CHAPTER 9

Estimating trends in ptarmigan numbers

Jenný Brynjarsdóttir and Ólafur K. Nielsen

9.1 Introduction

Gyrfalcons have specialized feeding habits and feed primarily on one or two species of ptarmigan (Nielsen 2003, Potapov and Sale 2005, Potapov 2011, Chapter 3 this volume). Gyrfalcon and ptarmigan have coupled predator-prey oscillations in some areas (Nielsen 1999) so it is essential for any study on Gyrfalcon populations to have an index that reflects ptarmigan abundance in the study area. Here we build on Chapter 8, which describes methods for estimating prey abundance, and offer a statistical method that can detect joint trends across multiple time series to create a quantitative representation of a joint abundance index. We acknowledge that there are many other methods for population abundance estimation, but a full treatment of these methods is beyond the scope of this chapter. Instead we present here a descriptive statistical approach that is non-parametric, such that no functional form is presumed for the trends.

Our example data were collected by one of us (ÓKN) as a part of a Gyrfalcon and prey population study (Nielsen 1986). The data are spring densities of territorial Rock Ptarmigan (*Lagopus muta*) males on six plots surveyed since 1981 in northeast Iceland (for methods see Nielsen 1996 and 2004). Ptarmigan numbers can be quantified using different methods, but we recommend that anyone starting afresh adopt Distance sampling (see Chapter 8, this volume). Count data like those collected in northeast

Iceland can be used for purposes other than to study predator-prey relationships. For instance, these data along with auxiliary ptarmigan data collected by the same field crew, including age ratios in spring and late summer, and regional hunting statistics (total harvest, age composition, number of hunters), have been used to model and reconstruct the Rock Ptarmigan population (Magnússon et al. 2004, Sturludóttir 2015). In this chapter we will focus on detecting time-trends in the count data.

9.2 Analysis

9.2.1 Data

We want to obtain an estimate of the changes in ptarmigan abundance over time. For the analysis we use Rock Ptarmigan densities from six plots in northeast Iceland (Nielsen et al. 2004 and unpubl. data). The layout of the data table is illustrated in Table 9.1, and Fig. 9.1 shows the six observed time-series. The average density of cocks varied between these six areas, e.g., the density was generally much higher at Holl than Hafurstadir. However, there was clearly a common temporal trend. The six time-series oscillated in synchrony over time, although the magnitude of the oscillations varied. It is therefore reasonable to treat these six areas as observations of a single population of Rock Ptarmigan on the Gyrfalcon study area in northeast Iceland (Nielsen 2011). Further, a logarithmic transformation of the data stabilizes the variation in oscillation magnitudes (Fig. 9.2), so on a logarithmic scale the time series can be assumed to have the same temporal trend but shifted up or down depending on area. In the following we refer to data transformed by the natural logarithm as "log-transformed" data.

### 9.2.2 Generalized additive modelling of count data

Let Y_{it} be the observed density of cocks (number per km^2) in area i and year t. We consider the following model for our data:

(1) $\quad \ln(Y_{it}) = \alpha_i + s(t) + \varepsilon_{it} \qquad i = 1, \ldots, 6, \quad t = 1981, \ldots, 2016.$

The α_i term is a site effect, representing the difference in average log-transformed density numbers between areas, $s(t)$ is a common trend function over time, and ε_{it} is a random error term. Our main objective is to estimate the $s(t)$ function, a function that represents the overall trend of Rock Ptarmigan abundance in northeast Iceland. There are many options for what form the function can take; here we want a non-parametric curve that is neither too smooth (not informative enough) nor too wiggly (too informative to be predictive), as well as a measure of confidence shown as error bars.

Table 9.1 The first three years of the data table showing density of Rock Ptarmigan (*Lagopus muta*) males on six census plots in northeast Iceland 1981–2016. The data are stored as a comma separated values file (csv) where the rows represent observations and columns are the variables Area, Year, and observed density of male ptarmigans.

Area	Year	Cock Density
Birningsstadir	1981	4.211
Burfellshraun	1981	2.400
Hafursstadir	1981	1.625
Hofstadaheidi	1981	3.111
Holl	1981	7.083
Laxamyri	1981	4.865
Birningsstadir	1982	4.386
Burfellshraun	1982	4.800
Hafursstadir	1982	1.875
Hofstadaheidi	1982	4.444
Holl	1982	12.500
Laxamyri	1982	8.378
Birningsstadir	1983	3.684
Burfellshraun	1983	4.000
Hafursstadir	1983	4.125
Hofstadaheidi	1983	6.222
Holl	1983	19.583
Laxamyri	1983	12.703
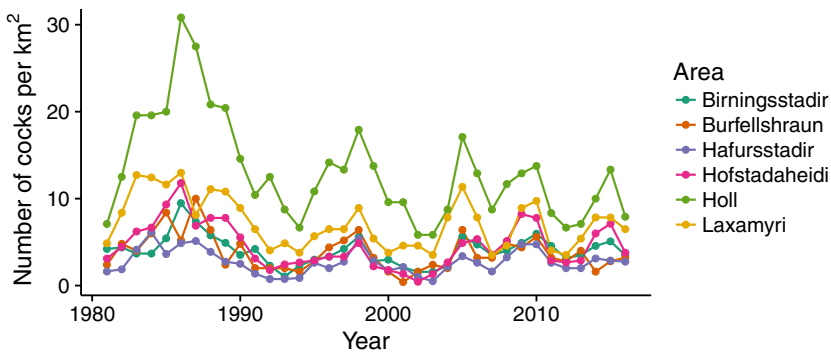
Figure 9.1 Density of Rock Ptarmigan (*Lagopus muta*) males on six census plots in northeast Iceland 1981–2016.
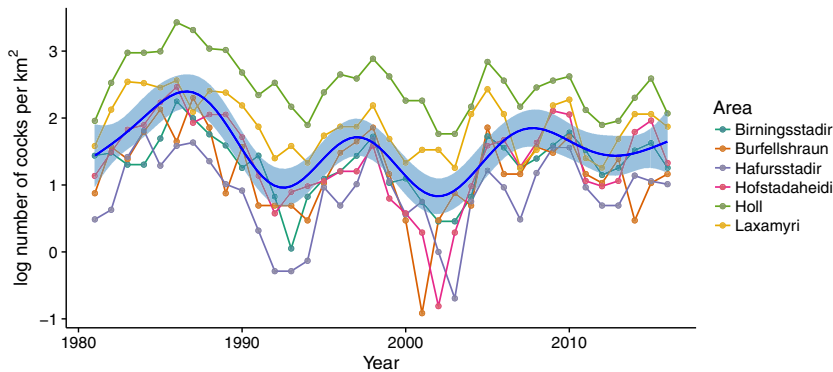
Figure 9.2 Log density of Rock Ptarmigan (*Lagopus muta*) males on six census plots in northeast Iceland 1981–2016. The fitted average trend curve $\hat{s}(b)$ is shown with a thick blue line and the shaded light blue area shows the 95% confidence bound.
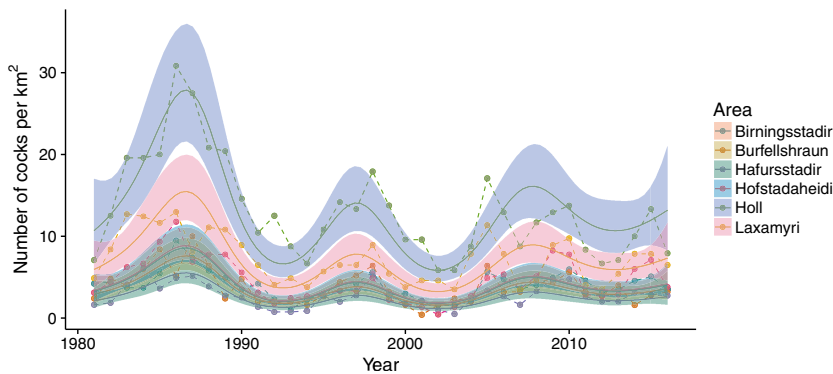
Figure 9.3 Fitted trend curves $e^{\hat{\alpha}_i + \hat{s}(t)}$ on the original scale for density of Rock Ptarmigan (*Lagopus muta*) males on six census plots in northeast Iceland 1981–2016. The shaded ribbons show the corresponding 95% confidence bounds.

Generalized additive models (GAMs) provide a flexible modeling framework for fitting non-parametric curves to data (see for example Hastie and Tibshirani 1990). A convenient route is to use the `gam()` function in the mgcv package (Wood 2000) in R (R Core Team 2016). This function can fit GAMs with different types of trend curves (smoothers) and can handle many different error distributions (i.e., not just the normal distribution). In addition, the `gam()` function automatically chooses the level of smoothness (degrees of freedom) in $\hat{s}(t)$, but the user can also manually

set the degree of smoothness. The sample R code below shows two possible strategies for analyzing our data:

1. log-transformed data, normal errors, and smoothness chosen automatically
2. log-transformed data, normal errors, but smoothness chosen manually by fixing the degrees of freedom of the trend curve.

```
# 1) fit gam to log transformed data
  Ptarmigan.gam <- gam(log(CockDensity) ~ factor(Area) +
      s(Year), data = Ptarmigan)

# plot the fitted s() curve
  plot(Ptarmigan.gam)

# estimated parameters, fitted values, and residuals
# output not shown
  coefficients(Ptarmigan.gam)
  fitted(Ptarmigan.gam)
  residuals(Ptarmigan.gam)

# summary table, output not shown
  summary(Ptarmigan.gam)

# obtain degrees of freedom for trend curve and aic
# output not shown
  summary(Ptarmigan.gam)$edf
  Ptarmigan.gam$aic

# 2) fix degrees of freedom for trend curve
  Ptarmigan.gam3 <- gam(log(CockDensity) ~ factor(Area) +
      s(Year, k = 17, fx = TRUE), data = Ptarmigan)
```

The `gam()` function returns a "gam object" in R that is similar to the object produced by the linear regression function `lm()`. We can, for example, access the estimated parameter values, fitted values and residuals as well as various information about the fit. The fitted curve (thick blue curve) for our log-transformed Rock Ptarmigan data (analysis 1 above) is shown in Fig. 9.2. Note that on the log scale, we assume that the trend curve is the same for all areas, only shifted up or down depending on the α_i parameter. When we transform back to the original scale the area specific α_i parameters allow for varying oscillation sizes between areas, as shown in Fig. 9.3. As shown in the code above, fitted curves can be easily plotted in R. The plots shown here, however, are made using the R package ggplot2 (Wickham 2009).

The **`gam()`** function in R automatically chooses the level of smoothness of $s(t)$ via cross-validation (other methods are also possible) which is convenient and produces reasonable results in most cases. Sometimes, however, the fitted trend curve does not capture all the desired features of our data. In our case, the fitted trend curve $\hat{s}(t)$ shown in Fig. 9.2 represents well the two major cycles in ptarmigan numbers from 1980–2002, but smooths over the next two (shorter) cycles. That is, although almost all data series exhibit a peak in 2005, 2010, and 2015, and a low in 2007 and 2012, the fitted trend curve only has one cycle over this same period with a peak around year 2007 (Figs. 9.2 and 9.3).

There is a simple statistical reason why the trend curve does not capture these short cycles in ptarmigan numbers. More oscillations (more "wiggles") in $s(t)$ require more degrees of freedom (df), which can be thought of as the number of parameters in the model. The fitted curve given by **`gam()`** is a trade-off between an overall goodness-of-fit criterion (e.g., mean squared prediction error) and a penalty term that increases with increasing degrees of freedom. The **`gam()`** function simply determined that the penalty for the extra degrees of freedom needed to represent the three cycles in 2005–2016 outweighed the effect of a better fit on the goodness-of-fit criterion. In our case we want a trend curve for ptarmigan abundance that can be used to reflect food conditions for Gyrfalcon and it is therefore necessary to include the different cycle frequencies of the two time periods.

The smoothness of the trend curve can be adjusted by fixing the degrees of freedom (df) for $s(t)$ (lower df gives a smoother curve), as shown in the R code above. The estimated df for the fitted curve was 8.855, so to make the curve less smooth we can set the df to a higher number. The challenge then is to choose a suitable value for the df. We fitted the model in equation (1) to our data with increasing df values and examined the fitted trend curve. For df values 16 and higher the curve exhibited the shorter cycles we wanted represented. To choose between df values that give similar looking trend curves we compared the Akaike Information Criterion (AIC, Akaike 1974), shown in Fig. 9.4, and found that AIC has a (local) minimum at 17 df. The fitted curve for 17 df is shown in Fig. 9.5.

```
# calculate AIC for models with different df
  dfs <- 5:25
  AIC <- c()
  for(i in 1:length(dfs)){
      tmp.gam.fit <- gam(log(CockDensity) ~
            factor(Area) + s(Year, k = dfs[i], fx = TRUE),
            data = Ptarmigan)
      AIC[i] <- tmp.gam.fit$aic
                        }
# plot AIC for each df
  plot(dfs, AIC)
```
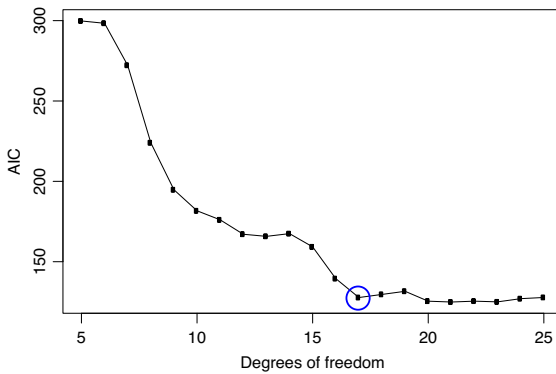
Figure 9.4 Akaike Information criterion (AIC) for different degrees of freedom for a trend curve calculated based on density of Rock Ptarmigan (*Lagopus muta*) males on six census plots in northeast Iceland 1981–2016.
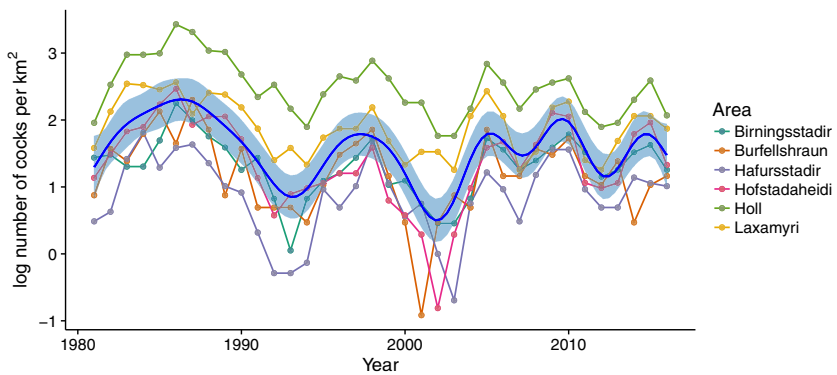
Figure 9.5 Log density of Rock Ptarmigan (*Lagopus muta*) males on six census plots in northeast Iceland 1981–2016 and a fitted average trend curve $\hat{s}(t)$ with 95% confidence bound (blue thick curve and the light blue ribbon around it). Here the degrees of freedom for $\hat{s}(t)$ were fixed to 17.

After fitting the GAM model in (1) we can provide an abundance index that may be helpful when modeling ptarmigan count data. The abundance index, with base year b, for each year t is calculated according to Fewster et al. (2000) as:

$$(2) \qquad \hat{I}(t) = \frac{\text{Expected density year } t}{\text{Expected density year } b} = \frac{\Sigma_i \exp(\hat{\alpha}_i)\exp(\hat{s}(t))}{\Sigma_i \exp(\hat{\alpha}_i)\exp(\hat{s}(b))} = \frac{\exp(\hat{s}(t))}{\exp(\hat{s}(b))}$$

The choice of base year b is arbitrary, but note that every index is a function of $\hat{s}(b)$ as well as $\hat{s}(t)$ so the uncertainty of every $\hat{I}(t)$ is influenced by the uncertainty of $\hat{s}(b)$. Therefore the year b should ideally be a year where we have the most data, since that makes the uncertainty in $\hat{s}(b)$ as low as possible.

```
# calculate the index
  # choose settings
    # years we want an index estimate
      PredYears <- 1981:2016
          # for better looking plots we can use
          # PredYears <- seq(1981,2016, by=0.2)

    # use first year as base year
      base <- 1

    # set the area, we chose Holl
      NewDat <- data.frame(Year = PredYears,
             Area = "Holl")

  # calculate index from the fitted trend curve
    s.term <- predict.gam(Ptarmigan.gam, newdata = NewDat,
        type = 'terms', terms = "s(Year)")
    index.hat <- exp(s.term - s.term[base])
```

In our case, we have six observations every year and choose the first year as a base year. An easy way to calculate the abundance index in R is to use the predicted values for one area to calculate $\hat{I}(t)$, as shown in the box above. The estimated index curve is shown in Fig. 9.6, both for the model chosen by the **gam()** function and the model that has 17 df for the trend curve.

The estimated abundance index as calculated in equation (2) can be used directly as an explanatory variable when modeling Gyrfalcon counts. However, it is important to keep in mind that using $\hat{I}(t)$ as a point estimate does not take into account the uncertainty of that estimate. We end this chapter by showing how we can calculate the error associated with $\hat{I}(t)$. Even though the error might be difficult to incorporate in a model for Gyr-

falcon counts, it is informative for the modeler to have information about the accuracy of the ptarmigan trend index. The error for the abundance index can be obtained via the mgcv package using Bayesian sampling methods. The posterior distribution of parameters can be approximated by a multivariate normal distribution whose mean vector and covariance matrix is given by the `gam()` function. By sampling from this posterior distribution, say N_{post} = 1000 times, we calculate the index in equation (2) for each sample to obtain 1000 samples from the posterior distribution of $I(t)$. These samples can then be used to obtain summaries of the posterior distribution of $I(t)$ (e.g., the posterior mean and 95% probability intervals (called credible intervals) for $I(t)$ for each year, as shown in Fig. 9.6). As an example of an extra insight gained from calculating error for our data, we notice in Fig. 9.6 that the 95% credible intervals for year 2005 (a peak) and 2007 (a low) overlap. This indicates that the indices for these two years are not significantly different. Example R code to obtain estimates and error for the population index is shown below. Note that we use the `mvrnorm()` function in the package MASS (Venables and Ripley 2002).

```
# uncertainty bounds for the index
  # set number of posterior samples
    nPost <- 1000

  # get nPost posterior samples of the index
    Xp <- predict(Ptarmigan.gam, newdata = NewDat,
        type = 'lpmatrix')
    nYears <- length(PredYears)
    PostCoef <- mvrnorm(n = nPost, coef(Ptarmigan.gam),
        Ptarmigan.gam$Vp)
    s.terms <- Xp %*% t(PostCoef)
    Index <- exp( s.terms - matrix(rep(s.terms[base, ],
        nYears), byrow = T, ncol = nPost) )

  # calculate posterior means and bounds for 95% CI
    IndexPostMean <- apply(Index, 1, mean)
    PostLB <- apply(Index, 1, quantile, prob = 0.025)
    PostUB <- apply(Index, 1, quantile, prob = 0.975)

  # build plot
    plot(PredYears, IndexPostMean, type = 'l',
        ylim=range(PostLB, PostUB))
    lines(PredYears, PostLB, type = 'l', lty = 2)
    lines(PredYears, PostUB, type = 'l', lty = 2)
    points(PredYears, index.hat)
```
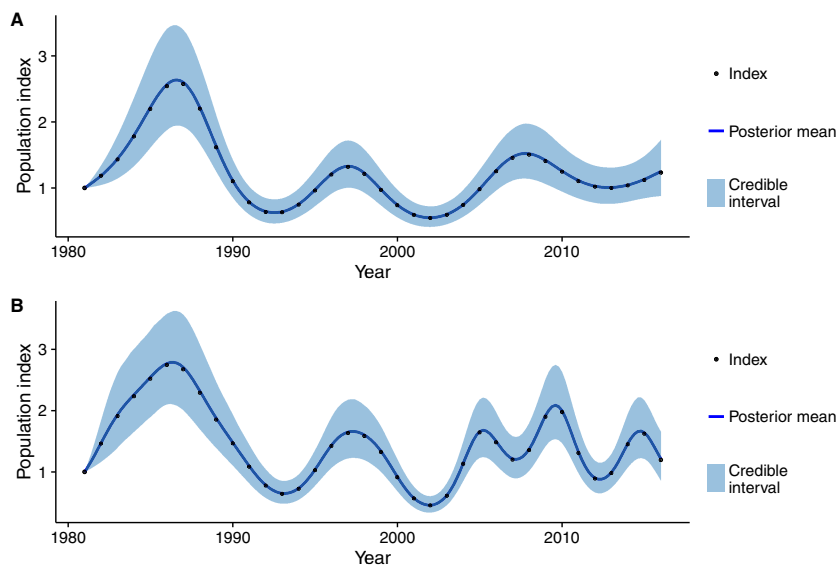
Figure 9.6 Estimated Rock Ptarmigan (*Lagopus muta*) population index based on counts on six census plots in northeast Iceland 1981–2016 with 95% confidence bound. Panel A shows results from the GAM model with automatic smoothness selection and Panel B shows results from setting the degrees of freedom to 17.

Literature cited

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control. AC-19:716–723.

Fewster, R. M., S. T. Buckland, G. M. Siriwardena, S. R. Baillie, and J. D. Wilson. 2000. Analysis of population trends for farmland birds using generalized additive models. Ecology 81:1970–1984.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. CRC Press, Boca Raton, Florida, USA.

Magnússon, K., J. Brynjarsdóttir, and Ó. K. Nielsen. 2004. Population cycles in Rock Ptarmigan *Lagopus muta*: modelling and parameter estimation. Science Institute University of Iceland, Reykjavík. RH-19-2004.

Nielsen, Ó. K. 1986. Population ecology of the Gyrfalcon in Iceland with comparative notes on the Merlin and the Raven. Ph.D. thesis, Cornell University, Ithaca, New York, USA.

Nielsen, Ó. K. 1996. Rock Ptarmigan censuses in northeast Iceland 1981 to 1994. Náttúrufræðingurinn 65:137–151. (In Icelandic with an English summary.)

Nielsen, Ó. K. 1999. Gyrfalcon predation on Ptarmigan: numerical and functional responses. Journal of Animal Ecology 68:1034–1050.

Nielsen, Ó. K. 2003. The impact of food availability on Gyrfalcon (*Falco rusticolus*) diet and timing of breeding. Pages 283–302 *in* D. B. A. Thompson, S. M. Redpath, A. H. Fielding, M. Marquiss, and C. A. Galbraith, editors. Birds of prey in a changing environment. Scottish Natural Heritage, Edinburgh, Scotland.

Nielsen, Ó. K. 2011. Gyrfalcon population and reproduction in relation to Rock Ptarmigan numbers in Iceland. Pages 21–48 *in* R. T. Watson, T. J. Cade, M. Fuller, G. Hunt, and E. Potapov, editors. Gyrfalcons and Ptarmigans in a Changing World, volume 2. The Peregrine Fund, Boise, Idaho, USA.

Nielsen, Ó. K., J. Brynjarsdóttir, and K. G. Magnússon. 2004. Monitoring of the Ptarmigan population in Iceland 1999–2003. Fjölrit Náttúrufræðistofnunar 47:1–110. (In Icelandic with an English summary.)

Potapov, E. 2011. Gyrfalcon diet: spatial and temporal variation. Pages 55–64 in R. T. Watson, T. J. Cade, M. Fuller, G. Hunt, and E. Potapov, editors. Gyrfalcons and Ptarmigan in a Changing World, volume 1. The Peregrine Fund, Boise, Idaho, USA.

Potapov, E., and R. Sale. 2005. The Gyrfalcon. T & A Poyser, London, UK.

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

Sturludóttir, E. 2015. Statistical analysis of trends in data from ecological monitoring. Ph.D. thesis, University of Iceland, Reykjavík.

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York, New York, USA.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. http://ggplot2.org.

Wood, S. N. 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62:413–428.