

The role of spatial dependency in niche shifts of invasive species

Isak Lerum

Main supervisor: Adam Klimes

Co-supervisors: Alistair William Robin Seddon, Joachim Paul Töpper
& Joseph Chipperfield

Master of Science in Biology – Biodiversity, Evolution and Ecology



Department of Biological Sciences

University of Bergen

September 2023

Abstract

Invasive species are an ongoing threat to nature and human interests. Species distribution models are used to predict the current and potential distribution of invasive species, but doing so has proved difficult. One possible cause of this difficulty are niche shifts, a change in the realised or fundamental niche between native and introduced ranges. Niche shifts have mostly been studied by approximating the realised niche, which often underestimates the true ecological constraints of a species as a result of biotic interactions and dispersal mechanisms. This study employs a new model that is more theoretically suitable to approximate the fundamental niche. The model features a spatial random effect that accounts for spatial processes in the data, distinguishing them from the environmental effects (fundamental niche). I find that in the datasets of previous niche shift studies the spatial effects dominate the environmental ones to such a degree that the fundamental niche cannot be reliably approximated. This is the case for both the native and introduced ranges for the invasive species, indicating that spatial effect domination is the norm for species in general, not just invasive species in their introduced ranges. Even if the fundamental niche could be estimated from some other method than species distribution modelling it would still have limited usefulness in predicting the species' distribution. The fundamental niche could still be used to predict the distribution of potentially suitable habitat, which is useful for selecting areas to establish preventative measures. But due to the strong spatial effects it will not produce reliable predictions of the distribution of the invasive species for selecting where restorative measures (eradication/culling of the invasive or support for harmed native species).

Acknowledgements

First, I'd like to thank Joseph Chipperfield, who developed the model used in this study. This thesis would have been impossible without his intricate knowledge on statistics and coding.

I'd like to thank Alistair William Robin Seddon, Joachim Paul Töpper whose guidance in writing and storytelling transformed my jumbled mess of thoughts into the document you're about to read. Their stark realism also ensured that this thesis finished on time.

I'd like to thank my family and Randi Hushagen from Sammen who supported me during the most stressful time of my life. I would've never been able to finish this thesis without their care.

Most of all, I'd like to thank main supervisor Adam Klimes, who has been with me through every step of the way. Guided me through every issue, no matter how complex or simple. In matters of literature, theory, coding & writing he has provided guidance through it all. Truly, thank you.

Table of Contents

Abstract	2
1. Introduction	4
1.1 Niche theory	5
1.2 Niche Shifts	5
1.3 Modelling species distributions	7
1.4 A spatial random effect model	9
1.5 Research questions	9
2. Methods	10
2.1 Literature search	10
2.2 Data importation & preparation	10
2.3 The spatial random effect model	13
2.4 Niche overlap analysis	13
2.5 Dataset & model assessment	17
3. Results	19
4. Discussion	22
4.1 Dataset assessments	23
4.2 Model assessments	23
4.3 Other considerations	26
5. Conclusion	26
6. References	27
Manual references	Error! Bookmark not defined.
7. Supplemental material	32
a) Angetter et al. (2011)	32
b) Battini et al. (2019)	37
c) Bidinger et al. (2012)	41
d) Manzoor et al. (2020)	45
e) Morehouse and Tobler (2013)	49
f) Rödder and Lötters (2009)	53
g) Rödder and Lötters (2010)	59
h) Zhang et al. (2022)	64
i) Ørsted and Ørsted (2019)	68

1. Introduction

The world is changing rapidly due to climate change and direct human activity which is leading to an increase in biological invasions (Bertelsmeier et al., 2013, Keller et al., 2011, Pyšek et al., 2010). Invasive species are generally considered to be species introduced to an area where they have no history of occurrence and that can sustain a population and expand beyond the point of introduction (Richardson et al., 2000). Some definitions additionally require invasive species to cause damage to the native ecosystem and/or human interests (Urban, 2020). An invasive species can harm the native ecosystem by directly harming other species (Medina et al., 2011), outcompeting them (Bøhn et al., 2008) or altering the environment (Barrios-Garcia and Ballari, 2012). An invasive species can harm human interests in the form of economic damage (Paini et al., 2016), harm to human health (Gucker & Corey, 2009) or damage to cultural values (Massengil, 2011). The total global damage caused by invasive species from 1960 to 2022 is estimated around US\$ 1130.6 billion (Cuthbert et al., 2022)

Many measures are taken to detect invasive species, mitigate their damage and spread, and to eradicate them where possible, but these efforts are complicated by a lack of knowledge of their ecology, climatic preferences, and their potential for spreading. A core effort with prevention work is to predict what species could become invasive and where they could invade. One method of predicting where a species could thrive outside its native geographical area (further as **native range**), is the use of **species distribution models (SDM)** (Guisan and Zimmermann, 2000). An SDM is a statistical approach estimating the relationship between a species' occurrences and the conditions in which it lives, which can be used to predict the species occurrence in unmapped areas, or for invasives in new locations.

For the purposes of prevention and mitigation, an SDM should be used as early in the invasion process as possible, but the earlier it is used the less data of occurrences are available for the invaded area. To bypass this issue, scientists may use the data from the species native range in the SDM and project the predictions onto new locations (further as **introduced range**) as seen in Peterson (2003) where the native (Southeast Asia) predictions were projected onto North America. Such projected predictions assume that species occurrences across the native and introduced ranges maintain the same relationships with the conditions they live in.

The relationship between species occurrences and the environment can change across space or time which is called a **niche shift** (Bates and Bertelsmeier, 2021). A niche shift may occur because of changes in species interactions or geographic barriers (a realised niche shift) or evolutionary processes that via physiology change the species' relationship with an environmental condition (a fundamental niche shift) (Bates and Bertelsmeier, 2021). While realised niche shifts are a widely accepted phenomenon in invasion biology, it is disputed how common they are (Petitpierre et al., 2012, Liu et al., 2020, Atwater and Barney, 2021). There is more disagreement about the existence of fundamental niche shifts in invasive species.

To explore the possibility of fundamental niche shifts in invasive species I will be using an SDM with a spatial random effect. This model should be able to disentangle the effects of species interactions and geographic barriers from environmental effects.

1.1 Niche theory

All species live with inherent biological constraints: A plant species has a minimum temperature it can survive in before its tissue freezes and it gets killed (Pocheville, 2015). Some reptiles depend on certain temperatures to determine the sex of their embryos, too hot and it will only result in female offspring (Shine, 1999). Many other factors, such as precipitation, humidity, and length of the day, contribute to what lets a species not only survive but also thrive and reproduce. The collective structure of these relationships is a niche.

In biogeography, we generally utilise the hutchinsonian niche theory (Evelyn Hutchinson, 1991). There, a niche is defined in environmental space, with each axis representing an environmental gradient e.g., annual precipitation or soil pH. There are 2 kinds of niches: the fundamental niche & the realised niche. A niche can be considered for an entire species or populations of the species.

The **fundamental niche** describes the environmental conditions under which a species can sustain a population indefinitely (Pocheville, 2015). For each climatic variable the species has a response curve, indicating how suitable each point along that climatic gradient is for the species. These response curves are typically unimodal in shape and interdependent. A species suitability in relation to one climatic variable (e.g., soil pH) can be affected by another variable (e.g., precipitation).

The **realised niche** is a version of the fundamental niche modified by biotic interactions (Peterson, 2011). In invasion biology the realised niche includes the additional restrictions imposed by geographic barriers to dispersal e.g., bodies of water for land-dwelling species (Rödder and Engler, 2011). The realised niche can be limited by exploitative interactions where one species exploits the other (such as parasitism or predator & prey) or by competitive interactions where both species are negatively affected by the competition over shared resources (Evelyn Hutchinson, 1991). The realised niche can also be expanded beyond the fundamental niche by positive interactions where one (commensalism) or both (mutualism) species benefit. These positive interactions can facilitate a species under otherwise less suitable conditions (Bruno et al., 2003).

1.2 Niche Shifts

The fundamental niche can only shift due to a change in internal conditions, that is evolution altering the climatic preferences/tolerances of the group (Bates and Bertelsmeier, 2021). The fundamental niche might be limited by a loss of alleles due to the founder effect when a species invades a new area, but the chance of the niche reverting to its original size increases with more introduction events improving genetic diversity. The fundamental niche can shift or expand due to several factors such as adaptation, polyploidisation or hybridisation with local species (Richardson and Pyšek, 2006, Sakai et al., 2001). A shift in the fundamental niche can also cause a shift in the realised niche as new pathways around dispersal barriers or the distributions of limiting species open through previously unsuitable climate or through alteration of biotic relationships.

The realised niche can shift due to a change in external or internal conditions. Changes in external conditions can be a change in the distribution of other species that define the limits of the niche or the establishment/destruction of a dispersal barrier (Bates and Bertelsmeier, 2021). A change in internal conditions that lead to a realised niche shift can be an increase in abundance in populations

abundance/density caused by more introduction events or positive population growth rate. The higher abundance and/or density can overcome or alter negative biotic interaction, increase the likelihood of some individuals crossing a dispersal barrier by chance (Richardson and Pyšek, 2006). Some species undergo a change in behaviour/physiology at higher population densities that improves the invasives ability to overcome dispersal barriers and biotic interactions e.g., locusts entering their swarming phase at higher populations densities (Topaz et al., 2012). Internal change would be evolution that alters the species biotic interactions or dispersal ability e.g., the invasive cane toad in Australia evolving longer legs and skeletal structure that aid in dispersal (Hudson et al., 2020).

A shift in the realised niche between a native and invasive population of the same species is to be expected. After all, a new geographic area (with presumably a different biological community) would be unlikely to present the same dispersal barriers and biotic interactions as the native range. Previous niche shift literature has been limited to studying realised niche shifts (Petitpierre et al., 2012, Atwater et al., 2017, Liu et al., 2020) and while mentioning the possibility of evolution they have not had the tools available to distinguish between the effects of biotic interactions and dispersal barriers from evolution.

As stated earlier, the realised niche shift as a phenomenon is widely accepted and studied (Petitpierre et al., 2012, Liu et al., 2020, Atwater and Barney, 2021). But partly because of the lack of tools to study it, the existence of fundamental niche shifts is experiencing more scepticism.

Proponents of fundamental niche shifts argue that introduction into a novel area can put the invasive under high selective pressure from the native environment and community (Novak, 2007). Genetic diversity (and possibly the breadth of suitable climate) will initially be reduced in the introduced range compared to the native range due to the founder effect, which in turn can lead to stronger selection pressure and local adaptation (Santos et al., 2012). However, over time as more introduction events occur genetic diversity will increase to match the diversity of the native range (Moran and Alexander, 2014). In fact, it may even become greater as alleles endemic to different locales of the native range are gathered in the introduced range and create novel genotypes (Novak, 2007). Invasives may also gain genetic diversity due to gene duplication (only plants) or hybridisation with local species. This increased genetic diversity allows for rapid adaptation to the selection pressures of the introduced range (Prentis et al., 2008). In the case of Chinook salmon introduced to New Zealand, it took only 26 generations for it to become locally adapted (Kinnison et al., 2008).

Those that are sceptical towards the prospect of fundamental niche shifts argue for niche conservatism. The time scale of fundamental niche shifts is a source of concern for sceptics. They find that niches are largely conserved for tens or hundreds of thousands of years (Peterson, 2011) while most invasive species have a history of invasion spanning from decades to centuries (McGrannachan et al., 2021). Sceptics also argue against fundamental niche shifts by stating that the case studies displaying niche shifts are unable to disentangle the effects of species interactions and geographic barriers (a shift limited to the realised niche) from an evolutionary event altering the species climatic tolerances/preferences in the introduced range (a fundamental niche shift) (Moran and Alexander, 2014).

1.3 Modelling species distributions

Species Distribution Models are statistical models that aim to predict the distribution of a taxon/species/group. The models take occurrence data of the target group and spatially mapped environmental predictors (usually climate variables such as annual precipitation but could include other predictors such as land use) in a specified geographic area, hereafter referred to as the **study area**, as input. The model estimates the relationships between occurrences and predictors in the form of **response curves**. The response curves can be combined to map the suitability in multidimensional climate space, hereafter referred to as the **suitability estimate**. The suitability estimate can then be reapplied to the study area to produce a geographical map of the environmental suitability of each point within the study area, hereafter called the **habitat estimate**. The higher the suitability the higher the likelihood that the group could thrive there.

The response curves calculated in an SDM can be used as an estimate of the group's niche (at least along the environmental gradients used in the model). If two models use the same environmental gradients then the response curves can be used for estimation of niche shift, approximating the similarity of the niches between the modelled groups (Lauzeral et al., 2011).

Commonly-applied SDMs rely on some key assumptions (Elith and Leathwick, 2009):

1. SDM's assume that the group being modelled is in equilibrium with its environment.
2. The occurrence and environmental data must be representative of the true distribution of occurrences and environment within the study area.
3. SDMs assume that the group retains the same niche across space. Let's say that an SDM is using the entire planet as the study area and the modelled group is a species present in North America and Europe. Individuals in North America have a different niche than individuals in Europe e.g., higher tolerance to low temperatures. The SDM derives the suitability from all members of the species resulting in a broader suitability estimate than one based only on individuals from Europe. The broader suitability estimate overpredict suitability in the colder parts of Europe because it assumes European individuals have as temperature tolerance as American ones.
4. The variables selected for the model needs to be the major factors limiting the group's distribution. If the major limiting factors are not present in the model, then the model will only be able to explain a limited amount of the underlying mechanisms determining the distribution.

SDMs on invasive species in their introduced ranges are particularly vulnerable to violations of assumption 1 (Gallien et al., 2012) and 2.

An invasive species currently undergoing an invasion process, not occupying all suitable area they are capable of dispersing to, will not be in equilibrium with its environment. This issue leads to an underprediction of the suitability and habitat estimates. This is likely to be the case in cases such as Battini et al. (2019) where the modelled species (*Pleurobranchaea maculata*) was first detected in its introduced range in 2009 (Farias et al., 2015) and has been known to expand its' range at a rate of ca. 330 km yr⁻¹ (Farias et al., 2016).

Invasives in their introduced range are also prone to having fewer and biased occurrence records. To effectively manage and mitigate an invasion one would want the model as early as possible after an invasion has begun, but the earlier it is run the less time to collect occurrence data. Once again Battini et al. (2019) serves as a good example. As the species history in the introduced range is so short only 12 occurrences were recorded in the Battini et al. (2019) dataset.

To bypass the issues with modelling invasive species in their introduced ranges, many have tried to model the species based on their native ranges and then transfer the estimated suitability onto their introduced ranges to generate the habitat estimate there (Peterson, 2003, Bates and Bertelsmeier, 2021). But this method comes with its own potential issues.

First, the introduced range may feature environmental conditions that are not present in the native range. The suitability estimate must then be extrapolated beyond its initial parameters, causing higher uncertainty in the habitat estimate in the introduced range.

Second, there is the possibility of a large niche shift between the native and introduced ranges. A smaller niche shift may not impact on the estimated distribution, but a larger niche shift will cause any prediction projected from one range onto the other to be inaccurate, making the model non-transferrable.

Many have attempted to determine whether niche shifts are present in invasive species, but the methodology (in terms of modelling approach and consequent calculations) of these studies have made it difficult to reach a conclusion on the commonness and nature of niche shifts.

First, most studies used the realised niche to assess niche shift (Petitpierre et al., 2012, Atwater et al., 2017, Liu et al., 2020), in which case it is impossible to determine whether the shift was caused by an evolutionary event for the species or simply a difference in geographic barriers and local community between native and introduced ranges.

Second, the calculations of niche overlap is inconsistent across the literature (Manzoor et al., 2020, Rödder and Lötters, 2010, Morehouse and Tobler, 2013) making it hard to generalise a trend across studies. Generally niche overlap analysis is performed on 1 of 3 data types: spatial predictions from an SDM, response curves derived from an SDM or a principal component analysis of the environment and occurrences in both study ranges (Guisan et al., 2014). The niche overlap is then calculated using the niche overlap index Schoener's D (Schoener, 1968) or its modified version Warren's I (Warren et al., 2008).

As of yet, most niche shift studies have been performed on the realised niche (Petitpierre et al., 2012, Atwater et al., 2017, Liu et al., 2020), but I'd like to argue that on the issue of transferability fundamental niche predictions would stand a better chance than realised niche predictions. Given that realised niches are expected to shift due to differing dispersal barriers and biotic interactions between the native and introduced ranges, they are unsuitable for the goal of accurately mapping the potential habitat of an invasive species. Fundamental niche predictions would not be as capable of predicting the invasive's current distribution, but an accurate map of suitable habitats would still be useful for management purposes. The only issue for SDMs that needs to be resolved is to distinguish between realised and fundamental niche, i.e., the spatial from the environmental effects on the occurrences.

In this study I will use a new model that can approximate the fundamental niche to investigate the rate and degree of niche shifts in datasets where realised niche shifts have been claimed.

1.4 A spatial random effect model

The model used in this study, hereafter referred to as the **SRE (spatial random effect) model**, was constructed with the intention of providing a toolset to separate the effect of the abiotic and biotic processes on determining a species' range. This, by extension, allows for the estimation both the actual distribution of the species (**distribution estimate**) and the extent of the species' suitable habitat (**habitat estimate**). In effect, this therefore allows for the separation of realised and fundamental niche processes governing the range of the species.

The model used in this study is based on the framework of an **Integrated Nested Laplace Approximation (INLA)** model (Bates and Bertelsmeier, 2021).

SREs have been used in SDMs before (Redding et al., 2017) but are not common. An SRE accounts for the spatial structure in the data and can be used to distinguish the effects of biotic interactions and dispersal barriers from the effects of the environment on the occurrences. As the SRE is estimated for the study area, the distribution estimate which is the combination of the suitability estimate and spatial effects, is not transferable outside the study area.

The SRE model allows for comparison of the fundamental niches between native and invasive study areas, which has previously been unavailable through the use of SDMs. However, since the spatial and environmental effects are kept separate other than in the distribution estimate, it cannot be used univariate comparison of the realised niche.

The model was developed as part of ECoMAP (<https://betweenthefjords.w.uib.no/ecomap>), a project that aims to inform the Norwegian state/public about the distribution and potential habitat for terrestrial species in Norway. ECoMAP intends to achieve this goal by creating a model that can be automatically run all species in national databases such as Artsdatabanken (<https://artsdatabanken.no>) and make the model results freely available for public use. The ECoMAP model is still under development and the model used in this study is based on the version from autumn 2022. Invasive species are particularly difficult to model and part of this study is to stress test the model, seeing what work and what limitations the model has.

1.5 Research questions

In this study I will use the SRE model to replicate previous case studies and compare niche overlaps with the previous studies.

The research questions this study aims to address are as follows:

1. Will the model with a spatial random effect, when using the same datasets as previous case studies, produce fundamental niche overlaps of similar magnitude to the realised niche shifts shown in previous case studies?
2. Does the model produce reliable niche approximations? A set of quality assessments will be made regarding the quality of the model results and the datasets used.

Should assessments for the model be poor then it will call into question the resulting overlap measures and the reliability for practical use (predicting actual species distributions and potential habitats in extrapolated conditions).

Because the datasets used will be replicated from previous studies the quality of the datasets matters less when comparing with the overlap measures of the previous studies, but still puts into question the actual value of the niche overlaps.

Should certain assessment criteria be poor across multiple datasets it may help identify limitations with the model and potential for improvement.

My initial expectations are that biotic interactions and dispersal barriers varying between native and invasive study ranges will cause the realised niche overlaps of the previous case studies to be lower than the fundamental overlap using the SRE model. I expect the fundamental niche overlap to be high (>60%) because the niche shift is measured between groups of the same species.

2. Methods

2.1 Literature search

First, I conducted a literature search for case studies where niche shifts are claimed and that could be replicated using the SRE model. I used the search engine **Web of Science**, with the keywords: **“Invasive”, “Niche shift” and “Species distribution model”**.

From there 5 criteria were used to select papers:

1. Only case studies using single species.
2. The native or invasive populations being modelled could not cover more than one continent to reduce modelling workload.
3. Occurrence and climate data had to be accessible through free online databases or included in the paper’s supplementary material.
4. The studies had to only model 1 primary native and introduced study areas each. A study may have secondary, smaller study areas that are modelled separately, but only the primary study areas will be used for this study.
5. The study ranges must appear to adhere to assumption 3 of SDMs (retaining niche across space). The native and invasive study ranges must look more likely to retain their niche within study ranges than between study ranges. Say a study has a native study area of Australia and an invasive study area of west-pacific islands spanning from New Zealand to Japan. The invasive study area does not appear at face value to be more likely to retain its fundamental niche across the entire study area than parts of it (New Zealand) matching with the native study area.

From this literature search 9 studies were selected: Angetter et al. (2011), Battini et al. (2019), Bidinger et al. (2012), Manzoor et al. (2020), Morehouse and Tobler (2013), Rödder and Lötters (2009), Rödder and Lötters (2010), Zhang et al. (2022) & Ørsted and Ørsted (2019)

2.2 Data importation & preparation

All datasets, except Battini et al. (2019), imported climate data from Worldclim (Fick and Hijmans, 2017). Battini, a marine dataset, used Bio-Oracle (Assis et al., 2018) through the r package “sdmpredictors” (Bosch S, Fernandez S, 2022). I imported the climate data from the same sources, selecting the same variables (See Table 1).

The source for occurrence data varied between studies. Some studies provided the occurrence data directly in supplemental material (Bidinger et al., 2012, Battini et al., 2019) or archived their dataset using the datadryad (<https://datadryad.org/stash>) service (Ørsted and Ørsted, 2019). Angetter et al. (2011), Manzoor et al. (2020), Rödder and Lötters (2009), Rödder and Lötters (2010) and Zhang et al. (2022) used the Global Biodiversity Information Facility, **GBIF** (<https://www.gbif.org>), as their primary source of occurrence data. GBIF provides a reference code for any exported dataset to aid in reproducibility, but none of the studies provided the reference code. I attempted to replicate the GBIF datasets based on only including occurrences up to the year of the studies’ publication and approximated the geographic extent of the study areas by visual inspection of figures of occurrences

in the previous studies. I have included the reference code for all replicated GBIF datasets in the supplementary material.

The Angetter et al. (2011), Rödder and Lötters (2009), Rödder and Lötters (2010) studies also utilised the VertNet database (David Bloom, 2016, <https://vertnet.org/resources/datatoolscode.html>) as a supplement to Gbif. In addition, these studies supplemented their datasets with additional occurrence records from “published references”. These published references consisted mostly of reports from *Herpetological review*, featuring 1-5 sightings with addresses but no coordinates. Other references were mostly books that were unavailable. Since these additional sources provided for only a small portion of the occurrence records (hence called supplemental by the original authors), I decided to only use the GBIF and VertNet data for replicating these studies.

Morehouse and Tobler (2009) used 2 databases: the U.S. Geological Survey (2023) and the Illinois Natural History Survey (2023). Unfortunately, since the publication the study species has undergone a taxonomic reclassification (Crandall and Grave, 2017) and been unified with another species. These databases have adopted this reclassification and it was not possible to discern what occurrences belonged to the old target species. I downloaded the datasets as is, accepting that the replicated dataset would contain additional records.

All occurrences recorded after the year of publication for each dataset were removed. The geographic extent of the study areas was rarely defined, and never with coordinates, in the previous studies. As such, the coordinates of the study areas were estimated by visual inspection of figures of occurrences in the previous studies, defining the occurrences at the most extreme points by latitude and the most extreme points by longitude respectively as the border of the study area, then a 5% buffer was added to each side of the study areas. The border coordinates were rounded to 1 decimal. All occurrences outside a study area for their respective study were excluded.

The climate data are imported in a raster format which is rectangular by nature. The SRE model is computationally intensive and with the resources available for this study it was not possible to run all the datasets at their original resolutions. To make the model run successfully, the climate data had to be aggregated to a lower resolution. The level of aggregation was set to different levels from study area to study area, based on resolution and size, with the goal of finishing a single model run within 4 hours. Some datasets could run at higher resolutions for up to 6 hours, but due to memory limitation all model runs extending beyond this timeframe would crash before finishing. The buffer was not included in cases where the extended area did not include any usable data (such as a buffer of a land-dwelling species only including ocean). In some datasets, the native and invasive study areas would overlap. The 5% buffer would not be added to the relevant sides of the study areas if that prevented the overlap. In some cases, the square defined by the most extreme latitudinal/longitudinal occurrence records of one study area would include occurrence records of the other study area (such as the Morehouse and Tobler dataset where the native study area is contained within the invasive study area). In these cases, the parts of the study area containing the undesirable occurrence records were excluded. Exclusion zones were drawn prioritizing excluding a distinguishable landmass containing only undesirable occurrences or excluding a minimal amount of area containing all undesirable occurrence records. The coordinates of all excluded areas are specified in supplementary materials.

The occurrence data was then mapped onto a grid of the same dimensions and resolution as the climate data. This grid only marked the presence of occurrences within cells, not the abundance of occurrences.

Table 1: Study area extent and occurrence & environmental data for each case study. Brackets in the resolution column refer to the aggregation factor required for the native and invasive study areas respectively. The variables column refers to the environmental predictors by the naming scheme of their respective sources.

Studies	Extent of study area (W, E, S, N)		Occurrence data	Environmental data		
	Native	Invasive	Sources	Sources	Resolution (Native/Invasive aggregation)	Variables
Angetter et al. 2011	-86, -73.7, 19.4, 27.5 ¹	-99.6, -75.6, 24.2, 37.9 ¹	GBIF & VertNet	Worldclim	30 arcsec (4/8)	bio1, bio2, bio5, bio8, bio9, bio10, bio12, bio13, bio14, bio18
Battini	139.7, 177.9, -47.9, -27.4	-66.3, -53, -45, -34.1	Supplementary material	Bio-Oracle	5 arcmin (3/1)	Mean depth, Current velocity maximum, Current velocity range, Dissolved oxygen minimum, Dissolved oxygen range, Temperature mean, Temperature range
Bidinger	86.7, 153.7, 21.9, 57.7	-7.1, 16.5, 42.4, 58.2	Appendix	Worldclim	5 arcmin (3/2)	bio1, bio3, bio5, bio6, bio7, bio8, bio9, bio15, bio16, bio17
Manzoor	-9, -3.4, 35.8, 41	-7.6, 1.9, 49.9, 59.4 ¹	GBIF	Worldclim	30 arcsec (2/2)	bio1, bio6, bio12, bio14
Morehouse	-87.3, -81.1, 36.6, 41.9 ¹	-122.5, -66.7, 33, 50.9 ¹	U.S. Geological Survey and the Illinois Natural History Survey	Worldclim	30 arcsec (3/13)	bio3, bio5, bio7, bio9, bio10, bio15, bio16, bio17, bio18, bio19
Rödder 2009	-11, 39.3, 21, 46.7	-120.2, -75.9, 15.3, 37.8	GBIF & VertNet	Worldclim	5 arcmin (2/2)	bio1, bio2, bio3, bio4, bio5, bio6, bio7, bio8, bio9, bio12, bio13, bio14, bio15, bio18, bio19
Rödder 2010	-86, -73.8, 19, 27	-93.8, -79.4, 24, 33.8 ¹	GBIF & VertNet	Worldclim	30 arcsec (3/4)	bio1, bio2, bio5, bio8, bio9, bio10, bio12, bio13, bio14, bio18
Zhang	-112.4, -37.8, -34.1, 31.3 ¹	79, 129.6, 22.3, 47.1	Supplementary material	Worldclim	2.5 arcmin (3/3)	bio5, bio6, bio18 & elevation
Ørsted	66.4, 148.8, 8.5, 49.8	-17.4, 44.2, 31.3, 60.9 ²	DataDryad	Worldclim	30 arcsec (17/ 13)	bio01, bio02, bio04, bio05, bio06, bio07, bio11, bio12, bio17

¹ Parts of this range was excluded, see supplementary material.

² The western buffer zone consisted only of ocean (irrelevant for the terrestrial study species) and was excluded to reduce computational strain.

2.3 The spatial random effect model

To model distribution of invasive species in their native and invasive study areas, I used an implementation of species distribution model with a spatial random effect (Redding et al., 2017). This model consists of a generalised linear model with beta-binomial family (hierarchical model with beta distributed probability of success of the binomial distribution) with two random effects. The first random effect being the spatial effect, so called Matérn model, which assumes that the occurrences have spatial covariance following a Matérn correlation function (Genton, 2001). This random effect was used to estimate spatial processes which were not captured by climatic predictors such as dispersal limitation or species interactions. The second random effect was for climatic predictors and their quadratic transformations whose effects were assumed to be normally distributed. Estimating the effects of climate predictors using random effects instead of the typically used fixed effects effectively shrinks posterior distributions of their parameters towards zero, which is equivalent to ridge regression (Hilt et al., 1977, Dormann et al., 2013) and was implemented in INLA using the method described in Wang et al. (2018). Ridge regression is regularization technique suited for estimation of parameters of multiple regression with correlated predictors as is often the case for climate variables (Dormann et al., 2013). Uninformative priors were used for each of the model parameters according to the INLA's default configurations (see INLA's user manual for more details: <https://inla.r-inla-download.org/r-inla.org/doc/inla-manual/inla-manual.pdf> (Rue et al., 2009)).

The model was implemented in R (R Core Team, 2022) using INLA package (Rue et al., 2009), which does integrated Laplace approximation for Bayesian inference (Gómez-Rubio, 2020).

Outputs of used model implementation include posterior mean of species response curve and 95% credible interval for each climate variable. Applying these climate response curves to the spatial climate data of the input results in a spatial prediction of the fundamental niche based on the available climate data.

By combining the fundamental niche prediction with spatial effect, the model implementation produces a spatial prediction of the realised niche. The realised niche prediction also has 95% credible interval (an uncertainty map).

2.4 Niche overlap analysis

To determine the potential presence and magnitude of any niche shifts I measured the niche overlap of the suitability estimates between the native and the invasive study areas for each case study. Guisan et al. (2014) classified 3 types of data niche overlap analysis can be performed on:

Niche overlap measured on geographic predictions of suitability (habitat or distribution estimates), hereafter referred to as **geographic comparison**, is a favoured method because it utilises the primary output of an SDM. Geographic comparison requires that the habitat/distribution estimate of two models cover the same area using the same grid. For cases such as niche shift studies on invasive species, this would require that one or both habitat estimates are projected onto the other study range or another area (usually global). Most of the time the estimate of the native model is projected onto the invasive study area for comparison since we are more interested in the transferability of the native model than the invasive one. Then for each cell, the difference in predicted values is measured.

While simple to perform, this method is inadequate because of what it measured isn't strictly the niche (fundamental for a habitat estimate or realised for a distribution estimate). A Hutchinsonian

niche is only described in environmental space. A spatial prediction overlap measure can tell you how much the spatial predictions will vary within an area, but it does not inform you how or how much the species' relationships with environmental factors vary between study areas. A geographic area is never evenly distributed in climate space and a measure of overlap using the same estimates in another area will yield different results.

Niche overlap measured on a Principal Component Analysis (PCA) generated space, hereafter referred to as **PCA comparison**, is a popular method among niche shift studies (Liu et al., 2020). This method does not utilise an SDM (but usually used in conjunction with one), but a PCA.

A PCA performs linear rotation of the axes in environmental space to maximise the amount of variation in the data accounted for in the higher dimensional axes (Jolliffe and Cadima, 2016). If the 2 first axes account for enough of the variation (at least a majority) then it is common to reduce the number of dimensions down to those 2 axes. For PCA comparison, the PCA is constructed with environmental data from the native and invasive study ranges and the dimensions are usually reduced to 2. Then the occurrences of the native and invasive are plotted onto the PCA separately and the overlap is compared. This method is unsuitable for this study since I rely on the SRE model to get at the fundamental niche. In theory, one could utilise the response curves calculated by an SDM to construct a multidimensional environmental space with as many axes as there are environmental gradients used in the model, though I have not encountered this method in the literature.

Niche overlap measured on the response curves calculated by an SDM, hereafter referred to as **univariate comparison**, is the method I have decided to use in this study. It utilises SDM outputs, which lets me get at the fundamental niche via the SRE model, and it measures in environmental space. A univariate comparison also displays what environmental preferences/tolerances have shifted e.g., a shift towards higher soil pH. To represent the total niche overlap, the mean overlap across environmental gradients will be used. While suitable for this study, this method is not suitable for models that calculate the interactions between environmental gradients.

The previous case studies calculate niche overlap in different ways.

First, is the overlap index Schoener's D (Schoener, 1968). The original use was to measure the niche overlap along a single gradient, prey size, in lizards. In a biogeographical context, Schoener's D calculates differences between suitability scores for each cell in a grid (spatial or multi-dimensional environmental space) for both niches (models), sums them and standardises them. In the case of measuring niche overlap of individual environmental gradients (response curves), the gradient must be divided into distinct bins, though there is no procedure to determine how many bins the gradient should be divided into. D can range from 0 (no overlap), to 1 (identical distribution), however D does not have a 1-1 relationship to niche overlap, a D of 0.3 does not equate to a 30% niche overlap.

$$D(p_x, p_y) = 1 - \frac{1}{2} \sum_i |p_{x,i} - p_{y,i}|$$

Where p is the suitability in a given cell of the area or bin of the gradient i divided by the sum of the suitability of all cells. x represents the predictions generated from native study range, while y is generated from invasive study range.

Second, is Warren's I (Warren et al., 2008), a modified version of Schoener's D meant to have a relationship to the niche overlap that is closer to 1-1. Warren et al. (2008) features a typographical error in the equation, but a corrected version can be found in Rödder and Engler (2011).

$$I(p_x, p_y) = 1 - \frac{1}{2} \left(\sqrt{\sum_i (\sqrt{p_{x,i}} - \sqrt{p_{y,i}})^2} \right)^2$$

(The symbols have the same meaning as in Schoener's D)

Bidinger et al. (2012) & Ørsted and Ørsted (2019) did not quantify their claims of niche shifts. Bidinger et al. (2012) looked at the distribution of native and invasive occurrences respectively along individual environmental gradients (as well as the distribution of those gradients in the native and introduced study areas). The authors noted that the native and invasive occurrences along most environmental gradients varied substantially, although the gradient with the highest contribution (40-50%) did not show such variation.

Ørsted and Ørsted (2019) looked at how the contribution of each environmental variable varied between the models based on the native/invasive study areas. They also noted the poor performance when one model was projected onto the other study area to predict the occurrence records there.

I decided to eschew these two overlap indexes for a more accurate calculation of the niche overlap that should still be closely comparable to Warren's I. The SRE model calculates suitability along independent environmental gradients, and I will be using the univariate comparison method. Schoener's D and Warren's I are more suitable for measurement in distinct "batches" (or cells in a grid) and become more abstract when measured on gradients. In univariate comparison, measuring the overlap of the response curves approximates the niche more closely than Schoener's D and Warren's I.

Signal strength of the response curves is influenced by the abundance of occurrences within the study range. To account for abundance, I standardized the response curves by dividing the probability by the area under each curve respectively. The native and the invasive curves are divided by their area under the curve in the available climate range they were fitted to. I did not have a functional form of the credible interval; therefore, I used only posterior mean estimate. To avoid extrapolation, the overlap is only measured in the climatic ranges shared by both native and invasive ranges (see figure 1).

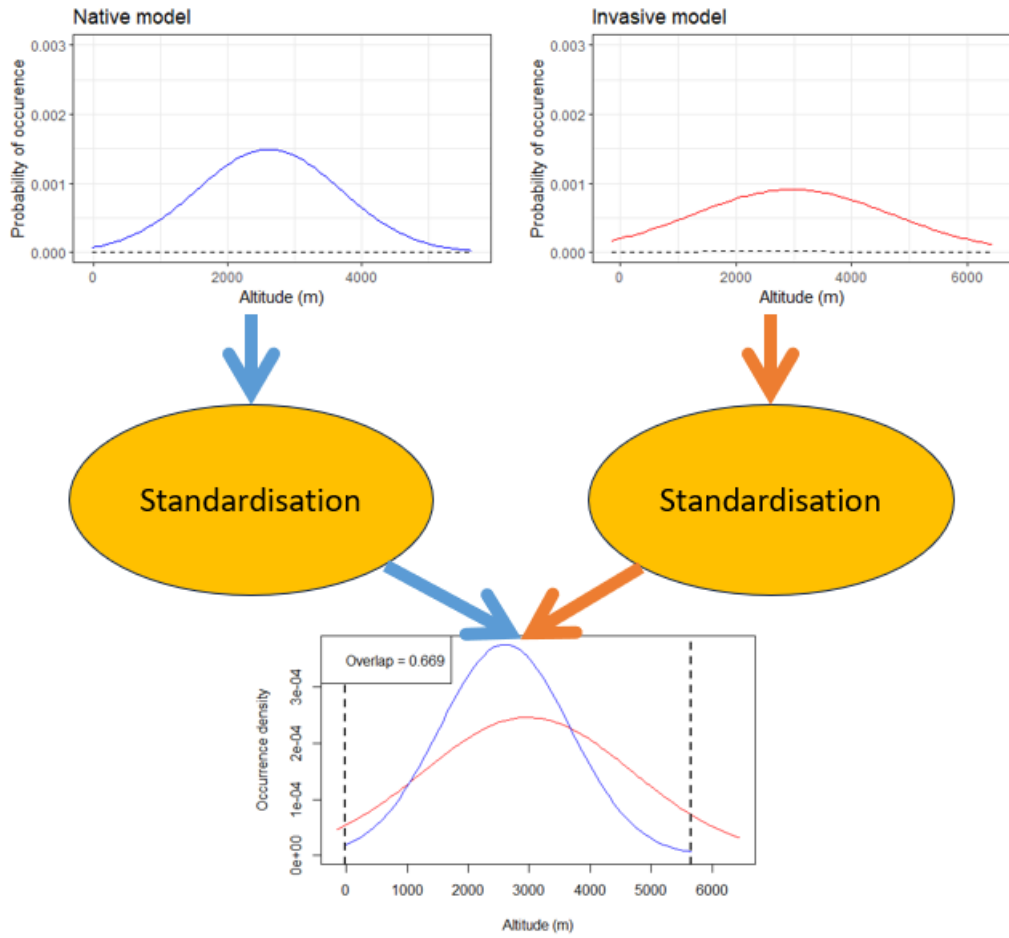


Figure 1: Response curves for an individual climatic variable in both native and invasive ranges are standardised by their respective area under the curve. The overlap of the 2 standardised response curves is only measured in the climatic range shared by both the native and invasive ranges (between the dashed lines).

2.5 Dataset & model assessment

To assess the reliability of the model results a series of quantitative and qualitative criteria assessing the datasets used and the model were collected. Each assessment has a quality measure of Satisfactory, Questionable & Unsatisfactory.

Dataset assessments:

Sample size is the amount of occurrence records in the dataset. A low number of occurrences casts suspicion on the results of this and the previous study, but it is important to note that the abundance of occurrence records alone is not that informative without considering the geographic extent of the model. 1000 occurrence records across France would be quite good for our models, but the same number of records across all of Europe would be considerably worse (see Prevalence below).

Prevalence refers to what proportion of valid cells in the model have occurrences in them. Valid cells refer to all cells where the species could occur, a terrestrial dataset would not have any values in oceanic cells making them “invalid”. Prevalence is an assessment that combines the factors of “Sample size” and “Model aggregation” and reflects upon both the quality of the model and dataset. A low prevalence means there is relatively little information to estimate the response curves, leading to higher uncertainty. A low prevalence can also negatively affect several model performance metrics, although AUC is unaffected (McPherson et al., 2004). Rödder and Engler (2011) show that Schoener’s D is sensitive to low prevalence. Prevalence should not impact the overlap measure of this study, as it uses standardised climate response curves, but the overlaps of previous studies will be less reliable at a low prevalence. The quality assessment thresholds are: Satisfactory > 0.1, Questionable 0.1-0.01, Unsatisfactory < 0.01.

Climate range overlap is a measurement of the proportion of climate that is shared by both native and invasive ranges compared to the total climatic range of both ranges combined. The coverage is measured for each climatic variable and the mean is computed. Should there be a high degree of shared climate and a low degree niche overlap it would be safe to assume the processes shaping the response curves are not limited by available climate (thus boosting the confidence in a niche shift). A low degree of shared climate would mean the 2 ranges are mostly non-analogous. The models would need to extrapolate climate relationships beyond their fitted parameters when projected onto the other range, casting uncertainty on the overlap.

The quality assessment thresholds are: Satisfactory > 0.7, Questionable 0.4-0.7, Unsatisfactory < 0.4.

Climate constraints is a qualitative assessment of how the recorded occurrences line up with the available climate range for each climatic variable. If many occurrences occur at the limit of the available climate range for a given variable, then it would lend credence to the argument that much of the species physiologically viable range may not be present, making the prediction of the niche uncertain.

Model assessments:

Data aggregation refers to the factor the climate grid had to be aggregated by. The higher the aggregation the more likely it is that patterns in the climate data are lost. From exploration of the datasets (usually with a reduced amount of climate variables to allow the model to run at lower aggregation), aggregations of 2-3 show little to no deviance from an unaggregated dataset. At an aggregation factor of 4-8 the amount of change is generally low but is context dependent on range size the initial resolution of the dataset. At aggregation factors of 9+ the changes between each degree of aggregation are very noticeable and spatial effect and habitat predictions tend to become a single hotspot radiating uniformly outwards. Aggregation also affects the occurrence data used in the model. The model only considers species presence in cells and not the abundance of occurrences within cells. Larger cells are more likely to contain more nearby occurrences, reducing the effective number of occurrences and reducing the density of occurrences of the local region of cells. The quality assessment thresholds are: Satisfactory < 3, Questionable 3-7, Unsatisfactory > 7.

Uncertainty is a measure of the uncertainty applies to the spatial predictions of the realised niche (the uncertainty of the spatial prediction of the fundamental niche was not available in the current version of the model). There was no single value of the uncertainty, such as the mean or maximum, that was truly representative of the patterns seen in the figures. A qualitative assessment was performed instead, looking at the patterns of uncertainty that had large enough aggregations of higher values to be visible in the figures. Sometimes small spots in the study area have much greater uncertainty than the rest of the model, obscuring any pattern of uncertainty in the rest of the area. These spots have been marked in grey and are described in each case study (See supplemental materials).

Spatial-climate contribution the mean of the proportional magnitude of the predictions for each cell using only the spatial components of the model divided by the predictions for each cell using only the climatic components of the model. If the spatial-climate contribution is above 1 then spatial effects are more important than environmental effects in determining the distribution of occurrences. If this is the case, then transferring a habitat estimate (the projection of the fundamental niche) onto that study range from another would be less useful since the distribution will diverge more from the estimate.

A high spatial-climate contribution also decreases the certainty that the fundamental niche can be accurately derived from this dataset.

The explained variability is calculated via the function:

$$Spatial_climate\ contribution = \frac{1}{n} \sum_i \frac{|s_i|}{|e_i|}$$

Where s_i is the value of the prediction (on the linear scale) for cell i using only the spatial components of the model, e_i is the value of the prediction (on the linear scale) for cell i using only the climatic components of the model and n is the number of cells.

The quality assessment thresholds are: Satisfactory < 1, Questionable 1-2, Unsatisfactory > 2.

AUC, the **Area Under the ROC-Curve** (Receiver Operating Characteristics), is a standard model performance metric for SDMs. The ROC-curve is a plot of true positive rate against false positive rate, each point in the plot is a different threshold for selecting the data as true or false. The area under the resulting curve is the AUC. AUC can have a value from 0 to 1 but in SDMs a value the relevant range is 0.5 (predictions are random) to 1 (prediction perfectly captures the testing data)

(Swets, 1988, Angetter et al., 2011). Due to low sample sizes in several datasets I decided to measure AUC on the training data (the data used in the model) instead of a separate set of testing data. This means that the AUC measures how well the model is fitted to the occurrence data rather than how well the model predicts known occurrences not used in the model.

The quality assessment thresholds are: Satisfactory > 0.8, Questionable 0.7-0.8, Unsatisfactory < 0.7.

Response curve shapes is a qualitative assessment of the shape of the climate response curves within their respective ranges. The model only calculates these curves as quadratic functions, but the quadratic term can be positive or negative. A positive quadratic term (a U-shaped curve) is theoretically unrealistic for most variables, a species will have physiological limits at both extreme ends of a gradient. It may not be a bad sign dependent on where it is cut off by the available climate. If a U-shaped curve is cut off somewhere in the middle so that it only has one rising tail, it may approximate one half of a n-shaped curve (with a negative quadratic term).

For practical purposes, n-shaped curves cut off before their peaks and U-shaped curves in general are bad for projecting a model onto a new range, as it will be prone to overprediction beyond the available climate it was fitted to. Flat curves are not necessarily indicative of an improper statistical model (it simply means that variable had no effect), but many flat curves in a model is a sign of a low effect of climate in general.

All code used for this study (function files, 1 file per dataset and 1 template file) are available at:

https://github.com/IsakLerum/Masters_niche_shift.

3. Results

The overlap measures showed an inconsistent pattern when compared to the measures of previous studies (Table 2). The studies that used Warren's I, which is one-to-one comparable with the new overlap, showed both lower and higher (Rödder and Lötters, 2009, Rödder and Lötters, 2010) overlaps in comparison to the new model.

Overlaps measured using Schoener's D are not one-to-one comparable with the new overlap, but generally indicate a similar range of values, as can be seen in the studies using both Schoener's D and Warren's I. Angetter et al. (2011) showed a higher overlap than the new model, Zhang et al. (2022) showed a lower overlap. Battini et al. (2019)'s overlap lied in a similar band to the new overlap. All new overlaps, except the one measured from the Zhang et al. (2022) dataset, were below the initial expectation.

Table 2: Niche overlap plus standard deviation (sd) estimates for the Spatial Random Effects (SRE) model in this study and compared to overlap values from the original case studies.

Study	SRE model Overlap (sd)	Original overlap	Measurement method	Niche measured	Data type for niche overlap
Angetter	0.43 (0.26)	0.68 ³	Schoener's D per climatic variable	Unspecified	Univariate comparison
Battini	0.47 (0.28)	0.13	Schoener's D	Realised	PCA comparison
Bidinger	0.44 (0.17)	Not quantified	Comparing environmental variable contribution, occurrence distributions along gradients.	Unspecified	Univariate comparison
Manzoor	0.25 (0.16)	D = 0.005 I = 0.004	Schoener's D & Warren's I	Unspecified	PCA comparison
Morehouse	0.14 (0.11)	0.30 ⁴	Schoener's D	Unspecified	Geographic comparison
Rödder 2009	0.41 (0.22)	D = 0.59 [‡] I = 0.75 [‡]	Schoener's D & Warren's I	Unspecified	Geographic/Univariate hybrid ⁵
Rödder 2010	0.44 (0.22)	D = 0.64 [‡] I = 0.77 [‡]	Schoener's D & Warren's I	Realised	Geographic/Univariate hybrid**
Zhang	0.70 (0.14)	0.33	Schoener's D	Realised	PCA comparison
Ørsted	0.55 (0.18)	Not quantified	Comparing environmental variable contribution each model's performance in predicting the occurrence records in the other range.	Unspecified	Geographic comparison

The assessments on the quality of the datasets had mixed results (Table 3).

Sample size assessment had an equal distribution of satisfactory, questionable & unsatisfactory results. Low sample size was just as likely to be a problem in the native as in the invasive study areas. But when looking at prevalence, all datasets were unsatisfactory except Rödder and Lötters (2009) which was questionable.

Climate range overlap was questionable to unsatisfactory except in the case of Rödder and Lötters (2009) which was satisfactory. In most datasets constraints on occurrences caused by available climate were minimal to non-existent. Manzoor et al. (2020) & Rödder and Lötters (2010) were scored as questionable due to climate constraints affecting distributions across some environmental gradients.

³ Total overlap was not measured in the study but overlap of individual climatic variables was. The mean of those overlaps is displayed here.

⁴ A comparison of available climate between ranges showed that the lack of shared climate was the cause of the low overlap. The niche overlap was deemed not significant.

⁵ See supplemental material for full description.

Table 3: Assessments of the datasets defined as satisfactory (blue), questionable (yellow) & unsatisfactory (red). Brackets in the Sample size column refer to how many cells of the study area had occurrences.

Study	Sample size: Native/Invasive (cells after aggregation)	Prevalence (Native / Invasive)	Mean climate range overlap (sd)	Climate constraints
Angetter	7594(122)/ 4221(273)	0.0148/ 0.0063	0.57 (0.20)	No climate constraints.
Battini	190(42)/ 12(8)	0.0014/ 0.0006	0.52 (0.28)	No climate constraints.
Bidinger	60(59)/ 171(162)	0.0005/ 0.0049	0.50 (0.21)	No climate constraints.
Manzoor	237(76)/ 40853(8446). Suspect poor reporting in native range.	0.0009/ 0.2128	0.36 (0.24)	Minor limitations on invasive occurrences
Morehouse	152(121)/ 2465(891)	0.0024/ 0.0123	0.28 (0.12)	Native occurrences often meet the edge of available climate.
Rödder 2009	1842(501)/ 1480(242)	0.0156/ 0.0133	0.71 (0.17)	No climate constraints.
Rödder 2010	3862 (320)/ 3633 (437)	0.0015/ 0.0039	0.56 (0.18)	Few variables limited by available climate.
Zhang	506(301)/ 393(232)	0.0031/ 0.0033	0.66 (0.27)	No climate constraints.
Ørsted	101(85)/ 306(185)	0.0009/ 0.0022	0.61 (0.15)	No climate constraints.

The assessments on the quality of the model had a distinction of mostly satisfactory assessment categories and mostly unsatisfactory assessment categories (Table 4).

Aggregation was kept within the satisfactory range for most datasets, but both Morehouse and Tobler (2013) & Ørsted and Ørsted (2019) utilised data with the highest resolution (30 arcseconds) of the entire USA or whole continents and were therefore highly aggregated in our analyses to achieve feasible computation times. AUC were almost all satisfactory, except the native model in Angetter et al. (2011) and the invasive model in Manzoor et al. (2020). Uncertainty was low to questionable in all case studies except (Battini et al., 2019) which had high uncertainty in the native model. Response curves on most models were unsatisfactory to questionable, with Rödder and Lötters (2009) and Zhang et al. (2022) assessed as satisfactory.

All case studies showed unsatisfactory results on the spatial-climate contribution assessment, except the Battini et al. (2019) dataset which had a questionable result (in the native study there was no spatial effect). Many case studies had models where the spatial effect contributed over 3 times as much as environmental effects.

Table 4: Assessments of the model defined as satisfactory (blue), questionable (yellow) & unsatisfactory (red).

Study	Model aggregation (Native / Invasive)	Uncertainty	Spatial/environmental contribution (Native / Invasive)	AUC (Native/ Invasive)	Response curve shapes
Angetter	4/8	Middling in invasive model.	34.995 / 15.919	0.731/ 0.973	Some curves are U-shaped in both models.
Battini	3/1	High in native model.	0 / 1.369	0.965/ 0.942	Invasive curves are all flat.
Bidinger	3/2	Low in models	29.323 / 16.808	0.936/ 0.803	Curves in native model are quite flat. Invasive model has a few U-shapes.
Manzoor	2/2	Low in models	3.308 / 15.995	0.915/ 0.759	Flat in native model. One U-shaped curve in invasive model.
Morehouse	3/13	Middling in invasive model	2.505 / 7.836	0.854/ 0.942	Some U-shapes in both models.
Rödder 2009	2/2	Middling in native model.	6.675 / 3.220	0.979/ 0.868	Some flat curves in invasive model.
Rödder 2010	2/3	Low in models.	3.130 / 40.975	0.876/ 0.931	Several U-shaped curves in both models.
Zhang	3/3	Low in models.	6.117 / 6.612	0.959/ 0.921	Bell shaped curves.
Ørsted	17/ 13	Middling where there is strongest occurrence signal.	3.797 / 48.112	0.816/ 0.864	Native curves are all flat.

4. Discussion

I expected that the niche overlap measures based on the spatial random effect model would be consistently higher than the measures from the original studies using models without a spatial effect. This expectation arises from the fact that the previous case studies all measured the overlap of the realised niche while this study theoretically measures the overlap of the fundamental niche. The realised niche would be expected to have a lesser overlap because it includes the spatial effect of dispersal barriers and biotic interactions. But the SRE model niche overlaps showed an inconsistent pattern of both greater and lesser values than the previous case studies.

I also expected that the new overlap measures would be above 0.6, which Rödder and Engler (2011) consider to be a high degree of overlap, since the groups being compared are still of the same species. All datasets except Zhang et al. (2022) produced a lower fundamental niche overlap. Across all datasets the mean overlap was 0.42 ± 0.16 , a value in-between other studies that have measured the native-invasive realised niche overlap of multiple species. Liu et al. (2020), which reviewed studies on 434 invasive species across different taxa, concluded that niche shifts were relatively rare with a mean niche overlap of 0.53. Atwater et al. (2017), a study that measured the niche overlap of 815 plant species, concluded that niche shifts are common with a mean niche overlap of 0.314 ± 0.185 . The results of this study leans more towards Atwater et al. (2017)'s conclusion, that niche shifts (in their case realised and in this case fundamental) niche shifts are common.

The impact the fundamental niche overlap (the univariate comparison of this study) has on the difference in spatial predictions (the geographic comparison) is uncertain, but the correlation between niche overlap and spatial predictions is known for the realised niche. Atwater and Barney (2021), a follow-up study to Atwater et al. (2017), showed a correlation between niche overlap and an overlap of spatial predictions of 0.456, with spatial overlap generally being higher. I would expect that the correlation between overlap of suitability estimates and the overlap of habitat estimates to be roughly similar, although it is a topic to be confirmed in further research.

As it stands, the low niche overlap suggests that spatial predictions would be non-transferable due to niche shifts, but the reliability of these overlaps is compromised by the results of the dataset and model assessments.

4.1 Dataset assessments

The quality of the sample sizes was evenly distributed, but all datasets had a very low prevalence. This low prevalence does put some uncertainty on the overlap measures of the previous studies. Low sample size was just as often in native as in invasive study areas.

Climate range overlap was only about 50% on average, meaning only half of the total climatic space was usable for niche overlap calculations. Especially Manzoor et al. (2020) (comparing Iberian Peninsula to Great Britain) and Morehouse and Tobler (2013) (comparing Illinois & Indiana to the rest of USA & southern Canada) with 36% and 28% climate range overlap respectively. Large portions of each model's response curve were outside the shared climatic space. This could lead to over- or underestimation of the niche overlap, but a cursory look on the response curves indicate that the overlap would be even lower should the niches be compared along the entire climatic space. The limits of available climate in the study ranges rarely seemed to impact the recorded occurrences, except in the Morehouse and Tobler (2013) dataset where the native occurrences often clustered at the edge of the available climate.

Overall, the datasets had some clear issues that might have affected the niche overlap analysis of this study and the previous case studies, especially the Manzoor et al. (2020) and Morehouse and Tobler (2013) datasets. Issues of sample size/prevalence could be solved by more rigorous surveying and reporting schemes, but with rare species they will likely stay a problem nevertheless due to scarcity of actual occurrence (Lawson et al., 2014). Issues of non-analogous climate between study areas and occurrence distributions limited by the available climate (such as study ranges of islands) are inherent to each case and cannot be solved. Cases of severe degrees of non-analogous climate and lack of available climate are simply unsuitable for SDM transferability and fundamental niche overlap analysis (though a lack of available climate only impacts transferability from a climate limited study range, not to a climate limited study range) (Stroud, 2021, Liu et al., 2021).

4.2 Model assessments

While aggregation is undesirable for the purpose of replicating previous studies most of the case studies could be replicated at a low factor of aggregation. Exploration of the datasets at different aggregation factors showed that the difference between low aggregation factors rarely had a noticeable effect on the results. This indicates that the environmental processes that define the niche of the modelled species operate on a coarser scale than the resolution used for the previous studies, especially Angetter et al. (2011), Manzoor et al. (2020), Morehouse and Tobler (2013), Rödder and Lötters (2010) and Ørsted and Ørsted (2019) that use a resolution of 30 arcsec. A reduction in the initial resolution would reduce computational strain.

Assessing uncertainty in the models was more difficult than initially expected. Measuring uncertainty (the width of the 95% credible interval) in the spatial predictions only proved meaningful under a qualitative assessment.

The spatial uncertainty showed a similar pattern across all models. Where predicted probability of occurrence was zero to low the uncertainty was zero to extremely low as well, with the exceptions of a few spots of extremely high uncertainty that are likely to be artefacts. Uncertainty tends to follow the distribution/habitat estimate in pattern and relative strength.

Uncertainty was also measured in the form of credible intervals on the response curves (example in figure 2), though not featured in the assessment table. These intervals were generally larger than the value of the mean. However, a functional form of these credible intervals was not available with the standardisation used to account for the abundance of the species within the study area (which I prioritised due to the difference in abundance between native and invasive study areas). I don't know the extent of uncertainty in the overlap scores since only the mean could be standardised for the niche overlap analysis.

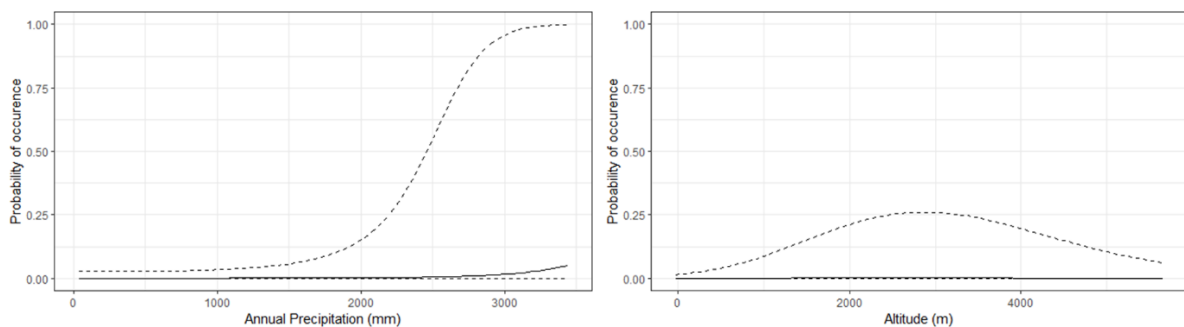


Figure 2: Example response curve (not standardised) with 95% credible interval from the Angetter et. al. (left) and Zhang et. al. (right) datasets. The solid line signifies the mean while the dashed lines signify the upper and lower credible intervals.

Contribution of spatial effects relative to environmental effects was quite high with only 3 study areas having a spatial-climate contribution below 3. This indicates that spatial effects are the main factors in determining the occurrence distribution. There are 3 possible causes of a high spatial effect contribution:

First, the species distribution in the study area is much more constrained by dispersal barriers and/or biotic interactions than environmental factors. This is to be expected in cases such as the Angetter et al. (2011) and Rödder and Lötters (2010) datasets where the native study areas are limited to Cuba and surrounding islands. Paquette and Hargreaves (2021) who discovered that biotic interactions tend to be the major limiting factor species' warm versus cool range edges. Amundrud et al. (2018) found that in the case of helicopter damselfly (*Mecistogaster modesta*) the species distribution was more limited by dispersal barriers than climate.

Second, the model is missing an environmental factor that is meaningful to the distribution of the species within the study area. The effects of the missing environmental factor will then be captured by the SRE.

Third, is the possibility of ongoing expansion. This is more likely to be the case in the invasive study range due to an ongoing invasion process but may also occur in the native study area if changes in the native range (mitigation of dispersal barriers, alteration of the community composition or climate change) loosen the constraints on the species distribution. Ongoing expansion would mean the distribution is defined by spatial effects of spread rather than environmental factors.

I have no way to distinguish between these 3 possibilities but no matter the cause, a high spatial-climate contribution of 3 puts a heavy suspicion on the model's ability to capture the fundamental niche.

AUC was generally high, showing that the model was generally well fitted to the data, but the high fit is more likely to be explained by the spatial effects (due to the high spatial-climate contribution) rather than environmental effects. The SRE model runs on the Battini et al. (2019) dataset are the exceptions, with lower spatial-climate contributions yet high AUC values.

The response curves show a lot of issues. Some cases had flat response curves along all/most climate variables in the model. Some datasets have response curves with a positive quadratic term. If only one tail of the curve is contained in the available climate it can approximate part of an n-shaped curve (the biologically realistic result), but if the curve is u-shaped, it is clearly unrealistic. Measurements of overlap with u-shaped curves are thus less reliable. Since independent response curves are supposed to be roughly bell-shaped (ter Braak, 1988, Guisan and Thuiller, 2005), the model could potentially be improved by disallowing the model to form response curves with positive quadratic terms. This would be especially helpful when projecting the habitat estimate onto a new study area with a non-analogous environment. Some response curves deviated from the occurrence record distribution along their environmental gradient, e.g. the invasive response curve for annual precipitation in the Manzoor et al. (2020) dataset. These deviations may be caused by a high occupancy rate of a rare climate (in this case high precipitation areas). The model interprets the high occupancy rate as a high suitability and shifts the response curve accordingly. These shifted curves tend to have their highest values at the edges of the available climate (and have positive quadratic terms), which may lower the overlap values and drag down the mean fundamental niche overlap.

The dominance of spatial effects in the model is by far the most interesting findings among the results. This dominance of spatial effects renders the SRE model unable to capture the fundamental niche and the measurement niche overlap becomes less insightful. Thus, based on these results we are unable to produce reliable niche approximations. Research question 1 (Does the SRE model replicate the pattern of niche shifts from previous studies) is therefore impossible to answer with confidence. Spatial effects were just as likely to be dominant in native ranges as in introduced ranges, indicating that spatial effects generally are more important for defining the geographic distribution of species than environmental effects. This is partially supported by previous niche shift studies (Petitpierre et al., 2012, Atwater et al., 2017, Atwater and Barney, 2021, Liu et al., 2020) which identify a majority of niche shifts to be caused by unfilling, the invasive niche not occupying niche space occupied by the native niche. They suggested that this unfilling is temporary as the niche will be filled out due to ongoing expansion in the introduced range. The importance of spatial processes for species distributions in general are further supported by Dormann et al. (2013) and Miller et al. (2007).

If spatial effect dominance is the norm, then it has 2 important implications for biogeography and invasion ecology. First, while the fundamental niche might be helpful for mapping out potentially suitable habitats for species, it will not be useful for predicting the distribution of species, whether native or invasive. Second, SDMs are generally unable to capture the relationship between species distributions and the environment without the confounding influence of spatial effects, making SDMs in general non-transferrable.

4.3 Other considerations

An issue I encountered during this study was selecting the geographic extent for each study area, since these parameters were not disclosed in the previous case studies. The geographic extent used in SDM should include all occurrence records of the modelled group, but also all the available climate within the population's range as well as the climatic conditions unsuitable for the species (to show the climatic limits of the group). The entire population is almost never completely recorded so it is not unlikely that there are more suitable habitats (and presumably more real occurrences) beyond the maximum spatial extent predicted from recorded occurrences. The climatic limits (should the groups be geographically limited by climate and not just dispersal barriers and biotic interactions) are also beyond this initial extent. To meet these standards, the extent of each study area ought to be defined on a case-by case basis. For reproducibility the geographic extent of a study area should be explicitly stated (something which is not commonly done with most niche shift studies).

I eschewed defining the extent on a case-by-case basis and opted for a standardised buffer of 5% from the most geographically extreme points (with exceptions for cases where a buffer intrudes on another population or contains entirely invalid data). This buffer is likely to be overly conservative, not containing enough of the surrounding climate, but necessary when weighing it against the impact of resolution and the computational strain of these 2 factors.

I would like to challenge the prospects of national modelling projects, like ECoMAP, that define study areas by political border that the modelled species are unlikely to respect. This reduces the sample size of occurrences and artificially limits the available climate for the model.

5. Conclusion

In this study I found the dominance of spatial effects in both native and invasive study areas for many species suggesting that the fundamental niche could not be accurately captured in any of the datasets. As such, it is not possible to measure the fundamental niche shift and whether evolution (in the form of genetic loss from the founder effect or gain from adaptation, hybridisation or polyploidisation) that alters environmental preferences of a species has occurred. If spatial effects (dispersal barriers, biotic interactions and/or ongoing expansion) are generally the dominant factors in determining the distribution of species, then determining the fundamental niche using SDMs would be a futile effort, attempting to transfer a native prediction onto an introduced area is therefore challenging. This limits the usefulness of SDMs as predictive tools for managing/fighting invasive species as the inability to capture the fundamental niche inhibits their ability to predict suitable habitats and ruins the model transferability between native and introduced ranges. In the management of invasive species SDMs should be limited to predicting the distribution of the invasive using data from the invasive study area which, while difficult, is still tremendously useful for management.

If the fundamental niche could be estimated from some other method than SDMs it would still have limited usefulness in predicting the species' distribution. The fundamental niche could still be used to predict the distribution of potentially suitable habitat, which is useful for selecting areas to establish preventative measures. But due to the strong spatial effects it will not produce reliable predictions of the distribution of the invasive species for selecting where restorative measures (eradication/culling of the invasive or support for harmed native species).

Further research should be conducted to determine to what degree spatial effects versus environmental effects determine the distribution of species across different taxa and in native species.

6. References

- AMUNDRUD, S. L., VIDELA, M. & SRIVASTAVA, D. S. 2018. Dispersal barriers and climate determine the geographic distribution of the helicopter damselfly *Mecistogaster modesta*. *Freshwater Biology*, 63, 214-223.
- ANGETTER, L.-S., LÖTTERS, S. & RÖDDER, D. 2011. Climate niche shift in invasive species: the case of the brown anole. *Biological Journal of the Linnean Society*, 104, 943-954.
- ASSIS, J., TYBERGHEIN, L., BOSCH, S., VERBRUGGEN, H., SERRÃO, E. A. & DE CLERCK, O. 2018. Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, 27, 277-284.
- ATWATER, D. Z., ERVINE, C. & BARNEY, J. N. 2017. Climatic niche shifts are common in introduced plants. *Nature Ecology & Evolution*, 2, 34-43.
- ATWATER, Z., DANIEL & BARNEY, J. N. 2021. Climatic niche shifts in 815 introduced plant species affect their predicted distributions. *Global Ecology and Biogeography*, 30, 1671-1684.
- BARRIOS-GARCIA, M. N. & BALLARI, S. A. 2012. Impact of wild boar (*Sus scrofa*) in its introduced and native range: a review. *Biological Invasions*, 14, 2283-2300.
- BATES, O. K. & BERTELSMEIER, C. 2021. Climatic niche shifts in introduced species. *Current Biology*, 31, R1252-R1266.
- BATTINI, N., FARÍAS, N., GIACHETTI, C., SCHWINDT, E. & BORTOLUS, A. 2019. Staying ahead of invaders: using species distribution modeling to predict alien species' potential niche shifts. *Marine Ecology Progress Series*, 612, 127-140.
- BERTELSMEIER, C., LUQUE, G. M. & COURCHAMP, F. 2013. Increase in Quantity and Quality of Suitable Areas for Invasive Species as Climate Changes. *Conservation Biology*, 27, 1458-1467.
- BIDINGER, K., LÖTTERS, S., RÖDDER, D. & VEITH, M. 2012. Species distribution models for the alien invasive Asian Harlequin ladybird (*Harmonia axyridis*). *Journal of Applied Entomology*, 136, 109-123.
- BOSCH S, FERNANDEZ S (2022). `_sdmpredictors`: Species Distribution Modelling Predictor Datasets `_R`, package version 0.2.14, <https://CRAN.R-project.org/package=sdmpredictors>.
- BRUNO, J. F., STACHOWICZ, J. J. & BERTNESS, M. D. 2003. Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution*, 18, 119-125.
- BØHN, T., AMUNDSEN, P.-A. & SPARROW, A. 2008. Competitive exclusion after invasion? *Biological Invasions*, 10, 359-368.
- CRANDALL, K. A. & GRAVE, S. D. 2017. An updated classification of the freshwater crayfishes (Decapoda: Astacidea) of the world, with a complete species list. *Journal of Crustacean Biology*, 37, 615-653.
- CUTHBERT, R. N., DIAGNE, C., HUDGINS, E. J., TURBELIN, A., AHMED, D. A., ALBERT, C., BODEY, T. W., BRISKI, E., ESSL, F., HAUBROCK, P. J., GOZLAN, R. E., KIRICHENKO, N., KOURANTIDOU, M., KRAMER, A. M. & COURCHAMP, F. 2022. Biological invasion costs reveal insufficient proactive management worldwide. *Science of The Total Environment*, 819, 153404.
- 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, 27-46.
- ELITH, J. & LEATHWICK, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.
- EVELYN HUTCHINSON, G. 1991. Population studies: Animal ecology and demography. *Bulletin of Mathematical Biology*, 53, 193-213.
- FARIAS, N., OBENAT, S. & GOYA, A. 2015. Outbreak of a neurotoxic side-gilled sea slug (*Pleurobranchaea* sp.) in Argentinian coasts. *New Zealand Journal of Zoology*, 42, 51-56.

- FARIAS, N., WOOD, S., OBENAT, S. & SCHWINDT, E. 2016. Genetic barcoding confirms the presence of the neurotoxic sea slug *Pleurobranchaea maculata* in southwestern Atlantic coast. *New Zealand Journal of Zoology*, 43, 292-298.
- FICK, S. E. & HIJMANS, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302-4315.
- GALLIEN, L., DOUZET, R., PRATTE, S., ZIMMERMANN, N. E. & THUILLER, W. 2012. Invasive species distribution models - how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, 21, 1126-1136.
- GENTON, M. 2001. Classes of Kernels for Machine Learning: A Statistics Perspective. *Journal of Machine Learning Research*, 2, 299-312.
- GÓMEZ-RUBIO, VIRGILIO (2020). Bayesian Inference with INLA. Chapman & Hall/CRC Press. Boca Raton, FL.
- GUCKER, COREY L. 2009. *Heracleum mantegazzianum*. In: Fire Effects Information System, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer). Available: <https://www.fs.usda.gov/database/feis/plants/forb/herman/all.html> [2023, June 27]
- GUISAN, A., PETITPIERRE, B., BROENNIMANN, O., DAEHLER, C. & KUEFFER, C. 2014. Unifying niche shift studies: insights from biological invasions. *Trends in Ecology & Evolution*, 29, 260-269.
- GUISAN, A. & THUILLER, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993-1009.
- GUISAN, A. & ZIMMERMANN, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-186.
- HILT, D. E., SEEGRIST, D. W., UNITED STATES. FOREST, S. & NORTHEASTERN FOREST EXPERIMENT, S. 1977. *Ridge, a computer program for calculating ridge regression estimates*, Upper Darby, Pa, Dept. of Agriculture, Forest Service, Northeastern Forest Experiment Station.
- HUDSON, C. M., VIDAL-GARCÍA, M., MURRAY, T. G. & SHINE, R. 2020. The accelerating anuran: evolution of locomotor performance in cane toads (*Rhinella marina*, Bufonidae) at an invasion front. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20201964.
- Illinois Natural History Survey - Crustacean Collection. Occurrence dataset (ID: 257ea482-2d6f-4cf4-846d-8efeb58f6727) https://biocoll.inhs.illinois.edu/portal/content/dwca/INHS-CRUSTACEAN_DwC-A.zip accessed via the INHS Collections Data Portal, biocoll.inhs.illinois.edu/portal, 2023-06-01).
- IPCC, 2021: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*[MASSON-DELMOTTE, V., P. ZHAI, A. PIRANI, S.L. CONNORS, C. PÉAN, S. BERGER, N. CAUD, Y. CHEN, L. GOLDFARB, M.I. GOMIS, M. HUANG, K. LEITZELL, E. LONNOY, J.B.R. MATTHEWS, T.K. MAYCOCK, T. WATERFIELD, O. YELEKÇI, R. YU, and B. ZHOU (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, In press, doi:[10.1017/9781009157896](https://doi.org/10.1017/9781009157896).
- IPCC, 2022: *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [H.-O. PÖRTNER, D.C. ROBERTS, M. TIGNOR, E.S. POLOCZANSKA, K. MINTENBECK, A. ALEGRÍA, M. CRAIG, S. LANGSDORF, S. LÖSCHKE, V. MÖLLER, A. OKEM, B. RAMA (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:[10.1017/9781009325844](https://doi.org/10.1017/9781009325844).
- JOLLIFFE, I. T. & CADIMA, J. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374, 20150202.
- KELLER, R. P., GEIST, J., JESCHKE, J. M. & KÜHN, I. 2011. Invasive species in Europe: ecology, status, and policy. *Environmental Sciences Europe*, 23, 23.

- KINNISON, M. T., UNWIN, M. J. & QUINN, T. P. 2008. Eco-evolutionary vs. habitat contributions to invasion in salmon: experimental evaluation in the wild. *Molecular Ecology*, 17, 405-414.
- LAUZERAL, C., LEPRIEUR, F., BEAUCHARD, O., DURON, Q., OBERDORFF, T. & BROSSE, S. 2011. Identifying climatic niche shifts using coarse-grained occurrence data: a test with non-native freshwater fish. *Global Ecology and Biogeography*, 20, 407-414.
- LAWSON, C. R., HODGSON, J. A., WILSON, R. J. & RICHARDS, S. A. 2014. Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, 5, 54-64.
- LIU, C., WOLTER, C., XIAN, W. & JESCHKE, J. M. 2020. Most invasive species largely conserve their climatic niche. *Proceedings of the National Academy of Sciences*, 117, 23643-23651.
- LIU, C., WOLTER, C., XIAN, W. & JESCHKE, J. M. 2021. Reply to Stroud: Invasive amphibians and reptiles from islands indeed show higher niche expansion than mainland species. *Proceedings of the National Academy of Sciences*, 118, e2020172118.
- MANZOOR, S. A., GRIFFITHS, G., OBIAKARA, M. C., ESPARZA-ESTRADA, C. E. & LUKAC, M. 2020. Evidence of ecological niche shift in *Rhododendron ponticum* (L.) in Britain: Hybridization as a possible cause of rapid niche expansion. *Ecology and Evolution*, 10, 2040-2050.
- MASENGILL, R. L. 2011. Control efforts for invasive northern pike *Esox lucius* on the Kenai Peninsula, 2007. Alaska Department of Fish and Game, Fishery Data Series No. 11-10, Anchorage. <https://www.adfg.alaska.gov/fedaidpdfs/fds11-10.pdf>
- MCGRANNACHAN, C. M., PAGAD, S. & MCGEOCH, M. A. 2021. A multiregional assessment of transnational pathways of introduction. *NeoBiota*, 64, 43-67.
- MCPHERSON, J. M., JETZ, W. & ROGERS, D. J. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, 41, 811-823.
- MEDINA, F. M., BONNAUD, E., VIDAL, E., TERSHY, B. R., ZAVALETA, E. S., JOSH DONLAN, C., KEITT, B. S., CORRE, M., HORWATH, S. V. & NOGALES, M. 2011. A global review of the impacts of invasive cats on island endangered vertebrates. *Global Change Biology*, 17, 3503-3510.
- MILLER, J., FRANKLIN, J. & ASPINALL, R. 2007. Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, 202, 225-242.
- MORAN, E. V. & ALEXANDER, J. M. 2014. Evolutionary responses to global change: lessons from invasive species. *Ecology Letters*, 17, 637-649.
- MOREHOUSE, R. L. & TOBLER, M. 2013. Invasion of Rusty Crayfish, *Orconectes Rusticus*, in the United States: Niche Shifts and Potential Future Distribution. *Journal of Crustacean Biology*, 33, 293-300.
- NOVAK, S. J. 2007. The role of evolution in the invasion process. *Proceedings of the National Academy of Sciences*, 104, 3671-3672.
- PAINI, D. R., SHEPPARD, A. W., COOK, D. C., DE BARRO, P. J., WORNER, S. P. & THOMAS, M. B. 2016. Global threat to agriculture from invasive species. *Proceedings of the National Academy of Sciences*, 113, 7575-7579.
- PAQUETTE, A. & HARGREAVES, A. L. 2021. Biotic interactions are more often important at species' warm versus cool range edges. *Ecology Letters*, 24, 2427-2438.
- PETERSON, A. T. 2003. Predicting the Geography of Species' Invasions via Ecological Niche Modeling. *The Quarterly Review of Biology*, 78, 419-433.
- PETERSON, A. T. 2011. Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography*, 38, 817-827.
- PETITPIERRE, B., KUEFFER, C., BROENNIMANN, O., RANDIN, C., DAEHLER, C. & GUISAN, A. 2012. Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. *Science*, 335, 1344-1348.
- POCHEVILLE, A. 2015. The Ecological Niche: History and Recent Controversies. In: HEAMS, T., HUNEMAN, P., LECOINTRE, G. & SILBERSTEIN, M. (eds.) *Handbook of Evolutionary Thinking in the Sciences*. Dordrecht: Springer Netherlands.

- PRENTIS, P. J., WILSON, J. R. U., DORMONTT, E. E., RICHARDSON, D. M. & LOWE, A. J. 2008. Adaptive evolution in invasive species. *Trends in Plant Science*, 13, 288-294.
- PYŠEK, P., JAROŠÍK, V., HULME, P. E., KÜHN, I., WILD, J., ARIANOUTSOU, M., BACHER, S., CHIRON, F., DIDŽIULIS, V., ESSL, F., GENOVESI, P., GHERARDI, F., HEJDA, M., KARK, S., LAMBTON, P. W., DESPREZ-LOUSTAU, M.-L., NENTWIG, W., PERGL, J., POBOLJŠAJ, K., RABITSCH, W., ROQUES, A., ROY, D. B., SHIRLEY, S., SOLARZ, W., VILÀ, M. & WINTER, M. 2010. Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences*, 107, 12157-12162.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- REDDING, D. W., LUCAS, T. C. D., BLACKBURN, T. M. & JONES, K. E. 2017. Evaluating Bayesian spatial methods for modelling species distributions with clumped and restricted occurrence data. *PLOS ONE*, 12, e0187602.
- 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, 409-431.
- RICHARDSON, D. M., PYSEK, P., REJMANEK, M., BARBOUR, M. G., PANETTA, F. D. & WEST, C. J. 2000. Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions*, 6, 93-107.
- RUE, H., MARTINO, S. & CHOPIN, N. 2009. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71, 319-392.
- RÖDDER, D. & ENGLER, J. O. 2011. Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Global Ecology and Biogeography*, 20, 915-927.
- RÖDDER, D. & LÖTTERS, S. 2009. Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). *Global Ecology and Biogeography*, 18, 674-687.
- RÖDDER, D. & LÖTTERS, S. 2010. Explanative power of variables used in species distribution modelling: an issue of general model transferability or niche shift in the invasive Greenhouse frog (*Eleutherodactylus planirostris*). *Naturwissenschaften*, 97, 781-796.
- SAKAI, A. K., ALLENDORF, F. W., HOLT, J. S., LODGE, D. M., MOLOFSKY, J., WITH, K. A., BAUGHMAN, S., CABIN, R. J., COHEN, J. E., ELLSTRAND, N. C., MCCAULEY, D. E., O'NEIL, P., PARKER, I. M., THOMPSON, J. N. & WELLER, S. G. 2001. The Population Biology of Invasive Species. *Annual Review of Ecology and Systematics*, 32, 305-332.
- SANTOS, J., PASCUAL, M., SIMÕES, P., FRAGATA, I., LIMA, M., KELLEN, B., SANTOS, M., MARQUES, A., ROSE, M. R. & MATOS, M. 2012. From nature to the laboratory: the impact of founder effects on adaptation. *Journal of Evolutionary Biology*, 25, 2607-2622.
- SCHOENER, T. W. 1968. The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna. *Ecology*, 49, 704-726.
- SHINE, R. 1999. Why is sex determined by nest temperature in many reptiles? *Trends in Ecology & Evolution*, 14, 186-189.
- STROUD, J. T. 2021. Island species experience higher niche expansion and lower niche conservatism during invasion. *Proceedings of the National Academy of Sciences*, 118, e2018949118.
- SWETS, J. A. 1988. Measuring the Accuracy of Diagnostic Systems. *Science*, 240, 1285-1293.
- TER BRAAK, C. J. F. 1988. *Biometrics*, 44, 631-632.
- TOPAZ, C. M., D'ORSOGNA, M. R., EDELSTEIN-KESHET, L. & BERNOFF, A. J. 2012. Locust Dynamics: Behavioral Phase Change and Swarming. *PLoS Computational Biology*, 8, e1002642.
- URBAN, M. C. 2020. Climate-tracking species are not invasive. *Nature Climate Change*, 10, 382-384.
- U.S. Geological Survey, 2023, Specimen observation data for *Faxonius rusticus* (GIRARD, 1852), Nonindigenous Aquatic Species Database, Gainesville, FL, <https://nas.er.usgs.gov/queries/CollectionInfo.aspx?SpeciesID=21>

- WANG, X., YU, Y., Faraway J. (2018) *Bayesian Regression Modeling with INLA*. Chapman & Hall/CRC Press. Boca Raton, FL.
- WARREN, D. L., GLOR, R. E. & TURELLI, M. 2008. ENVIRONMENTAL NICHE EQUIVALENCY VERSUS CONSERVATISM: QUANTITATIVE APPROACHES TO NICHE EVOLUTION. *Evolution*, 62, 2868-2883.
- ZHANG, W., CHEN, X., LIU, R., SONG, X., LIU, G., ZOU, J., QIAN, Z., ZHU, Z. & CUI, L. 2022. Realized niche shift associated with *Galinsoga quadriradiata* (Asteraceae) invasion in China. *Journal of Plant Ecology*, 15, 538-548.
- ØRSTED, I. V. & ØRSTED, M. 2019. Species distribution models of the Spotted Wing Drosophila (*Drosophila suzukii*, Diptera: Drosophilidae) in its native and invasive range reveal an ecological niche shift. *Journal of Applied Ecology*, 56, 423-435.

7. Supplemental material

a) Angetter et al. (2011)

Species: *Anolis sagrei*

Common name: Brown anole

Habitat: Terrestrial

Native range: Cuba and the Bahamas

Introduced range: The continental coast of the southern coast of the USA

Occurrence data: Data recreated from 2 sources, including only records up until 20011:

1. GBIF.org (29 March 2023) GBIF Occurrence Download: <https://doi.org/10.15468/dl.jh2rpg>
2. David Bloom (2016). VertNet_Retilia_Sept2016. CyVerse Data Commons. DOI 10.7946/P2Z59J

Occurrence records (Occurrence cells after aggregation): Native = 7594 (122), Invasive = 4221(273)

Climate data: Data retrieved from WORLDCLIM database, version 2.1. Resolution: 30 arcsec (Fick and Hijmans, 2017)

Aggregation factor: Native = 4, Invasive = 8

Native study area extent (W, E, S, N): -86, -73.7, 19.4, 27.5, Excluded: (-84, -79.5, 24, 27.5) & (-81, -79, 19.4, 20)

Invasive study area extent (W, E, S, N): -99.6, -75.6, 24.2, 37.9, Excluded: (-79.5, -72, 24.2, 28)

Predictors:

- bio1: Annual Mean Temperature (°C)
- bio2: Mean Diurnal Range (Mean of monthly (max temp - min temp)) (°C)
- bio5: Max Temperature of Warmest Month (°C)
- bio8: Mean Temperature of Wettest Quarter (°C)
- bio9: Mean Temperature of Driest Quarter (°C)
- bio10: Mean Temperature of Warmest Quarter (°C)
- bio12: Annual Precipitation (mm)
- bio13: Precipitation of Wettest Month (mm)
- bio14: Precipitation of Driest Month (mm)
- bio18: Precipitation of Warmest Quarter (mm)

The brown anole (*Anolis sagrei*) is a ground-dwelling lizard native to Cuba and surrounding island and invasive on the North American continent (Angetter et al., 2011)

Limitations:

The original paper utilised GBIF database but did not record a reference code. A new GBIF query was constructed, taking occurrences up to 2011 from Cuba, the Cayman Islands, the Bahamas and the norther coast of the Gulf of Mexico.

The original paper used occurrence data from HerpNet, which is now defunct, but VertNet (HerpNet's parent "entity") still keeps the records here:

<http://www.vertnet.org/resources/datatoolscode.html#t-tab1>

The original paper included additional occurrences from various sources such as Herpetological Review. These sources generally include few occurrences and with poor locality information. These additional sources were not included in this study.

The original paper states: "For modelling, we used georeferenced records of *A. sagrei* representing established populations only." While the VertNet data has a column labelled "establishmentmeans" with entries like "wild caught", "native" and "field photo" that could conceivably be used to determine established populations, there is no such identifier for the gbif data. The paper does not provide any method for determining what datapoints come from established populations, so all datapoints have been included.

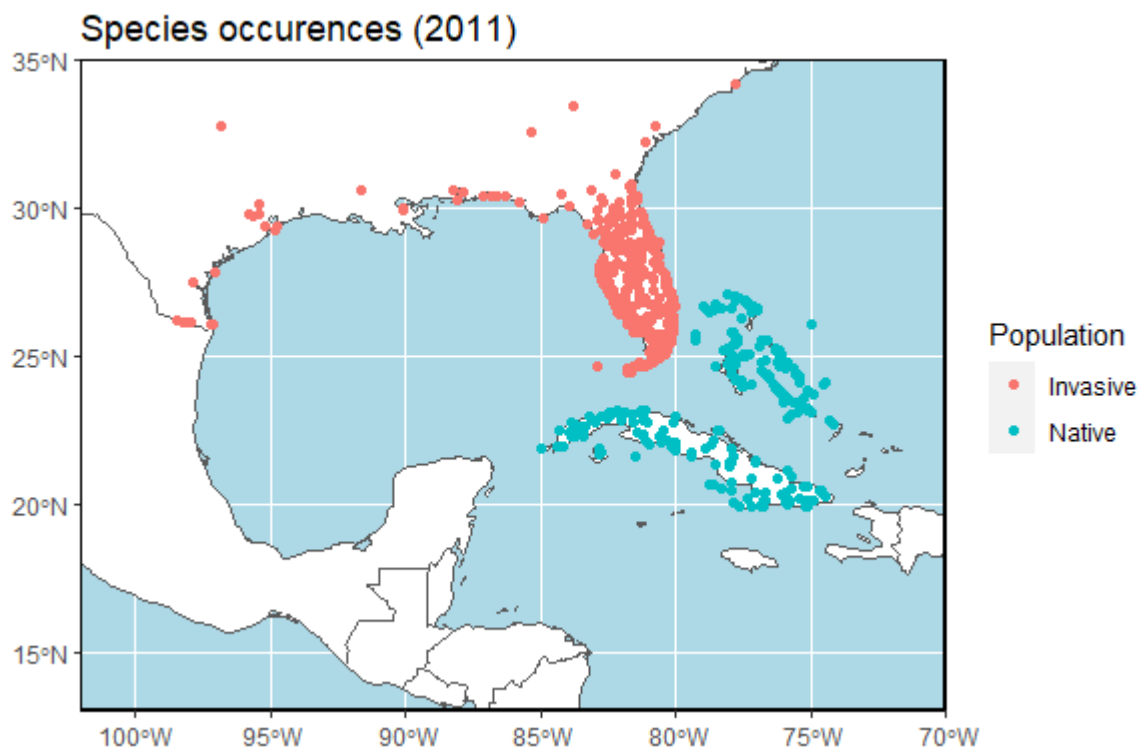
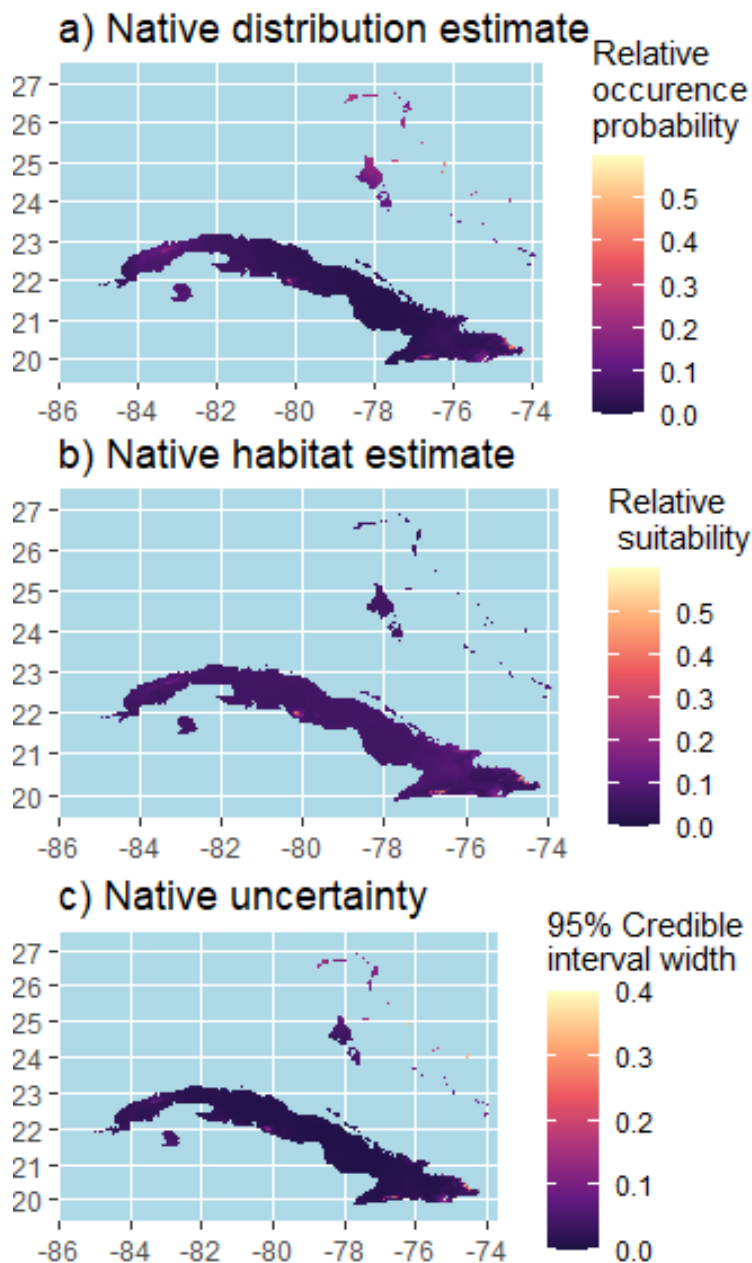


Figure a1: Occurrence records of *Anolis sagrei* up to 2011, native occurrences in blue and invasive occurrences in red.

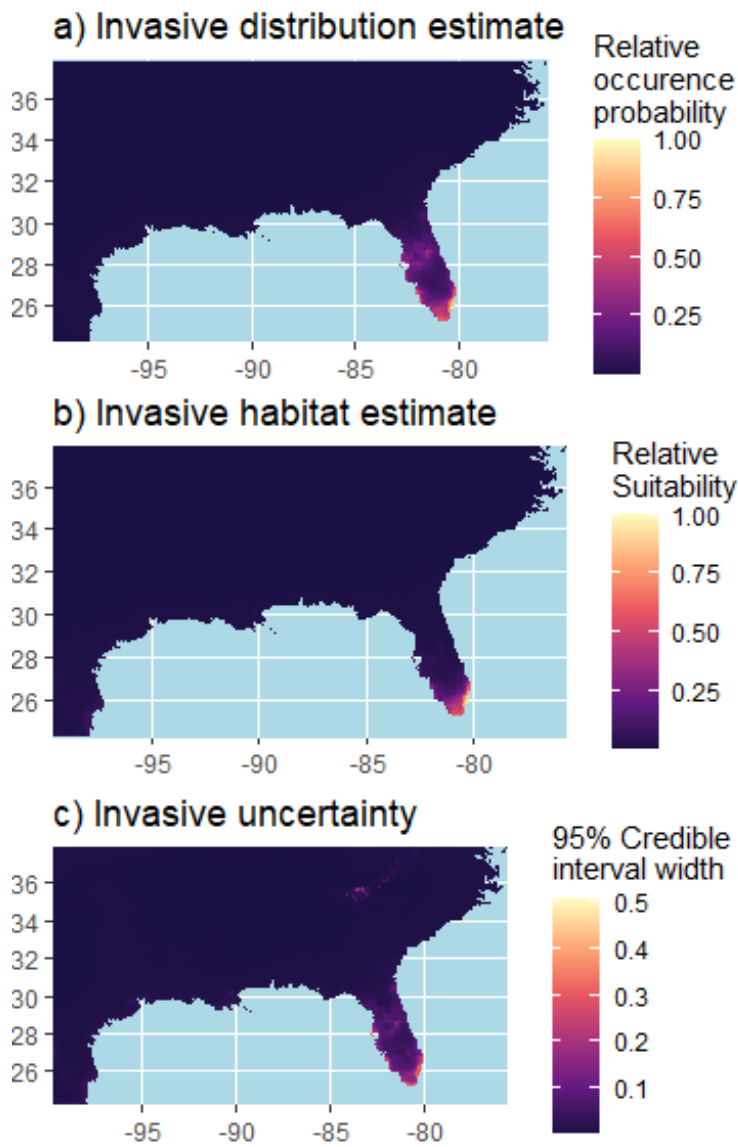


The distribution estimate shows the greatest signal in 3 clusters in Cuba: strongest in Baracoa, second in Turquino National Park (ca. 60% of max signal strength) and the weakest in Topes de Collantes (ca. 25%). A signal about the strength of the weakest cluster can be found on Grand Bahama and Andros.

The habitat estimate shows a similar pattern in Cuba compared to the realised niche, but with a somewhat more even distribution. The Bahamas has a low signal for the fundamental niche.

The uncertainty has the same pattern as the distribution estimate, with uncertainty at 0.05-0.1 around the clusters and reaching 0.6 in the centre of the clusters. The uncertainty on Grand Bahama is around 0.15.

Figure a2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model.



The distribution estimate shows the greatest signal on the southern coast of Florida and a middling signal around Tampa (ca. 30% of the maximum signal).

The habitat estimate shows the same signal in southern Florida, but without the signal around Tampa.

The uncertainty roughly matches the pattern of the distribution estimate, except for an uncertainty cluster around 36°N and 84°W with values around 0.4. The southern coast of Florida has uncertainty values around 0.3 and around Tampa the uncertainty was around 0.1.

Figure a3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model.

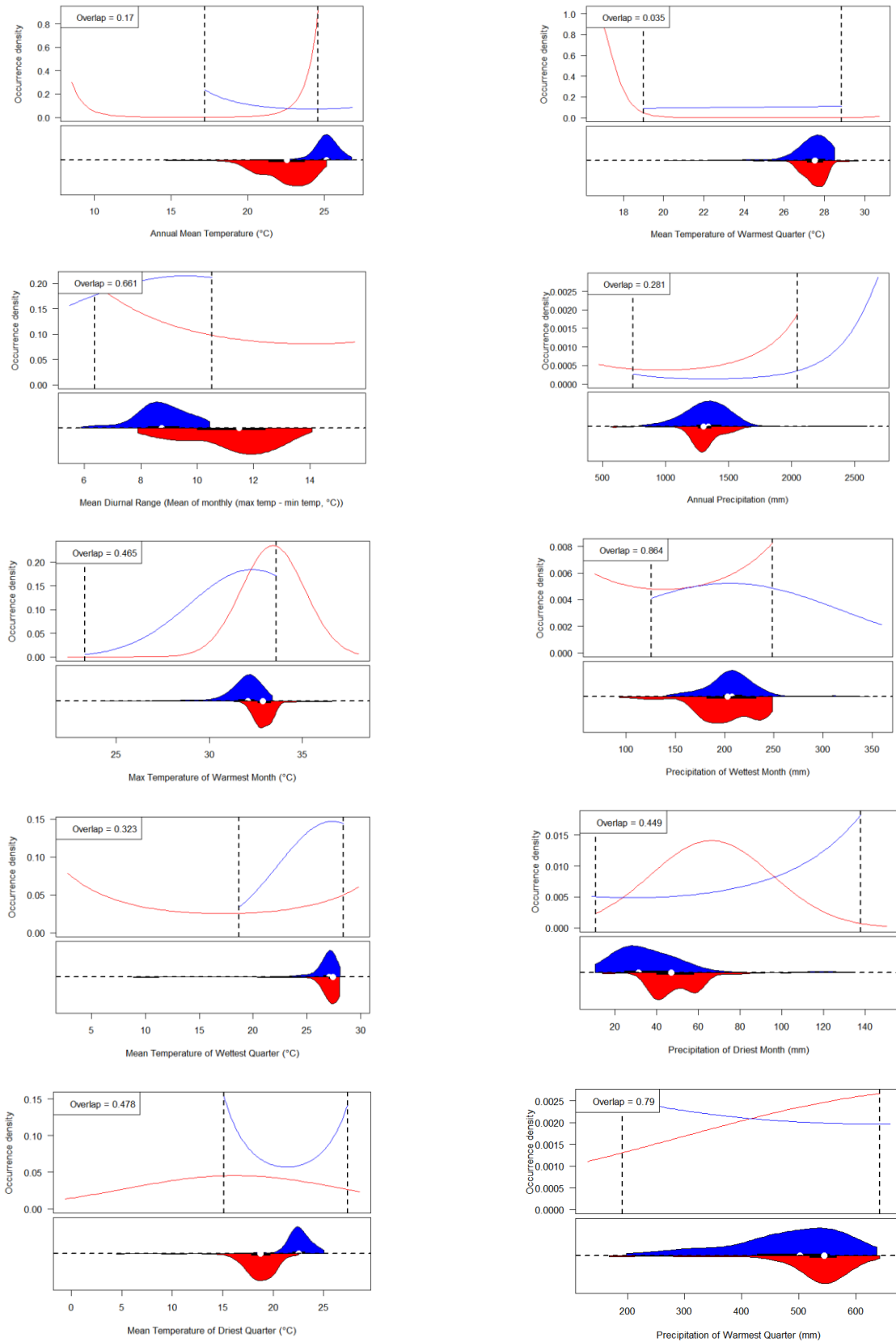


Figure a4: The response curves of the native (blue) and invasive (red) models. The area between the dashed lines represents the climate shared between the native and invasive study areas. Below are the distributions of occurrence records along the climatic gradient.

Many response curves in both the native and invasive model feature positive quadratic terms (U-shaped curves) that would lead to overpredictions if projected into non-analogue environments. The shared environmental space is mostly defined by the native study area as it tends to be smaller in the temperature gradients (this was somewhat expected due to its nature as a smaller island when compared to the invasive study area).

b) Battini et al. (2019)

Species: *Pleurobranchaea maculata*

Common name: Grey side-gilled slug

Habitat: Marine

Native range: The coasts of New Zealand and West-Australia

Introduced range: The Argentinian coast

Occurrence data: Data retrieved directly from the original paper (see Supplementary material, Table 1 in Battini et al. 2019)

Climate data: Data retrieved from Bio-Oracle database (Assis et al., 2018) via R package “sdmpredictors”. Resolution at 5 arcmin.

139.7, 177.9, -47.9, -27.4

Native study area extent (W, E, S, N): 139.7, 177.9, -47.9, -27.4

Invasive study area extent (W, E, S, N): -66.3, -53, -45, -34.1

Aggregation factor: Native = 2, Invasive = 1

Occurrence records (Occurrence cells after aggregation): Native = 190 (42), Invasive = 12 (8)

The Grey side-gilled slug (*Pleurobranchaea maculata*) is a marine scavenger/predator native to New Zealand and the eastern coast of Australia. In 2009 the species was discovered along the coast of Argentina and has since spread with a speed 330 km yr⁻¹. The species has been found down to a depth of 300m (Battini et al., 2019)

Limitations: Some species occurrences were recorded very close to land, causing them to be assigned cells with no climate data (land) in the aggregation process. Even without the aggregation, many occurrences were recorded inland (presumably in lakes and waterways) where there is no available climate data from the source used for the original study. The invasive occurrences have only 12 records, 4 were recorded inland.

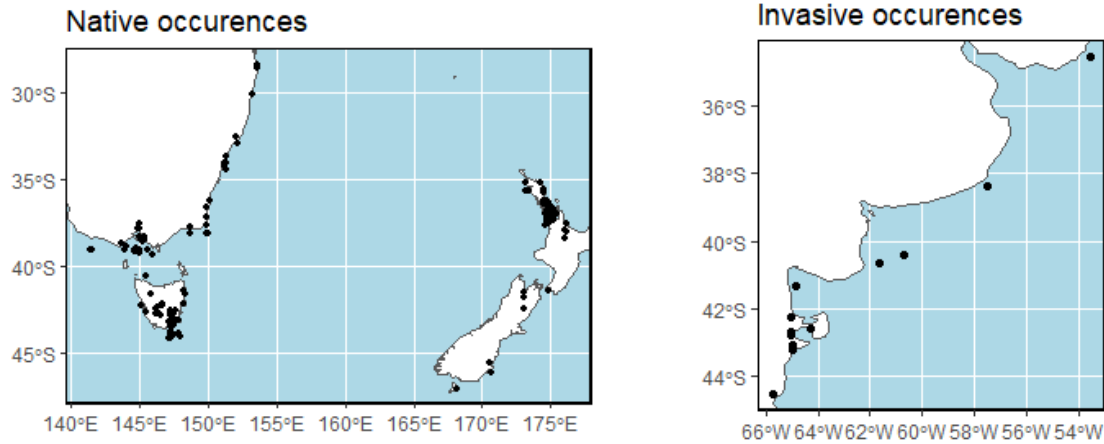
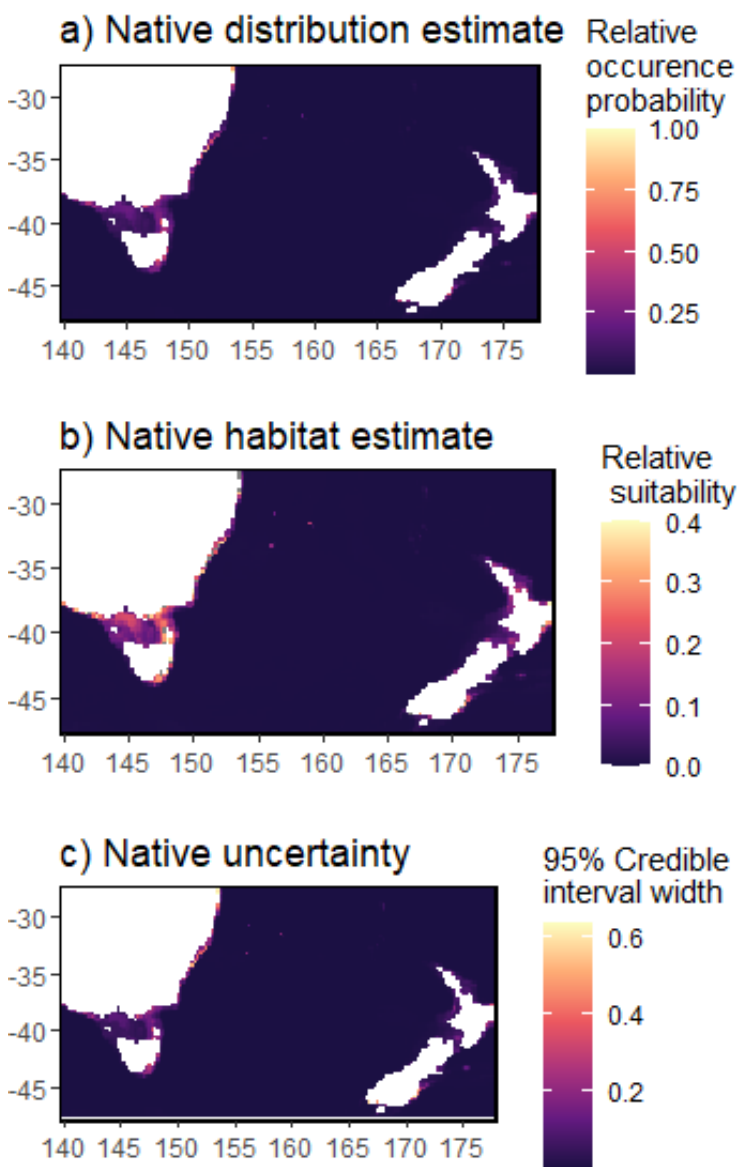


Figure b1: Occurrence records of *Pleurobranchaea maculata*, native occurrences (left) and invasive occurrences (right).

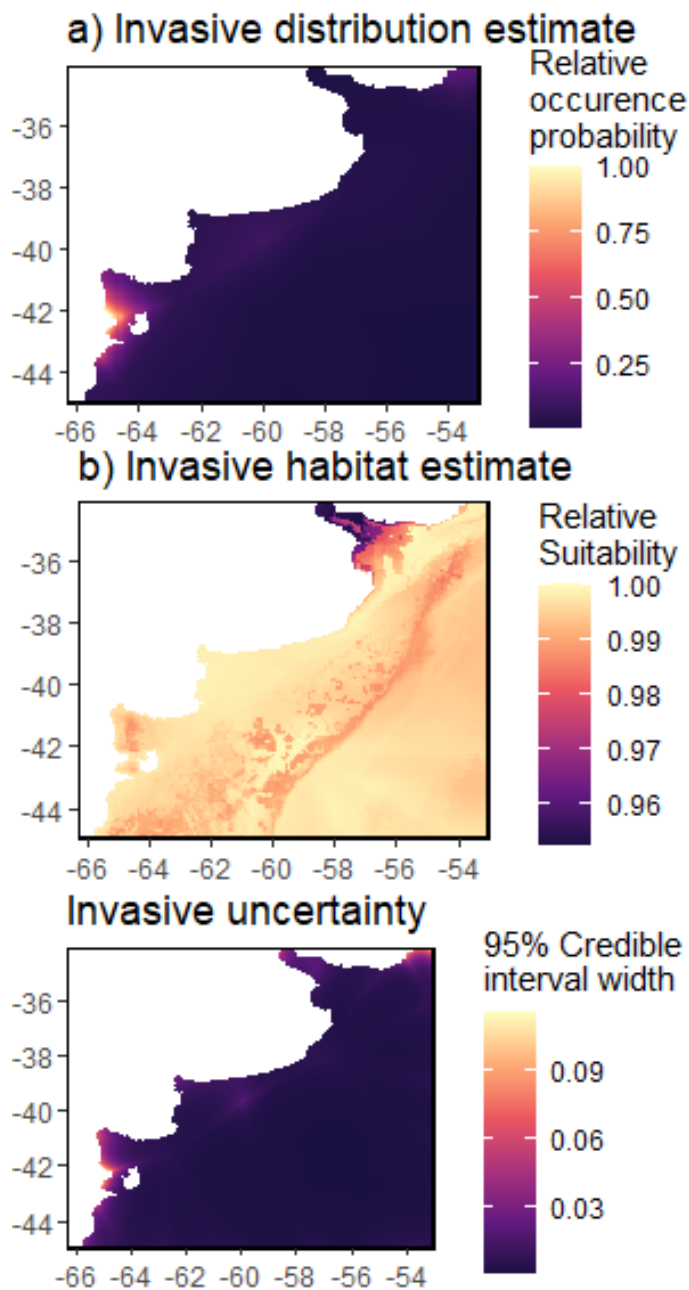


The distribution estimate shows the strongest signal along the eastern coast of Australia. The Bass-strait and the south coast of Tasmania along with the eastern and northern coast of New Zealand only showing up to 10% the strength of the maximum signal.

The habitat estimate is identical to the realised prediction (meaning there are no spatial effects).

The uncertainty shows the same pattern as the distribution estimate with uncertainties of up to 0.6 showing up on the Australian east coast, uncertainty around 0.2 on the southern coast and uncertainties at ca. 0.1 around New Zealand.

Figure b2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model.



The distribution estimate has the strongest signal clustered around the Valdes peninsula with a weaker signal (ca. 0.25 relative to maximum signal) evenly distributed along the coast.

The habitat estimate shows a relatively uniform signal across the entire study range (except Samborombón Bay which has a slightly weaker signal). This indicates a poor capture of the fundamental niche.

Uncertainty shows the same pattern as the distribution estimate with a maximum uncertainty at around 0.1.

Figure b3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model.

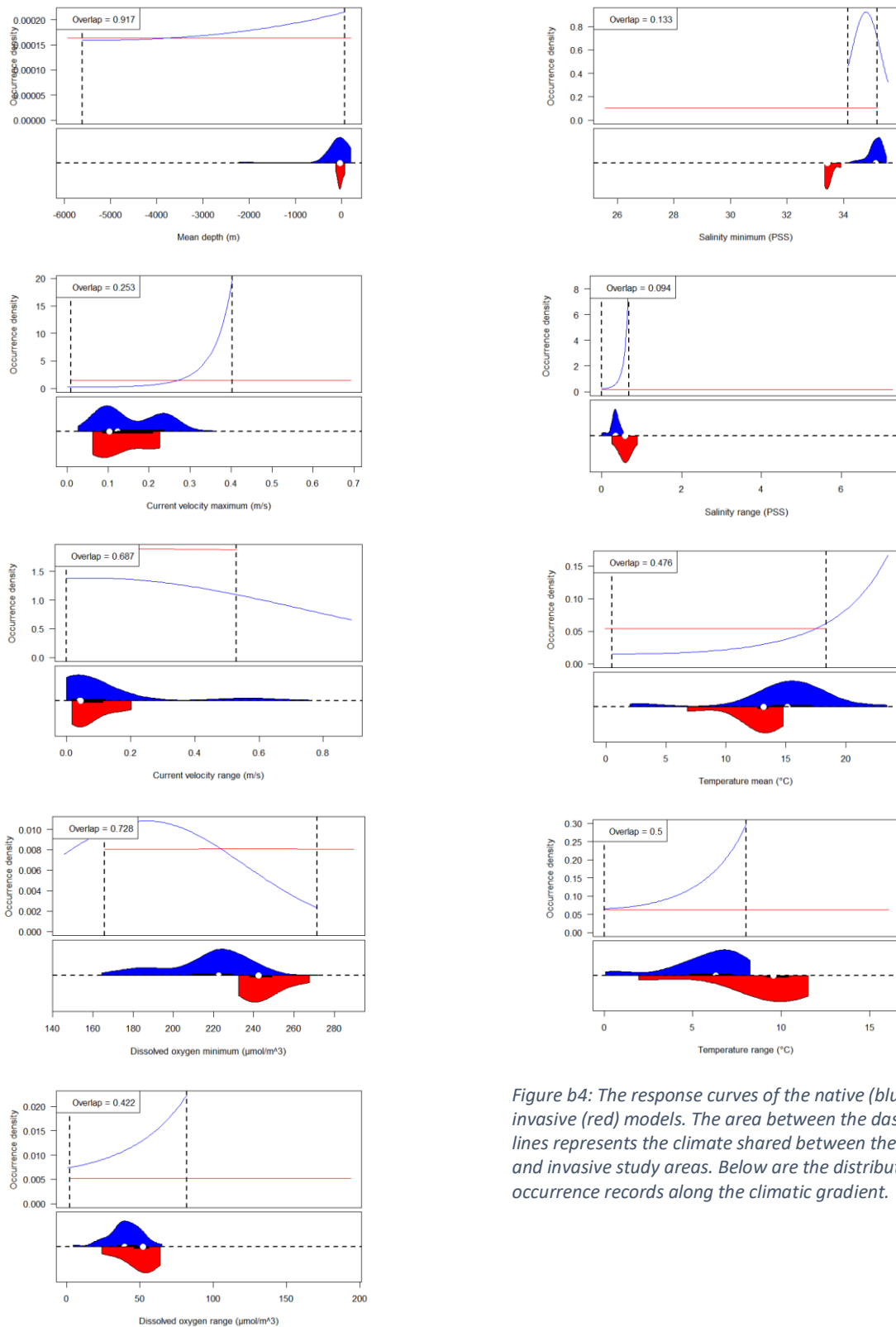


Figure b4: The response curves of the native (blue) and invasive (red) models. The area between the dashed lines represents the climate shared between the native and invasive study areas. Below are the distributions of occurrence records along the climatic gradient.

Many curves in the native model have positive quadratic terms, marking them as unsuitable for projection onto non-analogous environments. All response curves in the invasive range are flat, indicating a total failure to capture the fundamental niche. The occurrence distributions might be somewhat warped toward extremes by the aggregation (especially the mean depth since it has been noted that the species have not been found deeper than 300m).

c) Bidinger et al. (2012)

Species: *Harmonia axyridis*

Common name: Harlequin ladybird

Habitat: Terrestrial

Native range: Japan, South & North-Korea, eastern China, Mongolia, southern Siberia and northern Kazakhstan.

Introduced range: Austria, Belgium, England, France, Germany, Ireland, Italy & Switzerland.

Occurrence data: Data retrieved directly from the original paper (see Appendix 1 of Bidinger et al. 2012)

Occurrence records (Occurrence cells after aggregation): Native = 60 (52), Invasive = 171 (162)

Climate data: Data retrieved from WORLDCLIM database, version 2.1. Resolution: 2.5 arcmin

Aggregation factor: Native = 3, Invasive = 2

Native study area extent (W, E, S, N): 86.7, 153.7, 21.9, 57.7

Invasive study area extent (W, E, S, N): -7.1, 16.5, 42.4, 58.2

Predictors:

- bio1: Annual Mean Temperature (°C)
- bio3: Isothermality (BIO2/BIO7) (×100)
- bio5: Max Temperature of Warmest Month (°C)
- bio6: Min Temperature of Coldest Month (°C)
- bio7: Temperature Annual Range (BIO5-BIO6) (°C)
- bio8: Mean Temperature of Wettest Quarter (°C)
- bio9: Mean Temperature of Driest Quarter (°C)
- bio15: Precipitation Seasonality (Coefficient of Variation)
- bio16: Precipitation of Wettest Quarter (mm)
- bio17: Precipitation of Driest Quarter (mm)

The harlequin ladybird (*Harmonia axyridis*) is an insect native to east Asia and intentionally introduced to Europe in 1982 for pest control. The species has become invasive, harming native insects and grape production (Bidinger et al., 2012).

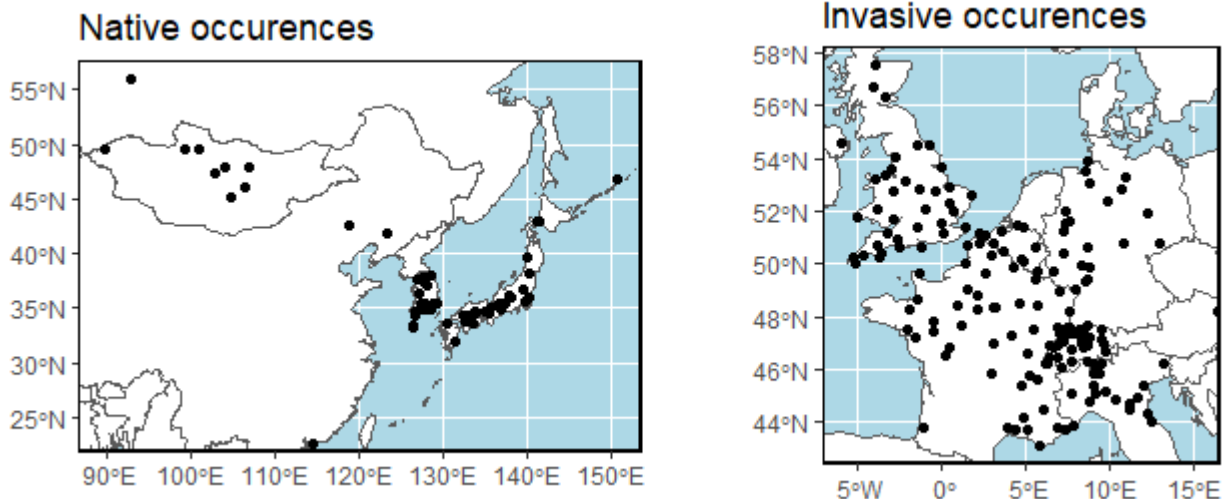


Figure c1: Occurrence records of *Harmonia axyridis*, native occurrences (left) and invasive occurrences (right).

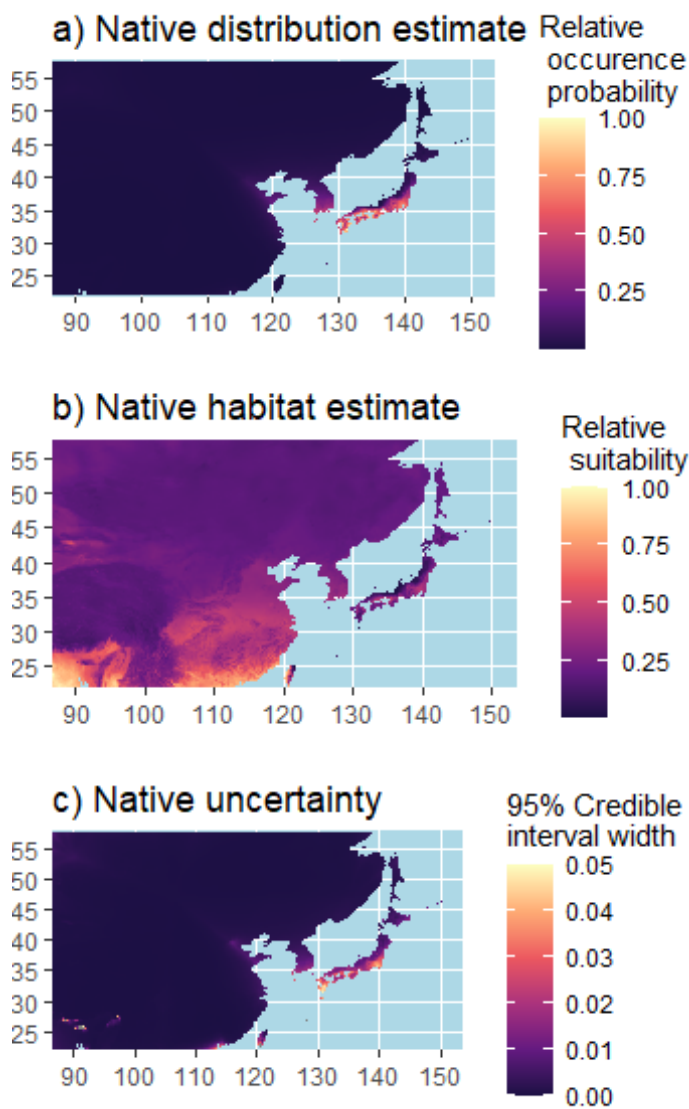
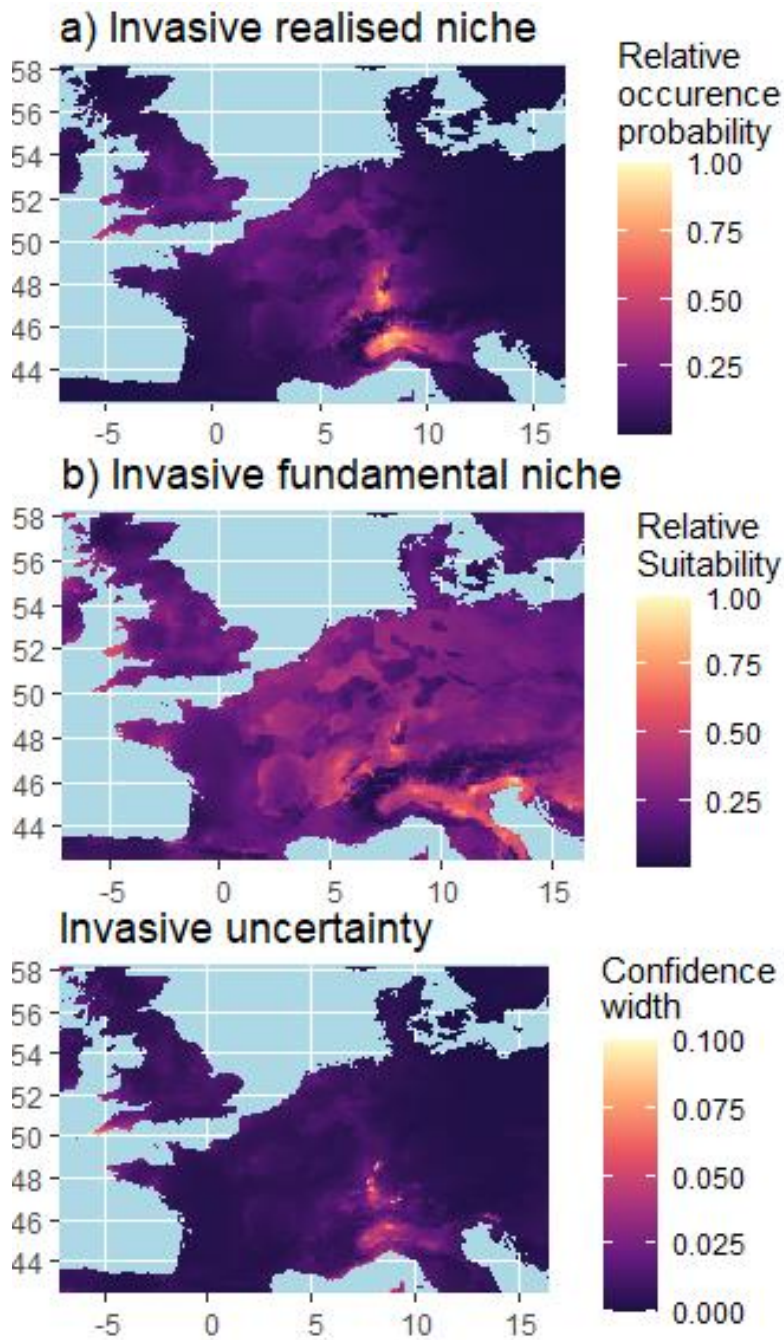


Figure c2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model.

The distribution estimate shows a pattern in accordance with the actual occurrences in Japan and Korea but fails to show any response in Mongolia. The strongest signal is in south Japan. The signal in South Korea is about 25% of the maximum signal strength. There is also a low (10%) signal in the Shandong province of China, where there are no recorded occurrences.

The habitat estimate shows a higher suitability in southern China and Bangladesh, with lower signal (<50%) in locations of recorded occurrences.

The uncertainty is minimal in continental Asia, but in Japan uncertainty reaches 0.05. Korea and Shandong have uncertainties around 0.02. Around Bangladesh there is an area where the confidence interval is 1 (marked in grey in figure c2) meaning total uncertainty.



The distribution estimate shows a pattern in accordance with the actual occurrences with hotspots in Switzerland and northern Italy. The model shows weak predictions at the western and eastern edges of the invasive population. Signals in the southern UK are around 40% of maximum signal strength.

The habitat estimate shows a similar pattern to the distribution, though somewhat less concentrated around the hotspots.

The uncertainty is concentrated in the hotspots shown in the realised prediction and has a size of 0.1-0.05 around the hotspots.

Figure c3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model.

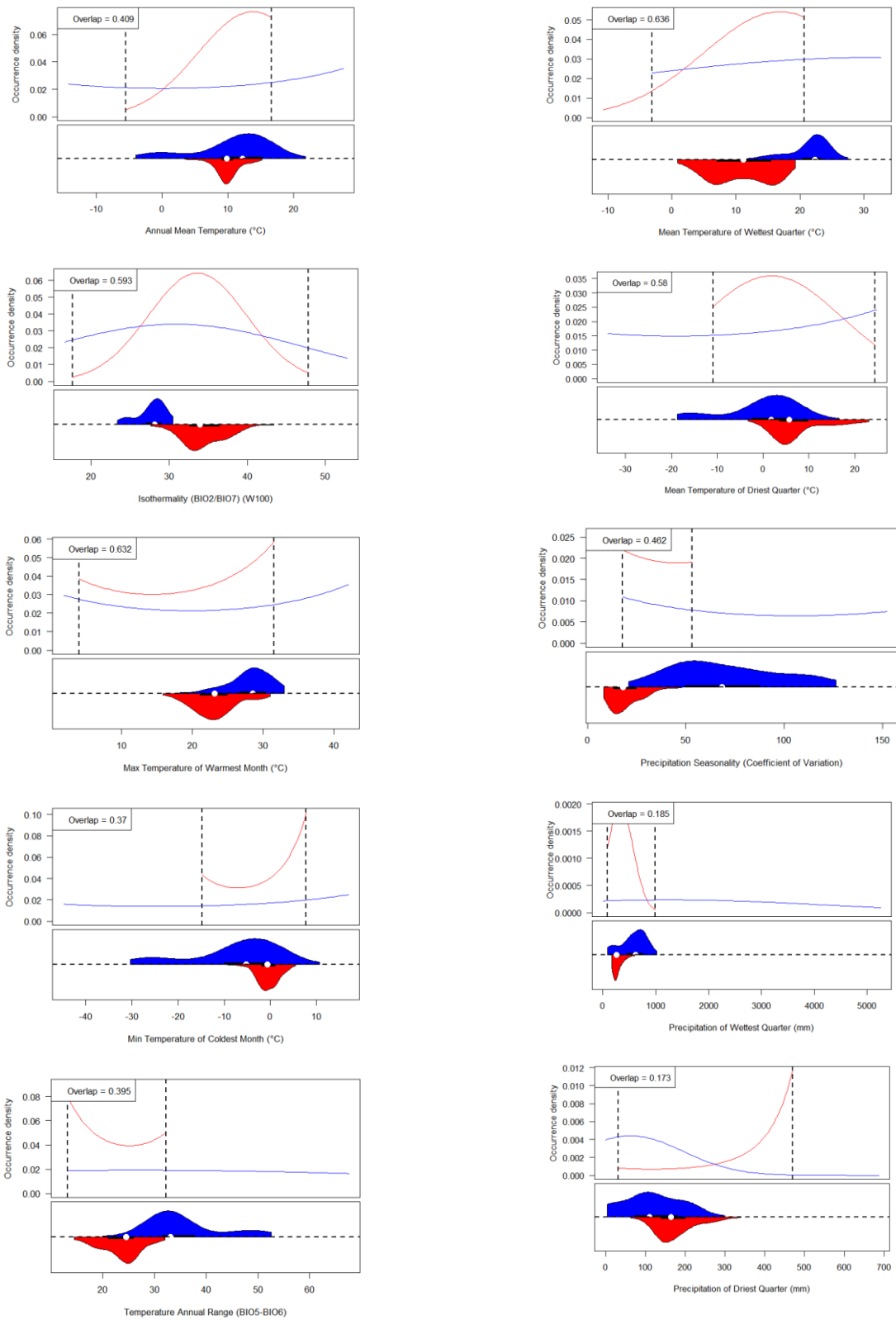


Figure c4: The response curves of the native (blue) and invasive (red) models. The area between the dashed lines represents the climate shared between the native and invasive study areas. Below are the distributions of occurrence records along the climatic gradient.

The response curves of the native model are mostly flat indicating a poor capture of the fundamental niche.

d) Manzoor et al. (2020)

Species: *Rhododendron ponticum*

Common name: Common rhododendron or Pontic rhododendron

Habitat: Terrestrial

Native range: Spain and Portugal

Introduced range: Great Britain

Occurrence data: Data recreated using Gbif, including only records up until 2020. GBIF.org (23 March 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.4yurdw>

Occurrence records (Occurrence cells after aggregation): Native = 237 (76), Invasive = 40853 (8446)

Climate data: Data retrieved from WORLDCLIM database, version 2.1. Resolution: 30 arcsec (Fick and Hijmans, 2017)

Aggregation factor: Native = 2, Invasive = 2

Native study area extent (W, E, S, N): -9, -3.4, 35.8, 41

Invasive study area extent (W, E, S, N): -7.6, 1.9, 49.9, 59.4, Excluded: (-7.6, -5.3, 51.7, 55.26), (1, 1.9, 49.9, 51) & (-7.6, -6.6, 55, 55.5)

Predictors:

- bio1: Annual Mean Temperature (°C)
- bio6: Min Temperature of Coldest Month (°C)
- bio12: Annual Precipitation (mm)
- bio14: Precipitation of Driest Month (mm)

Rhododendron ponticum (L.) is a perennial, evergreen shrub that generally invades woodlands. Introduced to the British Isles from mainland Europe in the eighteenth century, it has since become a widespread invasive species that impacts nearby flora by releasing chemicals into the soil that prevents germination. It's success as an invasive have largely been attributed to large reproductive output and tolerance for shade and nutrient poor soil. Genetic markers have shown that the *Rhododendron ponticum* in Britain primarily originate from the Iberian Peninsula (Manzoor et al., 2020).

Limitations:

Manzoor collected the data from GBIF but did not record a reference code. A new GBIF query was constructed, taking occurrences up to 2020 in the Iberian Peninsula and on Great Britain (<https://doi.org/10.15468/dl.4yurdw>).

The invasive study area was Great Britain. Any landmass other than Great Britain was not considered in the original study and all nearby landmasses contained occurrences. To minimise undesired data within the invasive study range no buffer zone was added and Parts of the polygon not including Great Britain was excluded (See code on GitHub).

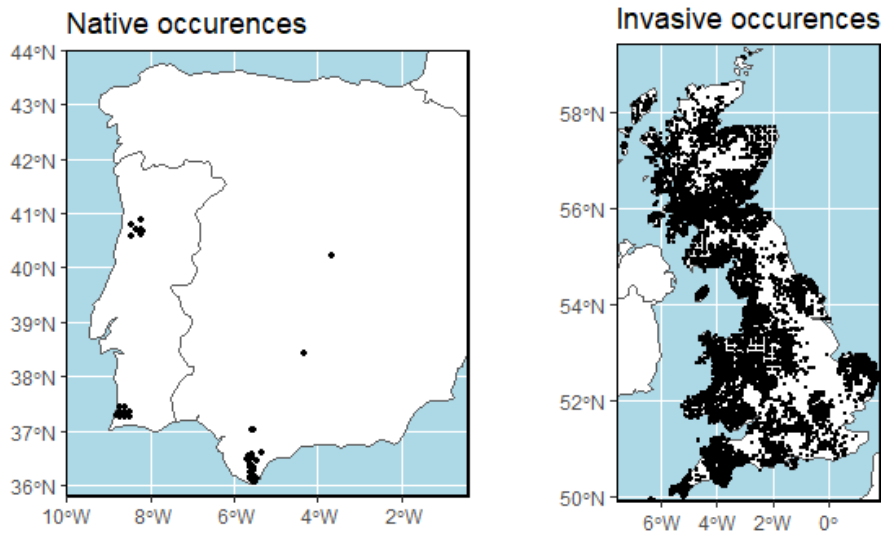
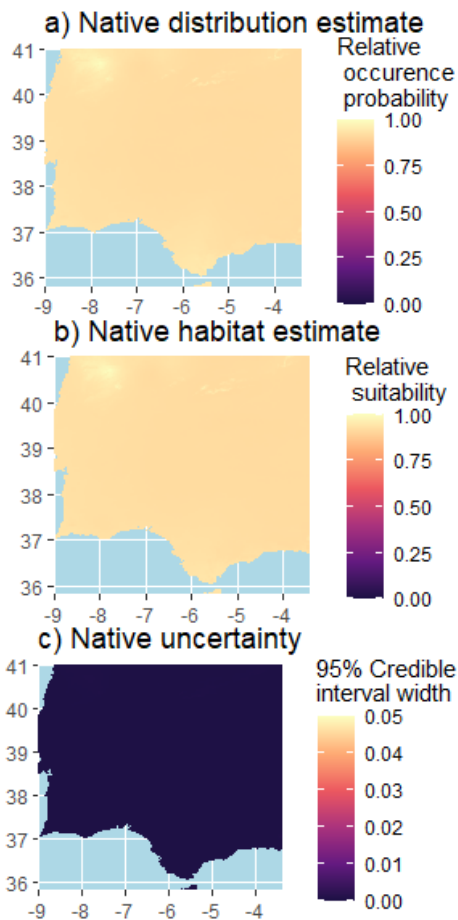


Figure d1: Occurrence records of *Rhododendron ponticum*, native occurrences (left) and invasive occurrences (right).

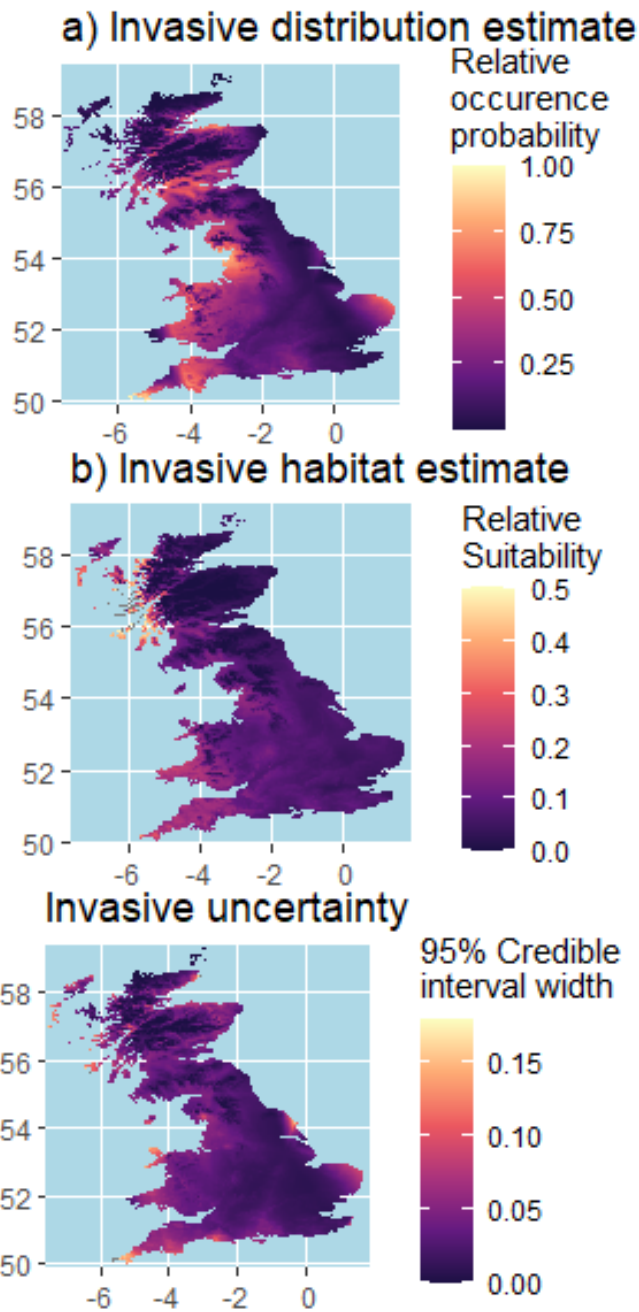
Occurrence records were scant in the native study area with most occurrences concentrated in 3 locations leading to a small effective sample size (number of occupied cells) and prevalence for the native model. The Invasive record on the other hand is the richest among all the datasets and occurrences are distributed widely over the study area.



The distribution and habitat estimates have the same pattern with a roughly uniform signal strength across the entire native study range with a slightly stronger signal in the northwest.

Uncertainty is at negligible levels across the entire study range.

Figure d2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model.



The distribution estimate shows a pattern like what can be observed in the occurrence records with a preference for lowlands.

The habitat estimate has the strongest signal on the Hebrides and the nearby Scottish coast. Otherwise matches the distribution estimate with the western part of England & Wales at values around 0.2 and eastern England around 0.1.

Uncertainty has a hotspot in the southwestern tip of Great Britain, reaching uncertainties up to 0.3, other western parts of Wales and the Hebrides have an uncertainty round 0.5. The rest of the introduced range follows the niche predictions with uncertainties around 0.1.

Figure d3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model.

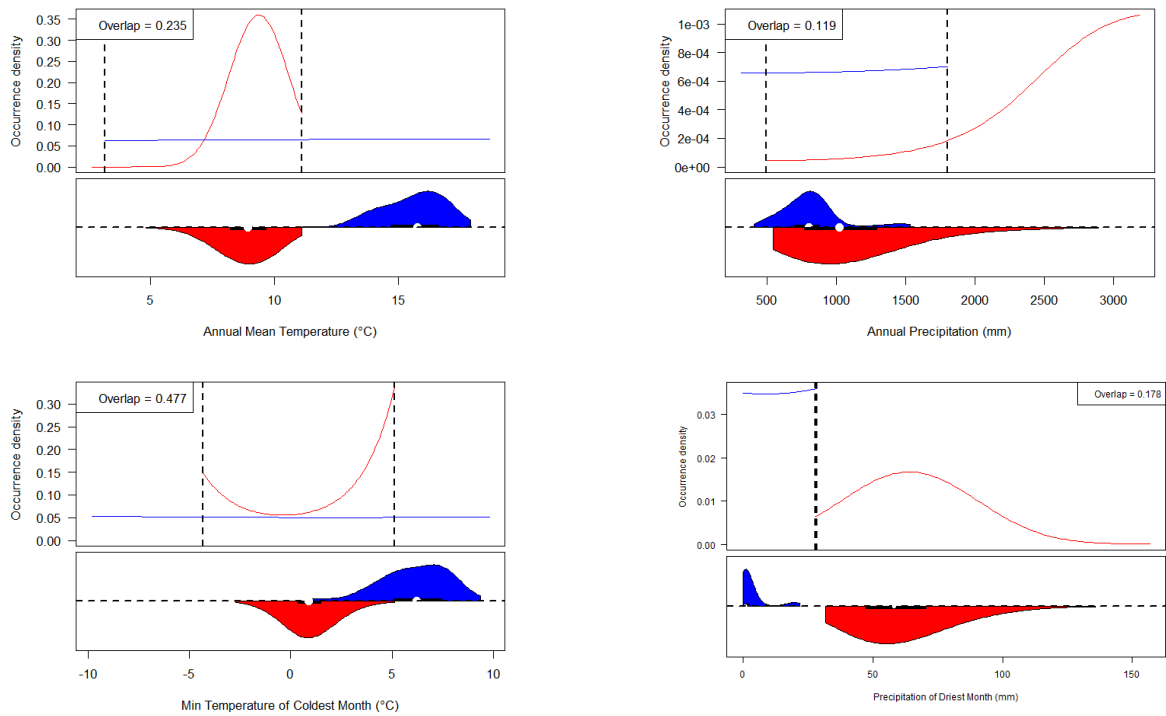


Figure d4: Standardised response curves of native (blue) and invasive (red) models. The area between the dashed lines is the climatic space shared in both native and invasive study areas. The violin plots below show the distribution of occurrence records along the climatic gradient.

The native model shows flat response curves indicating a null signal. The model is barely capturing any climatic response. There is almost no shared climate along the “precipitation of driest month” gradient, the overlap measure is highly unreliable. Along the annual precipitation gradient, the invasive response curve shows the opposite pattern from occurrence records distribution, but this may be caused by a high occupancy rate of a rare climate (high precipitation) in the study area.

e) Morehouse and Tobler (2013)

Species: *Faxonius rusticus* (formerly *Orconectes rusticus*)

Common name: Rusty Crayfish

Native range: The Ohio River and the lower Maumee River drainages.

Introduced range: The USA and southern Canada

Occurrence data: Data retrieved from 2 sources:

1. U.S. Geological Survey, 2023, Specimen observation data for *Faxonius rusticus* (Girard, 1852), Nonindigenous Aquatic Species Database, Gainesville, FL, <https://nas.er.usgs.gov/queries/CollectionInfo.aspx?SpeciesID=21>
2. Illinois Natural History Survey - Crustacean Collection. Occurrence dataset (ID: 257ea482-2d6f-4cf4-846d-8efeb58f6727) https://biocoll.inhs.illinois.edu/portal/content/dwca/INHS-CRUSTACEAN_DwC-A.zip accessed via the INHS Collections Data Portal, biocoll.inhs.illinois.edu/portal, 2023-06-01).

Climate data: Data retrieved from Data retrieved from WORLDCLIM database, version 2.1.

Resolution: 30 arcsec (Fick and Hijmans, 2017)

Aggregation factor: Native = 3, Invasive = 13

Native study area extent (W, E, S, N): -87.3, -81.1, 36.6, 41.9, Excluded: (-82.6, -81.7, 38, 38.9)

Invasive study area extent (W, E, S, N): -122.5, -66.7, 33, 50.9, Excluded: (-82.6, -80, 36, 38.9)

Occurrence records (Occurrence cells after aggregation): Native = 152 (121), Invasive = 2465 (891)

Predictors:

- bio3: Isothermality (BIO2/BIO7) (×100) (°C)
- bio5: Max Temperature of Warmest Month (°C)
- bio7: Temperature Annual Range (BIO5-BIO6) (°C)
- bio9: Mean Temperature of Driest Quarter (°C)
- bio10: Mean Temperature of Warmest Quarter (°C)
- bio15: Precipitation Seasonality (Coefficient of Variation)
- bio16: Precipitation of Wettest Quarter (mm)
- bio17: Precipitation of Driest Quarter (mm)
- bio18: Precipitation of Warmest Quarter (mm)
- bio19: Precipitation of Coldest Quarter (mm)

The rusty crayfish (*Faxonius rusticus*, formerly *Orconectes rusticus*) is a crustacean native to the Ohio River and the lower Maumee River drainages, but has been intentionally and accidentally introduced in bodies of water across the United States and Canada (Morehouse and Tobler, 2013).

Limitations:

In 2017 the genus (*Orconectes rusticus*, the modelled species in Morehouse and Tobler (2013)) was re-evaluated and changed to *Faxonius*, this included unifying several species. The databases now contain the expanded definition of the species, making it impossible to distinguish which occurrences belong to the old *Orconectes rusticus*. Since the native and invasive populations were so

close it was not possible to create a 5% buffer at the edges of the native range, the border was made prioritizing excluding the introduced range.

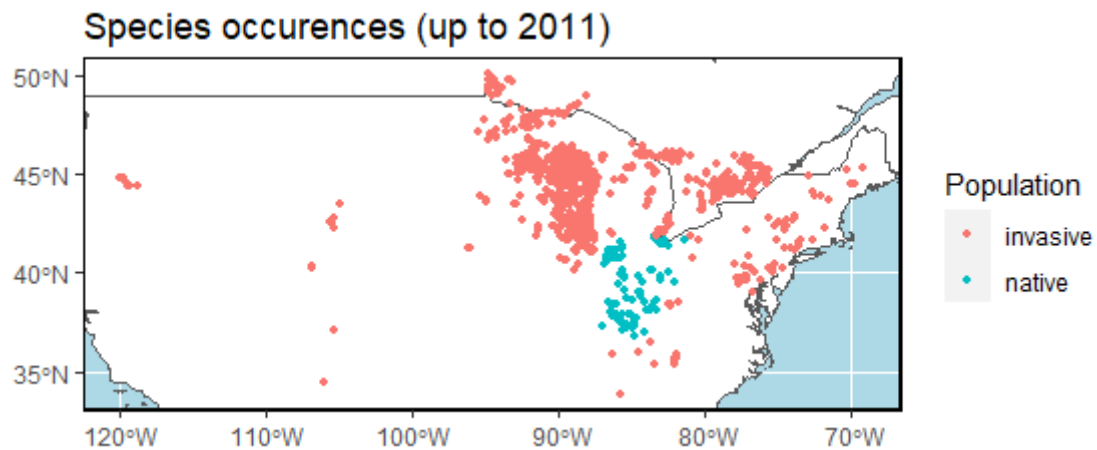


Figure e1: Occurrence records of *Faxonius rusticus*, native occurrences (left) and invasive occurrences (right).

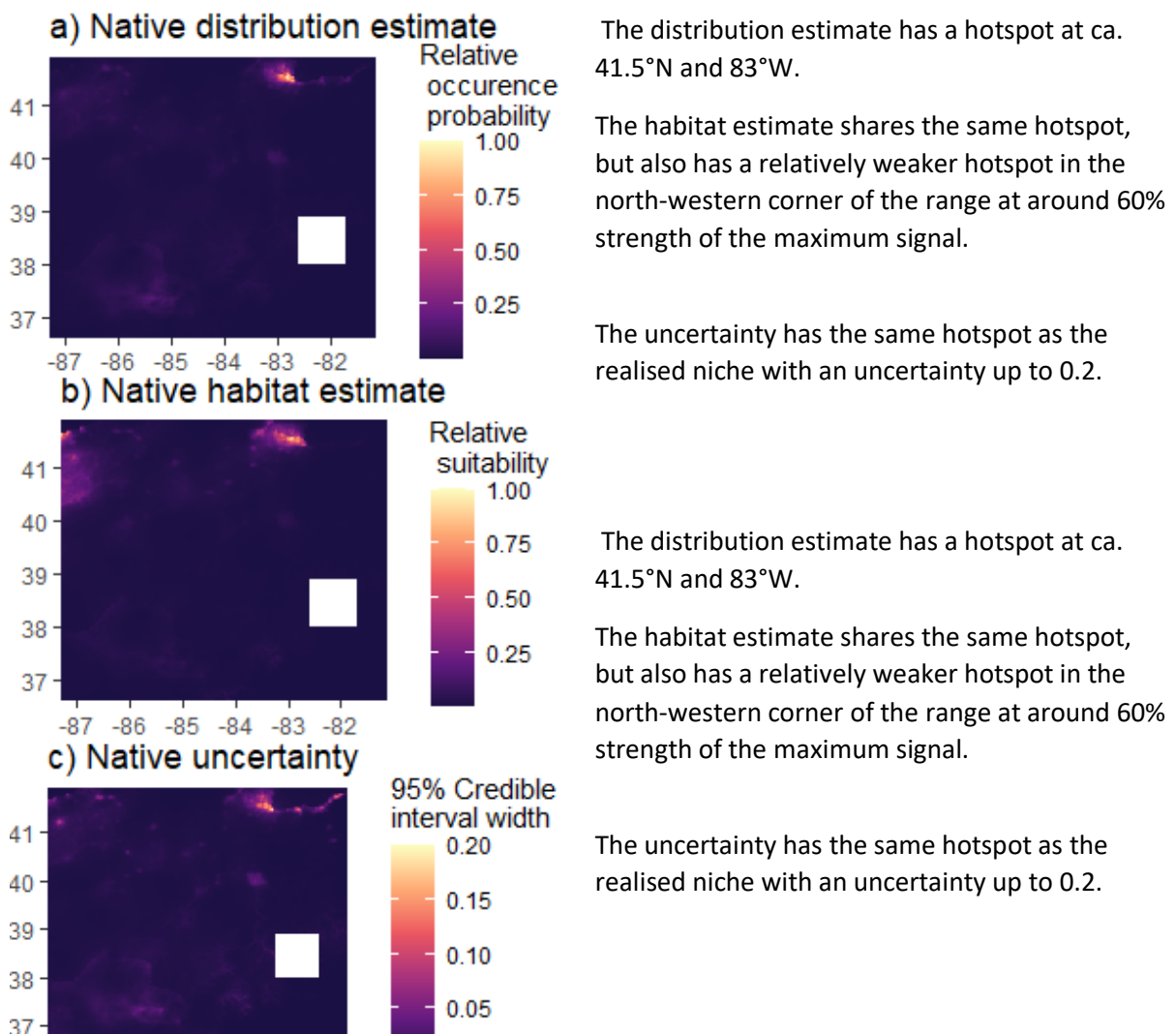


Figure e2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model. The white square is the exclusion of an area with invasive occurrences.

The distribution estimate has a hotspot at ca. 41.5°N and 83°W.

The habitat estimate shares the same hotspot, but also has a relatively weaker hotspot in the north-western corner of the range at around 60% strength of the maximum signal.

The uncertainty has the same hotspot as the realised niche with an uncertainty up to 0.2.

The distribution estimate has a hotspot at ca. 41.5°N and 83°W.

The habitat estimate shares the same hotspot, but also has a relatively weaker hotspot in the north-western corner of the range at around 60% strength of the maximum signal.

The uncertainty has the same hotspot as the realised niche with an uncertainty up to 0.2.

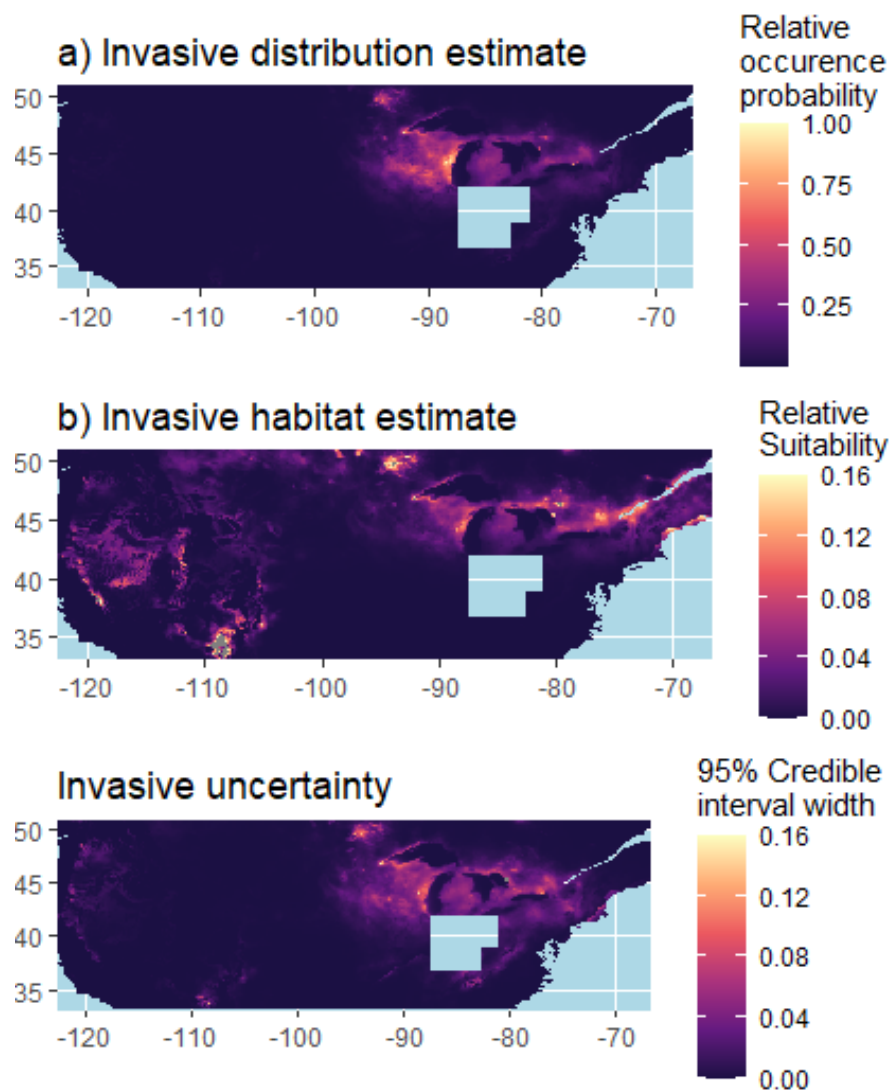


Figure e3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model. The blue polygon is the exclusion of an area containing the native study area.

Most signals in the distribution estimate are concentrated around the Great Lakes, with the strongest hotspot west of Lake Michigan. Between and west of the Great Lakes the signal is at around 30% of maximum signal strength.

The habitat estimate has its maximum signal in a hotspot around 33°N and 109°W (scale in figure was reduced to show patterns beyond this one hotspot). Otherwise, a weaker signal can be seen around the Great Lakes and the Sierra Nevada Mountain range with values up to 16% of maximum signal strength.

Uncertainty kept to a similar pattern to the distribution estimate with the highest uncertainty reaching 0.36 in a spot on the east edge of Lake Huron. Otherwise, the uncertainty around the Great Lakes is around 0.1.

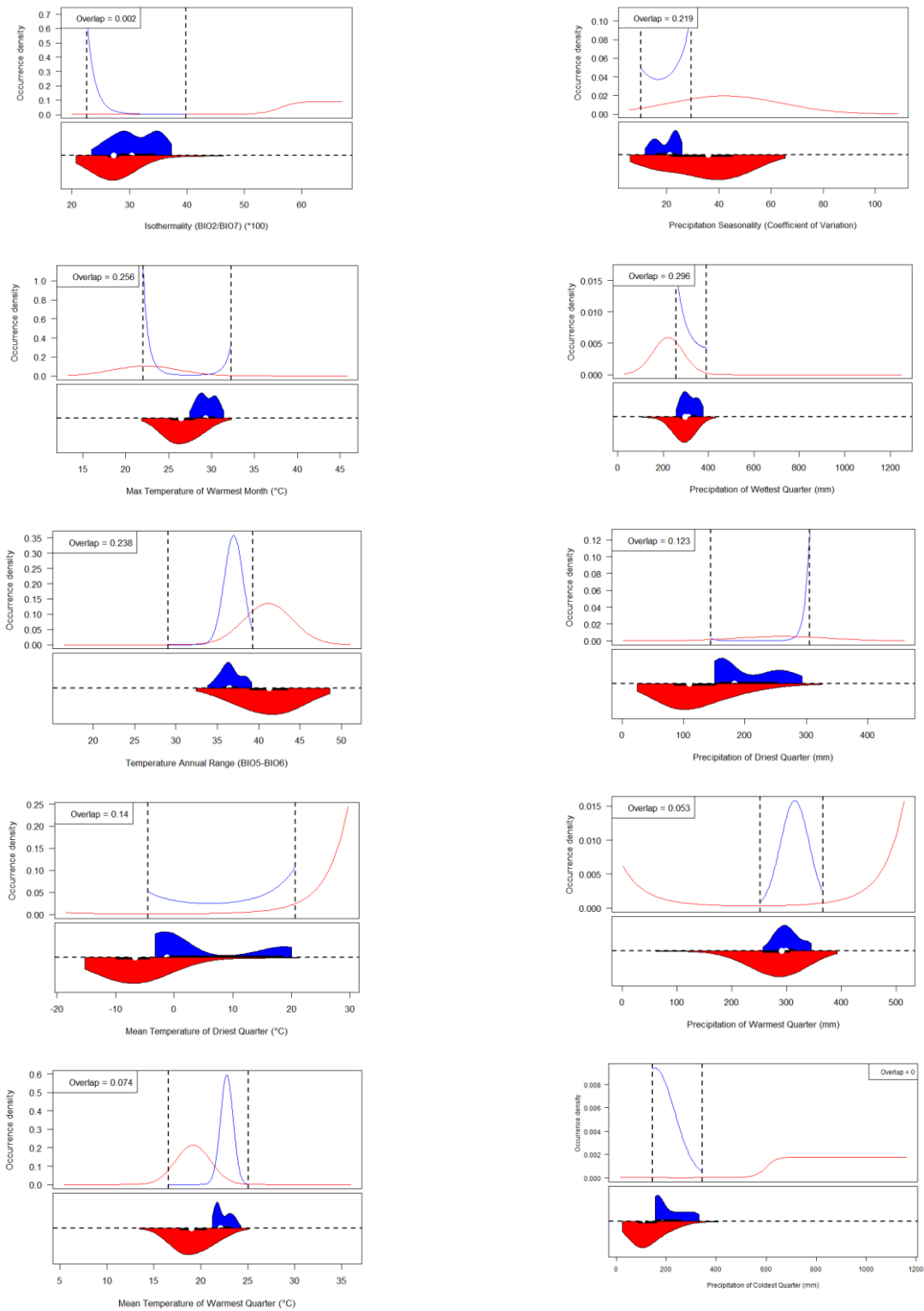


Figure a4: The response curves of the native (blue) and invasive (red) models. The area between the dashed lines represents the climate shared between the native and invasive study areas. Below are the distributions of occurrence records along the climatic gradient.

There is a small proportion of shared climate due to the size difference between native and invasive study areas. The distribution of occurrence records in the native study area is also limited by the lack of available climate. The native model features many curves with a positive quadratic term, marking it as unfit for projection.

f) Rödder and Lötters (2009)

Species: *Hemidactylus turcicus*

Common name: Mediterranean house gecko

Native range: The Mediterranean basin

Introduced range: North & Central America

Occurrence data: Data recreated from 2 sources, including only records up until 2009:

1. GBIF.org (28 March 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.ek48jv>
2. David Bloom (2016). VertNet_Retilia_Sept2016. CyVerse Data Commons. DOI 10.7946/P2Z59J

Climate data: Data retrieved from WORLDCLIM database, version 2.1. Resolution: 5 arcminutes (Fick and Hijmans, 2017)

Aggregation factor: Native = 2, Invasive = 2

Native study area extent (W, E, S, N): -11, 39.3, 21, 46.7

Invasive study area extent (W, E, S, N): -120.2, -75.9, 15.3, 37.8

Occurrence records (Occurrence cells after aggregation): Native = 1842 (501), Invasive = 1480 (242)

Predictors:

- bio1: Annual Mean Temperature (°C)
- bio2: Mean Diurnal Range (Mean of monthly (max temp - min temp)) (°C)
- bio3: Isothermality (BIO2/BIO7) (×100)
- bio4: Temperature Seasonality (standard deviation ×100)
- bio5: Max Temperature of Warmest Month (°C)
- bio6: Min Temperature of Coldest Month (°C)
- bio7: Temperature Annual Range (BIO5-BIO6) (°C)
- bio8: Mean Temperature of Wettest Quarter (°C)
- bio9: Mean Temperature of Driest Quarter (°C)
- bio12: Annual Precipitation (mm)
- bio13: Precipitation of Wettest Month (mm)
- bio14: Precipitation of Driest Month (mm)
- bio15: Precipitation Seasonality (Coefficient of Variation)
- bio18: Precipitation of Warmest Quarter (mm)
- bio19: Precipitation of Coldest Quarter (mm)

The Mediterranean house gecko (*Hemidactylus turcicus*) is a reptile native to the Mediterranean basin, but has been introduced to North America no later than 1910, when it was observed in Florida. Since then the invasive has spread across the southern half of the USA, Mexico and Panama (Rödder and Lötters, 2009).

Niche overlap analysis:

Rödger and Lötters (2009) and Rödger and Lötters (2010) used niche overlap analysis that can be described as a hybrid between the geographic and univariate comparison methods. They derived individual response curves from an SDM and then mapped each response curve onto the invasive study range individually. Then a geographic overlap comparison was performed per environmental variable.

Limitations:

The original paper utilised gbif but did not record a reference code. A new gbif query was constructed, taking occurrences up to 2008 around the Mediterranean and in Central & North America.

The original paper used occurrence data from HerpNet, which is now defunct, but VertNet (HerpNet's parent "entity") still keeps the records here:

<http://www.vertnet.org/resources/datatoolscode.html#t-tab1>

The original paper included additional occurrences from various sources such as Herpetological Review. These sources generally include few occurrences and with poor locality information. These additional sources were not included in this study.

The original paper made 9 datasets per range with different combinations of climate variables (We used dataset 1).

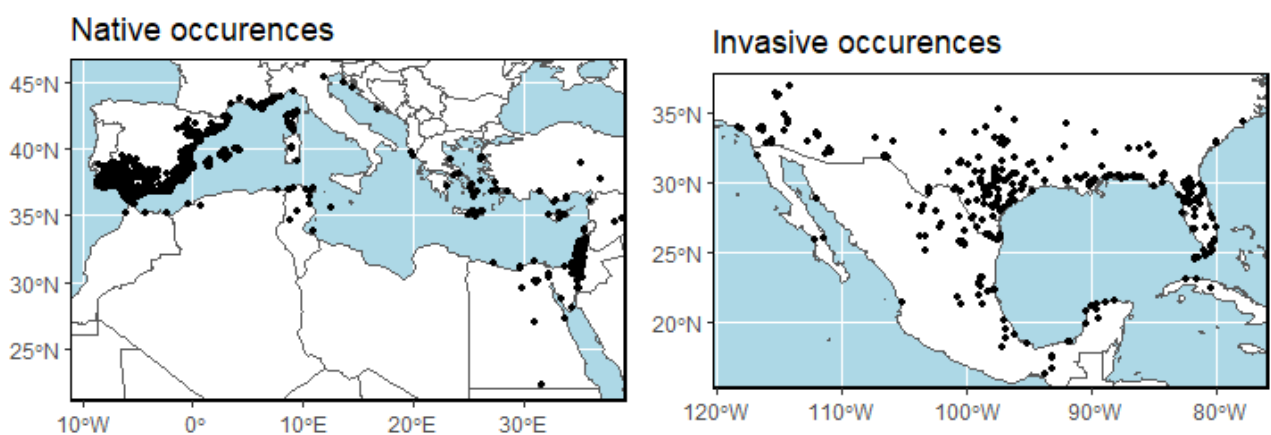
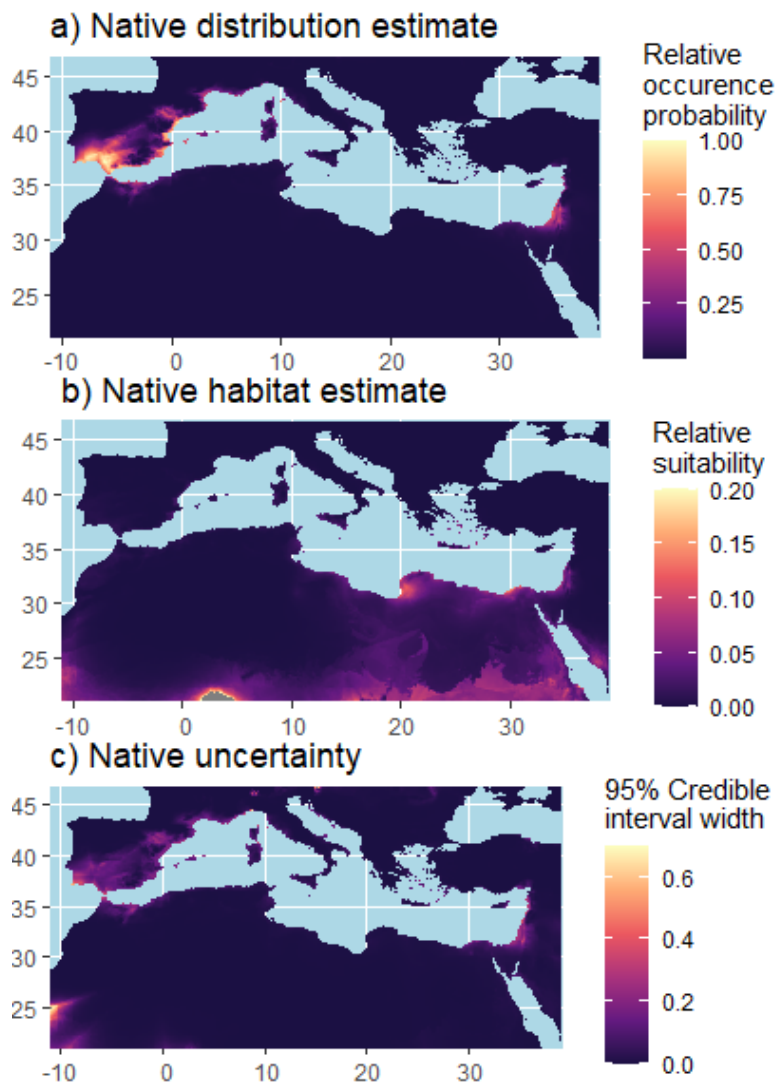


Figure f1: Occurrence records of *Hemidactylus turcicus*, native occurrences (left) and invasive occurrences (right).

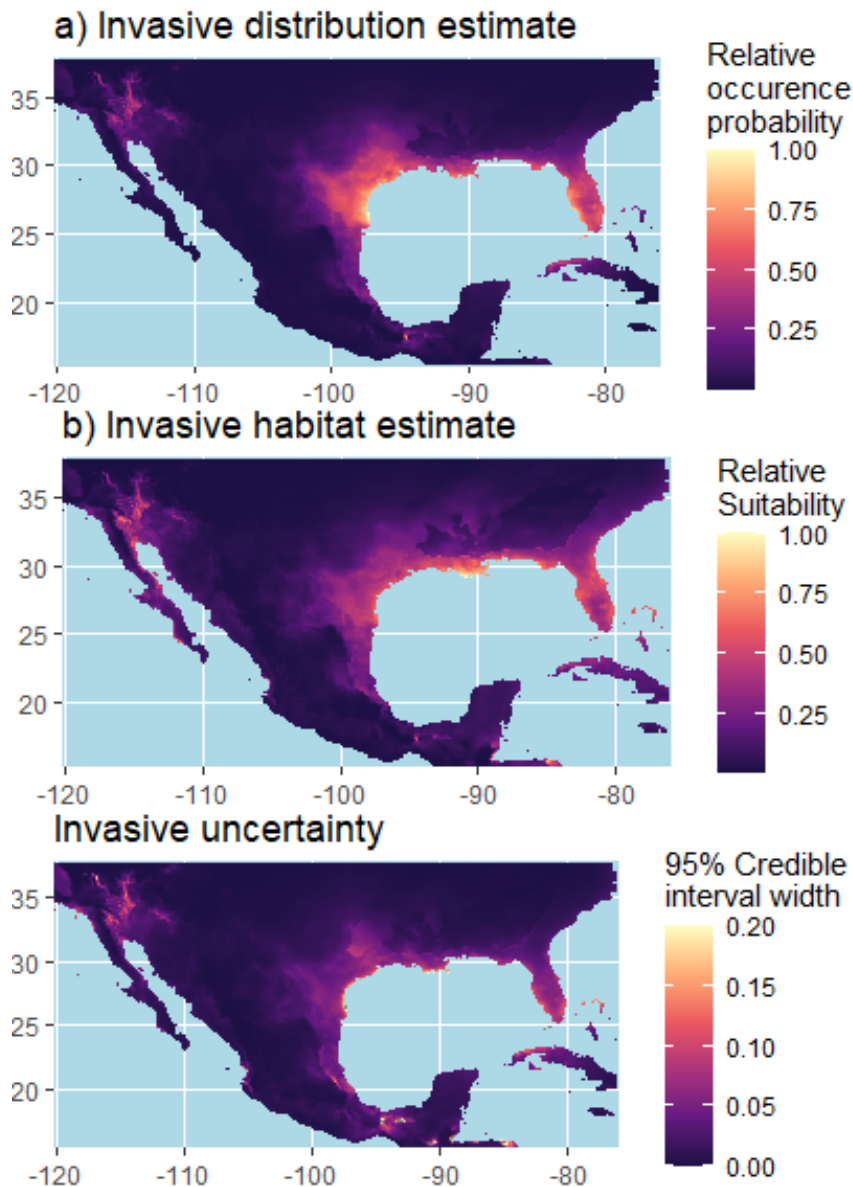


The distribution estimate has the strongest signal concentrated in the southwestern regions of the Iberian Peninsula, getting weaker as it goes along the Mediterranean coast up to France and a weaker signal on the eastern coast of the Mediterranean Sea (ca. 50% of maximum signal strength).

The habitat estimate has its strongest signal concentrated on the southern edge of the study area in the Saharan desert, Algeria (greyed out in the figure). A weaker signal continues along the southern edge of the study area and on the southeastern coast of the Mediterranean at around 10% the strength of the maximum signal.

Uncertainty maintains the same pattern as the distribution estimate (with uncertainty values around 0.2), except in the southwestern corner of the study area where values exceed 0.6.

Figure f2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model.

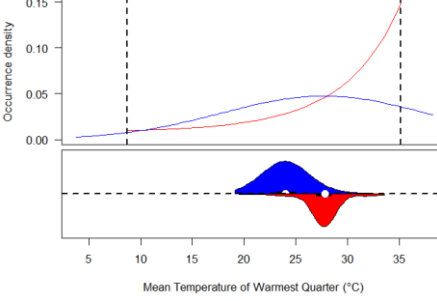
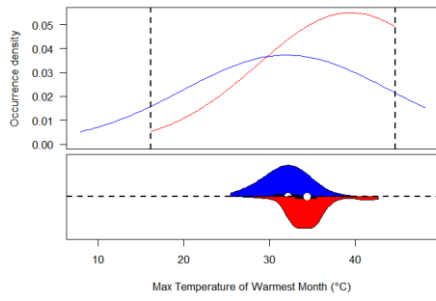
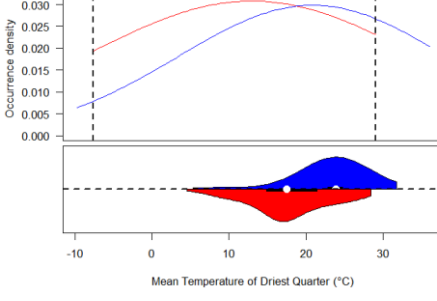
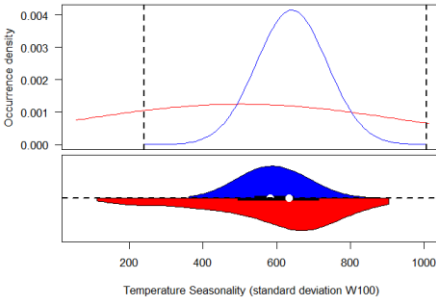
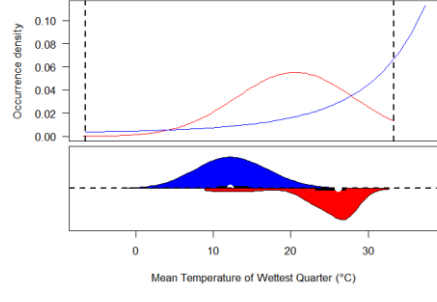
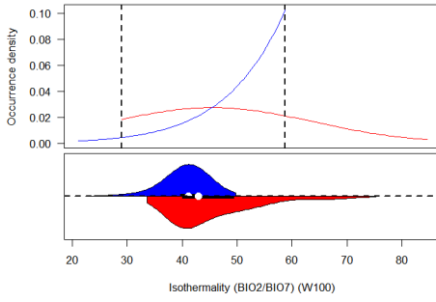
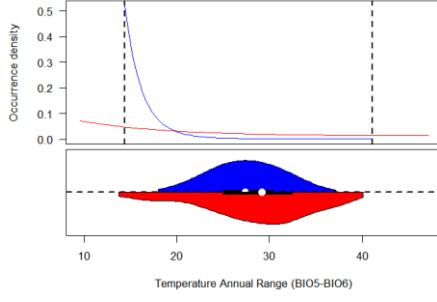
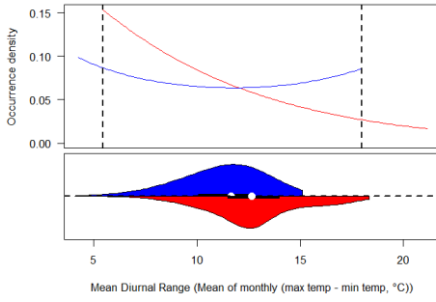
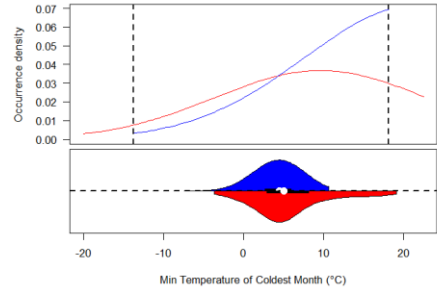
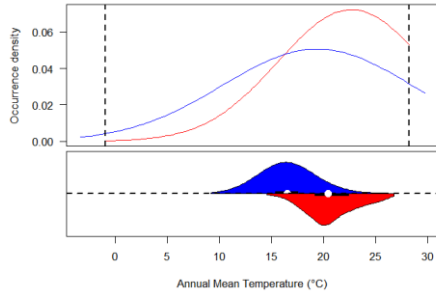


The distribution estimate shows the strongest signals around the continental coast of the Gulf of Mexico and a weaker signal north of the Gulf of California at around 15% of maximum signal strength.

The habitat estimate has the maximum signal in the Gulf of Mexico has shifted eastwards, and the relative strength of the Californian hotspot has increased to around 30% of maximum signal strength.

Uncertainty shows the same pattern as the distribution estimate, with uncertainty around the hotspots being around 0.05.

Figure f3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model.



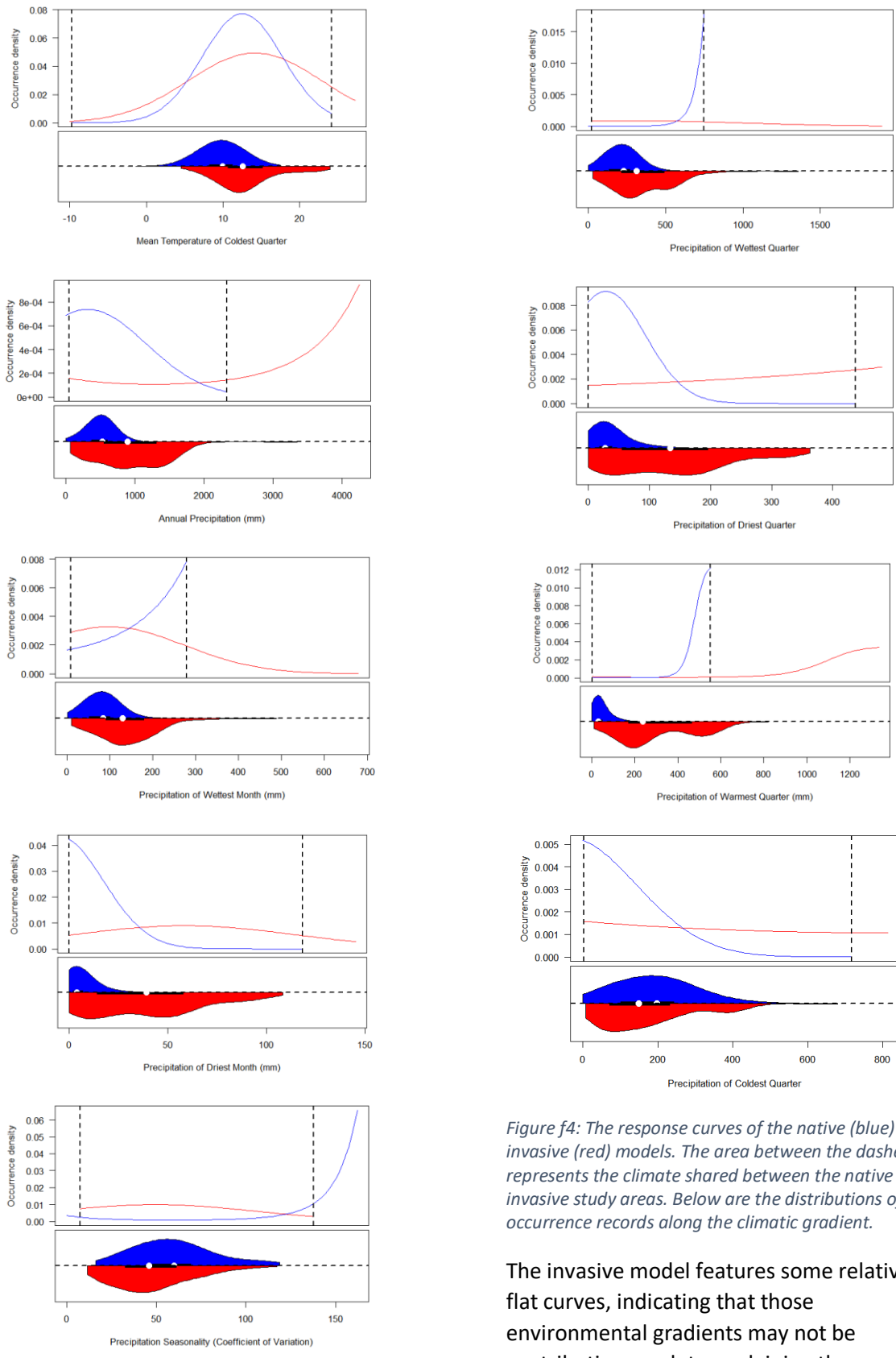


Figure f4: The response curves of the native (blue) and invasive (red) models. The area between the dashed lines represents the climate shared between the native and invasive study areas. Below are the distributions of occurrence records along the climatic gradient.

The invasive model features some relatively flat curves, indicating that those environmental gradients may not be contributing much to explaining the distribution in the invasive study area.

g) Rödder and Lötters (2010)

Species: *Eleutherodactylus planirostris*

Common name: Greenhouse frog

Native range: Cuba and the Bahamas

Introduced range: Southern coast of USA

Occurrence data: Occurrence data: Data recreated from 2 sources, including only records up until 2009:

1. GBIF.org (06 August 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.97przp>
2. David Bloom (2016). VertNet_Amphibia_Sept2016. CyVerse Data Commons. DOI 10.7946/P2F59W, 10.7946/P2F59W
3. U.S. Geological Survey, 2023, Specimen observation data for *Eleutherodactylus planirostris* (Cope, 1862), Nonindigenous Aquatic Species Database, Gainesville, FL, <https://nas.er.usgs.gov/queries/collectioninfo.aspx?SpeciesID=61>, Access Date: 3/24/2023.

Climate data: Data retrieved from WORLDCLIM database, version 2.1. Resolution: 30 arcsec (Fick and Hijmans, 2017)

Aggregation factor: Native = 3, Invasive = 4

Native study area extent (W, E, S, N): -86, -73.8, 19, 27

Invasive study area extent (W, E, S, N): -93.8, -79.4, 24, 33.8, Excluded: (-83, -79.5, 24.4, 27)

Occurrence records (Occurrence cells after aggregation): Native = 3862 (320), Invasive = 3633 (437)

Predictors:

- bio1: Annual Mean Temperature (°C)
- bio2: Mean Diurnal Range (Mean of monthly (max temp - min temp)) (°C)
- bio5: Max Temperature of Warmest Month (°C)
- bio8: Mean Temperature of Wettest Quarter (°C)
- bio9: Mean Temperature of Driest Quarter (°C)
- bio10: Mean Temperature of Warmest Quarter (°C)
- bio12: Annual Precipitation (mm)
- bio13: Precipitation of Wettest Month (mm)
- bio14: Precipitation of Driest Month (mm)
- bio18: Precipitation of Warmest Quarter (mm)

The greenhouse frog (*Eleutherodactylus planirostris*) is a medium-sized, brown frog native to Cuba and the Bahamas found in broadleaf forest and gardens. The species has become invasive on several Caribbean islands, the southern coast of the USA and even on the islands of Hawaii. The species reached to continental USA over 140 year ago and has become well established there (Rödder and Lötters, 2010).

Niche overlap analysis:

Rödder and Lötters (2009) and Rödder and Lötters (2010) used niche overlap analysis that can be

described as a hybrid between the geographic and univariate comparison methods. They derived individual response curves from an SDM and then mapped each response curve onto the invasive study range individually. Then a geographic overlap comparison was performed per environmental variable.

Limitations:

The original paper utilised gbif but did not record a reference code. A new gbif query was constructed, taking occurrences up to 2009 from Cuba, the Cayman Islands, the Bahamas and the norther coast of the Gulf of Mexico (spanning a latitude from Florida to the southern border of North Carolina).

The original paper used occurrence data from HerpNet, which is now defunct, but VertNet (HerpNet's parent "entity") still keeps the records here:

<http://www.vertnet.org/resources/datatoolscode.html#t-tab1>

The original paper included additional occurrences from 3 books. These sources generally include few occurrences and with poor locality information. These additional sources were not included in this study.

The original paper did study an additional introduced range on the Hawaiian Islands which was not included in this study.

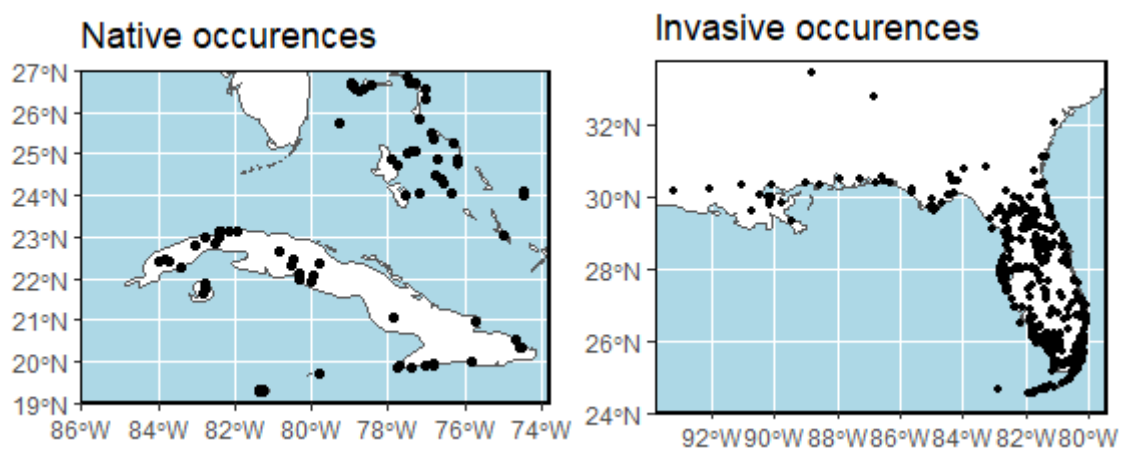
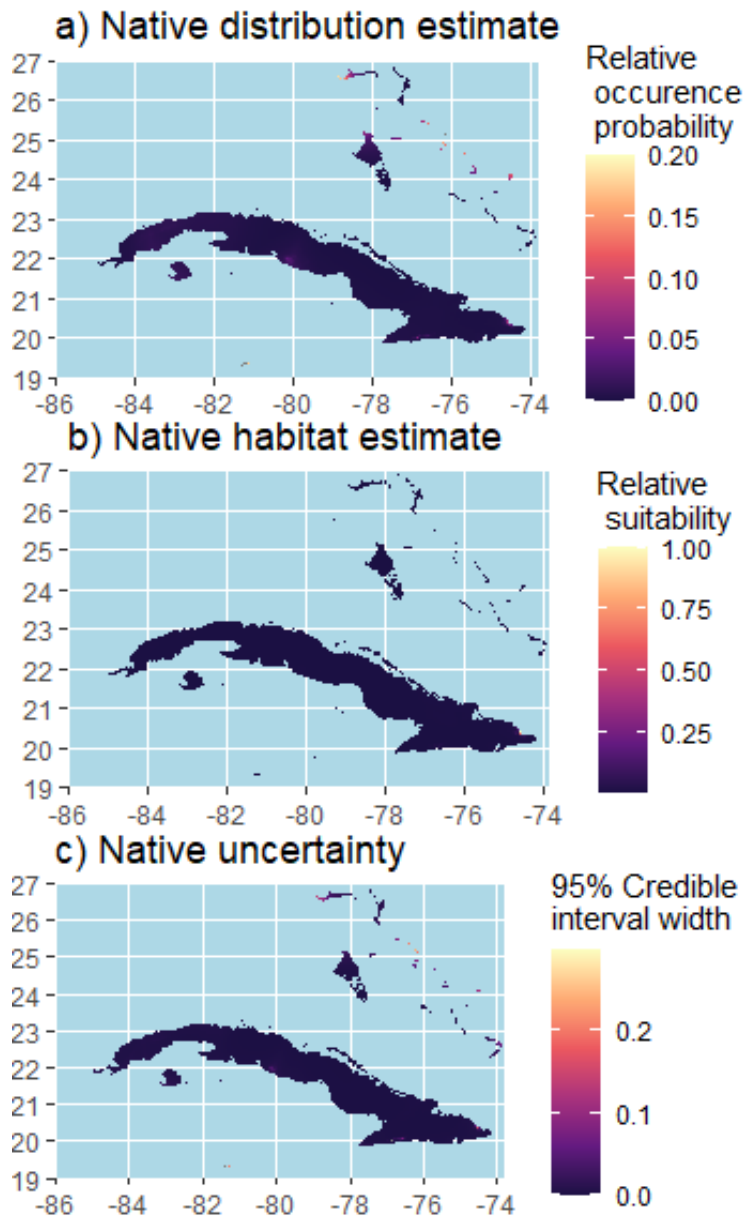


Figure g1: Occurrence records of *Eleutherodactylus planirostris*, native occurrences (left) and invasive occurrences (right).



The distribution estimate has its strongest signal on Grand Cayman Island with the next strongest signals (20% of max signal strength) on Grand Bahama Island and around Baracoa.

The habitat estimate only shows a signal around Baracoa, Cuba.

Uncertainty is largest around Baracoa and 2 points in Turquino National Park (ca. 0.7). Grand Cayman Island has an uncertainty around 0.2, while Grand Bahama Island has an uncertainty around 0.1.

Figure g2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model.

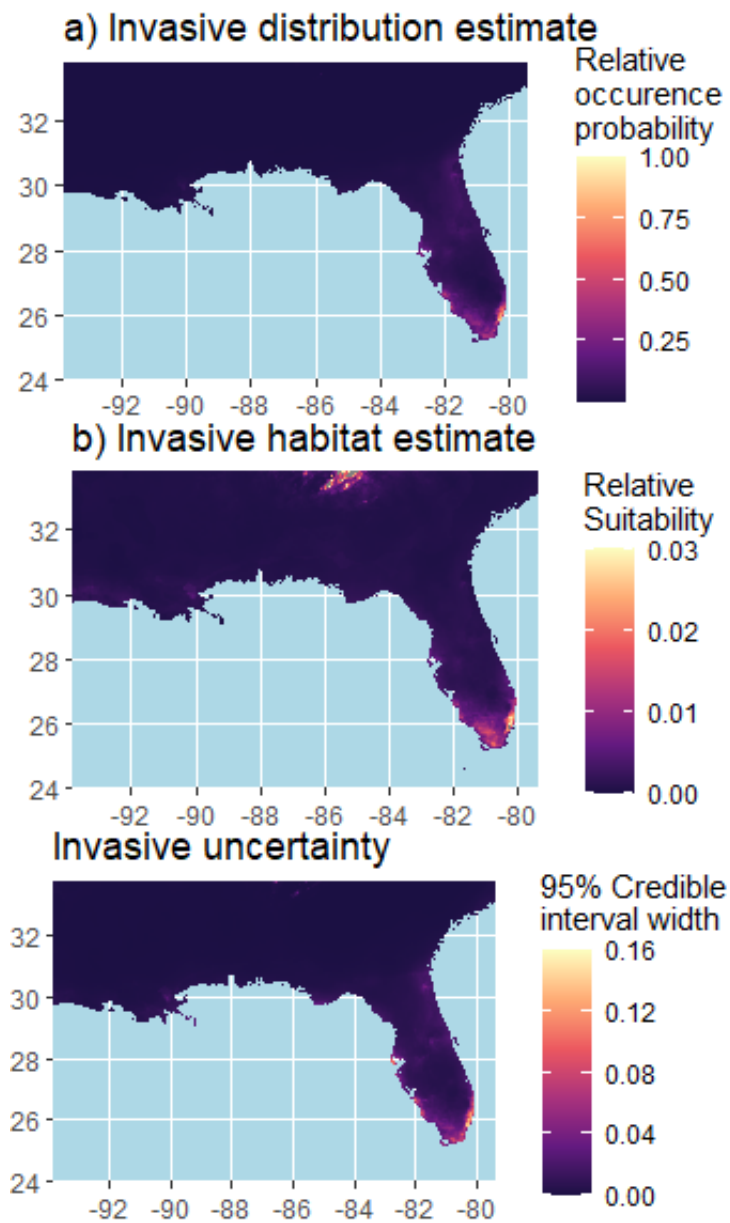


Figure g3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model.

The distribution estimate has a hotspot in southern Florida and a weaker signal on the western coast of Florida.

The strongest signal in the habitat estimate is in a few pixels at around 33.5°N and 86°W. Otherwise the next strongest signals are clustered around the maximum signal and southern Florida at around 3% of maximum signal strength.

Uncertainty at the maximum habitat estimate signal is around 0.7. Otherwise, the southern tip of Florida has uncertainties ranging from 0.06-0.15 in a similar pattern to the distribution estimate.

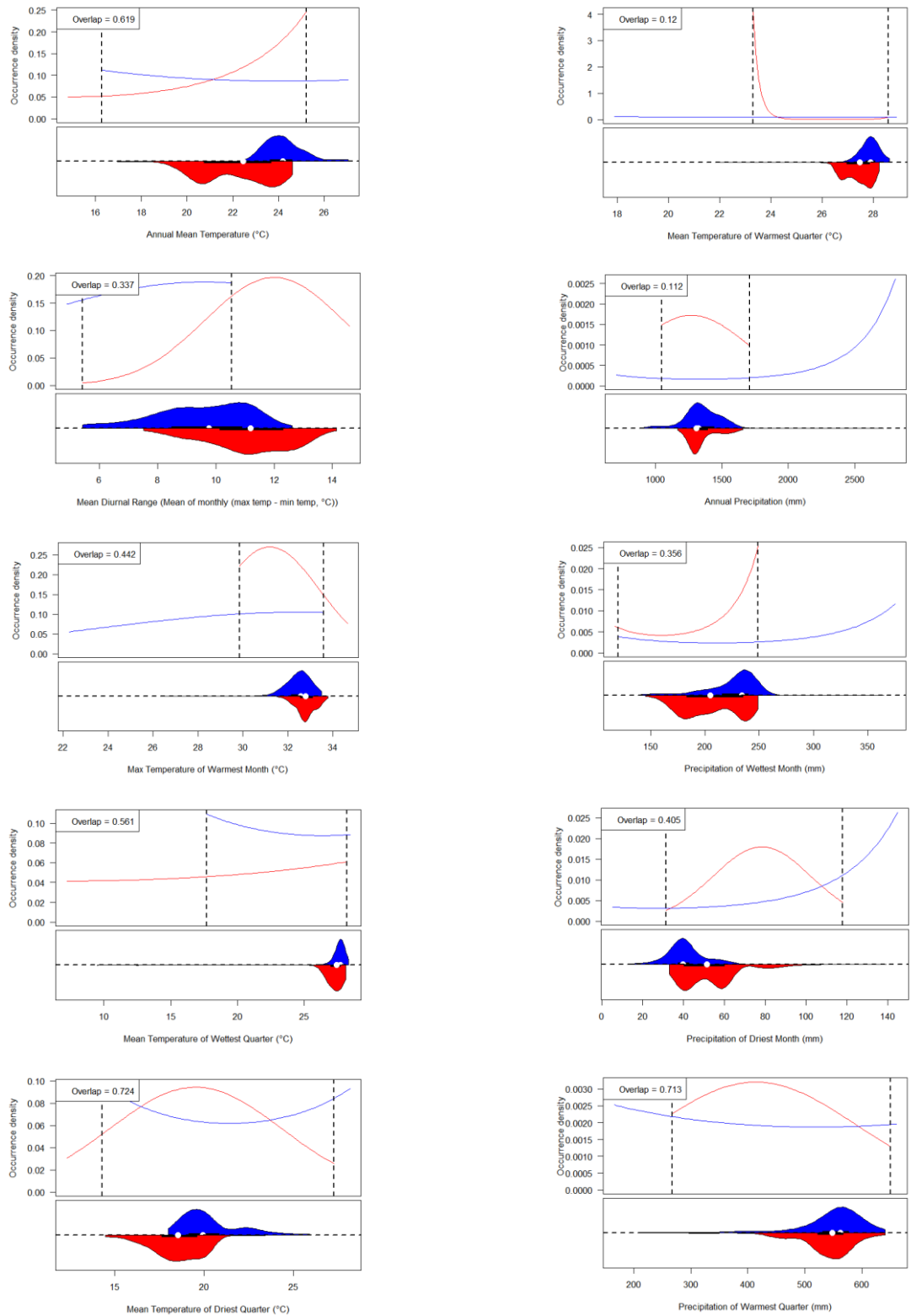


Figure g4: The response curves of the native (blue) and invasive (red) models. The area between the dashed lines represents the climate shared between the native and invasive study areas. Below are the distributions of occurrence records along the climatic gradient.

Both native and invasive models feature several response curves with positive quadratic terms which will lead to overprediction of projected onto non-analogous environments.

h) Zhang et al. (2022)

Species: *Galinsoga quadriradiata*

Common name: Shaggy soldier

Native range: Central & South America

Introduced range: China

Occurrence data: Data retrieved directly from the original paper (see Supplementary material, Table 2 in Zhang et al. 2022)

Climate data: Data retrieved from WORLDCLIM database, version 2.1. Resolution: 2.5 arcminutes (Fick and Hijmans, 2017)

Aggregation factor: Native = 3, Invasive = 3

Native study area extent (W, E, S, N): -112.4, -37.8, -34.1, 31.3, Excluded: (-87, -79.5, 25, 33)

Invasive study area extent (W, E, S, N): 79, 129.6, 22.3, 47.1

Occurrence records (Occurrence cells after aggregation): Native = 506 (301), Invasive = 393 (232)

Predictors:

- bio5: Max Temperature of Warmest Month (°C)
- bio6: Min Temperature of Coldest Month (°C)
- bio18: Precipitation of Warmest Quarter (mm)
- Elevation (m)

The shaggy soldier (*Galinsoga quadriradiata*) is an annual herbaceous plant native to South and Central America. The species has become invasive in China where it has caused both ecological and agricultural damage (Zhang et al., 2022).

Limitations:

The original study randomly diluted data so that each occurrence had a minimum distance between each other so as to reduce the effect of spatial autocorrelation. The spatial random effect included in the model for this study accounts for spatial autocorrelation, so the data was not diluted here.

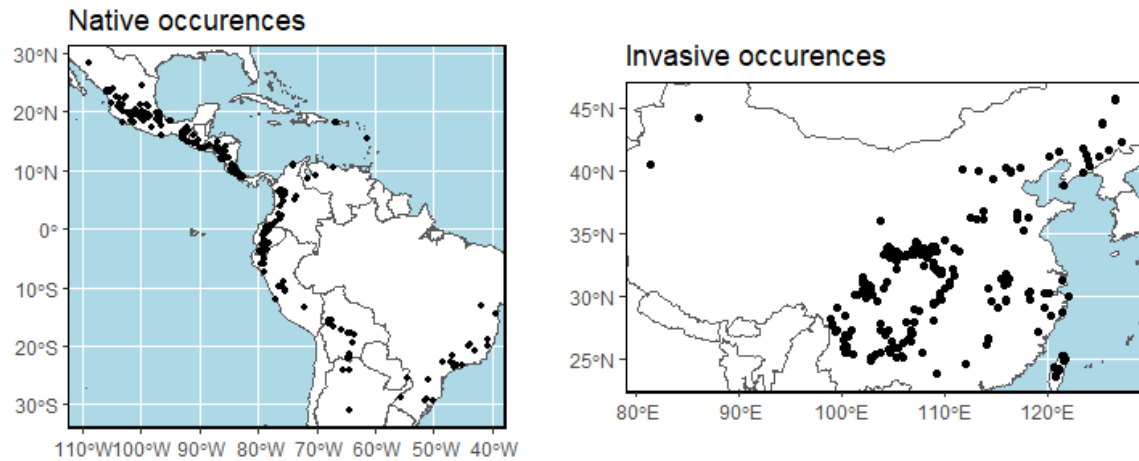


Figure h1: Occurrence records of *Galinsoga quadriradiata*, native occurrences (left) and invasive occurrences (right).

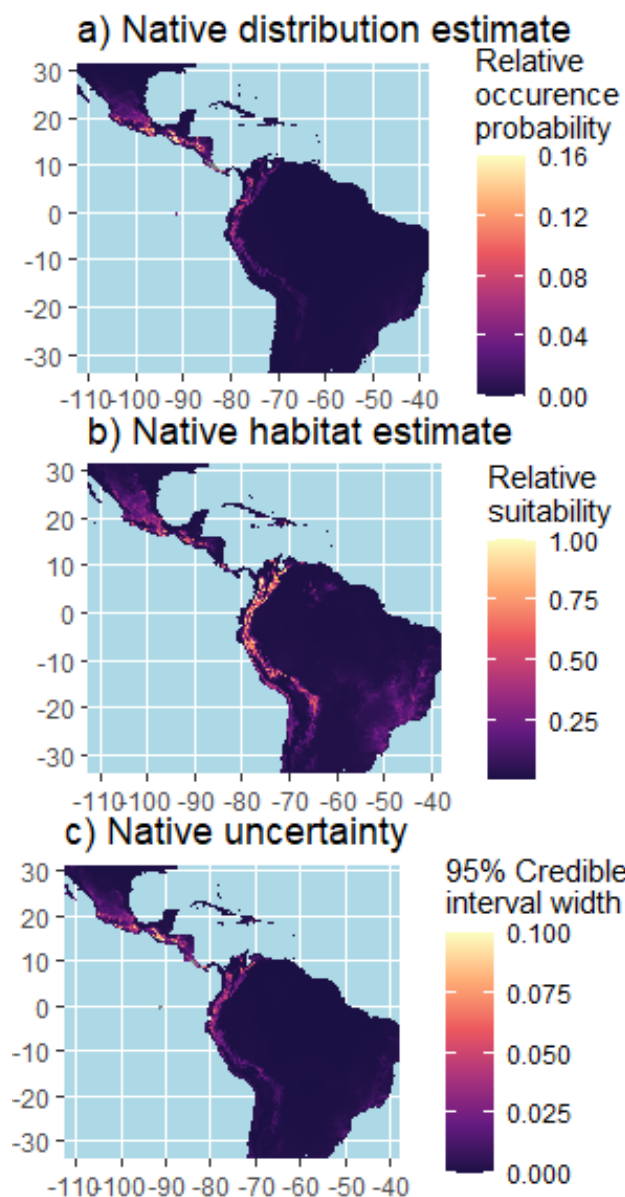
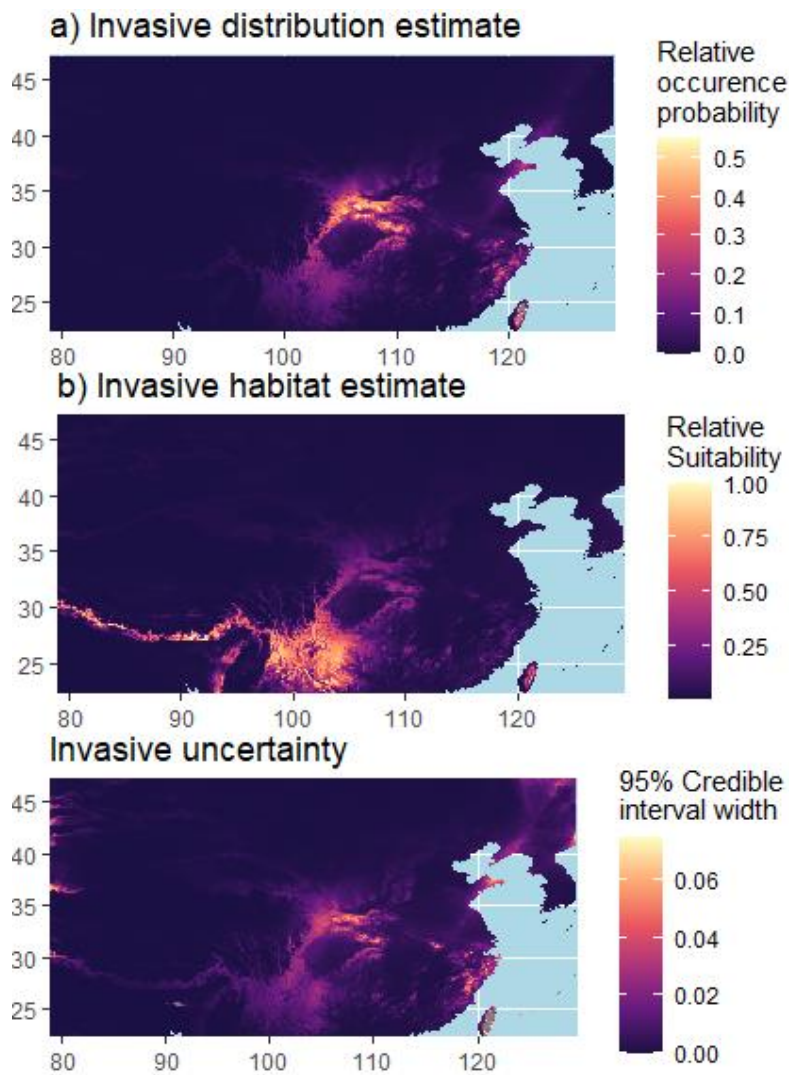


Figure h2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model.

The distribution estimate seems to highly correlate with elevation. The strongest signal covers most of Costa Rica. The weaker signals follow elevation north- and southwards at around 15% of maximum signal strength.

The habitat estimate pattern matches with elevation, with the strongest signal along the Andes Mountain range.

Uncertainty is highest in Costa Rica with values around 0.3 up to 0.4. Outside of Costa Rica the uncertainty follows elevation with the values up to 0.1.



The distribution estimate has a strong signal in Taiwan and a weaker signal around Sichuan province, China (up to 55% of maximum signal strength).

The habitat estimate pattern matches with elevation, except for the Tibetan plateau which has low to no signal and The Gobi Desert that has a lower signal.

Uncertainty matches the pattern of the distribution estimate with uncertainty in Taiwan reaching 0.2. The grey spot at around 25°N and 90°E has an uncertainty of 0.95. Otherwise, the uncertainty follows elevation (without the Tibetan plateau and Gobi Desert) up to 0.07 and some additional uncertainty at the western and northeastern borders of the study area around the same value.

Figure h3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model.

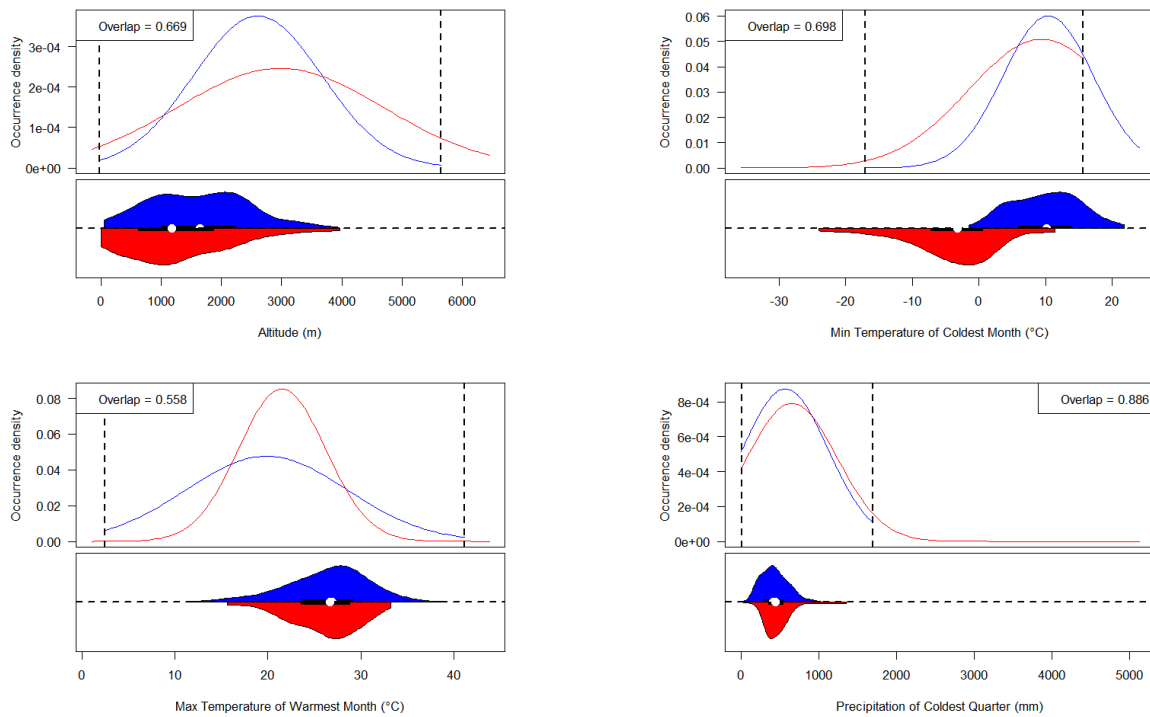


Figure h4: The response curves of the native (blue) and invasive (red) models. The area between the dashed lines represents the climate shared between the native and invasive study areas. Below are the distributions of occurrence records along the climatic gradient.

The response curves all appear to be reasonable. However, considering the dominance of spatial effects in both the native and invasive models (spatial effects contribution 6 times more than environmental effects) these curves cannot be relied upon.

i) Ørsted and Ørsted (2019)

Species: *Drosophila suzukii*

Common name: Spotted wing drosophila

Native range: Southeast Asia

Introduced range: Europe

Occurrence data: Data dryad: Ørsted, Iben V.; Ørsted, Michael (2018), Data from: Species distribution models of the Spotted Wing Drosophila (*Drosophila suzukii*, Diptera: Drosophilidae) in its native and invasive range reveal an ecological niche shift, Dryad, Dataset, <https://doi.org/10.5061/dryad.mn0254p>

Climate data: Data retrieved from WORLDCLIM database, version 2.1. Resolution: 30 arcsec (Fick and Hijmans, 2017)

Aggregation factor: Native = 17, Invasive = 13

Native study area extent (W, E, S, N): 66.4, 148.8, 8.5, 49.8

Invasive study area extent (W, E, S, N): -17.4, 44.2, 31.3, 60.9 (No western buffer)

Occurrence records (Occurrence cells after aggregation): Native = 101 (85), Invasive = 306 (185)

Predictors:

- bio1: Annual Mean Temperature (°C)
- bio2: Mean Diurnal Range (Mean of monthly (max temp - min temp)) (°C)
- bio4: Temperature Seasonality (standard deviation ×100)
- bio5: Max Temperature of Warmest Month (°C)
- bio6: Min Temperature of Coldest Month (°C)
- bio7: Temperature Annual Range (BIO5-BIO6) (°C)
- bio11: Mean Temperature of Coldest Quarter (°C)
- bio12: Annual Precipitation (mm)
- bio17: Precipitation of Driest Quarter (mm)

The spotted wing drosophila (*Drosophila suzukii*) is a species of fly native to Southeast Asia. In 2009 it was discovered both in Europe and North America. Since then, it has also been found in South America. The species prefers to lay its eggs in ripe fruit, which has caused great economic damage to the fruit industry in its introduced ranges.

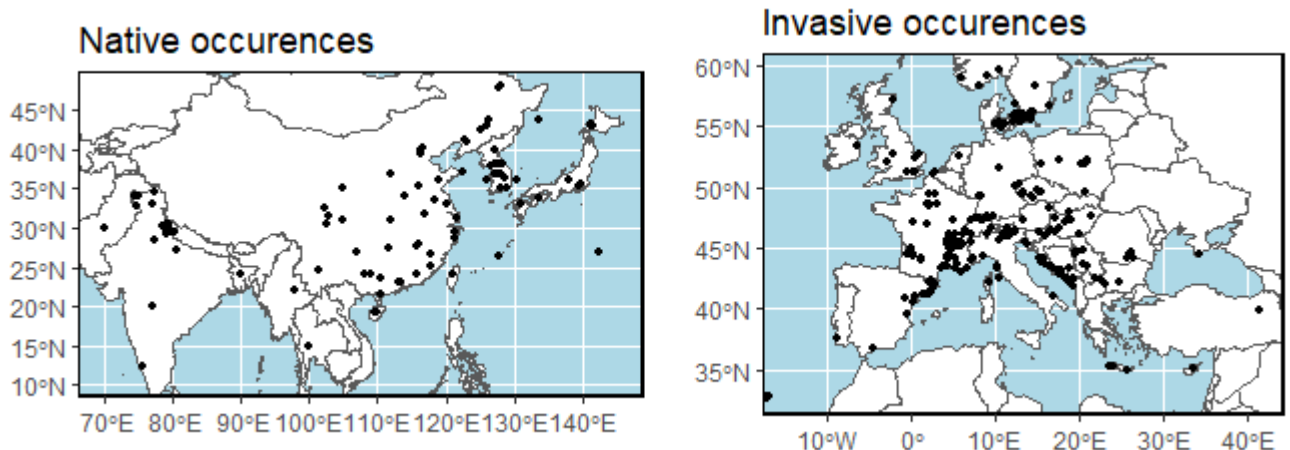


Figure i1: Occurrence records of *Drosophila suzukii*, native occurrences (left) and invasive occurrences (right).

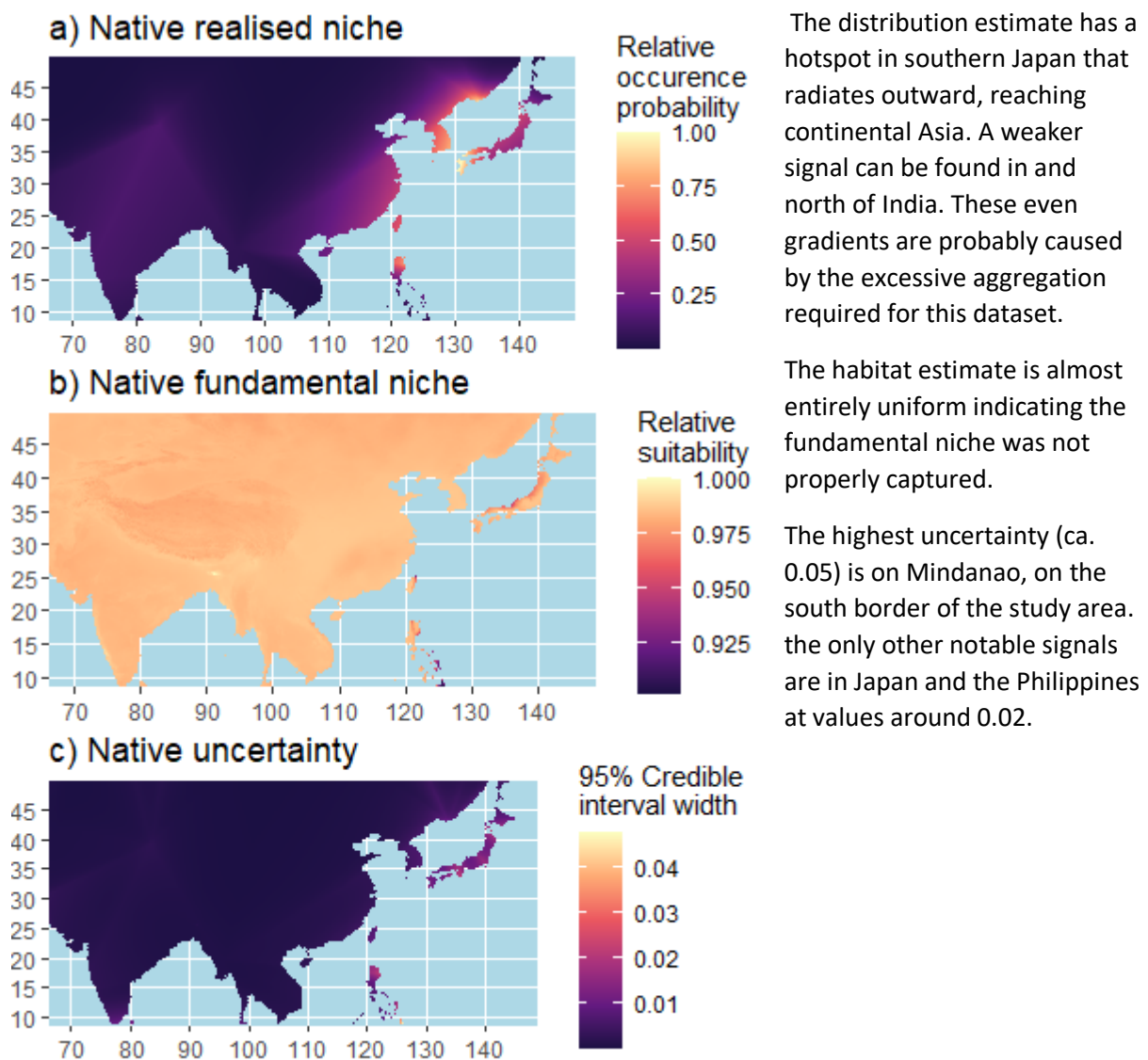
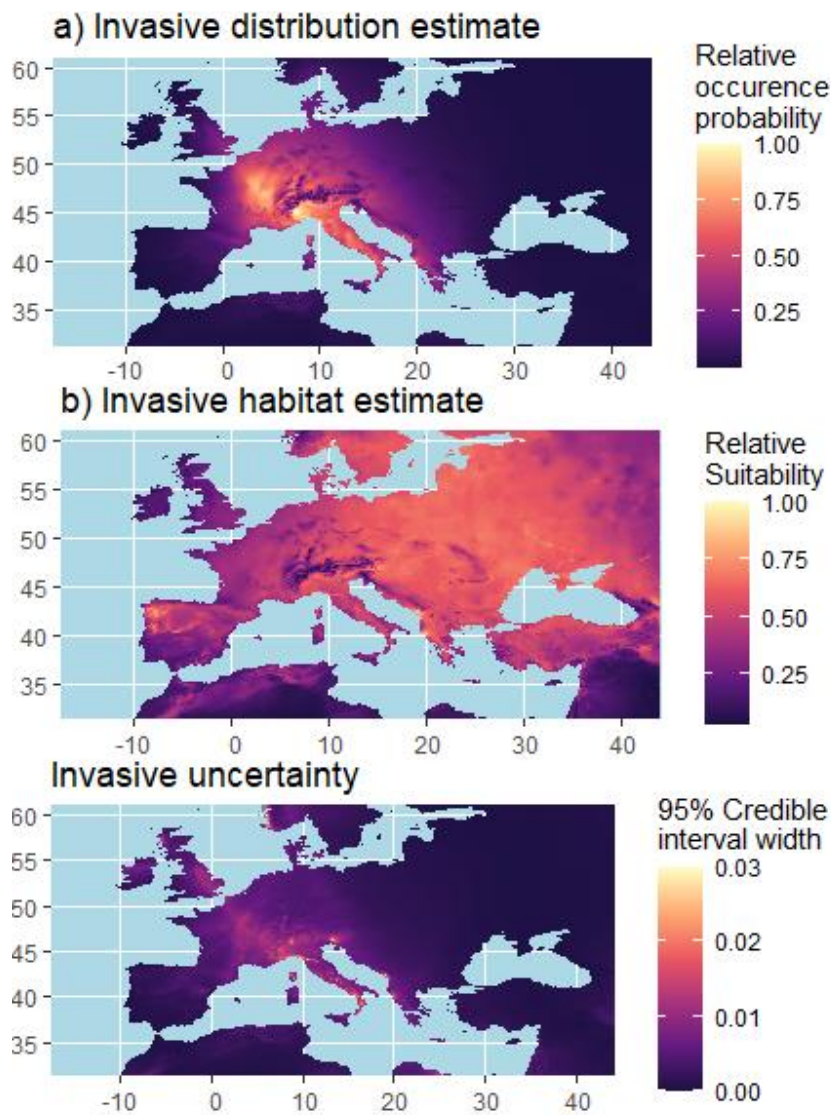


Figure i2: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from native model.



Due to the high aggregation factor, the results highly suspect. Patterns have been known to change when testing lower aggregation factors.

The distribution estimate has a hotspot west & south of the Alps that radiates outwards.

The habitat estimate has the strongest signal in eastern Europe and seems to have a negative relationship with elevation.

Uncertainty was mostly negligible with a low signal west and south of the Alps with values around 0.02. the Norwegian coast has a hotspot the reaches 0.27.

Figure i3: Relative distribution estimate (a), relative habitat estimate (b) and uncertainty of distribution estimate (c) from invasive model.

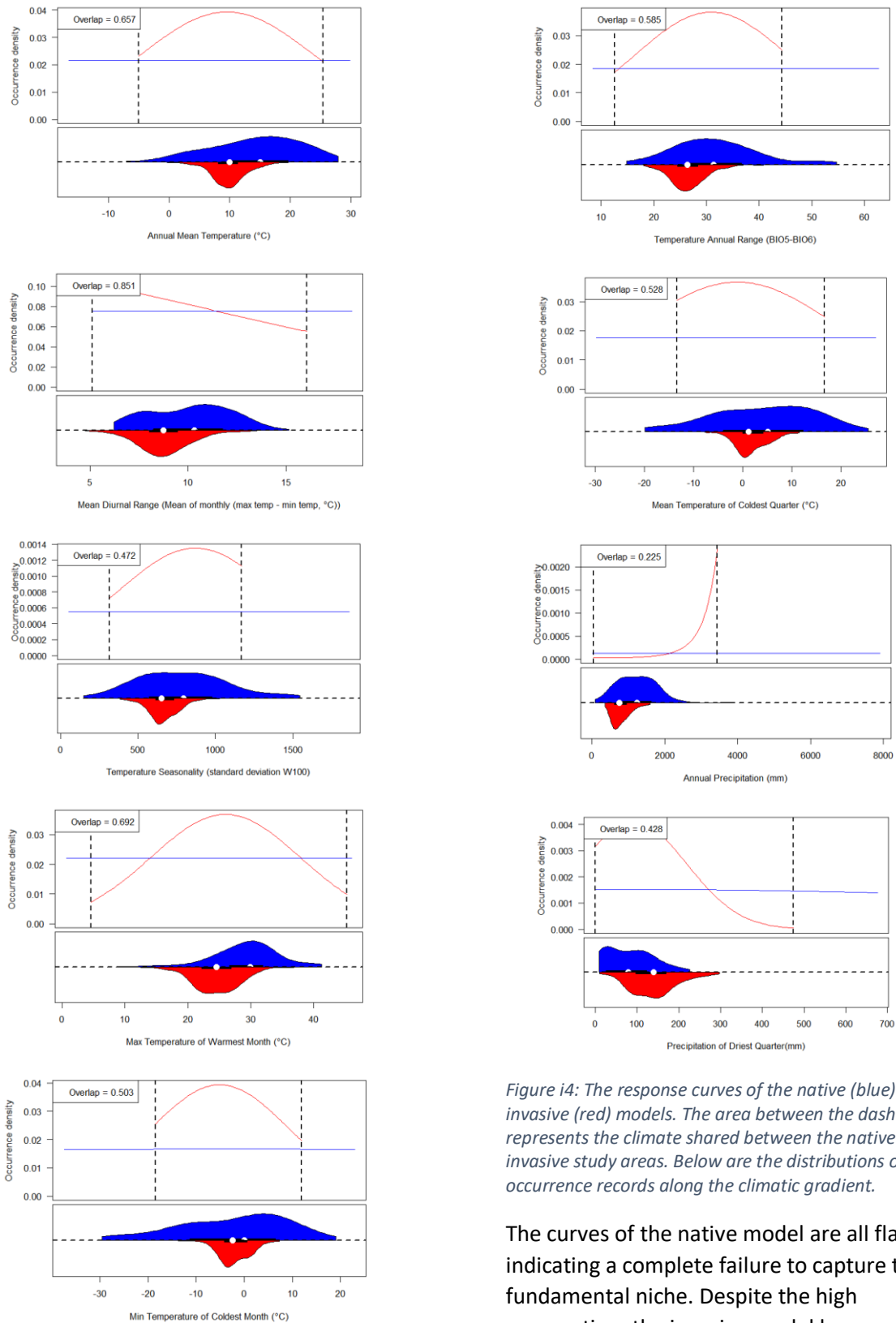


Figure i4: The response curves of the native (blue) and invasive (red) models. The area between the dashed lines represents the climate shared between the native and invasive study areas. Below are the distributions of occurrence records along the climatic gradient.

The curves of the native model are all flat indicating a complete failure to capture the fundamental niche. Despite the high aggregation, the invasive model has reasonable response curves and a relatively low proportion of spatial-to-climate effects (3.8) amongst the datasets.

