

# Species Density ~ SST

Cullen Molitor

2021-12-01



# EDS 222 Final Project

## Motivation

- Motivate your question. Why is this important? Is there existing evidence on this question? If so, why is it inconclusive? If not, why not?

California's Northern Channel Islands sit at the transition between two biogeographic provinces, the cold-water North Pacific and the warm-water Gulf of California to the south. San Miguel and Santa Rosa Islands have species representative of the North Pacific province while Anacapa and Santa Barbara Islands sit firmly in the Gulf of California province. Santa Cruz Island lies in the transition zone with the western end of the island favoring North Pacific species and the eastern end favoring Gulf of California species. This makes the Northern Channel Islands a unique place to study how species distributions respond to ocean temperatures.

```
knitr:::include_graphics("biogeo.png")
```

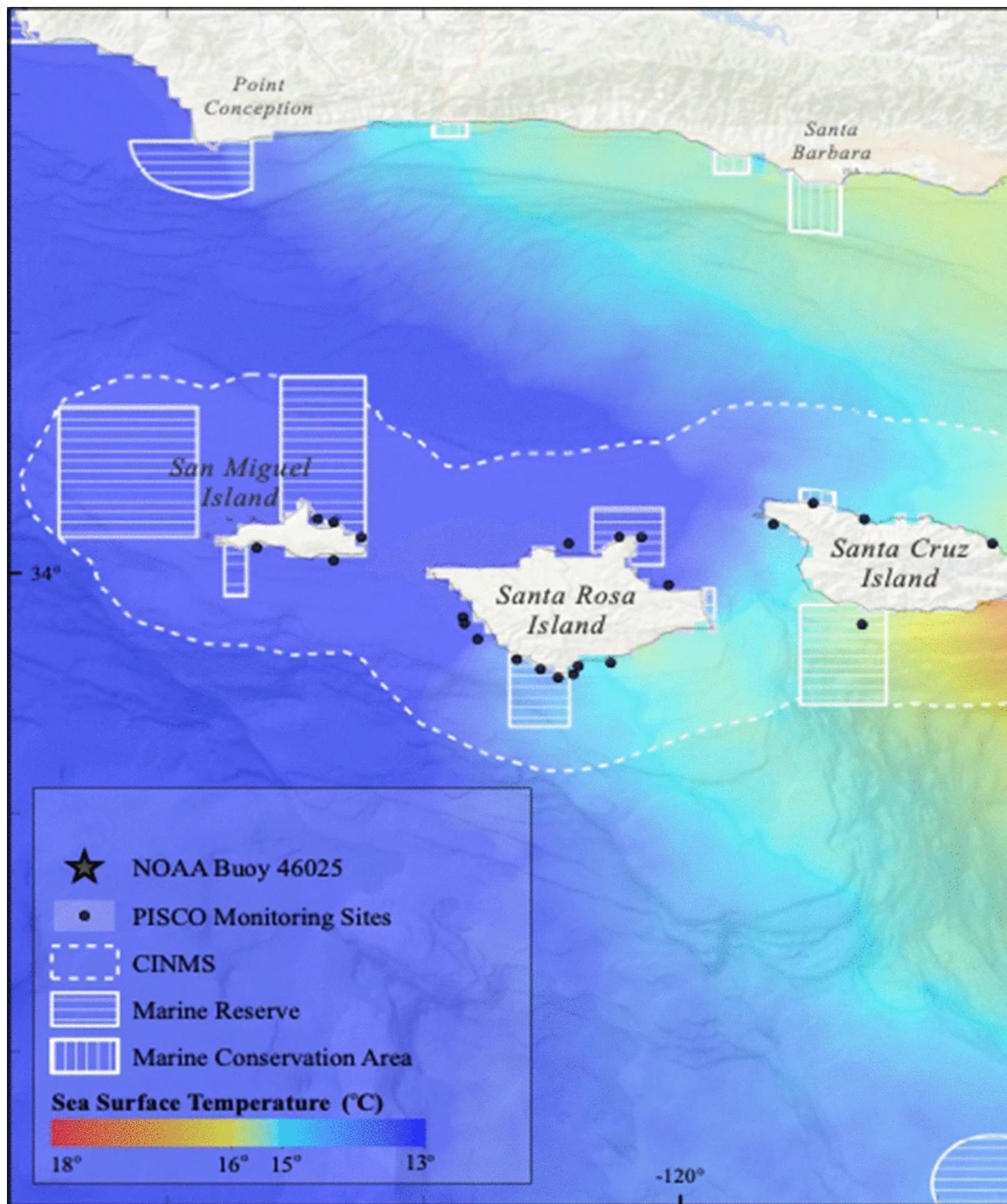


Figure 1. The Southern California Bight (SCB) shown with a composite sea surface temperatures (SST) color gradient for 2009. This SST is typical of the region and illustrates the transition between the North Pacific and Gulf of California biogeographic provinces (Image taken from Freedman 2020).

The El Niño Southern Oscillation (ENSO) Oceanic Niño Index calculates an anomaly value for each month going back to 1950. These values are derived from SST in the equatorial pacific ( $\pm 5^\circ$  latitude,  $120^\circ - 170^\circ W$  longitude). The oscillation seen in this cycle is responsible for driving certain global weather and climate patterns. Despite the distance from this region, ENSO is responsible for driving the water temperatures of the SCB and the Channel Islands.

The impetus of this analysis is to investigate which species are sensitive to either warm-water or cold-water events. Using a series of dynamic linear model with lag independent variables (SST anomalies)

## Data

- Describe your data. Where did you access it? What are its spatial and temporal features? What are its limitations? What do you know about the sampling strategy and what biases that may introduce? If helpful, you can use a histogram, scatterplot, or summary statistics table to describe your data.

The Kelp Forest Monitoring (KFM) program at Channel Islands National Park has been collecting abundance and size distribution data of more than 150 species for nearly 40 years (sampled annually) of 37 sites at the five Islands in the National Park.

Using species density, we can see the trends observed over multiple El Niño (warm-water) and La Niña (cold-water) events. Using a dynamic linear model with lag independent variables we can see the effect of sea surface temperature anomalies during and in the years following these events. recruitment of certain species either increases or decreases.

```
# Load libraries
library(tidyverse)
library(here)
library(broom)
library(tsibble)
library(dLagM)
library(arrows)
library(lubridate)

# Load and Tidy Data
oni_yearly <- read.table( # Read in ONI to be added to all data
  "https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/detrend.nino34.ascii.txt",
  header = T) %>%
  dplyr::mutate(Date = as.Date(ISOdate(YR, MON, 1)),
    DateStart = as.Date(ISOdate(YR, MON, 1)),
    DateEnd = ceiling_date(DateStart, "month")) %>%
  dplyr::rename(SST_Anom = ANOM,
    Month = MON,
    SurveyYear = YR) %>%
  dplyr::select(SurveyYear, Month, Date, DateStart, DateEnd, SST_Anom) %>%
  dplyr::group_by(SurveyYear) %>%
  dplyr::summarise(SST_Anom = mean(SST_Anom)) %>%
  dplyr::mutate(SST_Anom_1 = lag(SST_Anom, n = 1),
    SST_Anom_2 = lag(SST_Anom, n = 2),
    SST_Anom_3 = lag(SST_Anom, n = 3),
```

```

SST_Anom_4 = lag(SST_Anom, n = 4),
SST_Anom_5 = lag(SST_Anom, n = 5))

Species_Info <- readr::read_csv("Species_Complete.csv")

Site_Info <- readr::read_csv("Site_Info.csv")

density <- arrow::read_feather("Density.feather") %>%
  dplyr::filter(Classification != "Fish",
                !CommonName %in% c("wakame, adult", "wakame, juvenile", "white abalone")) %>%
  dplyr::left_join(oni_yearly)

```

## Analysis

- Clearly describe your analysis plan. What is your analysis plan? Why did you choose this analysis, given your data and question? What are the limitations?

## Model

```

Results_lag <- density %>%
  dplyr::group_by(ScientificName) %>%
  dplyr::summarise(
    generics::tidy(
      stats::lm(
        Mean_Density ~ SST_Anom + SST_Anom_1 + SST_Anom_2 + SST_Anom_3 + SST_Anom_4 + SST_Anom_5
      )) %>%
    # tidyrr::drop_na(p.value) %>%
    dplyr::filter(term != "(Intercept)") %>%
    dplyr::mutate(significant = ifelse(p.value <= .05, "yes", "no"))

Results_filtered <- Results_lag %>%
  dplyr::filter(p.value <= .05) %>%
  dplyr::arrange(estimate) %>%
  # dplyr::arrange(p.value) %>%
  dplyr::mutate(statistic = round(statistic, 3),
                p.value = round(p.value, 3),
                p.value = ifelse(p.value < 0.001, "< 0.001", as.character(p.value)))

```

## Results

- Summarize your results visually and in words. Show us your results in figure(s) and/or table(s) that are carefully labeled and captioned. Describe in the text (and orally when presenting) what you found, and how these results either do or do not help you answer your question.

```

prop <- Results_lag %>%
  group_by(ScientificName) %>%
  summarise(est_sum = sum(estimate))

```

```

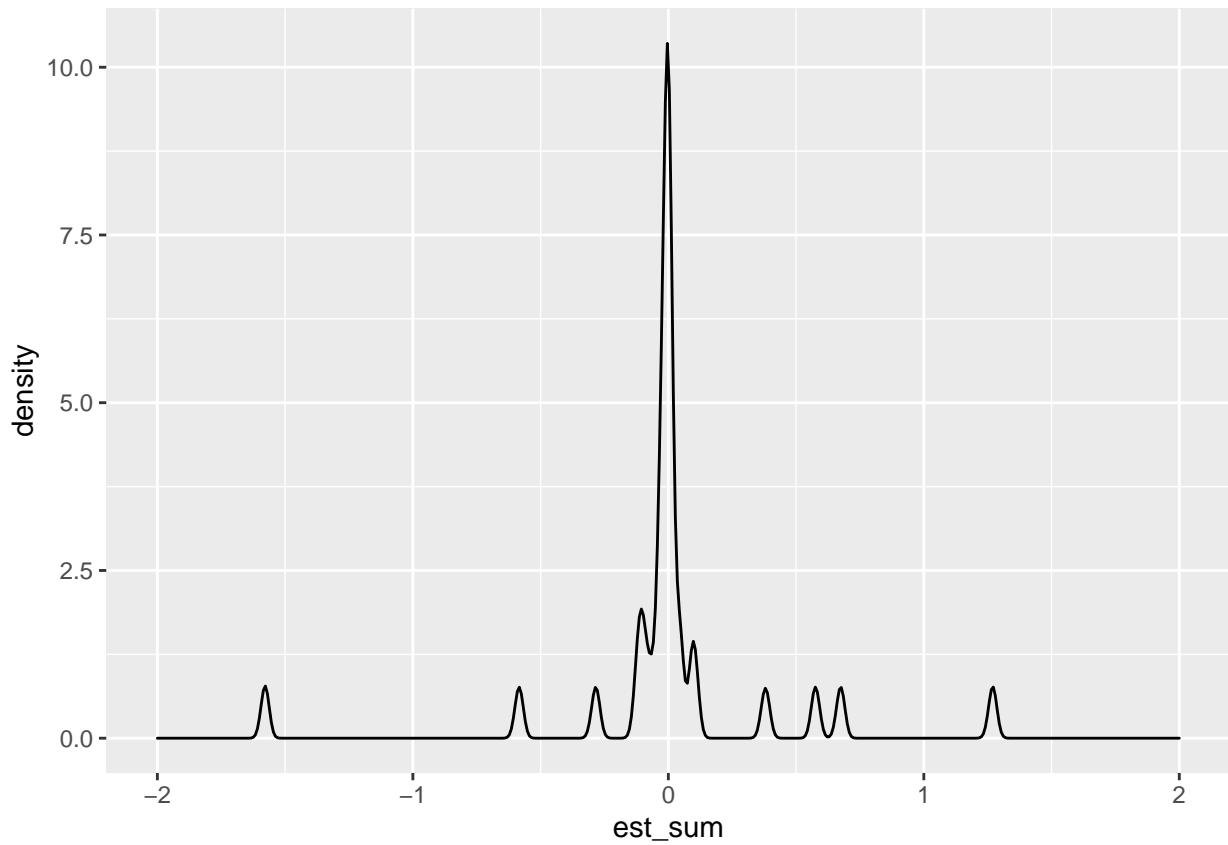
total <- length(Results_lag$ScientificName)
warm <- Results_lag %>%
  filter(estimate > 0) %>%
  # distinct(CommonName) %>%
  pull(ScientificName) %>%
  length()
cold <- Results_lag %>%
  filter(estimate < 0) %>%
  # distinct(CommonName) %>%
  pull(ScientificName) %>%
  length()

```

```

ggplot(data = prop, aes(x = est_sum)) +
  geom_density() +
  xlim(c(-2, 2))

```



```

# for (sp in unique(Results_filtered$CommonName)) {
#   p <- density %>%
#     filter(CommonName == sp) %>%
#     ggplot(aes(y = Mean_Density)) +
#     # geom_point() +
#     geom_smooth(aes(x = SST_Anom), se = F, method = lm, formula = "y~x", color = "red") +
#     geom_smooth(aes(x = SST_Anom_1), se = F, method = lm, formula = "y~x", color = "blue") +
#     geom_smooth(aes(x = SST_Anom_2), se = F, method = lm, formula = "y~x", color = "green") +

```

```

#     geom_smooth(aes(x = SST_Anom_3), se = F, method = lm, formula = "y~x", color = "black") +
#     geom_smooth(aes(x = SST_Anom_4), se = F, method = lm, formula = "y~x", color = "purple") +
#     labs(title = sp) +
#     theme_classic()
#
#   print(p)
# }

```

```

# for (sp in unique(Results_filtered$CommonName)) {
#   p <- Results_lag %>%
#     filter(CommonName == sp) %>%
#     ggplot() +
#     geom_line(aes(x = term, y = estimate, color = CommonName, group = CommonName), show.legend = F) +
#     geom_point(aes(x = term, y = estimate, shape = significant), size = 4) +
#     labs(title = sp) +
#     theme_classic()
#
#   print(p)
# }

```

## Future Work

- What might you do next? One short analysis cannot fully answer an interesting scientific question. If you had time to collect more data or conduct more analysis, what would help you answer this question better?

## Literature Cited

Costello, M.J., Tsai, P., Wong, P.S. et al. Marine biogeographic realms and species endemicity. Nat Commun 8, 1057 (2017). <https://doi.org/10.1038/s41467-017-01121-2>

Freedman, R.M., Brown, J.A., Caldow, C. et al. Marine protected areas do not prevent marine heatwave-induced fish community structure changes in a temperate transition zone. Sci Rep 10, 21081 (2020). <https://doi.org/10.1038/s41598-020-77885-3>