glmmTMB models

Sarah Popov

2023-02-05

WESA ZIP + ZINB base models

By station

Dataset: one count record per station per survey date, `nrow(dat)` records. `dat_p0`% of the records are zeroes.

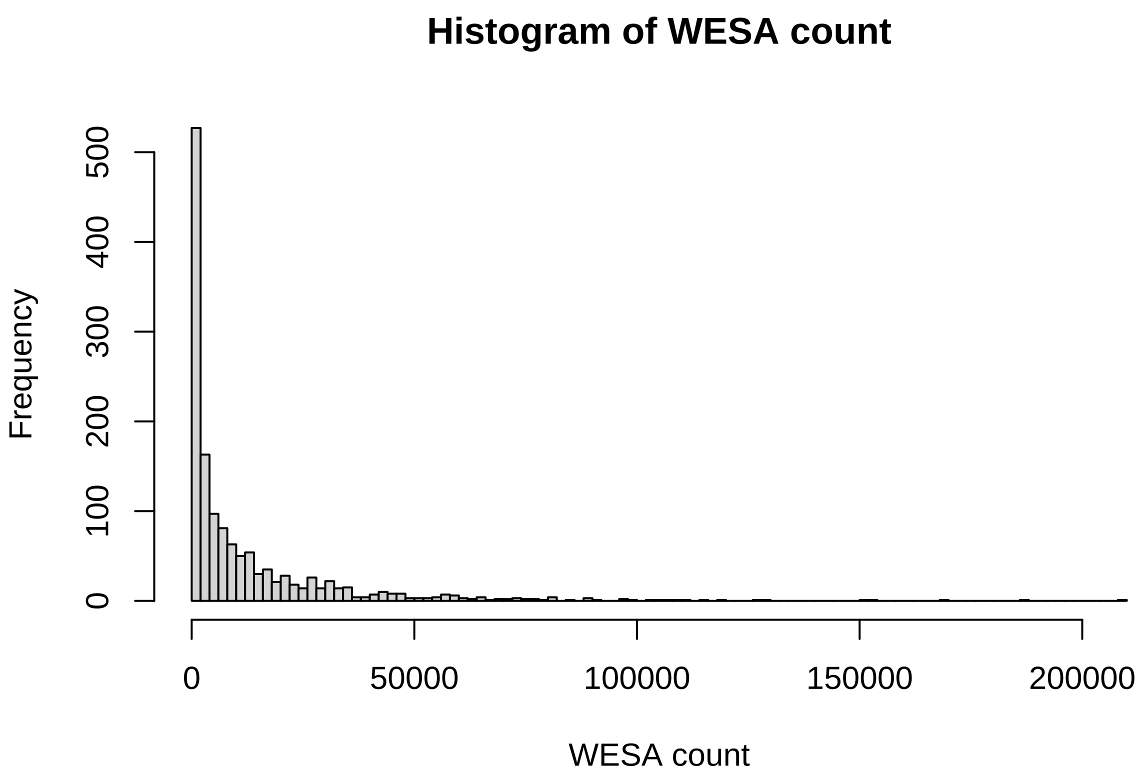


Figure 1: Histogram of WESA count per station per survey date. Plenty of zeroes...

Base models

```
##                dAIC df
## fit_poisson    NA   16
## fit_binom1     NA   17
## fit_binom2     NA   17
```

```
## fit_zipoisson NA   17
## fit_zinbinom1 NA   18
## fit_zinbinom2 NA   18
```

- Results of `glmmTMB::diagnose()` suggest certain random effects structures fail to converge: `(dos + I(dos^2) | n_s)` and `(dos + I(dos^2) | station_n)` both fail.
- A zero-inflation model is definitely the way to go, as the non-zi formula models fail to converge.
- There are potentially too many zeroes in this dataset than would be expected with a Poisson distribution. While it's certainly possible to run these models with Poisson distributions, it seems assuming a *negative binomial* distribution in the response variable makes more sense.

Simulate best-fit model

Simulate the best-fit model, `fit_zinbinom1`, 10,000 times to get the expected distribution of zero-values. The red line is the true proportion of zero-values from the data itself (`dat_p0`).

```
##    user  system elapsed
##  16.992   0.134  17.143
```

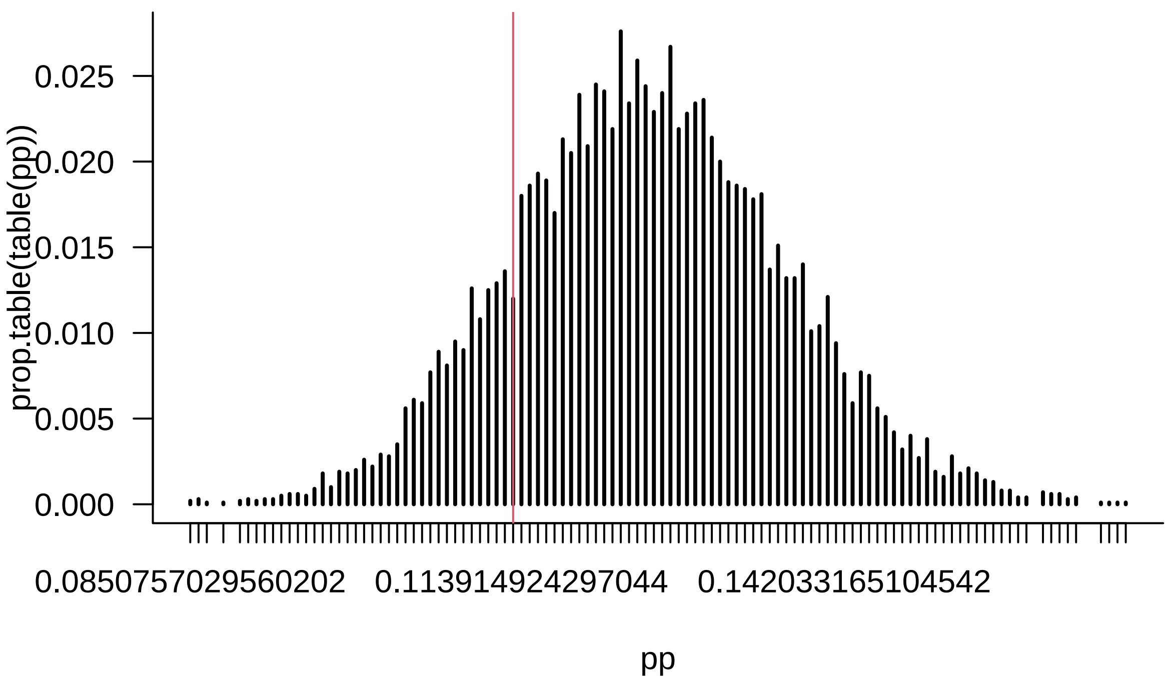


Figure 2: Best-fit model distribution of zeroes vs. observed proportion of zeroes (red line). The model overestimates the amount of zeroes observed in our dataset.

```
## numeric(0)
```

By North vs. South

Dataset: one count record per N/S division per survey date, `nrow(dat3)` records.

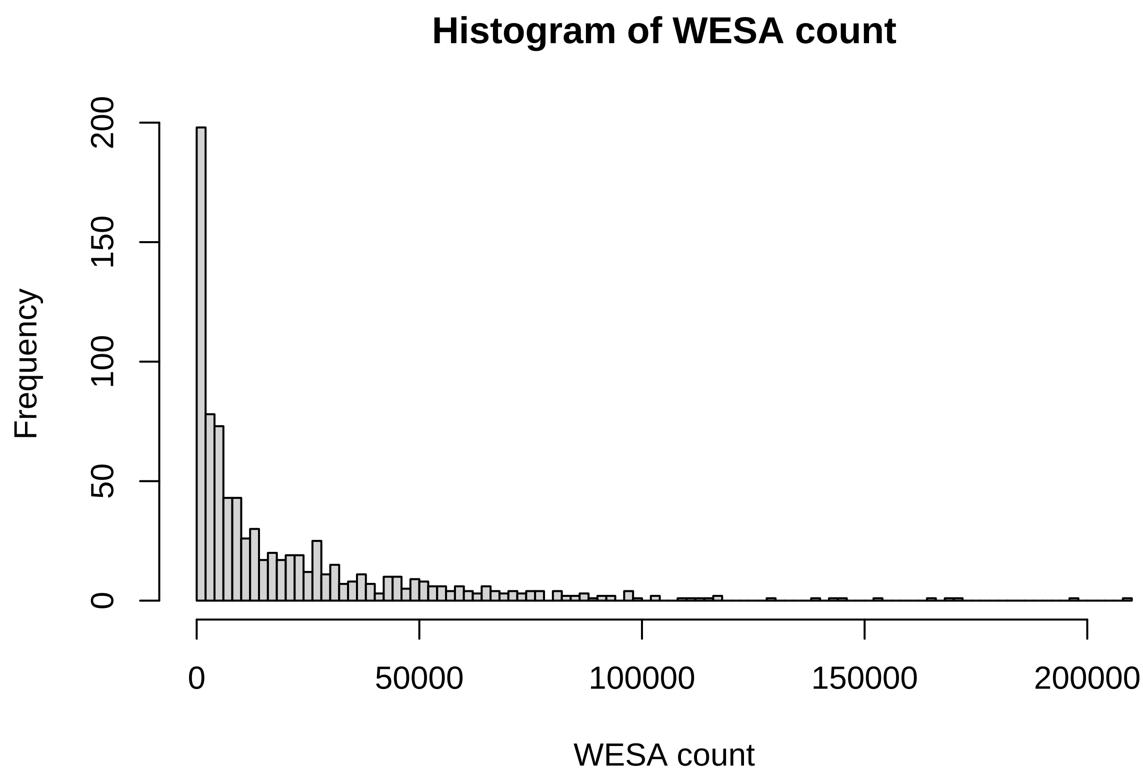


Figure 3: Histogram of WESA count per north/south region per survey date. Also plenty of zeroes. . . !

Base models

```
##                dAIC df
## fit_zipoisson   0   13
## fit_poisson    NA   12
## fit_binom1     NA   13
## fit_binom2     NA   13
## fit_zinbinom1 NA   14
## fit_zinbinom2 NA   14
```

- Similar to above, best-fit model is a zi negative binomial model.

Simulate best-fit model

Simulate the best-fit model, **`fit_zinbinom2`**, 10,000 times to get the expected distribution of zero-values. The red line is the true proportion of zero-values from the data itself (`dat3_p0`).

```
##    user  system elapsed
##  11.129   0.030  11.163
```

```
## numeric(0)
```

The residuals are also looking really good for the base model:

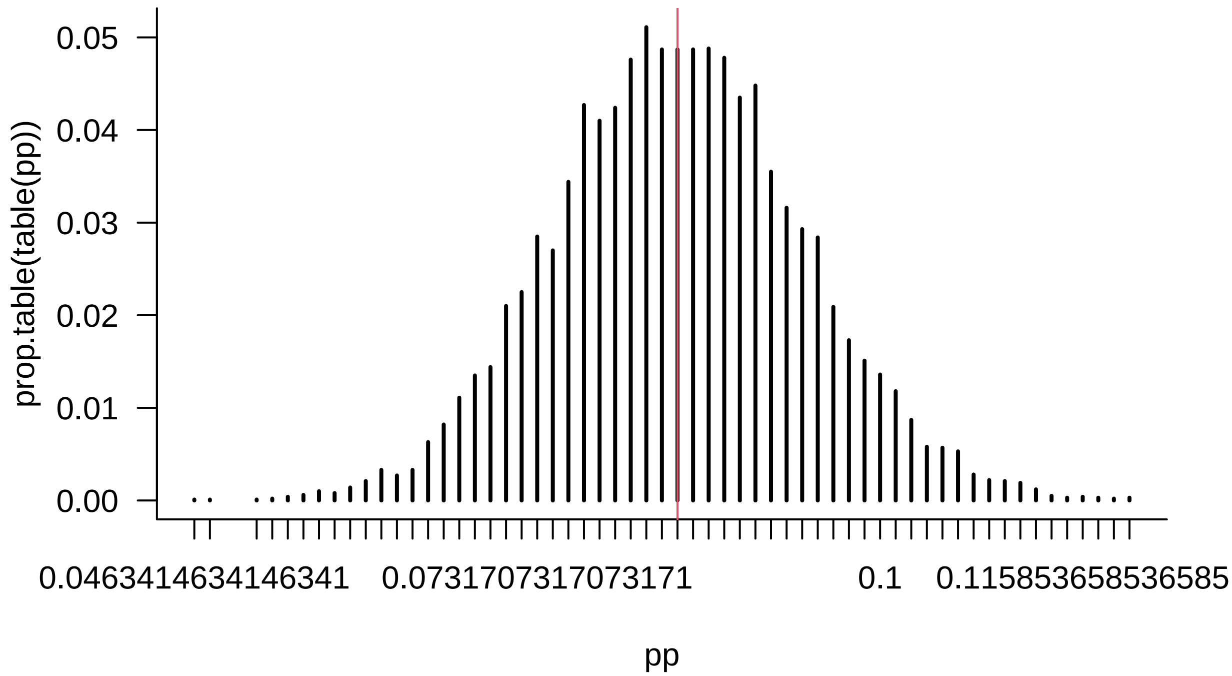
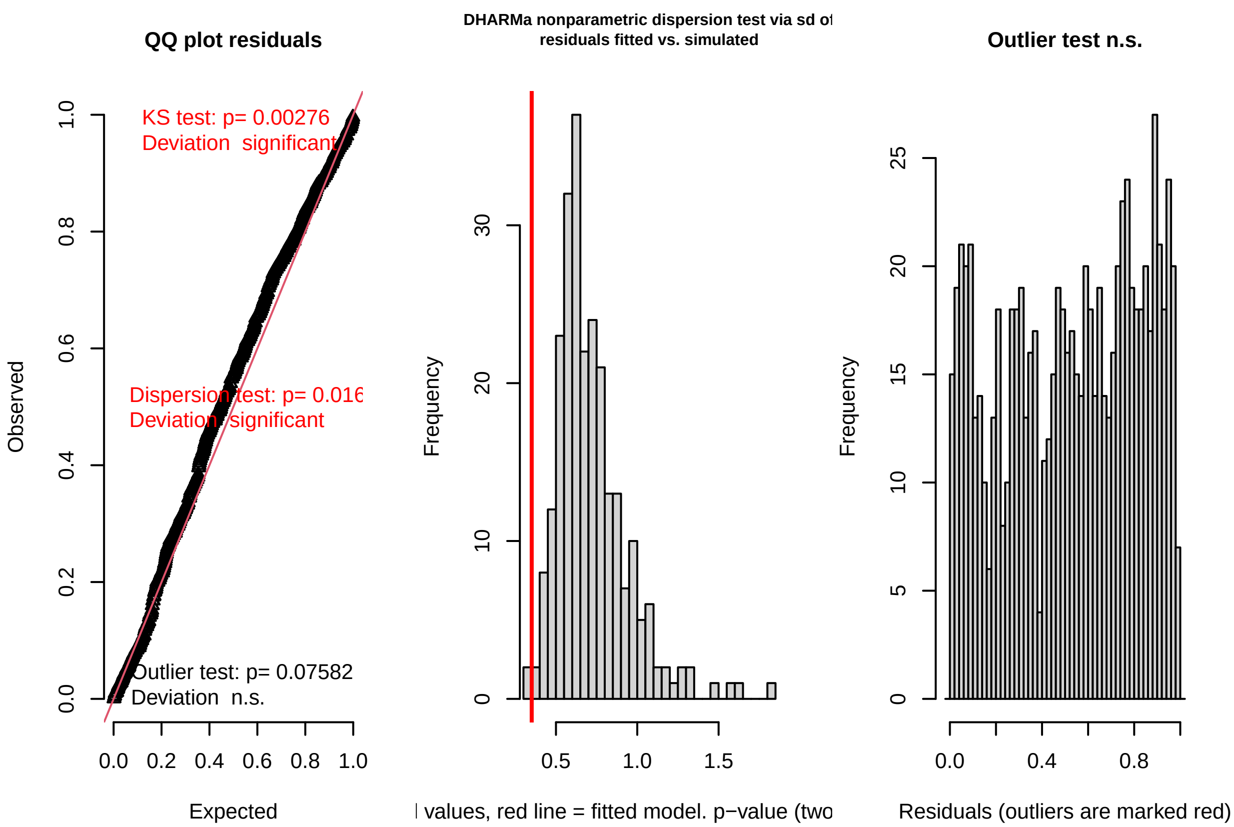


Figure 4: Best-fit model distribution of zeroes vs. observed proportion of zeroes (red line). The base model actually does a great job of estimating the correct proportion of zeroes!




```
## $uniformity
##
```

```
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.063366, p-value = 0.002762
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## dispersion = 0.48742, p-value = 0.016
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa outlier test based on exact binomial test with approximate
##  expectations
##
## data:  simulationOutput
## outliers at both margin(s) = 2, observations = 820, p-value = 0.07582
## alternative hypothesis: true probability of success is not equal to 0.007968127
## 95 percent confidence interval:
##  0.0002955138 0.0087825707
## sample estimates:
## frequency of outliers (expected: 0.00796812749003984 )
##                                             0.002439024

## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.063366, p-value = 0.002762
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## dispersion = 0.48742, p-value = 0.016
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa outlier test based on exact binomial test with approximate
##  expectations
```

```
##
## data:  simulationOutput
## outliers at both margin(s) = 2, observations = 820, p-value = 0.07582
## alternative hypothesis: true probability of success is not equal to 0.007968127
## 95 percent confidence interval:
##  0.0002955138 0.0087825707
## sample estimates:
## frequency of outliers (expected: 0.00796812749003984 )
##                                            0.002439024
```

Full model

Since the base model seems to actually be doing a good job of predicting our zeroes correctly, lets add the full suite of important variables identified in Canham et al. (2021).

```
## predicted_wesa ~ 1 + I(dos^2) + n_s + year_c + scale(windspd) +
##     scale(total_precip) + dos + scale(mean_temp) + tide + scale(flow) +
##     scale(elev_range) + scale(v) + scale(u) + n_s:scale(flow)
## <environment: 0x7fec3b2c3e20>
```

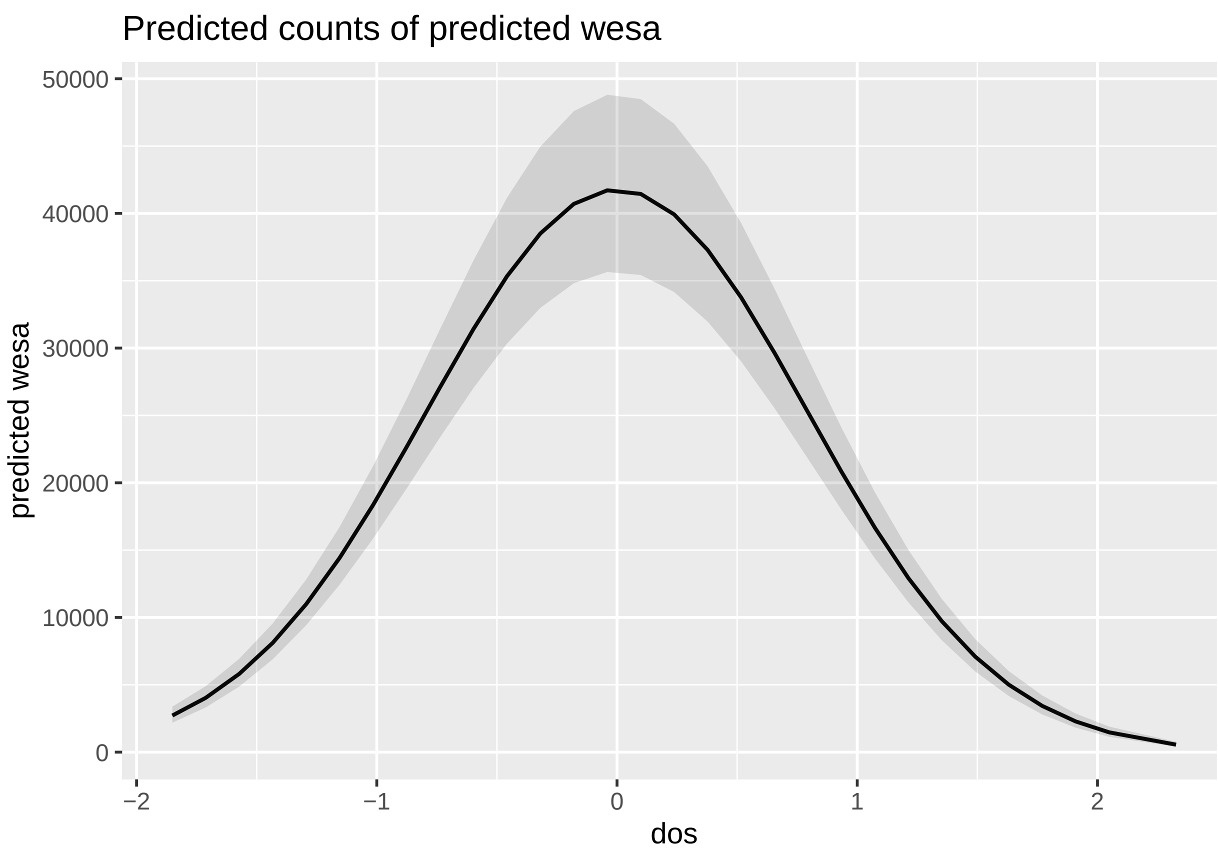
```
##    grouping                term                     block Iteration
## 1        NA                   1                   NA NA 1         1
## 2        NA            I(dos^2)            NA NA I(dos^2)         1
## 3        NA                 n_s                 NA NA n_s         1
## 4        NA              year_c              NA NA year_c         1
## 5        NA      scale(windspd)      NA NA scale(windspd)         1
## 6        NA scale(total_precip) NA NA scale(total_precip)         1
## 7        NA                 dos                 NA NA dos         1
## 8        NA    scale(mean_temp)    NA NA scale(mean_temp)         1
## 9        NA                tide                NA NA tide         1
## 10       NA         scale(flow)         NA NA scale(flow)         1
## 11       NA   scale(elev_range)   NA NA scale(elev_range)         1
## 12       NA            scale(v)            NA NA scale(v)         1
## 13       NA            scale(u)            NA NA scale(u)         1
## 14       NA     n_s:scale(flow)     NA NA n_s:scale(flow)         1
##             LRT
## 1            NA
## 2  2.721586e-41
## 3            NA
## 4  1.608086e-02
## 5  2.297421e-01
## 6  2.890690e-01
## 7  1.779782e-01
## 8  6.028775e-02
## 9  1.709183e-01
## 10           NA
## 11 1.536027e-02
## 12 4.125689e-01
## 13 1.756334e-01
## 14 6.751721e-03
##    grouping                term                     block Iteration
## 1        NA                   1                   NA NA 1         2
## 2        NA            I(dos^2)            NA NA I(dos^2)         2
```

```
## 3        NA                n_s                 NA NA n_s         2
## 4        NA             year_c              NA NA year_c         2
## 5        NA     scale(windspd)      NA NA scale(windspd)         2
## 6        NA scale(total_precip) NA NA scale(total_precip)         2
## 7        NA                dos                 NA NA dos         2
## 8        NA    scale(mean_temp)    NA NA scale(mean_temp)         2
## 9        NA               tide                NA NA tide         2
## 10       NA        scale(flow)         NA NA scale(flow)         2
## 11       NA   scale(elev_range)   NA NA scale(elev_range)         2
## 13       NA           scale(u)            NA NA scale(u)         2
## 14       NA     n_s:scale(flow)     NA NA n_s:scale(flow)         2
##             LRT
## 1            NA
## 2  3.595218e-41
## 3            NA
## 4  1.539885e-02
## 5  3.062086e-01
## 6  1.000000e+00
## 7  3.744783e-01
## 8  1.145044e-01
## 9  2.611901e-01
## 10           NA
## 11 2.714473e-02
## 13 1.997285e-01
## 14 8.649581e-03
##    grouping              term                   block Iteration          LRT
## 1        NA                 1                 NA NA 1         3           NA
## 2        NA          I(dos^2)          NA NA I(dos^2)         3 3.413419e-41
## 3        NA               n_s               NA NA n_s         3           NA
## 4        NA            year_c            NA NA year_c         3 1.584331e-02
## 5        NA    scale(windspd)    NA NA scale(windspd)         3 2.642618e-01
## 7        NA               dos               NA NA dos         3 3.439226e-01
## 8        NA  scale(mean_temp)  NA NA scale(mean_temp)         3 1.046603e-01
## 9        NA              tide              NA NA tide         3 2.287986e-01
## 10       NA       scale(flow)       NA NA scale(flow)         3           NA
## 11       NA scale(elev_range) NA NA scale(elev_range)         3 2.431653e-02
## 13       NA          scale(u)          NA NA scale(u)         3 1.740044e-01
## 14       NA   n_s:scale(flow)   NA NA n_s:scale(flow)         3 8.545909e-03
##    grouping              term                   block Iteration          LRT
## 1        NA                 1                 NA NA 1         4           NA
## 2        NA          I(dos^2)          NA NA I(dos^2)         4 1.424817e-42
## 3        NA               n_s               NA NA n_s         4           NA
## 4        NA            year_c            NA NA year_c         4 1.976681e-02
## 5        NA    scale(windspd)    NA NA scale(windspd)         4 2.524396e-01
## 8        NA  scale(mean_temp)  NA NA scale(mean_temp)         4 6.115978e-02
## 9        NA              tide              NA NA tide         4 1.708642e-01
## 10       NA       scale(flow)       NA NA scale(flow)         4           NA
## 11       NA scale(elev_range) NA NA scale(elev_range)         4 1.199084e-02
## 13       NA          scale(u)          NA NA scale(u)         4 1.572248e-01
## 14       NA   n_s:scale(flow)   NA NA n_s:scale(flow)         4 5.683203e-03
##    grouping              term                   block Iteration          LRT
## 1        NA                 1                 NA NA 1         5           NA
## 2        NA          I(dos^2)          NA NA I(dos^2)         5 5.848703e-43
## 3        NA               n_s               NA NA n_s         5           NA
```

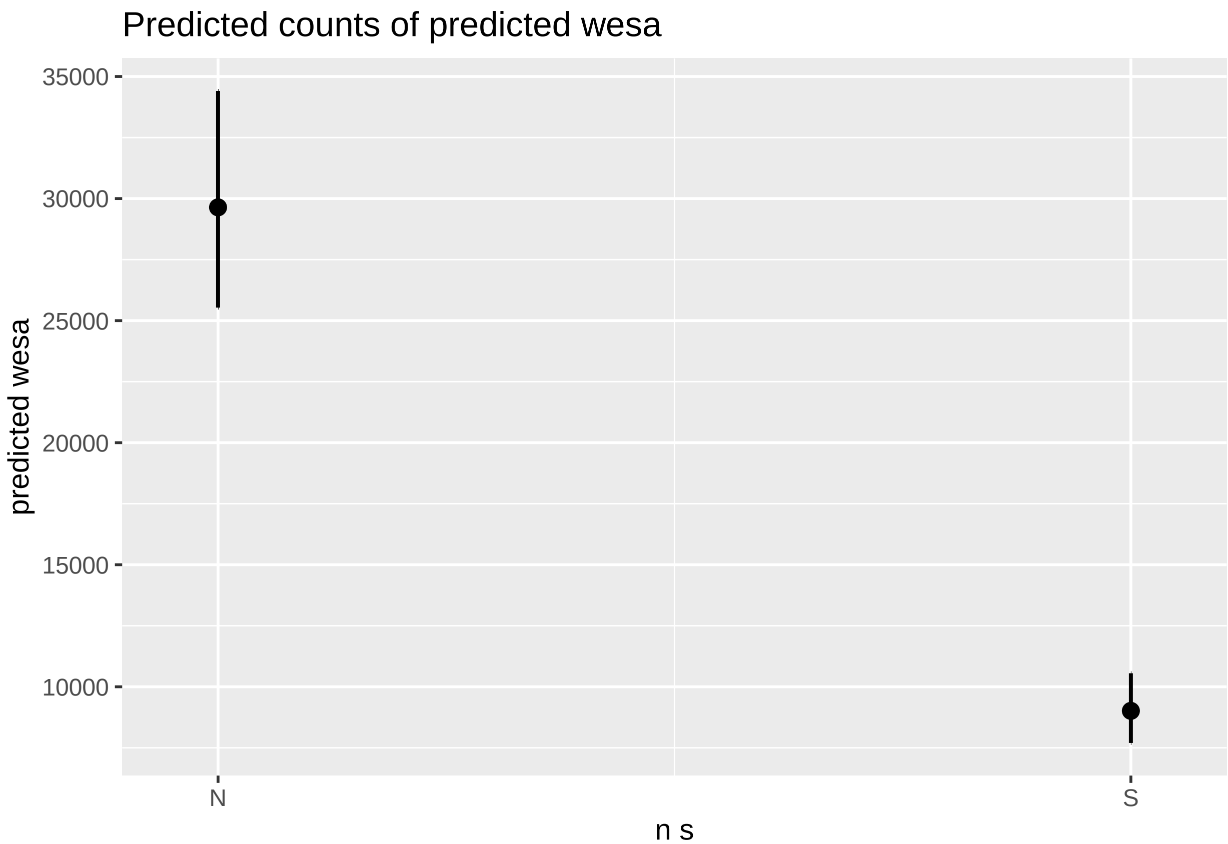
```
## 4        NA            year_c            NA NA year_c         5 2.263401e-02
## 8        NA  scale(mean_temp)  NA NA scale(mean_temp)         5 5.055212e-02
## 9        NA              tide              NA NA tide         5 1.514715e-01
## 10       NA       scale(flow)       NA NA scale(flow)         5           NA
## 11       NA scale(elev_range) NA NA scale(elev_range)         5 1.312859e-02
## 13       NA          scale(u)          NA NA scale(u)         5 1.182987e-01
## 14       NA   n_s:scale(flow)   NA NA n_s:scale(flow)         5 5.681264e-03
##    grouping              term                   block Iteration          LRT
## 1        NA                 1                 NA NA 1         6           NA
## 2        NA          I(dos^2)          NA NA I(dos^2)         6 1.109769e-42
## 3        NA               n_s               NA NA n_s         6           NA
## 4        NA            year_c            NA NA year_c         6 7.097449e-03
## 8        NA  scale(mean_temp)  NA NA scale(mean_temp)         6 5.548838e-02
## 10       NA       scale(flow)       NA NA scale(flow)         6           NA
## 11       NA scale(elev_range) NA NA scale(elev_range)         6 7.867718e-04
## 13       NA          scale(u)          NA NA scale(u)         6 1.657814e-01
## 14       NA   n_s:scale(flow)   NA NA n_s:scale(flow)         6 6.815652e-03
##    grouping              term                   block Iteration          LRT
## 1        NA                 1                 NA NA 1         7           NA
## 2        NA          I(dos^2)          NA NA I(dos^2)         7 2.155277e-42
## 3        NA               n_s               NA NA n_s         7           NA
## 4        NA            year_c            NA NA year_c         7 9.630618e-03
## 8        NA  scale(mean_temp)  NA NA scale(mean_temp)         7 3.931913e-02
## 10       NA       scale(flow)       NA NA scale(flow)         7           NA
## 11       NA scale(elev_range) NA NA scale(elev_range)         7 1.018990e-03
## 14       NA   n_s:scale(flow)   NA NA n_s:scale(flow)         7 5.374905e-03

##  Family: nbinom2  ( log )
## Formula:
## predicted_wesa ~ 1 + I(dos^2) + n_s + year_c + scale(mean_temp) +
##     scale(flow) + scale(elev_range) + n_s:scale(flow)
## Zero inflation:                 ~. - n_s
## Data: dat3
##
##      AIC      BIC   logLik deviance df.resid
##  16329.5  16404.9  -8148.8  16297.5      804
##
##
## Dispersion parameter for nbinom2 family (): 0.994
##
## Conditional model:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)       10.8964348  0.0597839  182.26  < 2e-16 ***
## I(dos^2)          -0.7976287  0.0326648  -24.42  < 2e-16 ***
## n_sS              -1.1904633  0.0750694  -15.86  < 2e-16 ***
## year_c            -0.1658925  0.0376284   -4.41 1.04e-05 ***
## scale(mean_temp)  -0.0856034  0.0383984   -2.23   0.0258 *
## scale(flow)       -0.0006394  0.0530568   -0.01   0.9904
## scale(elev_range) -0.0729741  0.0381219   -1.91   0.0556 .
## n_sS:scale(flow)  -0.0680396  0.0795012   -0.86   0.3921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Zero-inflation model:
```

```
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -3.19854    0.23380 -13.681  < 2e-16 ***
## I(dos^2)           0.37639    0.11952   3.149  0.00164 **
## year_c            -0.67383    0.15234  -4.423 9.73e-06 ***
## scale(mean_temp)  -0.25337    0.14280  -1.774  0.07602 .
## scale(flow)       -0.09636    0.18040  -0.534  0.59325
## scale(elev_range) -0.66036    0.14933  -4.422 9.78e-06 ***
## scale(flow):n_sS   0.07771    0.24987   0.311  0.75578
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

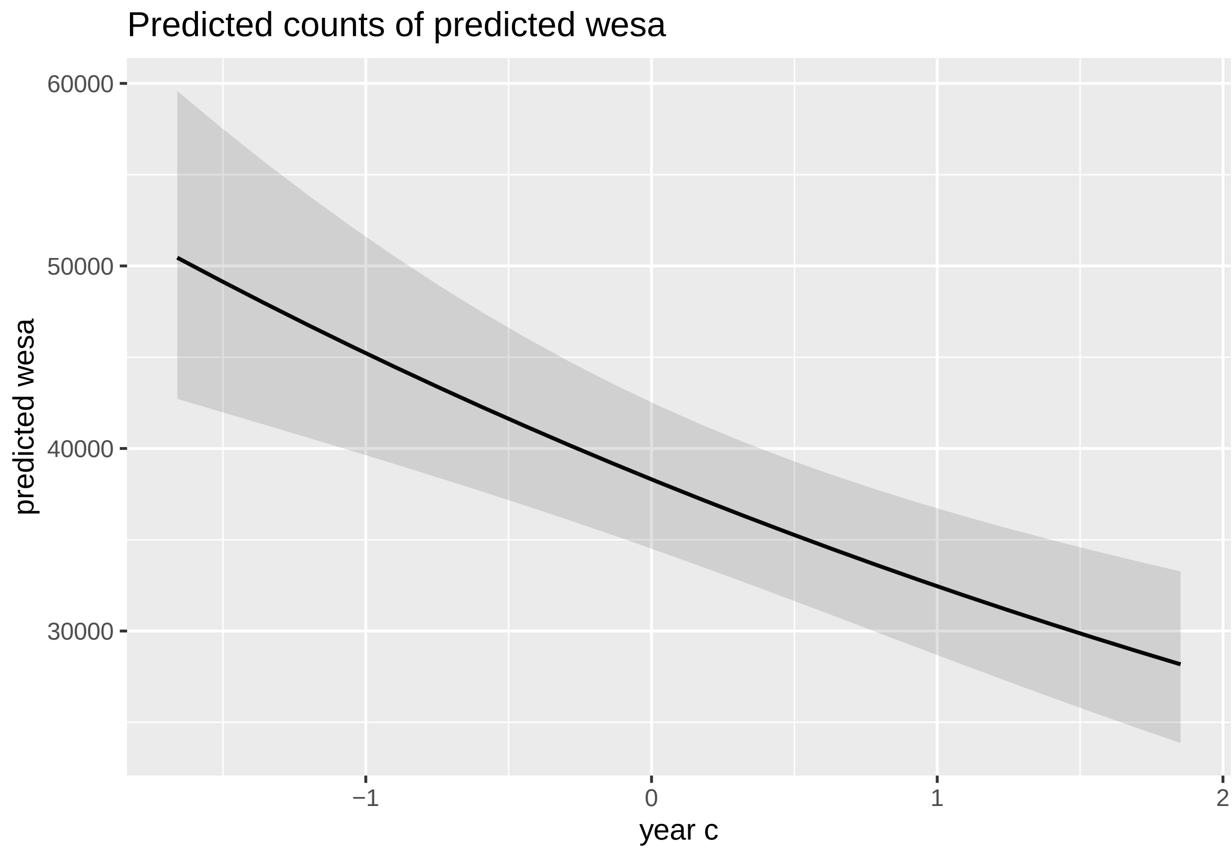
```
## $dos
```



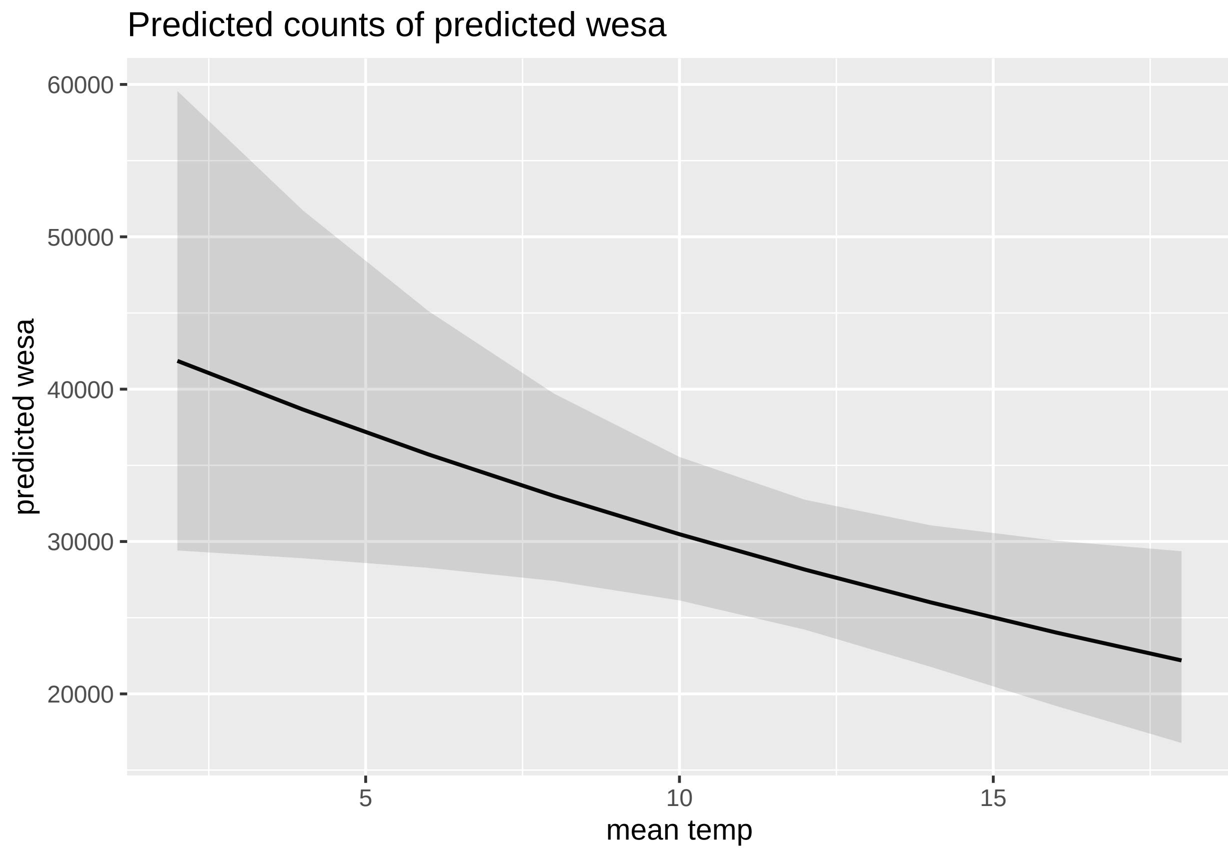

```
##
## $n_s
```


Predicted counts of predicted wesa
predicted wesa
35000
30000
25000
20000
15000
10000
N
S
n s

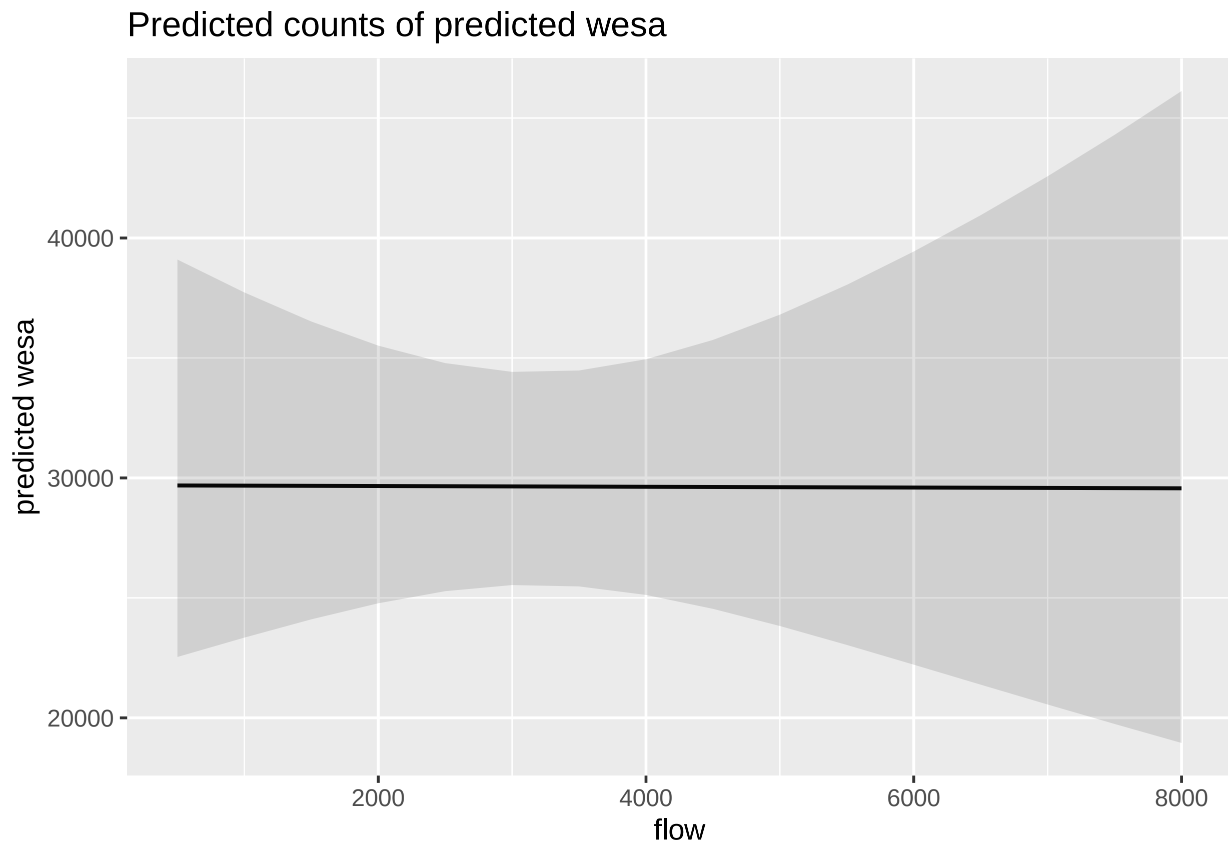
```
##
## $year_c
```


Predicted counts of predicted wesa
60000
50000
40000
30000
predicted wesa
−1
0
1
2
year c

```
##
## $mean_temp
```

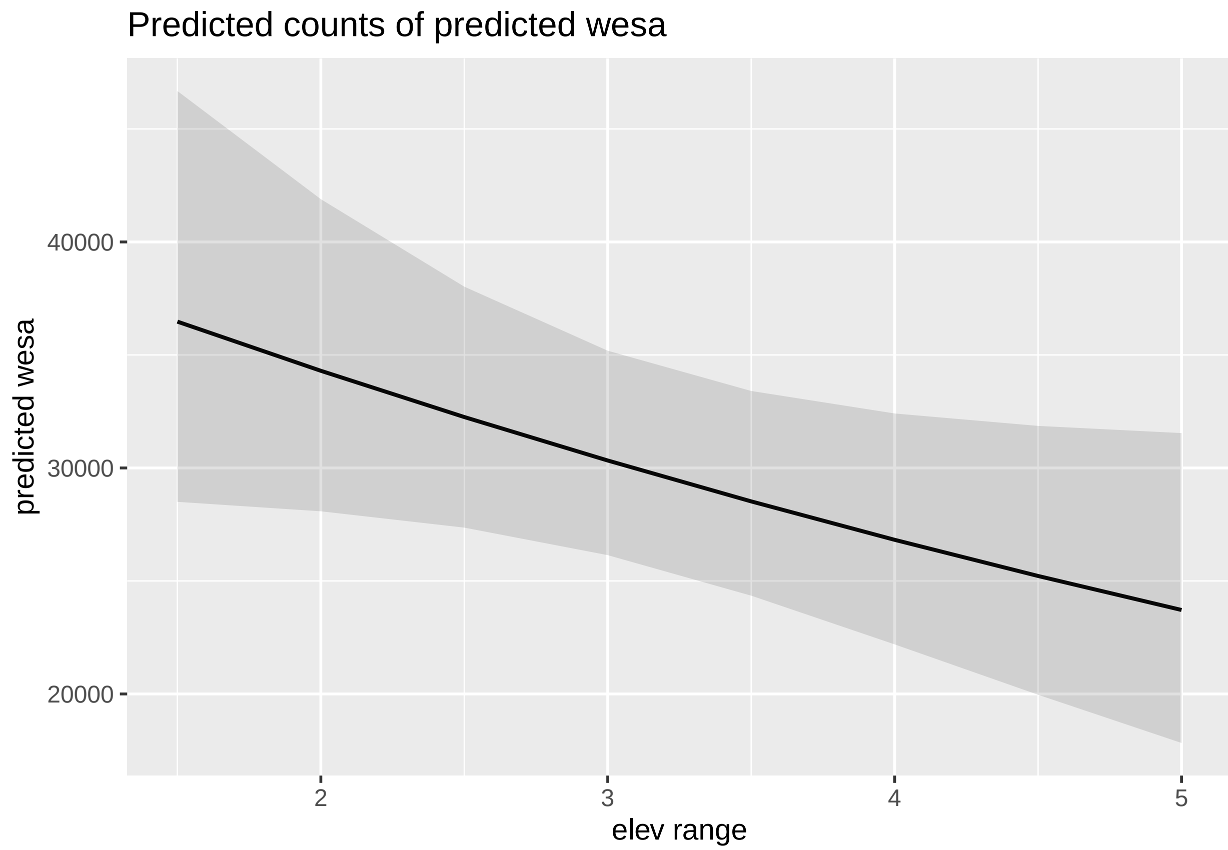

Predicted counts of predicted wesa
60000
50000
40000
30000
20000
predicted wesa
5
10
15
mean temp

```
##
## $flow
```

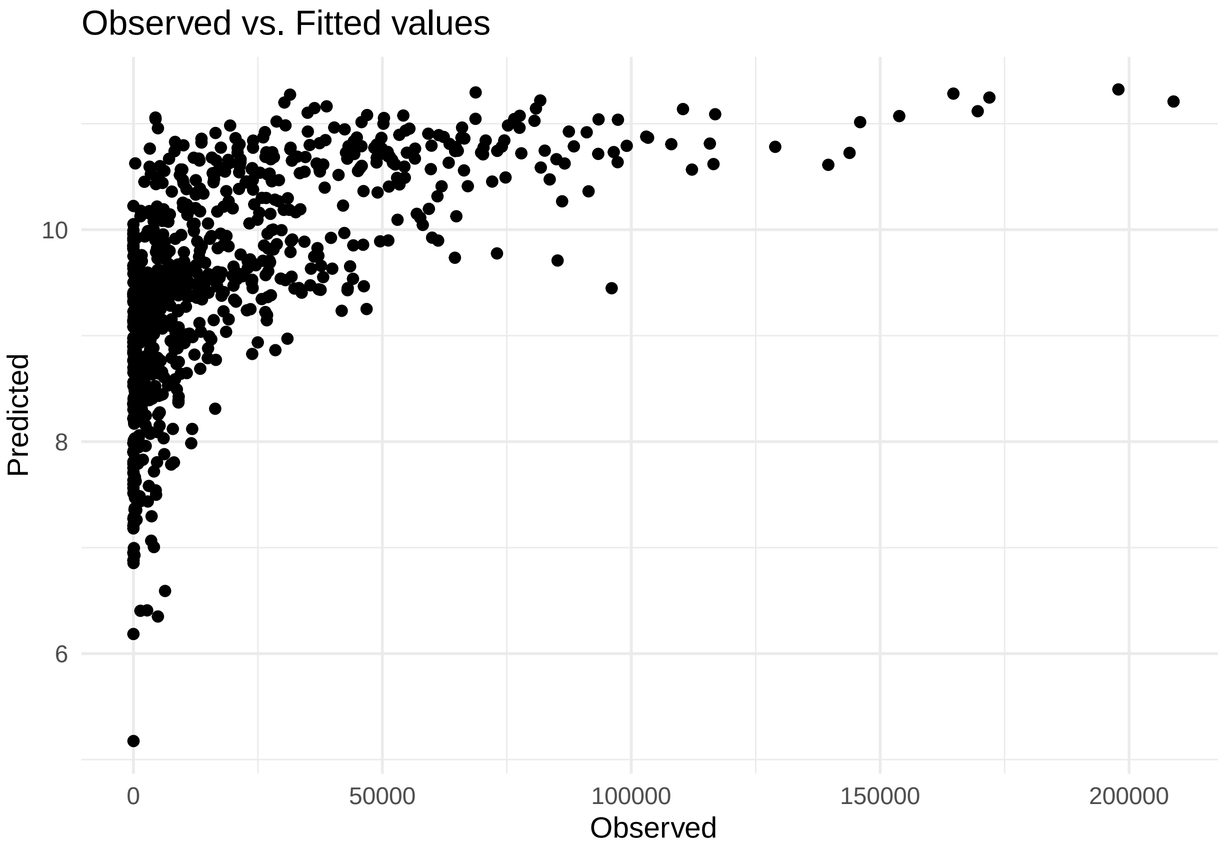


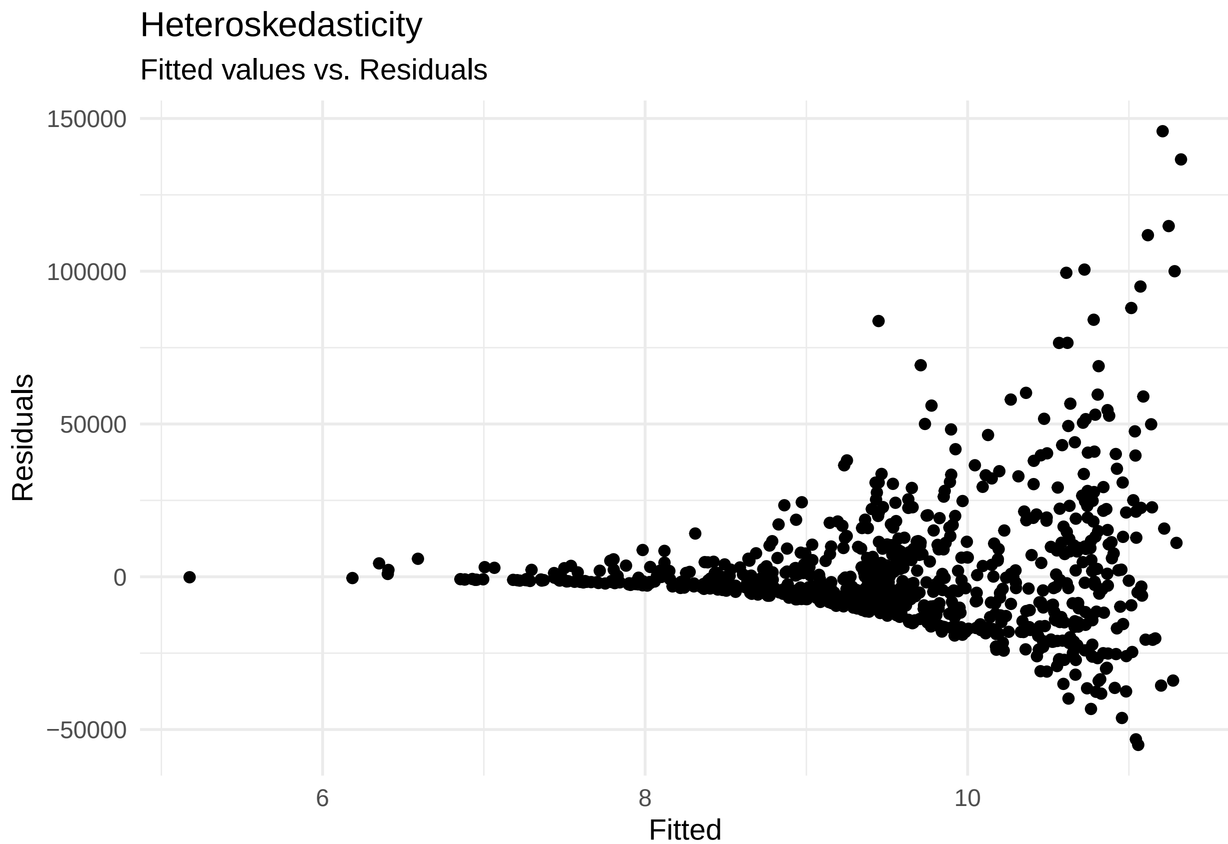
Predicted counts of predicted wesa
predicted wesa
40000
30000
20000
2000
4000
6000
8000
flow

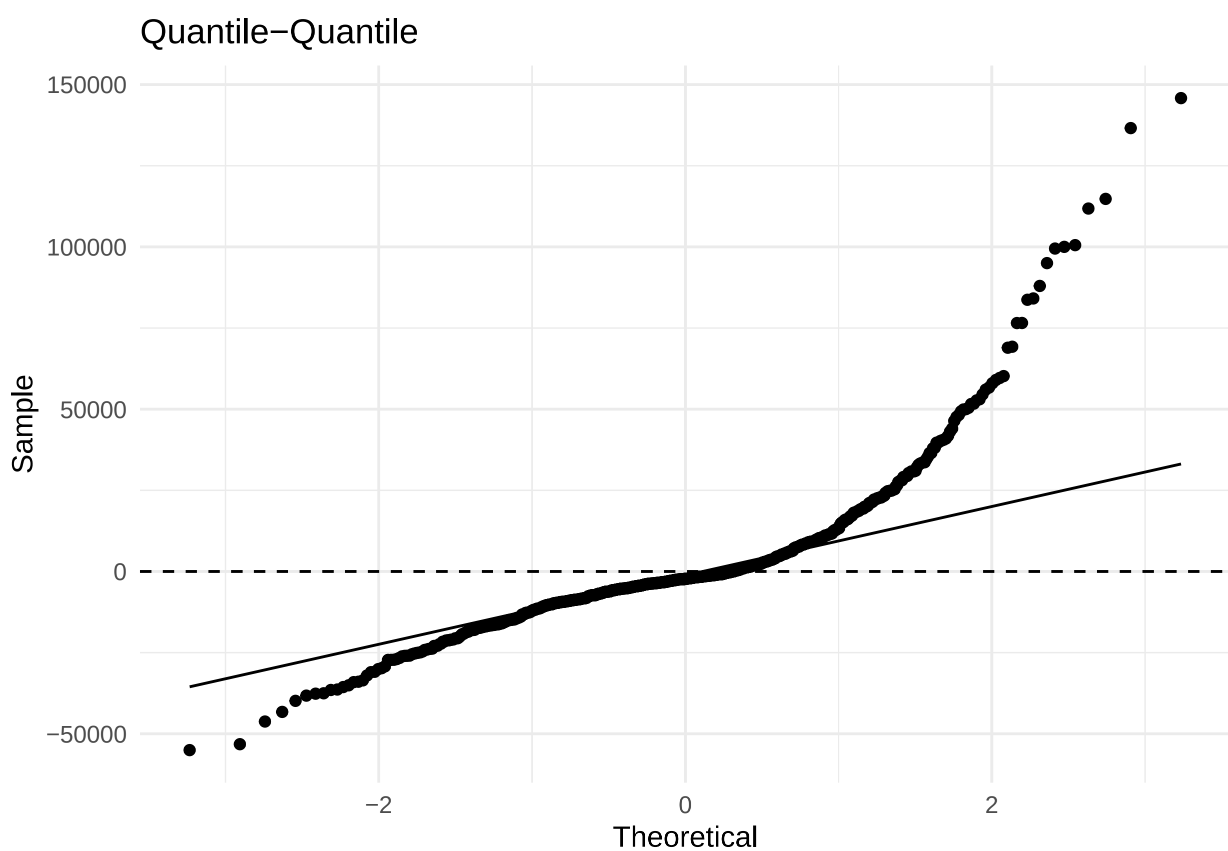

```
##
## $elev_range
```

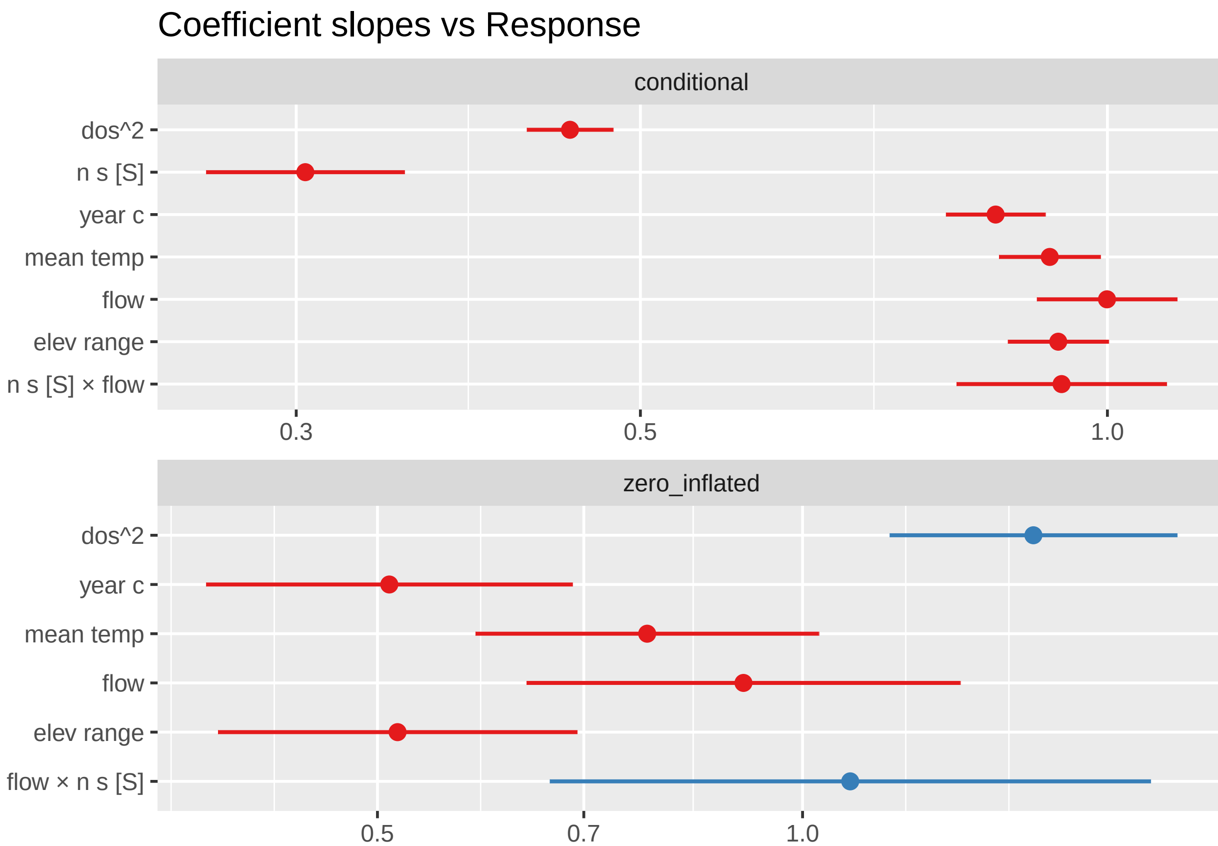

Predicted counts of predicted wesa
predicted wesa
40000
30000
20000
2
3
4
5
elev range

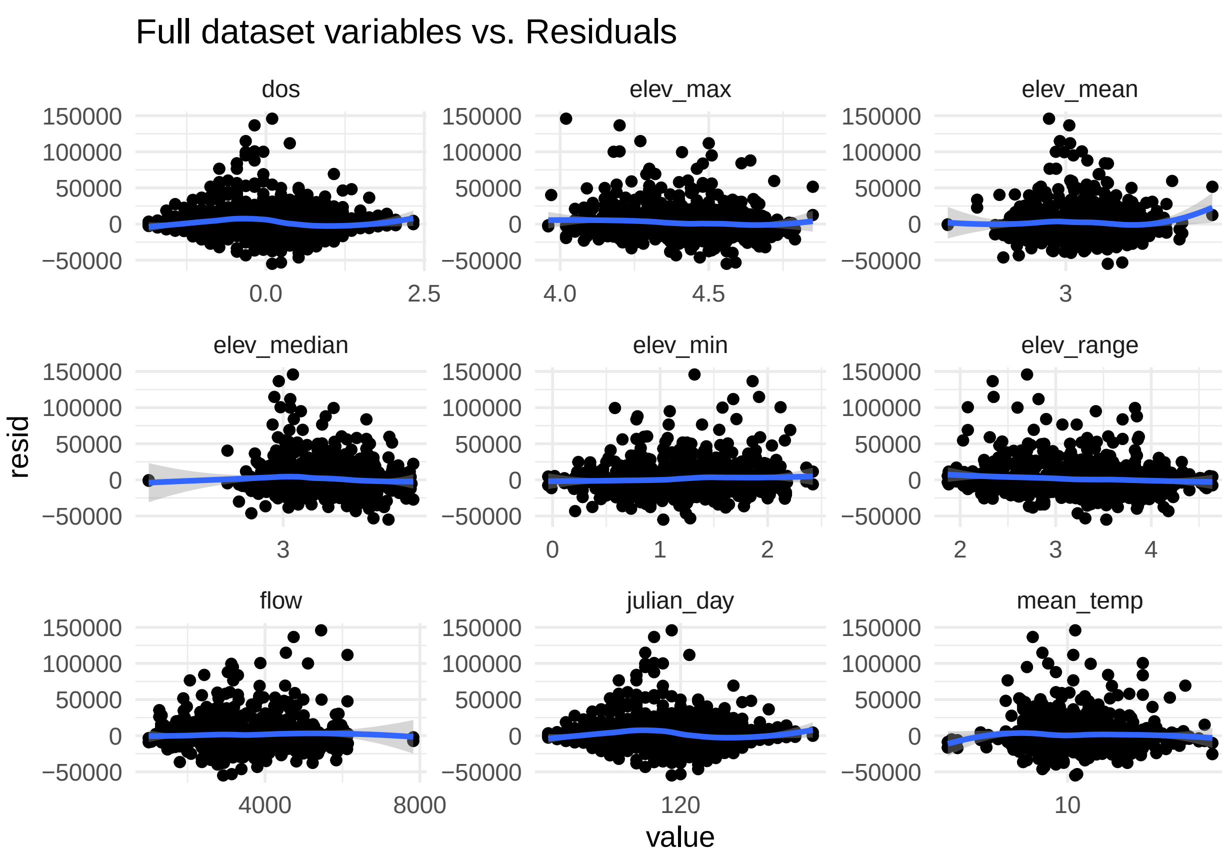
Diagnostics

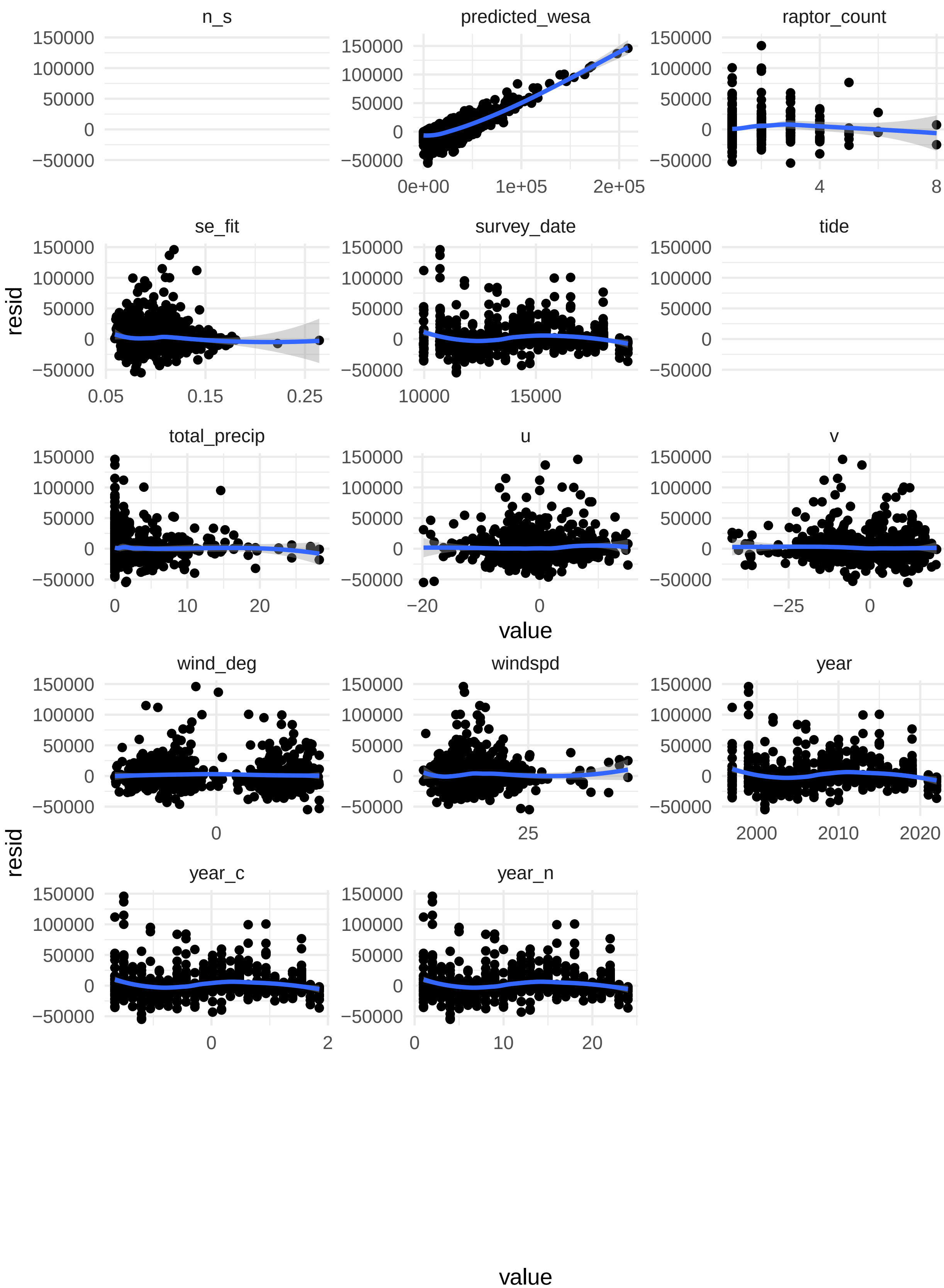



Heteroskedasticity
Fitted values vs. Residuals
150000
100000
50000
0
−50000
Residuals
6
8
10
Fitted


Quantile−Quantile
150000
100000
50000
0
−50000
Sample
−2
0
2
Theoretical


Coefficient slopes vs Response
conditional
dos^2
n s [S]
year c
mean temp
flow
elev range
n s [S] × flow
0.3
0.5
1.0
zero_inflated
dos^2
year c
mean temp
flow
elev range
flow × n s [S]
0.5
0.7
1.0


Full dataset variables vs. Residuals
dos
elev_max
elev_mean
elev_median
elev_min
elev_range
flow
julian_day
mean_temp
150000
100000
50000
0
−50000
resid
0.0
2.5
4.0
4.5
3
3
0
1
2
2
3
4
4000
8000
120
10
value



n_s
predicted_wesa
raptor_count
se_fit
survey_date
tide
total_precip
u
v
wind_deg
windspd
year
year_c
year_n
resid
value