UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Living on the edge in a changing world: modelling global change effects on the distribution of Red-backed Shrike (Lanius collurio) in Iberia

Manuel João Cabral Saldanha Ribeiro

Mestrado em Bioestatística

Trabalho de Projeto orientado por:
Doutor Rui Martins
Doutor Luís Reino

2024

Acknowledgements

Throughout the makings of this work, several challenges were made lighter thanks to the contribution of others willing to put time and effort into helping me.

First and foremost, Dr. Laura Cardador for being the soul and body of this work. If it were not for you, this thesis would have been much harder than it already was. Thank you for teaching the ropes to a beginner such as myself, and me not being able to put your name on the front page is a complete injustice for all your contributions. Thank you very much for being the catalyst of this thesis, and I hope this work shows a little bit of how much I learned from you.

Dr. Luís Reino and Dr. Joana Santana for always supporting me through the ups and downs. It is extremely rare to find people willing to help, endure and provide as much as you two did for me. I am extremely thankful for all your support, advice, reminders and overall patience throughout all of this time.

Dr. Rui Martins for his understanding and availability in helping me tie up this thesis. Not a lot of people would have shown the same attitude to my situation, and I'm very thankful for your contribution in this work.

My parents, Aníbal and Gabriela for all the support and care provided. I am glad that in the end I was able to complete this chapter, and that wouldn't have been possible without you. Thank you very much for your patience and understanding.

Nuno Farinha for giving authorization to use his illustration in this thesis.

Prof. Dr. Henrique Pereira for his advice and availability in helping me write this thesis.

My friends for the checkups, advice and cheering provided throughout the development of this thesis. You may feel like you haven't done much but believe me, some of the things you said in the course of this thesis helped me work out a lot of issues, and put things into perspectives much easier to deal with.

Last but definitely not least, Annika for going above and beyond anything I could have asked for. If I could I would also add your name to the front page, but unfortunately, I cannot do that (yet another injustice). You truly deserved a spot there and I couldn't be more grateful for everything you have done for me and this thesis. Vielen Danke für alles.

This thesis was already hard enough without the contributions of everyone mentioned here. I cannot thank you enough for everything you have done.

Data availability

All data and code used in this thesis are openly available in GitHub at https://github.com/MJCSRibeiro/Lanius_biomod2.git

Resumo

O aumento de temperaturas e modificação de usos do solo, como consequência de atividades de origem antropogénica, têm sido documentadas como alguns dos principais componentes na alteração de qualidade de habitats. Estas mudanças radicais têm impactos profundos na composição de comunidades biológicas, podendo levar a perturbações no nicho ecológico de determinadas espécies.

As distribuições geográficas das espécies muitas vezes são vistas como uma representação fisiológica entre a tolerância das mesmas a determinados fatores ecológicos e a forma como estes fatores variam ao longo do espaço. Em particular, territórios na margem da distribuição das espécies frequentemente marcam o limite dessa tolerância, existindo nesses locais um equilíbrio frágil entre populações biológicas e disponibilidade de habitat. Consequentemente, é espectável que os impactos das alterações globais sejam amplificados nas populações presentes em territórios no limite da distribuição geográfica.

Entre os grupos biológicos, as aves migradoras são frequentemente dos mais afetados por alterações na qualidade do habitat. Por percorrerem uma grande quantidade de território, este grupo enfrenta potenciais ameaças em territórios de reprodução primaveris, migração e invernada, tornando-o especialmente sensível a alterações na qualidade dos habitats a nível global.

Para identificar fatores limitantes da distribuição das espécies, assim como possíveis impactos originados por alterações na qualidade do habitat, ecólogos começaram a utilizar modelos de previsão de distribuição de espécies. Este tipo de modelação permitiu relacionar a forma como as espécies se distribuem com gradientes climáticos e topográficos, assim como obter previsões de mudanças de distribuição com alterações previstas de habitat.

Este trabalho focou-se em modelar a distribuição do Picanço-de-dorso-ruivo (*Lanius collurio*), uma ave migradora que inverna no sudeste de África e na primavera migra para a Europa e Ásia para se reproduzir, voltando para os territórios de invernada no outono. Esta espécie apresenta uma ampla distribuição em território de reprodução, estendendo-se desde o oeste da Europa até à Ásia Central. Na Península Ibérica, esta ave ocupa principalmente territórios no norte e centro, frequentemente associados a altitudes mais elevadas. O Picanço-de-dorso-ruivo depende de áreas abertas com alguma vegetação de médio porte. Arbustos e postes servem como postos de vigia, permitindo melhor deteção e captura de presas, tornando-se assim um componente importante para a sobrevivência desta espécie.

Em particular, certas populações no limite da sua distribuição sofreram grandes declínios, levando mesmo a extinções regionais em locais como o Reino Unido. Estas situações realçam a importância de compreender a forma como as espécies são limitadas pelo seu habitat, podendo guiar esforços de conservação de maneira mais eficiente em regiões críticas. Assim, a Península Ibérica foi escolhida como o foco espacial deste estudo, por representar o limite sudoeste da distribuição geográfica do Picanço-de-dorso-ruivo no período reprodutivo.

Em concreto, este trabalho teve como objetivos: determinar quais os fatores bioclimáticos e de uso do solo mais limitantes na distribuição do Picanço-de-dorso-ruivo; prever a distribuição geográfica da espécie baseada em modelos probabilísticos, usando os fatores previamente mencionados; usar esses mesmos modelos para estimar o impacto das alterações globais, baseadas em cenários previstos pelo

Painel Internacional de Alterações Climáticas (IPCC). Assim, os resultados permitiram averiguar os fatores condicionantes do seu nicho ecológico, assim como a variação expectável da distribuição sob condições futuras, permitindo uma análise do impacto das alterações globais no território reprodutivo total desta espécie.

Os dados usados neste trabalho foram condensados em grelhas geográficas de 0,5º em latitude e longitude, permitindo o mapeamento geográfico da informação. Para modelar a distribuição do Picanço-de-dorso-ruivo foram recolhidos dados de ocorrência da espécie em cada célula da grelha de 0,5º, só sendo necessário 1 presença dentro de cada célula para a espécie ser considerada presente. Caso contrário, a espécie foi considerada ausente dentro das células. Duas bases de dados online foram usadas neste trabalho, dados de ciência cidadã obtidos da plataforma GBIF, e dados de censos de reprodução do segundo Atlas Europeu das Aves Nidificantes (EBBA2). Visto que em grande parte do território é incerto se a espécie não foi detetada por ausência de condições favoráveis ou por falta de dados, foram usadas pseudo-ausências dentro de uma área delimitada para mapear os gradientes ambientais relevantes para a ocorrência da espécie. Estes gradientes consistiram em condições bioclimáticas e de uso do solo. Para as condições bioclimáticas foram calculadas as médias dentro de cada célula de 0,5º para: temperatura mínima, temperatura média, temperatura máxima, variação sazonal da temperatura, precipitação e variação sazonal da precipitação. Para os usos do solo, em cada célula de 0,5º foi atribuído o uso predominante dentro de 5 categorias: usos considerados não favoráveis à ocorrência da espécie (deserto, gelo, urbanizações intensas), matagal intenso, espaços maioritariamente abertos, zonas florestais e produções agrícolas. Ambos os grupos de covariáveis foram criados para dois períodos temporais distintos: o período *current* (1961-1990) e cenários futuros em 2050.

O primeiro passo focou-se em executar modelos baseados em 4 algoritmos distintos dentro do pacote R *biomod2*: regressão linear (GLM), máxima entropia (MaxEnt), *boosting* (GBM) e *bagging* (RF). Estes modelos relacionaram a ocorrência (presença ou ausência) do Picanço-de-dorso-ruivo com as condições bioclimáticas e de uso do solo presentes em cada célula da grelha geográfica, atribuindo uma estimativa de probabilidade de ocorrência. Após obter as estimativas de probabilidade para cada célula de 0,5º em todas a área em análise, foi possível converter os valores contínuos de probabilidade em categorias binárias de ausência (0) e presença (1), usando um valor limite separando valores abaixo do limite e acima do limite em dois grupos distintos. As previsões binárias foram comparadas com os dados originais de ocorrência, usando um método de *cross-validation* em que 70% dos dados foram usados para calibrar os coeficientes dos modelos, e as previsões dos modelos foram comparadas com um grupo de dados independente obtidos de censos para Portugal e Espanha. Isto permitiu obter várias métricas de avaliação dos modelos, entre os quais destacou-se o *True Skill Statistic* usado para avaliar a capacidade preditiva dos modelos.

Em seguida, as previsões probabilísticas dos modelos foram combinadas usando uma técnica conhecida como *ensemble*, em que é calculada a média dos valores contínuos de estimativa para a ocorrência da espécie em cada célula de 0, 5º. Apenas os modelos com valores de *True Skill Statistic* de pelo menos 0,7 foram considerados precisos o suficiente para integrarem o modelo *ensemble*. Desta forma, obteve-se estimativas de ocorrência da espécie baseadas em algoritmos diferentes, o que permitiu obter um modelo mais equilibrado e verificar onde existem zonas com maior variabilidade entre previsões dos diferentes modelos.

Após obter o modelo *ensemble*, o último passo consistiu em modelar a distribuição do picanço-de-dorso-ruivo em cenários de alterações bioclimáticas e de uso do solo previstas para o ano de 2050. Todos os

modelos que foram selecionados para integrar o modelo *ensemble* foram usados para prever as novas probabilidades de ocorrência sob condições futuras, e posteriormente calculada a média novamente. Assim, obteve-se uma previsão de como as alterações da qualidade de habitat irão afetar esta espécie em todo o seu território de reprodução e, particularmente, na Península Ibérica.

Os resultados obtidos demonstraram que os algoritmos de *boosting* (GBM) e *bagging* (RF) tiveram melhores previsões da distribuição do picanço-de-dorso-ruivo, e que utilizar a técnica *ensemble* criou um modelo com boa capacidade preditiva e adicionou informação relevante sobre áreas com maior incerteza sobre as estimativas. O modelo *ensemble* ainda atribuiu maior valores de influência à temperatura máxima e variação sazonal de temperatura sobre a variação das estimativas de probabilidade de ocorrência ao longo de todo o território. Em grande parte, o modelo final obteve uma previsão bastante precisa da distribuição do picanço-de-dorso-ruivo, tanto em toda a área de distribuição como na Península Ibérica.

No final, foram obtidas previsões que as alterações globais irão levar a uma expansão da distribuição para novos locais no norte da Europa e Ásia, e uma contração nos territórios a sul próximos do limite da distribuição. Na Península Ibérica os resultados foram mais incertos, uma vez que foi prevista um abandono dos locais de reprodução associados ao sistema Central, mas também o estabelecimento de novas populações reprodutoras em territórios a sul.

Palavras-chave: picanço-de-dorso-ruivo, modelos de distribuição de espécies, nicho ecológico, Península Ibérica, alterações globais

Abstract

Biodiversity is constantly under threat due to human-made global change. Habitat qualities, such as climate and land-use, are changing faster than species can adapt, leading to shifts in geographic distributions. These changes can have very negative impacts on local populations, especially for those at the margin of a species' distribution, often corresponding to the limits of their physiological tolerance to environmental pressures. This work focused on modeling the current and future distribution of the Red-backed Shrike (*Lanius collurio*) using species distribution models (SDMs). The Iberian Peninsula at the southwestern margin of its range was selected to assess the impacts of altering habitat conditions in the edge of its breeding distribution. Four modeling algorithms (GLM, MaxEnt, GBM and RF) included in the *biomod2* package were used to relate occurrence data to a set of bioclimatic and land-use variables. All available data was aggregated in 0.5º resolution grid cells across the study area. Model coefficients were calibrated using citizen science and breeding atlas occurrence data, and predictive accuracy was evaluated on independent field census data for the Iberian Peninsula. High performance models with True Skill Statistic values of at least 0.7 were combined into an ensemble model by averaging habitat suitability predictions. The results showed that decision tree algorithms (GBM and RF) had the best predictive performance, and averaging model predictions into an ensemble provided the highest evaluation metric scores. Maximum temperature was the dominant effect on influencing the distribution of this species, and projections into future scenarios predicted a significant northward expansion in most of the Red-backed Shrike's breeding range, and a fractional retraction from southern territories. In the Iberian Peninsula, global change was predicted to lead to contractions in the central system, with new areas in southern mountainous regions becoming more suitable for the occurrence of this species.

Keywords: Red-backed Shrike, species distribution models, ecological niche, Iberian Peninsula, global change

Contents

List of Figures

List of Tables

1 Introduction

Human driven global changes on climate and landscapes have been a rapidly growing concern, with well-documented negative effects on biodiversity across the globe, mainly habitat degradation through land-use changes and alterations in large scale climatic patterns (Jetz et al., 2007; Neto & Araújo, 2011; Pereira et al., 2012; Muluneh, 2021). Various anthropogenic impacts, such as the increase of greenhouse gases (e.g., CO2, methane), are linked to a gradual increase in global temperatures, while economic scenarios also foresee major reorganizations of land-use in the future (IPCC, 2008). Bioclimatic conditions, such as temperature and precipitation, have been shown to impact species ecological niches, and several species changed their geographic distributions in response to altering climate (Araújo & Rahbek, 2006; Chen et al., 2011). In addition, anthropogenic land-use changes, such as agricultural intensification, afforestation and urbanization, have shown indirect effects on biodiversity like altering food availability, fragmenting existing habitat and changing interactions between climate and landscape (Jetz et al., 2007; Reino et al., 2018; Sándor & Domşa, 2018).

As global change alters habitat availability and quality, these impacts often induce changes in species distributions, primarily in the form of range shifts, caused by population movements in response to altering habitat suitability (Pearson et al., 2006; Elith et al., 2010; Guisan et al., 2017). These range shifts take the form of expansions when species are predicted to occupy new areas, or contractions where local populations abandon previously occupied areas (Guisan et al., 2017). The derived effect from these changes are especially evident in locations at the edges of species geographic distributions, where local populations are more constrained by environmental pressures, increasing the impact that global change will have at local scales (Reino et al., 2006; Fernández et al., 2015). In some cases these effects have been so severe that they have led to local extinctions at regional levels (Araújo et al., 2005; Araújo & Rahbek, 2006; Tellería, 2018b).

Given these challenges, ecologists have started using species distribution models (SDMs) to assess the factors influencing geographical distributions and understand how global change will potentially affect range sizes of target species (Guisan & Zimmermann, 2000; Pearson et al., 2006; Elith et al., 2010). Conceptually, SDMs relate field observations to environmental predictor variables, based on statistically or theoretically derived response surfaces (Guisan & Zimmermann, 2000). These models quantify how species occurrences (presence or absence) respond to biotic factors (e.g., abundance of prey, predator density) and abiotic factors (e.g., climatic conditions such as temperature and precipitation, as well as landscape and topography) across different locations. Early SDMs relied on presence records of species to define environmental niches, using environmental envelopes to estimate suitable conditions (Guisan et al., 2017). Later, absences were also taken into consideration, enabling the application of probabilistic models, which facilitated the incorporation of parametric and semiparametric statistical models (e.g., GLM and GAM), and later the addition of machine-learning processes, further improving flexibility and predictive performance (Elith et al., 2006; Guisan et al., 2017).

By quantifying the response of a species to its environment, SDMs are able to assess habitat suitability at given locations with known conditions, providing an estimated probability of the species' presence in these areas. These probabilities highlight areas of high suitability for the occurrence of the target species,

aiding in targeted conservation efforts (Guisan et al., 2017). Additionally, this probability can then be converted into a binary prediction of presence or absence, which can further be compared with observed occurrence data to assess model accuracy through selected evaluation metrics such as Area Under the Curve (AUC), enabling accuracy performance comparisons between different model predictions and the observed data (Guisan & Thuiller, 2005; Elith et al., 2006).

Model evaluation was a crucial step for SDMs since these models were mostly seen as unreliable and merely informative tools (Guisan & Thuiller, 2005). Performance comparisons paved the way for SDMs to evolve into reliable predictive frameworks of species distributions, allowing researchers to assess the strength of different modeling algorithms, sampling designs and modeling frameworks (Elith et al., 2006). With advancements in computational tools and extensive data availability, SDMs now provide increasingly accurate predictions of suitable habitat and are recognized as powerful tools in ecological modeling (Elith & Leathwick, 2009).

While SDMs provide a link between species and their habitat preferences, they are limited to the available biological and environmental information (Guisan et al., 2017). Nowadays, multiple datasets are available online, easing accessibility to the required data (Elith et al., 2010). Species records are often acquired through open access online databases, with one of the most widely used being the Global Biodiversity Information Facility, also known as GBIF. Due to the lack of targeted surveys, observations usually incorporate citizen science, where records can be registered by scientific sampling approaches or opportunistic public records (Guisan et al., 2017). Citizen-science data is becoming increasingly important in bridging information gaps, due to the extensive amount of volunteers and minimal cost of data gathering compared to targeted surveys, which tend to be high cost and cover very small areas. However, opportunistic records often show uneven sampling intensities across the study area, introducing sampling bias in the occurrence data, which influences statistical models derived from it (Meyer et al., 2015).

Furthermore, these records are usually restricted to observed “presences”, as documenting observed “absences” requires intensive field surveys. These limitations complicate building probabilistic SDMs from citizen-science data, since the probabilistic distribution across environmental gradients is not represented (Guisan et al., 2017). To overcome this, researchers resort to generating pseudo-absences, where environmental information in the whole study area is used to compare with environmental conditions of existing presence data (Barbet-Massin, Jiguet, et al., 2012; Guisan et al., 2017). This method enables the application of probabilistic statistical models such as GLMs, further enhancing the accuracy and applicability of SDMs in predicting species distributions.

Similar to occurrence data, climate and land-use information is widely available through online datasets such as Worldclim and IMAGE, which provide geographic projections of environmental conditions over various time intervals (Guisan et al., 2017). These projections rely on extensive networks of climate stations and satellite imaging, allowing for holistic coverage of environmental factors across the globe. However, climate data often shows high levels of uncertainty due to uneven geographic coverage of climate stations. To address this, statistical techniques such as smoothing splines are employed to estimate conditions in unsampled locations, improving data consistency (Guisan et al., 2017).

Technological advancements in sampling techniques, such as remote sensing and high-resolution satellite imagery, allowed for the creation of highly accurate and detailed spatial and temporal data for environmental factors, increasing the reliability of models (Rew et al., 2020). For future conditions, modern tools are also available to offer powerful climatic and landscape change predictions. For

example, global circulation models (GCMs) simulate interactions between the atmosphere, oceans and land, to project probable climate conditions into future scenarios, based on carbon emissions from anthropogenic activities (IPCC, 2008). Additionally, land-use changes are projected by integrating socio-economic practices and policies, providing a more complete understanding of possible future habitat dynamics (Reino et al., 2018). This evolution of knowledge helps create more robust predictions of global change impacts, increasing the effectiveness of targeted conservation practices at local scale and improving global scale assessment of biodiversity shifts (Guisan et al., 2017).

Despite their widespread use and utility, SDMs are constrained by a large amount of uncertainties that complicate their interpretation and application (Guisan et al., 2017). These uncertainties stem from various sources, including the choice of modeling algorithms, the design of sampling strategies, the selection and scale of environmental predictors, and the quality of occurrence data (Thuiller, 2004; Araújo et al., 2005; Pearson et al., 2006). Each of these steps introduces variability in model predictions, leading to discrepancies in outcomes across studies and approaches. This issue is amplified under novel environmental conditions, as projections for future scenarios often extend beyond the ecological and climatic ranges where model accuracy has been validated (Elith et al., 2010). The inability to empirically verify such predictions creates additional challenges, especially when assessing the impacts of global change on species distributions (Elith et al., 2010).

To address these uncertainties, SDMs rely on robust modeling frameworks designed to enhance prediction reliability and reduce variability. The process begins with data preparation, to mitigate the issues such as presence of sampling bias in the species records. Strategies such as resampling occurrence data, or thinning oversampled areas by removing neighboring presences in the core of the distribution are some of the methods used to reduce bias in occurrence data (Guisan et al., 2017). Selecting a few ecologically meaningful covariates that directly influence species distributions is also crucial for maintaining model interpretability and ecological validity (Elith & Leathwick, 2009; Guisan et al., 2017). Predictive reliability is further strengthened through cross-validation techniques, which test the model's performance by evaluating it on unseen data (Guisan et al., 2017). This process ensures that the model's predictions remain robust across different subsets of the data and reduces the risk of overfitting (Hastie et al., 2009). Moreover, the introduction of ensemble forecasting has significantly reduced variance in model outputs. By averaging predictions from multiple modeling algorithms, ensemble methods provide a consensus view that balances individual model strengths and weaknesses, thereby increasing overall prediction accuracy and reliability (Araújo & New, 2007). These improvements enable SDMs to provide more reliable insights for understanding species distributions, such as unraveling the ecological factors that drive the geographical distribution of species.

In particular, the exploration of SDMs at the edge of the distribution range can provide crucial insights into the gradient between favorable and unfavorable environmental conditions. Locations at the edge of the distribution range are often representative of a delicate balance between suitable and unsuitable habitats for the target species (Reino et al., 2006). As such, species populations in these areas frequently show stronger responses to environmental factors than populations closer to the core of the distribution (B. J. Anderson et al., 2009). Additionally, impacts caused by global change can have profound negative consequences for populations in these areas since they suffer from diminished influx of individuals due to increased distance from populations at the core of the distribution range (Tellería, 2018a). As such, any impacts caused by global change are expected to have an increased effect on populations at the edge of the distribution, making SDMs a valuable asset in assessing habitat suitability changes on the margins of the distribution (B. J. Anderson et al., 2009).

This thesis will examine the impacts of global change on the Red-backed Shrike (*Lanius collurio*, Linnaeus 1758), focusing on its breeding range on the Iberian Peninsula. This bird species has a large breeding range across Europe and Asia (Lefranc & Worfolk, 1997), with the Iberian Peninsula representing the western edge of its breeding distribution (Tellería, 2018a). Though it is considered "Least Concern" in terms of conservation status due to its large range (IUCN, 2016), it has historically suffered declines in several parts of its range, as for instance in the Iberian Peninsula with population declines as high as 58% in some areas (Tellería, 2018b).

The Red-backed Shrike primarily inhabits temperate areas with a mosaic of land-uses, where open areas are complemented with perching spots (bushes, branches of trees, fencing wire) to improve detection and capture of prey (Pedersen et al., 2018; Yosef et al., 2020). Consequently, declines associated with this species are associated with land-use changes and altering climatic conditions (Tellería, 2018b). Even though it is possible that these breeding populations were also affected by other factors (e.g., habitat degradation in wintering grounds), radical changes in climate and land-use are considered to be the main driving force of range shifts in biodiversity (Araújo et al., 2005; Tellería, 2018a; Lehikoinen et al., 2019). For instance, in northwest Portugal land-use intensification (e.g., homogenization of the landscapes and intensive maize crops), has been described as an important factor that might explain the decline or disappearance of this species (Reino, 2005).

To better understand these shifts, this study will use SDMs to assess how environmental conditions drive habitat selection of the Red-backed Shrike. This will allow for a reliable evaluation of the potential impacts that global change might have on its breeding distribution, and highlight areas with high levels of habitat suitability, improving local conservation efforts.

Objectives

Understanding the ecological factors influencing species distributions is fundamental in anticipating populational responses to global change. The Red-backed Shrike was chosen as the model species for this thesis due to its extensive breeding range and sensitivity to climatic and land-use changes (Reino et al., 2006; Pedersen et al., 2018). As such, this study focuses on modeling the breeding distribution of this species across its entire range, with particular attention to the Iberian Peninsula in the edge of its distribution. The results of this work may contribute to a better understanding of how the Iberian populations of the Red-backed Shrike could be impacted by ongoing environmental changes.

More specifically, this study aimed to answer the following key questions:

1. How did the different model algorithms perform?

Given the different mathematical expressions and model parameters, it is expected that using multiple modeling algorithms will result in different predictions. Therefore, the first objective is to determine which of the models had the highest predictive accuracy. Previous works (e.g., Valavi et al., 2022)) using different model algorithms, documented that species distribution models using machine-learning approaches scored higher than linear regression, and that building an ensemble from models with good performance led to increased predictions. Therefore, it is expected that machine-learning algorithms will have higher predictive accuracy, and that creating an ensemble forecast will produce an even more accurate prediction.

2. How was the Red-backed Shrike's breeding range affected by bioclimatic and land-use variables?

By analyzing the relative importance of the different explanatory variables such as bioclimatic and land-use factors, it was assessed how these variables shape the species' breeding distribution. Previous studies (e.g., Reino et al., 2006; Tellería, 2018a) have shown that temperature, precipitation and land-use highly influence habitat suitability for the Red-backed Shrike. Therefore, it is expected that these variables will have high impact on habitat suitability, and that locations with temperate climate, moderate annual precipitation and land-uses such as agriculture and open spaces will provide optimal breeding conditions.

3. What was the model's predicted distribution of the Red-backed Shrike?

Using probabilistic modeling approaches, distribution maps were created representing the likelihood of presence of the Red-backed Shrike across its breeding range. SDMs often show increased variance in predictions at the edges of species range, since environmental conditions in these areas tend to differ from locations near the core of the distribution (B. J. Anderson et al., 2009). This will likely cause model predictions to be accurate near the core of the breeding range, while predictions on the edges of the distribution will probably show greater uncertainty. As such, the predicted distribution is expected to be very accurate in most of the European range, with regions near the margins (e.g., Scandinavia, Iberian Peninsula) showing more divergence between model predictions. Moreover, due to the nature of limited sampling in Asian countries, the eastern breeding range may show higher levels of uncertainty. In the

Iberian Peninsula, northern regions are expected to have high suitability for the Red-backed Shrike, due to their climatic similarities to central Europe (Reino et al., 2006). In contrast, mountainous ranges in the center of the Iberian Peninsula are also known to represent the southern limit of the breeding range, although the extent of which will likely vary between modeling approaches.

4. How would global change affect the breeding distribution?

Human-driven global change has been shown to impact biodiversity, including species distributions (Araújo & Rahbek, 2006). Other studies (e.g., Barbet-Massin, Thuiller, et al., 2012) show that the majority of European breeding birds will abandon southern breeding grounds, while expanding their northern range into new territories. As such, it is expected that global change will impact the future range size of the Red-backed Shrike, with southern populations suffering the highest decreases of habitat suitability, and previously unsuitable areas in the North will be occupied by new populations. In a similar pattern, it is expected that the breeding populations in the Iberian Peninsula will concentrate in areas associated with lower temperatures in the northern regions, abandoning most of their breeding grounds in the central system.

2 Methods

2.1 Model species

The Red-backed Shrike (*Lanius collurio*, Linnaeus 1758) is a small passerine from the Laniidae family, a strictly carnivorous family. Its main diet is composed of a large array of prey species, such as small invertebrates, micromammals, smaller passerines and reptiles (Tryjanowski et al., 2003). The Laniidae family is known for its behavior of impaling prey into thorns or spikes found in plants or even barbed wire, which can also be observed in the Red-backed Shrike (Lefranc & Worfolk, 1997). This species is classified as "Least Concern" in the IUCN Red List due to its very large and relatively stable distribution and population size, despite a trend of decreasing number of individuals between 1970-1990 ((IUCN, 2016). The breeding distribution has a widespread range, extending from Portugal to Russia, and from Finland to Israel (Lefranc & Worfolk, 1997).

Figure 2.1: Picture of male Red-backed Shrike (*Lanius collurio*) perching on top of bush. This species depends on elevated perching spots associated with open areas. *Author: Manuel Ribeiro*

Within the Iberian Peninsula, populations are mostly restricted to coastal and mountainous regions in the north, and strictly mountainous regions in the center (Martí & Sociedad Española de Ornitología, 2003; Reino et al., 2006). The southern distribution margin is restricted by very dry and warm summers (16 °C July isotherm), and more regionally in the Iberian Peninsula its range is primarily limited by temperature (20 °C August isotherm) and precipitation (600 mm August isoline) (Reino et al., 2006).

Overall, in its southwestern breeding distribution edge, this species tends to be restricted to higher elevations, largely avoiding areas of altitudes below 750-800 m (Reino et al., 2006).

The Red-backed Shrike breeds in a variety of open habitats, preferring ecotones between forests, grazing areas and regeneration forests (Lefranc & Worfolk, 1997; Harris, 2000). It inhabits a great variety of warm and even semi-arid habitats with open areas, such as shrubby grasslands, low-intensity farmland interspersed with heaths, hedgerows, bogs and marshes, forest clearings, orchards and even city parks, roads and railway-lines (Lefranc & Worfolk, 1997; Harris, 2000). As a general rule, suitable areas have low trees and bushes (between 1-3m of height) covering 5-60% of the territory, and diverse substrates with mix of high and low vegetation, with patches of naked soil or very short grass to improve prey detection and capture (Lefranc & Worfolk, 1997; Harris, 2000; Reino, 2005). Vegetation height and cover thus become very important factors for the occurrence of this species, as it cannot hunt without vantage points from which it detects prey, and chooses shrubs and small trees around 1m above ground to build its nest (Lefranc & Worfolk, 1997; Pedersen et al., 2018). It largely avoids breeding in low-lying coastal plains, very dry areas and cold rainy regions (Lefranc & Worfolk, 1997; Harris, 2000).

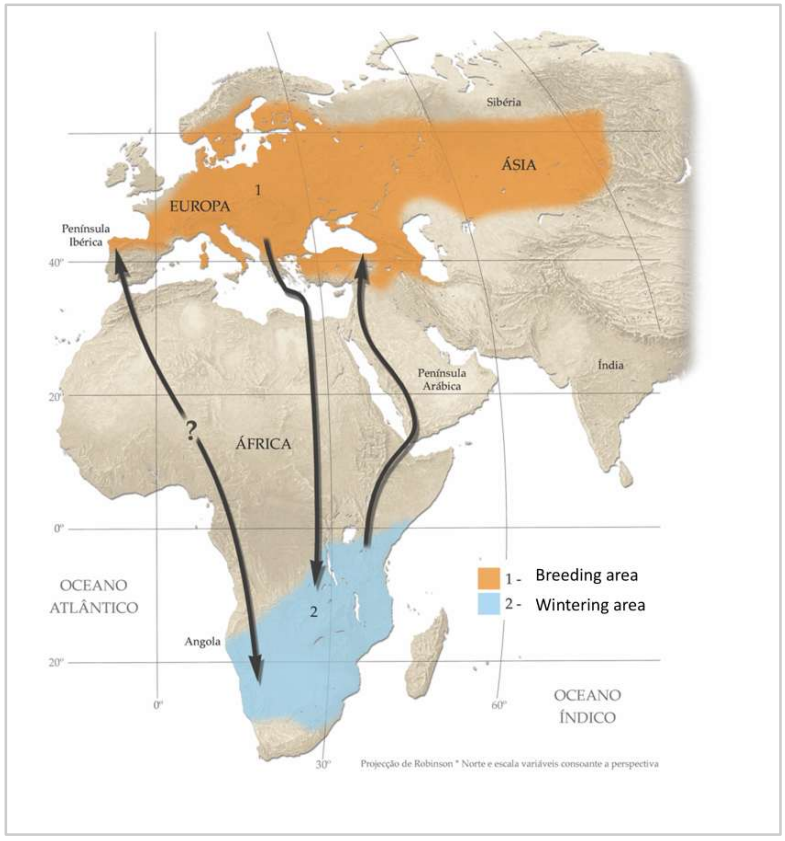

Figure 2.2: Map of breeding and wintering grounds, and migration route taken. Picture adapted from (Reino, 2005). Original illustrator: *Nuno Farinha*.

As a long-distance trans-Saharan migrant, the Red-backed Shrike spends its non-reproductive season in the eastern and southern part of Africa (Figure 2.2; Lefranc & Worfolk, 1997). It has one of the longest migration routes of all shrikes, covering thousands of kilometers per migration (Reino, 2005). In spring,

the species follows an eastern route through Africa to the Red Sea coast, from where each population spreads into their breeding territories (Lefranc & Worfolk, 1997). The first individuals arrive in their Palearctic territories between late April and late May, with some populations on the edges of the distribution, such as northwest Portugal, arriving as late as mid-June (Reino, 2005). Post-breeding movements begin in early July, with most European adults leaving their territories between mid-July and mid-August (Lefranc & Worfolk, 1997). Western European populations follow a migration path known as loop migration, where instead of directly crossing the Mediterranean Sea, they fly East until reaching Greece and Turkey, before heading South to their wintering grounds in Africa (Lefranc & Worfolk, 1997; Harris, 2000). Eastern populations migrate towards the Middle East, before continuing South to their wintering grounds (Lefranc & Worfolk, 1997). The peak of passage through these regions occurs between late-August and early-September (Lefranc & Worfolk, 1997; Harris, 2000; Reino, 2005).

2.2 Data

Predicting the distribution of the Red-backed Shrike using SDMs relies on linking occurrence data to relevant environmental factors, such as climate and land-use. By analyzing patterns between species presence or absence and observed habitat conditions over a wide area, these models attempt to reveal how climatic and/or land-use conditions influence the distribution of the Red-backed Shrike.

2.2.1 Occurrence Data

Initially only citizen science data was considered for this study, which was obtained through the Global Biodiversity Information Facility (GBIF.Org User, 2024; last accessed 06 June 2024) providing human-made presence-only observations for all available years. The obtained data was restricted to locations where the Red-backed Shrike was detected during June, which represents a period of the breeding season where the likelihood of migratory individuals being included is reduced. Furthermore, a mask was applied based on an expert breeding range of BirdLife International & NatureServe (IUCN, 2016), to discard observations of birds that deviated from known breeding grounds. This way, observations were considered to likely be breeding individuals in suitable locations.

Most of the data obtained from GBIF had a noticeable sampling bias, with Central and Western Europe being very well covered, while Eastern Europe and Asia were undersampled. The missing information from Eastern breeding grounds could severely limit the reliability of model predictions, as SDMs require the full range of a species to capture all the suitable habitat requirements (Albert et al., 2010). However, with the release of the second European Breeding Bird Atlas (Keller et al., 2020, last accessed 13 October 2021), a significant amount of missing information, mainly from eastern Europe, was made available.

As such the complete dataset for the whole breeding range was compiled from two sources: (1) Presence-only data from GBIF (GBIF.Org User, 2024) and (2) Presence-absence data from the European Breeding Bird Atlas (EBCC, 2022). Both datasets were aggregated at 0.5º resolution (0.5 x 0.5 º grid cells), where longitudinal and latitudinal coordinates are 0.5º apart from neighboring cells. Density was not taken into

consideration, meaning that only one presence needed to be reported for the species to be considered present in any given 0.5 x 0.5 º cell. In order to preserve data with higher accuracy, locations where both citizen science and breeding atlas data sets had information, breeding atlas data was prioritized (Elith et al., 2006).

Adding information from the whole breeding range of the Red-backed Shrike was essential to include the full habitat conditions that are suitable for the species (Araújo et al., 2019). Excluding either GBIF or EBBA2 presences could result in models failing to capture the full gradient of suitable bioclimatic and land-use conditions, severely compromising the reliability of the models (Albert et al., 2010).

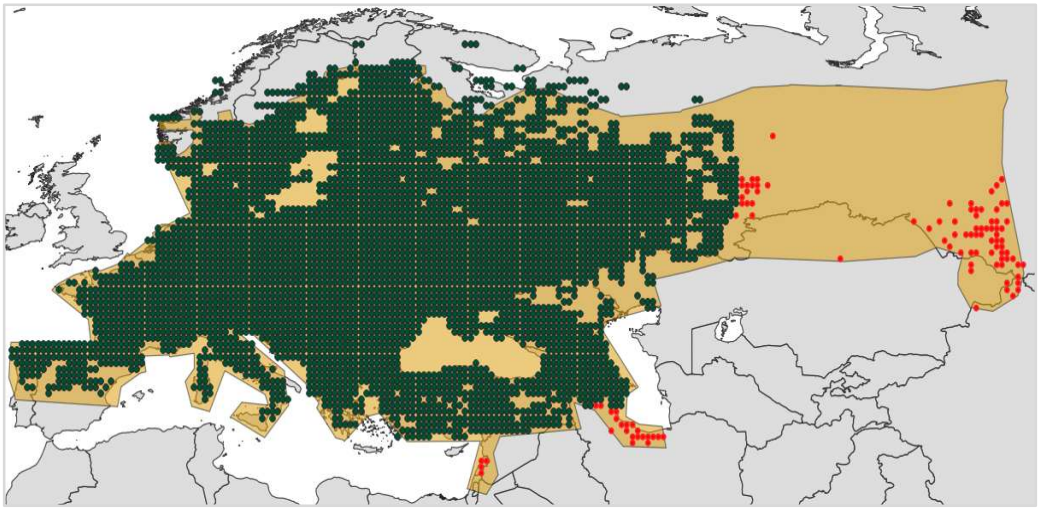

Figure 2.3: Map of GBIF (red) and EBBA2 (green) combined presence points across the whole breeding range. Expert knowledge range polygon is represented in light orange, and all GBIF observations that were outside of this area were removed to reduce the inclusion of dispersed individuals. Presence points from EBBA2 outside of the polygon were retained due to it being a coordinated effort to detect breeding birds. Figure created with QGIS, version 3.34.1.

Even with the inclusion of data from the European Breeding Bird Atlas, presence records for the Red-backed Shrike were heavily concentrated in its Central European breeding areas (Figure 2.3). To address this sampling bias and reduce spatial autocorrelation, a spatial filter was applied (Segurado et al., 2006). This filter removes presence data in locations close to each other in the core of the breeding range, to balance representation across the species' range. By reducing the influence of oversampled areas, undersampled locations and their associated environmental conditions gain greater weight in the modeling process. Using the package *spThin* (Aiello-Lammens et al., 2015) the function *thin* randomly selects presence points inside the breeding range, and all grid cells in a 50km radius around those points were removed from the dataset. The thinning continued until all cells were at least 50km apart from each other. A total of 2102 presences were removed from the initial 4313 presences in the data. A buffer was added in the above procedure to avoid coordinates with latitude and longitude values that represented less than 5% of the overall area, preserving presence data in the edges of the breeding range. In the end, the dataset of the study species comprised a total of 2211 presences.

Since there was no focused effort to determine the occurrence throughout the whole extent of the breeding range, the full occurrence dataset only shows observed presences and no observed absences

(Guisan et al., 2017). Although presence-only models have been used to predict species distributions, these are generally outperformed by regression and machine-learning approaches that require absence data (Elith et al., 2006). An alternative to observed absences is to treat locations where the species was not detected as so-called pseudo-absences (Ponder et al., 2001; R. P. Anderson, 2003). This allows for the models to have an environmental contrast between where the species is present and where it is, supposedly, absent (Phillips et al., 2006). With this approach, modeling techniques that were previously impossible with only presence data, such as regression techniques like Generalized Linear Models, are now available to be used (Guisan et al., 2017). A group of 10 000 randomly selected pseudo-absences was created inside of the study area (Barbet-Massin, Jiguet, et al., 2012; Guisan et al., 2017).

When determining the study area from which to draw the pseudo-absences, there is no consensus or set universal method (Guisan et al., 2017). Pseudo-absences should reflect locations within reach of the Red-backed Shrike but present unsuitable habitat conditions for the occurrence of the species (Barbet-Massin, Jiguet, et al., 2012). For this study four different pseudo-absences designs were considered (Figure 2.4):

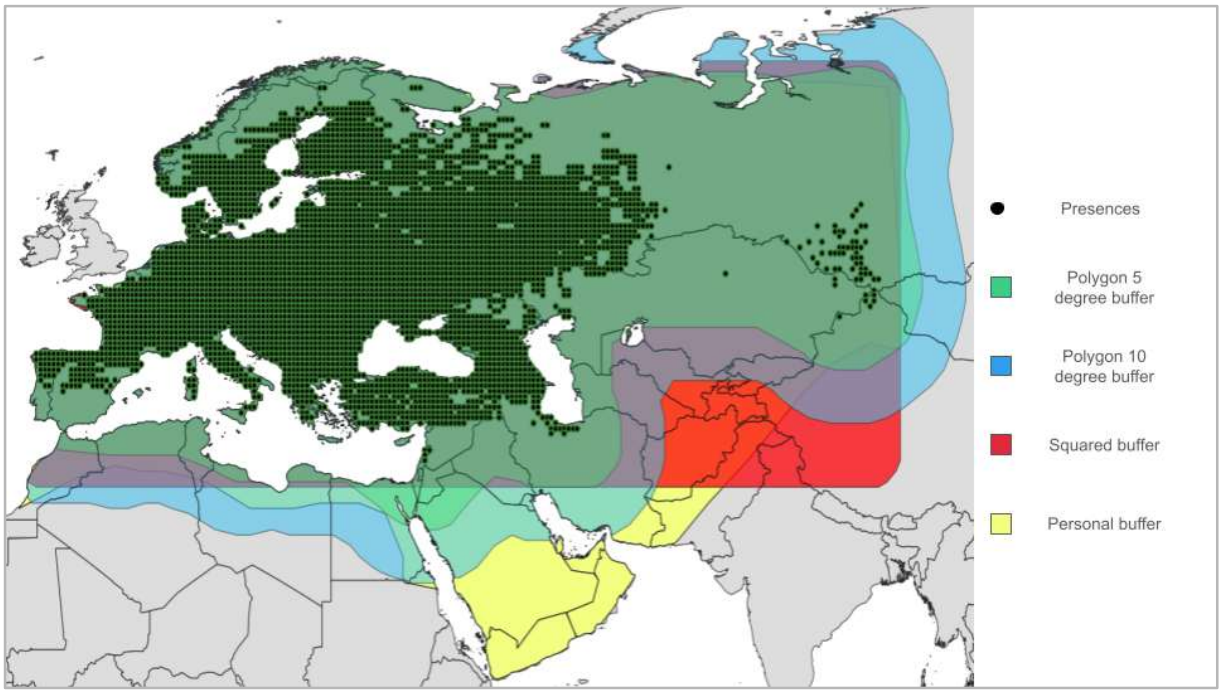

Figure 2.4: Map of all sampling area designs considered. Points correspond to presences in the date, and the rest of the cells inside the colored polygons correspond to locations from where to sample pseudo-absences. Figure created with QGIS, version 3.34.1.

- A squared pseudo-absence area created from maximum and minimum coordinates values of longitude and latitude for presence points. A buffer of 2.5º was added to include locations proximal to the margins of the distribution.
- Two polygon pseudo-absence areas derived from the expert knowledge breeding map. A buffer of 5º and 10º was added to expand from the periphery of the known breeding area, creating an equal distance between the edge of the breeding area and the periphery of the pseudo-absence polygon.

- A custom pseudo-absence area derived from the migration route. If populations of Red-backed Shrike are crossing over territories, but not stopping in any of them to breed, it was assumed that the locations were likely unsuitable. This comes from an assumption that migratory animals aim to balance energy wasted with maximized success of breeding. Parts of North Africa were also added due to the proximity to presence points in the Mediterranean basin, ensuring a contrasting environmental gradient (Barbet-Massin, Jiguet, et al., 2012).

Though methods for using the same dataset to calibrate and evaluate the models were available, it is recommended to test model's predictive performance in independent datasets (Guisan et al., 2017). Since the spatial focus of this study was in the Iberian Peninsula, an independent dataset was compiled using geographical coordinates of reported Red-backed Shrike breeding sites from the Portuguese Breeding Atlas (Instituto da Conservação da Natureza, 2008) and the Spanish Breeding Atlas (Martí & Sociedad Española de Ornitología, 2003). By restricting evaluation data to census, the occurrence data was expected to more accurately represent the real distribution of the Red-backed Shrike in the Iberian Peninsula. Furthermore, the use of independent data sets for evaluation is recommended due to the innate independence from the calibration set, ensuring that model predictive performance is not influenced by overfitting (Guisan & Zimmermann, 2000).

However, the Iberian Peninsula is part of the breeding range of the Red-backed Shrike and should be included in the calibration dataset. This was expected to probably reduce the independence between calibration and evaluation datasets, defeating the purpose of using an independent dataset to begin with. This highlights the trade-off in SDMs between including as much information as possible vs introducing biases (Guisan et al., 2017). As such, an additional analysis was done where the Iberian Peninsula was excluded from the calibration dataset to maintain independence from the evaluation census data. Should the resulting model show high predictive accuracy, then excluding the Iberian Peninsula from the calibration data would be a reasonable procedure. Otherwise, the information within this region would be included into the model calibration at the expense of reducing independence to the evaluation dataset.

2.2.2 Habitat covariates

To uncover the relationship between the Red-backed Shrike and its breeding range, a group of explanatory variables was used to understand how environmental conditions drive the species' occurrence. All of the selected variables were shown to potentially influence the distribution of bird species (Araújo et al., 2005; Neto & Araújo, 2011; Strubbe et al., 2013), and are commonly used in species distribution modeling. Two different groups of independent variables were used (Table 2.1): bioclimatic and land-use factors.

For bioclimatic variables, the focus was on selecting variables that reflect water and energy in the form of temperature, since these are known to have a widespread impact on the physiological limitations of several species (Araújo et al., 2005). The variables chosen for this study include: annual mean temperature ('tmean'), mean temperature of the warmest month ('tmax'), mean temperature of the coldest month ('tmin'), temperature variation with the seasons calculated with standard deviation*100 ('tsd'), annual precipitation ('prec') and precipitation seasonality through coefficient of variation ('precsd'). Variables that fall out of the breeding season, such as 'tmin', are still added since they impact other habitat qualities (vegetation for example) potentially leading to an indirect effect on the

distribution of the Red-backed Shrike (Reino et al., 2006). These variables were derived and resampled from monthly raw data provided by the CliMond dataset (Kriticos et al., 2012) at 0.5º resolution for both the current period (1961-1990) and the future scenarios (2050).

Table 2.1: Table of all environmental covariates considered for this work. The table is divided into two parts for bioclimatic and land-use variables. All variables have their abbreviated name, what they represent, measurement used for bioclimatic covariates and categorical level for land-use covariates.

Type	Name	Description	Measurement/Category
Bioclimatic (multiple variables)	tmin	Mean temperature of coldest month	ºC
	tmean	Mean of annual temperature	ºC
	tmax	Mean temperature of warmest month	ºC
	tsd	Temperature seasonality	standard deviation x 100
	prec	Annual precipitation	mm / 1 m^2
	precsd	Precipitation seasonality	coefficient of variation
Land (single categorical variable)	other	Unsuitable land	land1
	shrub	Shrub land	land2
	open	Open land	land3
	forest	Forested land	land4
	agri	Agricultural land	land5

Landscape structure, including the organization of vegetation and substrate, is also a crucial habitat component for the Red-backed Shrike (Harris, 2000; Pedersen et al., 2018). As such, a land-use variable ('land') was incorporated into the set of explanatory variables. It was obtained from the IMAGE project (IMAGE contributors, 2017; accessed April 2017) and reclassified as a categorical variable with five categories: forest areas, open land (such as grassland and savannas), shrub areas, agricultural areas and others (ice, desert and urban areas that are considered unsuitable for the species). The 'other' category was used as the reference level, to determine how different land-use types affect habitat suitability in comparison with unsuitable land-use. The complete land-use composition was compiled at 0.5° resolution for both current (1961-1990) and future (2050) periods.

To avoid introducing noise to the model and unnecessarily increasing complexity, a variance inflation factor (VIF) analysis was performed to account for correlation and multicollinearity in bioclimatic variables due to the close association of them (Guisan et al., 2017). The function *vifstep* from the package *usdm* (Naimi et al., 2014) applies a stepwise elimination of highly inflating variables. Furthermore, a

McNemar's test was also conducted in the case of persisted moderate correlation between variables, to check if the binary predictions between models with and without correlated variables produce significantly different results (McNemar, 1947).

SDMs also benefit from assessing a metric called variable importance, which measures the influence of each covariate in changing model predictions (Smith & Santos, 2020). In this work variable importance was checked using a method called variable permutation. If "permuting" (randomizing) a variable leads to substantial changes in model predictions, then the correlation between the original model predictions and the "permuted" model predictions will have a low value (Smith & Santos, 2020). Higher changes between model predictions lead to lower correlation, implying a greater influence of the covariate. Pearson's correlation coefficient was calculated, and each variable is dropped in turn and replaced with a "noise" variable to assess a variable's overall importance on model fit and predictive capability (Abdul-Wahab et al., 2024).

2.2.3 Sampling effort (BIAS variable)

Sampling bias is one of the most pervasive issues in SDMs (Beck et al., 2014; Guisan et al., 2017). It is relatively common for observations in online databases to have well surveyed areas near densely populated locations, while more remote and hard to traverse areas suffer from a lack of existing data (Ponder et al., 2001; Beck et al., 2014). This could cause problems for inferring the habitat suitability of the Red-backed Shrike since the Eastern part of its breeding range is poorly surveyed (Figure 2.3) and its real distribution probably wasn't correctly represented in these areas (Albert et al., 2010).

To help account for sampling bias in the data, a variable 'bias' was created, which represents the density of bird species detected for each location. The variable was created by extracting occurrences of bird species from GBIF during the breeding months (Ponder et al., 2001; R. P. Anderson, 2003), and employing a kernel-density function (Strubbe et al., 2013) to rescale the variable values between 1 (low density) to 10 (high density), reducing numeric disparities (Elith et al., 2010). This creates a density map of common bird species presences as a representation of sampling effort, where locations with higher number common bird species reported are considered to have higher sampling effort (Meyer et al., 2015). Training the models with this variable will likely lead to a high association of detecting the Red-backed Shrike with high density of common birds (higher sampling effort corresponds to higher probability of detecting the target species). Afterwards, when projecting the distribution of the Red-backed Shrike, the density of common bird species is set at its maximum value (in this case 10) for the whole sampling area, in an effort to create a uniform and intense sampling effort and counter the effects of sampling bias (R. P. Anderson, 2003; Albert et al., 2010).

2.3 Statistical Modeling

With both the occurrence data and the variables created, a modeling procedure was applied to analyze how the response variable (occurrence) responds to different value combinations of the explanatory variables (temperatures, precipitation, land and bias). There are multiple modeling algorithms that could be used to achieve this, each of them with its own strengths and limitations (Guisan et al., 2017). For

this work, the algorithms selected were composed of 2 different approaches: a statistical approach using Generalized Linear Models (GLM); and machine-learning approaches using Maximum Entropy (MaxEnt), Gradient Boosting Machine (GBM) and Random Forests (RF). The main difference between these groups is that statistical approaches often assume the data follows an appropriate distribution and the parameters for the distribution model are then estimated from the data, while machine-learning approaches normally don't have this assumption and just learn the relationship between the response variable and the predictors directly from the data (Elith et al., 2008). Additionally, in machine-learning literature covariate transformations are often referred to as 'features' and were addressed as such in this work (Elith et al., 2011). A detailed description of each modeling algorithm is presented in the following section.

Once the algorithms were chosen, the amount of information from the available data to calibrate the models needed to be selected. In this thesis two calibration sets were used: 'full models' calibrated with 100% of the data and 'single models' calibrated with 70% of the data. Models calibrated with 100% of the data are usually not recommended since this increases the likelihood of overfit (Guisan et al., 2017). As such, 'full models' were restricted to exploratory analysis of predicted distribution and overall performance between algorithms, while 'single models' were used to create the final ensemble model.

To reduce uncertainty caused by sampling artifacts, 10 replicates were created for each of the four algorithm 'single models' (Araújo & New, 2007). Every replicate's performance was evaluated through external validation using 70% of the compiled occurrence dataset to calibrate (train) the model parameters and an independent census dataset for the Iberian Peninsula to evaluate (test) the model's predictive performance. In the end, 4 'full models' and 40 'single models' were created.

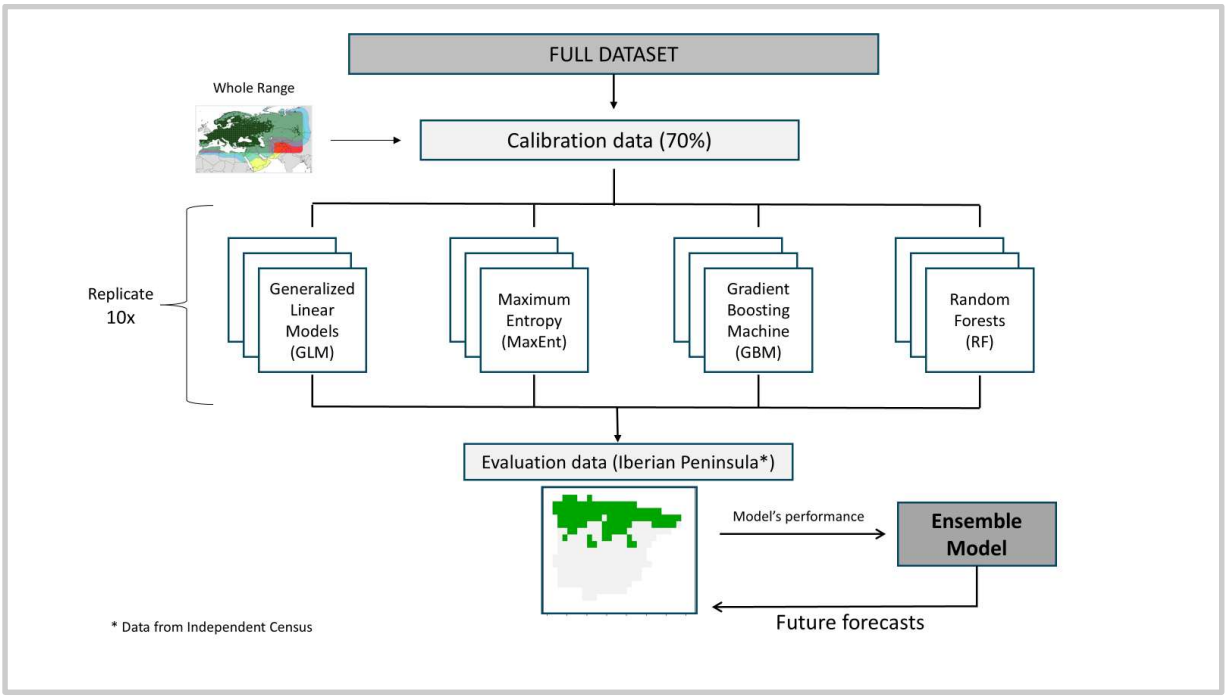

Figure 2.5: Illustration of the modeling procedure used in this work.

The modeling process was done in four steps (Figure 2.5): (1) run the four different techniques (2) use cross-validation to check each model's performance (3) apply an ensemble technique to aggregate predictions into a more robust final model; and (4) use the final ensemble model to project the Red-backed Shrike's distribution into current and future conditions. All modeling steps were done in R programming language (v4.1.1; (R Core Team, 2021)) using the RStudio statistical software (RStudio Team, 2021), mainly using functions from the *biomod2* (Thuiller et al., 2012) and *SDMTools* (VanDerwal et al., 2014) packages, two commonly used packages for species distribution modeling.

2.3.1 biomod2 package

With the increased development of analytical tools to study the relationship between organisms and their preferred habitat conditions, the R package *biomod2* was created to model species distributions (Thuiller et al., 2009). This package is highly flexible, with a large selection of different modeling techniques, such as linear regression and decision trees, including the option to tune specific model algorithm parameters, and the ability to create subsets of the data to fit different models, creating an unlimited amount of model replicates trained on different samples of the data (Thuiller et al., 2009). Outputs include a diversity of model accuracy metrics, from which model replicates are selected to incorporate a final ensemble model, with different options to aggregate single model predictions (e.g., simple average or weighted average), and response curves to visualize how the occurrence varies with each explanatory variable (Thuiller et al., 2009). The ability to customize a variety of modeling steps and streamline the whole analysis made *biomod2* one of the most used packages for modeling species distributions (Valavi et al., 2022).

In this work *biomod2* functions were used in all major steps of the modeling framework:

1. Formatting data

The function *BIOMOD_FormatingData* enables a quick preparation of all available data into the correct format required for modeling species distributions. This includes identifying presence locations, which covariates to be used, and where and how pseudo absences should be created. As previously mentioned, 10 000 pseudo absences were randomly generated inside the study area.

2. Tuning model options

Before running any models, modeling options needed to be set using the *BIOMOD_ModelingOptions* function. This included selecting if linear regression should use strictly linear functions, how many decision trees to run or how algorithms should account for overfit (e.g., regularization). Should users not have any preferences or knowledge of which options to set, the *biomod2* package already incorporates predetermined 'default' options for each algorithm. In this work, all modeling options were set to 'default'.

3. Running models

After formatting the data and tuning modeling options, models were fitted to the occurrence data using the *BIOMOD_Modeling* function. In this function, model algorithms to be fitted were chosen, as well as determining how many replicates to create and what percentage of data should be randomly subset to

calibrate model coefficients, were some of the settings available. Evaluation metrics were also automatically calculated and stored as an output of this function.

4. Aggregating predictions

Once all models were fitted and performance metrics made available, the *BIOMOD_EnsembleModeling* function was used to create an ensemble from the best performing models. This function allows users to select an evaluation metric and an associated threshold value, and automatically aggregates probability predictions from single model replicates that meet this condition.

5. Projecting to future scenarios

Finally, the ensemble model was used to project into environmental conditions predicted by future global change scenarios. This was done using the functions *BIOMOD_Projection* which fits the previously chosen single models into future environmental predictions, and *BIOMOD_EnsembleForecasting* which combines the newly acquired model projections into the future using the settings of the ensemble model established in the previous step. The final output was the ensemble model projections into future scenarios, enabling the assessment of global change impacts on the distribution of the Red-backed Shrike.

2.3.2 Modeling Algorithms

To assess how the presence or absence of the Red-backed Shrike relates to the explanatory variables, four different modeling algorithms were used. Using different algorithms to predict the occurrence of a species can give a more complete insight on the drivers of the species distribution (Araújo & New, 2007). All of the following modeling algorithms are commonly used in SDMs and have shown to be able to produce reliable predictions (Elith et al., 2006; Guisan et al., 2017). Model calibrations were done using the 'biomod2' package (Thuiller et al., 2012) in RStudio.

2.3.2.1 Generalized Linear Models

Regression approaches are the most commonly used in species distribution modeling (Elith et al., 2008; Guisan et al., 2017). In these models, the presence or absence of the study species is related to a set of pre-selected environmental predictors through regression analysis (Guisan et al., 2017). Compared to classical linear models, Generalized Linear Models (GLMs) present a more flexible approach since they allow the response variable to follow distributions from the exponential family (Hastie et al., 2009), such as the binomial distribution used for presence-absence data. This is achieved by the use of a link-function which relates the linear combination of predictors to the mean of the response variable, while constraining the predicted values within the range of the binary data (Guisan et al., 2017). Furthermore, relationships between the linear predictors and species occurrence can be transformed into higher-order polynomials to capture non-linear relationships, such as unimodal or bimodal response curves (Hastie

et al., 2009; Guisan et al., 2017). The default setting of the *biomod2* package restricted transformations up to quadratic functions with no interaction between covariates.

The occurrence data of the Red-backed Shrike followed a binomial distribution, which typically uses the logit link function to fit a linear model to the data (Hastie et al., 2009). The result was a binary logistic regression that modeled the probability (π) of the Red-backed Shrike being present in a location ($Y = 1$), given the values of the k explanatory variables (x):

$$\pi(x1,\ldots,xk) = P(Y = 1 \mid x1,\ldots,xk), \text{ or simply '}\pi\text{' for convenience.} \quad (2.1)$$

The logit model was expressed as the log-odds of the probability of the occurrence of the Red-backed Shrike:

$$log\left(\frac{\pi_i}{(1-\pi_i)}\right) = \beta 0 + \beta 1 \cdot x1_i + \beta 2 \cdot x2_i + \beta 3 \cdot x3_i + \ldots + \beta k \cdot xk_i \quad (2.2)$$

where π_i is the probability of the target species being present at location i, β0 is the intercept term that gives the log-odds when all covariate values are 0, (β1, β2, β3,...) are the coefficients of the covariates and $x(1,2,3\ldots,k)$ are the values of each covariate for location i (Guisan et al., 2017).

The logit equation can be further solved to express the probability of occurrence at each location given the values of the explanatory covariates:

$$P(Y = 1 \mid x_i) = \pi_i = \frac{e^{(\beta o+\beta 1\cdot x1_i+\ldots+\beta k\cdot xk_i)}}{(1+e^{(\beta o+\beta 1\cdot x1_i+\ldots+\beta k\cdot xk_i)})} \quad (2.3)$$

From this final equation GLMs produced the predicted probability values of occurrence of the Red-backed Shrike.

Without previous information on which variables should be included in the model, a variable selection procedure was implemented to produce the most parsimonious combination of explanatory variables (Guisan et al., 2017). For this, the GLM algorithm in *biomod2* applied a stepwise regression for examining the relative importance of each explanatory variable to explain the occurrence of the Red-backed Shrike (Guisan et al., 2017). The *biomod2* package used a combination of forward and backwards selection which started with the intercept model and at each step added and removed variables based on AIC (Akaike Information Criterion; Akaike 1974) (Thuiller et al., 2012; Guisan et al., 2017). At each step an AIC value was calculated for each model with different explanatory variable combinations, until a final model with the lowest AIC value was retained (Guisan et al., 2017).

2.3.2.2 Maximum Entropy

Maximum Entropy (MaxEnt) is a widely used machine learning algorithm known for its high predictive performance (Phillips et al., 2009; Elith et al., 2011; Guisan et al., 2017; Valavi et al., 2022). MaxEnt relies on the principle of maximum entropy, it assumes a uniform species distribution across the study area and adjusts this distribution according to the available environmental constraints at sampled locations (Phillips et al., 2006; Elith et al., 2011). It models the occurrence of the Red-backed Shrike by defining a probability distribution over all the grid cells in the study area, where presences serve as

sample points, and their associated environmental characteristics are the explanatory variables (Guisan et al., 2017).

Conceptually, in MaxEnt $f(z)$ is defined as the probability density of habitat conditions across the whole study area and $f1(z)$ as the probability density of predictors at locations where the Red-backed Shrike was present, given the set of climate and land-use covariates selected (z) (Elith et al., 2011; Guisan et al., 2017). The probability of presence of the Red-backed Shrike ($y = 1$) can be obtained by applying the Bayes' rule:

$$P(y = 1 \mid z) = \frac{f1(z) \cdot P(y = 1)}{f(z)} \tag{2.4}$$

Before considering the presence data, $f(z)$ can be seen as a null model since there is no reason to expect the Red-backed Shrike to prefer any particular environmental conditions, and it is expected to occupy suitable conditions proportionally to their availability in the region (Guisan et al., 2017). However, the probability of being present is not directly extractable from presence-only data (Elith et al., 2011). Instead, MaxEnt obtains the so called "raw output" by estimating the ratio $f1(z)/f(z)$, which gives insights into the importance of habitat features in determining the occurrence of the Red-backed Shrike. The raw output is then transformed into a probability estimate using its logistic output, which, similarly to the logit function in GLMs, expresses the log-odds of occurrence and adjusts it to approximate the probability of presence of the Red-backed Shrike (Elith et al., 2011):

$$\eta(z) = log(\frac{f1(z)}{f(z)}) \tag{2.5}$$

The coefficients of the covariates are then readjusted, so that the probability of occurrence of the Red-backed Shrike for average conditions at locations where the species is present (expressed as the parameter τ in MaxEnt) is 50% (Elith et al., 2011). MaxEnt then uses a number of transformation features to establish complex relationships between the response of occurrence to habitat conditions (Phillips et al., 2006; Elith et al., 2011). A total of 5 feature transformations are available - linear, quadratic, product, threshold and hinge. Within the *biomod2* package all of them were used to model the distribution of the study species. The set of available features $h(z)$ was combined with a vector of coefficients β using a Gibbs distribution from the exponential family (Elith et al., 2011):

$$f1(z) = f(z) \cdot e^{\eta(z)} \tag{2.6}$$

where $\eta(z) = \alpha + \beta.h(z)$, with α being a normalizing constant to ensure that $f1(z)$ sums to 1. The target of a MaxEnt model is $e^{\eta(z)}$, which estimates the ratio $f1(z)/f(z)$ (Elith et al., 2011).

In addition, MaxEnt calculates the difference between the probability density functions $f1(z)$ and $f(z)$, also referred to as the relative entropy (r), which measures how the habitat conditions at presence sites differ from the rest of the study area (Elith et al., 2011; Guisan et al., 2017). This difference can be calculated using the Kullback-Leibler divergence (Bishop, 2006):

$$r = -\int \quad f(z) \cdot log(\frac{f1(z)}{f(z)})\, dz \tag{2.7}$$

A logistic transformation essentially transforms the model from an exponential family equation to a logistic model, giving the model's probability prediction for the occurrence of the Red-backed Shrike:

$$P(y = 1 \mid z) = \frac{\tau \cdot e^{\eta(z)-r}}{(1-\tau+\tau \cdot e^{\eta(z)-r})} \quad (2.8)$$

where $\eta(z)$ is the logit score, r is the relative entropy of MaxEnt's estimate of $f1(z)$ to $f(z)$, and τ is the probability of presence under average conditions which was set to 0.5 (Elith et al., 2011).

MaxEnt further implements a regularization term (λ) to trade off model complexity and model fit (Elith et al., 2011). This term penalizes the features coefficients (β) by shrinking their values, aiming to only retain features with high coefficients (and therefore high impact), creating a balance between fit and complexity (Hastie et al., 2009). This helps avoid overfit and produces a more generalizable and accurate model (Elith et al., 2011; Guisan et al., 2017). All available features had an optimal regularization term automatically set by the *biomod2* package.

2.3.2.3 Gradient Boosting Machine

Within machine learning, recursive partition methods using decision trees have shown to be reliable in predicting species distributions (Guisan et al., 2017). Decision trees (Figure 2.6) apply binary recursive partitions that iteratively split the data into two nodes at a time for a given variable and an associated condition (Elith et al., 2008). Classification trees are used for discrete response variables, while regression trees are fitted to continuous responses (Hastie et al., 2009). The splitting condition involves setting a threshold for an explanatory variable, with observations that meet the established condition moving to the left node, and the remaining to the right (Elith et al., 2008). Each split is chosen based on an evaluation metric that reduces impurity (classification) or variance (regression), selecting the variable with the greatest improvement (Hastie et al., 2009). Since splitting conditions are dependent on previous conditions higher in the tree, interactions between predictors are automatically included in the modeling process (Elith et al., 2008). This process continues recursively, using different combinations of variables and conditions, until a stopping criterion, such as minimum number of observations at terminal nodes, is met (Bühlmann & Hothorn, 2007; Elith et al., 2008). The resulting decision tree represents a structured set of rules that define species habitat suitability (Elith et al., 2008).

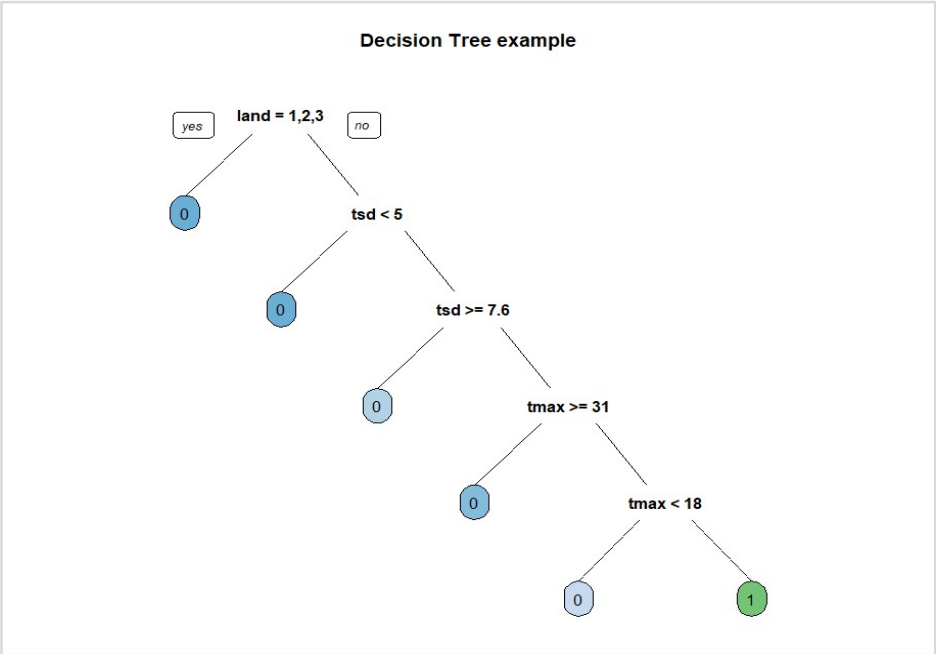

Figure 2.6: Illustration of classification decision tree showing how different covariate conditions split the occurrence data into different categories (0 for absence, 1 for presence).

A single decision tree is often not capable of capturing all of the complexity required to model the distribution of a species (Guisan et al., 2017). To address this issue, Gradient Boosting Machine (GBM), also known as Boosted Regression Trees (BRT), uses an approach called boosting, which combines several weak models into a single strong model, and is commonly used in species distribution modeling (Guisan et al., 2017). It consists of a forward stage-wise procedure, where each new model (in this case a decision tree) is fitted to the collection of previous models without changing earlier models (Guisan et al., 2017).

Boosting incorporates stochastic resampling without replacement, randomly selecting 50% of the available data at each iteration, which ensures that each final model is different every run (Elith et al., 2008). The residuals of the previous collection model, which represent variance unexplained by the models, are used to refine predictions (Elith et al., 2008). The next iteration fits a regression tree to the residuals of the previous models, with the objective of minimizing the remaining unexplained variance (Guisan et al., 2017). By repeatedly fitting simple trees to residuals, the algorithm incrementally improves performance, focusing on observations that were previously poorly predicted (Guisan et al., 2017). This iterative procedure continues until the model reaches optimal performance and the results from each tree are then combined to generate a robust model, reducing bias and variance, even with complex relationships (Elith et al., 2008; Guisan et al., 2017).

GBM optimizes performance by minimizing a loss function at each step of the iteration, which represents the differences between model predictions and observed occurrences and penalizes errors in predictions (Elith et al., 2008; Hastie et al., 2009). The *biomod2* package uses the logistic loss function for binary classifications:

$$L\big(y_i, f(x_i)\big) = -(y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i)), \quad (2.9)$$

where y_i is the observed occurrence at location i, $f(x_i)$ is the raw score of a regression function that minimizes the loss for location i and $p_i = 1 \div (1 + e^{-f(x_i)})$ and represents the probability of occurrence of the Red-backed Shrike at location i (Ridgeway, 2007). By minimizing the logistic loss, the boosting model gradually increases its predictive ability (Ridgeway, 2007). The gradient (z) represents the direction and intensity required to minimize residuals:

$$z_i = y_i - p_i \tag{2.10}$$

At each step, the modeling algorithm identifies the gradient (z), which represents the residual error, and indicates where it needs to improve its fit to the data. This gradient is positive for observed presences and negative for observed absences (Ridgeway, 2007). The algorithm then fits a regression tree to the gradients (residuals) that best aligns with their corresponding direction, reducing residual errors and improving performance (Ridgeway, 2007; Hastie et al., 2009).

Some of the critical parameters when fitting GBMs are the learning rate, the number of trees to be fitted, and the interaction depth (Elith et al., 2008). The learning rate (λ) is similar to the regularization term in MaxEnt, as it regularizes the weight given to the successive trees, to avoid fitting an overly complex model (Guisan et al., 2017). *Biomod2* set λ to 0.001, slowing down the rate of loss reduction by multiplying λ with the gradient (z), giving the boosting algorithm enough time to capture important patterns, while at the same time avoiding overfit (Ridgeway, 2007; Elith et al., 2008). The number of trees represents the total number of iterations. A large number of trees should always be used to minimize variation between each GBM model's output and to counteract the slow learning rate given by λ (Guisan et al., 2017). The total number of iterations was set at 2500 by default in *biomod2*. The interaction depth refers to the maximum depth of the trees, which relates to the number of nodes in the tree (Guisan et al., 2017). An interaction depth of 1 (single decision stump with two terminal nodes) fits an additive model, and consecutive depths of n fit a model with up to n-way interactions (Elith et al., 2008). This means that any tree with more than three nodes is able to capture interactions between predictors (Guisan et al., 2017). Interaction depth was set to 7 by the *biomod2* package.

2.3.2.4 Random Forests

Similar to GBM, Random Forests (RF) is a decision tree-based algorithm, which applies a technique called bagging (short for bootstrap aggregation), an ensemble modeling technique for classification and regression trees that reduces variance and prevents overfitting by averaging multiple decision trees trained on different subsets of the data (Guisan et al., 2017). In RF, a large number of bootstrap samples are drawn from the data with replacement, meaning some observations may be sampled multiple times while others are left out, and a decision tree is fitted to each sample (Guisan et al., 2017). Each tree is grown to maximum size, by continuously splitting until it reaches a minimum number of observations at a terminal node, which *biomod2* automatically sets at 5 (Guisan et al., 2017). In the end, a final ensemble model is created from combining (e.g., averaging) the results, which reduces the associated variance of the prediction (Guisan et al., 2017).

Random Forests were developed to avoid overfitting by adding some stochasticity to the process of building the decision trees (Guisan et al., 2017). Not only does this algorithm randomly subset parts of the data for each bootstrap sample, it also chooses a random subset of predictors to decide each best split (node) in the tree, hence the name “random” forests (Prasad et al., 2006). By randomizing the set of

predictors to choose from at each node, the correlation between each model reduces significantly, making the derived ensemble model more robust (Prasad et al., 2006). The *biomod2* package randomly selects the square root of the total number of variables as the number of variables subsetted at each node. From the set of randomly selected predictors, the split is done by the feature and associated condition that minimizes impurity in the child nodes (Hastie et al., 2009). Fitting a large number of trees, combined with the randomness introduced through bootstrapping and feature selection, stabilizes the model's predictions, making RF less likely to overfit (Breiman, 2001). By growing each tree to maximum size and selecting only the best split among a random subset at each node, RF tries to maintain some prediction strength while inducing diversity among trees (Breiman, 2001).

In this work the RF algorithm in the *biomod2* package computes a total of 500 classification trees. In classification decision trees the aggregation is done by majority vote where each tree makes a prediction (presence or absence) and the final model chooses the prediction with the most votes (Guisan et al., 2017). For each grid cell, the probability of the species being present is determined by the proportion of trees in the ensemble that vote for a presence (Prasad et al., 2006). The higher the number of trees that predict a presence for a grid cell, the higher the probability of presence of the Red-backed Shrike predicted by the model.

2.3.3 Model evaluation

To check model performance and reduce overfit a cross-validation technique was performed. Cross-validation consists of partitioning the occurrence data into two different groups, one group for training the modeling algorithm (calibration) and another for testing the predictions (evaluation) (Guisan et al., 2017). The way the data is partitioned into different samples can be done in multiple ways, depending on the cross-validation technique used. In this work, an external sample cross-validation consisted of a calibration set made from 70% of the whole aggregated data from online databases, which was used to train the model parameters, and an evaluation set made from the presence-absence data of independent censuses for the Iberian Peninsula (Instituto da Conservação da Natureza, 2008; Martí & Sociedad Española de Ornitología, 2003) to test each model’s predictive performance.

To assess the predictive accuracy of the implemented models in their ability to predict the distribution of the Red-backed Shrike in the Iberian Peninsula, a variety of evaluation metrics were calculated. Some metrics, like the Area Under the Curve (AUC) measure the general agreement between the model's predictions and the Iberian census data, while others, like sensitivity, focus on more specific aspects, such as the proportion of correctly predicted presences (Guisan et al., 2017). To ensure a comprehensive assessment of model performance, multiple evaluation metrics were computed (Elith & Leathwick, 2009).

2.3.3.1 Prevalence

Before analyzing any model, a prevalence value needs to be set. Prevalence describes the assumed likelihood that a species is present in a given location (Manel et al., 2001). In this work prevalence was set at 0.5, meaning the models assumed that the species was as likely to be present as it was to be absent from each location (Elith et al., 2010). This ensured that presences and pseudo-absences were given equal weight when calibrating the models.

2.3.3.2 Threshold

To assess model predictive performance, all of the fitted model's probability outputs needed to be converted into binary predicted presences and predicted absences, to be compared to the occurrence data. For this, a cutoff percentage that determines whether a predicted probability of occurrence corresponds to a presence (above the cutoff) or absence (below the cutoff) needed to be selected. Since any given number between 0 and 1 can be used, the applied practice is to choose a threshold that maximizes a target evaluation metric (Guisan et al., 2017). This is done by calculating the chosen evaluation metric for a range of possible thresholds, and selecting the threshold that gives the highest target metric value (Liu et al., 2016).

One of the most commonly used thresholds in SDMs is the maximization of the sum of sensitivity and specificity (Guisan et al., 2017). This gives the most accurate prediction of the real distribution of the Red-backed Shrike, allowing for a more accurate assessment of the impact global change will have in the distribution of local populations (Guisan et al., 2017). As such, the threshold selected was the one that maximized this metric for each model in this study.

Once the model's probabilistic predictions were available and a threshold selected, all of the information could be condensed into a binary contingency table (Table 2.2). From this contingency table, various evaluation metrics could be derived. The term "Absence" in this table were replaced with pseudo-absences in the data, due to a lack of controlled survey samples as previously mentioned. Nevertheless, the term "Absence" was used in this section to remain consistent with literature.

Table 2.2: Contingency table obtained after converting probabilistic predictions into binary (0 for absence, 1 for presence) predictions, to compare observed presence-absence to predicted presence-absence. Values from this table were used to calculate evaluation metrics in this section.

		Observed		
		Presence	Absence	Total
Predicted	Presence	True presence (TP)	False presence (FP)	Total predicted presences $TP + FP$
	Absence	False absence (FA)	True absence (TA)	Total predicted absences $FA + TA$
	Total	Total observed presences $TP + FA$	Total observed absences $FP + TA$	Total observations (N) $TP + FA + FP + TA$

2.3.3.3 Sensitivity and Specificity

Sensitivity measures the proportion of actual presences correctly predicted by the model, while specificity indicates the proportion of true absences accurately identified (Allouche et al., 2006).

$$Sensitivity = \frac{TP}{(TP + FA)} \tag{2.11}$$

$$Specificity = \frac{TA}{(TA + FP)} \tag{2.12}$$

Their counterparts are omission rate, reflecting the proportion of actual presences predicted as absences (false absences), and commission rate, showing the proportion of actual absences predicted as presences (false presences) (Allouche et al., 2006).

$$Omission\ rate = \frac{FA}{(TP+F\ )} = 1 - Sensitivity \tag{2.13}$$

$$Commission\ rate = \frac{FP}{(TA+FP)} = 1 - Specificity \tag{2.14}$$

Additionally, the proportion correct can be calculated, representing the overall accuracy by comparing correctly predicted outcomes to the total number of observations.

$$Proportion\ correct = \frac{(TP+TA)}{N} \tag{2.15}$$

2.3.3.4 AUC

The Area Under the Curve (AUC) is a threshold independent metric derived from the receiver operating characteristic curve (ROC curve), and it is the most commonly used integrated discrimination metric for SDMs (Guisan et al., 2017). It uses a graph to display all values of sensitivity against the corresponding value of commission rate ($1 - specificity$) across all thresholds between 0-1 (Allouche et al., 2006). This graph is then used to calculate the area under the curve, where a value of 1 means perfect prediction of presence and absence, 0.5 means that the model is no better than 'random', and 0 means that the model is giving a perfect contrary prediction (Guisan et al., 2017). It is suggested that good models typically have AUC values above 0.8 (Araújo et al., 2005).

Although the use of a threshold independent metric enables a general evaluation of model performance without setting an arbitrary threshold, many SDM practical applications require model probabilistic predictions to be transformed into binary predictions (Allouche et al., 2006). In these cases, a threshold should be set in order to evaluate model predictive ability (Allouche et al., 2006).

2.3.3.5 Kappa

Unlike AUC, Cohen's kappa is a metric that is dependent on the selection of a threshold to transform predictions into binary data (Allouche et al., 2006). Kappa is one of the most widely used metrics to evaluate model predictions to presence-absence data, and it measures the accuracy of the model

predictions relative to random predictions (Allouche et al., 2006). When used in a binary contingency table it can be calculated with the formula:

$$Kappa = \frac{(TP+TA) - \frac{(TP+\quad)\cdot(TP+F\quad)+(FP+TA)\cdot(FA+TA)}{N}}{N - \frac{(TP+F\quad)*(TP+F\quad)+(FP+T\quad)*(FA+T\quad)}{N}} \quad (2.16)$$

Kappa range from -1 to +1, where +1 indicates perfect agreement between model predictions and observed occurrence, and any value below 0 suggests a performance no better than random (Allouche et al., 2006). As a rule of thumb, values of 0.75 or higher are considered to be excellent, those between 0.4-0.75 are seen as good, and values below 0.4 indicate poor performance (Guisan et al., 2017). A documented limitation of Cohen's kappa is that it is sensitive to the prevalence of the species, showing a unimodal response to prevalence values (Allouche et al., 2006). This limits Cohen's kappa use unless prevalence is well known across the whole study region, which is often not possible in presence-only datasets (Allouche et al., 2006; Guisan et al., 2017).

2.3.3.6 True Skill Statistic (TSS)

The true skill statistic was proposed as an alternative evaluation metric to Cohen's kappa, since it was shown not to be affected by the species prevalence (Allouche et al., 2006). It measures how well the model separates presences and absences from random predictions and then compares that to a hypothetical set of perfect predictions (Allouche et al., 2006). Like Kappa, TSS requires the selection of a threshold to transform predictions into binary data. It can be simplified when used in a binary contingency table as:

$$TSS = \frac{(TP\cdot TA)-(FP\cdot FA)}{(TP+FA)\cdot(FP+T\quad)} = Sensitivity + Specificity - 1 \quad (2.17)$$

TSS values vary between -1 and +1, and their corresponding interpretation is closely related with Cohen's kappa statistic (Allouche et al., 2006). The simplicity of the formula, its interpretation, and the fact that it is not affected by prevalence, make TSS one of the preferred metrics for measuring predictive performance of presence-absence models (Guisan et al., 2017). As such, it was the main evaluation metric selected for assessing each model's predictive performance in this work.

All evaluation metrics were calculated in the internal algorithm of the *biomod2* package. However, a problem occurred when calculating the accuracy metrics of the single model predictions for the independent census data in the Iberian Peninsula. A workaround was found by manually extracting the model predictions and using the package *SDMTools* (VanDerwal et al., 2014) to calculate the accuracy metrics.

2.3.4 Ensemble

Selecting the correct model to predict a species' distribution is a complex and challenging task because of the large number of available modeling algorithms and optimization parameters (Araújo & New, 2007; Pearson et al., 2006; Guisan et al., 2017). When any of the different models is used it is essential to understand that each of them has limitations in its predictive abilities. No single model is capable of encompassing all the necessary information to fully predict the distribution of a species (Araújo & New, 2007).

A commonly used approach to address this challenge is what is called an ensemble forecast (Araújo & New, 2007; Guisan et al., 2017). As previously mentioned in the section describing the GBM and RF algorithms, an ensemble is a combination of single models that gives a general prediction based on the compilation of their outcomes (Guisan et al., 2017). This compiling can be done in a number of different ways, such as averaging the probabilities of occurrence, selecting the median of the predictions, or through a committee vote where each individual model votes on a presence or absence with the majority determining the outcome (Guisan et al., 2017). As a result, ensemble forecast predictions have shown to increase model predictive performance and also decrease the variance of the results, when compared to single-model approaches (Neto & Araújo, 2011; Guisan et al., 2017). Ensemble modeling also has the advantage of retaining the importance of variables and modeled response curves, while mapping the variance (uncertainty) across all models, allowing to identify areas where there is more disagreement between predictions (Araújo & New, 2007).

Since the ensemble model is created from an aggregation of other models, it is crucial to implement a pre-selection of the most accurate single models (Araújo & New, 2007). This ensures that the ensemble model predictions are as reliable as possible (Araújo & New, 2007). To achieve this, evaluation metrics are calculated for every 10 model replicates from each algorithm, calibrated with 70% of the whole dataset, and a value filter is applied afterwards to prevent models with weak predictive performance from being included in the ensemble (Guisan et al., 2017). For this work only single models, calibrated with 70% of the whole dataset, with a TSS score of 0.7 or higher were integrated into the final ensemble model.

Two outputs were derived from the ensemble modeling procedure for each grid cell: the 'mean ensemble model' representing the continuous probability of occurrence as the simple mean average of selected single model predictions, and the 'variance model', expressed in percentage, calculated from the coefficient of variation (CV) which highlights the uncertainty of the ensemble model.

2.3.5 Future projections

Once a robust ensemble model was created, the final step was to analyze how global change scenarios could potentially impact the distribution of the Red-backed Shrike. This involved projecting the previously calibrated ensemble model onto future bioclimatic and land-use conditions. For this general circulation models (GCMs) are commonly used to simulate potential climatic conditions, and land-use states are derived from predicted economic scenarios (Guisan et al., 2017).

The associated anthropogenic impacts that form the basis of these scenarios, like economic development leading to increased agricultural expansion and deforestation, energy consumption from fossil fuels or

green energy, and economic policies focusing on self-reliance or international trade, are expected to impact habitat suitability of several species (Thuiller et al., 2005). To assess how the bioclimatic and land-use variables will change, two carbon emission scenarios were selected from the Special Report on Emissions Scenarios published by the Intergovernmental Panel on Climate Change's Fourth Assessment Reports (IPCC, 2008). The scenarios chosen are widely used in SDMs, and attempt to represent reasonable optimistic and pessimistic outcomes for possible future environmental conditions:

1. A1B considers a future of rapid economic growth, where the global population peaks mid-century and declines thereafter. It assumes a balance across fossil and non-fossil energy sources, with technological advancements driving the adoption of cleaner energy. This scenario represents a world with moderate emissions, where economic prosperity is balanced with environmental awareness (IPCC, 2008).
2. A2 scenario's main driver is an increased interest in self-reliance and regionalism, with less emphasis on global cooperation. Economic development is slower and more regionally oriented, leading to higher emissions due to the continued reliance on fossil fuels. The slower technological advancements and regional focus could lead to more pronounced environmental degradation (IPCC, 2008).

To predict the values of the environmental variables in the future, two GCMs were fitted to these emission scenarios, the CSIRO-Mk3.0 (Gordon et al., 2002) henceforth referred as CS, and the MIROC-H (K-1 model developers, 2004) referred as MR. Both models simulate future climatic conditions based on the carbon emission scenarios, and are widely used in projecting model predictions into future climate conditions due to their relatively good predictions (Thuiller et al., 2005; Perkins et al., 2007). All of the climatic variables projections from these models were available and obtained from the CliMond dataset (Kriticos et al., 2012). Furthermore, land-use reorganizations under economic scenarios of the IPCC report were obtained from the IMAGE project (IMAGE contributors, 2017).

When projecting a distribution into future conditions an assumption needs to be made on how likely the study species is to disperse in response to changes in habitat suitability (Elith et al., 2010). For this, two dispersion metrics were used to evaluate different assumptions on how the species would possibly respond to global change. 'No dispersal' range shifts assumed that the species had limited mobility or high fidelity to breeding grounds and therefore measured how much of the predicted range was preserved from the current prediction to future predictions, while 'full dispersal' represented high mobility and ecological plasticity to colonize new suitable areas previously unsuitable (Guisan et al., 2017).

3 Results

3.1 Variable selection

Initial VIF analysis showed elevated multicollinearity between temperature variables, with ‘tmin’ and ‘tmean’ showing very high VIF scores (Table 3.1). The stepwise VIF elimination performed by the function *vifstep* resulted in the selection of the bioclimatic variables ‘tmax’, ‘tsd’, ‘prec’ and ‘precsd’, all showing VIF scores lower than 5 (maximum score 2.3 by ‘precsd’). However, ‘tmax’ and ‘precsd’ showed relatively high correlation (Figure 7.1). Though this is below the recommended 0.7 threshold to remove a variable, it was still high enough that a more in-depth verification was performed (Guisan et al., 2017).

Table 3.1: VIF scores of all explanatory covariates considered. Values above 10 were considered to introduce high amounts of multicollinearity and were removed.

Variable	tmin	tmax	tmean	tsd	prec	precsd	bias
VIF	134	56	214	36	1	3	1

Model variable importance output showed ‘precsd’ as the variable with the least influence for all algorithms except RF (Table 7.3). Additionally, binary ensemble predictions with and without ‘precsd’ in the set of covariates showed similar predictions between ensemble forecasts (Figure 7.4). Given that the variable ‘tmax’ was the variable with the highest importance in the ensemble model (Table 3.4), ‘precsd’ high correlation probably introduced more noise rather than adding relevant information into the model. As such, a decision was made to remove ‘precsd’ from the set of explanatory variables, leaving ‘tmax’, ‘tsd’, ‘prec’, ‘bias’, and ‘land’ as the final covariates to model the distribution of the Red-backed Shrike in this work.

3.2 Sampling design comparison

From the different sampling areas considered, the custom design based on the migration route yielded the best evaluation metric scores (Table 3.2). All sampling designs showed very similar results, with the migratory based pseudo-absence design showing the highest score in every performance metric. As such, it was the sampling design used in this analysis henceforth.

Table 3.2: Evaluation metrics of ensemble models from different sampling area designs to map pseudo-absences.

Design	Thresh	AUC	Omission	Sensitivity	Specificity	Prop correct	Kappa	TSS
Custom	0.70	0.918	0.061	0.939	0.896	0.912	0.814	0.835
Poly 5	0.69	0.913	0.071	0.929	0.896	0.908	0.806	0.825
Poly 10	0.69	0.897	0.101	0.899	0.896	0.897	0.781	0.795
Squared	0.68	0.909	0.061	0.939	0.879	0.901	0.792	0.818

3.3 Removal of the Iberian Peninsula from calibration

Models fitted without the Iberian Peninsula in the calibration dataset showed overall low predictive accuracy. The scores of the True Skill Statistic (TSS; evaluation metric chosen in this thesis to assess predictive accuracy performance) were below the pre-established value threshold of 0.7 (Table 3.3). Consequently, no models were included in the final ensemble model. To create an ensemble model, the value threshold was reduced to 0.4, which allowed for the inclusion of a selection of models. The resulting model had a TSS score of 0.578, a value far lower than TSS scores when the Iberian Peninsula was included in the calibration dataset (Table 3.4). As such, model analysis from this point on were performed with the Iberian Peninsula included in the calibration dataset.

Table 3.3: Evaluation metrics of every modeling algorithm calibrated with 100% of the data ('full models') and ensemble model without Iberian Peninsula data in the calibration dataset.

Model	Run	Thresh	AUC	Omission	Sensitivity	Specificity	Prop correct	Kappa	TSS
Maxent	full	0.04	0.767	0.212	0.788	0.746	0.761	0.508	0.534
GLM	full	0.91	0.718	0.131	0.869	0.566	0.676	0.384	0.435
GBM	full	0.90	0.638	0.313	0.687	0.590	0.625	0.254	0.276
RF	full	0.53	0.755	0.091	0.909	0.601	0.713	0.452	0.510
Ensemble	TSS>=0.4	0.58	0.789	0.121	0.879	0.699	0.765	0.532	0.578

3.4 Modeling algorithms comparison

3.4.1 Full models

All models calibrated with 100% of the species' occurrence data ('full models') successfully predicted the existence of suitable habitat in the western, central and eastern areas of the Red-backed Shrike's known range (Figure 3.1). A connection between the Central European and the Eastern Asian breeding area in the limit of the range is visible in every map, although the extent and proportion of suitable habitat varied significantly between models. The central part of the range in Europe is unanimously considered highly suitable while edge areas exhibit substantial variation between predictions. Additionally, novel areas where the species is not known to breed, such as India and Northern Africa, also show large divergences between the model predictions.

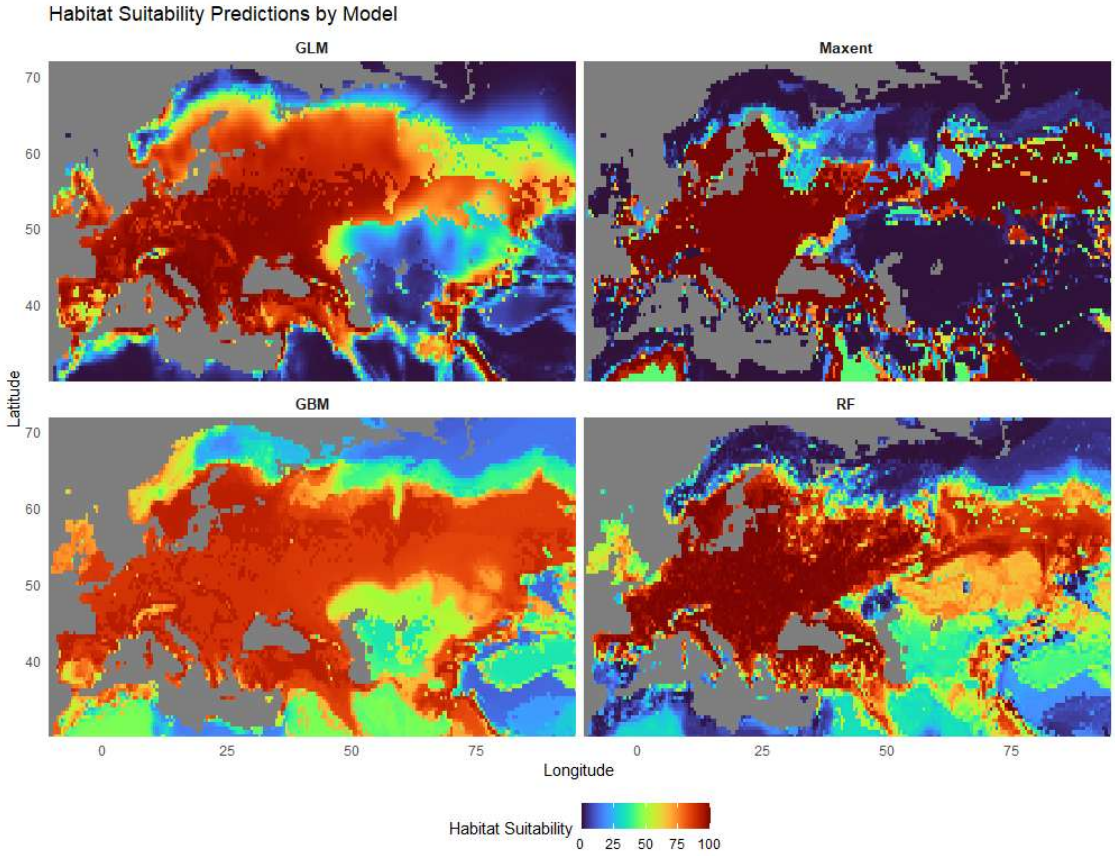

Figure 3.1: Map showing all algorithm full model's continuous probability (habitat suitability) projection for the whole range. Each 0.5x0.5 º grid cell has a value ranging from 0-100 representing probability as percentage, with lower values showing colder (blue) colors in the map and higher values showing warmer (red) colors. Dark gray color represents water and doesn't have any values associated.

MaxEnt produced the most conservative predictions among all models, seeming to restrict the highly suitable locations for the Red-backed Shrike mostly to central Europe and very scarcely towards edge areas of its known range, like the Iberian Peninsula. Habitat suitability also showed huge discrepancies between unsuitable and suitable areas, going from very low values (near 0) to high values (near 100) in most of the breeding range. On the contrary, GBM presented the most optimistic projection, predicting high habitat suitability not only in edge areas but also in novel locations. A gradient-like pattern could be observed in the habitat suitability predictions, where values change in "steps". GLM and RF's projections restricted high habitat suitability mostly to the species' known breeding range, with differences being more noticeable in areas near the edge of the breeding range.

Full model performance in predicting the distribution in the Iberian Peninsula varied greatly. GLM and MaxEnt demonstrated considerably low predictive ability with TSS values of 0.346 and 0.495 (Table 3.4), respectively. In contrast, GBM and RF exhibited strong predictive performance, with TSS values of 0.744 and 0.811, respectively. AUC and Kappa, which were calculated to serve as control metrics of TSS, showed a similar pattern for all models. GLM and GBM had high threshold values (0.89 and 0.86, respectively; Table 3.4) indicating a relatively high tendency for overprediction. MaxEnt's restricted predictions were reflected in its low threshold value (0.33). RF showed a relatively balanced prediction with the threshold value (0.60) being close to the established prevalence value of 0.5. Since the selected threshold was the one that maximizes sensitivity + specificity, all models had a good balance between sensitivity and specificity. However, GLM and MaxEnt lower TSS values are reflected in their low sensitivity (0.687 and 0.778) and specificity (0.659 and 0.717). Omission rates in these models were also high (0.313 and 0.222) leading to the removal of large portions of true presences from their predictions. In contrast to that, GBM and RF had high values of sensitivity (0.929 and 0.909) and specificity (0.815 and 0.902), with considerably low omission rates (0.071 and 0.091). These values indicate that GBM tended to overpredict given the highest sensitivity and comparatively low specificity, while RF had the best overall prediction with the highest specificity (0.902) and proportion of correct predictions (0.904).

Table 3.4: Evaluation metrics for each modeling algorithm calibrated with the whole dataset ('full model') and the ensemble model created from 'single model' predictions with TSS values above 0.7.

Model	Run	Threshold	AUC	Omission rate	Sensitivity	Specificity	Proportion correct	Kappa	TSS
GLM	full	0.89	0.673	0.313	0.687	0.659	0.669	0.326	0.346
MaxEnt	full	0.33	0.747	0.222	0.778	0.717	0.739	0.467	0.495
GBM	full	0.86	0.872	0.071	0.929	0.815	0.857	0.706	0.744
RF	full	0.60	0.905	0.091	0.909	0.902	0.904	0.797	0.811
Ensemble	TSS>=0.7	0.70	0.918	0.061	0.939	0.896	0.912	0.814	0.835

Predictions of the distributions of the Red-backed shrike showed varied ranges between algorithms (Figure 3.2). GLM and MaxEnt showed erroneous predictions to the southern region, with GLM even failing to accurately represent several presence points in the northern coast of Spain. In contrast, GBM

and RF predictions were very accurate. The previously acknowledged overprediction of GBM was shown with an expansion of the distribution range to novel areas in the South, while RF showed a very accurate prediction compared with the known distribution of the Red-backed Shrike.

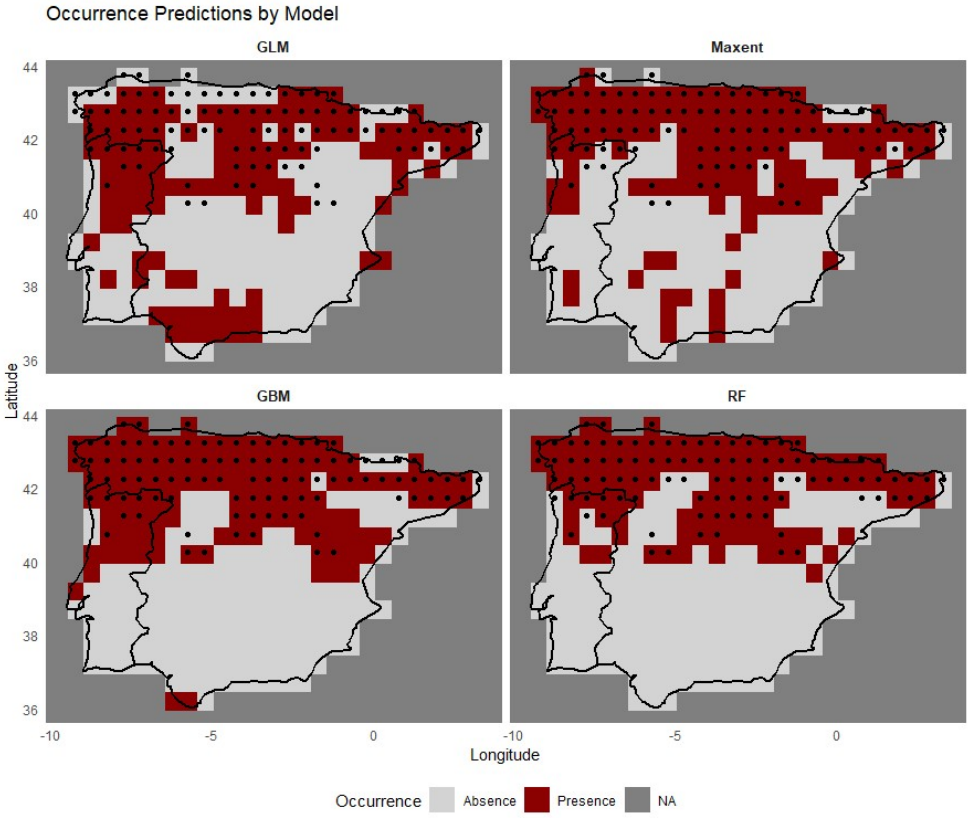

Figure 3.2: Map of each modelling algorithm 'full model' binary occurrence projection under future conditions for the Iberian Peninsula. Each 0.5x0.5 º grid cell has a value of either 0 (absence) or 1 (presence), with absences showing light gray in the map and presences showing dark red colors. Black dots represent presence points from the census data. Dark gray color represents water and does not have any values associated.

3.4.2 Single models

Table 3.5: TSS, sensitivity and specificity values for modeling algorithm replicates calibrated with 70% of the whole dataset ('single models'). Various values were calculated to highlight differences between prediction results.

Model	TSS				Sensitivity				Specificity			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
MaxEnt	0.472	0.047	0.386	0.533	0.755	0.077	0.636	0.889	0.718	0.068	0.555	0.798
GLM	0.391	0.077	0.284	0.476	0.750	0.106	0.636	0.879	0.642	0.040	0.590	0.711
GBM	0.706	0.054	0.612	0.762	0.891	0.063	0.788	0.960	0.814	0.030	0.763	0.867
RF	0.780	0.026	0.736	0.811	0.894	0.018	0.869	0.929	0.886	0.014	0.867	0.902

Replicates of each algorithm calibrated with only 70% of the data ('single models') showed a similar pattern to the 'full models' with two noticeably different groups in predictive accuracy: lower performances in linear regression (GLM and MaxEnt) and high accuracy in decision trees approaches (GBM and RF).

RF demonstrated the least variation between iterations for both sensitivity (sd = 0.018) and specificity (sd = 0.014), while also achieving the highest score of specificity (max = 0.902), and the best overall result in sensitivity (mean= 0.894). This high performance made RF have the highest rank in modeling accuracy performance from all the chosen algorithms (Figure 3.3). On the other hand, GBM scored higher in maximum sensitivity (max = 0.960; Table 3.5), at the cost of trading off lower values in specificity (mean = 0.814). These results gave GBM the highest rank scores in sensitivity at the cost of specificity (Figure 3.3).

In contrast, the maximum values of sensitivity and specificity of neither GLM nor MaxEnt were able to reach even the minimum values achieved by GBM or RF, highlighting a big discrepancy in performance between the decision tree algorithms and linear regression (Table 3.5). Moreover, MaxEnt tended to have higher ranks for specificity, while GLM was stronger in sensitivity (Figure 3.3). For the true skill statistic (TSS) values, RF showed to be more consistent in the predictions (sd = 0.026), while GLM had the highest variation across model runs (sd= 0.077). MaxEnt and GBM had similar intermediate values of standard deviation (0.047 and 0.054, respectively). The highest score of TSS in all value categories was achieved by RF, followed by GBM, MaxEnt and, finally, GLM (Table 3.5).

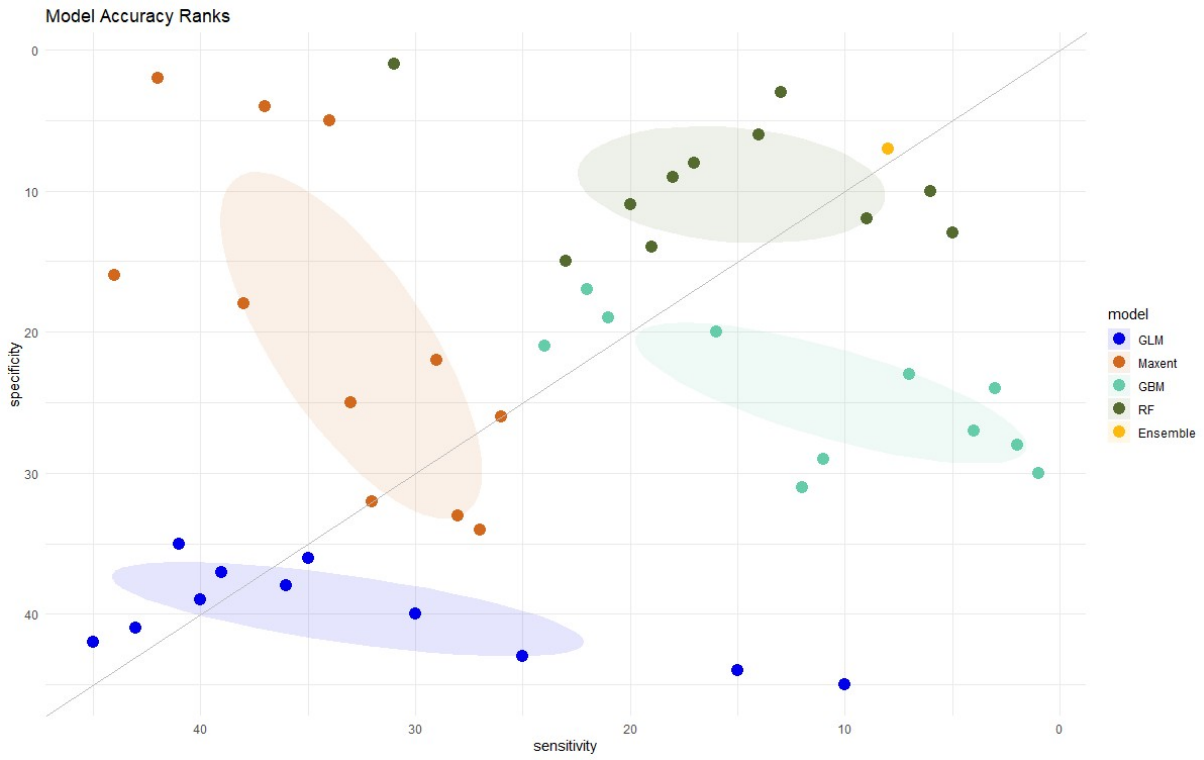

Figure 3.3: Graphical representation of model ranks for sensitivity and specificity. Each algorithm has an ellipse that represents the cluster of the associated ranks. Axis ranges are reversed so lower values (higher ranks) are represented to the right (sensitivity) and top (specificity).

The maximum TSS values of GLM (0.476) and MaxEnt (0.533) were below the pre-established value of 0.7, meaning no model from these algorithms was selected for the ensemble model. In contrast, the boosting and bagging algorithms (GBM and RF) were able to produce predictions accurate enough (TSS>=0.7) to be incorporated into the final ensemble model. Following this, from the 40 model replicates computed a total of 17 model replicates were selected, composed of 7 GBMs and 10 RFs.

3.5 Ecological drivers

The explanatory variables had considerably different impacts on the models. The 'bias' variable, representing birding effort, had the highest contribution for all modeling algorithms, ranging from 0.277 in GLM to 0.586 in GBM (Table 3.6). For the ensemble mean and variance models, 'bias' also had the strongest influence (0.551 and 0.460, respectively). These results highlight the pervasive impact of sampling bias correction in the predictions.

Table 3.6: Influence of each environmental covariate for each corresponding model, measured in variable importance. Higher values of variable importance correspond to higher influence of that covariate in the associated model.

Model	Run	tmax	tsd	prec	bias	land
GLM	full	0.166	0.049	0.133	0.277	0.048
MaxEnt	full	0.351	0.192	0.002	0.384	0.103
GBM	full	0.123	0.044	0.022	0.586	0.038
RF	full	0.223	0.155	0.076	0.573	0.078
Ensemble	mean	0.167	0.092	0.039	0.551	0.049
Ensemble	cv	0.300	0.284	0.274	0.460	0.225

Among the bioclimatic variables, maximum temperature ('tmax') was the most influential covariate, with contributions from 0.123 in GBM to 0.351 in MaxEnt. For the ensemble mean model, 'tmax' contributed 0.167, making it the most significant environmental variable in that context. Highest suitability for the Red-backed Shrike was generally associated with maximum temperatures ranging between 20ºC and 30ºC, with high variation between model algorithms at higher and lower temperatures (Figure 3.4). As a result, the ensemble model also showed a suitability spike at around 20ºC, slowly decreasing as temperatures increased (Figure 7.4). Notably, MaxEnt showed very high variation of maximum temperature influence at higher values of this covariate.

In comparison, temperature seasonality ('tsd') had moderate influence in most models, with MaxEnt showing the highest importance (0.192) and GBM the lowest (0.044), while the ensemble mean model showed an intermediate contribution of 0.092. Areas with seasonal temperature variation between 4ºC and 12ºC were consistently predicted as highly suitable by all models, including the ensemble (Figure

3.4). Yet again MaxEnt showed an odd pattern, where suitability had abrupt decreases of between very similar values of ‘tsd’, proceeded by sharp increases of suitability shortly after.

The last bioclimatic variable, annual precipitation (‘prec’), had relatively low influence across all models, with the highest contribution observed in GLM (0.133). In contrast, MaxEnt predictions showed negligible influence from annual precipitation (0.002). For the ensemble mean model, ‘prec’ contributed only 0.039. Suitability generally increased with precipitation intensity, plateauing above 500 mm for most models (Figure 3.4).

Land-use (‘land’) exhibited some variation of importance across models, with MaxEnt attributing the highest contribution (0.103), and GBM giving it the lowest importance (0.038). The ensemble mean model assigned a contribution of 0.049 to land-use. Despite these differences, all models generally predicted high suitability across most land-use types. However, ‘shrub’ consistently showed lower suitability compared to other categories, while ‘other’ also contributed relatively less suitability in some models (Figure 3.4).

In the ensemble variance predictions, all environmental variables contributed similarly, with values ranging between 0.225 (‘land’) and 0.300 (‘tmax’). This suggests that while these variables showed limited individual influence on mean predictions, their variability contributed notably to the overall uncertainty of ensemble model predictions.

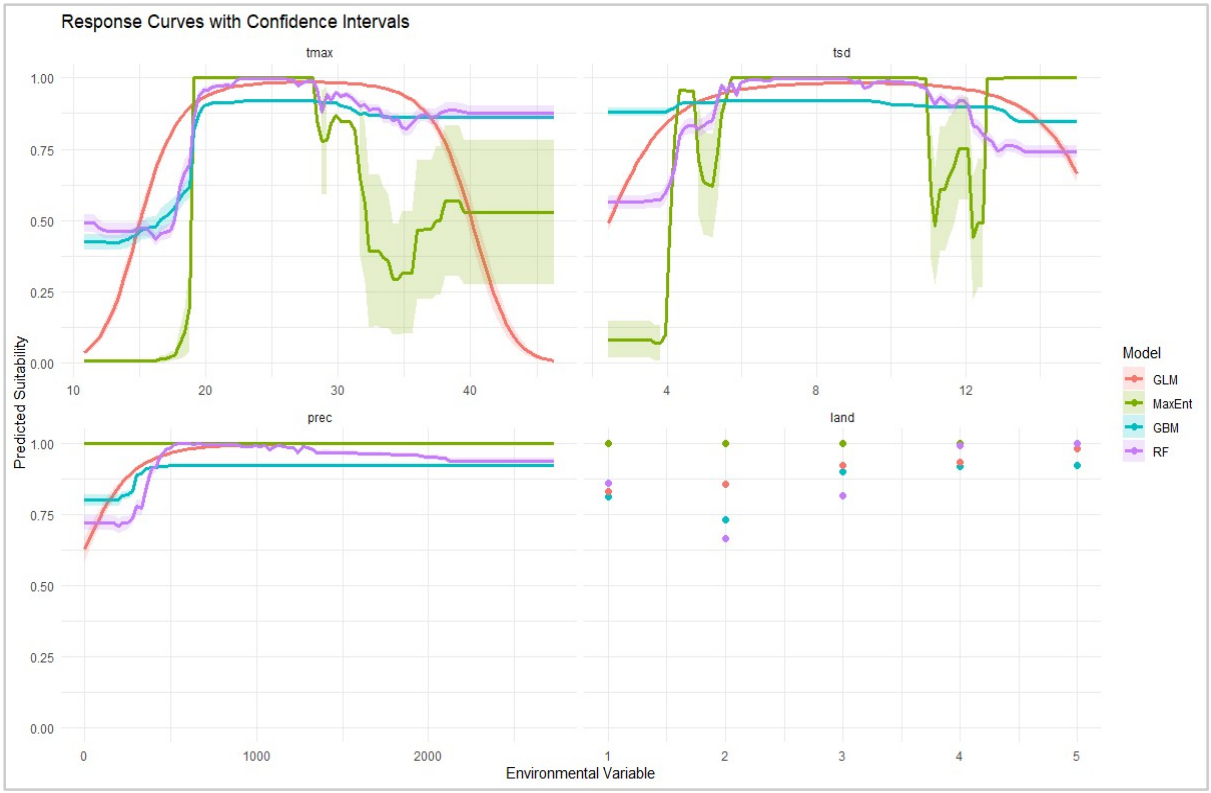

Figure 3.4: Response curves of the likelihood of occurrence of the Red-backed Shrike for each environmental predictor. Each graph illustrates how the probability of occurrence (values between 0-1) varies across different values of each covariate (top left: maximum temperature (‘tmax’), top right: temperature seasonality (‘tsd’), bottom left: annual precipitation (‘prec’) and bottom right: land-use (‘land)) for the replicates of different modeling algorithms. Colors correspond to a specific model, and 95% confidence intervals are drawn around the lines to show variation of each modeling algorithm.

3.6 Ensemble model

The ensemble model outperformed the individual model predictions (Table 3.4), achieving the highest TSS value among all models (0.835), as well as in control metrics AUC (0.918) and Kappa (0.814). Although some GBM models had higher sensitivity and some RF models exhibited higher specificity, the ensemble model had very high and similar values for both sensitivity and specificity (0.939 and 0.896, respectively), establishing a high rank in both metrics (Figure 3.3). This along with the highest proportion of correctly predicted occurrences (0.912), indicated an accurate and balanced prediction.

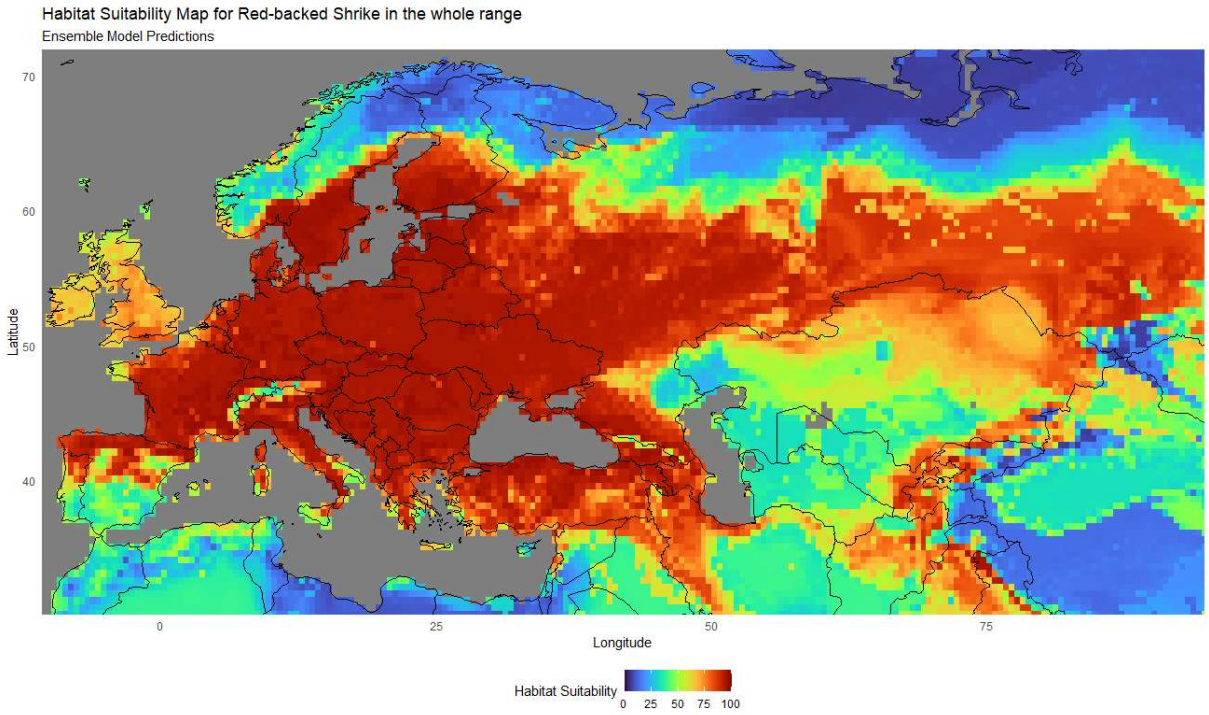

Figure 3.5: Map showing ensemble model's continuous probability (habitat suitability) projection for the whole range. Each 0.5x0.5 º grid cell has a value ranging from 0-100 representing habitat suitability as percentage, with lower values showing colder colors in the map and higher values showing warmer colors. Dark gray color represents water and does not have any values associated.

Ensemble predictions attributed high habitat suitability mainly to locations within the known breeding distribution of the Red-backed Shrike (Figure 3.5). Habitat suitability gradually decreased as distance from the known breeding range increased.

The ensemble model had very small uncertainty where most of the presence data converged, mainly in Central, Eastern and Western Europe (Figure 7.5). A large amount of uncertainty was found around edge areas of the breeding range, namely in Scandinavia, small pockets around the Caspian Sea and Altai mountains, and more importantly in the Iberian Peninsula. Areas outside of the known breeding range varied in disagreement between models, from low disagreement in arid areas in Central Asia, to high variance when projecting to the tundra near the Artic pole and southern half of the Mediterranean basin.

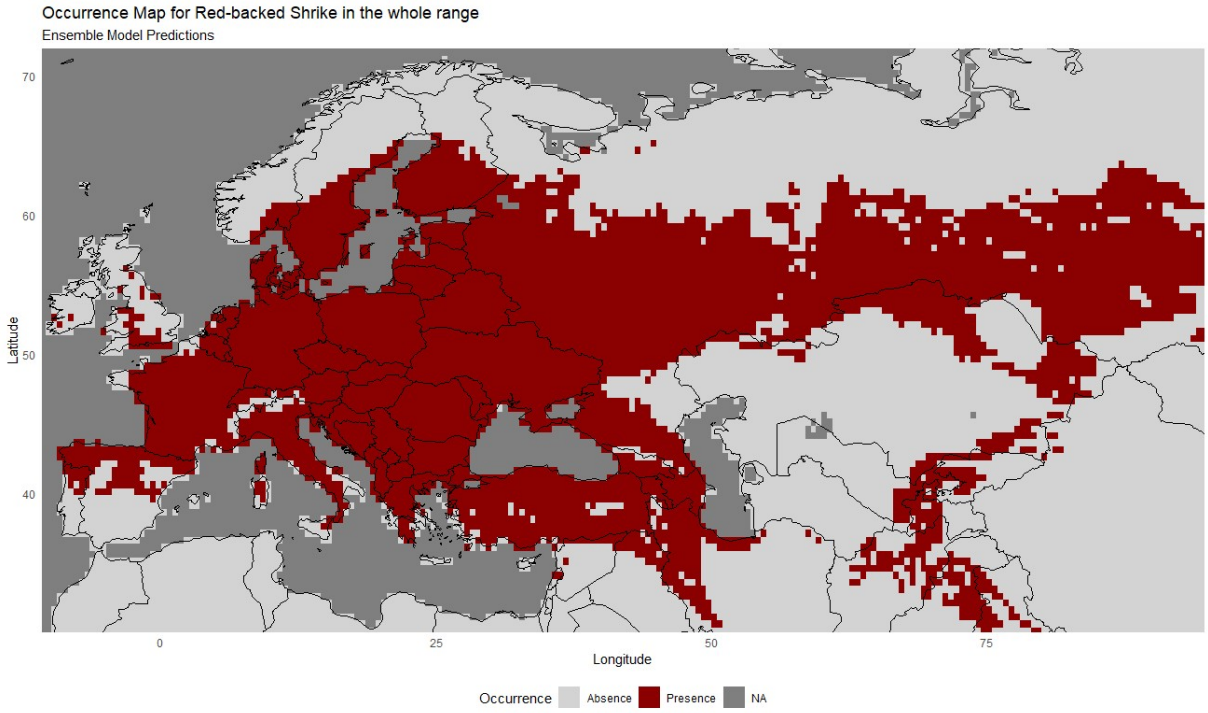

Figure 3.6: Map of ensemble model's binary occurrence prediction for the whole range, using the threshold (0.7) that maximizes TSS for the census data in the Iberian Peninsula. Each 0.5x0.5 º grid cell has a value of either 0 (absence) or 1 (presence), with absences showing light gray in the map and presences showing dark red colors. Dark gray color represents water and doesn't have any values associated.

After applying a binary transformation using the threshold that maximizes sensitivity + specificity in the Iberian Peninsula (0.7, in this study), the ensemble model's binary occurrence projection still provided a relatively accurate representation of the known breeding distribution across the entire range of the Red-backed Shrike (Figure 3.6). A connection can be seen between European breeding grounds and Asian breeding grounds, with no gaps within the known breeding range.

Similar to the whole range, the ensemble model predictions in the Iberian Peninsula estimated high suitable areas mostly to locations within the known range of the Red-backed Shrike (Figure 3.7). Habitat suitability decreased in southern areas associated with drier climate, restricting higher likelihood of occurrence to locations in mountainous regions.

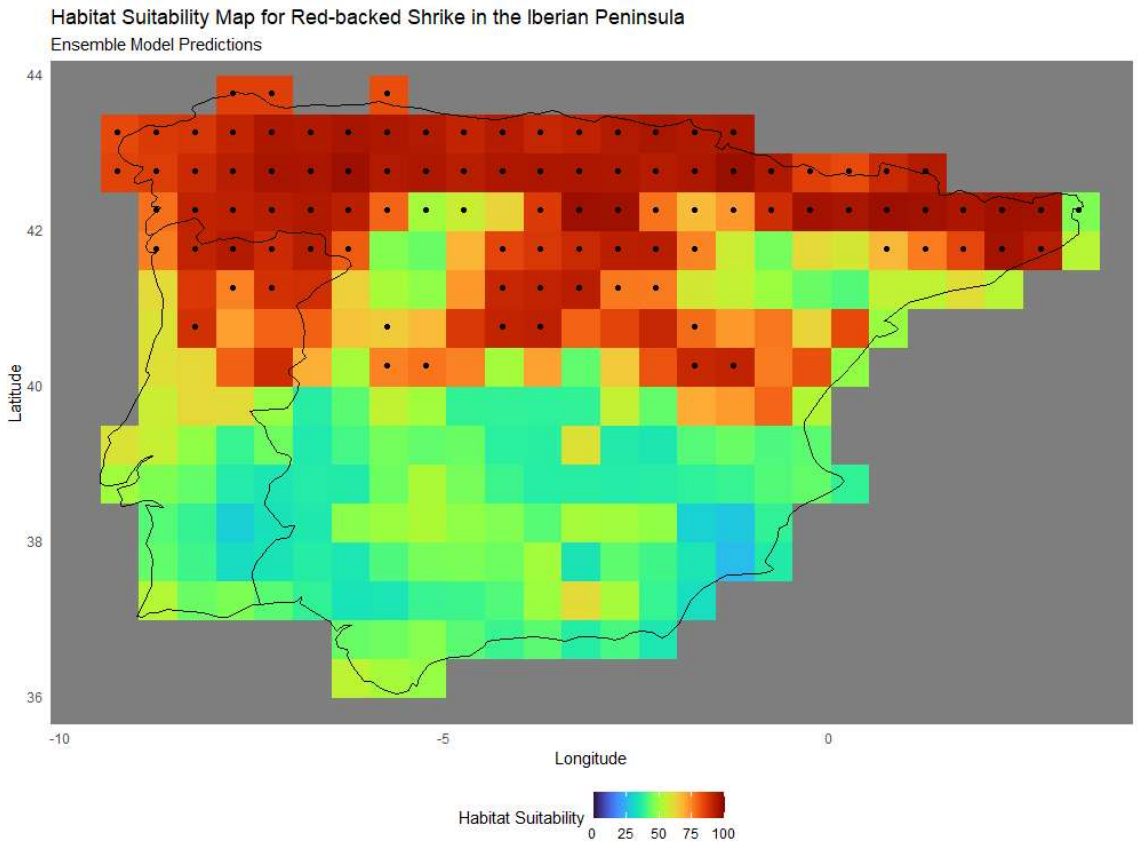

Figure 3.7: Map showing ensemble model's continuous probability (habitat suitability) projection for the Iberian Peninsula. Each 0.5x0.5 º grid cell has a value ranging from 0-100 representing a percentage, with lower values showing colder colors in the map and higher values showing warmer colors. Black dots represent presence points in the independent census data. Gray color represents water and doesn't have any values associated.

3.7 Global Change Projections

3.7.1 Probability predictions

Global change altered habitat suitability in the edge of the Red-backed Shrike's breeding range, especially along the northern limits and in the Iberian Peninsula (Figure 3.8). The predicted bioclimatic and land changes are expected to increase the likelihood of new highly suitable areas in northern Europe and Asia, while territories in the southern limit of the breeding showed an overall decrease in suitability.

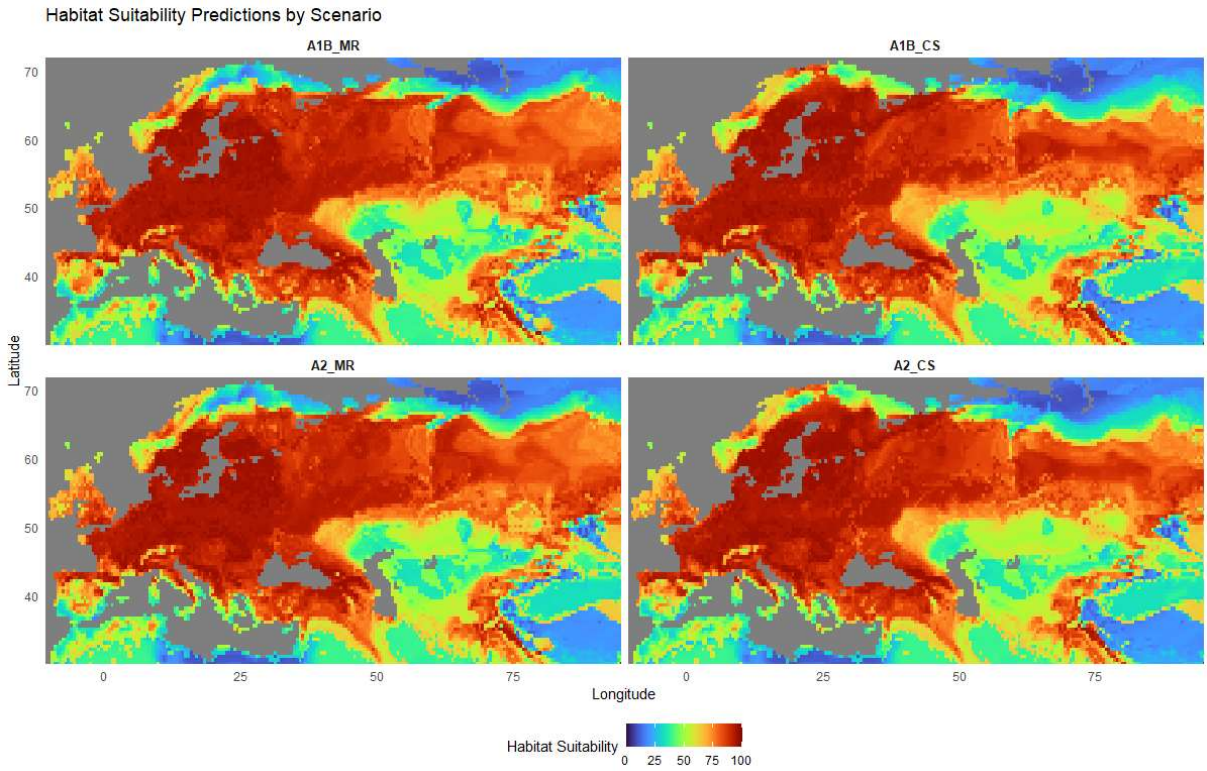

Figure 3.8: Map showing all ensemble model's projections into future scenarios continuous probability (habitat suitability) projection for the whole range. Each 0.5x0.5 º grid cell has a value ranging from 0-100 representing probability as percentage, with lower values showing colder colors in the map and higher values showing warmer colors. Dark gray color represents water and doesn't have any values associated.

Future projections overall uncertainty also seemed to be much lower than in current conditions in the study area, with edge areas of the breeding range still showing higher CV values than areas in the center of the breeding range (Figure 7.6). The Iberian Peninsula continued to show relatively high levels of disagreement between model predictions in the southern region, albeit to a much lesser extent than in the current conditions' projection.

3.7.2 Range shifts

The ensemble model mostly considers that under future global change scenarios environmental conditions will push the breeding range of the Red-backed Shrike further north, while reducing its southern breeding range (Figure 3.9).

Table 3.7: Predicted grid cell count changes under the environmental conditions of each associated future scenario.

Scenario	Loss	Stable		Gain	Percentage Change			Range Size		
		Absence	Presence		Loss	Gain	Total	Current	Future	
									No Dispersal	Full Dispersal
A1B_MR	426	12,560	5,283	1,991	7.46%	34.89%	27.41%	5,709	5,283	7,274
A1B_CS	525	12,899	5,184	1,652	9.20%	28.94%	19.74%	5,709	5,184	6,836
A2_MR	387	12,615	5,322	1,936	6.79%	33.91%	27.13%	5,709	5,322	7,258
A2_CS	490	12,934	5,219	1,617	8.58%	28.32%	19.74%	5,709	5,219	6,836

Range shifts varied more between Global Climate Models (MR and CS) than Carbon Emission (A1B and A2) future scenarios (Table 3.7). For the lost area, scenarios predicted by MR model showed lower range contractions than CS. Furthermore, the A2 economic scenario also showed lower range losses than A1. Nevertheless, range expansion surpassed predicted contractions, with territory gains of up to 1991 grid cells in A1B_MR. These differences led to positive range shifts from 19.74% (A1B_CS and A2_CS) to 27.41% (A1B_MR). Most of the current range was preserved in future scenarios, with the lowest future range size assuming no dispersal of 5184 for A1B_CS being very close to the current range size of 5709 grid cells. Assuming full dispersal capabilities, the future range size increases greatly between 6836 to 7274 grid cells, with noticeable differences between predictions under MR and CS scenarios.

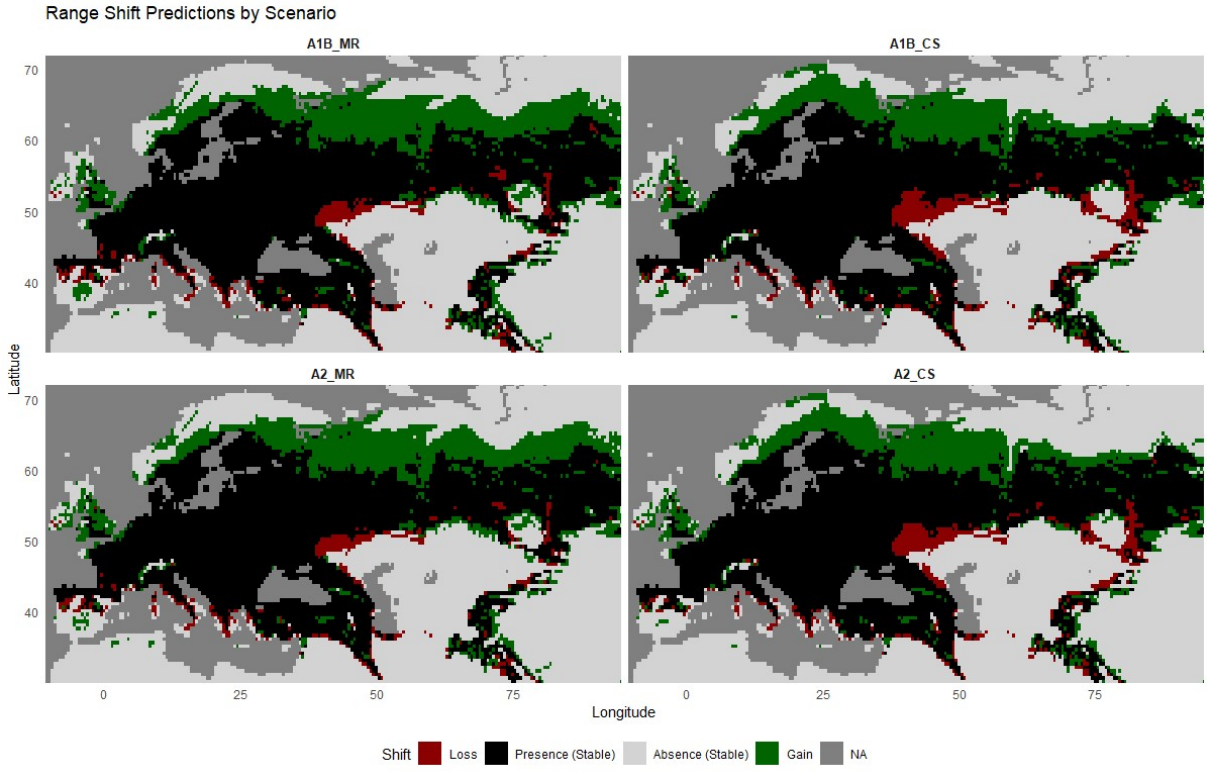

Figure 3.9: Map of range shift projections under future conditions. Each 0.5x0.5 º grid cell has a value of either -2 (loss), -1 (presence), 0 (absence) or 1 (gain). Loss and gain represent the shifts in range due to predicted changes in habitat conditions, while stable presences and absences are projections that remain unaltered through time. Dark gray color represents water and doesn't have any values associated.

Following the McNemar's test, model binary prediction differences for the whole range showed higher differences between global circulation models than between emission scenarios (Figure 7.4). The p-values for the comparison between MR and CS in the A1B and A2 carbon emission scenarios (medians 0.06 and 0.04, respectively; Table 7.4) were much lower than those for the comparison of the MR and CS predictions between emission scenarios (medians 0.72 and 0.68, respectively).

Meanwhile, model binary predictions for the Iberian Peninsula showed little difference between predictions. Unlike the predictions for the whole range, the inclusion of 'precsd' to the set of covariates lead to a higher divergence between ensemble models under current conditions for the Iberian Peninsula (median p-value 0.13; Table 7.6). The McNemar' test p-values for comparisons between GCMs and carbon emission scenarios were relatively high, with the differences between MR and CS for the A1B scenario showing the overall lowest values (median p-value 0.69; Table 7.5). However, a visible difference between GCMs projections was evident in the predicted range shifts for the target area (Figure 3.10). CSIRO (CS) showed little change from the predicted distribution under current conditions in both emission scenarios, with minor losses in the border of the Northern range and a small novel location in the Southern mountains. Meanwhile, MIROC (MR) predicted a greater contraction in the northern territories and a big increase in the novel southern territories.

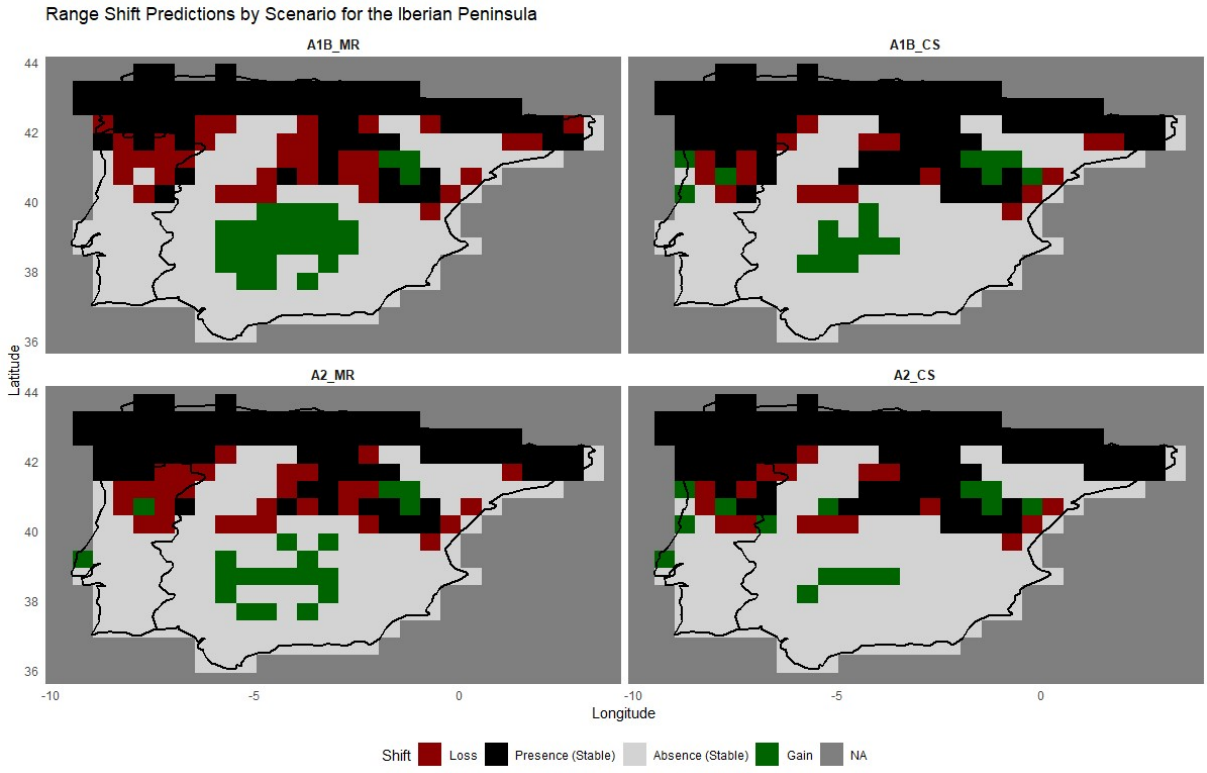

Figure 3.10: Map of range shift projections under future conditions for the Iberian Peninsula. Each 0.5x0.5 º grid cell has a value of either -2 (loss), -1 (presence), 0 (absence) or 1 (gain). Loss and gain represent the shifts in range due to predicted changes in habitat conditions, while stable presences and absences are projections that remain unaltered through time. Dark gray color represents water and does not have any values associated.

4 Discussion

The main objective of this study was to assess different SDM algorithms, while aiming to understand the driving factors of the Red-backed Shrike breeding distribution, as well as the effects global change could have on this species. By evaluating each algorithm performance and aggregating models into an ensemble, distribution projections were created for the whole Eurasian breeding range and more locally in the Iberian Peninsula.

This work highlighted the environmental factors that shape the distribution of the Red-backed Shrike. The results indicate that SDMs can provide viable insights into understanding the distribution of the Red-backed Shrike, giving a reliable and robust projection that can be used in combination with future scenarios (e.g., climate change) to predict the range shift caused by global change. However, the accuracy of such claims are dependent on factors such as bias in the data, choice of explanatory variables, as well as the spatial resolution of the data (Guisan et al., 2017). Furthermore, the wide range of algorithms available for SDMs increase variance in results (Araújo & New, 2007). For all of these reasons, interpretation of distribution predictions and range shifts should be done with caution, especially in areas with higher levels of uncertainty, such as locations in the margins of the distribution (Araújo et al., 2005).

4.1 Comparison of modeling approaches

GLMs had the poorest results of all algorithms, with the lowest predictive accuracy values (mean TSS = 0.391; Table 3.5) and lowest overall ranks in specificity and sensitivity (Figure 3.3). These results could stem from the fact that GLM assumed that the occurrence of the Red-backed Shrike followed, at most, unimodal responses to environmental gradients. Although this is relatively common in ecological systems (Albert et al., 2010), if the predictors had a more complex relationship to the occurrence of the Red-backed Shrike, then GLMs probably failed to capture this.

Furthermore, the exclusion of interactions between predictors could also have limited information captured by these models. In general, models capable of fitting complex functions and interactions between covariate tend to obtain higher predictive performance in SDMs (Valavi et al., 2022). As such, incorporation of higher degree polynomial terms and interactions between environmental data could potentially improve GLM's distribution predictions, at the expense of complicating interpretation of the results. Nevertheless, GLMs can provide a useful baseline in SDMs for comparison with other modeling algorithms.

MaxEnt's results also showed weak predictive performance (mean TSS = 0.472; Table 3.5), and especially low ranks in sensitivity (Figure 3.3). This further aligns with the notion that MaxEnt tended to underpredict habitat suitability in the Iberian Peninsula. With all the available feature transformations as well as a tunable regularization parameter, further exploration of these combinations might yield better results (Valavi et al., 2022).

The ‘default’ setting in the *biomod2* package automatically selects all available features to model the distribution of the Red-backed Shrike. This could have led to a known issue with MaxEnt, where the more features available in the MaxEnt algorithm, the more the model tends to overfit the data (Elith et al., 2011). Overfit is further evidenced by patterns in the response curves (Figure 3.4), where small increments of ‘tmax’ and ‘tmin’ led to abrupt changes in suitability, in a way that is not consistent with expected ecological patterns. Therefore, selecting fewer ecologically relevant feature transformations and changing the regularization multiplier might have been more beneficial. There were available methods (e.g., using R package *ENMeval* to fine tune MaxEnt settings) to select a more accurate combination though these are highly computationally demanding, and in order to maintain consistency between the different models to ensure comparability, in this analysis the settings were set to ‘default’ in *biomod2* options.

MaxEnt also outputs graphs that allow users to diagnose potential problems in the data (Phillips & Dudík, 2008). One of these graphs, the Omission and Predicted areas (Figure 7.2), was supposed to closely follow the 1-1 diagonal line. However, the pattern shown was consistent with insufficient independence between calibration and testing data (Elith et al., 2011), reducing the model’s reliability when projecting to new conditions in future scenarios. As such, exploration of sampling design could also prove crucial to MaxEnt’s performance.

The GBM algorithm performed better than GLM and MaxEnt, achieving relatively high predictive accuracy (mean TSS = 0.706; Table 3.5). However, its high sensitivity (mean 0.891) and comparatively lower specificity (mean 0.814) suggested a tendency to overpredict habitat suitability. This overprediction was further evidenced with a high threshold value (0.86) needed to differentiate presences and absences, which exceeded the default prevalence value of 0.5. This high threshold value suggests that GBM tended to overestimate habitat suitability, potentially inflating predictions in locations with marginal habitats (Allouche et al., 2006).

Other potential problems were identified through graphical diagnostics, especially by examining the error rate vs. number of trees graph (Figure 7.3). This graph showed that GBM was still in the process of improving predictive performance when it reached the maximum number of trees (2500) set by the *biomod2* package. Therefore, giving a higher number of trees or increasing learning rate between iterations could have increased GBM’s ability to accurately predict the distribution of the Red-backed Shrike (Hastie et al., 2009). Like MaxEnt, the use of ‘default’ modeling setting in the *biomod2* package likely limited this algorithm’s performance, since machine-learning algorithms frequently require fine tuning to achieve optimal performances (Valavi et al., 2022). Nevertheless, GBM demonstrated strong predictive power, aligning with previously established utility in SDMs (Elith et al., 2006; Valavi et al., 2022).

RF consistently showed the most accurate predictive performance, with the highest value of TSS for any given single model algorithm (mean TSS = 0.780; Table 3.5). A threshold value of 0.60 for the full model (Table 3.4), which is close to the default prevalence of 0.5, suggests that the modeling algorithm did not overpredict or underpredict habitat suitability for the Iberian Peninsula (Hanberry & He, 2013).

Additionally, RF scored very high ranks both in sensitivity and specificity (Figure 3.3) indicating a balanced prediction of presences and absences, reducing the risk of overestimating or overlooking locations with suitable habitat. Moreover, the standard deviation of the single model predictions TSS was very small (0.026), showing that RF was a very consistent modeling algorithm capable of capturing similar environmental information with different sets of data.

An advantage of RF is the ability to produce very accurate predictions with very little tuning of parameters, which likely contributed to its superior performance when using the 'default' settings from *biomod2* (Hastie et al., 2009). In the end, as shown in the results of this work, RF's high predictive accuracy, balanced sensitivity and specificity, and consistency reinforced its positions as a powerful tool in SDMs (Elith et al., 2006; Valavi et al., 2022).

The ensemble model, composed from the simple mean of seven GBM and ten RF predictions, had the best predictive performance, with the highest score of TSS (0.835; Table 3.4) and very high rank in both sensitivity and specificity (Figure 3.4) for the Iberian Peninsula. These results showed that building an ensemble model from single algorithm models with high predictive performances can lead to a more accurate and robust predictive model, while also highlighting areas of high uncertainty, providing valuable guidance for conservation strategies (Araújo & New, 2007).

Other ways of implementing an ensemble could have been performed, such as weighted average where probability values from models with higher TSS values have a higher contribution than weaker models (Guisan et al., 2017). While this approach could lead to an increase in predictive performance, it also might reduce generalizability, since the ensemble model would benefit single models that more closely align with the Iberian census data. In this work the simple mean was the preferred option since it can help mitigate the effects of bias in the occurrence data, ensuring more reliable predictions of ensemble model projections into future scenarios.

Overall, decision tree machine-learning algorithms (GBM and RF) significantly outperformed linear regression and maximum entropy approaches in this work. As previously mentioned, machine-learning algorithms were expected to have better performance and both GBM and RF have consistently shown to be very reliable algorithms in SDMs (Elith et al., 2006; Valavi et al., 2022). However, MaxEnt's low predictive performance was not expected, and the discrepancy between decision tree models and the remaining algorithms also came as a surprise. In other works (e.g., Valavi et al., 2022) MaxEnt was one of the best performing algorithms in a large selection of modeling approaches, demonstrating the importance of parameter tuning and data preparation in SDMs. In this work, the superior performance of the ensemble approach showed the advantage of aggregating multiple algorithms to improve predictive accuracy. Additionally, highlighting areas of high uncertainty allowed for a more cautious interpretation of results.

4.2 Environmental factors influencing distribution

From the bioclimatic covariates, maximum temperature and temperature seasonality had the highest influence in the final ensemble model (0.167 and 0.092, respectively; Table 3.6). In comparison, annual precipitation had very small influence in the model (0.039). These results are in line with previous works, where temperature had a higher influence than precipitation in modeling the breeding distribution of the Red-backed Shrike in the Iberian Peninsula (Reino et al., 2006; Tellería, 2018a).

In the land-use categories, locations with mainly forested, open or agricultural landscapes showed a higher suitability for the Red-backed Shrike (Figure 7.4). Open and agricultural land types were expected to be more suitable since this species often breeds in this kind of habitat (Tellería, 2018a; Yosef et al., 2020). The lower suitability in locations with extensive shrubs was also expected, since these

areas tend to not have patches of bare ground from which the Red-backed Shrike depends on for foraging (Lefranc & Worfolk, 1997). However, forested areas were not expected to show high levels of habitat suitability.

Furthermore, land-use seemed to have the least influence on the ensemble model (Table 3.6), which could have been caused by its very coarse resolution only showing the most prevalent land-use in a 0.5x0.5 º (approximately 50km) square area (Thuiller et al., 2004). Since species like the Red-backed Shrike only require a relatively open area with some available perches (e.g., small trees, bushes), a grid cell largely dominated by forested areas could have some smaller open areas which might be suitable for this species. However, that does not mean that forests are a suitable land-use for the presence of the Red-backed Shrike. Since this species is often found in a mosaic of habitats, like open areas with some scattered trees and shrubs, the coarse land-use data probably did not fully capture the complexity of its preferred habitats. Due to the very broad range of the study area, situations like this were inevitable, limiting the influence of land-use on habitat suitability. To better capture this ecological nuance, reducing grid cells to a higher resolution (smaller grid cells) or using different scales to represent land-use (e.g., percentage of vegetation cover, more 'land' categories) could be better ways of incorporating land information when trying to model the distribution of the Red-backed Shrike. A possible downside of reducing grid cells is the substantial increase in data volume, and adding more 'land' categories would trade-off model simplicity for complexity.

In addition, predicting the factors associated with distribution changes in migratory species such as the Red-backed Shrike is particularly challenging. Factors influencing the species' migration routes and habitat suitability during winter months were not considered in this work but could have a noticeable influence in future studies (Abdul-Wahab et al., 2024). Effects of global change on both breeding and wintering grounds could change migration routes, and potentially shift the timing of arrival and survival rate of individuals, ultimately affecting the Red-backed Shrike's breeding range (Pedersen et al., 2018).

4.3 Predicted range

Despite using different methods to relate occurrence data to environmental conditions, all models were able to estimate high likelihoods of occurrence of the Red-backed Shrike throughout its entire known breeding range (Figure 3.2). This result was in line with previous assertions that different modeling algorithms can be used in SDMs with varying degrees of accuracy (Elith et al., 2010; Guisan et al., 2017; Valavi et al., 2022). Models mainly predicted high suitability in central and eastern Europe, southern Scandinavia, in the border between Russia and Kazakhstan and in the northern regions of the Iberian Peninsula. This was likely due to maximum temperature (most influential environmental covariate) values in these regions ranging between 20-30ºC, which corresponds to high levels of habitat suitability in all modeling algorithms (Figure 7.9). In consequence, the ensemble model also showed high levels of habitat suitability in these areas, with low values of uncertainty.

Disagreements between model predictions were evident in locations near the edges of the breeding range, such as the Middle East, central Scandinavia, the Altai mountain range, and, more importantly, in the Iberian Peninsula. It is expected that locations in the edge of the distribution to have higher variance in model predictions, since these regions might represent the subtle gradient between suitable and unsuitable conditions for the Red-backed Shrike, leading to high levels of uncertainty in model

predictions (Pearson et al., 2006; Nenzén & Araújo, 2011). In the ensemble model, this uncertainty, measured as the coefficient of variation (CV), reached very high levels (max CV =123%) in some edge areas and also locations outside the known breeding range (Figure 7.5). This likely resulted from relatively optimistic GBM predictions conflicting with the more conservative predictions of RF, further amplified by the fact that these two groups were the only algorithms included in the final ensemble model. If there were single models from other algorithms in the ensemble, the discrepancy would probably decrease since there would either be more intermediate values close to the mean of the predictions, or more models would side with either GBM or RF predictions, bringing the standard deviation down. These findings highlight the caution needed to be taken when trying to interpret range shifts and their ecological or conservation implications, in particular for edge areas with high levels of uncertainty (Jetz et al., 2007; Elith et al., 2010).

There were, however, some glaring errors. For example, the Red-backed Shrike is known to breed in the Alps region, yet all models estimated a low habitat suitability in that region (Figure 3.5). This could be caused by the fact that breeding spots at lower altitudes have very different environmental pressures from spots at higher altitudes. This is a common issue with using a coarse resolution (0.5x0.5 º grid cells in this work) to map the data, where micro habitats such as valleys in the middle of a mountain range get obfuscated by the predominant environmental conditions at higher altitudes (Guisan et al., 2017). Using lower resolution could likely alleviate the problem, though with the large extent of the Red-backed Shrike's breeding range, this would lead to enormous datasets, drastically increasing computational power requirements.

In the Iberian Peninsula, all modeling algorithms used predicted high suitability to northern regions, although to varying degrees. The ensemble model predictions closely resembled the known distribution of the Red-backed Shrike, restricting high suitable areas to mostly inland and mountainous areas in the northern and central regions. These areas are known to have bioclimatic characteristics similar to central and northern Europe, and also represent the range edges of several Euro-Siberian bird species (Reino et al., 2006). Furthermore, moderately suitable areas were also noticeable in central Portugal, mountainous regions in southern Spain and near the Gibraltar strait, which closely resemble the results of Neto & Araújo (2011). These locations were likely on the limit of the tolerable bioclimatic conditions for the Red-backed Shrike, representing the interface between Atlantic and Mediterranean climate conditions. These unexpected results were likely erroneous extrapolations from the models, but could also signal areas where new populations are established in the future if the number of individuals in the Iberian Peninsula rises, driving individuals to seek new breeding spots with lower competition

4.4 Future scenarios of global change

As previously mentioned, SDMs consider two dispersion capacities of the study species to measure how global change impacts distribution ranges (Elith et al., 2010). When considering the two different dispersion metrics (No Dispersal vs. Full Dispersal), the value of the no dispersal metric remained fairly close to the value of the area covered by the current prediction range (Table 3.5). These results suggest that in the worst-case example the model predicts that the Red-backed Shrike should not lose much of its initial range. However, this metric is theoretically unsuitable in measuring how much of the current range of the Red-backed Shrike could be affected in future scenarios because it is a highly dispersive migratory bird. Therefore, the full dispersal range shift was assumed to be a more accurate representation

of the predicted future range changes (Barbet-Massin, Thuiller, et al., 2012). From this metric, the Red-backed Shrike was predicted to substantially increase its breeding range in all future scenarios (Table Results 3.5). However, 'no dispersal' range shifts can be very useful for conservation planning. Areas predicted to remain suitable to a target species under global change ensure that conservation actions taken will likely endure even in future conditions (Guisan et al., 2017).

Range shift predictions varied very little between the same GCMs in A1B and A2 emission scenarios (Figure 3.9), with most of the differences seeming rather small-scale local prediction disagreements. Most of the environmental differences between emission scenarios seemed to be from a land-use change rather than climate (Figures 7.10 - 7.13). Seeing as the variable 'land' had low influence on the ensemble model, the bioclimatic factors ended up being the primary drivers of range shifts. Consequently, differences between the model projections for the two carbon emission scenarios were not significantly different.

In contrast, range shift predictions varied more between the two GCMS used (Figure 7.7): MIROC-H (MR) showed a higher percentage of suitable habitat gain and also lower percentages of habitat suitability loss, when compared with CSIRO-Mk3.0 (CS) (Table 3.5). This was likely a product of the different environmental predictions between both GCMs, especially in temperature. MR scenarios showed localized increases in maximum temperature ('tmax') in regions like central Iberian Peninsula, western Turkey and northern Russia (Figure 7.10 and 7.12). The latter saw an increase in maximum temperature ranging between 2-3ºC, coupled with a decrease of 1-2ºC in temperature seasonality ('tsd') and a change in land-use from 'shrubland' into 'forest'. These changes were probably a key factor for the expansion of habitat suitability in the northeastern part of the range predicted by the ensemble model in MR scenarios.

The CS predictions showed a large increase in maximum temperature (3-6ºC) in a large array of territories such as the Baltic Sea, northern Scandinavia, and the Caucasus and Altai mountain ranges (Figure 7.11 and 7.13). However, CS maximum temperature increases in the northeastern part of the Red-backed Shrike's breeding distribution were more moderate compared to MR. Since maximum temperature had the highest importance of all environmental variables for the ensemble predictions, the more moderate increase in 'tmax' for the northeastern region under CS scenarios led to a lesser expansion in suitable habitat when compared with MR. Precipitation changes predicted for this area were minimal (0-100mm) and had negligible effects due to the low variable importance given to 'prec' in the ensemble model predictions.

Overall, model projections under global change scenarios predicted a northward colonization of the Red-backed Shrike into new breeding areas, while abandoning southern regions (Figure 3.9). Areas in northern Europe and Russia previously considered unsuitable for this species, were predicted to become more tolerable for the Red-backed Shrike, while the southern range was predicted to become more limited. These results were in line with previous assertions that European migratory birds are likely to shift towards northern territories as a consequence of global change (Jetz et al., 2007; Barbet-Massin, Thuiller, et al., 2012). The loss of suitable areas in the southern parts of the breeding range were much lesser than the gain from the northern areas (Table 3.5). If anything, the results of this work predicted that the Red-backed Shrike's breeding range will likely increase due to the effects of global change.

More locally in the Iberian Peninsula, in all future scenarios the ensemble model predicted contractions towards northern regions, but also an establishment of new southern populations in novel areas (Figure 3.10). The retreat of central populations into colder territories in the north as an impact of global change

was seen as a probable event in previous works (e.g., Neto & Araújo, 2011; Tellería, 2018a). However, the appearance of new suitable areas in the south was completely unexpected. In fact, with rising temperatures, southern territories that resemble Mediterranean climate would likely turn even more unsuitable for the Red-backed Shrike, as this species tends to avoid hot Mediterranean climate (Tellería, 2018a). A possible explanation could be the change in land-use predicted by these scenarios, even if it had seemingly low influence on the ensemble model. Compared to current conditions (Figure 7.14), future scenarios predicted a major transformation all over the Iberian Peninsula, particularly into agricultural land (Figures 7.15 – 7.18). These changes could have drastically influenced the ensemble projections in the south into higher habitat suitability, though this might be of limited consequence in the future.

Furthermore, interaction between covariates might play a role due to the effect of indirect gradients. Indirect gradients assume that a specific value of an environmental predictor (say land-use) does not have the same meaning across the study area (Albert et al., 2010). Agricultural areas in European countries could have extensively different characteristics (intensity, use of fertilizers, area of patches, local policies on maintenance of surrounding shrubs) than agricultural lands in Asian countries. Likewise, a maximum temperature value of 30º will probably have different impacts on animals living in forested areas (where shade is abundant), than animals living in open areas (more exposure to direct heat). Using a large extent of area to draw environmental information inevitably introduced these problems, although it was unknown to which degree. As such, a direct interpretation of these results would likely require a more in-depth analysis of how the ensemble models interactions between variables, and comparison between locations with similar conditions to the Iberian Peninsula under future scenarios.

Uncertainty, measured by coefficient of variation, was relatively low in future scenarios predictions when compared to current conditions (Figure 7.5 – 7.6). Edge areas, such as the Iberian Peninsula, Scandinavia, and the Mediterranean Basin, retained slightly higher uncertainty than locations near the core of the breeding distribution. These results could stem from the fact that future scenarios created a more uniform set of conditions in the study area, reducing the amount of variance in the model. Moreover, global change in edge areas may have pushed key environmental variables beyond a threshold that under current conditions led the GBM and RF to disagree in their predictions, leading to a higher agreement between future projections. This argument is further strengthened by the fact that under current conditions the model predictions for the same areas showed very high levels of uncertainty (Figure 7.5).

4.5 General remarks of SDMs and limitations of this work

Despite being a very powerful tool, SDMs applicability faces considerable challenges. From issues in the data to the huge array of methods in the modeling procedure, this type of analysis is filled with arbitrary choices that can end up influencing model predictions. Where were the pseudo-absences sampled? What were the modeling algorithms used? Which metric was selected to evaluate model predictive accuracy? The list goes on, and it can become very overwhelming. These points, coupled with the fact one of the main applications of SDMs is to project distribution into time periods or geographical spaces that differ from the ones used to calibrate the models, induce a high amount of variability and uncertainty in the results and interpretation (Guisan et al., 2017).

One of the main issues of SDMs is transferability (Elith et al., 2010; Guisan et al., 2017; Araújo et al., 2019). Evaluation metrics like TSS and AUC only show that the model is accurately predicting the available data, but verifying if the full range of ecological factors that influence the occurrence of the species has been captured is very limited (Elith et al., 2006; Pearson et al., 2006). The implication this has when trying to evaluate future scenarios of global change is that one has to exercise caution on whether the model has captured the entire environmental niche requirements in order to produce reliable predictions under novel conditions. This requires an in-depth understanding of the methods and assumptions taken within the modeling framework, to avoid issues such as overfit and extrapolation, assuring the production of a robust model capable of giving good predictions in novel environmental conditions (Araújo et al., 2019). In line with this, the results in this study showed that the ensemble model's threshold selected for the Iberian Peninsula still resulted in a relatively accurate prediction of the full breeding range, signaling that the model was able to capture a wide enough spectrum of environmental conditions needed to predict the distribution of the Red-backed Shrike (Manel et al., 2001; Guisan et al., 2017). If this was not the case, the model would have extrapolated, either by omission or commission, to compensate for the imbalance of its predictions. This was reasonable evidence that the final ensemble model was robust enough to make reliable projections into future novel environmental scenarios.

Another potential limitation was the possibility of genetic differences between populations across the breeding range. Genetic adaptations make populations within a confined region more adapted to habitat conditions in that area, so that individuals from two different populations could show different tolerances to environmental pressures (Elith & Leathwick, 2009). If the Red-backed Shrike showed high genetic adaptations to particular locations (e.g., Iberian Peninsula, central Europe, eastern Asia) such differences would need to be accounted for when projecting the distribution (Guisan et al., 2017). However, the Red-backed Shrike individuals have shown low percentages of individuals returning to the same breeding area in Poland (Tryjanowski et al., 2007), indicating a flow of breeding individuals to different areas. Furthermore, Pârâu et al. (2019) conducted a genetic analysis of Red-backed Shrike individuals across the Western Palearctic and revealed a high genetic diversity within the same populations, indicating a low level of local genetic adaptation. As such, modeling considerations for the Red-backed Shrike should be to include the whole breeding area, since breeding populations are expected to have the same (or very similar) environmental niches.

Threshold selection is yet another step that could introduce high variability in results, and consequently, in the interpretation (Nenzén & Araújo, 2011). Since the selected threshold was the one that maximizes sensitivity + specificity (hence maximizing TSS), the model results should be interpreted as trying to identify the areas where the species is more likely to occur (Pearson et al., 2006; Guisan et al., 2017). This threshold selection is usually used for conservation studies since it identifies the most suitable areas for the occurrence of the species, and therefore allows researchers to concentrate conservation efforts and resources into strategic key areas where the species is more likely to occur. It should be noted that in ecology it's expected for a species to not occupy its full potential range (Guisan & Zimmermann, 2000). The full range of locations with ecological conditions favorable to the occurrence of the species is called the fundamental niche, while the range that the species does occupy is called the realized niche. Presence data gives information about where the species is observed (realized niche), not the full range of locations where the environmental conditions are suitable (fundamental niche). The implication this has for the models is that sometimes maximizing sensitivity may be preferred, so as to remain as close as possible to the realized niche (Pearson et al., 2006). The gain of commission errors (false presences) derived from it can then be seen as locations where the conditions are considered suitable but the species

was not detected or that are out of the actual range of the species, and therefore part of the fundamental niche. However, the threshold that maximizes TSS is the most commonly used in literature, as well as the most indicated to project the effects of global change in the distribution of the species since it best represents the observed distribution of the species (Guisan et al., 2017). As such, the use of this threshold allowed future projections to more accurately represent probable range shifts, hence making it more justified given the objectives of this study.

Seeing as SDMs capture the environmental drivers of species distributions, there needs to be an assumption that the species is being mostly, if not solely, constrained by its environment into occupying its distribution range (Araújo et al., 2019). This is called being in equilibrium with its environment, and it's a very important assumption when trying to model the distribution of a species (Guisan et al., 2017). If a species is not at equilibrium with its environment, its occupied range is not mainly determined by its environment but by other external factors. This is most commonly the case in invasive species where the distribution observed in the novel territories cannot be assumed to be environmentally constrained, due to the possibility that the invading species simply did not have enough time to expand and establish populations in suitable areas nearby (Gallien et al., 2012). Given that the Red-backed Shrike is a long-distance migratory bird species, its capability of dispersing is incredibly high. Therefore, it is reasonable to assume that its observed realized niche is probably being constrained mainly by environmental conditions (Brotons et al., 2004). If there are available suitable areas, the Red-backed Shrike will probably occupy them, and if there are changes in the conditions of its environment, its mobility gives it the ability to rapidly respond to these changes. For these reasons, it was assumed that the breeding distribution of the Red-backed Shrike was being conditioned mainly by environmental factors, and therefore the application of SDMs should accurately represent the drivers of its breeding distribution (Elith & Leathwick, 2009).

Like most modeling procedures, data quality is one of the most important considerations into building reliable models. Most SDMs in literature use large online data collection databases (e.g., GBIF) that produce exclusively presence-only data, or use census data usually restricted to smaller areas that are able to reliably produce presence-absence data (Guisan et al., 2017). In this work a hybrid method was used, with the EBBA2 data mapping the data in the European part of the range and GBIF to also map occurrences in breeding locations in eastern Asia. This means the data used is a mix of presence-absence and presence-only data. This approach is not commonly used for valid reasons. One of the problems is the selection of pseudo-absences. Pseudo-absences rely on the assumption that true absences cannot reliably be determined, but with the inclusion of EBBA2 data the absences in the European part of the study area are considered true absences (Barbet-Massin, Jiguet, et al., 2012). Another issue was the independence between calibration and evaluation data sets, seeing as including census data from EBBA2 in the calibration dataset would reduce its independence from the Iberian Peninsula census data of the evaluation set (Elith et al., 2006). This raises the question of why analyses were not performed using only one of the databases (EBBA2 or GBIF). While these are valid concerns, there were stronger arguments that the methods applied were justifiable in both an ecological and modeling perspective.

From the ecological perspective, using only EBBA2 data raises a big problem: niche truncation. Niche truncation happens when part of the range of the study species is not included in the data, leading to models failing to capture the full range of environmental conditions that are suitable for the species (Albert et al., 2010). For example, if the data used for modeling was restricted to areas where cells with species presences had a maximum temperature between 10º-30ºC, the models would likely assume that a location with a maximum temperature of 40ºC is unsuitable for the occurrence of the Red-backed

Shrike. However, if it was shown that cells with presences that were not included in the data had maximum temperature of 40ºC, the model would be considered unreliable since it predicts these points as unsuitable when in fact the species does occur under these conditions. This is especially problematic when trying to model global change effects in the future, where maximum temperatures could increase throughout the whole range, leading models to likely give incorrect projections of the species distribution (Albert et al., 2010; Araújo et al., 2019; Elith et al., 2010). For this reason, it was crucial that the whole range of the Red-backed Shrike was included in the data when modeling the distribution, rendering using only the EBBA2 data option as improper.

In contrast, since GBIF collects data through the whole extent of the breeding range, using this database exclusively could be a viable option. However, from a modeling perspective using a citizen science database raises yet another challenge: sampling bias. Sampling bias is one of, if not the biggest, issues in SDMs so countering it was one of the main challenges (Beck et al., 2014; Guisan et al., 2017). If the presence data is largely condensed into well sampled spots near urbanized areas in developed countries, the models will overfit to this kind of data while presences in underrepresented spots, such as hard to traverse terrain (like bogs or mountains) or areas in countries where data collection is limited, will have very little influence in the model. Following this, the information in the models would probably only capture the environmental information of the well sampled areas, leading to unreliable projections similar to niche truncation. While introducing EBBA2 data would be thought to increase the problem, since the European part of the range was already well represented, the EBBA2 census extended to the western part of Russia, an area that was underrepresented in the GBIF dataset, adding missing information into the models. The combined data from EBBA2 and GBIF was not considered an effort to determine the Red-backed Shrike occurrence in the whole extent of the breeding range, since most of the eastern part of the breeding range was undersampled. Therefore, the data was considered as presence-only and sampled pseudo-absences could still be considered theoretically valid (Elith et al., 2006).

Due to the proximity sampled locations, SDMs often show spatial autocorrelation, where nearby locations tend to have similar environmental values, potentially leading to overfit (Guisan et al., 2017). This problem usually arises from the geographical continuity within the breeding range of species, where populations are interconnected, creating continuous spatial patterns characterized by similar environmental conditions. Addressing spatial autocorrelation usually involves methods that also counter sampling bias, such as spatial filtering of occurrence data used in this work. This reduces the likelihood that model predictions are not influenced by spatial patterns in the occurrence data (Segurado et al., 2006). Spatial autocorrelation can be assessed using metrics such as Moran's I that verifies if model residuals are spatially related to each other (Guisan et al., 2017). Though it was not assessed, spatial autocorrelation is likely to happen in SDMs and could have had an effect in the results of this thesis.

The fact that the ensemble model predictions were able to accurately represent the sub-sampled eastern areas of the breeding range showed that the implemented method of countering sampling bias was effective. It should be noted that using a variable as a way to counter sampling bias in the data is not the most commonly used method in literature. Other methods such as using weights for presences and absences (Elith et al., 2010), as well as using other species that are likely to be sampled with the target-species as pseudo-absences (Phillips et al., 2009) are more widespread methods in SDM works. Unfortunately, the R package *biomod2* seemed to be unable to work with weighted presence and absence data, since it did not incorporate the weights into the MaxEnt and Random Forests algorithms (https://github.com/biomodhub/biomod2/issues/403, last accessed 13 of May 2024), leaving target-

group pseudo-absences as the only viable alternative. Nonetheless, the results showed that the implementation of a sampling bias variable seemed to be effective, although it makes the interpretation of variable importance more difficult, since now there was a variable that outweighs other variables in the model, obfuscating the contribution that other variables might have in the model (Guisan et al., 2017). Using different R packages (e.g., *dismo*) to overcome the weighted data problem might be a more suitable option to model the distribution of the target species in future works.

5 Conclusion

This thesis focused on predicting the distribution of the Red-backed Shrike, as well as assessing the impacts of global change on range shifts. This was accomplished through the use of relevant bioclimatic and land-use covariates to estimate the probability of occurrence at each 0,5º cell across the study area. The result was a reasonably accurate prediction of the Red-backed Shrike's real breeding distribution, and a deeper understanding of how each covariate affects habitat quality for the occurrence of this species. Furthermore, the future global change impacts were predicted to lead to an expansion of northern suitable breeding areas at the cost of southern ones. In the Iberian Peninsula this led to a peculiar result that was very different from the overall prediction.

Most SDM works, such as this one, usually focus on smaller scales. The inclusion of modeling the whole distribution came later on, as the results were also obtained for that scale, and it ended up adding a much deeper dynamic than exclusively modeling the distribution for the Iberian Peninsula. As such, the work done in this thesis could highlight the importance of modeling both the whole range and specific target areas in SDMs to get a more inclusive perspective of the results generated.

Overall, this work put on display some of the major challenges of SDMs. This type of modeling has a lot of subjective procedures from start to finish, that could ultimately influence the results, such as choice of models, extent of study area and which covariates to use. Furthermore, the uncontrollable nature of the data made it much harder to understand what actually impacts the distribution of a species, especially one with such a wide range as the Red-backed Shrike. This was further amplified by the use of relatively big grid cells (0.5º) that probably obfuscated more detailed information.

In the end, this thesis represented the learning experience acquired through this study. Some errors occurred and missteps were made, but they ended up being counterbalanced by applying corrections and deeply analyzing the results. If this whole work had to be redone some changes would inevitably have to be implemented, but for the most part the procedure performed was well done and the results provided made sense with what was expected. As a final note, this kind of study highlights the importance of complementing the application of statistical analysis with knowledge from other fields, such as ecology to make more informed decisions and further interpret the results obtained.

6 Bibliographic references

Abdul-Wahab, C., Costa, J. S., D'Mello, F., & Häkkinen, H. (2024). Connected impacts: Combining migration tracking data with species distribution models reveals the complex potential impacts of climate change on European bee-eaters. *Journal of Ornithology*. https://doi.org/10.1007/s10336-024-02190-z

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

Albert, C. H., Yoccoz, N. G., Edwards, T. C., Graham, C. H., Zimmermann, N. E., & Thuiller, W. (2010). Sampling in ecology and evolution – bridging the gap between theory and practice. *Ecography*, *33*(6), 1028–1037. https://doi.org/10.1111/j.1600-0587.2010.06421.x

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Anderson, B. J., Akçakaya, H. R., Araújo, M. B., Fordham, D. A., Martinez-Meyer, E., Thuiller, W., & Brook, B. W. (2009). Dynamics of range margins for metapopulations under climate change. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1661), 1415–1420. https://doi.org/10.1098/rspb.2008.1681

Anderson, R. P. (2003). Real vs. artefactual absences in species distributions: Tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *Journal of Biogeography*, *30*(4), 591–605. https://doi.org/10.1046/j.1365-2699.2003.00867.x

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

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, *22*(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Araújo, M. B., & Rahbek, C. (2006). How Does Climate Change Affect Biodiversity? *Science*, *313*(5792), 1396–1397. https://doi.org/10.1126/science.1131758

Araújo, M. B., Whittaker, R. J., Ladle, R. J., & Erhard, M. (2005). Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, *14*(6), 529–538. https://doi.org/10.1111/j.1466-822X.2005.00182.x

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, *3*(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barbet-Massin, M., Thuiller, W., & Jiguet, F. (2012). The fate of European breeding birds under climate, land-use and dispersal scenarios. *Global Change Biology*, *18*(3), 881–890. https://doi.org/10.1111/j.1365-2486.2011.02552.x

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, *19*, 10–15. https://doi.org/10.1016/j.ecoinf.2013.11.002

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brotons, L., Thuiller, W., Araújo, M. B., & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, *27*(4), 437–448. https://doi.org/10.1111/j.0906-7590.2004.03764.x

Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, *333*(6045), 1024–1026. https://doi.org/10.1126/science.1206432

EBCC. (2022). *European Breeding Bird Atlas 2 website*. http://ebba2.info

Elith, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., … E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*(2), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

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

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*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

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

Fernández, P., Jordano, D., & Haeger, J. F. (2015). Living on the edge in species distribution models: The unexpected presence of three species of butterflies in a protected area in southern Spain. *Ecological Modelling*, *312*, 335–346. https://doi.org/10.1016/j.ecolmodel.2015.05.032

Gallien, L., Douzet, R., Pratte, S., Zimmermann, N. E., & Thuiller, W. (2012). Invasive species distribution models – how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, *21*(11), 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x

GBIF.Org User. (2024). *Occurrence Download*. The Global Biodiversity Information Facility. https://doi.org/10.15468/DL.D2K9EH

Gordon, H., Rotstayn, L., McGregor, J., Dix, M., Kowalczyk, E., O'farrell, S., Waterman, L., Hirst, A., Wilson, S., & Collier, M. (2002). *The CSIRO Mk3 climate system model*.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Thuiller, W., & Zimmermann, N. E. (with Di Cola, V., Georges, D., & Psomas, A.). (2017). *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press. https://doi.org/10.1017/9781139028271

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*(2–3), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Hanberry, B. B., & He, H. S. (2013). Prevalence, statistical thresholds, and accuracy assessment for species distribution models. *Web Ecology*, *13*(1), 13–19. https://doi.org/10.5194/we-13-13-2013

Harris, T. (2000). *Shrikes and bush-shrikes: Including wood-shrikes, helmet-shrikes, flycatcher-shrikes, philentomas, batises, and wattle-eyes*. Princeton University Press.

Hastie, T. J., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed). Springer.

IMAGE contributors. (2017, April). *Download packages*. http://themasites.pbl.nl/models/image/index.php/Welcome_to_IMAGE_3.0_Documentation

Instituto da Conservação da Natureza (Ed.). (2008). *Atlas das aves nidificantes em Portugal (1999-2005)*. Assírio & Alvim.

IPCC (Ed.). (2008). *Climate Change 2007: Synthesis report: a report of the Intergovernmental Panel on Climate Change*. IPCC.

IUCN. (2016). *Lanius collurio: BirdLife International: The IUCN Red List of Threatened Species 2017: e.T22705001A110988087* [Dataset]. https://doi.org/10.2305/IUCN.UK.2017-1.RLTS.T22705001A110988087.en

Jetz, W., Wilcove, D. S., & Dobson, A. P. (2007). Projected Impacts of Climate and Land-Use Change on the Global Diversity of Birds. *PLoS Biology*, *5*(6), e157. https://doi.org/10.1371/journal.pbio.0050157

K-1 model developers. (2004). *K-1 Coupled GCM (MIROC) Description*.

Keller, V., Herrando, S., Voříšek, P., Franch, M., Kipson, M., Milanesi, P., Martí, D., Anton, M., Klvaňová, A., Kalyakin, M. V., Bauer, H.-G., & Foppen, R. P. B. (2020). *European Breeding Bird Atlas 2: Distribution, Abundance and Change*. Lynx Edicions.

Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J., & Scott, J. K. (2012). CliMond: Global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, *3*(1), 53–64. https://doi.org/10.1111/j.2041-210X.2011.00134.x

Lefranc, N., & Worfolk, T. (1997). *Shrikes: A guide to the shrikes of the world*. Pica Press [u.a.].

Lehikoinen, A., Brotons, L., Calladine, J., Campedelli, T., Escandell, V., Flousek, J., Grueneberg, C., Haas, F., Harris, S., Herrando, S., Husby, M., Jiguet, F., Kålås, J. A., Lindström, Å., Lorrillière, R., Molina, B., Pladevall, C., Calvi, G., Sattler, T., … Trautmann, S. (2019). Declining population trends of European mountain birds. *Global Change Biology*, *25*(2), 577–588. https://doi.org/10.1111/gcb.14522

Liu, C., Newell, G., & White, M. (2016). On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution*, *6*(1), 337–348. https://doi.org/10.1002/ece3.1878

Manel, S., Williams, H. C., & Ormerod, S. J. (2001). Evaluating presence–absence models in ecology: The need to account for prevalence. *Journal of Applied Ecology*, *38*(5), 921–931. https://doi.org/10.1046/j.1365-2664.2001.00647.x

Martí, R. & Sociedad Española de Ornitología (Eds.). (2003). *Atlas de las aves reproductoras de España*. OAPN & SEO.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157. https://doi.org/10.1007/BF02295996

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*(1), 8221. https://doi.org/10.1038/ncomms9221

Muluneh, M. G. (2021). Impact of climate change on biodiversity and food security: A global perspective—a review article. *Agriculture & Food Security*, *10*(1), 36. https://doi.org/10.1186/s40066-021-00318-5

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, *37*(2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Nenzén, H. K., & Araújo, M. B. (2011). Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling*, *222*(18), 3346–3354. https://doi.org/10.1016/j.ecolmodel.2011.07.011

Neto, D., & Araújo, M. B. (2011). *Impactos, Vulnerabilidad y Adaptación al Cambio Climático de la Biodiversidad Española. 2 Fauna de Vertebrados.* Unpublished. https://doi.org/10.13140/RG.2.1.3766.3200

Pârâu, L. G., Frias-Soler, R. C., & Wink, M. (2019). High Genetic Diversity among Breeding Red-Backed Shrikes Lanius collurio in the Western Palearctic. *Diversity*, *11*(3), 31. https://doi.org/10.3390/d11030031

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P., & Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography*, *33*(10), 1704–1711. https://doi.org/10.1111/j.1365-2699.2006.01460.x

Pedersen, L., Schnedler-Meyer, N. A., Ekberg, P., & Tøttrup, A. P. (2018). Effects of forest management practices in clearings on breeding performance of the Red-backed Shrike (Lanius collurio). *Ornis Fennica*, *95*(4). https://doi.org/10.51812/of.133940

Pereira, H. M., Navarro, L. M., & Martins, I. S. (2012). Global Biodiversity Change: The Bad, the Good, and the Unknown. *Annual Review of Environment and Resources*, *37*(1), 25–50. https://doi.org/10.1146/annurev-environ-042911-093511

Perkins, S. E., Pitman, A. J., Holbrook, N. J., & McAneney, J. (2007). Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions. *Journal of Climate*, *20*(17), 4356–4376. https://doi.org/10.1175/JCLI4253.1

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, *31*(2), 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

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

Ponder, W. F., Carter, G. A., Flemons, P., & Chapman, R. R. (2001). Evaluation of Museum Collection Data for Use in Biodiversity Assessment. *Conservation Biology*, *15*(3), 648–657. https://doi.org/10.1046/j.1523-1739.2001.015003648.x

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, *9*(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reino, L. (2005). *O Picanço-de-dorso-ruivo*. João Azevedo.

Reino, L., Beja, P., & Heitor, A. C. (2006). Modelling spatial and environmental effects at the edge of the distribution: The red-backed shrike *Lanius collurio* in Northern Portugal. *Diversity and Distributions*, *12*(4), 379–387. https://doi.org/10.1111/j.1366-9516.2006.00245.x

Reino, L., Triviño, M., Beja, P., Araújo, M. B., Figueira, R., & Segurado, P. (2018). Modelling landscape constraints on farmland bird species range shifts under climate change. *Science of The Total Environment*, *625*, 1596–1605. https://doi.org/10.1016/j.scitotenv.2018.01.007

Rew, J., Cho, Y., Moon, J., & Hwang, E. (2020). Habitat Suitability Estimation Using a Two-Stage Ensemble Approach. *Remote Sensing*, *12*(9), 1475. https://doi.org/10.3390/rs12091475

Ridgeway, G. (2007). Generalized Boosted Models: A guide to the gbm package. *Update*, *1*(1), 2007.

RStudio Team. (2021). *RStudio: Integrated Development Environment for R* [Computer software]. RStudio, PBC. http://www.rstudio.com/

Sándor, A. D., & Domşa, C. (2018). Climate change, predictive modelling and grassland specialists: Assessing impacts of changing climate on the long-term conservation of Lesser Grey Shrikes (Lanius minor) in Romania. *Journal of Ornithology*, *159*(2), 413–424. https://doi.org/10.1007/s10336-017-1510-3

Segurado, P., Araújo, M. B., & Kunin, W. E. (2006). Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, *43*(3), 433–444. https://doi.org/10.1111/j.1365-2664.2006.01162.x

Smith, A. B., & Santos, M. J. (2020). Testing the ability of species distribution models to infer variable importance. *Ecography*, *43*(12), 1801–1813. https://doi.org/10.1111/ecog.05317

Strubbe, D., Broennimann, O., Chiron, F., & Matthysen, E. (2013). Niche conservatism in non-native birds in E urope: Niche unfilling rather than niche expansion. *Global Ecology and Biogeography*, *22*(8), 962–970. https://doi.org/10.1111/geb.12050

Tellería, J. L. (2018a). Distribution of the Red-Backed Shrike Lanius Collurio at Its Western Range Boundary: Patterns and Conservation Prospects. *Ardeola*, *65*(2), 221. https://doi.org/10.13157/arla.65.2.2018.ra2

Tellería, J. L. (2018b). Old Counts Suggest the Collapse of Two Red-Backed Shrike Lanius collurio Populations. *Ardeola*, *65*(2). https://doi.org/10.13157/arla.65.2.2018.sc1

Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, *10*(12), 2020–2027. https://doi.org/10.1111/j.1365-2486.2004.00859.x

Thuiller, W., Araújo, M. B., & Lavorel, S. (2004). Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*, *31*(3), 353–361. https://doi.org/10.1046/j.0305-0270.2003.00991.x

Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., Lafourcade, B., Patin, R., & Blancheteau, H. (2012). *biomod2: Ensemble Platform for Species Distribution Modeling* (p. 4.2-5–2) [Dataset]. https://doi.org/10.32614/CRAN.package.biomod2

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., & Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences*, *102*(23), 8245–8250. https://doi.org/10.1073/pnas.0409902102

Tryjanowski, P., Goławski, A., Kuźniak, S., Mokwa, T., & Antczak, M. (2007). Disperse or Stay? Exceptionally High Breeding-Site Infidelity in the Red-Backed Shrike *Lanius collurio*. *Ardea*, *95*(2), 316–320. https://doi.org/10.5253/078.095.0214

Tryjanowski, P., Karg, M. K., & Karg, J. (2003). Diet composition and prey choice by the red-backed shrike Lanius collurio in western Poland. *Belgian Journal of Zoology*, *133*(2), 157–162.

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

VanDerwal, J., Falconi, L., Januchowski, S., Shoo, L., & Collin, S. (2014). *SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises* [Dataset]. https://CRAN.R-project.org/package=SDMTools

Yosef, R., International Shrike Working Group, I., & Christie, D. (2020). Red-backed Shrike (Lanius collurio). In S. M. Billerman, B. K. Keeney, P. G. Rodewald, & T. S. Schulenberg (Eds.), *Birds of the World*. Cornell Lab of Ornithology. https://doi.org/10.2173/bow.rebshr1.01

7 Annex

Table 7.1: Evaluation metrics for all models fitted.

Model	Run	Threshold	AUC	Omission rate	Sensitivity	Specificity	Proportion correct	Kappa	TSS
MaxEnt	1	0.33	0.767	0.131	0.869	0.665	0.739	0.486	0.533
GLM	1	0.89	0.714	0.283	0.717	0.711	0.713	0.409	0.428
GBM	1	0.86	0.806	0.192	0.808	0.803	0.805	0.592	0.612
RF	1	0.57	0.892	0.111	0.889	0.896	0.893	0.773	0.785
MaxEnt	2	0.41	0.705	0.364	0.636	0.775	0.724	0.408	0.411
GLM	2	0.90	0.656	0.364	0.636	0.676	0.662	0.300	0.313
GBM	2	0.85	0.879	0.040	0.960	0.798	0.857	0.710	0.757
RF	2	0.60	0.879	0.121	0.879	0.879	0.879	0.743	0.757
MaxEnt	3	0.35	0.736	0.263	0.737	0.734	0.735	0.452	0.471
GLM	3	0.88	0.673	0.313	0.687	0.659	0.669	0.326	0.346
GBM	3	0.87	0.843	0.182	0.818	0.867	0.849	0.678	0.685
RF	3	0.62	0.900	0.101	0.899	0.902	0.901	0.789	0.801
MaxEnt	4	0.47	0.757	0.283	0.717	0.798	0.768	0.507	0.515
GLM	4	0.84	0.738	0.131	0.869	0.607	0.702	0.425	0.476
GBM	4	0.86	0.863	0.101	0.899	0.827	0.853	0.696	0.726
RF	4	0.55	0.868	0.131	0.869	0.867	0.868	0.720	0.736
MaxEnt	5	0.19	0.722	0.111	0.889	0.555	0.676	0.388	0.444
GLM	5	0.81	0.734	0.121	0.879	0.590	0.695	0.415	0.468
GBM	5	0.85	0.880	0.061	0.939	0.821	0.864	0.721	0.760
RF	5	0.61	0.898	0.071	0.929	0.867	0.890	0.770	0.796
MaxEnt	6	0.41	0.747	0.253	0.747	0.746	0.746	0.474	0.493
GLM	6	0.82	0.729	0.131	0.869	0.590	0.691	0.407	0.458
GBM	6	0.85	0.860	0.101	0.899	0.821	0.849	0.689	0.720
RF	6	0.55	0.879	0.121	0.879	0.879	0.879	0.743	0.757
MaxEnt	7	0.31	0.693	0.313	0.687	0.699	0.695	0.370	0.386
GLM	7	0.83	0.736	0.141	0.859	0.613	0.702	0.423	0.471
GBM	7	0.86	0.881	0.071	0.929	0.832	0.868	0.727	0.762
RF	7	0.58	0.905	0.091	0.909	0.902	0.904	0.797	0.811
MaxEnt	8	0.42	0.734	0.273	0.727	0.740	0.735	0.450	0.467
GLM	8	0.91	0.642	0.364	0.636	0.647	0.643	0.269	0.284
GBM	8	0.86	0.850	0.081	0.919	0.780	0.831	0.657	0.700
RF	8	0.67	0.905	0.091	0.909	0.902	0.904	0.797	0.811
MaxEnt	9	0.38	0.747	0.222	0.778	0.717	0.739	0.467	0.495
GLM	9	0.89	0.650	0.354	0.646	0.653	0.651	0.284	0.300
GBM	9	0.86	0.810	0.212	0.788	0.832	0.816	0.610	0.620
RF	9	0.54	0.895	0.101	0.899	0.890	0.893	0.774	0.789
MaxEnt	10	0.41	0.755	0.242	0.758	0.751	0.754	0.489	0.509
GLM	10	0.89	0.684	0.303	0.697	0.671	0.680	0.347	0.367
GBM	10	0.86	0.856	0.051	0.949	0.763	0.831	0.661	0.713
RF	10	0.58	0.879	0.121	0.879	0.879	0.879	0.743	0.757
MaxEnt	full	0.33	0.747	0.222	0.778	0.717	0.739	0.467	0.495
GLM	full	0.89	0.673	0.313	0.687	0.659	0.669	0.326	0.346
GBM	full	0.86	0.872	0.071	0.929	0.815	0.857	0.706	0.744
RF	full	0.60	0.905	0.091	0.909	0.902	0.904	0.797	0.811
Ensemble	TSS>=0.7	0.70	0.918	0.061	0.939	0.896	0.912	0.814	0.835

Table 7.2: Variable importance for all models fitted.

Model	Run	tmax	tsd	prec	bias	land
MaxEnt	1	0.355	0.178	0.052	0.386	0.081
GLM	1	0.177	0.051	0.125	0.268	0.045
GBM	1	0.132	0.053	0.019	0.575	0.037
RF	1	0.214	0.163	0.067	0.561	0.080
MaxEnt	2	0.369	0.243	0.012	0.329	0.128
GLM	2	0.159	0.040	0.137	0.297	0.046
GBM	2	0.133	0.050	0.024	0.599	0.033
RF	2	0.215	0.149	0.063	0.548	0.077
MaxEnt	3	0.371	0.208	0.022	0.389	0.102
GLM	3	0.143	0.054	0.149	0.277	0.046
GBM	3	0.128	0.045	0.024	0.582	0.038
RF	3	0.222	0.156	0.083	0.528	0.076
MaxEnt	4	0.334	0.178	0.047	0.375	0.108
GLM	4	0.148	0.041	0.146	0.273	0.054
GBM	4	0.106	0.048	0.029	0.575	0.042
RF	4	0.197	0.142	0.007	0.543	0.083
MaxEnt	5	0.346	0.169	0.023	0.384	0.118
GLM	5	0.151	0.046	0.149	0.276	0.052
GBM	5	0.122	0.047	0.026	0.579	0.037
RF	5	0.207	0.148	0.007	0.543	0.084
MaxEnt	6	0.336	0.165	0.062	0.394	0.091
GLM	6	0.161	0.036	0.105	0.299	0.043
GBM	6	0.129	0.038	0.028	0.621	0.037
RF	6	0.202	0.149	0.075	0.539	0.078
MaxEnt	7	0.382	0.197	0.012	0.342	0.104
GLM	7	0.168	0.040	0.127	0.275	0.053
GBM	7	0.133	0.055	0.018	0.579	0.037
RF	7	0.215	0.158	0.077	0.505	0.075
MaxEnt	8	0.363	0.190	0.026	0.353	0.096
GLM	8	0.147	0.052	0.132	0.259	0.062
GBM	8	0.138	0.053	0.002	0.557	0.044
RF	8	0.235	0.152	0.061	0.541	0.086
MaxEnt	9	0.371	0.139	0.019	0.408	0.103
GLM	9	0.156	0.048	0.139	0.279	0.051
GBM	9	0.122	0.043	0.027	0.560	0.046
RF	9	0.215	0.148	0.078	0.541	0.084
MaxEnt	10	0.379	0.195	0.001	0.428	0.094
GLM	10	0.105	0.045	0.141	0.296	0.041
GBM	10	0.127	0.048	0.003	0.588	0.030
RF	10	0.207	0.148	0.063	0.523	0.070
MaxEnt	full	0.351	0.192	0.002	0.384	0.103
GLM	full	0.166	0.049	0.133	0.277	0.048
GBM	full	0.123	0.044	0.022	0.586	0.038
RF	full	0.223	0.155	0.076	0.573	0.078
Ensemble	mean	0.167	0.092	0.039	0.551	0.049
Ensemble	cv	0.300	0.284	0.274	0.460	0.225

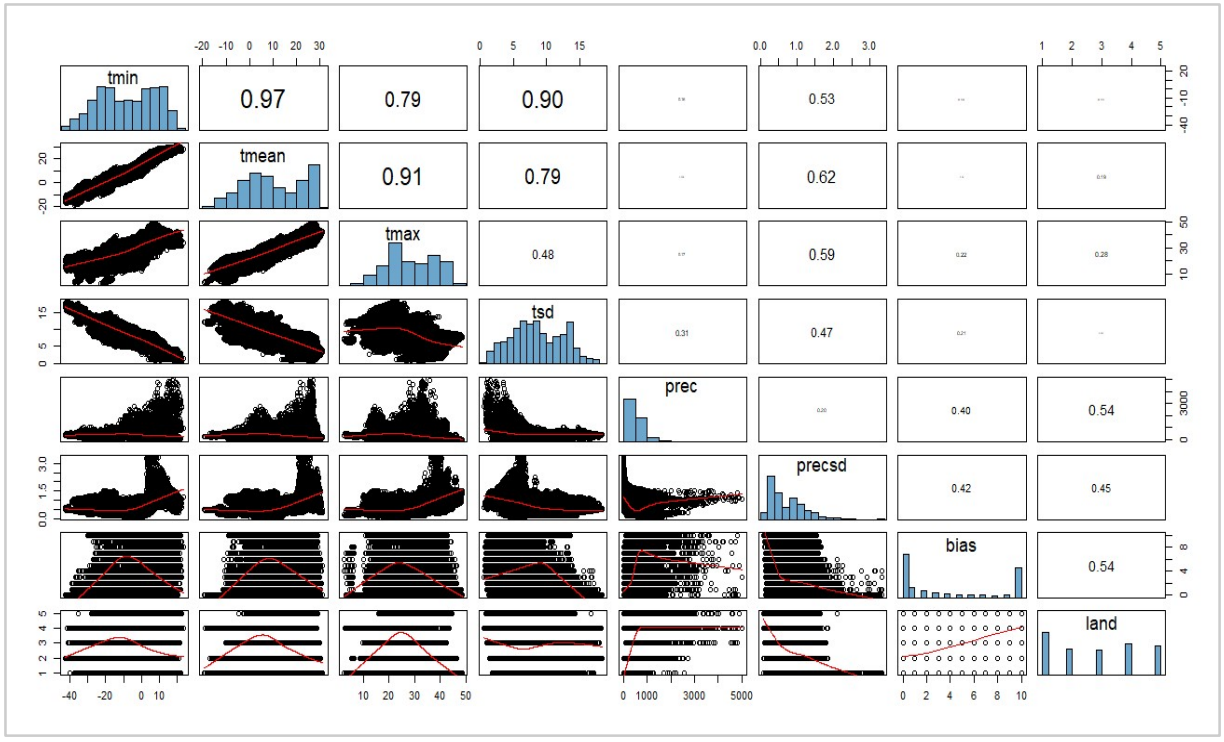

Figure 7.1: Pairwise correlation plot of all environmental covariates considered in this thesis. Lower diagonal shows bivariate scatterplots with a line representing the relationship, diagonal displays histograms with the distribution of each covariate, and upper diagonal reports Pearson correlation between each pair of covariates with bigger font size corresponding to stronger correlations.

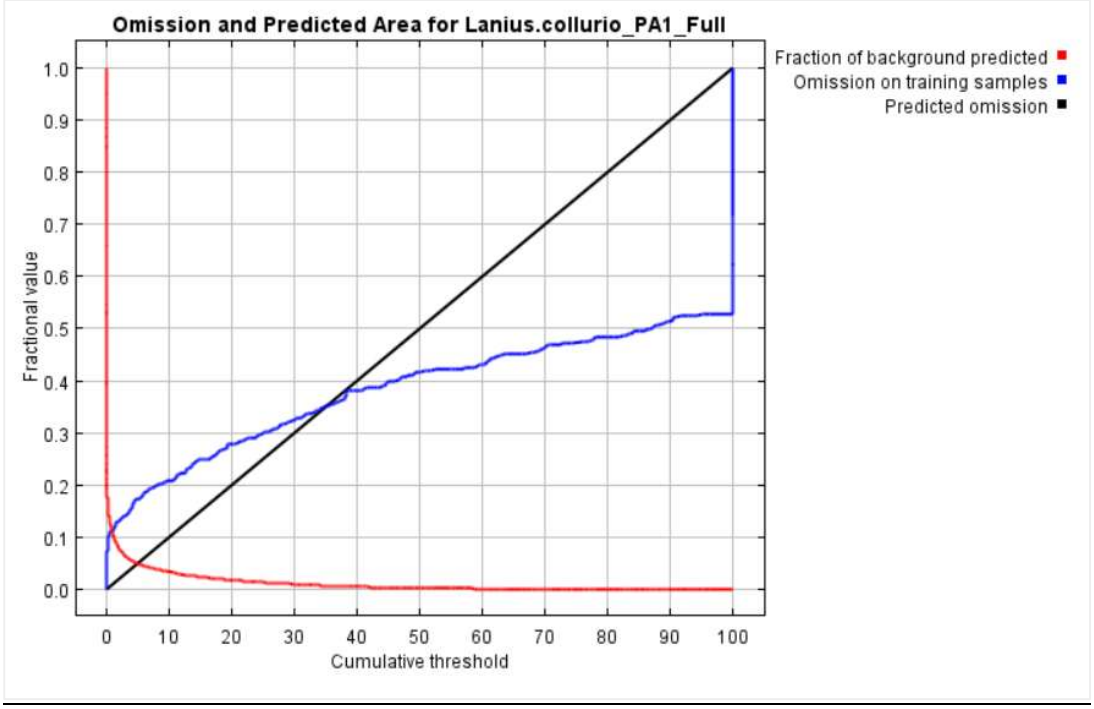

Figure 7.2: Omission and predicted area graph produced by MaxEnt. As threshold values (x-axis) increase, the fraction of predicted area should also increase, and lines closely following the 1-1 diagonal predicted omission usually show well calibrated features coefficients.

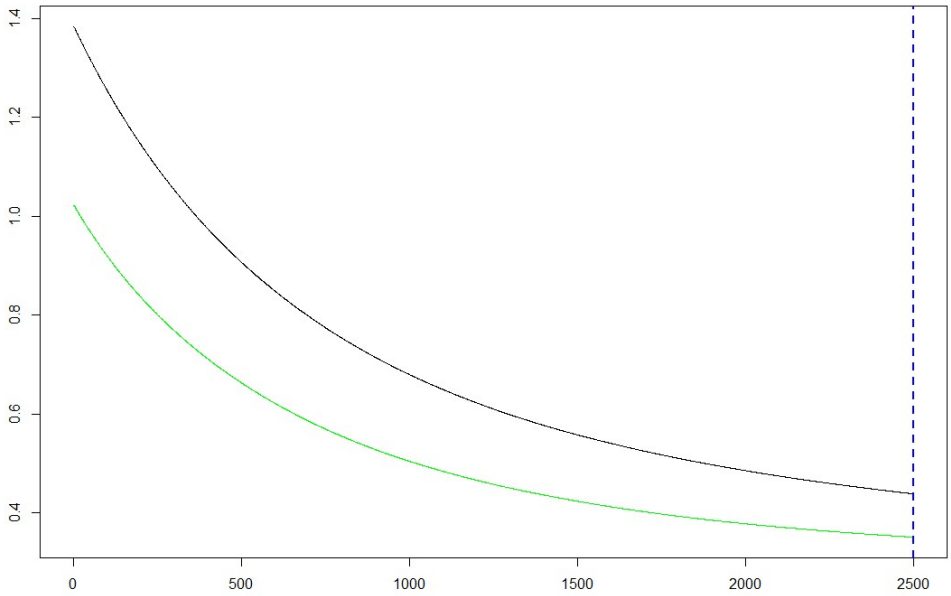

Figure 7.3: GBM graph of error rate (y-axis) against number of trees computed (x-axis). As the forward stage wise procedure runs, the model learns from previous mistakes and iteratively reduces error rate with each tree added. The goal is to run enough trees so that the error rate doesn't change significantly as more trees are added. Black line represents training data used to estimate the coefficients of the trees, and the green line represents the Iberian Peninsula evaluation data. Dashed blue line represents lowest error rate achieved for the evaluation dataset, and marks the number of trees that should be run.

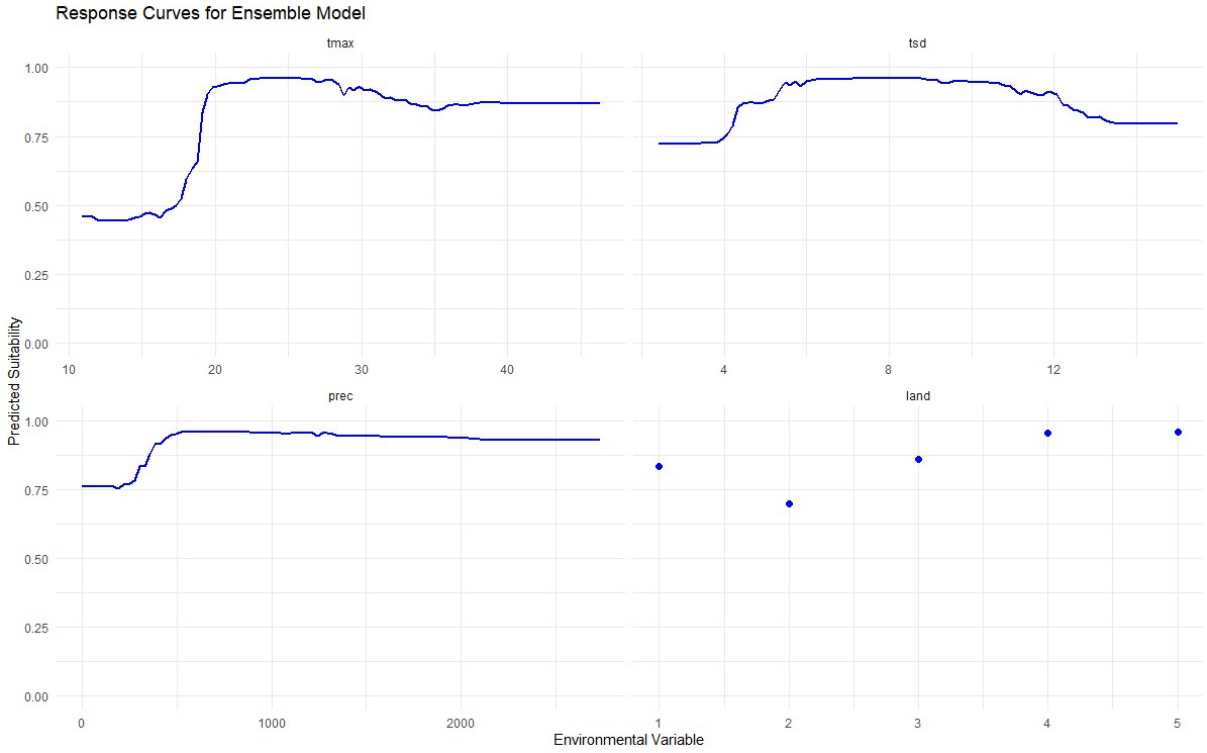

Figure 7.4: Response curves of the ensemble model predictions for each environmental covariate.

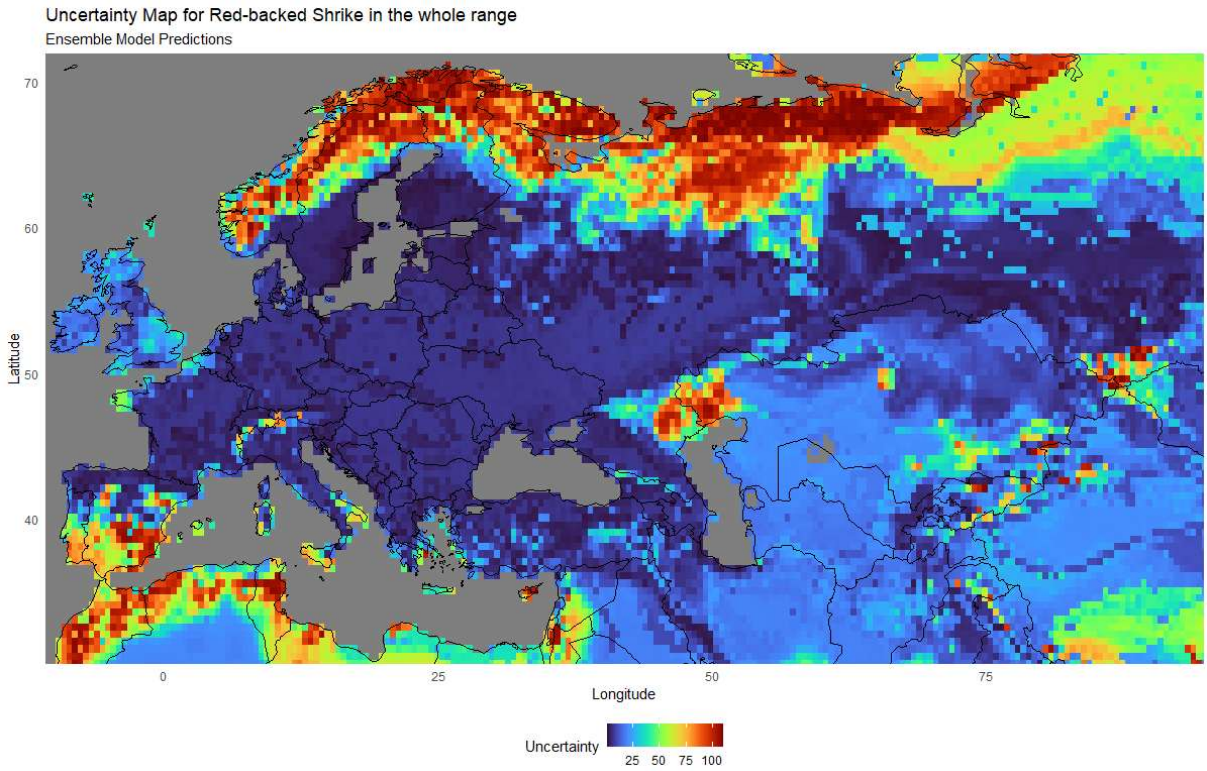

Figure 7.5: Map of coefficient of variation (CV) of ensemble model continuous predictions for current environmental conditions. Uncertainty values represented as percentages. Cold colors represent low uncertainty and warm colors show high variance between single model predictions that were aggregated into the final ensemble model.

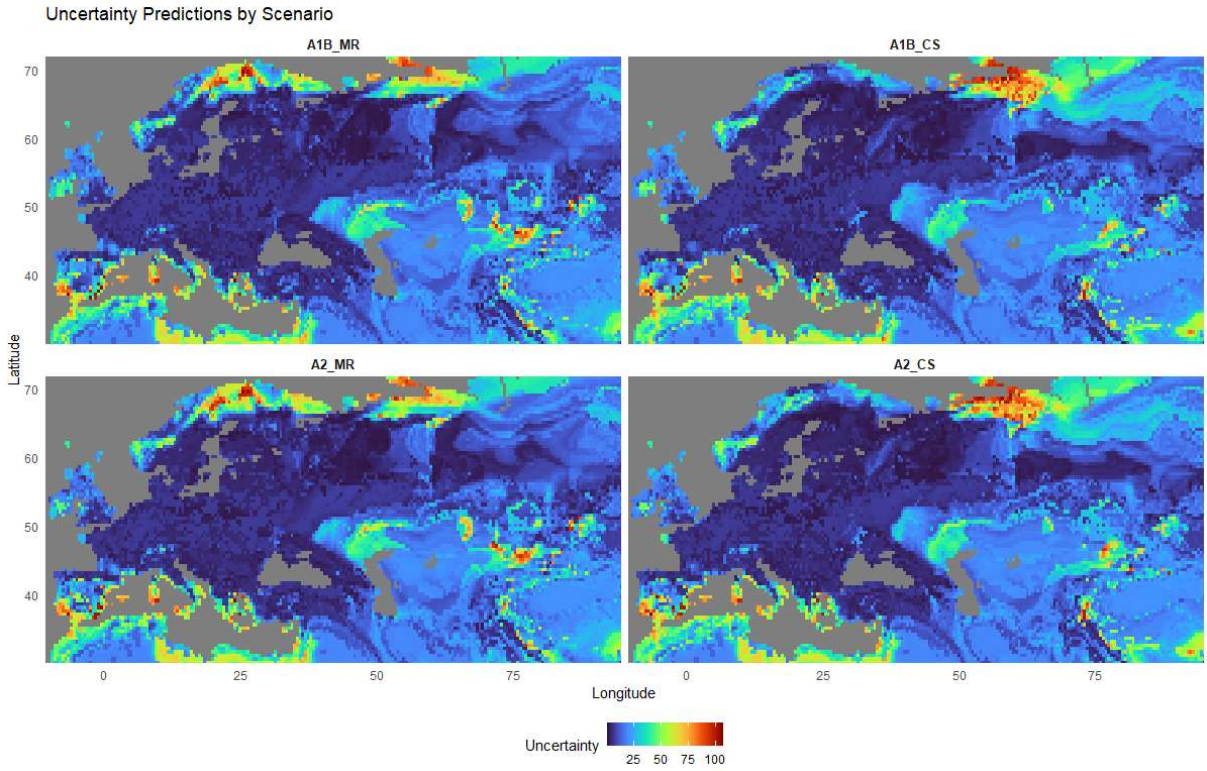

Figure 7.6: Map of coefficient of variation (CV) of ensemble model predictions for future scenarios. Values are represented as percentages. Cold colors represent low uncertainty and warm colors show high variance between model predictions.

Table 7.3: Variable importance values for each modeling algorithm, with 'precsd' included in the set of explanatory covariates.

Model	**tmax**	**tsd**	**prec**	**precsd**	**bias**	**land**
GLM	0.138	0.060	0.146	0.018	0.209	0.043
MaxEnt	0.335	0.201	0.020	0.013	0.312	0.081
GBM	0.122	0.045	0.016	0.002	0.712	0.027
RF	0.193	0.138	0.058	0.068	0.574	0.069

Table 7.4: p-values of McNemar's test for comparing scenarios (A1B and A2) and global circulation models (MR and CS) for the whole range; values below 0.05 indicate distinct predictions between models.

	A1B CS			A2 MR		
	min	median	max	min	median	max
A1B MR	<0.01	0.06	0.84	<0.01	0.72	1
A2 CS	0.01	0.68	1	<0.01	0.04	1

Table 7.5: p-values of McNemar's test for comparing scenarios (A1B and A2) and global circulation models (MR and CS) for the Iberian Peninsula; values below 0.05 indicate distinct predictions between models.

	A1B CS			A2 MR		
	min	median	max	min	median	max
A1B MR	0.04	0.69	1	0.03	1	1
A2 CS	0.18	1	1	0.15	0.82	1

Table 7.6: p-values of McNemar's test for comparing ensemble model with and without 'precsd' in the set of covariates under current conditions, for the whole range and Iberian Peninsula; values below 0.05 indicate distinct predictions between models.

	min	median	max
Whole Range	0.05	0.76	1
Iberian Peninsula	0.07	0.13	1

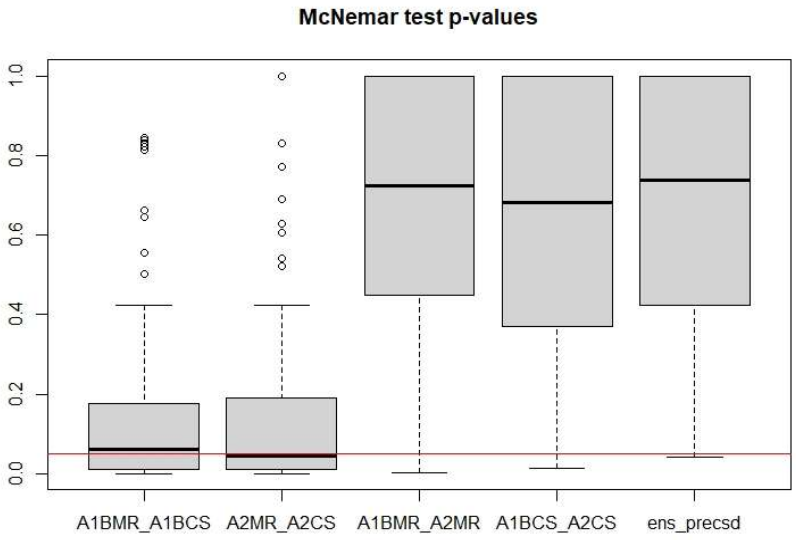

Figure 7.7: Boxplot of McNemar's test (null hypothesis: model predictions are similar) p-values for comparisons between model binary predictions for the whole range, based on 300 randomly sampled locations with 100 repetitions. Red line is set at 0.05 and signals the significance level. [ens_precsd - current prediction ensemble with vs without 'precsd' in the model]

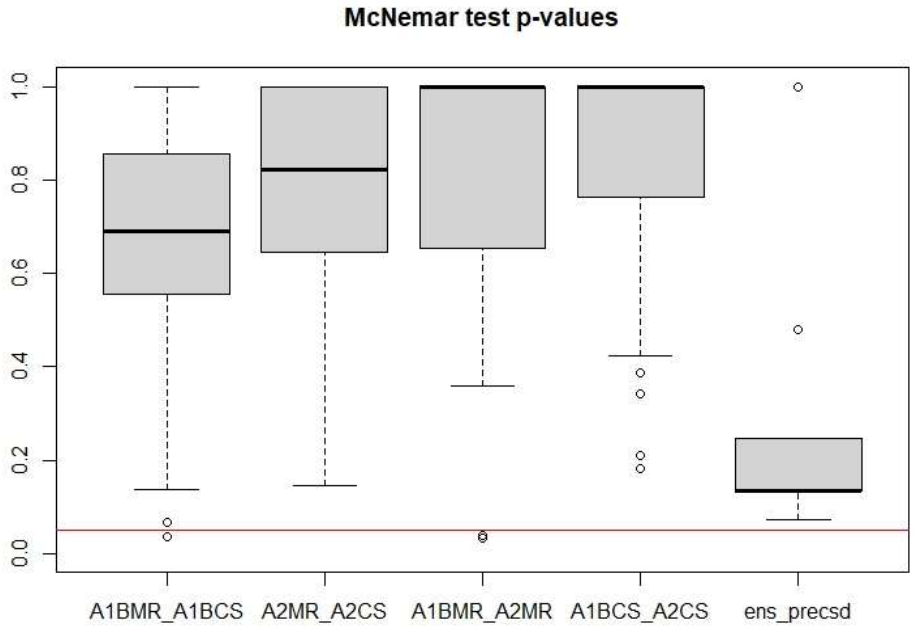

Figure 7.8: McNemar's test (null hypothesis: model predictions are similar) p-values for comparisons between model binary predictions for the Iberian Peninsula, based on 70% of data subsets of randomly sampled locations with 100 repetitions. Red line is set at 0.05 and signals the significance level. [ens_precsd - current prediction ensemble with vs without 'precsd' in the model]

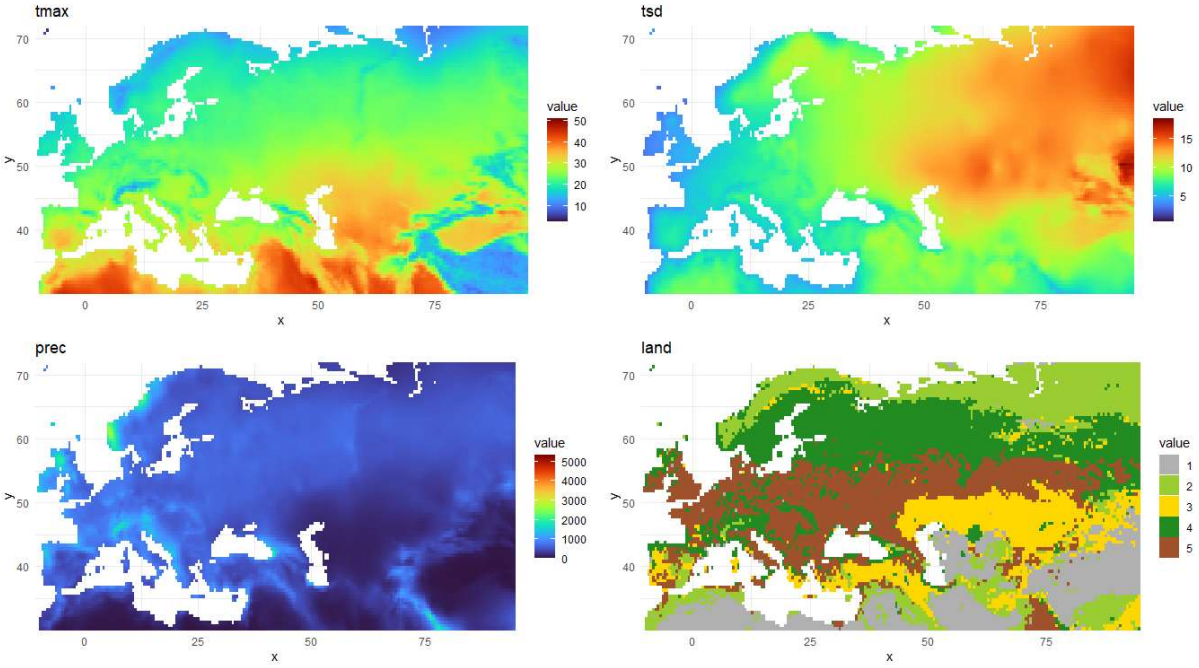

Figure 7.9: Maps of environmental covariates for current conditions.

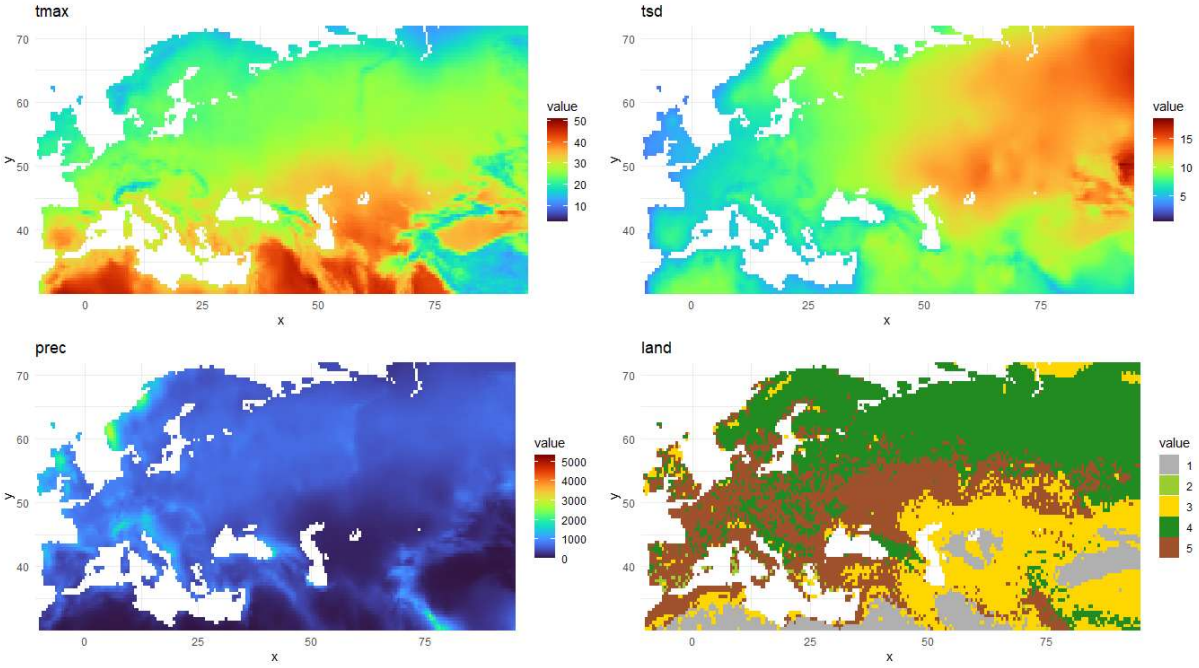

Figure 7.10: Maps of environmental covariates for A1B_MR scenario.

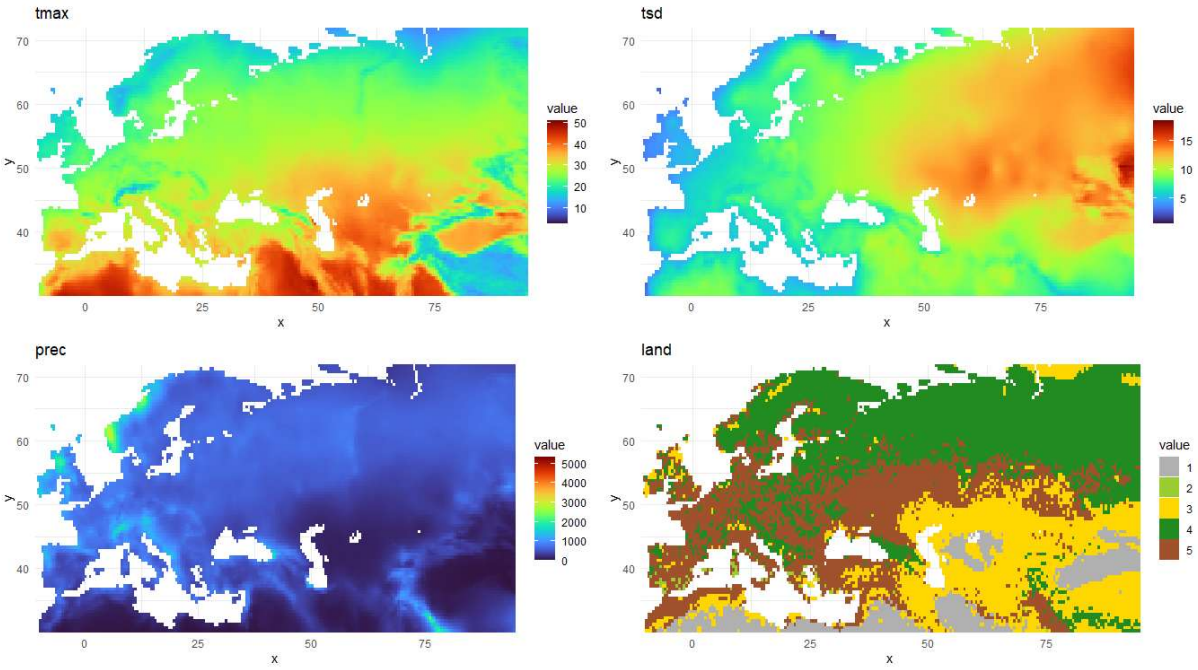

Figure 7.11: Maps of environmental covariates for A1B_CS scenario.

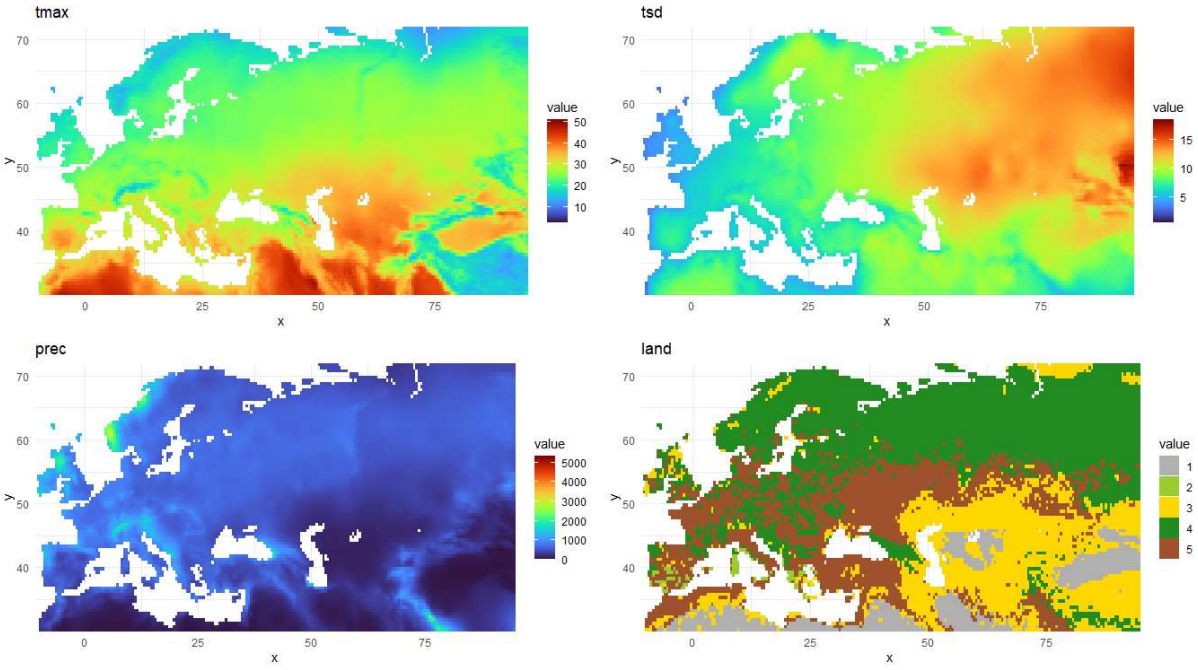

Figure 7.12: Maps of environmental covariates for A2_MR.

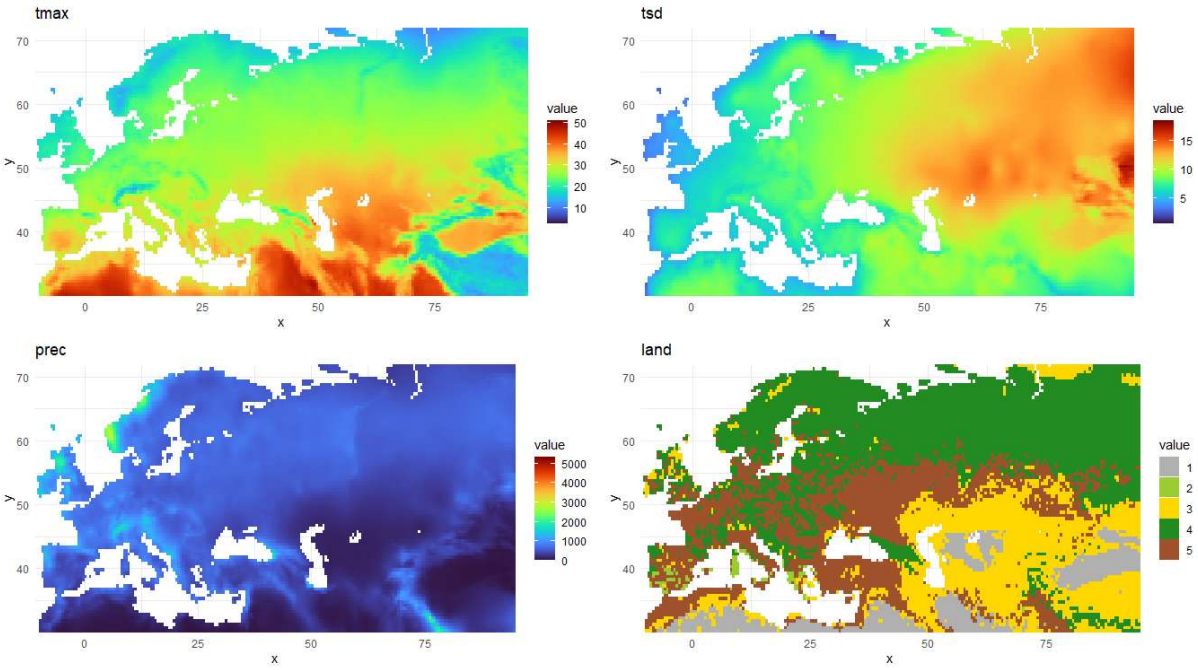

Figure 7.13: Maps of environmental covariates for A2_CS.

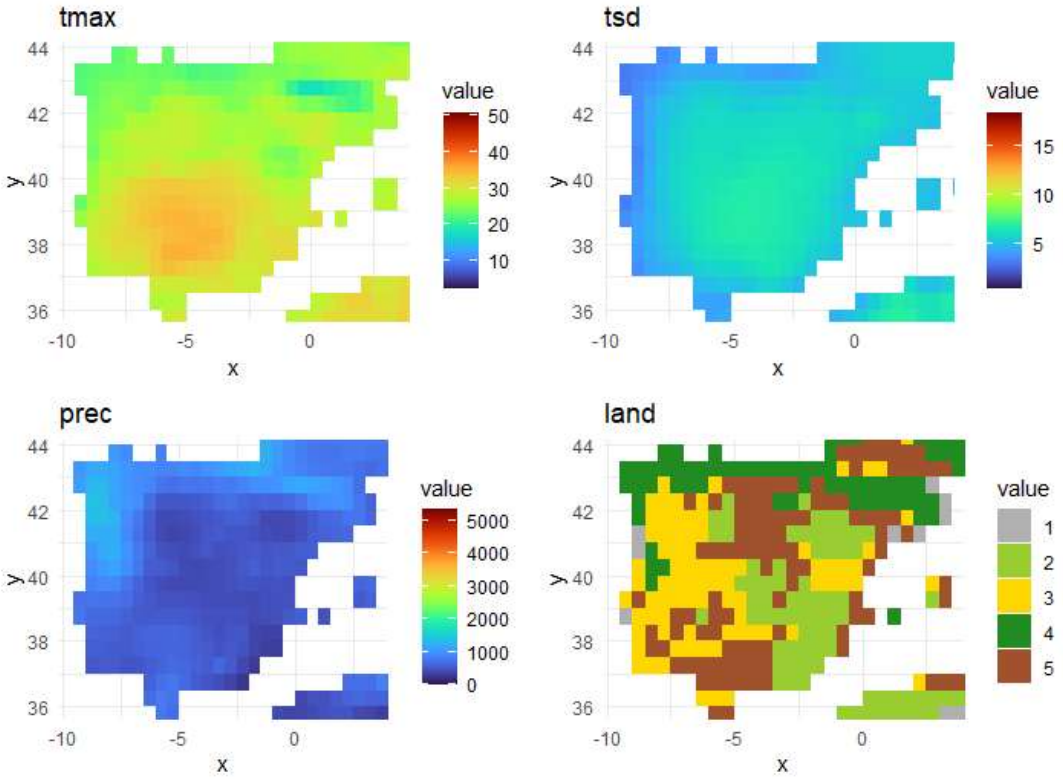

Figure 7.14: Maps of environmental covariates under current conditions in the Iberian Peninsula.

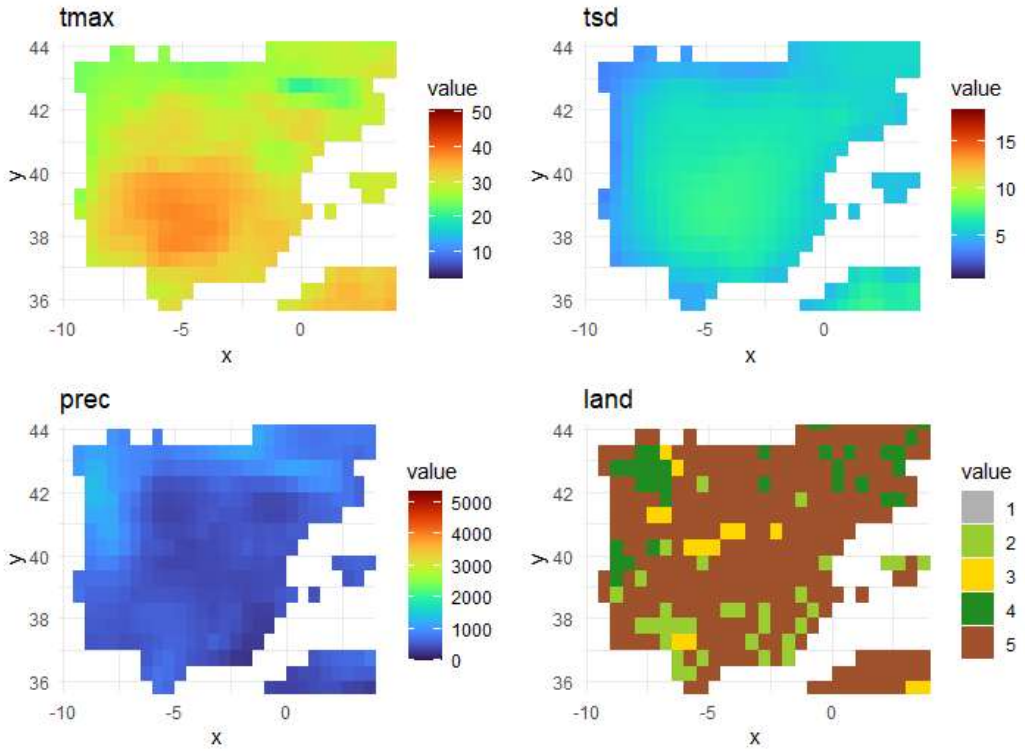

Figure 7.15: Maps of environmental covariates under A1B_MR scenario in the Iberian Peninsula.

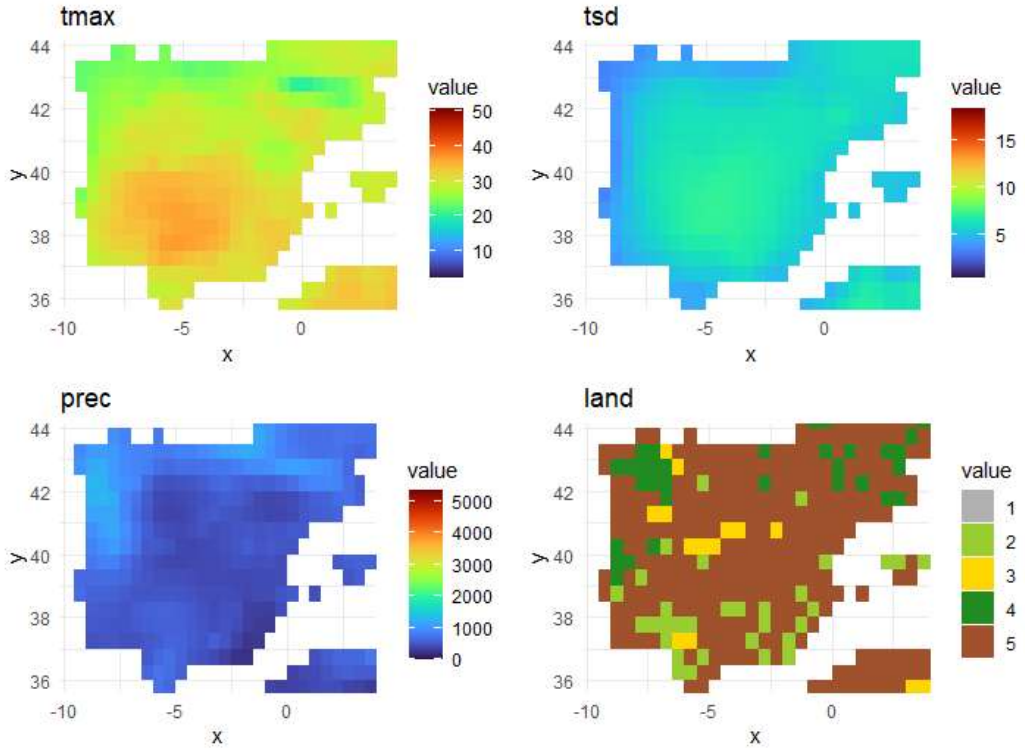

Figure 7.16: Maps of environmental covariates under A1B_CS scenario in the Iberian Peninsula.

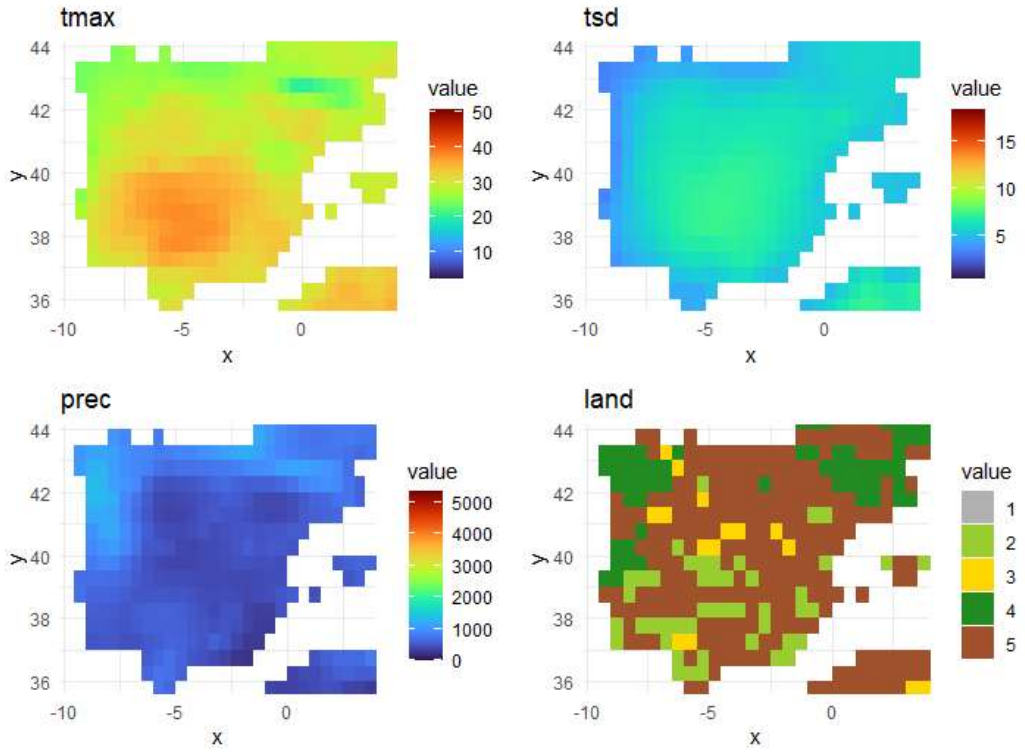

Figure 7.17: Maps of environmental covariates under A2_MR scenario in the Iberian Peninsula.

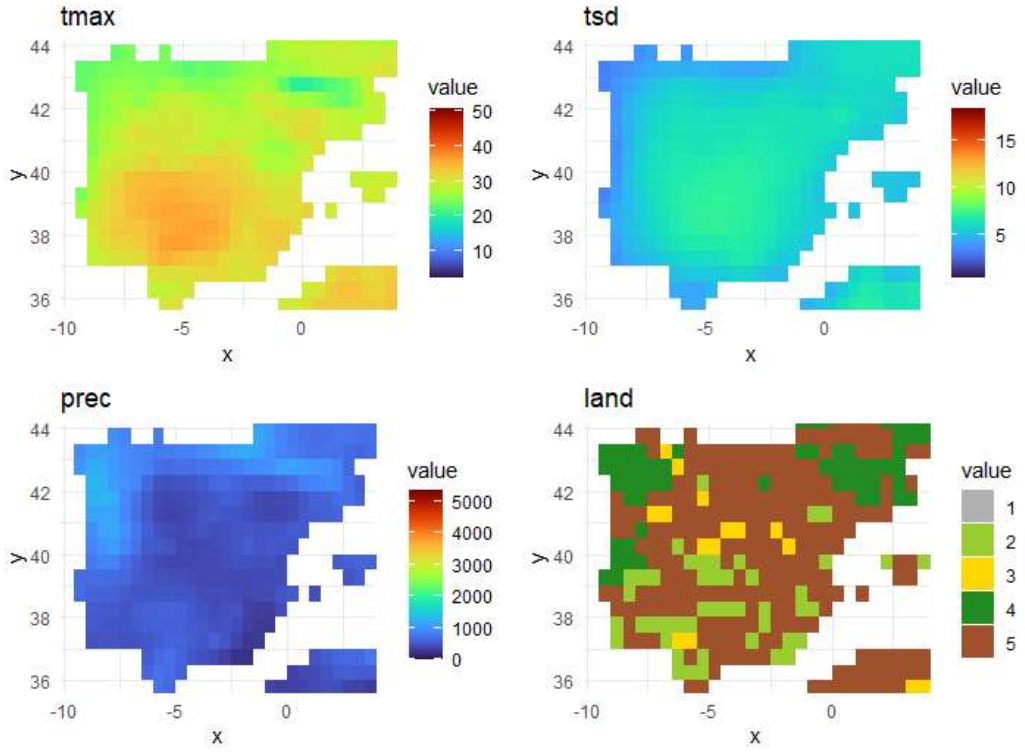

Figure 7.18: Maps of environmental covariates under A2_CS scenario in the Iberian Peninsula.