SDM for future scenarion Bristish columbia area

..

12/15/2021

```
# Close all connections and remove existing variables

rm(list = ls())
```

Adding R libraries

```
library(tidyverse) # normal data manipulating
library(dismo) # maxent
library(randomForest) # random forest
library(rpart) # rtree
library(kernlab) # support vector
library(raster) # raster files manipulating
library(rgdal) # for shapefile
```

Date from Future climate projection from Canadian Earth System Model version 5 (CanESM5)

2021-240————————————

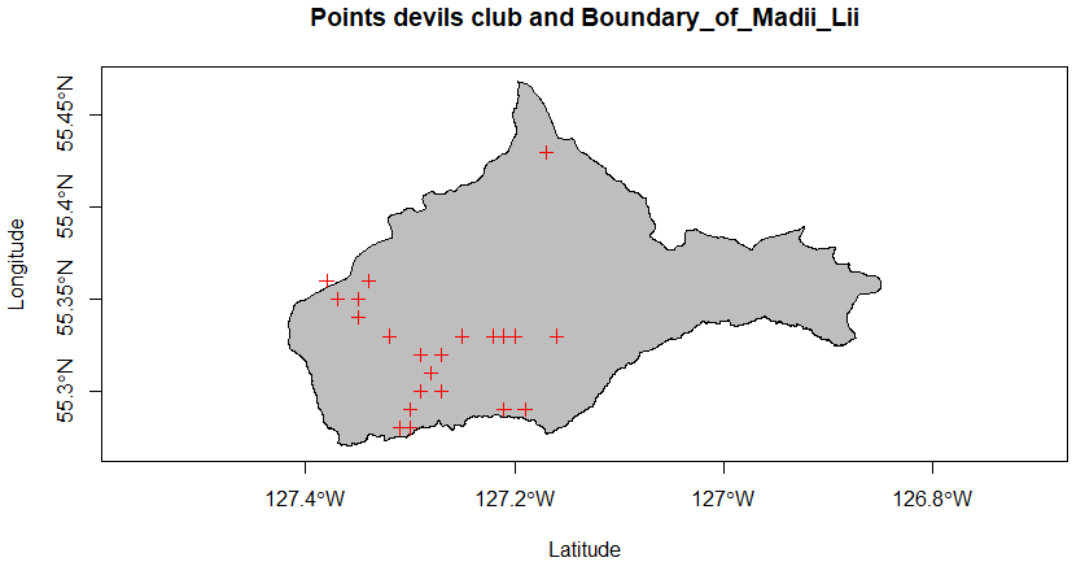
Data Processing for ssp126-2021-2040

```
# loading shapefiles of Boundary_WGS84
study <- readOGR('E:/online
work/SDM/newsdm15_2/shapefile/Boundary_of_Madii_Lii.shp')

## OGR data source with driver: ESRI Shapefile
## Source: "E:\online
work\SDM\newsdm15_2\shapefile\Boundary_of_Madii_Lii.shp", layer:
"Boundary_of_Madii_Lii"
## with 1 features
## It has 13 fields
## Integer64 fields read as strings:  OBJECTID

# Load the data [latitude longitude]
data <- read_csv("E:/online work/SDM/newsdm15_2/Book1.csv")
```

Making Map of species distribution

```
newproj<-"+proj=longlat +datum=WGS84"
study<- spTransform(study,crs(newproj))
#making point from lat and long
coord<- SpatialPoints(cbind(data$Long,data$Lat),proj4string =
CRS("+proj=longlat"))
```

```
plot(study,axes=TRUE,col='grey',border='black',xlab="Latitude",
ylab="Longitude",
     main="Points devils club and Boundary_of_Madii_Lii",xlim=c(-127.4168,-
126.8487),ylim=c(55.27017,55.46836 ))
box()
plot(coord,col='red',add=TRUE,xlim=c(-127.4168,-
126.8487),ylim=c(55.27017,55.46836 ))
```



1. Reading First scenario of 2021-2041 for ssp245

the resolution of original data was too low, we used downscale method so that we cen get higher resolution data.

```
biolayer<- Sys.glob('E:/online
work/SDM/newsdm15_2/ssp245/downscale/2021_40/*.tif')
# make it raster layers with all , it is called stack or brick
biolayer<-brick(lapply(paste(biolayer,sep=''),raster))
```

Extraction their values from raster layers of spp245

```
predictors<- biolayer
# extract value from predictors rasters for the present
presvals <- raster::extract(predictors, coord@coords)
```

Use identifiable name variables, This is very critical,you need to verify variables

names and assigning names. I am giving 1- 19 their short names ## Find their name [Variable name,click here] (https://www.worldclim.org/data/bioclim.html) I used short name

1.'AMTemp','MDRange','Isothermal', 4 'Tseasonal','MxTempW','MnTempC','TARange', 8.'MTempWt','MTempD','MTempW','MTempC', 12.'Apreci','PreWm','PreDM','Pseasonal', 16.'PreWtQ','PreDQ','PreWQ','PreCQ'

Renaming 19 Bioclimatic variables, when I am preparing script, variables were disorder, I need to do, I am ordering them according to raster layer # check in console typing "names(biolayer)"

```
climate_variables<- c('AMTemp','MTempW',
                      'MTempC','Apreci','PreWm','PreDM','Pseasonal',
'PreWtQ','PreDQ','PreWQ','PreCQ','MDRange','Isothermal','Tseasonal','MxTempW'
,'MnTempC','TARange','MTempWt','MTempD')
length(climate_variables)

## [1] 19

colnames(presvals)<- climate_variables
```

Generating pseudo absence values random points

```
set.seed(0)
# generate random points
backgr <- randomPoints(predictors, 5000)

## Warning in randomPoints(predictors, 5000): changed n to ncell of the mask
## (extent)

## Warning in randomPoints(predictors, 5000): generated random points =
## 0.442401960784314 times requested number

# extract data value from predictors for the absent points
absvals <- raster::extract(predictors, backgr)
# Rename with identifiable Acronym
colnames(absvals)<- climate_variables
# generate the training set for Observed Devilsclubs [present data ]
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
# convert into a data frame
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
# view some extract Bioclimate variable for Devilsclubs in the
head(sdmdata )

##   pb   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 5.509173 14.01654 -5.729412 798.3729 100.5418 35.51082  28.84465
278.6028
## 2  1 5.509173 14.01654 -5.729412 798.3729 100.5418 35.51082  28.84465
278.6028
## 3  1 5.509173 14.01654 -5.729412 798.3729 100.5418 35.51082  28.84465
278.6028
## 4  1 5.509173 14.01654 -5.729412 798.3729 100.5418 35.51082  28.84465
278.6028
## 5  1 5.650334 14.14806 -5.729412 787.3924 100.2354 35.51082  29.00895
278.6816
## 6  1 5.650334 14.14806 -5.729412 787.3924 100.2354 35.51082  29.00895
278.6816
##      PreDQ    PreWQ   PreCQ  MDRange Isothermal Tseasonal  MxTempW
```

```
MnTempC
## 1 129.5256 205.1483 174.106 9.759814   29.72275  831.7426 23.16986 -
11.47551
## 2 129.5256 205.1483 174.106 9.759814   29.72275  831.7426 23.16986 -
11.47551
## 3 129.5256 205.1483 174.106 9.759814   29.72275  831.7426 23.16986 -
11.47551
## 4 129.5256 205.1483 174.106 9.759814   29.72275  831.7426 23.16986 -
11.47551
## 5 129.1353 203.3728 172.877 9.761105   29.72275  831.7426 23.16986 -
11.33336
## 6 129.1353 203.3728 172.877 9.761105   29.72275  831.7426 23.16986 -
11.33336
##    TARange  MTempWt     MTempD
## 1 32.56379 5.444732 0.08860078
## 2 32.56379 5.444732 0.08860078
## 3 32.56379 5.444732 0.08860078
## 4 32.56379 5.444732 0.08860078
## 5 32.56379 5.563607 0.08860078
## 6 32.56379 5.563607 0.08860078
```

K-fold cross validation

It is so much important for Predictive model. It ensured that each observation from training and testing dataset (Presence or pseudo Absence) has the chance of appearing.K-Fold CV is where a given data set is split into a K number of sections/folds where each fold is used as a testing set at some point. Lets take the scenario of 5-Fold cross validation(K=5).This process is repeated until each fold of the 5 folds have been used as the testing set

```
set.seed(0)
# make 5 k-fold cross validation
group <- kfold(coord@coords, 5)
# split into training and test set
pres_train <- coord@coords[group != 1, ]
pres_test <- coord@coords[group == 1, ]
# extract value from predictors rasters for the present
ext <- extent(study)

# add seed number for same result for every run time
set.seed(10)
# generate random points, we will restrict the background points within 12.5%
of extent
backg <- randomPoints(predictors, n=5000, ext=ext, extf = 1.25)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
changed n
## to ncell of the mask (extent)
```

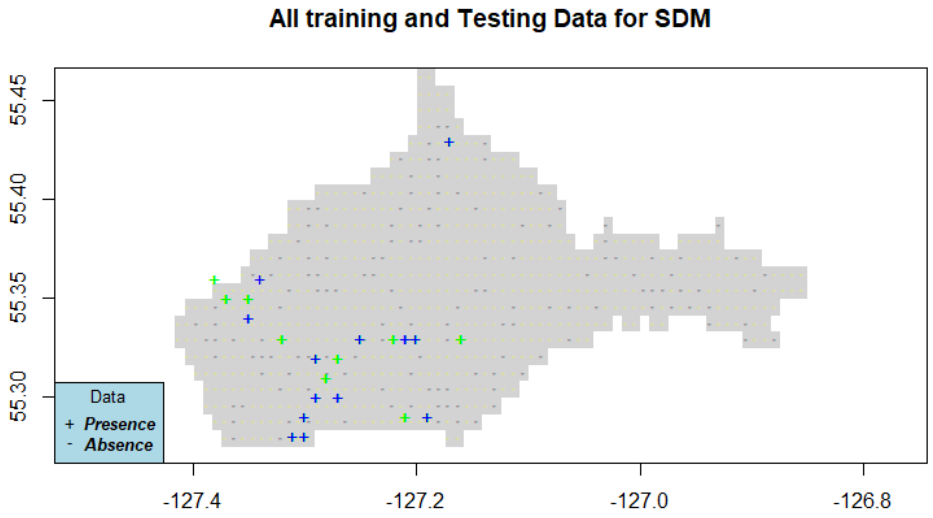
```
## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
generated
## random points = 0.442401960784314 times requested number

colnames(backg) = c('lon', 'lat')
group <- kfold(backg, 5)
backg_train <- backg[group != 1, ]
backg_test <- backg[group == 1, ]
```

View some rows of background train

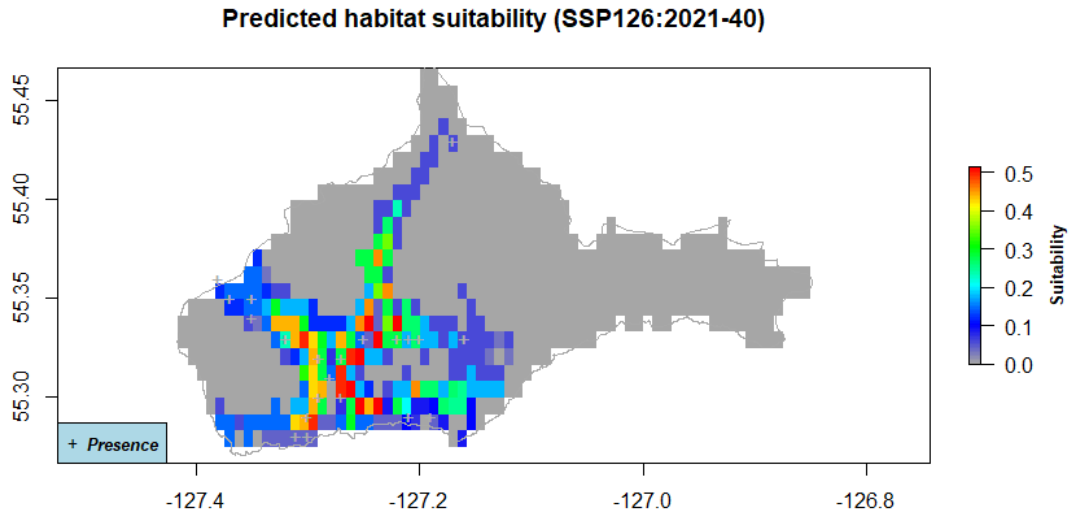
```
head(backg_train,2)

##            lon      lat
## [1,] -127.1042 55.32917
## [2,] -127.0542 55.34583
```

Now see what they are produced and where they are. we will use several models, then we will choose best performing models.

```
# make raster
r <- raster(predictors, 1)
# plot raster with removing na values
plot(!is.na(r), col=c('white', 'light grey'), legend=FALSE,main='All training
and Testing Data for SDM')
# plot the training points for absence
points(backg_train, pch='-', cex=0.5, col='yellow')
# plot the test points
points(backg_test, pch='-',  cex=0.5, col='black')
# Training Presence for absence
points(pres_train, pch= '+', col='green')
# Testing Presence
points(pres_test, pch='+', col='blue')
legend(x="bottomleft",legend = c('Presence','Absence'),pch=c('+','-'),
       cex=0.8,title="Data", text.font=4, bg='lightblue')
```




```
# the profiles method basic Suitability Modeling: they are presence only
methods
bc <- bioclim(predictors, pres_train)
# prediction of habitat suitability
pb <- predict(predictors, bc, ext=ext, progress='')
# Making color palette from gray to red
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pb,col=pal(100), main='Predicted habitat suitability (SSP126:2021-40)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
#plot(pb > tr, main='Presence/Absence')
plot(study, add=TRUE, border='dark grey')
points(pres_train, pch='+',col='grey67')
legend(x="bottomleft",legend = c('Presence'),pch=c('+'),
       cex=0.8, text.font=4, bg='lightblue')
```



###

Model evaluation , we have received good AUC values Here , n presences are training set which was generated from presence data [we splitted it into two parts, training and testing]. The absence data are not real absence. it is produced for the sake of measuring model accuracy. It is valid way actually.

```
evaluate(bc, p=pres_test, a=backg_test, x=predictors)

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.7853009
## cor            : 0.1940929
## max TPR+TNR at : 0.02931176
```

Receiving operating characteristics (ROC) and area under curve (AUC) are based on the confusion matrix . the confusion matrix is computed converting the habitat suitability maps into binary presence and absence map.shyyyy ### save raster file in the local (SSP245:2021-40) Low

```
writeRaster(pb, 'E:/online
work/SDM/newsdm15_2/ssp245/predicted/SSP245_2021_40_b.tif', overwrite=TRUE)
```

Apply MaxEnt (Maximum Entropy from Phillips et al., 2006).

We used Profile method but it only considers ‘presence’ data and does consider not absence or background data.Now we are going use machine learning methods use both presence and absence or background data.

```
# train data
train <- rbind(pres_train, backg_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(backg_train)))
# Data extraction from raster scenario : SSP126_2021_40
```

```
envtrain <- raster::extract(predictors, train)
# Transforming Dataframe
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
# Changing column names
colnames(envtrain)[2:20]<- climate_variables
# view 2 rows
head(envtrain,2)
```

```
##   pa   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 5.509173 14.01654 -5.729412 798.3729 100.5418 35.51082  28.84465
278.6028
## 2  1 5.509173 14.01654 -5.729412 798.3729 100.5418 35.51082  28.84465
278.6028
##      PreDQ    PreWQ   PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 129.5256 205.1483 174.106 9.759814   29.72275  831.7426 23.16986 -
11.47551
## 2 129.5256 205.1483 174.106 9.759814   29.72275  831.7426 23.16986 -
11.47551
##    TARange  MTempWt     MTempD
## 1 32.56379 5.444732 0.08860078
## 2 32.56379 5.444732 0.08860078
```

```
# testing data
testpres <- data.frame(raster::extract(predictors, pres_test) )
# Changing column names
colnames(testpres)<- climate_variables
#view top 2 rows
head(testpres,2)
```

```
##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 5.509173 14.01654 -5.729412 798.3729 100.5418 35.51082  28.84465
278.6028
## 2 5.720914 14.01654 -4.436096 789.4916 100.0822 35.45531  29.00895
278.6816
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 129.5256 205.1483 174.1060 9.759814   29.72275  831.7426 23.16986 -
11.47551
## 2 128.9401 202.7810 171.6481 9.758523   29.73263  831.7426 23.16986 -
11.47551
##    TARange  MTempWt     MTempD
## 1 32.56379 5.444732 0.08860078
## 2 32.56379 5.563607 0.08860078
```

Testing set for Absence data [randomly generated]

```
testbackg <- data.frame( raster::extract(predictors, backg_test) )
# Changing column names
```

```
colnames(testbackg)<- climate_variables
# View top 2 rows
head(testbackg,2)
```

```
##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 3.391768 14.41109 -7.022727 817.9117 105.1370 37.06514  28.68036
278.9440
## 2 3.532928 14.41109 -5.729412 810.9681 104.6775 37.00963  28.84465
279.0227
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal MxTempW
MnTempC
## 1 127.5740 222.9034 206.0582 9.743034   29.71286  823.5709 20.4192 -
11.19122
## 2 126.7934 221.1279 203.6003 9.740454   29.71286  823.5709 20.4192 -
11.19122
##    TARange  MTempWt    MTempD
## 1 32.74643 3.899362 -1.087001
## 2 32.74643 3.899362 -1.087001
```

Fit with Maxent modeling for Habitat suitability

```
maxent()
```

```
## Loading required namespace: rJava
```

```
## This is MaxEnt version 3.4.3
```

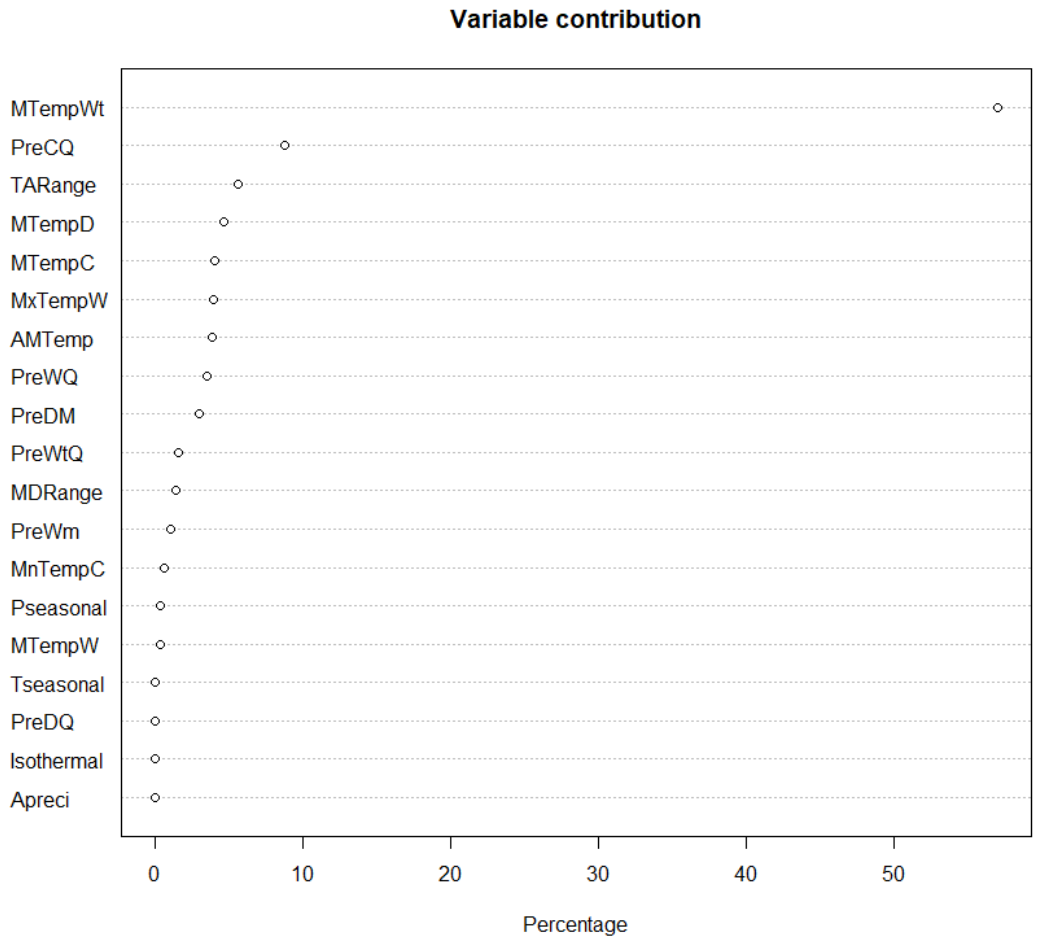
```
names(predictors)<- climate_variables
## Loading required namespace: rJava
xm <- maxent(predictors, pres_train)
```

```
## Warning in .local(x, p, ...): 1 (4.76%) of the presence points have NA
predictor
## values
```

```
## Warning in .local(x, p, ...): only got:702random background point values;
Small
## exent? Or is there a layer with many NA values?
```

```
## This is MaxEnt version 3.4.3
```

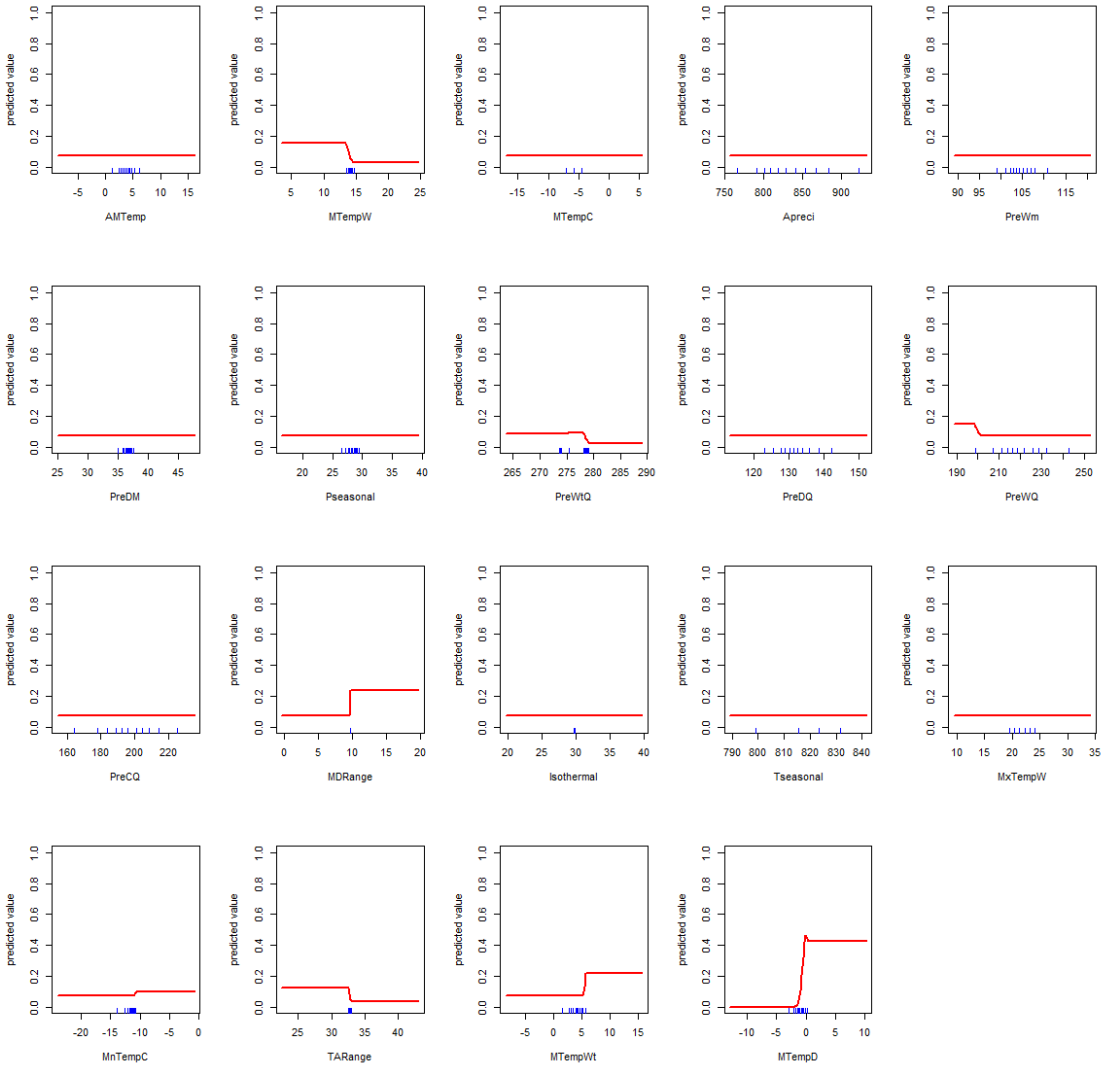
```
## This is MaxEnt version 3.4.3
plot(xm)
```



This will show what are variables are so important for modeling. they are MTempWt,PreCQ,TARange,MTempD,MTempC,MxTempW,AMTemp,PreDQ 1.MTempWt:Mean Temperature of Wettest Quarter 2.PreCQ: Precipitation of Coldest Quarter 3.TARange:Temperature Annual Range 4.MTempD:Mean Temperature of Driest Quarter 5.MTempC:Mean Temperature of Warmest Quarter 6.AMTemp: Annual Mean Temperature 7.PreDQ:Precipitation of Driest Quarter 8.MxTempW:Max Temperature of Warmest Month

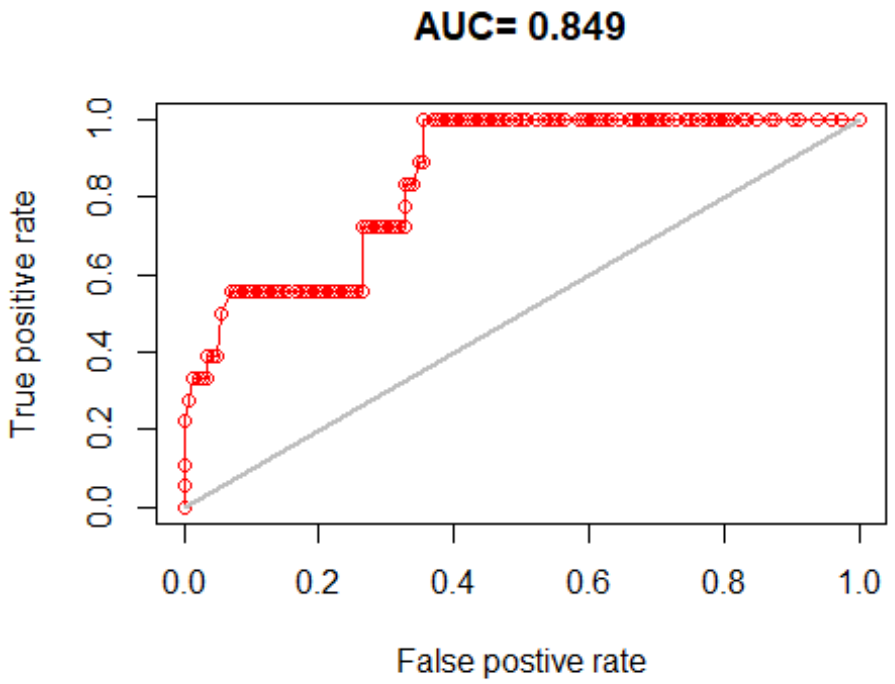
###A response plot:

```
response(xm)
```



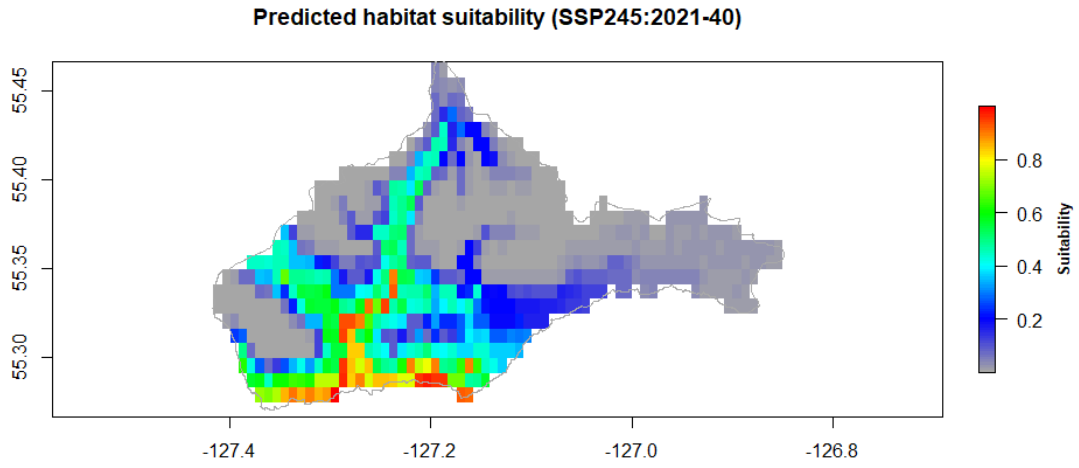
Maxent model evaluation values and Plot with AUC values

```
e <- evaluate(pres_test, backg_test, xm, predictors)
e

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.8493441
## cor            : 0.4484789
## max TPR+TNR at : 0.2340791

plot(e,'ROC')
```



The Maxent model is the better than profile model(bioclime) becuase its AUC is 0.92 Now, Predict from Maxent model

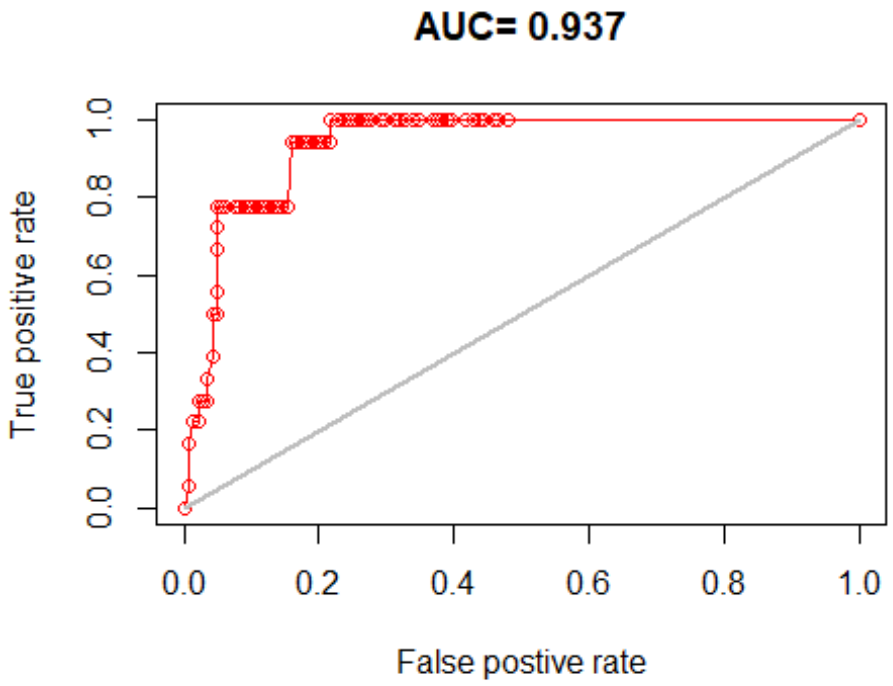
```
px <- predict(predictors, xm, ext=ext, progress='')
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(px, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted
habitat suitability (SSP245:2021-40)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



Now checking for the random forest Machine learning approach.It is used widely used. The

Random Forest (Breiman, 2001b) method is an extension of Classification and regression trees (CART; Breiman et al., 1984). This will show what are variables are so important for modeling. they are MTempWt,PreCQ,TARange,MTempD,MTempC,MxTempW,AMTemp,PreDQ

```
# removing missing values
envtrain<- na.omit(envtrain)
rf1 <-
randomForest(pa~MTempWt+PreCQ+TARange+MTempD+MTempC+MxTempW+AMTemp+PreDQ,
data=envtrain)
erf <- evaluate(testpres, testbackg, rf1)
erf

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.9371142
## cor            : 0.6198775
## max TPR+TNR at : 0.0612254

plot(erf,'ROC')
```



Looks promising AUC (0.97).Genenerally 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding. Now Maxent and random forest are the best performing.

Now Prediction for the random forest model

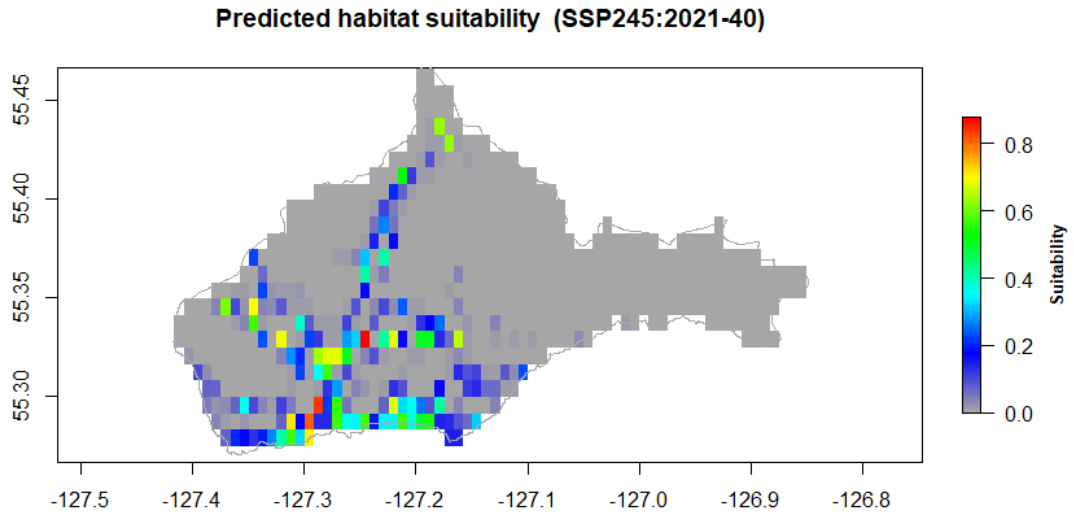
```
pr1 <- predict(predictors, rf1, ext=ext)
```

Save raster file random forest predicted in the local (SSP126:2021-40)

```
writeRaster(pr1, 'E:/online
work/SDM/newsdm15_2/predicted/SSP245_2021_40_random forest.tif',
overwrite=TRUE)
```

Map of the predicted habitat suitability from random forest

```
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pr1, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted
habitat suitability  (SSP245:2021-40)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



##Support Vector Machines (SVMs; Vapnik, 1998)

```
svm<- ksvm(pa~MTempWt+PreCQ+TARange+MTempD+MTempC+MxTempW+AMTemp+PreDQ,
data=envtrain)
esvm <- evaluate(testpres, testbackg, svm)
esvm

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.8746142
## cor            : 0.4081064
## max TPR+TNR at : 0.03057114
```

AUC of support vector machine is same as maxent. Now we have decided that we will use random forest method for the rest of the scenarios.

#2 ### Data Processing for :ssp245-2041-2060

```
# read layers
biolayer<- Sys.glob('E:/online
work/SDM/newsdm15_2/ssp245/downscale/2041_60/*.tif')
biolayer<-brick(lapply(paste(biolayer,sep=''),raster))
```

Extraction their values

```
# reassign as predictors
predictors<- biolayer
# extract value from predictors rasters for the present
presvals <- raster::extract(predictors, coord@coords)

# assign correct short names
colnames(presvals)<- climate_variables
```

Generating pseudo absence values random point in the

```
set.seed(0)
# generate random points
backgr <- randomPoints(predictors, 5000)

## Warning in randomPoints(predictors, 5000): changed n to ncell of the mask
## (extent)

## Warning in randomPoints(predictors, 5000): generated random points =
## 0.442401960784314 times requested number

# extract data value from predictors for the absent points
absvals <- raster::extract(predictors, backgr)
# Rename with identifiable Acronym
colnames(absvals)<- climate_variables
# generate the training set for Observed Devilsclubs [present data ]
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
# convert into a data frame
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
# view some extract Bioclimate variable for Devilsclubs in the
head(sdmdata )

##   pb   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 6.837648 15.30695 -4.181373 841.6835 108.2375 36.66298  29.70530
294.9459
## 2  1 6.837648 15.30695 -4.181373 841.6835 108.2375 36.66298  29.70530
294.9459
## 3  1 6.837648 15.30695 -4.181373 841.6835 108.2375 36.66298  29.70530
294.9459
## 4  1 6.837648 15.30695 -4.181373 841.6835 108.2375 36.66298  29.70530
294.9459
## 5  1 6.978954 15.43753 -4.181373 830.2888 107.9159 36.66298  29.88137
295.0347
## 6  1 6.978954 15.43753 -4.181373 830.2888 107.9159 36.66298  29.88137
295.0347
```

```
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 133.4288 216.6454 182.9133 9.585340   29.84047  826.0107 24.39612 -
9.521162
## 2 133.4288 216.6454 182.9133 9.585340   29.84047  826.0107 24.39612 -
9.521162
## 3 133.4288 216.6454 182.9133 9.585340   29.84047  826.0107 24.39612 -
9.521162
## 4 133.4288 216.6454 182.9133 9.585340   29.84047  826.0107 24.39612 -
9.521162
## 5 133.0197 214.8053 181.6309 9.586892   29.84047  826.0107 24.39612 -
9.378476
## 6 133.0197 214.8053 181.6309 9.586892   29.84047  826.0107 24.39612 -
9.378476
##   TARange  MTempWt   MTempD
## 1 31.8322 6.711839 1.204645
## 2 31.8322 6.711839 1.204645
## 3 31.8322 6.711839 1.204645
## 4 31.8322 6.711839 1.204645
## 5 31.8322 6.829849 1.204645
## 6 31.8322 6.829849 1.204645
```

k-fold cross validation

```
set.seed(0)
# make 5 k-fold cross validation
group <- kfold(coord@coords, 5)
# split into training and test set
pres_train <- coord@coords[group != 1, ]
#presence test
pres_test <- coord@coords[group == 1, ]
# extract value from predictors rasters for the present
ext <- extent(study)
```

add seed number for same result for every run time

```
set.seed(10)
# generate random points, we will restrict the background points within 12.5%
of extent
backg <- randomPoints(predictors, n=5000, ext=ext, extf = 1.25)
```

```
## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
changed n
## to ncell of the mask (extent)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
generated
## random points = 0.442401960784314 times requested number
```

```
# change column name
colnames(backg) = c('lon', 'lat')
group <- kfold(backg, 5)
```

```
# absence train
backg_train <- backg[group != 1, ]
# absence test
backg_test <- backg[group == 1, ]
```

View some rows of background train

```
head(backg_train,2)
```

```
##            lon      lat
## [1,] -127.1042 55.32917
## [2,] -127.0542 55.34583
```

```
# train data
train <- rbind(pres_train, backg_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(backg_train)))
# Data extraction from raster scenario : SSP126_2021_40
envtrain <- raster::extract(predictors, train)
# Transforming Dataframe
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
# Changing column names
colnames(envtrain)[2:20]<- climate_variables
# view 2 rows
head(envtrain,2)
```

```
##   pa   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 6.837648 15.30695 -4.181373 841.6835 108.2375 36.66298   29.7053
294.9459
## 2  1 6.837648 15.30695 -4.181373 841.6835 108.2375 36.66298   29.7053
294.9459
##      PreDQ    PreWQ    PreCQ MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 133.4288 216.6454 182.9133 9.58534   29.84047  826.0107 24.39612 -
9.521162
## 2 133.4288 216.6454 182.9133 9.58534   29.84047  826.0107 24.39612 -
9.521162
##   TARange  MTempWt   MTempD
## 1 31.8322 6.711839 1.204645
## 2 31.8322 6.711839 1.204645
```

```
# testing data
testpres <- data.frame(raster::extract(predictors, pres_test) )
# Changing column names
colnames(testpres)<- climate_variables
#view top 2 rows
head(testpres,2)
```

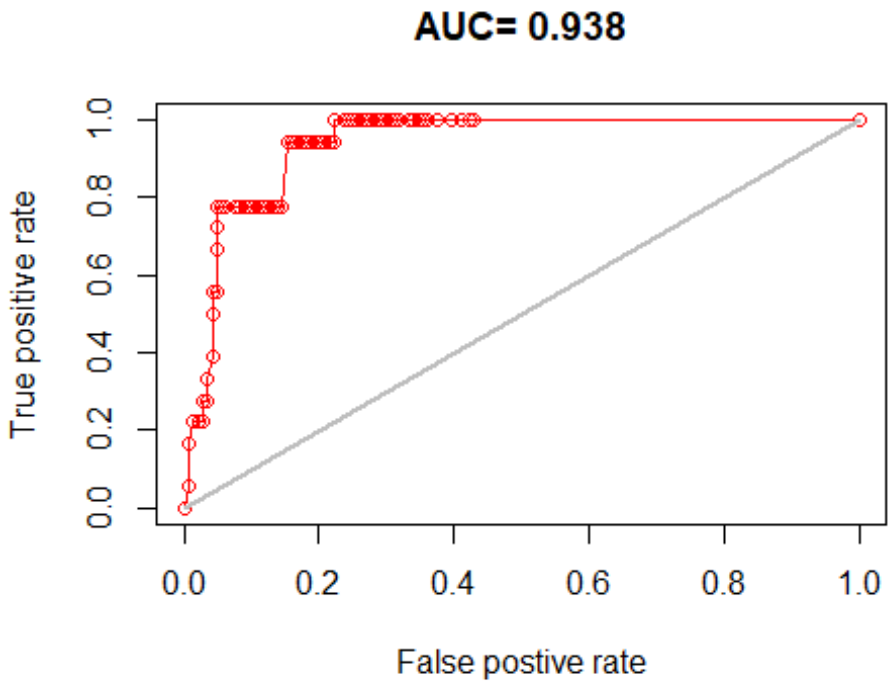
```
##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 6.837648 15.30695 -4.181373 841.6835 108.2375 36.66298  29.70530
```

```
294.9459
## 2 7.049608 15.30695 -2.877897 832.4672 107.7551 36.58691  29.88137
295.0347
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 133.4288 216.6454 182.9133 9.585340   29.84047  826.0107 24.39612 -
9.521162
## 2 132.8152 214.1919 180.3484 9.583789   29.84516  826.0107 24.39612 -
9.521162
##   TARange  MTempWt   MTempD
## 1 31.8322 6.711839 1.204645
## 2 31.8322 6.829849 1.204645
```

Testing set for Absence data [randomly generated]

```
testbackg <- data.frame( raster::extract(predictors, backg_test) )
# Changing column names
colnames(testbackg)<- climate_variables
# View top 2 rows
head(testbackg,2)

##     AMTemp  MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal   PreWtQ
## 1 4.718055 15.6987 -5.484848 861.9594 113.0612 38.79297  29.52923 295.3307
## 2 4.859361 15.6987 -4.181373 854.7538 112.5788 38.71690  29.70530 295.4195
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 131.3835 235.0468 216.2568 9.565172   29.83578  818.1239 21.64403 -
9.23579
## 2 130.5654 233.2066 213.6919 9.562070   29.83578  818.1239 21.64403 -
9.23579
##    TARange  MTempWt   MTempD
## 1 32.01786 5.177717 0.017931
## 2 32.01786 5.177717 0.017931

# Random forest
envtrain<- na.omit(envtrain)
rf1 <-
randomForest(pa~MTempWt+PreCQ+TARange+MTempD+MTempC+MxTempW+AMTemp+PreDQ,
data=envtrain)

## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?

# model evaluation
erf <- evaluate(testpres, testbackg, rf1)
erf

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.9378858
```

```
## cor             : 0.6072567
## max TPR+TNR at : 0.1627574

# ROC plot
plot(erf,'ROC')
```



Now Prediction for the random forest model

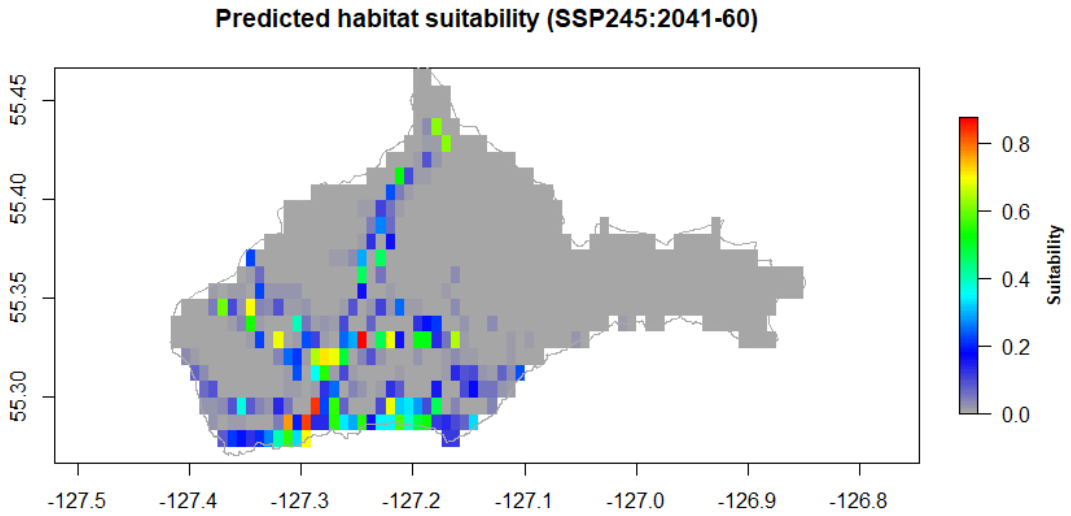
```
names(predictors)<- climate_variables
pr2 <- predict(predictors, rf1, ext=ext)
```

save raster file random forest predicted in the local (ssp245-2041-2060)

```
writeRaster(pr2, 'E:/online work/SDM/newsdm15_2/predicted/SSP245_2041_60_random forest.tif', overwrite=TRUE)
```

Map of the predicted habitat suitability from random forest

```
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pr2, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted habitat suitability (SSP245:2041-60)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



#3 ### Data Processing for :ssp245-2061-2080

```
# read layers
biolayer<- Sys.glob('E:/online
work/SDM/newsdm15_2/ssp245/downscale/2061_80/*.tif')
# make brick [multilayers]
biolayer<-brick(lapply(paste(biolayer,sep=''),raster))
```

Extraction their values

```
# reassign as predictors
predictors<- biolayer
names(predictors)<- climate_variables
# extract value from predictors rasters for the present
presvals <- raster::extract(predictors, coord@coords)

# assign correct short names
colnames(presvals)<- climate_variables
```

Generating pseudo absence values random points

```
set.seed(0)
# generate random points
backgr <- randomPoints(predictors, 5000)

## Warning in randomPoints(predictors, 5000): changed n to ncell of the mask
## (extent)

## Warning in randomPoints(predictors, 5000): generated random points =
## 0.442401960784314 times requested number

# extract data value from predictors for the absent points
absvals <- raster::extract(predictors, backgr)
```

```r
# Rename with identifiable Acronym
colnames(absvals)<- climate_variables
# generate the training set for Observed Devilsclubs [present data ]
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
# convert into a data frame
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
# view some extract Bioclimate variable for Devilsclubs
head(sdmdata )
```

```
##   pb   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 7.834166 16.26295 -3.244118 870.5339 113.0975 38.38541  30.28587
305.8267
## 2  1 7.834166 16.26295 -3.244118 870.5339 113.0975 38.38541  30.28587
305.8267
## 3  1 7.834166 16.26295 -3.244118 870.5339 113.0975 38.38541  30.28587
305.8267
## 4  1 7.834166 16.26295 -3.244118 870.5339 113.0975 38.38541  30.28587
305.8267
## 5  1 7.975402 16.39421 -3.244118 858.7692 112.7493 38.38541  30.45611
305.9229
## 6  1 7.975402 16.39421 -3.244118 858.7692 112.7493 38.38541  30.45611
305.9229
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 136.3641 225.6921 186.6696 9.391451   29.39304  826.7591 25.29128 -
8.467213
## 2 136.3641 225.6921 186.6696 9.391451   29.39304  826.7591 25.29128 -
8.467213
## 3 136.3641 225.6921 186.6696 9.391451   29.39304  826.7591 25.29128 -
8.467213
## 4 136.3641 225.6921 186.6696 9.391451   29.39304  826.7591 25.29128 -
8.467213
## 5 135.9559 223.7355 185.3699 9.392643   29.39304  826.7591 25.29128 -
8.326760
## 6 135.9559 223.7355 185.3699 9.392643   29.39304  826.7591 25.29128 -
8.326760
##    TARange  MTempWt   MTempD
## 1 31.70635 7.683828 2.181124
## 2 31.70635 7.683828 2.181124
## 3 31.70635 7.683828 2.181124
## 4 31.70635 7.683828 2.181124
## 5 31.70635 7.802104 2.181124
## 6 31.70635 7.802104 2.181124
```

k-fold cross validation

```r
set.seed(0)
# make 5 k-fold cross validation
group <- kfold(coord@coords, 5)
# split into training and test set
```

```
pres_train <- coord@coords[group != 1, ]
#presence test
pres_test <- coord@coords[group == 1, ]
# extract value from predictors rasters for the present
ext <- extent(study)
```

add seed number for same result for every run time

```
set.seed(10)
# generate random points, we will restrict the background points within 12.5%
of extent
backg <- randomPoints(predictors, n=5000, ext=ext, extf = 1.25)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
changed n
## to ncell of the mask (extent)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
generated
## random points = 0.442401960784314 times requested number

# change column name
colnames(backg) = c('lon', 'lat')
group <- kfold(backg, 5)
# absence train
backg_train <- backg[group != 1, ]
# absence test
backg_test <- backg[group == 1, ]
```

View some rows of background train

```
head(backg_train,2)

##            lon      lat
## [1,] -127.1042 55.32917
## [2,] -127.0542 55.34583
```

creating train set

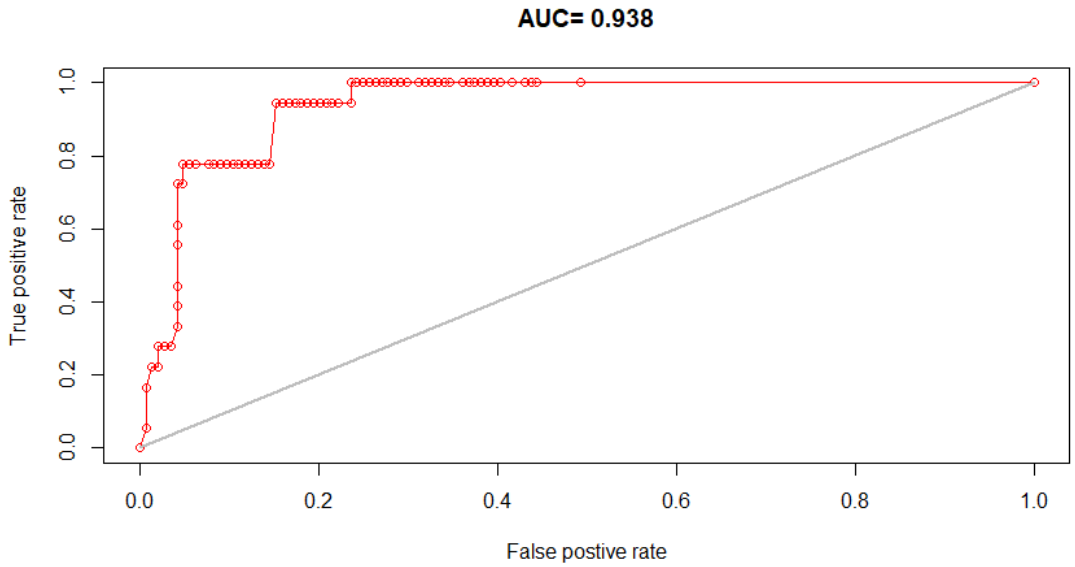
```
# train data
train <- rbind(pres_train, backg_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(backg_train)))
# Data extraction from raster scenario : SSP126_2021_40
envtrain <- raster::extract(predictors, train)
# Transforming Dataframe
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
# Changing column names
colnames(envtrain)[2:20]<- climate_variables
# view 2 rows
head(envtrain,2)
```

```
##   pa   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 7.834166 16.26295 -3.244118 870.5339 113.0975 38.38541  30.28587
305.8267
## 2  1 7.834166 16.26295 -3.244118 870.5339 113.0975 38.38541  30.28587
305.8267
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 136.3641 225.6921 186.6696 9.391451   29.39304  826.7591 25.29128 -
8.467213
## 2 136.3641 225.6921 186.6696 9.391451   29.39304  826.7591 25.29128 -
8.467213
##    TARange  MTempWt   MTempD
## 1 31.70635 7.683828 2.181124
## 2 31.70635 7.683828 2.181124

# testing data
testpres <- data.frame(raster::extract(predictors, pres_test) )
# Changing column names
colnames(testpres)<- climate_variables
#view top 2 rows
head(testpres,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 7.834166 16.26295 -3.244118 870.5339 113.0975 38.38541  30.28587
305.8267
## 2 8.046020 16.26295 -1.951872 861.0183 112.5752 38.29222  30.45611
305.9229
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 136.3641 225.6921 186.6696 9.391451   29.39304  826.7591 25.29128 -
8.467213
## 2 135.7519 223.0833 184.0701 9.390259   29.40336  826.7591 25.29128 -
8.467213
##    TARange  MTempWt   MTempD
## 1 31.70635 7.683828 2.181124
## 2 31.70635 7.802104 2.181124
```

Testing set for Absence data [randomly generated]

```
testbackg <- data.frame( raster::extract(predictors, backg_test) )
# Changing column names
colnames(testbackg)<- climate_variables
# View top 2 rows
head(testbackg,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 5.715621 16.65673 -4.536364 891.4681 118.3204 40.99458  30.11564
306.2437
```

```
## 2 5.856858 16.65673 -3.244118 884.0287 117.7981 40.90140  30.28587
306.3399
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 134.3235 245.2584 220.4628 9.375950   29.38271  818.4671 22.54995 -
8.186309
## 2 133.5073 243.3018 217.8633 9.373565   29.38271  818.4671 22.54995 -
8.186309
##    TARange  MTempWt   MTempD
## 1 31.86786 6.146236 1.002638
## 2 31.86786 6.146236 1.002638

# Random forest
envtrain<- na.omit(envtrain)
rf1 <-
randomForest(pa~MTempWt+PreCQ+TARange+MTempD+MTempC+MxTempW+AMTemp+PreDQ,
data=envtrain)

## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?

# model evaluation
erf <- evaluate(testpres, testbackg, rf1)
erf

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.9382716
## cor            : 0.6148736
## max TPR+TNR at : 0.1687366

# ROC plot
plot(erf,'ROC')
```



Now

Prediction for the random forest model

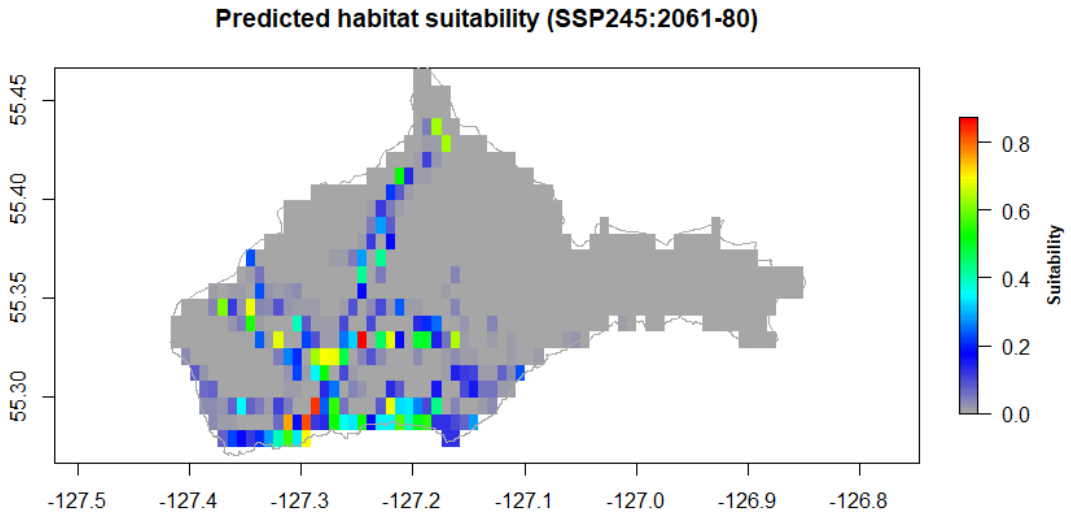
```
pr3 <- predict(predictors, rf1, ext=ext)
```

save raster file random forest predicted in the local (ssp245-2061-2080)

```
writeRaster(pr3, 'E:/online
work/SDM/newsdm15_2/predicted/SSP245_2061_80_random forest.tif',
overwrite=TRUE)
```

Map of the predicted habitat suitability from random forest

```
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pr3, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted
habitat suitability (SSP245:2061-80)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



#4 ### Data Processing for :ssp245-2081-2100

```
# read layers
biolayer<- Sys.glob('E:/online
work/SDM/newsdm15_2/ssp245/downscale/2081_2100/*.tif')
# make brick [multilayers]
biolayer<-brick(lapply(paste(biolayer,sep=''),raster))
```

Extraction their values

```
# reassign as predictors
predictors<- biolayer
names(predictors)<- climate_variables
# extract value from predictors rasters for the present
presvals <- raster::extract(predictors, coord@coords)

# assign correct short names
colnames(presvals)<- climate_variables
```

Generating pseudo absence values random point in the

```
set.seed(0)
# generate random points
backgr <- randomPoints(predictors, 5000)

## Warning in randomPoints(predictors, 5000): changed n to ncell of the mask
## (extent)

## Warning in randomPoints(predictors, 5000): generated random points =
## 0.442401960784314 times requested number

# extract data value from predictors for the absent points
absvals <- raster::extract(predictors, backgr)
```

```r
# Rename with identifiable Acronym
colnames(absvals)<- climate_variables
# generate the training set for Observed Devilsclubs [present data ]
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
# convert into a data frame
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
# view some extract Bioclimate variable for Devilsclubs
head(sdmdata )
```

```
##   pb   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 8.576319 16.86812 -2.321568 897.1217 116.1401 38.99702  30.34110
314.9118
## 2  1 8.576319 16.86812 -2.321568 897.1217 116.1401 38.99702  30.34110
314.9118
## 3  1 8.576319 16.86812 -2.321568 897.1217 116.1401 38.99702  30.34110
314.9118
## 4  1 8.576319 16.86812 -2.321568 897.1217 116.1401 38.99702  30.34110
314.9118
## 5  1 8.717624 16.99677 -2.321568 885.0045 115.7852 38.99702  30.50929
315.0045
## 6  1 8.717624 16.99677 -2.321568 885.0045 115.7852 38.99702  30.50929
315.0045
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 140.9368 230.6743 194.1588 9.265616   29.20835  818.3338 25.84083 -
7.699549
## 2 140.9368 230.6743 194.1588 9.265616   29.20835  818.3338 25.84083 -
7.699549
## 3 140.9368 230.6743 194.1588 9.265616   29.20835  818.3338 25.84083 -
7.699549
## 4 140.9368 230.6743 194.1588 9.265616   29.20835  818.3338 25.84083 -
7.699549
## 5 140.5288 228.6889 192.7846 9.266988   29.20835  818.3338 25.84083 -
7.559233
## 6 140.5288 228.6889 192.7846 9.266988   29.20835  818.3338 25.84083 -
7.559233
##    TARange  MTempWt   MTempD
## 1 31.48216 8.326390 3.017774
## 2 31.48216 8.326390 3.017774
## 3 31.48216 8.326390 3.017774
## 4 31.48216 8.326390 3.017774
## 5 31.48216 8.445184 3.017774
## 6 31.48216 8.445184 3.017774
```

k-fold cross validation

```r
set.seed(0)
# make 5 k-fold cross validation
group <- kfold(coord@coords, 5)
# split into training and test set
```

```
pres_train <- coord@coords[group != 1, ]
#presence test
pres_test <- coord@coords[group == 1, ]
# extract value from predictors rasters for the present
ext <- extent(study)
```

add seed number for same result for every run time

```
set.seed(10)
# generate random points, we will restrict the background points within 12.5%
of extent
backg <- randomPoints(predictors, n=5000, ext=ext, extf = 1.25)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
changed n
## to ncell of the mask (extent)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
generated
## random points = 0.442401960784314 times requested number

# change column name
colnames(backg) = c('lon', 'lat')
group <- kfold(backg, 5)
# absence train
backg_train <- backg[group != 1, ]
# absence test
backg_test <- backg[group == 1, ]
```

View some rows of background train

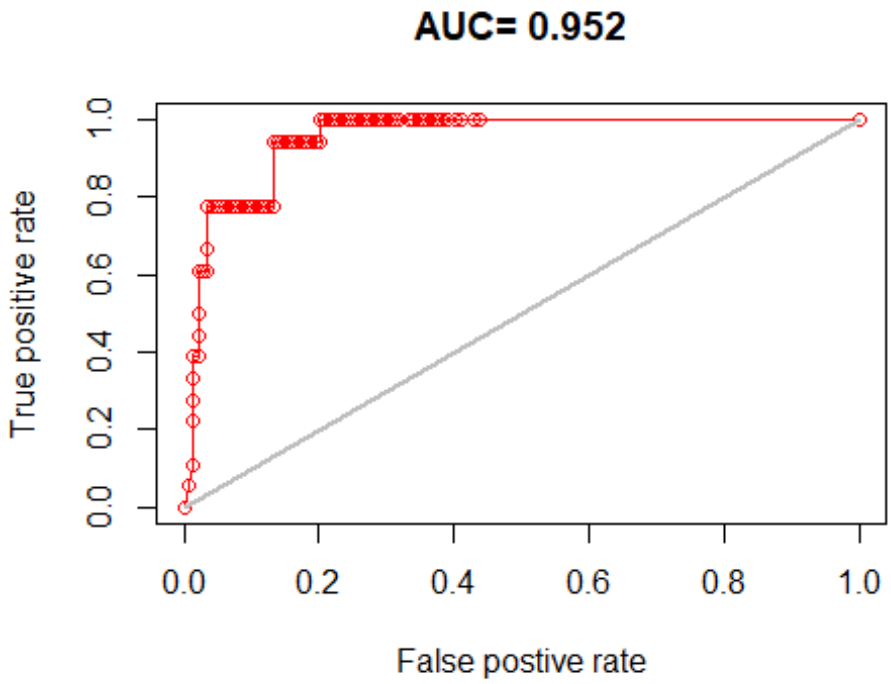
```
head(backg_train,2)

##            lon      lat
## [1,] -127.1042 55.32917
## [2,] -127.0542 55.34583

# train data
train <- rbind(pres_train, backg_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(backg_train)))
# Data extraction from raster scenario : SSP126_2021_40
envtrain <- raster::extract(predictors, train)
# Transforming Dataframe
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
# Changing column names
colnames(envtrain)[2:20]<- climate_variables
# view 2 rows
head(envtrain,2)

##   pa   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 8.576319 16.86812 -2.321568 897.1217 116.1401 38.99702   30.3411
```

```
314.9118
## 2  1 8.576319 16.86812 -2.321568 897.1217 116.1401 38.99702   30.3411
314.9118
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 140.9368 230.6743 194.1588 9.265616   29.20835  818.3338 25.84083 -
7.699549
## 2 140.9368 230.6743 194.1588 9.265616   29.20835  818.3338 25.84083 -
7.699549
##    TARange MTempWt   MTempD
## 1 31.48216 8.32639 3.017774
## 2 31.48216 8.32639 3.017774

# testing data
testpres <- data.frame(raster::extract(predictors, pres_test) )
# Changing column names
colnames(testpres)<- climate_variables
#view top 2 rows
head(testpres,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 8.576319 16.86812 -2.321568 897.1217 116.1401 38.99702  30.34110
314.9118
## 2 8.788276 16.86812 -1.026471 887.3210 115.6078 38.90931  30.50929
315.0045
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 140.9368 230.6743 194.1588 9.265616   29.20835  818.3338 25.84083 -
7.699549
## 2 140.3249 228.0271 191.4104 9.264245   29.21396  818.3338 25.84083 -
7.699549
##    TARange  MTempWt   MTempD
## 1 31.48216 8.326390 3.017774
## 2 31.48216 8.445184 3.017774
```

Testing set for Absence data [randomly generated]

```
testbackg <- data.frame( raster::extract(predictors, backg_test) )
# Changing column names
colnames(testbackg)<- climate_variables
# View top 2 rows
head(testbackg,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 6.456748 17.25408 -3.616667 918.6833 121.4631 41.45295  30.17292
315.3134
## 2 6.598053 17.25408 -2.321568 911.0209 120.9308 41.36524  30.34110
315.4060
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
```

```
MnTempC
## 1 138.8971 250.5287 229.8884 9.247788   29.20274   810.175 23.07354 -
7.418917
## 2 138.0812 248.5432 227.1399 9.245046   29.20274   810.175 23.07354 -
7.418917
##    TARange  MTempWt   MTempD
## 1 31.62857 6.782079 1.835272
## 2 31.62857 6.782079 1.835272

# Random forest
envtrain<- na.omit(envtrain)
rf1 <-
randomForest(pa~AMTemp+MTempD+PreCQ+MxTempW+MnTempC+MTempC+MTempW+Pseasonal+T
seasonal+MDRange, data=envtrain)

## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?

# model evaluation
erf <- evaluate(testpres, testbackg, rf1)
erf

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.9517747
## cor            : 0.6761039
## max TPR+TNR at : 0.2274768

# ROC plot
plot(erf,'ROC')
```



Now Prediction for the random forest model

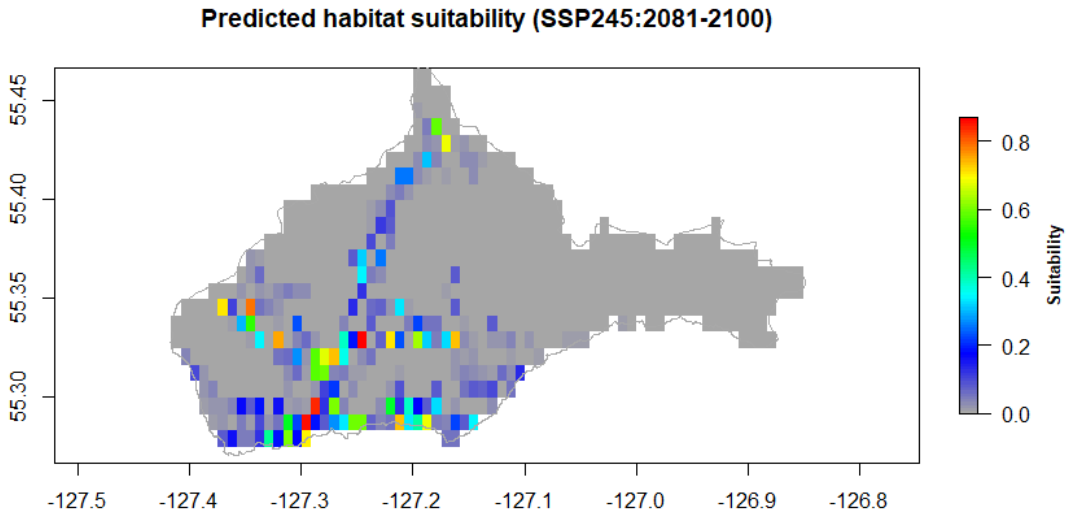
```
pr4 <- predict(predictors, rf1, ext=ext)
```

save raster file random forest predicted in the local (ssp245-2081-2100)

```
writeRaster(pr4, 'E:/online work/SDM/newsdm15_2/predicted/SSP245_2081_2100_random forest.tif', overwrite=TRUE)
```

Map of the predicted habitat suitability from random forest

```
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pr4, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted habitat suitability (SSP245:2081-2100)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



We only work on Medium and High forcing (ssp245 and ssp585)

#5 ### Data Processing for :ssp585-2021-2040

```
# read layers
biolayer<- Sys.glob('E:/online
work/SDM/newsdm15_2/ssp585/downscale/2021_40/*.tif')
# make brick [multilayers]
biolayer<-brick(lapply(paste(biolayer,sep=''),raster))
```

Extraction their values

```
# reassign as predictors
predictors<- biolayer
names(predictors)<- climate_variables
# extract value from predictors rasters for the present
presvals <- raster::extract(predictors, coord@coords)

# assign correct short names
colnames(presvals)<- climate_variables
```

Generating pseudo absence values random point in the

```
set.seed(0)
# generate random points
backgr <- randomPoints(predictors, 5000)

## Warning in randomPoints(predictors, 5000): changed n to ncell of the mask
## (extent)

## Warning in randomPoints(predictors, 5000): generated random points =
## 0.442401960784314 times requested number
```

```
# extract data value from predictors for the absent points
absvals <- raster::extract(predictors, backgr)
# Rename with identifiable Acronym
colnames(absvals)<- climate_variables
# generate the training set for Observed Devilsclubs [present data ]
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
# convert into a data frame
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
# view some extract Bioclimate variable for Devilsclubs in the
head(sdmdata )
```

```
##   pb   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 5.846623 14.38688 -5.237255 814.6859 103.4320 36.32305  29.39730
283.3717
## 2  1 5.846623 14.38688 -5.237255 814.6859 103.4320 36.32305  29.39730
283.3717
## 3  1 5.846623 14.38688 -5.237255 814.6859 103.4320 36.32305  29.39730
283.3717
## 4  1 5.846623 14.38688 -5.237255 814.6859 103.4320 36.32305  29.39730
283.3717
## 5  1 5.987736 14.51993 -5.237255 803.5814 103.1255 36.32305  29.55532
283.4494
## 6  1 5.987736 14.51993 -5.237255 803.5814 103.1255 36.32305  29.55532
283.4494
##      PreDQ    PreWQ    PreCQ MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 130.9769 211.3019 180.0695 9.76647   30.03062  832.3823 23.53762 -
10.79955
## 2 130.9769 211.3019 180.0695 9.76647   30.03062  832.3823 23.53762 -
10.79955
## 3 130.9769 211.3019 180.0695 9.76647   30.03062  832.3823 23.53762 -
10.79955
## 4 130.9769 211.3019 180.0695 9.76647   30.03062  832.3823 23.53762 -
10.79955
## 5 130.5999 209.4869 178.8159 9.76806   30.03062  832.3823 23.53762 -
10.65923
## 6 130.5999 209.4869 178.8159 9.76806   30.03062  832.3823 23.53762 -
10.65923
##    TARange  MTempWt    MTempD
## 1 32.27089 5.733197 0.2392894
## 2 32.27089 5.733197 0.2392894
## 3 32.27089 5.733197 0.2392894
## 4 32.27089 5.733197 0.2392894
## 5 32.27089 5.851650 0.2392894
## 6 32.27089 5.851650 0.2392894
```

k-fold cross validation

```
set.seed(0)
# make 5 k-fold cross validation
```

```
group <- kfold(coord@coords, 5)
# split into training and test set
pres_train <- coord@coords[group != 1, ]
#presence test
pres_test <- coord@coords[group == 1, ]
# extract value from predictors rasters for the present
ext <- extent(study)
```

add seed number for same result for every run time

```
set.seed(10)
# generate random points, we will restrict the background points within 12.5% of extent
backg <- randomPoints(predictors, n=5000, ext=ext, extf = 1.25)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
changed n
## to ncell of the mask (extent)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
generated
## random points = 0.442401960784314 times requested number

# change column name
colnames(backg) = c('lon', 'lat')
group <- kfold(backg, 5)
# absence train
backg_train <- backg[group != 1, ]
# absence test
backg_test <- backg[group == 1, ]
```

View some rows of background train

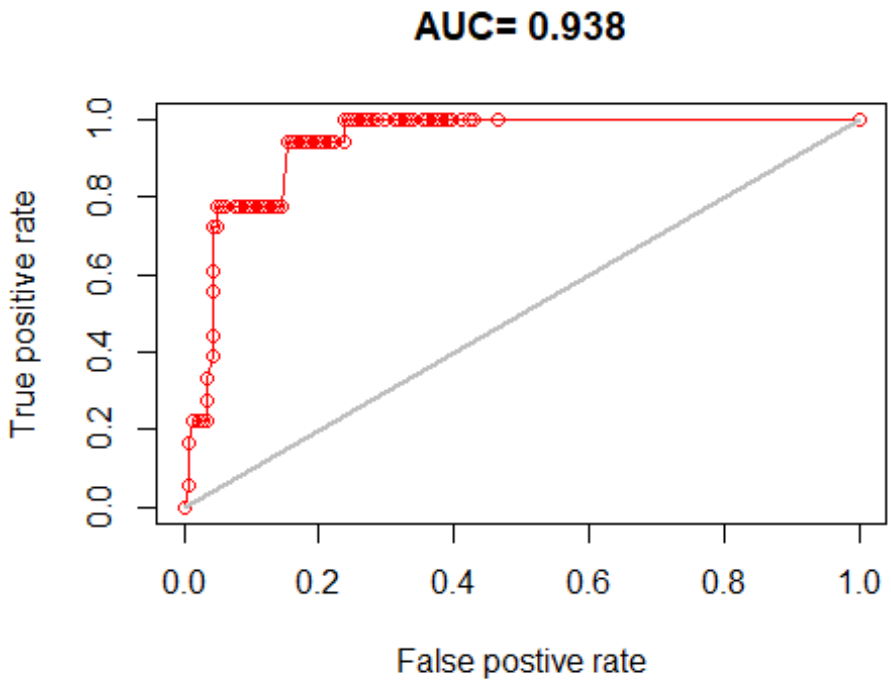
```
head(backg_train,2)

##            lon      lat
## [1,] -127.1042 55.32917
## [2,] -127.0542 55.34583

# train data
train <- rbind(pres_train, backg_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(backg_train)))
# Data extraction from raster scenario : SSP126_2021_40
envtrain <- raster::extract(predictors, train)
# Transforming Dataframe
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
# Changing column names
colnames(envtrain)[2:20]<- climate_variables
# view 2 rows
head(envtrain,2)
```

```
##   pa   AMTemp   MTempW    MTempC   Apreci   PreWm    PreDM Pseasonal
PreWtQ
## 1  1 5.846623 14.38688 -5.237255 814.6859 103.432 36.32305   29.3973
283.3717
## 2  1 5.846623 14.38688 -5.237255 814.6859 103.432 36.32305   29.3973
283.3717
##      PreDQ    PreWQ    PreCQ MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 130.9769 211.3019 180.0695 9.76647   30.03062  832.3823 23.53762 -
10.79955
## 2 130.9769 211.3019 180.0695 9.76647   30.03062  832.3823 23.53762 -
10.79955
##    TARange  MTempWt    MTempD
## 1 32.27089 5.733197 0.2392894
## 2 32.27089 5.733197 0.2392894

# testing data
testpres <- data.frame(raster::extract(predictors, pres_test) )
# Changing column names
colnames(testpres)<- climate_variables
#view top 2 rows
head(testpres,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 5.846623 14.38688 -5.237255 814.6859 103.4320 36.32305  29.39730
283.3717
## 2 6.058293 14.38688 -3.939661 805.7043 102.9722 36.26008  29.55532
283.4494
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 130.9769 211.3019 180.0695 9.766470   30.03062  832.3823 23.53762 -
10.79955
## 2 130.4115 208.8819 177.5623 9.764881   30.03862  832.3823 23.53762 -
10.79955
##    TARange  MTempWt    MTempD
## 1 32.27089 5.733197 0.2392894
## 2 32.27089 5.851650 0.2392894
```

Testing set for Absence data [randomly generated]

```
testbackg <- data.frame( raster::extract(predictors, backg_test) )
# Changing column names
colnames(testbackg)<- climate_variables
# View top 2 rows
head(testbackg,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 3.729928 14.78603 -6.534849 834.4453 108.0295 38.08615  29.23929
283.7085
```

```
## 2 3.871041 14.78603 -5.237255 827.4233 107.5697 38.02318  29.39730
283.7863
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 129.0919 229.4515 212.6635 9.745807   30.02261  824.3605 20.75631 -
10.51892
## 2 128.3380 227.6365 210.1563 9.742628   30.02261  824.3605 20.75631 -
10.51892
##    TARange  MTempWt     MTempD
## 1 32.41428 4.193304 -0.9427593
## 2 32.41428 4.193304 -0.9427593

# Random forest
envtrain<- na.omit(envtrain)
rf1 <-
randomForest(pa~MTempWt+PreCQ+TARange+MTempD+MTempC+MxTempW+AMTemp+PreDQ,
data=envtrain)

## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?

# model evaluation
erf <- evaluate(testpres, testbackg, rf1)
erf

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.9378858
## cor            : 0.6150403
## max TPR+TNR at : 0.1848691

# ROC plot
plot(erf,'ROC')
```



Now Prediction for the random forest model

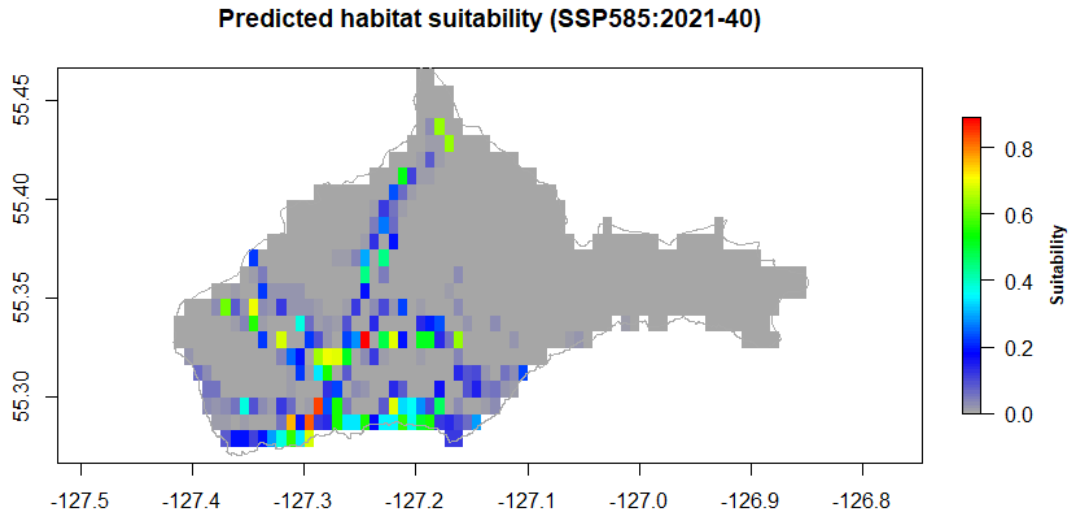
```
pr5 <- predict(predictors, rf1, ext=ext)
```

save raster file random forest predicted in the local (ssp585-2021-2040)

```
writeRaster(pr5, 'E:/online work/SDM/newsdm15_2/predicted/SSP585_2040_random forest.tif', overwrite=TRUE)
```

Map of the predicted habitat suitability from random forest

```
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pr5, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted habitat suitability (SSP585:2021-40)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



#6 ### Data Processing :ssp585-2041-2060

```
# read layers
biolayer<- Sys.glob('E:/online
work/SDM/newsdm15_2/ssp585/downscale/2041_60/*.tif')
# make brick [multilayers]
biolayer<-brick(lapply(paste(biolayer,sep=''),raster))
```

Extraction their bioclimatic variables values

```
# reassign as predictors
predictors<- biolayer
names(predictors)<- climate_variables
# extract value from predictors rasters for the present
presvals <- raster::extract(predictors, coord@coords)

# assign correct short names
colnames(presvals)<- climate_variables
```

Generating pseudo absence values random point in the

```
set.seed(0)
# generate random points
backgr <- randomPoints(predictors, 5000)

## Warning in randomPoints(predictors, 5000): changed n to ncell of the mask
## (extent)

## Warning in randomPoints(predictors, 5000): generated random points =
## 0.442401960784314 times requested number

# extract data value from predictors for the absent points
absvals <- raster::extract(predictors, backgr)
```

```r
# Rename with identifiable Acronym
colnames(absvals)<- climate_variables
# generate the training set for Observed Devilsclubs [present data ]
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
# convert into a data frame
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
# view some extract Bioclimate variable for Devilsclubs in the
head(sdmdata )
```

```
##   pb   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 7.872044 16.45744 -3.228431 863.6147 112.7023 37.32885  30.63447
306.6312
## 2  1 7.872044 16.45744 -3.228431 863.6147 112.7023 37.32885  30.63447
306.6312
## 3  1 7.872044 16.45744 -3.228431 863.6147 112.7023 37.32885  30.63447
306.6312
## 4  1 7.872044 16.45744 -3.228431 863.6147 112.7023 37.32885  30.63447
306.6312
## 5  1 8.013330 16.58659 -3.228431 851.8801 112.3498 37.32885  30.81558
306.7300
## 6  1 8.013330 16.58659 -3.228431 851.8801 112.3498 37.32885  30.81558
306.7300
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 135.7376 221.7675 185.9451 9.435161   29.26704  837.9612 25.54142 -
8.502540
## 2 135.7376 221.7675 185.9451 9.435161   29.26704  837.9612 25.54142 -
8.502540
## 3 135.7376 221.7675 185.9451 9.435161   29.26704  837.9612 25.54142 -
8.502540
## 4 135.7376 221.7675 185.9451 9.435161   29.26704  837.9612 25.54142 -
8.502540
## 5 135.3318 219.8673 184.6378 9.436659   29.26704  837.9612 25.54142 -
8.361817
## 6 135.3318 219.8673 184.6378 9.436659   29.26704  837.9612 25.54142 -
8.361817
##   TARange  MTempWt   MTempD
## 1 31.9893 7.717571 2.159852
## 2 31.9893 7.717571 2.159852
## 3 31.9893 7.717571 2.159852
## 4 31.9893 7.717571 2.159852
## 5 31.9893 7.835923 2.159852
## 6 31.9893 7.835923 2.159852
```

k-fold cross validation

```r
set.seed(0)
# make 5 k-fold cross validation
group <- kfold(coord@coords, 5)
# split into training and test set
```

```
pres_train <- coord@coords[group != 1, ]
#presence test
pres_test <- coord@coords[group == 1, ]
# extract value from predictors rasters for the present
ext <- extent(study)
```

add seed number for same result for every run time

```
set.seed(110)
# generate random points, we will restrict the background points within 12.5% of extent
backg <- randomPoints(predictors, n=5000, ext=ext, extf = 1.25)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
changed n
## to ncell of the mask (extent)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
generated
## random points = 0.442401960784314 times requested number

# change column name
colnames(backg) = c('lon', 'lat')
group <- kfold(backg, 5)
# absence train
backg_train <- backg[group != 1, ]
# absence test
backg_test <- backg[group == 1, ]
```

View some rows of background train

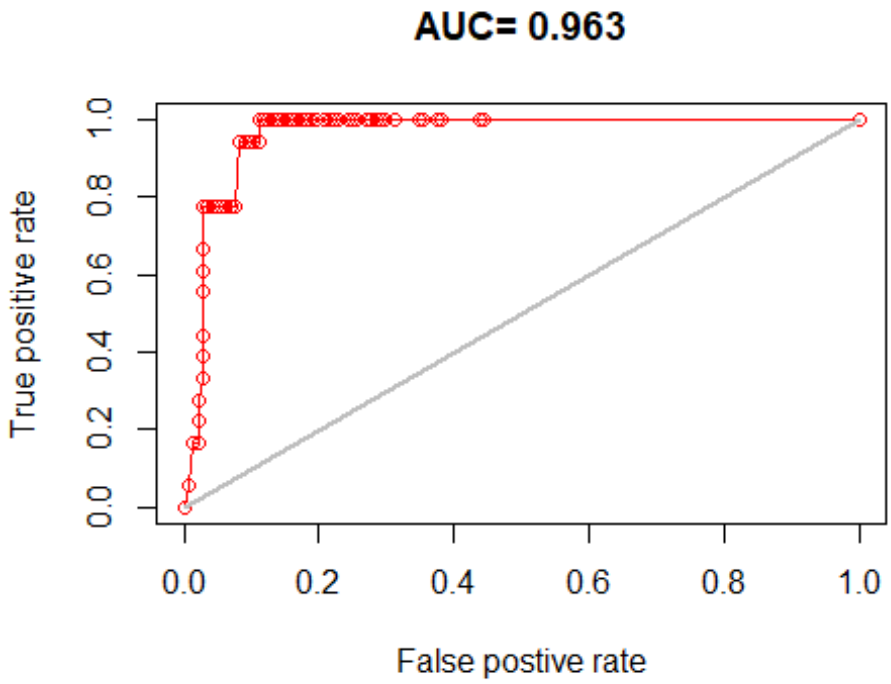
```
head(backg_train,2)

##            lon      lat
## [1,] -127.2042 55.32083
## [2,] -127.2042 55.40417

# train data
train <- rbind(pres_train, backg_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(backg_train)))
# Data extraction from raster scenario : SSP126_2021_40
envtrain <- raster::extract(predictors, train)
# Transforming Dataframe
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
# Changing column names
colnames(envtrain)[2:20]<- climate_variables
# view 2 rows
head(envtrain,2)

##   pa   AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1  1 7.872044 16.45744 -3.228431 863.6147 112.7023 37.32885  30.63447
```

```
306.6312
## 2  1 7.872044 16.45744 -3.228431 863.6147 112.7023 37.32885  30.63447
306.6312
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 135.7376 221.7675 185.9451 9.435161   29.26704  837.9612 25.54142 -
8.50254
## 2 135.7376 221.7675 185.9451 9.435161   29.26704  837.9612 25.54142 -
8.50254
##   TARange  MTempWt   MTempD
## 1 31.9893 7.717571 2.159852
## 2 31.9893 7.717571 2.159852

# testing data
testpres <- data.frame(raster::extract(predictors, pres_test) )
# Changing column names
colnames(testpres)<- climate_variables
#view top 2 rows
head(testpres,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 7.872044 16.45744 -3.228431 863.6147 112.7023 37.32885  30.63447
306.6312
## 2 8.083972 16.45744 -1.931105 854.1235 112.1735 37.25216  30.81558
306.7300
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 135.7376 221.7675 185.9451 9.435161   29.26704  837.9612 25.54142 -
8.50254
## 2 135.1289 219.2339 183.3305 9.433661   29.27120  837.9612 25.54142 -
8.50254
##   TARange  MTempWt   MTempD
## 1 31.9893 7.717571 2.159852
## 2 31.9893 7.835923 2.159852
```

Testing set for Absence data [randomly generated]

```
testbackg <- data.frame( raster::extract(predictors, backg_test) )
# Changing column names
colnames(testbackg)<- climate_variables
# View top 2 rows
head(testbackg,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 5.328898 16.71573 -4.525758 911.7612 119.3995 39.62967  29.91003
306.8288
## 2 6.741756 16.45744 -3.228431 883.4601 115.5222 38.32587  30.36281
307.2570
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
```

```
MnTempC
## 1 144.0564 245.2032 222.5491 9.406689   29.27120  829.8932 22.77699 -
8.783984
## 2 139.3898 231.9019 202.9398 9.420176   29.25873  829.8932 23.69847 -
7.939650
##    TARange  MTempWt    MTempD
## 1 32.17232 5.705584 0.8055785
## 2 32.06251 6.770753 1.4827152

# Random forest
envtrain<- na.omit(envtrain)
rf1 <-
randomForest(pa~MTempWt+PreCQ+TARange+MTempD+MTempC+MxTempW+AMTemp+PreDQ,
data=envtrain)

## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?

# model evaluation
erf <- evaluate(testpres, testbackg, rf1)
erf

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.9631559
## cor            : 0.6889676
## max TPR+TNR at : 0.06749697

# ROC plot
plot(erf,'ROC')
```



Now Prediction for the random forest model

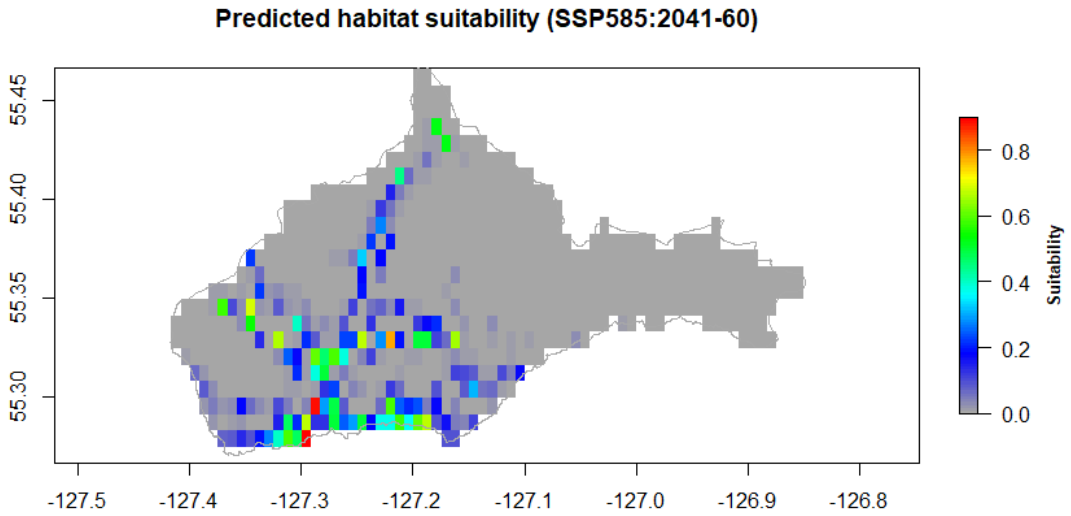
```
pr6 <- predict(predictors, rf1, ext=ext)
```

save raster file random forest predicted in the local (ssp585-2041-2060)

```
writeRaster(pr6, 'E:/online work/SDM/newsdm15_2/predicted/SSP585_2041_60_random forest.tif', overwrite=TRUE)
```

Map of the predicted habitat suitability from random forest

```
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pr6, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted habitat suitability (SSP585:2041-60)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



#7 ### Data Processing :ssp585-2061-2080

```
# read layers
biolayer<- Sys.glob('E:/online
work/SDM/newsdm15_2/ssp585/downscale/2061_80/*.tif')
# make brick [multilayers]
biolayer<-brick(lapply(paste(biolayer,sep=''),raster))
```

Extraction their values

```
# reassign as predictors
predictors<- biolayer
names(predictors)<- climate_variables # rename  for simplification
# extract value from predictors rasters for the present
presvals <- raster::extract(predictors, coord@coords)

# assign correct short names
colnames(presvals)<- climate_variables
```

Generating pseudo absence values random point in the

```
set.seed(10)
# generate random points
backgr <- randomPoints(predictors, 5000)

## Warning in randomPoints(predictors, 5000): changed n to ncell of the mask
## (extent)

## Warning in randomPoints(predictors, 5000): generated random points =
## 0.442401960784314 times requested number

# extract data value from predictors for the absent points
absvals <- raster::extract(predictors, backgr)
```

```r
# Rename with identifiable Acronym
colnames(absvals)<- climate_variables
# generate the training set for Observed Devilsclubs [present data ]
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
# convert into a data frame
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
# view some extract Bioclimate variable for Devilsclubs in the
head(sdmdata )
```

```
##   pb   AMTemp   MTempW     MTempC   Apreci    PreWm   PreDM Pseasonal
PreWtQ
## 1  1 10.23487 18.72111 -0.7901961 936.0084 131.7054 40.1906  33.30191
345.3145
## 2  1 10.23487 18.72111 -0.7901961 936.0084 131.7054 40.1906  33.30191
345.3145
## 3  1 10.23487 18.72111 -0.7901961 936.0084 131.7054 40.1906  33.30191
345.3145
## 4  1 10.23487 18.72111 -0.7901961 936.0084 131.7054 40.1906  33.30191
345.3145
## 5  1 10.37571 18.85001 -0.7901961 923.2843 131.2836 40.1906  33.46173
345.4300
## 6  1 10.37571 18.85001 -0.7901961 923.2843 131.2836 40.1906  33.46173
345.4300
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 144.3736 227.0371 200.9654 9.091190   28.43177  831.8701 28.14928 -
5.633973
## 2 144.3736 227.0371 200.9654 9.091190   28.43177  831.8701 28.14928 -
5.633973
## 3 144.3736 227.0371 200.9654 9.091190   28.43177  831.8701 28.14928 -
5.633973
## 4 144.3736 227.0371 200.9654 9.091190   28.43177  831.8701 28.14928 -
5.633973
## 5 143.9264 225.0053 199.5605 9.092649   28.43177  831.8701 28.14928 -
5.492754
## 6 143.9264 225.0053 199.5605 9.092649   28.43177  831.8701 28.14928 -
5.492754
##    TARange  MTempWt   MTempD
## 1 31.73709 10.10138 4.368202
## 2 31.73709 10.10138 4.368202
## 3 31.73709 10.10138 4.368202
## 4 31.73709 10.10138 4.368202
## 5 31.73709 10.21901 4.368202
## 6 31.73709 10.21901 4.368202
```

k-fold cross validation

```r
set.seed(05)
# make 5 k-fold cross validation
group <- kfold(coord@coords, 5)
# split into training and test set
```

```
pres_train <- coord@coords[group != 1, ]
#presence test
pres_test <- coord@coords[group == 1, ]
# extract value from predictors rasters for the present
ext <- extent(study)
```

add seed number for same result for every run time

```
set.seed(14)
# generate random points, we will restrict the background points within 12.5% of extent
backg <- randomPoints(predictors, n=5000, ext=ext, extf = 1.25)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
changed n
## to ncell of the mask (extent)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
generated
## random points = 0.442401960784314 times requested number

# change column name
colnames(backg) = c('lon', 'lat')
group <- kfold(backg, 5)
# absence train
backg_train <- backg[group != 1, ]
# absence test
backg_test <- backg[group == 1, ]
```

View some rows of background train

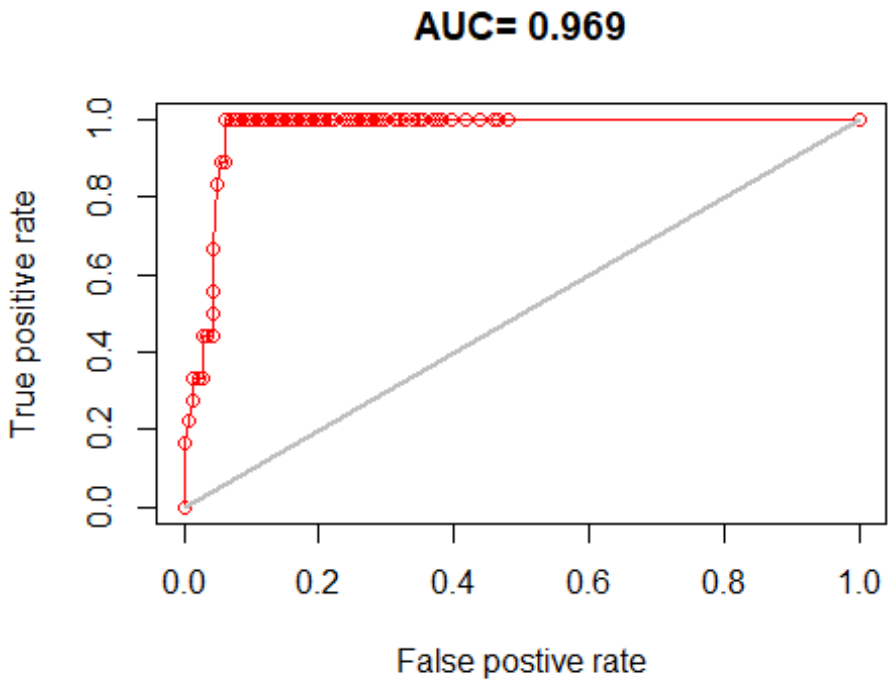
```
head(backg_train,2)

##            lon      lat
## [1,] -127.2958 55.32917
## [2,] -127.1208 55.36250

# train data
train <- rbind(pres_train, backg_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(backg_train)))
# Data extraction from raster scenario : SSP126_2021_40
envtrain <- raster::extract(predictors, train)
# Transforming Dataframe
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
# Changing column names
colnames(envtrain)[2:20]<- climate_variables
# view 2 rows
head(envtrain,2)

##   pa   AMTemp   MTempW     MTempC   Apreci    PreWm   PreDM Pseasonal
PreWtQ
## 1  1 10.23487 18.72111 -0.7901961 936.0084 131.7054 40.1906  33.30191
```

```
345.3145
## 2  1 10.23487 18.72111 -0.7901961 936.0084 131.7054 40.1906  33.30191
345.3145
##      PreDQ    PreWQ    PreCQ MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 144.3736 227.0371 200.9654 9.09119   28.43177  831.8701 28.14928 -
5.633973
## 2 144.3736 227.0371 200.9654 9.09119   28.43177  831.8701 28.14928 -
5.633973
##    TARange  MTempWt   MTempD
## 1 31.73709 10.10138 4.368202
## 2 31.73709 10.10138 4.368202

# testing data
testpres <- data.frame(raster::extract(predictors, pres_test) )
# Changing column names
colnames(testpres)<- climate_variables
#view top 2 rows
head(testpres,2)

##     AMTemp   MTempW    MTempC   Apreci    PreWm    PreDM Pseasonal PreWtQ
## 1 10.44613 18.72111 0.4983957 925.7169 131.0728 40.09661  33.46173 345.43
## 2 10.44613 18.72111 0.4983957 925.7169 131.0728 40.09661  33.46173 345.43
##      PreDQ   PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 143.7028 224.328 198.1557 9.089732    28.4387  831.8701 28.14928 -
5.633973
## 2 143.7028 224.328 198.1557 9.089732    28.4387  831.8701 28.14928 -
5.633973
##    TARange  MTempWt   MTempD
## 1 31.73709 10.21901 4.368202
## 2 31.73709 10.21901 4.368202
```

Testing set for Absence data [randomly generated]

```
testbackg <- data.frame( raster::extract(predictors, backg_test) )
# Changing column names
colnames(testbackg)<- climate_variables
# View top 2 rows
head(testbackg,2)

##     AMTemp   MTempW     MTempC   Apreci    PreWm    PreDM Pseasonal
PreWtQ
## 1 8.896887 18.97891 -0.7901961 944.4287 135.5013 41.97624  33.38182
338.3859
## 2 8.615208 18.97891 -0.7901961 951.5392 136.5558 42.25818  33.22200
338.2319
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 149.0689 239.2284 223.4434 9.078065   28.41791  823.9312 26.30276 -
5.069097
```

```
## 2 150.4104 242.6149 229.0629 9.073690   28.41791  823.9312 25.37951 -
5.069097
##    TARange  MTempWt   MTempD
## 1 31.82105 9.160386 3.691663
## 2 31.85464 8.925137 3.522528

# Random forest
envtrain<- na.omit(envtrain)
rf1 <-
randomForest(pa~MTempWt+PreCQ+TARange+MTempD+MTempC+MxTempW+AMTemp+PreDQ,
data=envtrain)

## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?

# model evaluation
erf <- evaluate(testpres, testbackg, rf1)
erf

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.96875
## cor            : 0.72225
## max TPR+TNR at : 0.4101547

# ROC plot
plot(erf,'ROC')
```



Now Prediction for the random forest model

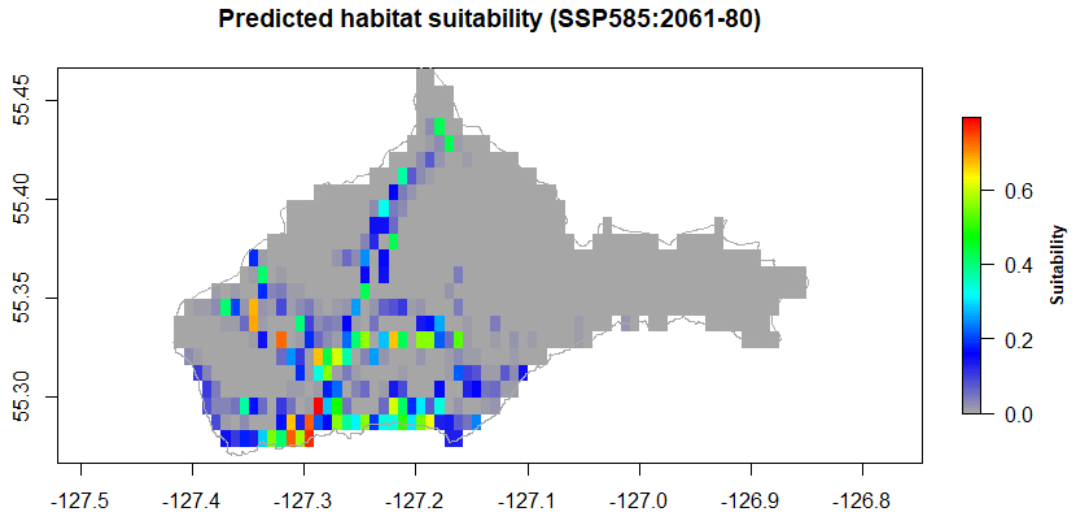
```
pr7 <- predict(predictors, rf1, ext=ext)
```

save raster file random forest predicted in the local (ssp585-2061-2080)

```
writeRaster(pr7, 'E:/online work/SDM/newsdm15_2/predicted/SSP585_2061_80_random forest.tif', overwrite=TRUE)
```

Map of the predicted habitat suitability from random forest

```
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pr7, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted habitat suitability (SSP585:2061-80)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



#8 ### Data Processing for BC:ssp585-2081-2100

```
# read layers
biolayer<- Sys.glob('E:/online
work/SDM/newsdm15_2/ssp585/downscale/2081_2100/*.tif')
# make brick [multilayers]
biolayer<-brick(lapply(paste(biolayer,sep=''),raster))
```

Extraction their values

```
# reassign as predictors
predictors<- biolayer
names(predictors)<- climate_variables # rename  for simplification
# extract value from predictors rasters for the present
presvals <- raster::extract(predictors, coord@coords)

# assign correct short names
colnames(presvals)<- climate_variables
```

Generating pseudo absence values random point in the

```
set.seed(0)
# generate random points
backgr <- randomPoints(predictors, 5000)

## Warning in randomPoints(predictors, 5000): changed n to ncell of the mask
## (extent)

## Warning in randomPoints(predictors, 5000): generated random points =
## 0.442401960784314 times requested number

# extract data value from predictors for the absent points
absvals <- raster::extract(predictors, backgr)
```

```
# Rename with identifiable Acronym
colnames(absvals)<- climate_variables
# generate the training set for Observed Devilsclubs [present data ]
pb <- c(rep(1, nrow(presvals)), rep(0, nrow(absvals)))
# convert into a data frame
sdmdata <- data.frame(cbind(pb, rbind(presvals, absvals)))
# view some extract Bioclimate variable for Devilsclubs in the
head(sdmdata )
```

```
##   pb   AMTemp   MTempW   MTempC   Apreci    PreWm   PreDM Pseasonal
PreWtQ
## 1  1 12.53817 21.18543 1.229412 987.3831 147.0497 42.8716  36.60263
382.4902
## 2  1 12.53817 21.18543 1.229412 987.3831 147.0497 42.8716  36.60263
382.4902
## 3  1 12.53817 21.18543 1.229412 987.3831 147.0497 42.8716  36.60263
382.4902
## 4  1 12.53817 21.18543 1.229412 987.3831 147.0497 42.8716  36.60263
382.4902
## 5  1 12.67833 21.31483 1.229412 973.9742 146.5897 42.8716  36.73838
382.6163
## 6  1 12.67833 21.31483 1.229412 973.9742 146.5897 42.8716  36.73838
382.6163
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 150.0817 221.9580 213.3616 8.965603   27.38491  849.8155 31.01765 -
3.532619
## 2 150.0817 221.9580 213.3616 8.965603   27.38491  849.8155 31.01765 -
3.532619
## 3 150.0817 221.9580 213.3616 8.965603   27.38491  849.8155 31.01765 -
3.532619
## 4 150.0817 221.9580 213.3616 8.965603   27.38491  849.8155 31.01765 -
3.532619
## 5 149.6337 219.9709 211.8671 8.966710   27.38491  849.8155 31.01765 -
3.390451
## 6 149.6337 219.9709 211.8671 8.966710   27.38491  849.8155 31.01765 -
3.390451
##    TARange  MTempWt   MTempD
## 1 32.50469 12.54937 6.347861
## 2 32.50469 12.54937 6.347861
## 3 32.50469 12.54937 6.347861
## 4 32.50469 12.54937 6.347861
## 5 32.50469 12.66657 6.347861
## 6 32.50469 12.66657 6.347861
```

k-fold cross validation

```
set.seed(0)
# make 5 k-fold cross validation
group <- kfold(coord@coords, 5)
# split into training and test set
```

```
pres_train <- coord@coords[group != 1, ]
#presence test
pres_test <- coord@coords[group == 1, ]
# extract value from predictors rasters for the present
ext <- extent(study)
```

add seed number for same result for every run time

```
set.seed(10)
# generate random points, we will restrict the background points within 12.5%
of extent
backg <- randomPoints(predictors, n=5000, ext=ext, extf = 1.25)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
changed n
## to ncell of the mask (extent)

## Warning in randomPoints(predictors, n = 5000, ext = ext, extf = 1.25):
generated
## random points = 0.442401960784314 times requested number

# change column name
colnames(backg) = c('lon', 'lat')
group <- kfold(backg, 5)
# absence train
backg_train <- backg[group != 1, ]
# absence test
backg_test <- backg[group == 1, ]
```

View some rows of background train

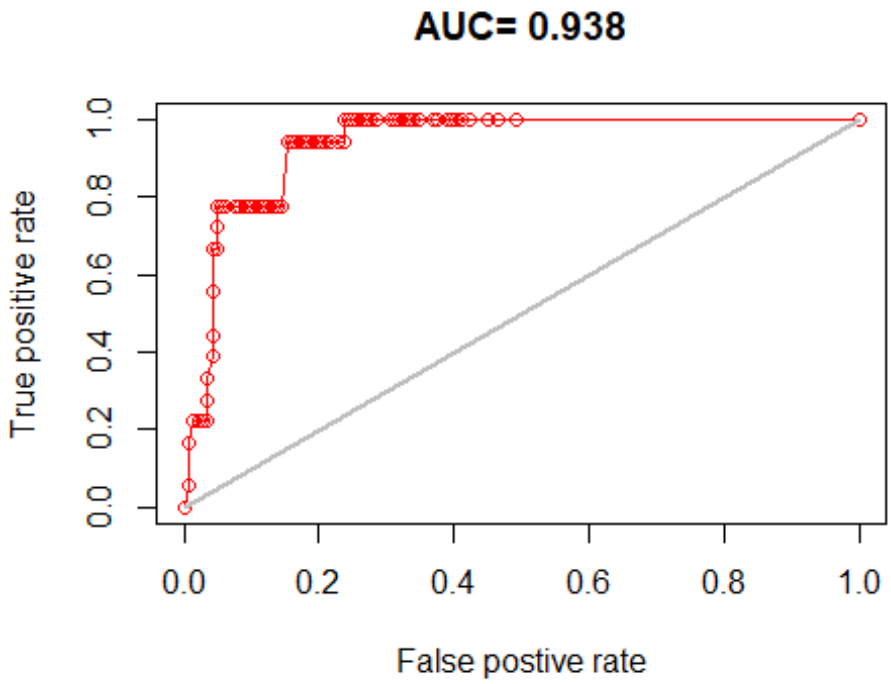
```
head(backg_train,2)

##            lon      lat
## [1,] -127.1042 55.32917
## [2,] -127.0542 55.34583

# train data
train <- rbind(pres_train, backg_train)
pb_train <- c(rep(1, nrow(pres_train)), rep(0, nrow(backg_train)))
# Data extraction from raster scenario : SSP126_2021_40
envtrain <- raster::extract(predictors, train)
# Transforming Dataframe
envtrain <- data.frame( cbind(pa=pb_train, envtrain) )
# Changing column names
colnames(envtrain)[2:20]<- climate_variables
# view 2 rows
head(envtrain,2)

##   pa   AMTemp   MTempW   MTempC   Apreci    PreWm   PreDM Pseasonal
PreWtQ
## 1  1 12.53817 21.18543 1.229412 987.3831 147.0497 42.8716  36.60263
```

```
382.4902
## 2  1 12.53817 21.18543 1.229412 987.3831 147.0497 42.8716  36.60263
382.4902
##      PreDQ   PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 150.0817 221.958 213.3616 8.965603   27.38491  849.8155 31.01765 -
3.532619
## 2 150.0817 221.958 213.3616 8.965603   27.38491  849.8155 31.01765 -
3.532619
##    TARange  MTempWt   MTempD
## 1 32.50469 12.54937 6.347861
## 2 32.50469 12.54937 6.347861

# testing data
testpres <- data.frame(raster::extract(predictors, pres_test) )
# Changing column names
colnames(testpres)<- climate_variables
#view top 2 rows
head(testpres,2)

##     AMTemp   MTempW   MTempC   Apreci    PreWm    PreDM Pseasonal   PreWtQ
## 1 12.53817 21.18543 1.229412 987.3831 147.0497 42.87160  36.60263 382.4902
## 2 12.74842 21.18543 2.513369 976.5377 146.3597 42.74622  36.73838 382.6163
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 150.0817 221.9580 213.3616 8.965603   27.38491  849.8155 31.01765 -
3.532619
## 2 149.4097 219.3085 210.3726 8.964497   27.39213  849.8155 31.01765 -
3.532619
##    TARange  MTempWt   MTempD
## 1 32.50469 12.54937 6.347861
## 2 32.50469 12.66657 6.347861
```

Testing set for Absence data [randomly generated]

```
testbackg <- data.frame( raster::extract(predictors, backg_test) )
# Changing column names
colnames(testbackg)<- climate_variables
# View top 2 rows
head(testbackg,2)

##     AMTemp   MTempW      MTempC   Apreci  PreWm    PreDM Pseasonal
PreWtQ
## 1 10.43570 21.57361 -0.05454544 1011.243 153.95 46.38240  36.46688
383.0367
## 2 10.57586 21.57361  1.22941172 1002.764 153.26 46.25701  36.60263
383.1628
##      PreDQ    PreWQ    PreCQ  MDRange Isothermal Tseasonal  MxTempW
MnTempC
## 1 147.8416 241.8292 252.2186 8.951218   27.37768  842.1006 28.26948 -
3.248284
```

```
## 2 146.9455 239.8421 249.2296 8.949004   27.37768  842.1006 28.26948 -
3.248284
##    TARange  MTempWt   MTempD
## 1 32.65714 11.02573 5.166661
## 2 32.65714 11.02573 5.166661

# Random forest
envtrain<- na.omit(envtrain)
rf1 <-
randomForest(pa~MTempWt+PreCQ+TARange+MTempD+MTempC+MxTempW+AMTemp+PreDQ,
data=envtrain)

## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?

# model evaluation
erf <- evaluate(testpres, testbackg, rf1)
erf

## class          : ModelEvaluation
## n presences    : 18
## n absences     : 144
## AUC            : 0.9375
## cor            : 0.6090044
## max TPR+TNR at : 0.1705785

# ROC plot
plot(erf,'ROC')
```



Now Prediction for the random forest model

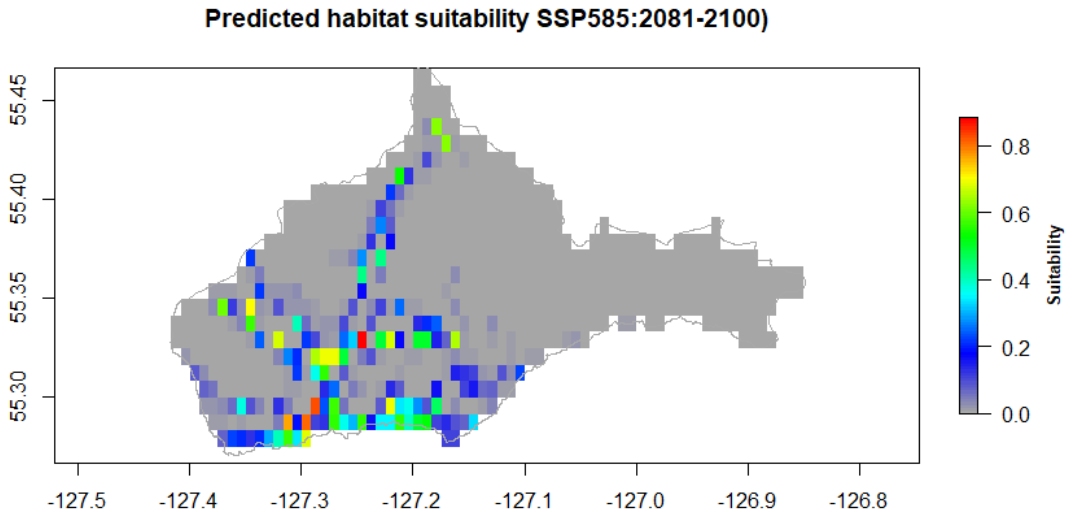
```
pr8 <- predict(predictors, rf1, ext=ext)
```

save raster file random forest predicted in the local (2081_2100/ssp585)

```
writeRaster(pr8, 'E:/online
work/SDM/newsdm15_2/predicted/SSP585_2081_2100_random forest.tif',
overwrite=TRUE)
```

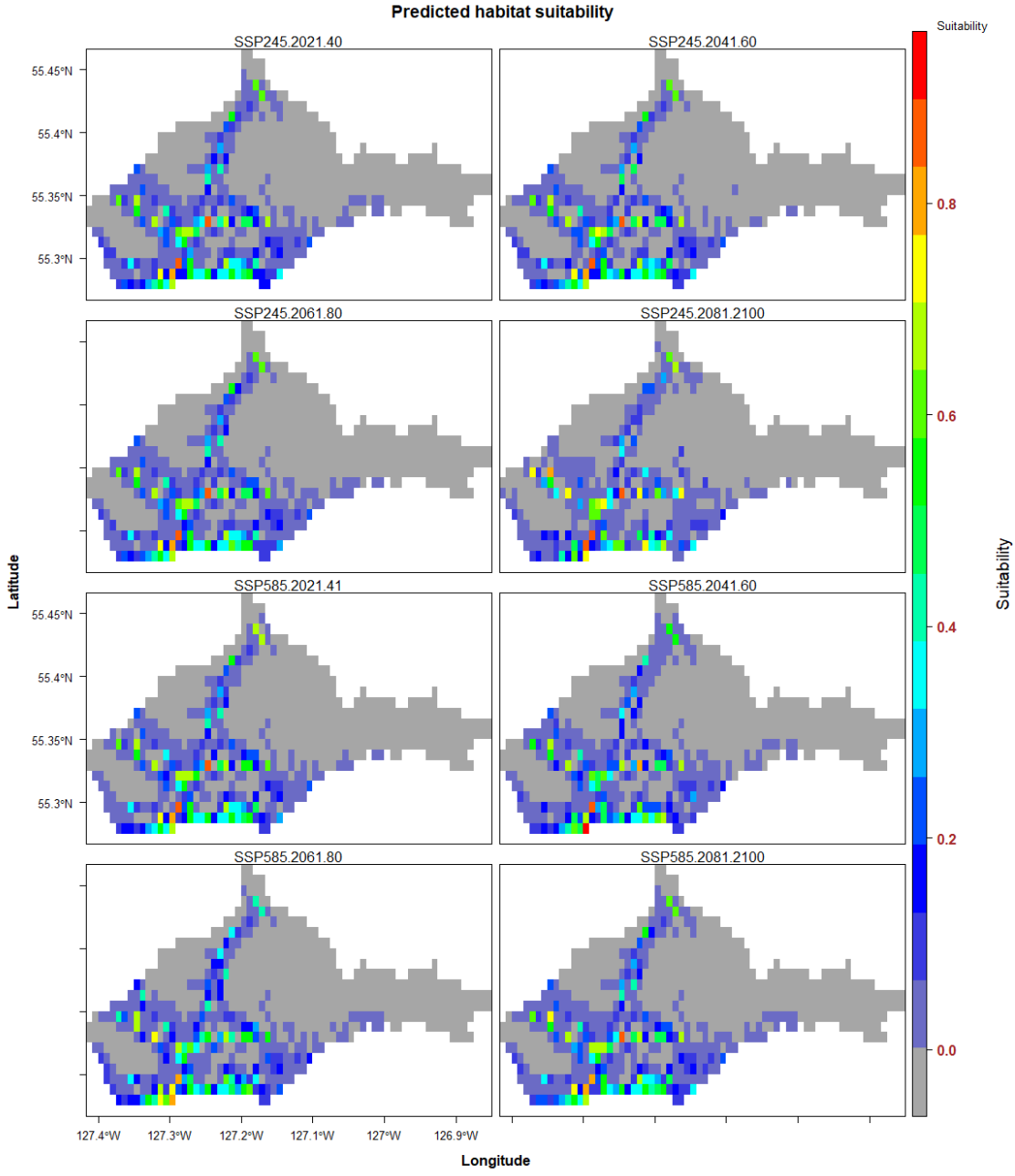
Map of the predicted habitat suitability from random forest

```
pal <- colorRampPalette(c("grey65","blue","cyan","green","yellow","red"))
plot(pr8, col=pal(100),legend.width=1, legend.shrink=0.75,main='Predicted
habitat suitability SSP585:2081-2100)',
     legend.args=list(text='Suitability', side=4, font=2, line=2.5, cex=0.8))
plot(study, add=TRUE, border='dark grey')
```



#See 8 scenarios together

```
library(rasterVis)

## Loading required package: lattice

# Making eight scenario together
rbri<- brick(pr1,pr2,pr3,pr4,pr5,pr6,pr7,pr8)
names(rbri)<- c('SSP245:2021-40','SSP245:2041-60','SSP245:2061-
80','SSP245:2081-2100','SSP585:2021-41','SSP585:2041-60','SSP585:2061-
80','SSP585:2081-2100')

levelplot(rbri,layout=c(2, 4),main='Predicted habitat suitability',
          col.regions=pal(100),ylab=list('Latitude',fontface='bold'),
          xlab=list('Longitude',fontface='bold'),
          colorkey=list(labels=list(cex=1, font=2, col="brown"),
                        height=1, width=1.4,
                        title='Suitability',
              main=list(label='b',side=1,line=0.5, cex=10)))
```


Predicted habitat suitability
SSP245.2021.40
SSP245.2041.60
SSP245.2061.80
SSP245.2081.2100
SSP585.2021.41
SSP585.2041.60
SSP585.2061.80
SSP585.2081.2100
55.45°N
55.4°N
55.35°N
55.3°N
127.4°W
127.3°W
127.2°W
127.1°W
127°W
126.9°W
Latitude
Longitude
Suitability
0.8
0.6
0.4
0.2
0.0
Suitability