

# Modeling Abundance in Relation to Covariates

*Jason D. Carlisle and Trent L. McDonald*

*July 21, 2015*

## Introduction

This tutorial builds on the introductory material presented in the **Rdistance Tutorial for Beginners**. The goal of a distance-sampling study is often to relate abundance to some other variable, e.g., a covariate describing the habitat at each survey site. **Rdistance** can provide you with site-level estimates of abundance that are corrected for imperfect detection. You can then use these estimates as the response variable in one of the many modeling frameworks available in Program R that you're likely already familiar with (e.g., `lm`, `glm`, `lme4`, etc.). Here, we make use of the example datasets already contained within **Rdistance** (i.e., line transect surveys of sparrows), to demonstrate an example modeling workflow. This tutorial is current as of version 1.3.2 of **Rdistance**.

## Estimate abundance at each site

The process of generating an overall, study-area estimate of abundance is outlined in the **Rdistance Tutorial for Beginners**. Use the `by.id` argument (`by.id=TRUE`) within `F.abund.estim` or `F.automated.CDA` to estimate abundance for each site surveyed.

```
require(Rdistance)
data(sparrow.detections)
data(sparrow.transects)
auto <- F.automated.CDA(detection.data=sparrow.detections,
                        transect.data=sparrow.transects,
                        by.id=TRUE, w.hi=150, area=10000, R=10, ci=0.95,
                        plot=FALSE, plot.bs=FALSE)
```

```
## Likelihood Series Expans Converged? Scale? AIC
## halfnorm cosine 0 Yes 0k 3263.4429
## halfnorm cosine 1 Yes 0k 3261.585
## halfnorm hermite 1 Yes 0k 3262.8066
## halfnorm simple 1 Yes 0k 3262.1526
## halfnorm cosine 2 Yes 0k 3263.3155
## halfnorm hermite 2 Yes 0k 3263.593
## halfnorm simple 2 Yes 0k 3263.4741
## halfnorm cosine 3 Yes 0k 3265.1285
## halfnorm hermite 3 Yes 0k 3265.5499
## halfnorm simple 3 Yes 0k 3265.216
## hazrate cosine 0 Yes 0k 3267.6246
## hazrate cosine 1 Yes 0k 3263.3092
## hazrate hermite 1 No NA NA
## hazrate simple 1 No NA NA
## hazrate cosine 2 Yes 0k 3265.0704
## hazrate hermite 2 Yes 0k 3268.9682
## hazrate simple 2 Yes 0k 3265.5796
## hazrate cosine 3 Yes 0k 3267.117
```

```

## hazrate      hermite 3    Bad    NA NA
## hazrate      simple 3    No     NA NA
## uniform      cosine 0    Yes    Ok 3260.7318
## uniform      cosine 1    Bad    NA NA
## uniform      hermite 1    Yes    Ok 3262.736
## uniform      simple 1    Yes    Ok 3262.2602
## uniform      cosine 2    Bad    NA NA
## uniform      hermite 2    No     NA NA
## uniform      simple 2    Bad    NA NA
## uniform      cosine 3    Bad    NA NA
## uniform      hermite 3    No     NA NA
## uniform      simple 3    Bad    NA NA
## negexp       cosine 0    Yes    Ok 3263.8153
## negexp       cosine 1    Yes    Ok 3260.0985
## negexp       hermite 1    Yes    Ok 3260.3108
## negexp       simple 1    Yes    Ok 3261.0097
## negexp       cosine 2    Yes    Ok 3262.0306
## negexp       hermite 2    No     NA NA
## negexp       simple 2    Yes    Ok 3262.2878
## negexp       cosine 3    Yes    Ok 3263.4898
## negexp       hermite 3    No     NA NA
## negexp       simple 3    Yes    Ok 3264.2055
## Gamma        0    Yes    Ok 3586.8919
## Note: Some models did not converge or had parameters at their boundaries.
## Computing bootstrap confidence interval on N...
## =====
##
##
## ----- Final Automated CDS Abundance Estimate -----
## Call: F.dfunc.estim(dist = dist, likelihood = fit.table$like[1], w.lo = w.lo,      w.hi = w.hi, expansion = expansion)
##
## Coefficients:
##      Beta      a1
## 0.02754839 -0.26542176
##
## Convergence: Success
## Function: NEGEXP with 1 expansion(s) of COSINE series
## Strip: 0 to 150
## Effective strip width: 44.72749
## Scaling: g(0) = 1
## Log likelihood: 1628.032
## AIC: 3260.098
##
## Abundance estimate: 1.152038 ; 95% CI=( 0.7236135 to 1.411445 )

```

Because `by.id=TRUE`, a data.frame of the site-level abundances is stored in the `auto` object under the name `nhhat.df`:

```
head(auto$nhhat.df)
```

```

##   siteID rawcount    nhhat
## 1     A1       15 3.3536424
## 2     A2       14 3.1300662

```

```
## 3      A3      10 2.2357616
## 4      A4      15 3.3536424
## 5      B1       1 0.2235762
## 6      B2       0 0.0000000
```

Because distances were measured in meters and `area=10000`, abundance (density) estimates are given as the number of sparrows per hectare.

## Model abundance estimates

The original transect dataset contained a variable named `sagemean` that represents the average sagebrush cover at each transect. Suppose we want to examine the relationship between sagebrush cover and sparrow abundance. First, merge the new data.frame of site-level abundance estimates back to the original transect data.frame that includes `sagemean`.

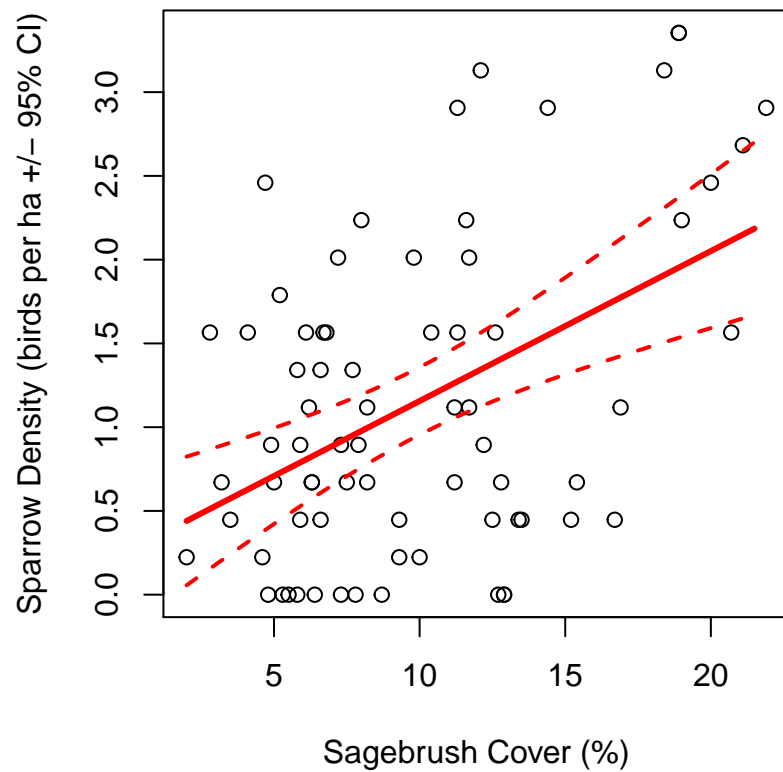
```
mydata <- merge(sparrow.transects, auto$nhathat.df, by="siteID")
head(mydata)
```

```
##   siteID length observer sagemean sage rawcount      nhathat
## 1     A1     500    obs4     18.9 High        15 3.3536424
## 2     A2     500    obs4     18.4 High        14 3.1300662
## 3     A3     500    obs5     19.0 High        10 2.2357616
## 4     A4     500    obs5     18.9 High        15 3.3536424
## 5     B1     500    obs3       4.6 Low         1 0.2235762
## 6     B2     500    obs3       4.8 Low         0 0.0000000
```

We now have both variables in the same data.frame. For the sake of demonstration, we will fit a simple linear regression model using the `lm` function to assess the relationship between sagebrush cover and sparrow density (per ha). We see that sparrow density increases with sagebrush cover, and density nearly quadruples across the range of sagebrush cover surveyed.

```
mod <- lm(nhathat ~ sagemean, data=mydata)
summary(mod)
```

```
##
## Call:
## lm(formula = nhathat ~ sagemean, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4157 -0.7356 -0.1005  0.6878  1.7860
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.2611     0.2286   1.142   0.257
## sagemean       0.0895     0.0206   4.344 4.64e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8569 on 70 degrees of freedom
## Multiple R-squared:  0.2123, Adjusted R-squared:  0.2011
## F-statistic: 18.87 on 1 and 70 DF, p-value: 4.639e-05
```



## Conclusion

By using `Rdistance` to estimate site-level abundance, you can account for imperfect detection, and use the modeling framework you choose to then relate site-level covariates to abundance. We demonstrated this workflow using `Rdistance`'s `F.automated.CDA` function to fit a detection function and provide abundance estimates for each site, and the standard `lm` function to model the relationship between a habitat covariate and sparrow density.