

UNIVERSIDADE FEDERAL DE SÃO CARLOS
BACHARELADO EM ENGENHARIA FLORESTAL

GABRIEL DE FREITAS PEREIRA

Climate change impact on *Bertholletia excelsa* distribution

SOROCABA - SP

2023

GABRIEL DE FREITAS PEREIRA

Climate change impact on *Bertholletia excelsa* distribution

Trabalho de conclusão de curso
Bacharelado em Engenharia Florestal
Orientadora: Dra. Karina Martins

SOROCABA - SP

2023

Acknowledgements

This work received financial support from Embrapa on the EcoGenCast project. FAPESP (São Paulo State Research Foundation) provided a scientific initiation scholarship to GFP (Grant 19/25968-4) and only was possible to use occurrences on such a large area with the help provided by the following authors: Beatriz Ribeiro, Daiana Tourne, Daniel Larrea, Daniela Borella, Daniela Granato, Dario Amaral, Diego Souza, Evert Thomas, Karina Martins, Katia Emídio, Lúcia Wadt, Lucieta Martorano, Marcelino Guedes, Quêzia Leandro, Raquel dos Santos, Ricardo Scoles e Walter Sousa.

Authors: Gabriel de Freitas Pereira, Evert Thomas, Viviana Ceccarelli, Tobias Fremout, Daiana Tourne, Lúcia Helena de Oliveira Wadt, Karina Martins

Abstract

Bertholletia excelsa, also known as the Brazil nut tree, stands out as a dominant and safeguarded species within the ecosystem, serving a pivotal function in the cycling of nutrients and the delivery of essential ecosystem services in the Amazon rainforest. This study aimed to model the suitable habitat distribution of this species, by comparing different ensemble models constructed using various spatial filtering resolutions. Additionally, we aimed to include an evaluation of the influence of human activities in promoting or dispersing *B. excelsa* seeds and/or stands. The results highlight the impact of model selection on the present and future projections, with a significant difference between the projections for the 2050s. We emphasise the importance of considering multiple models for different climatic conditions to better assess the uncertainty of future projections when making climate change policies. The best model (AUC = 0.94) was achieved with geographic plus environmental filtering and a resolution of 2.5 arcminutes. The model predicted a contraction of the suitable habitat area for Brazil nut trees in the present and an even more severe scenario in the future, with precipitation seasonality being the most impactful variable in the species distribution. Future studies should also examine the expected impacts of climate change on the seed production of Brazil nut trees, including indirect effects such as changes in pest and disease incidence.

Key-words brazil nut tree, ensemble modelling, habitat suitability modelling, climate change.

Bertholletia excelsa, também conhecida como castanheira-do-brasil, destaca-se como uma espécie dominante e protegida dentro do ecossistema, desempenhando uma função crucial no ciclo de nutrientes e na prestação de serviços ecossistêmicos essenciais na Floresta Amazônica. Este estudo teve como objetivo modelar a distribuição do habitat adequado para a espécie, comparando diferentes ensembles construídos com diversas resoluções de filtragem espacial. Além disso, buscamos avaliar a influência das atividades humanas na promoção ou dispersão de sementes e/ou

populações de *B. excelsa*. Os resultados destacam o impacto da seleção de modelos nas projeções atuais e futuras, com diferenças significativas entre as projeções para os anos de 2050. Enfatizamos a importância de considerar múltiplos modelos para diferentes condições climáticas, a fim de avaliar melhor a incerteza das projeções futuras ao tomar decisões relacionadas a políticas de mudanças climáticas. O melhor modelo (AUC = 0.94) foi obtido com filtragem geográfica e ambiental, com uma resolução de 2,5 minutos de arco. O modelo previu uma contração da área de habitat adequado para as castanheiras-do-brasil no presente e um cenário ainda mais severo no futuro, sendo a sazonalidade de precipitação a variável de maior impacto na distribuição da espécie. Estudos futuros também devem examinar os impactos esperados das mudanças climáticas na produção de sementes das castanheiras-do-brasil, incluindo efeitos indiretos, como mudanças na incidência de pragas e doenças.

Palavras-chave castanheira, método de ensemble, modelagem de adequabilidade, mudanças climáticas.

Summary

Introduction	6
Material and Methods	8
1. Presence data for distribution modelling	8
2. Explanatory variables	9
3. Habitat suitability modelling	10
4. Future projections	13
Results	14
Discussion	15
Conclusions	18
References	20
Supporting information	24

Introduction

The Amazon is the world's largest rainforest and holds the richest biodiversity worldwide, including between 6,000 and 16,000 tree species (CARDOSO et al., 2017; ter STEEGE et al., 2016). More than 15% of its area were deforested during the past 50 years, and a large extent is under strong pressure due to agriculture, illegal mining, and by road and power plant constructions (RAISG 2020). Significant alterations in climate in the last century have been documented in Amazonia, such as an increase in average temperatures (NOBRE et al., 2016), a strengthening of dry events and an expansion in dry-season length in Southern Amazonia (FU et al., 2013; MARENGO et al., 2011). Future projections have confirmed this tendency of drier and longer dry seasons over the Amazon (BOISIER et al., 2015; POKHREL et al., 2014), higher frequency of drought and rainfall extreme events (MARENGO et al., 2011; MARENGO & ESPINOZA, 2016), increased fire activity (FONSECA et al., 2019) and decreased soil moisture (POKHREL et al., 2014), with a higher impact in Southeastern Amazonia (POKHREL et al., 2014). How the tropical tree community would respond to the higher water stress and fire frequency is still under debate (AMIGO, 2020) since studies that evaluate the consequences of climate change on tree composition are lacking. Esquivel-Muelbert et al. (2019) showed a trend of a compositional change towards a more dry-tolerant flora but at a slower pace than climate change.

The Amazon nut or Brazil nut tree, *Bertholletia excelsa* (Lecythidaceae), is a canopy-emergent tree broadly distributed throughout the Amazon Basin and the Guianas (PRANCE 1990), predominantly in non-flooded regions. The Amazon nuts are almost exclusively obtained from natural forests by forest-based communities. The hard indehiscent fruits fall from the crowns and are harvested and opened at the end of the rainy season (WADT et al. 2008) and processed in factories in Bolivia, Brazil and Peru, the main exporters. According to the International Nut and Dried Fruit Council (2021), total Amazon nut world production in 2020/2021 crop summed up 27,400 metric tons, with 77% of exports from Bolivia, followed by Peru (15%) and Brazil (8%). Since thousands of local families in these countries rely a substantial part of their income from the harvesting, processing and commercialization of Amazon nuts (CLAY 1997;

COSLOVSKY 2014; KALLIOLA e FLORES 2011), this species is considered a cornerstone for the conservation of the Amazon Forest (GUARIGUATA et al. 2017).

Due to its wide distribution range, the species is found in a plethora of edaphic, topographic and climate conditions. A Species Distribution Model (SDM) study indicated that some soil and topographic characteristics, and annual potential evapotranspiration are the most important environmental determinants of *B. excelsa* current distribution (TOURNE et al. 2019). Although using a much smaller data set (only 255 presence points) in their SDM, Sales et al. (2020) predicted a greater influence of climate than soil predictors on *B. excelsa* distribution. The authors concluded that temperature seasonality was the most important determinant of *B. excelsa* habitat suitability. Historical climate fluctuations are also predicted to have caused several range contractions and expansions during the Pleistocene which might have significantly impacted *B. excelsa* current distribution. Thomas et al. (2014) predicted that the drier and colder weather in the Amazon during Last Glacial Maximum (LGM, ~21,000 yr BP) resulted in a dramatic reduction in *B. excelsa* habitat suitability. Only a few small, scattered populations were predicted to remain in LGM, mostly in southern Amazonia. As the climate conditions progressively became suitable, the species gradually dispersed from these few refugia and expanded its range.

Evidences from different disciplines suggest that, along with ecological processes determined by past and current environmental conditions, indigenous human populations have substantially influenced not only the current distribution of *B. excelsa* but also population density and structure through incidental or deliberated seed dispersal or by favouring seedling recruitment (GRIBEL et al. 2007; LEVIS et al. 2017; RIBEIRO et al. 2014; SCOLES e GRIBEL 2011; SHEPARD-JR e RAMIREZ 2011; THOMAS et al. 2014, 2015). The oldest documented record of carbonised fruits was dated 11,000 yr BP in a Paleoindian campsite in Eastern Brazilian Amazonia (ROOSEVELT et al. 1996), revealing that indigenous peoples have a longstanding relationship with *B. excelsa*. Thomas et al. (2015) found higher habitat suitability in their SDM and greater stand density of *B. excelsa* in the vicinity of Anthropogenic Dark Earth (ADE) sites, regardless environmental conditions, suggesting that human interventions have influenced *B. excelsa* current distribution, but none studies have ever

included proxies of human interventions in *B. excelsa* SDM to try to figure out the extent of their influence.

Here, we used the dataset of archaeological sites from McMichael et al. (2014) as proxies of human interventions to model *B. excelsa* distribution. The archaeological sites were included as predictors along with a large array of environmental variables. We also rely on the most complete *B. excelsa* occurrence dataset to date, which was augmented from the Tourne et al. (2019) data to include occurrence points from underrepresented locations. However, the great challenge is to cover a significant number of data that are definitely conclusive for interpretation on the estimate of unique genetic diversity of a given species in relation to the different locations of its occurrence. The lack of data related to genomic distribution makes it difficult to successfully determine priority sites for conservation. Therefore, the objective of this study is to (1) Provide an accurate *B. excelsa* suitability map; (2) Evaluate the influence of humans in promoting/dispersing *B. excelsa* seeds and/or promoting *B. excelsa* stands; (3) Predict the effect of climate change on future *B. excelsa* suitability and identify priority areas for conservation.

Material and Methods

1. Presence data for distribution modelling

Presence data were collected from different sources. *Bertholletia excelsa* presence data were compiled from GBIF (Global Biodiversity Information Facility; www.gbif.org), BIEN (Botanical Information and Ecology Network; <http://biendata.org/>), Tourne et al. (2019), and from a variety of authors partners from Brazilian Agricultural Research Corporation (Embrapa). The final data set of *B. excelsa* occurrence points for this study is available at Link S1. We also obtained presence points of all species of Amazon trees and shrubs from the Brazilian checklist (TER STEEGE et al., 2019), to provide better confidence on the *B. excelsa* pseudo-absence data. In short, we used the whole compiled occurrence records as input in the distribution modelling for the selection of environmental variables and for the target-group selection of pseudo-absence and background points, as detailed in the

following sections. We only modelled the output distribution for *Bertholletia excelsa*. The Amazon ecosystem was defined by aggregating the Amazon ecoregions in the Olson terrestrial ecoregions map (OLSON et al., 2001) and all the presence points occur within the Amazon ecosystem.

2. Explanatory variables

Initially we have used the same 15 variables of Tourne et al. (2019) best model, which were selected by a strategy that combined the expertise of scientists experienced in *B. excelsa* ecology and MAXENT (Table S1). Furthermore we also have used our own approach to select other variables to test, which included 35 climate, soil and terrain variables (Table S1), consisting of 19 bioclimatic variables (HIJMANS et al., 2005), annual potential evapotranspiration, aridity index, seven continuous soil variables (HENGL et al., 2017), five terrain variables, geoglyphs sites (MCMICHAEL; PALACE; GOLIGHTLY, 2014) and Anthropogenic Dark Earth sites (MCMICHAEL et al., 2014). Collinear variables under present conditions were removed based on stepwise calculations of variance inflation factors (VIF), retaining only variables with VIFs smaller than 10 (Table 1).

Since the correlations between variables can change in both strength and direction under future climate conditions, correcting for collinearity based exclusively on climate variables under present conditions can substantially impact future habitat suitability projections (BRAUNISCH et al., 2014; DORMANN et al., 2013). Therefore, we also assessed collinearity under future conditions and retained all variables resulting from the stepwise VIF procedure in at least one time horizon-by-emission scenario combination. Multicollinearity analyses were based on the values of the explanatory variables at the locations of the presence and pseudo-absence points used to calibrate the ensemble models. For *Bertholletia excelsa*, we performed the same analysis by using all the locations of the compiled occurrence records within the convex hull surrounding all presence points.

Table 1: Variables selected by VIF (VIF < 10)

Variables	VIF
pH index measured in aqueous solution	7.23
Precipitation Seasonality	6.24
Precipitation of Wettest Month	6.13
Bulk density	5.60
Stream flow power index	5.01
Mean Temperature of Wettest Quarter	4.05
Isothermality	3.95
Precipitation of Warmest Quarter	3.51
Average percentage of coarse fragments	3.46
Precipitation of Coldest Quarter	3.25
Average of silt particles	3.11
Mean Diurnal Range	2.92
Cation exchange capacity	2.65
Slope	2.63
Temperature Seasonality	2.55
Soil organic carbon content	2.14
Average percentage of clay particles	1.47
Geoglyph records	1.09
Dark Earth records	1.14
Terrain flow	1.03
Topographic position index	1.03
Terrain aspect	1.00

3. Habitat suitability modelling

For the suitability modelling, we used the presence points of *B. excelsa* at 2.5 arcminutes resolution (ca. 4 km at Equator). This modelling resolution was defined after testing the resolutions 2.5, 5, and 10 arcminutes using the selected variables of Tourne et al. (2019) (Table S1). The 2.5 arcminutes resolution was chosen because it was more

consistent when models built under the ensemble algorithm were compared. . For each tested resolution and for all the models presented in this study, two different areas were modelled in order to cover major climate changes, which were: Pan-Amazon extent and larger Pan-Amazon extent (here and on referred to as LAM, see the extent difference between areas on the Count maps from Link S2). And also two different filtering approaches were developed: geographical (or spatial) filtering and a combination of geographical and environmental filtering (referred as environmental). Therefore, the geographical filtering was applied to reduce the effects of spatially biased presence points on model calibration (KRAMER-SCHADT et al., 2013; TOURNE et al., 2019) and was implemented by randomly retaining one presence point per grid cell for each of the selected resolutions of spatial filtering. As model accuracy metrics are easily inflated when the geographical, and thereby the environmental, range of absence or pseudo-absence points is too wide (CHEFAOUI; LOBO, 2008), we selected pseudo-absence points from the area enclosed by a convex hull polygon around all presence points extended with a buffer corresponding to 10% of the polygon's largest axis (ACEVEDO et al., 2012). We also restricted the convex hull area to sites within the altitudinal range of the species, calculated as the maximum altitude among the presence points plus 200m (hereafter called the 'convex hull area'). This procedure aimed to improve the ability of the suitability models to accurately distinguish between presence and absence at lower altitudes. We separately selected pseudo-absence points for the presence-absence algorithms in our ensembles and background points for the Maxent algorithm. Pseudo-absence and background points were selected from the convex hull area according to target group method (MATEO et al., 2010; PHILLIPS et al., 2009). This method involves the selection of points from grid cells with presences of species belonging to a similar group as the target species. These locations are expected to reflect a similar bias to the sampling bias of the target species, therefore reducing the effects of spatially biased presence points on model calibration (PHILLIPS et al., 2009). Pseudo-absence points were selected from grid cells that did not contain any presence point of the modelled species but where a reasonable number of tree and shrub species has been recorded by past biodiversity surveys (i.e. above ecoregion-specific minimum sampling intensities specified below). This strategy provides some confidence on the species absence because in locations where several species have been sampled but the

modelled species, it's assumed that the species is likely to be absent, while in locations where only few species have been sampled, we cannot be confident on absence as the modelled species could have been missed because of insufficient sampling. In environmental filtering, instead of using geographic cells, cells are constructed around the principal components generated with Principal Component Analysis (PCA) performed with environmental variable values. Then, one point is selected for each of these "environmental" cells. Thus, for example, it is possible to ensure that the points are spread between lower and higher precipitation, rather than being clustered in high precipitation areas.

The sampling intensity was estimated as the number of species records per grid cell, using all the compiled occurrence records obtained from all tree and shrub species from the Brazilian checklists. Because species richness varies by ecoregion, minimum sampling intensities were set as the 90% percentile of species numbers per grid cell in each of the Olson terrestrial ecoregions. Background points were selected from grid cells with records of at least one tree or shrub species. As background points do not distinguish between presence and absence, background points were selected from any grid cell with at least one species without using a threshold for the sampling intensity. From here on, both pseudo-absence and background points are referred to as 'pseudo-absences' for brevity. Pseudo-absence and background points selected with the target group were filtered to the 2.5 arcminutes resolution to reflect the spatial bias of the filtered presences.

Suitability modelling was performed using ensembles composed of up to 10 modelling algorithms, as implemented in the BiodiversityR package for R (KINDT, 2018). The modelling algorithms considered were: maximum entropy algorithm (MAXENT); random forests (RF); generalised boosted regression models (GBM); generalised linear models (GLMSTEP, GLMNET); generalised additive models (GAMSTEP, MGCV); flexible discriminant analysis (FDA); support vector machines (SVM); multivariate adaptive regression splines (EARTH). The algorithms in the ensemble models were cross-validated using spatial blocks with the blockCV package for R (VALAVI et al., 2019), following Fremout et al. (2020). The weights of the algorithms in each of the ensemble models were optimised using the ensemble.tune

function of the BiodiversityR package, which maximises the AUC value of ensemble models by optimising the weights of the different algorithms through a factorial procedure (KINDT, 2018). With the spatial block cross-validation, presence and pseudo-absence points were partitioned in training and testing points using a set of spatial blocks. This strategy has the advantage that the obtained AUC values provide a better measure of model transferability, which is crucial when projecting species distributions to future climates (MUSCARELLA et al., 2014; WENGER; OLDEN, 2012). Presence and pseudo-absence points were partitioned using 10 folds, each fold consisting of one or more 100 km-wide squared blocks. The ensemble models were then calibrated again using all presence and pseudo-absence points and the final ensemble suitability predictions were made using the previously optimised weights.

4. Future projections

With all present models except the model using only the variables from Tourne et al. (2019), we made future projections for the period 2050s (average for 2040-2069) for two Socioeconomic Shared Pathways (SSP2-4.5 and SSP5-8.5) from Coupled Model Intercomparison Project Phase 6 (CMIP6) and also using the previous version (CMIP5) with Representative Concentration Pathways (RCP4.5 and RCP8.5). The older version was also considered due to the drastic predictions we had and initially aiming to confirm what we found. The SSP2-4.5 scenario represents the medium part of the range of future forcing pathways and updates the RCP4.5 pathway (O'NEILL et al., 2016). It is considered the emission scenario that fits most closely to the scenario in which all countries would comply with their Intended Nationally Determined Contributions of the Paris Agreement (TRIBETT et al., 2017), whereas the SSP5-8.5 scenario represents a worst-case scenario (updated version of RCP8.5). Future climatic conditions were characterised using eight General Circulation Models (GCMs): (BCC_CSM2_MR, CanESM5, CNRM_CM6_1, CNRM_ESM2_1, IPSL_CM6A_LR, MIROC_ES2L, MIROC6 and MRI_ESM2_0). They were selected from all the downscaled GCMs available at the WorldClim from the respective versions (WorldClim 2.1 - CMIP6 projections and WordlClim 2.1 - CMIP5 projections). The suitability maps for the present and the future projections were converted to presence-absence maps by using the suitability threshold at which model sensitivity and specificity are equal, which is

one of the two threshold selection methods recommended by Jiménez-Valverde and Lobo (2007). The presence-absence maps were masked by the convex hull of the species to exclude areas without any occurrence records. Suitable conditions under future climatic conditions were predicted in grid cells where at least 75% of the GCMs coincide.

The definition of priority areas for conservation was based on the future projections of habitat loss and suitability changes for Brazil nut trees. These projections guided the identification and selection of areas that are likely to be most adversely affected by changing conditions. By focusing on these priority areas, conservation efforts can be strategically directed towards preserving and restoring critical habitats, ensuring the long-term survival and resilience of the Brazil nut tree species and the broader ecosystem.

Results

Among the three resolutions tested (2.5, 5, 10 arcminutes) with the variables of Tourne et al. (2019), the best one was 2.5 arcminutes since the resulting models reached greater AUC values (0.87 when using spatial filtering and environmental filtering, see Link S3). Despite the commendable AUC scores recorded by these models, their projection was deemed inappropriate due to the inherent limitations posed by the inclusion of only three bioclimatic variables (Table S1), which would lead to a poor prediction of *B. excelsa* future distribution. Besides that, the models developed using the variables selected by the VIF method showcased even more promising outcomes, as elucidated in Link S4.

Consequently, the most optimal model was achieved through the implementation of geographic plus environmental filtering at a resolution of 2.5 arcminutes referring strictly to the area of Pan-Amazon for modelling. This refined model reached an elevated AUC value of 0.94, emphasising its efficacy in capturing the intricacies of habitat suitability for Brazil nut trees in the face of changing climatic conditions (Figure 1). Notably, within the spectrum of models developed, those harnessing CMIP5 data generally yielded inferior AUC scores compared to models employing the current

iteration (CMIP6). For an interactive display of the models developed for each version, alongside their corresponding AUC values, please refer to Link S4.

The ensuing projections bear implications of extensive areas potentially becoming unsuitable for the species by the year 2050, even under the most optimistic scenario (SSP2-4.5) (Figure 1). It is worth noting that the eastern/northeastern regions of the Amazon, along with the northwest and the northern of Amazon, emerge as the areas most significantly affected by climatic changes. Precipitation seasonality emerged as the most influential variable within the model, thereby underscoring its pivotal role in shaping these projections (Figure S3). Eminently, the variable importance analysis unveiled compelling insights, with geoglyphs ranking as the sixth most important variable, and the significance of Anthropogenic Dark Earth Sites, further reinforcing their role as a crucial factor in the model's predictions (Figure S3).

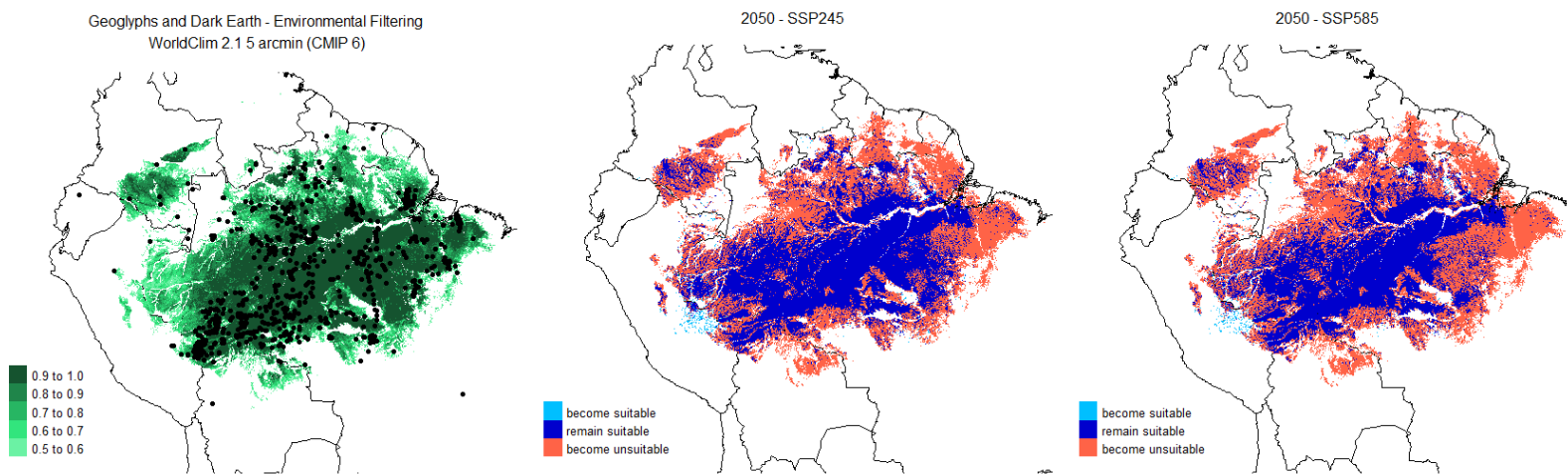


Fig 1. Model using geographic plus environmental filtering with a resolution of 2.5 arcminutes (AUC=0.94), and its future projections in the SSP2-4.5 and SSP5-8.5 scenarios for the average period of 2050.

Discussion

The best model yielded an impressive AUC of 0.94, showcasing significant improvement when compared with Tourne et al. (2019) best model, where the highest

AUC was 0.80. This marked enhancement can be attributed to the more robust approach used in the present study, which involved ensemble techniques that combined results from 10 algorithms, instead of only one (MAXENT) of Tourne et al. (2019). Additionally, our model benefited from a larger database consisting of 4200 presence points, in contrast to other studies such as Thomas et al. (2014), with 936 points, Tourne et al. (2019), with 3252 points, and Sales et al. (2020) with 255 points. This large data volume contributed to the model's accuracy and consistency in predictions, resulting in a superior AUC compared to existing literature.

Furthermore, our approach incorporated a broader array of variables (Table 1), which, although not always indicative of model improvement, allowed us to include as predictors proxies of human interventions, which have been hotly debated in the literature as potential strong determinants of *B. excelsa* distribution. Notably, the inclusion of Geoglyph records proved significant in assessing variable importance within the model as much as Anthropogenic Dark Earth sites (Figure S3). These sites are characterised by their fertile soils in the Amazon rainforest, and offer insight into the agricultural practices of pre-Columbian societies. The high potential of ADEs for sustainable agriculture and carbon sequestration has been demonstrated (BARLOW et al., 2018), further emphasising their impact on *B. excelsa* models. Their contribution aids in understanding the geographic distribution of ancestral populations, underscoring the urgency of protecting these areas and raising awareness to counteract human activities conflicting with environmental policies. Beyond that, it is important to mention that water availability played a critical role in the model as underscored by the variable importance analysis. Noticeably, the first three variables—precipitation seasonality, stream power index, and pH in H₂O—directly correlate with this essential factor. This highlights the significance of water-related parameters in shaping habitat suitability for Brazil nut trees. By prominently featuring these variables, the model inherently acknowledges the profound influence of water dynamics on the species' distribution and underscores the need to consider hydrological factors when assessing and conserving their habitats. This insight is crucial for future projections, particularly given that climate change has markedly altered rainfall patterns in the Amazon region (BOISIER et al., 2015; DE SOUZA et al., 2016; POKHREL et al., 2014). This

emphasises the importance of considering multiple bioclimatic variables and diverse scenarios in species suitability modelling in the Amazon. Such comprehensive approaches could enhance model performance, resulting in more reliable predictions.

As mentioned before, in our best model (Figure 1), the primary variable influencing habitat suitability was precipitation seasonality. Interestingly, certain variables that were less prominent in the best model of Tourne et al. (2019) gained significance here, with the top six influential variables in our model being: precipitation seasonality, stream power index, pH in H₂O, volume percentage of coarse fragments, temperature seasonality and geoglyphs. This subtle reordering of influential variables serves to underscore the intricate nature of the species-environment relationship. It also highlights the potential for variability across distinct models. This shift in results when compared to previous studies can likely be attributed to a range of factors, including variations in data sources and modelling techniques. These differences, however, offer valuable insights into the multifaceted factors influencing *B. excelsa*'s distribution. As an example, it is possible to highlight the topographic variable, which was one of the less influential variables differently from other studies such as Tourne et al. (2019), where it was really impactful.

Moving on to the projected outcomes, the projected substantial loss of suitable habitat displayed consistent behaviour across all future scenarios explored in this study. It indicates a significant reduction in the occurrence area, with displacement towards the central and southwestern regions of the Amazon. Notably, even in the face of this grim scenario, the best model with the highest AUC (0.94 - Figure 1) predicts a comparatively smaller unsuitable area than other models. Nonetheless, it is vital to acknowledge that future climate change may lead to local extinctions and population isolation as indicated by the model results, raising an alarm for the species, especially concerning the 2050s. Hence, the imperative to designate priority conservation areas becomes clear, especially for regions expected to encounter future losses. Eminently, the eastern/northeastern Amazon areas, along with the northwest and northern reaches of the Amazon, emerge as the zones most profoundly affected by shifting climate patterns. These particular regions stand out as high-priority candidates for conservation, primarily due to their heightened vulnerability to the anticipated effects of climate

change. Safeguarding these locales becomes pivotal in upholding the Brazil nut tree's habitat, as they are poised to undergo more pronounced climate shifts, potentially compounding the existing challenges faced by the species. In light of this, channelling conservation efforts towards these regions could make a significant contribution to the Brazil nut tree's long-term viability amidst the evolving climatic panorama. This proactive strategy entails pinpointing and protecting habitats that could become increasingly pivotal for the species' survival and interconnectivity. By doing so, the aim is to mitigate potential repercussions and ensure the endurance of Brazil nut trees and their encompassing ecosystems in the midst of ever-changing environmental dynamics.

Conclusions

We modelled the suitable habitat distribution of *Bertholletia excelsa* by comparing 10 ensemble models constructed using different spatial filtering resolutions, projected for two future scenarios. Therefore we found that model selection impacted present and mainly future modelling results, with a huge difference between the projections for the 2050s. This underscores the significance of customising the modelling methodology to align with the goals and attributes of the focal species. Employing various models to project into forthcoming climatic scenarios is crucial to enhancing the evaluation of projection uncertainties. These uncertainties, in turn, must be taken into account when making decisions pertinent to climate change policies (THUILLER et al., 2019). Moreover, we could evaluate the influence of human activities on species distribution by investigating the Anthropogenic Dark Earth sites and Geoglyphs. This analysis would underscore the importance of these areas in shaping the distribution of the Brazil nut tree, yielding favourable outcomes for the species.

In conclusion, our modelling efforts reveal a current contraction in the suitable habitat for Brazil nut trees, with a more concerning outlook for the future. Particularly, our findings underscore the significance of prioritising conservation efforts, mainly for regions projected to face heightened impacts. The eastern/northeastern Amazon regions, alongside the northwest and northern reaches of the Amazon, emerge as the most

vulnerable to climate changes. By directing conservation initiatives towards these areas, we can proactively address the potential challenges and ensure the long-term survival of the Brazil nut tree amidst evolving environmental conditions. It's worth noting that suitability, as revealed by our models, doesn't necessarily correlate directly with production quantity or quality, as highlighted by RAMIREZ-VILLEGAS et al. (2013). Therefore, as an avenue for future exploration, further studies should delve into the projected effects of climate change on Brazil nut tree production. This examination should encompass not only direct impacts but also indirect consequences, such as potential alterations in the incidence of pests and diseases. Such comprehensive research will provide a more holistic understanding of the potential ramifications of climate change on this crucial species.

References

- ACEVEDO, P. et al. Delimiting the geographical background in species distribution modelling. **Journal of Biogeography**, 2012.
- BARLOW, J., et al. A large-scale assessment of plant dispersal in Amazonian secondary forests. **Journal of Ecology**, 106(1), 238-249, 2018.
- BRAUNISCH, V. et al. Temperate mountain forest biodiversity under climate change: Compensating negative effects by increasing structural complexity. **PLoS ONE**, 2014.
- CHEFAOUI, R. M.; LOBO, J. M. Assessing the effects of pseudo-absences on predictive distribution model performance. **Ecological Modelling**, 2008.
- CLAY, J. W. “Brazil nuts: the use of a keystone species for conservation and development”. P. 246–82 in Harvesting wild species implications for biodiversity conservation, organizado por C. H. Freese. Baltimore: **John Hopkins University Press**, 1997.
- COSLOVSKY, V.. “Economic Development without Pre-Requisites: How Bolivian Producers Met Strict Food Safety Standards and Dominated the Global Brazil-Nut Market”. **World Development** 54:32–45, 2014.
- DORMANN, C. F. et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. **Ecography**, 2013.
- FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, 2017.
- FREMOUT, T. et al. Mapping tree species vulnerability to multiple threats as a guide to restoration and conservation of tropical dry forests. **Global Change Biology**, 2020.
- GRIBEL, R., M. R. LEMES, L. G. PINTO A.E., e G. H. SHEPARD-JR. “Phylogeography of Brazil-nut tree (*Bertholletia excelsa*, Lecythidaceae): evidence of human influence on the species distribution”. in Association for Tropical Biology and Conservation Annual Meeting. **Morelia**, Mexico, 2007.

GUARIGUATA, M R., PETER C, AMY E. DUCHELLE, e PIETER A. ZUIDEMA. “Revisiting the ‘cornerstone of Amazonian conservation’: a socioecological assessment of Brazil nut exploitation”. **Biodiversity and Conservation**, 2017.

HENGL, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. **PLoS ONE**, 2017.

HIJMANS, R. J. et al. WORLDCLIM - a set of global climate layers (climate grids). **International Journal of Climatology**, 2005.

(INC), International Nut and Dried Fruit Council. 2021. **Nuts & dried fruits statistical yearbook - 2020/2021**.

INGLIS, P.W., PAPPAS, M.C.R., RESENDE, L.V., and GRATTAPAGLIA, D.. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. **PLoS One** 13, e0206085, 2018.

KALLIOLA, R., e P. FLORES. “Brazil nut harvesting in Peruvian Amazonia from the perspective of ecosystem services”. **FENNIA** 189(2):1–13, 2011.

KINDT, R. Ensemble species distribution modelling with transformed suitability values. **Environmental Modelling and Software**, 2018.

KRAMER-SCHADT, S. et al. The importance of correcting for sampling bias in MaxEnt species distribution models. **Diversity and Distributions**, 2013.

LEVIS, C., F. R. et al. “Persistent effects of pre-Columbian plant domestication on Amazonian forest composition”. **Science** 355(6328):925–31, 2017.

LOMBARDO, U., C. MCMICHAEL, e E. K. TAMANAHA. “Mapping pre-Columbian land use in Amazonia”. **Past Global Change Magazine** 26(1):14–15, 2018.

MATEO, R. G. et al. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. **Diversity and Distributions**, 2010.

MCMICHAEL, C. H. et al. Predicting pre-Columbian anthropogenic soils in Amazonia. **Proceedings of the Royal Society B: Biological Sciences**, 2014.

MCMICHAEL, C. H.; PALACE, M. W.; GOLIGHTLY, M. Bamboo-dominated forests and pre-Columbian earthwork formations in south-western Amazonia. **Journal of Biogeography**, 2014.

MUSCARELLA, R. et al. A well-resolved phylogeny of the trees of Puerto Rico based on DNA barcode sequence data. **PLoS ONE**, 2014.

PRANCE, G. T. "*Bertholletia*". P. 114–18 in *Lecythidaceae. Part II. The Zygomorphic-Flowered New World Genera*, organizado por S. A. Mori e G. T. Prance. Bronx, NY: **New York Botanical Garden**. 1990.

RAMIREZ-VILLEGAS, J., JARVIS, A., & LÄDERACH, P.. Empirical approaches for assessing impacts of climate change on agriculture: The EcoCrop model and a case study with grain sorghum. **Agricultural and Forest Meteorology**, 2013.

RIBEIRO, M. B., et al. "Anthropogenic landscape in southeastern Amazonia: contemporary impacts of low-intensity harvesting and dispersal of Brazil nuts by the Kayapo Indigenous people". **PLoS One** 9(7):e102187, 2014.

ROOSEVELT, A. C., M. L. COSTA, L. MACHADO, M. MICHAEL, e et al. "Paleoindian cave dwellers in the Amazon: the peopling of the Americas". **Science** 272(5260):373–84, 1996.

RUSSELLO, M.A., WATERHOUSE, M.D., ETTER, P.D., JOHNSON, E.A..From promise to practice: pairing non-invasive sampling with genomics in conservation. **PeerJ**, 2015.

SALES, LILIAN P., LUCIRENE R, RÔMULO M, e ARNDT H. "Climate change drives spatial mismatch and threatens the biotic interactions of the Brazil nut". **Global Ecology and Biogeography** 30(1):117–27, 2020.

SCOLES, RICARDO, GRIBEL.. "Population Structure of Brazil Nut (*Bertholletia excelsa*, *Lecythidaceae*) Stands in Two Areas with Different Occupation Histories in the Brazilian Amazon". **Human Ecology** 39(4):455–64, 2011.

- Shepard-Jr, G. H., e H. Ramirez.. “‘Made in Brazil’: Human dispersal of the Brazil nut (*Bertholletia excelsa*, Lecythidaceae) in ancient Amazonia”. **Economic Botany** 65(1):44–61, 2011.
- O’NEILL, B. C. et al. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. **Geoscientific Model Development**, 2016.
- OLSON, D. M. et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth. **BioScience**, 2001.
- PHILLIPS, S. J. et al. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. **Ecological Applications**, 2009.
- TER STEEGE, H. et al. Towards a dynamic list of Amazonian tree species. **Scientific Reports**, 2019.
- THUILLER, W., GUÉGUEN, M., RENAUD, J., KARGER, D. N., & ZIMMERMANN, N. E. (2019). Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, 10(1), 1–9.
- THOMAS, E., C. ALCÁZAR-CAICEDO, J. LOO, e R. KINDT. “The distribution of the Brazil nut (*Bertholletia excelsa*) through time: from range contraction in glacial refugia to anthropogenic climate change”. *Boletim do Museu Paraense Emilio Goeldi, Ciências Naturais* 9(2):267–91, 2014.
- THOMAS, E., C. ALCÁZAR-CAICEDO, J. LOO, e R. KINDT. R. CORVERA,. “Uncovering spatial patterns in the natural and human history of Brazil nut (*Bertholletia excelsa*) across the Amazon Basin”. *Journal of Biogeography* 42(8):1367–82, 2015.
- TOURNE, D. C. M. et al. Strategies to optimize modeling habitat suitability of *Bertholletia excelsa* in the Pan-Amazonia. **Ecology and Evolution**, 2019.
- TRIBETT, W. R. et al. Paris INDCs. In: **Springer Climate**. [s.l: s.n.].
- VALAVI, R. et al. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. **Methods in Ecology and Evolution**, 2019.

WADT, L. H. O., K. A. KAINER, C. L. STAUDHAMMER, e R. O. P. SERRANO. “Sustainable forest use in Brazilian extractive reserves: Natural regeneration of Brazil nut in exploited populations”. **Biological Conservation** 141(1):332–46, 2008

WENGER, S. J.; OLDEN, J. D. Assessing transferability of ecological models: An underappreciated aspect of statistical validation. **Methods in Ecology and Evolution**, 2012.

Supporting information

Table S1: Explanatory variables selected by Tourne et al., 2019.

Type of variables	Variable	Source
Climatic variables	BIO1: Annual Mean Temperature BIO4: Temperature Seasonality (standard deviation*100) BIO6: Minimum Temperature of Coldest Month	WorldClim (Hijmans et al., 2005)
	PET: Annual potential evapotranspiration	Calculated using the Hargreaves formula (Hargreaves & Allen, 2003) for monthly reference evapotranspiration values.
Soil variables	Bulk density (fine earth) Clay content (0-2 micrometer) mass fraction Silt content (2-50 micrometer) mass fraction Sand content (2-50 micrometer) mass fraction Sand content (50-2000 micrometer) mass fraction Cation exchange capacity Organic carbon content (fine earth fraction) pH in H ₂ O	ISRIC SoilGrids250m data (Hengl et al., 2017).

Terrain variables	Aspect Direction of waterflow Slope	Calculated in raster package for R (Hijmans & van Etten, 2020) using the GMTED2010 digital elevation model (Danielson & Gesch, 2011).
-------------------	---	---

Table S2: Explanatory variables considered in the suitability modelling with their corresponding source

Type of variables	Variable	Source
Climatic variables	BIO1: Annual Mean Temperature BIO2: Mean Diurnal Range (Mean of monthly (Maximum Temperature - Minimum Temperature) BIO3: Isothermality ((BIO2/BIO7)*100) BIO4: Temperature Seasonality (standard deviation*100) BIO5: Maximum Temperature of Warmest Month BIO6: Minimum Temperature of Coldest Month BIO7: Temperature Annual Range (BIO5-BIO6) BIO8: Mean Temperature of Wettest Quarter BIO9: Mean Temperature of Driest Quarter BIO10: Mean Temperature of Warmest Quarter BIO11: Mean Temperature of Coldest Quarter BIO12: Annual Precipitation BIO13: Precipitation of Wettest Month	WorldClim (Hijmans et al., 2005)

	BIO14: Precipitation of Driest Month BIO15: Precipitation Seasonality (Coefficient of Variation) BIO16: Precipitation of Wettest Quarter BIO17: Precipitation of Driest Quarter BIO18: Precipitation of Warmest Quarter BIO19: Precipitation of Coldest Quarter	
	PET: Annual potential evapotranspiration Aridity index	Calculated using the Hargreaves formula (Hargreaves & Allen, 2003) for monthly reference evapotranspiration values. Calculated as annual precipitation / annual reference potential evapotranspiration.
Soil variables	Bulk density (fine earth) Clay content (0-2 micrometer) mass fraction Silt content (2-50 micrometer) mass fraction Sand content (50-2000 micrometer) mass fraction Cation exchange capacity Organic carbon content (fine earth fraction) pH in H ₂ O	ISRIC SoilGrids250m data (Hengl et al., 2017).
Terrain variables	Aspect Direction of waterflow Slope TPI: Topographic position index TRI: Terrain roughness index	Calculated in raster package for R (Hijmans & van Etten, 2020) using the GMTED2010 digital elevation model (Danielson & Gesch, 2011).

Table S3: Algorithms used.

Algorithm	Description	R package
-----------	-------------	-----------

MAXENT	maximum entropy algorithm	dismo
GBM	generalised boosted regression models	gbm
RF	random forests	randomForest
GLMSTEP	generalised linear models	dismo
GLMNET	generalised linear models	glmnet
GAMSTEP	generalised additive models	gam
MGCV	generalised additive models	mgcv
EARTH	multivariate adaptive regression splines	earth
FDA	flexible discriminant analysis	mda
SVM	support vector machines	kernlab

Link S1: https://github.com/Gabrielforest/dataset_amazon_nut_tree Amazon nut trees dataset.

Link S2: https://bideco.shinyapps.io/vif_vars_env/ Contains the suitability map of the best model and the second-best model with the three types of map for each (the difference between LAM can be observed in the count map).

Link S3: https://bideco.shinyapps.io/leaflet_active/ resolutions tested with Tourne variables.

Link S4: https://gabrielforest.github.io/maps_grid/ All the models achieved in this study.

Link S5: https://bnutmodelling.shinyapps.io/future_env/ Projections of the best model and the second-best model.

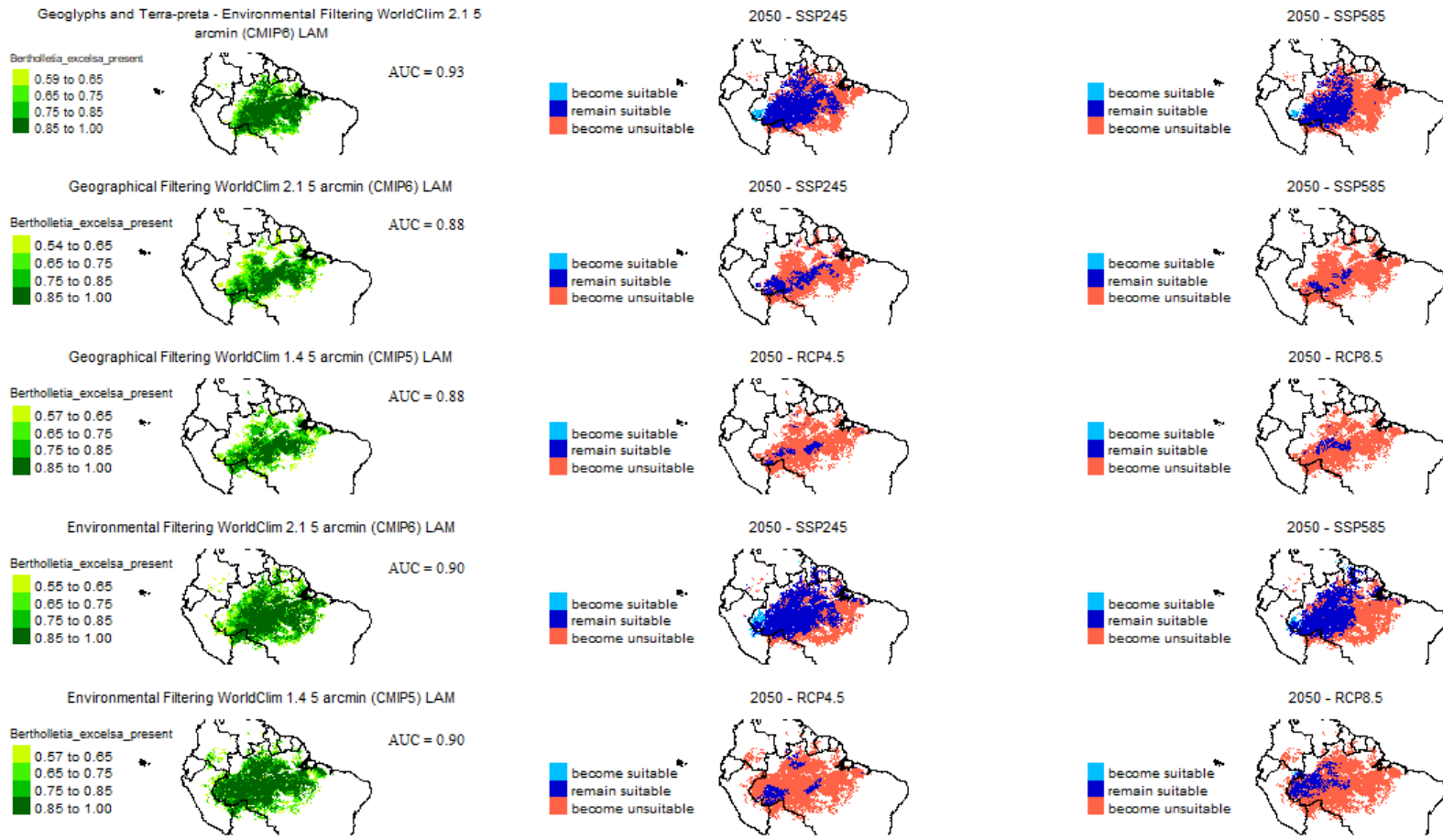


Fig S1. Model results with different filtering (geographical or geographical + environmental) and their respective projections (SSP2-4.5 and SSP5-8.5 when CMIP6; or RCP2-4.5 and RCP5-8.5 when CMIP5) scenarios for the average period of 2050.

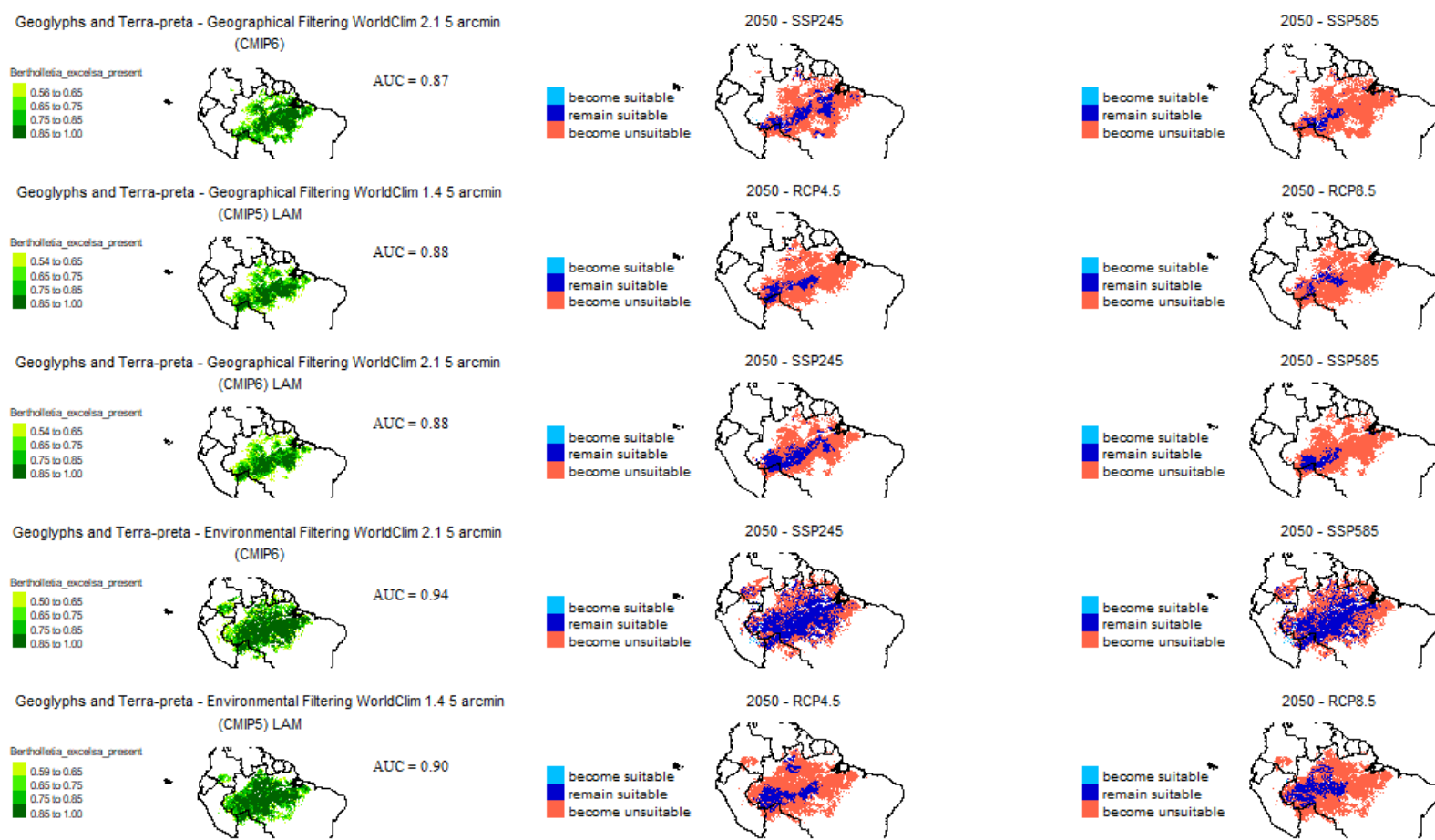


Fig S2. Model results with different filtering (geographical or geographical + environmental) and their respective projections (SSP2-4.5 and SSP5-8.5 when CMIP6; or RCP2-4.5 and RCP5-8.5 when CMIP5) for the average period of 2050.

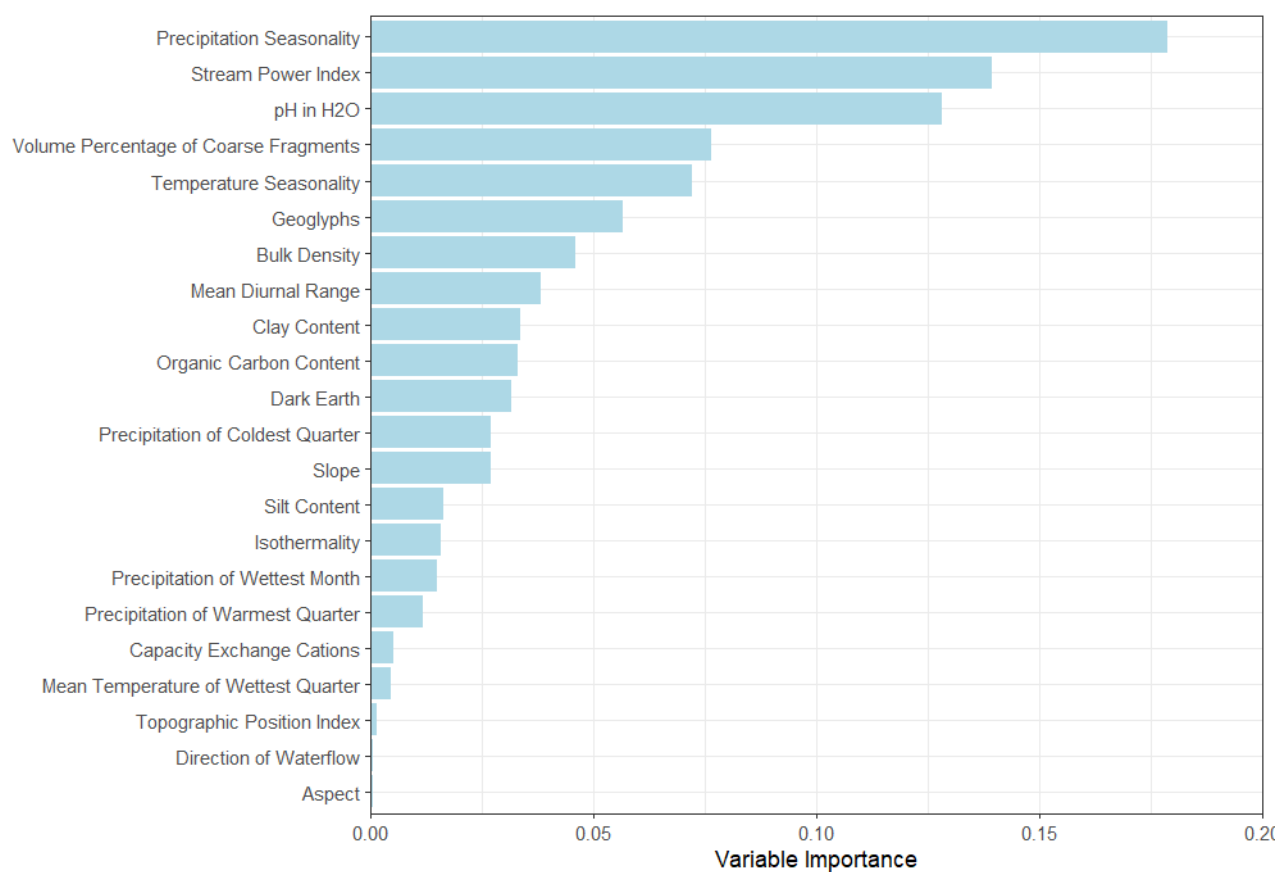


Fig S3. : Importance scores of the environmental variables for the best ensemble model.