Broad and fine scale factors predicting microclimates on coral reefs

Jaelyn Bos

Last updated: May 03, 2021

Introduction

Despite their small geographic area, coral reefs play an outsize role in maintaining marine biodiversity.[1] Unfortunately, these spectacular ecosystems are seriously threatened by rising ocean temperatures[2].
However, coral reefs are thermally heterogeneous on multiple scales, and corals vary significantly in their thermal tolerance[3]. Some corals thrive in shallow seas or intertidal pools with routine high temperatures well above the bleaching thresholds of corals in more temperate locations[4]. In fact, conspecific corals inhabiting different temperature regimes can have different thermal tolerances, as measured in bleaching thresholds and growth at high temperatures[5]. This can be true not only at the inter-reef scale, but also at the intra-reef scale: corals inhabiting hot microclimates within reefs can have higher bleaching resistance than corals inhabiting more temperate areas of the same reef[6].
Intriguingly, these effects sometimes persist intergenerationally[7,8] , indicating that hot microclimates might function as reservoirs of heat adapted genotypes. Recent research even indicates that input of larvae from hotter reefs can boost bleaching tolerances on recipient reefs[9], adding to the importance of locating these potential sources of heat adapted genotypes. Since these patches of heat adapted corals can probably occur on multiple scales, we need to consider both broad scale geographic and fine scale geomorphic factors as predictors of reef temperature regimes.

Methods

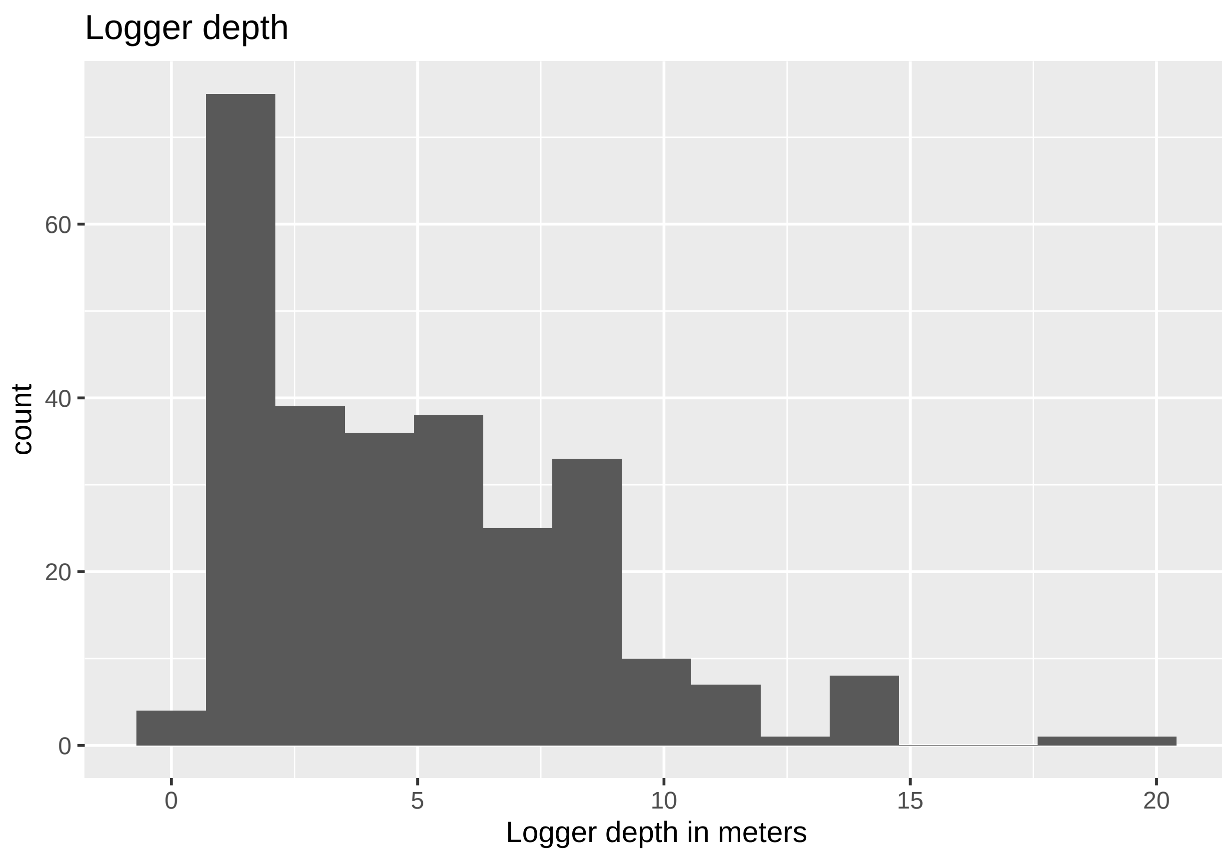
This analysis relies on temperature data from submerged temperature loggers placed by the Australian Institute of Marine Sciences (AIMS) and the United States' National Oceanographic and Atmospheric Administration (NOAA).
NOAA's National Coral Reef Monitoring program deploys submerged SeaBird Electronics brand temperature loggers on coral reef throughout US and territorial waters[10]. I downloaded the Coral Reef Monitoring program data through the NOAA National Centers for Environmental Information portal. The Hawaii data used in this project spans a timeframe from March 2005 to September 2019, though most of the individual points only have data for a subset of this time. Most loggers recorded data at 30 minute intervals.
I downloaded the AIMS data from the AIMS Sea Water Temperature Observing System website. The data was recorded by a combination of Odyssey brand and Sensus Altra brand loggers, mostly recording temperature every five or ten minutes[11]. I used data points from January 2010 to present, though many of the individual logger points only include data for part of that timeframe.
For both the NOAA and the AIMS data, logger metadata included the depth for each submerged logger in meters, as well as latitude and longitude for each logger.
I downloaded geomorphic class data from the Allen Coral Atlas. All logger points used in this analysis fell in the Allen Coral Atlas regions of "Hawaiian Islands," "Great Barrier Reef and Torres Strait," "Timor and Arafura Seas," and "Western Australia." The Allen Coral Atlas relies on Planet Dove data, which in turn relies on imagery from Dove, Sentinel 2 and Landsat satellites. The geomorphic class algorithm also uses wave exposure data from the National Oceanic and Atmospheric Associate (NOAA) Wave Watch III global wave model. The Allen Coral Atlas creates maps of geomorphic classes using a combination of Random Forest classifiers and Object Based Analysis[12].
All of the data preprocessing was run in Python. I used the package Geopandas to extract geomorphic class values to the logger points, and excluded points not associated with geomorphic classes. I also excluded days

with fewer than 12 temperature logs, or gaps of more than six hours between temperature logs, then excluded months with fewer than 25 days worth of data. Finally, I downsampled all of the temperature logs at each point to even thirty minute spacing.

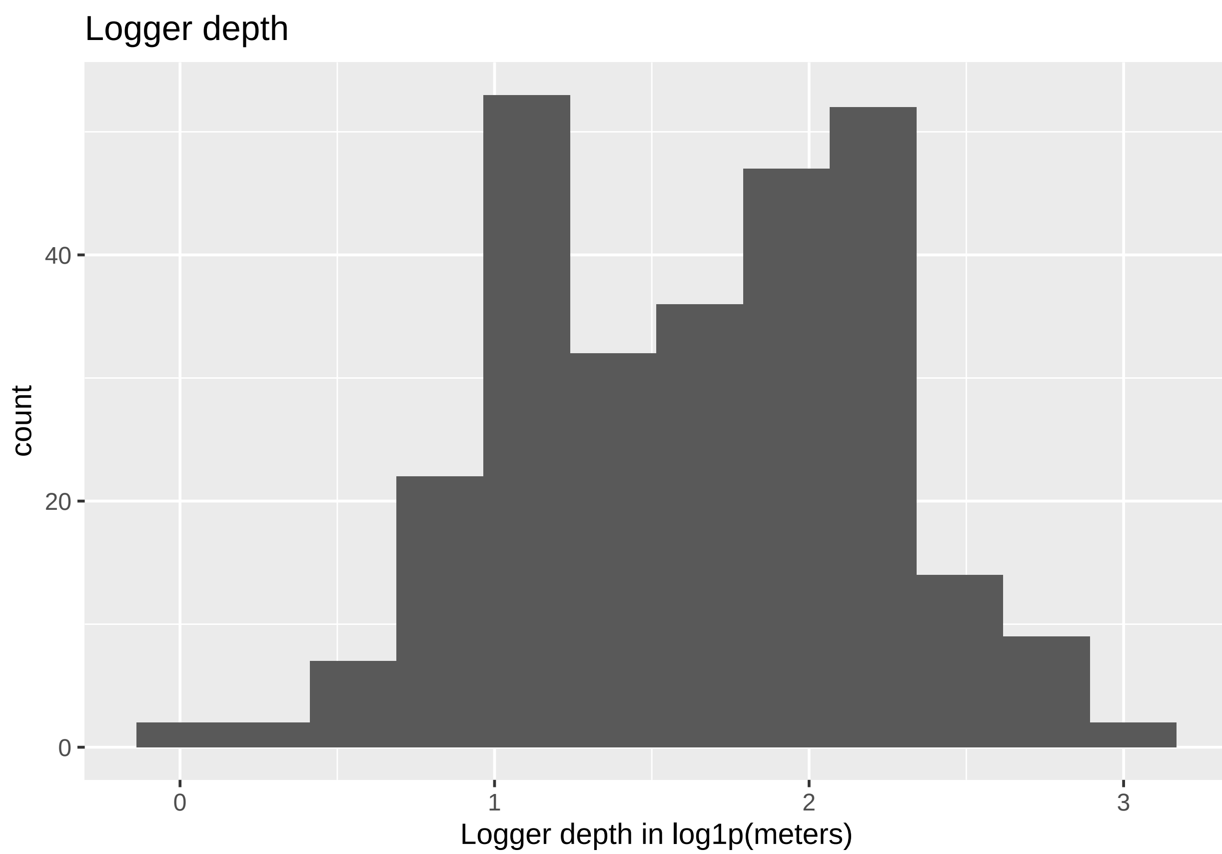
I used the preprocessed data to calculate maximum monthly mean, a statistic typically used to help assess coral bleaching likelihoods. Maximum monthly mean is a description of baseline summer temperature, and higher maximum monthly means are typically associated with higher bleaching thresholds, or lower probability of bleaching at a given temperature. I calculated maximum monthly mean by pooling all of the data for each month across at each logger and taking the mean. For example, if a given logger had data from January of 2012 to December of 2014, I calculated a January mean encompassing data from January 2012, January 2013, and January 2014, and so on for each month. Then, for each logger, I selected the month with the highest mean temperature. This is my maximum monthly mean temperature, abbreviated MMM.

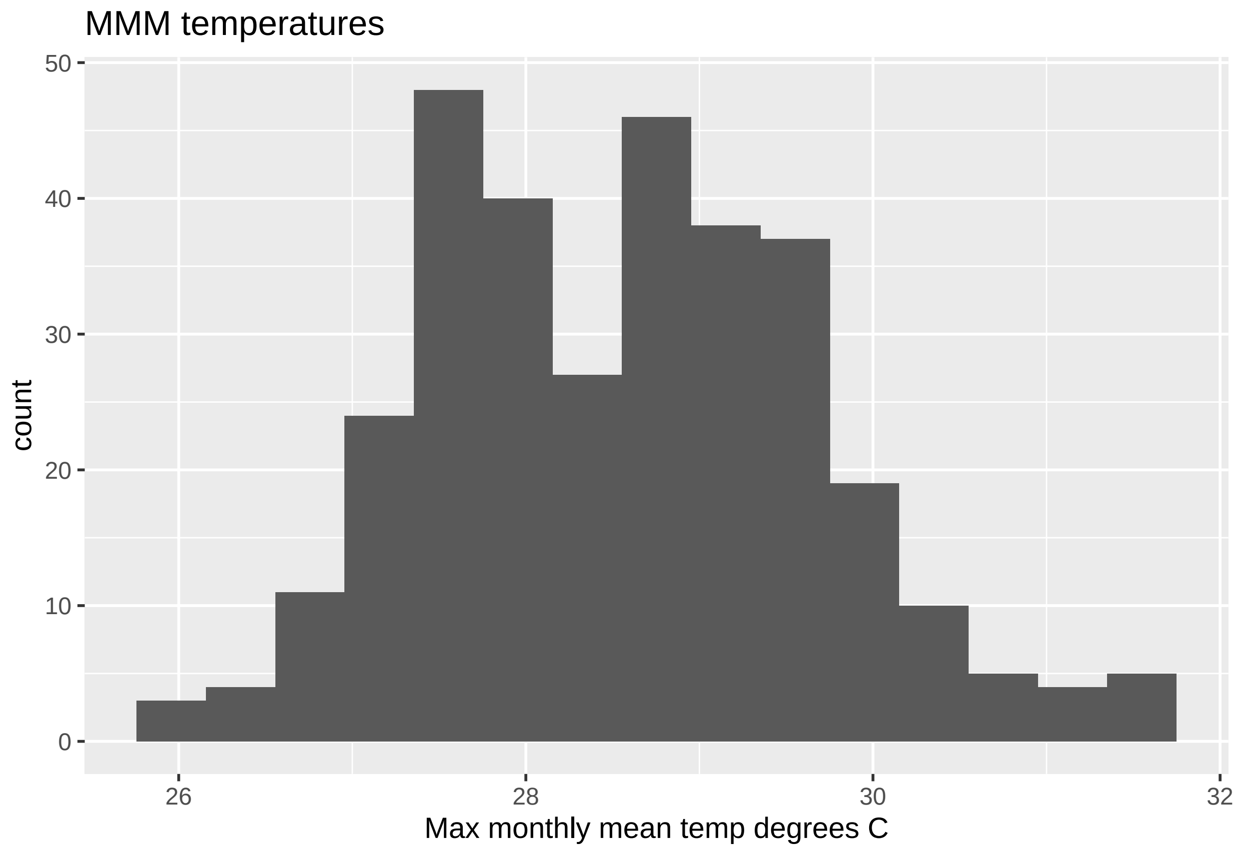
I log transformed the depth variable before analysis to make it more normal.

```
## Warning: Removed 43 rows containing non-finite values (stat_bin).
```




```
## Warning: Removed 43 rows containing non-finite values (stat_bin).
```


Logger depth
count
40
20
0
0
1
2
3
Logger depth in log1p(meters)

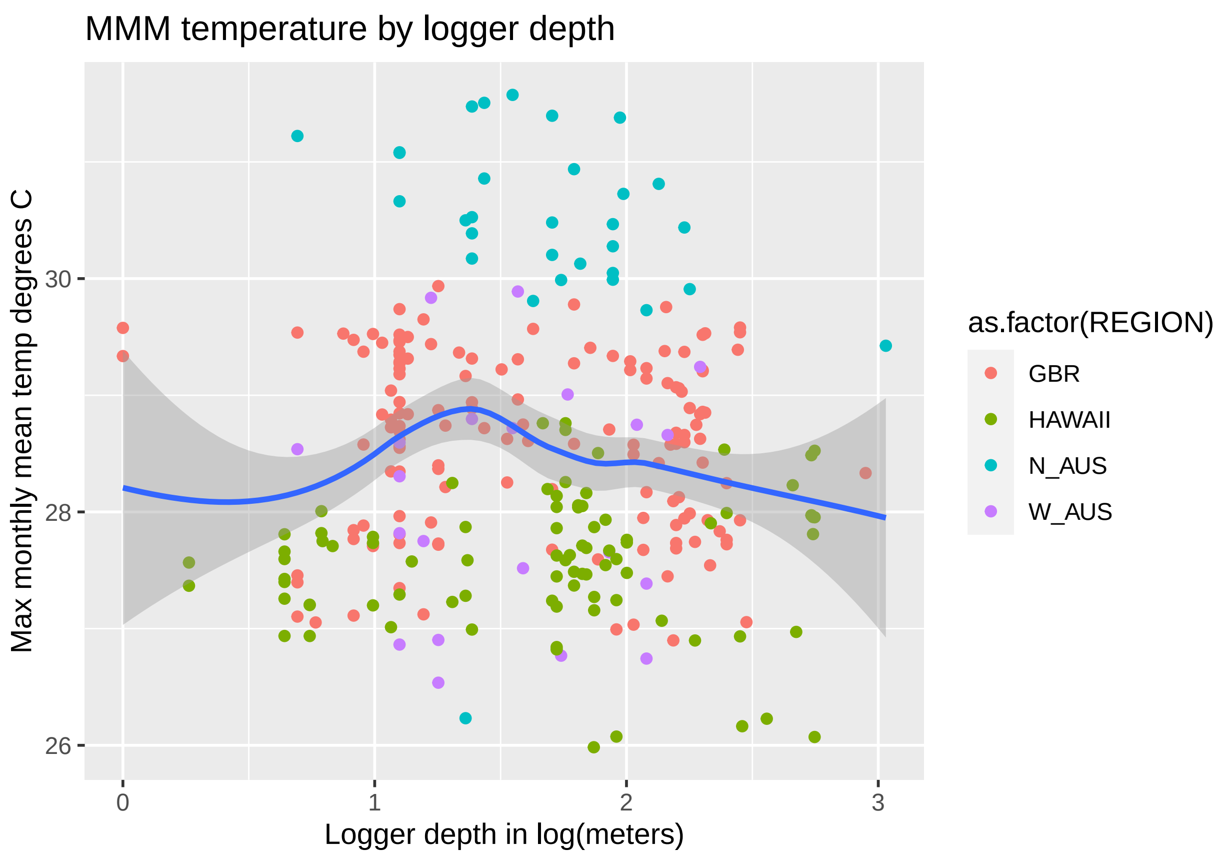


Results

I visually examined the relationships between my independent variables and my dependent variables.

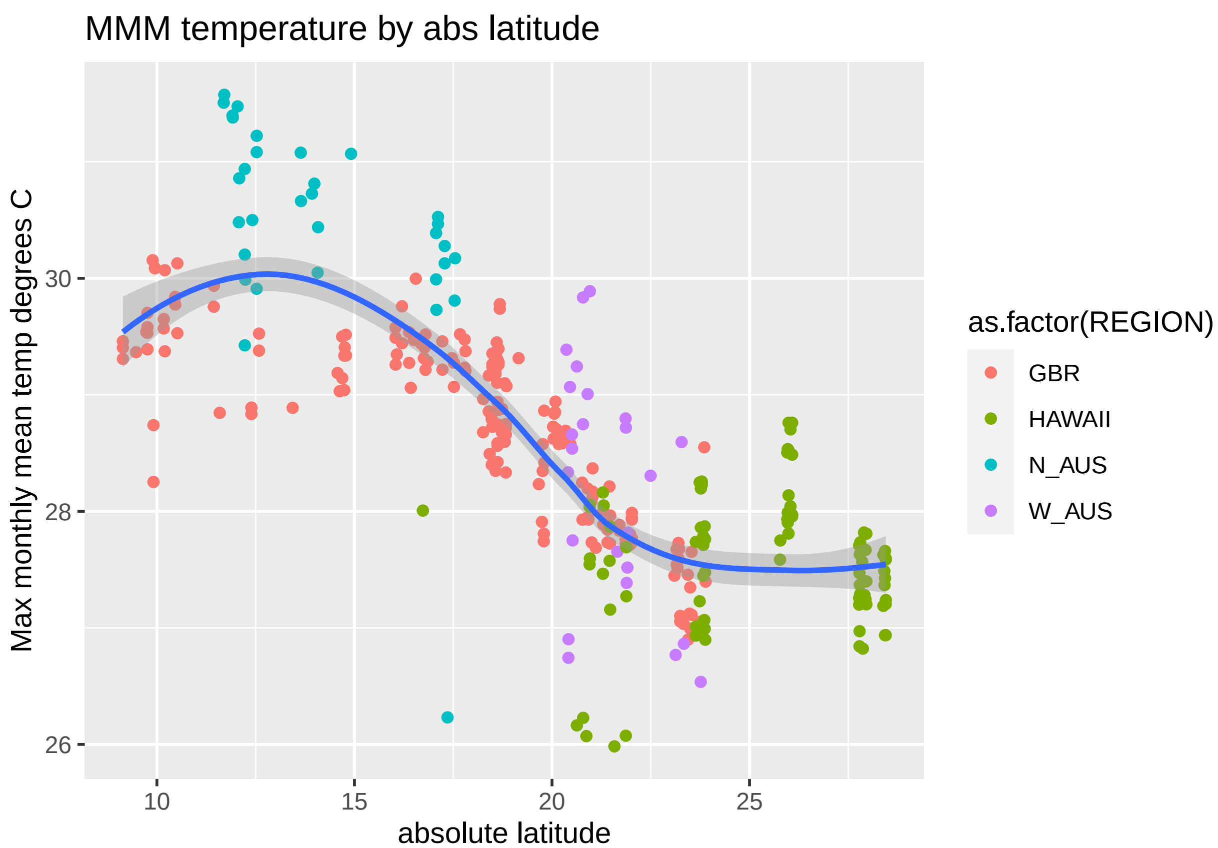
The relationship between depth and MMM temperature lacks an obvious pattern.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Conversely, depth and MMM temperature have a clear, though nonlinear, relationship.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



I ran some single variable linear models as a basic way of looking at the relationships between my variables.

```
#Rsquared MMM temp ~ geomorphic class linear model
print(summary(rclass_lm)$r.squared)
```

```
## [1] 0.13369
```

```
#Rsquared MMM temp ~ depth linear model
print(summary(depth_lm)$r.squared)
```

```
## [1] 0.0007003301
```

```
#Rsquared MMM temp ~ absolute latitude linear model
print(summary(lat_lm)$r.squared)
```

```
## [1] 0.5765584
```

Preliminarily, latitude looks important. Unfortunately, my data does not appear to meet all of the assumptions for linear models, and in particular does not demostrate a linear relationship between the dependent variable and each of its predictors.

To avoid this issue, I also ran several nonparametric Generalized Additive Models (GAMs).I included absolute latitude as a smooth term, and depth and geomorphic class as parametric terms.

I checked model assumptions using the gam.check function from the package "mgcv" and compared the models in term of both deviance explained and AIC.

```
print(AIC(lat_gam,depth_gam,rclass_gam,combo_gam,lat_rclass_gam))
```

```
##                        df      AIC
## lat_gam          9.847409 596.2000
```

```
## depth_gam       3.000000 859.7288
## rclass_gam     11.000000 953.8645
## combo_gam      19.442389 511.3320
## lat_rclass_gam 18.409602 566.4462
```

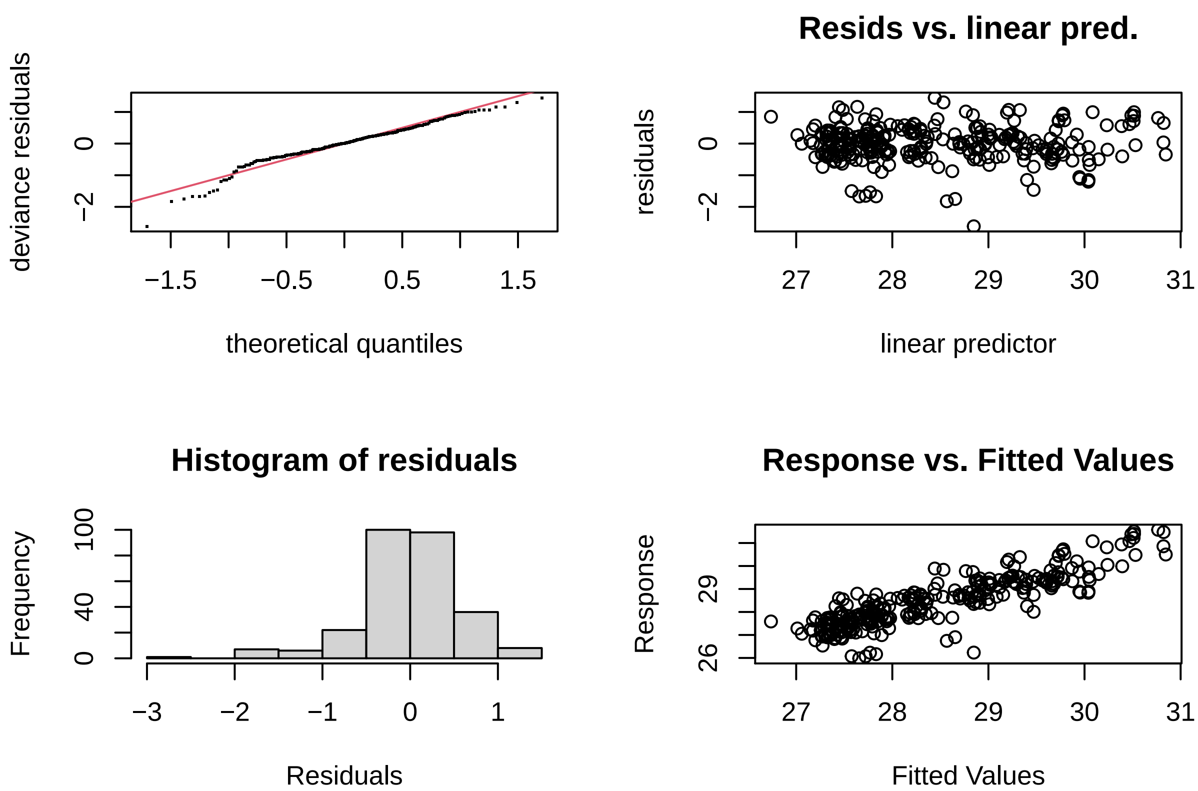
By AIC, the best model includes all three variables, though the depth variable never appears as significant.

```
print(summary(combo_gam))
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## TEMP_C ~ s(ABS_LAT) + LOG_DEPTH + as.factor(RCLASS)
##
## Parametric coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         29.35627    0.25301 116.029  < 2e-16 ***
## LOG_DEPTH                           -0.10606    0.07411  -1.431 0.153598
## as.factor(RCLASS)Shallow Lagoon     -1.02211    0.37755  -2.707 0.007235 **
## as.factor(RCLASS)Deep Lagoon        -0.40682    0.25312  -1.607 0.109217
## as.factor(RCLASS)Back Reef Slope    -0.26126    0.26101  -1.001 0.317796
## as.factor(RCLASS)Inner Reef Flat    -0.87931    0.26589  -3.307 0.001076 **
## as.factor(RCLASS)Outer Reef Flat    -0.84026    0.24603  -3.415 0.000739 ***
## as.factor(RCLASS)Reef Crest         -1.28808    0.30235  -4.260 2.86e-05 ***
## as.factor(RCLASS)Sheltered Reef Slope -0.70863    0.25075  -2.826 0.005080 **
## as.factor(RCLASS)Reef Slope         -0.88253    0.24828  -3.555 0.000450 ***
## as.factor(RCLASS)Plateau            -0.83597    0.27832  -3.004 0.002928 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df    F p-value
## s(ABS_LAT) 7.442  8.393 72.7  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.729   Deviance explained = 74.6%
## GCV = 0.36746  Scale est. = 0.34308   n = 278
```

The assumptions seem reasonably met.

```
gam.check(combo_gam)
```




```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 6 iterations.
## The RMS GCV score gradient at convergence was 4.257598e-07 .
## The Hessian was positive definite.
## Model rank =  20 / 20
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##               k'  edf k-index p-value
## s(ABS_LAT) 9.00 7.44    0.88    0.03 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I also ran a Random Forest model using the same data as the GAM models.

```
##  [1] "X"          "level_0"    "index"      "MONTH"      "POINTS"
##  [6] "DAILY_DIFF" "TEMP_C"     "LONGITUDE"  "LATITUDE"   "DEPTH"
## [11] "RCLASS"     "COORDS"     "REGION"     "ABS_LAT"    "LOG_DEPTH"
```

The Random Forest model ultimately produced a maximum R-squared value very similar to the best GAM model.

```
print(mmm_rf$results$Rsquared %>% max)
```

```
## [1] 0.7332973
```

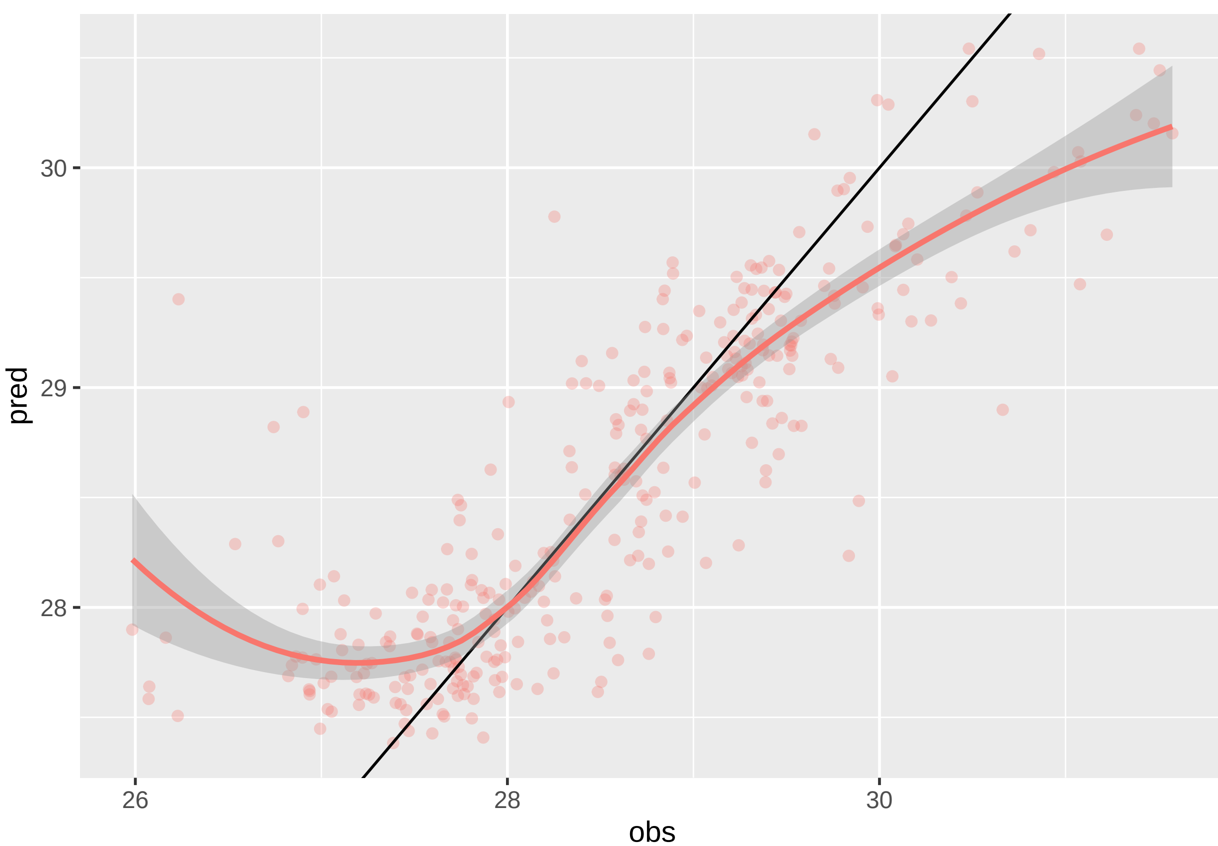
Once again, absolute latitude appears as the most important variable.

```
varImp(mmm_rf)
```

```
## ranger variable importance
##
##                              Overall
## ABS_LAT                    100.00000
## LOG_DEPTH                   14.35849
## RCLASSBack Reef Slope        1.93039
## RCLASSInner Reef Flat        1.70206
## RCLASSDeep Lagoon            1.02160
## RCLASSOuter Reef Flat        0.50467
## RCLASSReef Slope             0.46611
## RCLASSPlateau                0.41404
## RCLASSReef Crest             0.35763
## RCLASSShallow Lagoon         0.04694
## RCLASSSheltered Reef Slope   0.00000
```

I also plotted the observed versus predicted variables for the Random Forest model.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Discussion

In both GAM and random forest models, latitude was the most imporant predictor of maximum monthly mean temperature, far outweighing the effects of within-reef differences. Latitude alone accounted for a majority of the variation in MMM temperatures, though this might change once I incorporate data from other ocean basins. In general, temperature increased as absolute latitude decreased. However, the relationship between temperature and absolute latitude is notably nonlinear.

However, even though latitude was the most significant predictor, it was not the only significant predictor. I also found significant relationships between reef geomorphic class and MMM temperatures in boht the GAM and random forest models. This finding confirms that within-reef factors do have significant effects on temperature variation, even when accounting for latitude and depth. Likewise, the significance of geomorphic class in models including multiple geographic areas shows that morphologic features may have predictable effects across reefs.

Also interesting is the consistent lack of significance contribution from depth. However, this may arise in part from the limited range of depth values in the dataset: the Allen Coral Atlas geomorphic data only goes down to twenty meters, and I threw out the points below that depth that could not be associated with a geomorphic class, or even definitively confirmed as being of a reef. Likewise, I calculated maximum monthly mean temperatures using both diurnal and nocturnal temperature readings, as is standard. In the data cleaning process (not included in this document), I found that depth is much more associated with metrics of variability, like daily maximum minus daily minimum, than metrics of central tendency, like MMM.

Unfortunately, both GAM and Random Forest models have some notable limitations. First, their nonparametric nature makes it difficult to interpret relationships between variable quantitatively. Second, GAMs and Random Forest models can perform very poorly when predicting outside the original sample, especially if the new data to be used includes extreme values of the original predictors. Since an ultimate goal of this project is to predict temperature regimes in areas without in situ loggers, this constitutes a serious drawback.

One way to mitigate, or at least measure, the issue of predicting outside the sample is to divide the data to be fitted into a testing dataset and a training dataset. I intend to do this in the future, as I incorporate more geographic regions into the analysis. Right now, my limitation on incorporating NOAA data from the central Pacific and the Caribbean is that the Allen Coral Atlas is still in the process of creating geomorphic datasets for those regions; however, I hope to have them incorporated by the end of the summer at the latest. Furthermore, I am working on incorporating sea surface temperature data, an additional that I expect to radically change the model, possibly making latitude irrelevant. However, running latitude-based models was still a useful intellectual excerise in that it helps illucidate the relationship between geographic and geomorphic factors in governing temperature regimes on coral reefs.

References

1. Plaisance, L. et al. (2011) The Diversity of Coral Reefs: What Are We Missing? PLoS ONE 6, e25026

2. van Hooidonk, R. et al. (2016) Local-scale projections of coral reef futures and implications of the Paris Agreement. Sci Rep 6, 39666

3. McClanahan, T.R. et al. (2007) Effects of climate and seawater temperature variation on coral bleaching and mortality. Ecol. Monogr. 77, 503–525

4. Craig, P. et al. (2001) High temperatures tolerated by a diverse assemblage of shallow-water corals in American Samoa. Coral Reefs 20, 185–189

5. Howells, E.J. et al. (2016) Host adaptation and unexpected symbiont partners enable reef-building corals to tolerate extreme temperatures. Glob Change Biol 22, 2702–2714

6. Schoepf, V. et al. (2015) Limits to the thermal tolerance of corals adapted to a highly fluctuating, naturally extreme temperature environment. Sci Rep 5, 17639

7. Palumbi, S.R. et al. (2014) Mechanisms of reef coral resistance to future climate change. Science 344, 895–898

8. Howells, E.J. et al. (2012) Coral thermal tolerance shaped by local adaptation of photosymbionts. Nature Clim Change 2, 116–120

9. Kleypas, J.A. et al. (2016) Larval connectivity across temperature gradients and its potential effect on heat tolerance in coral populations. Glob. Change Biol. 22, 3539–3549

10. Ecosystem Sciences Division, Pacific Islands Fisheries Science Center (2020). National Coral Reef Monitoring Program: Water temperature data from subsurface temperature recorders (STRs) deployed at coral reef sites in the Hawaiian Archipelago (NCEI Accession 0210383). NOAA National Centers for Environmental Information. Dataset.

11. Australian Institute Of Marine Science Sea Temperature Observing System (SeaTemps Temperature Logger Program). . (2017) , Australian Institute of Marine Science

12. Allen Coral Atlas (2020). Imagery, maps and monitoring of the world's tropical coral reefs. doi.org/10.5281/zenodo.3833242