**Phytoplankton Community Shifts off the Coast of Cape Canaveral, FL**

Ben Stelling

**Methods**

The Bureau of Ocean Energy Management (BOEM) is using an area off the coast of Cape Canaveral, FL as the sand resource to replenish areas of the Northeast U.S. that were eroded by Hurricane Sandy (Weisskohl, 2013). As a part of BOEM’s effort to monitor how marine life recovers from dredging activity, phytoplankton samples were taken at 24 sites around the dredging sites. At these sampling sites, water samples were taken at two depth levels: integrated and bottom. The integrated water sample was taken using a 3-meter, integrated pole sampler that retrieved the sample of the top 3 meters of the water column. Five of these samples were taken, mixed in a bucket, and then 100 mL of this mixture was placed into an amber glass bottle and preserved with acidic Lugol’s solution to a final concentration of 0.1%. For the bottom samples, 5 samples were taken using a Niskin bottle allowing it to sink to the seafloor, pulled up off the sediment slightly, opened to allow water to flow in, then closed and pulled to the surface. These samples were also mixed in a bucket, and 100 mL was placed in an amber bottle and preserved with acidic Lugol’s solution as stated above. These samples were then stored at ambient temperature (72oF/22oC) for analysis.

The phytoplankton were counted using the Utermohl method (Utermohl 1931 and 1958; Edler & Elbrächter, 2010) and cyanobacteria counts were performed using epifluorescence microscopy. Samples were collected on five dates: October 17th, 2013 and February 3rd, May 22nd, July 30th, September 10th, 2014. The ultimate goal of this investigation is to see if there are notable changes in the community structure with differing depths and/or seasons. The coastal ocean, which is still affected by weather patterns and run-off, has a small water residence time. So this could mean that fluctuations in species composition vary explicitly due to seasonal shifts in light and temperature.

The data consisted of a site by species/size category matrix where the categories covered multiple size classes of common genera totaling 333 categories. The counts were also converted to cells mL-1 for comparison. Because of the differing sizes, this leads to abundances spanning many orders of magnitude and with many 0 values for rarer species. Euclidian distance lacks accuracy in these types of datasets because it can show more similarity between samples with no shared species than samples with similar compositions at differing densities (Orloci, 1978). The Bray-Curtis index is better suited for these types of datasets, but suffers in accuracy with large differences in orders of magnitude (Bray & Curtis, 1957). The Bray-Curtis index gives values from 0 to 1 for completely identical and no species shared respectively, and is controlled by more abundant species. The high cell densities for small species like Prasinophytes or cyanobacteria, therefore, will skew the results—overshadowing effects from important large species. For this reason, the counts were transformed (Equation 1) where *x* is the original values. This transformation decreases the intense effect of large values and allowed the values to be in the same order of magnitude**.** This made comparisons based on differences in uncommon species possible. The *vegdist*function in the vegan package in R 3.2.3 (Oksanen, et al., 2016; R core team, 2015)was used to calculate the dissimilarity matrix.

(1)

Once the data were transformed, a z-standardization was performed to equalize differences in variance among the different species. A k-means clustering algorithm (Hartigan & Wong, 1979) (function: *kmeans;* package: *stats*) was used on the z-standardized data to separate the data into groups to be analyzed for ecological significance. The within-group sum of squares and silhouette width of the k-means results were compared to choose the number of clusters to be analyzed.

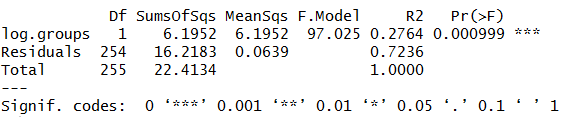
Non-metric multidimensional scaling (NMDS) (Kruskal, 1964; Kruskal and Wish, 1978**)** was used as the ordination technique the density data does not meet the assumptions that may be needed other ordination techniques (principle coordinates analysis) that assume an underlying relationship between the dissimilarities and component axes. NMDS does not call for assumptions like these because it ranks the data points distances making it flexible for multiple types of data and appropriate for my dissimilarities (Salmaso, 1996; Triantafillos & Economou-Amilli, 1991).

The significance of the groupings was tested by permutational multivariate analysis of variance (PerMANOVA), which uses the distance matrix (dissimiliarities) and permutations to test the differences between clusters (Anderson, et al., 2005) (function: *adonis*; package: *vegan*). This can test the significance of the k-means clusters and see if there are any community change patterns in the dataset. After the k-means clusters are represented, I visualized the seasons and water depths on the ordination plot to compare the clustering results to potential patterns.

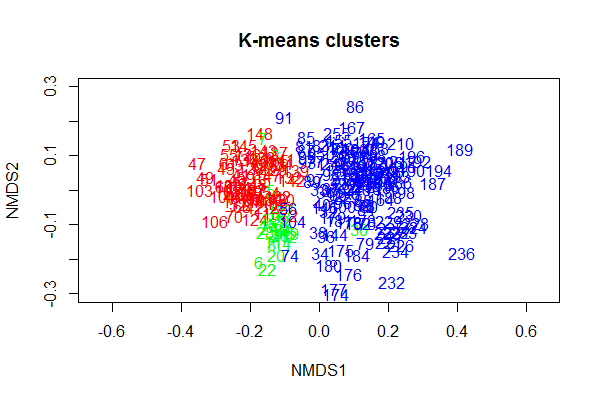
**Results**

3 clusters maximized the variance explained, while keeping a high silhouette width, so 3 clusters were chosen for the analysis (Figure 1). It is evident that the data could have gone into 2 or 3 clusters seeing how the red and blue clusters have a clear separation in between them with very little overlap. The green cluster is more interesting taking up a smaller area. This appears to have a tight population and may align with a specific species composition. The two larger groups are separated along the NMDS 1 axis which seems to signify the two majors clusters in this coastal region. When coloring the same plot with the seasonal groups (Figure 2) there appears to be some seasonality in the clusters. The green cluster appears to consist of mostly fall samples, which could symbolize a transitionary community with the larger winter-spring group. The largest group contains all of the summer values (Both September and July sampling dates), splits the spring values and contains some winter values as well—which is unexpected. The k-means clusters were significant as tested by the PerMANOVA (Table 1). The NMDS ordination also appeared to give a fair representation of the dissimilarities between the values (stress = 0.177).

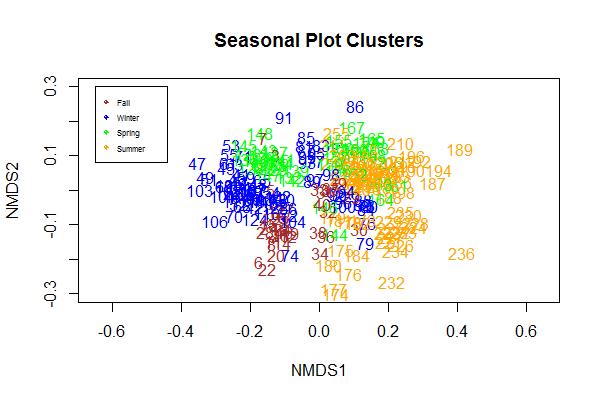
**PerMANOVA Results**



**Table 1:** Results of PerMANOVA test for differences among the three clusters. F-Model gives the calculated F-value and Pr(>F) denotes the probability of obtaining a value higher than modelled F-value.



**Figure 1**: k-means clusters results for all sites, given arbitrary number values. 3 clusters were obtained with the large right blue cluster encapsulating the majority of the samples, a tight small cluster all with similar species in green as an apparent subset of the large red cluster.



**Figure 2**: Seasonal variations in the sites to serve as a comparison to the k-means clustering. The majority of winter and some spring appear in the left large cluster, while fall makes up most of the small tight cluster. The large right cluster contains all of the summer, and parts of the other seasons.

**Discussion**

The results showed promise in the investigation of phytoplankton community dynamics on the Atlantic coast of Florida. It illustrates how the ocean community structure is not as volatile as some lake, bays, or estuarine systems that are at the whim of large stochastic weather events and large water residence times. In these systems, the input can be long lasting and often shift the community drastically. The open ocean is generally affected by these events but on a much smaller time scale because nutrients will generally flow away from any one site quickly. The results showed some seasonality in the data, but the overlap of winter data into the summer cluster is suspect. This paper being a preliminary investigation, the sites at which these data overlap may show some interesting patterns if looked at more closely. There may be other patterns at work like temperature of the water column, but without more data it is difficult to make hypotheses. In further sampling, the seasonality of sampling dates will be taken into account (i.e. attempting to sample around the summer and winter solstice). This will give a better picture of the light differences experienced by the different communities. Another thing to consider is how the data in categorized. Biovolumes among the species in this dataset vary widely and their role in the food web varies as well. Though most are primary producers, differentiation between large and small species may be useful in illuminating patterns that cell densities would not uncover. Also, including more abiotic factors into the model may provide a clearer picture of what is causing shifts in oceanic phytoplankton communities. Knowing that there is some shift in community structure is interesting, but revealing what drives this change could prove more useful.

**Literature Cited**

1. Allen, T. F. H., & Koonce, J. F. (1973). Multivariate approaches to algal stratagems and tactics in systems analysis of phytoplankton. Ecology, 54(6), 1234-1246.
2. Anderson, M. J., Gorley, R. N., & Clarke, R. K. (2005). Permanova. Permutational multivariate analysis of variance, a computer program. Department of Statistics, University of Auckland, 24.
3. Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. Ecological monographs, 27(4), 325-349.
4. Edler, L., & Elbrächter, M. (2010). The Utermöhl method for quantitative phytoplankton analysis. Microscopic and molecular methods for quantitative phytoplankton analysis, 110.
5. Hartigan, J. A. and Wong, M. A. (1979). A K-means clustering algorithm. Applied Statistics 28, 100–108
6. Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29(1), 1-27.
7. Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling (Vol. 11). Sage.
8. Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2016). vegan: Community Ecology Package. R package version 2.4-1. <https://CRAN.R-project.org/package=vegan>
9. Orlóci, L. (1978). Multivariate analysis in vegetation research. Springer.
10. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
11. Salmaso, N. (1996). Seasonal variation in the composition and rate of change of the phytoplankton community in a deep subalpine lake (Lake Garda, Northern Italy). An application of nonmetric multidimensional scaling and cluster analysis. Hydrobiologia, 337(1-3), 49-68.
12. Taylor, W. D., & Wetzel, R. G. (1988). Phytoplankton community dynamics in Lawrence Lake of southwestern Michigan. Archiv für Hydrobiologie. Supplementband. Monographische Beiträge, (81), 491-532.
13. Triantafillos, P. T., & ECONOMOU-AMILLI, A. (1991). Evaluation of the phytoplankton variation in Lake Trichonis (Greece) by means of multivariate analysis. Memorie Dell'Istituto Italiano Di Idrobiologia Dott. Marco De Marchi, 48, 99.
14. Utermöhl H (1931) Neue Wege in der quantitativen Erfassung des Planktons. (Mit besonderer Berücksichtigun des Ultraplanktons). Verh. Int. Ver. Theor. angew. Limnol. vol. 5, no. 2. p. 567-596
15. Utermöhl H (1958) Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. Mitt int. Verein. theor. angew. Limnol. 9: 1-38
16. Weisskohl, M (2013, October 21). BOEM Announces $2 Million in Awards to the University of Florida and the Navy for Hurricane Sandy Environmental Monitoring Along the Southeast Coast. Bureau of Ocean Energy Management (BOEM). Retrieved from <https://www.boem.gov>