Analysis of cross-validation tests

Gerardo Martin

11 de octubre de 2019

This is the cross validation analysis using the Partial ROC tests for the snake models. Data was partitioned spatially following a chessboard pattern with squares of 50km. For reference I show the tests performed to the models fitted with the full unpartitioned datasets. Each test was run at increasing omission thresholds to find if there is an optimum omission rate that maximises performance according to the Partial ROC tests. Each test was run with a 50% bagging and 1000 sampling iterations.

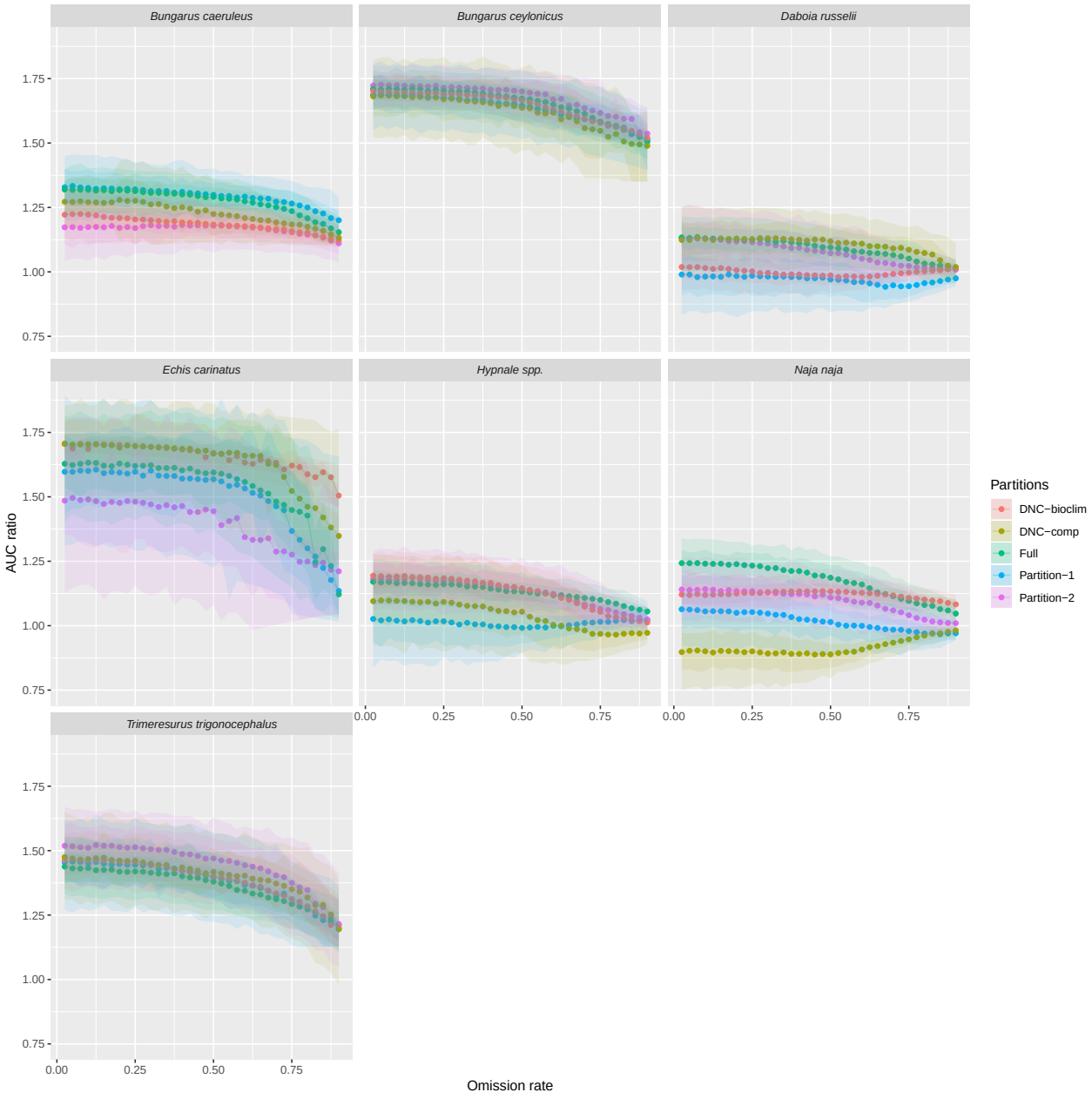
Extracting the AUC ratios

Here I extract the median of AUC ratios by species and omission rate. Then I assess statistical significance of the AUC ratios by calculating the proportion of ratios in each test run that are smaller than 1. For visualisation of variability I compute the confidence intervals with the `HPDinterval` function of the `coda` package at two levels, `0.95` and `0.68`.

Plots of the results

Here I show how the performance of the model changed with the omission rate to assess visually the optimal cutoff value (omission rate in the x axis and ratio (performance) in the y axis). The colour of the dots indicates the proportion of AUC ratios that were < 1, which is interpreted as statistical significance (AUC ratio significantly different from zero).

```
ggplot(ratios) + geom_point(aes(x = omis, y = AUC, colour = as.factor(model))) +
      geom_line(aes(x = omis, y = AUC, colour = as.factor(model)), alpha = 0.3) +
      geom_ribbon(aes(x = omis, ymin = lower, ymax = upper, fill = as.factor(model)), alpha = 0.1) +
      geom_ribbon(aes(x = omis, ymin = lower.1, ymax = upper.1, fill = as.factor(model)), alpha = 0.1) +
      facet_wrap(facets = "sp") +
      labs(colour = "Partitions", fill = "Partitions", x = "Omission rate", y = "AUC ratio") +
      theme(strip.text.x = element_text(face = "italic"))
```



These plots display the AUC ratio estimates as a function of the omission rate. In all cases AUC ratios decrease in an accelerating fashion to increasing values of the omission rate. The shaded areas show the 95 and 68% confidence intervals at each of the points where AUC ratios were estimated. The colour scale of the points indicates whether AUC ratios were significantly greater than 1, that is, the probability that the test estimated an AUC ratio smaller than 1. In all cases `p = 0`, which means that all of the estimated ratios were above the random prediction threshold.

Finally, looks like in all cases, the 0.05 omission threshold is optimal. To wrap it up, it makes sense to use as threshold the 5th percentile of suitability estimates for presence locations.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
d.perf <- ratios %>% group_by(sp,
                              model) %>% summarise(
                                    AUC = max(AUC))

p <- c()
om <-c()

for(i in 1:nrow(d.perf)){
      p[i] <- ratios$p[with(ratios,
                  which(AUC == d.perf$AUC[i] &
                        sp == d.perf$sp[i] &
                        model == d.perf$model[i]))]
      om[i] <- ratios$omis[with(ratios,
                  which(AUC == d.perf$AUC[i] &
                        sp == d.perf$sp[i] &
                        model == d.perf$model[i]))]
}

d.perf$p <- p
d.perf$omis <- om

write.csv(d.perf, "Omission-rates-cross-valid.csv")

print(as.data.frame(d.perf))
```

```
##                             sp       model       AUC           p  omis
## 1            Bungarus caeruleus DNC-bioclim 1.2246393 0.000000000 0.075
## 2            Bungarus caeruleus    DNC-comp 1.2791616 0.000000000 0.200
## 3            Bungarus caeruleus        Full 1.3194699 0.000000000 0.025
## 4            Bungarus caeruleus Partition-1 1.3335217 0.000000000 0.050
## 5            Bungarus caeruleus Partition-2 1.1817060 0.000000000 0.400
## 6           Bungarus ceylonicus DNC-bioclim 1.7013699 0.000000000 0.025
## 7           Bungarus ceylonicus    DNC-comp 1.6846463 0.000000000 0.050
## 8           Bungarus ceylonicus        Full 1.7110134 0.000000000 0.150
## 9           Bungarus ceylonicus Partition-1 1.6885820 0.000000000 0.150
## 10          Bungarus ceylonicus Partition-2 1.7263337 0.000000000 0.050
## 11              Daboia russelii DNC-bioclim 1.0189013 0.340659341 0.025
## 12              Daboia russelii    DNC-comp 1.1315932 0.020979021 0.075
## 13              Daboia russelii        Full 1.1352019 0.000999001 0.075
## 14              Daboia russelii Partition-1 0.9896070 0.549450549 0.050
## 15              Daboia russelii Partition-2 1.1316537 0.020979021 0.025
## 16              Echis carinatus DNC-bioclim 1.7068164 0.000000000 0.025
## 17              Echis carinatus    DNC-comp 1.7052724 0.000000000 0.025
## 18              Echis carinatus        Full 1.6326975 0.000000000 0.100
## 19              Echis carinatus Partition-1 1.6055801 0.000000000 0.125
## 20              Echis carinatus Partition-2 1.4958956 0.007992008 0.050
## 21                 Hypnale spp. DNC-bioclim 1.1940645 0.000000000 0.025
## 22                 Hypnale spp.    DNC-comp 1.0978283 0.147852148 0.050
```

```
## 23                 Hypnale spp.        Full 1.1701345 0.000000000 0.025
## 24                 Hypnale spp. Partition-1 1.0261218 0.396603397 0.025
## 25                 Hypnale spp. Partition-2 1.1858520 0.003996004 0.150
## 26                    Naja naja DNC-bioclim 1.1342285 0.000000000 0.475
## 27                    Naja naja    DNC-comp 0.9814785 0.853146853 0.900
## 28                    Naja naja        Full 1.2428388 0.000000000 0.075
## 29                    Naja naja Partition-1 1.0637352 0.181818182 0.025
## 30                    Naja naja Partition-2 1.1415491 0.011988012 0.100
## 31 Trimeresurus trigonocephalus DNC-bioclim 1.4650114 0.000000000 0.050
## 32 Trimeresurus trigonocephalus    DNC-comp 1.4751883 0.000000000 0.025
## 33 Trimeresurus trigonocephalus        Full 1.4379265 0.000000000 0.025
## 34 Trimeresurus trigonocephalus Partition-1 1.4575961 0.000000000 0.050
## 35 Trimeresurus trigonocephalus Partition-2 1.5215754 0.000000000 0.125
```