Bob Bastiaans

S1547127

# Species Distribution Models (SDM)

## Introduction

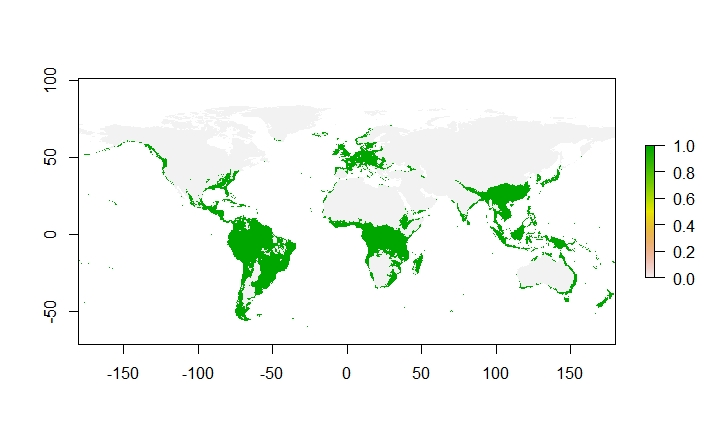
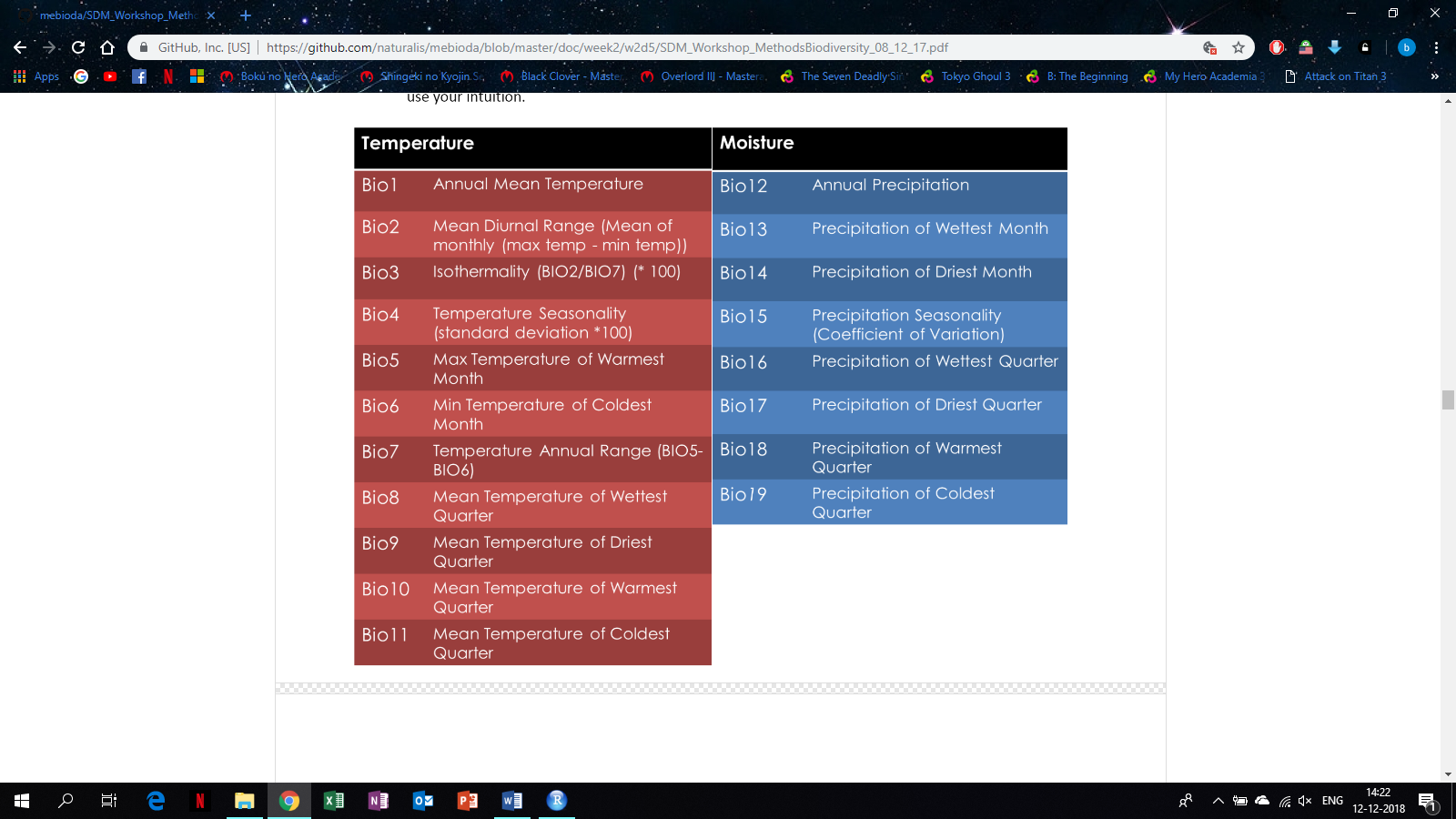
This report gives an insight about the cultivated root crop *Ipomoea batatas,* commonly known as the sweet potato. The *I. batatas* is globally cultivated but finds its origin in Central and South America (fig. 1), where it is still cultivated more than anywhere else in the world. The data used in this report was obtained from the Global Biodiversity Information Facility (GBIF). Bioclimatic variables were obtained from WorldClim. The SDMs were made in R and Maxent respectively.

Figure : Ipomoea batatas distribution at occurrence scale with integrated threshold (maximum training sensitivity and specificity threshold=0.319).

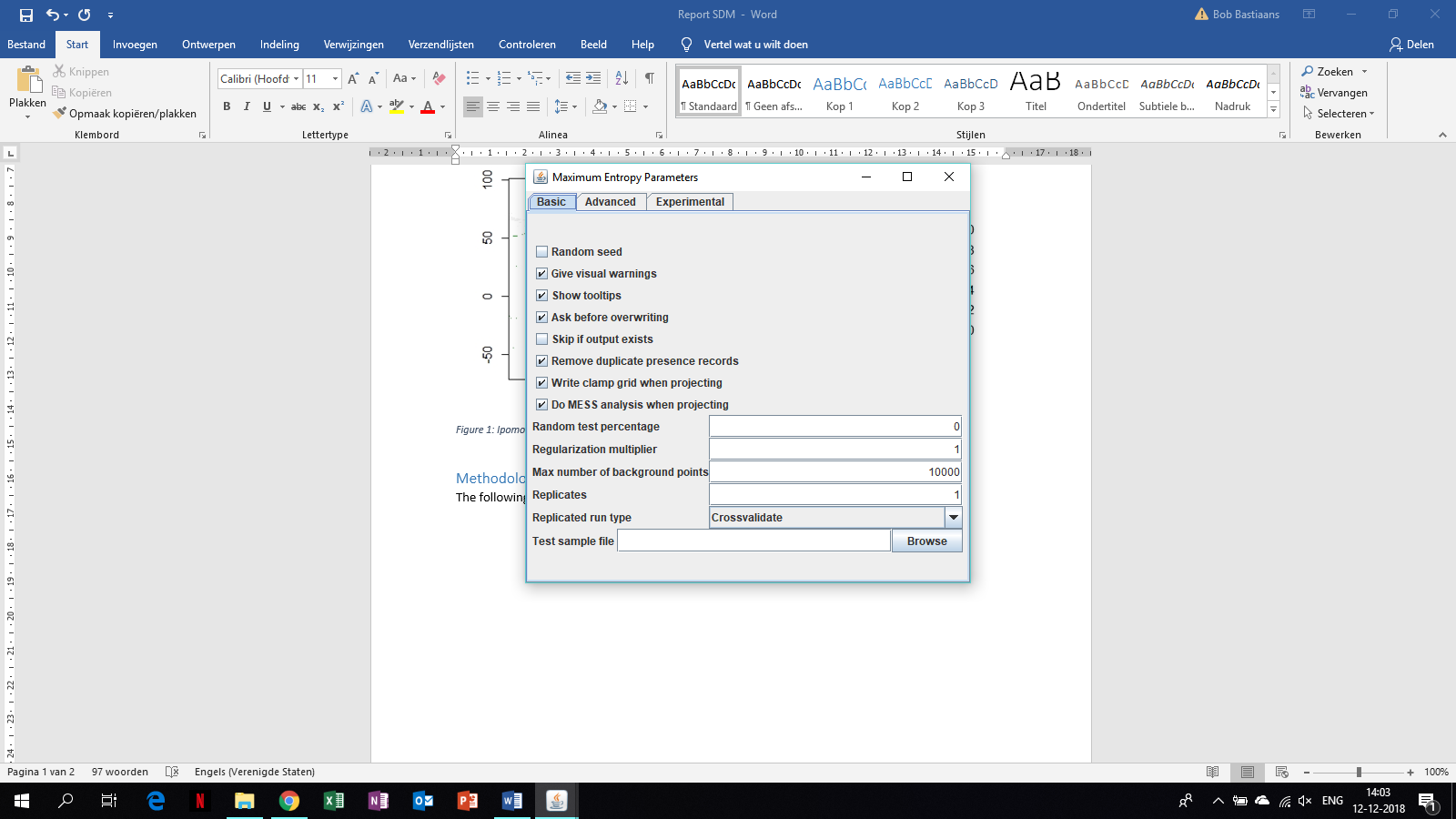
## Methodology

The following settings were used in MaxEnt (fig. 2). Remove duplicate presence records was checked to avoid pseudo replication and build a good model. The variables used in this model were: Bio2, Bio7, Bio8, Bio9, Bio18, Bio19 (Table I). These variables were chosen because of their lack of correlation and are therefore statistically strong. All variables obtained with a 5-minute resolution.

Tabel : Bioclimatic variables. Parameters are chosen based on their lack of correlation with each other (correlation<0.7). Bio2, 7 ,8, 9, 18 & 19 were used in this model.



Bio2 was used because it is the mean temperature per day. *I. batatas is* originally cultivated under tropical conditions, hence much data on temperature is needed. As is described, not only the mean temperature per day but also the variation in temperature per year (Bio7) is needed to establish the most suitable conditions for *I. batatas* in the present and future. Bio8, Bio9, Bio18 and Bio19 are used because these variables could be of most influence on the seed quality and biomass production of the *I. batatas*.



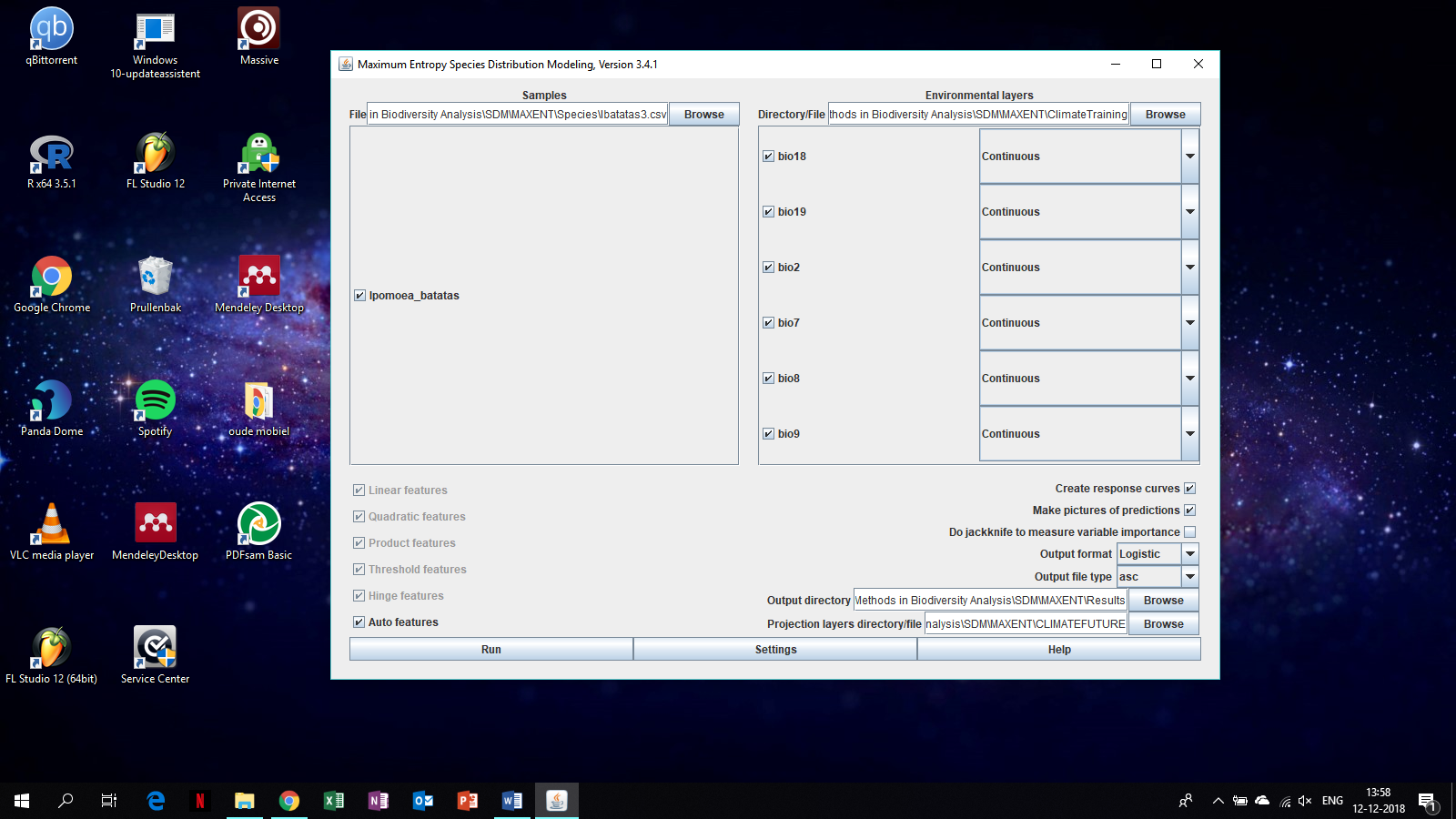


Figure : Settings used in MaxEnt.

Figure : Bioclimatic Variables used in MaxEnt model of I. batatas

## Model Output

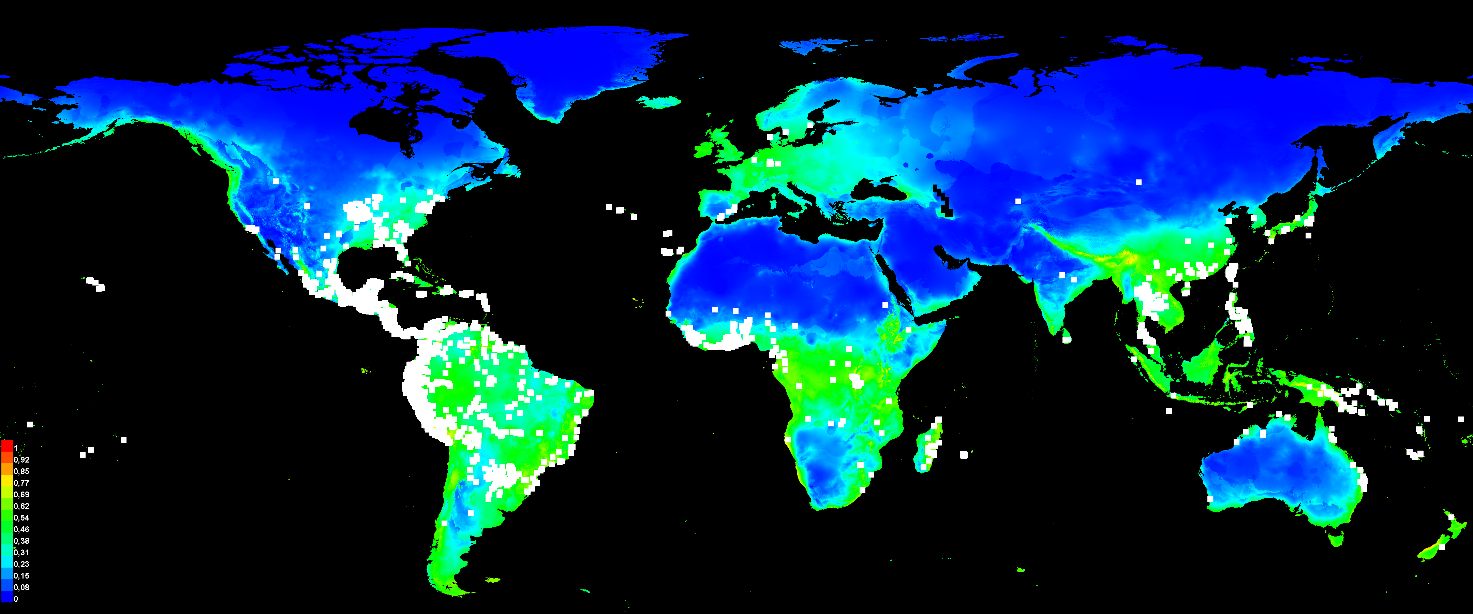
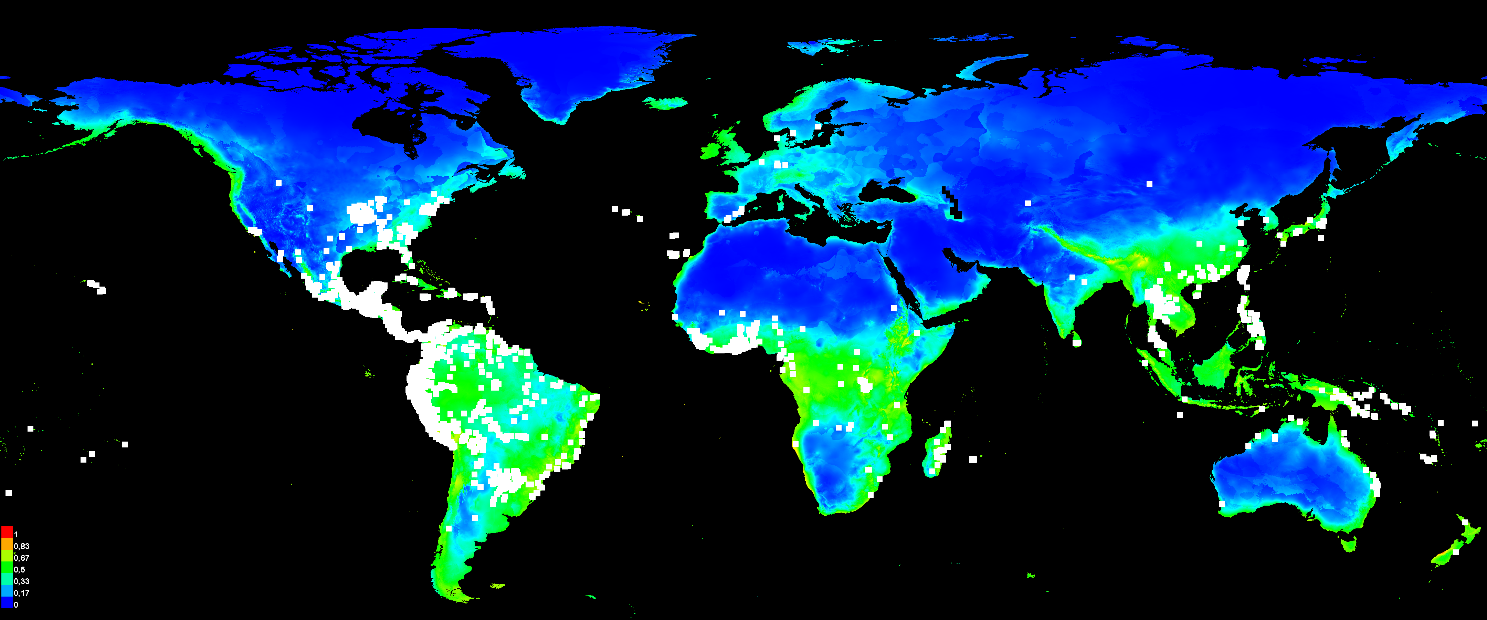
The SDM in fig. 1 is now combined in MaxEnt with the previously described bioclimatic variables, respectively 2, 7, 8, 9, 18, 19, resulting in a model that shows how favorable and suitable the conditions globally are for the *I. batatas* (fig. 4). A prediction of the future distribution of the *I. batatas* also has been made for RCP4.5 (fig. 5).

Figure 5: Future habitat suitability at occurrence scale of I. batatas. legend: see figure 4. distribution is modelled following the RCP4.5 scenario.

Figure : Present habitat suitability at occurrence scale for I. batatas. Legend: The warmer the colors, the more suitable the habitat. whit dots are test samples used to build the model.

In order to establish the reliability of the models, a Receiver Operating Characteristic (ROC) curve is made with the same data that is used in fig. 4 & 5. This will give the Area Under the Curve (AUC) value (fig. 6). The standard sensitivity of AUC=0.7 (reliable) and AUC>0.8 (good) is used to determine the reliability of the model. Since the AUC=0.794, the model is reliable but nearly as good for future predictions. The importance of the 6 bioclimatic variables used to model the species distribution is also calculated (Table II). The annual range of temperature is concluded to be the most important contribution to the present and future SDM of *I. batatas.*

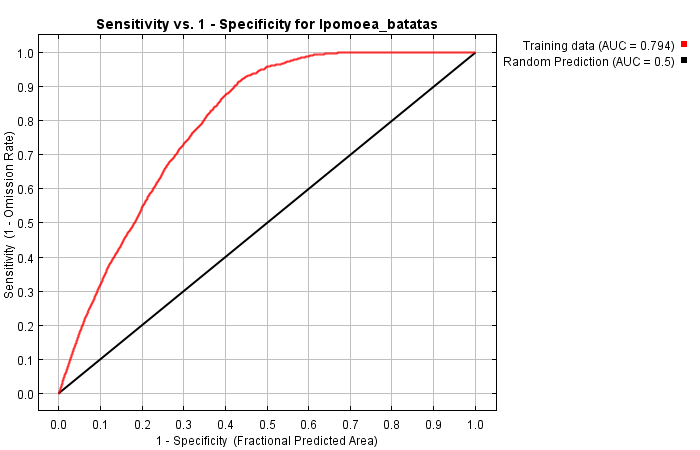
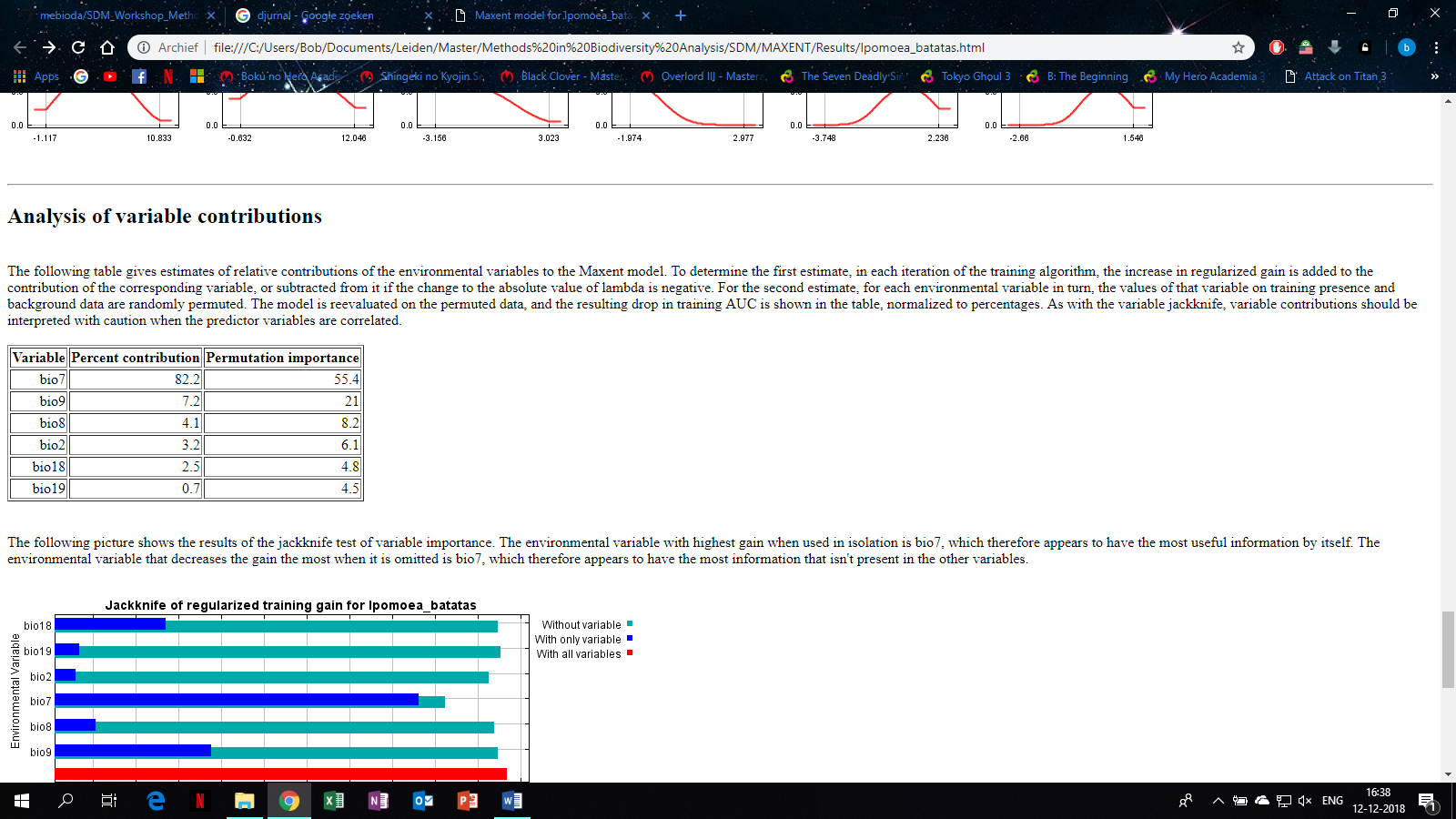


Figure 6: receiver operating characteristic (ROC) curve for the same data. This implies that the maximum achievable AUC is less than 1. AUC=0.794

Table II: Variable Importance in percentages of contribution for building the SDM. Permutation in descending order of importance for reliability of future SDM.



## Response to future scenario

In this section, the results from MaxEnt and occurrence data from GBIF are combined to build a map showing the change in suitable habitats between present and future climate conditions for *I. batatas* (fig. 7).

## 

Figure 7: Difference in present and future habitat suitability at occurrence scale of I. batatas. Legend: green=area gained, grey=never suitable, yellow=remains suitable, red=area lost, black dots=sampling locations of occurrence.

## Biological interpretation

The map shows the most profound losses of habitat at RCP4.5 are in Europe, at the east coast of North America and the middle of the Amazon rainforest. The most profound gained regions of habitat are in Alaska and in the tropical regions of Africa in between the equator and the tropic of cancer. Needless to say, the suitable habitat that is lost is far greater than the suitable habitat gained, but still the vast majority of future suitable habitat remains suitable. The loss of habitat in Europe doesn’t imply a major biological shift since this isn’t the natural habitat of *I. batatas.* The loss of habitat in the Amazon doesn’t have a major impact either because of the absence of intensive cultivation and agriculture of crops in this area. The loss of suitable habitat doesn’t have a major impact on the biology of this particular species, but the gained area could cause a shift in the biology. The *I. batatas* is a tropical species and therefore grows the best under tropical conditions. The RCP4.5 scenario allows tropical conditions to occur in formerly subtropical regions and coastal regions with a higher than average precipitation rate. This allows the *I. batatas* to distribute to coastal and subtropical regions as are shown on the map, or montane habitats (fig. 7). Nonetheless, one might argue if the model is that good of a representation for the future distribution of *I. batatas.* The AUC can be defined as nearly good, but that doesn’t account for the potential bias of the SDM. The obtained GBIF data only contained georeferenced data, which is just a portion of all the occurrence data on *I. batatas* in GBIF, and the data on GBIF is just only a small portion of all occurrence data (digitalized and non-digitalized) existing in science. The gained habitat area in Alaska is a good example of the possible errors the GBIF and WorldClim data. Additionally, only presences of the species are used in the SDM whereas pseudo-absences or absences are left out. Therefore, the AUC isn’t a reliable indication of this model. A null model could counteract this problem. Hence, the model is probably biased and not a reliable representation of the present and future SDM of *I. batatas*.

## References

GBIF.org (6th December 2018) GBIF Occurrence Download <https://doi.org/10.15468/dl.htadl3>

<http://www.worldclim.org/version1>

Steven J. Phillips, Miroslav Dudík, Robert E. Schapire. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: <http://biodiversityinformatics.amnh.org/open_source/maxent/>. Accessed on 2018-12-12.

R Version 3.5.1