Time Series Analysis of Arctic Sea Ice

Hector G. Flores Rodriguez*, Hernando Ombao†

Abstract

TODO: abstract

Keywords: Climate modeling; Spline regression analysis; Global warming; Arctic sea ice concentration.

*Department of Computer Science, University of California, Irvine, CA 92697, USA; `hfloresr@uci.edu`

†Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia; `hernando.ombao@kaust.edu.sa`

1 Introduction

scientific issues, why is this important

brief description of approach, summary of results

2 Linear Splines Trend Model

In order to determine significant changes in linear trends, let us consider a linear spline model with K change points, $\xi_1, \ldots, \xi_K$, such that they lie on the axis of abscissas and that they represent either:

- significant change in time.
- significant visual structural change in the data.

We then define a linear spline basis function with change point at ξ_k to be

$$(t-\xi_k)_+ = \begin{cases} 0, & \text{if } t < \xi_k \\ (t-\xi_k), & \text{if } t \geq \xi_k \end{cases}$$

where $(t-\xi_k)_+$ is the positive part of the function since the "+" sets the function to zero for values of t where $t-\xi_k$ is negative, illustrated in Figure 1.

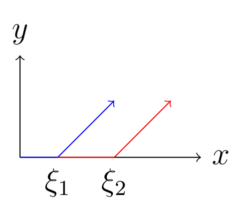

Figure 1: Linear spline basis.

To estimate the trend of sea ice concentration with respect to time, the following model is constructed

$$y_t = \mu_t + s_t + \epsilon_t \tag{1}$$

where,

- y_t is the observed dependent variable at time t
- μ_t denotes the piecewise linear trend, from previous definition of linear splines, is defined to be

$$\mu_t = \beta_0 + \beta_1 t + \sum_{k=1}^{K} b_k (t - \xi_k)_+ \tag{2}$$

- s_t is the seasonality component, which is expressed as

$$s_t = \sum_{j=1}^{m} [\alpha_j cos(2\pi\omega_j t) + \delta_j sin(2\pi\omega_j t)] \tag{3}$$

- ϵ_t is the random component that accounts for temporal correlation in $\{y_t\}$.

In equation (2), β_0, β_1, and b_k (for $k = 1, \ldots, K$) are the corresponding linear trend coefficients for each spline. Note that when $b_k \neq 0$, there is a change in the slope (linear trend) at time ξ_k.

2.1 Trend Analysis

In general, the change points, $\xi_1, \ldots, \xi_K$, are unknown. However, we initially fixed a pre-defined a set of change points and its determination is later treated as a problem of model selection. The method of iteratively reweighted least squares (IRLS) is used to estimate the piecewise linear trend coefficients. To determine our optimal set of change points, we employed a backward selection approach to our initial set of K candidate change points until all remaining change points had a significant p-value at the 5% significance level.

For further analysis, we removed the trends from our original series so that the detrended series can be expressed as

$$e_t = y_t - \hat{\mu_t}$$

where, $\hat{\mu_t}$ is estimated by using the IRLS algorithm. By removing the piecewise regression line, we can further study the periodicity and temporal correlation since

$$e_t \approx s_t + \epsilon_t$$

where e_t are the residuals after computing our estimates for the linear trend model and s_t is the seasonality component described in equation (3).

2.2 Spectral Analysis

The purpose of spectral analysis is to study oscilations present in the time series. In particular, we shall identify periodicity trends in the ice concentration. To pursue the investigation, we consider the set of harmonic frequencies

$$\omega_j = \frac{j}{T} \qquad \text{for } \ j = 1, \dots, T/2$$

were T is the number of time points in our data set. We then utilized an estimate of the power spectrum, $\hat{P}$, to identify the dominant harmonic frequencies in the time series. A simple and fast estimate of the power spectrum can be computed by the periodogram

$$\hat{P}(e^{j\omega}) = \sum_{h=-n+1}^{n-1} \hat{\gamma}(h) e^{-j\omega h}$$

however, it has been shown that the periodogram is not a consistent estimator of the power spectrum density (PSD). Instead, we utilized Welch's method

- the residuals, e_t, are split into K overlapping segments of length L
- apply Hanning window $w(n) = \frac{1}{2}(1 - \cos{(2\pi \frac{n}{N})})$ to each of the segments
- all K periodograms are averaged

$$\hat{P}_{welch}(e^{j\omega}) = \frac{1}{K} \sum_{k=1}^{K} \hat{P}_y^{(k)}(e^{j\omega}) \tag{4}$$

where,

$$\hat{P}_y^{(k)} = \frac{1}{N} \sum_{n=0}^{L-1} \left| w(n) y^{(k)}(n) e^{-j\omega n} \right|^2$$

To capture seasonal patterns, we consider the model in equation (3) where coefficients α_j and δ_j are estimated but the frequencies ω_j are obtained from the power spectrum estimate in equation (4). In our analysis, we chose the top five harmonic frequecies from the power spectrum estimate, and estimated the coefficients using the IRLS algorithm. Similarly to the trend analysis, we utilized a backward selection approach to our top five candidate harmonic frequencies until all remaining harmonics had a significant p-value at the 5% significance level.

The fitted seasonality component is then removed from our residuals, e_t, such that

$$e_t^* = e_t - \hat{s}_t$$

where, $\hat{s}_t$ is estimated via IRLS and using only the reamining significant harmonic frequencies. By removing the seasonality component, we can further investigate autoregressive (AR) models.

2.3 AR Models

A common approach for modeling univariate time series data is the autoregressive (AR) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

where y_t is the stationary time series, ϵ_t is white noise, $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$), and p denotes the order of the AR process. The purpose of AR models are based on the idea that past values might predict current observations. The AR process models y_t as a function of p past observations, $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$. To determine the order, p, of our AR model, we initially analyze the plots for the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of our detrended and deseasonalized series, e_t^*. From the ACF and PACF plots, we determine an initial set of models by examining the significance of each of the lags. We then fit the models and determine our final model according to Akaike information criterion (AIC) and Bayesian information criterion (BIC).

3 Analysis of Arctic Sea Ice Concentration

The autoregressive (AR) model is implemented on the sea ice concentration data set provided by the National Snow and Ice Data Center (NSIDC) (Walsh et al., 2015). We are interested in estimating the parameters for the changes in climate trends, seasonal patterns, and developing an AR model. Since the arctic sea ice data set spans a long range of years, various climate phenomonons that have occured in the past may greatly influence the trend in the series. The aim of our analysis is to investigate any correlation between any climate phenomenons and change points in our spline regression model and to detect whether global warming is a natural seasonal pattern or if there is any evidence that it may be human-induced.

3.1 Data Analysis

The Arctic sea ice concentration data set consists of monthly ice concentration from the beginning of January 1850 to the end of December 2013. The ice concentration are given as a percent from 0 to 100, inclusive. The spatial resolution of the monthly ice concentration are given on a quarter-degree latitude by quarter-degree longitude grid. Prior to 1979, the historical observations come in many forms: ship observations, compilations by naval oceanographers, analyses by national ice services, and others. From 1979 and onward, sea ice concentration came from a single source: satellite passive microwave data.

Our initial exploration of the data set was to aggregate the data into yearly averages to visualize how sea ice concentration changes from 1850 to 2013. As shown in Figure 2, a slight oscillation in sea ice concentration appears during the beginning of the series. However, a clear trend in decreasing in ice concentration is visible following the years after 1990.

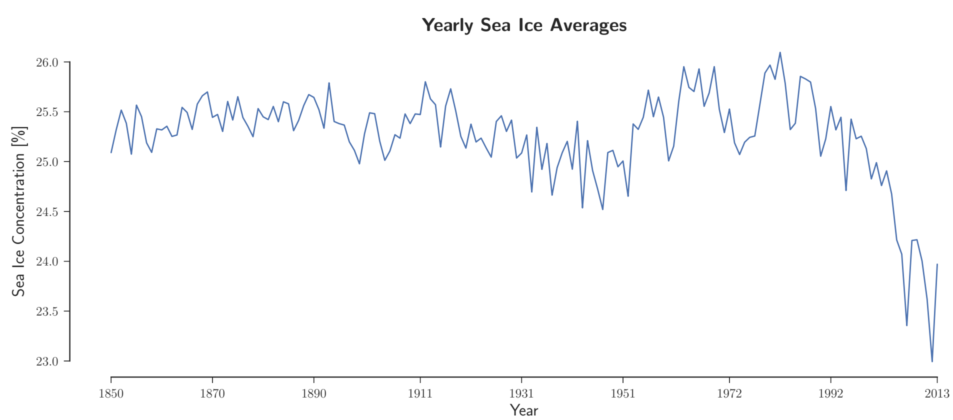

Figure 2: Yearly sea ice concentration averages.

To further examine the decreasing trend, we decoupled the series into four seasons:

- December, January, February (DJF)
- March, April, May (MAM)
- June, July, August (JJA)
- September, October, November (SON)

By seperating the seasons, we can visually examine the time series for each season, as illustrated in Figure 3. Interestingly enough, the colder seasons (DJF and MAM) seem to be visually stable. Whereas the warmer seasons (JJA and SON) capture the sharp decreasing trend following the years after 1990.

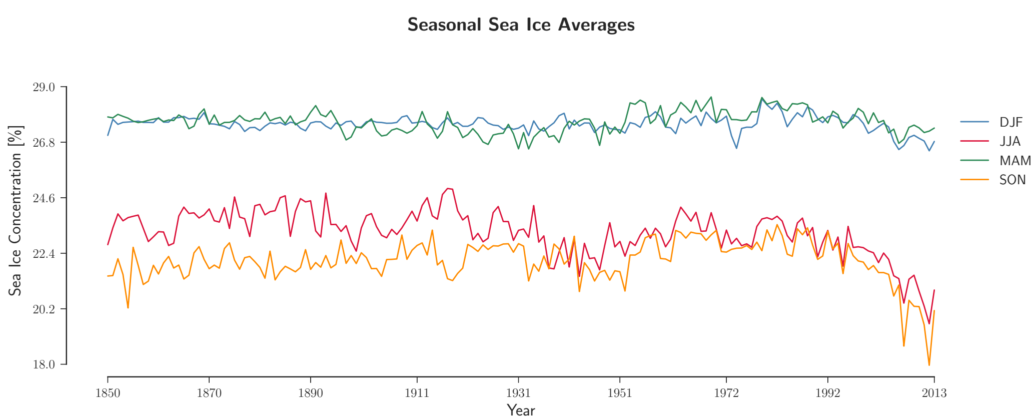

Figure 3: Seasonal sea ice concentration averages.

Furthermore, the variability is significantly greater in the months JJA and SON in comparison to DJF and MAM. With clear differences between the various series, we aim to determine where are the significant change points in each of the seasons and fit a regression model that can accurately capture the trends for each season.

3.2 Summer and Winter Trends

In Figure 3, we visually recognize interesting structural changes around the years: mid 1990's, late 1970's, 1940's, and early 1900's for each of the seasons. To determine our initial set of change points (CP), we utilized a non-parametric approach to change point detection, as outlined in (Matteson and James, 2013). As our aim is to analyze any correlation between significant climate phenomenons, we also included time points where significant heat waves have been reported. As a starting point, our initial set of change points (CP) for our linear splines trend model in equation (2), are

$$CP_{djf} = \{1935, 1944, 1979, 1997\}$$

$$CP_{mam} = \{1896, 1929, 1935, 1953, 1979, 1997\}$$

$$CP_{jja} = \{1935, 1943, 1963, 1979, 1997\}$$

$$CP_{son} = \{1915, 1950, 1979, 1997\}$$

We then allowed our greedy backward selection algorithm determine which change points contributed to a significant trend for each season. At a significance level of 5%, the resulting significant change points are shown in Table 1

Season	**Change Points**
DJF	$\{1997\}$
MAM	$\{1933, 1979\}$
JJA	$\{1997\}$
SON	$\{1915, 1950, 1979, 1997\}$

Table 1: Resulting change points for linear spline model after multiple comparison test.

Plot yearly averages, seasonal averages, and linear splines trends

3.3 Harmonic Regression

plot PSD, harmonic frequencies,

3.4 Results

ACF/PACF plots, AR models, residual diagnostics

Why are the results interesting

How can the scientist use your results to inform them on their work

Can your results be used to conirm some hypotheses?

4 Conclusion

limitations of current analysis

next steps (further analysis)

References

Matteson, D. S. and James, N. A. (2013). A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. *ArXiv e-prints.*

Walsh, J. E., Chapman, W. L., and Fetterer, F. (2015). Gridded monthly sea ice extent and concentration, 1850 onward, version 1.1. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: http://dx.doi.org/10.7265/N5833PZ5.