Model run (rmd file #2)

Debora

2024-05-06

Download environmental predictors

```
options(timeout=600)
# Lake water temperature raster based on satellite measurements (Armitage, 2023;
# https://onlinelibrary.wiley.com/doi/full/10.1111/ecog.06595#bib-0003)

url <- "https://datadryad.org/stash/downloads/file_stream/1895801"
temp_file <- tempfile()
temp_unzipped1 <- tempfile()
temp_unzipped2 <- tempfile()
download.file(url, destfile = temp_file, mode="wb")
unzip(temp_file, exdir = temp_unzipped1)
unzip(paste0(temp_unzipped1,"/LakeTemps_Code/rasters.zip"), exdir = temp_unzipped2)
lake <- raster::brick(paste0(temp_unzipped2,"/rasters/bioclim_lakes_10km.tif"))

air <- geodata::worldclim_global(var = 'bio', res = 5, download = T, path = 'data')
air <- as(air, "Raster")
air <- brick(air)
```

Future air temperature model (ACCESS)

```
# Future climate scenario from 'ACCESS-ESM1-5' climate model
clim_fut_access <- geodata::cmip6_world(model='ACCESS-ESM1-5', ssp='245', time='2041-2060', var='bioc',
clim_fut_access <- as(clim_fut_access, "Raster")
```

Future air temperature model (MIROC)

```
# Future climate scenario from 'MIROC-ES2L' climate model
clim_fut_miroc <- geodata::cmip6_world(model='MIROC-ES2L', ssp='245', time='2041-2060', var='bioc', dow
clim_fut_miroc <- as(clim_fut_miroc, "Raster")
```

Select bioclimatic variables

```
# Showing code for a single environmental predictor source
# (the same selection is done for all environmental rasters)
lake_crop1 <- lake_crop[[c(1,7,15)]]
air_crop11 <- air_crop[[c(10,11,15)]]

air_crop1 <- air_crop[[c(1,7,15)]]
air_crop11 <- air_crop[[c(10,11,15)]]
```

Load sampled datasets obtained in previous steps of the analysis

Training dataset (global records except Europe)

```
##          lon       lat
## 1 -87.72003  38.72377
## 2 -88.62159  38.99591
## 3 -87.76000  38.70583
## 4 -53.24036   3.91451
## 5  55.53241 -21.13317
## 6 -90.54842  35.46343
```

Testing dataset (European records)

```
##        lon      lat
## 1 0.62002 46.49774
## 2 0.68900 46.40908
## 3 0.99256 49.43957
## 4 0.97333 49.46973
## 5 1.01420 49.43349
## 6 1.01214 49.45276
```

Create study area, sample background points, and run models

```
raster_list <- list(lake_crop1, air_crop1, lake_crop11, air_crop11)
crop_raster_list <- list()
backg_train_list <- list()
maxent_output_list <- list()

count = 1
for (i in raster_list){
# Crop environment to correspond to species distribution range
model.extent<-extent(min(pres_train$lon)-10,max(pres_train$lon)+10,
                     min(pres_train$lat)-10,max(pres_train$lat)+10)

crop_raster <- crop(i,model.extent)

crop_raster_list[[count]] <- crop_raster

occ_buff <- buffer(pres_train, 600000, dissolve=TRUE) # width parameter = 600000m (600km)

# crop study area to buffer extent
studyArea <- crop(crop_raster, extent(occ_buff))

# mask the non buffer areas
studyArea <- mask(studyArea, occ_buff)
# output will still be a raster stack, just of the study area

# Randomly sample points
# Sample same number as our observed points inside the buffer
# to create pseudo-absences, or hypothetical areas where species
# could either be found or not

set.seed(2022)
backg_train = randomPoints(studyArea, n=length(pres_train$lon), p=pres_train, extf=1)
colnames(backg_train) <- c("lon","lat")
```

```r
backg_train_list[[count]] <- backg_train

# Run candidate models
e.mx.l <- ENMevaluate(occs = pres_train, envs = crop_raster, bg = backg_train,
                                  algorithm = 'maxent.jar', partitions = 'block',
                                  tune.args = list(fc = c("L","LQ"),
                                  rm = seq(0.5, 4, by = 0.5)))

maxent_output_list[[count]] <- e.mx.l


count=count+1
}
```

```
## Package ecospat is not installed, so Continuous Boyce Index (CBI) cannot be calculated.
## *** Running initial checks... ***
## * Removed 4 occurrence localities that shared the same grid cell.
## * Clamping predictor variable rasters...
## * Model evaluations with spatial block (4-fold) cross validation and lat_lon orientation...
##
## *** Running ENMeval v2.0.4 with maxent.jar v3.4.3 from dismo package v1.3.14 ***
##   |                                                                              |
## ENMevaluate completed in 4 minutes 40.4 seconds.
## Package ecospat is not installed, so Continuous Boyce Index (CBI) cannot be calculated.
## *** Running initial checks... ***
## * Removed 3 occurrence localities that shared the same grid cell.
## * Clamping predictor variable rasters...
## * Model evaluations with spatial block (4-fold) cross validation and lat_lon orientation...
##
## *** Running ENMeval v2.0.4 with maxent.jar v3.4.3 from dismo package v1.3.14 ***
##   |                                                                              |
## ENMevaluate completed in 4 minutes 25.6 seconds.
## Package ecospat is not installed, so Continuous Boyce Index (CBI) cannot be calculated.
## *** Running initial checks... ***
## * Removed 4 occurrence localities that shared the same grid cell.
## * Clamping predictor variable rasters...
## * Model evaluations with spatial block (4-fold) cross validation and lat_lon orientation...
##
## *** Running ENMeval v2.0.4 with maxent.jar v3.4.3 from dismo package v1.3.14 ***
##   |                                                                              |
## ENMevaluate completed in 4 minutes 10.6 seconds.
```

```
## Package ecospat is not installed, so Continuous Boyce Index (CBI) cannot be calculated.
## *** Running initial checks... ***
## * Removed 3 occurrence localities that shared the same grid cell.
## * Clamping predictor variable rasters...
## * Model evaluations with spatial block (4-fold) cross validation and lat_lon orientation...
##
## *** Running ENMeval v2.0.4 with maxent.jar v3.4.3 from dismo package v1.3.14 ***
##   |                                                                               |
## ENMevaluate completed in 4 minutes 24 seconds.
```

Calculate model performance metrics

Select models with lowest AICc

Lake model (Bio 1,7,15)

Features	Regularization.multiplier	Training.AUC	Mean.testing.AUC	AICc	Delta AICc	Parameters
L	0.5	0.73	0.70	18419.14	616.72	3
LQ	0.5	0.81	0.75	17802.42	0.00	6
L	1	0.73	0.70	18424.04	621.62	3
LQ	1	0.81	0.76	17860.95	58.53	6
L	1.5	0.73	0.70	18429.06	626.65	3
LQ	1.5	0.80	0.76	17953.27	150.86	6
L	2	0.73	0.70	18434.19	631.78	3
LQ	2	0.80	0.76	18037.61	235.19	5
L	2.5	0.73	0.70	18439.47	637.05	3
LQ	2.5	0.79	0.75	18122.97	320.55	5
L	3	0.73	0.70	18440.85	638.43	2
LQ	3	0.78	0.75	18203.27	400.86	5
L	3.5	0.73	0.70	18443.40	640.98	2
LQ	3.5	0.77	0.74	18288.97	486.56	5
L	4	0.73	0.71	18445.99	643.57	2
LQ	4	0.75	0.74	18387.38	584.96	5

Air temperature model (Bio 1,7,15)

Features	Regularization.multiplier	Training.AUC	Mean.testing.AUC	AICc	Delta AICc	Parameters
L	0.5	0.74	0.73	18331.20	755.59	3
LQ	0.5	0.83	0.77	17575.61	0.00	6
L	1	0.74	0.73	18336.95	761.34	3
LQ	1	0.82	0.77	17642.03	66.42	6
L	1.5	0.74	0.74	18342.87	767.26	3
LQ	1.5	0.82	0.77	17724.33	148.72	6
L	2	0.74	0.74	18348.97	773.36	3
LQ	2	0.82	0.77	17789.08	213.47	5
L	2.5	0.74	0.74	18352.18	776.57	2
LQ	2.5	0.81	0.78	17872.84	297.23	5
L	3	0.74	0.74	18355.72	780.11	2
LQ	3	0.80	0.77	17949.27	373.66	5

Features	Regularization.multiplier	Training.AUC	Mean.testing.AUC	AICc	Delta AICc	Parameters
L	3.5	0.74	0.74	18359.31	783.70	2
LQ	3.5	0.79	0.77	18035.28	459.67	5
L	4	0.74	0.74	18362.96	787.35	2
LQ	4	0.78	0.77	18118.11	542.50	5

Lake model (Bio 10,11,15)

Features	Regularization.multiplier	Training.AUC	Mean.testing.AUC	AICc	Delta AICc	Parameters
L	0.5	0.73	0.70	18429.15	680.64	3
LQ	0.5	0.80	0.75	17748.51	0.00	5
L	1	0.73	0.70	18430.34	681.83	3
LQ	1	0.80	0.76	17774.54	26.03	5
L	1.5	0.73	0.70	18431.64	683.12	3
LQ	1.5	0.80	0.76	17801.64	53.13	5
L	2	0.73	0.70	18433.05	684.54	3
LQ	2	0.80	0.77	17832.54	84.02	4
L	2.5	0.73	0.70	18434.57	686.06	3
LQ	2.5	0.80	0.77	17867.50	118.99	4
L	3	0.73	0.70	18436.20	687.69	3
LQ	3	0.80	0.76	17903.00	154.48	4
L	3.5	0.73	0.70	18437.95	689.44	3
LQ	3.5	0.80	0.76	17937.78	189.27	4
L	4	0.72	0.70	18439.80	691.29	3
LQ	4	0.80	0.76	17973.28	224.77	4

Air temperature model (Bio 10,11,15)

Features	Regularization.multiplier	Training.AUC	Mean.testing.AUC	AICc	Delta AICc	Parameters
L	0.5	0.74	0.72	18331.75	758.73	3
LQ	0.5	0.81	0.75	17573.02	0.00	6
L	1	0.74	0.72	18335.51	762.49	3
LQ	1	0.81	0.76	17626.14	53.12	5
L	1.5	0.74	0.72	18339.44	766.42	3
LQ	1.5	0.81	0.76	17644.19	71.16	5
L	2	0.74	0.72	18343.55	770.53	3
LQ	2	0.81	0.76	17666.33	93.30	4
L	2.5	0.74	0.72	18347.83	774.81	3
LQ	2.5	0.81	0.76	17693.73	120.70	4
L	3	0.74	0.72	18352.29	779.26	3
LQ	3	0.81	0.77	17721.36	148.33	4
L	3.5	0.74	0.72	18356.92	783.89	3
LQ	3.5	0.81	0.77	17749.20	176.18	4
L	4	0.74	0.72	18361.72	788.69	3
LQ	4	0.81	0.77	17777.25	204.23	4

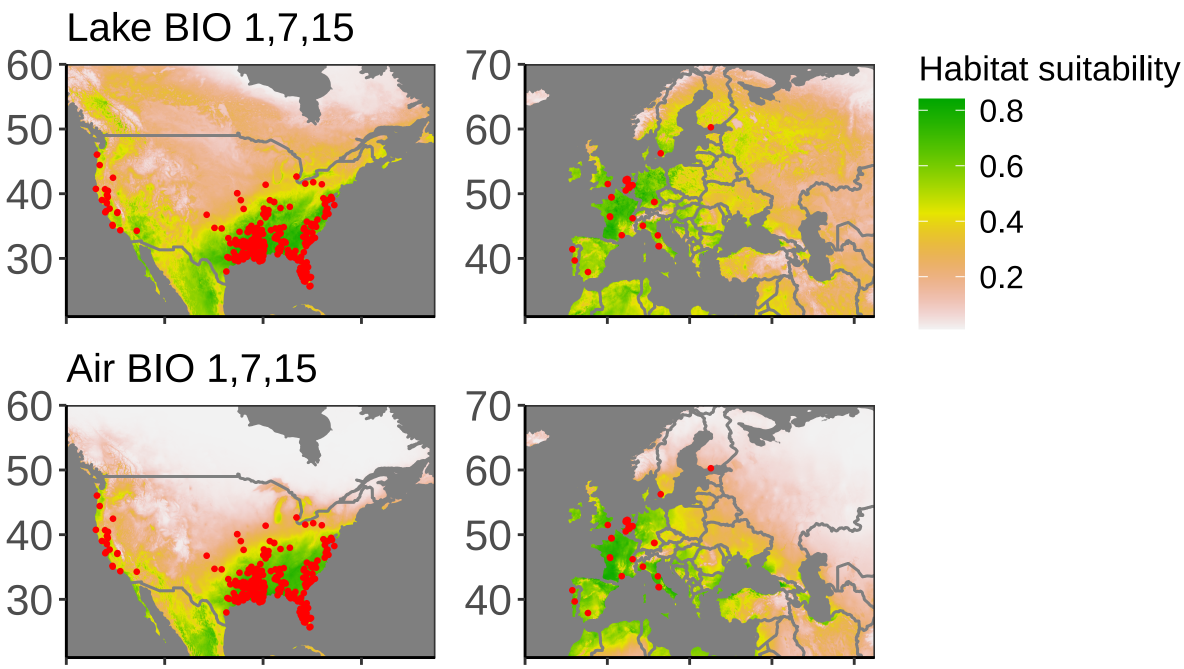
Obtain background points for the testing region (Europe)

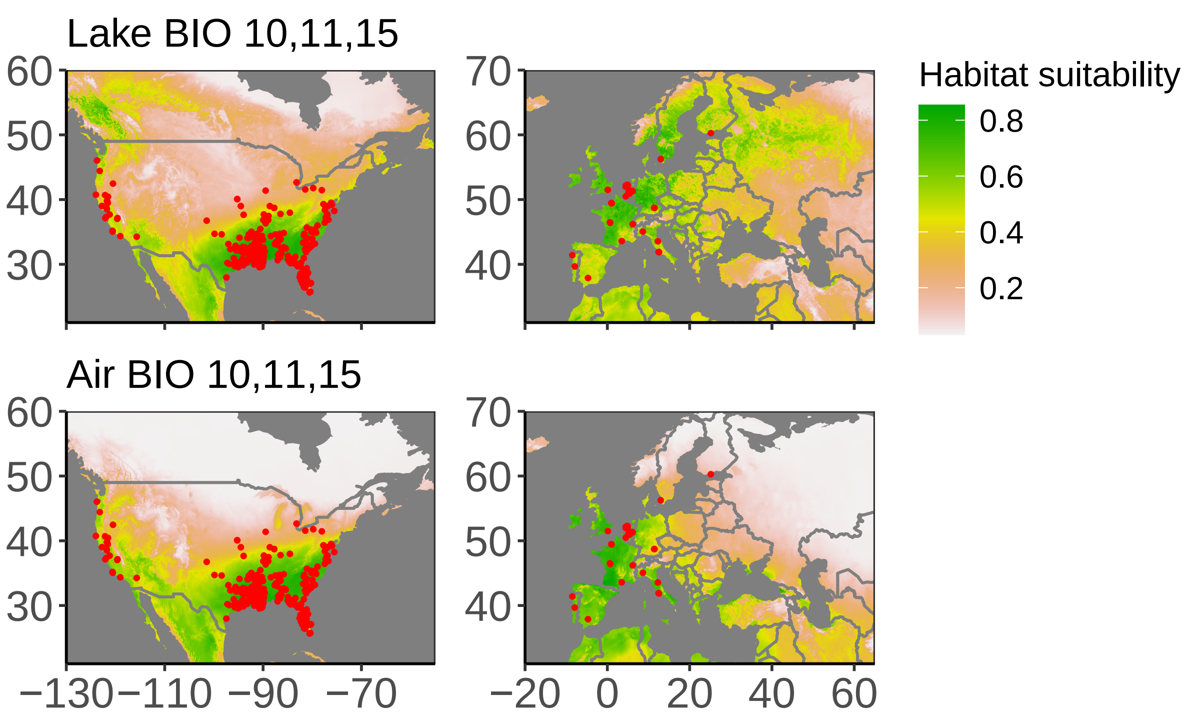
Here we followed the same process as shown for the training region

Evaluate models

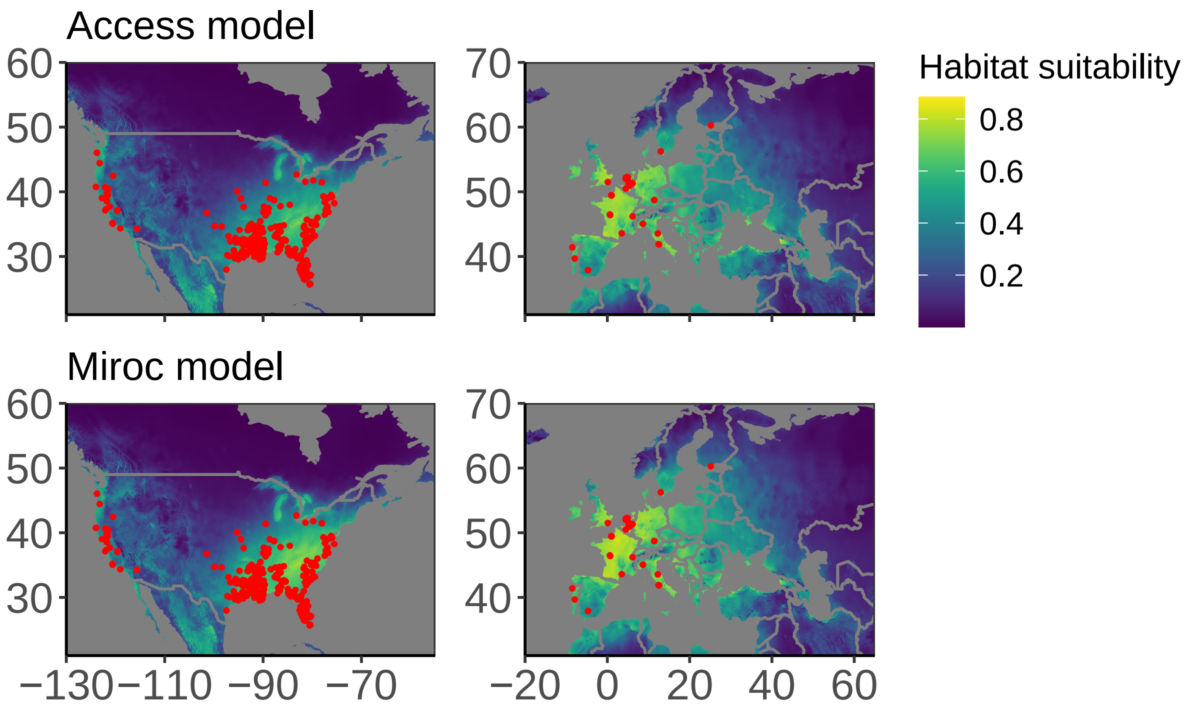
Model	AUC_train	AUC_test	TSS_train	TSS_test	Thresh_train	Thresh_test
Lake_BIO1,7,15	0.81	0.86	0.49	0.74	0.54	0.59
Air_BIO1,7,15	0.83	0.78	0.51	0.54	0.56	0.60
Lake_BIO10,11,15	0.81	0.77	0.49	0.49	0.59	0.54
Air_BIO10,11,15	0.82	0.68	0.50	0.35	0.58	0.61

Plot model predictions

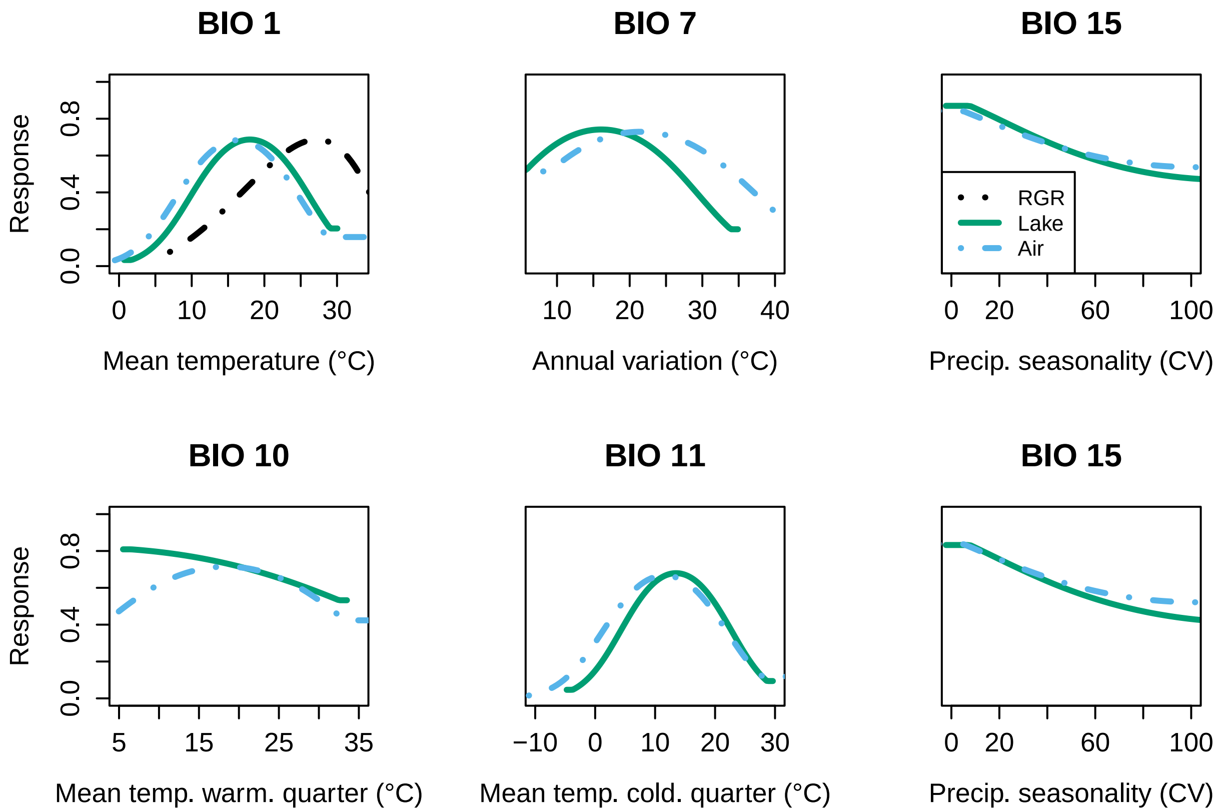




Making future predictions based on best air temperature model



Compare TPC with response curves for BIO1



Get minimum, maximum, and optimum values for each response curve

Predictor	Minimum	Optimum	Maximum
Lake_BIO1	0.69	17.89	30.05
Lake_BIO7	4.74	16.01	34.90
Lake_BIO15	-2.28	-2.28	144.72
Air_BIO1	-12.99	16.24	38.68
Air_BIO7	-1.03	21.21	56.92
Air_BIO15	-5.20	-5.20	143.96
Lake_BIO10	5.50	5.50	33.51
Lake_BIO11	-4.75	13.66	29.64
Lake_BIO15_2	-2.28	-2.28	144.72
Air_BIO10	-6.08	18.35	43.28
Air_BIO11	-22.12	11.76	37.78
Air_BIO15_2	-5.20	-5.20	143.96

Get variable importance for best models

Table 7: Lake model (BIO 1,7,15)

	variable	percent.contribution	permutation.importance
2	bioclim_lakes_10km_15	53.60	40.31
3	bioclim_lakes_10km_7	27.97	26.74
1	bioclim_lakes_10km_1	18.43	32.95

Table 8: Air model (BIO 1,7,15)

	variable	percent.contribution	permutation.importance
2	wc2.1_5m_bio_15	45.92	25.63
3	wc2.1_5m_bio_7	29.16	27.70
1	wc2.1_5m_bio_1	24.92	46.67

Table 9: Lake model (BIO 10,11,15)

	variable	percent.contribution	permutation.importance
2	bioclim_lakes_10km_11	49.29	50.06
3	bioclim_lakes_10km_15	49.04	42.77
1	bioclim_lakes_10km_10	1.67	7.16

Table 10: Air model (BIO 1,11,15)

	variable	percent.contribution	permutation.importance
2	wc2.1_5m_bio_11	50.02	60.43
3	wc2.1_5m_bio_15	46.60	29.68
1	wc2.1_5m_bio_10	3.38	9.89