



**Inland Norway  
University of  
Applied Sciences**

Faculty of Applied Ecology, Agricultural Sciences and Biotechnology

Léa Bataillard

**MSc thesis**

**Using ecological niche modelling to prioritise areas  
for conservation of the critically endangered Buffy-  
Headed marmoset (*Callithrix flaviceps*).**

Master in Applied Ecology

6EV399

**2022**

In collaboration with



Federal University of Viçosa, Minas Gerais, Brazil

## **Abstract**

Endemic to the Atlantic Forest in South-eastern Brazil, the Buffy-Headed marmoset (*Callithrix flaviceps*) has the smallest geographical range of all South American primates. Considered Critically Endangered by the IUCN Red List, this montane primate species is seriously lacking the required attention for effective conservation. However, being the least studied marmoset species, its niche is not well defined. In this thesis, I aimed to model its ecological niche using existing data. Presence-only points of *Callithrix flaviceps* occurrence as well as environmental variables from external databases were included in the modelling, with the main objectives of (1) defining suitable habitat, and (2) prioritizing areas for conservation and/or restoration.

I used Ensemble Modelling to define current suitability considering four climate and two landscape variables. Climate suitability and landscape suitability were modelled first separately, then together to determine overall habitat suitability. To identify areas to prioritize for conservation and/or restoration, I generated predictions for future suitability considering the intermediate (RCP4.5) and extreme (RCP8.5) climate change scenarios for the years 2050 and 2070.

From the six environmental variables used in this study, Precipitation Seasonality (BIO15) and tree canopy cover were the most important to predict *Callithrix flaviceps* distribution. On the other hand, Precipitation of Wettest Month (BIO13) was the least important. Results showed highest probability of occurrence in areas with 70% of Precipitation Seasonality. As expected, tree canopy cover > 80% was positively related with probability of occurrence. Ensemble model output showed that *Callithrix flaviceps* were most likely to occur at elevations above 1200 m and in areas with Temperature Seasonality (BIO4) between 21-23°C. Results were less clear for Annual Mean Temperature (BIO1) and Precipitation of Wettest Month (BIO13), however, probability of presence was higher for areas with Annual Mean Temperature between 10-17°C and around 210 mm of Precipitation of Wettest Month. Overall, present results show that seasonality is important and that *Callithrix flaviceps*, being a montane species, prefer colder environment and higher altitude.

Maps created to show hotspots for future suitability considering the two climate change scenarios suggest that preservation of the Caparaó National Park is needed, as it is likely to be a key area for the preservation of the species. High climatic suitability but low landscape suitability suggests that restoration of habitat in Serra Das Torres might be a useful strategy. Maintenance and restoration around Augusto Ruschii and Goiapaba-Açu protected areas might help to reach a higher overall habitat suitability. However, the Atlantic Forest being highly fragmented, habitat connectivity must be improved.

Results must be interpreted with caution, considering the poor reliability of the occurrence dataset. Moreover, results are limited to the set of variables used, but other variables such as biotic factors and connectivity are likely to affect the *Callithrix flaviceps* distribution. Future studies including species interaction with *C. aurita* and the invasive marmosets *C. penicillata* and *C. jacchus* are needed to get a better picture of *C. flaviceps* realised niche. Given the critical status of the species, it is crucial to benefit from all the data available (even if the quality is not ideal).

## **Contents**

<b>Introduction</b>	1
<b>Methods</b>	4
Study Area	4
Occurrence points	4
Environmental variables	5
Data preparation	7
Modelling	8
a) Preparing data	8
b) Calibrating models	8
c) Variable importance	9
d) Response curves	9
e) Ensemble modelling	9
f) Current and future projections and Ensemble Forecasting	9
<b>Results</b>	10
Climatic suitability	10
a) Individual models	10
b) Response curves	10
c) Ensemble modelling	11
d) Variable importance	11
e) Current and future projections and Ensemble forecasting	12
Landscape suitability	15
a) Individual models	15
b) Response curves	15
c) Ensemble modelling	16
d) Variable importance	16
e) Current and future projections and Ensemble forecasting	17
Overall habitat suitability	18
a) Individual models	18
b) Response curves	18
c) Ensemble modelling	19
d) Variable importance	19
e) Current and future projections and Ensemble forecasting	20
<b>Discussion</b>	23
Current projections	24

Projections under future climate change scenarios	24
Description of the future projections	25
Relevant factors for defining the niche	27
<b>Acknowledgements</b>	29
<b>References</b>	30
<b>Additional materials:</b>	
Part 1: Response curves for climatic suitability models	
Part 2: Response curves for landscape suitability models	
Part 3: Response curves for habitat suitability models	
Part 4: Current projection for suitability models	

## **List of figures**

Figure 1: Study area corresponding to the <i>Callithrix flaviceps</i> geographical range (2020) and the protected areas (UNEP-WCMC).....	4
Figure 2: Number of presence records of <i>Callithrix flaviceps</i> collected between 1964 and 2020 .....	5
Figure 3: Correlation matrix between 8 bioclimatic variables for predicting current distribution of <i>Callithrix flaviceps</i> .....	6
Figure 4: Models scores graph for each algorithm for predicting current species distribution of <i>Callithrix flaviceps</i> from climatic variables.....	10
Table 1: Ensemble model scores for predicting current species distribution of <i>Callithrix flaviceps</i> ....	11
Figure 5: Mean variable importance for each algorithm and ensemble models .....	12
Figure 6: Future (2050) prediction of climatic suitability for <i>Callithrix flaviceps</i> under climate scenario RCP45 .....	13
Figure 7: Future (2070) prediction of climatic suitability for <i>Callithrix flaviceps</i> under climate scenario RCP45.....	13
Figure 8: Future (2050) prediction of climatic suitability for <i>Callithrix flaviceps</i> under climate scenario RCP85 .....	14
Figure 9: Future (2070) prediction of climatic suitability for <i>Callithrix flaviceps</i> under climate scenario RCP85.....	14
Figure 10: Model scores for each algorithm for predicting current species distribution of <i>Callithrix flaviceps</i> from landscape variables .....	15
Table 2: Ensemble models scores .....	16
Figure 11: Mean of variable importance for each algorithm and ensemble models .....	16
Figure 12: Model scores for each algorithm for predicting current species distribution of <i>Callithrix flaviceps</i> from climate and landscape variables .....	18
Table 3: Ensemble model scores.....	19
Figure 13: Mean of variable importance for each algorithm and ensemble models .....	20
Figure 14: Future (2050) prediction of overall habitat suitability for <i>Callithrix flaviceps</i> under climate scenario RCP45.....	21
Figure 15: Future (2070) prediction of overall habitat suitability for <i>Callithrix flaviceps</i> under climate scenario RCP45.....	21
Figure 16: Future (2050) prediction of overall habitat suitability for <i>Callithrix flaviceps</i> under climate scenario RCP85.....	22
Figure 17: Future (2070) prediction of overall habitat suitability for <i>Callithrix flaviceps</i> under climate scenario RCP85.....	22
Figure 18: Projected overall habitat suitability for <i>Callithrix flaviceps</i> , combining both climate and landscape variables for the year 2070 considering the extreme climate change scenario (RCP8.5). ...	26

## **List of figures - Additional materials**

Table 1: Method of data collection for the <i>C. flaviceps</i> occurrence points.....	1
---	---

### Part 1: Response curves for climatic suitability models

Figure 1: Response Curves from generalised linear models (GLM).....	2
Figure 2: Response Curves from generalised additive models (GAM) .....	2
Figure 3: Response Curves from multivariate adaptive regression splines(MARS).....	3
Figure 4: Response Curves from random forest (RF).....	3
Figure 5: Response Curves from generalised boosting models (GBM).....	4
Figure 6: Response Curves from classification tree analysis (CTA).....	4
Figure 7: Response Curves from maximum entropy (MAXENT).....	5

### Part 2: Response curves for landscape suitability models

Figure 8: Response Curves plot from generalised linear models (GLM).....	6
Figure 9: Response Curves plot from generalised additive models (GAM) .....	6
Figure 10: Response Curves plot from multivariate adaptive regression spline (MARS) .....	6
Figure 11: Response Curves plot from random forest (RF) .....	7
Figure 12: Response Curves plot from generalised boosting models (GBM) .....	7
Figure 13: Response Curves plot from classification tree analysis (CTA). .....	7
Figure 14: Response Curves plot from maximum entropy (MAXENT).....	8

### Part 3: Response curves for overall habitat suitability models

Figure 15: Response Curves plot from generalised linear models (GLM).....	9
Figure 16: Response Curves plot from generalised additive models (GAM) .....	9
Figure 17: Response Curves plot from multivariate adaptive regression splines (MARS).....	10
Figure 18: Response Curves plot from random forest (RF) .....	10
Figure 19: Response Curves plot from generalised boosting models (GBM) .....	11
Figure 20: Response Curves plot from classification tree analysis (CTA) .....	11
Figure 21: Response Curves plot from maximum entropy (MAXENT).....	12

### Part 4: Current projections for suitability models

Figure 22: Current prediction of climatic suitability for <i>Callithrix flaviceps</i> .....	13
Figure 23: Current prediction of landscape suitability for <i>Callithrix flaviceps</i> .....	14
Figure 24: Current prediction of overall habitat suitability for <i>Callithrix flaviceps</i> .....	15

## Introduction

Non-human primates (NHP) play essential roles in their ecosystems, including forest and ecosystem health regeneration (Estrada, et al., 2017). However, due to industrialisation and illegal trade of wild animals, they suffer from population decline, and face strong loss and fragmentation of habitat (Estrada, et al., 2017). According to the global risk assessment conducted by Carvalho et al. (2019), comparing worst and best scenario on climate and land use cover (LUC) on primates' species worldwide, the neotropical primates will be suffering the most, regardless of the scenario (Carvalho, et al., 2019). In the neotropics, the Atlantic Forest is home of the highest primate species diversity (Pinto, et al., 1993). With a total area of 1.6 million hectares (Mha) (Muylaert, et al., 2018), it is mainly distributed in Brazil (93%) but also in Paraguay (5.3%) and Argentina (1.7%) (Carlucci, Marcilio-Silva, & Torezan, 2021). The Brazilian Atlantic Forest is highly fragmented, with a large distance between fragments (average 1440 m) (Ribeiro, et al., 2009). The first national park in Brazil, Itatiaia, was created in 1937 between the states of Rio de Janeiro and Minas Gerais. Since then, many parks and protected areas have been implemented in the southern Atlantic Forest (Carlucci, Marcilio-Silva, & Torezan, 2021).

Callitrichidae is a family of 7 genera and 60 species of neotropical monkeys that includes marmosets (*Callithrix* spp.) and tamarins (*Saguinus* spp.) (Rylands, Coimbra-Filho, & Mittermeier, 2009). They are all arboreal (Rylands, Coimbra-Filho, & Mittermeier, 2009) and show a typical parapatric distribution pattern (Braz, Lorini, & Vale, 2019). Indeed, they can coexist with each other, but the species occupy distinct niches and are separated by ecological factors. *Callithrix* have strong ecological plasticity (Rodrigues & Martinez, 2014). Previous findings highlight that the genus *Callithrix* can do rapid niche differentiation after species isolation and suggest that some species can adapt to warmer environments (especially *C. jacchus*, but also *C. geoffroyi* and to a lower degree, *C. kuhlii*) (Braz, Lorini, & Vale, 2019). However, it is important to note that additionally to climate change, the ongoing deforestation might also reduce their ability to adapt to changing climate (Braz, Lorini, & Vale, 2019).

Compared to other NHP, arboreal monkey species have limited dispersal capacities (da Silva, Ribeiro, Hasui, da Costa, & da Cunha, 2015). Indeed, they are more sensitive to habitat fragmentation and climate change, being less likely to find more suitable conditions than NHP that can disperse on the ground (Braz, Lorini, & Vale, 2019); (Rezende, Sobral-Souza, & Culot, 2020). Callitrichidae are great seed dispersers (Ferrari, Corrêa, & Coutinho, 1996; Corrêa, Coutinho, & Ferrari, 2000). The study of Simas et al. (2001) highlights the role of *Callithrix* species in the Caratinga Biological Station, Southeast Brazil, concluding that *C. flaviceps* was the only disperser of *Siparuna guianensis* in this area (Simas, et al., 2001). *Callithrix* spp. have morphological and behavioural adaptations for stimulating the flow of gum on trees (gouging), but, among the six *Callithrix* species, *C. flaviceps* is the least specialised for tree gouging (The IUCN Red List of Threatened Species 2021).

Among the genus *Callithrix*, most research in the past decades has focused on the common marmosets *C. jacchus* and *C. penicillata* (Hannibal, et al., 2019). These species are not originally from the Atlantic Forest (Braz, Lorini, & Vale, 2019). However, they have become invasive in this region, being serious threats for the endemic, montane species *C. aurita* and *C. flaviceps* (Vale, Menini Neto, & Prezoto, 2020). Indeed, hybridisation between common marmosets and *C. aurita* and *C. flaviceps* lead to offspring that have low fitness or are unable to reproduce (Malukiewicz, 2019). The montane species *C. aurita* and *C. flaviceps* are the least studied *Callithrix* species

(Hannibal, et al., 2019), yet the species with the most alarming conservation status (Endangered and Critically Endangered respectively (de Melo, et al., 2021a; de Melo, et al., 2021b)).

Currently, some *C. aurita* individuals are kept in captivity in the Fazenda Macedônia, eastern Minas Gerais, as part of the Mountain Marmoset Conservation Program (MMCP) (Malukiewicz, et al., 2021a). Located next to the famous river Rio Doce (Ipaba, MG), the Fazenda Macedônia has been recognized as Private Reserve of Natural Heritage (RPPN: Reserva Particular do Patrimônio Natural) (Amaral, et al., 2009). Today, the RPPN protects around 3,000 ha (50% of it being covered by native vegetation). The project has recently (November 2021) welcomed 2 new *C. aurita* offspring, born in captivity. In addition, to prevent the spread of common marmosets across *C. flaviceps* and *C. aurita*'s geographical ranges, a programme of sterilisation has been started by the MMCP (Personal communication, Fabiano Melo). These measures are great news for *C. aurita*, however, *C. flaviceps* still seriously needs attention (Hannibal, et al., 2019); (Braz, Lorini, & Vale, 2019); (de Melo, et al., 2021b).

The Buffy-Headed marmoset *Callithrix flaviceps* is a small sized monkey (average weight: 406 g (Rosenberger, 1992)) endemic to the southern Atlantic rainforest of Brazil, mostly between the states of Minas Gerais (MG) and Espírito Santo (ES) (Ferrari, Corrêa, & Coutinho, 1996); (Braz, Lorini, & Vale, 2019). With a geographical distribution of 30 815 km<sup>2</sup> (Malukiewicz, et al., 2021a), *Callithrix flaviceps* has the smallest original geographical range of all South American primates (Mittermeier, Coimbra-Filho, & Constable, 1980); (Mendes & de Melo, 2007); (Hannibal, et al., 2019). It has been considered Endangered for over 30 years (Hannibal, et al., 2019), and is classified as Critically Endangered since 2020 (The IUCN Red List of Threatened Species 2021). Its name on the threatened list is due to (1) intensive deforestation across the states of Minas Gerais and Espírito Santo (Hirsch, et al., 1999), (2) climate change (Townsend, 2008), resulting in dryer and warmer climate within the specie's range (Carlucci, Marcilio-Silva, & Torezan, 2021), (3) yellow fever outbreaks, which have highly affected *Callithrix* populations across the states of ES and MG in 2017 (Mares-Guia, Horta, Romano, & al., 2020), and (4) invasive species, leading to hybridization between *Callithrix* species (Malukiewicz, 2019); (Braz, Lorini, & Vale, 2019); (Moraes, et al., 2019).

*Callithrix flaviceps* usually live in groups of 2 to 15 individuals (Malukiewicz, et al., 2021a); (Trindade Do Carmo, 2022), prefer secondary forests (Ferrari & Mendes, 1991); (Pinto, et al., 1993) and occupy altitudes up to 1200m (a.s.l) or more (Ferrari, Corrêa, & Coutinho, 1996); (Moraes, et al., 2019). Their upper altitudinal range is limited by a significant decline in precipitation in winter (Ferrari, Corrêa, & Coutinho, 1996). Lower precipitation level is known to negatively impact the abundance of some of the *C. flaviceps* food resources (arthropods) (Corrêa, Coutinho, & Ferrari, 2000). *Callithrix* species are sensitive to extreme temperature (Ferrari S. F., 1988), and only a small increase in temperature would cause strong damages to all marmoset species, including *Callithrix flaviceps* (Braz, Lorini, & Vale, 2019); (Carvalho, et al., 2019).

Due to the expected impact of climate change on the distribution of species (Pearson & Dawson, 2003); (Serra-Diaz & Franklin, 2019); (Braz, Lorini, & Vale, 2019), it is urgent to implement relevant conservation areas, i.e. areas that will remain suitable considering future climate change scenarios, to increase the chances of long-term species survival (Mendes & Pereira, 2015). In their study from 2020, Rezende et al., suggest that prioritized areas for conservation of another endangered arboreal primate (the black lion tamarin *Leontopithecus chrysopygus*) can be both (1) the ones with high current climate suitability but low landscape suitability (because habitat



management and restoration of landscape is possible), and (2) low current climate but high landscape suitability (because it highlights areas to be considered in light of future climate change scenarios) (Rezende, Sobral-Souza, & Culot, 2020). They conclude that niche modelling based on both climatic and landscape variables are useful tools for conservation strategies and can provide realistic habitat suitability predictions (Rezende, Sobral-Souza, & Culot, 2020).

Protecting the remaining habitat as much as possible is the key management response to habitat loss (Townsend, 2008). However, what defines a suitable habitat for *Callithrix flaviceps* is not well defined yet. In this study, I aimed to map habitat suitability of *C. flaviceps* by modelling the species' ecological niche (including climate and landscape variables) within its geographical range, and thus, investigate how environmental variables influence habitat suitability. In addition, I used future climate scenarios to identify likely future suitable habitat. The following research questions were addressed:

1. Considering climate and landscape variables, what defines a suitable niche for *Callithrix flaviceps*?
2. How will climate change affect the climatic suitability in the next five decades within the *Callithrix flaviceps* geographical range?
3. Which area(s) should be prioritised for conservation of the *Callithrix flaviceps* population within its geographical range?
4. Where should we focus future restoration effort within the *Callithrix flaviceps* geographical range?

Based on presence-only points collected within the species' geographical range I modelled the *current* Buffy-Headed marmoset (*Callithrix flaviceps*) niche, including four climate and two landscape variables (**Research question1**). Predictions using two Representative Concentration Pathways (RCPs) scenarios were done for the years 2050 and 2070 to see how climate change would affect the climatic suitability (**Research question2**). Maps were then created to show area(s) to prioritize for conservation (**Research question3**) and restoration (**Research question4**).

## Methods

### Study Area

The study area corresponds to the most recent version of the IUCN polygon representing *Callithrix flaviceps* geographical range (AURICEPS WORKSHOP May 2020). The whole geographical range mostly sits between the states of Minas Gerais (MG), Espírito Santo (ES), and for a small part, in Rio de Janeiro (RJ) (Figure 1). The tropical climate of the states of Minas Gerais and Espírito Santo offers a yearly mean temperature of 23.9°C and 25°C respectively. Annual maximum temperature in MG is 26.9°C, while annual minimum temperature is 9.5°C. Espírito Santo seems to offer less variation in temperature, with an annual maximum temperature of 28.0°C and annual minimum temperature of 21.0°C. Being in the Atlantic Forest, the geographical range of *Callithrix flaviceps* offers humid and seasonal forest. Dry season is typically from May to August (August being the driest month), while rainy season occurs in the summer months from October to March (with a peak around December) (Weather and Climate, 2022).

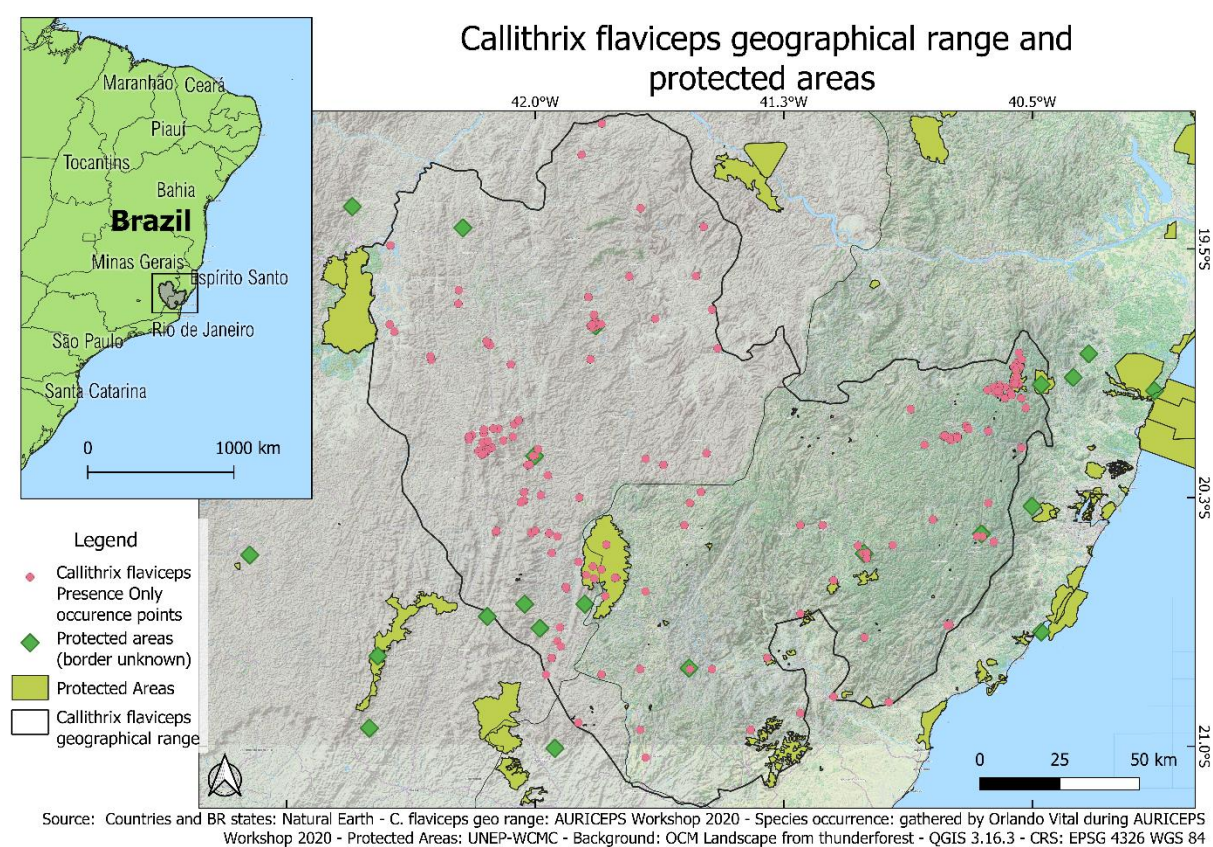


Figure 1: Study area corresponding to the *Callithrix flaviceps* geographical range (2020) and the protected areas (UNEP-WCMC).

### Occurrence points

Ecological niche modelling is a type of spatial analysis that requires two types of data: (1) the occurrence points of the species and (2) environmental data (Merow, Smith, & Silander Jr, 2013). The database of occurrence points was created in May 2020, during the AURICEPS WORKSHOP (by Orlando Vital from the Federal University of Viçosa (UFV) and MMCP, Brazil). This work gathered over 4800 presence-only observations points of all 6 *Callithrix* species found in Brazil (*C. penicillata*, *C. jacchus*, *C. geoffroyi*, *C. khulii*, *C. aurita*, *C. flaviceps*, and *C. sp* hybrids). Data from different sources were compiled, with a focus on taxon, location, and year of occurrence.

For this study, I focused only on *C. flaviceps*. Therefore, 272 presence-only points from published and unpublished records were used. These *C. flaviceps* occurrences were collected during the last 60 years (1964-2020) within the whole geographical range (study area).

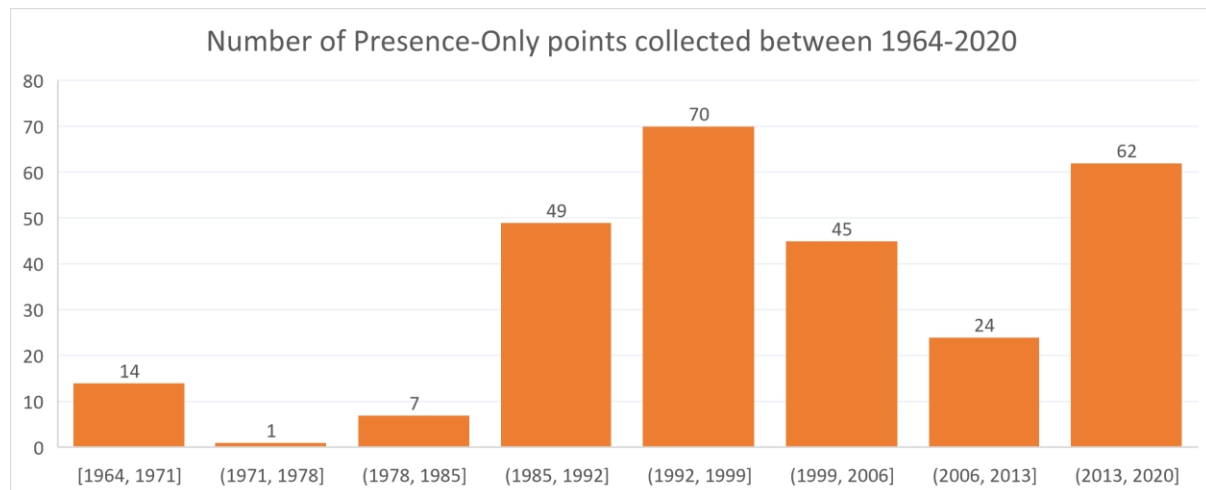
Published data came from:

- "Atlantic Primates: a dataset of communities and occurrences of primates in the Atlantic Forests of South America" (Culot, et al., 2019),
- "A Survey of Wild and Introduced Marmosets (*Callithrix: Callitrichidae*) in the Southern and Eastern Portions of the State of Minas Gerais, Brazil 2018." (Silva, et al., 2018),
- "Predicting the potential hybridization zones between native and invasive marmosets within Neotropical biodiversity hotspots" (Moraes, et al., 2019)

Unpublished data came from the researchers:

- Maria Cecilia Kierulff,
- Carla Possamai,
- Mariane Kaizer,
- Jorge Guimarães,
- Natan Perico
- Fernanda Silva.

Data were collected unevenly throughout the 60 years (*Figure 2*) and gathered from different sources. Method for data collection was not always provided. In this study, method was Not Applicable for 162 points out of 272. For the rest, presence was recorded by observations (sighting) in the wild (n = 73), line transect (n = 14), citizen science surveys (n = 6), museum (n = 5), personal records (n = 5), camera trap in canopy (n = 4), and interviews (n = 3) (*Table 1 of Additional materials*).



*Figure 2: Number of presence records of Callithrix flaviceps collected between 1964 and 2020.*

### Environmental variables

To be more accurate, modelling a niche must be multidimensional (Townsend, 2008). Therefore, a set of environmental variables (climate + landscape) was selected.

## Climate variables

I used current bioclimatic variables from the external database WorldClim v2.1 ([www.worldclim.org](http://www.worldclim.org)) at 30 arc seconds per pixel, referred to as 1km spatial resolution.

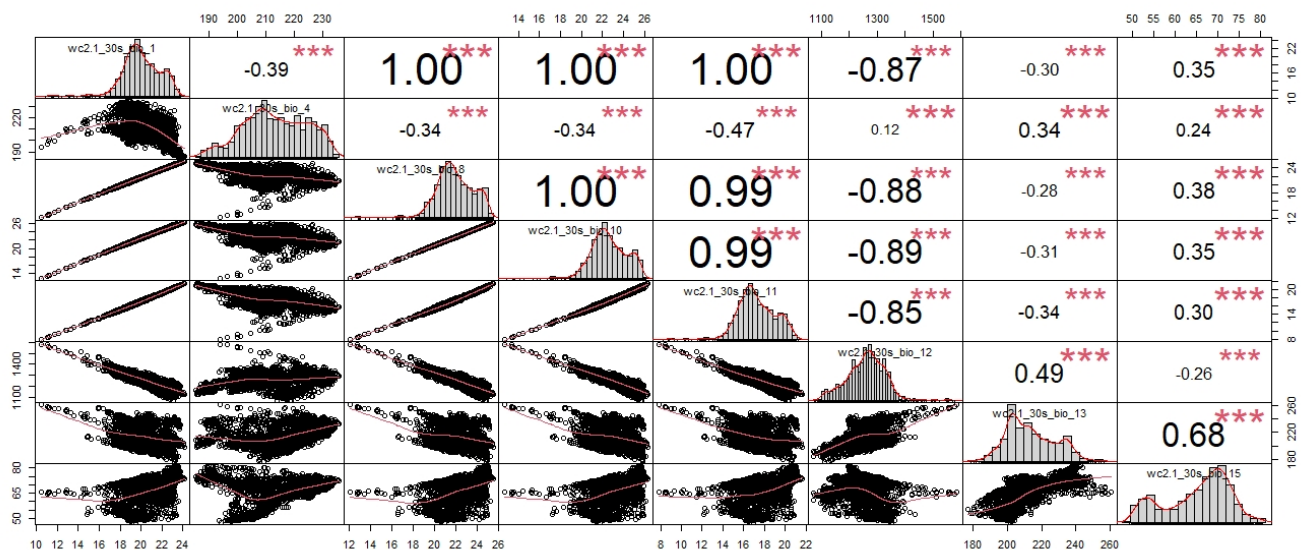
WorldClim provides layers for *current* (1970-2000) and *future* (2050 and 2070) climate data (Fick & Hijmans, 2017). More information on how these variables are derived is available on [www.worldclim.org](http://www.worldclim.org).

Variables were pre-selected based on the ecology of the Brazilian marmosets. *Callithrix flaviceps* are sensitive to extreme temperature (Ferrari, 1988); (Braz, Lorini, & Vale, 2019). Furthermore, precipitation level is known to impact the abundance of food (arthropods) (Corrêa, Coutinho, & Ferrari, 2000), and have an impact on the species distribution. Finally, *Callithrix* species seem to prefer seasonal environment (Corrêa, Coutinho, & Ferrari, 2000); (Braz, Lorini, & Vale, 2019). Based on this information, the following 8 variables were pre-selected:

- BIO1 = Annual Mean Temperature
- BIO4 = Temperature Seasonality (standard deviation ×100)
- BIO8 = Mean Temperature Wettest Quarter
- BIO10 = Mean Temperature of Warmest Quarter
- BIO11 = Mean Temperature of Coldest Quarter
- BIO12 = Annual Precipitation
- BIO13 = Precipitation of Wettest Month
- BIO15 = Precipitation Seasonality (Coefficient of Variation)

Climatic variables are expected to be correlated; hence, a correlation matrix on current data was run using the R package *PerformanceAnalytics* v2.0.4 (Peterson, et al., 2020).

Based on the correlation plot (*Figure 3*) and on study design from previous studies on Brazilian marmoset, I selected 4 climate variables.



*Figure 3: Correlation matrix between 8 bioclimatic variables for predicting current distribution of Callithrix flaviceps. wc2.1\_30s\_bio\_1: Annual Mean Temperature, wc2.1\_30s\_bio\_4: Temperature Seasonality (standard deviation ×100), wc2.1\_30s\_bio\_8: Mean Temperature Wettest Quarter, wc2.1\_30s\_bio\_10: Mean Temperature of Warmest Quarter, wc2.1\_30s\_bio\_11: Mean Temperature of Coldest Quarter,*

*wc2.1\_30s\_bio\_12: Annual Precipitation, wc2.1\_30s\_bio\_13: Precipitation of Wettest Month, wc2.1\_30s\_bio\_15: Precipitation Seasonality (Coefficient of Variation).*

Decision was taken to select Annual Mean Temperature (BIO1), Temperature Seasonality (BIO4), Precipitation of Wettest Month (BIO13) and Precipitation Seasonality (BIO15).

The same set of climatic variables (n=4) was used for the projections of future climate change scenarios. Data from 2050 (average of the years 2041-2060) and 2070 (average of the years 2061 and 2080) were selected.

Different Global Climate Models (GCM) and Representative Concentration Pathways (RCP) combinations are available in WorldClim (IPCC, 2022). GCMs mathematically represent major climate system components such as atmosphere, land surface, ocean, etc., and their interactions. The RCPs represent greenhouse gas concentration trajectories (IPCC, 2022). The four RCPs available in WorldClim (RCP2.6, RCP6.0, RCP4.5 and RCP8.5) represent different climate change scenarios, from the most optimistic (RCP2.6) (now considered impossible (IPCC, 2022)), to the most pessimistic scenario (RCP8.5).

For simplicity, only 1 GCM model and 2 RCP scenarios were considered in this study. Data from the GCM model HADGEM2-ES (as supported by Yin et al. (2013) (Yin, Fu, Shevliakova, & Dickinson, 2013)) was coupled with the "intermediate" (RCP4.5) and "extreme" (RCP8.5) RCP scenarios (IPCC, 2022). In the extreme scenario (RCP8.5), emissions keep rising throughout the 21<sup>st</sup> century and stabilize around 2050. Emissions in the intermediate scenario (RCP4.5) keep rising until 2040 but then drop (IPCC, 2022).

The Coupled Model Intercomparison Project Phase (CMIP) represents a collaborative framework, where different research groups compare multiple model simulations, with the objective to improve climate models (IPCC, 2022). In this study, high resolution data (30 arc-seconds per pixel) from the Coupled Model Intercomparison Project Phase (CMIP5) from WorldClim *version 1.4* ([www.worldclim.org](http://www.worldclim.org)) were used for projected climate variables.

### **Landscape variables**

Landscape variables included a Digital Elevation Model (hereafter DEM), at 3 arc-second per pixel (approximately 90 m spatial resolution) from EarthEnv ([www.earthenv.org](http://www.earthenv.org)) (Robinson, Regetz, & Guralnick, 2014) and Tree Canopy Cover for the year 2000 (*treecover2000*), at 1 arc-second per pixel (approximately 30m spatial resolution) from Hansen Global Forest Change v1.7 ([www.earthenginepartners.appspot.com](http://www.earthenginepartners.appspot.com)). The *treecover2000* raster layer shows a percentage per grid cell of vegetation taller than 5 m (Hansen, et al., 2013).

### **Data preparation**

Data preparation was done in QGIS (v3.16.3). All raster files, i.e., 6 corresponding to *current* environmental variables, and 16 (4 bioclimatic variables x 2 years x 2 RCPs) corresponding to *future* climate data, were clipped to the study area and exported as *.asc* files for further use in R. All variables were set to the *EPSG 4326 WGS84* geographical coordinate system.

*Callithrix flaviceps* observations (n = 272), all from 1964-2020, were analysed with the 1970-2000 WorldClim data (four climate variables), as well as the elevation (DEM) and the tree canopy cover percentage for the year 2000 (two landscape variables).

## Modelling

Although individual modelling can perform better in many cases (Marmion, et al., 2009), Ensemble Modelling where a number of different algorithms or training datasets are combined and result in one final prediction (Kotu & Deshpande, 2014) is useful when dealing with uncertainties of extrapolation (Marmion, et al., 2009) and to avoid the “choosing the best model” bias (Guisan, Thuiller, & Zimmermann, 2017). In this study, I used the *biomod2* package v3.5.1 (Thuiller, Georges, Gueguen, Engler, & Breiner, 2021) in the software R (R Core Team, 2019; RStudio Team, 2020) to generate ensemble modelling.

As recommended by De Kort et al. (2020), models were built first separately, based on climate and landscape niche, respectively, then together (De Kort, et al., 2020). Based on this approach, 3 maps of *current* suitability (climate suitability, landscape suitability, and overall habitat suitability), and 4 maps for *future* climate suitability (2 RCP scenarios x 2 years) were created. To help visualising future areas of both high climatic and high landscape suitability, I generated 4 extra maps, each combining one of the 4 future climate scenarios with the (current) landscape variables.

Note that, in this study, Ecological Niche Modelling focuses only on abiotic variables. Biotic factors such as species interactions and potential diseases outbreak were not included in the models.

### a) Preparing data

All algorithms available in *biomod2* need presence/absence data, except Maxent (maximum entropy) that uses presence data and a selection of background points. Therefore, 2000 pseudo-absence points were randomly generated within the study area. As recommended by Guisan et al. (2017), pseudo-absence selection was repeated 3 times to avoid sampling bias, resulting in 3 different datasets of 2000 pseudo-absence points each (Guisan, Thuiller, & Zimmermann, 2017).

### b) Calibrating models

Before ensemble modelling, single models were generated. Seven algorithms were run and evaluated. Algorithms included three statistical regression methods, i.e., generalised linear models (GLM) (Guisan, Edwards Jr, & Hastie, 2002), generalised additive models (GAM) (Guisan, Edwards Jr, & Hastie, 2002) and multivariate adaptive regression splines (MARS) (Friedman, 1991), and four machine learning models, i.e., random forest (RF) (Breiman, 2001), generalised boosting models (GBM/BRT) (Elith, Leathwick, & Hastie, 2008), classification tree analysis (CTA) (Vayssières, Plant, & Allen-Diaz, 2000), and maximum entropy (MAXENT) (Phillips, Anderson, & Schapire, 2006).

For all variables, the quadratic terms and two-way interactions were used in GLMs. The number of tree limits in RFs, and GBMs were kept as default, at 500 and 2500 respectively (refer to Rdocumentation on *biomod2* package for more details on default settings). MAXENT's maximum iterations were set to 1000.

Independent data were not available to evaluate the models, thus, models were calibrated on 80% of the data (training set), and 20% (validation set) was kept for model testing. This data-split procedure (cross-validation) was repeated 3 times.

I used two different metrics to evaluate each model. The relative operative characteristic (ROC) and the true skill statistics (TSS). ROC is a widely used and suitable measure of performance. Being threshold-independent, it cannot be done for presence-absence predictions (Allouche, Tsoar, &

Kadmon, 2006). TSS is a simple and accurate threshold-dependent measure of performance for predictions expressed as presence-absence, representing  $Sensitivity + Specificity - 1$  (Allouche, Tsoar, & Kadmon, 2006). Sensitivity represents the probability that a model correctly predicts the presence of a species, while specificity represents the probability that a model correctly predicts the absence of a species (Allouche, Tsoar, & Kadmon, 2006).

To assess visually the parameters used to build the models, 3 graphs focusing on (1) choice of algorithms, (2) the cross validation run, and (3) the pseudo-absence sampling procedure were created according to the evaluation metric used (TSS and ROC).

#### c) *Variable importance*

The variable importance was checked for each of the models. To make results more readable and to get the variable importance per algorithm, results were averaged among the cross-validation procedures (runs) and the pseudo-absences selections. I set the number of permutations for variable importance to 7.

#### d) *Response curves*

Once the variable importance was known, I investigated how each of them was related to the species' probability of presence. Therefore, the evaluation strip procedure proposed by Elith et al. (2005) from the R package *biomod2* (Thuiller, et al., 2021) was used to plot response curves for each variable, for each algorithm. Note that each response curve is modelled considering all other variables included in the model being held constant (Guisan, Thuiller, & Zimmermann, 2017).

#### e) *Ensemble modelling*

To avoid "selection of the best model" bias and to optimise accuracy, ensemble modelling was done based on a TSS threshold of 0.7, meaning that only models with a  $TSS \geq 0.7$  were kept for the final ensemble models. Final ensemble models were then evaluated using the same metrics as the individual models (ROC and TSS).

Five "ensembling" options were considered in the `BIOMOD_EnsembleModeling()` function: the *mean*, giving the mean probabilities across predictions; the *weighted mean* (*wmean*), estimating the weighted sum of probabilities; the *committee averaging* (*ca*), giving both a prediction and a measure of uncertainty, and the *confidence interval* (95%) (*ci*), showing 2 estimations (the lower one: *ci inf* and the upper one: *ci sup*) of the confidence interval around the *mean* probability.

Variable importance was also generated for the Ensemble Models.

#### f) *Current and future projections and Ensemble Forecasting*

After generating 5 ensemble models (mean, weighted mean, committee averaging, and inferior and superior confidence intervals) in the steps outlined above, I focused on *current* and *future* spatial distribution. The ensemble of models for current distribution built under the weighted mean rule was used to generate the projections. The functions `BIOMOD_Projection()` and `BIOMOD_EnsembleForecasting()` were used, for *single algorithms models* and *ensemble models* projections respectively.

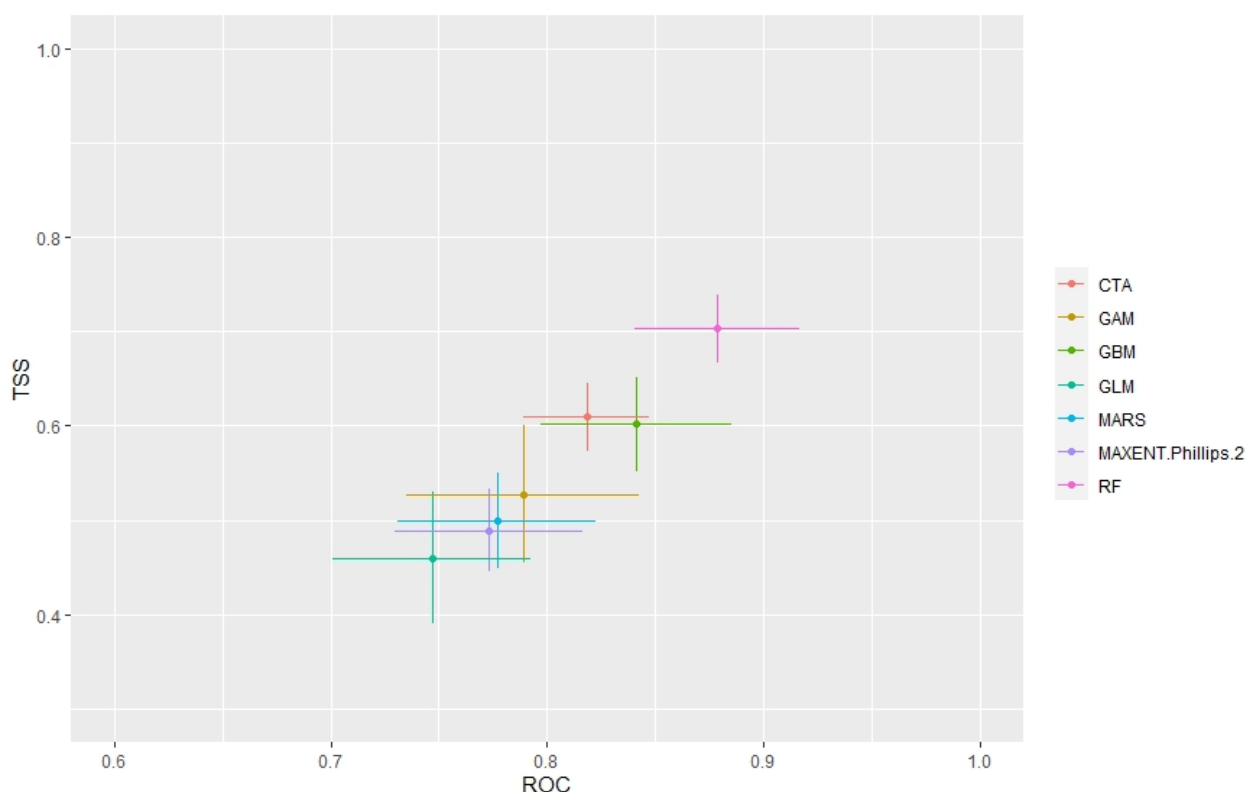
To find the areas of greatest habitat suitability for future occurrence of the species, I ran this step for climate suitability in the years 2050 and 2070, each based on current projections and trained models (from the *Ensemble modelling* step) with 2 climate change scenarios (RCP4.5 & RCP8.5).

## Results

### Climatic suitability

#### a) Individual models

Current climatic suitability modelling generated 63 models (3 pseudo-absence selections x 3 cross-validation procedures (runs) x 7 algorithms). *Figure 4* shows the range of TSS and ROC scores for each algorithm. On average, models from random forest (RF), classification tree analysis (CTA) and generalised boosting model (GBM) stand out with the highest performance. With a ROC > 0.8, the accuracy of the models generated with RF, CTA and GBM is considered “good” (Araújo et al., 2005). There are differences between algorithms, however, all seven gave “fair” predictive accuracy (ROC > 0.7) (Araújo et al., 2005). All TSS scores are considered “fair” (TSS > 0.4) or “good” (0.6 < TSS < 0.8) (Thuiller, Araujo, & Lafourcade, 2010).



*Figure 4: Models scores graph for each algorithm for predicting current species distribution of Callithrix flaviceps from climatic variables (BIO1, BIO4, BIO13, BIO15) according to the TSS (True Skill Statistics) and the ROC (Relative Operative Characteristic). Algorithms used were CTA: Classification Tree Analysis; GAM: Generalised Additive Models; GBM: Generalised Boosting Models, GLM: Generalised Linear Models; MARS: Multivariate Adaptive Regression Splines, MAXENT: Maximum Entropy and RF: Random Forest. Error bars represent the range of the different models of the same algorithm, central point corresponds to average between models.*

#### b) Response curves

Response curves for each algorithm are shown in *Figures 1-7 of Additional materials*. All models suggest that *Callithrix flaviceps* are found in areas where Annual Mean Temperature (BIO1) is below 17°C, except the GLMs and MAXENTs that show a more linear negative effect. Some models suggest a higher probability of presence above 23°C. Regarding Precipitation Seasonality (BIO15), all algorithms except GLMs show a strong peak at 70%. Additionally, most models (from GAM, MARS, RF, GBM, MAXENT and some models from GLM) show another peak around 50-55% of Precipitation Seasonality (*Additional materials, Figures 1-7*).



More variation among algorithms was found for Temperature Seasonality (BIO4). Models from the machine learning methods (RF, GBM, CTA and MAXENT) did not show much variation. However, according to GAMs, ideal Temperature Seasonality sits around 21-23°C. Similarly, some MARS models showed a peak around 21°C, other models showed a peak around 22.5°C. Interestingly, 3 models out of 9 generated by CTA found a peak, either at 21.2, 21.8, or 22.8°C.

Comparably to Temperature Seasonality, Precipitation of Wettest Month (BIO13) also show variations among algorithms. All GAM models showed a peak at 210 mm and another around 245 mm. Models from GBM suggested a higher probability around 210 mm as well, but with almost no variation. Some models from MARS and MAXENT showed higher probability of occurrence below 190 mm (and between 200 and 220 mm. GLMs suggested a linear negative relationship between Precipitation of Wettest Month and probability of presence. Some models from MARS also showed higher suitability from 200 to 220mm. And with the exception of one model, CTA did not show any pattern nor variation. Random Forest (RF) showed almost no variation but suggested higher relative probability of presence above 250 mm.

#### c) Ensemble modelling

Out of the 63 models created for current climatic suitability, 5 had TSS  $\geq 0.7$  and were kept for ensemble modelling: All of these were Random Forest (RF) models: "PA1\_RUN1\_RF", "PA1\_RUN2\_RF", "PA1\_RUN3\_RF", "PA2\_RUN3\_RF", "PA3\_RUN1\_RF" and "PA3\_RUN3\_RF".

Table 1 shows the 5 ensemble model scores. The weighted mean (*wmean*) ensemble model output was kept for running projections.

*Table 1: Ensemble model scores for predicting current species distribution of Callithrix flaviceps. The mean gives the mean probabilities across predictions; the weighted mean (wmean) estimates the weighted sum of probabilities; the committee averaging (ca) gives both a prediction and a measure of uncertainty, and the confidence interval (ci) shows 2 estimations (ci inf and ci sup) of the confidence interval around the mean probability. ROC: Relative Operative Characteristic, TSS: True Skill Statistics.*

	<b>ROC</b> TESTING DATA	<b>TSS</b> TESTING DATA	<b>SENSITIVITY</b> (TSS)	<b>SPECIFICITY</b> (TSS)
<b>MEAN</b>	0.994	0.952	98.897	96.335
<b>WMEAN</b>	<b>0.994</b>	<b>0.952</b>	<b>98.897</b>	<b>96.335</b>
<b>CA</b>	0.991	0.949	98.897	95.981
<b>CI INF</b>	0.988	0.921	95.956	96.140
<b>CI SUP</b>	0.993	0.961	100	96.052

#### d) Variable importance

The variable importance was run for each algorithm as well as for the 5 ensemble models (Figure 5). All 7 algorithms considered Precipitation Seasonality (BIO15) the most important variable to predict the *Callithrix flaviceps* distribution. With the exception of one algorithm (MAXENT), all algorithms agreed on Temperature Seasonality (BIO4) being the second most important climate variable. Overall, Precipitation of Wettest Month (BIO13) and Annual Mean Temperature (BIO1) were considered the least important variables for all algorithms except MAXENT, that considers Temperature Seasonality the least important variable to predict the species distribution.

Unsurprisingly (because all models kept for EM were from random forest), the variable importance given by the ensemble models concurs with the RF individual algorithm average. Precipitation Seasonality (BIO15) clearly stands out as the most important variable, followed by Temperature Seasonality (BIO4) and Annual Mean Temperature (BIO1) (Figure 5). Among the 4 climate variables used in this study, Precipitation of Wettest Month (BIO13) was the least important variable to predict the species distribution.

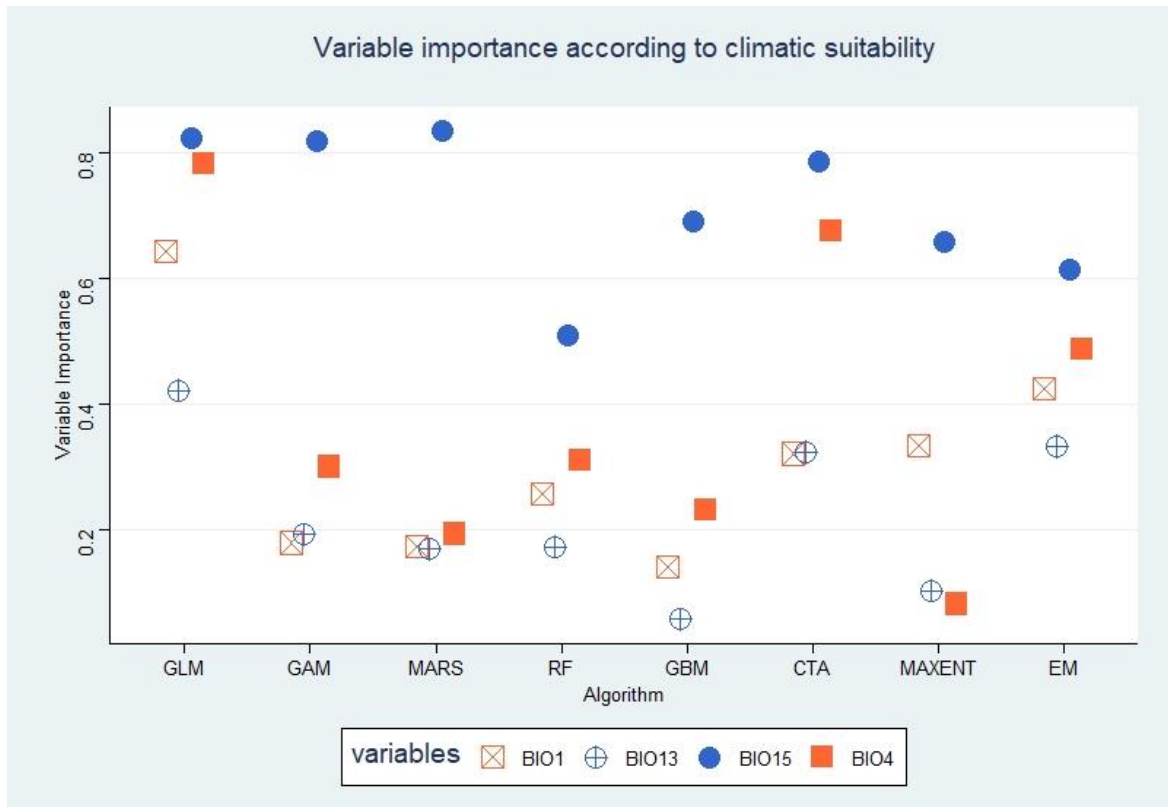


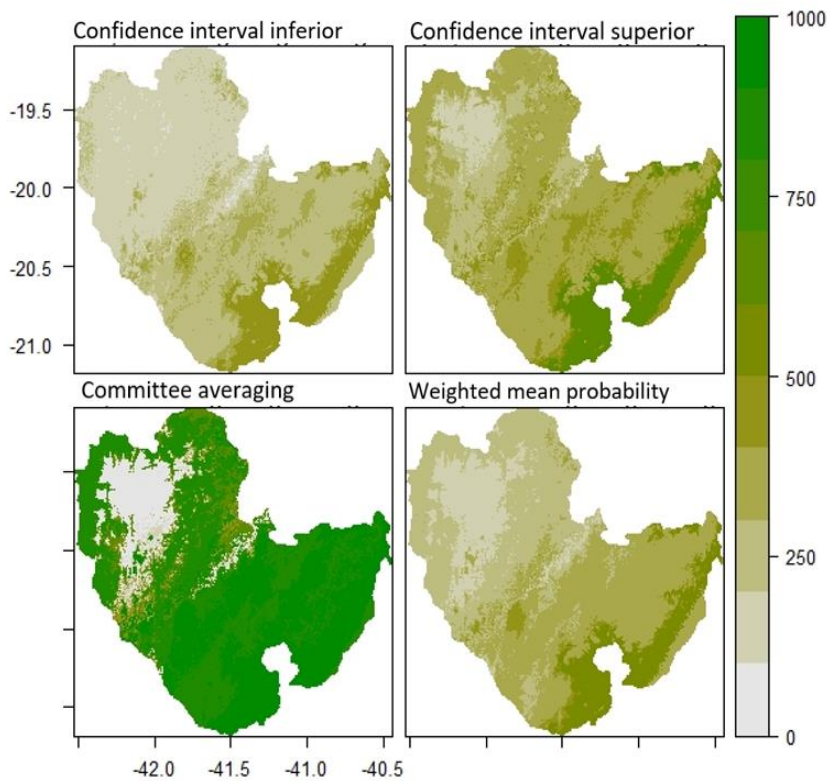
Figure 5: Mean variable importance for each algorithm and ensemble models for predicting current species distribution of *Callithrix flaviceps*. Values are relative to each other, values closer to 1 are considered more important for predicting the species distribution. Climatic variables: BIO1 (Annual Mean Temperature), BIO4 (Temperature Seasonality), BIO13 (Precipitation of Wettest Month), BIO15 (Precipitation Seasonality). Algorithms: GLM: Generalised Linear Models; GAM: Generalised Additive Models, MARS: Multivariate Adaptive Regression Splines, RF: Random Forest, GBM: Generalised Boosting Models, CTA: Classification Tree Analysis, MAXENT: Maximum Entropy, EM: Ensemble Models.

#### e) Current and future projections and Ensemble forecasting

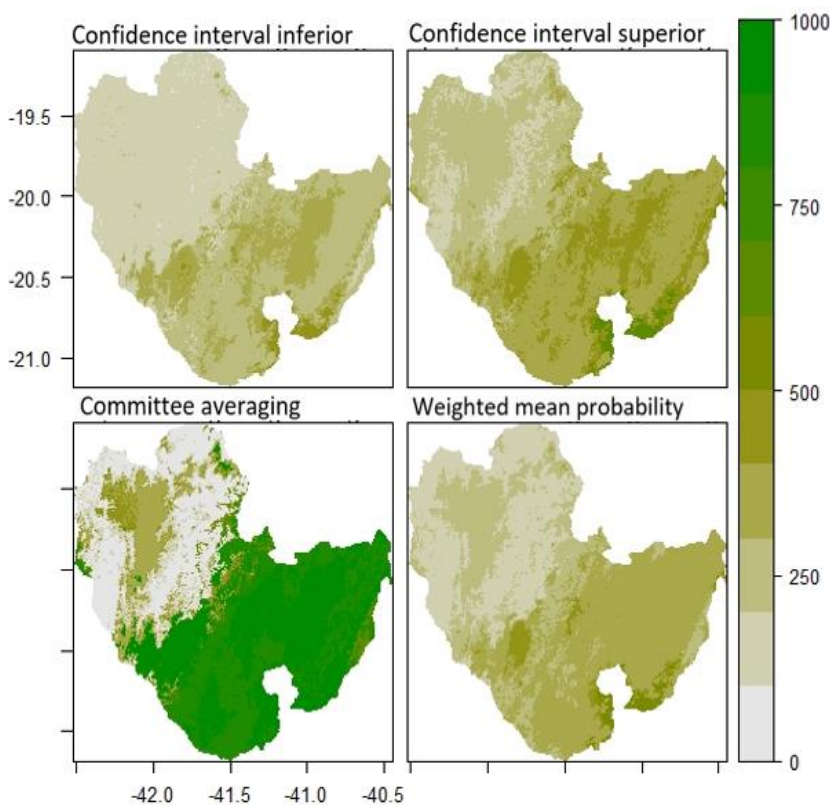
Figure 22 of Additional materials shows the predicted current climatic suitability for *Callithrix flaviceps*. The ensemble model output (bottom right) shows the weighted mean probability of suitability. Measures of uncertainty included committee averaging (bottom left) and 2 confidence levels around the mean of probability (top left and right). A high committee averaging value means that all models agreed to predict presence of the species. On the other hand, committee averaging of 0 means that all models agree to predict absence. Values around 0.5 mean that half of the models predicted presence, the other half predicted absence.

Note that, for storage issues, predicted suitability in the *biomod2* package is given on a scale 0-1000. Simply divide it by 1000 to have a 0 to 1 probability scale (Guisan, Thuiller, & Zimmermann, 2017).

Future climatic projections for the years 2050 and 2070 based on intermediate Representative Concentration Pathways (RCP4.5) are shown in *Figures 6 and 7*. Future climatic projections for the years 2050 and 2070 based on extreme Representative Concentration Pathways (RCP8.5) are shown in *Figures 8 and 9*.



*Figure 6: Future (2050) prediction of climatic suitability for Callithrix flaviceps under climate scenario RCP4.5. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.*



*Figure 7: Future (2070) prediction of climatic suitability for Callithrix flaviceps under climate scenario RCP4.5. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.*

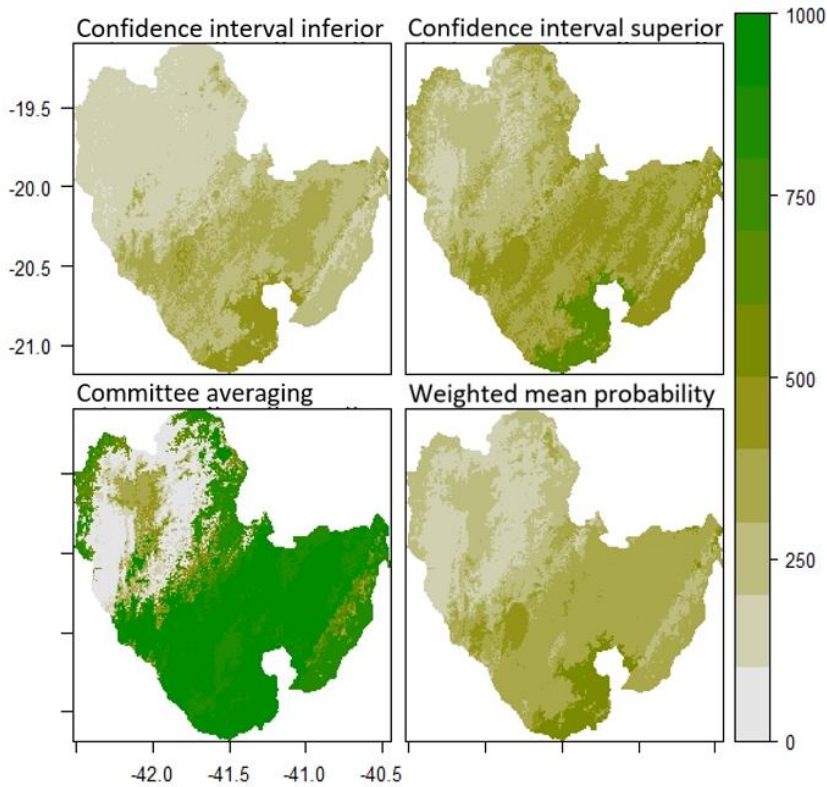


Figure 8: Future (2050) prediction of climatic suitability for *Callithrix flaviceps* under climate scenario RCP85. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.

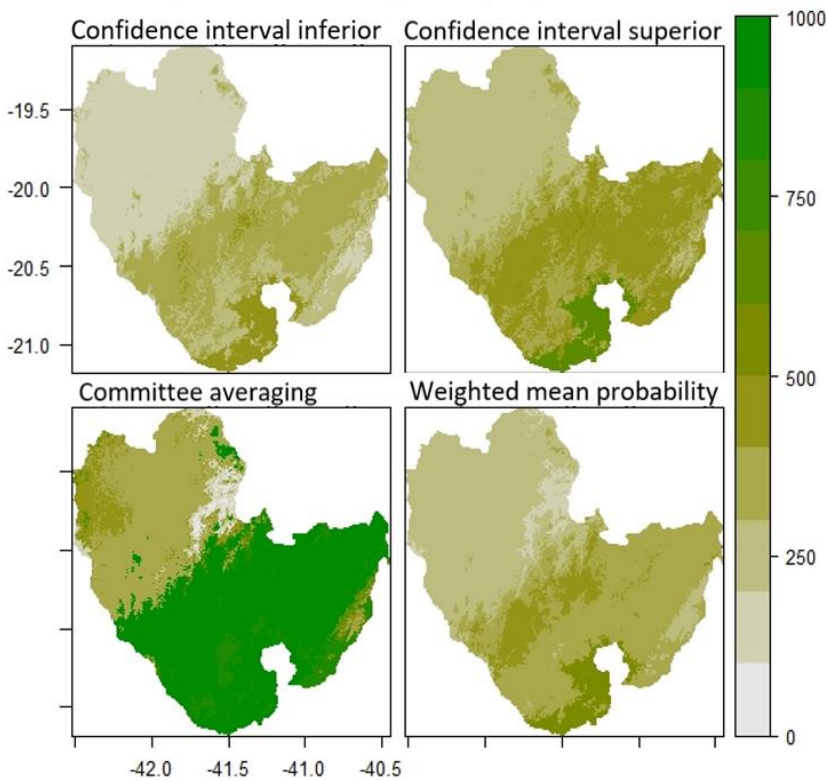


Figure 9: Future (2070) prediction of climatic suitability for *Callithrix flaviceps* under climate scenario RCP85. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.

## Landscape suitability

### a) Individual models

The same method and set of algorithms (n=7) were used for modelling the current landscape suitability including elevation from the digital elevation model (DEM) and tree canopy cover (Treecover) as predictor variables. Hence, 63 models were generated (3 pseudo-absence selections x 3 cross validation procedures (runs) x 7 algorithms).

Figure 11 shows that RF, GBM and CTA have similar scores and stand out from the rest of the algorithms.

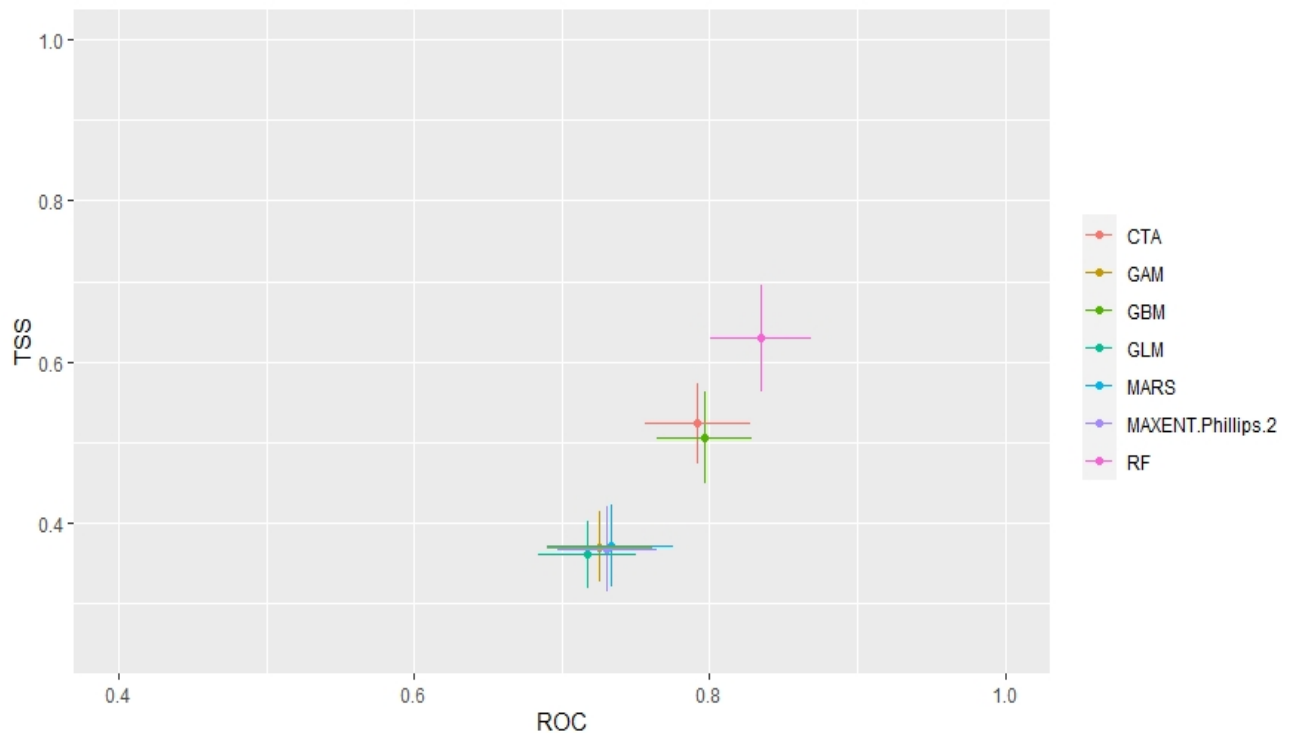


Figure 10: Model scores for each algorithm for predicting current species distribution of *Callithrix flaviceps* from landscape variables (elevation and tree cover), according to the TSS (True Skill Statistics) and the ROC (Relative Operative Characteristic). Algorithms: CTA: Classification Tree Analysis; GAM: Generalised Additive Models; GBM: Generalised Boosting Models; GLM: Generalised Linear Models; MARS: Multivariate Adaptive Regression Splines; MAXENT: Maximum Entropy and RF: Random Forest. Error bars represent the range of the different models of the same algorithm, central point corresponds to average between models.

### b) Response curves

Graphs showing response curves for each algorithm are shown in Figures 8-14 of *Additional materials*. All algorithms with no exception showed a positive relationship between probability of occurrence and tree canopy cover. In addition, a drop at 30% of tree cover was observed for most models.

For every algorithm, higher elevation (elevation closer to 2500 m) was associated with higher probability of occurrence for *Callithrix flaviceps*. The machine learning methods GBM, CTA and RF suggested a plateau above 1400 m (*Additional Materials, Figures 8-14*).

c) Ensemble modelling

Following the same method as for climatic suitability, five ensemble models were created. Considering the TSS threshold, two individual models were kept for the Ensemble: "PA1\_RUN2\_RF" and "PA3\_RUN3\_RF".

The weighted mean (*wmean*) had slightly better scores than the mean across predictions, so I kept this ensemble model output for projections (Table 2).

Table 2: Ensemble models scores. The mean, giving the mean probabilities across predictions; the weighted mean (*wmean*), estimating the weighted sum of probabilities; the committee averaging (*ca*), giving both a prediction and a measure of uncertainty, and the confidence interval (*ci*), showing 2 estimations (*ci inf* and *ci sup*) of the confidence interval around the mean probability. ROC: Relative Operative Characteristic, TSS: True Skill Statistics.

	ROC TESTING DATA	TSS TESTING DATA	SENSITIVITY (TSS)	SPECIFICITY (TSS)
MEAN	0.983	0.872	93.750	93.433
WMEAN	<b>0.983</b>	<b>0.876</b>	<b>94.118</b>	<b>93.462</b>
CA	0.974	0.903	97.794	92.524
CI INF	0.934	0.821	85.662	96.335
CI SUP	0.962	0.847	95.956	88.654

d) Variable importance

All algorithms and the ensemble models considered tree cover to be more important than elevation (DEM) to define the *Callithrix flaviceps* distribution (Figure 11).

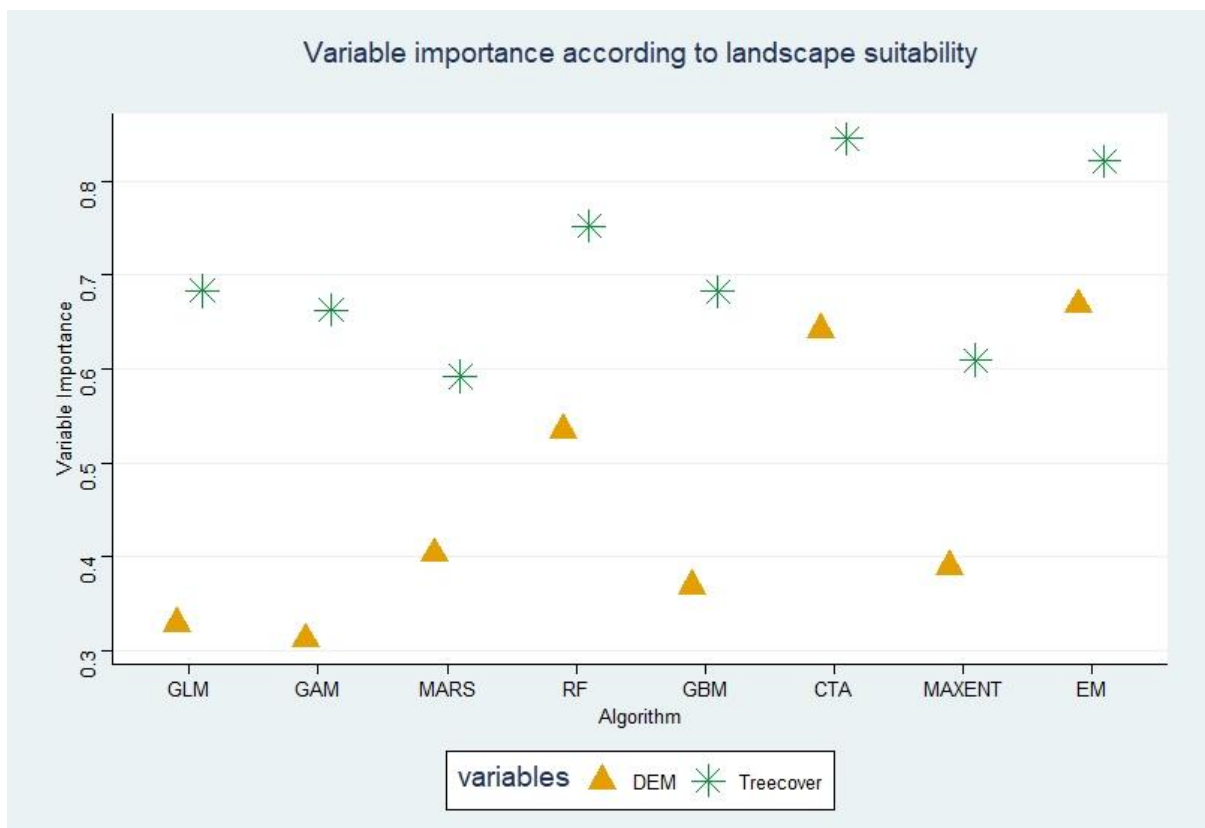


Figure 11: Mean of variable importance for each algorithm and ensemble models for predicting current species distribution of *Callithrix flaviceps*. Values are relative to each other, values closer to 1 are considered more important for predicting the species distribution. DEM (Digital Elevation Model), Treecover (Tree Canopy Cover).

*GLM: Generalised Linear Models; GAM: Generalised Additive Models, MARS: Multivariate Adaptive Regression Splines, RF: Random Forest, GBM: Generalised Boosting Models, CTA: Classification Tree Analysis, MAXENT: Maximum Entropy, EM: Ensemble Models.*

*e) Current and future projections and Ensemble forecasting*

The current projection for landscape suitability is shown in *Figure 23 of Additional materials*. Comparable to the climatic suitability presented above, the committee averaging (bottom left) showing the inconsistency between models and 2 confidence intervals around the mean of probability (top maps) are shown. The ensemble model representing the weighted mean is shown on the bottom right map.

Hotspots of suitability considering only landscape variables correspond to the Caparaó National Park (between the states of MG and ES), the Biological Reserve Augusto Ruschi, and Goiapaba-Açu area.

## Overall habitat suitability

### a) Individual models

Similarly to the climate-based and landscape-based niche suitability modelling, the total number of models created for the overall habitat suitability (combining both climate and landscape variables) was 63 (3 pseudo-absence selections x 3 cross validation procedures (runs) x 7 algorithms).

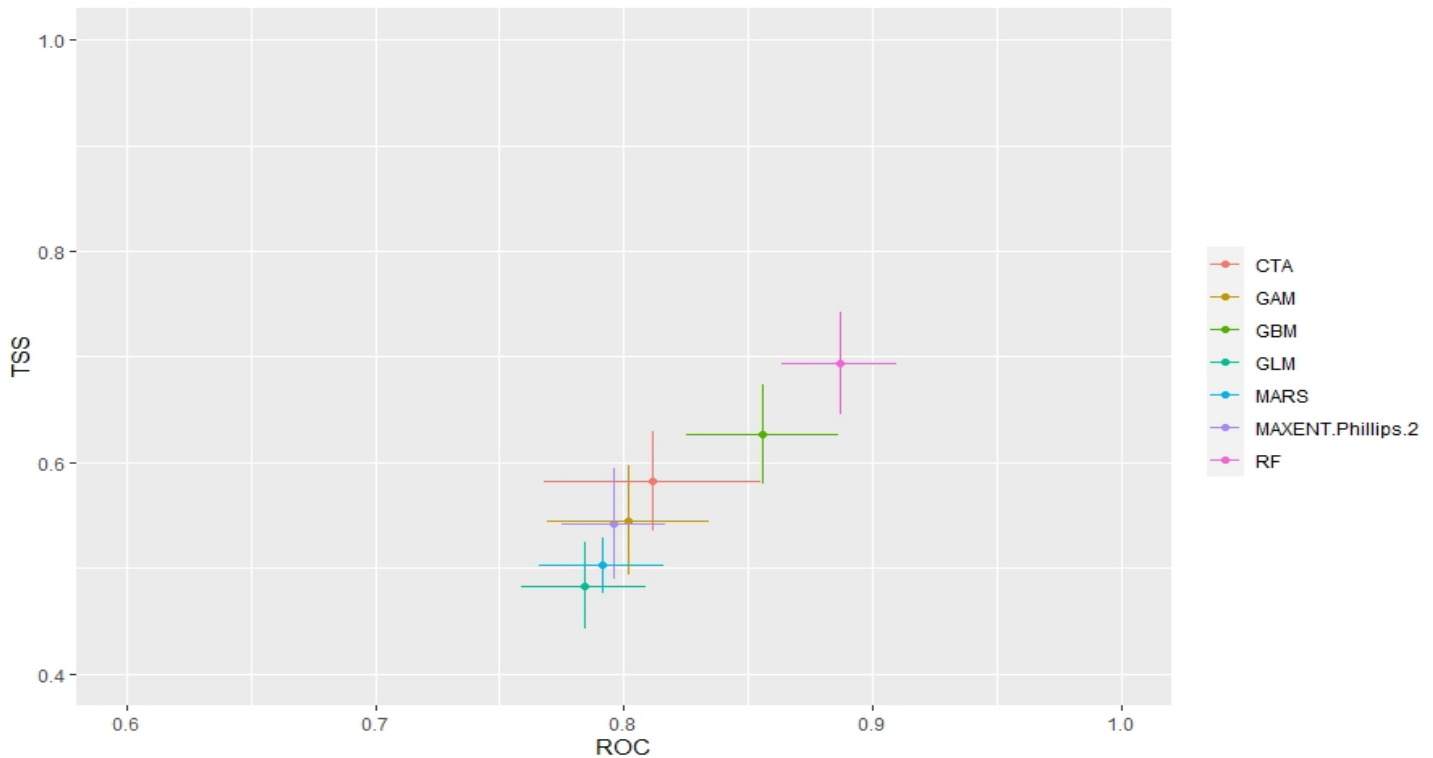


Figure 12: Model scores for each algorithm for predicting current species distribution of *Callithrix flaviceps* from climate and landscape variables according to the TSS (True Skill Statistics) and the ROC (Relative Operating Characteristic). Algorithms: CTA: Classification Tree Analysis, GAM: Generalised Additive Models, GBM: Generalised Boosting Models, GLM: Generalised Linear Models, MARS: Multivariate Adaptive Regression Splines, MAXENT: Maximum Entropy and RF: Random Forest. Error bars represent the range of the different models of the same algorithm, central point corresponds to average between models.

As for both climatic and landscape suitability modelling, the machine learning methods (RF, CTA and GBM) tended to perform better. In this case, RF stood out from all others.

### b) Response curves

Response curves from MARS were different in the habitat suitability modelling. Suitability clearly dropped at Annual Mean Temperature (BIO1) of 20°C (drop was at 17/18°C in the climate-only models). At Annual Mean Temperature of 20°C or higher, all algorithms showed low relative probability of presence, except random forest models.

All algorithms except GLMs and CTAs showed higher probability of species occurrence below 55% and at or above 70% of Precipitation Seasonality (BIO15).

According to GLM, MARS and MAXENT, when combining all environmental variables, elevation above 1000m seemed to be unsuitable for *Callithrix flaviceps*. This is surprising because all algorithms showed positive relationship between elevation and probability of presence when



modelling landscape variables only. The machine learning methods GBMs and RFs seemed more consistent and showed similar pattern as they did in landscape-only modelling.

*c) Ensemble modelling*

Based on the TSS threshold, five individual models were kept for EM: "PA1\_RUN2\_RF", "PA1\_RUN3\_RF", "PA3\_RUN1\_RF", "PA3\_RUN2\_RF", "PA3\_RUN3\_RF". Here again, five ensemble models were created.

Table 3 shows the five ensemble model scores. For consistency, the ensemble model weighted mean was kept for the projections.

*Table 3: Ensemble model scores. The mean, giving the mean probabilities across predictions; the weighted mean (wmean), estimating the weighted sum of probabilities; the committee averaging (ca), giving both a prediction and a measure of uncertainty, and the confidence interval (ci), showing 2 estimations (1 high and 1 low) of the confidence interval around the mean probability. ROC: Relative Operative Characteristic, TSS: True Skill Statistics.*

	<b>ROC</b> TESTING DATA	<b>TSS</b> TESTING DATA	<b>SENSITIVITY</b> (TSS)	<b>SPECIFICITY</b> (TSS)
<b>MEAN</b>	0.993	0.931	98.893	94.220
<b>WMEAN</b>	<b>0.993</b>	<b>0.931</b>	<b>99.262</b>	<b>93.794</b>
<b>CA</b>	0.989	0.945	99.262	95.224
<b>CI INF</b>	0.983	0.877	96.679	90.995
<b>CI SUP</b>	0.993	0.945	100	94.433

*d) Variable importance*

Precipitation Seasonality (BIO15) and Tree cover were the most important variables in climate-based modelling and landscape-based modelling respectively. Interestingly, statistical methods (GLM, GAM and MARS) gave varying results while modelling overall habitat suitability: In landscape-based modelling, statistical methods (GLM, GAM and MARS) considered tree cover to be more important than elevation (DEM). In the habitat modelling however, when combined with climate variables, elevation was considered more important than tree cover by all statistical methods (GLM, GAM, MARS).

Similarly, in climate-based modelling, all algorithms agreed on Precipitation Seasonality (BIO15) being the most important climate variable. However, when combined with landscape variables (modelling overall habitat suitability), statistical methods (GLMs, GAMs, MARSs) and MAXENT considered Annual Mean Temperature (BIO1) to be the most important variable (*Figure 13*).

Note however that statistical methods and MAXENT had the lowest scores. The machine learning methods RF and GBM seemed more consistent. They considered Precipitation Seasonality (BIO15) and tree cover the two most important variables (*Figure 13*), which is in accordance with the climate-based and landscape-based models done separately.

As expected because ensemble models were built exclusively on RF models, the most important variable to predict the species' distribution was tree cover, followed by Precipitation Seasonality (BIO15) and Temperature Seasonality (BIO4). Elevation (DEM) and Precipitation of Wettest Month (BIO13) were considered the least important variables.

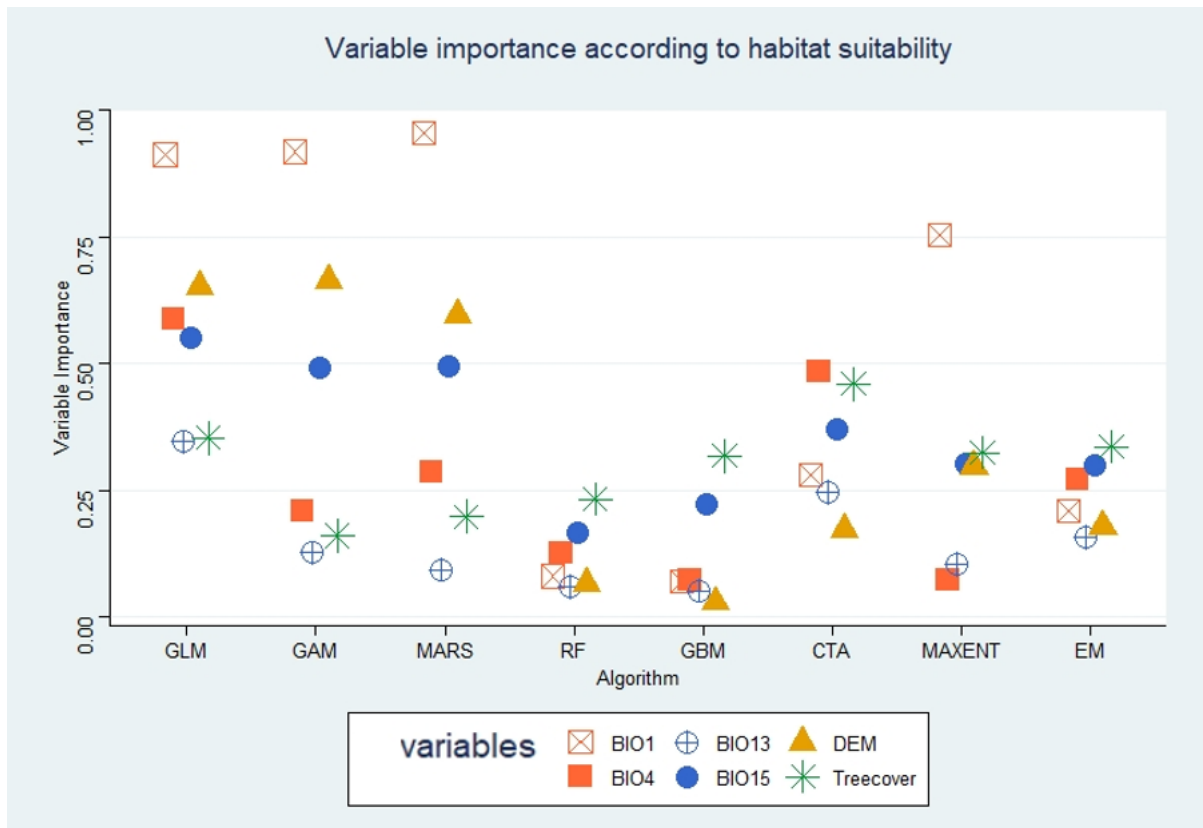


Figure 13: Mean of variable importance for each algorithm and ensemble models for predicting current species distribution of *Callithrix flaviceps* using both climatic and landscape variables. Values closer to 1 are considered more important for prediction distribution. Variables were BIO1 (Annual mean Temperature), BIO4 (Temperature Seasonality), BIO13 (Precipitation Wettest Month), BIO15 (Precipitation Seasonality), DEM (Digital Elevation Model), Treecover (Tree Canopy Cover). Algorithms were GLM: Generalised Linear Models, GAM: Generalised Additive Models, MARS: Multivariate Adaptive Regression Splines, RF: Random Forest, GBM: Generalised Boosting Models, CTA: Classification Tree Analysis, MAXENT: Maximum Entropy, EM: Ensemble Models.

#### e) Current and future projections and Ensemble forecasting

The current projection for overall habitat suitability is shown in Figure 24 of Additional materials.

Combining climate and landscape variables, the Caparaó National Park is still considered a suitable area as well the Biological Reserve Augusto Ruschi. However, the North-East part of the geographical range and Goiapaba-Açu area is now considered less suitable.

The following maps (Figures 14-17) were created to visualise the projected impact of the different climate scenarios on overall habitat suitability (climate + landscape). Please note that maps combining landscape and future climatic projections were created using *current* landscape layers. In each figure, the weighted mean of the probability of occurrence from the ensemble model is shown on the bottom right map. Inconsistency between models is shown on the committee averaging map (bottom left), as well as the two confidence levels around the mean of probability (top maps).

Regardless of the scenario, the Caparaó National Park and the southeast of the range (around Serra das Torres protected area) are projected to be highly suitable and at high consistency between models in 2050. However, committee averaging for the future predictions, show

inconsistency between models, especially considering the year 2070 for the extreme scenario (RCP8.5) (light green/yellow colours).

All areas in white (committee averaging, bottom left map) show high certainty of unsuitable areas. Most areas located in Espírito Santo show high inconsistency (for all projections).

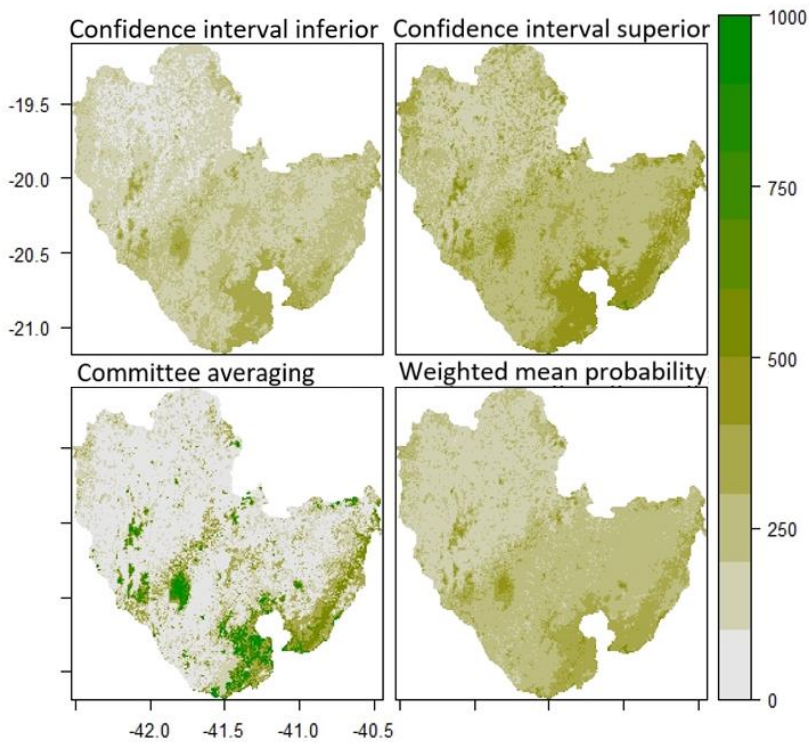


Figure 14: Future (2050) prediction of overall habitat suitability for *Callithrix flaviceps* under climate scenario RCP45. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.

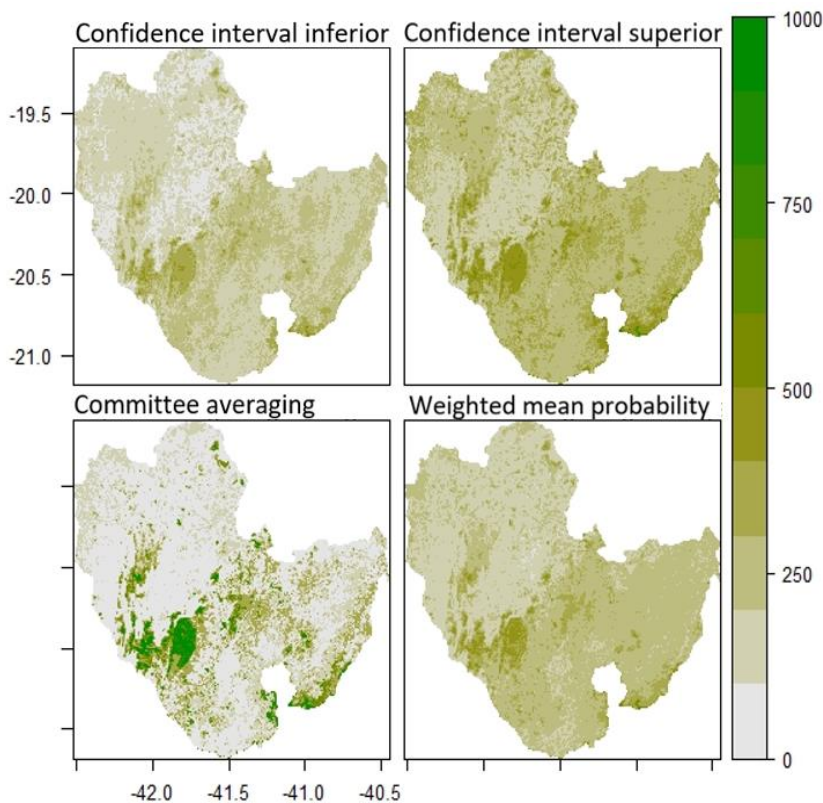


Figure 15: Future (2070) prediction of overall habitat suitability for *Callithrix flaviceps* under climate scenario RCP45. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.

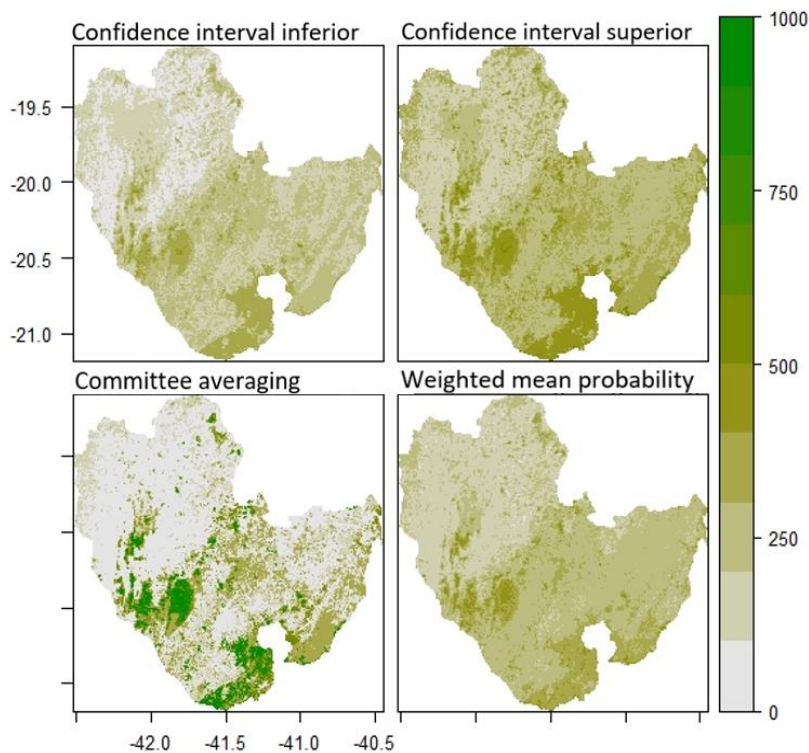


Figure 16: Future (2050) prediction of overall habitat suitability for *Callithrix flaviceps* under climate scenario RCP85. Each map represents a different Ensemble Model. The 2 top maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.

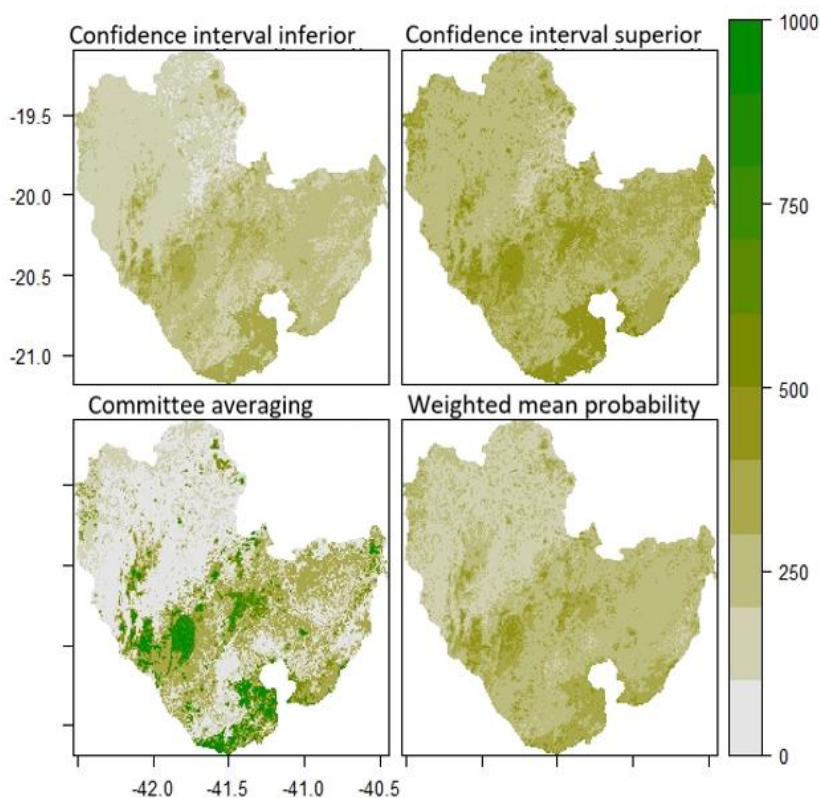


Figure 17: Future (2070) prediction of overall habitat suitability for *Callithrix flaviceps* under climate scenario RCP85. Each map represents a different Ensemble Model. The 2 top maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.

## Discussion

The aims of this study were to determine what defines a suitable ecological niche for *Callithrix flaviceps* and to map likely suitable areas for future climate change scenarios. To do so, climate niche-based, and landscape niche-based modelling were run first separately, then together. Climate change scenarios from two Representative Concentration Pathways (RCP4.5 and RCP8.5) for the years 2050 and 2070 were generated to determine the future climatic suitability of the focal species. Based on these results, and with the caution of the limitations outlined below, I will propose recommendations for conservation and restoration effort.

The reliability of the occurrence dataset was not optimal. Bias was introduced in many ways: in some cases (citizen science), there was no measure of effort, i.e., the results may influence the detectability of the species. This can introduce a bias toward observations in more accessible areas or areas closer to human infrastructure. On the other hand, the species might be present in areas not registered simply because nobody was actively searching for them. Another source of bias is the several decade time span of the dataset. The groups might already be gone from areas where they were found in the 1960s. Moreover, the Atlantic Forest being highly fragmented and connectivity already an issue, some areas where the species currently exists, may not be a real source for dispersal to adjacent areas that are assumed to be suitable in the future. Furthermore, the occurrence points were not evenly distributed within the study area.

Using Ensemble modelling in this study has shown the importance of including both climate and landscape variables to model ecological niche (especially when using statistical methods). Indeed, four climate variables and two landscape variables were included in the models. Interpretation from statistical methods (GLMs, GAMs, MARSs) and MAXENTs was different when modelling climate and landscape variables separately (climatic suitability and landscape suitability) versus together (overall habitat suitability). On the other hand, the machine learning methods Random Forest (RFs) and Generalised Boosting Method (GBMs) performed better and were more consistent, which was supported by Elith et al. (2006) stating that boosting and bagging methods usually have higher predictive performance (Elith, et al., 2006).

Landscape only niche-based modelling had lower scores than climate only niche-based and overall habitat modelling, which can be explained by the low number (2) of variables. For future work, more significant landscape variables could be added to the models, (i.e., future deforestation). In this study however, I aimed to predict the species niche and its potential distribution over decades. For these specific aims, Rezende et al. (2020) stated that climate variables were shown to be more relevant to use than landscape variables (Rezende, Sobral-Souza, & Culot, 2020). It is also important to mention that a model and/or algorithm that performs better than another under certain conditions, might not be the best for projections under future climate scenarios (Guisan, Thuiller, & Zimmermann, 2017). In addition, in this present work, the ecological niche modelling did not include any biotic factors.

In their study on geographic distribution of all *Callithrix* species in southeast Brazil, Braz et al. (2019) concluded that, unlike most *Callithrix* species, parapatry of *Callithrix flaviceps* was mostly maintained by biotic factors (i.e., interspecific interactions with *C. aurita* and *C. penicillata*) (Braz, Lorini, & Vale, 2019). Furthermore, the effect from the invasive species *C. penicillata* and *C. jacchus* and their range expansion leading to hybridisation are not to be negligible (Malukiewicz, et al., 2021a).

In other words, suitable areas highlighted by the modelling might not be suitable in reality, due to lack of connectivity and species interactions (hybridisation, competition).

#### Current projections

Precipitation Seasonality (BIO15) and Temperature Seasonality (BIO4) were considered the most important climatic variables to predict the species' distribution (*Figure 5 and 13*). These results align with previous hypotheses highlighting the importance of seasonality for the *Callithrix* spp. (Vale, Menini Neto, & Prezoto, 2020) and the tendency for montane species (*C. aurita* and *C. flaviceps*) to occupy colder environments than other *Callithrix* species (Braz, Lorini, & Vale, 2019). On the other hand, Precipitation of Wettest Month (BIO13) was considered the less decisive variables (*Figure 5 and 13*). Another Precipitation variable could be more relevant to use, i.e., Precipitation of the Driest Month (BIO14) to investigate if drought/low precipitation is a limiting factor, considering the effect of the prey species.

Regarding the landscape variables, tree cover played a major role in predicting distribution and was considered more decisive than elevation (DEM). It is known however, that elevation is more likely to have an indirect effect on species distribution because altitude influences temperature. Hence, for modelling, it is better to use variables that have direct effects on distribution only.

Overall, while modelling habitat suitability combining the climate and landscape variables, tree cover was the most important variable, closely followed by Precipitation Seasonality (BIO15).

As expected, most hotspots of current high climate and landscape suitability correspond to the already existing protected areas (*Figures 22 and 23 of Additional materials, see Figure 18 for protected area names*). This is also the case for many of the clusters of presence-only points that the analyses were based on (*Figure 1*). For example, the Caparaó National Park (between the states of MG and ES), the Biological Reserve Augusto Ruschi, the Goiapaba-Açú area (including both the environmental protection area and municipal park), the RPPN Águia Branca and the State parks of Forno Grande and Pedra Azul located in Espírito Santo are hotspots of suitability.

#### Projections under future climate change scenarios

Projections show fair scores (TSS > 0.4 and ROC > 0.7) despite the poor quality of the dataset. Nevertheless, results must be interpreted with caution, keeping in mind the sampling bias (observations and methods used were not evenly spread between 1964-2020), as well as in space (effort cannot be assumed to have been evenly distributed in space), and the temporal resolution mismatch (environmental variables were from 1970-2000 but occurrence points are from 1964-2020). In addition, committee averaging maps showed variation between models. Conservation and restoration decisions made on areas with inconsistency must be well reasoned. As seen in many previous studies on ecological niche modelling, the difference between current and closer scenarios (2050) can be huge, then tends to fade a bit (in 2070) (Guisan, Thuiller, & Zimmermann, 2017).

Regardless of the scenario, projections for future suitability between climate-only and overall suitability showed similar results. However, the committee averaging showed way more inconsistency in the overall habitat suitability (*Figures 14-17*) than in climate-only modelling (*Figures 6-9*), meaning that models had more difficulties agreeing whether the species would be present or absent in specific areas. In addition, committee averaging from overall habitat

suitability modelling predicted more areas of absence of the species. This points out that landscape variables had a big impact and suggests that tree cover should be maintained/restored.

In the climate-only modelling, Espírito Santo seemed highly suitable and with high consistency between models to predict presence. In the overall habitat modelling, most areas of the geographical range showed inconsistency between models. However, hotspots such as the Caparaó National Park, the south-eastern region of the geographical range (around Serra das Torres) and the State Park of Pedra Azul and Forno Grande were predicted as presence with high consistency between models.

In their study on the effects of climate change on all six *Callithrix* species in eastern Brazil, Braz et al. (2019) produced binary projections (suitable/non-suitable) and predicted the east side of Espírito Santo to be the most suitable for *C. flaviceps* in the extreme scenario (RCP8.5) for the year 2050, using a mean of two GCMs: HADGEM-ES & MIROC5, and considering the WorldClim climatic variables: Mean Diurnal Range (BIO2), Temperature Seasonality (BIO4), Mean Temperature of Warmest Quarter (BIO10), Annual Precipitation (BIO12), Precipitation Seasonality (BIO15) and Precipitation of Warmest Quarter (BIO18) (Braz, Lorini, & Vale, 2019). Indeed, they predicted that the eastern side of geographical range would be the only remaining suitable area. In these present results, the same area showed high inconsistency between models (*Figure 16*, committee averaging) but was projected as poorly suitable (*Figure 16*, weighted mean). Their findings predicted that all *Callithrix* species would maintain their current distribution areas, with the exception of *C. flaviceps*, for which the area would decrease by 95% (1,505 km<sup>2</sup>) (Braz, Lorini, & Vale, 2019). Results from this thesis differ, some explanations can be the lack of consistency in the occurrence dataset of our studies (they also used existing data, gathered from different sources), the very lower number of observations (n = 57), and the different climatic variables and spatial resolution used. Moreover, ensemble model outputs from this thesis were all built from random forest models, Braz et al. (2019) however, did not include this algorithm in their ensemble modelling analysis.

Overall, present results show more suitability within the state of Espírito Santo than in Minas Gerais (and Rio de Janeiro). Accordingly, Pinto et al. (2009) predicted that hotspots for the genus *Callithrix* correspond to the forest of Espírito Santo (Pinto, Lasky, Bueno, Keitt, & Galetti, 2009).

#### Description of the future projections

Considering the pressure from anthropogenic activities, the extreme RCP8.5 scenario is the most likely to happen in South America (Cuyckens, et al., 2016). Hence, habitat suitability projection for 2070 considering the extreme climate scenario (RCP8.5) and protected areas already in place, is shown in *Figure 18*.

None of the areas showed high overall habitat suitability within this scenario (all areas < 60% suitability, however, some hotspots are found).

The Caparaó National Park represents a moderately suitable area with high consistency in all scenarios. Among all four scenarios (2 RCP x 2 years), models predict suitability at 30-60% depending on the scenario, considering both climate-only and habitat suitability modelling. Climate-only and overall habitat suitability predictions show the same pattern. Based on these

results, this already protected area should be **maintained**, and is likely to be a key area for the persistence of *Callithrix flaviceps*.

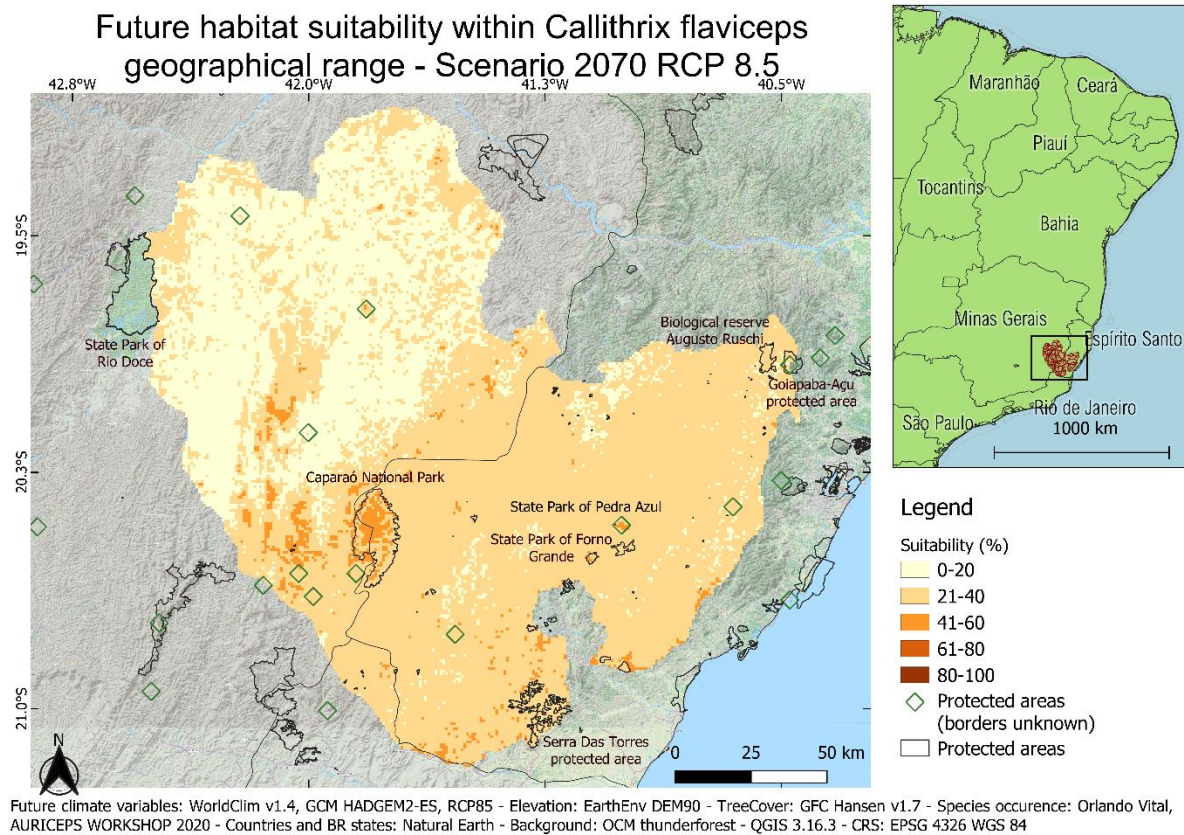


Figure 18: Projected overall habitat suitability for *Callithrix flaviceps*, combining both climate and landscape variables for the year 2070 considering the extreme climate change scenario (RCP8.5).

Results from the current study do not indicate a need to prioritize Northeast region of the current geographical range (around Augusto Ruschii and Goiapaba-Açu protected areas) for conservation of *Callithrix flaviceps*, even though this is one of the areas with the largest number of occurrence points. When considering only climate variables, future suitability was moderate (around 40%). The committee averaging maps from climate-only models show low variations between models. However, when combining climate and landscape variables together (overall habitat suitability modelling), suitability drops to 25% and inconsistency among models varies from “highly certain of absence” to “uncertain”. This is surprising because the landscape variables used in overall habitat modelling are the exact same as for landscape-only modelling. Current modelling in this area showed high suitability (with high consistency) (Figures 22, 23 and 24 of *Additional materials*). From this study, I can say that the future of the species in this area is highly uncertain, and further studies would be recommended. Note that most occurrence points collected in this region are fairly old (1960s-1980s) except inside the protected area of Augusto Ruschi, for which occurrences were more recent (2000s, 2010s).

**Restoration** of Serra Das Torres should be considered since this protected area was more suitable considering the climate-only variables regardless of the scenario, and consistently indicating presence from committee averaging from both habitat and climate-only modelling.



Suitability hotspots did not seem to follow the terrain elevation pattern except for the year 2050 considering RCP8.5. Indeed, except for Caparaó, the areas of highest projected current and future suitability seemed to match with lowest altitudes. Interestingly, around the Fazenda Macedônia east side of Minas Gerais, elevation is low (<400/500 m a.s.l.) and on the south-eastern part of the current range around Serra das Torres, where projected suitability is good, elevation is also low, especially compared to Caparaó National Park (highest point at 2800 m a.s.l.). This is in accordance with previous findings from Braz et al. (2019), who concluded that the montane species *C. flaviceps* and *C. aurita* under the RCP8.5 scenario are expected to lower their elevation by 2050 (Braz, Lorini, & Vale, 2019).

#### Relevant factors for defining the niche

Precipitation level is directly related with food availability, less precipitation meaning less arthropods (Corrêa, Coutinho, & Ferrari, 2000). This could explain why *C. flaviceps* seem to prefer higher precipitation seasonality (70%). However, the wide ecological plasticity of *Callithrix* spp. has been investigated (Ferrari, Corrêa, & Coutinho, 1996), and it is understood that they can easily change their diet according to the change of environment (Rodrigues & Martinez, 2014). For example, *Callithrix flaviceps* is known to switch from a gummivorous/insectivorous diet to feeding more on fruits when these latter were more abundant in the Caratinga Biological Station (Ferrari, Corrêa, & Coutinho, 1996).

The current and previous studies have provided knowledge about the fundamental niche of *Callithrix flaviceps*, but it is challenging to know the realised niche because of constraints from biotic factors such as resource competition, disease outbreaks and hybridisation. Hence, to have a broader picture and more accurate projections of the realised niche, further studies are needed, including species interactions with *C. aurita*, but also with the common and invasive marmosets *C. penicillata* and *C. jacchus*, as well as metrics such as landscape variables to predict constraints on *C. flaviceps* movement and dispersal capacity (Rezende, Sobral-Souza, & Culot, 2020). In addition, as proposed by Hannibal et al. (2019), agricultural land and urban areas should be included in the next conservation strategies to better consider habitat fragmentation (Hannibal, et al., 2019).

Silva et al. (2018), highlighted that being a rare species, the geographical distribution of *Callithrix flaviceps* may be underestimated (Silva, et al., 2018). A population of *C. flaviceps* has recently (2021) been observed in Santa Bárbara municipality, MG (personal communication, Orlando Vital). This is promising and suggests potential suitable areas outside current geographical range.

Genetic studies seem a great way for long term monitoring and conservation of species. However, to my knowledge, no genetic data is available for *Callithrix flaviceps*. As suggested by Malukiewicz et al. (2021), *Callithrix flaviceps* would highly benefit from the genomic skimming method and low-cost desktop sequencing to rapidly increase its genomic resources, and thus, obtain deeper coverage of genomes (including mitogenomes) (Malukiewicz, et al., 2021b). Monitoring the genetic variation within the population and inbreeding levels, to start a healthy population in captivity but also to verify the hybridisation level of the individuals seem necessary. Recently, Malukiewicz et al. (2021) discovered that an individual with a *C. aurita* phenotype was in reality a cryptic hybrid with a *C. penicillata* mitogenome lineage. The hybrid population might therefore be underestimated (Malukiewicz, et al., 2021b).

Despite the lack of consistency in the presence records leading to a poor-quality dataset, present results have highlighted some important variables affecting the *Callithrix flaviceps* distribution. I found that most hotspots of overall habitat suitability were matching with already in place protected areas, which is promising. However, the Atlantic Forest being a fragmented environment, these potential suitable areas will be lost if we don't improve connectivity. Moreover, species interactions and hybridisation with the common marmosets are making the situation more challenging. Further work is needed to confront present findings and try to investigate the realised niche of *Callithrix flaviceps* more in depth. Efforts for conservation are still desperately needed, it is therefore important to use the data available.

## **Acknowledgements**

First, I would like to express my sincere gratitude to Ane Eriksen, from Inland Norway University of Applied Sciences (INN, *Evenstad, Norway*), for her availability and willingness (even at long distance) to provide detailed feedback on the writing. Thank you for the support, encouragements and for our helpful discussion on my work.

This thesis would not have been possible without Rodrigo Salles de Carvalho, Sally Fransen and James Hall, from the Mountain Marmoset Conservation Program (MMCP) in Brazil, who guided me toward the Federal University of Viçosa (UFV, *Minas Gerais, Brazil*).

I am grateful to Fabiano Melo for making this partnership between INN and UFV possible and for trusting me to complete this thesis.

I'd like to thank Adriana Pereira Milagres, "Drica" from UFV for her guidance on ecological niche and ensemble modelling, as well as Olivier Devineau, from INN for his direct advice and feedback on various steps on the statistical analyses.

A warm thank you goes to Orlando Vital (from MMCP & UFV), for gathering the occurrence data and for our occasional chats on *Callithrix flaviceps* and the *Auriceps* project.

Lastly, I'd like to thank the INN team, especially the researchers, professors and students present during our ecology research group meetings on campus Evenstad (INN), with a special thanks to Mélanie Spedener for her time and guidance throughout this Master's thesis path.

## References

- 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*, 1223-1232.
- Amaral, M. V., de Souza, A. L., Soares, V. P., Soares, C. P., Martins, S. V., Leite, H. G., . . . de Oliveira Gaspar, R. (2009). Dinâmica da Cobertura Vegetal (1980-2004) em um Projeto de Produção Florestal, nos Municípios de Bugre e Ipaba, MG. *R. Árvore, Viçosa-MG*, pp. 135-125.
- Braz, A. G., Lorini, M. L., & Vale, M. M. (2019). Climate change is likely to affect the distribution but not parapatry of the Brazilian marmoset monkeys (*Callithrix* spp.). *Diversity and Distributions*, 536-550.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 5–32.
- Carlucci, M., Marcilio-Silva, V., & Torezan, J. (2021). The Southern Atlantic Forest: Use, Degradation, and Perspectives for Conservation. In M. C. Marques, & C. E. Grelle, *The Atlantic Forest : History, Biodiversity, Threats and Opportunities of the Mega-Diverse Forest* (pp. 91-111). Springer International Publishing.
- Carvalho, J. S., Graham, B., Rebelo, H., Bocksberger, G., Meyer, C. F., Wich, S., & Kühl, H. S. (2019). A global risk assessment of primates under climate and land use/cover scenarios. *Glob Change Biol.*, 3163-3178.
- Corrêa, H. M., Coutinho, P., & Ferrari, S. (2000). Between-year differences in the feeding ecology of highland marmosets (*Callithrix aurita* and *Callithrix flaviceps*) in south-eastern Brazil. *J. Zool., Lond.*, 421-427.
- Culot, L., Pereira, L., Agostini, I., de Almeida, M., Ribas Bella, T., & B.-M. J. (2019). Atlantic - Primates: a dataset of communities and occurrences of primates in the Atlantic Forests of South America. *Ecology*.
- Cuyckens, G., Christie, D., Domic, A., Malizia, L., & Renison, D. (2016). Climate change and the distribution and conservation of the world's highest elevation woodlands in the South American Altiplano. *Global and Planetary Change*, 79-87.
- da Silva, L. G., Ribeiro, M., Hasui, É., da Costa, C. A., & da Cunha, R. G. (2015). Patch Size, Functional Isolation, Visibility and Matrix Permeability Influences Neotropical Primate Occurrence within Highly Fragmented Landscapes. *PLOS ONE*, 1-20.
- De Kort, H., Baguette, M., Lenoir, J., & Stevens, V. (2020). Toward reliable habitat suitability and accessibility models in an era of multiple environmental stressors. *Ecology and Evolution*, 10937-10952.
- de Melo, F., Hilário, R., Ferraz, D., Pereira, D., Bicca-Marques, J., Jerusalinsky, L., . . . M.M., V.-M. (2021b). *The IUCN Red List of Threatened Species 2021*. Retrieved 01 11, 2021, from *Callithrix flaviceps*, Buffy-headed Marmoset: <https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T3571A191700879.en>.
- de Melo, F., Port-Carvalho, M., Pereira, D., Ruiz-Miranda, C., Ferraz, D., Bicca-Marques, J., . . . Mittermeier, R. (2021a). *The IUCN Red List of Threatened Species*. Retrieved from *Callithrix aurita* (amended version of 2020 assessment): <https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T3570A191700629.en>.

- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., . . . Overton, J. M. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 129-151.
- Elith, J., Leathwick, J., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 802-813.
- Estrada, A., Garber, P., Rylands, A., Roos, C., Fernandez-Duque, E., Di Fiore, A., . . . Barelli, C. (2017). Impending extinction crisis of the world's primates: Why primates matter. *Estrada et al. Sci. Adv.*, 1-16.
- Ferrari, S. F. (1988). *The Behaviour and Ecology of the Buffy-Headed Marmoset, Callithrix flaviceps (O. Thomas, 1903)*. London: The Department of Anthropology of University College London.
- Ferrari, S. F., & Mendes, S. L. (1991). Buffy-headed marmosets 10 years on. *Oryx*, 25, 105-109.
- Ferrari, S. F., Corrêa, H. K., & Coutinho, P. E. (1996). Ecology of the "Southern" Marmosets (*Callithrix aurita* and *Callithrix flaviceps*). *Adaptive Radiations of Neotropical Primates*, 157-171.
- Fick, S., & Hijmans, R. (2017, October). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 4302-4315. Retrieved from <https://doi.org/10.1002/joc.5086>
- Friedman, J. (1991). Multivariate Adaptive Regression Splines. *Annals of statistics*, 1-67.
- Guisan, A., Edwards Jr, T., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 89-100.
- Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat Suitability and Distribution Models with applications in R*. Cambridge: Cambridge University Press.
- Hannibal, W., Renon, P., Figueiredo, V., Oliveira, R., Moreno, A., & Martinez, R. (2019). Trends and biases in scientific literature about marmosets, genus *Callithrix* (Primates Callitrichidae): biodiversity and conservation perspectives. *Neotropical Biology and Conservation* 14, 529-538.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., . . . Townshend, J. R. (2013, 11 15). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* , 342(6160), 850-853. doi:DOI: 10.1126/science.1244693
- Hirsch, A., Brome Rylands, A., Paixão Toledo, P., De Brito, B. F., Landau, E., & Resende, N. (1999). Re-analysis of *Callithrix flaviceps* geographical distribution based on new records and through the use of a GIS. *Congresso Brasileiro de Primatologia* (pp. 1-16). Santa Teresa, ES: ResearchGate.
- IPCC. (2022). *IPCC*. Retrieved from Intergovernmental Panel on Climate Change (IPCC): <https://www.ipcc.ch/assessment-report/ar6/>
- Kotu, V., & Deshpande, B. (2014). *Predictive Analytics and Data Mining Concepts and Practice with RapidMiner*. Morgan Kaufmann is an imprint from Elsevier.
- Malukiewicz, J. (2019). A Review of Experimental, Natural, and Anthropogenic Hybridization in *Callithrix* Marmosets. *International Journal of Primatology*, 72-98.

- Malukiewicz, J., Boere, V., Borstelmann de Oliveira, M., D'Arc, M., Ferreira, J. A., French, J., . . . Tardif, S. (2021a). *An Introduction to the Callithrix Genus and Overview of Recent Advances in Marmoset Research*. Preprints.
- Malukiewicz, J., R. A., C., & al., e. (2021b). Genomic Skimming and Nanopore Sequencing Uncover Cryptic Hybridization in One of World's Most Threatened Primates. *Nature*.
- Mares-Guia, M., Horta, M., Romano, A., & al., e. (2020). Yellow fever epizootics in non-human primates, Southeast and Northeast Brazil (2017 and 2018). *Parasites & Vectors*.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 59-69.
- Mendes, C. L., & Pereira, A. (2015). Situação do sagui-da-serra (*Callithrix flaviceps*) em remanescentes Florestais dos distritos de São Sebastião do Sacramento, Dom Corrêa e Palmeiras, pertencentes ao município de Manhuaçu-MG. *Seminário Científico da FACIG* (pp. 1-7). MN, Brazil: Sociedade, Ciência e Tecnologia.
- Mendes, C. S., & de Melo, F. R. (2007). Situação atual do sagüi-da-serra (*Callithrix flaviceps*) em fragmentos florestais da Zona da Mata de Minas Gerais. *A primatologia no Brasil*, 163-180.
- Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 001-012.
- Mittermeier, R. A., Coimbra-Filho, A. F., & Constable, I. D. (1980). Range Extension for an Endangered Marmoset. *Oryx*, 15, 380-383.
- Moraes, A. M., Vancine, M. H., Moraes, A. M., De Oliveira Cordeiro, C., Pinto, M. P., Lima, A. A., . . . Sobral-Souza, T. (2019). Predicting the potential hybridization zones between native and invasive marmosets within Neotropical biodiversity hotspots. *Global Ecology and Conservation*, 1-13.
- Muylaert, R., Vancine, M., Bernardo, R., Oshima, J., Sobral-Souza, T., Tonetti, V., . . . Ribeiro, M. (2018). Uma Nota sobre os Limites Territoriais da Mata Atlântica. *Oecologia Australis*, 302-311.
- Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology & Biogeography*, 361-371.
- Peterson, B. G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., . . . Balkissoon, K. (2020). *CRAN.R*. Retrieved from <https://github.com/braverock/PerformanceAnalytics>
- Phillips, S., Anderson, R., & Schapire, R. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 231-259.
- Pinto, L. S., Costa, C. M., Strier, K. B., & da Fonseca, G. A. (1993). Habitat, Density and Group Size of Primates in a Brazilian Tropical Forest. *Folia Primatol*, 61:135-143.
- Pinto, N., Lasky, J., Bueno, R., Keitt, T. H., & Galetti, M. (2009). Primate Densities in the Atlantic Forest of Southeast Brazil: The Role of Habitat Quality and Anthropogenic Disturbance. In G. Paul A., E. Alejandro, B.-M. Júlio César, H. Eckhard W., & K. B. Strier, *South American Primates Comparative Perspectives in the Study of Behavior, Ecology, and Conservation* (pp. 413-431). Springer, New York, NY.

- Rezende, G. C., Sobral-Souza, T., & Culot, L. (2020). Integrating climate and landscape models to prioritize areas and conservation strategies for an endangered arboreal primate. *Am J Primatol.*, 82:e23202.
- Ribeiro, M., Metzger, J., Martensen, A., Camargo, A., & Hirota, M. (2009). The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, 1141-1153.
- Robinson, N., Regetz, J., & Guralnick, R. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57-67. Retrieved from <http://www.earthenv.org/DEM.html>
- Rodrigues, M. N., & Martinez, R. A. (2014). Wildlife in our backyard: interactions between Wied's marmoset *Callithrix kuhlii* (Primates: Callitrichidae) and residents of Ilhéus, Bahia, Brazil. *Wildlife Biology*, 91-96.
- Rosenberger, A. (1992). Evolution of feeding niches in new world monkeys. *American Journal of Biological Anthropology*, 525-562.
- Rylands, A., Coimbra-Filho, A., & Mittermeier, R. (2009). The Systematics and Distributions of the Marmosets (*Callithrix*, *Callibella*, *Cebuella*, and *Mico*) and *Callimico* (*Callimico*) (Callitrichidae, Primates). *The Smallest Anthropoids*, 25-61.
- Serra-Diaz, J. M., & Franklin, J. (2019). What's hot in conservation biogeography in a changing climate? Going beyond species range dynamics. *Divers Distrib.*, 25:492-498.
- Silva, F., Malukiewicz, J., Silva, L., Carvalho, R., Ruiz-Miranda, C., da Silva Coelho, F., . . . de Oliveira Silva, I. (2018). A Survey of Wild and Introduced Marmosets (*Callithrix*: Callitrichidae) in the Southern and Eastern Portions of the State of Minas Gerais, Brazil. *Primate Conservation*.
- Simas, N. K., Ferrari, S. F., Pereira, S. N., & Leitão, G. G. (2001). Chemical Ecological Characteristics of Herbivory of *Siparuna guianensis* Seeds by Buffy-Headed Marmosets (*Callithrix flaviceps*) in the Atlantic Forest of Southeastern Brazil. *Journal of Chemical Ecology*, pp. 93-107.
- Thuiller, W., Araujo, M., & Lafourcade, B. (2010). *Presentation Manual for BIOMOD*.
- Thuiller, W., Georges, D., Gueguen, M., Engler, R., & Breiner, F. (2021, June 11). *Package 'biomod2'*. Retrieved from [cran.r-project: https://cran-r-project.org/web/packages/biomod2/biomod2.pdf](https://cran.r-project.org/web/packages/biomod2/biomod2.pdf)
- Townsend, C. R. (2008). *Ecological Applications: toward a sustainable world*. Oxford: Blackwell Publishing.
- Trindade Do Carmo, S. (2022). *MSc thesis: Probabilidade de ocupação e detecção do sagui-da-serra (Callithrix flaviceps Thomas, 1903) na RPPN Fazenda Macedônia e em fragmentos de Mata Atlântica no seu entorno, Ipaba, MG*. Viçosa, Minas Gerais: Universidade Federal de Viçosa.
- Vale, C., Menini Neto, L., & Prezoto, F. (2020). Distribution and invasive potential of the black-tufted marmoset *Callithrix penicillata* in the Brazilian territory. *Scientia Plena*, 1-19.
- Vayssières, M., Plant, R., & Allen-Diaz, B. (2000). Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, 679-694.

*Weather and Climate*. (2022). Retrieved from The Global Historical Weather and Climate Data:  
<https://tcktcktck.org/brazil>

Yin, L., Fu, R., Shevliakova, E., & Dickinson, R. (2013). How well can CMIP5 simulate precipitation and its controlling processes over tropical South America? *Clim Dyn*, 3127-3143.



## **Additional materials:**

*Table 1: Method of data collection for the C. flaviceps occurrence points. Data were gathered by Orlando Vital as part of the Auriceps Workshop, 2020.*

	<b>Sighting</b>	<b>Line transect</b>	<b>Citizen science</b>	<b>Museum</b>	<b>Personal records</b>	<b>Camera trap</b>	<b>Interview</b>	<b>NA</b>
<b>1964</b>	1							12
<b>1969</b>				1				
<b>1977</b>				1				
<b>1980</b>	1							2
<b>1982</b>								1
<b>1984</b>								3
<b>1986</b>	3		1					
<b>1987</b>	1							
<b>1988</b>	1							3
<b>1991</b>	8	1						30
<b>1992</b>	1							
<b>1993</b>	11	1					2	27
<b>1995</b>	1			1				
<b>1996</b>	5							
<b>1997</b>	1			1				14
<b>1999</b>	4							2
<b>2000</b>	1			1				8
<b>2001</b>								2
<b>2002</b>								1
<b>2003</b>	3	11						1
<b>2005</b>	14							
<b>2006</b>	1		1				1	
<b>2009</b>								2
<b>2010</b>								1
<b>2011</b>								1
<b>2012</b>		1						7
<b>2013</b>								12
<b>2014</b>								5
<b>2015</b>	3							18
<b>2016</b>	1		1					
<b>2017</b>	1					3		
<b>2018</b>	4		3		5	1		9
<b>2019</b>	6							
<b>2020</b>	1							1
<b>Grand Total</b>	<b>73</b>	<b>14</b>	<b>6</b>	<b>5</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>162</b>

## Part 1: Response curves for climatic suitability models

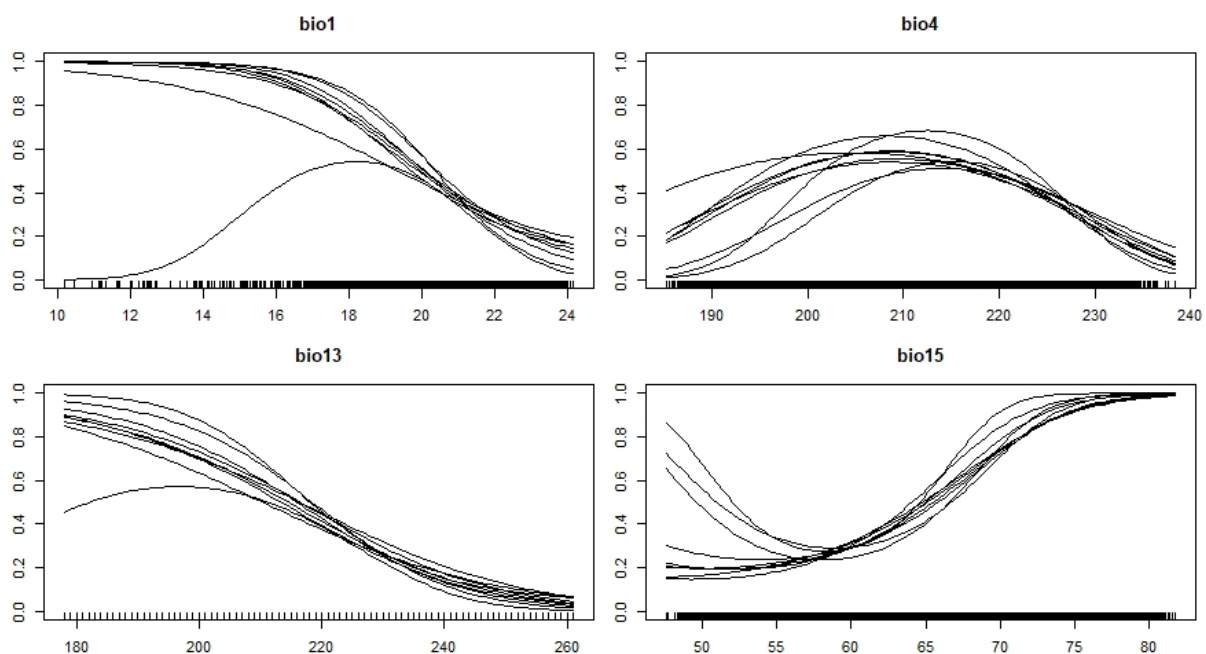


Figure 1: Response Curves from generalised linear models (GLM). Where BIO1: Annual Mean Temperature (°C), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

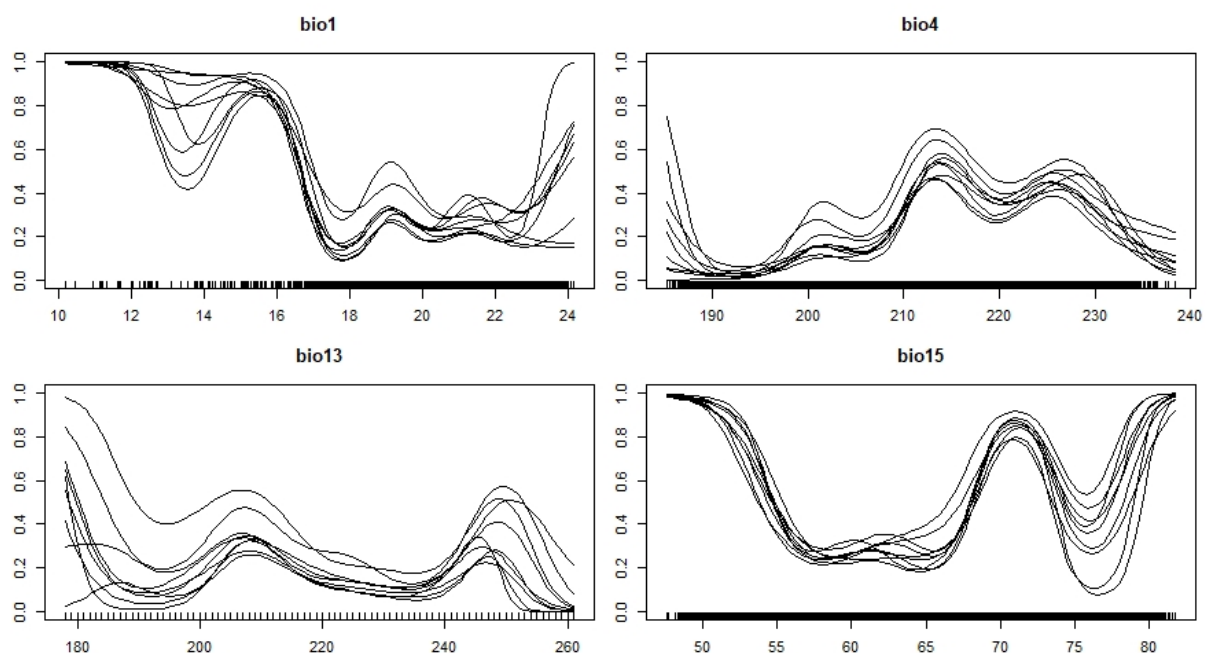


Figure 2: Response Curves from generalised additive models (GAM). Where BIO1: Annual Mean Temperature (°C), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

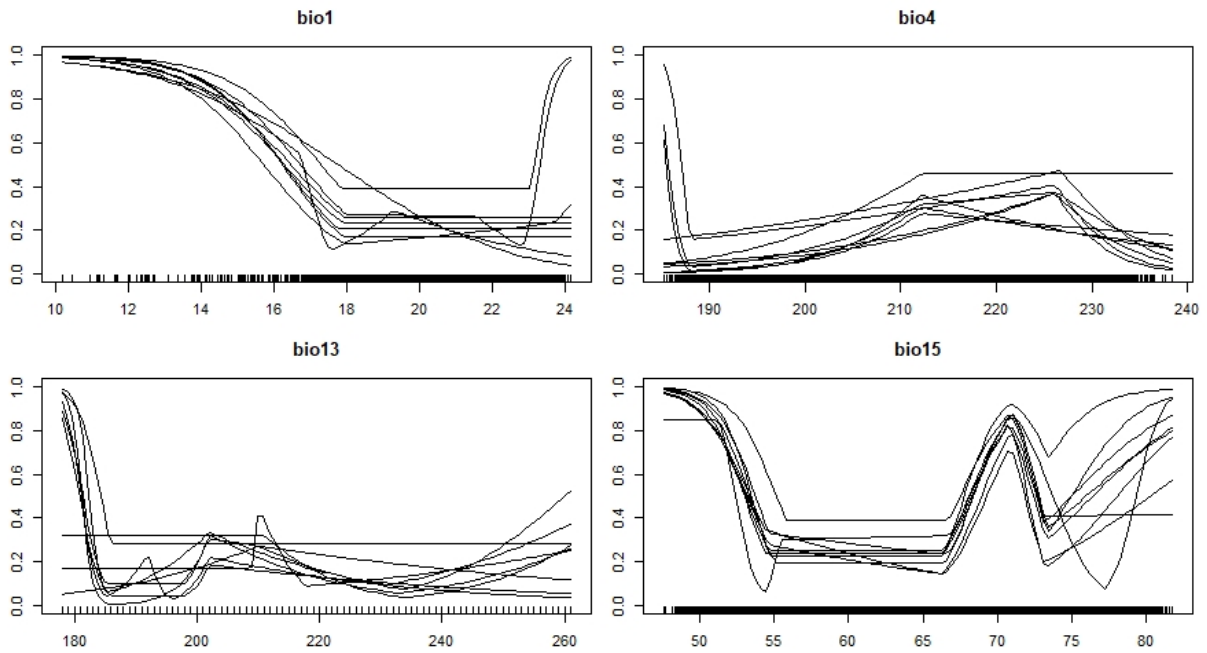


Figure 3: Response Curves from multivariate adaptive regression splines(MARS). Where BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

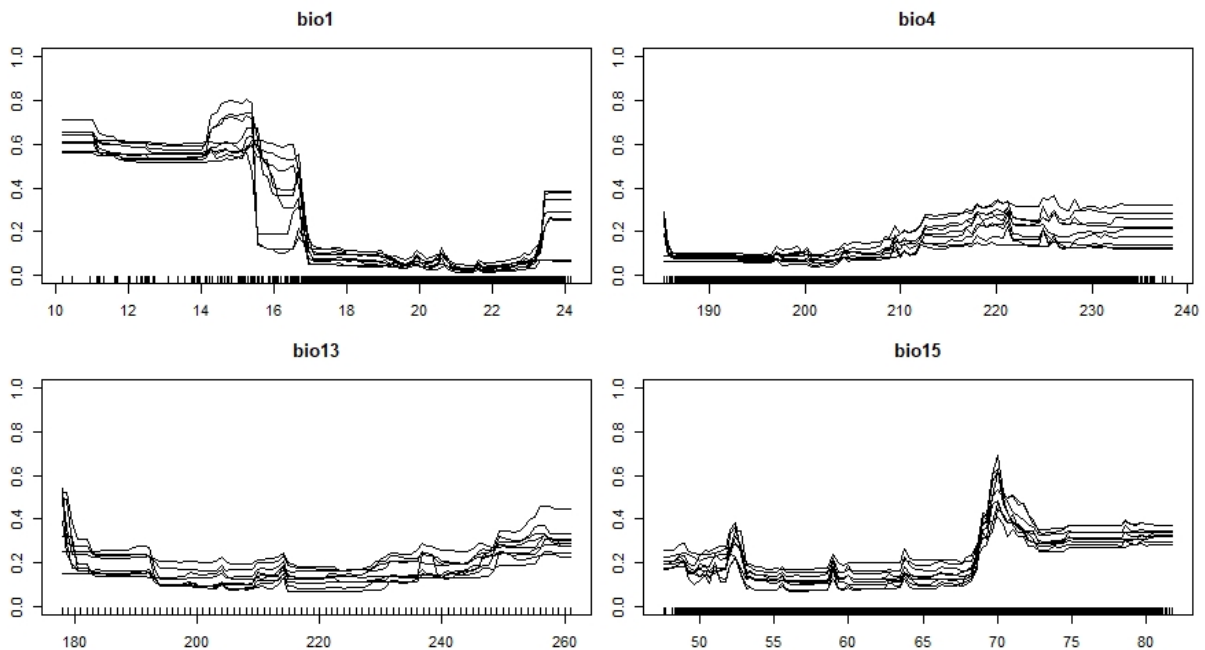


Figure 4: Response Curves from random forest (RF). Where BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

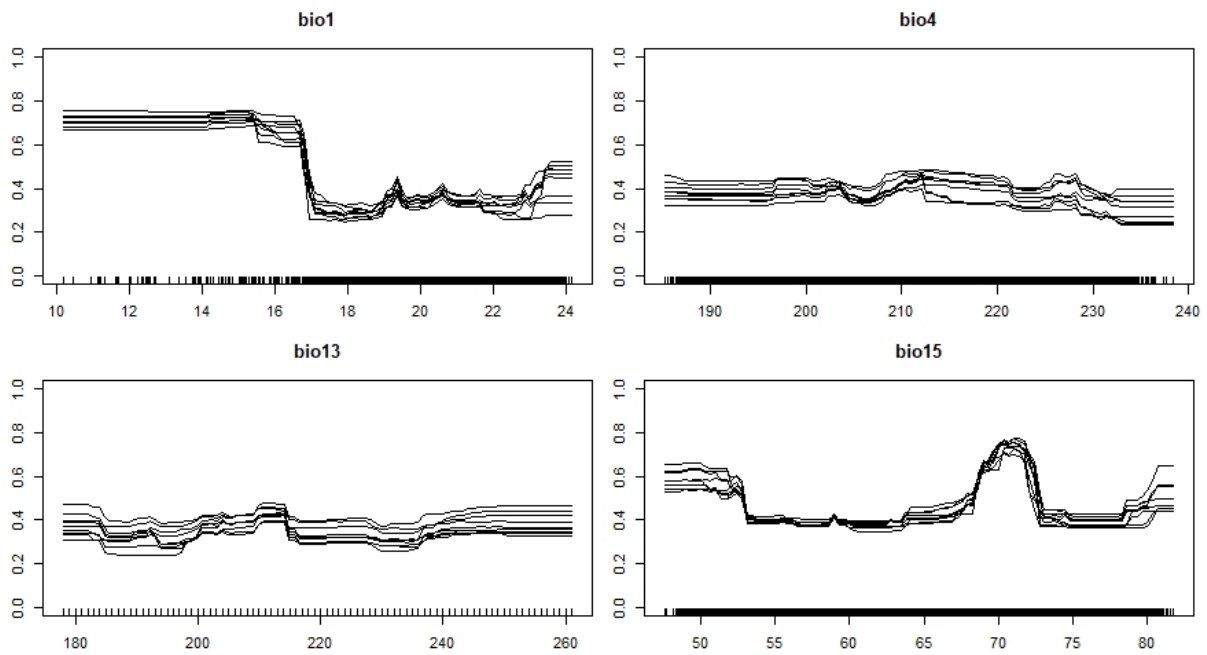


Figure 5: Response Curves from generalised boosting models (GBM). Where BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

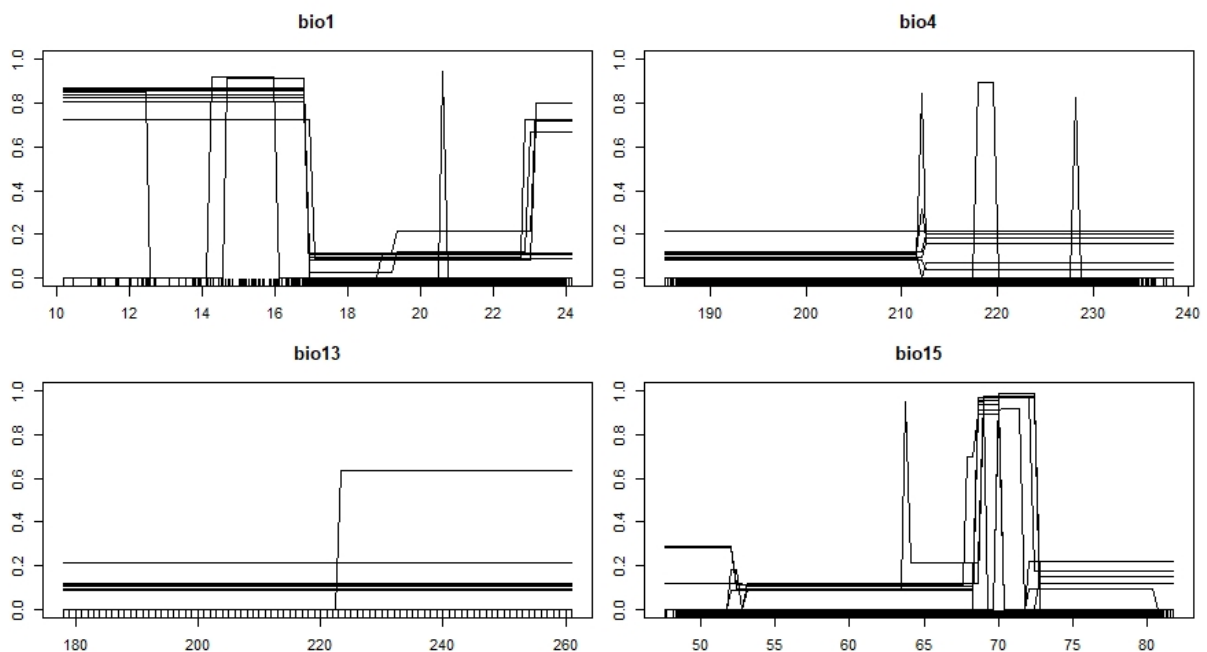


Figure 6: Response Curves from classification tree analysis (CTA). Where BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

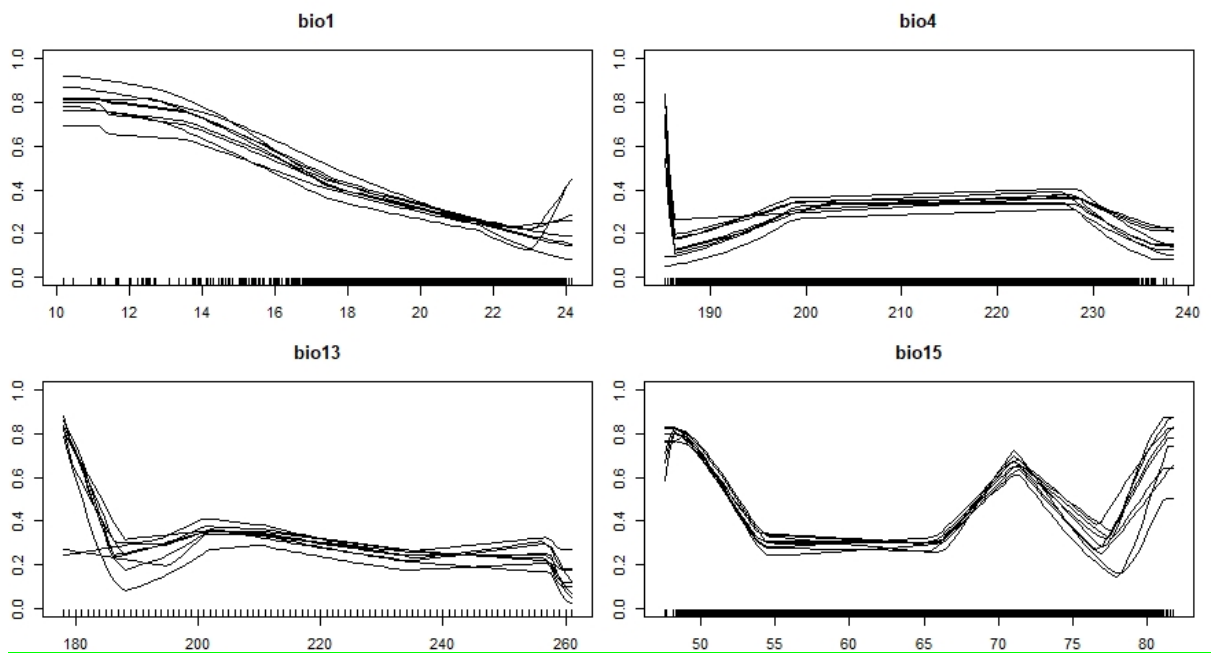


Figure 7: Response Curves from maximum entropy (MAXENT). Where BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

## Part 2: Response curves for landscape suitability models

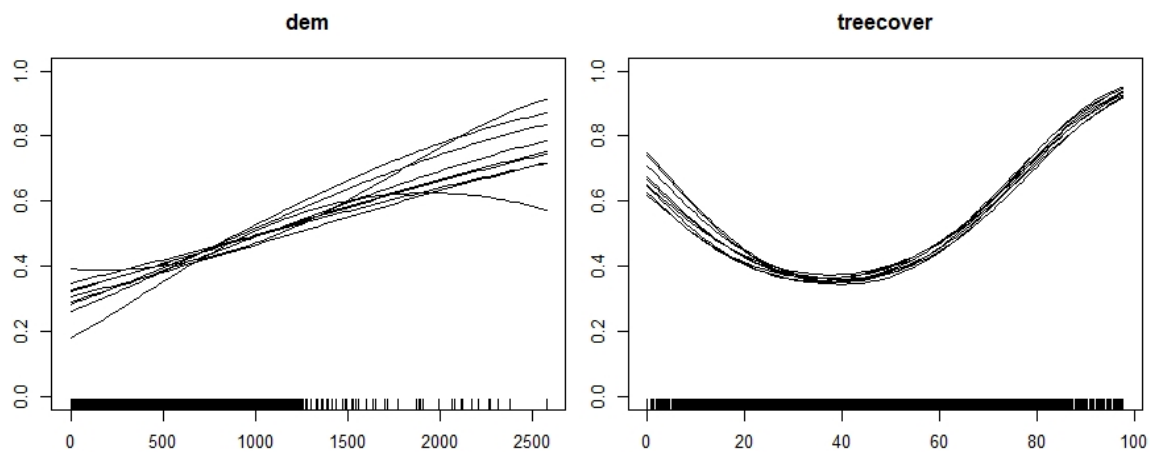


Figure 8: Response Curves plot from generalised linear models (GLM). Where dem: Digital Elevation Model (m) and treecover: Tree Canopy Cover (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

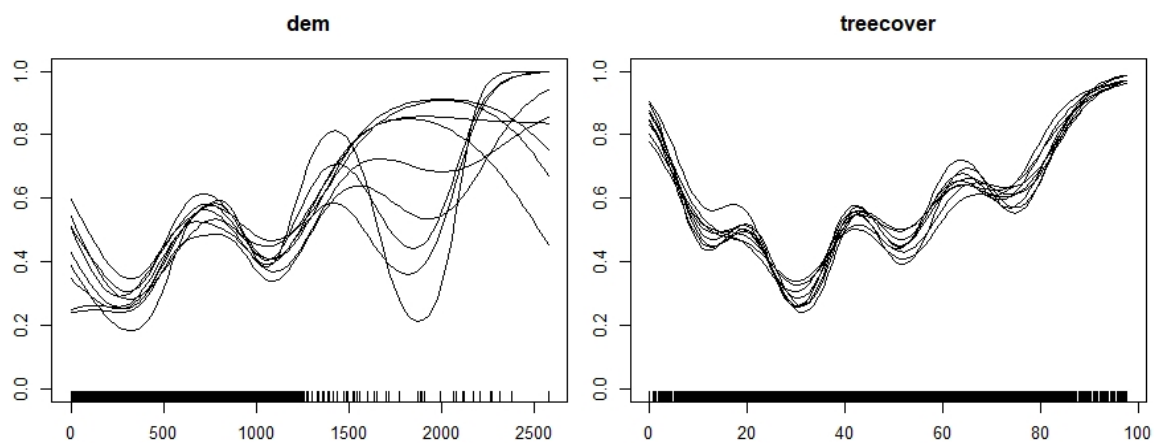


Figure 9: Response Curves plot from generalised additive models (GAM). Where dem: Digital Elevation Model (m) and treecover: Tree Canopy Cover (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

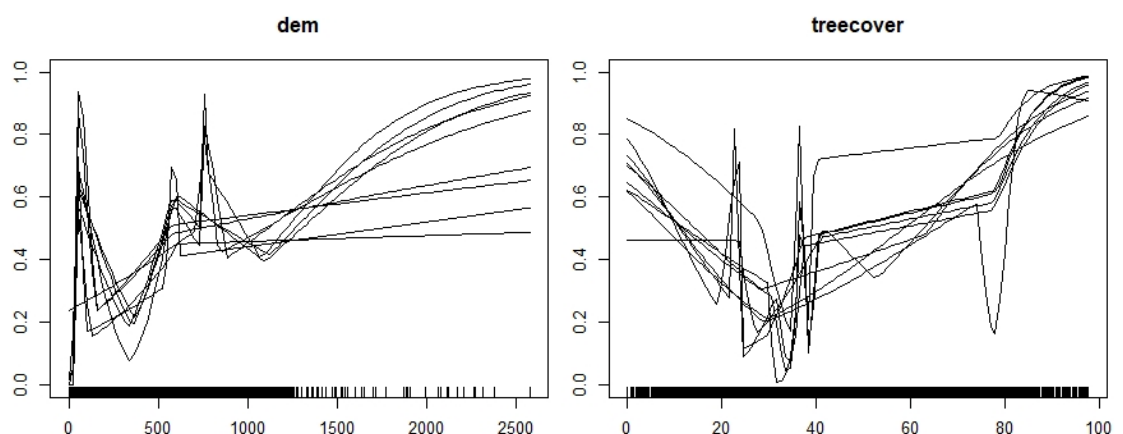


Figure 10: Response Curves plot from multivariate adaptive regression spline (MARS). Where dem: Digital Elevation Model (m) and treecover: Tree Canopy Cover (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

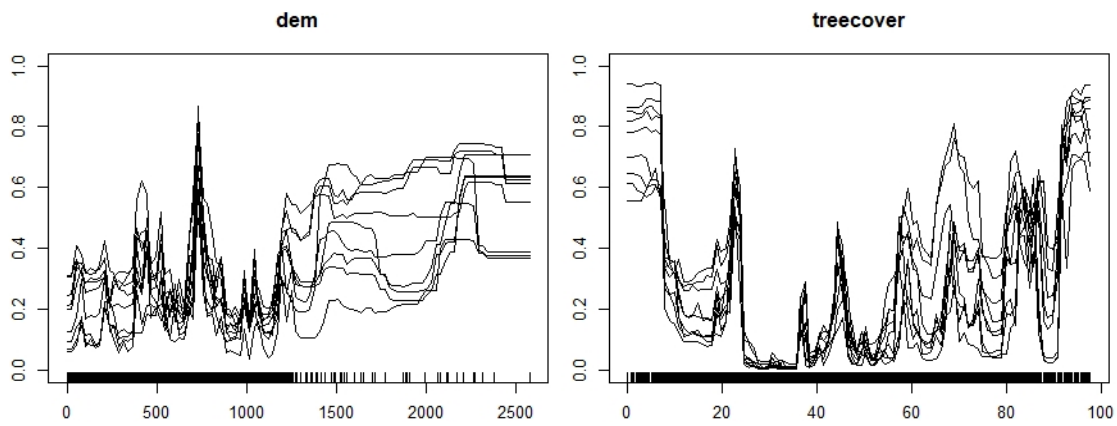


Figure 11: Response Curves plot from random forest (RF). Where dem: Digital Elevation Model (m) and treecover: Tree Canopy Cover (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

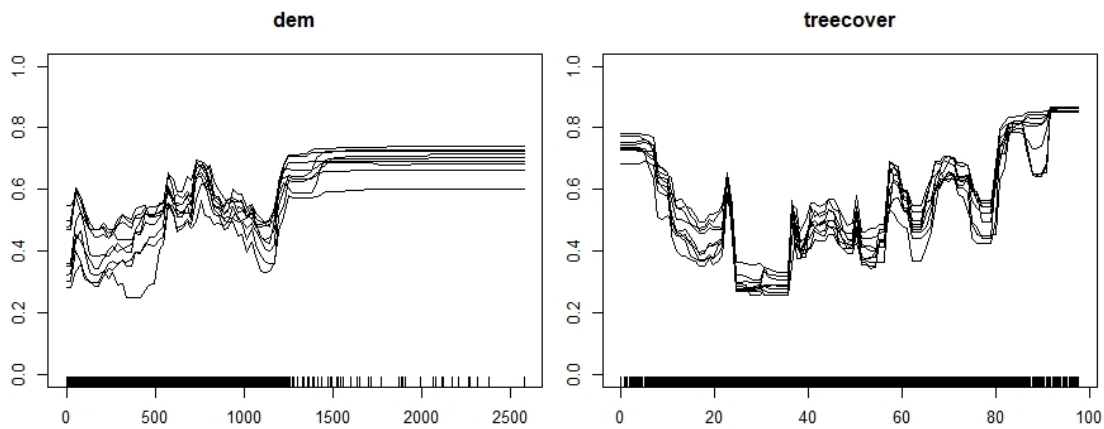


Figure 12: Response Curves plot from generalised boosting models (GBM). Where dem: Digital Elevation Model (m) and treecover: Tree Canopy Cover (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

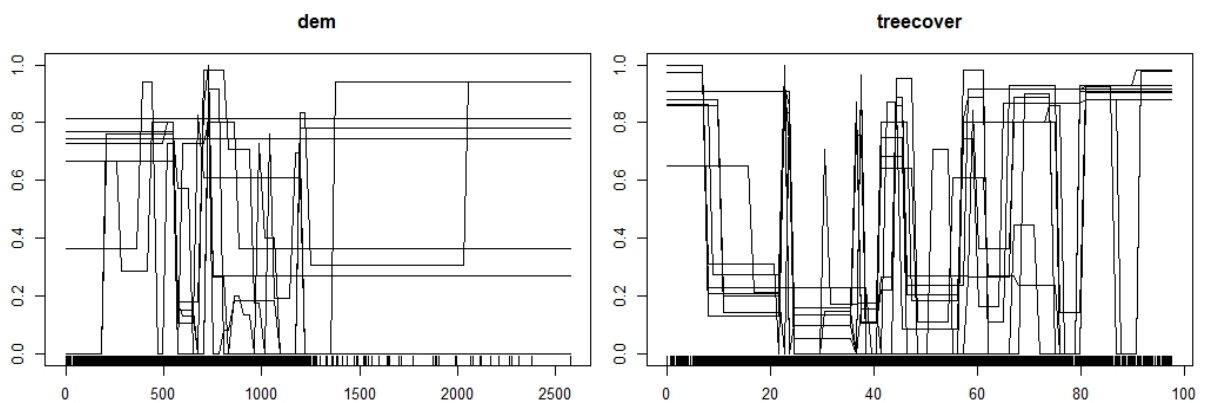


Figure 13: Response Curves plot from classification tree analysis (CTA). Where dem: Digital Elevation Model (m) and treecover: Tree Canopy Cover (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.

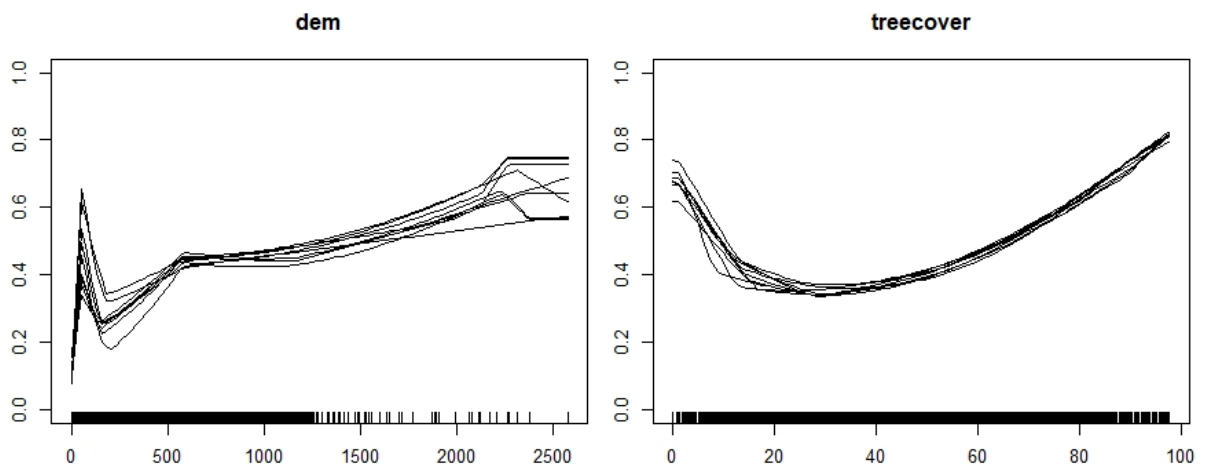


Figure 14: Response Curves plot from maximum entropy (MAXENT). Where dem: Digital Elevation Model (m) and treecover: Tree Canopy Cover (%). Each line represents a different model, with a different pseudo absence selection and cross validation run.



### Part 3: Response curves for overall habitat suitability models

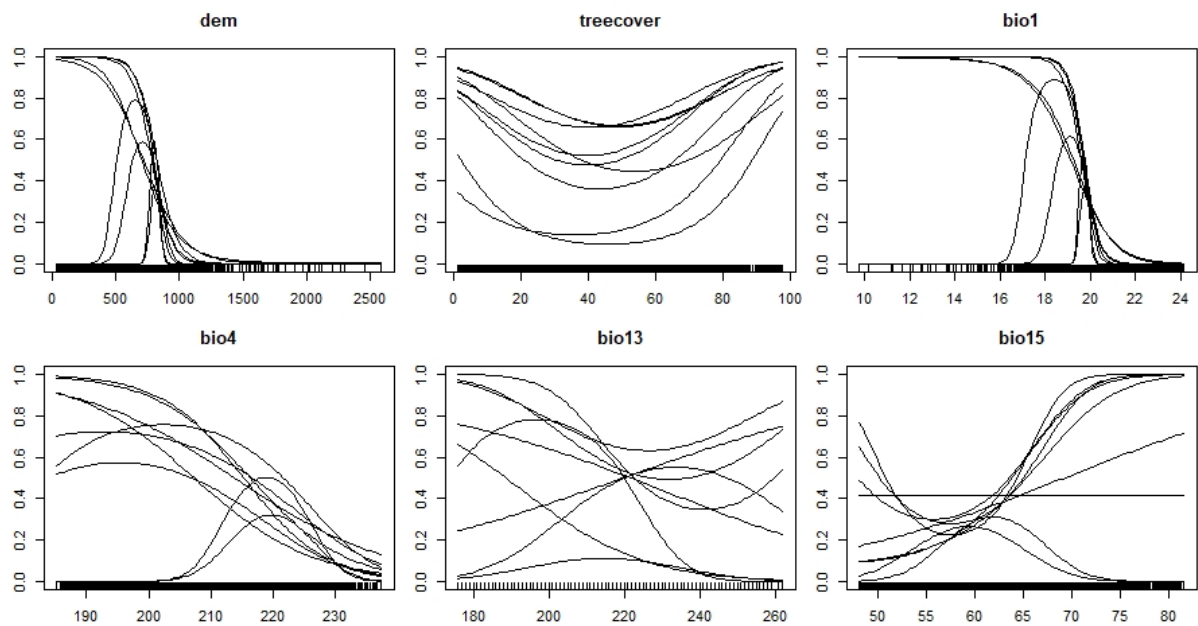


Figure 15: Response Curves plot from generalised linear models (GLM). Where dem: Digital Elevation Model (m), treecover: Tree Canopy Cover (%), BIO1: Annual Mean Temperature ( $^{\circ}$ C), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo-absence selection and cross validation run.

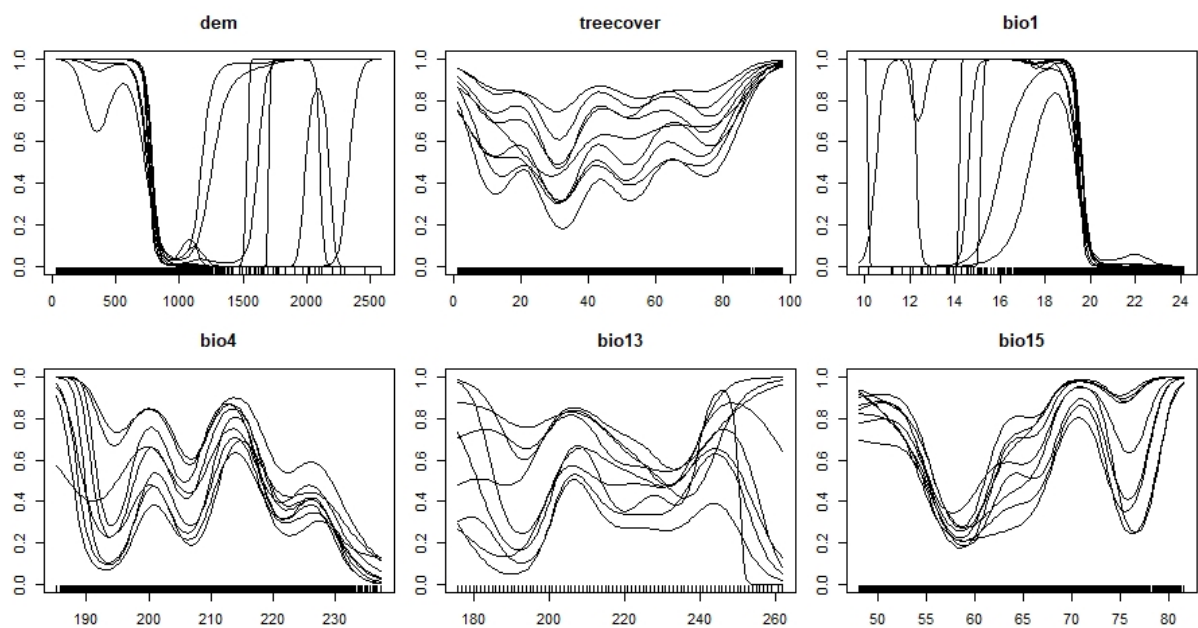


Figure 16: Response Curves plot from generalised additive models (GAM). Where dem: Digital Elevation Model (m), treecover: Tree Canopy Cover (%), BIO1: Annual Mean Temperature ( $^{\circ}$ C), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo-absence selection and cross validation run.

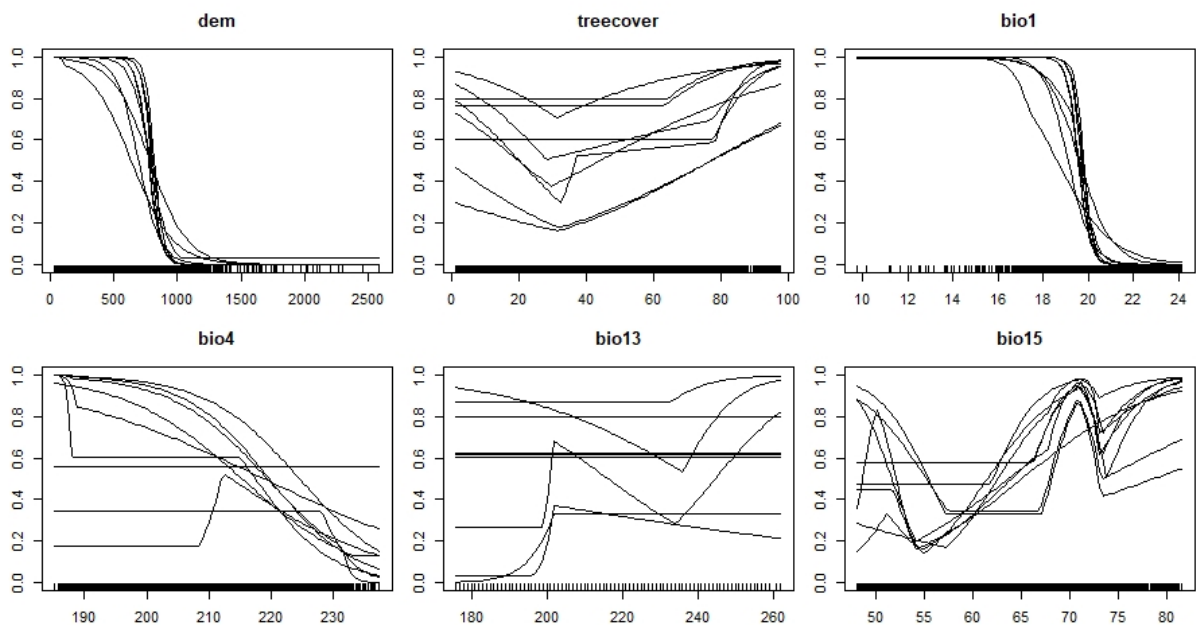


Figure 17: Response Curves plot from multivariate adaptive regression splines (MARS). Where dem: Digital Elevation Model (m), treecover: Tree Canopy Cover (%), BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo-absence selection and cross validation run.

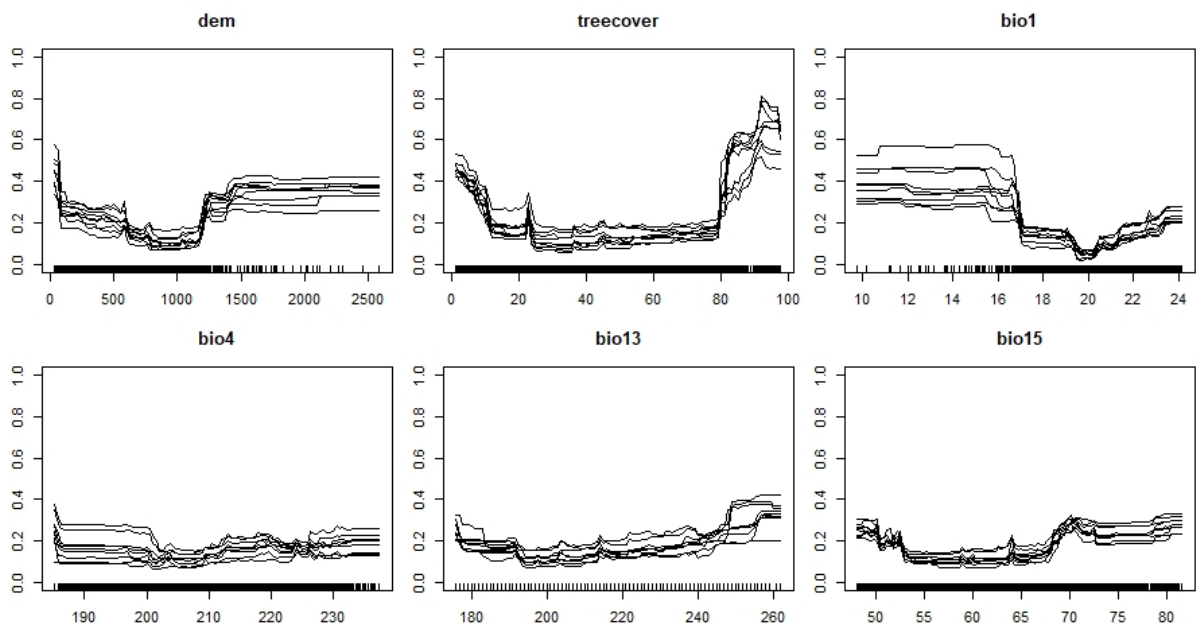


Figure 18: Response Curves plot from random forest (RF). Where dem: Digital Elevation Model (m), treecover: Tree Canopy Cover (%), BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo-absence selection and cross validation run.

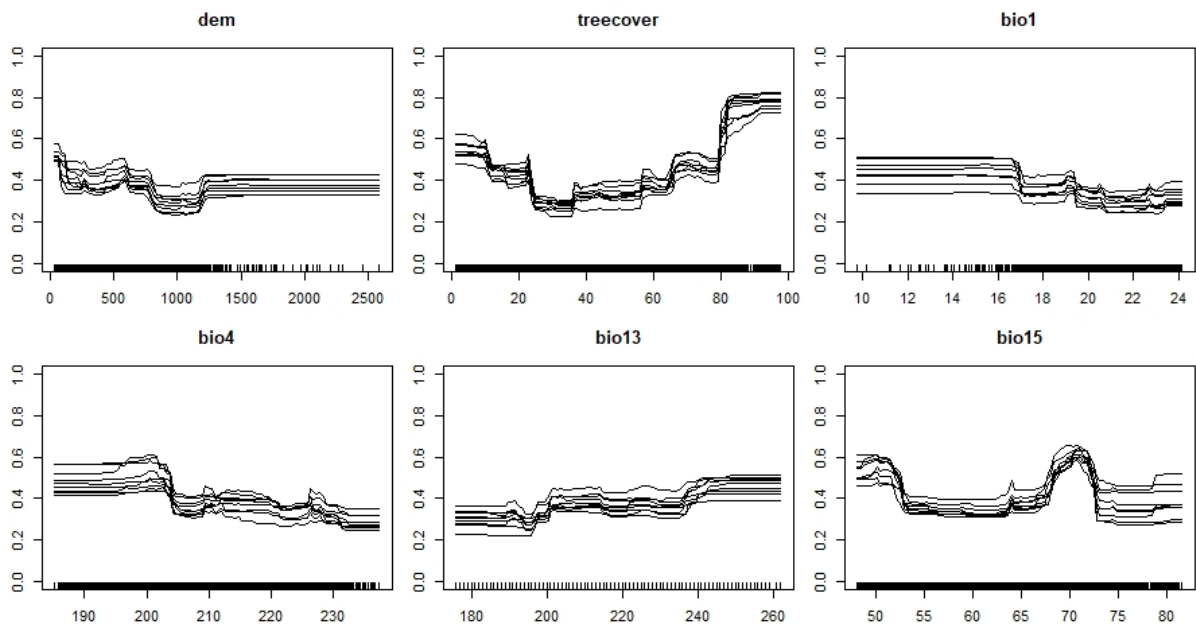


Figure 19: Response Curves plot from generalised boosting models (GBM). Where dem: Digital Elevation Model (m), treecover: Tree Canopy Cover (%), BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo-absence selection and cross validation run.

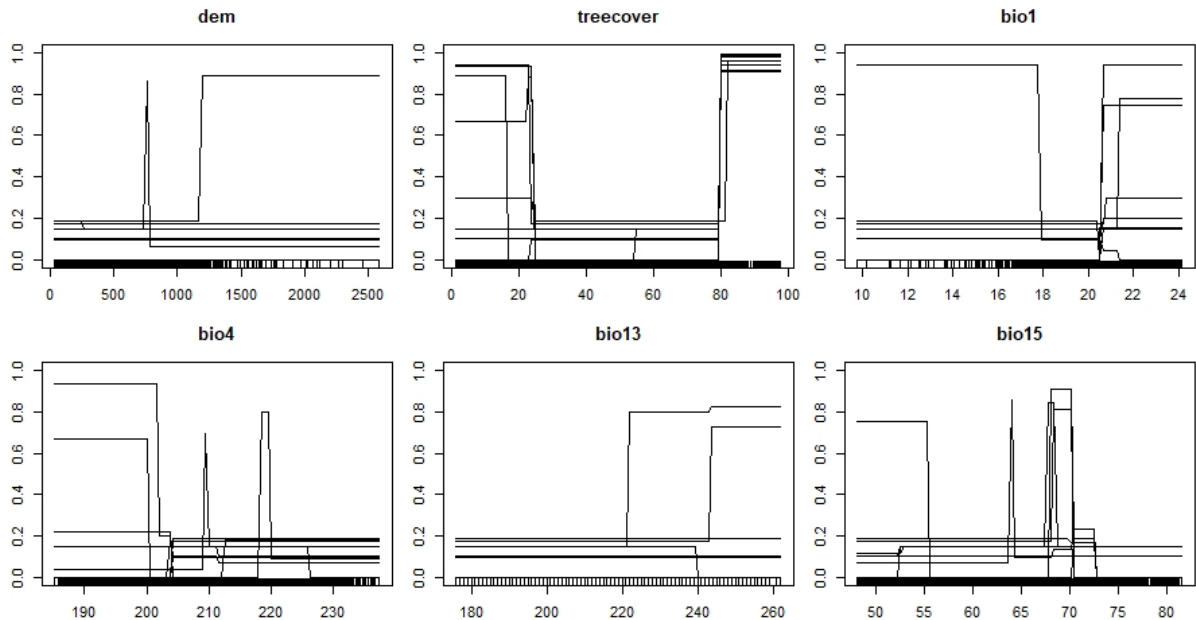


Figure 20: Response Curves plot from classification tree analysis (CTA). Where dem: Digital Elevation Model (m), treecover: Tree Canopy Cover (%), BIO1: Annual Mean Temperature ( $^{\circ}\text{C}$ ), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo-absence selection and cross validation run.

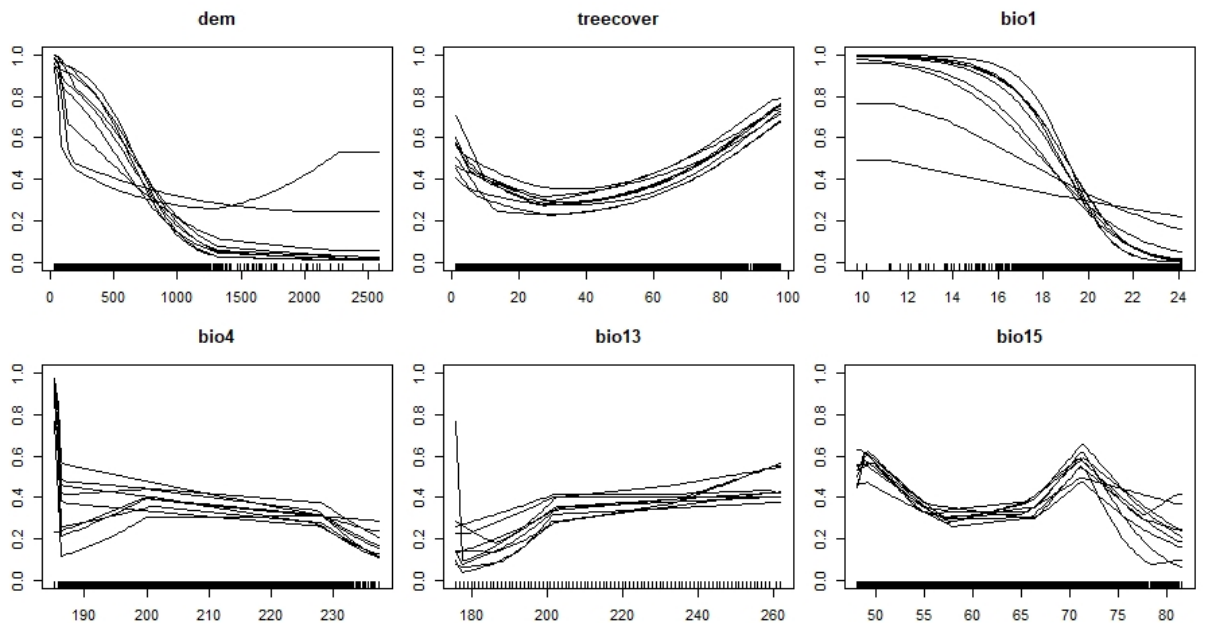


Figure 21: Response Curves plot from maximum entropy (MAXENT). Where dem: Digital Elevation Model (m), treecover: Tree Canopy Cover (%), BIO1: Annual Mean Temperature (°C), BIO4: Temperature Seasonality (standard deviation\*100), BIO13: Precipitation Wettest Month (mm), BIO15: Precipitation Seasonality (%). Each line represents a different model, with a different pseudo-absence selection and cross validation run.

Part 4: Current projection for suitability models

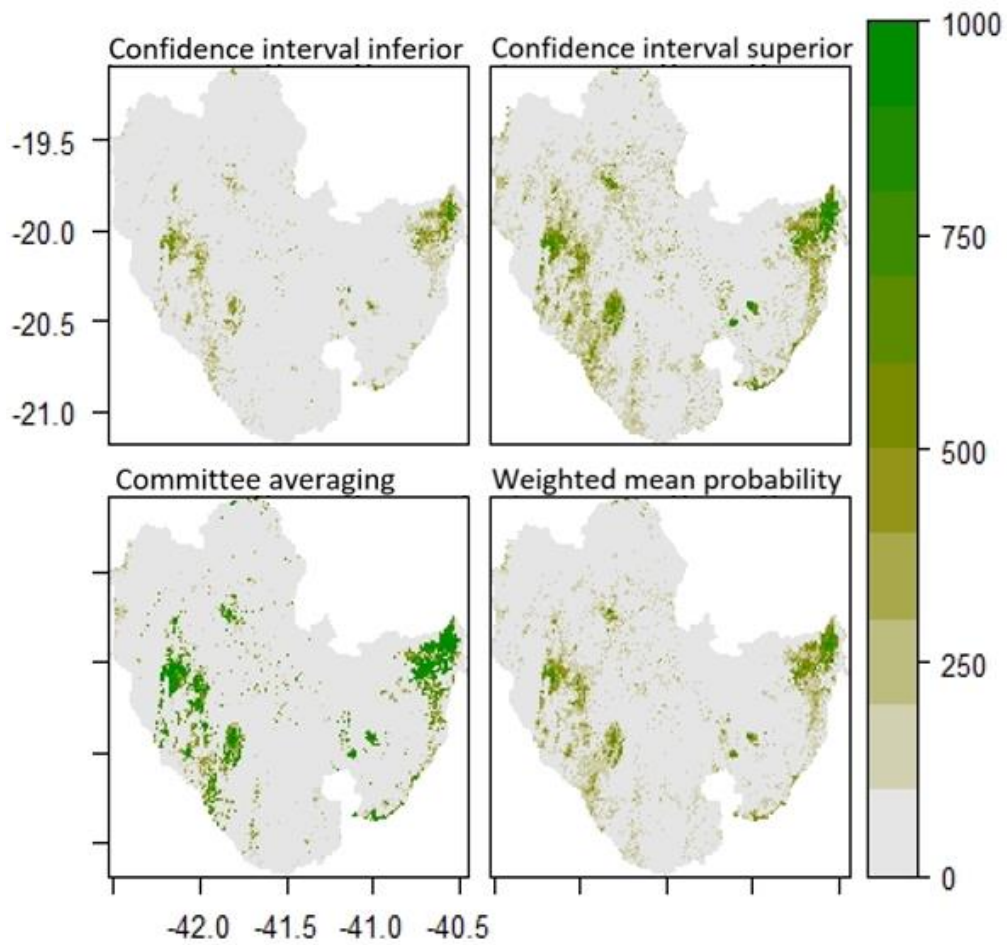


Figure 22: Current prediction of climatic suitability for *Callithrix flaviceps*. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.

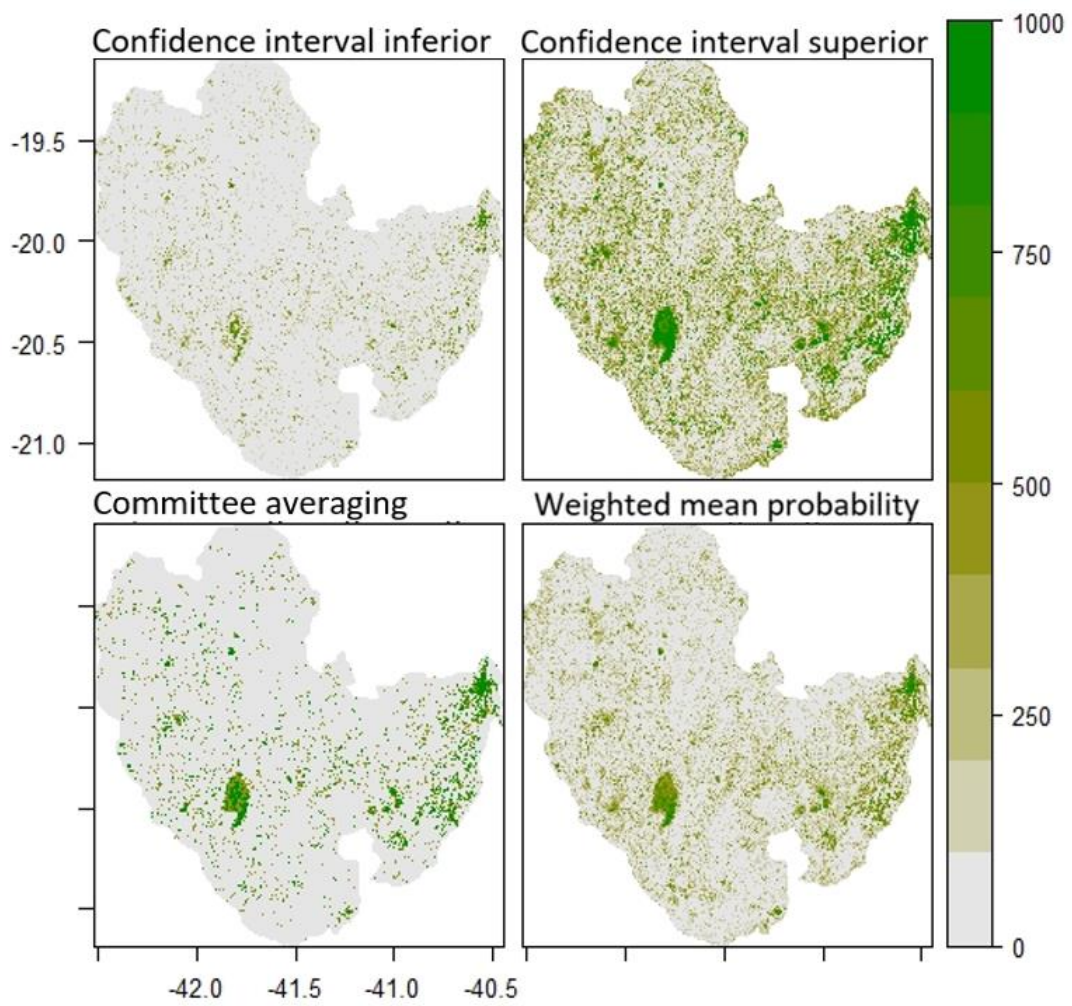


Figure 23: Current prediction of landscape suitability for *Callithrix flaviceps*. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.

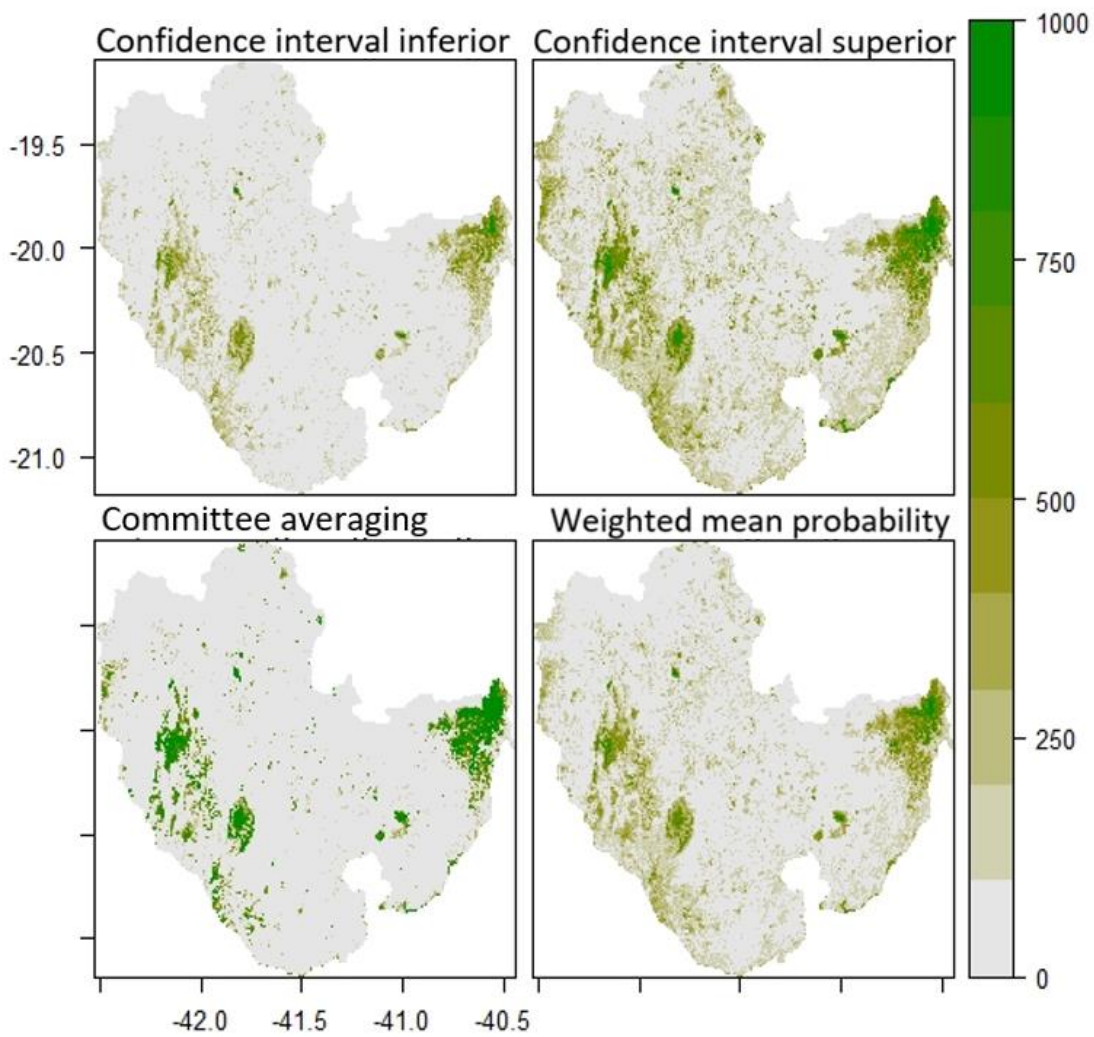


Figure 24: Current prediction of overall habitat suitability for *Callithrix flaviceps*. Each map represents a different Ensemble Model. The 2 tops maps show 2 levels of confidence interval around the mean probability. The committee averaging shows the consistency between models, value of 1 (dark green) means that all models agreed to predict presence, while value of 0 (grey) means that all models agreed to predict absence of the species, yellow/light green shows inconsistency between models. The weighted mean probability represents the actual projection, dark green represents high suitable areas.