Code ▾

Risk and confidence maps for Vaccinium corymbosum

Amy J.S Davis

2024-01-13

    - Reomve highly correlated climate predictors from training data
    - Add habitat and anthropogenic predictors
    - Identify highly correlated predictors from the habitat/anthropogenic/climate stack (full stack)
    - Remove highly correlated predictors from full stack
    - Identify and remove near zero variance predictors
    - Build models with climate and habitat data
    - Display model evaluation statistics
    - Create ensemble model
    - Use EU level ensemble models (each using a separate pseudoabsence draw) to predict at European level
    - Use EU level ensemble models to predict for Belgium only
    - Evaluate the performance of each the EU level ensemble models based on results from CV
- 2. Using thresholds identified for each model in the previous step, assess performance of each model
    - plot the best EU level ensemble model
    - Subset Belgium occurrences
    - plot the best EU level ensemble model showing only Belgium
    - Clip habitat raster stack to Belgium
    - Create individual RCP (2.6, 4.5, 8.5) climate raster stacks for Belgium
    - Combine habitat stacks with climate stacks for each RCP scenario
    - Create and export RCP risk maps for each RCP scenario
    - Create and export RCP risk maps for each RCP scenario
    - Create and export “difference maps”: the difference between predicted risk by each RCP scenario and historical climate
    - Check spatial autocorrelation of residuals to assess whether occurrence data should be thinned
    - Check Morans I.
    - Code for Mondrian conformal prediction functions
    - Quantify confidence of predicted values using class conformal prediction
    - Create confidence maps
    - Mask areas of below a set confidence level
    - confidence map of best model at EU level
    - Get variable importance of best european model
    - Generate and export response curves in order of variable importance
    - Plot response curves
    - Evaluate the performance of each the EU level ensemble models using independent data set from the future

File failed to load: /extensions/MathZoom.js

```
Warning: package ‘maps’ was built under R version 4.2.3Warning: package ‘spocc’ was built und
er R version 4.2.3Warning: package ‘SDMPlay’ was built under R version 4.2.3Warning: package
‘ggplot2’ was built under R version 4.2.3Warning: package ‘tibble’ was built under R version
4.2.3Warning: package ‘readr’ was built under R version 4.2.3Warning: package ‘dplyr’ was bui
lt under R version 4.2.3Warning: package ‘caret’ was built under R version 4.2.3Warning: pack
age ‘caretEnsemble’ was built under R version 4.2.3Warning: package ‘gbm’ was built under R v
ersion 4.2.3Warning: package ‘rgbif’ was built under R version 4.2.3Warning: package ‘maptool
s’ was built under R version 4.2.3Warning: package ‘dismo’ was built under R version 4.2.3War
ning: package ‘sf’ was built under R version 4.2.3Warning: package ‘geoR’ was built under R v
ersion 4.2.3Warning: package ‘pdp’ was built under R version 4.2.3Warning: package ‘here’ was
built under R version 4.2.3Warning: package ‘CoordinateCleaner’ was built under R version 4.
2.3Warning: package ‘knitr’ was built under R version 4.2.3Warning: package ‘kableExtra’ was
built under R version 4.2.3Warning: package ‘rmarkdown’ was built under R version 4.2.3
```

1, Download global occurrence data from GBIF

Retrieve the `taxonKey` s we want to use to download occurrences:

Hide

```
# TO DO: specify scientific name of species to be modelled
  species<- "Vaccinium corymbosum"

# retrieve taxon key from GBIF (which is returned here as the "usageKey")
taxon.data<-name_backbone(name=species)

taxonName<-species
taxon_key<-taxon.data$usageKey

gbif_filename<- paste(taxonName,".csv",sep="")
taxon.data
```

	usage…	scientificName	canonicalName	rank	status	confidenc
	<int>	<chr>	<chr>	<chr>	<chr>	<int
1	2882849	Vaccinium corymbosum L.	Vaccinium corymbosum	SPECIES	ACCEPTED	9

1 row | 1-8 of 23 columns

Hide

```
NA
```

Basis of record

Hide

File failed to load: /extensions/MathZoom.js

```
#All types of occurrences are downloaded, except `FOSSIL SPECIMEN` and `LIVING SPECIMEN`, which can have misleading location information (e.g. location of captive animal).

basis_of_record <- c(
  "OBSERVATION",
  "HUMAN_OBSERVATION",
  "MATERIAL_SAMPLE",
  "PRESERVED_SPECIMEN",
  "UNKNOWN",
  "MACHINE_OBSERVATION",
  "OCCURRENCE"
)
```

Specify time period to download occurrence data

Hide

```
year_begin <- 1971
year_end <-2010
```

Download only georeferenced points

Hide

```
hasCoordinate <- TRUE
```

Trigger download

Note: GBIF credentials are required in the next step.

Trigger download:

Hide

```

 gbif_download_key <- occ_download(
   pred_in("taxonKey", taxon_key),
   pred_in("basisOfRecord", basis_of_record),
   pred_gte("year", year_begin),
   pred_lte("year", year_end),
  pred("hasCoordinate", hasCoordinate),
   user = rstudioapi::askForPassword("GBIF username"),
   pwd = rstudioapi::askForPassword("GBIF password"),
   email = rstudioapi::askForPassword("Email address for notification")
 )
```

Check status of download

File failed to load: /extensions/MathZoom.js

Hide

```
metadata <- occ_download_meta(key = gbif_download_key)
metadata$key
metadata$status
```

Hide

```
occ_download_get(paste(metadata$key), path = here("./data/raw"))
```

```
Download file size: 1.08 MB
file exists & overwrite=FALSE, not overwriting...
```

```
<<gbif downloaded get>>
  Path: C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/raw/0047696-231120084113126.zip
  File size: 1.08 MB
```

Hide

```
raw.path<- here("data/raw//")
unzip(paste0(raw.path,metadata$key,".zip"),exdir=paste0(raw.path,metadata$key))

global<-as.data.frame(data.table::fread(paste0(raw.path,metadata$key,"/occurrence.txt"),header=TRUE))
```

2. Create a global SDM

2. Specify paths for output (defaults to file structure in ReadMe)

####3. Filter global occurrence data

Hide

File failed to load: /extensions/MathZoom.js

```
#remove unverified records
identificationVerificationStatus_to_discard <- c("unverified", "unvalidated","not able to val
idate","control could not be conclusive due to insufficient knowledge")

#enter value for max coordinate uncertainty in meters.

global.occ<-global %>%
  filter(speciesKey==taxonkey) %>%   #using taxonKey filters out accepted synonyms
  filter(is.na(coordinateUncertaintyInMeters)| coordinateUncertaintyInMeters<= 1000) %>%
  filter(!str_to_lower(identificationVerificationStatus) %in% identificationVerificationStatu
s_to_discard)

 global.occ$lon_dplaces<-sapply(global.occ$decimalLongitude, function(x) decimalplaces(x))
 global.occ$lat_dplaces<-sapply(global.occ$decimalLatitude, function(x) decimalplaces(x))
 global.occ[global.occ$lon_dplaces < 4& global.occ$lat_dplaces < 4 , ]<-NA
 global.occ<-global.occ[ which(!is.na(global.occ$lon_dplaces)),]
 global.occ<-within(global.occ,rm("lon_dplaces","lat_dplaces"))
global.occ<-global.occ[which( global.occ$year > 1970 & global.occ$year < 2011),]
```

Convert global occurrences to spatial points needed for modelling

Flag and remove centroids and invalid georeferenced points

```
Testing coordinate validity
Flagged 0 records.
Testing zero coordinates
Warning: GEOS support is provided by the sf and terra packages among othersFlagged 0 records.
Testing country capitals
Flagged 3 records.
Testing country centroids
Flagged 0 records.
Testing sea coordinates
trying URL 'https://naturalearth.s3.amazonaws.com/50m_physical/ne_50m_land.zip'
Content type 'application/zip' length 457183 bytes (446 KB)
downloaded 446 KB

Flagged 61 records.
Testing GBIF headquarters, flagging records around Copenhagen
Flagged 0 records.
Testing biodiversity institutions
Flagged 2 records.
Flagged 66 of 1678 records, EQ = 0.04.
Testing coordinate validity
Flagged 0 records.
Testing zero coordinates
Warning: GEOS support is provided by the sf and terra packages among othersFlagged 0 records.
Testing country capitals
Flagged 3 records.
Testing country centroids
Flagged 0 records.
Testing sea coordinates
trying URL 'https://naturalearth.s3.amazonaws.com/50m_physical/ne_50m_land.zip'
Content type 'application/zip' length 457183 bytes (446 KB)
downloaded 446 KB

Flagged 61 records.
Testing GBIF headquarters, flagging records around Copenhagen
Flagged 0 records.
Testing biodiversity institutions
Flagged 2 records.
Flagged 66 of 1678 records, EQ = 0.04.
```

Create global rasterstack using CHELSA data for model building

Hide

```
globalclimrasters <- list.files((here("./data/external/climate/trias_CHELSA")),pattern='tif',
full.names = T) #import CHELSA data
globalclimpreds <- stack(globalclimrasters)
```

Use SDMtab command from the SDMPlay package to remove duplicates per grid cell

Hide

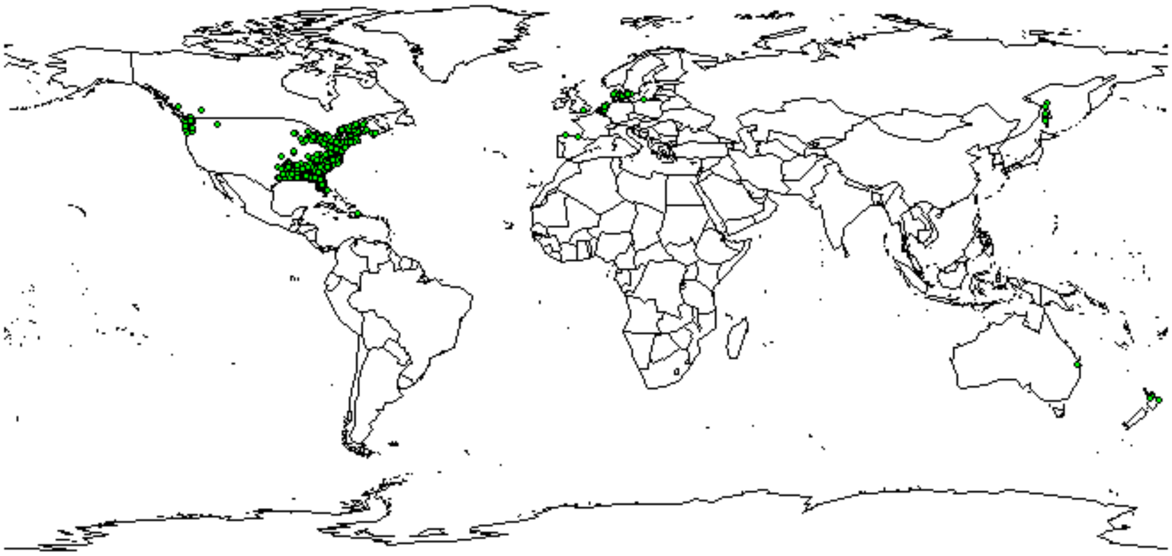
File failed to load: /extensions/MathZoom.js

```
global.SDMtable<- SDMPlay:::SDMtab(global.occ.LL.cleaned, globalclimpreds, unique.data = TRUE,background.nb= 0) #
numb.global.pseudoabs <-length(global.SDMtable$id) #sets the number of pseudoabsences equal to number of unique presences


global.occ.sp<-global.SDMtable[c("longitude", "latitude")]
coordinates(global.occ.sp)<- c("longitude", "latitude")
global.occ.sp$species<- rep(1,length(global.occ.sp$latitude)) #adds columns indicating species presence needed for modeling
```

plot distribution of cleaned global occurrences

Hide

```
maps::map('world', fill = FALSE, wrap=c(-180,180))
plot(global.occ.sp,pch=21,bg="green",cex=.5,add=TRUE)
```



Select wwf ecoregions that contain global occurrence points

Specify and import bias grids for relevant taxonomic group (e.g vascular plants)

Hide

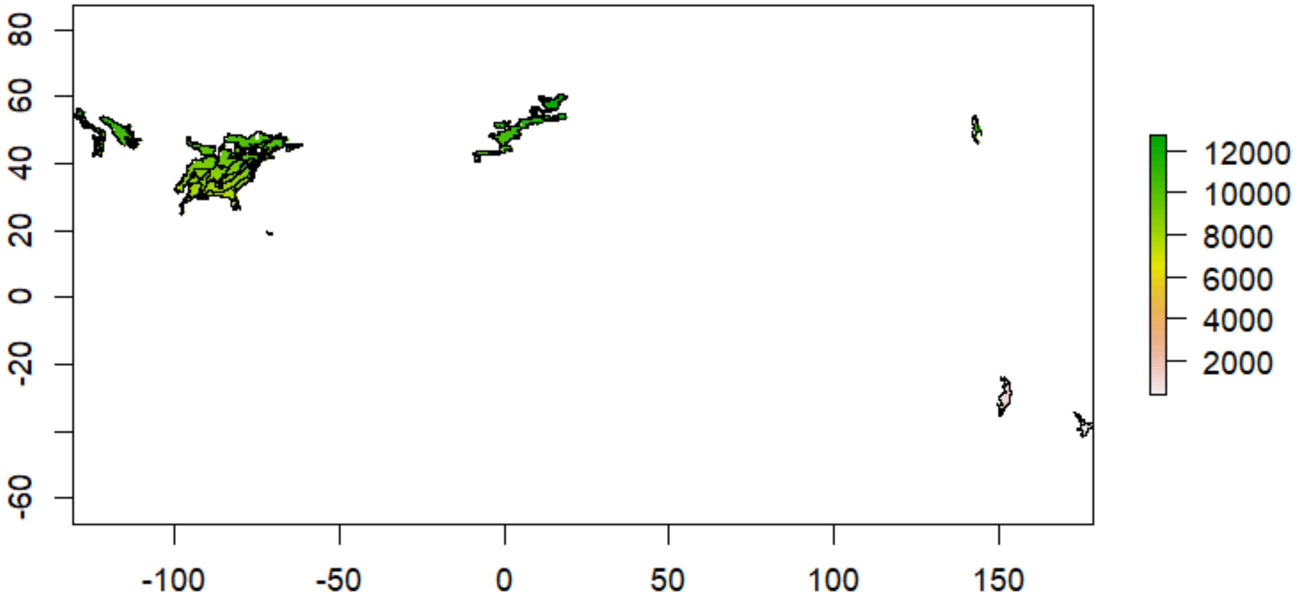
File failed to load: /extensions/MathZoom.js

```
biasgrid<-raster(here("./data/external/bias_grids/final/trias/plants_1deg_min5.tif"))### spec
ify appropriate bias grid here
```

Subset bias grid by ecoregions containing occurrence points

Hide

```
ext_wwf_ecoSub<-extent(wwf_ecoSub1)
biasgrid_crop<-crop(biasgrid,ext_wwf_ecoSub)
biasgrid_sub<-mask(biasgrid_crop,wwf_ecoSub1)
 plot(biasgrid_sub)
 plot(wwf_ecoSub1,add=TRUE)
```



Hide

```
NA
```

Use randomPoints function from dismo package to locate pseduobasences within the bias grid subset

Hide

File failed to load: /extensions/MathZoom.js

```
# generates pseudo absences equal to (or close to) the number of presences.
set.seed(728)
global_points<-randomPoints(biasgrid_sub,numb.global.pseudoabs, global.occ.sp, ext=NULL, extf=1.1, excludep=TRUE, prob=FALSE, cellnumbers=FALSE, tryf=70, warn=2, lonlatCorrection=TRUE)
# will throw a warning if randomPoints generated is less than numb.pseudoabs. If this happens, increase the number of tryf or ignore bias grid and sample from ecoregion only.
```

OPTIONAL: Sample from ecoregion only

run if the bias grid subset of ecoregions results in too small of an area for sampling

Hide

```
#  wwf_grid<-raster(here("./data/external/GIS/wwf_ecoregions_v1.tif"))
#  ecoregions_raster<-mask(wwf_grid,wwf_ecoSub1)
#  set.seed(768)
#  global_points<-randomPoints(ecoregions_raster, numb.pseudoabs, global.occ.sp, ext=NULL, extf=1.1, excludep=TRUE, prob=FALSE, cellnumbers=FALSE, tryf=150, warn=2, lonlatCorrection=TRUE)
```

Extract generated pseudo absences and create presence-pseudobasence dataset

Hide

```
global_pseudoAbs<-as.data.frame(global_points)
coordinates(global_pseudoAbs)<-c("x","y")
crs(global_pseudoAbs)<-CRS("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0")
global_pseudoAbs$species<-rep(0,length(global_pseudoAbs$x))
global_presabs<- spRbind(global.occ.sp,global_pseudoAbs) # join pseudoabsences with presences (occurrences)
```

Extract climate data for global scale modelling

Hide

```

global.data <- sdmData(species~.,train=global_presabs, predictors=globalclimpreds)
```

```
Warning: package ‘mda’ was built under R version 4.2.3Warning: package ‘glmnet’ was built under R version 4.2.3Warning: package ‘earth’ was built under R version 4.2.3Warning: package ‘plotmo’ was built under R version 4.2.3Warning: package ‘TeachingDemos’ was built under R version 4.2.3Warning: package ‘randomForest’ was built under R version 4.2.3
```

Hide

```
global.data.df<-as.data.frame(global.data)
```

File failed to load: /extensions/MathZoom.js

Identify highly correlated predictors

highlyCorrelated
CHELSA_minTmpColdestMon
CHELSA_meantemp
CHELSA_temp_seasonality
CHELSA_precipWettestMon

Remove highly correlated predictors from dataframe

Correct global clim preds values from integer format

Use caretList from Caret package to run multiple machine learning models

Hide

```
GlobalModelResults<-resamples(global_train)
Global.Mod.Accuracy<-summary(GlobalModelResults)# displays accuracy of each model
kable(Global.Mod.Accuracy$statistics$Accuracy,digits=2) %>%
kable_styling(bootstrap_options = c("striped"))
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
glm	0.59	0.63	0.67	0.66	0.68	0.72	0
gbm	0.70	0.72	0.74	0.74	0.75	0.78	0
rf	0.70	0.75	0.79	0.78	0.81	0.81	0
earth	0.67	0.68	0.70	0.70	0.72	0.76	0

Hide

```
GlobalModelResults<-resamples(global_train)
kable(Global.Mod.Accuracy$statistics$Kappa,digits=2) %>%
kable_styling(bootstrap_options = c("striped"))
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
glm	0.19	0.26	0.34	0.32	0.36	0.44	0
gbm	0.41	0.44	0.48	0.47	0.49	0.57	0

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
rf	0.41	0.51	0.57	0.55	0.61	0.63	0
earth	0.35	0.36	0.40	0.40	0.43	0.52	0

Hide

```
Global.Mod.Cor<-modelCor(resamples(global_train))# shows correlation among models.Weakly corr
elated algorithms are persuasive for stacking them in ensemble.
kable(Global.Mod.Cor,digits=2)%>%
kable_styling(bootstrap_options = c("striped"))
```

	glm	gbm	rf	earth
glm	1.00	0.65	0.39	0.23
gbm	0.65	1.00	0.59	0.39
rf	0.39	0.59	1.00	0.11
earth	0.23	0.39	0.11	1.00

Create ensemble model (combine individual models into one)

Hide

```
set.seed(478)
global_stack <- caretEnsemble(
  global_train,
  trControl=trainControl(method="cv",
                         number=10,
                         savePredictions= "final",classProbs=TRUE ))
print(global_stack)
```

Function to return threshold where sens=spec from caret results

Hide

File failed to load: /extensions/MathZoom.js

```
findThresh<-function(df){
  df[c("rowIndex","obs","present")]
  df<-df %>%
    mutate(observed= ifelse(obs == "present",1,0)) %>%
    select(rowIndex,observed,predicted=present)
  result<-PresenceAbsence::optimal.thresholds(df,opt.methods = 2)
  return(result)
}

#accuracy measures
accuracyStats<-function(df,y){
  df[c("rowIndex","obs","present")]
  df<-df %>%
    mutate(observed= ifelse(obs == "present",1,0)) %>%
    select(rowIndex,observed,predicted=present)
  result<-PresenceAbsence::presence.absence.accuracy(df,threshold = y,st.dev=FALSE)
  return(result)
}
```

Identify threshold and performance of global ensemble model

Hide

```
global.ens.thresh<-findThresh(global_stack$ens_model$pred)
accuracyStats(global_stack$ens_model$pred,global.ens.thresh$predicted)
```

model <chr>	threshold <dbl>	PCC <dbl>	sensitivity <dbl>	specificity <dbl>	Kappa <dbl>	AUC <dbl>
predicted	0.54	0.7694118	0.7706767	0.7681433	0.5388214	0.8535057

1 row

Create rasterstack of CHELSA climate data clipped to European modeling extent for prediction

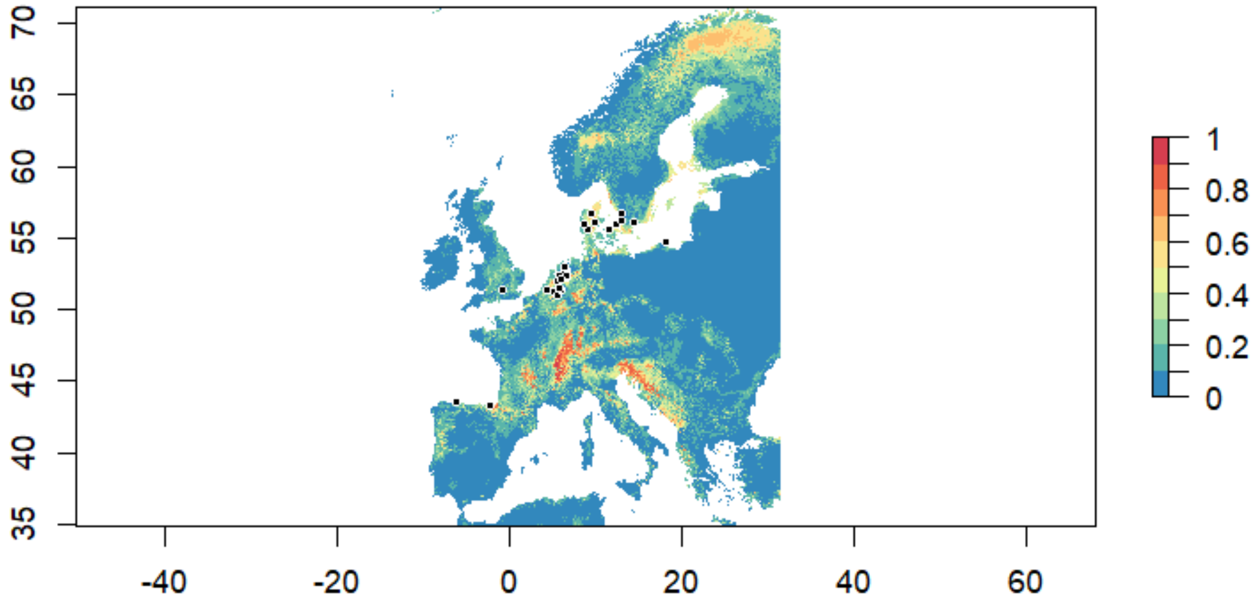
Hide

```
euclimrasters <- list.files((here("./data/external/climate/chelsa_eu_clips")),pattern='tif',f
ull.names = T)
eu_climpreds<-stack(euclimrasters)
eu_climpreds.10<-divide10(eu_climpreds) # correct for integer format of Chelsa preds
```

Restrict global model prediction to the extent of Europe

Plot global model prediction

File failed to load: /extensions/MathZoom.js



Export global model prediction

Hide

```
writeRaster(global_model, filename=file.path(rasterOutput,paste("GlobalEnsEU_",taxonkey, ".ti
f",sep="")),format="GTiff",overwrite=TRUE)
```

Get variable importance of global model

Hide

```
variableImportance_global<-varImp(global_stack)
kable(variableImportance_global,digits=2,caption="Variable Importance") %>%
kable_styling(bootstrap_options = c("striped"))
```

Variable Importance

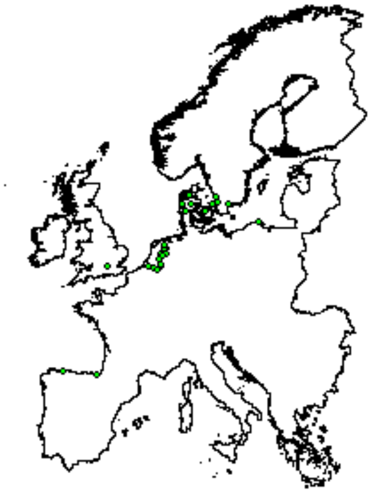
	overall	glm	gbm	rf	earth
CHELSA_precipSeasonality	2.23	6.98	0.00	0.00	26.27
CHELSA_annPrecip	21.56	0.00	14.56	21.15	39.88
CHELSA_temp_annRange	24.11	35.92	26.78	24.08	18.45
CHELSA_maxTmpWarmestMon	25.68	33.43	25.55	26.72	15.40

	overall	glm	gbm	rf	earth
CHELSA_precipDriestMon	26.42	23.67	33.11	28.04	0.00

Hide

```
write.csv(variableImportance_global,file = paste0(genOutput,taxonkey,"_varImp_global_model.cs
v"))
```

Create European subset



Create RasterStack of European climate variables from RMI

stack climate data

Hide

```
rmiclimrasters <- list.files((here("./data/external/climate/rmi_corrected")),pattern='tif',fu
ll.names = T)
rmiclimrasters #shows all available climate data
```

File failed to load: /extensions/MathZoom.js

```
 [1] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/anngdd100.tif"
 [2] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/annprecip_eea.tif"
 [3] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/annpvarrecip_eea.tif"
 [4] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/anntemp_eea.tif"
 [5] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/dristprec.tif"
 [6] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/maxtemp.tif"
 [7] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/mintemp.tif"
 [8] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/pet100.tif"
 [9] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/SolRad100.tif"
[10] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/temprang.tif"
[11] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/tempseas.tif"
[12] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/varSolRad100.tif"
[13] "C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/./data/external/climate/rmi_correct
ed/wettprec.tif"
```

Hide

```
rmiclimpreds <- stack(rmiclimrasters) #includes all available climate data
```

Transform eu occurrence dataset with unique presences back to a SpatialPoints dataframe.

Hide

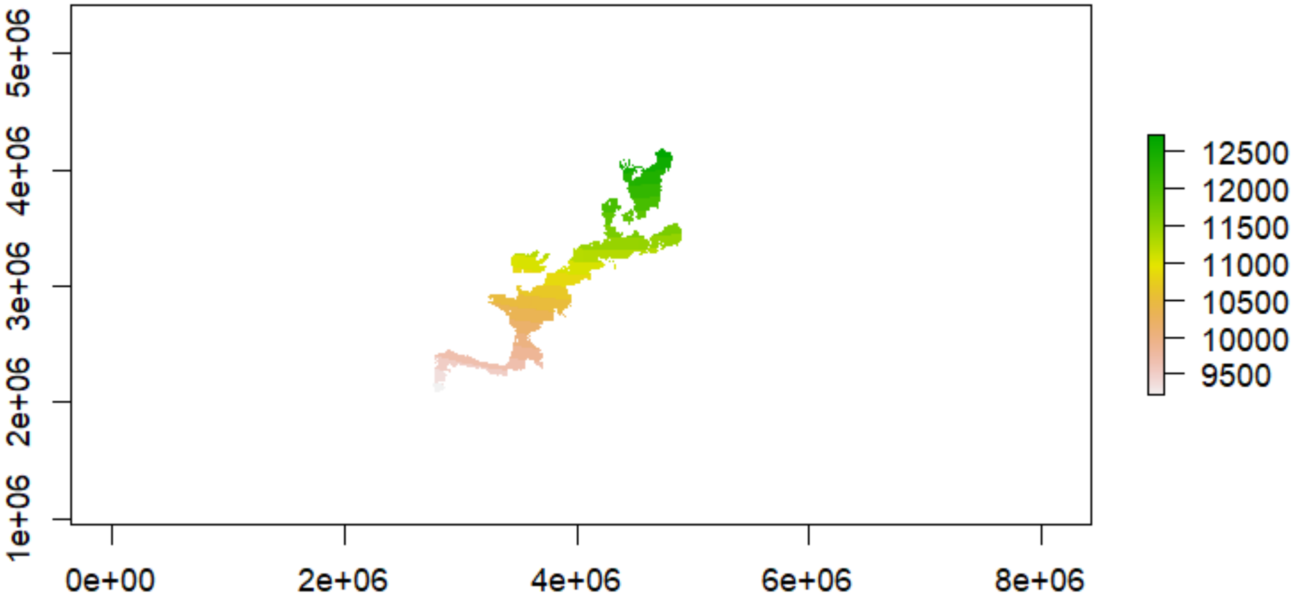
```
euocc<-as.data.frame(occ.eu@coords)
coordinates(euocc)<- c("longitude", "latitude")
euocc$occ<- rep(1,length(euocc$latitude))#adds columns indicating species presence needed for
modeling
proj4string(euocc)<-CRS("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0")#sp
ecify here the existing projection of the data
LLproj<-CRS("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0")

rmiproj<-CRS("+proj=laea +lat_0=52 +lon_0=10 +x_0=4321000 +y_0=3210000 +ellps=GRS80 +towgs84=
0,0,0,-0,-0,-0,0 +units=m +no_defs")
euocc1<-spTransform(euocc,rmiproj)
```

File failed to load: /extensions/MathZoom.js

Clip bias grid to European extent

Hide

```
studyextent<-euboundary
ecoregions_eu<-crop(biasgrid_sub,studyextent)
biasgrid_eu<-projectRaster(ecoregions_eu,rmiclimpreds)
plot(biasgrid_eu)
plot(studyextent,add=TRUE)
```

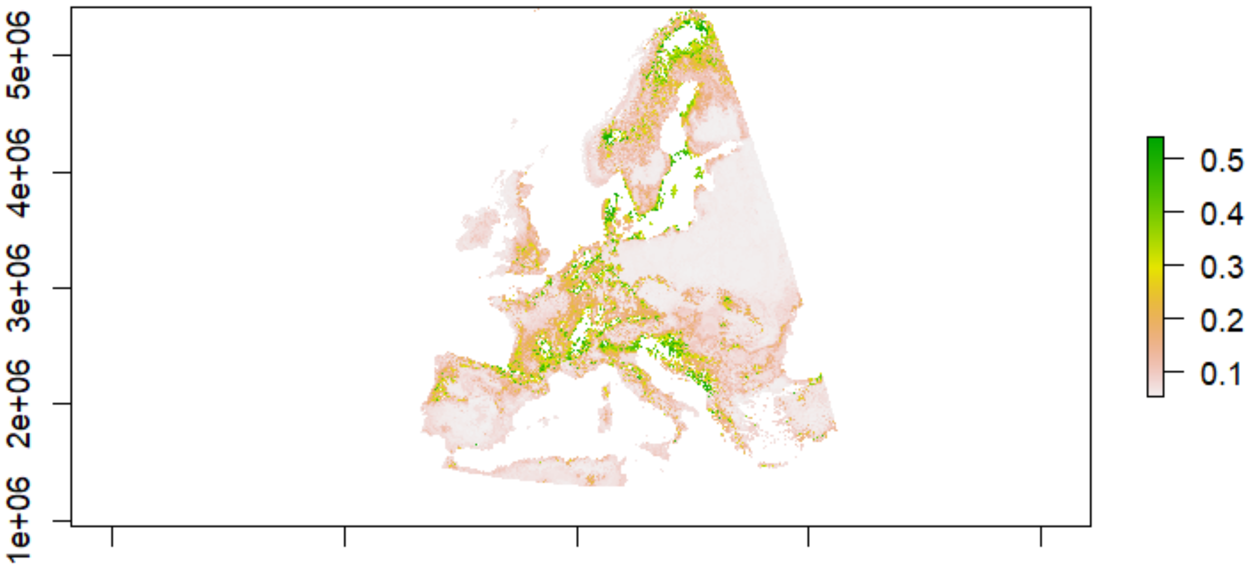


Mask areas of high habitat suitability from global climate model

Hide

```
wgs84_gcs<-CRS("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0")
crs(global_model)<-wgs84_gcs
#m<-global_model >.5
m<-global_model >= global.ens.thresh$predicted
global_mask<-mask(global_model,m,maskvalue=TRUE)
global_masked_proj<-projectRaster(global_mask,biasgrid_eu)
plot(global_masked_proj)
```

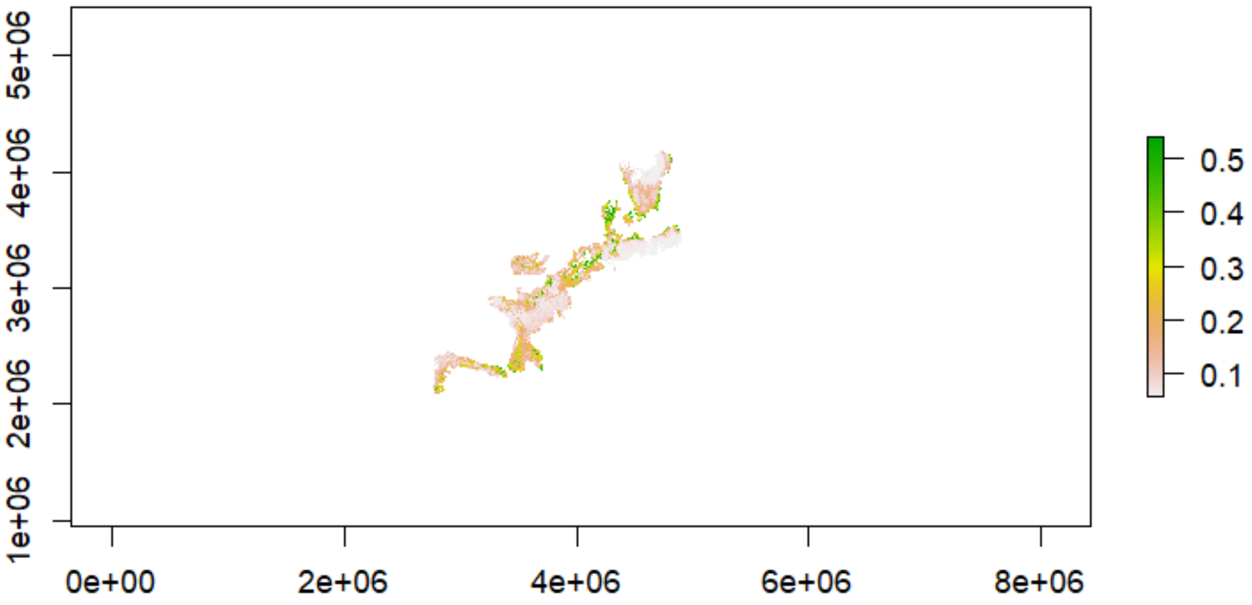
File failed to load: /extensions/MathZoom.js



Combine areas of low predicted habitat suitability with bias grid to exclude low sampled areas and areas of high suitability

Hide

```
pseudoSamplingArea<-mask(global_masked_proj,biasgrid_eu)
plot(pseudoSamplingArea)
```



Randomly locate pseudo absences within “pseudoSamplingArea”

Hide

```
# set number of pseudoabsences equal to the number of presences
numb.eu.pseudoabs<-nrow(euocc1)

# takes 10 draws of random pseudoabsences, returns as dataframes and names them X1-X10
setlist<-seq(1,10,1)
set.seed(120)
pseudoabs_pts<-lapply(setlist,function(x) as.data.frame(randomPoints(pseudoSamplingArea, num
b.eu.pseudoabs , euocc, ext=NULL, extf=1.1, excludep=TRUE, prob=FALSE,cellnumbers=FALSE, tryf
=50, warn=2, lonlatCorrection=TRUE)))
names(pseudoabs_pts)<-paste0("X",setlist)
```

Prepare occurrence (presence-pseudoabsence) datasets for modelling

Hide

File failed to load: /extensions/MathZoom.js

```
# extract data from predictors for absences
pseudoabs_pts1<-lapply(pseudoabs_pts, function(x) raster::extract(rmiclimpreds,x))

# add absence indicator
add.occ<-function(x,y){
occ<-rep(y,nrow(x))
cbind(x,occ)
}

pseudoabs_pts2<-lapply(pseudoabs_pts1, function(x) add.occ(x,0))

# extract eu presences and add presence indicator
presence<-as.data.frame(euocc1@coords)
names(presence)<- c("x","y")
presence1<-raster::extract(rmiclimpreds,presence)
occ<-rep(1,nrow(presence1))
presence1<-cbind(presence1,occ)

# join each pseudoabsence set with presences
eu_presabs.pts<-lapply(pseudoabs_pts2, function(x) rbind(x,presence1))
eu_presabs.coord<-lapply(pseudoabs_pts, function(x) rbind(x,presence))
```

Identify highly correlated climate predictors from training data

Hide

```
# convert eu data to dataframe
eu_presabs.pts.df<-lapply(eu_presabs.pts,function(x) as.data.frame(x))

# find attributes that are highly corrected
highlyCorrelated_climate <-lapply(names(eu_presabs.pts.df),function(x) findCorrelation(cor(eu_presabs.pts.df[[x]],use = 'complete.obs'), cutoff=0.7,exact=TRUE,names=TRUE))

highlyCorrelated_climate
```

File failed to load: /extensions/MathZoom.js

```
[[1]]
[1] "pet100"        "anntemp_eea"   "anngdd100"     "mintemp"       "annprecip_eea" "tempsea
s"      "SolRad100"    
[8] "wettprec"      "dristprec"    

[[2]]
[1] "pet100"        "anntemp_eea"   "mintemp"       "SolRad100"     "anngdd100"     "tempsea
s"      "annprecip_eea"
[8] "wettprec"     

[[3]]
[1] "pet100"        "anntemp_eea"   "tempseas"      "mintemp"       "annprecip_eea" "wettpre
c"      "dristprec"    

[[4]]
[1] "pet100"        "anntemp_eea"   "anngdd100"     "mintemp"       "tempseas"      "annpreci
p_eea" "wettprec"     
[8] "dristprec"    

[[5]]
[1] "anntemp_eea"   "anngdd100"     "pet100"        "mintemp"       "SolRad100"     "tempsea
s"      "annprecip_eea"
[8] "wettprec"      "dristprec"    

[[6]]
[1] "anntemp_eea"   "pet100"        "anngdd100"     "mintemp"       "tempseas"      "SolRad10
0"     "annprecip_eea"
[8] "wettprec"     

[[7]]
[1] "pet100"        "tempseas"      "anngdd100"     "anntemp_eea"   "mintemp"       "SolRad10
0"     "annprecip_eea"
[8] "wettprec"     

[[8]]
[1] "anntemp_eea"   "tempseas"      "pet100"        "mintemp"       "SolRad100"     "annpreci
p_eea" "wettprec"     

[[9]]
[1] "anntemp_eea"   "anngdd100"     "mintemp"       "pet100"        "tempseas"      "dristpre
c"     "annprecip_eea"

[[10]]
[1] "anntemp_eea"   "anngdd100"     "pet100"        "mintemp"       "annprecip_eea" "tempsea
s"      "dristprec"    
```

Hide

File failed to load: /extensions/MathZoom.js

```
eupreds<-as.data.frame(highlyCorrelated_climate[1])
kable(eupreds) %>%
kable_styling(bootstrap_options = c("striped"))
```

c..pet100....anntemp_eea....anngdd100....mintemp....annprecip_eea...
pet100
anntemp_eea
anngdd100
mintemp
annprecip_eea
tempseas
SolRad100
wettprec
dristprec

Reomve highly correlated climate predictors from training data

Hide

```
drop_climate<-highlyCorrelated_climate[[1]]
rmiclimpreds_uncor<-dropLayer(rmiclimpreds,drop_climate)
```

Add habitat and anthropogenic predictors

Identify highly correlated predictors from the habitat/anthropogenic/climate stack (full stack)

Hide

```
# find attributes that are highly correlated
highlyCorrelated_full <-lapply(names(occ.full.data),function(x) findCorrelation(cor(occ.full.
data[[x]],use = 'complete.obs'), cutoff=0.7,exact=TRUE,names=TRUE))
highlyCorrelated_vec<-unlist(highlyCorrelated_full)
eupreds1<-as.data.frame(highlyCorrelated_vec)
kable(eupreds1) %>%
kable_styling(bootstrap_options = c("striped"))
```

File failed to load: /extensions/MathZoom.js

highlyCorrelated_vec

Remove highly correlated predictors from full stack

Hide

```
occ.full.data<-sapply(names(occ.full.data),function (x) occ.full.data[[x]][,!(colnames(occ.full.data[[x]]) %in% highlyCorrelated_vec)],simplify=FALSE)
```

Identify and remove near zero variance predictors

Hide

```
# identify low variance predictors
nzv_preds<-lapply(names(occ.full.data),function(x) nearZeroVar(occ.full.data[[x]],names=TRUE))
nzv_preds
```

```
[[1]]
character(0)

[[2]]
character(0)

[[3]]
character(0)

[[4]]
character(0)

[[5]]
character(0)

[[6]]
character(0)

[[7]]
character(0)

[[8]]
character(0)

[[9]]
character(0)

[[10]]
character(0)
```

File failed to load: /extensions/MathZoom.js

Hide

```
nzv_preds.vec<-unique(unlist(nzv_preds))
nzv_preds.vec
```

```
character(0)
```

Hide

```
# remove near zero variance predictors. They don't contribute to the model.
occ.full.data<-sapply(names(occ.full.data),function (x) occ.full.data[[x]][,!(colnames(occ.fu
ll.data[[x]]) %in% nzv_preds.vec)],simplify=FALSE)
```

Build models with climate and habitat data

Hide

File failed to load: /extensions/MathZoom.js

```
# prepare data for modeling

occ.full.data.df<-lapply(occ.full.data, function(x) as.data.frame(x))

occ.full.data.df<- sapply(names(occ.full.data.df), function (x) cbind(occ.full.data.df[[x]],o
cc=eu_presabs.pts.df[[x]]$occ, deparse.level=0),simplify=FALSE)




factorVars<-function(df,var){
df[,c(var)]<-as.factor(df[,c(var)])
levels(df[,c(var)])<-c("absent","present")
df[,c(var)]<-relevel(df[,c(var)], ref = "present")
return(df)
}



occ.full.data.factor<-sapply(names(occ.full.data.df), function (x) factorVars(occ.full.data.d
f[[x]], "occ"),simplify=FALSE)
occ.full.data.forCaret<-sapply(names(occ.full.data.factor), function (x) replace(occ.full.dat
a.factor[[x]], is.na(occ.full.data.factor[[x]]),0),simplify=FALSE)




# uncomment 2nd control options for  LOOCV (leave one out cross validation, which is aka as "
jacknife" ) which should be used when occurrences are smaller than n=10 for each predictor in
the model)

#control<-trainControl(method="LOOCV",savePredictions="final", preProc=c("center","scale"),cl
assProbs=TRUE)
control <- trainControl(method="cv",number=4,savePredictions="final", preProc=c("center","sca
le"),classProbs=TRUE)
mylist<-list(
  glm =caretModelSpec(method = "glm",maxit=100),
  gbm= caretModelSpec(method = "gbm"),
  rf = caretModelSpec(method = "rf", importance = TRUE),
  earth= caretModelSpec(method = "earth"))

# set.seed(167)
 eu_models<-sapply(names(occ.full.data.forCaret), function(x) model_train_habitat <- caretLis
t(
   occ~temprang + maxtemp + annpvarrecip_eea + corine_perWetland, data= occ.full.data.forCare
t[[x]],
   trControl=control,
    tuneList=mylist), simplify=FALSE)
```

Display model evaluation statistics

Hide

File failed to load: /extensions/MathZoom.js

```
EU_ModelResults1<-sapply(names(eu_models), function(x) resamples(eu_models[[x]]),simplify=FALSE)
Results.summary<-sapply(names(EU_ModelResults1), function(x) summary(EU_ModelResults1[[x]]),simplify=FALSE)
Results.summary
```

File failed to load: /extensions/MathZoom.js

```
$X1

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth 
Number of resamples: 4 

Accuracy 
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.6875000 0.7173295 0.7312834 0.7420622 0.7560160 0.8181818    0
gbm   0.7941176 0.8079044 0.8304924 0.8334726 0.8560606 0.8787879    0
rf    0.7812500 0.8089489 0.8355615 0.8327902 0.8594029 0.8787879    0
earth 0.7500000 0.8011364 0.8649733 0.8548351 0.9186720 0.9393939    0

Kappa 
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.3750000 0.4335290 0.4618135 0.4836641 0.5119485 0.6360294    0
gbm   0.5882353 0.6158088 0.6605663 0.6664564 0.7112138 0.7564576    0
rf    0.5625000 0.6166398 0.6702843 0.6651054 0.7187500 0.7573529    0
earth 0.5000000 0.6010148 0.7291079 0.7091111 0.8372043 0.8782288    0


$X2

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth 
Number of resamples: 4 

Accuracy 
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.6764706 0.6764706 0.6819853 0.6897978 0.6953125 0.7187500    0
gbm   0.7352941 0.8166360 0.8630515 0.8419118 0.8883272 0.9062500    0
rf    0.7647059 0.8239890 0.8593750 0.8414522 0.8768382 0.8823529    0
earth 0.7647059 0.7867647 0.8345588 0.8272059 0.8750000 0.8750000    0

Kappa 
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.3529412 0.3529412 0.3639706 0.3795956 0.3906250 0.4375000    0
gbm   0.4705882 0.6332721 0.7261029 0.6838235 0.7766544 0.8125000    0
rf    0.5294118 0.6479779 0.7187500 0.6829044 0.7536765 0.7647059    0
earth 0.5294118 0.5735294 0.6691176 0.6544118 0.7500000 0.7500000    0


$X3

Call:
summary.resamples(object = EU_ModelResults1[[x]])
```

File failed to load: /extensions/MathZoom.js

```
Models: glm, gbm, rf, earth
Number of resamples: 4

Accuracy
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.6060606 0.6588681 0.6819853 0.6667502 0.6898674 0.6969697    0
gbm   0.7878788 0.7925579 0.8501838 0.8569101 0.9145360 0.9393939    0
rf    0.7878788 0.8366756 0.8795956 0.8716160 0.9145360 0.9393939    0
earth 0.6470588 0.7299465 0.8319129 0.8125696 0.9145360 0.9393939    0

Kappa
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.2070240 0.3164619 0.3639706 0.3326424 0.3801511 0.3956044    0
gbm   0.5776965 0.5856006 0.7003676 0.7142771 0.8290441 0.8786765    0
rf    0.5792350 0.6742205 0.7591912 0.7440734 0.8290441 0.8786765    0
earth 0.2941176 0.4608920 0.6644918 0.6254444 0.8290441 0.8786765    0


$X4

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth
Number of resamples: 4

Accuracy
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.5000000 0.5568182 0.6060606 0.5999053 0.6491477 0.6875000    0
gbm   0.7272727 0.7911932 0.8180147 0.8180983 0.8449198 0.9090909    0
rf    0.7187500 0.7933239 0.8502674 0.8320939 0.8890374 0.9090909    0
earth 0.7187500 0.7752757 0.8061497 0.7873078 0.8181818 0.8181818    0

Kappa
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.0000000 0.1107011 0.2124869 0.1999934 0.3017792 0.3750000    0
gbm   0.4510166 0.5815042 0.6360294 0.6351888 0.6897140 0.8176796    0
rf    0.4375000 0.5853898 0.6996961 0.6636429 0.7779493 0.8176796    0
earth 0.4375000 0.5505515 0.6107843 0.5734387 0.6336716 0.6346863    0


$X5

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth
Number of resamples: 4

Accuracy
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
```

```
glm   0.6060606 0.6588681 0.6976103 0.7124415 0.7511837 0.8484848    0
gbm   0.6470588 0.7242647 0.7537879 0.7507799 0.7803030 0.8484848    0
rf    0.7058824 0.7219251 0.7698864 0.7735350 0.8214962 0.8484848    0
earth 0.4705882 0.6176471 0.7239583 0.6917474 0.7980587 0.8484848    0

Kappa 
             Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm    0.21284404 0.3179169 0.3952206 0.4251332 0.5024369 0.6972477    0
gbm    0.29411765 0.4485294 0.5064576 0.5013469 0.5592750 0.6983547    0
rf     0.41176471 0.4427202 0.5390193 0.5467628 0.6430619 0.6972477    0
earth -0.05882353 0.2339129 0.4469959 0.3833807 0.5964637 0.6983547    0


$X6

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth 
Number of resamples: 4 

Accuracy 
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.6176471 0.6465993 0.6718750 0.6741728 0.6994485 0.7352941    0
gbm   0.7187500 0.8593750 0.9090074 0.8694853 0.9191176 0.9411765    0
rf    0.7812500 0.8570772 0.8823529 0.8630515 0.8883272 0.9062500    0
earth 0.7187500 0.8193934 0.8639706 0.8322610 0.8768382 0.8823529    0

Kappa 
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.2352941 0.2931985 0.3437500 0.3483456 0.3988971 0.4705882    0
gbm   0.4375000 0.7187500 0.8180147 0.7389706 0.8382353 0.8823529    0
rf    0.5625000 0.7141544 0.7647059 0.7261029 0.7766544 0.8125000    0
earth 0.4375000 0.6387868 0.7279412 0.6645221 0.7536765 0.7647059    0


$X7

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth 
Number of resamples: 4 

Accuracy 
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.6470588 0.6844920 0.7078598 0.7353916 0.7587595 0.8787879    0
gbm   0.7272727 0.8215241 0.8795956 0.8564645 0.9145360 0.9393939    0
rf    0.8181818 0.8221925 0.8336397 0.8486380 0.8600852 0.9090909    0
earth 0.7187500 0.7478693 0.8052585 0.8020137 0.8594029 0.8787879    0
```

File failed to load: /extensions/MathZoom.js

```
Kappa
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.2941176 0.3685662 0.4154412 0.4703644 0.5172394 0.7564576    0
gbm   0.4510166 0.6421659 0.7591912 0.7119069 0.8289322 0.8782288    0
rf    0.6346863 0.6439657 0.6672794 0.6968984 0.7202122 0.8183486    0
earth 0.4375000 0.4940613 0.6093987 0.6034126 0.7187500 0.7573529    0


$X8

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth
Number of resamples: 4

Accuracy
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.6969697 0.7477718 0.7729779 0.7577011 0.7829072 0.7878788    0
gbm   0.7272727 0.7500000 0.7905526 0.7880320 0.8285846 0.8437500    0
rf    0.7575758 0.8070410 0.8511586 0.8493483 0.8934659 0.9375000    0
earth 0.7187500 0.7251420 0.7424242 0.7714879 0.7887701 0.8823529    0

Kappa
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.3888889 0.4942810 0.5459559 0.5130573 0.5647321 0.5714286    0
gbm   0.4489796 0.4969312 0.5799870 0.5741134 0.6571691 0.6875000    0
rf    0.5092937 0.6126175 0.7013072 0.6967270 0.7854167 0.8750000    0
earth 0.4375000 0.4476375 0.4819659 0.5415344 0.5758628 0.7647059    0


$X9

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth
Number of resamples: 4

Accuracy
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.4848485 0.5075758 0.5909091 0.5833333 0.6666667 0.6666667    0
gbm   0.7272727 0.7954545 0.8484848 0.8257576 0.8787879 0.8787879    0
rf    0.7878788 0.8106061 0.8333333 0.8333333 0.8560606 0.8787879    0
earth 0.8181818 0.8181818 0.8484848 0.8560606 0.8863636 0.9090909    0

Kappa
             Min.    1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   -0.02185792 0.01120219 0.1768570 0.1694505 0.3351053 0.3459459    0
gbm    0.45504587 0.59178345 0.6973578 0.6517786 0.7573529 0.7573529    0
rf     0.57614679 0.62305130 0.6679669 0.6675806 0.7124962 0.7582418    0
```

File failed to load: /extensions/MathZoom.js

```
earth  0.63602941 0.63802195 0.6984639 0.7128265 0.7732685 0.8183486    0


$X10

Call:
summary.resamples(object = EU_ModelResults1[[x]])

Models: glm, gbm, rf, earth
Number of resamples: 4

Accuracy
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.5000000 0.6323529 0.6867201 0.6501783 0.7045455 0.7272727    0
gbm   0.7058824 0.7624081 0.8300189 0.8111770 0.8787879 0.8787879    0
rf    0.7058824 0.7155331 0.7987689 0.7955520 0.8787879 0.8787879    0
earth 0.6060606 0.6588681 0.6976103 0.7200173 0.7587595 0.8787879    0

Kappa
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm   0.0000000 0.2647059 0.3697791 0.2986510 0.4037243 0.4550459    0
gbm   0.4117647 0.5248162 0.6590278 0.6217933 0.7560049 0.7573529    0
rf    0.4117647 0.4310662 0.5965278 0.5900940 0.7555556 0.7555556    0
earth 0.2070240 0.3164619 0.3952206 0.4387045 0.5174632 0.7573529    0
```

Hide

```
Model.cor<-sapply(names(eu_models), function(x) modelCor(resamples(eu_models[[x]])),simplify=
FALSE)
Model.cor
```

File failed to load: /extensions/MathZoom.js

```
$X1
             glm        gbm         rf      earth
glm   1.00000000 0.30326929 0.91594845 0.09031046
gbm   0.30326929 1.00000000 0.06596895 0.32898325
rf    0.91594845 0.06596895 1.00000000 0.32786495
earth 0.09031046 0.32898325 0.32786495 1.00000000

$X2
            glm       gbm        rf     earth
glm   1.0000000 0.1736679 0.4472209 0.7522523
gbm   0.1736679 1.0000000 0.8228504 0.3250041
rf    0.4472209 0.8228504 1.0000000 0.1859766
earth 0.7522523 0.3250041 0.1859766 1.0000000

$X3
            glm       gbm        rf     earth
glm   1.0000000 0.7401168 0.9341484 0.4460490
gbm   0.7401168 1.0000000 0.9276338 0.9293759
rf    0.9341484 0.9276338 1.0000000 0.7244947
earth 0.4460490 0.9293759 0.7244947 1.0000000

$X4
             glm        gbm         rf       earth
glm    1.0000000 0.24853091 -0.5756973 -0.55266626
gbm    0.2485309 1.00000000  0.4859857  0.04054434
rf    -0.5756973 0.48598574  1.0000000  0.84129193
earth -0.5526663 0.04054434  0.8412919  1.00000000

$X5
            glm       gbm        rf     earth
glm   1.0000000 0.6471093 0.8566327 0.6427260
gbm   0.6471093 1.0000000 0.8335193 0.9298555
rf    0.8566327 0.8335193 1.0000000 0.9297146
earth 0.6427260 0.9298555 0.9297146 1.0000000

$X6
              glm        gbm         rf       earth
glm    1.00000000 -0.2725299 -0.2135904 -0.03371465
gbm   -0.27252988  1.0000000  0.9491523  0.95262057
rf    -0.21359038  0.9491523  1.0000000  0.97537787
earth -0.03371465  0.9526206  0.9753779  1.00000000

$X7
            glm       gbm        rf     earth
glm   1.0000000 0.5831945 0.9665362 0.4195661
gbm   0.5831945 1.0000000 0.7527433 0.3655588
rf    0.9665362 0.7527433 1.0000000 0.5367592
earth 0.4195661 0.3655588 0.5367592 1.0000000

$X8
            glm       gbm         rf       earth
```

File failed to load: /extensions/MathZoom.js

```
glm   1.0000000 0.6335944  0.8702158  0.1901749
gbm   0.6335944 1.0000000  0.6873973  0.3302051
rf    0.8702158 0.6873973  1.0000000 -0.2211102
earth 0.1901749 0.3302051 -0.2211102  1.0000000

$X9
             glm       gbm         rf      earth
glm    1.0000000 0.7821110 -0.1209717 -0.2949949
gbm    0.7821110 1.0000000  0.4917225  0.2586267
rf    -0.1209717 0.4917225  1.0000000  0.9467293
earth -0.2949949 0.2586267  0.9467293  1.0000000

$X10
              glm       gbm        rf       earth
glm    1.00000000 0.3923705 0.6598885 -0.06515531
gbm    0.39237052 1.0000000 0.9489769  0.26297052
rf     0.65988854 0.9489769 1.0000000  0.23162709
earth -0.06515531 0.2629705 0.2316271  1.00000000
```

Create ensemble model

Hide

```
set.seed(458)

#hideoutput<-capture.output(
set.seed(458)
lm_ens_hab<-sapply(names(eu_models), function (x) caretEnsemble(eu_models[[x]], trControl=tra
inControl(method="cv",                                                              number=1
0,savePredictions= "final",classProbs = TRUE)),simplify=FALSE)
```

PDF export function

PNG export function

Use EU level ensemble models (each using a separate pseudoabsence draw) to predict at European level

Hide

```
 ens_pred_hab_eu1<-sapply(names(lm_ens_hab), function(x) raster::predict(fullstack,lm_ens_hab
[[x]],type="prob"),simplify=FALSE)
```

Use EU level ensemble models to predict for Belgium only

Evaluate the performance of each the EU level ensemble models based on results from CV

File failed to load: /extensions/MathZoom.js

2. Using thresholds identified for each model in the previous step, assess performance of each model

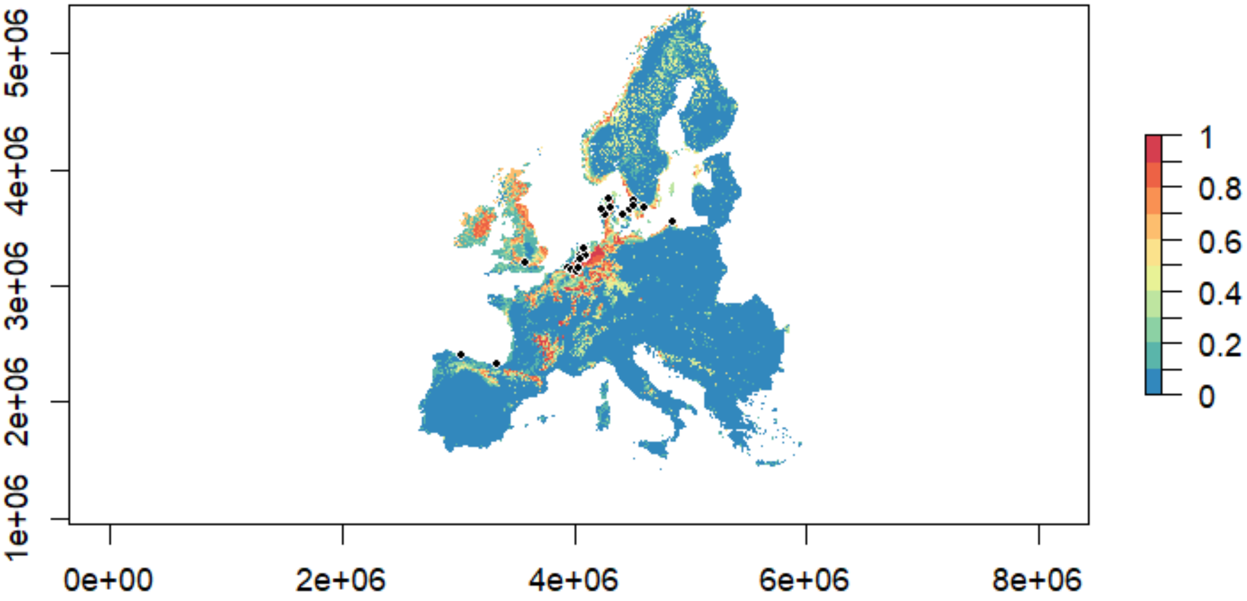
	model	threshold	PCC	sensitivity	specificity	Kappa	AUC
X1	predicted	0.48	0.85	0.85	0.85	0.70	0.89
X2	predicted	0.51	0.85	0.85	0.85	0.70	0.88
X3	predicted	0.50	0.85	0.85	0.85	0.70	0.89
X4	predicted	0.47	0.82	0.82	0.82	0.64	0.87
X5	predicted	0.52	0.77	0.77	0.76	0.53	0.85
X6	predicted	0.42	0.88	0.88	0.88	0.76	0.89
X7	predicted	0.46	0.83	0.83	0.83	0.67	0.91
X8	predicted	0.54	0.84	0.85	0.83	0.68	0.87
X9	predicted	0.55	0.82	0.82	0.82	0.64	0.88
X10	predicted	0.55	0.78	0.79	0.77	0.56	0.85

plot the best EU level ensemble model

Hide

```
# specify best model below
bestModel<-lm_ens_hab$X6



  brks <- seq(0, 1, by=0.1)
  nb <- length(brks)-1
  pal <- colorRampPalette(rev(brewer.pal(8, 'Spectral')))
   cols<-pal(nb)
  plot(ens_pred_hab_eu1$X6, breaks=brks, col=cols,lab.breaks=brks)# specify best model
  plot(euocc1,pch=21,cex=.8,col="white",add=TRUE)#plots species presences in 10 fold cv comme
nt this line to hide
```

File failed to load: /extensions/MathZoom.js

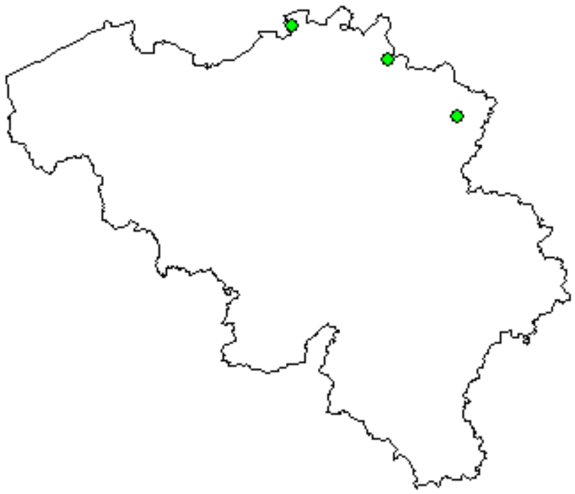


Hide

```
NA
```

Subset Belgium occurrences

File failed to load: /extensions/MathZoom.js

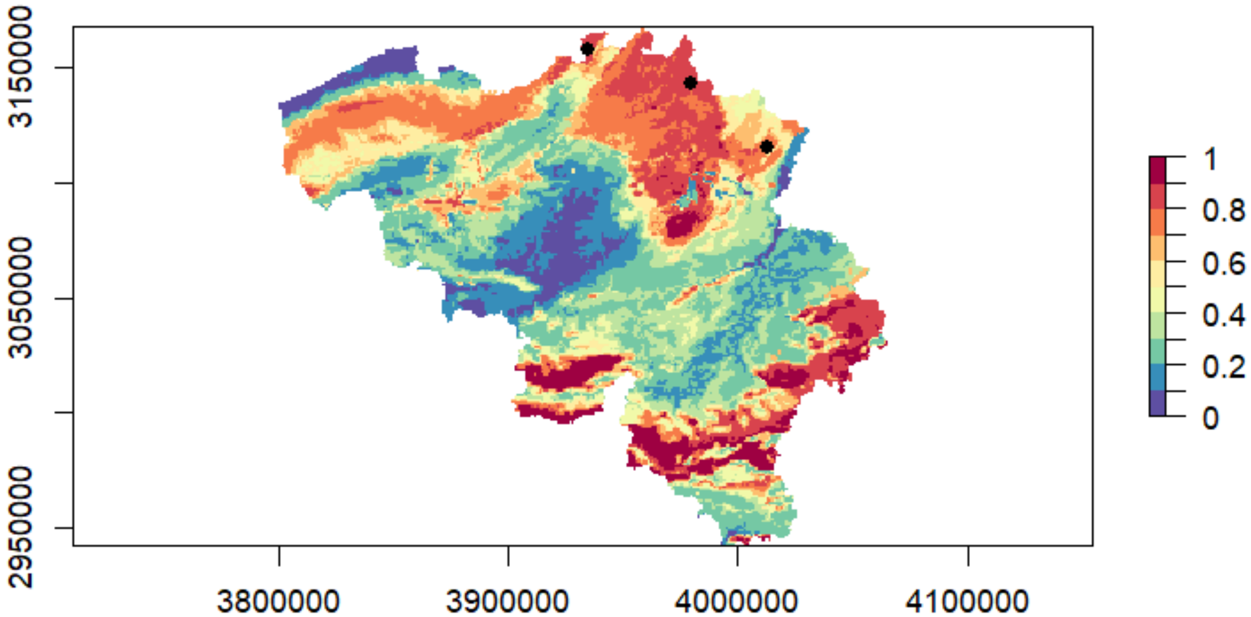


plot the best EU level ensemble model showing only Belgium

Hide

```
brks <- seq(0, 1, by=0.1)
  nb <- length(brks)-1
  pal <- colorRampPalette(rev(brewer.pal(11, 'Spectral')))
  cols<-pal(nb)
  plot(ens_pred_hab_be$X6, breaks=brks, col=cols,lab.breaks=brks) # specify best model
  plot(occ.country,pch=21,cex=1,add=TRUE)
```

File failed to load: /extensions/MathZoom.js



Hide

```
NA
```

Clip habitat raster stack to Belgium

Hide

```
habitat_stack<-stack(habitat)
habitat_only_stack<-crop(habitat_stack,country)
habitat_only_stack_be<-crop(habitat_only_stack,country)
```

Create individual RCP (2.6, 4.5, 8.5) climate raster stacks for Belgium

Hide

File failed to load: /extensions/MathZoom.js

```
be26 <- list.files((here("./data/external/climate/byEEA_finalRCP/belgium_rcps/rcp26")),pattern='tif',full.names = T)
belgium_stack26 <- stack(be26)

be45 <- list.files((here("./data/external/climate/byEEA_finalRCP/belgium_rcps/rcp45")),pattern='tif',full.names = T)
belgium_stack45 <- stack(be45)

be85 <- list.files((here("./data/external/climate/byEEA_finalRCP/belgium_rcps/rcp85")),pattern='tif',full.names = T)
belgium_stack85 <- stack(be85)
```

Combine habitat stacks with climate stacks for each RCP scenario

Hide

```
fullstack26<-stack(be26,habitat_only_stack_be)
fullstack45<-stack(be45,habitat_only_stack_be)
fullstack85<-stack(be85,habitat_only_stack_be)
```

Create and export RCP risk maps for each RCP scenario

Hide

```
ens_pred_hist<-raster::predict(fullstack_be,bestModel,type="prob")
ens_pred_hab26<-raster::predict(fullstack26,bestModel,type="prob")
crs(ens_pred_hab26)<-laea_grs80
writeRaster(ens_pred_hab26, filename=file.path(rasterOutput,paste("be_",taxonkey, "_rcp26.tif",sep="")), format="GTiff",overwrite=TRUE)
exportPDF(ens_pred_hab26,taxonkey,taxonName=taxonName,"rcp26.pdf")
```

```
null device 
          1 
```

Hide

```
ens_pred_hab45<-raster::predict(fullstack45,bestModel,type="prob")
crs(ens_pred_hab45)<-laea_grs80
writeRaster(ens_pred_hab45, filename=file.path(rasterOutput,paste("be_",taxonkey, "_rcp45.tif",sep="")), format="GTiff",overwrite=TRUE)
exportPDF(ens_pred_hab45,taxonkey,taxonName=taxonName,"rcp45.pdf")
```

```
null device 
          1 
```

Hide

File failed to load: /extensions/MathZoom.js

```
ens_pred_hab85<-raster::predict(fullstack85,bestModel,type="prob")
crs(ens_pred_hab85)<-laea_grs80
writeRaster(ens_pred_hab85, filename=file.path(rasterOutput,paste("be_",taxonkey, "_rcp85.ti
f",sep="")), format="GTiff",overwrite=TRUE)
exportPDF(ens_pred_hab85,taxonkey,taxonName=taxonName,"rcp85.pdf")
```

```
null device 
          1 
```

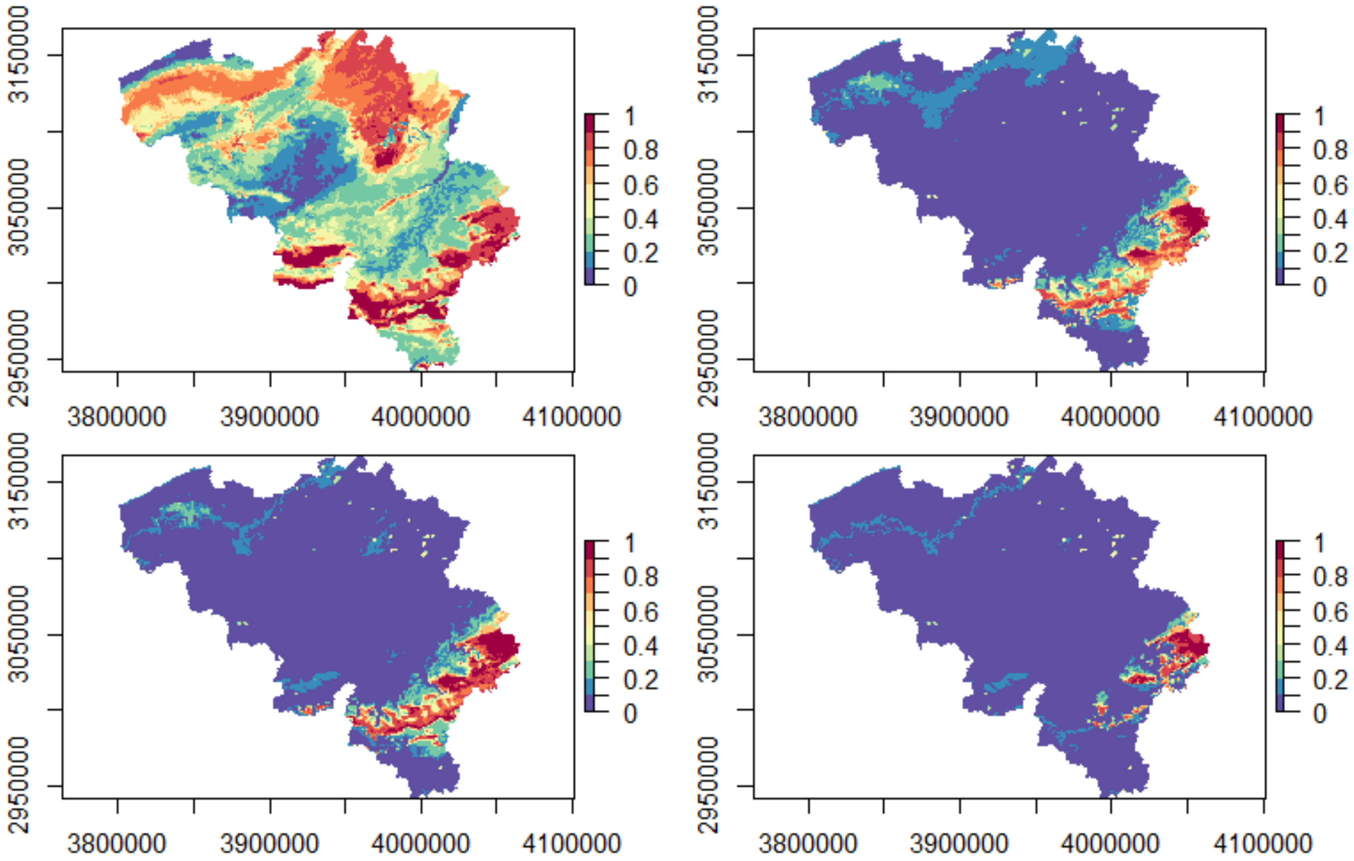
Create and export RCP risk maps for each RCP scenario

Hide

```


par(mfrow=c(2,2), mar= c(2,3,0.8,0.8))
plot(ens_pred_hist,breaks=brks, col=cols,lab.breaks=brks)
plot(ens_pred_hab26,breaks=brks, col=cols,lab.breaks=brks)
```

Hide

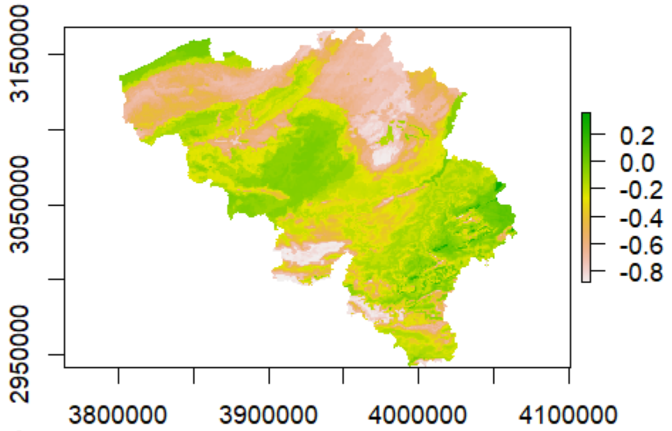
```
plot(ens_pred_hab45,breaks=brks, col=cols,lab.breaks=brks)
plot(ens_pred_hab85,breaks=brks, col=cols,lab.breaks=brks)
```

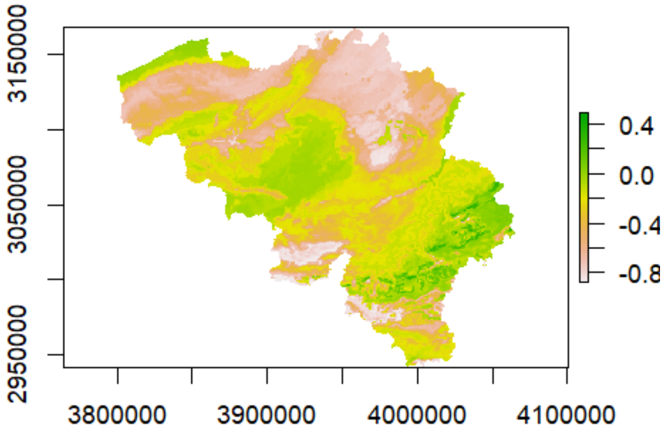


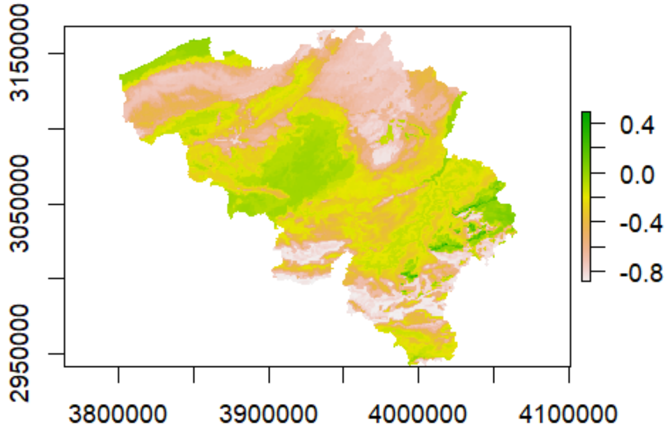
Create and export “difference maps”: the difference between predicted risk by each RCP scenario and historical climate

File failed to load: /extensions/MathZoom.js

```
null device 
          1 
null device 
          1 
null device 
          1 
```







Check spatial autocorrelation of residuals to assess whether occurrence data should be thinned

derive residuals from best model

Hide

```
predEns1<-bestModel$ens_model$pred
obs.numeric<-ifelse(predEns1$obs == "absent",0,1)
```

standardize residuals

Hide

File failed to load: /extensions/MathZoom.js

```
stdres<-function(obs.numeric, yhat){
  num<-obs.numeric-yhat
  denom<-sqrt(yhat*(1-yhat))
  return(num/denom)
}
hab.res<-stdres(obs.numeric,predEns1$present)

# specify corresponding model number from eu_presabs.coord datafile to join data with xy locations. If best model is "X1", join with eu_presabs.coord$X1



res.best.coords1<-cbind(coordinates(eu_presabs.coord$X1),occ.full.data.forCaret$X1)
removedNAs.coords<-na.omit(res.best.coords1)
res.best.coords<-cbind(removedNAs.coords,hab.res)
res.best.geo<-as.geodata(res.best.coords,coords.col=1:2,data.col = 3)
summary(res.best.geo) #note distance is in meters
```

```
Number of data points: 132

Coordinates summary
          x       y
min 2786500 2144500
max 4849027 4162500

Distance summary
         min          max
    519.7506 2766294.2721

Data summary
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
 8.71000 15.38500 16.31000 20.57402 24.82250 67.62000
```

Check Morans I.

Hide

```
#If Moran's I is very low (<0.10), or not significant, do not need to thin occurrences.
library(ape)
```

File failed to load: /extensions/MathZoom.js

```
Warning: package ‘ape’ was built under R version 4.2.3
Attaching package: ‘ape’

The following object is masked from ‘package:dplyr’:

    where

The following objects are masked from ‘package:raster’:

    rotate, zoom
```

Hide

```
res.best.df<-as.data.frame(res.best.coords)
occ.dists <- as.matrix(dist(cbind(res.best.df[1], res.best.df[2])))
occ.dists.inv <- 1/occ.dists
diag(occ.dists.inv) <- 0
Moran.I(res.best.df$hab.res,occ.dists.inv,scaled=TRUE,alternative="greater")
```

```
$observed
[1] 0.005079273

$expected
[1] -0.007633588

$sd
[1] 0.02993293

$p.value
[1] 0.3355235
```

Code for Mondrian conformal prediction functions

Quantify confidence of predicted values using class conformal prediction

Hide

File failed to load: /extensions/MathZoom.js

```
# quantify confidence for country level predictions based on historical climate and under RCP scenarios of climate change

set.seed(1609)
pvalsdf_hist<-classConformalPrediction(bestModel,ens_pred_hist)
set.seed(447)
pvalsdf_rcp26<-classConformalPrediction(bestModel,ens_pred_hab26)
set.seed(568)
pvalsdf_rcp45<-classConformalPrediction(bestModel,ens_pred_hab45)
set.seed(988)
pvalsdf_rcp85<-classConformalPrediction(bestModel,ens_pred_hab85)

# option to export confidence and pvals as csv
# write.csv(pvalsdf_hist,file=paste(genOutput,"confidence_",taxonkey, "_hist.csv",sep=""))
```

Create confidence maps

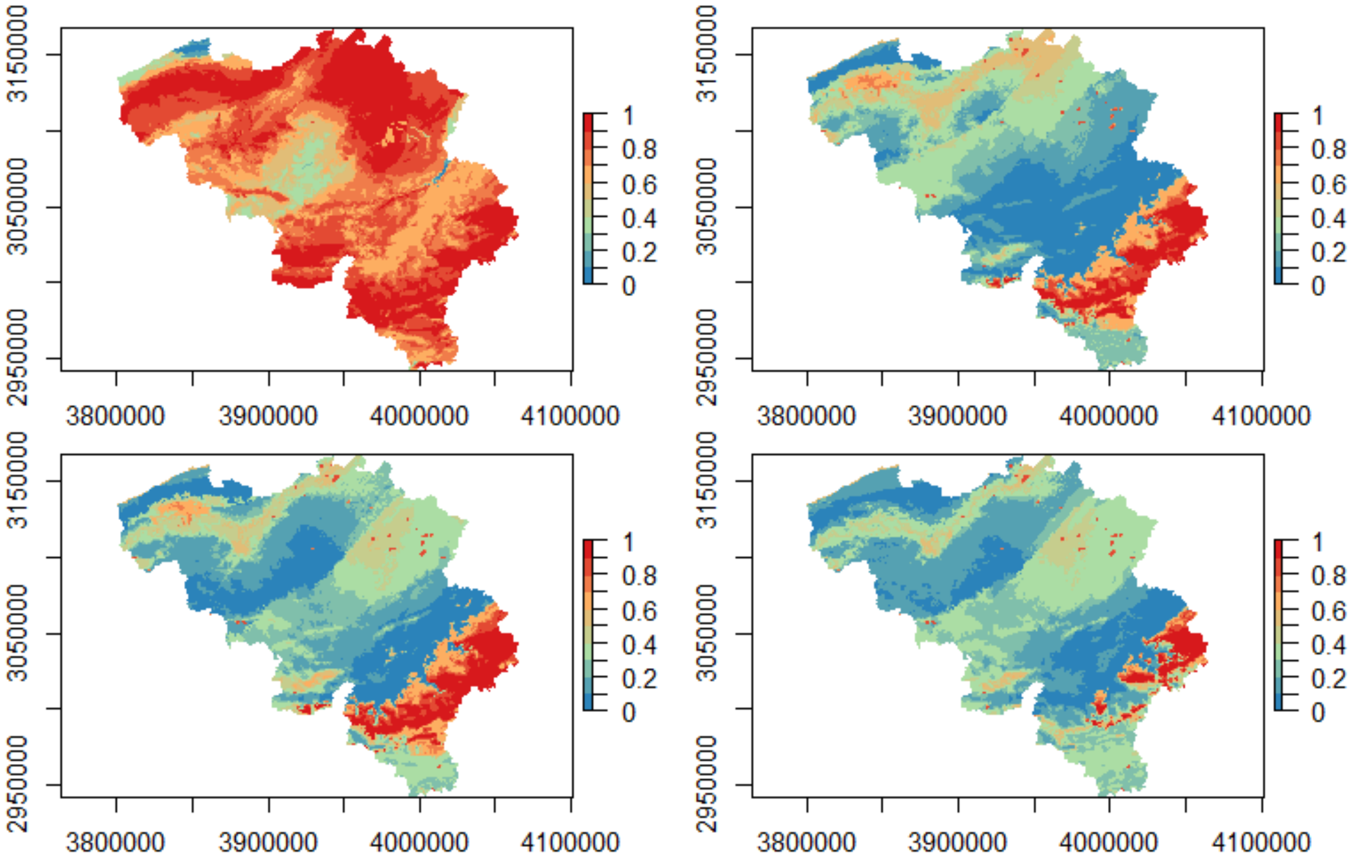
Hide

```
brks <- seq(0, 1, by=0.1)
  nb <- length(brks)-1
  pal <- colorRampPalette(rev(brewer.pal(4, 'Spectral')))
  cols<-pal(nb)


confidenceMaps<-function(x,taxonkey,taxonName,maptype){
pvals_dataframe<-get("x")
data.xyz <- pvals_dataframe[c("x","y","conf")]
rst <- rasterFromXYZ(data.xyz)
crs(rst)<-CRS("+proj=laea +lat_0=52 +lon_0=10 +x_0=4321000 +y_0=3210000 +ellps=GRS80 +units=m +no_defs")
plot(rst,breaks=brks, col=cols,lab.breaks=brks)
writeRaster(rst, filename=file.path(rasterOutput,paste("be_",taxonkey, "_",maptype,".tif",sep="")), format="GTiff",overwrite=TRUE)
exportPDF(rst,taxonkey,taxonName=taxonName,nameextension= paste(maptype,".pdf",sep=""))
return(rst)
}

par(mfrow=c(2,2), mar= c(2,3,0.8,0.8))
hist.conf.map<-confidenceMaps(pvalsdf_hist,taxonkey,taxonName,maptype="hist_conf")
rcp26.conf.map<-confidenceMaps(pvalsdf_rcp26,taxonkey,taxonName,maptype="rcp26_conf")
```

Hide

```
rcp45.conf.map<-confidenceMaps(pvalsdf_rcp45,taxonkey,taxonName,maptype="rcp45_conf")
rcp85.conf.map<-confidenceMaps(pvalsdf_rcp85,taxonkey,taxonName,maptype="rcp85_conf")
```

File failed to load: /extensions/MathZoom.js



Mask areas of below a set confidence level

Hide

```
# Cutoff for "high" confidence can be modified below. Cutoff should be a value between 0 and
1. Values that are less than the cutoff are shown in gray.
cutoff<-0.70

conf.brks <- seq(0,1, by=0.1)
  nb <- length(conf.brks)
  pal <- colorRampPalette(rev(brewer.pal(4, 'Spectral')))
  cols<-pal(nb)

par(mfrow=c(2,2), mar= c(2,3,0.9,0.8))
m1<-hist.conf.map < cutoff
hist_masked<-mask(ens_pred_hist,m1,maskvalue=TRUE)
plot(hist_masked,breaks=conf.brks, col=cols,lab.breaks=conf.brks)
plot(country,add=TRUE,border="dark gray")
```

Hide

```
m2<-rcp26.conf.map < cutoff
rcp26_masked<-mask(ens_pred_hab26,m2,maskvalue=TRUE)
plot(rcp26_masked,breaks=conf.brks, col=cols,lab.breaks=conf.brks)
plot(country,add=TRUE,border="dark gray")
```

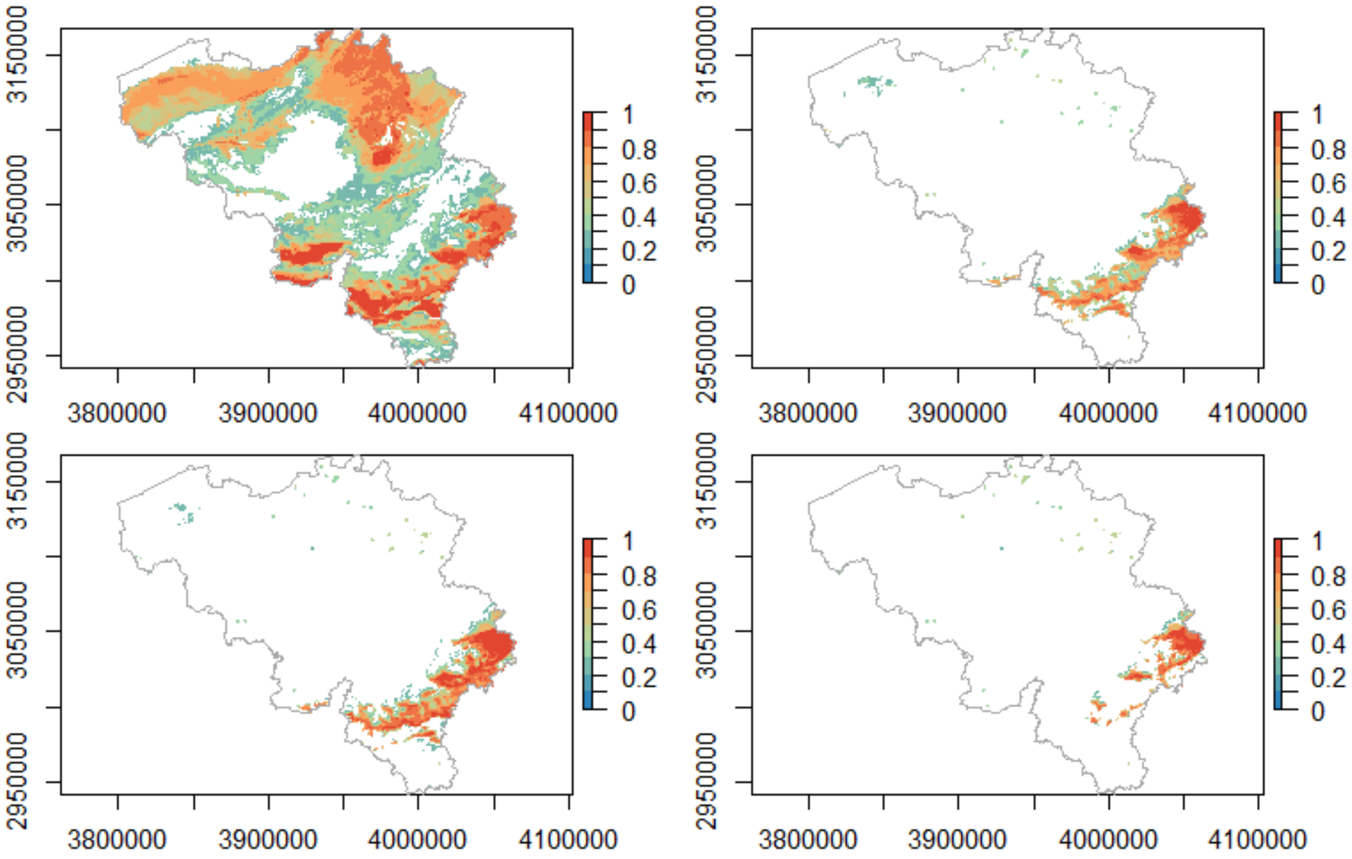
File failed to load: /extensions/MathZoom.js

Hide

```
m3<-rcp45.conf.map < cutoff
rcp45_masked<-mask(ens_pred_hab45,m3,maskvalue=TRUE)
plot(rcp45_masked,breaks=conf.brks, col=cols,lab.breaks=conf.brks)
plot(country,add=TRUE,border="dark gray")
```

Hide

```
m4<-rcp85.conf.map < cutoff
rcp85_masked<-mask(ens_pred_hab85,m4,maskvalue=TRUE)
plot(rcp85_masked,breaks=conf.brks, col=cols,lab.breaks=conf.brks)
plot(country,add=TRUE,border="dark gray")
```

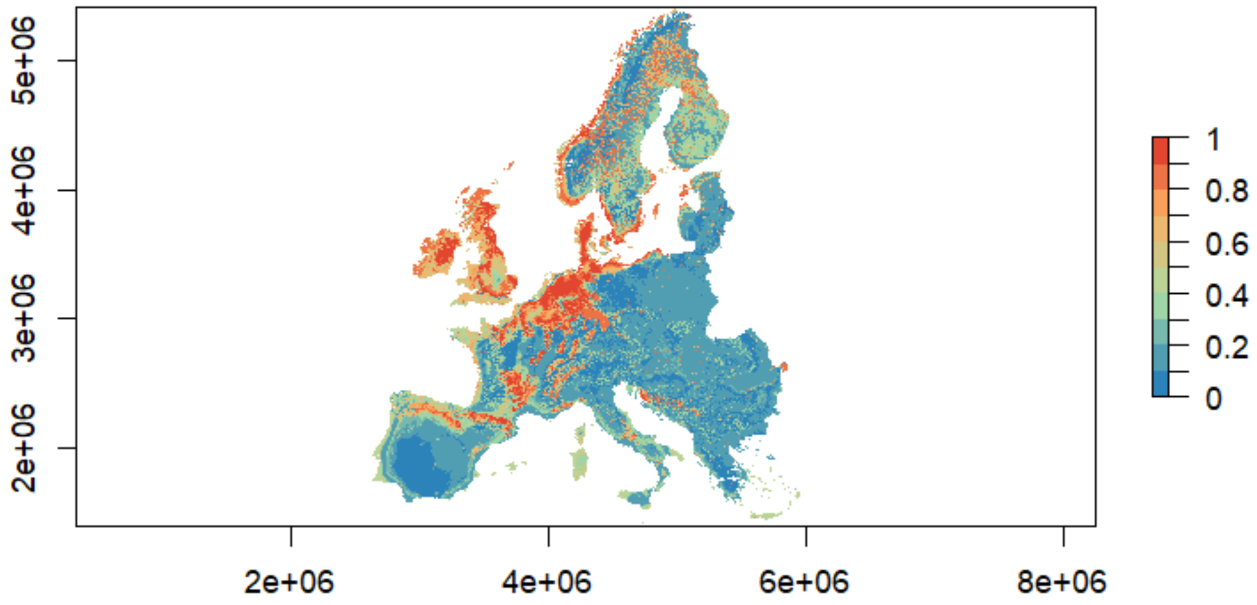


confidence map of best model at EU level

Hide

```
brks <- seq(0, 1, by=0.1)
  nb <- length(brks)-1
  pal <- colorRampPalette(rev(brewer.pal(4, 'Spectral')))
set.seed(792)
pvalsdf_hist_eu<-classConformalPrediction(bestModel,ens_pred_hab_eu1$X6)
hist.conf.map.eu<-confidenceMaps(pvalsdf_hist_eu,taxonkey,taxonName,maptype="hist_conf_eu")
```

File failed to load: /extensions/MathZoom.js



Get variable importance of best european model

Hide

```
variableImportance<-varImp(bestModel)
kable(variableImportance,digits=2,caption="Variable Importance") %>%
kable_styling(bootstrap_options = c("striped"))
```

Variable Importance

	overall	glm	gbm	rf	earth
corine_perWetland	0.73	8.57	0.00	0.00	0.29
annpvarrecip_eea	30.06	53.22	30.43	35.40	0.00
temprang	32.53	0.00	42.10	26.39	63.12
maxtemp	36.67	38.21	27.47	38.21	36.59

Hide

```
write.csv(variableImportance,file = paste0(genOutput,taxonkey,"_varImp_EU_model.csv"))
```

Generate and export response curves in order of variable

File failed to load: /extensions/MathZoom.js

importance

Hide

File failed to load: /extensions/MathZoom.js

```
topPreds <- variableImportance[with(variableImportance,order(-overall)),]
varNames<-rownames(topPreds)
## combine predictions from each model for each variable
## train data needs to be the training data used in the individual models used to build the ensemble model. This info can be extracted from the best ensemble model (ie. bestModel)
bestModel.train<-bestModel$models[[1]]$trainingData

partial_gbm<-function(x){
  m.gbm<-pdp::partial(bestModel$models$gbm$finalModel,pred.var=paste(x),train = bestModel.train,type="classification",
                      prob=TRUE,n.trees= bestModel$models$gbm$finalModel$n.trees, which.class = 1,grid.resolution=nrow(bestModel.train))
}



gbm.partial.list<-lapply(varNames,partial_gbm)

partial_glm<-function(x){
m.glm<-pdp::partial(bestModel$models$glm$finalModel,pred.var=paste(x),train = bestModel.train,type="classification",
              prob=TRUE,which.class = 1,grid.resolution=nrow(bestModel.train))
}

glm.partial.list<-lapply(varNames,partial_glm)

partial_rf<-function(x){
  pdp::partial(bestModel$models$rf$finalModel,pred.var=paste(x),train = bestModel.train,type="classification",
              prob=TRUE,which.class = 1,grid.resolution=nrow(bestModel.train))
}

rf.partial.list<-lapply(varNames,partial_rf)


partial_mars<-function(x){
m.mars<-pdp::partial(bestModel$models$earth$finalModel,pred.var=paste(x),train = bestModel.train,type="classification",
              prob=TRUE,which.class = 2,grid.resolution=nrow(bestModel.train)) # class=2 because in earth pkg, absense is the first class
}

mars.partial.list<-lapply(varNames,partial_mars)


names(glm.partial.list)<-varNames
names(gbm.partial.list)<-varNames
names(rf.partial.list)<-varNames
names(mars.partial.list)<-varNames

glm.partial.df<-as.data.frame(glm.partial.list)
```

```
gbm.partial.df<-as.data.frame(gbm.partial.list)
rf.partial.df<-as.data.frame(rf.partial.list)
mars.partial.df<-as.data.frame(mars.partial.list)

predx<-data.frame()
predy<-data.frame()

for (i in varNames){
  predx <- rbind(predx, as.data.frame(paste(i,i,sep=".")))
  predy<- rbind(predy,as.data.frame(paste(i,"yhat",sep=".")))
}
names(predx)<-""
names(predy)<-""

predx1<-t(predx)
predy1<-t(predy)


glm.partial.df$data<-'GLM'
gbm.partial.df$data<-'GBM'
rf.partial.df$data<-'RF'
mars.partial.df$data<-'MARS'

all_dfs<-rbind.data.frame(glm.partial.df,gbm.partial.df,rf.partial.df,mars.partial.df)


responseCurves<-function(x,y) {
  colors <- c("GLM" = "gray", "GBM"="red","RF"="blueviolet","MARS"= "hotpink")
  ggplot(all_dfs,(aes(x=.data[[x]],y=.data[[y]]))) +
    geom_line(aes(color = data), size =1.2, position=position_dodge(width=0.2))+
   theme_bw()+
    labs(y="Partial probability", x= gsub("//..*","",x),color="Legend") +
    scale_color_manual(values = colors)
}

allplots<-map2(predx1,predy1, ~responseCurves(.x,.y))
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
Please use `linewidth` instead.
```

Hide

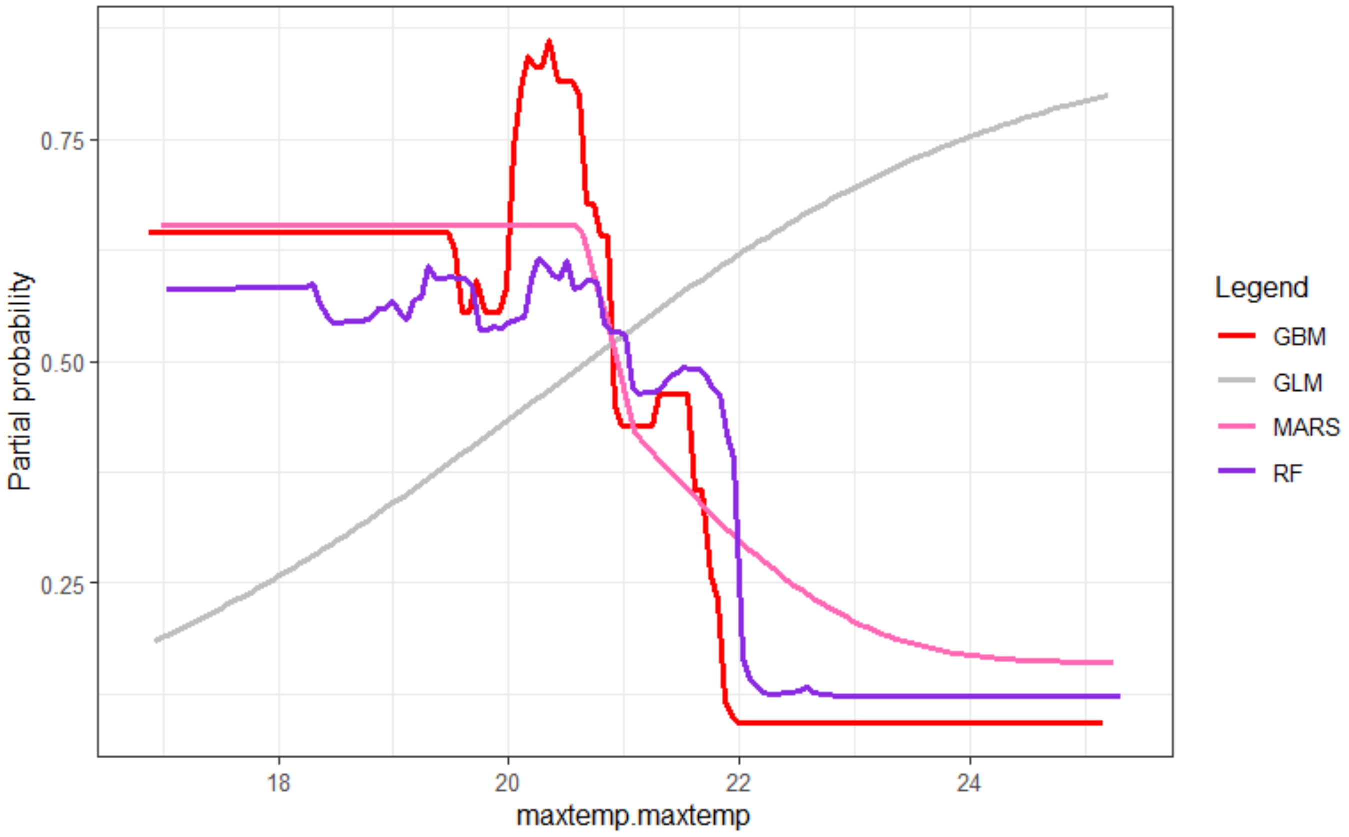
```
#export plots as PNGs
for(i in seq_along(allplots)){
  png(paste0(genOutput,taxonkey,"_",i,".png"),width = 5, height = 5, units = "in",res=300)
  print(allplots[[i]])
  dev.off()
}
```

Plot response curves

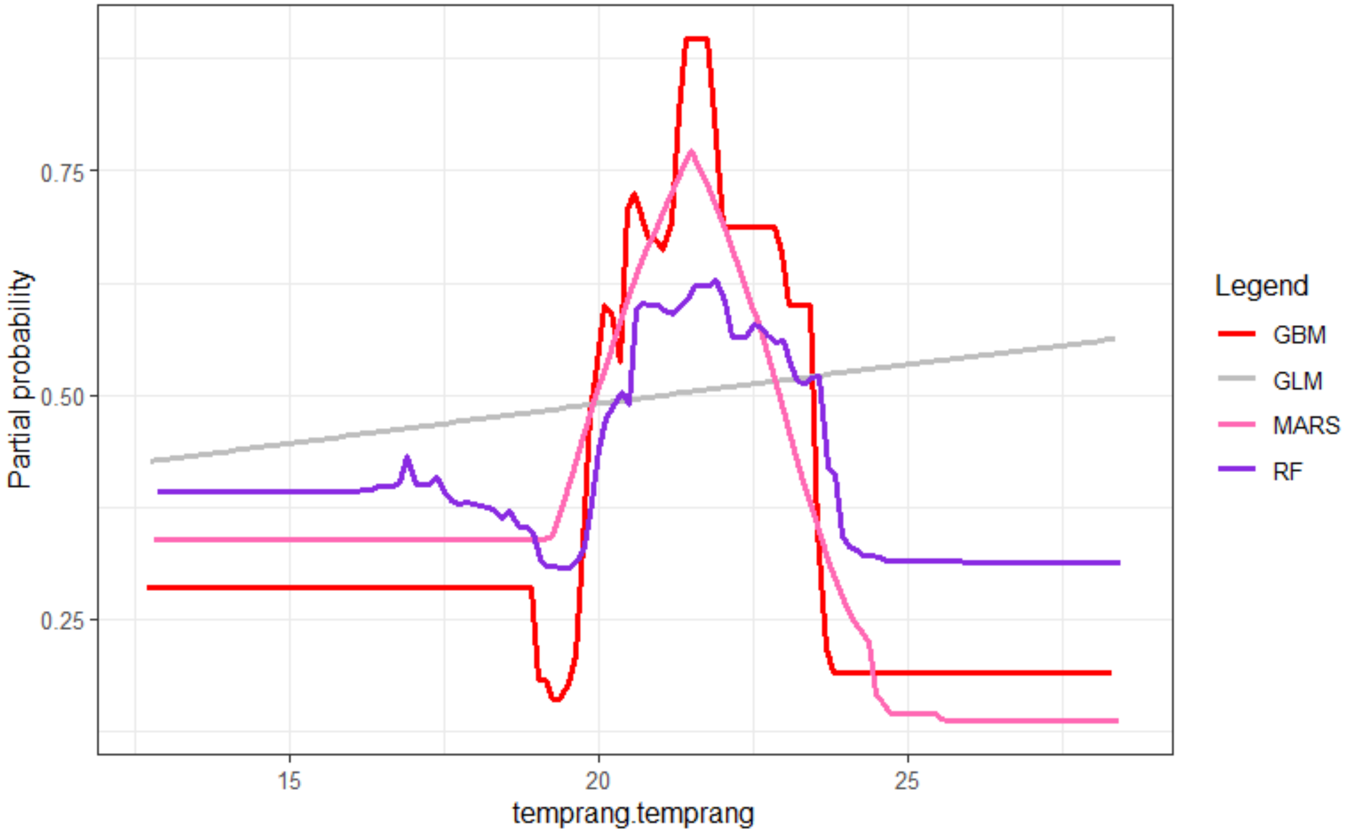
File failed to load: /extensions/MathZoom.js

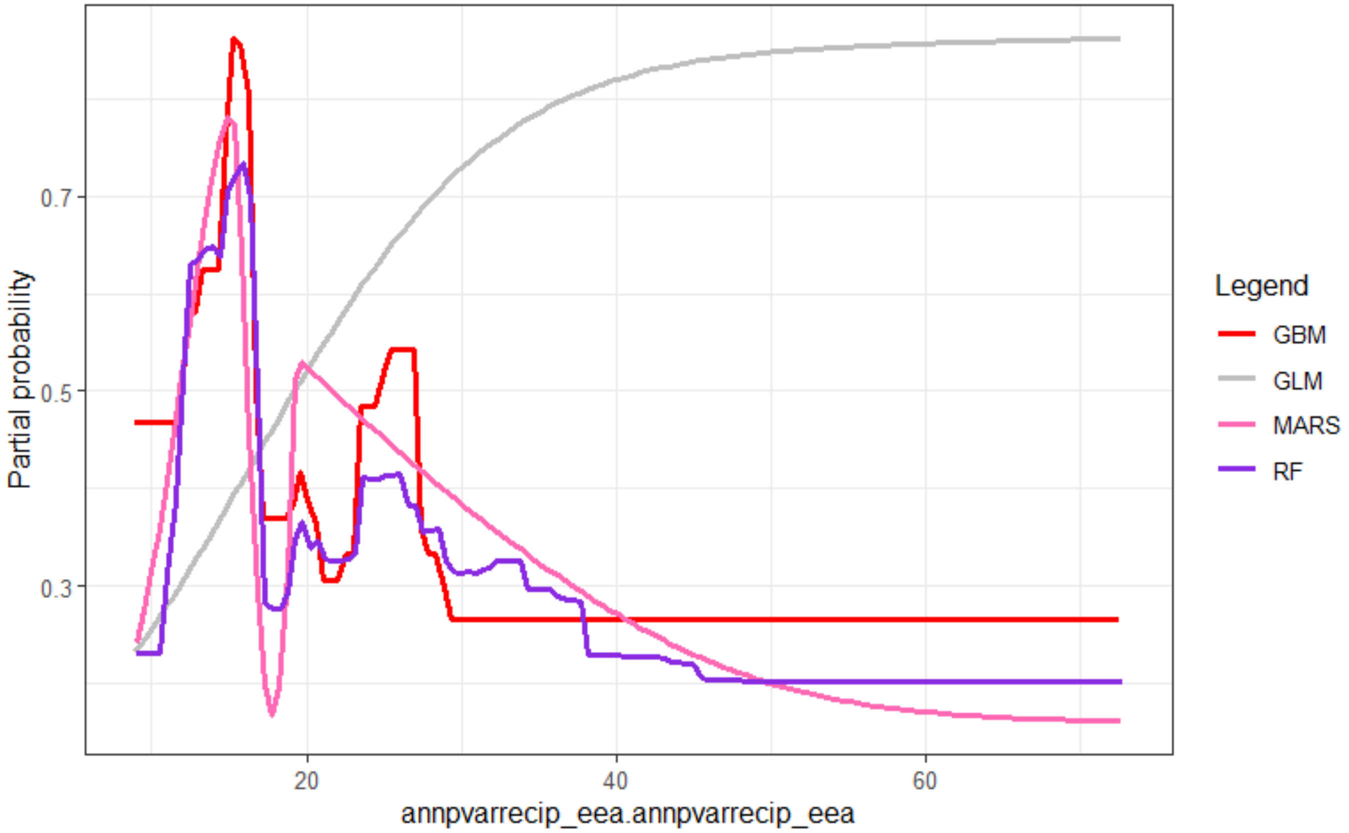
Hide

```
par(mfrow=c(3,4))
for(i in seq_along(allplots)){
  print(allplots[[i]])
}
```

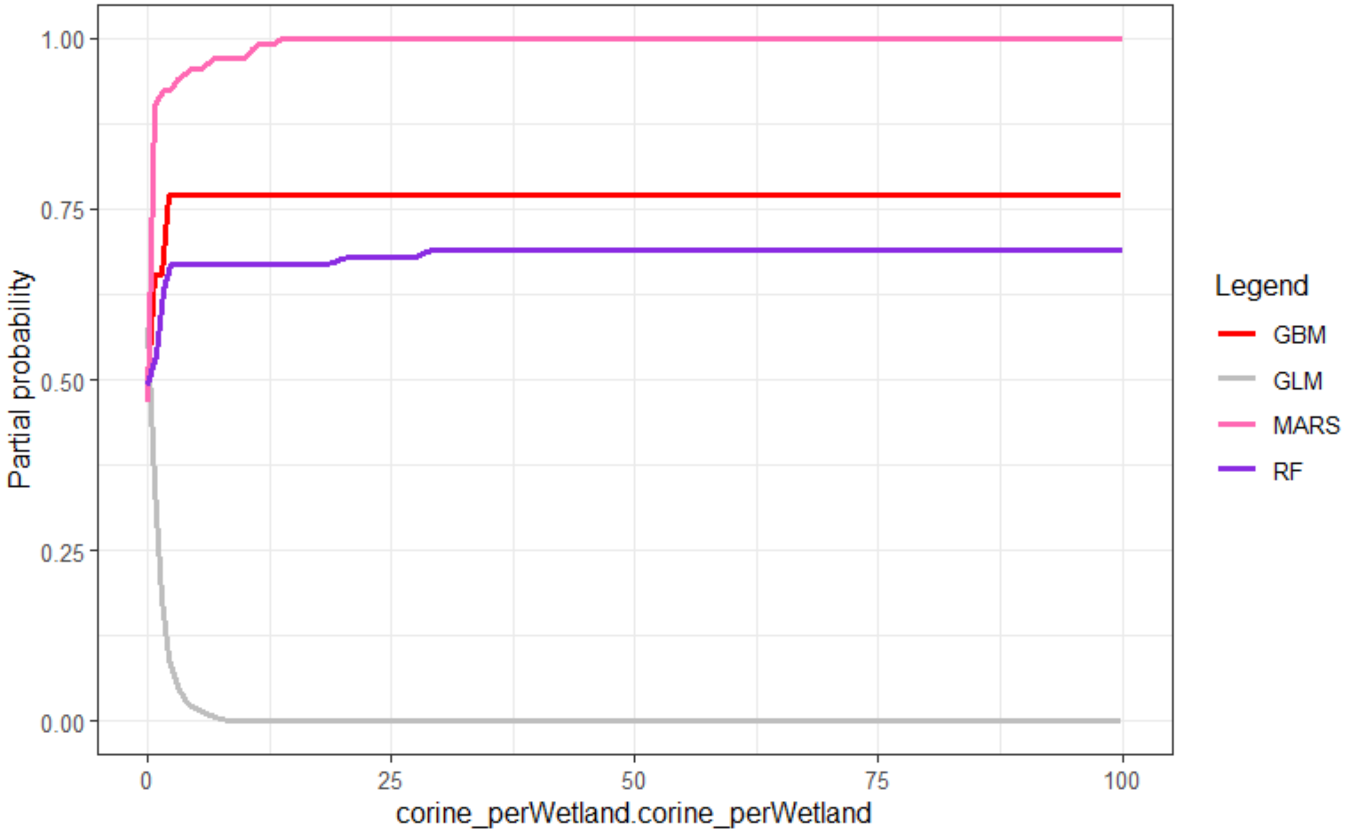


File failed to load: /extensions/MathZoom.js





File failed to load: /extensions/MathZoom.js



Evaluate the performance of each the EU level ensemble models using independent data set from the future

Hide

File failed to load: /extensions/MathZoom.js

```
# read in and prepare independent data
#2011-2021
eval.data<-read.csv("C:/Users/amyjs/Documents/projects/xps15/xps15/wiSDM/data/external/0001753-230828120925497/0001753-230828120925497.csv",header=TRUE,sep ="\t",quote="")

#enter value for max coordinate uncertainty in meters.

eval.data.occ<-eval.data %>%
  filter(is.na(coordinateUncertaintyInMeters)| coordinateUncertaintyInMeters< 1000)

eval.data.occ$lon_dplaces<-sapply(na.omit(eval.data.occ$decimalLongitude), function(x) decimalplaces(x))
eval.data.occ$lat_dplaces<-sapply(eval.data.occ$decimalLatitude, function(x) decimalplaces(x))
eval.data.occ[eval.data.occ$lon_dplaces < 4& eval.data.occ$lat_dplaces < 4 , ]<-NA
eval.data.occ<-eval.data.occ[ which(!is.na(eval.data.occ$lon_dplaces)),]
eval.data.occ<-within(eval.data.occ,rm("lon_dplaces","lat_dplaces"))

eval.data.occ<-eval.data.occ[c("decimalLongitude", "decimalLatitude")]
coordinates(eval.data.occ)<- c("decimalLongitude", "decimalLatitude")
proj4string(eval.data.occ)<-CRS("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0")#specify here the existing coord.sys of the data
eval.data.occ.proj<-spTransform(eval.data.occ,rmiproj)
```

```
$X1


          present
  absent      138
  present     131

$X2


          present
  absent      137
  present     132

$X3


          present
  absent      145
  present     124

$X4


          present
  absent      137
  present     132

$X5


          present
  absent      188
  present      81

$X6


          present
  absent      118
  present     151

$X7


          present
  absent      173
  present      96

$X8


          present
  absent      172
  present      97

$X9


          present
```

File failed to load: /extensions/MathZoom.js

```
  absent      149
  present     120

$X10

          present
  absent      138
  present     131
```

File failed to load: /extensions/MathZoom.js

```
$X1

         present
  absent       13
  present      98

$X2

         present
  absent       13
  present      98

$X3

         present
  absent       43
  present      68

$X4

         present
  absent       16
  present      95

$X5

         present
  absent       62
  present      49

$X6

         present
  absent       10
  present     101

$X7

         present
  absent       49
  present      62

$X8

         present
  absent       51
  present      60

$X9

         present
```

File failed to load: /extensions/MathZoom.js

```
  absent       25
  present      86

$X10

         present
  absent       11
  present     100
```

File failed to load: /extensions/MathZoom.js