# RESEARCH PROPOSAL: MODELING KELP POPULATION DYNAMICS AROUND THE CHANNEL ISLANDS NATIONAL PARK FROM 2010 TO 2022 USING A GENERALIZED ADDITIVE MODEL

BY DING LI

**1. Introduction.** Kelp forests in the nearshore ocean are extremely important to the local marine ecology, economics, and carbon cycle [Saccomanno, Bell and Pawlak (2023)]. Nonetheless, due to more and more severe climate change problems, such as more frequent heat waves that damage the food chain at which kelps function, the population of kelp forests has decreased dramatically in recent years, which causes serious issues [Sills et al. (2020)]. Therefore, researchers utilize a wide variety of advanced techniques to monitor kelp populations, which can be helpful in doing trend and driver analyses for kelp population dynamics using statistical tools [Saccomanno, Bell and Pawlak (2023); Schroeder et al. (2019); Cavanaugh et al. (2021)]. By conducting such analyses based on statistical models combined with some mechanistic knowledge, better policies and measure can be taken more appropriately for protecting, restoring, and predicting kelp population dynamics. In this research, we try to analyze the trend of the kelp population dynamics and fit a regression model using time, temperature, and sunlight data as the covariates, which are considered the main factors of kelp population dynamics mechanistically [Simonson, Scheibling and Metaxas (2015); Cavanaugh et al. (2011)]. Thus, we may check if the fitted statistical models based on real-world data agree with the theoretical explanation of kelp population dynamics.

**2. Data.** We use three time-series gridded datasets, including kelp canopy extent, land surface temperature (LST), and sunlight distribution. The kelp canopy dataset is obtained from KelpWatch, which is an online data library maintained by researchers from UCLA, USSB, WHOI, etc. The data are cell-gridded kelp canopy extent, retrieved by some machine learning algorithms applied on the satellite remote sensing data with high enough accuracy, where each cell contains a number from 0 to 1 representing how much proportion the cell is occupied by kelp canopy [Bell et al. (2023)]. For example, a 0 indicates that the grid does not contain any kelp canopy, whereas a 1 means that the grid is completely occupied by kelp canopies. In this project, we download the annual kelp canopy dataset around the Channel Islands National Park (CINP) in the United States from 2010 to 2022. For each year, the overall kelp canopy extent can be calculated by the sum of the products of the resolution of each cell (area per cell) times the extent value recorded in the cell. The time-series gridded temperature is downloaded from NASA EARTH OBSERVATIONS, which is retrieved based on the observations of a spaceborne satellite called TERRA/MODIS. The time-series gridded sunlight distribution is obtained from the National Solar Radiation Database (NSRDB) Physical Solar Model (PSM). The LST and the sunlight distribution datasets are all adjusted to having consistent spatiotemporal resolutions and coverages with the kelp canopy extent dataset. That is, we take the average LST and sunlight distributions around CINP over the whole year as the annual LST and sunlight in each year between 2010 and 2022. Hence, the response variate is the annual kelp canopy extent around CINP, denoted by $y_i,\ i = 1,\ 2,\ \ldots,\ 13$. The covariates are time $t_1 = 2010,\ 2011,\ \ldots,\ 2022$ for $i = 1,\ 2,\ \ldots,\ 13$, annual LST $l_i$ for $i = 1,\ 2,\ \ldots,\ 13$, and annual sunlight distribution $s_i$ for $i = 1,\ 2,\ \ldots,\ 13$. These are the

*Keywords and phrases:* Kelps, Climate change.

basic covariates that we will be using and deriving from the raw dataset. More covariates based on them, such as the lagged variables, will be discussed in the next section as part of the analysis plan.

Here, we analyze the ability of the covariates to explain the response variate. According to a broad literature review, many factors can influence kelp population dynamics, including nutrient availability, temperature, light, the strength of current, and some more complicated factors, such as some special events that hugely affect the food chain (e.g., heat waves, grazer booms) [Zimmerman and Kremer (1986); Bekkby et al. (2014); Saccomanno, Bell and Pawlak (2023); Simonson, Scheibling and Metaxas (2015); Cavanaugh et al. (2011)]. Hence, we select annual LST as the proxy for the temperature and annual sunlight volume as the proxy for light. Nonetheless, other factors such as nutrient availability, the strength of current, and more complicated mechanisms are too hard to be collected in a dataset. Fortunately, they all somehow follow some patterns or trends over time. For example, the nutrient cycling in the ocean has a seasonality pattern, the current also has a pattern in terms of time [Giovannoni and Vergin (2012)]. Other more complicated factors can usually be explained with more and more serious climate change over time, so it might have trends [Trenberth and Fasullo (2012)]. In addition, LST, as one of the quantitative indicators of climate change, might help explain the factors. Therefore, we may expect the existing covariates, especially time and LST, to explain most of the factors. All in all, we expect the existing set of covariates we choose can explain the kelp population dynamics around CINP from 2010 to 2022, at least based on the mechanistic knowledge that we know about kelp population dynamics.

**3. Plan for statistical models and analysis.** In the study, we use a generalized additive model (GAM) to estimate the relationships between the covariates and the kelp population dynamics. This is because the underlying real model for kelp population dynamics is expected to be complicated, which requires more than one covariate and a flexible enough model to capture. And the relationships between each covariate and the response variate are different, which requires a customized function basis for each covariate. Also, we know more information about the relationship between the kelp population dynamics and each covariate from the previous studies. Based on the knowledge, we can determine what functions should be contained in the basis for each part of the model corresponding to one of the covariates constituting the GAM. The model can be more interpretable since GAMs are themselves interpretable, and it is a very close form of translation from the theoretical mechanism of how the kelp population changes [Hastie et al. (2009)].

First, the population dynamics at any time should rely on its previous conditions and time. For example, according to the population growth mechanical models, they are usually modeled by an ordinary differential equation about time, whose solutions must be functions of time [Braun and Golubitsky (1983)]. Also, as a creature, the population of kelps has patterns, especially periodic patterns, such as seasonality, biological pump, and so on [Bengtsson, Sjøtun and Øvreås (2010)]. Hence, we may use some sine and cosine functions from the frequency basis with the largest Fourier coefficients of the data as the basis. Or we may simply use B-spline functions to fit the patterns since there is no need for extrapolation. Therefore, the final set of covariates is expected to be time, LST, sunlight, lagged population, lagged LST, and lagged sunlight. For each covariate, there will be a corresponding function with a specific basis customized using mechanistic knowledge in the GAM.

There will also be a problem where sunlight may have multicollinearity with LST. Hence, a ridge regression penalty term or some other feature engineering techniques may be employed to solve the potential problem.

After we train a GAM, we may assess the accuracy of the model using cross-validation with an appropriate accuracy indicator, such as the Mean Absolute Error, Root Mean Squared

Error, R-squared, etc. Once the accuracy is satisfying to us, we can interpret the model. For the covariate time, we may set one of $t_5 = 2014,\ t_6 = 2015$, or $t_7 = 2016$ as the reference point since these are the years when the heat wave events occurred, which caused the dramatic kelp population to decrease and changes its overall trends and patterns according to previous studies. By inferring the uncertainty and seeing the patterns before and after the reference point, we may validate whether the theory is consistent with our observations and draw some conclusions about the population trends and pattern changes before and after the events. Similarly, for other covariates, we just set the reference points as the function values in $t_5 = 2014,\ t_6 = 2015$, or $t_7 = 2016$. Finally, we may validate the mechanistic studies and give some essential information about the relationships between the covariates and the population dynamics that can guide more proper policy making and adaptive measures regarding different environments (i.e., covariate values) in the future, which answers the research questions that we care about.

## REFERENCES

BEKKBY, T., RINDE, E., GUNDERSEN, H., NORDERHAUG, K. M., GITMARK, J. K. and CHRISTIE, H. (2014). Length, strength and water flow: relative importance of wave and current exposure on morphology in kelp Laminaria hyperborea. *Marine Ecology Progress Series* **506** 61–70.

BELL, T. W., CAVANAUGH, K. C., SACCOMANNO, V. R., CAVANAUGH, K. C., HOUSKEEPER, H. F., EDDY, N., SCHUETZENMEISTER, F., RINDLAUB, N. and GLEASON, M. (2023). Kelpwatch: A new visualization and analysis tool to explore kelp canopy dynamics reveals variable response to and recovery from marine heatwaves. *Plos one* **18** e0271477.

BENGTSSON, M. M., SJØTUN, K. and ØVREÅS, L. (2010). Seasonal dynamics of bacterial biofilms on the kelp Laminaria hyperborea. *Aquatic Microbial Ecology* **60** 71–83.

BRAUN, M. and GOLUBITSKY, M. (1983). *Differential equations and their applications* **2**. Springer.

CAVANAUGH, K. C., SIEGEL, D. A., REED, D. C. and DENNISON, P. E. (2011). Environmental controls of giant-kelp biomass in the Santa Barbara Channel, California. *Marine Ecology Progress Series* **429** 1–17.

CAVANAUGH, K. C., BELL, T., COSTA, M., EDDY, N. E., GENDALL, L., GLEASON, M. G., HESSING-LEWIS, M., MARTONE, R., MCPHERSON, M., PONTIER, O. et al. (2021). A review of the opportunities and challenges for using remote sensing for management of surface-canopy forming kelps. *Frontiers in Marine Science* 1536.

GIOVANNONI, S. J. and VERGIN, K. L. (2012). Seasonality in ocean microbial communities. *Science* **335** 671–676.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H. and FRIEDMAN, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* **2**. Springer.

SACCOMANNO, V. R., BELL, T. W. and PAWLAK, C. (2023). Using Unoccupied Aerial Vehicles (UAVs) to map and monitor changes in emergent kelp canopy after an ecological regime shift. *Remote Sensing in Ecology and Conservation* **9** 62–75.

SCHROEDER, S. B., DUPONT, C., BOYER, L., JUANES, F. and COSTA, M. (2019). Passive remote sensing technology for mapping bull kelp (Nereocystis luetkeana): a review of techniques and regional case study. *Global Ecology and Conservation* **19** e00683.

SILLS, J., ARAFEH-DALMAU, N., SCHOEMAN, D. S., MONTAÑO-MOCTEZUMA, G., MICHELI, F., ROGERS-BENNETT, L., OLGUIN-JACOBSON, C. and POSSINGHAM, H. P. (2020). Marine heat waves threaten kelp forests. *Science* **367** 635-635.

SIMONSON, E., SCHEIBLING, R. and METAXAS, A. (2015). Kelp in hot water: I. Warming seawater temperature induces weakening and loss of kelp tissue. *Marine Ecology Progress Series* **537** 89–104.

TRENBERTH, K. E. and FASULLO, J. T. (2012). Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010. *Journal of Geophysical Research: Atmospheres* **117**.

ZIMMERMAN, R. C. and KREMER, J. N. (1986). In situ growth and chemical composition of the giant kelp, Macrocystis pyrifera: response to temporal changes in ambient nutrient availability. *Marine Ecology Progress Series* **27** 277–285.