

CHAPTER 10

Accounting for imperfect detection in estimates of yearly site occupancy

Erik Hedlin and Alastair Franke

Hedlin, E., and A. Franke. 2017. Accounting for imperfect detection in estimates of yearly site occupancy. Pages 183–206 in D.L. Anderson, C.J.W. McClure, and A. Franke, editors. *Applied raptor ecology: essentials from Gyrfalcon research*. The Peregrine Fund, Boise, Idaho, USA. <https://doi.org/10.4080/are.2017/010>

10.1 Introduction

Estimating site occupancy is an essential component of any monitoring program because changes in occupancy over time can serve as a metric of population status (MacKenzie et al. 2003). Detection is often imperfect during occupancy surveys, however, and estimating the proportion of occupied sites without accounting for detection error can lead to an underestimation of occupancy (Kéry and Schmidt 2008). Hierarchical occupancy models have become a popular tool to address such issues because they provide a framework for estimating parameters that drive occupancy while accounting for detection probability that is <1 (Marsh and Trenham 2008).

In this chapter we use a simulated data set that includes survey observations for 30 Gyrfalcon breeding sites across 20 years. The goal of this analysis is to gain a better understanding of the observed patterns of yearly occupancy in this simulated sample population, and to derive parameters of occupancy that tell us about population status using the multi-season occupancy model outlined in MacKenzie et al. (2003).

10.2 Occupancy model

The multi-season occupancy model is said to be hierarchical because it is composed of two submodels: 1) a model that describes the occupancy state, and 2) a model that describes the detection process (i.e., our observations in the field).

10.2.1 Occupancy state model

There are only two possible outcomes regarding occupancy (a site is either occupied or not), and we therefore have a two-level, or binary, response variable (i.e., a Bernoulli trial). We could use logistic regression (see Chapter 6) to determine factors that influence the probability of occupancy within a year, but we can gain valuable inference regarding population status by examining occupancy changes over many years. To do this we model a series of Bernoulli trials.

The structure of this occupancy model is based on the metapopulation model proposed by Hanski (1998) and uses the parameters colonization (γ) and extinction (ϵ) to explain changes in annual occupancy. We note here that many sources use the term survival probability (ϕ), or the probability that a site remains occupied across two consecutive years, rather than the term extinction. However, extinction is merely the inverse of survival ($\phi = 1 - \epsilon$) and thus both terms can be used interchangeably.

Because colonization and extinction in the current year are dependent on the previous year's occupancy state, this model requires a third parameter that describes the probability of occupancy in the first year (ψ_1). When combined, these three parameters describe a time series process that begins with initial occupancy probability (ψ_1), followed by occupancy that is driven by colonization and extinction rates, as follows:

If we let z indicate the occupancy state of site i at time $t = 1$ (i.e., the first survey year), our model begins with:

$$z_{i1} = \text{Bernoulli}(\psi_{i1})$$

Following the first year, the occupancy state at $t = 2$ is:

$$z_{i2} = \text{Bernoulli}(z_{i,t-1}\phi_{it} + (1 - z_{i,t-1})\gamma_{it})$$

We have two components in this second year Bernoulli trial: 1) the probability of survival given occupancy in the previous year: $z_{i,t-1}\phi_{it}$, and 2) the probability of colonization given the site is unoccupied in the first year: $(1 - z_{i,t-1})\gamma_{it}$. In other words, if site i was occupied in year 1 and we therefore substitute $z_{i,t-1}$ with 1, we get an occupancy probability that is fully dependent on the probability of survival:

$$z_{i2} \sim \text{Bernoulli}(1 * \phi_{it} + (1 - 1)\gamma_{it})$$

Or

$$z_{i2} \sim \text{Bernoulli}(1 * \phi_{it} + \cancel{(1 - 1)}\gamma_{it})$$

If site i is unoccupied at t_1 ($z_{i,t-1} = 0$) then we get an occupancy probability that is described only by the probability of colonization:

$$z_{i2} \sim \text{Bernoulli}(0 * \phi_{it} + (1 - 0)\gamma_{it})$$

Or

$$z_{i2} \sim \text{Bernoulli}(\cancel{0 * \phi_{it}} + (1)\gamma_{it})$$

This is because a site can only be colonized if it is unoccupied at $t-1$. By applying this model to our observed data, we can obtain estimates of annual colonization (i.e., an unoccupied site becomes occupied), extinction (i.e., an occupied site becomes unoccupied), and survival (i.e., an occupied site remains occupied), but would also like to investigate whether certain covariates are influencing these parameters. We accomplish this by linking specific covariates to each of the three parameters using a logit link function. All combined, our state occupancy model (Fig. 10.1) can be summarized as follows:

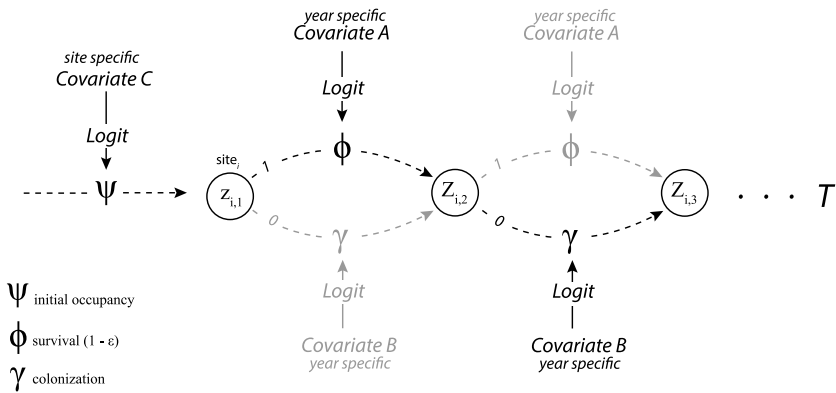


Figure 10.1 An illustration of the occupancy state model. As displayed, covariates are linked to initial occupancy, survival, and colonization using logit link functions. Following the first season (initial occupancy), the model can follow one of two pathways depending on the current season’s occupancy state. If a site is occupied in the current season (1), occupancy in the following season depends on the probability of survival (or conversely, 1 – the probability of extinction). If a site is unoccupied, occupancy in the following season is dependent on the probability of colonization.

10.2.2 Observation process

It is unrealistic to assume a perfect detection rate when surveying for Gyrfalcon occupancy. Many factors can contribute to detection error, and it is therefore important to include a submodel that accounts for such error. For detection rate to be assessed, each site must be surveyed more than once in each year to construct a detection history (Table 10.1, Fig. 10.2).

Occupancy modeling assumes that there are no false positives (i.e., indicating that a site is occupied when the true state is vacant), and the detection model is therefore dependent on the outcome of the occupancy state. If we let y_{ijt} denote the detection at site i during survey j in year t , then:

$$y_{ijt} = \text{Bernoulli}(z_{it} p_{ijt})$$

where z_{it} is the occupancy state at site i in year t , and p_{ijt} is the detection probability. Based on this equation, it is again not possible to detect occupancy if the true state is zero (i.e., $z_{it} = 0$), and thus one of the assumptions of the detection submodel is that there are no species misidentifications. If one is concerned about species misidentification, we point the reader to Miller et al. (2011). As with the parameters in the occupancy state model, detection probability may vary according to covariates. We can again use a logit link function to assess the relationship between covariates and detection probability.

10.2.3 Assumption of closure

The goal of this analysis is to determine how occupancy changes among primary sampling periods (i.e., years), however an important assumption of this analysis is that true occupancy doesn't change within each primary sampling period. If, for example, a breeding site is found to be occupied during the pre-laying and incubation periods, but then the nest fails and the adults either leave the breeding area or become less conspicuous, detection probability in all of the following surveys will decrease considerably. If this occurs frequently enough, detection probabilities become biased and cause overestimation of occupancy parameters (Rota et al. 2009, Kendall et al. 2013, Otto et al. 2013).

The violation of closure is a relatively common issue, and a number of methods have been proposed to relax this assumption (Bailey et al. 2014). One way to address closure is to conduct additional surveys within each secondary sampling period, which would re-assign the assumption of closure to a shorter period when closure could be confidently assumed (Williams et al. 2002, Rota et al. 2009). In adapting this method to the survey scheme we've proposed in this chapter (Fig. 10.2), one would have to conduct at least twice as many surveys per year (a minimum of two surveys within the pre-laying, incubation, and brood rearing periods).

Such increases in survey effort may not be feasible given the logistical issues associated with Gyrfalcon occupancy surveys. An appealing alternative is to introduce additional model parameters that indicate an individual's entry to, and exit from, the primary sampling period (Kendall et al. 2013). This provides the boundaries within which detection is possible, and thus reduces the bias caused from the violation of closure. Although we do not include these methods here, we recommend that users consider the assumption of closure in their study design.

10.3 Data

Here we examine the simulated occupancy history of 30 Gyrfalcon nesting territories over a period of 20 years. We generated the data within a surveying scheme that assumed each site was visited three times per year with the first visit occurring during the pre-laying period, the second during incubation, and the third during brood rearing (Fig. 10.2). We introduced three levels of covariates: 1) site-level covariates that vary among sites and relate to initial occupancy (ψ_1), 2) year-specific site-level covariates that describe colonization and extinction rates (γ and ϵ), and 3) observation covariates that vary across site, year, and survey to describe the detection process (p). Site-level covariates included distance from a nest site to anthropogenic disturbance (e.g., road; to test whether distance to disturbance influenced occupancy), the slope of the nesting cliff, the horizontal length of the nesting cliff, the distance to water, and the distance to the nearest neighbor. Year-specific site covariates (i.e., site covariates that changed among years) included the presence of a camera with ten cameras deployed each year, ptarmigan abundance, and lemming abundance. We included two observation covariates, one that indicated whether a survey was completed during the morning (A = 9 a.m. to 12 p.m.), midday (B = 12 p.m. to 3 p.m.), or afternoon (C = 3 p.m. to 6 p.m.), and a second that identified the survey period: pre-laying (pl), incubation (in), or brood rearing (br).

Although we need to format our data when we import it into R, we can start with a single spreadsheet that includes all of the information we would like to model. Each row of this spreadsheet contains data that relate to a specific Gyrfalcon nesting territory (labeled site1 to site30), and the information in each column depends on the matrix it is in. The observation matrix and the related observation covariate matrix has I rows with $T*m$ columns where I is the total number of sites, T is the total number of years, and m is the total number of surveys conducted per site per year (Table 10.1 and Fig. 10.2). Each site-level covariate is a vector of length i (Table 10.2), and year-specific site-level covariates are matrices with I rows, and T columns (Table 10.3).

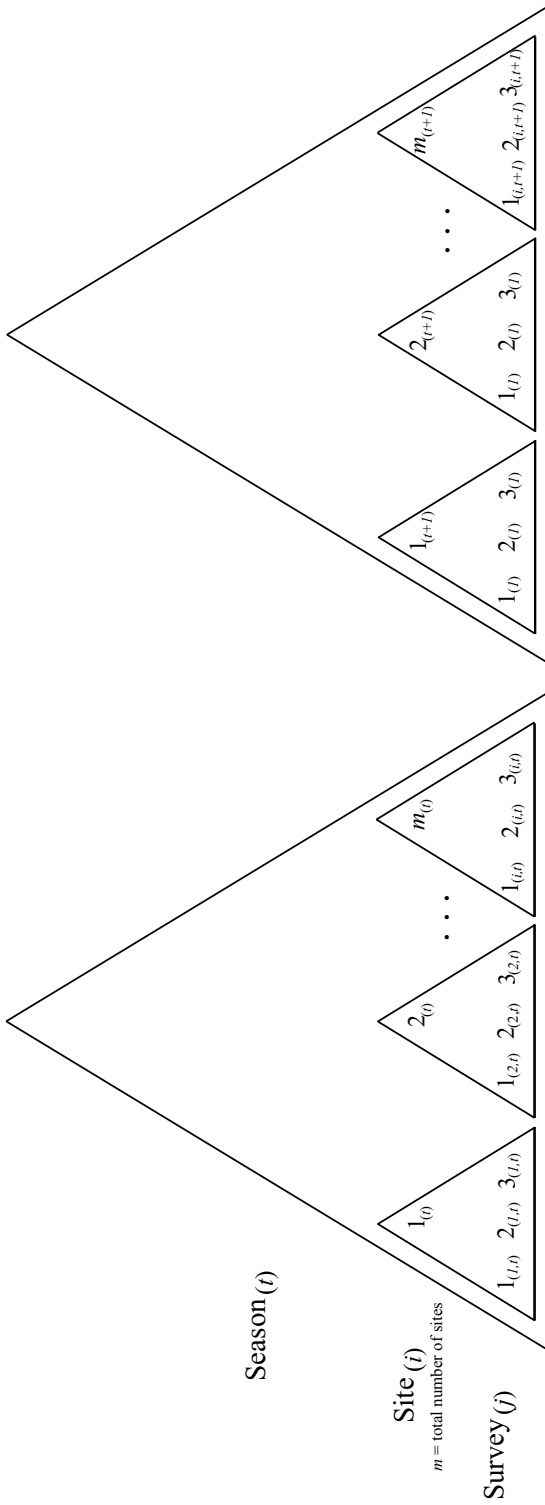


Figure 10.2 An illustration of the survey structure for this analysis. To build a detection history and therefore calculate detection probabilities, more than one survey must be conducted per site per year. In this example we chose to include three surveys.

Table 10.1 An example of the observation matrix where 1 = detected occupancy, and 0 = no detected occupancy. The observation covariate matrix is similar in structure, but is composed of information relating to each observation (i.e., the time of day the observation occurred).

| | Year 1 | | | Year 2 | | | Year 3 | | | ... | t |
|--------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----|-----------|
| | Obs1 | Obs2 | Obs3 | Obs1 | Obs2 | Obs3 | Obs1 | Obs2 | Obs3 | ... | <i>j</i> |
| Site1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | $t_{j,1}$ |
| Site2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | $t_{j,2}$ |
| Site3 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | ... | $t_{j,3}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Site _{<i>j</i>} | $1,1,i$ | $1,2,i$ | $1,3,i$ | $2,1,i$ | $2,2,i$ | $2,3,i$ | $3,1,i$ | $3,2,i$ | $3,3,i$ | ... | t_{ji} |

Table 10.2 Site-level covariates used to model initial occupancy. Each covariate is a single vector of values that apply to specific sites.

| | Slope |
|--------------------------|---------------------------|
| Site1 | 70 |
| Site2 | 45 |
| Site3 | 58 |
| ⋮ | ⋮ |
| Site _{<i>j</i>} | Slope _{<i>j</i>} |

Table 10.3 An example data table of year-specific covariates that indicates whether a motion-sensitive camera was present in a given year.

| | Year1 | Year2 | Year3 | ... | t |
|--------------------------|--------------|--------------|--------------|-----|----------|
| Site1 | no | no | no | ... | $t,1$ |
| Site2 | no | yes | no | ... | $t,2$ |
| Site3 | yes | yes | yes | ... | $t,3$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Site _{<i>j</i>} | $1,i$ | $2,i$ | $3,i$ | ... | t,i |

10.4 Analysis

Before we can start fitting occupancy models, we need to organize and transform our data. We will use the package `Unmarked` (Fiske and Chandler 2011) in R (R Core Team 2016) for our analysis, and this package requires that our data are coerced into a unique type of data frame to fit our models. We first extract all of the model components so we can organize them.

```
# extract covariates from our data
yMat      <- occdat[,2:61]      # detection matrix
cam       <- occdat[,87:106]   # camera presence
year      <- as.character(1996:2015) # year Matrix
year      <- matrix(year, nrow(occdat), 20, byrow=TRUE)
time      <- occdat[,147:206]  # survey time (within day)
season    <- occdat[,207:266]  # survey time (within season)
dtw       <- occdat$dtw        # distance of nest to water
cl.slope  <- occdat$cl.slope   # cliff slope
dtn       <- occdat$dtn        # distance to nearest territory
cl.length<- occdat$cl.length   # cliff horizontal length
disturb   <- occdat$disturb    # distance to disturbance
lem       <- occdat[,107:126]  # lemming abundance
ptar      <- occdat[,127:146]  # ptarmigan abundance
```

The next step is to transform some of the variables. Standardization of variables like 'dtn' that vary widely (468m to 3838m) can improve our model convergence in subsequent steps. Standardization consists of rescaling the variable's range by subtracting the mean and dividing by the standard deviation. Such transformations make the variable more numerically manageable throughout the model fitting process. Note that we do not show the standardization for all variables; see Online Appendix 10.1 for more details.

```
# scale variables to improve the model fitting process

# distance to disturbance
dist.sc <- as.data.frame((disturb -
  mean(disturb)) / sd(disturb))
# renames variable to "disturbance"
names(dist.sc)[names(dist.sc)=="(disturb -
  mean(disturb)) / sd(disturb)"] <- "disturbance"
```



```

# cliff slope
cl.slope.sc <- as.data.frame((cl.slope -
  mean(cl.slope)) / sd(cl.slope))
# renames variable to "cliff.slope"
names(cl.slope.sc)[names(cl.slope.sc)=="(cl.slope -
  mean(cl.slope)) / sd(cl.slope)"] <- "cliff.slope"

# cliff length
cl.length.sc <- as.data.frame((cl.length -
  mean(cl.length)) / sd(cl.length))
# renames variable to "cliff.length"
names(cl.length.sc)[names(cl.length.sc)=="(cl.length
  - mean(cl.length)) / sd(cl.length)"] <- "cliff.length"

# distance to water
dtw.sc <- as.data.frame((dtw - mean(dtw)) / sd(dtw))
# renames variable to "dtw"
names(dtw.sc)[names(dtw.sc)=="(dtw -
  mean(dtw)) / sd(dtw)"] <- "dtw"

# distance to neighbour
dtn.sc <- as.data.frame((dtn - mean(dtn)) / sd(dtn))
# renames variable to "dtn"
names(dtn.sc)[names(dtn.sc)=="(dtn -
  mean(dtn)) / sd(dtn)"] <- "dtn"

```

Now that our variables have been extracted and standardized where necessary, we need to organize them into the unique data frame required by Unmarked. We do this by grouping our covariates into data frames according to their type. In this example we have 3 types, 1) site-level covariates, 2) year-specific site covariates, and 3) observation covariates.

```

# site level covariates
s.covs <- data.frame(
  disturbance = matrix(t(dist.sc), ncol=1),
  cl.slope    = matrix(t(cl.slope.sc), ncol=1),
  cl.length   = matrix(t(cl.length.sc), ncol=1),
  dtw         = matrix(t(dtw.sc), ncol=1),
  dtn         = matrix(t(dtn.sc), ncol=1)
)

```

```

# year-specific site covariates
year.covs <- data.frame(
  cam      = matrix(t(cam), ncol = 1),
  year     = matrix(t(year), ncol = 1),
  ptar     = matrix(t(ptar.sc), ncol = 1),
  lem      = matrix(t(lem.sc), ncol = 1)
)

# observation covariates
obs.covs <- data.frame(
  day      = matrix(t(time), ncol = 1),
  season   = matrix(t(season), ncol = 1)
)

```

The next step is to use the `unmarkedMultFrame` function to coerce the data into the unique data frame required to fit the models.

```

# prepare data for analysis
gyrUMF <- unmarkedMultFrame(
  y = yMat,                # detection matrix
  siteCovs = s.covs,      # site level covs
  obsCovs = obs.covs,    # observation covs
  yearlySiteCovs = year.covs, # site-year covs
  numPrimary = 20        # number of seasons
)

# summary of our covariates and plot of yearly occupancy
summary(gyrUMF)
plot(gyrUMF)

```

Note that although we are using three types of covariates in this example, the minimum information needed to fit occupancy models in an `unmarkedMultFrame` is the observation matrix (`y`), and the specified number of years in which observations were gathered (`numPrimary`).

The next step in this analysis is to begin fitting models. The model structure used in the function `colext` requires specifying the four parameters and indicating the data source. Here is the null model that specifies constant parameter estimates across all years (i.e., the colonization, extinction, and detection probabilities are the same across all 20 years).

```

# initial (i.e., first-year) occupancy, colonization,
# extinction, detection
m0 <- colext(psiformula = ~1,      # 1st-year occupancy
             gammaformula = ~1,    # colonization
             epsilonformula = ~1,  # extinction
             pformula = ~1,       # detection
             data = gyrUMF # data source
            )

```

It is unlikely that colonization and extinction probabilities remain static across all 20 years, and we are interested in determining which of our covariates best describe the observed occupancy. To do this, we must build a set of candidate models that each test a particular hypothesis that is ecologically justifiable. As in Chapter 6, we use an information theoretic approach to rank our models according to Akaike's Information Criterion (AIC; Akaike 1974). A model's AIC score is driven by how well it fits the observed data, but is penalized according to the number of parameters within the model (a measure of parsimony). Ecological studies using AIC for model selection often use a correction factor to account for small sample size (Hurvich and Tsai 1989, Burnham and Anderson 2002), but because sample size is not known in dynamic occupancy models we leave this correction factor out.

```

# block 1: initial (i.e., first year) occupancy, coloniza-
# tion, extinction, detection
m1 <- colext(~dtn, ~year, ~year, ~day, data = gyrUMF)
m2 <- colext(~dtn, ~ptar, ~year, ~season, data = gyrUMF)
m3 <- colext(~dtn, ~1, ~cam, ~year, data = gyrUMF)
m4 <- colext(~disturbance, ~year, ~year, ~season,
             data = gyrUMF)
m5 <- colext(~disturbance, ~ptar, ~cam, ~1, data = gyrUMF)
m6 <- colext(~disturbance, ~lem, ~1, ~season,
             data = gyrUMF)
m7 <- colext(~cl.slope, ~year, ~year, ~day,
             data = gyrUMF)
m8 <- colext(~cl.slope, ~1, ~cam, ~season,
             data = gyrUMF)
m9 <- colext(~cl.slope, ~lem, ~1, ~year,
             data = gyrUMF)
m10 <- colext(~cl.length, ~year, ~year, ~1,
             data = gyrUMF)
m11 <- colext(~cl.length, ~ptar, ~cam, ~season,
             data = gyrUMF)

```

```

m12<- colext(~cl.length, ~lem, ~1, ~year,
             data = gyrUMF)
m13<- colext(~dtw, ~1, ~cam, ~day, data = gyrUMF)
m14<- colext(~dtw, ~ptar, ~1, ~year, data = gyrUMF)
m15<- colext(~dtw, ~lem, ~year, ~1, data = gyrUMF)

# block 2: coerce models into a list
m.rank <- fitList(
  'dtn.year.year.day'           = m1,
  'dtn.ptar.year.season'       = m2,
  'dtn.-.cam.season'           = m3,
  'disturb.year.year.season'   = m4,
  'disturb.ptar.cam.-'         = m5,
  'disturb.lem.-.season'       = m6,
  'slope.year.year.day'        = m7,
  'slope.-.cam.season'         = m8,
  'slope.lem.-.year'           = m9,
  'length.year.year.-'         = m10,
  'length.ptar.cam.season'     = m11,
  'length.lem.-.year'          = m12,
  'dtw.-.cam.day'              = m13,
  'dtw.ptar.-.year'           = m14,
  'dtw.lem.year.-'             = m15
)

# block 3: rank models in the list
modSel(m.rank)
# output

```

| | nPars | AIC | delta | AICwt | cumltvWt |
|--------------------------|-------|---------|-------|---------|----------|
| dtn.ptar.year.season | 26 | 1438.18 | 0.00 | 9.6e-01 | 0.96 |
| length.ptar.cam.season | 9 | 1444.84 | 6.66 | 3.4e-02 | 1.00 |
| slope.-.cam.season | 8 | 1451.28 | 13.10 | 1.4e-03 | 1.00 |
| disturb.lem.-.season | 8 | 1451.46 | 13.28 | 1.3e-03 | 1.00 |
| disturb.ptar.cam.- | 7 | 1453.20 | 15.03 | 5.3e-04 | 1.00 |
| dtn.-.cam.season | 25 | 1456.65 | 18.48 | 9.4e-05 | 1.00 |
| dtw.ptar.-.year | 25 | 1457.77 | 19.59 | 5.4e-05 | 1.00 |
| dtw.lem.year.- | 24 | 1460.51 | 22.34 | 1.4e-05 | 1.00 |
| dtw.-.cam.day | 8 | 1461.01 | 22.83 | 1.1e-05 | 1.00 |
| dtn.year.year.day | 43 | 1464.65 | 26.48 | 1.7e-06 | 1.00 |
| length.lem.-.year | 25 | 1464.81 | 26.63 | 1.6e-06 | 1.00 |
| disturb.year.year.season | 43 | 1465.91 | 27.74 | 9.1e-07 | 1.00 |
| slope.lem.-.year | 25 | 1465.94 | 27.76 | 9.0e-07 | 1.00 |
| length.year.year.- | 41 | 1471.69 | 33.51 | 5.1e-08 | 1.00 |
| slope.year.year.day | 43 | 1475.10 | 36.92 | 9.3e-09 | 1.00 |

The first block of code runs the fitting process for our candidate set, the second block organizes the models into a list with a designated title, and the last block produces our model selection output. Going from left to right in our model selection table, we list the names we provided for each model in the first column, followed by the number of parameters. Next we give the AIC score that is used to rank the models (lower score = better ranking based on parsimony), followed by the ‘delta’, or delta AIC. The delta is the AIC score of the respective model relative to the top (most parsimonious) model, and indicates the AIC distance between our top model and the set. Next we list the AIC weights that range between 0 and 1, and indicate the likelihood that the respective model is the best in the set. Our last column is the cumulative weight, which is the summation of our AIC weights.

From the set of a priori candidate models, our analysis indicates that the best ranked model included distance to nearest neighbor, ptarmigan abundance, year, and season (i.e., dtn.ptar.year.season). Moreover, the AIC weight (i.e., the relative likelihood of a model) of this model is 0.96 and is therefore likely the best model within the set of candidate models we proposed. In this model, distance to nearest neighbor best explained the probability of initial occupancy, whereas ptarmigan abundance best explained probability of colonization. Extinction probabilities varied among years, and detection probabilities varied according to the time of year (prelaying, incubation, and brood rearing).

We can inspect this model in more detail by examining the parameter estimates. We used a logit link to model each of the parameters using our covariates, so these estimates are on the log odds scale.

```
# model coefficients converted to probabilities
estimates <- round(coef(m2),2)
SE <- round(SE(m2),2)
(coef.table <- rbind(estimates,SE))

# output
      psi(Int) psi(dtn) col(Int) col(ptar) ext(Int)
estimates  1.15    2.08   -1.56    0.37   -2.44
SE         0.77    1.00    0.15    0.16    1.04
      ext(year1997) ext(year1998) ext(year1999)
estimates         1.97           1.12           2.09
SE                1.24           1.28           1.21
      ext(year2000) ext(year2001) ext(year2002)
estimates         1.89           2.18          -5.80
SE                1.30           1.38          39.77
      ext(year2003) ext(year2004) ext(year2005)
estimates         1.36           3.21           3.68
SE                1.32           1.21           1.54
```

| | ext(year2006) | ext(year2007) | ext(year2008) | | |
|-----------|---------------|---------------|---------------|--------|--|
| estimates | 4.08 | 2.53 | 2.05 | | |
| SE | 1.51 | 1.26 | 1.27 | | |
| | ext(year2009) | ext(year2010) | ext(year2011) | | |
| estimates | 1.43 | 0.39 | 1.12 | | |
| SE | 1.35 | 2.08 | 1.29 | | |
| | ext(year2012) | ext(year2013) | ext(year2014) | p(Int) | |
| estimates | 2.00 | 1.19 | -0.78 | 0.53 | |
| SE | 1.24 | 1.34 | 3.08 | 0.15 | |
| | p(seasonin) | p(seasonpl) | | | |
| estimates | 0.0 | -0.55 | | | |
| SE | 0.2 | 0.19 | | | |

We used effects parameterization in our models and each covariate was therefore fit with an intercept, and an effect or slope. The intercept indicates the log odds of occupancy at the reference level. The reference level is the mean for continuous variables, and is determined either alphabetically or numerically for categorical variables (unless stated otherwise using the `relevel` function). For example, if we look at our parameter `Psi`, we can see that `psi(Int)`, or the intercept for that parameter was 1.15. Distance to the nearest neighbor (`dtn`) was the variable used for `psi`, and because it is continuous, the intercept indicates that the log odds for initial occupancy was 1.15 at `dtn`'s mean. Next we have `psi(dtn)`, which is the slope of the straight line that describes the relationship between log odds of initial occupancy, and the distance to the nearest neighbor covariate. Log odds is difficult to interpret, so we can use the `plogis` function for a quick conversion to the probability scale. We can see below that when the `psi(int)` is converted, the intercept of `Psi` on the probability scale is 0.76. In other words, sites located 1774 m (the mean for `dtn`) away from the nearest breeding Gyrfalcon had a 76% chance of being occupied in the initial year of occupancy (see Fig. 10.3).

```
# convert dtn(int) from log odds to probability
plogis(1.15)
# output
0.76
```

Recall that `dtn` and `ptarmigan` abundance were both standardized to improve the model fitting process. Therefore, to begin making conclusions from our parameter estimates, we must first backtransform the covariates to their original scale. Additionally, visual representation of the parameter estimates may provide us with a more intuitive depiction of how our covariates relate to the respective parameters.

Starting with initial occupancy (ψ_1), we begin by creating a data frame that contains the range of values for our scaled dtn variable. We then use the `predict` function to add the predicted occupancy probabilities across the range of dtn. Finally, we convert the standardized dtn values to their original values and plot the relationship.

```
# visualize psi across distance to neighbor
nd    <- data.frame(dtn = seq(-1.5,2.1,length = 20))
E.psi <- predict(m2, type = "psi", newdata = nd,
                appendData = TRUE)

# convert distance to neighbor to original scale
E.psi$dtn <- E.psi$dtn * (sd(dtn)) + mean(dtn)

# plot Psi across distance to neighbor
graphics.off()
windows(7,5)
par(mar = c(4,4.8,0.7,2.1),mgp = c(1.8,0.2,0))

with(E.psi,
{
  plot(dtn, Predicted, ylim = c(0,1.1), type = "n",
       xlab = "Distance to Neighbour (m)",
       ylab = expression(hat(psi)),las = 1,
       xaxt = "n",yaxt = "n", bty = "l",font.lab = 2)
  lines(dtn, Predicted,lwd = 1)
  lines(dtn, Predicted + 1.96*SE, col = rgb(0,0,0,0.4))
  lines(dtn, Predicted - 1.96*SE,col = rgb(0,0,0,0.4))
  axis(1, at = seq(500,3500,500), labels =
        seq(500,3500,500) ,tck = 0.02,cex.axis = 0.75)
  axis(2, at = seq(0,1,0.1), labels = seq(0,1,0.1),
        las = 1,tck = 0.02,cex.axis = 0.75)
})
```

The resulting output (Fig. 10.3) indicates that the probability of occupancy at a site is relatively low if there is another occupied site nearby. Occupancy probability increased with nearest neighbor distance, and becomes asymptotic between 2,000 and 2,500 m.

Because ptarmigan abundance was also standardized, we can use the code above to plot colonization probability. Although more subtle, we see a positive relationship between ptarmigan and colonization probability (Fig. 10.4).

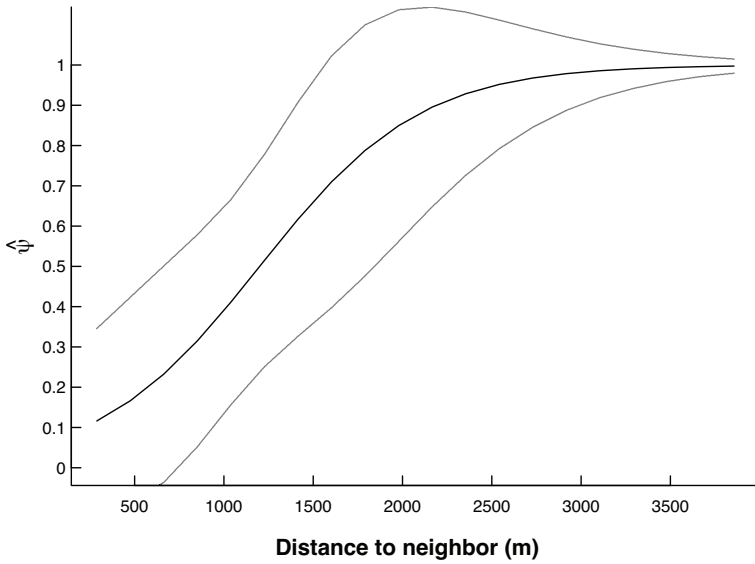


Figure 10.3 A visualization of how initial occupancy probability changes across the range of our 'territory proximity' variable in our top model. In the first year of this study, the close proximity of occupied territories negatively influenced occupancy.

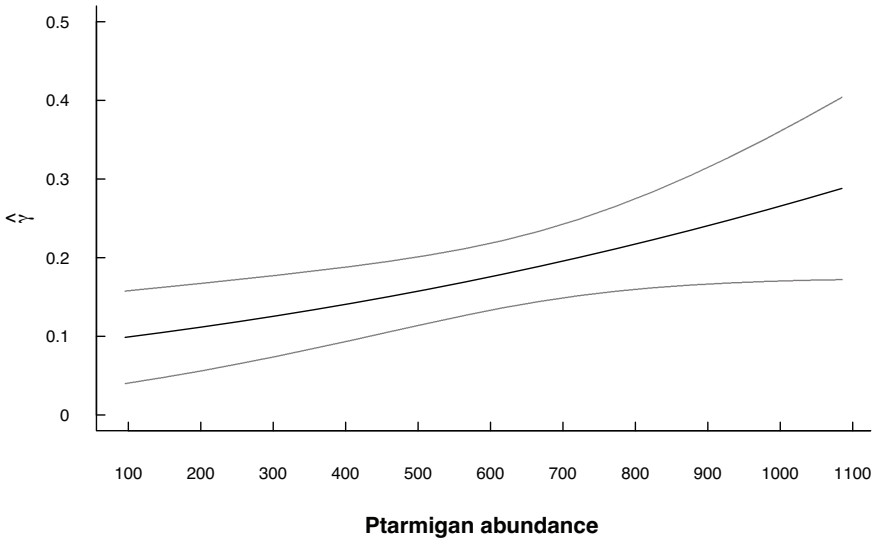


Figure 10.4 A visualization of how increasing ptarmigan abundance influences colonization probability.

Visualizing initial occupancy probability and colonization within the context of their respective covariates is useful, but our objective was to obtain some inference about population status by determining occupancy across the 20 years for which data were collected. We can approach this by first plotting colonization and extinction rates across that time period.

```
# plot extinction and colonization probabilities across year
# new data frame to store extinction/colonization probs
nd <- data.frame(year = c('year1997' , 'year1998' ,
  'year1999' , 'year2000' , 'year2001' , 'year2002' ,
  'year2003' , 'year2004' , 'year2005' , 'year2006' ,
  'year2007' , 'year2008' , 'year2009' , 'year2010' ,
  'year2011' , 'year2012' , 'year2013' , 'year2014' ,
  'year2015'))
# use predict function get extinction, colonization, and
# survival probs across year
E.ext <- predict(m1, type = 'ext', newdata = nd)
E.col <- predict(m1, type = 'col', newdata = nd)
# not plotted, this is how we get survival probs
E.surv <- 1-E.ext

#plot colonization and extinction across years
graphics.off()
windows(7,6)
par(mar=c(4,4.8,0.7,2.1),mfrow=c(2,1),mfg=c(1.5,0.2,0))

with(E.ext,
{
  plot(1:19, Predicted, pch =1, xaxt = 'n', xlab = 'Year',
    ylab = expression(paste('Extinction Prob (',
      'epsilon,' )')),las = 1,
    ylim = c(0,1),font.lab = 2,tck =0.03,
    cex.axis = 0.75)
  axis(1, at = 1:19, labels = seq(1997,2015,1),
    tck = 0.03 , cex.axis = 0.75)
  lines(1:19, Predicted, col=rgb(0,0,0,0.8),lty=2,lwd=2)
  arrows(1:19, lower, 1:19, upper, code=3, angle=90,
    length=0.03, col=rgb(0,0,0,0.2))
})

with(E.col,
{
```

```

plot(1:19, Predicted, pch = 1, xaxt = 'n', xlab = 'Year',
     ylab = expression(paste('Colonization Prob (',
                              '\gamma, ' )')),
     ylim = c(0,1), las = 1, font.lab = 2, tck = 0.03,
     cex.axis = 0.75)
lines(1:19, Predicted, col = rgb(0,0,0,0.8),
      lty = 2, lwd = 2)
axis(1, at = 1:19, labels = seq(1997,2015,1),
     tck = 0.03, cex.axis = 0.75)
arrows(1:19, lower, 1:19, upper, code = 3, angle = 90,
       length = 0.03, col = rgb(0,0,0,0.2))
})

```

The resulting output (Fig. 10.5) displays the variation in colonization and extinction by year, and gives us a better idea of dynamics as they are driven by these two parameters.

To depict yearly occupancy in this population, we project the predicted proportion of occupied sites across all years (Ψ_t) using our colonization, extinction, and occupancy probabilities at time $t-1$ using the following equation from MacKenzie et al. (2003):

$$\hat{\Psi}_t = \hat{\Psi}_{t-1} (1 - \hat{\epsilon}_{t-1}) + (1 - \hat{\Psi}_{t-1}) \hat{\gamma}_{t-1}$$

where $\hat{\Psi}_t$ is the estimated occupancy probability for a typical site at time t , $\hat{\epsilon}_{t-1}$ is the estimated probability that an occupied site at time $t-1$ becomes unoccupied at time t , and $\hat{\gamma}_t$ is the estimated probability that an unoccupied site at time $t-1$ becomes occupied at time t . The estimated occupancy probability at each time step ($\hat{\Psi}_t$) was automatically calculated during model fitting, and each of the values for $\hat{\Psi}_t$ were stored within our model object and can be called using '*modelname@smoothed*' (in our case we could code this as `m2@smoothed`). We also want to estimate the error around $\hat{\Psi}_t$, and use bootstrapping to resample our model for this purpose (Davison and Hinkley 1997). Finally, we want to plot $\hat{\Psi}_t$ along with its estimated error. To assess the difference between our observations and the estimation of $\hat{\Psi}_t$, we've included the proportion of sites occupied based on our surveys alone (Fig. 10.5).

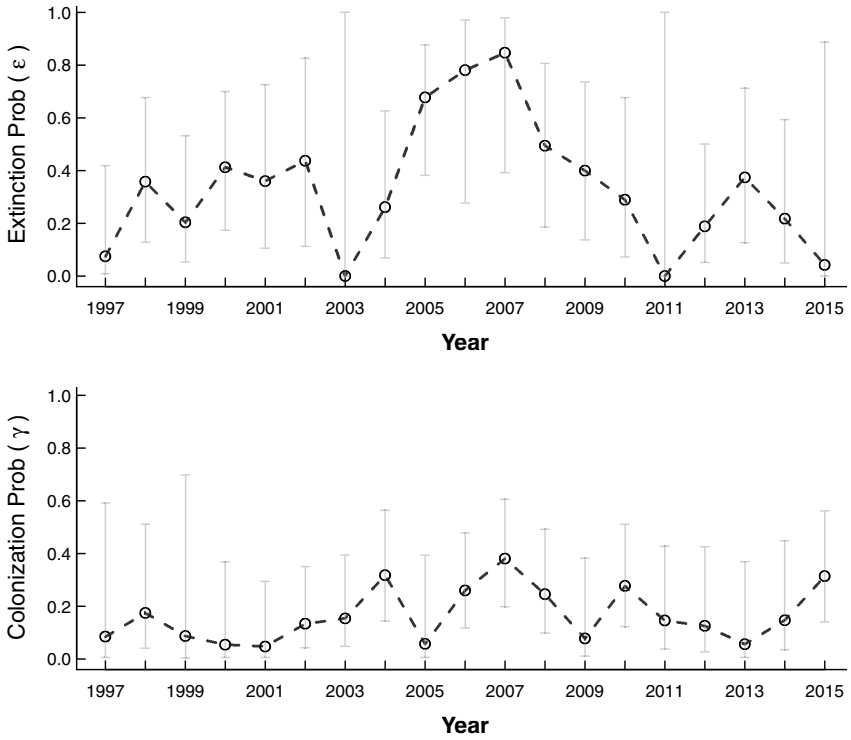


Figure 10.5 Colonization and extinction probabilities in the sampled population over the course of twenty years.

```

# visualize difference in estimated and observed occupancy
# observed occupancy does not account for detection error
det          <- detHist(m2)
det          <- as.data.frame(det[["out.freqs"]])
observed.props <- det$detected/det$sampled

# estimated occupancy accounts for detection error
#time consuming operation! decrease B if short on time
m2b         <- nonparboot(m2, B = 10)
PAO.boot    <- as.data.frame(cbind(
  smoothed = smoothed(m2b)[2,],
  SE = m2b$smoothed.mean.bsse[2,]))
upper      <- PAO.boot$smoothed + 1.96 * PAO.boot$SE
lower      <- PAO.boot$smoothed - 1.96 * PAO.boot$SE
PAO.boot    <- as.data.frame(cbind(PAO.boot,upper,lower))
years      <- seq(1996,2015,1)

# plot the observed vs. estimated occupancy
graphics.off()

windows(7,5)
par(mar = c(4,4.8,0.7,2.1),mgp = c(1.8,0.2,0))

plot(1996:2015,PAO.boot$smoothed, ylim = c(0,1),
  las = 1,pch = 16,font.lab = 2,
  xlab = "Year", ylab = "Proportion of Occupied Sites",
  bty = "l",xaxt = "n",yaxt = "n")
lines(1996:2015,PAO.boot$smoothed,lwd = 1)
points(1996:2015, observed.props,pch = 1)
lines(1996:2015, observed.props,lwd = 1, lty = 2)
arrows(1996:2015, PAO.boot$upper, 1996:2015,
  PAO.boot$lower, code = 3, angle = 90,
  length = 0.03,col = rgb(0,0,0,0.3))
axis(1, at = seq(1996,2015,2) ,
  labels = seq(1996,2015,2),tck = 0.02,
  cex.axis = 0.75)
axis(2, at = seq(0,1,0.1),
  labels = seq(0,1,0.1),las = 1,tck = 0.02,
  cex.axis = 0.75)
legend("topright", inset = 0.05,
  legend = c("Estimated","Observed"),bty = "n",
  lty = c(1,2), pch = c(16,1),cex = 0.75)

```

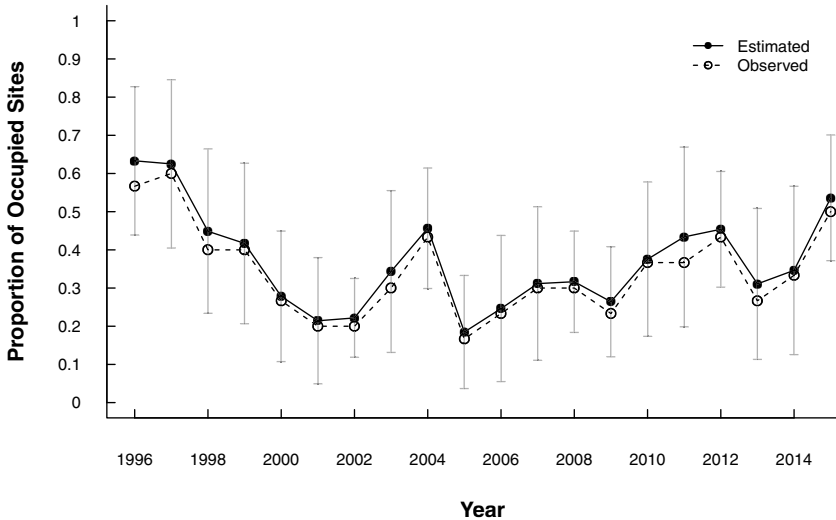


Figure 10.6 Estimated probability of occupancy plotted against the observed proportion of occupied sites. We can see that our observations underestimate the true occupancy due to detection error.

As we can see, our estimated occupancy probability matches our observed occupancy quite well (Fig. 10.6). However, throughout the entire range of years, our estimated occupancy is higher than our observed because of the detection error estimated in our model. In other words, determining occupancy without accounting for detection error leads to an underestimation in this example, and provides evidence for the importance of including the detection sub-model.

As a final step, we can characterize the overall trend in occupancy at the population level using the estimated annual occupancy probabilities to determine the average rate of change in our population as defined by MacKenzie et al. (2003). We do this by first calculating the change in occupancy (λ) between successive study years using the following equation, and then calculating the average and standard error:

$$\lambda_t = \frac{\hat{\Psi}_{t+1}}{\hat{\Psi}_t}$$

where λ_t equals the change in occupancy at time t , $\hat{\Psi}_{t+1}$ equals the estimate of occupancy at time $t+1$, and $\hat{\Psi}_t$ equals the estimate of occupancy at time t .

```

# extract lambda across all years.
PAO <- PAO.boot$smoothed
lambda <- rep(NA, times = (length(PAO)-1))
for(i in 1:(length(PAO )-1)){
  lambda[i] <- PAO[i+1]/PAO[i]
}

# average lambda and standard error.
mean(lambda)
(SE <- sd(lambda)/sqrt(length(lambda)))

# output
mean(lambda)
[1] 1.042452
(SE <- sd(lambda)/sqrt(length(lambda)))
[1] 0.07260853

```

Noting that a mean $\lambda < 1$ indicates population contraction and >1 indicates expansion, we can see that our population is stable at $\lambda = 1.04 \pm 0.07$ (standard error).

10.5 Conclusions

As demonstrated in this chapter, this analysis provides us with the tools to assess important drivers of yearly occupancy, and to more accurately assess long-term status of occupancy at the population-level while accounting for imperfect detection. Although the estimation of detection probability is dependent on an increase in survey effort, we feel that it is an important consideration that produces results that are more representative of true occupancy. Furthermore, it enables the investigation of factors that affect survey accuracy, and therefore the adjustment of survey protocols that maximize occupancy detection.

Literature cited

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19:716–723.
- Bailey, L. L., D. I. Mackenzie, and J. D. Nichols. 2014. Advances and applications of occupancy models. *Methods in Ecology and Evolution* 5:1269–1279.
- Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach*, Second Edition. Springer-Verlag, New York, New York, USA.
- Davison, A. C., and D. V Hinkley. 1997. *Bootstrap Methods and their Application*. Page Technometrics. Cambridge University Press, Cambridge, UK.
- Fiske, I., and R. Chandler. 2011. {unmarked}: An {R} Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software* 43:1–23.
- Hanski, I. 1998. Metapopulation dynamics. *Nature* 396:41–49.
- Hurvich, C. M., and C. L. Tsai. 1989. Regression and time-series model selection in small sample sizes. *Biometrika* 76:297–307.
- Kendall, W. L., J. E. Hines, J. D. Nichols, and E. H. C. Grant. 2013. Relaxing the closure assumption in occupancy models: Staggered arrival and departure times. *Ecology* 94:610–617.
- Kéry, M., and B. R. Schmidt. 2008. Imperfect detection and its consequences for monitoring for conservation. *Community Ecology* 9:207–216.
- MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* 84:2200–2207.
- Marsh, D., and P. Trenham. 2008. Current trends in plant and animal population monitoring. *Conservation Biology* 22:647–655.
- Miller, D. A., J. D. Nichols, B. T. McClintock, E. H. C. Grant, L. L. Bailey, and L. A. Weir. 2011. Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology* 92:1422–1428.
- Otto, C. R. V., L. L. Bailey, and G. J. Roloff. 2013. Improving species occupancy estimation when sampling violates the closure assumption. *Ecography* 36:1299–1309.
- R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rota, C. T., R. J. Fletcher, R. M. Dorazio, and M. G. Betts. 2009. Occupancy estimation and the closure assumption. *Journal of Applied Ecology* 46:1173–1181.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations: modeling, estimation, and decision making. Page Academic Press, San Diego, California, USA.