

Scripps's Murrelet - Lagged Covariates

Amelia J. DuVall & Marcela Todd

5/16/2021

This is v.2021-05-18

Introduction

This document tests the inclusion of lagged environmental covariates in the Scripps's Murrelet egg size model. We initially selected the time period January-June for covariates available at a monthly sampling rate due to the overlap with the murrelet breeding season. However, we felt it was pertinent to tests lags on all environmental covariate data since the effects of some environmental conditions may take time to manifest in terms of prey availability for murrelets during the breeding season. For each environmental covariate, we used single covariate models to test the support of a non-lagged covariate (2009-2017 data) versus a one-year lagged covariate (2008-2016 data) via AIC model selection. Most covariates were available at a monthly sampling rate, with the exception of larval anchovy. For covariates at a monthly sampling rate, we also tested a "full" covariate model with monthly sampling data averaged across 12 months versus a "half" covariate model with monthly sampling data averaged across 6 months (January-June).

We tested 6 covariates: 1. Larval anchovy - SCC (ANCHL)
2. Biologically Effective Upwelling Transport Index (BEUTI)
3. Sea Surface Temperature (SST)
4. North Pacific Gyre Oscillation Index (NPGO)
5. Pacific Decadal Oscillation Index (PDO)
6. Oceanic Nino Index (ONI)

See SCMU_egg_covariates.Rmd for more information on covariate sourcing.

```
## load egg size data
egg <- read.csv(here("data", "SCMU_egg_data.csv")) %>%
  filter(TrueOrder == TRUE) %>% # egg order known only
  select(Year, Observer, Plot, Size, EggOrder)
```

Environmental Covariates

ANCHL

```
ANCHLraw <- read.csv(here("data", "covariates", "cciea_EI_FBS_2020_a29e_1fd0_409d.csv"), na.strings = "NA")

ANCHL <- ANCHLraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
```

```

mutate(Year = year(time)) %>%
mutate(ANCHL = as.numeric(relative_abundance)) %>%
arrange(Year) %>%
filter(Year >= 2008 & Year <= 2017) %>%
dplyr::select(Year, ANCHL)

ANCHL_nolag <- ANCHL %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(ANCHL_nolag = scale(ANCHL)) %>%
  dplyr::select(Year, ANCHL_nolag)

ANCHL_lag <- ANCHL %>%
  filter(Year >= 2008 & Year <= 2016) %>%
  mutate(TrueYear = Year,
         Year = TrueYear + 1,
         ANCHL_lag = scale(ANCHL)) %>%
  dplyr::select(Year, ANCHL_lag)

ANCHL1 <- full_join(ANCHL_nolag, ANCHL_lag, by = "Year")

## join with egg size data
ANCHLdf1 <- left_join(egg, ANCHL1, by = "Year")

## run models
ANCHL_nolag_mod <- lmer(Size ~ ANCHL_nolag + (1 | Plot), data = ANCHLdf1)
ANCHL_lag_mod <- lmer(Size ~ ANCHL_lag + (1 | Plot), data = ANCHLdf1)

## compare AIC
ANCHL_nolag_AIC <- AIC(ANCHL_nolag_mod)
ANCHL_lag_AIC <- AIC(ANCHL_lag_mod)

## True or False, is the ANCHL no lag model AIC smaller than the ANCHL lag AIC?
ANCHL_nolag_AIC < ANCHL_lag_AIC

## [1] TRUE

## What is the delta AIC?
ANCHL_lag_AIC - ANCHL_nolag_AIC

## [1] 4.261118

```

There is support to use the non-lagged ANCHL data.

BEUTI

First, we tested against using data from the entire year versus from January-June.

```

BEUTIrrow <- read_csv(here("data", "covariates", "cciea_OC_BEUTI_784c_ef9f_af6f.csv"))

## Parsed with column specification:
## cols(

```

```

##   time = col_character(),
##   beuti = col_character(),
##   latitude = col_character()
## )

## Averaged over all 12 months
BEUTI_full <- BEUTIrrow %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time)) %>%
  mutate(beuti = as.numeric(beuti)) %>%
  group_by(Year) %>%
  summarise(BEUTI = mean(beuti, na.rm = TRUE)) %>%
  dplyr::select(Year, BEUTI) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(BEUTI_full = scale(BEUTI)) %>%
  arrange(Year)

## Averaged for January-June only
BEUTI_half <- BEUTIrrow %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(beuti = as.numeric(beuti)) %>%
  filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5 | Month == 6) %>%
  # filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5) %>%
  group_by(Year) %>%
  summarise(BEUTI = mean(beuti, na.rm = TRUE)) %>%
  dplyr::select(Year, BEUTI) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(BEUTI_half = scale(BEUTI)) %>%
  arrange(Year)

## join 2 datasets together
BEUTI1 <- full_join(BEUTI_full, BEUTI_half, by = "Year")

## join with egg size data
BEUTIdf1 <- left_join(egg, BEUTI1, by = "Year")

## run models
BEUTI_full_mod <- lmer(Size ~ BEUTI_full + (1 | Plot), data = BEUTIdf1)
BEUTI_half_mod <- lmer(Size ~ BEUTI_half + (1 | Plot), data = BEUTIdf1)

## compare AIC
BEUTI_full_AIC <- AIC(BEUTI_full_mod)
BEUTI_half_AIC <- AIC(BEUTI_half_mod)

## True or False, is the Jan-Jun BEUTI model AIC smaller than the BEUTI full AIC?
BEUTI_half_AIC < BEUTI_full_AIC

## [1] FALSE

```

```
## What is the delta AIC?
BEUTI_half_AIC - BEUTI_full_AIC
```

```
## [1] 0.2591508
```

Both models are supported. We selected the Jan-June model for ecological reasons. Then, using the top model, we tested a non-lagged versus one-year lagged dataset.

```
## no lag, averaged for Jan-June (same as BEUTI_half above)
BEUTI_nolag <- BEUTIrrow %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(beuti = as.numeric(beuti)) %>%
  filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5 | Month == 6) %>%
  group_by(Year) %>%
  summarise(BEUTI = mean(beuti, na.rm = TRUE)) %>%
  dplyr::select(Year, BEUTI) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(BEUTI_nolag = scale(BEUTI)) %>%
  arrange(Year) %>%
  dplyr::select(Year, BEUTI_nolag)

## one year lag, averaged for Jan-June
BEUTI_lag <- BEUTIrrow %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(beuti = as.numeric(beuti)) %>%
  filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5 | Month == 6) %>%
  group_by(Year) %>%
  summarise(BEUTI = mean(beuti, na.rm = TRUE)) %>%
  dplyr::select(Year, BEUTI) %>%
  filter(Year >= 2008 & Year <= 2016) %>%
  mutate(BEUTI_lag = scale(BEUTI),
         TrueYear = Year,
         Year = TrueYear + 1) %>%
  arrange(Year) %>%
  dplyr::select(Year, BEUTI_lag)

## join 2 datasets together
BEUTI2 <- full_join(BEUTI_nolag, BEUTI_lag, by = "Year")

## join with egg size data
BEUTIdf2 <- left_join(egg, BEUTI2, by = "Year")

## run models
BEUTI_nolag_mod <- lmer(Size ~ BEUTI_nolag + (1 | Plot), data = BEUTIdf2)
BEUTI_lag_mod <- lmer(Size ~ BEUTI_lag + (1 | Plot), data = BEUTIdf2)
```

```
## compare AIC
BEUTI_nolag_AIC <- AIC(BEUTI_nolag_mod)
BEUTI_lag_AIC <- AIC(BEUTI_lag_mod)

## True or False, is the no lag BEUTI model AIC smaller than the BEUTI lag AIC?
BEUTI_nolag_AIC < BEUTI_lag_AIC
```

```
## [1] FALSE
```

```
## What is the delta AIC?
BEUTI_nolag_AIC - BEUTI_lag_AIC
```

```
## [1] 5.868553
```

There is more support for a one-year lagged BEUTI from January-June.

Just to check, let's compare a full one-year lagged BEUTI versus half one-year lagged BEUTI.

```
## one year lag, all 12 months

BEUTI_full_lag <- BEUTIrrow %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time)) %>%
  mutate(beuti = as.numeric(beuti)) %>%
  group_by(Year) %>%
  summarise(BEUTI = mean(beuti, na.rm = TRUE)) %>%
  dplyr::select(Year, BEUTI) %>%
  filter(Year >= 2008 & Year <= 2016) %>%
  mutate(BEUTI_full_lag = scale(BEUTI),
         TrueYear = Year,
         Year = TrueYear + 1) %>%
  arrange(Year) %>%
  dplyr::select(Year, BEUTI_full_lag)

## one year lag, averaged for Jan-June (same above BEUTI_lag above)
BEUTI_half_lag <- BEUTIrrow %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(beuti = as.numeric(beuti)) %>%
  filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5 | Month == 6) %>%
  group_by(Year) %>%
  summarise(BEUTI = mean(beuti, na.rm = TRUE)) %>%
  dplyr::select(Year, BEUTI) %>%
  filter(Year >= 2008 & Year <= 2016) %>%
  mutate(BEUTI_half_lag = scale(BEUTI),
         TrueYear = Year,
         Year = TrueYear + 1) %>%
  arrange(Year) %>%
  dplyr::select(Year, BEUTI_half_lag)
```

```

## join 2 datasets together
BEUTI3 <- full_join(BEUTI_full_lag, BEUTI_half_lag, by = "Year")

## join with egg size data
BEUTIdf3 <- left_join(egg, BEUTI3, by = "Year")

## run model
BEUTI_full_lag_mod <- lmer(Size ~ BEUTI_full_lag + (1 | Plot), data = BEUTIdf3)
BEUTI_half_lag_mod <- lmer(Size ~ BEUTI_half_lag + (1 | Plot), data = BEUTIdf3)

## compare AIC
BEUTI_full_lag_AIC <- AIC(BEUTI_full_lag_mod)
BEUTI_half_lag_AIC <- AIC(BEUTI_half_lag_mod)

## True or False, is the half lag BEUTI model AIC smaller than the BEUTI full lag AIC?
BEUTI_half_lag_AIC < BEUTI_full_lag_AIC

```

```
## [1] FALSE
```

```

## What is the delta AIC?
BEUTI_full_lag_AIC - BEUTI_half_lag_AIC

```

```
## [1] -0.04785196
```

Both models are supported (which is what we would expect based on the first two tests). There is support to use 12-month averaged lagged data OR 6-month averaged lagged data for BEUTI.

SST

First, we tested against using data from the entire year versus from January-June.

```
SSTraw <- read_csv(here("data", "covariates", "cciea_OC_SST3_91cf_d165_213f-46025.csv"))
```

```

## Parsed with column specification:
## cols(
##   time = col_character(),
##   SST = col_character()
## )

```

```

## Averaged over all 12 months
SST_full <- SSTraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(sst = as.numeric(SST)) %>%
  group_by(Year) %>%
  summarise(SST = mean(sst, na.rm = TRUE)) %>%
  dplyr::select(Year, SST) %>%
  filter(Year >= 2009 & Year <= 2017) %>%

```

```

mutate(SST_full = scale(SST)) %>%
  arrange(Year)

## Averaged for January-June only
SST_half <- SSTraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(sst = as.numeric(SST)) %>%
  filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5 | Month == 6) %>%
  group_by(Year) %>%
  summarise(SST = mean(sst, na.rm = TRUE)) %>%
  dplyr::select(Year, SST) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(SST_half = scale(SST)) %>%
  arrange(Year)

## join 2 datasets together
SST1 <- full_join(SST_full, SST_half, by = "Year")

## join with egg size data
SSTdf1 <- left_join(egg, SST1, by = "Year")

## run models
SST_full_mod <- lmer(Size ~ SST_full + (1 | Plot), data = SSTdf1)
SST_half_mod <- lmer(Size ~ SST_half + (1 | Plot), data = SSTdf1)

## compare AIC
SST_full_AIC <- AIC(SST_full_mod)
SST_half_AIC <- AIC(SST_half_mod)

## True or False, is the Jan-Jun SST model AIC smaller than the SST full AIC?
SST_half_AIC < SST_full_AIC

## [1] FALSE

## What is the delta AIC?
SST_full_AIC - SST_half_AIC

## [1] -4.59945

```

The full SST model has slightly more support.

Then, using the top model, we tested a non-lagged versus one-year lagged dataset.

```

## no lag, averaged for 12 months (same as SST_full above)
SST_nolag <- SSTraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%

```

```

mutate(sst = as.numeric(SST)) %>%
group_by(Year) %>%
summarise(SST = mean(sst, na.rm = TRUE)) %>%
dplyr::select(Year, SST) %>%
filter(Year >= 2009 & Year <= 2017) %>%
mutate(SST_nolag = scale(SST)) %>%
arrange(Year)

## one year lag, averaged for 12 months
SST_lag <- SSTraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(sst = as.numeric(SST)) %>%
  group_by(Year) %>%
  summarise(SST = mean(sst, na.rm = TRUE)) %>%
  dplyr::select(Year, SST) %>%
  filter(Year >= 2008 & Year <= 2016) %>%
  mutate(SST_lag = scale(SST),
         TrueYear = Year,
         Year = TrueYear + 1) %>%
  arrange(Year) %>%
  dplyr::select(Year, SST_lag)

## join 2 datasets together
SST2 <- full_join(SST_nolag, SST_lag, by = "Year")

## join with egg size data
SSTdf2 <- left_join(egg, SST2, by = "Year")

## run models
SST_nolag_mod <- lmer(Size ~ SST_nolag + (1 | Plot), data = SSTdf2)
SST_lag_mod <- lmer(Size ~ SST_lag + (1 | Plot), data = SSTdf2)

## compare AIC
SST_nolag_AIC <- AIC(SST_nolag_mod)
SST_lag_AIC <- AIC(SST_lag_mod)

## True or False, is the no lag SST model AIC smaller than the SST lag AIC?
SST_nolag_AIC < SST_lag_AIC

## [1] TRUE

## What is the delta AIC?
SST_nolag_AIC - SST_lag_AIC

## [1] -9.502538

```

The non-lagged SST data has more support. There is support for a non-lagged SST data from the entire year (12 months).

NPGO

First, we tested three single covariate models: 1) data averaged from the entire year, 2) data from January-June, and 3) data from December-March (as supported by Chenillat et al. 2012, “the statistical relationship between NPGO index and nearshore wind variability (most upwelling favorable) along the U.S. West coast is strongest in the wintertime (December to March) off Central California”).

```
NPGOraw <- read_csv(here("data", "covariates", "cciea_OC_NPGO_712b_5843_9069.csv"))
```

```
## Parsed with column specification:
## cols(
##   time = col_character(),
##   NPGO = col_double()
## )
```

```
## Averaged over all 12 months
```

```
NPGO_full <- NPGOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time)) %>%
  mutate(npgo = as.numeric(NPGO)) %>%
  group_by(Year) %>%
  summarise(NPGO = mean(npgo, na.rm = TRUE)) %>%
  dplyr::select(Year, NPGO) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(NPGO_full = scale(NPGO)) %>%
  arrange(Year) %>%
  as_tibble()
```

```
## Averaged for January-June only
```

```
NPGO_half <- NPGOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(npgo = as.numeric(NPGO)) %>%
  filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5 | Month == 6) %>%
  group_by(Year) %>%
  summarise(NPGO = mean(npgo, na.rm = TRUE)) %>%
  dplyr::select(Year, NPGO) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(NPGO_half = scale(NPGO)) %>%
  arrange(Year)
```

```
## Averaged for December-March only
```

```
NPGO_delay <- NPGOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(npgo = as.numeric(NPGO)) %>%
  filter(Month == 12 | Month == 1 | Month == 2 | Month == 3) %>%
  filter(Year >= 2008 & Year <= 2017) %>%
```

```

slice(-c(1:3)) %>%
slice(-37) %>%
mutate(Period = c(rep("A", 4), rep("B", 4), rep("C", 4), rep("D", 4),
                    rep("E", 4), rep("F", 4), rep("G", 4), rep("H", 4), rep("I", 4))) %>%
group_by(Period) %>%
summarise(NPGO = mean(npgo, na.rm = TRUE)) %>%
mutate(Year = 2009:2017) %>%
mutate(NPGO_delay = scale(NPGO)) %>%
dplyr::select(Year, NPGO_delay) %>%
arrange(Year)

## join 2 datasets together
NPGO1 <- full_join(NPGO_full, NPGO_half, by = "Year")
NPGO2 <- full_join(NPGO1, NPGO_delay, by = "Year")

## join with egg size data
NPGOdf1 <- left_join(egg, NPGO2, by = "Year")

## run models
NPGO_full_mod <- lmer(Size ~ NPGO_full + (1 | Plot), data = NPGOdf1)
NPGO_half_mod <- lmer(Size ~ NPGO_half + (1 | Plot), data = NPGOdf1)
NPGO_delay_mod <- lmer(Size ~ NPGO_delay + (1 | Plot), data = NPGOdf1)

## compare AIC
NPGO_full_AIC <- AIC(NPGO_full_mod)
NPGO_half_AIC <- AIC(NPGO_half_mod)
NPGO_delay_AIC <- AIC(NPGO_delay_mod)

## True or False, is the full NPGO model AIC smaller than the half NPGO model AIC?
NPGO_full_AIC < NPGO_half_AIC

## [1] TRUE

## What is the delta AIC?
NPGO_full_AIC - NPGO_half_AIC

## [1] -1.661432

## True or False, is the delay NPGO model AIC smaller than the full NPGO model AIC?
NPGO_delay_AIC < NPGO_full_AIC

## [1] FALSE

## What is the delta AIC?
NPGO_delay_AIC - NPGO_full_AIC

## [1] 3.879544

```

The model that used all data from 12 months averaged for NPGO had the best fit.
Then, using the full NPGO model, we tested a non-lagged versus one-year lagged dataset.

```

## no lag, averaged for 12 months (same as NPGO_full above)
NPGO_nolag <- NPGOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time)) %>%
  mutate(npgo = as.numeric(NPGO)) %>%
  group_by(Year) %>%
  summarise(NPGO = mean(npgo, na.rm = TRUE)) %>%
  dplyr::select(Year, NPGO) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(NPGO_nolag = scale(NPGO)) %>%
  arrange(Year) %>%
  as_tibble()

## one year lag, averaged for 12 months
NPGO_lag <- NPGOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(npgo = as.numeric(NPGO)) %>%
  group_by(Year) %>%
  summarise(NPGO = mean(npgo, na.rm = TRUE)) %>%
  dplyr::select(Year, NPGO) %>%
  filter(Year >= 2008 & Year <= 2016) %>%
  mutate(NPGO_lag = scale(NPGO),
         TrueYear = Year,
         Year = TrueYear + 1) %>%
  arrange(Year) %>%
  dplyr::select(Year, NPGO_lag)

## join 2 datasets together
NPGO2 <- full_join(NPGO_nolag, NPGO_lag, by = "Year")

## join with egg size data
NPGOdf2 <- left_join(egg, NPGO2, by = "Year")

## run models
NPGO_nolag_mod <- lmer(Size ~ NPGO_nolag + (1 | Plot), data = NPGOdf2)
NPGO_lag_mod <- lmer(Size ~ NPGO_lag + (1 | Plot), data = NPGOdf2)

## compare AIC
NPGO_nolag_AIC <- AIC(NPGO_nolag_mod)
NPGO_lag_AIC <- AIC(NPGO_lag_mod)

## True or False, is the no lag NPGO model AIC smaller than the NPGO lag AIC?
NPGO_nolag_AIC < NPGO_lag_AIC

## [1] TRUE

## What is the delta AIC?
NPGO_nolag_AIC - NPGO_lag_AIC

```

```
## [1] -17.49543
```

The non-lagged NPGO data has more support. There is support for a non-lagged NPGO data from the entire year (12 months).

PDO

```
PDOraw <- read_csv(here("data", "covariates", "cciea_OC_PDO_712b_5843_9069.csv"))
```

```
## Parsed with column specification:
## cols(
##   time = col_character(),
##   PDO = col_double()
## )
```

```
PDO_full <- PDOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time)) %>%
  mutate(pdo = as.numeric(PDO)) %>%
  group_by(Year) %>%
  summarise(PDO = mean(pdo, na.rm = TRUE)) %>%
  dplyr::select(Year, PDO) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(PDO_full = scale(PDO)) %>%
  arrange(Year) %>%
  as_tibble()

## for January-June only
PDO_half <- PDOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(pdo = as.numeric(PDO)) %>%
  filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5 | Month == 6) %>%
  group_by(Year) %>%
  summarise(PDO = mean(pdo, na.rm = TRUE)) %>%
  dplyr::select(Year, PDO) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(PDO_half = scale(PDO)) %>%
  arrange(Year) %>%
  as_tibble()

## join 2 datasets together
PDO1 <- full_join(PDO_full, PDO_half, by = "Year")

## join with egg size data
PDOdf1 <- left_join(egg, PDO1, by = "Year")

## run models
```

```

PDO_full_mod <- lmer(Size ~ PDO_full + (1 | Plot), data = PDOdf1)
PDO_half_mod <- lmer(Size ~ PDO_half + (1 | Plot), data = PDOdf1)

## compare AIC
PDO_full_AIC <- AIC(PDO_full_mod)
PDO_half_AIC <- AIC(PDO_half_mod)

## True or False, is the Jan-Jun PDO model AIC smaller than the PDO full AIC?
PDO_half_AIC < PDO_full_AIC

```

```
## [1] FALSE
```

```

## What is the delta AIC?
PDO_full_AIC - PDO_half_AIC

```

```
## [1] -2.730244
```

The full PDO model has more support than the half model.

Then, using the full PDO model, we tested a non-lagged versus one-year lagged dataset.

```
## no lag, averaged for 12 months (same as PDO_full above)
```

```

PDO_nolag <- PDOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(pdo = as.numeric(PDO)) %>%
  group_by(Year) %>%
  summarise(PDO = mean(pdo, na.rm = TRUE)) %>%
  dplyr::select(Year, PDO) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(PDO_nolag = scale(PDO)) %>%
  arrange(Year)

```

```
## one year lag, averaged for 12 months
```

```

PDO_lag <- PDOraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(pdo = as.numeric(PDO)) %>%
  group_by(Year) %>%
  summarise(PDO = mean(pdo, na.rm = TRUE)) %>%
  dplyr::select(Year, PDO) %>%
  filter(Year >= 2008 & Year <= 2016) %>%
  mutate(PDO_lag = scale(PDO),
         TrueYear = Year,
         Year = TrueYear + 1) %>%
  arrange(Year) %>%
  dplyr::select(Year, PDO_lag)

```

```

## join 2 datasets together
PDO2 <- full_join(PDO_nolag, PDO_lag, by = "Year")

## join with egg size data
PDOdf2 <- left_join(egg, PDO2, by = "Year")

## run models
PDO_nolag_mod <- lmer(Size ~ PDO_nolag + (1 | Plot), data = PDOdf2)
PDO_lag_mod <- lmer(Size ~ PDO_lag + (1 | Plot), data = PDOdf2)

## compare AIC
PDO_nolag_AIC <- AIC(PDO_nolag_mod)
PDO_lag_AIC <- AIC(PDO_lag_mod)

## True or False, is the no lag PDO model AIC smaller than the PDO lag AIC?
PDO_nolag_AIC < PDO_lag_AIC

```

```
## [1] FALSE
```

```

## What is the delta AIC?
PDO_nolag_AIC - PDO_lag_AIC

```

```
## [1] 4.740066
```

There is support for including a one year lag for the PDO model averaged for 12 months.

ONI

```
ONIrrow <- read_csv(here("data", "covariates", "cciea_OC_ONI_712b_5843_9069.csv"))
```

```

## Parsed with column specification:
## cols(
##   time = col_character(),
##   ONI = col_double()
## )

```

```

ONI_full <- ONIrrow %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time)) %>%
  mutate(oni = as.numeric(ONI)) %>%
  group_by(Year) %>%
  summarise(ONI = mean(oni, na.rm = TRUE)) %>%
  dplyr::select(Year, ONI) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  arrange(Year) %>%
  mutate(ONI_full = scale(ONI)) %>%
  as_tibble()

```

```

## for January-June only
ONI_half <- ONIraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(oni = as.numeric(ONI)) %>%
  filter(Month == 1 | Month == 2 | Month == 3 | Month == 4 | Month == 5 | Month == 6) %>%
  group_by(Year) %>%
  summarise(ONI = mean(oni, na.rm = TRUE)) %>%
  dplyr::select(Year, ONI) %>%
  filter(Year >= 2009 & Year <= 2017) %>%
  mutate(ONI_half = scale(ONI)) %>%
  arrange(Year) %>%
  as_tibble()

## join 2 datasets together
ONI1 <- full_join(ONI_full, ONI_half, by = "Year")

## join with egg size data
ONIdf1 <- left_join(egg, ONI1, by = "Year")

## run models
ONI_full_mod <- lmer(Size ~ ONI_full + (1 | Plot), data = ONIdf1)
ONI_half_mod <- lmer(Size ~ ONI_half + (1 | Plot), data = ONIdf1)

## compare AIC
ONI_full_AIC <- AIC(ONI_full_mod)
ONI_half_AIC <- AIC(ONI_half_mod)

## True or False, is the Jan-Jun ONI model AIC smaller than the ONI full AIC?
ONI_half_AIC < ONI_full_AIC

```

```
## [1] FALSE
```

```

## What is the delta AIC?
ONI_full_AIC - ONI_half_AIC

```

```
## [1] -4.821609
```

The full ONI model has more support.

Using the full ONI model, we tested a non-lagged versus one-year lagged dataset.

```

## no lag, averaged for 12 months (same as ONI_full above)

ONI_nolag <- ONIraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(oni = as.numeric(ONI)) %>%
  group_by(Year) %>%

```

```

summarise(ONI = mean(oni, na.rm = TRUE)) %>%
dplyr::select(Year, ONI) %>%
filter(Year >= 2009 & Year <= 2017) %>%
mutate(ONI_nolag = scale(ONI)) %>%
arrange(Year) %>%
as_tibble()

## one year lag, averaged for 12 months
ONI_lag <- ONIraw %>%
  slice(-1) %>%
  mutate(time = ymd_hms(time)) %>%
  mutate(Year = year(time),
         Month = month(time)) %>%
  mutate(oni = as.numeric(ONI)) %>%
  group_by(Year) %>%
  summarise(ONI = mean(oni, na.rm = TRUE)) %>%
  dplyr::select(Year, ONI) %>%
  filter(Year >= 2008 & Year <= 2016) %>%
  mutate(ONI_lag = scale(ONI),
         TrueYear = Year,
         Year = TrueYear + 1) %>%
  arrange(Year) %>%
  dplyr::select(Year, ONI_lag)

## join 2 datasets together
ONI2 <- full_join(ONI_nolag, ONI_lag, by = "Year")

## join with egg size data
ONIdf2 <- left_join(egg, ONI2, by = "Year")

## run models
ONI_nolag_mod <- lmer(Size ~ ONI_nolag + (1 | Plot), data = ONIdf2)
ONI_lag_mod <- lmer(Size ~ ONI_lag + (1 | Plot), data = ONIdf2)

## compare AIC
ONI_nolag_AIC <- AIC(ONI_nolag_mod)
ONI_lag_AIC <- AIC(ONI_lag_mod)

## True or False, is the no lag ONI model AIC smaller than the ONI lag AIC?
ONI_nolag_AIC < ONI_lag_AIC

## [1] TRUE

## What is the delta AIC?
ONI_nolag_AIC - ONI_lag_AIC

## [1] -3.01292

```

The no lag 12 month ONI model has more support.