

Determining the Relationship Between Coral Uranium Concentrations and Seawater Environmental Conditions Using a Compilation of Uranium Isotope Data

Ella Hagopian, Katie Murney, Brenden Ireland

Advisors: Anne Gothmann, Matt Richey

Abstract

An understanding of past climate change is critical for understanding how future climate will evolve. All of Earth's surface environments (ocean, atmosphere, and land) play a role in the climate system and it is possible to use reconstructions of past oceanic environments to infer past changes in the global climate system. A primary approach toward reconstructing past environmental change is to use paleoenvironmental "proxies" as indirect archives. Recent experimental studies have suggested that the uranium concentration of fossil coral skeletons -- if well-preserved -- may be used as a proxy in order to glean information about environmental variables (e.g. seawater temperature, pH, and salinity) dating back thousands of years. In this study, we aimed to further test the robustness of the relationship between uranium concentration in fossil corals and indicators of the oceanic environment, primarily pH, in order to determine whether uranium concentrations in coral could be used as a proxy for the global climate in the past.

A compilation of previous measurements of uranium concentrations in 1,907 fossil coral skeleton samples were utilized to explore the above mentioned relationships. Secondary datasets with temperature, salinity, and pH were also joined with the primary coral sample dataset. We examined the relationship between pH and uranium concentrations in fossil coral using multivariate linear regression. Our analysis demonstrated that though there is a relationship between the pH of the ocean and uranium concentrations in fossil coral, there are other factors that affect this relationship. Factors such as genus should be considered before using fossil coral as a proxy for seawater pH, salinity, and temperature when reconstructing the past.

Introduction

The oceans play an important role in the climate system and respond to changes in global temperatures and atmospheric carbon dioxide concentrations (Sarmiento and Gruber, 2006). As demonstrated in Figure 1, there has been an observed inverse relationship between atmospheric CO₂ and seawater pH in the past 30 years (Keeling et al., 2004). Insight into how oceans respond to past climate change is critical for understanding how future ocean conditions evolve. A primary approach

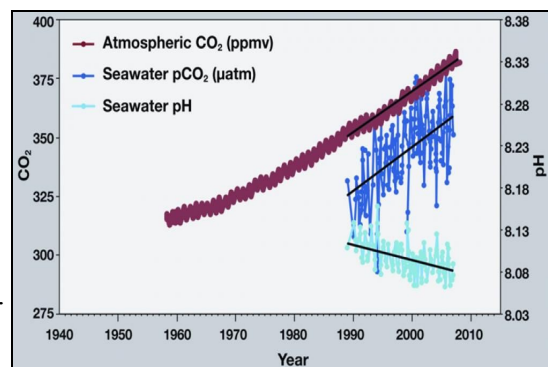


Figure 1. Relationship between atmospheric CO₂, seawater CO₂, and seawater pH from modern measurements

toward reconstructing past environmental change is to use paleoenvironmental “proxies” as indirect archives. For example, Saenger and Evans (2018) used the Magnesium/Calcium ratio in planktic foraminifera as a proxy for measuring oceanic temperature. Though they found other secondary effects such as mean size fraction that altered this relationship in specific species, the study validated an overall significant relationship between temperature and Mg/Ca ratios in planktic foraminifera.

Similarly, Robinson et al. (2014) examined methods for using properties of fossil coral as a proxy for paleoenvironmental change. Corals are animals in the cnidarian class that exhibit a wide range of morphologies and life habitats. Fossil corals have been identified as promising archives of past oceanic environments because of their growth distribution, longevity and ability to be accurately dated using radiometric techniques (Robinson et al., 2014).

Recent experimental studies suggest that the uranium concentration of fossil coral skeletons -- if well preserved -- may be used to glean environmental pH, temperature, and salinity information. Shen and Dunbar (1995) discovered a possible temperature dependence with uranium concentrations in two species of coral found in the Galapagos. Further, Inoue et al. (2011) examined the relationship between uranium concentration in *Acropora digitifera* and oceanic pH and found a strong inverse relationship between the two as demonstrated in Figure 2.

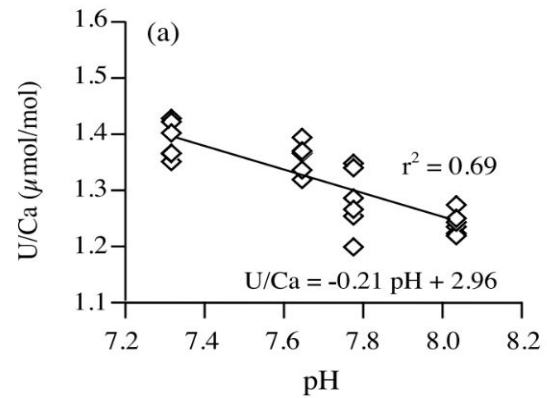


Figure 2. Experimental results comparing U/Ca concentrations in laboratory-cultured corals to pH (Inoue et al, 2011)

The current study aimed to quantify the significance of the relationship between uranium concentrations in fossil coral and oceanic environmental indicators like pH, salinity, and temperature in order to determine if coral sample data can be used as a proxy for the past oceanic environment, primarily pH. The coral samples came from past datasets compiled by Chutcharavan et al. (2018) by adding 2,184 previous U-series measurements to compilations from Hibbert et al. (2016) and Chen et al. (2016). Fig. 3 shows the distribution of collection locations of fossil coral samples for the three most abundant genera in our coral dataset: *Favia*, *Porites*, and *Acropora*. The shapes differentiate the different genera and the ocean is colored by pH level, showing that pH levels are higher closer to the equator.

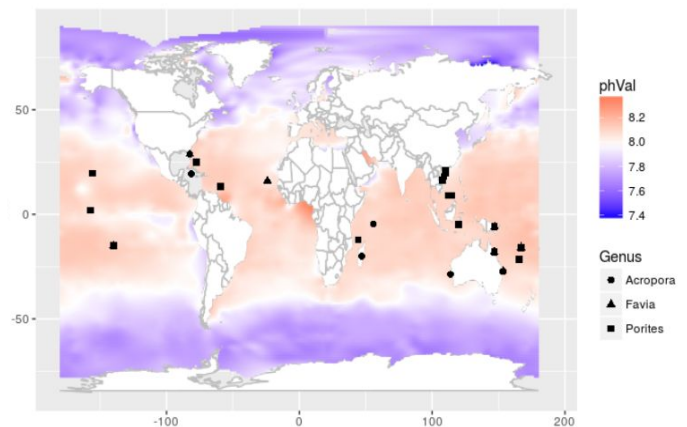


Figure 3. World map containing sites of coral samples with color variation according to average pH and shape variation according to the top three genera represented in the data.

Methods

The compiled coral dataset from Chutcharavan et al. (2018) describes fossil coral samples and their attributes, including variables such as ocean basin and location from which the coral samples were collected, genus, species, percent calcite in the coral skeleton, age of the coral, and uranium concentration of the coral skeleton. Other variables that were not used in our analysis but were still present in the original dataset included the U238 standard deviation value, the isotope measurement delta U234, and place and site variables that gave more specific information about the locations of the samples. The ages of the coral samples range from modern (0) to 57,943 years, with a median of 6,156 years. The original dataset of coral samples had 28 genera, but the final dataset narrowed the genus variable to the four most abundant genera to make modeling more meaningful. Additionally, the percent calcite of the fossil corals ranged from 0% to 23%. We considered uranium concentrations labeled as “not reported” and genera reported as “nd”, “unknown”, or “unidentified” as NA -- not available. All of these observations were missing information and were not useful to our analysis so we grouped them together as NA and excluded them from our final dataset.

To supplement the compiled coral data, one-degree gridded World Ocean Atlas (WOA) datasets were used. The WOA datasets contain information on the temperature and salinity of the ocean (Locarini et al., 2018; Zweng et al., 2018). Additionally, a dataset of oceanic pH was obtained from the Global Ocean Data Analysis Project (GLODAP) (Lauvset et al., 2016). We used these three datasets because there is not enough environmental information for the coral corresponding to the time of collection; instead, we are using data representing average environmental conditions for the present. A distance function was developed in R in order to measure the distance from each coral sample to each GLODAP and WOA factor, taking into account the earth's curvature. Using that function, the three closest temperature, pH, and salinity values were determined for each fossil coral observation in the original coral dataset. These three closest estimates were averaged to yield final values for temperature, salinity, and pH for all coral samples. The range of pH values across all of the coral samples was 7.80 to 8.00, which we considered relatively small considering the modern ocean range of pH is around 7.4 to 8.2 (Figure 3). The range of salinity values was 33.34 to 36.53, and the range of temperature values was 18.55 to 29.02 degrees Celsius.

Our first step in the data cleaning process was selecting out the relevant variables for the study; we excluded variables such as site, ocean basin, and location because they were not significant when we put them in multivariate models. When we encountered NAs or “not reported” in the Genus variable, we excluded them from the dataset because we couldn't use them to represent Genus. When we found NAs in the uranium concentration variable (U238), we grouped them together and excluded them from the dataset. Another challenge with our dataset was that many observations were misspelled or labeled differently. To address this problem, we manually renamed the misspelled genera. Additionally, we used R's geocoding to identify

latitude and longitude for the coral observations because they were not provided. In some cases, we manually searched for longitudes and latitudes.

Because the data was compiled from several sources, we had to make assumptions in order to clean it. One assumption we made was that using the three closest temperature, salinity, and pH readings would be appropriate to estimate a single reading for each fossil coral sample. We also decided that for any percent calcite reported as a range of values, we would select the maximum of that range as the single value for that observation. We made this assumption because we wanted to account for the highest possible amount of calcite in each fossil coral sample and not underestimate any of them.

After the data was curated and cleaned, we analyzed our data using multivariate linear regression models. Uranium concentration in the coral samples was the response variable and oceanic pH was the main explanatory variable. We explored other explanatory variables including age, genus, and percent calcite. Additionally, interaction terms between genus and environmental factors temperature, pH, and salinity were created to explore how uranium concentration relationships differed by genus. We created a final multivariate linear model using backwards selection (Table 1). Once we had identified significant predictors based on a 0.05 significance level, we added interaction terms with the four most abundant genera: *Acropora*, *Favia*, *Montastraea*, and *Porites*.

Results

The multivariate regression model predicted the uranium concentration of fossil coral. The main explanatory variables that were significant predictors were pH, salinity, age, and interactions between temperature, salinity, pH, and genus (Table 1). In addition, the interaction term for pH and *Porites* is positive and statistically significant ($t = 4.78$, $t < .001$), which is notable because this relationship is the opposite of the other three most abundant genera, *Acropora*, *Favia*, and *Montastraea*, indicating that further investigation is required.

Our results suggest that there is a relationship between uranium concentration and pH of seawater ($t = -2.616$, $p < 0.01$). Plots comparing pH of the oceanic environment and uranium concentrations of coral samples were compared. Each plot was grouped by the four most abundant genera, *Acropora*, *Favia*, *Montastraea*, and *Porites*. Each observation was also colored by

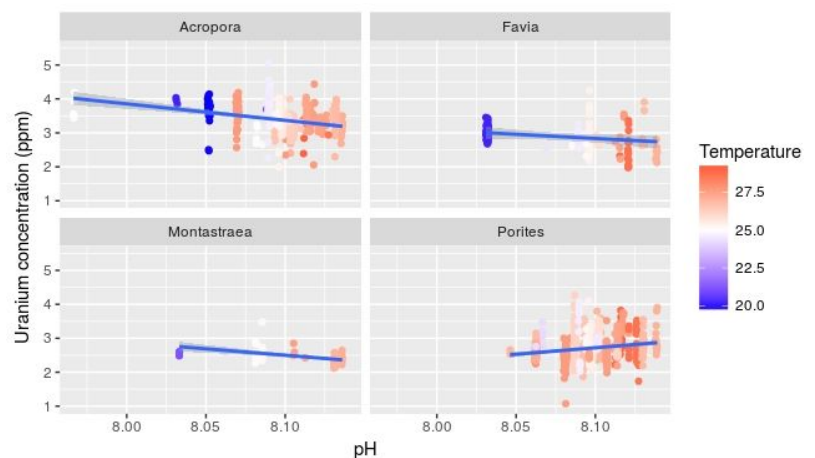


Figure 4. Scatterplots demonstrating observed relationship between uranium concentration and pH, colored by temperature.

temperature (Figure 4). This plot demonstrates that there is an inverse relationship between pH and uranium concentration in coral for *Acropora*, *Favia*, and *Montastraea* genera, indicating that increases in pH are associated with decreased levels of uranium concentration. However, there is a direct relationship between pH and uranium concentration for the *Porites* genus, showing that higher pH levels are associated with higher uranium concentrations. Thus, we cannot generalize an inverse relationship between pH and uranium concentration for all coral samples. Our results suggest that genus plays a role in the interaction between pH of the ocean and uranium concentrations of the fossil corals.

Similarly, there is a strong relationship between uranium concentration and salinity of the ocean ($t = -4.630$, $p < 0.001$). Salinity was plotted against uranium concentrations of coral samples where each observation was

colored by temperature and the plot was separated by the top four most abundant genera (Figure 5). In general, there was a positive relationship between salinity and uranium concentrations. However, when examined by genus, *Acropora* had an overall negative trend unlike *Favia*, *Montastraea*, or *Porites* (Table 1). There was no obvious pattern with temperature that could impact this relationship. Based on these findings, we cannot generalize a positive relationship between salinity and uranium concentration and must consider genus when examining this relationship.

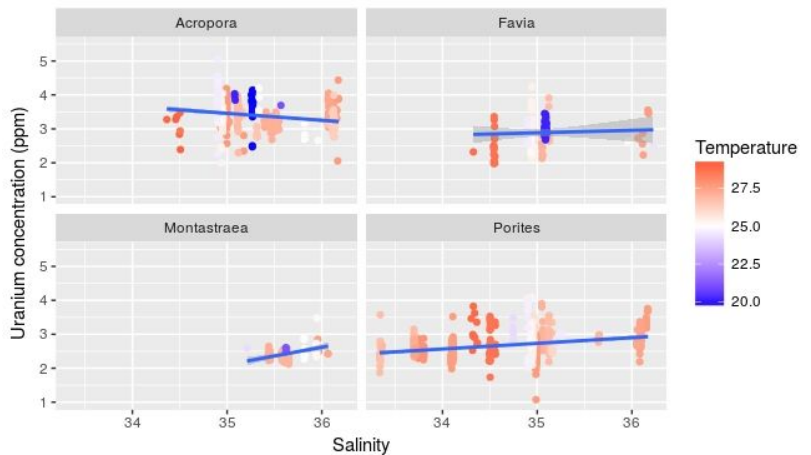


Figure 5. Scatterplots demonstrating observed relationship between uranium concentration and salinity, colored by temperature.

In a previous study, Inoue et al. (2011) found a negative relationship between pH and uranium concentration in coral. Their experiment measured uranium concentrations at pH levels of 7.8 and 8.0. The experimental data from Inoue et al. (2011) yielded uranium concentration values of 3.15 at a pH of 7.8 and 3.05 at a pH of 8.0. In order to explore the predicting capabilities of our model, the multiple regression model was used to predict uranium concentrations for two different pH levels: 7.8 and 8.0. These calculations were done holding temperature and salinity constant at 27°C and 33 parts per thousand (ppt), respectively, as performed in the Inoue et al. (2011) experiment. The model predictions yielded a uranium concentration value of 4.58 at a pH of 7.8 and 3.99 at a pH of 8.0. Thus, the model predicted uranium values that were slightly offset at higher levels than that found in the Inoue et al. (2011) experimental data. The predicted slope from the model was also slightly steeper than that of Inoue et al. (2011). Our model predictions project values that are outside of the calibration scope of Inoue et al. (2011) experiment. However, the model was able to predict an inverse relationship

between uranium concentration and pH of similar magnitude to that found by Inoue et al. (2011) (Table 2).

Discussion

This study was focused on evaluating the significance of the relationship between oceanic environmental indicators and fossil coral uranium concentrations in order to determine if fossil coral data can provide a clear indicator of the past oceanic environment. Based on our findings both in the models and the data visuals, it is clear that although there is a relationship between oceanic pH and uranium concentration of fossil coral, other factors play a role in that relationship. For example, the relationship between uranium concentration and ocean pH had a negative slope for *Acropora*, *Montastraea*, and *Favia*, but a positive slope for *Porites*. Additionally, the negative relationship seemed to be influenced by the relatively cold oceanic temperatures for *Acropora* and *Favia* samples and the absence of the colder oceanic temperatures for *Montastraea* and *Porites*. The roles of genus and temperature in the relationship between pH and uranium concentration suggests that there are likely additional factors that are influencing that relationship. Thus, using uranium concentration of fossil corals as a proxy for pH of the ocean is not quite enough to accurately represent the pH levels of the oceanic environment. Just as other proxies such as the Mg/Ca temperature proxy in foraminifera are influenced by secondary factors, there are other variables that affect the relationship between fossil coral uranium concentrations and oceanic pH levels which cannot be ignored (Saenger and Evans 2018). Taking into account the importance of secondary effects, information about fossil coral uranium concentrations is not enough on its own to make meaningful predictions of past oceanic environments.

Future Directions

One future direction that this research could explore is implementing methods used in Saenger and Evans (2018) experiments. Specifically, using an additive method to select the most appropriate regression models could be used in order to select variables that are significant predictors of uranium concentration. To assess the performance of the model, the predicted values could be compared to the observed uranium concentration values. In addition, the model could be validated using root mean square error (RMSE) and confidence intervals (CI). Additionally, the model could be used on 10,000 samples by replacement in order to assess the usefulness of the model (Saenger & Evans, 2018). We would like to use these methods to learn more about why these dependencies may have appeared in our results.

Future research should also explore other variables that may change the relationship between uranium concentration of fossil corals and the pH of the ocean. For example, the current study focused on fossil coral samples from the surface of the ocean, which includes all coral samples at a maximum of 5 meters of depth. Using deeper fossil coral samples would allow us to

examine depth as a variable because as depth increases, temperature and pH decrease. Lower depths would allow us to examine a greater range of environmental factors.

Our current dataset used a distance function to calculate the closest temperature, salinity, and pH measurements from the WOA and GLODAP datasets. Future research could find fossil coral samples with measurements of exact temperature, salinity, and pH for each sample, which would eliminate the need to compute estimated values for each sample. Finally, each individual dataset should be examined separately. Exploring each dataset separately would allow us to gain a greater understanding of the original fossil coral samples and the oceanic environmental data, which may lead to insight into the reasoning behind the relationships between uranium concentration in fossil coral and the oceanic temperature, salinity, and pH.

References

- Chutcharavan, P.M., A. Dutton, and M. J. Ellwood, 2017. Seawater $^{234}\text{U} / ^{238}\text{U}$ recorded by modern and fossil corals. *Geochimica et Cosmochimica Acta*, v. 224, 1-17.
- Inoue, M., R. Suwa, A. Suzuki, K. Sakai, and H. Kawahata, 2011. Effects of seawater pH on growth and skeletal U/Ca ratios of *Acropora digitifera* coral polyps. *Geophysical Research Letters*, v. 38, L12809.
- Keeling, C. D., Brix, H., and Gruber, N., 2004. Seasonal and long-term dynamics of the upper ocean carbon cycle at Station ALOHA near Hawaii, *Global Biogeochemical Cycles*, v. 18, GB4006.
- Lauvset, S. K., R. M. Key, A. Olsen, S. van Heuven, A. Velo, X. Lin, C. Schirnick, A. Kozyr, T. Tanhua, M. Hoppema, S. Jutterström, R. Steinfeldt, E. Jeansson, M. Ishii, F. F. Pérez, T. Suzuki and S. Watelet, 2016. A new global interior ocean mapped climatology: the 1°x1° GLODAP version 2, *Earth Syst. Sci. Data*, 8, 325–340, 2016, [doi:10.5194/essd-8-325-2016](https://doi.org/10.5194/essd-8-325-2016).
- Locarnini, R. A., A. V. Mishonov, O. K. Baranova, T. P. Boyer, M. M. Zweng, H. E. Garcia, J. R. Reagan, D. Seidov, K. Weathers, C. R. Paver, and I. Smolyar, 2018. World Ocean Atlas 2018, Volume 1: Temperature. A. Mishonov Technical Ed.; in preparation.
- NOAA PMEL Carbon Program. “Ocean Acidification Graph.” *Smithsonian Institutions* <https://ocean.si.edu/conservation/acidification/ocean-acidification-graph>. Accessed 30 April 2019.
- Robinson, L.F., J.F. Adkins, N. Frank, A.C. Gagnon, N.G. Prouty, E.B. Roark, and T. van de Flierdt, 2014. The geochemistry of deep-sea coral skeletons: A review of vital effects and applications for palaeoceanography. *Deep-Sea Research II*, v. 99, 184-198.
- Sarmiento, J.L., and N. Gruber, 2006. *Ocean Biogeochemical Dynamics*. Princeton University Press.
- Shen, G.T., and R.B. Dunbar, 1995. Environmental controls on uranium in reef corals. *Geochimica et Cosmochimica Acta*, v. 59, 2009-2024. <https://www.sciencedirect.com/science/article/pii/001670379500123>

Zweng, M. M., J. R. Reagan, D. Seidov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, A. V. Mishonov, O. K. Baranova, K. Weathers, C. R. Paver, and I. Smolyar, 2018. World Ocean Atlas 2018, Volume 2: Salinity. A. Mishonov Technical Ed.; in preparation.

Appendix

Table 1: *Output from final multivariate model*

	Estimate	Std. Error	t value	Pr(> t)
Intercept	34.11	8.62	3.96	<.001 ***
Temperature	-0.03	0.01	-2.40	0.02 *
Favia	-27.32	33.79	-0.81	0.42
Montastraea	38.93	70.50	0.55	0.58
Porites	-74.43	13.37	-5.57	<.001 ***
pH	-2.91	1.11	-2.62	0.01 **
Salinity	-0.18	0.04	-4.63	<.001 ***
Age	0.01	0.00	4.34	<.001 ***
Temp:Favia	0.00	0.05	-0.05	0.96
Temp:Montastraea	0.12	0.13	0.94	0.35
Temp:Porites	-0.06	0.02	-2.68	0.01 **
Favia:pH	2.18	4.37	0.50	0.62
Montastraea:pH	-6.39	7.62	-0.84	0.40
Porites:pH	8.23	1.72	4.78	<0.001 ***
Favia:Salinity	0.26	0.11	2.29	<0.1*
Montastraea:Salinity	0.25	0.42	0.59	0.56
Porites:Salinity	0.25	0.05	5.20	<0.001 ***

Table 2: Uranium concentration estimates from our model in comparison to Inoue et al. (2011) experimental data

pH level	Temperature	Salinity	Uranium concentration - Inoue et al. (2011)	Uranium concentration - model prediction
7.8	27°C	33 ppt	3.15	4.58
8.0	27°C	33 ppt	3.05	3.99

Code from R (also available in .rmd format):

```
``{r, warning=FALSE, message=FALSE}
#reading in necessary libraries and original dataset
library(ggplot2)
library(Rcpp)
library(readr)
library(readxl)
library(tidyverse)
library(ggmap)
Coral <- read_excel("~/FossilCoral/Datasets/Surface Coral Matlab Compiled Data.xls")
longlat <- read_csv("~/FossilCoral/Archives/LongLats.csv")
WOATemp1Deg <- read_csv("~/FossilCoral/Good Data/WOATemp1Deg.csv")
WOASalinity1Deg <- read_csv("~/FossilCoral/Good Data/WOASalinity1Deg.csv", skip = 1)
ph.df <- read_csv("~/FossilCoral/Datasets/pHVals_Lat_Long.csv")
sourceCpp("~/FossilCoral/Archives/distFunc.cpp") #distance function
``
```

```
#Steps/Code Leading to Final Clean Dataset
##Cleaning original data and adding longitude/latitude
```

These next several sections are all of the code that was written in order to clean the variables in the original coral dataset and merge pH, temperature, and salinity. This first section selects variables of interest, adds in longitudes & latitudes, and cleans the site, species, U238, and calcite variables.

```

``{r, warning=FALSE, message=FALSE}
#selecting and renaming columns
keepers<-c(3,4,12,14,15,16,38,40,47)
Coral<-Coral[,keepers]
mynames<-c("Location","Site","Species","Calcite","U238","U238sig","Age","U234delta","Basin")
names(Coral)<-mynames

#Cleaning Site column
Coral<- Coral %>%
  mutate(Site=ifelse(Site %in% "Huon Penninsula","Huon Peninsula",Site),
         Site=ifelse(Site %in% "Ironshore formation, wells drilled offshore from George
Town","George Town",Site),
         Site=ifelse(Site %in% "Not given, most likely San Salvador Island","San Salvador",Site),
         Site=ifelse(Site %in% "Off-shore",NA,Site),
         Site=ifelse(Site %in% "cored off 'south coast of Barbados","South Barbados",Site),
         Site=ifelse(Site %in% "cored off 'south coast of Barbados'", "South Barbados",Site),
         Site=ifelse(Site %in% "unknown",NA,Site))

#cleaning species & making new genus column
#also creating place, which is Location + site in one variable
##(makes it easier for geocode to find correct coordinates)
badnd<-c("n.d.", "N.D.", "unknown", "unidentified", "not", "", "branching") #values that will become
NA for genus
Coral<- Coral %>%
  mutate(Genus=stringr::str_split_fixed(Species, " ", 2)[,1],
         Genus=stringr::str_replace(Genus, "\\", ""),
         Genus=stringr::str_replace(Genus, "\\?", ""),
         Genus=ifelse(Genus %in% badnd, NA, Genus), #creating NAs
         Genus=ifelse(Genus %in% "Porities", "Porites", Genus),
         Place=paste(Location, Site, sep=" ")) #creating location + site variable

#cleaning U238 columns
Coral$U238 <- ifelse(Coral$U238 == "not reported", NA, Coral$U238) #creating NAs
Coral$U238 <- as.numeric(Coral$U238)

Coral1 <- Coral %>%
  mutate(U238 = as.numeric(as.character(U238)),
         U238Final=ifelse(stringr::str_detect(U238, "not reported"),str_replace(U238, "not
reported", ""),U238))

#Cleaning calcite column
calNames <- with(Coral, unique(Calcite))

```

```

oldcalcite <- calNames[c(3,4,5,6,7,8,16,28,47)] #values that need to be changed
newcalcite <- c("0","0","0","0","NA","NA","5","0","NA") #the new values
newcalcite1 <- data.frame(Calcite=oldcalcite,newcalcite)
Coral <- Coral %>% #joining the new calcite values to coral
  left_join(newcalcite1) %>%
  mutate(newcalcite=as.character(newcalcite))

Coral <- Coral %>%

mutate(newcalcite=ifelse(stringr::str_detect(Calcite,"<"),stringr::str_replace(Calcite,"<",""),newcal
cite),

newcalcite=ifelse(stringr::str_detect(Calcite,"_"),stringr::str_replace(Calcite,"_",""),newcalcite))

Coral <- Coral %>%
  mutate(newcalcite = ifelse(!is.na(newcalcite), newcalcite, Calcite),
         CalciteFinal = as.numeric(newcalcite))

#adding longitude and latitude to the dataset
#originally used geocode to get the longs & lats
##we wrote the geocode longs & lats to the csv file longlat.csv and now we're just using that file
Coral <- cbind(longlat, Coral)

#Grouping genus values
Acrop <- c("Acropora", "Acroporidae")
Favia <- c("Favia", "Faviid", "Faviidae", "Favites")
Goniastrea <- c("Goniastrea", "Goniastrea")
Isopora <- c("Isopora", "Isopora")
Montastraea <- c("Montastraea", "Montastrea")
Platygyra <- c("Platygyra", "Platygyra")
Porites <- c("Porites", "Poritidae")

Coral <- Coral %>%
  mutate(
    Genus1 = as.character(Genus), #was giving me problems if I didn't convert to character
    GenusGrouped = ifelse(Genus1 %in% Acrop, "Acropora",
      ifelse(Genus1 %in% Favia, "Favia",
        ifelse(Genus1 %in% Goniastrea, "Goniastrea",
          ifelse(Genus1 %in% Isopora, "Isopora",
            ifelse(Genus1 %in% Montastraea, "Montastraea",
              ifelse(Genus1 %in% Platygyra, "Platygyra",
                ifelse(Genus1 %in% Porites, "Porites", Genus1)))))))))
    Genus3 = as.factor(GenusGrouped) #wanted to change it back to factor at the end
  )

```

)

...

##Fixing longitudes and latitudes

In this section, we are going back into the longitudes and latitudes and manually fixing ones that the geocode incorrectly gave. We fix Palmyra Island, Galapagos, New Caledonia Island, and several around China.

```
``{r, warning=FALSE, message=FALSE}
##Palmyra Island & Galapagos
fixeruplatlon <- data.frame(Location=c("Palmyra Island", "Galapagos"), newlon=c(162.08,
-90.9656), newlat=c(5.889, -0.9538)) #correct longs & lats

# Add the new lat longs to the coral dataset
newcoral <- Coral %>%
  left_join(fixeruplatlon)

newcoral <- newcoral %>%
  mutate(lat=ifelse(is.na(newlat),lat,newlat),
         lon=ifelse(is.na(newlon),lon,newlon)) %>%
  select(-newlon,-newlat)

Coral <- newcoral
totCoral <- nrow(Coral)

##Repeating the process for more locations
fixeruplatlon1 <- data.frame(Site=c("Hainan Island", "Yongshu Reef", "Meji Reef", "Leizhou
Peninsula", "Nanwan Island", "Moreton Bay", "Houtman Abrolhos Islands", "Shark Bay"),
                           newlon=c(109.9497, 112, 115, 110, 109, 153.3447, 113.8, 113.7849),
                           newlat=c(19.5664, 9, 9, 21, 18, -27.3124, -28.7167, -25.9848))
Coral <- Coral %>%
  left_join(fixeruplatlon1)

Coral <- Coral %>%
  mutate(lat=ifelse(is.na(newlat),lat,newlat),
         lon=ifelse(is.na(newlon),lon,newlon)) %>%
  select(-newlon,-newlat)

##Repeating the process for New Caledonia
fixeruplatlon <- data.frame(Location="New Caledonia Island", newlon=165.5, newlat= -21.5)
```

```
Coral <- Coral %>%  
  left_join(fixeruplatlon)
```

```
Coral <- Coral %>%  
  mutate(lat=ifelse(is.na(newlat),lat,newlat),  
         lon=ifelse(is.na(newlon),lon,newlon)) %>%  
  select(-newlon,-newlat)  
  ...
```

```
###Finding closest temperature data
```

In this section, we use the WOA temperature data and match each WOA site to each Coral site in order to find every WOA site's distance from every Coral site. We then take the three closest WOA sites to each Coral site and average those to find a temperature for each Coral observation.

```
``{r, warning=FALSE, message=FALSE}  
#Selecting variables of interest (long, lat, & temp)  
surfacetemp <- WOATemp1Deg[,1:3]  
nm <- names(surfacetemp)
```

```
#Creating matrices that we will use in the for loop and fill with distances  
longsT <- as.matrix(surfacetemp[,"long"])  
latsT <- as.matrix(surfacetemp[,"lat"])  
nT <- nrow(surfacetemp)  
allDists <- matrix(nrow=totCoral,ncol=nT)
```

```
#using distance function to match sites to sites  
for(i in 1:totCoral){  
  res <- distFuncOneMany(as.double(unlist(Coral[i,"lon"])),  
                        as.double(unlist(Coral[i,"lat"])),  
                        longsT,  
                        latsT, nT)  
  allDists[i,]<-res  
}
```

```
allTemps <- as.matrix(surfacetemp[,"X0"])
```

```
#taking 3 closest temperature locations for each coral sample & averaging them  
closest <- matrix(nrow=totCoral,ncol=nT)  
nclosest <- 3
```

```

temps <- array(dim=totCoral)
for(i in 1:totCoral){
  ords <- order(allDists[i,])
  ordtemp <- allTemps[ords]
  #ordsalt <- ordsalt[Coral]
  Temps2<- mean(ordtemp[i:nclosest])
  temps[i] <- Temps2 #rename to n closest temps
}
...

```

##Finding closest salinity data

Much like for temperature, we use the WOA salinity data and match each WOA site to each Coral site in order to find every WOA site's distance from every Coral site. We then take the three closest WOA sites to each Coral site and average those to find a salinity for each Coral observation.

```

```{r, warning=FALSE, message=FALSE}
#selecting variables of interest
surfacesalt <- WOASalinity1Deg[, 1:3]
names(surfacesalt) <- nm #same names as from temperature

#setting up matrix
totCoral <- nrow(Coral)
longs <- as.matrix(surfacesalt[, "long"])
lats <- as.matrix(surfacesalt[, "lat"])
n <- nrow(surfacesalt)
allDists <- matrix(nrow=totCoral, ncol=n)

#using distance function to match
for(i in 1:totCoral){
 res <- distFuncOneMany(as.double(unlist(Coral[i, "lon"])),
 as.double(unlist(Coral[i, "lat"])),
 longs,
 lats, n)
 allDists[i,]<-res
}

allSalts <- as.matrix(surfacesalt[, "X0"])

#taking n closest temperatures for each coral sample
closest <- matrix(nrow=totCoral, ncol=n)

```

```

nclosest <- 3
salts <- array(dim=totCoral)
for(i in 1:totCoral){
 ords <- order(allDists[i,])
 ordsalt <- allSalts[ords]
 #ordsalt <- ordsalt[Coral]
 Salts2<- mean(ordsalt[i:nclosest])
 salts[i] <- Salts2 #rename to n closest temps
}

```

```

#combining temperature and salt info
salttemp <- cbind(temps, salts)
...

```

##Finding closest pH data

We repeat the same process as above for pH.

```

``{r, warning=FALSE, message=FALSE}
#creating matrix
longsP <- as.matrix(ph.df[, "lon"])
latsP <- as.matrix(ph.df[, "lat"])
nP <- nrow(ph.df)
allDists <- matrix(nrow=totCoral, ncol=nP)

#mapping distances to each coral sample
for(i in 1:totCoral){
 res <- distFuncOneMany(as.double(unlist(Coral[i, "lon"])),
 as.double(unlist(Coral[i, "lat"])),
 longsP,
 latsP, nP)
 allDists[i,]<-res
}

```

```

allpH <- as.matrix(ph.df[, "pHVal"])

```

```

#taking n closest pH samples for each coral sample
closest <- matrix(nrow=totCoral, ncol=nP)
nclosest <- 3
pH <- array(dim=totCoral)
for(i in 1:totCoral){
 ords <- order(allDists[i,])

```



```

ordpH <- allpH[ords]
#ordpH <- ordpH[Coral]
pH2<- mean(ordpH[i:nclosest])
pH[i] <- pH2 #rename to n closest pHs
}
...

```

##Renaming variables, adding pH/temp/salinity to coral data, creating final dataset

This code combines the Coral data with temperature, salinity, and pH info. It also renames variables just so we follow the same convention for each variable name/it's easier to work with. We commented out the line that writes the csv, but if you want to access the final dataset easily, just use CoralFinal.csv in Fossil Coral > Good Data.

```

``{r, warning=FALSE, message=FALSE}
Coral <- cbind(Coral,salttemp, pH)
Coral <- Coral %>%
 mutate(
 Lon = lon,
 Lat = lat,
 Temperature = temps,
 Salinity = salts,
 Genus = Genus3,
 Calcite = CalciteFinal) %>%
 select(Lon, Lat, Temperature, Salinity, pH, Calcite, U238, U238sig, Age, U234delta, Basin,
 Genus, Place, Location, Site)

#write.csv(Coral, "~/FossilCoral/Good Data/CoralFinal.csv", row.names=FALSE)
#creates a csv with our final clean data
#if you don't want to run everything over and over, just use CoralFinal
...

```

After all of the above steps, we created our final, clean dataset that we used for modeling. Below is a variable table indicating the variables still in our Coral dataset and detailing which we used in our final model.

Variable	Source	Used in Final Model?	Description
Lon	Google Maps	No	Longitude of coral sample, grabbed using geocode
Lat	Google Maps	No	Latitude of coral sample, grabbed using geocode
Genus	Original coral data	Yes	Taken from original species column, used only top 4 (Acropora, Porites, Montastraea, Favia)

Age	Original coral data	Yes	Age of coral, thousands of years
Calcite	Original coral data	No	Percent calcite in fossil coral, ppt
U238	Original coral data	Yes	Uranium/Calcium concentration in fossil coral, main response variable
U238sig	Original coral data	No	Error in U238 measurement
U234delta	Original coral data	No	?
Basin	Original coral data	No	Atlantic, Pacific, or Indian
Place	Original coral data	No	Combination of location + site, used to find longitude & latitude

#Analysis Results

##Models

``{r}

```
Coral <- read.csv("~/FossilCoral/Good Data/CoralFinal.csv")
top3genus <- c("Acropora", "Porites", "Favia", "Montastraea")
CoralG <- Coral[which(Coral$Genus %in% top3genus),]
CoralG1 <- CoralG %>%
 filter(!is.na(U238))
```

```
FinalModel <- lm(U238 ~ pH + Genus + Age + Temperature + Salinity + Temperature:Genus +
pH:Genus + Salinity:Genus, data=CoralG1) #final model used on poster
summary(FinalModel)
``
```

##Comparing Inoue et al. (2018) values to our estimated values

``{r}

```
#values that Inoue et al. controlled for
Temperature <- c(27, 27)
Salinity <- c(33, 33)
Age <- c(0,0)
Genus <- c("Acropora", "Acropora")
pH <- c(7.8, 8.0)
newdf <- data.frame(Temperature, Salinity, Age, Genus, pH)
#predicting U238 values from our model
preds <- predict(FinalModel, newdata=newdf)
preds
``
```

##Visualizations

###Visualizations from our poster

``{r}

```

#wrapped histograms with U238, temperature, pH, and genus
CoralG1 %>%
 filter(Genus %in% c("Porites", "Favia", "Acropora", "Montastraea")) %>%
 ggplot(aes(x=pH, y=U238, color=Temperature))+
 geom_point()+
 geom_smooth(method=lm)+
 coord_cartesian(ylim=c(1,5.5))+
 scale_color_gradient2(low = "blue", mid = "white", high = "red", midpoint = 25)+
 facet_wrap(~Genus)+
 labs(x="pH", y="Uranium concentration (ppm)", color="Temperature")

```

```

#wrapped histograms with U238, temperature, salinity, and genus
CoralG1 %>%
 filter(Genus %in% c("Porites", "Favia", "Acropora", "Montastraea")) %>%
 ggplot(aes(x=Salinity, y=U238, color=Temperature))+
 geom_point()+
 geom_smooth(method=lm)+
 coord_cartesian(ylim=c(1,5.5))+
 scale_color_gradient2(low = "blue", mid = "white", high = "red", midpoint = 25)+
 facet_wrap(~Genus)+
 labs(x="Salinity", y="Uranium concentration (ppm)", color="Temperature")

```

```

#comparing our model predictions to Inoue et al. (2011) predictions
pH <- c(7.8, 8.0, 7.8, 8.0)
U238 <- c(4.581011, 3.999881, 3.15, 3.05)
Models <- c("Our model", "Our model", "Inoue et al.", "Inoue et al.")
df <- data.frame(pH, U238, Models)
ggplot(df, aes(x=pH, y=U238, color=Models))+
 geom_point(size=2)+
 labs(y="Uranium concentration (ppm)")+
 coord_cartesian(ylim = c(1, 5.5))

```

```

#map with pH colors and genus overlaid by shape
phMed<-with(ph.df,median(phVal))
mp <- NULL
world_map <- maps::map("world", ".", exact = FALSE, plot = FALSE, fill = TRUE) %>% fortify()
#fixes tidyverse problem
ggplot(world_map)+
 geom_polygon(aes(long, lat, group=group),fill="white", colour="gray") +
 #ggtitle("Map of World with Temp Levels")+
 geom_point(data=ph.df,
 aes(x=lon, y=lat,color=phVal) , size=.25) +
 scale_color_gradient2(low="blue",mid = "white",midpoint=phMed,high="red") +

```

```

 geom_point(data=CoralG, aes(x=Lon, y=Lat, shape=Genus), size=2)
 ...

##Other visualizations (not included in poster)
####Basic histograms with single variables
```{r}
# Histogram of U238
ggplot(Coral, aes(U238)) + geom_histogram(bins = 50, color="white")

#Histogram of pH
ggplot(Coral, aes(pH))+ geom_histogram(bins = 50, color="white")

#Histogram of salinity
ggplot(Coral, aes(Salinity))+ geom_histogram(bins = 50, color="white")

#Histogram of temperature
ggplot(Coral, aes(Temperature))+ geom_histogram(bins = 50, color="white")

#Histogram of age
ggplot(Coral, aes(Age))+ geom_histogram(bins = 50, color="white")
...

####Basic visualizations with non-environmental variables
```{r}
#U238 vs. Age
ggplot(Coral, aes(y=U238, x=Age))+
 geom_point()+
 geom_smooth(method=lm)

#Histogram of U238 in each Basin
ggplot(Coral, aes(x = Basin, y = U238)) +
 geom_boxplot() + coord_flip()

#U238 vs. Calcite Level
ggplot(Coral, aes(y=U238, x=Calcite)) +
 geom_point()+
 geom_smooth(method=lm)
...

####Visualizations looking at pH
```{r}
#U238 and pH

```

```
ggplot(Coral, aes(x=pH, y=U238)) +  
  geom_point() +  
  geom_smooth(method=lm)
```

```
#pH and latitude
```

```
Coral %>%  
  filter(Lat>20) %>%  
  ggplot(aes(y = U238, pH, color = Lat)) +  
  geom_point() +  
  geom_smooth(method = lm)
```

```
#pH and Genus
```

```
ggplot(CoralG1, aes(y = U238, x=pH, color=Genus)) +  
  geom_point() +  
  geom_smooth(method = lm)
```

```
#another way to look at pH and Genus
```

```
ggplot(CoralG1, aes(y = U238, x=pH)) +  
  geom_point() +  
  geom_smooth(method = lm) +  
  facet_wrap(~Genus)
```

```
#pH and Calcite
```

```
ggplot(Coral, aes(y=U238, x=pH, color=Calcite)) +  
  geom_point()+  
  geom_smooth(method=lm)
```

```
#Map of pH only at coral locations
```

```
pHMed <- with(Coral, median(pH))  
mp <- NULL  
world_map <- maps::map("world", ".", exact = FALSE, plot = FALSE, fill = TRUE) %>% fortify()  
#fixes tidyverse problem  
ggplot(world_map)+  
  geom_polygon( aes(long, lat, group=group),fill="white", colour="gray") +  
  ggtitle("Map of World with pH")+  
  geom_point(data=Coral,  
             aes(x=Lon, y=Lat, color=pH) , size=1.8) +  
  scale_color_gradient2(low="blue",mid = "white",midpoint=8.05,high="red")
```

```
#Map with just pH values
```

```
ggplot(world_map)+  
  geom_polygon(aes(long, lat, group=group),fill="white", colour="gray") +  
  #ggtitle("Map of World with Temp Levels")+
```

```

geom_point(data=ph.df,
           aes(x=lon, y=lat,color=phVal) , size=.25) +
scale_color_gradient2(low="blue",mid = "white",midpoint=phMed,high="red")+
geom_point(data=Coral, aes(x=Lon, y=Lat), color="black")
...

```

```

###Visualizations looking at salinity

```

```

``{r}

```

```

#Salinity and U238

```

```

ggplot(Coral, aes(x=U238, y=Salinity)) +
  geom_point()+
  geom_smooth(method=lm)

```

```

#Salinity and latitude filtered by latitude

```

```

Coral %>%
  filter(Lat>20) %>%
  ggplot(aes(y = U238, Salinity, color=Lat)) +
  geom_point() +
  geom_smooth(method = lm)

```

```

#Salinity and latitude (unfiltered)

```

```

ggplot(Coral, aes(y = U238, Salinity, color=Lat)) +
  geom_point() +
  geom_smooth(method = lm)

```

```

#Salinity and Genus with U238 and Temperature

```

```

CoralG %>%
  filter(Temperature > 16) %>%
  ggplot(aes(y = U238, Salinity, shape = Genus, color=Temperature)) +
  geom_point() +
  geom_smooth(method = lm)+
  scale_color_gradient2(low="blue", mid="white", high="red", midpoint=21)

```

```

#another way of looking at salinity, genus, u238, temperature (filtered temperature)

```

```

CoralG %>%
  filter(Temperature > 16) %>%
  ggplot(aes(y = U238, Salinity, color=Temperature)) +
  geom_point() +
  geom_smooth(method = lm)+
  scale_color_gradient2(low="blue", mid="white", high="red", midpoint=21)+
  facet_wrap(~Genus)

```

```

#Salinity and Calcite

```

```

ggplot(Coral, aes(y=U238, x=Salinity, color=Calcite)) +
  geom_point()+
  geom_smooth(method=lm)

#Map of Salinity only at coral locations
world_map_2 <- maps::map("world", ".", exact = FALSE, plot = FALSE, fill = TRUE) %>%
fortify()
ggplot(world_map_2) +
  geom_polygon(aes(long,lat,group=group),fill="white", colour = "gray") +
  ggtitle("Map of Salinity") +
  geom_point(data=Coral,
             aes(x=Lon, y = Lat, color = Salinity), size = 1) +
  scale_color_gradient2(low="green",mid="yellow", midpoint = median(na.omit(Coral$Salinity)),
high="red")

#Map with just salinity
summary(WOASalinity1Deg)
world_map_2 <- maps::map("world", ".", exact = FALSE, plot = FALSE, fill = TRUE) %>%
fortify()
WOASalinity1Deg$Salinity <- WOASalinity1Deg[,3]
Filtered <- WOASalinity1Deg %>%
  filter(Salinity > 20)
ggplot(world_map) +
  geom_polygon(aes(long,lat,group=group),fill="white", colour = "gray") +
  ggtitle("Map of Salinity") +
  geom_point(data=Filtered, aes(x=Filtered[,2], y = Filtered[,1], color = Filtered[,3]), size = 1) +
  scale_color_gradient2(low="green",mid="yellow", midpoint = 34.4, high="red")+
  labs(color="Salinity")
```



```

###Visualizations looking at temperature
```{r}
#Temps
ggplot(Coral, aes(x=U238, y=Temperature)) +
 geom_point()

#Temps and latitude
Coral %>%
 filter(Lat>20) %>%
 ggplot(aes(y = U238, Temperature, color=Lat)) +
 geom_point() +
 geom_smooth(method = lm)

```


```

```

#Temperature and Genus
ggplot(CoralG, aes(y = U238, x=Temperature, color=Genus)) +
  geom_point() +
  geom_smooth(method = lm)

#Temperature and Calcite
Coral %>%
  filter(!is.na(Calcite),
         Calcite < 15) %>%
ggplot(aes(y=U238, x=Temperature, color=Calcite)) +
  geom_point()+
  geom_smooth(method=lm)

#Temperature & basin
ggplot(Coral, aes(y = U238, Temperature, color = Basin)) +
  geom_point() +
  geom_smooth(method = lm)

#Map of Temp only at coral locations
tempMed <- with(Coral, median(Temperature))
mp <- NULL
world_map <- maps::map("world", ".", exact = FALSE, plot = FALSE, fill = TRUE) %>% fortify()
#fixes tidyverse problem
ggplot(world_map)+
  geom_polygon( aes(long, lat, group=group),fill="white", colour="gray") +
  ggtitle("Map of World with Temperature")+
  geom_point(data=Coral,
            aes(x=Lon, y=Lat, color=Temperature) , size=.5, position="jitter") +
  scale_color_gradient2(low="blue", high="red", midpoint=tempMed, mid="orange")

#calcite quantiles
with(Coral, hist(Calcite))
with(Coral, quantile(log(Calcite+.01), na.rm=TRUE, probs=seq(0,1,by=0.1)))

#Map with just temperature
Temp <- read.csv("~/FossilCoral/Good Data/WOATemp1Deg.csv")
TempMed <- with(Temp, mean(X0, na.rm=TRUE))
ggplot(world_map)+
  geom_polygon(aes(long, lat, group=group),fill="white", colour="gray") +
  ggtitle("Map of World with Temp Levels")+
  geom_point(data=Temp,
            aes(x=long, y=lat,color=X0) , size=.25) +

```



```
scale_color_gradient2(low="blue",mid = "white",midpoint=TempMed,high="red") +  
geom_point(data=Coral, aes(x=Lon, y=Lat), color="black")  
...
```