JOURNAL OF
EMERGING INVESTIGATORS

Article

Using machine learning to develop a global coral bleaching predictor

Indeever Madireddy[1], Rachel Bosch[2], Serena McCalla[3]

[1] BASIS Independent Silicon Valley, San Jose, California

[2] AECOM, Cincinnati, Ohio

[3] iResearch Institute, Glen Cove, New York

SUMMARY

Coral bleaching is a fatal process that reduces coral diversity, leads to habitat loss for marine organisms, and is a symptom of climate change. This process occurs when corals expel their symbiotic dinoflagellates, algae that photosynthesize within coral tissue providing corals with glucose. Restoration efforts have attempted to repair damaged reefs; however, there are over 360,000 square miles of coral reefs worldwide, making it challenging to target conservation efforts. Thus, predicting the likelihood of bleaching in a certain region would make it easier to allocate resources for conservation efforts. We developed a machine learning model to predict global locations at risk for coral bleaching. Data obtained from the Biological and Chemical Oceanography Data Management Office consisted of various coral bleaching events and the parameters under which the bleaching occurred. Sea surface temperature, sea surface temperature anomalies, longitude, latitude, and coral depth below the surface were the features found to be most correlated to coral bleaching. Thirty-nine machine learning models were tested to determine which one most accurately used the parameters of interest to predict the percentage of corals that would be bleached. A random forest regressor model with an R-squared value of 0.25 and a root mean squared error value of 7.91 was determined to be the best model for predicting coral bleaching. In the end, the random model had a 96% accuracy in predicting the percentage of corals that would be bleached. This prediction system can make it easier for researchers and conservationists to identify coral bleaching hotspots and properly allocate resources to prevent or mitigate bleaching events.

INTRODUCTION

Coral bleaching is a deadly problem that affects corals worldwide (1). Corals form symbiotic relationships with photosynthetic dinoflagellates in the genus *Symbiodinium.* The corals provide their symbiotic dinoflagellates with protection from predators and provide trace elements to the dinoflagellates. In return, the symbiote provides the coral with photosynthetically produced glucose (2). However, stressors in the ocean such as rising ocean temperatures and increasing irradiance often lead to corals expelling their vibrantly colored dinoflagellates in a process called coral bleaching (3). Other stressors include weather events such as increased hurricane activity due to the onset of anthropogenically-induced climate change (4). It is theorized that these stressors lead to the damage of photosystem II, a protein complex that aids in photosynthesis, in the chloroplasts of the photosynthetic symbiotes. Photosystem II is involved in the electron transport chain on the thylakoid membrane. The damage to this photosystem leads to the production of dangerous reactive oxygen species (ROS). These ROS damage both the dinoflagellate and the coral, inducing the coral to expel the dinoflagellate (5). Since over 90% of a coral's energy needs are supplied by its symbiotes, a coral depends heavily on its symbiotes (6). Bleached corals no longer have their primary energy source, and if symbiotes do not return to the corals, the corals will starve to death. In the Indian Ocean, coral cover has declined by up to 90% since 2016, and 14% of corals have been lost worldwide in just the last decade (3).

Coral reefs function as an effective storm surge barrier and protect coastal populations from devastating waves (7). Studying thriving reefs is also beneficial to humans since they harbor many compounds that can be used medicinally (8, 9). Corals are also vital to a stable ocean ecosystem, as corals provide shelter and food to a variety of ocean organisms such as clownfish and parrotfish. Furthermore, marine organisms that live in coral reefs are also a food source for human populations. In summary, coral reefs are vital to not only humans but also the multitude of creatures living in the oceans.

Marine scientists currently implement a variety of methods to prevent coral bleaching. Methods include repopulating bleached regions, selectively breeding bleaching tolerant corals, or creating artificial reefs (9). These restorative methods are only implemented to recover previously bleached reefs. However, it is crucial to identify a potential bleaching site before damage occurs. This can be a daunting task since coral reefs are massive, spanning over 360,000 square miles. Thus, due to the vast size of reefs, it can be challenging to quickly identify at-risk regions and implement the appropriate preventative methods before bleaching occurs. This work aims to solve the problem of locating high bleaching risk areas in the expansive oceans.

Developing an algorithm that can predict locations at risk for coral bleaching will allow scientists, researchers, and conservationists to implement targeted prevention or restoration methods, ensuring regions most at risk of bleaching are protected. Existing coral bleaching prediction methods are limited as they only pertain to specific regions of the planet. For example, the predictor developed by Kumagai *et al.* specifically focuses on the region around Japan, the predictor developed by Williams *et al.* focuses on the Palmyra

Atoll, and the coral bleaching predictor developed by Done and Wooldridge specifically focuses on the Great Barrier Reef (11-13). Although local models are highly accurate in predicting bleaching in these specific regions, corals grow across the world, and thus, a global model is crucial to provide better coverage. Thus, in this work, a global coral bleaching predictor was developed. This system considers various ocean parameters to determine the likelihood of bleaching occurring at any given specific global location. When water parameters are entered into this prediction system, the predictor calculates both the percentage of coral colonies at risk for bleaching at that location and whether the corals are at high risk or low risk of bleaching.

This project explores different machine learning models to predict future bleaching patterns and determine specifically which model is most accurate. We hypothesized that sea surface temperatures, larger sea surface temperature anomalies, and geographic location would be the most significant factors correlated to coral bleaching. These parameters are similarly explored in other bleaching predictors (11-13). In addition, we believe that either a probabilistic linear regression Bayesian model or random forest regressor model would be most accurate in predicting the likelihood of a coral bleaching event, as these models have been successfully implemented in prior works (11, 13-14). Through this work, a random forest model was determined to be the most accurate machine learning model and sea surface temperature, sea surface temperature anomalies, longitude, latitude, and depth were the parameters found to be most correlated to coral bleaching.

RESULTS

Modeling and Bleaching Correlation

The dataset included coral bleaching events recorded across the world (**Figure 1**). To determine the most accurate predictors of coral bleaching, we performed a linear correlation analysis to model the relationship between all the parameters provided in the global dataset (**Figure 2**). We found certain variables such as wind speed to have a high correlation (r = 0.2) to coral bleaching. However, multicollinear parameters needed to be removed by calculating Pearson's correlation coefficient to prevent the overfitting of the model to the dataset. This would limit inaccurate biases in the model. When parameters such as wind speed or thermal stress anomaly were found multicollinear (r>0.8) to each other, we removed both. We found that sea surface temperature (ClimSST), sea surface temperature anomalies (SSTA), depth (Depth_m), longitude, and latitude were the most significant parameters correlated to bleaching (**Figure 3**). According to the results of the linear analysis, SSTA, sea surface temperature and depth exhibited the highest positive degree of correlation with bleaching, 0.06, 0.05, and 0.03 respectively, while longitude and latitude showed the highest negative degree of correlation with bleaching, -0.06 and -0.01 respectively. Through a follow-up principal component analysis (PCA) on all parameters found in the original dataset, it was once again found that the same parameters were most correlated to bleaching. Four components were needed for the PCA (**Figure 4**). This redundancy better validates the selection of our parameters of interest.

We selected the random forest model to model these parameters of interest because it resulted in the R-squared value of 0.25, which was the highest R-squared value for all the models tested (**Table 1**) with the lazypredict library (15). R-squared refers to the goodness-of-fit of the model to the data. In addition, the random forest model had the lowest root mean squared error (RMSE) value, 7.91, indicating a closer relationship between the model and the data. The random forest did take significantly longer to make its prediction, however, at a run-time of 2.08 seconds, compared with

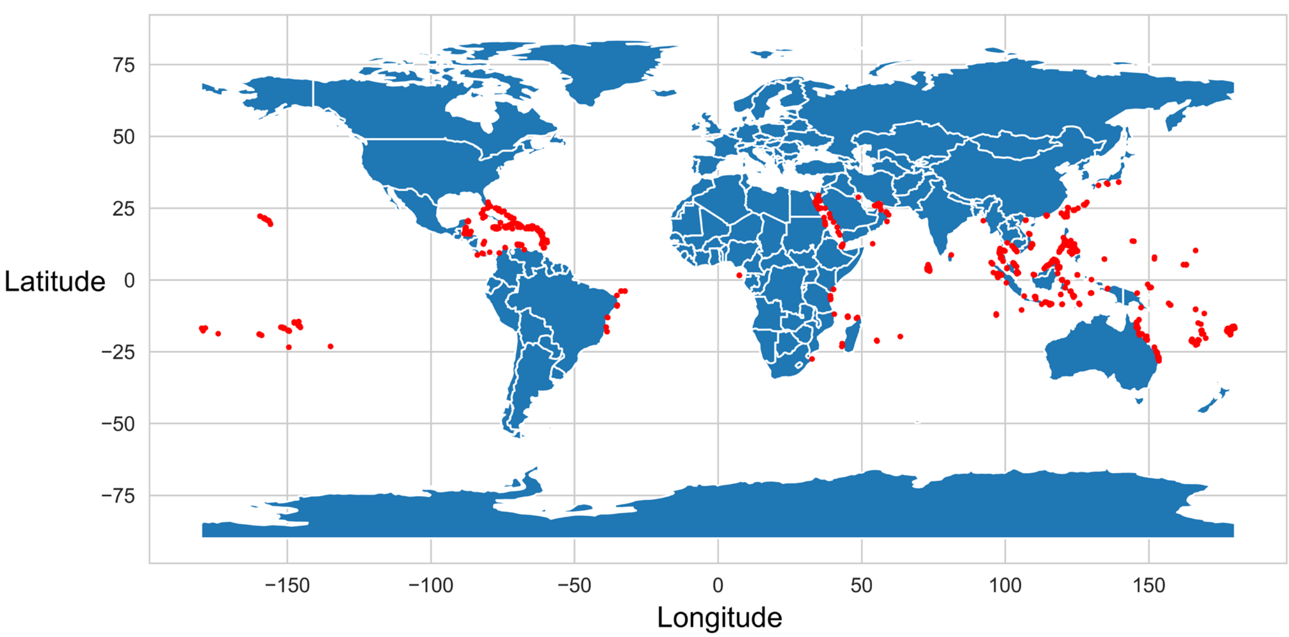

Figure 1: Global Map of Coral Reef Bleaching and non-Bleaching Events Included in Dataset. Red dots label the location of the bleaching events included in the dataset. Geopandas package (34) in python was used for plotting.

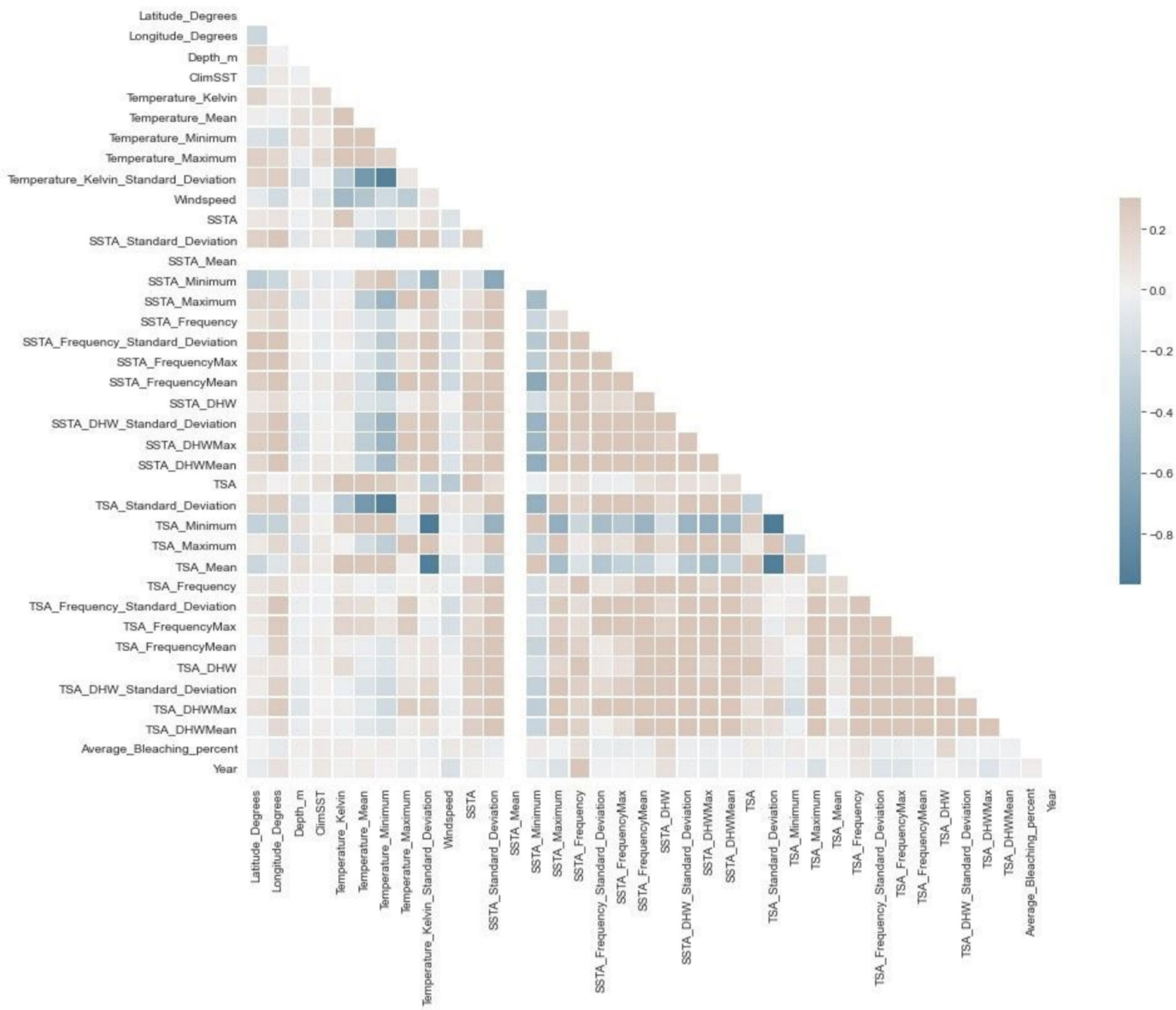

Figure 2: Linear Correlations Between all Parameters. Numerical parameters were taken from data produced by van Woesik et al. (24). The dataset may be accessed through the Biological and Chemical Oceanography Data Management Office. These linear correlations were used to identify multilinear parameters. TSA = thermal stress anomaly, SSTA_DHWMean = sea surface temperature anomaly degree heating weeks mean, SSTA_DHWMax = sea surface temperature anomaly degree heating weeks max.

a median run-time of 0.05 seconds for all models tested. GridSearchCV, a function in the scikit-learn python package, was used to determine the optimal hyperparameters for the random forest model. Tuning the hyperparameters, which help optimize the learning process, improved the model's accuracy. GridSearchCV found that 300 estimators, a depth of 16, a square root number of maximum features, and a mean squared error criterion were the optimal hyperparameters. The number of estimators is another way to quantify the number of trees in the random forest model, and the depth is the number of levels found in each decision tree. This model used the correlation coefficients and the results of the PCA, which determined that the longitude, latitude, depth, ClimSST, and SSTA were the least multicollinear parameters. Furthermore, k-means clustering with 2 clusters predicted the percent bleaching of a coral under the above conditions and whether there was a high or low likelihood of this bleaching occurring (**Figure 5**). Overall, 80% of the original dataset was used to train the model and the remaining 20% was used to test the random forest model. The random forest model was 96% accurate in predicting the bleaching percentage, which meant that, on average, it was 4% off in measuring the percentage of coral colonies expected to bleach under certain water parameters.

DISCUSSION

Random forest models take advantage of decision trees to form their predictions. However, not all datasets will work best with a random forest model as a large number of trees can slow prediction time. In the future, more data points will be collected from other sources that were not covered in this dataset in order to improve the model's global applicability.

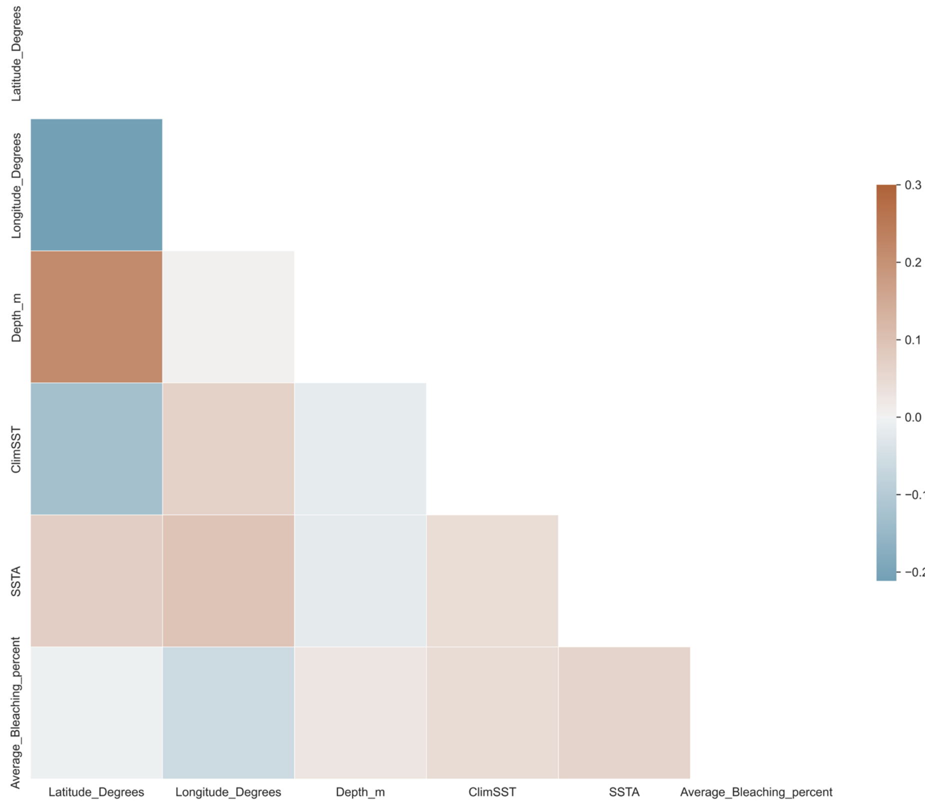

Figure 3: Parameters Most Closely Linearly Correlated to Bleaching. The figure depicts the parameters determined to be least multicollinear. The axes list the parameters of interest. The bottom row specifically focuses on the linear correlation of average bleaching to the other parameters of interest. Depth_m, ClimSST and SSTA have a positive linear relationship with coral bleaching while longitude and latitude have a negative correlation. SSTA = sea surface temperature anomalies, ClimSST = surface temperature.

The model outputted the predicted percentage of coral that would be expected to be bleached in that region of interest. The major benefit of this random forest model was its accuracy.

We identified the parameters of interest most closely correlated to coral bleaching, while removing multicollinear ones, and we developed a random forest model that predicts the percentage of corals bleached under these conditions. Longitude, latitude, depth, sea surface temperature anomalies, and temperature were the parameters of interest used to train the random forest model. In other words, the physical location of the coral in 3-dimensional space and the surrounding temperature are the most critical factors in determining the extent of coral bleaching. Other parameters could have been used in training the model, but they were removed to prevent overfitting the random forest model. For example, although thermal stress anomaly degree heating week (TSA_DHW), which is the sum of the previous twelve weeks when the TSA is found to be above one, was highly correlated to bleaching, it needed to be removed as it was highly multicollinear to other parameters. Interestingly, wind speed seemed to be highly linearly correlated with coral bleaching but was highly collinear with other parameters. An increase in wind speed is often correlated with hurricane activity, which could increase the likelihood of coral bleaching. Previous literature has suggested that hurricanes tend to track warmer waters (17). This might indicate that corals in warmer waters are not only more at risk from bleaching from the water temperature itself but also from the higher risk of damaging hurricanes. It has been established that hurricanes negatively affect coral reefs in various ways (18). One study has indicated a reduction in the coral recruitment of non-branching corals

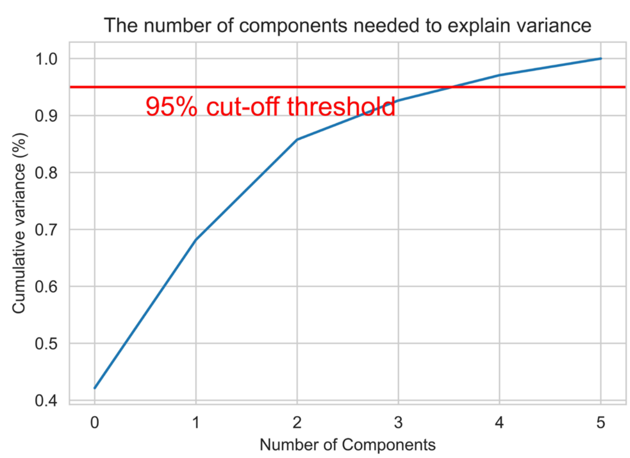

Figure 4: Number of Components Versus the Percent Cumulative Variance from PCA. The red line depicts the threshold at which 95% of the variance in the data is explained by the principal component analysis (PCA). The blue line depicts the variance explained versus the number of components used in the PCA.

Model	R-Squared	RMSE	Time Taken (s)
RandomForestRegressor	0.25	7.91	2.08
ExtraTreesRegressor	0.24	7.96	0.75
HistGradientBoostingRegressor	0.22	8.06	1.13
GBMRegressor	0.22	8.07	0.09
XGBRegressor	0.19	8.20	0.34
LinearSVR	-0.08	9.52	0.08
PassiveAggressiveRegressor	-0.20	10.03	0.02
DecisionTreeRegressor	-0.21	10.07	0.05
ExtraTreeRegressor	-0.33	10.56	0.02
AdaBoostRegressor	-0.44	10.97	0.10

Table 1. The Performance of Various Models on Coral Bleaching Prediction. The top five models are shaded gray while the bottom five models are shown unshaded. Models were ranked by their R-squared and RMSE values. A prediction time is provided for each model.

was linked to storm damage, in which increased storm or hurricane activity negatively correlated with coral recruitment (19). Another study found a coupling of both bleaching and hurricane intensity to be the most negative influence on coral recruitment and health (19).

There was a positive correlation between sea surface temperature anomalies and bleaching percentage. This indicates that when temperatures are higher than expected, an increased percentage of corals are expected to bleach. In agreement with previous research (1, 20, 21), ClimSST also had a positive correlation with bleaching, indicating that as temperatures increased, the likelihood of coral bleaching also increased. However, it seems that SSTA had a stronger influence on bleaching than ClimSST, indicating that deviations from the expected water temperature may be a larger factor for bleaching than the temperature itself. There also seems to be a slightly positive relationship between depth and bleaching risk. The deeper the coral, the higher risk of bleaching is. This might be because deeper corals are more sensitive to temperature fluctuations or stress in general. However, this does not necessarily imply that corals further below sea level are always more likely to get bleached since the Biological and Chemical Oceanography Data Management Office (BCO-DMO) dataset used in this work only includes bleaching events that have occurred between 0 and 12 meters underwater.

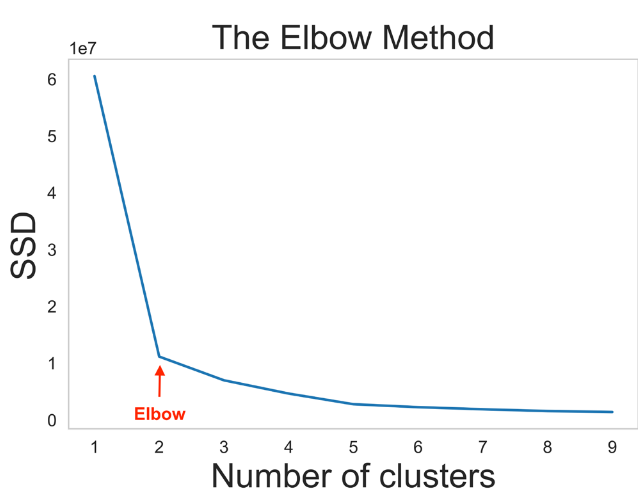

Figure 5: Optimal Number of Clusters. SSD, also known as SSE, is the sum squared error for each number of clusters. The curve shows the relationship between the error and the number of clusters tested. The arrow points to the optimal number of clusters which was two.

For k-means clustering of the data, clustering was employed in order to avoid any biases from manually classifying the data. K-means grouped the data points into two distinct clusters and determined that a bleaching percentage of 30% or above was enough to classify a data point in the high-risk group over the lower-risk group.

To develop a faster, more efficient, and more geographically extensive method of determining the likelihood of coral bleaching, a random forest model for determining the percentage of coral bleaching occurring under sea surface temperature, sea surface temperature anomalies, longitude, latitude, and depth was developed. This model used these parameters and outputted the predicted percentage of coral that would be bleached under this condition in addition to whether this percentage would be classified as low or high risk for bleaching. Highly multicollinear parameters such as wind speed or thermal stress were removed, and through a principal component analysis and linear correlation matrix, the parameters from the dataset most highly correlated to the percent of corals bleached were determined. Sea surface temperature, sea surface temperature anomalies, coral depth, longitude, and latitude were the parameters most highly correlated to bleaching. This work supports the initial hypothesis in which we claimed a random forest model, or a Bayesian model would be the most accurate. The lazypredict library was used to determine the best model for the predictions. After the random forest model was trained on the parameters correlated to bleaching, the model was then integrated into an application to make it accessible and easy to use. The major application of this project is that it allows for easy, global determination of at-risk coral bleaching locations from readily accessible data. NOAA regularly tracks both global sea surface temperature and sea surface temperature anomalies. This live data can be integrated into the model in a future version to provide real time updates on worldwide bleaching risk allowing for better monitoring and tracking of coral health worldwide (23, 24).

A future direction would be to apply this prediction model to a global map to pinpoint current locations at risk of bleaching. This heat map could color code the ocean with the relative percentage of bleaching expected at that location due to current ocean temperature parameters, depth, and location. A real-time map, as aforementioned, that imports data from multiple live sources could give real-time information as to what regions are at the highest risk could also be developed. Coral bleaching is a serious problem that damages ecosystems, hurts wildlife and humans. Better prediction tools like this model will help address coral bleaching by better identifying at risk areas.

MATERIALS AND METHODS

Dataset

Coral bleaching data was collected from the dataset assembled by van Woesek *et al.* (25) from BCO-DMO. The dataset included 9,666 instances of coral bleaching events and non-bleaching events (locations without measured bleaching) and the associated water parameters for both events. For each coral bleaching event, the dataset provided the percent of the coral colonies observed that were bleached by taking the average of four transected (cut) segments or slices of the coral. A bleaching value of 10 in the dataset would indicate that on average 10% of the coral colonies analyzed at that specific location were bleached. The dataset, which includes data from 1998 to 2017, provides parameters such as ClimSST, depth from the surface, windspeed, SSTA, longitude, latitude, and the name of the ocean in which the coral bleaching event occurred.

Data Cleaning

First, bleaching events with negative SSTA values were removed from the dataset. SSTA values were the difference between the weekly average sea surface temperature and the weekly climatological average sea surface temperature for that specific bleaching event. In short, an anomaly occurs when the sea surface temperature departs from the temperature expected for that region. To put this in context, a sea surface anomaly of 1.5-3.5 degrees Celsius is characteristic of the El Niño climate pattern (26). Because corals can tolerate lower temperatures, all data points with negative SSTA were removed to prevent bias in the analysis (27). Furthermore, locational data was cleaned up. Blanks in the numerical data set were labeled with N/A and counted in each column. 5,511 bleaching events were missing data in the city_3 column, 3142 bleaching events were missing data in the City_Town_2 column, and 289 bleaching events were missing data in the City_Town parameter. All three of these columns were meant to provide additional identifying information about where the bleaching event occurred, but due to the large number of missing data points, these three columns were removed. In addition, all non-numerical data such as the "Ocean," "Ecoregion," and "Realm," which all further provide more locational data, were also removed. These parameters were unnecessary because the datasets provided numerical longitudes and latitudes for each bleaching event. Finally, the spread of the data over time was also accounted for. This dataset provided data from 1998 to 2017. On average, 500-700 bleaching events were recorded each year. However, data was lacking between 1998 and 2002 when only 5 to 63 bleaching events were recorded in a year. Bleaching events from these years were also removed to allow an even distribution of data points over time. After data cleaning was completed, 6,112 bleaching events remained for further analysis.

Model Development

The aim of this work was to provide a numerical percentage value of the corals at risk of bleaching under certain conditions and to provide a non-numerical low- or high-risk classification to this prediction. To achieve the first objective, regression modeling was performed on the dataset to determine the parameters of interest most closely correlated to bleaching. Then these parameters were used in conjunctions with a random forest model to calculate the predicted bleaching percentage. For the second objective, these parameters of interest were clustered and then classified into two categories: high-risk and low-risk. When parameters are provided to the prediction system, the model determines whether it would be classified in a low- or high-risk cluster and thus whether there is a low or high risk of bleaching. A user interface was then built to use the developed model as an interactive system. This work was coded in the Python programming language (version 3.9.7) in a Jupyter notebook with open-source libraries such as scikit-learn (sklearn) (28) to perform regression modeling, seaborn for data visualization, and the tkinter library for app development (28-29).

Identifying Features of Interest

To identify features of interest, we tested the collinearity between the parameters (except for bleaching) provided in the dataset. Highly correlated pairs are parameters highly linearly or non-linearly correlated to each other. Average bleaching remained as the dependent variable throughout the course of this work. Highly correlated or multicollinear parameters need to be removed as they can cause overfitting of the machine learning model and increase computation time. The second figure shows a heat map representing the linear correlation between all the numerical parameters in the dataset. This map was generated with the seaborn package in Python. The Pearson correlation coefficient, r_{xy}, for each parameter was calculated according to the method of Pearson (30):

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad (1)$$

where n is the sample size, x_i is the independent sample point, and y_i is the dependent sample point, such that parameters with a Pearson correlation coefficient greater than 0.8 between each other were excluded from the model to prevent multicollinearity.

The original parameters were then simplified to only include parameters that were not highly correlated to each other. A heat map with just the remaining noncollinear parameters was generated. From this, it was determined that SSTA, ClimSST, Depth, Longitude, and Latitude were the parameters of interest most closely correlated to bleaching, and they were not highly correlated to each other. This identification of features of interest was further validated with principal component analysis (PCA). The Python package sklearn (31) was used to determine the number of principal components. Four components crossed the 95% threshold recommended for PCA. It can be seen that 4 components were the optimal number because it is the lowest number of components that either meets or exceeds the 95% threshold of explained variance. Once the PCA was run on the dataset (32), the same parameters as identified by the Pearson's correlation coefficient—SSTA, ClimSST, Depth, Longitude, and Latitude—were identified to be least multicollinear with the other parameters in the dataset and most closely correlated to bleaching.

Regression Modeling

To understand which models would work best to predict the percent bleaching, the lazypredict Python library (15) was used. This regression library used the parameters of interest, determined before, to predict the most accurate model. The library was used to run thirty-nine models automatically. All of the models in the library were regression models, such as the Tweedie regressor, Bagging regressor, and Huber regressor. For each model tested, the appropriate R-squared and root mean squared error (RMSE) values were calculated. R-squared is the square of the correlation coefficient, while RMSE is an indicator of the dispersion of the data and how far it deviates from the line of best fit or expectation. A higher R-squared and lower RMSE value is ideal. The R-squared was the highest for the random forest regressor, and the RMSE was simultaneously the lowest for this model out of all the models tested. The parameters of the random forest were hypertuned with GridSearchCV.

Clustering Features

After determining parameters of interest, the data was grouped to determine clusters of high bleaching percentages and low bleaching percentages. K-means clustering was employed. The elbow method helped determine the optimal number of clusters necessary to model the data. The number of clusters is chosen at the point where diminishing marginal returns in error reduction occurs when adding a new cluster. It can be seen that an elbow occurs at two clusters where the graph is no longer linear. At this point, additional clusters do not necessarily help to group the data better. From this, it can be concluded that two clusters best fit the data: one for high bleaching levels and one for low bleaching levels. A silhouette analysis (33), which determines the best degree of separation between a certain number of clusters, further validated the results of the elbow method. A silhouette coefficient was then determined for the each number of clusters. A larger coefficient which indicates better separation between the clusters is ideal. When tested on a range of one to fifteen clusters, two clusters had the largest silhouette coefficient of 0.7485. This metric indicates how well the clustering was performed. A higher coefficient correlates to a better clustering performance Eight iterations of k-means were then performed to find the centroids of these two clusters. Once the two clusters of data based on the parameters of interest were mapped, each bleaching event from the data set was assigned to a cluster.

Clustering for high- and low-risk areas was done with the k-means library (16). When the parameters were provided to the regression model, the program determined which cluster a bleaching event fell into.

App Development

A Python app was developed for the random forest regressor model with the tkinter Python package (22). Users can input the relevant parameters, and the app provides the percent of coral that would be expected to bleach at that location. The code for the app can be accessed here: www.github.com/IndeeverM/Global-Coral-Bleaching-Predictor/

ACKNOWLEDGEMENTS

The authors would like to thank Mr. YuanYu Lin, a graduate student from the University of North Carolina - Chapel Hill for his advice and thorough review of the manuscript.

Received: March 7, 2022
Accepted: June 28, 2022
Published: February 21, 2023

REFERENCES

1. Heron, Scott F., *et al*. "Warming trends and bleaching stress of the world's coral reefs 1985–2012." *Scientific Reports* 6.1, 2016, pp. 1-14, doi:10.1038/srep38402.
2. Xiang, Tingting, *et al*. "Symbiont population control by host-symbiont metabolic interaction in Symbiodiniaceae-cnidarian associations." *Nature Communications* 11.1, 2020, pp. 1-9, doi: 10.1038/s41467-019-13963-z.
3. Douglas, A. E. "Coral bleaching––how and why?" *Marine Pollution Bulletin* 46.4, 2003, pp. 385-392, doi: 10.1016/S0025-326X(03)00037-7.
4. Mann, Michael E., and Kerry A. Emanuel. "Atlantic hurricane trends linked to climate change." *Eos, Transactions American Geophysical Union* 87.24, 2006, pp. 233-241, doi: 10.1029/2006EO240001.
5. Smith, David J., *et al*. "Is photoinhibition of zooxanthellae photosynthesis the primary cause of thermal bleaching in corals?" *Global Change Biology* 11.1, 2005, pp. 1-11.F(4), doi: 10.1111/j.1529-8817.2003.00895.x.
6. Burriesci, Matthew S., *et al*. "Evidence that glucose is the major transferred metabolite in dinoflagellate–cnidarian symbiosis." *Journal of Experimental Biology* 215.19, 2012, pp. 3467-3477, doi: 10.1242/jeb.070946.
7. Beck, Michael W., *et al*. "The global flood protection savings provided by coral reefs." *Nature Communications* 9.1, 2018, pp. 1-9, doi: 10.1038/s41467-018-04568-z.
8. Cooper, Edwin L., *et al*. "Corals and their potential applications to integrative medicine." *Evidence-based Complementary and Alternative Medicine,* 2014, doi: 10.1155/2014/184959.
9. Harris, Daniel L., *et al*. "Coral reef structural complexity provides important coastal protection from waves under rising sea levels." *Science Advances* 4.2, 2018, doi: 10.1126/sciadv.aao4350
10. Boström-Einarsson, Lisa, *et al*. "Coral restoration–A systematic review of current methods, successes, failures and future directions." *PloS One* 15.1, 2020, doi: 10.1371/journal.pone.0226631
11. Kumagai, Naoki H., and Hiroya Yamano. "High-resolution modeling of thermal thresholds and environmental influences on coral bleaching for local and regional reef management." *PeerJ* 6, 2018, doi: 10.7717/peerj.4382
12. Williams, Gareth J., *et al*. "Modeling patterns of coral bleaching at a remote Central Pacific atoll." *Marine Pollution Bulletin* 60.9, 2010, pp. 1467-1476, doi: 10.1016/j.marpolbul.2010.05.009
13. Wooldridge, Scott, and Terry Done. "Learning to predict large-scale coral bleaching from past events: A Bayesian approach using remotely sensed data, in-situ data, and environmental proxies." *Coral Reefs* 23.1, 2004, pp. 96-108, doi: 10.1007/s00338-003-0361-y
14. Knudby, Anders, *et al*. "Mapping coral reef resilience indicators using field and remotely sensed data." *Remote Sensing* 5.3, 2013, pp. 1311-1334, doi: 10.3390/rs5031311
15. Pandala, Shankar Rao. "Welcome to Lazy Predict's Documentation!" Edited by Breno Batista da Silva, *Lazy Predict 0.2.9 Documentation*, 2019, www.lazypredict.readthedocs.io/en/latest/.
16. Lloyd, Stuart. "Least squares quantization in PCM." *IEEE transactions on information theory* 28.2, 1982, pp. 129-137, doi: 10.1109/TIT.1982.1056489
17. Fisher, Edwin L. "Hurricanes and the sea-surface temperature

field." *Journal of Atmospheric Sciences* 15.3, 1958, pp. 328-333, doi: 10.1175/1520-0469(1958)015<0328:HATSST>2.0.CO;2

18. Mallela, Jennie, and Crabbe, Michael James C. "Hurricanes and coral bleaching linked to changes in coral recruitment in Tobago." *Marine Environmental Research* 68.4, 2009, pp. 158-162, doi: 10.1016/j.marenvres.2009.06.001.
19. Crabbe, M. James C., *et al*. "Lack of recruitment of non-branching corals in Discovery Bay is linked to severe storms." *Bulletin of Marine Science* 70.3, 2002, pp. 939-945, www.uobrep.openrepository.com/handle/10547/293941.
20. Downs, C. A., *et al*. "Oxidative stress and seasonal coral bleaching." *Free Radical Biology and Medicine* 33.4, 2002, pp. 533-543, doi: 10.1016/s0891-5849(02)00907-3
21. Bonesso, Joshua Louis, *et al*. "Exposure to elevated sea-surface temperatures below the bleaching threshold impairs coral recovery and regeneration following injury." *PeerJ* 5, 2017, doi: 10.7717/peerj.3719.
22. Lundh, Fredrik. "An Introduction to Tkinter - TCL/TK." *An Introduction to Tkinter*, 29 Mar. 2003, tcltk.co.kr/files/TclTk_Introduction_To_Tkiner.pdf.
23. US Department of Commerce; NOAA; National Environmental Satellite Data and Information Service; Office of Satellite and Product Operations. "NOAA's Office of Satellite and Product Operations." *OSPO*, 31 Jan. 2013, www.ospo.noaa.gov/Products/ocean/sst/anomaly/.
24. US Department of Commerce; NOAA; National Environmental Satellite Data and Information Service; Office of Satellite and Product Operations. "NOAA's Office of Satellite and Product Operations." *Office of Satellite and Product Operations*, 6 Feb. 2014, www.ospo.noaa.gov/Products/ocean/sst.html.
25. van Woesik, Robert. "Identifying coral reef 'bright spots' from the global 2015-2017 thermal-stress event.", 2018, www.bco-dmo.org/project/762952.
26. Pierce, David W. "Hey! What Is an 'SST Anomaly'?" *El Niño Forecast, Sio Climate Research*, SCRIPPS Institution of Oceanography, 1997, meteora.ucsd.edu/~pierce/elnino/elnino.html.
27. Schoepf, Verena, *et al*. "Stress-resistant corals may not acclimatize to ocean warming but maintain heat tolerance under cooler temperatures." *Nature Communications* 10.1, 2019, pp.1-10, doi: 10.1038/s41467-019-12065-0.
28. Trappenberg, Thomas P. "Machine learning with sklearn." *Fundamentals of Machine Learning*. Oxford University Press, 2019. pp. 38-65, doi: 10.1093/oso/9780198828044.003.0003
29. Waskom, Michael L. "Seaborn: statistical data visualization." *Journal of Open Source Software* 6.60, 2021, 10.21105/joss.03021.
30. Kirch, Wilhelm. "Pearson's correlation coefficient." *Encyclopedia of Public Health*, 2008, pp. 1090-1091, doi: 10.1007/978-1-4020-5614-7_2569.
31. Pedregosa, Fabian, *et al*. "Scikit-learn: Machine learning in Python." *The Journal of Machine Learning Research* 12, 2011, pp. 2825-2830, www.jmlr.org/papers/v12/pedregosa11a.html
32. Smith, Lindsay I. "A tutorial on principal components analysis.", 2002, www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
33. Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of Computational and Applied Mathematics* 20, 1987, pp. 53-65, doi: 10.1016/0377-0427(87)90125-7.
34. Graser, Anita, and Melitta Dragaschnig. "Exploring movement data in notebook environments." *IEEE VIS 2020 Workshop on Information Visualization of Geospatial Networks, Flows and Movement (MoVis)*. 2020, move.geog.ucsb.edu/wp-content/uploads/2020/10/MoVIS20_paper_4.pdf