Report on the Species Distribution Model
The case of the *Zea mays*

Eva Lafay – **s2673479** – 2019-2020

Introduction

The Maize (*Zea mays)* is grown throughout the world and is the third most traded cereal (www.fao.org and Ranum, P., *et al*., 2014). Whereas there are a lot of morphological and genetical diversity among the maize landraces (Matsuoka, Y., *et al.* 2002), there is only one single domestication event. It took place 9.000 years ago in Mexico, when humans were selecting the harvestable grain-producing inflorescences (Benz, B.F., 2001). This cereal has nowadays different uses: mainly for animal feeding but is also used in human alimentation and has agro-industrial purposes (Ranum, P., *et al*., 2014). Nevertheless, in the context of climate change, agriculture is very vulnerable: the increasing of temperatures might induce a decline of the productions in the next coming years (Nabout, J.C., *et al.* 2012 and Ureta, C., *et al.* 2012). Maize is notably sensitive to temperature and to precipitation (Nabout, J.C., *et al.* 2012), thus a disturbance in the climate may lead to economic losses. Therefore, using the Species Distribution Model of the *Zea mays* would allow us a better understanding of the impacts of the human on these crops in terms of geographical distribution shifts and productivity scenario for the future, and could help in elaborating a management plan for this crop.

We hence want to investigate whether the potential distribution of the Maize will be transformed in the near future in the scenario of a Representative Concentration Pathway (RCP) of 8.5. We predict a shifting and a decline of the potential distribution areas for the year 2050.

Methodology and Results

- Overview:

The flowchart below (Figure 1) shows the steps followed to answer our research question.

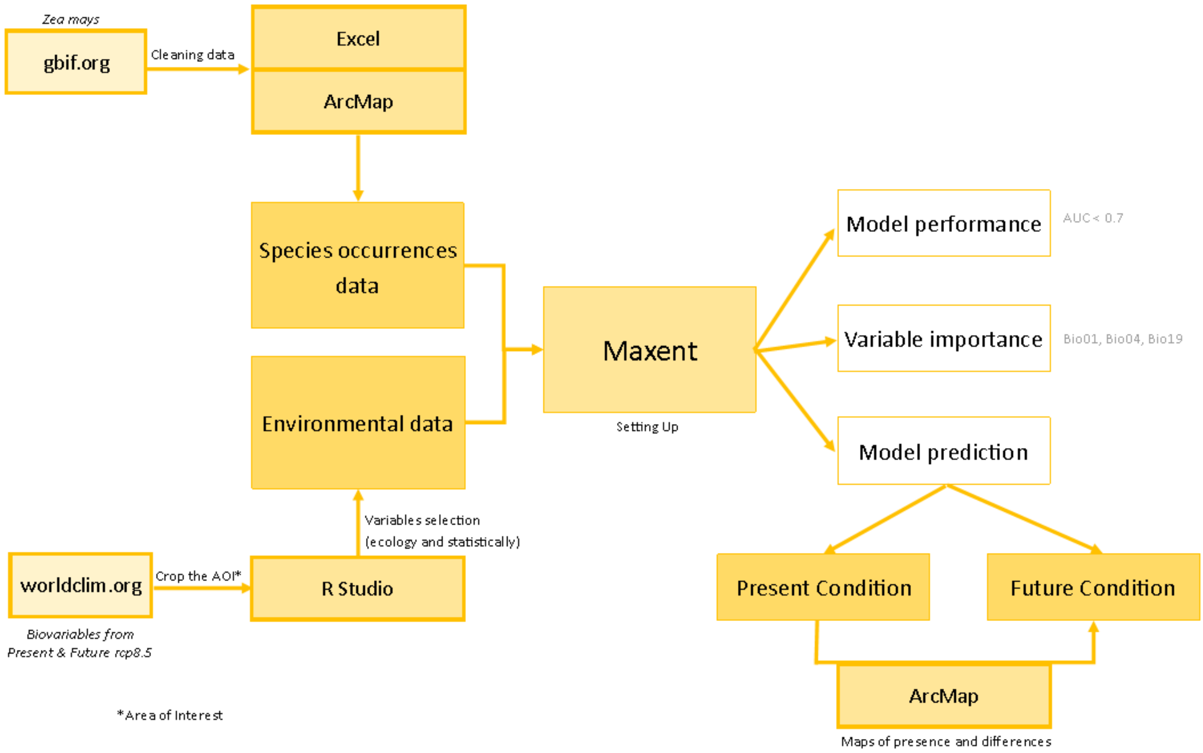

Figure 1 - Flowchart with the steps performed to test the research hypothesis

- Occurrence data

We download the data of the occurrence of *Zea mays* from gbif.com. We then open the file in Excel to conserve only the species name, the longitude and latitude and delete the other columns. We observe a huge amount of missing, duplicates or incorrect data that we erase (letters in the spatial coordinates for instance). We subsequently investigate spatial biases through ArcMap: all points situated in the water are removed; we also delete the points when there is only one occurrence in a country, because it doesn't indicate correctly the environment where the individual is found. We obtain at the end a global distribution of our species across the world that reflects properly the species ecological niche.

- Environmental data

We download all the Biovariables through the worldclim.org website. According to Nabout (2012), the important variables for the maize in terms of ecology are temperature and precipitation: Bio01, Bio04, Bio06, Bio19 they are respectively the annual Mean Temperature, the Temperature Seasonality, the Minimum Temperature of the Coldest Month, and the Precipitation of Coldest Quarter. Besides, we test the correlations of those with statistical tests in R Studio. For the autocorrelation we base us on the Pearson's pairwise correlation r, if $|r|<0.7$ we should exclude variables (see red color in Figure 2). For the multicollinearity we look at the Variance Inflaction Factor (VIF), if $VIF<10$ we can accept those variables (Figure 3). Removing the Bio06 leads us to obtain better results (Figure 4), hence we only keep Bio01, Bio04 and Bio19 for our research.

	Bio01	Bio04	Bio06	Bio19
Bio01	1	-0,8634503	0,96961966	0,24488965
Bio04	-0,8634503	1	-0,9488659	-0,3871551
Bio06	0,96961966	-0,9488659	1	0,35457911
Bio19	0,24488965	-0,3871551	0,35457911	1

Figure 2 - Pearson's pairwise correlation r between biovariables

```
vif(df.stk.AOI, maxobservations=nrow(df.stk.AOI))
Variables        VIF
    Bio01   45.645296
    Bio04   22.959415
    Bio06 112.460577
    Bio19    1.490364
```

Figure 3 - Variance Inflaction Factor (VIF) with Bio01, Bio04, Bio06 and Bio19 are almost all above 10.

```
vif(df.stk.AOI, maxobservations=nrow(df.stk.AOI))
Variables      VIF
    Bio01 4.080764
    Bio04 4.512396
    Bio19 1.221447
```

Figure 4 - Variance Inflaction Factor (VIF,) without Bio06, become all lower than 10.

- Model settings

The occurrence of our mays, cleaning during the previous steps, is put in our sample: 10904 presence records used for training and 2726 for testing. The environmental layers used are Bio01, Bio04 and Bio19, all continuous. The settings chosen for the model are the creation of 5 replicates of maxent because we want to train the model and report the average results. We select cross-validation for the run type of these replicates in order to estimate how accurately our predictive model will perform in practice. We choose to extrapolate because we want to predict the regions of environmental space outside the limits encountered during the training and we clamp because we want to restrict our variables to range of values used for training. We choose the responsecurves, the jackknife and randomseed to be true. The projection layers taken are the Present WLD, referring to the uncropped environmental variables, and the Future WLD and AOI (Area of Interest, cropped previously with R) for the scenario rcp8.5 to allow us to compare the distribution between now -2019- and 2050. For more information about the settings, see Figure 5.

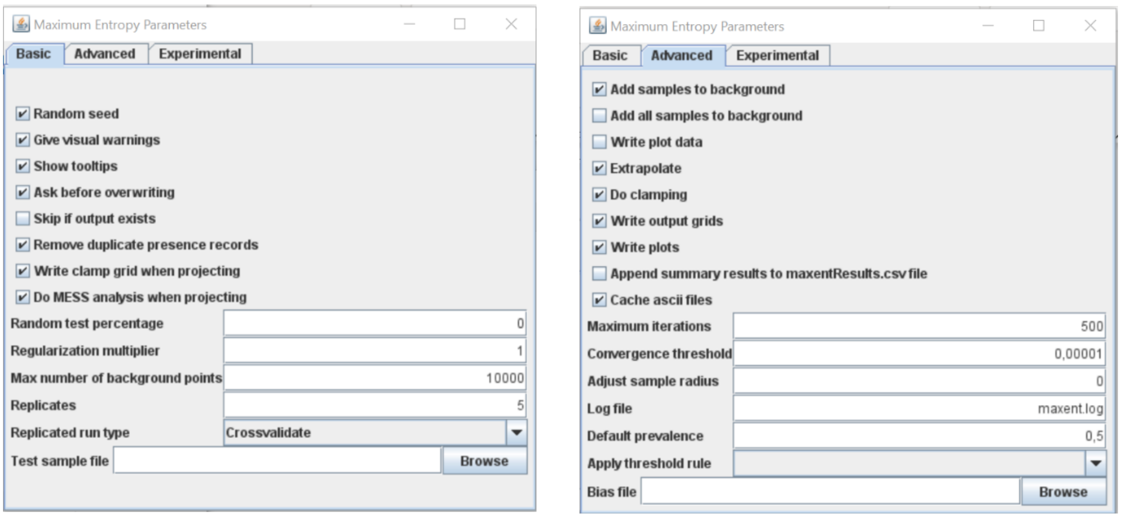

Figure 5 - Settings of Maxent

- Model output
    - Performance

The Figure 6 shows the area of predicted probability of *Zea mays*: it decreases on test data with the rising of the cumulative threshold.

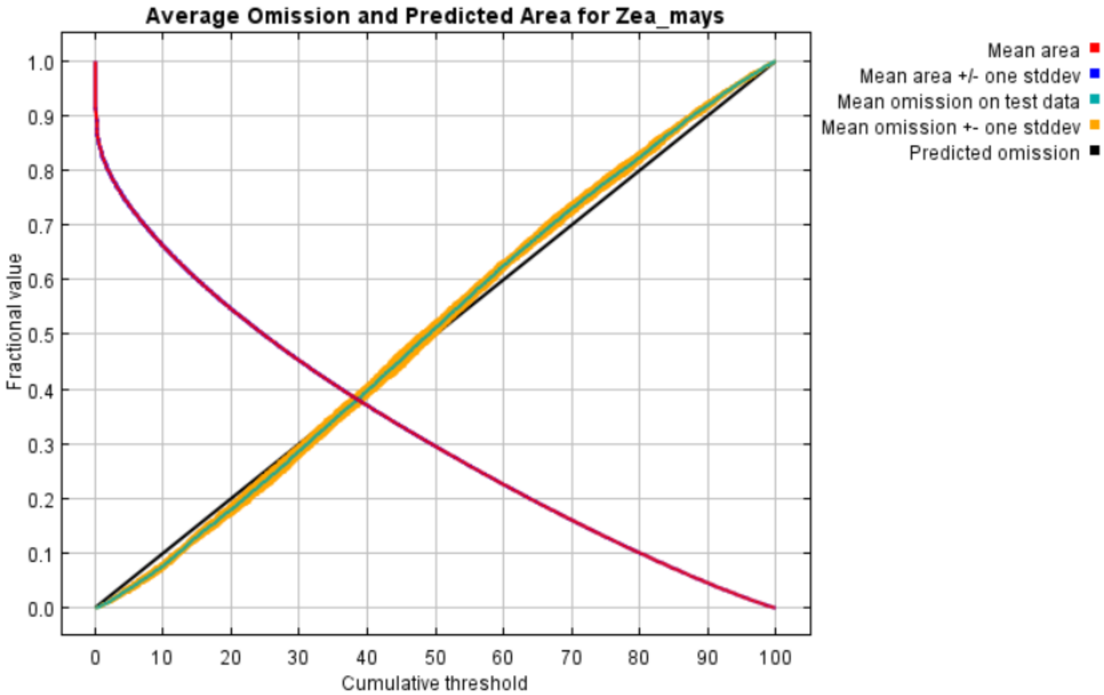

Figure 6 - Average Omission and Predicted Area for Zea mays

The **AUC** (Area under the receiver-operator curve) obtained is 0.674, that is lower than 0.7. It means that our prediction is not reliable and not accurate because close to a random prediction (Figure 7). The standard deviation is 0.005.

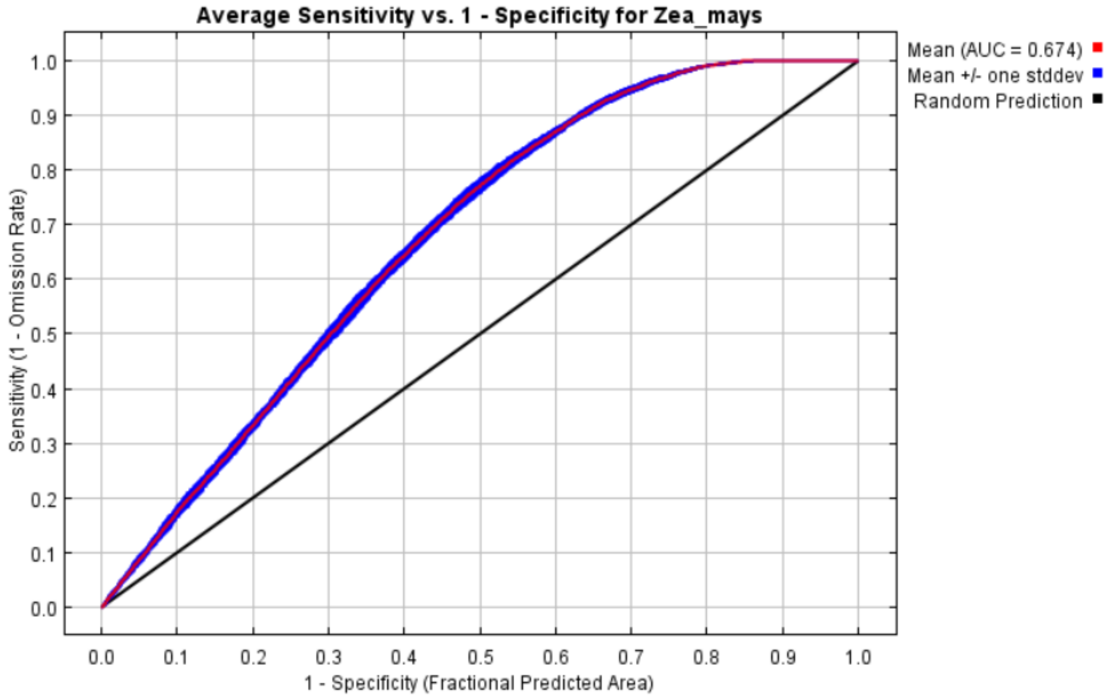

Figure 7 - ROC curve averaged over the replicate models

Other settings have been tried (with or without clamping, with or without extrapolation, with more replicates) the AUC remains lower than 0.7.

The **response curves** obtained tell us about the contribution of each variables to the Maxent prediction, either by keeping all the other environmental variables at their average sample value (above parts of the Figure 9) or by using only the corresponding variable (below parts of the Figure 9). The Y-axis shows the probability of having a suitable environment for your species. In both cases we have the similar curves, meaning that our variables are independent from each other. These curves allow us to know which environment is the most suitable for the mays: moderate temperature (Bio01), low temperature seasonality variation (Bio04). The precipitation of coldest quarter curve (Bio19) has a constant line, meaning that it isn't the most important biovariables to consider in habitat prediction.

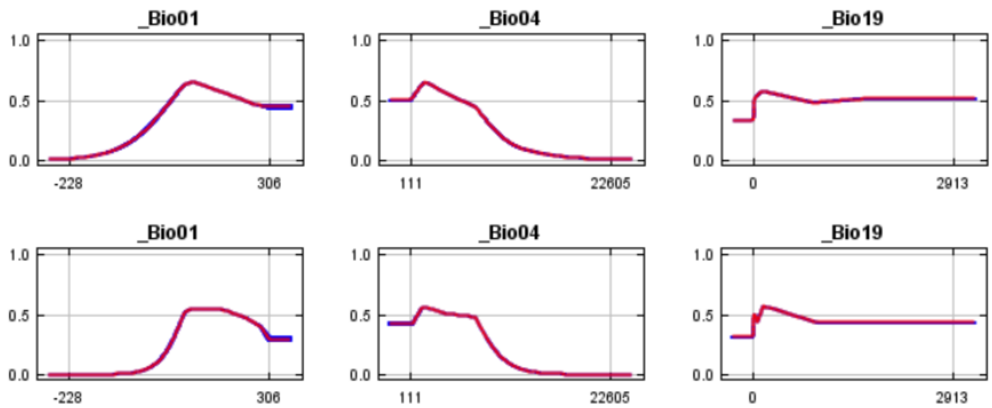

Figure 8 - Response curves of the Biovariables

Furthermore, we have access to the relative contributions of the environmental variables to Maxent model through this Table 1 below. The temperature seasonality (Bio04) contributes to 66.1%, while the Annual Mean Temperature contributes to 24.9% and Precipitation of the Coldest Quarter only contributes to 8.9% of the suitability of the environment for the *Zea mays*.

Table 1 - Variable contributions

Variable	Percent contribution	Permutation importance
_Bio04	66.1	56.9
_Bio01	24.9	30.7
_Bio19	8.9	12.4

Our entire model predictions map can't really be trusted since our AUC is very low.

- SDM projections

To create our maps, we use ArcMap with the outputs of Maxent. We base us on the threshold below (see Table 2) by using the Maximum Training Sensitivity plus Specificity, which is 0.441 to create binary maps of presence and absence of mays.

Table 2 - Threshold table

Logistic threshold	Description
0.171	Fixed cumulative value 1
0.314	Fixed cumulative value 5
0.373	Fixed cumulative value 10
0.002	Minimum training presence
0.389	10 percentile training presence
0.502	Equal training sensitivity and specificity
0.434	Maximum training sensitivity plus specificity
0.499	Equal test sensitivity and specificity
0.441	Maximum test sensitivity plus specificity
0.120	Balance training omission, predicted area and threshold value
0.241	Equate entropy of thresholded and original distributions

- Interpret the results of the following scenarios:

  - **Present world data**

The *Zea mays* has, in the present, a wide distribution throughout the world (Figure 9). Africa, South America, South Asia, Oceania and Western Europe are suitable habitats for the mays.

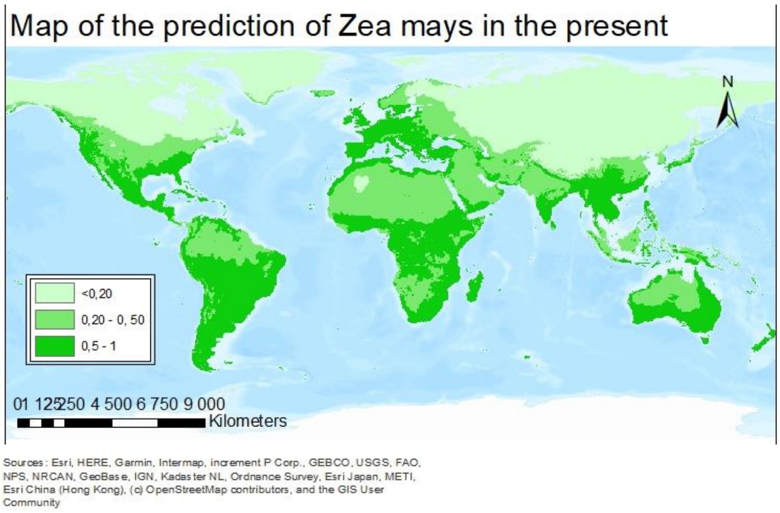

Figure 9 - Map of the prediction of Zea mays *in the present*

  - **Future scenario - rcp85**

In the scenario of rcp8.5, in 2050, the distribution of the maize appears to be quite similar to the present, but with less high-quality habitats in Africa, and more potential distribution areas in East Europe (Figure 10).

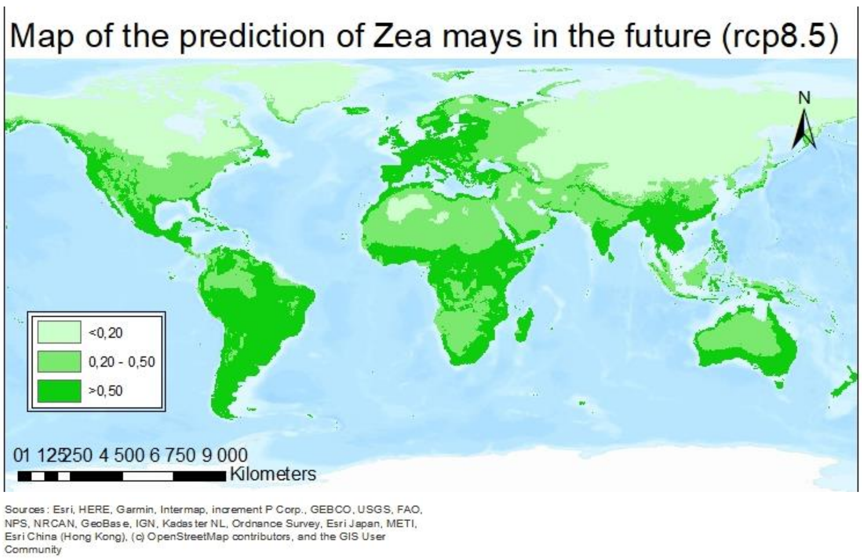

Figure 10 - Map of the prediction of Zea mays in the future, scenario rcp8.5

- Did the geographical space available for the species increased or decreased?

To see if the geographical space available for the species has increased or decreased, we make a map of the differences of presence and absent in present and future conditions (Figure 11). We observe more decrease areas than increase areas. The decreases occur mostly in the South part of the Earth, while the increase areas are present in the North countries.

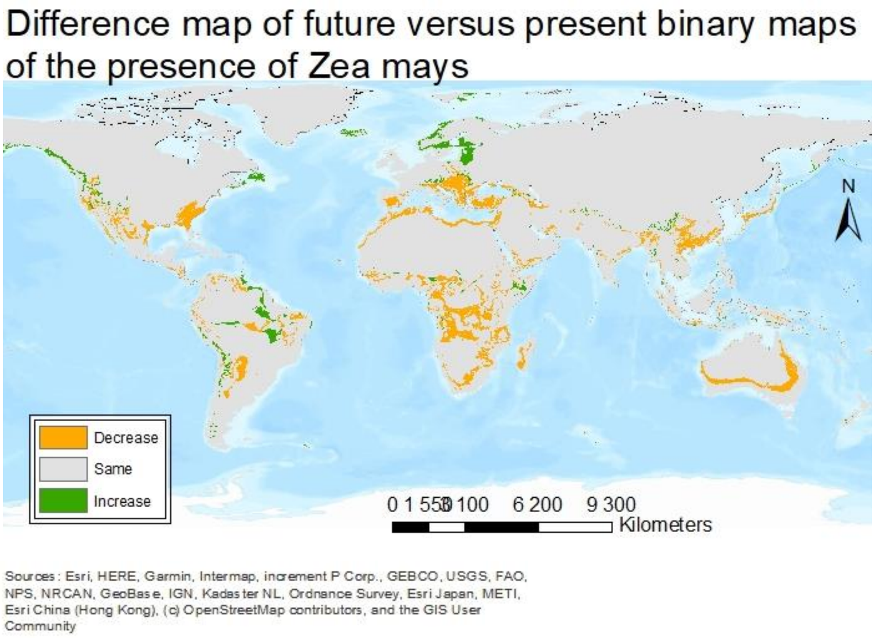

Figure 11 - Difference map of future vs present binary maps of the presence of Zea mays

The species could reach these new locations on its own because there are no islands constraints, and their present location are near the potential new ones. But it is more probable that the farmers plant them there if the environmental conditions are indeed changing this way.

Discussion and Conclusion

We have investigated the potential distribution of the Maize in the present compared to the future with the rcp8.5 scenario. The maps we obtained have confirmed our hypothesis: there is a shifting and a decline of the potential distribution areas for the year 2050 compared to now. There will be less suitable habitats for the *Zea mays* in the southern hemisphere of the globe, but some new potential areas of distribution, notably in northern Europa. That could imply that the farmers would lose their job in the south countries and/ or should move further north to still be able to work in mays field. We have also seen that the most important environmental variable in the presence of the maize is the temperature seasonality. Nevertheless, all the outputs given here are not totally reliable (the AUC is closed to a random prediction). To improve this model, we should pursue the research; by cleaning even more drastically the occurrences data for instance or looking at more specific cultivar from maize, in order to avoid the large spectrum of environmental needs of the species *Zea mays*, and taken into account the genetic diversity (Ureta, C., *et al.* 2012). We could also differentiate domesticated and wild *Zea mais*. Moreover, in our current predictions we have only considered climate changes, but one other main driver of biodiversity loss is the habitat change. Hence to improve our predictions for the future we should also adding this to our models.

Bibliography

Benz, B. F. (2001). Archaeological evidence of teosinte domestication from Guilá Naquitz, Oaxaca. Proceedings of the National Academy of Sciences, 98(4), 2104-2106.

GBIF.org (3 December 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.dkvw0n

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J., Buckler, E., & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences, 99(9), 6080-6084.

Nabout, J. C., Caetano, J. M., Ferreira, R. B., Teixeira, I. R., & Alves, S. D. F. (2012). Using correlative, mechanistic and hybrid niche models to predict the productivity and impact of global climate change on maize crop in Brazil. *Natureza & Conservacao*, *10*, 177-183.

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. Annals of the New York Academy of Sciences, 1312(1), 105-112.

Ureta, C., Martínez-Meyer, E., Perales, H. R., & Álvarez-Buylla, E. R. (2012). Projecting the effects of climate change on the distribution of maize races and their wild relatives in Mexico. Global Change Biology, 18(3), 1073-1082.

Zea mays L. in GBIF Secretariat (2019). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2019-12-10.