

Perils of multispecies and multisession distance sampling analysis

Example using Montrave data employing `region.table` and `sample.table` construct

AUTHOR

Eric Rexstad

AFFILIATION

CREEM, Univ of St Andrews

PUBLISHED

Jan. 20, 2022

Contents

A multispecies data set with multiple visits

Survey design

Analysis of only one species (incorrectly)

Use explicit data hierarchy

Simple detection function model

Consequence of incorrect analysis

A multispecies data set with multiple visits

It is increasingly common for investigators to conduct surveys in which multiple species are detected and density estimates for several species are of interest. There are many ways of analysing such data sets, but care must be taken. Not all approaches will produce correct density estimates. To demonstrate one of the ways to produce incorrect estimates, we will use the line transect survey data reported in [Buckland \(2006\)](#). This survey (and data file) recorded detections of four species of songbirds. We conduct an analysis of chaffinch (*Fringilla coelebs*) (coded c in the data file), but similar results would arise with the other species.

Begin by reading the flat file in a comma delimited format:

```
birds <- read.csv (
  "https://synergy.st-andrews.ac.uk/ds-manda/files/2016/11/montrave-line.csv")
```

Survey design

Buckland's design consisted of visiting each of the 19 transects in his study twice. To examine some of the errors that can arise from improper analysis, I choose to treat the two visits as strata for the express purpose of generating stratum (visit) -specific density estimates. Density estimates reported in [Buckland \(2006\)](#) are in units of birds $\cdot \text{hectare}^{-1}$.

```
birds$Region.Label <- birds$visit
cu <- convert_units("meter", "kilometer", "hectare")
```

Analysis of only one species (incorrectly)

The direct approach to producing a density estimate for the chaffinch would be to subset the original data frame and use the species-specific data frame for analysis. Begin by performing the subset operation.

```
chaf <- birds [ birds$species == "c", ]
```

When the data are subset, the integrity of the survey design is not preserved. A simple frequency table of the species-specific data frame flags up a number of transect/visit combinations where no chaffinches were detected. The result is that the subset data frame suggests 3 of the 19 transects lacked chaffinch detections on the first visit and one of the 19 transects lacked chaffinch detections on the second visit.



```
detects <- table ( chaf$Sample.Label, chaf$visit )
detects <- as.data.frame( detects )
names ( detects ) <- c ( "Transect", "Visit", "Detections" )
detects$Detections <- cell_spec( detects$Detections,
  background = ifelse ( detects$Detections == 0, "red", "white" ) )
knitr::kable ( detects, escape=FALSE ) %>%
  kable_paper( full_width=FALSE )
```

Transect	Visit	Detections
1	1	3
2	1	3
3	1	4
4	1	3
5	1	5
6	1	4
7	1	2
8	1	0
9	1	1
10	1	1
11	1	0
13	1	1
14	1	1
15	1	3
16	1	2
17	1	3
18	1	3
19	1	0

	19	1	0
1	2	1	
2	2	4	
3	2	3	
4	2	2	
5	2	4	
6	2	3	
7	2	3	
8	2	1	
9	2	0	
10	2	2	
11	2	1	
13	2	1	
14	2	1	
15	2	1	
16	2	1	
17	2	1	
18	2	4	
19	2	1	



But wait, the situation is even **worse** than the above table suggests. Transect 12 does not appear in the table above because there were no detections on *either* visit. Consequently, there were 4 transects without chaffinches on the first visit and 2 transects without chaffinches on the second visit.

Let's see what the `ds()` function thinks about the survey effort using information from the species-specific data frame.

```
chaf.wrong <- ds(chaf, key="hn", convert.units=
  = cu, truncation=95, formula=~Region.Label)
knitr::kable(chaf.wrong$dht$individuals$,
summary)%>%
  kable_paper(full_width=FALSE)%>%
  column_spec(6, background="salmon")%>%
  column_spec(7, background="steelblue")
```

Region	Area	CoveredArea	Effort	n	k	ER	se.ER	cv.ER
--------	------	-------------	--------	---	---	----	-------	-------

1	33.2	82.061	4.319	39	15	9.029868	1.1159303	0.1235821
2	33.2	83.562	4.398	34	17	7.730787	0.9798153	0.1267420
Total	66.4	165.623	8.717	73	32	8.374441	0.7396266	0.0883195

Examine the column labelled `k` (the number of transects) for each of the visits. Rather than the 19 transects that were surveyed on each visit, the `ds()` function erroneously believes there were only 15 transects surveyed on the first visit and 17 transects surveyed on the second visit.



Note also the number of detections per kilometer; roughly 9 on the first visit and 7.7 on the second visit. These encounter rates exclude kilometers of effort on transects where there were no detections. We will return to this comparison later

Use explicit data hierarchy

Additional arguments can be passed to `ds()` to resolve this problem. Consulting the `ds()` documentation

- `region.table` data.frame with two columns:
 - `Region.Label` label for the region
 - `Area` area of the region
 - `region.table` has one row for each stratum. If there is no stratification then `region.table` has one entry with `Area` corresponding to the total survey area. If `Area` is omitted density estimates only are produced.
- `sample.table` data.frame mapping the regions to the samples (i.e. transects). There are three columns:
 - `Sample.Label` label for the sample
 - `Region.Label` label for the region that the sample belongs to.
 - `Effort` the effort expended in that sample (e.g. transect length).

This analysis that produces erroneous results can be remedied by explicitly letting the `ds()` function know about the study design; specifically, how many strata and the number of transects within each stratum (and associated transect lengths).

Construct the `region` table and `sample` table showing the two strata with equal areas and each labelled transect (of given length) is repeated two times.

```
birds.regiontable <- data.frame(
  Region.Label = as.factor(
    c(1, 2)
  ), Area = c(33.2, 33.2)
)
birds.sampletable <- data.frame(
  Region.Label = as.factor(
    rep(c(1, 2), each = 19)
  ), Sample.Label = rep(1:19, times = 2),
  Effort = c(0.208, 0.401, 0.401, 0.299, 0.350, 0.401, 0.299, 0.350, 0.401, 0.299, 0.350, 0.401, 0.299, 0.350, 0.401, 0.299, 0.350, 0.401, 0.299, 0.350)
)
```

```

,
0.401 , 0.393 , 0.405 , 0.385 , 0.204
,
0.039 , 0.047 , 0.204 , 0.271 , 0.236
,
0.189 , 0.177 , 0.200 , 0.020 )
)

```

Simple detection function model

The chaffinch analysis is performed again, this time supplying the `region.table` and `sample.table` information to `ds()`. The correct number of transects (19) sampled on both visits (even though chaffinch was not detected on 4 transects on visit 1 and 2 transects on visit 2) is now recognised. Hence, the use of `region.table` and `sample.table` **solves the problem** of effort miscalculation if a species is not detected on all transects.

```

tr      <- 95 # as per Buckland (2006)
onlycf  <- ds ( data= birds [ birds $
species == "c" , ] ,
            region.table = birds.regiontable,
            sample.table = birds.sampletable,
            trunc= tr , convert.units= cu , key= "hn" ,
formula = ~ Region.Label)
knitr :: kable ( onlycf $ dht $ individuals$
summary ) %>%
kable_paper( full_width= FALSE ) %>%
column_spec( 6 , background= "salmon" ) %>%
column_spec( 7 , background= "steelblue")

```

Region	Area	CoveredArea	Effort	n	k	ER	se.ER	cv.ER
1	33.2	91.77	4.83	39	19	8.074534	1.2196305	0.1510465
2	33.2	91.77	4.83	34	19	7.039338	1.0612781	0.1507639
Total	66.4	183.54	9.66	73	38	7.556936	0.8031613	0.1062813

Consequence of incorrect analysis

To drive home the consequence of failing to properly specify the survey effort, contrast the encounter rate for the two visits from the incorrect calculations above (9.0 and 7.7 respectively), with the correct calculation (8.1 and 7.0 respectively). The number of transects is incorrect with the knock-on effect of effort being incorrect. If effort is incorrect then so too is covered area.

The ripple effect from incomplete information about the survey design results in positively biased estimates of density.

Buckland, S. T. (2006). Point-transect surveys for songbirds: Robust methodologies. *The Auk*, 123(2), 345–357. [https://doi.org/10.1642/0004-8038\(2006\)123\[345:PSFSRM\]2.0.CO;2](https://doi.org/10.1642/0004-8038(2006)123[345:PSFSRM]2.0.CO;2)

References