REGRADING DEGRADING:

ESTIMATING THE CURRENT DEGRADATION OF CORAL REEFS FOR WOODS HOLE

PRESENTED TO	PRESENTED ON	AS PROPOSED BY
Brigham Frandsen	Feb 24, 2025	Nate Leary, Audrey Moessing, Sam Lee, Andrew Goldston, Aidan Quigley

For the purpose of:

Building a predictive model that estimates coral bleaching from photographs taken by Woods Hole to aid their research efforts in coral preservation and assist their mission in advancing knowledge of the ocean

Overview & Scope

Coral reefs are some of the most biologically diverse ecosystems on the planet, but also some of the most threatened by increasing temperatures, ocean acidity, and other chemical or physical disruptions to their environment. In order to protect these rare and valuable aquatic communities, conservationists need to understand the extent of the damage that has occurred as well as identify any possible patterns between reef bleaching and any number of possible damage variables.

For our project, we plan to advance proactive reef care by generating usable data related to coral reef damage. Specifically, we want to use machine learning to identify data points that correspond to live coral, partially bleached coral, and fully bleached coral based on a series of photographs from the Majuro Reef. Once this data is generated, scientists can overlay a range of variables on the coral damage map and identify correlations with bleaching.

Data

The data for our project is based on several reef projects. As we develop our machine learning model, our training data will come from an open-source dataset associated with CoralScope. CoralScope currently operates as an automated coral segmentation program that differentials between coral/non-coral, growth form, and genus based on images. As a similar model to our intended program, we believe this coral data will act as an effective training set. The test set is drawn from single images from the Yellowfin Surfzone ASV in Majuro, Marshall Islands.

These photographs are obtained by a robot boat with a GoPro on the bottom that takes a picture of the seafloor every 10 seconds or so. Images captured per day range from 4,000 a day to 13,000 a day, with a total of 25,000 images to process. These photos were captured by the Woods Hole Oceanographic Institute, who has given us access in order to build this predictive model for use in their research efforts.

Method

We proceed with the methodology in three steps: data preprocessing, model fitting, and model evaluation. To process the data, we will use standard Python preprocessing libraries (namely, OpenCV and Pillow) to load our sample images, crop them into a square image, scale them into a 128 by 128 grid, and add contrast to bolster the signal that bleached spots on the coral images have with respect to the colors on the image as a whole. We will extract the RGB values to obtain 49,152 (128*128*3) image predictors. We will also augment this data with geospatial covariates such as where the image was taken, at what depth, and autoregressive spatial information (e.g. the classification—*Normal Coral, Partially Bleached Coral,* and *Bleached Coral*—of nearby observations). In this step, we propose an 80-20 training-test split. The training data will be further split into training and validation data sets to fine-tune our models. The test data set will only be fit to our final, fine-tuned model to assess final out-of-sample performance to select the best model.

We propose four main methods for prediction. For computational feasibility, we first propose an ordered logistic regression constrained by a ridge penalty. Next, we propose two machine learning approaches to strengthen predictive power. We propose a random forest model to fit our covariates to the three distinct classifications. Third, we propose a convolutional neural network (CNN) which has been known to predict well for image and spatial classification tasks. The hyperparameters of all models will be fine-tuned through randomized grid search methods using k-fold cross-validation on our validation set. Finally, we will compare our predictions to the expert-recommended (pre-trained) *CoralSCOP* model.

This model also provides hyperparameters which we will tune. For the model fitting, we will impose a loss function that penalizes false negatives the most. Although the classification set consists of three distinct classes, the classes are ordered: falsely classifying a coral that is unbleached that is, in reality, bleached, will result in a higher penalty than the converse given that our goal is to identify bleached coral and missing these patches has greater ecological consequences. Similarly, classifying a coral that is unbleached, that is, in reality, *partially bleached,* will again result in a higher penalty than the converse (although not as high of a penalty for misclassifying the fully bleached coral patches). For the computationally intensive tasks, we will attempt to fit the models via batch processing and or parallelization.

Lastly, after selecting the best model as given by the best out-of-sample loss, we will seek to rigidify its robustness by (1) if computationally feasible, bootstrapping the data to obtain uncertainty for our predictions that comes as a result of sample selection, and (2) depending on the model chosen, assess the relative importance of our covariates, and in particular, the relative importance for our augmented covariates (aside from the image data itself). We will do this in a variety of ways. We will assess the out-of-sample volatility by leaving each augmented covariate of interest out of the model fitting process and assess the change to the loss function. If our chosen model is a random forest model, we will use SHAPLY values to determine the relative importance of the augmented covariates. This will provide interpretable results in addition to strong predictive power.

Conclusion

In conclusion, over the next two months we will deliver a robust model that predicts the presence of coral and classifies it for Wood Hole's research. Our model will be based and tested on various methods for accuracy and computation time to find the model that works the best. Lastly, our model will determine the importance or predictive power of various features to provide further insight into interpreting what factors are contributing to the growing problem of coral degradation.