Biodiversity in Coral Reefs

Tim Hogan

12/6/2019

Introduction

Coral reefs are an important marine ecosystem. Their structure allows them to host many species of organisms, which can act as an important resource for humans. However, as climate change and human impact apply pressure to these reefs, the diversity and number of these species may be threatened. Understanding the impact of coral reef degradation on the organisms within it is imperative, as it can allow management to more protectively protect these ecosystems.

This report will investigate how species richness can be affected by multiple features of coral reefs by analyzing data from the NOAA Pacific RAMP fish SPC 2010-2017 dataset (Heenan et al. 2017). This dataset compiles ~4700 reef surveys in 39 islands in the western Pacific Ocean, from 2010 to 2017. To assess species richness, I will calculate the Menhinick index at each site of the reef, and use a linear mixed effects model to assess how the composition of the seabed, substrate height, visibility, and year influence a site's Menhinick index.

Results

Estimations

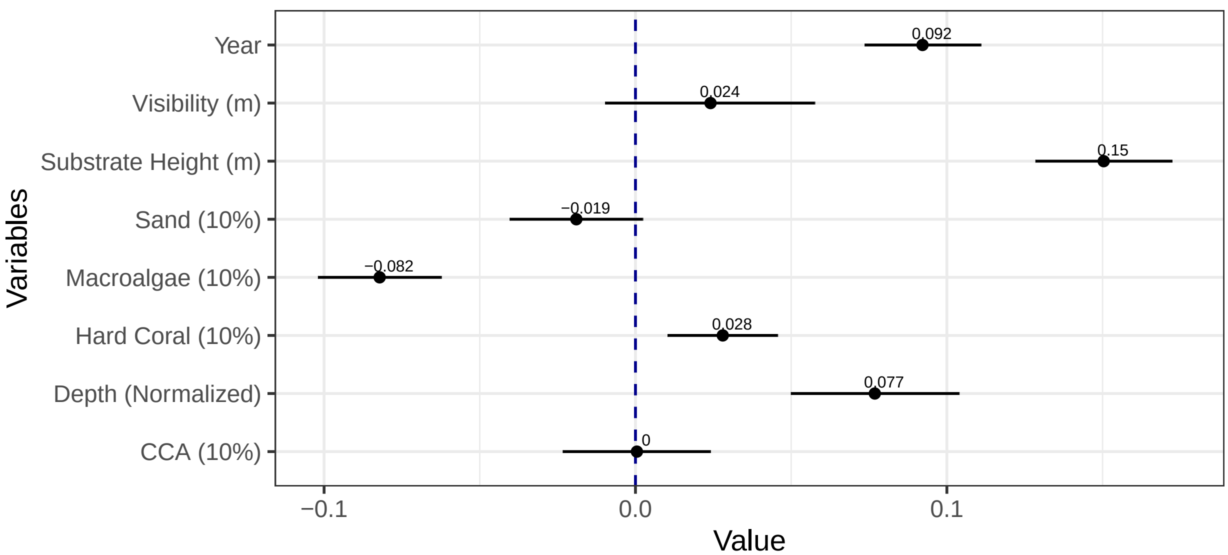

Figure 1: The predicted coefficients and 95% confidence interval for each fixed effect in the linear mixed effects model. The estimated value is displayed above each point estimate. 0 is highlighted with a dotted blue line.

Figure 1 graphically summarizes the estimated effects of the linear mixed effects model, as well as their 95% confidence intervals. Depth, substrate height, year, and hard coral have strong evidence of a positive effect on species richness. The effect of hard coral seems relatively small, with the Menhinick index estimated to increase by ~0.028 for each 10% increase in hard coral cover. Additionally, macroalgae is displayed to have a strong negative effect on species richness, with an estimated decrease of 0.082 in the Menhinick index for each increase of 10% of macroalgae cover. The effect of sand, crustose coralline algae, and visibility are found to be small, with confidence intervals crossing 0, indicating that our model does not have strong enough evidence to support that they significantly influence species richness.

Predictions

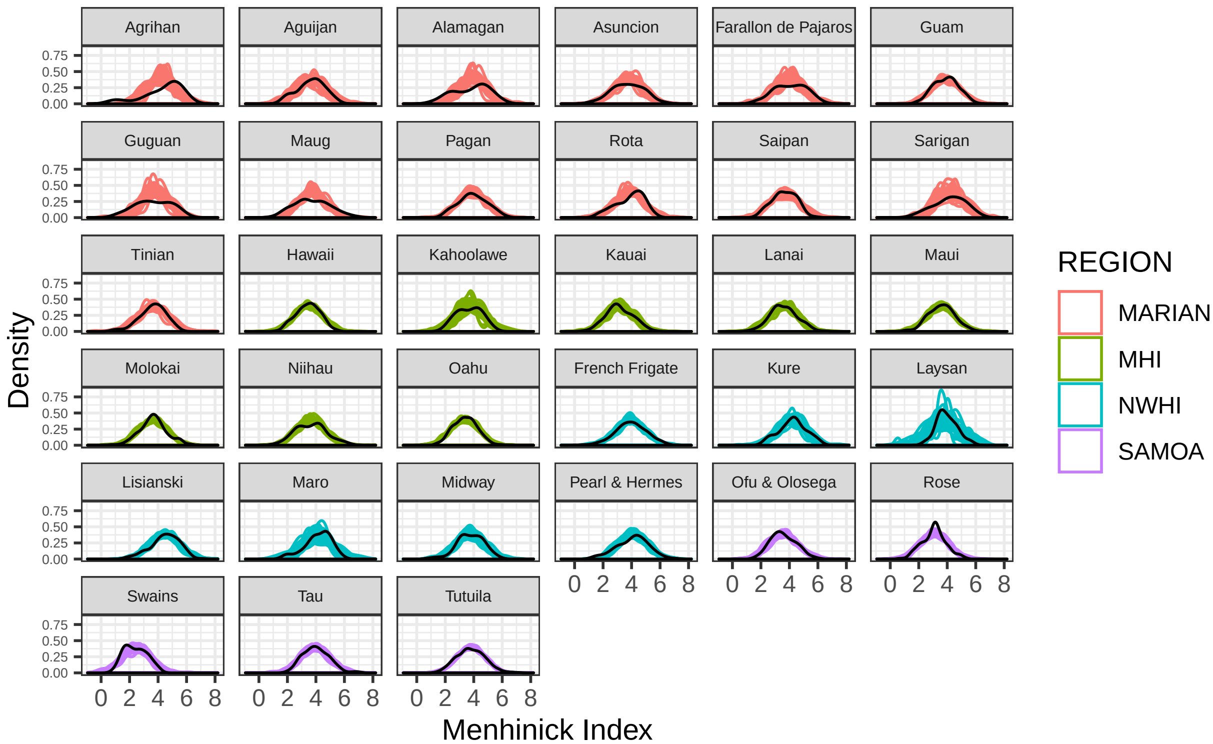

Figure 2: Density plots of posterior predicted values for Menhinick Indices on each island, colored by region, as well as the actual data's density (black).

Figure 3 displays the predicted range of values of the linear mixed effects model compared to the rest of the data, separated for each island. The range of estimated values seems to mostly capture the data from the islands, but does not always capture smaller features, such as the peak in "Rose", the skew to the right in "Maro", or the plateau close to the center in "Alamagan". Higher uncertainty ranges are also seen for some islands, mostly likely showing a lower number of sampled sites.

Model Checking

Figure 2 displays the residual and Q-Q plots for the linear mixed effects models. The residuals in the plot appear to distribute around 0. Additionally, there appear to be no obvious patterns or signs of heteroscedasticity. The Q-Q plot shows a linear relationship between the theoretical and sample quantities, showing that the model's predicted values follow a similar distribution as the data.

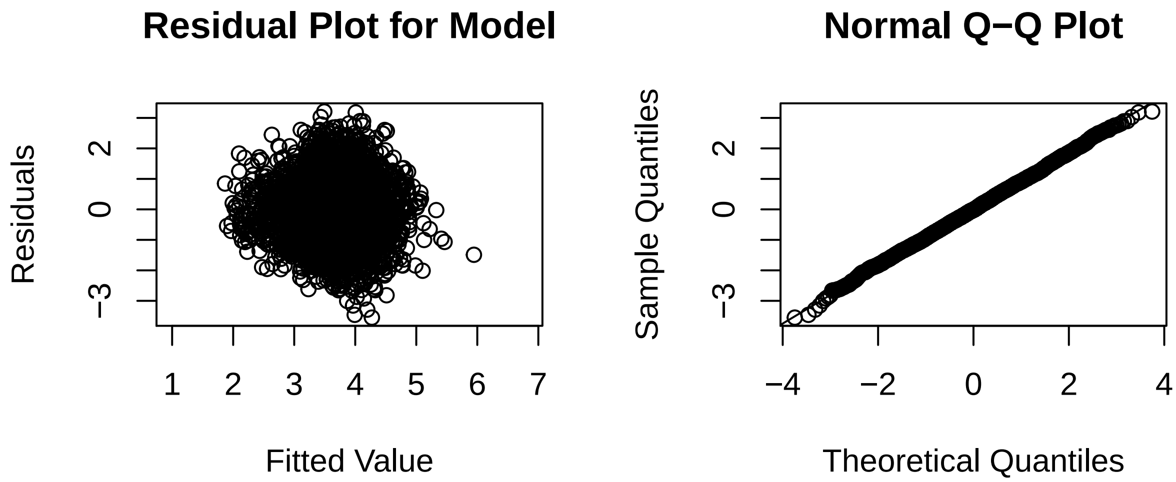

Figure 3: Residual and Q-Q plot of the linear mixed effects model.

Discussion

Additionally, it is possible that these coefficients may be influenced by the properties of the data. The predictions shown in Figure 4 do not capture all of the features of the data, and also consistently produce a higher number of Menhinick Indices at the mean. This can showcase the model utilizing the grouping effect of Islands to effectively, or may be showing some other underlying effects, such as low sample size. Additionally, before making conclusions, it may be important to consider what other events may have been occuring at these reefs, such as conservation efforts or changes in climactic effects. The impacts on each of these island regions may differ.

To improve this

Additionally, this model most likely only fits for island chains in the West Pacific. Coral reefs, as sensitive ecosystems, can be greatly influence c

Methods

Data Cleaning and Selection

Data was downloaded and cleaned using the tidyverse package in R. The data selected for the model was selected from 2012 to 2017, since earlier surveys did not consistently measure substrate height (Heenan et al. 2017).

The data contains survey conducted in five listed regions, being the main Hawaiian Islands (MHI), northwestern Hawaiian Islands (NWHI), American Samoa (SAMOA), the Northern Mariana Islands (MARIAN), and Pacific Remote Island Areas (PRIA). The distribution of these islands is displayed in Figure 4. In it, it shows that four of the regions are contained in island chains, whereas PRIA appears to be scattered across a large region. In order to account for this difference, PRIA was not considered in the model. The final dataset contained data from ~2800 survey, crossing four regions and 33 islands.

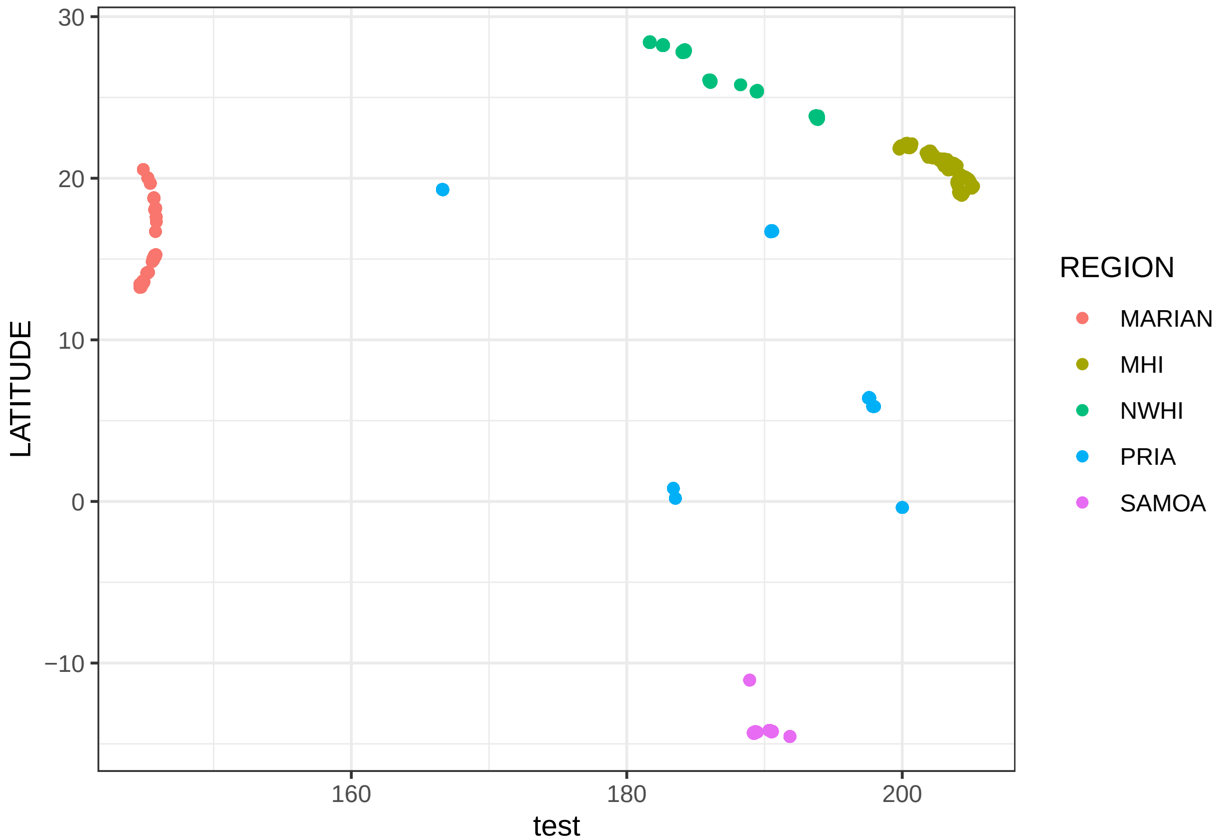

Figure 4: Map displaying each island surveyed in the dataset from 2012 onwards, colored by region.

For each survey, the Menhinick index was calculated at each site usign the equation $M = \frac{n}{\sqrt{N}}$, where n is the observed number of species at that site, and N is the observed number of organisms. The distribution of calculated Menhinick Index is displayed below in Figure 5.

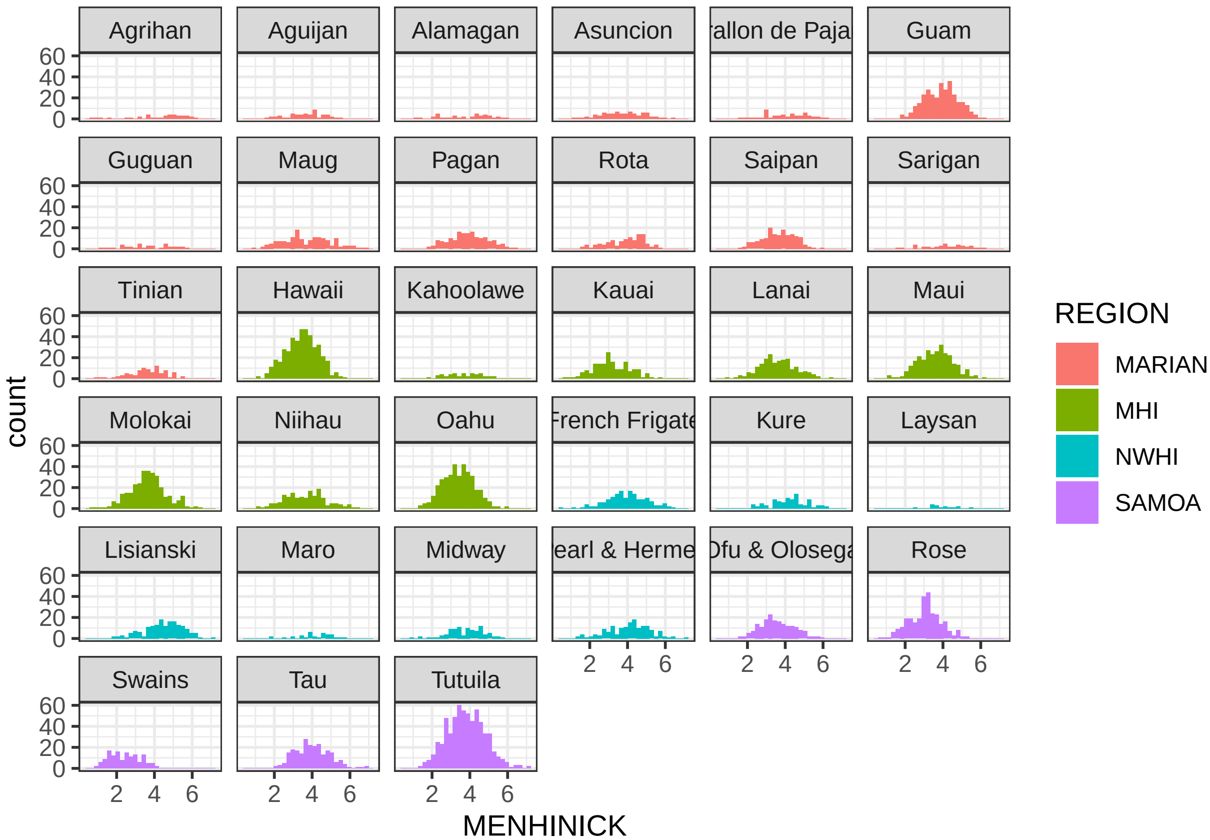

Figure 5: Histogram displaying distribution of Menhinick indices at each island, colored by region.

Model selection and transformations

Because of the nature of coral reefs, surveys done at the same islands will likely to have correlation with one another. They will most likely also have correlation with other islands in the same region. In order to account for this correlation, a linear mixed effects model was constructed using islands as random effects. Year, depth, visibility, the measured height of the substrate, and the percent coverage of hard coral, macroalgae, sand, and crustose coralline algae were used as fixed effects for the model. To improve interpretability of variables, depth was standardized by using a z-score, year was centered at 2014, and the percent coverage of each feature was divided by 10. Models were constructed using the lme4 and rstanarm packages in r.

Predictions were created using a Bayesian posterior predictive method with rstanarm.

Bibliography

Appendix