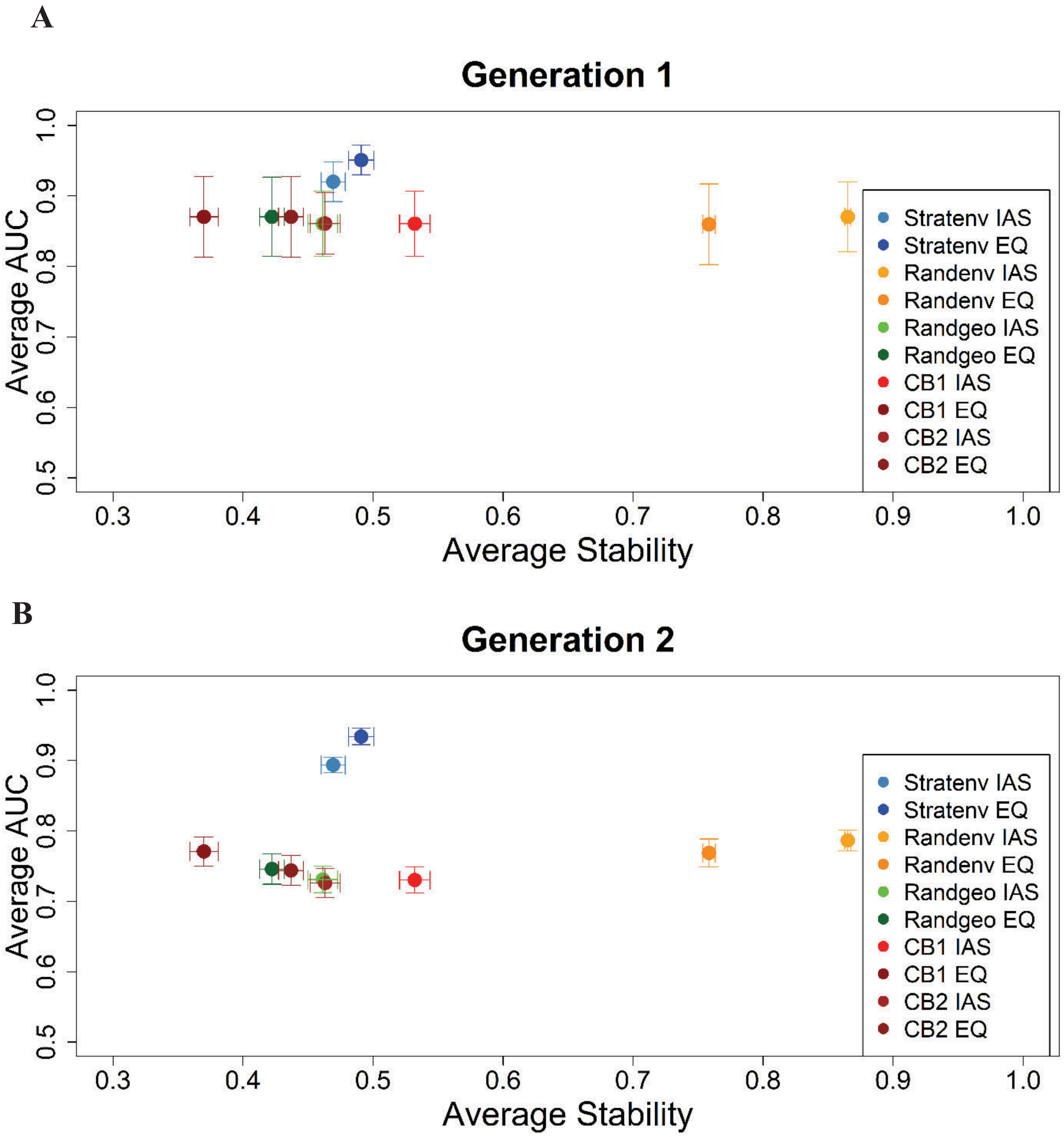


Figure 6: Average model similarity vs average AUC for real species models of generation 1 (the models that were run using real occurrence data) (A) and generation 2 (the models that were run using virtual occurrence data generated from the output suitability maps of generation 1 models) (B). .

Discussion

This research focuses on two questions: (1) Which method of background points sampling produces the best-performing SDMs for invasive and equilibrium species in terms of model accuracy and stability? (2) Do these background point sampling strategies differ between native and invasive alien species?

What drives model stability?

Our results indicate that the classic randgeo sampling of background points produces less stable models (i.e., models with lower similarity values) than both strategies of sampling in environmental space. This is logical and intuitive because the environmental space is used to model the environmental niche (Guisan et al. 2017). In addition, it is also optimal to sample species observations in the field based on environmental space (Hirzel & Guisan 2002). The finding likewise concurs with previous findings that randgeo sampling can cause prediction biases (Botella et al. 2020, Freeman et al. 2022). However, sampling background points in geographic space is still much more common than sampling in environmental space (examples of the latter are e.g. Fragnière et al. 2022, Bazzichetto et al. 2023).

However, the randenv method of background point selection produces by far the most stable (measured by similarity) models for both invasive alien species and species at equilibrium, though the model stability is significantly higher for invasive alien species. This is consistent with the results of Bazzichetto et al. 2023, who used virtual species to show that sampling background points uniformly in environmental space (which is intended to also be fully random in environmental space) produces models that most closely approximate reality for species with wide environmental tolerances, which, in the real world, would include invasive alien species. Therefore, the superior model similarity we found for invasive alien species using this background sampling strategy could be explained by the fact that the stratenv data points are very good at capturing the extremes of the environment, whilst the randenv background points are more continuously spread along the core parts of the environmental gradients. The latter probably informs the models better, particularly the models on invasive alien species, because these are often not present in the extremes of the environment (e.g. there are very few IAS at high elevations (Fuentes-Lillo et al. 2021, Barros et al. 2022)). Therefore, the distribution of invasive alien species and the randenv BP are probably similar. This could be further reinforced by the fact that citizen science datasets like GBIF are subject to human sampling bias (Chauvier et al. 2021) and therefore contain less observations in high mountains, even for alpine species.

What drives model accuracy?

On the side of performance metrics, the stratenv background method yields higher AUC and AUC+CBI (our combined index) for equilibrium species than any other background treatment, particularly in generation 2 models. This is almost entirely due to higher AUC and the effect is noticeable in all equilibrium species models. Still, the ability of the stratenv models to yield higher AUC indicates that they are better at discerning suitable area from unsuitable area. Namely, a possible explanation that they yield higher AUC but not higher Boyce is that the stratenv models include fewer false positives than the other methods, likely due to their ability to classify the absences of the species.

Our results therefore show that although randenv background treatment yields by far the most stable models, it is less good at modelling realistic species distributions than the stratenv treatment. This is likely because the randenv treatment distributes its background points more heavily in environmental values that are common, rather than in the extremes. Both presence points and randenv BP find themselves often in the most common environmental values. Hence, it is often easier for randenv models to identify relationships between the presence points and the environment and therefore create stable models, but they are not necessarily the right relationships. The latter are more likely to be identified by the stratenv background model treatments.

Do optimal BP selection strategies differ between equilibrium and non-equilibrium species?

When answering question 2 (if different BP selection strategies produce the most accurate and stable models for EQ species and IAS), we would expect that EQ species would yield higher performance metrics and more stable models, since SDMs assume that the species is at equilibrium with the environment and that the species occupies its whole realized niche. In this study, it is assumed that the latter condition is satisfied for EQ species, but not for IAS. Stratified sampling in environmental space was the only method where SDMs trained on occurrence data of equilibrium species yielded both higher performance metrics and higher stability values than for non-equilibrium species.

This might be because the stratenv background points are agglomerated into areas where environmental gradients are steep, such as in mountains. Previous studies indicate that model performance is better when background point sampling biases match the bias of the occurrence records (Botella et al. 2020, Freeman et al. 2022, Schartel & Cao (2024)). This may explain why the stratenv background point sampling method, unlike all other methods, yielded higher model similarity for equilibrium species than for invasive alien species: the distribution pattern of these background points matches the distribution pattern of the species. This agrees with the observation that modelling species not at equilibrium with the environment can lead to inaccurate predictions (Freeman et al. 2022). It is therefore possible that the stratenv sampling works better for equilibrium species than for alien species because the stratenv background points are much more abundant around mountains. Hence, the models may be more stable not because the species are at equilibrium, but because they are primarily mountain species. In addition, many of the equilibrium species chosen in this study have relatively narrow distributions, meaning that stratenv background points, which are agglomerated where the environment is heterogeneous, can provide the model with a more complete view of the habitat preferences of narrow-ranged species. This may indicate that the methods have great conservation applications, as narrow-ranged species are both vulnerable to climate change and notoriously difficult to model due to low sample size (Hernández-Lambraño et al. 2020). However, environmental strata are very strongly correlated to topography and elevation, so likely, the random stratified sampling in environmental space informs SDMs better on the margins of the environmental tolerance of each species. Still, we cannot refute the possibility that stratenv sampling is better suited to equilibrium species, regardless of whether these occur primarily in mountains or not.

An additional criticism might be that the spatial autocorrelation of the stratenv BP dataset would lead to errors in model training and predictions. Since this has been proven to be true for occurrence datasets (Veloz 2009, Miller et al. 2012), it might also be true for BP datasets. However, model training is about statistical correlations between occurrences and environmental values, it takes place in environmental space, not in geographic space. Therefore, we do not see how model errors resulting from spatial autocorrelation of the stratenv BP models could artificially inflate the stability values. In addition, our sampling of the stratenv BP data should be (most) uniform along environmental gradients, not biased.

The better stability and performance results for invasive alien species than for equilibrium species yielded by all sampling strategies except stratenv can be explained by their distributions being spread relatively evenly over the study area, just like the distributions of the selected species. The species used for this study are widespread, at an advanced stage of invasion. Sampling strategies that cover the more common environmental values therefore inform the model better. It is also possible that the stratenv background points produce models that are less sensitive to sample size, as the invasive alien species have on average more than twice as many occurrence points than the equilibrium species (see Table 1). A complementary argument is that since the chosen invasive alien species are at an advanced stage of invasion, they could be closer to equilibrium with the environment than initially assumed. This may also explain why the result of stratenv background points yielding higher model similarity results for equilibrium species is significant, but only barely: the invasive alien species are relatively close to equilibrium, as well.

We found that no performance metrics could be generated from many models of generation 2 (i.e., the models built using occurrences generated from the generation 1 habitat suitability maps) that were trained with only 10 or 30 occurrences. This is an additional finding that models trained with small sample size may still yield performance metrics (but see Collart & Guisan 2023) but fail our additional stability test. Our results therefore suggest that it is advisable to not run SDMs with fewer than 50 occurrence records (which has already been shown to be problematic (Adde et al. 2023)), and that testing with performance metrics may not be sufficient to evaluate SDMs. Null models are a possible measure of reliability that may alleviate this concern, as proposed in Collart and Guisan (2023).

Possible expansions

Further investigations expanding on our methods might yield interesting results. For instance, different methods of assessing overlap of species' predictions, e.g., using Schoener's D (e.g. Warren et al. 2008) or Kulczynski's coefficient (e.g. Randin et al. 2006) instead of our correlation method used to calculate model stability could be used in future studies. This might help put our findings in the larger context of other studies, e.g. on calculating niche and distribution overlaps (Warren et al. 2008), or to compare model predictions in transferability assessments (Randin et al. 2006). Another potentially promising method to assess the effect of spatial autocorrelation on models would be to use Dutilleul's modified t-test, which is especially designed to account for the degree of spatial autocorrelation in statistical inference (Dutilleul et al. 1993). This test would thus be mostly ideal when one is interested in comparing two samples in the presence of spatial autocorrelation.

Secondly, limiting the SDM study area to the home range of the respective species might provide valuable insights, as large home ranges may artificially inflate AUC (Lobo et al. 2008). In this study, we have not done so, as the comparability of the models, particularly of the widespread IAS and the narrow-ranged EQ species, would be compromised.

Finally, alternate methods of thinning occurrence data might strongly influence performance metrics. We have already presented an alternate thinning method and the effect on model stability and performance metrics for two species in the supplementary materials (appendix 3), however environmental thinning of occurrence data has been shown to improve model performance metrics (Varela et al. 2014).

In conclusion, from our findings, we suggest sampling background points stratified randomly in environmental space when dealing with 1) low sample sizes, 2) species at equilibrium with the environment and 3) species with narrow ranges. Conversely, it is better to sample the background points randomly in environmental space when species are widespread and have a broad environmental niche. In addition, our results show that stratified random in environmental space background sampling yields both more accurate and more stable models than the classic fully random in geographic space sampling, whilst fully random in environmental space background sampling yields by far the most stable, but not necessarily the most correct models. These findings should however be confirmed by further studies.

Acknowledgements

We give thanks to the Swiss Confederation and to La Sapienza University of Rome for awarding the doctoral grants that made this research possible, as well as to the Institute of Earth Surface Dynamics of the University of Lausanne for providing additional financial support. BS and AG also benefitted from the Swiss national Valpar.CH project for additional support. In addition, we thank the University of Lausanne for granting access to their High Performance Computing (HPC) cluster. Finally, we thank both of our anonymous reviewers for their swift responses, thoroughness, and insights, which have allowed us to greatly improve the scientific work presented here.

References

Alexander, J. M., Chalmandrier, L., Lenoir, J., Burgess, T. I., Essl, F., Haider, S., Kueffer, C., McDougall, K., Milbau, A., Nuñez, M. A., Pauchard, A., Rabitsch, W., Rew, L. J., Sanders, N. J., & Pellissier, L. (2018). Lags in the response of mountain plant communities to climate change. Global Change Biology, 24(2), 563–579. https://doi.org/10.1111/gcb.13976

Anderson, R. P., Araújo, M. B., Guisan, A., Lobo, J. M., Martinez-Meyer, E., Peterson, A. T., & Soberón, J. (2016). Are species occurrence data in global online repositories fit for modeling species distributions. The case of the Global Biodiversity Information Facility (GBIF). Final report of the task force on GBIF data fitness for modelling. Global Biodiversity Information Facility (GBIF), Secretary Office, Aarhus, Denmark.

Araujo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., . . . Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. Science Advances, 5(1). Doi:ARTN eaat485810.1126/sciadv.aat4858

Barber, R. A., Ball, S. G., Morris, R. K. A., & Gilbert, F. (2022). Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Diversity and Distributions, 28*(1), 128-141. Doi:10.1111/ddi.13442

Barbet-Massin, M., Jiguet, F., Albert, C.H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barros, A., Haider, S., Mullerova, J., Alexander J.M., Alvarez M.A., Aschero V., Daehler C., Peyre G., Ratier Backes A., Arévalo J.R., Cavieres L., Dar P., Fuentes-Lillo E., Liedtke R., McDougall K., Milbau A., Morgan J.W., NaylorB.J., Nuñez M.A., Pauchard A., Rashid I., Reshi Z.A., Rew L., Sandoya V., Seipel T., Vorstenbosch T., Vítková M., Neville W., Wedegärtner R.E. M., Zong S., Lembrechts J. J. 2022. The role of roads and trails for facilitating mountain plant invasions. In: Barros, A., Shackleton, R., Rew, L., Pauchard, A. (Eds.). Tourism, Recreation and Biological Invasions. CABI publishing, London. pp 14-26.

Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., . . . Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling, 222* (11), 1810-1819. doi:DOI 10.1016/j.ecolmodel.2011.02.011

Bazzichetto, M., Lenoir, J., Da Re, D., Tordoni, E., Rocchini, D., Malavasi, M., Barták, V., & Sperandii, M. G. (2023). Sampling strategy matters to accurately estimate response curves' parameters in species distribution models. Global Ecology and Biogeography, 32(10), 1717–1729. https://doi.org/10.1111/geb.13725

Botella, C., Joly, A., Monestiez, P., Bonnet, P., & Munoz, F. (2020). Bias in presence-only niche models related to sampling effort and species niches: Lessons for background point selection. PLOS ONE, 15(5), e0232078. https://doi.org/10.1371/journal.pone.0232078

Broennimann O, DiCola V, Guisan A (2023). _ecospat: Spatial Ecology Miscellaneous Methods_.R package version 4.0.0, <http://www.unil.ch/ecospat/home/menuguid/ecospat-resources/tools.html>.

CBD, 2014. Pathways of introduction of invasive species, their prioritization and management. —European Environment Agency. (n.d.). [Policy Document]. Retrieved May 31, 2023, from https://www.eea.europa.eu/policy-documents/cbd-2014-pathways-of-introduction

Chapman, A. D., Belbin, L., Zermoglio, P. F., Wieczorek, J., Morris, P. J., Nicholls, M., . . . Schigel, D. (2020). Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data. *Biodiversity Information Science and Standards, 4*. Doi:10.3897/biss.4.50889

Chauvier, Y., Zimmermann, N. E., Poggiato, G., Bystrova, D., Brun, P., & Thuiller, W. (2021). Novel methods to correct for observer and sampling bias in presence-only species distribution models. *Global Ecology and Biogeography, 30*(11), 2312-2325. Doi:10.1111/geb.13383

Chiocchio, A., Arntzen, J. W., Martínez-Solano, I., de Vries, W., Bisconti, R., Pezzarossa, A., Maiorano, L., & Canestrelli, D. (2021). Reconstructing hotspots of genetic diversity from glacial refugia and subsequent dispersal in Italian common toads (Bufo bufo). *Scientific Reports 2021 11:1*, *11*(1), 1–14. https://doi.org/10.1038/s41598-020-79046-y

Collart F., Guisan A. 2023. Small to train, small to test: Dealing with low sample size in model evaluation. *Ecological Informatics* 75: 102106.
Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 19651-19658. Doi:DOI 10.1073/pnas.0901650106

Cosentino, F., Seamark, E. C. J., Van Cakenberghe, V., & Maiorano, L. (2023). Not only climate: The importance of biotic interactions in shaping species distributions at macro scales. *Ecology and Evolution*, 13(3), e9855. https://doi.org/10.1002/ece3.9855

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(6), 774-787.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1), 27-46. DOI: 10.1111/j.1600-0587.2012.07348.x

Downloading and cleaning GBIF data with R | R-bloggers (n.d.). Retrieved October 6, 2022, from https://www.r-bloggers.com/2021/03/downloading-and-cleaning-gbif-data-with-r/

Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Fan, W.-Y., & Wang, Z.-G. (2014). The Predictive Performance and Stability of Six Species Distribution Models. PLOS ONE, 9(11), e112764. https://doi.org/10.1371/journal.pone.0112764

Dutilleul, P., Clifford, P., Richardson, S., & Hemon, D. (1993). Modifying the t Test for Assessing the Correlation Between Two Spatial Processes. Biometrics, 49(1), 305–314. https://doi.org/10.2307/2532625

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1), 43–57. https://doi.org/10.1111/J.1472-4642.2010.00725.X

European Lepidoptera and their ecology: Erebia meolans. (n.d.). Retrieved February 8, 2024, from http://www.pyrgus.de/Erebia_meolans_en.html

Fletcher, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology, 100*(6). Doi:ARTN e0271010.1002/ecy.2710

Fragnière Y., Gremaud J., Pesenti E., Bétrisey S., Petitpierre B., Guisan A., Kozlowski G., 2022. Mapping habitats sensitive to overgrazing in the Swiss Northern Alps using ecological niche modeling. *Biological conservation*, 274 p. 109742. https://doi.org/10.1016/j.biocon.2022.109742

Franklin, J. (2010). Mapping Species Distributions: Spatial Inference and Prediction: Cambridge University Press.

Freeman, M. S., Dick, J. T. A., & Reid, N. (2022). Dealing with non-equilibrium bias and survey effort in presence-only invasive Species Distribution Models (iSDM); predicting the range of muntjac deer in Britain and Ireland. Ecological Informatics, 69, 101683. https://doi.org/10.1016/j.ecoinf.2022.101683

Fuentes-Lillo, E., Lembrechts, J. J., Cavieres, L. A., Jiménez, A., Haider, S., Barros, A., & Pauchard, A. (2021). Anthropogenic factors overrule local abiotic variables in determining non-native plant invasions in mountains. Biological Invasions, 23(12), 3671–3686. https://doi.org/10.1007/s10530-021-02602-8

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*(11), 1126-1136. Doi:DOI 10.1111/j.1466-8238.2012.00768.x

GBIF.org (03 June 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.tc6fqe

GBIF.org (21 July 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.ns4jzr

GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.mpcxws

GBIF.org (20 May 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.cbxen3

Gomes, V. H. F., Ijff, S. D., Raes, N., Amaral, I. L., Salomão, R. P., de Souza Coelho, L., de Almeida Matos, F. D., Castilho, C. V., de Andrade Lima Filho, D., López, D. C., Guevara, J. E., Magnusson, W. E., Phillips, O. L., Wittmann, F., de Jesus Veiga Carim, M., Martins, M. P., Irume, M. V., Sabatier, D., Molino, J.-F., … ter Steege, H. (2018). Species Distribution Modelling: Contrasting presence-only models with plot abundance data. Scientific Reports, 8(1), Article 1. https ://doi.org/10.1038/s41598-017-18927-1

Gormley, A. M., Forsyth, D. M., Griffioen, P., Lindeman, M., Ramsey, D. S. L., Scroggie, M. P., & Woodford, L. (2011). Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. Journal of Applied Ecology, 48(1), 25-34. doi:10.1111/j.1365-2664.2010.01911.x

Grimmett, L., Whitsed, R., & Horta, A. (2020). Presence-only species distribution models are sensitive to sample prevalence: Evaluating models using spatial prediction stability and accuracy metrics. Ecological Modelling, 431, 109194. https://doi.org/10.1016/j.ecolmodel.2020.109194

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., . . . Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography, 24*(3), 276-292. Doi:Doi 10.1111/Geb.12268

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: Insights from biological invasions. Trends in Ecology & Evolution, 29(5), 260–269. https://doi.org/10.1016/j.tree.2014.02.009

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., . . . Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. Ecology Letters, 16(12), 1424-1435. Doi:Doi 10.1111/Ele.12189

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models, with applications in R.* Cambridge: Cambridge University Press.

Hernández-Lambraño, R. E., Carbonell, R., & Sánchez-Agudo, J. Á. (2020). Making the most of scarce data: Mapping distribution range and variation in population abundance of a threatened narrow-range endemic plant. Journal for Nature Conservation, 57, 125889. https://doi.org/10.1016/j.jnc.2020.125889

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2019). Package 'dismo'. R package version 1.1-4. CRAN. Retrieved from https://cran.r-project.org/web/packages/dismo/index.html

Hirzel, A., & Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling, 157(2), 331–341. https://doi.org/10.1016/S0304-3800(02)00203-X

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(2), 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

hseebens. (2022). Hseebens/DASCO workflow [HTML]. https://github.com/hseebens/DASCOworkflow/blob/30d6467a6f3259c5a0f622fe1d6a8ef4dbf5b159/run_DASCO_workflow.R (Original work published 2021)

Husson, F., Josse, J., Le, S., & Mazet, J. (2020). Package 'FactoMineR'. Multivariate Exploratory Data Analysis and Data Mining. Available on CRAN: https://cran. R-project. Org/web/packages/FactoMineR/FactoMineR. Pdf.

InfoSpecies. (n.d.). Retrieved April 7, 2023, from https://www.infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte invasive_alien_speciesinswitzerlandfactsheets.pdf (n.d.). Retrieved October 6, 2022, from https://www.ecosia.org/search?q=invasive_alien_speciesinswitzerlandfactsheets.pdf

Iturbide, M., Bedia, J., & Gutiérrez, J. M. (2018). Background sampling and transferability of species distribution model ensembles under climate change. Global and Planetary Change, 166, 19–29. https://doi.org/10.1016/j.gloplacha.2018.03.008

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

Jimenez, L., & Soberon, J. (2020). Leaving the area under the receiving operating characteristic curve behind: An evaluation method for species distribution modelling applications based on presence-only data. *Methods in Ecology and Evolution, 11*(12), 1571-1586. Doi:10.1111/2041-210x.13479

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. Scientific Data, 4(1), 170122. https://doi.org/10.1038/sdata.2017.122

Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C. L., Pinilla-Buitrago, G. E., Boria, R. A., Soley-Guarda, M., Anderson, R. P. (2021). ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. Methods in Ecology and Evolution, 12(9), 1602–1608. https://doi.org/10.1111/2041-210X.13628

Land Cover 2050 – Country – Overzicht. (n.d.). Retrieved October 6, 2022, from https://www.arcgis.com/home/item.html?id=3cce97cba8394287bcaf60f7618a5500

List of Invasive Alien Species of Union concern – Environment – European Commission. (n.d.). Retrieved October 6, 2022, from https://ec.europa.eu/environment/nature/invasivealien/list/index_en.htm

Liu, C. R., Newell, G., & White, M. (2019). The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography, 42*(3), 535-548. Doi:10.1111/ecog.03188

Lobo, J. M., Jimenez-Valverde, A., & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. Ecography, 33(1), 103-114. doi:10.1111/j.1600-0587.2009.06039.x

AUC: a misleading measure of the performance of predictive distribution models—Lobo—2008—Global Ecology and Biogeography—Wiley Online Library. (n.d.). Retrieved February 28, 2024, from https://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2007.00358.x

Lyu, S., & Alexander, J. M. (2022). Competition contributes to both warm and cool range edges. Nature Communications, 13(1), Article 1. https ://doi.org/10.1038/s41467-022-30013-3

Mateo, R. G., Felicísimo, Á. M., & Muñoz, J. (2010). Effects of the number of presences on reliability and stability of MARS species distribution models: The importance of regional niche variation and ecological heterogeneity. Journal of Vegetation Science, 21(5), 908–922. https://doi.org/10.1111/j.1654-1103.2010.01198.x

Menéndez, R., Megías, A. G., Hill, J. K., Braschler, B., Willis, S. G., Collingham, Y., Fox, R., Roy, D. B., & Thomas, C. D. (2006). Species richness changes lag behind climate change. Proceedings of the Royal Society B: Biological Sciences, 273(1593), 1465–1470. https://doi.org/10.1098/rspb.2006.3484

Miller, J. A. (2012). Species distribution models: Spatial autocorrelation and non-stationarity. Progress in Physical Geography: Earth and Environment, 36(5), 681–692. https://doi.org/10.1177/0309133312442522

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. *Proceedings of the Royal Society B-Biological Sciences, 278*(1725), 3644-3653. Doi:DOI 10.1098/rspb.2010.2769

OFEV, O. fédéral de l'environnement. (n.d.). Espèces exotiques en Suisse. Retrieved April 7, 2023, from https://www.bafu.admin.ch/bafu/fr/home/themen/thema-biodiversitaet/biodiversitaet–publikationen/publikationen-biodiversitaet/gebietsfremde-arten-in-der-schweiz.html

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., Kassem, K. R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience 51(11):933-938. Url: https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R., Martínez-Meyer, E., Nakamura, M., & Araújo, M. P. (2011). Ecological niches and geographic distributions. Princeton: Princeton University Press.

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., & Guisan, A. (2012). Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. Science, 335(6074), 1344–1348. https://doi.org/10.1126/science.1215933

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. *Ecological Applications, 19*(1), 181-197. Doi:10.1890/07-2153.1

Poland, T. M., Patel-Weynand, T., Finch, D. M., Miniat, C. F., Hayes, D. C., & Lopez, V. M. (Eds.). (2021). Invasive Species in Forests and Rangelands of the United States: A Comprehensive Science Synthesis for the United States Forest Sector. Springer International Publishing. https://doi.org/10.1007/978-3-030-45367-1

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(10), 1689–1703. https://doi.org/10.1111/j.1365-2699.2006.01466.x

Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and poisson point process models for species distribution modeling in ecology. Biometrics, 69(1), 274-281. doi:10.1111/j.1541-0420.2012.01824.x

Santana Marques, P., Resende Manna, L., Clara Frauendorf, T., Zandonà, E., Mazzoni, R., & El-Sabaawi, R. (2020). Urbanization can increase the invasive potential of alien species. Journal of Animal Ecology, 89(10), 2345–2355. https ://doi.org/10.1111/1365-2656.13293

Schartel, T.E. and Cao, Y., 2024. Background selection complexity influences Maxent predictive performance in freshwater systems. Ecological Modelling, 488, p.110592.

scrubr source : R/scrubr-package.R (n.d.). Retrieved October 6, 2022, from https://rdrr.io/cran/scrubr/src/R/scrubr-package.R

Senay SD, Worner SP, Ikeda T (2013) Novel Three-Step Pseudo-Absence Selection Technique for Improved Species Distribution Modelling. PLoS ONE 8(8): e71218. https://doi.org/10.1371/journal.pone.0071218

Slodowicz, D., Descombes, P., Kikodze, D., Broennimann, O., & Müller-Schärer, H. (2018). Areas of high conservation value at risk by plant invaders in Georgia under climate change. Ecology and Evolution, 8(9), 4431–4442. Doi:10.1002/ece3.4005

SpThin: An R package for spatial thinning of species occurrence records for use in ecological niche models—Aiello-Lammens—2015—Ecography—Wiley Online Library. (n.d.). Retrieved May 30, 2023, from https://onlinelibrary.wiley.com/doi/10.1111/ecog.01132

Steen, B., Cardoso, A. C., Tsiamis, K., Nieto, K., Engel, J., & Gervasini, E. (2019). Modelling hot spot areas for the invasive alien plant Elodea nuttallii in the EU. Management of Biological Invasions, 10(1), 151–170. https://doi.org/10.3391/MBI.2019.10.1.10

Steen, V. A., Tingley, M. W., Paton, P. W. C., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution, 12*(2), 216-226. Doi:10.1111/2041-210x.13525

Svenning, J. C., & Sandel, B. (2013). Disequilibrium vegetation dynamics under future climate change. *American Journal of Botany, 100*(7), 1266-1286. Doi:Doi 10.3732/Ajb.1200469

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(9), 947-955. doi:10.1111/2041-210x.12203

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography, 44*(12), 1731-1742. Doi:10.1111/ecog.05615

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs, 92(1). doi:ARTN e0148610.1002/ecm.1486

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(11), 1084–1091. https://doi.org/10.1111/j.1600-0587.2013.00441.x

Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. Journal of Biogeography, 36(12), 2290–2299. https://doi.org/10.1111/j.1365-2699.2009.02174.x

Venables WN, Ripley BD (2002). Modern Applied Statistics with S, Fourth edition. Springer, New York. ISBN 0-387-95457-0, https://www.stats.ox.ac.uk/pub/MASS4/.

Vetaas, O. R. (2002). Realized and potential climate niches: a comparison of four Rhododendron tree species. Journal of Biogeography, 29(4), 545-554.

Vicente, J. R., Vaz, A. S., Queiroz, A. I., Buchadas, A. R., Guisan, A., Kueffer, C., Marchante, E., Cabral, J.A., Nesper, M., Broennimann, O., Godoy, O., Alves, P., Castro-Díez, P., Henriques, R., Honrado, J. P. (2019). Alien Plant Species: Environmental Risks in Agricultural and Agro-Forest Landscapes Under Climate Change. Climate Change Management, 215–234. https ://doi.org/10.1007/978-3-319-75004-0_13/FIGURES/1

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution. Evolution, 62(11), 2868–2883. https://doi.org/10.1111/j.1558-5646.2008.00482.x

Whitford, A.M., Shipley, B.R. and McGuire, J.L., 2024. The influence of the number and distribution of background points in presence-background species distribution models. Ecological Modelling, 488, p.110604.

Wisz, M. S., & Guisan, A. (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data BMC Ecology, 9(8), online. Doi:10.1186/1472-6785-9-8.

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(5), 763-773. doi:10.1111/j.1472-4642.2008.00482.x

CHAPTER III

Optimizing invasive species distribution models: a test with the giant hogweed in Switzerland

Bart Steen*[1,2], Laura Celesti-Grapow[3], Luigi Maiorano**[2], Antoine Guisan**[1,4]

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

[2] Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, viale dell'Università 32, 00185 Roma, Italy

[3] Department of Environmental Biology, Sapienza University of Rome, Piazzale A. Moro 5, 00197 Roma, Italy

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

* Corresponding author: bart.steen@unil.ch / b.r.l.steen2@gmail.com

** Shared last authorship

This paper has not yet been submitted for publication.

My contribution to the project: I co-collected (along with Laura Celesti-Grapow), analyzed and cured the data, performed the analyses and led the writing of the manuscript.

Abstract

Invasive alien species are one of the primary drivers of biodiversity loss and cause great economic damage worldwide. They can be managed, but anticipating their invasions is paramount. This can be accomplished by modelling their current and future distributions. Species Distribution Models are a staple tool for this purpose and recent advances to use them to model invasive alien species have been made and are applied here. In this paper, two common species distribution modelling methods, MaxEnt and Random Forests, are employed to simulate the present and future potential distributions dangerous invasive species, *Heracleum mantegazzianum* Sommier & Levier, using opportunistically collected occurrence datasets, a nested modelling approach and background points sampled stratified randomly in environmental space. Possible bias in the occurrence datasets are treated in several ways and the differences in output distributions are discussed, as well as the uses of the outputs for guiding eradication efforts.

Introduction

Invasive alien species (IAS) are species that are causing economic or natural damage and are exotic in parts of their ranges, causing an estimated $391 billion economic damage per year (IPBES 2023) and posing a great threat to biodiversity (UNEP 2023). Their spread has been greatly facilitated by the world trade and human movements. Moreover, ongoing climate change increasingly allow them to access areas that have remained mostly invasion free until recently, like the polar regions and high mountains (Haider et al. 2022, Pauchard et al. 2009, Petitpierre et al. 2016).

Species Distribution Models (SDMs; Franklin 2010, Guisan et al., 2017) are becoming increasingly more advised as a tool in conservation planning and management (Guisan et al. 2013) but often remain underused in practice (Cassini 2011, Tulloch et al. 2016, Steen et al. 2024a). Recent advancements on creating standards for SDMs (Araujo et al. 2019) and producing accurate predictions of species distributions have been made (Kass et al. in press) and need to be applied in practice to create accurate distribution maps for practitioners, a step that is especially paramount for monitoring IAS (Lake et al. 2020).

For instance, it was determined that sampling background points stratified randomly in environmental space yields the best performance metrics (Steen et al. 2024b). In addition, the use of a spatially-nested approach was shown to be crucial for modelling IAS (Gallien et al. 2012, Adde et al. 2023). The reason is, an optimized SDM for the purposes of management would ideally make use of high-resolution data and cover the entire species range, so that all known environmental conditions in which the species is known to exist are included in the model. Unfortunately, such high-resolution data is only available at local and regional scales, making the task of creating accurate SDMs for the purposes of conservation efforts harder. A spatially-nested approach allows both capturing the whole realized climatic niche over the entire species range at coarse resolution and large extent and fitting the species' fine environmental requirements at high resolution in a smaller extent. This allows overcoming the risk of niche truncation (Chevalier et al. 2021, Adde et al. 2023; Guisan et al. in review) and ensures that all known conditions for the species are accounted for in the case of niche shifts between the native and invaded ranges (Broennimann & Guisan 2008), which can affect IAS (Guisan et al. 2014; Fernandez & Hamilton 2015, Bates et al. 2020, Aravind et al. 2022). Moreover, species presence is likely predicted by different environmental variables at different scales and resolutions, e.g. climate being hypothesized to be more important at large scales while soil or land cover being hypothesized to be more important toward finer scales (Pearson & Dawson 2003, & Dawson 2004, Ni et al. 2007, McGill 2010, Vicente et al. 2014). The nested approach can link the largest scale to the smallest scale, thus allowing for different covariate selection at different scales (Guisan et al. in review).

Another potential drawback of most SDMs is that they are fitted using citizen science datasets, which often suffer from spatial and temporal sampling bias (Bird et al. 2014, Callaghan et al. 2019, Anderson et al. 2020). However, the exact nature of this bias is hardly possible to determine (Di Cecco et al. 2021). Therefore, any efforts to correct the bias by, for instance, subsampling the data might also eliminate valuable ecological information (Steen et al. 2021), making it hard to determine the appropriate treatment of the occurrence data (if any).

Heracleum mantegazzianum is an extensively studied IAS (Pyšek et al. 2007, Shackleton et al. 2020). It causes severe health hazards due to its photoreactive poison (Bhowmik & Chandran 2015) and obstructs ecosystem services like recreational fishing (Caffrey 1999, Nasadiuk & Mamchur 2024). In addition, its ability to form dense monocultures and alter the nutrient environment makes it a severe threat to biodiversity (Caffrey 1999). It invades areas left open by clearcuts and abandoned agricultural ground, mesic grasslands, roadsides, forest edges and eutrophied and iron-polluted soils forming tall-herb stands (Pyšek et al. 2007, Thiele & Otte 2007). For these reasons, *H. mantegazzianum* is both on the List of Union Concern (European Commission 2024) and on the Swiss blacklist of IAS (Infoflora.ch.

The species is spread mainly by humans (Bhowmik & Chandran 2015). It attaches itself to human clothes and shoes, and therefore, spreads through any vehicle humans use. Its natural reproduction is through wind-borne spores (Müllerová 2024). It is a low to mid elevation mountain species originally from the Caucasus mountains and displays generalist behavior in its invaded range. In Germany, it favors abandoned agricultural areas, likely profiting from high nutrient content of soils and absence of land management (Thiele et al. 2007, Thiele & Otte 2008, Thiele et al. 2008). In Europe, it is already found at altitudes above 1500 m (Schackleton et al. 2020; Müllerová 2024). It will probably follow the trend of invaders to spread to still higher elevations (Pauchard et al. 2009, Petitpierre et al. 2016), where it can potentially disturb the many hotspots of alpine biodiversity, which provide invaluable ecosystem services (Grêt-Regamey & Kytzia 2007, Ramel et al. 2020).

In this study, we aim at using novel clues on optimizing SDM modelling approaches (Steen et al. 2024a; Steen et al. 2024b) to build the most optimized SDM for the potential present and future distributions of the IAS *H. mantagazzianum*.

We gathered the most recent findings on building accurate SDMs (e.g. Araujo et al. 2019), making use of multiple treatment strategies, and test them on the giant hogweed. In particular, we implemented a spatially-nested approach (Chevalier et al. 2021; Guisan et al. in review) focusing on 3 different spatial scales: i) Global (i.e. all biomes where *H. mantegazzianum* is present), ii) Europe and iii) Switzerland. We also employed different classes of SDM algorithms and sampled background points in an environmentally random-stratified way (Steen et al. 2024b). In addition, we ran extensive model tuning and overfitting analyses, as well as different methods to deal with sampling bias. Finally, since it is not only important to know the current distribution of invasive species, but also its future distribution, we also projected the models in future climates.

Materials and Methods

Overall analytical framework

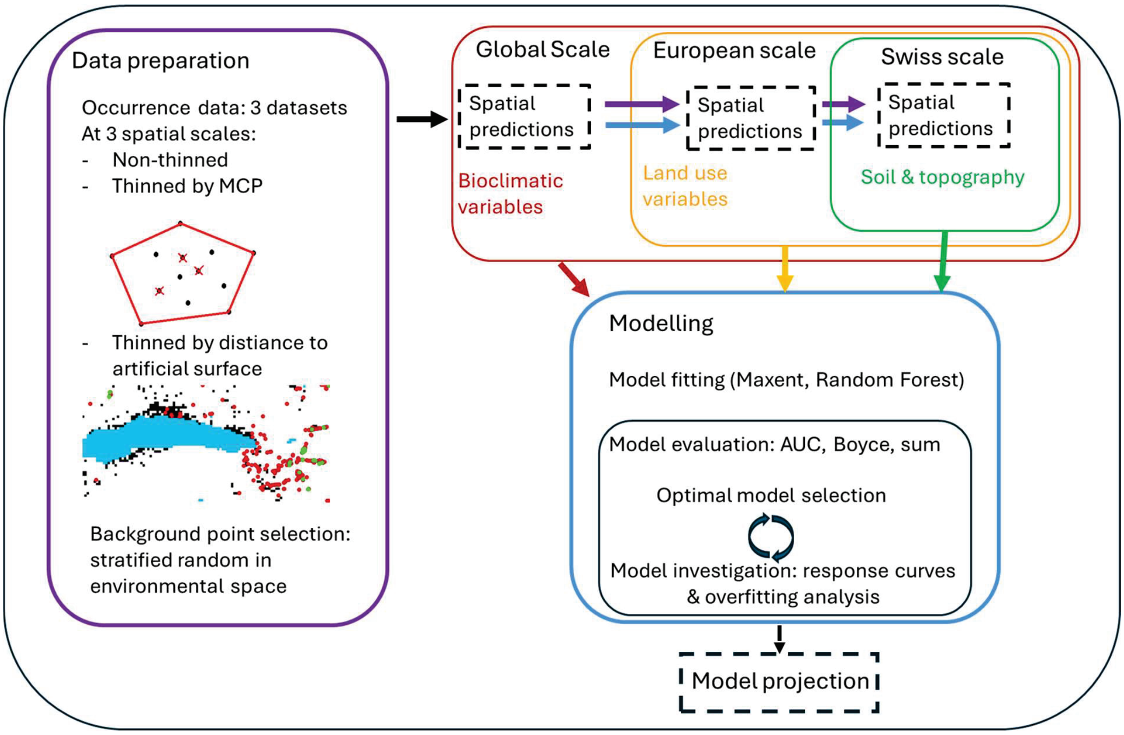


Figure 1: Overall analytical framework of the analyses performed.

Environmental variables

Global bioclimatic variables were downloaded from CHELSA, version 2.1 (Karger et al. 2017), at a resolution of 0.0083 degrees (30 arc seconds). All variables are calculated as the average values from the time series of 1981 – 2010. Karger et al. (2017) provide also multiple projections of the CHELSA 2.1 variables according to several scenarios. In this study, the SSP 370 pathway was chosen for climatic projections.

At European scale, land use variables were used because in Central Europe, *H. mantegazzianum* grows mainly on anthropogenically disturbed areas, nutrient rich soil and abandoned agricultural ground (Thiele et al. 2008, Pergl et al. 2011). For present-day land cover variables, we used the 2015 datasets provided by the European Space Agency (Global ESA CCI land cover classification map). Binary variables included were swampy and often flooded vegetation, mostly cropland and grass, shrub and scrubland. In addition, the gridDistance() function from the R raster package was used to calculate the distance to forest, water, snow and bare/sparsely vegetated areas. Finally, the world population density (cropped to Europe) (Earth Science Data Systems 2024) was inserted as a variable.

At Swiss scale, environmental layers on soil and topography were used. They were taken from the SWECO25 dataset (Külling et al. 2024) and from the Swiss Federal Office of Agriculture (FOAG 2012).

We employed the Ecospat R package (Di Cola et al. 2017) to assess collinearity, at each scale, among all these environmental variables. A threshold of 0.70 for correlation between variables was established, in accordance with Dormann et al. (2013). When two variables had a correlation >0.7, we kept only one of them. The definitive roster of variables chosen for model calibration across all three tiers is outlined in Table 1. For an ecological justification of the inclusion of each variable, see Appendix 2. The biomes where *H. mantegazzianum* does not occur, as well as all water bodies, were removed from all environmental variables.

Table 1: Variables used at each spatial scale to run the Maxent and Random Forest models.

Variable name	**Unit**	**Description**
World scale (30 arc seconds)		
bio2_V.2.1	°C	Annual temperature range
bio12_V.2.1	kg m^-2 year-1	Annual precipitation amount
bio15_V.2.1	kg m^-2	Precipitation seasonality
bio19_V.2.1	kg m^-2 month^-1	Mean monthly precipitation amount of the coldest quarter
Surface wind mean	m/s	Average surface wind velocity
Surface wind max	m/s	Maximum surface wind velocity
Growth degree days	number of days	Annual growth days - days with temperatures above 0 degrees
Growth season precipitation	kg m^-2 gsl^-1	precipitation sum accumulated on all days during the growing season. Gsl = growth season length in days
Growth season temperature	°C	Mean temperature of the growing
Net primary productivity	g C m^−2 yr^−1	Net primary production
Europe scale (30 arc seconds)		
Bioclimatic envelope	N/A	Sum of thresholded world-scale models cropped to Europe scale
Distance to surface water	m	Distance to water. Calculated for every point on land inside the study area
Distance to sparse vegetation	m	Distance from bare or sparse vegetated areas. *H. mantegazzianum* was assumed to not occur in sparse vegetations.
Distance to snow and ice	m	-
Swampy or often flooded vegetation	N/A	-
Mostly cropland	N/A	Vegetation largely consisting of cropland
Grass, scrubs and shrubs	N/A	-
Distance to forest	m	Distance from nearest forest pixel (value = 0 if pixel is forest)
Worldpop_2000	people/km^2	Population density
Swiss scale (25m)		
Europe model		Sum of thresholded Europe-scale models cropped to Swiss scale
Digital elevation model (DEM)	m	Average elevation per 2x2 m square
Aspect	Degrees	Orientation of slope, averaged over 2x2 m squares, varying from 0 to 360
Slope	Degrees	Slope averaged over 2x2 m squares
Hillshade	Index (shaded to bright)	Measure of irradiation per 2x2m square.
Nutrient holding capacity	Index (1-6)	Ordinal measure of soil capacity to hold nutrients
Kultur	Index (1-5)	Ordinal measure of land productivity

Occurrence data

Data were collected from one citizen science dataset (GBIF 2021) and from an unbiased occurrence dataset collected using a scientific sampling scheme (Celesti-Grapow et al. unpublished data). Only occurrence records from the year 2000 or later were selected, in order to match the CHELSA version 2.1 time series. We assumed that the time difference between 1981-2010 and 2000-2021 was representative of the delayed response of the species to shifting climate (Steen et al. 2024b). From the full database we eliminated improbable coordinates, occurrences characterized by coordinate uncertainty exceeding 50 meters, observations reporting zero abundance, and the removal of entries denoting an occurrence status of "absent" across multiple languages. Additionally, a visual inspection, checking for occurrence records where they should not be, was conducted. These data cleaning operations were executed as per Steen et al. (2024b). The three datasets were all used to run the SDMs at all three spatial scales.

Bias correction and data thinning

Given the different sampling designs of the individual datasets, three different versions of the full combined dataset were used. Firstly, a fully non-thinned dataset (so the entire dataset, except for selecting only 1 occurrence record per environmental variable raster grid), a MCP thinned dataset and a dataset based on the distance to artificial surface and urban area. These treatments were chosen because of the differences in sampling design and the different sampling biases. Secondly, an MCP-based thinning method (Steen et al. 2024b) was applied, which counters spatial autocorrelation but also might eliminate ecologically relevant information (Steen et al., 2021). Therefore, the third dataset was thinned based on the assumption that sampling intensity in citizen science databases is higher by human infrastructure. On the other hand, higher abundances of *H. mantegazzianum* along roads and cities might be realistic since humans are a major vector in spreading the species. Due to this uncertainty, we also used the entire, non-thinned occurrence point dataset. In the spatially nested approach, only SDMs run using the same occurrence dataset treatment at larger scale were included in the lower scale models using that treatment.

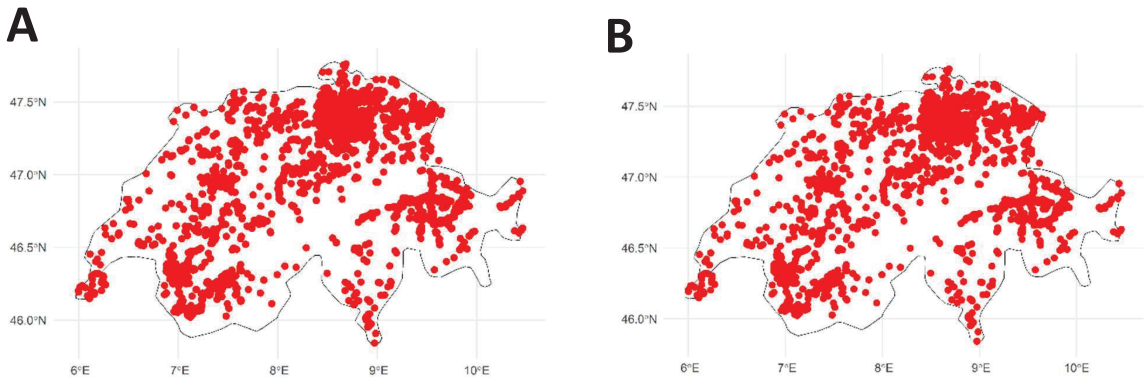
The occurrence data were thinned as per Steen, et al. (2024b) as follows:

we delineated a Minimum Convex Polygon (MCP; Mohr 1947) encompassing the data points. This serves to define the boundaries of the species' home range. Subsequently, within this MCP, we implemented a random point selection method. This procedure was iterated 50 times, and on each iteration, we computed the average nearest neighbor distance between the randomly selected points.

The overall average of these individual averages was then calculated. To reduce sampling bias and align the distribution of species occurrence records with what would be expected in a random distribution, we employed the spThin R package (Aiello-Lammens et al. 2015, Chiocchio et al. 2021).

The data thinned based on distance to artificial surface and urban area were created by extracting the "artificial surface and urban area" from the present-day land use data (Global ESA CCI land cover classification map) and calculating the distance from for each pixel that was not artificial surface or urban area using the gridDistance() function from the R raster package. The values of this raster dataset were then used as a "weight" to sample occurrence records from the non-thinned dataset using the sample() function in R. The number of occurrence points sampled was the same as the number of points remaining after the MCP thinning.

Both thinning steps were repeated at all 3 spatial scales, generating 6 thinned datasets. Since the non-thinned occurrence datasets were also used, a total of nine occurrence datasets have been used to generate the models (three at each spatial scale). The ones built for the Swiss scale are given in Figure 2.



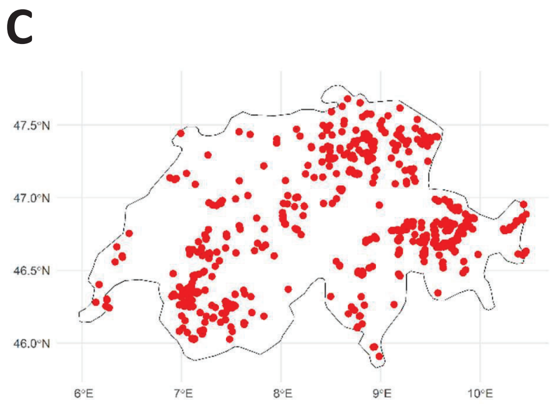


Figure 2: The three occurrence datasets used to fit the SDMs at Swiss scale: no thinning (A), thinned with MCP (B) and thinned by proximity to artificial surface and urban area (C).

Species Distribution Models

Spatially-nested approach

The spatially-nested approach was implemented by running Species Distribution Models (SDMs) using the Maxent algorithm (Phillips et al. 2017) and Random Forests (standard implementation, using the randomForest() function from the randomForest R library (Breiman et al. 2024)). Models were run at three spatial scales: (i) global (from which all the biomes where *H. mantegazzianum* can occur were extracted), (ii) Europe and (iii) Switzerland. The output habitat suitability map of the world model was used as input for the European model, and the output of the European model was in turn used as input for the Swiss model (Adde et al. 2023).

This was done for three reasons. 1) modelling the whole range, at least climatic, at which a species occurs is essential to avoid niche truncation, i.e., underestimation of the realized niche of the species by modelling only a fraction of the range of the environmental variables where the species occurs, 2) it allows for using the fine-resolution data available only at local (here Swiss) scale whilst also using coarser data at higher scales, and 3) it allows for selection of environmental variables most appropriate for predicting the distribution of *H. mantegazzianum* at each of the three scales.

The Maxent and Random Forest algorithms were used to fit the SDMs. These were chosen because they are both widely adopted algorithms (Gomes et al. 2018, Rathore & Sharma 2023) and because we found it appropriate to use one regression-based and one tree-based algorithm.
Maxent draws upon presence-only data, which is much more widely available than data on the absence of species (Guillera-Arroita et al. 2015, Fletcher et al. 2019). The version of the algorithm based on an infinitely logistical regression (Phillips et al. 2017) was used here.

To avoid model overfitting of the MaxEnt models, we varied the mathematical function types (linear, quadratic, hinge, product, threshold) and the Regularization Multiplier (RM), a controller of model complexity: function types were varied linear, quadratic and hinge (LQH), 2) linear, quadratic and product (LQP), and 3) linear, quadratic product and hinge (LQPH). Threshold functions were never used because hinge functions have been shown to be a better replacement (Elith et al. 2011). In addition, the RM was varied from 1 to 5, giving us 15 combinations of settings.

We also used Random Forests, another widely used SDM algorithm. Tuning was done similar to the Maxent models, but the number of trees ("ntree" variable in randomForest function in randomForest package (Random Forests: https://www.stat.berkeley.edu/users/breiman/RandomForests/) was varied between 1000, 1500 and 2000 and the degree of random variation ("mtry" variable in randomForest function in randomForest package) was varied between 1 and 5. The selection of the optimal models was otherwise done in precisely the same way as for the Maxent models, but the performance metrics of the final models at all scales were performed using the out-of-bag (OOB) measures.

Background points selection

At each level of spatial analysis, the environmental variable raster layers were segmented into three strata using the following approach: first, we determined the range between the lowest and highest values within each layer. Next, we selected all the cells whose values fell within each one-third quantile of the dataset, creating three distinct strata within each layer. Each of these strata received a unique numerical identifier that did not repeat in any other layer. Subsequently, we generated a new raster layer in which these numeric codes were summed, resulting in a distinctive code for every strata combination. If the dataset was a factor (as is the case for the binary variables at European level), the cells containing "0" and "1" variables were replaced by a new unique code. Finally, we employed the 'sampleStratified' function from the raster R package to perform sampling across this newly created layer, thus generating stratified random points within the environmental space. The minimum number of points per stratum that yielded at least 10000 background points total were used. Since the model output at higher spatial scales was used as input for the model at lower spatial scale, the background points had to be resampled based on the occurrence point dataset used to generate the former. Therefore, each model at the European and Swiss spatial scales had a unique set of background points.

Performance metrics

We calculated the AUC (Hanley & McNeal 1984) and continuous Boyce index (Hirzel et al. 2006) performance metrics, as well as their sum (Steen at al. 2024b), 100x for each combination and selected the model settings (i.e., rm and function types for MaxEnt, ntree and mtry for RF) that had the highest metrics, splitting the occurrence data randomly between occurrence and validation bins each time. Before computing the sum, the AUC values were first recalculated to the gini AUC (2*AUC-1), in order to give it the same "weight" as the continuous Boyce index. The models with the highest performance metrics were projected on the matching environmental variables and the resulting rasters were summed in order to create the final product at each spatial scale.

If one or more of the models had the same settings, the output map of the model with those settings was used only once in the sum to create the final output maps. This is because the sum of the performance metrics was included in order to include models that had both high AUC and high Boyce, i.e., both high calibration and high discerning measures. If two models have the same settings, then the sum has the same settings as the maximum AUC model, or the Maximum Boyce model, or both. Therefore, it would add no new information to the output to include it more than once and the final summed maps only included models with unique settings.

Variable importance

After selection of the models with the optimal performance metrics, the explanatory power of each of the predictors was determined for all the models. The varImportance() function of the fitMaxnet R package (Wilson 2024) was used for the Maxent models. For Random Forests, the necessary data is automatically saved when using the randomForest() function (Breiman et al. 2024).

Future projections

The CHELSA datasets that were projected by Karger et al. (2017) to the future time period of 2041-2060 using the SSP3-7.0 scenario were used in this study. This scenario was chosen because it occupies a middle ground between medium-high and very-high emissions. The scenario also assumes a fragmented world in which nations largely pursue their own interest and bars game-changing technological advancements (Karger et al. 2017). The world population model projections according to the SSP3-7.0 scenario for the year 2050 were used.

The future land use data were designed by Clarke lab (Esri Land Cover 2050) for the year 2050 (building on the data provided by the European Space Agency, which we used for present day land cover data) and were based on observed trends of historical land use change over the period 2010-2018. Whilst it

does not particularly follow any RCP or SSP scenario, we believed that there were no conflicting assumptions made in any of the datasets.
The future land use categories however did need to be harmonized with the current land use. The code used to accomplish this in the QGIS (QGIS Development Team 2009) raster calculator software is presented in Appendix 1.

There are however no future projections for most of the Swiss-scale model data, as the variables relating to topography were assumed to not change in the future. However, soil and edaphic variables will likely change. Therefore, we made the simplified assumption that the edaphic variables will not change. The future projections at the Swiss scale were therefore done by inserting the outputs at the higher scales as an input variable, as part of our spatially nested approach.

Results

Fit of the models

Variable importance

The predictive power of all the included environmental variables differed slightly for each model fitted at each scale but showed similar patterns overall. For the Maxent models fitted at world scale with non-thinned occurrence data, the annual temperature seasonality (bio15) had by far the highest predictive power, followed by the number of days in the growing season (gdd0). The same was true for the Maxent models fitted with occurrence data thinned with an MCP, but in the models fitted by data thinned by distance to artificial surfaces, the average temperature of the growth season (gst) was the second-strongest predictor (after bio15).

The RF models at world scale were also most powerfully predicted by bio15, with growth season temperature coming second. One of the two RF models fitted with MCP thinned data had Bio19 as the second-most powerful predictor and the other one had bio19 (mean monthly precipitation amount of the coldest quarter).

At Europe scale, by far the most powerful predictor was the world-scale model output. This held true across almost all models, except for Maxent and RF models fitted with data thinned by distance to artificial surface, in which it was distance to permanent snow and ice. Swampy or often flooded vegetation was the second-most powerful predictor for the other Maxent models. In RF models, distance to snow and ice ranked as the second most powerful predictor overall, though for one non-thinned dataset, population density was the second-most powerful predictor.

At Swiss scale, the Maxent models were most powerfully predicted by the slope, with the Europe-scale model variable coming second. For RF models, the most powerful predictor overall was the model fitted at European scale, though the slope was the strongest for RF models that used data thinned with an MCP, and the second-strongest predictor for the other models.

Model evaluation

Overfitting analysis

One Random Forest model at European scale (made with the MCP thinned occurrence dataset) was found to be overfitted (paired t-test: p = 0.047). However, as it was summed with other MCP outputs at that scale and the distribution maps at Swiss scale are the main output of the study, this model was included in the results.

Performance metrics

Both Maxent and RF models showed good performance metrics, with average AUC at world scale always over 0.95 and average Boyce index always over 0.85. On all three spatial scales, RF models had the highest average AUC, sometimes with values of around 0.999, especially for non-thinned data models, indicating an effect of high sample size. One RF model at European scale with non-thinned data (mtry = 2 and ntree = 2000) had only a Boyce index of 0.75, but this was a great outlier.

Present predictions

The final present-day outputs at Swiss scale are presented in Figure 3.

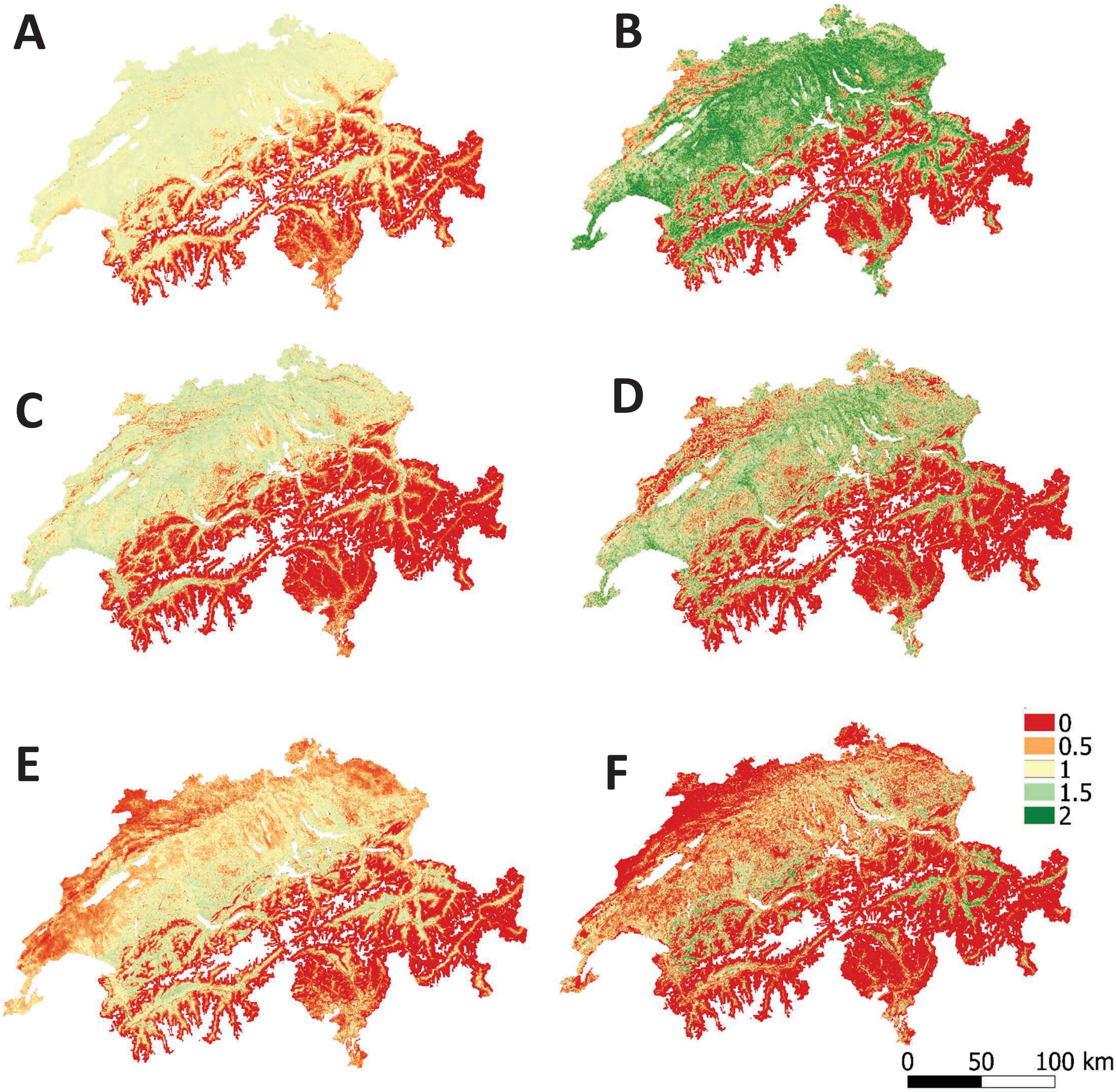


Figure 3: *Summed SDM outputs for* Heracleum mantegazzianum *for the present day, for both Maxent (A, C, E) and Random Forest (B, D, F) models, for all three occurrence datasets: non-thinned (A, B), thinned using a Minimum Convex Polygon (C, D) and thinned by distance to artificial surface and urban area (E, F). The scale bars denote habitat suitability and vary from 0-2 because there were always two models, with a habitat suitability range from 0 to 1.*

Future predictions

The final outputs at Swiss scale for the future are presented in Figure 4.

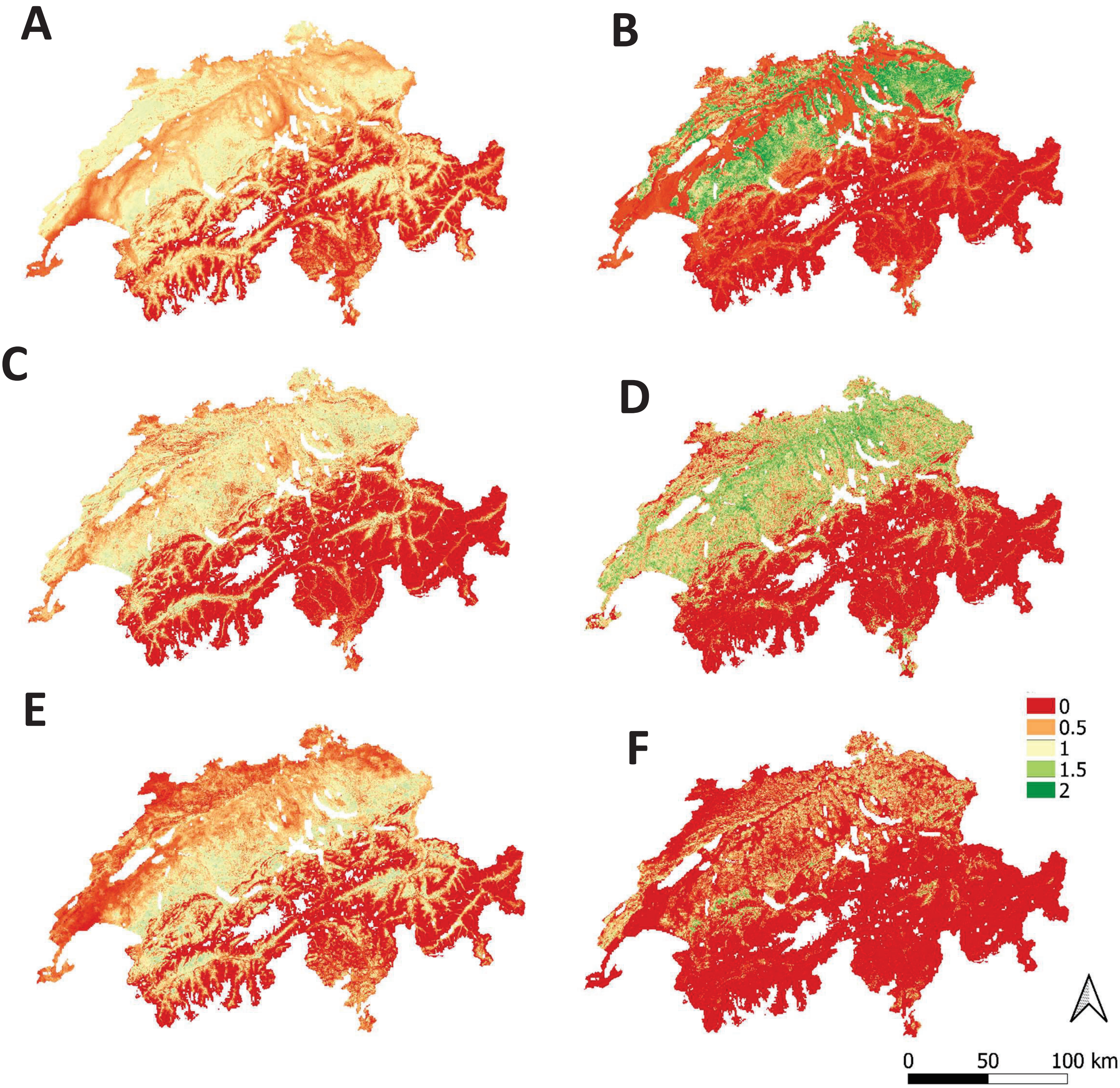


Figure 4: *Summed SDM outputs for* Heracleum mantegazzianum *for the year 2050, for both Maxent (A, C, E) and Random Forest (B, D, F) models, for all three occurrence datasets: non-thinned (A, B), thinned using a Minimum Convex Polygon (C, D) and thinned by distance to artificial surface and urban area (E, F). The scale bars denote habitat suitability and vary from 0-2 because there were always two models, with a habitat suitability range from 0 to 1.*

Discussion

Need for predicting and monitoring

Predicting and monitoring IAS is crucial for mitigating their negative impacts on biodiversity, ecosystem services, and human health (Shackleton et al. 2020). *Heracleum mantegazzianum* is a prime example of an IAS that poses significant threats to nature and economy and effectively managing it is therefore essential. Thankfully, it is possible to effectively manage the species (Nasadiuk & Mamchur 2024). These findings underscore the importance of continuous monitoring and adaptive management strategies to control the spread of invasive species like *H. mantegazzianum* and protect ecological integrity and ecosystem services. Here, we assessed how optimal predictive models could be built for this problematic IAS.

The Swiss Alps are refuges for biodiversity and provide invaluable ecosystem services (Ramel et al. 2020). With climate change, invasive alien plants are likely to shift their distributions to higher altitudes (Petitpierre et al. 2016), jeopardizing the biodiversity hotspots in and ecosystem services provided by the Swiss alps. For long, many mountains were thought to be invasion-free (Pauchard et al. 2009), but global warming is changing the story, and since *H. mantegazzian*um is a mountain species in its native range, its spread to higher altitudes is likely to happen where it invades (Cuddington et al. 2022). This is especially true considering that the species has demonstrated its ability to settle in high mountain areas when escaping from high-altitude botanical gardens (e.g. up to 2200 m; Müllerová 2024).

SDMs are increasingly recommended to be used in conservation studies (Guisan et al. 2013; Steen et al. 2024b), with even the International Union for the Conservation of Nature (IUCN) using them to estimate the species' distributions and to investigate how climate change will affect them (Cassini 2011). Studies have shown that SDMs can predict the potential spread of *H. mantegazzianum*, guiding targeted monitoring and management efforts (Shackleton et al. 2020). This study tested different settings to model the current and future distributions for *H. mantegazzianum* in Switzerland and look for the most optimal approach.

Current distributions

In the present-day models (Figure 3), the habitat suitability for *H. mantegazzianum* is lowest for the models thinned by proximity to artificial surface and urban areas. This approach assumes that citizen science data, such as those from GBIF and Infoflora, are biased to proximity to roads and also more easily accessible areas (Hughes et al. 2021, Díaz-Calafat et al. 2024 for insects). On the flipside, this approach truncates the niche of the species, by biasing the available data to areas further from infrastructure, which in Switzerland are more likely to be high-altitude areas, as the country is very mountainous and the vast majority of the population lives outside of the Alps (EDA, n.d.). It therefore stands to reason that the models that use data thinned by proximity to cities and roads model more rare environmental values, as mountain habitats have very variable climate and topography.

The fact that *H. mantegazzianum*'s distribution is most powerfully predicted by slope in Switzerland may reflect the fact that the species grows the most in the plains and lower lands (as seen in our occurrence data), presently not favoring the mountainous areas that cover much of Switzerland. This is supported by the fact that the plant indeed is more likely to occur at low slopes (see the response curves in Appendix 3).

Future distributions

Looking at the future predictions of habitat suitability in Switzerland (figure 4), the habitat suitability decreases for all models, both Maxent and RF. However, in Maxent models, it does spread to higher up in mountains. This probably reflects the fact that *H. mantegazzianum* is a generalist species in its invaded range, meaning its distribution is mostly inside the most common climatic conditions found in Europe. This is supported by the fact that the model at world scale, fitted with climate variables, is the strongest predictor for the European models. The upward shift is likely representative of the future climatic shift, which would be driven by precipitation and the conditions of the growth season, as indicated by the most powerful predictors, the seasonality of precipitation, the length of the growth season and the temperature of the growth season.

The radical decrease of habitat suitability from present to future for the RF models fitted at Swiss scale with non-thinned occurrence data may be telling of model overfitting. This is more likely to occur with large datasets. The very sharp limit between suitable and unsuitable areas, particularly in the model fitted with non-thinned occurrence data, further supports this theory. In addition, our method of sampling background points stratified randomly in environmental space implies splitting each covariate up in three strata and sample equally from those three. The strongest predictor of the non-thinned data RF models is the summed model output at European scale. For both RF Swiss-scale models with non-thinned data, the probability of occurrence of the species is almost entirely concentrated in the upper third of the variable values (see Appendix 3). It could therefore be that our stratified random in environmental space sampling allows the model to draw a very rigid line between suitable and unsuitable habitat, explaining the dramatic decrease in *H. mantegazzianum* habitat suitability in the future. The same holds true for the RF projections of the models fitted with data thinned by artificial surface and urban area.

It must also be noted that the decreases are more significant for RF models than for Maxent models, again perhaps indicating too strong fitting of data. However, the overfitting analyses of the RF models showed only one overfitted model on the European scale. Therefore, the models may perform very well, statistically, but put overly strict boundaries to the modelled niches. The more lenient predictions for Maxent models would seem to be more realistic for an invasive species, indicating that this algorithm is less inclined to put overly strict constraints on the modelled realized niche.

Implications for management

The usefulness of the output maps likely varies based on the exact spatial focus of conservation efforts. For instance, the models fitted with data from artificial surface and urban area do not match the present-day distribution of *H. mantegazzianum*. Therefore, these are likely not realistic. However, the Maxent models fitted with this dataset do demonstrate un upward shift of the species distribution in the future, and the output maps may therefore still be relevant when identifying areas to subject to scanning for the species in high-mountain habitats. Therefore, as all Maxent models fitted with all data display an upwards spread, they may all be useful for this purpose.

In general, the models fitted with MCP thinned data display the least dramatic shift, which is likely to match reality, due to *H. mantegazzianum*'s invasive and generalist nature. Still, the models fitted with non-thinned datasets display the highest suitability values in Switzerland and are therefore least likely to underestimate *H. mantegazzianum*'s potential distribution, which is important in the case of invasive alien species. However, to use this study's results to identify areas likely to be invaded, and therefore prioritize conservation efforts, it would be more prudent to use the slightly more restrictive MCP model outputs, also due to their moderate future decrease in habitat suitability. Lastly, the output maps in this study could be crossed with other relevant information, such as the spatial locations of biodiversity hotspots, the delivery of ecosystem services known to be impeded by *H. mantegazzianum* (such as recreational fishing) and locations where the health hazards posed by the species could be particularly harmful, such as in the proximity of schools.

References

Adde, A., Rey, P. L., Brun, P., Külling, N., Fopp, F., Altermatt, F., Broennimann, O., Lehmann, A., Petitpierre, B., Zimmermann, N.E., Pellissier, L., & Guisan, A. (2023). N-SDM: A high-performance computing pipeline for nested species distribution modelling. *Ecography*, *2023*, e06540. https://doi.org/10.1111/ecog.06540

Anderson, R. P., Araujo, M. B., Guisan , A., Lobo, J. M., Martinez-Meyer, E., Peterson, A. T. & Soberon, J. M. (2020) Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. Frontiers of Biogeography, 12

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. Science Advances, 5(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Aravind, N.A., Shaanker, M. U., Bhat H. N., P., Charles, B., Shaanker R., U., Shah, M. A., & G, R. (2022). Niche shift in invasive species: Is it a case of "home away from home" or finding a "new home"? Biodiversity and Conservation, 31(11), 2625–2638. https://doi.org/10.1007/s10531-022-02447-0

Bates, O. K., Ollier, S., & Bertelsmeier, C. (2020). Smaller climatic niche shifts in invasive than non-invasive alien ant species. Nature Communications, 11(1), 5213. https://doi.org/10.1038/s41467-020-19031-1

Bhowmik, P. C., Chandran, R. S. (2015) 'Biology, ecology, distribution and current status of Heracleum mantegazzianum Sommier & Levier', Journal of Crop and Weed, Vol. 11(1), pp. 1–17

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., & Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. Biological Conservation, 173, 144–154. https://doi.org/10.1016/j.biocon.2013.07.037

Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. Ecological Modelling, 157(2), 281–300. https://doi.org/10.1016/S0304-3800(02)00200-4

Broennimann, O., Guisan, A., Predicting current and future biological invasions: both native and invaded ranges matter. Biol. Lett. 4, 585-589 (2008).

Caffrey, J. M. (1999). Phenology and long-term control of Heracleum mantegazzianum. In J. Caffrey, P. R. F. Barrett, M. T. Ferreira, I. S. Moreira, K. J. Murphy, & P. M. Wade (Eds.), Biology, Ecology and

Management of Aquatic Plants (pp. 223–228). Springer Netherlands. https://doi.org/10.1007/978-94-017-0922-4_31

Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B., & Major, R. E. (2019). Improving big citizen science data: Moving beyond haphazard sampling. PLOS Biology, 17(6), e3000357. https://doi.org/10.1371/journal.pbio.3000357

Cassini, M. H. (2011). Ranking threats using species distribution models in the IUCN Red List assessment process. Biodiversity and Conservation, 20(14), 3689–3692. https://doi.org/10.1007/s10531-011-0126-9

Celesti-Grapow, L., Barni, E., Brundu, G., Citterio, S., Villani C., et al. & Montagnani C. Distribution of Heracleum mantegazzianum Sommier & Levier in Italy. (in preparation).

Chevalier, M., Broennimann, O., Cornuault, J., & Guisan, A. (2021). Data integration methods to account for spatial niche truncation effects in regional projections of species distribution. Ecological Applications: A Publication of the Ecological Society of America, 31(7), e02427. https://doi.org/10.1002/eap.2427

Cuddington, K., Sobek-Swant, S., Drake, J., Lee, W., & Brook, M. (2022). Risks of giant hogweed (Heracleum mantegazzianum) range increase in North America. Biological Invasions, 24(1), 299–314. https://doi.org/10.1007/s10530-021-02645-x

Díaz-Calafat, J., Jaume-Ramis, S., Soacha, K., Álvarez, A., & Piera, J. (2024). Revealing biases in insect observations: A comparative analysis between academic and citizen science data. PLOS ONE, 19(7), e0305757. https://doi.org/10.1371/journal.pone.0305757

Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., & Hurlbert, A. H. (2021). Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. BioScience, 71(11), 1179–1188. https://doi.org/10.1093/biosci/biab093

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 (6), 774–787.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carr´e, G., Marqu´ez, J.R.G., Gruber, B., Lafourcade, B., Leit˜ao, P.J., Münkemüller, T., Mcclean, C., Osborne, P.E., Reineking, B., Schr¨oder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013.

Esri Land Cover 2050. (n.d.). Retrieved December 5, 2024, from https://github.com/vannizhang/land-cover-2050#readme

Earth Science Data Systems, N. (2024). Global 1-km Downscaled Population Base Year and Projection Grids Based on the Shared Socioeconomic Pathways, Revision 01 | NASA Earthdata [Dataset]. Earth Science Data Systems, NASA. https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-pd-sspbsyr-1km-1.01

EASIN - European Alien Species Information Network. (n.d.). Retrieved December 9, 2024, from https://easin.jrc.ec.europa.eu/spexplorer/search/searchpaged

Fernández M, Hamilton H (2015) Ecological Niche Transferability Using Invasive Species as a Case Study. PLoS ONE 10(3): e0119891. https://doi.org/10.1371/journal.pone.0119891

Fletcher, R.J., Hefley, T.J., Robertson, E.P., Zuckerberg, B., McCleery, R.A., Dorazio, R. M., 2019. A practical guide for combining data to model species distributions. Ecology 100 (6), e02710. https://doi.org/10.1002/ecy.2710. ARTN.

FOAG (2012). *Carte des aptitudes des sols de la Suisse*. (n.d.). Retrieved December 5, 2024, from https://www.blw.admin.ch/fr/carte-des-aptitudes-des-sols

Franklin, J. (2010). Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press. https://doi.org/10.1017/CBO9780511810602

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(11), 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x

GBIF.org (09 October 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.34phaw

Geography – Facts and Figures. (n.d.). Retrieved December 9, 2024, from https://www.eda.admin.ch/aboutswitzerland/en/home/umwelt/geografie/geografie---fakten-und-zahlen.html

Global ESA CCI land cover classification map (1992-2015)- A Big Earth Data Platform for Three Poles. (n.d.). Retrieved December 9, 2024, from https://poles.tpdc.ac.cn/en/data/c205fc4f-4847-4a7d-bb04-7c60f27438ae/

Gomes, V.H.F., Ijff, S.D., Raes, N., Amaral, I.L., Salom˜ao, R.P., de Souza Coelho, L., de Almeida Matos, F.D., Castilho, C.V., de Andrade Lima Filho, D., L´opez, D.C., Guevara, J.E., Magnusson, W.E., Phillips, O.L., Wittmann, F., de Jesus Veiga Carim, M., Martins, M.P., Irume, M.V., Sabatier, D., Molino, J.-F., ter Steege, H., 2018. Species Distribution Modelling: contrasting presence-only models with plot abundance data. Sci. Rep. 8 (1), 1 https://doi.org/10.1038/s41598-017-18927-1.

Grêt-Regamey, A., & Kytzia, S. (2007). Integrating the valuation of ecosystem services into the Input–Output economics of an Alpine region. Ecological Economics, 63(4), 786–798. https://doi.org/10.1016/j.ecolecon.2007.02.026

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Wintle, B.A., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Glob. Ecol. Biogeogr. 24 (3), 276–292. https://doi.org/ 10.1111/Geb.12268.

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). Habitat suitability and distribution models: with applications in R. https://doi.org/10.1017/9781139028271

Guisan, A., Chevalier, M., Adde, A., Zarzo-Arias, A., Goicolea, T., Broennimann, O., Petitpierre, B., Scherrer, D., Rey, P.-L., Collart, F., Riva, F., Steen, B., Mateo, Rubén G. (in Review). Spatially-nested species distribution models (N-SDM): an effective tool to overcome niche truncation for better projections. Journal of Ecology

Haider, S., Lembrechts, J. J., McDougall, K., Pauchard, A., Alexander, J. M., Barros, A., Cavieres, L. A., Rashid, I., Rew, L. J., Aleksanyan, A., Arévalo, J. R., Aschero, V., Chisholm, C., Clark, V. R., Clavel, J., Daehler, C., Dar, P. A., Dietz, H., Dimarco, R. D., … Seipel, T. (2022). Think globally, measure locally: The MIREN standardized protocol for monitoring plant species distributions along elevation gradients. Ecology and Evolution, 12(2), e8590. https://doi.org/10.1002/ece3.8590

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

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.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. Ecography, 44(9), 1259–1269. https://doi.org/10.1111/ecog.05926

Infoflora Lists. (n.d.). Retrieved December 9, 2024, from https://www.infoflora.ch/en/neophytes/lists.html

IPBES. (2023). Summary for Policymakers of the Thematic Assessment Report on Invasive Alien Species and their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (H. E. Roy, A. Pauchard, P. Stoett, T. Renard Truong, S. Bacher, B. S. Galil, P. E. Hulme, T. Ikeda, K. V. Sankaran, M. A. McGeoch, L. A. Meyerson, M. A. Nuñez, A. Ordonez, S. J. Rahlao, E. Schwindt, H. Seebens, A. W. Sheppard, & V. Vandvik, Eds.). IPBES Secretariat. https://doi.org/10.5281/zenodo.7430692

European Commission - Invasive alien species (2024, December 3). https://environment.ec.europa.eu/topics/nature-and-biodiversity/invasive-alien-species_en

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P. & Kessler, M. (2017) Climatologies at high resolution for the earth's land surface areas. Scientific Data 4, 170122. https://doi.org/10.1038/sdata.2017.122

Kass J., Smith A., Warren D., Vignali S., Schmitt S., Aiello-Lammens M., Arlé E., Barbosa A.M., Broennimann O., Cobos M.E., Guégen M., Guisan A., Merow C., Naimi B., Nobis M., Ondo I., Osorio-Olvera L., Owens A., Pinilla-Buitrago G., Sánchez-Tapia A., Thuiller W., Valavi R., Velazco S., Zizka A., Zurell D. (In press). Achieving higher standards in species distribution modeling by leveraging the diversity of available software. Ecography.

Külling, N., Adde, A., Fopp, F., Schweiger, A. K., Broennimann, O., Rey, P.-L., Giuliani, G., Goicolea, T., Petitpierre, B., Zimmermann, N. E., Pellissier, L., Altermatt, F., Lehmann, A., & Guisan, A. (2024). SWECO25: A cross-thematic raster database for ecological research in Switzerland. *Scientific Data*, *11*(1), 21. https://doi.org/10.1038/s41597-023-02899-1

T. A. Lake, R. B. D. Runquist, D. A. Moeller (2020). Predicting range expansion of invasive species: Pitfalls and best practices for obtaining biologically realistic projections. Diversity and Distributions 26, 1767-1779.

Land Cover 2050—Country—Aperçu. (n.d.). Retrieved December 7, 2024, from https://www.arcgis.com/home/item.html?id=3cce97cba8394287bcaf60f7618a5500

McGill, B. J. (2010) Matters of Scale. Science, 328, 575-576.

Mohr, C. O. (1947). Table of Equivalent Populations of North American Small Mammals. American Midland Naturalist, 37(1), 223. https://doi.org/10.2307/2421652

Müllerová, J. (2024). Heracleum mantegazzianum (giant hogweed). CABI Compendium, CABI Compendium, 26911. https://doi.org/10.1079/cabicompendium.26911

Nasadiuk, P., & Mamchur, Z. (2024). Review of modern approaches to the management of invasive species of the genus Heracleum in Europe. Visnyk of Lviv University. Biological series, 92, Article 92. https://doi.org/10.30970/vlubs.2024.92.06

Ni, J., Wang, G. H., Bai, Y. F., & Li, X. Z. (2007). Scale-dependent relationships between plant diversity and above-ground biomass in temperate grasslands, south-eastern Mongolia. *Journal of Arid Environments*, *68*(1), 132–142. https://doi.org/10.1016/j.jaridenv.2006.05.003

Pearson, R. G., Dawson, T. E. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology & Biogeography 12, 361-372.

Pearson, D. & Dawson, T. P. (2004) Modelling species distribution in Britain: a hierarchical integration of climate and land-cover data. Ecography, 27, 285-298.

Pauchard, A., Kueffer, C., Dietz, H., Daehler, C.C., Alexander, J., Edwards, P.J., Arévalo, J.R., Cavieres. L.A., Guisan, A., Haider, S., Jakobs, G., McDougall, K., Millar, C.I., Naylor, B.J., Parks, C.G., Rew, L.J., Seipel, T. (2009). Ain't no mountain high enough: Plant invasions reaching high elevations. Frontiers in Ecology and the Environment, 7, 479-486.

Pergl, J., Müllerová, J., Perglová, I., Herben, T., & Pyšek, P. (2011). The role of long-distance seed dispersal in the local population dynamics of an invasive plant species. Diversity and Distributions, 17(4), 725–738. https://doi.org/10.1111/j.1472-4642.2011.00771.x

Petitpierre, B., McDougall, K., Seipel, T., Broennimann, O., Guisan, A., & Kueffer, C. (2016). Will climate change increase the risk of plant invasions into mountains? Ecological Applications, 26(2), 530–544. https://doi.org/10.1890/14-1871

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. Ecography, 40(7), 887–893. https://doi.org/10.1111/ecog.03049

Pyšek, P., Cock, M. J. W., Nentwig, W., & Ravn, H. P. (2007). Ecology and Management of Giant Hogweed (Illustrated edition). CABI.

QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation. URL http://qgis.org

Ramel, C., Rey, P.-L., Fernandes, R., Vincent, C., Cardoso, A. R., Broennimann, O., Pellissier, L., Pradervand, J.-N., Ursenbacher, S., Schmidt, B. R., & Guisan, A. (2020). Integrating ecosystem services within spatial biodiversity conservation prioritization in the Alps. *Ecosystem Services*, *45*, 101186. https://doi.org/10.1016/j.ecoser.2020.101186

Rathore, M. K., & Sharma, L. K. (2023). Efficacy of species distribution models (SDMs) for ecological realms to ascertain biological conservation and practices. Biodiversity and Conservation, 32(10), 3053–3087. https://doi.org/10.1007/s10531-023-02648-1

Random Forests. (n.d.). Retrieved December 9, 2024, from https://www.stat.berkeley.edu/users/breiman/RandomForests/

Breiman, L., Cutler, A., Liao, A. Wiener, M. (2024). randomForest: Breiman and Cutlers Random Forests for Classification and Regression (Version 4.7-1.2) [Computer software]. https://cran.r-project.org/web/packages/randomForest/index.html

a. Steen, B., Adde, A., Schlaepfer, M. A., Guisan, A., & Maiorano, L. (2024). Distributions of non-native and native plants are not determined by the same environmental factors. Ecological Solutions and Evidence, 5(4), e12374. https://doi.org/10.1002/2688-8319.12374

b. Steen, B., Broennimann, O., Maiorano, L., & Guisan, A. (2024). How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels. Ecological Modelling, 493, 110754. https://doi.org/10.1016/j.ecolmodel.2024.110754

Steen, V. A., Tingley, M. W., Paton, P. W. C., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. Methods in Ecology and Evolution, 12(2), 216–226. https://doi.org/10.1111/2041-210X.13525

Thiele, J., Otte, A., Eckstein, R.L., 2007. Ecological needs, habitat preferences and plant communities invaded by Heracleum mantegazzianum. In: *Ecology and management of giant hogweed (Heracleum mantegazzianum)* [ed. by Pyšek, P., Cock, M.J.W., Nentwig, W., Ravn, H.P.]. Wallingford, UK: CABI. 126-143.

Thiele, J., Otte, A., 2008. Invasion patterns of Heracleum mantegazzianum in Germany on the regional and landscape scales. Journal for Nature Conservation, 16(2):61-71.

Thiele, J., Schuckert, U., Otte, A., 2008. Cultural landscapes of Germany are patch-corridor-matrix mosaics for an invasive megaforb.*Landscape Ecology*, 23(4):453-465.

Tulloch, A. I. T., Sutcliffe, P., Naujokaitis-Lewis,I., Tingley, R., Brotons, L., Ferraz, K., Possingham, H. P., 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. https://doi.org/10.1016/j.biocon. 2016. 04. 023

United Nations Environment Program (2023, August 29). Invasive Alien Species Report | UNEP - UN Environment Programme. https://www.unep.org/resources/report/invasive-alien-species-report

Steen, V. A., Tingley, M. W., Paton, P. W. C., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. Methods in Ecology and Evolution, 12(2), 216–226. https://doi.org/10.1111/2041-210X.13525

Vicente, J. R., Gonçalves, 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. https://doi.org/10.1016/j.ecocom.2014.10.005

Wilson, P. D. (2024). Peterbat1/fitMaxnet [R]. https://github.com/peterbat1/fitMaxnet (Original work published 2020)

CHAPTER IV

How much do out-of-the-native-range occurrences of alien invasive species reflect their native niche?

Bart Steen*[1,2], Luigi Maiorano**[2], Antoine Guisan**[1,3]

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

[2] Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, viale dell'Università 32, 00185 Roma, Italy

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

* Corresponding author: bart.steen@unil.ch / b.r.l.steen2@gmail.com

** Shared last authorship

This paper is under review in Ecography.

My contribution to the project: I conceived and developed (in cooperation with Luigi Maiorano and Antoine Guisan) the idea, analyzed the data, drew the figures and led the writing of the manuscript.

Abstract

Invasive alien species (IAS) present a major threat to biodiversity and ecosystem services. Managing and controlling them is of paramount importance. To guide conservation efforts, detailed maps of potential distributions of IAS, as for instance produced by species distribution models (SDMs), are very useful, but lacking in the dimension of probability of invasion success. An important barrier to the successful establishment of IAS is climate. The Niche Margin Index provides a measure of (dis)similarity between the native climatic niche of IAS and the environmental conditions in another site (which, for these purposes, would be a (potential) invasion site). This measure was shown to be positively correlated with invasion success in invasive vertebrates. For invasive plants, however, special caution is advised to pick the method to define the native climatic niche. This study examines the effect on NMI of using three different methods to define the native range: the full native range polygons, the geographical occurrences of the species inside the native range, and a thresholded SDM inside the native range. The last method occupies the middle ground between the first and second, which present the highest and lowest NMI, respectively (i.e. occurrences < SDMs < range polygons). This has implications for nature conservation and invasion science, as it points out new considerations that must be made before the probability of invasion success is assessed in plant species.

Introduction

Invasive alien species (IAS), defined as species that are not native to the region in question and have been proven to cause ecological and/or economic damage (Polce et al. 2023) are one of the five main drivers of biodiversity loss (IPBES 2023). Species Distribution Models (SDMs; Guisan et al. 2017, Peterson et al. 2011, Franklin 2010) are increasingly advised to be used in nature conservation planning (Guisan et al. 2013, Tulloch et al. 2016, Steen et al. 2024a), and are now being used for this purpose, e.g. by IUCN (Cassini 2011). Whilst SDMs have great potential for foreseeing invasions by IAS, using them implies making assumptions. One such presupposition is that the potential of the species to establish itself in any location with suitable habitat is unlimited, which is unrealistic. There are many biological and geographical impediments to the successful establishment of colonizing plants (Richardson & Pyšek 2006, Saul et al. 2017, Broennimann et al. 2021).

In truth, for a biotic invasion to be successful, several hurdles need to be crossed. Firstly, the species must disperse to the site of invasion, then it must find adequate local environmental conditions (e.g., presence of soil, nutrition, sunlight, shelter, favorable climate, etc.) and face biotic barriers (e.g., competition, predation, etc.) (Richardson & Pyšek 2006).

Once these barriers have been crossed, meaning once IAS have established themselves firmly, they become more difficult and costly to combat (Plues et al., 2012, as noted in Broennimann et al. 2017). Adding the dimension of probability of successful establishment therefore has great conservation applicability. One key filter that a species needs to pass through is climate (Steen et al. 2024). Broennimann et al. (2021) have shown that for invasive mammals, the successful invasion of a new area is positively correlated to the similarity of the native climatic niche (NCN; climatic niche in the native range of a species) to the climatic niche in the invaded range (Broennimann et al. 2021). They designed a new measure called the Niche Margin Index (NMI) that measures how much the conditions at a site are inside (innerness) but also, and this is new, outside (outerness) the NCN. The NMI is a continuous measure from niche innerness to outerness, i.e., how much inside or outside of the NCN a site in the invaded range lies. Positive NMI implies that it lies within the NCN, whilst a negative one indicates it lies outside the margin, and the absolute value tells us how much inside or outside it lies. Averaging NMI values across many occurrences of a species in its invaded range can therefore inform on how much the invaded climatic niche (ICN) is similar or not to the NCN, i.e. inside or equal to the NCN or partly outside (i.e., expanding).

Broennimann et al. (2021) used the native range maps (not the SDM outputs) produced by IUCN based on expert judgment to calculate the NMI. This may be realistic for mammals, as they are highly mobile. However, for plants, using the whole native range is likely an overestimation of the species' true native realized niche (Díaz-Calafat et al. 2024). This can be remedied by calculating the NMI using only occurrence points in the native range of the species, however, this is likely to be an underestimation. Scientific fieldwork does not provide enough information to quantify the niche (Araújo & Guisan 2006) and citizen science datasets likely suffer from biases (Bird et al. 2014, Callaghan et al. 2019, Anderson et al. 2020, Hughes et al. 2021, Díaz-Calafat et al., 2024). The reality probably lies somewhere in between and may be approximated by SDMs, since the latter would not be needed if either expert-drawn maps or occurrence points of the species were adequate to typify a species' realized niche. Yet, this has never been assessed.

In this study, we investigate how using all three previously described types of data (occurrences, SDMs, range polygons) to calculate the NMI affects the NCN-matching of occurrences outside the native range. We collected occurrence data for 45 invasive plant species and calculated the NMI for each occurrence point outside of the native range of the species. Then, we compared the results among the three methods. We hypothesize that 1) the high extreme will be the full native range values, 2) the lowest will be the occurrences only, and 3) that the SDMs will occupy the middle ground. We then discuss what the most realistic approximation of the NCN might be, and how this study is a seed for further research in adding the dimension of probability of establishment to IAS management plans.

Materials and Methods

NMI is a value of distance to niche margin in environmental hyperspace (Broennimann et al. 2021). It ranges between -1 and -∞, with the exact limit of the species' niche, i.e., the NCN margin, lying at 0. An NMI value of 1 indicates that the evaluated site has conditions at the center of the NCN, whilst negative NMI values indicate that the conditions of the evaluated site are outside the NCN, and the more negative the NMI, the farther away the site is from the NCN margin. Niche dissimilarity can in theory be infinite as environmental hyperspace can be, too. However, when using climatic values available on earth, the distance to the margin is within known values.

From the website of the European Alien Species Information Facility (EASIN), a list of 88 tracheophytes that are invasive in at least one member state of the European Union was obtained. For each of these species, occurrence data was downloaded from Global Biodiversity Information Facility (GBIF) (GBIF.org 2024). The downloaded data had a coordinate inaccuracy of 50m or less (Steen et al. 2019, Steen et al. 2024b), an occurrence status of "Present" and the basis of the record had to be "Observation", "Human Observation", "Living Specimen", or "Occurrence evidence". The data were subsequently cleaned as per Steen et al. 2024b. Subsequently, the GIFT R package was used to determine each species' native range (Denelle et al. 2023). The GIFT database contained no data on the native range of some species. In addition, sometimes there were less than 10 occurrences per species either inside or outside of the native range. Finally, some species had less than 70 occurrence points in the native range, which was deemed insufficient for running a SDM (Steen et al. 2024b). Those species had to be excluded, and 45 species remained. For the final list of species, see Table 1. The total number of occurrence points among the 45 species was 1089468, of which there were 252372 in the native ranges and 837096 in the invaded ranges. For the number of occurrence points per species, see Appendix 1.

Table 1: Final list of species included for NMI calculation

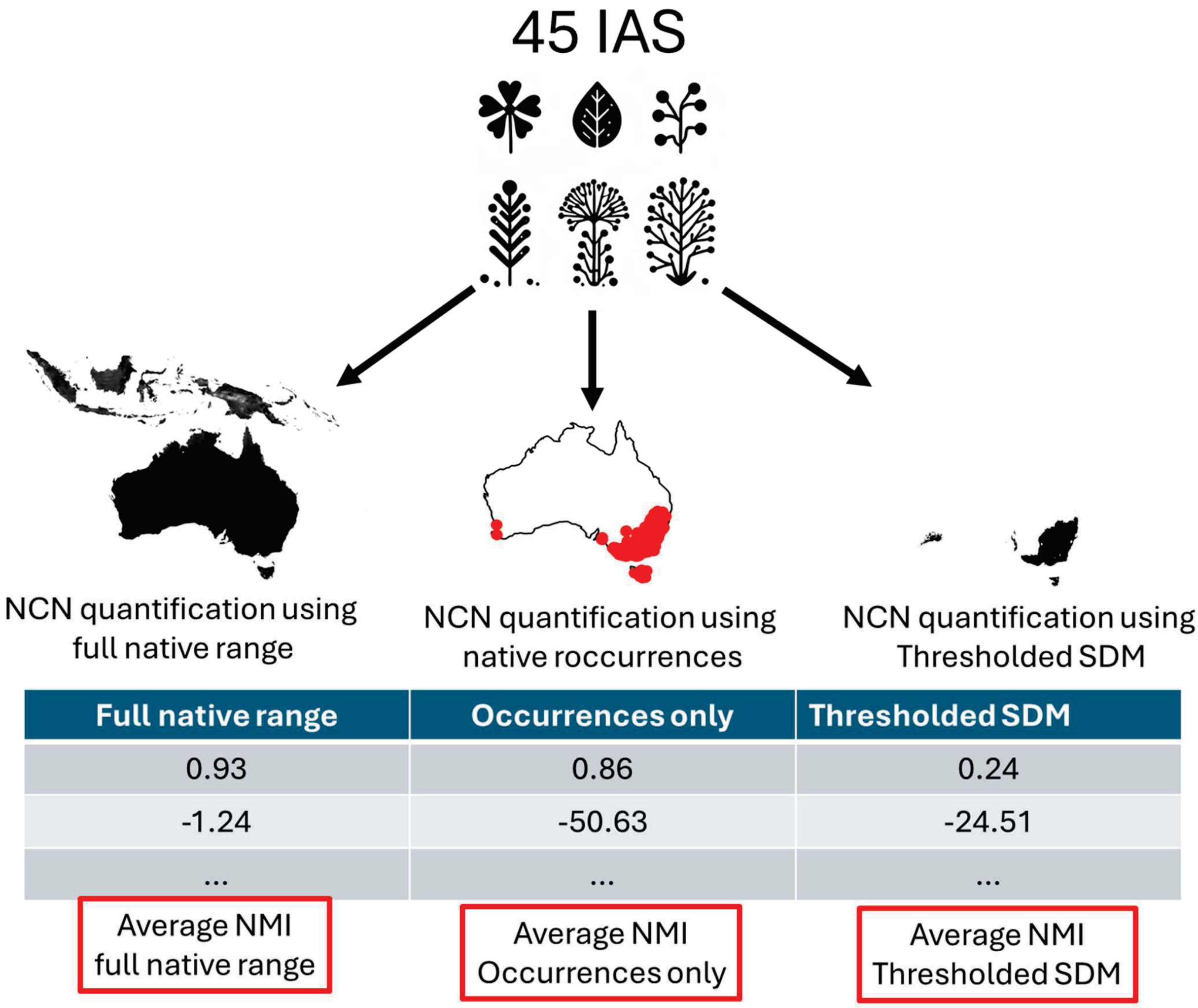
Acacia dealbata
Acer negundo
Ageratina riparia
Ailanthus altissima
Allium triquetrum
Alternanthera philoxeroides
Ambrosia artemisiifolia
Araujia sericifera
Arbutus unedo
Arundo donax
Baccharis halimifolia
Buddleja davidii
Cortaderia selloana
Cotula coronopifolia
Crassula helmsii
Cylindropuntia fulgida
Cylindropuntia imbricata
Cylindropuntia prolifera
Echinocystis lobata
Erigeron annuus
Eschscholzia californica
Gypsophila paniculata
Helianthus tuberosus
Heracleum mantegazzianum
Leucaena leucocephala
Ludwigia peploides
Lupinus polyphyllus
Lysichiton americanus
Prunus serotina
Reynoutria japonica
Rhododendron ponticum
Ricinus communis
Robinia pseudoacacia
Rumex confertus
Senecio inaequidens
Solidago canadensis

Calotropis procera
Carpobrotus edulis
Cenchrus setaceus
Hyacinthoides hispanica
Impatiens glandulifera
Impatiens parviflora
Spartium junceum
Tradescantia fluminensis
Vachellia farnesiana

For the sake of comparability, no thinning of occurrence points of any kind was performed before calculating NMI or before running SDMs.

The NMI value was calculated using the code available on Github (ecospat 2023) for the occurrence data of all species outside the native range (so for all their worldwide introductions), as determined by the GIFT package. We therefore considered each occurrence point in the introduced range of a species a separate invasion event. The NMI was calculated in three different ways: 1) by using all environmental data inside the native range of the species, 2) by using only the occurrence data in the native range of the species, 3) by using the environmental data inside a thresholded SDM of the native range. This threshold was the maxTSS() function from the Ecospat R package (Di Cola et al. 2017; Broennimann et al. 2025). The SDMs were run with the maxnet R package (Phillips et al. 2017) and tuned as per Steen et al. (2024b). In addition, the background points of the model were sampled stratified randomly in environmental space, and all occurrence points of the species inside its native range were used for fitting the models. Finally, the variables CHELSA (V.2.1) bioclimatic variables at 30 arc second resolution were downloaded (Karger et al. 2017): daily temperature range (bio2), temperature seasonality (bio4), temperature of the coldest quarter (bio10), temperature of the warmest quarter (bio11), precipitation of the driest quarter (bio16), precipitation of the coldest quarter, (bio17) and precipitation of the warmest quarter (bio18). These variables were chosen because they are commonly used to quantify invasive species' niches (including plants, animals and fungi) at global scale (Bellard et al. 2013; Broennimann et al. 2021).

For an overview of the study design, see Figure 1.



Full native range	Occurrences only	Thresholded SDM
0.93	0.86	0.24
-1.24	-50.63	-24.51
…	…	…

Figure 1: Design of the study. NCN = Native climatic niche.

As the NMI values were calculated for all occurrence points in the invaded ranges of the species, the main result of the study was three sets of 837096 NMI values.

Results

The main results are presented in Figure 2. Paired t-tests (with equal variances assumed) showed that all three datasets of NMI values were very significantly different from each other (p < 2.2e-16).

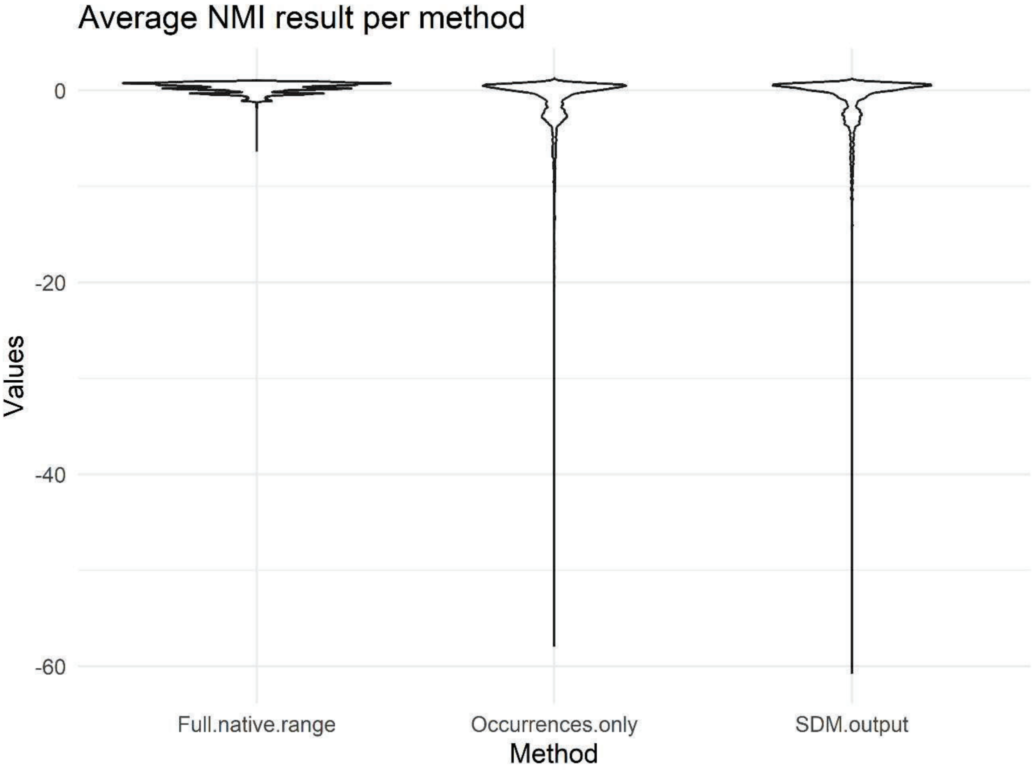


Figure 2: Violin plots of all introduced occurrence points per method of native range selection.

The precise NMI averages were 0.298558 for the full native range, -0.9106878 for the native occurrences, and -0.857308 for the thresholded SDM.

Discussion

Our results confirm the hypothesis. As expected, the average NMI output calculated with the SDM output occupies the middle ground between the outputs calculated using the whole native range, which shows the highest average NMI, and the occurrence points only, which shows the most negative average NMI.

The high level of niche similarity between occurrence points in the native range is explained by the (comparatively) high variation in climatic conditions inside the native range. As the species colonize new areas across a range of environmental conditions (Richardson & Pyšek 2006, Broennimann et al. 2007, Steen et al. 2024b), these are more likely to be included already in the native range.

The index calculated using only the occurrence points in the native range is negative, likely because much less variation can be included in the occurrence points as compared to the whole climatic niche, even if there are many occurrence records. In addition, IAS often experience a niche shift when spreading to a new area and occupying new niches compared to their home range (Broennimann et al. 2007). This may also be contributory to the native range NMI results, though the broad range of environmental values in this method still compensates for this, to a degree.

Finally, the SDM output lies in between the other two, but is far closer to the output with only native occurrences than to the output of the whole native range. Apparently, the thresholded SDM output is far more restrictive than the full home range and therefore includes a far smaller heterogeneity of environmental conditions. The similarity to the results obtained using occurrence points only may be caused by the fact that the SDMs were run on the exact same data as were used in this dataset. An additional explanation could be that no thinning of records was done at any stage, thus maintaining the possible biases of the citizen science datasets, which could make it easier for the models to discern patterns. However, this cannot be verified, as it is unknown if the same biases are persistent across all 45 target species. This absence of thinning may have induced the SDMs of some species with many occurrence records to be tightly fit to the occurrence data, thus drawing a too restrictive border between suitable and unsuitable habitat. However, as there is a high variability in the number of occurrence data (See Appendix 1), this is unlikely to be the case for all species.

It must also be noted that the NMI value only pertains to the potential start of an invasion, not to when the species is already well established and has spread, which is certainly the case for many of the focal species in large parts of their invaded range. As such, the dissimilarity may be artificially reduced (i.e., becoming closer to 0 or more negative) in all the three sets of results.

Nevertheless, a possible interpretation of our results is that using the whole native range is a vast overestimation of the environmental values where the species occurs. The contrary, that this method produces the best results as it shows more similarity with the native range, is also a possible interpretation, but an unlikely one, since invasive plants are well known to spread to very different environmental conditions when invading (Broennimann et al. 2007, Bates & Bertelsmeier 2021).

Whilst we cannot categorically conclude the best way to quantify the native climatic niche from our results, theory suggests that the SDM method would be the closest approximation, since expert-based maps are likely overestimations of the native range of plant species (Díaz-Calafat et al. 2024) and occurrence data provide an insufficient overview, regardless of whether they are sampled by citizen scientists (which suffer from many biases (Bird et al. 2014, Callaghan et al. 2019, Anderson et al. 2020, Hughes et al. 2021, Díaz-Calafat et al., 2024)) or using scientific protocols (Araújo & Guisan 2006). In our results, SDMs provide the middle ground, even despite their strong similarity to NMI values calculated from occurrence points only (though they were still significantly different from each other).

This has implications for invasion science and conservation planning. The potential distribution of IAS, as modelled by for instance SDMs, is very valuable, but can be refined to more closely approximate the true distribution of IAS. For IAS to invade successfully, a number of environmental and biotic barriers need to be crossed (Richardson & Pyšek 2006, Saul et al. 2017, Broennimann et al. 2021). Climate is probably among the most important ones of these (Steen et al. 2024b). This study presents new considerations relating to assessing the probability of invasion.

In conclusion, the method of quantifying the native niche produces vastly different results. Therefore, we advise caution when trying to add the dimension of establishment success to SDMs, even despite the great conservation applications (Plues et al., 2012, as noted in Broennimann et al. 2017). Steps that must be considered are taxonomy (vertebrates, tracheophyte, etc.) and the method of defining a species' native range. Further research is required to investigate the effects of native niche quantification for different taxa, as well as to validate whether the SDM method indeed approximates the native range the best.

References

Anderson, R. P., Araújo, M. B., Guisan, A., Lobo, J. M., Martínez-Meyer, E., Peterson, A. T., & Soberón, J. M. (2020). Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. Frontiers of Biogeography, 12(3). https://doi.org/10.21425/F5FBG47839

Bates, O. K., & Bertelsmeier, C. (2021). Climatic niche shifts in introduced species. Current Biology, 31(19), R1252–R1266. https://doi.org/10.1016/j.cub.2021.08.035

Bellard, C., Thuiller, W., Leroy, B., Genovesi, P., Bakkenes, M., & Courchamp, F. (2013). Will climate change promote future invasions? Global Change Biology, 19(12), 3740–3748. https://doi.org/10.1111/gcb.12344

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., & Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. Biological Conservation, 173, 144–154. https://doi.org/10.1016/j.biocon.2013.07.037

Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T., & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. Ecology Letters, 10(8), 701–709. https://doi.org/10.1111/j.1461-0248.2007.01060.x

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(4), 481–497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Broennimann, O., Petitpierre, B., Chevalier, M., González-Suárez, M., Jeschke, J. M., Rolland, J., Gray, S. M., Bacher, S., & Guisan, A. (2021). Distance to native climatic niche margins explains establishment success of alien mammals. Nature Communications, 12(1), Article 1. https://doi.org/10.1038/s41467-021-22693-0

Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B., & Major, R. E. (2019). Improving big citizen science data: Moving beyond haphazard sampling. PLOS Biology, 17(6), e3000357. https://doi.org/10.1371/journal.pbio.3000357

Cassini, M. H. (2011). Ranking threats using species distribution models in the IUCN Red List assessment process. Biodiversity and Conservation, 20(14), 3689–3692. https://doi.org/10.1007/s10531-011-0126-9

Denelle, P., Weigelt, P., & Kreft, H. (2023). GIFT—An R package to access the Global Inventory of Floras and Traits. Methods in Ecology and Evolution, 14(11), 2738–2748. https://doi.org/10.1111/2041-210X.14213

Díaz-Calafat, J., Jaume-Ramis, S., Soacha, K., Álvarez, A., & Piera, J. (2024). Revealing biases in insect observations: A comparative analysis between academic and citizen science data. PLOS ONE, 19(7), e0305757. https://doi.org/10.1371/journal.pone.0305757

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 (6), 774–787.

ecospat. (2023). Ecospat/NMI [R]. https://github.com/ecospat/NMI (Original work published 2020)

ecospat: Vignettes/vignette_ecospat_package.Rmd. (n.d.). Retrieved January 17, 2025, from https://rdrr.io/cran/ecospat/f/vignettes/vignette_ecospat_package.Rmd

Franklin, J., 2010. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press.

GBIF.org (2 September 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.czn8hv

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). Habitat suitability and distribution models: with applications in R. https://doi.org/10.1017/9781139028271

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. Ecol. Lett. 16 (12), 1424–1435. https://doi.org/10.1111/ Ele.12189.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. Ecography, 44(9), 1259–1269. https://doi.org/10.1111/ecog.05926

IPBES (2023). Summary for Policymakers of the Thematic Assessment Report on Invasive Alien Species and their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Roy, H. E., Pauchard, A., Stoett, P., Renard Truong, T., Bacher, S., Galil, B. S., Hulme, P. E., Ikeda, T., Sankaran, K. V., McGeoch, M. A., Meyerson, L. A., Nuñez, M. A., Ordonez, A., Rahlao, S. J., Schwindt, E., Seebens, H., Sheppard, A. W., and Vandvik, V. (eds.). IPBES secretariat, Bonn, Germany. https://doi.org/10.5281/zenodo.7430692

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P. & Kessler, M. (2017) Climatologies at high resolution for the earth’s land surface areas. Scientific Data 4, 170122. https://doi.org/10.1038/sdata.2017.122

NMI/script_NMI.R at master · ecospat/NMI. (n.d.). GitHub. Retrieved January 6, 2025, from https://github.com/ecospat/NMI/blob/master/script_NMI.R

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). Ecological Niches and Geographic Distributions. Princeton University Press. https://doi.org/10.1515/9781400840670

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. Ecography, 40(7), 887–893. https://doi.org/10.1111/ecog.03049

Pluess, T., Jarošík, V., Pyšek, P., Cannon, R., Pergl, J., Breukers, A., & Bacher, S. (2012). Which Factors Affect the Success or Failure of Eradication Campaigns against Alien Species? PLOS ONE, 7(10), e48157. https://doi.org/10.1371/journal.pone.0048157

Polce, C., Cardoso, A. C., Deriu, I., Gervasini, E., Tsiamis, K., Vigiak, O., Zulian, G., & Maes, J. (2023). Invasive alien species of policy concerns show widespread patterns of invasion and potential pressure across European ecosystems. Scientific Reports, 13(1), 8124. https://doi.org/10.1038/s41598-023-32993-8

Richardson, D. M., & Pyšek, P. (2006). Plant invasions: Merging the concepts of species invasiveness and community invasibility. Progress in Physical Geography: Earth and Environment, 30(3), 409–431. https://doi.org/10.1191/0309133306pp490pr

Saul, W.-C., Roy, H. E., Booy, O., Carnevali, L., Chen, H.-J., Genovesi, P., Harrower, C. A., Hulme, P. E., Pagad, S., Pergl, J., & Jeschke, J. M. (2017). Assessing patterns in introduction pathways of alien species by linking major invasion data bases. Journal of Applied Ecology, 54(2), 657–669. https://doi.org/10.1111/1365-2664.12819

Steen, B., Cardoso, A.C., Tsiamis, K., Nieto, K., Engel, J., Gervasini, E. (2019) Modelling hot spot areas for the invasive alien plant Elodea nuttallii in the EU. Management of Biological Invasions 10(1): 151–170, https://doi.org/10.3391/mbi.2019.10.1.10

A Steen, B., Adde, A., Schlaepfer, M. A., Guisan, A., & Maiorano, L. (2024). Distributions of non-native and native plants are not determined by the same environmental factors. Ecological Solutions and Evidence, 5(4), e12374. https://doi.org/10.1002/2688-8319.12374

B Steen, B., Broennimann, O., Maiorano, L., & Guisan, A. (2024). How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels. Ecological Modelling, 493, 110754. https://doi.org/10.1016/j.ecolmodel.2024.110754

Tulloch, A. I. T., Sutcliffe, P., Naujokaitis-Lewis, I., Tingley, R., Brotons, L., Ferraz, K., Possingham, H. P., 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. https://doi.org/10.1016/j.biocon.2016.04.023

Synthesis and perspectives

Overview of the thesis findings

The work presented in this PhD thesis provides novel insights to fill the knowledge gaps related to building niche-based statistical models that can be used to explore the spread and establishment of invasive alien species (IAS). Through this thesis I explored 4 main research questions (see the Introduction for more details) organized into 4 chapters and with clear answers published or submitted to peer-reviewed journals (see Fig. 1 for a visual illustration of the most important findings).

Research question 1) What are the differences in environmental variables most effective in building SDMs for native and non-native species?

Taking an original modeling approach based on modelling congeneric pairs of species, Chapter 1 has pointed out that climate is a very strong determinant for the distribution of both native and non-native species. In addition, the distributions of non-native species are more powerfully predicted including anthropogenic variables and human-modified land cover types when compared to native congenerics. These results were published in Ecological Solutions and Evidence (Steen et al. 2024a).

Research question 2) Does sampling background points in environmental space improve SDM performance?

In Chapter 2 I compared four ways to sample background points and assessed the corresponding performance of the SDMs with two measures: accuracy and stability. I tested four background point sampling schemes: geographically random, geographically stratified random, environmentally random and environmentally stratified random. I found that sampling background points stratified randomly in environmental space creates the most accurate models and sampling them fully randomly in environmental space creates the most stable models. Both accuracy and stability of predictions are vital in conservation planning, regardless of whether they focus on IAS or not. To perform the analyses, we also developed a novel method to evaluate model stability. These findings have been published in Ecological Modelling (Steen et al. 2024b).

Research question 3) Does the optimal method of background point sampling differ between species that are at equilibrium and species that are at disequilibrium (e.g., IAS)?

Chapter 2 also pointed out that species that are (or can be assumed to be) at equilibrium with the environment show better model accuracy and stability than IAS (which are at disequilibrium with the environment by definition) only when sampling background points stratified randomly in environmental space. All the other methods showed the opposite pattern, which was surprising, because in theory SDMs are expected to work better for species that are at equilibrium. The explanations invoked for that unexpected result were that the IAS considered in this study had by far higher sample sizes on average, making it easier for the models to determine patterns, and that the equilibrium species were (in this study) primarily mountain species, which had their occurrences concentrated where the environmentally stratified background points were more abundant than those of the other sampling schemes. Finally, stratified random sampling in environmental space is likely to inform SDMs better of the extremes of species' tolerances, making them possibly more useful to model rare species.

Chapter 3 was the application of the above findings. It demonstrated the use of the stratified random in environmental space method of background point sampling in a nested SDM approach to model the potential distributions of the IAS *Heracleum mantegazzianum* in Switzerland and presents projections of the models to the future. This chapter also touched on whether this method of background point sampling could draw a too rigid line between suitable and unsuitable habitat and thus could overly constrain the environmental conditions where the species can occur, i.e., underestimate the species' ability to live outside of its known environmental niche. Thus, results may not be transferable to other geographic and temporal ranges, as suggested by others (e.g., Goicolea et al. 2024, Luan et al. 2024, Randin et al. 2020).

Research question 4) How can we quantify the susceptibility of an area to invasion? In other words, how do we add the dimension of native niche outerness, and not just innerness, to the probability that a certain site will be invaded once an IAS arrives there?

Chapter 4 focused on assessing the probability of invasion success also outside the native niche conditions of the species (i.e., including native niche outerness), a dimension that is not covered by SDMs (which include only native niche innerness) and which addresses the climatic barriers that IAS can cross under some circumstances to potentially establish in a new area and especially in new conditions. The Niche Margin Index (NMI) method, designed by Broennimann et al. (2021), was used here to measure the niche (dis)similarity between the climatic conditions at the location of the occurrence points of 45 invasive alien plants species outside their native range and the climatic niche in their native range. I.e., the degree by which the climatic conditions at each occurrence point outside the native range lay inside or outside the native niche was quantified. Subsequently, all the NMI values were averaged. More specifically, the native niche was quantified in three different ways - using the full native range polygons, using only the occurrence of the species inside the native range and using a thresholded SDM inside the native range. The gradual differences in the quantified niche and distribution sizes between the three measures, with occurrence points resulting in underestimating niches and distributions and range polygons in overestimating them, pointed out that caution is advised when quantifying the native climatic niche with different types of data. Specifically, the highest similarity is acquired when using the full native range polygons, the lowest is acquired when using only the native occurrence records and the SDM method occupies the middle ground between the two. Though the SDM method occupies the middle ground, more research is required to point out the nuances involved with selecting different data types for different groups.

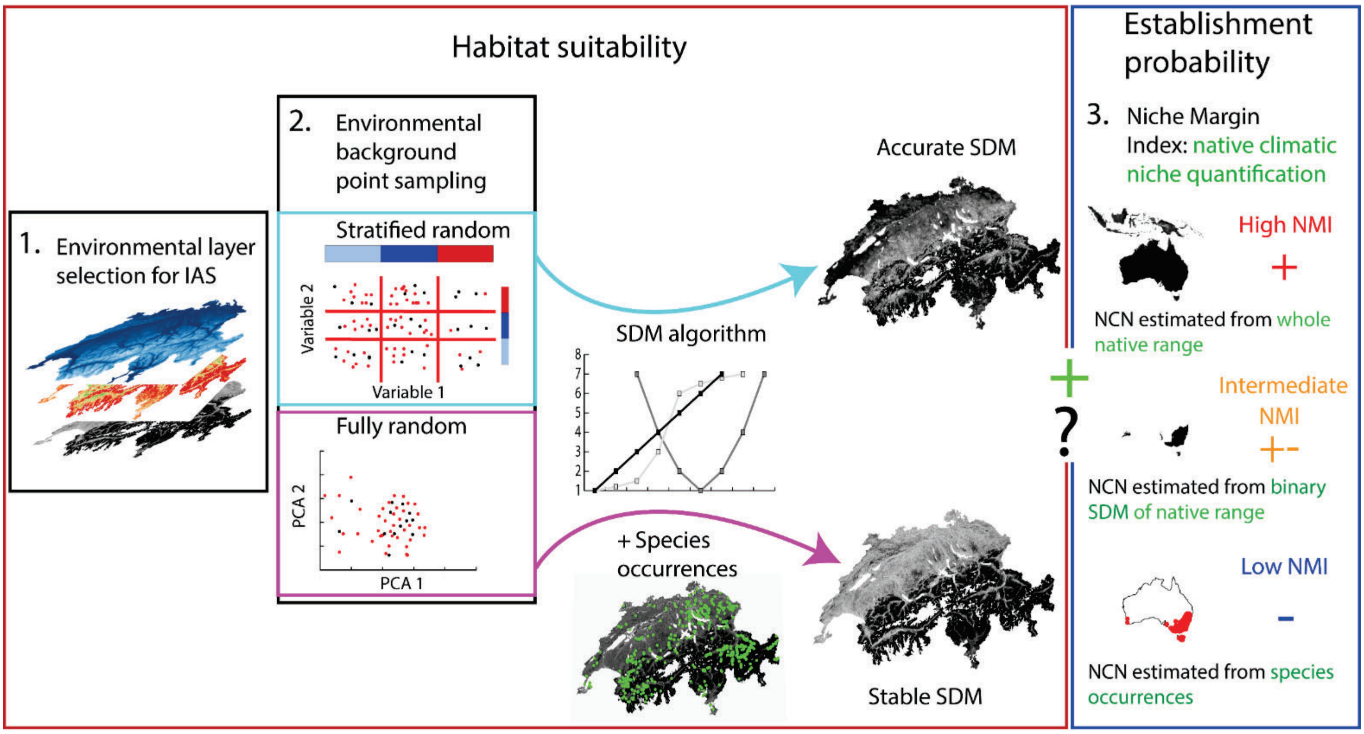


Figure 1: Summary of main findings of the thesis. The study area is presented as Switzerland because it is the scale of the final outputs of chapter 3 and because management purposes are most meaningful at small(er) scale. Point 1 refers to chapter 1, on covariates for non-native species. Note that the figure remarks on IAS instead of non-native species, because the latter may turn into the former without our knowledge and the principles are transferable. Point 2 illustrates the environmental background point selection strategies (chapter 2) and the figure explains the main findings on model accuracy and stability yielded from the respective methods. Finally, point 3 illustrates the Niche Margin Index results, which could potentially give a more complete view of plant establishment risk than SDMs, but due to the greatly different NMI value results (see text), the usefulness of this technique is yet ambiguous, as illustrated by the question mark.

Relevance to the management of IAS

Chapter 1 provides empirical evidence that disturbed habitats and areas close to infrastructure and roads should be surveyed for IAS. If a species displays invasive behavior in other areas, the probability that it will do so in its newly invaded range is high (Lockwood et al. 2001). Therefore, drawing on catalogues and knowledge synthesis on IAS, such as the ones done by the European Alien Species Information Network (EASIN, n.d.), to inform management decisions is prudent.

Chapter 2 points out some considerations regarding the building of SDMs for IAS. Whilst sampling background points fully randomly in environmental space yielded both higher stability and higher accuracy for IAS than for species at equilibrium, this sampling method is likely most suitable for running SDMs for IAS that are beginning to spread and/or are displaying generalist behaviors, as the background points in those cases would be the most abundant in the areas the species can occupy (Bazzichetto et al. 2023). Conversely, sampling stratified randomly in environmental space might be the best when researching if an IAS will spread to areas of extreme environments, such as areas of high altitude and/or latitude.

Chapter 3 points out possible pitfalls when applying the new findings to IASs, touching particularly on whether model overfitting can be associated with sampling the background points in environmental space. This is useful, as SDMs are increasingly advised as methods in conservation planning studies (Araujo et al. 2019, Barbet-Massin et al. 2012, Guisan et al. 2013, Low et al. 2021, Tulloch et al. 2016).

Chapter 4 paves the way for the use of NMI (Broennimann et al. 2021; see above) as a measure for probability of invasion success. It is a method that is relatively easy to use and that can be used on many species in a short time, once the necessary considerations have been accurately determined and researched.

Future studies and perspectives

It would be interesting to put the findings of chapter 1 in the context of the COUE (centroid shift, overlap, unfilling and expansion) framework (Guisan et al. 2014) which provides a common structure for analyzing niche overlaps. Doing this for all our congeneric species pairs (or others) would allow analysis of niche divergence between native and non-native species. Overall trends of niche divergence could indicate coexistence, whilst persistent overlap or congruence might imply continued competition. In addition, the subdivision of the target species by time for residence in Switzerland, growth form, etc. in chapter 1 likely often yielded non-significant results because of a too small sample size. The methods could be repeated with more pairs of congenerics. Particularly, an open question is why the distributions of those non-native species in chapter 1 that are invasive are more powerfully predicted by edaphic and vegetation-based factors? A logical explanation is their residence time, meaning they were given more time to spread, and occupy the available niches in the environment (Carboni et al. 2024, Menuz et al. 2014). This could mean that they have had more time to reach environmental equilibrium. Alternatively, they may have had time to shift and expand their niches, whilst not (yet) reaching any equilibrium (Menuz et al. 2014). This question of equilibrium links to uncertainties also remaining in chapter 2. It is possible indeed that the IAS modelled in that study have been spreading across the world for a long time and thus have reached their own equilibrium (Václavík & Meentemeyer 2012, Foster et al. 2020, Steen et al. 2024b). This would explain the small (but still significant) differences between IAS and equilibrium species for all the background point selection strategies. Conversely, whilst there is good reason to believe that the mountain species I modelled are indeed at equilibrium, this also cannot be verified. It can only be verified for those species for which the fundamental niche would be fully known, which is currently not the case – or even possible - for any real-world species (Matthiopoulos 2022, Takola & Schielzeth, 2022).

A possible solution therefore is to apply the methodology to virtual species. These are the only species for which the environmental niche (fundamental and/or realized) can be fully known (Qiao et al. 2015). There are R packages available that allow for the creation of such virtual species (e.g., virtualSpecies (Leroy et al. 2016), sdmvspecies (Duan et al. 2015)). Repeating the methods of chapter 2 for virtual species might add to our understanding of background point selection strategies (as e.g. done for other issues; e.g. Chevalier et al. 2021, Hirzel & Guisan 2002, Thibaud et al. 2014).

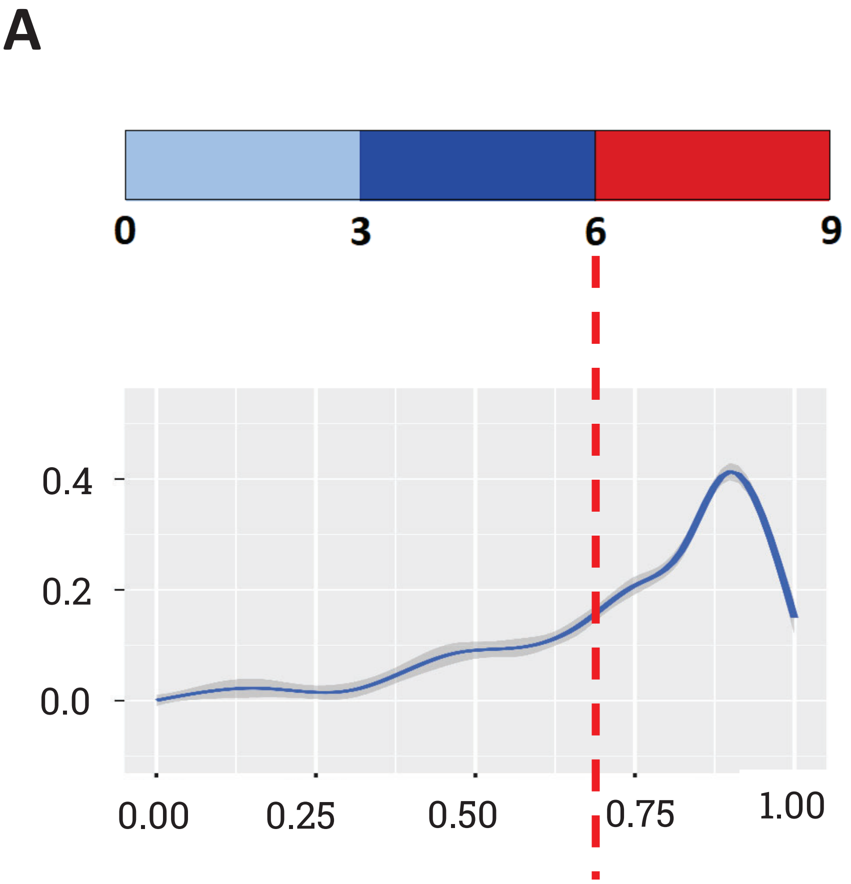
However, previous studies have made advances in analyzing the state of (dis)equilibrium of real-world IAS with the environment, as well. Uncertainty regarding state of equilibrium undermines confidence in identification of key areas for IAS management (Araújo et al. 2005, Pili et al. 2020). Gallien et al. (2012) have used a hierarchical approach similar to the one in chapter 3 of this thesis, in which they compared the degree which 27 IAS filled the outputs of a global and regional SDM. If both were filled to a high degree, the likelihood of further spread was small, but if one or both niches were filled to a lower degree, the probability of spreading was higher. The possible invaded areas could also be constrained by the global and regional SDMs. Again, the COUE framework could provide a unifying structure of analysis here.

In addition, background points have been used before to improve SDMs for IAS purposes (e.g., by assigning weights to background points in the regional range according to the SDM output in the global range (Gallien et al. 2012). In addition, our findings present a way to link SDMs and ecological theory together, which could be seen as a way to insert causality into SDMs, which are purely correlative. Different ways have been designed to do this for IAS (e.g., "hybrid models"; Elith et al. 2010, Gallien et al. 2010).

Appendix 3 of chapter 2 also contains an interesting seed for future research. One IAS (*Heracleum mantegazzianum*) and one equilibrium species (*Aster alpinus*) were selected from the roster of species used in the main study and their occurrence points were thinned by distance to artificial surface and urban area. *A. alpinus* shows a decrease in stability and AUC for the environmental sampling methods, but a vast increase in both dimensions for the classical fully random in geographic space sampling. Since *A. alpinus* is already a mountain species, the thinning might bring its distribution closer to geographically random whilst truncating its true environmental niche. This might explain the result. The differences in AUC and stability in *H. mantegazzianum* models are marked, as this treatment eliminates the (possible) effect of it being spread through human movement, as is typical for invasive plants (Arianoutsou et al. 2021, Nunes et al. 2015, Pyšek et al. 2020, Tsiamis et al. 2017), which are more present at low elevations. Hence, if the low elevation occurrence points get thinned, the species moves closer to the mountain ranges, where the SDMs may be able to discern correlations between the occurrence points and climate more easily (Steen et al. 2024b).

These findings should be verified with more species, but this thread of research is already further explored in chapter 3, where habitat suitability for *H. mantegazzianum* across Switzerland is modelled using 3 different sets of occurrence data: non-thinned, thinned by MCP and thinned by proximity to infrastructure. The last dataset shows the lowest habitat suitability in both present and future. This shows that thinning occurrence data in different ways can truncate the niche of a species (Aiello-Lammens et al. 2015), as can the omission of certain areas in the species range (as explored in Adde et al. 2023, Chevalier et al. 2021, Goicolea et al. 2024. Shackleton et al. 2020).

Chapter 3 is also by our knowledge the first scientific work that explores the use of environmentally stratified background points (Steen et al. 2024b) in combination with the spatially nested (i.e., the hierarchical) approach (Adde et al. 2023, Chevalier et al. 2021, Goicolea et al. 2024. Shackleton et al. 2020). Signs of overfitting have been found. There was one model fitted with Random Forest algorithms European scale that was overfitted. Random Forests are well known to pose a risk of overfitting data (Luan et al. 2020). The background point selection may have contributed to this. It is for instance possible that the most suitable habitat of the species is concentrated in one of the strata of a strong environmental predictor; if many background points are sampled in the environmental space of high habitat suitability, the model may be able to draw a very rigid line between suitable and unsuitable habitat. See Figure 2 for a visual presentation of this.


A
0
3
6
9
0.4
0.2
0.0
0.00
0.25
0.50
0.75
1.00

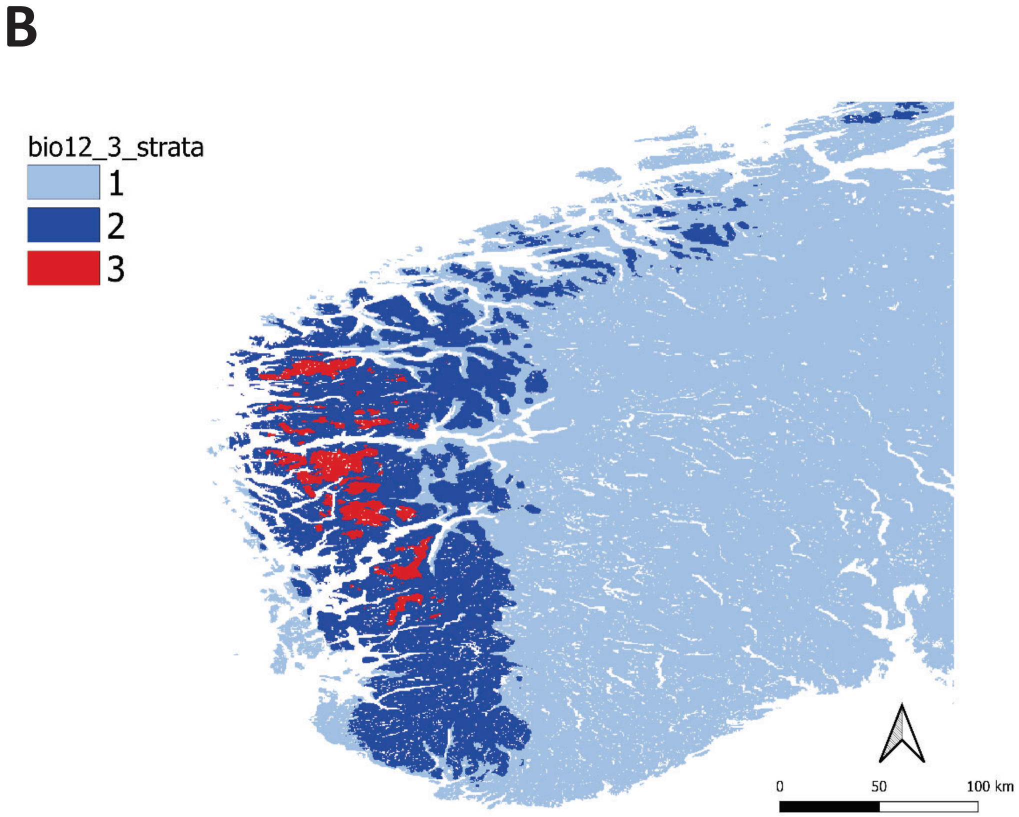

B
bio12_3_strata
1
2
3
0
50
100 km

Figure 2: graphical representation of a hypothetical case where the border between two strata of a powerful predictor of a fitted SDM corresponds to a drastic increase in habitat suitability for the species. Panel A presents a comparison between the strata and a response curve corresponding to the variable, i.e., the habitat suitability plotted as a function of variable value. Panel B gives a geographic representation of the three strata expressed in A.

In such a case, the distinction between suitable and unsuitable habitat drawn is potentially too rigid; the models may be too accurate and overfitted to the data, which poses problems, e.g., making the results of study non-transferable to other areas (Helmstetter et al. 2021), or to the future (Goicolea et al. 2024, Luan et al. 2024, Randin et al. 2020), and excessive conservatism in even the projections close to the occurrence data (Valavi et al. 2022). This theory may be supported by the fact that the Random Forest models in chapter 3 show a greater loss of habitat suitability between the present and 2050 projections than the Maxent ones. More research with more species is needed to investigate this, again possibly by using virtual species. The work on the Niche Margin Index (NMI; Broennimann et al. 2021) in chapter 4 meditates on the method of quantifying the native niche of a species. This may differ greatly per species group, e.g., between mammals and tracheophytes, as the former are mobile, and the latter are not. The degree of mobility is, however, only one axis of distinction. Others might be more or less sensitivity to climate (e.g., amphibians, reptiles and birds; García-Rodriquez et al. 2023). Repeating the methodology of chapter 4 for these species groups could lead to new insights into the optimal method for native niche quantification and into usefulness of the NMI method. Further research with more species groups (as e.g. in García-Rodriquez et al. 2023) is required to test this.

Finally, in addition to the probability of invasion success once an IAS arrives, the probability of it spreading to a certain area in the first place can also be added to SDM output (e.g. through dynamic dispersal modelling; Snell 2014). Whilst this thesis did not directly deal with this dimension, our results may nevertheless indicate that proxies for the probability of IAS spreading are proximity to infrastructure and disturbance of habitats, though bioclimate was a more powerful predictor (Steen et al. 2024a). This interpretation is supported by the fact that the risk of invasion is increased in disturbed areas (El-Barougy et al. 2021, Rai 2022, Steen et al. 2019, Thiele et al. 2007) and often spread via human movement (Arianoutsou et al. 2021, Nunes et al. 2015, Pyšek et al. 2020, Tsiamis et al. 2017). There is abundant scientific literature on additional pathways of invasion (Essl et al. 2015, Roy et al. 2017, Pyšek et al. 2020) and likewise efforts have been made to model dispersal and establishment probability of IAS through e.g. propagule pressure simulations, land use types, etc. (e.g. Thomas & Moloney 2014, Froese et al. 2019). This thesis provides results that are complementary with those threads of research but could be further researched and through additional computational methods, of which there are many (Thompson et al. 2021), each based on different assumptions and philosophies. For instance, circuit theory (Etherington 2011) can assess the potential spread of IAS through analyzing landscape connectivity, which can in turn be used in conservation management plans to satisfy different objectives (Weerasena et al. 2022). In addition, Reaction–advection–diffusion models (Baker 2017) and integrodifference equations (Gharouni et al. 2015) model the stream of propagules from high density to low density in a time-explicit manner in order to assess IAS spread and population growth. These types of models do not incorporate dispersal by humans, but this can be accomplished by using Gravity Models that can incorporate dispersal over geographic cells that are not spatially adjacent (though they do not include population dynamics; Thompson et al. (2021)). Non-spatially explicit methods also exist, such as nonspatial discrete-time population growth models (Shyu et al. 2013). This is by no means an exhaustive overview, though it must be noted that models centering on population growth dynamics require additional accurate data on population density. Still, areas of high suitability for IAS, as identified by SDMs, could be priority areas for conducting IAS population presence and density scans.

If all the above methods could be applied, it could vastly improve the future predictions of IAS. We would have SDMs, optimized by appropriate background point selection, aided by mechanistic models built using ecology specific to the IAS, plus an estimation of the likelihood of expansion of individual IAS (by comparing the global to the regional models as per Gallien et al. 2012), thereby increasing confidence in determination of ideal protected areas, and finally, adding the likelihood of spreading and probability of establishment. All these are steppingstones on the way of optimizing the forecasting of future biological invasions.

Technical limitations

Many of the studies in this thesis would have been impossible without the use of the High-Performance Computing (HPC) clusters of the University of Lausanne (https://wiki.unil.ch/ci/books/high-performance-computing-hpc). Often, several hundred GB of RAM memory were needed, for instance to run the PCA analyses on 109+ million cells of the CHELSA datasets at world scale, even after cropping them by the *H. mantegazzianum* biomes. Not all studies could have access to such high computing power.

Even with the HPC clusters, it was impossible to create sets of fully random in environmental space background points for chapter 3. This was because the generation of random combinations of points at world scale took too long. The reason for this is as follows: In our methodology, random sampling in environmental space happens by sampling along the axes of the Principal Component Analysis (PCA) dataset created from the environmental variables. For a new random point to be sampled, it had to satisfy three conditions: 1) match a combination of factors present in the datasets, 2) be random across all PCA axes and 3) have unique values on all PCA axes. The primary and secondary axes of the environmental PCA at world scale did not explain enough of the variation, so the tertiary axis had to be used, as well. Consequently, kernel density estimations had to be performed in three dimensions, which meant that each new background point had to be sampled randomly across three axes instead of two, reducing the chances of getting a match considerably. The rate of sampling of new combinations reached a plateau long before 10000 background points were generated.

Conclusion

The work presented in this thesis offers novel insights into invasion science and the use of Species Distribution Models (SDMs) for conservation efforts. Firstly, it was found that the climatic envelope is the strongest determinant of both native and non-native species distributions in Switzerland, but the former are more strongly associated with more pristine habitats and the latter are more associated with anthropogenic variables. The implications for conservation are discussed. Secondly, this thesis lays out that background point selection can be tailored to the species and purposes of an SDM study. Specifically, sampling fully randomly in environmental space is optimal for generalist species with a large range and sampling stratified randomly in environmental space is optimal for species that have narrow ranges, and/or are at equilibrium with the environment and/or have a low number of occurrence records. These findings were applied to building an optimal SDM for *Heracleum mantegazzianum*. Even when using these new methods, problems of overfitting have been encountered in this work and are cautioned against. Finally, the possibility of assessing establishment probability of IAS is explored using the Niche Margin Index (NMI). This technique quantifies similarity between native climatic niche and the climatic conditions in the invaded range, which is an indication for invasion success in mammals. I show that different ways of quantifying the native range yield very different results, and that caution is therefore advised.

All these findings are new methods and clues that aid the science of predicting IAS distributions.

References

Adde, A., Rey, P.-L., Brun, P., Külling, N., Fopp, F., Altermatt, F., Broennimann, O., Lehmann, A., Petitpierre, B., Zimmermann, N. E., Pellissier, L., & Guisan, A. (2023). N-SDM: A high-performance computing pipeline for Nested Species Distribution Modelling. Ecography, 2023(6), e06540. https://doi.org/10.1111/ecog.06540

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography, 38(5), 541–545. https://doi.org/10.1111/ecog.01132

Araujo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. Science Advances, 5(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Araújo, 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. https://doi.org/10.1111/j.1365-2486.2005.01000.x

Arianoutsou, M., Bazos, I., Christopoulou, A., Kokkoris, Y., Zikos, A., Zervou, S., Delipetrou, P., Cardoso, A. C., Deriu, I., Gervasini, E., & Tsiamis, K. (2021). Alien plants of Europe: Introduction pathways, gateways and time trends. PeerJ, 9, e11270. https://doi.org/10.7717/peerj.11270

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. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Baker, C. M. (2017). Target the Source: Optimal Spatiotemporal Resource Allocation for Invasive Species Control. Conservation Letters, 10(1), 41–48. https://doi.org/10.1111/conl.12236

Bazzichetto, M., Lenoir, J., Da Re, D., Tordoni, E., Rocchini, D., Malavasi, M., Bart´ak, V., Sperandii, M.G., 2023. Sampling strategy matters to accurately estimate response curves' parameters in species distribution models. Glob. Ecol. Biogeogr. 32 (10), 1717–1729. https://doi.org/10.1111/geb.13725.

Broennimann, O., Petitpierre, B., Chevalier, M., González-Suárez, M., Jeschke, J. M., Rolland, J., Gray, S. M., Bacher, S., & Guisan, A. (2021). Distance to native climatic niche margins explains establishment success of alien mammals. Nature Communications, 12(1), Article 1. https://doi.org/10.1038/s41467-021-22693-0

Carboni, L. J., Biganzoli, F., & Cipriotti, P. A. (2024). Niche shifts drive the global distribution of a widespread plant invader of rangelands. Biological Invasions, 27(1), 25. https://doi.org/10.1007/s10530-024-03496-y

Chevalier, M., Broennimann, O., Cornuault, J. & Guisan, A. (2021) Data integration methods to account for spatial niche truncation effects in regional projections of species distribution. Ecological Applications, 31.

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(1), 108–110. https://doi.org/10.1111/ecog.01080

EASIN - European Alien Species Information Network. (n.d.). Retrieved June 14, 2023, from https://easin.jrc.ec.europa.eu/easin/EASINNutshell

El-Barougy, R. F., Dakhil, M. A., Halmy, M. W., Gray, S. M., Abdelaal, M., Khedr, A.-H. A., & Bersier, L.-F. (2021). Invasion risk assessment using trait-environment and species distribution modelling techniques in an arid protected area: Towards conservation prioritization. Ecological Indicators, 129, 107951. https://doi.org/10.1016/j.ecolind.2021.107951

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. Methods in Ecology and Evolution, 1(4), 330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Essl, F., Bacher, S., Blackburn, T. M., Booy, O., Brundu, G., Brunel, S., Cardoso, A.-C., Eschen, R., Gallardo, B., Galil, B., García-Berthou, E., Genovesi, P., Groom, Q., Harrower, C., Hulme, P. E., Katsanevakis, S., Kenis, M., Kühn, I., Kumschick, S., … Jeschke, J. M. (2015). Crossing Frontiers in Tackling Pathways of Biological Invasions. BioScience, 65(8), 769–782. https://doi.org/10.1093/biosci/biv082

Foster, S. L., Kharouba, H. M., & Smith, T. W. (2022). Testing the assumption of environmental equilibrium in an invasive plant species over a 130 year history. Ecography, 2022(12), e12933. https://doi.org/10.1111/ecog.06284

Froese, J. G., Pearse, A. R., & Hamilton, G. (2019). Rapid spatial risk modelling for management of early weed invasions: Balancing ecological complexity and operational needs. Methods in Ecology and Evolution, 10(12), 2105–2117. https://doi.org/10.1111/2041-210X.13284

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(11), 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x

Gallien, L., Münkemüller, T., Albert, C. H., Boulangeat, I., & Thuiller, W. (2010). Predicting potential distributions of invasive species: Where to go from here? Diversity and Distributions, 16(3), 331–342. https://doi.org/10.1111/j.1472-4642.2010.00652.x

García-Rodríguez, A., Lenzner, B., Marino, C., Liu, C., Velasco, J. A., Bellard, C., Jeschke, J. M., Seebens, H., & Essl, F. (2023). Patterns and drivers of climatic niche dynamics during biological invasions of island-endemic amphibians, reptiles, and birds. Global Change Biology, 29(17), 4924–4938. https://doi.org/10.1111/gcb.16849

Gharouni, A., Barbeau, M. A., Locke, A., Wang, L., & Watmough, J. (2015). Sensitivity of invasion speed to dispersal and demography: An application of spreading speed theory to the green crab invasion on the northwest Atlantic coast. Marine Ecology Progress Series, 541, 135–150. https://doi.org/10.3354/meps11508

Goicolea, T., Adde, A., Broennimann, O., García-Viñas, J. I., Gastón, A., José Aroca-Fernández, M., Guisan, A., & G. Mateo, R. (n.d.). Spatially-nested hierarchical species distribution models to overcome niche truncation in national-scale studies. Ecography, n/a(n/a), e07328. https://doi.org/10.1111/ecog.07328

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: Insights from biological invasions. Trends in Ecology & Evolution, 29(5), 260–269. https://doi.org/10.1016/j.tree.2014.02.009

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., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. Ecology Letters, 16(12), 1424–1435. https:// doi. org/ 10. 1111/ Ele. 12189

Helmstetter, N. A., Conway, C. J., Stevens, B. S., & Goldberg, A. R. (2021). Balancing transferability and complexity of species distribution models for rare species conservation. Diversity and Distributions, 27(1), 95–108. https://doi.org/10.1111/ddi.13174

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

Leroy, B., Meynard, C. N., Bellard, C., & Courchamp, F. (2016). Virtualspecies, an R package to generate virtual species distributions. Ecography, 39(6), 599–607. https://doi.org/10.1111/ecog.01388

Lockwood, J. L., Simberloff, D., McKinney, M. L., & Von Holle, B. (2001). How Many, And Which, Plants Will Invade Natural Areas? Biological Invasions, 3(1), 1–8. https://doi.org/10.1023/A:1011412820174

Low, B. W., Zeng, Y., Tan, H. H., & Yeo, D. C. J. (2021). Predictor complexity and feature selection affect Maxent model transferability: Evidence from global freshwater invasive species. Diversity and Distributions, 27(3), 497–511. https:// doi. org/ 10. 1111/ DDI. 13211

Luan, J., Xu, B., Ji, Y., Zhang, C., Xue, Y., & Ren, Y. (2024). Improving the spatial transferability of species distribution models to inform biological conservation of two piscivore fish species. Biodiversity and Conservation, 33(14), 4215–4235. https://doi.org/10.1007/s10531-024-02947-1

Luan, J., Zhang, C., Xu, B., Xue, Y., & Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. Fisheries Research, 227, 105534. https://doi.org/10.1016/j.fishres.2020.105534

Matthiopoulos, J. (2022). Defining, estimating, and understanding the fundamental niches of complex animals in heterogeneous environments. Ecological Monographs, 92(4), 1–28.

Menuz, D. R., Kettenring, K. M., Hawkins, C. P., & Cutler, D. R. (2015). Non-equilibrium in plant distribution models – only an issue for introduced or dispersal limited species? Ecography, 38(3), 231–240. https://doi.org/10.1111/ecog.00928

Nunes, A.L., Tricarico, E., Panov, V.E., Cardoso, A.C., Katsanevakis, S. (2015). Pathways and gateways of freshwater invasions in Europe. Aquatic Invasions 10 (4), 359-370. http://dx.doi.org/10.3391/ai.2015.10.4.01

Pili, A. N., Tingley, R., Sy, E. Y., Diesmos, M. L. L., Diesmos, A. C. (2020). Niche shifts and environmental non-equilibrium undermine the usefulness of ecological niche models for invasion risk assessments. Scientific Reports, 10(1), 7972. https://doi.org/10.1038/s41598-020-64568-2

Pyšek, P., Hulme, P. E., Simberloff, D., Bacher, S., Blackburn, T. M., Carlton, J. T., Dawson, W., Essl, F., Foxcroft, L. C., Genovesi, P., Jeschke, J. M., Kühn, I., Liebhold, A. M., Mandrak, N. E., Meyerson, L. A., Pauchard, A., Pergl, J., Roy, H. E., Seebens, H., … Richardson, D. M. (2020). Scientists' warning on invasive alien species. Biological Reviews, 95(6), 1511–1534. https://doi.org/10.1111/brv.12627

Qiao, H., Peterson, A. T., Campbell, L. P., Soberón, J., Ji, L., & Escobar, L. E. (2016). NicheA: Creating virtual species and ecological niches in multivariate environmental scenarios. Ecography, 39(8), 805–813. https://doi.org/10.1111/ecog.01961

Rai, P. K. (2022). Environmental Degradation by Invasive Alien Plants in the Anthropocene: Challenges and Prospects for Sustainable Restoration. Anthropocene Science, 1(1), 5–28. https://doi.org/10.1007/s44177-021-00004-y

Randin, C. F., Ashcroft, M. B., Bolliger, J., Cavender-Bares, J., Coops, N. C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., … Payne, D. (2020). Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. Remote Sensing of Environment, 239, 111626. https://doi.org/10.1016/j.rse.2019.111626

Roy, H. E., Hesketh, H., Purse, B. V., Eilenberg, J., Santini, A., Scalera, R., Stentiford, G. D., Adriaens, T., Bacela-Spychalska, K., Bass, D., Beckmann, K. M., Bessell, P., Bojko, J., Booy, O., Cardoso, A. C., Essl, F., Groom, Q., Harrower, C., Kleespies, R., … Dunn, A. M. (2017). Alien Pathogens on the Horizon: Opportunities for Predicting their Threat to Wildlife. Conservation Letters, 10(4), 477–484. https://doi.org/10.1111/conl.12297

Shackleton, R. T., Petitpierre, B., Pajkovic, M., Dessimoz, F., Brönnimann, O., Cattin, L., Čejková, Š., Kull, C. A., Pergl, J., Pyšek, P., Yoccoz, N., & Guisan, A. (2020). Integrated Methods for Monitoring the Invasive Potential and Management of Heracleum mantegazzianum (giant hogweed) in Switzerland. Environmental Management, 65(6), 829–842. https://doi.org/10.1007/s00267-020-01282-9

Snell, R. S. (2014). Simulating long-distance seed dispersal in a dynamic vegetation model. Global Ecology and Biogeography, 23(1), 89–98. https://doi.org/10.1111/geb.12106

a. Steen, B., Adde, A., Schlaepfer, M. A., Guisan, A., & Maiorano, L. (2024). Distributions of non-native and native plants are not determined by the same environmental factors. Ecological Solutions and Evidence, 5(4), e12374. https://doi.org/10.1002/2688-8319.12374

b. Steen, B., Broennimann, O., Maiorano, L., & Guisan, A. (2024). How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels. Ecological Modelling, 493, 110754. https://doi.org/10.1016/j.ecolmodel.2024.110754

Shyu, E., Pardini, E. A., Knight, T. M., & Caswell, H. (2013). A seasonal, density-dependent model for the management of an invasive weed. Ecological Applications, 23(8), 1893–1905. https://doi.org/10.1890/12-1712.1

Steen, B., Cardoso, A. C., Tsiamis, K., Nieto, K., Engel, J., & Gervasini, E. (2019). Modelling hot spot areas for the invasive alien plant *Elodea nuttallii* in the EU. Management of Biological Invasions, 10(1), 151–170. https://doi.org/10.3391/mbi.2019.10.1.10

Takola, E., & Schielzeth, H. (2022). Hutchinson's ecological niche for individuals. Biology & Philosophy, 37(4), 25. https://doi.org/10.1007/s10539-022-09849-y

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.

Thiele, J., Otte, A., & Eckstein, R. L. (2007). Ecological needs, habitat preferences and plant communities invaded by Heracleum mantegazzianum. In Ecology and management of giant hogweed (Heracleum mantegazzianum) (pp. 126–143). https://doi.org/10.1079/9781845932060.0126

Thomas, S. M., & Moloney, K. A. (2015). Combining the effects of surrounding land-use and propagule pressure to predict the distribution of an invasive plant. Biological Invasions, 17(1), 477–495. https://doi.org/10.1007/s10530-014-0745-7

Thompson, B. K., Olden, J. D., & Converse, S. J. (2021). Mechanistic invasive species management models and their application in conservation. Conservation Science and Practice, 3(11), e533. https://doi.org/10.1111/csp2.533

Tsiamis, K., Cardoso, A. C., & Gervasini, E. (2017). The European Alien Species Information Network on the Convention on Biological Diversity pathways categorization. NeoBiota, 32, 21–29. https://doi.org/10.3897/neobiota.32.9429

Tulloch, A. I. T., Sutcliffe, P., Naujokaitis-Lewis, I., Tingley, R., Brotons, L., Ferraz, K., Possingham, H. P., 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. https://doi.org/10.1016/j.biocon.2016.04.023

Václavík, T., & Meentemeyer, R. K. (2012). Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. Diversity and Distributions, 18(1), 73–83. https://doi.org/10.1111/j.1472-4642.2011.00854.x

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. Ecological Monographs, 92(1), e01486. https://doi.org/10.1002/ecm.1486

Weerasena, L., Hunt, N., Bandara, D., & McKnight, M. (2022). Spatially explicit multi-objective mathematical model for invasive species management. Biological Invasions, 24(6), 1839–1862. https://doi.org/10.1007/s10530-022-02754-1

Appendix I

Supplementary Material of Chapter I

Supporting information

Publication title:

Distributions of non-native and native plants are not determined by the same environmental factors

Authors:

Bart Steen*, bart.steen@unil.ch, Institute of Earth Surface Dynamics, University of Lausanne, CH-1015 Lausanne Switzerland & Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, viale dell'Università 32, 00185 Roma, Italy

Antoine Adde, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133 CH-8600 Dübendorf

Martin A. Schlaepfer,** Institute for Environmental Sciences, Université de Genève, 66 Boulevard Carl-Vogt, CH-1211 Genève 5

Antoine Guisan****,** Institute of Earth Surface Dynamics, University of Lausanne, CH-1015 Lausanne Switzerland & Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne Switzerland

Luigi Maiorano****,** Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, Viale dell'Università 32, 00185 Roma, Italy

* Corresponding author

** Shared last authorship

Published in:

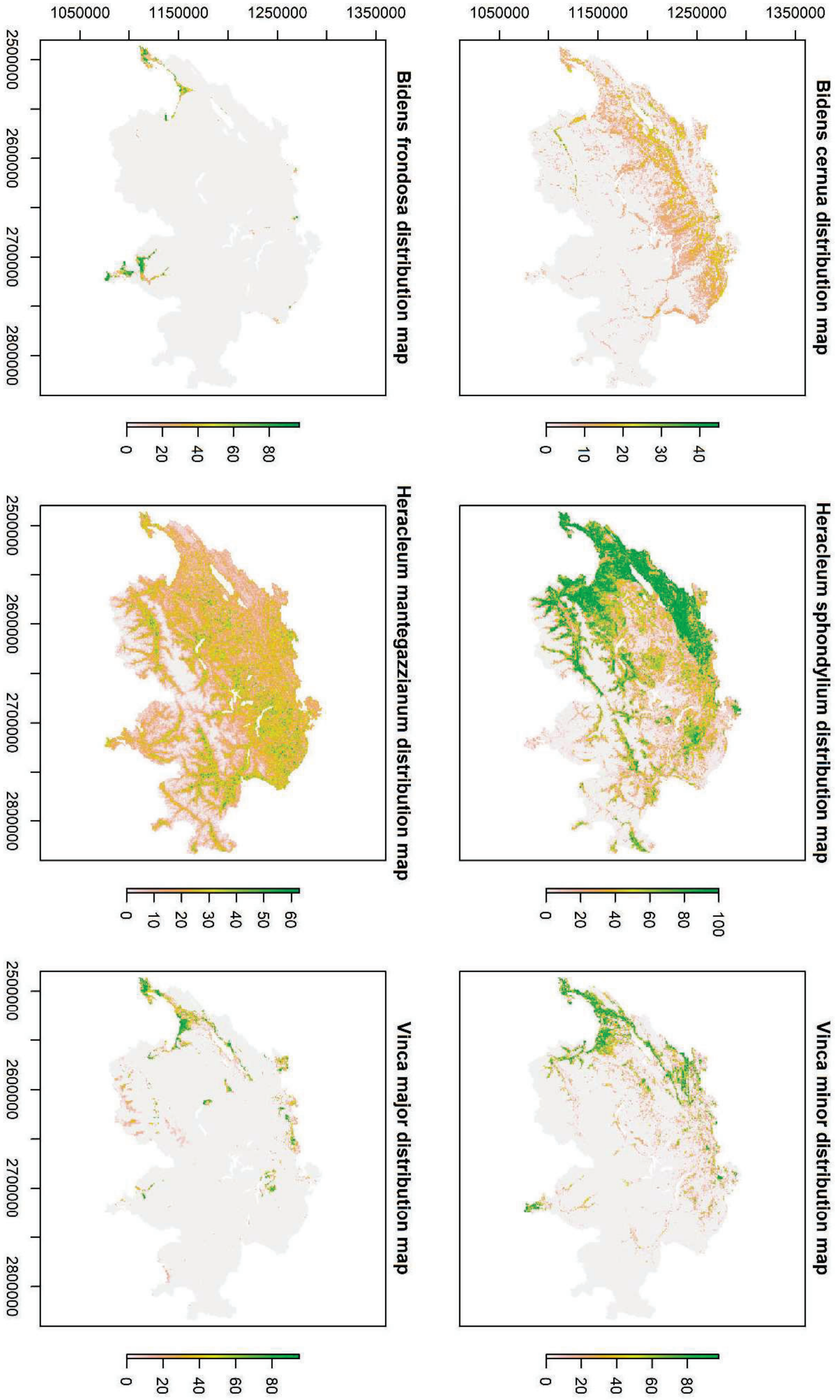

Bidens cernua distribution map
Heracleum sphondylium distribution map
Vinca minor distribution map
Bidens frondosa distribution map
Heracleum mantegazzianum distribution map
Vinca major distribution map
1050000
1150000
1250000
1350000
2500000
2600000
2700000
2800000
0
10
20
30
40
0
20
40
60
80
100
0
20
40
60
80
0
10
20
30
40
50
60

Figure S1*: Example output Species Distribution Model from a species pair, for the native - nonnative species pairs* Bidens cernua and Bidens frondosa*,* Heracleum sphondylium *and* Heracleum mantegazzianum *and finally* Vinca minor *and* Vinca major*. The scale bars refer to habitat suitability, i.e., likelihood of occurrence of the species given the environment and the projection is EPSG:2056. The distance measurements therefore are in meters.*

Table S1: *List of target species: non-native and native congeneric species pairs.*

Non-native	**Native**
Acer negundo	Acer platanoides
Alchemilla mollis	Alchemilla pentaphyllea
Amaranthus albus	Amaranthus blitum
Anemone blanda	Anemone nemorosa
Arabis alpina caucasica	Arabis nova
Artemisia annua	Artemisia vulgaris
Artemisia verlotiorum	Artemisia absinthium
Aster lanceolatus	Aster bellidiastrum
Atriplex micrantha	Atriplex prostrata
Avena sativa	Avena fatua
Berberis julianae	Berberis vulgaris
Bidens frondosa	Bidens cernua
Bromus carinatus	Bromus ramosus
Bromus catharticus	Bromus benekenii
Bromus inermis	Bromus arvensis
Bromus madritensis	Bromus japonicus
Bromus sitchensis	Bromus tectorum
Calendula officinalis	Calendula arvensis
Calystegia silvatica	Calystegia sepium
Camelina sativa	Camelina microcarpa
Campanula poscharskyana	Campanula cervicaria
Campylopus introflexus	Campylopus flexuosus
Cardamine occulta	Cardamine impatiens
Cardaminopsis arenosa arenosa	Cardaminopsis halleri
Cerastium tomentosum	Cerastium glutinosum
Chenopodium giganteum	Chenopodium album
Chenopodium pumilio	Chenopodium glaucum
Chenopodium strictum	Chenopodium botrys
Cicerbita macrophylla	Cicerbita plumieri
Cornus sanguinea australis	Cornus mas
Cornus sericea	Cornus sanguinea
Coronopus didymus	Coronopus squamatus
Corylus maxima	Corylus avellana
Cotoneaster bullatus	Cotoneaster integerrimus
Cotoneaster horizontalis	Cotoneaster tomentosus
Crepis pulchra	Crepis vesicaria taraxacifolia
Crepis setosa	Crepis mollis
Crocus tommasinianus	Crocus albiflorus
Cyclamen coum	Cyclamen hederifolium
Cyperus eragrostis	Cyperus fuscus
Digitalis purpurea	Digitalis lutea
Epilobium ciliatum	Epilobium tetragonum tetragonum

Eragrostis virescens	Eragrostis pilosa
Erigeron annuus annuus	Erigeron atticus
Erigeron annuus	Erigeron alpinus
Erysimum cheiri	Erysimum cheiranthoides
Euphorbia humifusa	Euphorbia dulcis
Euphorbia maculata	Euphorbia cyparissias
Euphorbia myrsinites	Euphorbia amygdaloides
Euphorbia nutans	Euphorbia seguieriana seguieriana
Euphorbia virgata	Euphorbia falcata
Fallopia aubertii	Fallopia dumetorum
Geranium sibiricum	Geranium dissectum
Glyceria striata	Glyceria maxima
Helleborus orientalis	Helleborus foetidus
Heracleum mantegazzianum	Heracleum sphondylium
Hypericum calycinum	Hypericum pulchrum
Iris foetidissima	Iris pseudacorus
Juncus tenuis	Juncus bufonius
Lamium galeobdolon argentatum	Lamium amplexicaule
Lemna minuta	Lemna minor
Lepidium densiflorum	Lepidium campestre
Lepidium virginicum	Lepidium graminifolium
Leucanthemum praecox	Leucanthemum adustum
Linaria genistifolia dalmatica	Linaria alpina alpina
Linaria genistifolia	Linaria repens
Linum usitatissimum	Linum alpinum
Lolium multiflorum	Lolium perenne
Lonicera caprifolium	Lonicera caerulea
Lonicera henryi	Lonicera alpigena
Lunaria annua	Lunaria rediviva
Lysimachia punctata	Lysimachia nummularia
Malus pumila	Malus sylvestris
Matricaria discoidea	Matricaria chamomilla
Medicago arabica	Medicago lupulina
Mentha spicata	Mentha suaveolens
Muscari armeniacum	Muscari comosum
Nepeta racemosa	Nepeta cataria
Nigella damascena	Nigella arvensis
Onobrychis viciifolia	Onobrychis montana
Oxalis dillenii	Oxalis corniculata
Oxalis stricta	Oxalis acetosella
Papaver croceum	Papaver rhoeas
Papaver somniferum	Papaver dubium lecoqii
Phalaris canariensis	Phalaris arundinacea
Picris echioides	Picris hieracioides hieracioides
Pimpinella peregrina	Pimpinella saxifraga

Pinus nigra	Pinus mugo uncinata
Pinus strobus	Pinus sylvestris
Plantago coronopus	Plantago atrata atrata
Polygonum polystachyum	Polygonum persicaria
Potentilla fruticosa	Potentilla pusilla
Potentilla norvegica	Potentilla erecta
Prunus cerasifera	Prunus avium
Prunus domestica domestica	Prunus padus
Prunus domestica	Prunus mahaleb
Prunus persica	Prunus spinosa
Pyrus communis	Pyrus pyraster
Quercus ilex	Quercus pubescens
Quercus rubra	Quercus petraea
Ribes aureum	Ribes petraeum
Ribes rubrum	Ribes alpinum
Rosa multiflora	Rosa spinosissima
Rosa rugosa	Rosa chavinii
Rubus armeniacus	Rubus ulmifolius
Rubus laciniatus	Rubus caesius
Rubus phoenicolasius	Rubus bifrons
Rumex longifolius	Rumex pulcher
Rumex patientia	Rumex conglomeratus
Sagittaria latifolia	Sagittaria sagittifolia
Salvia officinalis	Salvia pratensis pratensis
Salvia verticillata	Salvia pratensis
Scilla forbesii	Scilla bifolia
Scutellaria altissima	Scutellaria alpina
Sedum hybridum	Sedum telephium fabaria
Sedum spurium	Sedum rubens
Senecio inaequidens	Senecio rupestris
Setaria italica	Setaria verticillata
Sisymbrium altissimum	Sisymbrium strictissimum
Solanum chenopodioides	Solanum dulcamara
Solanum tuberosum	Solanum nigrum nigrum
Solidago canadensis	Solidago virgaurea minuta
Solidago gigantea	Solidago virgaurea
Stachys byzantina	Stachys pradica
Symphytum asperum	Symphytum tuberosum
Tanacetum parthenium	Tanacetum vulgare
Thlaspi alliaceum	Thlaspi caerulescens
Thymus vulgaris	Thymus longicaulis
Trifolium alexandrinum	Trifolium patens
Trifolium resupinatum	Trifolium ochroleucon
Verbena bonariensis	Verbena officinalis
Veronica filiformis	Veronica arvensis

Veronica peregrina	Veronica alpina
Veronica persica	Veronica dillenii
Viburnum rhytidophyllum	Viburnum opulus
Vicia faba	Vicia sylvatica
Vicia lutea	Vicia tetrasperma
Vicia pannonica	Vicia cracca cracca
Vicia villosa varia	Vicia cracca
Vicia villosa	Vicia sepium
Vinca major	Vinca minor

Table S2*: Full list of environmental variables (Külling et al. 2024) used to fit models in the N-SDM pipeline.*

Class	**Dataset**	**Variable**	**Metric**
bioclim	chclim25	annual mean temperature	raw value (°C)
bioclim	chclim25	mean diurnal range (mean of monthly (max temp - min temp))	raw value (°C)
bioclim	chclim25	temperature seasonality (standard deviation *100)	raw value (°C)
bioclim	chclim25	min temperature of coldest month	raw value (°C)
bioclim	chclim25	mean temperature of wettest quarter	raw value (°C)
bioclim	chclim25	mean temperature of warmest quarter	raw value (°C)
bioclim	chclim25	annual precipitation	raw value (mm)
bioclim	chclim25	precipitation of driest month	raw value (mm)
bioclim	chclim25	precipitation of wettest quarter	raw value (mm)
bioclim	chclim25	precipitation of warmest quarter	raw value (mm)
bioclim	chclim25	isothermality (bio2/bio7) (* 100)	raw value (°C)
bioclim	chclim25	max temperature of warmest month	raw value (°C)
bioclim	chclim25	temperature annual range (bio5-bio6)	raw value (°C)

bioclim	chclim25	mean temperature of driest quarter	raw value (°C)
bioclim	chclim25	mean temperature of coldest quarter	raw value (°C)
bioclim	chclim25	precipitation of wettest month	raw value (mm)
bioclim	chclim25	precipitation seasonality (coefficient of variation)	coefficient of variation
bioclim	chclim25	precipitation of driest quarter	raw value (mm)
bioclim	chclim25	precipitation of coldest quarter	raw value (mm)
edaphic	eivdescombes	soil aeration	continuous index (low to high)
edaphic	eivdescombes	soil moisture	continuous index (dry to water)
edaphic	eivdescombes	soil humus	continuous index (humus poor to rich)
edaphic	eivdescombes	soil nutrients	continuous index (nutrient poor to rich)
edaphic	eivdescombes	soil ph	continuous index (acidic to alkaline)
edaphic	eivdescombes	soil moisture variability	continuous index (low to high moisture intraannual variability)
edaphic	modiffus3	nitrogen load	raw value (kg N/ha/year)

edaphic	modiffus3	phosphorus load	raw value (kg P/ha/year)
hydro	gwn07	distance to lake (by size class 1 to 3 + all)	raw value (distance in meters)
hydro	gwn07	distance to river (by strahler class 1 to 4 + all)	raw value (distance in meters)
hydro	bafu	river oekomorphology (class 1 to 5)	raw value (m^3/s)
hydro	swisstopo	slope (mean; max)	raw value (m^3/s)
hydro	bafu	annuals discharge (mean; max; var)	raw value (m^3/s)
hydro	zappa	baseflow index	continuous index (low to high)
hydro	zappa	groundwater recharge	raw value (mm)
hydro	zappa	runoff coefficient (total runoff/precipitation)	continuous index (low to high)
hydro	zappa	snowmelt portion (snowmelt/total runoff)	continuous index (low to high)
hydro	zappa	snow melt	continuous index (mm)
topo	alti3d2016	aspect (min; max; mean; med)	raw value (direction in degrees)
topo	alti3d2016	hill shade (min; max; mean; med)	continuous index (shades of grey)
topo	alti3d2016	slope (min; max; mean; med)	raw value (degrees)

popdensity	statpop13	human population density	raw value (inhabitants)
transport	sonbase	noise pollution index	continuous index (quiet to noisy)
transport	tlm3d	distance to road (by road class)	raw value (distance in meters)
transport	tlm3d	topo distance to road (by road class)	raw value (distance in meters)
vege	nfi	canopy cover (min, max, med)	percentage cover
vege	copernicus	proportion of coniferous forest	percentage cover
vege	copernicus	proportion of deciduous forest	percentage cover
lulc	geostat2	forest	percentage cover
lulc	geostat2	agriculture	percentage cover
lulc	geostat65	intensive orchards	percentage cover
lulc	geostat65	scattered fruit trees	percentage cover
lulc	geostat65	regular vineyards	percentage cover
lulc	geostat65	horticulture	percentage cover
lulc	geostat65	favourable arable land and meadows	percentage cover
lulc	geostat65	other arable land and meadows	percentage cover
lulc	geostat65	farm pastures	percentage cover
lulc	geostat65	brush meadows and farm pastures	percentage cover
lulc	geostat65	mountain meadows	percentage cover

lulc	geostat65	favourable alpine pastures	percentage cover
lulc	geostat65	brush alpine pastures	percentage cover
lulc	geostat65	rocky alpine pastures	percentage cover
lulc	geostat65	remote and steep alpine meadows and pastures	percentage cover
lulc	geostat65	normal dense forest	percentage cover
lulc	geostat65	forest stripes, edges	percentage cover
lulc	geostat65	cut forests	percentage cover
lulc	geostat65	forest fresh cuts	percentage cover
lulc	geostat65	devastated forests	percentage cover
lulc	geostat65	open forest (on agricultural areas)	percentage cover
lulc	geostat65	open forest (on unproductive areas)	percentage cover
lulc	geostat65	brush forest	percentage cover
lulc	geostat65	groves, hedges	percentage cover
lulc	geostat65	clusters of trees (on agricultural areas)	percentage cover
lulc	geostat65	trees on unproductive areas	percentage cover
lulc	geostat65	scrub vegetation	percentage cover
lulc	geostat65	unproductive grass and shrubs	percentage cover
lulc	geostat65	wetlands	percentage cover
lulc	geostat65	bare rock	percentage cover

lulc	geostat65	glacier	percentage cover
lulc	geostat8	closed forest	percentage cover
lulc	geostat8	open forest	percentage cover
lulc	geostat8	settlement_urban_amenities	percentage cover
lulc	geostat8	permanent crops	percentage cover
lulc	geostat8	intensive agriculture	percentage cover
lulc	geostat8	grassland_meadows	percentage cover
lulc	geostat8	alpine pastures	percentage cover
lulc	geostat8	hydrological	percentage cover

Table S3: *All land use/land cover variables and their associated naturality score.*

Category	**metric**	Naturality score
forest	percentage cover	5
agriculture	percentage cover	2
intensive orchards	percentage cover	2
scattered fruit trees	percentage cover	3
regular vineyards	percentage cover	2
horticulture	percentage cover	3
favourable arable land and meadows	percentage cover	3
other arable land and meadows	percentage cover	3
farm pastures	percentage cover	2
brush meadows and farm pastures	percentage cover	3
mountain meadows	percentage cover	4

favourable alpine pastures	percentage cover	4
brush alpine pastures	percentage cover	4
rocky alpine pastures	percentage cover	4
remote and steep alpine meadows and pastures	percentage cover	4
normal dense forest	percentage cover	5
forest stripes, edges	percentage cover	4
cut forests	percentage cover	3
forest fresh cuts	percentage cover	2
devastated forests	percentage cover	2
open forest (on agricultural areas)	percentage cover	3
open forest (on unproductive areas)	percentage cover	4
brush forest	percentage cover	4

groves, hedges	percentage cover	3
clusters of trees (on agricultural areas)	percentage cover	3
trees on unproductive areas	percentage cover	3
scrub vegetation	percentage cover	4
unproductive grass and shrubs	percentage cover	4
wetlands	percentage cover	5
bare rock	percentage cover	5
glacier	percentage cover	5
closed forest	percentage cover	5
open forest	percentage cover	5
settlement_urban_amenities	percentage cover	1
permanent crops	percentage cover	2

intensive agriculture	percentage cover	2
grassland_meadows	percentage cover	3
alpine pastures	percentage cover	4
hydrological	percentage cover	4

Table S4: *List of target species and subdivision by growth form, year of introduction ot Switzerland, and origin region.*

Species name	**Growth form**	**Raunkiær growth form**	**First year observed in CH**	**Less than 40 yrs ago?**	**Origin region**
Acer negundo	Tree	Phanerophyte	1898	No	North America
Alchemilla mollis	Herb	Herbal chamaephyte	1950	No	Asia
Amaranthus albus	Herb	Therophyte	1882	No	North America
Anemone blanda	Geophyte	Geophyte	Before 1930	No	Europe
Arabis alpina caucasica	Herb	Herbal chamaephyte	Before 1930	No	Europe
Artemisia annua	Herb	Therophyte	1892	No	Asia
Artemisia verlotiorum	Geophyte	Geophyte	1920	No	Asia
Aster lanceolatus	Geophyte	Geophyte	1882	No	North America
Atriplex micrantha	Herb	Therophyte	2010	Yes	Asia
Avena sativa	Grass	Therophyte	Before 1930	No	Asia
Berberis julianae	Shrub	Nanophanerophyte	1990	Yes	Asia
Bidens frondosa	Herb	Therophyte	1970	No	North America
Bromus carinatus	Grass/Geophyte	Geophyte	1995	Yes	North America
Bromus catharticus	Grass	Therophyte	1911	No	South America
Bromus inermis	Grass	Perennial Hemikryptophyte	1857	No	Asia
Bromus madritensis	Grass	Therophyte	1903	No	Africa
Bromus sitchensis	Grass	Perennial Hemikryptophyte	2000	Yes	North America
Calendula officinalis	Herb	Therophyte	Before 1930	No	Europe
Calystegia silvatica	Geophyte	Geophyte	1945	No	Northern hemisphere
Camelina sativa	Herb	Therophyte	Before 1930	No	Europe
Campanula poscharskyana	Herb	Perennial Hemikryptophyte	1990	Yes	Europe
Campylopus introflexus	Moss	N/A	1980	No	Sourthern hemisphere
Cardamine occulta	Herb	Hemikryptophyte	2007	Yes	Asia
Cardaminopsis arenosa arenosa	Herb	Monocarpic Hemikryptophyte	1867	No	Europe

Cerastium tomentosum	Herb	Herbal chamaephyte	1809	No	Europe
Chenopodium giganteum	Shrub	Therophyte	1932	No	Asia
Chenopodium pumilio	Herb	Therophyte	1938	No	Oceania
Chenopodium strictum	Herb	Therophyte	1874	No	Asia
Cicerbita macrophylla	Geophyte	Geophyte	1930	No	Asia
Cornus sanguinea australis	Shrub	Phanerophyte	2016	Yes	Europe
Cornus sericea	Shrub	Nanophanerophyte	1882	No	North America
Coronopus didymus	Herb	Therophyte	1856	No	South America
Corylus maxima	Tree	Phanerophyt	1920	No	Europe
Cotoneaster bullatus	Shrub	Nanophanerophyte	1997	Yes	Asia
Cotoneaster horizontalis	Shrub	Nanophanerophyte	1954	No	Asia
Crepis pulchra	Herb	Therophyte	1882	No	Europe
Crepis setosa	Herb	Therophyte	1833	No	Europe
Crocus tommasinianus	Geophyte	Geophyte	1998	Yes	Europe
Cyclamen coum	Geophyte	Geophyte	1998	Yes	Asia
Cyperus eragrostis	Herb/Geophyte	Geophyte	1996	Yes	South America
Digitalis purpurea	Herb	Therophyte	1856	No	Europe
Epilobium ciliatum	Herb	Herbal chamaephyte	1968	No	North America
Eragrostis virescens	Grass	Therophyte	1993	Yes	South America
Erigeron annuus annuus	Herb	Therophyte	1865	No	North America
Erigeron annuus	Herb	Therophyte	1865	No	North America
Erysimum cheiri	Herb	Herbal chamaephyte	1857	No	Europe
Euphorbia humifusa	Herb	Therophyte	1878	No	Asia
Euphorbia maculata	Herb	Therophyte	1870	No	North America
Euphorbia myrsinites	Herb	Herbal chamaephyte	1948	No	Asia
Euphorbia nutans	Herb	Therophyte	1915	No	North America
Euphorbia virgata	Herb	Perennial Hemikryptophyte	1876	No	Asia
Fallopia aubertii	Epiphyte	Phanerophyt, Liana	1904	No	Asia

Geranium sibiricum	Herb	Therophyte	1945	No	Asia
Glyceria striata	Geophyte	Geophyte	1950	No	North America
Helleborus orientalis	Geophyte/herb	Geophyte	1990	Yes	Asia
Heracleum mantegazzianum	Herb	Monocarpic Hemikryptophyte	1899	No	Asia
Hypericum calycinum	Shrub	Woody chamaephyte	Before 1930	No	Europe
Iris foetidissima	Herb	Perennial Hemikryptophyte	Around 1968	No	Europe
Juncus tenuis	Grass	Perennial Hemikryptophyte	1900	No	North America
Lamium galeobdolon argentatum	Herb	Hemikryptophyte	1874	No	Europe
Lemna minuta	Aquatic	Aquaphyt	1972	No	North America
Lepidium densiflorum	Herb	Therophyte	1899	No	North America
Lepidium virginicum	Herb	Therophyte	1856	No	North America
Leucanthemum praecox	Herb	Perennial Hemikryptophyte	1879	No	Asia
Linaria genistifolia dalmatica	Geophyte	Geophyte	1904	No	Europe
Linaria genistifolia	Herb	Unknown	1904	No	Asia
Linum usitatissimum	Herb	Therophyte	Before 1930	No	Europe
Lolium multiflorum	Herb	Therophyt	1848	No	Asia
Lonicera caprifolium	Epiphyte	Nanophanerophyte, Liana	1860	No	Europe
Lonicera henryi	Epiphyte	Nanophanerophyte, Liana	1997	Yes	Asia
Lunaria annua	Herb	Therophyte	1846	No	Europe
Lysimachia punctata	Herb	Perennial Hemikryptophyte	1892	No	Europe
Malus pumila	Tree	Phanerophyt	1874	No	Asia
Matricaria discoidea	Herb	Therophyte	1885	No	North America
Medicago arabica	Herb	Therophyte	1881	No	Africa
Mentha spicata	Geophyte	Geophyte	Before 1930	No	Europe
Muscari armeniacum	Geophyte	Geophyte	1990	Yes	Asia
Nepeta racemosa	Herb	Hemikryptophyte	1930	No	Asia

Nigella damascena	Herb	Therophyte	1832	No	Europe
Onobrychis viciifolia	Herb	Perennial Hemikryptophyte	1872	No	Europe
Oxalis dillenii	Herb	Therophyte	1950	No	North America
Oxalis stricta	Geophyte	Geophyte, Therophyte	1858	No	North America
Papaver croceum	Herb	Perennial Hemikryptophyte	1881	No	Asia
Papaver somniferum	Herb	Therophyte	Before 1930	No	Europe
Phalaris canariensis	Grass	Therophyte	1866	No	Africa
Picris echioides	Herb	Therophyte	1849	No	Africa
Pimpinella peregrina	Herb	Perennial Hemikryptophyte	1990	Yes	Europe
Pinus nigra	Tree	Phanerophyt	1861	No	Europe
Pinus strobus	Tree	Phanerophyt	1862	No	North America
Plantago coronopus	Herb	Therophyte	1904	No	Africa
Polygonum polystachyum	Geophyte	Geophyte	1917	No	Asia
Potentilla fruticosa	Shrub	Nanophanerophyte	1834	No	Northern hemisphere
Potentilla norvegica	Herb	Therophyte	1895	No	Northern hemisphere
Prunus cerasifera	Shrub and tree	Phanerophyt	1943	No	Asia
Prunus domestica domestica	Shrub and tree	Phanerophyt	1990	Yes	Asia
Prunus domestica	Shrub and tree	Phanerophyt	Before 1930	No	Asia
Prunus persica	Tree	Phanerophyt	1843	No	Asia
Pyrus communis	Tree	Phanerophyt	Before 1930	No	Europe
Quercus ilex	Tree	Phanerophyt	1909	No	Europe
Quercus rubra	Tree	Phanerophyt	1899	No	North America
Ribes aureum	Shrub	Phanerophyt	1885	No	North America
Ribes rubrum	Shrub	Nanophanerophyte	1873	No	Europe
Rosa multiflora	Epiphyte	Phanerophyt, Liana	1913	No	Asia
Rosa rugosa	Shrub	Nanophanerophyte	1900	No	Asia
Rubus armeniacus	Shrub	Nanophanerophyte	1990	Yes	Asia
Rubus laciniatus	Shrub	Nanophanerophyte	1882	No	Europe
Rubus phoenicolasius	Shrub	Nanophanerophyte	1957	No	Asia

Rumex longifolius	Herb	Perennial Hemikryptophyte	1955	No	Europe
Rumex patientia	Herb	Perennial Hemikryptophyte	1843	No	Europe
Sagittaria latifolia	Aquatic	Hydrophyt, Geophyte	1955	No	North America
Salvia officinalis	Shrub	Woody chamaephyte	Before 1930	No	Europe
Salvia verticillata	Herb	Perennial Hemikryptophyte	1835	No	Europe
Scilla forbesii	Geophyte	Geophyte	1996	Yes	Asia
Scutellaria altissima	Geophyte	Geophyte	1950	No	Europe
Sedum hybridum	Herb	Herbal chamaephyte	Around 1998	Yes	Asia
Sedum spurium	Herb	Herbal chamaephyte	1886	No	Asia
Senecio inaequidens	Herb	Herbal chamaephyte	1924	No	Africa
Setaria italica	Herb	Herbal chamaephyte	Before 1930	No	Asia
Sisymbrium altissimum	Herb	Therophyte	1878	No	Asia
Solanum chenopodioides	Herb	Therophyte	1971	No	South America
Solanum tuberosum	Geophyte	Geophyte	1861	No	South America
Solidago canadensis	Herb	Perennial Hemikryptophyte	1860	No	North America
Solidago gigantea	Herb	Perennial Hemikryptophyte	1881	No	North America
Stachys byzantina	Herb	Herbal chamaephyte	1882	No	Asia
Symphytum asperum	Herb	Perennial Hemikryptophyte	1895	No	Asia
Tanacetum parthenium	Herb	Perennial Hemikryptophyte	1827	No	Asia
Thlaspi alliaceum	Herb	Therophyte	1966	No	Europe
Thymus vulgaris	Shrub	Woody chamaephyte	Before 1930	No	Europe
Trifolium alexandrinum	Herb	Therophyte	1950	No	Asia
Trifolium resupinatum	Herb	Therophyte	1877	No	Asia
Verbena bonariensis	Herb	Hemikryptophyte	1917	No	South America
Veronica filiformis	Herb	Perennial Hemikryptophyte	1893	No	Asia
Veronica peregrina	Herb	Therophyte	1880	No	North America
Veronica persica	Herb	Therophyte	1842	No	Asia
Viburnum rhytidophyllum	Shrub	Nanophanerophyte	1935	No	Asia
Vicia faba	Herb	Therophyte	1882	No	Asia

Vicia lutea	Herb	Therophyte, Liana	1818	No	Europe
Vicia pannonica	Herb	Therophyte, Liana	1884	No	Europe
Vicia villosa varia	Herb	Therophyte, Liana	1882	No	Europe
Vicia villosa	Herb	Therophyte, Liana	1882	No	Asia
Vinca major	Shrub	Woody chamaephyte	1845	No	Europe

Table S5: *Frequency of selection of the three vegetation-based covariates and three transportation covariates by both native and non-native models.*

Frequency of variable selection		
	Native species	Non-native species
Canopy cover (min, max, med)	21	22
Proportion of coniferous forest	43	28
Proportion of deciduous forest	43	29
Noise pollution index	31	45
Euclidian distance to road	48	41
Topographic distance to road	22	14

Appendix II

Supplementary Material of Chapter II

Appendix 1

Variable selection rationale:

The environmental variable selection was reasoned from the standpoint that many of the studied species (all the equilibrium species and some of the invasive species) are mountains species. As such, we focused on powerful limiters of the harsh mountain climate – drought and cold.

We therefore chose precipitation seasonality, as it is a proxy for the harsh mountain climate. We considered it to be a better one than high or low temperature alone.

We also included the ranges of temperatures, as these variables incorporate the elevational gradient that many of the target species live on.

In addition, since species distributions often lag years or even decades behind the environmental changes (see section 2.4), we placed the focus on variables that summarized the variables over a year, or a quarter, if available. We reasoned that those variables that averaged values over a year or a quarter included the information of variables that averages values over a single month.

Table S1: Performance metrics comparisons across treatments and across species group (invasive alien species (IAS) and equilibrium species (EQ)) for real species. AUC = Area Under Curve, CBI = Continuous Boyce Index, with the p-value of t-tests (either paired or unpaired, depending on the variances being equal or unequal) comparing the values between the two species groups.

		Generation 1			***Generation 2***		
		AUC	*CBI*	*AUC+CBI*	*AUC*	*CBI*	*AUC+CBI*
Randgeo	*IAS*	0.86074	0.95544	1.808078	0.73092	0.8519	1.575146
	EQ	0.87046	0.92581	1.785798	0.74581	0.86201	1.599333
	p – value	0.8569	0.2033	0.7623	0.3543	0.5602	0.3952
Randenv	*IAS*	0.87037	0.95173	1.815333	0.78652	0.87625	1.655695
	EQ	0.85989	0.92082	1.772955	0.76866	0.88615	1.646117
	p – value	0.8504	0.2248	0.5803	0.2374	0.4652	0.6977
Stratenv	*IAS*	0.92012	0.95299	1.866679	0.89386	0.88616	1.770245
	EQ	0.95095	0.93607	1.880653	0.93418	0.88956	1.813893
	p – value	0.1064	0.4035	0.6919	2.2^{e}-16	0.7965	0.006357
cb1	*IAS*	0.86064	0.95668	1.809102	0.73013	0.85677	1.578555
	EQ	0.87035	0.92389	1.783834	0.7708	0.87817	1.639858

	p – value	0.8576	0.161	0.7336	0.01232	0.1935	0.0277
cb2	*IAS*	0.86096	0.95926	1.811899	0.72612	0.85539	1.573475
	EQ	0.87044	0.92531	1.783574	0.74393	0.86881	1.603456
	p – value	0.8609	0.1151	0.6961	0.2994	0.4388	0.3159

Effects of number of generated occurrences

The number of generated occurrence points seems to have a positive correlation in model stability across all real-world species and strategies ()

Table S2: Effect of nr of generated occurrence records on real-world species. Average similarity values and correlation of those averages with the number of generated occurrences.

Generated occurrences	Stratenv			Randenv			Randgeo			cb1			cb2		
	Total	EQ	IAS	Total	EQ	IAS	Total	EQ	IAS	Total	EQ	IAS	Total	EQ	IAS
1000	0.493	0.503	0.484	0.837	0.780	0.893	0.451	0.424	0.478	0.459	0.381	0.534	0.466	0.457	0.475
500	0.494	0.509	0.479	0.837	0.782	0.892	0.451	0.428	0.473	0.453	0.368	0.533	0.455	0.442	0.467
300	0.486	0.496	0.477	0.844	0.800	0.888	0.448	0.428	0.469	0.458	0.380	0.532	0.459	0.453	0.464
100	0.469	0.482	0.456	0.822	0.771	0.872	0.445	0.429	0.459	0.460	0.389	0.529	0.446	0.443	0.449
50	0.474	0.495	0.452	0.784	0.728	0.833	0.441	0.414	0.466	0.436	0.346	0.521	0.464	0.440	0.485
30	0.483	0.484	0.483	0.748	0.695	0.811	0.402	0.388	0.417	0.467	0.343	0.584	0.407	0.383	0.432
10	0.439	0.441	0.436	0.693	0.607	0.766	0.433	0.457	0.412	0.424	0.362	0.469	0.409	0.386	0.442
Correlation	0.635	0.582	0.614	0.625	0.566	0.689	0.554	0.052	0.650	0.352	0.494	0.120	0.594	0.613	0.507

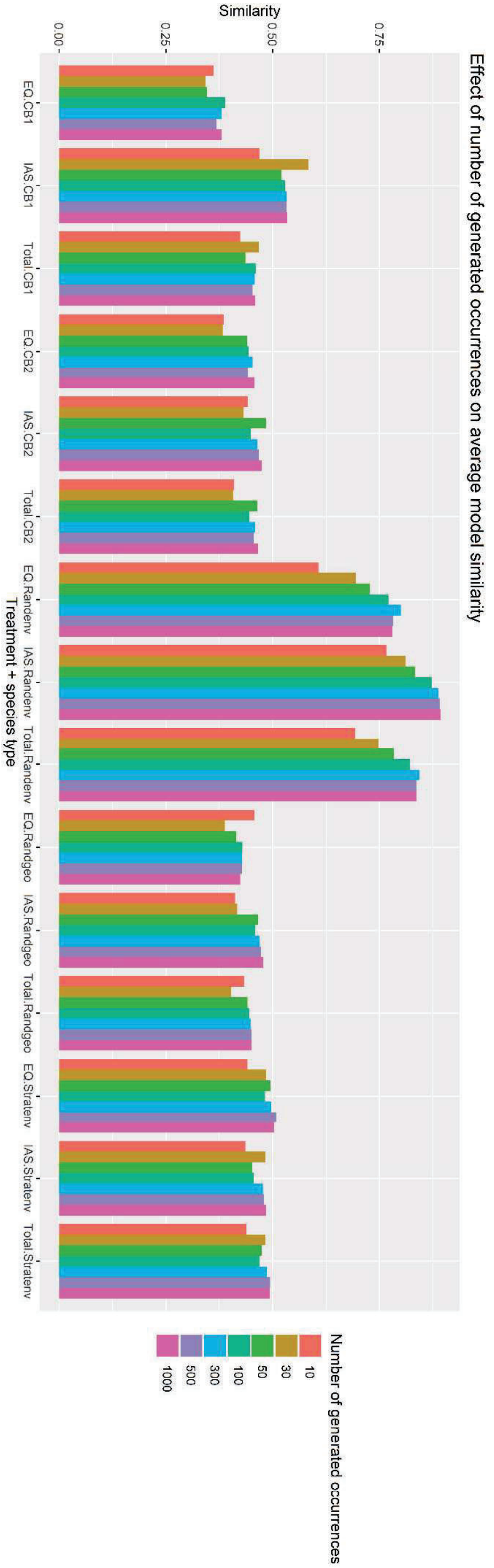

Effect of number of generated occurrences on average model similarity
Similarity
0.00
0.25
0.50
0.75
EQ.CB1
IAS.CB1
Total.CB1
EQ.CB2
IAS.CB2
Total.CB2
EQ.Randenv
IAS.Randenv
Total.Randenv
EQ.Randgeo
IAS.Randgeo
Total.Randgeo
EQ.Stratenv
IAS.Stratenv
Total.Stratenv
Treatment + species type
Number of generated occurrences
10
30
50
100
300
500
1000

Figure S1: Model similarity of species models per number of generated occurrence points in the generation 2 models, across treatments and for both equilibrium species (EQ) and invasive alien species (IAS).

Table S3: Within-treatment comparison of average model similarity, i.e. stability for real equilibrium species and real invasive alien species. Averages and p-values of t-tests assuming unequal variances.

	Stratenv	Randgeo	Randenv	cb1	cb2
IAS vs EQ (p – value)	0.1089561	0.008676833	< 2.2^{e}-16	< 2.2^{e}-16	0.08221
Average EQ	0.4908881	0.4222804	0.7583992	0.3699268	0.4369632
Average IAS	0.4694295	0.4614234	0.8652017	0.5321942	0.463163
STD error EQ	0.00967071	0.009600233	0.004921816	0.01088874	0.00958933
STD error IAS	0.009242786	0.01138823	0.002234767	0.01176738	0.01161806

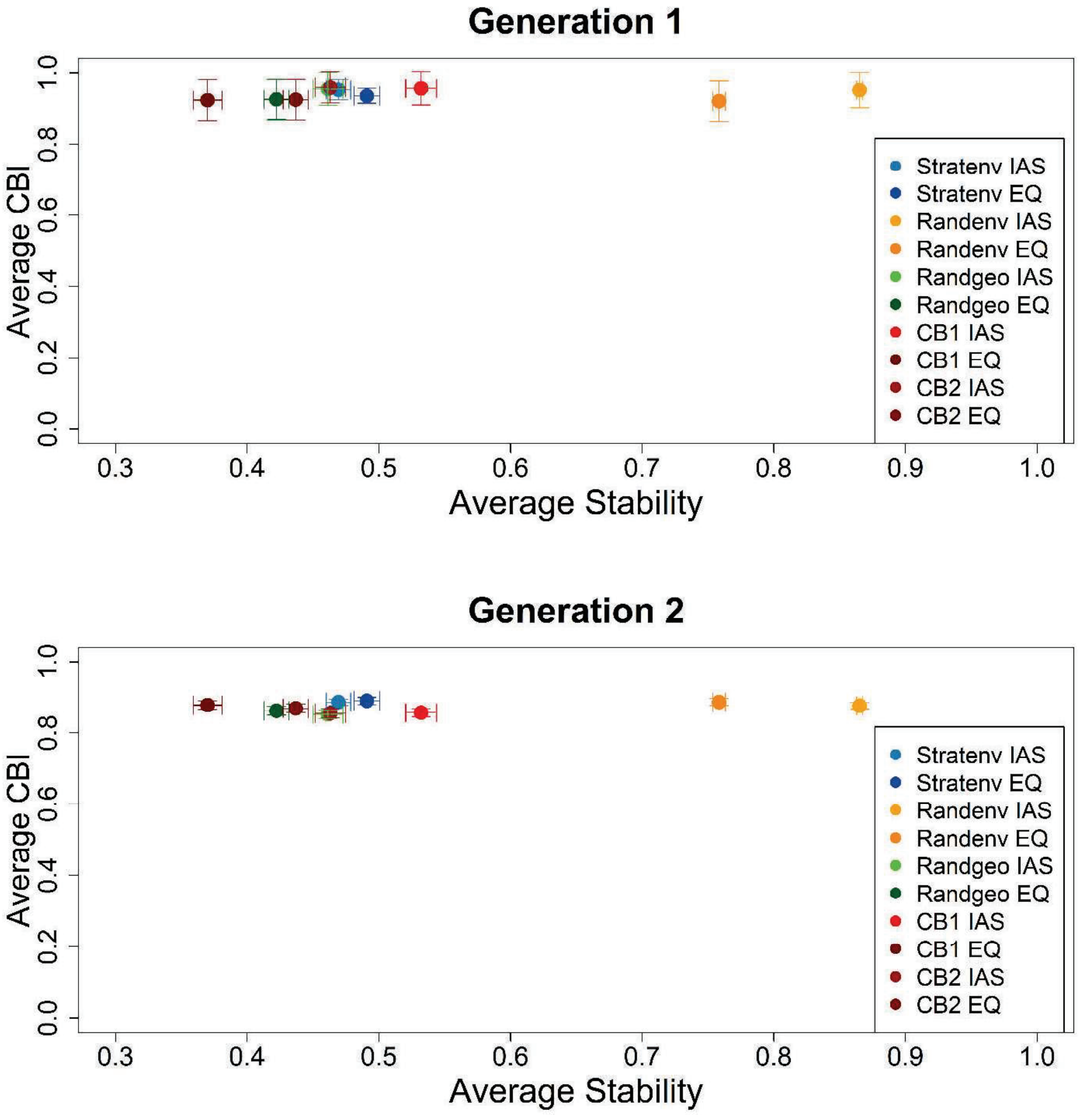


Figure S2: Average model similarity vs average Boyce index (CBI) for real species models of generation 1 (the models that were run using real occurrence data) (A) and generation 2 (the models that were run using virtual occurrence data generated from the output suitability maps of generation 1 models) (B).

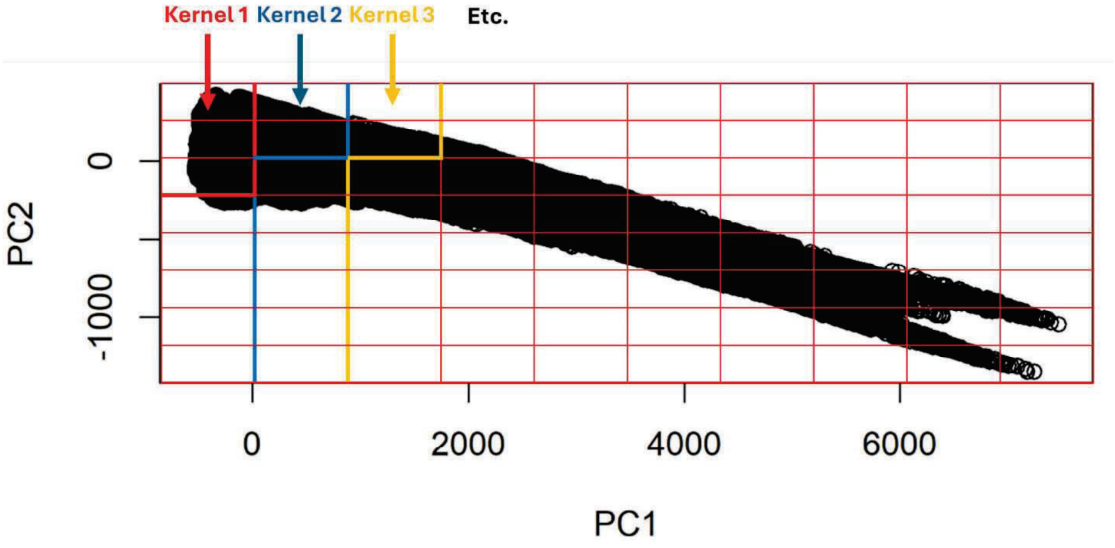


Figure S3: Graphical representation of the sampling of the fully random in environmental space background data: division of the environmental PCA axes of the whole study area into kernels and summing of the appropriate squares as per the density of data points.

Table S4: number of generation 2 models that did not yield performance metrics.

	randgeo	cb1	cb2	stratenv	randenv	Total
Aedes albopictus	12	8	8	4	8	40
Asclepias syriaca	12	12	12	4	8	48
Branta canadensis	4	4	8	2	8	26
Cameraria ohridella	6	4	4	2	2	18
Harmonia axyridis	4	4	4	2	8	22
Heracleum mantegazzianum	4	2	8	2	10	26
Impatiens glandulifera	8	8	6	2	12	36
Ips typographus	4	8	4	4	8	28
Lespedeza cuneata	4	6	4	8	4	26
Lupinus polyphyllus	8	6	8	2	12	36
Myocastor coypus	4	6	4	6	4	24
Ondatra zibethicus	4	4	4	6	8	26
Popillia japonica	8	12	6	2	10	38
Sciurus niger	6	6	10	8	8	38
Tamias sibiricus	14	14	12	6	6	52
Alnus alnobetula	12	12	12	2	6	44
Aster alpinus	14	6	12	6	8	46
Boloria pales	4	6	4	4	4	22
Colias phicomone	6	6	6	2	6	26
Erebia meolans	6	4	4	4	4	22
Helictrotrichon pratense	6	4	4	6	4	24
Lagopus muta	8	4	10	2	10	34
Leontopodium nivale	6	8	6	10	6	36
Lepus timidus	6	4	6	4	10	30
Marmota marmota	6	6	4	2	6	24
Nucifraga caryocatactes	6	8	6	4	10	34
Parnassius apollo	8	8	4	2	8	30
Parnassius mnemosyne	10	12	12	4	10	48
Ranunculus glacialis	6	6	6	0	4	22
Tetrao urogallus	8	2	2	4	4	20
	214	200	200	116	216	946

Table S5: Additional rationale of why each species was included into the study.

Category	**Species group**	**Species name**	**Rationale**	**Number of occurrences**
Invasive	Plants	*Asclepias syriaca*	On the List of Invasive Alien species of Union Concern. Likely sensitive to temperature. (https://www.cabi.org/isc/datasheet/7249#toclimate)	1342
		Heracleum mantegazzianum	Mentioned in both the EU list of Union Concern and by BAFU. Mountain species in native range. Already present at 1500m altitude.	833
		Impatiens glandulifera	Mentioned in both the EU list of Union Concern and by BAFU. Mountain species in native range. Already present at high altitudes.	769
		Lespedeza cuneata	Generalist invasive species that is likely to spread to	236

			higher altitudes along with developing infrastructure.	
		Lupinus polyphyllus	Mentioned in both the EU list of Union Concern and by BAFU. Mountain species in native range. Already present at 1500m altitude.	820
	Vertebrates	*Branta canadensis*	High ecological impact ([CABI] [DAISIE] [NOBANIS])	920
		Myocastor coypus	Mentioned in the EU list of Union Concern. Generalist species that will likely spread to high altitudes.	400
		Ondatra zibethicus	Mentioned in both the EU list of Union Concern and by BAFU. Generalist species that will likely spread.	281
		Sciurus niger	Mentioned in the EU list of Union Concern. Generalist species that will likely spread to high altitudes.	686

		Tamias sibiricus	Mentioned in the EU list of Union Concern. Generalist species that will likely spread to high altitudes.	77
	Insects	*Aedes albopictus*	Vector of many diseases, including malaria and Denge fever. Also likely to spread to higher altitudes.	95
		Cameraria ohridella	Pest of forests cold-tolerant, and as an insect, it is likely to react to climate.	225
		Harmonia axyridis	List of Union Concern, seemed not to be occurring everywhere. Is also an insect, likely bound to temperature, as states its CABI page: https://www.cabi.org/isc/datasheet/26515#toclimate	1377
		Ips typographus	Invasive species and pest particularly of Picea abies, which occurs at high altitudes.	77

			Insect species which likely responds to climate.	
		Popillia japonica	Invasive species and its distribution is likely to spread more northward, meaning likely also to high altitudes.	794
Equilibrium	Plants	*Alnus alnobetula*	Typical of high altitudes and latitudes (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	164
		Aster alpinus	Mountain species, adapted to high environments. Indicator species of high mountain pastures (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	86
		Helictrotrichon pratense	Indicator species of high mountain pastures	70

			(infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	
		Leontopodium nivale	Mountain species, adapted to cold, habitat likely to shrink (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	423
		Ranunculus glacialis	Indicator species of high mountain pastures (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	107
	Vertebrates	*Lagopus muta*	Indicator species of high mountain pastures (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	64
		Lepus timidus	Mountain species (infospecies.ch/fr/projets/infr	286

			astructure-ecologique.html#berichte)	
		Marmota marmota	Mountain species (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	123
		Nucifraga caryocatactes	Indicator species of high-altitude coniferous forests (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	433
		Tetrao urogallus	Indicator species of high mountain pastures (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	445
	Insects	*Boloria pales*	Mountain species, typical of high-altitude pastures. (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	75

		Colias phicomone	Mountain species, adapted to high environments. Indicator species of high mountain pastures. (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	87
		Erebia meolans	Alpine species (http://www.pyrgus.de/Erebia_meolans_en.html)	140
		Parnassius apollo	Indicative species of high mountain pastures infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	212
		Parnassius mnemosyne	Indicative species of high mountain pastures (infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte)	93

References

List of Invasive Alien Species of Union concern – Environment – European Commission. (n.d.). Retrieved October 6, 2022, from https://ec.europa.eu/environment/nature/invasivealien/list/index_en.htm

OFEV, O. fédéral de l'environnement. (n.d.). Espèces exotiques en Suisse. Retrieved April 7, 2023, from https://www.bafu.admin.ch/bafu/fr/home/themen/thema-biodiversitaet/biodiversitaet–publikationen/publikationen-biodiversitaet/gebietsfremde-arten-in-der-schweiz.html

InfoSpecies. (n.d.). Retrieved April 7, 2023, from https://www.infospecies.ch/fr/projets/infrastructure-ecologique.html#berichte

DAISIE - Inventory of alien invasive species in Europe. (n.d.). Retrieved May 15, 2024, from https://ipt.inbo.be/resource?r=daisie-checklist

NOBANIS - European Network on Invasive Species. (n.d.). Retrieved May 15, 2024, from https://www.nobanis.org/

European Lepidoptera and their ecology: Erebia meolans. (n.d.). Retrieved May 16, 2024, from http://www.pyrgus.de/Erebia_meolans_en.html

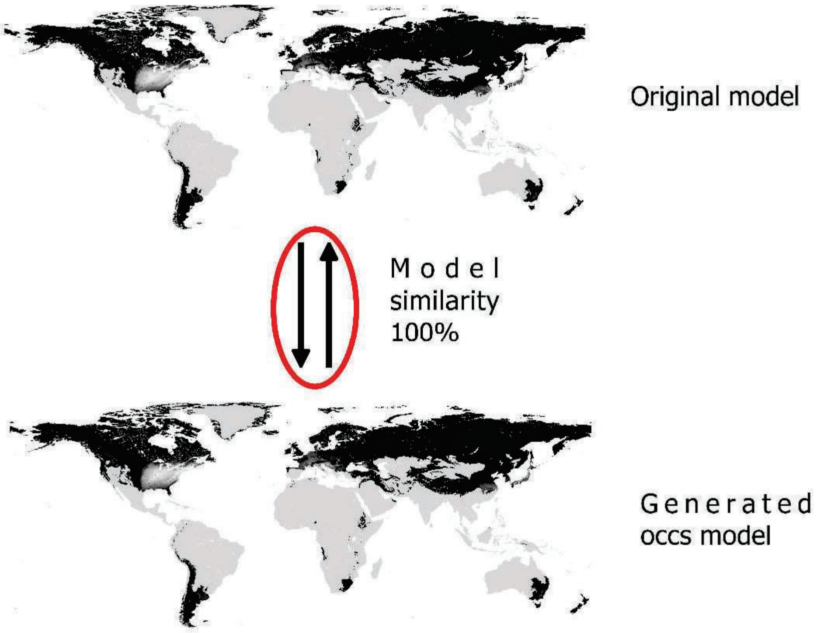


Figure S4: An idealized example of model stability, i.e., perfect similarity between a model of generation 1 and one of generation 2.

Appendix 2. ODMAP reporting protocol.

How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels

– ODMAP Protocol –

Author list hidden for peer-review

Journal: Ecological Modelling

2024-02-29

Overview

Authorship

Contact: hidden for peer-review

Study link: [doi]

Model objective

Model objective: Analysis of model performance and stability

Target output: current habitat suitability maps for 30 species across the world at 30 arc seconds resolution

Focal Taxon

Focal Taxon: 30 species (see main text and appendix 1 for details)

Location

Location: Seven biomes: temperate broadleaf and mixed forests, temperate coniferous forests, temperate grasslands, savannahs and shrublands, montane grasslands and shrublands, boreal forests/taiga and tundra

Scale of Analysis

Spatial extent: -180.000, 179.9999, -55.77514, 83.64153 (xmin, xmax, ymin, ymax)

Spatial resolution: 30 arc seconds

Temporal extent: 1979-2013 for the environmental variables and 2000-2022 for the occurrence data.

Boundary: geographical, biomes specified by Olson et al. 2001.

Biodiversity data

Observation type: observation, human observation, Material sample, Living specimen, specimen (source: GBIF)

Response data type: presence / background data

Predictors

Predictor types: bioclimatic

Hypotheses

Model assumptions: 1) candidate covariates potentially drive the distribution of the species and key covariates are included in the models, 2) The observed distributions of some target species are in equilibrium with the environment, and for others this is not the case, 3) The species maintain their niches across the specified time period; 4) The correlation structure between covariates does not change over time; 5) The background datasets used give the model a good impression of the available environment (hypothesis tested)

Algorithms

Modelling techniques: Maxnet (MAX) (Phillips et al. 2017)

Model complexity: hyperparameters set using ENMeval 2.0.3 (Kass et al. 2021)

Model averaging/ensemble modelling: N/A

Workflow

Model workflow: Only weakly correlated predictors (<70% correlation) were retained in the analysis. We performed parameter tuning for the regularisation parameter (1 to 5 in increments of 1) and combinations of modelfitting functions (LQH, LQP or LQHP) (b) in MaxEnt models based on AUC, Boyce and the sum of both using the ENMeval R package.

Software

Software: ENMeval v.2.0.3 (Kass et al., 2021) and R v4.2.2

Code availability: TBD

Data availability: https://zenodo.org/records/10255654

Data

Biodiversity data

Taxon names: 30 species, see publication for details

Ecological level: species

Data sources: Occurrence records for all the world for matching species were obtained from the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/) in several installments (on 03 June, 2022 (https://doi.org/10.15468/dl.tc6fqe; https://doi.org/10.15468/dl.2m6kf8), July 21, 2022 (https://doi.org/10.15468/dl.ns4jzr), October 31, 2022 (https://doi.org/10.15468/dl.mpcxws), 20 May 2023 (https://doi.org/10.15468/dl.cbxen3)).

Sampling design: Citizen science data. No sampling design can be mentioned.

Sample size: Only species with more than 50 occurrence records in the 80% bin after spatial disaggregation and were retained for modelling.

Cleaning: Removal of non-native records for half of the study species using Dasco R package, R cleaning, 1 point per 30 arc second raster grid cell.

Background data: 10,000 background absences were generated using special strategies of sampling is geographic and environmental space.

Data partitioning

Occurrence data: Split-sample approach repeated 50 times with 20% of the data withheld for models of generation 1. The 80%bin was split 50-50 in training and validation bins for models of generation 1(see manuscript) and in models of generation 2, the full 100% data was split 50-50 in. No selection of truly independent test data was done.
Background data: 50-50 training-validation bins, or using checkerboard sampling design (aggregation factor 5 or 10).

Predictor variables

Predictor variables: Suite of 19 bioclimatic candidate covariates CHELSA version 1.2 (bio1, bio2, bio7, bio12, bio15, bio17, bio19).

Data sources: CHELSA (Karger et al. 2017).

Spatial extent: -180.000, 179.9999, -55.77514, 83.64153 (xmin, xmax, ymin, ymax)

Spatial resolution: 30 arc seconds

Coordinate reference system: WGS84

Temporal extent: 1979-2013

Transfer data

N/A (no future projections were done)

Model

Selected predictors

Covariate selection: Use of ecological knowledge; see additional paper attachments.

Multicollinearity

Multicollinearity: Only variables less than 70% correlated were included.

Model settings

Hyperparameter settings were varied: regularization multiplier (1 to 5, increments of 5) and used combinations of function types (LQH, LQP, LQHP)

Model estimates

N/A

Threshold selection

N/A

Assessment

Performance statistics

Performance on training data: Model accuracy was evaluated using a split-sample approach repeated 50 times with 50% of the data kept for validation. For each model, the best combination of hyperparameters was identified using the average "Score" of three evaluation metrics including the Area Under the Curve' (AUC'), the Continuous Boyce Index (CBI) and the sum of both, with AUC first adapted to have the same "weight" as the CBI (as AUC^'=AUC*2-1).

Performance on validation data: same as for training data

Plausibility check

No plausibility checks were performed.

Prediction

Prediction output

Prediction unit: Pixel-level habitat suitability values

Uncertainty quantification

No uncertainty quantification was done.

References

Downloading and cleaning GBIF data with R | R-bloggers (n.d.). Retrieved October 6, 2022, from https://www.r-bloggers.com/2021/03/downloading-and-cleaning-gbif-data-with-r/

hseebens. (2022). Hseebens/DASCO workflow [HTML]. https://github.com/hseebens/DASCOworkflow/blob/30d6467a6f3259c5a0f622fe1d6a8ef4dbf5b159/run_DASCO_workflow.R (Original work published 2021)

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. Scientific Data, 4(1), 170122. https://doi.org/10.1038/sdata.2017.122

Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C. L., Pinilla-Buitrago, G. E., Boria, R. A., Soley-Guarda, M., Anderson, R. P. (2021). ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. Methods in Ecology and Evolution, 12(9), 1602–1608. https://doi.org/10.1111/2041-210X.13628

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., Kassem, K. R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience 51(11):933-938. Url: https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world

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. Ecological Applications, 19(1), 181-197.

Appendix 3: *supplementary methods*

- *Supplementary analysis with alternate sampling scheme*

The main analysis of the paper was repeated on a different set of occurrence records of the invasive alien species *Heracleum mantegazzianum* and the equilibrium species *Aster alpinus*. Instead of the Minimum Convex Polygon thinning in the main paper, the cleaned (and for *Aster alpinus* records that were out of its non-native range, as determined by the DASCO R package) occurrence records were subjected to a thinning based on the distance from artificial surface and urban area, i.e., the more distant an occurrence point is from human infrastructure or urban area, the more likely it is to be retained. The resulting datasets contained the same number of occurrence records as the MCP thinned ones. This method of thinning is referred to below as "arturb".

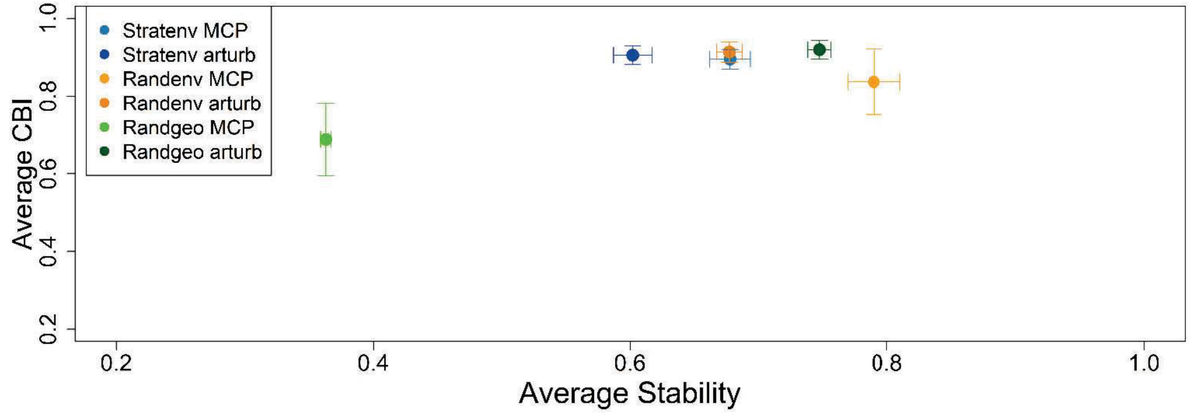
We compared the resulting Area under Curve (AUC), Continuous Boyce Index (CBI) and similarity scores to the results of the original MCP thinned analyses of the two target species for the randgeo, randenv and stratenv BP selection strategies, meaning we used the exact same BP datasets used in the main paper for the new analysis.

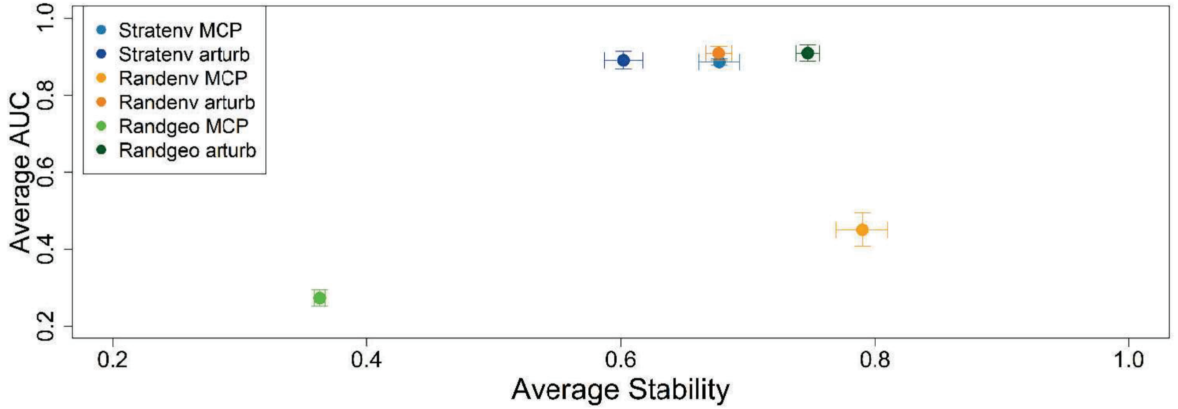
The thinning by distance to human infrastructure strategy was chosen because it eliminates a likely sampling bias of the GBIF citizen science sampling: the predilection of citizen scientists to sample close to the infrastructure.

Aster alpinus and *Heracleum mantegazzianum* were chosen because they are both mountain species in their home range and might therefore be strongly affected by the sampling bias correction of the new thinning method.

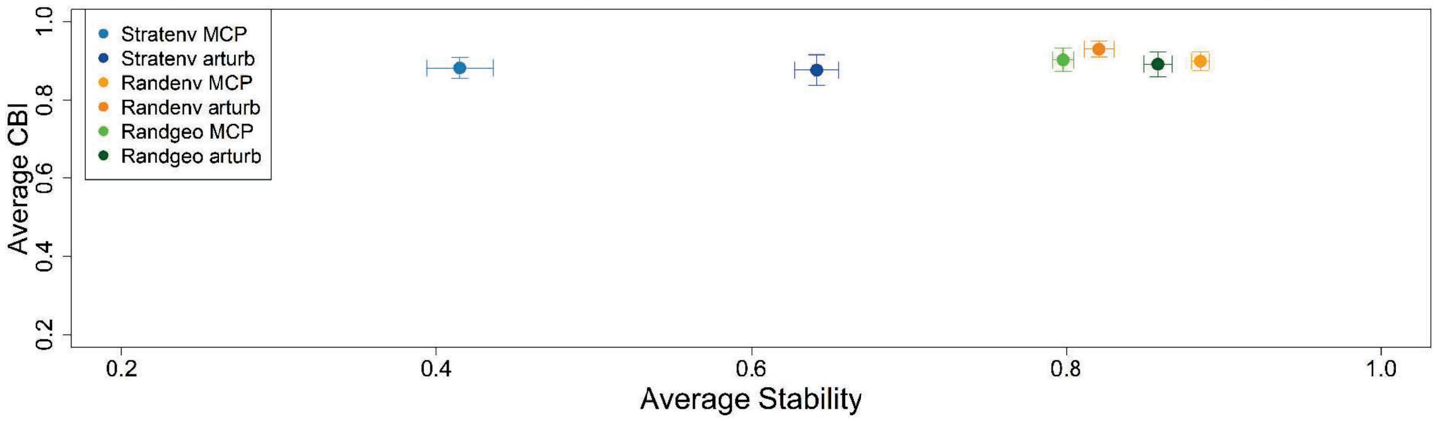
The new method of thinning the occurrence records by distance to infrastructure was chosen because it is likely to eliminates citizen science sampling bias.

The results of the arturb thinning versus the MCP thinning are presented in Figure S31.


Average model similarity vs average model CBI per treatment Aster alpinus


Average model similarity vs average model AUC per treatment Aster alpinus

Average model similarity vs average model CBI per treatment Heracleum mantegazzianum



Average model similarity vs average model AUC per treatment Heracleum mantegazzianum

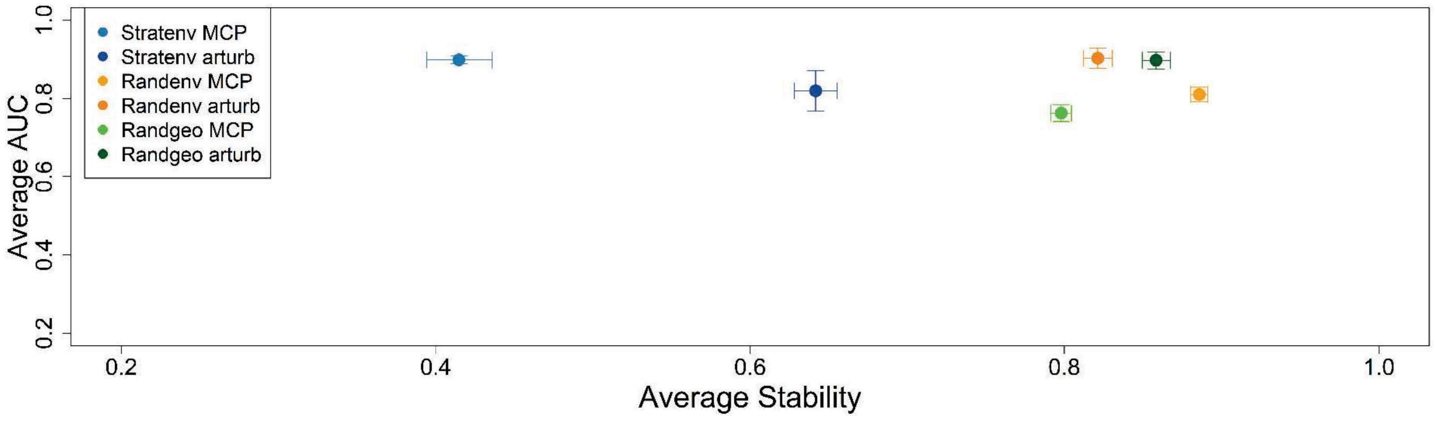


Figure S31: Scatterplots comparing the AUC and Boyce index to similarity (i.e., stability score) results for the arturb thinning method versus the MCP thinning method, and the randgeo, randenv and stratenv strategies.

The vastly improved model stability of the *Heracleum mantegazzianum* arturb models for the stratenv method may be explained by the fact that the arturb thinning of this species' occurrence records has left relatively more occurrence records in mountains, where the stratenv BP are mostly located. It may therefore be easier for the stratenv models to identify relationships of species occurrence data with the environment.

It is worth mentioning that *H. mantegazzianum* spreads primarily via roads and human infrastructure and that therefore, the arturb method of thinning likely eliminates relevant ecological information.

In addition, the randenv and randgeo performance metrics, particularly AUC, for *Aster alpinus* are very significantly higher when the arturb thinning method is used. This would indicate a far higher ability of these models to discern suitable from unsuitable area. This may be because the environmental variables included in the models become more meaningful when the occurrence records of the species are further from infrastructure, i.e., likely higher in mountains. Perhaps therefore, the occurrence records provide a closer approximation of the species' actual niche.

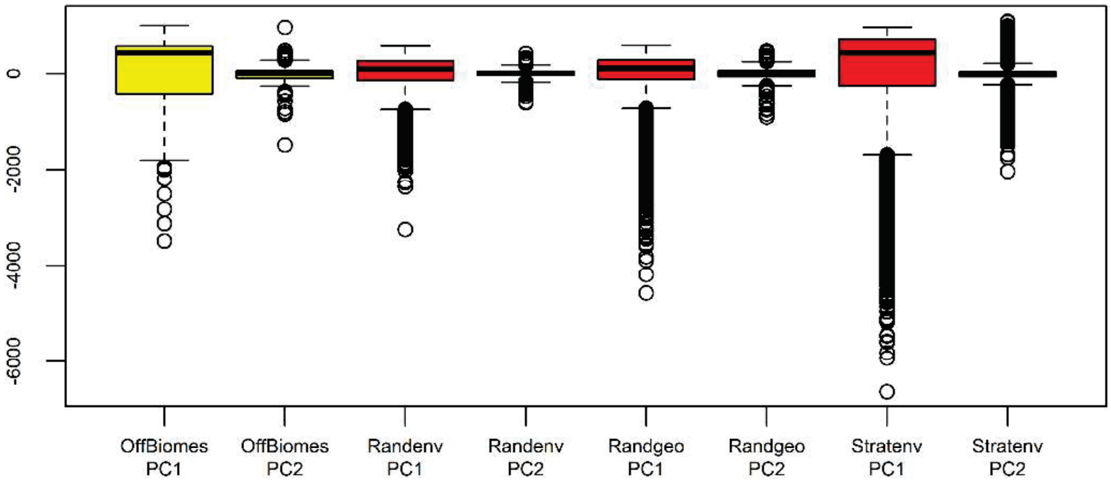
This discerning capacity of the models might also explain the greatly improved model stability of the *Aster alpinu*s arturb models for the randgeo strategy: the Generation 2 virtual species occurrence models are more likely to be generated in high mountains, where the Generation 2 models can easily discern an environmental gradient.

Therefore, the new arturb thinning method might artificially inflate AUC and stability values, rather than present a better ecological approximation. Nevertheless, this supplementary analysis highlights that the thinning of occurrence records vastly influences the model performance and similarity.

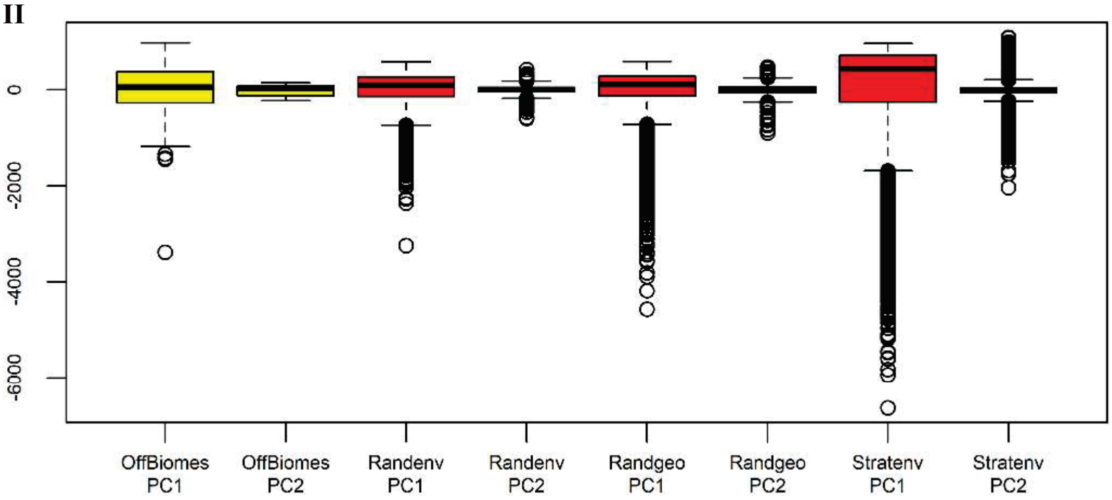
We would like to underline that unlike the MCP thinning method, the arturb thinning method makes an inherent assumption about the distribution of the occurrence records. Whilst this assumption may hold up for many records in the GBIF database, it cannot be assumed that it is ubiquitous across all taxa.

- *Additional analyses on the effect of omitting the occurrence data that fall outside the study area*

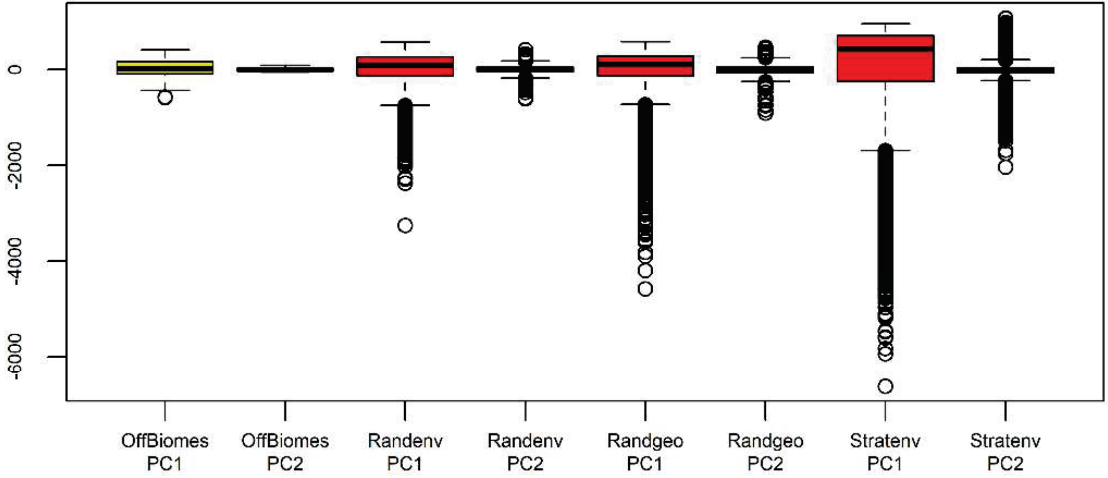
I



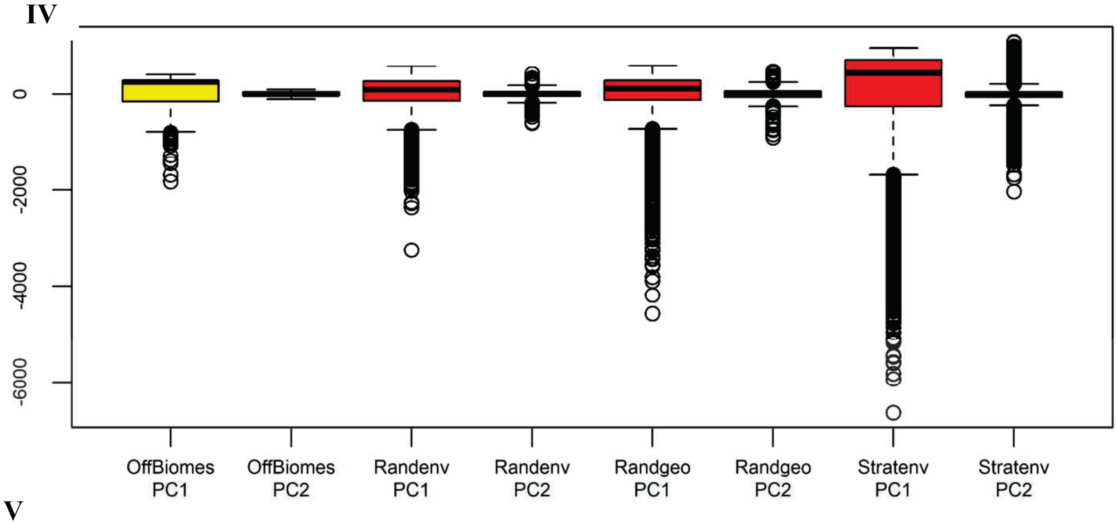
II



III


0
-2000
-4000
-6000
OffBiomes PC1
OffBiomes PC2
Randenv PC1
Randenv PC2
Randgeo PC1
Randgeo PC2
Stratenv PC1
Stratenv PC2

IV


0
-2000
-4000
-6000
OffBiomes PC1
OffBiomes PC2
Randenv PC1
Randenv PC2
Randgeo PC1
Randgeo PC2
Stratenv PC1
Stratenv PC2

V

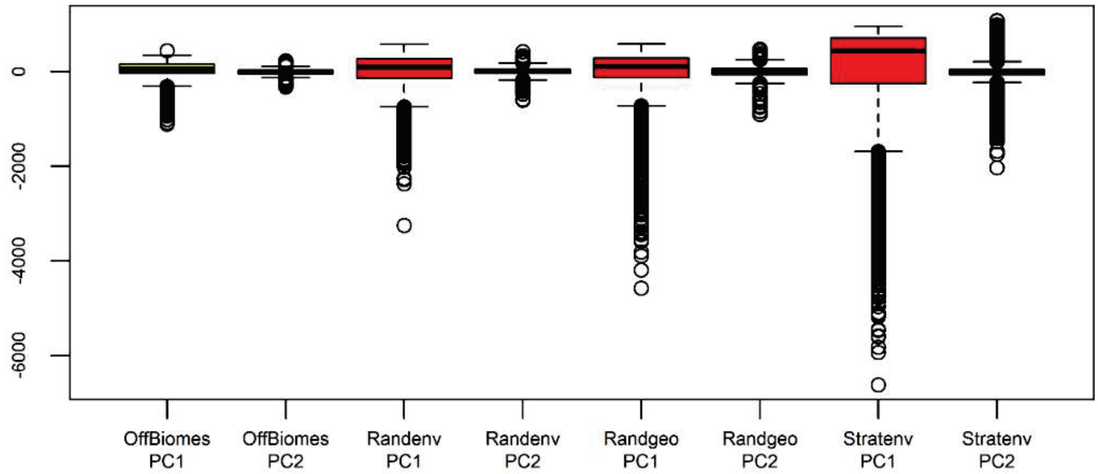
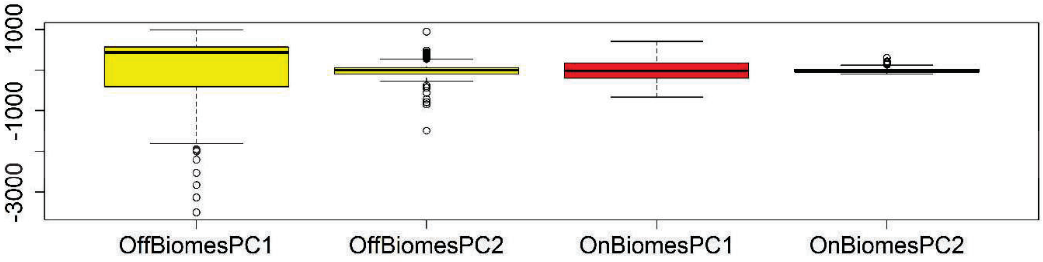

0
-2000
-4000
-6000
OffBiomes PC1
OffBiomes PC2
Randenv PC1
Randenv PC2
Randgeo PC1
Randgeo PC2
Stratenv PC1
Stratenv PC2

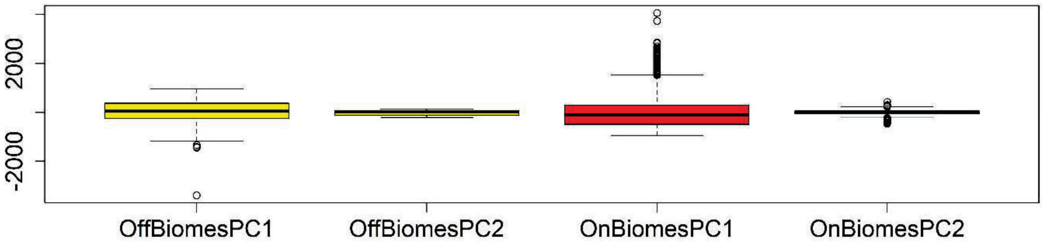
Figure S32: boxplots of PCA analyses of the data off the biomes for Aedes albopictus *(I),* Alnus alnobetula *(II)*, Aster alpinus *(III)*, Lepus timidus *(IV) and* Sciurus niger *(V), (the two invasive alien species and the three equilibrium species with the highest proportion of occurrence data off the biomes)*, *compared to the PCA values of the randgeo, randenv and stratenv background point datasets. These PCA analyses were performed independently on each occurrence and background points dataset and the proportion of explained variation was always over 70%.*

Results discussion: We see that the BP boxplots encompass most of the species presence data and therefore, we assume that not many values at the species presence points that were omitted are different from values that were included in the models. A possible exception to this are the results of *Aedes albopictus*. We however emphasize that invasive alien species like this one were specifically selected as target species because they are not at equilibrium with the environment, i.e., their niche is expanding. Therefore, a potential partial truncation of the niche should not be a problem.

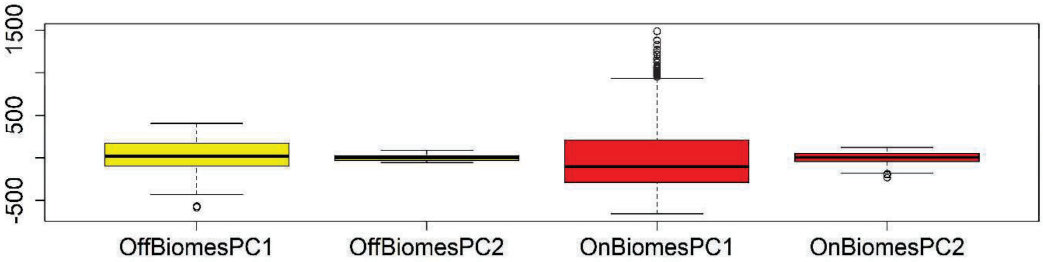
Aedes albopictus



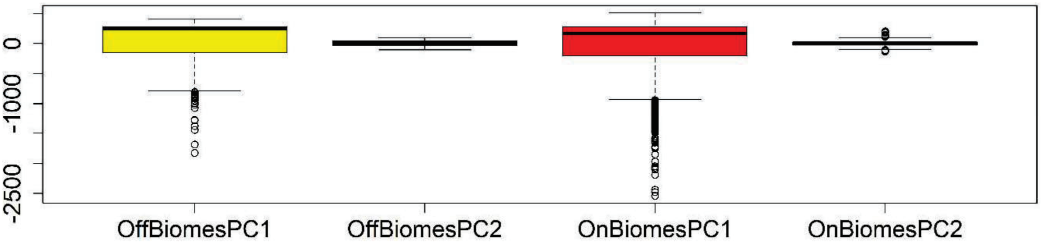
Alnus alnobetula



Aster alpinus



Lepus timidus



Sciurus niger

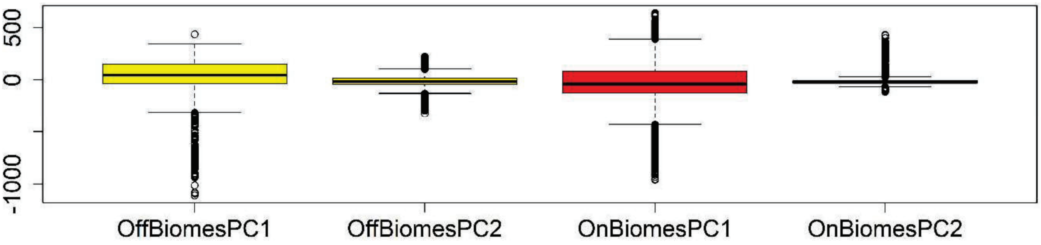


Figure S33: boxplots of PCA analyses of the data off the biomes for Aedes albopictus*,* Alnus alnobetula, Aster alpinus, Lepus timidus *and* Sciurus niger*, (the two invasive alien species and the three equilibrium species with the highest proportion of occurrence data off the biomes), compared to the PCA values of the same species on the biomes. These PCA analyses were performed independently on each occurrence and background points dataset and the proportion of explained variation was always over 70%.*

Table S31: number of occurrence points on and off the biomes for all target species, plus the calculated proportion of occurrence points off the biomes.

Species	Nr before biomes extraction	Number after biomes extraction	Prop off biomes
Alnus alnobetula	1919	1791	0.066701407
Lagopus muta	2449	2402	0.019191507
Tetrao urogallus	12024	11924	0.0083167
Aster alpinus	856	793	0.073598131
Colias phicomone	654	650	0.006116208
Nucifraga caryocatactes	14338	13555	0.054610127
Erebia meolans	957	921	0.037617555
Lepus timidus	6172	5671	0.08117304
Helictotrichon pratense	938	909	0.030916844
Leontopodium nivale	2922	2831	0.031143053
Parnassius apollo	22546	21803	0.032954848
Parnassius mnemosyne	709	709	0
Ranunculus glacialis	2649	2646	0.001132503
Marmota marmota	2199	2199	0
Boloria pales	973	964	0.009249743
Aedes albopictus	1512	1105	0.269179894
Asclepias syriaca	25658	24877	0.030438849
Branta canadensis	212895	199835	0.061344794
Cameraria ohridella	226324	226296	0.000123716
Harmonia axyridis	139389	130898	0.060915854
Heracleum mantegazzianum	47198	46826	0.00788169
Impatiens glandulifera	133134	131751	0.01038803
Ips typographus	1126	1109	0.015097691
Lespedeza cuneata	1304	1253	0.039110429
Lupinus polyphyllus	78498	76152	0.029886112
Myocastor coypus	14070	12922	0.08159204
Ondatra zibethicus	457234	454499	0.00598162
Popillia japonica	7288	7156	0.018111965
Sciurus niger	18967	11828	0.376390573
Tamias sibiricus	1590	1585	0.003144654

Appendix III

Supplementary Material of Chapter III

Appendix 1: Resampling of the present-day land use datasets (created by the European Space Agency) to match the categories of the future (2050) land use dataset (created by Clark lab of Clark university, USA)

 # Mostly cropland from LUC selection:

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 10 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 11 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 12 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 20 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 30

 # Forest

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 40 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 50 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 60 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 61 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 62 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 70 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 71 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 72 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 80 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 81 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 82 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 90 OR

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 100

 # Grassland, scrubs and shrub

"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 110 OR

```
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 120 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 121 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 122 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 130 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 140
# Need to fuse category 40 (Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland(<50%)) with forest, since forest includes areas of >15% tree cover.
# Sparse vegetation
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 150 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 152 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 153
# Swampy or Often Flooded Vegetation
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 160 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 170 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 180
# Artificial Surface or Urban Area
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 190
# Bare areas
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 200 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 201 OR
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 202
# Surface water
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 210
# Permanent snow and ice
"ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7@1" = 220
```

Appendix 2: Variable selection rationale

World scale

The selection of environmental variables was based on the premise that *Heracleum mantegazzianum* is a mid-elevation mountain species within its native range. Consequently, we concentrated on significant climatic constraints typical of mountainous regions, specifically drought and cold.

We selected precipitation seasonality as a key variable, considering it a robust indicator of the harsh mountain climate, superior to high or low temperature alone. For the same reason, the number of annual growth degree days (ggd0) were used. Additionally, we included temperature ranges to capture the elevational gradient inhabited by many target species.

Recognizing that species distributions often lag behind environmental changes by years or even decades, we prioritized variables that summarized conditions over extended periods, such as annual or quarterly averages. We posited that these aggregated variables encapsulate the information provided by monthly averages.

Finally, the average and maximum surface wind temperatures were chosen because the natural way of dispersal of *H. mantegazzianum* is to have its propagules carried by the wind.

Europe scale

Since *H. mantegazzianum* is often spread by humans and cities may be a particular hotspot, as it was used as a decorative plant (EASIN, n.d.), the world population was used at European scale. (We could not use distance to infrastructure, as we thinned one of our datasets using that distance). In addition, because our Europe scale lies in the invaded range of *H. mantegazzianum*, and the species only spreads inside forests in its native range, but invades forest edges, the distance to forests was used as an indicator.

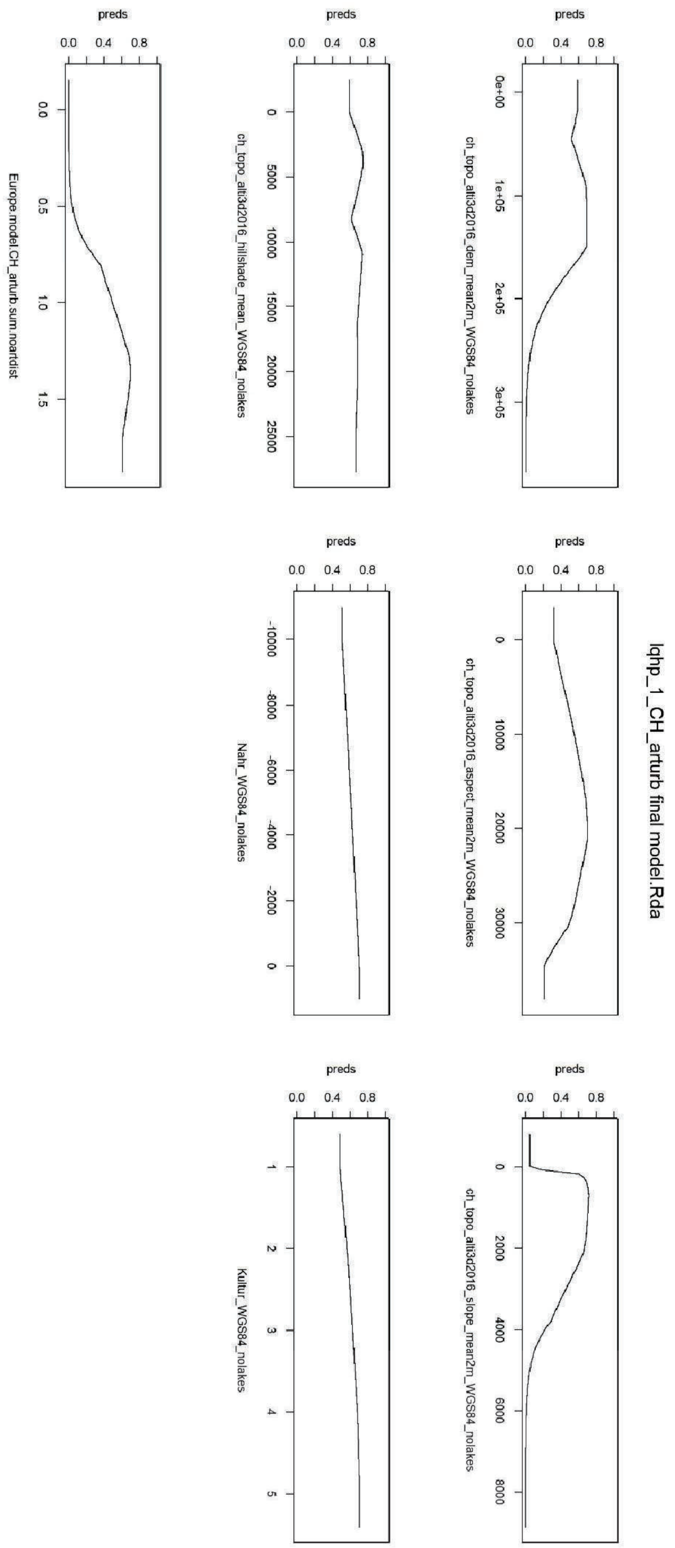
Moreover, distance to bare areas and to snow were used, since it is a mountain species in its home range. Importantly, crops were importantly used as a predictor as well, since the species grows very well on (abandoned) agricultural soil (Thiele & Otte 2008). The species also spreads a lot along water bodies, impeding fisheries, so distance to surface water was included in the roster.

Finally, we aimed to have a holistic view of vegetation on the European continent and therefore included also swampy or often flooded vegetation and grassland, scrubland and shrubland.

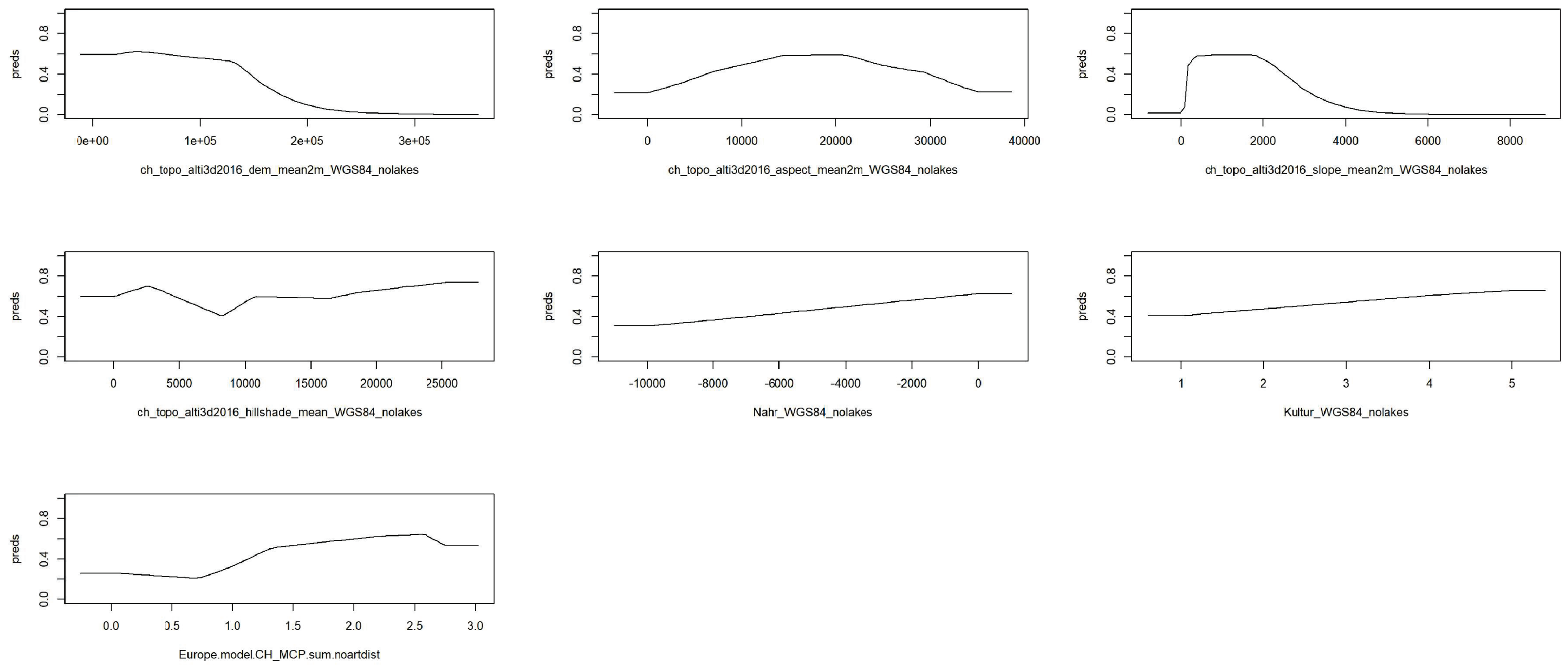
Swiss scale

The Swiss-scale data has a resolution of 25m. At that scale, the greatest determinants of a plant species' ability to thrive are light, food, water and temperature. The latter is always very strongly correlated with altitude, and therefore we used the Swiss Digital Elevation Model. Light is determined by aspect and shadow, so we included aspect, slope and hillshade variables. Nutrients were approximated by the ability of the ground to hold nutrients, and the potential of the soil to support crops. No proxy data was included for water at this scale, but no literature suggests that *H. mantegazzianum* is sensitive to drought or excessive wetness and since precipitation is included in the world-scale model, and soil water level could be influenced by slope, it is included indirectly through these variables.

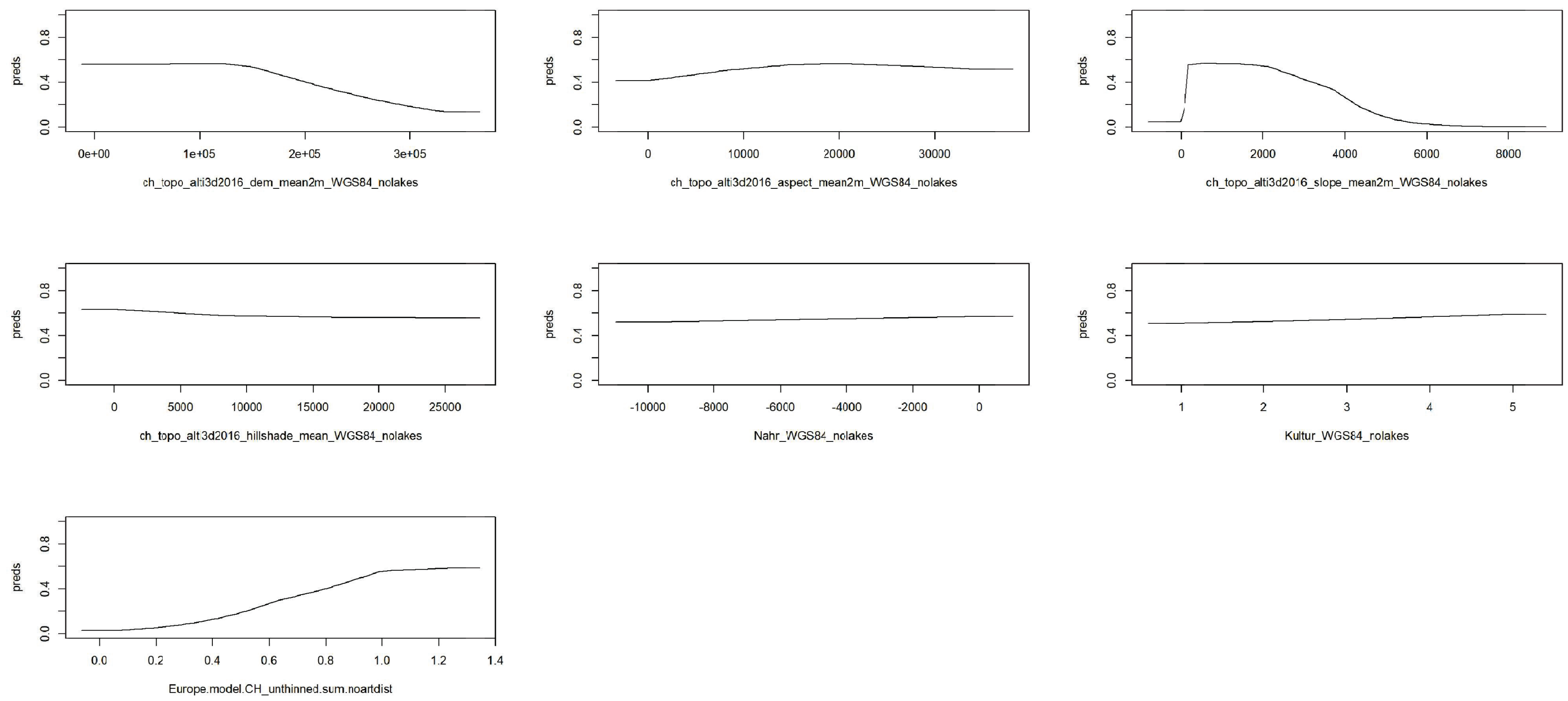
Appendix 3: Response curves of Maxent models



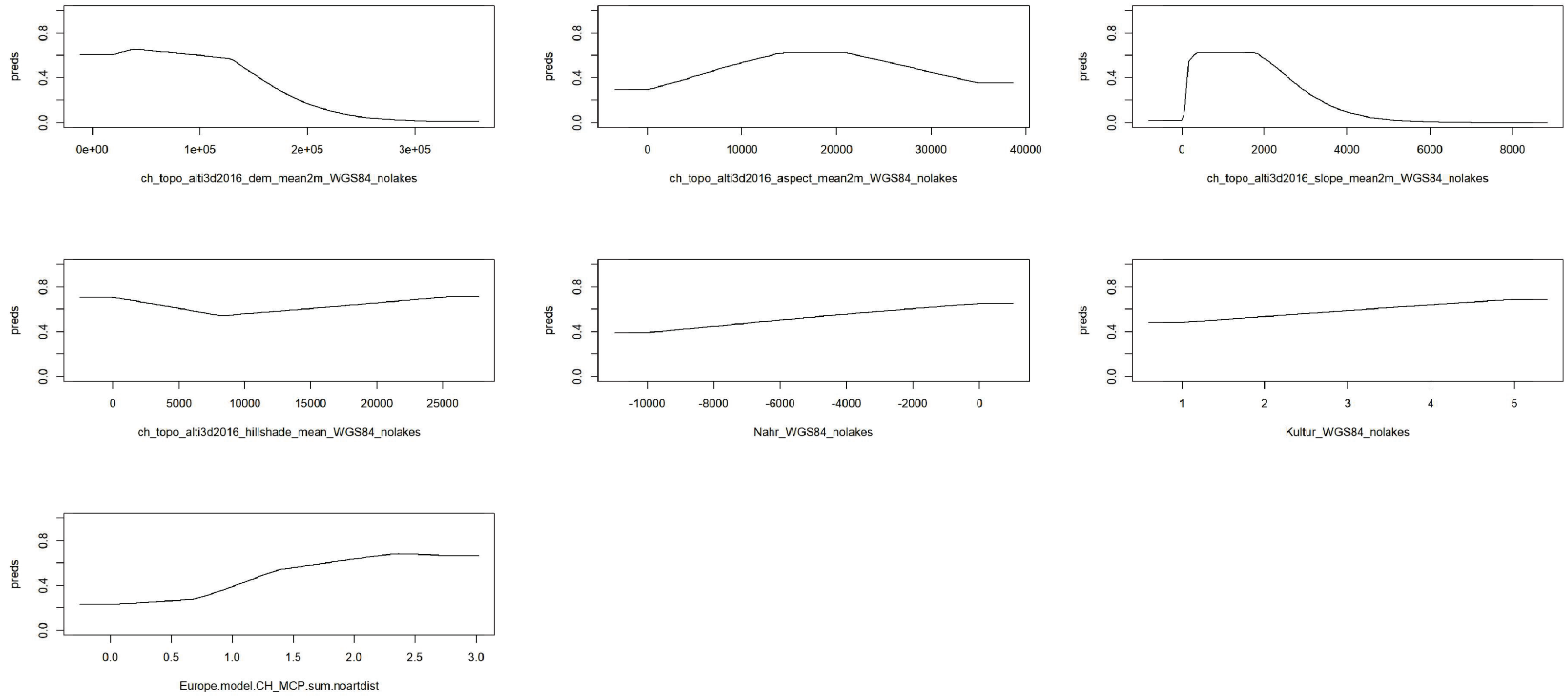
lqhp_1_CH_MCP final model.Rda



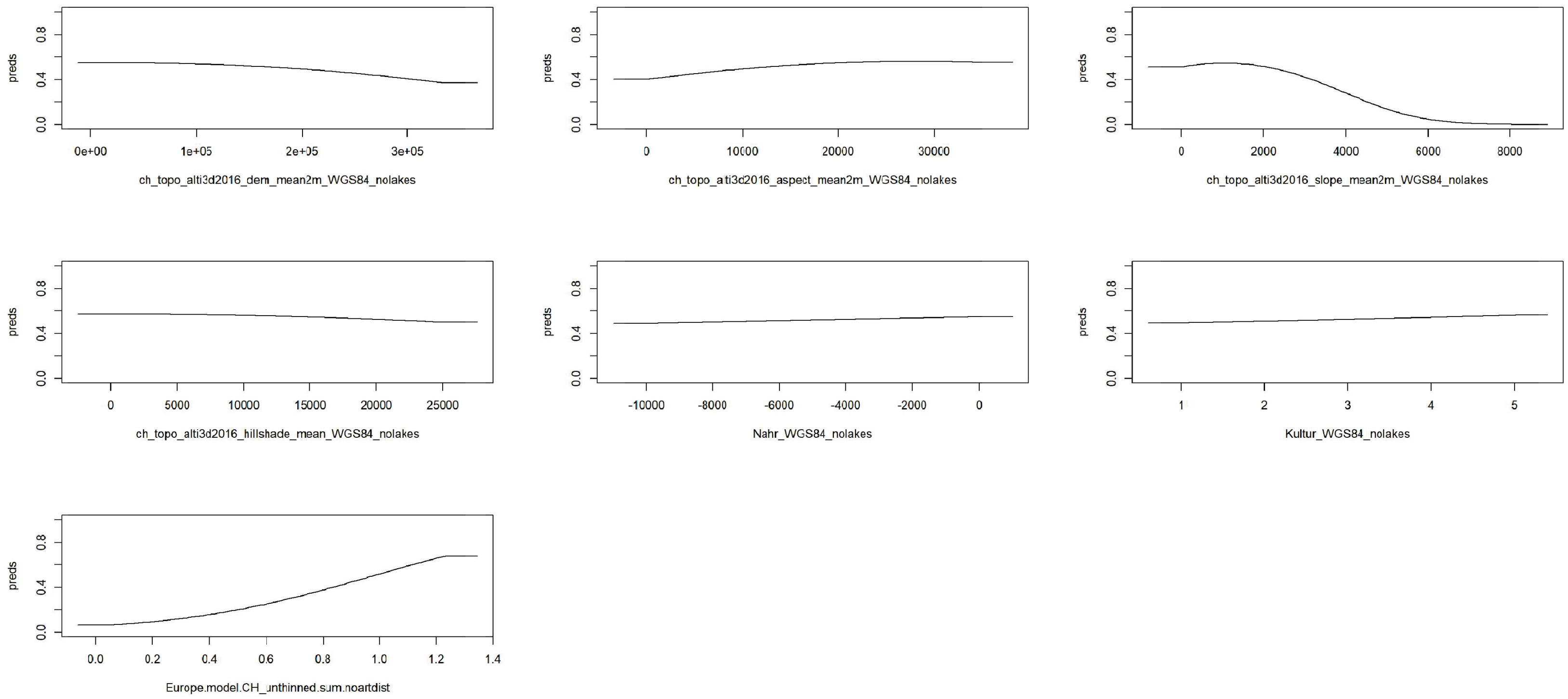
lqhp_1_CH_unthinned final model.Rda



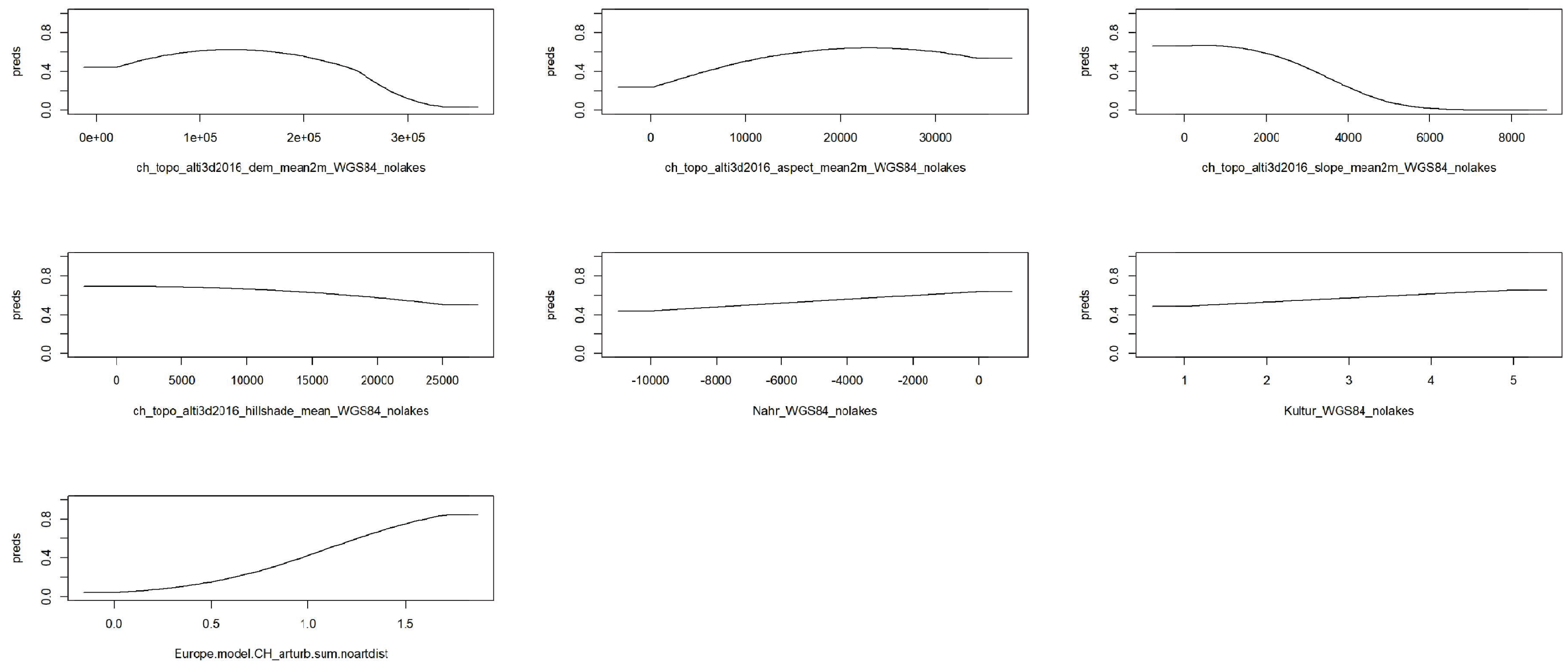
lqhp_2_CH_MCP final model.Rda



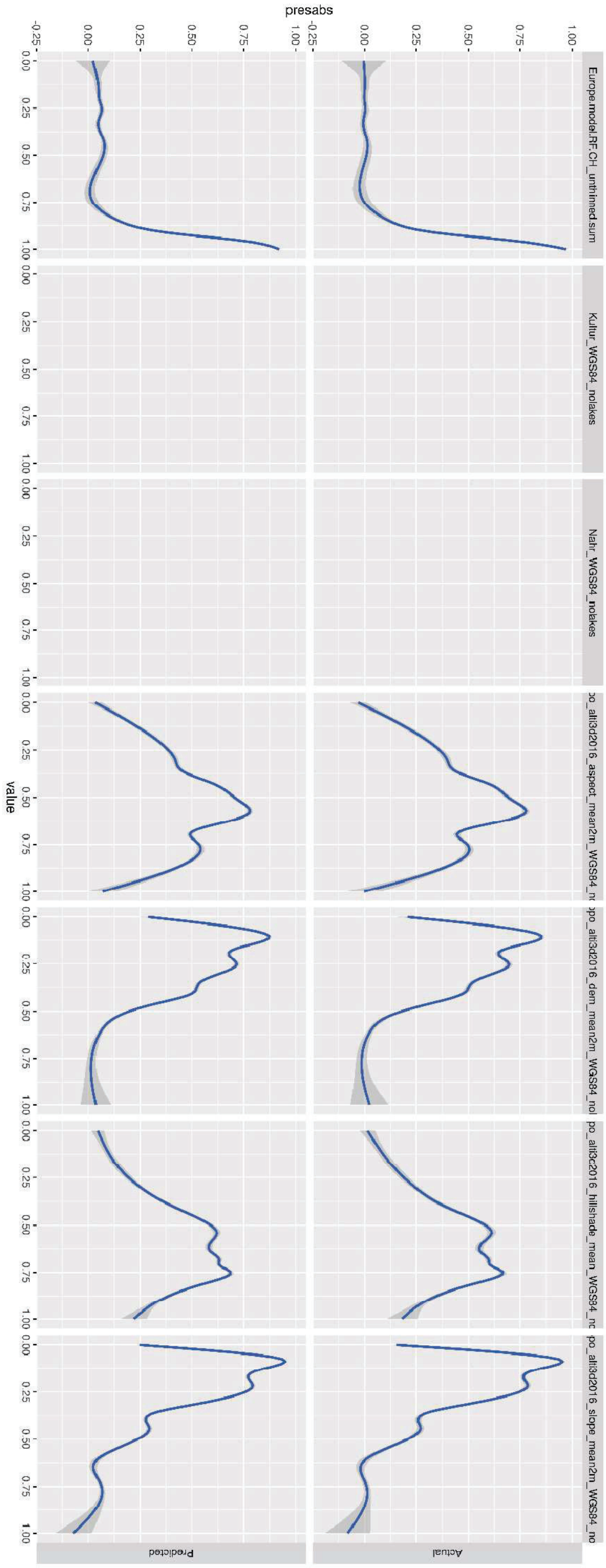
lqp_2_CH_unthinned final model.Rda

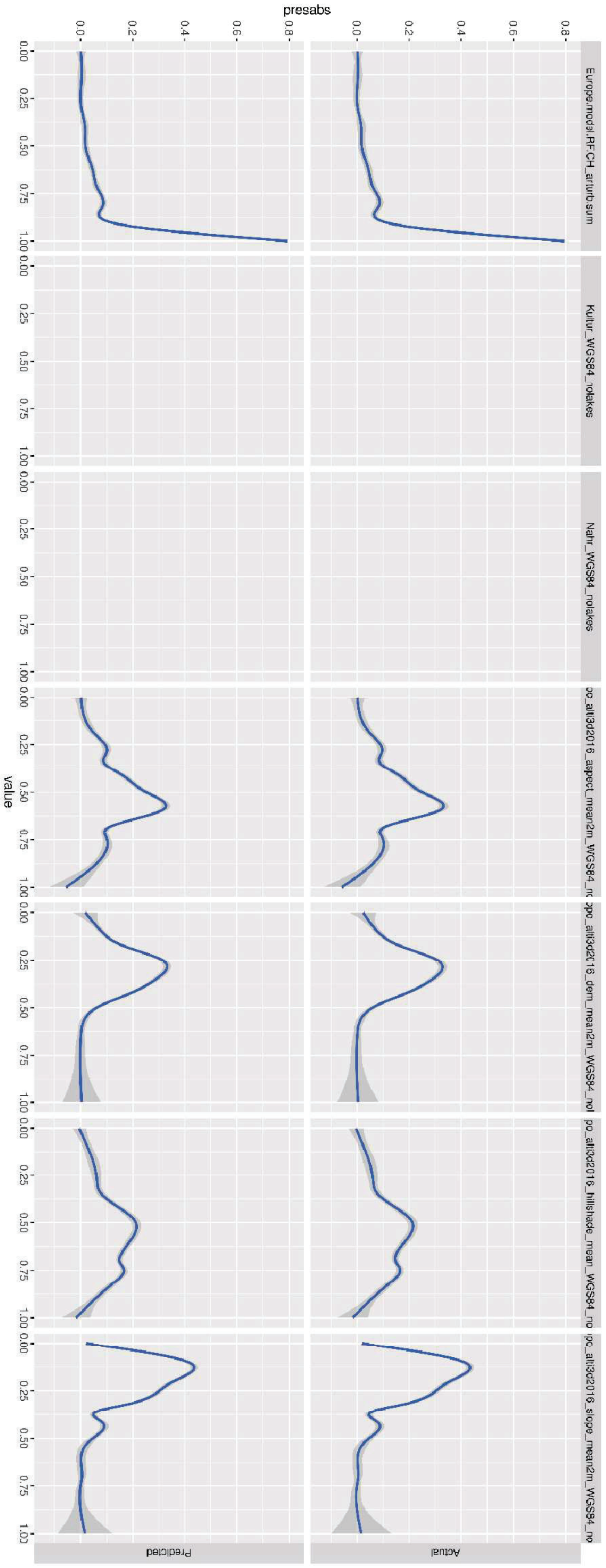


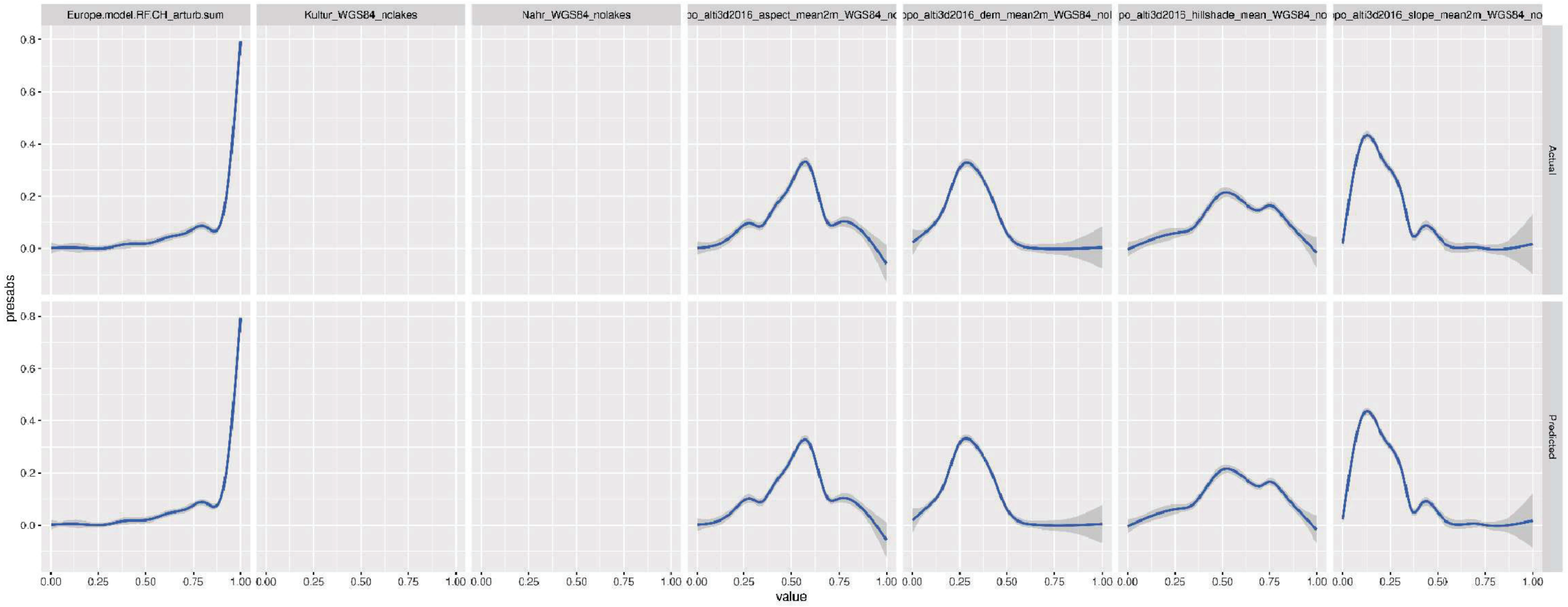
lqp_5_CH_arturb final model.Rda

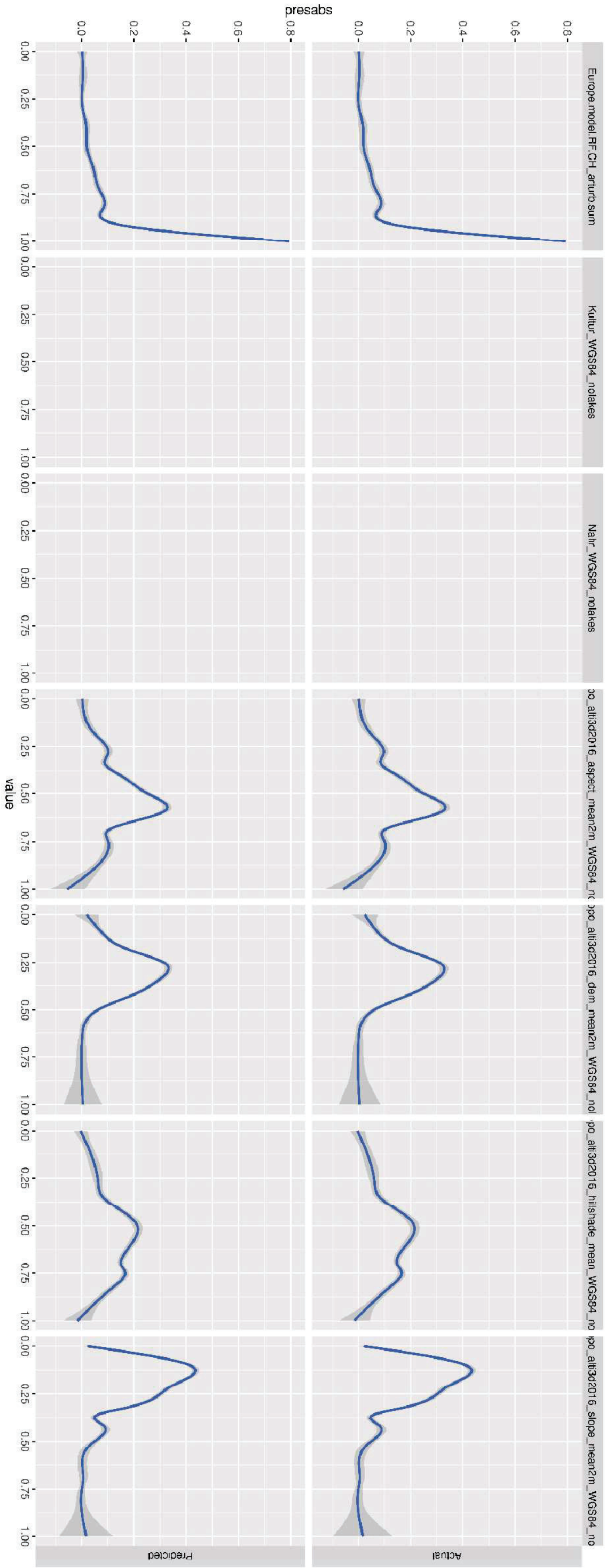


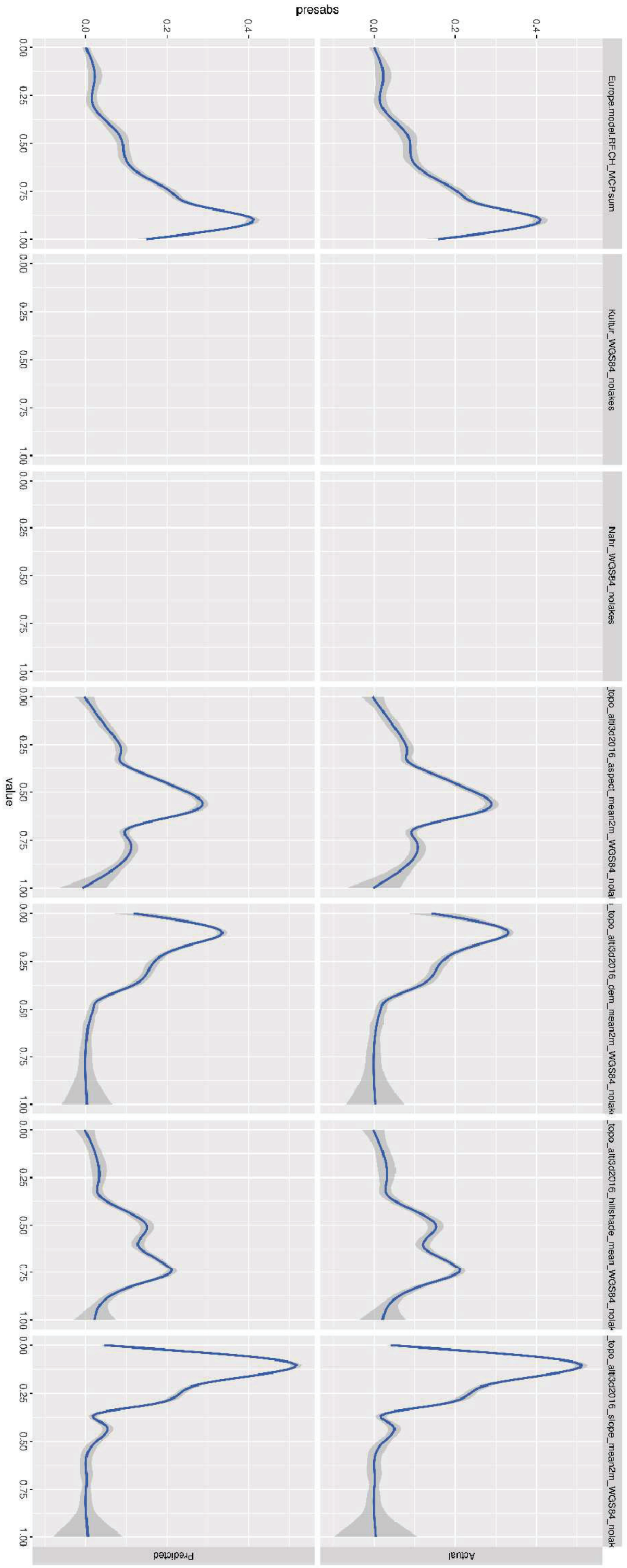
Appendix 4: Response curves of Random Forest models

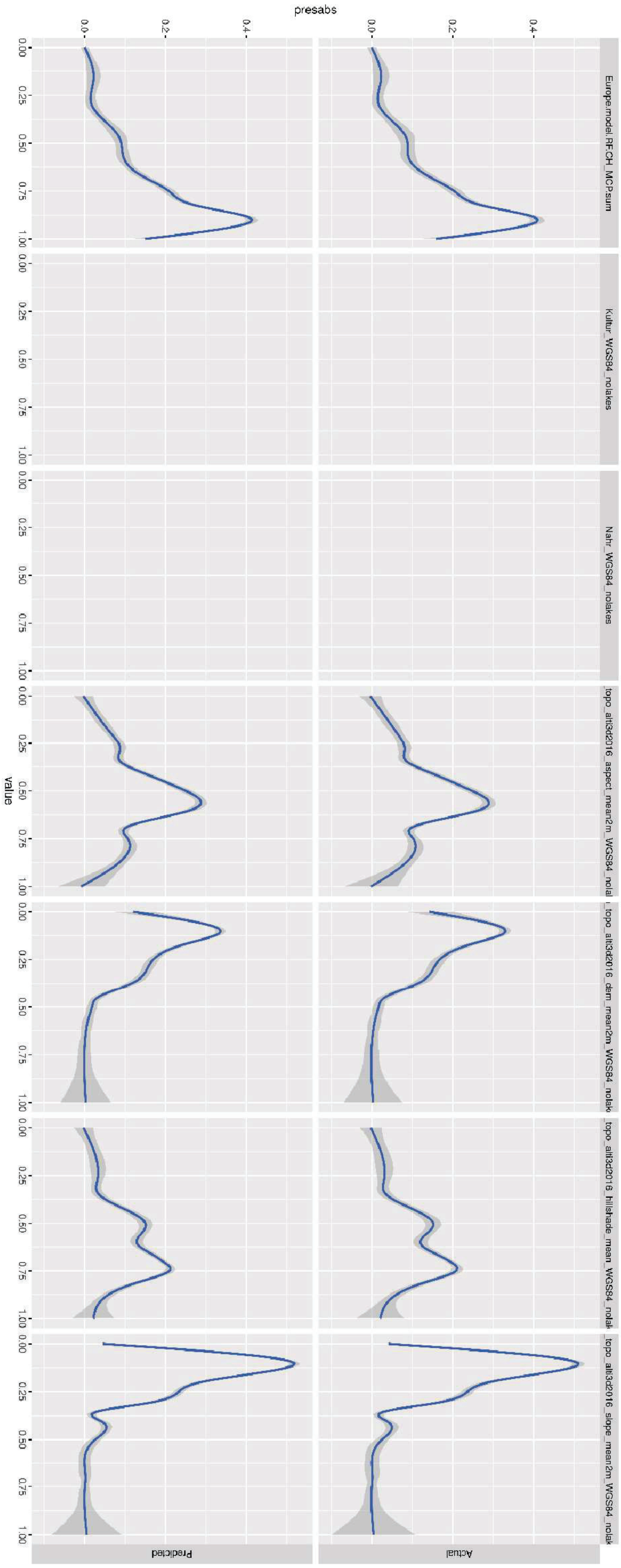


presabs
value
Europe.modal.RFCH_arturb.sum
Kultur_WGS84_nolakes
Nahr_WGS84_nolakes
po_alti3d2016_aspect_mean2m_WGS84_nc
po_alti3d2016_dem_mean2m_WGS84_nol
po_alti3d2016_hillshade_mean_WGS84_no
po_alti3d2016_slope_mean2m_WGS84_no
Actual
Predicted
0.0
0.2
0.4
0.6
0.8
0.00
0.25
0.50
0.75
1.00


Europe.model.RF.CH_arturb.sum
Kultur_WGS84_nolakes
Nahr_WGS84_nolakes
po_alti3d2016_aspect_mean2m_WGS84_no
po_alti3d2016_dem_mean2m_WGS84_nol
po_alti3d2016_hillshade_mean_WGS84_no
po_alti3d2016_slope_mean2m_WGS84_no
Actual
Predicted
presabs
0.0
0.2
0.4
0.6
0.8
value
0.00
0.25
0.50
0.75
1.00


presabs
value
Europe.model.RF.CH_arturo.sum
Kultur_WGS84_nolakes
Nahr_WGS84_nolakes
po_alti3d2016_aspect_mean2m_WGS84_nc
po_alti3d2016_dem_mean2m_WGS84_nol
po_alti3d2016_hillshade_mean_WGS84_no
po_alti3d2016_slope_mean2m_WGS84_no
Actual
Predicted
0.0
0.2
0.4
0.6
0.8
0.00
0.25
0.50
0.75
1.00


presabs
value
Actual
Predicted
0.0
0.2
0.4
0.00
0.25
0.50
0.75
1.00
Europe.model.RF.CH_MCPsum
Kultur_WGS84_nolakes
Natur_WGS84_nolakes
_topo_alti3d2016_aspect_mean2m_WGS84_nolal
_topo_alti3d2016_dem_mean2m_WGS84_nolak
_topo_alti3d2016_hillshade_mean_WGS84_nolak
_topo_alti3d2016_slope_mean2m_WGS84_nolak


presabs
0.0
0.2
0.4
Europe.model_RFCH_MCPsum
Kultur_WGS84_nolakes
Nahr_WGS84_nolakes
topo_alti3d2016_aspect_mean2m_WGS84_nolal
topo_alti3d2016_dem_mean2m_WGS84_nolak
topo_alti3d2016_hillshade_mean_WGS84_nolak
topo_alti3d2016_slope_mean2m_WGS84_nolak
Predicted
Actual
value
0.00
0.25
0.50
0.75
1.00

Appendix IV

Supplementary Material of Chapter IV

Appendix 1: Number of occurrences in the native range and invaded range for each species.

Species	**Native range**	**Intros**
Acacia dealbata	6759	7687
Acer negundo	4047	14812
Ageratina riparia	492	594
Ailanthus altissima	19726	18362
Allium triquetrum	6879	5358
Alternanthera philoxeroides	1280	456
Ambrosia artemisiifolia	17094	12701
Araujia sericifera	233	7108
Arbutus unedo	23883	1398
Arundo donax	6567	3260
Baccharia halimifolia	1282	3285
Buddleja davidii	275	71897
Calotropis procera	690	4328
Carpobrotus edulis	375	3282
Cenchrus setaceus	344	533
Cortaderia selloana	246	6080
Cotula coronopifolia	282	1514
Crassula helmsii	485	1223
Cylindropuntia fulgida	1166	319
Cylindropuntia imbricata	2187	900
Cylindropuntia prolifera	260	26
Echinocystis lobata	1238	1019
Erigeron annuus	83397	11152
Eschscholzia californica	2158	14679
Gypsophila paniculata	182	122
Helianthus tuberosus	245	3382
Heracleum mantegazzianum	1311	43734
Hyacinthoides hispanica	463	1269
Impatiens glandulifera	7255	136961
Impatiens parviflora	3878	26934
Leucaena leucocephala	115	3433
Ludwigia peploides	4589	1228
Lupinus polyphyllus	747	64177
Lysichiton americanus	1203	2424
Prunus serotina	4563	14037
Reynoutria japonica	30436	25670
Rhododendron ponticum	185	8333
Ricinus communis	1597	8443
Robinia pseudoacacia	1778	42185
Rumex confertus	488	27
Senecio inaequidens	928	200319
Solidago canadensis	1084	56904
Spartium junceum	4500	426
Tradescantia fluminensis	150	5081

Vachellia farnesiana	5330	34
Total	252372	837096

Acknowledgments

First, I give thanks to my two supervisors – Professor Antoine Guisan of the University of Lausanne and Professor Luigi Maiorano of La Sapienza, University of Rome. Their creative ideas, scientific expertise and endless willingness to help were invaluable in creating this contribution to science and nature.

I also thank my colleagues Olivier Broennimann and Antoine Adde for lending their expertise, ideas and the fruits of their own hard work to my publications. Further thanks to Antoine Adde, Pierre-Louis Rey and Nora Khelidj, who had the misfortune of being my office mates – though I don't think I bothered you too much.

Special thanks also to Martin Schlaepfer for postulating the idea that gave rise to chapter 1 and to Laura Celesti-Grapow for providing the occurrence dataset in chapter 3. I thank both also for their input during the writing process.

In addition, thanks to Prof. Anouschka Hof, Prof. Sylvia Haider, Dr. Ana-Cristina Cardoso, Prof. Gianalberto Losapio and Prof. Moreno Di Marco for agreeing to act as experts during one or both of my defenses. I thank Prof. Haider also for her input during my mid-term evaluations, for which I must be grateful to Prof. Franz Essl, as well.

I likewise give thanks to the Swiss Confederation and to La Sapienza University of Rome for awarding the doctoral grants that made this research possible. My thanks as well to the UNIL Institute of Earth Surface Dynamics of the University of Lausanne for providing additional financial support, especially Prof. Georgina King, Prof. Stuart Lane and Prof. Antoine Guisan, whose considerable and selfless efforts have significantly improved the financial situation of several PhD students at the institute, including me. I also thank the members of Swiss national Valpar.CH project for additional support and for organizing the very enjoyable sorties. In addition, I thank the University of Lausanne for granting access to their High Performance Computing (HPC) cluster. Furthermore, I thank all our anonymous reviewers for their swift responses, thoroughness, and insights, which have allowed us to greatly improve the scientific work presented here.

Finally, massive thanks to all members of the Ecospat and Biodiversity Change groups at UNIL and the Maiorano lab at Sapienza for the kindness, hospitality, generosity, laughter and scientific discussions. Without you, I would not have enjoyed doing this thesis nearly as much or explored so many different avenues of research.

I already could not stop myself from making some jokes above but below comes an even less formal part. Thanks so much to Dr. Adde, for massively improving my physical condition by taking me on all those grueling, but fulfilling, runs. Thanks also to Wolfgang, Pauline and Salome, for joining me on language lunch and teaching me better German and French. My thanks likewise to Samuel, for the quizzes and the Britishness. Thanks also to Alice and Giulia for the calm, relaxing coffee meetings and to Milton for sharing such culinary and occasionally alcoholic delights. Finally, my thanks to the Ecospat and IDYST members for organizing the retreats that make each year much more interesting.