# The frequency model of blooms of plankton in the coastal region of Kimberly, Western Australia by time-series analysis


## Abstract

The plankton community plays a critically important role in marine ecosystems as it contributes to the primary production and provides and cycles nutrients for the other underwater species. Plankton biomass, inferred by chlorophyll concentration from satellite data, is closely related to the naturally changing environment and the acute hazards, thus, modeling the plankton biomass cycles can spot the abnormal time in which unexpected extreme events occurred. Based on the fixed harmonic terms representing the replenishment of plankton and identification of hazards, further attempts can be invested to capture these unpredictable influences on plankton growth in the region, providing insights to improve conservation and resource management.

The approach Dominant Frequency State Analysis is used to decide the dominant frequency and to test the influence of high and low frequency on plankton records. The model result shows that multiple-year frequency does not have any statistically significant influence, while the annual cycle is dominant, supported by months' trend. The frequency states range from 3 months to 12 months. The model is a good fit, however, highlighting abnormal peaks and lows of plankton biomass, indicating acute hazards around the time.


## 1. Methodology of analysis

a. Data collection and the influence of data science

The phytoplankton biomass was inferred by NASA's surface chlorophyll concentration. The data which will be used in this analysis report was combined from multiple satellites and carried out over 24 years from 1998 until 2021 at the sampling point near the Kimberley coastline, North-western Australia. The data records are collected and published in The Spatiotemporal Data & Time Series Toolkit by the COPEPOD project - a free and compiled resource for plankton and ecosystems data research. The blue and green dots represent the density and type of plankton of the sample location, and the red dot is the sample location where the dataset of chlorophyll concentration is compiled.

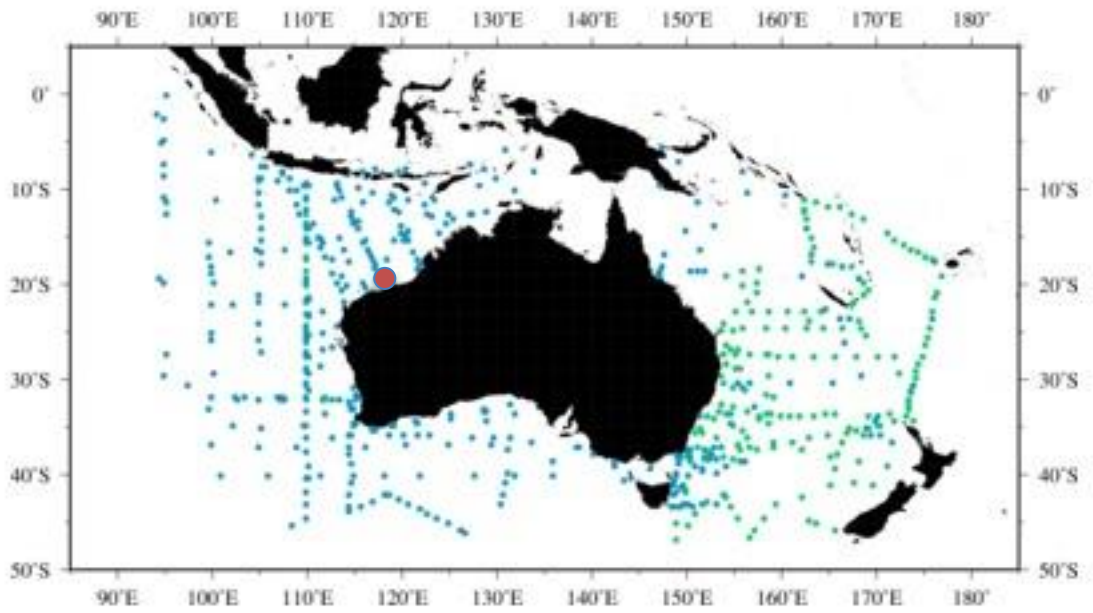


**Figure 1:** The chlorophyll concentration of Australian coastal lines and the sample place.

b. Methodology of analysis - Literature review

In this report, we will use the Dominant Frequency State Analysis (DFSA) approach, which is employed in several studies about earth system time series. This approach was first introduced in the study analysing the dynamic trend of the Pacific ENSO phenomenon (Bruun, Icarus Allen & Smyth, 2017). This study used long-term oscillations (up to 200 years) and tropical and extratropical interactive components, however, under the scope of this report and the limitation of historic data, the model will test the contributions of cycles of up to 10 years.

According to the article, the DFSA uses eigenvalue equations to characterise the dynamic components of the earth system process. All possible eigenvalues will form a spectrum in which the dominant term will stand out. The combination of eigenvalue and eigenvector can be expanded to reflect the frequency states at a specific location. Thus, DFSA possesses certain advantages over other analysis approaches such as spectral tapering methods, singular spectral analysis, and analysis with empirical orthogonal functions. It provides a complete representation of the process by estimating the signal structure without explicitly specifying different frequency bands, dynamic operators, and noise functions in another step.

The spectral tapering methods have been developed and widely used in several pieces of research about climatic phenomena. These methods, including the recently used multitaper method, evolve around smoothing and tapering assumptions to detangle different signal structures and assess frequency bands (Mann and Lees, 1996; Ghil, et al., 2002; Li, et al. 2013). An advantage of this act is a stable spectral estimate without noise and fine-scale structure as it invests in separating the phenomenon from underlying noise after representing the noise

structure. On the other hand, DFSA does not break into steps but estimates the noise with signal components simultaneously.

Without previous knowledge of an underlying physical phenomenon, information can be obtained by singular spectrum analysis (SSA). It can recognise and estimate systems with eigenvalues and is applied independently to search high and low-frequency signal bands. On the other hand, while the standard empirical orthogonal function (EOF) initially defines a multiplicative time and space function, the Complex and Complex Hilbert EOFs develop from the original method by capturing the structure of time signals.

By reviewing the previous methods and analysing their advantages and disadvantages, the chlorophyll concentration fluctuations will be examined by the DFSA method due to its fitness. Compared to the other methods, DFSA is stronger in time signal structure and can perform without spatial-temporal data.

## 2. Model development result

The result will show the attempts and performance of the model fitting with the chlorophyll dataset. Based on the current state of the chlorophyll concentration and the 24-year history, attempts to develop a model have been made to conclude a good-fit model which can capture possible linear and cyclic trends. This part will decide which linearity and cycles have significant impacts on the chlorophyll concentration.

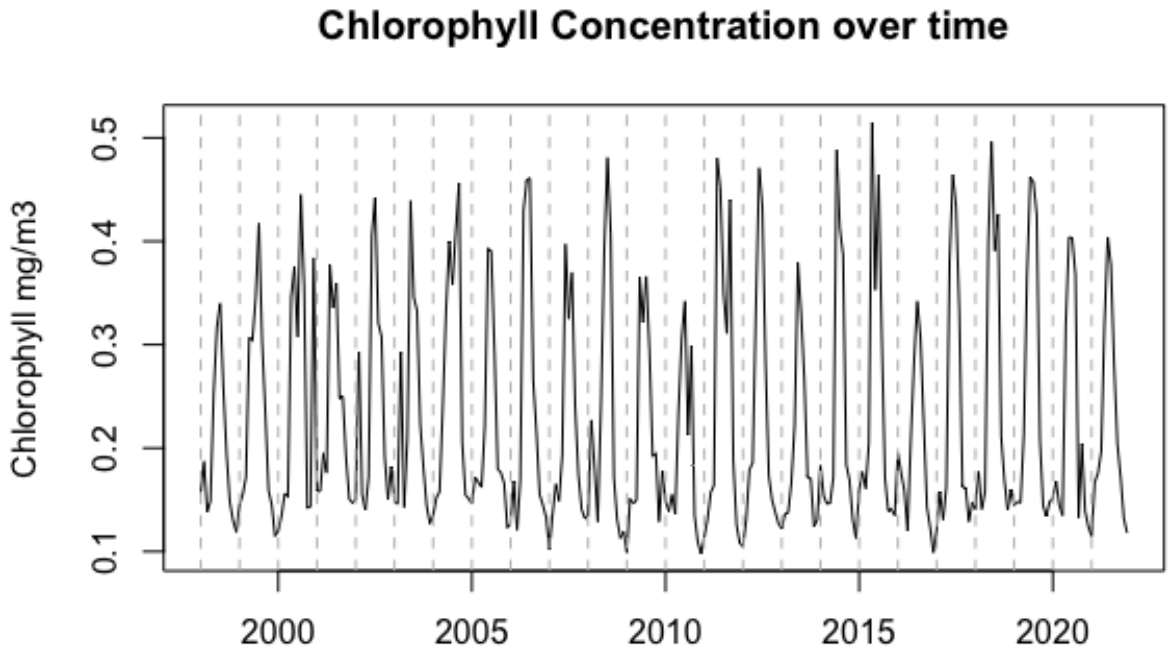

**Figure 2:** Chlorophyll concentration records from 1998-2021 (raw data)

Figure 2 plots the chlorophyll concentration by month from 1998 to 2021. As can be seen from the raw data, the plankton population normally blooms only once a year, except for the period 2000-2004 in which there were multiple population surges. However, the yearly peaks are not stable: one or two high peaks followed by a plunge peak. In 2009-2010 and 2016, the blooms were significantly lower than in other years around.

As the data is asymmetric and all are positive, log transformation of the raw data would be conducted to make it symmetric. We will apply a seasonal oscillation to explain the records, and experimental trials with multiple frequencies to test which frequencies best capture the variation. The most influential frequencies will be retained and compared in a frequency spectrum tool to conclude which is the dominant harmonic cycle.

Residual plots will be in use with trials to test if there is any pattern left out by the model. A good model will result in a sensible high R-squared index, high maximum likelihood, and low AIC score. Besides, randomly distributed residuals would indicate that the model has captured most cyclic patterns and left only random noises.

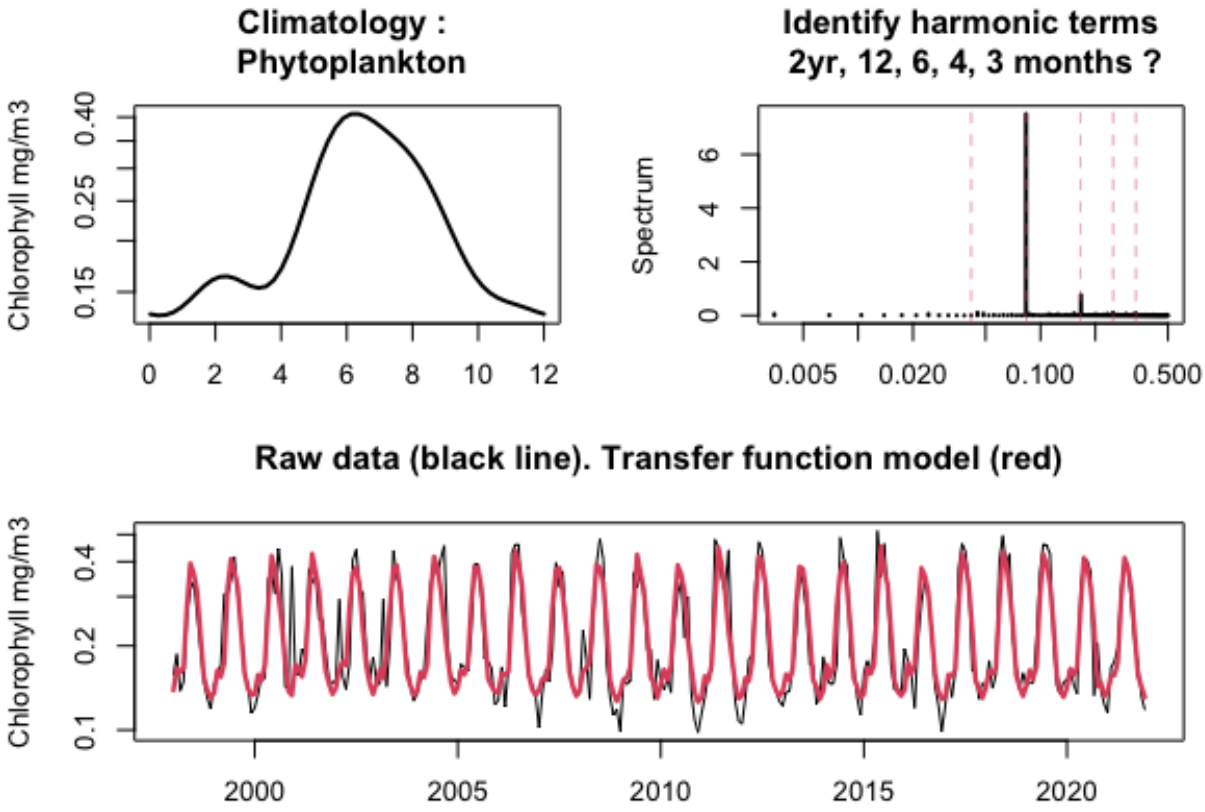


**Figure 3:** Annual harmonic terms and transfer function model fit

The figure above shows the result of the transfer function model (red line) by cyclic variation of 3, 4, 6, 12, and 24 months and the annual harmonic cycles (excluding the 2-year cycle). The phytoplankton biomass, indicated by the chlorophyll concentration, blooms only once in June every year, then the population nearly disappears from November until January next year.

The frequency spectrum tool was used to examine the annual harmonic cycles with the Periodogram. From the Periodogram, the dominant cycle is 12 months, followed by 6 months. This indicates the annual and seasonal trends in the plankton biomass within a year.

Looking at the original data and the red model fit, the model can be considered a good fit for the raw data. However, there are some abnormal spikes that the model cannot capture. The first one is in 2000 in which plankton surged twice a year and the model missed the second bloom. In the next year 2002-2003, the model cannot capture the sudden surge before June's peak, especially in 2003 even though this happened two years consecutively. However this pattern did not last for the next 20 years, thus, other features might be required to capture these abnormalities. It can be seen from the plot that there are multiple consecutive years from 2007 in which the model failed to predict the lowest point in the low season despite succeeding in reaching the peak in those years. However, in the last 5 years, the model looks like a good fit with the raw data.

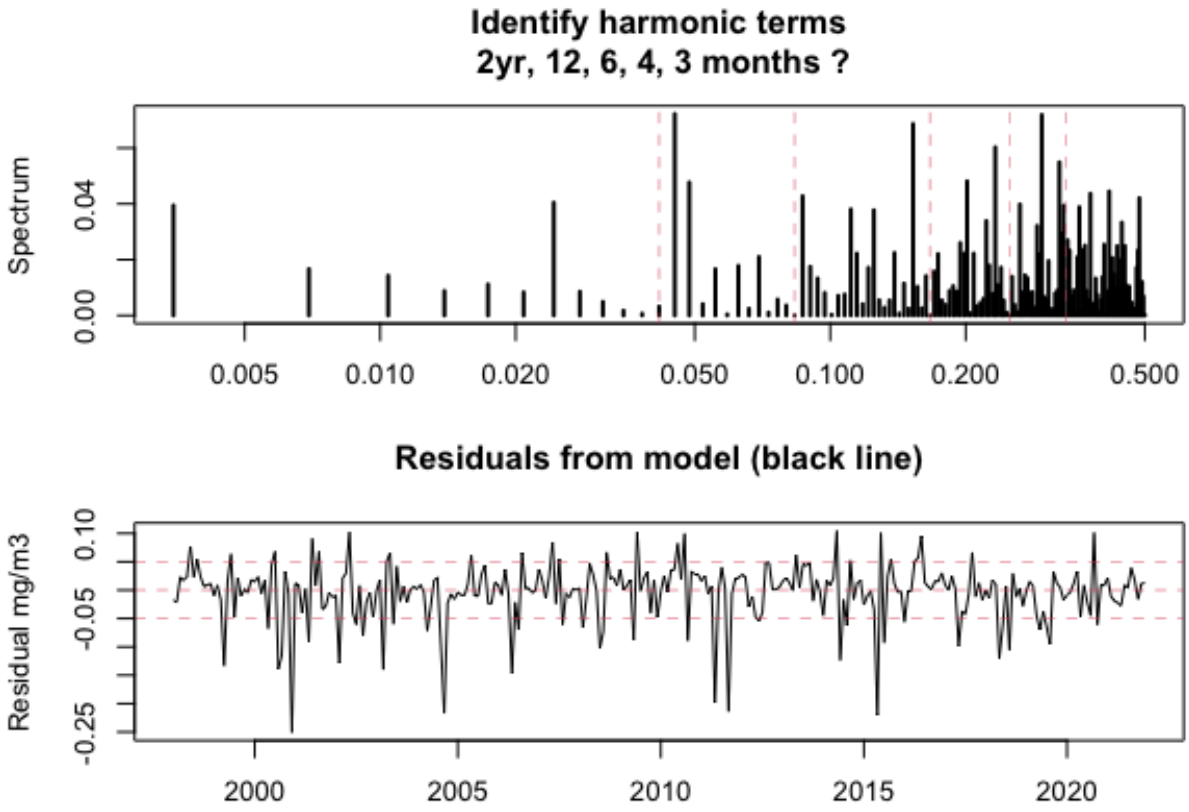


**Figure 4:** All harmonic terms of residuals (including 2 years) and residual plot of the model

From figure 4, it can be seen that the harmonic terms of residuals are not noticeably significant with the highest terms reaching around 0.05 of the spectrum. These most influential cycles surround 2 years, 6 months, and somewhere between 3-4 months - which has been captured in the model fit. On the other hand, these harmonic terms of residuals are not as significantly dominant as that of the fit data in figure 3. There are a lot of other frequencies scattering from 2 to 24 months recording comparative impact (the black bars from 2 months - at 0.5 - until over 6 months - the middle red dashes).

The residuals plot the difference between fitted values and the actual values of chlorophyll concentration. As can be seen from the plot, the absolute value of residuals at some minimum remains high, especially in 2000-2004 when the model failed to predict the second bloom in the year and the peak residual before 2005 is since the bloom came later that year. The two consecutive peak residuals after 2010 and after 2015 refer to the second explosive bloom after June's bloom in 2010-2011, 2015. Except for these deep plunges, most residuals scatter quite symmetrically around 0 and within 0.05 distance. The residuals do not cluster around at any time and do not show any clear patterns.

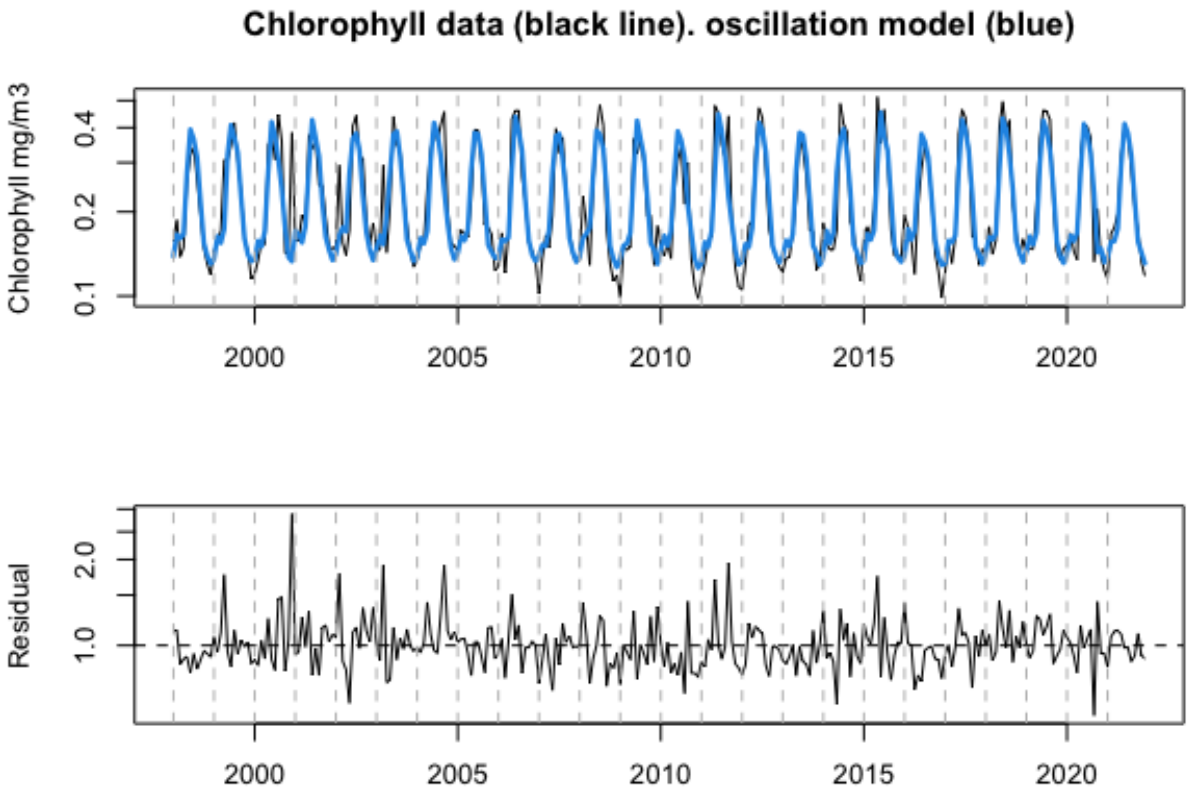


**Figure 5:** The oscillation model (blue line) and the 10^(residuals of transfer function)

The final model is best fitted with no linear trend, and no Seasonal Autoregressive, with influence from Autoregressive and cyclic trends of 3 months, 4 months, 6 months, 12 months, and 24 months. The R-squared records 79.8%, indicating that the model covers nearly 80% of the

variation. From the plot and the result of the residuals test, with a p-value of 0.113 > 0.05, there is enough evidence to reject the dependence in residuals. With a balanced number of values below and above 0 (131 and 157 respectively), residuals are confirmed to distribute independently around the mean = 0.

| | | 2.5 % | 97.5 % |
|---|---|---|---|
| ar1 | 0.1981960 | 0.0850745 | 0.3113175 |
| sar1 | 0.0000000 | NA | NA |
| intercept | -0.6873307 | -0.6998372 | -0.6748243 |
| h.c1-MA0 | 0.0444329 | 0.0289662 | 0.0598997 |
| h.s1-MA0 | 0.0575530 | 0.0420748 | 0.0730311 |
| h.c2-MA0 | -0.2150093 | -0.2319990 | -0.1980196 |
| h.s2-MA0 | -0.0769700 | -0.0939838 | -0.0599562 |
| h.c3-MA0 | 0.0000000 | NA | NA |
| h.s3-MA0 | 0.0000000 | NA | NA |
| h.c4-MA0 | -0.0265693 | -0.0404938 | -0.0126448 |
| h.s4-MA0 | 0.0047065 | -0.0092180 | 0.0186311 |
| h.c5-MA0 | 0.0068705 | -0.0058976 | 0.0196386 |
| h.s5-MA0 | -0.0231623 | -0.0359225 | -0.0104021 |
| t.seq-MA0 | 0.0000000 | NA | NA |

**Table 6:** Coefficients and 95% Confidence interval of the model

Table 6 shows the magnitude of coefficients and their confidence intervals of them. After running multiple trials of frequencies with Autoregressive (ar1), Seasonal autoregressive (sar1), and linear components (t), the coefficients of sar1 and t show no significant impact on the final result of the model, thus removed without noticeable impact on R-squared, likelihood and AIC score. This indicates that the chlorophyll concentration is significantly related to the lagged values, but not related to the seasonal lagged values (the lagged values at the same month in the previous year). On the other hand, 12 months cyclic is statistically significant with the concentration.

Frequencies of more than 1 year (from 2-10 years) have been tested but show no significant influence on the final result of this model. Thus, the model will base on annual and month-periodic harmonic cycles to predict the values of chlorophyll concentration.

## 3. Discussion

a. The model coefficients

The AR1, denoted by autoregressive coefficients, suggests the strength and direction of the linear relationship between chlorophyll's current value and its most recent past values. As the value has been log-transformed, this coefficient shows the linear relationship between the log value between

the current and last record. AR1 = 0.18 with a 95% confidence interval implies the positive relationship between two log values, and thus, a positive relationship between two values.

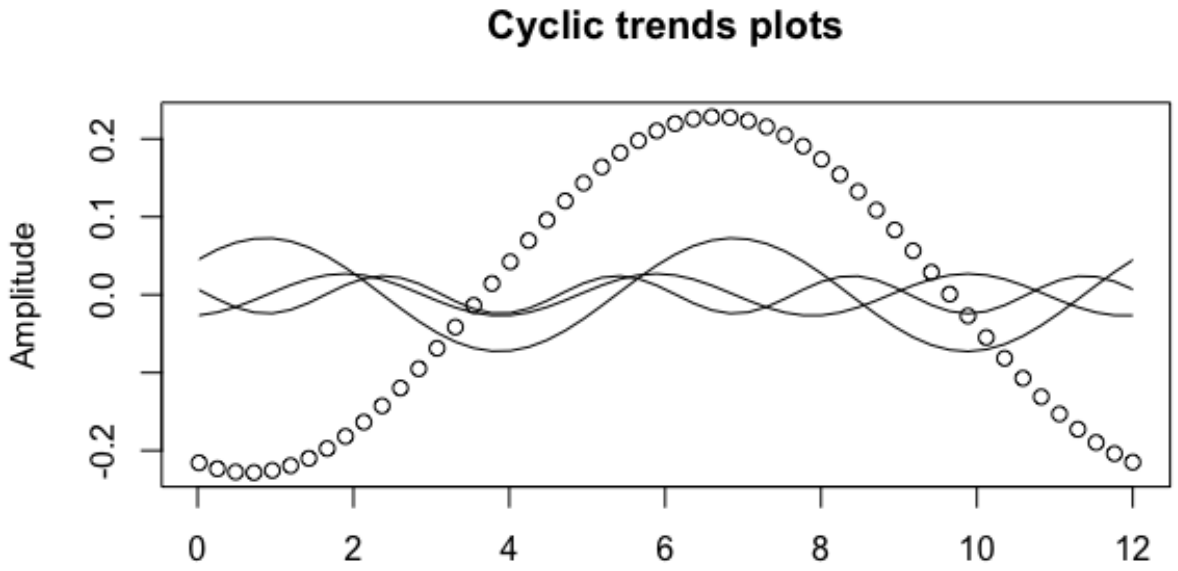


**Figure 6:** Plots of cyclic terms (3 months, 4 months, 6 months, and 12 months)

Figure 6 resembles the in-year concentration of chlorophyll in figure 3. The peak cycle of all 4 trends evolves between June and July, representing June's peak in the record. On the left of the peak, February has two short-term frequencies reaching the upper cycle, while April witnesses the down cycle of all three terms, resembling the small hump in February followed by the plunge in April. Meanwhile, on the right of the peak, the three trends vary between up and down, thus, despite being dominated by the downward of the 1-year frequency, the decreasing trend on this side is not as steep as on the left side.

b. The remaining uncertainties, their reasons

The abnormal spikes in the residuals are closely related to the regional climatic acute events as similar natural disasters coincide with the sudden plunge of the plankton community. One hypothesis to explain the drastic residuals in 2002 in figure 4 can be cyclone Chris which struck straightly the Western Australia in February. The Board of Meteorology of Australia considered this one of the most influential and harmful tropical cyclones, causing intensive landfall and the worst flood in the Nullagine River.

In 2010-2012, La Nina caused tremendously high downpours, flooding the regions and increasing the fresh water from coastal areas in the La Nina seasons for two consecutive years. According to the Australian Board of Meteorology (2012), the northwestern regions experienced the highest and second-highest rainfall volume in history since the catastrophic flood in 1900. In December 2010, Western Australia also experienced the Gascoyne River flood, considered the most severe

in the region as it happened suddenly in the third-driest month of the year in the driest region of Australia (Macqueen, 2011).

Another considerable flood in the Western occurred in February 2017, referred to as the Western Australia floods (Morrison, 2017). This region is characterised by arid land and dry weather type, thus, the salinity concentration dropped at a historically serious rate compared to the other year. Another consequence of these years' extreme La Nina is the temperature decrease in the region. Changing the standard water quality and temperature of the sea surface may negatively hinder plankton growth and reproduction (Yuan, et al., 2020). This could explain the weakening bloom of plankton in the area in both June's peak month and the low season. These prolonged flooding years coincide with the abnormally decreasing records of the plankton population compared to other years.

Apart from the suspicious lowest points in the plankton population, the model of chlorophyll concentration could not capture the second blooms apart from the annual June peak. Another unexplained event is the plankton bloom in 2004 which was unexpectedly delayed and occurred later than the yearly June peak, shown in figure 4. In 2004, the annual growth of plankton was delayed but reached the normal peak, however, suddenly plunged sharply afterward, coming back to the annual trend in late that year. No empirical explanation has been found for the second blooms in the extreme weather years and the delayed bloom of plankton. There has been no report about natural disasters happening in that region around that time. Until now, hypotheses about these might have to be left unsolved.

c. Possible data science solutions to the remaining uncertainties

The main causes of the remaining noise which has not been captured by the model are due to unexpected extreme events, thus, it would be difficult to incorporate these acute catastrophes into our original model. A noise function can be added to our function if we can model the time and effects of these acute events and make our noise function flexible enough to capture the impacts of sudden hazards.

Some efforts have been made to predict, as Damle and Yalcin (2007) tried to predict flood as early as possible, however, it comes with an increased risk of missing the starting time. Geetha and Nasira (2016) attempted to predict the tropical cyclone from 6-year records of 14 features and Salmun and Molod's study (2011) provided estimates and insights for predicting floods in the coastal regions. In the study introducing the method used in this analysis by Bruun, Allen, and Smyth (2017), ENSO patterns were modeled, which would provide valuable information for later studies to predict the influence of El Nino and La Nina. By inheriting the outcomes of previous studies, an understanding of the underlying patterns of natural phenomena can be achieved and incorporated into later model development.

## 4. Conclusion

The DFSA and the model have successfully exhibited and analysed the oscillation of plankton bloom over the years. There is no significant presence of long-term cycle resonances (more than 1 year) but only short-term cycles (1 year and less). The dominant frequency is 12 months, supported by the cyclic fluctuation every 3, 4, and 6 months. The model outcome suggests that seasonal autoregressive and linear components have no significant impact on the model performance, thus, these terms are removed without affecting the performance.

The model shows that the plankton community on the Kimberly coast has a quite stable pattern, however, could not capture some sudden plunges probably caused by natural hazards. The years in which the model failed to predict the lows of plankton coincide with years when there were severe catastrophes including floods and tropical cyclones in the region. The direct relationship between the two has not been confirmed, and there are possibly other hypotheses around the sudden change or delay in plankton productivity. Several studies have modelled and predicted the emergence of natural disasters, and solutions lie in how we can inherit those earlier studies and incorporate those predictions into our exhibition of plankton activity. Further attempts would be required to further develop this model.

**Reference**


Alheit, J. and Niquen, M. (2004) Regime shifts in the Humboldt Current ecosystem, *Progress in Oceanography*, 60(2-4):201-222

Bruun, JT., Icarus Allen, J., and Smyth, TJ. ( 2017), Heartbeat of the Southern Oscillation explains ENSO climatic resonances, *Journal of Geophysical Research: Oceans*, 122(8), pp.6746-6772.

Bureau of Meteorology - Australian Government (2012) Record-breaking La Niña events, Achieved at: http://www.bom.gov.au/climate/enso/history/La-Nina-2010-12.pdf, viewed on 28 Mar 2023.

Bureau of Meteorology - Australian Government (2002) Tropical Cyclone Chris. Available at: <http://www.bom.gov.au/cyclone/history/pdf/chris.pdf> [Accessed 15 Mar 2023]

Cheng, Y., Bhoot, VN., Kumbier, K., Sison-Mangus, MP., Brown, JB., Kudela, R. and Newcomer, ME., (2021) A novel random forest approach to revealing interactions and controls on chlorophyll concentration and bacterial communities during coastal phytoplankton blooms, *Scientific reports*, 11(1):19944.

Cheung, WWL., Dunne, J., Sarmiento, JL., and Pauly, D. (2011) Integrating ecophysiology and plankton dynamics into projected maximum fisheries catch potential under climate change in the Northeast Atlantic, *ICES Journal of Marine Science*, 68: 1008–1018.

COPEPOD Database, Available at: <https://www.st.nmfs.noaa.gov/copepod/about/about-copepod.html> [Accessed 15 Mar 2023]

Damle, C. and Yalcin, A. (2007) Flood prediction using time series data mining. *Journal of Hydrology*, 333(2-4):305-316.

Geetha, A. and Nasira, GM. (2016) Time series modeling and forecasting: Tropical cyclone prediction using ARIMA model, *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom),* New Delhi, India, 3080-3086.

Ghil, M., et al. (2002), Advanced spectral methods for climatic time series, *Reviews of geophysics*, 40(1):3-1

Hays, GC., Richardson, AJ. and Robinson, C, (2005) Climate change and marine plankton, *Trends in ecology & evolution*, 20(6):337-344.

Lebourges-Dhaussy, A., Huggett, J., Ockhuis, S., Roudaut, G., Josse, E. and Verheye, H. (2014) Zooplankton size and distribution within mesoscale structures in the Mozambique Channel: a comparative approach using the TAPS acoustic profiler, a multiple net sampler and ZooScan image analysis, *Deep Sea Research Part II: Topical S*

Li, J., et al. (2013), El Niño modulations over the past seven centuries, *Nature climate change,* 3(9):822–826.

Luo, JY., Irisson, JO., Graham, B., Guigand, C., Sarafraz, A., Mader, C. and Cowen, RK. (2018) Automated plankton image analysis using convolutional neural networks, *Limnology and Oceanography: methods*, 16(12):814-827.

Macqueen, CL. (2011) Survival, recovery and Aussie spirit-Gascoyne locals share their stories, *ABC News,* 04 Feb. Available at: <https://www.abc.net.au/local/stories/2011/02/04/3129867.htm> [Accessed 15 Mar 2023]

Mann, ME., and Lees, JM. (1996), Robust estimation of background noise and signal detection in climatic time series, *Climate Change*, 33(3):409–445.

Morrison, L. (2017) WA floods: Second man's body recovered as receding waters reveal extent of damage, *ABC News*, 14 Feb. Available at: <https://www.abc.net.au/news/2017-02-14/second-man-dies-in-wa-floods/8269992 > [Accessed 15 Mar 2023]

Racault, MF., Platt, T., Sathyendranath, S., Ağirbaş, E., Martinez Vicente, V. and Brewin, R. (2014) Plankton indicators and ocean observing systems: support to the marine ecosystem state assessment, J*ournal of Plankton Research*, 36(3):621-629.

Salmun, H. and Molod, A. (2015) The use of a statistical model of storm surge as a bias correction for dynamical surge models and its applicability along the u.s. east coast, J*ournal of Marine Science Engineering*, 3(1):73-86.

Yuan, Z., Liu, D., Masqué, P., Zhao, M., Song, X. and Keesing, JK. (2020) Phytoplankton responses to climate-induced warming and interdecadal oscillation in North-Western Australia, *Paleoceanography and Paleoclimatology*, 35(3).