Predicting Community Assemblage in The Rocky Intertidal of Acadia National Park

Preliminary Proposal

DSE6311 – SU2 2025
Nick Lagoni
7/15/2025

1 Introduction

1.1 Background

The Gulf of Maine is changing rapidly, warming at an accelerated rate compared to most of the globe (Friedland et al. 2020). Intertidal communities are experiencing a similarly rapid change, including changes in vertical zonation (Trott 2022) and introduction of invasive and range expanding species (Cohen et al. 1995, Yamada 2001, Epifanio 2013, Johnson 2015, Cheng et al. 2025). These changes may compound over time into significant changes to community structure. Because of this and the importance of these ecosystems ecologically and as a source of ecosystem services (including as interpretive spaces in National Parks, such as Acadia), it is important that we have an understanding of what the future holds for the rocky intertidal in the Gulf of Maine, and which factors contribute most to that future.

1.2 Research Question and Hypothesis

The primary research question posed by this analysis is: can we accurately predict how the community assemblage of the rocky intertidal in Acadia National Park will change in the near future, and what variables (including percent cover, count abundance, and temperature) are the strongest predictors of that. While long-term monitoring has been ongoing through the Northeast Temperate Network (NETN) in the Inventory and Monitoring Division of the National Park Service (NPS) since 2013, no significant analysis has been done on this subject since that protocol began.

I hypothesize that temperature and substrate percent cover will significantly predict the presence and abundance of motile invertebrates in the Acadia rocky intertidal over time. I also hypothesize that temperature and the presence and abundance of motile invertebrates will significantly predict algal percent cover in the Acadia rocky intertidal over time.

I predict that an increase in algal cover, including *Fucus vesiculosa*, *Ascophyllum nodosa*, several species of green algae, and several species of red algae, will be correlated with an increase in the abundance of motile invertebrates *Littorina littorea*, *Littorina obtusata*, *Carcinus maenas*, *and Tectura testudinalis*. I predict that increased temperature will be correlated with a decrease in *L. littorea*, *L. obtusata,* and *T. testudinalis*. I predict that algal percent cover of *Fucus vesiculoa* and *Ascophyllum nodosa* will be the strongest predictors of the various motile invertebrate abundances.

2 Data and Analysis

2.1 Data Source

NETN has a dataset spanning 8-years (2013 – 2021) of percent cover, count abundance, and temperature data for several sites in Acadia (Northeast Temperate Network 2021). I may be able to get more recent years data, all the way through 2025, directly from the Data Manager for NETN, if it has been input. This dataset is fairly robust but may not be large enough for machine learning. I will attempt to use it and pivot to a different dataset, location, or question altogether if necessary.

2.2 Variables

I'd like to examine every different species contained within the dataset to see if I can predict each over time. This may require multiple models to accommodate multiple response variables, or I may need to focus down to one response variable for simplicity's and interpretability's sake. A particularly impactful species may be one of the ecosystem engineers like *F. vesiculosus* (Westerbom and Koivisto 2022) or *A. nodosum* (Råberg and Kautsky 2007), or the recent and devastating invasive *C. maenas* (Cohen et al. 1995, Rayner and McGaw 2019), so these are all candidates for being the primary focus of this study. The predictor variables will include temperature, percent cover of each algal species and sessile invertebrates like *Semibalanus balanoides* and *Mytilus edulis*, and abundance of each motile invertebrate that is not the target variable.

2.3 Analytical Plan

My tentative analytical plan is to start with simple regression machine learning models, such as logistic regression, before transitioning to more robust and complex decision tree models. I anticipate random forest being a good fit for this data for determining importance scores and the large number of features contained in this dataset. Logistic regression (and the variations therein such as ridge regression and LASSO regression) may prove valuable for interpretability's sake and reduced computational power, especially if I need to run an individual model for each target species. I could see my dataset not being sufficient for any great modeling, or I could see one of the auxiliary variables (i.e, the person recording the data) turns out to be confounding.

2.4 Evaluating the Question and Hypotheses

I will know that my question has been answered whether or not the model performs. If the model performs well, then yes, we can predict the changes in community assemblage. If the model predicts poorly, then, at least with the current data available, we cannot.

The hypothesis relies on first the model performing significantly well and then relies on that model producing an interpretable and meaningful list of important variables. If the model is able to do both of these things, and our hypothesized predictors out to be strong predictors of whichever target variable I select, we can consider the hypothesis supported.

3 Technical Details

3.1 Language and GitHub Repo

I intend to code exclusively in R via RStudio, though I may dip into Python if there is a significant demonstrated need for it in my project. I may do some data cleaning work in Excel as well, depending on the structure of the dataset. My GitHub repo can be found here.

3.2 Additional Resources

I may benefit from seeking out some more Gulf of Maine related datasets that include the timespan from 2013 – 2021, to use as additional covariates for the model. This may include things like mean-high high-water levels, meteorological data, and park visitation data.

References

Cheng, H., M. D. McMahan, S. B. Scyphers, L. McClenachan, and J. H. Grabowski. 2025. Observations, perceptions and concerns of the American lobster industry regarding the range-expansion of Black Sea Bass. Marine Policy 173:106517.

Cohen, A. N., J. T. Carlton, and M. C. Fountain. 1995. Introduction, dispersal and potential impacts of the green crab Carcinus maenas in San Francisco Bay, California. Marine Biology 122:225–237.

Epifanio, C. E. 2013. Invasion biology of the Asian shore crab Hemigrapsus sanguineus: A review. Journal of Experimental Marine Biology and Ecology 441:33–49.

Friedland, K. D., R. E. Morse, J. P. Manning, D. C. Melrose, T. Miles, A. G. Goode, D. C. Brady, J. T. Kohut, and E. N. Powell. 2020. Trends and change points in surface and bottom thermal environments of the US Northeast Continental Shelf Ecosystem. Fisheries Oceanography 29:396–414.

Johnson, D. S. 2015. The Savory Swimmer Swims North: A Northern Range Extension of the Blue Crab Callinectes Sapidus? Journal of Crustacean Biology 35:105–110.

Northeast Temperate Network. 2021, January 1. Long-term Rocky Intertidal - Database. https://irma.nps.gov/DataStore/Reference/Profile/2289832.

Råberg, S., and L. Kautsky. 2007. A comparative biodiversity study of the associated fauna of perennial fucoids and filamentous algae. Estuarine, Coastal and Shelf Science 73:249–258.

Rayner, G., and I. J. McGaw. 2019. Effects of the invasive green crab *(Carcinus maenas)* on American lobster *(Homarus americanus)*: Food acquisition and trapping behaviour. Journal of Sea Research 144:95–104.

Trott, T. J. 2022. Mesoscale Spatial Patterns of Gulf of Maine Rocky Intertidal Communities. Diversity 14:557.

Westerbom, M., and M. Koivisto. 2022. Mussels and canopy-forming algae as ecosystem engineers: their contribution to community organization in the rocky sublittoral. Frontiers in Marine Science 9.

Yamada, S. B. 2001. Global invader: the European green crab.