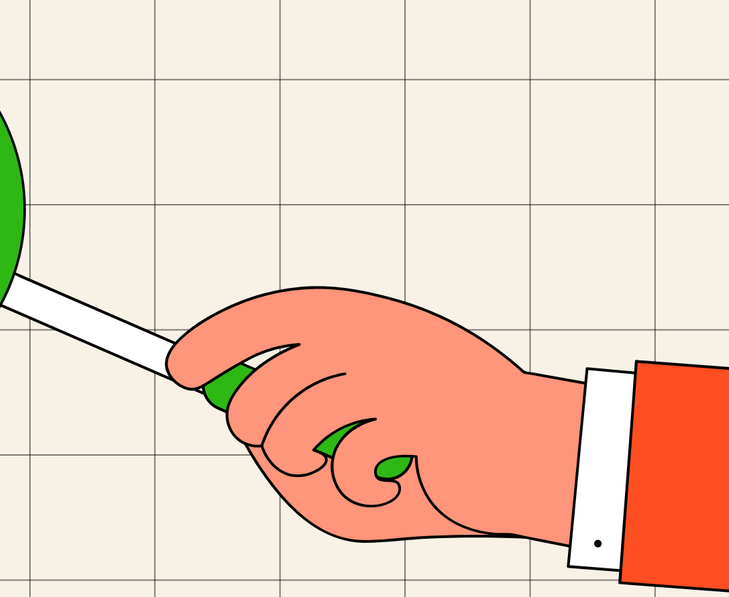
LEARNING GLARMA



Delores Chan

Jaden Husser

Generalized Linear Autoregressive Moving Average Models

WHAT DOES THE PACKAGE DO?

GLARMA has many uses beyond population ecology, but for the purposes of ESM 211, it's great for:

Fitting Models for Observation Driven Count Series Data

In short, that means: time series data that is not normal!

The GLARMA package takes data and models changes in population against explanatory variables. As we have learned, populations over time are often not linear or normally distributed.

What types of problems can you solve?

With the Package

- Not just populations, observation driven data
- Not normally distributed data (accounts for variations and environmental influences)
  - Can be crime, disease etc.
- Best fit model that accounts for variability

With the Models

- Generalized Linear Autoregressive Moving Average Models
  - Poisson ⭐
  - Negative Binomial
  - Binomial

Model Overview

Explanatory variables

Initial value of the regression coefficient

$$\log \mu_t = \mathbf{x}_t^{\mathrm{T}} \boldsymbol{\beta} + \nu_t + \sum_{i=1}^{\infty} \psi_i e_{t-i}.$$

Log of the Poisson Distribution

Linear Regression

Lag: information from previous observation

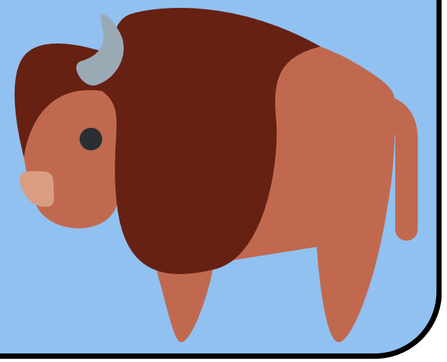
Types of Data Needed

1. Observational Count Data

One vector:

- Observational population count

*Time needed for plots



2. Explanatory Variables

Intercept:

- Value of 1, but needs to be a matrix

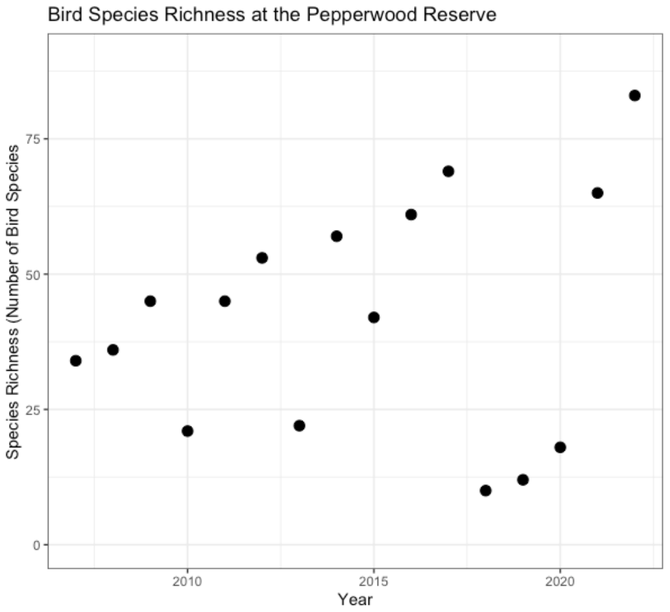
Major events:

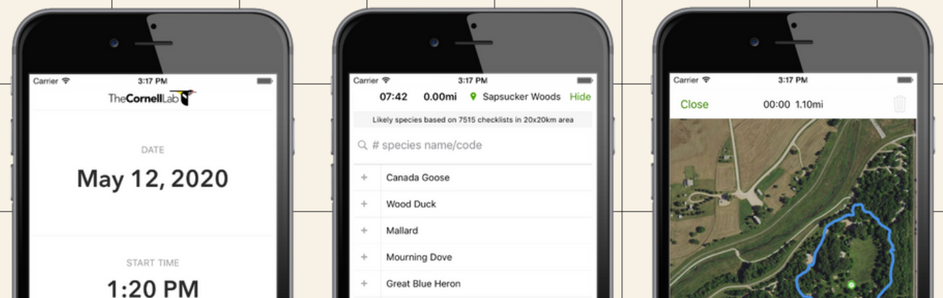
- Technically optional, if you want to compare model factoring in events or not
- Also a matrix

Ex: Bison Data

	year	bison	intercept	major_events
12	1981	2397	1	0
13	1982	2245	1	0
14	1983	2157	1	0
15	1984	2259	1	0
16	1985	2465	1	0
17	1986	2776	1	0
18	1987	2976	1	0
19	1988	3255	1	1
20	1989	2716	1	0
21	1990	3109	1	0
22	1991	3426	1	0
23	1992	3357	1	0
24	1993	3329	1	0
25	1994	4114	1	0
26	1995	3928	1	1
27	1996	3584	1	0
28	1997	2797	1	0

Data and Problem Example





Pepperwood Preserve, Santa Rosa, CA

Model and Code Example

```r
#GLARMA example
#CL Jerde
################

library(glarma)
library(here) # makes a local directory for the R project
library(tidyverse)

#imports eBird species richness data for Pepperwood Preserve
SR<-read.csv(here("ESM_211", "Portfolio Assignments", "Glarma", "Data","SR_df.csv"))

Y<-SR$S # makes the response variable only the species richness
X.1<-SR %>% select(intercept, dist) #makes model 1's explanatory variables of the intercept (1)
#and presence absence of a fire event
X.1<-as.matrix(X.1) #makes the data frame into matrix


X.0<-SR %>% select(intercept) #makes a null model without fire information
X.0<-as.matrix(X.0) #makes the data frame into a matrix
```

- Brings in data
- Assigns vectors
- Creates matrices

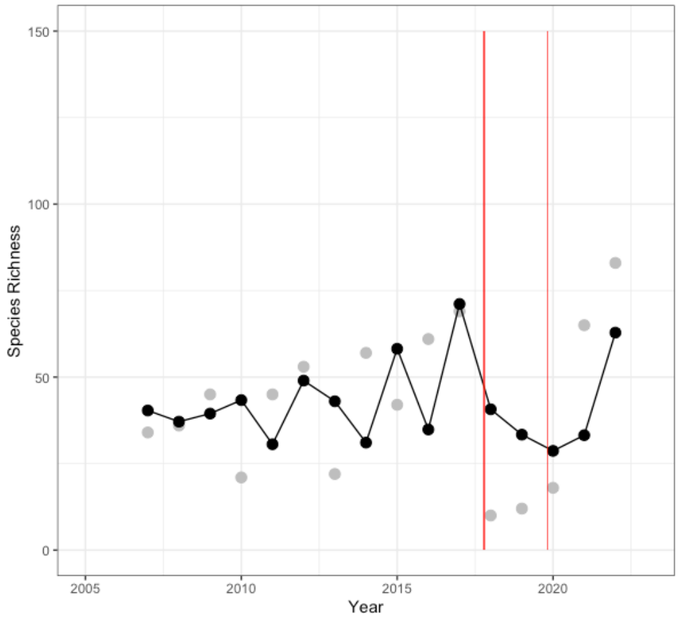
```r

# null model without fire but with a 1 time step lag
glarmamod.0 <- glarma(Y, X.0, phiLags = c(1), type = "Poi", method = "FS",
                      residuals = "Pearson", maxit = 100, grad = 1e-6)

# model with fire events included
glarmamod.1 <- glarma(Y, X.1, phiLags = c(1), type = "Poi", method = "FS",
                      residuals = "Pearson", maxit = 100, grad = 1e-6)

#summary statistics just like getting information out of linear regression
summary(glarmamod.0)
summary(glarmamod.1)

```

- Glarma null model using the intercept
- Glarma explanatory model
- Summary of the two models

```r
#### Modified SR plot for Pepperwood
SR_mod<-SR

SR_mod$est<-glarmamod.1$fitted.values #adds the fitted values of the Best fit GLARMA model to the data

#plots the model (black) and the species richness data (gray)
SR_plot_mod<-ggplot(SR_mod, aes(x=year)) +
  geom_point(col="gray",aes(y=S),size=3) +
  geom_point(col="black",aes(y=est),size=3) +
  geom_line(col="black",aes(y=est)) +
  labs(x="Year", y="Species Richness") +
  ylim(0,150) + xlim(2005,2023)+
  theme_bw()

#Adds two red horizontal lines for the occurance and duration of the fires
SR_plot_mod<-SR_plot_mod +
  annotate("rect", xmin = 2017.767, xmax =2017.833 , ymin = 0, ymax = 150, alpha = .75,fill = "red") +
  annotate("rect", xmin = 2019.808, xmax =2019.855 , ymin = 0, ymax = 150, alpha = .75,fill = "red")


SR_plot_mod

```

- Plotting the model with the explanatory variables

Output and Inference Example




```
GLARMA Coefficients:
     phi_1
0.08332709

Linear Model Coefficients:
intercept       dist
3.6975915  0.2102572

Degrees of Freedom: 15 Total (i.e. Null);  13 Residual
Null Deviance: 181.4732
Residual Deviance: 154.2258
AIC: 251.6174
```

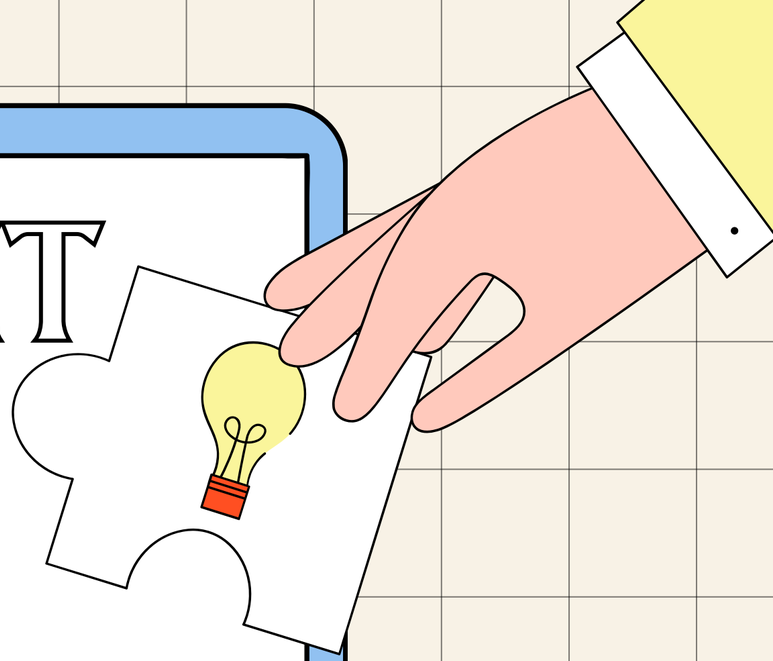
the variable you chose

Good for comparing models best fit

UNIQUE INSIGHT

FORECAST VS GLARMA

Want to learn more? Check out the forecast vignette!



01 Both are autoregressive: predicting future values based on past values

02 Forecast instead is integrated, so it uses the difference between past and future values

03 Forecast is not suited for long term predictions and the parameters are subjective

Advantages

1. Great for observational count data; there's no need to convert/calculate your data
2. Can compare model fit in order to choose the best one
3. Great for weird data

Limitations

1. It's important to know your species biology well in order to implement an appropriate time lag
2. Very technical and stats heavy
3. Limited resources for application and interpretation
    a. Vignette not very helpful

Resources

Vignette:

"The glarma Package for Observation Driven Time Series Regression of Counts" by William T.M. Dunsmuir and David J. Scott

Package Description:

"Package: Glarma" by William T.M. Dunsmuir and Cenanning Li

Studies that use Glarma:

"Trends of Canine Rabies Lyssavirus and Impact of the Intensified Rabies Control Program in Davao City, Philippines: 2006–2017" by Zython Lachica et al.

"Spawning periodicity and synchrony of bluehead chub (Nocomis leptocephalus) and a nest associate, yellowfin shiner (Notropis lutipinnis), across local streams" by Seonghyun Kim and Yoichiro Kanno

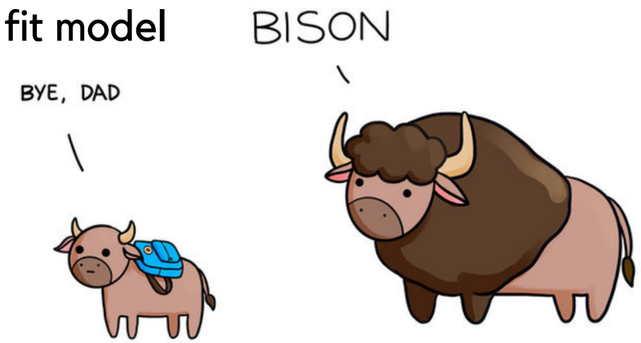
PORTFOLIO ASSIGNMENT

Comparing Time Lags for Bison Population Modeling

We're giving you the code! ;) you're welcome

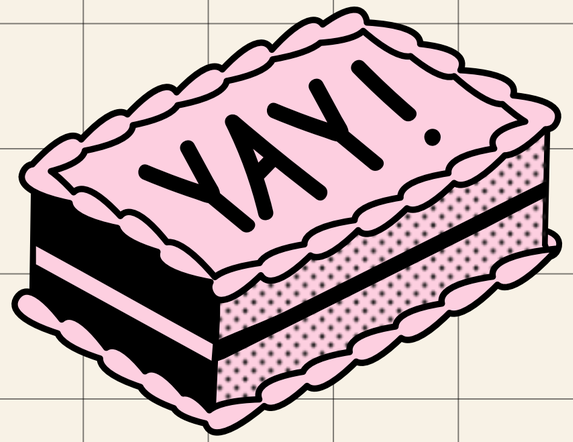
Questions:

1. After you create your models, visually compare the plots. Which looks like the best fit?
2. Compare the AIC for the models. Which is actually the best fit?
3. Create a model with major events factored in and compare it with the best fit model identified in question 2. Is the model with major events a better fit?



Some hints:

- Think about bison reproduction and how it would affect time lags:
  - When do they reach sexual maturity? (2 years for females)
  - How long is gestation? (~ 9 months)
  - How often do bison breed? (Once a year)
- Lower AIC = good



THANK YOU!