Colocasia esculenta species distribution model

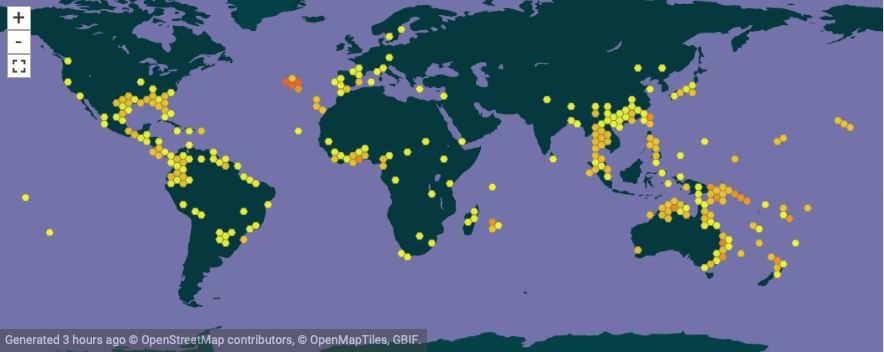
By: Jan Lankhorst

# Introduction

The Taro (*Colocasia esculenta*) is a tuberous plant with a main distribution and origin in the tropics. It is thought to have originated in Asia, around India, Nepal and Bangladesh. Due to its edible roots, it has been cultivated and consequently spread westward throughout Asia and Oceania. It also spread eastward to Africa, the Mediterranean and even the Americas.

The Taro has a few varieties which differ in their environmental tolerances. The “original” Taro prefers very wet soils, while the Eddoe is more drought and cold resistant. This leads to a very broad range of possible environments where the *Colocasia esculenta* can survive.

In figure 1. the current distribution is shown. Already a very broad distribution can be seen, from the wet tropics in South-East Asia and South-America to the drier Mediterranean basin and eastern United States.

**

*Figure 1. Occurences as downloaded from Gbif.org.*

# Methodology

Occurrence data was obtained from GBIF, with a search on *Colocasia esculenta* with coordinates included: <https://doi.org/10.15468/dl.crfhse>. This found 7026 georeferenced records which were downloaded and loaded into R with the R-studio IDE. To link the occurrences to climatic variables and make prediction for future distributions, both present and predicted climate data was needed. Both of these data files were downloaded from WorldClim.org. 5-minute (present) bioclimatic variables were downloaded from: <http://biogeo.ucdavis.edu/data/climate/worldclim/1_4/grid/cur/bio_5m_bil.zip>

Predicted bioclimatic variables for 2050 were downloaded at 5-minute resolution with a RCP45 scenario from:

<http://biogeo.ucdavis.edu/data/climate/cmip5/5m/hd45bi50.zip>

RCP 4,5 was chosen as this is more of a midway between the most optimistic (RCP 2,6) and the worst-case scenario (RCP 8,5). The year 2050 was chosen because the Taro is an important staple crop for many farmers in the tropical regions so near-future predictions are important for these farmers.

In order to link species distribution to climatic variables, a first step was to remove highly correlated variables from the model. All correlations (Spearman rho) higher than 0,7 or lower than -0,7 were omitted, leaving a model with 6 bioclimatic variables.

The variables used were:

bio18 precipitation of warmest quarter

bio19 precipitation of coldest quarter

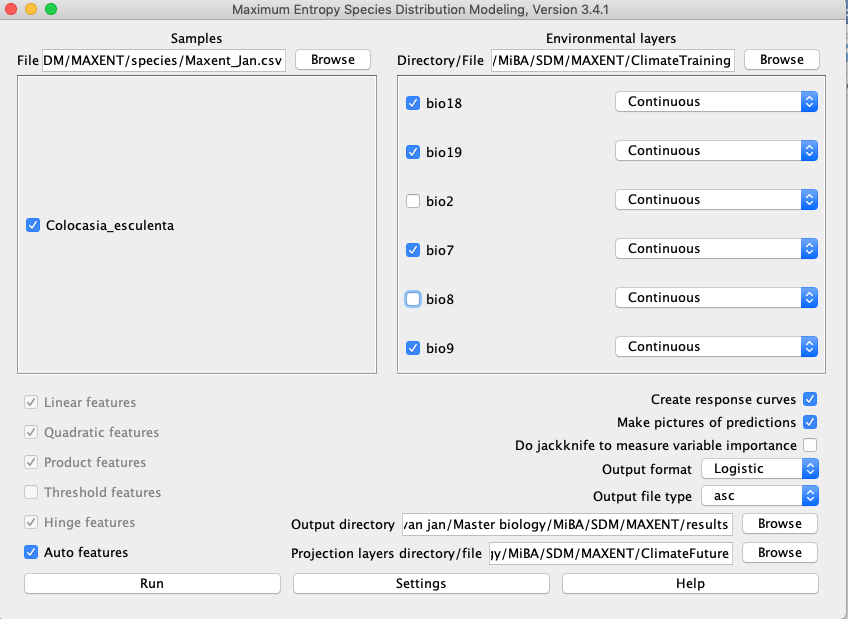
bio2 Mean of monthly max temp/min temp

bio7 Temperature annual range

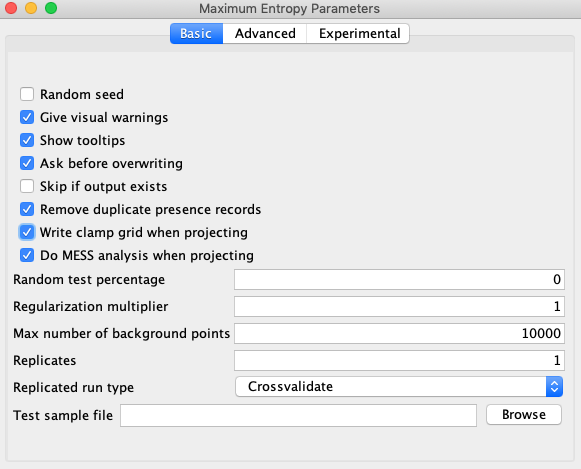
bio8 Mean temperature of wettest quarter

bio9 Mean temperature of coldest quarter

Using these variables and Maxent version 3.4.1. Species Distribution Models and the following analysis could be done. Maxent settings were as seen in figure 2.1 and 2.2. The relative importance of each variable can be seen in figure 3. The percent contribution of each variable is reasonably spread, with no variables contributing above 35% or below 5%.

**

*Figure 2.1. Maxent settings. “Projection layers directory” has both present and future climate data folders.*

**

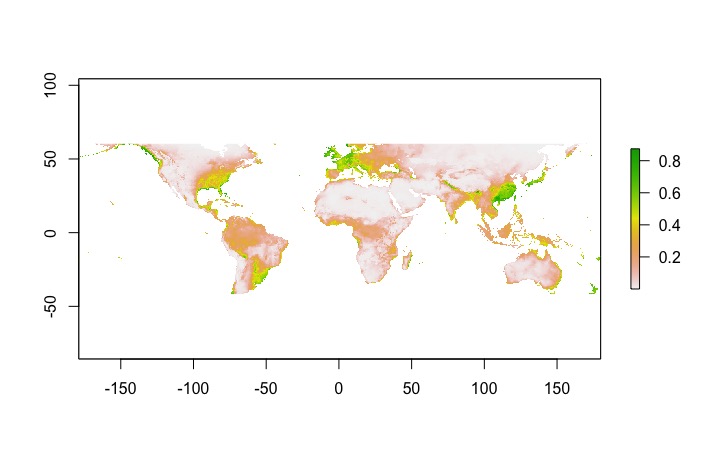


*Figure 3. Variable Importance Table*

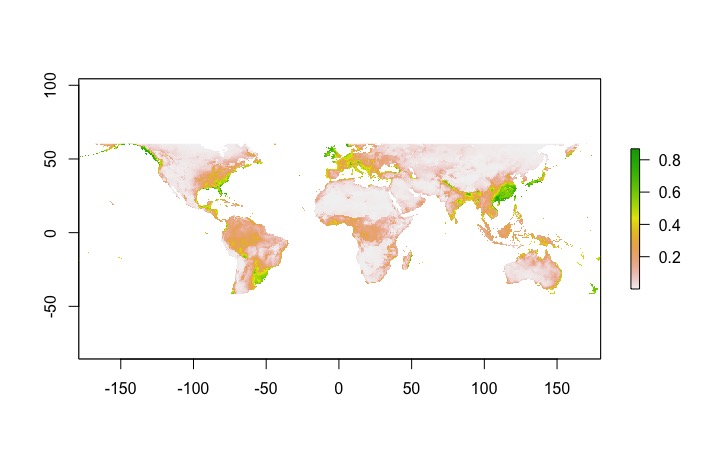
*Figure 2.2 “settings” tab from Maxent*

# Model output

Based on the aforementioned variables, a current habitat suitability and a predicted suitability for 2050 could be modeled and in R-studio visualized. The results are in figure 4 and 5 respectively.

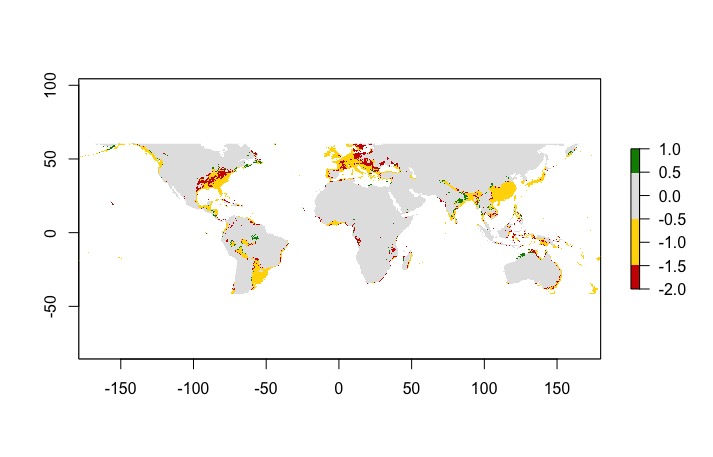


*Figure 4. Present habitat suitability, green colors showing more suitable habitat, red colors showing less suitable habitat.*

**

*Figure 5. Future habitat suitability, green colors showing more suitable habitat, red colors showing less suitable habitat.*

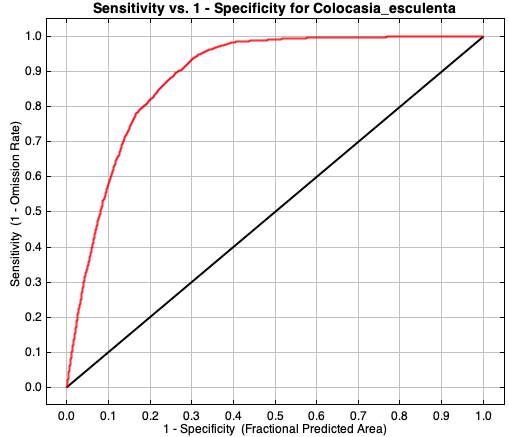
Not much difference can be seen between the two figures. Europe and North America lose the most habitat suitability and some parts of Oceania lose a little. When overlaying the two suitability maps, differences are more clearly visible (figure 6.). A general loss of habitat suitability caused by climate change can be seen. Some parts of the world gain possible habitats for the Taro (green parts) most of the current habitat is still suitable in the future (yellow) but also a large part is not suitable in the future (red).



*Figure 6. Overlap map of present and future habitat suitability. Green is gained habitat, yellow is maintained habitat and red is lost habitat.*

This model seems not very predictive and it might be caused by the variables which were chosen based on the lack of correlation and not necessarily on biologic relevance. On the other hand, precipitation in warmest and coldest quarters are very relevant factors for plant growth and should be able to explain habitat suitability to at least a small extent.

The model’s AUC is 0.888 with the ROC curve shown in figure 7. This AUC value is considered pretty good, but AUC’s are not all that matter and should not blindly be followed.

**

*Figure 7. ROC curve with training data (red) with an AUC of 0.888 and random prediction (black) with an AUC of 0.5*

In general, this model might be useful for farmers who are dependent on Taro for their income, to be able to see that there are changes in climate coming which influences the possible production area for this species. Exact conclusions should not be drawn as this model with only 6 climate variables has most likely not captured all relevant data and needs some extent of optimization before it can be commercially used.