

Triatoma predicted distribution

Lauren Ash

May 24, 2016

```
library(sp)
library(raster)
library(dismo)
library(gbm)
library(maptools)
require(pscl)
library(rgeos)
```

```
dimPoints<-read.csv("T_dimidiata.csv") #122 points
infPoints<-read.csv("T_infestans.csv") #67 points
proPoints<-read.csv("T_protracta.csv") #27 points
sanPoints<-read.csv("T_sanguisuga.csv") #35 points
sorPoints<-read.csv("T_sordida.csv") #58 points
```

```
# For T. sanguisuga and T. protracta, please include Canada and Mexico/Central America
# For T. dimidiata, please include North and South America
# For T. sordida and T. infestans, please include Central America.
```

```
coordinates(dimPoints)<-~Lon+Lat
coordinates(infPoints)<-~Lon+Lat
coordinates(proPoints)<-~Lon+Lat
coordinates(sanPoints)<-~Lon+Lat
coordinates(sorPoints)<-~Lon+Lat
```

```
bioclim <- getData("worldclim", var = "bio", res = 10)
stack(bioclim)
```

```
## class      : RasterStack
## dimensions  : 900, 2160, 1944000, 19  (nrow, ncol, ncell, nlayers)
## resolution  : 0.1666667, 0.1666667  (x, y)
## extent     : -180, 180, -60, 90  (xmin, xmax, ymin, ymax)
## coord. ref. : +proj=longlat +datum=WGS84 +ellps=WGS84 +towgs84=0,0,0
## names       : bio1, bio2, bio3, bio4, bio5, bio6, bio7, bio8, bio9, bio10, bio11, bio12, bi
## min values  : -269, 9, 8, 72, -59, -547, 53, -251, -450, -97, -488, 0,
## max values  : 314, 211, 95, 22673, 489, 258, 725, 375, 364, 380, 289, 9916, 2
```

```
# extent(left long, right long, lowest lat, highest lat)
layers <- crop(bioclim, extent(-170, -33, -60, 75))
#plot(layers$bio1) #north, central, and south america
#points(dimPoints)
```

```
data("wrld_simpl")
ws1<-rasterize(wrld_simpl,layers)
values(layers)[is.na(values(layers))]<-0
layers1<-mask(layers,ws1)
```

```

for (j in 1:25) {
  ##divide data into random groups
  group <- dismo::kfold(dimPoints, 5)
  pres_train <- dimPoints[group != 1, ]
  pres_test <- dimPoints[group == 1, ]

  #extract data for presence data
  presvals <- raster::extract(layers1, pres_train)
  presvals[is.na(presvals)]<-0
  presvals2 <- raster::extract(layers1, pres_test)
  presvals2[is.na(presvals2)]<-0

  #####back ground points from areas nearby - PERHAPS A BETTER METHOD FOR BACKGROUND SAMPLING
  # circles with a radius of 50 km
  #x <- circles(dimPoints, d=500000, lonlat=TRUE)
  #pol <- gUnaryUnion(x@polygons)
  #samp1 <- spsample(pol, 300, type="random", iter=25)
  #insea<-extract(ws1,samp1)
  #backg<-samp1[!is.na(insea)]

  ##### create background set with 500 random points
  backg <- dismo::randomPoints(layers1, n=500, warn=0)
  colnames(backg) = c('lon', 'lat')
  group <- dismo::kfold(backg, 5)
  backg_train <- backg[group != 1, ]
  backg_test <- backg[group == 1, ]

  #extract values for background data
  absvals <- raster::extract(layers1, backg_train)
  absvals[is.na(absvals)]<-0
  absvals2 <- raster::extract(layers1, backg_test)
  absvals2[is.na(absvals2)]<-0

  ### create datatable to run regression on
  pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
  sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
  pb2 <- c(rep(1, nrow(pres_test)), rep(0, nrow(absvals2)))
  testdata <- data.frame(cbind(pb2, rbind(presvals2, absvals2)))
}

```

```

#run MAXENT
xm <- dismo::maxent(x=layers1, removeDuplicatess=T, p=pres_train,a=backg_train)

```

```

## Warning in .local(x, p, ...): 4 (4.17%) of the presence points have NA
## predictor values

```

```

## Loading required namespace: rJava

```

```

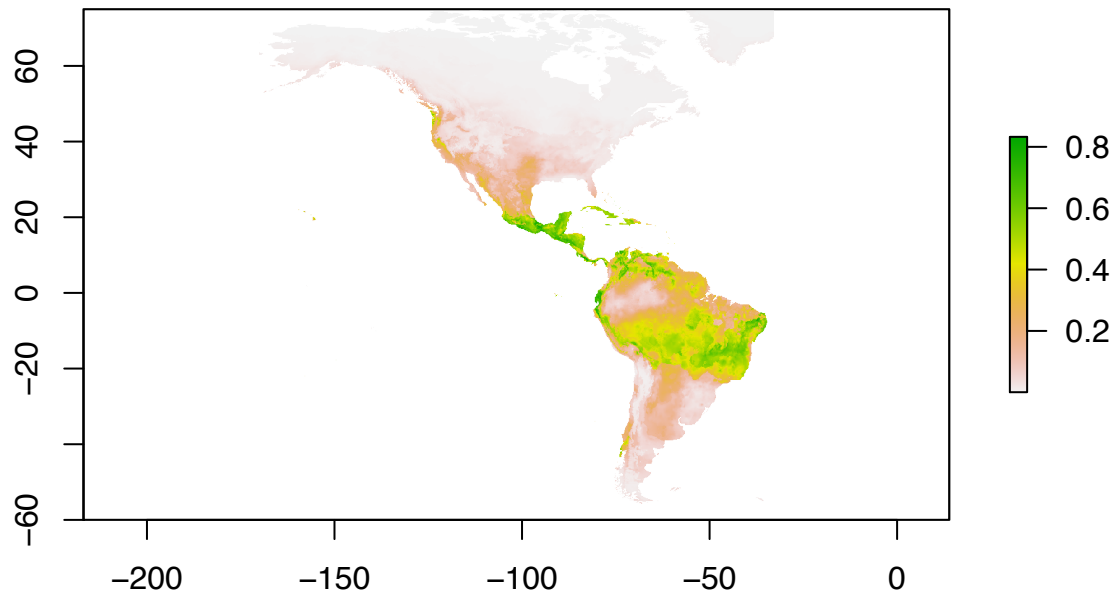
#evaluate
e2 <- evaluate(pres_test, backg_test, xm, layers1)
#e2

```

```
#predict
px <- predict(layers1, xm, progress="")

#plot maxent
plot(px, main="T. dimidiata predicted distribution")
```

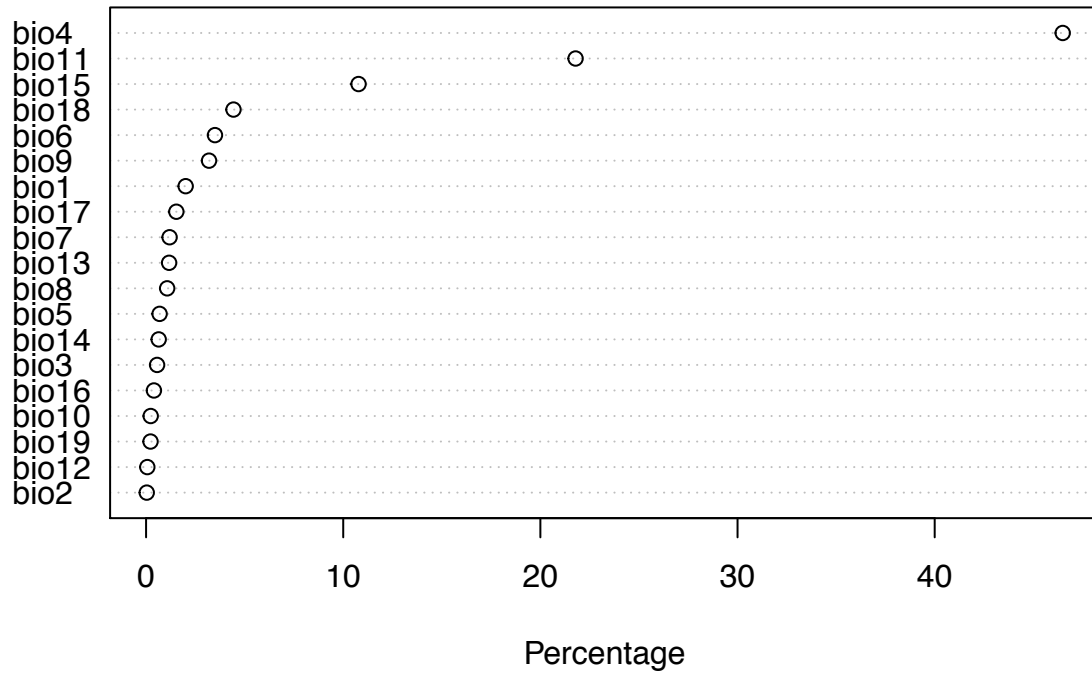
T. dimidiata predicted distribution



```
#points(backg_train)

# plot showing importance of each variable
plot(xm)
```

Variable contribution



Mean temperature of coldest quarter ~50% variable contribution

Temperature seasonality ~25% variable contribution

Annual mean temperature ~7% variable contribution