Analyzing Bird Species Distribution along the California Current Ecosystem

ANNA F HARVEY

Bellevue University

DSC 680: Applied Data Science

Prof. Catherine Williams

May 2, 2021

Abstract

The California Current Ecosystem plays an important role in the overall health of the eastern Pacific and holds many important fisheries. Every type of plant and animal within the ecosystem plays a role in maintaining its overall balance. Understanding what animals are present within the ecosystem and where they are primarily located can help guide key decisions, such as creating protected areas and tracking endangered populations. By studying specific types or species of animals within an ecosystem, it is possible to determine behavioral trends that in turn offer insight into environmental factors that may play a role in the distribution of species.

The purpose of this study is to examine the distribution of different animal groups observed during ecosystem surveys. 108 different groups were recorded at various stations. The animals recorded are primarily different bird species along with various broad categories of animals such as sharks and seals. Historical information about the locations of these animal is available on a daily basis between 2003 and 2012, providing foundational information to predict the appearance of animals at specific stations in the future. Accurate predictions of animal appearances at specific stations could allow for preemptive efforts to maintain environmental factors and resources in the areas to allow for those animals to flourish. This is particularly important to consider for salmonids and other fisheries in the area.

Introduction

Ecosystem-based management is a strategy employed by the National Oceanic and Atmospheric Administration to provide holistic awareness of ecosystem health.[1] When addressing an issue such as the number of salmon returning to spawn, multiple factors must be analyzed. While conditions such as commercial fishing rates and predator presence may seem like obvious drivers of salmon population change, the impact of each piece of the ecosystem must be analyzed as well. The relationship between

fisheries and the environment can be tenuous but there are ways to manage the impact fisheries have on an ecosystem by analyzing ecosystem variables. A broad-scale approach is necessary to fully understand the past, current, and future condition of an ecosystem.

Ecosystem-based management applies this concept by focusing on how different variables impact each other. The presence of a predator species may impact the population of their preferred source of prey. However, environmental conditions may also impact the population of a prey species which would then in turn affect the number of predators in the area. Climate effects, change in current systems, fisheries, predator-prey relationships, and competition and other factors play a role in the overall health of an ecosystem.

Problem Statement

Sea bird populations are impacted by the presence of prey species in a given area. As prey availability grows, so should the population of predatory sea birds. Additionally, if sea bird populations are particularly high in a given area, that may drive down the population of a prey species. The population of sea birds demonstrates an ecosystem's ability to sustain that population. Therefore, if the presence of sea bird species in specific locations can be predicted based on prior data, those predictions can be used as a baseline to determine overall ecosystem health.

Methods

The primary dataset for this project is the "Bird Density by Station in the California Current Ecosystem" dataset from the ongoing project *Juvenile Salmon & Ocean Ecosystem Survey and Salmon Ocean Behavior and Distribution*.[2] The dataset provides information about species (or other objects) observed at oceanographic stations along with details about the segment surveyed, density of the animals observed, latitude and longitude, distance from shore, and date observed. There are 108 different groups represented by species codes. 51 of those groups are identifiable bird groups that

follow the American Ornithologists' Union codes for bird species.[3] The other 57 groups are codes created by the project team to identify other animal and object groups such as sharks, seals, and boats. As I was unable to reach anyone on the project team regarding the codebooks for those miscellaneous groups, I chose to focus on the identifiable bird groups.

Data was cleaned and visualized in various formats to understand the data distribution for species. It was discovered that two species of birds were observed at much higher rates than the other bird species. These two species are the Common Murre (code: COMU) and the Sooty Shearwater (code: SOSH). In observation of overall species appearances at different stations, these two species appeared as outliers (Figure 1). An attempt was made to visualize the species density as a map, however the high number of species led to a map that was difficult to read and did not provide significant insight.

Attempts were made to apply Naïve Bayes models to the data. A categorical Naïve Bayes model was chosen as well as a Gaussian Naïve Bayes model, due to the categorical nature of the data. The target variable was the station. The feature variables were the species and species density. The goal was to determine if species appearances at stations could be predicted based on the data provided. These models resulted in inconclusive results. A different type of predictive model may need to be utilized to accurately predict species occurrences. Cluster models and time series models may be potential candidates for this data.

Results and Conclusions

Initial attempts at predictive modeling show no strong predictive accuracy for the appearance of specific species at a station. Further analysis must be rendered to understand if this is due to the nature of the model or because the data cannot be predicted in this fashion. The visualizations of the various species at stations do seem to indicate that certain species show up at certain stations more than others. This suggests that prediction may be possible. However, the nature of an ecosystem is

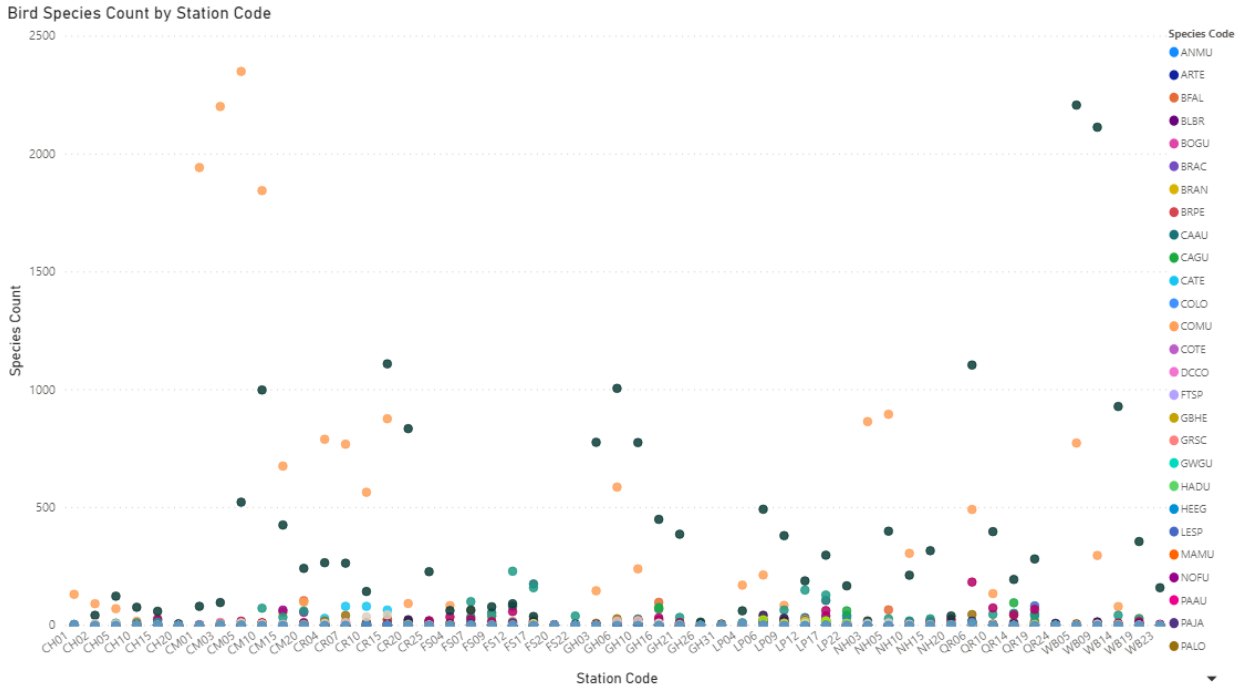
unpredictable and continuously changing. It is possible that the number of factors influencing the distribution of species may cause location predictions to be unreliable.

Future Investigations

Due to the high occurrence of the COMU and SOSH bird species, it may be valuable to focus predictive analytics on those two species alone. It would be valuable to investigate the relationship of these species to the ecosystem to understand possible contributing factors to their distribution as well as their impact on the environment. I will attempt to utilize a clustering model and/or a time series model to discover if a different type of model can yield higher accuracy.

APPENDIX

Figure 1. Scatter plot showing bird species COMU and SOSH with significantly higher counts at most stations compared to other bird species.



RESOURCES

1. NOAA, F. (2020). California Current Regional Ecosystem. https://www.fisheries.noaa.gov/west-coast/ecosystems/california-current-regional-ecosystem.

2. NOAA. (n.d.). Bird Density by Station in the California Current Ecosystem. NWFSC PARR Data - Bird Distribution and Abundance - Bird Density By Station in the California Current Ecosystem. https://www.webapps.nwfsc.noaa.gov/apex/parrdata/inventory/tables/table/bird_density_by_station_in_the_california_current_ecosystem.

3. IBP. (2020, September 16). Standardized 4- and 6-letter Bird Species ("Alpha") Codes. https://www.birdpop.org/pages/birdSpeciesCodes.php.

https://github.com/anhar421/Portfolio

10 Questions

1. Which stations showed the most bird activity?
2. What would it mean for a species to have high accuracy for a predicted appearance at a station?
3. What would it mean for a species to have low accuracy for a predicted appearance at a station?
4. What additional variables could increase the ability to predict species appearances at a station?
5. Do Common Murres (COMU) and/or Sooty Shearwaters (SOSH) play significant roles in the local ecosystem?
6. What factors might impact the ability for accurate data gathering when doing ecosystem surveys?
7. What strategies could be employed to collect more accurate data during ecosystem surveys?
8. What ecosystem variables that impact bird species appearances are controllable by ecosystem managers?
9. What ecosystem variables that impact bird species appearances are not controllable by ecosystem managers?
10. How would it help ecosystem managers to be able to accurately predict the appearance of specific species in specific locations?