PUCRS

ESCOLA DE CIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA
MESTRADO EM ZOOLOGIA

CHRISTIAN SCHÖBEL

NICHE MODELING OF THE ECONOMICAL IMPORTANT *Mahanarva* SPECIES IN SOUTH AND CENTRAL AMERICA (HEMIPTERA, CERCOPIDAE)

Porto Alegre

2018

PÓS-GRADUAÇÃO - STRICTO SENSU

Pontifícia Universidade Católica
do Rio Grande do Sul

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL

Escola de Ciências

PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

Niche Modeling of the economical important *Mahanarva* species in South and Central America (HEMIPTERA, CERCOPIDAE)

Christian Schöbel

DISSERTAÇÃO DE MESTRADO

2018

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL

Escola de Ciências

PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

Niche Modeling of the economical important *Mahanarva* species in South and Central America (HEMIPTERA, CERCOPIDAE)

Christian Schöbel

Orientador: Dr. Gervasio Silva Carvalho

DISSERTAÇÃO DE MESTRADO

Porto Alegre – RS - Brasil

2018

Summary

Figures

Tables

Acknowledgements

Resumo

Mahanarva fimbriolata, M. spectabilis, M. liturata and *M. posticata* (Hemiptera: Cercopidae) são conhecidas como pragas de plantações de cana-de-açúcar e pastagem em todo Brasil. Por alimentarem-se diretamente da seiva das plantas, esses cercopídeos causam fitotoxicidade e devido a isso diminuem a produção. A modelagem da distribuição de espécies permite analisar a possível occurencia das quatro espécies na América do Sul e Central. Para criar modelos de distribuição de espécies foram utilizados em R, os algoritmos Bioclim, Domain, diferentes modelos lineares generalizados e Maxent. Nesses modelos foram utilizadas variáveis bioclimáticas atuais e futuras, além da elevação e outras variáveis agrícolas. As variáveis climáticas futuras são para os anos 2050 e 2070 com diferentes *repentant concentration pathways*. As espécies apresentam habitats adequados em diferentes países da América do Sul e Central, onde as plantações de cana-de-açúcar são abundantes. Os resultados das análises climáticas futuras não apresentaram diferenças em relação às análises climáticas atuais. No geral, o algoritmo Maxent mostrou os maiores valores de AUC e o Bioclim os menores. As variáveis que mais contribuíram para os modelos são: elevação, *isothermality* e diferentes variáveis de precipitação. As mudanças climáticas e ciclos de vida de insetos adicionais não têm impacto em habitats adequados dos insetos. Em geral, o Maxent é o melhor algoritmo para realizar modelos de distribuição de espécies com um número baixo de pontos de ocorrência e análises de mudanças climáticas.

Palavras chave: Spittlebug; Cana de açúcar; pastagem; Dano agrícola

Abstract

Mahanarva fimbriolata, M. spectabilis, M. liturata and *M. posticata* (Hemiptera: Cercopidae) are known pests for sugarcane and pasture plantations throughout Brazil. By direct sap feeding on the plants they cause phytotoxicity and due to this they decrease the production of plantations. With species distribution modeling it is possible to analyze the possible occurence of the four species in South and Central America. To create species distribution models the algorithms *Bioclim, Domain,* different generalized linear models and *Maxent* were used in R. For those models current and future bioclimatic variables as well as elevation and other agricultural variables were used. The future climatic variables are for the years 2050 and 2070 with different repentant concentration pathways. The species show suitable habitats in different countries in South and Central America where sugarcane plantations are abundant. The results of the future climate analyzes do not show differences compared to the current climate analyzes. Overall the *Maxent* algorithm showed the highest AUC scores and *Bioclim* the lowest. The variables which contributed the most to the models are elevation, isothermality and different precipitation variables. Climate change and therefore additional insect lifecycles do not have an impact on the insect's suitable habitats. Overall *Maxent* is the best algorithm to perform species distribution models with a low number of occurrence points and for climate change analyzes.

Keywords: Spittlebug; Sugarcane; Pasture; Agricultural Damage

Presentation

Mahanarva fimbriolata, *M. spectabilis*, *M. Liturata* and *M. Posticata* (Hemiptera: Cercopidae) are known pests on sugarcane and pasture plantations. Therefore those four species are causing high economic damage every year on the plantations throughout Brazil. In this study species distribution models were created to analyze where possible habitats of the four species in South and Central America are. Furthermore, possible habitats regarding the climate change are examined. With this project it was possible to show that there are different suitable habitats in South and Central America where the species still do not occur. The impact of the climate change on suitable habitats of the four species is minimal. The article will be send to the Journal "Diversity and Distributions". Consequently the article is following the rules of the stated journal.

Niche Modeling of the economical important *Mahanarva* species in South and Central America (HEMIPTERA, CERCOPIDAE)

Schöbel, C. & Carvalho, G. S.

Laboratório de Entomologia, Faculdade de Biociências, Programa de Pos-Graduacão em Zoologia, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga 6681, 90619-900 Porto Alegre, RS, Brazil, cschoebel@hotmail.com, gervasio@pucrs.br

Abstract

Aim

Mahanarva fimbriolata, M. spectabilis, M. liturata and *M. posticata* (Hemiptera: Cercopidae) are known pests for sugarcane and pasture plantations and by direct sap feeding on the plants they cause phytotoxicity. The aim is to analyze the possible occurence of the four species in South and Central America.

Location

The four species are known in Brazil but whole South and Central America are analyzed.

Methods

To create species distribution models the algorithms *Bioclim, Domain,* different generalized linear models and *Maxent* were used in R. For those models current and future bioclimatic variables as well as elevation and other agricultural variables were used. The future climatic variables are for the years 2050 and 2070 with different repentant concentration pathways.

Results

The species show suitable habitats in different countries in South and Central America where sugarcane plantations are abundant. The results of the future climate analyzes do not show differences compared to the current climate analyzes. Overall the *Maxent* algorithm showed the highest AUC scores and *Bioclim* the lowest. The variables which contributed the most to the models are elevation, isothermality and different precipitation variables.

Main conclusion

Climate change and therefore additional insect lifecycles do not have an impact on suitable habitats. Overall *Maxent* is the best algorithm to perform species distribution models with a low number of occurrence points and for climate change analyzes.

Introduction

Mahanarva Distant, 1909 is one of 173 genera of the Cercopidae Leach, 1815 (Soulier-Perkins, 2016) and this family in turn, is the biggest of the Cercopoidea Leach, 1815 (Hemiptera – Auchenorrhyncha). Totally there are 44 species of *Mahanarva.* Their common name (spittlebug) is referred to the foamy excretion which the nymphs excrete for their protection (Carvalho & Webb, 2005). Generally Cercopidae is one of the most diverse families of the Cercopoidea. The species are characterized by a variability of bright colors.

Studies of *M. fimbriolata* showed that the complete lifecycle of the species takes about 60 days (egg to adult). In about 37 of the 60 days the insect is in the nymph stage followed by an adult longevity of about 18 days in case of males and about 23 days in females (Garcia *et al.,* 2006). Overall, factors like temperature, the availability of food or the day length have an influence on the development time of nymphs and also on the duration of the adult longevity (Simões *et al*. 2013).

Spittlebugs are damaging plants by direct sap feeding. Therefore the foliage turns yellow due to a lack of sap and in general the development of the host plant is impaired (Carvalho, 1978). In 2016 38.9 million tons of sugar were produced. With that Brazil is the world's largest producer of sugar (International Sugar Organization, 2017). The sugarcane planted area in Brazil, in 2006, covered 8.6 million hectares. Furthermore, in 2014 there were already 10.4 million hectares planted (IBGE, 2016). This is an increase of about 21% (1.8 million hectares) in eight years. This shows that the sugarcane production is an important and increasing part of the Brazilian agriculture in recent years. Some *Mahanarva* species are known agricultural pests. Especially the species *Mahanarva fimbriolata* (Stål, 1854*), M. posticata* (Stål, 1855), *M. spectabilis* (Distant, 1909) and *M. liturata* (Le Peletier & Serville 1825) are well known to damage sugarcane and pasture plantations (Alves & Carvalho, 2014; Resende e*t al.*, 2013; Almeida *et al.*, 2003). *Mahanarva fimbriolata* has the largest distribution of the four species and is present in most of the Brazilian states. The other species have a more distinct distribution to the coastal and the central states. *M. liturata* and *M. posticata* occur only in the states at the Atlantic coast and *M. spectabilis* occurs mostly in the central Brazilian states.

Following experiments of Almeida *et al.* in 2003 and 2006 significant savings in sugarcane plantations due to biological pest control of spittlebugs are possible. In the experiments, the fungus *Metarhyzium anisopliae* was applied to about 25 hectares of sugarcane plantation to reduce the sugarcane damage based on *Mahanarva fimbriolata*. This biological pest control is much cheaper than common chemical pest treatments. In these experiments it was possible to save about R$ 7 million. Furthermore the estimated profit of the biological control of the spittlebug is approximately R$ 212000 per hectare. These studies show that biological pest control is a cheap alternative to common chemical pest control.

Even more important is the possible profit that can be achieved due to the control of *Mahanarva fimbriolata*. An extrapolation of the R$ 212000 savings per hectare to the 2014 in sugarcane covered area of Brazil is very impressive. In theory a profit of about R$ 220.48 million each year, only because of a higher sugarcane production due to the biological pest control of *Mahanarva fimbriolata* is possible. This calculation has to be considered with caution due to a serious lack of information about the pests of sugarcane and pasture plantations.

With distribution modeling it is possible to analyze if these species can spread and which regions can potentially be affected. This analyzes can be done by the study of the three main types of influences on species. Those are limiting factors like temperature barriers, disturbances (for example human influences) on the species and resources which are all compounds that can be assimilated by organisms (e.g. energy, water). In Geographical Information Systems (GIS) this set of data, including the known distribution of the species can be visualized in different layers (Guisan & Thuiller, 2005). By projecting this model onto an actual map, geographical regions inside and outside of the species niche are visible. Now evaluations of possible invasion can be made by the detection of accessible regions with similar conditions in comparison to the ecological niche of the species (Peterson, 2003).

The distribution of the four mentioned spittlebugs is not well known and there is a lack of distribution records. With the species distribution modelling it is possible to discover suitable habitats in South and Central America to delimit areas with possible or already existing spittlebug infestation. Besides that the influence of the climate change on those habitats will be studied. This is the first ever attempt to use species distribution modelling on *Mahanarva* spittlebugs.

Despite knowing that *Maxent* will probably outperform the other algorithms, the comparison with other algorithms within this study is necessary (Elith, *et. al.*, 2006; Hijmans & Elith, 2017). Due to the low number of presence records and the aim to analyze the species distribution with the aspect of climate change it is not clear which algorithm would perform best and because of this also other algorithms will be carried out to achieve the best possible results. *Bioclim* is known to not perform very well in comparison to other algorithms but it is still used in cases when the climate change is analysed (Elith, *et. al.*, 2006; Hijmans & Graham, 2006). The *Domain* algorithm is frequently used and therefore should be useful to compare the results to other studies (Hijmans & Elith, 2017).

Material & Methods

The occurrence data was collected from the websites Species link and the Global Biodiversity Information Facility (GBIF, 2016; Species Link, 2017). Furthermore data was collected throughout Brazil by solicitation of Antonio Soares (Consultor de Desenvolvimento de Mercado – Bayer S.A.). Data without coordinates and only a city or region as occurrence were georeferenced using Global Gazetteer Version 2.3 (Falling Rain Software, Ltd 2017). For *M. fimbriolata* 19, for *M. spectabilis* 29, for *M. liturata* 17 and for *M. posticata* only 9 different occurrence points are available to analyze (Figure 1; Attachments: Distribution Coordinates).
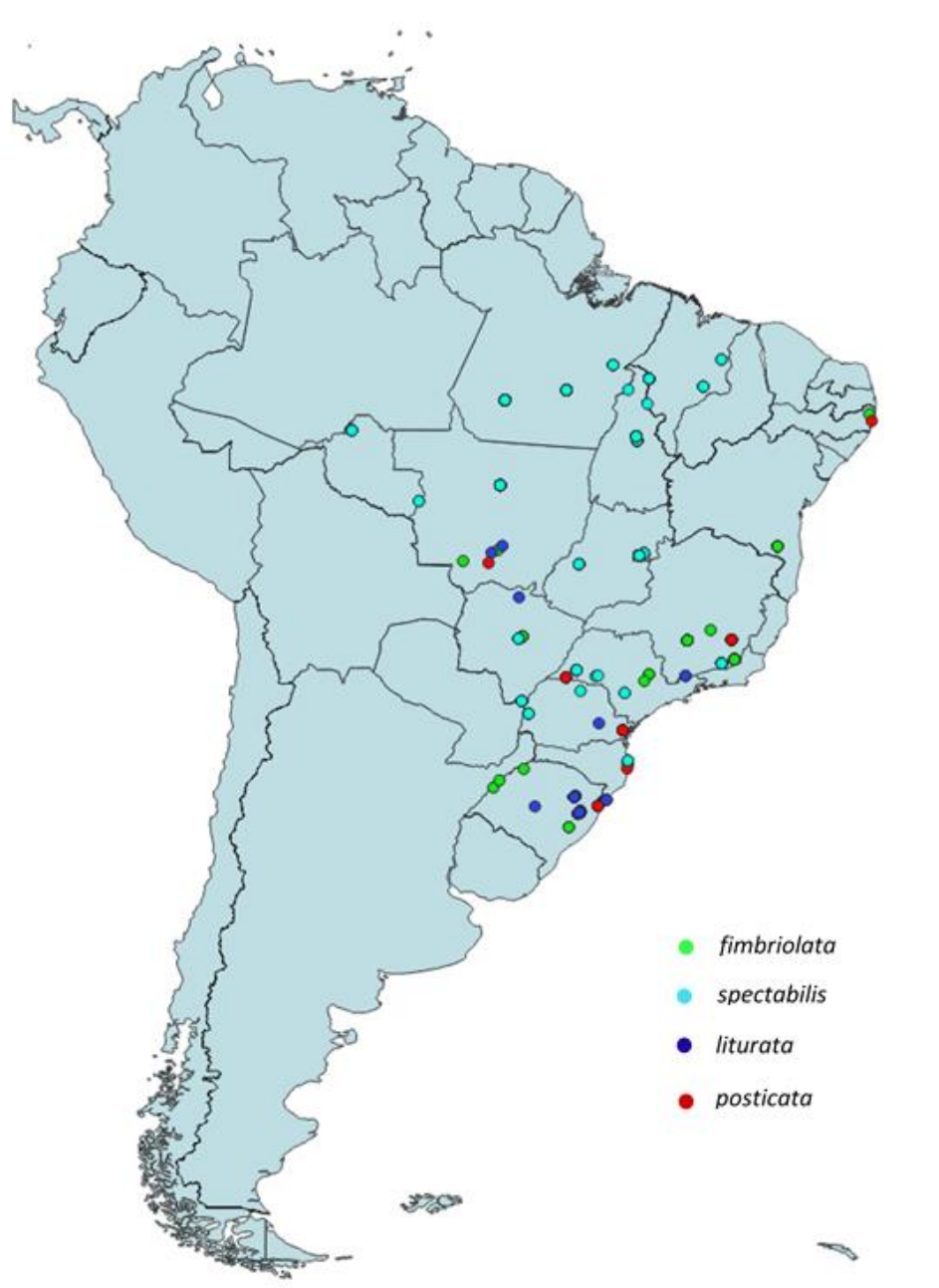
Figure 1: Distribution records of *M. fimbriolata* (green), *M. spectabilis* (turquois), *M. liturata* (blue) and *M. posticata* (red) (Attachments:Table 3).

With R version 3.3.2 and RStudio version 1.0.136 including the packages *dismo*, *sdm*, *sp*, *raster, rgdal* and *maptools* it was possible to create species distribution models of the species following the guide *Species distribution modeling with R* (Bivand, *et. al.*, 2017; Bivand, *et. al.*, 2017; Hijmans, *et. al.*, 2011; Hijmans & Van Etten, 2012; Naimi & Araújo, 2016; Pebesma & Bivand, 2005; R Development Core Team, 2008; RStudio Team, 2016). Climatic data of Worldclim was used for the algorithms. For the current climate data the Version 2 Bioclimatic variables with a spatial resolution of 2.5 minutes were selected (Fick & Hijmans, 2017; Attachments: Variable Layers). For future data WorldClim 1.4 downscaled (CMIP5) data - as well with a spatial resolution of 2.5 minutes - was chosen and the bioclimatic variables of BCC-CSM1-1 for the representative concentration pathways 26, 45, 60 and 85 for the years 2050 and 2070 were downloaded (Hijmans, *et. al.*, 2005). Furthermore elevation data, global cover and agricultural land use data was used. Those maps present the global areas used as pasture and sugarcane plantations and as general croplands (Ramankutty, *et. al.*, 2010a; Ramankutty, *et. al.*, 2010b; Ramankutty, *et. al.*, 2008; You, *et. al.*, 2014; Attachments: Variable Layers). The GlobCover 2009 version 2.3 map was used and the selected elevation map is the Natural Earth I with Shaded Relief and has a scale of 1:10 million (Natural Earth, 2017; Arino, *et. al.* 2012; Attachments: Variable Layers). All maps were cut in the shape of South and Central America using Quantum GIS (QGIS Development Team 2009).

Within R the collinearity – the correlation between two or more predictor variables - was analyzed. For this the *Variance inflation factor* with Pearson correlation coefficients was carried out. This method calculates the correlation coefficient between two variables and detects the pair with the highest coefficient. Afterwards the one variable with the highest correlation coefficient will be excluded and the method will be repeated until there are no remaining pairs with a high correlation coefficient (Naimi & Araújo, 2016). The models were generated using the algorithms *Domain, Bioclim, Maxent* and *Generalized Linear Models* executing *linear*, *logistic* and *Poisson* regressions (Phillips, *et. al.*, 2006; Busby, 1991). With those presence only algorithms only the actual distribution data is used and no Absence Points are going to be created or analyzed.

To analyze the performance of the models the dataset will be divided into a "training" and a "testing" group. Bootstrap models will be generated for every species to predict the accuracy of the model as a result of the training data. Probably 75% of the dataset will be used to train the model (training) and 25% to estimate the performance (Franklin, 2009). Furthermore, the area under the curve (AUC) of all algorithms was calculated for each species. The AUC is an indicator for the quality of the model (Hijmans & Elith 2017). The AUC has a value between 0 and 1 by which a value of 1 indicates a perfect

model and 0.5 represents bad model, not better than a random guess (Elith, *et al.*, 2006, Peterson, *et. al*, 2008). The results of the models for the four species are presented in maps.

The maps using future climate data will be compared to the maps of the current climate using Quantum GIS v. 2.18 (QGIS Development Team 2009). With the raster calculator tool the future maps will be subtracted from the current climate maps. This shows positive values when the futures habitat suitability is higher compared to the current one. Vice versa negative values are the result. Afterwards they will be colorized using the singleband pseudocolor presentation with a coloration from red (negative values) to green (positive values) with unchanged values in white. In total fifteen same interval classes will be used to cover values from -60 to 60. The number of fifteen interval classes was chosen to still show distinctive classes. A higher number results in higher coloration which makes it even more difficult to distinguish different classes. Even if there are higher or lower values in the maps the same intervals of -60 to 60 are used to guarantee a bright coloring because very high or low values tend to be infrequent and not abundant. Afterwards the Raster layer statistics Tool of the QGIS Toolbox will be used to analyze the mean pixel value of each map.

Results

Area Under the Curve

A total of 216 analyzes were made. All AUC values are between 0.42 and 1 with a mean value of 0.88. Only one analysis shows a value beneath 0.7 (0.42) and therefore a bad performance. Sixty analyzes have an AUC of over 0.95 which shows a high performance and the remaining 155 analyzes with an AUC over 0.7 have a moderate performance. Throughout all analyzes the AUC of *M. fimbriolata* has an arithmetic mean of 0.81, *M. spectabilis* of 0.86, *M. liturata* of 0.96 and *M. posticata* of 0.92. *M. liturata* shows the best and *M. fimbriolata* the worst results in comparison between the four species. Only the *Binomial* and the *Poisson* GLMs reached values of 1, exclusively in the analyzes of *M. liturata* and *M. posticata* (AUC Attachments Table 2).

Overall the *Maxent* algorithm performed best. Within the 36 analyzes 21 showed the highest AUC results at the *Maxent* algorithm. In case of the rcp26 and rcp45 future climate data of the year 2070 and the rcp60 future climate data of 2050 two different algorithms have the same AUC value which was the highest. In those cases both algorithms are counted as the best performing algorithm. This only occurred in the analyzes of *M. liturata* and *M. posticata*. In the analyzes of 2070 *Maxent* and the *Gaussian* generalized linear model have the same values and in the one analyzes of the 2050 climate data the *Binomial* and the *Poisson* generalized linear model (GLM) share the same result. The *Domain* and the *Binomial* GLM have the highest AUC in each four analyzes, the *Gaussian* GLM in nine and the *Poisson* GLM in three analyzes. The *Bioclim* algorithm has the worst results without reaching the highest AUC in any analyzes. Furthermore in 20 of the 36 analyzes the algorithm scored the lowest AUC in comparison to the other algorithms. In all analyzes of *M. spectabilis Maxent* was the best performing algorithm. Within the results of the other species this is not so obvious. In as well six analyzes *Maxent* best performed for *M. liturata*. The other analyzes show the best results in one of the generalized linear models. *M. fimbriolata,* especially with the climate data of 2050, shows the best performance with the *Domain* algorithm in four cases, in each two cases with the *Maxent* and the *Poisson* GLM and once with the *Binomial* GLM algorithm. Finally *M. posticata* shows the best performance in five analyzes with the *Gaussian* GLM and in three cases with *Maxent*. One analysis shows – as mentioned before – both algorithms with the same, highest AUC value (AUC Attachments: Table 2).

Distribution Maps

The different algorithms create maps with certain characteristic and repetitive schemes. The created maps do not show the species occurrence. They rather show more or less suitable habitats created with the used variables. There are color patterns with scales between 0 and 1, whereby 0 refers to a very low suitable habitat and 1 refers to a very high suitability. There are different scales within the created maps by different algorithms. All *Maxent* maps show a scale between 0 and 1 whereby the scales of *Domain* and *Bioclim* start with 0 but only reach 0.6 and 0.5. The three with GLM algorithms created maps have a lot of different scales. Many scales possess negative values and are far beyond the range of 0 to 1 (Figure 2, A-D). All maps created with the different GLM's *Binomial, Gaussian* and *Poisson* show throughout all species and different climatic data nearly the same pattern of suitable habitats. Only the presented color changes but all three different colors refer to the same poor habitat suitability of about 0%. The colors vary due to the different scales used. One small patch in nearly every map presents a higher suitability approximately at the location of the state São Paulo, Brazil (Figure 2, C). In comparison to this the *Bioclim* maps tend to create only small areas of higher habitat suitability and the rest of the map presents very low habitat suitability (Figure 2, A). The patches created by the *Domain* algorithm are larger but still smaller than those of *Maxent* (Figure 2, B D).
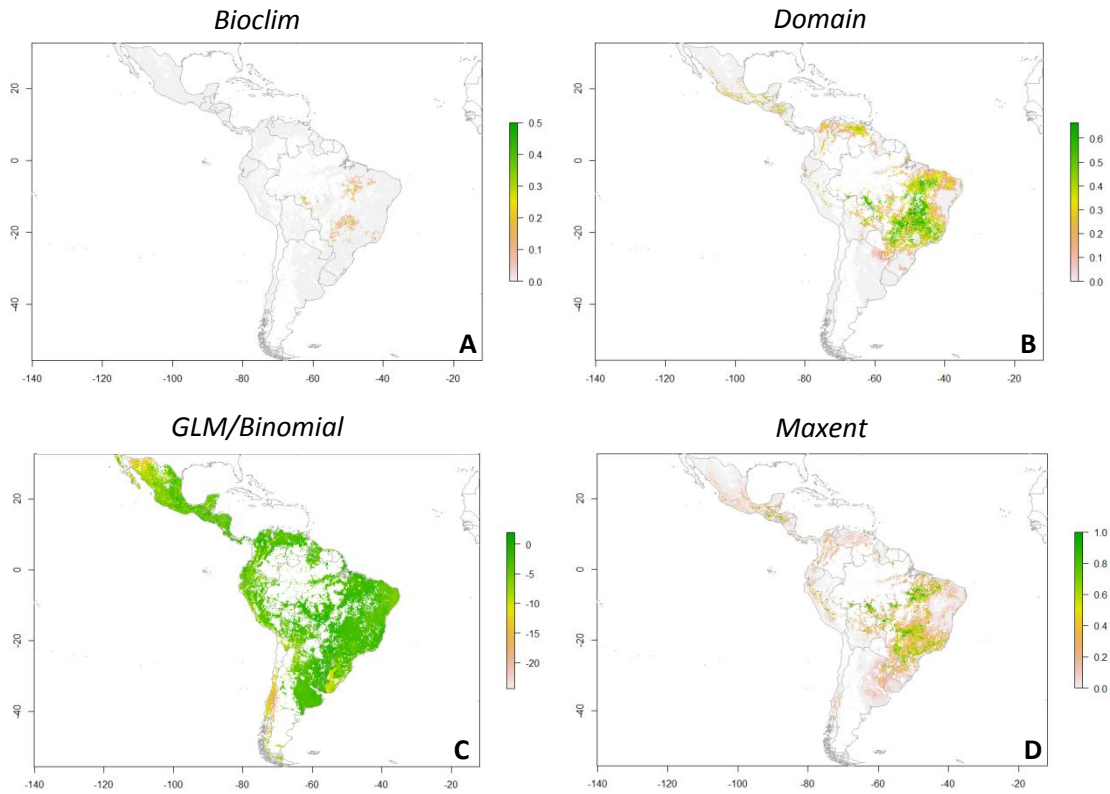
Figure 2: Species Distribution Models of *M. fimbriolata* using the algorithms *Bioclim* (**A**), *Domain* (**B**), *Generalized Linear model/Binomial* (**C**) and *Maxent* (**D**). The scales show the habitat suitability, referring to a higher suitability with a higher score.

Regarding the high AUC scores particularly the results of *Maxent* will be analyzed as this algorithm performed the best in comparison to the others. Only in case of some future climate analyzes of *M. fimbriolata* the *Domain* results will be shown due to the same reason. The *M. fimbriolata* results show patches with high suitable habitats at the Atlantic coast of Brazil until north to Paraiba, and within the states, Minas Gerais, São Paulo, Rio de Janeiro, Paraná, Santa Catarina and Rio Grande do Sul. There is also high habitat suitability in Uruguay and in the northeast part of Argentina as well as in middle Chile. Furthermore there are small patches in the center of Bolivia, Ecuador, Colombia and in the north of Venezuela. In Central America Nicaragua, Honduras, El Salvador and Guatemala small patches are present. In addition, at the Mexican Coast of the Gulf of Mexico there are as well some small patches located (Figure 3). The habitat suitability of *M. liturata* is quite similar to those shown by *M. fimbriolata*. The patch at the Atlantic coast in South Brazil to Argentina is more pronounced but only extends up to São Paulo. Further north only in the state of Bahia the habitat suitability is higher. The other states in the northeast of Brazil only show very low habitat suitability. In Ecuador and Peru the patch is bigger in comparison to *M. fimbriolata*. Central America has fewer suitable habitats for *M.*

liturata (Figure 6). The pattern of *M. spectabilis* is much smaller and different. The north tip of Argentina and central Chile bear low suitable habitats but further south there is no habitat suitability at all. Within Brazil the pattern is not distributed in the South and at the Atlantic Coast. Moreover the states Maranhão, Piauí, Tocantis, Mato Grosso, Goiás, Minas Gerais and São Paulo have a higher habitat suitability for this species. In the center of Bolivia, Peru and Colombia there is each a small stripe with medium to high habitat suitability. In Central America higher habitat suitability concentrates on Nicaragua, Honduras, El Salvador and Guatemala (Figure 4). *M. posticata* shows higher habitat suitability in the west and south of Mexico. Besides that there are patches with higher suitable habitats in Guatemala and Nicaragua. In South America Ecuador, the south of Chile, Paraguay, Uruguay, the north of Argentina and small patches in the east of Peru and in the center of Bolivia mark higher suitable habitats. The Atlantic coast in Brazil from the south to Rio de Janeiro and especially the states Minas Gerais, Goiás and Mato Grosso do Sul are showing higher habitat suitability (Figure 5).
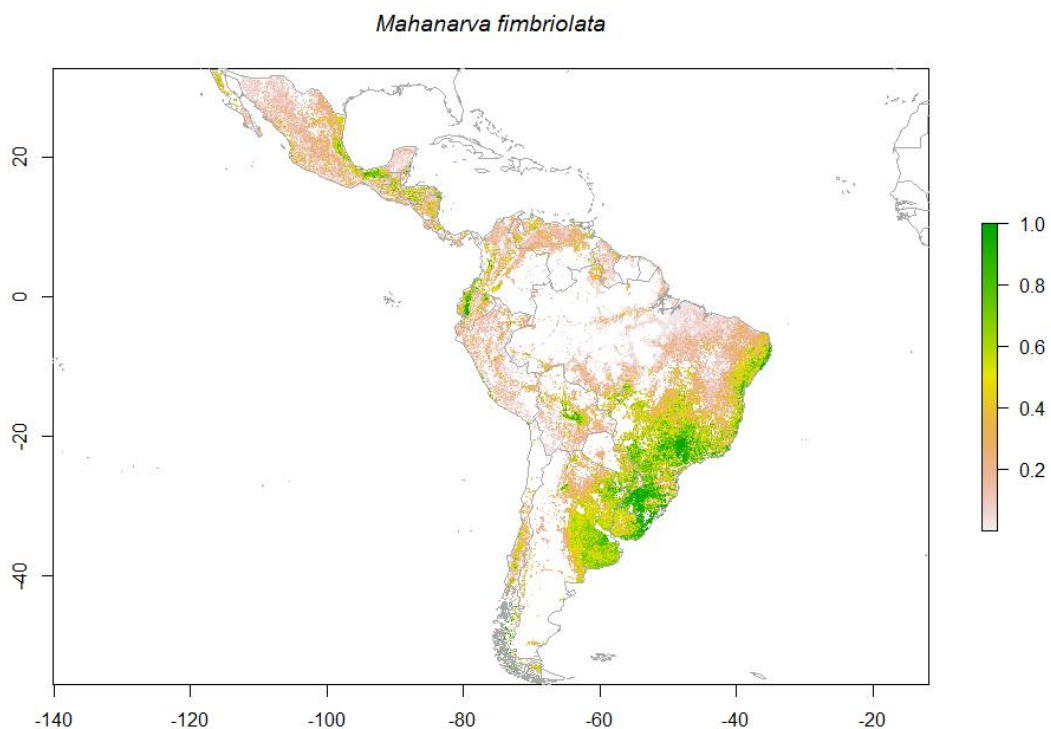
Figure 3: Maxent species distribution model of *M. fimbriolata* using current climate data. The scale shows the habitat suitability, referring to a higher suitability with a higher score.
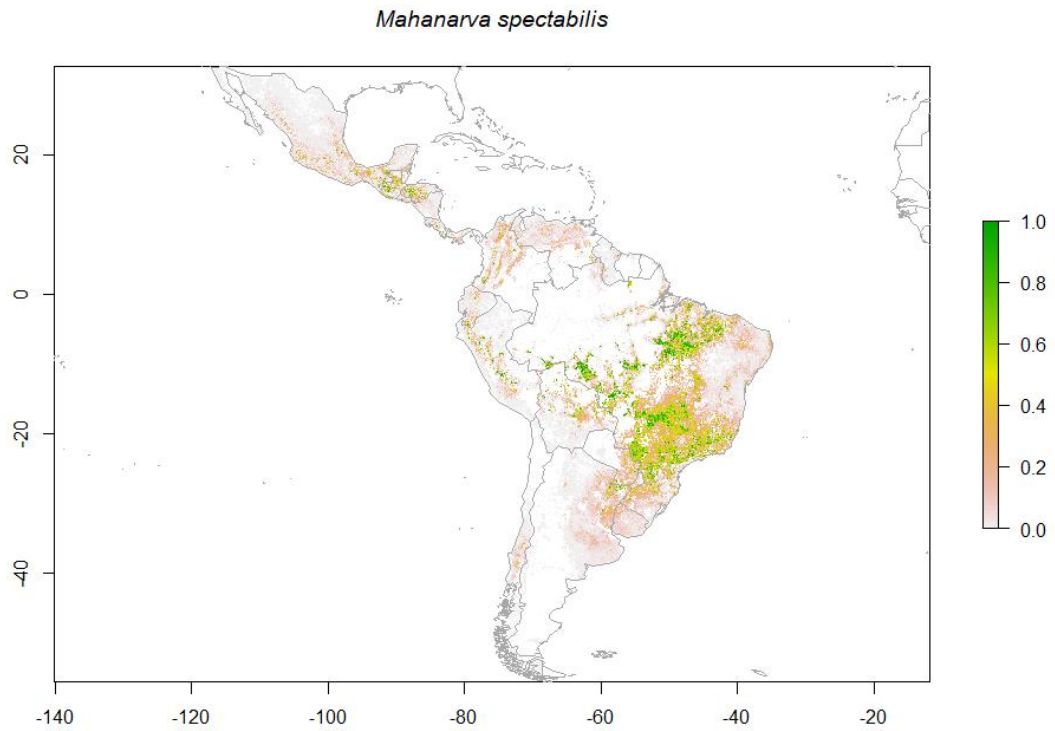
Figure 4: *Maxent* species distribution model of *M. spectabilis* using current climate data. The scale shows the habitat suitability, referring to a higher suitability with a higher score.
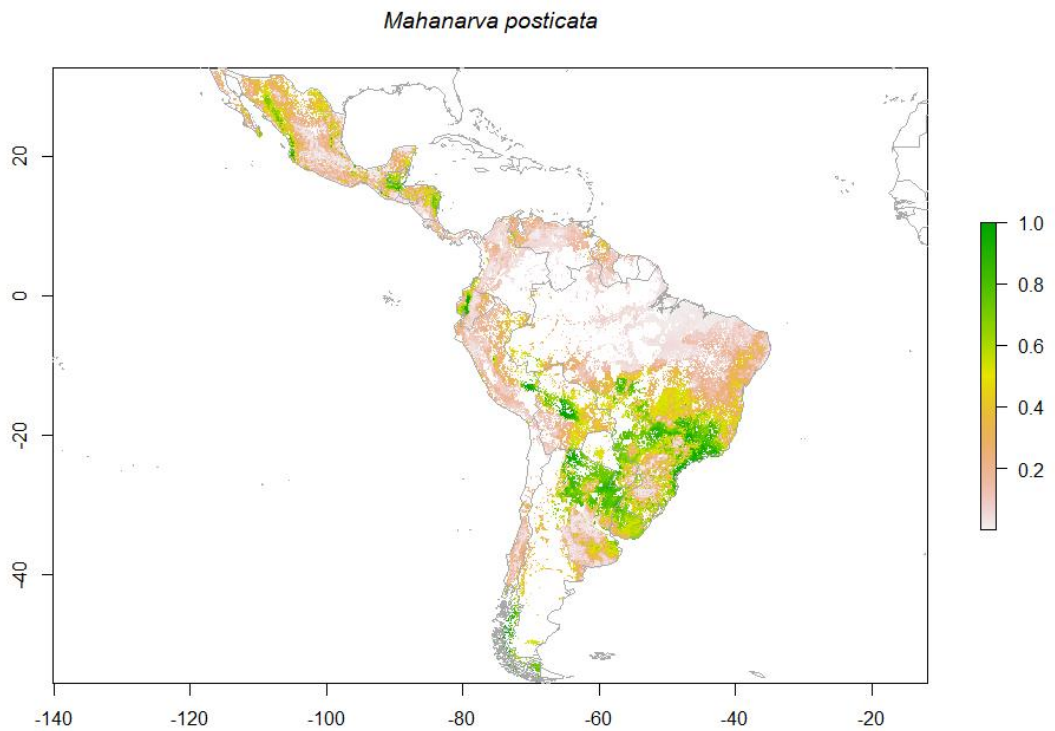
Figure 5: *Maxent* species distribution model of *M. posticata* using current climate data. The scale shows the habitat suitability, referring to a higher suitability with a higher score.
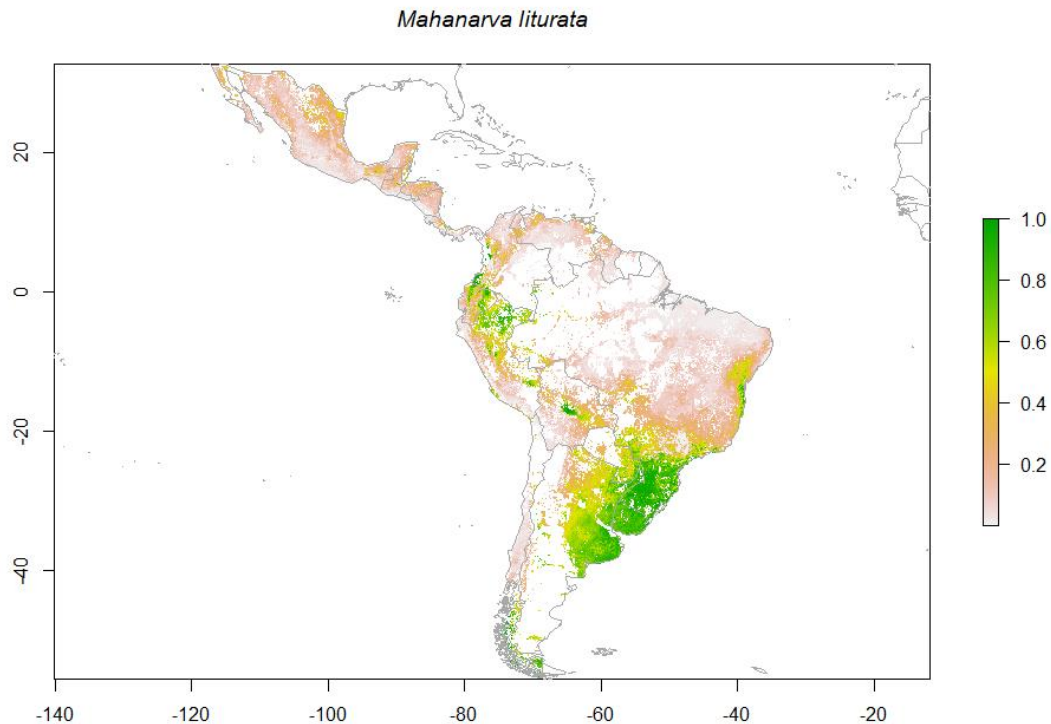
Figure 6: *Maxent* species distribution model of *M. liturata* using current climate data. The scale shows the habitat suitability, referring to a higher suitability with a higher score.

Future Climate Data

In comparison to the current climate results the future climate analyzes of *M. fimbriolata* are quite similar. Overall all different models for the years 2050 and 2070 only change the intensity of the colors shown on the map and thereby the habitat suitability a little bit. For example the rcp26 data of 2050 shows minor habitat suitability in the northeast of Argentina. But the rcp45 data of 2050 already shows higher suitable habitats for this area. In total, the different maps vary little but the patches are basically the same. The colored areas within the maps of the *Domain* algorithm are in general smaller compared to *Maxent.* The comparison between the different maps shows that the colored area with higher habitat suitability is located very similar. Only habitats in Mexico, Peru, Ecuador and Venezuela seem less suitable using *Domain.* Furthermore the results of *Maxent* show bigger areas of low habitat suitability around patches with higher suitability. In *Domain* those low habitat suitability areas do mostly not exist. As in the results of *M. fimbriolata*, maps of *M. spectabilis* generated with future data do not differ very much from the results of the model with the current climate. Only minor changes of the color intensity and the suitability patterns are visible. The results of *M. liturata* and *M. posticata* show the same pattern. Only minimal differences are determinable. Across all analyzes it stands out that the rcp45 data year 2050 analysis shows minimal lower habitat suitability in the northeast of Argentina and south Brazil.

Future Climate Data Comparison

By comparing the species distribution maps using future climate scenarios to the maps using current climate data the differences caused by different future climate scenarious will be visualized. In the resulting maps, green colored areas imply higher future habitat suitability compared to current climate habitat suitability for a given location. And the opposite, red colored areas denote lower future habitat suitability. All of the folowing explanations of changes of habitat suitability are refering to the comparison of future climate species distribution maps to current climate distribution maps.

The resulting maps show different areas of increasing or decreasing habitat suitabilty. But a few changes seem to be present in all future scenarios of one species. The maps of *M. fimbriolata* show decrasing habitat suitability in the northwest of South America. Especially a strip north the amazon in Venezuela, Colombia and Ecuador. Furthermore the the habitat suitability seems to increase in central america especially in Mexico. Those changes are more visible in the representative concentration pathway (rcp) scenarios 26 and 45 than in the other ones. The rcp 85 scenario even does show a decreasing habitat suitability. Those changes are comprehensible in the comparisons of the years 2050 and 2070. The results of *M. liturata* are more uniform. All scenarios show increasing habitat suitability in Chile, northeast Brazil, South Colombia and Venezuela and in Central America Honduras, Nicaragua, Costa Rica and Panama. On the other hand suitable habitats in Uruguay, North Argentina, Peru, the central and in some parts of the southern Atlantic coast of Brazil will decrease throughout all future climate scenarios. *Mahanarva posticata* shows very irregular changes of suitable habitats, comparing future climate scenarios with the current climate. Only a decrease of habitat suitability in the Brazilian states Mato Grosso, Goiás, Minas Gerias e São Paulo is visible in all scenarios. In contrast the suitable habitats of *M. spectabilis* does not change very much between the analyses. Comparing the 2070 scenarios to the 2050 scenarios there are bigger areas of less suitable habitats in the year 2070 maps. In general all maps show more suitable habitats in Colombia and Venezuela as well as in Chile. In South Brazil the States Paraná and Santa Catarina also demonstrate more suitable habitats. On the other hand habitats at the Pacific Coast of Peru decrease their suitability.

Figure 7: *Mahanarva fimbriolata* species distribution map comparison between future climate scenarios and current climate. For the years 2050 and 2070 the representative concentration pathway (rcp) scenarios 26 and 85 are shown. Green areas illustrate positive changes of suitable habitats in comparison between the future climate scenario and current climate. Vice versa red areas indicate negative changes.
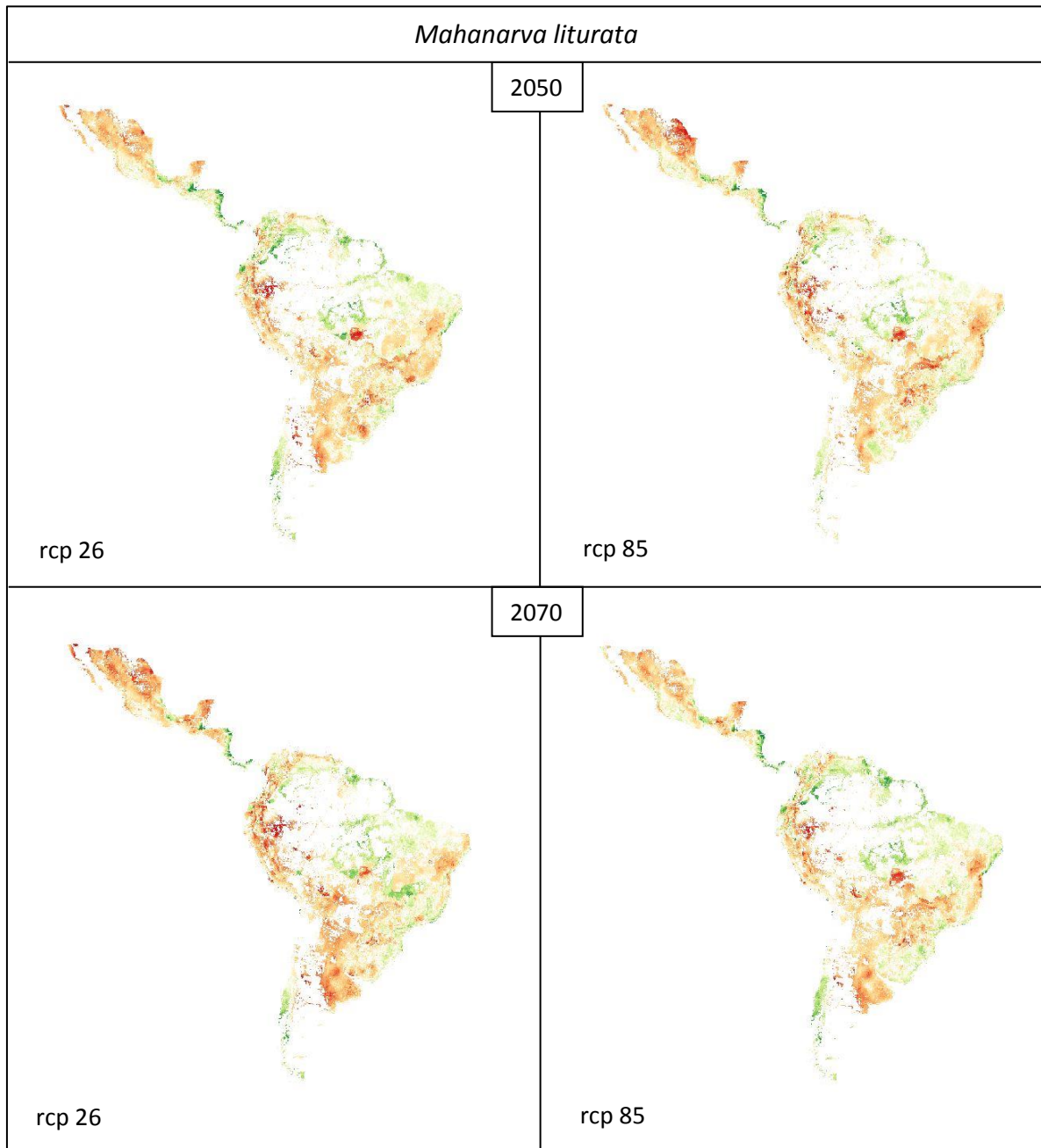
Figure 8: *Mahanarva liturata* species distribution map comparison between future climate scenarios and current climate. For the years 2050 and 2070 the representative concentration pathway (rcp) scenarios 26 and 85 are shown. Green areas illustrate positive changes of suitable habitats in comparison between the future climate scenario and current climate. Vice versa red areas indicate negative changes.

Mahanarva posticata	
2050	

rcp 26

rcp 85

2070

rcp 26

rcp 85
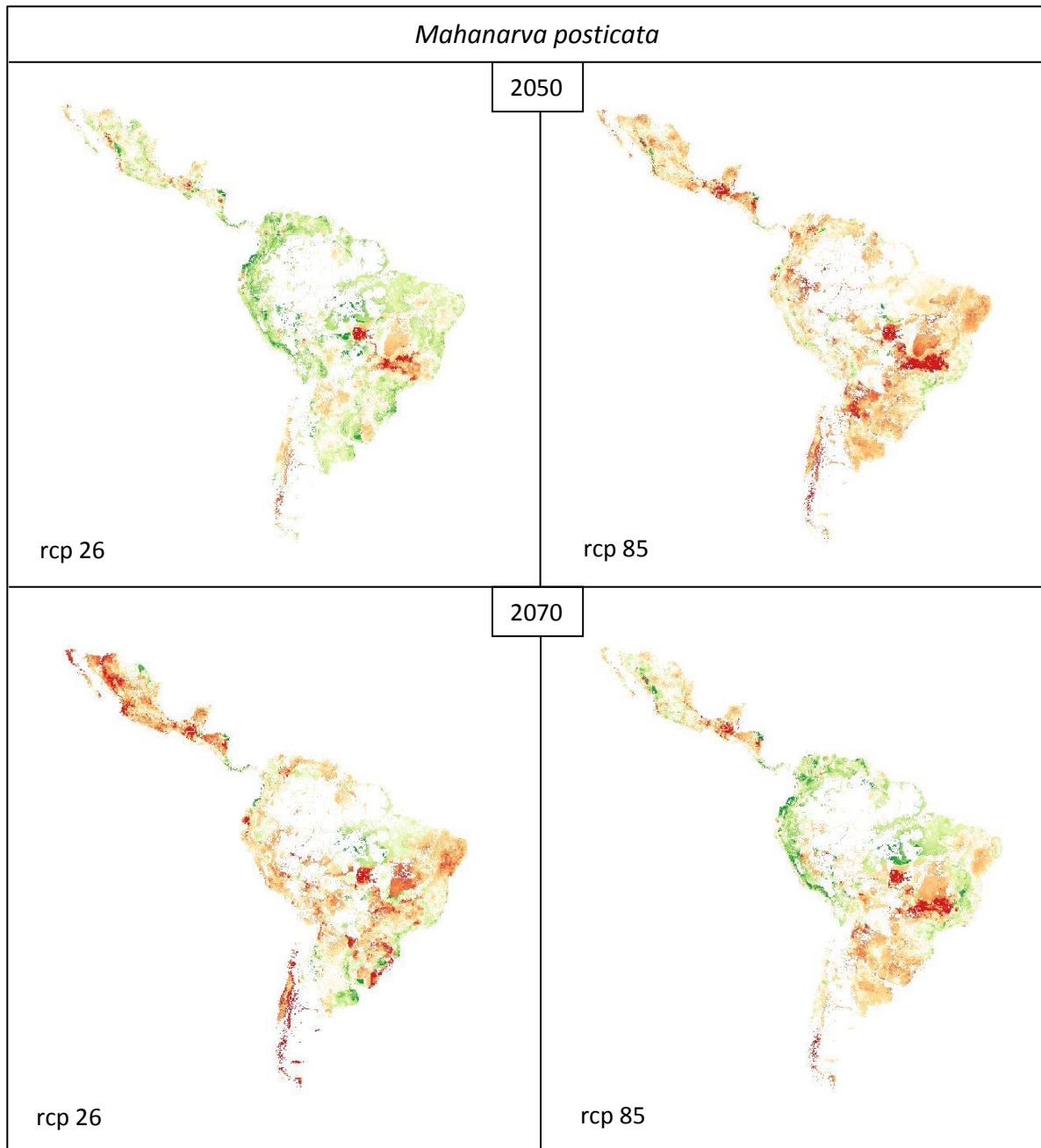
Figure 9: *Mahanarva posticata* species distribution map comparison between future climate scenarios and current climate. For the years 2050 and 2070 the representative concentration pathway (rcp) scenarios 26 and 85 are shown. Green areas illustrate positive changes of suitable habitats in comparison between the future climate scenario and current climate. Vice versa red areas indicate negative changes.

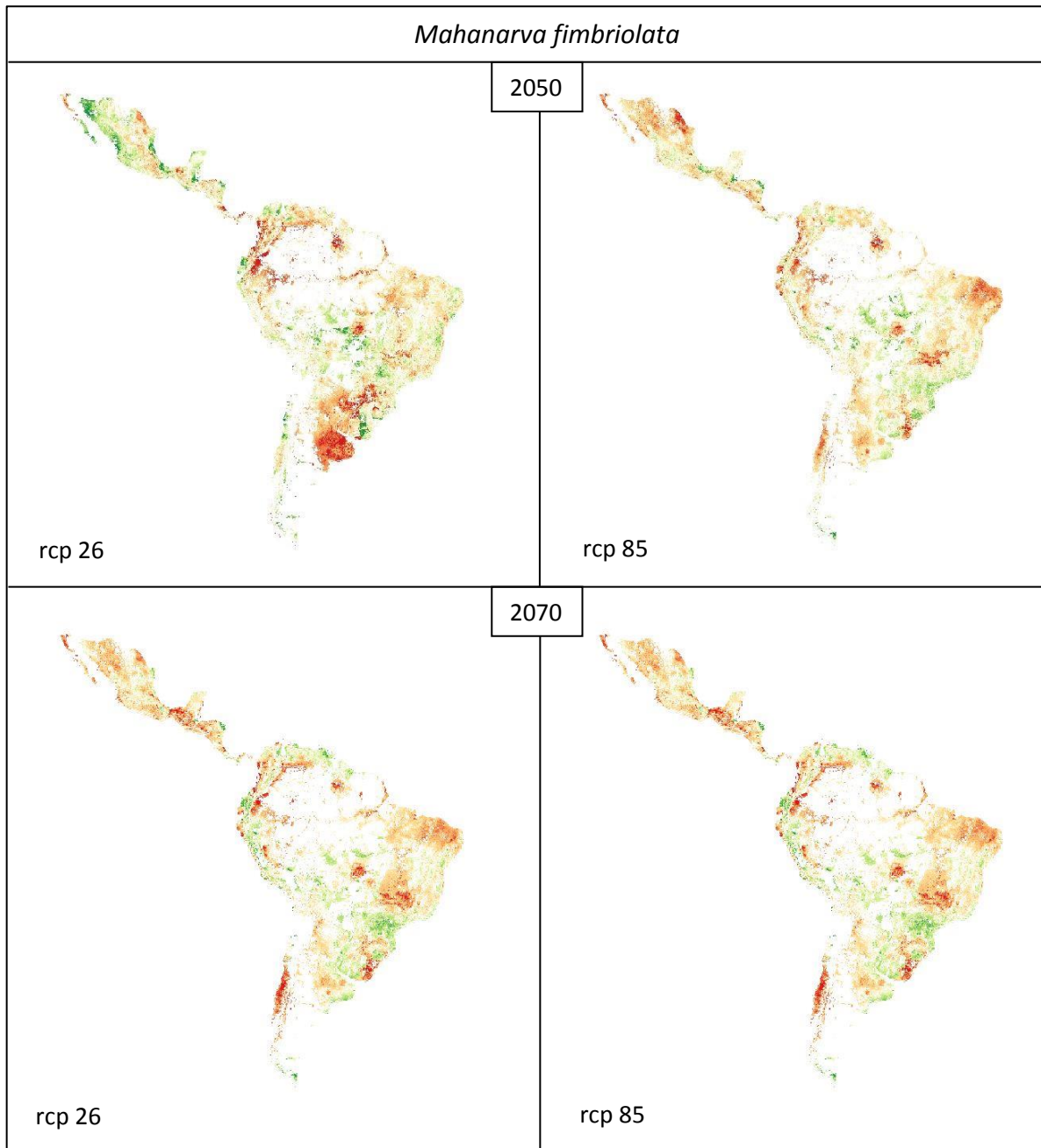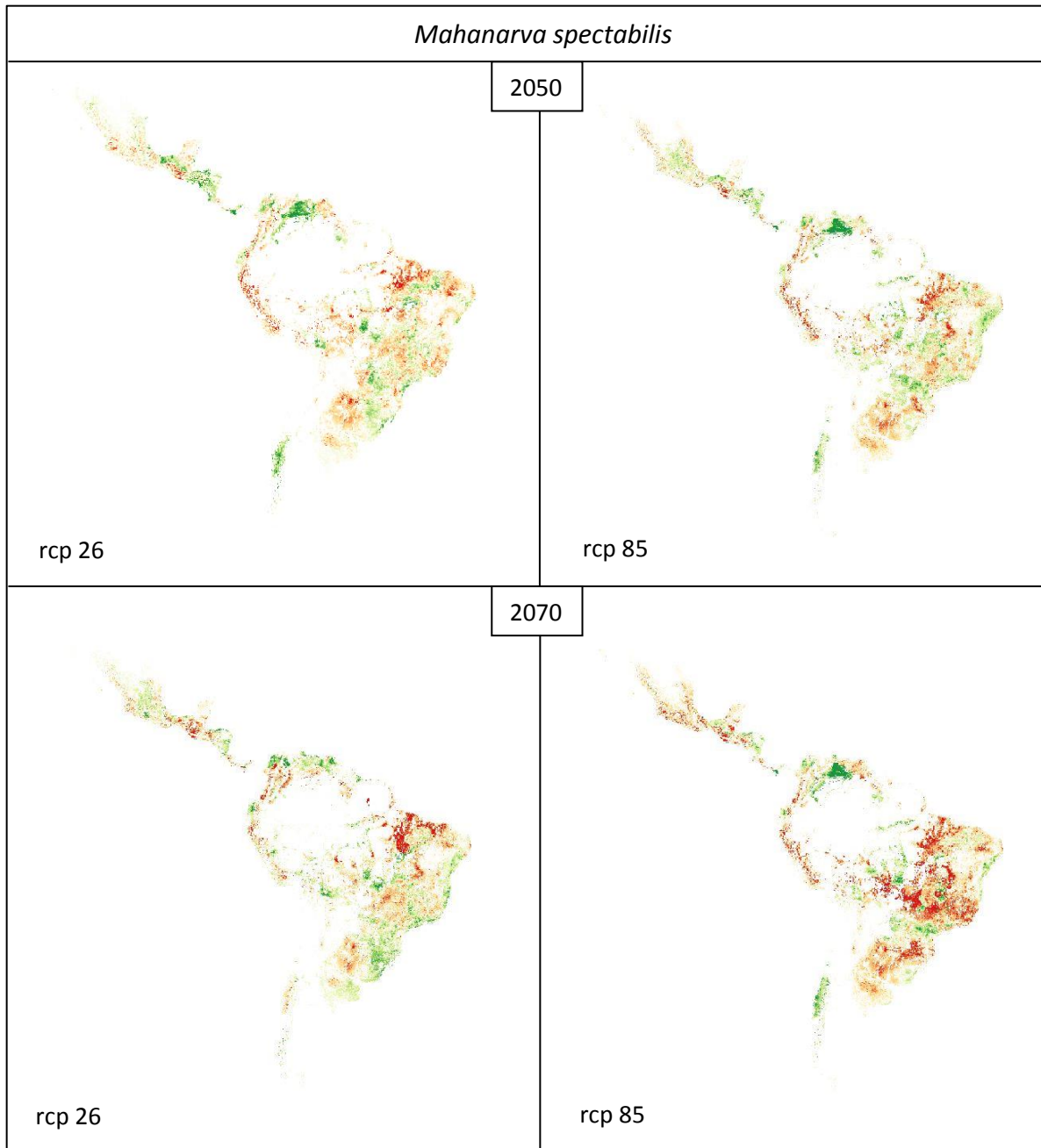Figure 10: *Mahanarva spectabilis* species distribution map comparison between future climate scenarios and current climate. For the years 2050 and 2070 the representative concentration pathway (rcp) scenarios 26 and 85 are shown. Green areas illustrate positive changes of suitable habitats in comparison between the future climate scenario and current climate. Vice versa red areas indicate negative changes.

Tabel 1: Mean pixel values of the Maxent Future climate maps in comparison to the actual climate maps.

Species	Year	rcp	Mean
M. fimbriolata	2050	26	-0.239538847663
M. fimbriolata	2050	45	-0.368699530546
M. fimbriolata	2050	60	0.139328733263
M. fimbriolata	2050	85	-0.283689115344
M. fimbriolata	2070	26	-0.417242042611
M. fimbriolata	2070	45	0.085014976799
M. fimbriolata	2070	60	-0.262382677741
M. fimbriolata	2070	85	-0.29218217623
M. liturata	2050	26	-0.251832974849
M. liturata	2050	45	-0.0823152003522
M. liturata	2050	60	-0.277847202773
M. liturata	2050	85	-0.289699402563
M. liturata	2070	26	-0.357690851721
M. liturata	2070	45	-0.21532392076
M. liturata	2070	60	-0.183541365686
M. liturata	2070	85	-0.204609173696
M. posticata	2050	26	0.237824350805
M. posticata	2050	45	-0.653312411404
M. posticata	2050	60	-0.0402623121323
M. posticata	2050	85	-0.20298115429
M. posticata	2070	26	-0.533634651379
M. posticata	2070	45	-0.51078916083
M. posticata	2070	60	-0.63062988494
M. posticata	2070	85	-0.128722914586
M. spectabilis	2050	26	-0.0562127041142
M. spectabilis	2050	45	0.0606828230846
M. spectabilis	2050	60	0.0536032633601
M. spectabilis	2050	85	-0.0524986530553
M. spectabilis	2070	26	-0.0766607792299
M. spectabilis	2070	45	-0.17760674566
M. spectabilis	2070	60	-0.130360908183
M. spectabilis	2070	85	-0.312027728801

Nearly all mean values in the comparison between future climate maps and current climate maps show negative mean values. Those values indicate that because of future climate conditions the area of suitable habitats for the four *Mahanarva* species will be reduced in comparison to the area under current climate conditions. Only five out of 32 comparisons show a positive mean value and therefore more suitable habitats in future climate scenarios. Overall there is no clear increase or decrease of the mean values between the different future rcp climate scenarios of one species. Just in case of *Mahanarva spectabilis* a clear difference between the scenarios of 2050 and 2070 are visible. Excluding the rcp60 scenario of 2050 all 2050 mean values are lower than the 2070 values.

Variable Contribution Analysis

The *Variable Contributions Analysis* of the *Maxent* algorithm shows that for every species and analysis there are a few important variables with the highest contribution percentages and furthermore about two thirds of the variables did not contribute anything to the analysis. The most important variable for *Mahanarva fimbriolata* is the elevation. Only in two of the nine analyzes the isothermality (mean monthly diurnal range (max temp – min temp) divided by the annual mean temperature) contributed more to the analysis but the elevation still stayed on the second place. The third most contributed variable is the precipitation of the warmest quarter. *Mahanarva spectabilis* shows not such a contribution of one variable. Overall the variables elevation, annual precipitation, precipitation of the coldest quarter and global land cover are contributing with different percentages in the different analysis. Using the future climate variables of 2050 the annual precipitation was the most important contributor and using the data of 2070 it was the precipitation of the coldest quarter. The elevation data always takes the second place only contributing more in the analysis of the current climate with almost identical contribution percentages like the global land cover variable. The precipitation seasonality variable is by far contributing the most in all analyzes of *Mahanarva liturata.* The isothermality variable is always on the second place. The most contributed variable for *Mahanarva posticata* analyzes is the precipitation of the coldest quarter followed by the isothermality. Regarding all species the elevation, the isothermality and the different variables of the precipitation are the most important variables used to analyze the distribution of these species.

Discussion

The results show a distribution potential along areas where sugarcane plantations exist and no potential where plantations are non-existent. For example there is no chance for distribution in the Amazon rainforest (Leff, Ramankutty, & Foley, 2004; Ramankutty, *et. al.*, 2010b). In total all results of the generated models show a slightly wider probable distribution area for *M. fimbriolata* than for the other species. This is not surprising because the species has the widest known distribution of all four species.

The Variable Contribution Analysis shows that besides the elevation, the isothermality – a temperature variable – and the precipitation are the most important variables for the models. Temperature and precipitation are very important abiotic factors for insects. Insects are poikilothermic and therefore the activity of insects depends on the temperature of its habitat. The metabolism of insects works at higher rates at an optimal temperature. This also has an effect on the development of eggs, and juvenile life forms (Jaworski & Hilszczanski, 2013). The precipitation also can have a huge influence on spittlebugs. The nymph produces a foamy excretion for their protection (Carvalho & Webb, 2005). During a rainfall the introduced water can dilute the excretion which leaves the nymph unprotected against predators. Therefore higher precipitation can have a negative effect on spittlebug populations. Overall the most important variable for the models was the elevation which is not included in the WorldClim package. The other important variables are a part of WorldClim. Following Bucklin *et. al.* in 2015 additional variables besides the bioclimatic variables provided by WorldClim only have a minor effect on the accuracy of climate based species distribution models. Because of this the authors suggest to only implement bioclimatic variables. But in this case with the high importance of the elevation and in some cases of the croplands variable it seems that those variables are necessary to create reliable results. Besides that the distribution of each species shows that they only occur in areas with sugarcane plantations. This shows the importance of the host plant for their distribution. But on the other side the other important variables are bioclimatic and therefore in total they seem to be more important.

Following estimations, an up to 4.8°C temperature increase by 2081 – 2100 relative to 1986 – 2005 caused by the climate change is possible (IPCC, 2013, RCP8.5). Due to this increase poikilothermic insects may experience up to five additional life cycles per season which would result in further pest infestation (Yamamura & Kiritani, 1998). In a habitat with limited food resources additional spittlebug's life cycles per season would result in a further distribution of the species. The lack of food forces the spittlebugs to search for new habitats with adequate resources. This also depends on the availability of near, reachable and suitable habitats. Besides that, habitats which do not possess the adequate temperature for the spittlebugs could turn habitable due to the climate change. But the results do not

show a further distribution of the *Mahanarva* species due to the climate change in comparison to the current climate. Even decreased suitable habitats due to the climate change are shown for all species in nearly all future climate scenarios. Because of this it seems that the niche of the analyzed spittlebugs is not limited by temperature but rather by other abiotic factors like precipitation as shown in the Variable Contribution Analysis of the future climate data. Furthermore the populations probably will not reach the food limits of their habitats. The analyzed spittlebugs occur on sugarcane and pasture plantations which present an enormous amount of accessible food sources. The populations are more likely controlled by the use of pesticides or predators. Only in the period between the harvest and the next growth of the plantations, food is not abundant and the spittlebugs could be forced to search for a new habitat. Therefore food resources probably do not force the species to look for a new habitat. This may be an explanation for very similar distribution models with current and future climate data. A higher temperature and additional life cycles do not have an effect on the distribution of those spittlebugs. The other mentioned abiotic factors like rainfall can as well have a great impact on the spittlebugs life. Due to the climate change light rainfall events will decrease and heavy rainfall events increase (Das *et. al.*, 2011). Those heavy rainfalls have a higher chance to wash away the protecting foam of spittlebugs nymphs exposing them to predation.

Which algorithm is the best?

The *Bioclim* algorithm shows the worst results compared to the other algorithms. In other studies the algorithm as well did not perform well (Elith, *et. al.*, 2006). *Bioclim* is mainly used for climate change analyzes. But in this area the algorithm is as well outperformed by *Domain* and *Maxent* (Hijmans & Graham, 2006). Therefore the, in this study presented poor results are not surprising. The *Domain* algorithm as well did not perform well in another study compared to *Maxent* and other algorithms but it was frequently used for species distribution modeling (Elith, *et. al.*, 2006; Hijmans & Elith, 2017). The generalized linear models (GLM's) scored intermediate AUC results in the study of Elith, *et.al.* in 2006 but in the cited study the number of used presence records for the species is significantly higher than in the present study. Therefore it seems that the GLM algorithms are not performing well with a low number of presence records. Besides that GLM's are the only algorithms with negative numbers on their scales. It is impossible to have a negative result. The algorithms are analyzing whether there are suitable areas for the distribution of the species. For this the niches of the species are analyzed and compared with habitats on the map. Not suitable habitats do not bear a chance of distribution. Therefore the minimal value of each scale should be zero. A negative value does not fit into this scale. In addition all equal high AUC values occur within the GLM algorithms (Attachments: Table 2 AUC).

Moreover only in the GLM's AUC values of one were achieved which is as well unlikely regarding the same pattern on each map for each species. Because of this reasons it seems that despite the sometimes high AUC, those algorithms are not adequate to analyze the distribution of the four *Mahanarva* species. Maxent is the most used species distribution algorithm and performed very well in other comparison studies (Hijmans & Elith, 2017; Elith, *et. al.*, 2006). In a comparison between *GARP* (genetic algorithm) models and *Maxent* using jackknife tests and only small numbers of species occurrence points the last mentioned algorithm was able to perform better with small sample sizes. Furthermore *Maxent* has a considerably higher ability to predict excluded localities. Therefore analyzes with a small sample sizes are best used to identify similar regions to where the species is known to occur and not to predict the range limits of species (Ashraf, *et. al.*, 2017; Pearson, *et. al.*, 2006). In general the AUC results of the algorithms regarding the distribution of the four *Mahanarva* species are quite similar compared to other studies. *Maxent* is the best performing algorithm not only using current but also future climate data.

Small Number of occurrence points

The small number of occurrence records of the species is problematic. As shown in a study in Africa a minimum number of 14 occurrence points for narrow-ranged and 25 for widespread species is necessary for the good quality of an algorithm. In some cases where species have a very low prevalence, the minimum number can also be lower than 14. Furthermore the researchers pointed out that model performance is increasing with an increasing number of occurrence records (Van Proosdij, *et. al*, 2016).The difficulty is to define what is a narrow-ranged and what is a widespread species. Especially regarding that the proper distribution of the four species and especially for *M. posticata* is not well known. All four species are found from south to central or northeast Brazil. Concerning the size of Brazil those species can be considered wide ranged species. Only *M. spectabilis* with 29 occurrence records passed the minimum number of 25 and only for this species all analyzes showed the best performance using the *Maxent* algorithm. All other species are below the minimum number of occurrence records. *M. posticata* only has 9 occurrence records and is thereby far beneath the level of a good performance of an algorithm.

Another study also shows decreasing model accuracy with a decreasing sample size. But none of the analyzed algorithms performed well with a sample size beneath 30. Nonetheless the *Domain* algorithm had an intermediate performance and *Maxent* was not very sensitive to the sample size which resulted in one of the best performances in comparison to the other algorithms studied (Wisz, *et. al.*, 2008). Also the results of the 2006 study by Pearson, *et. al.* demonstrated that analyzes with a small sample

size have a great value because of significant predictive abilities in the jackknife test (Pearson, *et. al.*, 2006).

Considering the comparison between the future and the current climate the maps of *Mahanarva spectabilis* show the most uniform results with very similar changes in habitat suitability. The maps of the other species with less distribution records are quite different. This results alone are no indicator for better results because all those analyses are based on the already present distribution maps.

Further studies

In further studies the aim is to implement the *Maxlike* algorithm. *Maxlike* is a simple likelihood-based alternative to *Maxent* (Chandler, Royle, & Kindt, 2017). The algorithm estimates the probability of occurrence unlike *Maxent*. Therefore *Maxlike* is a promising likelihood-based alternative to *Maxent* and it will be interesting how the algorithm performs with a low number of presence records.

In general more presence records are necessary to create better analyzes. Especially *M. posticata* has a very low number of known occurrence points. Furthermore it would be interesting which algorithm performed the best with a sufficient number of occurrence records. Would *Maxent* score the highest AUC values for every species or would other algorithms score good values as well? Besides that analyzes with different sets of variables could be carried out to resolve the question if non bioclimatic variables have an influence on the quality and accuracy of the algorithms.

Literature

Almeida, J. E. M. de, Batista Filho, A., & Santos, A. S. (2003). Avaliação do controle biológico de *Mahanarva fimbriolata* (Hom.,Cercopidae) com o fungo *Metarhizium anisopliae* em variedades de cana-de-açucar e differentes épocas de corte. *Arquivos do Instituto Biológico* 70(1): 101–103.

Almeida, C. L. de, Stingel, E., Primo, S. C. V., Sanguino, A., Burnquist, W.L., De Beni Arrigoni, E., & Figueiredo Filho, O. (2006). O controle biológico da cigarrinha das raízes da cana-de-açucar, *Mahanarva fimbriolata*. WorkShop STAB, Piracicaba.

Alves, R. T., & Carvalho, G. S. (2014). Primeiro registro das espécies de cigarrinhas-da-raiz da cana-de-açúcar *Mahanarva spectabilis* (Distant) e *Mahanarva liturata* (Le Peletier & Serville) atacando canaviais na região de Goianésia (GO), Brasil. *Arquivos do Instituto Biológico*, 81(1): 82–84.

Arino, O., Ramos, P., Jose, J., Kalogirou, V., Bontemps, S., Defourny, P., & Van Bogaert, E. (2012). Global Land Cover Map for 2009 (GlobCover 2009). © European Space Agency (ESA) & Université catholique de Louvain (UCL), PANGAEA. https://doi.org/10.1594/PANGAEA.787668.

Ashraf, U., Peterson, A. T., Chaudhry, M. N., Ashraf, I., Saqib, Z., Ahmad, S. R., & Ali, H. (2017). Ecological niche model comparison under different climate scenarios: a case study of *Olea* spp. in Asia. *Ecosphere*, 8(5): e01825.10.1002/ecs2.1825.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., & Rouault, E. (2017). Package 'rgdal'. Available online at: https://cran.r-project.org/web/packages/rgdal/index.html.

Bivand, R., Lewin-Kohl, N., Pebesma, E.,Archer, E., Baddeley, A., Bearman, N., Bibiko, H., Brey, S., Callahan, J., Carillo, G., Dray, S., Forrest, D., Friendly, M., Giraudoux, P., & Golicher, D. (2017). Package'maptools'. Available at: https://cran.r-project.org/web/packages/maptools/index.html.

Bucklin, D. N., Basille, M., Benscoter, A. M., Brandt, L. A., Mazotti, F. J., Romanach, S. S., Speroterra, C., & Watling, J. I. (2015). Comparing species distribution models constructed with different subsets of environmental predictors. *Diversity and Distributions*, 21:23–35.

Busby, J. R. (1991). BIOCLIM- a bioclimate analysis and prediction system. In C. R. Margulles & M. P. Austin (Eds.) *Nature Conservation: Cost Effective Biological Surveys and Data Analysis* (pp. 64–68) CSIRO Australia.

Carvalho, G. S. (1978). *As cigarrinhas das pastagens*. Porto Alegre. Pontifícia Universidade Católica do Rio Grande do Sul, RS, Porto Alegre.

Carvalho, G. S., & Webb, M. D. (2005). *Cercopid Spittle Bugs of the New World*. Pensoft, Sofia.

Chandler, R., Royle, A., & Kindt, R. (2017). Package 'maxlike'.Available online at: https://cran.r-project.org/web/packages/maxlike/maxlike.pdf. [Accessed: 18.12.2017].

Das, D. K, Singh, J., & Venilla, S. (2011) Emerging Crop Pest Scenario under the Impacto f CLimate Change – A Brief Review. *Journal of Agricultural Physics*, 11:13–20.

Distant, W. L. (1909). Rhynchotal notes xlvi. *Annals and Magazine of Natural History*, 8(3):187–213.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettemann,

F., Leathwick. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Moritz, G. M. C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Petersen, A. T., Phillips, S. T., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., & Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29:129–151.

Falling Rain Software, Ltd. (2017). Global Gazeteer Version 2.3. (Available at: www.fallingrain.com/world/index.html). (Access: May 12, 2016).

Franklin, J, (2009). *Mapping species distributions*., Cambridge University Press, New York.

Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. doi: http://dx.doi.org/10.1002/joc.5086.

Garcia, J. F., Botelho, P. S. M., & Parra, J. R. P. (2006). Biology and fertility life table of *Mahanarva fimbriolata* (Stal) (Hemiptera: Cercopidae) in sugarcane. *Scientia Agricola*, 63(4):317–320.

GBIF. (2016). Global Biodiversity Information Facility (GBIF). (Available at: www.gbif.org). (Access: July 23, 2016).

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

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25:1965–1978.

Hijmans, R. J., & Elith, J. (2017). Species distribution modeling with R. (Available at: https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf). (Accessed May 30, 2017).

Hijmans, R. J., & Graham, C. H. (2006). Testing the ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology,* 12:2272-2281, DOI: http://dx.doi.org/10.1111/j.1365-2486.2006.01256.x.

Hijmans, R. J, Phillips, S., Leathwick, J., & Elith, J. (2011). Package 'dismo'. (Available at: http://cran.r-project.org/web/packages/dismo/index.html). (Accessed: November 11, 2017).

Hijmans, R. J., & Van Etten, J. (2012). Raster: Geographic analysis and modeling with raster data. R package version 2.0-12. (Available at: http://CRAN.R-project.org/package=raster.). (Accessed: November 4, 2017].

IBGE. (2016). Instituto Brasileiro de Geografia e Estatística – Sistema IBGE de Recuperação de Dados (SIDRA). Available online at: www.sidra.ibge.gov.br [Accessed: April 14, 2016].

International Sugar Organization. (2017). Sugar Year Book 2017. Available online at: https://www.isosugar.org/content/pages/SYB%202017-69d562.pdf [Accessed: March 09, 2017].

IPCC. (2013). *Climate change 2014: the physical science basin.* Cambridge University Press, New York. USA.

Jaworski, T., & Hilszczanski, J. (2013). The effect of temperature and humidity changes on insects development and their impact on forest ecosystems in the context of expected climate change. *Forest Research Papers*, 74(4):345–355. DOI: 10.2478/frp-2013-0033.

Leff, B., Ramankutty, N., & Foley, J. A. (2004). Geographic distribution of major crops across the world. *Global Biogeochemical Cycles,* 18(1). Doi:10.1029/2003GB002108.

Naimi, B., & Araujo, M. B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368–375, DOI: 10.1111/ecog.01881.

Natural Earth. (2017). Natural Earth I with Shaded Relief. (Available at: http://www.naturalearthdata.com/downloads/10m-raster-data/10m-natural-earth-1/) . (Accessed: November 04, 2017).

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. (2006). Predicting species distribution from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34:102–117.

Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. R News 5 (2). (Available at: https://cran.r-project.org/doc/Rnews/). (Accessed: November 04, 2017).

Peterson, A. T. (2003). Predicting the geography of species' invasions via ecological niche modeling.*The Quarterly Review of Biology*, 78(4):419–433.

Peterson, A. T., Papes, M., & Soberon, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niches modeling. *Ecological Modeling*, 213:63–72.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190:231–259.

QGIS Development Team. (2009). QGIS Geographic Information System. Open Source Geospatial Foundation. (Available at: http://qgis.osgeo.org). (Accessed: November 04, 2017).

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2010a). Global Agricultural Lands: Croplands, 2000. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). (Available at: http://dx.doi.org/10.7927/H4C8276G). (Accessed: November 04, 2017).

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2010b). Global Agricultural Lands: Pastures, 2000. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). (available at: http://dx.doi.org/10.7927/H47H1GGR). (Accessed: November 05, 2017).

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000. *Global Biogeochemical Cycles* 22 (1). (Available at: http://dx.doi.org/10.1029/2007GB002952). (Accessed November 05, 2017).

Resende, T. T., Auad, A. M., Fonseca, M. G., Sobrinho, F. S., Ribeiro dos Santos, D., & Barbosa da Silva, S. E. (2013).The Damage Capacity of *Mahanarva spectabilis*(Distant, 1909) (Hemiptera: Cercopidae) Adults on *Brachiaria ruziziensis* Pasture. *The ScientificWorld Journal*. http://dx.doi.org/10.1155/2013/281295.

R Development Core Team. (2008). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. (Available at: http://www.R-project.org). (Accessed: November 04, 2017).

RStudio Team. (2016). *RStudio: Integrated Development for R.* RStudio, Inc., Boston, MA. (Available at: http://www.rstudio.com/). (Accessed: November 04, 2017).

Soulier-Perkins, A. (2016). COOL - Cercopoidea Organised On Line. (Available at: www.hemiptera-databases.org/cool). (Accessed: April 18, 2016).

Simões, A. D., Lima, E. R., Auad, A. M., Resende, T. T., &Leite, M. V. (2013). Development of the spittlebug *Mahanarva fimbriolata* under varying photophase conditions. *Journal of Insect Science*, 13:105. (Available online: http://www.insectscience.org/13.105).

Species Link. (2017). Species Link. (Available at: splink.cria.org.br). (Accessed: November 04, 2017).

Van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, 39(6):542–552.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisam, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14:763–773.

Yamamura, K., & Kiritani, K. (1998). A simple method to estimate the potential increase in the number of generations under global warming in temperate zones. *Applied Entomology and Zoology*, 33(2):289–298.

You, L., Wood-Sichra, U., Fritz, S., Guo, Z., See, L., & Koo, J. I. (2014). Spatial Production Allocation Model (SPAM) 2005 v2.0. (Available at: http://mapspam.info). (November 5, 2017).

Attachments

AUC

Tabel 2: Area Under the Curve results of all four species regarding the different algorithms *Bioclim, Domain, GLM Binomial, GLM Gaussian, GLM Poisson* and *Maxent*. Separation of the tables by current, 2050 and 2070 climate data and further distinction by the four *rcp* concentrations 26, 45, 60 and 85 in case of the future climate data. Yellow boxes each present the highest and red boxes the lowest AUC scores.

Current Climate

AUC

	M. fimbriolata	*M. spectabilis*	*M. liturata*	*M. posticata*
Bioclim	0,715812	0,8432075	0,95	0,9166667
Domain	0,8568376	0,8379245	0,9644068	0,9153846
GLM Binomial	0,8739316	0,8701887	1	0,7384615
GLM Gaussian	0,8397436	0,8539623	0,9864407	0,8410256
GLM Poisson	0,8611111	0,8698113	0,98305	0,7692308
Maxent	0,8824786	0,9358491	0,9966102	0,974359

2050 rcp26

AUC

	M. fimbriolata	*M. spectabilis*	*M. liturata*	*M. posticata*
Bioclim	0,7222222	0,8252174	0,9459184	0,9166667
Domain	0,8688889	0,8504348	0,9795918	0,9166667
GLM Binomial	0,8111111	0,8752174	0,9693878	0,9070513
GLM Gaussian	0,7266667	0,8704348	0,9979592	0,9839744
GLM Poisson	0,7977778	0,8708696	0,9346939	0,7980769
Maxent	0,7688889	0,9269565	0,9959184	0,9647436

2050 rcp45

AUC

	M. fimbriolata	*M. spectabilis*	*M. liturata*	*M. posticata*
Bioclim	0,7222222	0,8721739	0,8181818	0,9166667
Domain	0,88	0,8491304	0,9404959	0,9166667
GLM Binomial	0,8666667	0,8826087	0,9603306	0,8621795
GLM Gaussian	0,8355556	0,8504348	0,9768595	0,9775641
GLM Poisson	0,84	0,8826087	0,9570248	0,8653846
Maxent	0,4240359	0,9165217	0,9938776	0,974359

2050 rcp60

AUC

	M. fimbriolata	M. spectabilis	M. liturata	M. posticata
Bioclim	0,7077778	0,8476087	0,9306122	0,9166667
Domain	0,86	0,8426087	0,9795918	0,9150641
GLM Binomial	0,8022222	0,8565217	0,9591837	1
GLM Gaussian	0,7666667	0,8513043	0,9938776	0,9839744
GLM Poisson	0,8177778	0,8573913	0,9326531	1
Maxent	0,7777778	0,9	0,9959184	0,9807692

2050 rcp85

AUC

	M. fimbriolata	M. spectabilis	M. liturata	M. posticata
Bioclim	0,7222222	0,8341304	0,9306122	0,9166667
Domain	0,8733333	0,8473913	0,9591837	0,9150641
GLM Binomial	0,8733333	0,7365217	1	0,8557692
GLM Gaussian	0,8466667	0,7082609	0,977551	0,9519231
GLM Poisson	0,9	0,71	0,9673469	0,9102564
Maxent	0,8555556	0,93	0,9959184	0,9775641

2070 rcp26

AUC

	M. fimbriolata	M. spectabilis	M. liturata	M. posticata
Bioclim	0,7222222	0,88	0,9306122	0,9166667
Domain	0,8533333	0,873913	0,9795918	0,9166667
GLM Binomial	0,8733333	0,8904348	0,9306122	0,8269231
GLM Gaussian	0,8155556	0,8813043	0,9959184	0,9807692
GLM Poisson	0,8711111	0,886087	0,9591837	0,8076923
Maxent	0,8577778	0,9169565	0,9959184	0,9423077

2070 rcp45

AUC

	M. fimbriolata	M. spectabilis	M. liturata	M. posticata
Bioclim	0,7222222	0,8432609	0,9306122	0,9166667
Domain	0,8236715	0,8513043	0,9795918	0,9166667
GLM Binomial	0,8550725	0,8982609	0,9510204	0,9070513
GLM Gaussian	0,7922705	0,8886957	0,9959184	0,9807692
GLM Poisson	0,8067633	0,8969565	0,9102041	0,9038462
Maxent	0,8574879	0,9378261	0,9959184	0,9807692

2070 rcp60

AUC

	M. fimbriolata	M. spectabilis	M. liturata	M. posticata
Bioclim	0,7222222	0,8265217	0,9459184	0,9166667
Domain	0,8755556	0,8519565	0,9795918	0,9166667
GLM Binomial	0,8288889	0,8617391	0,8979592	0,9775641
GLM Gaussian	0,76	0,8547826	0,9795918	0,9839744
GLM Poisson	0,8355556	0,8621739	0,9183673	0,9807692
Maxent	0,8	0,9278261	0,9959184	0,9839744

2070 rcp85

AUC

	M. fimbriolata	M. spectabilis	M. liturata	M. posticata
Bioclim	0,7222222	0,8245652	0,9306122	0,9166667
Domain	0,8244444	0,8495652	0,9571429	0,8974359
GLM Binomial	0,8711111	0,8917391	0,965306	0,8685897
GLM Gaussian	0,8311111	0,863913	0,9877551	0,9775641
GLM Poisson	0,8933333	0,8747826	0,955102	0,8685897
Maxent	0,8695652	0,9382609	0,9959184	0,9583333

Distribution Coordinates

Tabel 3: Occurence records of *Mahanarva fimbriolata*, *Mahanarva spectabilis*, *Mahanarva liturata* and *Mahanarva posticata*.

| M. fimbriolata		M. spectabilis		M. liturata		M. posticata	
longitude	latitude	longitude	latitude	longitude	latitude	longitude	latitude
-35.17	-7.9	-43.35	-21.75	-45.35	-22.45	-35.03	-8.28
-40.23	-15.25	-43.36	-4.87	-49.72	-29.33	-42.80	-20.43
-42.64	-21.53	-44.36	-6.37	-50.15	-25.08	-43.35	-21.76
-42.80	-20.43	-47.37	-5.93	-51.18	-29.93	-48.54	-27.16
-43.35	-21.76	-47.46	-7.33	-51.20	-30.03	-48.58	-27.59
-43.96	-19.9	-47.65	-15.62	-51.23	-30.02	-48.83	-25.48
-45.25	-20.47	-47.91	-15.76	-51.32	-30.11	-50.21	-29.68
-47.38	-22.36	-48.03	-9.36	-51.33	-30.10	-51.98	-22.53
-47.64	-22.36	-48.03	-9.37	-51.45	-29.10	-56.27	-16.17
-48.54	-27.31	-48.08	-9.14	-51.51	-29.19		
-51.82	-30.85	-48.52	-6.54	-51.52	-29.17		
-51.98	-22.53	-48.55	-27.15	-51.98	-22.53		
-54.32	-27.62	-48.72	-23.39	-53.70	-29.70		
-54.38	-20.26	-49.36	-5.17	-53.89	-30.81		
-55.67	-28.25	-50.20	-22.42	-54.58	-18.08		
-55.75	-57.68	-50.33	-22.46	-55.50	-15.22		
-56	-28.65	-51.17	-23.28	-56.10	-15.60		
-57.68	-16.07	-51.26	-16.23				
-60.13	-12.74	-51.38	-22.12				
		-51.95	-6.56				
		-54.04	-24.54				
		-54.43	-23.84				
		-54.61	-20.37				
		-55.37	-7.14				
		-55.36	-7.12				
		-55.62	-11.85				
		-60.13	-12.74				
		-63.84	-8.81				
		-63.90	-8.77				

Variable Layers

Elevation

Globecover

Pasture
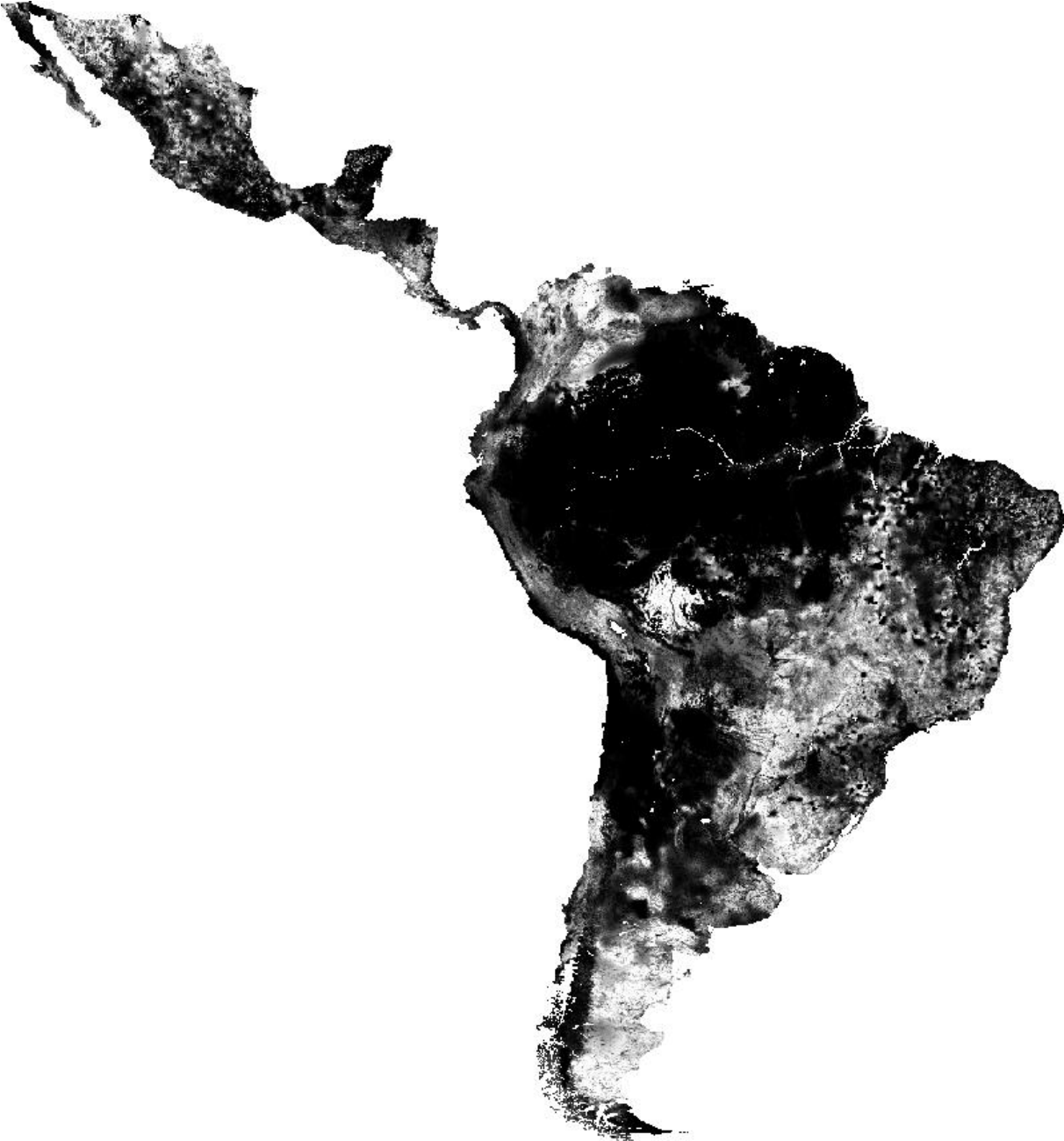
Croplands

Sugarcane Physical Area

Bioclim Current Climate
BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) (* 100)

BIO4 = Temperature Seasonality (standard deviation *100)

BIO5 = Max Temperature of Warmest Month

BIO6 = Min Temperature of Coldest Month
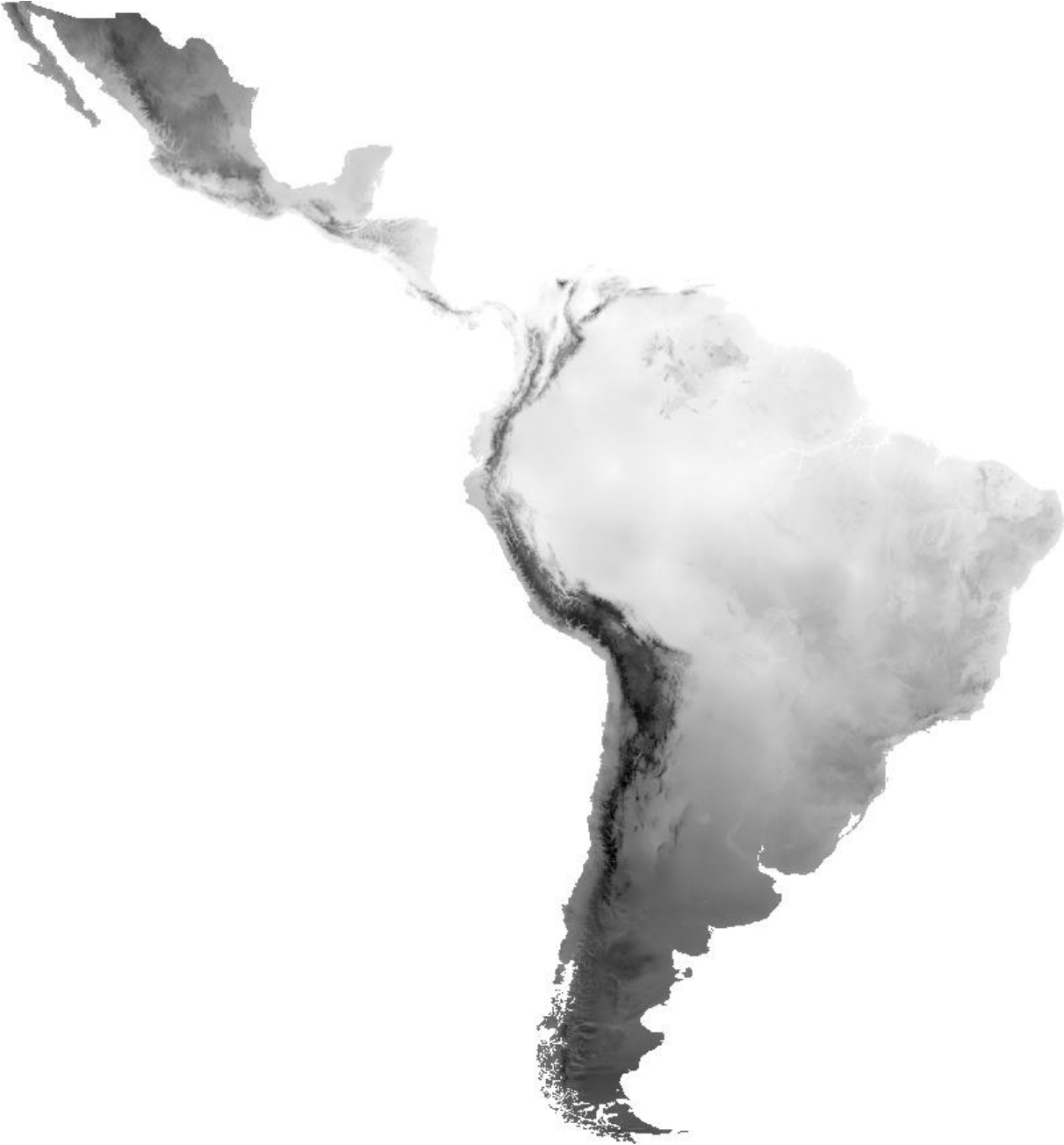
BIO7 = Temperature Annual Range (BIO5-BIO6)

BIO8 = Mean Temperature of Wettest Quarter
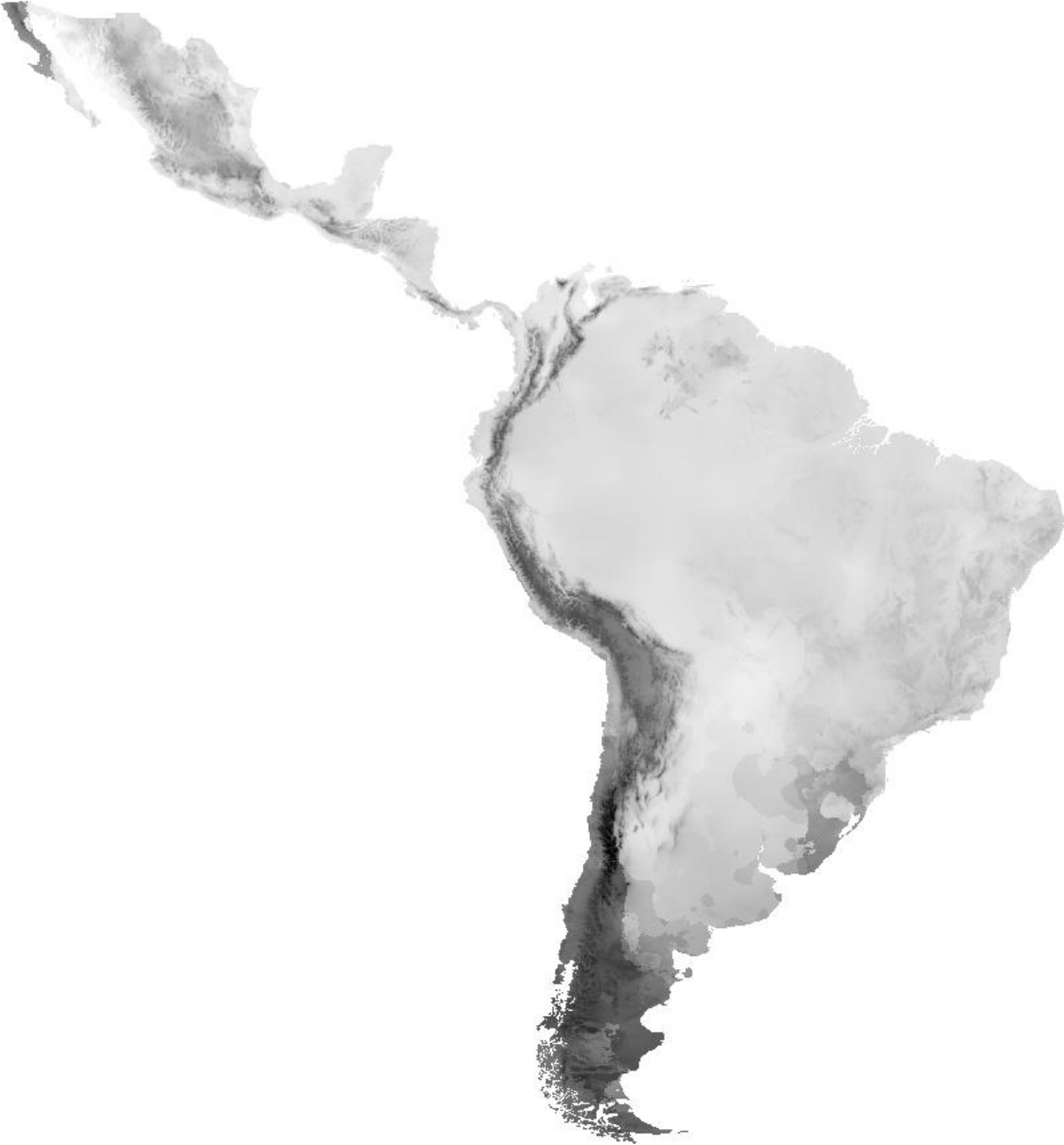
BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

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

R Code

###################################

#Based on Hijmans & Elith 2017

#Hijmans, R.J. & Elith, J. (2017) Species distribution modeling with R. Available online at

https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf (accessed May 30, 2017).

```
library(dismo)
library(raster)
library(sp)
library(rgdal)
library(maptools)
data("wrld_simpl")
library(rJava)
library(sdm)
library(usdm)
```

```
fimbriolata<-read.table("mahanarva_fimbriolata_coordinates_clean.txt", header=TRUE)
```

```
#Deleting repeatetly coordinates.
#Here all duplicates with the same coordinates are selected. The species name does not matter.
fimbriolata_duplicate<-duplicated(fimbriolata[, c('lon', 'lat')])
sum(fimbriolata_duplicate)
#Now creating a table with only one of every coordinate.
fimbriolata_clean<- fimbriolata[!fimbriolata_duplicate, ]
#fimbriolata_clean has only one of every coordinate! From now on only use fimbriolata_clean!
```

#Bioclim Raster Layer and other Current Climate Variables (Bioclim Version 2.0).

```
bio1_2.5min<-file.choose(); x <- GDAL.open(bio1_2.5min)
bio2_2.5min<-file.choose(); x <- GDAL.open(bio2_2.5min)
bio3_2.5min<-file.choose(); x <- GDAL.open(bio3_2.5min)
bio4_2.5min<-file.choose(); x <- GDAL.open(bio4_2.5min)
bio5_2.5min<-file.choose(); x <- GDAL.open(bio5_2.5min)
bio6_2.5min<-file.choose(); x <- GDAL.open(bio6_2.5min)
```

```
bio7_2.5min<-file.choose(); x <- GDAL.open(bio7_2.5min)
bio8_2.5min<-file.choose(); x <- GDAL.open(bio8_2.5min)
bio9_2.5min<-file.choose(); x <- GDAL.open(bio9_2.5min)
bio10_2.5min<-file.choose(); x <- GDAL.open(bio10_2.5min)
bio11_2.5min<-file.choose(); x <- GDAL.open(bio11_2.5min)
bio12_2.5min<-file.choose(); x <- GDAL.open(bio12_2.5min)
bio13_2.5min<-file.choose(); x <- GDAL.open(bio13_2.5min)
bio14_2.5min<-file.choose(); x <- GDAL.open(bio14_2.5min)
bio15_2.5min<-file.choose(); x <- GDAL.open(bio15_2.5min)
bio16_2.5min<-file.choose(); x <- GDAL.open(bio16_2.5min)
bio17_2.5min<-file.choose(); x <- GDAL.open(bio17_2.5min)
bio18_2.5min<-file.choose(); x <- GDAL.open(bio18_2.5min)
bio19_2.5min<-file.choose(); x <- GDAL.open(bio19_2.5min)
pasture<-file.choose(); x <- GDAL.open(pasture)
croplands<-file.choose(); x <- GDAL.open(croplands)
sugarcane<-file.choose(); x <- GDAL.open(sugarcane)
elevation<-file.choose(); x <- GDAL.open(elevation)
globcover<-file.choose(); x <- GDAL.open(globcover)

#2050 Climate Data Bioclim
#rcp26

rcp26bi01<-file.choose(); x <- GDAL.open(rcp26bi01)
rcp26bi02<-file.choose(); x <- GDAL.open(rcp26bi01)
rcp26bi03<-file.choose(); x <- GDAL.open(rcp26bi03)
rcp26bi04<-file.choose(); x <- GDAL.open(rcp26bi04)
rcp26bi05<-file.choose(); x <- GDAL.open(rcp26bi05)
rcp26bi06<-file.choose(); x <- GDAL.open(rcp26bi06)
rcp26bi07<-file.choose(); x <- GDAL.open(rcp26bi07)
rcp26bi08<-file.choose(); x <- GDAL.open(rcp26bi08)
rcp26bi09<-file.choose(); x <- GDAL.open(rcp26bi09)
rcp26bi10<-file.choose(); x <- GDAL.open(rcp26bi10)
rcp26bi11<-file.choose(); x <- GDAL.open(rcp26bi11)
rcp26bi12<-file.choose(); x <- GDAL.open(rcp26bi12)
rcp26bi13<-file.choose(); x <- GDAL.open(rcp26bi13)
```

```
rcp26bi14<-file.choose(); x <- GDAL.open(rcp26bi14)

rcp26bi15<-file.choose(); x <- GDAL.open(rcp26bi15)

rcp26bi16<-file.choose(); x <- GDAL.open(rcp26bi16)

rcp26bi17<-file.choose(); x <- GDAL.open(rcp26bi17)

rcp26bi18<-file.choose(); x <- GDAL.open(rcp26bi18)

rcp26bi19<-file.choose(); x <- GDAL.open(rcp26bi19)

pasture<-file.choose(); x <- GDAL.open(pasture)

croplands<-file.choose(); x <- GDAL.open(croplands)

sugarcane<-file.choose(); x <- GDAL.open(sugarcane)

elevation<-file.choose(); x <- GDAL.open(elevation)

globcover<-file.choose(); x <- GDAL.open(globcover)


#Repeat the procedure with the other variable for the different rcp's and years. Don't forget to
#adjust the names.


#2.5 Current Climate Data
predictors_2.5min<-stack(bio1_2.5min, bio2_2.5min, bio3_2.5min, bio4_2.5min, bio5_2.5min,
        bio6_2.5min, bio7_2.5min, bio8_2.5min, bio9_2.5min, bio10_2.5min, bio11_2.5min,
        bio12_2.5min, bio13_2.5min, bio14_2.5min, bio15_2.5min, bio16_2.5min, bio17_2.5min,
        bio18_2.5min, bio19_2.5min, croplands, pasture, sugarcane, globcover, elevation)
names(predictors_2.5min)


#rcp 26
predictors_2050rcp26<-stack(rcp26bi01, rcp26bi02, rcp26bi03, rcp26bi04, rcp26bi05, rcp26bi06,
        rcp26bi07, rcp26bi08, rcp26bi09, rcp26bi10, rcp26bi11, rcp26bi12, rcp26bi13,
        rcp26bi14,rcp26bi15, rcp26bi16, rcp26bi17, rcp26bi18, rcp26bi19, croplands, pasture,
        sugarcane, globcover, elevation)
names(predictors_2050rcp26)


#rcp 45
predictors_2050rcp45<-stack(rcp45bi01, rcp45bi02, rcp45bi03, rcp45bi04, rcp45bi05, rcp45bi06,
        rcp45bi07, rcp45bi08, rcp45bi09, rcp45bi10, rcp45bi11, rcp45bi12, rcp45bi13,
        rcp45bi14,rcp45bi15, rcp45bi16, rcp45bi17, rcp45bi18, rcp45bi19, croplands, pasture,
        sugarcane, globcover, elevation)
names(predictors_2050rcp45)
```

```
#rcp 60
predictors_2050rcp60<-stack(rcp60bi01, rcp60bi02, rcp60bi03, rcp60bi04, rcp60bi05, rcp60bi06,
        rcp60bi07, rcp60bi08, rcp60bi09, rcp60bi10, rcp60bi11, rcp60bi12, rcp60bi13,
        rcp60bi14,rcp60bi15, rcp60bi16, rcp60bi17, rcp60bi18, rcp60bi19, croplands, pasture,
        sugarcane, globcover, elevation)
names(predictors_2050rcp60)


#rcp85
predictors_2050rcp85<-stack(rcp85bi01, rcp85bi02, rcp85bi03, rcp85bi04, rcp85bi05, rcp85bi06,
        rcp85bi07, rcp85bi08, rcp85bi09, rcp85bi10, rcp85bi11, rcp85bi12, rcp85bi13,
        rcp85bi14,rcp85bi15, rcp85bi16, rcp85bi17, rcp85bi18, rcp85bi19, croplands, pasture,
        sugarcane, globcover, elevation)
names(predictors_2050rcp85)


v2 <- vifcor(predictors_2050rcp85, th=0.9)
v2


#Bioclim Current Cimate Data
predictors<-stack(bio2_2.5min, bio3_2.5min, bio6_2.5min, bio8_2.5min, bio9_2.5min, bio12_2.5min,
        bio13_2.5min, bio14_2.5min, bio15_2.5min, bio18_2.5min, bio19_2.5min, croplands,
        pasture, sugarcane, elevation, globcover)


#Bioclim Future Climate Data
predictors<-stack(rcp85bi02, rcp85bi03, rcp85bi04, rcp85bi05, rcp85bi08, rcp85bi09, rcp85bi12,
        rcp85bi13, rcp85bi14, rcp85bi15, rcp85bi18, rcp85bi19, croplands, pasture, sugarcane,
        globcover, elevation)
names(predictors)
#Before the predictors were selected analysing the collinearity (vifcor)



presvals <- extract(predictors, fimbriolata_clean)
set.seed(0)
background <- randomPoints(predictors, 500)
absvals <- extract(predictors, background)
```

```
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))


pred_nf <- predictors


#Training and Testing sets


group <- kfold(fimbriolata_clean, 5)
pres_train <- fimbriolata_clean[group !=1, ]
pres_test <- fimbriolata[group ==1, ]


colnames(background) = c('lon', 'lat')
group <- kfold(background, 5)
background_train <- background[group !=1, ]
background_test <- background[group ==1, ]


# Within every algorithm I only created the first plot and not the second!
# The second plot is based on the presence and absence data and for the four-
# species there is only a small amount of present and nearly none absent data available.
# I still left the second plot included for possible further analyzes.



### Profile Methods using Bioclim algorithm ###


bc <- bioclim(pred_nf, pres_train)
plot(bc, a=1, b=2, p=0.85)


e <- evaluate(pres_test, background_test, bc, pred_nf)
e
tr <- threshold(e, 'spec_sens')
tr


pb <- predict(pred_nf, bc, progress='')
pb
```

```
par(mfrow=c(1,1))
plot(pb, main='Bioclim, raw values')
plot(wrld_simpl, add=TRUE, border='Dark grey')
plot(pb > tr, main='presence/absence')
plot(wrld_simpl, add=TRUE, border='Dark grey')
points(pres_train, pch='+')
```

Domain

```
dm <- domain(pred_nf, pres_train)
e <- evaluate(pres_test, background_test, dm, pred_nf)
e
pd = predict(pred_nf, dm, extent = extent, progress='')
par(mfrow=c(1,1))
plot(pd, main='Domain, raw values')
plot(wrld_simpl, add=TRUE, border='dark grey')
tr <- threshold(e, 'spec_sens')
plot(pd > tr, main='presence/absence')
plot(wrld_simpl, add=TRUE, border='dark grey')
points(pres_train, pch='+')
```

Regression models
```
train <- rbind(pres_train, background_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(background_train)))
envtrain <- extract(predictors, train)
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
head(envtrain)

testpres <- data.frame( extract(predictors, pres_test) )
testbackground <- data.frame( extract(predictors, background_test) )
```

Generalized Linear models

```
#At first select Future or Current Climate Data. Only then calculate pg.
```

#Current Climate Data

#The Evaluate functions calculate the AUC for every model.

```
 gm1 <- glm(pa ~ bio2 + bio3 + bio6 + bio8 + bio9 + bio12 + bio13 + bio14 + bio15 + bio18 + bio19 +
        croplands + pasture + sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation,
        family = binomial(link ="logit"), data = envtrain)
summary(gm1)
coef(gm1)
evaluate(testpres, testbackground, gm1)
ge1 <- evaluate(testpres, testbackground, gm1)


gm2 <- glm(pa ~ bio2 + bio3 + bio6 + bio8 + bio9 + bio12 + bio13 + bio14 + bio15 + bio18 + bio19 +
        croplands + pasture + sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation,
        family = gaussian(link ="identity"), data = envtrain)
evaluate(testpres, testbackground, gm2)
ge2 <- evaluate(testpres, testbackground, gm2)
ge2


gm3 <- glm(pa ~ bio2 + bio3 + bio6 + bio8 + bio9 + bio12 + bio13 + bio14 + bio15 + bio18 + bio19 +
        croplands + pasture + sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation,
        family = poisson(link ="log"), data = envtrain)
evaluate(testpres, testbackground, gm3)
ge3 <- evaluate(testpres, testbackground, gm3)
ge3



# Future Climate Data

#rcp26
gm1 <- glm(pa ~ rcp26bi02 + rcp26bi03 + rcp26bi05 + rcp26bi07 + rcp26bi08 + rcp26bi09 + rcp26bi12
        + rcp26bi13 + rcp26bi14 + rcp26bi15 + rcp26bi18 + rcp26bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = binomial(link
        ="logit"), data = envtrain)
summary(gm1)
```

```
coef(gm1)

evaluate(testpres, testbackground, gm1)

ge1 <- evaluate(testpres, testbackground, gm1)


gm2 <- glm(pa ~ rcp26bi02 + rcp26bi03 + rcp26bi05 + rcp26bi07 + rcp26bi08 + rcp26bi09 + rcp26bi12
        + rcp26bi13 + rcp26bi14 + rcp26bi15 + rcp26bi18 + rcp26bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = gaussian(link
        ="identity"), data = envtrain)

evaluate(testpres, testbackground, gm2)

ge2 <- evaluate(testpres, testbackground, gm2)

ge2


gm3 <- glm(pa ~ rcp26bi02 + rcp26bi03 + rcp26bi05 + rcp26bi07 + rcp26bi08 + rcp26bi09 + rcp26bi12
        + rcp26bi13 + rcp26bi14 + rcp26bi15 + rcp26bi18 + rcp26bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = poisson(link
        ="log"), data = envtrain)

evaluate(testpres, testbackground, gm3)

ge3 <- evaluate(testpres, testbackground, gm3)

ge3



#rcp45

gm1 <- glm(pa ~ rcp45bi02 + rcp45bi03 + rcp45bi05 + rcp45bi08 + rcp45bi09 + rcp45bi12 + rcp45bi13
        + rcp45bi14 + rcp45bi15 + rcp45bi18 + rcp45bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = binomial(link
        ="logit"), data = envtrain)

summary(gm1)

coef(gm1)

evaluate(testpres, testbackground, gm1)

ge1 <- evaluate(testpres, testbackground, gm1)


gm2 <- glm(pa ~ rcp45bi02 + rcp45bi03 + rcp45bi05 + rcp45bi08 + rcp45bi09 + rcp45bi12 + rcp45bi13
        + rcp45bi14 + rcp45bi15 + rcp45bi18 + rcp45bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = gaussian(link
        ="identity"), data = envtrain)
```

```
evaluate(testpres, testbackground, gm2)

ge2 <- evaluate(testpres, testbackground, gm2)

ge2


gm3 <- glm(pa ~ rcp45bi02 + rcp45bi03 + rcp45bi05 + rcp45bi08 + rcp45bi09 + rcp45bi12 + rcp45bi13
        + rcp45bi14 + rcp45bi15 + rcp45bi18 + rcp45bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = poisson(link
        ="log"), data = envtrain)

evaluate(testpres, testbackground, gm3)

ge3 <- evaluate(testpres, testbackground, gm3)

ge3



#rcp60

gm1 <- glm(pa ~ rcp60bi02 + rcp60bi03 + rcp60bi05 + rcp60bi07 + rcp60bi08 + rcp60bi09 + rcp60bi12
        + rcp60bi13 + rcp60bi14 + rcp60bi15 + rcp60bi18 + rcp60bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = binomial(link
        ="logit"), data = envtrain)

summary(gm1)

coef(gm1)

evaluate(testpres, testbackground, gm1)

ge1 <- evaluate(testpres, testbackground, gm1)


gm2 <- glm(pa ~ rcp60bi02 + rcp60bi03 + rcp60bi05 + rcp60bi07 + rcp60bi08 + rcp60bi09 + rcp60bi12
        + rcp60bi13 + rcp60bi14 + rcp60bi15 + rcp60bi18 + rcp60bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = gaussian(link
        ="identity"), data = envtrain)

evaluate(testpres, testbackground, gm2)

ge2 <- evaluate(testpres, testbackground, gm2)

ge2


gm3 <- glm(pa ~ rcp60bi02 + rcp60bi03 + rcp60bi05 + rcp60bi07 + rcp60bi08 + rcp60bi09 + rcp60bi12
        + rcp60bi13 + rcp60bi14 + rcp60bi15 + rcp60bi18 + rcp60bi19 + croplands + pasture +
        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = poisson(link
        ="log"), data = envtrain)
```

```
evaluate(testpres, testbackground, gm3)

ge3 <- evaluate(testpres, testbackground, gm3)

ge3



#rcp85

gm1 <- glm(pa ~ rcp85bi02 + rcp85bi03 + rcp85bi04 + rcp85bi05 + rcp85bi08 + rcp85bi09 + rcp85bi12

        + rcp85bi13 + rcp85bi14 + rcp85bi15 + rcp85bi18 + rcp85bi19 + croplands + pasture +

        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = binomial(link

        ="logit"), data = envtrain)

summary(gm1)

coef(gm1)

evaluate(testpres, testbackground, gm1)

ge1 <- evaluate(testpres, testbackground, gm1)


gm2 <- glm(pa ~ rcp85bi02 + rcp85bi03 + rcp85bi04 + rcp85bi05 + rcp85bi08 + rcp85bi09 + rcp85bi12

        + rcp85bi13 + rcp85bi14 + rcp85bi15 + rcp85bi18 + rcp85bi19 + croplands + pasture +

        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = gaussian(link

        ="identity"), data = envtrain)

evaluate(testpres, testbackground, gm2)

ge2 <- evaluate(testpres, testbackground, gm2)

ge2


gm3 <- glm(pa ~ rcp85bi02 + rcp85bi03 + rcp85bi04 + rcp85bi05 + rcp85bi08 + rcp85bi09 + rcp85bi12

        + rcp85bi13 + rcp85bi14 + rcp85bi15 + rcp85bi18 + rcp85bi19 + croplands + pasture +

        sugarcane_physical_area_SPAM + Globcover_cut + NE_elevation, family = poisson(link

        ="log"), data = envtrain)

evaluate(testpres, testbackground, gm3)

ge3 <- evaluate(testpres, testbackground, gm3)

ge3


pg1 <- predict(predictors, gm1, extent=extent)

pg2 <- predict(predictors, gm2, extent=extent)

pg3 <- predict(predictors, gm3, extent=extent)
```

```r
#GLM Binomial
par(mfrow=c(1,1))
plot(pg1, main='GLM/Binomial, raw values')
plot(wrld_simpl, add=TRUE, border='dark grey')
tr<-threshold(ge1, 'spec_sens')
plot(pg1 > tr, main='presence/absence')
plot(wrld_simpl, add=TRUE, border='dark grey')
points(pres_train, pch='+')
points(background_train, pch='-, cex=0.25')


#GLM Gaussian
par(mfrow=c(1,1))
plot(pg2, main='GLM/Gaussian, raw values')
plot(wrld_simpl, add=TRUE, border='dark grey')
tr<-threshold(ge2, 'spec_sens')
plot(pg2 > tr, main='presence/absence')
plot(wrld_simpl, add=TRUE, border='dark grey')
points(pres_train, pch='+')
points(background_train, pch='-, cex=0.25')


#GLM Poisson
par(mfrow=c(1,1))
plot(pg3, main='GLM/Poisson, raw values')
plot(wrld_simpl, add=TRUE, border='dark grey')
tr<-threshold(ge3, 'spec_sens')
plot(pg3 > tr, main='presence/absence')
plot(wrld_simpl, add=TRUE, border='dark grey')
points(pres_train, pch='+')
points(background_train, pch='-, cex=0.25')


##### MAXENT #####


#Put the file maxent.jar into the java folder in the dismo directory


system.file("java", package="dismo")
```

```r
#Checking if the jar file is present
jar <- paste(system.file(package="dismo"), "/java/maxent.jar", sep='')
if (file.exists(jar)) {
  xm <- maxent(predictors, pres_train)
  plot(xm)
} else {
  cat('cannot run this example because Maxent is not available')
  plot(1)
}


# The 64bit Java Version is necessary!!


if (file.exists(jar)) {
  response(xm)
}else{
  cat('Cannot run this example because maxent is not available')
  plot(1)
}



if (file.exists(jar)) {
  e <- evaluate(pres_test, background_test, xm, predictors)
  e
  px <- predict(predictors, xm, extend=extend, progress='')
  par(mfrow=c(1,1))
  plot(px, main= substitute(paste(italic('Mahanarva fimbriolata'))))
  plot(wrld_simpl, add=TRUE, border='dark grey')
  tr <- threshold(e,'spec_sens')
  plot(px > tr, main='presence/absence')
  plot(wrld_simpl, add=TRUE, border='dark grey')
  points(pres_train, pch='+')
}else{
  plot(1)
}
```

summary(xm)

#html data with further information

xm

Biosketch

Christian Schöbel is master student of the Zoology Pos-graduate Program at Pontifícia Universidade Católica do Rio Grande do Sul (Brazil) on the distribution and biodiversity of spittlebugs.

Gervasio Silva Carvalho is a professor at Pontifícia Universidade Católica do Rio Grande do Sul (Brazil) with expertise in Cercopidae taxonomy.