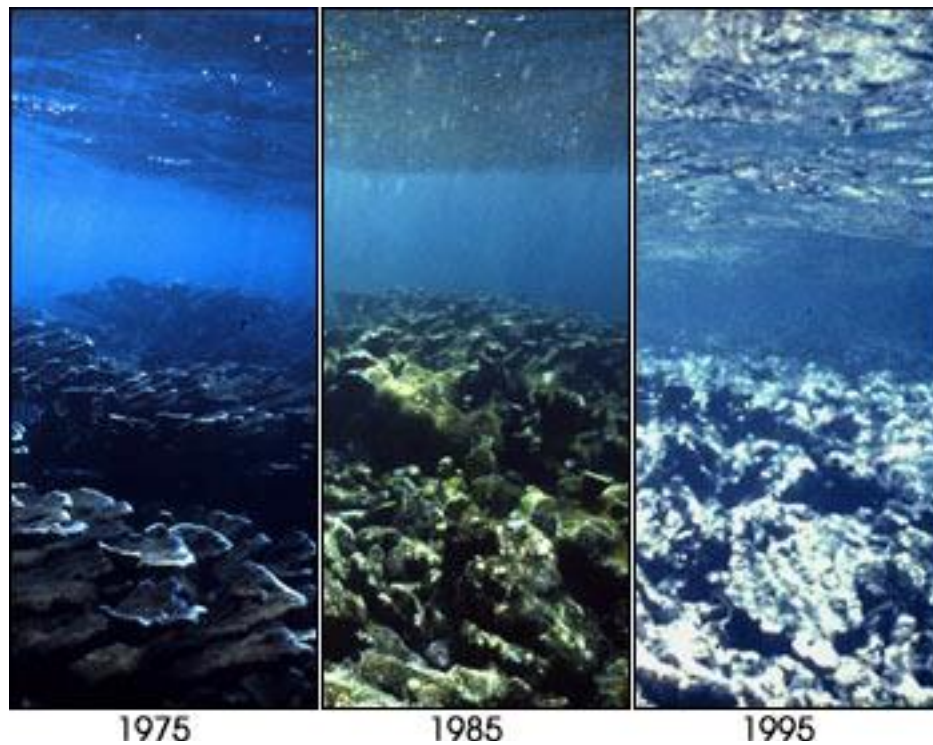


Investigating environmental variables correlating with recent coral mortality and disease in the Florida Keys National Marine Sanctuary using geostatistical analysis

Marie Strader
April 2014



Coral health has decreased dramatically in the past 40 years. This photo shows the decline of elkhorn coral in Carysfort Reef, FL. In 1975 corals are abundant and healthy, by 1985 they are visibly sick and by 1995 the reef is merely algae covered rubble. Photographs courtesy Phillip Dustan, College of Charleston

Introduction:

Coral reefs are declining at an astonishing rate. Major factors that contribute to this decline are temperature stress from rising sea surface temperatures, increasing rates of disease and physical harm from storms (Bruno et al., 2007; D'Angelo & Wiedenmann, 2014). At local scales, water quality can also have an impact on the persistence of coral reefs (Fabricius, 2005). However, it's been difficult to pinpoint spatial relationships between water quality measurements and coral decline, particularly because coral exist along environmental gradients and therefore can vary greatly in their tolerance to environmental stressors (Kenkel et al., 2013).

The coral populations in the Florida Keys National Marine Sanctuary (FKNMS) have experienced substantial loss in the last 30 years (Gardner, Côté, Gill, Grant, & Watkinson, 2003). Environmental and water quality components vary drastically between inshore and offshore reef tracts. For example, due to the buffering of the Gulf Stream, temperature variation on offshore reefs is substantially less than temperature variation on inshore reefs (Kenkel et al., 2013; Lirman et al., 2011). Considering the variation in nutrient enrichment and temperature in the Florida Keys marine system, it is critical to investigate how these water quality measures spatially correlate with recent coral mortality and disease.

The Southeast Environmental Research Center (SERC) has been monitoring multiple water quality measurements throughout sites in the FKNMS and neighboring systems for the past 8 years. I plan to use the geostatistical modeling analysis, kriging, to extrapolate these water quality measurements throughout the entire boundary of the FKNMS in order to correlate these measurements with recent coral decline. Kriging estimates spatial variables based on three assumptions: spatial dependency, constant data variance (stationarity) and a Gaussian distribution (Krivoruchko, n.d.). Because it uses the properties of the data, kriging is more precise than other interpolation methods such as inverse distance weighting. Although there are multiple kriging predictors, ordinary kriging assumes a constant and unknown mean and produces continuous predictions and therefore will be the most appropriate for this analysis (Krivoruchko, n.d.).

Problem:

How does the spatial variation of benthic dissolved nutrients and maximum summer temperature help explain the prevalence of coral disease and recent mortality in the FKNMS? What are the correlation coefficients between these environmental variables and the spatial extent of coral disease and recent mortality?

Data Collection:

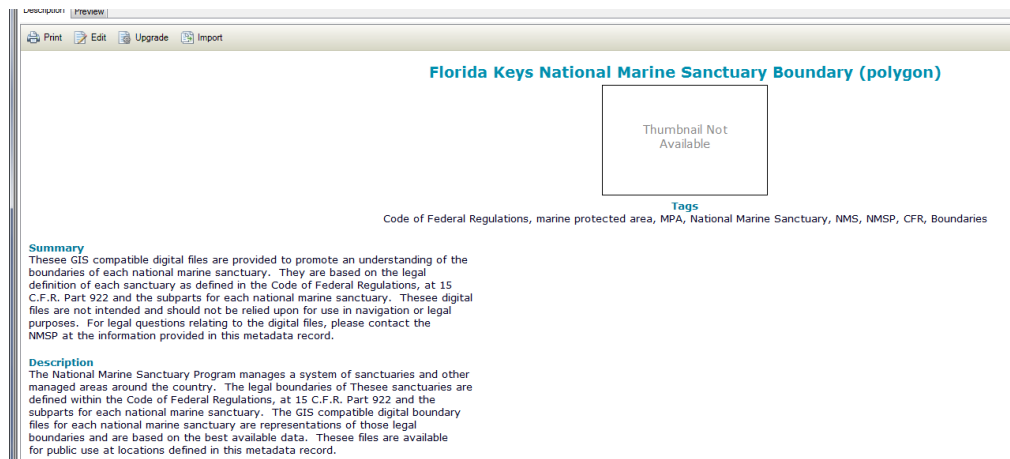
Four datasets were collected for this analysis.

1: The Southeast Environmental Research Center (SERC) has an ongoing water quality-monitoring project for the water quality protection program for the FKNMS. The EPA implemented this protection program in 1996 to assess trends in water

quality. This data is freely available for research and academic purposes (<http://serc.fiu.edu/wqmnetwork/FKNMS-CD/>).

2: The Florida Reef Resilience Program (FRRP) has completed 1568 coral surveys since 1995. The FRRP monitors shallow coral reefs in the Florida Keys and surveys are completed by teams from Universities and numerous governmental agencies. The survey data is available in many different formats and is freely available (<http://frfp.org/data/>). I downloaded two specific data sets. The first was coral disease by site (which includes long and lat coordinates for each site). The second was coral mortality by site (also includes long and lat coordinates for each site).

3: Since the SERC data includes data sites far beyond the sites of the coral surveys, I chose to limit my analysis to the boundaries of the FKNMS. To do this I downloaded the ESRI shapefile of the FKNMS (http://sanctuaries.noaa.gov/library/imast_gis.html).



4: I also needed an outline of the Florida coastline. The source for this data is FMRI, Florida Geographic Data Library. The data was found at (<http://fcelter.fiu.edu/data/GIS/?layer=state>)

Data Preprocessing:

1: For the SERC data I decided limit my analysis to the year 2013. The data was downloaded in .xlm format where each row was a sampling point with corresponding site (in long and lat), date and numerous measurements of water quality. I deleted the rows with dates prior to 2013 and formatted all numerical items to "numeric" and all the text items to "text" and saved it as a .csv file. I opened a blank map in ArcMap. To add this data, I imported the .csv file by File -> Add Data -> Add XY data. The SERC2013.csv file was imported and x and y were set as 'lat' and 'long' columns in the .csv file. The coordinate system was set to a geographic coordinate system (NAD 1893 (2011)). After this file was imported as an 'events' file, I then exported the data as a shapefile: Data -> Export Data and saved it as

“SERC2013.shp”. This shapefile was uploaded directly into the map and the events file (XY data) was removed.

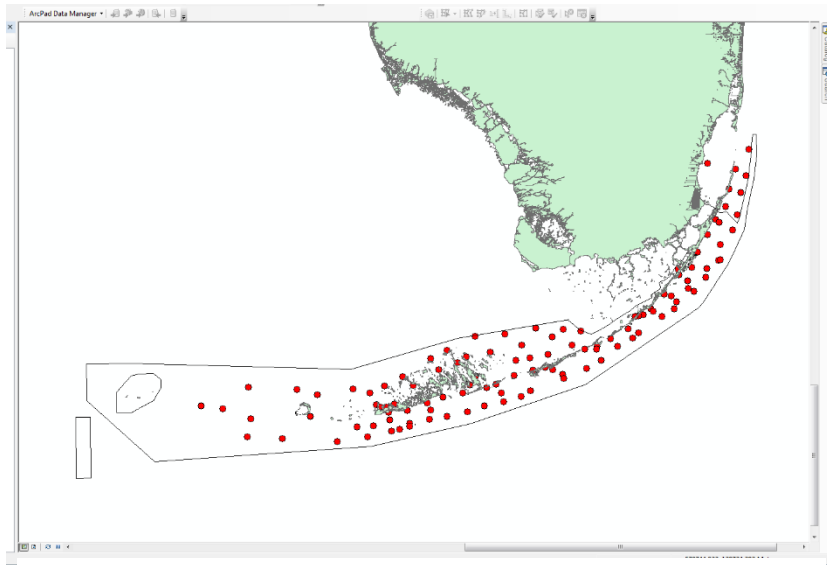


Figure 2: Red dots are sites where SERC data was taken in 2013. The green polygon is the southern tip of FL state and the black polygon is the outline of the FKNMS

2: I had two data sets from the FRRP: coral mortality and coral disease. For both, I removed the data before 2013 and formatted number items as “numeric” and text items as “text” and saved both as .csv files. I uploaded the .csv files as XY data as with the SERC data, set the X and Y as columns ‘lat’ and ‘long’ in the .csv file. I then set the geographic coordinate

system to NAD 1983 (2011). Both of these events files were exported into shapefiles as “2013disease.shp” and “2013mortality.shp” and uploaded directly into the map. The XY event files were removed from the map.

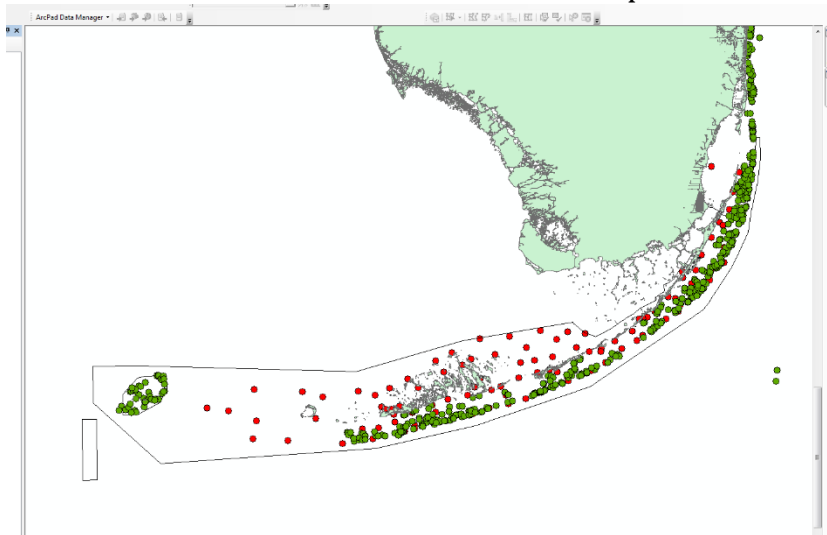


Figure 3: Red dots are sites where SERC data was taken in 2013 and green dots are sites where coral disease surveys were taken in 2013

3 and 4: The FKNMS and FL state coastline were already downloaded as shapefiles so they were added to the map using the “Add Data” option. I made sure the coordinate system matched with the other files NAD1983 (2011).

ArcGIS processing:

1: Kernel Density:

I aimed to visualize the spatial extent of coral disease and recent mortality within the FKNMS boundaries. The coral disease shapefile has multiple fields corresponding to counts of individual coral colonies with specific diseases at the sites surveyed. The count of individuals with any disease at a site is summed in the field "total". The density kernel function can be calculated from point data and calculates the density of features in the neighborhood of those features (ArcGIS help files <http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How+>

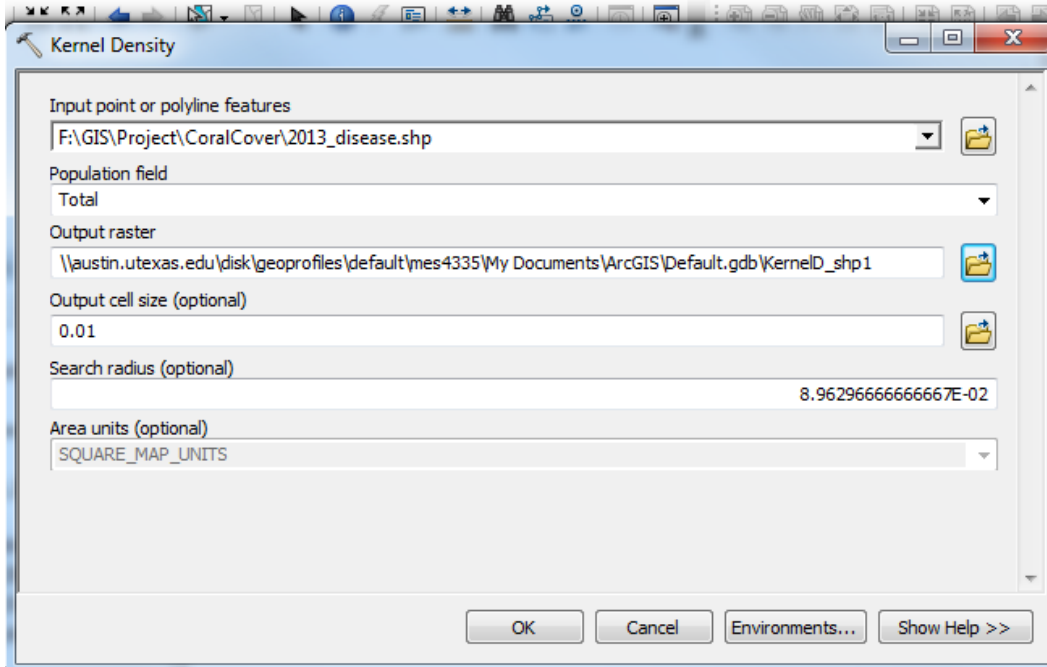


Figure 4: Kernel Density tool for total count of diseased individuals at a site. Resolution of output raster set at 0.01

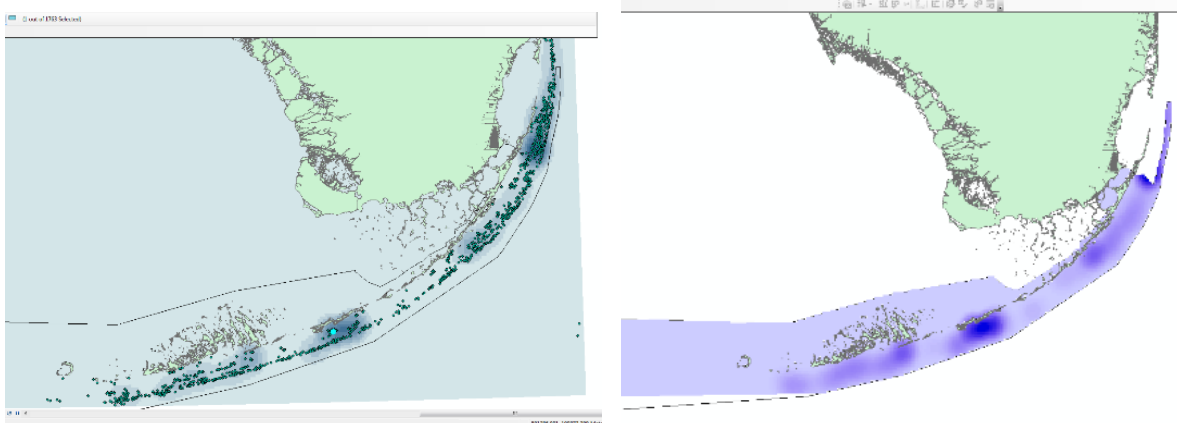


Figure 5: Kernel Density output raster of recent coral mortality in 2013 before and after masking using the FKNMS boundary and change in symbology

Kernel+Density+works). In my case, the feature I want to calculate density from is the 'total' field in the disease dataset, or total number of individuals with disease at a site. The output of this function is a raster. I used the '2013disease.shp' shapefile as the input file, the field 'total' as the 'population' and set the cell size to 0.01.

I used the same process for the mortality dataset. The mortality shapefile has multiple fields that contain different metrics for assessing mortality. The field that I decided to use was NRM or 'Total number of corals with recent mortality'. Again using the kernel density function, the input file was '2013mortality.shp' and the population field was 'NRM'. Again the cell size of the output raster was set to 0.01.

Both of these output rasters were uploaded into the map. It is interesting to note that many of the regions with high recent mortality also have high disease, although these two parameters do not necessarily overlap.

Masking: In order to restrict the density raster to the boundary of the FKNMS, I used the 'extract by mask' tool in the Arc toolkit.

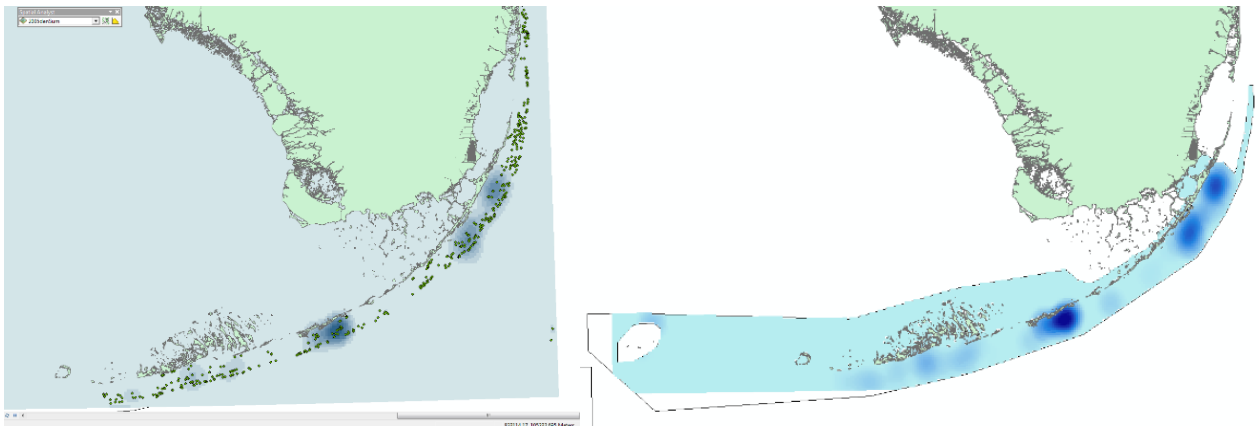


Figure 6: Kernel Density output raster of total coral disease in 2013 before and after masking using the FKNMS boundary and change in symbology

2: Kriging interpolation

I decided to model 3 different benthic parameters that have been known to affect coral health: maximum summer temperature (temp), dissolved oxygen (DO) and dissolved inorganic nitrogen (DIN).

Maximum benthic temperature:

I first explored the data to determine how appropriate it was for kriging. Since data was taken at each site more than one time in the data set I used the maximum value at a site because corals here can live on the cusp of their upper thermal limits. Since this is somewhat randomly sampled data in a natural environment it doesn't perfectly fit the assumptions of the model. However, the histogram (Figure 7) shows a relatively normal (Gaussian) distribution with the exception of some outliers. Transforming the data did not change the overall shape of the distribution so I decided not to do a transformation. Trend analysis showed that there are two polynomials that fit the data. The semivariogram cloud shows that this data is spatially correlated but there is an obvious anisotropy which was incorporated into the model. I ran ordinary kriging, removed a second order polynomial and changed anisotropy to "true" in order to generate my prediction map. The cross validation results show that the model does not robustly predict all the values, but also does not completely fail either.

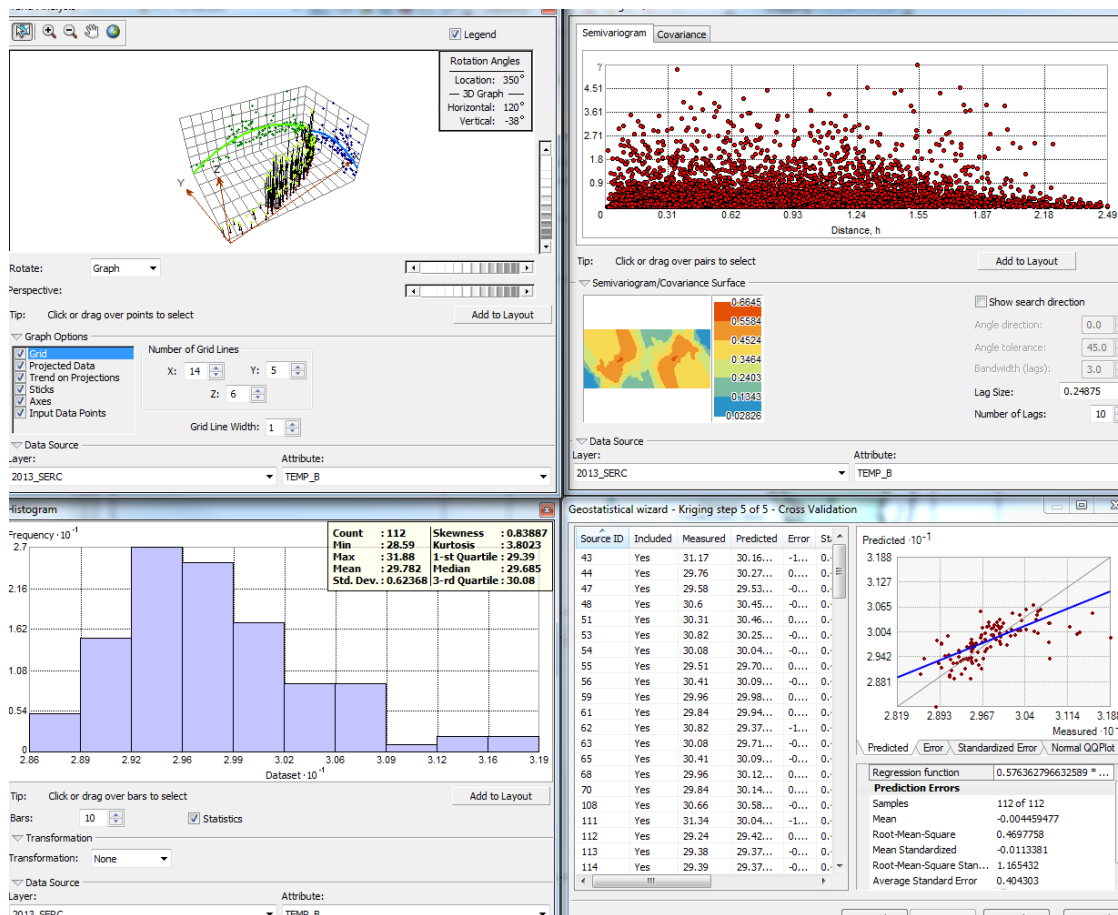


Figure 7: Data exploration and kriging for maximum benthic temperature in 2013. Panel in the top left shows the trend analysis. Panel in the top right shows the semivariogram cloud. Panel in the bottom left shows the histogram of the data. Panel in the bottom right shows the cross validation analysis of the model.

Benthic dissolved oxygen (DO):

I used the mean value at each site over the course of the year for this data set. Exploring this dataset revealed a distribution that was not typically Gaussian. However, none of the possible transformations helped it become more normal so I decided not to use a transformation. Trend analysis showed a similar trend to the benthic maximum surface temperature with a two polynomials fitting to the data. The semivariogram was also less than perfect but still showed that these data points are not independent from each other and that there is a strong anisotropy. For this set of data I ran ordinary kriging without a transformation and removing the second order polynomial. I removed the anisotropy (true) and generated a prediction model. The cross validation results show that the predicted values do not exactly match the actual values. This model is generally less robust than maximum benthic temperature.

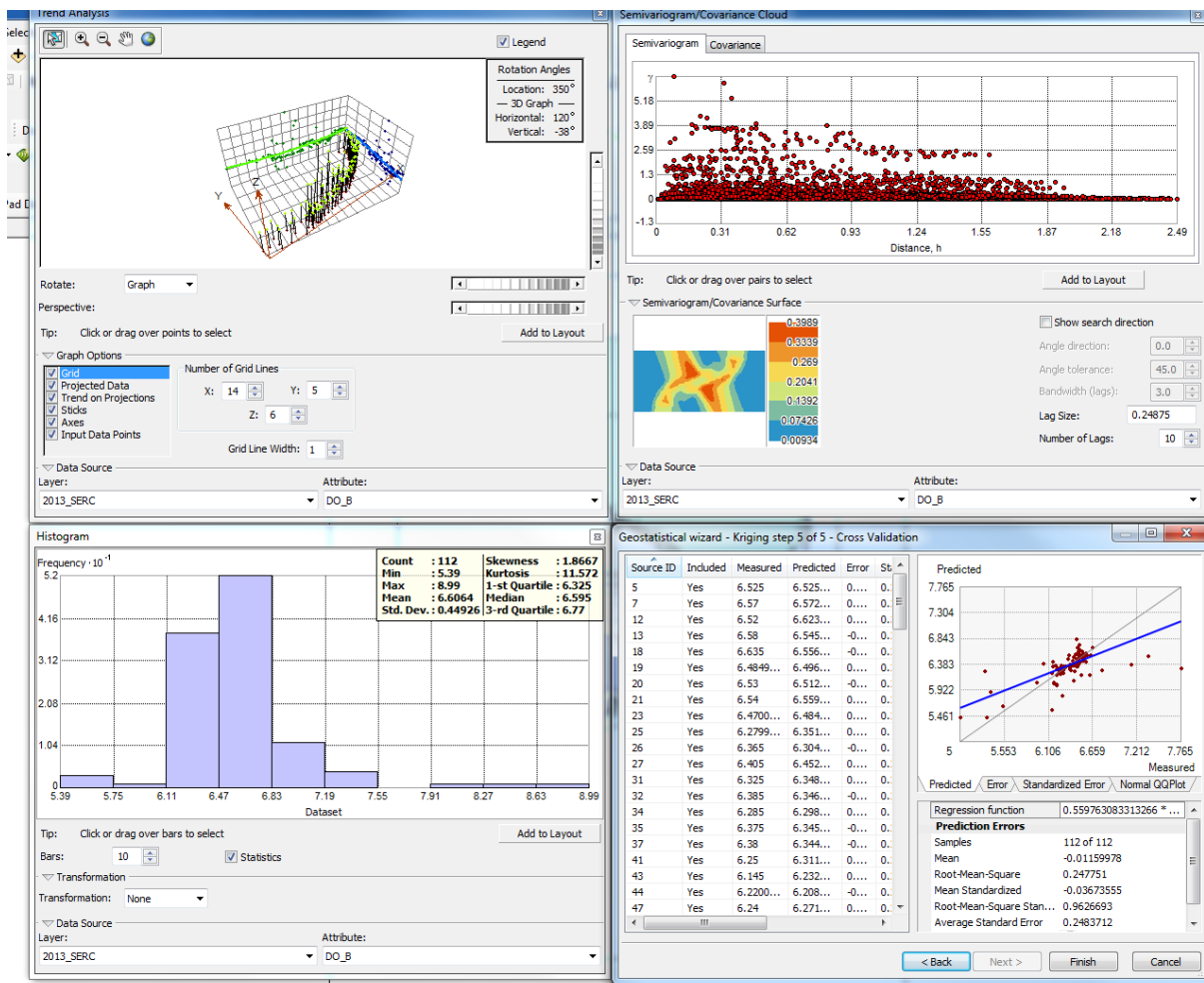


Figure 8: Data exploration and kriging for mean benthic dissolved oxygen in 2013. Panel in the top left shows the trend analysis. Panel in the top right shows the semivariogram cloud. Panel in the bottom left shows the histogram of the data. Panel in the bottom right shows the cross validation analysis of the model.

Benthic dissolved organic nitrogen (DIN):

I also used the mean value at each site for dissolved organic nitrogen (DIN). The data set was even less normal than the previous two sets of data. However I was unable to log-transform the data, probably since there were multiple zero values. Therefore I wasn't able to perform any transformation on this data. However, there did appear to be some spatial autocorrelation by looking at the semivariogram cloud, although there is a strong anisotropy. The trend analysis reveals the same trend as in the DO data. The kriging model used was ordinary with removal of second order polynomials. A prediction map was generated and anisotropy was coded as 'true'. The cross validation plots show that although the predicted values did not exactly match actual values, there was a positive correlation.

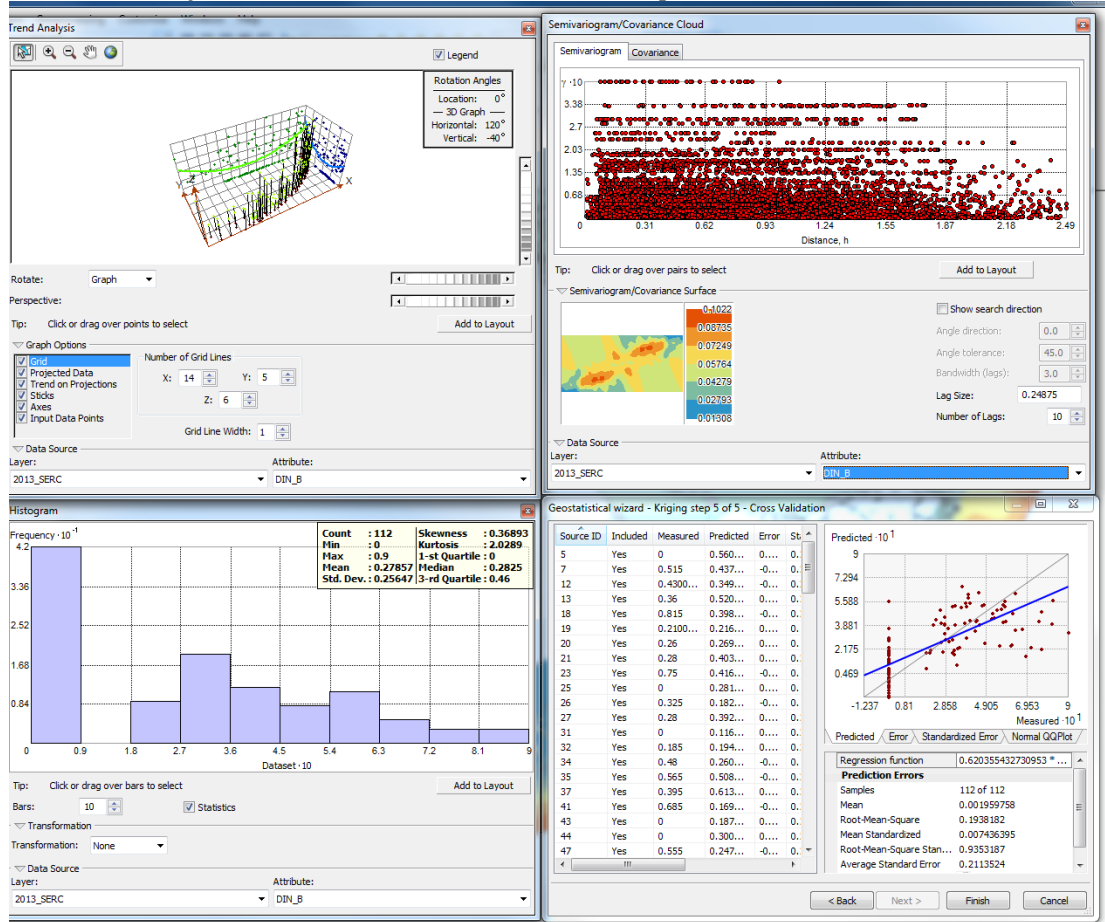


Figure 8: Data exploration and kriging for mean benthic dissolved inorganic nitrogen in 2013. Panel in the top left shows the trend analysis. Panel in the top right shows the semivariogram cloud. Panel in the bottom left shows the histogram of the data. Panel in the bottom right shows the cross validation analysis of the model.

There are multiple ways to examine how robust these models are. Ideally the predictions should be as close to the measured values as possible (Krivoruchko,

n.d.). One way to check this is by the cross-validation method predicted plot where we hope to see a 1:1 correlation between actual and predicted values. Another way is with the root-mean-square prediction error. The smaller the root-mean-square prediction error is, the closer the predictions are to true values. A third way is with the mean prediction error. You want your prediction values to be unbiased, therefore you want your prediction error mean to be as close to zero as possible. A fourth way is to compare the average standard error to the root-mean-square value which tells you if your uncertainty is valid. In the case of these three models: temperature, DO and DIN had root-mean-square values of 0.469, 0.247 and 0.19, respectively. Using this proxy, the DIN model best predicts values. Mean prediction error was -0.0044, -0.012 and 0.0019 for temperature, DO and DIN, respectively. Again, this shows that the DIN model shows the smallest error. Average standard error was 0.40, 0.25 and 0.21 for temperature, DO and DIN, respectively. Since these values are similar to the root-mean-square values, we can say that the uncertainty in our models is valid.

For each of these kriging models, I exported the data as a raster with a cell size of 0.01 and used the boundary of the FKNMS for masking.

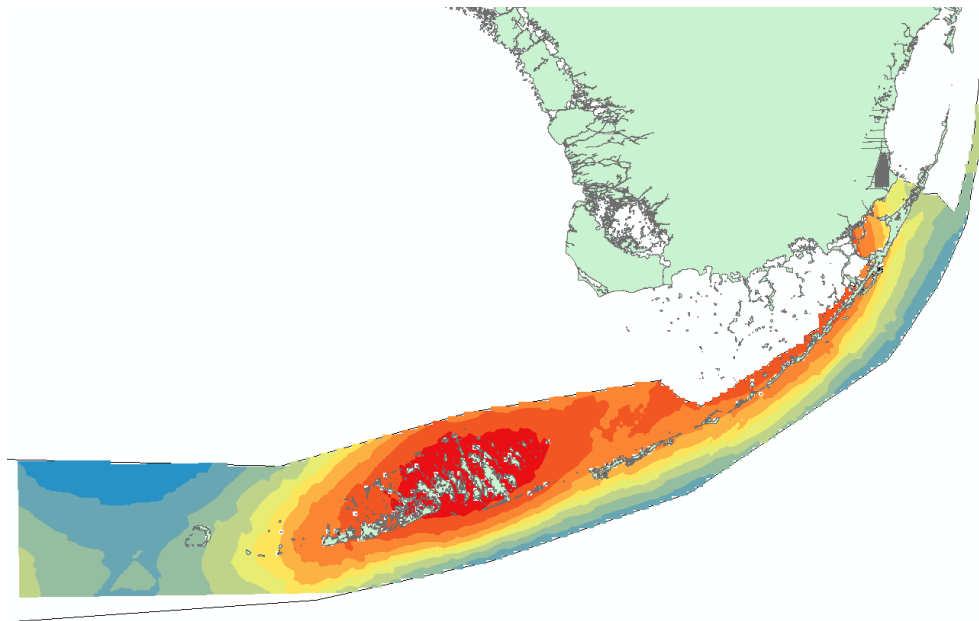


Figure 9: Prediction map of maximum benthic temperature in the FKNMS

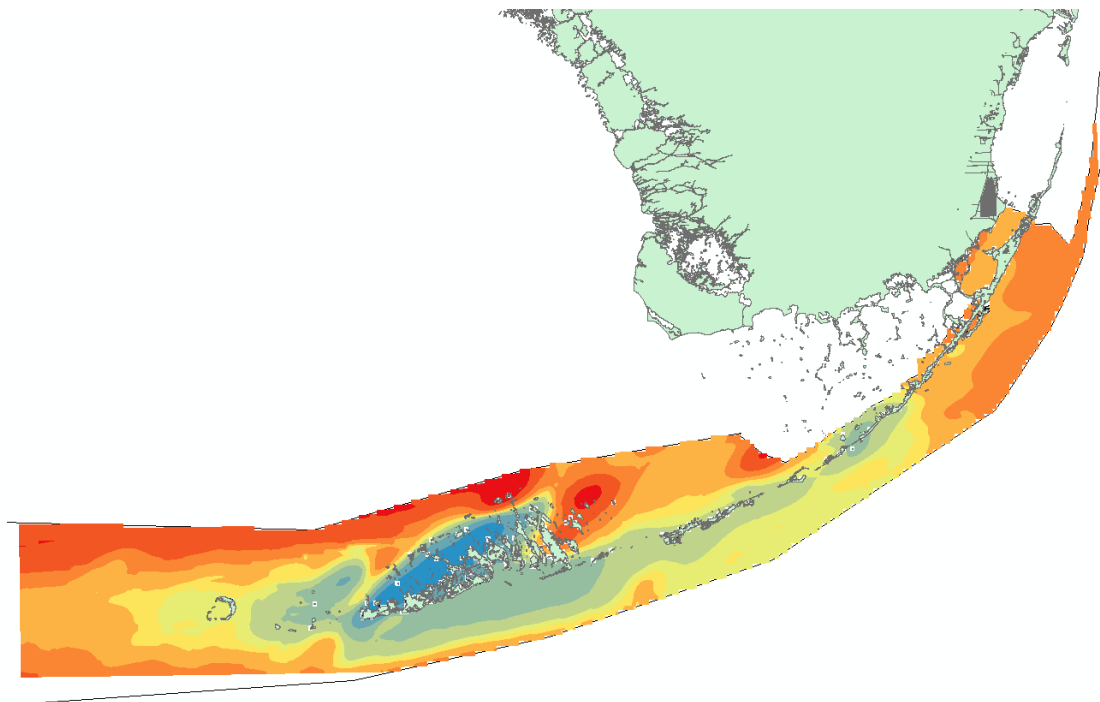


Figure 10: Prediction map of mean benthic dissolved oxygen in the FKNMS

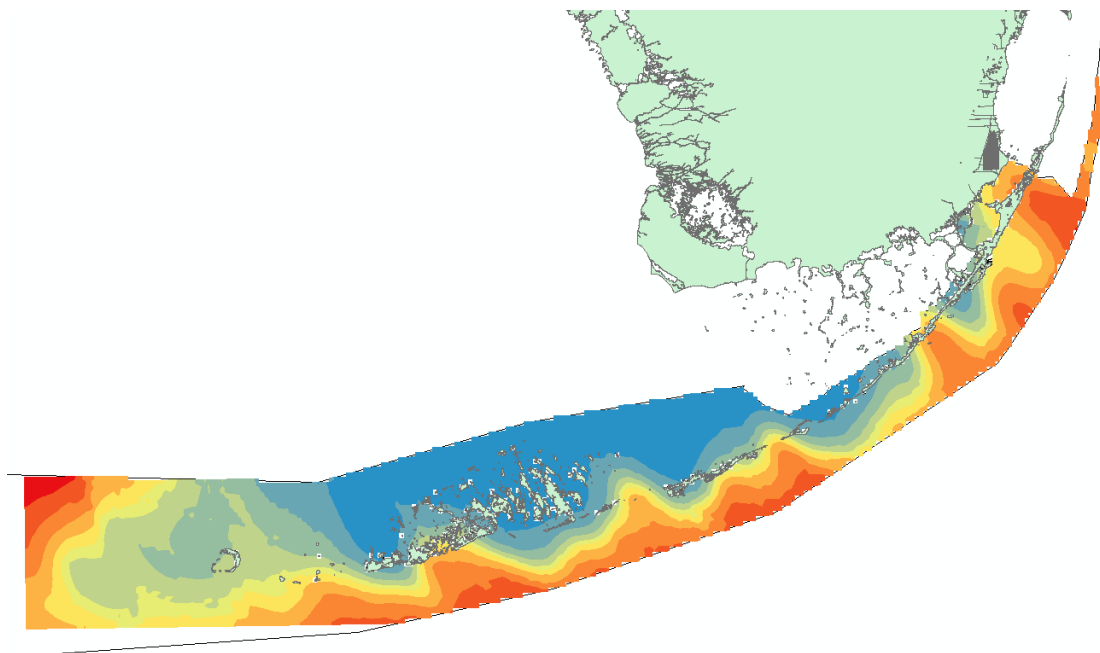


Figure 11: Prediction map of benthic dissolved inorganic nitrogen in the FKNMS

3: Correlation analysis

My goal for this project was to find which of these three environmental variables spatially correlate with total coral disease and recent mortality. One way to do this is to take the kernel density rasters of disease and mortality and create contours to overlay on the temperature, DO and DIN rasters. This will allow me to visualize which environmental variables appear to spatially overlap.

Disease contour:

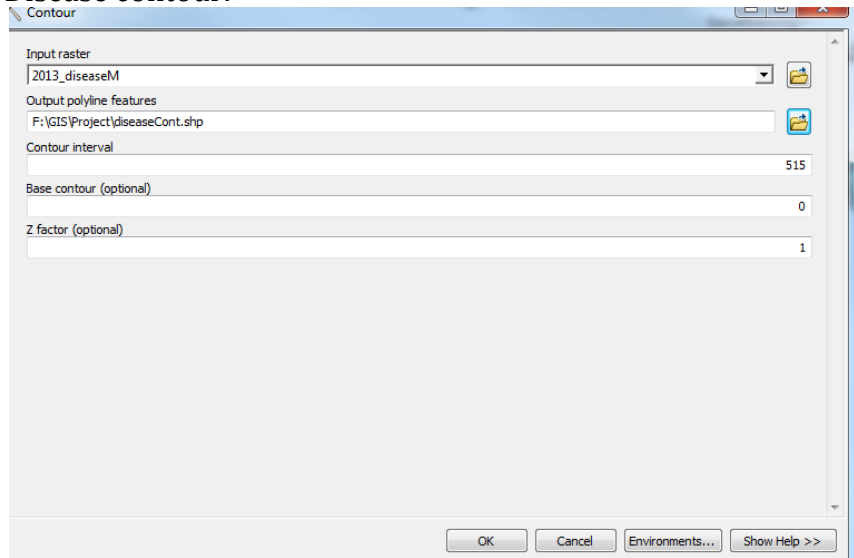


Figure 12: Using the Contour tool to generate 10 evenly spaced contours that show the areas with highest coral disease

Mortality Contour:

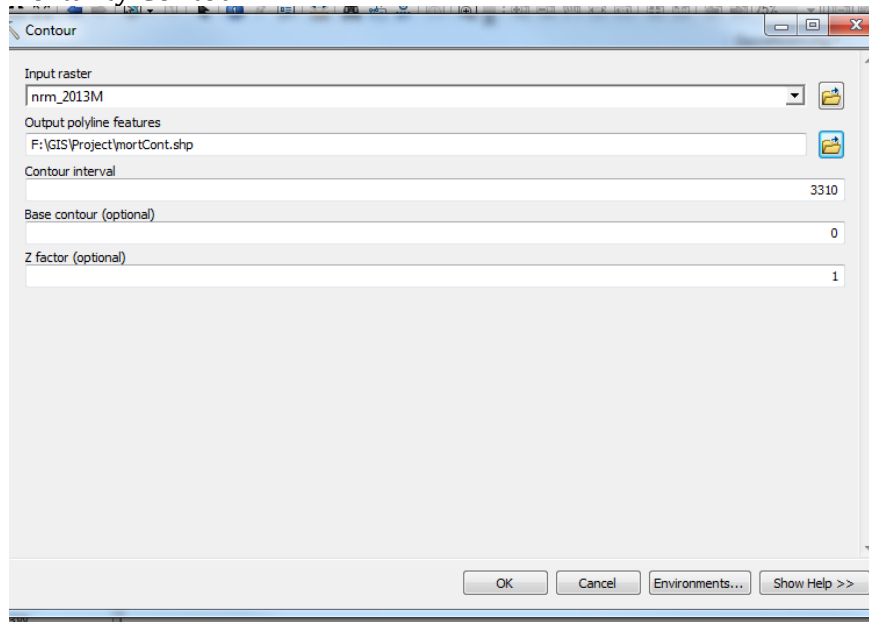


Figure 13: Using the Contour tool to generate 10 evenly spaced contours that show the areas with highest recent coral mortality

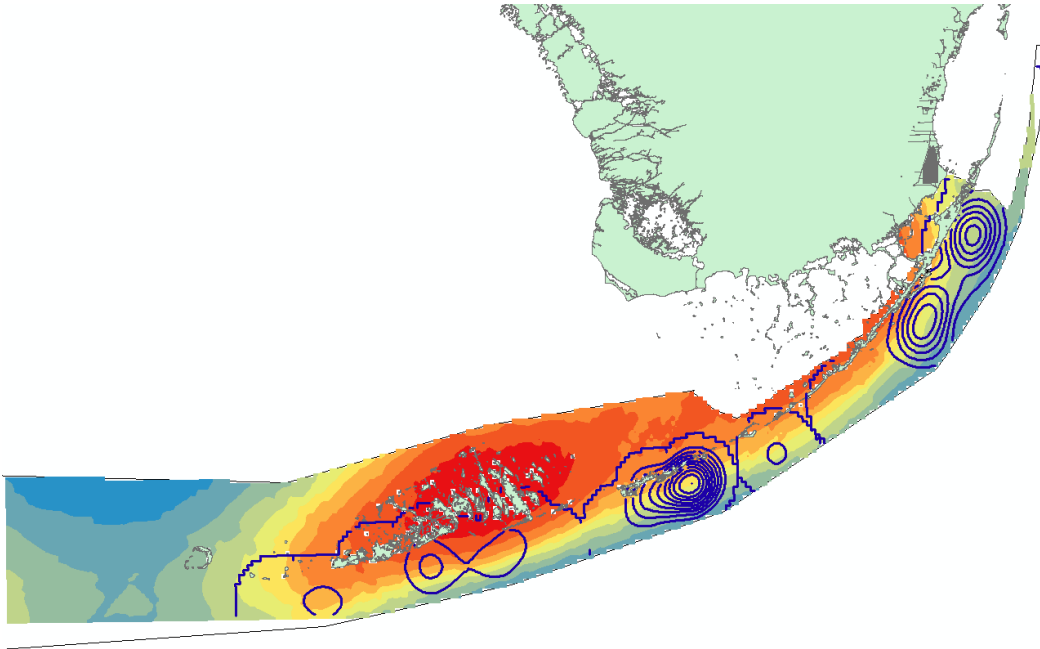


Figure 14: Contour of diseased coral density overlaid with maximum benthic temperature

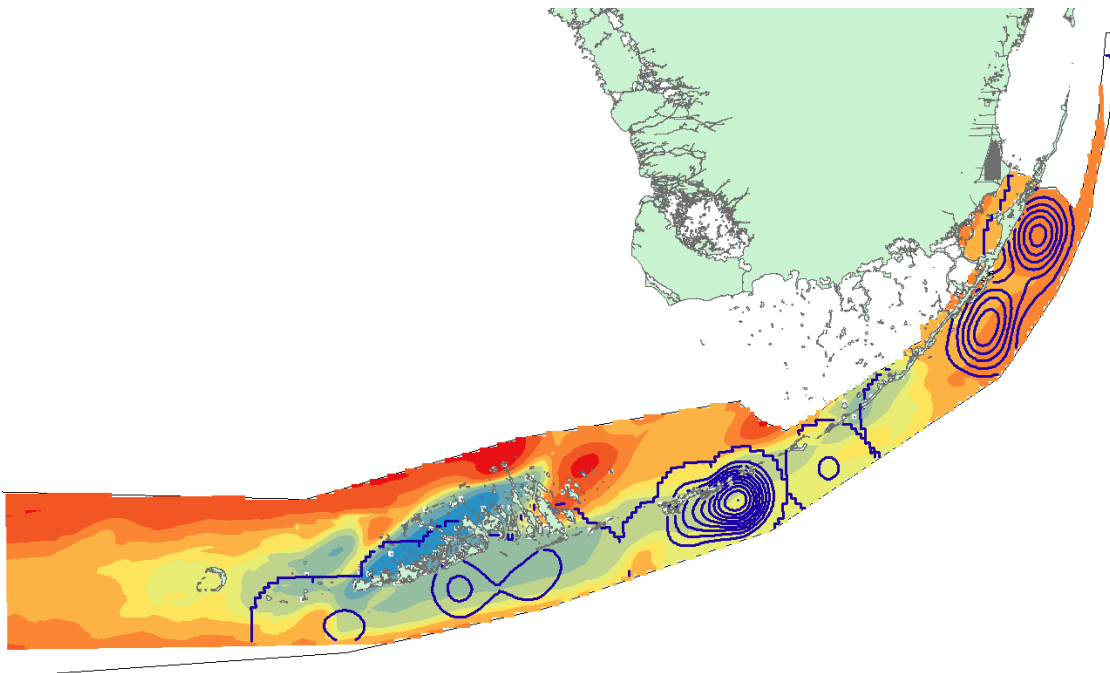


Figure 15: Contour of diseased coral density overlaid with mean benthic dissolved oxygen

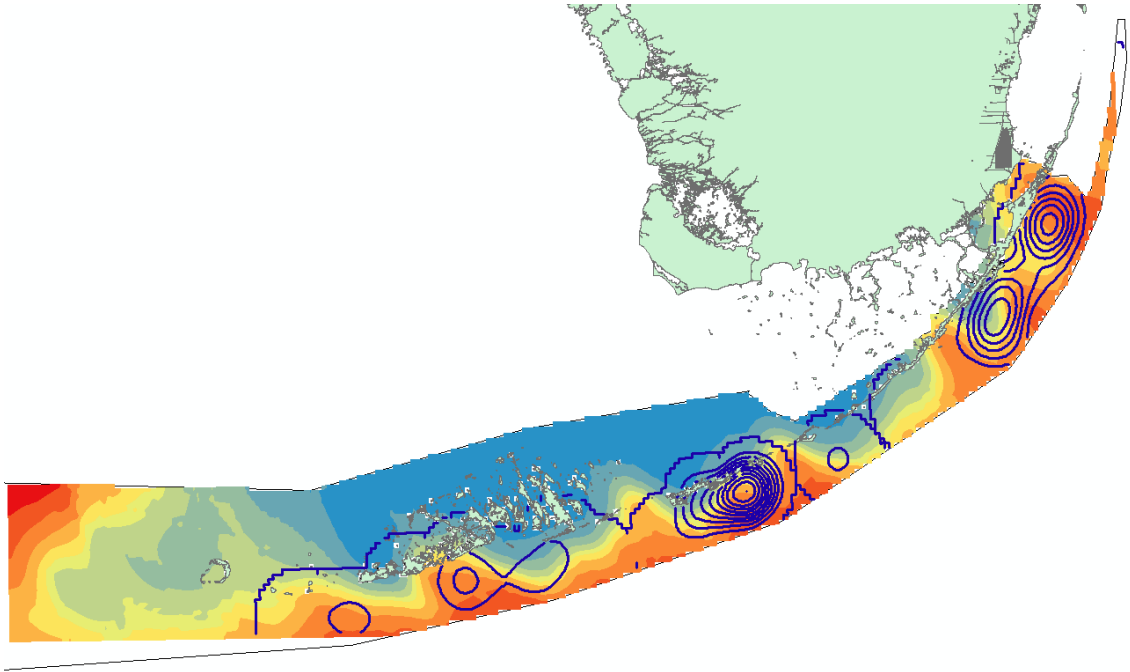


Figure 16: Contour of diseased coral density overlaid with mean benthic dissolved inorganic nitrogen

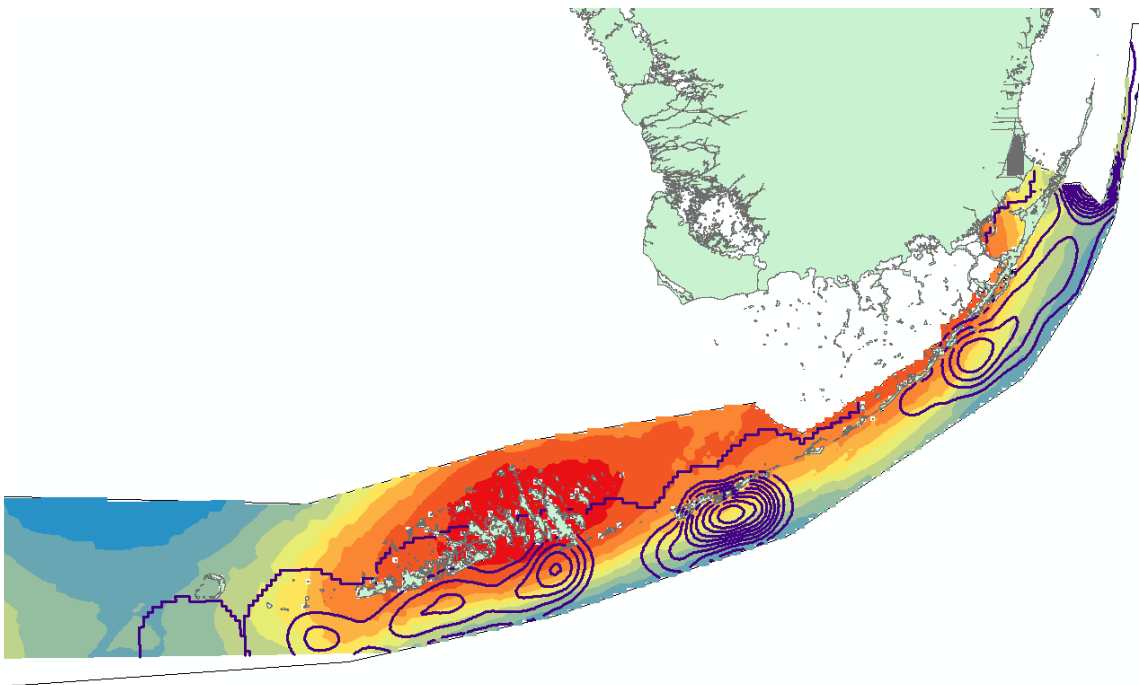


Figure 17: Contour of recent coral mortality overlaid with maximum benthic temperature

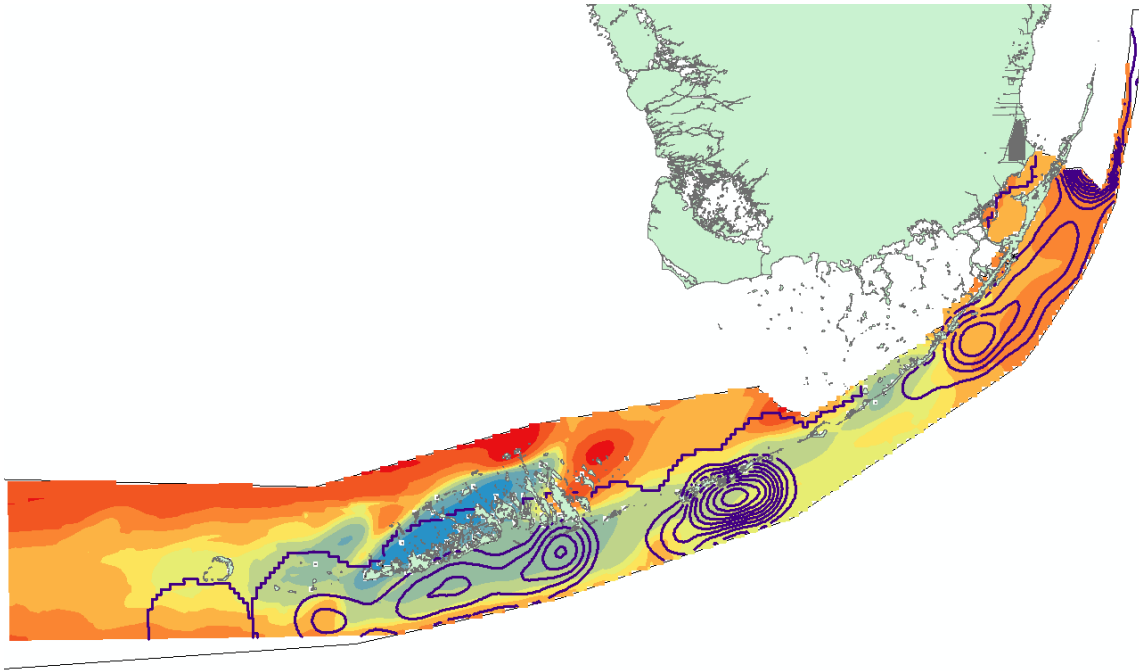


Figure 18: Contour of recent coral mortality overlaid with mean benthic dissolved oxygen

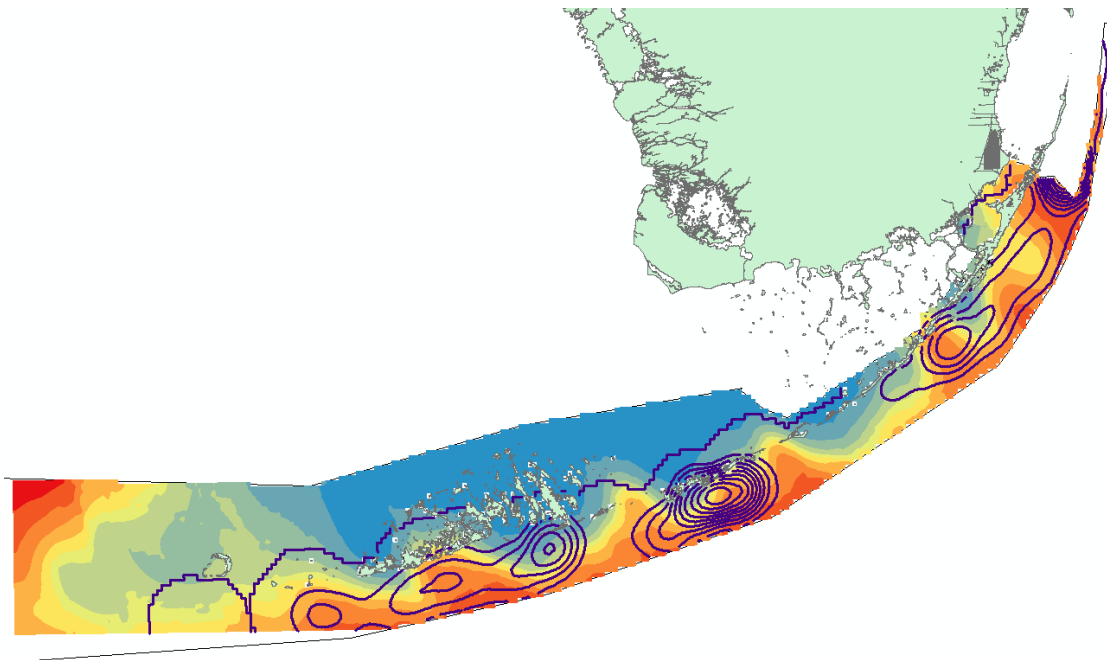


Figure 19: Contour of recent coral mortality overlaid with mean benthic dissolved inorganic nitrogen

Results and Conclusions:

This analysis shows that none of the three environmental variables modeled here fully explain the spatial extent of coral disease and recent mortality in the FKNMS. However, there do appear to be variables that correspond to these types of coral decline more than others. For example, patterns of dissolved inorganic nitrogen seem to correlate with patches of coral disease in the upper and lower keys (Figure 16). Maximum summer temperature varies along the coast where temperatures are higher closer to shore and lower far from shore. This pattern does not correlate with the density of coral disease (Figure 14). Dissolved oxygen appears to vary in a different direction than temperature where lower levels of dissolved oxygen exist in the lower keys where the upper keys appear to have higher levels of dissolved oxygen. This pattern does not correlate with the density pattern of diseased corals (Figure 15). However, future analysis should find a way to include more variables, weight and combine them into a sophisticated hazard analysis. The prevalence of coral disease is a complex process that is not likely attributed to a single measurement of water quality. The spatial pattern of coral disease we see in the FKNMS is likely a combination of numerous abiotic and biotic factors.

Recent coral mortality also seems to correlate with the high levels of dissolved inorganic nitrogen (Figure 19) and to some extent, low levels of dissolved oxygen in the lower keys (Figure 18). Again, the spatial pattern of maximum benthic temperature does not correlate at all with recent mortality or disease in the system. This could be explained by the ability of coral populations to adapt at very local scales to this environmental gradient. However, future analysis should consider modeling numerous other variables.

One source of uncertainty in this analysis could be from data sampling. Sampling data in both the SERC and FRRP datasets are available due to a massive effort involving numerous researchers and institutions. This inherent sampling bias could affect the modeling presented here. It's also possible that there is a lot of missing data in both these large datasets. Any missing data that was entered as a zero value could have dramatically affected the prediction and density maps generated in this project.

- Bruno, J. F., Selig, E. R., Casey, K. S., Page, C. a, Willis, B. L., Harvell, C. D., ... Melendy, A. M. (2007). Thermal stress and coral cover as drivers of coral disease outbreaks. *PLoS Biology*, 5(6), e124. doi:10.1371/journal.pbio.0050124
- D'Angelo, C., & Wiedenmann, J. (2014). Impacts of nutrient enrichment on coral reefs: new perspectives and implications for coastal management and reef survival. *Current Opinion in Environmental Sustainability*, 7, 82–93. doi:10.1016/j.cosust.2013.11.029
- Fabricius, K. E. (2005). Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Marine Pollution Bulletin*, 50(2), 125–46. doi:10.1016/j.marpolbul.2004.11.028
- Gardner, T. a, Côté, I. M., Gill, J. a, Grant, A., & Watkinson, A. R. (2003). Long-term region-wide declines in Caribbean corals. *Science (New York, N.Y.)*, 301(5635), 958–60. doi:10.1126/science.1086050
- Kenkel, C. D., Goodbody-Gringley, G., Caillaud, D., Davies, S. W., Bartels, E., & Matz, M. V. (2013). Evidence for a host role in thermotolerance divergence between populations of the mustard hill coral

(*Porites astreoides*) from different reef environments. *Molecular Ecology*, 22(16), 4335–48.
doi:10.1111/mec.12391

Krivoruchko, K. (n.d.). Introduction to Modeling Spatial Processes Using Geostatistical Analyst, 1–27.

Lirman, D., Schopmeyer, S., Manzello, D., Gramer, L. J., Precht, W. F., Muller-Karger, F., ... Thanner, S. (2011). Severe 2010 cold-water event caused unprecedented mortality to corals of the Florida reef tract and reversed previous survivorship patterns. *PLoS One*, 6(8), e23047.
doi:10.1371/journal.pone.0023047