Assessing the anthropological impacts on Coral Reefs in the Pacific

Lavinia Schlyter
EPF Lausanne, Switzerland

June 15, 2021

Abstract

Various are the studies which underline the impact of climate change on coral reefs though anthropological effects remain less studied. By collecting data from various sources, a baseline dataframe was set up as a means to establish whether anthropological factors affect coral cover, more specifically hard coral cover. Basic models are shown in order to give an insight in preliminary results. Known results were successfully reproduced, some anthropological factors were deemed significant in relation to coral cover and predictive modelling is shown to be limited in this framework.

Keywords: Coral reefs, Great Barrier Reef, DHW, predictive modelling, anthropological factors

1 Introduction

Over the years, the increasing threat on corals by natural and anthropoligical factors [1], [2], [3], has caused a surge in interests to study [4] and protect them. Ecosystems provide a range of services that are fundamental, with up to one third of marine wildlife depending on coral reefs for survival and at least 500 million people's livelihoods [5]. Despite the numerous efforts; the abundance and diversity of corals worldwide continue to decline [6], [7]. Its threats are numerous and vary in space and time. The most notable natural drivers for coral health are temperature, photosynthetic active radiation, UV radiance, but also high wind speeds which by their force prove destructive not only for humans [8], [9], [10]. Anthropological effects on corals have been modestly studied where factors such as over-fishing, terrestrial run-off, caostal development and tourism activities have all been identified as high drivers of coral reef decay [3], [11], [12]. In this project, the main goal is to set up a framework using existing data by merging them and encoding specific features. We then proceed with data analysis such as to better understand the collected data; this represents the major part of the project. The encodings are used in a model taken from literature in order to compare and form a solid baseline for further work. Several predictive models are additionally presented to show a baseline result, though as discussed in the further sections, the presented models only scratch the surface of the real work at hand.

2 Data collection and pre-processing

2.1 Study area

The area of focus is located in the Western Pacific in the Banda, Arafura, Solomon and Coral Sea. The countries of focus are Australia, Timor Leste, Indonesia and Solomon, all of which have been surveyed at least once by the Caitlin expedition [13]. Coral reefs in the region include The Great Barrier Reef and the coral triangle region reefs. The subset of the region used in this study is 100-160°E and -24-4°N.

2.2 Literature review

We first started by exploring papers for a literature review such as to select the environmental variables known to cause harm to hard corals, these are listed in Table 1. You may also see the different factors significance as defined by the various papers. The `X` symbol denote a factor which was used/discussed but no significance level is given. We then collected anthropological data based off literature or curiosity. These are discussed later but a summary is given in Table (6).

Health Drivers			Temperature [C°]	Chlorophyll [mg/m^3]	PAR (photosynthetic active radiation) (Einstein/(m^3/day))	Wind speed [m/s]	UV irradiance [mW/(m^2 nm)]	Deep Waters [m]
Papers	Paper Date	Location						
Bright spots among the world's coral reefs [JE Cinner] [12]	2016	World						High (positive with depth)
Global Effects of Local Human Population Density and Distance to Markets on the Condition of Coral Reef Fisheries [JE Cinner] [11]	2012	World						
Modelling susceptibility of coral reefs to environmental stress using remote sensing data and GIS models [J Maina] [14]	2007	Western Indian Ocean	High	Low	High	High	High	
The 27–year decline of coral cover on the Great Barrier Reef and its causes [De'ath] [15]	2012	Great Barrier Reef, Australia	High			High		
Remote sensing of coral reefs and their physical environment [Hedley [16]	2004	World	High	X	High	X	High	

Table 1: Health drivers according to literature review

2.3 Coral Coverage data

The coral data is provided by the *XL Catlin Seaview Survey Project* that was developed as a collaboration between the University of Queensland and ocean conservation non-profit Underwater Earth, available here [13].

The goal of their project is to collect rapid, detailed, globally distributed scientific surveys of coral reefs to support research and conservation. The data collection took place between 2012-2018 with a custom camera that produced high quality photographs of the reef at transects that were usually 1.5-2 km long. Images were taken between 0.5-2m to ensure consistent spatial resolution.

The raw data is a collection of over one million images covering around $1m^2$ of the sea floor. Deep learning algorithms were used to estimate the benthic cover from each photo. The relative abundance of each benthic group was also estimated.

For this project we are interested in the benthic cover of the corals derived from the images and we will thus not focus on the raw images.

The data set of interest contains the proportional cover of reef cover between *hard and soft corals, other invertebrates, algae or others.* The data set also contains the start and end latitude and longitude mapping for each survey. Australia (AUS) contains 261 surveys, Indonesia (IDN) 114, Timor Leste (TLS) 26 and Solomon (SLB) 20, totalling to 421 over a range of 5 years (2012, 2014, 2016, 2017, 2018). The distribution of the surveys are shown in Figure (1). We note that Australia has data taken at four different years, Indonesia two whereas Timor Leste and Solomon have single year surveys. The issues that arise from this particular distribution are discussed in further sections.

The different points where the surveys were taken are highlighted in Figure (2).

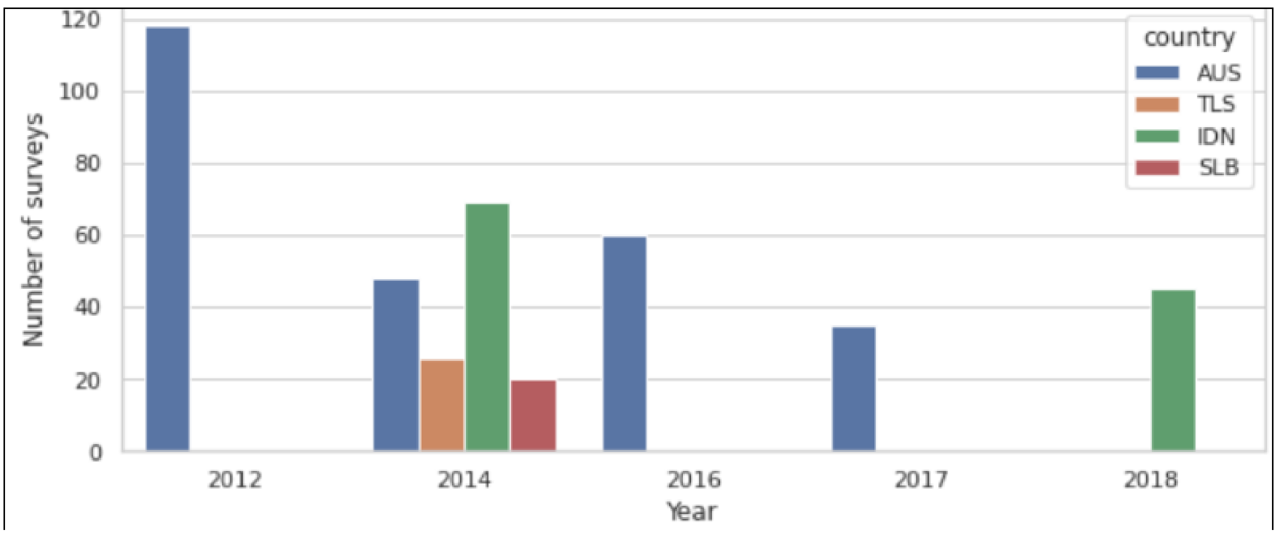

Figure 1: Distribution of surveys in countries of interest

In this study, the focus is on the hard corals proportion given in percentage. Data processing included filtering out the surveys from countries of interest out of the 860 surveys and the coordinates of the start point for the survey were taken as the geometry point for the respective survey. This estimation compared to more precise methods is chosen because of the relatively short distance between end and start geometries; thus not imitating a large mistake. The distance between two survey points from the same year are around 1km apart. The GeoDataFrame containing the discussed data will be the base sheet and further data will be added to the frame. Because 5 surveys have 4 duplicates (same date, same survey) the choice of keeping the first one is only arbitrary. Only the month and year of the survey were kept as a means to facilitate the merging with additional monthly data. The data frame currently consists of 416 rows.

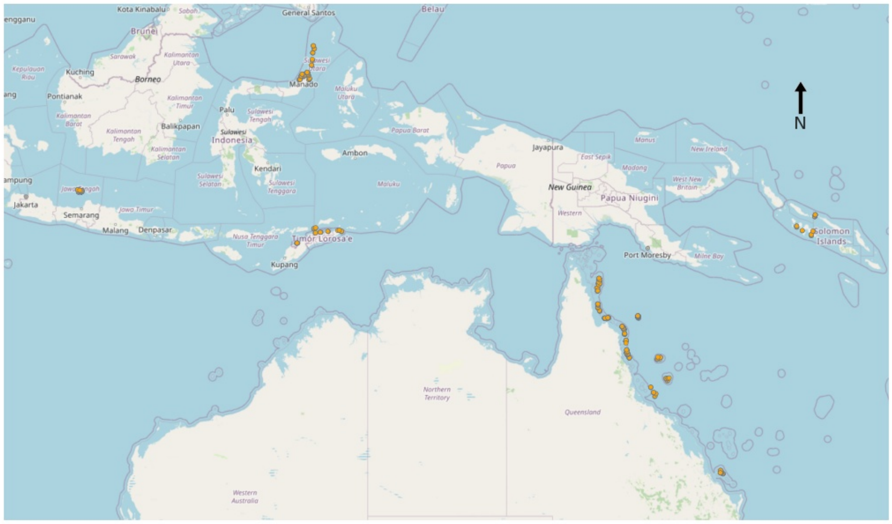

Figure 2: Map showing the area of interest

Environmental data

2.4 Sea Surface Temperature

The effects of temperature on coral reefs are long studied and have proved to be an essential component for the well being of hard corals [17], [15], [14]. The daily sea surface temperature (SST) measures were averaged by month at a 5 km resolution for data ranging from 1985-present taken from the NOAA Coral Reef Watch website [18] under a CSV format with degree Celsius units. Using shape files from the different countries at a 5 km resolution and using their respective "id", we merge the SST and its geometry based on the "id" present in both the SST and the shapefiles. This is done separately for each of the 4 countries. We then find the intersections between the polygons and the survey points. When these were zero due to the limit size of the shapefile, we take the centroid of the polygons and find the nearest point to each. Because the effects of temperature on the corals are not direct, several variables were encoded which take into account the delayed response to the variable.

The variables were encoded for 3, 6, 12, 120 (10 years), 240 (20 years) months before the survey was taken. Following results in O. Selmoni [in press] as well as some preliminary outcomes found using these months; only the 1, 10, 20 year(s) were kept.

2.5 Degree Heating Week

The degree heating week (DHW) is a measure how much heat stress has been accumulated in an area over the past 3 months it is computed using eq (1)

$$DHW_i = \sum_{j=i-83}^{i} \frac{HS_j}{7} \tag{1}$$

where HS_j is the Coral Bleaching Hotspot whose value is $\geq 1C$. The factor $\frac{1}{7}$ is used to express the DHW in terms of $(C - weeks)$ and we sum over 84, because we have around 84 days in 3 months (12 weeks).

The pre-processing steps are the same as for SST.

2.6 Wind speed

High wind speed associated with cyclones are known to be a significant factor influencing coral cover [15], [14]. The 10m wind speed, ocean surface currents were downloaded through the Copernicus API using the delimited area outlined in (2.1) for averaged monthly data ranging from 2000-2021 with a resolution of 0.5°x0.5° using a GRIB format. The feature corresponds to the horizontal speed of the wind at a height of ten meters above the surface of the Earth with units of m/s.

Aiming at pre-processing the data file, the format was translated to NetCDF using `Xarray` and then into a GeoDataFrame using the coordinates and including the variable of interest is "si10". The wind data is then merged to the survey locations where the closest wind measure is taken for the survey data of interest. Here again, after a first iteration through the model, we extended the data by merging only by location. That is; we kept all the monthly measures of the wind speed preceding the survey date and encode them in the same way as for SST and DHW.

2.7 Solar radiation

Solar radiation was deemed highly important in both [14] and [16]. This data set is the same as the wind speed but the variables are *surface net solar radiation and surface net solar radiation clear sky* with units $[\frac{J}{m^2}]$. It thus goes through the same pre-processing steps.

2.8 Chlorophyll

The concentration of chlorophyll is characterized as a non significant feature in [14] for coral cover, though we deemed it interesting to add. The mass concentration of Chlorophyll α $[\frac{mg}{m-3}]$ data were downloaded through the Copernicus API using the NetCDF format and taking the same delimited area for the years 2012, 2014, 2016, 2017, 2018 with a horizontal resolution "Sinusoidal equal-area grid" 4km x 4km using version v5.0 which corresponds to the updated processing chain. Because the data is very heavy and taken daily, the data were taken arbitrary on the 1st day of each month. `NaN` values were removed, the data was then merged to the dataframe by taking the month of interest and computing its closest values.

(Due to the size of the data, the laptop used was not powerful enough to encode the values as done for SST, DHW, Wind speed.) No extra time was spent into encoding as it was deemed as a feature of low impact in [14].

2.9 Ocean depth

The depth of the ocean is regarded as a beneficial condition for corals as it may provide a refuge from disturbances. The variable is downloaded through Copernicus in NetCDF format with units in meters. Because we consider that within 20 years the depth of the ocean is not varying, we take the most recent map available for the most accurate data [2020] with a resolution of 15 arc seconds corresponding to around 450 meters at the latitudes of interest. The closest computed depth to each survey point is then attributed to it in order to approximate the depth. Some heights were slightly positive and were thus set to zero.

2.10 Photosynthetically Active Radiation

Photosynthetically active radiation (PAR) is a measure of the light available for photosynthesis and has been shown to be a stress indicator for coral reefs [14], [9].

The PAR data was downloaded from NASA under the NetCDF format with units of $\frac{\mu Einstein}{m^2 \cdot s}$ with a monthly temporal resolution and 4km spatial resolution. As done in previous steps, they are merged based on nearest measurement to survey and the dates which coincide with the survey dates. The limitations of this approach is discussed in section 5.

Anthropological data

In this section we discuss data which was collected for the purpose of assessing the research question; whether anthropological data influences the percentage of coral cover. Various papers inspired which data seemed valuable to collect in addition to ones collected due to interest and curiosity.

2.11 Distance to closest agricultural land

Human activities related to agricultural land use are known to affect corals in a negative way [2].

In order to encode this variable, we compute the distance from each survey point to the closest agriculture land, where we expect a decrease in coral cover for smaller distances.

The land cover data was downloaded using Google Earth Engine from the Copernicus data base. It was downloaded at a 5 km resolution and masked with $1's$ in places where values corresponded to agricultural land and $0's$ for the rest. Because the coordinates were given in (x,y); they were reprojected using `gdal wrap` to get latitudes and longitudes. We then compute the closest distance to points where the value is one, in order to compute the distance to the closest agricultural land.

2.12 Night lights

The VIIRS nighttime layer is used with the aim to show patterns of human activity and energy behaviors.

The *tiff* layer was downloaded from the Google Earth Engine at a 15 km resolution with the averaged monthly radiance over a period of a year between 2012-2018. Negative and `NaN` values were removed. The data was encoded such as to have a column with the values for the survey's year. Pre 2012 data was not available to enable the computation of more long term averages though we do not expect the night light values to drastically change in the time range used.

2.13 Distance to marine protected areas

A number of marine protected areas have emerged with the intention of providing a safe space for sea life [19].

In an attempt to encode this data, it was chosen to use the distance to the closest marine protected area (MAP), this would enable to have a certain leverage on values which are just at the border of the MAP but not inside.

Shapefiles containing polygons and points of marine protected areas were downloaded from Protected Planet [19]. For the polygons the centroid is taken and we then compute the closest distance to the survey points in meters. In order to calculate the distance we must first project the latitudes and longitudes onto a Cartesian frame. This is done in Python using `rasterio transform` with the appropriate transform function [20].

Improvements to this encoding are discussed in section 5.

2.14 Fishing

With overfishing being considered as one of the three most significant threats to coral reef ecosystems [21], we were interested to see whether we could quantify this impact.

The fishing effort data which contains the coordinates and the number of hours spent fishing at a coordinate with a 100th degree resolution was downloaded from the Global Fishing Watch datasets. The years downloaded were 2012, 2014, 2016, 2017, 2018. The coordinates must be projected into (x,y) in order to compute the distance to the fishing stops. This is done using `PyProj` from the inital $CRS = epsg : 4326$ to $epsg : 3112$ which is used as an approximation for areas around Northern Australia. All distances to the survey points are then computed and for distances of less than 50 km, we consider their fishing hours to be summed; this number is chosen arbitrarily and further research should be taken in order to find out a meaningful threshold for the distance. Further limitations to this encoding are discussed in the section 5.

2.15 Distance to closest port

It has been shown has distance to market has a strong explanatory role in reef structure globally [11]. Instead of looking at the markets we will focus on the distance to the closest port. The ports around the world were downloaded through the Global Fishing Watch database. The *tiff* file was converted into NetCDF using `gdal translate`; this format enables for efficient processing of voluminous data. We bound the data using the area of study and compute the closest distance to each survey point. In this case, we take the latest available dataset making the hypothesis that in the time frame and in the region of interest, the ports will not vary.

2.16 Table of summary

A table of summary for the data collected is completed in Table (6).

3 Explanatory data analysis

Following the collection and pre-processing of the data, we proceeded into some data exploration and analysis.

3.1 Distributions

The first step was to look at the distribution of the different coral data variables (averaged) as shown in Figure (3).

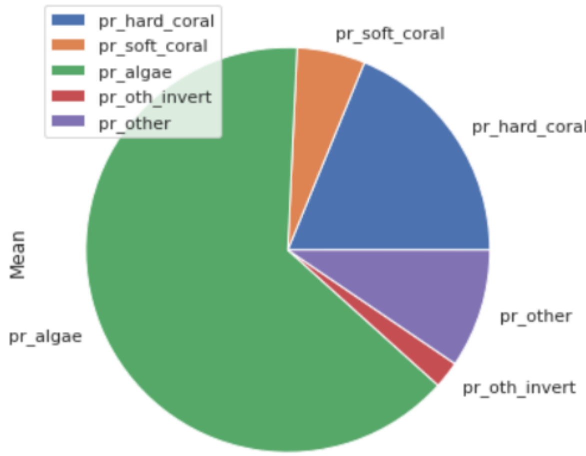

Figure 3: Coral data distribution across the study area

We highlight the greatest proportion is composed of algae followed by hard coral and other. This concurs with analysis pursued in the *coral triangle area* by Burke [22] who mentions an overgrowth of algae.

The second step was to look at the variation of hard coral and algae across the years for all surveys. This is presented in Figure (4). It seems as if the hard corals and algae have opposite trends. As one increases the other decreases. Following a discussion with Oliver Selmoni, this behavior is rather typical and expected. Figure (5) show the distributions through time. This leads us to believe in a decrease in both hard and soft corals.

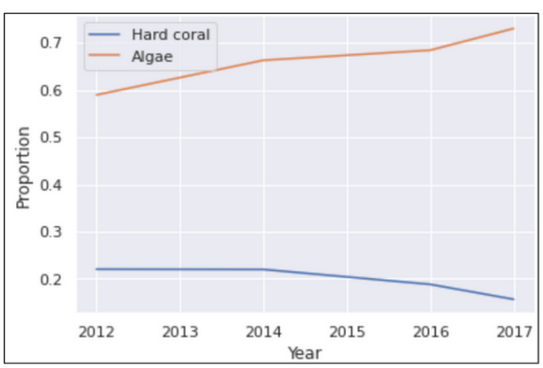

Figure 4: The proportion of algae and hard coral as a function of the years for all surveys

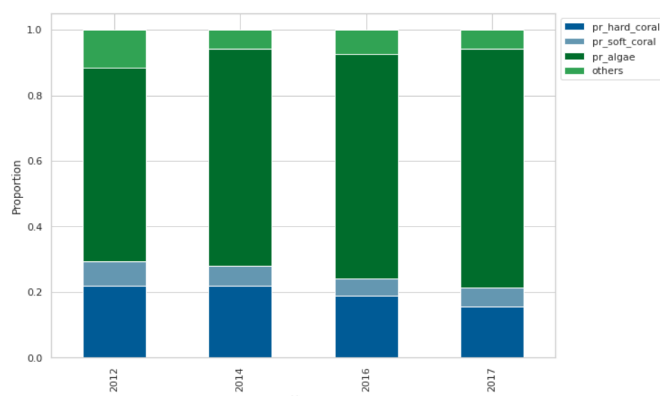

Figure 5: Proportion of hard, soft, algae and others (other + other invertebrates grouped) for all surveys

It is important to highlight the fact that not all countries are sampled in the same year and even when it the case; it may not be at the same location. We therefore plotted the same figures but for locations with were surveyed 4 times, this enables us to get a "real" insight into the change of proportions. This is shown in Figure (6) and (7). Please note that 4 surveys at the same location were only available for Australia and this for only 20 locations (80 surveys). This limits any conclusion for the area of study though it does seem to show an interesting trend. This trend is characterized by a decrease in hard and soft corals, increase in algae and other throughout the years.

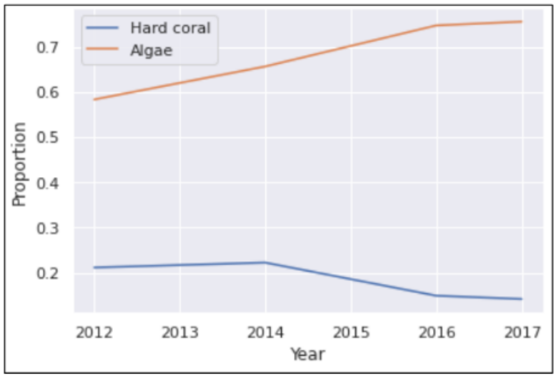

Figure 6: The proportion of algae and hard coral as a function of the years for locations with 4 surveys

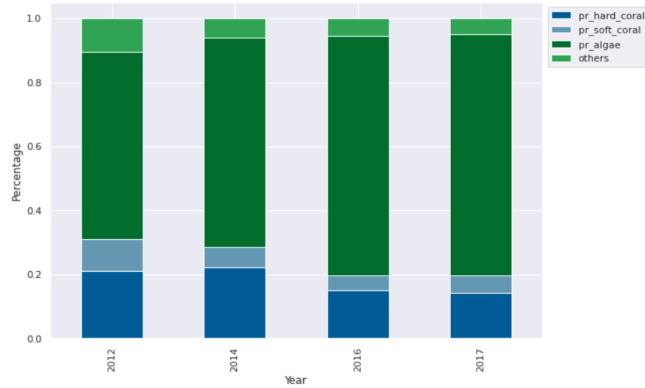

Figure 7: Proportion of hard, soft, algae and others (other + other invertebrates grouped) for locations with 4 surveys

We then decided to look into the change in corals between two years, this is only possible for Australia and Indonesia, as can be seen in Figure (3), they contain surveys for at least two years. Considering the fact that the reefs are big we cannot look at two surveys not sampled at the same location. This reduced the data and left 75 locations (150 surveys). In order to better visualize the change in hard corals, we plot it based off latitudes, as presented in Figure (8).

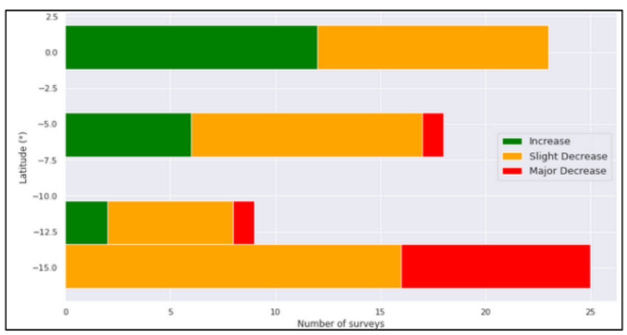

Figure 8: Change in hard corals for locations sampled twice based on latitudes

An increase (in green) is defined as any positive change between the latest year and first year of survey. A slight decrease is defined here as a decrease of 10% in hard corals and a Major decrease is any decrease above 10%.

Above 10 latitude, the surveys are exclusively located in Indonesia. The figure suggests that the Northern parts have less decrease in hard coral cover and even some increase. In order to see that the increase is not negligible, refer to Figure (19) in the Appendix.

Statistical tests should be included in further work in order to to confirm this visual tendency.

3.2 Correlations

Figure (9) depicts the linear correlations between the variables used in the modelling discussed in further sections. Our focus is on the correlations with the proportion of hard corals.

It seems that no significant correlation is detected for the target variable, though it is interesting to note a slight correlation with the distance to urban places, that is, the further we are from urban areas the more cover coral we might encounter. We also not a slight negative linear correlation with distance to marine protected areas (MAP). The further we are from an MAP (greater distance), the less hard corals we tend to find.

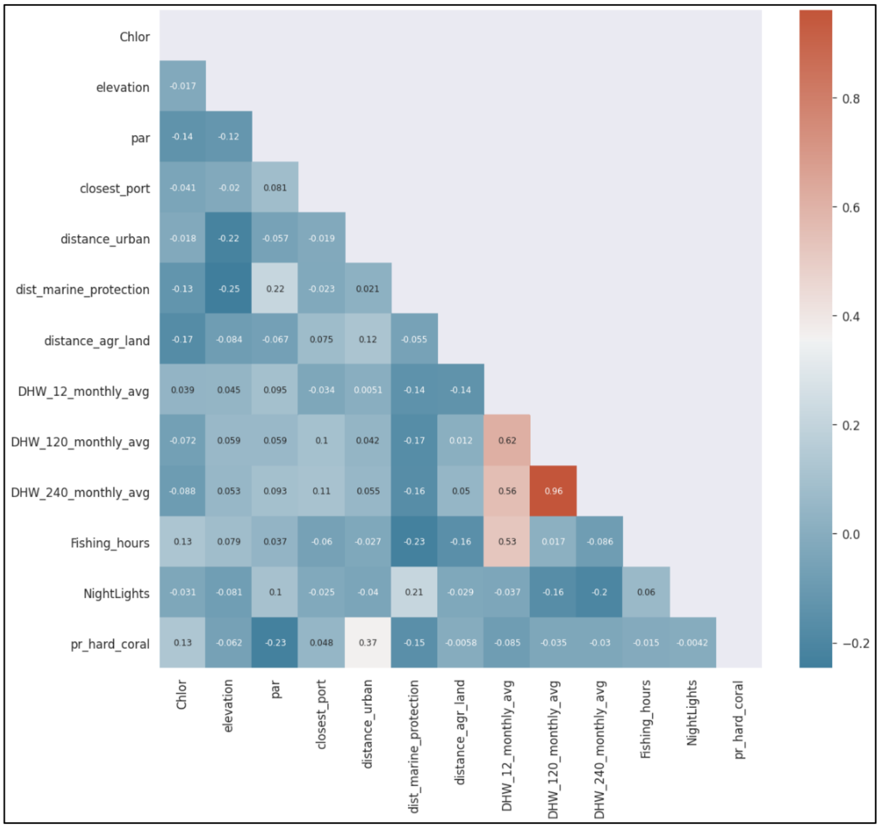

Figure 9: Linear correlation matrix for variables used in the basic modelling

4 Modelling Methods and results

So as to give a first insight into the data, further explanatory data analysis were undertaken as well as some basic predictive model. The idea of the latter is to see whether adding anthropological features would add any valuable information to the model with only DHW.

4.1 Replicating results method

With the intention to assess the encoding, we compare to a paper by O. Selmoni who looks are the effects of the DHW worldwide preceding 1, 10 and 20 years before the survey. The analysis is conducted using a uni-variate generalized linear model with beta regression; a distribution which models a probability with a domain bounded between 0 and 1. This method does not exist yet on `python`, we therefore use `Rpy2` so as to use the R software for the `GLMMTMB` package [23]. Using the spatial encoding discussed further, GLMM with beta regression will first try to "understand" the differences in coral cover by looking at spatiality and then explain the rest of the difference by looking at the variables given [24].

4.2 Addition of features

The next steps involve adding the anthropological data onto the model.

For the predictive models, we use two dataframes; one which is named environmental (env.) and the other anthropological (anth.). These are shown in Table (2). Because the target variable (*pr_hard_coral*) is continuous we must a regressive model; in this project, we use linear regression and decision trees.

Env. df	Anth. df
Spatial Encoding	Spatial Encoding
DHW 1 yr	DHW 1, 10, 20 yr(s)
DHW 10 yrs	Distance to: Marine Protected Areas Ports Urban centers Agricultural land
DHW 20 yrs	Fishing hours
	Night Lights

Table 2: The different dataframes (df) used

4.3 Spatial encoding

In order to take into account spatiality into the model, the inter-distances between survey points were computed. This was plotted with histograms (see appendix) the peaks were taken as thresholds for spatial distances (divided by two for a radius). The distance at the peaks are displayed in Table 3, a country encoding was also added. These thresholds are used in order to apply hierarchical clustering using the Ward distance method [25].

	Distance	Name
Peak 1	10	Local
Peak 2	500	Regional
Peak 3	1 500	State
Peak 4	2 200	-
Country	-	Country

Table 3: Distances for threshold + country

Peak 4 and peak 3 had similar clusters, it is thus kept at peak 3 (state). The country and the state encoding are shown in Figure (10) and (11).

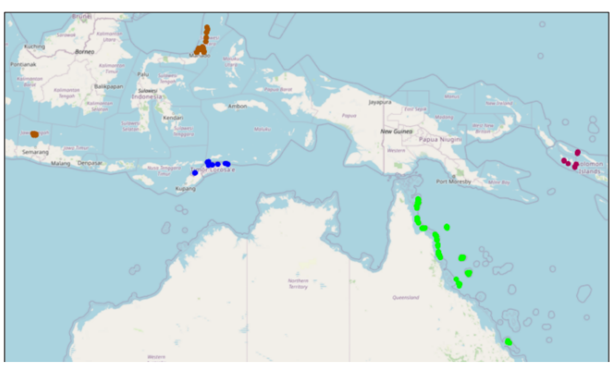

Figure 10: Country encoding

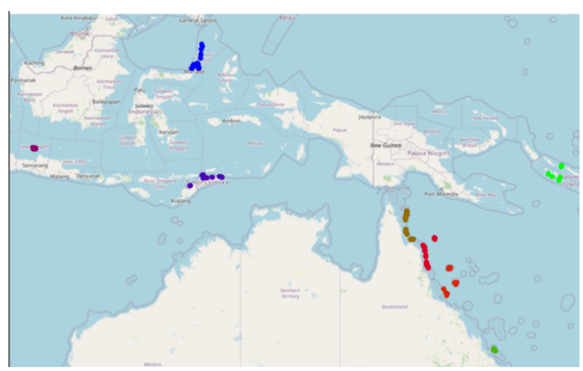

Figure 11: State encoding

In the beta regression model these values are categorically encoded and for the predictive model each is one-hot encoded.

Modelling Results

4.4 Replicating results

In the worldwide study conducted by O. Selmoni [in press], the team found a significant relation for DHW 1, 10, 20 years before the survey was taken in explaining the differences in coral cover. Using the same method but focusing on a smaller subset of countries we find for the DHW preceding one year before survey, Figure (12).

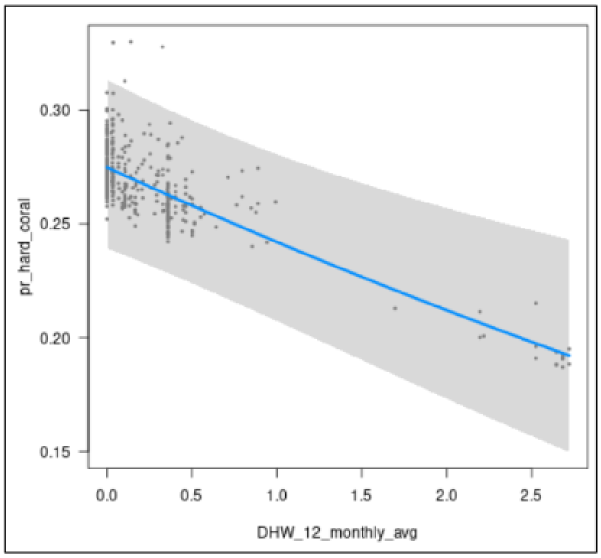

Figure 12: Coral cover as a function of DHW (1 year) with beta regression

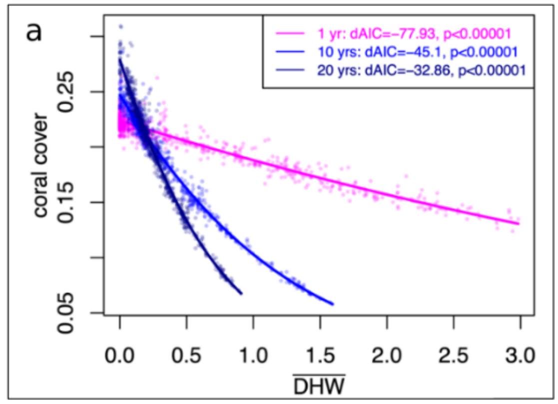

Figure 13: Selmoni results

Figure (12) highlights the impact of DHW indicator, the higher the value the less coral cover is present. This matches the paper of comparison. The same was done for DHW 10 and 20 years before the survey was taken but these do not yield significant results. Table 4 contains the results from the current model and the one used by O. Selmoni [in press]. The results are also shown in Figure (13).

The AIC stands for Akaike information criterion and is a measure which computes the relative quality of statistical models and thus provides a mean for model selection. The lower the AIC the better a fit to the data; we compare the AIC to the null model. The ΔAIC in Table 4 computes the null model AIC minus the current model; where a difference of -2 is significant.

	DHW 1 year		DHW 10 years		DHW 20 years	
ΔAIC	-11.8	-77.93	-1.6	-45.1	-1.9	-32.86
p_value	0.000348	<0.00001	0.532	<0.00001	0.761	<0.00001

Table 4: Comparing results with current model in blue and O. Selmoni in white

We note that both models conclude that DHW average over 1 year before the survey is the most important indicator of coral cover decay. In contrast with the paper to which we are comparing the results, DHW average over 10 or 20 years before the survey is deemed non significant and the null model nearly does as well. This difference to the paper may be caused by several factors, the first is that the number of surveys is multiplied by 4 over a much larger area than just the current area of study, another reason may be due to the great amount of bleaching which occurred in Australia in 2015-2016 [26]. This would fit within the average of a year but not 10 and 20, as the average would damp the significant decay due to the mass bleaching event.

4.5 Explanatory data analysis with GLMM

We now add several anthropological features and see how this affects the model.

The new model adds night-lights, fishing hours as well as the distance to closest port. This results in a model with a ΔAIC$= -27.3$ (compared to the null model) which is a decrease to the previous model indicating a better fit to the data.

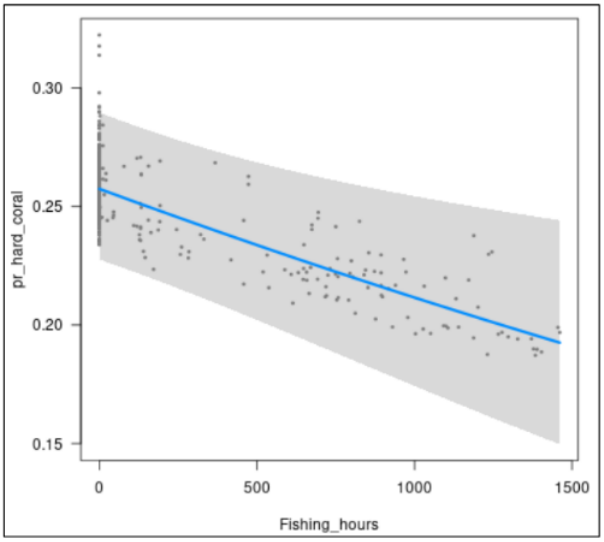

Figure 14: Coral cover as a function of number fishing hours

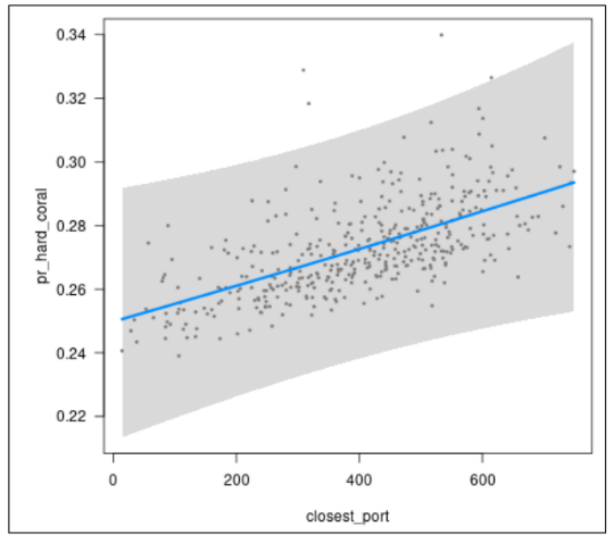

Figure 15: Coral cover as a function of distance to port

It is interesting to note in Figure (14) that as the number of fishing hours increase, it seems the less coral cover we find; though further tests and more efficient encoding would need to be conducted for definite conclusion. As one may see, many points are zero. On the other hand, Figure (15) tends to portray an increase in coral cover the further we travel from the ports. This indication seems plausible as the closer we are to the surface, the closer we are to human activity which is known to cause damage. Though additional research should be conducted in order to verify possible correlations with other variables. Night lights were deemed insignificant and DHW (1 year before survey) was less significant than in the previous univariate model as some anthropological factor seem to explain part of the distribution. Additional anthropological data did not add any explanatory power.

4.6 Predictive modelling

For the predictive modelling, we use linear regression as well as tree regression. The results are grouped in Table 5 with repeated k-fold cross validation (12 splits and repeated 5 times) in order to test the results.

	Local, Regional, state and country encoding		Only Country encoding		No spatial encoding	
Models	Env. df	Anth df	Env. df	Anth df	Env. df	Anth df
Linear Reg.	0.51 ±0.12	0.54 ±0.13	0.026 ±0.08	0.16 ±0.14	-0.03 ±0.05	0.15 ±0.14
Random For.	0.35 ±0.13	0.45 ±0.16	0.09 ±0.18	0.33 ±0.15	0.02 ±0.12	0.34 ±0.14

Table 5: R^2 results for predictive modelling (mean ± std)

The different types of encoding were meant to test how additional spatial information helps the models to predict the coral cover. The first column (pink) encodes all spatial data, the second (yellow) encodes only the four countries and the third (orange) does not encode any spatial information. The model was tested with 20% test set. We note the high standard deviation of the model results as well as a decreasing fit to the data as spatial encoding is removed. We also note an increase in fit when adding the anthropological data. Especially when encoding only the countries.

5 Limitations and further work

The monthly average does not capture high wind speeds as they usually last less than a day but are destructive for the reefs [27], it would be more insighful to look into the hourly data and filter out the number of hours at which various thresholds were exceeded.

The depths could be grouped by categories instead of by absolute value, this method would be inspired by Cinner [12] who found that depth provide refuges and are thus beneficial for the growth of corals.

Due to the size of the photosynthetically active radiation data it could not be encoded (using my computer) similarly to the DHW, SST etc... though this may not be the scope of this particular project, building a more powerful model for environmental factors is also of great interest.

Encoding the night lights as described in the *data collection section* leads inconsistencies such as two surveys within 15 km having the same value. A better way to encode would be to download the data with a finer resolution based off the distribution in space and make a radius around the survey and compute the mean in that area. What is important is to have surveys close to the land pick up the values from the human activity operating on land.

The encoding of the marine protected areas by approximating the polygons to their centroid is limiting. It would be interesting to also add a boolean variable on whether the survey points are within a marine protected area, but also keep the distance for points which are not within the polygons, but take the points as a distance with respect to the closest polygon's edge instead of the centroid, which in this case was deemed an acceptable approximation due to the sizes of the polygon.

For the fishing data, inversely proportional weights for several threshold for the distances should be considered. The number of fishing hours would then be scaled to the factor. Another improvement to this would be to consider differently, the various forms of fishing as some are known to cause more harm to the reefs such as *blast fishing* or *trawling* [28], [29].

Some basic modelling was tested and further time should be invested in order to better encode but also find more models which could be used for this particular set up. For the predictive model, it would be interesting to test the models with instead of the spatial encoding presented above, simply using latitude and longitude.

Evidently, having more surveys but also having more surveys taken at the same location but different years would help build better models.

6 Conclusion

Following a thorough literature review, data was collected, cleaned and examined through the data analysis section presented in section (3). The data seemed to show a tendency of the algae proportion to increase as the hard corals decrease. By comparing the change in hard corals through two separate years for Indonesia and Australia, it suggests a greater decrease of hard corals in the South and even a slight increase of hard corals in the North. The data after collection, cleaning and merging was used in order to set up a baseline for modelling.

Firstly in order to faithfully reproduce known results, secondly to add anthropological data to the explanatory data analysis and lastly to test predictive modelling excluding/including anthropological data. Adding anthropological data sounds promising in both the explanatory and predictive models.

Future research should work on the data with the modifications presented in Section (5), as well as dig further into the modelling and try additional multi-variate models as well as testing these in other parts of the world and not only the given area in the Pacific. Once a faithful predictive model with less variance is established, comparing the important features with the explanatory model should also be included in the next steps.

Appendix

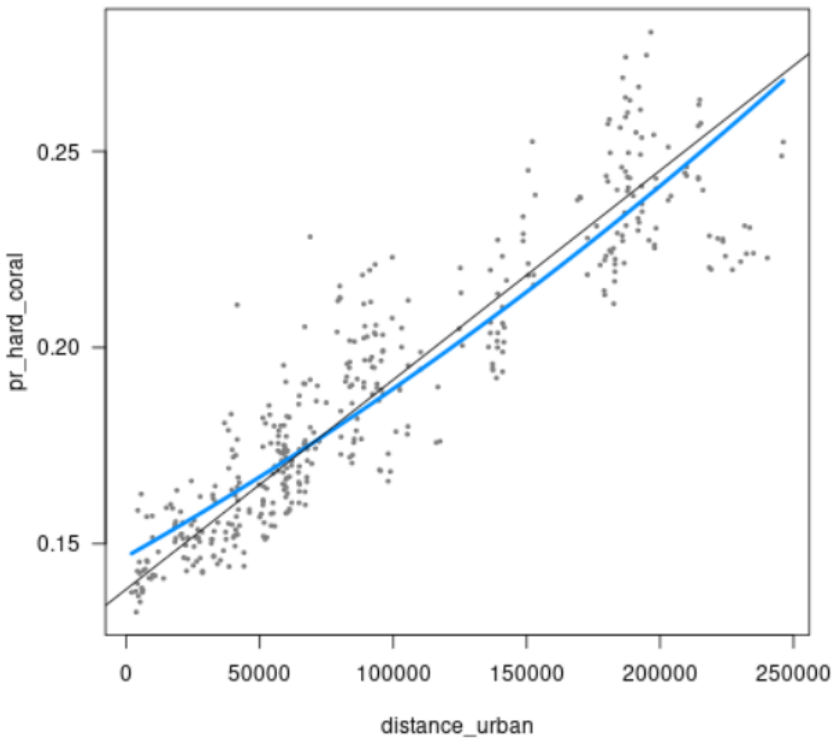

Figure 16: Distance to Urban centers

Figure (16) depicts beta regression in blue and linear regression in black. Linear regression on this uni-variate model yielded an $R^2 = 0.1314$. Unfortunately due to an inconsistency in the Hessian it could not be combined with other variables (reason why it does not appear on the report). Though the tendency shown in the figure and the dispersion of points seem to indicate that as we go further away from urban centers, the more hard corals we have.

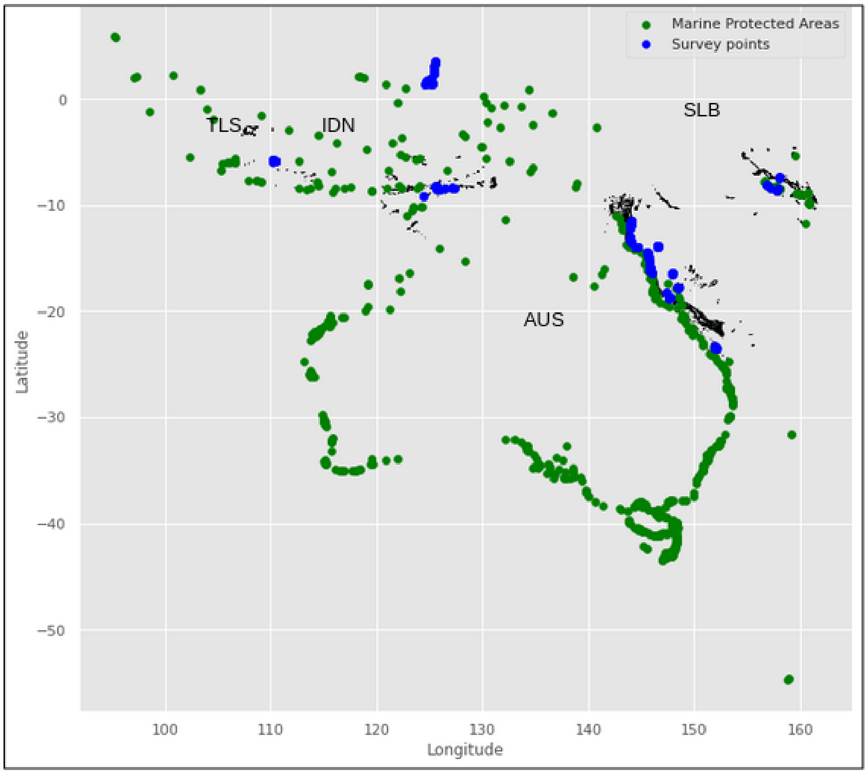

Figure 17: Marine Protected areas and survey points

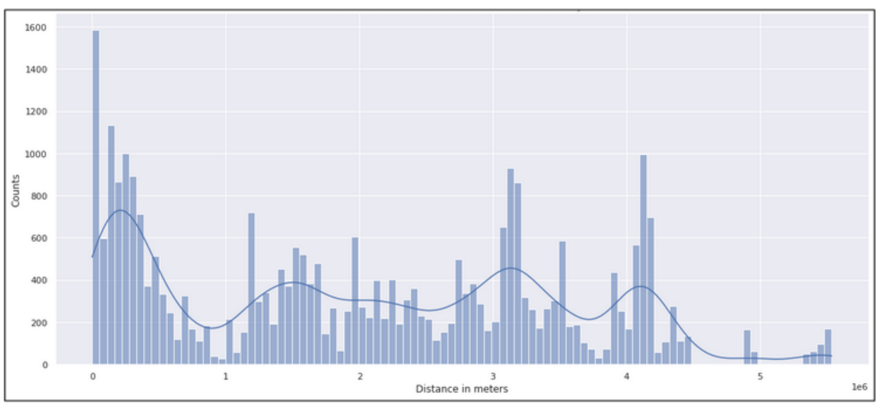

Figure 18: Inter-survey distribution used for picking the peaks (spatial encoding)

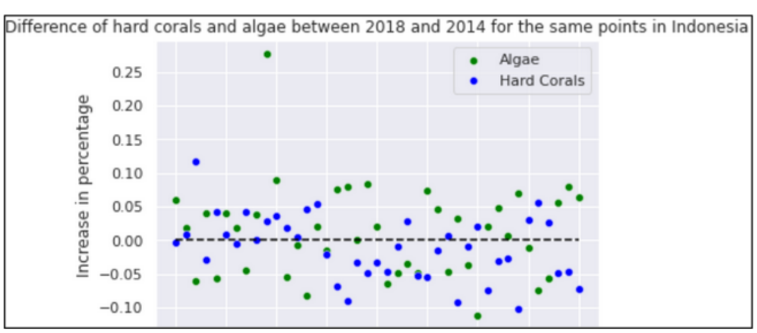

Figure 19: Difference in hard corals and algae

Variable	Units	Format	Projection	Temporal resolution	Temporal coverage	Spatial resolution [deg or m]	Spatial Coverage	Data source
Corals								
Biomass per taxonomy	Proportion	CSV	EPSG:4326	Not constant	2012-2018	0.5-2m	Global	Caitlin survey
Coral Coverage	Proportion	CSV	EPSG:4326	Nor constant	2012-2018	0.5-2m	Global	Caitlin survey
Environmental drivers								
SST (Sea surface temperature) [30]	Degrees	CSV	EPSG:4326	Monthly	35 past years	5000 m	Global	NOAA
DHW (Degree heating week) [30]	°C-weeks	CSV	EPSG:4326	Monthly average	35 past years	5000 m	Global	NOAA
Wind speed [31]	m/s	Grid GRIB	EPSG:4326	Monthly	From Jan 1981	0.5° x 0.5°	Global	Copernicus
Chlorophyll-a [32]	mg/m^3	Grid NetCDF	EPSG:4326	Daily	1997 to present	4000	Global	Copernicus
Solar radiation [31]	J/m^2	Grid GRIB	EPSG:4326	Monthly	From Jan 1981	0.5° x 0.5°	Global	Copernicus
Photosynthetically Active Radiation PAR [33]	µEinstein.m-2. s-1	NETCDF	EPSG:4326	Monthly	Since 2000	40000	Global	NASA
Ocean Depth [34]	Meters	NETCDF	EPSG:4326	Yearly	-	15 arc seconds	Global	GEBCO
Anthropogenic drivers								
Nightlights [35]	Average radiance	TIFF	WGS 84	Monthly	2012-2020	15 arc seconds (sampled at 15km)	Global	EOG
Land Cover [36]	Classification and cover fraction	TIFF	WGS 84	Yearly	2015-2019	100m at equator (sampled at 5km)	Global	Copernicus
Protected Areas [19]	-	Shapefile	EPSG:4326	Monthly	1981-2021	-	Global	Copernicus
Fishing Hours [37]	Hours	CSV	EPSG:4326	Daily	From 2012	100th degree	Global	Global Fishing Watch
Urban mapping [38]	-	Shapefile	EPSG:4326	Once	Since 1995	-	Global	NASA
Distance from port [39]	Meters	NETCDF	EPSG:4326	Once	2020	1000	Global	Global Fishing Watch

Table 6: Data collected and sources

References

[1] Mebrahtu Ateweberhan, David A Feary, Shashank Keshavmurthy, Allen Chen, Michael H Schleyer, and Charles RC Sheppard. Climate change impacts on coral reefs: synergies with local effects, possibilities for acclimation, and management implications. *Marine pollution bulletin*, 74(2):526–539, 2013.

[2] Camilo Mora. A clear human footprint in the coral reefs of the caribbean. *Proceedings of the Royal Society B: Biological Sciences*, 275(1636):767–773, 2008.

[3] Lauretta Burke, Kathleen Reytar, Mark Spalding, and Allison Perry. *Reefs at risk revisited*. World Resources Institute, 2011.

[4] Leon C Braat and Rudolf De Groot. The ecosystem services agenda: bridging the worlds of natural science and economics, conservation and development, and public and private policy. *Ecosystem services*, 1(1):4–15, 2012.

[5] Ove Hoegh-Guldberg. Coral reef ecosystems and anthropogenic climate change. *Regional Environmental Change*, 11(1):215–227, 2011.

[6] Nancy Knowlton and Jeremy BC Jackson. Shifting baselines, local impacts, and global change on coral reefs. *PLoS Biol*, 6(2):e54, 2008.

[7] Shahid Naeem, J Emmett Duffy, and Erika Zavaleta. The functions of biological diversity in an age of extinction. *science*, 336(6087):1401–1406, 2012.

[8] Barbara E Brown. Adaptations of reef corals to physical environmental stress. *Advances in marine biology*, 31:221–299, 1997.

[9] Ove Hoegh-Guldberg. Climate change, coral bleaching and the future of the world's coral reefs. *Marine and freshwater research*, 50(8):839–866, 1999.

[10] Peter J Mumby, John RM Chisholm, Alasdair J Edwards, Serge Andrefouet, and Jean Jaubert. Cloudy weather may have saved society island reef corals during the 1998 enso event. *Marine Ecology Progress Series*, 222:209–216, 2001.

[11] Joshua E Cinner, Nicholas AJ Graham, Cindy Huchery, and M Aaron MacNeil. Global effects of local human population density and distance to markets on the condition of coral reef fisheries. *Conservation Biology*, 27(3):453–458, 2013.

[12] Joshua E Cinner, Cindy Huchery, M Aaron MacNeil, Nicholas AJ Graham, Tim R McClanahan, Joseph Maina, Eva Maire, John N Kittinger, Christina C Hicks, Camilo Mora, et al. Bright spots among the world's coral reefs. *Nature*, 535(7612):416–419, 2016.

[13] Manuel González-Rivero, Alberto Rodriguez-Ramirez, Oscar Beijbom, Peter Dalton, Emma V. Kennedy, Benjamin P. Neal, Julie Vercelloni, Pim Bongaerts, Anjani Ganase, Dominic E.P. Bryant, Kristen Brown, Catherine Kim, Sebastian Lopez-Marcano, Sophie Dove, Christophe Bailhache, Hawthorne L. Beyer, and Ove Hoegh-Guldberg. Seaview survey photo-quadrat and image classification dataset, 2019.

[14] Joseph Maina, Valentijn Venus, Timothy R McClanahan, and Mebrahtu Ateweberhan. Modelling susceptibility of coral reefs to environmental stress using remote sensing data and gis models. *Ecological modelling*, 212(3-4):180–199, 2008.

[15] Glenn De'ath, Katharina E Fabricius, Hugh Sweatman, and Marji Puotinen. The 27–year decline of coral cover on the great barrier reef and its causes. *Proceedings of the National Academy of Sciences*, 109(44):17995–17999, 2012.

[16] John D Hedley, Chris M Roelfsema, Iliana Chollett, Alastair R Harborne, Scott F Heron, Scarla Weeks, William J Skirving, Alan E Strong, C Mark Eakin, Tyler RL Christensen, et al. Remote sensing of coral reefs for monitoring and management: a review. *Remote Sensing*, 8(2):118, 2016.

[17] Peter J Mumby, William Skirving, Alan E Strong, John T Hardy, Ellsworth F LeDrew, Eric J Hochberg, Rick P Stumpf, and Laura T David. Remote sensing of coral reefs and their physical environment. *Marine pollution bulletin*, 48(3-4):219–228, 2004.

[18] William Skirving, Benjamin Marsh, Jacqueline De La Cour, Gang Liu, Andy Harris, Eileen Maturi, Erick Geiger, and C. Mark Eakin. CoralTemp and the coral reef watch coral bleaching heat stress product suite version 3.1, data accessed on feb. 2021. *Remote Sensing*, 12(23):3856, November 2020.

[19] Website. Marine Protected Areas. https://www.protectedplanet.net/en/thematic-areas/marine-protected-areas, 2021. [Online; accessed 08-March-2021].

[20] Sean Gillies et al. Rasterio: geospatial raster i/o for Python programmers, 2013–.

[21] Callum M Roberts. Effects of fishing on the ecosystem structure of coral reefs. *Conservation biology*, 9(5):988–995, 1995.

[22] Lauretta Burke, Katie Reytar, Mark Spalding, and Allison Perry. *Reefs at risk revisited in the Coral Triangle*. 2012.

[23] Mollie E Brooks, Kasper Kristensen, Koen J Van Benthem, Arni Magnusson, Casper W Berg, Anders Nielsen, Hans J Skaug, Martin Machler, and Benjamin M Bolker. glmmtmb balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, 9(2):378–400, 2017.

[24] Silvia Ferrari and Francisco Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7):799–815, 2004.

[25] Fionn Murtagh and Pierre Legendre. Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. *arXiv preprint arXiv:1111.6285*, 2011.

[26] TP Hughes, JT Kerry, and T Simpson. Large-scale bleaching of corals on the great barrier reef, 2018.

[27] Toby A Gardner, Isabelle M Côté, Jennifer A Gill, Alastair Grant, and Andrew R Watkinson. Hurricanes and caribbean coral reefs: impacts, recovery patterns, and role in long-term decline. *Ecology*, 86(1):174–184, 2005.

[28] John W McManus, Rodolfo B Reyes Jr, and Cleto L Nanola Jr. Effects of some destructive fishing methods on coral cover and potential rates of recovery. *Environmental management*, 21(1):69–78, 1997.

[29] Malcolm R Clark and Derek P Tittensor. An index to assess the risk to stony corals from bottom trawling on seamounts. *Marine Ecology*, 31:200–211, 2010.

[30] Noaa coral reef watch. 2019, updated daily. noaa coral reef watch version 3.1 daily 5km satellite regional virtual station time series data for southeast florida, mar. 12, 2013-mar. 11, 2014. college park, maryland, usa: Noaa coral reef watch. https://coralreefwatch.noaa.gov/product/vs/data.php. Accessed: 2021-04-03.

[31] Hersbach, h., bell, b., berrisford, p., biavati, g., horányi, a., muñoz sabater, j., nicolas, j., peubey, c., radu, r., rozum, i., schepers, d., simmons, a., soci, c., dee, d., thépaut, j-n. (2019): Era5 monthly averaged data on single levels from 1979 to present. copernicus climate change service (c3s) climate data store (cds).

[32] Thomas jackson, ben calton. (2020): Ocean colour daily data from 1997 to present derived from satellite observations.

[33] Modis aqua, monthly photosynthetically active radiation; global measures data, nasa ob.daac.

[34] Gebco compilation group (2020) gebco 2020 grid (doi:10.5285/a29c5465-b138-234d-e053-6c86abc040b9).

[35] Viirs nighttime day/night band composites version 1.

[36] Copernicus global land cover layers: Cgls-lc100 collection 3.

[37] Global datasets of ais-based fishing effort and vessel presence v2.0.

[38] The global rural-urban mapping project, version 1 (grumpv1): Urban extent polygons, revision 02 contains 74,445 urban extent polygons (in esri shapefile format) with the settlement names (or names of the largest cities in the case of multi-city agglomerations).

[39] Distance from port in meters (2020).