North Atlantic Ocean Chlorophyll a Variability

Yizhou Aria Jiang

April 18, 2019

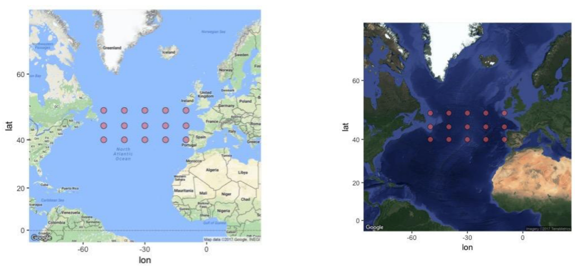
1 Executive Summary

The main goal of this paper is to examine the variability of Chlorophyll a in the North Atlantic Ocean. Wavelet analysis, time series decomposition with moving average and PCA analysis are performed on pre-processed data. The main conclusion regarding data analysis method is that time series decomposition is very effective for eliminating seasonal influence and showing the underlying trend in data. In terms of actual findings, there are two conclusions. First, longitude seems to be a better metric in explaining how data from different extraction points are correlated. Second, Chlorophyll, temperature and salinity vary together. Specifically, Chlorophyll and temperature vary in the same direction while Chlorophyll and salinity vary in opposite directions.

2 Background

The data used for analysis contains time series information across fifteen extraction points in the North Atlantic Ocean on Chlorophyll, salinity and temperature. All data is measured at sea surface. The longitude of the extraction points range from -50 to -10 with an increment of 10 while the latitude of the extraction points range from 40 to 50 with an increment of 5. The extraction points are plotted on both terrain and satellite map in figure 1.

The salinity data was measured daily while both temperature and chlorophyll data were measured every eight days. As shown in figure 2, the length of time series on each variable varies from four to eighteen years. While performing PCA analysis among variables, only the time span during which values of all three variables are recorded



(a) Terrain

(b) Satellite

Figure 1: Data Extraction Points on Maps

is considered. Besides, eight-day averages are calculated for salinity data so that all three variables have the same frequency of measurement.

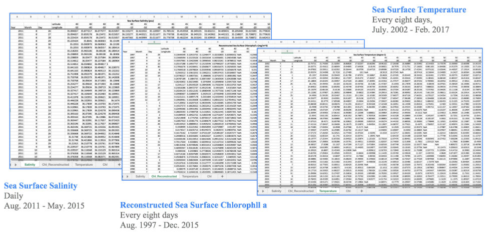


Figure 2: Data Preview

3 Data Analysis

Of interest is Chlorophyll's spatial variability as well as its variability in the temperature-salinity space without seasonal influence. The following two subsections will address the two topics separately.

3.1 Spatial Variability

As a first step, wavelet analysis is performed on Chlorophyll times series across all points. Figure 3 shows wavelet power spectrum at the extraction point (40, -50) as well as a plot showing reconstructed times series superimposed on the original one. As mentioned in the background section, chlorophyll data is measured every eight days, which then logically is used to define the length of a period. Thus, there are forty-six periods in a year, twenty-three periods in six months and eleven point five periods in one season. As shown in the wavelet power spectrum, at forty-six, twenty-three and eleven point five the plot shows the highest power. In other words, the data shows strong cycles every three, six and twelve months.

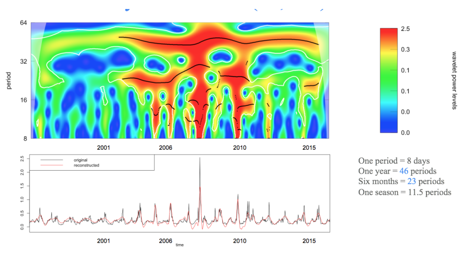


Figure 3: Wavelet Analysis

Next, the same time series of Chlorophyll at the extraction point (40, -50) is decomposed into seasonal, trend, and irregular components by moving averages [1]. It is assumed that 1) the amplitude of the seasonal effect is the same each cycle, and 2)the additive model is used instead of the multiplicative model. During the decomposition process, the trend component is first determined using a moving average. The trend component is then removed from the time series. Subsequently, the seasonal component is computed by averaging each period over all periods. After centering the seasonal component and removing it from the time series, the irregular/error component is what is left. Figure 4 shows the decomposition of point (40, -50). By comparing the trend graph with the reconstruction plot in figure 3, one can see that decomposition seems to be able to remove more seasonal influence.

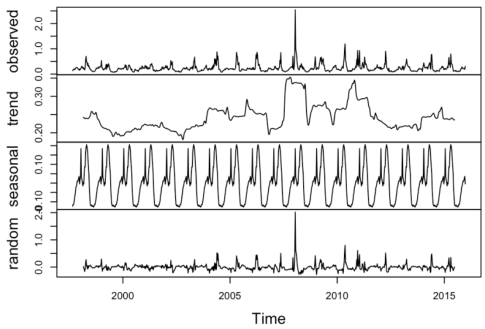


Figure 4: Decomposition of a Time Series

After computing Chlorophyll trends at all extraction points, the Chlorophyll variability across all points at every point in time is computed and plotted in figure 5 (left). In figure 5 on the right is a plot of spatial variability with seasonal influence, that is, variability computed from the given data instead of trends. Note that the scale of variability in the plot on the left is a lot smaller than that in the plot on the right because seasonal influence has plenty of variability and is removed from the left plot. However, the plot on the left shows an upward trend in Chlrophyll spatial variability while the plot on the right shows a flat general trend.

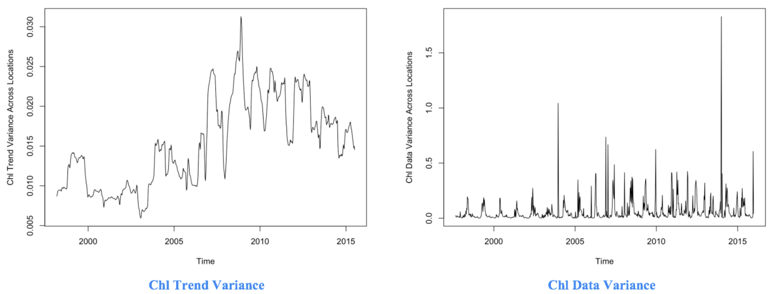


Figure 5: Time Series of Chlorophyll Variability Across Locations

As a next step, PCA analysis is conducted among Chlorophyll trends of all locations. As shown in figure 6, the first two principle components each explains significantly more variability in the data than all other principle components. Together, the first two components roughly explain 55 percent of total variability. Figure 7 shows the loadings of all points on each of the first five principle components.

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Standard deviation	2.136	1.6065	1.1765	1.04118	0.89688	0.79791	0.75043	0.71086	0.58770	0.5249	0.34745	0.28428	0.23509
Proportion of Variance	0.351	0.1985	0.1065	0.08339	0.06188	0.04897	0.04332	0.03887	0.02657	0.0212	0.00929	0.00622	0.00425
Cumulative Proportion	0.351	0.5496	0.6561	0.73944	0.80132	0.85029	0.89361	0.93248	0.95905	0.9802	0.98953	0.99575	1.00000

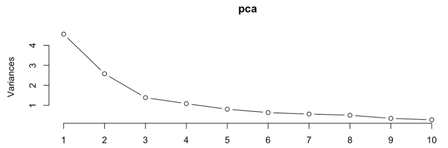


Figure 6: PCA Analysis Among All Locations on Chlorophyll

	PC1	PC2	PC3	PC4	PC5
40, -50	-0.06803905	0.456928582	0.30592996	-0.077354047	0.0001502238
40, -40	-0.30196652	0.275426202	0.31844358	0.094266434	-0.1248378977
40, -30	-0.23196612	-0.242052930	0.43943978	-0.384902248	-0.3276185925
40, -20	-0.27304689	-0.268598707	-0.06552598	-0.239348292	-0.4962019332
40, -10	-0.19449334	0.379010413	-0.14098436	-0.307391346	-0.1328747316
45, -40	-0.22138716	-0.005183285	0.07913541	0.773861045	-0.2920189479
45, -30	-0.40022000	-0.219686699	0.08562420	0.086429867	0.1135027191
45, -20	-0.39193324	-0.241360218	-0.17212778	0.001635159	0.0211972621
45, -10	-0.25854623	0.211695348	-0.54263904	-0.082648243	-0.0560821290
50, -50	-0.11001310	0.489966119	-0.03426070	-0.090011503	-0.0972934263
50, -40	-0.30599817	0.121941654	[illegible]	0.183544306	-0.0242733654
50, -30	-0.28245601	0.111520682	[illegible]	0.060091842	0.4259143559
50, -20	-0.35047265	-0.149133111	-0.10378589	-0.163209523	0.5656843412

Figure 7: Loadings of Variables on Principle Components

The biplot in figure 8 visualizes the loadings of all locations in the first two principle components. In the biplot, the points correspond to the PC1 and PC2 scores of each observation; the arrows represent the correlation of the variables (points) with PC1 and PC2; the circle indicates the theoretical maximum extent of the arrows. By analyzing the spatial distribution of variables that correlate with one another, one can conclude that points with the same or similar longitudes tend to correlate with one another while points with the same or similar latitudes don't seem to be more likely to correlate with one another. Thus, longitude seems to be a better metric in explaining how data from different extraction points are correlated.

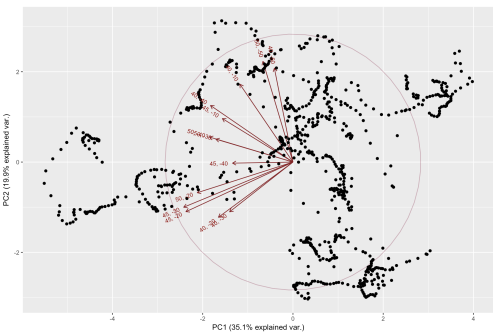


Figure 8: PCA1 & PC2 Biplot

3.2 Variability in temperature - salinity space

In this subsection, PCA analysis is first conducted on three variables - temperature, salinity and Chlorophyl - across all points. Figure 9 and 10 show the details of the PCA analysis conducted at point (40, -30). In the biplot shown in figure 10, data points are grouped by seasons and are colored differently. It is obvious that data observed in winter and summer demonstrate very different characteristics. Data observed in winter gather at low temperature and high Chlorophyll while data observed in summer suggest the opposite condition. On the other hand, data observed in both spring and fall fall in between those observed in winter and summer. Thus, it is reasonable to suspect seasonal influence in the PCA analysis.

	PC1	PC2	PC3
Standard deviation	1.3542	0.9976	0.41348
Proportion of Variance	0.6113	0.3317	0.05699
Cumulative Proportion	0.6113	0.9430	1.00000

	PC1	PC2	PC3
Temp	-0.6942778	-0.1826758	0.6961379
Chl	0.7060117	0.0149292	0.7080428
Sal	-0.1397351	0.9830598	0.1186063

Figure 9: PCA Analysis Without Seasonal Adjustment

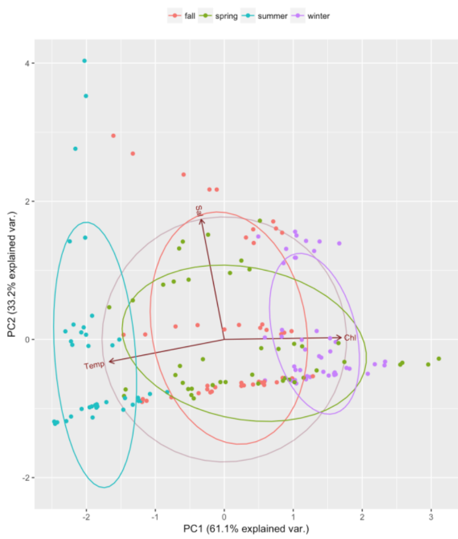


Figure 10: PCA1 & PC2 Biplot Without Seasonal Adjustment

Naturally, the next step is to conduct the same PCA analysis, but on trends across all variables and all points. Figure 11 and 12 show the results of the PCA analysis at the same point (40, -30), which turns out to be very different from that of previous ones. There are two main differences. First, salinity has very high loading (-0.64) compared with its previous loading (-0.14). Second, while previously temperature and Chlorophyll vary in opposite directions, they vary in the same direction in the analysis on trends.

	PC1	PC2	PC3
Standard deviation	1.4478	0.8203	0.48051
Proportion of Variance	0.6987	0.2243	0.07696
Cumulative Proportion	0.6987	0.9230	1.00000

	PC1	PC2	PC3
Chl	0.5349176	0.72973857	-0.4258459
Temp	0.5492514	-0.68332523	-0.4810297
Sal	-0.6420172	0.02341477	-0.7663326

Figure 11: PCA Analysis With Seasonal Adjustment

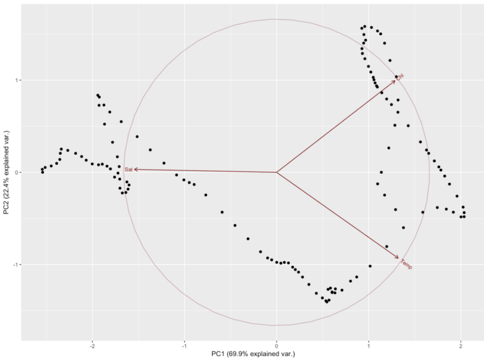


Figure 12: PCA1 & PC2 Biplot With Seasonal Adjustment

4 Conclusion

From the data analysis section, one can conclude that time series decomposition is very effective for adjusting data for seasonal influence and letting the underlying trend contained in the data show. In terms of more specific findings, there are two conclusions. First, when it comes to spatial variability, longitude seems to be a better metric in explaining how data from different extraction points are correlated than latitude. Second, Chlorophyll, temperature and salinity vary together. Specifically, Chlorophyll and temperature vary in the same direction while Chlorophyll and salinity vary in opposite directions.

5 Reference

[1]M. Kendall and A. Stuart (1983) The Advanced Theory of Statistics, Vol.3, Griffin. pp. 410–414.