Analyzing acidity in the California Current System (CCS)

Namrata Kolla

EAS 4480 | Yuhang Wang

Environmental Data Analysis | Final Project

Abstract

This paper studies how pH changes in the California Current System (CCS) across two major regimes—before 1980 and after 1980—in the model dataset from NOAA's Geophysical Fluid Dynamics Laboratory. It conducts autocorrelation analysis, least-squares spectral analysis, linear and higher order regressions, and chi-squared tests on residuals to determine the strength of cycles and trends in each regime. From the first regime to the second, the cycles become longer and less frequent and the trend becomes more negative and stronger. In its entirety, this analysis supports the hypothesis that natural impacts on ocean pH in the CCS were stronger prior to 1980 and anthropogenic impacts were stronger after 1980.

Introduction

Ocean acidity has been a major focus of study for climate scientists due to its effects on physiological processes such as respiration, calcification, photosynthesis, and reproduction (Administration, 2015). Altering the ability of marine organisms to undergo these processes has a direct impact on primary productivity, fishing profits, eco-tourism, and the overall health of the oceans, which consequently have a direct impact on human beings. Changes in acidity can happen due to numerous reasons, the major one being fluctuating levels of atmospheric carbon dioxide (CO2). Gaseous CO2 reacts with water to

create carbonic acid, whose deprotonations make up the core chemical processes behind ocean acidification. This study examines a model data set of pH from 1861 to 2100 in the California Current System (CCS), the eastern edge of the North (or Central) Pacific gyre, to see if the data contains distinguishable trends or cyclic properties. This is an important question because it can help discern whether natural or anthropogenic processes have a greater impact on acidity in the CCS. The assumptions being made are that the ocean is a major sink for carbon dioxide emitted by human activities and that the CCS will experience increased acidity immediately or after some delay due to ocean circulation.

Figure 1. Location of the California Current System (CCS) (Turner, 2011)

The original data set can be found on the website for the Geophysical Fluid Dynamics Laboratory in the National Oceanic and Atmospheric Administration (NOAA) (2014). It is an Earth System Model that uses the Generalized Ocean Layer Dynamics (GOLD) code base (NOAA, 2014). This code base improves upon previous code bases by better incorporating vertical mixing (NOAA, 2014). While this is a model of pH, it comes

very close to collected experimental data and therefore still has a large range of idealized and realistic applications. To give a few examples, this model has already been successfully used to predict the impacts of the Deepwater Horizon oil spill, the effects of wave-driven mixing on global circulation, and the large-scale dynamics of overflow (NOAA, 2014).

Before conducting my analysis on CCS acidity, the pH data had to be prepared in a few ways. First, the pH data for the CCS region was extracted and saved in a separate structure to minimize the size of the data file and the amount of time it takes to execute multiple iterations. Second, the pH was calculated along isopycnal 26.5 $kg*m^{-3}$ because water parcels prefer to move along layers of equal density, and isopycnals are a more reliable way of comparing pH data over time (Talley, 2000). This particular isopycnal was chosen because it is in the subsurface where plankton preside and it is deep enough that it never surfaces, so any influx of carbonic acid must come from adjoining currents, not direct interactions with the atmosphere. The third step in preparing the data was deriving and using pH anomalies for all of the calculations in the place of pH. pH anomalies describe how much pH values deviate from the baseline pH—the average pH over the complete time span—and help minimize problems that occur when outliers are removed (i.e. anomalies have lower sensitivity) (Center, 2015).

Analysis

The main diagram this report analyzes is Figure 2, a time series of mean pH anomalies in the CCS from 1861 to 2100. Values above 0 connote pH being higher than average (more basic than usual) and values below 0 connote pH being lower than average (more acidic than usual).

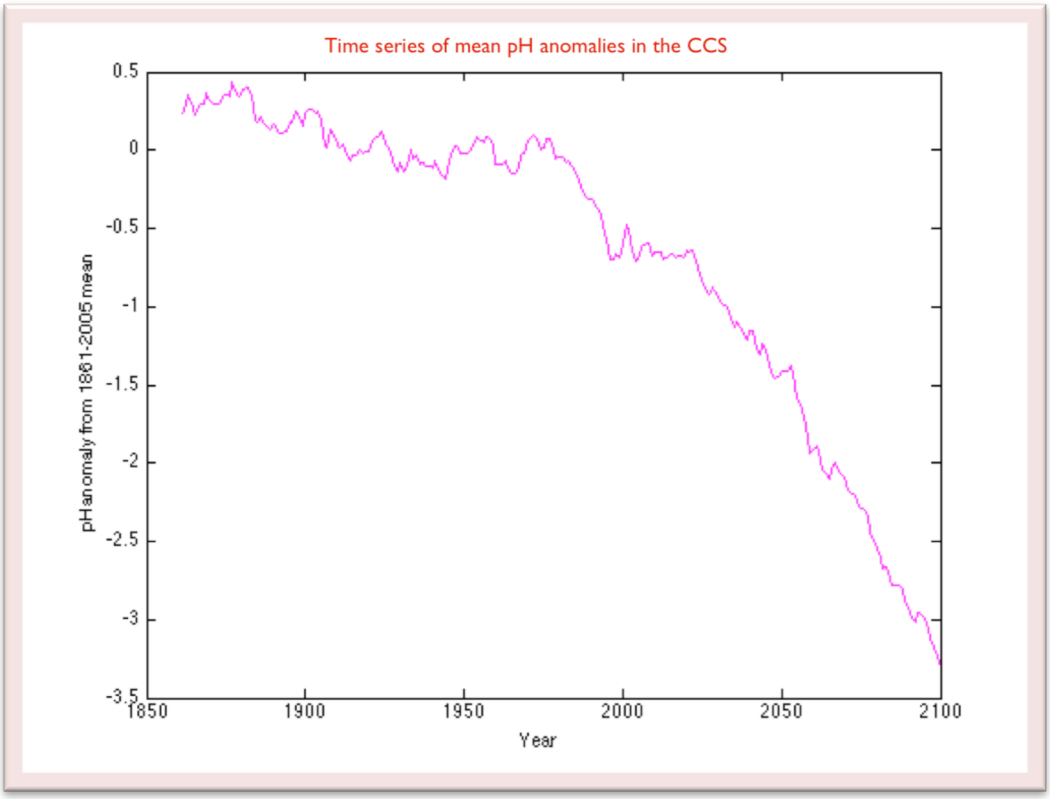

Figure 2. Time series of mean pH anomalies in the CCS

It is clear that pH anomalies become increasingly negative over time, meaning the CCS will experience higher-than-average acidity more often than before. However, the rate at which pH becomes more negative seems to increase significantly after approximately 1980. The peaks and troughs evident in the pre-1980 data also become less evident. A change in the rate of decrease suggests that different processes are dominating the period before the 1980s and the period after the 1980s. In climate science, cycles are indicative of natural changes while long-term positive or negative trends are indicative of anthropogenic impacts (Open Source Systems, 2015). If the first regime of the CCS data can be proven to be primarily cyclic and the second regime can be proven to primarily follow a trend, it would suggest that anthropogenic impacts began to dominate natural processes over the control of ocean pH starting in the 1980s. This is also the report's hypothesis.

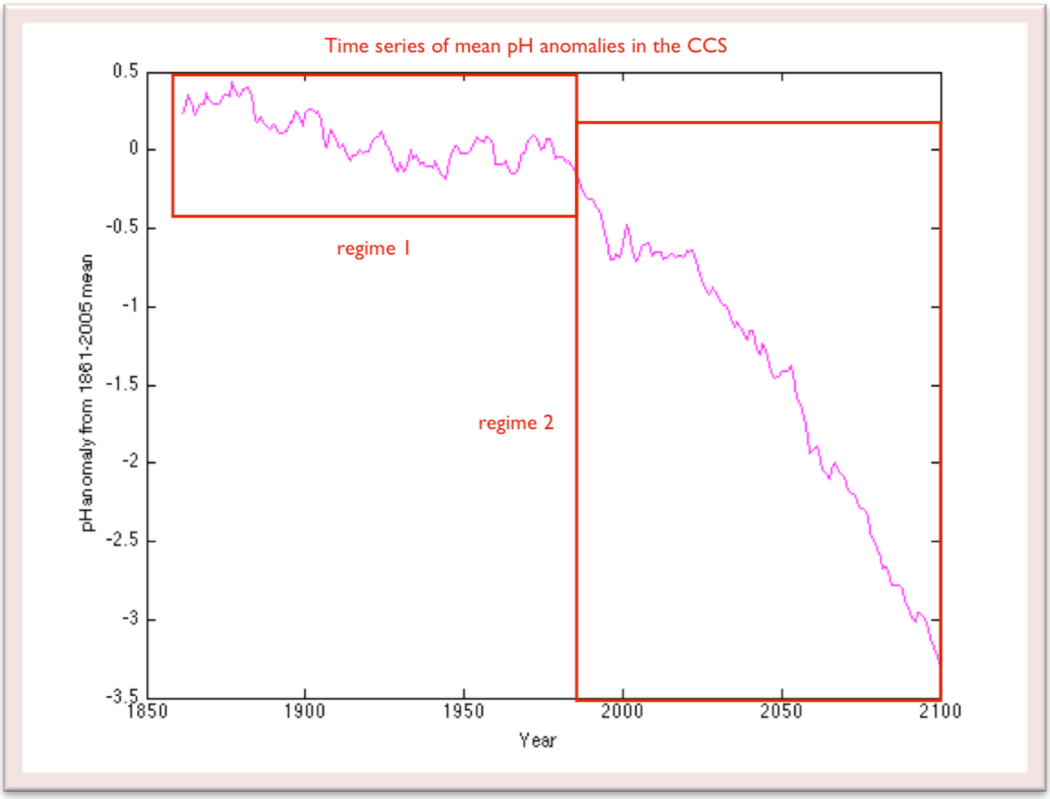

Figure 3. Time series of mean pH anomalies in CCS with regimes made distinct

Regime 1

The strength of natural processes was determined by searching for cyclicality in the data. First, the data was detrended and underwent a test for autocorrelation. If a series of data is autocorrelated, it means that the data is correlated with itself given some time lag (Trauth, 2006). Therefore, data that is cyclic should be autocorrelated and have a time lag near in value to the period of the cycle.

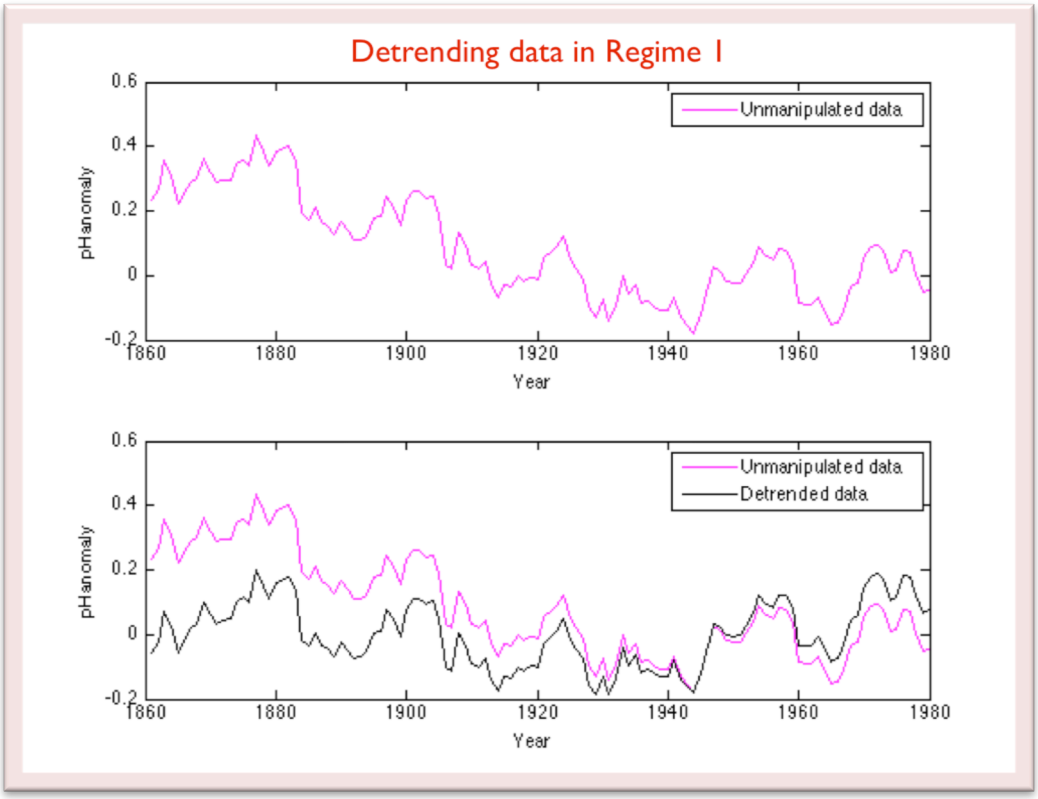

Figure 4. Effect of detrending data in Regime 1

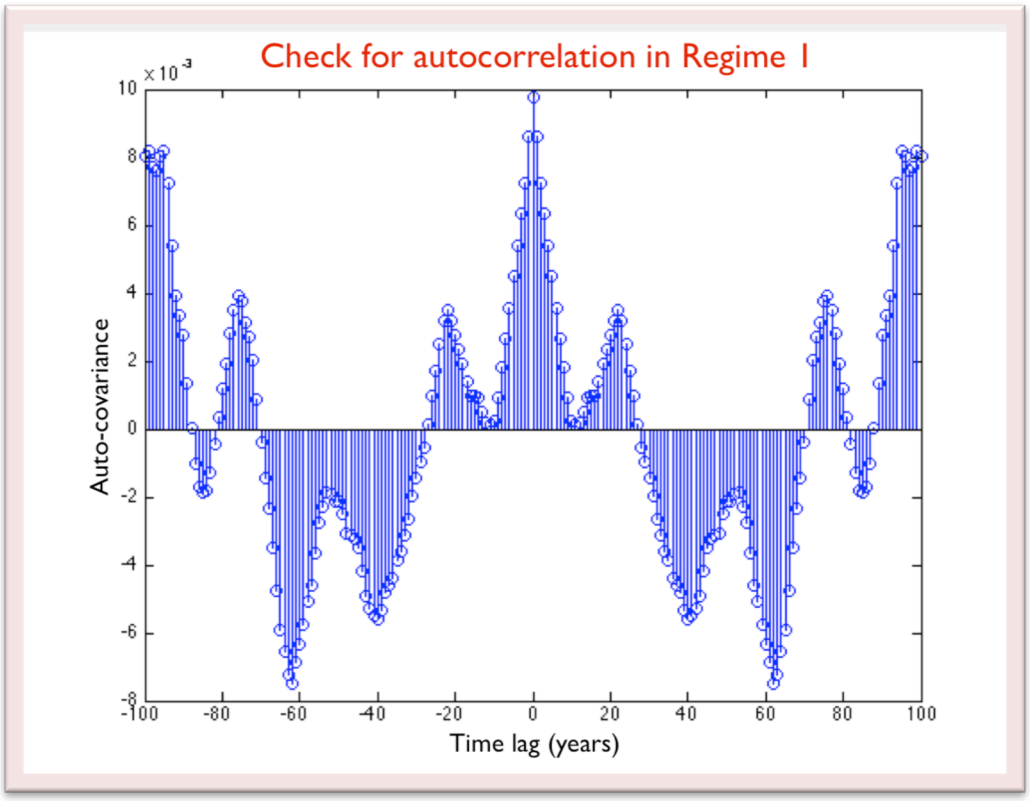

Figure 5. Autocorrelation plot for Regime 1

Figure 5 shows positive correlation at 20, 75, and 100-year time lags and negative correlation at 40, 60, and 85-year time lags. 0 time lag can be disregarded because it is when the data overlaps with itself and will naturally have 100% autocorrelation. The autocorrelation results suggest that Regime 1 not only has cyclic characteristics, but also that it carries multiple, short-period cycles.

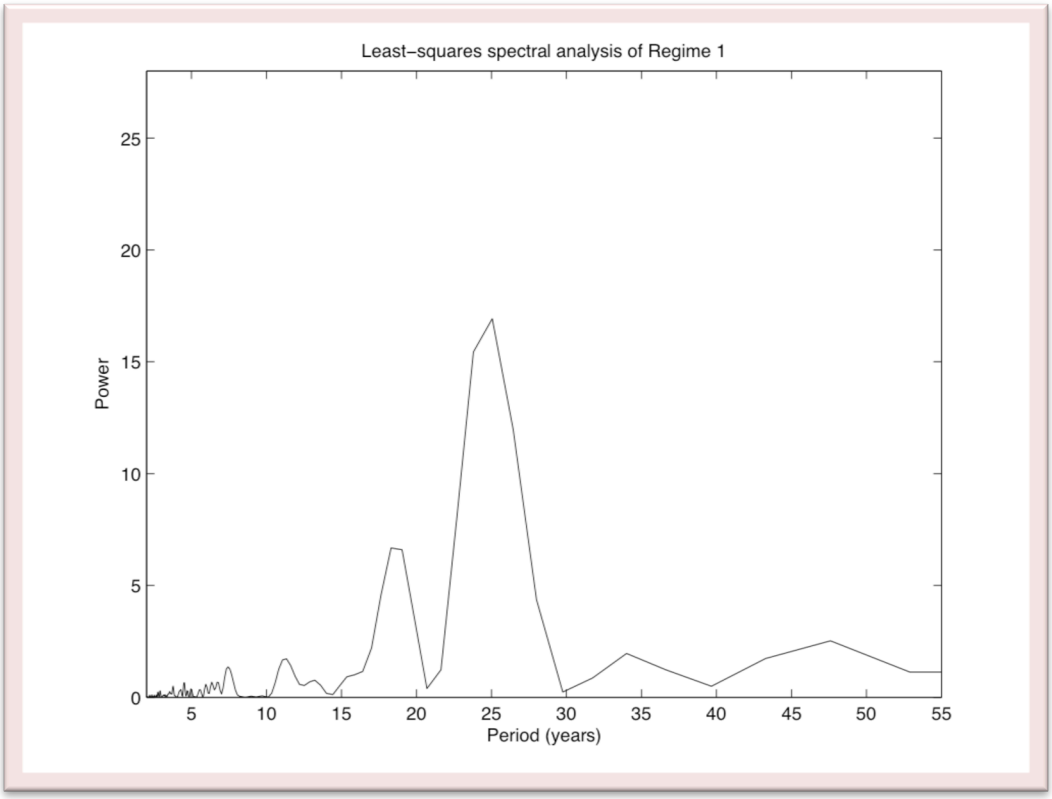

Figure 6. Power plot of Regime 1 using Lomb-Scargle method

In Figure 6, least-squares spectral analysis is applied to discern the exact period of these cycles. Though the data is evenly spaced and a fast-Fourier transform periodogram could have been used, the Lomb-Scargle method was employed because it created a clearer plot. At a confidence level of 99%, there is a cycle with a period between 20 and 30 years, and at a confidence level of 95%, there is another cycle with a period between 15 and 20 years. These results correspond with the results of the autocorrelation analysis. While the long periods identified showed up in the power plot, the uncertainty level was too high to mark them as significant cycles.

The next portion of analysis searched for significant large-scale trends in Regime 1 to determine the impact of human activities on CCS pH anomalies. pH anomalies (non-

detrended) were plotted against years and the polyfit function was used to create a first-order fit for the data (Figure 7).

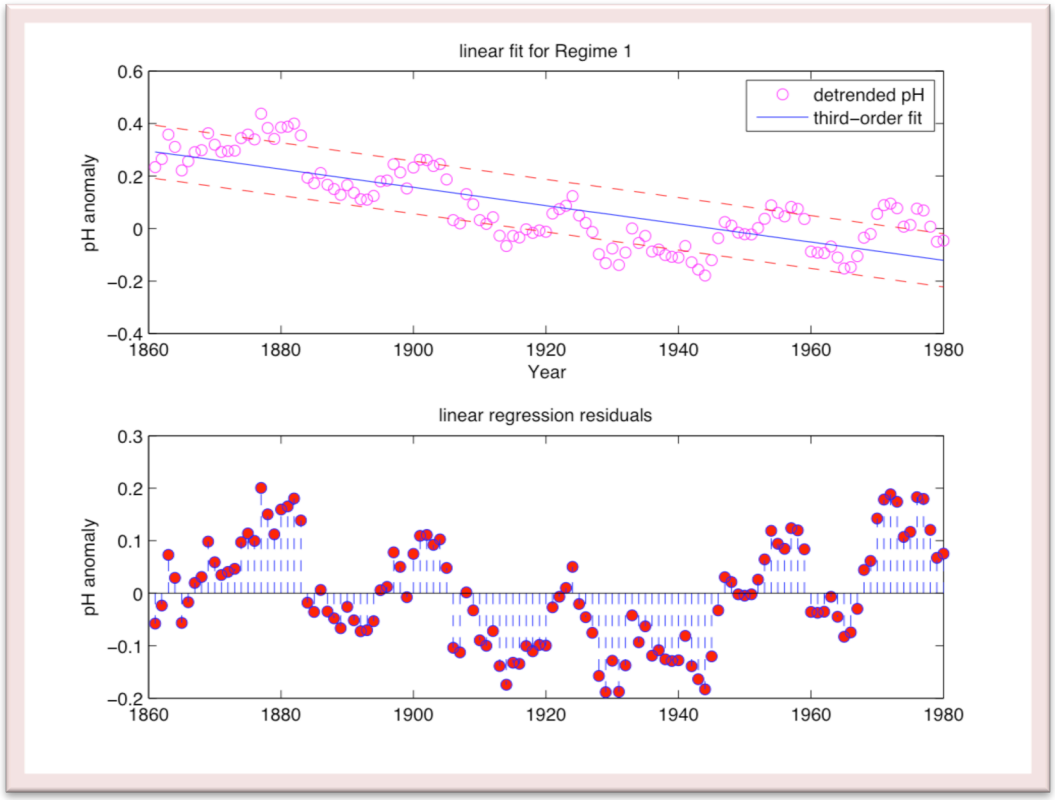

Figure 7. Trendline analysis of Regime 1

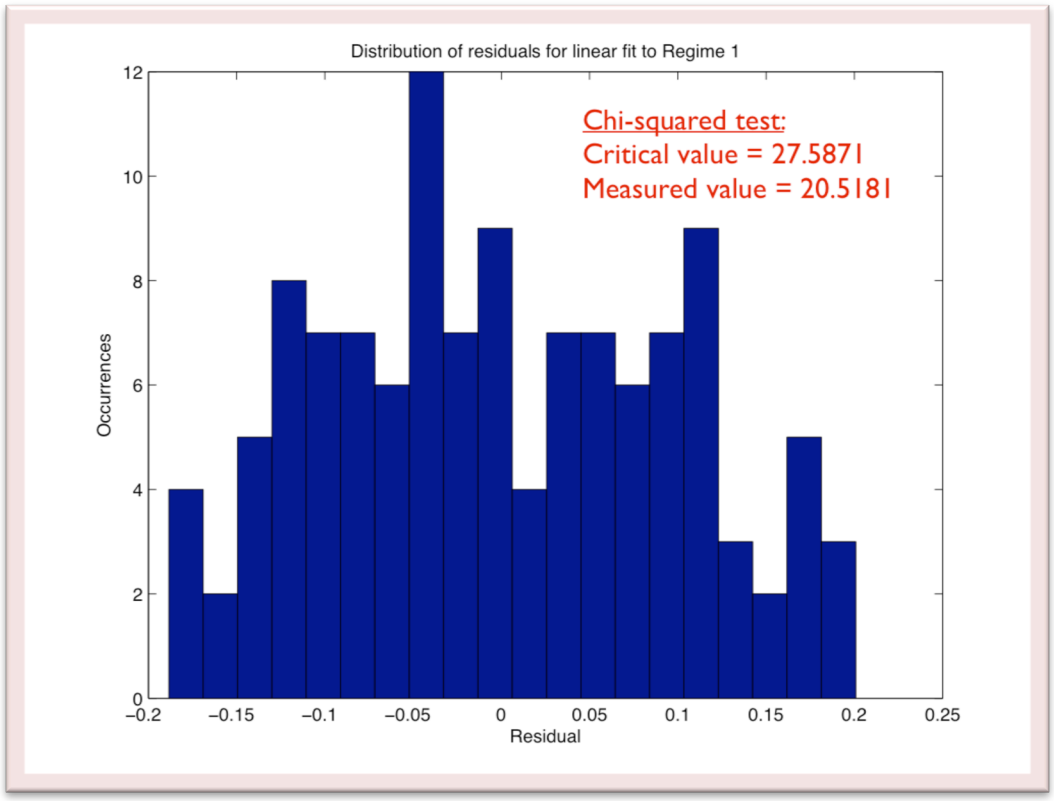

Figure 8. Distribution of residuals for first-order fit to Regime 1

The equation of the first-order fit is $y = -0.0035x + 6.7535$

The slope of the model is -0.0035, meaning that the value of the anomaly falls by 0.0035 each year. The dashed lines depict the 95% confidence interval for each point. Most of the data falls within the dashed lines, but there are several years where it deviates outside the interval. To further determine the strength of the linear model, the residuals were analyzed for randomness and normality. A stem plot reveals that the residuals are distributed on both sides of the x-axis, but there is a clear non-random pattern to their distribution due to the cyclic nature of this regime. A chi-squared test was conducted to check for normality of the residuals (Figure 8). The calculated value (20.518) was lower than the critical value (27.587), signifying that the null hypothesis—the residuals are

normally distributed—cannot be rejected. These results justified the use of the linear model.

Finally, the slope between the real values and the predicted (model) values was calculated to determine the strength of the model (Figure 9). A slope near 1 would suggest that the model was very good at predicting the real values while a slope near 0 would suggest that the model was a poor predictor. The slope was found to be 0.5968, denoting a moderately weak strength for the model.

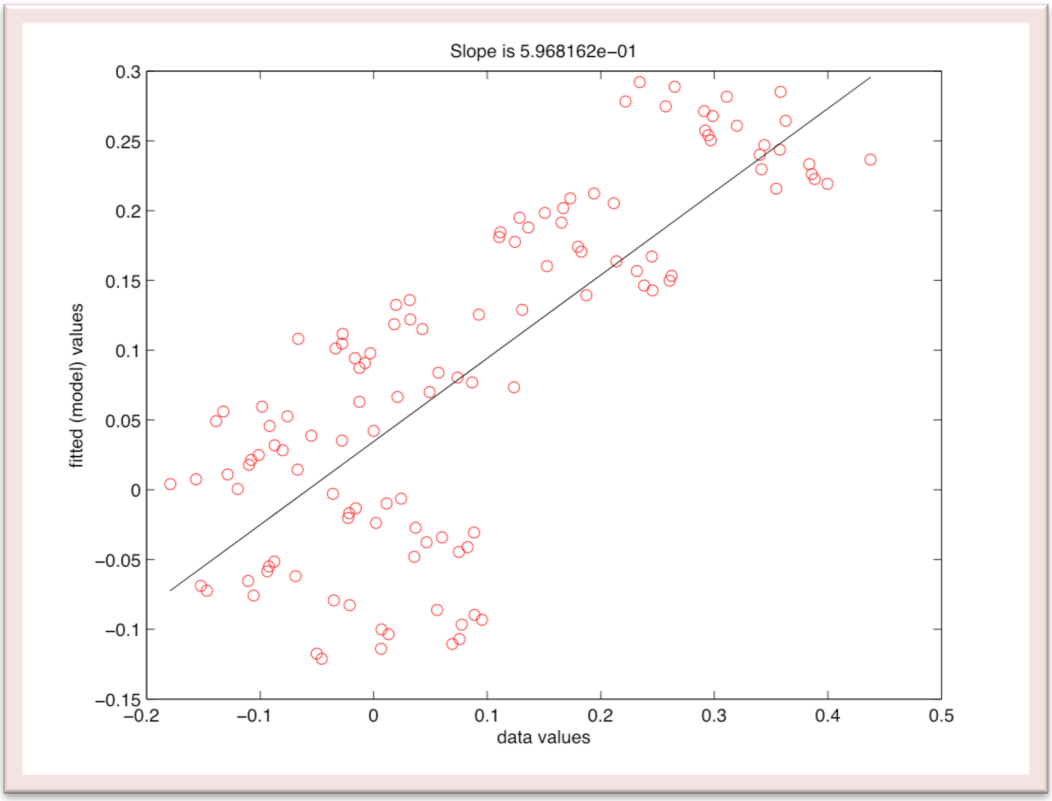

Figure 9. Fitted values (as per linear model) vs. data values for Regime 1

Regime 2

The same analysis conducted on Regime 1 was repeated for Regime 2: an analysis of cyclicality to determine the impact of natural processes and a trend analysis to determine

the anthropogenic impact. Figure 10 displays the effect of detrending the data from Regime 2 and Figure 11 graphs the autocorrelation.

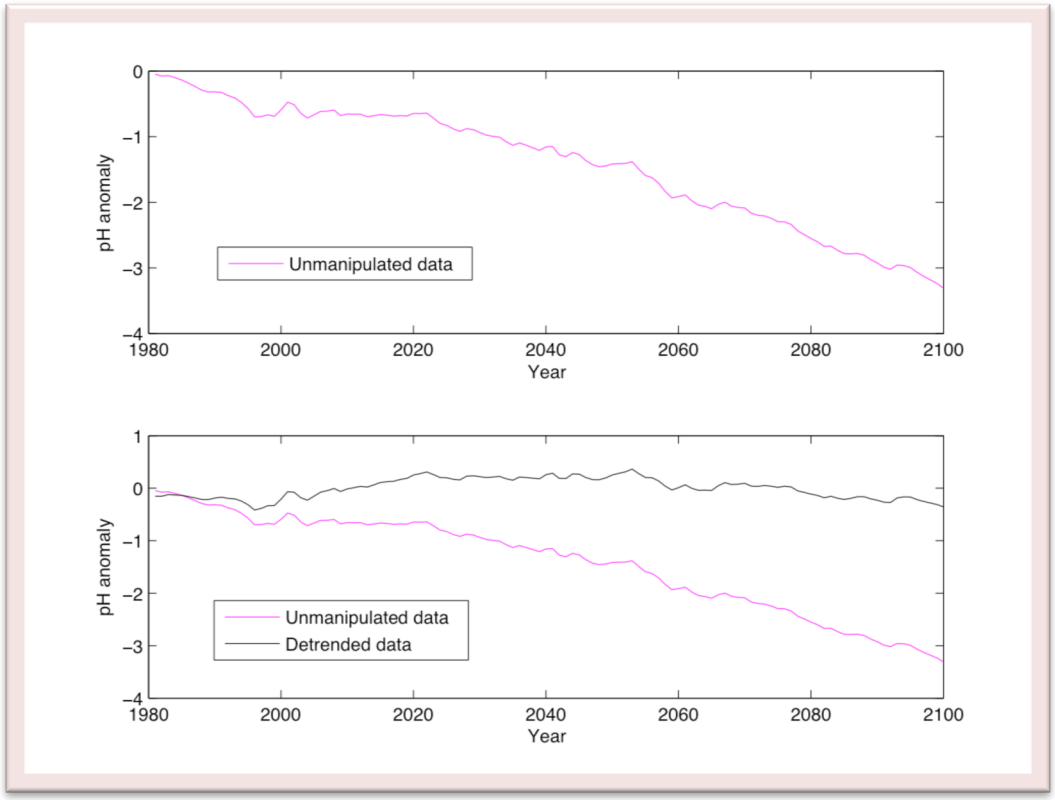

Figure 10. Effect of detrending data in Regime 2

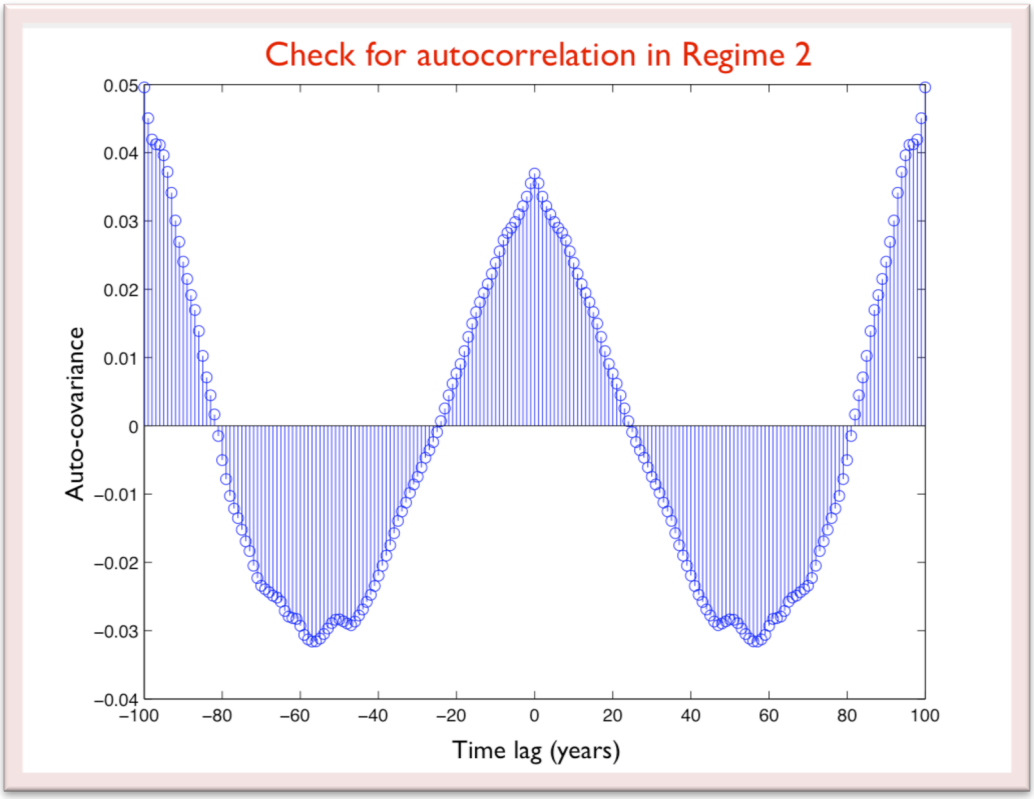

Figure 11. Autocorrelation plot for Regime 2

Figure 11 shows strong correlations at 60 and 100-year time lags. The results show that Regime 2 also displays cyclic behavior, but the cycles are fewer and the time lags are larger compared to Regime 1.

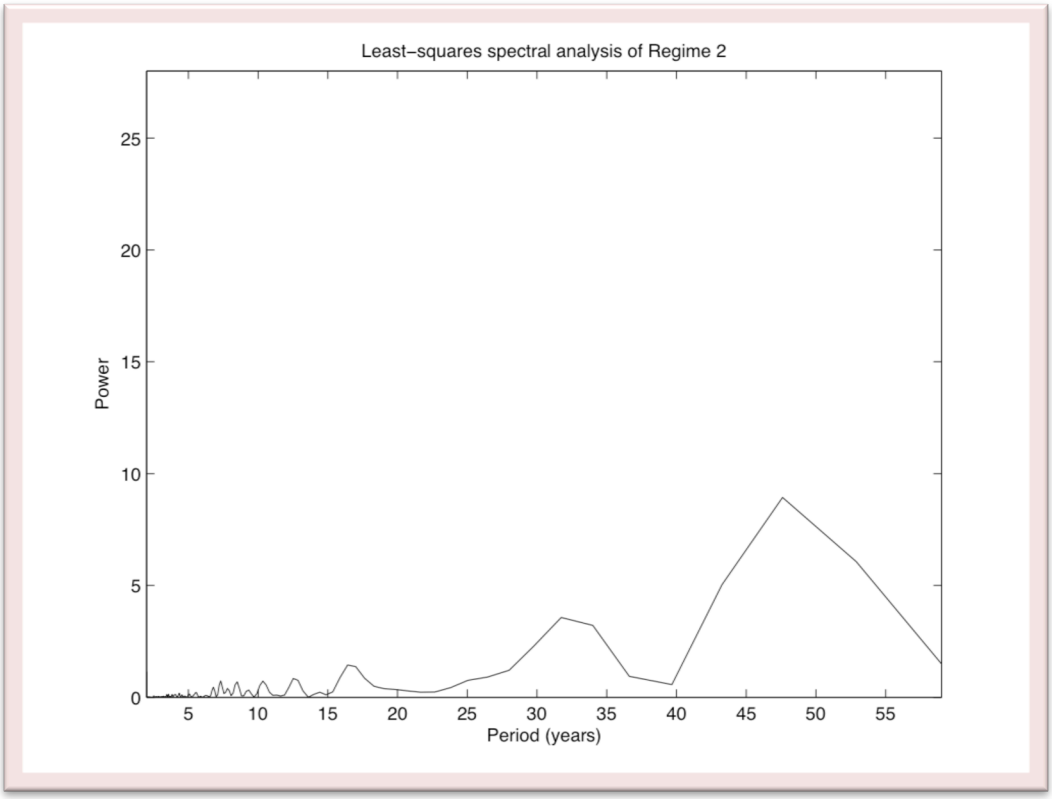

Figure 12. Power plot of Regime 2 using Lomb-Scargle method

In Figure 12, least-squares spectral analysis is applied again to discern the period of these cycles. At a confidence level of 90%, there is a cycle with a period between 30 and 35 years and another cycle with a period between 45 and 50 years. Both of these cycles are longer and determined at a lower confidence level than those in Regime 1.

The next portion of analysis searched for significant large-scale trends in Regime 2 to determine the impact of human activities on CCS pH anomalies. Once again, pH anomalies (non-detrended) were plotted against years and the polyfit function was used to create a first-order fit for the data (Figure 13).

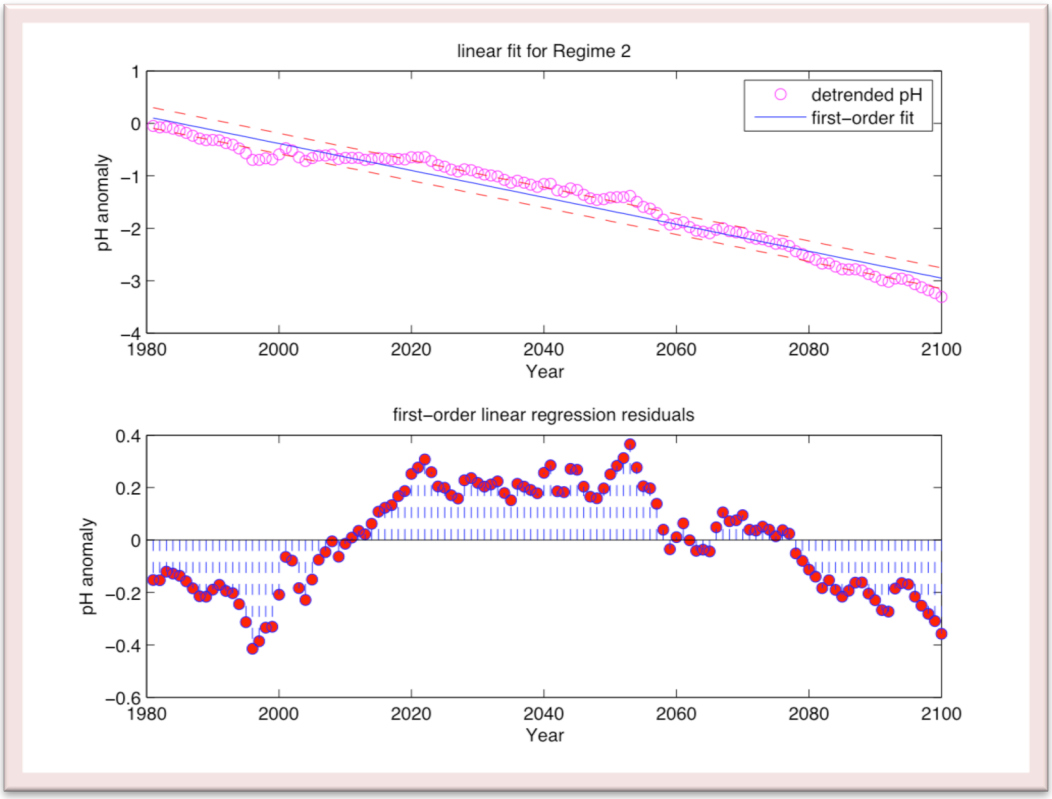

Figure 13. Trend line analysis of Regime 2, first-order fit

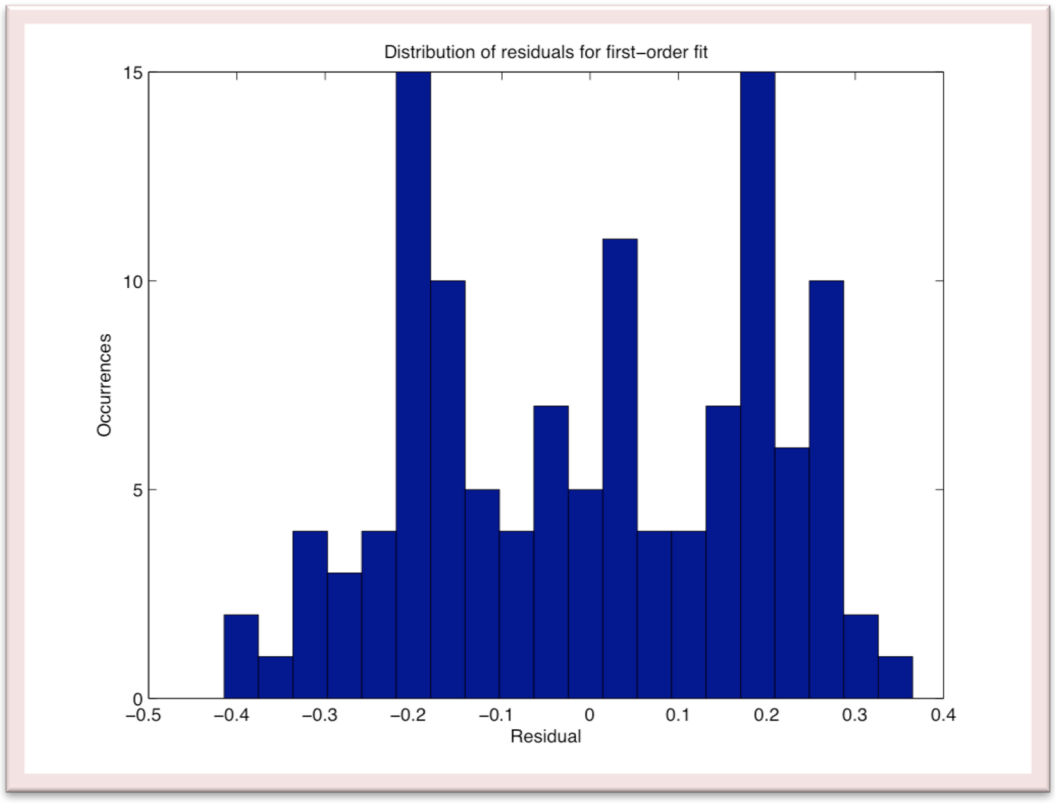

Figure 14. Distribution of residuals to first-order fit to Regime 2

The slope of the model is -0.0257, meaning that the value of the anomaly falls by 0.0257 each year. While this could have justified that the acidity trend is greater in Regime 2 than Regime 1, the results of the residual analysis nullifies our ability to use the linear model. Unfortunately, the residuals are not normally distributed as per the chi-squared test (Figure 14). The calculated value (53.1994) is much higher than the critical value (27.5871), signifying that the null hypothesis—the residuals are normally distributed—must be rejected. A second-order function was attempted next to better model the data, but its residuals also fail to pass the chi-squared test. The third-order model fits best and passes the residual analysis; the trend line and residuals are displayed in Figure 15 and Figure 16.

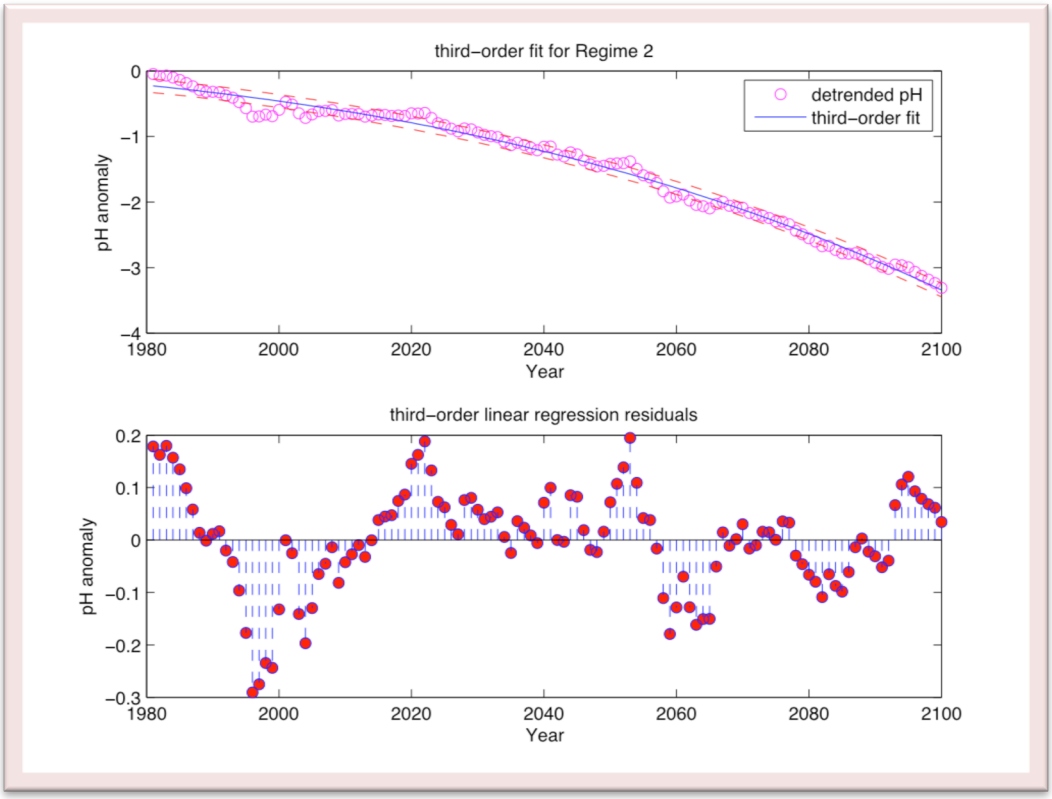

Figure 15. Trend line analysis of Regime 2, third-order fit

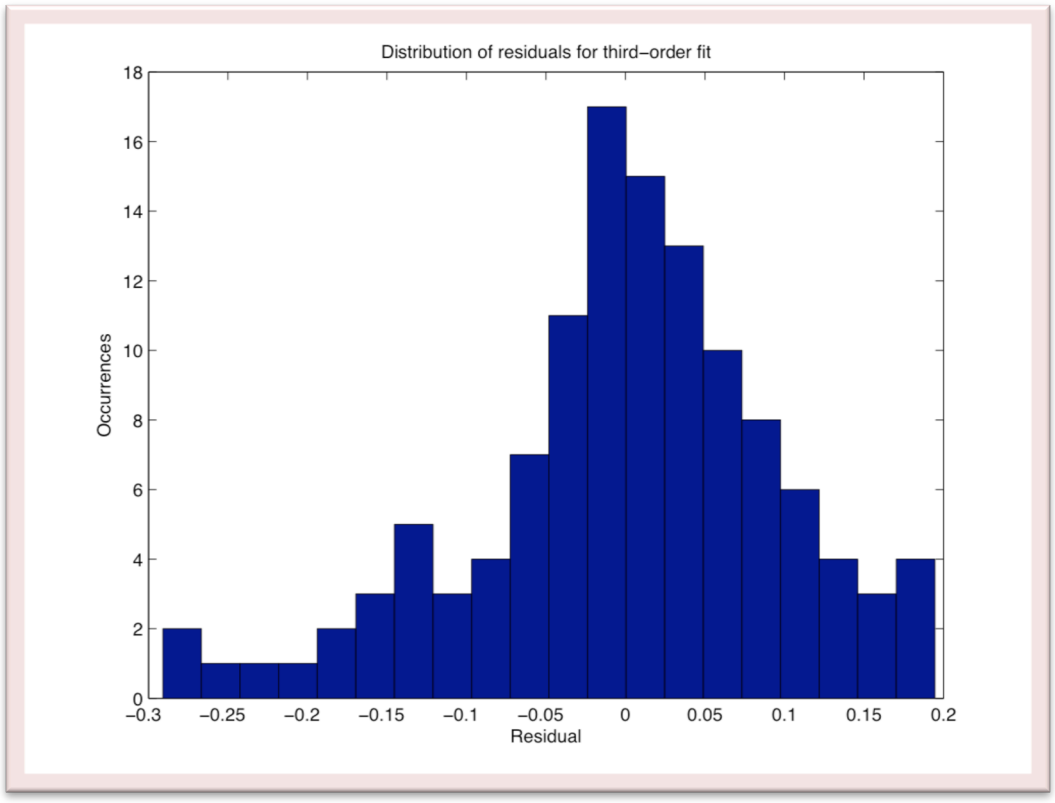

Figure 16. Distribution of residuals to third-order fit to Regime 2

The equation of the third-order fit is

$$y = -0.000001x^3 + 0.000001x^2 - 0.0041x + 2.5821$$

The first two coefficients are so small that the derivative of this equation (for slope) approximately amounts to -0.0041. This slope is greater in magnitude than Regime 1, suggesting that the rate at which pH anomalies are dropping below 0 is greater in Regime 2 than it was in Regime 1. The dashed lines depict the 95% confidence interval for each point, and more of the data falls within the dashed lines as compared to the Regime 1 model. This implies that a trend line is a better fit for Regime 2 than it was for Regime 1, which was better described by cycles. To further determine the strength of the linear model for Regime 2, the residuals were analyzed for randomness and normality.

The stem plot in Figure 15 reveals that the residuals are distributed on both sides of the x-axis. There is still some non-random pattern to their distribution due to the cyclic nature of this regime, but this pattern is less distinct than that of Regime 1. A chi-squared test was conducted to check for normality of the residuals (Figure 16). The calculated value (27.393) was lower than the critical value (27.587), signifying that the null hypothesis—the residuals are normally distributed—cannot be rejected. These results justified the use of the third-order model. Finally, the slope between the real values and the predicted (model) values was calculated to determine the strength of the model.

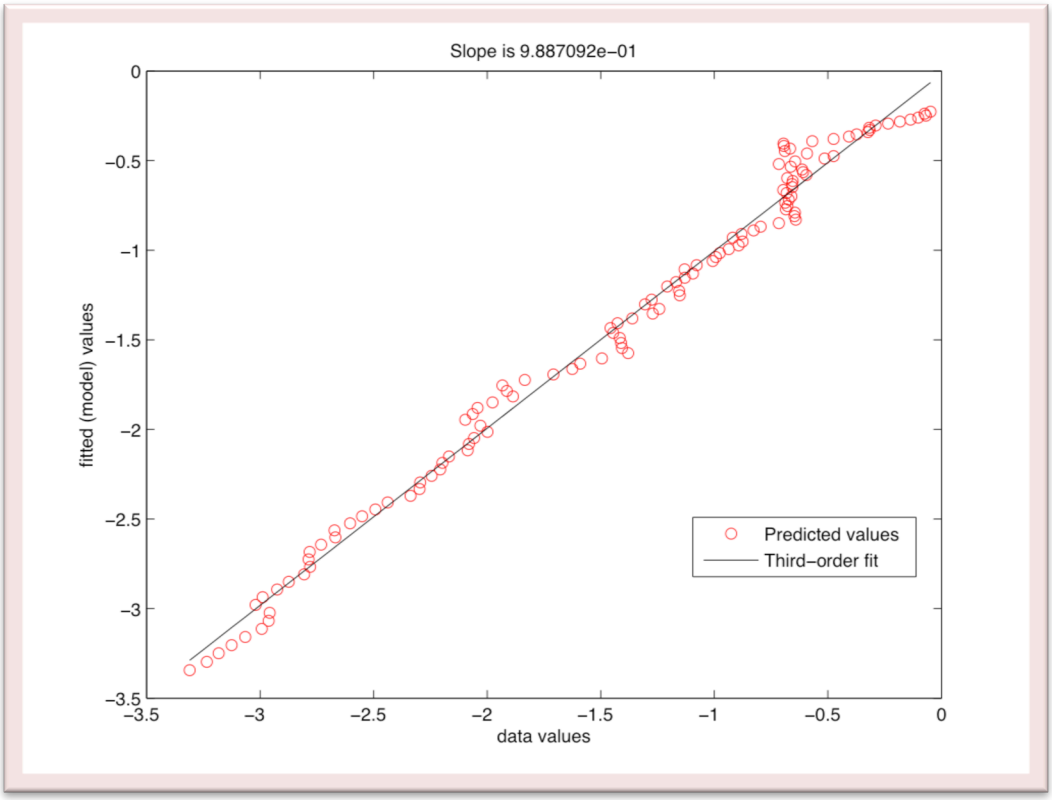

Figure 17. Fitted values (as per third-order model) vs. data values for Regime 2

The slope was found to be 0.989, suggesting this model is very accurate. The slope is much closer to 1 than model for Regime 1, implying that Regime 2 follows a trend much better than Regime 1 does and carries less error.

Conclusion

Analysis	Regime 1	Regime 2
Autocorrelation for cyclic analysis	Cycles exist; short cycles As many as 6	Cycles exist; long cycles As many as 3
Lomb-Scargle test for periodicity	99% confidence – cycle with 20-30 year period 95% confidence – cycle with 15-20 year period	90% confidence – cycle with 30-35 year period 90% confidence – cycle with 45-50 year period
Chi-squared on residuals of first-order model	Passes Measured value: 20.518 Critical value: 27.587	Fails Measured value: 53.199 Critical value: 27.587
Order of best trend line and its slope	Trend exists First-order Slope = -0.0035	Trend exists Third-order Slope ≈ 0.0041
Strength of best-fit trend line	Moderately weak Slope = 0.597	Very strong Slope = 0.989

Figure 18. Summary of analysis

In summary, I found natural and anthropogenic impacts in my two regimes, but the natural impacts were stronger in Regime 1 and the anthropogenic impacts were stronger in Regime 2. From Regime 1 to Regime 2, the cycles signifying natural impacts become fewer and longer, and the confidence level for the periods of the most prominent cycles drops by at least 5%. While a negative trend exists in both regimes, the rate of decrease is greater and more exponential in Regime 2 than Regime 1. This result alongside the strength of the best-fit trend line being greater for Regime 2 than Regime 1 suggests that anthropogenic impacts play an increasingly larger role in Regime 2. In its entirety, this analysis supports the hypothesis that anthropogenic impacts began dominating natural processes over the control of ocean pH in the California Current System in the 1980s, the boundary between Regimes 1 and 2.

In the future, it would be worthwhile to analyze the same data after dividing it into more than two regimes. Considering the Industrial Revolution was already in full swing by 1900 and major CO2 emissions had been happening for several decades, anthropogenic impacts should have been dominating natural impacts prior to the decade of 1980. Future research can also try keeping two regimes and moving the boundary from 1980 to an earlier or later year. The year 1980 was chosen somewhat arbitrarily after initial observation of the pH data. Comparing the results of the same analyses before and after moving the boundary year may provide new knowledge in the field of ocean carbon circulation, particularly in regards to time lag/delay in the subsurface CCS.

References

Administration, N. O. a. A. (2015). A primer on pH. *PMEL Carbon Program.* 2015, from http://pmel.noaa.gov/co2/story/A+primer+on+pH

Center, N. C. D. (2015). Anomalies vs. Temperature. *Monitoring References.* 2015, from http://www.ncdc.noaa.gov/monitoring-references/dyk/anomalies-vs-temperature

NOAA. (2014). GOLD. *Models.* 2015, from http://www.gfdl.noaa.gov/gold-ocean-model

Open Source Systems, S., and Solutions. (2015). Global Warming Natural Cycle. *Climate Change, Natural Cycle, Global Warming.* 2015, from http://ossfoundation.us/projects/environment/global-warming/natural-cycle

Talley, L. (2000). Talley Topic 2: Properties of seawater. *SIO 210.* 2015

Trauth, M. H. (2006). *MATLAB Recipes for Earth Sciences* (pp. 237). Retrieved from http://staff.on.br/puxiu/MatLab_Pack/MATLAB Recipes for Earth Sciences - M.H.Trauth.pdf

Turner, E. (2011). U.S. Global Ocean Ecosystems Dynamics (GLOBEC) Northeast Pacific. *Climate Change.* 2015, from http://www.cop.noaa.gov/stressors/climatechange/current/fact-globecpne.aspx