

Brunswick Point Peep Models

Sarah Popov

2022-11-29

Roberts Bank, Delta, BC, is situated in the great Pacific Flyway and serves as an important stopover for peeps migrating north in the spring. It therefore hosts a large seasonal population of peeps: namely, Western sandpiper (WESA) and Dunlin (DUNL), that rely on the seasonal nutritional bounty provided by the Fraser River delta.

This document describes the suite of models used to estimate yearly changes in spatial distribution and abundance of peeps in the Roberts Bank estuary.

There are two datasets used in this modelling pipeline:

1. `sr`, which contains species ratio data (WESA:DUNL)
2. `dat`, which contains bird counts + environmental covariates

For an interactive interface with the `dat` dataset used in this document, see <https://popovs.shinyapps.io/peepr/>.

Species composition model

The daily ratio of Western sandpiper (WESA) to Dunlin (DUNL) across the entire study period is first modelled using a dataset of known species ratios (species ratios are not measured during every survey).

The ratios are modelled using a binomial generalized linear mixed model (binomial GLMM). The resulting predicted ratios are then used to estimate the number of WESA vs. DUNL per day.

sr data summary

```
## survey_date          wesa          dunl          total
## Min.   :1997-04-21   Min.    :  0.0   Min.    :  0.0   Min.    :  48.0
## 1st Qu.:2005-04-20   1st Qu.: 126.8   1st Qu.:  42.0   1st Qu.: 287.2
## Median :2009-04-23   Median : 355.0   Median :  93.0   Median : 524.5
## Mean   :2009-09-05   Mean    : 470.1   Mean    : 176.7   Mean    : 646.8
## 3rd Qu.:2015-04-16   3rd Qu.: 719.0   3rd Qu.: 203.5   3rd Qu.: 943.8
## Max.   :2022-05-04   Max.    :2605.0   Max.    :5000.0   Max.    :5047.0
##
##          p_wesa          p_dunl          year          julian_day
## Min.    :  0.00   Min.    :  0.00   2006    : 23   Min.    :101.0
## 1st Qu.: 53.04   1st Qu.:  8.83   2019    : 23   1st Qu.:112.0
## Median : 81.32   Median : 18.68   2005    : 18   Median :117.0
## Mean    : 68.94   Mean    : 31.06   2007    : 18   Mean    :117.6
## 3rd Qu.: 91.17   3rd Qu.: 46.96   2008    : 18   3rd Qu.:123.0
## Max.    :100.00   Max.    :100.00   2012    : 18   Max.    :137.0
##
##                                (Other):200
##
##          dos.V1
## Min.    : -2.2763470
## 1st Qu.: -0.7798519
## Median : -0.0996268
## Mean    : -0.0221924
## 3rd Qu.:  0.7166433
## Max.    :  2.6212735
##
```

Five models are built and compared

Response variable:

- y - WESA:DUNL ratio

Predictor variables:

- dos - day of season (recentered/scaled Julian date)
- year - year of survey

```
## Data: sr
## Models:
## lme5: y ~ 1 + (1 | year)
## lme4: y ~ dos + (1 | year)
## lme2: y ~ dos + I(dos^2) + (1 | year)
```

```
## lme3: y ~ dos + (dos | year)
## lme1: y ~ dos + I(dos^2) + (dos + I(dos^2) | year)
##      npar    AIC    BIC logLik deviance   Chisq Df Pr(>Chisq)
## lme5     2 64756 64764 -32376    64752
## lme4     3 25804 25815 -12899    25798 38954.1  1    <2e-16 ***
## lme2     4 16295 16310  -8144    16287  9510.6  1    <2e-16 ***
## lme3     5 22521 22540 -11255    22511    0.0  1         1
## lme1     9 11083 11117  -5532    11065 11446.2  4    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

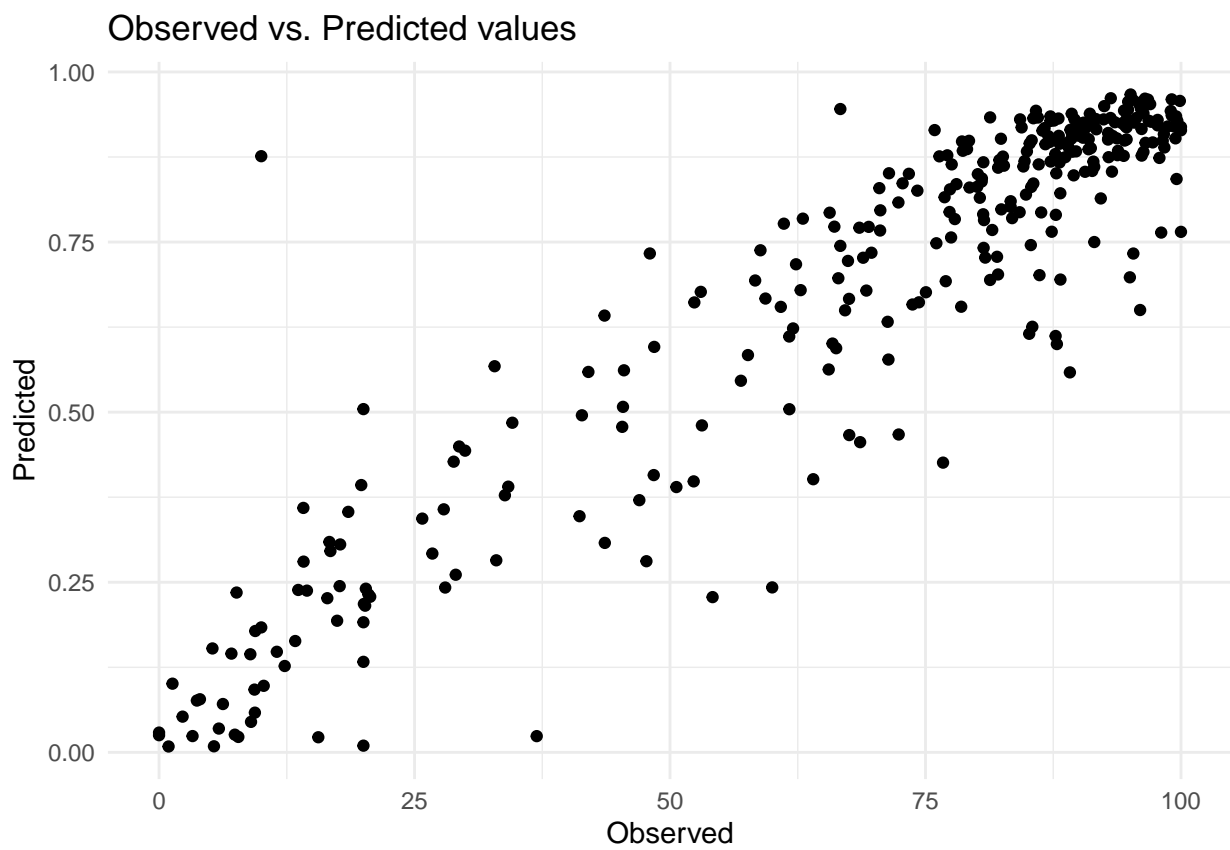
The best-fit model is `lme1`. The residuals from `lme1` are appended to the `sr` dataset and another model is re-fit in order to estimate overdispersion. Because the standard deviation of the residuals is < 1 , the model is deemed an appropriate candidate for predicting daily species ratios.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: y ~ dos + I(dos^2) + (dos + I(dos^2) | year) + (1 | resid)
## Data: sr
##      AIC      BIC    logLik deviance df.resid
## 3409.879 3447.499 -1694.939  3389.879      308
## Random effects:
## Groups Name      Std.Dev. Corr
## resid (Intercept) 0.94716
## year  (Intercept) 0.65139
##      dos          0.06454 -1.00
##      I(dos^2)      0.41479 -0.65 0.68
## Number of obs: 318, groups: resid, 318; year, 24
## Fixed Effects:
## (Intercept)      dos      I(dos^2)
##      1.8826      1.3770     -0.8085
```

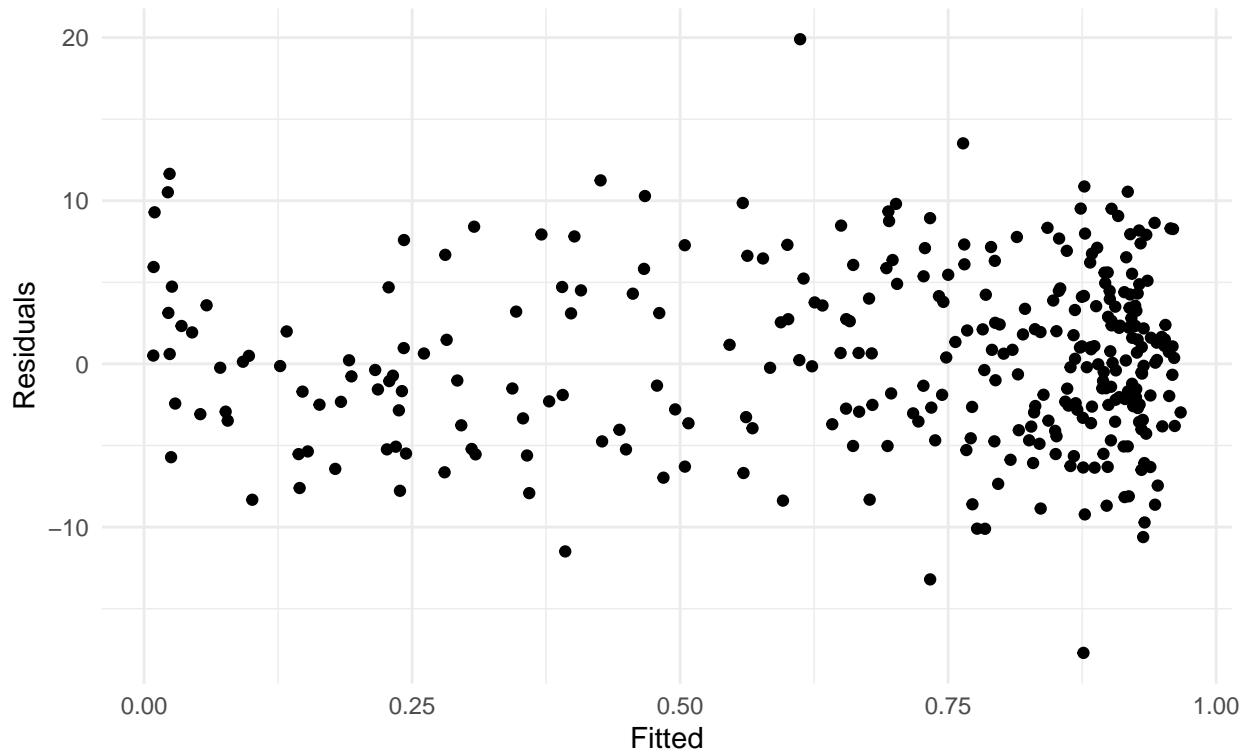
Summary of best fit model (lme1)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: y ~ dos + I(dos^2) + (dos + I(dos^2) | year)
## Data: sr
##
##      AIC      BIC   logLik deviance df.resid
## 11082.7 11116.6 -5532.4 11064.7      309
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -23.5674  -3.5412   0.2172   3.6901  21.7396
##
## Random effects:
##   Groups Name      Variance Std.Dev. Corr
##   year  (Intercept) 0.6045   0.7775
##         dos         0.4629   0.6803   0.18
##         I(dos^2)    0.6650   0.8155  -0.71 -0.53
## Number of obs: 318, groups: year, 24
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.7477     0.1593  10.974 < 2e-16 ***
## dos           1.3365     0.1398   9.558 < 2e-16 ***
## I(dos^2)      -0.8168     0.1676  -4.874 1.09e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) dos
## dos       0.177
## I(dos^2) -0.705 -0.522
```

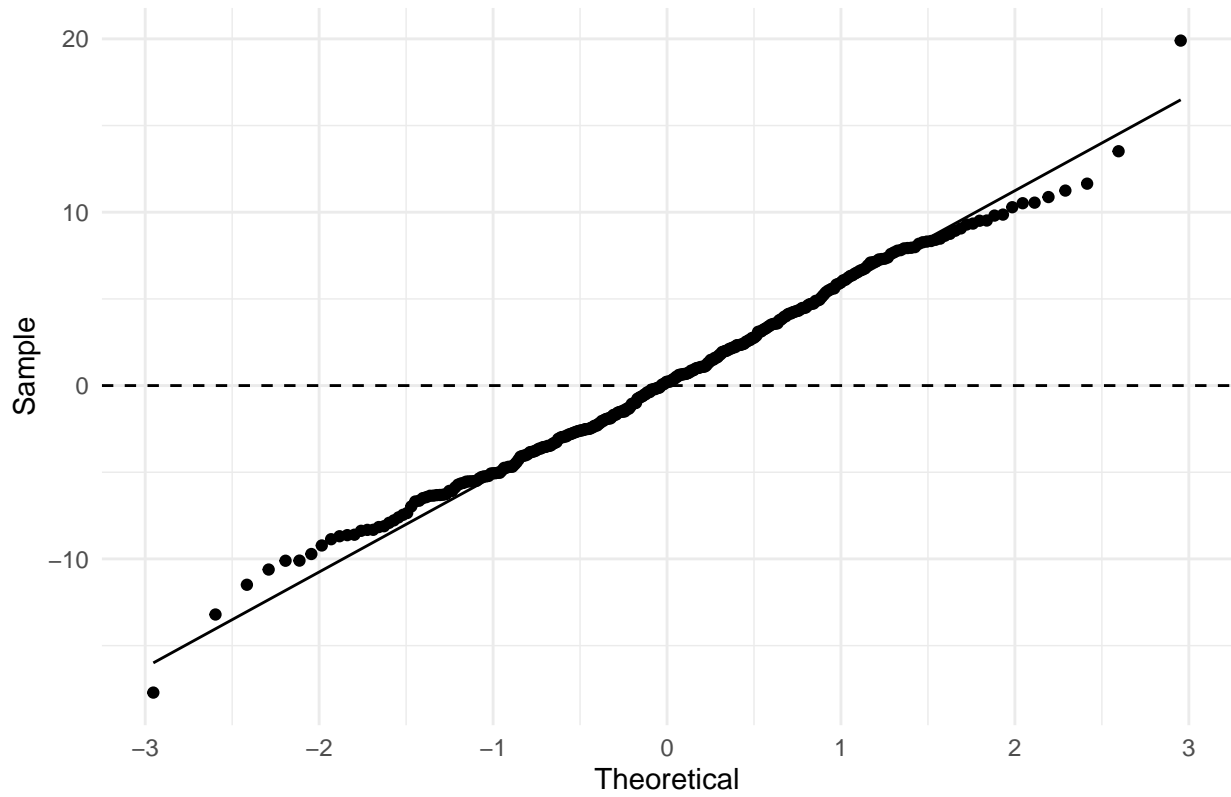
Check assumptions of best fit model

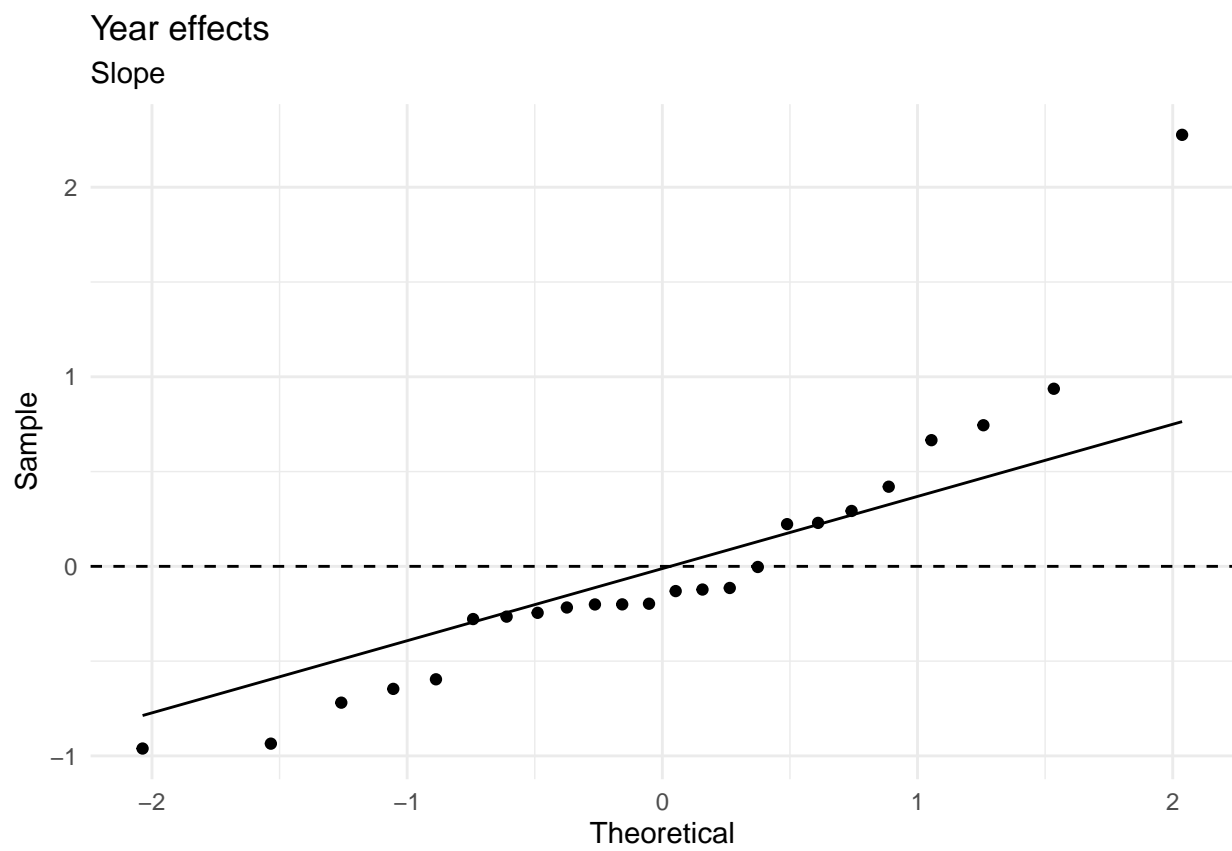
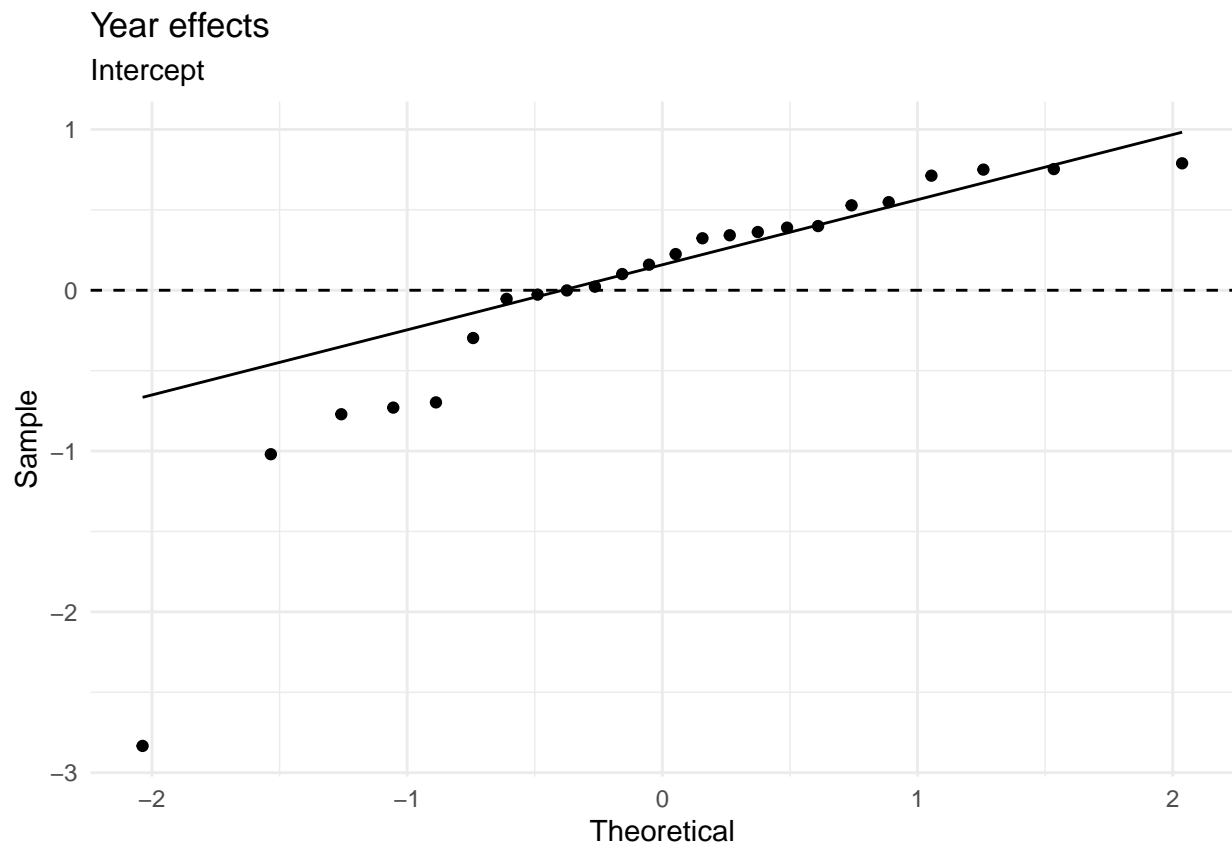


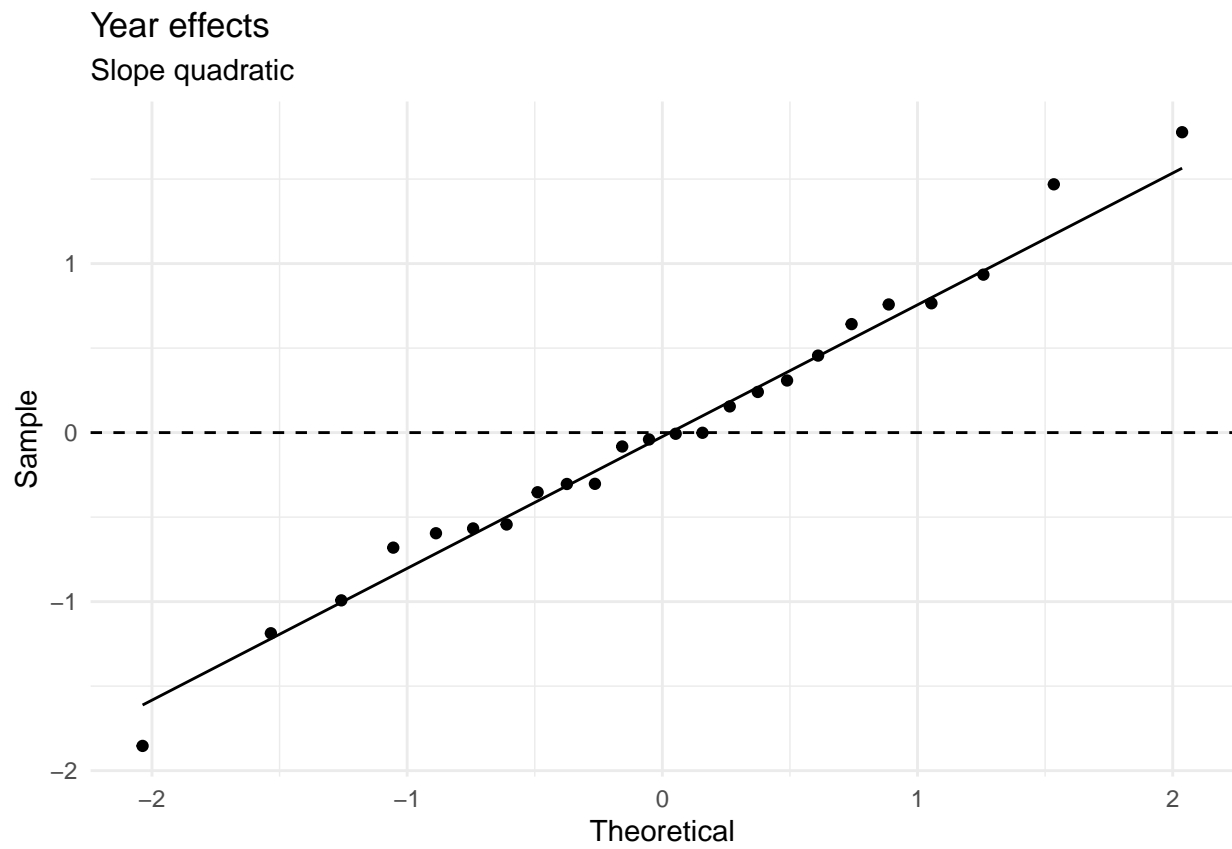
Heteroskedasticity
Fitted values vs. Residuals



Quantile–Quantile

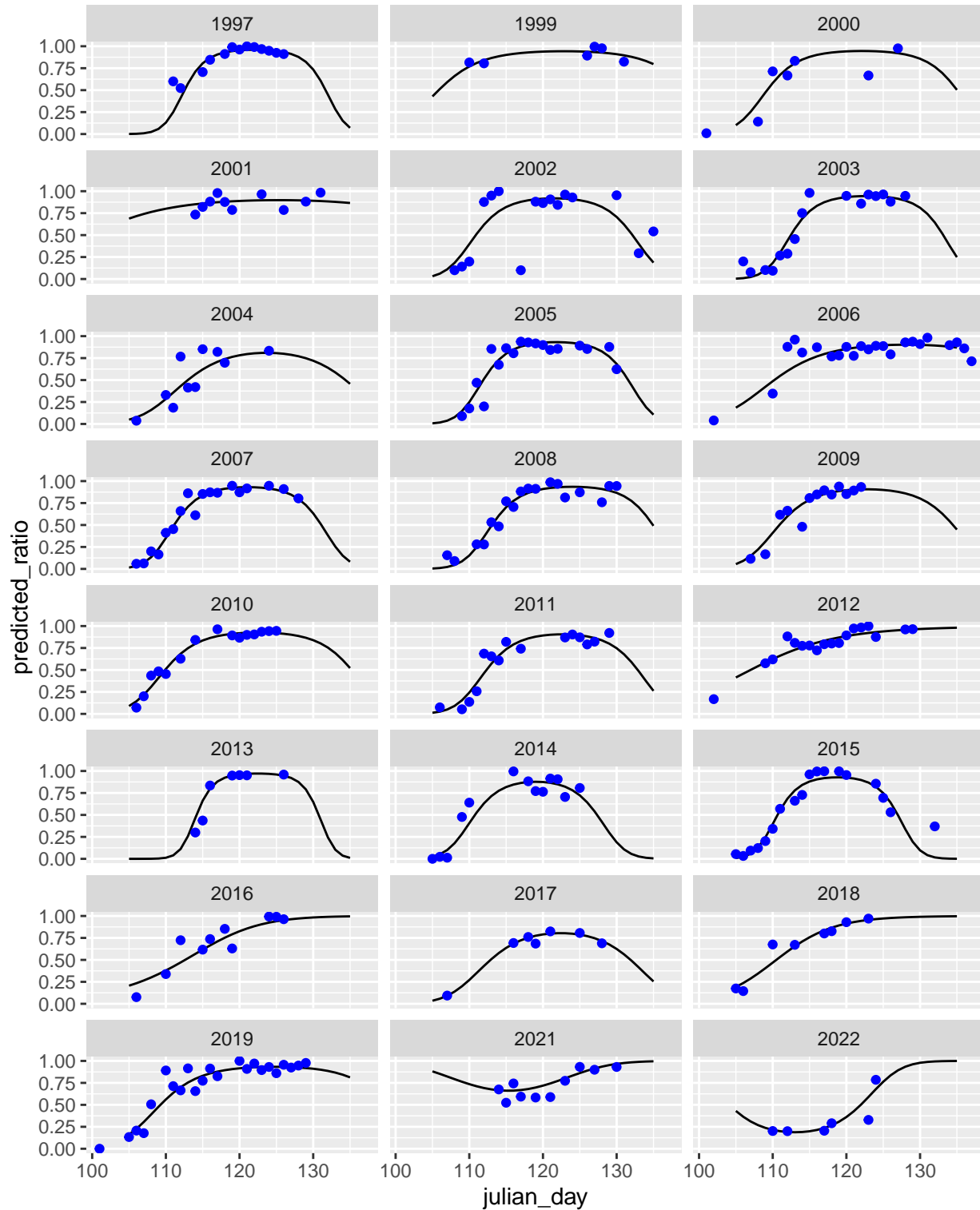






Predict WESA/DUNL population

Using the derived binomial GLMM above, we will predict the amount of WESA and DUNL each day. For any years that are missing from the bGLMM we will assign the mean proportion of WESA:DUNL.



North-South distribution model

The primary goal of this analysis is to determine whether melt flow regimes alter the north-south distribution of peeps. Birds are surveyed from a route of standardized survey stations. Any birds counted in or north of the “view corner” are labelled as **N**, while any birds counted south of the “view corner” are labelled as **S** for this analysis.¹

The **dat** dataset, which contains survey data and associated environmental covariates, will be used in this model. The **dat** dataset can be explored in greater detail in the PeepR Shiny app.

¹For now, this includes birds counted in Canoe Pass, but those will be excluded in a later analysis to compare results.

dat filtering

The **dat** dataset underwent some filtering steps to exclude unwanted data prior to this analysis. First, only data that was originally included in Canham et al. (2021) is extracted from the **bppeeps** database. While survey data oftentimes includes multiple sweeps, low-quality or reconnaissance sweeps were excluded from the daily totals used in Canham et al. (2021). As such, they were excluded from this dataset as well.

The total bird count for several surveys were obtained by taking the average total count from multiple sweeps. In these cases, it is difficult to obtain accurate location-level numbers because birds often moved locations between sweeps. These records are excluded for now (45 survey dates out of a total of 538).

Additionally, certain survey dates only included a total bird count for the day, but no location-level information. These records were excluded from the initial database query (24 survey dates out of a total of 538).

After excluding these surveys, the initial dataset queried from the **bppeeps** dataset included 2207 records from a total of 469. This is termed the ‘full dataset’ and was filtered further in **R**.

Filtering step	No. survey dates	No. records	No. survey dates lost	No. records lost
Full dataset	469	2207	NA	NA
Remove NA count records	469	2198	0	-9
Exclude dates where total # of birds < 1000	443	2085	-26	-113
Exclude dates outside of survey period (<04-15 or >05-15)	430	2023	-13	-62
Exclude Intercauseway and NA stations (e.g. location was simply ‘inner mud’, ‘mumbles’, ‘flying’, etc.)	429	1892	-1	-131
Exclude records where only bird count occurs in location that spans >2 stations (e.g., ‘BP to CP’)	425	1882	-4	-10

dat data summary after filtering

```
##      survey_date      start_time      station
## Min.   :1997-04-21   Length:1882      Length:1882
## 1st Qu.:2004-04-19   Class :character   Class :character
## Median :2008-05-07   Mode  :character   Mode  :character
## Mean   :2009-03-12
## 3rd Qu.:2014-04-18
## Max.   :2022-05-04
##
##      station_n      station_s      mumblies_yn      mud_yn
## Coal Port      :536   Bend           : 70   Mode :logical   Mode :logical
## Pilings        :465   Coal Port      : 55   FALSE:1700     FALSE:1697
## Brunswick Point:311   View corner   : 45   TRUE :182      TRUE :185
## 34th St pullout:204   34th St pullout: 17
## View corner    :174   Pilings        : 15
## Bend           :118   (Other)        : 1
## (Other)        : 74   NA's           :1679
## marsh_yn      tide_edge_yn      flying_yn      final_count
## Mode :logical   Mode :logical   Mode :logical   Min.   : 0
## FALSE:1815     FALSE:1807     FALSE:1866     1st Qu.: 500
## TRUE :67       TRUE :75       TRUE :16       Median : 4125
##                                     Mean    : 12745
##                                     3rd Qu.: 14500
##                                     Max.    : 222500
##
##      p_wesa      elev_min      elev_max      elev_median
## Min.   : 0.00   Min.   :-0.040   Min.   :3.960   Min.   :2.240
## 1st Qu.: 52.32   1st Qu.: 0.830   1st Qu.:4.300   1st Qu.:3.071
## Median : 80.66   Median : 1.210   Median :4.430   Median :3.255
## Mean   : 68.69   Mean   : 1.223   Mean   :4.418   Mean   :3.239
## 3rd Qu.: 90.97   3rd Qu.: 1.620   3rd Qu.:4.560   3rd Qu.:3.415
## Max.   :100.00   Max.   : 2.420   Max.   :4.850   Max.   :3.735
## NA's    :589
##      elev_mean      elev_range      flow      total_precip      mean_temp
## Min.   :2.660   Min.   :1.870   Min.   : 996   Min.   : 0.00   Min.   : 3.80
## 1st Qu.:2.967   1st Qu.:2.712   1st Qu.:2430   1st Qu.: 0.00   1st Qu.: 9.20
## Median :3.037   Median :3.210   Median :3090   Median : 0.00   Median :10.50
## Mean   :3.041   Mean   :3.195   Mean   :3227   Mean   : 2.09   Mean   :10.64
## 3rd Qu.:3.118   3rd Qu.:3.607   3rd Qu.:3990   3rd Qu.: 2.00   3rd Qu.:11.90
## Max.   :3.422   Max.   :4.640   Max.   :7830   Max.   :28.20   Max.   :17.50
##                                     NA's    :5      NA's    :4
##      u      v      windspd      wind_deg
## Min.   :-19.8067   Min.   :-42.3467   Min.   : 5.292   Min.   :-177.35
## 1st Qu.: -3.8058   1st Qu.: -9.9586   1st Qu.:10.542   1st Qu.: -77.07
## Median : -1.0261   Median : -0.4341   Median :12.958   Median : -22.79
## Mean   : -0.7134   Mean   : -1.8065   Mean   :14.309   Mean   : 15.03
## 3rd Qu.: 2.5179   3rd Qu.: 8.0889   3rd Qu.:16.906   3rd Qu.: 117.95
## Max.   : 15.0489   Max.   : 20.4541   Max.   :43.750   Max.   : 179.93
##
##      station_n_no      station_s_no      station_diff      year      julian_day
## Min.   :1.000   Min.   :3.000   Min.   :0.00   2008 : 124   Min.   :105
## 1st Qu.:4.000   1st Qu.:5.000   1st Qu.:1.00   2019 : 121   1st Qu.:112
## Median :5.000   Median :6.000   Median :1.00   2012 : 120   Median :118
```

```

## Mean      :5.554   Mean      :6.094   Mean      :1.01   2005      : 116   Mean      :118
## 3rd Qu.   :8.000   3rd Qu.   :8.000   3rd Qu.   :1.00   2001      : 109   3rd Qu.   :124
## Max.      :8.000   Max.      :8.000   Max.      :2.00   2006      :  95   Max.      :135
##              NA's    :1679   NA's      :1679   (Other):1197
##      dos.V1      n_s
## Min.      :-1.8238057   N: 559
## 1st Qu.   :-0.8404703   S:1323
## Median    : 0.0023885
## Mean      : 0.0000000
## 3rd Qu.   : 0.8452474
## Max.      : 2.3904887
##

```

Build and test model

In this initial analysis, the model will be kept relatively simple: use the same model as in Canham et al. (2021), but add a “north vs. south” term. The response variable will be the total count for that location/day * the predicted WESA ratio generated by the bGLMM above.

Base model

Response variable:

- `log_wesa` - Log-transformed predicted WESA count

Predictor variables:

- `year_c` - Scale-transformed survey year
- `dos` - Scale-transformed Julian date, aka Day of Season

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: log_wesa ~ year_c + dos + I(dos^2) + (dos + I(dos^2) | year)
##   Data: dat
##
## REML criterion at convergence: 9666.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.9584 -0.3512  0.2321  0.6717  2.0999
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   year      (Intercept)  0.4644    0.6815
##             dos          0.4704    0.6859    0.05
##             I(dos^2)     0.1626    0.4033   -0.51  0.03
## Residual                9.5098    3.0838
## Number of obs: 1882, groups: year, 24
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  7.91981    0.17559 14.09445  45.104 < 2e-16 ***
## year_c       0.43653    0.13529 21.39075   3.227  0.00398 **
## dos          0.05938    0.16521 17.09441   0.359  0.72368
## I(dos^2)     -1.28574    0.12112 11.08547 -10.615 3.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) year_c dos
## year_c      -0.049
## dos          0.034  0.025
## I(dos^2)    -0.577  0.018  0.031
```

Full model

NOTE: at the time of this writing, IR is *unavailable* in this model. I have not gotten any response from the team that manages the UBC Totem Station data.

Response variable:

- `log_wesa` - Log-transformed predicted WESA count

Predictor variables:

- `year_c` - Scale-transformed survey year
- `dos` - Scale-transformed Julian date, aka Day of Season
- `elev_range` - Tidal amplitude (m)
- `total_precip` - Total daily precipitation (mm)
- `mean_temp` - Daily mean temperature (C°)
- `flow` - Fraser River discharge (m³/s)
- `u, v` - Westerly and Southerly wind vectors (km/h)
- `n_s` - Location ('North' or 'South')

Additionally, the dataset is reduced down to only complete cases of all predictor variables of interest (1882 -> 1873).

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## log_wesa ~ year_c + dos + I(dos^2) + scale(elev_range) + scale(total_precip) +
##   scale(mean_temp) + scale(flow) + scale(u) + scale(v) + n_s +
##   (dos + I(dos^2) | year) + (scale(flow) | n_s)
##   Data: dat2
##
## REML criterion at convergence: 9295.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.6037 -0.3664  0.1907  0.6530  2.6563
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   year     (Intercept)  0.62485   0.7905
##           dos          0.58201   0.7629    0.02
##           I(dos^2)     0.18182   0.4264   -0.49  0.18
##   n_s      (Intercept)  6.57361   2.5639
##           scale(flow)  0.07199   0.2683    0.99
## Residual                7.85024   2.8018
## Number of obs: 1873, groups:  year, 24; n_s, 2
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    8.151e+00  1.950e+00  1.950e-08   4.180   1.0000
## year_c         4.857e-01  1.488e-01  2.043e+01   3.265   0.0038 **
## dos           -7.627e-04  2.049e-01  3.067e+01  -0.004   0.9971
## I(dos^2)       -1.339e+00  1.219e-01  1.169e+01 -10.990  1.63e-07 ***
```

```

## scale(elev_range)      3.994e-01  7.721e-02  9.786e+02   5.173 2.79e-07 ***
## scale(total_precip)    4.194e-02  7.384e-02  1.774e+03   0.568  0.5701
## scale(mean_temp)       1.680e-01  8.911e-02  9.056e+02   1.885  0.0597 .
## scale(flow)            1.741e-02  2.457e-01  2.140e+00   0.071  0.9496
## scale(u)               8.213e-02  8.665e-02  1.377e+03   0.948  0.3434
## scale(v)               2.105e-01  8.887e-02  1.581e+03   2.368  0.0180 *
## n_sS                   6.674e-01  1.376e+00  1.977e-08   0.485  1.0000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) year_c dos      I(d^2) scl(l_) scl(t_) scl(m_) scl(f) scal()
## year_c      -0.004
## dos          0.011  0.024
## I(dos^2)     -0.052  0.019  0.051
## scl(lv_rng) -0.001  0.012 -0.050  0.036
## scl(ttl_pr)  0.008 -0.009 -0.013  0.014  0.044
## scl(mn_tmp)  0.000 -0.061 -0.151  0.003 -0.070  0.068
## scale(flow)  0.667 -0.005 -0.306  0.074  0.035  0.020  0.014
## scale(u)     0.000 -0.039  0.021 -0.068 -0.102 -0.042  0.071  0.025
## scale(v)     -0.006 -0.015  0.042 -0.066  0.018 -0.251 -0.079  0.000  0.522
## n_sS         -0.354 -0.002 -0.011 -0.007 -0.002 -0.023 -0.002  0.104  0.010
##      scl(v)
## year_c
## dos
## I(dos^2)
## scl(lv_rng)
## scl(ttl_pr)
## scl(mn_tmp)
## scale(flow)
## scale(u)
## scale(v)
## n_sS          0.025
## optimizer (nloptwrap) convergence code: 0 (OK)
## unable to evaluate scaled gradient
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues

```

Backwards step-wise selection

```

## Backward reduced random-effect table:
##
##      Eliminated npar  logLik    AIC    LRT Df
## <none>                21 -4647.7 9337.4
## dos in (dos + I(dos^2) | year)      0  18 -4668.6 9373.3 41.822  3
## I(dos^2) in (dos + I(dos^2) | year)  0  18 -4651.2 9338.4  6.943  3
## scale(flow) in (scale(flow) | n_s)   0  19 -4650.2 9338.4  4.964  2
##      Pr(>Chisq)
## <none>
## dos in (dos + I(dos^2) | year)      4.376e-09 ***
## I(dos^2) in (dos + I(dos^2) | year)  0.07373 .
## scale(flow) in (scale(flow) | n_s)   0.08360 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



```

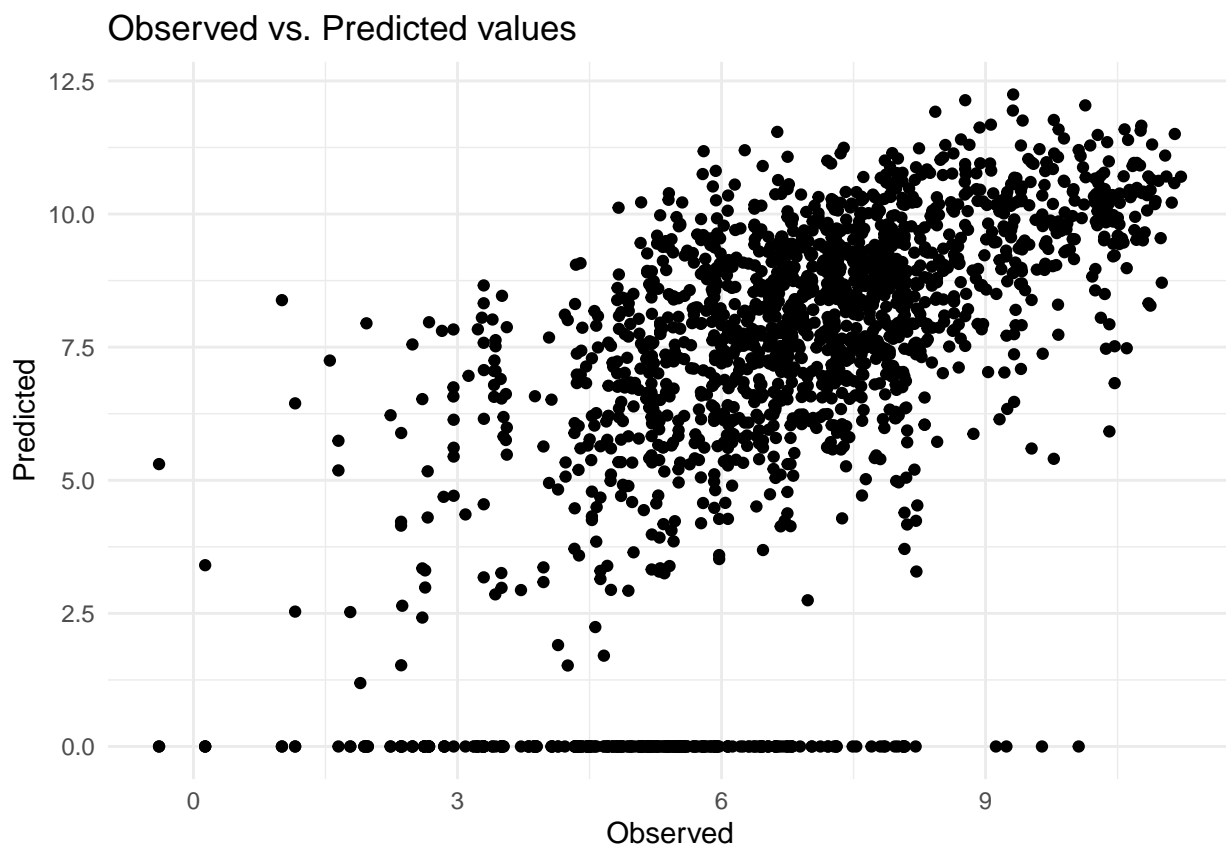
##
## Backward reduced fixed-effect table:
## Degrees of freedom method: Satterthwaite
##
##           Eliminated Sum Sq Mean Sq NumDF   DenDF F value   Pr(>F)
## n_s           1    1.85    1.85     1     0.00   0.2352  1.000000
## scale(flow)    2    0.09    0.09     1     1.97   0.0116  0.924327
## scale(total_precip) 3    2.49    2.49     1 1773.49   0.3179  0.572966
## scale(u)       4    7.46    7.46     1 1405.21   0.9517  0.329455
## scale(mean_temp) 5   24.48   24.48     1  961.84   3.1231  0.077505
## year_c        0   78.39   78.39     1   20.45  10.0000  0.004806
## dos          0    0.93    0.93     1   22.76   0.1188  0.733482
## I(dos^2)      0 869.36  869.36     1   11.26 110.9087 3.584e-07
## scale(elev_range) 0 227.24  227.24     1 1067.75  28.9905 8.942e-08
## scale(v)      0   57.11   57.11     1 1663.84   7.2857  0.007021
##
## n_s
## scale(flow)
## scale(total_precip)
## scale(u)
## scale(mean_temp)  .
## year_c           **
## dos
## I(dos^2)         ***
## scale(elev_range) ***
## scale(v)         **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Model found:
## log_wesa ~ year_c + dos + I(dos^2) + scale(elev_range) + scale(v) + (dos + I(dos^2) | year) + (scale

```

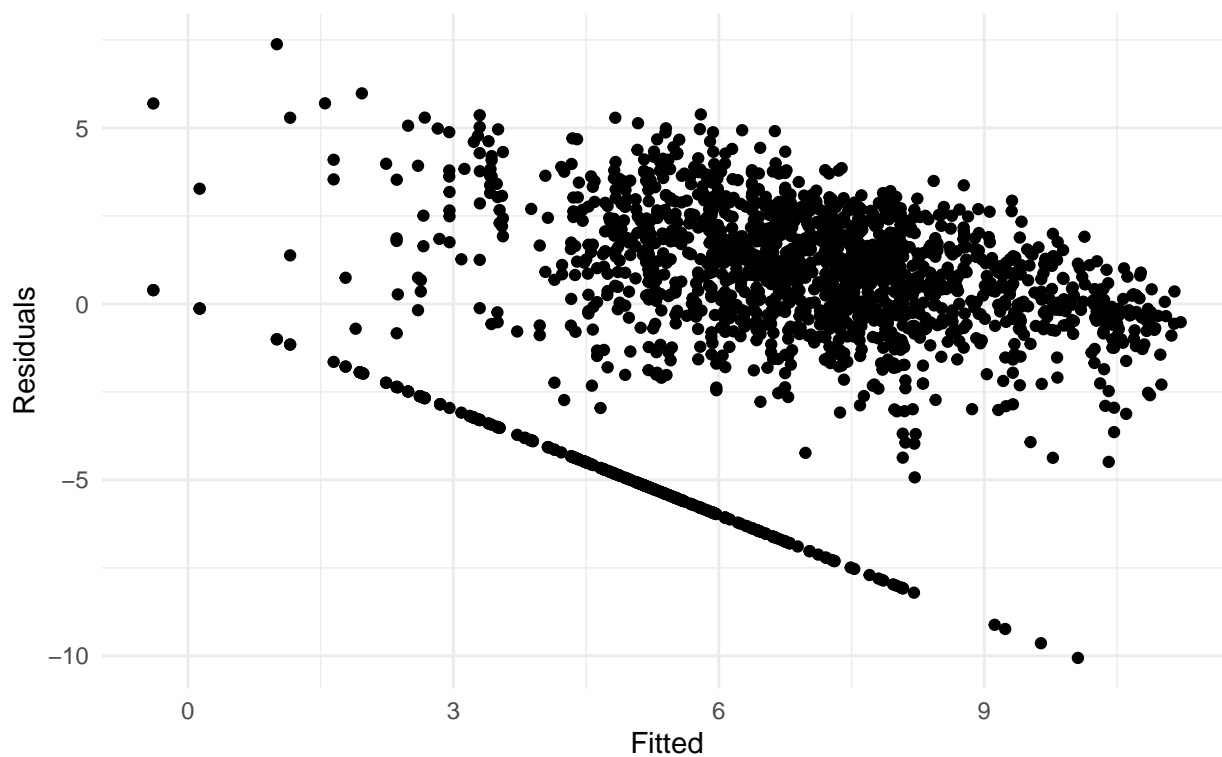
Summary of best-fit model

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: log_wesa ~ year_c + dos + I(dos^2) + scale(elev_range) + scale(v) +
##      (dos + I(dos^2) | year) + (scale(flow) | n_s)
## Data: dat2
##
## REML criterion at convergence: 9290.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5926 -0.3549  0.1977  0.6622  2.6351
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## year    (Intercept)  0.66453   0.8152
##          dos         0.56830   0.7539  -0.06
##          I(dos^2)    0.21228   0.4607  -0.51  0.14
## n_s      (Intercept)  2.40110   1.5495
##          scale(flow) 0.04795   0.2190  1.00
## Residual              7.83853   2.7997
## Number of obs: 1873, groups: year, 24; n_s, 2
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    8.48259    0.81189    4.70595  10.448 0.000196 ***
## year_c         0.47802    0.15116   20.44910   3.162 0.004806 **
## dos            0.06505    0.18871   22.76153   0.345 0.733482
## I(dos^2)       -1.33576    0.12684   11.26124 -10.531 3.58e-07 ***
## scale(elev_range) 0.41344    0.07679 1067.75043   5.384 8.94e-08 ***
## scale(v)        0.19582    0.07255 1663.83794   2.699 0.007021 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) year_c dos      I(d^2) scl( )
## year_c      -0.013
## dos          0.353  0.016
## I(dos^2)     -0.213  0.022  0.068
## scl(lv_rng) -0.043  0.008 -0.048  0.026
## scale(v)     0.011  0.001  0.013 -0.032  0.091
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Check assumptions of best fit model



Heteroskedasticity
Fitted values vs. Residuals



Quantile-Quantile

