**Species distribution modelling *Manihot esculenta* Crantz**

**Introduction**

The cassava (*Manihot esculenta* Crantz) crop is a woody shrub, native to South America. Nowadays it’s extensively cultivated all around the equator. Its edible roots commonly eaten in Afrika and South America. Cassava is native to the savannah grounds of brazil, Subsequently cassava is able to grow well on nutrient poor and dry habitats. It can be harvested all year round, unlike most crops. Hence cassava can compensate for food shortages in periods when other crops cannot be harvested. Since climate change will increase the number and frequency of extreme periods, cassava can become increasingly important for a stable food supply. The main purpose of the simulation is to see whether the suitable habitat of the cassava crop in Africa changes in the near future if the wold continues to warm the way it does now.

**Methodology**

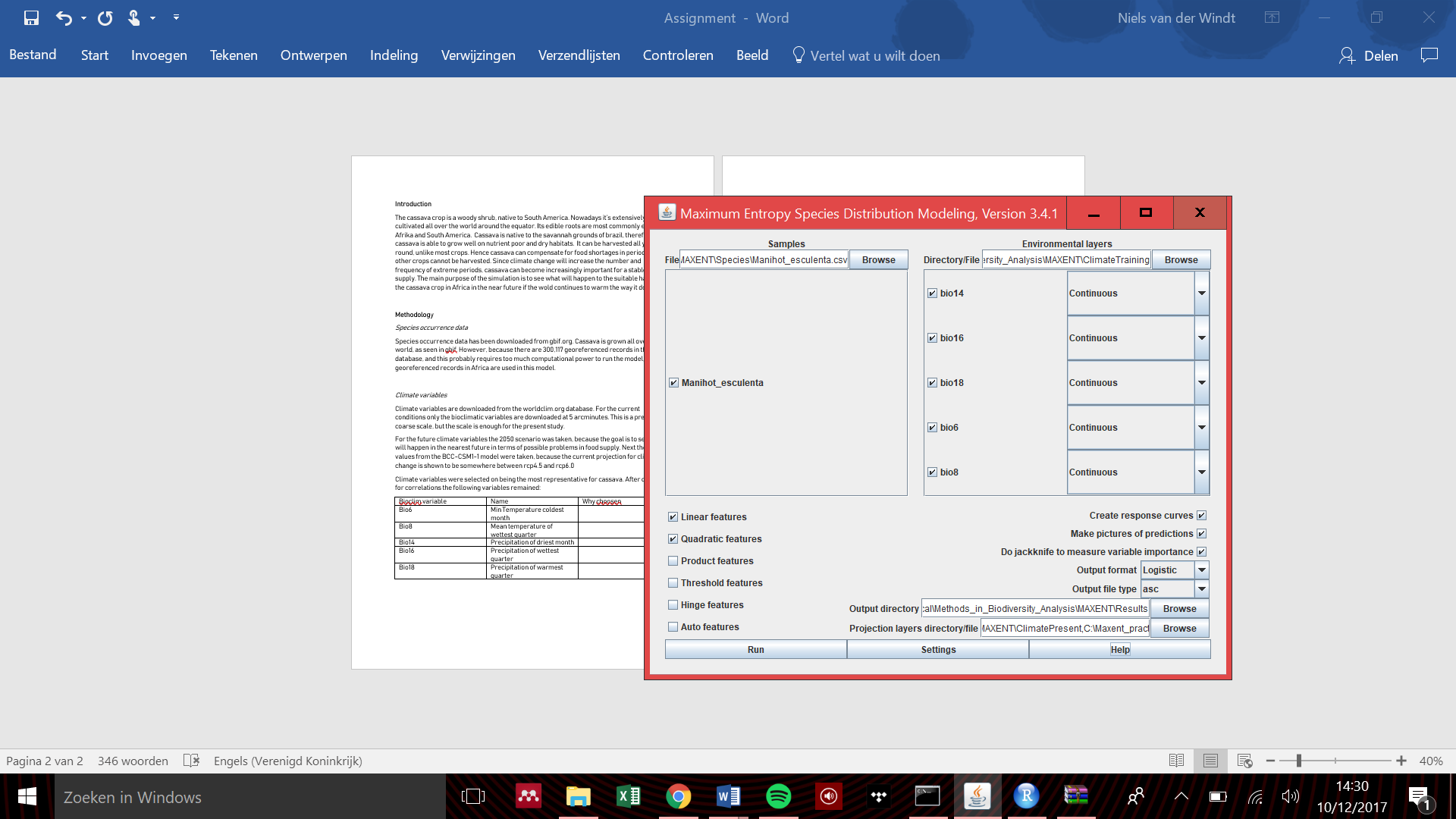
*Species occurrence data*

Species occurrence data has been downloaded from gbif.org. Although cassava is grown all over the world, as seen in GBIF, not all records have been included in the model. This was done because there are 300,117 georeferenced records in the GBIF database. Including all this data would probably require too much computational power to run the model. Therefore only the georeferenced records located in Africa were used in this model.

*Climate variables*

Climate variables were downloaded from the worldclim.org database. For the current conditions the bioclimatic variables were downloaded at 5 arcminutes. This is a pretty coarse scale, but the scale is enough for the present study.

For the future climate variables the 2050 scenario was taken, because the goal is to see what will happen in the near future in terms of possible problems in food supply. Thereafter the rcp6.0 values from the BCC-CSM1-1 model were taken, because the current projection for climate change is shown to be somewhere between rcp4.5 and rcp6.0. Using the 6.0 scenario would therefore give a representation of the current development with a lower risk of underestimating the effects.

Spatial climate variables we selected on the basis of correlations (| Spearman rho | > 0.7). Highly correlated variables were compared one to another and variables that were expected to have the least effect on the cassava, based on its ecology, were removed from the model. Because cassava is a tropical plant growing around the equator and it is relatively drought resistant, variables that included minimal temperatures and extreme precipitation events were considered most important. This includes the following variables: Bio6 (Min temperature coldest month), Bio8 (Mean temperature of wettest quarter), Bio14 (Precipitation of driest month), Bio16 (Precipitation of wettest quarter) and Bio18 (Precipitation of warmest quarter).

*Species distribution modelling*

The algorithm used for the creation of the model is MAXENT. The settings used in the model can be seen in figure 1. Only linear + quadratic features were used for calculations. This means the output distribution is constraint for the species to have the same expectation for each of the continuous climate variables as the sample locations for that species and the output distribution is to have the same expectation and variance of the climatic variables as the samples.

Figure 1: Model settings MAXENT

**Model output**

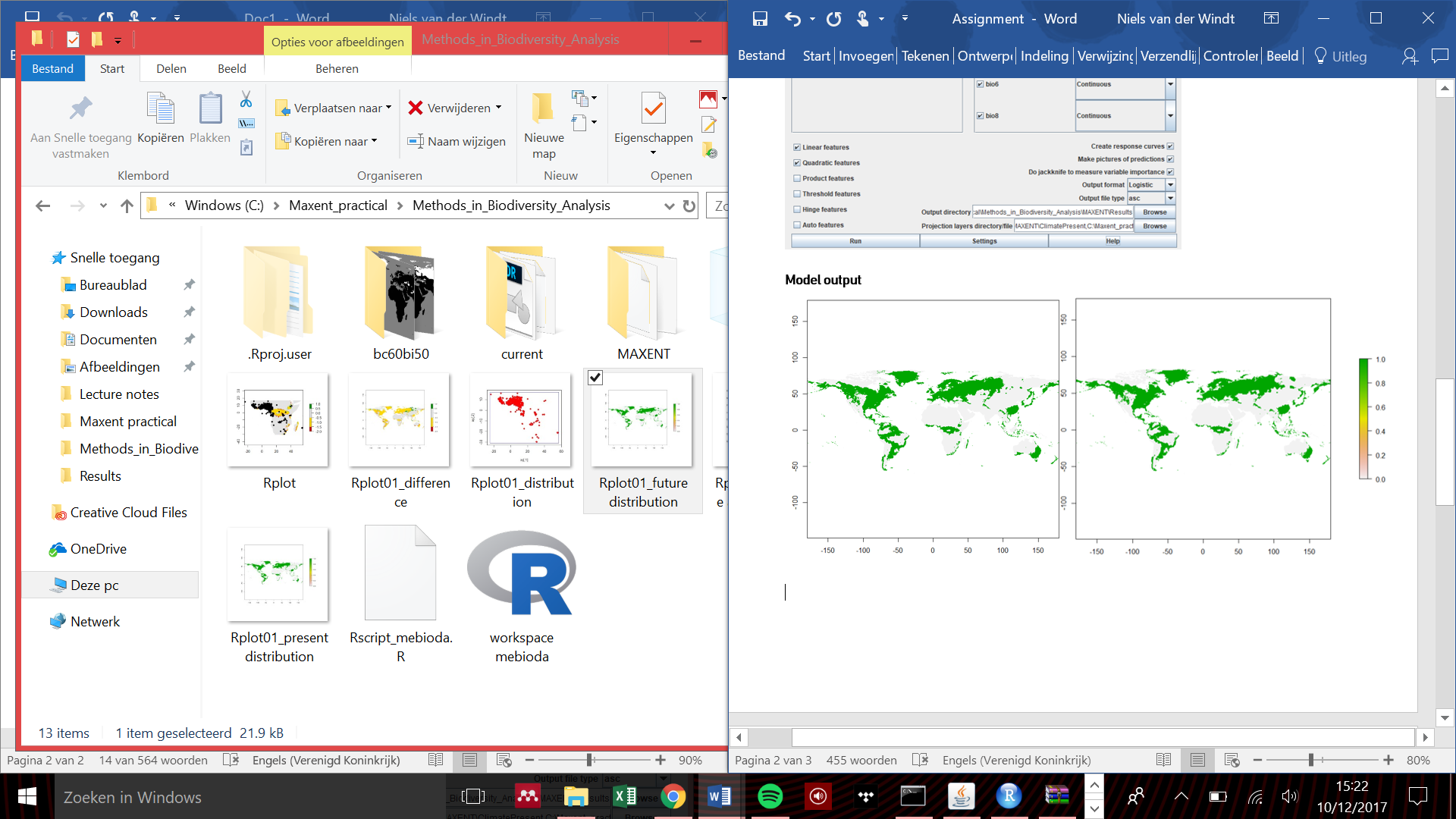
The figure below shows the present and future distribution maps for Manihot esculenta according to the model. Differences between the present and future seem to be minimal.

Figure 2: Present (l) and future (r) distribution map of Manihot esculenta

The AUC of the receiver operator curve of the map gives a value of 0.703. The model can therefore be considered as a useful model ( > 0.7). The maximum training sensitivity plus specificity gives a logistic threshold value of 0.444

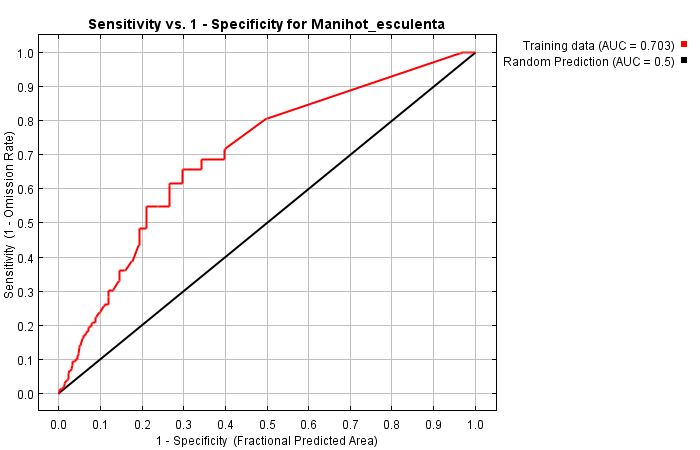


Figure 3: Receiver Operator Characteristic curve

The analysis of variable contributions show that bioclim variable 14 gives the largest contribution to the Maxent model (table 1). Therefore precipitation in the driest month explains the distribution shown in the map the best and therefore it is probably worth keeping an eye on this variable for the future.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Percent contribution** | **Permutation importance** |
| bio14 | 88 | 92.7 |
| bio16 | 8.7 | 5.3 |
| bio18 | 3.1 | 2 |
| bio6 | 0.2 | 0 |
| bio8 | 0 | 0 |

Table 1: Relative contributions of the environmental variables to the Maxent model.

**Response to future scenario**

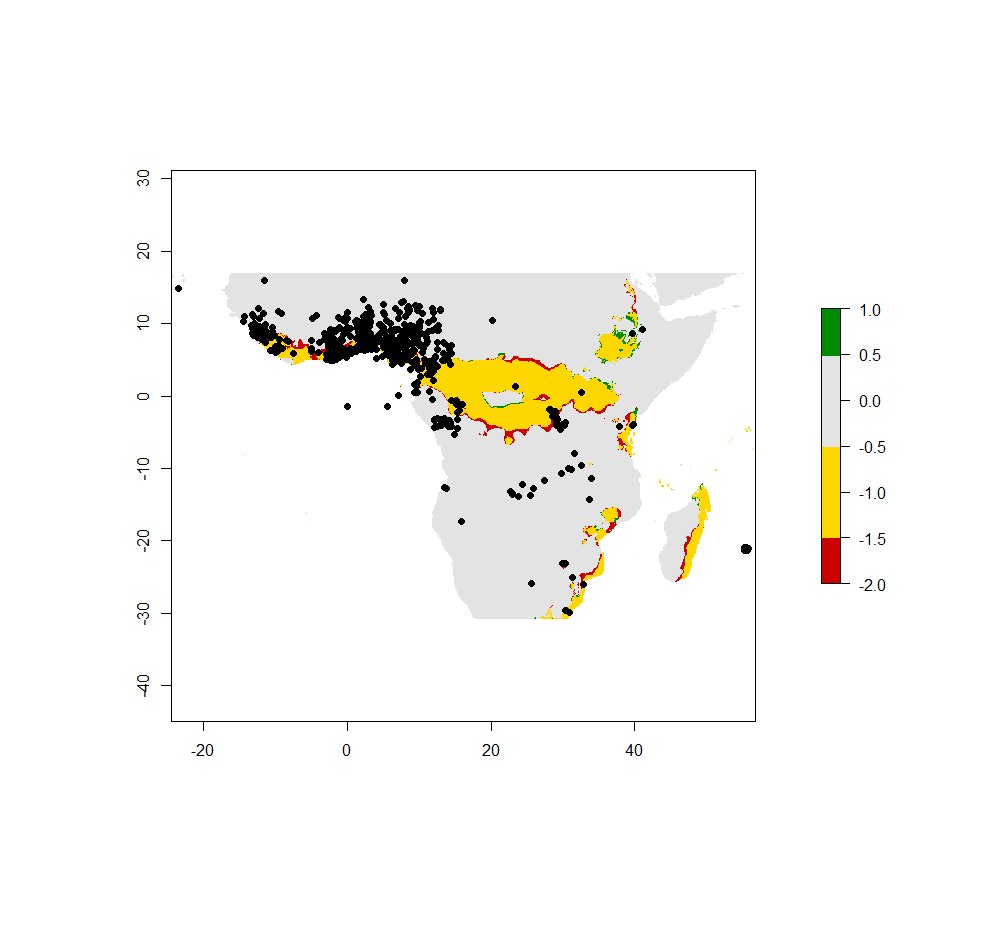
Figure 4 shows the generated map that shows the change in distribution for the present compared to the future and reported sightings (as black dots). Changes in the future are minimal

Figure 4: Change in distribution area for Manihot esculenta in the future

**Biological interpretation:**

The model shows the distribution area to largely remain the same in the future. There are some gains and losses on the edges, as can be expected, however the changes are minor. Presence locations however do not match the present and future distribution map very well. This probably means there are other factors involved in the distribution for the species, such as humans cultivation of the crop.

According to this model the cassava crop is mostly affected by feature changes in precipitation in the driest month. It appears to be near its limit for drought tolerance. So even though cassava is known to grow well in dry habitats, a further decrease in precipitation is likely limiting for the species distribution.

This model is quite limited and is therefore not really reliable or useful. Reason being that the model is only run once and only verified by AUC values. And even though this is the mostly widely used method for testing models, a better verification method would be verification by a Null model. Furthermore, for more reliable outcomes the model needs to be tested for a multitude of times. Another problem is the spatial limitation of the model. The data used is only from Africa and it is therefore, when used on a global level, extrapolated to climatic values outside the scope of the reference data. Finally biotic and other factors are not taken into account by this model.