Modelling Birds Population

Ignacio Alvarez

Fall 2013

1 Introduction

According to (REF, Etterson) one of the basic research goals in Ecology consist on understand the distribution and abundance of the animal population. In this work, the particular goal will be to explore alternatives to model the time trends in Western Great Lakes Birds population over 1994 to 2011.

1.1 Data description

The data source for this study came from "an extensive, long-term monitoring program with over 1600 off-road sampling points designed to track regional population trends and investigate the response of forest birds to regional land use patterns"(REF:).

This monitoring program is carry out by Natural Resources Research Institute (University of Minnesota Duluth) with the objective of "sustain forest resources and bird diversity in western Great Lakes forests" (REF:).

For this report in particular the dataset which will be used in this report consist in the yearly bird count for 73 species on three National Forest from 1995 to 2013. There are several interesting covariate that are relavated in the sampling procedure, however most of them are site characteristics at the moment when the sampling was done, therefore they are not so meaningfully when we consider yearly aggregated data.

We can see a first look of the data in Figure 1, where the total bird count per year is plotted for all three forest. We can see Superior is consistently over all years the forest with more bird counts (why? is bigger ?) reaching 10000 counts on three time during the monitoring period. The others two forest, Chequamegon and Chippewa show a similar trend in the total bird count.

Overall, it seems to be a first sub-period from 1995 to 2003 where the total bird count is increasing every year but from that this increment stop. From 2003 on Superior forest it seem to oscillate around a little more than 8000 birds while for Chequamegon the count are getting smaller each year.

Figure 1: Raw trend in the data

There is a big variability of the counts among species. Table 1 presents the total bird count on year 2007 for the most abundant species, (just to compare it with the on line annual report). We can see that OVEN and REVI are very abundant species among all forest.

However the structure of abundant is different on each forest. Species like NAWA and WTSP are abundant only in Superior forest, while VEER it is only abundant for Chippewa. In order to model the species count along time it might be better to work separately by forest (at least in a initial steps).

Specie	Chequamegon	Chippewa	Superior
OVEN	1003	835	1168
REVI	823	997	771
BTNW	254	177	219
NAWA	240	348	867
BLJA	222	199	230
CSWA	211	330	375
RBGR	208	70	181
WTSP	180	387	940
HETH	175	249	265
AMRO	156	102	154
LEFL	155	368	120
YBSA	132	150	129
COYE	126	209	71
VEER	91	402	264
WIWR	48	50	195

Table 1: Total bird count on 2007 for the most abundant species. Values are ordered according the count on Chequamegon forest.

1.2 Initial models

As a starting point we fit quadratic regression model separately for each specie and forest. Equation 1 represent one of the regressions, for one species in one forest.

$$\begin{aligned} Y_{tfs} &= \beta_{0fs} + \beta_{1fs} t + \beta_{2fs} t^2 + \epsilon_{tfs} \\ \epsilon_{tfs} &\sim N(0, \sigma^2_{fs}) \end{aligned} \quad (1)$$

where Y_{tfs} represent the average bird count on year t in the forest f for the specie s. There are 73 species and 3 forest in the data set so there are 219 models in total. We can use the results of these models to get idea of the distribution for intercept, slope and quadratic term on each forest.

Using R function `density` we get an estimate for each parameter based on 73 observations, these are shown on Figure 2. There are 12 panels, each row represent one of the coefficient (intercept, slope, quadratic term and variance) and each column represent one forest.

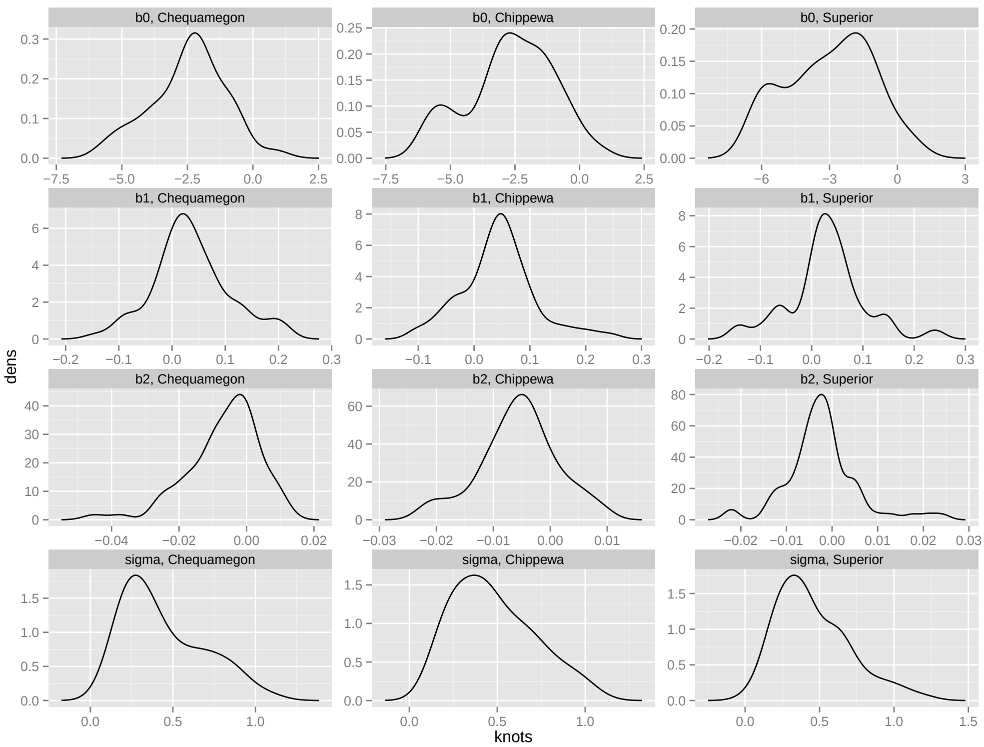

Figure 2: Densities for coefficients of model 1.

The distribution fore each coefficient is similar on the 3 forest. The Intercept is mostly on negative values, as the year covariate is centered on 2000, this represent the expected count in that year which is less than 1 bird for most of the species. The slope densities show the slope is centered arround 0 for most of the species but with some values far from it on each forest.

In the quadratic term the densities show some differences among forest. There is a left skewed in Chequamegon, relatively simeric on Chippewa and right skewed in Superior forest (meaning ????).

Table **??** shows the summary statistics for each coefficient and figure **??** presents histograms for each one of the model coefficients.

The next step would be to explore the bivariate relation among the coefficients in the model. With this in mind we can see a sccatter matrix of the estimated coefficients in Figure **??**. The

forestN	parameter	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Chequamegon	b0	-5.797	-3.453	-2.328	-2.463	-1.691	1.009
Chequamegon	b1	-0.143	-0.004	0.030	0.036	0.072	0.211
Chequamegon	b2	-0.045	-0.011	-0.004	-0.007	0.000	0.012
Chequamegon	sigma	0.110	0.250	0.364	0.443	0.625	1.099
Chippewa	b0	-5.859	-3.474	-2.637	-2.658	-1.507	0.741
Chippewa	b1	-0.104	0.012	0.048	0.044	0.077	0.245
Chippewa	b2	-0.023	-0.010	-0.005	-0.006	-0.002	0.010
Chippewa	sigma	0.127	0.294	0.451	0.482	0.646	1.057
Superior	b0	-6.177	-4.310	-2.846	-3.005	-1.567	0.815
Superior	b1	-0.150	0.000	0.028	0.028	0.061	0.248
Superior	b2	-0.022	-0.006	-0.003	-0.002	0.000	0.024
Superior	sigma	0.033	0.270	0.392	0.458	0.615	1.199

main point to see here is a negative relation between the slope and the quadratic term for all forest, is .67, .82 and .83 on Chequamegon, Chippewa and Superior respectively. (KAISER dixit: while polynomials are wonderfully flexible functions for describing data patterns, they do not lend themselves to interpretation of coefficient values -different combinations of constant, linear, and quadratic terms can lead to quite similar functions over a finite range of covariate values)

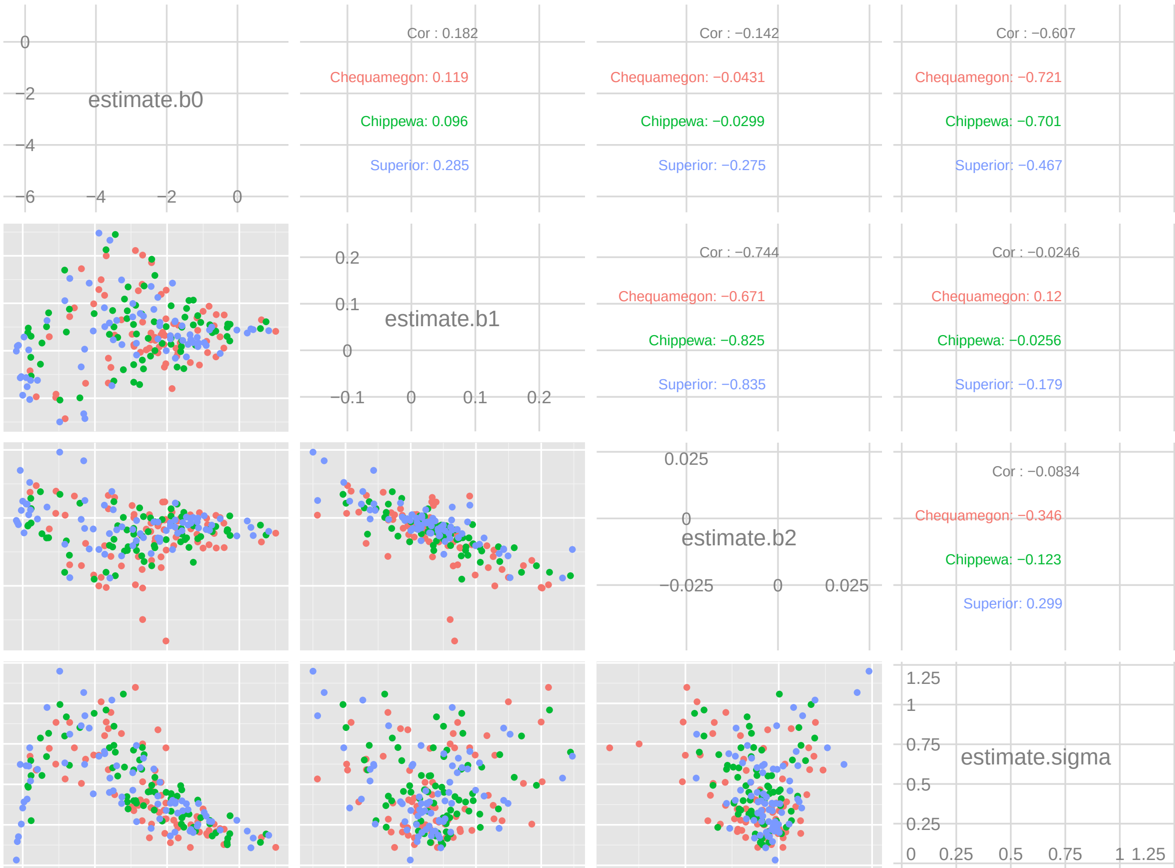

Figure 3: Bivarieate relation among coefficients of model 1.

Also, there is a negative relation between between intercept and variance, specially for Chequamegon and Chippewa forest. Maybe the main message for this plot is that we should not treat this parameters as independent and try to include their dependence into the model we are fitting.

A second model considered is a regression using data from all 73 species in each forest and including random terms for the species coefficients.

$$
\begin{aligned}
log(Y_{tfs}) &= \beta_{0fs} + \beta_{1fs}t + \beta_{2fs}t^2 + \epsilon_{tfs} \\
\beta_{fs} &\sim N\left(\begin{bmatrix} \beta_{0fs} \\ \beta_{1fs} \\ \beta_{2fs} \end{bmatrix}, \Sigma_f\right) \qquad (2) \\
\epsilon_{tfs} &\sim N(0, \sigma_f^2) \qquad (3)
\end{aligned}
$$

where the matrix Σ is fomrmed as $(\Sigma)_{ij} = \rho_{ij}\sigma_i\sigma_j$ when $i \neq j$ and $(\Sigma)_{ii} = \sigma_i^2$, having 6 parameter in total.

Forest	b0	b1	b2
Chequamegon	-2.463	0.036	-0.007
Chippewa	-2.658	0.044	-0.006
Superior	-3.005	0.028	-0.002

Forest	sd0	sd1	sd2
Chequamegon	1.418	0.058	0.008
Chippewa	1.653	0.050	0.005
Superior	1.906	0.061	0.006

1	2	3
Chequamegon	Chippewa	Superior
0.5078357	0.5346165	0.5191165

Forest	rho_01	rho_02	rho_12
Chequamegon	0.131	0.001	-0.610
Chippewa	0.122	-0.006	-0.803
Superior	0.344	-0.313	-0.827

2 Statistical Model

$$\begin{aligned}
log(Y_{tfs}) &= \beta_{0fs} + \beta_{1fs}t + \beta_{2fs}t^2 + \epsilon_{tfs} \\
\beta_{fs} &= \begin{pmatrix} \beta_{0fs} \\ \beta_{1fs} \\ \beta_{2fs} \end{pmatrix} \sim N(0, \Sigma_{fs}) \\
\Sigma_{fs} &\sim \text{inv-gamma, scaled inv-gamma, ...} \\
\epsilon_{tfs} &\sim N(0, \sigma_\epsilon^2) \\
\sigma_\epsilon^2 &\sim inv - gama(\alpha, \gamma)
\end{aligned} \tag{4}$$