Harbor seals analysis

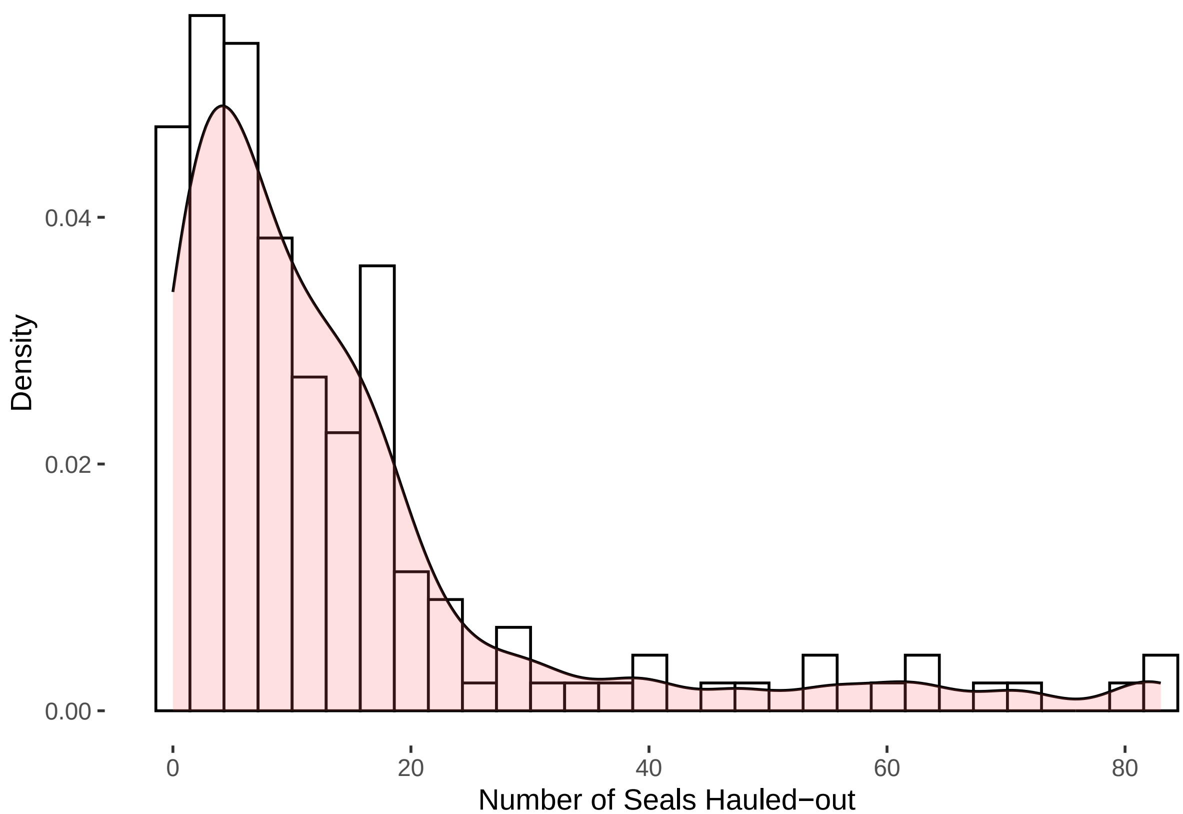
Kyra Bankhead

2022-12-14

In this markdown I will:

1. Find the best distribution to use for GLMMs.
2. Run GLMMs and AICcs to find the most appropriate model and predictors.
3. Create visualization graphs for each site.

Check for Appropriate Distribution



Negative binomial or poisson would be the best fit. Let's check if the mean is equal to the variance and if there is zero inflation.

```r
require(MASS) # for glm
require(performance) # overdispersion
require(DHARMa) # auto cor and zero-inflation


var(full.data$seals)
```

```
## [1] 299.2771
```

```r
mean(full.data$seals)
```

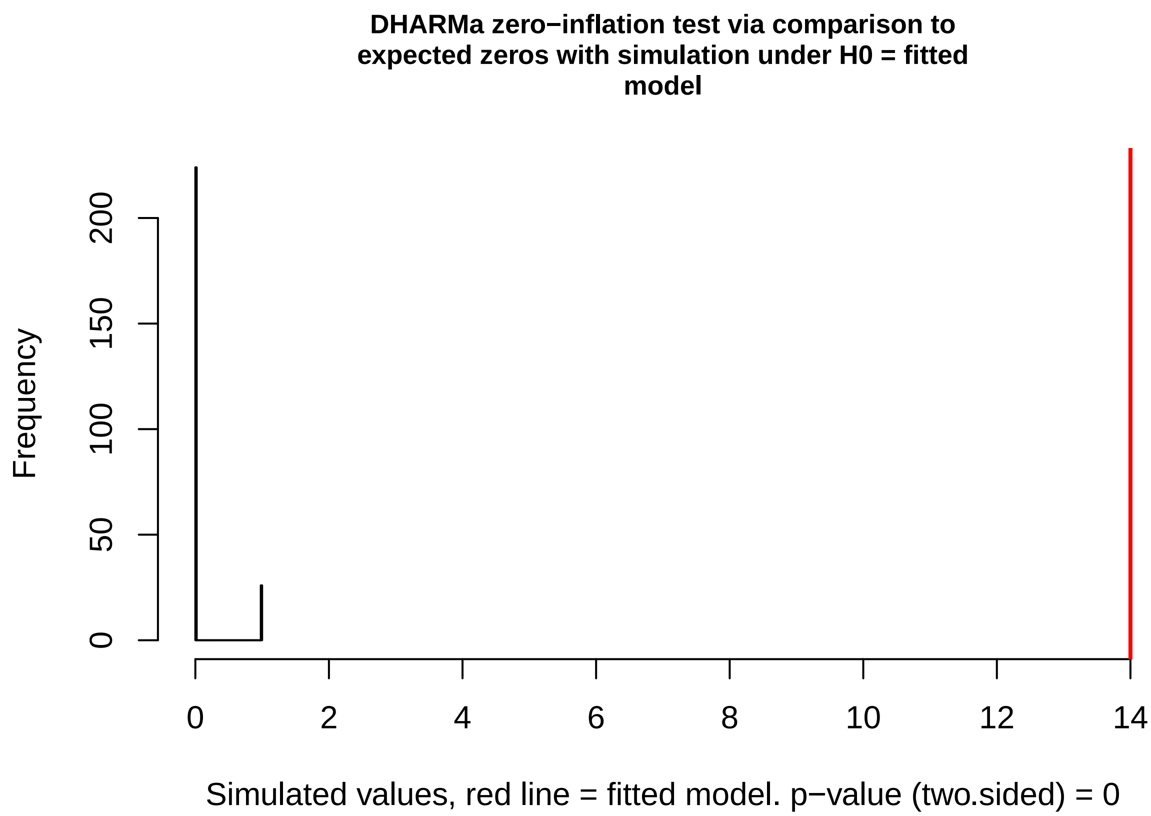
```
## [1] 14.12258
```

```r
# Check poisson model
mod<- glm(seals ~ site*noise + month + tide + time, data = full.data, family = "poisson")
simulationOutput <- simulateResiduals(fittedModel = mod)

# Check for overdispersion
check_overdispersion(mod)
```

```
## # Overdispersion test
##
##        dispersion ratio =   6.737
##   Pearson's Chi-Squared = 997.052
##                 p-value = < 0.001
```

```r
# Check for zero inflation
testZeroInflation(simulationOutput)
```



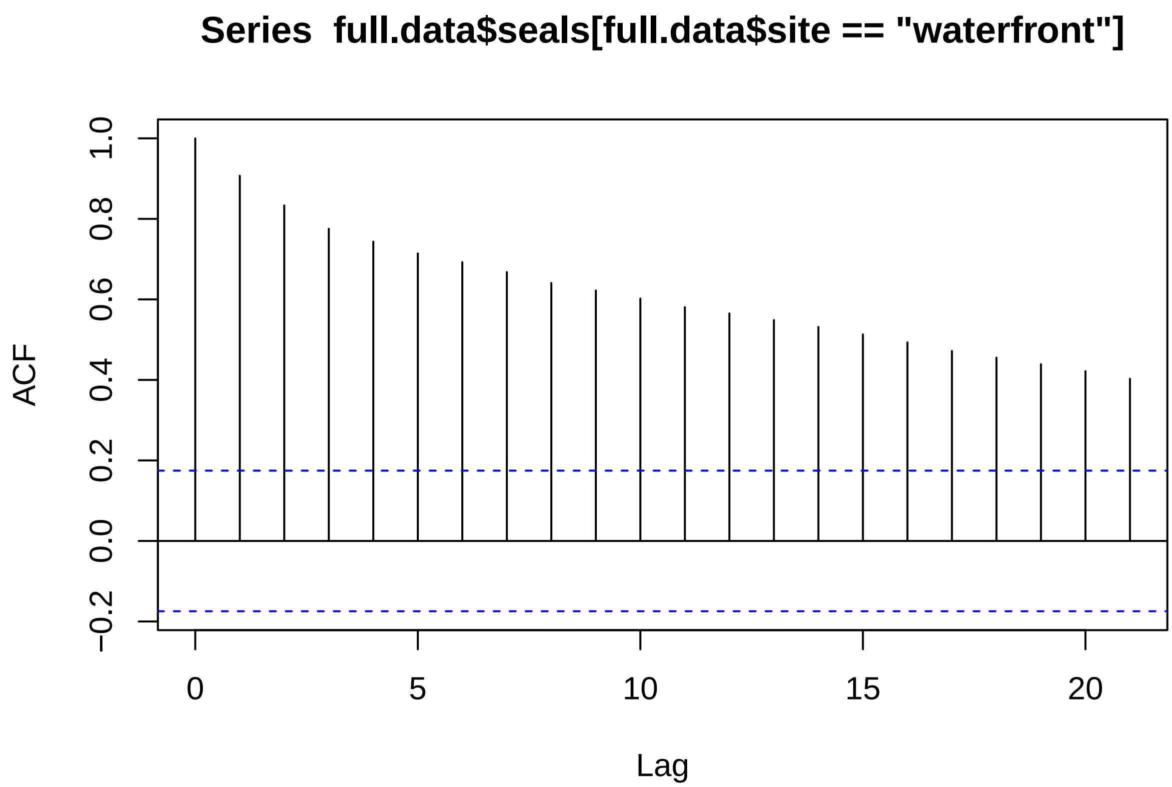

```
##
##  DHARMa zero-inflation test via comparison to expected zeros with
##  simulation under H0 = fitted model
##
## data:  simulationOutput
## ratioObsSim = 134.62, p-value < 2.2e-16
## alternative hypothesis: two.sided
```

The variance is way higher than the mean which was confirmed by the Pearson chi-squared test. Additionally there was zero-inflation present in the model. Therefore the Poisson distribution is not applicable here and we will continue with the negative binomial distribution.

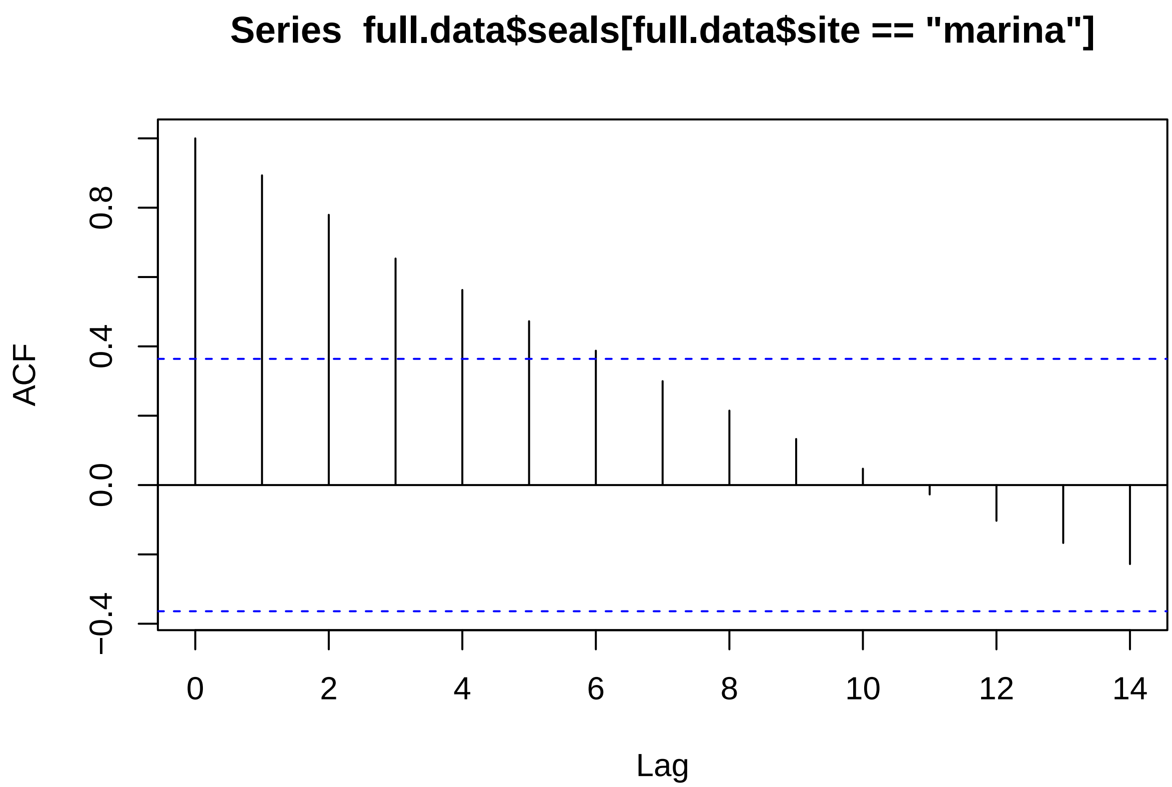
Check for autocorrelation

```
## Loading required package: glmmTMB
```

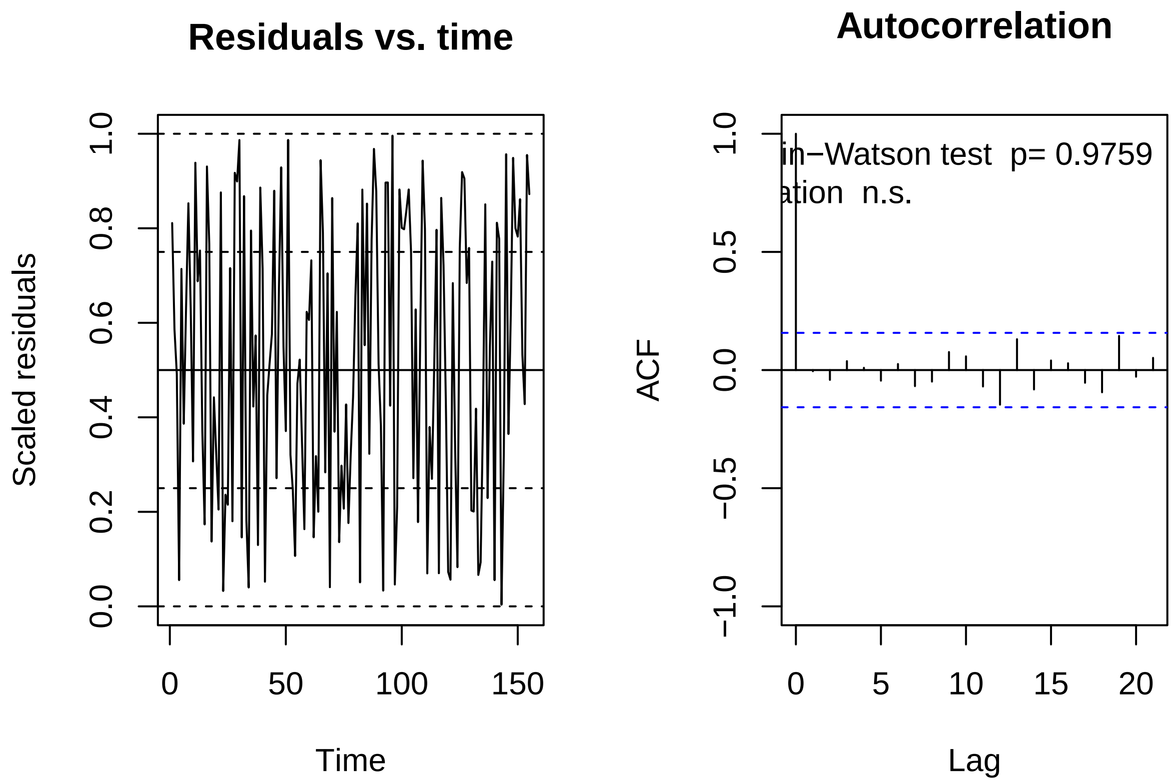
```
## Warning in checkMatrixPackageVersion(): Package version inconsistency detected.
## TMB was built with Matrix version 1.4.1
## Current Matrix version is 1.5.3
## Please re-install 'TMB' from source using install.packages('TMB', type = 'source') or ask CRAN for a
```



Series full.data$seals[full.data$site == "waterfront"]
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
Lag




```
## DHARMa::testTemporalAutocorrelation - no time argument provided, using random times for each data poi
```




```
##
##  Durbin-Watson test
##
## data:  simulationOutput$scaledResiduals ~ 1
## DW = 1.9952, p-value = 0.9759
## alternative hypothesis: true autocorrelation is not 0
```

The x-axis corresponds to the different lags of the residuals. The y-axis shows the correlation of each lag. Finally, the dashed blue line represents the significance level. Both sites show high autocorrelation. Therefore I have incorporated a ar1() effect between month and year to correct for the correlation between observations as seen in the third graph.

Create GLMMs and find best model with AICc

To test whether noise affects the number of seals hauled-out by site, I will insert an interaction between noise level and site.

```
##
## Model selection based on AICc:
##
##                                   K    AICc Delta_AICc AICcWt Cum.Wt      LL
## seals ~ site+noise                8 1043.29       0.00   0.46   0.46 -513.15
## seals ~ site*noise                9 1044.19       0.90   0.29   0.75 -512.48
## seals ~ site*noise + time        10 1045.02       1.73   0.19   0.94 -511.75
```

```
## seals ~ site*noise + tide + time 11 1047.28       4.00   0.06   1.00 -511.72
## seals ~ 1                        3 1145.72     102.43   0.00   1.00 -569.78
```

Looks like the best model will contain month, noise, site and time as predictors. This is the summary of that model:

```
##  Family: nbinom2  ( log )
## Formula:
## seals ~ noise * site + (1 | date) + ar1(as.factor(month) + 0 |      year)
## Zero inflation:         ~1
## Data: full.data
##
##      AIC      BIC   logLik deviance df.resid
##   1043.0   1070.3   -512.5   1025.0      146
##
## Random effects:
##
## Conditional model:
##  Groups Name              Variance Std.Dev. Corr
##  date   (Intercept)       0.03188  0.1786
##  year   as.factor(month)6 0.15094  0.3885   0.06 (ar1)
## Number of obs: 155, groups:  date, 135; year, 2
##
## Dispersion parameter for nbinom2 family (): 3.47
##
## Conditional model:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)           4.96740    0.94127   5.277 1.31e-07 ***
## noise                -0.03403    0.02376  -1.432   0.1521
## sitewaterfront       -2.79924    1.12998  -2.477   0.0132 *
## noise:sitewaterfront  0.03146    0.02683   1.173   0.2410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Zero-inflation model:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.7374     0.4156  -6.587  4.5e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

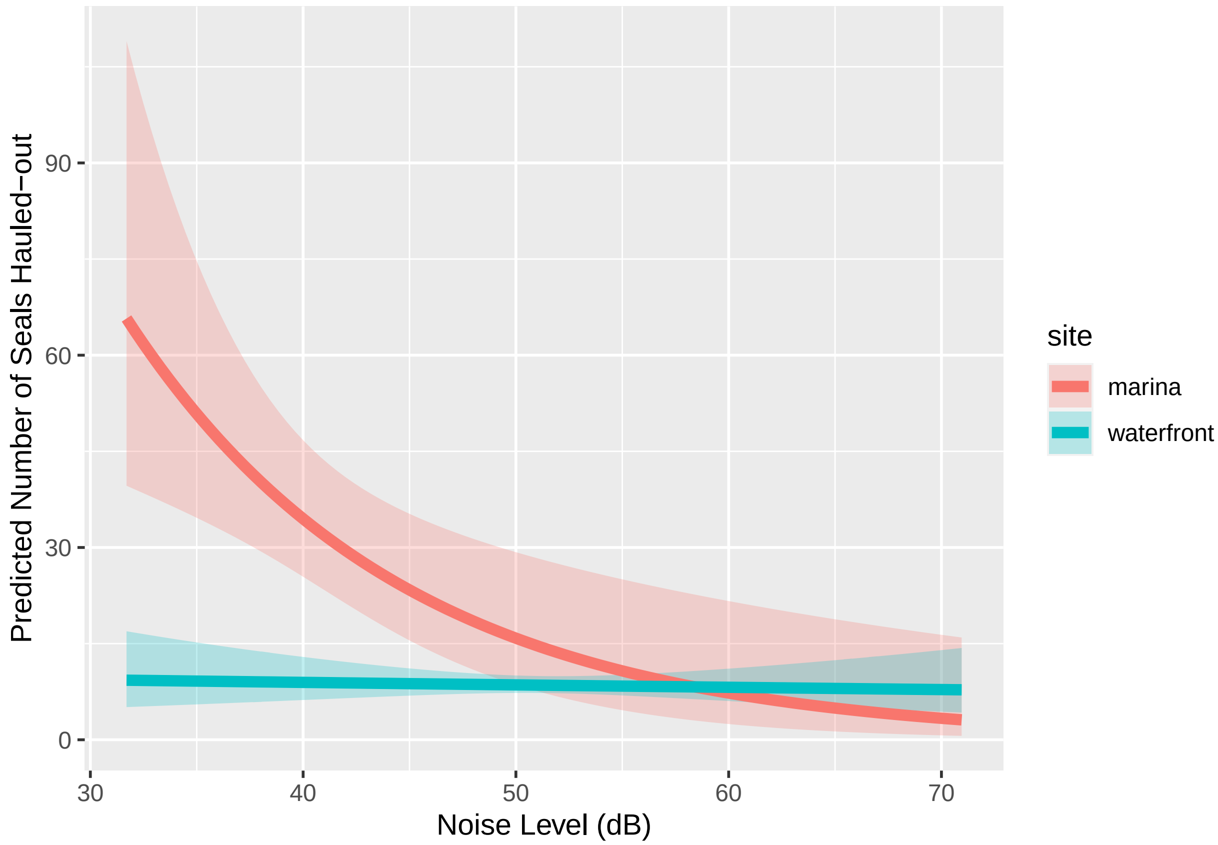
- Despite what the p-values show, since and interaction between site and noise were chosen over the other models we will say that the effect of noise on the number of seals haul-out depends on what site they are located in.

Use model output to predict response variable

```
##   noise       site      phat
## 1 49.07     marina 17.034396
## 2 49.07 waterfront  8.595981
```

In the output above, we see that the predicted number of seals hauled-out for the Marina is about 17.03, holding noise level at its mean. The predicted number of events for the Waterfront is lower at 8.60.

Below we will obtain the mean predicted number of seals hauled-out for values of noise across its entire range for each site and graph these.



The graph shows the log linear model of the expected seals counts across the range of noise levels, for each site along with 95 percent confidence intervals. Note that what is plotted are the expected values, not the log of the expected values.