Scripps's Murrelet Egg Size Model

Amelia J. DuVall & Marcela Todd Zaragoza

2021

This is v.2021-12-30

Introduction

This document includes code for likelihood ratio tests to test inclusion of random effects, model selection, and model diagnostics for a linear-mixed model of Scripps's Murrelet *Synthliboramphus scrippsi* egg size as a function of egg order and oceanographic indices at Santa Barbara Island, Channel Islands National Park, USA from 2009-2017.

```
knitr::opts_chunk$set(echo = TRUE)

## load libraries
library(here)
library(tidyverse)
library(janitor)
library(ggplot2)
library(lubridate)
library(viridis)
library(lme4)
library(RLRsim)
library(faraway)
library(sjPlot)
library(forecast)

## load egg size data and covariate data
egg <- read.csv(here("data", "SCMU_egg_data.csv"))
covars <- read.csv(here("data", "covariates", "covars.csv"))

## join covariate data with egg size data
SCMUdf <- left_join(egg, covars, by = "Year") %>% # join by year
  filter(TrueOrder == TRUE) %>% # select eggs with order known only
  filter(Size > 1210) %>% # remove the outliers
  dplyr::select(Year, Observer, Plot, Size, EggOrder, ANCHL, BEUTI, NPGO, ONI, PDO, SST)
```

Global Model

```
## create data frame specifying predictors to include
predictors <- as.data.frame(matrix(c(FALSE, TRUE), 2, 7))
# 7 potential predictors (includes EggOrder)
```

```r
## add column names
cov_names <- colnames(predictors) <- colnames(SCMUdf[,5:11])

## create set of all possible combinations
full_set <- expand.grid(predictors)

## select models with correlated predictors
ii <- which(full_set$ANCHL + full_set$NPGO == 2 |
              full_set$ANCHL + full_set$ONI == 2 |
              full_set$ANCHL + full_set$PDO == 2 |
              full_set$BEUTI + full_set$ONI == 2 |
              full_set$BEUTI + full_set$PDO == 2 |
              full_set$NPGO + full_set$PDO == 2 |
              full_set$NPGO + full_set$SST == 2 |
              full_set$ONI + full_set$PDO == 2)

## create reduced set of models and convert to a matrix for easier indexing
use_set <- as.matrix(full_set[-ii,])

## number of models in set
(n_mods <- nrow(use_set))
```

```
## [1] 30
```

```r
## find max number of predictors in a model
max(rowSums(use_set))
```

```
## [1] 4
```

```r
## covariates in global model
cov_names[use_set[which.max(rowSums(use_set)),]]
```

```
## [1] "EggOrder" "ANCHL"    "BEUTI"    "SST"
```

Likelihood Ratio Tests

We used likelihood ratio tests using the RLRsim package on the global model to test the support for inclusion of two random effects:

1) Plot: the monitoring plot from which the egg was obtained and measured (n = 8)

2) Observer: the person who measured the egg (n = 27)

```r
## remove NAs in Observer field (e.g., unknown Observer)
SCMUdf2 <- SCMUdf[-c(which(is.na(SCMUdf$Observer==TRUE))),]

## global model with both random effects
bm_both <- lmer(Size ~ EggOrder + ANCHL + BEUTI + SST +
                  (1 | Observer) + (1 | Plot), data = SCMUdf2, REML = FALSE)

## run model with plot random effect only
bm_plot <- lmer(Size ~ EggOrder + ANCHL + BEUTI + SST +
                  (1 | Plot), data = SCMUdf2, REML = FALSE)
```

```
## run model with observer random effect only
bm_obs <- lmer(Size ~ EggOrder + ANCHL + BEUTI + SST +
                 (1 | Observer), data = SCMUdf2, REML = FALSE)
```

boundary (singular) fit: see ?isSingular

 # Error "(boundary(singular) fit: see ?isSingular" due to very small or zero variance; does not affect

```
## Exact RLRT test
# m is the reduced model containing only the RE to be tested with
# the random effect set to zero under the null hypothesis.
# mA and Mo are the models under the alternative and the null, respectively.

# observer set to zero under the null hypothesis
exactRLRT(m = bm_obs, mA = bm_both, m0 = bm_plot, seed = 16)
```

Using restricted likelihood evaluated at ML estimators.

Refit with method="REML" for exact results.

##
simulated finite sample distribution of RLRT.
##
(p-value based on 10000 simulated values)
##
data:
RLRT = 0.52989, p-value = 0.1836

```
# plot set to zero under the null hypothesis
exactRLRT(m = bm_plot, mA = bm_both, m0 = bm_obs, seed = 16)
```

Using restricted likelihood evaluated at ML estimators.
Refit with method="REML" for exact results.

##
simulated finite sample distribution of RLRT.
##
(p-value based on 10000 simulated values)
##
data:
RLRT = 1.6371, p-value = 0.0598

Model Diagnostics

These diagnostics are done for the global model.

```
## run global model
global_mod <- lmer(Size ~ EggOrder + ANCHL + BEUTI + SST + (1 | Plot),
                   data = SCMUdf, REML = TRUE)
```

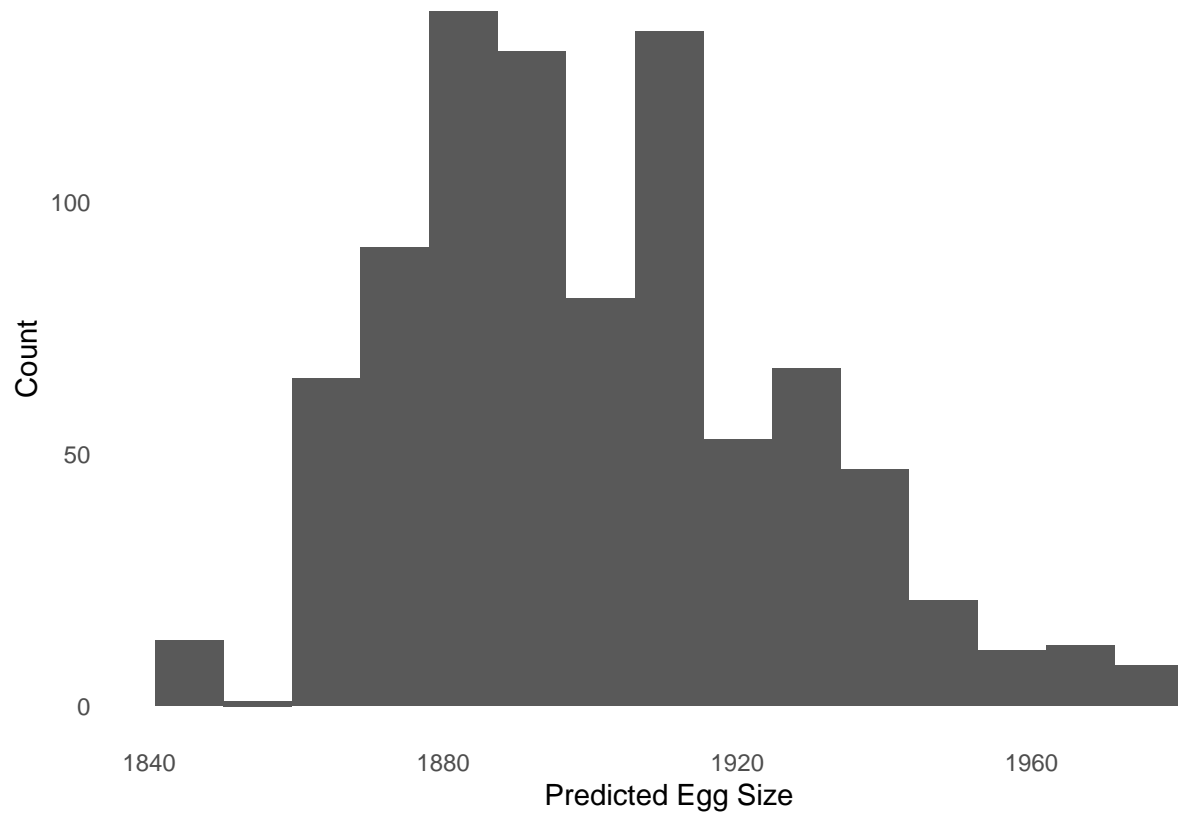### Predicted Egg Size Values

```
## extract predicted egg size values and plot
preds <- predict(global_mod)
p1 <- ggplot() +
```

```
geom_histogram(mapping = aes(preds), bins = 15) + # set bins
theme_minimal() +
xlab("Predicted Egg Size") + ylab("Count") +
theme(panel.grid.major = element_blank(),
      panel.grid.minor = element_blank())
p1
```
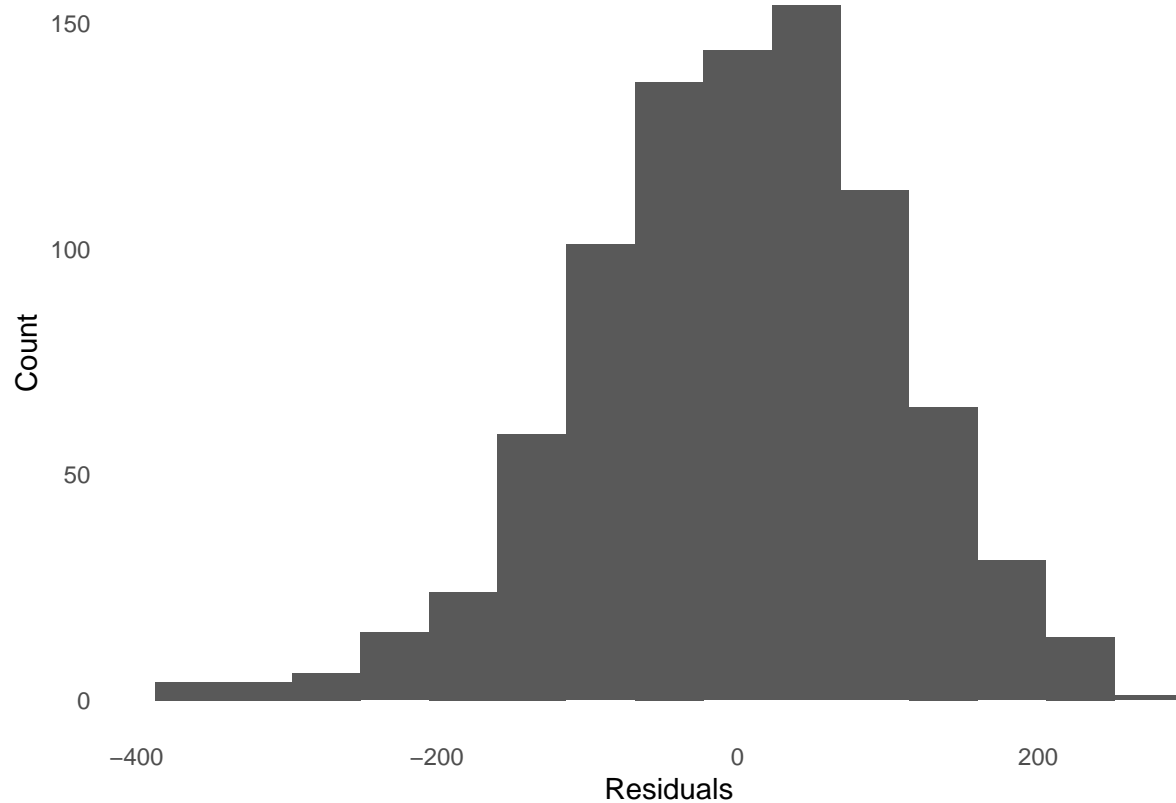

Residuals

```
## extract residuals and plot
resids <- residuals(global_mod)
p2 <- ggplot() +
  geom_histogram(mapping = aes(resids), bins = 15) +
  theme_minimal() +
  xlab("Residuals") + ylab("Count") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
p2
```

Model Coefficients

```
## extract coeffs and random effects
coef(global_mod) # this include fixed and random effects
```

```
## $Plot
##        (Intercept) EggOrderEgg2    ANCHL      BEUTI       SST
## APNC     1893.779      36.17656 4.813278 -14.30913 8.977387
## BH       1898.347      36.17656 4.813278 -14.30913 8.977387
## BT       1855.839      36.17656 4.813278 -14.30913 8.977387
## CC       1901.490      36.17656 4.813278 -14.30913 8.977387
## DO       1882.390      36.17656 4.813278 -14.30913 8.977387
## ESC      1883.689      36.17656 4.813278 -14.30913 8.977387
## LC       1881.975      36.17656 4.813278 -14.30913 8.977387
## WC       1875.124      36.17656 4.813278 -14.30913 8.977387
##
## attr(,"class")
## [1] "coef.mer"
```

```
ranef_pl <- ranef(global_mod)$Plot # plot random effect only

## look at data going into random effects
table(SCMUdf$Plot)
```
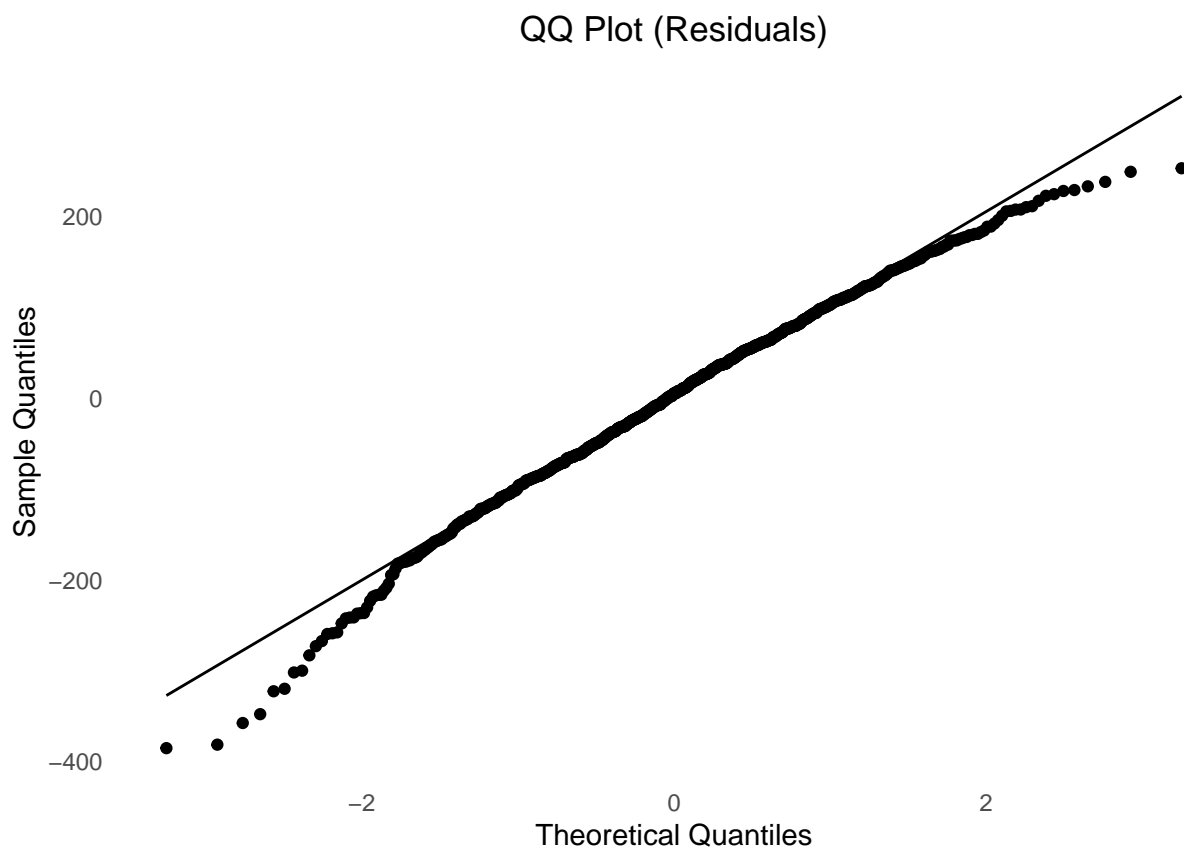
```
##
## APNC   BH   BT   CC   DO  ESC   LC   WC
```

```
## 115    31    16  275  192    1  238    4
```

Q-Q Plots

```
qqresids <- as.data.frame(resids)
p3 <- ggplot(qqresids, aes(sample = resids)) +
  geom_qq() + geom_qq_line() +
  theme_minimal() +
  xlab("Theoretical Quantiles") + ylab("Sample Quantiles") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.title = element_text(hjust = 0.5)) +
  ggtitle("QQ Plot (Residuals)")
p3
```



```
plotRE <- as.data.frame(unlist(ranef_pl))
p4 <- ggplot(plotRE, aes(sample = unlist(ranef_pl))) +
  geom_qq() + geom_qq_line() +
  theme_minimal() +
  xlab("Theoretical Quantiles") + ylab("Sample Quantiles") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.title = element_text(hjust = 0.5)) +
  ggtitle("QQ Plot (Plot Random Effect)")
p4
```

QQ Plot (Plot Random Effect)
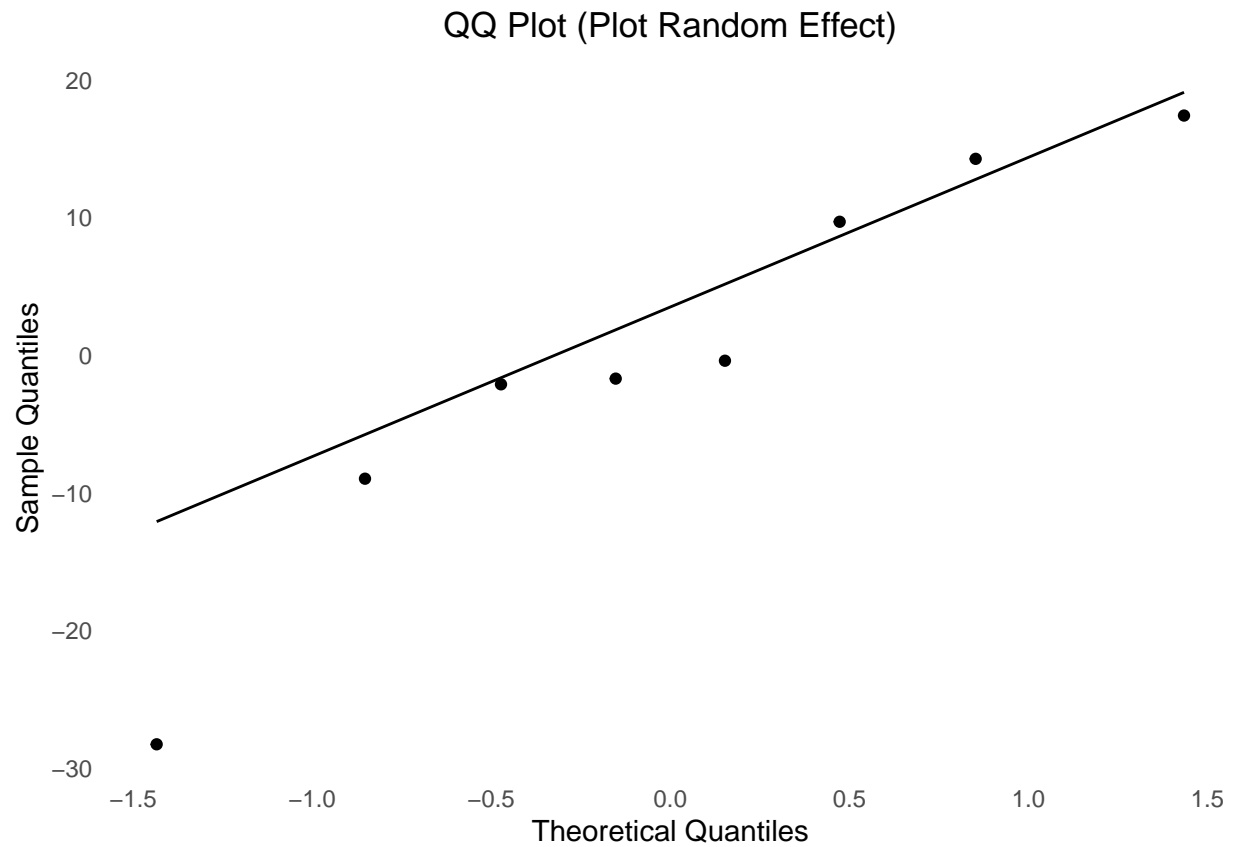
Fitted Values Versus Residuals

```
## plot residuals versus fitted values
yh <- fitted(global_mod)
plot(yh, resids, las = 1, pch = 16,
     xlab = "Fitted", ylab = "Residuals",
     main = "Residuals vs fitted")
abline(h=0, lty = "dashed")
```

Residuals vs fitted
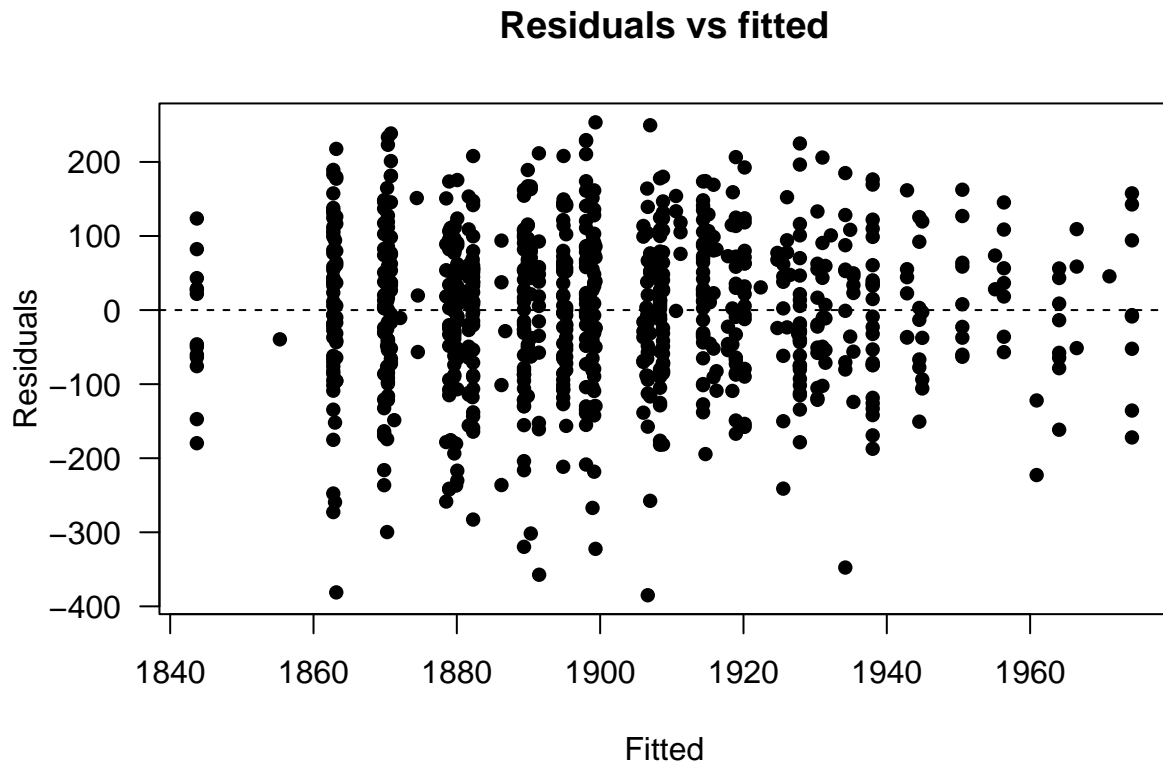
Levene's Test

We can formally test the assumption of homogenous variance via the Levene's Test, which compares the absolute values of the residuals among groups.

```
## split residuals into 2 groups
g1 <- resids[yh <= median(yh)]
g2 <- resids[yh > median(yh)]

## Levene's test
var.test(g1, g2)
```

```
##
##  F test to compare two variances
##
## data:  g1 and g2
## F = 1.1289, num df = 437, denom df = 433, p-value = 0.2066
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.9352029 1.3626040
## sample estimates:
## ratio of variances
##            1.1289
```

Fit Candidate Models

```r
## create empty matrix for storing results
mod_res <- matrix(NA, n_mods, 1)
colnames(mod_res) <- c("AIC")

## fit models and store AIC
for(i in 1:n_mods) {
  if(i == 1) {
    fmla <- "Size ~ 1 + (1 | Plot)"
  } else {
    fmla <- paste("Size ~ (1 | Plot) +", paste(cov_names[use_set[i,]], collapse = " + "))
  }
  mod_fit <- lmer(as.formula(fmla), data = SCMUdf, REML = TRUE)
  mod_res[i,"AIC"] <- AIC(mod_fit)
}

## create empty matrix for storing results
delta_res <- matrix(NA, n_mods, 1)
colnames(delta_res) <- c("deltaAIC")

## convert IC to deltaIC
delta_res[,"deltaAIC"] <- mod_res[,"AIC"] - min(mod_res[,"AIC"])
(delta_res <- round(delta_res, 2)) # round results
```

```
##        deltaAIC
##  [1,]    58.26
##  [2,]    37.82
##  [3,]    40.37
##  [4,]    19.37
##  [5,]    42.07
##  [6,]    21.23
##  [7,]    32.89
##  [8,]    11.79
##  [9,]    31.76
## [10,]    10.78
## [11,]    21.25
## [12,]     0.00
## [13,]    37.23
## [14,]    15.70
## [15,]    22.94
## [16,]     1.37
## [17,]    29.14
## [18,]     7.92
## [19,]    44.95
## [20,]    24.26
## [21,]    35.55
## [22,]    14.56
## [23,]    29.63
## [24,]     8.54
## [25,]    25.68
## [26,]     4.54
## [27,]    31.55
## [28,]    10.07
```

```
## [29,]     25.99
## [30,]      4.78
# top 5 models
top6 <- order(delta_res)[1:6]

## "best" models from our set
cov_names[use_set[top6[1],]] # EggOrder, BEUTI, NPGO

## [1] "EggOrder" "BEUTI"    "NPGO"
cov_names[use_set[top6[2],]] # EggOrder, NPGO, ONI

## [1] "EggOrder" "NPGO"     "ONI"
cov_names[use_set[top6[3],]] # EggOrder, ANCHL, BEUTI, SST (>2 AIC)

## [1] "EggOrder" "ANCHL"    "BEUTI"    "SST"
cov_names[use_set[top6[4],]] # EggOrder, PDO, SST (>2 AIC)

## [1] "EggOrder" "PDO"      "SST"
cov_names[use_set[top6[5],]] # EggOrder, PDO (>2 AIC)

## [1] "EggOrder" "PDO"
cov_names[use_set[top6[6],]] # EggOrder, BEUTI, SST (>2 AIC)

## [1] "EggOrder" "BEUTI"    "SST"
## run top models
topmod1 <- lmer(Size ~ EggOrder + BEUTI + NPGO + (1 | Plot),
                data = SCMUdf, REML = TRUE)
topmod2 <- lmer(Size ~ EggOrder + NPGO + ONI + (1 | Plot),
                data = SCMUdf, REML = TRUE)
topmod3 <- lmer(Size ~ EggOrder + ANCHL + BEUTI + SST + (1 | Plot),
                data = SCMUdf, REML = TRUE)
topmod4 <- lmer(Size ~ EggOrder + PDO + SST + (1 | Plot),
                data = SCMUdf, REML = TRUE)
topmod5 <- lmer(Size ~ EggOrder + PDO + (1 | Plot),
                data = SCMUdf, REML = TRUE)
topmod6 <-lmer(Size ~ EggOrder + BEUTI + SST + (1 | Plot),
                data = SCMUdf, REML = TRUE)


## Model selection table
AIC.tab <- matrix(NA, nrow = 6, ncol = 3) # 6 rows for 6 top models
AIC.tab[1,1] <- AIC(topmod1) # AIC for topmod1 in first row, first column
AIC.tab[2,1] <- AIC(topmod2) # AIC for topmod2 in second row, first column
AIC.tab[3,1] <- AIC(topmod3) # AIC for topmod3 in second row, first column
AIC.tab[4,1] <- AIC(topmod4) # AIC for topmod4 in second row, first column
AIC.tab[5,1] <- AIC(topmod5) # AIC for topmod5 in second row, first column
AIC.tab[6,1] <- AIC(topmod6) # AIC for topmod6 in second row, first column
AIC.tab[,2] <- AIC.tab[,1] - min(AIC.tab[,1]) # calculate delta AIC
AIC.tab[,3] <- exp(-0.5*AIC.tab[,2])/
  (sum(exp(-0.5*AIC.tab[,2]))) # calculate model weights
colnames(AIC.tab) <- c("AIC", "deltaAIC", "model_weights")
print(AIC.tab)
```

```
##             AIC deltaAIC model_weights
## [1,] 10558.27 0.000000   0.577498180
## [2,] 10559.64 1.372339   0.290770975
## [3,] 10562.81 4.539334   0.059682453
## [4,] 10563.05 4.778723   0.052949771
## [5,] 10566.19 7.915187   0.011035437
## [6,] 10566.81 8.542793   0.008063185
```

Top Models

Top Model 1

```
topmod1 <- lmer(Size ~ EggOrder + BEUTI + NPGO + (1 | Plot),
                data = SCMUdf, REML = TRUE)
summary(topmod1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Size ~ EggOrder + BEUTI + NPGO + (1 | Plot)
##    Data: SCMUdf
##
## REML criterion at convergence: 10546.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8401 -0.6528  0.0465  0.6904  2.4192
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Plot     (Intercept)   161.8   12.72
##  Residual             10701.0  103.45
## Number of obs: 872, groups:  Plot, 8
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1888.337      7.019 269.047
## EggOrderEgg2   36.179      8.713   4.152
## BEUTI         -13.338      4.713  -2.830
## NPGO          -15.613      3.522  -4.433
##
## Correlation of Fixed Effects:
##            (Intr) EggOE2 BEUTI
## EggOrdrEgg2 -0.258
## BEUTI        0.086 -0.004
## NPGO        -0.063  0.000 -0.230
```

```
confint(topmod1)
```

```
## Computing profile confidence intervals ...
```

```
##                   2.5 %      97.5 %
## .sig01          0.00000   35.186121
## .sigma         98.63360  108.388816
## (Intercept) 1869.66658 1901.989483
## EggOrderEgg2   19.19568   53.351177
## BEUTI         -22.91441   -4.318019
```

11

```
## NPGO           -22.67392   -8.856872
```

Top Model 2

```
topmod2 <- lmer(Size ~ EggOrder + NPGO + ONI + (1 | Plot),
                data = SCMUdf, REML = TRUE)
summary(topmod2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Size ~ EggOrder + NPGO + ONI + (1 | Plot)
##    Data: SCMUdf
##
## REML criterion at convergence: 10547.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8019 -0.6556  0.0455  0.6769  2.5062
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Plot      (Intercept)   178.4   13.36
##  Residual              10715.2  103.51
## Number of obs: 872, groups:  Plot, 8
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1888.353      7.238 260.899
## EggOrderEgg2   36.542      8.721   4.190
## NPGO          -13.431      3.847  -3.492
## ONI           12.399      4.855   2.554
##
## Correlation of Fixed Effects:
##            (Intr) EggOE2 NPGO
## EggOrdrEgg2 -0.252
## NPGO        -0.078  0.009
## ONI         -0.082  0.022  0.452
```

```
confint(topmod2)
```

```
## Computing profile confidence intervals ...
```

```
##                    2.5 %      97.5 %
## .sig01         0.000000   37.370181
## .sigma        98.718950  108.487913
## (Intercept) 1868.448016 1902.242601
## EggOrderEgg2  19.565650   53.761071
## NPGO         -21.088561   -6.037404
## ONI            3.152548   22.321170
```

Top Model 3

```
topmod3 <- lmer(Size ~ EggOrder + ANCHL + BEUTI + SST + (1 | Plot),
                data = SCMUdf, REML = TRUE)
summary(topmod3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Size ~ EggOrder + ANCHL + BEUTI + SST + (1 | Plot)
##    Data: SCMUdf
##
## REML criterion at convergence: 10548.8
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -3.7095 -0.6318  0.0537  0.6870  2.4419
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Plot     (Intercept)   393.6   19.84
##  Residual             10772.4  103.79
## Number of obs: 872, groups:  Plot, 8
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 1884.079      9.530 197.700
## EggOrderEgg2  36.177      8.744   4.137
## ANCHL          4.813      5.457   0.882
## BEUTI        -14.309      5.455  -2.623
## SST            8.977      4.213   2.131
##
## Correlation of Fixed Effects:
##            (Intr) EggOE2 ANCHL  BEUTI
## EggOrdrEgg2 -0.192
## ANCHL        0.094  0.007
## BEUTI        0.090 -0.001  0.532
## SST         -0.062 -0.002 -0.579 -0.273
```

```
confint(topmod3)
```

```
## Computing profile confidence intervals ...
```

```
##                     2.5 %      97.5 %
## .sig01          0.0000000   47.030247
## .sigma         99.0003359  108.841747
## (Intercept)  1859.7243438 1901.850097
## EggOrderEgg2   19.1992218   53.500458
## ANCHL          -5.6373662   15.756027
## BEUTI         -25.5857618   -4.011488
## SST             0.7582154   17.270787
```

Top Model 4

```
topmod4 <- lmer(Size ~ EggOrder + PDO + SST + (1 | Plot),
                data = SCMUdf, REML = TRUE)
summary(topmod4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Size ~ EggOrder + PDO + SST + (1 | Plot)
##    Data: SCMUdf
##
## REML criterion at convergence: 10551.1
```

```
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7474 -0.6550  0.0462  0.6908  2.4459
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Plot     (Intercept)   186.8   13.67
##  Residual             10753.9  103.70
## Number of obs: 872, groups:  Plot, 8
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 1887.528      7.318 257.919
## EggOrderEgg2  36.222      8.735   4.147
## PDO           19.275      4.528   4.256
## SST            2.893      4.025   0.719
##
## Correlation of Fixed Effects:
##            (Intr) EggOE2 PDO
## EggOrdrEgg2 -0.248
## PDO          0.003  0.003
## SST         -0.020  0.001 -0.525
```

confint(topmod4)

```
## Computing profile confidence intervals ...
```

```
##                   2.5 %      97.5 %
## .sig01         0.000000   38.14533
## .sigma        98.899099  108.68695
## (Intercept) 1867.226202 1901.54131
## EggOrderEgg2  19.206046   53.45492
## PDO           10.736526   28.75875
## SST           -5.125251   10.65872
```