Thaumatotibia\_leucotreta

Nico

2/18/2022

## Front Matter

The following script was developed cooperatively by the SHSU SDM working group, including Laura Bianchi, Austin Brenek, Jesus Castillo, Nick Galle, Kayla Hankins, Kenneth Nobleza, Chris Randle, Nico Reger, Alyssa Russell, Ava Stendahl based on [tutorials](https://rspatial.org/) provided by Robert Hijmans and Jane Elith. Chris Randle composed the following script from many scripts developed by the SHSU SDM working group.

*This works best if your environment is empty at the start.*

I have tried to set this up to eliminate required changes to the code. When you see text in **BOLD** below, that will be an indication that you need to make a decision.

# Libraries

library(dismo)

## Loading required package: raster

## Loading required package: sp

library(sp)  
library(raster)  
library(stats)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:raster':  
##   
## intersect, select, union

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(knitr)  
library(rgeos)

## rgeos version: 0.5-8, (SVN revision 679)  
## GEOS runtime version: 3.9.1-CAPI-1.14.2   
## Please note that rgeos will be retired by the end of 2023,  
## plan transition to sf functions using GEOS at your earliest convenience.  
## GEOS using OverlayNG  
## Linking to sp version: 1.4-5   
## Polygon checking: TRUE

library(maptools)

## Checking rgeos availability: TRUE  
## Please note that 'maptools' will be retired by the end of 2023,  
## plan transition at your earliest convenience;  
## some functionality will be moved to 'sp'.

library(rgdal)

## Please note that rgdal will be retired by the end of 2023,  
## plan transition to sf/stars/terra functions using GDAL and PROJ  
## at your earliest convenience.  
##   
## rgdal: version: 1.5-27, (SVN revision 1148)  
## Geospatial Data Abstraction Library extensions to R successfully loaded  
## Loaded GDAL runtime: GDAL 3.2.1, released 2020/12/29  
## Path to GDAL shared files: C:/Users/User/Documents/R/win-library/4.1/rgdal/gdal  
## GDAL binary built with GEOS: TRUE   
## Loaded PROJ runtime: Rel. 7.2.1, January 1st, 2021, [PJ\_VERSION: 721]  
## Path to PROJ shared files: C:/Users/User/Documents/R/win-library/4.1/rgdal/proj  
## PROJ CDN enabled: FALSE  
## Linking to sp version:1.4-5  
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,  
## use options("rgdal\_show\_exportToProj4\_warnings"="none") before loading sp or rgdal.  
## Overwritten PROJ\_LIB was C:/Users/User/Documents/R/win-library/4.1/rgdal/proj

library(ecospat)

## Loading required package: ade4

## Loading required package: ape

##   
## Attaching package: 'ape'

## The following objects are masked from 'package:raster':  
##   
## rotate, zoom

## Loading required package: gbm

## Loaded gbm 2.1.8

## Registered S3 methods overwritten by 'adehabitatMA':  
## method from  
## print.SpatialPixelsDataFrame sp   
## print.SpatialPixels sp

library(usdm)  
library(mgcv)

## Loading required package: nlme

##   
## Attaching package: 'nlme'

## The following object is masked from 'package:usdm':  
##   
## Variogram

## The following object is masked from 'package:dplyr':  
##   
## collapse

## The following object is masked from 'package:raster':  
##   
## getData

## This is mgcv 1.8-38. For overview type 'help("mgcv-package")'.

# Genus and Species strings

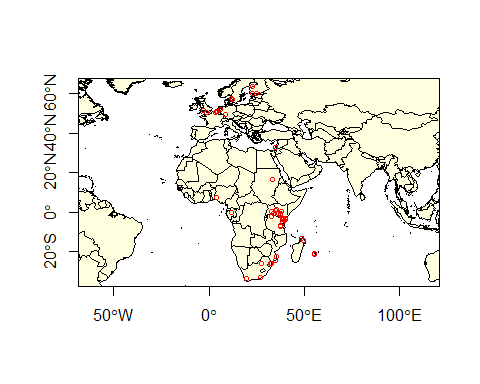
Their are many places in this code where you will need to save files with filenames including the genus and species. We’ll save these as strings to automate the creation of file names. Enter your genus name and specific epithet in the quotes below.

genus<-"Thaumatotibia"  
species<-"leucotreta"

## Occurrence Data

Import occurrence data from csv file already generated (2020-2021), or using the script “Occurrence\_Data.rmd” and visualize it.

sdmdata <- read.csv(file = "Thaumatotibia\_leucotreta\_clean.csv")  
##and visualize the data  
#first lets get the extent of the data (the coordinates of the smallest box needed to encapsulate the data) To do this I first need to convert sdmdata into a spatial points dataframe with the same crs as "wrldsmpl", a giant spatial polygons data frame available from maptools  
sdmdataframe<-data.frame(sdmdata)  
data(wrld\_simpl)  
coordinates(sdmdataframe) <- ~lon+lat  
crs(sdmdataframe) <- crs(wrld\_simpl)  
#And then extract the extent  
e<-extent(sdmdataframe)  
xmin<-xmin(e)  
xmax<-xmax(e)  
ymin<-ymin(e)  
ymax<-ymax(e)  
# and then plot a map and add the points from sdmdata  
plot(wrld\_simpl, xlim=c(xmin,xmax), ylim=c(ymin,ymax), axes=TRUE, col="light yellow")  
box()  
points(sdmdata$lon, sdmdata$lat, col='red', cex=0.75)



Let’s divide the data into training and testing data sets. The following code divides the data set into 80% training and 20% testing.

#let's make sdmdata into a dataframe  
data(wrld\_simpl)  
coordinates(sdmdata) <- ~lon+lat  
crs(sdmdata) <- crs(wrld\_simpl)  
  
#let's extract just the coordinates  
presence <- coordinates(sdmdata)  
#First we'll make a random list of integers from 1-5 as long as our presence data. Setting the seed results in a repeatable random process  
set.seed(0)  
#now make a list as long as the number of rows in presence consisting of a random series of integers from 1-5  
group <- kfold(presence, 5)  
#Then we want to use this to retrieve the number of the rows in the presence data that are associated with the number 1 in our group index.  
test\_indices <- as.integer(row.names(presence[group == 1, ]))  
#and create a new list of coordinates including only those rows that are NOT in test indices. These are all the row numbers NOT corresponding with the test\_indices (which is ~80% of the data).  
pres\_train <- presence[-test\_indices,]  
#and those that do correspond with test indices (20%) of the data  
pres\_test <- presence[group ==1,]

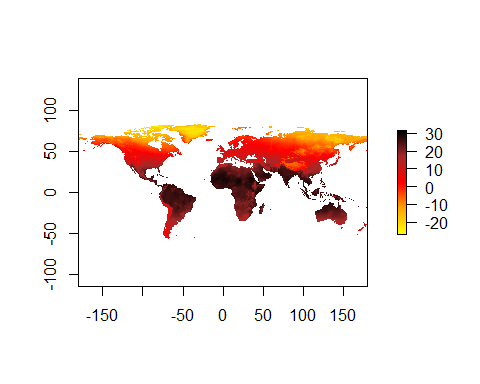
Save pres\_data and test\_data as csv files just in case.

#first presdata\_train  
outdata<-data.frame(pres\_train)  
colnames(outdata)<-c("lon","lat")  
write.csv(outdata, file=paste0(genus,"\_",species,"\_train.csv"), row.names=FALSE)  
  
#and then presdata\_test  
outdata<-data.frame(pres\_test)  
colnames(outdata)<-c("lon","lat")  
write.csv(outdata, file=paste0(genus,"\_",species,"\_test.csv"), row.names=FALSE)

## Predictor data

Let’s get the giant predictor file, name the bands, and generate our raster color schemes. This predictor set consists of all 35 Climond layers and elevation. Get it from Randle and keep it in your directory.

predictors<-stack('Climond\_elev\_HI.tif')  
bands<-c("Ann\_mean\_temp", "Mean\_durnal\_temp\_range", "Isothermality", "Temp\_Seasonality", "MaxTemp\_WarmestWeek", "MinTemp\_ColdestWeek", "Temp\_Annual\_Range", "Mean\_temp\_wettest\_q", "Mean\_temp\_driest\_q", "Mean\_temp\_warmiest\_q", "Mean\_temp\_coldest\_q", "Ann\_ precip", "Precip\_driest\_week", "Precip\_wettest\_week", "Precip\_Seasonality", "Precip\_wettest\_q", "Precip\_driest\_q", "Precip\_warmest\_q", "Precip\_coldest\_q", "Ann\_mean\_rad", "Hghest\_weekly\_rad", "Lowest\_weekly\_rad", "Rad\_seasonality", "Rad\_wettest\_q", "Rad\_driest\_q", "Rad\_warmest\_q", "Rad\_coldest\_q", "Ann\_mean\_moisture", "Highsets\_weekly\_moisture", "Lowest\_weekly\_moisture", "Moisture\_seasonality", "Mean\_moisture\_wettest\_q", "Mean\_moisture\_driest\_q", "Mean\_moisture\_warmest\_q", "Mean\_moisture\_coldest\_q", "Elev", 'human\_impact')  
names(predictors)<-bands  
cool<-colorRampPalette(c('gray','green','dark green',"blue"))  
warm<-colorRampPalette(c('yellow', 'orange', 'red', 'brown', 'black'))  
plot(predictors[["Ann\_mean\_temp"]], col=warm(100))

 And now we will use the VIFstep function to identify layers contributing most to collinearity (variance inflation factor). Rather than do this from a raster, I think it makes much more sense to do this from a dataframe in which we have sampled all the layers at the presence points only. This is because the larger a species distribution is, the lower the probability of collinearity across the range, even if layers are collinier where the species actually exists in the range.

#extract environmental data using the points in sdmdata  
env\_data<-extract(predictors,sdmdata)  
#give names to the columns  
colnames(env\_data)<-bands  
#run the vif  
vif<-vifstep(env\_data)  
#and let's find the layers that were excluded and drop them  
excluded<-vif@excluded  
predictors<-dropLayer(predictors,excluded)  
#and let's just go ahead and see which layers were dropped.  
NClayers<-names(predictors)  
NClayers

## [1] "Mean\_durnal\_temp\_range" "Ann\_.precip"   
## [3] "Precip\_Seasonality" "Precip\_driest\_q"   
## [5] "Precip\_coldest\_q" "Hghest\_weekly\_rad"   
## [7] "Rad\_wettest\_q" "Rad\_driest\_q"   
## [9] "Highsets\_weekly\_moisture" "Mean\_moisture\_warmest\_q"   
## [11] "Elev" "human\_impact"

## General additive model

# Data preparation

Generally speaking, we want to sample absence data from the region in which the presence data occur. There are two ways to do that, and one of them is better than the other. The first is to sample randomly. That may seem like a good idea, but its counter-intuitively not. The reason is that the presence data likely includes sampling bias. This will be inherent in p(hypothesis). If we include the same sampling bias in p(data), they cancel out in Bayes Theorem. The way that we’ll do that is to create circles around our data and sample absence points from within those. The circles will have a diameter equal to the average distance between points.

#convert presence training data into a spatial points dataframe.  
pres\_train\_SPDF<-SpatialPoints(pres\_train)  
crs(pres\_train\_SPDF) <- crs(wrld\_simpl)  
#Let's get the average distance between points (great circle distance--takes into account the curvature of the earth). spDists creates a matrix of distances between points. This includes zeros.   
dist<-spDists(pres\_train\_SPDF,longlat = TRUE)  
#replace the zeros with NA  
dist[dist == 0]<-NA  
#and calculate the mean--this is the average distance between points...the result will be in kilometers, but we need to convert it to meters so we multiply by 1000  
dist<-1000\*mean(dist, na.rm=TRUE)  
#now we are going to make circles using the average distance between points as the diameter.   
x <- circles(pres\_train\_SPDF, d=dist, lonlat=TRUE)  
#and convert those into polygons  
pol <- polygons(x)  
#and draw a number of samples from that approximately three times the number of presence points. We'll chop that down at the end.  
samp1 <- spsample(pol, nrow(pres\_train)\*3, type='random', iter=25)

## Warning in proj4string(obj): CRS object has comment, which is lost in output

#and get the cell numbers from the raster stack (right to left, up to down)  
cells <- cellFromXY(predictors, samp1)  
#and transform each of those to the center of its cell.  
abs\_train <- xyFromCell(predictors, cells)  
#You'll get a warning saying that your CRS object has lost a comment. This is unimportant and can be ignored.

And let’s go ahead and extract the presence data, remove rows with NA values, and add a column of 1s.

pres\_train\_data<-extract(predictors,pres\_train)  
complete<-complete.cases(pres\_train\_data)  
pres\_train\_data<-pres\_train\_data[complete,]  
pres\_train\_data<-cbind(pres\_train\_data,1)

Now we want to extract predictors for the absence data, remove rows with NA values and chop it down to the size of our presence training data, and combine these into one data frame with column names (pa is the last column of 0,1 which indicates presence or absence)

abs\_train\_data<-extract(predictors,abs\_train)  
#remove rows with NA values  
complete<-complete.cases(abs\_train\_data)  
abs\_train\_data<-abs\_train\_data[complete,]  
#and select a number of rows equal to the presence training data  
abs\_train\_data<-abs\_train\_data[1:nrow(pres\_train\_data),]  
#and add a column of zeros to the end.  
abs\_train\_data<-cbind(abs\_train\_data,0)  
#put the two matrices together and name the colmns  
train\_data<-rbind(pres\_train\_data,abs\_train\_data)  
colnames(train\_data)<-c(names(predictors),"pa")  
train\_data<-as.data.frame(train\_data)

[1] “Mean\_durnal\_temp\_range” “Ann\_.precip” “Precip\_Seasonality” “Precip\_driest\_q” “Precip\_coldest\_q” “Hghest\_weekly\_rad” “Rad\_wettest\_q”  
[8] “Rad\_driest\_q” “Highsets\_weekly\_moisture” “Mean\_moisture\_warmest\_q” “Elev” “human\_impact”

# Training the GAM and making predictions.

**This is a pain in the neck because all of the layers have to be specified. I recommend printing the column names in the console colnames(train\_data) and then copying them and formatting them**

gam <- gam(pa ~ Mean\_durnal\_temp\_range + Ann\_.precip + Precip\_Seasonality +   
Precip\_driest\_q + Precip\_coldest\_q + Hghest\_weekly\_rad + Hghest\_weekly\_rad + Rad\_wettest\_q + Rad\_driest\_q + Rad\_driest\_q+ Highsets\_weekly\_moisture + Mean\_moisture\_warmest\_q, Elev, human\_impact, ,  
 family = binomial(link = "logit"), data=train\_data)

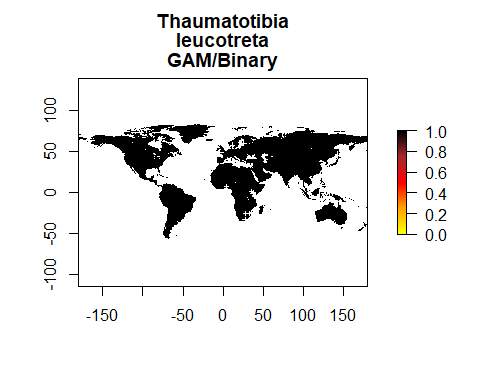
## Warning in gam.fit3(x = args$X, y = args$y, sp = lsp, Eb = args$Eb, UrS =  
## args$UrS, : fitted probabilities numerically 0 or 1 occurred

summary(gam)

##   
## Family: binomial   
## Link function: logit   
##   
## Formula:  
## pa ~ Mean\_durnal\_temp\_range + Ann\_.precip + Precip\_Seasonality +   
## Precip\_driest\_q + Precip\_coldest\_q + Hghest\_weekly\_rad +   
## Hghest\_weekly\_rad + Rad\_wettest\_q + Rad\_driest\_q + Rad\_driest\_q +   
## Highsets\_weekly\_moisture + Mean\_moisture\_warmest\_q  
##   
## Parametric coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 2.318e+01 2.899e+06 0 1  
## Mean\_durnal\_temp\_range 6.431e-01 1.245e+05 0 1  
## Ann\_.precip -9.269e-04 1.071e+03 0 1  
## Precip\_Seasonality -3.572e+00 8.725e+05 0 1  
## Precip\_driest\_q 1.076e-03 2.915e+03 0 1  
## Precip\_coldest\_q -7.075e-03 4.336e+03 0 1  
## Hghest\_weekly\_rad -1.698e-02 8.527e+03 0 1  
## Rad\_wettest\_q 1.560e-02 2.315e+03 0 1  
## Rad\_driest\_q 2.728e-02 4.346e+03 0 1  
## Highsets\_weekly\_moisture -4.240e-01 1.753e+06 0 1  
## Mean\_moisture\_warmest\_q 3.699e+00 1.282e+06 0 1  
##   
##   
## R-sq.(adj) = -Inf Deviance explained = -Inf%  
## UBRE = -0.8836 Scale est. = 1 n = 189

Let’s make some predictions and export them to a file

GAMpreds <- predict(predictors, gam, type = 'response')  
writeRaster(GAMpreds, filename = paste0(genus,"\_",species,"\_GAM.tif"), overwrite=TRUE)  
plot(GAMpreds, main=c(genus,species,'GAM/Binary'),col=warm(100), zlim=c(0,1))  
points(pres\_test, col='white', cex =.4, pch=3)



## MaxEnt

We need many more background points for MaxEnt and BRT than we needed for GAM. Let’s go ahead and generate those.

samp1 <- spsample(pol, 10000, type='random', iter=25)

## Warning in proj4string(obj): CRS object has comment, which is lost in output

#and get the cell numbers from the raster stack (right to left, up to down)  
cells <- cellFromXY(predictors, samp1)  
#and transform each of those to the center of its cell.  
background\_train <- xyFromCell(predictors, cells)  
#The backround data has too many NA values so to fix this:  
#First I get the predictor data associated with my points  
background\_train\_data<-extract(predictors,background\_train)  
#Then I remove all the points that don't have data  
complete<-complete.cases(background\_train\_data)  
background\_train<-background\_train[complete,]

Let’s go ahead and set a locations for java **This will obviously be specialized for your computer. Try to find the ‘home’ folder in java and specify the path below**

Sys.setenv(JAVA\_HOME='/Library/Java/JavaVirtualMachines/jdk-11.0.1.jdk/Contents/Home')

First we let the program know to start up maxent using the command maxent. After that, all we need to do is to make a model oject (me\_model), from the raster data and the presence training data.

maxent()

## Loading required namespace: rJava

## java.home option:

## JAVA\_HOME environment variable: /Library/Java/JavaVirtualMachines/jdk-11.0.1.jdk/Contents/Home

## Warning in fun(libname, pkgname): Java home setting is INVALID, it will be ignored.  
## Please do NOT set it unless you want to override system settings.

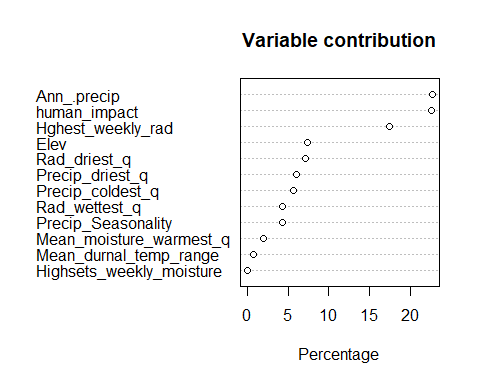
## This is MaxEnt version 3.4.3

me\_model <- maxent(predictors, pres\_train, a=background\_train)

## Warning in .local(x, p, ...): 3 (6.25%) of the presence points have NA predictor  
## values

## This is MaxEnt version 3.4.3

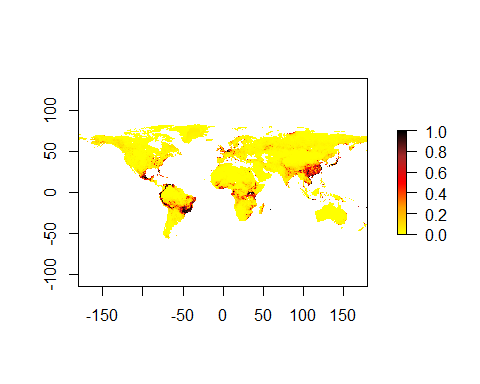
#and plot the models most important layers  
par(mfrow=c(1,1))  
plot(me\_model)

 Let’s go ahead and make some predictions

MEpreds<-predict(predictors, me\_model, type='response')

## This is MaxEnt version 3.4.3

writeRaster(MEpreds, filename=paste0(genus,"\_",species,"\_ME.tif"),overwrite=TRUE)  
#and plot  
plot(MEpreds, col=warm(100), zlim=c(0,1))



## Boosted regression trees

We need to prepare data for BRT in much the same way that we did for GAM, with the exception that we will need a lot more background data. We can use the 10,000 points that we already generated for ME

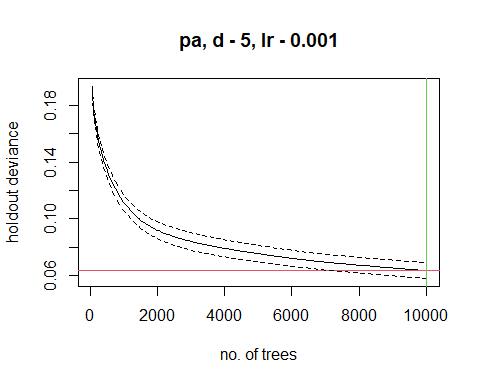
#let's get the data from our predictors  
bg\_train\_data<-extract(predictors,background\_train)  
#and bind a column of 0 to the end of it  
bg\_train\_data<-cbind(bg\_train\_data,0)  
#and convert it to a data frame  
bg\_train\_data<-as.data.frame(bg\_train\_data)  
#and then combine it withe the presence training data  
pres\_train\_data<-as.data.frame(pres\_train\_data)  
BRT\_data<-rbind(pres\_train\_data, bg\_train\_data)  
colnames(BRT\_data)<-c(names(predictors),"pa")

And now we can train a model using the first X columns of train\_data and the 6th as the response. Let’s start with tree complexity of 5 and learning rate of 0.001

sdm.tc5.lr001 <- gbm.step(data=BRT\_data, gbm.x = 1:nlayers(predictors), gbm.y = ncol(BRT\_data), family = "bernoulli", tree.complexity = 5, learning.rate = 0.001, bag.fraction = 0.5)

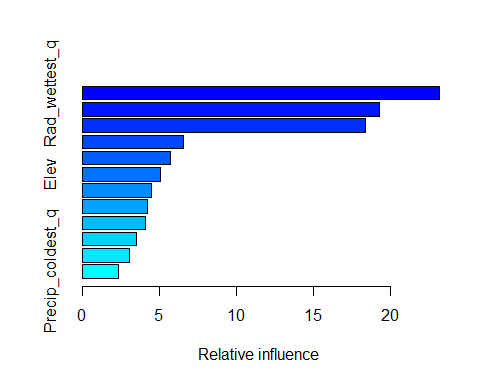
##   
##   
## GBM STEP - version 2.9   
##   
## Performing cross-validation optimisation of a boosted regression tree model   
## for pa and using a family of bernoulli   
## Using 4323 observations and 12 predictors   
## creating 10 initial models of 50 trees   
##   
## folds are stratified by prevalence   
## total mean deviance = 0.2148   
## tolerance is fixed at 2e-04   
## ntrees resid. dev.   
## 50 0.1904   
## now adding trees...   
## 100 0.1767   
## 150 0.1676   
## 200 0.1605   
## 250 0.1545   
## 300 0.1495   
## 350 0.1451   
## 400 0.1412   
## 450 0.1376   
## 500 0.1344   
## 550 0.1314   
## 600 0.1286   
## 650 0.126   
## 700 0.1235   
## 750 0.1212   
## 800 0.119   
## 850 0.1169   
## 900 0.115   
## 950 0.1132   
## 1000 0.1115   
## 1050 0.1099   
## 1100 0.1084   
## 1150 0.107   
## 1200 0.1057   
## 1250 0.1045   
## 1300 0.1033   
## 1350 0.1022   
## 1400 0.1011   
## 1450 0.1001   
## 1500 0.0992   
## 1550 0.0983   
## 1600 0.0975   
## 1650 0.0967   
## 1700 0.0959   
## 1750 0.0952   
## 1800 0.0945   
## 1850 0.0938   
## 1900 0.0932   
## 1950 0.0926   
## 2000 0.0921   
## 2050 0.0916   
## 2100 0.091   
## 2150 0.0905   
## 2200 0.0901   
## 2250 0.0896   
## 2300 0.0892   
## 2350 0.0887   
## 2400 0.0883   
## 2450 0.0879   
## 2500 0.0875   
## 2550 0.0872   
## 2600 0.0868   
## 2650 0.0864   
## 2700 0.0861   
## 2750 0.0857   
## 2800 0.0854   
## 2850 0.0851   
## 2900 0.0848   
## 2950 0.0845   
## 3000 0.0841   
## 3050 0.0838   
## 3100 0.0835   
## 3150 0.0832   
## 3200 0.083   
## 3250 0.0827   
## 3300 0.0825   
## 3350 0.0822   
## 3400 0.0819   
## 3450 0.0817   
## 3500 0.0815   
## 3550 0.0812   
## 3600 0.081   
## 3650 0.0807   
## 3700 0.0805   
## 3750 0.0803   
## 3800 0.0801   
## 3850 0.0798   
## 3900 0.0797   
## 3950 0.0795   
## 4000 0.0793   
## 4050 0.079   
## 4100 0.0788   
## 4150 0.0786   
## 4200 0.0784   
## 4250 0.0782   
## 4300 0.078   
## 4350 0.0778   
## 4400 0.0777   
## 4450 0.0775   
## 4500 0.0773   
## 4550 0.0771   
## 4600 0.0769   
## 4650 0.0768   
## 4700 0.0766   
## 4750 0.0765   
## 4800 0.0763   
## 4850 0.0761   
## 4900 0.0759   
## 4950 0.0758   
## 5000 0.0756   
## 5050 0.0753   
## 5100 0.0751   
## 5150 0.0749   
## 5200 0.0748   
## 5250 0.0746   
## 5300 0.0745   
## 5350 0.0743   
## 5400 0.0742   
## 5450 0.074   
## 5500 0.0737   
## 5550 0.0736   
## 5600 0.0735   
## 5650 0.0733   
## 5700 0.0732   
## 5750 0.0731   
## 5800 0.0729   
## 5850 0.0727   
## 5900 0.0726   
## 5950 0.0724   
## 6000 0.0723   
## 6050 0.0721   
## 6100 0.072   
## 6150 0.0718   
## 6200 0.0717   
## 6250 0.0715   
## 6300 0.0714   
## 6350 0.0713   
## 6400 0.0711   
## 6450 0.071   
## 6500 0.0708   
## 6550 0.0707   
## 6600 0.0705   
## 6650 0.0704   
## 6700 0.0703   
## 6750 0.0701   
## 6800 0.07   
## 6850 0.0699   
## 6900 0.0697   
## 6950 0.0696   
## 7000 0.0695   
## 7050 0.0694   
## 7100 0.0693   
## 7150 0.0692   
## 7200 0.0691   
## 7250 0.069   
## 7300 0.0689   
## 7350 0.0687   
## 7400 0.0686   
## 7450 0.0686   
## 7500 0.0685   
## 7550 0.0683   
## 7600 0.0683   
## 7650 0.0681   
## 7700 0.068   
## 7750 0.0678   
## 7800 0.0678   
## 7850 0.0677   
## 7900 0.0675   
## 7950 0.0674   
## 8000 0.0674   
## 8050 0.0673   
## 8100 0.0671   
## 8150 0.0671   
## 8200 0.067   
## 8250 0.0669   
## 8300 0.0668   
## 8350 0.0667   
## 8400 0.0666   
## 8450 0.0665   
## 8500 0.0664   
## 8550 0.0663   
## 8600 0.0662   
## 8650 0.0661   
## 8700 0.066   
## 8750 0.0659   
## 8800 0.0658   
## 8850 0.0657   
## 8900 0.0656   
## 8950 0.0655   
## 9000 0.0654   
## 9050 0.0653   
## 9100 0.0652   
## 9150 0.0651   
## 9200 0.065   
## 9250 0.065   
## 9300 0.0649   
## 9350 0.0647   
## 9400 0.0646   
## 9450 0.0646   
## 9500 0.0645   
## 9550 0.0644   
## 9600 0.0643   
## 9650 0.0642   
## 9700 0.0641   
## 9750 0.064   
## 9800 0.0639   
## 9850 0.0638   
## 9900 0.0638   
## 9950 0.0637   
## 10000 0.0636

## fitting final gbm model with a fixed number of 10000 trees for pa



##   
## mean total deviance = 0.215   
## mean residual deviance = 0.027   
##   
## estimated cv deviance = 0.064 ; se = 0.006   
##   
## training data correlation = 0.936   
## cv correlation = 0.804 ; se = 0.019   
##   
## training data AUC score = 1   
## cv AUC score = 0.985 ; se = 0.007   
##   
## elapsed time - 0.16 minutes   
##   
## ########### warning ##########   
##   
## maximum tree limit reached - results may not be optimal   
## - refit with faster learning rate or increase maximum number of trees

summary(sdm.tc5.lr001)



## var rel.inf  
## Rad\_wettest\_q Rad\_wettest\_q 23.183674  
## human\_impact human\_impact 19.256334  
## Hghest\_weekly\_rad Hghest\_weekly\_rad 18.390458  
## Precip\_Seasonality Precip\_Seasonality 6.580573  
## Precip\_driest\_q Precip\_driest\_q 5.692429  
## Elev Elev 5.074596  
## Mean\_moisture\_warmest\_q Mean\_moisture\_warmest\_q 4.474409  
## Ann\_.precip Ann\_.precip 4.249135  
## Mean\_durnal\_temp\_range Mean\_durnal\_temp\_range 4.114975  
## Rad\_driest\_q Rad\_driest\_q 3.526498  
## Highsets\_weekly\_moisture Highsets\_weekly\_moisture 3.094648  
## Precip\_coldest\_q Precip\_coldest\_q 2.362273

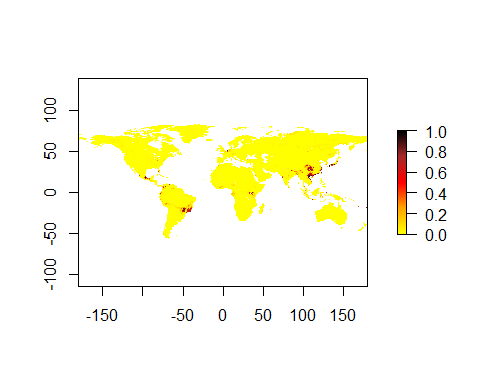
**Note: you may want to try different combinations! If your trees are converging too slowly, raise the tree complexity by 1 or two, and back the learning rate down. On the other hand of your holdout deviance drops very quickly and slowly starts to rise, you are overfitting. Drop the tree complexity and raise the learning rate.**

Let’s make predictions and save them

BRTpreds<-predict(predictors, sdm.tc5.lr001, type='response')

## Using 10000 trees...  
##   
## Using 10000 trees...  
##   
## Using 10000 trees...  
##   
## Using 10000 trees...

writeRaster(BRTpreds, filename=paste0(genus,"\_", species,"\_BRT.tif"), overwrite=TRUE)  
#and plot  
plot(BRTpreds, col=warm(100), zlim=c(0,1))



## Evaluation

We want to generate the following metrics for each of the three models: AUC, COR, maximum Kappa, TRS, and it wouldn’t kill us to have a Boyce graph either.

#Absence Testing Data First we’ll use the pres\_test data to generate absence test data. This time we want about the same number of points for both. To do that, we’ll generate 4x the number of absence points as presence points and chop it to size.

pres\_test\_SPDF<-SpatialPoints(pres\_test)  
data("wrld\_simpl")  
crs(pres\_test\_SPDF) <- crs(wrld\_simpl)  
#now we are going to make circles of about a degree (110000 meters at the equator). I'm working in a relatively small area, but if your data are widespread, you can increase this by changing d.  
x <- circles(pres\_test\_SPDF, d=dist, lonlat=TRUE)  
#and convert those into polygons  
pol <- polygons(x)  
#and draw a number of samples from that...because   
samp1 <- spsample(pol, 4\*length(pres\_test), type='random', iter=25)

## Warning in proj4string(obj): CRS object has comment, which is lost in output

#and get the cell numbers from the raster stack (right to left, up to down)  
cells <- cellFromXY(predictors, samp1)  
#and transform each of those to the center of its cell.  
abs\_test <- xyFromCell(predictors, cells)  
#You'll get a warning saying that your CRS object has lost a comment. This is unimportant and can be ignored.

GAM evaluation

p<-extract(GAMpreds,pres\_test)  
a<-extract(GAMpreds,abs\_test)  
#And let's get rid of nasty NA values and shrink a to the size of p  
p<-p[!is.na(p)]  
a<-a[!is.na(a)]  
a<-a[1:length(p)]  
#Let's look at the shape of these data  
#lets weld all the data together  
all\_vals<-c(p,a)  
e<-evaluate(p=p,a=a)  
AUC\_GAM<-e@auc  
COR\_GAM<-e@cor  
pa<-c(replicate(length(p),1),replicate(length(a),0))  
kappaGAM<-ecospat.max.kappa(all\_vals,pa)  
TSS\_GAM<-ecospat.max.tss(all\_vals,pa)

## Warning in max(table$tss, na.rm = TRUE): no non-missing arguments to max;  
## returning -Inf

print(paste('Max kappa: ', kappaGAM[2] ))

## [1] "Max kappa: 0"

print(paste('TSS:', TSS\_GAM[[2]]))

## [1] "TSS: -Inf"

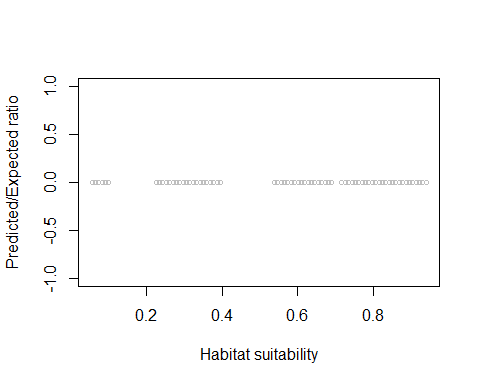
e

## class : ModelEvaluation   
## n presences : 23   
## n absences : 23   
## AUC : 0.4584121   
## cor : 0.04416923   
## max TPR+TNR at : 0.9999

And let’s go ahead and estimate the Boyce Index

ecospat.boyce(fit=GAMpreds,pres\_test,nclass=0,PEplot = TRUE)

## Warning in if (class(obs) == "data.frame" | class(obs) == "matrix") {: the  
## condition has length > 1 and only the first element will be used



## $F.ratio  
## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## [39] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
##   
## $Spearman.cor  
## [1] NA  
##   
## $HS  
## [1] 0.05484367 0.06379779 0.07275190 0.08170601 0.09066012 0.09961424  
## [7] 0.22497181 0.23392592 0.24288004 0.25183415 0.26078826 0.26974237  
## [13] 0.27869649 0.28765060 0.29660471 0.30555882 0.31451294 0.32346705  
## [19] 0.33242116 0.34137528 0.35032939 0.35928350 0.36823761 0.37719173  
## [25] 0.38614584 0.39509995 0.53836575 0.54731986 0.55627398 0.56522809  
## [31] 0.57418220 0.58313631 0.59209043 0.60104454 0.60999865 0.61895276  
## [37] 0.62790688 0.63686099 0.64581510 0.65476922 0.66372333 0.67267744  
## [43] 0.68163155 0.69058567 0.71744800 0.72640212 0.73535623 0.74431034  
## [49] 0.75326445 0.76221857 0.77117268 0.78012679 0.78908090 0.79803502  
## [55] 0.80698913 0.81594324 0.82489735 0.83385147 0.84280558 0.85175969  
## [61] 0.86071380 0.86966792 0.87862203 0.88757614 0.89653025 0.90548437  
## [67] 0.91443848 0.92339259 0.93234670 0.94130082

ME Evaluation

p<-extract(MEpreds,pres\_test)  
a<-extract(MEpreds,abs\_test)  
#And let's get rid of nasty NA values and shrink a to the size of p  
p<-p[!is.na(p)]  
a<-a[!is.na(a)]  
a<-a[1:length(p)]  
#Let's look at the shape of these data  
#lets weld all the data together  
all\_vals<-c(p,a)  
e<-evaluate(p=p,a=a)  
AUC\_ME<-e@auc  
COR\_ME<-e@cor  
pa<-c(replicate(length(p),1),replicate(length(a),0))  
kappaME<-ecospat.max.kappa(all\_vals,pa)  
TSS\_ME<-ecospat.max.tss(all\_vals,pa)  
print(paste('Max kappa: ', kappaME[2] ))

## [1] "Max kappa: 0.826086956521739"

print(paste('TSS:', TSS\_ME[[2]]))

## [1] "TSS: 0.826086956521739"

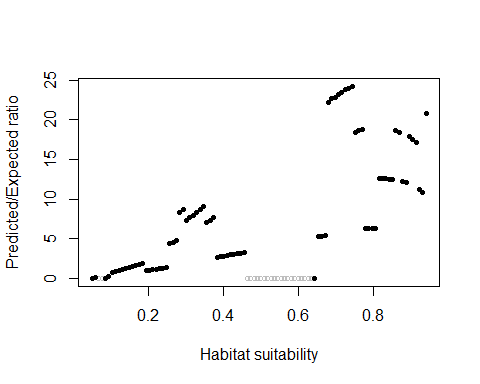
e

## class : ModelEvaluation   
## n presences : 23   
## n absences : 23   
## AUC : 0.9508507   
## cor : 0.7576243   
## max TPR+TNR at : 0.2481069

And let’s go ahead and estimate the Boyce Index

ecospat.boyce(fit=MEpreds,pres\_test,nclass=0,PEplot = TRUE)

## Warning in if (class(obs) == "data.frame" | class(obs) == "matrix") {: the  
## condition has length > 1 and only the first element will be used



## $F.ratio  
## [1] 0.04752642 0.09883926 0.00000000 0.00000000 0.00000000 0.29571149  
## [7] 0.70539554 0.82309278 0.94611492 1.07543412 1.20531789 1.32991864  
## [13] 1.46026639 1.59351673 1.72334164 1.84281404 0.98886660 1.05427151  
## [19] 1.11654628 1.17904522 1.24157331 1.31179361 1.38162906 4.35364370  
## [25] 4.55789089 4.75626975 8.30805855 8.74580123 7.33158049 7.65300417  
## [31] 7.99251778 8.32097742 8.69437524 9.06248387 7.08095778 7.31210064  
## [37] 7.65061614 2.63604973 2.72357731 2.78914867 2.86597297 2.95528731  
## [43] 3.04497494 3.12499444 3.20612839 3.26813436 0.00000000 0.00000000  
## [49] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000  
## [55] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000  
## [61] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000  
## [67] 0.00000000 5.25853065 5.32960873 5.44319690 22.14827297 22.66192085  
## [73] 22.88109952 23.25998448 23.52783041 23.90317469 23.96051173 24.23829604  
## [79] 18.47763223 18.70315673 18.74996663 6.29727094 6.32207032 6.29198208  
## [85] 6.31319101 12.59454189 12.61575077 12.64769837 12.54183036 12.54533073  
## [91] 18.72393224 18.44729959 12.22129418 12.14206375 17.87036311 17.52205821  
## [97] 17.09555781 11.16213558 10.78194291 20.78609017  
##   
## $Spearman.cor  
## [1] 0.792  
##   
## $HS  
## [1] 0.05000005 0.05900005 0.06800005 0.07700005 0.08600005 0.09500005  
## [7] 0.10400005 0.11300005 0.12200005 0.13100004 0.14000004 0.14900004  
## [13] 0.15800004 0.16700004 0.17600004 0.18500004 0.19400004 0.20300004  
## [19] 0.21200004 0.22100004 0.23000004 0.23900004 0.24800004 0.25700004  
## [25] 0.26600004 0.27500004 0.28400004 0.29300004 0.30200004 0.31100004  
## [31] 0.32000004 0.32900003 0.33800003 0.34700003 0.35600003 0.36500003  
## [37] 0.37400003 0.38300003 0.39200003 0.40100003 0.41000003 0.41900003  
## [43] 0.42800003 0.43700003 0.44600003 0.45500003 0.46400003 0.47300003  
## [49] 0.48200003 0.49100003 0.50000003 0.50900003 0.51800002 0.52700002  
## [55] 0.53600002 0.54500002 0.55400002 0.56300002 0.57200002 0.58100002  
## [61] 0.59000002 0.59900002 0.60800002 0.61700002 0.62600002 0.63500002  
## [67] 0.64400002 0.65300002 0.66200002 0.67100002 0.68000002 0.68900002  
## [73] 0.69800002 0.70700002 0.71600001 0.72500001 0.73400001 0.74300001  
## [79] 0.75200001 0.76100001 0.77000001 0.77900001 0.78800001 0.79700001  
## [85] 0.80600001 0.81500001 0.82400001 0.83300001 0.84200001 0.85100001  
## [91] 0.86000001 0.86900001 0.87800001 0.88700001 0.89600001 0.90500000  
## [97] 0.91400000 0.92300000 0.93200000 0.94100000

BRT Evaluation

p<-extract(BRTpreds,pres\_test)  
a<-extract(BRTpreds,abs\_test)  
#And let's get rid of nasty NA values and shrink a to the size of p  
p<-p[!is.na(p)]  
a<-a[!is.na(a)]  
a<-a[1:length(p)]  
#Let's look at the shape of these data  
#lets weld all the data together  
all\_vals<-c(p,a)  
e<-evaluate(p=p,a=a)  
AUC\_BRT<-e@auc  
COR\_BRT<-e@cor  
pa<-c(replicate(length(p),1),replicate(length(a),0))  
kappaBRT<-ecospat.max.kappa(all\_vals,pa)  
TSS\_BRT<-ecospat.max.tss(all\_vals,pa)  
print(paste('Max kappa: ', kappaBRT[2] ))

## [1] "Max kappa: 0.608695652173913"

print(paste('TSS:', TSS\_BRT[[2]]))

## [1] "TSS: 0.608695652173913"

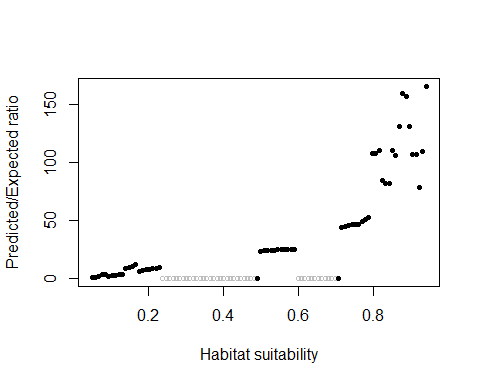
e

## class : ModelEvaluation   
## n presences : 23   
## n absences : 23   
## AUC : 0.9111531   
## cor : 0.6025518   
## max TPR+TNR at : 0.00342633

And let’s go ahead and estimate the Boyce Index

ecospat.boyce(fit=BRTpreds,pres\_test,nclass=0,PEplot = TRUE)

## Warning in if (class(obs) == "data.frame" | class(obs) == "matrix") {: the  
## condition has length > 1 and only the first element will be used

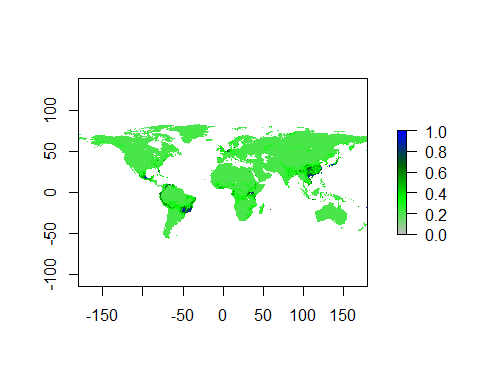


## $F.ratio  
## [1] 0.3690409 1.0083884 1.8322227 3.7921755 3.1717415 1.9306726  
## [7] 2.4867183 2.8302430 3.2911056 3.7241027 8.4081407 9.5597448  
## [13] 10.6972680 11.6874467 6.3632390 6.7859179 7.2994349 7.7101063  
## [19] 8.1995476 8.5847823 9.0116119 0.0000000 0.0000000 0.0000000  
## [25] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## [31] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## [37] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## [43] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## [49] 0.0000000 0.0000000 23.3627443 23.7830265 24.4027796 24.3763124  
## [55] 24.4558868 24.9444617 24.9721778 25.0556968 25.1679283 25.0277951  
## [61] 25.2811699 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## [67] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## [73] 0.0000000 0.0000000 43.9823092 44.5489792 45.8205097 46.1972456  
## [79] 46.1498152 46.7741103 48.9650545 50.6763472 52.6962720 108.1827196  
## [85] 108.3130602 110.8506042 84.2811000 82.0254015 82.1253106 110.3065521  
## [91] 106.5163981 131.4325146 159.9641281 156.9845867 131.2789720 107.4072162  
## [97] 106.6427521 78.9518501 109.3672019 165.8668635  
##   
## $Spearman.cor  
## [1] 0.929  
##   
## $HS  
## [1] 0.05034320 0.05933264 0.06832207 0.07731151 0.08630094 0.09529037  
## [7] 0.10427981 0.11326924 0.12225867 0.13124811 0.14023754 0.14922698  
## [13] 0.15821641 0.16720584 0.17619528 0.18518471 0.19417415 0.20316358  
## [19] 0.21215301 0.22114245 0.23013188 0.23912131 0.24811075 0.25710018  
## [25] 0.26608962 0.27507905 0.28406848 0.29305792 0.30204735 0.31103679  
## [31] 0.32002622 0.32901565 0.33800509 0.34699452 0.35598395 0.36497339  
## [37] 0.37396282 0.38295226 0.39194169 0.40093112 0.40992056 0.41890999  
## [43] 0.42789942 0.43688886 0.44587829 0.45486773 0.46385716 0.47284659  
## [49] 0.48183603 0.49082546 0.49981490 0.50880433 0.51779376 0.52678320  
## [55] 0.53577263 0.54476206 0.55375150 0.56274093 0.57173037 0.58071980  
## [61] 0.58970923 0.59869867 0.60768810 0.61667754 0.62566697 0.63465640  
## [67] 0.64364584 0.65263527 0.66162470 0.67061414 0.67960357 0.68859301  
## [73] 0.69758244 0.70657187 0.71556131 0.72455074 0.73354018 0.74252961  
## [79] 0.75151904 0.76050848 0.76949791 0.77848734 0.78747678 0.79646621  
## [85] 0.80545565 0.81444508 0.82343451 0.83242395 0.84141338 0.85040281  
## [91] 0.85939225 0.86838168 0.87737112 0.88636055 0.89534998 0.90433942  
## [97] 0.91332885 0.92231829 0.93130772 0.94029715

# Making the ensemble and evaluation

The ensemble is simply the average of GAM, ME, and BRT predictions weighted by AUC.

ENSpreds<-(GAMpreds\*AUC\_GAM+MEpreds\*AUC\_ME+BRTpreds\*AUC\_BRT)/(AUC\_GAM+AUC\_ME+AUC\_BRT)  
writeRaster(ENSpreds, filename=paste0(genus,"\_",species,"\_ENS.tif"), overwrite=TRUE)  
plot(ENSpreds, col=cool(100), zlim=c(0,1))

 And let’s evaluate

p<-extract(ENSpreds,pres\_test)  
a<-extract(ENSpreds,abs\_test)  
#And let's get rid of nasty NA values and shrink a to the size of p  
p<-p[!is.na(p)]  
a<-a[!is.na(a)]  
a<-a[1:length(p)]  
#Let's look at the shape of these data  
#lets weld all the data together  
all\_vals<-c(p,a)  
e<-evaluate(p=p,a=a)  
AUC\_ENS<-e@auc  
COR\_ENS<-e@cor  
pa<-c(replicate(length(p),1),replicate(length(a),0))  
kappaENS<-ecospat.max.kappa(all\_vals,pa)  
TSS\_ENS<-ecospat.max.tss(all\_vals,pa)  
print(paste('Max kappa: ', kappaENS[2] ))

## [1] "Max kappa: 0.782608695652174"

print(paste('TSS:', TSS\_ENS[[2]]))

## [1] "TSS: 0.782608695652174"

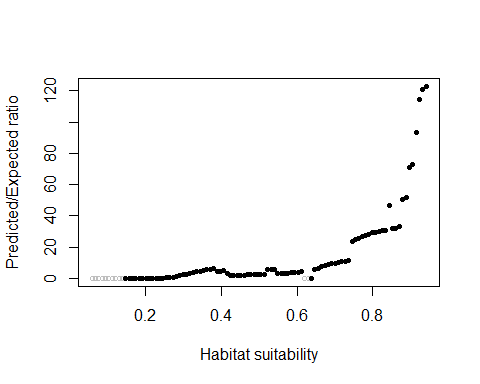
e

## class : ModelEvaluation   
## n presences : 23   
## n absences : 23   
## AUC : 0.9432892   
## cor : 0.7006542   
## max TPR+TNR at : 0.3336248

and the Boyce ploy for the ensemble

ecospat.boyce(fit=ENSpreds,pres\_test,nclass=0,PEplot = TRUE)

## Warning in if (class(obs) == "data.frame" | class(obs) == "matrix") {: the  
## condition has length > 1 and only the first element will be used



## $F.ratio  
## [1] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000  
## [6] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000  
## [11] 0.00000000 0.07212691 0.05666866 0.05148914 0.04906597  
## [16] 0.04758599 0.04657926 0.09167242 0.09051389 0.13435667  
## [21] 0.13318811 0.13216133 0.41424107 0.53266614 0.78259973  
## [26] 1.02121774 2.10392421 2.50859005 2.32461511 3.31723934  
## [31] 3.72459647 4.13751105 4.56566855 5.00912900 5.47475397  
## [36] 5.97261759 6.47357567 4.18580084 4.46788682 4.75761219  
## [41] 3.34722764 1.76523406 1.86483239 1.96562533 2.07257101  
## [46] 2.17107419 2.26470778 2.34995399 2.44612103 2.52925501  
## [51] 2.62988065 2.74352539 5.67262998 5.83613607 6.00854431  
## [56] 3.07833995 3.15703891 3.21438215 3.34299569 3.50077259  
## [61] 3.74270774 3.98350939 4.25662121 0.00000000 0.00000000  
## [66] 0.00000000 5.94261237 6.52204295 7.43465432 8.17568570  
## [71] 8.77585318 9.26420445 9.62937446 10.05141324 10.62645863  
## [76] 11.02794897 11.53153412 23.80822034 24.84793809 25.77403670  
## [81] 26.55045481 27.50913097 28.07615240 29.05618617 29.37903268  
## [86] 30.00662216 30.39210277 30.74549932 46.59632343 31.74429379  
## [91] 32.19908309 33.24698225 50.69539850 51.70619632 70.95488556  
## [96] 72.97064935 93.02549669 114.57073917 120.61695886 122.81398907  
##   
## $Spearman.cor  
## [1] 0.902  
##   
## $HS  
## [1] 0.05804051 0.06696147 0.07588242 0.08480338 0.09372433 0.10264528  
## [7] 0.11156624 0.12048719 0.12940815 0.13832910 0.14725006 0.15617101  
## [13] 0.16509197 0.17401292 0.18293387 0.19185483 0.20077578 0.20969674  
## [19] 0.21861769 0.22753865 0.23645960 0.24538056 0.25430151 0.26322246  
## [25] 0.27214342 0.28106437 0.28998533 0.29890628 0.30782724 0.31674819  
## [31] 0.32566915 0.33459010 0.34351105 0.35243201 0.36135296 0.37027392  
## [37] 0.37919487 0.38811583 0.39703678 0.40595774 0.41487869 0.42379964  
## [43] 0.43272060 0.44164155 0.45056251 0.45948346 0.46840442 0.47732537  
## [49] 0.48624633 0.49516728 0.50408823 0.51300919 0.52193014 0.53085110  
## [55] 0.53977205 0.54869301 0.55761396 0.56653492 0.57545587 0.58437682  
## [61] 0.59329778 0.60221873 0.61113969 0.62006064 0.62898160 0.63790255  
## [67] 0.64682351 0.65574446 0.66466541 0.67358637 0.68250732 0.69142828  
## [73] 0.70034923 0.70927019 0.71819114 0.72711210 0.73603305 0.74495400  
## [79] 0.75387496 0.76279591 0.77171687 0.78063782 0.78955878 0.79847973  
## [85] 0.80740069 0.81632164 0.82524260 0.83416355 0.84308450 0.85200546  
## [91] 0.86092641 0.86984737 0.87876832 0.88768928 0.89661023 0.90553119  
## [97] 0.91445214 0.92337309 0.93229405 0.94121500

and finally let’s make a table of evaluation metrics

#Let's go in this order of columns, left to right: AUC, COR, Kappa, TSS  
eGAM<-c(AUC\_GAM,COR\_GAM,kappaGAM[2], TSS\_GAM[[2]])  
eME<-c(AUC\_ME, COR\_ME, kappaME[2],TSS\_ME[[2]])  
eBRT<-c(AUC\_BRT, COR\_BRT, kappaBRT[2],TSS\_BRT[[2]])  
eENS<-c(AUC\_ENS, COR\_ENS, kappaENS[2], TSS\_ENS[[2]])  
all\_evals<-rbind(eGAM,eME,eBRT,eENS)  
colnames(all\_evals)<-c("AUC", "COR","MaxKappa","TSS")  
rownames(all\_evals)<-c("GAM","MaxEnt", "BRT", "Ensemble")  
write.csv(all\_evals, file=paste0(genus,"\_",species, '\_eval.csv'))