Interim Report S25

Introduction

Since the start of the Industrial Revolution, atmospheric carbon dioxide levels have spiked upward due to practices like deforestation and the burning of fossil fuels, causing an ever increasing amount of carbon dioxide to dissolve into the ocean. Once dissolved, carbon dioxide undergoes a series of chemical equilibrium reactions that ultimately lowers the pH of the water in a process known as ocean acidification. This process has numerous ramifications on marine ecosystems, including but not limited to the impairment of specific pH dependent biological processes and the leaching of calcium carbonate required to build the shells and skeletons of many marine species.[1]

This project is sponsored by California Cooperative Oceanic Fisheries Investigations (Cal-COFI), an organization founded in 1949 to study the ecological aspects of the Pacific sardine collapse off of the coast of California. CalCOFI is committed to studying California's coastal marine environment and collecting relevant oceanographic data in order to provide insight on important climate change related topics such as renewable energy, integrated ocean management, and marine spatial planning.[2]

We aim to extend the research done in "A 37 year record of ocean acidification in the Southern California current" by Wolfe et al.[3] on the yearly rate of change of certain carbonate chemistry variables by examining all available carbonate chemistry data collected across CalCOFI observation stations rather than only surface data collected at station 90.90, and study the impact of ocean acidification on zooplankton and krill biovolumes.

---

[1] Ocean acidification: https://www.nature.com/scitable/knowledge/library/ocean-acidification-25822734/
[2] CalCOFI report: https://calcofi.org/downloads/publications/CalCOFI3YrReports/CalCOFI_Review_2017_2021.pdf
[3] Wolfe Paper: https://www.nature.com/articles/s43247-023-01065-0

Problems of Interest

The goals of this project can be split into two core parts: the analysis of ocean carbonate chemistry variables, and the study of zooplankton and krill biovolumes off of the California coast.

For the carbonate chemistry portion, it is of interest to examine how important ocean carbon chemistry and oceanographic variables, namely total alkalinity (TA), total dissolved inorganic carbon (DIC), the Revelle Factor, pH, pCO_2, Omega aragonite ($\Omega_{aragonite}$), Omega calcite ($\Omega_{calcite}$), temperature, salinity, and CO^{2-}. Additionally, we wish to assess the performance of Empirical Seawater Property Estimation Routines (ESPER)[4] in predicting carbonate chemistry variables across different depths with easy to collect oceanographic variables such as temperature and salinity as inputs.

Regarding zooplankton and krill biovolumes, we want to conduct a cross-comparison of the hydrographic and biological datasets to assess the amount of overlap spatially and temporally. Using co-located measurements, we wish to model the effects of ocean acidification, using the carbonate chemistry and oceanographic variables aforementioned, on zooplankton and krill abundance. In addition, we are interested in exploring how pH and related environmental factors affect the abundance of calcifying versus non-calcifying species.

Materials and Methods

Oceanographic and Carbonate Chemistry Data[5]

CalCOFI samples from a predetermined sampling grid off the coast of California on a quarterly basis. Typical stations are set 40 nautical miles apart. At each sampling point, identified by a station and line number, CalCOFI lowers a carousel of 24 bottles into the water, which collect seawater samples from around 20 different depths (typically ranging from 20 to 515 meters). Researchers on the ship then measure oceanographic values such as the temperature, salinity, macronutrient concentration and other properties of these samples. This results in the oceanographic dataset used in this study.

While oceanographic data is measured on every CalCOFI cruise, carbonate chemistry values such as TA and DIC are only occasionally measured from the collected water samples and are stored in their own dataset.

We merged these two datasets through a left join on the carbonate chemistry dataset to create merged_bottle_data.csv.

---

[4]ESPER: https://github.com/BRCScienceProducts/ESPER
[5]Bottle Database: https://calcofi.org/data/oceanographic-data/bottle-database/

Biological Data

There are three datasets that we are focusing on for the biological data: (1) CalCOFI NOAA Zooplankton Volume, (2) BTEDB (Krill) Abundances, and (3) PRPOOS Data (for Zooplankton Calcifiers/Non-Calcifiers Abundance). The zooplankton and krill biovolume data are obtained using net tows (Bongo and/or Pairovet) at each standard CalCOFI station. The PRPOOS data is also obtained using a net tow but rather than sampling at station, it is conducted during transits between stations. These three datasets have each been merged with the bottle data to create zoop_data/zooplankton_pH.csv, krill_data/CV_merged_krill.csv, and PRPOOS/prpoos_summary.csv, respectively.

CalCOFI NOAA Zooplankton Volume[6]:

The zooplankton biovolume data measures the amount of "plankton" (the small and microscopic organisms floating in the sea, consisting chiefly of diatoms, protozoans, small crustaceans, and the eggs and larval stages of larger animals) in the volume of sea water sampled, which is useful for research on the ecology, ecosystems, and fisheries of different fish species. In particular, we are interested in the variables `total_plankton` and `small_plankton`.

BTEDB (Krill) Abundances[7]:

The krill abundance dataset provides information on krill abundance from the Brinton and Townsend Euphausiid Database (BTEDB) based on samples collected by the CalCOFI program since 1951. The samples were analyzed by experienced taxonomists, including Edward Brinton, Annie Townsend, and Linsey Sala. Most of the data comes from springtime research cruises, particularly in Southern California and Central California. Since 2004, taxonomic analysis has been carried out as part of the CCE-LTER (California Current Ecosystem Long-Term Ecological Research) program, supported by the U.S. National Science Foundation. Samples are collected by specialized nets towed from depths of up to 210 meters. It includes species such as *Euphausia pacifica*, *Nematoscelis difficilis*, and *Thysanoessa spinifera*, with individuals categorized by size and developmental phase (e.g., calyptopis, furcilia, juvenile, adult).

---

[6]CalCOFI Zooplankton Volume Database: https://oceanview.pfeg.noaa.gov/erddap/tabledap/erdCalCOFIzoovol.html

[7]BTEDB (Krill Volume) Data: https://portal.edirepository.org/nis/mapbrowse?packageid=knb-lter-cce.313.1

PRPOOS (Calcifiers/Non-Calcifiers)[8]:

The PRPOOS (Planktonic Rate Processes in Oligotrophic Ocean Systems) dataset contains abundance and estimated biomass values for various zooplankton taxa, which can be categorized into calcifying and non-calcifying groups. The calcifying taxa are defined as *byrozoan larvae*, *pteropoda heteropoda*, *ostracods*, and *rhizaria*; the remaining taxa are considered non-calcifying.

Carbon Chemistry Methods

ESPER Model Validation

A mixture of qualitative and quantitative analyses were applied to assess ESPER model performance. To determine the best performing model, we compared the RMSE and relative errors across models. To identify areas of weakness in ESPER's predictive ability, we generated plots of predictions and/or residuals against the model's inputs. Finally, we performed t-tests of the residuals to ascertain whether a statistically significant bias in the model predictions exists. Specifically, given that μ_ϵ is the mean residual value for a particular model's predictions, we tested the following hypotheses:

$$H_0 : \mu_\epsilon = 0 \quad \text{and} \quad H_a : \mu_\epsilon \neq 0 \tag{1}$$

To account for alpha inflation from multiple testing, we applied the Benjamini-Hochberg procedure to adjust our obtained p-values.

Ocean Acidification Trend Analysis

CO2SYS, a program used to mechanistically calculate other carbonate chemistry values given oceanographic values and at least two carbonate chemistry values as input[9], was used on the merged bottle dataset to obtain values of pH, the Revelle Factor, pCO_2, CO_3^{2-}, Omega aragonite, and Omega calcite for each observation. These values were then merged with the bottle data set to create merged_bottle_co2sys.csv. The values of pH, the Revelle Factor, pCO_2, TA, DIC, temperature, salinity, Omega aragonite, Omega calcite, and CO_3^{2-} were then seasonally detrended using the procedure recommended in "Advancing best practices for assessing trends of ocean acidification time series" by Sutton et al[10]. Since the relationship between depth and different carbonate chemistry variables can be hard to model, and since we

---

[8]PRPOOS (Zooplankton Calcifiers/Non-Calcifiers Volume) Zooscan Database: https://oceaninformatics.ucsd.edu/zooscandb/

[9]CO2SYS: https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/CO2SYS/co2rprt.html

[10]Seasonally Detrend: https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2022.1045667/full

would expect that there is an interaction between time trends and depth, we have decided to limit our study to observations with depths of 20 meters or less for the time being. Observations from stations with less than 20 observations were then filtered out in order to ensure more accurate model fits. Linear regression models were then fit for each variable of interest against time at each station. The Benjamini-Hochberg procedure was used to adjust p-values to account for multiple testing. Additionally, a mixed effects model was fit to each variable of interest regressed against time, depth, and with a random intercept for station. Maximum likelihood estimation was used to estimate the parameters for every model.

Biological Systems Methods

The main methods we have implemented so far are a combination of linear models and spatial models. Our initial model consisted of the carbonate chemistry variables of interest as predictors (TA, DIC, temperature, and salinity) with the appropriate abundance variables as our response variables. To build on these models, we included more variables from the CO2SYS output, such as pH, and performed LASSO regression to identify any other potentially significant predictors. For spatial modelling, we used a spatial mesh model as well as GAM with spatial splines to incorporate the effects of latitude and longitude on the generalized linear models.