|  |
| --- |
| Methods in biodiversity analysis  Habitat suitability prediction of the groundnut (Arachis hypogaea)  11-12-2017  C:\Users\Sam\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Peanut-Arachis+hypogaea+2.jpgSam Boerlijst S1483439 |

**Introduction to the species**

The groundnut (*Arachis hypogaea*), also known as the peanut, is a widely cultivated annual crop. It is mainly grown as a source for proteins and oil, both for human consumption and as fodder. The groundnut is primarily cultivated in semi-arid tropical Asia and sub-Saharan Africa and with a global production estimated at 60 million tons, it is ranked fifth among the most important sources of vegetable oils (FAO, 2013).

Major losses in groundnut production are caused by combinations of high temperatures and drought (Sharma, K. et al., 2002; Hamidou, H. et al., 2013). This could cause problems in the near future, since these abiotic factors are expected to increase with climate change.

Since *A. Hypogaea* is a tetraploid unlike its close relatives, cross-breeding for drought resistance could be troublesome (Moretzsohn et al., 2013). This, in combination with the low polymorphism detected in the species, make the groundnut likely to be more susceptible to abiotic stressors.

The objective of this short study was to access the influence of climate change on habitat suitability for the groundnut, focussing on the regions where it is currently cultivated.

**Methodology**

Habitat suitability was modelled with data acquired from the Bioclim website.

Bioclim predictions for the scenario rcp45 at 5m (HadGEM2-CC) were chosen since these represent the predictions for the current trend in climate change. Predictions for 2050 were used to model short term effects due to the expected vulnerability of the crop.

Suitable uncorrelated variables were selected in R by identifying correlations greater than 0.7 or lower than -0.7 with spearman’s correlation. Subsequently variables were chosen based on the ecology of the groundnut, mainly focussing on its expected sensitivity to drought and the combined effect of drought and high temperatures on evaporation. Due to its current distribution centred around the equator, mean temperatures were expected to have a lesser effect than fluctuations in temperature. Also, short term fluctuations were chosen over long term fluctuations because the groundnut is an annual plant.

This lead to a subset based on the variables: mean diurnal range (Bio2), mean temperature of the driest quarter (Bio9), annual precipitation (Bio12), precipitation of the driest month (Bio14).

A test set was created with jackknife resampling for model validation.

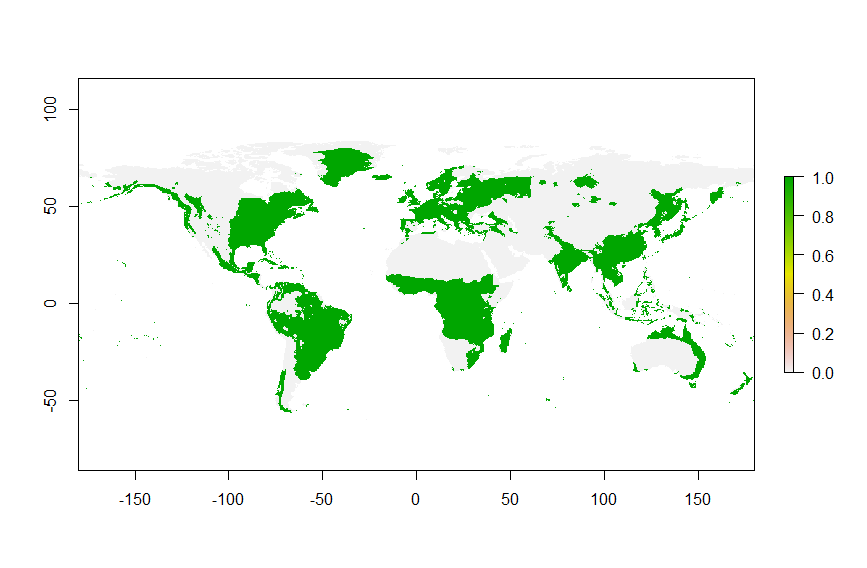
With the created subsets a species distribution model was created in Maxent. The output was set to logistic and the option to do jackknife to measure variable importance was disabled.

Only linear and quadratic features were used since these are expected to represent the future distribution the most accurate.

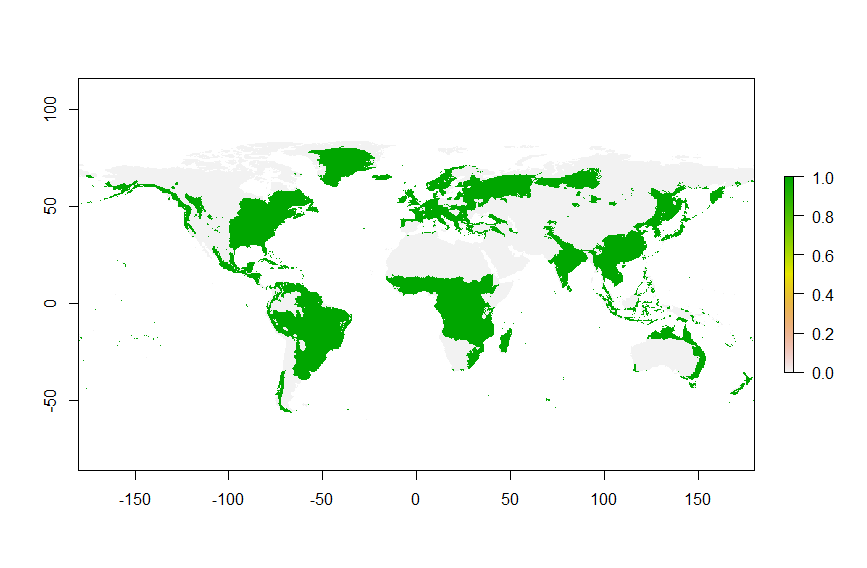
**Model output**

The model resulted in distributions of suitable habitat far greater than the actual distribution of the species. This is likely due to its cultivation, which hinders both the amount of offspring and dispersal.

At first glance, not much difference can be found between the present and future distribution of suitable habitat. Only some difference is visible in Russia. This suggests that there is no strong effect caused by climate change under the RCP 4.5 2050 scenario.



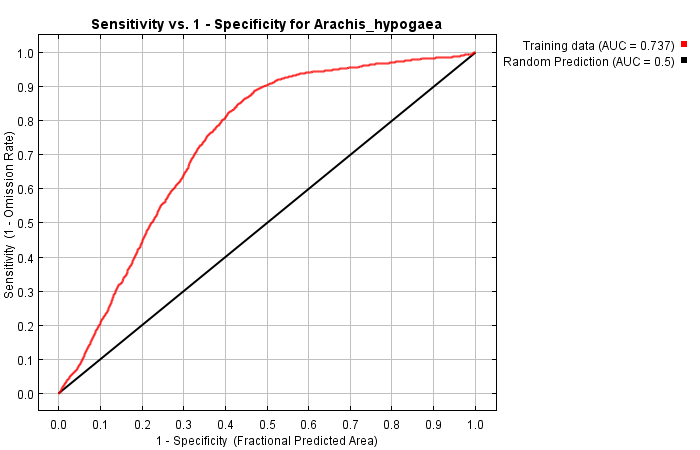
**Figure 2** Current distribution of suitable habitat



**Figure 3** Future distribution of suitable habitat

**Model performance**

The AUC of the receiver operator curve of the model is 0.737. The model can thus be seen as reliable. The logistic threshold for maximum training sensitivity plus specificity of the model is 0.472.



**Figure 4** Receiver operator curve of the model

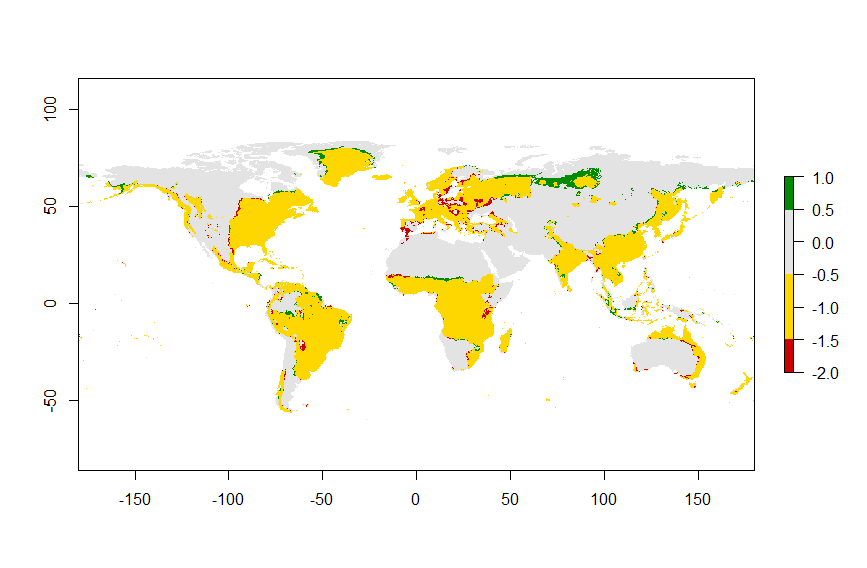
According to the variable contribution table, annual precipitation (bio12) is the main driver of the distribution pattern.

**Table 1** variable contributions

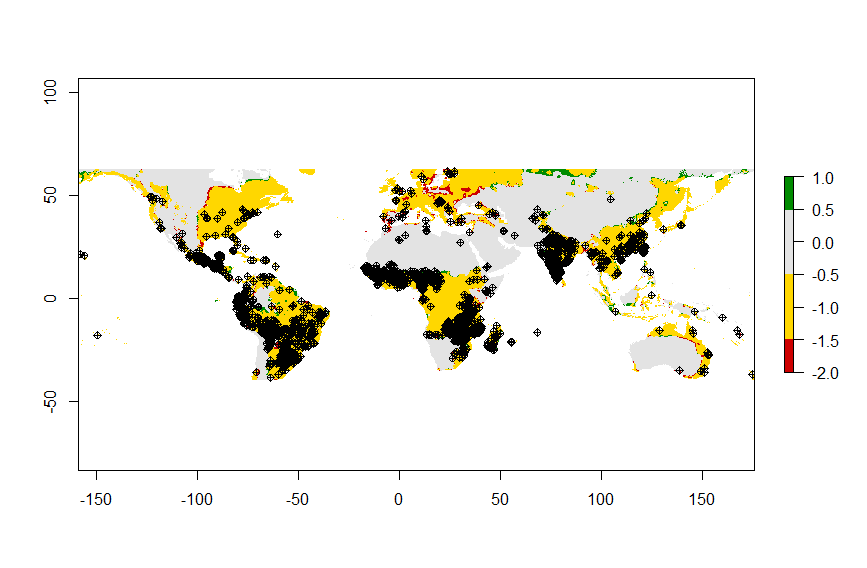
|  |  |  |
| --- | --- | --- |
| Variable | Percent contribution | Permutation importance |
| bio12 | 79.1 | 90 |
| bio14 | 16.1 | 5 |
| bio2 | 2.9 | 0.9 |
| bio9 | 2 | 4.1 |

**Response to future distribution**

Indeed, there is only little difference in the present and future habitat suitability. Climate change appears not to be a great threat for groundnut cultivation under the RCP 4.5 2050 scenario.



**Figure 5** Differences in suitable habitat according to the model. Gray areas are never suitable, yellow areas remain suitable, red indicates where suitable habitat is lost and green indicates where suitable habitat is gained.



**Figure 6** Differences in suitable habitat. The map is cropped to the extent of the GBIF data and the entries are overlayed

**Biological interpretation**

According to the model only little change in distribution is expected. Peanut cultivation therefore does not have to shift. Annual precipitation was identified to be the main driver of the distribution, which was to be expected, but is still striking, since precipitation in the driest month – according to the variable contributions - has only little effect. Possibly the duration of the drought is more important than the amount of drought.

Also, the size difference between the climatic envelope and the actual distribution indicates that there is a variable driving the distribution which is outside the model. Being a crop, this is likely to be the result of cultivation.

In the end the model is not all that useful, since it has not been verified with a null model or back-testing. However, the outcome is not surprising and could very well be correct. The groundnut is quite widely distributed and the effect of temperature change should therefore not be a great stressor. However, since the groundnut is cultivated, the main driver of its distribution is likely the human. Another limitation of the model is the course scale used for the bioclim variables. The 5m scale suffices for the practical, but for an accurate model a finer scale is needed. Additionally, the model could be run multiple times to build a consensus, thereby decreasing the effect of stochastity on the pseudo absences. Also, like many models it does not encompass biotic interactions. The groundnut is a flowering plant and must be pollinated. Climate change could cause a mismatch with its pollinators.

Hamidou, F., Halilou, O., & Vadez, V. (2013). Assessment of Groundnut under Combined Heat and Drought Stress: Heat and{Bibliography} Drought Stress in Groundnut. *Journal of Agronomy and Crop Science*, *199*(1), 1–11. https://doi.org/10.1111/j.1439-037X.2012.00518.x

Moretzsohn, M. C., Gouvea, E. G., Inglis, P. W., Leal-Bertioli, S. C. M., Valls, J. F. M., & Bertioli, D. J. (2013). A study of the relationships of cultivated peanut (Arachis hypogaea) and its most closely related wild species using intron sequences and microsatellite markers. *Annals of Botany*, *111*(1), 113–126. https://doi.org/10.1093/aob/mcs237

Sharma, K K and Lavanya, M (2002) Recent developments in transgenics for abiotic stress in legumes

of the semi-arid tropics. JIRCAS Working Report No. 23, 23. pp. 61-73. ISSN 1341-710X

**Appendix R-script**

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

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

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

############## SDM Workshop - R Script ##############

############## Clipping + Variable Selection ##############

############## Friday 8th December 2017 ##############

############## Leon Marshall - Naturalis Biodiversity Center ##############

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

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

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

#Packages

library(sp)

library(rgdal)

library(raster)

library(biomod2)

#functions

cor.prob <- function (X, dfr = nrow(X) - 2) {

R <- cor(X, use="pairwise.complete.obs", method ="spearman")

above <- row(R) < col(R)

r2 <- R[above]^2

Fstat <- r2 \* dfr/(1 - r2)

R[above] <- 1 - pf(Fstat, 1, dfr)

R[row(R) == col(R)] <- NA

R

}

# Use this to dump the cor.prob output to a 4 column matrix

flattenSquareMatrix <- function(m) {

if( (class(m) != "matrix") | (nrow(m) != ncol(m))) stop("Must be a square matrix.")

if(!identical(rownames(m), colnames(m))) stop("Row and column names must be equal.")

ut <- upper.tri(m)

data.frame(i = rownames(m)[row(m)[ut]],

j = rownames(m)[col(m)[ut]],

cor=t(m)[ut],

p=m[ut])

}

# Load species occurrence file

sp <- read.csv("D:/opdracht MBA/MAXENT/Species/AH.csv",header=T) #load csv of occurrence

head(sp) #check table looks correct

sp\_shp <- sp #rename table

coordinates(sp\_shp) <- ~longitude+latitude #convert table to points shapefile

proj4string(sp\_shp) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no\_defs") #define projection WGS1984

#Create bounding box around points

bbox <- extent(sp\_shp) #create bounding box of points

bbox <- bbox+2 #increase border so we do not truncate data

plot(bbox, col='blue') #check if box surrounds points

plot(sp\_shp,add=T,pch=19,col='red') #add points

#Load Bioclim rasters

bioclims <- stack(list.files("D:/bioclim\_now",

pattern="\*.bil$", full.names=TRUE,

ignore.case=TRUE)) #create a stack of present bioclims

bioclims\_fut <- stack(list.files("D:/bioclim\_future",

pattern="\*.tif$", full.names=TRUE,

ignore.case=TRUE)) #create a stack of future bioclims

#Clip rasters by bounding box

bioclims\_bbox <- stack() #create empty raster stack

for(i in 1:19){

ras <- crop(bioclims[[i]],bbox) #clip rasters by bbox

bioclims\_bbox <- stack(bioclims\_bbox,ras) #stack rasters together

}

plot(bioclims\_bbox)

#Variable Selection

bioclims\_corr <- as.data.frame(bioclims\_bbox, xy=T) #convert raster stack to a table

bioclims\_corr <- na.omit(bioclims\_corr[3:21]) #remove NA values (ocean) and remove coordiantes

sc <- flattenSquareMatrix(cor.prob(bioclims\_corr)) #Table summary correlations

sc1 <- sc[with(sc, order(cor)), ]

sc2 <- sc1[sc1$cor>=0.7|sc1$cor<=-0.7,] #High Correlations

sc2[with(sc2, order(i)), ] #list correlated variables

bioclims\_corr2 <- subset(bioclims\_corr, select=-c(bio1, bio3, bio4, bio5, bio6, bio7, bio8, bio10, bio11, bio13, bio15, bio16, bio17, bio18, bio19))

#remove unwanted variables

sc <- flattenSquareMatrix(cor.prob(bioclims\_corr2)) #Table summary correlations

sc1 <- sc[with(sc, order(cor)), ]

sc2 <- sc1[sc1$cor>=0.7|sc1$cor<=-0.7,] #High Correlations

sc2[with(sc2, order(i)), ] #list correlated variables

#Make selection of chosen variables

list.names <- c("bio1","bio10","bio11","bio12","bio13","bio14",

"bio15","bio16","bio17","bio18","bio19","bio2",

"bio3","bio4","bio5","bio6","bio7","bio8","bio9") #list of names in same order as rasterstacks

select.names <- names(bioclims\_corr2) #list of selected rasters

numbers <- match(select.names,list.names) #numbers corresponding to chosen variables

pres\_bioclims <- bioclims[[numbers]] #select only variables interested in

fut\_bioclims <- bioclims\_fut[[numbers]] #select only variables interested in

clip\_bioclims <- bioclims\_bbox[[numbers]] #select only variables interested in

#Save bioclims as .asc

dir.create("D:/opdracht MBA/MAXENT/ClimatePresent") #create folder

dir.create("D:/opdracht MBA/MAXENT/ClimateFuture") #create folder

dir.create("D:/opdracht MBA/MAXENT/ClimateTraining") #create folder

for(i in 1:nlayers(pres\_bioclims)){

ras\_sc=scale(pres\_bioclims[[1]],center=TRUE, scale=TRUE) #scale variables

writeRaster(ras\_sc,

file=paste("D:/opdracht MBA/MAXENT/ClimatePresent/",select.names[i],".asc",sep=""))

}

for(i in 1:nlayers(fut\_bioclims)){

ras\_sc=scale(fut\_bioclims[[1]],center=TRUE, scale=TRUE) #scale variables

writeRaster(ras\_sc,

file=paste("D:/opdracht MBA/MAXENT/ClimateFuture/",select.names[i],".asc",sep=""))

}

for(i in 1:nlayers(clip\_bioclims)){

ras\_sc=scale(clip\_bioclims[[1]],center=TRUE, scale=TRUE) #scale variables

writeRaster(ras\_sc,

file=paste("D:/opdracht MBA/MAXENT/ClimateTraining/",select.names[i],".asc",sep=""))

}

#Run maxent

#Create Binary map based on Threshold

pres <- raster("D:/opdracht MBA/MAXENT/Results/Arachis\_hypogaea\_ClimatePresent.asc")

# load present global distribution

plot(pres) #visualize

pres\_clip <- crop(pres,bbox) #clip to training area

plot(pres\_clip) #visualize

fut <- raster("D:/opdracht MBA/MAXENT/Results/Arachis\_hypogaea\_ClimateFuture.asc")

# load future global distribution

plot(fut)

fut\_clip <- crop(fut,bbox) #clip to training area

plot(fut\_clip) #visualize

th <- 0.387 #define threshold

m <- c(0, th,0, th, 1, 1) #matrix everything before th as 0 and everything after th as 1

bin\_mat <- matrix(m, ncol=3, byrow=TRUE) #convert to correct matrix

pres\_bin <- reclassify(pres, bin\_mat) #reclassify by matrix present

fut\_bin <- reclassify(fut, bin\_mat) #reclassify by matrix future

plot(pres\_bin)

plot(fut\_bin)

#Calculate Change in Sutiable Climate Conditions

range\_change<-BIOMOD\_RangeSize(CurrentPred=pres\_bin,FutureProj=fut\_bin,SpChange.Save=NULL)

#calculate range change between two maps

col.lst <- c("red3" , "gold", "grey89" , "green4") #define plot colours

plot(range\_change$Diff.By.Pixel,col=col.lst) #plot range change

#Calculate occurrence area change maps

pres\_bin\_clip <- crop(pres\_bin,bbox) #clip binary map

fut\_bin\_clip <- crop(fut\_bin,bbox) #clip binary map

range\_change\_clip<-BIOMOD\_RangeSize(CurrentPred=pres\_bin\_clip,FutureProj=fut\_bin\_clip,SpChange.Save=NULL)

#calculate range change between two maps

col.lst <- c("red3" , "gold", "grey89" , "green4") #define plot colours

plot(range\_change\_clip$Diff.By.Pixel,col=col.lst) #plot range change

plot(sp\_shp,add=T, col="black", pch=10) #show points on top