Interpreting MRDS output: making sense of all the numbers

ML Burt, CREEM, University of St Andrews (lb9@st-andrews.ac.uk)

INTRODUCTION

The MRDS package (Laake *et al*. 2015) for R (R core Team 2016) was written to allow the user to estimate abundance from a mark-recapture line transect survey (i.e. taking account of imperfect detection both on and away from the transect centreline). On running an analysis lots of output is generated and it can be a bit daunting for a first-time user. This document aims to help the user to understand the output and find key bits of information. The MRDS package can be used in R directly or via program Distance 7 Beta 4 (Thomas *et al*. 2010). The same output is available in both but in Distance output is generated automatically to a 'Results' tab and in R the user is required to do a bit of work to obtain the information (e.g. using the `summary` and `plot` commands). Some knowledge of conventional distance sampling (see Buckland *et al*. 2001) and MRDS is assumed; for details on undertaking a MRDS analysis see Burt *et al*. (2014).

The data used here as an example is from a survey of faecal pellets (Jenkins and Manly 2008; Example 1 of Burt *et al.* 2014). The Distance project of these data is available to download from http://distancesampling.org/Distance/example-projects/MEE_Burtetal_Example1.zip. Of interest in this analysis was the probability of detection of pellet groups rather than estimating density or abundance. Outputs using the example data are annotated (in red text) in the following sections but first terms and quantities are defined.

This is a work in progress; comments and suggestions to improve the document are welcome.

GLOSSARY

Covered region	Region covered during the survey i.e. $2wL$.
Study region	Area of interest.
Detected object	This could be a group (cluster) of objects and group size is recorded or individual objects if cluster size is one for all objects.
Observer	One or more people performing the same role or could be an acoustic or digital observer.
Observer 1	Also known as the Primary observer in a trial configuration setup.
Observer 2	Also known as the Secondary observer in a trial configuration setup.
DS model	Distance sampling model; fitted assuming $g(0)=1$ i.e. certain detection on the transect centreline. This could be a conventional distance sampling model or a multiple covariate distance sampling model (Marques and Buckland 2004).
MR model	Mark-recapture model; logistic regression model $p_{j\|3-j}(y,\underline{z}) = \frac{\exp(\beta_0+\beta_1 y+\sum_{k=1}^{K}\beta_{k+1}z_k)}{1+\exp(\beta_0+\beta_1 y+\sum_{k=1}^{K}\beta_{k+1}z_k)}$ where j (j=1 or 2) is observer, the β's are model coefficients, y is perpendicular distance and z are covariates.
IO configuration	Independent observer configuration; both observers search independently of the other observer. The probability of detection by either, or both, of the observers is of interest.
Trial configuration	One observer (often called the primary) searches independently. A second observer (often called the tracker) searches for animals and tracks them in order to determine more easily if the primary has also detected them. The probability of detection of the primary observer is of interest.
Full independence	Detections between observers are assumed to be independent at all perpendicular distances. This assumption requires only a MR model to be fitted.

Point independence — Detections between observers are assumed to be independent only at the point where perpendicular distance is zero (i.e. on the transect centreline). This assumption requires both a DS and MR model to be fitted.

NOTATION

Observed values

n_1 — total number of detected objects seen by observer 1 (also Primary observer)

n_2 — total number of detected objects seen by observer 2 (also Secondary observer)

n_D — total number of detected objects seen by both observers (Duplicate detections)

$n_P = n_1+n_2-n_D$ — total number of detected objects (Pooled detections)

$p_{1|2} = n_D/n_2$ — proportion detected by observer 1 of those seen by observer 2

$p_{2|1} = n_D/n_1$ — proportion detected by observer 2 of those seen by observer 1

Estimated values

The estimated probabilities are the probabilities of detection for detected objects. The model used to estimate them is given in parentheses.

$\hat{p}_j(0)$ (MR) Estimate of probability of detection (of objects) on the trackline for observer *j* (*j*=1 or 2). If the MR model is of the form $\hat{p}_{j|3-j}(y) = \frac{\exp(\hat{\beta}_0+\hat{\beta}_1 y)}{1+\exp(\hat{\beta}_0+\hat{\beta}_1 y)}$ i.e. no covariates (except distance) then

$\hat{p}_{j|3-j}(0) = \frac{\exp(\hat{\beta}_0)}{1+\exp(\hat{\beta}_0)}$. Similar calculations hold if observer is included (with the coefficient for observer included) but if other covariates are included, then the function is averaged over all covariates and a more complicated formula is used (see Laake and Borchers 2004).

$\hat{p}_P(0)$ (MR) Estimate of probability of detection on the trackline (for both observers combined). When the MR model is simple (i.e. only contains covariates for distance (and/or observer in an IO configuration)), then $\hat{p}_P(0) = \hat{p}_1(0) + \hat{p}_2(0) - \hat{p}_1(0)\hat{p}_2(0)$. This equation does not hold when other covariates are included in the MR model; in this case, the intercept is obtained by averaging over all covariates (see Laake and Borchers 2004).

$\hat{p}_{P.DS}$ (DS) Estimate of probability of detection (over all distances) for both observers pooled

$\hat{p}_{1.DS}$ (DS) Estimate of probability of detection (over all distances) for observer 1

$\hat{p}_P$ Estimate of probability of detection (over all distances) for both observers pooled taking into imperfect account detection on the trackline. Under the point independence assumption $\hat{p}_P = \hat{p}_P(0).\hat{p}_{P.DS}$

$\hat{p}_1$ Estimate of probability of detection (over all distances) for observer 1 taking into imperfect account detection on the trackline. Under the point independence assumption $\hat{p}_1 = \hat{p}_1(0)\hat{p}_{1.DS}$

$\hat{N}_{cIO} = \frac{n_P}{\hat{p}}$ Estimated number of groups in the covered region for IO configuration

$\hat{N}_{cT} = \frac{n_1}{\hat{p}_1}$ Estimated number of groups in the covered region for Trial configuration

$\hat{N}$ Estimated number of individuals in the study region

$\hat{N}_g$ Estimated number of groups, or clusters, in the study region

$E[\hat{s}] = \frac{\hat{N}}{\hat{N}_g}$ Expected group size

OUTPUT FROM MRDS

Output in Distance goes to the Results tab. In R the user requests output using `summary` and `plot` commands. The exact information provided in the output will depend on the observer configuration and the independence assumption used. Here, we follow the order of the output used in Distance.

Summary of the observations

The numbers of detected objects are tabulated and also plotted in histograms. The tabulated data in Distance is found on the Observation/Summary tab and the histograms are on the Observation/Plot tab. In R use `det.tables(ddfmodel)` to list these tables (for a fitted model called `ddfmodel`) and to plot the histograms use `plot(det.tables(ddfmodel))`.

The tabulated data consist of the numbers of objects detected in each perpendicular distance interval used for the histograms for observer 1, observer 2, pooled and duplicate detections. This information is useful because it illustrates the data that underpin the fitted models. Table 1 shows an example of tabulated data for three distance intervals (there are many more intervals in the actual data) and provides a summary of the key bits of information that can be found in these tables.

The data used for the histograms of the number of objects are described in Table 2a.

Detection function summary

In Distance the detection function(s) is summarised on the Detection Fct/Summary tab. In R use `summary(ddfmodel)`. The estimated coefficients of the fitted models are listed along with the probabilities of detection. The information included in the output depends on the configuration and independence assumption chosen. For an IO point independence model see Figure 1; for an IO full independence model see Figure 2; for a trial point independence model see Figure 3 and for a trial full independence model see Figure 4.

The detection function plots are described in Table 2b. In R use `plot(ddfmodel)`. The intercepts of the fitted models are also given in Table 2a.

Density and abundance estimates

Density and abundance estimates (if requested) are found in Distance on the 'Density Estimates and associated quantities' tab. In R, data frames containing information on strata (`region.data`), transects (`sample.data`) and observations (`obs.data`) are required as input to obtain density and abundance estimates using the `dht` function i.e. `dht(ddfmodel,region.data,sample.data,obs.data)`. These data link objects to transects and transects to survey regions and provide data on search effort and area of survey strata.

Summary data and estimates (density and abundance) are provided for groups (clusters) and individuals and also expected group size for each strata. In 'Summary statistics' (for either clusters or individuals) the number of objects (n) will depended on whether an IO configuration (n_P) or a trial configuration (n_1) is selected.

REFERENCES

Buckland ST, DR Anderson, KP Burnham. JL Laake, DL Borchers and L Thomas (2001) Introduction to Distance Sampling. Oxford University Press, Oxford, UK

Burt ML, DL Borchers, KJ Jenkins and TA Marques (2014) Using mark-recapture distance sampling methods on line transect surveys. Methods in Ecology and Evolution. doi: 10.1111/2041-210X.12294

Laake J, DL Borchers, L Thomas, D Miller and J Bishop (2015) mrds: Mark-Recapture Distance Sampling. R package version 2.1.14. https://CRAN.R-project.org/package=mrds

Laake JL and DL Borchers (2004) Methods for incomplere detection at distance zero. Advanced Distance Sampling (eds) ST Buckland DR Anderson, KP Burnham. JL Laake, DL Borchers and L Thomas, Oxford University Press, Oxford, UK

Marques FFC and ST Buckland (2004) Covariates models for the detection function. Advanced Distance Sampling (eds) ST Buckland DR Anderson, KP Burnham. JL Laake, DL Borchers and L Thomas, Oxford University Press, Oxford, UK

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Thomas L, ST Buckland, EA Rexstad, JL Laake, S Strindberg, SL Hedley, JRB Bishop, TA Marques and KP Burnham (2010) Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology 47: 5-14. DOI: 10.1111/j.1365-2664.2009.01737.x

Table 1 Observation summary tables: a) key information extracted from b) example output. The symbol ':' indicates that there are more distance intervals in the actual data.

a) Summary of key information for three distance intervals.

Interval	n_1	n_2	n_D	n_P	$p_{1\|2}$	$p_{2\|1}$	Results tab
[0, 5.17]	78	85	66	97	0.776	0.846	Observation/ Summary
[5.17, 10.3]	40	42	30	52	0.714	0.750	
[10.3, 15.5]	41	40	30	51	0.750	0.731	
:							
Total	1094	1102	816	1380	0.740[1]	0.745[1]	Detection Fct/Summary

[1] Not given in Detection function summary – these are Petersen estimates

b) Example output

```
Observer 1 detections
             Detected
              Missed Detected (n1)
  [0,5.17]        19       78          (19 + 78 = 97 = np)
  (5.17,10.3]     12       40
  (10.3,15.5]     10       41
     :

Observer 2 detections
             Detected
              Missed Detected (n2)
  [0,5.17]        12       85          (12 + 85 = 97 = np)
  (5.17,10.3]     10       42
  (10.3,15.5]     11       40
     :

Duplicate detections (nD)
   [0,5.17] (5.17,10.3] (10.3,15.5] (15.5,20.7] (20.7,25.9]   (25.9,31]
         66          30          30          53          35          46
     :

Pooled detections (nP)
   [0,5.17] (5.17,10.3] (10.3,15.5] (15.5,20.7] (20.7,25.9]   (25.9,31]
         97          52          51          86          64          64
     :

Observer 1 detections of those seen by Observer 2
            Missed Detected Prop. Detected (p1|2)
[0,5.17]        19       66      0.7764706       (66/(19+66) = 66/85 = 0.7764)
(5.17,10.3]     12       30      0.7142857
(10.3,15.5]     10       30      0.7500000
     :

Observer 2 detections of those seen by Observer 1
            Missed Detected Prop. Detected (p2|1)
[0,5.17]        12       66      0.8461538       (66/(12+66)= 66/78 = 0.8461)
(5.17,10.3]     10       30      0.7500000
(10.3,15.5]     11       30      0.7317073
     :
```

Table 2 Information plotted for each observer configuration (IO and Trial). A dash indicates that figure is not plotted.

a) Observation/Plot tab

Summary plot #		Histogram colour		Numbers of objects for who?
IO	Trial	Black	Blue	
1	1	n_P	n_1	Pooled and observer 1
2	2	n_P	n_2	Pooled and observer 2
3	3	n_D		Duplicates
4	-	n_P		Pooled
5	4	n_2	n_D	Observer 2 and duplicates
6	-	n_1	n_D	Observer 1 and duplicates

b) Detection Function/Plot tab

The points on the plots are estimated values for individual detections and the line is the average value (taking into account all covariates in the model).

Detection probability plot #		Histogram	Which model used for independence assumption?		Intercept of the line is at?
IO	Trial		Point	Full	
1	1	Scaled n_1	DS model	MR model	$\hat{p}_1(0)$
2	-	Scaled n_2	DS model	MR model	$\hat{p}_2(0)$
3	-	Scaled n_P	DS model	MR model	$\hat{p}_P(0)$
4	-	Scaled n_D	DS model	MR model	?
5	2	$p_{1\|2}$	MR model	MR model	$\hat{p}_1(0)$
6	-	$p_{2\|1}$	MR model	MR model	$\hat{p}_2(0)$

Figure 1 Example detection function summary for an **IO point independence** model: MR model contains distance and a factor for observer (this is a Petersen model); the DS model uses a hazard rate form with no covariates (apart from distance).

Summary for io.fi object (MR model)

```
Number of observations   :  1380                          nP
Number seen by primary   :  1094                          n1
Number seen by secondary :  1102                          n2
Number seen by both      :  816                           nD
AIC                      :  2652.566

Conditional detection function parameters:
               estimate          se
(Intercept)  1.334518220 0.107556941
distance    -0.004843781 0.001385673
observer2    0.028370866 0.084224532

                        Estimate          SE         CV
Average primary p(0)   0.7915870 0.017744426 0.02241627     p̂1(0)
Average secondary p(0) 0.7962288 0.017526680 0.02201211     p̂2(0)
Average combined p(0)  0.9575314 0.006690943 0.00698770     p̂P(0)
```

Summary for ds object (DS model)

```
Number of observations :  1380                            nP
Distance range         :  0   -  150
AIC                    :  13612.95

Detection function:
 Hazard-rate key function

Detection function parameters
Scale coefficient(s):
            estimate         se
(Intercept) 4.425513 0.05855335

Shape coefficient(s):
             estimate        se
(Intercept) 0.6851006 0.1247415

           Estimate         SE         CV
Average p 0.6924608 0.02190796 0.03163784                 p̂P.DS
```

Summary for io object (MR + DS model combined)

```
Total AIC value :  16255.2                          = 2652.566 + 13612.95

                       Estimate          SE         CV
Average p              0.663053  0.02148313 0.03240032     p̂P
N in covered region 2081.281660 74.86672579 0.03597145     N̂cIO
```

Figure 2 Example detection function summary for an **IO full independence** model: MR model contains covariates distance and observer (as a factor).

```
Summary for io.fi object (MR model)
Number of observations   :  1380                                  nP
Number seen by primary   :  1094                                  n1
Number seen by secondary :  1102                                  n2
Number seen by both      :  816                                   nD
AIC                      :  16481.92

Conditional detection function parameters:
               estimate          se
(Intercept)  1.334518220 0.107556941
distance    -0.004843781 0.001385673
observer2    0.028370866 0.084224532

                           Estimate           SE          CV
Average p                 0.9233260  0.007189382 0.007786396      p̂P
Average primary p(0)      0.7915870  0.016272902 0.020557313      p̂1(0)
Average secondary p(0)    0.7962288  0.016064551 0.020175796      p̂2(0)
Average combined p(0)     0.9575314  0.005181690 0.005411509      p̂P(0)
N in covered region    1494.5966586 16.110394124 0.010779091      N̂cIO
```

Figure 3 Example detection function summary for a **Trial point independence** model: DS model uses a hazard rate form with no covariates in the scale parameter; MR model contains distance only.

Summary for trial.fi object (MR model)

```
Number of observations                :  1380                    nP
Number seen by primary                :  1094                    n1
Number seen by secondary (trials)     :  1102                    n2
Number seen by both (detected trials):  816                      nD
AIC                                   :  1260.732

Conditional detection function parameters:
                estimate          se
(Intercept)  1.279522703 0.124363484
distance    -0.003960919 0.001732436

                      Estimate         SE         CV
Average primary p(0) 0.7823685 0.02117513 0.02706542              p̂1(0)
```

Summary for ds object (DS model)

```
Number of observations :  1094                                   n1

Distance range         :  0  -  150
AIC                    :  10770.29

Detection function:
 Hazard-rate key function

Detection function parameters
Scale coefficient(s):
            estimate         se
(Intercept) 4.442346 0.05685968

Shape coefficient(s):
             estimate       se
(Intercept) 0.8301251 0.133593

           Estimate         SE      CV
Average p 0.6936849 0.02237827 0.03226                           p̂1.DS
```

Summary for trial object (MR + DS model combined)

```
Total AIC value =  12031.02                                      = 10770.29 + 1260.73
                         Estimate          SE         CV
Average p                0.5427173  0.02285377 0.04210991        p̂1
N in covered region 2015.7825642 94.36006632 0.04681064          N̂cT
```

Figure 4 Example detection function summary for a **Trial full independence** model: MR model contains distance only.

Summary for trial.fi object (MR model)

```
Number of observations                :  1380                    nP
Number seen by primary                :  1094                    n1
Number seen by secondary (trials)     :  1102                    n2
Number seen by both (detected trials):  816                      nD
AIC                                   :  12185.06
```

Conditional detection function parameters:

	estimate	se
(Intercept)	1.279522703	0.124363484
distance	-0.003960919	0.001732436

	Estimate	SE	CV	
Average p	0.7262759	0.01521478	0.02094904	$\hat{p}_1$
Average primary p(0)	0.7823685	0.01621225	0.02072201	$\hat{p}_1(0)$
N in covered region	1506.3146420	39.23133973	0.02604458	$\widehat{N}_{cT}$