# A brief tutorial on runing Maxent in R

*Xiao Feng, Cassondra Walker, Fikirte Gebresenbet*

*July 4, 2017*

## Contents



## 1. Set up the working environment

### 1.1 Load packages

Running Maxent in R requires several packages. Specifically, the "dismo" package, which contains the functions for maximum entropy species distribution modeling. However, the "rgbif" package gives access to the GBIF database for occurrence location downloads, (https://cran.r-project.org/web/packages/rgbif/rgbif.pdf), the "raster" package provides functions for working with gridded data ( https://cran.r-project.org/web/packages/raster/raster.pdf), the "rgeos" package provides the ability to manipulate spatial data (https://cran.rstudio.com/web/packages/rgeos/rgeos.pdf), and the "knitr" package allows for report generation and easy display of modeling output (https://cran.r-project.org/web/packages/knitr/knitr.pdf), which are needed to complete the modeling process.

**Thread 1**

```
library("dismo")
library("raster")
#library("rgbif")
library("knitr")
require("rgeos")
```

```
## Warning: package 'rgeos' was built under R version 3.4.2
```

If you are using a Mac machine, an additional step may be needed #####Thread 2

```
dyn.load('/Library/Java/JavaVirtualMachines/jdk1.8.0_144.jdk/Contents/Home/jre/lib/server/libjvm.dylib'
require("rJava")
```

**Thread 3**

```
knitr::opts_knit$set(root.dir = '/Users/iel82user/Google Drive/1_osu_lab/projects/2017_7_workshop_enm_R/
```

### 1.2 Set up the Maxent path

In order for Maxent to work properly in R, the .jar file associated with Maxent needs to be accessible.

**Thread 4**

## 2. Prepare data input

### 2.1 Load environmental layers

In our example, we used bioclimatic variables (downloaded from worldclim.org) as input environmental layers for our SDM. We suggest saving all environmental layers in one folder to make access to these easier. We stack our environmental layers so that they may be processed simultaneously (batch processing) to decrease errors that may occur when processed individually.

**Thread 5**

```
# This searches for all files that have "data/bioclim/" in the path name and have a file extension of .
clim_list <- list.files("../data/bioclim/",pattern=".bil$",full.names = T)

# stacking the bioclim variables to process them at one go
clim <- raster::stack(clim_list)
```

### 2.2 Occurrence data

#### 2.2.1 Download occurrence data

For our example, the nine banded armadillo, we downloaded occurrence data from GBIF, the Global Biodiversity Information Facility. We have provided an if/else statement that checks for a file with "data/occ-raw" in the pathname. If a file does exist, R will load this file, otherwise it will download occurrence locations for *Dasypus novemcinctus*, the nine banded armadillo, from gbif and save it as a .csv file named "data/occ_raw.csv".

**Thread 6**

```
if(file.exists("../data/occ_raw")){
  load("../data/occ_raw")
}else{
  occ_raw <- gbif("Dasypus novemcinctus")
  save(occ_raw,file = "../data/occ_raw")
  write.csv("../data/occ_raw.csv")
```

```
}

# View the first few lines of the occurrence dataset
head(occ_raw)
```

```
##   acceptedNameUsage accessRights adm1 adm2 associatedReferences
## 1              <NA>         <NA> <NA> <NA>                 <NA>
## 2              <NA>         <NA> <NA> <NA>                 <NA>
## 3              <NA>         <NA> <NA> <NA>                 <NA>
## 4              <NA>         <NA> <NA> <NA>                 <NA>
## 5              <NA>         <NA> <NA> <NA>                 <NA>
## 6              <NA>         <NA> <NA> <NA>                 <NA>
##       basisOfRecord behavior bibliographicCitation catalogNumber    class
## 1 HUMAN_OBSERVATION     <NA>                 <NA>       4990630 Mammalia
## 2 HUMAN_OBSERVATION     <NA>                 <NA>       4879238 Mammalia
## 3 HUMAN_OBSERVATION     <NA>                 <NA>       4934903 Mammalia
## 4 HUMAN_OBSERVATION     <NA>                 <NA>       5025320 Mammalia
## 5 HUMAN_OBSERVATION     <NA>                 <NA>       5253808 Mammalia
## 6 HUMAN_OBSERVATION     <NA>                 <NA>       5253797 Mammalia
##   classKey          cloc collectionCode collectionID continent
## 1      359        Mexico   Observations         <NA>      <NA>
## 2      359 United States   Observations         <NA>      <NA>
## 3      359 United States   Observations         <NA>      <NA>
## 4      359 United States   Observations         <NA>      <NA>
## 5      359        Mexico   Observations         <NA>      <NA>
## 6      359        Mexico   Observations         <NA>      <NA>
##   coordinatePrecision coordinateUncertaintyInMeters       country crawlId
## 1                  NA                         11211        Mexico      73
## 2                  NA                            NA United States      73
## 3                  NA                            61 United States      73
## 4                  NA                            15 United States      73
## 5                  NA                            31        Mexico      73
## 6                  NA                            31        Mexico      73
##   datasetID                            datasetKey
## 1      <NA> 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 2      <NA> 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 3      <NA> 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 4      <NA> 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 5      <NA> 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 6      <NA> 50c9509d-22c7-4a22-a47d-8c48425ef4a7
##                              datasetName              dateIdentified day
## 1 iNaturalist research-grade observations 2017-01-22T20:53:11.000+0000  20
## 2 iNaturalist research-grade observations 2017-01-01T22:35:30.000+0000   1
## 3 iNaturalist research-grade observations 2017-01-13T00:13:21.000+0000   3
## 4 iNaturalist research-grade observations 2017-01-30T02:21:40.000+0000  29
## 5 iNaturalist research-grade observations 2017-03-09T00:45:24.000+0000  20
## 6 iNaturalist research-grade observations 2017-03-09T00:45:21.000+0000   1
##   depth depthAccuracy disposition dynamicProperties
## 1    NA            NA        <NA>              <NA>
## 2    NA            NA        <NA>              <NA>
## 3    NA            NA        <NA>              <NA>
## 4    NA            NA        <NA>              <NA>
## 5    NA            NA        <NA>              <NA>
## 6    NA            NA        <NA>              <NA>
```

```
##   earliestEonOrLowestEonothem earliestEpochOrLowestSeries
## 1                        <NA>                        <NA>
## 2                        <NA>                        <NA>
## 3                        <NA>                        <NA>
## 4                        <NA>                        <NA>
## 5                        <NA>                        <NA>
## 6                        <NA>                        <NA>
##   earliestEraOrLowestErathem earliestPeriodOrLowestSystem elevation
## 1                       <NA>                        <NA>        NA
## 2                       <NA>                        <NA>        NA
## 3                       <NA>                        <NA>        NA
## 4                       <NA>                        <NA>        NA
## 5                       <NA>                        <NA>        NA
## 6                       <NA>                        <NA>        NA
##   elevationAccuracy endDayOfYear establishmentMeans
## 1                NA         <NA>               <NA>
## 2                NA         <NA>               <NA>
## 3                NA         <NA>               <NA>
## 4                NA         <NA>               <NA>
## 5                NA         <NA>               <NA>
## 6                NA         <NA>               <NA>
##                     eventDate eventID eventRemarks eventTime      family
## 1 2017-01-20T18:07:06.000+0000    <NA>         <NA> 00:07:06Z Dasypodidae
## 2 2017-01-01T15:05:23.000+0000    <NA>         <NA> 21:05:23Z Dasypodidae
## 3 2017-01-03T00:00:00.000+0000    <NA>         <NA>      <NA> Dasypodidae
## 4 2017-01-29T17:24:00.000+0000    <NA>         <NA> 23:24:00Z Dasypodidae
## 5 2017-01-20T01:53:00.000+0000    <NA>         <NA> 07:53:00Z Dasypodidae
## 6 2017-01-01T03:05:00.000+0000    <NA>         <NA> 09:05:00Z Dasypodidae
##   familyKey fieldNotes fieldNumber   fullCountry     gbifID genericName
## 1      9369       <NA>        <NA>        Mexico 1453372346     Dasypus
## 2      9369       <NA>        <NA> United States 1453323155     Dasypus
## 3      9369       <NA>        <NA> United States 1453348189     Dasypus
## 4      9369       <NA>        <NA> United States 1453388402     Dasypus
## 5      9369       <NA>        <NA>        Mexico 1453490727     Dasypus
## 6      9369       <NA>        <NA>        Mexico 1453490719     Dasypus
##     genus genusKey geodeticDatum geologicalContextID georeferencedBy
## 1 Dasypus  2440775         WGS84                <NA>            <NA>
## 2 Dasypus  2440775         WGS84                <NA>            <NA>
## 3 Dasypus  2440775         WGS84                <NA>            <NA>
## 4 Dasypus  2440775         WGS84                <NA>            <NA>
## 5 Dasypus  2440775         WGS84                <NA>            <NA>
## 6 Dasypus  2440775         WGS84                <NA>            <NA>
##   georeferencedDate georeferenceProtocol georeferenceRemarks
## 1              <NA>                 <NA>                <NA>
## 2              <NA>                 <NA>                <NA>
## 3              <NA>                 <NA>                <NA>
## 4              <NA>                 <NA>                <NA>
## 5              <NA>                 <NA>                <NA>
## 6              <NA>                 <NA>                <NA>
##   georeferenceSources georeferenceVerificationStatus habitat
## 1                <NA>                           <NA>    <NA>
## 2                <NA>                           <NA>    <NA>
## 3                <NA>                           <NA>    <NA>
## 4                <NA>                           <NA>    <NA>
```

```
## 5                <NA>                            <NA>    <NA>
## 6                <NA>                            <NA>    <NA>
##   higherClassification higherGeography higherGeographyID
## 1                 <NA>            <NA>              <NA>
## 2                 <NA>            <NA>              <NA>
## 3                 <NA>            <NA>              <NA>
## 4                 <NA>            <NA>              <NA>
## 5                 <NA>            <NA>              <NA>
## 6                 <NA>            <NA>              <NA>
##   highestBiostratigraphicZone
## 1                        <NA>
## 2                        <NA>
## 3                        <NA>
## 4                        <NA>
## 5                        <NA>
## 6                        <NA>
##                  http://unknown.org/occurrenceDetails identificationID
## 1    https://www.inaturalist.org/observations/4990630         10088385
## 2    https://www.inaturalist.org/observations/4879238          9778997
## 3    https://www.inaturalist.org/observations/4934903          9942563
## 4    https://www.inaturalist.org/observations/5025320         10188611
## 5 http://conabio.inaturalist.org/observations/5253808         10832444
## 6 http://conabio.inaturalist.org/observations/5253797         10832433
##   identificationQualifier identificationReferences identificationRemarks
## 1                    <NA>                     <NA>                  <NA>
## 2                    <NA>                     <NA>                  <NA>
## 3                    <NA>                     <NA>                  <NA>
## 4                    <NA>                     <NA>                  <NA>
## 5                    <NA>                     <NA>                  <NA>
## 6                    <NA>                     <NA>                  <NA>
##   identificationVerificationStatus identifiedBy identifier individualCount
## 1                             <NA>         <NA>    4990630              NA
## 2                             <NA>         <NA>    4879238              NA
## 3                             <NA>         <NA>    4934903              NA
## 4                             <NA>         <NA>    5025320              NA
## 5                             <NA>         <NA>    5253808              NA
## 6                             <NA>         <NA>    5253797              NA
##   informationWithheld infraspecificEpithet institutionCode institutionID
## 1                <NA>                 <NA>     iNaturalist          <NA>
## 2                <NA>                 <NA>     iNaturalist          <NA>
## 3                <NA>                 <NA>     iNaturalist          <NA>
## 4                <NA>                 <NA>     iNaturalist          <NA>
## 5                <NA>                 <NA>     iNaturalist          <NA>
## 6                <NA>                 <NA>     iNaturalist          <NA>
##   island islandGroup ISO2        key  kingdom kingdomKey language
## 1   <NA>        <NA>   MX 1453372346 Animalia          1     <NA>
## 2   <NA>        <NA>   US 1453323155 Animalia          1     <NA>
## 3   <NA>        <NA>   US 1453348189 Animalia          1     <NA>
## 4   <NA>        <NA>   US 1453388402 Animalia          1     <NA>
## 5   <NA>        <NA>   MX 1453490727 Animalia          1     <NA>
## 6   <NA>        <NA>   MX 1453490719 Animalia          1     <NA>
##                    lastCrawled              lastInterpreted
## 1 2017-06-08T07:11:11.177+0000 2017-06-08T07:25:29.690+0000
## 2 2017-06-08T07:10:24.826+0000 2017-06-08T07:23:00.742+0000
```

```
## 3 2017-06-08T07:10:48.729+0000 2017-06-08T07:24:16.084+0000
## 4 2017-06-08T07:11:28.574+0000 2017-06-08T07:26:17.003+0000
## 5 2017-06-08T07:13:03.329+0000 2017-06-08T07:31:08.982+0000
## 6 2017-06-08T07:13:03.015+0000 2017-06-08T07:31:07.830+0000
##                    lastParsed      lat latestEonOrHighestEonothem
## 1 2017-06-08T07:11:38.246+0000 21.32681                       <NA>
## 2 2017-06-08T07:10:55.702+0000 30.62777                       <NA>
## 3 2017-06-08T07:11:21.304+0000 29.65055                       <NA>
## 4 2017-06-08T07:11:53.035+0000 33.06473                       <NA>
## 5 2017-06-08T07:13:27.284+0000 18.54291                       <NA>
## 6 2017-06-08T07:13:26.930+0000 18.54279                       <NA>
##   latestEpochOrHighestSeries latestEraOrHighestErathem
## 1                       <NA>                      <NA>
## 2                       <NA>                      <NA>
## 3                       <NA>                      <NA>
## 4                       <NA>                      <NA>
## 5                       <NA>                      <NA>
## 6                       <NA>                      <NA>
##   latestPeriodOrHighestSystem
## 1                        <NA>
## 2                        <NA>
## 3                        <NA>
## 4                        <NA>
## 5                        <NA>
## 6                        <NA>
##                                                  license lifeStage
## 1 http://creativecommons.org/licenses/by-nc/4.0/legalcode      <NA>
## 2 http://creativecommons.org/licenses/by-nc/4.0/legalcode      <NA>
## 3 http://creativecommons.org/licenses/by-nc/4.0/legalcode      <NA>
## 4 http://creativecommons.org/licenses/by-nc/4.0/legalcode      <NA>
## 5 http://creativecommons.org/licenses/by-nc/4.0/legalcode      <NA>
## 6 http://creativecommons.org/licenses/by-nc/4.0/legalcode      <NA>
##   locality locationAccordingTo locationID locationRemarks        lon
## 1     <NA>                <NA>       <NA>            <NA>  -88.34585
## 2     <NA>                <NA>       <NA>            <NA>  -87.91988
## 3     <NA>                <NA>       <NA>            <NA> -100.06927
## 4     <NA>                <NA>       <NA>            <NA>  -96.97565
## 5     <NA>                <NA>       <NA>            <NA>  -95.15737
## 6     <NA>                <NA>       <NA>            <NA>  -95.15733
##   lowestBiostratigraphicZone                     modified month
## 1                       <NA> 2017-01-22T20:57:57.000+0000     1
## 2                       <NA> 2017-01-02T18:40:05.000+0000     1
## 3                       <NA> 2017-01-19T16:23:03.000+0000     1
## 4                       <NA> 2017-01-31T18:53:35.000+0000     1
## 5                       <NA> 2017-03-09T04:16:17.000+0000     1
## 6                       <NA> 2017-03-09T04:18:41.000+0000     1
##   municipality nameAccordingTo namePublishedIn namePublishedInYear
## 1         <NA>            <NA>           <NA>                <NA>
## 2         <NA>            <NA>           <NA>                <NA>
## 3         <NA>            <NA>           <NA>                <NA>
## 4         <NA>            <NA>           <NA>                <NA>
## 5         <NA>            <NA>           <NA>                <NA>
## 6         <NA>            <NA>           <NA>                <NA>
##   nomenclaturalCode                                    occurrenceID
```

```
## 1              <NA>    https://www.inaturalist.org/observations/4990630
## 2              <NA>    https://www.inaturalist.org/observations/4879238
## 3              <NA>    https://www.inaturalist.org/observations/4934903
## 4              <NA>    https://www.inaturalist.org/observations/5025320
## 5              <NA> http://conabio.inaturalist.org/observations/5253808
## 6              <NA> http://conabio.inaturalist.org/observations/5253797
##                          occurrenceRemarks occurrenceStatus     order
## 1                                     <NA>             <NA> Cingulata
## 2                                     <NA>             <NA> Cingulata
## 3                                     <NA>             <NA> Cingulata
## 4                                     <NA>             <NA> Cingulata
## 5 PROCER2016/ CEIBA JAGUAR A.C.-RB TUXTLAS             <NA> Cingulata
## 6 PROCER2016/ CEIBA JAGUAR A.C.-RB TUXTLAS             <NA> Cingulata
##   orderKey organismID organismRemarks originalNameUsage
## 1      735       <NA>            <NA>              <NA>
## 2      735       <NA>            <NA>              <NA>
## 3      735       <NA>            <NA>              <NA>
## 4      735       <NA>            <NA>              <NA>
## 5      735       <NA>            <NA>              <NA>
## 6      735       <NA>            <NA>              <NA>
##   otherCatalogNumbers ownerInstitutionCode parentNameUsage   phylum
## 1                <NA>                <NA>            <NA> Chordata
## 2                <NA>                <NA>            <NA> Chordata
## 3                <NA>                <NA>            <NA> Chordata
## 4                <NA>                <NA>            <NA> Chordata
## 5                <NA>                <NA>            <NA> Chordata
## 6                <NA>                <NA>            <NA> Chordata
##   phylumKey preparations previousIdentifications    protocol
## 1        44         <NA>                    <NA> DWC_ARCHIVE
## 2        44         <NA>                    <NA> DWC_ARCHIVE
## 3        44         <NA>                    <NA> DWC_ARCHIVE
## 4        44         <NA>                    <NA> DWC_ARCHIVE
## 5        44         <NA>                    <NA> DWC_ARCHIVE
## 6        44         <NA>                    <NA> DWC_ARCHIVE
##   publishingCountry                      publishingOrgKey
## 1                US 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 2                US 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 3                US 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 4                US 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 5                US 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 6                US 28eb1a3f-1c15-4a95-931a-4af90ecb574d
##             recordedBy recordNumber
## 1        bernardo_zg21         <NA>
## 2              Jessica         <NA>
## 3 Diana-Terry Hibbitts         <NA>
## 4       Jennifer Linde         <NA>
## 5    CEIBA JAGUAR A.C.         <NA>
## 6    CEIBA JAGUAR A.C.         <NA>
##                                      references reproductiveCondition
## 1 https://www.inaturalist.org/observations/4990630                 <NA>
## 2 https://www.inaturalist.org/observations/4879238                 <NA>
## 3 https://www.inaturalist.org/observations/4934903                 <NA>
## 4 https://www.inaturalist.org/observations/5025320                 <NA>
## 5 https://www.inaturalist.org/observations/5253808                 <NA>
```

```
## 6 https://www.inaturalist.org/observations/5253797                 <NA>
##                                       rights         rightsHolder
## 1        © bernardo_zg21 some rights reserved        bernardo_zg21
## 2              © Jessica some rights reserved              Jessica
## 3 © Diana-Terry Hibbitts some rights reserved Diana-Terry Hibbitts
## 4       © Jennifer Linde some rights reserved       Jennifer Linde
## 5    © CEIBA JAGUAR A.C. some rights reserved    CEIBA JAGUAR A.C.
## 6    © CEIBA JAGUAR A.C. some rights reserved    CEIBA JAGUAR A.C.
##   samplingEffort samplingProtocol                      scientificName
## 1           <NA>             <NA> Dasypus novemcinctus Linnaeus, 1758
## 2           <NA>             <NA> Dasypus novemcinctus Linnaeus, 1758
## 3           <NA>             <NA> Dasypus novemcinctus Linnaeus, 1758
## 4           <NA>             <NA> Dasypus novemcinctus Linnaeus, 1758
## 5           <NA>             <NA> Dasypus novemcinctus Linnaeus, 1758
## 6           <NA>             <NA> Dasypus novemcinctus Linnaeus, 1758
##   scientificNameID  sex source              species speciesKey
## 1             <NA> <NA>   <NA> Dasypus novemcinctus    2440779
## 2             <NA> <NA>   <NA> Dasypus novemcinctus    2440779
## 3             <NA> <NA>   <NA> Dasypus novemcinctus    2440779
## 4             <NA> <NA>   <NA> Dasypus novemcinctus    2440779
## 5             <NA> <NA>   <NA> Dasypus novemcinctus    2440779
## 6             <NA> <NA>   <NA> Dasypus novemcinctus    2440779
##   specificEpithet startDayOfYear taxonID taxonKey taxonomicStatus
## 1    novemcinctus           <NA>   47075  2440779            <NA>
## 2    novemcinctus           <NA>   47075  2440779            <NA>
## 3    novemcinctus           <NA>   47075  2440779            <NA>
## 4    novemcinctus           <NA>   47075  2440779            <NA>
## 5    novemcinctus           <NA>   47075  2440779            <NA>
## 6    novemcinctus           <NA>   47075  2440779            <NA>
##   taxonRank taxonRemarks type typeStatus typifiedName
## 1   SPECIES         <NA> <NA>       <NA>         <NA>
## 2   SPECIES         <NA> <NA>       <NA>         <NA>
## 3   SPECIES         <NA> <NA>       <NA>         <NA>
## 4   SPECIES         <NA> <NA>       <NA>         <NA>
## 5   SPECIES         <NA> <NA>       <NA>         <NA>
## 6   SPECIES         <NA> <NA>       <NA>         <NA>
##   verbatimCoordinateSystem verbatimElevation
## 1                     <NA>              <NA>
## 2                     <NA>              <NA>
## 3                     <NA>              <NA>
## 4                     <NA>              <NA>
## 5                     <NA>              <NA>
## 6                     <NA>              <NA>
##                         verbatimEventDate
## 1 Fri Jan 20 2017 18:07:06 GMT-0600 (CST)
## 2 Sun Jan 01 2017 15:05:23 GMT-0600 (CST)
## 3                              2017-01-03
## 4                  2017/01/29 5:24 PM CST
## 5                  2017/01/20 1:53 AM CST
## 6                  2017/01/01 3:05 AM CST
##                       verbatimLocality verbatimSRS verbatimTaxonRank
## 1                  Panab, Panab, YUC, MX        <NA>              <NA>
## 2  Village Park Preserve, Daphne, AL, US        <NA>              <NA>
## 3 Texas: Edwards County, Camp Wood Hills        <NA>              <NA>
```

```
## 4                    Lewisville, TX, USA        <NA>             <NA>
## 5         San Andrés Tuxtla, Ver., México       <NA>             <NA>
## 6         San Andrés Tuxtla, Ver., México       <NA>             <NA>
##   vernacularName waterBody year downloadDate
## 1           <NA>      <NA> 2017   2017-07-04
## 2           <NA>      <NA> 2017   2017-07-04
## 3           <NA>      <NA> 2017   2017-07-04
## 4           <NA>      <NA> 2017   2017-07-04
## 5           <NA>      <NA> 2017   2017-07-04
## 6           <NA>      <NA> 2017   2017-07-04
```

### 2.2.2 Clean occurrence data

Since some of our records do not have appropriate coordinates and some have missing locational data, we need to find these records and remove them from our dataset. To do this, we create a new dataset named "occ_clean", which is a subset of the "occ_raw" dataset where records with missing latitude and/or longitude are removed. This particular piece of code also returns the number of records that were removed from the dataset. Additionally, we remove duplicate records and create a subset of the cleaned data with the duplicates removed.

**Thread 7**

```
# remove bad coordinates, where either the lat or long coordinate is missing
occ_clean <- subset(occ_raw,(!is.na(lat))&(!is.na(lon)))
cat(nrow(occ_raw)-nrow(occ_clean), "records are removed")
```

```
## 2426 records are removed
```

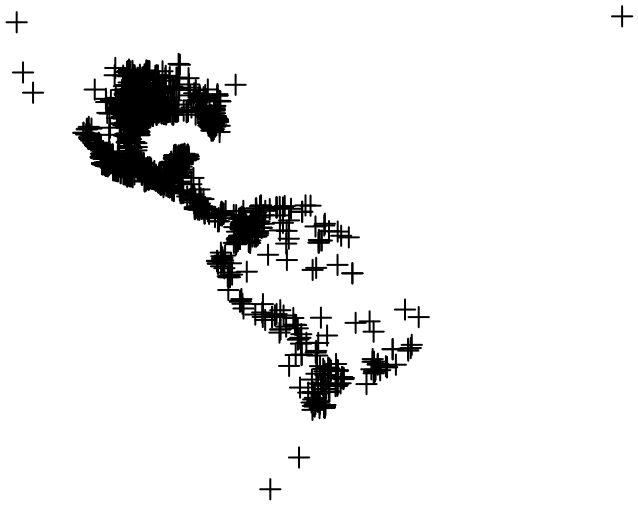
```
# remove duplicated data based on latitude and longitude
dups <- duplicated(occ_clean[c("lat","lon")])
occ_unique <- occ_clean[!dups,]
cat(nrow(occ_clean)-nrow(occ_unique), "records are removed")
```

```
## 1506 records are removed
```

Up to this point we have been working with a data frame, but it has no spatial relationship defined in R, so we needed to make the data spatial. Once our data is spatial we can use the plot() function to see the occurrence data and allow us to check for data points that appear to be erroneous.

**Thread 8**

```
# make occ spatial
coordinates(occ_unique) <- ~ lon + lat
## look for erroneous points
plot(occ_unique)
```



```
#Figure *****????
```

In Figure #, we can see several points that appear outside the known distribution of *Dasypus novemcinctus* (North and South America) and we need to remove these from our occurrence data set. To do this we select points that have longitudes greater than -110 or lower than -40 (outside our area of interest) and remove them from the data set.

### Thread 9

```
# remove some errors (i.e., only keep good records)
occ_unique <- occ_unique[which(occ_unique$lon>-110 &
                                 occ_unique$lon < -40),]
```

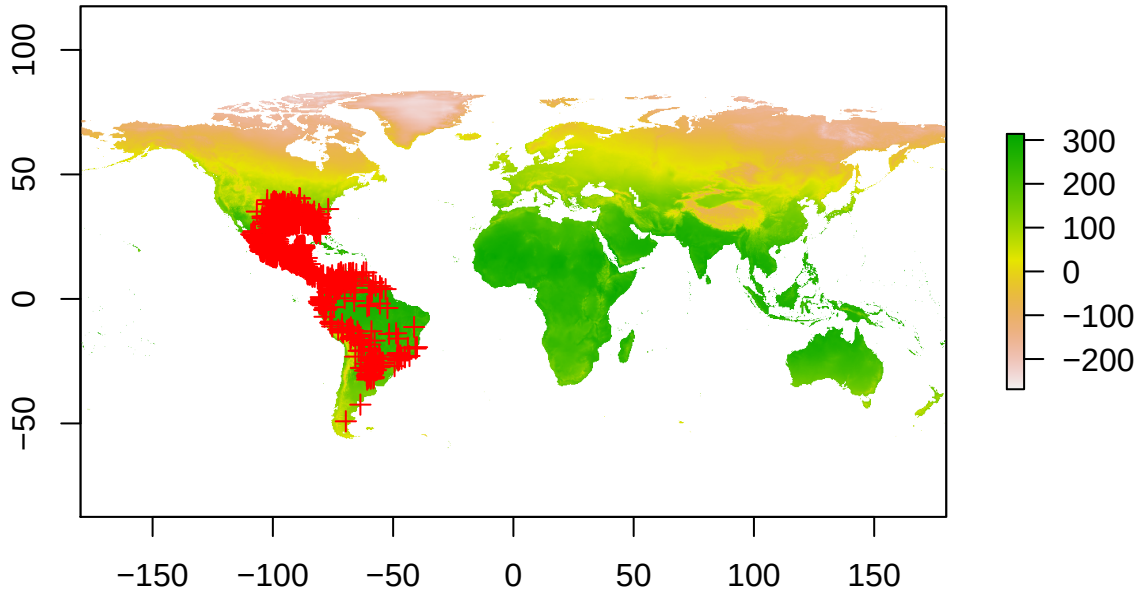
Maxent only utilizes only one occurrence location per pixel or cell for the environmental data when creating models, so we need to thin our occurrence data so that only one location falls within each cell.

### Thread 10

```
# thin occ data (keep one occ per cell)
cells <- cellFromXY(clim[[1]],occ_unique)
dups <- duplicated(cells)
occ_final <- occ_unique[!dups,]
cat(nrow(occ_unique)-nrow(occ_final), "records are removed")
```

```
## 1124 records are removed
```

```r
# plot the first climatic layer (or )
plot(clim[[1]])
# replace [[1]] with any nth number of the layer of interest from the raster stack

# plot the final occurrence data on the enviromental layer
plot(occ_final,add=T,col="red")
```




```r
# the 'add=T' tells R to put the incoming data on the existing layer
```

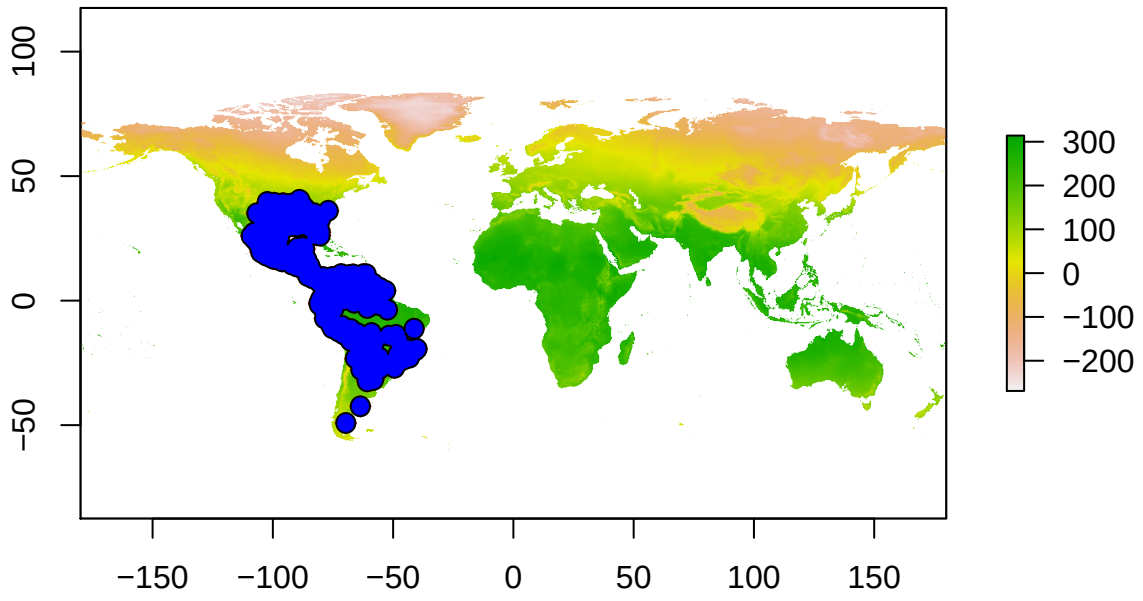
### 2.3 Set up study area

In setting up our study area we adopt a methodology that better samples the background and presence environmental conditions for modeling. we create a buffer around our occurrence locations and define this as our study region, which will allows us to better train the model to distinguish between conditions associated with species presence and background locations without over sampling from the background. We establish a 4 decimal degree buffer around the occurrence points and to make sure that our buffer encompasses the appropriate area, we plot the occurrence points, the first environmental layer, and the buffer polygon.

**Thread 11**

```r
# this creates a 4 decimal degree buffer around the occurrence data
occ_buff <- buffer(occ_final,4)

# plot the first element ([[1]]) in the raster stack
plot(clim[[1]])
```

```
# this adds the occurrence data
plot(occ_final,add=T,col="red")
# this adds the buffer polygon
plot(occ_buff,add=T,col="blue")
```




With a defined study area and the environmental layers stacked, we then clip the layers to the extent of our study area. However, for ease of processing, we do this in two steps rather than one. First we create a coarse rectangular shaped study area around the occurrence data and study area to reduce environmental data raster size and then extract by mask using the buffer we created to more accurately clip environmental layers. We have found that this approach keeps the computer from slowing down by trying to clip the world extent bioclim data layers to a smaller study area. We save the cropped environmental layers as .asc (ascii files) as inputs for Maxent.

**Thread 12**

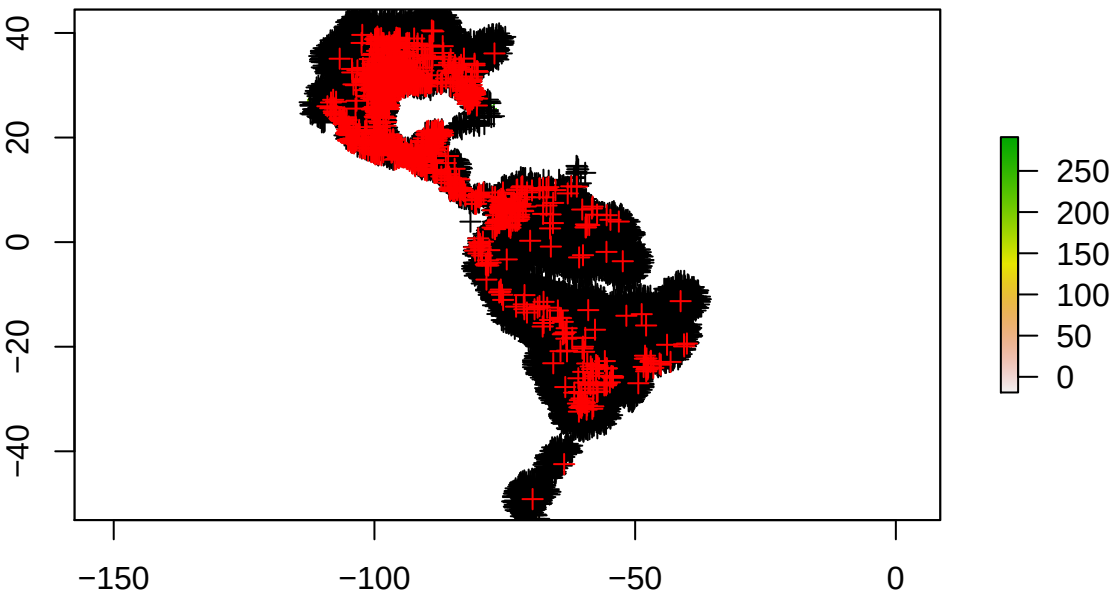
```
# crop study area to a manageable extent (rectangle shaped)
studyArea <- crop(clim,extent(occ_buff))

# the 'study area' created by extracting the buffer area from the raster stack
studyArea <- mask(studyArea,occ_buff)
# output will still be a raster stack, just of the study area

# save the new study area rasters as ascii
writeRaster(studyArea,
            filename=paste0("../data/studyarea/",names(studyArea),".asc"), ## a series of names for outp
            format="ascii", ## the output format
```

```
            bylayer=TRUE, ## this will save a series of layers
            overwrite=T)
```

After processing our environmental data, we now need to define our background points and establish training and testing points. We want define a random sample that will be selected each time we used the sampleRandom function in R, so that we can re-run models. This will allow the same set of points to be selected every time the code is run rather than generating a new set of random points. The set.seed(1) function accomplished this. With an established a random selection technique we select 10,000 background points from the study area, ignoring pixels with no data. To visualize the background points selected, we plot the study area, occurrence data, and background data.

**Thread 13**

```
# select background points from this buffered area
# when the number provided to set.seed() function, the same random sample will be selected in the next
set.seed(1)

# use this code before the sampleRandom function everytime, if you want to get the same "random samples
bg <- sampleRandom(x=studyArea,
                   size=10000,
                   na.rm=T, ## na.rm removes the 'Not Applicaple' points
                   sp=T) ## sp is telling R to give us spatial points

plot(studyArea[[1]])
# add the background points to the plotted raster
plot(bg,add=T)
# add the occurrence data to the plottted raster
plot(occ_final,add=T,col="red")
```




### 2.4 Split occurrence data into training & testing

We then use the same set.seed(1) function to select 50% of our data. We define the first 50% selected as training data and the other 50% as testing.

**Thread 14**

```
# get the same random sample for training and testing
set.seed(1)

# randomly select 50% for training
selected <- sample(1:nrow(occ_final),nrow(occ_final)*0.5)
# this is the selection
occ_train <- occ_final[selected,]

# this is the opposite of the selection
occ_test <- occ_final[-selected,]
```

### 2.5 Format data for Maxent

The last step before model creation is to format Maxent inputs for modeling since the dismo package requires a table/dataframe for model inputs. We extract environmental data from the raster stack for the backround, training, and testing points in a dataframe format.

**Thread 15**

```
# extracting env conditions for training occ from the raster stack; a data frame is returned (i.e multi
p <- extract(clim,occ_train)
# env conditions for testing occ
p_test <- extract(clim,occ_test)
# extracting env conditions for background
a <- extract(clim,bg)
```

Maxent reads a "1" as presence and "0" as pseudo-absence. Thus, we need to assign a "1" to the training environmental conditions and a "0" for the background. We first create a set of rows with the same number as the training and testing data, and put the value of "1" each cell and a "0" for background. We then combine the "1"s and "0"s into a vector that will be added to the dataframe containing the environmental conditions associated with the testing and background conditions.

**Thread 16**

```
# repeat the number 1 as many numbers as the number of rows in p, and repeat 0 as the rows of backgroun
pa <- c(rep(1,nrow(p)), rep(0,nrow(a)))

# (rep(1,nrow(p)) creating the number of rows as the p data set to have the number one as the indicator
#  rep(0,nrow(a)) creating the number of rows as the a data set to have the number zero as the indicato
# The c combines these ones and zeros into a new vector that can be added to the Maxent table data fram
pder <- as.data.frame(rbind(p,a))
```

# 3 Maxent models

## 3.1 Simple implementation

To run simple distribution models using the "dismo"" package, only three function specifications are needed; the climatic conditions, the occurrence data, and an output location for the results. The environmental conditions and occurrence data are the databases we created earlier. Model parameters can also be specified in the maxent function, when not speficied default parameters are used. We provide details about modeling parameters later in this document. You can view model results in an html browser.

**Thread 17**

```
# mod <- maxent(x=climatic data, p=occurrence data, path=output location)
mod <- maxent(x=pder, ## env conditions
              p=pa,   ## 1:presence or 0:absence

              path=paste0("../output/maxent_outputs"), ## folder for maxent output; if we do not specify
              args=c("responsecurves") ## parameter specification
              )
## the maxent functions runs a model in the default settings..to change these parameters, you have to te

# view the maxent model in a html brower
mod
```

```
## class    : MaxEnt
## variables: bio1 bio10 bio11 bio12 bio13 bio14 bio15 bio16 bio17 bio18 bio19 bio2 bio3 bio4 bio5 bio6
```

```
# view detailed results
mod@results
```

```
##                                                                  [,1]
## X.Training.samples                                           655.0000
## Regularized.training.gain                                      0.7265
## Unregularized.training.gain                                    0.9604
## Iterations                                                   500.0000
## Training.AUC                                                   0.8596
## X.Background.points                                        10575.0000
## bio1.contribution                                             17.1627
## bio10.contribution                                            20.4753
## bio11.contribution                                             8.7616
## bio12.contribution                                             8.4875
## bio13.contribution                                             1.8276
## bio14.contribution                                             0.7496
## bio15.contribution                                             9.2740
## bio16.contribution                                             0.6694
## bio17.contribution                                             0.6045
## bio18.contribution                                             0.9334
## bio19.contribution                                             0.9610
## bio2.contribution                                              1.0134
## bio3.contribution                                             15.1186
## bio4.contribution                                              1.3084
## bio5.contribution                                              8.2928
## bio6.contribution                                              3.3366
## bio7.contribution                                              0.3255
## bio8.contribution                                              0.1911
## bio9.contribution                                              0.5069
## bio1.permutation.importance                                   16.4025
## bio10.permutation.importance                                  10.6769
## bio11.permutation.importance                                   3.4186
## bio12.permutation.importance                                   7.1080
## bio13.permutation.importance                                   3.0867
## bio14.permutation.importance                                   3.9966
## bio15.permutation.importance                                  20.4166
## bio16.permutation.importance                                   0.3416
## bio17.permutation.importance                                   0.5438
## bio18.permutation.importance                                   0.6744
## bio19.permutation.importance                                   1.3227
## bio2.permutation.importance                                    1.5541
## bio3.permutation.importance                                    2.9937
## bio4.permutation.importance                                   20.8722
## bio5.permutation.importance                                    1.2303
## bio6.permutation.importance                                    3.2102
## bio7.permutation.importance                                    1.0008
## bio8.permutation.importance                                    0.0737
## bio9.permutation.importance                                    1.0766
## Entropy                                                        8.5492
## Prevalence..average.of.logistic.output.over.background.sites.  0.2410
## Fixed.cumulative.value.1.cumulative.threshold                  1.0000
## Fixed.cumulative.value.1.logistic.threshold                    0.0625
## Fixed.cumulative.value.1.area                                  0.8119
## Fixed.cumulative.value.1.training.omission                     0.0046
## Fixed.cumulative.value.5.cumulative.threshold                  5.0000
## Fixed.cumulative.value.5.logistic.threshold                    0.1355
## Fixed.cumulative.value.5.area                                  0.6447
```

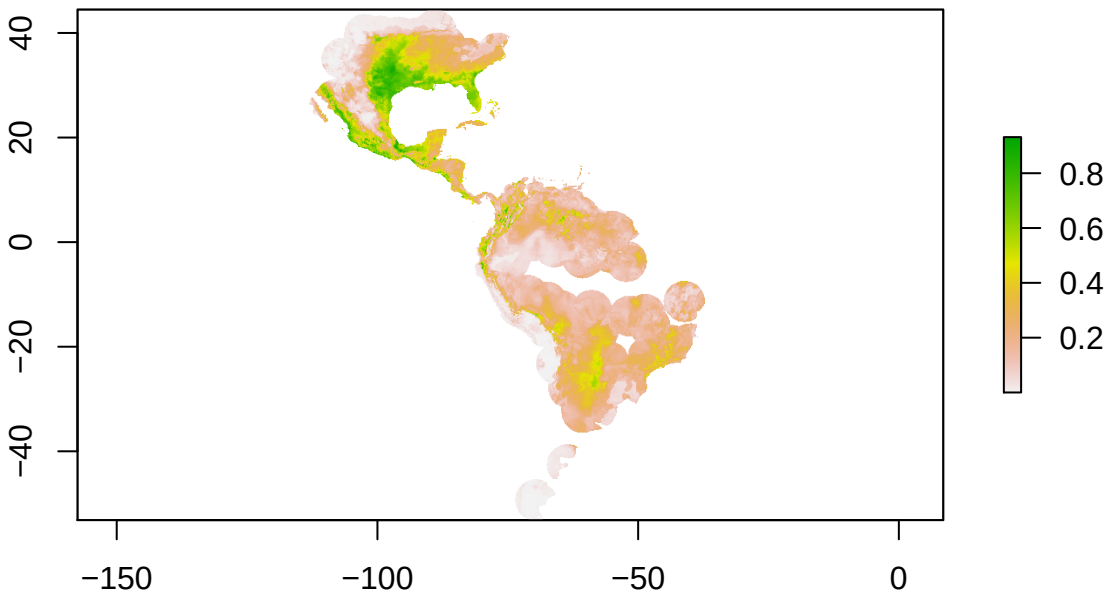
```
## Fixed.cumulative.value.5.training.omission                                    0.0229
## Fixed.cumulative.value.10.cumulative.threshold                               10.0000
## Fixed.cumulative.value.10.logistic.threshold                                  0.1833
## Fixed.cumulative.value.10.area                                                0.5157
## Fixed.cumulative.value.10.training.omission                                   0.0473
## Minimum.training.presence.cumulative.threshold                                0.0200
## Minimum.training.presence.logistic.threshold                                  0.0054
## Minimum.training.presence.area                                                0.9608
## Minimum.training.presence.training.omission                                   0.0000
## X10.percentile.training.presence.cumulative.threshold                        17.8502
## X10.percentile.training.presence.logistic.threshold                           0.2531
## X10.percentile.training.presence.area                                         0.3762
## X10.percentile.training.presence.training.omission                            0.0992
## Equal.training.sensitivity.and.specificity.cumulative.threshold              32.3187
## Equal.training.sensitivity.and.specificity.logistic.threshold                 0.3702
## Equal.training.sensitivity.and.specificity.area                               0.2168
## Equal.training.sensitivity.and.specificity.training.omission                  0.2168
## Maximum.training.sensitivity.plus.specificity.cumulative.threshold           25.9220
## Maximum.training.sensitivity.plus.specificity.logistic.threshold              0.3177
## Maximum.training.sensitivity.plus.specificity.area                            0.2765
## Maximum.training.sensitivity.plus.specificity.training.omission               0.1389
## Balance.training.omission..predicted.area.and.threshold.value.cumulative.threshold     2.0374
## Balance.training.omission..predicted.area.and.threshold.value.logistic.threshold      0.0965
## Balance.training.omission..predicted.area.and.threshold.value.area            0.7536
## Balance.training.omission..predicted.area.and.threshold.value.training.omission       0.0076
## Equate.entropy.of.thresholded.and.original.distributions.cumulative.threshold        11.3240
## Equate.entropy.of.thresholded.and.original.distributions.logistic.threshold           0.1949
## Equate.entropy.of.thresholded.and.original.distributions.area                         0.4881
## Equate.entropy.of.thresholded.and.original.distributions.training.omission            0.0534
```

### 3.2 Predict function

Once a model has been constructed, it will not provide predictions unless specified in the model parameters. However, by using the predict function we can predict the model to raster layers or occurrence data frames.

**Thread 18**

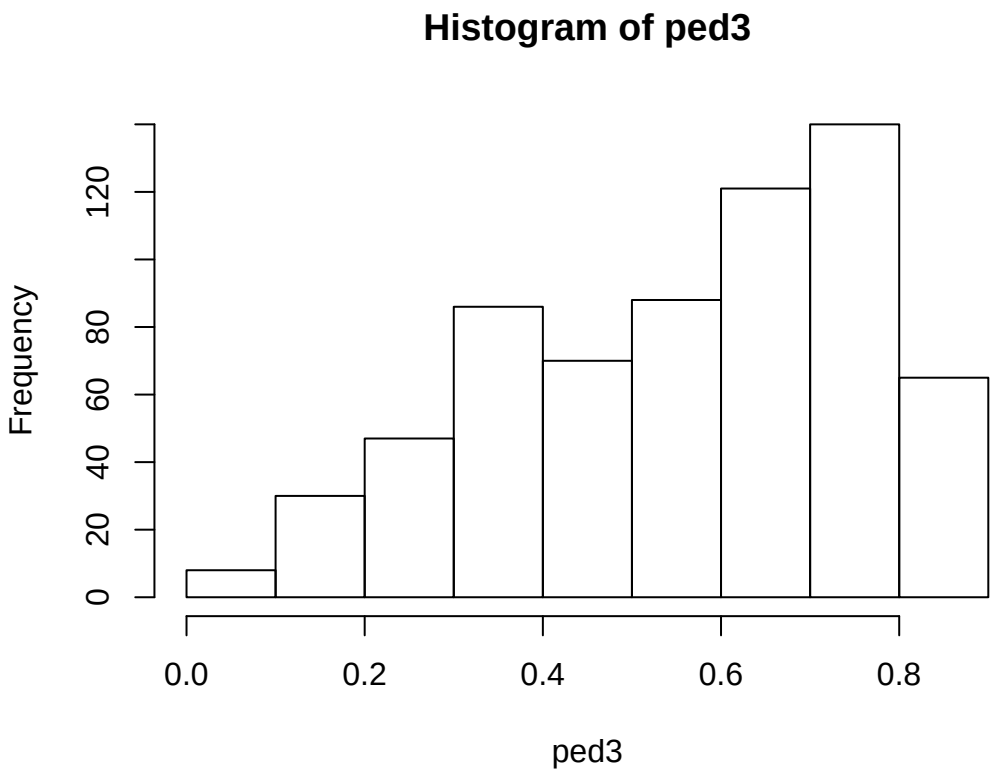
```r
# maxent.R doesnt give us a prediction of training data/layers (unless you specify the projection layer

# example 1, project to study area [raster]
ped1 <- predict(mod,studyArea) # studyArea is the clipped rasters we used to extract environmental cond
plot(ped1) # plot the continuous prediction
```




```
# example 2, project to the world
#ped2 <- predict(mod,clim)
#plot(ped2)

# example 3, project with training occurrences [dataframes]
ped3 <- predict(mod,p)
head(ped3)
```

```
## [1] 0.7553921 0.3420225 0.5019929 0.5993227 0.7655950 0.7684774
```

```
hist(ped3)# creates a histogram of the prediction
```




### 3.3 Model evaluation

To evaluate models, we use the evaluate function from the "dismo" package, which gives model performance using the AUC metric. Training and testing AUC can be calculated using the evaluate function.

**Thread 19**

```
# using "training data" to evaluate
mod_eval_train <- dismo::evaluate(p=p,a=a,model=mod)
print(mod_eval_train)#p & a are dataframe/s (the p and a are the training presence and background point
```

```
## class          : ModelEvaluation
## n presences    : 655
## n absences     : 10000
## AUC            : 0.8807075
## cor            : 0.4044683
## max TPR+TNR at : 0.3175671
```

```
mod_eval_test <- dismo::evaluate(p=p_test,a=a,model=mod)
print(mod_eval_test) # training AUC may be higher than testing AUC
```

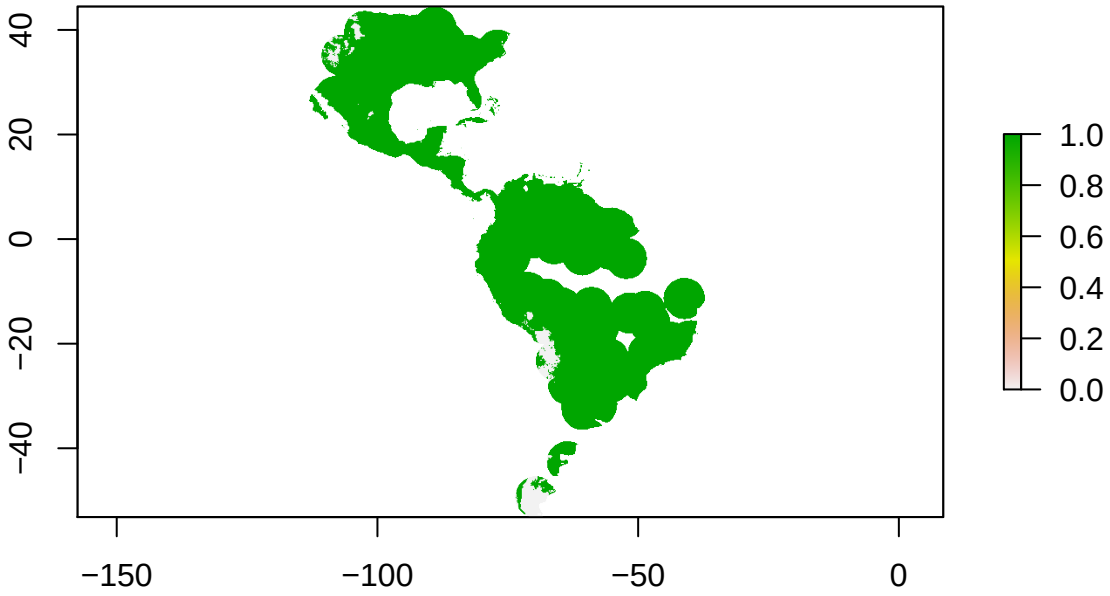
```
## class          : ModelEvaluation
## n presences    : 657
## n absences     : 10000
## AUC            : 0.8401474
## cor            : 0.3532565
```

```
## max TPR+TNR at : 0.3763733
```

To threshold our continuous predictions of suitability into binary predictions we use the threshold function of the "dismo" package. To plot the binary prediction, we plot the predictions that are larger than the threshold.

**Thread 20**

```
# calculate thresholds of models
thd1 <- threshold(mod_eval_train,"no_omission") # 0% omission rate [minimum training presence]
thd2 <- threshold(mod_eval_train,"spec_sens") # hiest TSS

# plotting points that are above the previously calculated tresholded value
plot(ped1>=thd1)
```




## 4 Maxent parameters

### 4.1 Select features

```
# load the function that prepares parameters for maxent
source("../code/Appendix2_prepPara.R")

mod1_autofeature <- maxent(x=pder[c("bio1","bio4","bio11")], ## env conditions, here we selected only 3
               p=pa, ## 1:presence or 0:absence
              path="../output/maxent_outputs1_auto", ## path of maxent output, this is the folder you w
              args=prepPara(userfeatures=NULL) )  ## default is autofeature
```

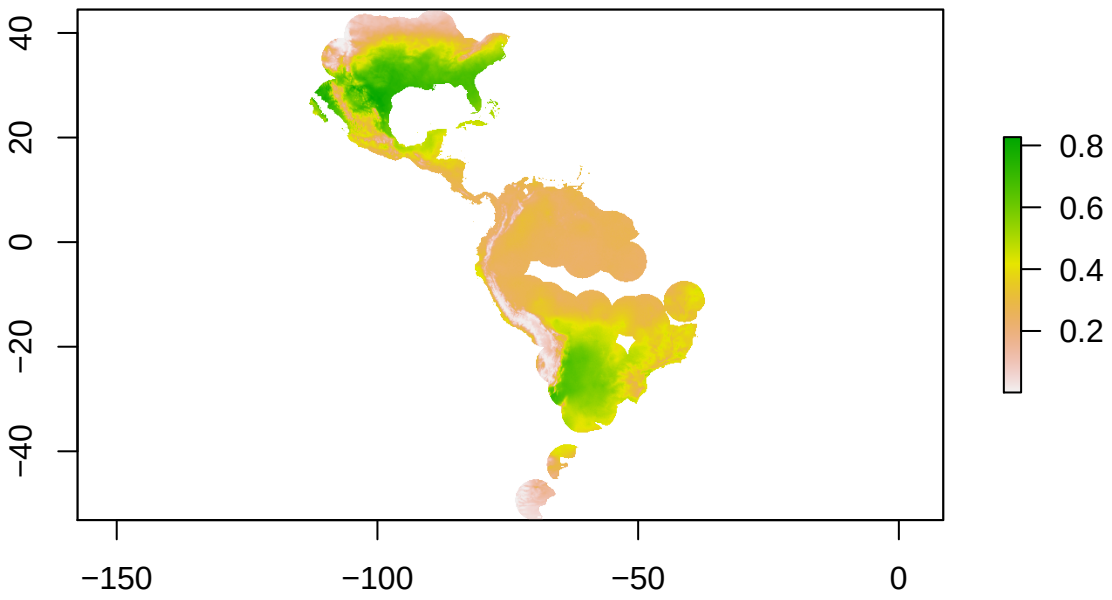
```
# or select Linear& Quadratic features
mod1_lq <- maxent(x=pder[c("bio1","bio4","bio11")], ## env conditions, here we selected only 3 predicto
                  p=pa, ## 1:presence or 0:absence
                 path=paste0("../output/maxent_outputs1_lq"), ## path of maxent output, this is the folder
                 args=prepPara(userfeatures="LQ") )  ## default is autofeature, here LQ represents Linear&
```

### 4.2 Change beta-multiplier

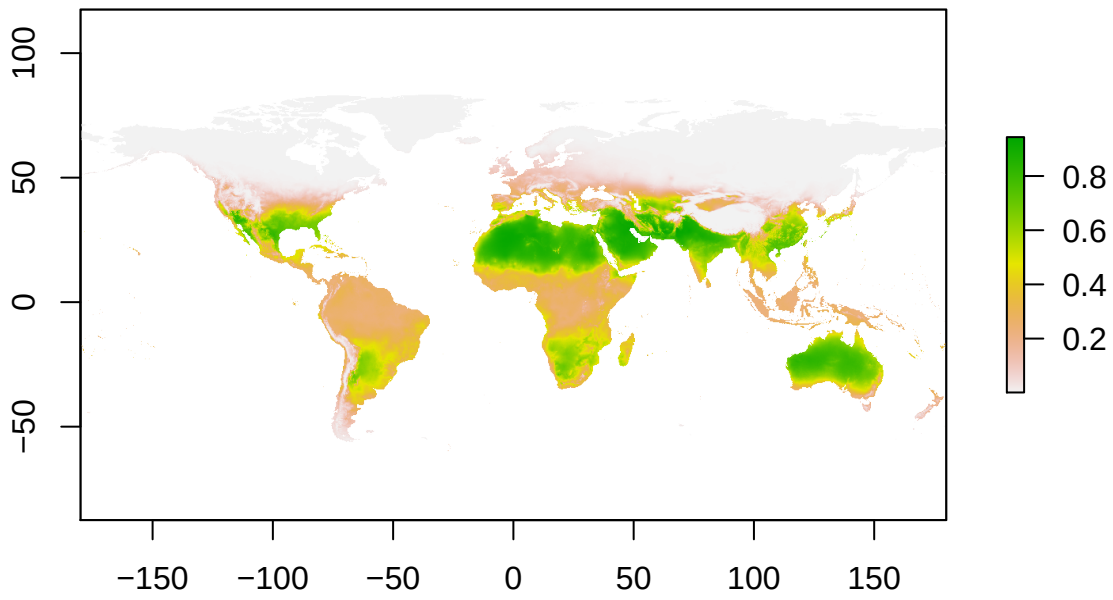
```
#change betamultiplier for all features
mod2 <- maxent(x=pder[c("bio1","bio4","bio11")],
                  p=pa,
                 path=paste0("../output/maxent_outputs2_0.5"),
                 args=prepPara(userfeatures="LQ",
                               betamultiplier=0.5) )

mod2 <- maxent(x=pder[c("bio1","bio4","bio11")],
                  p=pa,
                 path=paste0("../output/maxent_outputs2_complex"),
                 args=prepPara(userfeatures="LQH", ## include L, Q, H features
                               beta_lqp=1.5, ## use different betamultiplier for different features
                               beta_hinge=0.5 ) )
```

### 4.3 Specify projection layers

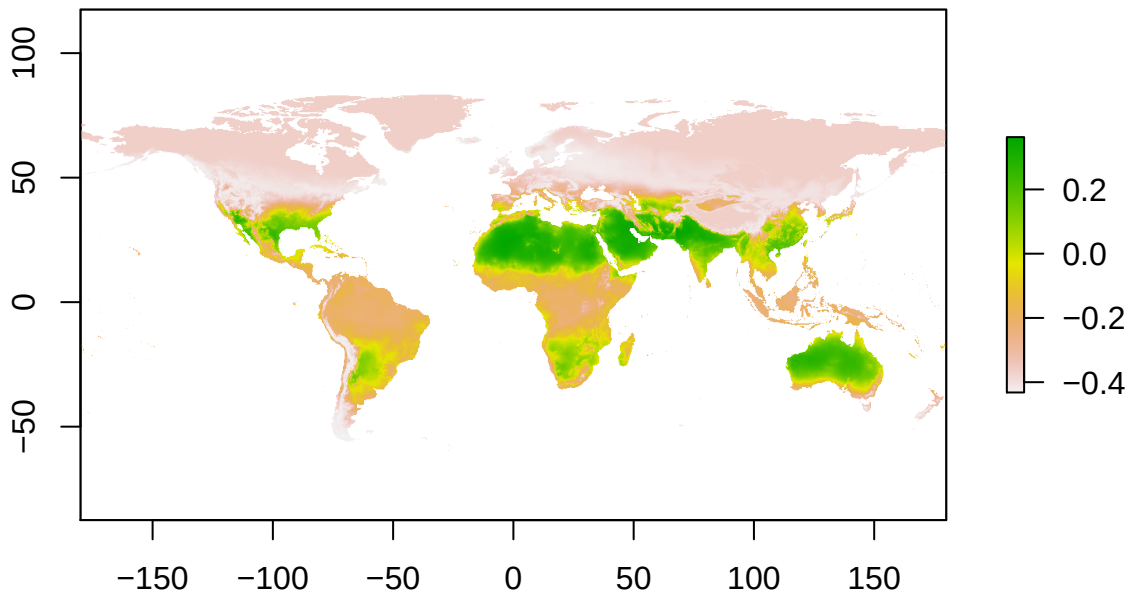
```
# note: (1)the projection layers must exist in the hard disk (as relative to computer RAM); (2) the nam
mod3 <- maxent(x=pder[c("bio1","bio11")],
                  p=pa,
                 path=paste0("../output/maxent_outputs3_prj1"),
                 args=prepPara(userfeatures="LQ",
                               betamultiplier=1,
                               projectionlayers="/Users/iel82user/Google Drive/1_osu_lab/projects/2017_7_w

# load the projected map
ped <- raster(paste0("../output/maxent_outputs3_prj1/species_studyarea.asc"))
plot(ped)
```




```
# we can also project on a broader map, but please caustion about the inaccuracy associated with model
mod3 <- maxent(x=pder[c("bio1","bio11")],
               p=pa,
              path=paste0("../output/maxent_outputs3_prj2"),
              args=prepPara(userfeatures="LQ",
                            betamultiplier=1,
                            projectionlayers="/Users/iel82user/Google Drive/1_osu_lab/projects/2017_7_w
# plot the map
ped <- raster(paste0("../output/maxent_outputs3_prj2/species_bioclim.asc"))
plot(ped)
```

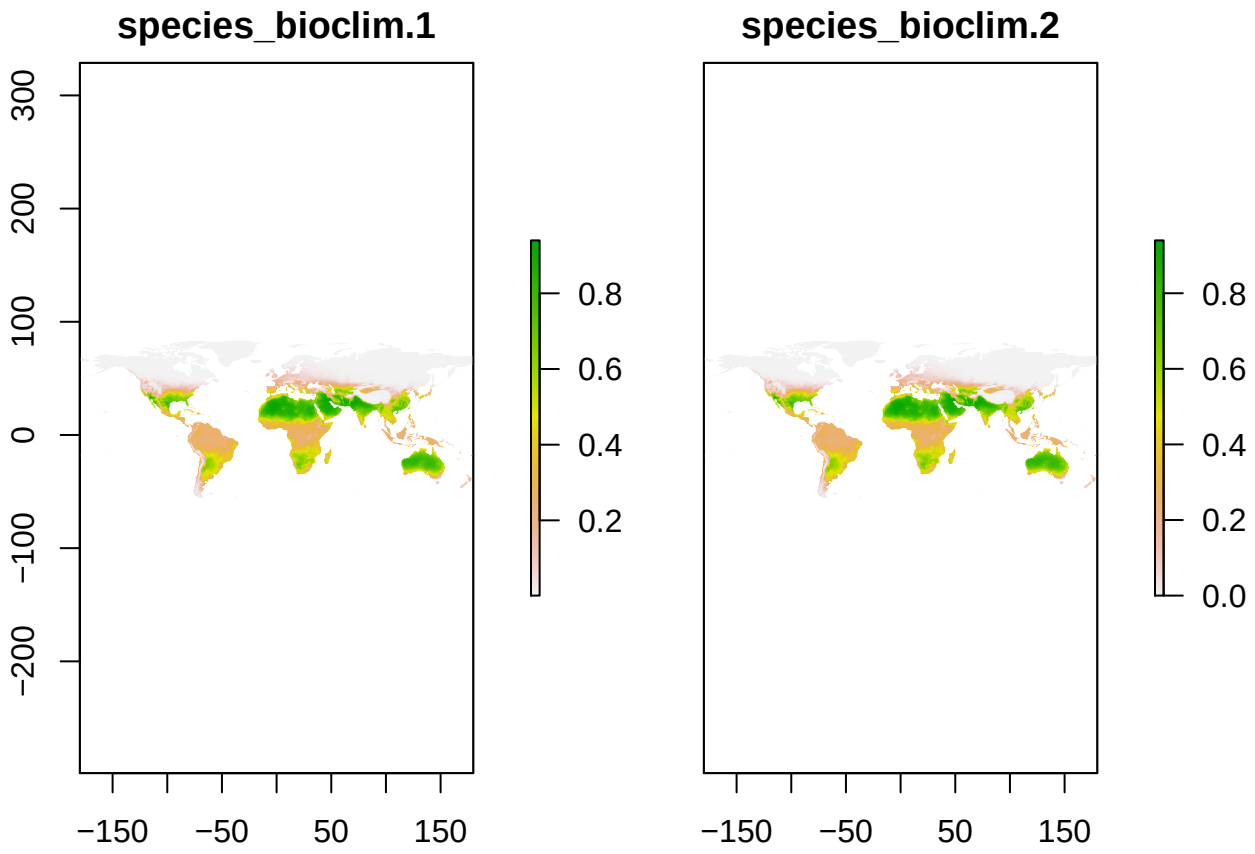



```
# simply check the difference if we used a different betamultiplier
mod3_beta1 <- maxent(x=pder[c("bio1","bio11")],
                     p=pa,
                    path=paste0("../output/maxent_outputs3_prj3"),
                    args=prepPara(userfeatures="LQ",
                                  betamultiplier=100, ## for an extreme example, set beta as 100
                                  projectionlayers="/Users/iel82user/Google Drive/1_osu_lab/projects/2017_7_w
ped3 <- raster(paste0("../output/maxent_outputs3_prj3/species_bioclim.asc"))
plot(ped-ped3) ## quickly check the difference between the two predictions
```

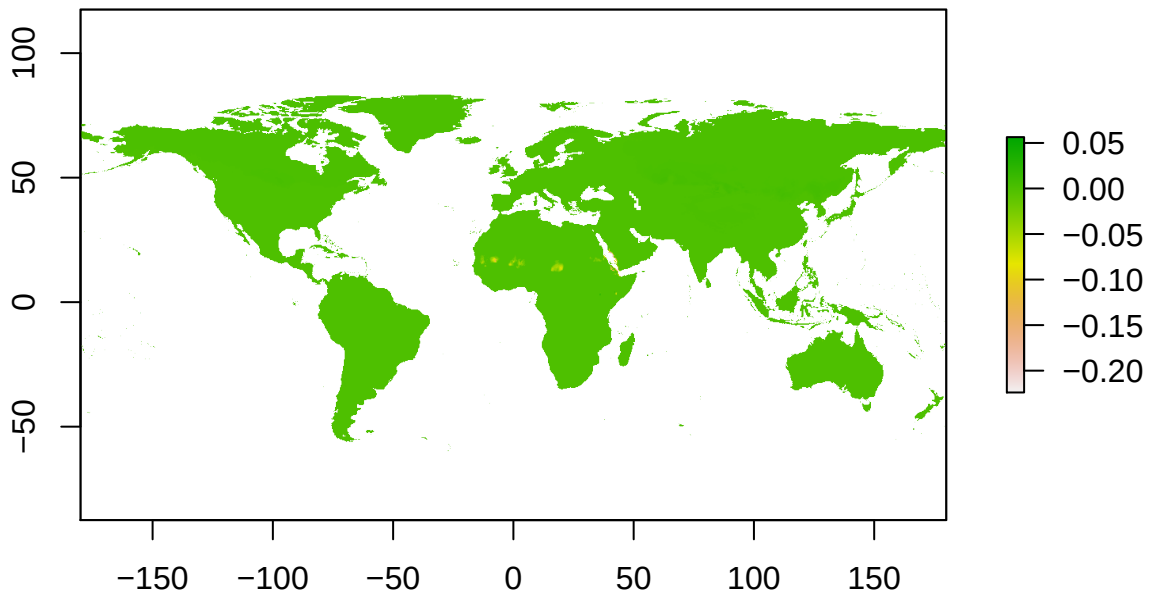



### 4.4 Clamping function

```
# enable or disable clamping function; note clamping function is involved when projecting
mod4_clamp <- maxent(x=pder[c("bio1","bio11")],
                     p=pa,
                     path=paste0("../output/maxent_outputs4_clamp"),
                     args=prepPara(userfeatures="LQ",
                                   betamultiplier=1,
                                   doclamp = TRUE,
                                   projectionlayers="/Users/iel82user/Google Drive/1_osu_lab/projects/2

mod4_noclamp <- maxent(x=pder[c("bio1","bio11")],
                       p=pa,
                       path=paste0("../output/maxent_outputs4_noclamp"),
                       args=prepPara(userfeatures="LQ",
                                     betamultiplier=1,
                                     doclamp = FALSE,
                                     projectionlayers="/Users/iel82user/Google Drive/1_osu_lab/projects


ped_clamp <- raster(paste0("../output/maxent_outputs4_clamp/species_bioclim.asc") )
ped_noclamp <- raster(paste0("../output/maxent_outputs4_noclamp/species_bioclim.asc") )
plot(stack(ped_clamp,ped_noclamp))
```




```
plot(ped_clamp - ped_noclamp) ## we may notice small difference, especially clamp shows higher predicti
```




### 4.5 Cross validation

```
mod4_cross <- maxent(x=pder[c("bio1","bio11")], p=pa,
                             path=paste0("../output/maxent_outputs4_cross"),
                             args=prepPara(userfeatures="LQ",
                                           betamultiplier=1,
                                           doclamp = TRUE,
                                           projectionlayers="/Users/iel82user/Google Drive/1_osu_lab/pro
                                           replicates=5, ## 5 replicates
                                           replicatetype="crossvalidate") ) ##possible values are: cross
```