P2

Extent of terrestrial alien species

Hanno Sandvik

15 March 2023

Contents

Introduction

This code calculates indicator P2 of NorInAliS.

Preparations

Load data from the Alien Species List 2018 (Sandvik et al. 2020):

```
fab  <- read.csv2(url("https://datadryad.org/stash/downloads/file_stream/359484"),
                  as.is=T)
```

The AOOs (areas of occupancy) provided in the above dataset are only the *best* estimates of the *total* AOO for each species (i.e. the *median* expert judgement of the *real* AOO, including "dark figures" or unreported occurrences). However, in order to quantify uncertainty, P2 needs more than that, viz. the *low* and *high* estimates of the total AOO (i.e. lower and upper *quartiles*) as well as the *known* AOO (i.e. excluding dark figures). These values are here read from a separate file. Their source is https://artsdatabanken.no/fremmedartslista2018.

```
aoo <- read.csv2("cache/aoo.txt", as.is=T)
head(aoo)
```

```
##                 Name known  low best high
## 1     Abies amabilis     4   20   40   60
## 2     Abies balsamea    64  256  512  768
## 3 Abies cephalonica    12   60  120  180
## 4     Abies concolor   100  500 1000 1500
## 5      Abies grandis    24  120  240  360
## 6      Abies koreana    12   24   60   96
```

Restrict data to alien species that are reproducing unaidedly in mainland Norway:

```
fab <- fab[which(fab$Status == "reproducing" & fab$Mainl),]
```

Restrict the data to terrestrial species:

```r
w <- which(fab$LifeSt %in%
             c("lim", "lim,mar", "lim,par", "lim,mar,par", "mar", "mar,par"))
fab <- fab[-w,]
aoo <- aoo[-w,]
```

Make sure that the two data frames are compatible:

```r
if (all(fab$Name == aoo$Name)) {
  cat("Everything is fine.\n")
} else {
  cat("ERROR: For some reason, the two dataframes are not compatible!\n")
}
```

```
## Everything is fine.
```

Define an auxiliary function:

```r
# Combines text strings
"%+%" <- function(string1, string2) paste0(string1, string2)
```

Functions

Define indicator-specific functions:

The first two functions use maximum-likelihood estimation to infer the standard deviation of the AOOs, based on the best estimate (median), low estimate (1st quartile) and high estimate (3rd quartile), and assuming a log-normal distribution.

```r
f <- function(s, mean, q1, q3) return(
  (q1 - qlnorm(0.25, log(mean) - exp(2*s)/2, exp(s)))^2 +
  (q3 - qlnorm(0.75, log(mean) - exp(2*s)/2, exp(s)))^2
)


findSD <- function(Ex, q1, q3) return(
  exp(optimise(f, c(-12, 12), mean=Ex, q1=q1, q3=q3)$min)
)


P2 <- function(dataset,
               column = c("known", "low", "best", "high"),
               nsim = 100000,
               maxArea = 323800) {
  # simulates AOOs for all species, which is the basis of indicator P2
  N <- nsim      # random numbers per species
  M <- maxArea   # maximum possible area in km^2
  aoo <- dataset[, column]
  AOO <- matrix(0, N, nrow(aoo))
  for (i in 1:nrow(aoo)) {
    if (any(is.na(aoo[i, c("low", "best", "high")]))) {
      if (is.na(aoo$best[i])) {
        if (is.na(aoo$known[i])) {
          # if no AOO is provided, assume it is 0
          AOO[, i] <- 0
        } else {
          # of no total AOOs are provided, assume they equal the known AOO
```

```
      AOO[, i] <- aoo$known[i]
      }
    } else {
      # if no low and high estimates are provided, use the best estimate
      AOO[, i] <- aoo$best[i]
    }
  } else {
    if (aoo$best[i] == 0) {
      # if the best estimate is 0, keep it 0
      AOO[, i] <- 0
    } else {
      # if low, best and high estimates are provided, estimate the standard
      # deviation and generate log-normally distributed random numbers
      SD <- findSD(aoo$best[i], aoo$low[i], aoo$high[i])
      AOO[, i] <- qlnorm(runif(N), log(aoo$best[i]) - SD * SD / 2, SD)
      AOO[, i] <-  sapply(AOO[, i], min, M)  # constrain to area of Norway
      AOO[, i] <- ceiling(AOO[, i] / 4) * 4  # ensure multiples  of 4 km^2
    }
  }
} # i
  return(AOO)
} # P2
```

Error checking and correction

Check for obvious errors in the original data:

(1) Is any low estimate greater than the corresponding best estimate?

```
w <- which(aoo$low > aoo$best)
if (length(w)) {
  print(aoo[w,])
} else {
  cat("Everything is fine.\n")
}
```

```
##                    Name known   low best   high
## 28 Acrotrichis cognata    80 40000  800 120000
```

This looks very much like a punching error in line 28. Most likely, the best estimate should have been in the middle between the low and high estimate (implying a dark figure of 1000 rather than 10). The error is corrected manually:

```
aoo$best[w] <- 80000
```

(2) Is any high estimate less than the corresponding best estimate?

```
w <- which(aoo$high < aoo$best)
if (length(w)) {
  print(aoo[w,])
} else {
  cat("Everything is fine.\n")
}
```

```
##                         Name known low  best high
## 55     Alphitophagus bifasciatus    20  20  1000   20
## 345       Dorytomus filirostris     4   4    40    4
```

```
## 472           Harmonia axyridis   128 128  2560  128
## 591        Lithocharis nigriceps   112 112  5600  112
## 644 Melampsoridium hiratsukanum   900 900  4500  900
## 680   Myrmecocephalus concinnus    12  12   240   12
## 704            Omalium rugatum   164 164 16400  164
## 726        Otiorhynchus armadillo    8   8   400    8
## 777        Philonthus rectangulus   60  60  3000   60
```

Here, it seems that the low and high estimate simply haven't been provided, and that, as a default, the known AOO was listed instead. In the absence of other information, we have to assume the low and high estimates to be equal to the best estimate:

```r
aoo$low[w] <- aoo$high[w] <- aoo$best[w]
```

(3) By definition, AOOs are multiples of 4 square kilometres. Some figures are incompatible with this definition. We solve this by rounding upwards:

```r
aoo[, 2:5] <- ceiling(aoo[, 2:5] / 4) * 4
```

(4) Is any AOO greater than the area of mainland Norway?

```r
w <- which(aoo$high > 323800)
if (length(w)) {
  print(aoo[w,])
} else {
  cat("Everything is fine.\n")
}
```

```
##                 Name known    low    best    high
## 112 Atomaria lewisi    176 88000 176000 880000
```

Indeed. That is clearly an overestimate. The high estimate is reduced to (a bit less than) the area of Norway. All other estimates are possible in principle, and are thus left unchanged.

```r
aoo$high[w] <- 323000
```

Show a graph of the distribution of the best estimates of total AOOs:

```r
par(mai=c(0.96, 0.96, 0.06, 0.06), lwd=1.8)
H <- hist(log10(as.numeric(aoo$best)), breaks=seq(0, 6, 1/3),
  xlab="Area of occupancy (km²)", ylab="Number of species", main="",
  xaxt="n", cex.axis=1.2, cex.lab=1.8, lwd=1.8, col=grey(0.84))
axis(1, 0:6, c(1, 10, 100, expression(10^3), expression(10^4), expression(10^5),
  expression(10^6)), cex.axis=1.2, lwd=1.8)
```
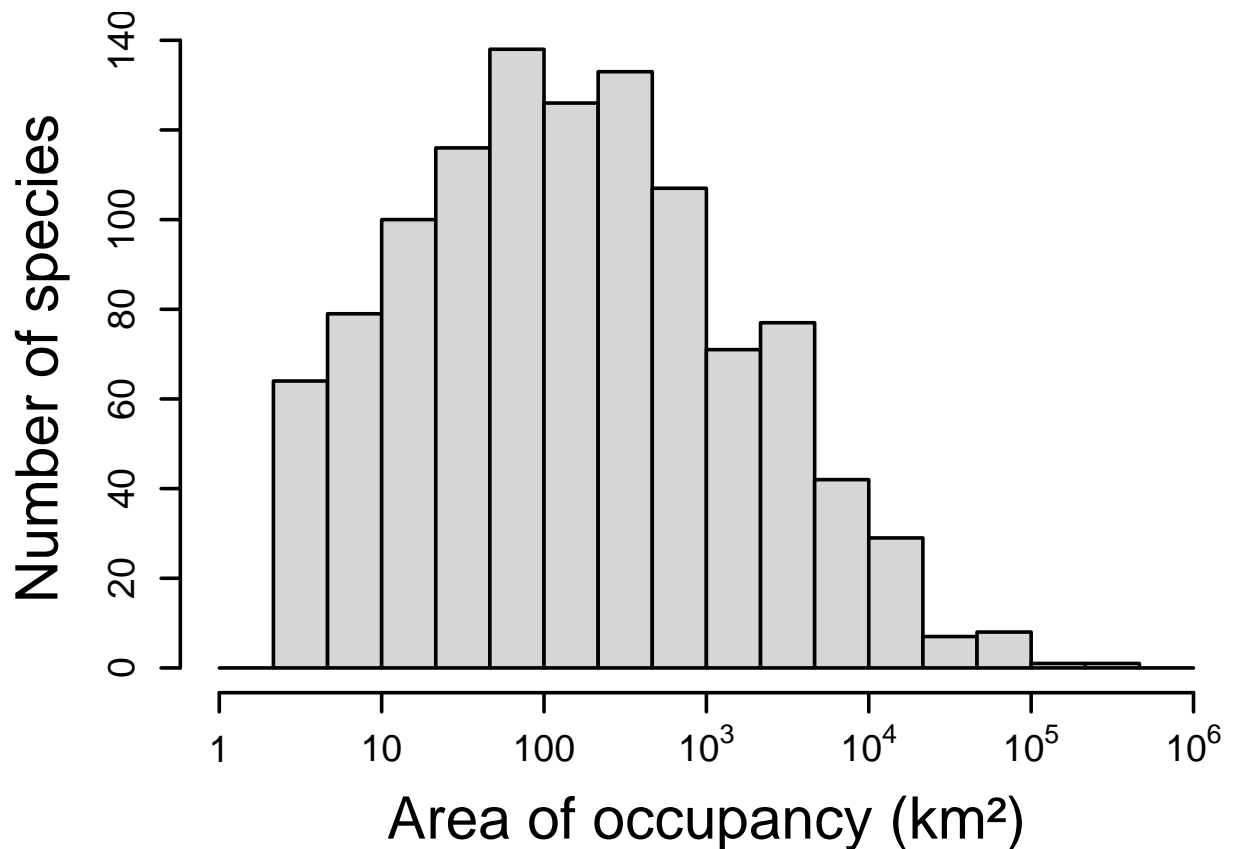
Simulations

If you want P2 for *high- and severe-impact* species only, do this first (not run):

```r
aoo <- aoo[which(fab$Impact %in% c("HI", "SE")),]
```

If you want P2 for the *remaining* species only, do this first (not run):

```r
aoo <- aoo[which(fab$Impact %in% c("NK", "LO", "PH")),]
```

If you want P2 for *all* alien species (**default**), do none of the above.

Then start the simulation:

```r
AOO <- P2(aoo)
avg   <-     mean(apply(AOO, 1, sum))
stdev <-       sd(apply(AOO, 1, sum))
conf  <- quantile(apply(AOO, 1, sum), c(0.025, 0.25, 0.5, 0.75, 0.975))
```

Output of the results:

```r
{
  cat("P2 is " %+% round(avg,    -2) %+% " km² ± " %+%
                  round(stdev, -2) %+% " km² (mean ± SD)\n\n")
  cat("Confidence levels (in km²):\n")
  print(round(conf, -2))
}
```

```
## P2 is 2410900 km² ± 197900 km² (mean ± SD)
```

```
## 
## Confidence levels (in km²):
##     2.5%      25%      50%      75%    97.5%
## 2054400 2271900 2400300 2538700 2828200
```