Distance sampling

Jez Bird

26 March 2019

## Cleaning raw detection data

Set working directory

knitr::opts\_knit$set(root.dir = normalizePath("C:/Users/Jez/Documents/PhD/Analysis/Data/Island-wide"))

Read in raw transect data - this contains all detections along transects. Transects have not been split into sections yet.

raw\_transect\_data <- read.csv("Raw\_transect\_data.csv")  
  
summary(raw\_transect\_data)

## Transect Start End Waypoint   
## Min. : 1.00 Min. : 288 Min. : 289.0 Min. : 1.0   
## 1st Qu.: 76.25 1st Qu.: 368 1st Qu.: 366.0 1st Qu.: 426.0   
## Median :114.00 Median : 514 Median : 510.0 Median : 880.0   
## Mean :108.62 Mean : 930 Mean : 917.3 Mean : 912.1   
## 3rd Qu.:144.00 3rd Qu.:1802 3rd Qu.:1750.0 3rd Qu.:1349.8   
## Max. :218.00 Max. :1992 Max. :1993.0 Max. :2000.0   
## NA's :1 NA's :55 NA's :120   
## lat lon ele   
## Min. :-54.77 Min. :158.8 Min. : 13.25   
## 1st Qu.:-54.71 1st Qu.:158.8 1st Qu.:178.82   
## Median :-54.66 Median :158.8 Median :220.49   
## Mean :-54.66 Mean :158.9 Mean :217.62   
## 3rd Qu.:-54.61 3rd Qu.:158.9 3rd Qu.:256.01   
## Max. :-54.49 Max. :158.9 Max. :396.17   
## NA's :123 NA's :123 NA's :123   
## time Species Mark Distance   
## : 123 AP :2848 : 120 Min. : 0.000   
## 2018-01-05T02:57:12Z: 1 WHP : 306 B:3051 1st Qu.: 0.800   
## 2018-01-05T03:00:46Z: 1 : 120 P: 195 Median : 1.600   
## 2018-01-05T03:06:42Z: 1 SPP? : 55 Mean : 2.328   
## 2018-01-05T03:09:06Z: 1 UNK : 15 3rd Qu.: 3.100   
## 2018-01-05T03:10:02Z: 1 WHP? : 13 Max. :40.000   
## (Other) :3238 (Other): 9 NA's :121   
## Slope Puff.age Veg.height Feather.sample  
## : 120 Min. :1.000 0: 617 :3342   
## D:1246 1st Qu.:2.000 S:2696 F? : 1   
## F: 552 Median :3.000 T: 53 F01 : 1   
## U:1448 Mean :2.626 F02 : 1   
## 3rd Qu.:3.000 F03 : 1   
## Max. :4.000 F04 : 1   
## NA's :3187 (Other): 19   
## Poo.sample Observer   
## :3349 JB:2126   
## 3 : 1 PP:1240   
## 4 : 1   
## POO 324 : 1   
## POO 353 VEG/RECENT: 1   
## POO327 DIRT/DRY : 1   
## (Other) : 12

colnames(raw\_transect\_data)

## [1] "Transect" "Start" "End" "Waypoint"   
## [5] "lat" "lon" "ele" "time"   
## [9] "Species" "Mark" "Distance" "Slope"   
## [13] "Puff.age" "Veg.height" "Feather.sample" "Poo.sample"   
## [17] "Observer"

Tidy the data

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

## Rename columns  
raw\_transect\_data = rename(raw\_transect\_data, veg\_height = `Veg.height`)  
raw\_transect\_data = rename(raw\_transect\_data, puff\_age = `Puff.age`)  
  
## Delete columns  
raw\_transect\_data$Feather.sample <- NULL  
raw\_transect\_data$Poo.sample <- NULL  
  
saveRDS(raw\_transect\_data, "Transect\_data\_renamed.Rds")  
  
transect\_data\_renamed = readRDS("Transect\_data\_renamed.Rds")  
  
sapply(transect\_data\_renamed, class)

## Transect Start End Waypoint lat lon   
## "integer" "integer" "integer" "integer" "numeric" "numeric"   
## ele time Species Mark Distance Slope   
## "numeric" "factor" "factor" "factor" "numeric" "factor"   
## puff\_age veg\_height Observer   
## "integer" "factor" "factor"

transect\_data\_renamed[["puff\_age"]] = as.factor(transect\_data\_renamed[["puff\_age"]])  
  
# reset missing veg\_height values as NA  
missing\_veg\_height <- transect\_data\_renamed[["veg\_height"]] == 0  
  
transect\_data\_renamed[["veg\_height"]][missing\_veg\_height] <- NA  
  
summary(transect\_data\_renamed)

## Transect Start End Waypoint   
## Min. : 1.00 Min. : 288 Min. : 289.0 Min. : 1.0   
## 1st Qu.: 76.25 1st Qu.: 368 1st Qu.: 366.0 1st Qu.: 426.0   
## Median :114.00 Median : 514 Median : 510.0 Median : 880.0   
## Mean :108.62 Mean : 930 Mean : 917.3 Mean : 912.1   
## 3rd Qu.:144.00 3rd Qu.:1802 3rd Qu.:1750.0 3rd Qu.:1349.8   
## Max. :218.00 Max. :1992 Max. :1993.0 Max. :2000.0   
## NA's :1 NA's :55 NA's :120   
## lat lon ele   
## Min. :-54.77 Min. :158.8 Min. : 13.25   
## 1st Qu.:-54.71 1st Qu.:158.8 1st Qu.:178.82   
## Median :-54.66 Median :158.8 Median :220.49   
## Mean :-54.66 Mean :158.9 Mean :217.62   
## 3rd Qu.:-54.61 3rd Qu.:158.9 3rd Qu.:256.01   
## Max. :-54.49 Max. :158.9 Max. :396.17   
## NA's :123 NA's :123 NA's :123   
## time Species Mark Distance   
## : 123 AP :2848 : 120 Min. : 0.000   
## 2018-01-05T02:57:12Z: 1 WHP : 306 B:3051 1st Qu.: 0.800   
## 2018-01-05T03:00:46Z: 1 : 120 P: 195 Median : 1.600   
## 2018-01-05T03:06:42Z: 1 SPP? : 55 Mean : 2.328   
## 2018-01-05T03:09:06Z: 1 UNK : 15 3rd Qu.: 3.100   
## 2018-01-05T03:10:02Z: 1 WHP? : 13 Max. :40.000   
## (Other) :3238 (Other): 9 NA's :121   
## Slope puff\_age veg\_height Observer   
## : 120 1 : 27 0 : 0 JB:2126   
## D:1246 2 : 57 S :2696 PP:1240   
## F: 552 3 : 51 T : 53   
## U:1448 4 : 44 NA's: 617   
## NA's:3187   
##   
##

library(ggplot2)  
library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

## Split the data by species and sign (burrow or puff)  
  
library(data.table)

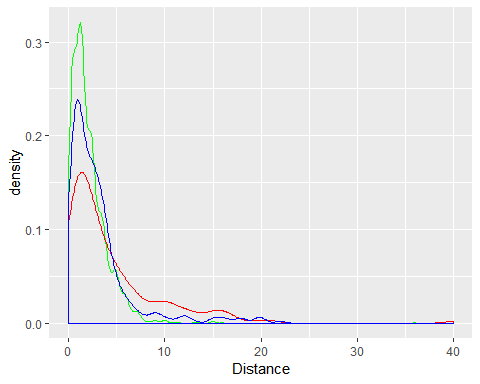
##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

setDT(transect\_data\_renamed)  
  
prion\_brws <- transect\_data\_renamed[Species == "AP" & Mark == "B",]  
prion\_puffs <- transect\_data\_renamed[Species == "AP" & Mark == "P",]  
WHP\_brws <- transect\_data\_renamed[Species == "WHP" & Mark == "B",]  
  
saveRDS(prion\_brws, "prion\_brws.Rds")  
prion\_brws = readRDS("prion\_brws.Rds")  
  
saveRDS(prion\_puffs, "prion\_puffs.Rds")  
prion\_puffs = readRDS("prion\_puffs.Rds")  
  
saveRDS(WHP\_brws, "WHP\_brws.Rds")  
WHP\_brws = readRDS("WHP\_brws.Rds")

Explore the influence of observer, slope, veg height on distance data

All\_sp\_plot <- ggplot() +   
 geom\_density(data=prion\_brws, aes(x=Distance), color='green') +  
 geom\_density(data = prion\_puffs, aes(x=Distance), color='red') +  
 geom\_density(data = WHP\_brws, aes(x=Distance), color='blue')  
All\_sp\_plot



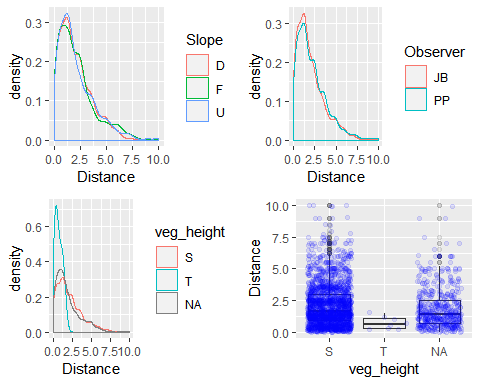
Prion puffs were detected at greater distances than burrows, and White-headed Petrel burrows were detected at greater distances than Antarctic Prion burrows

AP\_brw\_slope\_plot <- ggplot(data = prion\_brws, aes(x=Distance, color=Slope)) + geom\_density() + xlim(0,10)  
  
AP\_brw\_observer\_plot <- ggplot(data = prion\_brws, aes(x=Distance, color=Observer)) + geom\_density() + xlim(0,10)  
  
AP\_brw\_veg\_height\_density\_plot <- ggplot(data = prion\_brws, aes(x=Distance, color=veg\_height)) + geom\_density() + xlim(0,10)  
  
AP\_brw\_veg\_height\_box\_plot <- ggplot(data = prion\_brws, aes(x=veg\_height, y=Distance)) + geom\_boxplot(alpha=0.1) +geom\_jitter(alpha = 0.1, color = "blue") + ylim(0,10)  
  
grid.arrange(AP\_brw\_slope\_plot, AP\_brw\_observer\_plot, AP\_brw\_veg\_height\_density\_plot, AP\_brw\_veg\_height\_box\_plot, nrow = 2)

## Warning: Removed 10 rows containing non-finite values (stat\_density).  
  
## Warning: Removed 10 rows containing non-finite values (stat\_density).  
  
## Warning: Removed 10 rows containing non-finite values (stat\_density).

## Warning: Removed 10 rows containing non-finite values (stat\_boxplot).

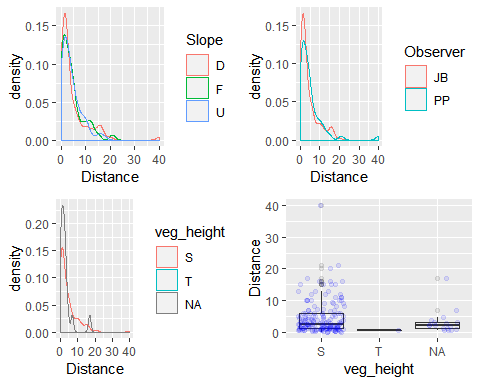
## Warning: Removed 40 rows containing missing values (geom\_point).



AP\_puff\_slope\_plot <- ggplot(data = prion\_puffs, aes(x=Distance, color=Slope)) + geom\_density() + xlim(0,40)  
  
AP\_puff\_observer\_plot <- ggplot(data = prion\_puffs, aes(x=Distance, color=Observer)) + geom\_density() + xlim(0,40)  
  
AP\_puff\_veg\_height\_density\_plot <- ggplot(data = prion\_puffs, aes(x=Distance, color=veg\_height)) + geom\_density() + xlim(0,40)  
  
AP\_puff\_veg\_height\_box\_plot <- ggplot(data = prion\_puffs, aes(x=veg\_height, y=Distance)) + geom\_boxplot(alpha=0.1) +geom\_jitter(alpha = 0.1, color = "blue") + ylim(0,40)  
  
grid.arrange(AP\_puff\_slope\_plot, AP\_puff\_observer\_plot, AP\_puff\_veg\_height\_density\_plot, AP\_puff\_veg\_height\_box\_plot, nrow = 2)

## Warning: Groups with fewer than two data points have been dropped.

## Warning: Removed 2 rows containing missing values (geom\_point).



WHP\_brw\_slope\_plot <- ggplot(data = prion\_puffs, aes(x=Distance, color=Slope)) + geom\_density() + xlim(0,25)  
  
WHP\_brw\_observer\_plot <- ggplot(data = prion\_puffs, aes(x=Distance, color=Observer)) + geom\_density() + xlim(0,25)  
  
WHP\_brw\_veg\_height\_density\_plot <- ggplot(data = prion\_puffs, aes(x=Distance, color=veg\_height)) + geom\_density() + xlim(0,25)  
  
WHP\_brw\_veg\_height\_box\_plot <- ggplot(data = prion\_puffs, aes(x=veg\_height, y=Distance)) + geom\_boxplot(alpha=0.1) +geom\_jitter(alpha = 0.1, color = "blue") + ylim(0,25)  
  
grid.arrange(WHP\_brw\_slope\_plot, WHP\_brw\_observer\_plot, WHP\_brw\_veg\_height\_density\_plot, WHP\_brw\_veg\_height\_box\_plot, nrow = 2)

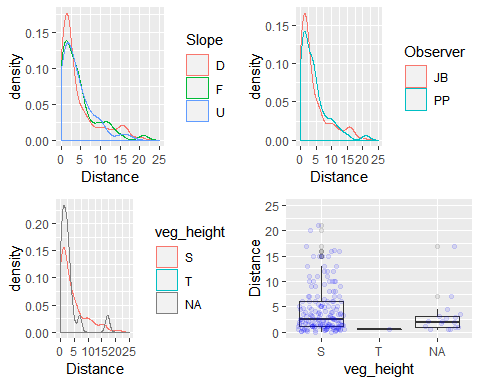
## Warning: Removed 1 rows containing non-finite values (stat\_density).

## Warning: Removed 1 rows containing non-finite values (stat\_density).  
  
## Warning: Removed 1 rows containing non-finite values (stat\_density).

## Warning: Groups with fewer than two data points have been dropped.

## Warning: Removed 1 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 4 rows containing missing values (geom\_point).



Observer and slope don’t appear to have much of an influence on detection distance so will perhaps be of limited importance when modelling detection functions. Veg height does seem to be important although there were too few records of puffs or White-headed Petrel burrows, and perhaps even prion burrows in tall veg to draw conclusions.

## Distance analysis

The value of distance sampling and analysis over plot sampling is allowing for abundance estimation despite imperfect detection. Distance analysis delivers model robustness, meaning that unbiased estimates are still obtained even if not all sources of variability in detection are incorporated. Hence, it is not common to fit an exhaustive candidate model set to data, but there are approaches to capture covariates that may generate models that fit the data better. So, I explore different covariates below.

Distance analysis generally uses one of four main functions to describe the drop off in detection with distance from the transect: 1. Uniform - effectively no drop off in detection. Unrealistic but works with an adjustment term (see below) 2. Half-normal 3. Hazard-rate - similar to half-normal - basically S-curves but slightly different shapes 4. Negative exponential - assumes detection starts dropping off immediately and then tails off with distance. Also unrealistic so not used here.